



(12)发明专利申请

(10)申请公布号 CN 106460046 A

(43)申请公布日 2017.02.22

(21)申请号 201580021610.2

D.W.马霍尼 G.P.利加德

(22)申请日 2015.03.26

H.T.阿拉维

(30)优先权数据

(74)专利代理机构 中国专利代理(香港)有限公司 72001

61/972942 2014.03.31 US

代理人 段菊兰 鲁炜

61/977954 2014.04.10 US

(85)PCT国际申请进入国家阶段日

(51)Int.Cl.

2016.10.31

G12Q 1/68(2006.01)

(86)PCT国际申请的申请数据

PCT/US2015/022751 2015.03.26

(87)PCT国际申请的公布数据

W02015/153284 EN 2015.10.08

(71)申请人 梅奥医学教育和研究基金会

地址 美国明尼苏达州

申请人 精密科学公司

权利要求书5页 说明书33页  
序列表13页

(72)发明人 D.A.阿尔奎斯特 W.R.泰勒

(54)发明名称

检测结直肠赘生物

(57)摘要

本文提供了涉及检测赘生物形成并且尤其，但不排他性地涉及用于检测恶变前和恶性赘生物如结直肠癌的方法、组合物和相关用途的技术。

1. 一种在得自受试者的样品中筛查结直肠赘生物的方法,所述方法包括:
  - 1) 测定得自受试者的样品中的标志物的甲基化状态;并且
  - 2) 当所述标志物的甲基化状态不同于在没有赘生物的受试者中测定的所述标志物的甲基化状态时,将所述受试者鉴定为有赘生物,  
其中所述标志物包含如表1A和B所提供的差异性甲基化区域(DMR)中的碱基。
2. 根据权利要求1所述的方法,其中所述赘生物为结直肠癌、大结直肠腺瘤和/或无蒂锯齿状息肉。
3. 根据权利要求1所述的方法,其中所述赘生物为癌前期的。
4. 根据权利要求1所述的方法,其包括测定多种标志物。
5. 根据权利要求1所述的方法,其中测定标志物的甲基化状态包括用选择性修饰所述样品中的未甲基化胞嘧啶残基以生成经修饰的残基,但是不修饰甲基化胞嘧啶残基的试剂处理获得的样品;并且  
确定已经历所述处理步骤的所述样品中的所述标志物的甲基化水平。
6. 根据权利要求1所述的方法,其包括测定2至10种标志物。
7. 根据权利要求1所述的方法,其中测定所述样品中的所述标志物的甲基化状态包括确定一个碱基的甲基化状态。
8. 根据权利要求1所述的方法,其中测定所述样品中的所述标志物的甲基化状态包括确定在多个碱基处的甲基化的程度。
9. 根据权利要求1所述的方法,其中所述标志物的甲基化状态包括相对于所述标志物的正常甲基化状态,所述标志物的甲基化增加或减少。
10. 根据权利要求1所述的方法,其中所述标志物的甲基化状态包括相对于所述标志物的正常甲基化状态,所述标志物的甲基化模式不同。
11. 根据权利要求1所述的方法,其包括测定正向链的甲基化状态或测定反向链的甲基化状态。
12. 根据权利要求1所述的方法,其中所述标志物为100个或更少碱基的区域。
13. 根据权利要求1所述的方法,其中所述标志物为500个或更少碱基的区域。
14. 根据权利要求1所述的方法,其中所述标志物为1000个或更少碱基的区域。
15. 根据权利要求1所述的方法,其中所述标志物为5000个或更少碱基的区域。
16. 根据权利要求1所述的方法,其中所述标志物为一个碱基。
17. 根据权利要求1所述的方法,其中所述标志物在高CpG密度启动子中。
18. 根据权利要求1所述的方法,其中所述样品为粪便样品、组织样品、结直肠囊肿样品、结直肠肿瘤样品、血液样品或尿液样品。
19. 根据权利要求1所述的方法,其中所述测定包括使用甲基化特异性聚合酶链反应、核酸测序、质谱法、甲基化特异性核酸酶、基于质量的分离或靶标捕获。
20. 根据权利要求1所述的方法,其中所述测定包括使用甲基化特异性寡核苷酸。
21. 一种寡核苷酸,其包含选自SEQ ID NO:1-57的序列。
22. 一种寡核苷酸,其包含与具有在DMR中的碱基的染色体区域互补的序列。
23. 一种对DMR的甲基化状态敏感的寡核苷酸。
24. 根据权利要求1所述的方法,其中具有选自NDRG4、BMP3、OPLAH、FLI1、PDGFD、

SFMBT2、CHST2、VAV3和DTX1的标注的染色体区域包含所述标志物。

25. 根据权利要求24所述的方法

其中具有所述SFMBT2标注的所述染色体区域包含选自SFMBT2\_895、SFMBT2\_896和SFMBT2\_897的一个或多个染色体区域。

其中具有所述CHST2标注的所述染色体区域包含选自CHST2\_7889和CHST2\_7890的一个或多个染色体区域。

26. 一种试剂盒,其包含:

1) 亚硫酸氢盐试剂;和

2) 对照核酸,其包含来自于选自表1A和B中提供的DMR的DMR的序列,并且具有与未患癌症的受试者相关的甲基化状态。

27. 一种试剂盒,其包含亚硫酸氢盐试剂和根据权利要求21-22中任一项所述的寡核苷酸。

28. 一种试剂盒,其包含:

1) 亚硫酸氢盐试剂;和

2) 对照核酸,其包含来自于如表1A和B中提供的DMR的序列,并且具有与患有癌症的受试者相关的甲基化状态。

29. 一种试剂盒,其包含用于从受试者获得样品的样品收集器;用于从所述样品分离核酸的试剂;亚硫酸氢盐试剂;和根据权利要求21-22中任一项所述的寡核苷酸。

30. 根据权利要求29所述的试剂盒,其中所述样品为粪便样品或结直肠组织样品。

31. 一种组合物,其包含含DMR的核酸和亚硫酸氢盐试剂。

32. 一种组合物,其包含含DMR的核酸和根据权利要求21-22中任一项所述的寡核苷酸。

33. 一种组合物,其包含含DMR的核酸和甲基化敏感性限制酶。

34. 一种组合物,其包含含DMR的核酸和聚合酶。

35. 根据权利要求34所述的组合物,其中所述组合物为反应混合物。

36. 一种在得自受试者的样品中筛查赘生物的方法,所述方法包括:

a) 确定所述样品中包含DMR中的碱基的标志物的甲基化状态,所述DMR选自如表1A和B所提供的DMR;

b) 将来自于所述受试者样品的所述标志物的甲基化状态与来自于未患癌症的受试者的正常对照样品的所述标志物的甲基化状态作比较;

c) 确定所述受试者样品和所述正常对照样品的甲基化状态的差异的置信区间和/或p值。

37. 根据权利要求36所述的方法,其中所述置信区间为90%、95%、97.5%、98%、99%、99.5%、99.9%或99.99%并且所述p值为0.1、0.05、0.025、0.02、0.01、0.005、0.001或0.0001。

38. 一种在得自受试者的样品中筛查赘生物的方法,所述方法包括使包含DMR的核酸与亚硫酸氢盐试剂反应以生成亚硫酸氢盐反应的核酸;为所述亚硫酸氢盐反应的核酸测序以提供所述亚硫酸氢盐反应的核酸的核苷酸序列;将所述亚硫酸氢盐反应的核酸的核苷酸序列与来自于未患癌症的受试者的包含所述DMR的核酸的核苷酸序列作比较以鉴定所述两个序列的差异;并且当存在差异时将所述受试者鉴定为有赘生物。

39. 一种用于在得自受试者的样品中筛查赘生物的系统,所述系统包括配置用于确定样品的甲基化状态的分析组件,配置用于比较所述样品的甲基化状态与数据库中记录的对照样品或参考样品甲基化状态的软件组件,及配置用于基于甲基化状态的组合确定单个值并且警告使用者癌症相关的甲基化状态的警报组件。

40. 根据权利要求39所述的系统,其中所述样品包含含DMR的核酸。

41. 根据权利要求39所述的系统,其还包含用于分离核酸的组件。

42. 根据权利要求39所述的系统,其还包含用于收集样品的组件。

43. 根据权利要求39所述的系统,其还包含用于收集粪便样品的组件。

44. 根据权利要求39所述的系统,其中所述数据库包含含DMR的核酸序列。

45. 根据权利要求39所述的系统,其中所述数据库包含来自于未患癌症的受试者的核酸序列。

46. 一种分离核酸的集合,每个核酸具有包含DMR的序列。

47. 根据权利要求46所述的核酸的集合,其中每个核酸具有来自于未患癌症的受试者的序列。

48. 一种系统,其包含根据权利要求46或47所述的核酸的集合和与所述核酸的集合相关的核酸序列的数据库。

49. 根据权利要求48所述的系统,其还包含亚硫酸氢盐试剂。

50. 根据权利要求48所述的系统,其还包含核酸测序仪。

51. 一种在得自受试者的样品中筛查结直肠赘生物的方法,所述方法包括:

1) 测定得自受试者的样品中的标志物的甲基化状态,

其中所述测定包括

使所述样品与甲基化特异性寡核苷酸接触,并且

确定所述标志物的甲基化状态,并且

2) 当所述标志物的甲基化状态不同于在没有赘生物的受试者中测定的所述标志物的甲基化状态时,将所述受试者鉴定为有赘生物,

其中所述标志物包含如表1A和B所提供的差异性甲基化区域(DMR)中的碱基。

52. 根据权利要求51所述的方法,其中所述赘生物为结直肠癌、大结直肠腺瘤和/或无蒂锯齿状息肉。

53. 根据权利要求51所述的方法,其中所述赘生物为癌前期的。

54. 根据权利要求51所述的方法,其包括测定多种标志物。

55. 根据权利要求51所述的方法,其包括测定2种或更多种标志物。

56. 根据权利要求51所述的方法,其包括测定2至10种标志物。

57. 根据权利要求51所述的方法,其中测定所述样品中的所述标志物的甲基化状态包括确定一个碱基的甲基化状态。

58. 根据权利要求51所述的方法,其中测定所述样品中的所述标志物的甲基化状态包括确定在多个碱基处的甲基化的程度。

59. 根据权利要求51所述的方法,其中所述标志物的甲基化状态包括相对于所述标志物的正常甲基化状态,所述标志物的甲基化增加或减少。

60. 根据权利要求51所述的方法,其中所述标志物的甲基化状态包括相对于所述标志

物的正常甲基化状态,所述标志物的甲基化模式不同。

61. 根据权利要求51所述的方法,其包括测定正向链的甲基化状态或测定反向链的甲基化状态。

62. 根据权利要求51所述的方法,其中所述标志物为100个或更少碱基的区域。

63. 根据权利要求51所述的方法,其中所述标志物为500个或更少碱基的区域。

64. 根据权利要求51所述的方法,其中所述标志物为1000个或更少碱基的区域。

65. 根据权利要求51所述的方法,其中所述标志物为5000个或更少碱基的区域。

66. 根据权利要求51所述的方法,其中所述标志物为一个碱基。

67. 根据权利要求51所述的方法,其中所述标志物在高CpG密度启动子中。

68. 根据权利要求51所述的方法,其中所述样品为粪便样品、组织样品、结直肠囊肿样品、结直肠肿瘤样品、血液样品或尿液样品。

69. 根据权利要求51所述的方法,其中所述测定包括使用甲基化特异性聚合酶链反应、核酸测序、质谱法、甲基化特异性核酸酶、基于质量的分离或靶标捕获。

70. 根据权利要求51所述的方法,其中所述甲基化特异性寡核苷酸是选自SEQ ID NO: 1-57的序列。

71. 一种检测受试者中的赘生物的方法,其包括:

a) 从受试者获得包含DNA的样品;

b) 用选择性修饰所获得的DNA中的未甲基化胞嘧啶残基以生成经修饰的残基,但是不修饰甲基化胞嘧啶残基的试剂处理所获得的DNA;

c) 确定已经历步骤b)的所述处理的所述DNA中的一个或多个DNA甲基化标志物的甲基化水平,其中一个或多个DNA甲基化标志物包含如表1A和B所提供的差异性甲基化区域(DMR)中的碱基,

d) 将确定的所述一个或多个DNA甲基化标志物的甲基化水平与没有赘生物的受试者的所述一个或多个DNA甲基化标志物的甲基化水平参考做比较,

e) 当所述DNA甲基化标志物中的或多个的甲基化状态不同于在没有赘生物的受试者中测定的所述标志物的甲基化状态时,将所述受试者鉴定为有赘生物。

72. 根据权利要求71所述的方法,其中确定所述DNA甲基化标志物中的一个或多个的甲基化升高包括确定选自CpG岛和CpG岛岸的区域内的甲基化改变。

73. 根据权利要求72所述的方法,其中确定所述CpG岛或CpG岸内的甲基化升高包括所述DNA甲基化标志物的编码区或调控区内的甲基化升高。

74. 根据权利要求71所述的方法,其中所述确定已经历步骤b)的所述处理的所述DNA中一个或多个DNA甲基化标志物的甲基化水平包括确定所述一个或多个DNA甲基化标志物的甲基化得分和/或甲基化频率。

75. 根据权利要求71所述的方法,其中步骤b)的所述处理通过对所获得的DNA的亚硫酸氢盐修饰来实现。

76. 根据权利要求71所述的方法,其中所述确定已经历步骤b)的所述处理的所述DNA中一个或多个DNA甲基化标志物的甲基化水平通过选自以下的技术实现:甲基化特异性PCR、定量甲基化特异性PCR、甲基化敏感性DNA限制酶分析、定量亚硫酸氢盐焦磷酸测序和亚硫酸氢盐基因组测序PCR。

78. 根据权利要求71所述的方法,其中所述赘生物为结直肠癌、大结直肠腺瘤和/或无蒂锯齿状息肉。

79. 一种检测受试者中的赘生物的方法,其包括:

确定来自于受试者的样品的甲基化比率,其中所述甲基化比率是相对于参考基因经亚硫酸氢盐处理的DNA拷贝数的水平,一个或多个DNA甲基化标志物经亚硫酸氢盐处理的区域的甲基化的水平,其中所述一个或多个DNA甲基化标志物包含如表1A和B所提供的差异性甲基化区域(DMR)中的碱基,其中所述参考基因为MYOD或ACTB,

当所述DNA甲基化标志物中的一种或多种的甲基化比率不同于在没有赘生物的受试者中测定的所述相应标志物的甲基化比率时,将所述受试者鉴定为有赘生物。

80. 根据权利要求79所述的方法,其中甲基化的所述水平通过使用实时聚合酶链反应(PCR)确定。

81. 根据权利要求79所述的方法,其中甲基化的所述水平通过使用实时聚合酶链反应(PCR)确定,其中所述PCR中使用的至少一种引物能够区别未甲基化和甲基化核酸。

82. 根据权利要求79所述的方法,其中甲基化的所述水平通过使用实时聚合酶链反应(PCR)确定,其中所述PCR中使用的两种引物均能够区别未甲基化和甲基化核酸。

83. 根据权利要求79所述的方法,其中甲基化的所述水平通过使用实时聚合酶链反应(PCR)确定,其中所述PCR中使用的探针能够区别未甲基化和甲基化核酸。

85. 根据权利要求79所述的方法,其中甲基化的所述水平通过使用实时聚合酶链反应(PCR)确定,其中所述PCR中使用的引物和探针两者都能够区别未甲基化和甲基化核酸。

85. 根据权利要求79所述的方法,其中所述参考基因中的所述区域的水平通过使用实时聚合酶链反应(PCR)确定。

86. 根据权利要求79所述的方法,其中所述参考基因中的所述区域的水平通过使用实时聚合酶链反应(PCR)确定,其中所述区域包含第一和第二引物结合位点和探针结合位点并且其中所述第一和第二引物结合位点和所述探针结合位点无CpG二核苷酸。

87. 根据权利要求79所述的方法,其中所述参考基因中的所述区域无CpG二核苷酸。

## 检测结直肠赘生物

### 技术领域

[0001] 本文提供了涉及检测赘生物形成并且尤其,但不排他性地涉及用于检测恶变前和恶性赘生物如结直肠癌的方法、组合物和相关用途的技术。

### [0002] 发明背景

[0003] 美国男性和女性加在一起结直肠癌仍然是第二最常见的癌症(Siegel R等人,CA Cancer J Clin 2013;63:11-30)。前期病变向癌症进展的基础生物学本身有利于筛查(Vogelstein B等人,Science 2013;339:1546-58)。有证据支持和指导方针认可了几种试验和策略中的任一种(Levin B等人,Gastroenterology 2008;134:1570-95;Rex DK等人,Am J Gastroenterol 2009;104:739-50;Karl J等人,Clin Gastroenterol Hepatol 2008;6:1122-8)。从社会的角度来看,筛查被认为是有成本效益的(Karl J等人,Clin Gastroenterol Hepatol 2008;6:1122-8;Heitman SJ等人,PLoS Med 2010;7:e1000370;Parekh M等人,Aliment Pharmacol Ther 2008;27:697-712;Sharaf RN等人,Am J Gastroenterol 2013;108:120-32)。

[0004] 结直肠癌起因于积累的遗传和表观遗传改变,为分析粪便的肿瘤特异性变化提供了基础(Berger BM等人,Pathology 2012;44:80-8)。先前在筛查环境下基于粪便的早一代DNA试验的大规模研究仅证明对结直肠癌的良好灵敏度和对晚期腺瘤的低灵敏度(Ahlquist DA等人,Ann Intern Med 2008;149:441-50,W81;Imperiale TF等人,N Engl J Med 2004;351:2704-14)。已经取得了重要进展,包括稳定缓冲液(Boynton KA等人,Clin Chem 2003;49:1058-65;Zou H等人,Cancer Epidemiol Biomarkers Prev 2006;15:1115-9),更多判别标志物(Ahlquist DA等人,Gastroenterology 2012;142:248-56;Bardan E等人,Israel journal of medical sciences 1997;33:777-80),具有更高分析灵敏度的平台(Ahlquist DA等人,Gastroenterology 2012;142:248-56;Aronchick CA等人,Gastrointestinal endoscopy 2000;52:346-52),使用逻辑回归分析而不是单独的标志物值进行结果判断,及自动化。

