



# (12)发明专利申请

(10)申请公布号 CN 106574294 A

(43)申请公布日 2017.04.19

(21)申请号 201580022938.6

(22)申请日 2015.03.03

(30)优先权数据

14382074.4 2014.03.03 EP

(85)PCT国际申请进入国家阶段日

2016.10.27

(86)PCT国际申请的申请数据

PCT/EP2015/054451 2015.03.03

(87)PCT国际申请的公布数据

W02015/132273 EN 2015.09.11

(71)申请人 何塞普特鲁塔博士赫罗纳生物医学  
研究所

地址 西班牙赫罗纳

申请人 赫罗纳大学

(72)发明人 马里奥娜·萨拉·佩吉斯

杰西·加西亚-吉尔

特里萨·马斯·德·萨萨尔斯

赛维尔·奥尔迪格尔

(74)专利代理机构 广州华进联合专利商标代理  
有限公司 44224

代理人 刘培培

(51)Int.Cl.

C12Q 1/68(2006.01)

权利要求书2页 说明书36页  
序列表6页 附图25页

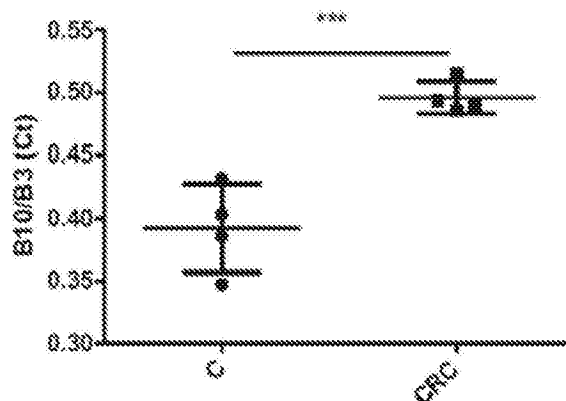
(54)发明名称

用于通过定量PCR从人粪便样本诊断结肠直  
肠癌的方法、引物及试剂盒

(57)摘要

本发明涉及结肠直肠癌(CRC)检测领域。具  
体地,其涉及用于基于粪便中的一种或多种16S  
rDNA细菌序列的量化水平,对人类对象的CRC和/  
或腺瘤性息肉进行早期检测、风险筛查和监控的  
方法。其还涉及使用所述细菌序列作为CRC和/  
或腺瘤性息肉的生物标记物的用途;涉及包括用于  
量化所述细菌序列的试剂和说明书的试剂盒;和  
涉及用于量化所述16S rDNA细菌序列的核酸。

B10/B3



1. 用于筛查人类对象的结肠直肠癌 (CRC) 和/或腺瘤性息肉的方法, 包括:

i. 从所述对象的粪便样本中, 量化选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列, 或其互补rRNA序列; 和

ii. 基于至少一个所述序列的量化水平, 对人类对象中的CRC和/或息肉进行诊断、早期检测、确定复发、确定发展风险或预测, 确定是否应当对所述人类对象进行结肠镜检查, 在所述人类对象被诊断患有CRC和/或息肉时确定预后, 或指导对患有CRC和/或息肉的患者的治疗。

2. 根据权利要求1所述的方法, 其特征在于, 所述方法包括量化所述细菌序列中的2、3或4个, 优选为量化所有4个细菌序列。

3. 根据权利要求1或2所述的方法, 其特征在于, 所述方法还包括: 在步骤i之后, 将对象样本水平与对照样本水平进行比较, 其中对象样本水平从对照样本水平偏离表明存在CRC和/或腺瘤性息肉。

4. 根据权利要求1至3中任一项所述的方法, 其特征在于, 在步骤i中量化至少两个所述细菌序列, 并确定所述序列的所述量化水平的至少一个比率, 还包括, 将所述对象中的至少一个所述比率与对照样本中的比率相比较, 其中从所述对照样本的比率偏离表明存在CRC和/或腺瘤性息肉。

5. 根据权利要求1至4中任一项所述的方法, 其特征在于, 在所述细菌序列量化之前, 通过借助任何方法对所述样本进行物理破坏、并以硅胶膜柱进行DNA纯化的方式, 从所述粪便样本中提取DNA。

6. 根据权利要求1至5中任一项所述的方法, 其特征在于, 所述细菌序列的量化通过定量PCR进行。

7. 根据权利要求1至6中任一项所述的方法, 其特征在于, 包括:

a. 通过定量PCR从所述人粪便样本中确定选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列的水平, 其中SEQ ID NO:1水平的测定使用了与SEQ ID NO:2-13具有至少90%一致性的引物; SEQ ID NO:4水平的测定使用了与SEQ ID NO:5-6具有至少90%一致性的引物; SEQ ID NO:7水平的测定使用了与SEQ ID NO:8-9具有至少90%一致性的引物; SEQ ID NO:10水平的测定使用了与SEQ ID NO:11-12具有至少90%一致性的引物; 和/或

b. 如步骤a所述地通过定量PCR从所述人粪便样本中确定选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10中的至少两种16S rDNA细菌序列的水平, 并确定选自以下的比率中的至少一种: SEQ ID NO:10/SEQ ID NO:4, SEQ ID NO:7/SEQ ID NO:4, SEQ ID NO:4/SEQ ID NO:1, SEQ ID NO:7/SEQ ID NO:1, SEQ ID NO:7/SEQ ID NO:10和SEQ ID NO:10/SEQ ID 1。

8. 根据权利要求7所述的方法, 其特征在于, SEQ ID NO:1浓度的测定使用了序列号为SEQ ID NO:2-13的引物; SEQ ID NO:4浓度的测定使用了序列号为SEQ ID NO:5-6的引物; SEQ ID NO:7浓度的测定使用了序列号为SEQ ID NO:8-9的引物; SEQ ID NO:10浓度的测定使用了序列号为SEQ ID NO:11-12的引物。

9. 根据权利要求1至8中任一项所述的方法, 其特征在于, 所述方法是用于筛查CRC的方法, 并包括对SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10的量化。

10. 根据权利要求1至8中任一项所述的方法,其特征在于,所述方法是用于筛查腺瘤性息肉的方法,包括对SEQ ID NO:1,SEQ ID NO:4/SEQ ID NO:1,SEQ ID NO:7/SEQ ID NO:4,SEQ ID NO:7/SEQ ID NO:1,SEQ ID NO:10/SEQ ID NO:4和SEQ ID NO:10/SEQ ID 1的量化。

11. 根据权利要求1至10中任一项所述的方法,其特征在于,所述方法还包括测定和/或量化已知指示CRC的一种或多种分子生物标记。

12. 根据权利要求1至11中任一项所述的方法,其特征在于,所述方法还包括将所述方法结果存储在电脑可读介质中。

13. 选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列或其互补rRNA序列作为生物标记用于以下各项的用途:对人类对象中的CRC和/或息肉进行诊断、早期检测、确定复发、确定发展风险或预测;或确定是否应当对所述人类对象进行结肠镜检查;或在所述人类对象被诊断患有CRC和/或息肉时确定预后;或指导对患有CRC和/或息肉的患者的治疗;其中所述细菌序列从所述人类对象的粪便样本中量化得到。

14. 试剂盒,包括:

a. 选自以下的试剂:

i. 能够与选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16SrDNA细菌序列或其互补rRNA序列特异杂交的核酸探针;或

ii. 能够特异扩增选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列的核酸引物对;和

b. 用于根据权利要求1至12中任一项所述的方法量化来自人类粪便样本中的一种或多种所述序列的水平的说明书。

15. 选自SEQ ID NO:2,SEQ ID NO:13,SEQ ID NO:5,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:9,SEQ ID NO:11和SEQ ID NO:12或与其具有至少90%一致性的核酸序列。

## 用于通过定量PCR从人粪便样本诊断结肠直肠癌的方法、引物及试剂盒

### 技术领域

[0001] 本发明涉及结肠直肠癌 (CRC) 检测领域。具体地,其涉及用于基于粪便中的一种或多种16S rDNA细菌序列的量化水平,对人类对象的CRC和/或腺瘤性息肉进行早期检测、风险筛查和监控的方法。其还涉及使用所述细菌序列作为CRC和/或腺瘤性息肉的生物标记物的用途;涉及包括用于量化所述细菌序列的试剂和说明书的试剂盒;和涉及用于量化所述16S rDNA细菌序列的核酸。

### 背景技术

[0002] 结肠直肠癌 (CRC) 是欧洲和美国的第二大癌症死亡原因,也是欧洲最频繁诊断的癌症,于2008年有超过40万新发病例和20万死亡人数。这些数据表明需要采取预防措施。

[0003] 减少CRC发病率和死亡率的最有效和最经济的措施,是CRC风险筛查和监测测试。筛查测试分为:主要检测癌症早期的测试;和,既能够早期发现癌症也能够检测腺瘤性息肉,从而通过息肉切除术(即息肉切除)提供更大预防潜力的测试,参见Levin et al等人的美国癌症协会 (ACS) 2008年筛查和监测指南 (CA Cancer J Clinicians, 2008;58 (3) 130-160)。平均风险人群和具有增加或高风险的人群应用不同的筛查指南

[0004] (<http://www.cancer.org/cancer/colonandrectumcancer/moreinformation/colonandrectumcancerearlydetection/colorectal-cancer-early-detection-acs-recommendations>)。

[0005] 当前,主要检测CRC的测试是粪便测试,其中包括:i) 粪便血液检查:粪便潜血试验 (“FOBT”) 和粪便免疫化学试验 (“FIT”);和ii) 粪便DNA (“sDNA”) 测试(参见,<http://www.cancer.org/cancer/colonandrectumcancer/moreinformation/colonandrectumcancerearlydetection/colorectal-cancer-early-detection-screening-tests-used>)。

[0006] i) 粪便血液检查

[0007] FOBT和FIT都是通过检测粪便样本中的血液量来筛查CRC。该测试基于的前提是:肿瘤组织,特别是恶性组织,其出血多于典型粘膜,且出血量随着息肉尺寸和癌症阶段而增加。推荐进行多重测试,因为是间歇性出血。虽然粪便血液检查可以检测出一些下部结肠的早期肿瘤,其无法检测到(i) 上部结肠的CRC,因为任何血液会被代谢;和/或(ii) 较小的腺瘤性息肉,从而造成假阴性。任何由于痔疮,裂隙,炎性疾病(溃疡性结肠炎,克罗恩病),感染性疾病,甚至长跑导致的胃肠出血,都将造成假阳性(Beg et al., J Indian Acad Clin Med, 2002;3 (2) 153-158)。当前的ACS指南(Levin et al., CA Cancer J Clinicians, 2008; 58 (3) 130-160) 建议50岁及以上的平均风险成人使用基于愈创木脂的粪便潜血试验 (gFOBT) 每年筛查。FOBT是欧盟唯一推荐的测试 (Segnan N. et al., 2010, European guidelines for quality assurance in colorectal cancer screening and diagnosis. Luxembourg: Office for Official Publications of the European

Communities)。这两个指南均建议,任何阳性FOBT测试后均应后续进行结肠镜检查。

[0008] ii) 粪便DNA (“sDNA”) 测试

[0009] 所述sDNA测试是测量粪便试样中的各种DNA标记。所述sDNA测试目前可用Cologuard™(Exact Sciences Corp. (Madison, WI)), 测量包括K-ras, APC和P53基因中的分离点突变的多标记物组;BAT-26探针;DNA完整性(DIA)标记;和波形蛋白基因甲基化(Levin et al., CA Cancer J Clinicians, 2008; 58(3) 130-160)。虽然一些指南中推荐sDNA测试, 其他指南则比较保守而不推荐sDNA测试。在一项研究中, sDNA测试的一个版本优于FOBT, 但仍然只检测到15%的晚期腺瘤(Imperiale et al., N Engl J Med, 2004; 351:2704-2714)。

[0010] ACS指南中所述的能够检测腺瘤性息肉和癌症的测试是结构测试, 其中包括: (i) 结肠镜检查, (ii) 乙状结肠镜检查 (“FSIG”), (III) 双对比钡灌肠 (“DCBE”), 以及 (iv) CT结肠成像 (“CTC”, 模拟结肠镜检查)。所有这些方法都需要病人的肠道净化, 并将空气泵入大肠以帮助可视化。目前的指南鼓励在条件允许和患者愿意接受侵入性测试的情况下使用设计为同时检测早期癌症和腺瘤性息肉的测试。

[0011] 结肠镜检查是一种侵入性技术, 可以在一次检查中直接检查所有粘膜提取活检并切除结肠和直肠息肉。这不仅改进了结肠直肠癌的预后, 还能防止疾病。但此技术相关主要缺点是: 成本较高(因为需要训练有素的人员), 麻醉相关的死亡风险和肠穿孔的风险(Garborg K. et al., Ann Oncol., 2013;

[0012] 24(8):1963-72)。

[0013] 结肠镜检查在有经验的专业人员的手中具有100%灵敏和100%特异性, 但由于结肠镜检查的风险和成本, 结肠镜检查前需要使用非侵入性测试对高危人群进行筛查, 如上述用于检测粪便中的血液的粪便测试, 即gFOBT或FIT, 其如上所述会导致显著数量的假阳性结果, 从而导致了等量的不必要结肠镜检查。这表明需要一种有效的非侵入性结肠直肠癌筛查工具。目前还缺乏一种有效的结肠直肠癌筛查工具, 导致了漫长的结肠镜检查候诊名单(>6周), 以及大量经济资源被分配给错误指示的结肠镜检查。这种情况代表了公共和私人医疗的显著支出, 并导致了系统饱和(Allison J. et al., Practical Gastroenterology, 2007, 3:21-32)。

[0014] 因此, 有必要改进用于CRC和/或腺瘤性息肉的早期检测和监测的实际方法的灵敏度、特异性、入侵性和/或成本效益。

[0015] CRC的遗传基础和自然历史已得到很好的限定, 考虑到CRC中不到5%为遗传性, 并且多数为散发, 诊断患者没有结肠肿瘤的个人或家族史。本病的病因仍是未知的, 虽然其疑为具有内源性和外源性因素都积极参与了肿瘤发展的多因素起源。这些因素包括: 年龄, 烟草, 炎性肠病的个人历史, 饮食, 生活方式, 以及微生物群。

[0016] 最近, 已经知晓在CRC患者的结肠粘膜细菌群落不同于健康个体, 且肠道微生物已被提出作为CRC各阶段的发展和进程中的决定因素(Chen W et al., PLoS One, 2012; 7(6): e39743; Zhu Q et al., Tumour Biol., 2013; 34(3):1285-300; Ahn J et al., J Natl Cancer Inst., 2013; 105(24):1907-11; Na wu et al., Microbial ecology, 2013; 66(2): 462-470)。

[0017] W02012/170478涉及使用细菌标志检测腺瘤和结肠直肠癌的方法, 进行了16rRNA基因焦磷酸测序以表征粘膜活检的附着细菌群落, 计算了腺瘤和非腺瘤主体之间的相对丰

度,得出的结论是:腺瘤的发展与存在于肠粘膜的各种细菌类群的相对丰度变化有关。具体地讲,通过对16S rRNA基因(16sDNA)的定量PCR测试,测定了具核梭杆菌(*Fusobacterium nucleatum*)在具有和不具有腺瘤的正常直肠黏膜中的丰度,得出的结论是在腺瘤病例中的丰度高于对照。根据上述数据,提出了使用具核梭杆菌作为结肠直肠癌的生物标志物的建议。

[0018] 粘膜活检样本的发明人之一进行的实验证实了肠道菌群在CRC各个阶段的发展和演变中的介导作用,显示了结肠直肠癌患者的厚壁菌门、放线菌和大肠杆菌的丰度以及健康对象的普拉梭菌(*Faecalibacterium prausnitzii*)的丰度的关联性(Mas de Xaxars T., 2012, Dipòsit legal:GI.1664-2012, <http://hdl.handle.net/10803/94513>)。

[0019] 但是,本领域中公知,粘膜活检和粪便样本之间的微生物群组成存在差异。Lepage等人(*Inflamm Bowel Dis.*, 2005;11(5):473-80)表明,在给定的个体中,粘膜相关和粪便菌群中的优势种不同。同样地,Eckburg等人(*Science*, 2005;308(5728):1635-1638)报告了粘膜和粪便样本中发现的系统发育谱系之间存在统计学显著差异,假定粪便菌群表示脱落粘膜细菌和单独非粘附管腔菌群的组合。

[0020] Nechvatal等人(*J Microbiol Methods.*, 2008;72(2):124-32)表明,粪便DNA可能用作肠内细菌和人类癌症标记物的流行病学研究来源。值得注意的是,其描述了来自人类粪便的DNA提取物的细菌和人序列的实时定量PCR量化。Balamurugan等人(*J Gastroenterol Hepatol.*, 2008;23:1298-303)描述了通过实时定量PCR对结肠直肠癌患者粪便中的特定细菌进行量化,该特定细菌的代谢产物已知具有保护作用(即直肠真杆菌(*Eubacterium rectale*)和普拉梭菌(*Faecalibacterium prausnitzii*),其为产丁酸细菌),或者是有毒或致癌(即脱硫(硫酸盐还原菌)和粪肠球菌(产生细胞外超氧化物))。然而,两者在对细菌序列作为CRC生物标志物的诊断或预测值测定中均沉默。

[0021] 在US2014/0024036中,Wang等人描述了用于CRC检测的方法,包括通过qPCR对来自CRC患者和对照的粪便中的*F. nucleatum* (FNN)的nusG基因进行量化,其检测CRC的灵敏度为57%的灵敏度,特异性为89.5%。

[0022] 虽然,已经描述了用于检测和/或量化粪便中的CRC的DNA生物标记的方法,仍然需要从粪便样本对人类个体进行CRC的早期诊断、风险筛查或患者监测的可靠方法。该方法的有效性能够通过早期检测有效地防止CRC和减少不必要的结构检查测试,如结肠镜检查,由于目前作为CRC检测的标准筛查方法的FOBT和FIT测试的准确性,该检查目前用于疑似CRC的个体中的次数高于必要次数。

## 发明内容

[0023] 在本发明的第一方面中,涉及用于筛查人类对象的结肠直肠癌(CRC)和/或腺瘤性息肉的方法,包括:

[0024] i. 从所述对象的粪便样本中,量化选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列,或其互补rRNA序列;和

[0025] ii. 基于至少一个所述序列的所述量化水平,对人类对象中的CRC和/或息肉进行诊断、早期检测、确定复发、确定发展风险或预测;或确定是否应当对所述人类对象进行结肠镜检查;或在所述人类对象为诊断患有CRC和/或息肉时确定预后;或指导对患有CRC和/

或息肉的患者的治疗。

[0026] 在本发明的第二方面中,涉及将选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列或其互补rRNA序列作为生物标记用于以下各项的用途:对人类对象中的CRC和/或息肉进行诊断、早期检测、确定复发、确定发展风险或预测;或确定是否应当对所述人类对象进行结肠镜检查;或在所述人类对象为诊断患有CRC和/或息肉时确定预后;或指导对患有CRC和/或息肉的患者的治疗;其中所述细菌序列从所述人类对象的粪便样本中量化得到。

[0027] 在本发明的第三方面中,涉及一种试剂盒,包括:

[0028] a. 选自以下的试剂:

[0029] i. 能够与选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列或其互补rRNA序列特异杂交的核酸探针;或

[0030] ii. 能够特异扩增选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列的核酸引物对;和

[0031] b. 用于量化根据权利要求1至12中任一项所述方法中的人类粪便样本中的一种或多种所述序列的水平的说明书。

[0032] 在本发明的第四方面中,涉及选自SEQ ID NO:2,SEQ ID NO:13,SEQ ID NO:5,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:9,SEQ ID NO:11和SEQ ID NO:12或与其具有至少90%一致性的核酸序列。

[0033] 在本发明的第五方面中,涉及将选自SEQ ID NO:2,SEQ ID NO:13,SEQ ID NO:5,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:9,SEQ ID NO:11和SEQ ID NO:12或与其具有至少90%一致性的核酸序列用于本发明的筛查方法的用途。

## 附图说明

[0034] 图1至图7中所用的以下符号含义如下:

[0035] \* 表示统计学显著性为 $p=0.05$

[0036] \*\* 表示统计学显著性 $P=0.01$

[0037] \*\*\* 表示统计显著性为 $p=0.005$

[0038] 图8至图26中的统计显著性以\*表示,且p值显示在相应的表格中。

[0039] 图1:两序列SEQ ID NO:10/SEQ ID NO:4的Ct值比率。CRC表示从结肠直肠癌患者粪便样本中获得的值,而C表示从健康患者中获得的对照值。

[0040] 图2:两序列SEQ ID NO:7/SEQ ID NO:4的Ct值比率。CRC表示从结肠直肠癌患者粪便样本中获得的值,而C表示从健康患者中获得的对照值。

[0041] 图3:两序列SEQ ID NO:4/SEQ ID NO:1的Ct值比率。CRC表示从结肠直肠癌患者粪便样本中获得的值,而C表示从健康患者中获得的对照值。

[0042] 图4:序列SEQ ID NO:4的Ct绝对值。CRC表示从结肠直肠癌患者粪便样本中获得的值,而C表示从健康患者中获得的对照值。

[0043] 图5:序列SEQ ID NO:7的Ct绝对值。CRC表示从结肠直肠癌患者粪便样本中获得的值,而C表示从健康患者中获得的对照值。

[0044] 图6:两序列SEQ ID NO:4/SEQ ID NO:1的Ct值比率。CRC表示从结肠直肠癌患者粪

便样本中获得的值,高风险表示从前一次结肠镜检查中有息肉并有结肠直肠癌风险增加的Lynch综合征患者获得的值,低风险表示从前一次结肠镜检查中无息肉的Lynch综合征患者获得的值。

[0045] 图7:序列SEQ ID NO:4的Ct绝对值。CRC表示从结肠直肠癌患者粪便样本中获得的值,高风险表示从前一次结肠镜检查中有息肉并有结肠直肠癌风险增加的Lynch综合征患者获得的值,低风险表示从前一次结肠镜检查中无息肉的Lynch综合征患者获得的值。

[0046] 图8:用于B3扩增的引物的验证:扩增曲线(8a)和解离曲线(8b)。

[0047] 图9:用于B10扩增的引物的验证:扩增曲线(9a)和解离曲线(9b)。

[0048] 图10:用于B41扩增的引物的验证:扩增曲线(10a)和解离曲线(10b)。

[0049] 图11:用于B46扩增的引物的验证:扩增曲线(11a)和解离曲线(11b)。

[0050] 图12:用于B48扩增的引物的验证:扩增曲线(12a)和解离曲线(12b)。

[0051] 图13:用于B50扩增的引物的验证:扩增曲线(13a)和解离曲线(13b)。

[0052] 图14:B3在CRC和C组中的Ct绝对值。

[0053] 图15:B48在CRC和C组中的Ct绝对值。

[0054] 图16:B10在CRC和C组中的Ct绝对值。

[0055] 图17:B46在CRC和C组中的Ct绝对值。

[0056] 图18:B3在CRC和C、L组中的Ct绝对值。

[0057] 图19:B48在CRC、L和C组中的Ct绝对值。

[0058] 图20:B10在CRC、L和C组中的Ct绝对值。

[0059] 图21:B46在CRC、L和C组中的Ct绝对值。

[0060] 图22:B3,B10,B46和B48在CRC+L vs.C组中的Ct绝对值。

[0061] 图23:B48/B10在CRC、高风险L、低风险L和C组的Ct绝对值比率。

[0062] 图24:B3/B10在CRC、高风险L、低风险L和C组的Ct绝对值比率。

[0063] 图25:B46/B10在CRC、高风险L、低风险L和C组的Ct绝对值比率。

[0064] 图26:B3/B48在CRC、高风险L、低风险L和C组的Ct绝对值比率。

[0065] 图27:B3,B10,B46和B48在健康vs.CRC分析中的ROC曲线。

[0066] 图28:B3,B10,B46和B48在健康vs.Lynch分析中的ROC曲线。

[0067] 图29:B3,B10,B46和B48在健康vs.CRC+Lynch分析中的ROC曲线。

[0068] 图30:B3,B10,B46和B48在CRC vs Lynch分析中的ROC曲线。

## 具体实施方式

[0069] 在本发明中,确定了CRC和/或腺瘤性息肉的诊断、早期发现、风险筛查和监测相关的特定细菌序列。

[0070] i.在本发明的第一方面中,涉及用于筛查人类对象的结肠直肠癌(CRC)和/或腺瘤性息肉的方法,包括:

[0071] ii.i.从所述对象的粪便样本中,量化选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列,或其互补rRNA序列;和

[0072] iii.ii.基于至少一个所述序列的所述量化水平,对人类对象中的CRC和/或息肉进行诊断、早期检测、确定复发、确定发展风险或预测;或确定是否应当对所述人类对象进

行结肠镜检查;或在所述

[0073] 人类对象为诊断患有CRC和/或息肉时确定预后;或指导对患有CRC和/或息肉的患者的治疗。

[0074] 在本申请中,SEQ ID NO:1所示序列也称为“B3”。类似地,SEQ ID NO:4也称为“B10”,SEQ ID NO:7也称为“B46”且SEQ ID NO:10也称为“B48”,所述名称可互换使用。B3,B10,B46和B48分别是本申请的发明人在为实现本发明所进行的所有实验操作过程中识别序列所用的名称。SEQ ID NO:1、4、7和10所示的序列可见于EMBL-EBI公共序列数据库中,其索引号如下:GQ411111.1 (SEQ ID NO:1),GQ411118.1 (SEQ ID NO:4),GQ411150.1 (SEQ ID NO:7),GQ411152.1 (SEQ ID NO:10)。

