



(12)发明专利

(10)授权公告号 CN 104777313 B

(45)授权公告日 2017.09.26

(21)申请号 201510123032.5

G01N 33/574(2006.01)

(22)申请日 2011.07.11

C12Q 1/68(2006.01)

(65)同一申请的已公布的文献号

申请公布号 CN 104777313 A

(43)申请公布日 2015.07.15

(30)优先权数据

61/363,122 2010.07.09 US

61/444,947 2011.02.21 US

(62)分案原申请数据

201180033907.2 2011.07.11

(73)专利权人 私募蛋白质体公司

地址 美国科罗拉多州

(72)发明人 S·威尔科克斯 D·艾尔斯

N·亚尼奇 L·戈尔德

M·里尔-米恩 T·贾维斯

(74)专利代理机构 永新专利商标代理有限公司

72002

代理人 王健

(51)Int.Cl.

G01N 33/68(2006.01)

(56)对比文件

WO 2008/063413 A2,2008.05.29,

WO 2010/030697 A1,2010.03.18,

WO 2008/063413 A2,2008.05.29,

Hans-Stefan Hofmann 等.Matrix

Metalloproteinase-12 Expression

Correlates with Local Recurrence and

Metastatic Disease in Non-Small Cell Lung

Cancer Patients.《Clinical Cancer

Research》.2005,第11卷(第3期),

Hans-Stefan Hofmann 等.Matrix

Metalloproteinase-12 Expression

Correlates with Local Recurrence and

Metastatic Disease in Non-Small Cell Lung

Cancer Patients.《Clinical Cancer

Research》.2005,第11卷(第3期),

审查员 段晓露

权利要求书1页 说明书104页 附图45页

(54)发明名称

肺癌生物标记及其用途

(57)摘要

本申请包括用于检测和诊断肺癌的生物标记、方法、装置、试剂、系统和试剂盒。在一方面,本申请提供了生物标记,其可以单独使用或以各种组合使用来诊断肺癌或允许区分诊断肺结节为良性或恶性的。在另一方面,本申请提供了诊断个体的肺癌的方法,其中所述方法包括在来自个体的生物学样品中检测至少一个生物标记值,所述至少一个生物标记值对应于选自表18、表20或表21提供的生物标记的组的至少一个生物标记,其中基于所述至少一个生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性。

1. 特异于生物标记蛋白的捕获试剂在制备试剂盒中的用途,所述试剂盒用于通过如下方法诊断个体患有或不患有非小细胞肺癌的似然性,所述方法包括:

提供由生物标记蛋白MMP-12以及至少一种选自表21的生物标记组成的生物标记组;以及

在来自个体的生物样品中检测生物标记蛋白,以给出生物标记值,所述生物标记值每个对应于该组中的生物标记蛋白,其中所述生物标记值提供所述个体患有或不患有非小细胞肺癌的似然性的指示,

其中所述生物标记组选自:(i) MMP-12和C9,(ii) MMP-12和MMP-7,(iii) MMP-12和ERBB1,(iv) MMP-12和SCF sR,(v) MMP-12、C9和SCF sR,以及(vi) MMP-12、C9和ERBB1。

2. 权利要求1的用途,其中检测所述生物标记值包括进行体外测定,所述体外测定包括对应于每个所述生物标记的至少一种捕获试剂,并且还包括从适配体、抗体和核酸探针的组选择所述至少一种捕获试剂。

3. 权利要求2的用途,其中所述至少一种捕获试剂为适配体。

4. 权利要求1-3中任一项的用途,其中所述生物样品为肺组织,并且其中所述生物标记值源自所述肺组织的组织学或细胞学分析。

5. 权利要求1-3中任一项的用途,其中所述生物样品选自全血、血浆和血清。

6. 权利要求5的用途,其中所述生物样品为血清。

7. 权利要求1-3中任一项的用途,其中所述个体为人。

8. 权利要求1-3中任一项的用途,其中所述个体为吸烟者或具有肺结节。

肺癌生物标记及其用途

[0001] 本申请是2011年7月11日提交的题为“肺癌生物标记及其用途”的中国专利申请201180033907.2的分案申请。

[0002] 相关申请

[0003] 本申请要求于2010年7月9日提交的美国临时申请系列号61/363,122以及于2011年2月21日提交的美国临时申请系列号61/444,947的权益,这两个临时申请的题目均为“肺癌生物标记及其用途”。为了所有目的,这些申请每个整体援引加入本文。

发明领域

[0004] 本申请一般涉及个体中生物标记的检测及癌症的诊断,更具体地涉及用于诊断个体的癌症,更特别是肺癌的一种或多种生物标记、方法、装置、试剂、系统和试剂盒。

背景技术

[0005] 下面的描述提供了本申请相关信息的概述,并非承认任何本文提供的信息或者引用的出版物是本申请的现有技术。

[0006] 肺癌是癌症相关死亡率的最常见的原因。这对于男性和女性都是如此。2005年在美国,肺癌死亡比乳腺癌、前列腺癌和结肠癌死亡的组合还多。在那一年,107,416名男性和89,271名女性诊断患有肺癌,并且90,139名男性和69,078名女性死于肺癌。在美国男性中,肺癌是白人男性、黑人男性、亚裔/太平洋岛男性、美国印第安男性/阿拉斯加土著男性以及西班牙裔男性中第二常见的癌症。在美国女性中,肺癌是白人女性、黑人女性和美国印第安女性/阿拉斯加土著女性中第二常见的癌症,并且是亚裔/太平洋岛女性和西班牙裔女性中第三常见的癌症。对于不戒烟的人,肺癌死亡率为15%,甚至对于50-59岁的戒烟的人仍为5%以上。仅美国在肺癌上每年保健费用就为950亿美元。

[0007] 91%由吸烟所致的肺癌是非小细胞肺癌(NSCLC),其占有所有肺癌的约87%。所有肺癌中剩余的13%是小细胞肺癌,虽然混合细胞肺癌确有发生。因为小细胞肺癌罕见且迅速致死,所以早期检测机会不大。

[0008] NSCLC有三种主要类型:鳞状细胞癌、大细胞癌和腺癌。腺癌是最常见形式的肺癌(30%-40%,且据报道高达50%),并且是在吸烟者和非吸烟者中最常发现的肺癌。鳞状细胞癌占有所有肺癌的25-30%,并且通常在近端支气管中发现。早期NSCLC倾向于局部化,如果早期检测到,其通常可以手术治疗,并且结果有利且存活率改善。其他治疗选择包括放疗、药物疗法和这些方法的组合。

[0009] NSCLC根据肿瘤大小及其在包括淋巴结在内的其他组织中的存在来分期。在隐藏期,癌细胞在痰样品或灌洗样品中发现,在肺中检测不到肿瘤。在0期,仅肺最深的内衬(lining)表现出癌细胞,肿瘤尚未生长通过内衬。在IA期,癌症被认为是侵袭性的,并且已生长深入肺组织,但是肿瘤横向(across)小于3cm。在这个时期,在支气管或淋巴结中未发现肿瘤。在IB期,肿瘤横向大于3cm或者已生长入支气管或胸膜,但是尚未生长入淋巴结。在IIA期,肿瘤横向大于3cm,并且已生长入淋巴结。在IIB期,肿瘤在淋巴结中发现并且横向大

于3cm,或者已生长入支气管或胸膜;或者癌症不在淋巴结中但是在胸壁、膈、胸膜、支气管或围绕心脏的组织中发现。在IIIA期,癌细胞在肺和支气管附近的淋巴结中以及肺之间的淋巴结中,但是在肿瘤所位于的胸的一侧上。在IIIB期,癌细胞位于胸的肿瘤的对侧及在颈中。肺附近的其他器官也可以具有癌细胞,并且在一个肺叶中可以发现多个肿瘤。在IV期,肿瘤在相同肺或两个肺的多于一个的肺叶中发现,并且癌细胞在身体其他部分发现。

[0010] 目前肺癌诊断方法包括测试痰的癌细胞、胸部X-射线、呼吸道的纤维光学评价和低剂量螺旋计算机断层扫描(CT)。痰细胞学的灵敏性非常低。胸部X-射线也较不灵敏,要求病灶的大小大于1cm以可见。支气管镜检查要求肿瘤在支气管镜可及的呼吸道内可见。最广泛认可的诊断方法是CT,但是常与X-射线一起,CT的应用涉及电离辐射,其本身可以导致癌症。CT还有明显的限制:扫描需要高技术水平以解释,并且许多观测的异常事实上不是肺癌,在跟踪CT发现中产生很大保健花费。最常见伴随发现是良性肺结节(lung nodule)。

[0011] 肺结节是较圆的病灶或异常组织区域,其位于肺中且大小可变。肺结节可以是良性或癌性的,但是大多数是良性的。如果小结小于4mm,则发病率仅为1.5%,如果小结为4-8mm,则发病率为约6%,如果超过20mm,则发病率为约20%。对于小和中等大小的小结,建议患者在3个月至1年内重复扫描。对于许多大的小结,患者接受生物活检(其是侵入性的并且可能导致并发症),即使大部分小结是良性的。

[0012] 因此,亟需可以替代或补充CT的诊断方法以减少进行的手术数量及降低手术并发症的风险。另外,甚至当不存在或未知肺结节时,需要在早期检测肺癌的方法以改善患者的结局。仅16%的肺癌病例诊断为局部早期癌症,其中5年存活率为46%,相比之下,在晚期诊断的为84%,其中5年存活率仅为13%。这证实依赖于症状诊断是无用的,因为它们中的许多是其他肺疾病常见的。这些症状包括久咳、血痰、胸痛和复发性支气管炎或肺炎。

[0013] 当存在癌症早期诊断方法时,益处通常是医疗行业所接受的。广泛使用筛查方案的癌症具有最高的5年存活率,如乳腺癌(88%)和结肠癌(65%),相比之下肺癌为16%。但是,如果癌症在1期通过筛查诊断,则88%的肺癌患者存活10年或更长。这证实明显亟需可以可靠地检测早期NSCLC的诊断方法。

[0014] 从健康状态发展至疾病伴随着受影响的组织中蛋白表达的变化。健康和疾病组织中蛋白质组的比较查询(interrogation)可以提供对疾病生物学的深入了解,并且导致发现用于诊断的生物标记、治疗性干预的新靶标以及鉴定最可能受益于靶向治疗的患者。特定疾病状态的生物标记的选择首先涉及鉴定标记,该标记对特定医疗应用在疾病群体中具有与对照群体相比可测量且统计学显著的差异。生物标记可以包括分泌或脱落(shed)的分子,其与疾病发展或进程平行,并且容易从肺组织或远端组织对病灶响应而扩散入血流。鉴定的生物标记或生物标记的集合(set)通常临床上进行验证,或者证实为对其所选的原始预期用途是可靠的指示物。生物标记可以包括小分子、肽、蛋白和核酸。影响生物标记鉴定的一些关键问题包括可用数据的过拟合(over-fitting)及数据偏差。

[0015] 已使用各种方法来试图鉴定生物标记及诊断疾病。对于基于蛋白的标记,这些方法包括二维电泳、质谱和免疫测定方法。对于核酸标记,这些方法包括mRNA表达谱、微RNA谱、FISH、基因表达系列分析(SAGE)和大规模基因表达阵列。

[0016] 二维电泳的应用由于如下问题而受限:低检测灵敏度;与蛋白溶解性、电荷及疏水性相关的问题;凝胶再现性;以及单个斑点代表多种蛋白的可能性。对于质谱,取决于所用

形式,限制围绕样品加工和分离、对低丰度蛋白的灵敏性、信噪比考虑及不能立即鉴定检测的蛋白而出现。免疫测定方法发现生物标记的限制集中在基于抗体的多重测定不能测量大量分析物。可以简单地印刷高质量抗体的阵列,并且无需夹心而测量与这些抗体结合的分析物。(这会是使用全基因组核酸序列经杂交测量有机体或细胞中的全部DNA或RNA序列的方式上的等同物。因为杂交可以是同一性的严紧测试,所以杂交实验可行。甚至非常好的抗体在选择它们的结合配偶体中也并非足够严紧来在血液或甚至是细胞提取物环境中工作,因为那些基质中的蛋白总体(ensemble)具有极其不同的丰度。)因此,必须使用不同的基于免疫测定的方法以发现生物标记-需要使用多重ELISA测定(即夹心)以获得足够严紧性来同时测量许多分析物从而决定哪些分析物的确是生物标记。夹心免疫测定不放大到高含量,因此使用标准阵列形式不能用严紧夹心免疫测定发现生物标记。最后,抗体试剂产生相当大的批次差异和试剂不稳定性。本发明的蛋白生物标记发现平台克服了这个问题。

[0017] 许多这些方法依赖或需要在分析前一些类型的样品的分级。因此进行设计为在一系列良好限定的样品群体中鉴定/发现统计学相关生物标记的足够有效的研究所需的样品制备是极其困难、昂贵和耗时的。在分级期间,大范围的变异性可能被引入各种样品。例如,一种潜在的标记可能对于方法是不稳定的,标记的浓度可能变化,不合适的聚集或解聚可能发生,无意的样品污染可能发生,并因此掩盖预期的早期疾病中的微小变化。

[0018] 广泛接受的是使用这些技术的生物标记发现和检测方法对于鉴定诊断性生物标记具有严重限制。这些限制包括不能检测低丰度生物标记,不能持续覆盖蛋白质组的完整动态范围,样品加工和分级中的不可再现性,以及方法的整体不可再现性和缺乏稳健性(robustness)。另外,这些研究在数据中引入了偏差,针对鉴定和验证靶疾病群体内的生物标记所需的分布和随机化方面,没有充分解决包括适当对照在内的样品群体的复杂性。

[0019] 尽管意图发现有效的新生物标记的努力已进行了几十年,但是这些努力大部分是不成功的。针对各种疾病的生物标记通常在实验室中鉴定,通常通过进行一些疾病过程的基础研究时偶然发现。基于所述发现及少量临床数据,发表的论文提示鉴定了新的生物标记。然而大多数这些建议的生物标记未证实是真实或有用的生物标记,这主要是因为测试的少量临床样品对于已确实发现有效的生物标记仅提供弱统计学证据。也就是说,最初的鉴定对于统计学的基本元素是不严格的。在1994-2003年的每一年中,检索科学文献表明公开了上千篇关于生物标记的参考文献。然而,在同时期内,FDA一年最多批准3种新蛋白生物标记的诊断应用,并且在若干年中没有批准新的蛋白生物标记。

[0020] 基于失败的生物标记发现努力的历史,已建议了数学理论以进一步促进通常理解,即针对疾病的生物标记很少且难以发现。基于2D凝胶或质谱的生物标记研究支持这些观点。通过这些方法鉴定了非常少的有用生物标记。然而,通常忽视2D凝胶和质谱测量血液中存在的约1nM或更高浓度的蛋白,这种蛋白的总体很可能是最不可能随疾病变化的。除了本发明的生物标记发现平台,尚不存在能够精确测量低得多的浓度的蛋白表达水平的蛋白质组生物标记发现平台。

[0021] 关于复杂的人生物学的生物化学途径已知许多。许多生物化学途径以在病理学内局部发挥作用的分泌的蛋白达到顶点或开始,例如分泌生长因子以刺激病理学中其他细胞的复制,分泌其他因子以避免免疫系统等。尽管许多这些分泌的蛋白以旁分泌方式发挥作用,但是一些在身体的远端运行。具有生物化学途径基本了解的本领域技术人员会理解,许

多病理学特异性蛋白应当以低于(甚至远低于)2D凝胶和质谱检测极限的浓度存在于血液中。在这种相对丰富数目的疾病生物标记的鉴定之前必须有一种蛋白质组平台,其可以分析低于2D凝胶或质谱可以检测的浓度的蛋白。

[0022] 因此,亟需生物标记、方法、装置、试剂、系统和试剂盒,其允许(a)区分良性肺结节(pulmonary nodule)与恶性肺结节;(b)检测肺癌生物标记;和(c)诊断肺癌。