[0005] 虽然筛查减少了结直肠癌死亡率(Mandel JS等人,N Engl J Med.1993,328:1365-71;Hardcastle JD等人,Lancet.1996,348:1472-7;Kronborg O等人,Scand J Gastroenterol.2004,39:846-51;Winawer SJ等人,J Natl Cancer Inst.1993,85:1311-8;Singh H等人,JAMA.2006,295:2366-73),但观察到的减少是适度的(Singh H等人,JAMA.2006;295,2366-73;Heresbach D等人,Eur J Gastroenterol Hepatol.2006,18:427-33)并且在美国一半以上的成人尚未接受筛查(Meissner HI,Cancer Epidemiol Biomarkers Prev.2006,15:389-94)。

[0006] 一种新兴的癌症筛查方法牵涉对来自于癌症患者的身体样品,如粪便、血清和尿液中肿瘤特异性DNA改变的测定(Osborn NK,Ahlquist DA.Gastroenterology 2005;128:192-206;Ahlquist DA等人,Gastroenterology 2000;119:1219-27;Ahlquist DA等人,Gastroenterology 2002;122:Suppl A40;Chen WD等人,J Natl Cancer Inst 2005;97:1124-32;Zou H等人,Cancer Epidemiol Biomarkers Prev 2006;15:1115-9;Zou HZ,Clin

Cancer Res 2002;8:188-91;Hoque MO, J Clin Oncol 2005;23:6569-75;Belinsky SA等人, Cancer Res 2006;66:3338-44;Itzkowitz SH等人, Clin Gastroenterol Hepatol 2007;5:111-7; Kann L等人, Clin Chem 2006;52:2299-302)。如果要在癌症筛查应用中实现效率和有效性,重要的是以高精度选择标志物。由于结直肠赘生物形成的分子异质性,高检测率往往需要一组标志物。

[0007] 已经在来自于结直肠癌患者的粪便和血清/血浆样品中检测出几个甲基化基因(Ahlquist DA, Gastroenterology 2002;122:Suppl A40;Chen WD等人, J Natl Cancer Inst 2005;97:1124-32;Zou HZ等人, Clin Cancer Res 2002;8:188-91;Itzkowitz SH等人, Clin Gastroenterol Hepatol 2007;5:111-7;Petko Z等人, Clin Cancer Res 2005;11:1203-9;Muller HM等人, Lancet 2004;363:1283-5;Leung WK等人, Clin Chem 2004;50:2179-82;Ebert MP等人, Gastroenterology 2006;131:1418-30;Grady WM等人, Cancer Res 2001;61:900-2)。尽管在大多数结直肠癌中已经发现一些甲基化基因,但基于体液的测定法的产率仍不是最理想的(Ahlquist DA等人, Gastroenterology 2002;122:Suppl A40;Chen WD等人, J Natl Cancer Inst 2005;97:1124-32;Zou H等人, Cancer Epidemiol Biomarkers Prev 2006;15:1115-9;Zou HZ, Clin Cancer Res 2002;8:188-91;Belinsky SA等人, Cancer Res 2006;66:3338-44;Itzkowitz SH等人, Clin Gastroenterol Hepatol 2007;5:111-7;Kann L等人, Clin Chem 2006;52:2299-302;Petko Z等人, Clin Cancer Res 2005;11:1203-9;Muller HM等人, Lancet 2004;363:1283-5;Leung WK等人, Clin Chem 2004;50:2179-82;Ebert MP等人, Gastroenterology 2006;131:1418-30;Grady WM等人, Cancer Res 2001;61:900-2)。

[0008] 需要更精确、用户友好并且可广泛经销的用于提高筛查有效性、可接受性和访问的工具。

### 发明概要

[0009] 甲基化DNA已经作为大多数肿瘤类型的组织中的一类潜在生物标志物进行研究。在许多情况下,DNA甲基转移酶将甲基基团添加到DNA的胞嘧啶-磷酸-鸟嘌呤(CpG)岛位点,作为基因表达的表观遗传控制。在生物学上有吸引力的机制中,认为肿瘤抑制基因的启动子区内的获得甲基化事件会使表达沉默,从而有助于肿瘤发生。DNA甲基化可能是比RNA或蛋白质表达在化学和生物学上更稳定的诊断工具(Laird(2010) Nat Rev Genet 11:191-203)。此外,在癌症如散发性结肠癌中,甲基化标志物提供了优良特异性并且比单独的DNA突变具有更广泛的信息并且更灵敏(Zou等人(2007) Cancer Epidemiol Biomarkers Prev 16:2686-96)。

[0010] 对CpG岛的分析当应用于动物模型和人类细胞系时已经取得重要发现。例如,Zhang和同事发现,来自于相同CpG岛的不同部分的扩增子可具有不同的甲基化的水平(Zhang等人(2009) PLoS Genet 5:e1000438)。进一步地,甲基化水平在高度甲基化和未甲基化序列间双峰分布,进一步支持DNA甲基转移酶活性的二进制转换样模式(Zhang等人(2009) PLoS Genet 5:e1000438)。对鼠体内组织和体外细胞系的分析证明,仅仅约0.3%的高CpG密度启动子(HCP,定义为在300个碱基对的区域内具有>7% CpG序列)甲基化,而低CpG密度的区域(LCP,定义为在300个碱基对的区域内具有<5% CpG序列)趋向于呈动态组



织特异性模式频繁甲基化 (Meissner等人 (2008) Nature 454:766-70)。HCP包括泛在管家基因和高度调控发育基因的启动子。在HCP位点中,几个确定的标志物如Wnt 2、NDRG2、SFRP2和BMP3 > 50%甲基化 (Meissner等人 (2008) Nature 454:766-70)。

[0011] 已经在结直肠癌患者的血液和粪便中检测出甲基化基因并且建议作为候选筛查标志物 (Ahlquist DA等人, Gastroenterology 2002;122:Suppl A40; Chen WD等人, J Natl Cancer Inst 2005;97:1124-32; Zou HZ, Clin Cancer Res 2002;8:188-91; Itzkowitz SH等人, Clin Gastroenterol Hepatol 2007;5:111-7; Kann L等人, Clin Chem 2006;52:2299-302; Petko Z等人, Clin Cancer Res 2005;11:1203-9; Muller HM等人, Lancet 2004;363:1283-5; Leung WK等人, Clin Chem 2004;50:2179-82; Ebert MP等人, Gastroenterology 2006;131:1418-30; Grady WM等人, Cancer Res 2001;61:900-2)。

[0012] Zou和同事,例如,评估了在结直肠赘生物形成中频繁甲基化的基因以鉴定最具判别性的一个 (Zou等人, 2007 Cancer Epidemiol Biomarkers Prev. 16 (12):2686-2696)。在结直肠癌中特异性甲基化的四个基因 (骨形态生成蛋白3 (BMP3)、EYA2、无芒样同源框-4 (ALX4) 和波形蛋白 (vimentin)) 选自41个候选基因并且对74个癌细胞、62个腺瘤细胞和70个正常上皮细胞进行了评估。定性和定量分析甲基化状况并且通过亚硫酸氢盐基因组测序来确认。在五个结肠癌细胞系中评估了甲基化对基因表达的影响。通过测序检测K-ras和BRAF突变。分别在66%、66%、68%和72%的癌细胞, 74%、48%、89%和84%的腺瘤细胞及7%、5%、11%和11%的正常上皮细胞中检测出BMP3、EYA2、ALX4或波形蛋白的甲基化 ( $P < 0.01$ , 癌症或腺瘤对比正常细胞)。推断BMP3、EYA2、ALX4和波形蛋白基因在大多数结直肠赘生物中甲基化,但是在正常上皮细胞中很少甲基化。

[0013] 癌症筛查需要在取自受试者的样品 (例如, 粪便样品、结直肠组织样品) 中检查时,在组织水平上具有广泛信息并且对结直肠癌表现出高特异性的结直肠癌标志物或标志物组。

[0014] 因此,本文提供了结直肠癌筛查标志物的技术,在从取自受试者的样品 (例如, 粪便样品、结直肠组织样品、血清样品、血液或血液制品) 中检测时,所述标志物提供高信噪比和低本底水平。

[0015] 在开发本发明的实施方案的过程中进行的实验中,在病例对照研究中通过比较来自于有结直肠赘生物形成、腺瘤和/或无蒂锯齿状息肉 (SSP) 的受试者的结直肠组织的DNA标志物的甲基化状态与来自于对照受试者 (例如, 正常组织如正常结肠) 的相同DNA标志物的甲基化状态来鉴定标志物 (参见, 实施例1、表1A和B)。

[0016] 例如,在开发本发明的实施方案的过程中进行的实验证明NDRG4、BMP3、OPLAH、FLI1、PDGFD、SFMBT2 (例如, SFMBT2\_895、SFMBT2\_896、SFMBT2\_897)、CHST2 (例如, CHST2\_7889和CHST2\_7890)、VAV3和DTX1为检测粪便样品内的结直肠癌的有效标志物 (参见, 实施例1及表1A和B)。

[0017] 如本文所述,所述技术提供了许多对结直肠赘生物形成 (例如, 结直肠癌、腺瘤、SSP) 有高判别力的甲基化DNA标志物及其子集 (例如, 2、3、4、5、6、7、8、9、10、11、12个或更多个标志物的集合)。实验将选择过滤器应用于候选标志物以鉴定提供高信噪比和低本底水平的标志物,以例如在为了结直肠癌筛查或诊断的目的而测定远侧介质 (例如, 粪便、血液、尿液、转移组织等) 时,提供高特异性。这样,所述技术提供了用于结直肠癌筛查或诊断目的

的特异性标志物和标志物组合。

[0018] 在一些实施方案中,所述技术与评估生物样品中本文鉴定的一个或多个标志物的存在和甲基化状态有关。这些标志物包括如本文所讨论的一个或多个差异性甲基化区域(DMR),例如,如表1A和B所提供的。在所述技术的实施方案中评估甲基化状态。这样,本文提供的技术不限于测量基因的甲基化状态的方法。例如,在一些实施方案中通过基因组扫描方法测量甲基化状态。例如,一种方法牵涉限制性标记基因组扫描(Kawai等人(1994) *Mol. Cell. Biol.* 14:7421-7427)而另一个实例牵涉甲基化敏感性任意引物PCR(Gonzalzo等人(1997) *Cancer Res.* 57:594-599)。在一些实施方案中,通过用甲基化敏感性限制酶消化基因组DNA,接着进行对目标区域的Southern分析(消化-Southern方法)监测在特定CpG位点的甲基化模式的变化。在一些实施方案中,分析甲基化模式的变化牵涉基于PCR的过程,该过程牵涉在PCR扩增之前用甲基化敏感性限制酶消化基因组DNA(Singer-Sam等人(1990) *Nucl. Acids Res.* 18:687)。另外,已经报道其它技术利用亚硫酸氢盐处理DNA作为甲基化分析的起点。这些包括甲基化特异性PCR(MSP)(Herman等人(1992) *Proc. Natl. Acad. Sci. USA* 93:9821-9826)和限制性酶消化由经亚硫酸氢盐转化的DNA扩增的PCR产物(Sadri和Hornsby(1996) *Nucl. Acids Res.* 24:5058-5059;及Xiong和Laird(1997) *Nucl. Acids Res.* 25:2532-2534)。已经开发PCR技术用于基因突变的检测(Kuppuswamy等人(1991) *Proc. Natl. Acad. Sci. USA* 88:1143-1147)和等位基因特异性表达的量化(Szabo和Mann(1995) *Genes Dev.* 9:3097-3108;及Singer-Sam等人(1992) *PCR Methods Appl.* 1:160-163)。此类技术使用内部引物,其退火为PCR产生的模板并且在要测定的单核苷酸的5'立即终止。在一些实施方案中使用如美国专利第7,037,650号中所述的使用“定量Ms-SNuPE测定”的方法。

[0019] 评估甲基化状态后,常常将甲基化状态表示为在特定位点(例如,在单核苷酸处,在特定区域或基因座处,在较长目标序列处,例如达~100-bp、200-bp、500-bp、1000-bp或更长的DNA子序列)甲基化的单独DNA链相对于样品中包含该特定位点的DNA总体的分数或百分比。传统上,通过PCR使用校准器确定未甲基化核酸的量。然后,亚硫酸氢盐处理已知量的DNA并且使用实时PCR或其它指数扩增,例如QuARTS测定(例如,如美国专利第8,361,720号、美国专利第8,916,344号;和美国专利申请公布第2012/0122088和2012/0122106号中所提供的)确定所得甲基化特异性序列。

[0020] 例如,在一些实施方案中方法包括通过使用外标生成未甲基化靶标的标准曲线。标准曲线由至少两个点构建并且将未甲基化DNA的实时Ct值与已知量化标准关联起来。然后,由至少两个点和外标构建甲基化靶标的第二标准曲线。这第二标准曲线将甲基化DNA的Ct与已知量化标准关联起来。接下来,为甲基化和未甲基化群体确定试样Ct值并且由通过前两个步骤产生的标准曲线计算DNA的基因组当量。由甲基化DNA的量,相对于群体中DNA的总量,例如(甲基化DNA的数量)/(甲基化DNA的数量+未甲基化DNA的数量)×100,计算目标位点处甲基化的百分比。

[0021] 根据本发明的另一方面,当相对于参考基因经亚硫酸氢盐处理的DNA拷贝数的水平,一个或多个DNA甲基化标志物的甲基化比率不同时,表明生物样品的赘生物形成,其中所述一个或多个DNA甲基化标志物包含如表1A和B提供的差异性甲基化区域(DMR)中的碱基。甲基化比率包括通过用于确定生物标志物的甲基化水平的相同方式确定的DNA甲基化

标志物的甲基化水平和参考基因中区域水平的比率。通常,甲基化比率用通过用于确定DNA甲基化标志物的甲基化水平的相同方式确定的DNA甲基化标志物的甲基化水平和参考基因中区域水平的比率表示。

[0022] 在一些实施方案中,甲基化比率是DNA甲基化标志物的甲基化水平和参考基因的区域水平的比率,所述二者是使用实时聚合酶链反应(PCR)定量测量的。例如,可以使用一对引物和寡核苷酸探针定量测量来自于受试者样品的DNA甲基化标志物的甲基化水平,其中一个引物、两个引物、寡核苷酸探针,或两个引物和寡核苷酸探针能够区别甲基化和未甲基化核酸,例如在用修饰剂,例如将未甲基化胞嘧啶转化为经转化核酸的亚硫酸氢盐修饰核酸之后。

[0023] 本发明的参考基因的所述区域可以是基因的具有一个或多个位点的任何区域或没有甲基化位点,例如无CpG二核苷酸的区域。例如,参考基因的所述区域可以是具有两个用于扩增如PCR的引物结合位点,无CpG二核苷酸的区域或具有至少一个用于实时PCR的特异性寡核苷酸探针结合位点,无CpG二核苷酸的区域。一方面,本发明的参考基因的所述区域是MYOD基因的区域。另一方面,本发明的参考基因的所述区域是ACTB基因的区域。再一个实施方案中,本发明的参考基因的所述区域是不经常受到拷贝数改变,如基因扩增或缺失的区域。

[0024] 一般而言,根据本发明,在亚硫酸氢盐转化之后但未直接或间接判别位点的甲基化状况,用特异性结合该位点的引物和特异性探针使用实时PCR定量测量参考基因的区域水平。

[0025] 在某些实施方案中,提供了检测受试者中的赘生物的方法。此类方法包括,例如,从受试者获得包含DNA的样品;用选择性修饰获得的DNA中的未甲基化胞嘧啶残基以生成经修饰的残基,但是不修饰甲基化胞嘧啶残基的试剂处理获得的DNA;确定已经历步骤b)的处理的DNA中的一个或多个DNA甲基化标志物的甲基化水平,其中一个或多个DNA甲基化标志物包含如表1A和B所提供的差异性甲基化区域(DMR)中的碱基;将确定的所述一个或多个DNA甲基化标志物的甲基化水平与没有赘生物的受试者的所述一个或多个DNA甲基化标志物的甲基化水平参考做比较;并且当所述DNA甲基化标志物中的一个或多个的甲基化状态不同于在没有赘生物的受试者中测定的所述标志物的甲基化状态时,将所述受试者鉴定为有赘生物。

[0026] 在一些实施方案中,确定所述DNA甲基化标志物中的一个或多个的甲基化升高包括确定选自CpG岛和CpG岛岸的区域内的甲基化改变。

[0027] 在一些实施方案中,确定所述CpG岛或CpG岸内的甲基化升高包括所述DNA甲基化标志物的编码区或调控区内的甲基化升高。

[0028] 在一些实施方案中,确定已经历步骤b)的处理的所述DNA中一种或多种DNA甲基化标志物的甲基化水平包括确定所述一个或多个DNA甲基化标志物的甲基化得分和/或甲基化频率。

[0029] 在一些实施方案中,步骤b)的处理通过对获得的DNA的亚硫酸氢盐修饰来实现。

[0030] 在一些实施方案中,确定已经历步骤b)的处理的所述DNA中一个或多个DNA甲基化标志物的甲基化水平通过选自以下的技术实现:甲基化特异性PCR、定量甲基化特异性PCR、甲基化敏感性DNA限制酶分析、定量亚硫酸氢盐焦磷酸测序和亚硫酸氢盐基因组测序PCR。

[0031] 在一些实施方案中,所述赘生物为结直肠癌、大结直肠腺瘤和/或无蒂锯齿状息肉。

[0032] 在某些实施方案,提供了检测受试者中的赘生物的方法。此类方法包括,例如,确定来自于受试者的样品的甲基化比率,其中所述甲基化比率是相对于参考基因经亚硫酸氢盐处理的DNA拷贝数的水平,一个或多个DNA甲基化标志物经亚硫酸氢盐处理的区域的甲基化水平,其中所述一个或多个DNA甲基化标志物包含如表1A和B所提供的差异性甲基化区域(DMR)中的碱基,其中参考基因为MYOD或ACTB,当所述DNA甲基化标志物中的一个或多个的甲基化比率不同于在没有赘生物的受试者中测定的相应标志物的甲基化比率时,将所述受试者鉴定为有赘生物。

