



(12) 发明专利

(10) 授权公告号 CN 102713919 B

(45) 授权公告日 2016. 07. 06

(21) 申请号 201080051528. 1

约翰·史蒂文·赖利

(22) 申请日 2010. 11. 12

(74) 专利代理机构 上海市华诚律师事务所

(30) 优先权数据

31210

61/261, 147 2009. 11. 13 US

代理人 谈晨雯

(85) PCT国际申请进入国家阶段日

(51) Int. Cl.

2012. 05. 14

G06F 19/24(2006. 01)

(86) PCT国际申请的申请数据

(56) 对比文件

PCT/US2010/056455 2010. 11. 12

US 6156507 A, 2000. 12. 05,

(87) PCT国际申请的公布数据

EP 1529847 A1, 2005. 05. 11,

W02011/060222 EN 2011. 05. 19

CN 101147067 A, 2008. 03. 19,

(73) 专利权人 贝克曼考尔特公司

WO 2008/050041 A1, 2008. 05. 02,

地址 美国加利福尼亚州

审查员 吴瑶裔

(72) 发明人 陆久留 子华·王

安东尼奥·阿雷瓦洛·雷耶斯

埃里克·艾伦·古斯塔夫森

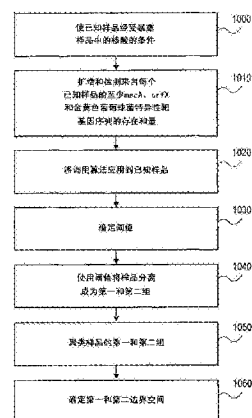
权利要求书3页 说明书17页 附图11页

(54) 发明名称

用于使用聚类来检测生物学状态的存在的系统和方法

(57) 摘要

一种用于判定生物学实体的存在的方法。该方法可以包含至少将与多个样品关联的与第一遗传因子(例如, *mecA*)关联的多个第一输入值、与第二遗传因子(例如, *femA*)关联的多个第二输入值、以及与第三遗传因子(例如, *orfX*)关联的多个第三输入值输入到数字计算机中。每个已知样品包括多个第一输入值中的一个第一输入值、多个第二输入值中的一个第二输入值、以及多个第三输入值中的一个第三输入值。该方法还包括确定与第三遗传因子关联的阈值, 使用该阈值将样品分离成为第一组样品和第二组样品, 在由第一遗传因子和第二遗传因子限定的特征空间中聚类第一组样品, 使用第一组样品来限定第一边界空间, 以及使用第二组样品来限定第二边界空间。第一和第二边界空间将生物学状态与其它生物学状态区分开来。其它实施例还可以包含使用诸如 SCCmec 的遗传因子。



1. 一种用于使用聚类来检测生物学状态的存在的方法, 其特征在于, 包括:

至少将与多个已知样品关联的与第一遗传因子关联的多个第一输入值、与所述多个已知样品关联的与第二遗传因子关联的多个第二输入值、以及与所述多个已知样品关联的与第三遗传因子关联的多个第三输入值输入到数字计算机中, 其中, 每个已知样品包括所述多个第一输入值中的一个第一输入值、所述多个第二输入值中的一个第二输入值、以及所述多个第三输入值中的一个第三输入值;

确定与所述第三遗传因子关联的阈值;

使用所述阈值将所述已知样品分离成为第一组已知样品和第二组已知样品;

在由所述第一遗传因子和所述第二遗传因子限定的特征空间中聚类所述第一组已知样品;

使用所述第一组已知样品限定第一边界空间, 其中, 所述第一边界空间将包含耐甲氧西林金黄色葡萄球菌(MRSA)状态的生物学状态与不包含耐甲氧西林金黄色葡萄球菌MRSA状态的生物学状态区分开来; 以及

使用所述第二组已知样品限定第二边界空间, 其中, 所述第二边界空间将包含耐甲氧西林金黄色葡萄球菌(MRSA)状态的生物学状态与不包含耐甲氧西林金黄色葡萄球菌MRSA状态的生物学状态区分开来,

其中, 所述第一遗传因子包含mecA, 所述第二遗传因子包含金黄色葡萄球菌特异性靶基因序列, 以及所述第三遗传因子包含OrfX, 以及

其中所述第一边界空间和所述第二边界空间是通过边界函数限定的空间。

2. 如权利要求1所述的方法, 其特征在于, 所述第二遗传因子包含femA。

3. 如权利要求1所述的方法, 其特征在于, 所述第一边界空间和所述第二边界空间分别由第一和第二椭圆限定。

4. 如权利要求1或3所述的方法, 其特征在于, 所述第一输入值、所述第二输入值和所述第三输入值是Ct值。

5. 如权利要求1所述的方法, 其特征在于, 进一步包括, 在输入之前:

使所述已知样品经受将暴露所述已知样品中的核酸的条件;

扩增至少所述第一遗传因子、所述第二遗传因子和所述第三遗传因子; 以及

检测至少所述第一遗传因子、所述第二遗传因子和所述第三遗传因子的存在和量。

6. 一种数字计算机, 其特征在于, 包括计算机可读介质和结合到所述计算机可读介质的处理器, 其中所述计算机可读介质包括能够由所述处理器执行的代码, 以实施如权利要求1、2或3所述的方法。

7. 一种用于确定耐甲氧西林金黄色葡萄球菌(MRSA)的存在的系统, 其特征在于, 所述系统包括结合到如权利要求6所述的数字计算机的测量模块。

8. 一种用于使用如权利要求1、2或3所述的方法产生的分析模型的方法, 其特征在于, 用于使用所述分析模型的方法包括:

将与未知样品中的所述第一遗传因子关联的第一输入值、与所述未知样品中的所述第二遗传因子关联的第二输入值、与所述未知样品中的所述第三遗传因子关联的第三输入值输入到所述数字计算机或其它数字计算机中; 和

使用所述数字计算机或其它数字计算机, 使用所述第一边界空间或所述第二边界空间

将所述未知样品分类为与所述生物学状态关联。

9. 一种产生分析模型的方法,所述分析模型将包含耐甲氧西林金黄色葡萄球菌(MRSA)状态的生物学状态与不包含耐甲氧西林金黄色葡萄球菌(MRSA)状态的生物学状态区分开来,其特征在于,所述方法包括:

至少将与第一遗传因子关联的多个第一输入值、与第二遗传因子关联的多个第二输入值以及与第三遗传因子关联的多个第三输入值输入到数字计算机中;

使用所述数字计算机,使用至少所述多个第一输入值以及与第二遗传因子关联的至少所述第二输入值,产生一个以上的中间值;和

使用所述数字计算机,使用所述一个以上的中间值和所述多个第三输入值产生用于所述生物学状态的边界空间,

其中,所述边界空间将包含耐甲氧西林金黄色葡萄球菌(MRSA)状态的生物学状态与不包含耐甲氧西林金黄色葡萄球菌(MRSA)状态的生物学状态区分开来;

其中,所述第一遗传因子包含mecA,所述第二遗传因子包含金黄色葡萄球菌特异性靶基因序列,以及所述第三遗传因子包含SCCmec,

其中所述边界空间是通过边界函数限定的空间。

10. 如权利要求9所述的方法,其特征在于,所述第二遗传因子包含femA。

11. 如权利要求9所述的方法,其特征在于,所述边界空间由椭圆限定。

12. 如权利要求11所述的方法,其特征在于,代价函数被用于使所述椭圆最优化。

13. 如权利要求9、10、11或12所述的方法,其特征在于,所述第一输入值、所述第二输入值和所述第三输入值是Ct值。

14. 如权利要求9所述的方法,其特征在于,进一步包括,在输入之前:

使样品经受将暴露所述样品中的核酸的条件;

扩增至少所述第一遗传因子、所述第二遗传因子和所述第三遗传因子;以及

检测至少所述第一遗传因子、所述第二遗传因子和所述第三遗传因子的存在和量。

15. 一种数字计算机,其特征在于,包括计算机可读介质和结合到所述计算机可读介质的处理器,其中所述计算机可读介质包括能够由所述处理器执行的代码,以实施如权利要求9、10、11或12所述的方法。

16. 一种用于确定耐甲氧西林金黄色葡萄球菌(MRSA)的存在的系统,其特征在于,所述系统包括结合到如权利要求15所述的数字计算机的测量模块。

17. 一种用于使用如权利要求书9、10、11或12所述的方法产生的分析模型的方法,其特征在于,用于使用所述分析模型的方法包括:

将与未知样品中的所述第一遗传因子关联的第一输入值、与未知样品中的所述第二遗传因子关联的第二输入值、与未知样品中的所述第三遗传因子关联的第三输入值输入到所述数字计算机或其它数字计算机中;和

使用所述边界空间将所述未知样品分类为与所述生物学状态关联。

18. 如权利要求1所述的方法,其特征在于,所述第一边界空间和所述第二边界空间两者都是椭圆形的边界空间。

19. 如权利要求1所述的方法,其特征在于,所述第一边界空间和所述第二边界空间两者都已经在原点周围经历平移和角位移。

20. 如权利要求18所述的方法,其特征在于,所述第一边界空间和所述第二边界空间两者都被限定作为mecA和所述金黄色葡萄球菌特异性靶基因序列的测量量的函数。

21. 如权利要求1所述的方法,其特征在于,代价函数被用于找到用于所述第一和第二边界空间的最优参数。

22. 如权利要求9所述的方法,其特征在于,每个中间值被限定作为mecA和金黄色葡萄球菌特异性靶基因序列SA的函数,其中,所述中间值是mecA和SA的加权线性组合,并且使用公式被限定:

中间值 = $\text{mecA} * \sin(-0.3) + \text{SA} * \cos(-0.3)$ 。

用于使用聚类来检测生物学状态的存在的系统和方法

[0001] 相关申请的交叉引用

[0002] 这个申请是非临时的,而且要求在2009年11月13日提交的美国临时申请第61/261,147号的利益,通过引用将其全部结合于此,用于所有目的。

背景技术

[0003] 耐甲氧西林金黄色葡萄球菌(MRSA)涉及伴有包含医院暴发的严重后果的感染,而且对各种各样的抗生素显现抵抗力,因此限制了治疗选择。与MRSA关联的保健有特殊的临床重要性,因为它不但对所有的青霉素和头孢菌素有显著的交叉抗性,而且还对多个其它通常使用的抗生素有典型的抗药性。MRSA感染的治疗通常需要更加昂贵的以及常常更加有毒性的抗生素,这通常被用作最后的防线。因此,MRSA的迅速检测对于治疗以及感染控制措施两者来说在临床上是至关紧要的。

[0004] 由于MRSA可能常常与多个其它相关的细菌相互移植的事实,MRSA的检测进一步是复杂的,多个其它相关的细菌包含甲氧西林敏感金黄色葡萄球菌(MSSA)、耐甲氧西林血浆凝固酶阴性葡萄球菌(MR-CoNS)和/或甲氧西林敏感血浆凝固酶阴性葡萄球菌(MS-CoNS)。

[0005] 在临床微生物学实验室中用于MRSA的检测的传统方法涉及将来自样品的细菌进行培养作为第一步骤,用于MRSA与MSSA以及MR-CoNS的隔离和区分。这个方法是消耗时间的,而且需要最少20到24小时,直至知道结果为止。

[0006] 对于耐甲氧西林金黄色葡萄球菌(MRSA)的检测以及将它与甲氧西林敏感金黄色葡萄球菌(MSSA)区分,许多基于分子的方法已经被出版。一个这样的方法靶向MRSA、葡萄球菌盒染色体的mecA基因(SCCmec,担负甲氧西林耐药性)以及金黄色葡萄球菌的矿泉基因的两个分离区域(美国专利5,702,895,Sinsimer等人,临床微生物学的期刊(Journal of Clinical Microbiology),2005年9月,4585-4591)。使用这个方法的MRSA的清楚的检测被非金黄色葡萄球菌菌株的相互存在所阻碍,非金黄色葡萄球菌菌株诸如是还持有用于甲氧西林耐药性的mecA基因的耐甲氧西林血浆凝固酶阴性葡萄球菌(MR-CoNS)(Becker等人,临床微生物学的期刊(Journal of Clinical Microbiology),2006年1月,p229-231)。

[0007] 更近来的分子的方法对于SCCmec以及金黄色葡萄球菌基因组的侧翼区域(flanking region)利用引物以及探针(美国专利6,156,507,Huletsky等人,临床微生物学的期刊(Journal of Clinical Microbiology),2004年5月,p1875-1884)。SCCmec是可动遗传因子,可动遗传因子携带mecA基因,并且在orfX基因的3'端插入在特异性位点attBsc。SCCmec的左末端邻接有attBsc的无orfX侧,同时SCCmec的右末端邻接有attBsc的orfX侧(Ito等人,Antimicrob.Agent Chemother.,2001,45,p1323-1336;Ito等人,Antimicrob.Agent Chemother.,2004,48,p2637-2651,Noto等人,J.Bacteriol.,2008,190:1276-1283)。这个方法从SCCmec/orfX接点的检测来推断mecA基因的存在。当已经有描述的许多不同类型的SCCmec时,这个方法需要多个引物的使用。这个方法还受到由没有包含mecA基因的SCCmec盒的存在所引起的假阳性结果(Farley等人,临床微生物学的期刊(Journal of Clinical Microbiology),2008年2月,p743-746)以及由新出现的SCCmec类

型没有被化验覆盖所引起的假阴性结果(Heusser等人, *Antimicrob. Agents Chemother.*, 2007年1月, p390-393)。