[0023] 为了满足这种需要,已开发用于发现生物标记的基于适配体的蛋白质组新技术,其能够从血浆或血清的小样品体积同时测量上千种蛋白(参见例如,U.S. Pub.No.2010/0070191;U.S. Pub.No.2010/0086948,Ostroff et al.Nature Precedings,<http://precedings.nature.com/documents/4537/version/1> (2010);Gold et al.Nature Precedings,<http://precedings.nature.com/documents/4538/version/1> (2010))。通过新产生的包含化学修饰的核苷酸的慢解离速率适配体(SOMAmer)使得能够进行称为SOMAscan的这种技术,这大大扩展选择所述适配体的大随机化核酸文库的物理化学多样性(参见美国专利第7,947,447号)。与SELEX相容的这种修饰将官能团引入适配体,所述适配体常在蛋白-蛋白相互作用、抗体-抗原相互作用以及小分子药物与它们的蛋白靶标的相互作用中发现。总的来说,这些修饰的使用扩展可能的适配体靶标的范围,改善它们的结合特性并促进具有慢解离速率的适配体的选择。

[0024] 具体地,用将蛋白浓度的特征转化为相应的DNA适配体浓度的特征然后利用DNA微阵列平台定量的方法来测量复杂基质如血浆中的蛋白(Gold et al.Nature Precedings,<http://precedings.nature.com/documents/4538/version/1> (2010))。该测定借助于平衡结合和动力学攻击(dynamic challenge)。两者均在溶液中而不是表面上进行,以便利用结合和解离的更有利的动力学。本质上,该测定利用适配体作为具有明确形状的折叠结合实体和特异性探针可识别的独特序列的双重性。

[0025] 该测定能够同时测量血清中低至高丰度的大量蛋白。例如,在长期暴露于烟草的群体中分析来自非小细胞肺癌(NSCLC)的4个独立研究的1,326个个体的样品。测量15 μ L血清中的超过800种蛋白,并且开发了12-蛋白板,其在训练组中以91%灵敏度和84%特异性区分NSCLC与对照,在盲的独立验证组中以89%灵敏度和83%特异性区分NSCLC与对照。重要的是,对于早期和晚期NSCLC,表现是相似的(Ostroff et al.Nature Precedings,<http://precedings.nature.com/documents/4537/version/1> (2010))。

[0026] 目前为止,已利用这种方法进行包括肺癌(U.S. Pub.No.2010/0070191)、卵巢癌(U.S. Pub.No.2010/0086948)和慢性肾疾病在内的人疾病的几种临床生物标记研究。这些研究已鉴定这些疾病中的每种以及通常癌症的新的潜在的疾病生物标记。

[0027] 发明概述

[0028] 本申请证实新发现的微阵列平台技术用来鉴定来自组织的疾病相关的生物标记的用途。本申请包括用于检测和诊断来自组织的癌症,更特别是肺癌的生物标记、方法、试剂、装置、系统和试剂盒。本申请的生物标记使用实施例6详述的基于多重适配体的测定鉴定。通过使用本文所述的基于适配体的生物标记鉴定方法,本申请描述了可用于检测和诊断肺癌的惊人的大量来自组织的肺癌生物标记。在鉴定这些生物标记中,测量了来自许多个体样品的超过800种蛋白,其中一些浓度在低毫微微摩尔(femtomolar)范围。这比用2D凝胶和/或质谱进行的生物标记发现实验低约4个数量级。

[0029] 尽管一些所述肺癌生物标记可单独用于检测和诊断肺癌,但是本文所述的方法用于分组用作一组生物标记的肺癌生物标记的多个子集。一旦鉴定了单独的生物标记或生物标记的子集,则个体中肺癌的检测或诊断可以用能够测量生物学样品中所选生物标记或多个生物标记的水平差异的任何测定平台或者形式完成。

[0030] 然而,仅仅通过使用本文所述的基于适配体的生物标记鉴定方法,其中超过800个单独的潜在生物标记值从先前已经诊断为患有或不患有肺癌的大量个体中逐个进行了筛查,才可能鉴定本文公开的肺癌生物标记。这种发现方法与从条件培养基或裂解的细胞发现生物标记截然相反,因为其询问无需翻译为人病理学的更加患者相关的系统。

[0031] 因此,本申请一方面提供了一种或多种生物标记以用于单独或以各种组合来诊断肺癌,特别是非小细胞肺癌(NSCLC),或者允许将肺结节区分诊断为良性或恶性。示例性实施方案包括表18提供的生物标记,如上所述,这些生物标记用实施例1中一般描述并在实施例6中更具体描述的基于多重适配体的测定鉴定。表18提供的标记可以用于区分良性小结与癌性小结。表18提供的标记还可以用于区分无症状吸烟者与患有肺癌的吸烟者。在一方面,所述生物标记为MMP-7。在另一方面,所述生物标记为MMP-12。

[0032] 尽管一些所述肺癌生物标记可以单独用于检测和诊断肺癌,但是本文所述的方法还用于分组肺癌生物标记的多个子集,其各自可用作两个或更多个生物标记的组。因此,本申请的各个实施方案提供了包含N个生物标记的组合,其中N是至少2个生物标记。在其他实施方案中,N选自2-36个生物标记中的任意数。

[0033] 仍然在其他实施方案中,N选自2-7、2-10、2-15、2-20、2-25、2-30、2-36中的任意数。在其他实施方案中,N选自3-7、3-10、3-15、3-20、3-25、3-30、3-36中的任意数。在其他实施方案中,N选自4-7、4-10、4-15、4-20、4-25、4-30、4-36中的任意数。在其他实施方案中,N选自5-7、5-10、5-15、5-20、5-25、5-30、5-36中的任意数。在其他实施方案中,N选自6-10、6-15、6-20、6-25、6-30、6-36中的任意数。在其他实施方案中,N选自7-10、7-15、7-20、7-25、7-30、7-36中的任意数。在其他实施方案中,N选自8-10、8-15、8-20、8-25、8-30、8-36中的任意数。在其他实施方案中,N选自9-15、9-20、9-25、9-30、9-36中的任意数。在其他实施方案中,N选自10-15、10-20、10-25、10-30、10-36中的任意数。应当理解N可以选自包含类似但更高级(order)的范围。

[0034] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物学样品中检测至少一个生物标记值,所述至少一个生物标记值对应于选自表18提供的生物标记的组的至少一个生物标记,其中所述个体基于所述至少一个生物标记值分类为患有肺癌。

[0035] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于选自表18所列的生物标记的组的至少N个生物标记之一,其中所述个体患有肺癌的似然性(likelihood)基于所述生物标记值确定。

[0036] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于选自表18所列的生物标记的组的至少N个生物标记之一,其中所述个体基于所述生物标记值分类为患有肺癌,并且其中N=2-10。

[0037] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于选自表18所列的生物标记的组的至少N个生物标记之一,其中所述个体患有肺癌的似然性基于所述生物标记值确定,并且其中 $N=2-10$ 。

[0038] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测至少一个生物标记值,所述至少一个生物标记值对应于选自表18所列的生物标记的组的至少一个生物标记,其中基于所述至少一个生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性。

[0039] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于选自表18所列的生物标记的组的至少N个生物标记之一,其中基于所述生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中 $N=2-10$ 。

[0040] 在另一方面,本发明提供了一种诊断个体不患有肺癌的方法,所述方法包括在来自个体的生物学样品中检测至少一个生物标记值,所述至少一个生物标记值对应于选自表18所列的生物标记的组的至少一个生物标记,其中基于所述至少一个生物标记值,所述个体分类为不患有肺癌。

[0041] 在另一方面,本发明提供了一种诊断个体不患有肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于选自表18所列的生物标记的组的至少N个生物标记之一,其中基于所述生物标记值,所述个体分类为不患有肺癌,并且其中 $N=2-10$ 。

[0042] 在另一方面,本发明提供了一种诊断肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表18所列的生物标记的组,其中所述生物标记值的分类指示所述个体患有肺癌,并且其中 $N=3-10$ 。

[0043] 在另一方面,本发明提供了一种诊断肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表18所列的生物标记的组,其中所述生物标记值的分类指示所述个体患有肺癌,并且其中 $N=3-15$ 。

[0044] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表18所列的生物标记的组,其中基于所述生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,并且其中 $N=3-10$ 。

[0045] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表18所列的生物标记的组,其中基于所述生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中 $N=3-15$ 。

[0046] 在另一方面,本发明提供了一种诊断肺癌不存在的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表18所列的生物标记的组,其中所述生物标记值的分类指

示所述个体中不存在肺癌,并且其中 $N=3-10$ 。

[0047] 在另一方面,本发明提供了一种诊断肺癌不存在的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表18所列的生物标记的组,其中所述生物标记值的分类指示所述个体中不存在肺癌,并且其中 $N=3-15$ 。

[0048] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值对应于选自表18所列的生物标记的组的至少 N 个生物标记之一,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,并且其中 $N=2-10$ 。

[0049] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表18所列的生物标记的组,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中 $N=3-10$ 。

[0050] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表18所列的生物标记的组,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中 $N=3-15$ 。

[0051] 在另一方面,本发明提供了一种诊断个体中不存在肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值对应于选自表18所列的生物标记的组的至少 N 个生物标记之一,其中基于偏离预定阈值的分类评分,所述个体分类为不患有肺癌,并且其中 $N=2-10$ 。

[0052] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机执行方法。所述方法包括:在计算机上检索个体的生物标记信息,其中所述生物标记信息包括生物标记值,所述生物标记值每个对应于选自表18所列的生物标记的组的至少 N 个生物标记之一,其中 N 如上定义;用计算机对每个所述生物标记值进行分类;以及基于多个分类指示所述个体患有肺癌的似然性。

[0053] 在另一方面,本发明提供了一种将个体分类为患有或不患有肺癌的计算机执行方法。所述方法包括:在计算机上检索个体的生物标记信息,其中所述生物标记信息包括生物标记值,所述生物标记值每个对应于选自表18提供的生物标记的组的至少 N 个生物标记之一;用计算机对每个所述生物标记值进行分类;以及基于多个分类指示所述个体是否患有肺癌。

[0054] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质,所述程序代码可由计算装置或系统的处理器执行,所述程序代码包括:对归因于来自个体的生物学样品的数据进行检索的代码,其中所述数据包括生物标记值,所述生物标记值每个对应于所述生物学样品中选自表18所列的生物标记的组的至少 N 个生物标记之一,其中 N 如上定义;以及执行分类方法的代码,所述分类方法指示作为所述生物标记值的函数的所述个体患有肺癌的似然性。

[0055] 在另一方面,本发明提供了一种指示个体的肺癌状态的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质,所述程序代码可由计算装置或系统的处理器执行,所述程序代码包括:对归因于来自个体的生物学样品的数据进行检索的代码,其中所述数据包括生物标记值,所述生物标记值每个对应于所述生物学样品中选自表18提供的生物标记的组的至少N个生物标记之一;以及执行分类方法的代码,所述分类方法指示作为所述生物标记值的函数的所述个体的肺癌状态。

[0056] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机执行方法。所述方法包括:在计算机上检索个体的生物标记信息,其中所述生物标记信息包括生物标记值,所述生物标记值对应于选自表18所列的生物标记的组的生物标记;用计算机对所述生物标记值进行分类;以及基于所述分类指示所述个体患有肺癌的似然性。

[0057] 在另一方面,本发明提供了一种将个体分类为患有或不患有肺癌的计算机执行方法。所述方法包括:从计算机检索个体的生物标记信息,其中所述生物标记信息包括生物标记值,所述生物标记值对应于选自表18提供的生物标记的组的生物标记;用计算机对所述生物标记值进行分类;以及基于所述分类指示所述个体是否患有肺癌。

[0058] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质,所述程序代码可由计算装置或系统的处理器执行,所述程序代码包括:对归因于来自个体的生物学样品的数据进行检索的代码,其中所述数据包括生物标记值,所述生物标记值对应于所述生物学样品中选自表18所列的生物标记的组的生物标记;以及执行分类方法的代码,所述分类方法指示作为所述生物标记值的函数的所述个体患有肺癌的似然性。

[0059] 在另一方面,本发明提供了一种指示个体的肺癌状态的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质,所述程序代码可由计算装置或系统的处理器执行,所述程序代码包括:对归因于来自个体的生物学样品的数据进行检索的代码,其中所述数据包括生物标记值,所述生物标记值对应于所述生物学样品中选自表18提供的生物标记的组的生物标记;以及执行分类方法的代码,所述分类方法指示作为所述生物标记值的函数的所述个体的肺癌状态。

[0060] 在本申请的另一实施方案中,示例性实施方案包括表20提供的生物标记,如上所述,这些生物标记用实施例1中一般描述并在实施例6中更具体描述的基于多重适配体的测定鉴定。表20提供的标记可以用于区分良性小结与癌性小结。表20提供的标记还可以用于区分无症状吸烟者与患有肺癌的吸烟者。关于表20,N选自2-25个生物标记中的任意数。已确定表20提供的标记可用于组织和血清样品。

[0061] 仍然在其他实施方案中,N选自2-7、2-10、2-15、2-20、2-25中的任意数。在其他实施方案中,N选自3-7、3-10、3-15、3-20、3-25中的任意数。在其他实施方案中,N选自4-7、4-10、4-15、4-20、4-25中的任意数。在其他实施方案中,N选自5-7、5-10、5-15、5-20、5-25中的任意数。在其他实施方案中,N选自6-10、6-15、6-20、6-25中的任意数。在其他实施方案中,N选自7-10、7-15、7-20、7-25中的任意数。在其他实施方案中,N选自8-10、8-15、8-20、8-25中的任意数。在其他实施方案中,N选自9-15、9-20、9-25中的任意数。在其他实施方案中,N选自10-15、10-20、10-25中的任意数。应当理解N可以选自包含类似但更高级的范围。

[0062] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个

体的生物样品中检测至少一个生物标记值,所述至少一个生物标记值对应于选自表20提供的生物标记的组的至少一个生物标记,其中所述个体基于所述至少一个生物标记值分类为患有肺癌。

[0063] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物样品中检测生物标记值,所述生物标记值每个对应于选自表20所列的生物标记的组的至少N个生物标记之一,其中所述个体患有肺癌的似然性基于所述生物标记值确定。

[0064] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物样品中检测生物标记值,所述生物标记值每个对应于选自表20所列的生物标记的组的至少N个生物标记之一,其中所述个体基于所述生物标记值分类为患有肺癌,并且其中 $N=2-10$ 。

[0065] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物样品中检测生物标记值,所述生物标记值每个对应于选自表20所列的生物标记的组的至少N个生物标记之一,其中所述个体患有肺癌的似然性基于所述生物标记值确定,并且其中 $N=2-10$ 。

[0066] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物样品中检测至少一个生物标记值,所述至少一个生物标记值对应于选自表20所列的生物标记的组的至少一个生物标记,其中基于所述至少一个生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性。

[0067] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物样品中检测生物标记值,所述生物标记值每个对应于选自表20所列的生物标记的组的至少N个生物标记之一,其中基于所述生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中 $N=2-10$ 。

[0068] 在另一方面,本发明提供了一种诊断个体不患有肺癌的方法,所述方法包括在来自个体的生物样品中检测至少一个生物标记值,所述至少一个生物标记值对应于选自表20所列的生物标记的组的至少一个生物标记,其中基于所述至少一个生物标记值,所述个体分类为不患有肺癌。

[0069] 在另一方面,本发明提供了一种诊断个体不患有肺癌的方法,所述方法包括在来自个体的生物样品中检测生物标记值,所述生物标记值每个对应于选自表20所列的生物标记的组的至少N个生物标记之一,其中基于所述生物标记值,所述个体分类为不患有肺癌,并且其中 $N=2-10$ 。

[0070] 在另一方面,本发明提供了一种诊断肺癌的方法,所述方法包括在来自个体的生物样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表20所列的生物标记的组,其中所述生物标记值的分类指示所述个体患有肺癌,并且其中 $N=3-10$ 。

[0071] 在另一方面,本发明提供了一种诊断肺癌的方法,所述方法包括在来自个体的生物样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表20所列的生物标记的组,其中所述生物标记值的分类指示所述个体患有肺癌,并且其中 $N=3-15$ 。

[0072] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自

吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表20所列的生物标记的组,其中基于所述生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,并且其中N=3-10。

[0073] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表20所列的生物标记的组,其中基于所述生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中N=3-15。

[0074] 在另一方面,本发明提供了一种诊断肺癌不存在的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表20所列的生物标记的组,其中所述生物标记值的分类指示所述个体中不存在肺癌,并且其中N=3-10。