[0033] 在一些实施方案中,甲基化水平通过使用实时聚合酶链反应(PCR)确定。在一些实施方案中,甲基化水平通过使用实时聚合酶链反应(PCR)确定,其中PCR中使用的至少一种引物能够区别未甲基化和甲基化核酸。在一些实施方案中,甲基化水平通过使用实时聚合酶链反应(PCR)确定,其中PCR中使用的两个引物均能够区别未甲基化和甲基化核酸。在一些实施方案中,甲基化水平通过使用实时聚合酶链反应(PCR)确定,其中PCR中使用的探针能够区别未甲基化和甲基化核酸。在一些实施方案中,甲基化水平通过使用实时聚合酶链反应(PCR)确定,其中PCR中使用的引物和探针两者都能够区别未甲基化和甲基化核酸。在一些实施方案中,参考基因中的所述区域的水平通过使用实时聚合酶链反应(PCR)确定。在一些实施方案中,参考基因中的所述区域的水平通过使用实时聚合酶链反应(PCR)确定,其中所述区域含有第一和第二引物结合位点和探针结合位点并且其中所述第一和第二引物结合位点和所述探针结合位点无CpG二核苷酸。在一些实施方案中,参考基因中的所述区域无CpG二核苷酸。

[0034] 本文还提供了用于实践所述方法的组合物和试剂盒。例如,在一些实施方案中,对一个或多个标志物有特异性的试剂(例如,引物、探针)单独地或呈集合(例如,用于扩增多个标志物的引物对集合)提供。也可提供用于进行检测测定的附加试剂(例如,用于进行QuARTS、PCR、测序、亚硫酸氢盐处理或其它测定的酶、缓冲液、阳性和阴性对照)。在一些实施方案中,提供含有对于进行一种方法所必需、足够或有用的一种或多种试剂的试剂盒。还提供含有所述试剂盒的反应混合物。进一步提供含有多种可相互添加和/或添加到试样中以使反应混合物完全的试剂的主混合试剂集合。

[0035] 在一些实施方案中,本文描述的技术与设计用于进行一系列通过本文描述的方法所提供的算术或逻辑操作的可编程机器相关。例如,所述技术的一些实施方案与计算机软件和/或计算机硬件相关(例如,在其中实现)。一方面,所述技术涉及包括一种形式的存储器、用于进行算术和逻辑操作的元件和用于执行一系列指令(例如,如本文所提供的方法)的处理元件(例如,微处理器)以读取、操纵和存储数据的计算机。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,例如表1A和B所提供的一个或多个DMR的甲基化状态);比较甲基化状态(例如,例如表1A和B所提供的一个或多个DMR的甲基化状态);生成标准曲线;确定Ct值;计算甲基化的分数、频率或百分比(例如,例如表1A和B所提供的一个或多个DMR的甲基化);鉴定CpG岛;确定测定法或标志物的特异性和/或灵敏度;计算ROC曲线和相关AUC;序列分析;全部如本文所述或在本领域中已知。

[0036] 在一些实施方案中,软件或硬件组件接收多次测定的结果并且基于多次测定的结

果确定指示癌症风险的单值结果向用户报告(例如,确定例如表1A和B所提供的多个DMR的甲基化状态)。有关实施方案基于来自于多次测定的结果的数学组合(例如,加权组合、线性组织),例如确定多种标志物(例如表1A和B所提供的多个DMR)的甲基化状态来计算危险系数。在一些实施方案中,DMR的甲基化状态限定一个维度并且可在多维空间上具有值并且由多个DMR的甲基化状态限定的坐标是例如要向用户报告,例如与结直肠癌风险有关的结果。

[0037] 一些实施方案包括存储介质和存储器组件。存储器组件(例如,易失性和/或非易失性存储器)用于存储指令(例如,本文所提供的方法的实施方案)和/或数据(例如,工件如甲基化测量、序列和与之相关的统计学描述)。一些实施方案涉及还包括CPU、图形卡 and 用户界面(例如,包括输出设备如显示器和输入设备如键盘)中的一种或多种的系统。

[0038] 与所述技术相关的可编程机器包括传统现存技术和正在开发或尚待开发的技术(例如,量子计算机、化学计算机、DNA计算机、光学计算机、基于自旋电子学的计算机等)。

[0039] 在一些实施方案中,所述技术包括用于传输数据的有线(例如,金属电缆、光纤)或无线传输介质。例如,一些实施方案涉及经过网络(例如,局域网(LAN)、广域网(WAN)、ad-hoc网络、互联网等)的数据传输。在一些实施方案中,可编程机器作为同级存在于此类网络上并且在一些实施方案中可编程机器具有客户端/服务器关系。

[0040] 在一些实施方案中,数据存储于计算机可读存储介质中,如硬盘、闪存存储器、光学介质、软盘等。

[0041] 在一些实施方案中,本文提供的技术与多个共同操作以进行如本文所述的方法的可编程设备相关。例如,在一些实施方案中,多台计算机(例如,通过网络连接)并行工作以收集和处理数据,例如在执行集群计算或网格计算中或在依赖于通过常规网络接口,如以太网(Ethernet)、光纤或通过无线网络技术连接到网络(专用、公用网络或互联网)的完整计算机(具有机载CPU、存储器、电源、网络接口等)的一些其它分布式计算机体系结构中。

[0042] 例如,一些实施方案提供了包括计算机可读介质的计算机。该实施方案包括与处理器耦合的随机存取存储器(RAM)。处理器执行存储器中存储的计算机可执行的程序指令。此类处理器可包括微处理器、ASIC、状态机或其它处理器,并且可以是许多计算机处理器中的任一种,如来自于Santa Clara的Intel Corporation、Schaumburg, Illinois的California and Motorola Corporation的处理器。此类处理器包括介质,或可与介质连通,例如计算机可读介质,其存储当由处理器执行时,使处理器进行本文描述的步骤的指令。

[0043] 计算机可读介质的实例包括但不限于电子、光、磁或其它存储器或能够为处理器提供计算机可读指令的传输设备。合适介质的其它实例包括但不限于软盘、CD-ROM、DVD、磁盘、存储器芯片、ROM、RAM、ASIC、已配置的处理器、所有光学介质、所有磁带或其它磁性介质或计算机处理器可以从中读取指令的任何其它介质。同样,各种其它形式的计算机可读介质也可向计算机传输或携带指令,包括有线和无线的路由器、专用或公用网络,或其它传输设备或通道。指令可包含来自任何合适的计算机编程语言,包括例如C、C++、C#、Visual Basic、Java、Python、Perl和JavaScript的代码。

[0044] 在一些实施方案中计算机连接到网络。计算机还可包括许多外部或内部设备如鼠标、CD-ROM、DVD、键盘、显示器或其它输入或输出设备。计算机的实例为个人计算机、数字助理、个人数字助理、蜂窝式电话、移动电话、智能电话、寻呼机、数字平板、膝上型计算机、网

络家电和其它基于处理器的设备。一般而言,与本文提供的技术的方方面面有关的计算机可以是在能够支持包含本文提供的技术的一种或多种程序的任何操作系统,如Microsoft Windows、Linux、UNIX、Mac OS X等上操作的任何类型的基于处理器的平台。一些实施方案包括执行其它应用程序(例如,应用软件)的个人计算机。应用软件可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、即时通讯应用程序、演示应用程序、互联网浏览器应用程序、日历/管理器应用程序和客户端设备能够执行的任何其它应用程序。

[0045] 本文描述为与所述技术相关的所有此类组件、计算机和系统可为逻辑性或虚拟的。

[0046] 因此,本文提供了与一种在得自受试者(例如,人类受试者)的样品(例如,粪便样品、结直肠组织样品、血液样品、血液制品样品)中筛查结直肠赘生物的方法有关的技术,所述方法包括测定得自受试者的样品中的标志物的甲基化状态;并且当所述标志物的甲基化状态不同于在没有结直肠赘生物的受试者中测定的所述标志物的甲基化状态时,将所述受试者鉴定为有结直肠赘生物,其中所述标志物包含选自如表1A和B所提供的DMR的差异性甲基化区域(DMR)中的碱基。所述技术还涵盖确定结直肠癌的状态或阶段,例如,在一些实施方案中赘生物为癌前期的。一些实施方案提供了包括测定多种标志物,例如包括测定2至10种标志物的方法。

[0047] 所述技术不限于评估甲基化状态。在一些实施方案中评估样品中的标志物的甲基化状态包括确定一个碱基的甲基化状态。在一些实施方案中,测定样品中的标志物的甲基化状态包括确定在多个碱基处的甲基化程度。而且,在一些实施方案中标志物的甲基化状态包括相对于所述标志物的正常甲基化状态,所述标志物的甲基化增加。在一些实施方案中,标志物的甲基化状态包括相对于所述标志物的正常甲基化状态,所述标志物的甲基化减少。在一些实施方案中,标志物的甲基化状态包括相对于所述标志物的正常甲基化状态,所述标志物的甲基化模式不同。

[0048] 此外,在一些实施方案中标志物为100个或更少碱基的区域,标志物为500个或更少碱基的区域,标志物为1000个或更少碱基的区域,标志物为5000个或更少碱基的区域,或在一些实施方案中,标志物为一个碱基。在一些实施方案中标志物在高CpG密度启动子中。

[0049] 所述技术不受相同类型限制。例如,在一些实施方案中样品为粪便样品、组织样品、结直肠组织样品、血液样品(例如,血浆、血清、全血)、排泄物或尿液样品。

[0050] 此外,所述技术不限于用于确定甲基化状态的方法中。在一些实施方案中所述测定包括使用甲基化特异性聚合酶链反应、核酸测序、质谱法、甲基化特异性核酸酶、基于质量的分离或靶标捕获。在一些实施方案中,所述测定包括使用甲基化特异性寡核苷酸。在一些实施方案中,所述技术使用用于确定甲基化状态的大规模平行测序(例如,下一代测序),例如通过合成测序、实时(例如,单分子)测序、微珠乳液测序、纳米孔测序等。

[0051] 所述技术提供了用于检测DMR的试剂,例如在一些实施方案中提供了包含SEQ ID NO:1-57提供的序列的寡核苷酸集合。在一些实施方案中提供了包含与具有在DMR中的碱基的染色体区域互补的序列的寡核苷酸,例如对DMR的甲基化状态敏感的寡核苷酸。

[0052] 所述技术提供了多组标志物,例如在一些实施方案中,所述标志物包含具有为NDRG4、BMP3、OPLAH、FLI1、PDGFD、CHST\_7889、SFMBT2\_895、SFMBT2\_896、SFMBT2\_897、CHST2\_

7890、VAV3和DTX1的标注并且包含标志物的染色体区域(参见,表1A和B)。

[0053] 另外,实施方案提供了分析来自于表1A和B的DMR的方法。一些实施方案提供了确定标志物的甲基化状态,其中具有为NDRG4、BMP3、OPLAH、FLI1、PDGFD、CHST\_7889、SFMBT2\_895、SFMBT2\_896、SFMBT2\_897、CHST2\_7890、VAV3和DTX1的标注的染色体区域包含标志物。

[0054] 提供了试剂盒实施方案,例如,一种试剂盒包含亚硫酸氢盐试剂;和对照核酸,其包含来自于选自表1A和B中提供的任何染色体区域的DMR的序列,并且具有与未患癌症的受试者相关的甲基化状态。在一些实施方案中,试剂盒包含亚硫酸氢盐试剂和如本文所述的寡核苷酸。在一些实施方案中,试剂盒包含亚硫酸氢盐试剂;和对照核酸,其包含来自于选自如表1A和B中提供的任何染色体区域的DMR的序列,并且具有与患有结直肠癌、腺瘤和/或SSP的受试者相关的甲基化状态。一些试剂盒实施方案包含用于从受试者获得样品(例如,粪便样品)的取样器;用于从所述样品分离核酸的试剂;亚硫酸氢盐试剂;和如本文所述的寡核苷酸。

[0055] 所述试剂盒与组合物(例如,反应混合物)的实施方案有关。在一些实施方案中提供了一种组合物,其包含含DMR的核酸和亚硫酸氢盐试剂。一些实施方案提供了一种组合物,其包含含DMR的核酸和如本文所述的寡核苷酸。一些实施方案提供了一种组合物,其包含含DMR的核酸和甲基化敏感性限制酶。一些实施方案提供了一种组合物,其包含含DMR的核酸和聚合酶。

[0056] 提供了另外的有关方法实施方案以在得自受试者的样品中筛查结直肠赘生物,例如一种方法包括确定所述样品中包含DMR中的碱基的标志物的甲基化状态,所述DMR选自如表1A和B所提供的任何染色体区域;将来自于所述受试者样品的标志物的甲基化状态与来自于未患癌症的受试者的正常对照样品的标志物的甲基化状态作比较;确定所述受试者样品和所述正常对照样品的甲基化状态上的差异的置信区间和/或p值。在一些实施方案中,置信区间为90%、95%、97.5%、98%、99%、99.5%、99.9%或99.99%并且p值为0.1、0.05、0.025、0.02、0.01、0.005、0.001或0.0001。方法的一些实施方案提供了以下步骤:使包含DMR的核酸与亚硫酸氢盐试剂反应以生成亚硫酸氢盐反应核酸;为所述亚硫酸氢盐反应核酸测序以提供所述亚硫酸氢盐反应核酸的核苷酸序列;将所述亚硫酸氢盐反应核酸的核苷酸序列与来自于未患癌症的受试者的包含所述DMR的核酸的核苷酸序列作比较以鉴定所述两个序列上的差异;并且当存在差异时将所述受试者鉴定为有结直肠赘生物。

[0057] 通过所述技术提供了用于在得自受试者的样品中筛查结直肠赘生物的系统。系统的示范性实施方案包括例如,一种用于在得自受试者的样品中筛查结直肠赘生物的系统,所述系统包括配置用于确定样品的甲基化状态的分析组件,配置用于比较所述样品的甲基化状态与数据库中记录的对照样品或参考样品甲基化状态的软件组件,及配置用于警告使用者癌症相关的甲基化状态的警报组件。在一些实施方案中,由软件组件接收来自于多次测定(例如,确定例如表1A和B所提供的多个标志物例如DMR的甲基化状态)的结果并且基于多个结果计算要报告的值或结果来确定警报。一些实施方案提供了与本文提供的每个DMR相关,用于计算要向使用者(例如,医师、护士、临床医生等)报告的值或结果和/或警报的加权参数的数据库。在一些实施方案中报告来自于多次测定的所有结果并且在一些实施方案中基于来自于多次测定的表明受试者中的结直肠癌风险的一个或多个结果的合成,使用一个或多个结果来提供得分、值或结果。

[0058] 在系统的一些实施方案中,样品包含含DMR的核酸。在一些实施方案中,所述系统还包含用于分离核酸的组件,用于采集样品的组件,如用于采集粪便样品的组件。在一些实施方案中,所述系统包含含DMR的核酸序列。在一些实施方案中,数据库包含来自于未患癌症的受试者的核酸序列。还提供了核酸,例如核酸的集合,每个核酸具有包含DMR的序列。在一些实施方案中,核酸的集合,其中每个核酸具有来自于未患癌症的受试者的序列。有关系统实施方案包含所描述的核酸的集合和与核酸的集合相关的核酸序列的数据库。一些实施方案还包含亚硫酸氢盐试剂。并且,一些实施方案还包含核酸测序仪。

[0059] 所述技术与组合物(例如,反应混合物)的实施方案有关。在一些实施方案中提供了一种组合物,其包含含DMR(例如,表1A和B所提供的DMR)的核酸和亚硫酸氢盐试剂。一些实施方案提供了一种组合物,其包含含DMR的核酸和如本文所述的寡核苷酸。一些实施方案提供了一种组合物,其包含含DMR的核酸和甲基化敏感性限制酶。一些实施方案提供了一种组合物,其包含含DMR的核酸和聚合酶。

[0060] 基于本文所含的教导,对于相关领域的技术人员而言另外的实施方案将显而易见。

[0061] 发明详述

[0062] 本文提供了涉及检测结直肠赘生物形成并且尤其,但不排他性地涉及用于检测恶变前和恶性结直肠癌的方法、组合物和相关用途的技术。本文描述所述技术时,所用章节标题仅仅是为了组织的目的而不得解释为以任何方式限制主题。

[0063] 在各个实施方案的这种详细描述中,出于解释的目的,阐述了许多具体细节以提供对公开的实施方案的透彻理解。然而,本领域的技术人员将认识到,可实践这些各个实施方案,需要或不需这些具体细节。在其它情况下,结构和设备以方框图形式示出。此外,本领域的技术人员可以容易地认识到,方法呈现和进行的具体顺序是说明性的并且考虑到该顺序可以改变并且仍然在本文公开的各个实施方案的精神和范围之内。

[0064] 本申请中引用的所有文献和类似材料,包括但不限于专利、专利申请、文章、书籍、论文和互联网网页明确地通过引用整体并入用于任何目的。除非另有定义,本文使用的所有技术和科学术语具有本文所述各个实施方案所属领域中的普通技术人员通常所理解的含义。并入的参考文献中的术语定义出现与本教导中提供的定义不同时,应以本教导中提供的定义为准。

[0065] 定义

[0066] 为利于对本技术的理解,下面定义了许多术语和短语。另外的定义在发明详述通篇有阐述。

[0067] 在说明书和权利要求书通篇,除非上下文另外明确指出,否则以下术语采用本文明确相关的含义。如本文中所用的短语“在一个实施方案中”尽管可能,但不一定是指相同实施方案。此外,如本文中所用的短语“在另一个实施方案中”尽管可能,但不一定是指不同实施方案。因此,如下所述,在不背离本发明的范围或精神的前提下,可容易地组合本发明的各个实施方案。