[0008] 另一个方法利用跨越不同的SCCmec类型的高同源性的区域中的一个引物以及侧翼金黄色葡萄球菌DNA中的一个引物(Cuny等人, *Clin. Microbiol Infect* 2005;11:834-837, 欧洲专利1529847B1)。这个方法也经受假阳性, 因为随着引物包围这个区域, 同样引发MSSA的概率是高的。

[0009] 最后, 已经描述一种方法, 使用特异性抗体以及磁性颗粒对于金黄色葡萄球菌进行积极地选择(Francois等人, 临床微生物学的期刊(*Journal of Clinical Microbiology*), 2003年1月, p254-260; 欧洲专利1,370,694B1)。这个方法使金黄色葡萄球菌富足, 但是需要三个引物/探针组的使用以便积极地识别MRSA并且降低检测CoNS的可能性。该方法需要离心步骤以及分离溶菌步骤以便恢复核酸。

[0010] 市场上可买到的MRSA化验靶向SCCmec右末端接点以及orfX。五个不同类型以及很多亚型的SCCmec已经被识别, 并且新的SCCmec亚型的出现的潜力是高的。此外, 来源于MRSA的MSSA可能保留一部分SCCmec序列, 而没有mecA基因是可能的。因此, 靶向具有orfX的SCCmec右末端接点的化验可能给出伴随MRSA衍生出的MSSA的假阳性结果, 以及伴随MRSA携带新的紧急的SCCmec型/亚型的假阴性结果。

[0011] 如此, 当前用于MRSA的检测的方法是麻烦的、消耗时间的, 而且不可能是特别准确的。因此, 需要快速、易用、可靠而且能够从包含MSSA和/或MR-CoNS的其它相关的细菌检测并且同时区分诸如MRSA的生物学实体或其它生物学实体的方法和系统。

发明内容

[0012] 本发明的实施例涉及用于判定生物学状态存在于或不存在于样品中的系统和方法。本发明的一些实施例可以指向包含检测样品中的耐甲氧西林金黄色葡萄球菌(MRSA)的方法, 样品可能另外包含甲氧西林敏感金黄色葡萄球菌(MSSA)、耐甲氧西林血浆凝固酶阴性葡萄球菌(MR-CoNS), 和/或细菌的其它菌株。

[0013] 本发明的一个实施例指向一种(产生分析模型的)方法, 包括: 将与多个样品关联的至少与第一遗传因子(例如, mecA)关联的多个第一输入值、与第二遗传因子(例如, femA)关联的多个第二输入值、以及与第三遗传因子(例如, orfX)关联的多个第三输入值输入到数字计算机中。每个已知样品包括多个第一输入值中的一个第一输入值、多个第二输入值中的一个第二输入值、以及多个第三输入值中的一个第三输入值。该方法还包括确定与第三遗传因子关联的阈值, 使用阈值将样品分离成为第一组样品和第二组样品, 在通过第一遗传因子和第二遗传因子限定的特征空间中聚类第一组样品, 使用第一组样品限定第一边界空间, 以及使用第二组样品限定第二边界空间。第一和第二边界空间区分生物学状态与其它生物学状态。

[0014] 本发明的另一个实施例指向产生分析模型的方法, 该分析模型将生物学状态与其它生物学状态区分开来。该方法包括: 将至少与第一遗传因子(例如, mecA)关联的多个第一输入值、与第二遗传因子(例如, femA)关联的多个第二输入值、以及与第三遗传因子(例如, SCCmec)关联的多个第三输入值输入到数字计算机中。该方法还包含使用数字计算机, 使用至少多个第一输入值以及与第二遗传因子关联的至少第二输入值来产生一个以上的中间

值,以及使用数字计算机,使用一个以上的中间值以及多个第三输入值产生用于生物学状态的边界空间。边界空间将生物学状态与其它生物学状态区分开来。

[0015] 本发明的其它实施例指向用于使用分析模型以及计算机可读介质(例如,诸如存储芯片以及存储磁盘的非暂时的计算机可读介质)的方法以及使用该分析模型的系统。

[0016] 在这个公开中,这些实施例以及其它实施例稍后将被更详细地描述。

附图说明

[0017] 图1显示说明根据一个实施例的用于形成分析模型所采取的步骤的流程图。

[0018] 图2显示说明根据一个实施例的大于和小于orfX的阈值的二维空间中的femA对比mecA的图。

[0019] 图3A和3B显示说明根据一个实施例的二维空间中的femA对比mecA的图。

[0020] 图4显示说明根据一个实施例的使用分析模型所采取的步骤的流程图。

[0021] 图5显示说明根据一个实施例的用于使用中间值来形成分析模型所采取的步骤的流程图。

[0022] 图6A和6B显示说明根据一个实施例的二维空间中的mecA对比SCCmec以及femA对比SCCmec的显示图。

[0023] 图7显示说明根据一个实施例的用于产生中间值以及使用中间值与SCCmec以产生边界空间的方法的图。

[0024] 图8显示根据一个实施例的二维空间中的中间值对比SCCmec的图。

[0025] 图9说明根据一个实施例的作为二维空间中的中间值对比SCCmec的函数的边界空间。

[0026] 图10显示说明根据一个实施例的用于使用分析模型所采取的步骤的流程图。

[0027] 图11是能够用于执行各种实施例的系统的方框图。

具体实施方式

[0028] 各种实施例公开了从样品中的诸如mecA、SCCmec、orfX和/或诸如femA的金黄色葡萄球菌特异性靶基因序列的遗传因子的测量量识别样品中的耐甲氧西林金黄色葡萄球菌(MRSA)的系统和方法。本发明的一些实施例通过当orfX的检测量小于阈值时使用第一边界空间判定MRSA的存在和当orfX的检测量大于阈值时使用第二边界空间判定MRSA的存在,来判定样品中的MRSA的存在。其它实施例从至少mecA和金黄色葡萄球菌特异性靶基因序列产生中间值,而且使用边界空间来判定MRSA的存在。该边界空间可以使用中间值和SCCmec来被限定。

[0029] 在这个公开中使用的所有科学和技术术语具有本领域中通常使用的含义,除非另有说明。如在这个公开中使用的,以下的词或短语具有指定的含义。

[0030] 如在此使用的,术语“样品”以它的最宽的意义被使用,而且指的是生物学起源的任何类型的物质,例如,可以是任何流体、组织或细胞。例如,样品可以是生物学流体,例如,小便、血液、血清、血浆、鼻水、脑脊液等等。做为选择,样品可以是培养细胞或组织、微生物的培养物、或从生物学物质产生或来源于生物学物质的任何碎片或产物。可选择地,样品可以被净化、局部地净化、未净化、浓缩或扩增。

[0031] 如在此使用的术语“遗传因子”可以涉及在本发明的方法中作为靶是有用的所关心的基因组中的子序列。在一些实施例中，遗传因子是开放阅读框或基因，诸如例如葡萄球菌中的orfX、femA或mecA。遗传因子还可以是可能或不可能包含mecA基因的可动遗传因子，诸如葡萄球菌盒染色体、SCCmec。

[0032] 如在此使用的，“输入值”可以是例如可以与遗传因子关联的任何适当的值。例如，输入值可以是与各种靶基因关联的Ct值。

[0033] 如在此使用的，“边界空间”可以通过“边界函数”限定的空间。“边界函数”可以是用于判定数据与生物学状态关联或与生物学状态不关联的数学函数。通过使用神经网络、代价函数等等，可以以包含手工的任何适当的方式产生边界函数。边界函数还可以由包含椭圆形、矩形、圆形等等的任何适当的形状或线条来表示。边界函数还可以是在形状上有规律的或无规律的。

[0034] 如在此使用的，术语“SCCmec”、“SCCmec序列”和“SCCmec盒”被可交替地使用，以涉及被称为葡萄球菌盒染色体的遗传因子，该遗传因子携带mecA基因并且被插入如Ito等人(2001, *Antimicrob. Agents Chemother.*, 45:1323-1336)中描述的葡萄球菌基因组中。在一个实施例中，SCCmec在orfX接点。

[0035] 在这个申请中，SCCmec插入位点被称为“orfX-ISS/attBsc”。插入位点是在此涉及的金黄色葡萄球菌基因的3'端，如“orfX”。发生SCCmec插入的染色体位点被称为“attBsc”。在插入位点的特异性序列在这里指的是“orfX-插入位点序列(orfX-ISS)”或“attBsc核心区域”。已知这个序列在金黄色葡萄球菌中是高保存的序列(Ito等人, *Antimicrob. Agent Chemother.*, 2001, 45, p323-1336, Noto等人, *J. Bacteriol.*, 2008, 190: 1276-1283)。

[0036] 在插入金黄色葡萄球菌的orfX-ISS/attBsc区域之后，SCCmec左末端接点区域被称为MRSA-LE并且右末端接点区域被称为MRSA-RE。在左末端接点，SCCmec序列与attBsc的非orfX侧邻接。中在右末端接点中，SCCmec序列与attBsc的orfX侧邻接。orfX-ISS/attBsc区域被详细地描述在Ito等人中(2001, *Antimicrob. Agents Chemother.*, 45:1323-1336; Ito等人, *Antimicrob. Agent Chemother.*, 2004, 48, p2637-2651, Noto等人, *J. Bacteriol.*, 2008, 190:1276-1283)以及美国专利第6,156,507号中，所有都通过引用被结合在此。如果SCCmec插入不存在，那么orfX-ISS/attBsc区域是不间断的。如果orfX-ISS/attBsc区域被识别作为完整的通过扩大方法论，那么这表示SCCmec盒没有被插入。然而，没有orfX-ISS attBsc区域的扩增不表示mecA基因存在。已知在SCCmec盒变得被插入之后，可能失去mecA基因。因此即使没有mecA，SCCmec盒仍然可以防止orfX-ISS/attBsc区域的扩增。

[0037] “低聚核苷酸”是具有两个以上的核苷酸亚基共价连接在一起的核苷聚合物。低聚核苷酸通常是大约10到大约100核苷酸。核苷酸亚基的糖类基团可以是核糖、脱氧核糖或它们的改性派生物，诸如OMe。核苷酸亚基可以通过诸如磷酸二酯键合、改性键合的键合或通过不防止低聚核苷酸与它的互补靶核苷酸序列杂交的非核苷酸部分被结合。改性键合包含其中标准磷酸二酯键合被诸如硫代磷酸酯键合、磷酸甲酯联接或中性的肽键代替的那些键合。含氮碱基类似物还可以是对应于本发明的低聚核苷酸的成分。“靶核苷酸”是包含靶核苷酸序列的核酸。“靶核苷酸序列”、“靶核苷酸序列”或“靶序列”是可以被杂交到互补

低聚核苷酸的特异性脱氧核糖核苷酸或核苷酸序列。

[0038] 如在此使用的,术语“探针”指的是能够杂交到所关心的标靶核酸的低聚核苷酸。作为探针通过互补碱基对结合到所关心的标靶核酸的结果,杂交出现。一个本领域的技术人员将理解的,依据杂交条件的严格性,探针将典型地基本上使缺乏完全互补的标靶序列与探针序列形成化学键。探针可以与适当的标签或报道基因部分(reporter moiety)关联,以使探针(以及因此它的标靶)可以被检测、目测、测量和/或测定。

[0039] 如在此使用的,术语“引物”指的是用于引发核酸综合物的低聚核苷酸。引物通过互补碱基对与模板杂交,并且因此用于发起复制。以与以上对于探针的描述的同样方式出现杂交。在PCR中,使用两个引物:典型地与有义链杂交的“正向引物”以及典型地与反义链杂交的“反向引物”。

[0040] 如在此使用的,术语“PCR”指的是用于通过在没有使用活体的情况下的DNA的酶的复制,在较长的双链DNA分子之内的短的DNA序列(通常50到600个碱基)的指数式扩增的技术(Mullis等人,Methods Enzymol.,1987;155:335-50)。其它体外扩增技术可被用于本发明,而且对于那些技术人员是众所周知的。这些方法例如包括连接酶链式反应(LCR)、核酸序列扩增法(NASBA)、链置换扩增法(SDA)、转录介导扩增法(TMA)、分支DNA(bDNA)以及滚环扩增技术(RCAT)。

[0041] 在此使用的术语“实时PCR”指的是PCR的类型,在每个扩增循环之后,当实时积聚在反应中时,扩增的DNA被量化(Heid等人,Genome Research,1996(10):986-994)。用于实现实时PCR的许多探针化学处理对于那些技术人员来说是众所周知的。一个通常使用的方法是TaqMan®化验(参见,例如,美国专利第5,210,015号;第5,487,972号;以及第5,804,375号)。可以使用并且可以被商业上采购的其它实时PCR探针包含FRET引物、Molecular Beacons、Scorpion引物®、Amplifluor引物®、LUX引物®、Eclipse®以及Ultimate探针®。对于实时PCR技术的综述,参见Bustin等人,J.Mol.Endocrin,34:597-601(2005)。

[0042] 如在此使用的,术语“多重PCR”指的是PCR的类型,其中,超过一个组的引物被包含在反应中,该反应允许两个以上的不同的标靶在单个反应试管中被扩增。术语“多重PCR”还指的是多个引物以及探针被使用,但只有一个标靶被扩增的PCR。在一个实施例中,本发明的多重PCR是实时PCR。

[0043] 如在此使用的,“生物学状态”可以涉及来源于例如病人的任何来源的样品的特殊的生物学状态。在大多数情况下,生物学状态涉及样品是否包含特殊的生物学实体,例如,标靶疾病生物体或与疾病关联的患者细胞。例如,一个生物学状态可以是包含MRSA细菌的样品,同时另一个生物学状态可以是不包含MRSA细菌的样品。在其它实例中,生物学状态可以涉及样品是否包含癌细胞。