[0075] 在另一方面,本发明提供了一种诊断肺癌不存在的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表20所列的生物标记的组,其中所述生物标记值的分类指示所述个体中不存在肺癌,并且其中N=3-15。

[0076] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值对应于选自表20所列的生物标记的组的至少N个生物标记之一,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,并且其中N=2-10。

[0077] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表20所列的生物标记的组,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中N=3-10。

[0078] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组N个生物标记中的生物标记,其中所述生物标记选自表20所列的生物标记的组,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中N=3-15。

[0079] 在另一方面,本发明提供了一种诊断个体中不存在肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值对应于选自表20所列的生物标记的组的至少N个生物标记之一,其中基于偏离预定阈值的分类评分,所述个体分类为不患有肺癌,并且其中N=2-10。

[0080] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机执行方法。所述方法包括:在计算机上检索个体的生物标记信息,其中所述生物标记信息包括生物标记值,所述生物标记值每个对应于选自表20所列的生物标记的组的至少N个生物标记之一,其中N如上定义;用计算机对每个所述生物标记值进行分类;以及基于多个分类指示所述个体患有肺癌的似然性。

[0081] 在另一方面,本发明提供了一种将个体分类为患有或不患有肺癌的计算机执行方

法。所述方法包括：在计算机上检索个体的生物标记信息，其中所述生物标记信息包括生物标记值，所述生物标记值每个对应于选自表20提供的生物标记的组的至少N个生物标记之一；用计算机对每个所述生物标记值进行分类；以及基于多个分类指示所述个体是否患有肺癌。

[0082] 在另一方面，本发明提供了一种指示肺癌的似然性的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质，所述程序代码可由计算装置或系统的处理器执行，所述程序代码包括：对归因于来自个体的生物样品的数据进行检索的代码，其中所述数据包括生物标记值，所述生物标记值每个对应于所述生物样品中选自表20所列的生物标记的组的至少N个生物标记之一，其中N如上定义；以及执行分类方法的代码，所述分类方法指示作为所述生物标记值的函数的所述个体患有肺癌的似然性。

[0083] 在另一方面，本发明提供了一种指示个体的肺癌状态的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质，所述程序代码可由计算装置或系统的处理器执行，所述程序代码包括：对归因于来自个体的生物样品的数据进行检索的代码，其中所述数据包括生物标记值，所述生物标记值每个对应于所述生物样品中选自表20提供的生物标记的组的至少N个生物标记之一；以及执行分类方法的代码，所述分类方法指示作为所述生物标记值的函数的所述个体的肺癌状态。

[0084] 在另一方面，本发明提供了一种指示肺癌的似然性的计算机执行方法。所述方法包括：在计算机上检索个体的生物标记信息，其中所述生物标记信息包括生物标记值，所述生物标记值对应于选自表20所列的生物标记的组的生物标记；用计算机对所述生物标记值进行分类；以及基于所述分类指示所述个体患有肺癌的似然性。

[0085] 在另一方面，本发明提供了一种将个体分类为患有或不患有肺癌的计算机执行方法。所述方法包括：从计算机检索个体的生物标记信息，其中所述生物标记信息包括生物标记值，所述生物标记值对应于选自表20提供的生物标记的组的生物标记；用计算机对所述生物标记值进行分类；以及基于所述分类指示所述个体是否患有肺癌。

[0086] 在另一方面，本发明提供了一种指示肺癌的似然性的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质，所述程序代码可由计算装置或系统的处理器执行，所述程序代码包括：对归因于来自个体的生物样品的数据进行检索的代码，其中所述数据包括生物标记值，所述生物标记值对应于所述生物样品中选自表20所列的生物标记的组的生物标记；以及执行分类方法的代码，所述分类方法指示作为所述生物标记值的函数的所述个体患有肺癌的似然性。

[0087] 在另一方面，本发明提供了一种指示个体的肺癌状态的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质，所述程序代码可由计算装置或系统的处理器执行，所述程序代码包括：对归因于来自个体的生物样品的数据进行检索的代码，其中所述数据包括生物标记值，所述生物标记值对应于所述生物样品中选自表20提供的生物标记的组的生物标记；以及执行分类方法的代码，所述分类方法指示作为所述生物标记值的函数的所述个体的肺癌状态。

[0088] 在本申请的另一实施方案中，示例性实施方案包括表21提供的生物标记，这些生物标记用实施例1中一般描述并在实施例2和6中更具体描述的基于多重适配体的测定鉴定。表21提供的标记可以用于区分良性小结与癌性小结。表21提供的标记还可以用于区分

无症状吸烟者与患有肺癌的吸烟者。关于表21，N选自2-86个生物标记中的任意数。表21中包括的所有生物标记可用于提供在组织和血清样品中探寻的信息。

[0089] 仍然在其他实施方案中，N选自2-7、2-10、2-15、2-20、2-25、2-30、2-35、2-40、2-45、2-50、2-55、2-60、2-65、2-70、2-75、2-80或2-86中的任意数。在其他实施方案中，N选自3-7、3-10、3-15、3-20、3-25、3-30、3-35、3-40、3-45、3-50、3-55、3-60、3-65、3-70、3-75、3-80或3-86中的任意数。在其他实施方案中，N选自4-7、4-10、4-15、4-20、4-25、4-30、4-35、4-40、4-45、4-50、4-55、4-60、4-65、4-70、4-75、4-80或4-86中的任意数。在其他实施方案中，N选自5-7、5-10、5-15、5-20、5-25、5-30、5-35、5-40、5-45、5-50、5-55、5-60、5-65、5-70、5-75、5-80或5-86中的任意数。在其他实施方案中，N选自6-10、6-15、6-20、6-25、6-30、6-35、6-40、6-45、6-50、6-55、6-60、6-65、6-70、6-75、6-80或6-86中的任意数。在其他实施方案中，N选自7-10、7-15、7-20、7-25、7-30、7-35、7-40、7-45、7-50、7-55、7-60、7-65、7-70、7-75、7-80或7-86中的任意数。在其他实施方案中，N选自8-10、8-15、8-20、8-25、8-30、8-35、8-40、8-45、8-50、8-55、8-60、8-65、8-70、8-75、8-80或8-86中的任意数。在其他实施方案中，N选自9-15、9-20、9-25、9-30、9-35、9-40、9-45、9-50、9-55、9-60、9-65、9-70、9-75、9-80或9-86中的任意数。在其他实施方案中，N选自10-15、10-20、10-25、10-30、10-35、10-40、10-45、10-50、10-55、10-60、10-65、10-70、10-75、10-80或10-86中的任意数。应当理解N可以选自包含类似但更高级的范围。

[0090] 在另一方面，本发明提供了一种诊断个体的肺癌的方法，所述方法包括在来自个体的生物学样品中检测至少一个生物标记值，所述至少一个生物标记值对应于选自表21提供的生物标记的组的至少一个生物标记，其中所述个体基于所述至少一个生物标记值分类为患有肺癌。

[0091] 在另一方面，本发明提供了一种诊断个体的肺癌的方法，所述方法包括在来自个体的生物学样品中检测生物标记值，所述生物标记值每个对应于选自表21所列的生物标记的组的至少N个生物标记之一，其中所述个体患有肺癌的似然性基于所述生物标记值确定。

[0092] 在另一方面，本发明提供了一种诊断个体的肺癌的方法，所述方法包括在来自个体的生物学样品中检测生物标记值，所述生物标记值每个对应于选自表21所列的生物标记的组的至少N个生物标记之一，其中所述个体基于所述生物标记值分类为患有肺癌，并且其中N=2-10。

[0093] 在另一方面，本发明提供了一种诊断个体的肺癌的方法，所述方法包括在来自个体的生物学样品中检测生物标记值，所述生物标记值每个对应于选自表21所列的生物标记的组的至少N个生物标记之一，其中所述个体患有肺癌的似然性基于所述生物标记值确定，并且其中N=2-10。

[0094] 在另一方面，本发明提供了一种筛查吸烟者的肺癌的方法，所述方法包括在来自吸烟者个体的生物学样品中检测至少一个生物标记值，所述至少一个生物标记值对应于选自表21所列的生物标记的组的至少一个生物标记，其中基于所述至少一个生物标记值，所述个体分类为患有肺癌，或者确定所述个体患有肺癌的似然性。

[0095] 在另一方面，本发明提供了一种筛查吸烟者的肺癌的方法，所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值，所述生物标记值每个对应于选自表21所列的生物标记的组的至少N个生物标记之一，其中基于所述生物标记值，所述个体分类为患有肺

癌,或者确定所述个体患有肺癌的似然性,其中 $N=2-10$ 。

[0096] 在另一方面,本发明提供了一种诊断个体不患有肺癌的方法,所述方法包括在来自个体的生物学样品中检测至少一个生物标记值,所述至少一个生物标记值对应于选自表21所列的生物标记的组的至少一个生物标记,其中基于所述至少一个生物标记值,所述个体分类为不患有肺癌。

[0097] 在另一方面,本发明提供了一种诊断个体不患有肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于选自表21所列的生物标记的组的至少 N 个生物标记之一,其中基于所述生物标记值,所述个体分类为不患有肺癌,并且其中 $N=2-10$ 。

[0098] 在另一方面,本发明提供了一种诊断肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表21所列的生物标记的组,其中所述生物标记值的分类指示所述个体患有肺癌,并且其中 $N=3-10$ 。

[0099] 在另一方面,本发明提供了一种诊断肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表21所列的生物标记的组,其中所述生物标记值的分类指示所述个体患有肺癌,并且其中 $N=3-15$ 。

[0100] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表21所列的生物标记的组,其中基于所述生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,并且其中 $N=3-10$ 。

[0101] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表21所列的生物标记的组,其中基于所述生物标记值,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中 $N=3-15$ 。

[0102] 在另一方面,本发明提供了一种诊断肺癌不存在的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表21所列的生物标记的组,其中所述生物标记值的分类指示所述个体中不存在肺癌,并且其中 $N=3-10$ 。

[0103] 在另一方面,本发明提供了一种诊断肺癌不存在的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表21所列的生物标记的组,其中所述生物标记值的分类指示所述个体中不存在肺癌,并且其中 $N=3-15$ 。

[0104] 在另一方面,本发明提供了一种诊断个体的肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值对应于选自表21所列的生物标记的组的至少 N 个生物标记之一,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,并且其中 $N=2-10$ 。

[0105] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标

记中的生物标记,其中所述生物标记选自表21所列的生物标记的组,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中 $N=3-10$ 。

[0106] 在另一方面,本发明提供了一种筛查吸烟者的肺癌的方法,所述方法包括在来自吸烟者个体的生物学样品中检测生物标记值,所述生物标记值每个对应于一组 N 个生物标记中的生物标记,其中所述生物标记选自表21所列的生物标记的组,其中基于偏离预定阈值的分类评分,所述个体分类为患有肺癌,或者确定所述个体患有肺癌的似然性,其中 $N=3-15$ 。

[0107] 在另一方面,本发明提供了一种诊断个体中不存在肺癌的方法,所述方法包括在来自个体的生物学样品中检测生物标记值,所述生物标记值对应于选自表21所列的生物标记的组的至少 N 个生物标记之一,其中基于偏离预定阈值的分类评分,所述个体分类为不患有肺癌,并且其中 $N=2-10$ 。

[0108] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机执行方法。所述方法包括:在计算机上检索个体的生物标记信息,其中所述生物标记信息包括生物标记值,所述生物标记值每个对应于选自表21所列的生物标记的组的至少 N 个生物标记之一,其中 N 如上定义;用计算机对每个所述生物标记值进行分类;以及基于多个分类指示所述个体患有肺癌的似然性。

[0109] 在另一方面,本发明提供了一种将个体分类为患有或不患有肺癌的计算机执行方法。所述方法包括:在计算机上检索个体的生物标记信息,其中所述生物标记信息包括生物标记值,所述生物标记值每个对应于选自表21提供的生物标记的组的至少 N 个生物标记之一;用计算机对每个所述生物标记值进行分类;以及基于多个分类指示所述个体是否患有肺癌。

[0110] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质,所述程序代码可由计算装置或系统的处理器执行,所述程序代码包括:对归因于来自个体的生物学样品的数据进行检索的代码,其中所述数据包括生物标记值,所述生物标记值每个对应于所述生物学样品中选自表21所列的生物标记的组的至少 N 个生物标记之一,其中 N 如上定义;以及执行分类方法的代码,所述分类方法指示作为所述生物标记值的函数的所述个体患有肺癌的似然性。

[0111] 在另一方面,本发明提供了一种指示个体的肺癌状态的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质,所述程序代码可由计算装置或系统的处理器执行,所述程序代码包括:对归因于来自个体的生物学样品的数据进行检索的代码,其中所述数据包括生物标记值,所述生物标记值每个对应于所述生物学样品中选自表21提供的生物标记的组的至少 N 个生物标记之一;以及执行分类方法的代码,所述分类方法指示作为所述生物标记值的函数的所述个体的肺癌状态。

[0112] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机执行方法。所述方法包括:在计算机上检索个体的生物标记信息,其中所述生物标记信息包括生物标记值,所述生物标记值对应于选自表21所列的生物标记的组的生物标记;用计算机对所述生物标记值进行分类;以及基于所述分类指示所述个体患有肺癌的似然性。

[0113] 在另一方面,本发明提供了一种将个体分类为患有或不患有肺癌的计算机执行方

法。所述方法包括：从计算机检索个体的生物标记信息，其中所述生物标记信息包括生物标记值，所述生物标记值对应于选自表21提供的生物标记的组的生物标记；用计算机对所述生物标记值进行分类；以及基于所述分类指示所述个体是否患有肺癌。

[0114] 在另一方面，本发明提供了一种指示肺癌的似然性的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质，所述程序代码可由计算装置或系统的处理器执行，所述程序代码包括：对归因于来自个体的生物学样品的数据进行检索的代码，其中所述数据包括生物标记值，所述生物标记值对应于所述生物学样品中选自表21所列的生物标记的组的生物标记；以及执行分类方法的代码，所述分类方法指示作为所述生物标记值的函数的所述个体患有肺癌的似然性。

[0115] 在另一方面，本发明提供了一种指示个体的肺癌状态的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质，所述程序代码可由计算装置或系统的处理器执行，所述程序代码包括：对归因于来自个体的生物学样品的数据进行检索的代码，其中所述数据包括生物标记值，所述生物标记值对应于所述生物学样品中选自表21提供的生物标记的组的生物标记；以及执行分类方法的代码，所述分类方法指示作为所述生物标记值的函数的所述个体的肺癌状态。

[0116] 在本申请的一方面，上述方法的每个中所述选自表21的N个生物标记中的至少一个是选自表20的生物标记。在另一实施方案中，所述选自表20的生物标记为MMP-12。

[0117] 在另一方面，本发明提供了一种诊断肺癌的方法，所述方法包括在来自个体的生物学样品中检测生物标记值，所述生物标记值每个对应于一组生物标记中的生物标记，所述一组生物标记选自表22-25所列的生物标记的组，其中所述生物标记值的分类指示所述个体患有肺癌。

[0118] 在另一方面，本发明提供了一种诊断不存在肺癌的方法，所述方法包括在来自个体的生物学样品中检测生物标记值，所述生物标记值每个对应于一组生物标记中的生物标记，所述一组生物标记选自表22-25提供的生物标记的组，其中所述生物标记值的分类指示所述个体中不存在肺癌。

附图说明

[0119] 图1A是检测生物学样品中的肺癌的示例方法的流程图。

[0120] 图1B是用朴素贝叶斯 (naïve Bayes) 分类方法检测生物学样品中的肺癌的示例方法的流程图。

[0121] 图2示出单个生物标记SCFsR的ROC曲线，其使用用于检测无症状吸烟者中的肺癌的测试的朴素贝叶斯分类器 (classifier)。

[0122] 图3示出1至10个生物标记的生物标记组的ROC曲线，其使用用于检测无症状吸烟者中的肺癌的测试的朴素贝叶斯分类器。

[0123] 图4示出随生物标记的数目从1增加到10的分类评分 (特异性+灵敏性) 的增加，其使用用于良性小结-肺癌组的朴素贝叶斯分类。

[0124] 图5示出对于良性小结对照组 (实线) 和肺癌疾病组 (虚线)，作为log转化的RFU形式的累积分布函数 (cdf) 的SCFsR的测量的生物标记分布，以及它们的曲线拟合为正态cdf (短划线)，以用于训练 (train) 朴素贝叶斯分类器。

