



(12) 发明专利申请

(10) 申请公布号 CN 115917009 A

(43) 申请公布日 2023.04.04

(21) 申请号 202080102436.5

P·科内尔·诺格尔 M·切西科拉

(22) 申请日 2020.09.21

P·科纳普 M·比滕茨

(30) 优先权数据

63/046,510 2020.06.30 US

(74) 专利代理机构 北京三友知识产权代理有限公司 11127

专利代理师 张培源 褚瑶杨

(85) PCT国际申请进入国家阶段日

2022.12.26

(51) Int.Cl.

C12Q 1/6886 (2006.01)

(86) PCT国际申请的申请数据

PCT/EP2020/076225 2020.09.21

(87) PCT国际申请的公布数据

W02022/002423 EN 2022.01.06

(71) 申请人 通用诊断股份公司

地址 西班牙塞维利亚

(72) 发明人 K·克鲁斯马

J·马丁内兹-巴雷亚 C·亨塞

P·索拉·德罗斯·桑托斯

权利要求书2页 说明书43页

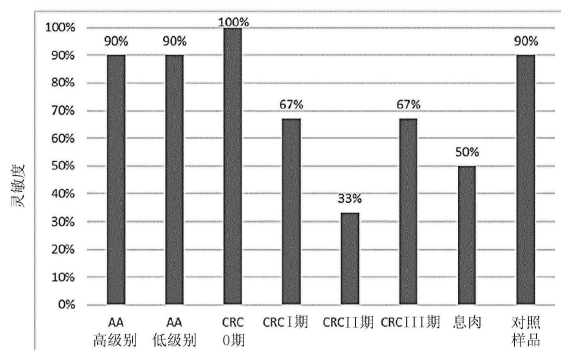
序列表14页 附图5页

(54) 发明名称

结肠直肠肿物的检测

(57) 摘要

本公开尤其提供了用于结肠直肠肿物检测(例如,筛查)的方法、和与之相关的组合物。在各种实施方式中,本公开提供了用于腺瘤和/或早期结肠直肠癌检测(例如,筛查)的方法和与之相关的组合物。在各种实施方式中,本公开提供了包括分析一种或多种甲基化生物标志物的甲基化状态的筛查方法,以及与之相关的组合物。在各种实施方式中,本公开提供了用于检测(例如,筛查)的方法,其包括检测(例如,筛查)cfDNA中(例如,ctDNA中)的一种或多种甲基化生物标志物的甲基化状态。在各种实施方式中,本公开提供了用于筛查的方法,其包括使用MSRE-qPCR和/或使用大规模平行测序(例如,下一代测序)检测(例如,筛查)cfDNA中(例如,ctDNA中)的一种或多种甲基化生物标志物的甲基化状态。



1. 一种体外筛查人类受试者的结肠直肠肿物的方法,所述方法包括:

(i) 将受试者分类为患有晚期腺瘤, (ii) 将受试者分类为患有息肉病, (iii) 将受试者分类为患有结肠直肠癌0、I、II、III期或未分化期, (iv) 在确定或不确定受试者患有下述病症中的哪种的情况下, 将受试者分类为患有晚期腺瘤、息肉病和结肠直肠癌病症中的至少一种病症, 或 (v) 在确定或不确定受试者患有这些病症中的哪一种的情况下, 将受试者分类为患有晚期腺瘤和结肠直肠癌病症中的至少一种病症], 所述方法包括:

- 确定人类受试者DNA样品中以下一项或多项各自的甲基化状态:

(i) 具有SEQ ID NO:6的基因SLC6A1内的甲基化基因座, 和

(ii) SEQ ID NO:51内的甲基化基因座, 并且

将获得的数据与从健康个体获得的参考值进行比较, 以及

如果与参考样品相比, 在一个或多个基因座中检测到超甲基化, 则诊断受试者的结肠直肠肿物。

2. 如权利要求1所述的方法, 所述方法进一步包括: 确定人类受试者的DNA样品中以下的甲基化状态:

(i) 具有SEQ ID NO:23的基因F13A1内的甲基化基因座; 和

(ii) 具有SEQ ID NO:35的基因BARHL1内的所述甲基化基因座, ; 并且

将获得的数据与从健康个体获得的参考值进行比较, 以及

如果与参考样品相比, 在一个或多个所述基因座中检测到超甲基化, 则诊断人类受试者的结肠直肠肿物。

3. 如权利要求1所述的方法, 所述方法进一步包括: 确定人类受试者的DNA样品中以下的甲基化状态:

(i) 具有SEQ ID NO:1的基因CSMD2内的甲基化基因座; 和

(ii) SEQ ID NO:25内的甲基化基因座; 并且

将获得的数据与从健康个体获得的参考值进行比较, 以及

如果与参考样品相比, 在一个或多个所述基因座中检测到超甲基化, 则诊断人类受试者的结肠直肠肿物。

4. 如前述权利要求中任一项所述的方法, 其中, 一个或多个标志物中的每一个是包含权利要求1至3中任一项所列的差异甲基化区域 (DMR) 的至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%的部分的甲基化基因座。

5. 如前述权利要求中任一项所述的方法, 其中, 所述DNA分离自人类受试者的血液、血浆、粪便或结肠直肠组织。

6. 如前述权利要求中任一项所述的方法, 其中, 所述DNA是人类受试者的无细胞DNA。

7. 如前述权利要求中任一项所述的方法, 其中, 使用定量聚合酶链反应 (qPCR) 确定甲基化状态。

8. 如前述权利要求中任一项所述的方法, 其中, 使用以下方式确定甲基化状态: 大规模平行测序, 例如下一代测序, 其包括合成测序, 实时测序, 如单分子测序、珠乳剂测序和纳米孔测序。

9. 如前述权利要求中任一项所述的方法, 其中, 每个甲基化基因座的长度等于或小于 5000bp、4,000bp、3,000bp、2,000bp、1,000bp、950bp、900bp、850bp、800bp、750bp、700bp、650bp、600bp、550bp、500bp、450bp、400bp、350bp、300bp、250bp、200bp、150bp、100bp、50bp、40bp、30bp、20bp或10bp。

10. 一种用于前述权利要求中任一项所述方法的试剂盒, 所述试剂盒包含一个或多个寡核苷酸引物对, 用于扩增一个或多个相应的甲基化基因座。

11. 一种在权利要求1至9中任一项所述的方法中筛选结肠直肠癌的诊断性qPCR反应, 所述诊断性qPCR反应包括: 人DNA、聚合酶、用于扩增一个或多个相应的甲基化基因座的一个或多个寡核苷酸引物对, 以及可选地, 至少一种甲基化敏感性限制酶。

12. 一种用于权利要求1至9中任一项所述方法的试剂盒, 所述试剂盒包含一种或多种寡核苷酸捕获诱饵, 例如一种或多种生物素化的寡核苷酸探针, 以用于捕获一个或多个相应的甲基化基因座以杂交到感兴趣的区域。

13. 如权利要求1至9中任一项所述的方法, 所述方法包括使用下一代测序(NGS)来确定所述一个或多个标志物各自的甲基化状态。

14. 如权利要求13所述的方法, 所述方法包括使用一种或多种寡核苷酸捕获诱饵, 如富集目标区域的生物素化寡核苷酸探针, 以捕获一个或多个相应的甲基化基因座, 然后进行文库制备和测序, 其中样品在捕获前进行亚硫酸氢盐转化或酶促转化。

结肠直肠肿物的检测

技术领域

[0001] 本发明一般涉及用于检测和/或预先筛查结肠直肠肿物的方法和试剂盒。在一些实施方式中,本文所述的方法和试剂盒利用人类基因组的已鉴定的差异甲基化区域作为标志物来确定结肠直肠肿物的存在和/或风险。

背景技术

[0002] 结肠直肠肿物(如结肠直肠癌)的筛查是癌症预防、诊断和治疗的重要组成部分。根据一些报告,结肠直肠癌(CRC)已被鉴定为世界上第三大最常见的癌症类型和第二大癌症死亡原因。根据一些报道,每年有超过180万例新发结肠直肠癌病例,约881,000人死于结肠直肠癌,约占癌症死亡人数的十分之一。建议进行定期结肠直肠癌筛查,尤其是50岁以上的个体。此外,50岁以下的个体的结肠直肠癌的发病率随着时间的推移而增加。统计数据表明,目前的结肠直肠癌筛查技术是不够的。

[0003] 此外,结肠癌的早期检测会降低死亡率。检测和去除结肠癌的前体,包括但不限于具有晚期特征的结肠息肉,将降低CRC的发病率,因为这些息肉被认为是恶性进展的最大风险。切除晚期结肠息肉将降低患癌症的风险。通常,约9-16%的50岁以上无症状患者会发现晚期腺瘤。

[0004] 因此,需要可以提供结肠直肠肿物的分类和/或诊断的方法、组合物和系统。特别是,需要在早期阶段对结肠直肠肿物进行诊断和/或分类。

发明内容

[0005] 本公开尤其提供了用于检测(例如,筛查)结肠直肠肿物的方法、和与之相关的组合物。在各种实施方式中,本公开提供了用于对患有和/或未患有恶变前的和/或恶性的结肠直肠肿物的受试者进行分类的方法,所述结肠直肠肿物包括但不限于晚期腺瘤、息肉病、结肠直肠癌(例如,0、I、II、III期或未分化期)和/或其各种组合。在各种实施方式中,本公开提供了用于结肠直肠肿物筛查的方法以及与之相关的组合物,所述方法包括确定来自受试者(例如,人类受试者)的样品(例如,血液样品、血液制品样品、粪便样品、结肠直肠组织样品)内的一个或多个标志物内发现的一个或多个甲基化位点处的甲基化状态(例如,甲基化的数量、频率或模式)。例如,标志物可以包括甲基化基因座,例如人类受试者的脱氧核糖核酸(DNA)的差异甲基化区域(DMR)。在各种实施方式中,本公开提供了将受试者分类为患有和/或未患有晚期腺瘤、息肉病、结肠直肠癌和/或其任何组合的方法,所述方法包括确定cfDNA(无细胞DNA),例如ctDNA(循环肿瘤DNA)中一个或多个甲基化基因座中每一个的甲基化状态。在各种实施方式中,本公开提供了用于结肠直肠肿物筛查的方法,所述方法包括使用以下方法来确定cfDNA(例如,ctDNA)中一个或多个甲基化基因座中的每一个的甲基化状态,例如,大规模平行测序(例如下一代测序),例如合成测序、实时(例如单分子)测序、乳液测序、纳米孔测序、定量聚合酶链反应(qPCR)(例如,甲基化敏感的限制酶定量聚合酶链反应,MSRE-qPCR)。本文提供的各种组合物和方法提供了足以用于临床应用以筛查病症的

灵敏度和特异性,这些病症包括但不限于晚期腺瘤、息肉病和/或早期结肠直肠癌。本文提供的各种组合物和方法可用于晚期腺瘤、息肉病和/或结肠直肠癌筛查,其通过分析受试者的可获取组织样品,例如是血液或血液成分(例如,cfDNA,例如,ctDNA)或粪便的组织样品。

[0006] 一方面,本发明涉及一种检测(例如,筛查)人类受试者的结肠直肠肿物的方法[例如,(i)将受试者分类为患有晚期腺瘤,(ii)将受试者分类为患有息肉病,(iii)将受试者分类为患有结肠直肠癌(例如,0、I、II、III期或未分化期),(iv)在确定或不确定受试者患有下述病症中的哪种的情况下,将受试者分类为具有晚期腺瘤、息肉病和结肠直肠癌病中的至少一种病症,或(v)在确定或不确定受试者患有下述病症中的哪种的情况下,将受试者分类为患有晚期腺瘤和结肠直肠癌中的至少一种病症],所述方法包括:

[0007] -确定人类受试者DNA样品中以下一项或多项中的每一个的甲基化状态:

[0008] (i)具有SEQ ID NO:6的基因SLC6A1内的甲基化基因座,和

[0009] (ii)SEQ ID NO:51内的甲基化基因座,并且

[0010] 将获得的数据与从健康个体获得的参考值进行比较,以及

[0011] 如果与参考样品相比,在一个或多个基因座中检测到超甲基化,则可以诊断受试者的结肠直肠肿物。

[0012] 另一方面,本发明涉及一种检测(例如,筛查)人类受试者的结肠直肠肿物的方法[例如,(i)将受试者分类为患有晚期腺瘤,(ii)将受试者分类为患有息肉病,(iii)将受试者分类为患有结肠直肠癌(例如,0、I、II、III期或未分化期),(iv)在确定或不确定受试者下述病症中的哪种的情况下,将受试者分类为具有晚期腺瘤、息肉病和结肠直肠癌病中的至少一种病症,或(v)在确定或不确定受试者下述病症中的哪种的情况下,将受试者分类为患有晚期腺瘤和结肠直肠癌中的至少一种病症],所述方法包括:确定从受试者获得的样品(例如,血液样品、血液制品样品、粪便样品、结肠直肠组织样品)中鉴定的一个或多个标志物中每一个的甲基化状态,并且至少部分地基于一个或多个标志物的每一个的所确定的甲基化状态来确定受试者是否患有结肠直肠肿物,其中一个或多个标志物中的每一个是甲基化基因座,其包含选自表1中列出的DMR的差异甲基化区域(DMR)的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%)。

[0013] 另一方面,本发明涉及一种检测(例如,筛查)人类受试者的结肠直肠肿物的方法[例如,(i)将受试者分类为患有晚期腺瘤,(ii)将受试者分类为患有息肉病,(iii)将受试者分类为患有结肠直肠癌(例如,0、I、II、III期或未分化期),(iv)在确定或不确定受试者下述病症中的哪种的情况下,将受试者分类为具有晚期腺瘤、息肉病和结肠直肠癌病中的至少一种病症,或(v)在确定或不确定受试者下述病症中的哪种的情况下,将受试者分类为患有晚期腺瘤和结肠直肠癌中的至少一种病症],所述方法包括:确定人类受试者脱氧核糖核酸(DNA)中以下一项或多项中的每一个的甲基化状态:(i)具有SEQ ID NO.6的SLC6A1基因内的甲基化基因座;(ii)具有SEQ ID NO:23的基因F13A1内的甲基化基因座;(iii)具有SEQ ID NO:35的基因BARHL1内的甲基化基因座;并且至少基于所述确定的甲基化状态来诊断人类受试者的结肠直肠肿物。

[0014] 在一些实施方式中,具有SEQ ID.No:6的基因SLC6A1内的甲基化基因座包含具有

SEQ ID NO:6的SLC6A1的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%)。

[0015] 在一些实施方式中,具有SEQ ID.N0:23的基因F13A1内的甲基化基因座包含具有SEQ ID NO:23的F13A1的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%)。

[0016] 在一些实施方式中,具有SEQ ID.N0:35的基因BARHL1内的甲基化基因座包含具有SEQ ID NO:35的BARHL1的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%)。

[0017] 在一些实施方式中,所述方法还包括确定甲基化基因座的甲基化状态,所述甲基化基因座包括人类受试者脱氧核糖核酸(DNA)中的SEQ ID NO:51的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%),并且其中,诊断步骤包括至少基于包含SEQ ID NO:51的所述至少一部分的甲基化基因座确定的甲基化状态,来诊断人类受试者的结肠直肠肿物。另一方面,本发明涉及检测(例如,筛查)人类受试者的结肠直肠肿物的方法[例如,(i)将受试者分类为患有晚期腺瘤,(ii)将受试者分类为患有息肉病,(iii)将受试者分类为患有结肠直肠癌(例如,0、I、II、III期或未分化期),(iv)在确定或不确定受试者下述病症中的哪种的情况下,将受试者分类为具有晚期腺瘤、息肉病和结肠直肠癌病中的至少一种病症,或(v)在确定或不确定受试者下述病症中的哪种的情况下,将受试者分类为患有晚期腺瘤和结肠直肠癌中的至少一种病症],所述方法包括:确定人类受试者脱氧核糖核酸(DNA)中以下一项或两项中的每一个的甲基化状态:(i)具有SEQ ID NO:1的基因CSMD2内的甲基化基因座;和(ii)具有SEQ ID NO:6的基因SLC6A1内的甲基化基因座;并且至少基于所述确定的甲基化状态来诊断人类受试者的结肠直肠肿物。

[0018] 在一些实施方式中,具有SEQ ID NO:1的基因CSMD2内的甲基化基因座包含具有SEQ ID NO:1的CSMD2的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%)。

[0019] 在一些实施方式中,具有SEQ ID NO:6的基因SLC6A1内的甲基化基因座包含具有SEQ ID NO:6的SLC6A1的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%)。

[0020] 在一些实施方式中,所述方法还包括确定甲基化基因座的甲基化状态,所述甲基化基因座包括人类受试者脱氧核糖核酸(DNA)中的SEQ ID NO:25的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%),并且其中,诊断步骤包括至少基于包含SEQ ID NO:25的所述至少一部

分的甲基化基因座确定的甲基化状态来诊断人类受试者的结肠直肠肿物。

[0021] 在一些实施方式中,所述方法还包括确定甲基化基因座的甲基化状态,所述甲基化基因座包括人类受试者脱氧核糖核酸(DNA)中的SEQ ID NO:51的至少一部分(例如,至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%),并且其中,诊断步骤包括至少基于包含SEQ ID NO:51的所述至少一部分的甲基化基因座确定的甲基化状态来诊断人类受试者的结肠直肠肿物。

[0022] 在一些实施方式中,DNA分离自人类受试者的血液或血浆。

[0023] 在一些实施方式中,DNA是人类受试者的无细胞DNA。

[0024] 在一些实施方式中,使用定量聚合酶链反应(qPCR)确定甲基化状态。

[0025] 在一些实施方式中,使用大规模平行测序(例如,下一代测序)[例如,合成测序、实时(例如,单分子)测序、珠乳剂测序、纳米孔测序等]来确定甲基化状态。

[0026] 在一些实施方式中,每个甲基化基因座的长度等于或小于5000bp、4,000bp、3,000bp、2,000bp、1,000bp、950bp、900bp、850bp、800bp、750bp、700bp、650bp、600bp、550bp、500bp、450bp、400bp、350bp、300bp、250bp、200bp、150bp、100bp、50bp、40bp、30bp、20bp、或10bp。

[0027] 在一些实施方式中,所述方法包括使用下一代测序(NGS)确定一个或多个标志物中的每一个的甲基化状态。在一些实施方式中,所述方法包括使用一种或多种对目标区域富集的寡核苷酸捕获诱饵(例如,生物素化寡核苷酸探针)以捕获一个或多个相应的甲基化基因座(例如,随后进行文库制备和测序,例如,其中,样品在捕获前经过亚硫酸氢盐转化或酶促转化)。

[0028] 另一方面,本发明涉及一种用于方法(例如,如本文所述方法)的试剂盒,所述试剂盒包含一个或多个寡核苷酸引物对,用于扩增一个或多个相应的甲基化基因座。

[0029] 另一方面,本发明涉及用于(例如,在本文所述的方法中)检测(例如,筛查)结肠直肠癌的诊断性qPCR反应,该诊断性qPCR反应包括人DNA、聚合酶、用于扩增一个或多个相应的甲基化基因座的一个或多个寡核苷酸引物对,以及可选地至少一种甲基化敏感性限制酶。

[0030] 另一方面,本发明涉及试剂盒(例如,用于本文所述的方法),其包含一种或多种寡核苷酸捕获诱饵(例如,一种或多种生物素化寡核苷酸探针),用于捕获一个或多个相应的甲基化基因座(例如,用于与感兴趣的区域/区域杂交)。

[0031] 在各个方面,本发明的方法和组合物可以与本领域已知的生物标志物组合使用,例如美国专利号10,006,925和美国专利号63,011,970中所公开的。

[0032] 定义

[0033] 一种或一种:冠词“a”和“an”在本文中用于指代一个或多个(即至少一个)冠词的语法对象。例如,“一种元素”是指一种元素或多于一种的元素。

[0034] 约:术语“约”,当在本文中用于提及值时,是指在上下文中与所引用的值相似的值。一般而言,熟悉上下文的本领域技术人员将理解在该上下文中“约”所涵盖的相关差异程度。例如,在一些实施方式中,例如如本文中所述,术语“约”可涵盖在参考值的25%、20%、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、

4%、3%、2%、1%或百分之几的值的范围。

[0035] 晚期腺瘤:如本文所用,术语“晚期腺瘤”通常是指表现出相对异常、不受控制和/或自主生长的最初迹象但尚未归类为癌性改变的细胞。在结肠组织的背景下,“晚期腺瘤”是指显示以下的肿瘤生长:高级别异型增生迹象,和/或尺寸 $\geq 10\text{mm}$,和/或绒毛组织学类型,和/或锯齿状组织学类型伴有任何类型的发育异常。

[0036] 施用:如本文所用,术语“施用”通常是指将组合物施用至受试者或系统,例如以实现包括在组合物中或以其他方式由组合物递送的试剂的递送。

[0037] 试剂:如本文所用,术语“试剂”是指实体(例如,小分子、肽、多肽、核酸、脂质、多糖、复合物、组合、混合物、系统或现象,例如热、电流、电场、磁力、磁场等)。

[0038] 改善:如本文所用,术语“改善”是指受试者状态的预防、减轻、缓和或改良。改善包括但不限于要求完全康复或完全预防疾病、病症或状况。

[0039] 扩增子或扩增子分子:如本文所用,术语“扩增子”或“扩增子分子”是指通过从模板核酸分子转录产生的核酸分子,或具有与其互补的序列的核酸分子,或双链核酸,包括任何此类核酸分子。转录可以从引物开始。

[0040] 扩增:如本文所用,术语“扩增”是指使用模板核酸分子与各种试剂组合以从模板核酸分子产生另外的核酸分子,这些另外的核酸分子可以相同或与模板核酸分子的片段同一(例如,具有至少70%同一性,例如至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性)和/或是与其互补的序列。

[0041] 扩增反应混合物:如本文所用,术语“扩增反应混合物”或“扩增反应”是指模板核酸分子连同足以扩增模板核酸分子的试剂。

[0042] 生物样品:如本文所用,术语“生物样品”通常是指从感兴趣的生物来源(例如,组织或生物体或细胞培养物)获得或衍生的样品,如本文所述。在一些实施方式中,例如,如本文所述,生物来源是或包括生物体,例如动物或人。在一些实施方式中,例如,如本文所述,生物样品是或包括生物组织或流体。在一些实施方式中,例如,如本文所述,生物样品可以是或包括细胞、组织或体液。在一些实施方式中,例如,如本文所述,生物样品可以是或包括血液、血细胞、无细胞DNA、自由漂浮的核酸、腹水、活检样品、手术试样、含有细胞的体液、痰、唾液、粪便、尿液、脑脊液、腹腔液、胸水、淋巴液、妇科液、分泌物、排泄物、皮肤拭子、阴道拭子、口腔拭子、鼻拭子、冲洗液或灌洗液,例如导管灌洗液或支气管肺泡灌洗液、抽吸物、刮屑、骨髓。在一些实施方式中,例如,如本文所述,生物样品是或包括从单个受试者或多个受试者获得的细胞。样品可以是直接从生物来源获得的“原始样品”,也可以是“处理过的样品”。生物样品也可称为“样品”。

[0043] 生物标志物:如本文所用,术语“生物标志物”与其在本领域中的使用一致,是指其存在、水平或形式与特定生物事件或感兴趣的状态相关的实体,因此其被认为是成为该事件或状态的“标志物”。本领域技术人员将理解,例如,在DNA生物标志物的上下文中,生物标志物可以是或包括基因座(例如一个或多个甲基化基因座)和/或基因座的状态(例如,一个或多个甲基化基因座的状态)。仅举几个生物标志物的例子,在一些实施方式中,例如,如本文所述,生物标志物可以是或包括特定疾病、病症或状况的标志物,或者可以是特定疾病、病症或状况可以例如在受试者中发展、发生或复发的定量概率的定性标志物。在一些实施方式中,例如,如本文所述,生物标志物可以是或包括特定治疗结果的标志物,或其定量概

率的定性。因此,在各种实施方式中,例如,如本文所述,生物标志物可以预测、预后和/或诊断相关生物事件或感兴趣的状态。生物标志物可以是任何化学类别的实体。例如,在一些实施方式中,例如,如本文所述,生物标志物可以是或包括核酸、多肽、脂质、碳水化合物、小分子、无机试剂(例如,金属或离子)或其组合。在一些实施方式中,例如,如本文所述,生物标志物是细胞表面标志物。在一些实施方式中,例如,如本文所述,生物标志物是细胞内的。在一些实施方式中,例如,如本文所述,在细胞外发现生物标志物(例如,在细胞外分泌或以其他方式产生或存在,例如在体液中,例如血液、尿液、眼泪、唾液、脑脊液等)。在一些实施方式中,例如,如本文所述,生物标志物是甲基化基因座的甲基化状态。在某些情况下,例如,如本文所述,生物标志物可称为“标志物”。

[0044] 仅举一个生物标志物的实例,在一些实施方式中,例如,如本文所述,该术语指由基因编码的产物的表达,其表达是特定肿瘤、肿瘤亚类、肿瘤分期等的特征。作为另选或另外,在一些实施方式中,例如,如本文所述,特定标志物的存在或水平可以与特定信号传导途径的活性(或活性水平)相关,例如,其活性是特定类别肿瘤特征的信号传导途径。

[0045] 本领域技术人员将理解,生物标志物可以单独确定特定生物事件或感兴趣的状态,或者可以代表或有助于确定特定生物事件或感兴趣状态的统计概率。本领域技术人员将理解,标志物在其与特定生物事件或感兴趣状态相关的特异性和/或灵敏度方面可能不同。

[0046] 血液成分:如本文所用,术语“血液成分”是指全血的任何成分,包括红细胞、白细胞、血浆、血小板、内皮细胞、间皮细胞、上皮细胞和无细胞DNA。血液成分还包括血浆成分,包括蛋白质、代谢物、脂质、核酸和碳水化合物,以及可存在于血液中的任何其他细胞,例如由于怀孕、器官移植、感染、损伤或疾病而存在于血液中的任何其他细胞。