[0044] 本发明的一个实施例涉及用于检测样品中的MRSA的化验,该样品可以包含MRSA、MSSA、MR-CoNS或其它细菌。本发明的实施例利用用于同时扩增和检测多个标靶的组合的多重PCR。

[0045] 根据一个实施例,标靶DNA的初始量通过PCR阈值循环(Ct)被测量。例如,对于要被分析的反应,确定限定的信号阈值。达到这个信号阈值所需的循环数(Ct)被确定,用于标靶核酸以及用于参考或标准核酸。靶分子的绝对的或相对的复制数可以根据对于标靶核酸获得的Ct值与参考核酸相比来确定。因此Ct与初始标靶DNA的量成反比,参见Heid等人,1996,

Genome Research 6(10):986,用于通过引用在此结合的全部讨论。可以采用其它数学方法,其考虑基于在一个识别方法期间扩增的基因的预定设置量或数量的指征的特殊的靶基因的初始量的外推法。

[0046] 在一个实施例中,本发明指向判定样品中的MRSA的存在的方法,所述方法包括使样品暂时并且在条件下经受实时PCR,以便产生扩增产物的水平,该扩增产物的水平足以通过荧光被检测,而且指示样品中的一个以上的MRSA特异性靶序列的初始水平。

[0047] 在另一个实施例中,利用一组引物(正向和反向)以及探针进行扩增。可以利用在它的5'端的荧光报道基因分子以及在它的3'端的猝灭分子标明该探针。该猝灭分子防止信号从荧光报道基因分子的发射。探针杂交到正向和反向引物杂交的区域之间的靶序列的区域。随着聚合酶沿着探针已经杂交的链移动,探针的5'端通过聚合酶的外切核酸酶活性被劈开,因此允许由于猝灭部分的分离引起的荧光信号的发射。

[0048] 在具体的实施例中,本发明的探针可以包含双标签的荧光探针,双标签的荧光探针包括荧光的报道基因(荧光团)和荧光的或非荧光的猝灭剂分子。本发明的荧光团可以在任何位置被附接到探针,包含在5'末端、3'末端或任何末端的内部。在本发明的实施例中,荧光团和猝灭剂分别被附接到探针的5'和3'末端。荧光团的实例包括但不限于FAM、ROX、HEX、NED、Cy5、Texas Red、Calfluor Red、CalFluor Orange、Quasar 670、Quasar705。猝灭剂的实例包括但不限于TAMARA、Blackhole猝灭剂BHQ-1, BHQ-2。

[0049] 在另一个实施例中,本发明提供一种利用三个靶化验来从MSSA、MR-CoNS或其它细菌检测和区分MRSA的方法,其中,化验中使用的靶(可以是遗传因子的实例)包含mecA基因序列、金黄色葡萄球菌特异性靶基因序列、SCCmec基因序列和/或orfX。在具体的实施例中,金黄色葡萄球菌特异性靶基因是femA。在以下的描述中,femA常常被明确地提及作为金黄色葡萄球菌特异性靶基因序列;然而,根据各种实施例,其它金黄色葡萄球菌特异性靶基因序列同样可以被使用。

[0050] 本发明的一些实施例指向使用至少第一和第二边界空间来形成分析模型,以及使用这样的分析模型。本发明的其它实施例指向使用至少一个中间值来形成边界空间的分析模型的产生和使用。另外,其它实施例涉及用于使用这样的分析模型的方法以及使用这样的分析模型的系统。以下进一步详细描述这些方法。

[0051] 使用至少两个边界空间的实施例

[0052] 图1显示流程图,该流程图图解能够用于构造根据本发明的实施例的分析模型的步骤。有时,分析模型能够用于判定MRSA是否存在于样品中。

[0053] 在步骤1000,选择量的已知样品经受暴露样品中的细菌的核酸的条件。在一些实施例中,已知样品是已知是否与特殊的生物学状态关联的一个样品。例如,已知样品可以是知道样品是否与MRSA关联的一个样品。如此,已知样品可以被用于构造模型,该模型可以判定以后的未知样品是否也包含MRSA。

[0054] 对于使样品经受暴露样品中的核酸的条件,有许多不同的方式。例如,样品中的细胞可以根据众所周知的技术被溶解。然后可以通过例如使样品的温度上升以分离核酸的链来使核酸变性。

[0055] 在步骤1010,在样品中,诸如mecA、orfX和金黄色葡萄球菌特异性靶基因序列的至少三个靶的特性(例如,相对或绝对量或基因显现量)。根据一个实施例,金黄色葡萄球菌

特异性靶基因序列是femA。有许多不同的方式来测量样品中的这种遗传因子的量。例如,多重PCR处理能够用于为测量的每个靶测量PCR阈值循环(Ct)。

[0056] 在测量至少三个靶的特性之后,与那些特性关联的值可以被输入到数字计算机中。以下提供示范性数字计算机的细节。各种输入值可以以任何适当的方式被输入到数字计算机中。在一些实施例中值可以自动地(例如,通过连接到产生输入值的测量模块的数据或用于产生输入值的数据)或通过用户人工地被输入到数字计算机中。

[0057] 在一些实施例中,至少与第一遗传因子(例如,诸如mecA的第一靶)关联的多个第一输入值、与第二遗传因子(例如,诸如femA的第二靶)关联的多个第二输入值、以及与第三遗传因子(例如,诸如orfX的第三靶)关联的多个第三输入值被输入到数字计算机中。第一第二以及第三输入值与多个已知样品关联。每个已知样品包括多个第一输入值中的一个第一输入值、多个第二输入值中的一个第二输入值以及多个第三输入值中的一个第三输入值。第一、第二以及第三值可以是与第一第二以及第三遗传因子关联的Ct值。

[0058] 在步骤1020,调用算法可以被应用于来自每个已知样品的每个测量的靶。调用算法可以具有任何适当的指令的组合。在一些实施例中,调用算法可以包含以任何适当组合的一个以上的图1中的步骤1030、1040、1050以及1060。

[0059] 在步骤1030,与第三遗传因子(例如,orfX)关联的阈值被确定。可以以任何适当的方式确定该阈值。例如,在一些实施例中,阈值可以简单地是诸如“45”的离散值。该阈值可以已经先前被输入到数字计算机中并且存储在数字计算机中的存储器中,或可以在第一输入值、第二输入值以及第三输入值被输入到数字计算机中的几乎相同的时间通过用户被输入。

[0060] 在步骤1040,调用算法使用阈值将已知样品分离成为第一组样品以及第二组已知样品。例如,调用算法可以通过判定哪个样品落在阈值以上以及哪个样品落在阈值以下来分离第一组已知样品与第二组已知样品。说明性地,如果第三遗传因子是orfX,而且阈值是“45”的Ct值,那么第一组已知样品可以对于orfX具有小于45的Ct值,而且第二组已知样品可以对于orfX具有大于或等于45的Ct值。

[0061] 在步骤1050,在样品被分离之后,第一组样品中的样品被聚类在由第一遗传因子和第二遗传因子限定的特征空间中。第一和第二组样品可以以任何适当的方式被聚类。例如,第一组样品可以被标绘在由第一遗传因子和第二遗传因子限定的二维空间中。第二组样品可以被标绘在由第一遗传因子和第二遗传因子限定的二维空间中。这个例如被图解在图2中,图2显示对于 $\text{orfX} < 45$ 和 $\text{orfX} \geq 45$,femA对比mecA的两个图表。

[0062] 在步骤1060,分别使用第一和第二组样品来限定第一边界空间和第二边界空间。第一和第二边界空间将生物学状态与其它生物学状态区分开来。在图2中,第一和第二边界空间被显示为椭圆。

[0063] 第一和第二边界空间可以以任何适当的方式被确定,而且可以具有任何适当的形状。例如,在一些实施例中,可以由椭圆来限定第一和第二边界空间。在其它实施例中,可以由矩形、多边形、平行六边形或其它形状来限定边界空间。这种边界空间可以通过神经网络及其他最优化算法被确定和最优化。第一和第二边界空间可以形成至少一部分分析模型,该分析模型能够用于将生物学状态与其它生物学状态区分开来。有时,为了辅助用户,可能理想的是在诸如LCD屏幕的显示器上在包含第一和第二样品组的二维图表上图解地显示第

一和第二边界空间。

[0064] 现在可以描述图1中所示的说明该方法的具体的实施例。对于3标靶的“策略A”算法,296个进行的样本被收集和提供,用于分析模型开发。在策略A执行中,三个标靶对应于包含mecA、femA和orfX的第一、第二和第三遗传因子。

[0065] 在任何分类问题中,当适当的特性被测量以将一个种类的事件与另一个区分开来时,唯一的特征空间将存在,允许数据被分类。这对于MRSA分类是正确的。为了产生用于其应用的分析程序,如图3A所示,观察特征空间。在图3A中,阳性和阴性数据两者被标绘在femA对比mecA的二维特征空间中。阳性数据点沿着左下角中的斜线聚类,并且阴性数据点尤其在上半部空间和右下角中更随机的散开。

[0066] 基于这个以及其他的确定的数据,数学模型被选择以概述该阳性特征空间。用于阳性事件的数据类似椭圆,该椭圆被选择以用公式表示阳性和阴性聚类之间的边界。椭圆形的边界可以被认为是理想的选择,因为阳性数据通常在较佳的特征空间中形成高斯分布(注意,表示特征空间的边界的高斯分布的横截面是椭圆形)。这证明图3b中的椭圆形的模型是正确的。数学上,给出用于椭圆形模型的等式为:

$$[0067] \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

[0068] 其中,a是半长轴,并且b是半短轴。在这种情况下,椭圆被假定为在起点周围。为了支持该MRSA检测,椭圆形的模型必须支持在由(x₀,y₀)限定的起点周围并且具有角度θ的平移和角位移。

[0069] 如此,在它的最普通的形式中,椭圆可以被完全地表示在二维的特征空间中,如:

$$[0070] \quad \frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1$$

$$[0071] \quad \text{其中} \begin{cases} x' = -\cos(\theta)x + \sin(\theta)y - x_0 \\ y' = \sin(\theta)x + \cos(\theta)y - y_0 \end{cases}$$

[0072] 一般化的椭圆可以取决于一组五个参数,a、b、x₀、y₀和θ。一旦该组参数被确定,就限定唯一的椭圆(或特征空间)。

[0073] 为了获得最优的一组椭圆形的参数,可以基于分类结果限定代价函数。对于这类应用,通常使用的代价函数是在接收器操作特性的曲线或ROC曲线之下的区域(关于ROC曲线的更多细节,参考http://en.wikipedia.org/wiki/Receiver_operating_characteristic)。ROC曲线提供敏感性对比1-特异性的图表,用于应用。对于MRSA应用,假阳性数量和假阴性数量的组合被采用作为代价函数:

[0074] 代价=c₁*FN#+c₂*FP#。

[0075] 加权系数c₁和c₂被选择以表示特殊问题中的偏好。

[0076] 给定用于特征空间和设计的代价函数定义的数学模型,用于特征空间的模型可以被最优化以使代价函数最小化。具有最低代价的模型的实现是用于分类问题的最优方案。有许多可以被利用的最优化过程,诸如爬山法(Hill Climbing)、模拟退火算法(Simulated Annealing)、遗传算法(Genetic Algorithms)等等。对于这个申请,利用遗传算法。更多关于遗传算法的详细信息可以在http://en.wikipedia.org/wiki/Genetic_algorithms中被找到。

[0077] 参考图2,观察到对于这个模型,在二维特征空间中由femA对比mecA测量的特征未必总是均匀的或理想的。例如,如图2所示的,当orfX等于或大于45时,阳性数据点形成“肥的”聚类(如描画轮廓的),但是当orfX小于45时,阳性数据点形成具有椭圆形的更加一列式的聚类。结果,分别对于orfX<45和orfX≥45建立两个椭圆形的模型。利用这个orfX相关的椭圆形的模型,对于MRSA阳性样品的数据基本上被分类在三维空间中。

[0078] 然后,图2中说明的分析模型的图解的描绘可以用来将未知样品分类作为与诸如MRSA的特殊生物学状态关联。

[0079] 通常,形成的分析模型的使用可以包含将与未知样品关联的第一输入值、第二输入值和第三输入值输入到数字计算机中。数字计算机可以是与用于形成分析模型相同的数字计算机,或可以是不同的数字计算机(例如,当分析模型被形成在数字计算机上时,但是被存储在使用的另一个数字计算机中)。第一输入值、第二输入值和第三输入值与未知样品中的第一、第二和第三遗传因子关联。在输入值被输入之后,数字计算机利用使用分析模型的第一边界空间或第二边界空间,将未知样品分类为与生物学状态关联。

[0080] 图4说明根据一个实施例的能够用于判定MRSA是否存在于样品中的步骤。如在此使用的,未知样品指的是不知道MRSA是否存在于其中的样品。图4中的步骤可以使用模型来判定未知样品是否包含MRSA,该模型诸如是使用来自图1的步骤产生的模型。未知样品可以具有各种测量的标靶,而且通过分析从未知样品的测量的标靶产生的中间矢量相对于模型的边界函数归于哪里,这些测量值能够用于检测MRSA的存在。

[0081] 在步骤1100,未知样品经受暴露样品中的细菌的核酸的条件。以上关于图1中的步骤1000描述的相同的技术同样可以在步骤1100被使用。

[0082] 在步骤1110,可以从未知样品确定与至少三个标靶,mecA、femA和orfX(或其它遗传因子)关联的特性。在以上图1中的步骤1010期间使用的相同的技术能够用于实现步骤1110。该特性可以和用于形成分析模型的特性一样。例如,如果Ct值被用于形成分析模型,那么可以在步骤1110中对于三个标靶确定Ct值。