- [0125] 图6示出与本文所述的各种计算机执行方法一起使用的示例计算机系统。
- [0126] 图7是一实施方案的指示个体患有肺癌的似然性的方法的流程图。
- [0127] 图8是一实施方案的指示个体患有肺癌的似然性的方法的流程图。
- [0128] 图9示出可以用于检测生物学样品中一个或多个肺癌生物标记的示例适配体测定。
- [0129] 图10示出从聚集的潜在生物标记的集合使用哪些生物标记来构建分类器以区分NSCLC与良性小结的频率的柱状图。
- [0130] 图11示出从聚集的潜在生物标记的集合使用哪些生物标记来构建分类器以区分NSCLC与无症状吸烟者的频率的柱状图。
- [0131] 图12示出从位点一致(site-consistent)的潜在生物标记的集合使用哪些生物标记来构建分类器以区分NSCLC与良性小结的频率的柱状图。
- [0132] 图13示出从位点一致的潜在生物标记的集合使用哪些生物标记来构建分类器以区分NSCLC与无症状吸烟者的频率的柱状图。
- [0133] 图14示出从得自聚集和位点一致的标记组合的潜在生物标记的集合使用哪些生物标记来构建分类器以区分NSCLC与良性小结的频率的柱状图。
- [0134] 图15示出从得自聚集和位点一致的标记组合的潜在生物标记的集合使用哪些生物标记来构建分类器以区分NSCLC与无症状吸烟者的频率的柱状图。
- [0135] 图16示出得自pull-down实验的凝胶图像,其示出作为蛋白LBP、C9和IgM的捕获试剂的适配体的特异性。对于每个凝胶,泳道1是来自链霉抗生物素蛋白-琼脂糖珠的洗脱物,泳道2是最终洗脱物,泳道是MW标记泳道(主要条带从上至下是110、50、30、15和3.5kDa)。
- [0136] 图17A示出一对柱状图,其总结了使用表1第5列(实线)所列的生物标记和随机标记的集合(虚线)的所有可能的单蛋白朴素贝叶斯分类器评分(灵敏性+特异性)。
- [0137] 图17B示出一对柱状图,其总结了使用表1第5列(实线)所列的生物标记和随机标记的集合(虚线)的所有可能的二蛋白朴素贝叶斯分类器评分(灵敏性+特异性)。
- [0138] 图17C示出一对柱状图,其总结了使用表1第5列(实线)所列的生物标记和随机标记的集合(虚线)的所有可能的三蛋白朴素贝叶斯分类器评分(灵敏性+特异性)。
- [0139] 图18A示出一对柱状图,其总结了使用表1第6列(实线)所列的生物标记和随机标记的集合(虚线)的所有可能的单蛋白朴素贝叶斯分类器评分(灵敏性+特异性)。
- [0140] 图18B示出一对柱状图,其总结了使用表1第6列(实线)所列的生物标记和随机标记的集合(虚线)的所有可能的二蛋白朴素贝叶斯分类器评分(灵敏性+特异性)。
- [0141] 图18C示出一对柱状图,其总结了使用表1第6列(实线)所列的生物标记和随机标记的集合(虚线)的所有可能的三蛋白朴素贝叶斯分类器评分(灵敏性+特异性)。
- [0142] 图19A示出使用选自完全组(◆)的2-10个标记的朴素贝叶斯分类器的灵敏性+特异性评分,以及通过在用于良性小结对照组的分类器产生期间放弃最好的5个(■)、10个(▲)和15个(x)标记获得的评分。
- [0143] 图19B示出使用选自完全组(◆)的2-10个标记的朴素贝叶斯分类器的灵敏性+特异性评分,以及通过在用于吸烟者对照组的分类器产生期间放弃最好的5个(■)、10个(▲)和15个(x)标记获得的评分。
- [0144] 图20A示出一组ROC曲线,其从表38和39中的数据对1至5个标记的组建模。

[0145] 图20B示出一组ROC曲线,其从对图19A的1至5个标记的组训练数据计算。

[0146] 图21示出邻近与远端组织之间(图21A)、肿瘤与邻近组织之间(图21B)以及肿瘤与远端组织之间(图21C)来自8个NSCLC切除样品的813种蛋白的蛋白表达的相对变化,其表示为 $\log_2$ 中值比例。虚线代表2倍变化($\log_2=1$)。

[0147] 图22示出肿瘤组织样品中蛋白水平的热图。将样品排成列并分为远端、邻近和肿瘤样品。在每种组织类型内,将样品分为腺癌(AC)和鳞状细胞癌(SCC)。每列上的数字对应于患者编号。将蛋白按行展示并利用分级群聚(hierarchical clustering)排序。

[0148] 图23(A-T)示出与邻近或远端组织相比,在肿瘤组织中具有升高水平的蛋白。

[0149] 图24(A-P)示出与用于本研究的8个NSCLC样品的邻近或远端组织相比,在肿瘤组织中具有降低水平的蛋白。

[0150] 图25示出这个研究中检测的所选的生物标记在冰冻组织切片上的SOMAmer组织化学。(A)染色瘤巢周围的纤维胶原基质的血小板反应蛋白-2(红色)。(B)用血小板反应蛋白-2SOMAmer(红色)染色的相应的正常肺标本。(C)染色肺腺癌中的分散巨噬细胞的巨噬细胞甘露糖受体SOMAmer(红色)。(D)染色正常肺实质切片中的许多肺泡巨噬细胞的巨噬细胞甘露糖受体SOMAmer(红色)。(E)突出巨噬细胞甘露糖受体SOMAmer染色的细胞形态学分布的多色图像:绿色=细胞角蛋白(AE1/AE3抗体),红色=CD31(EP3095抗体),而橙色=SOMAmer。该图中的所有核用DAPI复染。

[0151] 图26示出与血清相比,NSCLC组织中蛋白表达的变化。最上面的两个图分别示出源自血清样品的 $\log_2$ 比例(LR)对源自邻近组织和远端组织的 $\log$ 比例。下面的4个图示出图的颜色所示的上图的放大部分(绿色为与非肿瘤组织相比减少的表达,红色为与非肿瘤组织相比增加的表达)。图23和24所示的分析物已标记,并且关于血清样品的出版物中提到的分析物如实心红色标记所示。

[0152] 图27示出组织样品中的血小板反应蛋白-2组织化学鉴定。在单一肺癌标本的连续冰冻切片中鉴定血小板反应蛋白-2,这通过:(A)自制的兔多克隆血小板反应蛋白-2多克隆抗体,(B)来自用来制备自制的多克隆抗体的兔的免疫前血清,(C)商业(Novus)兔多克隆血小板反应蛋白-2抗体,以及(D)血小板反应蛋白-2SOMAmer。然后血小板反应蛋白-2SOMAmer用来染色正常和恶性肺组织的冰冻切片,用标准抗生物素蛋白-生物素-过氧化物酶显色,以证实不同形态学分布:(E)瘤巢周围的纤维化基质的强染色,与癌细胞的最小染色,(F)粘液腺癌中瘤巢周围的纤维化基质的强染色,与癌细胞没有显著染色,(G)正常肺组织,显示支气管上皮细胞和分散的肺泡巨噬细胞的强胞质染色,以及(H)腺癌强胞质染色,与非纤维化的明显炎症性基质没有显著染色。

[0153] 发明详述

[0154] 除非另有说明,本文公开的发明的实施采用本领域技术水平内的化学、微生物学、分子生物学和重组DNA技术的常规方法。这类技术在文献中充分解释。参见,例如, Sambrook, et al. *Molecular Cloning: A Laboratory Manual* (Current Edition); *DNA Cloning: A Practical Approach*, vol. I&II (D. Glover, ed.); *Oligonucleotide Synthesis* (N. Gait, ed., Current Edition); *Nucleic Acid Hybridization* (B. Hames & S. Higgins, eds., Current Edition); *Transcription and Translation* (B. Hames & S. Higgins, eds., Current Edition); *Histology for Pathologists* (S. E. Mills, Current Edition)。本说明

书中引用的所有出版物、公开的专利文件和专利申请指示本发明所属领域的技术水平。本文引用的所有出版物、公开的专利文件和专利申请援引加入本文,与每个单独的出版物、公开的专利文件或专利申请具体地和单独地指明援引加入本文的程度相同。

[0155] 现在详细描述本发明的代表性实施方案。虽然本发明结合列举的实施方案进行描述,但是应理解本发明并不限于这些实施方案。相反,本发明旨在涵盖可以包括在如权利要求书所限定的本发明范围内的所有替代、修饰和等价物。

[0156] 除非特别说明,本文所用的技术和科学术语具有本发明所属领域技术人员通常理解的相同含义。尽管与本文所述的方法、装置和材料相似或等价的任何方法、装置和材料可用于实施或测试本发明,但是现在描述优选的方法、装置和材料。

[0157] 如在包括所附权利要求书在内的本申请中所用,除非特别说明,单数形式“一个(a)”、“一个(an)”和“这个(the)”包括复数形式,且与“至少一个”和“一个或多个”可以互换使用。因此,提及的“一个适配体”包括适配体的混合物,提及的“探针”包括探针的混合物等。

[0158] 如本文所用,术语“约”表示数值的不明显更改或变化,由此该数值所涉及的项目的基本功能未改变。

[0159] 如本文所用,术语“包含(comprises)”、“包含(comprising)”、“包括(includes)”、“包括(including)”、“含有(contains)”、“含有(containing)”及它们的任何变体意图覆盖非排他的包含,由此包含、包括或含有一个元件或者一系列元件的过程、方法、方法限定产品或组成(composition of matter)不仅包括这些元件,而且可以包括未明确列举或这样的过程、方法、方法限定产品或组成固有的其他元件。

[0160] 本申请包括用于检测和诊断肺癌的生物标记、方法、装置、试剂、系统和试剂盒。

[0161] 在一方面,本发明提供了一种或多种生物标记,其单独或以各种组合用于诊断肺癌,允许区分诊断肺结节是良性或恶性的,监测肺癌复发或寻址(address)其他临床指征。在其他方面,所述生物标记可以用于确定关于个体的肺癌的信息,例如预后、癌症分类、疾病风险的预测或治疗的选择。如下文详细描述,示例性实施方案包括表18、20和21提供的生物标记,这些生物标记使用基于多重适配体的测定来鉴定,所述测定在实施例1中,更特别是在实施例2和6中描述。每个生物标记可用于测定如下文定义的任何类型的样品。

[0162] 表1第2列列出得自以下分析的发现:来自NSCLC癌症病例的几百个个体血液样品以及来自吸烟者和来自诊断为具有良性肺结节的个体的几百个等价个体血液样品。吸烟者和良性小结组的设计使得肺癌测试与群体匹配可以具有最大益处。(这些病例和对照得自多个临床场所以模拟这样的测试可以应用的真实世界条件的范围)。潜在的生物标记在单独的样品而不是在混合疾病和对照血液中测量;这允许更好地理解与疾病(在这种情况下是肺癌)的存在和不存在相关的表型中个体和组的变化。由于对每个样品进行800个以上的蛋白测量,并且单独测量来自每个疾病和对照群体的几百个样品,所以表1第2列得自非常大的数据集合的分析。使用本文“生物标记的分类和疾病评分计算”章节中描述的方法分析测量结果。

[0163] 表1第2列列出了发现可用于区分得自NSCLC个体的样品与得自吸烟者和良性肺结节个体的“对照”样品的生物标记。使用本文所述的多重适配体测定,发现了38个生物标记,其区分得自患有肺癌的个体的样品与得自吸烟者对照组中个体的样品(见表1第6列)。相似

地,使用多重适配体测定,发现了40个生物标记,其区分得自NSCLC个体的样品与得自具有良性肺结节的人的样品(见表1第5列)。总之,这两列38和40个生物标记由61个独特的生物标记组成,因为区分NSCLC与良性小结的生物标记的列表和区分NSCLC与不患有肺癌的吸烟者的生物标记的列表之间有相当大的重叠。

[0164] 表18列出如实施例6所述得自分析诊断患有NSCLC的吸烟者的8个个体组织样品的发现。所有患者均为47-75岁的吸烟者,并且覆盖NSCLC1A期至3B期。从每个个体获得3个样品:肿瘤组织、邻近健康组织(在肿瘤的1cm内)和远端无关肺组织。选择样品以匹配肺癌诊断测试可以具有最大益处的群体。潜在的生物标记在单独的样品而不是在混合疾病和对照组织中测量;这允许更好地理解与疾病(在这种情况下是肺癌)的存在和不存在相关的表型中个体和组的变化。利用Mann-Whitney检验分析测量结果。

[0165] 表18列出发现可用于区分得自NSCLC个体的样品与得自相同个体的邻近和远端无关肺组织的“对照”样品的生物标记。使用本文所述的多重适配体测定,发现了36个生物标记,其区分肿瘤组织样品与得自诊断患有NSCLC的个体的邻近和远端肺组织的样品。关于表1第2列,可以看到这些生物标记中的11个与如实施例2所述在血清样品中鉴定的生物标记重叠。在原血清谱中未测量的额外的标记MMP-12已被发现是血清和组织中有用的生物标记。表21提供合并的血清和肿瘤组织样品中鉴定的生物标记的总数(86个)的列表。表20提供鉴定的肿瘤组织样品独特的生物标记的列表(25个)。

[0166] 虽然某些所述肺癌生物标记可单独用于检测和诊断肺癌,但是本文还描述了肺癌生物标记的多个子集的分组,其中每个分组或者子集选择可作为一组三个或更多个生物标记使用,这在本文中互换地称为“生物标记组”和一组。因此,本申请的各个实施方案提供了包含N个生物标记的组合,其中N是至少2个生物标记。在其他实施方案中,N选自2-86个生物标记(表21);2-36个生物标记(表18)或2-25个生物标记(表20)。在其他实施方案中,N选自2-86(表21),并且所述N个生物标记中至少一个为MMP-12。在其他实施方案中,N选自2-25(表20),并且所述N个生物标记中至少一个为MMP-12。包括MMP-12作为标记之一的2-5个生物标记的代表性组如表22-25所示。

[0167] 仍然在其他实施方案中,所述生物标记选自表18所列的生物标记,并且N选自2-7、2-10、2-15、2-20、2-25、2-30、2-36中的任意数。在其他实施方案中,N选自3-7、3-10、3-15、3-20、3-25、3-30、3-36中的任意数。在其他实施方案中,N选自4-7、4-10、4-15、4-20、4-25、4-30、4-36中的任意数。在其他实施方案中,N选自5-7、5-10、5-15、5-20、5-25、5-30、5-36中的任意数。在其他实施方案中,N选自6-10、6-15、6-20、6-25、6-30、6-36中的任意数。在其他实施方案中,N选自7-10、7-15、7-20、7-25、7-30、7-36中的任意数。在其他实施方案中,N选自8-10、8-15、8-20、8-25、8-30、8-36中的任意数。在其他实施方案中,N选自9-15、9-20、9-25、9-30、9-36中的任意数。在其他实施方案中,N选自10-15、10-20、10-25、10-30、10-36中的任意数。应当理解N可以选自包含类似但更高级的范围。

[0168] 仍然在其他实施方案中,所述生物标记选自表20所列的生物标记,并且N选自2-7、2-10、2-15、2-20、2-25中的任意数。在其他实施方案中,N选自3-7、3-10、3-15、3-20、3-25中的任意数。在其他实施方案中,N选自4-7、4-10、4-15、4-20、4-25中的任意数。在其他实施方案中,N选自5-7、5-10、5-15、5-20、5-25中的任意数。在其他实施方案中,N选自6-10、6-15、6-20、6-25中的任意数。在其他实施方案中,N选自7-10、7-15、7-20、7-25中的任意数。在其

他实施方案中,N选自8-10、8-15、8-20、8-25中的任意数。在其他实施方案中,N选自9-15、9-20、9-25中的任意数。在其他实施方案中,N选自10-15、10-20、10-25中的任意数。在其他实施方案中,N选自9-15、9-20、9-25中的任意数。在其他实施方案中,N选自10-15、10-20、10-25中的任意数。应当理解N可以选自包含类似但更高级的范围。