[0075] B3,B10,B46和B48序列对应于先前从未培养的细菌分离物中分离的变性梯度凝胶电泳(DGGE)凝胶带。使用BLAST分析来鉴定相应的细菌种类:

[0076] B3 16S rDNA的最佳匹配BLAST为:产气柯林斯菌(*Collinsella aerofaciens*);

[0077] B10 16S rDNA的最佳匹配BLAST为:普拉梭菌(*Faecalibacterium prausnitzii*);

[0078] B46 16S rDNA的最佳匹配BLAST为:普拉梭菌(*Faecalibacterium prausnitzii*),*Subdoligranulum variabil*;

[0079] B48 16S rDNA的最佳匹配BLAST为:瘤胃球菌(*Ruminococcus*),罗氏菌(*Roseburia*),粪球菌(*Coprococcus*)。

[0080] 如本申请所用的术语“结肠直肠癌”、“CCR”或“CRC”是用于始于结肠或直肠的癌症。根据其起源地,这些癌症也可以分别称为结肠癌或直肠癌。结肠癌和直肠癌有许多共同的特点。

[0081] 根据ACS,几种类型的癌症可能在结肠或直肠开始。结肠直肠癌的95%以上的是一种被称为腺癌的癌症。这些癌症在构成产生粘液以润滑结肠和直肠内部的腺体的细胞内开始。其他不太常见类型的肿瘤也可能开始在结肠和直肠中。其包括:类瘤癌,胃肠道间质瘤(间质瘤),淋巴瘤和肉瘤。

[0082] 癌筛查的目标是通过降低晚期疾病的发病率来降低死亡率。为此,现代的CRC筛查可通过检测早期腺癌和检测并切除腺瘤性息肉来实现这个目标,腺瘤性息肉通常被认为是非专性前期病变。

[0083] “腺瘤性息肉”是器官如结肠的内表面上的异常腺细胞非癌性生长。例如,三种类型可以在结肠中生长的腺瘤性息肉为管状、绒毛状和绒毛管状腺瘤。腺瘤性息肉常见于50岁以上成人中,但大多数息肉不会发展成腺癌;其组织学和大小确定其临床重要性。

[0084] 根据ACS,优选具有寻找息肉和癌症的最佳概率的测试,因为发现并除去其息肉能使得部分人群免得结肠直肠癌。优选地,本发明的方法是用于CRC和腺瘤性息肉筛查的方法。

[0085] 在一个具体的实施例中,本发明的方法是用于诊断或检测CRC和/或腺瘤性息肉的筛查方法,优选为用于CRC和/或腺瘤性息肉的早期检测。术语“早期检测”指的是临床症状存在之前的检测。术语“检测”和“诊断”在本申请中可互换使用。据发现,本发明的方法是用于通过早期阶段的疾病诊断防止结肠直肠癌的有力工具。

[0086] 在一个具体的实施例中,本发明的筛查方法在50岁及以上的对象中执行。鼓励50岁及以上的平均风险男女遵循结肠直肠癌的筛查试验(见Levin et al.,*CA Cancer J*

Clinicians,2008;58(3)130-160的表2,其通过引用并入本申请)。

[0087] 在另一个实施例中,本发明的筛查方法在增加风险和/或高风险的对象中进行。将特定指南应用于增加或高风险的患者。

[0088] 根据ACS(见Levin et al.,CA Cancer J Clinicians,2008;58(3)130-160的表3,其通过引用并入本申请),增加风险和高风险的对象包括以下内容:

[0089] 增加风险

[0090] 具有先前结肠镜息肉史的对象

[0091] 患有结肠直肠癌的对象;和

[0092] 具有结肠直肠癌或腺瘤性息肉家族史的对象。

[0093] 高风险

[0094] 通过基因检测诊断为家族性腺瘤性息肉病(FAP)或未经基因检测而疑似FAP的对象;

[0095] 患有遗传性非息肉性结肠直肠癌(HNPCC或Lynch综合征)或未经基因检测而基于家族史具有增加的HNPCC风险的对象;和

[0096] 患有炎症性肠病的对象,例如慢性溃疡性结肠炎或克罗恩病。

[0097] 本发明的目的之一是提供一种预诊断工具。具体地,本发明提供了一种通过鉴定具有发展CRC和/或腺瘤性息肉的倾向或风险的个体来在产生临床征兆之前诊断结肠直肠癌的方法。

[0098] 在一个具体的实施例中,本发明的筛查方法是一种用于确定人类对象的发展CRC和/或腺瘤性息肉的风险的方法,包括:

[0099] i.从所述对象的粪便样本中,量化选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列,或其互补rRNA序列;和

[0100] ii.基于至少一个所述序列的所述量化水平,确定发展CRC和/或腺瘤性息肉的风险。

[0101] 在一个优选的实施例中,本发明的筛查方法是一种用于确定增加风险和/或高风险的对象(优选为Lynch综合症携带者)发展CRC和/或腺瘤性息肉的风险的方法。

[0102] 术语“Lynch综合症”是指遗传性非息肉性结肠直肠癌或HNPCC(一种大大增加个体患结肠直肠癌风险的遗传疾病),以及子宫内膜癌(子宫内膜癌)、卵巢癌和其他一些癌症。患有此症的人往往在年轻时发展成癌症,往往不会首先产生许多息肉。

[0103] 本发明的筛查方法中在实施例4(表3和4)中展现出了以统计学显著方式在具有息肉(高危人群)和没有息肉(低风险人群)的Lynch综合症患者之间加以区分的能力。根据实施例4中的结果,SEQ ID NO:4(B10)的量化和SEQ ID NO:7(B46)/SEQ ID NO:4(B10)比率的确是良好指标。相应地,SEQ ID NO:4(B10)和SEQ ID NO:7(B46)/SEQ ID NO:4(B10)比率是用于本发明的方法中筛查息肉存在的优选生物标记。

[0104] 在先前结肠镜检查中具有息肉史或具有癌症切除后的结肠直肠癌病史的对象中,为监控或监督的目的进行筛查试验,即在所述人类对象中检测癌症和/或腺瘤性息肉复发。

[0105] 在进一步的具体实施例中,本发明的筛查方法是用于检测先前具有息肉和/或结肠直肠癌历史的人类对象的结肠直肠癌(CRC)和/或腺瘤性息肉的监控方法,包括:

[0106] i.从所述对象的粪便样本中,量化选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7

和SEQ ID NO:10中的至少一种16S rDNA细菌序列,或其互补rRNA序列;和

[0107] ii. 基于至少一个所述序列的所述量化水平,确定CRC和/或腺瘤性息肉的复发风险。

[0108] 本发明的筛查方法在实施例8(表8)中表现出了以统计学显著方式在上一次结肠镜检查中发现CRC和息肉(高危人群)和上一次结肠镜检查中未发现CRC和息肉(低危人群)之间加以区分的能力。SEQ ID NO:1/SEQ ID NO:4(B3/B10),SEQ ID NO:4/SEQ ID NO:7(B10/B46),SEQ ID NO:1/SEQ ID NO:10(B3/B48)和SEQ ID NO:10/SEQ ID NO:4(B48/B10)比率的量化得到的结果尤其出色。

[0109] 相应地,SEQ ID NO:1/SEQ ID NO:4(B3/B10),SEQ ID NO:4/SEQ ID NO:7(B10/B46),SEQ ID NO:1/SEQ ID NO:10(B3/B48)和SEQ ID NO:10/SEQ ID NO:4(B48/B10)比率是用于本发明的筛查/监测息肉存在的方法的优选生物标记物,优选为在增加风险或高风险的患者,如具有CRC病史的Lynch患者中。

[0110] 在进一步的具体实施例中,本发明的筛查方法是一种用于确定是否应当对人类对象进行结肠镜检查的方法,包括:

[0111] i. 从所述对象的粪便样本中,量化选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列,或其互补rRNA序列;和

[0112] ii. 基于至少一个所述序列的所述量化水平,是否应当对人类对象进行结肠镜检查。

[0113] 本领域的普通技术人员知晓用于本发明的细菌标志物的量化和分析的若干方法和装置。术语“量化”指的是,量化样本中的特定核酸序列的量的能力。用于测定目标核酸序列的量的分子生物学方法在本领域中是公知的。这些方法包括但不限于终点PCR、竞争PCR、逆转录酶PCR(RT-PCR)、定量PCR(qPCR),逆转录酶qPCR(RT-qPCR)、PCR-ELISA、DNA微阵列、原位杂交测定例如斑点印迹或荧光原位杂交法(FISH)、分支DNA(Nolte,Adv.Clin.Chem.,1998,33:201-235),并复用所述方法(参见例如Andoh et al.,Current Pharmaceutical Design,2009;15,2066-2073)。优选的引物和/或探针以可预测的方式发生反应,通常是通过提供直接线性响应来增加细菌核酸序列量。通过制备相应标样并与之比较,可以容易地量化样本中的给定核酸序列的量。

[0114] 优选地,所述核酸序列是选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10的16S rDNA序列。

[0115] 一种特别优选的定量方法是FISH,其结合了探针杂交和荧光光镜、激光共聚焦显微镜或流式细胞仪来直接量化个别细菌序列。对于FISH方法的综述,参见例如,Harmsen et al.,Appl Environ Microbiol,2002;68 2982-2990,Kalliomaki et al.,J AllergClinImmunol,2001;107 129-134,Tkachuk et al.,Genet.Anal.Tech.Appl.,1991;8:67-74;Trask et al.,Trends Genet.,1991;7(5):149-154;和Weier et al.,Expert Rev.Mol.Diagn.,2002,2(2):109-119;和美国专利6,174,681。

[0116] 另一种特别优选的定量方法是定量PCR(qPCR),也称为实时PCR。PCR过程是本领域公知的,因此不详细描述。Q-PCR技术的概述和实验方案可以从供应商如Sigma-Aldrich公司的SYBR Green qPCR应用获得,例如参见

[0117] <http://www.sigmaaldrich.com/technical-documents/protocols/biology/>

sybr-green-qpcr.HTML或<http://www.sigmaldrich.com/life-science/molecular-biology/pcr/quantitative-pcr/qpcr-technical-guide.html>。对于用于量化细菌基因标记物的丰度和表达的qPCR方法的综述,可以参见,例如,Smith CJ and Osborn AM., FEMS Microbiol Ecol., 2009; 67(1): 6-20。

[0118] 术语“量化水平”可以是浓度(每单位体积的DNA量),每细胞数量的DNA量,循环阈值(Ct值)或其任何数学变换。在一个优选的实施例中,所述细菌序列的量化通过qPCR进行。在一个更优选的实施例中,所述细菌序列的定量是通过qPCR进行的,且量化等级是Ct值。Ct值(循环阈值)定义为荧光信号越过阈值所需的qPCR循环数。Ct水平与样本中的目标核酸的量成反比(即,Ct水平越低,样本中目标核酸的量越大)。

[0119] 粪便样本中的目标核酸序列丰度的定量(例如SEQ ID NO:1)可以是绝对或相对的。相对定量是基于一个或多个内部参照基因,即来自参照菌株的16S rRNA基因,如使用通用引物测定总细菌数(真细菌)并将该目标核酸序列的丰度表达为真细菌百分比(如SEQ ID NO:1/真细菌比率)。绝对定量通过与DNA标准比较得到目标分子的确切数目。

[0120] 在一个具体的实施例中,本发明的方法进一步在步骤i)后包括将对象样本水平与对照样本中的水平相比较,其中,当选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10的一个或多个序列的所述水平从所述对照样本的各序列水平偏离(优选地,统计学显著偏差)时,表明存在CRC和/或腺瘤性息肉。例如,当量化水平偏离对照样本中测得的各水平的中位值(临界)值加(+)/减(-)标准偏差时,则表明存在CRC和/或腺瘤性息肉。

[0121] 在一个优选的实施例中,当所述对象样本中的一个或一个以上所述序列的浓度低于对照样本中测得的各序列水平的中位值(临界)值减去(-)标准偏差时,表明存在CRC和/或腺瘤性息肉。在另一个优选实施例中,当所述对象样本中的一个或一个以上所述序列的Ct水平高于对照样本中测得的各序列水平的中位值(临界)值加上(+)标准偏差时,表明存在CRC和/或腺瘤性息肉。

[0122] 术语“对照样本”可以指参照人群的对照样本的集合,例如,对照样本可以是来自健康对象的样本,来自没有腺瘤性息肉的对象的样本,来自没有腺瘤性息肉和/或结肠直肠癌的既往史及其组合的对象的样本。

[0123] 上述参照种群之一的人类对象的分类,是通过结肠镜检查查探进行的。例如,“健康对象”是在先前的结肠镜检查中不存在腺瘤性息肉或结肠直肠癌的对象。

[0124] 优选地,本发明的方法包括量化选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10的2、3、4个序列,优选为全部4个序列。

[0125] 在一个具体的实施例中,本发明的方法包括量化选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10的至少两个序列,并确定至少两个所述序列之间的关系。优选地,确定2、3、4个所述序列之间的关系(例如比率,多元分析等)。

[0126] 在一个优选的实施例中,通过将第一序列的量化水平除以第二序列的量化水平,获得所述至少两个序列之间的比率。例如,序列SEQ ID NO:10/SEQ ID NO:4的浓度比率,由序列SEQ ID NO:10的浓度除以序列SEQ ID NO:4的浓度获得。SEQ ID NO:10/SEQ ID NO:4比率也可以通过由称为SEQ ID NO:10的序列的Ct值除以称为SEQ ID NO:4的序列的Ct值来获得。

[0127] 本发明的方法的另一个实施例中,在步骤i)中量化至少两个所述细菌序列,且确

定对象样本中的所述序列的水平的比率中的至少一个,还包括:还包括,将所述对象中的至少一个所述比率与对照样本中的所述比率相比较,其中从所述对照样本的偏离(优选为统计学显著偏离)表明存在CRC和/或腺瘤性息肉。例如,当一个或多个的所述比率偏离对照样本中测得的各水平的中位值(临界)值加(+)/减(-)标准偏差时,则表明存在CRC和/或腺瘤性息肉。

[0128] 当涉及测试样本和对照样本之间的差异时,术语“统计显著性”指的是:当使用合适的统计分析时各组相同的概率小于5%的情况,例如 $p < 0.05$ 。换句话说,在完全随机的基础上获得相同结果的概率为每100次尝试中小于5次。本领域技术人员知晓如何选择合适的统计分析。通常情况下,适当的统计分析是基于所研究的变量是否具有正态分布,例如通过使用Kolmogorov-Smirnov检验来确定。优选地,在存在正态分布的情况下,可以使用参数模型例如为t检验或ANOVA检验;而在不存在正态分布时,可以使用非参数模型,如Mann-Whitney U检验或Kruskal-Wallis检验。

[0129] 在一个具体的实施例中,使用灵敏度、特异性、准确性、ROC分析结果或其组合来描述本发明的筛查方法。特别是,用来量化该方法的良好性和可靠性。

[0130] 还有一些与灵敏度、特异性和准确性的描述一起使用的若干术语,即真阳性(TP)、真阴性(TN)、假阴性(FN)、假阳性(FP)。如果证明患者患有疾病,且给定的筛查试验业表明该疾病的存在,则测试的结果被认为是真阳性。同样地,如果证明患者未患有疾病,且筛查试验业表明该疾病不存在,则测试结果为真阴性(TN)。真阳性和真阴性均表明筛查试验和验证情况(也称为真实标准)之间的结果一致。然而,医疗测试都不是完美的。如果筛查试验表明实际上未患有疾病的患者患有该疾病,则测试结果为假阳性(FP)。类似地,如果筛查试验的结果表明实际上患有该疾病的患者未患有该疾病,则测试结果为假阴性(FN)。假阳性和假阴性均表明测试结果与实际情况相反。

[0131] 灵敏度,特异性和准确性以术语TP, TN, FN和FP来描述。

[0132]  $\text{灵敏度} = \text{TP} / (\text{TP} + \text{FN}) = (\text{真阳性评估的数量}) / (\text{所有阳性评估的数量})$ ;

[0133]  $\text{特异性} = \text{TN} / (\text{TN} + \text{FP}) = (\text{真阴性评估的数量}) / (\text{所有阴性评估的数量})$ ;

[0134]  $\text{准确度} = (\text{TN} + \text{TP}) / (\text{TN} + \text{TP} + \text{FN} + \text{FP}) = (\text{正确的评估数}) / (\text{所有评估的数量})$ 。

[0135] 在一个优选的实施例中,本发明的筛查方法以统计学显著的方式,以至少65%、至少70%、至少75%、至少80%的中、至少85%、至少90%、至少95%或优选为100%的灵敏度和/或特异性,对人类对象中的CRC和/或息肉进行诊断、早期检测、确定复发、确定发展风险或预测;或确定是否应当对所述人类对象进行结肠镜检查;或在所述人类对象为诊断患有CRC和/或息肉时确定预后;或指导对患有CRC和/或息肉的患者的治疗。

[0136] 如以上等式所示,本申请中使用的术语“灵敏度”指的是筛查试验正确确定真阳性的比例。其显示了测试用于检测一种疾病的良好程度。灵敏度(“sens”)可以是在 $0 (0\%) < \text{sens} < 1 (100\%)$ 的范围内,理想地,假阴性的数量等于零或近似等于零而灵敏度等于1 (100%)或近似等于1 (100%)。优选地,其灵敏度为70%至90%,75%至95%,80%至95%,85%至100%,或90%至100%。更优选地,本发明的方法的灵敏度值为至少85%,如约86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99%或100%。

[0137] 本申请所用的术语“特异性”是指通过筛查试验正确识别真阴性的比例。其显示了测试用于鉴定正常(阴性)状态的良好程度。特异性(“spec”)可以在 $0 (0\%) < \text{spec} < 1 (100\%)$

的范围内,理想地,假阳性的数量等于零或近似等于零而特异性等于1 (100%) 或近似等于1 (100%)。优选地其特异性为70%至90%,75%至95%,80%至95%,85%至100%,或90%至100%。更优选地,本发明的方法的特异性值为至少85%,例如约86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99%或100%。

[0138] 本申请中的术语“准确性”是群体中真实结果(真阳性或真阴性)的比例。其衡量筛查试验对一种情况的准确性程度,即,确定或排除给定情况的正确程度。准确性(“acc”)可以在 $0 (0%) < acc < 1 (100%)$ 的范围内,理想地,假阳性的数量等于零或近似等于零而准确性等于1 (100%) 或近似等于1 (100%)。优选地,本发明的方法的准确度为至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%或优选为100%。在一个优选的实施例中,其准确度为70%至90%,75%至95%,80%至95%,85%至100%,或90%至100%。优选地,本发明的方法的准确度值为至少85%,例如约86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99%或100%..

[0139] “ROC曲线”是灵敏度和特异性之间关系的图形表示,有助于通过确定筛查试验的最佳阈值(最佳临界点)来决定最佳模型。ROC曲线下的面积(AUC)提供了一种测量筛查试验准确性的方法。优选地,本发明的方法的AUC范围值为从0.6至1,更优选为0.7至1,更优选的值为0.8至1的范围内,更优选为0.9至1。在一个具体的实施例中,AUC为从0.6到0.8,例如从0.6至0.75或0.65至0.75。

[0140] 此外,灵敏度和特异性为比例,因此,相应的置信区间可以通过使用本领域中公知的比例标准方法来计算。两种类型的95%置信区间通常围绕比例限定。确切的置信区间是通过使用二项分布达成确切估计来限定的。渐近置信区间是通过假设样本分布近似正态来计算的。本领域的技术人员知晓如何定义相应置信区间。选择一种或另一种类型的置信区间,通常取决于样本比例是否是良好的近似正态分布。

[0141] 在本发明的方法的一个具体实施例中,在量化细菌序列之前,从粪便样本中提取DNA。从粪便样本提取DNA的若干方法是公知的,所有这些方法均依靠化学或机械破碎,使用洗涤剂进行裂解,或这些方法的组合(Kennedy A. et al., PLoS One, 2014; 9 (2) : e88982)。粪便细菌DNA的提取方法是已知的,例如从M Corist et al., Journal of Microbiological Methods, 2002; 50 (2) : 131-139, Whitney D et al., Journal of Molecular Diagnostics, American Society for Investigative Pathology, 2004; 6 (4) : 386-395和WO2003/068788。

[0142] 优选的,所述方法使用机械破碎(如高速珠击提取)、化学裂解和最终纯化步骤的组合,优选使用硅胶膜柱,如包括在市售DNA提取试剂盒“**MobioPowerSoil®** DNA extraction procedure”(Mo-Bio Laboratories Inc.)、**FastDNA®** SPIN Kit for soil procedure (MP biomedical) 和**NucleoSpin®** Soil (Macherey-Nagel GmbH&Co. KG) 中的硅胶膜柱。来自粪便样本的DNA提取物中存在PCR抑制剂,如胆红素,胆汁盐和复合碳水化合物,是确定来自粪便样本的DNA提取物中的DNA生物标记的难点之一(Fleckna et al., Mol Cell Probes, 2007; 21 (4) : 282-7)。优选的DNA提取方法是提供低量PCR抑制剂的粪便提取物的方法,如少于5%,优选为小于2%,更优选为小于1%,更优选为小于0.5%,例如小于0.25%,0.1%,0.05%或0.01%。

[0143] 在一个具体的实施例中,本发明的方法包括:

[0144] a. 通过定量PCR从所述人粪便样本中确定选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列水平, 其中SEQ ID NO:1水平的测定使用了与SEQ ID NO:2-13具有至少90%一致性的引物; SEQ ID NO:4水平的测定使用了与SEQ ID NO:5-6具有至少90%一致性的引物; SEQ ID NO:7水平的测定使用了与SEQ ID NO:8-9具有至少90%一致性的引物; SEQ ID NO:10水平的测定使用了与SEQ ID NO:11-12具有至少90%一致性的引物; 和/或

[0145] b. 如步骤a所述地通过定量PCR从所述人粪便样本中确定选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10中的至少两种16S rDNA细菌序列水平, 并确定选自以下的比率的至少一种: SEQ ID NO:10/SEQ ID NO:4, SEQ ID NO:7/SEQ ID NO:4, SEQ ID NO:4/SEQ ID NO:1, SEQ ID NO:7/SEQ ID NO:1, SEQ ID NO:7/SEQ ID NO:10和SEQ ID NO:10/SEQ ID 1。

[0146] 在一个优选的实施例中, 本发明的从人粪便样本筛查CRC和/或腺瘤性息肉的方法, 包括:

[0147] a. 通过定量PCR从所述人粪便样本中确定选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列的浓度, 其中SEQ ID NO:1水平的测定使用了与SEQ ID NO:2-13具有至少90%一致性的引物; SEQ ID NO:4水平的测定使用了与SEQ ID NO:5-6具有至少90%一致性的引物; SEQ ID NO:7水平的测定使用了与SEQ ID NO:8-9具有至少90%一致性的引物; SEQ ID NO:10水平的测定使用了与SEQ ID NO:11-12具有至少90%一致性的引物;

[0148] b. 可选地, 如步骤a所述地, 通过定量PCR从所述人粪便样本中确定选自SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10中的至少两种16S rDNA细菌序列水平, 并确定选自以下的比率的至少一种: SEQ ID NO:10/SEQ ID NO:4, SEQ ID NO:7/SEQ ID NO:4, SEQ ID NO:4/SEQ ID NO:1, SEQ ID NO:7/SEQ ID NO:1, SEQ ID NO:7/SEQ ID NO:10和SEQ ID NO:10/SEQ ID 1; 和(c) 基于来自步骤(a)的序列的至少一个的浓度和/或来自步骤(b)的至少一个比率的值, 对CRC和/或腺瘤性息肉进行诊断、确定风险或确定复发。

[0149] 在本发明的方法的另一个实施例中, 所述相对于SEQ ID NO:2-13具有至少90%一致性的引物是相对于SEQ ID NO:2-13具有至少95%一致性的引物, 所述引物相对于SEQ ID NO:5-6具有至少90%一致性的引物为至少90%一致性具有至少95%一致性的引物, 所述相对于SEQ ID NO:8-9具有至少90%一致性的引物是相对于SEQ ID NO:8-9具有至少95%一致性的引物, 且所述相对于SEQ ID NO:11-12具有至少90%一致性的引物是相对于SEQ ID NO:11-12具有至少95%一致性的引物。

[0150] 在一个特别优选的实施例中, 所述相对于SEQ ID NO:2-13具有至少95%一致性的引物是由SEQ ID NO:2-13所示的引物, 所述相对于SEQ ID NO:5-6具有至少95%的一致性的引物是由SEQ ID NO:5-6所示的引物, 所述相对于SEQ ID NO:8-9具有至少95%的一致性的引物是由SEQ ID NO:8-9所示的引物, 且所述相对于SEQ ID NO:11-12具有至少95%的一致性的引物是由SEQ ID NO:11-12所示的引物。