[0083] 在步骤1120,分析模型可以被应用于与三个标靶关联的输入值。与分析模型关联的处理可以包含至少图4中的步骤1130和1140。

[0084] 在步骤1130,例如,该方法判定与orfX关联的第三输入值是否大于或小于该分析模型中的阈值(例如,“45”)。

[0085] 在步骤1140,如果与未知样品关联的第三输入值小于阈值(例如,45),那么然后先前形成的第一边界空间被用于判定样品是否与所关心的生物学状态(例如,MRSA)关联。做为选择,如果与未知样品关联的第三输入值大于阈值(例如,45),那么然后先前形成的第二边界空间被用于判定样品是否与所关心的生物学状态(例如,MRSA)关联。

[0086] 如果希望,其他的规则能够用于进一步分类未知样品。例如,一个、两个或三个以上的其他的标靶(例如,与MR-CoNs关联的标靶)可以被用作其他的数据,其他的数据可以被另外用于分类未知样品。

[0087] 使用中间值的实施例

[0088] 本发明的另一个实施例可以指向用于产生分析模型的方法,该分析模型将生物学状态与其它生物学状态区分开来。该方法使用至少一个中间值。可以参考图5描述这种实施例。

[0089] 在步骤5000,选择量的已知样品经受暴露样品中的细菌的核酸的条件。步骤5000的细节可以与关于图1中的步骤1000描述的那些相同或不同,并且以上描述不需要被重复。

[0090] 在步骤5010,对于每个已知样品,测量诸如mecA、femA和SCCmec的至少三个标靶的特性。步骤5010的细节可以与以上关于图1中的步骤1010描述的那些类似或不同,并且以上描述不需要被重复。然而,注意,在这个实例中,SCCmec代替orfX被识别作为标靶。

[0091] 在确定输入值之后,它们被输入到数字计算机中。在一些实施例中,输入值包含进入数字计算机中的与已知样品关联的至少与第一遗传因子(例如,诸如mecA的第一标靶)关联的多个第一输入值、与第二遗传因子(例如,诸如femA的第二标靶)关联的多个第二输入值、以及与第三遗传因子(例如,诸如SCCmec的第三标靶)关联的多个第三输入值。第一第二和第三输入值可以在任何适当的时间以及以任何适当的顺序被输入到数字计算机中。

[0092] 在步骤5020,调用算法可以被应用于来自每个已知样品的每个测量的标靶。在这个实例中的调用算法与以上关于图1描述的调用算法不同。

[0093] 在步骤5030,可以使用数字计算机使用至少多个第一输入值以及至少与第二遗传因子关联的第二输入值,确定一个以上的中间值。例如,在一些实施例中,与第一遗传因子(mecA)和第二遗传因子(femA)关联的第一和第二输入值(例如,第一和第二Ct值)可以用来产生一个以上的中间值,一个以上的中间值可以和与第三遗传因子(例如,SCCmec)关联的第三输入值结合。

[0094] 如步骤5040中所示的,在产生一个以上的中间值之后,使用数字计算机,使用一个以上的中间值和多个第三输入值,产生用于生物学状态的边界空间。边界空间将生物学状态与其它生物学状态区分开来。

[0095] 说明性地,对于3标靶的“策略C”算法,199个进行的样本被收集和提供,用于算法开发。在策略“C”执行中,三个标靶是mecA、femA和SCCmec。

[0096] 在图6A和6B中显示mecA对比SCCmec和femA对比SCCmec的二维特征空间。可以看出,在两个二维特征空间中,有灰色区域,在灰色区域中,阳性和阴性被混合而且不能被区分(由黑色圆圈描画轮廓的)。

[0097] 为了克服该灰色区域,新的参数能够用于将事件(event)区分为阳性或阴性。这个新参数被建立,如:

[0098] 新参数= $\text{mecA} * \sin(-0.3) + \text{femA} * \cos(-0.3)$

[0099] “新参数”是中间值的实例,因为它来源于与第一遗传因子(例如,mecA)关联的第一输入值以及第二遗传因子(例如,femA)关联的第二输入值。

[0100] 图7显示了新参数如何可以被产生和使用的图解。图7显示与mecA 810和femA 820关联的值可以被组合以形成中间值Y 840。这个中间值Y 840和SCCmec 830可以形成二维的特征空间,二维的特征空间能够用于限定边界空间,边界空间用于将未知样品分类作为与MRSA或非MRSA关联。

[0101] 在图8中提供新参数对比SCCmec特征空间的图表。在这个说明中,灰色消失,而且新参数与SCCmec一起构成更好的特征空间用于分类。

[0102] 给定“新参数”和SCCmec,椭圆形的模型(如图9中所示)被建立,以限定阳性和阴性数据点之间的边界。可以以与先前描述的同样的方式,利用遗传算法最优化该椭圆形的模型。(例如,任何先前描述的边界形成技术,诸如以上步骤1060中的那些,可被用于本发明的

实施例。)位于椭圆之内的所有的数据点被认为是MRSA阳性。

[0103] 图10说明根据一个实施例的能够用于判定MRSA是否存在于样品中的步骤。如在此使用的,未知样品指的是不知道MRSA是否存在于其中的样品。该方法可以使用分析模型,诸如使用图5中步骤产生的模型,以判定未知样品是否包含MRSA(或与另一个生物学状态关联)。未知样品中的各种标靶的特性可以被测量以形成第一、第二和第三输入值。使用输入值和分析模型,可以检测未知样品中的MRSA的存在。

[0104] 参考图10,在步骤6000,未知样品经受暴露样品中的细菌的核酸的条件。在图1中的步骤1000期间使用的相同的技术可被用于步骤6000。

[0105] 在步骤6010,从未知样品中,测量至少三个标靶,mecA和SCCmec的特性。对于未知样品,可以确定第一第二和第三输入值。以上关于步骤1010描述的相同的技术能够用于实现步骤6010。

[0106] 在步骤6020,分析模型和第一第二和第三输入值能够用于判定特殊的生物学状态(例如,MRSA或非MRSA)是否存在于未知样品中。

[0107] 在步骤6030,当将分析模型应用到第一、第二和第三输入值时,可以使用第一和第二输入值确定中间值。例如,以下等式可以用来确定中间值:

[0108] $Y = \text{新参数} = \text{mecA} * \sin(-0.3) + \text{femA} * \cos(-0.3)$

[0109] 如说明的,假设样品具有mecA=28.49、femA=27.90、以及SCCmec=27.26, $Y = 18.2345$ 。

[0110] 在步骤6040,在该方法中,通过使用分析模型中的边界函数,判定样品是否对于MRSA呈阳性。例如,在以上实例中,值 $Y = 18.2345$ 以及SCCmec=27.26可以与图9中的椭圆形的边界函数相比。因为这个实例将落在边界函数之内,所以未知样品将被分类为MRSA阳性。

[0111] 系统

[0112] 图11显示系统,该系统包含数字计算机300以及操作地结合(可以包含电子结合)到数字计算机300的测量模块301。

[0113] 在这个实例中,数字计算机300可以包含各种典型的计算机部件,各种典型的计算机部件包含操作地结合在一起的系统总线304、一个以上的磁盘驱动器305、RAM 306以及处理器307。根据本实施例的准确的特性,还可以存在其它部件。图11还显示显示器308、键盘302以及鼠标303。一些实施例中,这些部件及其他部件还可以是数字计算机的一部分。

[0114] 系统还可以具有用于测量样品(例如,已知的或未知的)中的选择标靶的特性的测量模块301。根据被选择用于测量标靶响应的测量方法,这个测量模块可以在本发明的不同的实施例之间变化。例如,根据一个实施例,测量模块可以实施关于样品的PCR分析,并且因此可以是实时PCR设备。实时PCR设备在市场上是可买到的。

[0115] 在本发明的一个实施例中,样品被放置在测量模块301中,在测量模块301中,样品被处理,而且来自样品的选择标靶的特性(例如,数量)被测量。然后,这个数据(例如,先前描述的输入值)沿着系统总线304被传送到数字计算机300中,而且使用处理器307,适当的调用算法或分析模型可以被应用于响应数据。指令使得处理器307执行调用算法或分析模型(如以上描述的),该指令可以被存储在诸如RAM 306或磁盘驱动器305的计算机可读介质上。代表调用算法和/或分析模型的数据也可以被存储在这个相同的介质上。然后,来自调用算法或分析模型的应用的输出可以被显示在显示器308或其它输出装置(例如,打印机)

上。例如,先前描述的边界函数以及它们的关联的图表可以被显示在显示器308上或以其它的方式被输出。如此,然后来自测量的样品的信息被用于帮助构造模型或判定样品是否包含MRSA。

[0116] 如上所述,在一些实施例中,计算机可读介质可以存储或包含代码,可以通过处理器执行该代码以实现用于形成分析模型的方法。在一个实施例中,该方法可以包括:将至少与第一遗传因子关联的多个第一输入值、与第二遗传因子关联的多个第二输入值以及与第三遗传因子关联的多个第三输入值输入到与多个已知样品关联的数字计算机中,其中,每个已知样品包括多个第一输入值中的一个第一输入值、多个第二输入值中的一个第二输入值以及多个第三输入值中的一个第三输入值;判定与第三遗传因子关联的阈值;使用阈值将已知样品分离成为第一组已知样品和第二组已知样品;在由第一遗传因子和第二遗传因子限定的特征空间中聚类第一组已知样品;使用第一组已知样品限定第一边界空间,其中第一边界空间将生物学状态与其它生物学状态区分开来;以及使用第二组已知样品限定第二边界空间,其中第二边界空间将生物学状态与其它生物学状态区分开来。在另一个实例中,该方法可以包含将至少与第一遗传因子关联的多个第一输入值、与第二遗传因子关联的多个第二输入值以及与第三遗传因子关联的多个第三输入值输入到数字计算机中;使用数字计算机,使用至少多个第一输入值以及至少与第二遗传因子关联的第二输入值,产生一个以上的中间值;以及使用数字计算机,使用一个以上的中间值以及多个第三输入值产生用于生物学状态的边界空间,其中,边界空间将生物学状态与其它生物学状态区分开来。

[0117] 如上所述,在一些实施例中,计算机可读介质可以存储或包含代码,可以通过处理器执行该代码以实现用于使用分析模型的方法。该方法可以包含将与未知样品关联的第一输入值、第二输入值以及第三输入值输入到数字计算机或其它数字计算机中,其中,第一输入值、第二输入值以及第三输入值与未知样品中的第一、第二和第三遗传因子关联;以及使用数字计算机或其它数字计算机,使用第一边界空间或第二边界空间,将未知样品分类为与生物学状态关联。在另一个实施例中,该方法可以包括:将与未知样品关联的第一输入值、第二输入值以及第三输入值输入到数字计算机或其它数字计算机中,其中,第一输入值、第二输入值以及第三输入值与未知样品中的第一、第二和第三遗传因子关联;以及使用数字计算机或另一个数字计算机,使用第一边界空间或第二边界空间,将未知样品分类为与生物学状态关联。

[0118] 实例

[0119] 199个试验样品被标签为“Long Beach数据收集”。在这个数据收集中,测试以下靶组合:1)orfX、mecA和femA,2)mecA、femA和SCCmec,以及3)mecA、femA、SCCmec和MR-Cons。对于这个数据开发的调用算法是基于2维椭圆形的数学模型。在一个实例中,该模型产生中间值。通过中间值上阈值确定MRSA分类结果。该数学模型用公式表示为

$$[0120] \quad \frac{[-x \cdot \cos(\theta) + y \cdot \sin(\theta) - x_0]^2}{a^2} + \frac{[x \cdot \sin(\theta) + y \cdot \cos(\theta) - y_0]^2}{b^2} = S,$$

[0121] 其中,S在中间值中,并且x和y是对这个模型的两个输入。对于orfX、femA和mecA的数据收集,x和y分别是mecA和femA。对于SCCmec、femA和mecA的数据收集,x和y分别是SCCmec和 $Y=f(\text{mecA}, \text{femA})$ 。具体地说, $Y=f(\text{mecA}, \text{femA})=\text{mecA} \cdot \sin(-0.3) + \text{femA} \cdot \cos(-0.3)$,其中,-0.3是弧度。 x_0 、 y_0 、a、b和 θ 是预定参数,通过利用具有给定标准的遗传算法训练这个

模型而获得预定参数。给定 x_0 、 y_0 、 a 、 b 和 θ ，每一对 x 和 y 将产生一个 S 。小的 S 意味着 (x, y) 靠近 (x_0, y_0) ，反之亦然。为了分类目的，理想的是选择 S 的阈值（例如， S_0 ）。如果样品产生小于 S_0 的中间值，那么这个样品被认为是MRSA阳性。这些调用算法是基于参数化的数学模型，而且利用遗传算法被训练，以达到最优性能，并且通过中间值上的阈值生成分类结果。

[0122] 样品制备

[0123] 199个Nasal药签被收集并且存储在斯图尔特(stuart)传送介质中。药签头被去除，而且每个药签头被转移到2ml的样品悬浮液试管中，该样品悬浮液试管具有基于样品缓冲剂的1200 μ L的Tris，样品缓冲剂具有10mM Tris pH8.0以及1mM EDTA，pH8.0~100mg的1mm氧化锆/硅质颗粒。通过以3000rpm的速度涡旋样品试管至少15秒，药签头上的细菌被逐出。

[0124] 然后药签头从样品试管被无菌地去除，并且被传送到具有1ml的胰酶大豆肉汤(Trypic Soy broth)TSB以及6.5%NaCl的15ml细菌培养试管中。接种过的细菌试管被传送到37°C的孵育箱中，并且随着以200rpm的速度摇动，被整夜培育。

