



(12) 发明专利申请

(10) 申请公布号 CN 112673114 A

(43) 申请公布日 2021.04.16

(21) 申请号 201980059667.X

(74) 专利代理机构 北京市柳沈律师事务所  
11105

(22) 申请日 2019.09.11

代理人 张涛

(30) 优先权数据

2018-171850 2018.09.13 JP

(51) Int.Cl.

C12Q 1/6869 (2006.01)

(85) PCT国际申请进入国家阶段日

2021.03.11

A61P 15/00 (2006.01)

A61P 35/00 (2006.01)

(86) PCT国际申请的申请数据

PCT/JP2019/035774 2019.09.11

G01N 33/50 (2006.01)

G01N 33/53 (2006.01)

(87) PCT国际申请的公布数据

W02020/054782 JA 2020.03.19

G01N 33/574 (2006.01)

A61K 45/00 (2006.01)

C12N 15/12 (2006.01)

(71) 申请人 国立研究开发法人国立癌症研究中心

地址 日本东京都

申请人 合同会社予幸集团中央研究所

权利要求书2页 说明书17页  
序列表6页 附图3页

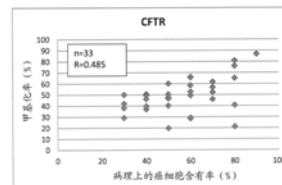
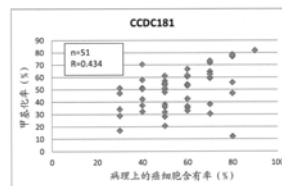
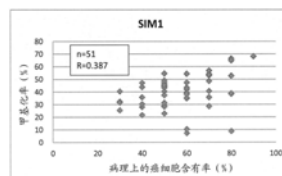
(72) 发明人 石原弘喜 牛岛俊和

(54) 发明名称

乳腺癌细胞存在率的推定方法

(57) 摘要

本发明提供一种使乳腺癌药物治疗效果预测指标的精度提高的方法。具体而言,本发明提供包含以下(1)和(2)的乳腺癌细胞存在率的推定方法等:(1)在采集自人类受试者的样本中,对以下:(a)包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;(b)包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;(c)包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或(d)这些的组合的甲基化率进行分析;以及(2)基于(1)中分析得到的甲基化率,对采集自人类受试者的样本中的乳腺癌细胞存在率进行推定。



1. 一种乳腺癌细胞存在率的推定方法,其包含以下(1)和(2):
  - (1) 在采集自人类受试者的样本中,对以下:
    - (a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
    - (b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
    - (c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或
    - (d) 这些的组合的甲基化率进行分析;以及
  - (2) 基于(1)中分析得到的甲基化率,对采集自人类受试者的样本中的乳腺癌细胞存在率进行推定。
2. 权利要求1记载的方法,其中,

分析是使用选自亚硫酸氢盐、一种以上的引物、一种以上的核酸探针、限制性内切酶、抗甲基化胞嘧啶抗体以及纳米孔中的一种或两种以上的手段进行的。
3. 权利要求1或2记载的方法,其中,

分析是通过亚硫酸氢盐测序法、亚硫酸氢盐·焦磷酸测序法、甲基化特异性PCR法、基因组限制性酶切扫描法(RLGS)、单核苷酸引物扩增(SNuPE)法、CpG岛·微阵列法、甲基化荧光检测法、COBRA方法、质量分析(质量阵列)法、使用甲基化特异性限制酶、高分辨率熔解分析(HRM)法、纳米孔分析法、ICON探针法、甲基化特异性MLPA法或免疫测定法进行的。
4. 一种对乳腺癌的药物治疗效果的预测方法,其包含以下(1)~(4):
  - (1) 对采集自人类受试者的样本中的乳腺癌标志物的值进行测定;
  - (2) 在采集自人类受试者的样本中,对以下:
    - (a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
    - (b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
    - (c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或
    - (d) 这些的组合的甲基化率进行分析;
  - (3) 对在(1)中测定得到的乳腺癌标志物的值通过在(2)中分析得到的甲基化率进行修正,计算出乳腺癌标志物的修正值;以及
  - (4) 基于在(3)中经过修正的乳腺癌标志物的修正值,预测对乳腺癌的药物治疗效果。
5. 一种乳腺癌的判定方法,其包含以下(1)和(2):
  - (1) 在采集自人类受试者的样本中,对以下:
    - (a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
    - (b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

(c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或

(d) 这些的组合

的甲基化率进行分析;以及

(2) 基于在(1)中分析得到的甲基化率,对罹患乳腺癌的可能性进行评估。

6. 一种用于推定乳腺癌细胞存在率的试剂,其包含对以下:

(a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

(b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

(c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或

(d) 这些的组合

的甲基化率的分析手段。

7. 一种用于判定乳腺癌的试剂,其包含对以下:

(a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

(b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

(c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或

(d) 这些的组合

的甲基化率的分析手段。

8. 一种试剂盒,其包含以下(1)和(2):

(1) 对以下:

(a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

(b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

(c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或

(d) 这些的组合

的甲基化率的分析手段;以及

(2) 乳腺癌标志物的测定手段。

## 乳腺癌细胞存在率的推定方法

### 技术领域

[0001] 本发明涉及乳腺癌细胞存在率的推定方法等。

### 背景技术

[0002] 癌组织样本中的癌细胞组分的评估对于排除使用了癌标志物的癌症诊断(举例而言,癌症发病的风险的判定、是否罹患癌症的判定、药物治疗对癌症的效果的预测(如非专利文献1)以及对癌症的药物治疗的预后评估)中由于混入不需要的组分(正常细胞)而引起的判定精度降低是很重要的。

[0003] 作为评估癌细胞组分的方法,例如可举出:(1)病理诊断和(2)使用预定方法从活检采集的癌组织样本(包括癌细胞和非癌细胞)中分离癌细胞的方法等。然而,方法(1)需要熟练的医生来执行,并且方法(2)存在操作复杂等问题。

[0004] 为了解决上述问题,现在被提出的方法有通过用癌细胞组分特异性标志物(“组分标志物”)分析癌组织样本来评价癌细胞组分。作为这样的组分标志物,举例而言,已有食管癌的甲基化标志物TFAP2B、ARHGEF4以及RAPG EFL1(非专利文献2)、以及胃癌的甲基化标志物OSR2、PPFIA3、VAV3(非专利文献3)被报道。

[0005] 现有技术文献

[0006] 非专利文献

[0007] 非专利文献1:Fujii S,Yamashita S,Yamaguchi T,Takahashi M,Hozu mi Y,Ushijima T et al.Pathological complete response of HER2-positive breast cancer to trastuzumab and chemotherapy can be predicted by HSD17B4 methylation.Oncotarget.2017;8(12):19039-48.Doi:10.18632/oncotarget.15118.

[0008] 非专利文献2:Takahashi T et al.,PLoS one,2013;8(12):e82302

[0009] 非专利文献3:Zong L et al.,Gastric cancer,2016;19(2):361-9

### 发明内容

[0010] 本发明所要解决的技术问题

[0011] 本发明的目的是提供一种判定乳腺癌组织中包含的乳腺癌细胞含有率的癌标志物。本发明的目的还在于通过测定乳腺癌细胞含有率来提高预测对乳腺癌的药物治疗效果的指标的精度。

[0012] 解决技术问题的技术手段

[0013] 首先,本发明人等为了评估乳腺癌细胞的组分,想到对在乳腺癌细胞中被高度甲基化,而在正常细胞(非乳腺癌细胞)中完全不被甲基化或只是低度甲基化的乳腺癌细胞特异性CpG位点进行了探索,并且在此探索中,选择了不在细胞之间以高概率发生拷贝数变化,且可以以广泛患者为对象的CpG位点。如此,通过对常染色体和性染色体上存在的超过450000个甲基化候选位点进行全面的锐意研究的结果,作为这样的CpG位点,本发明人等在SIM1基因、CCDC181基因和CFTR基因中发现了特定的CpG位点。本发明人等成功地将这些CpG

位点的甲基化率用作乳腺癌细胞存在率的指标,从而完成了本发明。

[0014] 也就是说,本发明如下。

[0015] [1]一种乳腺癌细胞存在率的推定方法,其包含以下(1)和(2):

[0016] (1)在采集自人类受试者的样本中,对以下:

[0017] (a)包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

[0018] (b)包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

[0019] (c)包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或

[0020] (d)这些的组合

[0021] 的甲基化率进行分析;以及

[0022] (2)基于(1)中分析得到的甲基化率,对采集自人类受试者的样本中的乳腺癌细胞存在率进行推定。

[0023] [2]项[1]的方法,其中,

[0024] 分析是使用选自亚硫酸氢盐、一种以上的引物、一种以上的核酸探针、限制性内切酶、抗甲基化胞嘧啶抗体以及纳米孔中的一种或两种以上的手段进行的。

[0025] [3]项[1]或[2]方法,其中,

[0026] 分析是通过亚硫酸氢盐测序法、亚硫酸氢盐·焦磷酸测序法、甲基化特异性PCR法、基因组限制性酶切扫描法(RLGS)、单核苷酸引物扩增(SNuPE)法、CpG岛·微阵列法、甲基化荧光检测法、COBRA方法、质量分析(质量阵列)法、使用甲基化特异性限制酶、高分辨率熔解分析(HRM)法、纳米孔分析法、ICON探针法、甲基化特异性MLPA法或免疫测定法进行的。

[0027] [4]一种对乳腺癌的药物治疗效果的预测方法,其包含以下(1)~(4):

[0028] (1)对采集自人类受试者的样本中的乳腺癌标志物的值进行测定;