[0047] 癌症:如本文所用,术语“癌症”、“恶性肿瘤”、“赘生物”、“肿瘤”和“癌”可互换使用,指其中细胞表现出或表现了相对异常的、不受控制的和/或自主生长的疾病、病症或状况,因此它们展现出或展现了异常升高的增殖率和/或异常生长表型。在一些实施方式中,例如,如本文所述,癌症可包括一种或多种肿瘤。在一些实施方式中,例如,如本文所述,癌症可以是或包括癌前(例如良性)、恶性、转移前、转移和/或非转移的细胞。在一些实施方式中,例如,如本文所述,癌症可以是或包括实体瘤。在一些实施方式中,例如,如本文所述,癌症可以是或包括血液肿瘤。一般而言,本领域已知的不同类型癌症的实例包括例如结肠直肠癌,造血系统癌症包括白血病、淋巴瘤(霍奇金和非霍奇金)、骨髓瘤和骨髓增生性疾病;肉瘤、黑色素瘤、腺瘤、实体组织癌、口腔、咽喉、喉癌和肺癌的鳞状细胞癌、肝癌、泌尿生殖系统癌症如前列腺癌、宫颈癌、膀胱癌、子宫癌和子宫内膜癌、以及肾细胞癌、骨癌、胰腺癌、皮肤癌、皮肤或眼内黑色素瘤、内分泌系统癌、甲状腺癌、甲状旁腺癌、头颈癌、乳癌、胃肠癌和神经系统癌、良性病变等如乳头状瘤等。

[0048] 化学治疗剂:如本文所用,术语“化学治疗剂”与其在本领域中的使用一致,是指一种或多种已知用于治疗癌症或有助于治疗癌症的或具有用于治疗癌症或有助于治疗癌症的已知特征的试剂。特别地,化学治疗剂包括促凋亡剂、细胞生长抑制剂和/或细胞毒性剂。在一些实施方式中,例如,如本文所述,化学治疗剂可以是或包括烷化剂、蒽环类、细胞骨架破坏剂(例如,微管靶向部分,例如紫杉烷、美登素及其类似物)、埃坡霉素、组蛋白脱乙酰酶抑制剂HDAC)、拓扑异构酶抑制剂(例如抑制剂拓扑异构酶I和/或拓扑异构酶II)、激酶抑制

剂、核苷酸类似物或核苷酸前体类似物、肽抗生素、铂类药物、维甲酸类、长春花生物碱了和/或具有相关抗增殖活性的类似物。在某些特定实施方式中,例如,如本文所述,化学治疗剂可以是或包括放线菌素、全反式视黄酸、奥瑞他汀、阿扎胞苷、硫唑嘌呤、博来霉素、硼替佐米、卡铂、卡培他滨、顺铂、氯霉素、环磷酰胺、姜黄素、阿糖胞苷、柔红霉素、多西紫杉醇、多西氟尿苷、多柔比星、表柔比星、埃坡霉素、依托泊苷、氟尿嘧啶、吉西他滨、羟基脲、伊达比星、伊马替尼、伊立替康、美登素和/或其类似物(如DM1)、氮芥、巯嘌呤、甲氨蝶呤、米托蒽醌、美坦素、奥沙利铂、紫杉醇、培美曲塞、泰尼泊苷、硫鸟嘌呤、托泊替康、戊柔比星、长春碱、长春新碱、长春地辛、长春瑞滨或其组合。在一些实施方式中,例如,如本文所述,可以在抗体-药物偶联物的上下文中使用化学治疗剂。在一些实施方式中,例如,如本文所述,化学治疗剂是在抗体-药物缀合物中发现的一种,其选自以下组成的组:hLL1-多柔比星、hRS7-SN-38、hMN-14-SN-38、hLL2-SN-38、hA20-SN-38、hPAM4-SN-38、hLL1-SN-38、hRS7-Pro-2-P-Dox、hMN-14-Pro-2-P-Dox、hLL2-Pro-2-P-Dox、hA20-Pro-2-P-Dox、hPAM4-Pro-2-P-Dox、hLL1-Pro-2-P-Dox、P4/D10-多柔比星、吉妥珠单抗奥佐米星、本妥昔单抗、曲妥珠单抗、奥英妥珠单抗、格巴妥莫单抗(glembatumomab vedotin)、SAR3419、SAR566658、BIIB015、BT062、SGN-75、SGN-CD19A、AMG-172、AMG-595、BAY-94-9343、ASG-5ME、ASG-22ME、ASG-16M8F、MDX-1203、MLN-0264、anti-PSMA ADC、RG-7450、RG-7458、RG-7593、RG-7596、RG-7598、RG-7599、RG-7600、RG-7636、ABT-414、IMGN-853、IMGN-529、玛汀-沃瑟妥珠单抗(vorsetuzumab mafodotin)和莫星-洛沃妥珠单抗(lorvotuzumab mertansine)。在一些实施方式中,例如,如本文所述,化学治疗剂可以是或包含法呢基-硫代水杨酸(FTS)、4-(4-氯-2-甲基苯氧基)-N-羟基丁酰胺(CMH)、雌二醇(E2)、四甲氧基芪(TMS)、 δ -生育三烯酚、盐霉素或姜黄素。

[0049] 联合疗法:如本文所用,术语“联合疗法”是指向受试者施用两种或更多种试剂或方案,使得两种或更多种试剂或方案一起治疗受试者的疾病、病症或状况。在一些实施方式中,例如,如本文所述,两种或更多种治疗剂或方案可以同时、顺序或以重叠给药方案施用。本领域技术人员将理解,联合疗法包括但不要求将两种试剂或方案在单一组合中一起施用,也不是同时施用。

[0050] 可比较:如本文所用,术语“可比较”是指两个以上条件、环境、试剂、实体、群体等的组,它们可能彼此不相同,但足够相似以允许在之间进行比较,使得本领域技术人员将理解可以基于观察到的差异或相似性合理地得出结论。在一些实施方式中,例如,如本文所述,条件、环境、试剂、实体、群体等的可比较组通常以多个基本上相同的特征和零个、一个或多个不同的特征为特征。本领域普通技术人员将理解,在上下文中,需要何种程度的同一性才能使组中的成员是可比较的。例如,本领域普通技术人员将理解,当以足够数量和类型的基本上相同的特征为特征以保证观察到的差异可以全部或部分归因于其不相同的特征的合理结论。

[0051] 可检测部分:如本文所用,术语“可检测部分”是指可检测的任何元素、分子、官能团、化合物、片段或其他部分。在一些实施方式中,例如,如本文所述,可检测部分单独提供或使用。在一些实施方式中,例如,如本文所述,提供和/或利用与另一试剂相关联(例如,连接至另一试剂)的可检测部分。可检测部分的实例包括但不限于各种配体、放射性同位素(例如 ^3H 、 ^{14}C 、 ^{18}F 、 ^{19}F 、 ^{32}P 、 ^{35}S 、 ^{135}I 、 ^{125}I 、 ^{123}I 、 ^{64}Cu 、 ^{187}Re 、 ^{111}In 、 ^{90}Y 、 $^{99\text{m}}\text{Tc}$ 、 ^{177}Lu 、 ^{89}Zr 等)、荧光染

料、化学发光剂、生物发光剂、光谱可分辨的无机荧光半导体纳米晶体(即量子点)、金属纳米颗粒、纳米团簇、顺磁性金属离子、酶、比色标记、生物素、地高辛(dioxigenin)、半抗原和可获得抗血清或单克隆抗体的蛋白质。

[0052] 诊断:如本文所用,术语“诊断”是指确定受试者是否患有或将发展疾病、病症、状况或状态,和/或定量概率的定性。例如,在癌症的诊断中,诊断可以包括关于癌症的风险、类型、阶段、恶性度或其他分类的确定。在某些情况下,例如,如本文所述,诊断可以是或包括与预后和/或对一种或多种一般或特定治疗剂或方案的可能响应有关的确定。

[0053] 诊断信息:如本文所用,术语“诊断信息”是指可用于提供诊断的信息。诊断信息可以包括但不限于生物标志物状态信息。

[0054] 差异甲基化:如本文所用,术语“差异甲基化”描述了甲基化状态在第一条件和第二条件之间不同的甲基化位点。差异甲基化的甲基化位点可称为差异甲基化位点。在某些情况下,例如,如本文所述,DMR由通过使用寡核苷酸引物扩增产生的扩增子定义,例如,选择用于扩增DMR或扩增存在于扩增子中的感兴趣的DNA区域的一对寡核苷酸引物。在某些情况下,例如,如本文所述,DMR被定义为由一对寡核苷酸引物扩增的DNA区域,包括具有寡核苷酸引物的序列或与寡核苷酸引物互补的序列的区域。在某些情况下,例如,如本文所述,DMR被定义为由一对寡核苷酸引物扩增的DNA区域,不包括具有寡核苷酸引物的序列或与寡核苷酸引物互补的序列的区域。如本文所用,具体提供的DMR可以通过相关基因的名称后接起始位置的三位数字来明确标识,使得例如在ALK的位置29921434处开始的DMR可以标识为ALK'434。如本文所用,具体提供的DMR可以通过染色体编号后接DMR的开始和终止位置来明确标识。例如,在表1中标识为具有uid 9_132579614_132579683的DMR也可以标识为chr9:132579614-132579683。

[0055] 差异甲基化区域:如本文所用,术语“差异甲基化区域”(DMR)是指包含一个或多个差异甲基化位点的DNA区域。在选定的感兴趣的条件下,例如癌症状态,包括更多数量或频率的甲基化位点的DMR可以被称为高甲基化DMR。在选定的感兴趣的条件下,例如癌症状态,包括较少数量或频率的甲基化位点的DMR可以被称为低甲基化DMR。作为结肠直肠癌甲基化生物标志物的DMR可称为结肠直肠癌DMR。在某些情况下,例如,如本文所述,DMR可以是单个核苷酸,该单个核苷酸是甲基化位点。在某些情况下,例如,如本文所述,DMR具有至少10、至少15、至少20、至少30、至少50或至少75个碱基对的长度。在某些情况下,例如,如本文所述,DMR的长度等于或小于5000bp、4,000bp、3,000bp、2,000bp、1,000bp、950bp、900bp、850bp、800bp、750bp、700bp、650bp、600bp、550bp、500bp、450bp、400bp、350bp、300bp、250bp、200bp、150bp、100bp、50bp、40bp、30bp、20bp或10bp(例如,使用定量聚合酶链反应(qPCR)确定甲基化状态,例如,甲基化敏感的限制酶定量聚合酶链反应(MSRE-qPCR))。在某些情况下,例如,如本文所述,作为晚期腺瘤的甲基化生物标志物的DMR也可用于鉴定结肠直肠癌。

[0056] DNA区:如本文所用,“DNA区”是指较大DNA分子的任何连续部分。本领域技术人员将熟悉用于确定第一DNA区和第二DNA区是否对应的技术,例如基于第一DNA区和第二DNA区的序列相似性(例如,序列同一性或同源性)和/或上下文(例如,第一DNA区和第二DNA区上游和/或下游的核酸的序列同一性或同源性)。

[0057] 除非本文另有说明,在人类中发现的或与人类相关的序列(例如,与人类DNA杂交的序列)发现于、基于和/或衍生自实例代表性的人基因组序列,其通常称为和本领域技术

人员已知为智人(人类)基因组组装GRCh38、hg38和/或基因组参考联盟Human Build 38。本领域技术人员将进一步理解, hg38的DNA区可以通过已知系统提及, 所述已知系统包括根据指定的编号鉴定特定核苷酸位置或其范围。

[0058] 给药方案: 如本文所用, 术语“给药方案”可以指一组施用于受试者的一个或多个相同或不同的单位剂量, 通常包括多个单位剂量施用, 其中每个单位剂量施用与其他单位剂量施用间隔一段时间。在各种实施方式中, 例如, 如本文所述, 给药方案的一个或多个或所有单位剂量可以相同或可以变化(例如, 随时间增加、随时间减少, 或根据受试者和/或医生的决定进行调整)。在各种实施方式中, 每个剂量之间的一个或多个或所有时间段可以相同或可以变化(例如, 随时间增加、随时间减少或根据受试者和/或医生的决定进行调整)。在一些实施方式中, 例如, 如本文所述, 给定的治疗剂具有推荐的给药方案, 其可以涉及一个或多个剂量。通常, 市售药物的至少一个推荐给药方案是本领域技术人员已知的。在一些实施方式中, 例如, 如本文所述, 当跨相关人群施用时, 给药方案与期望的或有益的结果相关(即, 是治疗性给药方案)。

[0059] 下游: 如本文所用, 术语“下游”是指第一DNA区相对于第二DNA区更靠近包括第一DNA区和第二DNA区的核酸的C末端。

[0060] 基因: 如本文所用, 术语“基因”是指例如在染色体中的单个DNA区, 其包括编码产物(例如, RNA产物和/或多肽产物)的编码序列, 连同所有、一些或没有有助于调节编码序列表达的DNA序列。在一些实施方式中, 例如, 如本文所述, 基因包括一个或多个非编码序列。在一些特定实施方式中, 例如, 如本文所述, 基因包括外显子和内含子序列。在一些实施方式中, 例如, 如本文所述, 基因包括一种或多种调控元件, 其例如可以控制或影响基因表达的一个或多个方面(例如, 细胞类型特异性表达、诱导性表达等)。在一些实施方式中, 例如, 如本文所述, 基因包括启动子。在一些实施方式中, 例如, 如本文所述, 基因包括(i)在编码序列上游延伸预定数量的核苷酸的DNA核苷酸和(ii)在编码序列下游延伸预定数量的核苷酸的DNA核苷酸中的一者或两者。在各种实施方式中, 例如, 如本文所述, 预定数量的核苷酸可以是500bp、1kb、2kb、3kb、4kb、5kb、10kb、20kb、30kb、40kb、50kb、75kb或100kb。

[0061] 同源性: 如本文所用, 术语“同源性”是指聚合性分子之间, 例如核酸分子(例如, DNA分子和/或RNA分子)之间和/或多肽分子之间的总体相关性。本领域技术人员将理解同源性可以定义为例如百分比同一性或百分比同源性(序列相似性)。在一些实施方式中, 例如, 如本文所述, 如果聚合性分子的序列具有至少25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或99%同一性, 则其被认为是具有“同源性”。在一些实施方式中, 例如, 如本文所述, 如果聚合物分子的序列具有至少25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或99%相似性, 则其被认为是具有“同源性”。

[0062] 杂交: 如本文所用, “杂交”是指第一核酸与第二核酸结合以形成双链结构, 该结合通过核苷酸的互补配对发生。本领域技术人员将认识到, 互补序列尤其可以杂交。在各种实施方式中, 例如, 如本文所述, 杂交可发生在例如具有至少70%互补性, 例如至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%互补性的核苷酸序列之间。本领域技术人员将进一步理解第一核酸和第二核酸的杂交是否发生可取决于各种反应条件。可以发生杂交的条件是本领域已知的。

[0063] 低甲基化:如本文所用,术语“低甲基化”是指与参考状态相比,在感兴趣的状态中甲基化核苷酸至少一个少一个的甲基化基因座的状态(例如,与健康对照相比在结肠直肠癌中甲基化核苷酸至少少一个)。

[0064] 高甲基化:如本文所用,术语“高甲基化”是指与参考状态相比在感兴趣的状态中甲基化核苷酸至少一个多一个的甲基化基因座的状态(例如,与健康对照相比在结肠直肠癌中甲基化核苷酸至少多一个)。

[0065] 同一性、同一的:如本文所用,术语“同一性”和“同一的”是指聚合性分子之间,例如核酸分子(例如,DNA分子和/或RNA分子)之间和/或多肽分子之间的总体相关性。计算两个提供的序列之间的百分比同一性的方法是本领域已知的。两个核酸或多肽序列的百分比同一性的计算,例如,可以通过比对两个序列(或一个或两个序列的互补序列)以实现最佳比较目的(例如,可以在第一序列和第二序列中的一个或两个中引入缺口以用于最佳比对,以及出于比较目的可以忽略不相同的序列)。然后比较相应位置处的核苷酸或氨基酸。当第一序列中的一个位置被与第二个序列中的相应位置相同的残基(例如,核苷酸或氨基酸)占据时,则分子在该位置是相同的。两个序列之间的百分比同一性是序列共享的相同位置数量的函数,并且可选地考虑缺口的数量和每个缺口的长度,可能需要引入其以实现两个序列的最佳比对。序列的比较和两个序列之间同一性百分比的确定可以使用计算算法来完成,例如BLAST(基本局部比对搜索工具)。

[0066] “改善”、“增加”或“减少”:如本文所用,这些术语或语法上可比较的比较术语表示相对于可比较的参考测量的值。例如,在一些实施方式中,例如,如本文所述,相对于使用可比较的参考试剂或没有试剂获得的评估值,使用感兴趣的试剂实现的评估值可以“提高”。作为选择或另外,在一些实施方式中,例如,如本文所述,相对于在不同条件下或在不同时间点(例如,在诸如施用感兴趣的试剂等事件之前或之后)在相同受试者或系统中获得的评估值,或在不同的比比较的受试者中(例如,在可比较的受试者或系统中,其与感兴趣的受试者或系统不同之处在于存在一种或多种特定疾病、感兴趣的障碍或状况,或之前接触过某种状况或试剂等),在感兴趣受试者或系统中的评估值可以“提高”。在一些实施方式中,例如,如本文所述,比较性术语指的是统计相关的差异(例如,足以实现统计相关性的普遍性和/或量级的差异)。本领域技术人员将意识到或将能够容易地确定在给定上下文中需要或足以实现此类统计显著性的差异的程度和/或普遍性。

[0067] 甲基化:如本文所用,术语“甲基化”包括在(i)胞嘧啶的C5位置;(ii)胞嘧啶的N4位置;(iii)腺嘌呤的N6位置中的任何位置处的甲基化。甲基化还包括(iv)其他类型的核苷酸甲基化。甲基化的核苷酸可称为“甲基化核苷酸”或“甲基化核苷酸碱基”。在一些实施方式中,例如,如本文所述,甲基化特指胞嘧啶残基的甲基化。在某些情况下,甲基化特指存在于CpG位点中的胞嘧啶残基的甲基化。

[0068] 甲基化测试:如本文所用,术语“甲基化测试”是指可用于确定甲基化基因座的甲基化状态的任何技术。

[0069] 甲基化生物标志物:如本文所用,术语“甲基化生物标志物”是指生物标志物,其是或包括至少一个甲基化基因座和/或至少一个甲基化基因座的甲基化状态,例如高甲基化基因座。特别地,甲基化生物标志物是一种生物标志物,其特征在于一个或多个核酸基因座的甲基化状态在第一状态和第二状态之间(例如,在癌性状态和非癌性状态之间)的变化。

[0070] 甲基化基因座:如本文所用,术语“甲基化基因座”是指包含至少一个差异甲基化区域的DNA区。在选定的感兴趣的条件下,例如癌症状态,包括更多数量或频率的甲基化位点的甲基化基因座可以被称为高甲基化基因座。在选定的感兴趣的条件下,例如癌症状态,包括较少数量或频率的甲基化位点的甲基化基因座可以被称为低甲基化基因座。在某些情况下,例如,如本文所述,甲基化基因座具有至少10、至少15、至少20、至少30、至少50或至少75个碱基对的长度。在某些情况下,例如,如本文所述,甲基化基因座的长度小于5000bp、4,000bp、3,000bp、2,000bp、1,000bp、950bp、900bp、850bp、800bp、750bp、700bp、650bp、600bp、550bp、500bp、450bp、400bp、350bp、300bp、250bp、200bp、150bp、100bp、50bp、40bp、30bp、20bp、或10bp(例如,使用定量聚合酶链反应(qPCR)确定甲基化状态,例如甲基化敏感的限制酶定量聚合酶链反应(MSRE-qPCR))。

[0071] 甲基化位点:如本文所用,甲基化位点是指在至少一种条件下被甲基化的核苷酸或核苷酸位置。在其甲基化状态下,甲基化位点可称为甲基化的位点。

[0072] 甲基化状态:如本文所用,“甲基化状态”、“甲基化态”或“甲基化谱”是指甲基化基因座内甲基化位点处甲基化的数量、频率或模式。因此,第一状态和第二状态之间甲基化状态的变化可以是或包括甲基化位点的数量、频率或模式的增加,或者可以是或包括甲基化位点的数量、频率或模式的减少。在各种情况下,甲基化状态的变化是甲基化值的变化。

[0073] 甲基化值:如本文所用,术语“甲基化值”是指甲基化状态的数字表示,例如,以代表甲基化基因座甲基化频率或比率的数字形式。在某些情况下,例如,如本文所述,甲基化值可以通过包括在用甲基化依赖性限制酶对样品进行限制性消化后定量样品中存在的完整核酸的量的方法产生。在某些情况下,例如,如本文所述,甲基化值可以通过包括在样品的亚硫酸氢盐反应之后比较扩增谱的方法产生。在某些情况下,例如,如本文所述,可以通过比较亚硫酸氢盐处理的和未处理的核酸的序列来产生甲基化值。在某些情况下,例如,如本文所述,甲基化值是或包括或基于定量PCR结果。

[0074] 核酸:如本文所用,在其最广泛的意义上,术语“核酸”是指被引入或可以引入寡核苷酸链中的任何化合物和/或物质。在一些实施方式中,例如,如本文所述,核酸是通过磷酸二酯键引入或可以引入寡核苷酸链中的化合物和/或物质。从上下文中将清楚,在一些实施方式中,例如,如本文所述,术语核酸指单体核酸残基(例如,核苷酸和/或核苷),并且在一些实施方式中,例如,如本文所述,指包含多个单体核酸残基的多核苷酸链。核酸可以是或包括DNA、RNA或其组合。核酸可包括天然核酸残基、核酸类似物和/或合成残基。在一些实施方式中,例如,如本文所述,核酸包括天然核苷酸(例如,腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷)。在一些实施方式中,例如,如本文所述,核酸是或包括一种或多种核苷酸类似物(例如,2-氨基腺苷、2-硫胸苷、肌苷、吡咯-嘧啶、3-甲基腺苷、5-甲基胞苷、C-5丙炔基-胞苷、C-5-丙炔基-尿苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔基-尿苷、C5-丙炔基-胞苷、C5-甲基胞苷、2-氨基腺苷、7-脱氮腺苷、7-脱氮鸟苷8-氧代腺苷、8-氧代鸟苷、0(6)-甲基鸟嘌呤、2-硫胞苷、甲基化碱基、嵌入碱基及其组合)。

[0075] 在一些实施方式中,例如,如本文所述,核酸具有编码功能基因产物例如RNA或蛋白质的核苷酸序列。在一些实施方式中,例如,如本文所述,核酸包括一个或多个内含子。在一些实施方式中,例如,如本文所述,核酸包括一个或多个基因。在一些实施方式中,例如,

如本文所述,核酸通过从天然来源分离、通过基于互补模板的聚合酶促合成(体内或体外)、在重组细胞或系统中繁殖和化学合成中的一种或多种来制备。

[0076] 在一些实施方式中,例如,如本文所述,核酸类似物与核酸的不同之处在于它不利用磷酸二酯骨架。例如,在一些实施方式中,例如,如本文所述,核酸可以包括一种或多种肽核酸,其是本领域已知的并且在主链中具有肽键而不是磷酸二酯键。作为选择或另外,在一些实施方式中,例如,如本文所述,核酸具有一个或多个硫代磷酸酯和/或5'-N-亚磷酰胺键而不是磷酸二酯键。在一些实施方式中,例如,如本文所述,与天然核酸中的那些相比,核酸包含一种或多种经修饰的糖(例如,2'-氟核糖、核糖、2'-脱氧核糖、阿拉伯糖和己糖)。

[0077] 在一些实施方式中,例如,如本文所述,核酸是或包括至少3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、110、120、130、140、150、160、170、180、190、200、225、250、275、300、325、350、375、400、425、450、475、500、600、700、800、900、1000、1500、2000、2500、3000、3500、4000、4500、5000或更多个残基。在一些实施方式中,例如,如本文所述,核酸部分或完全是单链的,或者部分或完全是双链的。

[0078] 核酸检测测试:如本文所用,术语“核酸检测测试”是指确定感兴趣核酸的核苷酸组成的任何方法。核酸检测测试包括但不限于DNA测序方法、基于聚合酶链式反应的方法、探针杂交方法、连接酶链反应等。

[0079] 核苷酸:如本文所用,术语“核苷酸”是指例如DNA和/或RNA聚合物的多核苷酸的结构组分或构件。核苷酸包括碱基(例如,腺嘌呤、胸腺嘧啶、尿嘧啶、鸟嘌呤或胞嘧啶)和糖分子和至少一个磷酸基团。如本文所用,核苷酸可以是甲基化的核苷酸或未甲基化的核苷酸。本领域技术人员将理解,核酸术语,例如“基因座”或“核苷酸”可以指单个核酸分子的基因座或核苷酸和/或指基因座的累积群体或代表基因座或核苷酸的多个核酸(例如样品和/或受试者的代表中的多个核酸)中的核苷酸(例如,具有相同的相同核酸序列和/或核酸序列背景,或具有基本上相同的核酸序列和/或核酸背景)。

[0080] 寡核苷酸引物:如本文所用,术语寡核苷酸引物或引物是指使用、能够使用或用于从模板核酸分子产生扩增子的核酸分子。在允许转录的条件下(例如,在核苷酸和DNA聚合酶的存在下,以及在合适的温度和pH值下),寡核苷酸引物可以提供从与寡核苷酸引物杂交的模板的转录起始点。通常,寡核苷酸引物是长度为5至200个核苷酸的单链核酸。本领域技术人员将理解用于从模板核酸分子产生扩增子的最佳引物长度可随包括温度参数、引物组成和转录或扩增方法在内的条件而变化。如本文所用,一对寡核苷酸引物是指分别与模板双链核酸分子的第一链和第二链互补的一组两条寡核苷酸引物。就模板核酸链而言,一对寡核苷酸引物的第一成员和第二成员可分别称为“正向”寡核苷酸引物和“反向”寡核苷酸引物,因为正向寡核苷酸引物能够和与模板核酸链互补的核酸链杂交,反向寡核苷酸引物能够与模板核酸链杂交,并且正向寡核苷酸引物相对于模板核酸链的位置为反向寡核苷酸引物序列相对于模板核酸链的位置的5'。本领域技术人员将理解,第一寡核苷酸引物和第二寡核苷酸引物分别作为正向寡核苷酸引物和反向寡核苷酸引物的鉴定是任意的,因为这些标识符取决于给定的核酸链或其互补物是否被用作模板核酸分子。

[0081] 重叠:术语“重叠”在本文中用于指代DNA的两个区域,每个区域包含一个子序列,该子序列与另一个区域中相同长度的子序列基本同一(例如,DNA的两个区域有一个共同的子序列)。“基本同一”是指两个相同长度的子序列相差少于给定数量的碱基对。在某些情况