[0151] 在本申请中, 如果问题序列残基与确定序列残基95%相同, 则问题序列相对于确定序列具有95%一致性。

[0152] 在一个优选的实施例中, 本发明的方法是用于结肠直肠癌的诊断方法, 且在存在

如下情况时,在步骤(c)中诊断为结肠直肠癌:如果SEQ ID NO:1的Ct值大于Ct临界值37.2,或者如果SEQ ID NO:4的Ct值高于Ct临界值16.56,或者如果SEQ ID NO:7的Ct值高于Ct临界值22.97,或者如果SEQ ID NO:10的Ct值高于Ct临界值19.24,或者如果比率值SEQ ID NO:10/SEQ ID NO:4低于临界值1.16,或者如果比率值SEQ ID NO:7/SEQ ID NO:4低于临界值1.40,或者如果比率值SEQ ID NO:4/SEQ ID NO:1高于临界值0.44,或者如果比率值SEQ ID NO:7/SEQ ID NO:1高于临界值0.62,或者如果比率值SEQ ID NO:10/SEQ ID:1高于临界值0.52,或者如果the SEQ ID NO:10/SEQ ID:7高于临界值1.2。

[0153] 在另一个优选的实施例中,本发明的方法是用于诊断发展结肠直肠癌的风险的方法,其中在存在如下情况时,在步骤(c)中诊断为有发展结肠直肠癌的风险:如果SEQ ID NO:1的Ct值高于Ct临界值36.45,或者如果SEQ ID NO:4的Ct值高于Ct临界值14.02,或者如果SEQ ID NO:7的Ct值高于Ct临界值24.76,或者如果SEQ ID NO:10的Ct值高于Ct临界值21.42,或者如果比率值SEQ ID NO:10/SEQ ID NO:4低于临界值1.56,或者如果比率值SEQ ID NO:7/SEQ ID NO:4低于临界值1.78,或者如果比率值SEQ ID NO:4/SEQ ID NO:1高于临界值0.38,或者如果比率值SEQ ID NO:7/SEQ ID NO:1高于临界值0.68,或者如果比率值SEQ ID NO:10/SEQ ID 1高于临界值0.59,或者如果SEQ ID NO:10/SEQ ID:7高于临界值1.16。

[0154] 在另一个实施例中,本发明的方法还包括测定和/或量化已知表明CRC的一种或多种分子生物标记。优选地,使用多重标记物组。所述多重标记物组可以包括测定存在于一个或多个下列基因中的点突变:APC,K-ras,BRAF,p53,NRAS,PIK3CA,CDK8,CMYC,CCNE1,CTNNB1,NEU(HER2),MYB,FBXW7,PTEN,SMAD4,SMAD2,SMAD3,TGFβ11R,TCF7L2,ACVR2和BAX(Fearon,E.R.,Annu.Rev.Pathol.Mech.Dis.,2011;6:479-507)。可能用于确定的其他公知CRC标记物为CpG岛甲基化,每粪便重量单位的人DNA的量,和每粪便重量单位的总DNA量。

[0155] 在进一步的实施例中,本发明的方法进一步包括将测定结果存储在计算机可读介质中。如本申请中所使用的“计算机可读介质”可以是可包含、存储、通信、传播或传送本发明的方法中确定的结果的任何设备。该介质可以是电子、磁、光、电磁、红外线或半导体系统(或装置或设备)或传播介质。

[0156] 在本发明的第二方面中,涉及将选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列或其互补rRNA序列作为生物标记用于以下各项的用途:对人类对象中的CRC和/或息肉进行诊断、早期检测、确定复发、确定发展风险或预测;或确定是否应当对所述人类对象进行结肠镜检查;或在所述人类对象为诊断患有CRC和/或息肉时确定预后;或指导对患有CRC和/或息肉的患者的治疗;其中所述细菌序列从所述人类对象的粪便样本中量化得到。

[0157] 优选地,所述核酸序列是选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10的16S rDNA序列。

[0158] 如本申请所用的术语“生物标志物”是指作为通常存在于身体样本中且可容易测定的物质的疾病标记物。所测量的量可以关联到潜在的疾病病理生理学,诸如CRC和/或息肉的存在或不存在,或未来CRC和/或息肉的存在或不存在的概率。在接受该疾病治疗的患者中,所述测量量也将关联到对治疗的响应性。

[0159] 在本发明的第三方面中,涉及一种试剂盒,包括:

[0160] a.选自以下的试剂:

[0161] i.能够与选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列或其互补rRNA序列特异杂交的核酸探针;或

[0162] ii.能够特异扩增选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10中的至少一种16S rDNA细菌序列的核酸引物对;和

[0163] b.用于量化根据权利要求1至12中任一项所述方法中的人类粪便样本中的一种或多种所述序列的水平的说明书。

[0164] 如本申请所用的术语“探针”是指合成或生物方法产生的核酸,长度为10和285个碱基对之间,含有允许在预定条件下对目标核酸序列进行特异性优先杂交的特异性核酸序列,并且可选地含有用于检测或提高分析性能的部分。通常至少需要10个核苷酸才能统计获得特异性并形成稳定的杂交产物,且285个核苷酸的最大值通常即为可以容易地调节反应参数以确定不匹配的序列和优先杂交的长度上限。探针可以可选地包含在一定的测定条件下促进其适当或最佳性能的某些成分。例如,探针可被修饰以用于改善其对核酸酶降解的抗性(例如通过封端),用于携带检测配体(例如,荧光素),或用于促进其被捕获到固体支持物上(例如,聚脱氧腺苷“尾”)。

[0165] 如本申请所用的术语“引物”是指可在扩增方法诸如聚合酶链式反应(“PCR”)中使用以扩增核苷酸序列的寡核苷酸。引物是基于特定目标序列的多核苷酸序列,例如具体的16S rDNA序列而设计的。引物和探针的设计和验证在本领域中是众所周知的。对于实时定量PCR方法,参见例如Rodriguez A et al. (Methods Mol Biol., 2015, 1275: 31-56)。

[0166] 本申请使用的“特异性”是指使用的核苷酸序列在测定条件(通常指严格条件)下杂交/扩增预定目标序列且基本上不会杂交/扩增非目标序列。

[0167] 本申请使用的“杂交”是指在预定的反应条件下,允许两条部分或完全互补的核酸链根据核酸碱基彼此配对明确的规则,以反平行方式结合以形成特异稳定氢键的双链核酸的处理过程,。

[0168] “实质性杂交”是指观察到的杂交量将令结果观察者认为该结果相对于阳性和阴性对照的杂交数据为阳性结果。“被认为是背景噪音”的数据则不是实质性杂交。

[0169] “严格杂交条件”是指约0.9摩尔浓度的NaCl的盐溶液中约35°C至65°C。严格性也可以通过如存在于杂交溶液中的离子物质的浓度和种类,所含变性剂的种类和浓度,以及杂交的温度这类反应参数来控制。通常随着杂交条件变得更加严格,若要形成稳定杂交物,则较长的探针是优选的。作为一项规则,杂交条件的严格性是将决定需使用的优选探针的某些特性。

[0170] 优选地,所述核酸序列是选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7和SEQ ID NO:10的16S rDNA序列。

[0171] 在一个具体的实施例中,所述试剂盒用于对人类对象中的CRC和/或息肉进行诊断、早期检测、确定复发、确定发展风险或预测;或确定是否应当对所述人类对象进行结肠镜检查;或在所述人类对象为诊断患有CRC和/或息肉时确定预后;或指导对患有CRC和/或息肉的患者的治疗。在一个优选的实施例中,所述试剂盒还可以包括DNA提取装置,用于进行杂交和/或扩增的装置,检测装置,和/或用于收集和/或保持生物样本的一个或多个容器。

[0172] 在一个具体的实施例中,本发明涉及一种用于从人粪便样本诊断结肠直肠癌的试剂盒,包括至少一对选自SEQ ID NO:2-13,SEQ ID NO:5-6,SEQ ID NO:8-9和SEQ ID NO:11-12的引物组对。

[0173] 在本发明的第四方面中,涉及选自SEQ ID NO:2,SEQ ID NO:13,SEQ ID NO:5,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:9,SEQ ID NO:11和SEQ ID NO:12或与其具有至少90%一致性的核酸序列;或与其具有至少90%的一致性,优选为具有95%的一致性,更优选为具有96%,97%,98%,99%或100%的一致性。优选地,所述序列在对选自SEQ ID NO:1,SEQ ID NO:4,SEQ ID NO:7and SEQ ID NO:10的至少一个16S rDNA序列进行量化的扩增测定法(如定量PCR)中用作引物或在杂交试验中用作核酸探针。

[0174] 在一个具体的实施例中,所述核酸序列是选自SEQ ID NO:2-13,SEQ ID NO:5-6,SEQ ID NO:8-9和SEQ ID NO:11-12的PCR引物对。在另一个具体实施例中,所述核酸序列是选自SEQ ID NO:2,SEQ ID NO:13,SEQ ID NO:5,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:9,SEQ ID NO:11和SEQ ID NO:12的核酸探针。

[0175] 在本发明的第五方面中,涉及将选自SEQ ID NO:2,SEQ ID NO:13,SEQ ID NO:5,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:9,SEQ ID NO:11和SEQ ID NO:12或与其具有至少90%一致性的核酸序列用于本发明的筛查方法的用途,其中SEQ ID NO:2和/或SEQ ID NO:13用于SEQ ID NO:1的定量;SEQ ID NO:5和/或SEQ ID NO:6用于SEQ ID NO:4的定量;SEQ ID NO:8和/或SEQ ID NO:9用于SEQ ID NO:7的定量;且SEQ ID NO:11和/或SEQ ID NO:12是用于SEQ ID NO:10的定量。

[0176] 应该理解的是,实施例中描述的特定实施例仅作为举例目的,而不限制本发明。本发明的原理特征可以用于各种实施例中,而不脱离本发明的范围。本领域的技术人员知晓或能够确定仅使用常规实验、本申请描述的具体程序的众多等同形式,这些等同形式应当认为在本发明的范围内,并且由权利要求书所涵盖。

[0177] 说明书中提及的所有出版物和专利申请均为本发明所属领域的技术人员水平。本申请中的所有出版物和专利申请通过引用方式并入本申请,其程度如同每个单独的出版物或专利申请具体且单独地表明通过引用方式并入本申请。

[0178] “一”或“一个”的使用,可以指“一个”,但也与“一个或多个”,“至少一个”和“一个或多个”的意思相一致。

[0179] 本申请全文中的术语“约”是指所指示数值±其值的5%,优选为指示值±其值的2%,最优选地,术语“约”是指恰好为指示值(±0%)。

[0180] 实施例

[0181] 实施例1.-材料和方法

[0182] 1.生物样本

[0183] 从27例最近被诊断为结肠直肠癌(CRC)(阶段0-1)的患者、24例Lynch综合征携带者(L)和19例健康个体(C)处得到粪便标本。所有患者在样本采集前至少15天进行结肠镜检查。

[0184] 所有人员都签署了相应的知情同意书。排除标准包括研究前的1个月内有过抗生素治疗以及年龄<18岁。

[0185] 2.16S rDNA细菌序列:

[0186] B3, B10, B41, B46, B48和B50序列对应于先前从未培养的细菌分离物中分离出的变性梯度凝胶电泳 (DGGE) 凝胶带。

[0187] -16S rDNA细菌序列SEQ ID NO:1 (B3), 在EMBL-EBI欧洲核苷酸存档 (ENA) 数据库中发表为GQ411111.1;

[0188] -16S rDNA细菌序列SEQ ID NO:4 (B10), 在EMBL-EBI欧洲核苷酸存档 (ENA) 数据库中发表为GQ411118.1;

[0189] -16S rDNA细菌序列SEQ ID NO:7 (B46), 在EMBL-EBI欧洲核苷酸存档 (ENA) 数据库中发表为GQ411150.1;

[0190] -16S rDNA细菌序列SEQ ID NO:10 (B48), 在EMBL-EBI欧洲核苷酸存档 (ENA) 数据库中发表为GQ411152.1;

[0191] -16S rDNA细菌序列B41, 在EMBL-EBI欧洲核苷酸存档 (ENA) 数据库中发表为GQ411145.1; 和

[0192] -16S rDNA细菌序列B50, 在EMBL-EBI欧洲核苷酸存档 (ENA) 数据库中发表为GQ411154.1

[0193] 3. 样本采集, 保存和存储

[0194] 24小时的粪便标本采集在粪便容器中并储存在-20℃下一个月, 并在ISO 9001条件下运输到-80℃冰柜。

[0195] 4. 样本处理和DNA提取

[0196] 对于粪便的16S-DNA提取, 使用NucleoSpin®Soil. (Macherey-Nagel)。

[0197] 在生物II级安全条件下, 将粪便样本200毫克-500毫克置于Nucleospin珠管中。取700微升SL1和150微升强化剂SX加入到珠管中。

[0198] 5. 然后, 根据制造商的说明提取DNA并进行纯化, 并用洗脱缓冲液50微升或10mM Tris盐酸pH值=7.4洗脱。

[0199] 6. 计算DNA浓度

[0200] 提取物的DNA浓度通过使用Qubit® 2.0 Fluorometer Catalog n° Q32866以Quant-IT dsDNA Broad-Range Assay Kit进行荧光分析确定。这种方法具有对dsDNA相对于RNA的高度选择性, 且其在2ng至1000ng的范围内提供了准确定量。在此范围内, 荧光线性相关于DNA浓度。用5微升提取物进行DNA测定。

[0201] 7. 粪便样本中提取的DNA的定量实时PCR (qPCR)

[0202] 通过定量实时PCR对来自粪便样本的DNA进行分析。特别是, 我们评估了每个样本中4种细菌种群的数量: B3, B10, B48和B46。

[0203] 用定量实时PCR以荧光dsDNA染色量化细菌序列。在这项研究中, 我们已经使用了预混BRYT® (Promega (Madison, USA))。通过Furet. et人描述的方法 (.FEMS Microbiol Ecol., 2009; 68 (3) : 351-62) 扩增真细菌。

[0204] 每个样本的细菌丰度表达为以总DNA浓度归一化的Ct。

[0205] Ct (循环阈值) 定义为荧光信号超过阈值所需的q-PCR循环数目。Ct水平与样本中的目标核酸浓度的对数成反比 (即Ct水平越低, 样本中的目标核酸的量越高)。实时分析经过40个扩增循环。

[0206] 8. 统计分析方法

[0207] 所述数据的统计正态分布通过Kolmogorov-Smirnov检验进行分析。根据数据是否存在统计正态分布,采用适当的统计检验来比较下列各组。

[0208] 经分析各组分别为:CRC vs. 健康;CRC vs.Lynch vs. 健康;CRC+Lynch vs. 健康;和CRC vs.Lynch-高风险vs.Lynch-低风险vs. 健康。

[0209] 对于所述组,分析变量为:

[0210] -4种细菌序列量化,表达为Ct;

[0211] -4种细菌序列比率的量化;

[0212] -重量;和

[0213] -DNA浓度。

[0214] 使用二元回归和ROC分析,确定健康和CRC条件及风险(表达为Outfit并分离为两种风险情况)之间的细菌序列量化相关性,以及检验的灵敏度和特异性。

[0215] 未观察到DNA和重量的组间差异。观察到Lynch和CRC+Lynch和健康志愿者(C)之间的真细菌(Eub)量不同,但这被认为是正常的,因为该疾病的状况意味着系统发育核心的变化,其中包括具有较少DNA拷贝的细菌。这并不意味着细菌总数的减少(参见例如Sobhani et al., PLoS One, 2011, 27; 6 (1) : e16393)。出于这个原因,没有把Eub量化用作归一化目的的内部参考。

[0216] 实施例2:引物设计和验证

[0217] 用于对活检中的细菌标记B3, B10, B46, B48, B41和B50进行量化的引物,其基于与此前使用以下生物信息工具从系统发育组获得的序列的对比分析进行设计:来自欧洲生物信息研究所(European Bioinformatics Institute(www.ebi.ac.uk))的ClustalX, Netprimer((Premier Biosoft)和PrimerExpress(Life Tecnologies-Thermo Fischer)。

[0218] 选择的检测系统是SybrGreen®。弃去了相对于相邻组具有小于3该不同位置的引物组,还弃去了解离曲线中的Tm值不同于期望值的引物组。在活检样本和粪便样本中进行引物验证。测定解离曲线以分析引物性能。

[0219] 在图8a至13b中,分别展示了用于验证为扩增B3, B10, B46, B48, B41与B50所设计的引物的扩增曲线和解离曲线。选择了展示出充分解离曲线(每个引物组为单峰)的引物用于粪便样本中的16S rDNA的细菌量化。值得注意的是,选择了扩增B3, B10, B46和B48的引物,其具体如下面的表9所示。B51和B40被弃去,因为缺乏特异性(溶解曲线每个引物组为多峰值)。

[0220]

B3正向(SEQ ID NO:2)	GGAGGCCTTCGGGTCGTAA
B3反向(SEQ ID NO:13)	AGGTTCCGGGGGCTTCGG
B10正向(SEQ ID NO:5)	CAACAAGGTAAGTGACGGC
B10反向(SEQ ID NO:6)	CGCCTACCTGTGCACTACTC
B46正向(SEQ ID NO:8)	TCCACGTAAGTCACAAGCG
B46反向(SEQ ID NO:9)	CGCCTACCTGTGCACTACTC
B48正向(SEQ ID NO:11)	GTACGGGGAGCAGCAGTG
B48反向(SEQ ID NO:12)	GACTACTAGATGCACAGTTTCC

[0221] 表9:选择用于粪便样本中的16S rDNA细菌序列B3, B10, B46和B48的qPCR扩增引物

对。

[0222] 实施例3:结肠直肠癌患者VS健康对象的16S rDNA生物标志物分析

[0223] 对7例对照和9例结肠直肠癌患者总共16个粪便样本进行了分析。分析粪便样本中的如实施例1所记载的每个细菌标志物的定量分析表示为Ct值。Ct(循环阈值)定义为荧光信号超过阈值所需的q-PCR循环数。Ct水平与样本中的目标核酸浓度的对数成反比(即CT水平越低,样本中的目标核酸的量越高)。实时分析经过40个扩增循环。所得结果如下所示在表1和2中。表1为Ct绝对值,表2为Ct比率。使用了双样本t检验。

[0224] 表1:Ct绝对值

[0225]

ID	组别	条件	B48	B46	B10	B3
F10	CRC	1	18.59	22.78	16.33	33.64
F11	CRC	1	20.33	26.45	20.37	41.57
F3	CRC	1	23.52	26.2	19.8	38.45
F4	CRC	1	16.41	20.87	14.88	
F5	CRC	1	19.04	24	17.61	35.68
F6	CRC	1	23.22	27.67	21.82	
F16	CRC	1		20.75	14.35	
F7	CRC	1	23.19	28.2	22.34	
F8	CRC	1	18.82	22.04	15.61	
F12	C	0	17.55	20.9	14.95	
F13	C	0	19.36	24.78	19.22	
F14	C	0	19.46	21.44	15.07	37.34
F15	C	0	17.55	19.85	12.84	36.98
F9	C	0	17.61	21.8	15.41	35.66
F1	C	0	18.09	21.97	14.79	38.3
F2	C	0	17.07	20.57	12.76	
平均值	CRC		20.39	24.329	18.12	37.34
	C		18.099	21.616	15.01	37.07

[0226]

标准偏差	CRC	2.65	2.89	3.04	3.44
	C	0.94	1.58	2.15	1.09
P 值	CRC vs C	0.025	0.021	0.019	0.444
中点(临界值)		19.244	22.972	16.56	37.2
灵敏度%		66.67	83.33	83.33	50
特异性%		55.56	60	60	50
准确度*		60	68.75	68.75	50

[0227] \*准确度的确定:  $(TN+TP) / (TN+TP+FN+FP)$ 。

[0228] 表2:Ct比率

[0229]

ID	组别	条件	B48/B10	B10/B3	B46/B10	B46/B48	B46/B3	B48/B3
F10	CRC	1	1.14	0.49	1.39	1.23	0.68	0.55
F11	CRC	1	1	0.49	1.3	1.3	0.64	0.49
F3	CRC	1	1.19	0.51	1.32	1.11	0.68	0.61
F4	CRC	1	1.1		1.4	1.27		
F5	CRC	1	1.08	0.49	1.36	1.26	0.67	0.53
F6	CRC	1	1.06		1.27	1.19		
F16	CRC	1			1.45			
F7	CRC	1	1.04		1.26	1.22		
F8	CRC	1	1.21		1.41	1.17		
F12	C	0	1.17		1.4	1.19		
F13	C	0	1.01		1.29	1.28		
F14	C	0	1.29	0.4	1.42	1.1	0.57	0.52
F15	C	0	1.37	0.35	1.55	1.13	0.54	0.47
F9	C	0	1.14	0.43	1.41	1.24	0.61	0.49
F1	C	0	1.22	0.39	1.49	1.21	0.57	0.47
F2	C	0	1.34		1.61	1.21		
平均值	CRC		1.1	0.5	1.35	1.22	0.67	0.55
	C		1.22	0.39	1.45	1.19	0.57	0.49
标准偏差	CRC		0.07	0.01	0.07	0.06	0.02	0.05
	C		0.13	0.04	0.11	0.06	0.03	0.02
P 值	CRC vs							
	C		0.02	0.001	0.018	0.224	0.001	0.045
中点 (临界值)			1.16123	0.4441	1.40243	1.20669	0.62	0.52
灵敏度%			75	100	77.78	55.56	100.00	75.00
特异性%			71.43	100	71.43	42.86	100.00	75.00
准确度*			73.33	100	75	50	100.00	75.00

[0230] \*准确度的确定:  $(TN+TP) / (TN+TP+FN+FP)$ 。

[0231] SEQ ID NO:10/SEQ ID NO:4比率的结肠直肠癌诊断的灵敏度和特异性分别是75和71%；对比率SEQ ID NO:7/SEQ ID NO:4分别为77%和71%；对比率SEQ ID NO:4/SEQ ID NO:1分别为100%和100%；对比率SEQ ID NO:7/SEQ ID NO:1分别为100%和100%；对比率SEQ ID NO:10/SEQ ID NO:1分别为75%和75%。

[0232] 图1,图2和3是Ct值的比率的图形表示 (SEQ ID NO 10/SEQ ID NO:4, SEQ ID NO:7/SEQ ID NO:4和SEQ ID NO:4/SEQ ID NO:1)。

[0233] 图4和5是Ct绝对值的图形表示 (SEQ ID NO:4, SEQ ID NO:7)。

[0234] 实施例4:Lynch综合征患者的16S rDNA生物标志物分析:CRC vs高风险(有息肉)与低风险(无息肉)

[0235] 此分析的目的是确定生物标志物预测值,以在产生临床迹象前检测结肠直肠癌风险。

[0236] 总共对8例Lynch综合征患者(具有发展结肠直肠癌的遗传性增加风险)进行了分析。所有个体都在最长1年的时期内进行了结肠镜检查,然后进行粪便样本采集。根据内镜

检查,其中4例曾经有过恶性息肉(命名为高风险),4例没有过任何息肉(命名为低风险)。

[0237] 使用实施例2中所述的引物,确定分析人类粪便样本中如实施例1所记载的每个细菌标志物的定量,并量化为Ct值。

[0238] 所得结果如下所示在表3和表4中。表3为Ct绝对值且表4为Ct比率。使用了双样本t检验。

[0239] 表3:Ct绝对值

[0240]

ID	组别	条件	B48	B46	B10	B3
F10	CRC	1	18.59	22.78	16.33	33.64
F11	CRC	1	20.33	26.45	20.37	41.57
F3	CRC	1	23.52	26.20	19.80	38.45
F4	CRC	1	16.41	20.87	14.88	
F5	CRC	1	19.04	24.00	17.61	35.68
F6	CRC	1	23.22	27.67	21.82	
F16	CRC	1		20.75	14.35	
F7	CRC	1	23.19	28.20	22.34	
F8	CRC	1	18.82	22.04	15.61	
MIL1	高风险	2	22.58	25.87	14.56	35.56
MIL2	高风险	2	24.53	25.79	14.73	37.06
MIL5	高风险	2	22.56	24.57	14.16	38.44
MIL4	高风险	2	20.00	29.90	18.41	37.99
MIL6	低风险	3	20.15	24.90	15.20	37.58
MIL3	低风险	3	18.11	20.45	10.34	31.16
MIL7	低风险	3	22.81	24.70	13.35	41.40
MIL8	低风险	3	20.60	21.88	11.38	32.42
平均值	CRC		20.39	24.33	18.12	37.34
	高风险		22.42	26.53	15.46	37.26
	低风险		20.42	22.98	12.57	35.64
标准偏差	CRC		2.65	2.89	3.04	3.44
	高风险		1.86	2.32	1.98	1.27
	低风险		1.93	2.18	2.15	4.74
P 值	CRC vs 高风险		0,10	0,10	0,070	0,48
	CRC vs 低风险		0,49	0,21	0,004	0,29
	高风险 vs 低风险		0,09	0,03	0,047	0,27
中点(临界值) 高风险 vs 低风险			21,42	24,76	14,02	36,45
灵敏度% 高风险 vs 低风险			75,00	75,00	80,00	60,00
特异性% 高风险 vs 低风险			75,00	75,00	100,00	66,67
准确度 高风险 vs 低风险			75,00	75,00	87,50	62,50

[0241] 表4 Ct比率

[0242]