[0029] (2)在采集自人类受试者的样本中,对以下:

[0030] (a)包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

[0031] (b)包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;

[0032] (c)包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或

[0033] (d)这些的组合

[0034] 的甲基化率进行分析;

[0035] (3)对在(1)中测定得到的乳腺癌标志物的值通过在(2)中分析得到的甲基化率进行修正,计算出乳腺癌标志物的修正值;以及

[0036] (4)基于在(3)中经过修正的乳腺癌标志物的修正值,预测对乳腺癌的药物治疗效果。

[0037] [5]一种乳腺癌的判定方法,其包含以下(1)和(2):

[0038] (1)在采集自人类受试者的样本中,对以下:

- [0039] (a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
- [0040] (b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
- [0041] (c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或
- [0042] (d) 这些的组合
- [0043] 的甲基化率进行分析;以及
- [0044] (2) 基于在(1)中分析得到的甲基化率,对罹患乳腺癌的可能性进行评估。
- [0045] [6] 一种用于推定乳腺癌细胞存在率的试剂,其包含对以下:
- [0046] (a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
- [0047] (b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
- [0048] (c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或
- [0049] (d) 这些的组合
- [0050] 的甲基化率的分析手段。
- [0051] [7] 一种用于判定乳腺癌的试剂,其包含对以下:
- [0052] (a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
- [0053] (b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
- [0054] (c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或
- [0055] (d) 这些的组合
- [0056] 的甲基化率的分析手段。
- [0057] [8] 一种试剂盒,其包含以下(1)和(2):
- [0058] (1) 对以下:
- [0059] (a) 包含序列编号1的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
- [0060] (b) 包含序列编号2的核苷酸序列中第1000和第1001个核苷酸的CG部分中的胞嘧啶残基;
- [0061] (c) 包含序列编号3的核苷酸序列中第1001和第1002个核苷酸的CG部分中的胞嘧啶残基;或
- [0062] (d) 这些的组合
- [0063] 的甲基化率的分析手段;以及
- [0064] (2) 乳腺癌标志物的测定手段。
- [0065] 发明效果

[0066] 本发明的推定方法举例而言,可用于提高利用乳腺癌标志物的乳腺癌判定(举例而言,乳腺癌风险的判定、是否罹患乳腺癌的判定、药物治疗效果的预测、预后评估)的精度。

[0067] 本发明的预测方法,举例而言,可以用作判断乳腺癌患者的外科手术必要性和选择药物治疗法时的指标。

[0068] 本发明的判定方法,举例而言,可用于乳腺癌的诊断。

[0069] 本发明的试剂以及试剂盒,举例而言,可用于本发明的方法的简便实施。

## 附图说明

[0070] [图1]图1是显示,通过显微镜检查评估的癌细胞含有率和SIM1、CCDC 181以及CFTR单独甲基化水平相关性的图。

[0071] [图2]图2是显示,通过显微镜检查评估的癌细胞含有率和选自SIM1、C CDC181以及CFTR中的2个基因的组的甲基化水平(选择甲基化水平高的一方)的相关性的图。

[0072] [图3]图3是显示,通过显微镜检查评估的癌细胞含有率和SIM1、CCDC 181以及CFTR的3个基因的组的甲基化水平(选择最高的甲基化水平)的相关性的图。

[0073] [图4]图4是显示,对病理学上的完全缓解(pCR)、非完全缓解(Non-pCR)的HER2阳性型乳腺癌组织样本的HSD17B4基因的甲基化率的测定值,用通过显微镜检查评估的癌细胞含有率以及SIM1和CCDC181的组的甲基化率进行修正而得到的值的图。

[0074] 本发明的具体实施方式

[0075] (一般性说明)

[0076] 本发明中作为对象的癌是乳腺癌。乳腺癌可以分为HER2阳性型、luminal A型、luminal B型以及三阴性型的亚型。本发明中作为对象的癌,可以为任意亚型的乳腺癌,但是在特定的情况下优选为HER2阳性型。

[0077] 在本发明中,“乳腺癌细胞存在率”是指,相对于样本中的所有细胞的个数(也就是乳腺癌细胞的个数以及非乳腺癌细胞的个数的合计)的乳腺癌细胞的个数的比例。也就是说,乳腺癌细胞存在率以下述公式表示。

[0078] 乳腺癌细胞存在率 = (乳腺癌细胞的个数) / [(乳腺癌细胞的个数) + (非乳腺癌细胞的个数)]

[0079] 在本说明书中,“组分标志物”是指,将采集自受试者的样本等的细胞集团分离为癌细胞组分和非癌性细胞(正常细胞)组分时,癌细胞组分特异性的标志物。组分标志物为基因时,也称为“组分标志物基因”。

[0080] 在本发明中主要分析的SIM1基因、CCDC181基因以及CFTR基因(乳腺癌组分标志物基因)的CpG位点的信息如下表1所示(基因组的位置基于人类基因组参考序列hg38)。

[0081] [表1]

[0082] 表1. 乳腺癌组分标志物基因

	编号	基因	染色体	核苷酸编号	探针ID
[0083]	1	<i>SIM1</i>	6	100465064	cg27252696
	2	<i>CCDC181</i>	1	169427630	cg24808280
	3	<i>CFTR</i>	7	117479583	cg09626894

[0084] 本发明中分析的SIM1基因的CpG位点可以定义为包含序列编号1的核苷酸序列中第1000和第1001个核苷酸残基(在SIM1基因的外显子1上游存在的标准核苷酸序列ggccgact中的第4以及第5个核苷酸)的CG部分(5'侧的ggc部分以及3'侧的act部分未变异的情况)。标准核苷酸序列是用于鉴定与上述表1的CpG位点对应的CpG位点而提供参照的序列(本发明中分析的其他基因的CpG位点也是同样)。如果是本领域技术人员,则对于观察到基因组DNA变异的人类受试者,可以在考虑基因组DNA的变异的情况下,适当决定与存在于上述位置的CpG位点相当的CpG位点。以下,在本说明书中简单称为“SIM1基因的CpG位点”时,只要没有特别说明,则表示所述CpG位点。

[0085] 本发明中分析的CCDC181基因的CpG位点可以定义为包含序列编号2的核苷酸序列中第1000和第1001个核苷酸残基(在CCDC181基因的内含子2中存在的标准核苷酸序列cagcggca中的第4以及第5个核苷酸)的CG部分(5'侧的cag部分以及3'侧的gca部分未变异的情况)。如果是本领域技术人员,则对于观察到基因组DNA变异的人类受试者,可以在考虑基因组DNA的变异的情况下,适当决定与存在于上述位置的CpG位点相当的CpG位点。以下,在本说明书中简单称为“CCDC181基因的CpG位点”时,只要没有特别说明,则表示所述CpG位点。

[0086] 本发明中分析的CFTR基因的CpG位点可以定义为包含序列编号3的核苷酸序列中第1001和第1002个核苷酸残基(存在于CFTR基因的外显子1上游的标准核苷酸序列ctccgggg中的第4以及第5个核苷酸)CG部分(5'侧的ctc部分以及3'侧的ggg部分未变异的情况)。如果是本领域技术人员,则对于观察到基因组DNA的变异人类受试者,可以在考虑基因组DNA的变异的情况下,适当决定与存在于上述位置的CpG位点相当的CpG位点。以下,在本说明书中简单称为“CFTR基因的CpG位点”时,只要没有特别说明,则表示所述CpG位点。

[0087] (乳腺癌细胞存在率的推定方法)

[0088] 本发明提供乳腺癌细胞存在率的推定方法。

[0089] 本发明的推定方法包含以下(1)以及(2):

[0090] (1) 在采集自人类受试者的样本中,对以下:

[0091] (a) SIM1基因的CpG位点中的胞嘧啶残基;

[0092] (b) CCDC181基因的CpG位点中的胞嘧啶残基;

[0093] (c) CFTR基因的CpG位点中的胞嘧啶残基;或

[0094] (d) 这些的组合

[0095] 的甲基化率进行分析;以及

[0096] (2) 基于(1)中分析得到的甲基化率,推定采集自人类受试者的样本中的乳腺癌细胞存在率。

[0097] 在步骤(1)中,作为样本,可以使用选自血液、体液、组织以及细胞中的一种或两种以上的样本,但是优选使用对应上述癌的组织样本。在本发明中,在这样的癌组织样本中含



有的基因组DNA中,对特定的CpG位点进行分析。基因组DNA,可以从癌组织样本中通过任意的方法进行适当提取。作为癌组织样本,举例而言,可以使用通过针刺活检等活检方法采集的样本。癌组织样本可以经过预处理。作为这样的预处理,例如可举出:提取、细胞固定、组织固定、组织薄片化、细胞解离处理、加热、冻结、冷藏、液状化。癌组织样本通常是包含癌细胞和正常细胞(非癌性细胞)的细胞集团的样本。

[0098] 本发明中分析的甲基化是胞嘧啶残基的5号位的甲基化。

[0099] 甲基化率的分析可以通过该领域中公知的任意方法进行。举例而言,分析可使用选自亚硫酸氢盐、引物、核酸探针、限制性内切酶、抗甲基化胞嘧啶抗体以及纳米孔中的一种或两种以上的分析手段进行。

[0100] 亚硫酸氢盐(bisulfite)将非甲基化胞嘧啶转化为尿嘧啶,而不将甲基化胞嘧啶转化为尿嘧啶。因此,由于亚硫酸氢盐具有这样的性质,其广泛地在甲基化胞嘧啶的分析中与其他分析手段组合使用。