[0125] 然后确定金黄色葡萄球菌和/或MRSA的存在或不存在。10 μ L的每一个整夜培养基在BBL™ CHROMagar MRSA以及BBL™ CHROMagar金黄色葡萄球菌板上被形成条纹。的来自每个试管的1200 μ L的样品溶液中的500 μ L然后经受如Agencourt VirNA试剂盒方案描述的DNA隔离过程。这个过程，简单地说，伴随200个CFU S. 猫属细菌作为处理控制开始。10单位的消色肽酶(Achrompeptidase)被添加到每个试管，很好地混合，并且在70°C的水浴槽中培育4分钟。然后，包含188 μ L的溶菌缓冲剂、1.0L的PolyA(600g/ml)以及100 μ L的蛋白酶K(6.4mg/ml)的新鲜制备的溶菌溶液289 μ L被加到每个样品中并且被很好地混合。然后每个样品在70°C被培育1分钟，然后被允许冷却2分钟。然后，10 μ L的磁性颗粒和575 μ L的100%异丙醇被添加，并且通过涡旋被很好地混合。反应含量被允许在室温下培育5分钟，然后，通过将样品试管放置在磁体架子上6分钟来收集磁性颗粒，从而将磁性颗粒与样品溶液分离，直到溶液变得清澈。

[0126] 接下来，上清液被吸出，同时注意在吸出期间不去除任何颗粒。500 μ L的清洗缓冲剂(3.3M氨基甲脒硫氰酸盐(Guanidine Thiocyanate)，1.7%氟核(Triton)X-100，167.5mM柠檬酸钠)被加到样品中，并且被涡旋10秒以混合。然后，试管在磁体上被培育4分钟(或直到清澈)。然后上清液被再次吸出样品。然后900 μ L的75%的新鲜制备乙醇被添加并且被涡旋10秒。然后试管在磁体上被培育4分钟，直到清澈。然后上清液再次被吸出样品，并且再一次重复乙醇清洗。然后颗粒在磁体上被干燥15-25分钟。当颗粒的环开始破裂时，样品被洗提。试管离开磁体，并且25 μ L的核酸酶自由水被添加。然后样品被涡旋以便混合。然后试管在70°C被培育5分钟。试管被放回在磁体上并且被培育1分钟。然后上清液被传送到干净的试管用于PCR扩增。

[0127] PCR引物和探针，PCR循环条件

[0128] 列举在Master混合表中的试剂制备就绪。根据总的反应数，可以通过简单地将标明容量的试剂在DNA/RNA/RNase-自由试管中添加在一起，制备充足的Master混合物。试管可以被涡旋以便混合，然后保持就绪用于以后的使用。20 μ L的每个洗提液被添加到Mx3000P 96孔PCR板(非边缘的)(Stratagene, Cat#401333)(一个洗提液，一个孔)。30 μ L的Master混合物被添加到填充有洗提液的每个孔中，然后通过轻轻地上上下下吸8次以上(多通道可能

是有用的)而被混合。该板利用MicroAmp™光学粘合剂膜(采用的生物系统)被紧密地覆盖,然后在流入PCR机器之前,以1100xg被离心3分钟。

[0129] Stratagene MX 3005P仪器上的PCR循环条件设置如下:4'@37°C(1x);1分钟@95°C(1x);15秒@95°C→10秒@62°C→30秒@58°C(40x)。监视的标靶被表示在以下的表1中:

标靶	函数	Oligo ID	序列
orfX-ISS	正向引物	OrfX-ISS/attBScc for-1	TGAGGGTTGTGTTAATTGAGCAAGTG
	正向引物	OrfX-ISS/attBScc for-2	TGCGGGTTGTGTTAATTGAACAAGTG
	反向引物	mecII512-1-sccmec- 3	TCACTTTTATTCTTCAAAGATTTGAGC
	反向引物	primer11-1-sccmec- 7	AAATTGCTACTAAAGAGGATATGGAAAACCATC
	反向引物	primer12-sccmec-8	CTCTGCTTTATATTATAAAATTACGGCTG
	反向引物	newtypeiii-1- sccmec-14	CGTATGATATTGCAAGGTATAATCCAATATTTTC
	反向引物	typeIVc-sccmec-2	CTTGAAATGAAAGACTGCGGAGGCTAAC
	反向引物	NEWPRIMERS	TGAGCTTTTTTCCACTCCCATTCTCTCCAAA
	反向引物	SCCmec-4nV	GCAATTCACATAAACCTCATATGTCTGATAC
	反向引物	SCCmec-3n	CATTCAATCATCCACCCTAAACTTAATCTTTC
	反向引物	SCCmec-5n	TATGGAAATCCATCTCTACTTTATTGTTTTCTTC
	反向引物	SCCmec-6n	AATATTTTCATATATGTAAATTCCTCCACATCTC
	反向引物	SE-7-11	CTATTTCTGTAATACTTAAAAACCTTTTCTTCC
	反向引物	SE-17	CCGTATGATTCATATTAATAATGAATCATACGGAGG
	反向引物	SE-13	CTTCTTATGAAATGTCTTTTTTCACTTATCC
	TaqMan 探针	orfX 探针-2	ATGCTTCTCCTCGCATAATCTTAAAYGCTC
	TaqMan 探针	ORFX 探针-1	ACGCTTCTCCACGCATAATCTTAAATGCTC
	TaqMan 探针	ORFX 探针	ACGCCTCTCCTCGCATAATCTTAAATGCTC
femA	正向引物	femA-3 正向 引物	GACCGTTATAATTTCTATGGTGTTAGTGG
	反向引物	femA-3 反向 引物	GTCACCAACATATTCAATAATTTGAGC
	TaqMan 探针	femA-sa- 探针	ACAGAAGATGCTGAAGATGCTGGTGT

[0130]

mecA	正向引物	mecA-2 正向引物	GCAGAAAGACCAAAGCATACATATTGA
	反向引物	MecA-2 反向引物	GCCTATCTCATATGCTGTTCTGT
	TaqMan 探针	mecA 探针	AGACCGAAACAATGTGGAATTGGCCA
S.felis (IC)	IC 正向引物	Sfforwardnew	TGCCAATGTAGATAGTCTTCCAGA
	IC 反向引物	sffreversenew	AAGTGCCCAAGAAGAATGAGTGG
	IC 探针	fSfelis	ACCGCCACCATTATTACGTACAGCTG
SCCmec	正向引物	OrfX-ISS/attBSec for-1	TGAGGGTTGTGTTAATTGAGCAAGTG
	正向引物	OrfX-ISS/attBSec for-2	TGCGGGTTGTGTTAATTGAACAAGTG
	反向引物	mecII512-1-sccmec-3	TCACTTTTATTCTTCAAAGATTTGAGC
	反向引物	primer11-1-sccmec-7	AAATTGCTACTAAAGAGGATATGGAAAACCATC
	反向引物	primer12-sccmec-8	CTCTGCTTTATATTATAAAATTACGGCTG
	反向引物	newtypeiii-1-sccmec-14	CGTATGATATTGCAAGGTATAATCCAATATTTTC
	反向引物	typeIVc-sccmec-2	CTTGAAATGAAAGACTGCGGAGGCTAAC
	反向引物	NEWPRIMERS	TGAGCTTTTTCCACTCCCATTTCTTCCAAA
	反向引物	SCCmec-4nV	GCAATTCACATAAACCTCATATGTTCTGATAC
	反向引物	SCCmec-3n	CATTCAATCATCCACCCTAAACTTAATCTTTC
	反向引物	SCCmec-5n	TATGGAAATCCATCTCTACTTTATTGTTTTCTTC
	反向引物	SCCmec-6n	AATATTTTATATATGTAATTCCTCCACATCTC
	反向引物	SE-7-11	CTATTTCTGTAATACTTAAAACCTTTTCTTCC
	反向引物	SE-17	CCGTATGATTCATATTAAAATGAATCATACGGAGG
	反向引物	SE-13	CTTCTTATGAAATGTCCTTTTTCACCTTATCC
	TaqMan 探针	orfx 探针-2	ATGCTTCTCCTCGCATAATCTTAAAYGCTC
	TaqMan 探针	ORFX 探针-1	ACGCTTCTCCACGCATAATCTTAAATGCTC
	TaqMan 探针	ORFX 探针	ACGCCTCTCCTCGCATAATCTTAAATGCTC

[0131]

[0132] 在以上的表中,具有对应于SCCmec的序列的探针与SCCmec的右末端互补。

[0133] 以下表显示使用Long Beach数据集中的199个样品并且使用根据本发明的实施例的分析模型(如以上描述的)以及使用其它调用算法的分析模型产生的假阳性和假阴性数据。

[0134]

富集培养数据1 (59个阳性, 140个阴性)				
	FP	FN		
1: 策略A, 3个靶标聚类(mecA, femA, orfX)	10	2		DxN调用算法
2: 策略C, 4个靶标聚类(mecA, femA, SCCmec (MRSA), SCCmec (MR-ConS))	3	2		DxN调用算法
3: 策略C, 3个靶标聚类(mecA, femA, SCCmec)	3	4		DxN调用算法
4: Xpert - 1个靶标 (SCCmec)	11	2		
5: 策略C, 1个靶标 (SCCmec)	11	4		公开的调用算法
6: 策略C, 2个靶标 (SCCmec, mecA)	10	5		公开的调用算法
7: 策略C (SCCmec, mecA) Ct 比较 (± 4 ct)	5	7		公开的调用算法
8: 策略C, 2个靶标 (SCCmec, femA)	3	5		公开的调用算法
9: 策略C (SCCmec, femA) Ct 比较	3	5		公开的调用算法
10: 策略C, 3个靶标 (SCCmec, mecA 和 femA)	2	5		公开的调用算法
11: 策略A, 2个靶标 (femA 和 mecA)	38	1		公开的调用算法
12: 策略A, 2个靶标 (orfX - ve) Ct 比较 (± 4 ct)	24	5		公开的调用算法
13: 策略A, 3个靶标 (mecA, femA, orfX)	18	25		公开的调用算法
14: 策略A, 3个靶标 (mecA, femA, orfX) Ct 比较 (± 4 ct)	17	5		公开的调用算法

[0135] 最初的三个分析模型(1到3)是根据本发明的实施例产生的那些,而且显示良好的结果。使用三个靶标聚类和以上描述的处理(例如,如图1中的)的题为策略A的第一分析模型得到10个假阳性和2个假阴性。使用策略C和4个靶标的第二分析模型得到3个假阳性和2个假阴性。使用策略C的第三分析模型得到3个假阳性和4个假阴性。以上描述了与策略A和策略C关联的方法(例如,如图5中的)。在4个靶标的策略C实例的情况下,MRConS被用作其他的靶标以将MRSA样品与非MRSA样品区分开来。

[0136] 余留的数据显示根据公开的调用算法被处理的数据。以下提供这种算法的一些细节。

[0137] 对于4:Xpert是已知的并且市场上可买到的测试。

[0138] 对于5:策略C,1个靶标;SCCmec是靶标。如果SCCmec Ct值小于32,那么MRSA存在。如果SCCmec Ct值大于32,那么没有MRSA。

[0139] 对于6:策略C,2个靶标(SCCmec, mecA):如果SCCmec和mecA两者都小于32,那么MRSA存在。如果SCCmec大于32或mecA大于32,那么没有MRSA。

[0140] 对于7:策略C(SCCmec, mecA)Ct比较(± 4 ct):如果SCCmec大于32或mecA大于32,那么没有MRSA。如果SCCmec和mecA两者的Ct值都小于32,而且SCCmec和mecA之间的德耳塔Ct小于4,那么存在MRSA。如果德耳塔Ct大于4,那么MRSA不存在。

[0141] 对于8:策略C,2个靶标(SCCmec, femA):如果SCCmec和mecA两者都小于32,那么MRSA存在。如果SCCmec大于32或femA大于32,那么没有MRSA。

[0142] 对于9:策略C(SCCmec, femA)Ct比较:如果SCCmec大于32或femA大于32,那么没有MRSA。如果SCCmec和femA两者的Ct值都小于32,而且SCCmec和femA之间的德耳塔Ct小于4,那么存在MRSA。如果德耳塔Ct大于4,那么MRSA不存在。

[0143] 对于10:策略C,3个标靶(SCCmec,mecA和femA):如果SCCmec、mecA、femA都大于32,那么MRSA存在。如果mecA大于32,那么没有MRSA。如果SCCmec大于32,那么没有MRSA存在。如果mecA和femA两者都小于32,而且SCCmec大于32,以及mecA和femA的德耳塔Ct值小于4,那么MRSA存在。

[0144] 对于11:策略A,2个标靶(femA和mecA):如果femA和mecA两者都小于32,那么MRSA存在。否则,没有MRSA。

[0145] 对于12:策略A,2个标靶(orfX-ve)Ct比较($\pm 4ct$):如果femA和mecA两者都小于32以及femA和mecA之间的德耳塔Ct小于4,那么MRSA存在。否则,没有MRSA。

[0146] 对于13:策略A,3个标靶(mecA,femA,orfX):如果femA和mecA两者都小于32以及orfX是阴性的,那么MRSA存在。否则,没有MRSA。

[0147] 对于14:策略A,3个标靶(mecA,femA,orfX)Ct比较($\pm 4ct$):如果femA和mecA两者都小于32以及orfX是阴性的,而且mecA和mecA之间的德耳塔Ct小于4,那么MRSA存在。否则,没有MRSA。