[0169] 仍然在其他实施方案中,所述生物标记选自表21所列的生物标记,并且N选自2-7、2-10、2-15、2-20、2-25、2-30、2-35、2-40、2-45、2-50、2-55、2-60、2-65、2-70、2-75、2-80或2-86中的任意数。在其他实施方案中,N选自3-7、3-10、3-15、3-20、3-25、3-30、3-35、3-40、3-45、3-50、3-55、3-60、3-65、3-70、3-75、3-80或3-86中的任意数。在其他实施方案中,N选自4-7、4-10、4-15、4-20、4-25、4-30、4-35、4-40、4-45、4-50、4-55、4-60、4-65、4-70、4-75、4-80或4-86中的任意数。在其他实施方案中,N选自5-7、5-10、5-15、5-20、5-25、5-30、5-35、5-40、5-45、5-50、5-55、5-60、5-65、5-70、5-75、5-80或5-86中的任意数。在其他实施方案中,N选自6-10、6-15、6-20、6-25、6-30、6-35、6-40、6-45、6-50、6-55、6-60、6-65、6-70、6-75、6-80或6-86中的任意数。在其他实施方案中,N选自7-10、7-15、7-20、7-25、7-30、7-35、7-40、7-45、7-50、7-55、7-60、7-65、7-70、7-75、7-80或7-86中的任意数。在其他实施方案中,N选自8-10、8-15、8-20、8-25、8-30、8-35、8-40、8-45、8-50、8-55、8-60、8-65、8-70、8-75、8-80或8-86中的任意数。在其他实施方案中,N选自9-15、9-20、9-25、9-30、9-35、9-40、9-45、9-50、9-55、9-60、9-65、9-70、9-75、9-80或9-86中的任意数。在其他实施方案中,N选自10-15、10-20、10-25、10-30、10-35、10-40、10-45、10-50、10-55、10-60、10-65、10-70、10-75、10-80或10-86中的任意数。应当理解N可以选自包含类似但更高级的范围。

[0170] 在一实施方案中,可用于生物标记子集或者组的生物标记的数目基于生物标记值的特定组合的灵敏性和特异性值。本文所用术语“灵敏性”和“特异性”是关于基于在个体的生物学样品中检测的一个或多个生物标记值来正确分类个体患有肺癌或不患有肺癌的能力。“灵敏性”指生物标记或多个生物标记关于正确分类患有肺癌的个体的性能。“特异性”指生物标记或多个生物标记关于正确分类不患有肺癌的个体的性能。例如,用于测试一组对照样品和肺癌样品的一组标记的85%特异性和90%灵敏性指85%的对照样品由该组正确分类为对照样品,并且90%的肺癌样品由该组正确分类为肺癌样品。期望或优选的最小值可以如实施例3所述确定。

[0171] 在一方面,在个体中通过以下方法检测或诊断肺癌:对来自所述个体的生物学样品进行测定,并检测生物标记值,所述生物标记值每个对应于生物标记MMP-7、MMP-12或IGFBP-2中的至少一个和选自表21的生物标记列表的至少N个额外的生物标记,其中N等于2、3、4、5、6、7、8、9、10、11、12、13、14或15。在另一方面,在个体中通过以下方法检测或诊断肺癌:对来自所述个体的生物学样品进行测定,并检测生物标记值,所述生物标记值每个对应于生物标记MMP-7、MMP-12或IGFBP-2中的至少一个和选自表21的生物标记列表的至少N个额外的生物标记之一,其中N等于1、2、3、4、5、6、7、8、9、10、11、12或13。在另一方面,在个体中通过以下方法检测或诊断肺癌:对来自所述个体的生物学样品进行测定,并检测生物标记值,所述生物标记值每个对应于生物标记MMP-7和选自表21的生物标记列表的至少N个额外的生物标记之一,其中N等于2、3、4、5、6、7、8、9、10、11、12、13、14或15。在另一方面,在个体中通过以下方法检测或诊断肺癌:对来自所述个体的生物学样品进行测定,并检测生物标记值,所述生物标记值每个对应于生物标记MMP-12和选自表21的生物标记列表的至少

N个额外的生物标记之一,其中N等于2、3、4、5、6、7、8、9、10、11、12、13、14或15。在另一方面,在个体中通过以下方法检测或诊断肺癌:对来自所述个体的生物学样品进行测定,并检测生物标记值,所述生物标记值每个对应于生物标记IGFBP-2和选自表21的生物标记列表的至少N个额外的生物标记之一,其中N等于2、3、4、5、6、7、8、9、10、11、12、13、14或15。

[0172] 本文鉴定的肺癌生物标记表示较大数目的可用于有效检测或诊断肺癌的生物标记的子集或组的选择。期望数目的这类生物标记的选择取决于所选生物标记的特定组合。重要的是记住:用于检测或诊断肺癌的生物标记的组还可以包括在表18、20或21中未发现生物标记,并且包括在表18、20或21中未发现的额外的生物标记可以减少选自表18、20或21的特定子集或组中的生物标记的数目。如果额外的生物医学信息与生物标记值联合用于建立对于给定测定可接受的灵敏性和特异性值,则用于子集或组的来自表18、20或21的生物标记的数目也可以减少。

[0173] 可以影响用于生物标记的子集或组的生物标记数目的另一因素是用于从进行肺癌诊断的个体中获得生物学样品的方法。在精心控制的样品获取环境中,符合期望的灵敏性和特异性值所必需的生物标记的数目会低于在样品收集、处理和贮存中可以存在更多变化的情况中的数目。在研究表18、20或21所列的生物标记列表中,利用多个样品收集位点来收集数据以进行分类器训练。这提供了更稳健的生物标记,其对于样品收集、处理和贮存中的变化较不敏感,但是如果训练数据都在非常相似条件下获得,则还可以要求子集或组中更大的生物标记数目。

[0174] 本申请的一方面可以参考图1A和B来一般性描述。生物学样品获得自所关注的一个或多个个体。然后测定该生物学样品以检测所关注的一个或多个(N个)生物标记的存在,并确定所述N个生物标记的每一个的生物标记值(在图1B中称为标记RFU)。一旦检测生物标记并指定生物标记值,则如本文详细描述地对每个标记进行评分或者分类。然后组合标记评分以提供总诊断评分,其表示获取样品的个体患有肺癌的似然性。

[0175] 如本文所用,“肺(lung)”与“肺(pulmonary)”可互换使用。

[0176] 如本文所用,“吸烟者”指具有吸烟历史的个体。

[0177] “生物学样品”、“样品”和“测试样品”在本文中可互换使用,指获得自或以另外的方式源自个体的任何材料、生物液体、组织或者细胞。这包括血液(包括全血、白细胞、外周血单核细胞、血沉棕黄层(buffy coat)、血浆和血清)、痰、泪液、粘液、洗鼻液(wash)、鼻抽吸物(aspirate)、呼吸物(breath)、尿、精液、唾液、脑膜液(meningeal fluid)、羊水、腺体液(glandular fluid)、淋巴液、乳头抽吸物、支气管抽吸物、滑液、关节抽吸物、细胞、细胞提取物和脑脊液。其还包括上述所有材料的实验分离级分。例如,血液样品可以分级分离为血清或含有诸如红细胞或白细胞(white blood cell)(白细胞(leukocyte))的特定类型血细胞的级分。如果需要,样品可以是来自个体的样品的组合,如组织与液体样品的组合。术语“生物学样品”还包括含有均质固体材料的材料,如来自粪便样品、组织样品或组织活检样品的材料。术语“生物学样品”还包括源自组织培养或者细胞培养的材料。可以使用获得生物学样品的任何合适方法;示例性方法包括如静脉切开放血术、拭子(如口腔拭子)以及细针抽吸活检方法。易受细针抽吸影响的示例性组织包括淋巴结、肺、肺洗液、BAL(支气管肺泡灌洗液)、甲状腺、乳腺和肝。样品还可以通过显微切割(如激光捕获显微切割(LCM)或激光显微切割(LMD))、膀胱冲洗、涂片(如PAP涂片)或导管灌洗收集。获得自或源自个体的

“生物样品”包括在获得自所述个体之后已经通过任何合适方式处理的任何此类样品。

[0178] “组织样品”或“组织”指上述生物样品的某子集。根据这个定义,组织是异质环境中的大分子的集合。如本文所用,组织指单一细胞类型、细胞类型的集合、细胞的聚集体或大分子的聚集体。在结构和组成方面,组织一般是物理矩阵,其可以是流体或刚性的。胞外基质是结构和组成上更加刚性的组织的实例,而膜双层在结构和组成上更加流体。组织包括但不限于通常为特定种类的细胞与它们的胞间物质一起的聚集体,其形成通常用来表示给定器官的一般细胞结构 (cellular fabric) 的结构材料之一,例如肾组织、脑组织、肺组织。组织的4个普通类别为上皮组织、结缔组织、神经组织和肌肉组织。鉴定组织靶标的慢解离速率适配体的方法描述于2011年1月13日公开的国际申请公开第W0 2011/006075号,其整体援引加入本文。

[0179] 在这个定义内的组织的实例包括但不限于大分子的异质聚集体,如非细胞的纤维蛋白凝块;细胞的均质或异质聚集体;包含具有特异性功能的细胞的高级结构,如器官、肿瘤、淋巴结、动脉等;以及单独的细胞。组织或细胞可以是在它们的天然环境中、分离的或者在组织培养中。组织可以是完整的或修饰的。修饰可以包括许多改变,例如转化、转染、激活和亚结构分离,如细胞膜、细胞核、细胞器等。

[0180] 组织、细胞或亚细胞结构的来源可以得自原核生物以及真核生物。这包括人、动物、植物、细菌、真菌和病毒结构。

[0181] 此外,应当认识到生物样品可以通过从许多个体中取得生物样品并将它们混合或混合每个个体的生物样品的等份而获得。混合的样品可以作为来自单个个体的样品进行处理,并且如果在混合的样品中确定癌症的存在,则可以将每个个体的生物样品再进行测试以确定哪个/哪些个体患有肺癌。

[0182] 为了本说明书的目的,短语“归因于来自个体的生物样品的数据”指所述数据以某种形式源自所述个体的生物样品或利用所述个体的生物样品产生。数据在产生后可以被重新格式化、修改或以数学方式改变至某种程度,例如通过从一种测量系统中的单位转变为另一测量系统中的单位;但是应当理解,数据源自所述生物样品或利用所述生物样品产生。

[0183] “靶标”、“靶分子”和“分析物”在本文中可互换使用,指可能存在于生物样品中的任何所关注的分子。“所关注的分子”包括特定分子的任何微小变化,如在蛋白的情况下,例如氨基酸序列的微小变化、二硫键形成、糖基化、脂质化、乙酰化、磷酸化或者任何其他操作或修饰,如与基本不改变分子同一性的标记组分偶联。“靶分子”、“靶标”或“分析物”是一种类型或种类的分子或多分子结构的一组拷贝。“靶分子”、“靶标”和“分析物”指一组以上这样的分子。示例性靶分子包括蛋白、多肽、核酸、碳水化合物、脂质、多糖、糖蛋白、激素、受体、抗原、抗体、affybodies、抗体模拟物 (mimic)、病毒、病原体、毒性物质、底物、代谢物、过渡态类似物、辅因子、抑制剂、药物、染料、营养素、生长因子、细胞、组织以及前述任何物质的任何片段或部分。

[0184] 如本文所用,“多肽”、“肽”和“蛋白”在本文中可互换使用,指任何长度的氨基酸聚合物。聚合物可以是线性或支化的,其可以包含修饰的氨基酸,并且其可以被非氨基酸中断。该术语还涵盖已经被天然修饰或者通过干预修饰的氨基酸聚合物;例如,二硫键形成、糖基化、脂质化、乙酰化、磷酸化或任何其他操作或修饰,如与标记组分偶联。该定义还包括

例如含有一个或多个氨基酸类似物(包括例如非天然氨基酸等)以及本领域已知的其他修饰的多肽。多肽可以是单链或缔合(associated)链。该定义还包括前蛋白和完整的成熟蛋白;源生自成熟蛋白的肽或多肽;蛋白的片段;剪接变体;蛋白的重组形式;具有氨基酸修饰、缺失或取代的蛋白变体;消化;以及翻译后修饰,如糖基化、乙酰化、磷酸化等。

[0185] 如本文所用,“标记”和“生物标记”可互换使用,指指示个体中正常或异常过程或者个体中疾病或其他疾病状况的迹象或者是个体中正常或异常过程或者个体中疾病或其他疾病状况的迹象的靶分子。更具体地,“标记”或“生物标记”是与无论正常与否的特定生理状态或过程的存在相关的解剖学、生理学、生物化学或分子参数,并且如果是异常的,则无论是慢性或急性的。生物标记可以通过各种方法检测和测量,包括实验室测定和医学成像。当生物标记是蛋白时,还可以使用相应基因的表达作为生物学样品中相应蛋白生物标记的量或存在或不存在或者编码该生物标记的基因或控制该生物标记表达的蛋白的甲基化状态的替代测量。

[0186] 如本文所用,“生物标记值”、“值”、“生物标记水平”和“水平”在本文中可互换使用,指使用任何分析方法来检测生物学样品中的生物标记而进行的测量,其示出所述生物学样品中的生物标记、对于所述生物标记或对应于所述生物标记的存在、不存在、绝对量或浓度、相对量或浓度、效价、水平、表达水平、测量水平的比率等。所述“值”或“水平”的确切性质取决于用于检测生物标记的特定分析方法的具体设计和组分。

[0187] 当生物标记表示个体中异常过程或疾病或其他疾病状况或者是个体中异常过程或疾病或其他疾病状况的迹象时,该生物标记通常描述为与表示个体中正常过程或不存在疾病或其他疾病状况或者是个体中正常过程或不存在疾病或其他疾病状况的迹象的生物标记的表达水平或值相比时是过表达或低表达的。

[0188] “上调”、“上调的”、“过表达”、“过表达的”及其任何变体在本文中可互换使用,指生物学样品中生物标记的值或水平高于通常在来自健康或正常个体的相似生物学样品中检测的所述生物标记的值或水平(或者值或水平的范围)。该术语还可以指生物学样品中生物标记的值或水平高于在特定疾病的不同阶段检测的所述生物标记的值或水平(或者值或水平的范围)。

[0189] “下调”、“下调的”、“低表达”或“表达低的”及其任何变体在本文中可互换使用,指生物学样品中生物标记的值或水平低于通常在来自健康或正常个体的相似生物学样品中检测的生物标记的值或水平(或者值或水平的范围)。该术语还可以指生物学样品中生物标记的值或水平低于在特定疾病的不同阶段检测的所述生物标记的值或水平(或者值或水平的范围)。

[0190] 此外,过表达的或低表达的生物标记还可以指与所述生物标记的“正常”表达水平或值相比是“差异表达的”或者具有“不同水平”或“不同值”,所述“正常”表达水平或值表示个体中正常过程或不存在疾病或其他疾病状况或者是个体中正常过程或不存在疾病或其他疾病状况的迹象。因此,生物标记的“差异表达”还可以指与所述生物标记的“正常”表达水平不同。

[0191] 术语“不同的基因表达”和“差异表达”在本文中可互换使用,指在患有指定疾病的对象中基因(或其相应的蛋白表达产物)的表达被激活至相对于其在正常或对照对象中的表达较高或较低的水平。该术语还包括基因(或其相应的蛋白表达产物)的表达在相同疾病

的不同阶段被激活至较高或较低水平。还应当理解差异表达的基因可以在核酸水平或蛋白水平激活或抑制,或者可以进行可变剪接以获得不同的多肽产物。这样的差异可以通过许多改变证实,包括多肽的mRNA水平、表面表达、分泌或其他分配(partitioning)。不同的基因表达可以包括比较两个或更多个基因或者它们的基因产物之间的表达;或者比较两个或更多个基因或者它们的基因产物之间的表达的比率;或者甚至比较相同基因的两种不同加工的产物,其在正常对象与患病对象之间或者在相同疾病的不同阶段之间是不同的。差异表达包括在例如正常和患病细胞或者经历不同疾病事件或疾病阶段的细胞中的基因或其表达产物在时间或细胞表达模式中的定量以及定性的差异。