[0068] 另外,如本文中所用,除非上下文另外明确指出,否则术语“或”是包含性“或”运算符并且与术语“和/或”等效。除非上下文另外明确指出,否则术语“基于”是非排他性并且允许基于未描述的另外的因素。另外,在说明书通篇,“一”、“一种(个)”和“所述(该)”的含义



包括复数指示物。“在……中”的含义包括“在……中”和“在……上”。

[0069] 如本文中所示，“核酸”或“核酸分子”通常是指任何核糖核酸或脱氧核糖核酸，其可为未修饰或经修饰的DNA或RNA。“核酸”包括但不限于单链和双链核酸。如本文中所示，“核酸”还包括如以上所描述的含有一个或多个修饰碱基的DNA。因此，出于稳定性或其它原因骨架经修饰的DNA为“核酸”。术语“核酸”当其用于本文中时包括此类经化学、酶或代谢修饰的核酸形式，以及病毒和细胞，包括例如单细胞和复杂细胞所特有的DNA的化学形式。

[0070] 术语“寡核苷酸”或“多核苷酸”或“核苷酸”或“核酸”是指具有两个或更多个脱氧核糖核苷酸或核糖核苷酸，优选三个以上并且通常十个以上的分子。确切大小将取决于许多因素，这又取决于寡核苷酸的最终功能或用途。寡核苷酸可以任何方式生成，包括化学合成、DNA复制、反转录或其组合。DNA的典型脱氧核糖核苷酸为胸腺嘧啶、腺嘌呤、胞嘧啶和鸟嘌呤。RNA的典型核糖核苷酸为尿嘧啶、腺嘌呤、胞嘧啶和鸟嘌呤。

[0071] 如本文中所示，术语核酸的“基因座”或“区域”是指核酸的亚区，例如染色体上的基因、单核苷酸、CpG岛等。

[0072] 术语“互补”和“互补性”是指按碱基配对原则相关联的核苷酸（例如，1个核苷酸）或多核苷酸（例如，核苷酸的序列）。例如，序列5'-A-G-T-3'与序列3'-T-C-A-5'互补。互补性可为“部分的”，其中仅一些核酸碱基根据碱基配对原则匹配。或者，核酸之间可存在“完全”或“全部”互补性。核酸链之间的互补性程度影响核酸链之间杂交的效率和强度。这在扩增反应和依赖核酸之间的结合的检测方法中特别重要。

[0073] 术语“基因”是指包含生成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。功能多肽可以由全长编码序列或该编码序列的任一部分编码，只要保持多肽的所需活性或功能性质（例如，酶活性、配体结合、信号转导等）。术语“部分”在关于基因使用时是指该基因的片段。片段在尺寸上范围可从几个核苷酸到整个基因序列减去一个核苷酸。因此，“包含基因的至少一部分的核苷酸”可包含该基因的片段或整个基因。

[0074] 术语“基因”还涵盖结构基因的编码区并且包括5'和3'末端位于编码区附近的序列，例如任一端上约1kb的距离，使得该基因对应于全长mRNA（例如，包含编码、调控、结构和其它序列）的长度。位于编码区的5'并且存在于mRNA上的序列称为5'非翻译或未翻译序列。位于编码区的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cDNA和基因的基因组形式两种。在一些生物体（例如，真核生物）中，基因的基因组形式或克隆物含有被称为“内含子”或“插入区”或“插入序列”的非编码序列间断的编码区。内含子是转录为核RNA（hnRNA）的基因片段；内含子可含调控元件如增强子。从核或初级转录产物去除或“剪掉”内含子；因此在信使RNA（mRNA）转录产物中不存在内含子。mRNA在翻译期间起到指定新生多肽中氨基酸的序列或顺序的功能。

[0075] 除含有内含子以外，基因的基因组形式还可包括位于RNA转录产物上存在的序列的5'和3'末端的序列。这些序列称为“侧翼”序列或区（这些侧翼序列位于mRNA转录产物上存在的非翻译序列的5'或3'）。5'侧翼区可含调控序列如启动子和增强子，其控制或影响基因的转录。3'侧翼区可含指导转录的终止、转录后裂解和聚腺苷酸化的序列。

[0076] 术语“野生型”关于基因使用时是指具有分离自天然存在的来源的基因的特征的基因。术语“野生型”关于基因产物使用时是指具有分离自天然存在的来源的基因产物的特征的基因产物。术语“天然存在的”应用于物体时是指物体可以在自然界中找到的事实。例

如,生物体(包括病毒)中存在的可以从自然界中的来源分离并且尚未在实验室经人工刻意修饰的多肽或多核苷酸序列是天然存在的。野生型基因常常是在群体中最频繁观察到的基因或等位基因并且因此被任意指定为基因的“正常”或“野生型”形式。相反,术语“经修饰”或“突变”关于基因或基因产物使用时分别指在与野生型基因或基因产物相比时,在序列和/或功能性质上展现出改变(例如,特征改变)的基因或基因产物。值得注意的是,天然存在的突变体可以分离;这些通过它们在与野生型基因或基因产物相比时具有改变的特征的事实来鉴定。

[0077] 术语“等位基因”是指基因的变异;变异包括但不限于变异体和突变体、多态性基因座和单核苷酸多态性基因座、移码和剪接突变。等位基因可天然出现在群体中或其可能在该群体的任何特定个体生存期间产生。

[0078] 因此,术语“变异体”和“突变体”关于核苷酸序列使用时是指与另一个通常有关的核苷酸序列有一个或多个核苷酸不同的核酸序列。“变异”是两个不同核苷酸序列之间的差异;通常,一个序列为参考序列。

[0079] “扩增”是牵涉模板特异性的核酸复制的特例。与非特异性模板复制(例如,依赖模板,但不依赖特异性模板的复制)完全不同。此处将模板特异性与复制精确性(例如,正确多核苷酸序列的合成)和核苷酸(核糖或脱氧核糖)特异性区别开。经常根据“靶”特异性来描述模板特异性。靶序列在旨在将其从另一核酸中分选出来的意义上来说是“靶”。为了这样分选出来已主要设计了扩增技术。

[0080] 核酸的扩增通常是指生成多个拷贝的多核苷酸或多核苷酸的一部分,该部分通常从少量的多核苷酸(例如,单个多核苷酸分子,10至100个拷贝的多核苷酸分子,其可能或可能不完全相同),其中扩增产物或扩增子通常可检测。多核苷酸的扩增涵盖多个化学和酶促过程。在聚合酶链反应(PCR)或连接酶链反应(LCR;参见例如,美国专利第5,494,810号)期间由一个或几个拷贝的靶或模板DNA分子生成多个DNA拷贝是扩增形式。另外的扩增类型包括但不限于等位基因特异性PCR(美国专利第5,639,611号)、组装PCR(美国专利第5,965,408号)、解旋酶依赖性扩增(美国专利第7,662,594号)、热启动PCR(美国专利第5,773,258和5,338,671号)、序列间特异性PCR、反向PCR(Triglia等人(1988) *Nucleic Acids Res.*, 16:8186)、连接介导的PCR(Guilfoyle,R.等人,*Nucleic Acids Research*,25:1854-1858(1997);美国专利第5,508,169号)、甲基化特异性PCR(Herman等人,(1996) *PNAS* 93(13) 9821-9826)、微型引物PCR、多重连接依赖性探针扩增(Schouten等人,(2002) *Nucleic Acids Research* 30(12):e57)、多重PCR(Chamberlain等人,(1988) *Nucleic Acids Research* 16(23) 11141-11156;Ballabio等人,(1990) *Human Genetics* 84(6) 571-573;Hayden等人,(2008) *BMC Genetics* 9:80)、巢式PCR、重叠延伸PCR(Higuchi等人,(1988) *Nucleic Acids Research* 16(15) 7351-7367)、实时PCR(Higuchi等人,(1992) *Biotechnology* 10:413-417;Higuchi等人,(1993) *Biotechnology* 11:1026-1030)、反转录PCR(Bustin,S.A.(2000) *J.Molecular Endocrinology* 25:169-193)、固相PCR、热不对称嵌套PCR和降落PCR(Don等人,*Nucleic Acids Research*(1991) 19(14) 4008;Roux,K.(1994) *Biotechniques* 16(5) 812-814;Hecker等人,(1996) *Biotechniques* 20(3) 478-485)。多核苷酸扩增也可以使用数字PCR完成(Kalinina等人,*Nucleic Acids Research*.25:1999-2004,(1997);Vogelstein和Kinzler,*Proc Natl Acad Sci USA*.96:9236-41,(1999);国际

专利公布第W005023091A2号;美国专利申请公布第20070202525号)。

[0081] 如本文中所用,术语“核酸检测测定法”是指确定目标核酸的核苷酸组成的任何方法。核酸检测测定法包括但不限于DNA测序法、探针杂交法、结构特异性裂解测定法(例如,INVADER测定法,Hologic,Inc.)并且例如在美国专利第5,846,717、5,985,557、5,994,069、6,001,567、6,090,543和6,872,816号;Lyamichev等人,Nat.Biotech.,17:292(1999),Hall等人,PNAS,USA,97:8272(2000)和US 2009/0253142)中有描述;聚合酶链反应;分支杂交法(例如,Chiron,美国专利第5,849,481、5,710,264、5,124,246和5,624,802号);滚环复制(例如,美国专利第6,210,884、6,183,960和6,235,502号);NASBA(例如,美国专利第5,409,818号);分子信标技术(例如,美国专利第6,150,097号);循环探针技术(例如,美国专利第5,403,711、5,011,769和5,660,988号);连接酶链反应(例如,Barnay Proc.Natl.Acad.Sci USA 88,189-93(1991));QuARTS测定法(例如,如美国专利第8,361,720;及美国专利申请公布第2012/0122088和2012/0122106号中所提供);和夹心杂交法(例如,美国专利第5,288,609号)。

[0082] 术语“引物”是指如同在纯化限制消耗中天然存在的或经合成产生的寡核苷酸,其在处于诱导与核酸链互补的引物延伸产物合成的条件下时,(例如,在核苷酸和诱导剂如DNA聚合酶的存在下并且在合适的温度和pH),能够起到合成起始点的作用。为了扩增的最高效率,引物优选为单链,但是可选地可为双链。若为双链,在用于制备延伸产物之前首先处理引物以将其链分离。优选地,因为为寡脱氧核糖核苷酸。引物必须足够长以在诱导剂的存在下引发延伸产物的合成。引物的确切长度将取决于许多因素,包括温度、引物来源和所述方法的使用。

[0083] 术语“探针”是指如同在纯化限制消耗中天然存在的或经合成、重组或通过PCR扩增产生的寡核苷酸,其能够与另一目标寡核苷酸杂交。探针可为单链或双链。探针用于特定基因序列的检测、鉴定和分离(例如,“捕获探针”)。考虑到,在一些实施方案中,本发明中使用的任何探针均可用任何“报告分子”标记,使其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法)、荧光、放射性和发光系统。并非旨在将本发明限于任何特定检测系统或标记。

[0084] 如本文中所用,“甲基化”是指胞嘧啶的位置C5或N4、腺嘌呤的N6位置的胞嘧啶甲基化,或其它类型的核酸甲基化。因为典型体外DNA扩增方法不保持扩增模板的甲基化模式,所以体外扩增的DNA通常未甲基化。然而,“未甲基化DNA”或“甲基化DNA”也可以指原始模板分别未甲基化或甲基化的扩增DNA。

[0085] 因此,如本文中所用,“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,其中甲基部分在公认的典型核苷酸碱基中不存在。例如,胞嘧啶在其嘧啶环上不含甲基部分,但是5-甲基胞嘧啶在其嘧啶环的位置5处含有甲基部分。因此,胞嘧啶不是甲基化核苷酸而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位置5处含有甲基部分;然而,为了本文的目的,当存在于DNA中时不将胸腺嘧啶视为甲基化核苷酸,因为胸腺嘧啶是DNA的典型核苷酸碱基。

[0086] 如本文中所用,“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。

[0087] 如本文中所用,核酸分子的“甲基化状态”、“甲基化性质”和“甲基化状况”是指在

核酸分子中存在或不存在一个或多个甲基化核苷酸碱基。例如,将含有甲基化胞嘧啶的核酸分子视为甲基化(例如,核酸分子的甲基化状态为甲基化)。将不含任何甲基化核苷酸的核酸分子视为未甲基化。

[0088] 特定核酸序列(例如,如本文所述的基因标志物或DNA区域)的甲基化状态可以指示该序列中每个甲基的甲基化状态或可以指示该序列中碱基(例如,一个或多个胞嘧啶)子集的甲基化状态,或可以指示关于该序列中区域甲基化密度的信息,提供或不提供该序列中发生甲基化的位置的信息。

[0089] 核酸分子中核苷酸基因座的甲基化状态是指在核酸分子内的特定基因座存在或不存在甲基化核苷酸。例如,当核酸分子中第7个核苷酸处存在的核苷酸为5-甲基胞嘧啶时,核酸分子中第7个核苷酸处的胞嘧啶的甲基化状态为甲基化。类似地,当核酸分子中第7个核苷酸处存在的核苷酸为胞嘧啶(而非5-甲基胞嘧啶)时,核酸分子中第7个核苷酸处的胞嘧啶的甲基化状态为未甲基化。

[0090] 甲基化状况可以任选地用“甲基化值”表示或指示(例如,表示甲基化频率、分数、比率、百分比等)。可以产生甲基化值,例如通过量化在用甲基化依赖性限制酶进行限制消化之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图谱或通过比较经亚硫酸氢盐处理和未处理的核酸的序列。因此,值例如甲基化值,表示甲基化状况并且因此可以用作基因座多个拷贝的甲基化状况的量化指标。当期望将样品中序列的甲基化状况与阈值或参考值做比较时,这特别有用。

[0091] 如本文中所用,“甲基化频率”或“甲基化百分比(%)”是指相对于分子或基因座未甲基化的情况的数量,分子或基因座甲基化的情况的数量。

[0092] 这样,甲基化状态描述核酸(例如,基因组序列)甲基化的状态。另外,甲基化状态是指在特定基因座处的核酸片段与甲基化有关的特征。此类特征包括但不限于,这个DNA序列内的任何胞嘧啶(C)残基是否甲基化,甲基化C残基的位置,遍及核酸任何特定区域的甲基化C的频率或百分比,及由于例如等位基因起源的差异引起的甲基化的等位差异。术语“甲基化状态”、“甲基化性质”和“甲基化状况”还指遍及生物样品中核酸的任何特定区域的相对浓度、绝对浓度或甲基化C或未甲基化C的模式。例如,如果核酸序列中的胞嘧啶(C)残基甲基化,则可称其为“高甲基化”或具有“增加的甲基化”,而如果DNA序列内的胞嘧啶(C)残基未甲基化,则可称其为“低甲基化”或具有“减少的甲基化”。同样,如果核酸序列内的胞嘧啶(C)残基与另一核酸序列(例如,来自于不同区域或来自于不同个体等)相比甲基化,则认为该序列与另一核酸序列相比高甲基化或具有增加的甲基化。可选地,如果DNA序列内的胞嘧啶(C)残基与另一核酸序列(例如,来自于不同区域或来自于不同个体等)相比未甲基化,则认为该序列与另一核酸序列相比低甲基化或具有减少的甲基化。另外,如本文中所述的术语“甲基化模式”是指在核酸的一个区域上甲基化和未甲基化核苷酸的集中位点。当整个区域上甲基化和未甲基化核苷酸的数量相同或相似,但甲基化和未甲基化核苷酸的位置不同时,两个核酸可具有相同或相似的甲基化频率或甲基化百分比,但是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具有增加或减少的甲基化)、频率或模式上不同时,称序列为“差异性甲基化”或具有“甲基化上的差异”或具有“不同的甲基化状态”。术语“差异性甲基化”是指与癌症阴性样品中核酸甲基化的水平或模式相比,癌症阳性样品中核酸甲基化的水平或模式上的差异。也可以指在术后癌症复发的患者与没有复

发的患者之间在水平或模式上的差异。差异性甲基化和DNA甲基化的特定水平或模式是预后和预测生物标志,例如,曾经已经定义了恰当的截止或预测特征。

[0093] 甲基化状态频率可用于描述一群个体或来自于单个个体的样品。例如,具有50%甲基化状态频率的核苷酸基因座在50%的情况下甲基化而在50%的情况下未甲基化。此类频率可用于,例如描述在一群个体或一堆核酸中核苷酸基因座或核酸区域甲基化的程度。因此,当第一群或批核酸分子中的甲基化不同于第二群或批核酸分子中的甲基化时,第一群或批的甲基化状态频率将不同于第二群或批的甲基化状态频率。此类频率也可用于,例如描述在单个个体中核苷酸基因座或核酸区域甲基化的程度。例如,此类频率可用于描述来自于组织样品的一组细胞在核苷酸基因座或核酸区域处甲基化或未甲基化的程度。

[0094] 如本文中所示“核苷酸基因座”是指核苷酸在核酸分子中的位置。甲基化核苷酸的核苷酸基因座是指甲基化核苷酸在核酸分子中的位置。

[0095] 通常,人DNA的甲基化发生在二核苷酸序列上,包括相邻鸟嘌呤和胞嘧啶,其中胞嘧啶位于鸟嘌呤的5' (也称为CpG二核苷酸序列)。在人类基因组中CpG二核苷酸内的大多数胞嘧啶是甲基化的,然而在称为CpG岛的特定CpG二核苷酸富集基因组区域中一些保持未甲基化(参见例如,Antequera等人(1990) *Cell* 62:503-514)。