ID	组别	条件	B48/B10	B10/B3	B46/B10	B46/B48	B46/B3	B48/B3
F10	CRC	1	1.14	0.49	1.39	1.23	0.68	0.55
F11	CRC	1	1.00	0.49	1.30	1.30	0.64	0.49
F3	CRC	1	1.19	0.51	1.32	1.11	0.68	0.61
F4	CRC	1	1.10		1.40	1.27		
F5	CRC	1	1.08	0.49	1.36	1.26	0.67	0.53
F6	CRC	1	1.06		1.27	1.19		
F16	CRC	1			1.45			
F7	CRC	1	1.04		1.26	1.22		
F8	CRC	1	1.21		1.41	1.17		
MIL1	高风险	2	1.55	0.41	1.78	1.15	0.73	0.63
MIL2	高风险	2	1.67	0.40	1.75	1.05	0.70	0.66
MIL5	高风险	2	1.59	0.37	1.74	1.09	0.64	0.59
MIL4	高风险	2	1.09	0.48	1.62	1.49	0.79	0.53
MIL6	低风险	3	1.33	0.40	1.64	1.24	0.66	0.54
MIL3	低风险	3	1.75	0.33	1.98	1.13	0.66	0.58
MIL7	低风险	3	1.71	0.32	1.85	1.08	0.60	0.55
MIL8	低风险	3	1.81	0.35	1.92	1.06	0.67	0.64
平均值	CRC		1.10	0.50	1.35	1.22	0.67	0.55
	高风险		1.47	0.41	1.72	1.20	0.71	0.60
	低风险		1.65	0.35	1.85	1.13	0.65	0.58
标准偏差	CRC		0.07	0.01	0.07	0.06	0.02	0.05
	高风险		0.26	0.05	0.07	0.20	0.06	0.06
	低风险		0.22	0.04	0.15	0.08	0.03	0.04
P 值	CRC vs 高风险		0.001491	0.009755	0.000001	0.38	0.11	0.10
	CRC vs 低风险		0.000029	0.000159	0.000002	0.02	0.19	0.21
	高风险 vs 低风险		0.172991	0.044382	0.087799	0.28	0.06	0.25
中点(临界值) 高风险 vs 低风险			1.56	0.38	1.78	1.16	0.68	0.59
灵敏度% 高风险 vs 低风险			66.67	75.00	80.00	50.00	75.00	75.00
特异性% 高风险 vs 低风险			60.00	75.00	100.00	50.00	75.00	75.00
准确度 高风险 vs 低风险			62.50	75.00	87.50	50.00	75.00	75.00

[0243] 微生物比率指标能够区分出具有发展结肠直肠癌的高风险的个体组(具有Lynch综合征和息肉的患者)。检测结肠直肠癌高风险人群的灵敏度和特异性对SEQ ID NO:4分别为80%和100%:对比率SEQ ID NO:4/SEQ ID NO:1分别为75%和75%;对比率SEQ ID NO:7/SEQ ID NO:4分别为80%和100%;对比率SEQ ID NO:7/SEQ ID NO:1分别为75%和75%;且对比率SEQ ID NO:10/SEQ ID NO:1分别为75%和75%。

[0244] 图6是Ct值的比率的图形表示(SEQ ID NO:4/SEQ ID NO:1),且图7是CT绝对值的图形表示(SEQ ID NO:4)。

[0245] 实施例5:结肠直肠癌患者VS健康对象的16S rDNA生物标志物分析

[0246] 粪便样本获取自最近诊断为结肠直肠癌(CRC)(阶段0-I)的27例患者和健康个体(C)。所有患者菌株样本采集前至少15天进行了结肠镜检查。

[0247] 所有人员都签署了相应的知情同意书。排除标准包括研究前的1个月内有过抗生素治疗以及年龄<18岁。

[0248] 表5(见下)所示为4个细菌序列的量化,CRC患者粪便中的Ct表现为显著增加。该事实意味着该组中的4个细菌含量较低。

[0249] 该4种细菌含量的显著减少可能是用于筛查结肠直肠癌患者的强大的工具。如果我们单独分析细菌标志物的性能,可以认识到,最好的结果获取自B3,其中鉴定CRC患者的特异性为94%,灵敏度为48%,且准确度为0.7;而排在第二位的是B46,具有84%的特异性、61.5%的灵敏度和0.698的准确度。

[0250] B48和B10也显示出0.69的准确性,但分别有100%的特异性、36%的灵敏度,和57%的特异性、89.5%的灵敏度。

[0251] 计算了细菌序列定量的Ct之间的比率。虽然没有观察到任何统计差异,但必须考虑到,将该比率应用到增加的样本中,可能是用于完成CRC筛查的有用算法。鉴于此,还必须考虑4个序列的组合,以增加试验的灵敏度和特异性。

[0252] 表5:所述4种细菌序列在健康个体和CRC患者中的定量(表示为Ct)以及用双样本t检验和ROC分析进行统计分析。

[0253]

ID	组别	条件	B3	B48	B10	B46
4	CRC	1	29.86	20.84	14.18	21.34
5	CRC	1	24.3	15.88	10.34	17.53
6	CRC	1	27.33	20.66	15.77	23.13
7	CRC	1	24.88	19.06	13.64	20.69
8	CRC	1	20.61	17.78	12.92	19.7
11	CRC	1	27.28	25.69	17.05	24.35
12	CRC	1	29.04	22.58	16.49	23.17
13	CRC	1	21.98	17.52	10.97	17.34
14	CRC	1	27.07	19.94	13.52	19.9
15	CRC	1	28.57	26.60	19.77	27.53
24	CRC	1	26.77	22.64	17.84	23.83
25	CRC	1	21.64	17.01	11.54	18.7
27	CRC	1	26.52	20.97	15.49	22.63
31	CRC	1	29.91	18.92	16.44	23.8
35	CRC	1	23.37	18.81	12.24	18.77
36	CRC	1	22.87	17.28	12.74	19.76
38	CRC	1		16.36	24.43	30.7
F10	CRC	1	20.49	18.59	16.33	22.78
F16	CRC	1	18.24		14.35	20.75
F11	CRC	1	29.19	20.33	20.37	26.45
F3	CRC	1	32.48	23.52	19.80	26.20

[0254]

F4	CRC	1	21.76	16.41	14.88	20.87
F5	CRC	1	20.56	19.04	17.61	24.00
F6	CRC	1	31.6	23.22	21.82	27.67
F7	CRC	1	27.14	23.19	22.34	28.20
F8	CRC	1	22.37	18.82	15.61	22.04
43	CRC	1	30.9	22.46	17.12	25.57
16	C	0	22.97	20.21	15.54	23.12
17	C	0	20.4	15.84	10.71	17.54
18	C	0	22.05	15.72	12.06	17.67
23	C	0	21.79	17.16	12.59	18.89
28	C	0	20	18.86	15.29	21.33
30	C	0	26.61	20.47	12.78	20.26
32	C	0	22.76	15.07	15.95	23.09
33	C	0	23.66	18.30	13.54	19.57
34	C	0	22.01	17.73	14.29	19.07
37	C	0	24.06	19.60	14.47	19.65
F1	C	0	24.71	18.09	14.79	21.97
F12	C	0	20.79	17.55	14.95	20.90
F13	C	0	25.1	19.36	19.22	24.78
F14	C	0	21.63	19.46	15.07	21.44
F2	C	0	20.42	17.07	12.76	20.57
F15	C	0	19.69	17.55	12.84	19.85
F9	C	0	23.29	17.61	15.41	21.80
39	C	0		18.03	11.95	18.97
平均值		CRC	25.64346154	20.15846154	16.13222222	22.86666667

	C	22.46705882	17.98222222	14.12111111	20.58166667
标准偏差	CRC	3.96679613	2.887129637	3.575465632	3.448853767
	C	1.948072011	1.514799756	1.957886432	1.927739456
P 值 (双样本 t 检验)	CRC vs C	0.038	0.019	0.049	0.021
CRC vs C	中点	26.69	20.90	15.75	22.0
	AUC	0.70	0.69	0.69	0.698
	灵敏度	48%	36%	57%	61.5%
	特异性	94%	100%	89.5%	84%

[0255] 图14至17分别提供了CRC组和健康 (C) 组中的B3, B10, B46和B48的Ct绝对值的图形表示。图27提供了健康vs CRC分析中的B3, B10, B46和B48的ROC曲线。

[0257] 实施例6: CRC VS Lynch (无临床症状) 与健康个体的16S rDNA生物标志物分析

[0258] 此分析的目的是确定生物标志物预测值, 以在产生临床迹象前检测结肠直肠癌风险。

[0259] 粪便样本获取自27例最近被诊断为结肠直肠癌 (CRC) (阶段0-I) 的患者、, 24例Lynch综合征携带者 (L) 和19例健康个体 (C) 处得到粪便标本。所有患者在样本采集前至少15天进行结肠镜检查。

[0260] 所有人员都签署了相应的知情同意书。排除标准包括研究前的1个月内有过抗生素治疗以及年龄<18岁。

[0261] 我们量化了健康个体、CRC最近确诊的患者和Lynch综合征携带者的粪便样本中的所述4种细菌序列。Lynch综合征携带者通常有发展成癌症的遗传倾向, 特别是结肠直肠癌。这一组由没有结肠直肠癌肿瘤和无临床标志的个人组成, 并且其认为是结肠直肠癌风险组。

[0262] 我们观察到, 与健康的患者群体相比, Lynch综合征组呈现出显著增加B3和B48Ct值。所以, Lynch组的粪便样本中B3和B48细菌含量低。

[0263] 事实上, 与CRC组相比, Lynch组并没有显示出除B10以外的细菌序列中的任何统计学差异。该组中观察到了Lynch组即使无临床体征和肿瘤仍与CRC组具有类似的微生物情况的暗示。这一事实意味着, 检测一个或多个所述4个细菌序列可能是筛查具有CRC风险的人和可能需要结肠镜检查的人的有力工具。

[0264] 表6: 所述4种细菌序列在健康个体、CRC患者和Lynch综合征携带者中的定量 (表示为Ct) 以及用单因素ANOVA和ROC分析进行统计分析。

[0265]

ID	组别	条件	B3	B48	B10	B46
4	CRC	1	29.86	20.84	14.18	21.34
5	CRC	1	24.3	15.88	10.34	17.53
6	CRC	1	27.33	20.66	15.77	23.13
7	CRC	1	24.88	19.06	13.64	20.69
8	CRC	1	20.61	17.78	12.92	19.7
11	CRC	1	27.28	25.69	17.05	24.35
12	CRC	1	29.04	22.58	16.49	23.17
13	CRC	1	21.98	17.52	10.97	17.34
14	CRC	1	27.07	19.94	13.52	19.9
15	CRC	1	28.57	26.60	19.77	27.53
24	CRC	1	26.77	22.64	17.84	23.83
25	CRC	1	21.64	17.01	11.54	18.7
27	CRC	1	26.52	20.97	15.49	22.63
31	CRC	1	29.91	18.92	16.44	23.8
35	CRC	1	23.37	18.81	12.24	18.77
36	CRC	1	22.87	17.28	12.74	19.76
38	CRC	1		16.36	24.43	30.7
F10	CRC	1	20.49	18.59	16.33	22.78
F16	CRC	1	18.24		14.35	20.75
F11	CRC	1	29.19	20.33	20.37	26.45

[0266]

F3	CRC	1	32.48	23.52	19.80	26.20
F4	CRC	1	21.76	16.41	14.88	20.87
F5	CRC	1	20.56	19.04	17.61	24.00
F6	CRC	1	31.6	23.22	21.82	27.67
F7	CRC	1	27.14	23.19	22.34	28.20
F8	CRC	1	22.37	18.82	15.61	22.04
43	CRC	1	30.9	22.46	17.12	25.57
16	C	0	22.97	20.21	15.54	23.12
17	C	0	20.4	15.84	10.71	17.54
18	C	0	22.05	15.72	12.06	17.67
23	C	0	21.79	17.16	12.59	18.89
28	C	0	20	18.86	15.29	21.33
30	C	0	26.61	20.47	12.78	20.26
32	C	0	22.76	15.07	15.95	23.09
33	C	0	23.66	18.30	13.54	19.57
34	C	0	22.01	17.73	14.29	19.07
37	C	0	24.06	19.60	14.47	19.65
F1	C	0	24.71	18.09	14.79	21.97
F12	C	0	20.79	17.55	14.95	20.90
F13	C	0	25.1	19.36	19.22	24.78
F14	C	0	21.63	19.46	15.07	21.44
F2	C	0	20.42	17.07	12.76	20.57
F15	C	0	19.69	17.55	12.84	19.85
F9	C	0	23.29	17.61	15.41	21.80
39	C	0		18.03	11.95	18.97

[0267]

1	L	2	23.08	18.76	12.03	19.21
2	L	2	26.58	18.47	13.60	21.02
3	L	2	23.51	19.36	10.85	18.33
9	L	2		16.94	12.89	19.77
10	L	2	19.91	17.68	10.44	17.84
19	L	2	32.81	21.60	15.72	21.67
20	L	2	19.9	19.77	17.81	24.2
21	L	2	19.66	17.05	13.55	19.7
22	L	2	21.28	16.04	10.22	17.48
26	L	2	19.57	20.73	11.79	19.36
29	L	2	23.67	17.92	11.30	18.2
40	L	2	29.14	23.47	19.42	26.3
41	L	2	32.14	20.18	17.27	23.72
42	L	2	35.19	18.34	19.63	27.07
44	L	2	28.83	23.16	16.15	23.42
45	L	2	24.79	16.85	10.84	19.02
MIL1	L	2	37.1	22.60	14.69	25.52
MIL2	L	2	32.74	24.60	14.63	25.53
MIL3	L	2	27.80	18.00	10.34	20.28
MIL4	L	2	29.09	20.00	18.17	29.66
MIL5	L	2	33.45	22.60		24.81
MIL6	L	2	31.69	20.20	14.66	24.54
MIL7	L	2	35.88	22.90	13.05	24.59
MIL8	L	2	26.2	20.60	10.85	21.87
平均值		CRC	25.64346154	20.15846154	16.13222222	22.86666667

	C	22.46705882	17.98222222	14.12111111	20.58166667
	Lynch	27.56565217	19.90916667	13.9076087	22.21291667
标准偏差	CRC	3.96679613	2.887129637	3.575465632	3.448853767
	C	1.948072011	1.514799756	1.957886432	1.927739456
	Lynch	5.628873475	2.403976627	3.02383331	3.369239985
P 值 (单因素 ANOVA)	CRC vs C	0.038	0.019	0.049	0.021
	CRC vs L	0.381	0.988	0.057	0.849
	L vs C	0.004	0.018	0.989	0.163
C vs L	中点	25.65	19.69	11.87	23.27
	AUC	0.73	0.746	0.55	0.629
	灵敏度	60.9	54.2	34.8	45.8
	特异性	88.9	94.1	94.7	94.7
L vs CRC	中点	31.65	19.21	14.78	21.95
	AUC	0.60	0.487	0.69	0.55
	灵敏度	96	52	65.4	61.5
	特异性	34.8	58.3	69.6	54.2
CRC vs C	中点	26.69	20.90	15.75	22.0
	AUC	0.70	0.69	0.69	0.698
	灵敏度	48%	36%	57%	61.5%
	特异性	94%	100%	89.5%	84%

[0269] 图14至17分别提供了CRC、L和C组中的SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10的Ct绝对值的图形表示。

[0270] 图28提供了健康vs Lynch (L) 分析中的B3, B10, B46和B48的ROC曲线。

[0271] 图30为提供了CRC vs Lynch (L) 分析中的B3, B10, B46和B48的ROC曲线。

[0272] 实施例7: CRC+Lynch vs 健康个体的16S rDNA生物标志物分析

[0273] 根据实施例6的分析, Lynch综合征携带者具有与CRC患者类似的微生物情况。因此, 在这个实验中, 我们的目标是测试, 在结肠镜检查前, 使用对所述4个细菌序列的定量检测来在健康个体和结肠直肠癌风险个体 (CRC+L) 筛查结肠直肠癌风险个体的潜力。

[0274] 在表7中, 我们用Ct表示细菌序列的量化。

[0275] 我们观察到,B3,B48和B46显著增加了结肠直肠癌风险组(CRC+L)的Ct值。所以,这些细菌在粪便中的水平降低。

[0276] 虽然没有统计学差异( $p=0.137$ ),但B10在CRC+L中也以较低的水平存在。

[0277] 在检测结肠直肠癌中评估细菌量化的性能,表明B3和B48具有最佳准确性(0.7),且分别具有56.2%的灵敏度、88.9%的特异性和51%的灵敏度、84%的特异性。

[0278] 排在第二位的B46具有0.665的准确性和94.7%的灵敏度、23.12%的特异性。最后,B10具有0.58的准确性和44.9%的灵敏度、89.5%的特异性。

[0279] 计算了细菌序列定量的Ct之间的比率。虽然没有观察到统计学差异(B3/B46 $p=0.061$ ,Mann-Whitney U检验;B10/B3 $p=0.132$ ,双样本t检验分析;B46/B10 $p=0.127$ ,双样本t检验分析);我们必须考虑到,将该比率应用到增加样本中,可能是用于完成CRC筛查的有用算法。

[0280] 鉴于此,还必须考虑四个序列的组合,以增加试验的灵敏度和特异性。

[0281] 表7:所述4种细菌序列在健康个体(C)和结肠直肠癌风险个体(CRC+L)中的定量(表示为Ct)以及用双样本t检验和ROC分析进行统计分析。

[0282]

ID	组别	条件	B3	B48	B10	B46
4	CRC+L	3	29.86	20.84	14.18	21.34
5	CRC+L	3	24.3	15.88	10.34	17.53
6	CRC+L	3	27.33	20.66	15.77	23.13
7	CRC+L	3	24.88	19.06	13.64	20.69

[0283]

8	CRC+L	3	20.61	17.78	12.92	19.7
11	CRC+L	3	27.28	25.69	17.05	24.35
12	CRC+L	3	29.04	22.58	16.49	23.17
13	CRC+L	3	21.98	17.52	10.97	17.34
14	CRC+L	3	27.07	19.94	13.52	19.9
15	CRC+L	3	28.57	26.60	19.77	27.53
24	CRC+L	3	26.77	22.64	17.84	23.83
25	CRC+L	3	21.64	17.01	11.54	18.7
27	CRC+L	3	26.52	20.97	15.49	22.63
31	CRC+L	3	29.91	18.92	16.44	23.8
35	CRC+L	3	23.37	18.81	12.24	18.77
36	CRC+L	3	22.87	17.28	12.74	19.76
38	CRC+L	3		16.36	24.43	30.7
F10	CRC+L	3	20.49	18.59	16.33	22.78
F16	CRC+L	3	18.24		14.35	20.75
F11	CRC+L	3	29.19	20.33	20.37	26.45
F3	CRC+L	3	32.48	23.52	19.80	26.20
F4	CRC+L	3	21.76	16.41	14.88	20.87
F5	CRC+L	3	20.56	19.04	17.61	24.00
F6	CRC+L	3	31.6	23.22	21.82	27.67
F7	CRC+L	3	27.14	23.19	22.34	28.20
F8	CRC+L	3	22.37	18.82	15.61	22.04
43	CRC+L	3	30.9	22.46	17.12	25.57
1	CRC+L	3	23.08	18.76	12.03	19.21
2	CRC+L	3	26.58	18.47	13.60	21.02

[0284]

3	CRC+L	3	23.51	19.36	10.85	18.33
9	CRC+L	3		16.94	12.89	19.77
10	CRC+L	3	19.91	17.68	10.44	17.84
19	CRC+L	3	32.81	21.60	15.72	21.67
20	CRC+L	3	19.9	19.77	17.81	24.2
21	CRC+L	3	19.66	17.05	13.55	19.7
22	CRC+L	3	21.28	16.04	10.22	17.48
26	CRC+L	3	19.57	20.73	11.79	19.36
29	CRC+L	3	23.67	17.92	11.30	18.2
40	CRC+L	3	29.14	23.47	19.42	26.3
41	CRC+L	3	32.14	20.18	17.27	23.72
42	CRC+L	3	35.19	18.34	19.63	27.07
44	CRC+L	3	28.83	23.16	16.15	23.42
45	CRC+L	3	24.79	16.85	10.84	19.02
MIL1	CRC+L	3	37.1	22.60	14.69	25.52
MIL2	CRC+L	3	32.74	24.60	14.63	25.53
MIL3	CRC+L	3	27.80	18.00	10.34	20.28
MIL4	CRC+L	3	29.09	20.00	18.17	29.66
MIL5	CRC+L	3	33.45	22.60		24.81
MIL6	CRC+L	3	31.69	20.20	14.66	24.54
MIL7	CRC+L	3	35.88	22.90	13.05	24.59
MIL8	CRC+L	3	26.2	20.60	10.85	21.87
16	C	0	22.97	20.21	15.54	23.12
17	C	0	20.4	15.84	10.71	17.54
18	C	0	22.05	15.72	12.06	17.67

[0285]

23	C	0	21.79	17.16	12.59	18.89
28	C	0	20	18.86	15.29	21.33
30	C	0	26.61	20.47	12.78	20.26
32	C	0	22.76	15.07	15.95	23.09
33	C	0	23.66	18.30	13.54	19.57
34	C	0	22.01	17.73	14.29	19.07
37	C	0	24.06	19.60	14.47	19.65
F1	C	0	24.71	18.09	14.79	21.97
F12	C	0	20.79	17.55	14.95	20.90
F13	C	0	25.1	19.36	19.22	24.78
F14	C	0	21.63	19.46	15.07	21.44
F2	C	0	20.42	17.07	12.76	20.57
F15	C	0	19.69	17.55	12.84	19.85
F9	C	0	23.29	17.61	15.41	21.80
39	C	0		18.03	11.95	18.97
平均值		CRC+L	26.54	20.03	15.10	22.56
		C	22.46	17.98	14.12	20.58
标准偏差		CRC+L	4.86	2.64	3.48	3.39
		C	1.94	1.51	1.96	1.93
P 值 (双样本 t 检验)		CRC+L vs C	0.0001	0.001	0.137	0.004
		中点	25.65	19.69	15.58	23.13
		AUC	0.71	0.70	0.580	0.665
		灵敏度	56.2	51	44.9	94.7
		特异性	88.9	84	89.5	23.12

[0286] 图22分别提供了CRC+Lynch (L) 和健康 (C) 组中的SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7和SEQ ID NO:10的Ct绝对值的图形表示。

[0287] 图29提供了健康vs CCR+Lynch分析中的B3, B10, B46和B48的ROC曲线。

[0288] 实施例8:CRC vs L-高风险vs L-低风险VS健康个体的16S rDNA生物标志物分析

[0289] 在本例中,我们的目标是利用细菌序列定量在Lynch综合征携带者中确定不同级别的结肠直肠癌风险。

[0290] 我们根据患者的结直肠肿瘤背景及其最近一次结肠镜检查中存在腺瘤的情况,将Lynch综合征携带者分为高风险(L-高风险)与低风险(L-低风险)。我们有6个高风险和18个低风险个体。

[0291] 对所述四种细菌序列进行了量化,在不同组之间没有观察到统计学差异。只有对于B3水平,比较C vs L-低风险( $p=0.088$ ,Kruskal-Wallis检验)和C vs L-高风险( $p=0.143$ ,Kruskal-Wallis检验),有Ct值下降的趋势。

[0292] 计算了细菌序列定量的Ct之间的比率。

[0293] 使用了Kruskal-Wallis检验,发现B3/B10和B10/B46的比率在C和L低风险组之间有显著差异。比较C vs L-高风险组的B3/B48比率,显示存在显著差异。B3/B10和B48/B10的比率在CRC vs L-高风险中均显著不同。

[0294] 因为样本大小的原因,没有执行ROC分析,但考虑到统计分析,我们可以预测最佳比例可能是B3/B10,因为能够区分L-低风险组。

[0295] 鉴于此,还必须考虑四个序列的组合,以使得测试更强大,并且必须进行多变量逻辑回归分析,但是需要增加样本大小。

[0296] 表8:所述4种细菌序列在健康个体的CT(C),L-低风险,L-高风险和CRC中的定量,表达为Ct。

[0297]