[0101] 引物是测序用引物或基因扩增用引物(例如PCR引物)或它们的组合。引物可以适当设计以能够分析目标CpG位点。举例而言,引物可设计为退火结合至目标CpG位点的下游区域(对于SIM1基因,为包含序列编号1的核苷酸序列中第1002~2001个核苷酸残基的部分中的任意区域,对于CCDC181基因,为包含序列编号2的核苷酸序列中第1002~2001个核苷酸残基的部分中的任意区域,对于CFTR基因,为包含序列编号3的核苷酸序列中第1003~2001个核苷酸残基的部分中的任意区域),或为了扩增包含目标CpG位点的基因区域,可设计为使目标CpG位点的上游区域(对于SIM1基因,为包含序列编号1的核苷酸序列中第1~999个核苷酸残基的部分中的任意区域,对于CCDC181基因,为包含序列编号2的核苷酸序列中第1~999个核苷酸残基的部分中的任意区域,对于CFTR基因,为包含序列编号1的核苷酸序列中第1~1000个核苷酸残基的部分中的任意区域)和下游区域(上述所示下游区域中的任意区域)中有义链和反义链退火。

[0102] 核酸探针,可以设计为与包含目标CpG位点的区域(对于SIM1基因,为含有包含序列编号1的核苷酸序列中第1000和第1001个核苷酸残基的CG部分的部分,对于CCDC181基因,为含有包含序列编号2的核苷酸序列中第1000和第1001个核苷酸残基的CG部分的部分,对于CFTR基因,为含有包含序列编号3的核苷酸序列中第1001和第1002个核苷酸残基的CG部分的部分)或其上游区域(上述所示的上游区域中的任意区域)或其下游区域(上述所示的下游区域中的任意区域)杂交。核酸探针可以以游离的形态或在固相中固定的形态使用。作为固相,举例而言可以举出:粒子(例如磁性粒子);阵列、膜(例如硝化纤维素膜、滤纸)、柱等支持体;以及板(例如井穴板)、离心管等容器。作为固相的材料,例如可举出:玻璃、塑料以及金属。此外,核酸探针也可以是,在使用抗甲基化胞嘧啶抗体的甲基化胞嘧啶分析中详述的核酸探针。

[0103] 限制性内切酶是甲基化特异性或非特异性限制性内切酶,可以和如上所述的亚硫酸氢盐、引物、核酸探针适当组合使用。

[0104] 具体而言,作为使用如上所述的分析手段的分析方法,例如可举出:DNA甲基化测序法、亚硫酸氢盐测序法、亚硫酸氢盐·焦磷酸测序法、甲基化特异性PCR法、基因组限制性酶切扫描法(RLGS)法、单核苷酸引物扩增(SNuPE)法、CpG岛·微阵列法、甲基化荧光检测法、COBRA法、质量分析(质量阵列)法、甲基化特异性限制性内切酶的使用、高分辨率熔解分

析 (HRM) 法、纳米孔分析分析法、ICON探针法、甲基化特异性MLPA法。这些方法在该领域中是公知的 (例如日本特开2012-090555号公报, 日本特开2014-036672号公报, 日本特表2010-538638号公报, 日本再表2009/136501号公报)。

[0105] 分析也可以使用抗甲基化胞嘧啶抗体进行。使用抗甲基化胞嘧啶抗体进行的甲基化胞嘧啶的分析在该领域中是公知的 (例如W02015/025862; W02015/025863; W02015/025864; W02016/052368; 日本特开2012-230019号公报; DNA Research 13, 37-42 (2006); Anal. Chem. 2012, 84, 7533-7538)。抗甲基化胞嘧啶抗体可以和如上所述的一种或两种以上的分析手段组合使用。具体而言, 作为使用这样的分析手段的方法, 例如可举出: 将抗甲基化胞嘧啶抗体和异种核酸探针 (例如普通RNA探针、修饰RNA探针) 组合使用的方法 (例如W02015/025862); 将抗甲基化胞嘧啶抗体和固相探针以及捕捉探针组合使用的方法 (例如W02015/025863); 将抗甲基化胞嘧啶抗体和吸收剂聚核苷酸以及捕捉探针组合使用的方法 (例如W02015/025864); 将抗甲基化胞嘧啶抗体和修饰核酸碱基配合性异种核酸探针组合使用的方法 (例如W02016/052368)。和抗甲基化胞嘧啶抗体组合使用的核酸探针可以设计为, 当与包含本发明中分析的所述CpG位点的DNA链杂交而形成包含该DNA链和核酸探针的双链结构部分时, CpG位点中的甲基胞嘧啶残基和核酸探针形成不配对部分 (换句话说, 不互补结合) (例如W02015/025862)。因此, 核酸探针, 作为与不配对部分对应的核苷酸残基, 可以具有可与甲基胞嘧啶残基互补结合的鸟嘌呤残基以外的核苷酸残基 (例如胞嘧啶残基、胸腺嘧啶残基、腺嘌呤残基、尿嘧啶残基)。或者, 核酸探针可以设计为, 不形成这样的不配对部分 (例如W02016/052368)。

[0106] 就使用抗甲基化胞嘧啶抗体的方法而言, 可以通过该领域中公知的任意免疫学方法进行。具体而言, 作为这样的方法, 例如可举出: 酶免疫测定法 (EIA) (例如CLEIA、ELISA)、荧光免疫测定法、化学发光免疫测定法、电化学发光免疫测定法、凝集法、免疫染色、流式细胞计数法、生物膜层干涉法、In Situ PLA法、放大化学发光亲和均相检测法、线印记法、免疫印记法。

[0107] 本发明中分析的甲基化率是癌细胞中的所述CpG中胞嘧啶残基的甲基化的比例。甲基化率的测定在该领域中是周知的 (例如日本特开2012-090555号公报, 日本特开2014-036672号公报, 日本特表2010-538638号公报, 日本再表2009/136501号公报)。

[0108] 本发明中分析的甲基化率也可以是SIM1基因、CCDC181基因以及CFTR基因的CpG位点中的胞嘧啶残基的甲基化率的组合。如所述 (d) 所显示的, 甲基化率的组合是包含这些的甲基化率的两种以上 (也就是说, 两种或三种) 的组合, 具体而言, 是SIM1基因和CCDC181基因的组合; SIM1基因和CFTR基因的组合; CCDC181基因和CFTR基因的组合; 以及SIM1基因、CCDC181基因以及CFTR基因的组合中的CpG位点中的胞嘧啶残基的甲基化率。组合中的甲基化率, 举例而言, 可以是组合中的最高甲基化率、最低甲基化率以及平均甲基化率的任意一种, 但是优选采用最高甲基化率。

[0109] 在步骤 (2) 中, 基于分析得到的甲基化率, 可以推定采集自人类受试者的样本中的乳腺癌细胞存在率。

[0110] (对乳腺癌的药物治疗效果的预测方法)

[0111] 本发明还提供对乳腺癌的药物治疗效果的预测方法。

[0112] 作为对乳腺癌的药物治疗, 例如可举出使用抗癌剂的治疗。作为抗癌剂, 例如可举

出：微管抑制剂、抗癌抗生物质、拓扑异构酶抑制剂、铂类药物、烷基化剂。抗癌剂也可以是分子靶向药。作为这样的分子靶向药，例如可举出：HER2抑制剂、EGFR抑制剂（例如吉非替尼、拉帕替尼、厄洛替尼、西妥昔单抗）、c-MET抑制剂（例如PHA-665752、SU11274、XL-880）、ALK抑制剂（例如WHI-P154、TAE684、PF-2341066）、PDGFR抑制剂（例如伊马替尼、达沙替尼、瓦他拉尼）、c-KIT抑制剂（例如舒尼替尼、马赛替尼、莫特沙尼）。

[0113] 对乳腺癌的药物治疗可根据乳腺癌的亚型适当选择。

[0114] 作为对HER2阳性型乳腺癌的药物治疗，例如可举出使用HER2抑制剂的治疗。HER2是和表皮生长因子(EGFR)属于同家族的受体酪氨酸激酶。HER2具有细胞外结构域、跨膜结构域和具有酪氨酸激酶活性的细胞内结构域以单链连接的结构（例如参照Coussens等，Science, 1985, Vol. 230, No. 4730, pp. 1132-1139, 和Genebank登记号：NP\_001005862.1）。作为HER2抑制剂，例如可举出：抗体（例如曲妥珠单抗、帕妥珠单抗）、低分子有机化合物（例如拉帕替尼、那拉替尼、阿法替尼等HER2酪氨酸激酶抑制剂）、药物结合抗体（例如T-DM1）。

[0115] 作为对luminal A型乳腺癌的药物治疗，例如可举出通过激素疗法或化学疗法进行的治疗。作为激素疗法，例如可举出使用了LH-RH激动剂、抗雌激素药物、黄体激素药物、芳香酶抑制剂的疗法。作为化学疗法，例如可举出使用了拓扑异构酶抑制药物、微管作用药物、烷基化药物、代谢拮抗药物的疗法。

[0116] 作为对luminal B型乳腺癌的药物治疗，例如可举出通过激素疗法或化学疗法进行的治疗。作为激素疗法，例如可举出使用了LH-RH激动剂、抗雌激素药物、黄体激素药物、芳香酶抑制剂的疗法。作为化学疗法，例如可举出使用了拓扑异构酶抑制药物、微管作用药物、烷基化药物、代谢拮抗药物的疗法。

[0117] 作为对三阴性型乳腺癌的药物治疗，例如可举出通过化学疗法进行的治疗。作为化学疗法，例如可举出使用了拓扑异构酶抑制药物、微管作用药物、烷基化药物、代谢拮抗药物的疗法。