[0096] 如本文中所示,“CpG岛”是指基因组DNA的G:C富集区,其含有相对于全基因组DNA数量增加的CpG二核苷酸。CpG岛长度可为至少100、200个或更多个碱基对,其中该区域的G:C含量为至少50%并且观测到的CpG频率与预期频率的比率为0.6;在一些情况下,CpG岛长度可为至少500个碱基对,其中该区域的G:C含量为至少55%)并且观测到的CpG频率与预期频率的比率为0.65。观测到的高于预期频率的CpG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,观测到的高于预期频率的CpG频率可以根据公式 $R = (A \times B) / (C \times D)$ 计算,其中R是观测到的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是分析序列中核苷酸的总数,C是分析序列中C核苷酸的总数,并且D是分析序列中G核苷酸的总数。通常确定在CpG岛中,例如在启动子区的甲基化状态。尽管将认识到人类基因组中的其它序列有DNA甲基化倾向,如CpA和CpT(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)。

[0097] 如本文中所示,根据核酸分子的甲基化状态修饰核酸分子的核苷酸的试剂,或甲基化特异性试剂,是指可以按影响核酸分子的甲基化状态的方式改变核酸分子的核苷酸序列的化合物或组合物或其它药剂。用此类试剂处理核酸分子的方法可以包括使核酸分子与该试剂接触,若需要,连同另外的步骤来实现核苷酸序列的所需变化。核酸分子的核苷酸序列上的此类变化可以产生其中每个甲基化核苷酸均被修饰为不同核苷酸的核酸分子。核酸核苷酸序列上的此类变化可以产生其中每个未甲基化核苷酸均被修饰为不同核苷酸的核酸分子。核酸核苷酸序列上的此类变化可以产生其中选定的每个未甲基化的核苷酸(例如,每个未甲基化的胞嘧啶)均被修饰为不同核苷酸的核酸分子。使用此类试剂改变核酸核苷酸序列可以产生其中为甲基化核苷酸的每个核苷酸(例如,每个甲基化胞嘧啶)均被修饰为不同核苷酸的核酸分子。如本文中所示,使用修饰选定核苷酸的试剂是指修饰核酸分子中

通常存在的四种核苷酸(对于DNA而言为C、G、T和A并且对于RNA而言为C、G、U和A)中的一种核苷酸的试剂,以致该试剂修饰所述一种核苷酸,而不修饰另外三种核苷酸。在一个示例性实施方案中,此类试剂修饰未甲基化的选定核苷酸以产生不同核苷酸。在另一个示例性实施方案中,此类试剂可以使未甲基化的胞嘧啶核苷酸脱氨基。示例性试剂为亚硫酸氢盐。

[0098] 如本文中所用,术语“亚硫酸氢盐试剂”是指在一些实施方案中包含亚硫酸氢盐、二亚硫酸盐、氢亚硫酸盐或其组合用于区别例如CpG二核苷酸序列中的甲基化和未甲基化胞苷的试剂。

[0099] 术语“甲基化测定法”是指用于确定核酸序列中一个或多个CpG二核苷酸序列的甲基化状态的任何测定法。

[0100] 术语“MS AP-PCR”(甲基化敏感性随机引物聚合酶链反应)是指允许使用CG富集引物进行基因组全局扫描以集中在最有可能含CpG二核苷酸的区域并且经Gonzalzo等人(1997) *Cancer Research* 57:594-599描述的本领域公认的技术。

[0101] 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。

[0102] 术语“HeavyMethyl™”是指其中覆盖介于扩增引物之间或被其覆盖的CpG位置的甲基化特异性阻断探针(本文也称为阻断物)使得核酸样品的甲基化特异性选择性扩增成为可能的测定法。

[0103] 术语“HeavyMethyl™ MethyLight™”测定法是指HeavyMethyl™ MethyLight™测定法,这是MethyLight™测定法的变型,其中将MethyLight™测定法与覆盖介于扩增引物之间的CpG位置的甲基化特异性阻断探针相结合。

[0104] 术语“Ms-SNuPE”(甲基化敏感性单核苷酸引物延伸)是指Gonzalzo和Jones(1997) *Nucleic Acids Res.* 25:2529-2531描述的本领域公认的测定法。

[0105] 术语“MSP”(甲基化特异性PCR)是指Herman等人(1996) *Proc. Natl. Acad. Sci. USA* 93:9821-9826和美国专利第5,786,146号描述的本领域公认的甲基化测定法。

[0106] 术语“COBRA”(联合亚硫酸氢盐限制分析)是指Xiong和Laird(1997) *Nucleic Acids Res.* 25:2532-2534描述的本领域公认的甲基化测定法。

[0107] 术语“MCA”(甲基化CpG岛扩增)是指Toyota等人(1999) *Cancer Res.* 59:2307-12和W000/26401A1描述的甲基化测定法。

[0108] 如本文中所用,“选定核苷酸”是指核酸分子中通常存在的四种核苷酸(对于DNA而言为C、G、T和A并且对于RNA而言为C、G、U和A)中的一种核苷酸,并且可以包括通常存在的核苷酸的甲基化衍生物(例如,当C为选定核苷酸时,在选定核苷酸的含义上包括甲基化和未甲基化C两者),而甲基化选定核苷酸特指甲基化的通常存在的核苷酸并且未甲基化选定核苷酸特指未甲基化的通常存在的核苷酸。

[0109] 术语“甲基化特异性限制酶”或“甲基化敏感性限制酶”是指根据其识别位点的甲基化状态选择性消化核酸的酶。在如果识别位点未甲基化或半甲基化,则特异性切割的限制酶的情况下,如果识别位点甲基化,则不会发生切割或将以显著降低的效率发生。在如果识别位点甲基化,则特异性切割的限制酶的情况下,如果识别位点未甲基化,则不会发生切割或将以显著降低的效率发生。优选甲基化特异性限制酶,其识别序列含有CG二核苷酸(举例来说诸如CGCG或CCCGGG的识别序列)。对于一些实施方案而言进一步优选如果这种二核

苷酸中的胞嘧啶在碳原子C5处甲基化,则不切割的限制酶。

[0110] 如本文中所用,“不同核苷酸”是指在化学性质上不同于选定核苷酸的核苷酸,通常使得不同核苷酸具有不同于选定核苷酸的沃森-克里克(Watson-Crick)碱基配对性质,由此与选定核苷酸互补的通常存在的核苷酸和与不同核苷酸互补的通常存在的核苷酸不相同。例如,当C为选定核苷酸时,U或T可为不同核苷酸,这由C与G的互补性和U或T与A的互补性为例。如本文中所用,与选定核苷酸互补或与不同核苷酸互补的核苷酸是指在高度严格条件下,以比互补核苷酸与通常存在的四种核苷酸中的三种碱基配对更高的亲和力,与选定核苷酸或不同核苷酸碱基配对的核苷酸。互补的实例为DNA(例如,A-T和C-G)和RNA(例如,A-U和C-G)中的沃森-克里克碱基配对。因此,例如,在高度严格条件下,G与C碱基配对的亲和力比G与G、A或T碱基配对更高并且,因此,当C为选定核苷酸时,G是与选定核苷酸互补的核苷酸。

[0111] 如本文中所用,指定标志物的“敏感性”是指报告高于区别赘生物和非赘生物样品的阈值的DNA甲基化值的样品百分比。在一些实施方案中,阳性定义为组织学确认的报告高于阈值(例如,与疾病相关的范围)的DNA甲基化值的赘生物形成,而假阴性定义为组织学确认的报告低于阈值(例如,与无疾病相关的范围)的DNA甲基化值的赘生物形成。因此,敏感性值反映得自己知患病样品的指定标志物的DNA甲基化测量将会在疾病相关测量范围内的可能性。如此处所定义,计算的敏感性值的临床相关性表示对指定标志物在应用于具有临床病状的受试者时将检测出该病状存在的可能性的估计。

[0112] 如本文中所用,指定标志物的“特异性”是指报告低于区别赘生物和非赘生物样品的阈值的DNA甲基化值的赘生物样品的百分比。在一些实施方案中,阴性定义为组织学确认的报告低于阈值(例如,与无疾病相关的范围)的DNA甲基化值的非赘生物样品,而假阳性定义为组织学确认的报告高于阈值(例如,与疾病相关的范围)的DNA甲基化值的非赘生物样品。因此,特异性值反映得自己知非赘生物样品的指定标志物的DNA甲基化测量将会在疾病相关测量范围内的可能性。如此处所定义,计算的特异性值的临床相关性表示对指定标志物在应用于没有临床病状的受试者时将检测出该病状不存在的可能性的估计。

[0113] 如本文中所述的术语“AUC”是“曲线下面积”的缩写。具体而言是指接受者操作特征(ROC)曲线下的面积。ROC曲线是对于诊断试验的不同可能分割点的真阳性率相对于假阳性率的绘图。其根据选定分割点示出了敏感性和特异性之间的平衡(任何敏感性上的增加都将伴随特异性的降低)。ROC曲线下面积(AUC)是对诊断试验的度量(面积越大越好;最佳为1;随机试验将具有位于对角线上的面积为0.5的ROC曲线;参考:J.P.Egan.(1975) Signal Detection Theory and ROC Analysis,Academic Press,New York)。

[0114] 如本文中所用,术语“赘生物”是指“生长超过正常组织的生长并且与之不协调的异常组织物质”。参见例如,Willis RA,“The Spread of Tumors in the Human Body”,London,Butterworth&Co,1952。

[0115] 如本文中所用,术语“腺瘤”是指腺起源的良性肿瘤。虽然这些生长是良性的,但随时间推移它们可能进展变成恶性的。

[0116] 术语“癌前期”或“赘生物形成前”及其等效术语是指正在经历恶性转化的任何细胞增生性病征。

[0117] 赘生物、腺瘤、癌症等的“部位”是受试者身体中赘生物、腺瘤、癌症等所在的组织、

器官、细胞类型、解剖区域、身体部分等。

[0118] 如本文中所示，“诊断”试验应用包括检测或鉴定受试者的疾病状态或情况，确定受试者会患上指定疾病或病状的可能性，确定有疾病或病状的受试者将对治疗反应的可能性，确定对有疾病或病状的受试者的预后(或其可能进展或消退)，及确定治疗对有疾病或病状的受试者的作用。例如，诊断可用于检测受试者患上赘生物的存在性或可能性或此类受试者将顺利地化合物(例如，药品，例如药物)或其它治疗反应的可能性。

[0119] 术语“标志物”，如本文中所示，是指能够通过区别癌细胞与正常细胞，例如基于其甲基化状态，诊断癌症的物质(例如，核酸或核酸的区域)。

[0120] 术语“分离的”关于核酸使用时，如同在“分离的寡核苷酸”中，是指核酸序列经鉴定并且和一般在其天然来源中与之缔合的至少一种污染核酸分开。分离的核酸呈不同于其在自然界中的形式或情形存在。相反，非分离的核酸，如DNA和RNA，呈其在自然界中存在的状态被发现。非分离的核酸的实例包括：在宿主细胞染色体上相邻基因的附近发现的指定DNA序列(例如，基因)；RNA序列，如编码特定蛋白质的特定mRNA序列，在细胞中作为与编码大量蛋白质的许多其它mRNA的混合物被发现。然而，编码特定蛋白质的分离的核酸包括，举例而言，一般表达该蛋白质的细胞中的此类核酸，其中核酸处于和天然细胞不同的染色体位置，或另外侧接与在自然界中所发现的不同的核酸序列。分离的核酸或寡核苷酸可以呈单链或双链形式存在。要利用分离的核酸或寡核苷酸表达蛋白质时，寡核苷酸将在最低限度上含有有义或编码链(即，寡核苷酸可为单链)，但可含有义链和反义链两者(即，寡核苷酸可为双链)。分离的核酸在从其天然或典型环境中分离之后，可与其它核酸或分子组合。例如，分离的核酸可存在于宿主细胞中，核酸已经位于其中，例如用于异源表达。

[0121] 术语“纯化的”是指分子，核酸或氨基酸序列是从其天然环境中取出的、分离的或分开的。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%不含且更优选至少90%不含与之天然缔合的其它组分。如本文中所示，术语“纯化的”或“纯化”还指从样品中去除污染物。污染蛋白质的去除导致样品中目标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；从而样品中重组多肽的百分比增加。

[0122] 术语包含指定多核苷酸序列或多肽的“组合物”广义上是指含有指定多核苷酸序列或多肽的任何组合物。所述组合物可包含含有盐(例如NaCl)、去垢剂(例如SDS)和其它组分(例如，Denhardt溶液、奶粉、鲑鱼精DNA等)的水溶液。

[0123] 术语“样品”在其最广义上使用。在某种意义上它可以指动物细胞或组织。在另一种意义上，意在包括从任何来源获得的样本或培养物，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)中获得并且涵盖液体、固体、组织和气体。环境样品包括环境材料如地表物质、土壤、水和工业样品。不得将这些实例解释为对适用于本发明的样品类型的限制。

[0124] 如本文中所示，“远程样品”当在一些情况下使用时涉及从不是样品的细胞、组织或器官来源的部位间接采集的样品。例如，在粪便样品(例如，不是来自于直接取自结肠组织的样品)中评估源自结肠或直肠的样品材料时，该样品为远程样品。

[0125] 如本文中所示，术语“患者”或“受试者”是指经受通过所述技术提供的各种试验的生物体。术语“受试者”包括动物，优选哺乳动物，包括人。在一个优选实施方案中，受试者为



灵长类动物。在一个甚至更加优选的实施方案中,受试者为人。

[0126] 如本文中所示,术语“试剂盒”是指用于递送材料的任何递送系统。在反应测定法的情况下,此类递送系统包括允许储存、运输或递送反应试剂(例如,适当容器中的寡核苷酸、酶等)和/或支撑材料(例如,缓冲液、进行该测定法的书面说明书等)从一个位置到另一位置的系统。例如,试剂盒包括一个或多个含有相关反应试剂和/或支撑材料的外壳(例如,盒子)。如本文中所示,术语“分体试剂盒”是指包含两个或更多个独立容器的递送系统,每个容器含有总试剂盒组分的子部分。可将容器一起或单独递送给预定接收者。例如,第一容器可含有用于测定的酶,而第二容器含有寡核苷酸。术语“分体试剂盒”旨在涵盖含有受联邦食品、药物和化妆品法案(Federal Food, Drug, and Cosmetic Act)第520(e)章监管的分析物特定试剂(ASR)的试剂盒,但不限于此。事实上,在术语“分体试剂盒”中包括包含两个或更多个独立容器的任何递送系统,每个容器含有总试剂盒组分的子部分。相反,“组合试剂盒”是指在单个容器中(例如,在封装每种所需组分的单个盒子中)含有反应测定的所有组分的递送系统。术语“试剂盒”包括分体和组合试剂盒两者。

[0127] 技术实施方案

[0128] 总的来说,胃肠癌占有比任何其它器官系统更高的癌症死亡率。结直肠癌(CRC)是第二最致命性的癌症,国内每年>600,000人死亡。虽然在美国对结直肠癌进行筛查,但是考虑到结肠镜检查的成本、不适和侵入性及当前非侵入性粪便血液试验项目的性能低下,遵从性很差。为了减轻CRC对个体和社会的负担,需要既有效且患者友好的新试验策略。利用广泛的信息生物标志物的非侵入性、精确分子试验可提供合理的方法。基于粪便的测定法就是一个这样的实例。结直肠癌和前期癌使细胞和DNA进入消化流中并且最终在粪便中排出。高度灵敏的测定法已用于检测患有CRC癌症和癌前息肉的患者粪便中的遗传和表观遗传标志物两者。在最近的多中心研究中,将这些标志物并入粪便测定法中,粪便测定法对CRC表现出与结肠镜检查基本上等同的灵敏度。新标志物如果经证实比现有标志物更灵敏、更具特异性或更能预测病变部位,则其价值将增加。此外,当适用于筛查应用时除CRC外,标志物将理想地检测出关键的癌前病变-腺瘤性息肉和锯齿状息肉。

[0129] 结直肠癌根本的基因组基质牵涉基因和染色体异常,包括单碱基突变、异倍性、缺失、易位、拷贝数改变和表达变化。所有这些事情正使用更新的基因组技术,包括大规模平行测序进行深入研究。然而,证明遗传改变是非常不均匀的,并且在一些方面是随机的,而不是复发性的。例如,APC基因是CRC病变中突变最多的基因(~90%),但是突变位点遍布15个编码外显子,使全基因分析成为必要。这增加了开发具有所需性能水平的测定法的复杂性。

[0130] DNA在胞嘧啶-磷酸-鸟嘌呤(CpG)岛位点通过DNA甲基转移酶的表现遗传甲基化已经作为大多数肿瘤类型的组织中一类潜在的生物标志物进行研究。在生物学上有吸引力的机制中,认为肿瘤抑制基因的启动子区内的获得甲基化事件会使表达沉默,从而有助于肿瘤发生。DNA甲基化可能是比RNA或蛋白质表达在化学和生物学上更稳定的诊断工具。此外,异常甲基化标志物比单独的DNA突变具有更广泛的信息并且更灵敏并且提供了优良特异性。

[0131] 临床应用有高度判别力的标志物可具有很大影响。例如,对远侧介质如粪便或血液中的此类标志物的测定可用于精确筛查或诊断测定中以检测结直肠赘生物形成。

[0132] 在开发本发明的实施方案的过程中进行的实验中,在病例对照研究中通过比较来自于有结直肠赘生物形成、腺瘤和/或无蒂锯齿状息肉(SSP)的受试者的结直肠组织的DNA标志物的甲基化状态与来自于对照受试者(例如,正常组织如正常结肠)的相同DNA标志物的甲基化状态来鉴定标志物(参见,实施例3、表1A和B)。

[0133] 在开发本发明的实施方案的过程中进行的另外的实验证明NDRG4、BMP3、OPLAH、FLI1、PDGFD、CHST\_7889、SFMBT2\_895、SFMBT2\_896、SFMBT2\_897、CHST2\_7890、VAV3和DTX1为用于检测粪便样品中的结直肠癌的有效标志物(参见,实施例1及表1A和B)。

[0134] 因此,本文提供了结直肠癌筛查标志物的技术,在从取自受试者的样品(例如,粪便样品、结直肠组织样品)中检测时,所述标志物提供高信噪比和低本底水平。在病例对照研究中通过比较来自于有结直肠赘生物形成和/或腺瘤的受试者的结直肠组织的DNA标志物的甲基化状态与来自于对照受试者(例如,正常组织如正常结肠)的相同DNA标志物的甲基化状态来鉴定标志物(参见,实施例1及表1A和B)。