ID	组别	条件	B3	B48	B10	B46	B3/B10	B48/B10	B3/B48	B10/B46
4	CRC	1	29,86	20,84	14,18	21,34	2,11	1,47	1,43	0,66
5	CRC	1	24,3	15,88	10,34	17,53	2,35	1,54	1,53	0,59
6	CRC	1	27,33	20,66	15,77	23,13	1,73	1,31	1,32	0,68
7	CRC	1	24,88	19,06	13,64	20,69	1,82	1,40	1,31	0,66
8	CRC	1	20,61	17,78	12,92	19,7	1,60	1,38	1,16	0,66
11	CRC	1	27,28	25,69	17,05	24,35	1,60	1,51	1,06	0,70
12	CRC	1	29,04	22,58	16,49	23,17	1,76	1,37	1,29	0,71
13	CRC	1	21,98	17,52	10,97	17,34	2,00	1,60	1,25	0,63
14	CRC	1	27,07	19,94	13,52	19,9	2,00	1,47	1,36	0,68
15	CRC	1	28,57	26,60	19,77	27,53	1,45	1,35	1,07	0,72
24	CRC	1	26,77	22,64	17,84	23,83	1,50	1,27	1,18	0,75
25	CRC	1	21,64	17,01	11,54	18,7	1,88	1,47	1,27	0,62
27	CRC	1	26,52	20,97	15,49	22,63	1,71	1,35	1,26	0,68
31	CRC	1	29,91	18,92	16,44	23,8	1,82	1,15	1,58	0,69
35	CRC	1	23,37	18,81	12,24	18,77	1,91	1,54	1,24	0,65
36	CRC	1	22,87	17,28	12,74	19,76	1,80	1,36	1,32	0,64
38	CRC	1		16,36	24,43	30,7		0,67		0,80
F10	CRC	1	20,49	18,59	16,33	22,78	1,25	1,14	1,10	0,72
F16	CRC	1	18,24		14,35	20,75	1,27			0,69
F11	CRC	1	29,19	20,33	20,37	26,45	1,43	1,00	1,44	0,77
F3	CRC	1	32,48	23,52	19,80	26,20	1,64	1,19	1,38	0,76
F4	CRC	1	21,76	16,41	14,88	20,87	1,46	1,10	1,33	0,71
F5	CRC	1	20,56	19,04	17,61	24,00	1,17	1,08	1,08	0,73
F6	CRC	1	31,6	23,22	21,82	27,67	1,45	1,06	1,36	0,79
F7	CRC	1	27,14	23,19	22,34	28,20	1,21	1,04	1,17	0,79
F8	CRC	1	22,37	18,82	15,61	22,04	1,43	1,21	1,19	0,71
43	CRC	1	30,9	22,46	17,12	25,57	1,80	1,31	1,38	0,67
16	C	0	22,97	20,21	15,54	23,12	1,48	1,30	1,14	0,67
17	C	0	20,4	15,84	10,71	17,54	1,90	1,48	1,29	0,61
18	C	0	22,05	15,72	12,06	17,67	1,83	1,30	1,40	0,68
23	C	0	21,79	17,16	12,59	18,89	1,73	1,36	1,27	0,67
28	C	0	20	18,86	15,29	21,33	1,31	1,23	1,06	0,72
30	C	0	26,61	20,47	12,78	20,26	2,08	1,60	1,30	0,63
32	C	0	22,76	15,07	15,95	23,09	1,43	0,94	1,51	0,69
33	C	0	23,66	18,30	13,54	19,57	1,75	1,35	1,29	0,69
34	C	0	22,01	17,73	14,29	19,07	1,54	1,24	1,24	0,75
37	C	0	24,06	19,60	14,47	19,65	1,66	1,35	1,23	0,74
F1	C	0	24,71	18,09	14,79	21,97	1,67	1,22	1,37	0,67
F12	C	0	20,79	17,55	14,95	20,90	1,39	1,17	1,18	0,72
F13	C	0	25,1	19,36	19,22	24,78	1,31	1,01	1,30	0,78

[0298]

F14	C	0	21,63	19,46	15,07	21,44	1,44	1,29	1,11	0,70
F2	C	0	20,42	17,07	12,76	20,57	1,60	1,34	1,20	0,62
F15	C	0	19,69	17,55	12,84	19,85	1,53	1,37	1,12	0,65
F9	C	0	23,29	17,61	15,41	21,80	1,51	1,14	1,32	0,71
39	C	0		18,03	11,95	18,97		1,51		0,63
1	L-低风险	4	23,08	18,76	12,03	19,21	1,92	1,56	1,23	0,63
3	L-低风险	4	23,51	19,36	10,85	18,33	2,17	1,78	1,21	0,59
9	L-低风险	4		16,94	12,89	19,77		1,31		0,65
10	L-低风险	4	19,91	17,68	10,44	17,84	1,91	1,69	1,13	0,58
19	L-低风险	4	32,81	21,60	15,72	21,67	2,09	1,37	1,52	0,73
20	L-低风险	4	19,9	19,77	17,81	24,2	1,12	1,11	1,01	0,74
21	L-低风险	4	19,66	17,05	13,55	19,7	1,45	1,26	1,15	0,69
26	L-低风险	4	19,57	20,73	11,79	19,36	1,66	1,76	0,94	0,61
29	L-低风险	4	23,67	17,92	11,30	18,2	2,09	1,59	1,32	0,62
40	L-低风险	4	29,14	23,47	19,42	26,3	1,50	1,21	1,24	0,74
42	L-低风险	4	35,19	18,34	19,63	27,07	1,79	0,93	1,92	0,73
44	L-低风险	4	28,83	23,16	16,15	23,42	1,79	1,43	1,24	0,69
MIL1	L-低风险	4	37,1	22,60	14,69	25,52	2,53	1,54	1,64	0,58
MIL2	L-低风险	4	32,74	24,60	14,63	25,53	2,24	1,68	1,33	0,57
MIL3	L-低风险	4	27,80	18,00	10,34	20,28	2,69	1,74	1,54	0,51
MIL6	L-低风险	4	31,69	20,20	14,66	24,54	2,16	1,38	1,57	0,60
MIL7	L-低风险	4	35,88	22,90	13,05	24,59	2,75	1,75	1,57	0,53
MIL8	L-低风险	4	26,2	20,60	10,85	21,87	2,41	1,90	1,27	0,50
2	L-高风险	5	26,58	18,47	13,60	21,02	1,95	1,36	1,44	0,65
22	L-高风险	5	21,28	16,04	10,22	17,48	2,08	1,57	1,33	0,58
41	L-高风险	5	32,14	20,18	17,27	23,72	1,86	1,17	1,59	0,73
45	L-高风险	5	24,79	16,85	10,84	19,02	2,29	1,55	1,47	0,57
MIL4	L-高风险	5	29,09	20,00	18,17	29,66	1,60	1,10	1,45	0,61
MIL5	L-高风险	5	33,45	22,60		24,81			1,48	
平均值	CRC		25,64	20,16	16,13	22,87	1,66	1,28	1,28	0,70
	C		22,47	17,98	14,12	20,58	1,60	1,29	1,25	0,68
	L-低风险		27,45	20,20	13,88	22,08	2,02	1,50	1,34	0,63
	L-高风险		27,89	19,02	14,02	22,62	1,96	1,35	1,46	0,63
标准偏差	CRC		3,97	2,89	3,58	3,45	0,29	0,21	0,14	0,05
	C		1,95	1,51	1,96	1,93	0,21	0,16	0,11	0,05
	L-低风险		6,08	2,40	2,96	3,09	0,44	0,27	0,25	0,08
	L-高风险		4,59	2,41	3,62	4,42	0,26	0,22	0,09	0,06
P 值 Kruskal-Wallis 检验	CRC vs C		0,32							
	CRC vs LL		1				0,017	0,022		
	CRC vs LH		1							
	C vs LL		0,088				0,016			0,01
	C vs LH		0,143						0,031	
LL vs LH		1								

[0299] 图23至26分别提供了CRC组,高风险L组,低风险L组和C组的以下Ct绝对值比率的图形表示:B48/B10,B3/B10,B46/B10和B3/B48。

## 序列表

- <110> 何塞普特鲁塔博士赫罗纳生物医学研究所  
赫罗纳大学
- <120> 用于通过定量 PCR 从人粪便样本诊断结肠直肠癌的方法、引物及试剂盒
- <130> 900 663
- <150> EP14382074
- <151> 2014-03-03
- <160> 13
- <170> PatentIn version 3.5
- <210> 1
- <211> 489
- <212> DNA
- <213> Artificial Sequence
- [0001] <220>
- <223> /mol\_type="unassigned DNA" /note="B3 16S ribosomal DNA sequence, partial  
sequence (accession  
number GQ411111.1)"
- <400> 1
- acggaggcct tcgggtcgta accgcttca gcagggaaga gtcaagactg tacctgcaga 60
- agaagccccg gctaactacg tgccagcagc cgcgtaata cgtagggggc gagcgttacc 120
- cggattcatt gggcgtaaag cgcgcgtagg cggccccgca ggccgggggt cgaagcgggg 180
- ggctcaacc cccgaagccc ccggaacctc cgcggcttgg gtccggtagg ggaggggtgga 240
- acaccgggtg tagcgggtgga atgcgcagat atcgggtgga acaccgggtg cgaaggcggc 300
- cctctgggcc gagaccgacg ctgaggcgcg aaagctgggg gagcgaacag gattagatac 360
- cctggtagtc ccagccgtaa acgatggacg ctaggtgtgg ggggacgata cccccgtgcc 420
- gcagccaacg cattaagcgt cccgcctggg gactacggcc gcaaggctaa aactcaaagg 480
- aattgacgg 489

	<210> 2		
	<211> 19		
	<212> DNA		
	<213> Artificial Sequence		
	<220>		
	<223> /mol_type="unassigned DNA" /note="B3 16S ribosomal DNA sequence forward primer"		
	<400> 2		
	ggaggccttc gggtcgtaa		19
	<210> 3		
	<211> 18		
	<212> DNA		
	<213> Artificial Sequence		
	<220>		
	<223> /mol_type="unassigned DNA" /note="B3 16S ribosomal DNA sequence reverse primer"		
[0002]	<400> 3		
	ccgaagcccc cggaacct		18
	<210> 4		
	<211> 485		
	<212> DNA		
	<213> Artificial Sequence		
	<220>		
	<223> /mol_type="unassigned DNA" /note="B10 16S ribosomal DNA sequence, partial sequence (accession number GQ411118.1)"		
	<400> 4		
	cttcgattg taaactcctg ttgttgagga gataatgacg gtactcaaca aggtaagtga		60
	cggtactacta cgtgccagca gccgcggtaa aacgtaggtc acaagcgttg tccggaatta		120
	ctgggtgtaa agggagcgca ggcgggaaga caagttggaa gtgaaatcca tgggtcaac		180

	ccatgaactg ctttcaaac tgttttctt gagtagtca caggtaggcg gaattcccgg	240
	tgtagcggtg gaatgcgtag atatcgggag gaacaccagt ggcgaaggcg gcctactggg	300
	caccaactga cgctgaggct cgaagtgtg ggtagcaaac aggattagat acctggtag	360
	tccacactgt aaacgatgat tactaggtgt tggaggattg acccctcag tgccgcagtt	420
	aaacaataa gtaatccacc tggggagtac gaccgcaagg ttgaaactca aaggaattga	480
	cggac	485
	<210> 5	
	<211> 19	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
[0003]	<223> /mol_type="unassigned DNA" /note="B10 16S ribosomal DNA sequence forward primer"	
	<400> 5	
	caacaaggta agtgacggc	19
	<210> 6	
	<211> 20	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
	<223> /mol_type="unassigned DNA" /note="B10 16S rDNA sequence reverse primer"	
	<400> 6	
	cgcttacctg tgcactactc	20
	<210> 7	
	<211> 474	
	<212> DNA	
	<213> Artificial Sequence	

<220>  
 <223> /mol\_type="unassigned DNA" /note="B46 16S ribosomal DNA sequence, partial  
 sequence (accession  
 number GQ411150.1)"

<400> 7  
 gtaacatttc tgaacaaga catacagcag tttctcagag ttcccaaact caaaggaatt 60  
 gacggaagcc gcggtccac gtaagtcaca agcgtgtcc ggaattactg ggtgtaaagg 120  
 gagcgcaggc gggaagacaa gttggaagtg aatccatgg gctcaacca taaactgctt 180  
 tcctaactgt ttttcttgag tagtcacag gtaggcggaa ttcccgggtg agcggtgga 240  
 tgcttagata tcgggaggaa caccattggc gaaggcggcc tactggactg caactgacgc 300  
 tgaggctcga aagtgtgggt agcaaacagg attagatacc ctggtagtcc actccgtaaa 360  
 ccatgattac tacgtgttg aggattgacc cttctgtgc cgagataac acaataagta 420  
 atccacctgg ggagtacgac cgcccgggtg agactcacag gaattgacgg actc 474

[0004]

<210> 8  
 <211> 19  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> /mol\_type="unassigned DNA" /note="B46 16S ribosomal DNA sequence forward  
 primer"

<400> 8  
 tccacgtaag tcacaagcg 19

<210> 9  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> /mol\_type="unassigned DNA" /note="B46 16S ribosomal DNA sequence reverse  
 primer"

<400> 9		
cgcttacctg tgcactactc		20
<210> 10		
<211> 566		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> /mol_type="unassigned DNA" /note="B48 16S ribosomal DNA sequence, partial sequence (accession number GQ411152.1)"		
<400> 10		
tctcttacgg ggagcagcag tggggaatat tgcacaatgg gggaaaccct gatgcagcga		60
cgccgcgtga gcgatgaagt atttcggtat gtaaagctct atcagcaggg aagaaaatga		120
cggtacctga ctaagaagcc ccggctaaat acgtgccagc agcccggtga atacgtatgg		180
[0005] tgcaagcgtt atccggattt actgggtgta aagggagcgt agacggagtg gcaagtctga		240
tgtgaaaacc cggggctcaa cccgggact gcattgaaa ctgtgcatct agagtgtcgg		300
agaggtaagc ggaattccta gtgtagcggg gaaatgcgta gatattagga ggaacaccag		360
tggcgaaggc ggcttactgg acgataactg acgctgaggc tcgaaagcgt ggggagcaaa		420
caggattaga taccctggta gtccagccg taaacgatga ctactaggtg tcggggagca		480
cagctcttcg gtgccgagc aaacgcaata agtattccac ctggggagta cgttcgcaag		540
aatgaaactc acagaatttg acggag		566
<210> 11		
<211> 18		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> /mol_type="unassigned DNA" /note="B48 16S ribosomal DNA sequence forward		

primer"

<400> 11

gtacggggag cagcagtg

18

<210> 12

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> /mol\_type="unassigned DNA"

/note="B48 16S ribosomal DNA sequence reverse

primer"

[0006]

<400> 12

gacactctag atgcacagtt tcc

23

<210> 13

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> /mol\_type="unassigned DNA"

/note="B3 16S ribosomal DNA sequence reverse

primer"