[0118] 在对癌的药物疗法中，可以并用多种抗癌剂。作为并用的多种抗癌剂，例如可举出上述两种以上的抗癌剂的组合以及上述一种以上的抗癌剂和其他抗癌剂的组合。作为其他抗癌剂，例如可举出：紫杉烷类抗癌剂（例如紫杉醇）等微管抑制剂、葱环类（例如阿霉素）、芳香酶抑制剂（例如来曲唑）。

[0119] 本发明的预测方法包含以下(1)～(4)：

[0120] (1) 对采集自人类受试者的样本中的乳腺癌标志物的值进行测定；

[0121] (2) 在采集自人类受试者的样本中，对以下：

[0122] (a) SIM1基因的CpG位点中的胞嘧啶残基；

[0123] (b) CCDC181基因的CpG位点中的胞嘧啶残基；

[0124] (c) CFTR基因的CpG位点中的胞嘧啶残基；或

[0125] (d) 这些的组合

[0126] 的甲基化率进行分析；

[0127] (3) 对(1)中测定得到的乳腺癌标志物的值，用(2)中分析得到的甲基化率进行修正，计算乳腺癌标志物的修正值；以及

[0128] (4) 基于(3)中经过修正的乳腺癌标志物的修正值，预测对乳腺癌的药物治疗效果。

[0129] 步骤(1)中的乳腺癌标志物是作为乳腺癌指标的标志物,例如可举出DNA甲基化率(例如HSD17B4的cg15896301:非专利文献3)。

[0130] 步骤(2)中的组分标志物基因(SIM1基因、CCDC181基因或CFTR基因)的CpG位点以及甲基化率的分析手段如上所述。

[0131] 步骤(3)中的乳腺癌标志物的值的修正是指,基于在采集自人类受试者的样本(包含乳腺癌细胞以及非乳腺癌细胞的样本)中测定得到的乳腺癌标志物的值,计算出更纯粹地与乳腺癌细胞对应的乳腺癌标志物的值。因此,乳腺癌标志物的修正值,举例而言,可以通过下述公式求出。

[0132] 乳腺癌标志物的修正值 = (采集自人类受试者的样本中测定得到的乳腺癌标志物的值) / (采集自人类受试者的样本中的乳腺癌细胞存在率)

[0133] 就步骤(4)而言,可在基于乳腺癌标志物的值对乳腺癌的药物治疗效果进行预测手法中,通过使用乳腺癌标志物的修正值代替乳腺癌标志物的值来进行。通过使用乳腺癌标志物的修正值,期待对药物治疗效果做出高精度预测。

[0134] (乳腺癌的判定方法)

[0135] 本发明还提供乳腺癌的判定方法。

[0136] 乳腺癌的判定是指,对人类受试者罹患乳腺癌的可能性高低进行评估。

[0137] 本发明的判定方法包含以下(1)以及(2):

[0138] (1) 在采集自人类受试者的样本中,对以下:

[0139] (a) SIM1基因的CpG位点中的胞嘧啶残基;

[0140] (b) CCDC181基因的CpG位点中的胞嘧啶残基;

[0141] (c) CFTR基因的CpG位点中的胞嘧啶残基;或

[0142] (d) 这些的组合

[0143] 的甲基化率进行分析;以及

[0144] (2) 基于(1)中分析得到的甲基化率,对罹患乳腺癌的可能性进行评估。

[0145] 步骤(1)中的采集自人类受试者的样本、组分标志物基因(SIM1基因、CCDC181基因或CFTR基因)的CpG位点以及甲基化率的分析手段如上所述。

[0146] 在步骤(2)中,甲基化率在基准值以上时,人类受试者有罹患乳腺癌的可能性,和/或甲基化率比基准值低时,人类受试者罹患乳腺癌可能性低,可以以此进行评估。作为这样的基准值,可以使用适当设定为能够鉴别乳腺癌患者和正常人的界限值。作为甲基化率的界限值,举例而言,可以在10~30%的范围内,采用任意百分率的值,优选采用15~25%的范围内的任意百分率的值,更优选20%。

[0147] (试剂以及试剂盒)

[0148] 本发明还提供用于推定乳腺癌细胞存在率的试剂。

[0149] 本发明的推定用试剂包含以下:

[0150] (a) SIM1基因的CpG位点中的胞嘧啶残基;

[0151] (b) CCDC181基因的CpG位点中的胞嘧啶残基;

[0152] (c) CFTR基因的CpG位点中的胞嘧啶残基;或

[0153] (d) 这些的组合

[0154] 的甲基化率的分析手段。

[0155] 本发明的推定用试剂中的组分标志物基因 (SIM1基因、CCDC181基因或CFTR基因) 的CpG位点以及甲基化率的分析手段如上所述。

[0156] 本发明的推定用试剂,能够使用甲基化率分析手段对采集自人类受试者的样本中的组分标志物基因的甲基化率进行分析,并基于分析所得的甲基化率,推定采集自人类受试者的样本中的乳腺癌细胞存在率。

[0157] 本发明还提供用于判定乳腺癌的试剂。

[0158] 本发明的判定用试剂包含对以下:

[0159] (a) SIM1基因的CpG位点中的胞嘧啶残基;

[0160] (b) CCDC181基因的CpG位点中的胞嘧啶残基;

[0161] (c) CFTR基因的CpG位点中的胞嘧啶残基;或

[0162] (d) 这些的组合

[0163] 的甲基化率的分析手段。

[0164] 本发明的判定用试剂中的组分标志物基因 (SIM1基因、CCDC181基因或CFTR基因) 的CpG位点以及甲基化率的分析手段如上所述。

[0165] 本发明的判定用试剂,能够使用甲基化率分析手段对采集自人类受试者的样本中的组分标志物基因的甲基化率进行分析,并基于分析得到的甲基化率,判定人类受试者是否罹患乳腺癌。

[0166] 本发明还提供试剂盒。

[0167] 本发明的试剂盒包含以下 (1) 以及 (2):

[0168] (1) 对以下:

[0169] (a) SIM1基因的CpG位点中的胞嘧啶残基;

[0170] (b) CCDC181基因的CpG位点中的胞嘧啶残基;

[0171] (c) CFTR基因的CpG位点中的胞嘧啶残基;或

[0172] (d) 这些的组合

[0173] 的甲基化率的分析手段;以及

[0174] (2) 乳腺癌标志物的测定手段。

[0175] 本发明的用于判定乳腺癌的试剂中的组分标志物基因 (SIM1基因、CCDC181基因或CFTR基因) 的CpG位点、甲基化率的分析手段以及乳腺癌标志物的测定手段如上所述。

[0176] 本发明的试剂盒,可以用于多种用途,举例而言,可以用于预测对乳腺癌的药物治疗效果。

## [实施例]

[0177] 以下,将通过实施例对本发明进行详细说明,但本发明不限于这些实施例。

[0178] 1. 样本

[0179] 1-1. 采集自生物体的样本

[0180] 作为采集自乳腺癌患者样本,从195个乳腺癌患者处取得了195个检测样品的乳腺癌组织样本。各患者的乳腺癌组织样本是从被诊断患有乳腺癌的患者处通过针刺活检收集的。样本使用PAX基因组织系统 (Qiagen, 希尔登, 德国) 固定,包埋到低熔点石蜡中。制备10片10 $\mu$ m的薄切切片,提取DNA。同时,对样本进行显微镜检查,确定癌细胞组分。上述的195个

检测样品中,关于采取自HER2阳性型乳腺癌患者的61个检测样品,对其对于曲妥珠单抗的治疗应答进行了确定,将样本中无残存癌细胞的状态定义为病理学上的完全缓解(pCR),此外的状态定义为非完全缓解(Non-pCR)。另外,将61个样本中的10个样本,通过激光捕获显微切割(LCM)将癌细胞和非癌细胞进行了分离。

[0181] 1-2. 培养细胞株样本

[0182] 作为培养乳腺癌细胞株样本,全部共20个乳腺癌细胞株(BT-474、SK-BR-3、MDA-MB-453、HCC38、MDA-MB-231、T-47D、Hs 578T、MCF7、UACC-3199、ZR-75-1、BT-20、MDA-MB-436、HCC1937、MDA-MB-468、HCC1428、BT-549、AU565、HCC1395、MDA-MB-157以及HCC1954)是从美国模式培养物集存库(ATCC,American Type Culture Collection)处购入的(Rockville,MD)。

[0183] 作为培养正常细胞株样本,从Cambrex (East Rutherford,NL)处购入人类乳房上皮细胞(HMECs)。

[0184] 2. 组分标志物的DNA甲基化水平的测定

[0185] 2-1. DNA甲基化水平的测定方法

[0186] 基因组特异性DNA甲基化水平是通过亚硫酸氢盐·焦磷酸测序法分析得到的。亚硫酸氢盐·焦磷酸测序法具体而言,是通过以下的步骤进行的。用BamHI消化基因组DNA,将1 $\mu$ g的BamHI消化基因组DNA,如报道(Yamashita S,Takahashi S,McDonnell N,Watanabe N,Niwa T,Hosoya K et al.Methylation silencing of transforming growth factor-beta receptor type II in rat prostate cancers.Cancer research.2008;68(7):2112-21)进行亚硫酸氢盐修饰。将亚硫酸氢盐修饰DNA悬浮在40 $\mu$ l的TE缓冲液中,将1 $\mu$ l的试样,如报道(Yoshida T,Yamashita S,Takamura-Enya T,Niwa T,Ando T,Enomoto S et al.Alu and Satalpha hypomethylation in Helicobacter pylori-infected gastric mucosae.International journal of cancer.2011;128(1):33-9)进行亚硫酸氢盐测序。通过生物素化引物对目标基因组区域进行扩增。将生物素标记PCR产物与0.2 $\mu$ M焦磷酸测序·引物退火结合,使用PSQ 96焦磷酸测序·系统(QIAGEN,Valencia,CA,USA)进行焦磷酸测序。甲基化水平使用PSQ Assay·design·software(QIAGEN)获得。