下,例如,如本文所述,每个子序列具有至少20个碱基对的长度,彼此之间的差异少于4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性或至少99.5%的相似性)。在某些情况下,例如,如本文所述,每个子序列具有至少24个碱基对的长度,其差异少于5、4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性或至少99.5%的相似性)。在某些情况下,例如,如本文所述,每个子序列具有至少50个碱基对的长度,其差异少于10、9、8、7、6、5、4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性,或至少99.5%的相似性)。某些情况下,例如,如本文所述,每个子序列具有至少100个碱基对的长度,其差异少于20、15、10、9、8、7、6、5、4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性或至少99.5%的相似性)。某些情况下,例如,如本文所述,每个子序列具有至少200个碱基对的长度,其差异少于40、30、20、15、10、9、8、7、6、5、4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性或至少99.5%的相似性)。某些情况下,例如,如本文所述,每个子序列具有至少250个碱基对的长度,其差异少于50、40、30、20、15、10、9、8、7、6、5、4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性或至少99.5%的相似性)。某些情况下,例如,如本文所述,每个子序列具有至少300个碱基对的长度,其差异少于60、50、40、30、20、15、10、9、8、7、6、5、4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性或至少99.5%的相似性)。某些情况下,例如,如本文所述,每个子序列具有至少500个碱基对的长度,其差异少于100、60、50、40、30、20、15、10、9、8、7、6、5、4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性或至少99.5%的相似性)。某些情况下,例如,如本文所述,每个子序列具有至少1000个碱基对的长度,其差异少于200、100、60、50、40、30、20、15、10、9、8、7、6、5、4、3、2或1个碱基对(例如,两个子序列具有至少80%、至少85%、至少90%、至少95%的相似性、至少97%的相似性、至少98%的相似性、至少99%的相似性或至少99.5%的相似性)。在某些情况下,例如,如本文所述,两个DNA区域的第一区域的子序列可以包含两个DNA区域的第二区域的全部(或反之亦然)(例如,共同的子序列可能包含一个或两个区域的全部)。

[0082] 药物组合物:如本文所用,术语“药物组合物”是指其中活性剂与一种或多种药学上可接受的载体一起配制的组合物。在一些实施方式中,例如,如本文所述,活性剂以适合施用于受试者的单位剂量量存在,例如,在当施用于相关人群时显示达到预定治疗效果的统计学显著概率的治疗方案中。在一些实施方式中,例如,如本文所述,药物组合物可以配制用于以特定形式(例如,以固体形式或液体形式)施用,和/或可以特别适用于,例如:口服施用(例如,作为药液)(水溶液或非水溶液或混悬液)、片剂、胶囊剂、大丸剂(bolus)、粉剂、颗粒剂、糊剂等,其可专门配制用于例如口腔、舌下或全身吸收);肠胃外施用(例如,通过皮

下、肌内、静脉内或硬膜外注射,例如无菌溶液或混悬液,或缓释制剂等);局部应用(例如,作为乳膏、软膏、贴剂或喷雾剂应用于例如皮肤、肺或口腔);阴道内或直肠内施用(例如,作为子宫托、栓剂、乳膏或泡沫);眼部施用;鼻或肺施用等。

[0083] 药学上可接受的:如本文所用,术语“药学上可接受的”适用于用于配制本文公开的组合物的一种或多种或所有组分,是指每种组分必须与组合物的其他成分相容。组合物并且对其接收者无害。

[0084] 药学上可接受的载体:如本文所用,术语“药学上可接受的载体”是指促进配制和/或改变试剂例如药物试剂的生物利用度的药学上可接受的材料、组合物或载体,例如液体或固体填充剂、稀释剂、赋形剂或溶剂包封材料。可用作药学上可接受的载体的材料的一些实例包括:糖类,例如乳糖、葡萄糖和蔗糖;淀粉,如玉米淀粉和马铃薯淀粉;纤维素及其衍生物,如羧甲基纤维素钠、乙基纤维素和醋酸纤维素;黄蓍胶粉;麦芽;明胶;滑石;赋形剂,如可可脂和栓剂蜡;油类,如花生油、棉籽油、红花油、芝麻油、橄榄油、玉米油和豆油;二醇类,如丙二醇;多元醇,如甘油、山梨糖醇、甘露糖醇和聚乙二醇;酯类,如油酸乙酯、月桂酸乙酯;琼脂;缓冲剂,如氢氧化镁和氢氧化铝;海藻酸;无热原水;等渗盐水;林格氏溶液;乙醇;pH缓冲溶液;聚酯、聚碳酸酯和/或聚醚;和其他用于药物制剂的无毒相容性物质。

[0085] 息肉病综合征:本文所用的术语“息肉病”和“息肉病综合征”是指遗传性疾病,其包括但不限于家族性腺瘤性息肉病(FAP)、遗传性非息肉病性结肠直肠癌(HNPCC)/林奇综合征、加德纳综合征、Turcot综合征、MUTYH息肉病、Peutz-Jeghers综合征、Cowden病、家族性幼年性息肉病和增生性息肉病。在一些实施方式中,息肉病包括锯齿状息肉病综合征。锯齿状息肉病的分类依据是受试者在乙状结肠近端有5个以上锯齿状息肉,其中两个以上大小至少10mm,在锯齿状息肉病家族史的背景下在乙状结肠近端有一个锯齿状息肉,并且/或整个结肠有20个以上锯齿状息肉。

[0086] 预防或防止:如本文所用的与疾病、病症或状况的发生相关的术语“预防”和“防止”是指降低发生疾病、病症或状况的风险;延迟疾病、病症或状况的发作;延迟疾病、病症或状况的一种或多种特征或症状的发作;和/或降低疾病、病症或状况的一种或多种特征或症状的频率和/或严重性。预防可以指对特定受试者的预防或对受试者群体的统计影响。当疾病、病症或状况的发作延迟了预定的时间段时,可以认为预防是完全的。

[0087] 探针:如本文所用,术语“探针”是指能够与互补靶标杂交并包括可检测部分的单链或双链核酸分子。在一些实施方式中,例如,如本文所述,探针是限制性消化产物或合成产生的核酸,例如通过重组或扩增产生的核酸。在某些情况下,例如,如本文所述,探针是可用于检测、鉴定和/或分离靶序列例如基因序列的捕获探针。在各种情况下,例如,如本文所述,探针的可检测部分可以是例如酶(例如ELISA,以及基于酶的组织化学分析)、荧光部分、放射性部分或与发光信号相关的部分。

[0088] 预后:如本文所用,术语“预后”是指确定至少一种可能的未来结果或事件的定量概率的定性。如本文所用,预后可以是对受试者中疾病、病症或状况例如癌症的可能进程的确定,关于受试者预期寿命的确定,或关于对治疗(例如特殊疗法)的响应的确定。

[0089] 预后信息:如本文所用,术语“预后信息”是指可用于提供预后的信息。预后信息可包括但不限于生物标志物状态信息。

[0090] 启动子:如本文所用,“启动子”可以指直接或间接(例如,通过启动子结合的蛋白

质或物质)与RNA聚合酶结合并参与编码序列转录起始的DNA调控区。

[0091] 参考:如本文所用,描述了相对于其进行比较的标准或对照。例如,在一些实施方式中,例如,如本文所述,将感兴趣的试剂、受试者、动物、个体、群体、样品、序列或值与参考或对照试剂、受试者、动物、个体、群体、样品、序列或值进行比较。在一些实施方式中,例如,如本文所述,与感兴趣样品中的特征的测试或确定基本上同时地测试和/或确定参考或其特征。在一些实施方式中,例如,如本文所述,参考是历史参考,可选地体现在有形媒介中。通常,如本领域技术人员将理解的,参考是在与评估中的那些可比较的条件或环境下确定或表征,例如,关于样品。当存在足够的相似性以证明对特定可能参考或对照的依赖和/或比较是合理的时,本领域技术人员将理解。

[0092] 风险:如本文所用,关于疾病、病症或状况,术语“风险”是指特定个体发展疾病、病症或状况的定量概率(无论以百分比或其他方式表示)的定性。在一些实施方式中,例如,如本文所述,风险表示为百分比。在一些实施方式中,例如,如本文所述,风险是等于或大于0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100%的定量概率的定性。在一些实施方式中,例如,如本文所述,风险被表达为相对于参考风险或水平的风险或归因于参考的相同结果的风险的定量水平的定性。在一些实施方式中,例如,如本文所述,相对风险与参考样品相比增加或减少了的系数为1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、3、4、5、6、7、8、9、10或更多。

[0093] 样品:如本文所用,术语“样品”通常是指从感兴趣的来源获得或衍生的材料的等分试样。在一些实施方式中,例如,如本文所述,感兴趣的来源是生物或环境来源。在一些实施方式中,例如,如本文所述,样品是直接从感兴趣的来源获得的“原始样品”。在一些实施方式中,例如,如本文所述,从上下文中将清楚,术语“样品”是指通过处理原始样品(例如,通过去除一种或多种成分和/或通过一种或多种试剂添加到原始样品中)获得的制剂。这样的“处理过的样品”可以包括例如从样品中提取的或通过使原始样品经受诸如核酸的扩增或逆转录、某些成分的分离和/或纯化等技术获得的细胞、核酸或蛋白质,等等。

[0094] 在某些情况下,例如,如本文所述,处理过的样品可以是已被扩增(例如,预扩增)的DNA样品。因此,在各种情况下,例如,如本文所述,鉴定的样品可以指样品的原始形式或样品的加工形式。在某些情况下,例如,如本文所述,酶消化的DNA样品可以指初级酶消化的DNA(酶消化的直接产物)或进一步加工的样品,例如酶消化的DNA,其已经经过扩增步骤(例如,中间扩增步骤,例如预扩增)和/或过滤步骤、纯化步骤或修饰样品的步骤以促进进一步的步骤,例如在确定甲基化状态(例如,DNA的原始样品的甲基化状态和/或存在于其原始来源背景中的DNA的原始样品的甲基化状态)。

[0095] 筛查:如本文所用,术语“筛查”是指旨在生成诊断信息和/或预后信息的任何方法、技术、过程或任务。因此,本领域技术人员将理解,术语筛查涵盖确定个体是否患有、可能患有或发展、或有风险患有或发展疾病、病症或状况,例如,结肠直肠癌的方法、技术、过程或任务。

[0096] 特异性:如本文所用,生物标志物的“特异性”是指以不存在感兴趣的事件或状态为特征的样品的百分比,对此,生物标志物的测量准确地表明不存在感兴趣的事件或状态(真阴性率)。在各种实施方式中,例如,如本文所述,阴性样品的表征不依赖于生物标志物,并且可以通过任何相关测量来实现,例如本领域技术人员已知的任何相关测量。因此,特异

性反映了当在未表征感兴趣的事件或状态的样品中测量时生物标志物检测不存在感兴趣的事件或状态的概率。在其中感兴趣的事件或状态是结肠直肠癌的特定实施方式中,例如,如本文所述,特异性是指生物标志物检测没有结肠直肠癌的受试者中没有结肠直肠癌的概率。例如,可以通过组织学来确定没有结肠直肠癌。

[0097] 灵敏度:如本文所用,生物标志物的“灵敏度”是指以存在感兴趣的事件或状态为特征的样品的百分比,对此,生物标志物的测量准确地表明存在感兴趣的事件或状态(真阳性率)。在各种实施方式中,例如,如本文所述,阳性样品的表征不依赖于生物标志物,并且可以通过任何相关测量来实现,例如本领域技术人员已知的任何相关测量。因此,灵敏度反映了当在以存在感兴趣的事件或状态为特征的样品中测量时,生物标志物将检测到感兴趣的事件或状态存在的概率。在其中感兴趣的事件或状态是结肠直肠癌的特定实施方式中,例如,如本文所述,灵敏度是指生物标志物检测患有结肠直肠癌的受试者中是否存在结肠直肠癌的概率。结肠直肠癌的存在可以例如通过组织学来确定。

[0098] 实体瘤:如本文所用,术语“实体瘤”是指包括癌细胞在内的异常组织块。在各种实施方式中,例如,如本文所述,实体瘤是或包括不包含囊肿或液体区域的异常组织块。在一些实施方式中,例如,如本文所述,实体瘤可以是良性的;在一些实施方式中,例如,如本文所述,实体瘤可以是恶性的。实体瘤的例子包括癌、淋巴瘤和肉瘤。在一些实施方式中,例如,如本文所述,实体瘤可以是或包括肾上腺、胆管、膀胱、骨、脑、乳腺、子宫颈、结肠、子宫内膜、食道、眼睛、胆囊、胃肠道、肾、喉、肝、肺、鼻腔、鼻咽、口腔、卵巢、阴茎、垂体、前列腺、视网膜、唾液腺、皮肤、小肠、胃、睾丸、胸腺、甲状腺、子宫、阴道和/或外阴肿瘤。

[0099] 癌症分期:如本文所用,术语“癌症分期”是指对癌症进展水平的定性或定量评估。在一些实施方式中,例如,如本文所述,用于确定癌症分期的标准可以包括但不限于癌症在身体中的位置、肿瘤大小、癌症是否已经扩散到淋巴结、癌症是否已经扩散到身体的一个或多个不同部位等。在一些实施方式中,例如,如本文所述,可以使用所谓的TNM系统对癌症进行分期,根据该系统,T是指主要肿瘤的大小和范围,通常称为原发肿瘤;N是指附近有癌症的淋巴结数量;M是指癌症是否已经转移。在一些实施方式中,例如,如本文所述,癌症可被称为0期(存在异常细胞但尚未扩散到附近组织,也称为原位癌或CIS;CIS不是癌症,但它可以变成癌症)、I-III期(存在癌症;数字越大,肿瘤越大,扩散到附近组织越多),或IV期(癌症已扩散到身体的远处部位)。在一些实施方式中,例如,如本文所述,癌症可被指定为选自以下组成的组的阶段:原位(存在异常细胞但未扩散至附近组织);局部(癌症仅限于它开始的地方,没有扩散的迹象);区域性(癌症已经扩散到附近的淋巴结、组织或器官);远处(癌症已经扩散到身体的远处部位);和未知(没有足够的信息来鉴定癌症分期)。

[0100] 对……易感:对疾病、病症或状况“易感”的个体有风险发展该疾病、病症或状况。在一些实施方式中,例如,如本文所述,对疾病、病症或状况易感的个体不表现出该疾病、病症或状况的任何症状。在一些实施方式中,例如,如本文所述,对疾病、病症或状况易感的个体尚未被诊断出患有该疾病、病症和/或状况。在一些实施方式中,例如,如本文所述,对疾病、病症或状况易感的个体是已经暴露于与疾病、病症或状况的发展相关的状况或呈现与疾病、病症或状况的发展相关的生物标志物状态(例如甲基化状态)的个体。在一些实施方式中,例如,如本文所述,发展疾病、病症和/或状况的风险是基于人群的风险(例如,患有疾病、病症或状况的个体的家庭成员)。

[0101] 受试者:如本文所用,术语“受试者”是指生物体,通常是哺乳动物(例如,人)。在一些实施方式中,例如,如本文所述,受试者患有疾病、病症或状况。在一些实施方式中,例如,如本文所述,受试者对疾病、病症或状况易感。在一些实施方式中,例如,如本文所述,受试者表现出疾病、病症或状况的一个或多个症状或特征。在一些实施方式中,例如,如本文所述,受试者未患有疾病、病症或状况。在一些实施方式中,例如,如本文所述,受试者不表现出疾病、病症或状况的任何症状或特征。在一些实施方式中,例如,如本文所述,受试者是具有对疾病、病症或状况的易感性或风险为特征的一种或多种特性的人。在一些实施方式中,例如,如本文所述,受试者是患者。在一些实施方式中,例如,如本文所述,受试者是已对其进行诊断和/或已对其进行治疗的个体。在某些情况下,例如,如本文所述,人类受试者可以互换地称为“个体”。

[0102] 治疗剂:如本文所用,术语“治疗剂”是指当施用于受试者时引起所需药理学作用的任何试剂。在一些实施方式中,例如,如本文所述,如果一种试剂在合适的人群中表现出统计学上显著的效果,则该试剂被认为是治疗剂。在一些实施方式中,例如,如本文所述,合适的群体可以是模型生物体群体或人类群体。在一些实施方式中,例如,如本文所述,合适的群体可以通过各种标准定义,例如特定年龄组、性别、遗传背景、预先存在的临床状况等。在一些实施方式中,例如,如本文所述,治疗剂是可用于治疗疾病、病症或状况的物质。在一些实施方式中,例如,如本文所述,治疗剂是在其可上市用于施用于人类之前已经或需要由政府机构批准的试剂。在一些实施方式中,例如,如本文所述,治疗剂是需要医学处方才能施用于人类的试剂。

[0103] 治疗有效量:如本文所用,术语“治疗有效量”是指产生所施用的所需效果的量。在一些实施方式中,例如,如本文所述,该术语是指当根据治疗给药方案向患有或易患疾病、病症或状况的群体施用足量以治疗该疾病、病症或状况的量。本领域普通技术人员将理解,术语治疗有效量实际上并不需要在特定个体中实现成功治疗。相反,治疗有效量可以是当施用于需要此类治疗的个体时在大量受试者中提供特定所需药理学响应的量。在一些实施方式中,例如,如本文所述,提及治疗有效量可以是提及在一种或多种特定组织(例如,受疾病、病症或状况影响的组织)或流体(例如,血液、唾液、血清、汗水、眼泪、尿液等)中测量的量。本领域普通技术人员将理解,在一些实施方式中,治疗有效量的特定试剂可以以单一剂量配制和/或施用。在一些实施方式中,例如,如本文所述,治疗有效的试剂可以以多个剂量配制和/或施用,例如,作为多剂量给药方案的一部分。

[0104] 治疗:如本文所用,术语“治疗”(也称为“治疗”或“治疗”)是指施用部分或完全减轻、改善、缓解、抑制、延迟特定疾病、病症或状况的发作、降低特定疾病、病症或状况的严重程度和/或降低特定疾病、病症或状况的一种或多种症状、特征和/或原因的发生率,或为实现任何此类结果的目的而施用。在一些实施方式中,例如,如本文所述,此类治疗可以针对不表现出相关疾病、病症或状况的迹象的受试者和/或仅表现出疾病、病症或状况的早期征兆的受试者。作为选择或另外,此类治疗可以针对表现出相关疾病、病症和/或状况的一种或多种确定迹象的受试者。在一些实施方式中,例如,如本文所述,治疗可以针对已被诊断患有相关疾病、病症和/或状况的受试者。在一些实施方式中,例如,如本文所述,治疗可以针对已知具有一种或多种易感因素的受试者,所述一种或多种易感因素与相关疾病、病症或状况的发展风险增加在统计学上相关。在各种实例中,治疗是针对癌症的。

[0105] 上游:如本文所用,术语“上游”是指第一DNA区相对于第二DNA区更靠近包括第一DNA区和第二DNA区的核酸的N末端。

[0106] 单位剂量:如本文所用,术语“单位剂量”是指作为单一剂量和/或以药物组合物的物理离散单位施用的量。在许多实施方式中,例如,如本文所述,单位剂量包含预定量的活性剂。在一些实施方式中,例如,如本文所述,单位剂量包含完整的单一剂量的试剂。在一些实施方式中,例如,如本文所述,施用多于一个单位剂量以实现总单一剂量。在一些实施方式中,例如,如本文所述,需要或预期需要施用多个单位剂量,以实现指定效果。单位剂量可以例如是一定体积的含有预定量的一种或多种治疗部分的液体(例如,可接受的载体)、预定量的一种或多种固体形式的治疗部分、缓释制剂或包含预定量的一种或多种治疗部分的递送装置,等等。应当理解,单位剂量可以存在于包括除治疗剂之外的多种组分中的任一种的制剂中。例如,可以包括可接受的载体(例如,药学上可接受的载体)、稀释剂、稳定剂、缓冲剂、防腐剂等。本领域技术人员将理解,在许多实施方式中,例如,如本文所述,特定治疗剂的总适当日剂量可以包括一部分或多个单位剂量,并且可以例如由执业医师在合理的医学判断范围内决定。在一些实施方式中,例如,如本文所述,任何特定受试者或生物体的特定有效剂量水平可取决于多种因素,包括所治疗的病症和病症的严重程度;所用特定活性化合物的活性;采用的特定成分;受试者的年龄、体重、一般健康状况、性别和饮食;施用时间,使用的特定活性化合物的排泄率;治疗持续时间;与使用的特定化合物组合或同时使用的药物和/或其他疗法,以及医学领域众所周知的类似因素。

[0107] 未甲基化:如本文所用,术语“未甲基化”和“非甲基化”可互换使用,是指鉴定的DNA区不包括甲基化核苷酸。

[0108] 变体:如本文所用,术语“变体”是指与参考实体显示显著结构同一性但与参考实体相比在一个或多个化学部分的存在、不存在或水平上与参考实体在结构上不同的实体。在一些实施方式中,例如,如本文所述,变体在功能上也不同于其参考实体。一般而言,特定实体是否被适当地视为参考实体的“变体”取决于其与参考实体的结构同一性程度。变体可以是与参考相当但不完全相同的分子。例如,变体核酸可以在核苷酸序列的一个或多个差异处不同于参考核酸。在一些实施方式中,例如,如本文所述,变体核酸显示与参考核酸的总体序列同一性为至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%或99%。在许多实施方式中,例如,如本文所述,如果感兴趣的核酸具有与参考的序列相同但在特定位置有少量序列改变的序列,则感兴趣的核酸被认为是参考核酸的“变体”。在一些实施方式中,例如,如本文所述,与参考相比,变体具有10、9、8、7、6、5、4、3、2或1个取代的残基。在一些实施方式中,例如,如本文所述,与参考相比,变体具有不超过5、4、3、2或1个残基添加、取代或缺失。在各种实施方式中,例如,如本文所述,添加、取代或缺失的数目少于约25、约20、约19、约18、约17、约16、约15、约14、约13、约10、约9、约8个、约7个、约6个,并且通常少于约5个、约4个、约3个或约2个残基。

[0109] 附图描述

[0110] 本发明的上述及其他目的、方面、特征及优点,通过以下结合附图的说明,将会变得更加明显易懂,其中:

[0111] 图1是显示基于4-区域组合的疾病组样品分布的箱形图。穿过图表的水平线表示90%特异性线。

[0112] 图2是表示所示病症的检测的灵敏度的图表。

[0113] 图3是显示基于4DMR组合的疾病组样品分布的箱形图。穿过图表的水平线表示90%特异性线。

[0114] 图4A-F是显示表示对应于所示病症的样品中DMR区域的甲基化百分比的箱形图。

具体实施方式

[0115] 据设想,要求保护的发明的系统、架构、设备、方法和过程包括使用来自本文所述的实施方式的信息开发的变化和适配。如本说明书所设想的,可以执行本文所述的系统、架构、设备、方法和过程的适配和/或修改。

[0116] 在整个说明书中,如果物品、设备、系统和架构被描述为具有、包括或包含特定组件,或者过程和方法被描述为具有、包括或包括特定步骤,则可以预期的是,另外,本发明的物品、设备、系统和架构基本上由或由所列举的组件组成,并且存在基本上由或由所列举的处理步骤组成的根据本发明的过程和方法。

[0117] 应当理解的是,步骤的顺序或执行特定动作的顺序并不重要,只要本发明保持可操作即可。此外,可以同时进行两个以上步骤或动作。

[0118] 本文中提及的任何出版物,例如,在背景部分中,并不承认该出版物作为本文提出的任何权利要求的现有技术。背景部分是出于清楚目的而呈现的,并不意味着作为关于任何权利要求的现有技术的描述。

[0119] 结肠直肠肿物筛查

[0120] 需要包括但不限于晚期腺瘤、息肉病和/或结肠直肠癌的结肠直肠肿物的改进的检测(例如,筛查)方法。这包括了筛查早期结肠直肠癌的需求。尽管有对个体(例如45岁以上)进行筛查的建议,但结肠直肠癌筛查计划通常无效或不能令人满意。改进的结肠直肠肿物筛查可改善诊断并降低结肠直肠癌死亡率。

[0121] DNA甲基化(例如,高甲基化或低甲基化)可以激活或失活基因,其包括影响肿物(包括癌症)发展的基因。因此,例如,高甲基化可以使一种或多种基因失活,这些基因通常起到抑制癌症的作用,导致或促进样品或受试者中癌症的发展。因此,在人类受试者的DNA样品中测量的来自表1、5和/或6的一种或多种标记物的高甲基化(增加的甲基化)状态指示癌前肿瘤(例如晚期腺瘤和/或息肉病)。

[0122] 本公开内容包括以下发现:确定本文提供的一个或多个甲基化基因座的甲基化状态和/或本文提供的一个或多个DMR的甲基化状态提供了对包括但不限于晚期腺瘤、息肉病和/或结肠直肠癌(例如,早期结肠直肠癌)的结肠直肠肿物的检测(例如,筛查)。在一些实施方式中,筛查可以以高度敏感和/或特异性地将受试者分类为患有结肠直肠肿物,即受试者患有一种或多种病症或不患有一种或多种病症。本公开提供了包括或涉及结肠直肠肿物生物标志物的组合物和方法,所述标志物包括晚期腺瘤、息肉病和/或结肠直肠癌甲基化生物标志物,其单独地或在包含两种以上甲基化生物标志物的各种组中,提供对晚期腺瘤、息肉病和/或结肠直肠癌(例如,早期结肠直肠癌)的筛查。

[0123] 在各种实施方式中,用于检测包括晚期腺瘤、息肉病和/或结肠直肠癌的结肠直肠肿物的本公开的甲基化生物标志物选自是表1中所列DMR或包括表1中所列DMR的至少一部分的甲基化基因座。表1列出了发现DMR的DNA区域,其中包括染色体编号(chr)、DMR在染色

体上的起始和终止位置,以及已知与该区域相关的基因(如果有)。如果目前已知没有基因与该区域相关,则术语“NA”列在相关基因列中。对每个DMR还提供了唯一的识别号(uid),可用于明确标识DMR。此外,还列出了DMR区域的大小(例如,长度) (“区域的大小”)。