<400> 13

aggttccggg ggcttcgg

18

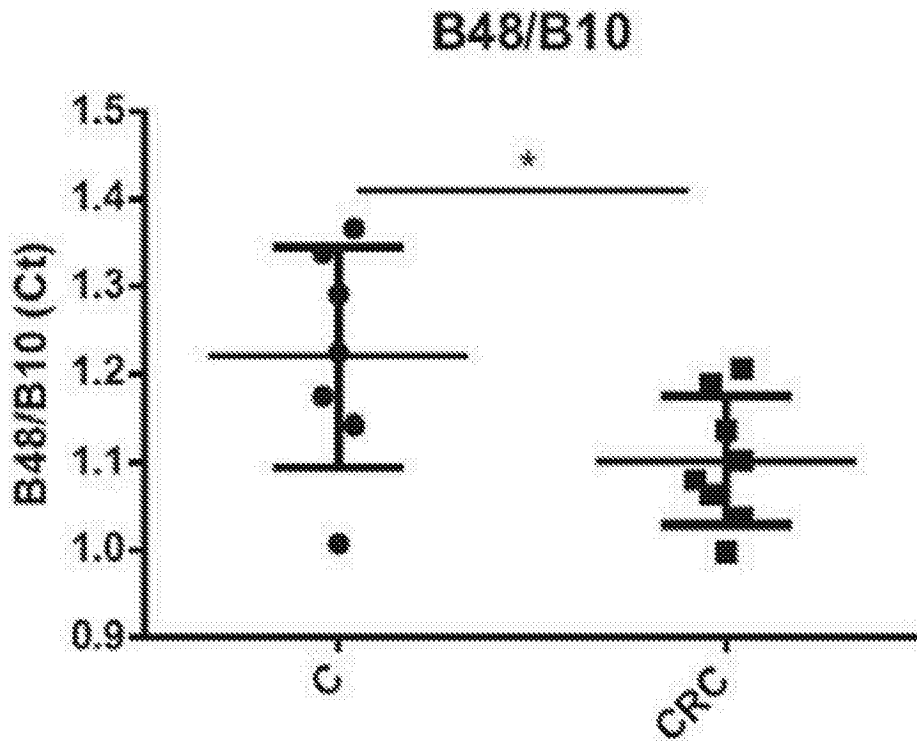


图1

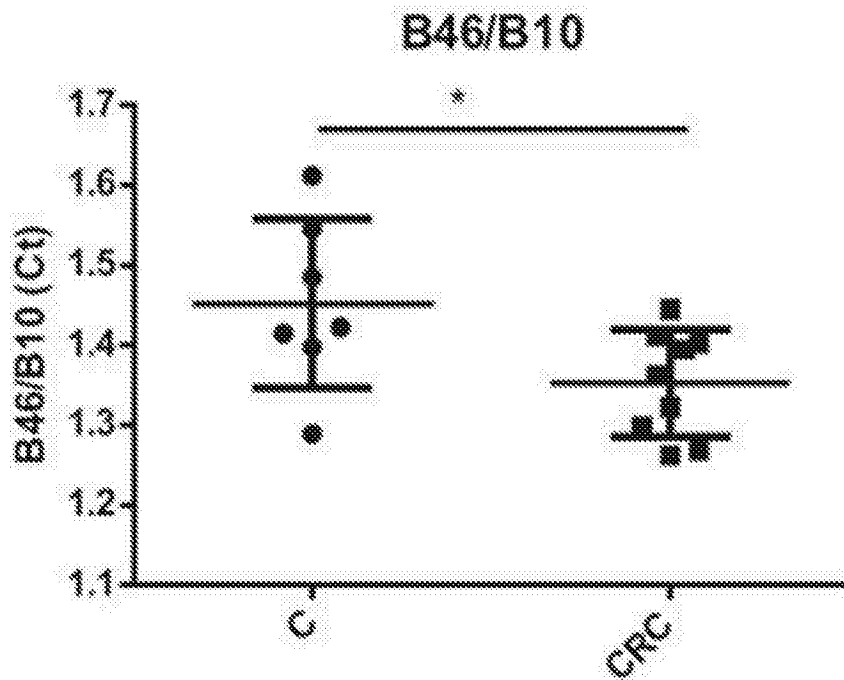


图2

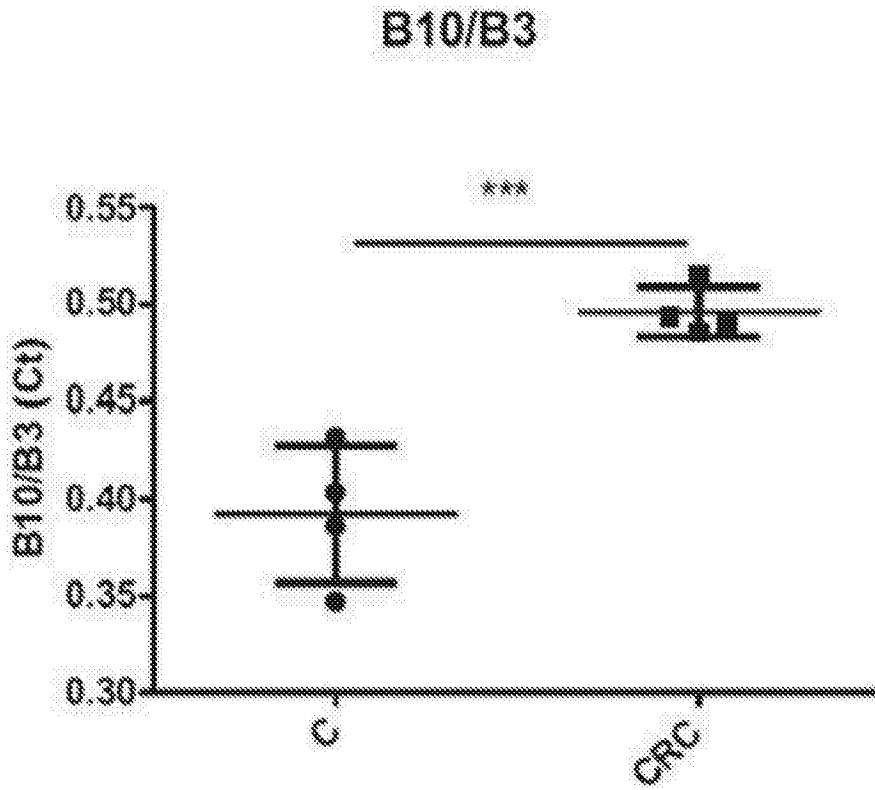


图3

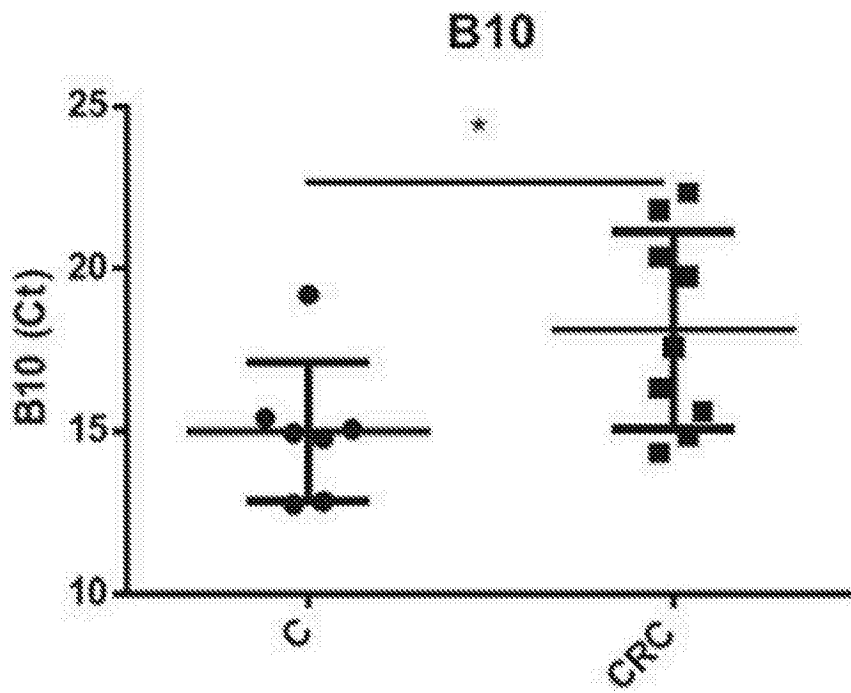


图4

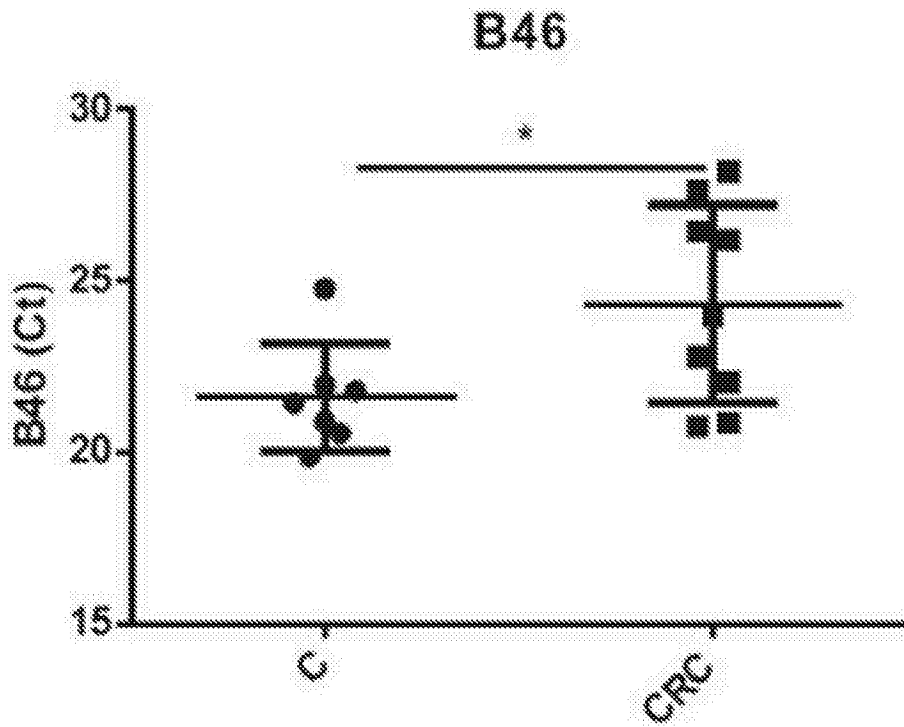


图5

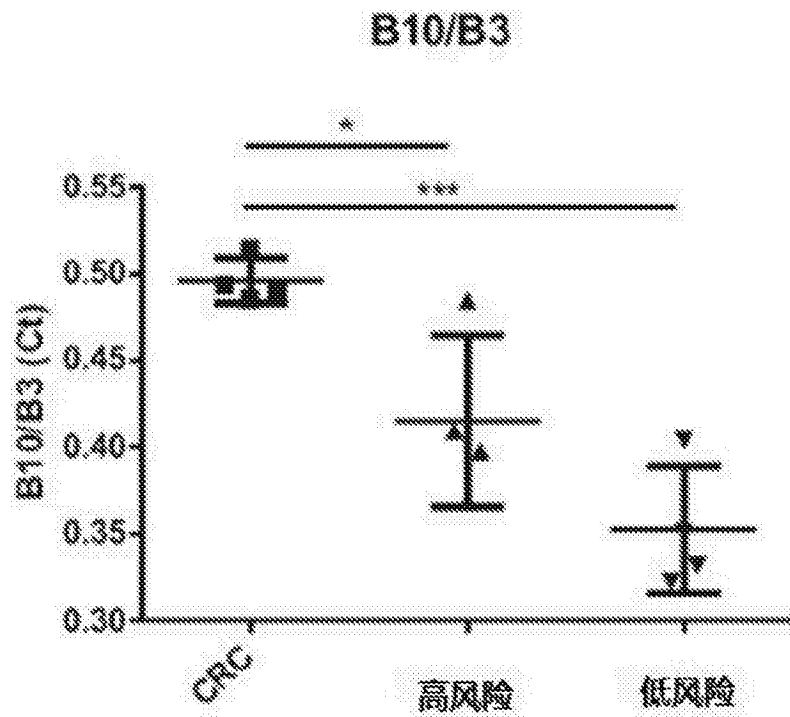


图6

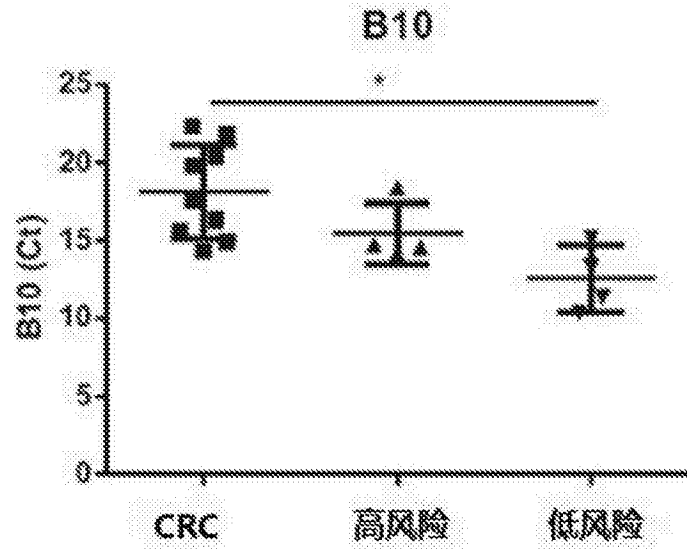


图7

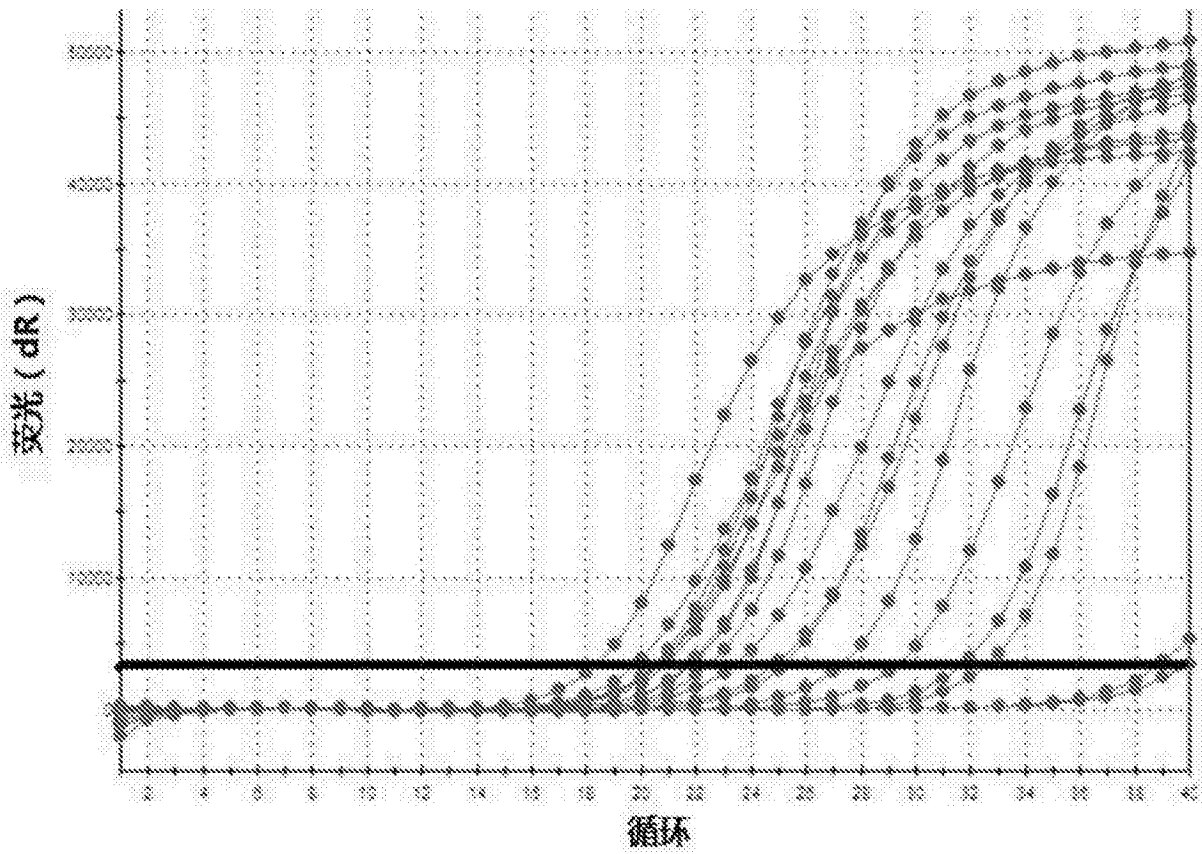


图8a

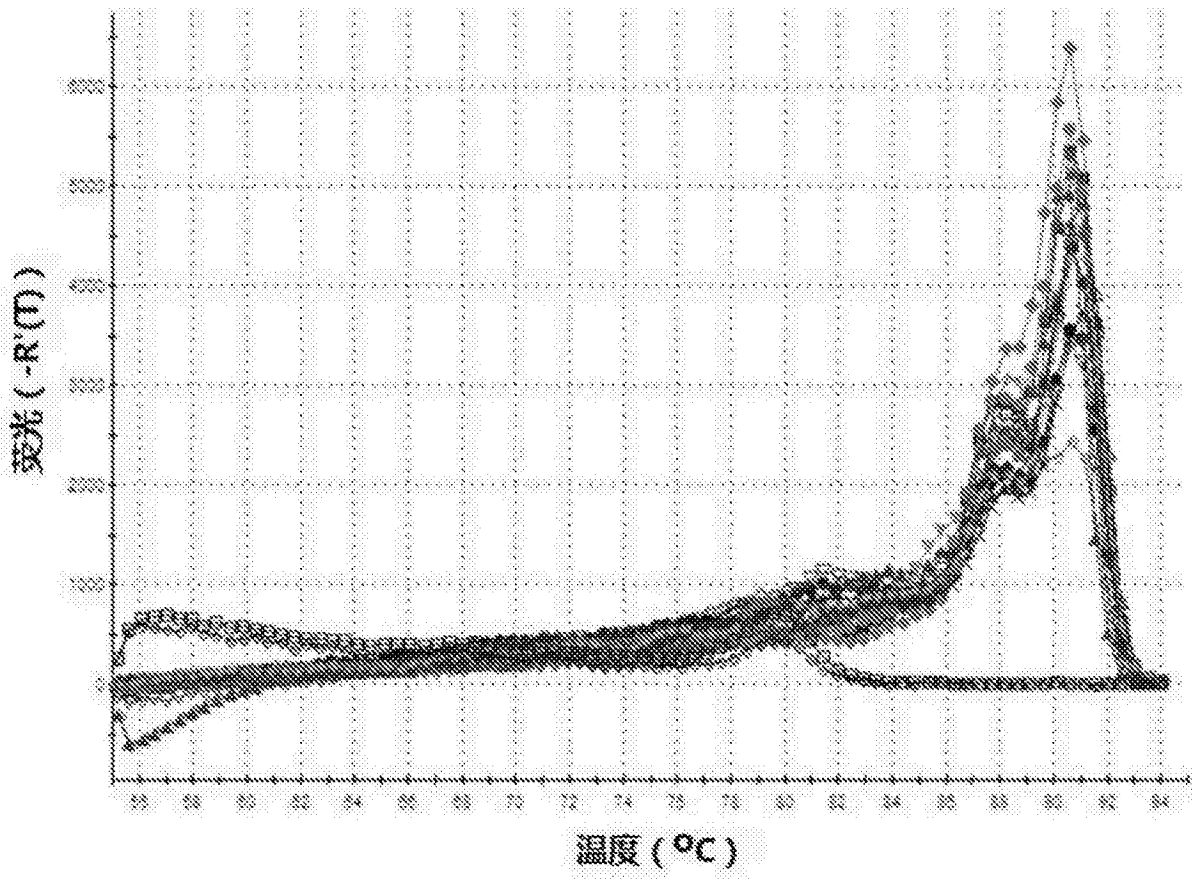


图8b

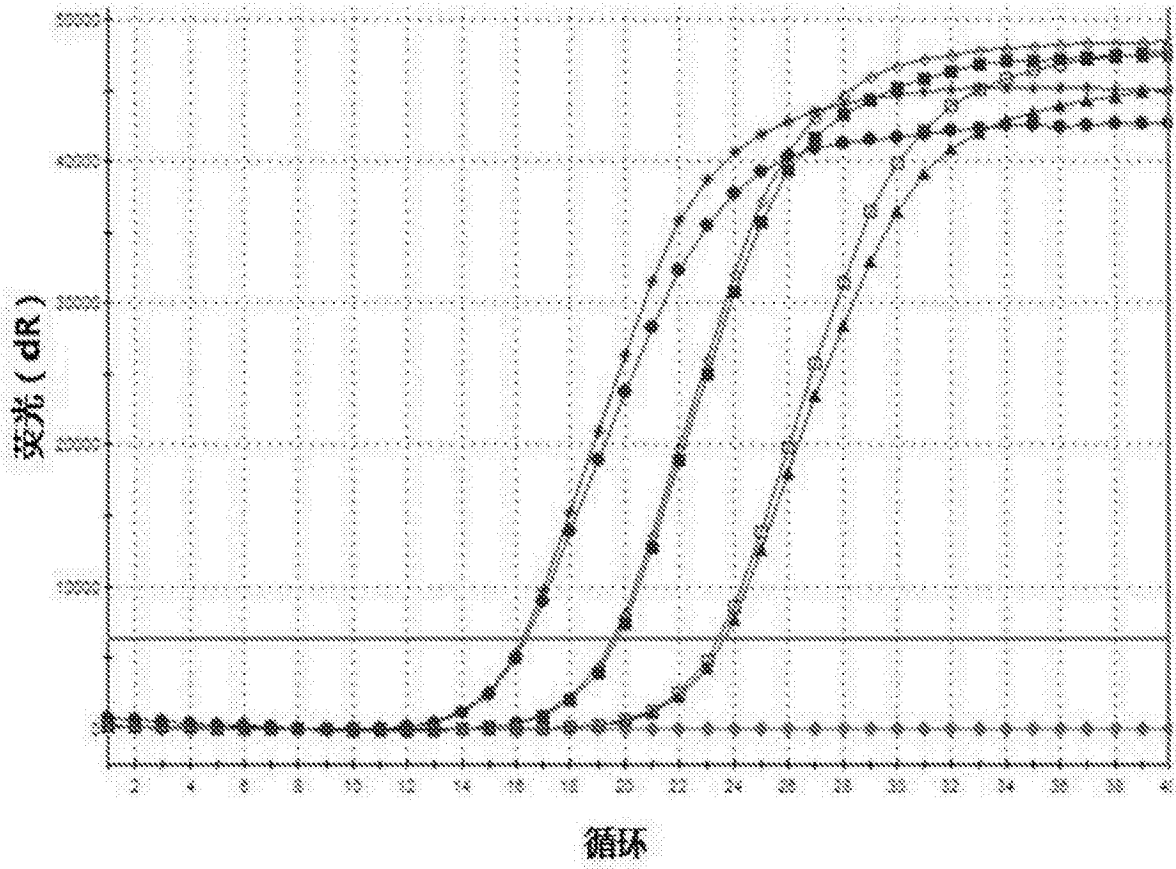


图9a

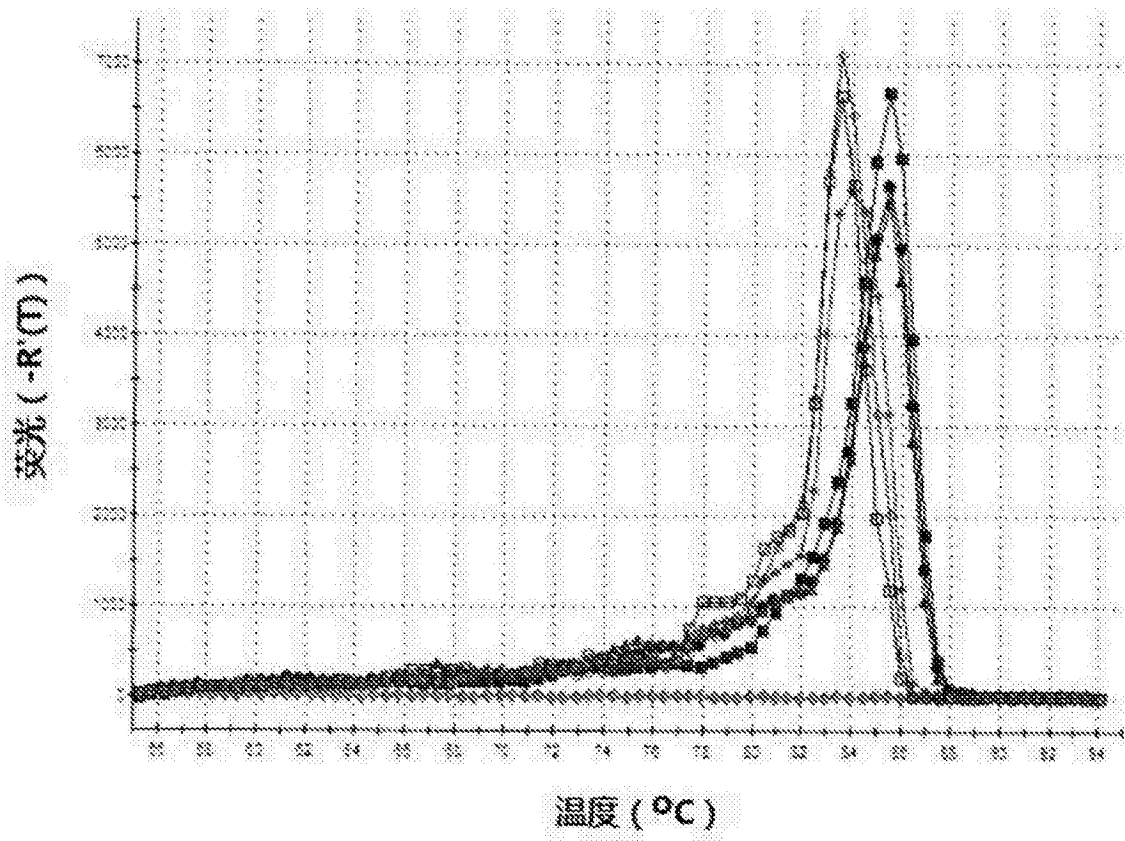


图9b

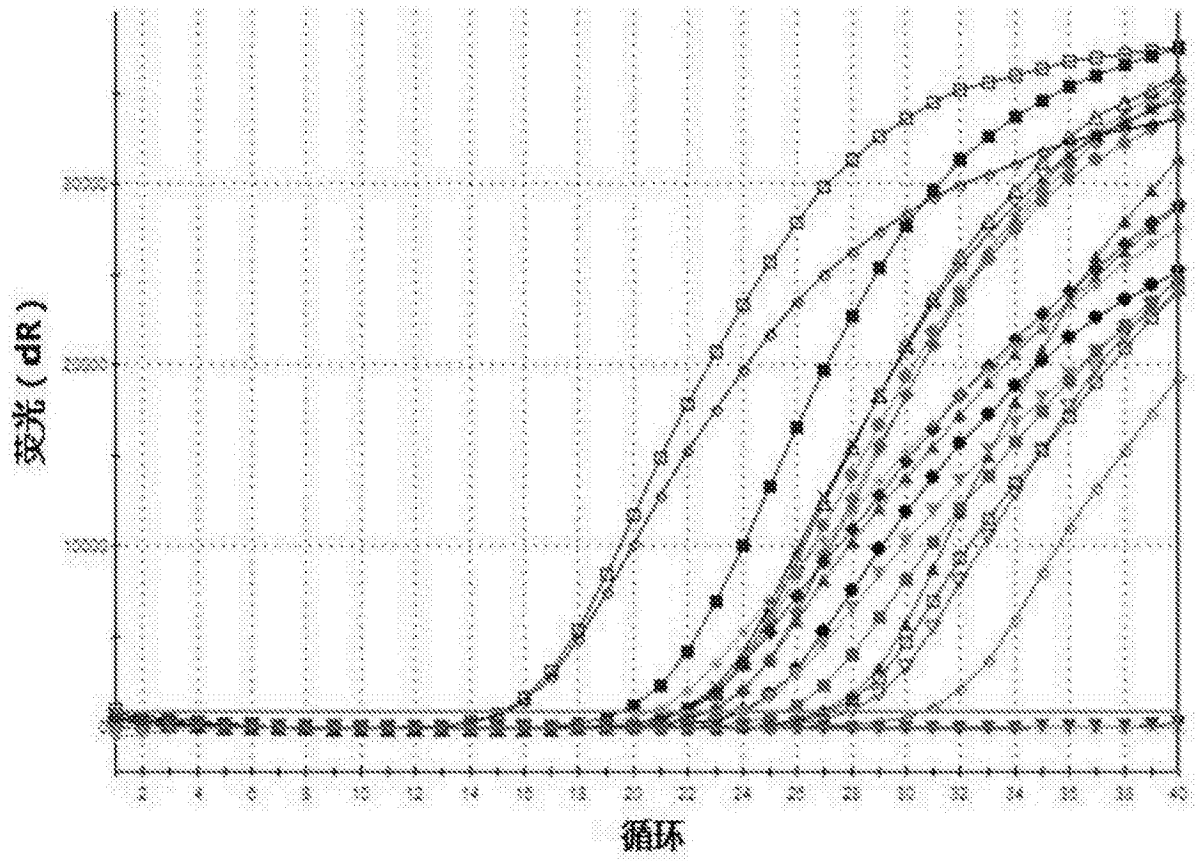


图10a

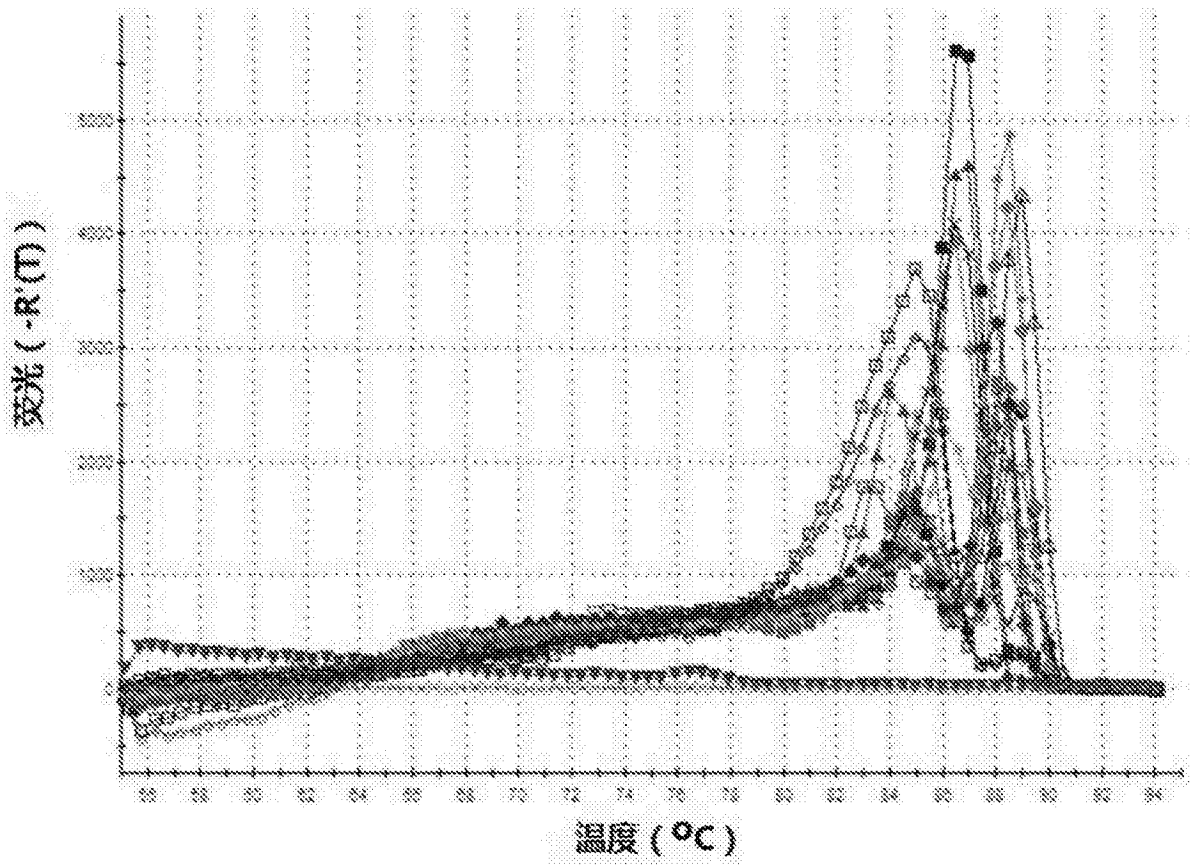


图10b