[0148] 在这个申请中描述的软件成分、步骤或函数可以被实现作为软件代码,通过一个以上的处理器使用任何适当的计算机语言,诸如,例如Java、C++或使用例如传统的或目标定向技术的Perl,来执行软件代码。软件代码可以被存储作为计算机可读介质上的一系列指令或命令,计算机可读介质诸如是随机存取存储器(RAM),只读存储器(ROM),诸如硬盘驱动器或软磁盘的磁性介质,或诸如CD-ROM的光学介质。任何这样的计算机可读介质还可以存在于单个计算的设备上或之内而且可以存在于系统或网络之内的不同的计算的设备上或之内。

[0149] 本发明的一些实施例可以作为软件或硬件或两者相结合的控制逻辑被实现。控制逻辑可以被存储在信息存储介质中,作为多个指令,多个指令适于引导信息处理装置以施行本发明的实施例中公开的一组步骤。基于在此提供的公开和教导,普通的本领域的技术人员将领会其它途径和/或方法来实现本发明。

[0150] 任何“一”、“一个”、“该”的叙述打算意指“一个以上”,除非具体地有相反的标明。

[0151] 以上描述是说明性的而不是限制性的。当查阅本公开时,本发明的许多变化对那些本领域的技术人员来说就变得显而易见了。因此,本发明的范围应当不是参考以上描述被确定,而是应当参考未决的权利要求书连通它们的全部范围或等效物一起被确定。