[0124] 表1:与对照组相比,发现在结肠直肠癌和/或晚期腺瘤患者的血液和/或组织中甲基化模式发生显著改变的DMR列表。

SEQ ID NO	相关基因	chr	起始	终止	区域的大小
SEQ ID NO: 1	CSMD2	1	34165443	34165675	233
SEQ ID NO: 2	MRPS21	1	150294178	150294265	31
SEQ ID NO: 3	FCGR2A	1	161539915	161540181	267
SEQ ID NO: 4	SLC30A10	1	219928356	219928386	88
SEQ ID NO: 5	MSH2,KCNK12	2	47569627	47569858	232
SEQ ID NO: 6	SLC6A1	3	10993689	10993900	268
SEQ ID NO: 7	NA	3	43998153	43998214	104
SEQ ID NO: 8	VEPH1,PTX3	3	157437029	157437296	62
SEQ ID NO: 9	NLGN1	3	173397284	173397387	212
SEQ ID NO: 10	CCDC39,LOC101928882	3	180679674	180679706	35
SEQ ID NO: 11	SOX2,SOX2-OT	3	181712580	181712614	33
SEQ ID NO: 12	NKX3-2	4	13545274	13545403	345
[0125] SEQ ID NO: 13	PCDH7	4	30718010	30718076	105
SEQ ID NO: 14	TMEM150C	4	82562331	82562477	1133
SEQ ID NO: 15	INTU	4	127623100	127623235	169
SEQ ID NO: 16	STOX2	4	183904880	183905140	100
SEQ ID NO: 17	STOX2	4	183905237	183905728	232
SEQ ID NO: 18	LOC642366,ISL1	5	51387638	51388080	492
SEQ ID NO: 19	HOMER1	5	79512985	79513164	261
SEQ ID NO: 20	VCAN	5	83474370	83474485	130
SEQ ID NO: 21	LINC00461	5	88660609	88660834	136
SEQ ID NO: 22	NA	6	1524106	1524152	67
SEQ ID NO: 23	F13A1	6	6320205	6320663	147
SEQ ID NO: 24	NA	6	27288050	27288544	226
SEQ ID NO: 25	NA	6	27670532	27670614	116
SEQ ID NO: 26	ADGRB3,LOC101928307	6	68635305	68635560	180
SEQ ID NO: 27	NA	6	129767354	129767459	443

[0126]

SEQ ID NO: 28	C6orf118	6	165309346	165309582	106
SEQ ID NO: 29	ZNF736	7	64314242	64314346	495
SEQ ID NO: 30	NA	7	65252405	65252749	237
SEQ ID NO: 31	RGS20	8	53877760	53878892	457
SEQ ID NO: 32	RGS20	8	53881543	53881774	256
SEQ ID NO: 33	CRH	8	66177597	66177765	83
SEQ ID NO: 34	TMEM74	8	108787618	108787717	47
SEQ ID NO: 35	BARHL1	9	132579614	132579683	214
SEQ ID NO: 36	C1QL3	10	16521408	16521447	817
SEQ ID NO: 37	NA	10	22475853	22476043	743
SEQ ID NO: 38	MIR3663HG	10	117165171	117165431	38
SEQ ID NO: 39	TENM4	11	79438095	79438308	473
SEQ ID NO: 40	LOC105369438,AMOTL1	11	94740046	94740862	103
SEQ ID NO: 41	RERG	12	15221303	15221461	70
SEQ ID NO: 42	SOX5	12	24563707	24564068	191
SEQ ID NO: 43	HOXC11,HOTAIR	12	53974529	53974571	261
SEQ ID NO: 44	SYT1	12	78865146	78865306	40
SEQ ID NO: 45	PCDH20	13	61414743	61414845	161
SEQ ID NO: 46	PCDH9	13	67231493	67231965	159
SEQ ID NO: 47	GPC6	13	93228702	93229444	362
SEQ ID NO: 48	SOX21,SOX21-AS1	13	94712995	94713032	43
SEQ ID NO: 49	RTN1	14	59870236	59870262	27
SEQ ID NO: 50	AQP4-AS1,CHST9	18	27184277	27184863	587
SEQ ID NO: 51	NA	19	20052466	20053193	78
SEQ ID NO: 52	LAMP5-AS1,LAMP5	20	9516555	9516632	57
SEQ ID NO: 53	TIAM1	21	31344104	31344160	728
SEQ ID NO: 54	SHOX	X	625391	625465	75

[0127] 区分腺瘤和对照受试者的DMR的优选组合, (表5)

[0128] 表5. 四DMR的组合1。

[0129]

SEQ ID NO	基因	chr	起始	终止
SEQ ID NO:6	SLC6A1	3	10993689	10993900
SEQ ID NO:51	NA	19	20052466	20053193
SEQ ID NO:23	F13A1	6	6320205	6320663
SEQ ID NO:35	BARHL1	9	132579614	132579683

[0130] 区分结肠直肠肿物和对照受试者的DMR的优选组合, (表6)

[0131] 表6四DMR的组合2。

[0132]	SEQ ID NO	基因	chr	起始	终止
	SEQ ID NO:1	CSMD2	1	34165443	34165675
	SEQ ID NO:6	SLC6A1	3	10993689	10993900
	SEQ ID NO:25	NA	6	27670532	27670614
	SEQ ID NO:51	NA	19	20052466	20053193

[0133] 为了避免任何疑问, 本文提供的任何甲基化生物标志物尤其可以是结肠直肠肿物标志物, 或包括在结肠直肠肿物标志物中。此外, 本文的任何甲基化生物标志物可以是晚期腺瘤、息肉病和/或和/或结肠直肠癌 (例如, 早期结肠直肠癌) 甲基化生物标志物或包括在晚期腺瘤、息肉病和/或和/或结肠直肠癌 (例如, 早期结肠直肠癌) 甲基化生物标志物中。

[0134] 在一些实施方式中, 所述甲基化生物标志物可以是单个甲基化基因座或包括单个甲基化基因座。在一些实施方式中, 甲基化生物标志物可以是两个以上甲基化基因座或包括两个以上甲基化基因座。在一些实施方式中, 甲基化生物标志物可以是或包括单个差异甲基化区域 (DMR) (例如, (i) 选自表1中所列的DMR, (ii) 包含选自表1中所列的DMR的DMR, (iii) 与选自表1所列的一个或多个DMR重叠的DMR, 或 (iv) 作为选自表1所列那些的DMR的一部分的DMR。在一些实施方式中, 甲基化基因座可以是或包括两个以上DMR (例如, 两个、三个、四个或更多个选自表1中所列的DMR, 或两个、三个、四个或更多个DMR, 其中每一个都与选自表1中所列的DMR重叠和/或包含选自表1中所列的DMR)。在一些实施方式中, 甲基化生物标志物可以是或包括单个甲基化位点。在其他实施方式中, 甲基化生物标志物可以是或包括两个以上甲基化位点。在一些实施方式中, 甲基化基因座可包括两个以上DMR并且还包括与一个或多个所包括的DMR相邻的DNA区域。

[0135] 在某些情况下, 甲基化基因座是或包括基因, 如表1中提供的基因。在某些情况下, 甲基化基因座是或包括基因的一部分, 如表1中提供的基因的一部分。在某些情况下, 甲基化基因座包括但不限于识别基因的核酸边界。

[0136] 在某些情况下, 甲基化基因座是或包括基因的编码区, 如表1中提供的基因的编码区。在某些情况下, 甲基化基因座是或包括基因编码区的一部分, 例如, 表1中提供的基因编码区的一部分。在某些情况下, 甲基化基因座包括但不限于识别基因编码区的核酸边界。

[0137] 在某些情况下, 甲基化基因座是或包括基因的启动子和/或其他调节区, 如表1中提供的基因的启动子和/或其他调节区。在某些情况下, 甲基化基因座是或包括基因的启动子和/或其他调节区的一部分, 例如表1中提供的基因的启动子和/或调控区的一部分。在某些情况下, 甲基化基因座包括但不限于识别基因的启动子和/或其他调控区的核酸边界。在一些实施方式中, 甲基化基因座是或包括高CpG密度启动子或其一部分。

[0138] 在一些实施方式中, 甲基化基因座是或包括非编码序列。在一些实施方式中, 甲基化基因座是或包括一个或多个外显子和/或一个或多个内含子。

[0139] 在一些实施方式中, 甲基化基因座包括在编码序列上游延伸预定数量的核苷酸的DNA区域, 和/或在编码序列下游延伸预定数量的核苷酸的DNA区域。在各种情况下, 预定数量的核苷酸在上游和/或下游并且是或包括例如500bp、1kb、2kb、3kb、4kb、5kb、10kb、20kb、30kb、40kb、50kb、75kb、或100kb。本领域技术人员将理解, 能够影响编码序列表达的甲基化

生物标志物通常可以在编码序列、上游和/或下游的任何这些距离内。

[0140] 本领域技术人员将理解,被鉴定为甲基化生物标志物的甲基化基因座不一定需要在单个实验、反应或扩增子中进行测定。可以分析鉴定为结肠直肠癌甲基化生物标志物的单个甲基化基因座,例如,在包括单独扩增(或提供寡核苷酸引物和足以扩增的条件)甲基化基因座内的一个或多个不同或重叠DNA区域(例如,一个或多个不同或重叠的DMR)的方法中。本领域技术人员将进一步理解,不需要分析被鉴定为甲基化生物标志物的甲基化基因座的每个核苷酸的甲基化状态,也不需要分析存在于甲基化基因座内的每个CpG。相反,可以分析作为甲基化生物标志物的甲基化基因座,例如,通过分析甲基化基因座内的单个DNA区域,例如,通过分析甲基化基因座内的单个DMR。

[0141] 本公开的DMR可以是甲基化基因座或包括甲基化基因座的一部分。在某些情况下,DMR是具有甲基化基因座的DNA区域,例如,长度为1至5,000bp。在各种实施方式中,DMR是长度等于或小于5000bp、4,000bp、3,000bp、2,000bp、1,000bp、950bp、900bp、850bp、800bp、750bp、700bp、650bp、600bp、550bp、500bp、450bp、400bp、350bp、300bp、250bp、200bp、150bp、100bp、50bp、40bp、30bp、20bp、或10bp的甲基化基因座的DNA区域。在一些实施方式中,DMR长度为1、2、3、4、5、6、7、8或9bp。

[0142] 包括但不限于本文提供的甲基化基因座和DMR的甲基化生物标志物可包括至少一个作为结肠直肠肿瘤生物标志物的甲基化位点。

[0143] 为清楚起见,本领域技术人员将理解,术语甲基化生物标志物被广泛使用,使得甲基化基因座可以是包括一个或多个DMR的甲基化生物标志物,其中每个DMR本身也是甲基化生物标志物,并且每个所述DMR可以包括一个或多个甲基化位点,每个所述甲基化位点本身也是甲基化生物标志物。此外,甲基化生物标志物可包括两个或更多个甲基化基因座。因此,作为甲基化生物标志物的状态并不取决于生物标志物中包含的核酸的连续性,而是取决于包含的DNA区域在第一状态和第二状态之间的甲基化状态变化的存在,如在结肠直肠癌和对照之间。

[0144] 如本文所提供的,甲基化基因座可以是一个或多个甲基化基因座中的任一个,其中每个甲基化基因座是、包括表1中鉴定的基因(或特定DMR)或者是表1中鉴定的基因的一部分。在一些实施方式中,结肠直肠肿瘤甲基化生物标志物(例如,晚期腺瘤、息肉病和/或结肠直肠癌(例如,早期结肠直肠癌)甲基化生物标志物)包括单个甲基化基因座,其是、包括表1中鉴定的基因或者是表1中鉴定的基因的一部分。

[0145] 在一些实施方式中,甲基化生物标志物包括两个以上甲基化基因座,每个甲基化基因座是、包括表1中鉴定的基因或者是表1中鉴定的基因的一部分。在一些实施方式中,结肠直肠肿瘤甲基化生物标志物(例如,晚期腺瘤、息肉病和/或结肠直肠癌甲基化生物标志物)包括多个甲基化基因座,每个甲基化基因座是、包括表1中鉴定的基因或者是表1中鉴定的基因的一部分。

[0146] 在各种实施方式中,甲基化生物标志物可以是或包括存在于本文提供的一个或多个甲基化基因座(例如,一个或多个DMR)中的一个或多个个体核苷酸(例如,CpG情况中的单个个体半胱氨酸残基)或多个个体半胱氨酸残基(例如,多个CpG的多个单独的半胱氨酸残基)。因此,在一些实施方式中,甲基化生物标志物是或包括多个个体甲基化位点的甲基化状态。

[0147] 在各种实施方式中,甲基化生物标志物是或包括或特征在于甲基化状态的变化,所述变化是一个或多个甲基化基因座(例如,一个或多个DMR)内的一个或多个甲基化位点的甲基化的变化。在各种实施方式中,甲基化生物标志物是或包括甲基化状态的变化,所述变化是一个或多个甲基化基因座(例如,一个或多个DMR)内甲基化位点数量的变化。在各种实施方式中,甲基化生物标志物是或包括甲基化状态的变化,所述变化是一个或多个甲基化基因座(例如,一个或多个DMR)内甲基化位点频率的变化。在各种实施方式中,甲基化生物标志物是或包括甲基化状态的变化,所述变化是一个或多个甲基化基因座(例如,一个或多个DMR)内甲基化位点模式的变化。

[0148] 在各种实施方式中,一个或多个甲基化基因座(例如,一个或多个DMR)的甲基化状态表示为被甲基化的样品中存在的一个或多个甲基化基因座(例如,一个或多个DMR)的分数或百分比,例如,作为在一个或多个特定甲基化基因座(例如,一个或多个特定DMR)处甲基化的样品中DNA的单个DNA链的数量的分数。本领域技术人员将理解,在某些情况下,甲基化的分数或百分比可以从例如样品内的一个或多个分析的DMR的甲基化DMR与未甲基化DMR的比率计算。

[0149] 在各种实施方式中,将一个或多个甲基化基因座(例如,一个或多个DMR)的甲基化状态与参考甲基化状态值和/或与参考样品中的一个或多个甲基化基因座(例如,一个或多个DMR)的甲基化状态进行比较。在某些情况下,参考是来自相同来源的非同期样品,例如来自相同来源的先前样品,例如来自相同受试者。在某些情况下,一个或多个甲基化基因座(例如,一个或多个DMR)的甲基化状态的参考是已知代表特定状态(例如,癌症状态或非癌症状态)的样品(例如,来自受试者的样品)或多个样品中一个或多个甲基化基因座(例如,一个或多个DMR)的甲基化状态。因此,参考可以是或包括一个或多个预定阈值,该预定阈值可以是定量的(例如甲基化值)或定性的。在某些情况下,DMR甲基化状态的参考是在不包括DMR的核苷酸的同一样品中存在的一个核苷酸或多个核苷酸(例如,多个连续寡核苷酸)的甲基化状态。本领域技术人员将理解,参考测量通常通过使用与进行非参考测量的方法相同、相似或相当的方法进行测量来产生。

[0150] 晚期腺瘤

[0151] 在一些实施方式中,本文提供的方法和组合物可用于筛查晚期腺瘤。晚期腺瘤包括但不限于:结肠和/或直肠内的肿物性腺瘤生长、位于结肠近端的腺瘤、位于结肠和/或直肠远端的腺瘤、低级别异型增生的腺瘤、高级别异型增生的腺瘤、显示任何大小的高级别异型增生迹象的结肠直肠组织的肿物性生长、任何组织学和/或异型增生等级的尺寸大于或等于10mm的结肠直肠组织的肿物性生长、任何类型异型增生和任何尺寸的绒毛组织学类型的结肠直肠组织的肿物性生长,以及具有任何异型增生等级和/或尺寸的锯齿状组织学类型的结肠直肠组织。

[0152] 癌症

[0153] 在一些实施方式中,本公开的方法和组合物可用于结肠直肠癌筛查。结肠直肠癌包括但不限于结肠癌、直肠癌及其组合。结肠直肠癌包括转移性结肠直肠癌和非转移性结肠直肠癌。结肠直肠癌包括位于结肠癌近端的癌症和位于结肠远端的癌症。

[0154] 结肠直肠癌包括本领域已知的各种可能分期中的任何一个的结肠直肠癌,包括例如I期、II期、III期和IV期结肠直肠癌(例如0、I、IIA、IIB、IIC、IIIA、IIIB、IIIC、IVA、IVB和

IVC期)。结肠直肠癌包括肿瘤/淋巴结/转移(TNM)分期系统的所有阶段。对于结肠直肠癌,T可以指肿瘤是否生长到结肠壁还是直肠壁中,如果如此的话多少层;N可以指肿瘤是否已经扩散到淋巴结,如果如此,有多少个淋巴结以及它们位于哪里;并且M可以指癌症是否已经扩散到身体的其他部位,如果如此,已经扩散到哪些部位,扩散到什么程度。T、N和M的特定阶段是本领域已知的。T阶段可以包括TX、T0、Tis、T1、T2、T3、T4a和T4b;N阶段可以包括NX、N0、N1a、N1b、N1c、N2a和N2b;M阶段可以包括M0、M1a和M1b。此外,结肠直肠癌的等级可包括GX、G1、G2、G3和G4。癌症、尤其是结肠直肠癌分期的各种手段是本领域众所周知的,例如在万维网上的cancer.net/cancer-types/colorectal-cancer/stages上总结的。

[0155] 在某些情况下,本公开内容包括早期结肠直肠癌的筛查。早期结肠直肠癌可以包括例如位于受试者内的结肠直肠癌,例如,因为它们尚未扩散至受试者的淋巴结,例如癌附近的淋巴结(N0期),并且尚未扩散至远端位点(M0期)。早期癌症包括对应于例如0至II C期的结肠直肠癌。

[0156] 因此,本公开的结肠直肠癌尤其包括恶性前结肠直肠癌和恶性结肠直肠癌。本公开的方法和组合物可用于筛查所有形式和阶段的结肠直肠癌,包括但不限于本文命名的或本领域已知的那些,以及其所有亚组。因此,本领域技术人员将理解,此处提供的对结肠直肠癌的所有参考包括但不限于其所有形式和阶段的结肠直肠癌,包括但不限于本文命名的或本领域已知的那些,以及所有其亚组。

[0157] 息肉病

[0158] 在一些实施方式中,本公开的方法和组合物可用于筛查息肉病(例如,息肉病综合征)。

[0159] 息肉病包括导致个体更容易形成多发息肉(例如超过10个息肉)的遗传病症。这些息肉通常存在于个体的结肠和/或直肠中。许多遗传性病症可归类为息肉病综合征,其包括但不限于:家族性腺瘤性息肉病(FAP)、遗传性非息肉病性结肠直肠癌(HNPCC)/Lynch综合征、Gardner综合征、Turcot综合征、MUTYH息肉病、Peutz-Jeghers综合征、Cowden病、家族性幼年性息肉病、锯齿状息肉病综合征(SPS)和增生性息肉病。

[0160] 锯齿状息肉病综合征(SPS)是一种息肉病综合征,可通过具有以下特征的个体来识别:5个以上靠近乙状结肠的锯齿状息肉,其中两个以上具有至少10mm尺寸,在锯齿状息肉病家族史的背景下具有靠近乙状结肠的锯齿状息肉,和/或整个结肠和直肠有20个以上锯齿状息肉。

[0161] 受试者和样品

[0162] 使用本文提供的方法和组合物分析的样品可以是任何生物样品和/或任何样品,包括核酸。在各种特定实施方式中,使用本文提供的方法和组合物分析的样品可以是来自哺乳动物的样品。在各种特定实施方式中,使用本文提供的方法和组合物分析的样品可以是来自人类受试者的样品。在各种特定实施方式中,使用本文提供的方法和组合物分析的样品可以是来自小鼠、大鼠、猪、马、鸡或牛的样品。

[0163] 在各种情况下,人类受试者是被诊断或寻求诊断为患有结肠直肠肿物(例如,结肠直肠癌)的受试者,被诊断为或寻求诊断为有风险患结肠直肠肿物(例如,结肠直肠癌)的受试者,和/或被诊断为或寻求诊断为有直接风险患结肠直肠肿物(例如,结肠直肠癌)的受试者。在各种情况下,人类受试者是被鉴定为需要结肠直肠肿物(例如,结肠直肠癌)筛查的受

试者。在某些情况下,人类受试者是被鉴定为需要由执业医师进行结肠直肠癌筛查的受试者。在各种情况下,人类受试者被鉴定为由于年龄而需要结肠直肠癌筛查,例如由于年龄等于或大于45岁,例如年龄等于或大于45、50、55、60、65、70、75、80、85或90岁。尽管在某些情况下,18岁以上的受试者可能被鉴定为有结肠直肠肿物(例如,结肠直肠癌)风险和/或需要结肠直肠肿物(例如,结肠直肠癌)筛查。在各种情况下,基于但不限于家族史、既往诊断和/或执业医师的医学评估,人类受试者被鉴定为高风险和/或需要结肠直肠肿物(例如,结肠直肠癌)筛查。在各种情况下,人类受试者是未被诊断为患有癌症(例如结肠直肠癌)的受试者,没有风险患癌症(例如结肠直肠癌)的受试者,或没有直接风险患癌症(例如结肠直肠癌)的受试者,未被诊断为患有癌症(例如结肠直肠癌)的受试者,和/或没有尝试诊断癌症(例如结肠直肠癌)的受试者,或其任何组合。

[0164] 来自受试者,例如人类或其他哺乳动物受试者的样品可以是例如血液、血液成分(例如,等离子,血沉棕黄层)、cfDNA(无细胞DNA)、ctDNA(循环肿瘤DNA)、粪便、或晚期腺瘤和/或结肠直肠组织的样品。在某些特定实施方式中,样品是受试者的排泄物或体液(例如受试者的粪便、血液、血浆、淋巴液或尿液)或结肠直肠肿物的组织样品,如结肠息肉、晚期腺瘤和/或结肠直肠癌的组织样品。来自受试者的样品可以是细胞或组织样品,例如具有癌症或包括癌细胞(例如具有肿瘤或转移性组织)的细胞或组织样品。在各种实施方式中,可以通过活组织检查(例如,结肠镜切除术、细针抽吸或组织活组织检查)或手术获得来自受试者例如人或哺乳动物受试者的样品。

[0165] 在各种特定实施方式中,样品是无细胞DNA(cfDNA)样品。cfDNA通常以短的双链片段形式存在于生物体液(例如血浆、血清或尿液)中。cfDNA的浓度通常很低,但在特定条件下会显著增加,包括但不限于怀孕、自身免疫性疾病、心肌梗塞和癌症。循环肿瘤DNA(ctDNA)是特异性源自癌细胞的循环DNA的组成部分。ctDNA可以存在于人类体液中。例如,在某些情况下,可以发现ctDNA与白细胞和红细胞结合和/或相关。在某些情况下,可以发现ctDNA不与白细胞和红细胞结合和/或相关。用于检测肿瘤来源的cfDNA的各种测试基于对癌症(例如相关癌症)特征的遗传或表观遗传修饰的检测。癌症特征的遗传或表观遗传修饰可包括但不限于肿瘤抑制基因中的致癌突变或癌症相关突变、活化的致癌基因、高甲基化和/或染色体疾病。检测癌症或癌前特征的遗传或表观遗传修饰可以确认所检测的cfDNA是ctDNA。

[0166] cfDNA和ctDNA可提供源组织甲基化状态的实时或近乎实时的指标。cfDNA和ctDNA在血液中的半衰期约为2小时,因此在给定时间采集的样品可以相对及时地反映源组织的状态。

[0167] 从样品中分离核酸(例如,从血液或血浆中分离cfDNA)的各种方法是本领域已知的。核酸可以通过例如但不限于标准DNA纯化技术,通过直接基因捕获(例如,通过澄清样品以去除测试抑制剂并用捕获剂从澄清样品中捕获靶核酸(如果存在的话),以产生捕获复合物,并分离捕获复合物以回收目标核酸)来分离。

[0168] 测量甲基化状态的方法

[0169] 甲基化状态可以通过本领域已知的多种方法和/或通过本文提供的方法测量。本领域技术人员将理解,用于测量甲基化状态的方法通常可应用于来自任何来源和任何种类的样品,并且将进一步理解可用于将样品修饰成适合通过以下方法进行测量的形式的处理

步骤。测量甲基化状态的方法包括但不限于,包括全基因组亚硫酸氢盐测序、靶向亚硫酸氢盐测序、靶向酶促甲基化测序、甲基化状态特异性聚合酶链式反应(PCR)的方法,包括质谱、甲基化阵列的方法,包括甲基化特异性核酸酶的方法,包括基于质量的分离的方法,包括目标特异性捕获的方法,以及包括甲基化特异性寡核苷酸引物在内的方法。甲基化的某些特定的测试使用亚硫酸氢盐试剂(例如,亚硫酸氢根离子)或酶促转化试剂(例如,Tet甲基胞嘧啶双加氧酶2)。

[0170] 亚硫酸氢盐试剂尤其可包括亚硫酸氢盐(bisulfite)、偏重亚硫酸盐(disulfite)、亚硫酸氢盐(hydrogen sulfite)或它们的组合等,这些试剂可用于区分甲基化和未甲基化的核酸。亚硫酸氢盐与胞嘧啶和5-甲基胞嘧啶的相互作用不同。在典型的基于亚硫酸氢盐的方法中,DNA与亚硫酸氢盐接触会使未甲基化的胞嘧啶脱氨基为尿嘧啶,而甲基化的胞嘧啶不受影响,选择性保留甲基化胞嘧啶,而不是未甲基化胞嘧啶。因此,在经亚硫酸氢盐处理的样品中,尿嘧啶残基代替未甲基化的胞嘧啶残基,并因此提供未甲基化胞嘧啶残基的识别信号,而残留的(甲基化)胞嘧啶残基因此提供甲基化胞嘧啶残基的识别信号。可以例如,通过下一代测序(NGS)分析经亚硫酸氢盐处理的样品。

[0171] 酶促转化试剂可包括Tet甲基胞嘧啶双加氧酶2(TET2)。TET2氧化5-甲基胞嘧啶,从而保护它免受APOBEC的连续脱氨作用。APOBEC将未甲基化的胞嘧啶脱氨为尿嘧啶,同时氧化5-甲基胞嘧啶不受影响。因此,在TET2处理的样品中,尿嘧啶残基代替未甲基化的胞嘧啶残基,从而提供未甲基化的胞嘧啶残基的识别信号,而剩余的(甲基化的)胞嘧啶残基由此提供甲基化胞嘧啶残基的识别信号。TET2处理的样品可以进行分析,例如,通过下一代测序(NGS)进行分析。

[0172] 测量甲基化状态的方法可以包括但不限于大规模平行测序(例如,下一代测序)以确定甲基化状态,例如,合成测序、实时(例如,单分子)测序、珠乳剂测序、纳米孔测序或本领域已知的其他测序技术。在一些实施方式中,测量甲基化状态的方法可以包括全基因组测序,例如,以碱基对分辨率测量来自亚硫酸氢盐或酶促处理的材料的全基因组甲基化状态。