[0192] 如本文所用,“个体”指测试对象或患者。个体可以是哺乳动物或非哺乳动物。在许多实施方案中,个体是哺乳动物。哺乳动物个体可以是人或非人。在许多实施方案中,个体是人。健康或正常个体是其中通过常规诊断方法不可检测出所关注的疾病或疾病状况(包括例如肺疾病、肺相关疾病或其他肺疾病状况)的个体。

[0193] “诊断(Diagnose)”、“诊断(diagnosing)”、“诊断(diagnosis)”及其变体指基于个体相关的一种或多种迹象、症状、数据或其他信息对所述个体的健康状态或疾病状况的检测、确定或识别。个体的健康状态可以诊断为健康/正常(即诊断为不存在疾病或疾病状况)或者诊断为患病/异常(即诊断为存在疾病或疾病状况或者对疾病或疾病状况的特征的评价)。对于特定疾病或疾病状况,术语“诊断(diagnose)”、“诊断(diagnosing)”、“诊断(diagnosis)”等涵盖对疾病的初始检测;对疾病的表征或分类;疾病的进展、缓解或复发的检测;以及在给予个体治疗或疗法后疾病应答的检测。肺癌的诊断包括区分患有癌症与不患有癌症的个体,包括吸烟者和非吸烟者。其还包括区分良性肺结节与癌性肺结节。

[0194] “预后(Prognose)”、“预后(prognosing)”、“预后(prognosis)”及其变体指预测患有疾病或疾病状况的个体中所述疾病或疾病状况的未来进程(如预测患者存活),这个术语涵盖在给予个体治疗或疗法后评价疾病的应答。

[0195] “评价(Evaluate)”、“评价(evaluating)”、“评价(evaluation)”及其变体涵盖“诊断”和“预后”,并且还涵盖对不患病个体的疾病或疾病状况的未来进程的确定或预测以及确定或预测在表面上已经治愈疾病的个体中所述疾病或疾病状况复发的似然性。术语“评价”还包括评价个体对疗法的应答,例如预测个体是否可能对治疗剂顺利地应答,或者不大可能对治疗剂应答(或者会例如经历毒性或其他不期望的副作用);选择给予个体的治疗剂;或者监测或确定个体对已经给予该个体的疗法的应答。因此,“评价”肺癌可以包括例如以下任何方面:预后个体中肺癌的未来进程;预测表面上已经治愈肺癌的个体中肺癌的复发;或者确定或预测个体对于肺癌治疗的应答;或者基于确定源自个体生物学样品的生物标记值来选择给予该个体的肺癌治疗。

[0196] 任何如下实例均可以称作“诊断”或“评价”肺癌:最初检测肺癌的存在或不存在;确定肺癌的具体阶段、类型或亚型或者其他分类或特征;确定肺结节是良性病变或恶性肺肿瘤;或者检测/监测肺癌进展(如监测肺肿瘤生长或转移扩散)、缓解或复发。

[0197] 如本文所用,“额外的生物医学信息”指除了使用本文所述的任何生物标记之外的对个体所做的与肺癌风险相关的一个或多个评价。“额外的生物医学信息”包括如下任何方面:个体的物理描述(physical descriptor)、通过CT成像观察到的肺结节的物理描述、个体的身高和/或体重、个体的性别、个体的种族、吸烟史、职业史、暴露于已知致癌物(如暴露

于任何石棉、氦气、化学品、来自火的烟以及空气污染,这可以包括来自静止或移动来源的排放物,如工业/工厂或汽车/海运/飞机排放物)、暴露于二手烟、肺癌(或其他癌症)家族史、肺结节的存在、小结的大小、小结的部位、小结的形态学(如通过CT成像、磨玻璃影(GGO)观察,实体、非实体)、小结的边界特点(如平滑、分叶、锐利及光滑、针状、浸润)等。吸烟史通常以术语“包年(pack year)”量化,其指一个人吸烟的年数乘以每天吸烟的平均包数。例如,平均每天吸一包烟共35年的人称作吸烟史为35包年。额外的生物医学信息可以通过使用本领域已知的常规技术得自个体,如通过使用常规患者问卷调查或健康史问卷调查等得自个体自身,或者得自医学从业人员等。或者,额外的生物医学信息可以得自常规成像技术,包括CT成像(如低剂量CT成像)和X-射线检查。生物标记水平的测试联合任何额外的生物医学信息的评价与单独测试生物标记或单独评价额外的生物医学信息的任何特定项目(如单独的CT成像)相比,可以例如改善检测肺癌(或其他肺癌相关用途)的灵敏性、特异性和/或AUC。

[0198] 术语“曲线下面积”或“AUC”指接受者操作特征(ROC)曲线下的面积,这两个术语均为本领域所熟知。AUC测量可以用于比较完整数据范围内的分类器的精确性。具有较大AUC的分类器具有较大的能力来正确分类两个所关注的组(如肺癌样品与正常或对照样品)之间的未知情况。ROC曲线可以用于对特定特征的性能作图(如本文描述的任何生物标记和/或任何额外的生物医学信息项目),以在两个群体之间进行区分(如患有肺癌的病例与无肺癌的对照)。通常,整个群体(如病例与对照)的特征数据基于单个特征的值递增分类。然后,对于该特征的每个值,计算数据的真阳性率和假阳性率。真阳性率通过计数高于该特征值的病例数,然后除以病例总数而确定。假阳性率通过计数高于该特征值的对照数,然后除以对照总数而确定。尽管这个定义指其中特征在病例中与在对照中相比升高的情况,但是这个定义还适用于其中特征在病例中与在对照中相比降低的情况(在这种情况下,计数低于该特征值的样品)。ROC曲线可以对单个特征以及其他单个输出产生,例如两个或更多个特征的组合可以是数学组合(如加、减、乘等)以提供单个的和值,该单个的和值可以在ROC曲线中绘制。此外,其中组合产生单个输出值的多个特征的任意组合可以在ROC曲线中绘制。这些特征的组合可以包括测试。ROC曲线是测试的真阳性率(灵敏性)对测试的假阳性率(1-特异性)的作图。

[0199] 如本文所用,“检测”或“确定”生物标记值包括使用观察和记录对应于生物标记值的信号所需的设备以及产生该信号所需的材料。在许多实施方案中,生物标记值使用任何合适的方法检测,包括荧光、化学发光、表面等离子共振、表面声波、质谱、红外线光谱、拉曼光谱、原子力显微术、扫描隧道显微术、电子化学检测方法、核磁共振、量子点等。

[0200] “固体支持物”在本文中指具有分子可以直接或间接,通过共价键或非共价键附着的表面的任何支持物。“固体支持物”可以具有各种物理形式,可以包括例如膜;芯片(如蛋白芯片);玻片(如载玻片或盖玻片);柱;空心、固体、半固体、有孔或有腔的颗粒,例如珠;凝胶;纤维,包括光学纤维材料;基质;及样品容器。示例性样品容器包括样品孔、管、毛细管、小瓶及能够容纳样品的任何其他容器、沟槽或凹陷。样品容器可以包含于多样品平台上,如微量滴定板、玻片、微流体装置等。支持物可以由天然或合成材料、有机或无机材料组成。其上附着捕获试剂的固体支持物的成分通常取决于附着方法(如共价附着)。其他示例性容器包括微滴和微流体控制的或大量的油/水性乳液,在其中可以进行测定和相关操作。合适的

固体支持物包括例如塑料、树脂、多糖、硅石或基于硅石的材料、官能化玻璃、改性的硅、碳、金属、无机玻璃、膜、尼龙、天然纤维(例如丝、羊毛和棉)、聚合物等。包含固体支持物的材料可以包含反应基团,例如羧基、氨基或羟基以用于捕获试剂的附着。聚合固体支持物可以包括如聚苯乙烯、聚对苯二甲酸乙二醇酯、聚乙酸乙烯酯、聚氯乙烯、聚乙烯吡咯烷酮、聚丙烯腈、聚甲基丙烯酸甲酯、聚四氟乙烯、丁基橡胶、苯乙烯丁二烯橡胶、天然橡胶、聚乙烯、聚丙烯、(聚)四氟乙烯、(聚)偏氟乙烯、聚碳酸酯和聚甲基戊烯。可以使用的合适的固体支持物颗粒包括例如编码的颗粒,如Luminex[®]-型编码的颗粒、磁性颗粒以及玻璃颗粒。

[0201] 生物标记的示例性用途

[0202] 在许多示例性实施方案中,本发明提供了诊断个体的肺癌的方法,所述方法通过检测对应于个体的肺组织中存在的一个或多个生物标记的一个或多个生物标记值来进行,并且通过任何数目的分析方法来进行,包括本文所述的任何分析方法。这些生物标记例如在肺癌个体中与在无肺癌个体中差异表达,特别是NSCLC。生物标记在个体中的差异表达的检测可以用于例如早期诊断肺癌、区分良性与恶性肺结节(例如在计算机断层(CT)扫描上观察到的小结)、监测肺癌复发或者用于其他临床指征,包括确定预后和治疗方法。

[0203] 本文所述的任何生物标记可以用于肺癌的各种临床指征,包括如下任何方面:检测肺癌(如在高危个体或人群中);表征肺癌(如确定肺癌类型、亚型或阶段),如通过区分非小细胞肺癌(NSCLC)与小细胞肺癌(SCLC)和/或区分腺癌与鳞状细胞癌(或者促进组织病理学);确定肺结节是良性小结或恶性肺肿瘤;确定肺癌预后;监测肺癌进展或缓解;监测肺癌复发;监测转移;治疗选择;监测对治疗剂或其他治疗的应答;对个体的计算机断层扫描(CT)筛查进行分层(stratification)(如鉴定面临较高肺癌风险的那些个体,从而最可能得益于螺旋CT筛查,因此增加CT的阳性预测值);将生物标记测试与诸如吸烟史等的额外的生物医学信息组合,或者与小结大小、形态学等组合(如提供与单独的CT测试或生物标记测试相比具有增加的诊断性能的测定);促进诊断肺结节为恶性或良性;促进在CT上观察到肺结节就做出临床决定(例如,如果小结被视为低风险的,例如如果基于生物标记的测试是阴性的,有或无小结大小的归类,则要求重复CT扫描;或者,如果小结被视为中等至高风险的,例如如果基于生物标记的测试是阳性的,有或无小结大小的归类,则考虑进行活组织检查);以及促进决定临床后续处理(如在CT上观察到未钙化小结后是否实施重复CT扫描、细针活组织检查或胸廓切开术)。生物标记测试可以改善单独的CT筛查的阳性预测值(PPV)。除了联合CT筛查之外,本文所述的生物标记还可以与用于与肺癌的任何其他成像方式如胸部X-射线检查联合使用。此外,所述生物标记还可以用于在通过成像方式或其他临床相关性检测肺癌指征之前或者在症状出现之前允许这些应用的某一些。

[0204] 本文所述的任何生物标记可以用于诊断肺癌的示例性方式是:未知患有肺癌的个体中一个或多个所述生物标记的差异表达可以表明该个体患有肺癌,从而使得可以在治疗最有效的疾病早期检测肺癌,也许在通过其他方式检测肺癌之前或者在症状出现之前检测肺癌。肺癌期间一个或多个生物标记的过表达可以指示肺癌的进展,如肺肿瘤生长和/或转移(因此表示不良预后);而一个或多个生物标记差异表达程度的降低(即在随后的生物标记测试中,个体中的表达水平趋向或接近“正常”表达水平)可以指示肺癌的缓解,如肺肿瘤缩小(因此提示良好或较好的预后)。相似地,在肺癌治疗期间一个或多个生物标记差异表达的程度增加(即在随后的生物标记测试中,个体中的表达水平进一步远离“正常”表达水

平)可以指示肺癌的进展,并因此表示所述治疗是无效的;而在肺癌治疗期间一个或多个生物标记的差异表达降低可以指示肺癌的缓解,并因此表示该治疗是成功的。此外,在个体看起来已经治愈肺癌之后一个或多个生物标记的差异表达的增加或降低可指示肺癌的复发。在这种情况下,例如可以在早期对个体重新启动治疗(或者如果个体维持治疗,则修改治疗方案以增加剂量和/或频率),否则直至晚期还未检测到肺癌的复发。此外,个体中一个或多个生物标记的差异表达水平可以预测个体对特定治疗剂的应答。在监测肺癌复发或进展中,生物标记表达水平的改变可以指示需要重复成像(如重复CT扫描),例如来确定肺癌活性或确定需要改变治疗方案。

[0205] 本文所述的任何生物标记的检测可以特别地在肺癌治疗后使用或者与肺癌治疗联合使用,如评价治疗的成功或者监测治疗后肺癌的缓解、复发和/或进展(包括转移)。肺癌治疗可以包括例如给予个体治疗剂、进行手术(如手术切除至少一部分肺肿瘤)、给予放疗或本领域所用的任何其他类型肺癌治疗方法以及这些治疗的任何组合。例如,任何生物标记可以在治疗后检测至少一次,或者可以在治疗后检测多次(如定期检测),或者可以在治疗之前和之后检测。个体中任何生物标记随时间的差异表达水平可以指示肺癌的进展、缓解或复发,其实例包括如下方面:生物标记的表达水平在治疗后与治疗前相比增加或降低;生物标记的表达水平在治疗后较晚时间点与治疗后较早时间点相比增加或降低;以及生物标记的表达水平在治疗后的一个时间点与该生物标记的正常水平相比不同。

[0206] 作为具体的实例,本文所述的任何生物标记的生物标记水平可以在手术前和手术后(例如手术后2-4周)的血清样品中确定。手术后样品与手术前样品相比生物标记表达水平的增加可以指示肺癌的进展(如不成功的手术);而手术后样品与手术前样品相比生物标记表达水平的降低可以指示肺癌的消退(如成功除去肺肿瘤的手术)。生物标记水平的相似分析可以在其他形式的治疗之前和之后进行,如在放疗或者给予治疗剂或癌症疫苗之前和之后进行。

[0207] 除了作为独立运行的诊断测试的生物标记水平测试之外,生物标记水平还可以联合SNP或者指示疾病易感性风险增加的其他遗传病变或变异性的确定来进行。(参见例如,Amos et al.,Nature Genetics 40,616-622(2009))。

[0208] 除了作为独立运行的诊断测试的生物标记水平测试之外,生物标记水平还可以联合CT筛查进行。例如,生物标记可以促进实施CT筛查的医学和经济理由,例如筛查面临肺癌风险的大量无临床症状群体(如吸烟者)。例如,生物标记水平的“CT前”测试可以用于分类CT筛查的高危个体,如基于个体生物标记水平来鉴定面临肺癌最高危险并且应当优先进行CT筛查的那些个体。如果实施CT测试,则可以测量一个或多个生物标记的生物标记水平(如通过血清或血浆样品的适配体测定来确定),并且可以联合额外的生物学信息来评价诊断评分(如通过CT测试确定肿瘤参数),以增强单独的CT或生物标记测试的阳性预测值(PPV)。确定生物标记水平的“CT后”适配体组可以用于确定通过CT(或其他成像方式)观察到的肺结节是恶性或良性的似然性。

[0209] 本文所述的任何生物标记的检测可以用于CT后测试。例如,生物标记测试可以显著消除或降低单独CT的假阳性测试数目。此外,生物标记检测可以促进患者的治疗。例如,如果肺结节的大小小于5mm,则生物标记测试的结果可以使患者在更早的时间从“观察和等待”进展至活组织检查;如果肺结节为5-9mm,则生物标记测试可以排除对假阳性扫描使用

活组织检查或胸廓切开术;以及如果肺结节大于10mm,则生物标记测试可以排除具有良性小结的这些患者亚群的手术。基于生物标记测试在一些患者中排除活组织检查的需求是有益的,因为存在与小结活组织检查相关的显著发病率以及难以根据小结的位置获得小结组织。相似地,在一些患者中排除手术的需要,如其小结实际上是良性的那些患者,会避免与手术相关的不必要风险和花费。

[0210] 除了联合CT筛查测试生物标记水平之外(如联合在CT扫描上观察到的肺结节的大小或其他特征评价生物标记水平),关于生物标记的信息还可以联合其他类型的数据进行评价,特别是指示个体患有肺癌风险的数据(如患者临床史、症状、癌症家族史、诸如个体是否是吸烟者的风险因素和/或其他生物标记的状况等)。这些不同数据可以通过自动化方法评价,如计算机程序/软件,其可以在计算机或其他设备/装置中实施。