[0187] 2-2. 组分标志物的甲基化水平的分析

[0188] 2-2-1. 组分标志物的筛选

[0189] 从常染色体以及性染色体中的477、344个CpG位点中,搜索在1-2中制备的培养通常细胞株(非癌细胞)中的甲基化率低,而在培养乳腺癌细胞株中甲基化率高的CpG位点,发现了3个CpG位点作为组分标志物。各CpG位点的详细信息如表1所示。

[0190] 2-2-2. 使用组分标志物进行的乳腺癌评估和病理上的癌细胞含有率的相关性

[0191] 为了判定组分标志物对乳腺癌评估的正确性,对于HER2阳性型乳腺癌样本的51个检测样品(CFTR则为33个检测样品),对通过各组分标志物(SIM1、CCDC181以及CFTR)以及这些的组合评估的癌细胞组分和通过显微镜检查评估的癌细胞含有率之间的相关性进行了判定。就组分标志物组合的情况而言,采用了3个标志物中的最高甲基化率。结果如图1~3所示。确认了通过组分标志物评估的癌细胞组分和病理上的癌细胞含有率之间有显著的相关性( $R=0.38\sim 0.77$ )。因此,可以认为这些组分标志物是能够对乳腺癌细胞组分进行评估的标志物。

[0192] 3.通过乳腺癌细胞组分对HSD17B4甲基化水平进行修正

[0193] 3-1.乳腺癌标志物的值的测定

[0194] 对上述1-1中判定为pCR和Non-pCR的HER2阳性型乳腺癌样本的61个检测样品,进行了HSD17B4基因的甲基化水平的测定。HSD17B4甲基化水平是参照上述2-1通过亚硫酸氢盐·焦磷酸测序法测定得到的(参照非专利文献3)。

[0195] 3-2.乳腺癌标志物的值的修正

[0196] 对于上述3-1中测定得到的HSD17B4的甲基化水平,以组分标志物或通过显微镜检查评估的乳腺癌细胞组分,如下进行了修正:[修正HSD17B4甲基化水平=100×(HSD17B4甲基化水平)/(癌细胞组分)]。

[0197] 3-3.统计分析

[0198] 相关性分析是使用皮尔逊积矩相关系数进行的。对于曲妥珠单抗应答者(pCR)和非应答者(Non-pCR)之间的HSD17B4的修正甲基化水平的差异,则使用曼-惠特尼U检验进行了评估。全部的分析使用PASW统计版本18.0(SPSS Japan Inc.,东京,日本)进行,当两侧p值<0.05时,认为具有统计学意义。

[0199] 3-4.癌细胞组分标志物对于HSD17B4甲基化水平的修正的应用

[0200] 对于通过评估癌细胞组分,能够对癌细胞HSD17B4甲基化水平进行多少程度的修正进行了评估。如果基于甲基化的原始数据,则pCR以及Non-pCR样品之间的HSD17B4甲基化水平没有观察到显著差异(p=0.245)(图4)。与之对照,在修正后的结果中,pCR样品的甲基化水平与Non-pCR样品相比显著性提高(显微镜检查:p=0.0001;组分标志物:p=0.0004)。关于预测pCR的灵敏度以及特异性(表2),修正前分别为13.6%以及94.9%。通过DNA甲基化标志物修正后的灵敏度以及特异性(59.1%以及84.6%),与通过显微镜检查修正的灵敏度以及特异性(59.1%以及87.2%)相当。在表2中,HSD17B4甲基化水平使用基于已有报道(前述的Fujii S et al.)的界限值(50%)分为了高水平和低水平。pCR是指病理学上的完全缓解。

[0201] [表2]

[0202] 表2.修正前、通过显微镜检查修正后以及通过甲基化组分标志物修正后的HSD17B4甲基化的预测性能

	HSD17B4 甲基化	pCR 样 本数	Non-pCR 样本数	灵敏度 (%)	特异性 (%)	阳性准确率 (%)
修正前	高	3	2	13.6	94.9	60.0
	低	19	37			
基于显微镜检查 的修正	高	13	5	59.1	87.2	72.2
	低	9	34			
基于甲基化组分 标志物的修正	高	13	6	59.1	84.6	68.4
	低	9	33			

[0204] 4.对各亚型的乳腺癌的适用性的评估

[0205] 关于除了1-1中进行了pCR/Non-pCR判定的61个检测样品外的134个检测样品的乳

腺癌样本 (HER2阳性型、luminal A型、luminal B型以及三阴性型), 在还混入有非癌细胞的状态下对组分标志物 (SIM1、CCDC181以及CFTR) 甲基化率进行了测定。以甲基化率20%作为界限值, 将3个标志物中任意一个甲基化率显示在界限值以上的检测样品判定为“阳性”。结果如表3所示。表3中, HER2表示HER2阳性型, TNBC表示三阴性型。

[0206] [表3-1]

[0207] 表3. 各亚型乳腺癌样本中的SIM1、CCDC181以及CFTR的甲基化率

[0208]

样本编号	组分标志物			亚型
	CFTR	CCDC181	SIM1	
1	42.07	37.94	29.58	HER2
2	33.06	31.54	6.88	
3	19.19	34.19	14.68	
4	18.66	15.78	9.37	
5	68.70	67.19	45.41	
6	51.59	55.62	48.43	
7	3.91	13.63	1.30	
8	55.12	63.72	42.88	
9	40.44	50.37	21.37	
10	28.99	59.85	28.19	
11	57.59	62.77	52.91	
12	45.95	47.05	36.35	
13	69.39	40.21	51.05	
14	24.48	19.51	11.81	
15	58.43	57.76	8.10	
16	23.97	25.65	10.92	
17	26.98	35.26	10.21	luminal A型
18	69.25	57.52	29.72	
19	61.31	48.92	11.68	
20	48.41	58.61	6.99	
21	6.19	56.46	22.51	
22	77.68	58.61	6.25	
23	4.82	19.86	15.69	
24	82.68	80.51	74.21	
25	51.97	63.21	37.92	
26	51.34	31.61	8.04	
27	52.10	64.00	55.91	
28	39.44	30.71	18.08	
29	49.89	45.83	25.38	
30	72.36	72.27	55.46	
31	15.27	52.53	11.16	
32	45.02	63.33	30.80	
33	33.63	34.80	27.24	
34	61.08	66.59	54.48	
35	3.96	61.97	41.23	
36	71.90	50.45	28.11	
37	34.39	12.50	19.67	
38	21.05	13.45	29.56	
39	3.48	10.76	7.37	
40	60.93	67.48	32.83	



[0209] [表3-2]

[0210] 表3接续

[0211]

样本编号	组分标志物			亚型	
	CFTR	CCDC181	SIM1		
41	31.98	14.53	8.73	luminal A型	
42	30.03	41.18	33.43		
43	33.90	40.26	21.07		
44	42.82	63.38	28.16		
45	44.60	44.70	33.34		
46	18.53	31.90	11.72		
47	25.16	50.38	15.76		
48	39.96	50.47	11.26		
49	70.60	61.07	39.86		
50	13.30	48.98	2.52		
51	38.84	49.35	11.99		
52	48.58	55.38	11.72		
53	46.48	61.26	27.55		
54	19.18	24.87	9.03		
55	24.79	25.84	54.06		
56	74.21	75.66	11.53		
57	8.97	13.89	6.33		
58	32.10	44.26	15.72		
59	50.81	63.22	32.64		
60	24.53	41.47	14.48		
61	37.95	36.71	11.08		
62	73.98	62.83	45.52		
63	33.80	24.44	23.29		
64	14.66	18.98	14.20		
65	78.07	72.58	65.08		
66	4.24	85.28	73.66		
67	15.95	55.32	18.63		
68	6.89	14.51	1.85		
69	48.85	13.42	21.29		luminal B型
70	35.46	54.00	35.37		
71	5.02	11.67	3.70		
72	46.05	70.31	46.25		
73	31.45	55.38	24.09		
74	46.48	48.81	4.51		
75	39.63	43.43	27.63		
76	57.40	13.48	39.74		
77	73.11	11.49	1.69		
78	48.86	38.63	38.94		
79	17.87	10.96	38.35		
80	23.83	22.61	5.51		

[0212] [表3-3]

[0213] 表3接续

样本编号	组分标志物			亚型
	CFTR	CCDC181	SIM1	
81	29.43	43.67	17.69	luminal B型
82	45.53	56.19	44.25	
83	59.17	49.79	38.90	
84	31.03	39.80	15.88	
85	22.18	21.40	35.22	
86	7.65	12.45	10.57	
87	47.71	24.02	6.55	
88	57.13	59.70	45.00	
89	30.09	27.03	17.36	
90	72.99	58.14	47.54	
91	50.89	55.05	35.10	
92	86.06	86.34	77.75	
93	53.60	54.20	23.13	
94	53.99	50.26	20.41	
95	3.76	24.18	10.20	
96	87.97	83.37	78.75	
97	57.53	52.67	33.35	
98	52.90	47.66	3.69	
99	21.99	5.90	1.04	
100	69.16	56.05	27.46	
101	71.49	62.97	43.73	
102	40.26	38.00	19.17	
103	21.98	56.71	29.25	
104	69.85	46.64	2.07	
105	64.01	44.33	2.38	
106	61.59	12.51	14.14	
107	25.23	29.12	16.86	
108	40.33	13.25	18.11	
109	66.49	35.87	2.27	
110	64.25	36.83	2.18	
111	26.04	28.60	4.63	
112	48.39	9.85	4.10	
113	13.12	44.79	2.76	
114	28.25	26.27	19.23	
115	10.07	12.75	8.84	
116	2.47	13.01	2.18	
117	77.03	69.01	8.36	
118	22.33	33.04	19.13	
119	53.88	18.13	4.84	
120	46.38	45.66	36.77	
				TNBC