[0173] 在一些实施方式中,测量甲基化状态的方法包括减少代表性的亚硫酸氢盐测序,例如,利用限制酶以碱基对分辨率测量来自亚硫酸氢盐或酶促处理的材料的高CpG含量区域的甲基化状态。

[0174] 在一些实施方式中,测量甲基化状态的方法可以包括靶向测序,例如,以碱基对分辨率对来自亚硫酸氢盐或酶促处理的材料的预选基因组位置测量甲基化状态。

[0175] 在一些实施方式中,可以通过互补的体外合成寡核苷酸序列(诱饵、引物或探针)来进行对感兴趣的区域的预选(捕获)。

[0176] 在一些实施方式中,用于测量甲基化状态的方法可以包括Illumina甲基化测定,例如,以单核苷酸分辨率定量测量基因组中超过850,000个甲基化位点。

[0177] 各种甲基化测试程序可与亚硫酸氢盐处理结合使用,以确定靶序列如DMR的甲基化状态。此类测试可尤其包括甲基化特异性限制酶qPCR、经亚硫酸氢盐处理的核酸的测序、PCR(例如,使用序列特异性扩增)、甲基化特异性核酸酶辅助小等位基因富集PCR和甲基化灵敏度高分辨率溶解。在一些实施方式中,从经亚硫酸氢盐处理的DNA样品中扩增DMR,并根据例如Illumina方案或基于转座的Nextera XT方案制备DNA测序文库以用于测序。在一些

实施方式中,高通量和/或下一代测序技术用于实现DNA序列的碱基对水平解析,允许分析甲基化状态。

[0178] 另一种可用于甲基化检测的方法包括使用甲基化特异性寡核苷酸引物的PCR扩增的方法(MSP方法),例如应用于经亚硫酸氢盐处理的样品(参见例如Herman 1992Proc.Natl.Acad.Sci.USA 93:9821-9826,其关于确定甲基化状态的方法通过引用并入本文)。使用甲基化状态特异性寡核苷酸引物来扩增经亚硫酸氢盐处理的DNA可以区分甲基化和非甲基化核酸。用于MSP方法的寡核苷酸引物对包括至少一种寡核苷酸引物,其能够与包括甲基化位点(例如CpG)的序列杂交。在与胞嘧啶残基互补的位置包含T残基的寡核苷酸引物将选择性地与其中胞嘧啶在亚硫酸氢盐处理之前未甲基化的模板杂交,而在与胞嘧啶残基互补的位置包含G残基的寡核苷酸引物将在亚硫酸氢盐处理之前,与其中胞嘧啶是甲基化胞嘧啶的模板选择性杂交。MSP结果可以在使用或不使用测序扩增子的情况下获得,例如使用凝胶电泳。MSP(甲基化特异性PCR)允许使用亚硫酸氢盐转化的DNA的PCR扩增对位点特异性DNA甲基化进行高灵敏度检测(检测水平为0.1%的等位基因,具有完全特异性)。

[0179] 可用于确定样品在亚硫酸氢盐处理后的甲基化状态的另一种方法是甲基化敏感高分辨率溶解(MS-HRM)PCR(参见,例如Hussmann 2018Methods Mol Biol.1708:551-571,其关于确定甲基化状态的方法通过引用并入本文)。MS-HRM是一种基于PCR的管内方法,其可基于杂交熔解检测感兴趣的特定基因座的甲基化水平。在执行MS-HRM之前对DNA进行亚硫酸氢盐处理可确保甲基化和未甲基化DNA之间的不同碱基组成,其用于通过高分辨率熔解分离所得扩增子。独特的引物设计促进了测试的高灵敏度,能够在未甲基化背景中检测低至0.1-1%的甲基化等位基因。用于MS-HRM测试的寡核苷酸引物设计为与甲基化等位基因互补,特定的退火温度使这些引物能够与甲基化和未甲基化等位基因退火,从而提高测试的灵敏度。

[0180] 另一种可用于在亚硫酸氢盐处理样品后确定甲基化状态的方法是定量多重甲基化特异性PCR(QM-MSP)。QM-MSP使用甲基化特异性引物对DNA甲基化进行灵敏定量(参见,例如Fackler 2018Methods Mol Biol.1708:473-496,其关于确定甲基化状态的方法通过引用并入本文)。QM-MSP是一种两步PCR方法,在第一步中,一对基因特异性引物(正向和反向)在一个PCR反应中同时和多重扩增同一基因的甲基化和未甲基化拷贝。在36个PCR循环后,该甲基化独立扩增步骤可产生每 μL 高达 10^9 个拷贝的扩增子。在第二步中,使用实时PCR和两个独立的荧光团检测同一孔中每个基因的甲基化/未甲基化DNA(例如6FAM和VIC),用标准曲线对第一次反应的扩增子进行量化。在100,000个参考基因拷贝中可检测到一个甲基化拷贝。

[0181] 另一种可用于确定在亚硫酸氢盐处理样品后甲基化状态的方法是甲基化特异性核酸酶辅助小等位基因富集(MS-Name)(参见,例如,Liu 2017Nucleic Acids Res.45(6):e39,其关于确定甲基化状态的方法通过引用并入本文)。Ms-Name基于对在双链(ds)DNA(DSN)具有特异性的DNA核酸酶的存在下探针与靶序列的选择性杂交,从而杂交产生随后被DSN消化的双链DNA的区域。因此,靶向未甲基化序列的寡核苷酸探针产生局部双链区域,导致未甲基化靶标的消化;能够与甲基化序列杂交的寡核苷酸探针会产生局部双链区,其导致甲基化靶标的消化,使甲基化靶标保持完整。此外,寡核苷酸探针可以同时将DSN活性导向经亚硫酸氢盐处理的DNA中的多个靶标。随后的扩增可以富集未消化的序列。Ms-Name可

以单独使用或与本文提供的其他技术组合使用。

[0182] 另一种可用于确定在亚硫酸氢盐处理样品后甲基化状态的方法是甲基化灵敏度单核苷酸引物延伸 (Ms-SNuPE™) (参见例如Gonzalgo 2007Nat Protoc.2(8):1931-6,其关于确定甲基化状态的方法通过引用并入本文)。在Ms-SNuPE中,执行链特异性PCR以生成DNA模板,以用于使用Ms-SNuPE进行定量甲基化分析。然后用寡核苷酸进行SNuPE,该寡核苷酸设计用于在被询问的CpG位点的紧接上游进行杂交。反应产物可以在聚丙烯酰胺凝胶上电泳,以用于通过磷光体图像分析进行可视化和定量。扩增子还可以携带直接或间接可检测的标记,例如荧光标记、放射性同位素或可脱离的分子片段或具有可通过质谱法区分的质量的其他实体。检测可以通过例如基质辅助激光解吸/电离质谱 (MALDI) 或使用电子喷雾质谱 (ESI) 来进行和/或可视化。

[0183] 可用于在亚硫酸氢盐处理样品后确定甲基化状态的某些方法在基于扩增的方法中利用第一寡核苷酸引物、第二寡核苷酸引物和寡核苷酸探针。例如,寡核苷酸引物和探针可用于实时聚合酶链式反应 (PCR) 或液滴数字PCR (ddPCR) 的方法中。在各种情况下,第一寡核苷酸引物、第二寡核苷酸引物和/或寡核苷酸探针选择性杂交甲基化DNA和/或未甲基化DNA,使得扩增或探针信号指示样品的甲基化状态。

[0184] 用于检测甲基化状态 (例如,5-甲基胞嘧啶水平的存在) 的其他基于亚硫酸氢盐的方法公开于例如Frommer (1992Proc Natl Acad Sci U S A.1;89(5):1827-31,其关于确定甲基化状态的方法通过引用并入本文)。

[0185] 在某些MSRE-qPCR实施方式中,使用例如实时PCR或数字PCR在天然 (例如未消化) 形式的样品的等分试样中测量总DNA的量。

[0186] 各种扩增技术可单独使用或与本文所述的其他技术结合使用以检测甲基化状态。本领域技术人员在阅读了本说明书后将理解如何将本领域已知的和/或本文描述的各种扩增技术与本领域已知的和/或本文提供的用于甲基化状态确定的各种其他技术结合起来。扩增技术包括但不限于PCR,例如定量PCR (qPCR)、实时PCR和/或数字PCR。本领域技术人员将理解聚合酶扩增可以在单个反应中多重扩增多个靶标。PCR扩增子的长度通常为100至2000个碱基对。在各种情况下,扩增技术足以确定甲基化状态。

[0187] 基于数字PCR (dPCR) 的方法涉及将样品分配和分布在具有96孔、384孔或更多孔的板的孔中,或分配和分布在单个乳液液滴 (ddPCR) 中,例如,使用微流体装置,使得一些孔包括一个或多个模板拷贝,其他不包含模板拷贝。因此,扩增前每孔的平均模板分子数小于1。发生模板扩增的孔的数量提供了模板浓度的量度。如果样品已与MSRE接触,发生模板扩增的孔的数量提供了甲基化模板浓度的量度。

[0188] 在各种实施方式中,基于荧光的实时PCR测试,例如MethyLight™,可用于测量甲基化状态 (参见,例如Campan 2018Methods Mol Biol.1708:497-513,其甲基化状态的测试方法通过引用并入本文)。MethyLight是一种基于荧光的定量实时PCR方法,其可灵敏地检测和量化基因组候选区的DNA甲基化。MethyLight特别适合在未甲基化DNA的高背景下检测低频甲基化DNA区,因为它结合了甲基化特异性引发和甲基化特异性荧光探测。此外,MethyLight可与数字PCR结合使用,用于高度灵敏地检测个体甲基化分子,用于疾病检测和筛查。

[0189] 用于确定甲基化状态的基于实时PCR的方法通常包括基于外部标准的分析生成未

甲基化DNA的标准曲线的步骤。标准曲线可以由至少两个点构建,并且可以将消化的DNA的实时Ct值和/或未消化的DNA的实时Ct值与已知的定量标准进行比较。在特定情况下,可以确定MSRE消化和/或未消化样品或样品等分试样的样品Ct值,并且可以从标准曲线计算DNA的基因组当量。可以评估MSRE消化和未消化DNA的Ct值,以鉴定消化的扩增子(例如,有效消化;例如,产生45的Ct值)。也可以鉴定在消化或未消化条件下未扩增的扩增子。然后可以跨条件直接比较感兴趣的扩增子的校正Ct值,以确定条件之间甲基化状态的相对差异。作为选择或另外,消化和未消化DNA的Ct值之间的delta差异可用于确定条件之间甲基化状态的相对差异。

[0190] 在某些特定实施方式中,其他技术中美,全基因组亚硫酸氢盐测序可用于确定是或包括单个甲基化基因座的结肠直肠肿物(例如,晚期腺瘤、息肉病和/或结肠直肠癌(例如,早期结肠直肠癌))甲基化生物标志物的甲基化状态。在某些特定实施方式中,全基因组亚硫酸氢盐测序以及其他技术可用于确定是或包括两个以上甲基化基因座的甲基化生物标志物的甲基化状态。

[0191] 本领域技术人员将进一步理解用于全基因组亚硫酸氢盐测序的方法、试剂和方案是本领域众所周知的。与传统的全基因组测序不同,全基因组亚硫酸氢盐测序能够检测胞嘧啶核苷酸的甲基化状态,这是由于亚硫酸氢盐试剂进行了脱氨基处理。

[0192] 本领域技术人员将理解,在以本文提供的结肠直肠癌筛查方法分析多个甲基化基因座(例如,多个DMR)的甲基化状态的实施方式中,每个甲基化基因座的甲基化状态可以是以多种形式中的任一种测量或表示,并且多个甲基化基因座的甲基化状态(优选地每个以相同、相似或可比较的方式测量和/或表示)一起或累积地以各种形式的任一种分析或表示。在各种实施方式中,每个甲基化基因座的甲基化状态可以测量为甲基化部分。在各种实施方式中,每个甲基化基因座的甲基化状态可以表示为与参考样品相比,总测序读数中甲基化读数的百分比值。在各种实施方式中,每个甲基化基因座的甲基化状态可以表示为与参考的定性比较,例如通过将每个甲基化基因座鉴定为高甲基化或低甲基化。

[0193] 在其中分析单个甲基化基因座的一些实施方式中,单个甲基化基因座的高甲基化构成受试者患有或可能患有病症(例如,晚期腺瘤、息肉病和/或结肠直肠癌,例如,早期结肠直肠癌)的诊断,而不存在单甲基化基因座的高甲基化构成受试者可能不患有病症的诊断。在一些实施方式中,多个分析的甲基化基因座的单个甲基化基因座(例如单个DMR)的高甲基化构成受试者患有或可能患有病症的诊断,而在多个分析的甲基化基因座的任何甲基化基因座不存在高甲基化构成受试者可能不患有病症的诊断。在一些实施方式中,多个分析的甲基化基因座中甲基化基因座确定百分比(例如,预定百分比)(例如,至少10%(例如,至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或100%))的高甲基化构成受试者患有或可能患有病症的诊断,而不存在多个分析的甲基化基因座中甲基化基因座确定百分比(例如,预定百分比)(例如,至少10%(例如,至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或100%))的高甲基化构成受试者不可能患有病症的诊断。在一些实施方式中,多个分析的甲基化基因座(例如,2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、或54个DMR)中确定数量(例如,预定数量)的甲基化基因座(例如,至少1、

2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、或54个DMR)的高甲基化构成受试者患有或可能患有病症的诊断,而不存在多个分析的甲基化基因座(例如,2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、或54个DMR)中确定数量(例如,预定数量)的甲基化基因座(例如,至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、或54个DMR)的高甲基化构成受试者不可能患有病症的诊断。

[0194] 在一些实施方式中,定性或定量测量多个甲基化基因座(例如,多个DMR)的甲基化状态,并且组合多个甲基化基因座中的每一个的测量以提供诊断。在一些实施方式中,多个甲基化基因座中的每一个的定量测量的甲基化状态的定性被单独加权,并且加权值被组合以提供可以与参考进行比较以提供诊断的单个值。

[0195] 应用

[0196] 本公开的方法和组合物可用于多种应用中的任何一种。例如,本公开的方法和组合物可用于筛查或帮助筛查结肠直肠肿物,例如,晚期腺瘤、息肉病和/或结肠直肠癌(例如,早期结肠直肠癌)。在各种情况下,使用本公开内容的方法和组合物进行筛查可以检测任何阶段的结肠直肠癌,包括但不限于早期结肠直肠癌。在一些实施方式中,使用本公开内容的方法和组合物进行的筛查适用于45岁以上,例如45、50、55、60、65、70、75、80、85或90岁以上的个体。在一些实施方式中,使用本公开的方法和方法组合物进行的筛查适用于20岁以上,例如20、25、30、35、40、45、50、55、60、65、70、75、80、85或90岁以上的个体。在一些实施方式中,使用本公开的方法和方法组合物进行的筛查适用于20至50岁,例如20至30岁、20至40岁、20至50岁、30岁至40岁、30至50岁或40至50岁的个体。在各种实施方式中,使用本公开内容的方法和组合物进行的筛查适用于经历腹痛或不适的个体,例如,经历未诊断或不完全诊断的腹痛或不适的个体。在各种实施方式中,使用本公开内容的方法和组合物进行的筛查适用于没有可能与诸如晚期腺瘤、息肉病和/或结肠直肠癌等结肠直肠肿物相关的症状的个体。因此,在一些实施方式中,使用本公开内容的方法和组合物进行的筛查是完全或部分预防性或防止性的,至少对于晚期或非早期的结肠直肠癌而言如此。

[0197] 在各种实施方式中,使用本公开内容的方法和组合物进行的结肠直肠肿物筛查可应用于无症状的人类受试者。如本文所用,如果受试者通过非侵入性可观察标记(例如,没有一个、几个或所有基于装置的探测、组织样品分析、体液分析、手术或结肠直肠癌筛查)没有报告和/或证明病症的充分特征以支持对受试者可能患有病症的医学合理怀疑,则受试者可以称为是“无症状的”。结肠直肠肿物,诸如晚期腺瘤、息肉病和/或早期结肠直肠癌特别可能在根据本公开内容的方法和组合物筛查的无症状个体中检测到。

[0198] 本领域技术人员将理解,对诸如晚期腺瘤和/或结肠直肠癌等结肠直肠肿物的定期、预防性和/或防止性筛查改善了诊断。如上所述,根据至少一种癌症分期系统,早期癌症包括结肠直肠癌的0至IIC期。因此,本公开尤其提供了用于包括晚期腺瘤、息肉病和/或早期结肠直肠癌的结肠直肠肿物的诊断和治疗的方法和组合物。通常,特别是在根据本公开每年进行筛查和/或其中受试者在筛查时无症状的实施方式中,本发明的方法和组合物特

别有可能检测到早期结肠直肠癌。

[0199] 在各种实施方式中,根据本公开的结肠直肠癌筛查对给定受试者进行一次或对给定受试者进行多次。在各种实施方式中,根据本公开定期进行结肠直肠癌筛查,例如每六个月、每年、每两年、每三年、每四年、每五年或每十年。

[0200] 在各种实施方式中,使用本文公开的方法和组合物进行的筛查将提供病症的诊断(例如,结肠直肠肿物的类型或类别)。在其他情况下,使用本文公开的方法和组合物进行的结肠直肠肿物筛查将指示具有一种或多种病症,但对于特定病症的诊断不是确定的。例如,筛查可用于将受试者分类为具有一种或多种病症或病症的组合,所述病症包括但不限于晚期腺瘤、息肉病和/或结肠直肠癌。筛查也可用于将受试者分类为患有结肠直肠肿物而不鉴定受试者患有哪种病症。在各种情况下,使用本公开的方法和组合物进行筛查之后可以进行进一步的诊断确认测试,该进一步的诊断确认测试可以确认、支持、破坏、或拒绝由先前筛查(例如,根据本公开的筛查)产生的诊断。

[0201] 如本文所用,诊断确认测试可以是提供被执业医师确认为确定性的诊断的结肠直肠癌测试,例如基于结肠镜检查的诊断,或显著增加或降低先前诊断的可能性的结肠直肠癌测试是正确的,例如,根据本公开的筛查产生的诊断。诊断确认测试可以包括现有的筛查技术,这些技术通常需要在灵敏度、特异性和非侵入性中的一个或多个方面进行改进,特别是在检测早期结肠直肠癌方面。

[0202] 在某些情况下,诊断确认测试是一种测试,它是或包括受试者组织的视觉或结构检查,例如通过结肠镜检查。在一些实施方式中,结肠镜检查包括或之后是组织学分析。结肠直肠癌的视觉和/或结构测试可以包括检查结肠和/或直肠的结构是否有任何异常组织和/或结构。例如,可以通过直肠使用内窥镜或通过CT扫描进行视觉和/或结构检查。在某些情况下,诊断确认测试是结肠镜检查,例如,包括或随后进行组织学分析。根据某些报告,结肠镜检查是目前主要的和/或最依赖于诊断确认测试。

[0203] 另一种基于计算机断层扫描(CT)的视觉和/或结构诊断确认测试是CT结肠成像,有时也称为虚拟结肠镜检查。CT扫描利用结肠和/或直肠的大量X射线图像来生成结肠的尺寸表示。尽管可用作诊断确认测试,但一些报告表明CT结肠成像不足以替代结肠镜检查,至少部分是因为医生没有实际接触受试者的结肠来获取组织以进行组织学分析。

[0204] 另一种诊断确认测试可以是乙状结肠镜检查。在乙状结肠镜检查中,乙状结肠镜用于通过直肠对结肠和/或直肠的部分进行成像。根据一些报道,乙状结肠镜检查并未得到广泛应用。

[0205] 在某些情况下,诊断确认测试是基于粪便的测试。通常,当使用基于粪便的测试代替目视或结构检查时,建议以比使用目视或结构检查所需的频率更高的频率使用。在一些情况下,诊断确认测试是基于guiaac的粪便潜血试验或粪便免疫化学试验(gFOBTs/FITs)(参见,例如Navarro 2017World J Gastroenterol.23(20):3632-3642,其关于结肠直肠癌测试通过引用并入本文)。FOBTs和FITs有时用于诊断结肠直肠癌(参见,例如,Nakamura 2010J Diabetes Investig.10月19日;1(5):208-11,其关于结肠直肠癌测试通过引用并入本文)。FIT是基于粪便中潜血的检测,潜血的存在通常预示着结肠直肠癌,但通常量不足以通过肉眼进行识别。例如,在典型的FIT中,该测试利用血红蛋白特异性试剂来测试粪便样品中的潜血。在各种情况下,FIT试剂盒适合个人在家中使用。当在没有其他诊断确认测试

的情况下使用时,建议每年使用一次FIT。通常不依赖FIT为结肠直肠癌的结论性诊断提供足够的诊断信息。

[0206] 诊断确认测试还包括gFOBT,其经设计以通过化学反应检测粪便中的潜血。与FIT一样,当在没有其他诊断确认测试的情况下使用时,可能建议每年使用gFOBT。通常不依赖gFOBT为结肠直肠癌的结论性诊断提供足够的诊断信息。

[0207] 诊断确认测试还可以包括粪便DNA检测。结肠直肠癌的粪便DNA检测可经设计以鉴定粪便样品中癌症的DNA序列特征。如果在没有其他诊断确认测试的情况下使用,建议每三年使用一次粪便DNA检测。通常不依赖粪便DNA检测来为结肠直肠癌的结论性诊断提供足够的诊断信息。

[0208] 一种特别的筛查技术是基于粪便的筛查测试(Cologuard® (Exact Sciences Corporation, Madison, WI, United States)),其将FIT分析与DNA的异常修饰(例如突变和甲基化)分析相结合。Cologuard®测试与单独的FIT检测相比,显示出改善的灵敏度,但由于依从率低,临床上可能不切实际或无效,所述低依从率至少部分是由于受试者不喜欢使用基于粪便的检测(参见,例如,doi:10.1056/NEJMc1405215(例如,2014N Engl J Med.371(2):184-188))。Cologuard®测试似乎将几乎一半的合格人群排除在筛查计划之外(参见,例如, van der Vlugt 2017Br J Cancer.116(1):44-49)。如本文提供的筛查的使用(例如通过基于血液的分析),将增加选择筛查结肠直肠癌的个体数量(参见,例如Adler 2014BMC Gastroenterol.14:183;Liles 2017Cancer Treatment and Research Communications 10:27-31)。就目前所知,只有一种现有的结肠直肠癌筛查技术Epirocolon获得FDA批准和CE-IVD准售,并且是基于血液的。Epirocolon基于SEPT9基因的高甲基化。Epirocolon测试的结肠直肠癌检测的准确度低,灵敏度为68%,晚期腺瘤灵敏度仅为22%(参见,例如, Potter 2014Clin Chem.60(9):1183-91)。本领域尤其需要一种非侵入性结肠直肠癌筛查,其可以实现高受试者依从性,并且具有高和/或改进的特异性和/或灵敏度。

[0209] 在各种实施方式中,根据本公开的方法和组合物的筛查降低了结肠直肠癌死亡率,例如通过早期结肠直肠癌诊断。数据支持结肠直肠癌筛查降低结肠直肠癌死亡率,这种影响持续了30多年(参见,例如,Shaukat 2013N Engl J Med.369(12):1106-14)。此外,结肠直肠癌特别难以治疗,至少部分是因为没有及时筛查的结肠直肠癌可能直到癌症过了早期阶段才被检测到。至少出于这个原因,结肠直肠癌的治疗通常是不成功的。为了使结肠直肠癌结果的全人群改善最大化,根据本公开的筛查的利用可以与例如合格受试者的招募配对以确保广泛筛查。

[0210] 在各种实施方式中,包括一种或多种本文公开的方法和/或组合物的结肠直肠肿瘤筛查之后是结肠直肠癌的治疗,例如早期结肠直肠癌的治疗。在各种实施方式中,结肠直肠癌例如早期结肠直肠癌的治疗包括施用治疗方案,所述治疗方案包括手术、放射疗法和化学疗法中的一种或多种。在各种实施方式中,结肠直肠癌例如早期结肠直肠癌的治疗包括施用包括本文提供的一种或多种治疗的治疗方案,以用于治疗0期结肠直肠癌、I期结肠直肠癌和/或II期结肠直肠癌。

[0211] 在各种实施方式中,结肠直肠癌的治疗包括通过以下中的一种或多种来治疗早期结肠直肠癌,例如0期结肠直肠癌或I期结肠直肠癌:手术切除癌组织(例如通过局部切除

(例如通过结肠镜)、部分结肠切除术或全结肠切除术)。

[0212] 在各种实施方式中,结肠直肠癌的治疗包括通过以下中的一种或多种来治疗早期结肠直肠癌,例如II期结肠直肠癌:手术切除癌组织(例如,通过局部切除(例如,通过结肠镜)、部分结肠切除术或全结肠切除术)、手术切除已识别的结肠直肠癌组织附近的淋巴结,以及化学疗法(例如,施用5-FU和亚叶酸、奥沙利铂或卡培他滨中的一种或多种)。

[0213] 在各种实施方式中,结肠直肠癌的治疗包括通过以下中的一种或多种来治疗III期结肠直肠癌:手术切除癌组织(例如,通过局部切除(例如,通过基于结肠镜检查的切除)、部分结肠切除术或全结肠切除术)、手术切除已识别的结肠直肠癌组织附近的淋巴结,以及化学疗法(例如,施用5-FU和亚叶酸、奥沙利铂或卡培他滨中的一种或多种,例如以下述组合:(i) 5-FU和亚叶酸,(ii) 5-FU、亚叶酸和奥沙利铂(例如,FOLFOX),或(iii) 卡培他滨和奥沙利铂(例如,CAPEOX)) 和放射疗法。

[0214] 在各种实施方式中,结肠直肠癌的治疗包括通过以下中的一种或多种来治疗IV期结肠直肠癌:手术切除癌组织(例如,通过局部切除(例如,通过基于结肠镜检查的切除)、部分结肠切除术或全结肠切除术)、手术切除已识别的结肠直肠癌组织附近的淋巴结,手术切除转移灶、化学疗法(例如,施用5-FU、亚叶酸、奥沙利铂、卡培他滨、伊立替康、VEGF靶向治疗剂(例如,贝伐单抗、阿柏西普、或雷莫芦单抗)、EGFR靶向治疗剂(例如,西妥昔单抗或帕尼单抗)、瑞戈非尼、三氟尿苷和替吡嘧啶中的一种或多种,例如以下述组合:(i) 5-FU和亚叶酸,(ii) 5-FU、亚叶酸和奥沙利铂(例如,FOLFOX), (iii) 卡培他滨和奥沙利铂(例如,CAPEOX)、以及(v) 氟尿嘧啶和替吡嘧啶(Lonsurf))、放射疗法、肝动脉输注(例如,如果癌已转移至肝脏)、肿瘤消融、肿瘤栓塞、结肠支架、结肠切除术、结肠造口术(例如,分流结肠造口术) 和免疫疗法(例如,派姆单抗)。