[0135] 虽然本文的公开提到某些说明的实施方案,但应理解这些实施方案是以举例的方式而不是以限制的方式提出。

[0136] 在特定方面中,本技术提供了用于鉴定、确定结直肠癌和/或为结直肠癌分类的组合物和方法。在一些实施方案中,所述方法包括确定分离自受试者的生物样品(例如,粪便样品或结直肠组织样品)中至少一种甲基化标志物的甲基化状况,其中所述标志物的甲基化状态变化表明结直肠癌的存在、类别或部位。特定实施方案涉及包含差异性甲基化区域(DMR,例如表1A和B中提供的DMR)的标志物,其用于诊断或筛查新生细胞增生性疾病(例如,结直肠癌),包括在疾病的癌前期阶段早期检测。此外,所述标志物用于区分新生与良性细胞增生性疾病。在特定方面中,本技术提供了一种方法,其中将新生细胞增生性疾病与良性细胞增生性疾病区分开。

[0137] 本技术的标志物在检测和区分结直肠增生性疾病中特别有效,从而提供了结直肠癌早期检测、分类和治疗的改进手段。

[0138] 除其中分析本文提供的至少一个标志物、标志物的区域或包含DMR(例如表1A和B中提供的DMR)的标志物的碱基的甲基化分析的实施方案外,所述技术还提供了多组对结直肠癌检测有实用性的标志物,其包含至少一个标志物、标志物的区域或包含DMR的标志物的碱基。除其中分析本文提供的至少一个标志物、标志物的区域或包含DMR(例如表1A和B中提供的DMR)的标志物的碱基的甲基化分析的实施方案外,所述技术还提供了多组对结直肠癌检测(例如,表达标志物、DNA的量、肽、血红蛋白等)有实用性的标志物,其包含任何类型或类别的标志物(例如,完整标志物、标志物的区域、标志物的碱基等)。

[0139] 所述技术的一些实施方案基于对至少一个标志物、标志物的区域或包含DMR的标志物的碱基的CpG甲基化状况的分析。

[0140] 在一些实施方案中,本技术提供亚硫酸氢盐技术连同一种或多种甲基化测定法用于确定包含DMR(例如表1A和B中提供的DMR)的至少一个标志物中CpG二核苷酸的甲基化状况的用途。基因组CpG二核苷酸可甲基化或未甲基化(可选地分别称为高和低甲基化)。然而,在远程样品(例如,血液、器官排放物或粪便)的背景下,本发明的方法适于分析异源性的生物样品,例如低浓度的肿瘤细胞,或来自其中的生物材料。因此,当分析此类样品内的CpG位置的甲基化状况时,可以使用定量测定法确定在特定CpG位置的甲基化水平(例如,百

分比、分散、比率、比例或程度)。

[0141] 对包含DMR的标志物中CpG二核苷酸序列的甲基化状况的确定在结直肠癌的诊断和表征中具有实用性。

[0142] 标志物的组合

[0143] 在一些实施方案中,所述技术涉及评估包含来自表1A和B的DMR的标志物(例如,2、3、4、5、6、7、8、9、10个)的组合或包含DMR的标志物的甲基化状态。在一些实施方案中,评估一个以上标志物的甲基化状态增加了筛查或诊断以鉴定受试者中的结直肠癌生物的特异性和/或敏感性。除其中分析本文提供的至少一个标志物、标志物的区域或包含DMR(例如表1A和B中提供的DMR)的标志物的碱基的甲基化分析的实施方案外,所述技术还提供了多组对结直肠癌检测(例如,表达标志物、DNA的量、肽、血红蛋白等)有实用性的标志物,其包含任何类型或类别的标志物(例如,完整标志物、标志物的区域、标志物的碱基等)。

[0144] 通过各种标志物的组合预测各种癌症,例如,如同通过与预测特异性和灵敏度有关的统计技术鉴定那样。所述技术提供了鉴定一些癌症的预测性组合和经验证的预测性组合的方法。

[0145] 测定甲基化状态的方法

[0146] 分析核酸中5-甲基胞嘧啶的存在的方法基于Frommer等人描述的用于检测DNA或其各种变型中的5-甲基胞嘧啶的亚硫酸氢盐方法(Frommer等人(1992) Proc.Natl.Acad.Sci.USA 89:1827-31)。为5-甲基胞嘧啶绘图的亚硫酸氢盐方法基于以下观测结果,胞嘧啶,而非5-甲基胞嘧啶与氢亚硫酸根离子(也称为亚硫酸氢盐)反应。在一些实施方案中,反应根据以下步骤进行:首先,胞嘧啶与氢亚硫酸盐反应形成磺化胞嘧啶。接着,磺化反应中间产物自发脱氨产生磺化尿嘧啶。最后,磺化尿嘧啶在碱性条件下去磺化形成尿嘧啶。因为尿嘧啶与腺嘌呤形成碱基对(因此表现类似胸腺嘧啶),而5-甲基胞嘧啶与鸟嘌呤形成碱基对(因此表现类似胞嘧啶),所以检测是可能的。这使得判别甲基化胞嘧啶与非甲基化胞嘧啶成为可能,例如通过亚硫酸氢盐基因组测序(Grigg G和Clark S, Bioessays(1994) 16:431-36;Grigg G,DNA Seq.(1996) 6:189-98)或例如美国专利第5,786,146号中公开的甲基化特异性PCR(MSP)。

[0147] 一些常规技术与包括将待分析的DNA封入琼脂糖基质中,从而防止DNA扩散和复性(亚硫酸氢盐仅与单链DNA反应),并且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)的方法有关。因此可能分析单个细胞的甲基化状况,说明了所述方法的实用性和敏感性。Rein,T.等人(1998) Nucleic Acids Res.26:2255提供了对检测5-甲基胞嘧啶的传统方法的综述。

[0148] 亚硫酸氢盐技术通常牵涉继亚硫酸氢盐处理之后扩增已知核酸的短特定片段,然后通过测序来测定产物(Olek和Walter(1997) Nat.Genet.17:275-6)或进行引物延伸反应(Gonzalgo和Jones(1997) Nucleic Acids Res.25:2529-31;WO 95/00669;美国专利第6,251,594号)以分析单个胞嘧啶位置。一些方法使用酶消化法(Xiong和Laird(1997) Nucleic Acids Res.25:2532-4)。在本领域中还已描述了通过杂交检测(Olek等人,WO 99/28498)。另外,已描述了亚硫酸氢盐技术对于单个基因的甲基化检测的用途(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)。

[0149] 各种甲基化测定程序是本领域已知的并且根据本技术可连同亚硫酸氢盐处理一起使用。这些测定法允许确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。除其它技术以外,此类测定法牵涉经亚硫酸氢盐处理的核酸的测序,PCR(对于序列特异性扩增而言)、Southern印迹分析和使用甲基化敏感性限制酶。

[0150] 例如,已经通过使用亚硫酸氢盐处理简化基因组测序用于分析甲基化模式和5-甲基胞嘧啶分布(Frommer等人(1992)Proc.Natl.Acad.Sci.USA 89:1827-1831)。另外,限制酶消化由亚硫酸氢盐转化的DNA扩增的PCR产物可用于评估甲基化状态,例如如Sadri和Hornsby(1997)Nucl.Acids Res.24:5058-5059所述或如称为COBRA(联合亚硫酸氢盐限制分析)的方法中所体现那样(Xiong和Laird(1997)Nucleic Acids Res.25:2532-2534)。

[0151] COBRA™分析是用于确定少量基因组DNA中特定基因座处的DNA甲基化水平的定量甲基化测定法(Xiong和Laird,Nucleic Acids Res.25:2532-2534,1997)。简言之,使用限制酶消化揭示经亚硫酸氢钠处理的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的亚硫酸氢盐处理向基因组DNA中引入甲基化依赖性序列差异。然后使用对目标CpG岛有特异性的引物进行经亚硫酸氢盐转化的DNA的PCR扩增,接着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检测。原DNA样品中的甲基化水平用在广泛的DNA甲基化水平范围呈线性定量方式,消化和未消化PCR产物的相对量表示。另外,这种技术可容易地适用于从显微切割的石蜡包埋组织样品获得的DNA。

[0152] 用于COBRA™分析的典型试剂(例如,可能在基于COBRA™的典型试剂盒中找到)可包括但不限于:特定基因座(例如,特定基因、标志物、DMR、基因区域、标志物区域、经亚硫酸氢盐处理的DNA序列、CpG岛等)的PCR引物;限制酶和适当缓冲液;基因杂交寡核苷酸;对照杂交寡核苷酸;寡核苷酸探针的激酶标记试剂盒;和经标记的核苷酸。另外,亚硫酸氢盐转化试剂可包括:DNA变性缓冲液、磺化缓冲液、DNA回收试剂或试剂盒(例如,沉淀、超滤、亲和柱)、去磺化缓冲液和DNA回收组分。

[0153] 优选地,测定法如“MethyLight™”(一种基于荧光的实时PCR技术)(Eads等人,Cancer Res.59:2302-2306,1999)、Ms-SNuPE™(甲基化敏感性单核苷酸引物延伸)反应(Gonzalzo和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;美国专利第5,786,146号)和甲基化CpG岛扩增(“MCA”;Toyota等人,Cancer Res.59:2307-12,1999)单独或结合这些方法中的一种或多种使用。

[0154] “HeavyMethyl™”测定技术是基于经亚硫酸氢盐处理的DNA的甲基化特异性扩增,评估甲基化差异的定量方法。覆盖介于扩增引物之间或被其覆盖的CpG位置的甲基化特异性阻断探针(“阻断物”)使得核酸样品的甲基化特异性选择性扩增成为可能。

[0155] 术语“HeavyMethyl™ MethyLight™”测定法是指HeavyMethyl™ MethyLight™测定法,这是MethyLight™测定法的变型,其中将MethyLight™测定法与覆盖介于扩增引物之间的CpG位置的甲基化特异性阻断探针相结合。HeavyMethyl™测定法也可结合甲基化特异性扩增引物使用。

[0156] 用于HeavyMethyl™分析的典型试剂(例如,可能在基于MethyLight™的典型试剂盒中找到)可包括但不限于:特定基因座(例如,特定基因、标志物、DMR、基因区域、标志物区域、经亚硫酸氢盐处理的DNA序列、CpG岛或经亚硫酸氢盐处理的DNA序列或CpG岛等)的PCR引物;阻断寡核苷酸;优化PCR缓冲液和脱氧核苷酸;及Taq聚合酶。

[0157] MSP(甲基化特异性PCR)允许评估CpG岛内几乎任一组CpG位点的甲基化状况,不依赖于使用甲基化敏感性限制酶(Herman等人Proc.Natl.Acad.Sci.USA 93:9821-9826, 1996;美国专利第5,786,146号)。简言之,DNA经亚硫酸氢钠修饰,亚硫酸氢钠将未甲基化而非甲基化的胞嘧啶转化为尿嘧啶,并且随后用相对于未甲基化DNA而言对甲基化DNA有特异性的引物扩增产物。MSP仅需要少量的DNA,对指定CpG岛基因座的0.1%甲基化等位基因灵敏,并且可以对从石蜡包埋样品提取的DNA进行。用于MSP分析的典型试剂(例如,可能在基于MSP的典型试剂盒中找到)可包括但不限于:特定基因座(例如,特定基因、标志物、DMR、基因区域、标志物区域、经亚硫酸氢盐处理的DNA序列、CpG岛等)的甲基化和未甲基化PCR引物;优化PCR缓冲液和脱氧核苷酸;及特异性探针。

[0158] MethyLight™测定法是一种高通量定量甲基化测定法,其利用基于荧光的实时PCR(例如,TaqMan®),在PCR步骤之后不需要进一步操纵(Eads等人,Cancer Res.59:2302-2306,1999)。简言之,MethyLight™方法以基因组DNA混合样品开始,根据标准程序在亚硫酸氢钠反应中所述混合样品转化为甲基化依赖性序列差异的混合物(亚硫酸氢盐方法将未甲基化的胞嘧啶转化为尿嘧啶)。然后在“偏向”反应中,例如用重叠已知CpG二核苷酸的PCR引物进行基于荧光的PCR。在扩增过程的水平上和荧光检测过程的水平上进行序列判别。

[0159] MethyLight™测定法用作核酸,例如基因组DNA样品中甲基化模式的定量试验,其中在探针杂交的水平上进行序列判别。在定量形式中,在重叠特定假定甲基化位点的荧光探针的存在下PCR反应提供甲基化特异性扩增。通过引物和探针都不在任何CpG二核苷酸上的反应提供输入DNA的量的无偏对照。可选地,基因组甲基化的定量试验通过探测具有未覆盖已知甲基化位点的对照寡核苷酸(例如,HeavyMethyl™和MSP技术基于荧光的形式)或具有覆盖潜在甲基化位点的寡核苷酸的偏向PCR混合物实现。

[0160] MethyLight™方法连同任何合适的探针(例如,“TaqMan®”探针、Lightcycler®探针等)使用。例如,在一些应用中双链基因组DNA经亚硫酸氢钠处理并且使用TaqMan®探针(例如)和MSP引物和/或HeavyMethyl阻断物寡核苷酸和TaqMan®探针进行两组PCR反应之一。TaqMan®探针经荧光“报告物”和“猝灭剂”分子双重标记并且设计为对GC含量相对高的区域有特异性,使其在PCR循环中在比正向和反向引物高约10℃的温度下解链。这允许TaqMan®探针在PCR退火/延伸步骤期间保持完全杂交。当Taq聚合酶在PCR期间酶促合成新链时,将最终到达退火的TaqMan®探针。然后Taq聚合酶5'至3'内切核酸酶活性将通过使其消化而置换TaqMan®探针以释放荧光报告物分子以使用实时荧光检测系统定量检测此刻其未猝灭的信号。

[0161] 用于MethyLight™分析的典型试剂(例如,可能在基于MethyLight™的典型试剂盒中找到)可包括但不限于:特定基因座(例如,特定基因、标志物、DMR、基因区域、标志物区域、经亚硫酸氢盐处理的DNA序列、CpG岛等)的PCR引物;TaqMan®或Lightcycler®探针;优化PCR缓冲液和脱氧核苷酸;及Taq聚合酶。

[0162] QM<sup>TM</sup> (定量甲基化) 测定法是基因组DNA样品中甲基化模式的一种备选定量试验, 其中在探针杂交的水平上进行序列判别。在这种定量形式中, 在重叠特定假定甲基化位点的荧光探针的存在下PCR反应提供无偏扩增。通过引物和探针都不在任何CpG二核苷酸上的反应提供输入DNA的量的无偏对照。可选地, 基因组甲基化的定量试验通过探测具有未覆盖已知甲基化位点的对照寡核苷酸 (HeavyMethyl<sup>TM</sup>和MSP技术基于荧光的形式) 或具有覆盖潜在甲基化位点的寡核苷酸的偏向PCR混合物实现。

[0163] QM<sup>TM</sup>方法可在扩增过程中连同任何合适的探针, 例如“TaqMan®”探针、Lightcycler®探针使用。例如, 双链基因组DNA经亚硫酸氢钠处理并且经受无偏引物和TaqMan®探针。TaqMan®探针经荧光“报告物”和“猝灭剂”分子双重标记, 并且设计为对GC含量相对高的区域有特异性, 使其在PCR循环中在比正向和反向引物高约10°C的温度下解链。这允许TaqMan®探针在PCR退火/延伸步骤期间保持完全杂交。当Taq聚合酶在PCR期间酶促合成新链时, 将最终到达退火的TaqMan®探针。然后Taq聚合酶5'至3'内切核酸酶活性将通过使其消化而置换TaqMan®探针以释放荧光报告物分子以使用实时荧光检测系统定量检测此刻其未猝灭的信号。用于QM<sup>TM</sup>分析的典型试剂 (例如, 可能在基于QM<sup>TM</sup>的典型试剂盒中找到) 可包括但不限于: 特定基因座 (例如, 特定基因、标志物、DMR、基因区域、标志物区域、经亚硫酸氢盐处理的DNA序列、CpG岛等) 的PCR引物; TaqMan®或Lightcycler®探针; 优化PCR缓冲液和脱氧核苷酸; 及Taq聚合酶。

[0164] Ms-SNuPE<sup>TM</sup>技术是一种用于评估特定CpG位点的甲基化差异的定量方法, 其基于对DNA的亚硫酸氢盐处理, 接着是单核苷酸引物延伸 (Gonzalvo&Jones, Nucleic Acids Res. 25: 2529-2531, 1997)。简言之, 基因组DNA与亚硫酸氢钠反应以将未甲基化的胞嘧啶转化为尿嘧啶, 同时让5-甲基胞嘧啶不变。然后使用对经亚硫酸氢盐转化的DNA有特异性的PCR引物进行所需靶序列的扩增, 并且分离所得产物并用作目标CpG位点甲基化分析的模板。可分析少量DNA (例如, 显微切割病理切片) 并且避免利用限制酶用于确定CpG位点的甲基化状况。

[0165] 用于Ms-SNuPE<sup>TM</sup>分析的典型试剂 (例如, 可能在基于Ms-SNuPE<sup>TM</sup>的典型试剂盒中找到) 可包括但不限于: 特定基因座 (例如, 特定基因、标志物、DMR、基因区域、标志物区域、经亚硫酸氢盐处理的DNA序列、CpG岛等) 的PCR引物; 优化PCR缓冲液和脱氧核苷酸; 凝胶提取试剂盒; 阳性对照引物; 特定基因座的Ms-SNuPE<sup>TM</sup>引物; 反应缓冲液 (对于Ms-SNuPE反应而言); 和经标记的核苷酸。另外, 亚硫酸氢盐转化试剂可包括: DNA变性缓冲液、磺化缓冲液、DNA回收试剂或试剂盒 (例如, 沉淀、超滤、亲和柱)、去磺化缓冲液和DNA回收组分。