[0215] [表3-4]

[0216] 表3接续

样本编号	组分标志物			亚型
	CFTR	CCDC181	SIM1	
121	40.23	22.12	3.22	TNBC
122	2.95	19.35	2.33	
123	7.73	43.47	10.65	
124	73.12	9.90	2.15	
125	68.06	19.54	3.08	
126	40.05	40.93	26.96	
127	41.84	60.84	35.12	
128	26.82	35.24	19.12	
129	70.52	20.94	1.25	
130	61.04	59.95	27.82	
131	24.44	18.14	0.87	
132	30.22	44.65	25.85	
133	29.36	10.33	7.10	
134	42.37	34.06	22.24	

[0217] 在上述的结果中,在对于各个乳腺癌亚型,SIM1、CCDC181、CFTR以及这些的组合的甲基化率(组合时为最高甲基化率)超过20%则判定为阳性的情况下,各个组分标志物以及乳腺癌判定率(甲基化阳性的乳腺癌患者的比例)如下记表4所示。

[0219] [表4]

[0220] 表4. 各组分标志物的乳腺癌判定率

组分标志物	乳腺癌判定率(甲基化阳性的乳腺癌患者的比例)(%)
SIM1	47.0
CCDC181	78.4
CFTR	81.3
SIM1+CCDC181	81.3
SIM1+CFTR	84.3
CCDC181+CFTR	90.3
SIM1+CCDC181+CFTR	91.0

[0221] 在所有组分标志物、所有组合的情况下都显示具有较高的乳腺癌判定率(甲基化阳性的乳腺癌患者的比例)。特别是上述结果显示了通过组合两种类以上的组分标志物,能够以较高概率对乳腺癌进行判定。

[0222] [序列自定义名称]

[0223] 序列编号1表示SIM1基因的甲基化水平测定CpG位点(1000以及第1001个核苷酸残基)前后1000bp的基因组区域的碱基序列。

[0224] 序列编号2表示CCDC181基因的甲基化水平测定CpG位点(1000以及第1001个核苷酸残基)前后1000bp的基因组区域的碱基序列。

[0225] 序列编号3表示CFTR基因的甲基化水平测定CpG位点(1001以及第1002个核苷酸残基)前后1000bp的基因组区域的碱基序列。

[0226] 序列编号4~6各自表示SIM1基因的CpG位点的甲基化水平测定用正向引物、反向引物以及测序引物的碱基序列。

[0227] 序列编号7~9各自表示CCDC181基因的CpG位点的甲基化水平测定用正向引物、反向引物以及测序引物的碱基序列。

[0228] 序列编号10~12各自表示CFTR基因的CpG位点的甲基化水平测定用正向引物、反

向引物以及测序引物的碱基序列。

## 序列表

<110> 国立研究开发法人国立癌症研究中心

合同会社予幸集团中央研究所

<120> 乳腺癌细胞存在率的推定方法

<130> PRBA-20807

<150> JP2018-171850

<151> 2018-09-13

<160> 12

<170> PatentIn version 3.5

<210> 1

<211> 2001

<212> DNA

<213> 智人

<400> 1

```

aacaactccc tccaaacgca cctctaggag ccgaatctgg tcaccggcgt accaaaacca 60
agccgggagt cgagtctgta gggggtgatt cgtgaggaca acaattgttt ttgggacaat 120
ttgacttttt cttgatgcta cctgttgtgt tttctcagat gagagaaaga gaaggcggtt 180
caaattagag ggtctctaaa agataattag tcagcaagag tcctggtgag cgtttgttta 240
aacagctcct attagggaac cgcgagcatc gttaagtcca ttcattgtaga ctgtatcctt 300
tctgtaagga ataatcatgt tggacttttg tttcctgaaa caaaaccga caaagagttg 360
gctaattgttt aagataaagg actgttggga gctatgtatc tctatatcaa tcattctatt 420
ttgccctgaa atgtaagctt ttcccctagt gtccccctag ctgctcccca taggaaaact 480
ccccggccga gtttccacgg tgaaaaaat cagccccaaa cccaagccat ttcgaaaaag 540
acgcacttct ttgggattaa tttaaagaca gttattgtat ttatttctc tgccaatgaa 600
ttccgtgtcg ccatggtttc agagaaacga gaaagagaaa ataactcccc aaaccggcct 660
cgcgcacgcc gtcggccagc gactgccagc gggatgggcc tttgggcgaa ccctgcctcc 720
tccgccggtg gacctaaccc gcaggcccga cctccgctgc ctccggacgt tcccgttct 780
gctgggccgc cgcgctggac tgagcgtca accaccgac ccaggaggc tcgaggagcg 840
cgggtcgcgc acgacgtcgg gaaggcccag gaggccgca gcagcggcg caaccggaca 900
gcgccgacct ggcaccctga gctcattagg agtccagcgg tcccgcgggt agtaagattc 960
agagcccgga gcctgggcgc gcggagggcg gctcggggcc gactcgtgtc actcagccct 1020
attggccaat gagcgctcg cccctggccg cgccaggcca atgggaggcg agggggcttg 1080
tgagtggcat tgagggaggg cgggagagag gcggccccgg agtgaagttg aagctaaacc 1140
cttaagctat aaagaagtta cggggggcag ttttcggctt ccaattagga ataatagtga 1200
actggcttcg tagcaactac ggaggaccag gattctaaaa tcacctact cgtcccaaag 1260
cttgcatect cctctttctg ccacgcacc ctctccaat tcatgatcca gaaaaggag 1320
ccgggaatgc tctgctctc tctgccgtg ggaagcggag aggccggcg tgctcgtggg 1380
ttggacggta ggcatgagaa cagttaagag atgggcgccc ccgaaacctc tgccgcttgt 1440

```

ggggactgaa ggtaggtgaa gcagaagacg ccccgcgccc gccagcagc cccgcagctc 1500  
 cgcggtgggtg tgggagaggc cgcggcgcct cccacccccg ggggagcctg cgaggggctg 1560  
 tcgagagcgc gccacctggt aacctggcgc tgctgggccc cgcttggtgc agcccctgct 1620  
 gggcagccag agcgctgggt cgccttggga gtcccagag acttggtgtg taagtgtcac 1680  
 tttttgagga atccctcaga cttgagacc aagttaaaga ggtagccaag gtccagatct 1740  
 gccaaagtag ggagctgaaa ggcccctgcc tcgccttggga ggaattctga tttgcggaaa 1800  
 cctgctttcc tgggacgtgc gccggcctg ggcgcgact cggggatccg cggcggaaga 1860  
 tcccttgccg gtcttcagaa atatgtattt tgttgttgg gcaacttccc aggggtctgg 1920  
 aggatgcggc tggcatgggt gagggaatgc cagcattatt tgcagggggc tgtctcgctg 1980  
 ccgatttacc cgtttcttta a 2001

<210> 2

<211> 2001

<212> DNA

<213> 智人

<400> 2

catattgggt acagtgtaca ctgctcaggt gatgggtgca ccaaaatctc agaaatcacc 60  
 gctaaagaac ttatccatgt aaccaaacac cacctgttcc cccaaaacct gatacttaaa 120  
 aaaaaagctg ctatcaccca ttatttga aaatacaatt taggaggatg gtcagaggaa 180  
 aaaactatta acctaggagt tctgaaacct agggcttact tccagctttg gtaagtaact 240  
 ctacaactct acatccatta aaggagattt gactaaaatt cttcaacct ttagcagtga 300  
 ataacctttt ttgggagcgt cttttttcaa gcgctgaagc agtagcaaca ttaggtcatt 360  
 acaactacac cacatctagg aagtgcactc aatattaca ttcatttta cttccattt 420  
 taaaagacct taagtctata acttgaagaa tatctaaaa ttaatttta cattgcctct 480  
 cttggagaaa aggtgctctc catcacatca agtgatagaa atcctaataa taacaattt 540  
 aaaaataatc cgaacacact ggattgtact gaggaatat ttttacttct cccagttctt 600  
 atttatttct ggctgggaat ggatagttgt taaaatcatc tcaactttca ccacaatcta 660  
 aattcatctt gctttgcctt tatacataaa gataagcaga atttgagaaa attctagttg 720  
 accattttga catgcctcaa agctagagca aaaatcctat tttctttggc aactctcttg 780  
 ggtaaagttt tggttttgat taggaaaaaa agaagctgag ggtcaaaaat atggttaact 840  
 attagcaatt tgcttatcta ctaatgaaaa ctaaaaacta cttttctacg ttaactgtga 900  
 agcacataac atatagaaat atagttttag aagttgtcct aagaaattct aaatctgacg 960  
 agagacagct acaagaggga aatttcacaa ccaacacagc ggcacttcgc gagtctctat 1020  
 aacgttagtg atgcgtggat tccttagtcg tcccctaga aacctagct tcaactggccc 1080  
 ttcgaagca ccgggatggg tgccgggagg ggtcgggtgt ctgaggatat tgcagcgcc 1140  
 tgcgcagatt tttgtcgtc tcgtagttag agaggcccg atggaggacg cagaggcacg 1200  
 ctgttgccat ggcagtgtgg tcctggctgc cgcggaggca ggtgccgggg tctcctttgc 1260  
 ctcaatgtga agagcttaga aagaggagga gaggagaact cccccggca tctctgtgat 1320  
 ccagccgcc gcattttaca caggcaagag ggatatagag ggagagaagg gaaggacgat 1380  
 gcggagatgt taatgggtgg atagagggta aaggtataaa cgtctggtat agctgggagg 1440