[0215] 本领域技术人员本文提供的结肠直肠癌的治疗可以例如如由执业医师确定的那样,单独或以任何组合,以任何顺序、方案和/或治疗程序使用。本领域技术人员将进一步理解,高级治疗选择可能适用于先前患有癌症或结肠直肠癌的受试者中的早期癌症,例如,被诊断为患有复发性结肠直肠癌的受试者。

[0216] 在一些实施方式中,本文提供的用于结肠直肠肿物筛查的方法和组合物可以告知例如由个体,医疗保健机构、医疗保健从业者、医疗保险提供者、政府机构或对医疗保健费用感兴趣的其他各方做出的治疗和/或支付(例如,医疗保健的费用的报销或减少,例如筛查或治疗) 决定和/或行动。

[0217] 在一些实施方式中,本文提供的用于结肠直肠肿物筛查的方法和组合物可以告知与健康保险提供者是否偿付医疗保健费用支付者或接受者(或不)有关的决策,例如,用于(1) 筛查本身(例如,报销筛查,除非不可用,仅适用于定期/定期筛查,或仅适用于临时和/或偶然动机的筛查);和/或用于(2) 治疗,包括例如,基于筛查结果启动、维持和/或改变治疗。例如,在一些实施方式中,本文提供的用于结肠直肠肿物筛查的方法和组合物用作基础、有助于或支持确定是否将向医疗保健费用支付者或接受者提供报销或成本降低。在一些情况下,寻求报销或降低成本的一方可以提供根据本说明书进行的筛查的结果以及对医疗保健费用的这种报销或成本降低的请求。在某些情况下,作出关于是否提供医疗费用的报销或费用降低的决定的一方将全部或部分基于接收和/或审查根据本说明书进行的筛查的结果来作出决定。

[0218] 为避免任何疑问,本领域技术人员将从本公开中理解,本说明书的用于结肠直肠癌诊断的方法和组合物至少用于体外使用。因此,本公开的所有方面和实施方式可以至少在体外进行和/或使用。

[0219] 试剂盒

[0220] 本公开内容尤其包括试剂盒,该试剂盒包括一种或多种如本文提供的用于筛查的组合物,可选地与其在筛查(例如,结肠直肠癌肿物,例如,晚期腺瘤和/或结肠直肠癌(例如,早期结肠直肠癌)的筛查)中的使用说明书组合。在各种实施方式中,用于结肠直肠肿物筛查的试剂盒可包括一种或多种寡核苷酸捕获诱饵(例如,一个或多个生物素化寡核苷酸探针)。在一些实施方式中,用于筛查的试剂盒可选地包括一种或多种本文公开的亚硫酸氢盐试剂。在一些实施方式中,用于筛查的试剂盒可选地包括一种或多种如本文所公开的酶促转化试剂。

[0221] 寡核苷酸捕获诱饵可用于下一代测序(NGS)技术,以靶向特定的感兴趣DNA区域。在一些实施方式中,一种或多种捕获诱饵被靶向以捕获对应于一个或多个甲基化基因座的感兴趣DNA区域(例如,甲基化基因座,其包含一个或多个DMR,例如,如表1、5和/或6所示DMR的至少一部分)。寡核苷酸捕获诱饵旨在富集目标DNA区域,并有助于DNA文库的制备。富集的目标区域然后将使用例如本文讨论的NGS测序技术进行测序。

[0222] 在各种实施方式中,用于筛查的试剂盒可以包括以下的一种或多种:一种或多种寡核苷酸引物(例如,一个或多个寡核苷酸引物对)、一种或多种MSRE、一种或多种用于qPCR的试剂(例如,足以完成qPCR反应混合物的试剂,包括但不限于dNTP和聚合酶),以及用于结肠直肠肿物筛查的试剂盒的一种或多种成分的使用说明书。在各种实施方式中,用于结肠直肠肿物筛查的试剂盒可包括以下中的一种或多种:一种或多种寡核苷酸引物(例如,一个或多个寡核苷酸引物对)、一种或多种亚硫酸氢盐试剂、一种或多种用于qPCR的试剂(例如,足以完成qPCR反应混合物的试剂,包括但不限于dNTP和聚合酶),以及用于结肠直肠癌筛查的试剂盒的一种或多种成分的使用说明书。

[0223] 在一些实施方式中,本公开的试剂盒包括用于扩增如本文公开的甲基化基因座和/或DMR的至少一个寡核苷酸引物对(例如,在表1、5和/或6中)。

[0224] 在某些情况下,本公开的试剂盒包括用于扩增本公开的一个或多个甲基化基因座的一个或多个寡核苷酸引物对。在某些情况下,本公开的试剂盒包括用于扩增一个或多个甲基化基因座的一个或多个寡核苷酸引物对,所述一个或多个甲基化基因座是或包括表1、5和/或6中鉴定的一个或多个基因的全部或部分。在某些特定情况下,本公开的试剂盒包括包括用于多个甲基化基因座的寡核苷酸引物对,每个甲基化基因座是或包括表1中鉴定的基因的全部或部分,多个甲基化基因座包括,例如2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、或54个甲基化基因座,例如,如表1、5和/或6所提供的。

[0225] 在某些情况下,本公开的试剂盒包括用于扩增本公开的一个或多个DMR的一个或多个寡核苷酸引物对。在某些情况下,本公开的试剂盒包括用于扩增一个或多个DMR的一个或多个寡核苷酸引物对,这些DMR是或包括表1、5和/或6中鉴定的基因的全部或部分或在表1、5和/或6中鉴定的基因内。在某些情况下,本公开的试剂盒包括一个或多个寡核苷酸引物

对,用于扩增与目前已知基因无关的一个或多个DMR。在某些特定实施方式中,本公开的试剂盒包括用于多个DMR的寡核苷酸引物对,例如2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、或54个DMR,例如,如表1、5和/或6中所提供的。

[0226] 本公开的试剂盒还可包括单独或在单一溶液中的一种或多种MSRE。在各种实施方式中,一种或多种MSRE选自包括AciI、Hin6I、HpyCH4IV和HpaII的MSRE组(例如,使得该试剂盒包括AciI、Hin6I和HpyCH4IV,单独或在单一溶液中)。在一些实施方式中,本公开的试剂盒包括一种或多种用于qPCR的试剂(例如,足以完成qPCR反应混合物的试剂,包括但不限于dNTP和聚合酶)。

[0227] 实施例

[0228] 实施例1. 结肠直肠肿物相关标志物的鉴定

[0229] 本实施例包括对与结肠直肠肿物(例如,结肠直肠癌、晚期腺瘤、息肉病)的诊断和/或分类相关的标志物的鉴定。这种鉴定是使用从本文所述的组织和血液成分中获得的基因组DNA(gDNA)样品的全基因组亚硫酸氢盐测序(WGBS)进行的。同样如本文所解释的,全基因组亚硫酸氢盐测序(WGBS)允许确定关于gDNA内各种基因座的甲基化状态。与来自对照受试者的gDNA中的相同基因座相比,从患有结肠直肠肿物的受试者的组织样品中获得的gDNA中的基因座的甲基化状态差异是可用于确定受试者是否患有结肠直肠肿物的差异甲基化区域(DMR)的指示。

[0230] 此外,鉴定出的DMR还可用于对诊断结肠直肠肿物进行分类。通过WGBS鉴定DMR还允许进一步开发更有针对性的试剂盒和检测方法来确定甲基化状态。例如,甲基化状态可以使用定量聚合酶链反应(qPCR)、甲基化敏感的限制性内切酶定量聚合酶链反应(MSRE-qPCR)、靶向下一代测序(NGS)技术或本领域技术人员将理解的其他等效技术来确定。

[0231] 使用具有不同组织学背景的组织样品以及血沉棕黄层样品以获得用于全基因组亚硫酸氢盐测序的gDNA。获得DNA的来源见下表2。来自晚期腺瘤结肠组织的样品被认为对该病症呈阳性。其他组织和样品类型被认为是评估晚期腺瘤组织甲基化状态的对照。其他组织类型(“其他对照”)包括来自身体其他器官的组织样品,例如肺、乳腺、胰腺和胃。其他癌症包括来自身体其他器官的癌组织的组织样品,包括乳腺癌、肺癌、胰腺癌和胃癌。血沉棕黄层样品包括从2名晚期腺瘤患者和19名非肿瘤性疾病患者的血液中提取的血沉棕黄层。

[0232] 表2. 分析样品来源。

[0233]

患者类型	血沉棕黄层	正常组织	病理学组织
结肠直肠腺瘤	2	57	90
乳腺癌		4	10
肺癌		10	10
胰腺癌		9	10
胃癌		11	12
小肠癌		2	
结肠直肠癌		85	
非肿瘤性	19		1

[0234] 根据制造商的方案,使用DNeasy Blood&Tissue试剂盒(Qiagen)从组织和血沉棕

黄层样本中提取基因组DNA (gDNA)。然后处理提取的gDNA以将其片段化。使用Covaris S220超声仪将gDNA片段化为长度约为400bp的片段。用于gDNA片段化的超声波发生器的示例性设置如下:峰值入射功率140;占空比5%;每突发循环200;处理时间55s。

[0235] 经提取和切碎的gDNA (基因组DNA) 用EZ DNA Methylation-Lightning试剂盒 (ZymoResearch) 进行亚硫酸氢盐转化。使用Accel-NGS甲基序列DNA文库试剂盒 (Swift Biosciences) 从亚硫酸氢盐转化的DNA制备测序文库, 然后使用NovaSeq6000 (Illumina) 设备以37.5x的平均深度进行测序, 使用配对末端测序 (2x150 bp)。使用带有Bowtie 2的亚硫酸氢盐读取映射器将测序读数与亚硫酸氢盐转化的人类基因组 (Ensembl 91组装) 进行比对。以下步骤用于将测序读数与亚硫酸氢盐转化的人类基因组进行比对:

- [0236] • 测序质量评价
- [0237] • 与参考基因组 (hg38) 的比对
- [0238] • 适配子二聚体的重复数据删除和清理
- [0239] • 甲基化调用 (例如, 甲基化核酸的鉴定)

[0240] DMRSeq分析包用于识别与其他对照样品 (例如, 健康结肠组织、其他癌症组织、其他对照组织和血沉棕黄层样品) 相比晚期腺瘤组织DNA的差异甲基化区域。评估该区域的q值, 即用组间标签置换检验校正的p值, 以从患有晚期腺瘤的受试者中选择DNA区域, 这些区域的甲基化程度与从对照受试者获得的DNA中的相同区域有显著差异。发现69740个区域的q值小于0.05。小于0.05的q值指示差异甲基化区域的高统计显著性。

[0241] 为在每个不同甲基化区域内发现的CpG构建百分比甲基化矩阵。在此阶段, 计算每个CpG位置的甲基化百分比。对于每个样品, 通过计算甲基化读数 (“M”) 占总读数 (M+UM, 其中 “UM” 是未甲基化读数的数量) 的百分比来计算DMR中特定CpG的甲基化百分比。使用完整数据的子集将每个层分开, 如下表3所示。

[0242] 表3. 数据子集。

[0243]

结肠直肠腺瘤_PAT	结肠直肠腺瘤_NORM	结肠直肠癌_NORM
57	44	33
肺癌_NORM	胃癌_PAT	胃_NORM
10	12	11
乳癌_PAT	乳癌_NORM	肺癌_PAT
10	4	10
胰腺癌_PAT	胰腺癌_NORM	血沉棕黄层
10	9	21

[0244] 乳癌每个DMR的甲基化百分比 (例如, 每个DMR的甲基化CpG百分比) 的平均值 (μ) 和标准偏差 (σ) 是按向量计算的, 不考虑DMR中的CpG的顺序信息 (例如, 位置、定位)。即, 取甲基化百分比矩阵 (行是CpG, 列是样品) 并将其转换为向量, 然后计算均值和标准差。该指标是针对每个层单独计算的, 这意味着它是针对晚期腺瘤组和对照组中的每个亚组计算的, 即来自如表3所示的非结肠直肠腺瘤_PAT的所有亚组的样品之间。此信息用作下一步的过滤条件。

[0245] 进行进一步过滤以选择与晚期腺瘤样品相比在对照样品中未高度甲基化的区域。使用在对照样品 (例如, 所有非晚期腺瘤样品) 中发现具有 $\mu_{\text{对照}} < 0.2$ (平均甲基化百分比小于

20%)的区域。为了确保所有对照样品中的区域的甲基化百分比较低,标准偏差($\sigma_{\text{对照}}$)的上限(upperbound)被定义为 $E(\sigma_{\text{对照}}, \mu_{\text{对照}} < 0.2)$,即仅使用满足 $\mu_{\text{对照}} < 0.2$ 的区域的对照中标准偏差的平均值。

[0246] 最后,选择的区域其晚期腺瘤与所有其他组织样品之间差异的效应大小(β)足够大, β 值小于-0.4。下表1中鉴定了54个DMR,它们满足 $\mu_{\text{对照}} < 0.2$ 和 $\beta < -0.4$ 的所有预定义标准。

[0247] 表1:与对照组相比,发现结肠直肠癌和/或晚期腺瘤患者血液中甲基化模式发生显著改变的DMR列表。

[0248]

SEQ ID NO	相关基因	chr	起始	终止	区域大小
SEQ ID NO: 1	CSMD2	1	34165443	34165675	233
SEQ ID NO:	MRPS21	1	150294178	150294265	31
SEQ ID NO: 3	FCGR2A	1	161539915	161540181	267
SEQ ID NO: 4	SLC30A10	1	219928356	219928386	88
SEQ ID NO: 5	MSH2,KCNK12	2	47569627	47569858	232
SEQ ID NO: 6	SLC6A1	3	10993689	10993900	268
SEQ ID NO: 7	NA	3	43998153	43998214	104
SEQ ID NO: 8	VEPH1,PTX3	3	157437029	157437296	62
SEQ ID NO: 9	NLGN1	3	173397284	173397387	212
SEQ ID NO: 10	CCDC39,LOC101928882	3	180679674	180679706	35
SEQ ID NO: 11	SOX2,SOX2-OT	3	181712580	181712614	33
SEQ ID NO: 12	NKX3-2	4	13545274	13545403	345
SEQ ID NO: 13	PCDH7	4	30718010	30718076	105
SEQ ID NO: 14	TMEM150C	4	82562331	82562477	1133
SEQ ID NO: 15	INTU	4	127623100	127623235	169
SEQ ID NO:16	STOX2	4	183904880	183905140	100
SEQ ID NO: 17	STOX2	4	183905237	183905728	232
SEQ ID NO: 18	LOC642366,ISL1	5	51387638	51388080	492
SEQ ID NO: 19	HOMER1	5	79512985	79513164	261
SEQ ID NO: 20	VCAN	5	83474370	83474485	130
SEQ ID NO: 21	LINC00461	5	88660609	88660834	136
SEQ ID NO: 22	NA	6	1524106	1524152	67
SEQ ID NO: 23	F13A1	6	6320205	6320663	147
SEQ ID NO: 24	NA	6	27288050	27288544	226
SEQ ID NO: 25	NA	6	27670532	27670614	116
SEQ ID NO: 26	ADGRB3,LOC101928307	6	68635305	68635560	180
SEQ ID NO: 27	NA	6	129767354	129767459	443
SEQ ID NO: 28	C6orf118	6	165309346	165309582	106

[0249]

SEQ ID NO: 29	ZNF736	7	64314242	64314346	495
SEQ ID NO: 30	NA	7	65252405	65252749	237
SEQ ID NO: 31	RGS20	8	53877760	53878892	457
SEQ ID NO: 32	RGS20	8	53881543	53881774	256
SEQ ID NO: 33	CRH	8	66177597	66177765	83
SEQ ID NO: 34	TMEM74	8	108787618	108787717	47
SEQ ID NO: 35	BARHL1	9	132579614	132579683	214
SEQ ID NO: 36	C1QL3	10	16521408	16521447	817
SEQ ID NO: 37	NA	10	22475853	22476043	743
SEQ ID NO: 38	MIR3663HG	10	117165171	117165431	38
SEQ ID NO: 39	TENM4	11	79438095	79438308	473
SEQ ID NO: 40	LOC105369438,AMOTL1	11	94740046	94740862	103
SEQ ID NO: 41	RERG	12	15221303	15221461	70
SEQ ID NO: 42	SOX5	12	24563707	24564068	191
SEQ ID NO: 43	HOXC11,HOTAIR	12	53974529	53974571	261
SEQ ID NO: 44	SYT1	12	78865146	78865306	40
SEQ ID NO: 45	PCDH20	13	61414743	61414845	161
SEQ ID NO: 46	PCDH9	13	67231493	67231965	159
SEQ ID NO: 47	GPC6	13	93228702	93229444	362
SEQ ID NO: 48	SOX21,SOX21-AS1	13	94712995	94713032	43
SEQ ID NO: 49	RTN1	14	59870236	59870262	27
SEQ ID NO: 50	AQP4-AS1,CHST9	18	27184277	27184863	587
SEQ ID NO: 51	NA	19	20052466	20053193	78
SEQ ID NO: 52	LAMP5-AS1,LAMP5	20	9516555	9516632	57
SEQ ID NO: 53	TIAM1	21	31344104	31344160	728
SEQ ID NO: 54	SHOX	X	625391	625465	75

[0250] 实施例2.对血浆样品验证选定的差异甲基化区域

[0251] 本实施例包括实验验证,即可以使用来自使用全基因组亚硫酸氢盐测序(WGBS)的受试者的血浆样品来识别DMR(不同甲基化区域)。出于筛查目的,重要的是能够使用容易获得的样品,如血液、尿液或粪便,以便于轻松确定结肠直肠肿物。在本文讨论的实验之前,不知道是否可以从cfDNA充分分析实施例1中发现的结肠直肠肿物生物标志物以成功捕获ctDNA部分,允许将受试者和/或样品识别和/或分类为结肠直肠肿物阳性。

[0252] 下面描述了从受试者血浆中提取无细胞DNA(cfDNA)的示例性方法。从2018-2019年期间在西班牙结肠直肠癌筛查中心和肿瘤诊所就诊的33名参与者收集了20mL血浆。样品获自患有晚期腺瘤、结肠直肠癌、息肉病的受试者或未被诊断为患有结肠直肠肿物的对照

受试者。下表4进一步描述了样品群组的细分。

[0253] 表4. 血浆样品群组细分。

	晚期腺瘤	结肠直肠癌	息肉病	对照
特征	n=10	n=11	n=2	n=10
年龄(岁, 平均 (IQR))	64.5 (52-71)	65.6 (52-77)	65.5 (60-71)	67.6 (53-76)
女性	5	6	1	5
男性	5	5	1	5
分期				
0期		2		
I期		3		
II期		3		
III期		3		
腺癌特征				
高级别异型增生	5			
>=10mm	5			
管状的	4			
管状绒毛	4			
锯齿状的	2			

[0255] 按照制造商的方案 (QIAamp® Circulating Nucleic Acid Handbook 10/2019, Qiagen), 使用 QIAamp® Circulating Nucleic Acid 试剂盒 (Qiagen) 从 20ml 血浆样品中提取 cfDNA (无细胞 DNA)。随后, 提取的 cfDNA 使用 EZ DNA Methylation-Lightning 试剂盒 (ZymoResearch) 直接进行亚硫酸氢盐转化。

[0256] 使用 Accel-NGS 甲基序列 DNA 文库试剂盒 (Swift Biosciences) 进行文库制备。简而言之, 每个样本中约有 25ng 亚硫酸氢盐转化的 cfDNA 发生了变性。带有第一个截短接头 (“截短接头1”) 的小尾巴被添加到 DNA 片段的 3' 端。在用于合成互补底部 DNA 链的引物延伸反应之后, 将第二个截短接头 (“截短接头2”) 连接到互补的底部 DNA 链。然后, 进行 qPCR 以确定最佳的 PCR 循环数。通过索引 PCR, 纳入了双索引的全长接头并提高了产量。通过 Agencourt AMPure XP 珠 (Beckman Coulter) 对接头连接的 DNA 片段进行清理, 并在低 EDTA (乙二醇四乙酸) TE 缓冲液中洗脱 DNA。

[0257] 使用 2% 琼脂糖凝胶 (ThermoFisher) 进行凝胶电泳并随后切除大约 280 至 320 个碱基对的 DNA 片段, 从而进行大小选择和小片段 (如接头二聚体) 的有效去除。使用 ZymoClean 凝胶 DNA 回收试剂盒 (ZymoResearch) 在不含 RNase-DNase 的水中从琼脂糖凝胶上洗脱 DNA。

[0258] 测序在 Illumina Novaseq 6000PE150 上进行, 每条泳道 25% PhiX。

[0259] 测序读数与亚硫酸氢盐转化的人类基因组 (Ensembl 91 组装) 比对, 使用具有 Bowtie 2 的亚硫酸氢盐读取映射器, 遵循以下步骤:

[0260] 评估测序质量

[0261] 与参考基因组 (hg38) 的比对

[0262] 接头二聚体的重复数据删除和清理

[0263] 甲基化调用

[0264] 表1中的所有54个DMR在来自受试者的血浆样品中进行了评估,如上表4中所示。通过评估表1中的所有54个DMR,实现了将晚期腺瘤与对照分离的AUC为90%。这一结果证实,所有这些区域都是有用的,并有助于血浆中的晚期腺瘤检测。

[0265] 然而,检测晚期腺瘤只需要一小部分DMR。进一步分析表明,组合1使用如下表5所示的四个DMR区域,在晚期腺瘤和对照样品之间提供了令人惊讶的良好区分。对表5的四个DMR区域还提供了替代标识符,通过这些标识符可以根据与DMR相关的基因(如果可用)和起始位置的后三位数字明确地识别DMR。提供了第二标识符,其基于发现DMR的染色体编号连同DMR的开始和终止位置来识别DMR。

[0266] 表5. 四个DMR的组合1。

[0267]

SEQ ID NO	基因	chr	起始	终止
SEQ ID NO:6	SLC6A1	3	10993689	10993900
SEQ ID NO:51	NA	19	20052466	20053193
SEQ ID NO:23	F13A1	6	6320205	6320663
SEQ ID NO:35	BARHL1	9	132579614	132579683

[0268] 组合1能够以90% (9/10) 的灵敏度和90% (9/10) 的特异性区分患有晚期腺瘤的受试者和对照受试者的样本。组合1也能够用于区分对照患者和其他结肠直肠肿物。例如,当使用组合1时,息肉病的灵敏度为50%,而结肠直肠癌检测的灵敏度为27%。基于4区域组合的患者分离的箱形图表示可以在图1中看到。在图1中,使用以下缩写:AA代表晚期腺瘤,CNTRL代表对照,CRC代表结肠直肠癌,PPS代表息肉病。图1中使用的术语“单个”表示实验中使用了来自个体受试者的样品,而不是混集样品。术语“N以上”(y轴)是指高于在晚期腺瘤组中计算的平均(μ)甲基化的读数的数量。

[0269] 表6中显示的第二种四DMR组合(组合2)也可用于检测结肠直肠肿物。

[0270] 表6. 四DMR的组合2。

[0271]

SEQ ID NO	基因	chr	起始	终止
SEQ ID NO:1	CSMD2	1	34165443	34165675
SEQ ID NO:6	SLC6A1	3	10993689	10993900
SEQ ID NO:25	NA	6	27670532	27670614
SEQ ID NO:51	NA	19	20052466	20053193

[0272] 组合2用于检测晚期腺瘤,灵敏度为80%,息肉病的灵敏度为50%,结肠直肠癌的灵敏度为54.5%。当查看结肠直肠病例的子分类时,组合2优先检测0期和I期结肠直肠癌(CRC),如图2所示。高级别和低级别异型增生腺瘤的分离灵敏度相同,进一步证实了这种标记物组合对于早期检测特别有用。

[0273] 基于组合2的患者分离的箱形图表示可以在图3中看到。图中使用的术语“单个”表示实验中使用了来自个体受试者的样品,而不是混集样品。术语“ $\mu\%$ 以上”(y轴)是指高于在晚期腺瘤组中计算的平均(μ)甲基化的读数的百分比。

[0274] 将特异性降低到80%将确保对100%的晚期腺瘤患者、100%的息肉病患者和64%的结肠直肠癌患者进行正确分类。这些结果再次表明使用四DMR的这种组合能够检测和分类结肠直肠肿物。

[0275] 对于患有晚期腺瘤 (AA) 的受试者,组合1和组合2中的每一个中的各DMR的统计结果分别呈现在下表7和8中。

[0276] 表7.基于“ $\mu\%$ 以上”值,与对照相比的晚期腺瘤的组合1中DMR的各DMR性能。

[0277]	SEQ ID NO	基因	c hr	起始	终止	灵敏度	特异性
	SEQ ID NO:6	SLC6A1	3	10993689	10993900	60%	90%
	SEQ ID NO:51	NA	19	20052466	20053193	40%	90%
	SEQ ID NO:23	F13A1	6	6320205	6320663	10%	90%
	SEQ ID NO:35	BARHL1	9	132579614	132579683	30%	90%

[0278] 表8.基于“ $\mu\%$ 以上”值,与对照相比的晚期腺瘤中组合2中区域的个体DMR性能。

[0279]	SEQ ID NO	基因	chr	起始	终止	灵敏度	特异性
	SEQ ID NO:1	CSMD2	1	34165443	34165675	30%	90%
	SEQ ID NO:6	SLC6A1	3	10993689	10993900	60%	90%
	SEQ ID NO:25	NA	6	27670532	27670614	60%	90%
	SEQ ID NO:51	NA	19	20052466	20053193	40%	90%

[0280] 从上表中可以看出,两个DMR (具有SEQ ID NO:6的SLC6A1,和SEQ ID NO:25) 达到了60%的灵敏度和90%的特异性。然而,与单独使用单个DMR相比,DMR的组合能够改善晚期腺瘤的分类和检测。在组合1中,晚期腺瘤检测和分类以90%的特异性实现了90%的灵敏度。在组合2中,晚期腺瘤检测和分类以90%的特异性实现了80%的灵敏度。此外,组合1和2都拥有具有SEQ ID NO:6的SLC6A1和SEQ ID NO:51,这表明它们在分类和检测晚期腺瘤中的重要性。具有多个DMR的晚期腺瘤检测的改进证明了晚期腺瘤和其他结肠直肠肿物的复杂性。