[0152] 为了所有的目的,以上提及的所有的专利、专利申请、公报和描述以其全部内容通过引用被结合在此。均没有被承认是现有技术。

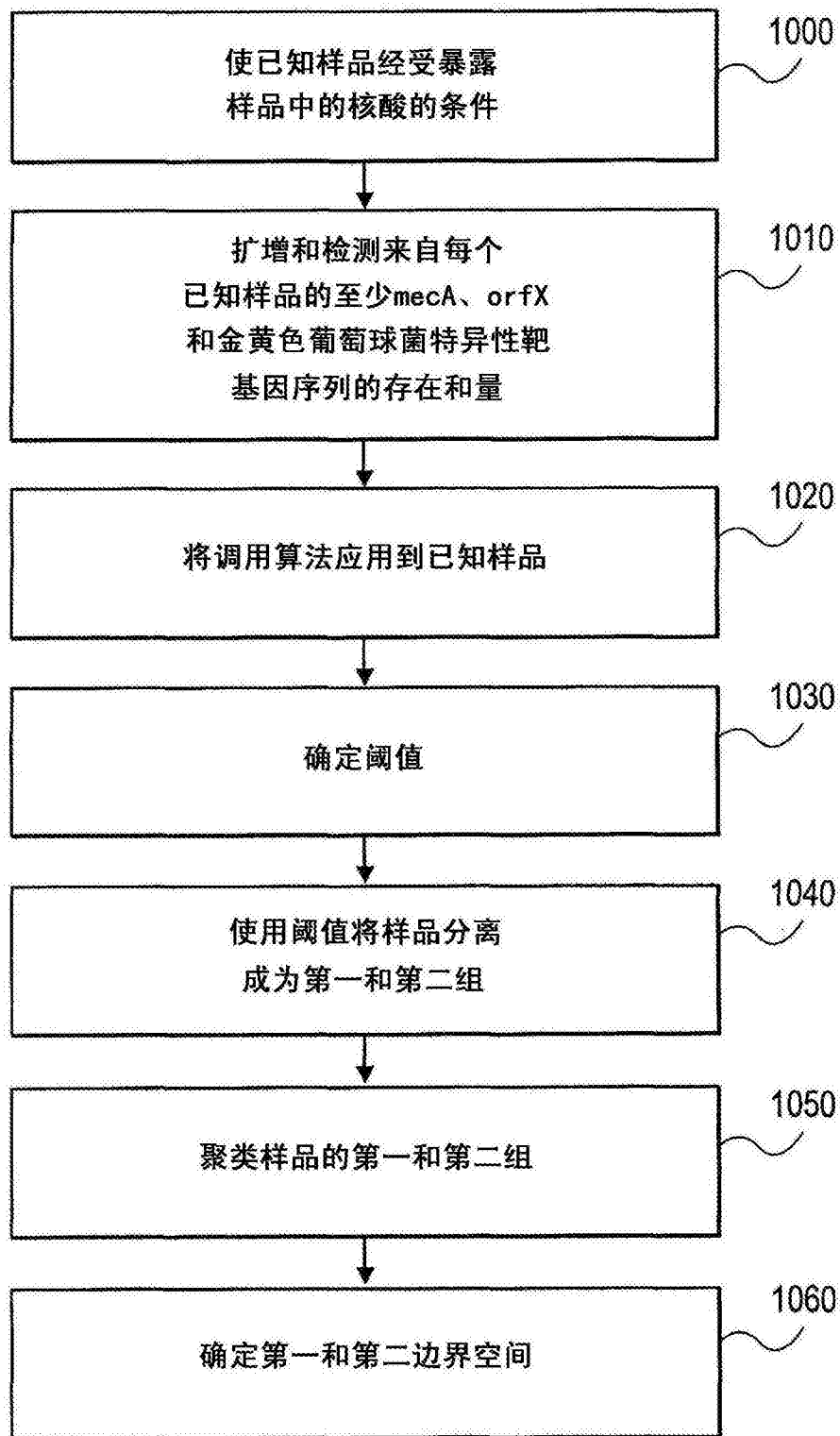


图1

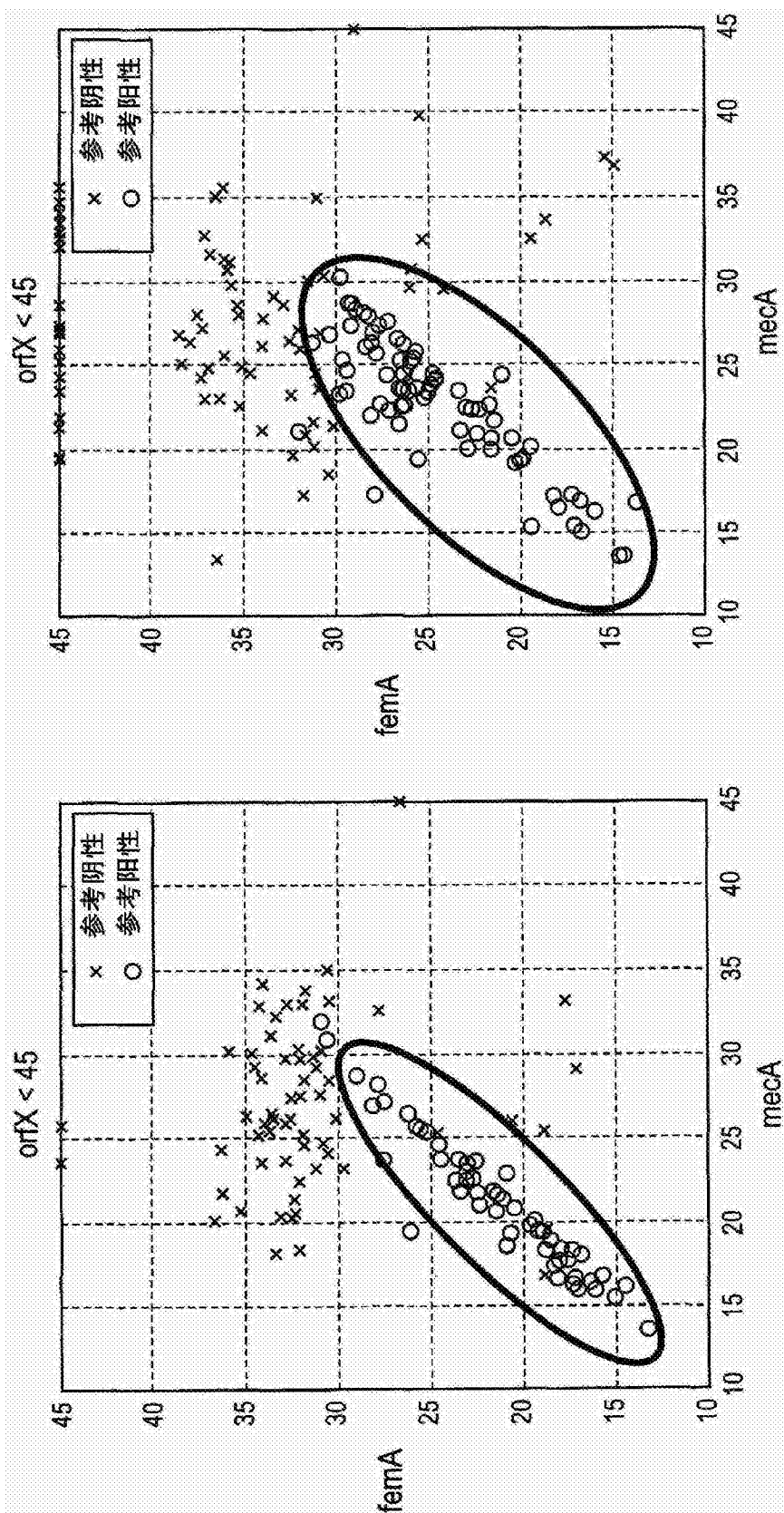


图2

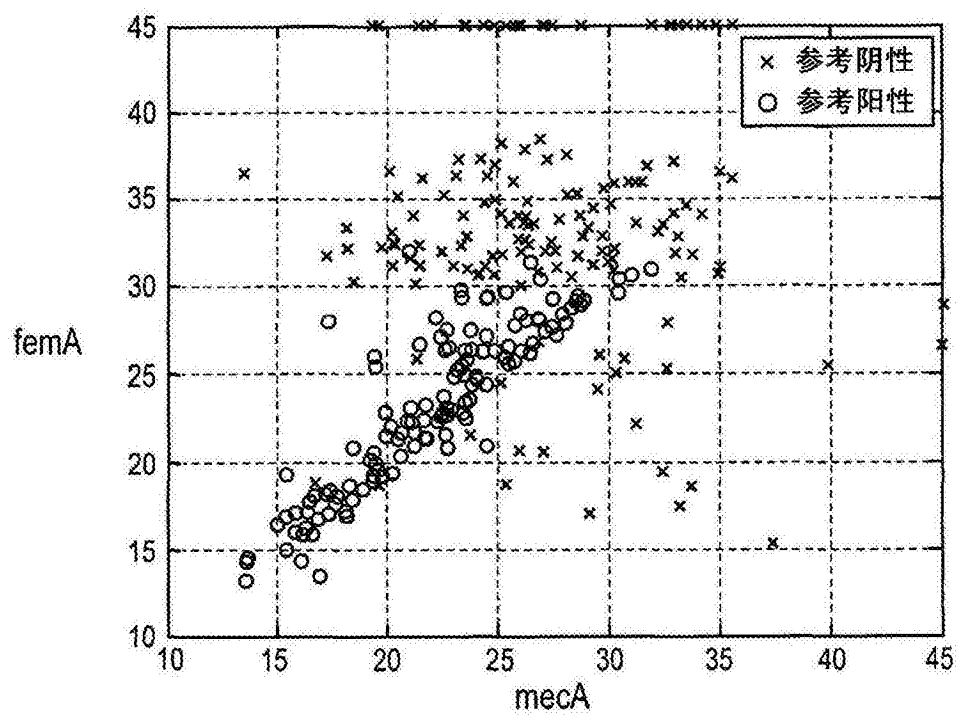


图3A

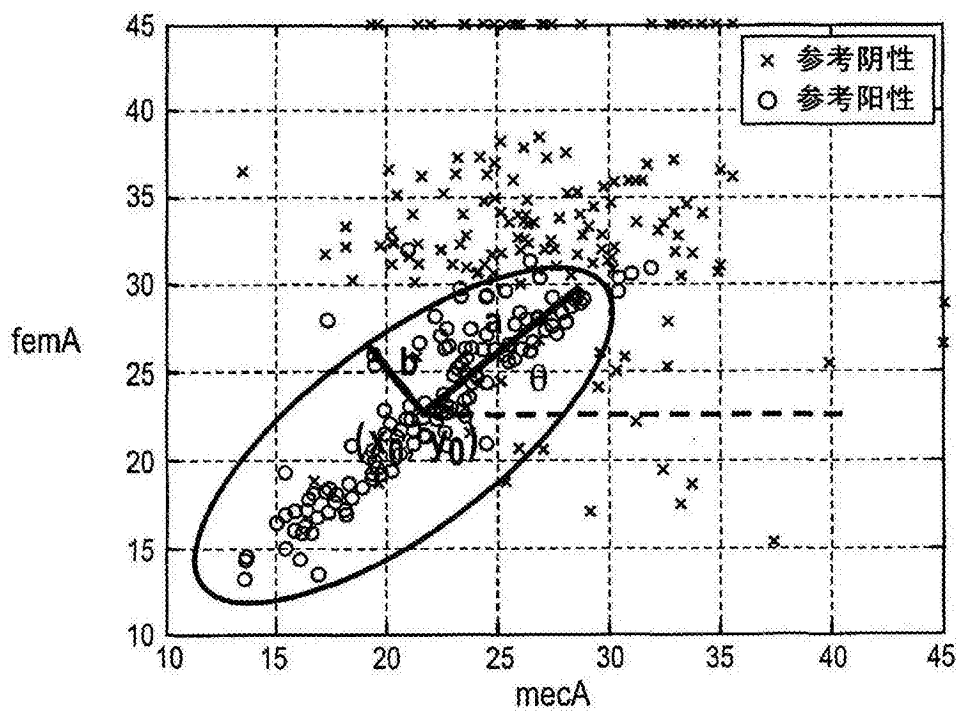


图3B

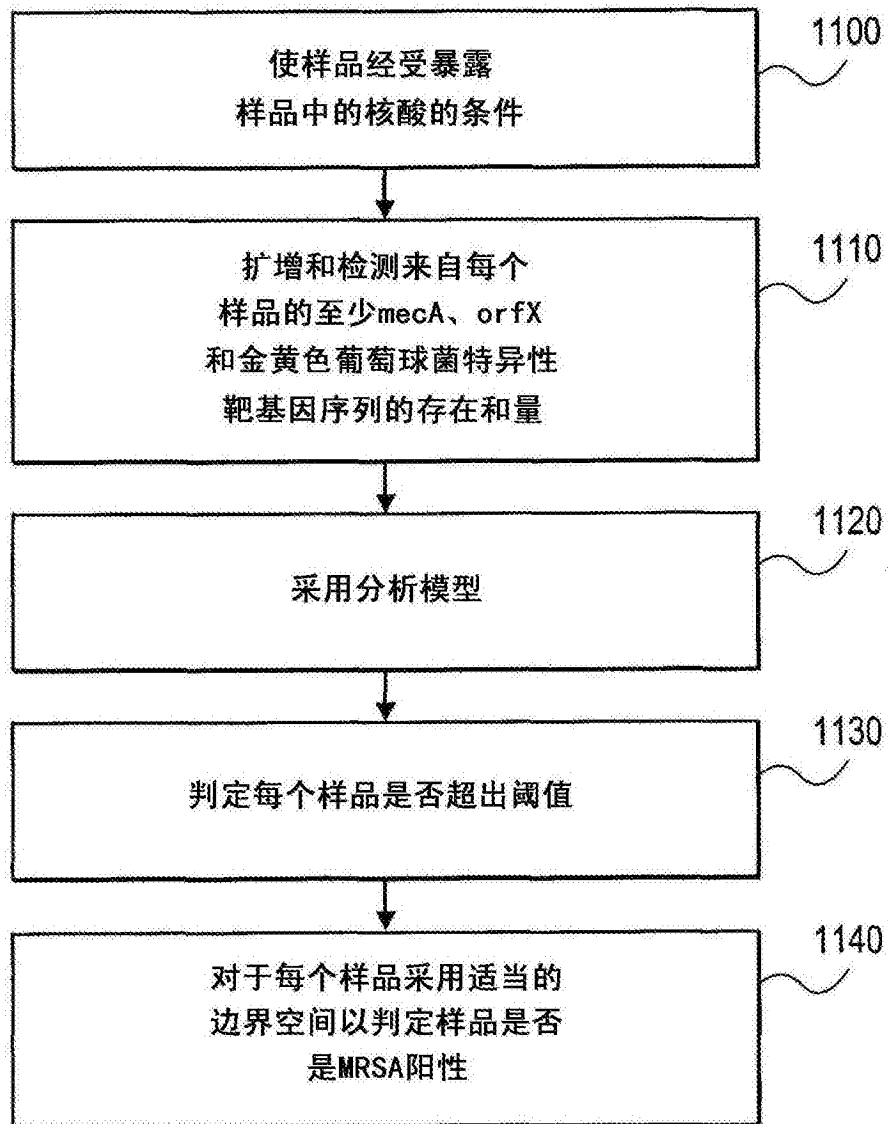


图4

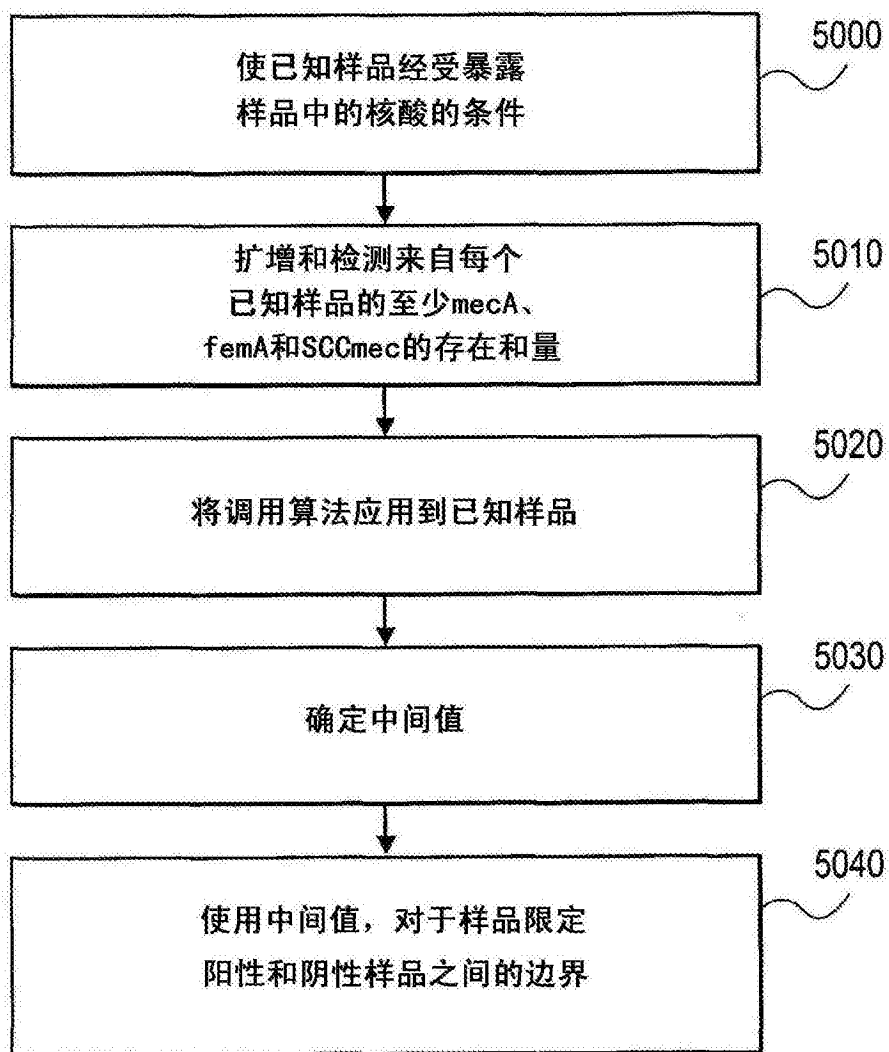