[0211] 任何所述生物标记还可以用于成像测试。例如,显像剂可以与任何所述生物标记偶联,这可以用于辅助肺癌诊断、监测疾病进展/缓解或转移、监测疾病复发或者监测对治疗的应答等。

[0212] 生物标记和生物标记值的检测和确定

[0213] 本文所述的生物标记的生物标记值可以使用任何已知的分析方法来检测。在一实施方案中,生物标记值使用捕获试剂(capture reagent)检测。如本文所用,“捕获剂(capture agent)”或“捕获试剂”指能够特异性结合生物标记的分子。在许多实施方案中,捕获试剂可以在溶液中暴露于生物标记,或者可以暴露于生物标记,同时该捕获试剂固定在固体支持物上。在其他实施方案中,捕获试剂含有与固体支持物上的第二特征反应的特征。在这些实施方案中,捕获试剂可以在溶液中暴露于生物标记,然后该捕获试剂上的特征可以联合固体支持物上的第二特征来将所述生物标记固定在固体支持物上。捕获试剂基于进行的分析类型加以选择。捕获试剂包括但不限于适配体、抗体、adnectin、锚蛋白、其他抗体模拟物(mimetic)及其他蛋白支架、自身抗体、嵌合物、小分子、F(ab')₂片段、单链抗体片段、Fv片段、单链Fv片段、核酸、凝集素、配体-结合受体、affybodies、纳米抗体(nanobodies)、印迹聚合物(imprinted polymer)、高亲合性多聚体(avimer)、肽模拟物(peptidomimetic)、激素受体、细胞因子受体及合成受体以及这些物质的修饰物和片段。

[0214] 在一些实施方案中,生物标记值使用生物标记/捕获试剂复合物来检测。

[0215] 在其他实施方案中,生物标记值得自生物标记/捕获试剂复合物,并且例如作为生物标记/捕获试剂相互作用之后的反应结果间接检测,但是依赖于生物标记/捕获试剂复合物的形成。

[0216] 在一些实施方案中,生物标记值从生物样品中的生物学标记直接检测。

[0217] 在一实施方案中,生物标记使用多重形式检测,这允许在生物样品中同时检测两个或更多个生物标记。在多重形式的一实施方案中,捕获试剂直接或间接、共价或非共价地固定在固体支持物上分散的位置。在另一实施方案中,多重形式使用分离的固体支持物,其中每个固体支持物具有与该固体支持物相关的独特捕获试剂,例如量子点。在另一实施方案中,单独的装置用于检测生物样品中待检测的多个生物标记的每一个。可以配置单独的装置以允许同时处理生物样品中的每个生物标记。例如,可以使用微量滴定板,由此该板中的每个孔用于独特地分析生物样品中待检测的多个生物标记之一。

[0218] 在一个或多个前述实施方案中,可以使用荧光标签(tag)来标记生物标记/捕获复

合物的组分以允许检测生物标记值。在许多实施方案中,使用已知技术可以将荧光标记(fluorescent label)与对本文所述的任何生物标记特异性的捕获试剂偶联,然后该荧光标记可以用于检测相应的生物标记值。合适的荧光标记包括稀土元素螯合物、荧光素及其衍生物、罗丹明及其衍生物、丹磺酰、别藻蓝蛋白、PBXL-3、Qdot 605、丽丝胺(Lissamine)、藻红蛋白、德克萨斯红及其他这样的化合物。

[0219] 在一实施方案中,荧光标记是荧光染料分子。在一些实施方案中,荧光染料分子包括至少一个取代的吲哚环(indolium ring)体系,其中吲哚环的3-碳上的取代基含有化学反应性基团或偶联的物质。在一些实施方案中,染料分子包括AlexaFluor分子,例如AlexaFluor 488、AlexaFluor 532、AlexaFluor 647、AlexaFluor 680或AlexaFluor 700。在其他实施方案中,染料分子包括第一类型和第二类型的染料分子,如两种不同的AlexaFluor分子。在其他实施方案中,染料分子包括第一类型和第二类型的染料分子,并且两种染料分子具有不同的发射光谱。

[0220] 荧光可以用与大范围的测定形式相容的许多仪器测量。例如,已经设计了分光荧光计来分析微量滴定板、显微镜载玻片、印刷阵列(printed array)、小杯等。参见Principles of Fluorescence Spectroscopy, by J.R.Lakowicz, Springer Science+Business Media, Inc., 2004; Bioluminescence&Chemiluminescence: Progress&Current Applications; Philip E.Stanley and Larry J.Kricka editors, World Scientific Publishing Company, January 2002。

[0221] 在一个或多个前述实施方案中,化学发光标签可以任选地用于标记生物标记/捕获复合物的组分以允许检测生物标记值。合适的化学发光材料包括任何草酰氯、Rodamin 6G、Ru(bipy)₃²⁺、TMAE(四三(二甲基氨基)乙烯(tetrakis(dimethylamino)ethylene))、连苯三酚(1,2,3-三羟基苯(1,2,3-trihydroxybenzene))、光泽精、过氧草酸酯(peroxyoxalate)、芳基草酸酯、吖啶酯(acridinium ester)、二氧杂环丁烷(dioxetane)等。

[0222] 仍然在其他实施方案中,检测方法包括酶/底物组合,其产生对应于生物标记值的可检测信号。通常,酶催化生色底物的化学改变,这种改变可以使用多种技术测量,包括分光光度法、荧光及化学发光。合适的酶包括例如萤光素酶、萤光素、苹果酸脱氢酶、脲酶、辣根过氧化物酶(HRPO)、碱性磷酸酶、β-半乳糖苷酶、葡糖淀粉酶、溶菌酶、葡萄糖氧化酶、半乳糖氧化酶及葡萄糖-6-磷酸脱氢酶、尿酸氧化酶、黄嘌呤氧化酶、乳过氧化物酶、微过氧化物酶等。

[0223] 仍然在其他实施方案中,检测方法可以是产生可测量信号的荧光、化学发光、放射性核素或酶/底物组合的组合。多种方式的信号在生物标记测定形式中可以具有独特且有利的特征。

[0224] 更特别地,本文所述的生物标记的生物标记值可以使用已知的分析方法来检测,包括单重适配体测定、多重适配体测定、单重或多重免疫测定、mRNA表达谱、miRNA表达谱、质谱分析、组织学/细胞学方法等,这在下文中详细地描述。

[0225] 使用基于适配体的测定确定生物标记值

[0226] 检测和定量生物学样品及其他样品中生理学上有意义的分子的测定在科学研究和卫生保健领域是重要的工具。一类这样的测定包括使用包含固定在固体支持物上的一个

或多个适配体的微阵列。所述适配体各自能够以高特异性方式和非常高的亲和力结合靶分子。参见例如题为“核酸配体”的美国专利第5,475,096号,还参见例如美国专利第6,242,246号、美国专利第6,458,543号和美国专利第6,503,715号,这些专利的题目均是“核酸配体诊断生物芯片”。一旦使微阵列与样品接触,则适配体结合所述样品中存在的它们各自的靶分子,从而允许确定对应于生物标记的生物标记值。

[0227] 如本文所用,“适配体”指对靶分子具有特异性结合亲和力的核酸。应当了解到亲和相互作用的问题关键是程度;然而在本文中,适配体对其靶标的“特异性结合亲和力”指适配体通常以与其结合测试样品中其他组分的亲和力相比更高层次的亲和力结合其靶标。“适配体”是一种类型或物种的核酸分子的一系列拷贝,其具有特定的核苷酸序列。适配体可以包含任何合适数目的核苷酸,包括任何数目的化学修饰的核苷酸。“适配体”指多于一个的这种系列的分子。不同的适配体可以具有相同或不同数目的核苷酸。适配体可以是DNA或RNA或化学修饰的核酸,并且可以是单链、双链的或者含有双链区,以及可以包含高级结构。适配体还可以是光适配体(photoaptamer),其中该适配体中包含光反应性或化学反应性官能团以允许其与其对应靶标共价连接。本文公开的任何适配体方法可以包括使用特异性结合相同靶分子的两种或更多种适配体。如下文进一步描述,适配体可以包含标签。如果适配体包含标签,则该适配体的所有拷贝不需要具有相同的标签。此外,如果不同的适配体各自包含标签,则这些不同的适配体可以具有相同的标签或者不同的标签。

[0228] 适配体可以使用任何已知方法鉴定,包括SELEX方法。一旦鉴定,则可以根据任何已知方法制备或合成适配体,这些已知方法包括化学合成方法和酶促合成方法。

[0229] 术语“SELEX”和“SELEX方法”在本文中可互换使用,通常指(1)与(2)的组合,其中(1)是选择以期望的方式与靶分子相互作用的适配体,例如以高亲和力结合蛋白,(2)是扩增那些选择的核酸。SELEX方法可以用于鉴定对特定靶标或生物标记具有高亲和力的适配体。

[0230] SELEX通常包括制备核酸的候选混合物;使所述候选混合物与期望的靶分子结合以形成亲和复合物;分离所述亲和复合物与未结合的候选核酸;使核酸与所述亲和复合物分开并分离所述核酸;纯化所述核酸;以及鉴定特异性适配体序列。所述方法可以包括多次循环以进一步精制所选适配体的亲和力。所述方法可以包括在该方法的一个或多个点的扩增步骤。参见例如题为“核酸配体”的美国专利第5,475,096号。SELEX方法可以用于产生与适配体的靶标共价结合的适配体,以及与适配体的靶标非共价结合的适配体。参见例如题为“通过指数富集的核酸配体的系统进化:Chemi-SELEX”的美国专利第5,705,337号。

[0231] SELEX方法可以用于鉴定含有修饰的核苷酸的高亲和力适配体,所述修饰的核苷酸赋予该适配体改善的特征,例如改善的体内稳定性或改善的递送特征。此类修饰的实例包括核糖和/或磷酸和/或碱基位置的化学取代。通过SELEX方法鉴定的含有修饰的核苷酸的适配体描述于题为“含有修饰的核苷酸的高亲和力核酸配体”的美国专利第5,660,985号,其描述了含有在嘧啶的5'-和2'-位置处经化学修饰的核苷酸衍生物的寡核苷酸。见上文,美国专利第5,580,737号描述了高特异性适配体,其含有用2'-氨基(2'-NH₂)、2'-氟(2'-F)和/或2'-O-甲基(2'-OMe)修饰的一个或多个核苷酸。还参见题为“SELEX和PHOTOSELEX”的美国专利申请公开20090098549,其描述了具有扩展的物理和化学性质的核酸文库及其在SELEX和photoSELEX中的用途。

[0232] SELEX还可以用于鉴定具有期望的解离速率(off-rate)特征的适配体。参见题为“产生具有改善的解离速率的适配体的方法”的美国专利申请公开20090004667,其描述了产生可以结合靶分子的适配体的改进SELEX方法。描述了产生与各自的靶分子具有较慢解离速率的适配体和光适配体的方法。所述方法包括使候选混合物与靶分子接触;允许形成核酸-靶标复合物;以及进行缓慢解离速率富集过程,其中具有快解离速率的核酸-靶标复合物解离并不再形成,而具有慢解离速率的复合物会保持完整。此外,所述方法包括在产生候选核酸混合物中使用修饰的核苷酸,以产生具有改善的解离速率性能的适配体。

[0233] 这种测定的变化使用包含光反应性官能团的适配体,这允许适配体与其靶分子共价结合或“光交联”。参见例如题为“核酸配体诊断生物芯片”的美国专利第6,544,776号。这些光反应性适配体也称作光适配体。参见例如美国专利第5,763,177号、美国专利第6,001,577号和美国专利第6,291,184号,所述专利的题目均是“通过指数富集的核酸配体的系统进化:核酸配体的光选择和溶液SELEX”;还参见例如题为“核酸配体的光选择”的美国专利第6,458,539号。在使微阵列与样品接触并使光适配体具有结合其靶分子的机会之后,将该光适配体光激活并洗涤固体支持物以除去任何非特异性结合的分子。可以使用严格洗涤条件,因为结合光适配体的靶分子由于该光适配体上光激活的官能团所产生的共价键而通常未被除去。在这种方式中,测定允许检测对应于测试样品中的生物标记的生物标记值。

[0234] 在这两种测定形式中,适配体在与样品接触之前固定在固体支持物上。然而,在某些情况下,在与样品接触之前固定适配体也许无法提供最佳的测定。例如,预固定适配体可能导致适配体与靶分子在固体支持物表面上的无效混合,这可能导致漫长的反应时间及因此延长的温育时间以允许适配体与其靶分子有效结合。此外,当光适配体用于测定并且取决于用作固体支持物的材料时,该固体支持物可能趋于分散或吸收用于实现光适配体与其靶分子之间的共价键形成的光。此外,根据所用的方法,结合适配体的靶分子的检测可能不准确,因为固体支持物的表面也可能暴露于且受所用的任何标记剂的影响。最后,适配体固定在固体支持物上通常包括在适配体暴露于样品之前的适配体制备步骤(即固定),这个制备步骤可能影响适配体的活性或功能性。

[0235] 还描述了适配体测定,其允许适配体在溶液中捕获其靶标,然后在检测之前使用设计为除去适配体-靶标混合物中特定组分的分离步骤(参见题为“测试样品的多重分析”的美国专利申请公开20090042206)。所述适配体测定方法允许检测和定量测试样品中的非核酸靶标(如蛋白靶标),这通过检测和定量核酸(即适配体)进行。所述方法产生核酸替代物(surrogate)(即适配体)以检测和定量非核酸靶标,由此允许包括扩增在内的许多核酸技术用于包括蛋白靶标在内的更大范围的期望靶标。

[0236] 可以构建适配体以促进从适配体生物标记复合物(或光适配体生物标记共价复合物)分离测定组分,以及允许分离适配体以进行检测和/或定量。在一实施方案中,这些构建体可以包含适配体序列中可裂解或可释放的元件。在其他实施方案中,可以在适配体中引入额外的官能性,例如标记的或可检测的组分、间隔组分或者特异性结合标签或固定元件。例如,适配体可以包含通过可裂解部分与适配体连接的标签、标记、分隔标记与可裂解部分的间隔组分。在一实施方案中,可裂解元件是光可裂解接头(linker)。光可裂解接头可以连接至生物素部分和间隔区段,可以包含NHS基团以用于胺的衍生化,以及可以用于在适配体中引入生物素基团,从而允许适配体在测定方法中较晚地释放。

[0237] 用溶液中所有测定组分进行的均质测定在检测信号之前不需要分离样品与试剂。这些方法是快速且易于使用的。这些方法基于分子捕获或与其特异性靶标反应的结合试剂产生信号。对于肺癌,分子捕获试剂是适配体或抗体等,特异性靶标是表20的肺癌生物标记。

[0238] 在一实施方案中,一种信号产生方法利用由于荧光团-标记的捕获试剂与其特异性生物标记靶标的相互作用而导致的各向异性信号改变。当标记的捕获剂与其靶标反应时,增加的分子量导致附着于该复合物的荧光团的旋转运动变得更慢,从而改变各向异性值。通过监测各向异性改变,结合事件可以用于定量测量溶液中的生物标记。其他方法包括荧光偏振测定、分子信标方法、时间分辨荧光猝灭法、化学发光、荧光共振能量转移等。

[0239] 可以用于检测对应于生物学样品中生物标记的生物标记值的基于溶液的示例性适配体测定包括如下步骤:(a)通过使所述生物学样品与适配体接触来制备混合物,所述适配体包含第一标签并具有对所述生物标记的特异性亲和力,其中当所述样品中存在所述生物标记时形成适配体亲和复合物;(b)使所述混合物暴露于包含第一捕获元件的第一固体支持物,并且允许所述第一标签与所述第一捕获元件结合;(c)除去未与所述第一固体支持物结合的混合物的任何组分;(d)使第二标签附着于所述适配体亲和复合物的生物标记组分;(e)从所述第一固体支持物释放所述适配体亲和复合物;(f)使释放的适配体亲和复合物暴露于包含第二捕获元件的第二固体支持物,并且允许所述第二标签与所述第二捕获元件结合;(g)通过分离未复合的适配体与所述适配体亲和复合物来从所述混合物除去任何未复合的适配体;(h)从固体支持物洗脱适配体;以及(i)通过检测所述适配体亲和复合物的适配体组分来检测所述生物标记。