[0281] 表示在组合1和2中每个DMR的高于晚期腺瘤组中计算的平均值(μ) 甲基化的读数的百分比的单独箱形图显示在图4A-F中。对于分类为患有晚期腺瘤 (AA)、结肠直肠癌 (CRC) 或息肉病 (PPS) 的每个受试者或分类为对照受试者 (CNTRL) 的受试者,显示了高于晚期腺瘤中该区域平均甲基化的读数的百分比。从箱形图和表7和表8中可以看出,在分类和检测已鉴定的结肠直肠肿物方面,单个标记没有达到与组合1和2一样高的灵敏度和特异性水平。

序列表

<110> 通用诊断公司

<120> 结肠直肠肿物的检测

<130> AX200123W0

<150> US63/046,510

<151> 2020-06-30

<160> 54

<170> BiSSAP 1.3.6

<210> 1

<211> 233

<212> DNA

<213> 智人

<400> 1

```
gccacctcca cctccagata aaccacaaat tacatctaaa gggttgttta tccgtgtctg      60
ttttgcaatt gaccagtttc ttttaagttc agtcctcctg ttttcattta taacatcacc      120
cattaatata cccccctctc cacacacaca cacacaaaca cacacacaca cacacagtga      180
cagagacaca cgcactcaca cacacaggca catacacgca cacctcttcc acg              233
```

<210> 2

<211> 88

<212> DNA

<213> 智人

<400> 2

```
gattcacgtc tactttctag gatgacttcc atgtgctcca tctcgcgcgt ccctgagcat      60
gttgaatttc caaatcctaa ataagccg                                          88
```

<210> 3

<211> 267

<212> DNA

<213> 智人

<400> 3

```
gtcgtacttc tctgacaaaa atcaacggaa cccccgaacc cacagagagc atccattttg      60
gattccccga agacccttcg ttcttgcatt ctctttctgt ctctttttct caacctcttc      120
tgctggaagc tgagtcctgc tccagatcta ggcaagtgtc agcgcagaaa aaagacctgc      180
ctcgctcagg gctatgagcc gcgccctgaa gcacggaaag ctaattgtgt cactggtttc      240
aaatcaacct caattttttt ggagacg                                          267
```

<210> 4

<211> 31

<212> DNA

<213> 智人

<400> 4	
cggagaccag ctccgccacg aagaaggcga c	31
<210> 5	
<211> 232	
<212> DNA	
<213> 智人	
<400> 5	
cggtagccct tggcacgtat tcttagagga gaaaacggag gtcacaaag gtcagatcac	60
agagccggcc agtgttgag cacaggcggc ccggggtgag cgccagaggt gggctttctt	120
ccctcactga aagccgggag ggagagagag agagagaacg ggggccggcg gagaagaggg	180
cgagacgaaa gtaagcaaag ggacattaga agggaaggca gagccgaggg ac	232
<210> 6	
<211> 212	
<212> DNA	
<213> 智人	
<400> 6	
ggtagcctcg ggcagtgcc attgggttct gagcacacgt cccacgggt ggcaccaca	60
gatgtcctgt tctaggcttg gtcggtctt cagacaagaa actcagaccg ggcagtcccc	120
tattgaggct ctgagcta atcctcccaa aatagacatg aaccacaagg agaatttttt	180
aaaagccaaa tgataacacc acgttctttc cg	212
<210> 7	
<211> 62	
<212> DNA	
<213> 智人	
<400> 7	
cgcccagatt cacgggtcccc acgtgttgct ggagttggcg cagacgcgtg tgcgggcata	60
gc	62
<210> 8	
<211> 268	
<212> DNA	
<213> 智人	
<400> 8	
cgaaatagac aatggactcc atcccactga ggaccgtaag ttcactttaa ctgtttctct	60
gctaaccctg actacatata cacctcttgg tctaaataac acacatatat tttgtggcca	120
gtgagacaag ttaaaaaatt atagcttggt atgcaaaagt gagaagcact tgaagaaaga	180
tggaggtttc aaagttatct ctgtaacgta cataatgggt tgaatcatat caaatcgtca	240
atatattgact gttttgagct acataacc	268
<210> 9	
<211> 104	

<212> DNA	
<213> 智人	
<400> 9	
cgtgtctaaa taagtacatg acactaaatt tccttttaaa tccacctttt acaccatggc	60
cagtcgcttg tttaactccc gttcaaggga cacggttttc aaac	104
<210> 10	
<211> 33	
<212> DNA	
<213> 智人	
<400> 10	
cgtaactttc cacgaaaagg cggtctctcg atc	33
<210> 11	
<211> 35	
<212> DNA	
<213> 智人	
<400> 11	
cgcctgggcg ccgagtggaa acttttgtcg gagac	35
<210> 12	
<211> 130	
<212> DNA	
<213> 智人	
<400> 12	
gacccaaacc acgtttctta cctctgcaga tgtatccact tattccagcg cttaacaga	60
acactgatac taagttgagt caaatctgcg gagaaaatcc aagataatgc agttccataa	120
atcatgctcg	130
<210> 13	
<211> 67	
<212> DNA	
<213> 智人	
<400> 13	
cgttccgata aacccgggtt cttgccaaat gtaagaggga tttggcttta ctgccactta	60
gccggcc	67
<210> 14	
<211> 226	
<212> DNA	
<213> 智人	
<400> 14	
cgtgccgcct gaggaagggtg ccctgtggca gggggtggcc gctgggagat gcctgcctct	60
aaccgacgtc caggcgtgac taaacctcga cgccaccccc attcacaagc tcaactcagg	120

gattcccaag caaccgcaga ccggaggtgg cgcagaggca gtgaccgagg tcgctgatta	180
ggggccgaga ggctggcaaa taataatttt aaaaaaaaaa aaaacc	226
<210> 15	
<211> 136	
<212> DNA	
<213> 智人	
<400> 15	
cgagccctgg tggggctggc ggcccacaga gccccacct gccccgagct cccacagcga	60
ggagtggccg cgccgcccgc cagtgcgccc ggctccgaga ccggcagggg agcacgcggg	120
cgaaggaggg gccgcc	136
<210> 16	
<211> 261	
<212> DNA	
<213> 智人	
<400> 16	
cgatatcagc ctgttagaag catatcccta taaagattta atatccctgt ctctgcatct	60
tggcacctgt gaatatgaaa caacagcata aatatgattt tgaacgttgc attgtcacag	120
atgaaaaaat gcaccaacat gtcaaatgca gcgctgaaaa aggaaatcgg gcttattttt	180
gtcgttgttt actgtaccaa agcatttttg aaaacccaaa tcgaggagat aaccgttttt	240
gaatgaacgg cagtgcaaag c	261
<210> 17	
<211> 492	
<212> DNA	
<213> 智人	
<400> 17	
cggcgcattg gcccctcctc ccctcgcgcg ccgcgcgcat tgttgctcctt tagcgattgg	60
ttgttgacc agaaacagct gtgcagagcc gtgccatcta aagagctgtg gacctgaatg	120
cagcgtagcg ggctggcgtt gacttacacc gggactccag agggagagag gaagcgctgc	180
aggccacttg cattgcgtct tccaggctgc gtggaccgg cgccccggcg tgtgcggttg	240
tgggggagct cgccgtggcc tcccctccct ctggcttttag cttccttttg ggttggcgca	300
ggtgggcccag gcagcgcacc gcagatctcc ccgttcccac gaaggctggc tcgctgtctc	360
tctccgagcg ggagggacca tcctaaaaat atgtaaatat ccaagcgctg gctccaggct	420
ggggcagctg ccaaggtccc cgcgccgccc ccgggtgttt tacatgaaaa tgagaagcct	480
gatgggaacc gc	492
<210> 18	
<211> 443	
<212> DNA	
<213> 智人	
<400> 18	

cgggaggacg gtctcttctg ccgagcagac cacgatgtgg tggagagggc cagtctaggc	60
gctggcgacc cgctcagtcc cctgcatcca gcgcggccac tgcaaatggc aggtactcct	120
ctgcccggct cgggtaggca ggcgccaggt taagccagcc tgtgtgccag cggccacaac	180
aactatggta gctacagggg tggtcgtagt gtttgccctgc agttaaata gaagtgtctgt	240
atgcaatttg cgctgtgctc tgctcctttg cagcaagggt caatgcactc actgtctccc	300
ttgattcccc gagcacacct acaccgtctg tgtgtctcta tatggttaca cataaatgta	360
caccacttgt gtacacgtgt atacacacgc ccaaacatta cttccagttc gctctggcct	420
ccaaaccttg gcttgctgaa aac	443
<210> 19	
<211> 180	
<212> DNA	
<213> 智人	
<400> 19	
cggcatttgc ttattcgagt taaaatgctt tgcggaggag acagcgatca actctattcc	60
acaaaatgag tctacaagta gggaaatgca gaatccggct tcaccgagtc ctataaaaaa	120
tgagttcgct ggtcatttca ctcatgtcct cctcgacact caggagagc caggtactc	180
<210> 20	
<211> 116	
<212> DNA	
<213> 智人	
<400> 20	
cgagactcgt tttaggatac tcttttcctt tcccagcgg caacaaatgt gactgcagag	60
gcgtgccggg atggaagggt gggaaagaac tagacaagcg gaggtggtcc agcccc	116
<210> 21	
<211> 226	
<212> DNA	
<213> 智人	
<400> 21	
cgtgccgcct gaggaagggt ccctgtggca gggggtggcc gctgggagat gcctgcctct	60
aaccgacgtc caggcgtgac taaacctcga cgccaccccc attcacaagc tcaactcagg	120
gattcccaag caaccgcaga ccggaggtgg cgcagaggca gtgaccgagg tcgctgatta	180
ggggccgaga ggctggcaaa taataatttt aaaaaaaaaa aaaacc	226
<210> 22	
<211> 47	
<212> DNA	
<213> 智人	
<400> 22	
cgcgacggcc ccaattccag caacgctaga gggcgcccgt gccaaagc	47
<210> 23	

<211> 459	
<212> DNA	
<213> 智人	
<400> 23	
ggggttccag tctagcaggc tgttctcact tggccccact ccctccacct tttgtgttcc	60
agctccataa tctgctccct gaggaaaggg ggctcgtccc ttggggaagc acctccaact	120
cccccatccc catttggtgg cattctaagc aagcaacggc ttcgggagag ctgcctcgag	180
agcctgagag aagtcccgt tagaagctgg gctgggcagg tgcggagttg ggggcgggaa	240
gccaggattg ggcaagtgga gctgcctgtg accggcgcca cagggccag agcaagccgc	300
ttgctggttc aaccaggaaa ccgaggtgca gaaggtggac gcagcgggcc ctggctcata	360
gggtgcaggg tcggtggctt acctgcaggc gtcacctcc agaggtgccc tcgcgtgggc	420
ttgctctgtg cgcctcgggg acttcctcaa acggactcg	459
<210> 24	
<211> 495	
<212> DNA	
<213> 智人	
<400> 24	
cgtggagggg gtgggggtgga aagaagtagg gaatggagaa gattactaag aaaagtttcc	60
tgtctggaac tgcggcagat ctctttggat agagatgact acttaacctc actctgcttt	120
ccttcgccgc ggttgcggcc gcgacctgt tcttaccacc agcaattcct ccagggactt	180
ggtcagcagc ccaacttgat ctgcgtctct ctgctaaggt gtttcgcgaa cagggtcaac	240
tccaagtctc acctttctag gaatcccggg cgcagcgcgg gggtcgggac tccgacctgt	300
atttccaggc ggaggtttcc ctgggtcagg cggccactct ctgccagaga ttgtcagtta	360
tccaactgtc aatagagccg ccgctccagc gagtttaatt taggcacaga aaagtcctgc	420
ctgggttgag gtgggcttag gatgagtta cttgagtgtg tgatttagaa atagatctat	480
gggacagaca gacac	495
<210> 25	
<211> 83	
<212> DNA	
<213> 智人	
<400> 25	
cggctttgaa gaaggaaaaa gtgagagcac aagcgagcca gccaggagtc gaacctagaa	60
tcttctgatc cgtagtcaga cgc	83
<210> 26	
<211> 256	
<212> DNA	
<213> 智人	
<400> 26	
gagacagagc ttactatct cgctccctct cgcgcctccc tcctcgctgg gcattcaaac	60

agctttccga catcaccagc caaggatttt tttccccgct ctccttagtc gccgtccgtc	120
catcagtacc tgcagggggg aggaggagga gggaggaaag cggaaagagg aaaaagcata	180
agcttgagcc ttccgatccg accacgaata ctcctgtaat aaaccaccg cccaacaaa	240
tctgccatag cagccg	256
<210> 27	
<211> 106	
<212> DNA	
<213> 智人	
<400> 27	
cggaaggacc gggctgaagc gtggccacga ggagggggat acccgtgcga gcgctgagcc	60
ggcagagcgg ctgcagccca cgggctcctc ggacccccgc tgctgc	106
<210> 28	
<211> 237	
<212> DNA	
<213> 智人	
<400> 28	
gcacccccca cctctgtgct ttgcgcgctg gtttcagatt ctctgaggag gcagcacaac	60
cccaatccag tcctcagccc tagtgaccct ggagccggcg tgcagccacc agaaaaagat	120
tttcacatitt caaatcagtc ttcagaagcc catccctcca caattgaaca tcaagcaaat	180
ctcacaagcc agcaagagcc gctccaggcc acttacattc aggtccccgc tcctccg	237
<210> 29	
<211> 105	
<212> DNA	
<213> 智人	
<400> 29	
gcggtctgct ctggagtctg catccccgag tccccgcggg cacagctcaa tcctcatttc	60
cctccatcgc agattagggg ctgagccagc agccaaaac ctacg	105
<210> 30	
<211> 345	
<212> DNA	
<213> 智人	
<400> 30	
gctgtgcggg tccgggactc agggttcccc gctgctatca aggctgcgta gcttccccct	60
cccctcctcc ttaggtggca acttgtggac acaccatta agcggtcagg cgtcaggttt	120
cctcccgaga ggtgggaggc gccctggcct tgattcatcg tgaagctagg caggagattt	180
cccagccacg gaggggtggaa agcttgcctt gacctcagca ggtcatgtca ctccgtgtgc	240
acaaggcctc gagaggcact ttttaaaaat tttttgaggc ggttttcttt tttgtctttg	300
tcttttcttt tttttctgag acagagacag agtctcgctc tgtcg	345
<210> 31	

<211> 1133

<212> DNA

<213> 智人

<400> 31

cgagttcaac	tcacccagga	gcaaacaaac	gacagcaaga	caaatcagcc	accgcactcg	60
eggcttccca	gaaagggcct	catgaatgag	aatgggttgc	taggtttcct	tccctctctc	120
ctgacaatcg	cttcccacaa	gacttccacc	gccgaaagaa	tacaggccgg	gcctggtgac	180
tgcggagtg	gggaaccgcg	ccaggcccac	gaggccgctc	gcgaccgctc	ccgccttcag	240
gaccttgag	agcggccgcc	gcgcccctgg	gacccacgcc	caacccagaa	cacccgcgct	300
cgcttccgc	gcctcaccac	cccagcactt	tattcgctgc	ttcccgcctc	caccttcatt	360
ttttttggca	accaccgctc	ttgtttttca	cagacagatt	aaattgtttt	ttgtttggca	420
tgagggggtg	tttggattgg	ccgagctaca	ttccggtttg	taactacact	aaacgttagt	480
gtaccaagaa	actaagagcg	tctcaaagat	gacagtctga	acgtggcagg	tgaccttaaa	540
aagccactga	gttcaacccc	tttataaatg	aactcaaaaa	ataaagagac	attcccaagg	600
ttagccatta	agttagtggc	aaggttgagt	gtagaaccgt	gggtttctga	cttctaaact	660
agtgttccc	tccactttcc	aaaagaaaaa	aaatatctca	ttgcttaggt	ctcctaagca	720
acaagtggaa	ggtggaaagg	ccggtgttcc	cggaataag	aaggacacat	ggagtttgga	780
aaggaggtct	gcagagggtt	cacttgtagt	agcagcccta	acatggaatc	ccggtgtccc	840
cagctccacg	caacatacct	acagccgcac	tggccagtcc	ctttttccaa	ccagcctgtg	900
gagggtctga	ggcccttcag	ccaccttcag	ccactatctg	gaccttccat	agacaacaat	960
cccaatagaa	tctagtagcc	aaggccaact	ctgtgcctct	ggcaggcaga	gccctgggca	1020
cttagggaga	tgggaaaggg	cagggggaac	aagggaagac	cagagaaggg	aggggaagag	1080
caacagttac	aaaagtgaac	taagattttc	atccaaattc	tgaagaacca	ggc	1133

<210> 32

<211> 232

<212> DNA

<213> 智人

<400> 32

cgctgggaag	ggtaaggaga	cagagcctcc	tggaatcgta	gcgcctcctt	ttaggagaag	60
tgcaaccagg	gcaggggcac	cgaggggcag	ggtgaggaag	tggacgcccc	acgcgtggac	120
cctagaagac	cgactaggta	tgggcgttca	ctcggagcct	ctgttcacgc	tcttgaaaac	180
cgaatgaaca	gagaagtgcc	ctccaccctt	cgctgcgcc	agggcagtta	ac	232

<210> 33

<211> 169

<212> DNA

<213> 智人

<400> 33

gggactgcct	tagacgtgct	gggctttggc	ctcagtgatt	cggatgggcg	ctgtccgcgg	60
tgctgaagcg	cctggggagc	ggggaaaggg	ggcgcttgca	gggctccgca	cctaagctca	120

tgcgctcacc ggccaggagt gacccttctt tctagactct gaaaggacg	169
<210> 34	
<211> 100	
<212> DNA	
<213> 智人	
<400> 34	
ggcgggcggg gtggggcggg gcggggcgct cccggagcat cccgggagtt gtaggccagg	60
ggcgggtccct gcatccctcc tgtctctgtc tcctcccacg	100
<210> 35	
<211> 70	
<212> DNA	
<213> 智人	
<400> 35	
cgaaatcgga ataaaacgat gttattgaga gagggctaaa tcccagagta aatttcaaac	60
gaaataaatc	70
<210> 36	
<211> 40	
<212> DNA	
<213> 智人	
<400> 36	
ggcgccggcg cggccaccgg gagggcagag ccagccgccg	40
<210> 37	
<211> 191	
<212> DNA	
<213> 智人	
<400> 37	
cgcggggtgc agcatgagtc ttcctttgtg gcgtgcggct ccatcggaac gcgcgttgcg	60
acgacaaatt ccttttttcc cccccgcagt taacagttct ggggcagagg ctggtggaga	120
ggtccagagc ccactcagac cgagatgaag atgaggaaaa gcatgagcag gaagaggctg	180
gcggctgcgg c	191
<210> 38	
<211> 261	
<212> DNA	
<213> 智人	
<400> 38	
cggagccgag tgctcggtct ctttccccgg gacgggacag ggagtggagt tcctagcccc	60
ttgggtggga aaagccccgc gcacgtcacc ggggtcccgtc tgctcactgc ttctgcatat	120
tttaagcctg gacacagcct ccttttaggat agaaaggcat ttcccaaaca acaccgattc	180
tgggggtgta gtgggcctgg cgctgggtcc tggagagaag gttcagcccc ctttctcatc	240

cctgtactttt ggggatggct c	261
<210> 39	
<211> 214	
<212> DNA	
<213> 智人	
<400> 39	
cgctggaatg tggcaaaaac aaaacactcc ccacccatgc acatcaccat ctctgactg	60
cgggctgggg ggaaggaggt tcagggccaa tgtgtcccag acttcagcgt tccccacggc	120
tgtgtcaggg ctgggggtgc ttatccccct acagacgaaa atcaagattt aaaagcatac	180
tcttactgtg gtttctctat caaagcccat caac	214
<210> 40	
<211> 817	
<212> DNA	
<213> 智人	
<400> 40	
gcataaatta acaacttctc agcctcgggt tcccaacctg tgcaatgcag ctcatattac	60
tttggccact tgagccaggt gatctcttga agagtaggaa ataatgttgg actggagttg	120
gaattcttgt tccaaaagga ggggattggg gggcatttgg gtctcctcct acctttctgc	180
aagctctgaa aatcgtccat ctctcagga gaagcccaca gaacaaacga tttcccaagc	240
atcagaggag ggcagacata acataacaga gcagggaatc gggttagaac gggttatgct	300
ctgatttttt ggaaaatggg caaacggggc ggggactagg gaggattccg ccagccggga	360
gttgggaggc cgcgcgcgcc tctgcggagc gtgacggcca cggtcgccta gcaacgagcg	420
gggatgcgcg gggcgccctgg gtgggcgagg ggtgctcgcc gccgccgccg cgcgccctcg	480
ggcccccggg accgcgggaa aacttcgcgg ctctcggcg ctggggcgtc ggcgccaggc	540
ccggcagaca gagcaatgcg ccccgcgggc tgagggcaga ggatgcggcc gcggcccagc	600
gcccggccgg gggagccgcg ggggtggcacg cggggaaagt tggcgcgcg ctgaccgcgc	660
ctggaagccg cgcggtgccg gggccgagtt gtccccaaag tttctgcggc gatttgtcac	720
tccctgggga tctggcggtc tgaatcccgc ggggcctccg gctcagggat tctgagcgct	780
gggagagaga agcccgcgct tttcccgggg acctgcg	817
<210> 41	
<211> 159	
<212> DNA	
<213> 智人	
<400> 41	
gtttgcgagt tgctgggctg cggtcgtggg tgggactcgc cgcagaagca agtgccagt	60
gcccggccgg ggtctcctca ctgcgctcg ctccgactag cggcggaggg actgcggcag	120
gacgcgagct gagccccggc aaggccgctg cgctcagcg	159
<210> 42	
<211> 362	

<212> DNA	
<213> 智人	
<400> 42	
gtcggtgatg ccctgagcag aagctgtggg gctgatttaa gtgtgtgctc tgacagctcc	60
cgggctgctc cggcgcttat ctcttctaata cttactttca ttgacttaaa acagctgccg	120
gtttgccagg ggaaaaaaat cctattaaat ctcaaaccag ggtggggcgg gggcgggggt	180
tcctgacagt gatcttccca gagctagtgt ccagaaaaga gcagggggaa gggagagcat	240
tcacgggggt ggctgcgtgt tgggaagggg gtgatggaaa ggggactaag gaaacaattc	300
gagcaacacg tttcttttcc tcgctgtagc cttctctcct tttgctttta tgccgaggtg	360
cg	362
<210> 43	
<211> 43	
<212> DNA	
<213> 智人	
<400> 43	
cggcgacagg ggaatggggc gaggcggcgc aggactccac tgc	43
<210> 44	
<211> 161	
<212> DNA	
<213> 智人	
<400> 44	
ggtgcctatc tgctcgaatc catcccaaag attttcttct cgatacactt gggttttagt	60
ggtgggagtc tgagacctag ggagggatcc ccgggtggcc tattgtgaga aatatgagac	120
tacatgtgca tctcctggaa aagcactttg cacagcgccc g	161
<210> 45	
<211> 103	
<212> DNA	
<213> 智人	
<400> 45	
cgatcccgaa cccctgtagc taaagcggat tgagcgcacc cccgatgccc tcgaccttat	60
ctggatacat ttcttgcttc agaaacttcc tctcatgacc cac	103
<210> 46	
<211> 473	
<212> DNA	
<213> 智人	
<400> 46	
gttttcatac gggaactgga tggaatgact ccccaaaaaa tacagcttta tttctcaaat	60
actgaccccc aaagcactat ctagtaatat atttgattga tctttcaaag tcagtaaacc	120
acaaaggttt gtgtaatggc ttgtacttaa cgcctgatac ctgagtaaag tttgaagcat	180

taacatagga acagttcact tggaacaaaa atttatittc tgaatgacct ataaagggtg	240
tcagagaagg tcttagtaca tgattgaata acttgggtcta acttacagga agaataaagg	300
gcattttattt taaacctgtg tgttccccat tctataatgg gctgcctagg cattaaaggc	360
tcttgacata ctttaagctct tgactgaggg cgtgttgagg attaccctt gtttttgagt	420
acaatagttt tgtttgcttt ttttctttt taagtaacat ttgttttttag acg	473

<210> 47

<211> 743

<212> DNA

<213> 智人

<400> 47

gggctgctca aaccgggcag ctgaagtcct tagtgacctc agatcgtcaa gtcaaaacgc	60
tgaattttcca ccagcctctg tctgcttttt gccaaataac tgggtgatgg atgaaaagca	120
ttttgcagat atcttagaac atcacagttt cgatacgttg aggaattact attttcttat	180
gattttcaag ctgtagaagt gagggttttt acttacactg aaatgaacac atttaaataa	240
atttgagcat tggcaaaggg ggaaaaaaag aggcgcaaat taccacgctc attatataga	300
aggagctttt tcagttcaga gccagacatt ccctttgctg agtctaagtt agaactctgtg	360
gtgaattata agcctacttt tctatccttg ttacttcttc cttcttttcc agaactcctt	420
aatttggttaa tcaatgaata gagagcgact gtccccacag ctcttaagt ttcttaactc	480
tccttctcct ttgtctactg ttatttcatt ctttttaatt aattgataag gatcagcttc	540
gctttttttt ttccctcccc aaatctcagg gaattcaact ttttaaagg tttatagtat	600
ggatcacttc ttctaggaac ttttctcct ttaatctggg ctttttcaaa cggtatcttt	660
taagcacaca agatatttcc caacatgatt tcagataaga tgtctcaaag aagaaaatag	720
ttaggtattt ttaaattgct tcg	743