图5

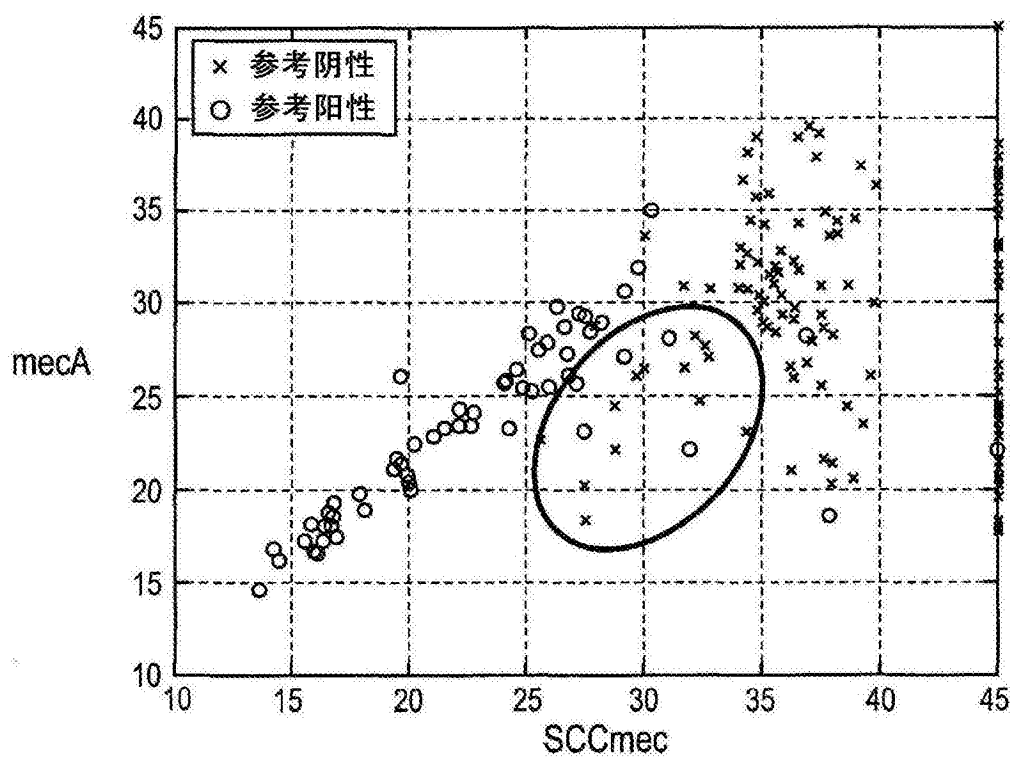


图6A

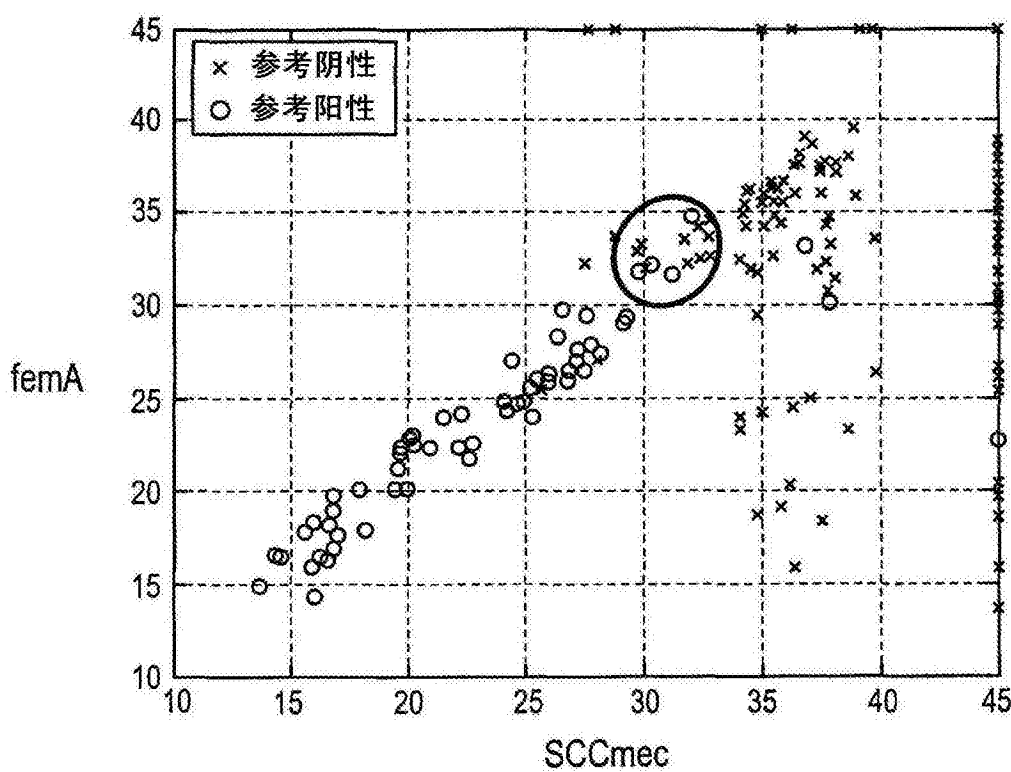


图6B

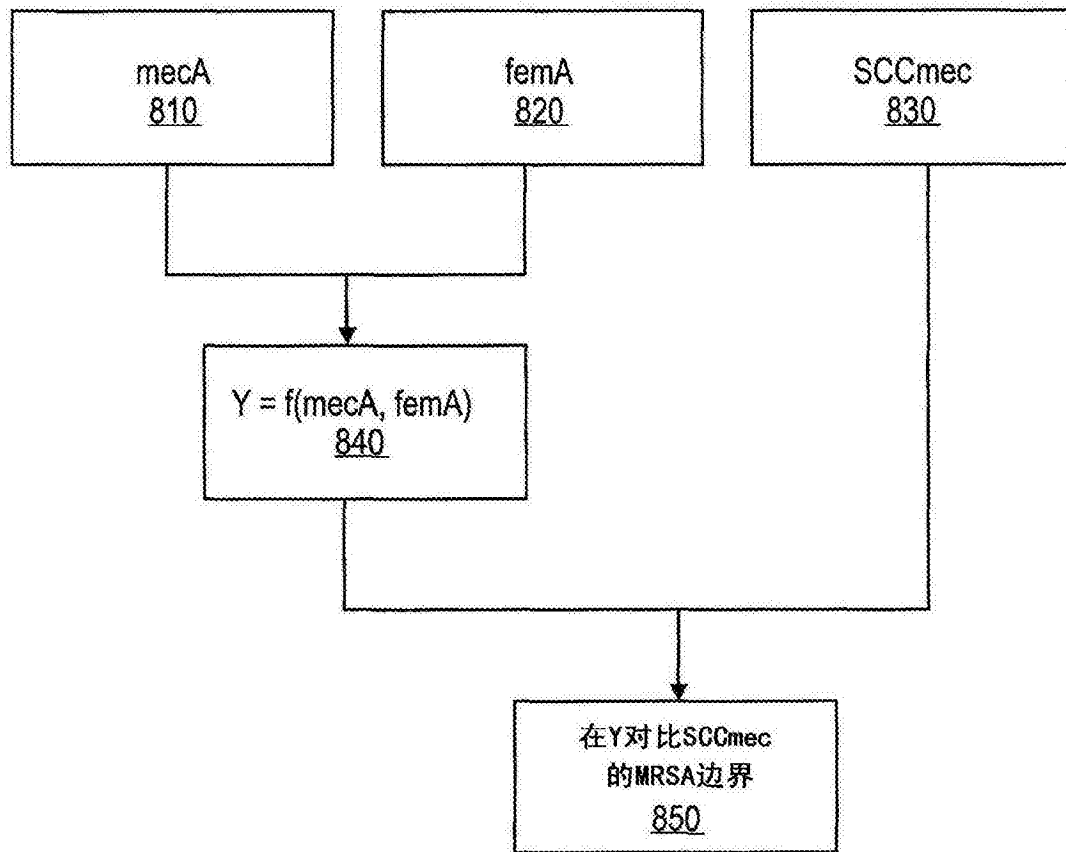


图7

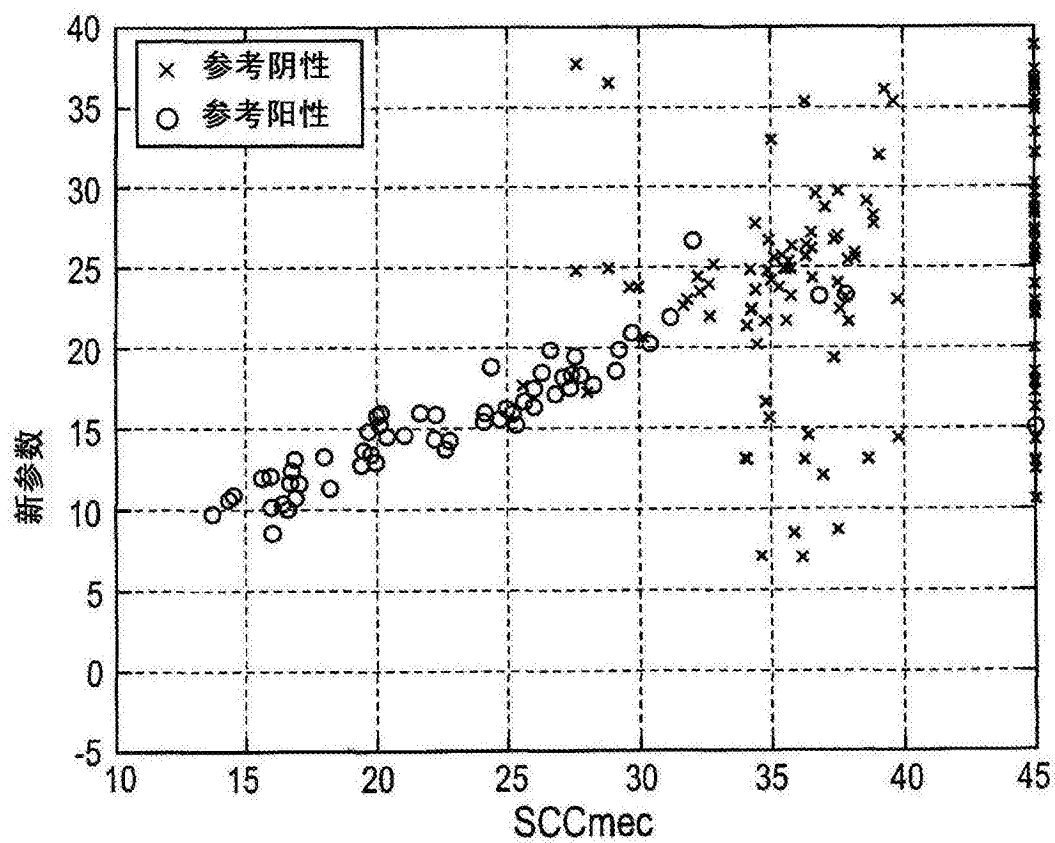


图8

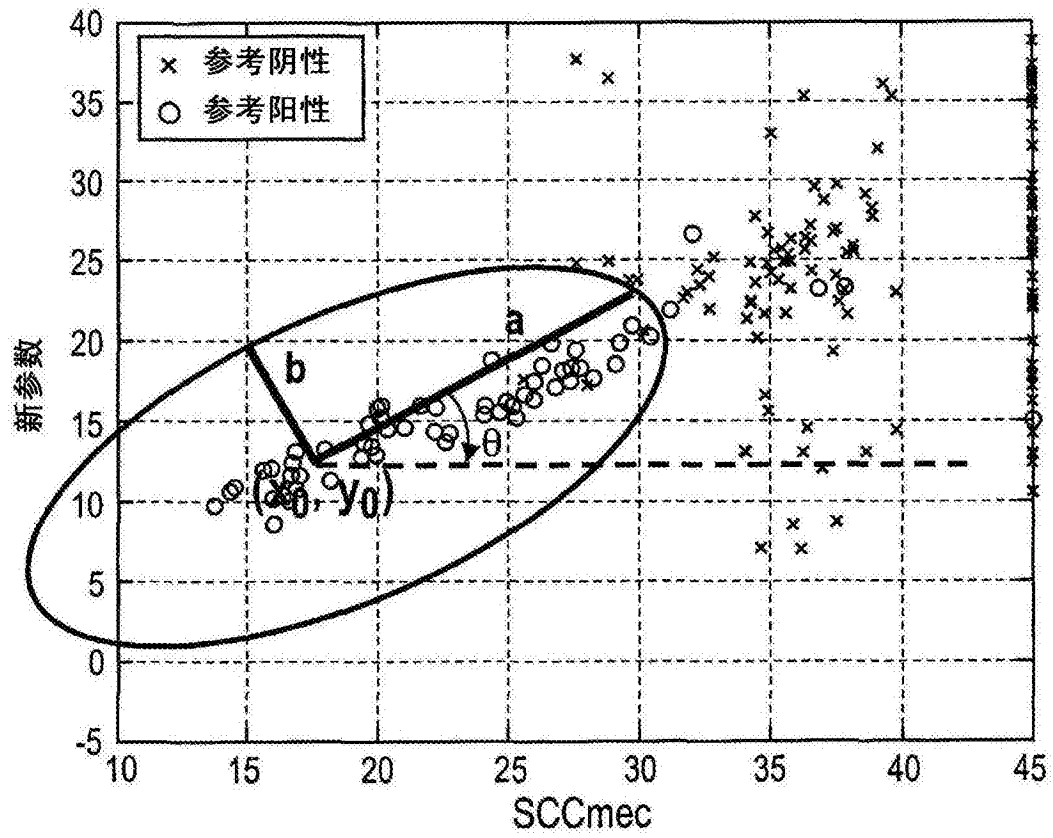


图9

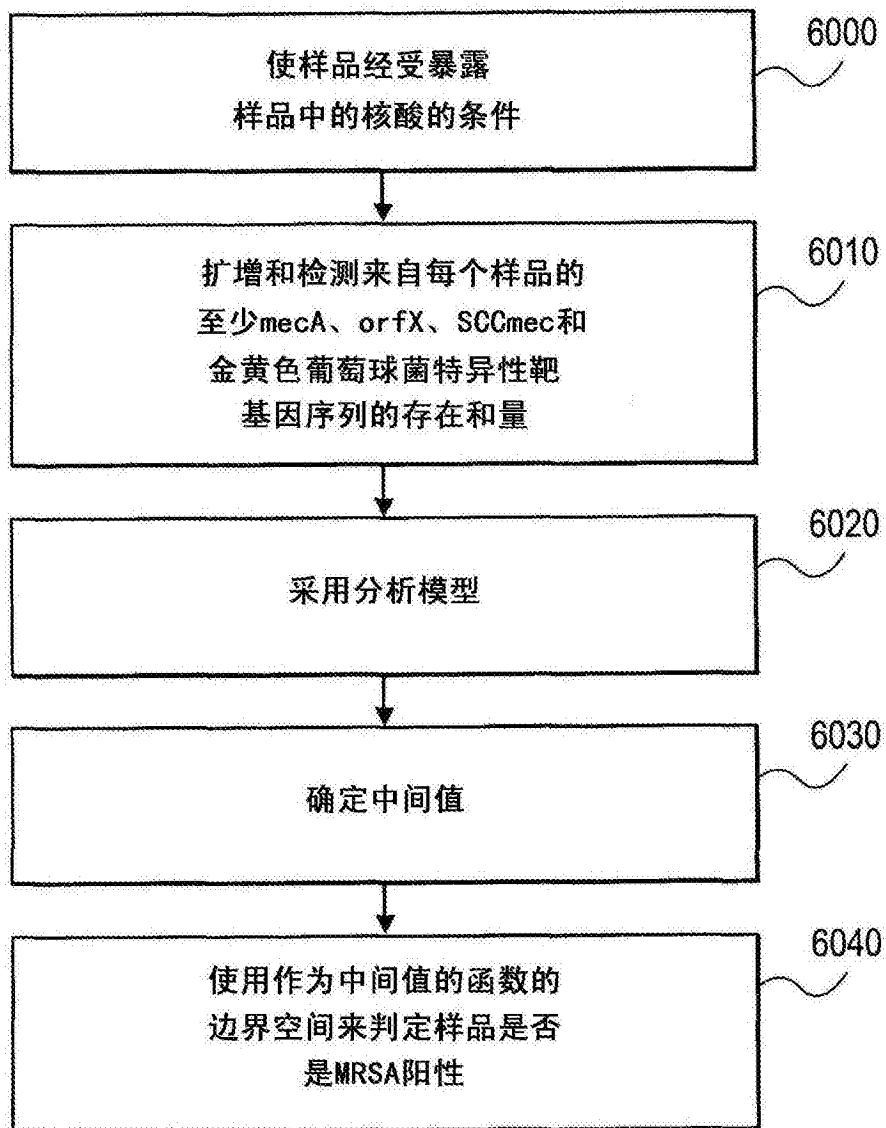


图10

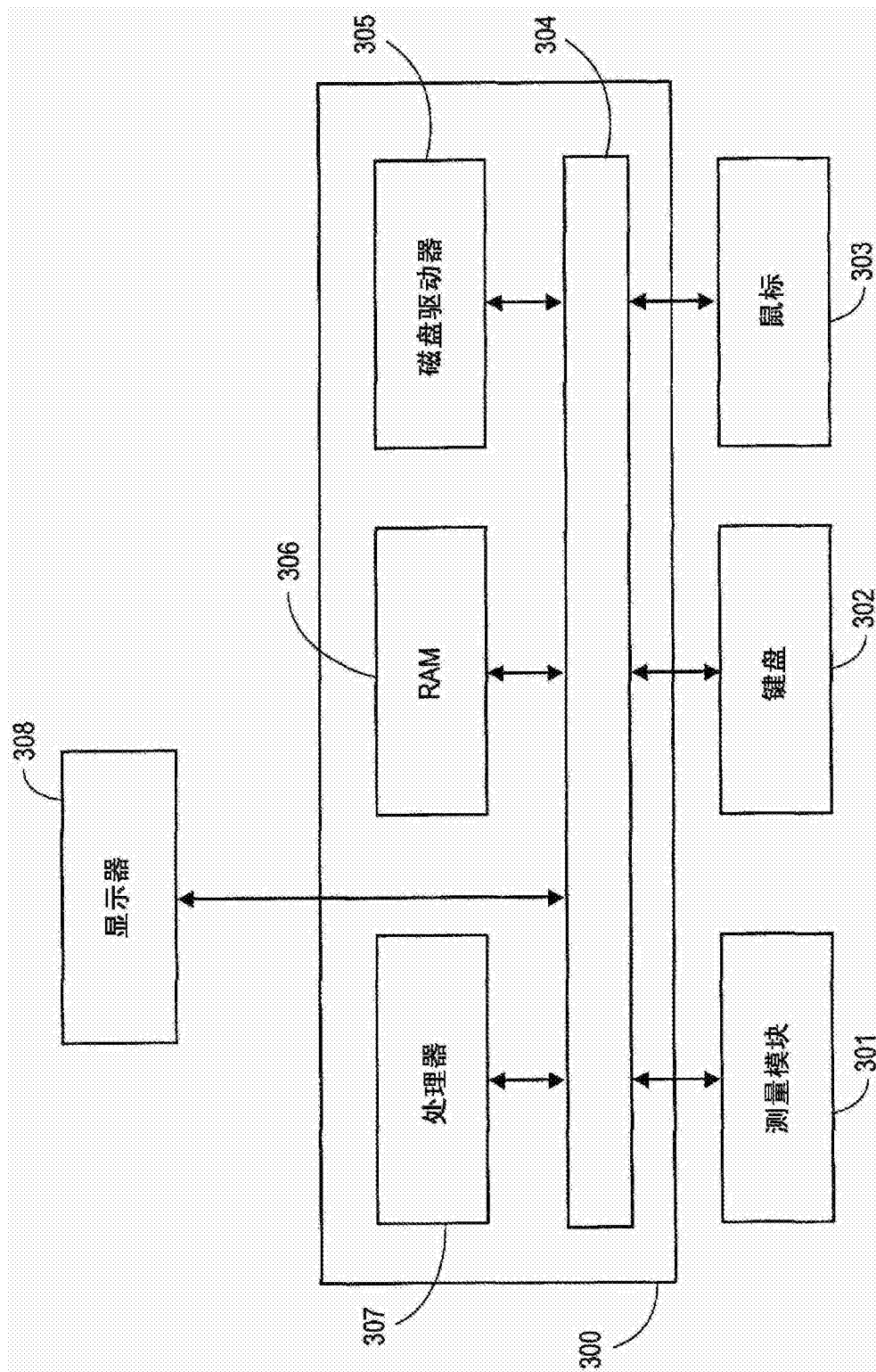


图11