[0166] 简化代表性亚硫酸氢盐测序 (RRBS) 从亚硫酸氢盐处理核酸开始以将所有未甲基化的胞嘧啶转化为尿嘧啶, 接着进行限制酶消化 (例如, 用识别包括CG序列的位点的酶如MspI) 并且在与适体配体偶合后进行片段的完全测序。限制酶的选择使片段富含CpG密集区, 减少了分析期间可以绘图到多个基因位置的冗余序列的数量。这样, RRBS通过选择用于测序的限制片段的子集 (例如, 使用制备凝胶电泳通过尺寸选择) 而降低了核酸样品的复杂性。与全基因组亚硫酸氢盐测序相反, 通过限制酶消化产生的每个片段均含有至少一个CpG二核苷酸的DNA甲基化信息。这样, RRBS使样品富含启动子、CpG岛和其它基因组特征, 这些区域中限制酶切割位点的频率较高, 并且因此提供了评估一个或多个基因组基因座的甲基

化状态的测定法。

[0167] RRBS的典型方案包括用限制酶如MspI消化核酸样品,补平突出端和A-尾部,连接适体,亚硫酸氢盐转化和PCR的步骤。参见例如,等人(2005) Nat Methods 7:133-6; Meissner等人(2005) Nucleic Acids Res.33:5868-77。

[0168] 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扩增(QuARTS)测定法估计甲基化状态。在每种QuARTS测定法中这些反应按序发生,包括主反应中的扩增(反应1)和靶探针裂解(反应2);和副反应中的FRET裂解和荧光信号产生(反应3)。用特异性引物扩增靶核酸时,具有翼序列的特异性检测探针与扩增子松散结合。靶结合位点处特异性侵入性寡核苷酸的存在使裂解酶通过在检测探针和翼序列之间切割而释放翼序列。翼序列与相应FRET盒的非发夹部分互补。因此,翼序列在FRET盒上起到侵入性寡核苷酸的作用并且影响FRET盒荧光团和猝灭剂之间的裂解,这样产生荧光信号。裂解反应可以切割每个靶标的多个探针并且因此释放每个翼的多个荧光团,提供指数信号扩增。QuARTS可以通过使用具有不同染料的FRET盒在单一反应中很好地检测多个靶标。参见例如,在Zou等人(2010) “Sensitive quantification of methylated markers with a novel methylation specific technology” Clin Chem 56:A199;美国专利第8,361,720号和美国专利申请序列号13/594,674、12/946,745和12/946,752号中。

[0169] 在一些实施方案中,通过例如直接基因捕获,从样品中分离靶核酸。例如,在一些实施方案中,通过以下从样品中分离靶核酸,例如去除测定抑制剂以产生澄清样品(例如,用PVP、PVPP和/或使用旋转过滤器),用捕获试剂从澄清样品捕获靶核酸(若存在)以形成捕获复合物,从澄清样品中分离捕获复合物,从核酸溶液中的捕获复合物回收靶核酸(若存在),并且任选地重复以分离不同靶标(参见例如,美国专利申请序列号14/145,082、14/145,087、14/145,070、14/145,056、13/470,251、13/470,018、13/469,999和13/469,989)。

[0170] 在一些实施方案中,使用根据本发明的引物寡核苷酸集合(例如,参见表1A和B)和扩增酶扩增经处理DNA的片段。几个DNA片段的扩增可以在同一个反应釜中同时进行。通常,使用聚合酶链反应(PCR)进行扩增。扩增子通常长度为100至2000个碱基对。

[0171] 在所述方法的另一个实施方案中,可通过使用甲基化特异性引物寡核苷酸检测包含DMR(例如,表1A和B中提供的DMR)内或附近的CpG位置的甲基化状况。这种技术(MSP)在Herman的美国专利第6,265,171号中有描述。使用甲基化状况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经亚硫酸氢盐处理的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。对非甲基化DNA有特异性的MSP引物在CpG的C位置处含有“T”。

[0172] 通过扩增的方式获得的片段可以携带直接或间接可检测标记。在一些实施方案中,标记为荧光标记、放射性核素或具有可以在质谱仪中检测的典型质量的可检测分子片段。所述标记为质量标记时,一些实施方案提供经标记扩增子具有单个正或负净电荷,在质谱仪中允许更佳的可检测性。可通过例如基质辅助激光解吸/电离质谱法(MALDI)的方式或使用电喷雾质谱法(ESI)进行检测并且可视化。

[0173] 在一些实施方案中,分离DNA的方法包括如美国专利申请序列号13/470,251 (“Isolation of Nucleic Acids”)中所述的核酸的分离。

[0174] 方法

[0175] 在所述技术的一些实施方案中,提供了包括以下步骤的方法:

[0176] 1) 使得自受试者的核酸(例如,基因组DNA,例如分离自体液如粪便样品或结直肠组织)与区别至少一个标志物DMR(例如,表1A和B中提供的DMR)内的甲基化和非甲基化CpG二核苷酸的至少一种试剂或一系列试剂接触,并且

[0177] 2) 检测结直肠赘生物或增生性疾病(例如,结直肠癌、大腺瘤、SSP)(例如,提供高于或等于80%的敏感性和高于或等于80%的特异性)。

[0178] 在所述技术的一些实施方案中,提供了包括以下步骤的方法:

[0179] 1) 使得自受试者的核酸(例如,基因组DNA,例如分离自体液如粪便样品或结直肠组织)与区别至少一个标志物内的甲基化和非甲基化CpG二核苷酸的至少一种试剂或一系列试剂接触,所述标志物选自具有选自以下标注的染色体区域:NDRG4、BMP3、OPLAH、FLI1、PDGFD、SFMBT2(例如,SFMBT2\_895、SFMBT2\_896、SFMBT2\_897)、CHST2(例如,CHST2\_7889和CHST2\_7890)、VAV3和DTX1,并且

[0180] 2) 检测结直肠癌(例如,提供高于或等于80%的敏感性和高于或等于80%的特异性)。

[0181] 优选地,敏感性为约70%至约100%,或约80%至约90%,或约80%至约85%。在一些实施方案中,特异性为约70%至约100%,或约80%至约90%,或约80%至约85%。在一些实施方案中,特异性为约90%至100%、91%至99%、93%至97%、94%至96%、95%至99%、96%至99.5%、97%至99.9%等。)

[0182] 可通过任何方式分离基因组DNA,包括使用可商购试剂盒。简言之,其中目标DNA被细胞膜封装,应通过酶、化学或机械方式破坏和溶解生物样品。然后例如通过用蛋白酶K消化,为DNA溶液清除蛋白质和其它污染物。然后从溶液中回收基因组DNA。这可借助于多种方法实现,包括盐析、有机提取或DNA与固相支持物结合。方法的选择将受几种因素影响,包括时间、费用和所需DNA的量。包含赘生物或新生前物质的所有样品类型适合于本方法中,例如细胞系、组织学载玻片、活检、石蜡包埋组织、体液、粪便、结肠排放物、尿液、血浆、血清、全血、分离的血细胞、从血液中分离的细胞及其组合。

[0183] 所述技术不限于用于制备样品和提供用于试验的核酸的方法。例如,在一些实施方案中,使用直接基因捕获从粪便样品或血液或血浆样品中分离DNA,例如美国专利申请序列号14/145,082、14/145,087、14/145,070、14/145,056、13/470,251、13/470,018、13/469,999和13/469,989中所详述。

[0184] 然后用区别至少一个包含DMR(例如,表1A和B中提供的DMR)的标志物内的甲基化和非甲基化CpG二核苷酸的至少一种试剂或一系列试剂处理基因组DNA样品。

[0185] 在一些实施方案中,所述试剂按照杂交行为将5'-位置处未甲基化的胞嘧啶碱基转化为尿嘧啶、胸腺嘧啶或不同于胞嘧啶的另一碱基。然而,在一些实施方案中,所述试剂可为甲基化敏感性限制酶。

[0186] 在一些实施方案中,以使得按照杂交行为将5'-位置处未甲基化的胞嘧啶碱基转化为尿嘧啶、胸腺嘧啶或不同于胞嘧啶的另一碱基的方式处理基因组DNA样品。在一些实施方案中,这种处理用硫酸氢盐(氢亚硫酸盐、二亚硫酸盐)实现,接着进行碱水解。

[0187] 然后分析经处理的核酸以确定靶基因序列(来自于包含DMR,例如表1A和B提供的DMR的标志物的至少一个基因、基因组序列或核苷酸)的甲基化状态。在一些实施方案中,分



析方法是如本文所描述的QuARTS和/或MSP。

[0188] 异常甲基化,更具体地说包含DMR(例如表1A和B提供的DMR)的标志物的高甲基化与结直肠癌相关。

[0189] 所述技术涉及对与结直肠癌相关的任何样品的分析。例如,在一些实施方案中样品包含得自患者的组织和/或生物体液。在一些实施方案中,样品包含分泌物。在一些实施方案中,样品包含血液、血清、血浆、胃分泌物、结直肠组织、结直肠肿瘤组织、结直肠活检样品、胰液、胃肠道活检样品、来自于胃肠道活检的显微切割细胞、脱落至胃肠腔的胃肠道细胞和/或回收自粪便的胃肠道细胞。在一些实施方案中,受试者为人。这些样品可源于上消化道、下消化道,或包含来自于上消化道和下消化道的细胞、组织和/或分泌物。样品可包括来自于肝脏、胆管、胰腺、胃部、结肠、直肠、食道、小肠、阑尾、十二指肠、息肉、胆囊、肛门和/或腹膜的细胞、分泌物或组织。在一些实施方案中,样品包含细胞液、腹水、尿液、粪便、胰液、内窥镜检查期间获得的体液、血液、粘液或唾液。在一些实施方案中,样品为粪便样品。

[0190] 此类样品可以通过任何类型的技术获得。例如,尿液和粪便样品可容易得到,全血、腹水、血清、结直肠或胰液样品例如可在肠胃外通过使用针和注射器获得。不含或基本上不含细胞的样品可以通过使样品经受各种技术获得,包括但不限于离心和过滤。虽然通常优选使用非侵入性技术获得样品,但是仍然可优选获得诸如组织匀浆、组织切片和活检样本的样品。

[0191] 在一些实施方案中,所述技术提供了一种治疗患者(例如,有结直肠癌、早期结直肠癌或可发展结直肠癌的患者)的方法,所述方法包括确定如本文所提供的的一个或多个DMR的甲基化状态并且基于确定甲基化状态的结果向患者施用治疗。所述治疗可是进行结肠镜检查,施用药物化合物、疫苗,进行手术,为患者成像,进行另一试验。优选地,所述用途是在临床筛查的方法,预后评估的方法,监测治疗结果的方法,鉴定最有可能对特定治疗性治疗反应的患者的方法,为患者或受试者成像的方法及药物筛选和开发的方法中。

[0192] 在一些实施方案中,临床癌症预后包括确定癌症的侵蚀性及肿瘤复发的可能性以计划最有效的疗法。如果可以做更准确的预后或甚至可以评估发展癌症的潜在风险,则可以选择适当疗法,并且在一些情况下可以为患者选择不太严重的疗法。癌症生物标志物的评估(例如,确定甲基化状态)对于将预后良好和/或发展癌症的风险低,将不需要治疗或需要有限治疗的受试者与很可能发展癌症或遭受癌症复发,可能受益于更强治疗或监测的受试者分开有用。

[0193] 在当前公开的主题的一些实施方案中,随时间推移可做生物标志物的多次确定以利于诊断和/或预后。使用生物标志物的时间变化预测临床结果,监测胃肠癌的进展,和/或监测针对癌症的适当疗法的功效。例如在此类实施方案中,可能会期望了解在有效治疗的过程中随时间推移生物样品中本文公开的一个或多个生物标志物(例如,DMR)(若监测,和潜在地一个或多个附加生物标志物)的甲基化状态的变化。

[0194] 在一些实施方案中,在早期,例如在疾病的症状出现之前采用本发明的方法和组合物治疗或诊断疾病。

[0195] 在一些实施方案中,统计分析将预后指标与不良结果的倾向联系在一起。例如,在一些实施方案中,通过统计显著性水平确定,与得自未患癌症的患者的正常对照样品中的甲基化状态不同的甲基化状态可以表示受试者比具有更类似于对照样品中的甲基化状态

的水平受试者更有可能患癌症。另外,甲基化状态从基线(例如,“正常”)水平的变化可以反映受试者预后,并且甲基化状态的变化程度可与不良事件的严重程度相关。常常通过比较两个或更多群体并且确定置信区间和/或p值来确定统计显著性。参见,例如,Dowdy和Wearden,Statistics for Research,John Wiley&Sons,New York,1983。当前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。

[0196] 在其它实施方案中,可以确定本文公开的预后或诊断生物标志物(例如,DMR)的甲基化状态变化的阈值程度,并且仅仅将生物样品中生物标志物的甲基化状态变化程度与甲基化状态变化的阈值程度做比较。本文提供的生物标志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%。在其它实施方案中,可以确定“列线图”,通过列线图将预后或诊断指示物(生物标志物或生物标志物的组合)的甲基化状态直接与针对指定结果的相关倾向直接相关。技术人员熟悉此类列线图用于将两个数值相关联的用途,理解这种测量中的不确定性与标志物浓度上的不确定性相同,因为参考的是单独样品处理,而非群体平均值。

[0197] 在一些实施方案中,与生物样品同时分析对照样品,使得可以将从生物样品获得的结果与从对照样品获得的结果做比较。另外,考虑到可以提供标准曲线,可将其与生物样品的测定结果做比较。此类标准曲线根据测定单位,例如荧光信号强度(若使用荧光标记)呈现生物标志物的甲基化状态。使用取自多个供体的样品,可以为正常组织中所述一个或多个生物标志物的对照甲基化状态,以及取自有化生的供体或有胃肠癌的供体的组织中所说一个或多个生物标志物的“风险”水平提供标准曲线。在所述方法的某些实施方案中,在鉴定得自受试者的生物样品中本文提供的一个或多个DMR的异常甲基化状态之后,将受试者鉴定为有化生。在所述方法的其它实施方案中,检测到得自受试者的生物样品中一个或多个此类生物标志物的异常甲基化状态导致受试者被鉴定为患有癌症。

[0198] 标志物的分析可以与一个试样中另外的标志物分开或同时进行。例如,可以将几个标志物组合到一次试验中以有效处理多个样品并且潜在地提供更高的诊断和/或预后精确性。另外,本领域的技术人员将认识到试验来自于相同受试者的多个样品(例如,在连续时间点)的价值。连续样品的此类试验可以允许鉴定随时间推移标志物甲基化状态的变化。甲基化状态的变化以及甲基化状态变化的不存在可以提供关于疾病状况的有用信息,包括但不限于鉴定事件发生开始的大致时间,可挽救组织的存在和量,药物疗法的适合程度,各种疗法的有效性,及对受试者结果,包括未来事件风险的鉴定。

[0199] 标志物的分析可以呈各种物理形式进行。例如,微量滴定板或自动化可用于促进对大量试样的处理。可选地,单样品形式可用于促进及时地,例如在移动运输或急诊室情况下的即时治疗和诊断。

[0200] 在一些实施方案中,如果与对照甲基化状态比较时,样品中至少一个生物标志物的甲基化状态存在可测量的差异,将受试者诊断为患有结直肠癌。相反,在生物样品中未鉴定出甲基化状态的变化时,可以将受试者诊断为未患结直肠癌,未处于癌症风险中,或诊断为具有低癌症风险。就这一点而言,因此可以将有癌症或癌症风险的受试者与具有低风险至基本上无癌症或癌症风险的受试者区分开。具有发展结直肠癌的风险的那些受试者可以采用更强和/或更规律的筛查方案,包括内窥镜检查监测。另一方面,具有低风险至基本上

无风险的那些受试者可以避免进行内窥镜检查,直至诸如进一步筛查,例如根据本技术进行筛查,表明在那些受试者中已经出现结直肠癌风险时为止。

[0201] 如以上所提到的,根据本技术的所述方法的实施方案,检测所述一个或多个生物标志物的甲基化状态的变化可以是定性确定或者可以是定量确定。这样,将受试者诊断为患有胃肠癌或处于发展胃肠癌风险的步骤表明,进行了某些阈值测量,例如生物样品中所述一个或多个生物标志物的甲基化状态不同于预定对照甲基化状态。在所述方法的一些实施方案中,对照甲基化状态是生物标志物的任何可检测甲基化状态。在所述方法的其它实施方案中,其中对照样品与生物样品同时试验,预定甲基化状态是对照样品中的甲基化状态。在所述方法的其它实施方案中,预定甲基化状态基于和/或通过标准曲线鉴定。在所述方法的其它实施方案中,预定甲基化状态是具体状态或状态范围。这样,可以在本领域技术人员将显而易见的可接受限制范围内,部分基于实践的所述方法的实施方案和所需特异性等,选择预定甲基化状态。

### 实施例

[0202] 实施例1.