<210> 48

<211> 38

<212> DNA

<213> 智人

<400> 48

gcctgccgac ttagggctcc cggagctcgc cggccgcg	38
---	----

<210> 49

<211> 27

<212> DNA

<213> 智人

<400> 49

cgaatattcc cagtcgcccc tggcgac	27
-------------------------------	----

<210> 50

<211> 587

<212> DNA

<213> 智人

<400> 50

cgcagagaga	gagaaaggag	gaaggcctgc	agctctagac	ttagcacctg	tgaacttagt	60
tggaataaac	tcctgacaga	catcaaccat	aacaccttct	aacagaaaca	tgctgatagg	120
caggatgttc	aaacggggca	gaagaggag	caggcaaccc	tgtgaaagta	acagcagcag	180
aaacaaacac	accattcctt	ctccccgcaa	cccacttgga	gcccctcctc	agatcgcccc	240
tcccaacccc	tctgctctgg	ccacgatcgc	acctggcctc	ccggtccccc	taacttcctt	300
cccacctctc	ccgcagcctg	cgcctgagcc	tgagccaggt	cgcggagttt	gagactgacg	360
caaaggaggc	acccccgcag	cagagatgct	cgtctttctg	ccacacaccc	tgaggagacc	420
gacagactgg	cagcagaaac	taaagactgt	tctgcccgtc	ctctttccaa	cctctgcatg	480
ccccaagagg	ggtctgaccc	cggaatcctg	gagtccaggg	tgccccgccg	gggcgcagga	540
aggaaactca	ggtagctgac	aagttcaggc	ggccgtcctt	ctccage		587

<210> 51

<211> 728

<212> DNA

<213> 智人

<400> 51

cgccccgttg	ccaggagag	tgaatacagc	tagggacgtg	aagggtaggt	ctgggctggg	60
cattgaggag	ggtattaggc	taggaagtat	acctcacctt	tgtccaaatc	gggtggttcg	120
gcctcctctc	attggcccta	tgcagctggg	ttgcctctct	caccgcacc	caagggtctt	180
tccagagtgt	tgcgtcattt	ccagcccagg	gagctgcctt	ctttcctaaa	ctgcaatgga	240
aactgtcctg	atgtctgaga	caatgtccgt	tgtgccgcag	ccctctgcct	tctctagcca	300
gagcgcgagc	tcagctgctt	ttgagagaaa	tcagccacct	ggcccacctg	tgcacaactt	360
cagagctttg	taggggggtga	caaggactgt	gtcttcaggg	aaatgtcacg	ggcattagcg	420
cctttttcgc	aaatgtgaaa	gttgagaaat	caagaagggt	aattattggg	ttgccaaga	480
tctgctaaga	gcaaaggaga	aaactccgtt	tcccaggcat	gtgtcttgtg	agccattttt	540
aaatcaaccc	tcttaagtgg	acaagctcca	gaacacaaca	tgaagctgat	gatgacttag	600
gcaatttatg	cttgaactca	ttggcctcat	ctcaagtcag	tgtctcagag	acacaggtgg	660
gacctgatcc	ccaaggaaca	gatagcattc	cagattcatg	ggagcaactt	ttgagatgtg	720
gagcaccc						728

<210> 52

<211> 78

<212> DNA

<213> 智人

<400> 52

gccaggaacc	gcaggcgtgg	ggacccaaac	gtcaccctg	gcctgaccc	agataagaag	60
tcccttgaag	gcctgtcg					78

<210> 53

<211> 57

<212> DNA

<213> 智人	
<400> 53	
ggtggaaaga atcgatttca aaattcaagc tcaccgctgc tcaacaaggc gcgcacg	57
<210> 54	
<211> 75	
<212> DNA	
<213> 智人	
<400> 54	
ggtcacatac gctaacaaga cacggtgaaa agtctcttct catcggcttg gtgtgctgct	60
ctcttctctc tctcg	75

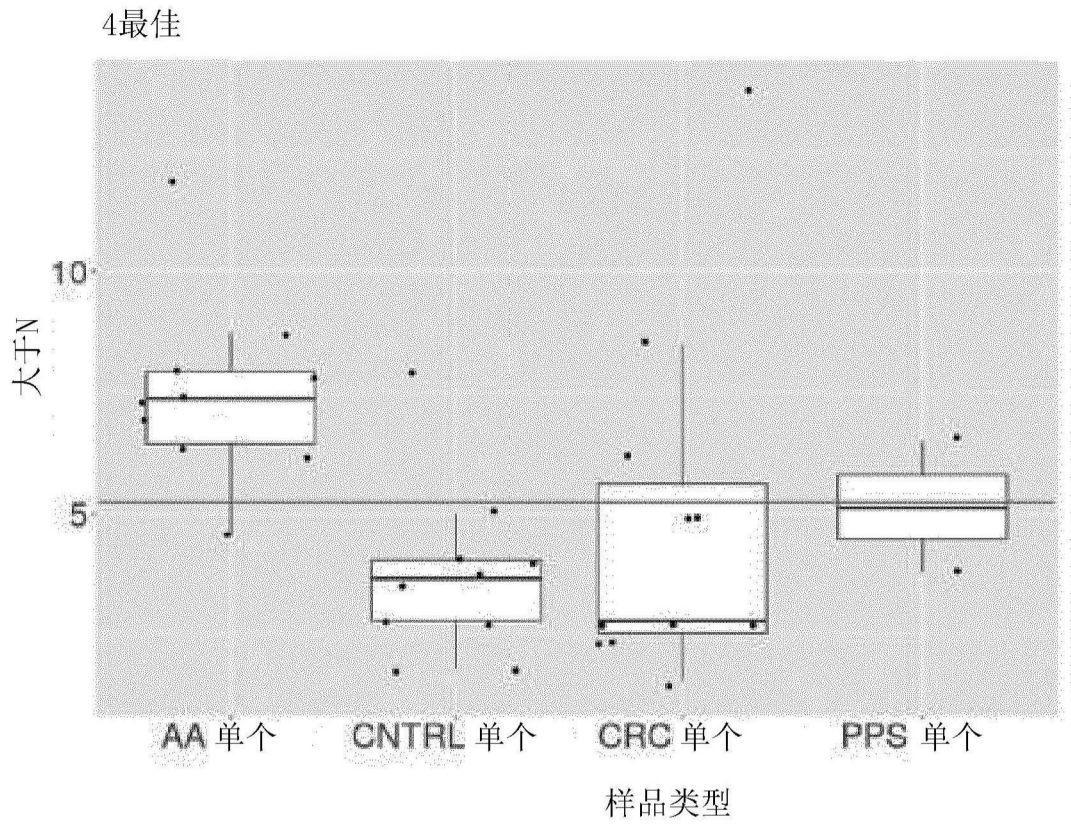


图1

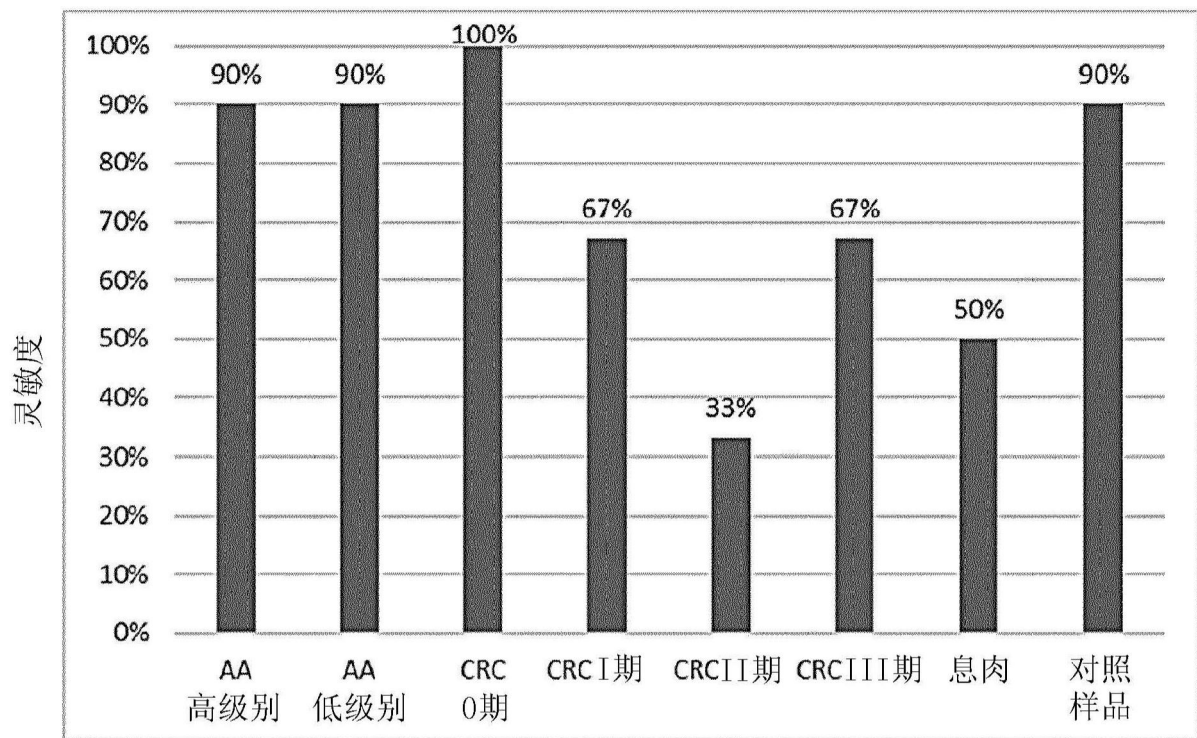


图2

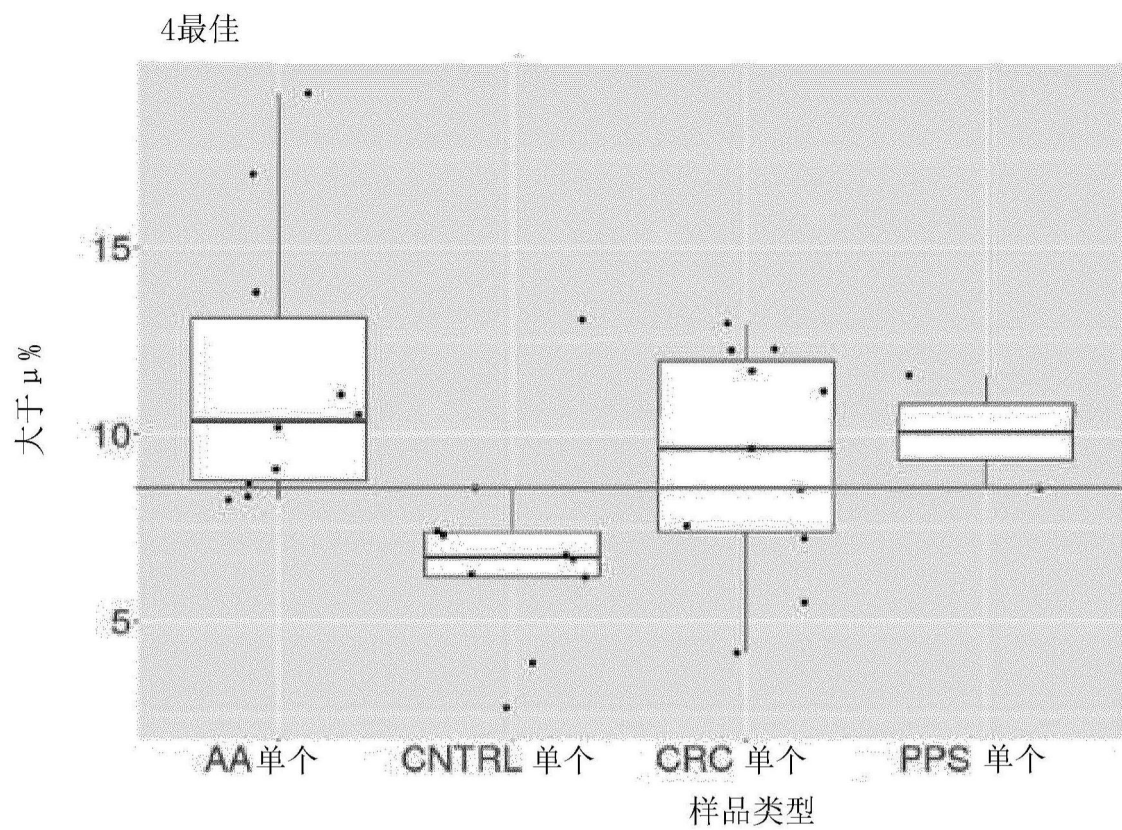


图3

SEQ. ID. NO.1

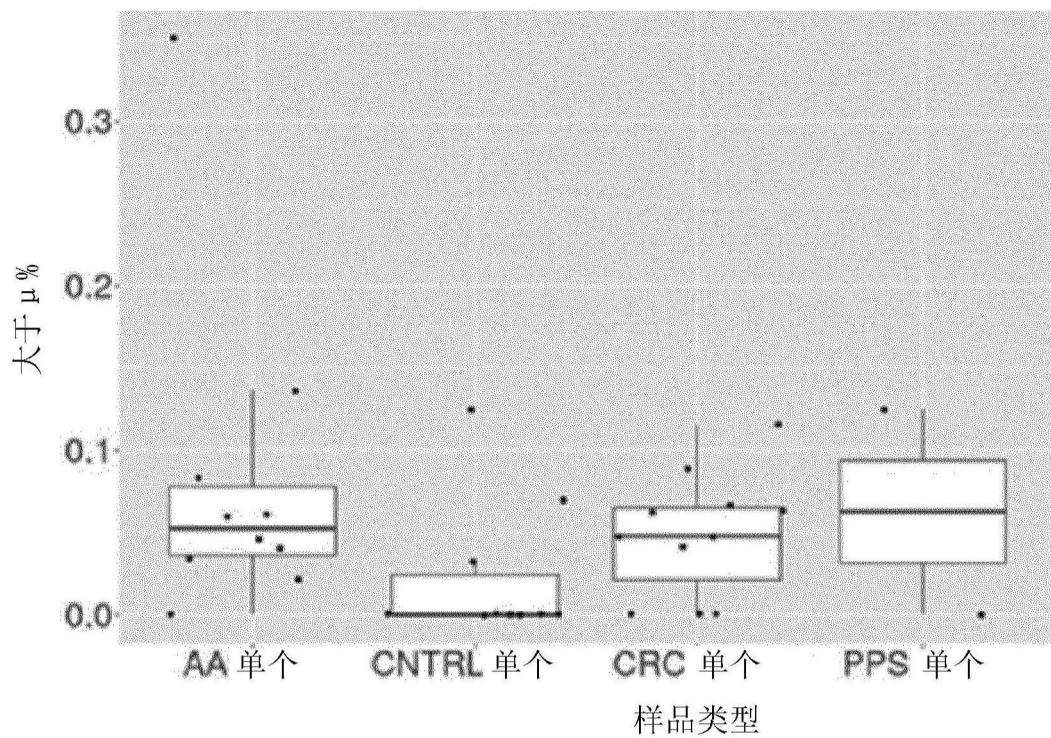


图4A

SEQ. ID. NO.25

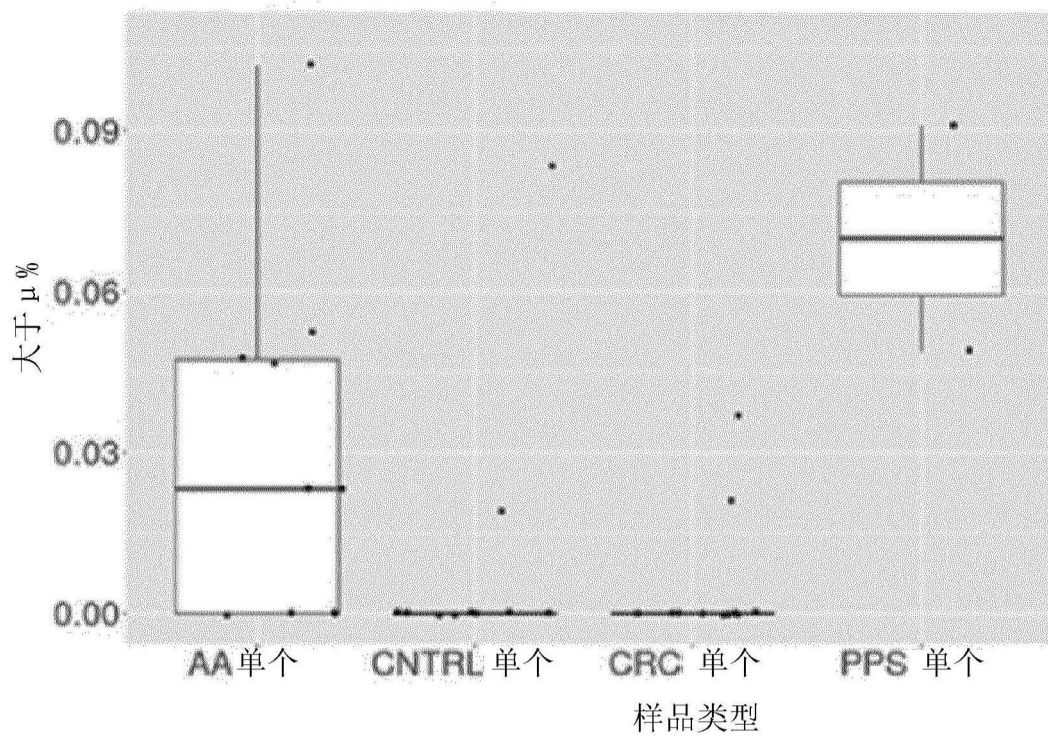


图4B

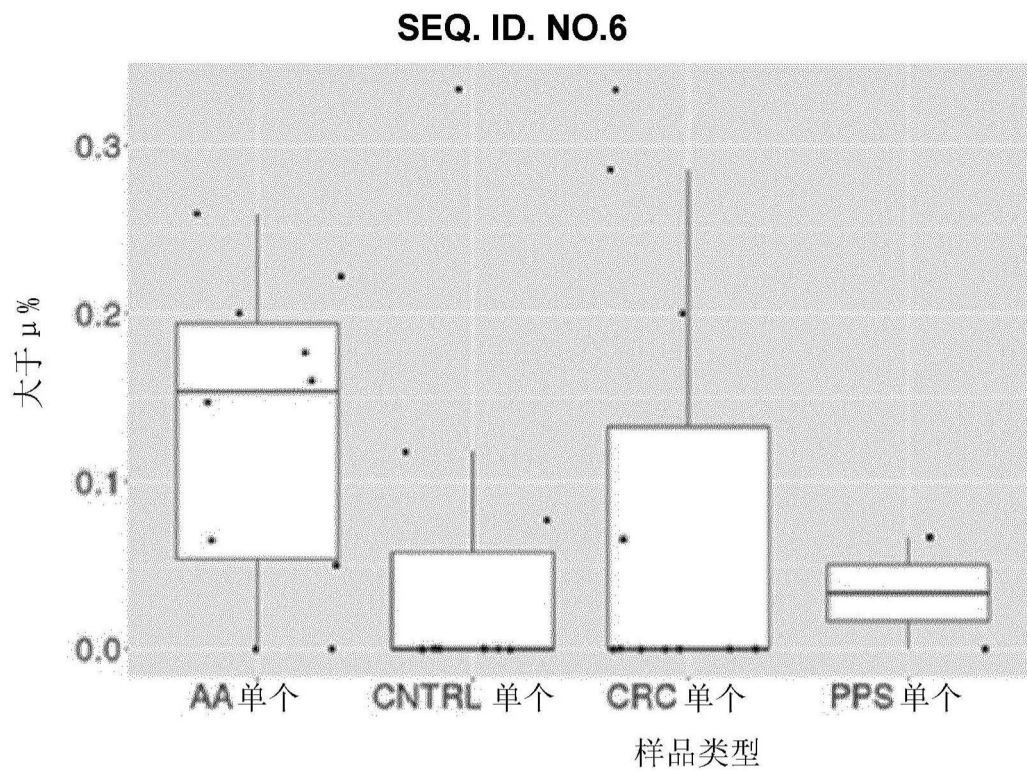


图4C

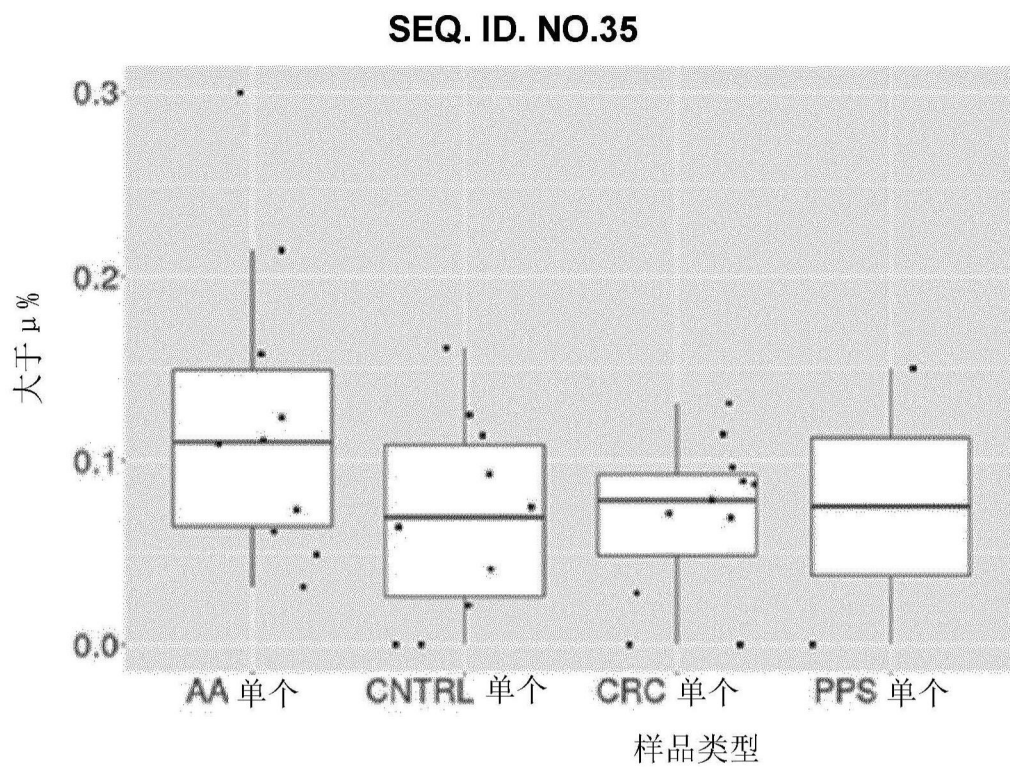


图4D

SEQ. ID. NO.23

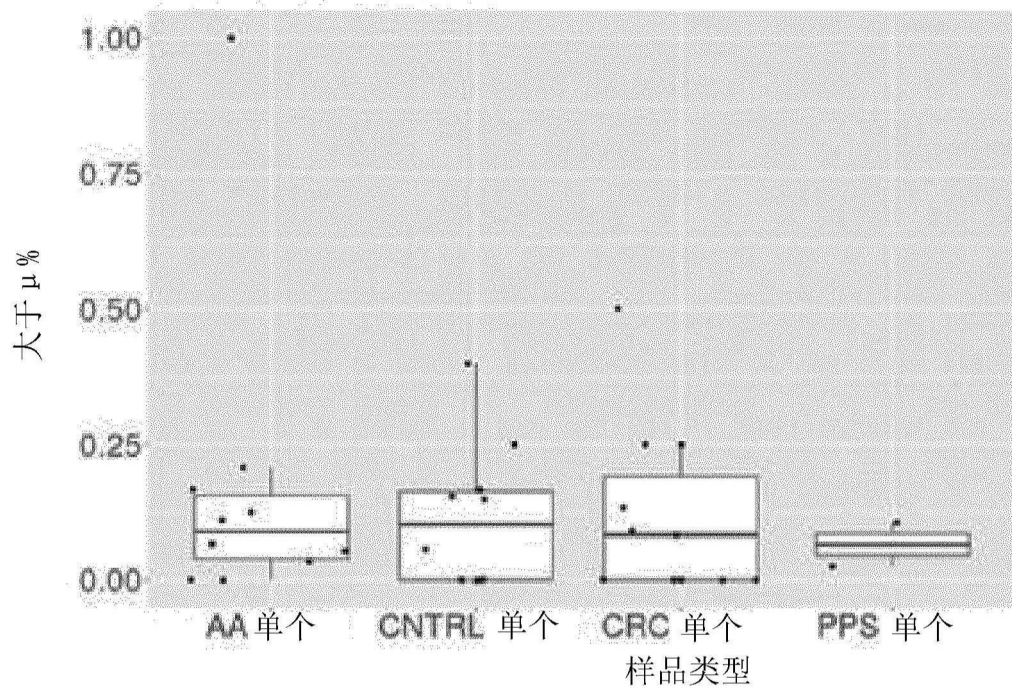


图4E

SEQ. ID. NO.51

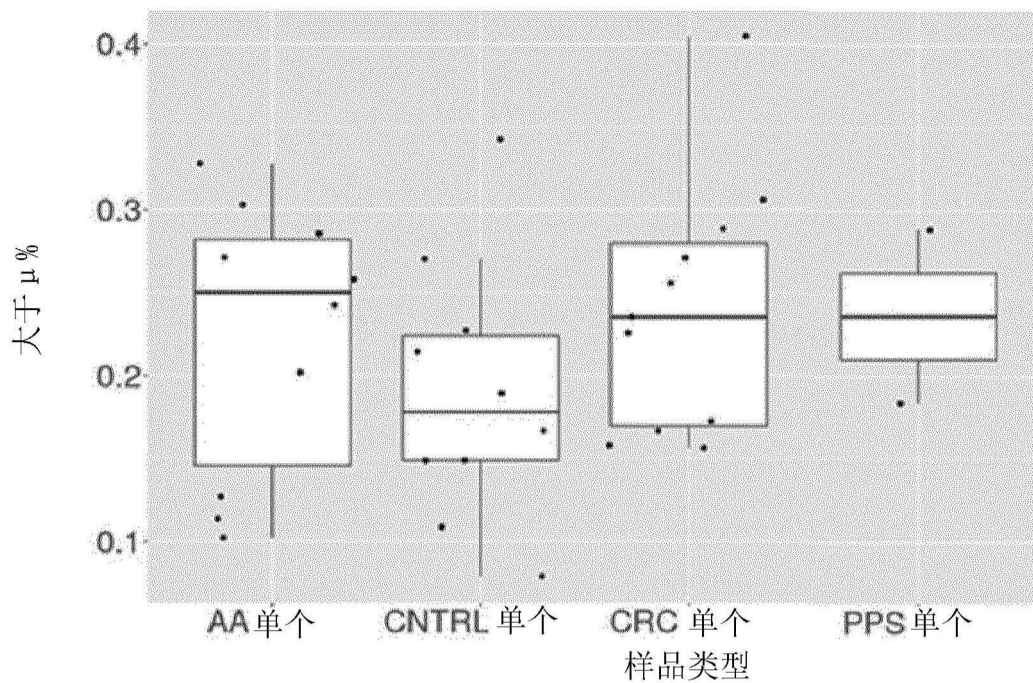


图4F