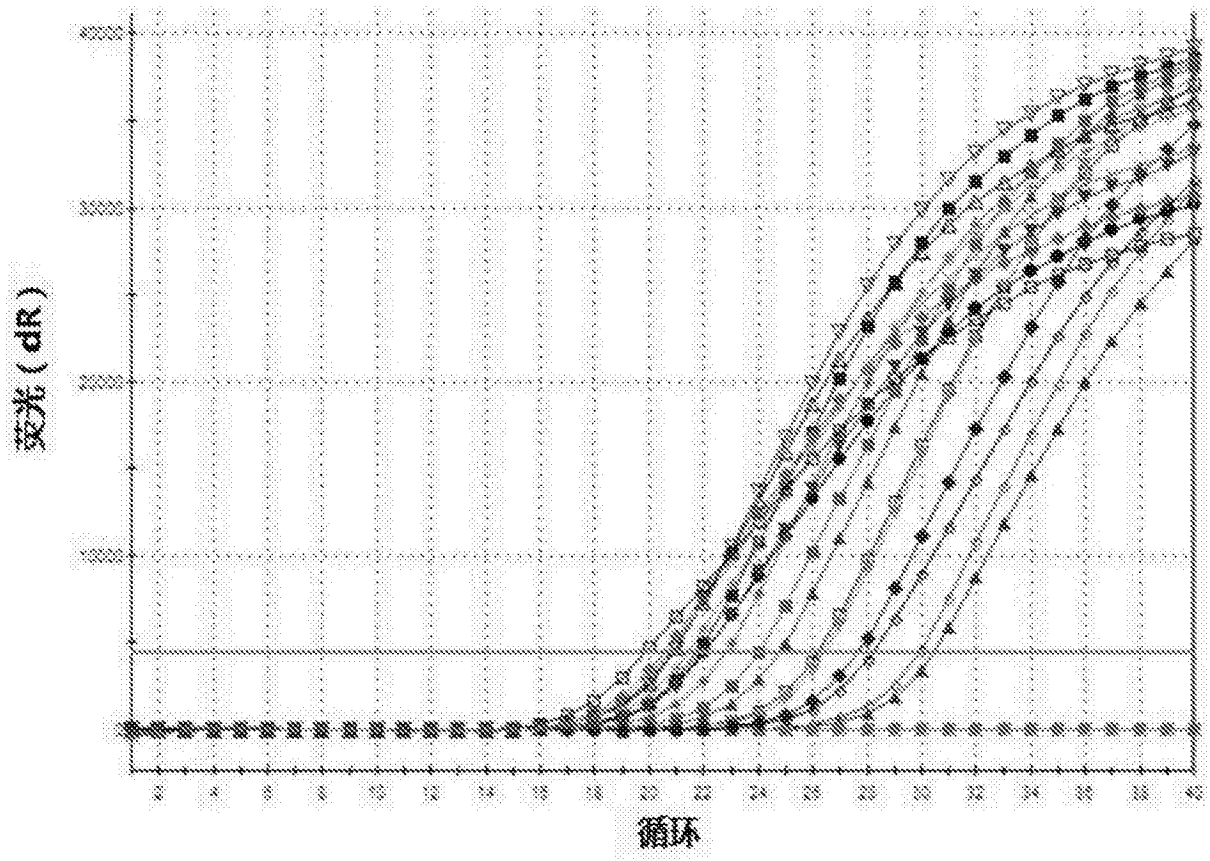


图11a

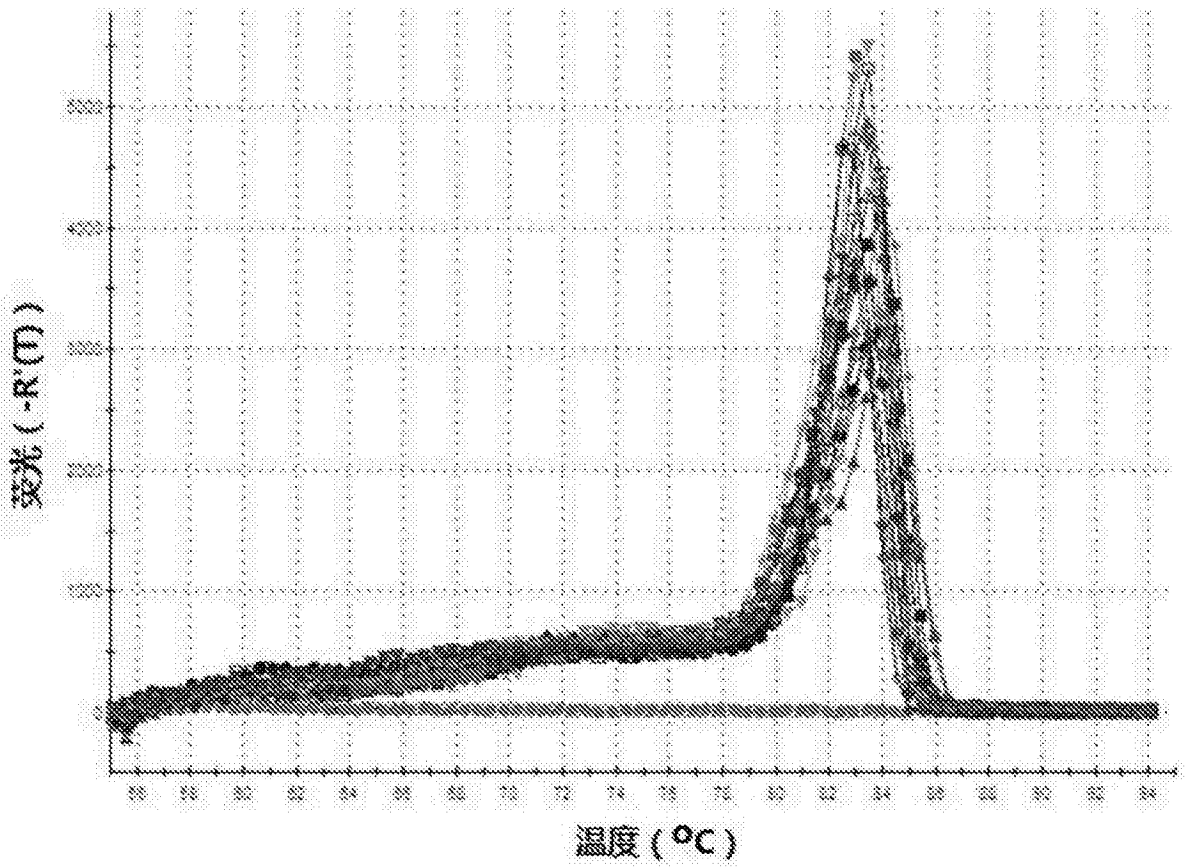


图11b

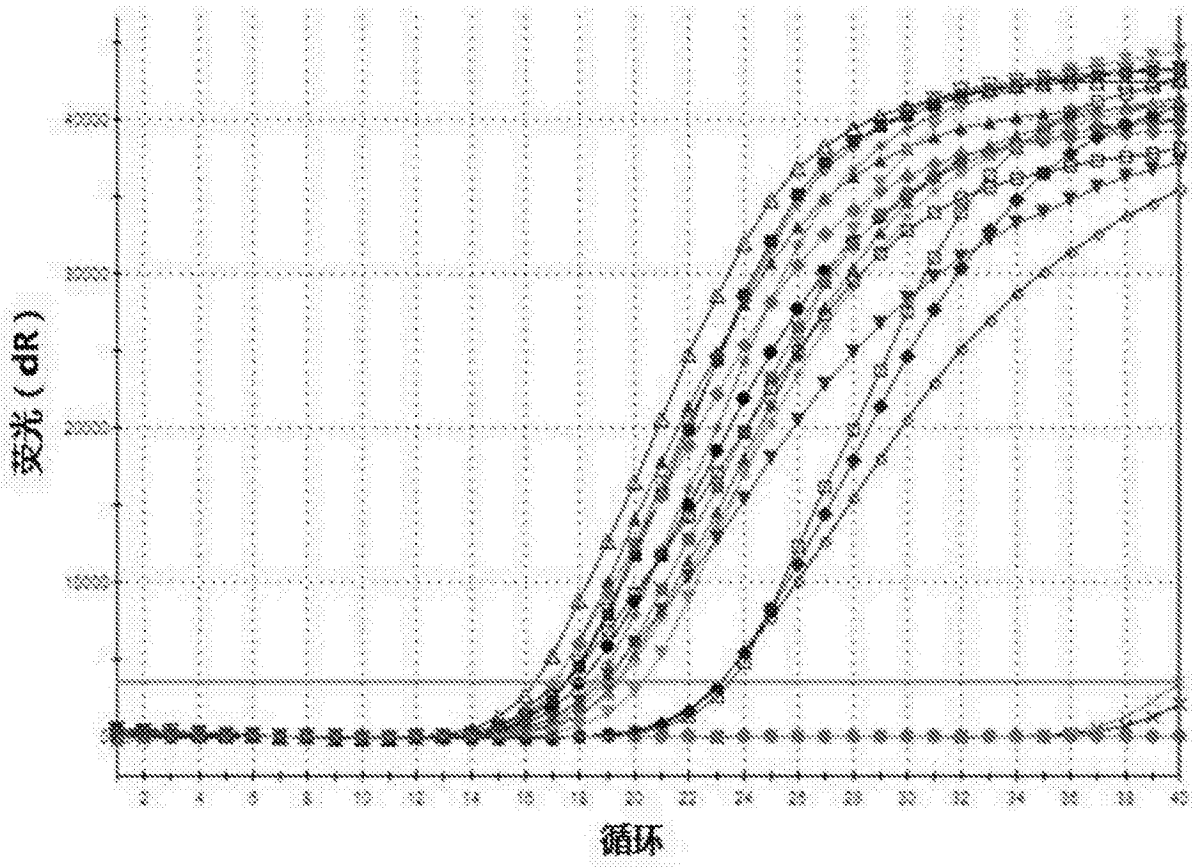


图12a

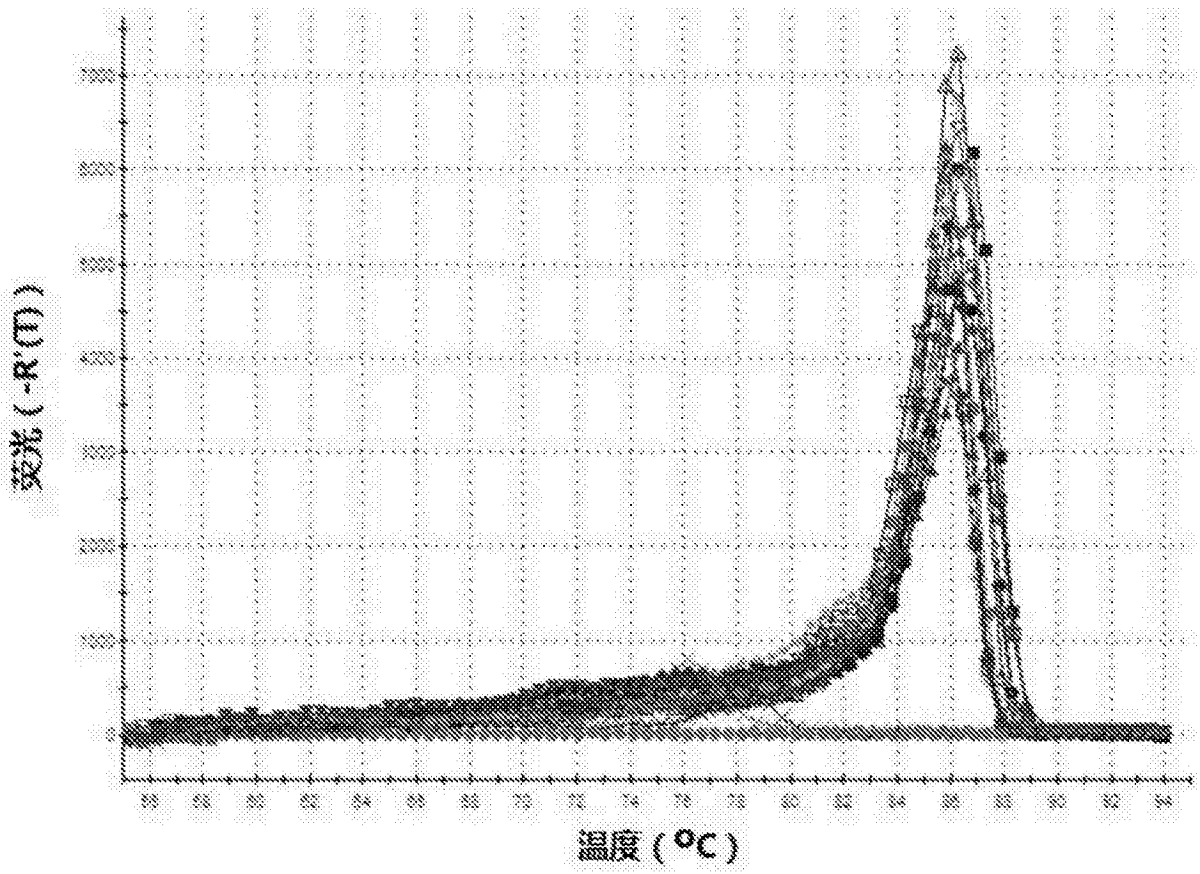


图12b

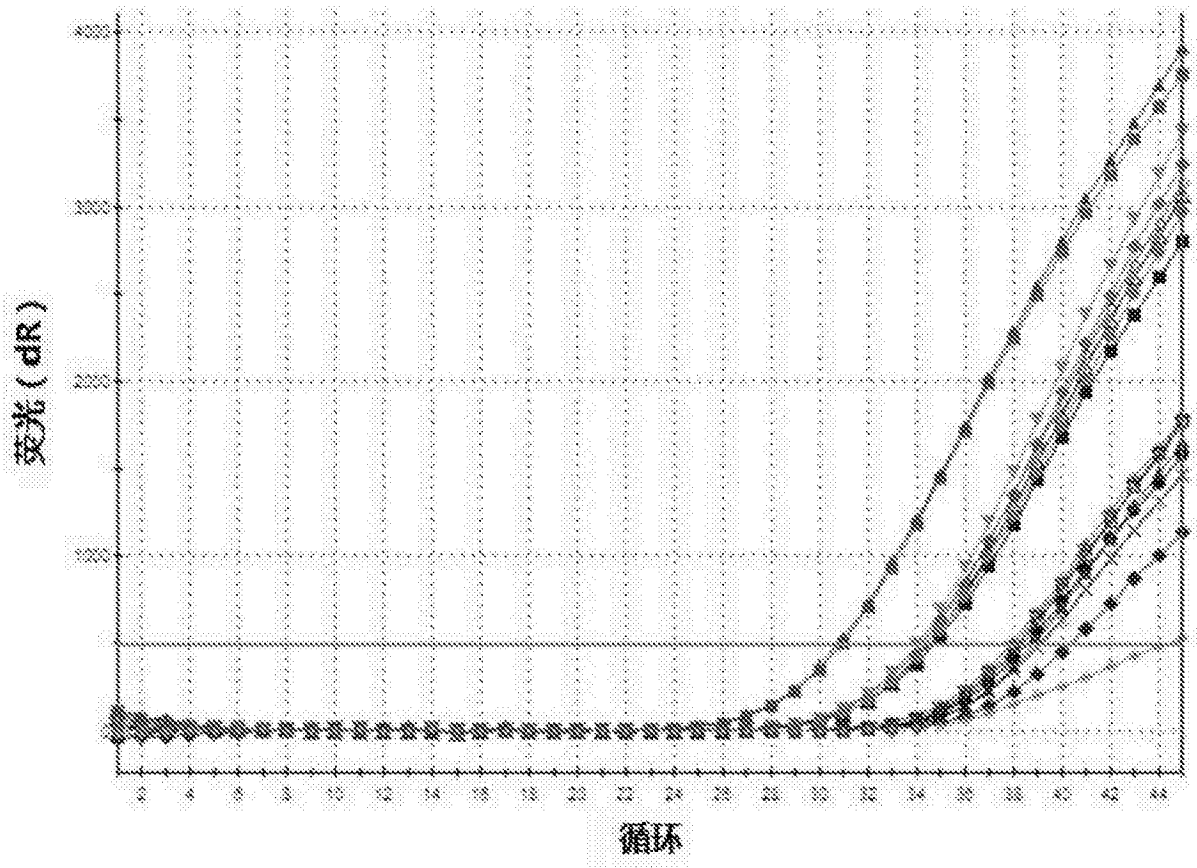


图13a

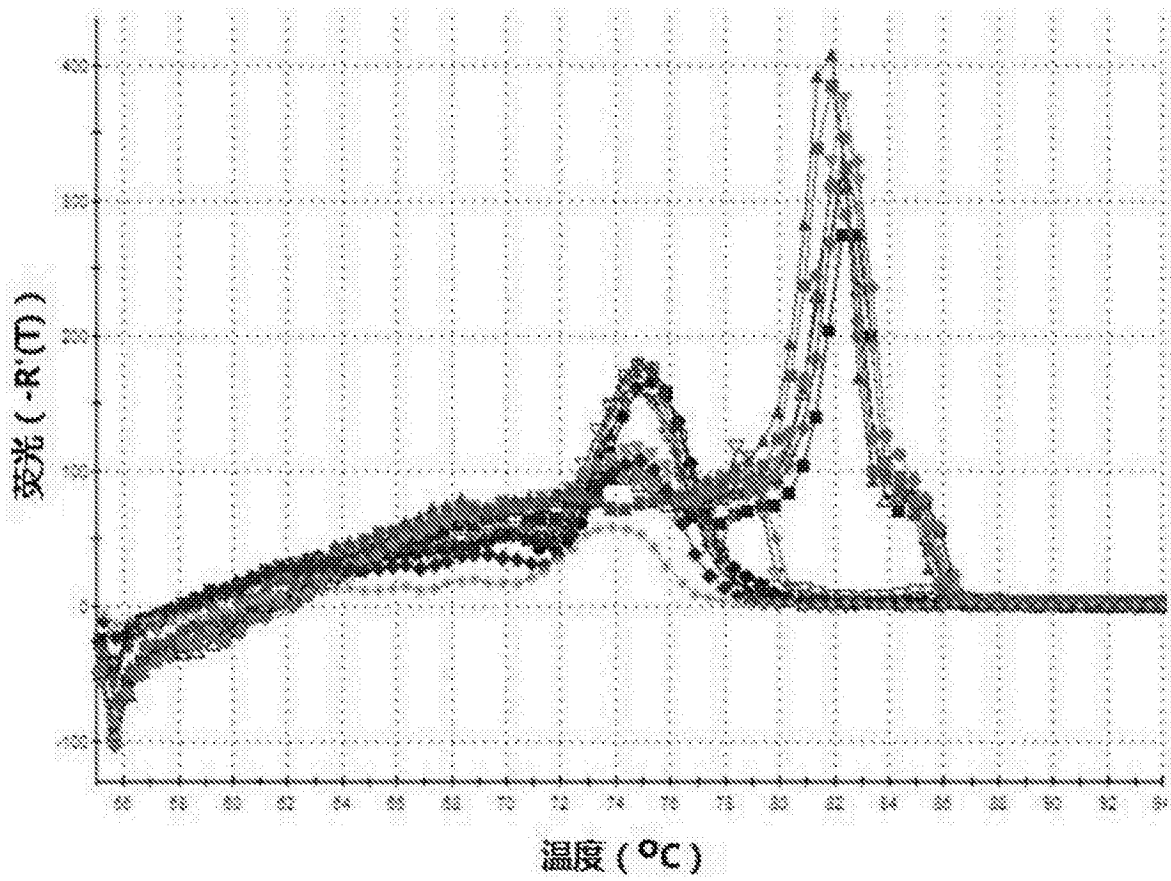


图13b

B3

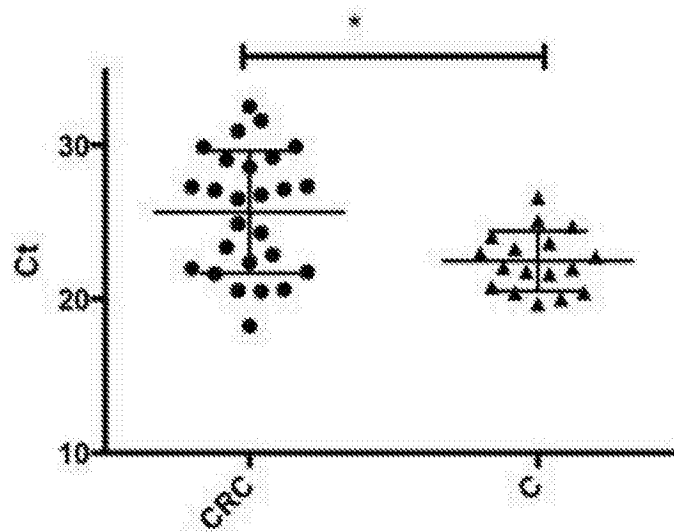


图14 CRC和C组中B3的Ct值

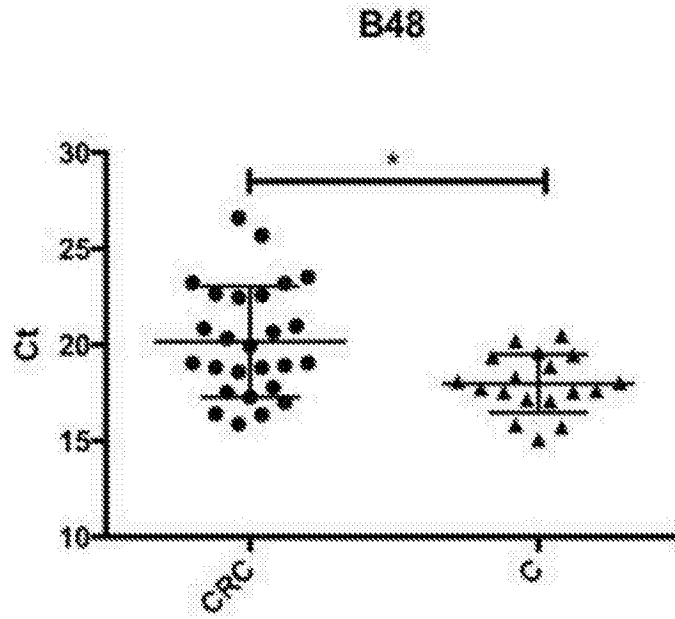


图15 CRC和C组中B48的Ct值

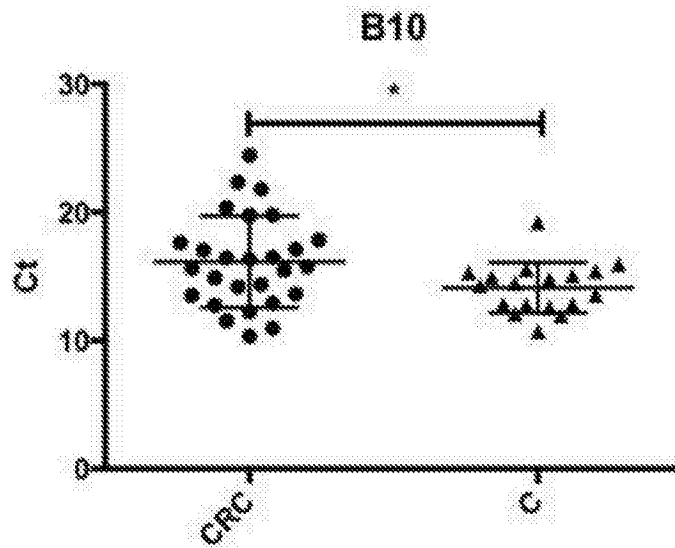


图16 CRC和C组中B10的Ct值

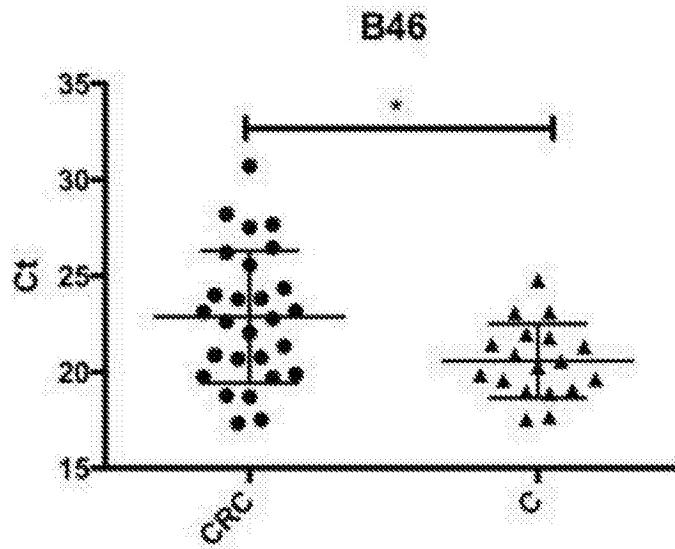


图17 CRC和C组中B46的Ct值

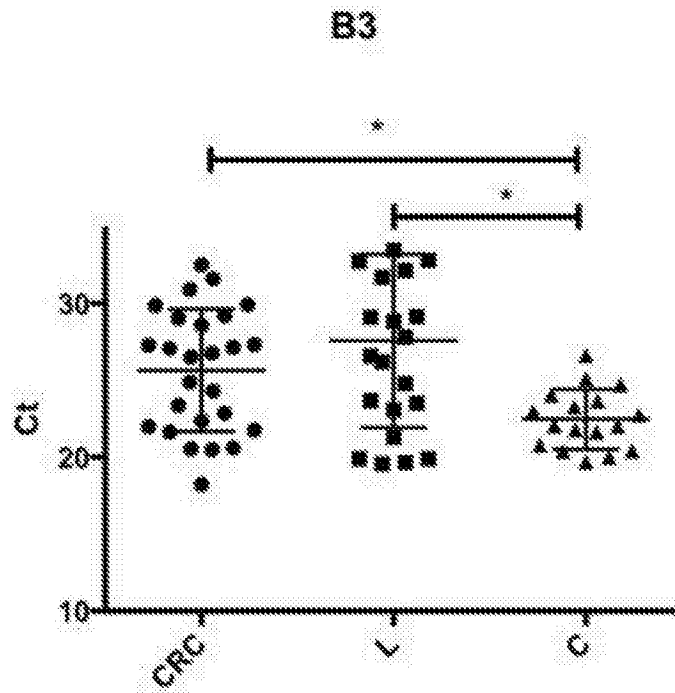


图18 CRC、L和C组中B3的Ct值

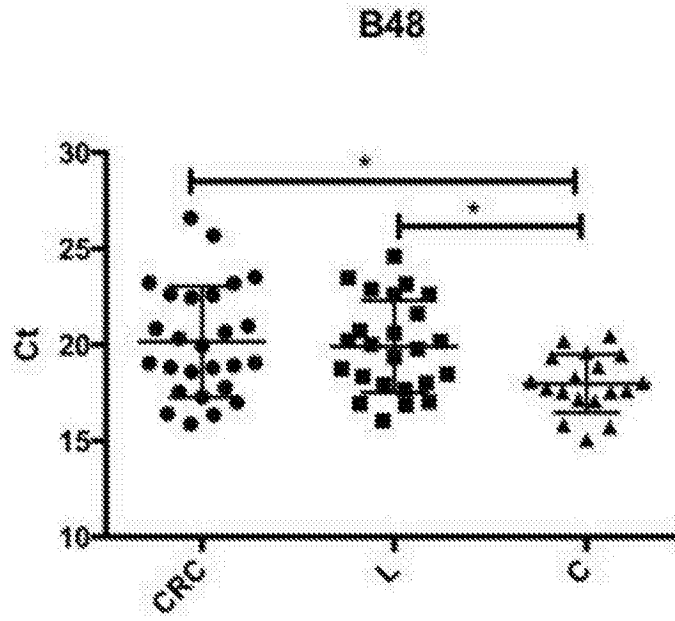


图19 CRC、L和C组中B48的Ct值