gagggtaggt agtgagggaa gaagctttct ggtggcgcct taccctctgg agtccgaggt 1500  
 ccaccgagtc tggcactggt ggggaagtct gtaacacct ttcttacta tgtatgaagg 1560  
 atactcgtct gagagtagct tagactaaat tttgtcata ttctttccc tctcaaagtt 1620  
 cttgaacgtg gaaagtaata atttgtccc tgggagaaa agtcactccc gaaattgtag 1680  
 cttccatgag gaggcaatgg tattcatttc gttggtgta cttgggatc ctgaatacag 1740  
 attaaaatta tcacctaaag tccataaaat atttatcatc atactttgtg acagtttatg 1800  
 attttttaaa aacttactta taggatcaca gtcttggcag gatacacagg aataatgggc 1860  
 cagatggttt catgatgtat ggattagttt ttgcttaagt acttacttta tattaaaagg 1920  
 cattctcttt gacatgcttt cagttcttag atttttaagc aagcatatgt ttaaacgct 1980  
 gtaagatggg acccaattat t 2001

<210> 3

<211> 2001

<212> DNA

<213> 智人

<400> 3

cggaagggca aagcagagct atgaaaacct gctgaacaca ttctttatth tcaacacagg 60  
 ttcttgtctt tccatcatga aatgcacatt ttatttgtac tgtatttggg tgaccacaag 120  
 tcaacaacaa gataattcac aagacccttg ccttagatgt gtcggcaata aagtaatcag 180  
 gccaaaattt ttactttcct ttgaattttt caattcaaac acaatgtatg cttgctttta 240  
 cacagtaggg ttcagggatt agagggttg ctctttaaaa accgtcagag acacaggcaa 300  
 tctacacaaa aattctcaga aggaaggcgc ctacgcctgg gaatgccag atgccctca 360  
 gagagttaa gatggcgttt ctctgagtca ggtcaaagtt aacacattac cttegttca 420  
 aagactgctt ggcttccttt cgggtgatta gtcaagatgt tttgctgact gagactagga 480  
 aatctatagg agggcgggtt agtttacatt gttccttgtc attatcgcta aaacactcca 540  
 aagccttcct taaaaatgcg cactgggcta aaaaggatag acaaggaaca catcctgggc 600  
 cggtaattac gcaaagcatt atctcctctt acctccttgc agatthtttt ttctctttca 660  
 gtacgtgtcc taagatttct gtgccacct tggagttcac tcacctaac ctgaaactaa 720  
 taaagcttgg ttcttttctc cgacacgcaa aggaagcgt aaggtaaatg catcagacc 780  
 acactgccgc ggaactttc ggctctctaa ggctgtatth tgatatacga aaggcacatt 840  
 ttcttccct tttcaaaatg caccttcaa acgtaacagg aaccgacta ggatcatcgg 900  
 gaaaaggagg aggaggagga aggcaggctc cggggaagct ggtggcagcg ggtcctgggt 960  
 ctggcggacc ctgacgcgaa ggagggtcta ggaagctctc cggggagccg gttctcccgc 1020  
 cgggtgcttc ttctgtctc cagcgttgc aactggacct aaagagaggc cgcgactgtc 1080  
 gccacactgc gggatgggccc tgggtgctggg cgtaaggac acggacctgg aaggagcgcg 1140  
 cgcgagggag ggaggctggg agtcagaatc gggaaagga ggtgcggggc ggcgagggag 1200  
 cgaaggagga gaggaggaag gagcgggagg ggtgctggcg ggggtgcgta gtgggtggag 1260  
 aaagccgcta gagcaaatth ggggccggac caggcagcac tcggctthta acctgggcag 1320  
 tgaaggcggg ggaagagca aaaggaagg gtggtgtgcg gtagtaggggt ggggtggggg 1380  
 aattggaagc aaatgacatc acagcaggtc agagaaaaag gtttagcgg caggcaccca 1440

gagtagtagg tctttggcat taggagcttg agcccagacg gccctagcag ggaccccagc 1500  
gcccagagaga ccatgcagag gtcgcctctg gaaaaggcca gcgttgtctc caaacttttt 1560  
ttcaggtgag aaggtggcca accgagcttc ggaaagacac gtgcccacga aagaggagg 1620  
cgtgtgtatg ggttgggttt ggggtaaagg aataagcagt ttttaaaaag atgcgctatc 1680  
attcattggt ttgaaagaaa atgtgggtat tgtagaataa aacagaaagc attaagaaga 1740  
gatggaagaa tgaactgaag ctgattgaat agagagccac atctacttgc aactgaaaag 1800  
ttagaatctc aagactcaag tacgctacta tgcacttgtt ttatttcatt tttctaagaa 1860  
actaaaaata cttgttaata agtacctaag tatggtttat tggttttccc cttcatgcc 1920  
ttggacactt gattgtcttc ttggcacata caggtgccat gcctgcatat agtaagtgtc 1980  
cagaaaacat ttcttgactg a 2001

<210> 4

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 4

ggttttagagg gtagtaagat ttagagtt 28

<210> 5

<211> 25

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 5

aactaccccc cctaacttct ttata 25

<210> 6

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 6

accaataaaa ctaaataaca 20

<210> 7

<211> 30

<212> DNA

<213> 人工序列

<220>



<223> 引物  
<400> 7  
gaagagagat agttataaga gggaaatttt 30  
<210> 8  
<211> 25  
<212> DNA  
<213> 人工序列  
<220>  
<223> 引物  
<400> 8  
accctctatc ccaccattaa catct 25  
<210> 9  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> 引物  
<400> 9  
gggaaatttt ataattaata 20  
<210> 10  
<211> 29  
<212> DNA  
<213> 人工序列  
<220>  
<223> 引物  
<400> 10  
aaggaagagt taaggtaa at gtattagat 29  
<210> 11  
<211> 25  
<212> DNA  
<213> 人工序列  
<220>  
<223> 引物  
<400> 11  
ccacctctct ttaa atccaa ttaac 25  
<210> 12  
<211> 18  
<212> DNA  
<213> 人工序列