[0203] 该实施例证明NDRG4、BMP3、OPLAH、FLI1、PDGFD、CHST\_7889、SFMBT2\_895、SFMBT2\_896、SFMBT2\_897、CHST2\_7890、VAV3和DTX1为粪便样品内用于检测结直肠癌的有效标志物。

[0204] 表1A中提供了NDRG4、BMP3、OPLAH、FLI1、PDGFD、CHST\_7889、SFMBT2\_895、SFMBT2\_896、SFMBT2\_897、CHST2\_7890、VAV3和DTX1的正向和反向引物及探针序列。表1B提供了关于表1A中提供的甲基化标志物、染色体和DMR基因组坐标的信息。

[0205] 将每个标志物的捕获探针设计为具有75-80°C的解链温度和介于25-35个碱基之间的长度(参见表1C)。另外,可以将捕获探针杂交区选择为在亚硫酸氢盐后QuARTS足迹内。表1C提供了甲基化标志物和各自的捕获探针序列。

[0206] 表1A:实施例1中利用的标志物的正向引物、探针、反向引物序列

[0207]

甲基化标志物	正向引物	探针序列	反向引物
VAV3	TCGGAGTCG AGTTTAGCG C (SEQ ID NO: 1)	CCACGGACG-CGG CGTTCGCGA/3C6/ (SEQ ID NO: 2)	CGAAATCGAAAA ACAAAAACCGC (SEQ ID NO: 3)
CHST2_7 889	CGAGTTCGG TAGTTGTAC GTAGA (SEQ ID NO: 4)	CGCCGAGG-TCGT CGATACCG/3C6/ (SEQ ID NO: 5)	CGAAATACGAACG CGAAATCTAAA ACT (SEQ ID NO: 6)
SFMBT2_ 897	GTcGTcGTcG AGAGGGTA (SEQ ID NO: 7)	CCACGGACG-ATC GGTTTCGTT/3C6/ (SEQ ID NO: 8)	CGAACAAAAACGA ACGAACGAA (SEQ ID NO: 9)
SFMBT2_ 896	GCGTTTAGG TTGGTCGGA GA (Version1) (SEQ ID NO: 10) GCGTTTAGG TTGGTCGGA G (形式 2) (SEQ ID NO: 49)	CGCCGAGG-CTAC GAACCGAA/3C6/ (形式 1) (SEQ ID NO: 11) CGCCGAGGCCGA AAA ACTAC/3C6/ (形式 2) (SEQ ID NO: 50)	CCTAACCAACGCA CTCAACC (形式 1) (SEQ ID NO: 12) ACGCACTCAACCT ACGAAC (形式 2) (SEQ ID NO: 51)
SFMBT2_ 895	TTAGCGAcGT AGTcGTcGTT G (形式 1) (SEQ ID NO: 13) GCGACGTAG TCGTCGTTG T (形式 2) (SEQ ID NO: 52)	CCACGGACG-CGA AAACGCGAA/3C6/ (形式 1) (SEQ ID NO: 14) CCACGGACGGAA AACGCGAAA/3C6/ (形式 2) (SEQ ID NO: 53)	CCCAACGCGAAAA AAACGC (形式 1) (SEQ ID NO: 15) CCAACGCGAAAA AACGCG (形式 2) (SEQ ID NO: 54)
CHST2_7 890	GTATAGCGC GATTCGTA GcG (SEQ ID NO: 16)	CGCCGAGG-CGAA CATCCTCC/3C6/ (SEQ ID NO: 17)	AATTACCTACGCTA TCCGCCC (SEQ ID NO: 18)
OPLAH	cGTcGcGTTTT	CCACGGACG-GCA	CGCGAAA ACTAAA

	TcGGTTATAC G (SEQ ID NO: 19)	CCGTAAAAC/3C6/ (SEQ ID NO: 20)	AAACCGCG (SEQ ID NO: 21)
	PDGFD AAACGTAA TTIGTTGTTT GTTTCGTT (形 式 1) (SEQ ID NO: 22) GCGAATAAA TAAACGTTA ATTTGTTGTT TGTTTCG (形 式 2) (SEQ ID NO: 55)	ACTTCCGAACGC GTATAAATACC (形 式 1) (SEQ ID NO: 23) CCACGGACGCGC ACTTCCTTA/3C6/ (形 式 2) (SEQ ID NO: 56)	GCGAATAAATAAAC GTAAATTTGTTGTT TGTTTCG (形式 1) (SEQ ID NO: 24) ACTTCCGAACGC GTATAAATACC (形 式 2) (SEQ ID NO: 57)
[0208]	FLI1 GTTGcGAGG TTAGGTTGT AATCG (SEQ ID NO: 25)	CGCCGAGG-CGTC CATTTAAC/3C6/ (SEQ ID NO: 26)	CGCCGCTTACCTTA ATAATCCC (SEQ ID NO: 27)
	DTX1 GAGTCGCGG TTTCGTTTTT (SEQ ID NO: 28)	CGCCGAGG-CGCG TTCGTTTT /3C6/(SEQ ID NO: 29)	GACGCGACGACCG AAAAAC (SEQ ID NO: 30)
	NDRG4 CGGTTTTTCG TTCGTTTTTT CG (SEQ ID NO: 31)	CCACGGACG GTTTCGTTTATCG/3 C6/ (SEQ ID NO: 32)	CCGCCTTCTACGCG ACTA (SEQ ID NO: 33)
	BMP3 GTTTAATTTT CGGTTTCGT CGTC (SEQ ID NO: 34)	CGCCGAGG CGGTTTTTTGCG/3 C6/ (SEQ ID NO: 35)	CGCTACGAAACAC TCCGA (SEQ ID NO: 36)

[0209] 表1B. 甲基化标志物、染色体和DMR基因组坐标。

甲基化标志物	染色体	DMR 基因组坐标
BMP3	4	81031173 - 81031262
NDRG4	16	58463478 - 58463588
VAV3	1	107964966 - 107965057
CHST2 7889	3	143119424 - 143119583
SFMBT2 897	10	7410903 - 7411014
SFMBT2 896	10	7410764 - 7410837
SFMBT2 895	10	7410331 - 7410490
CHST2 7890	3	143119999 - 143120158
OPLAH	8	144051847 - 144052006
PDGFD	11	104164082 - 104164186

[0211]	FLII	11	128694158 - 128694317
	DTX1	12	113056762 - 113056895

[0212] 表1C. 甲基化标志物和相应的捕获探针序列。

甲基化标志物	捕获探针序列
VAV3	/5AmMC6/GATCGAGGGAGCAGGAGCCGCGGCTGACGGGTCGCG (SEQ ID NO: 37)
CHST2_7889	/5AmMC6/CGGTGCCGAGAGCTGCCAGAGAGTTGGATTCTGCG (SEQ ID NO: 38)
SFMBT2_897	/5AmMC6/GCGAGCGGGCAAGGGCGGGCGAGC (SEQ ID NO: 39)
SFMBT2_896	/5AmMC6/ACCTGCGGGCCGAAGGGCTGCTCTCCGG (SEQ ID NO: 40)
SFMBT2_895	/5AmMC6/AGGAGACGCGGGAGCGCGGGGTAGGTAGC (SEQ ID NO: 41)
[0213] CHST2_7890	/5AmMC6/GGCATCCTCCCGGTGATGGAAGCAGCCGCCGCG (SEQ ID NO: 42)
OPLAH	/5AmMC6/GGAAGGCGCGGCGCTCGGTCAGCACTGACAGCAG (SEQ ID NO: 43)
PDGFD	/5AmMC6/TCGCCGAGCTCTCCCCAACTTCCCTGCATGCTGAACIT (SEQ ID NO: 44)
FLII	/5AmMC6/CCGTCCATTTGGCCAAGTCTGCAGCCGAGCC (SEQ ID NO: 45)
DTX1	/5AmMC6/CTGCGTCCGTCCGTCCGCGGGCAGTCTGTCCA (SEQ ID NO: 46)
NDRG4	/5AmMC6/TCCCTCGCGCGTGGCTTCCGCCTTCTGCGCGGCTGGGGTGCCCGGTGG (SEQ ID NO: 47)
BMP3	/5AmMC6/GCGGGACACTCCGAAGGCGCAAGGAG (SEQ ID NO: 48)

[0214] 每个捕获探针经5' -NH<sub>2</sub>修饰合成以允许通过标准碳二亚胺偶合化学法偶合至经-COOH修饰的磁性颗粒上。同样,捕获探针的互补寡核苷酸合成为含有5' -Cy3标记。这种互补探针用于确认捕获探针偶合至磁性颗粒上。

[0215] 为试验每个探针的捕获效率以及评估标志物性能,制备正常和癌症患者的两份粪便混合物。癌症粪便混合物来自于6名患者(3名为CRC,1名为AA而2名未知)。类似地,正常粪便来自于6名非CRC正常患者。在匀浆离心之后通过混合上清液制备粪便。然后将混合上清液等分为含有14mL上清液的单一捕获样品。

[0216] 将捕获探针设计为具有75-80°C的解链温度和介于25-35个碱基之间的长度。另外,可以将捕获探针杂交区选择为在亚硫酸氢盐后QuARTS足迹内。

[0217] 为进行捕获,混合两种标志物的捕获珠粒(具有共价连接的捕获探针的磁性颗粒)加上ACTB捕获珠粒以形成三捕获珠粒混合物。

[0218] 对14mL的阳性和正常粪便上清液进行捕获,一式三份(因为混合物不足,所以VAV3和DTX1除外,进行一式两份)。

[0219] 捕获完成之后,在42°C下用0.1N NaOH从捕获珠粒将粪便DNA洗脱20分钟,接着在

56°C下使用亚硫酸氢铵进行亚硫酸氢盐转化1小时。然后为粪便DNA去磺化并且使用涂硅磁珠纯化并且在70uL的10mM Tris-HCl (pH 8)、0.1mM EDTA中洗脱。然后试验10uL洗脱液并且在QuARTS测定法中量化。

[0220] 为能够量化样品,使用具有对应于QuARTS足迹的DNA插入物的pUC57质粒。DNA插入物侧接EcoRI位点以允许使用260nm下的吸光度线性化和量化。

[0221] 为允许反算洗脱样品中的链,对10uL洗脱液和消化质粒的稀释液进行QuARTS测定法。

[0222] 表1D示出了每个试验标志物获得的结果。这些结果显示在阳性和正常粪便混合物间甲基化标志物具有高链差异,表明这些标志物是粪便中用于CRC检测的候选物。

[0223] 表1D.

甲基化标志物	阳性混合物均链	正常混合物均链	倍数差异(阳性/正常)
[0224] NDRG4	1,568	140	11.2
BMP3	395	8	51.4
OPLAH	840	279	3.0
FLI1	1,715	167	10.2
[0225] PDGFD	843	67	12.6
CHST2_7889	945	17	56.4
SFMBT2_895	837	5	152.3
SFMBT2_896	856	150	5.7
SFMBT2_897	844	45	18.7
CHST2_7890	1,396	62	22.6
VAV3	367	21	17.9
DTX1	751	105	7.2

[0226] 实施例2.

[0227] 在人类受试者内筛查结直肠癌的示例性程序。

[0228] 使得自人类受试者的核酸(例如,基因组DNA,例如分离自体液如粪便样品或结直肠组织)与区别至少一个标志物DMR内的甲基化和非甲基化CpG二核苷酸的至少一种试剂或一系列试剂接触,所述标志物DMR选自:

[0229] • NDRG4、BMP3、OPLAH、FLI1、PDGFD、SFMBT2(例如,SFMBT2\_895、SFMBT2\_896、SFMBT2\_897)、CHST2(例如,CHST2\_7889和CHST2\_7890)、VAV3和DTX1(如表1A和B中所述)。

[0230] 当所述标志物的甲基化状态不同于在没有赘生物的受试者中测定的所述标志物的甲基化状态时,将所述受试者鉴定为有结直肠癌。

[0231] 以上说明书中提到的所有出版物和专利通过引用整体并入本文用于所有目的。在不背离所述技术的范围和精神的前提下,描述的所述技术的组合物、方法和用途的各种修改和变型对于本领域技术人员将显而易见。虽然已经连同具体示例性实施方案描述了所述技术,但是应理解要求保护的本发明不应过度限于此类具体实施方案。事实上,对于药理学、生物化学、医学科学或有关领域的技术人员明显的,对描述的用于实施本发明的模式的各种修改旨在属于以下权利要求的范围之内。

[0001]

## 序列表

- <110> EXACT SCIENCES CORPORATION  
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH
- <120> 检测结直肠赘生物
- <130> EXCTM-33841/WO-1/ORD
- <150> US 61/977,954  
<151> 2014-04-10
- <150> US 61/972,942  
<151> 2014-03-31
- <160> 57
- <170> PatentIn版本3.5
- <210> 1  
<211> 19  
<212> DNA  
<213> 人工序列
- <220>  
<223> 合成
- <400> 1  
tcggagtcga gtttagcgc 19
- <210> 2  
<211> 21  
<212> DNA  
<213> 人工序列
- <220>  
<223> 合成
- <400> 2  
ccacggacgc ggcgttcgcg a 21
- <210> 3  
<211> 24  
<212> DNA  
<213> 人工序列
- <220>  
<223> 合成
- <400> 3  
cgaaatcgaa aaaacaaaa cgc 24
- <210> 4



## [0002]

<211> 23  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 4  
 cgagttcggg agttgtacgt aga 23

<210> 5  
 <211> 20  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 5  
 cgccgaggtc gtcgataccg 20

<210> 6  
 <211> 27  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 6  
 cgaaatacga acgcaaatc taaaact 27

<210> 7  
 <211> 19  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 7  
 gtcgtcgttc gagaggta 19

<210> 8  
 <211> 21  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 8

[0003]

ccacggacga teggtttcgt t 21

<210> 9  
 <211> 22  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 9  
 cgaacaaaaa cgaacgaacg aa 22

<210> 10  
 <211> 20  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 10  
 gcgtttaggt tggtcggaga 20

<210> 11  
 <211> 20  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 11  
 egccgaggct acgaaccgaa 20

<210> 12  
 <211> 20  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 12  
 cetaaccaac gcactcaacc 20

<210> 13  
 <211> 21  
 <212> DNA  
 <213> 人工序列

[0004]

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 13

tttagcgacgt agtcgctggtt g

21

&lt;210&gt; 14

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 14

ccacggacgc gaaaacgcga a

21

&lt;210&gt; 15

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 15

cccaacgcga aaaaaacgc

19

&lt;210&gt; 16

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 16

gtatagcgcg atttcgtagc g

21

&lt;210&gt; 17

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 17

cgccgaggcg aacatcctcc

20

&lt;210&gt; 18

## [0005]

<211> 21  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 18  
 aattacctac gctatccgcc c 21

<210> 19  
 <211> 22  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 19  
 cgtcgegitt ttcggtata cg 22

<210> 20  
 <211> 21  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 20  
 ccacggacgg caccgtaaaa c 21

<210> 21  
 <211> 21  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 21  
 cgcgaaaact aaaaaaccgc g 21

<210> 22  
 <211> 27  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 22

[0006]

aaacgttaat ttgttgittg tttcgtt 27

<210> 23  
 <211> 24  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 23  
 actttcegaac cgcgtataaa tacc 24

<210> 24  
 <211> 35  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 24  
 gcgaataaat aaacgttaat ttgttgittg tttcg 35

<210> 25  
 <211> 23  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 25  
 gttgcgaggt taggttgtaa tcg 23

<210> 26  
 <211> 20  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 26  
 cgccgagcgc tccatttaac 20

<210> 27  
 <211> 22  
 <212> DNA  
 <213> 人工序列

[0007]

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 27

cgccgcttac ctttaataatc cc

22

&lt;210&gt; 28

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 28

gagtcgcggt ttcgttttc

19

&lt;210&gt; 29

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 29

cgccgagcgc cgttcgtttt

20

&lt;210&gt; 30

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 30

gacgcgacga ccgaaaaac

19

&lt;210&gt; 31

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 31

cggttttcgt tcgttttttc g

21

&lt;210&gt; 32

## [0008]

<211> 21  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 32  
 ccacggacgg ttcgtttatc g 21

<210> 33  
 <211> 18  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 33  
 ccgcctteta cgegacta 18

<210> 34  
 <211> 23  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 34  
 gtttaatttt cggtttcgtc gtc 23

<210> 35  
 <211> 20  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 35  
 cgccgaggcg gttttttgcg 20

<210> 36  
 <211> 18  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 36

[0009]

cgctacgaaa cactccga 18

<210> 37  
 <211> 36  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 37  
 gatcgaggga gcaggagccg cggctgacgg gtcgcg 36

<210> 38  
 <211> 34  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 38  
 cggtgccgag agctgccaga gagttggatt ctgc 34

<210> 39  
 <211> 24  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 39  
 gcgagcgggc aagggcgggc gaggc 24

<210> 40  
 <211> 28  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 40  
 acctgcgggc cgaagggctg ctctccgg 28

<210> 41  
 <211> 29  
 <212> DNA  
 <213> 人工序列



[0010]

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 41

aggagacgcg ggagcgcggg gtaggtagc

29

&lt;210&gt; 42

&lt;211&gt; 34

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 42

ggcctcctcc cggatgatgga agcagccgcc gccg

34

&lt;210&gt; 43

&lt;211&gt; 34

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 43

ggaaggcgcg gcgctcggtc agcaetgaca gcag

34

&lt;210&gt; 44

&lt;211&gt; 39

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 44

tcgccgagct ctececaaac ttctgcatg ctgaacttt

39

&lt;210&gt; 45

&lt;211&gt; 31

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 45

ccgtccattt ggccaagtct gcagccgagc c

31

&lt;210&gt; 46

## [0011]

- <211> 33  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 合成  
  
 <400> 46  
 ctgctccgt ccgtcgccg ggcagtctgt cca 33
- <210> 47  
 <211> 48  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 合成  
  
 <400> 47  
 tcctcgcgc gtggettccg cttctgcgc gctggggtg cccggtgg 48
- <210> 48  
 <211> 26  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 合成  
  
 <400> 48  
 gcgggacaact ccgaaggcgc aaggag 26
- <210> 49  
 <211> 19  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 合成  
  
 <400> 49  
 gcgttaggt tggteggag 19
- <210> 50  
 <211> 20  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 合成  
  
 <400> 50

[0012]

cgccgaggcc gaaaaactac 20

<210> 51  
 <211> 19  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 51  
 acgcaactcaa cctacgaac 19

<210> 52  
 <211> 19  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 52  
 gcgacgtagt cgtcgttgt 19

<210> 53  
 <211> 21  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 53  
 ccacggacgg aaaacgcgaa a 21

<210> 54  
 <211> 19  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 合成

<400> 54  
 ccaacgcgaa aaaaacgcg 19

<210> 55  
 <211> 35  
 <212> DNA  
 <213> 人工序列

## [0013]

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 55

gcgaataaat aaacgttaat ttgttgtttg tttcg

35

&lt;210&gt; 56

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 56

ccacggacgc gcaattcett a

21

&lt;210&gt; 57

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 合成

&lt;400&gt; 57

actttccgaa cgcgtataaaa tacc

24