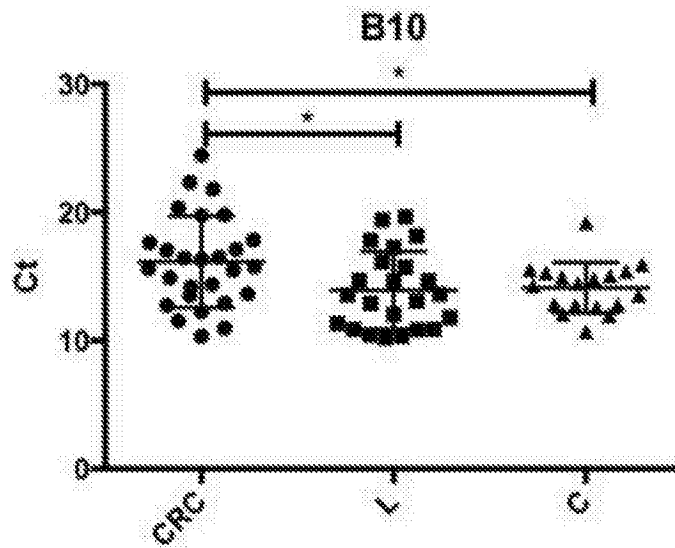


图20 CRC、L和C组中B10的Ct值



图21 CRC、L和C组中B46的Ct值

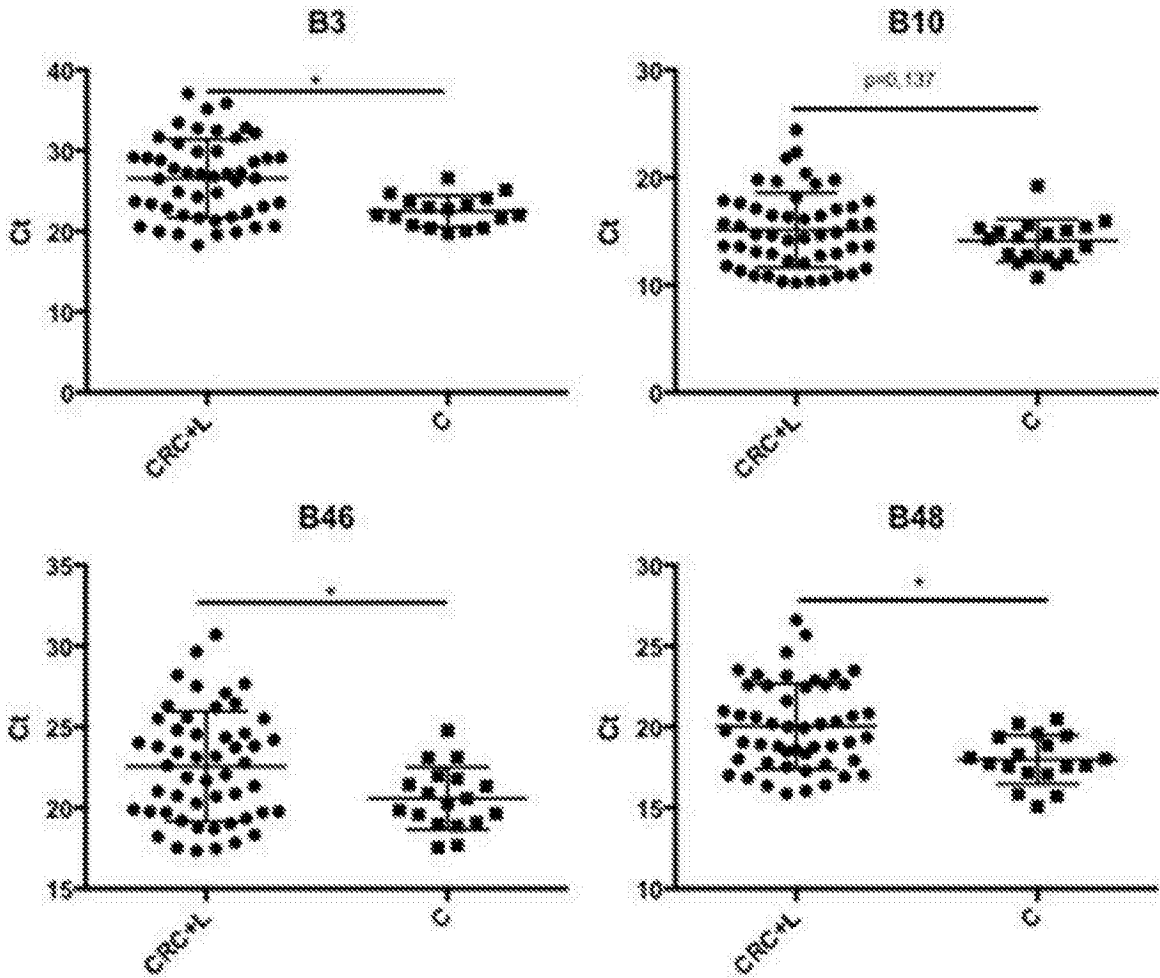


图22 B3,B10,B46和B48在CRC+L vs C组中的Ct值

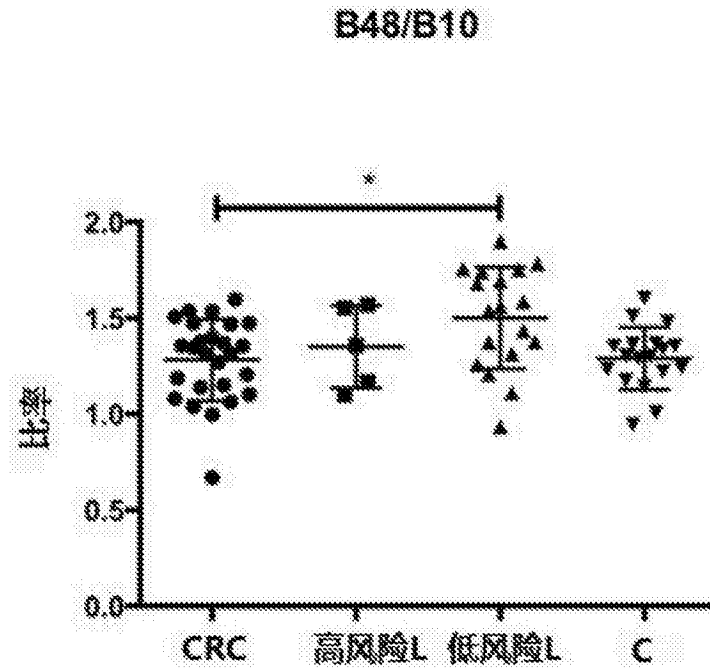


图23 B48/B10比率值在CRC、高风险L、低风险L和C组中的Ct值

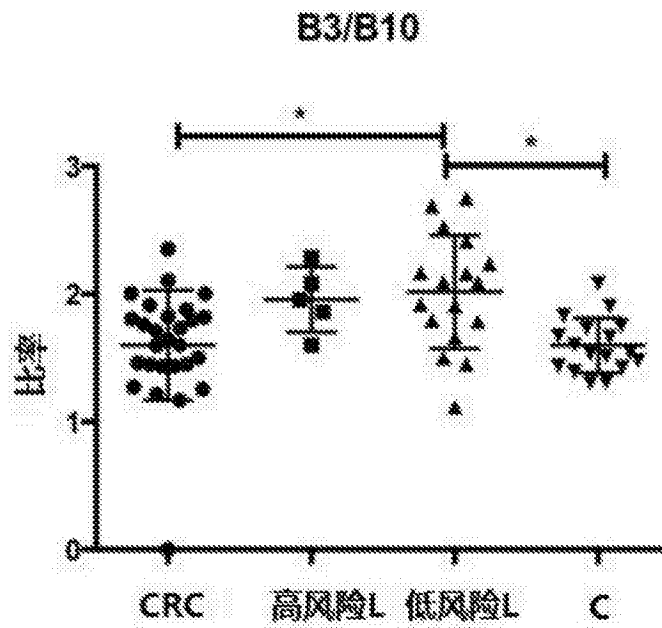


图24 B3/B10比率值在CRC、高风险L、低风险L和C组中的Ct值

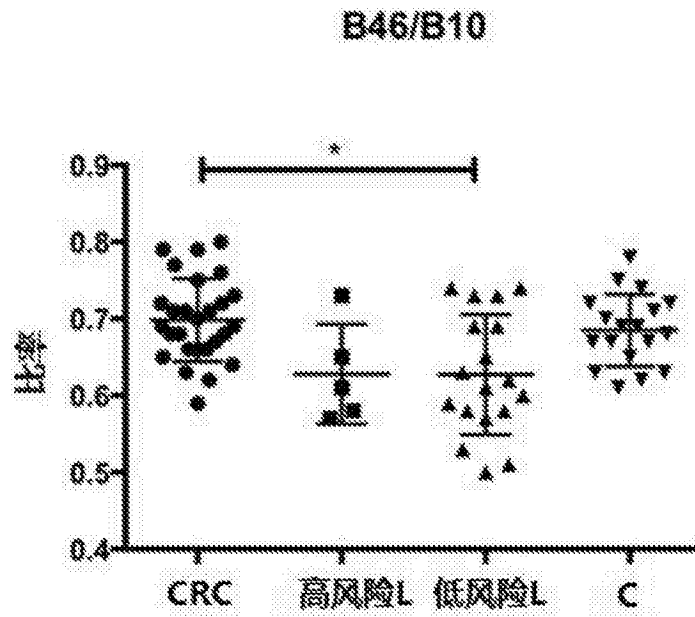


图25 B46/B10比率值在CRC、高风险L、低风险L和C组中的Ct值

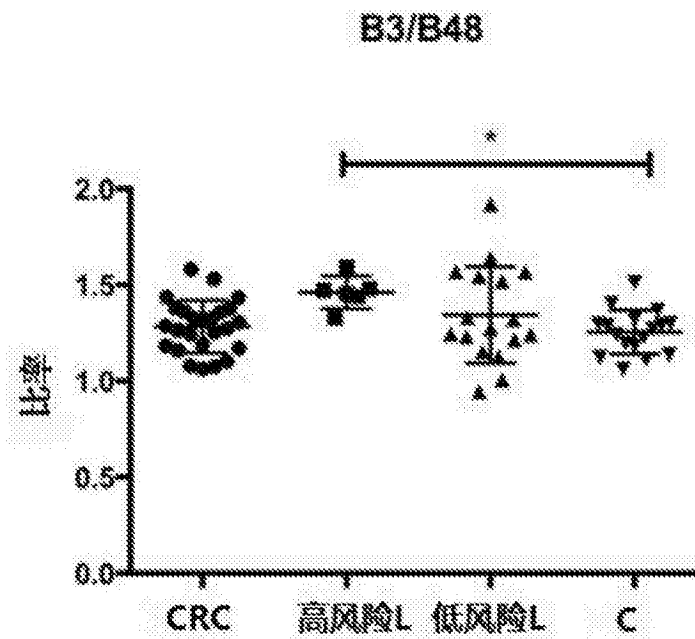


图26 B3/B48比率值在CRC、高风险L、低风险L和C组中的Ct值

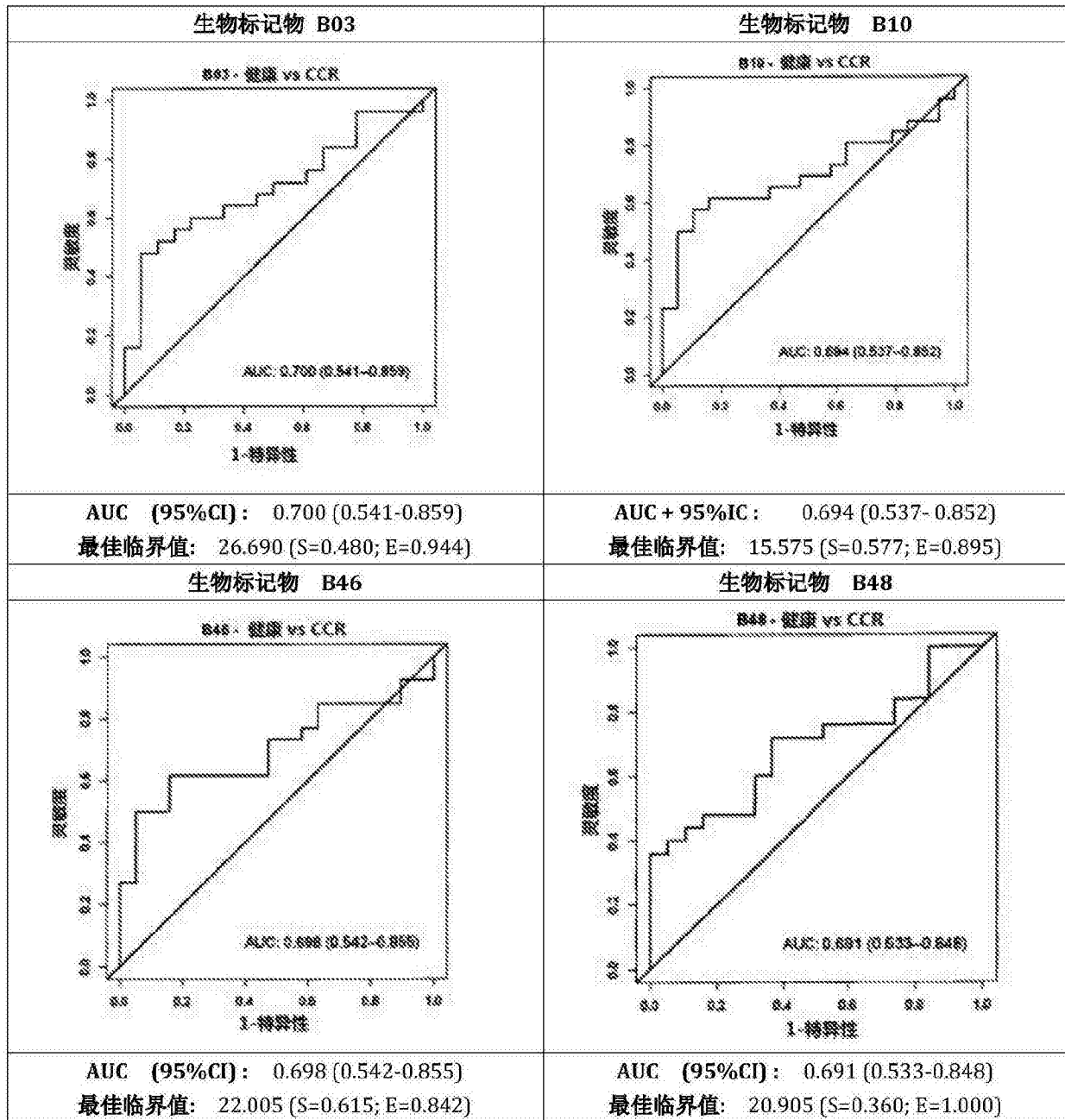


图27

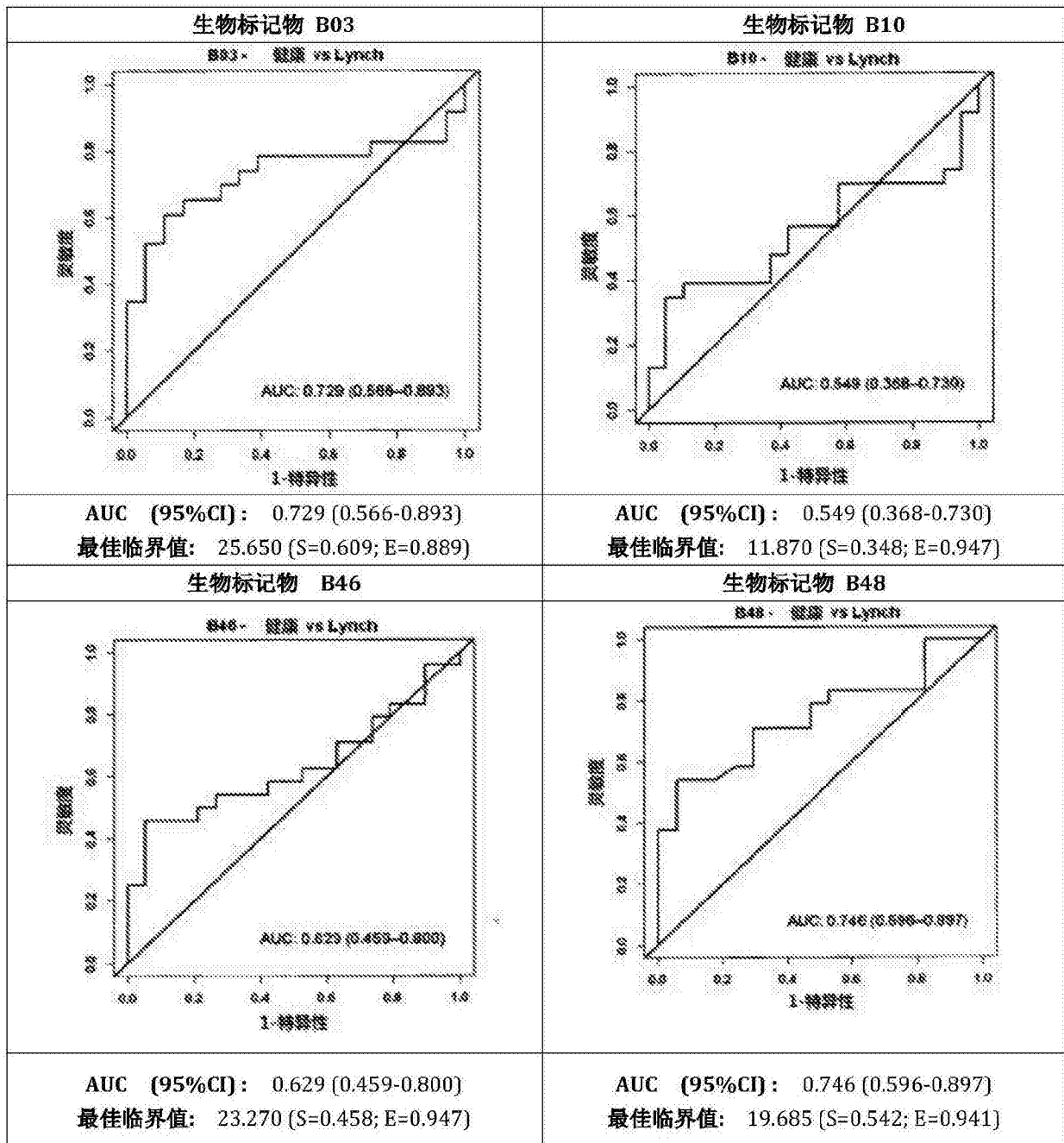


图28

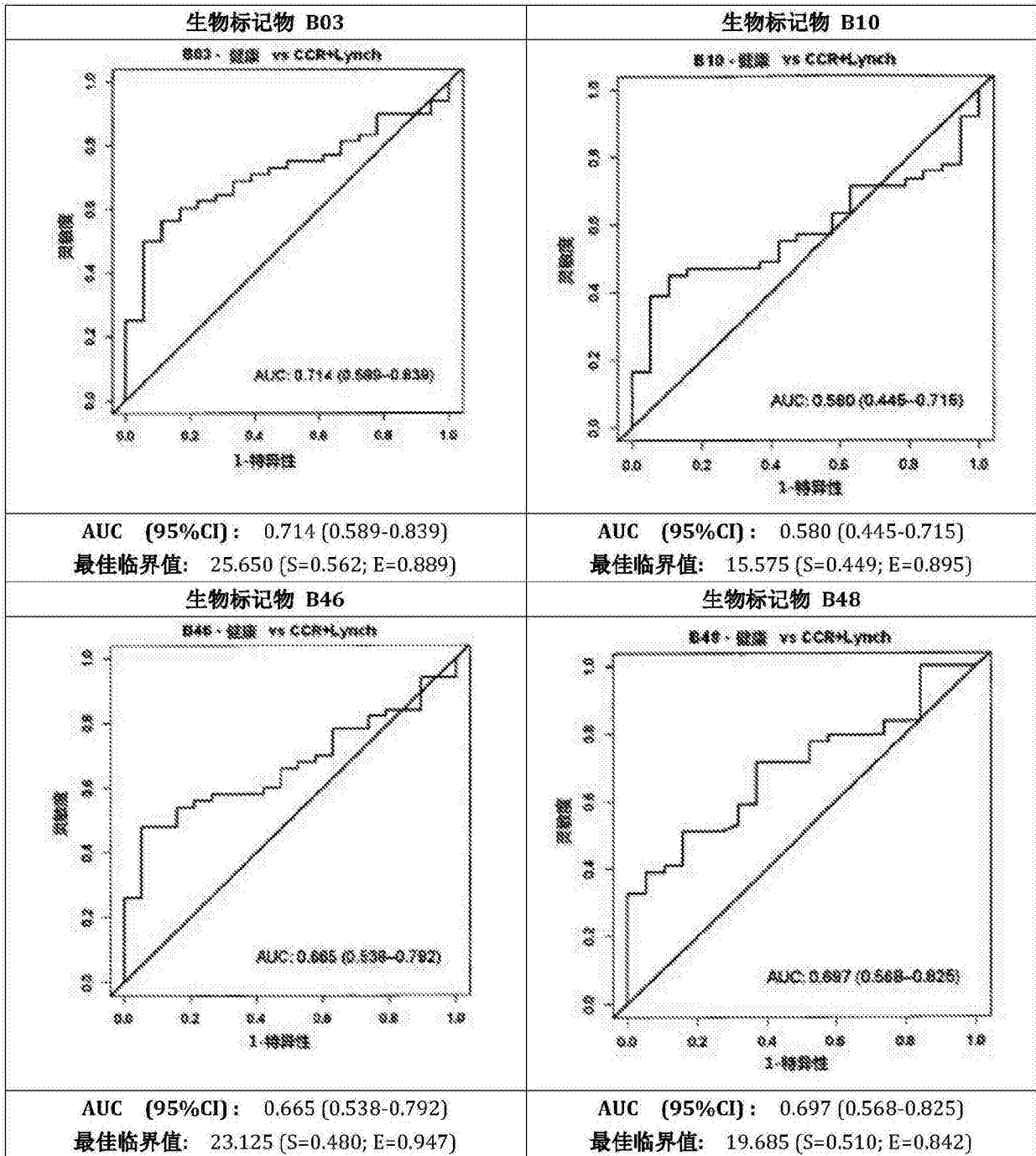


图29

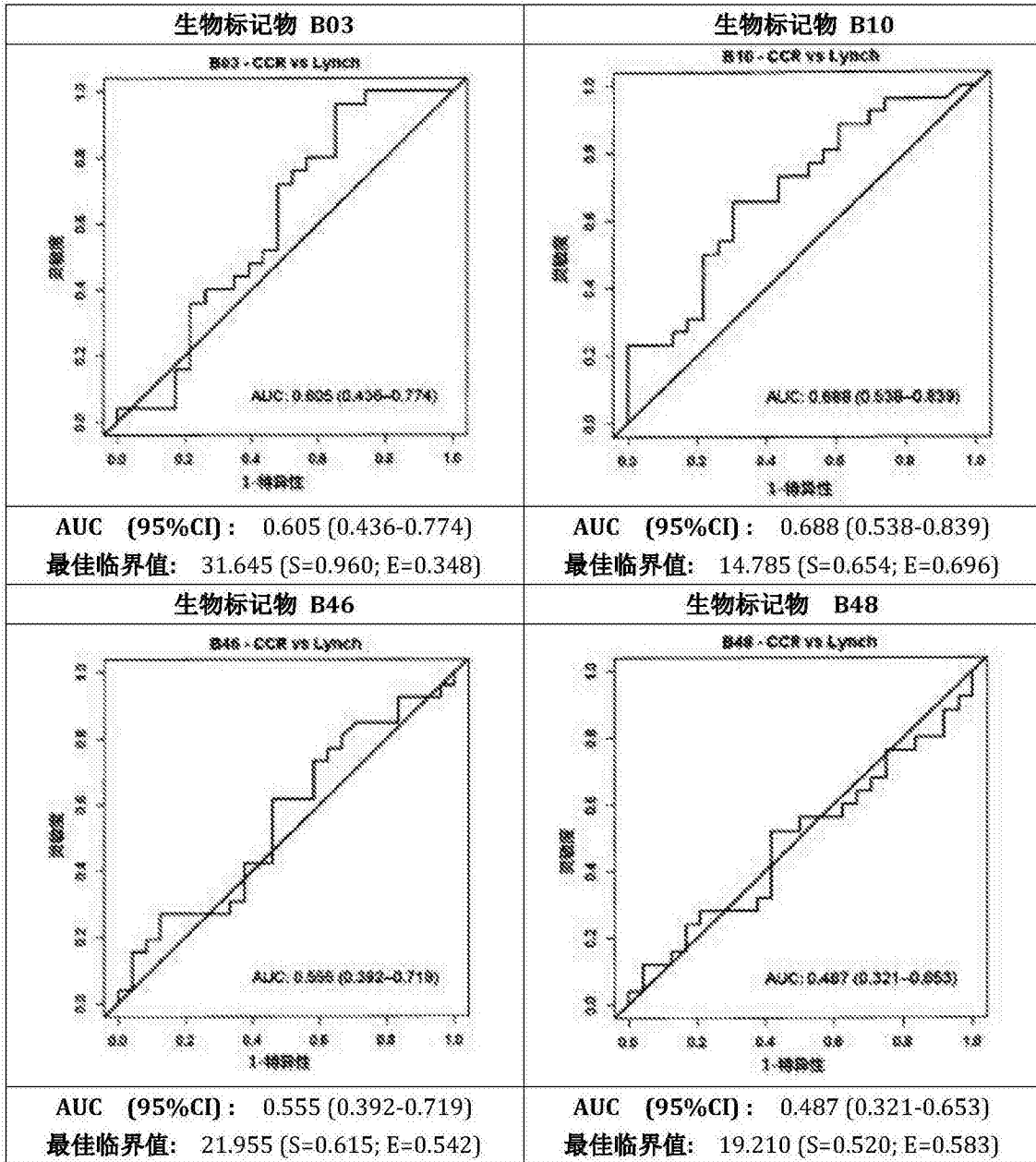


图30