<220>

<223> 引物

<400> 12

ggagggttta ggaagttt 18

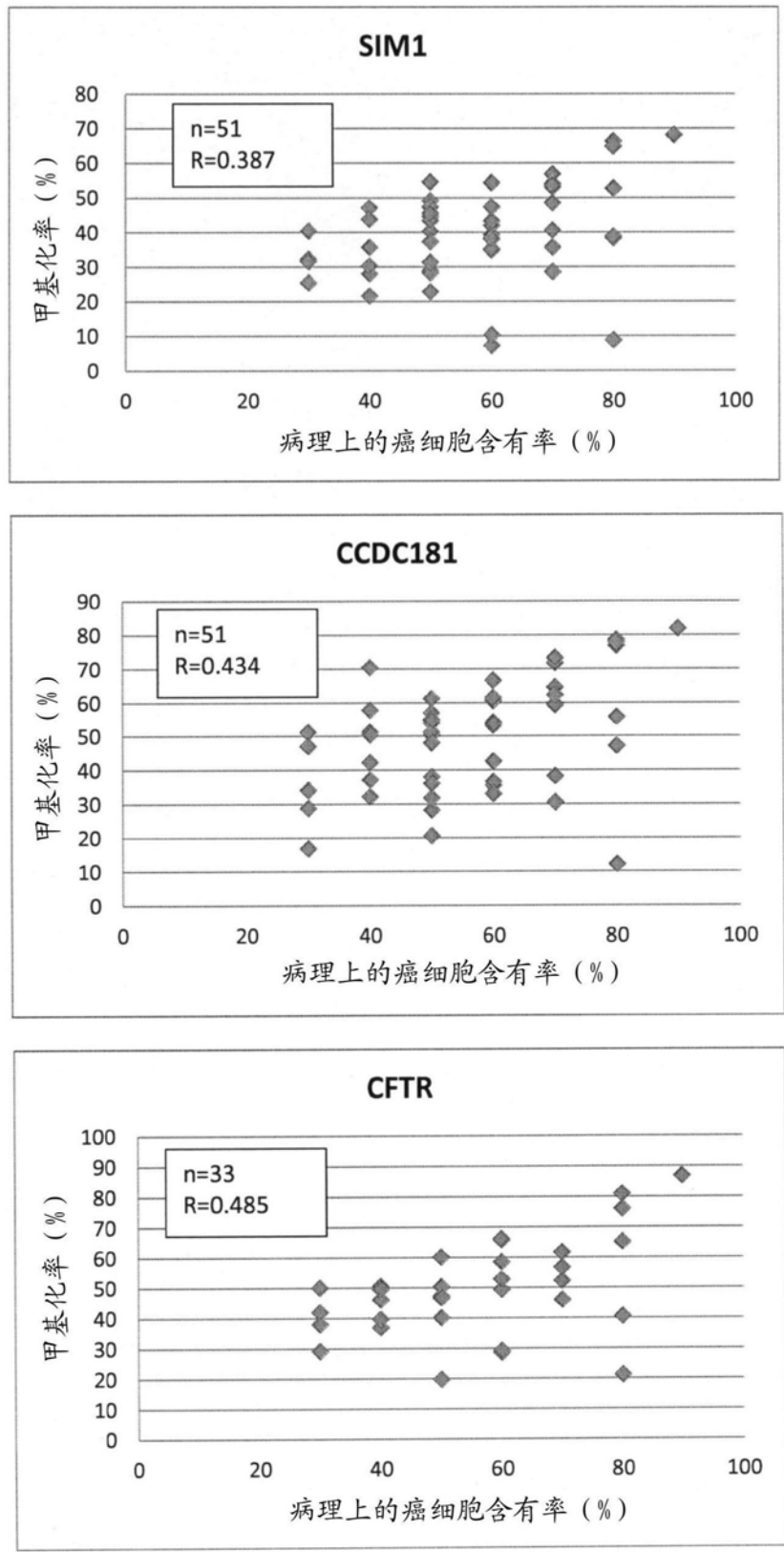


图1

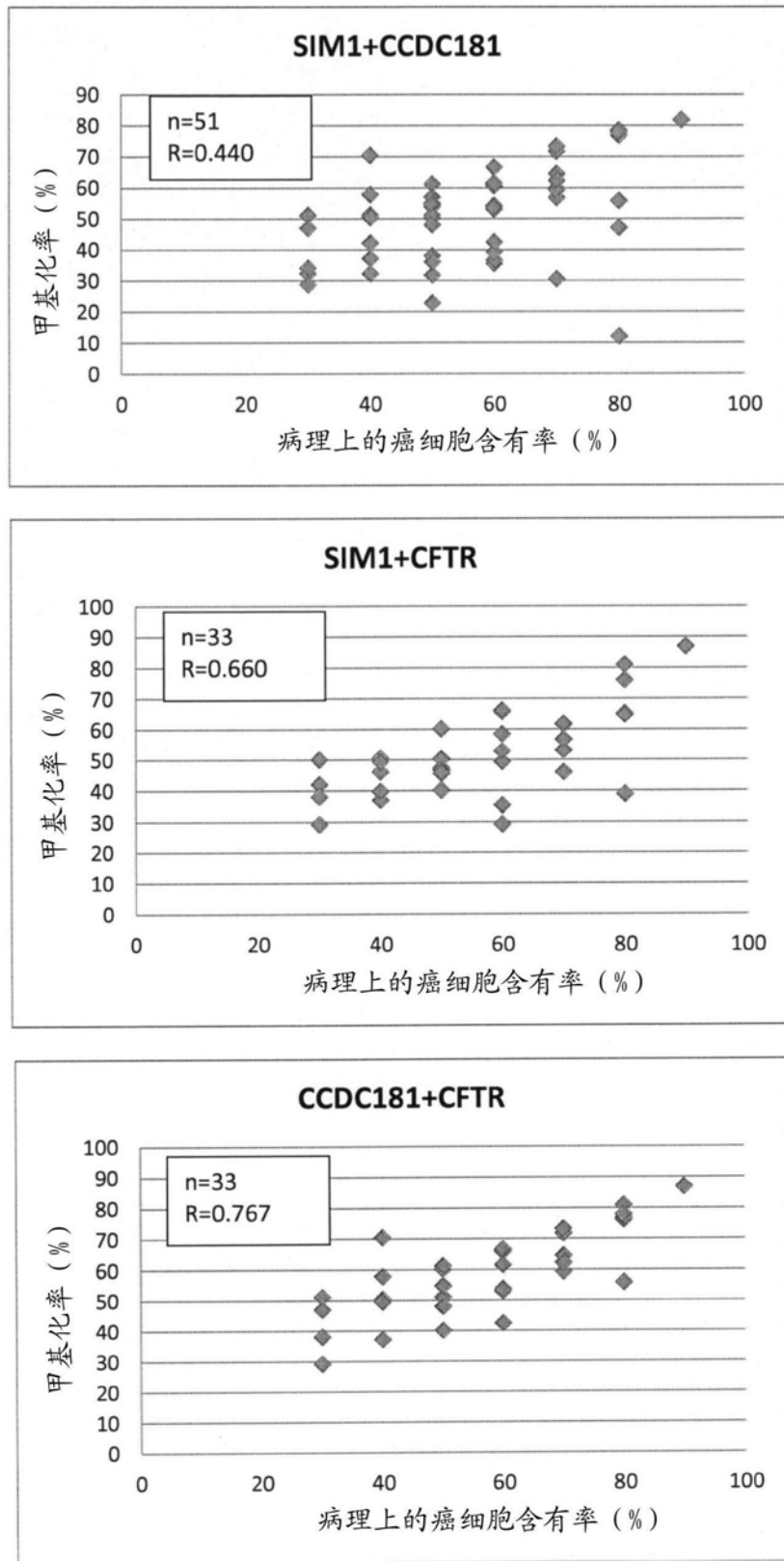


图2

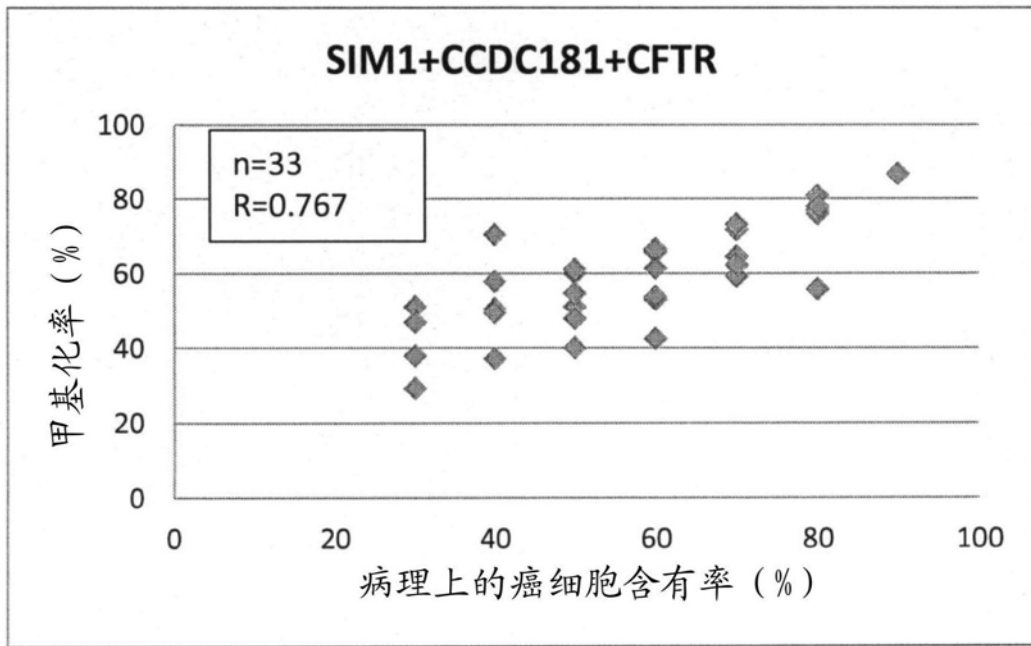


图3

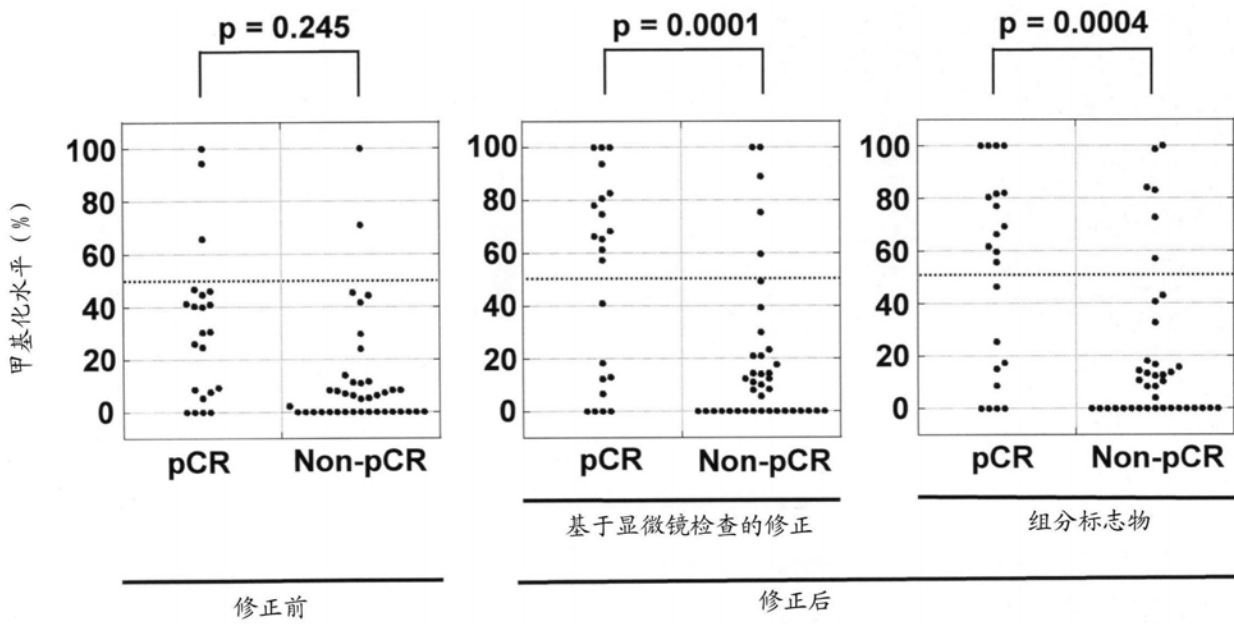


图4