[0240] 使用免疫测定确定生物标记值

[0241] 免疫测定方法基于抗体与其对应靶标或分析物的反应,并且根据特定测定形式可以检测样品中的分析物。为了改进基于免疫反应性的测定方法的特异性和灵敏性,通常由于单克隆抗体的特异性表位识别而使用单克隆抗体。多克隆抗体由于其与单克隆抗体相比增加的靶标亲和力而成功地用于各种免疫测定。免疫测定已经设计为用于大范围生物学样品基质。免疫测定形式已经设计为提供定性、半定量及定量结果。

[0242] 定量结果通过使用已知浓度的待检测的特定分析物产生的标准曲线来产生。将来自未知样品的应答或信号在标准曲线上作图,并确定该未知样品中对应于靶标的量或值。

[0243] 已经设计了许多免疫测定形式。ELISA或EIA可以定量检测分析物。这种方法依赖于标记对分析物或抗体的附着,标记组分直接或者间接包括酶。ELISA测试可以设计为直接、间接、竞争性或者夹心检测分析物。其他方法依赖于标记,如放射性同位素(I^{125})或荧光。其他技术包括例如凝集反应、浊度测定法、比浊法、蛋白印迹、免疫沉淀、免疫细胞化学、免疫组织化学、流式细胞术、Luminex测定等(参见ImmunoAssay:A Practical Guide, edited by Brian Law, published by Taylor&Francis, Ltd., 2005 edition)。

[0244] 示例性测定形式包括酶联免疫吸附测定(ELISA)、放射性免疫测定、荧光、化学发光以及荧光共振能量转移(FRET)或时间分辨的-FRET(TR-FRET)免疫测定。检测生物标记的方法的实例包括生物标记免疫沉淀及随后允许辨别大小和肽水平的定量方法,如凝胶电泳、毛细管电泳、平面电色谱等。

[0245] 检测和/或定量可检测标记或信号产生材料的方法取决于所述标记的性质。由合

适的酶催化的反应产物(其中所述可检测标记是酶,见上文)可以是但不限于荧光、发光或放射性的,或者它们可以吸收可见光或紫外光。适合于检测这样的可检测标记的检测仪的实例包括但不限于X光照片、放射性计数器、闪烁计数器、分光光度计、比色计、荧光计、发光计和光密度计。

[0246] 可以通过允许适当准备、处理和分析反应的任何方式来进行任何检测方法。这可以例如在多孔测定板(如96孔或者384孔)中进行,或者使用任何合适的阵列或者微阵列进行。可以人工或自动化制备各种试剂的储液,使用能够检测可检测标记的可商购的分析软件、机器人技术和检测仪器自动化进行所有随后的移液、稀释、混合、分配、洗涤、温育、样品读取、数据收集和分析。

[0247] 使用基因表达谱确定生物标记值

[0248] 测量生物学样品中的mRNA可以用作检测该生物学样品中相应的蛋白水平的替代。因此,本文所述的任何生物标记或生物标记的组还可以通过检测适当的RNA来检测。

[0249] mRNA表达水平通过逆转录定量聚合酶链式反应(RT-PCR及随后的qPCR)测量。RT-PCR用于从mRNA产生cDNA。cDNA可用于qPCR测定以随DNA扩增过程的进展而产生荧光。通过与标准曲线比较,qPCR可以产生绝对测量度,如每细胞的mRNA拷贝数。RNA印迹、微阵列、Invader测定以及与毛细管电泳组合的RT-PCR全部已经用于测量样品中mRNA的表达水平(参见Gene Expression Profiling:Methods and Protocols,Richard A.Shimkets, editor,Humana Press,2004)。

[0250] miRNA分子是小RNA,其不编码但是可以调节基因表达。适合测量mRNA表达水平的任何方法均可以用于相应的miRNA。最近,许多实验室已经研究了miRNA作为疾病的生物标记的用途。许多疾病涉及广泛的转录调节,并且毫不意外地发现miRNA可以作为生物标记。miRNA浓度与疾病之间的关联通常不如蛋白水平与疾病之间的关联明确,但是miRNA生物标记值可能是重要的。当然,随着疾病期间任何RNA的不同表达,开发体外诊断产品所面临的问题包括需要miRNA在患病细胞中存活及易于提取以进行分析,或者miRNA被释放进入血液或其他基质中,在此它们必须存活足够长的时间以进行测量。蛋白生物标记具有相似的要求,尽管许多潜在的蛋白生物标记以旁分泌方式在疾病期间于病变和功能部位有意地分泌。许多潜在的蛋白生物标记设计为在合成那些蛋白的细胞外起作用。

[0251] 使用体内分子成像技术检测分子标记

[0252] 任何所述的生物标记(见表20)还可以用于分子成像测试。例如,显像剂可以与任何所述生物标记偶联,这可以用于辅助肺癌诊断、监测疾病进展/缓解或转移、监测疾病复发或者监测对治疗的应答等。

[0253] 体内成像技术提供了用于确定个体体内特定疾病状态的非侵入性方法。例如,身体的所有部分或者甚至整个身体均可以作为三维图像观察,从而提供关于身体内形态学和结构的有价值的信息。这样的技术可以与检测本文所述的生物标记组合以提供关于个体的癌症状态,特别是肺癌状态的信息。

[0254] 体内分子成像技术的应用由于该技术的各种进展而得以扩展。这些进展包括新造影剂或标记的开发,如放射性标记和/或荧光标记,其可以在身体内提供强信号;以及开发更强的新成像技术,其可以从身体外部检测和分析这些信号,并且具有足够的灵敏性和精确度以提供有用的信息。造影剂可以在适当的成像系统中观察,从而提供所述造影剂所处

位置的身体部分或多个部分的图像。造影剂可以与捕获试剂结合或缔合,例如适配体或抗体,例如和/或结合或缔合肽或蛋白,或寡核苷酸(例如为了检测基因表达),或者复合物,所述复合物含有任何这些物质及一种或多种大分子和/或其他颗粒形式。

[0255] 造影剂还是可以用于成像的放射性原子的特征。对于闪烁照相研究合适的放射性原子包括镓-99m或碘-123。其他易于检测的部分包括例如磁共振成像(MRI)的自旋标记物,如碘-123、碘-131、铟-111、氟-19、碳-13、氮-15、氧-17、钆、锰或铁。这样的标记为本领域熟知,并且可以由本领域技术人员容易地选择。

[0256] 标准成像技术包括但不限于磁共振成像、计算机断层扫描、正电子发射断层扫描(PET)、单光子发射计算机断层扫描(SPECT)等。对于诊断性体内成像,可用的检测设备的类型是选择指定造影剂的主要因素,如用于靶标(蛋白、mRNA等)的指定放射性核素和特定生物标记。所选的放射性核素通常具有通过指定类型设备可检测的衰变类型。此外,当选择用于体内诊断的放射性核素时,其半衰期应当足够长以允许在靶组织最大吸收时进行检测,但是也应当足够短,以最小化宿主所受的有害辐射。

[0257] 示例性成像技术包括但不限于PET和SPECT,这是将放射性核素全身(synthetically)或局部地给予个体的成像技术。随后,随时间测量放射性示踪剂的吸收,并用于获得关于靶向的组织与生物标记的信息。由于所用的特定同位素的高能(γ -射线)发射以及用于检测它们的设备的灵敏性和完善(sophistication),可以从身体外部推导出放射性的二维分布。

[0258] PET中常用的正电子发射核素包括例如碳-11、氮-13、氧-15和氟-18。通过电子捕获和/或 γ -发射衰变的同位素用于SPECT中,并且包括例如碘-123和镓-99m。用镓-99m标记氨基酸的示例性方法是在螯合前体的存在下还原高镓酸盐离子以形成不稳定的镓-99m-前体配合物,其又与双官能修饰的趋化肽的金属结合基团反应,形成镓-99m-趋化肽偶联物。

[0259] 抗体常用于这样的体内成像诊断方法。用于体内诊断的抗体的制备和用途为本领域熟知。特异性结合表20的任何生物标记的标记的抗体可以注入疑似患有某种类型癌症(如肺癌)的个体,并根据所用的特定生物标记的可检测性来诊断或评价所述个体的疾病状态。如上文所述,使用的标记根据所用的成像形式来选择。标记的定位允许确定癌症的扩散。器官或组织内标记的量还允许确定该器官或组织中癌症的存在与否。

[0260] 相似地,适配体可以用于这样的体内成像诊断方法。例如,用于鉴定表20所述的特定生物标记的适配体(并且因此特异性结合该特定生物标记)可以适当地进行标记并注入疑似患有肺癌的个体,并根据该特定生物标记的可检测性来诊断或评价所述个体的肺癌状态。如上文所述,使用的标记根据所用的成像形式来选择。标记的定位允许确定癌症的扩散。器官或组织内标记的量还允许确定该器官或组织中癌症的存在与否。适配体定向的显像剂与其他显像剂相比可以具有关于组织渗透、组织分布、动力学、消除、效力和选择性方面独特且有利的特征。

[0261] 这样的技术还可以任选地用标记的寡核苷酸进行,例如通过用反义寡核苷酸成像检测基因表达。这些方法用于原位杂交,例如用荧光分子或放射性核素作为标记。检测基因表达的其他方法包括例如检测报道基因的活性。

[0262] 另一种常见类型的成像技术是光学成像,其中对象体内的荧光信号通过所述对象体外的光学设备检测。这些信号可以是由于实际的荧光和/或生物发光。光学检测设备灵敏

性的改进增加了光学成像在体内诊断测定中的应用。

[0263] 体内分子生物标记成像的用途日益增加,包括临床试验,例如在新癌症疗法临床试验中更快速地测量临床效力,和/或避免对诸如多发性硬化的那些疾病的长期安慰剂治疗,其中这样的长期治疗可能被认为是存在伦理问题。

[0264] 关于其他技术的综述,参见N.Blow, *Nature Methods*, 6, 465-469, 2009。

[0265] 使用组织学/细胞学方法确定生物标记值

[0266] 对于肺癌的评价,许多组织样品可用于组织学或细胞学方法。样品选择取决于原发肿瘤位置和转移的部位。例如,支气管内和经支气管活组织检查、细针抽吸、切割针及核心活组织检查可以用于组织学。支气管洗涤(washing)和刷检(brushing)、胸膜抽吸以及痰可以用于细胞学。虽然细胞学分析仍用于诊断肺癌,但是已知组织学方法提供癌症检测更好的灵敏性。本文鉴定的在肺癌个体中表现出上调的任何生物标记(见表19)可以用于染色组织学样本作为疾病的指征。

[0267] 在一实施方案中,对于相应的生物标记是特异性的一种或多种捕获试剂用于肺细胞样品的细胞学评价,并且可以包括如下一个或多个方面:收集细胞样品、固定细胞样品、脱水、透明(clearing)、将细胞样品固定在显微镜载玻片上、使细胞样品透化、分析物检索处理、染色、脱色、洗涤、封闭以及在缓冲溶液中与一种或多种捕获试剂反应。在另一实施方案中,细胞样品从细胞块(cell block)中产生。

[0268] 在另一实施方案中,对于相应的生物标记是特异性的一种或多种捕获试剂用于肺组织样品的组织学评价,并且可以包括如下一个或多个方面:收集组织样本、固定组织样品、脱水、透明、将组织样品固定在显微镜载玻片上、使组织样品透化、分析物检索处理、染色、脱色、洗涤、封闭、再水合以及在缓冲溶液中与一种或多种捕获试剂反应。在另一实施方案中,固定和脱水用冷冻代替。

[0269] 在另一实施方案中,使对于相应的生物标记是特异性的一种或多种适配体与组织学或细胞学样品反应,并且可以作为核酸扩增方法中的核酸靶标。合适的核酸扩增方法包括例如PCR、q- β 复制酶、滚环扩增、链置换、解旋酶依赖性扩增、环介导的等温扩增、连接酶链式反应以及限制和环化辅助的滚环扩增。

[0270] 在一实施方案中,将对于用于组织学或细胞学评价的相应生物标记是特异性的一种或多种捕获试剂在缓冲溶液中混合,所述缓冲溶液可以包含任何如下成分:封闭材料、竞争剂、去污剂、稳定剂、载体核酸、聚阴离子材料等。

[0271] “细胞学方案”通常包括样品收集、样品固定(fixation)、样品固定(immobilization)和染色。“细胞制备”可以包括样品收集后的一些处理步骤,包括使用一种或多种慢解离速率的适配体来染色制备的细胞。

[0272] 样品收集可以包括直接将样品置于未处理的转运容器中,将样品置于含有一些类型介质的转运容器中,或者将样品直接置于玻片上(固定)而不进行任何处理或固定。

[0273] 样品固定可以通过将一部分收集的样本涂在用聚赖氨酸、明胶或硅烷处理的载玻片上而改进。玻片可以通过在玻片上涂有薄且均匀的细胞层而制备。通常采取小心操作以使最小化机械扭转和干燥假象。液体样本可以通过细胞块方法处理。或者,液体样本可以与固定溶液在室温下1:1混合约10分钟。

[0274] 细胞块可以从剩余的积液、痰、尿液沉淀、胃肠液、细胞刮取物或细针抽吸物中制

备。通过离心或膜过滤浓缩或压实细胞。已经开发了许多细胞块制备方法。代表性方法包括固定的沉淀、细菌琼脂或膜过滤方法。在固定的沉淀方法中,将细胞沉淀与诸如鲍音液(Bouins)、苦味酸或缓冲的福尔马林的固定剂混合,然后将混合物离心以沉淀固定的细胞。除去上清,尽可能完全地干燥细胞团块(pellet)。收集团块并包在镜头纸中,然后置于组织盒(tissue cassette)中。将组织盒置于含其他固定剂的罐子中,作为组织样品进行处理。琼脂方法与上述方法非常相似,只是取出团块并在纸巾上干燥,然后切成两半。将切面置于载玻片上一滴熔化的琼脂中,然后将该团块用琼脂包被,保证琼脂中无气泡形成。使琼脂变硬,然后除去任何过多的琼脂。将其置于组织盒中,完成组织处理。或者,可以将团块直接悬浮于在65℃的2%液体琼脂中并离心样品。使琼脂细胞团块在4℃下固化1小时。可以从离心管中取出固体琼脂并切成两半。将琼脂包在滤纸中,然后置于组织盒中。从这点开始的处理与上述方法相同。在任何这些方法中可以用膜过滤代替离心。任何这些方法均可以用于产生“细胞块样品”。

[0275] 细胞块可以使用专门的树脂制备,包括Lowicryl树脂、LR White、LR Gold、Unicryl和MonoStep。这些树脂具有低粘度,并且可以在低温下及用紫外(UV)光聚合。包埋方法依赖于在脱水期间逐渐冷却样品,将样品转移至树脂以及于最终低温下在合适UV波长处聚合细胞块。

[0276] 细胞块切片可以用苏木精-伊红染色以进行细胞形态学检查,而其他切片用于特异性标记检查。

[0277] 无论方法是细胞学方法或组织学方法,可以在进一步处理之前将样品固定以防止样品降解。这种方法称作“固定”,并且描述了可以互换使用的许多材料和方法。基于待检测的靶标和待分析的特定细胞/组织类型,根据经验最佳地选择样品固定方案和试剂。样品固定依赖于试剂,如乙醇、聚乙二醇、甲醇、福尔马林或异丙醇。样品应当尽可能在收集及附着在玻片上后很快固定。然而,选择的固定剂可以在各种分子靶标中引入结构改变,这使得随后更难以检测。固定(fixation)和固定(immobilization)方法及其顺序可以改变细胞的外观,并且这些改变必须是由细胞学技术人员预期及认可的。固定剂可以导致某些类型细胞收缩,并且导致细胞质出现颗粒或网状物。许多固定剂通过使细胞组分交联而起作用。这可以破坏或改变特异性表位,产生新表位,导致分子缔合以及降低膜通透性。福尔马林固定是一种最常用的细胞学/组织学方法。福尔马林在相邻蛋白之间或在蛋白内形成甲基桥。沉淀或凝固也用于固定,乙醇常用于这种类型的固定。交联与沉淀的组合也可以用于固定。牢固的固定方法在保留形态学信息方面是最佳的,而较弱的固定方法对于保留分子靶标方面是最佳的。

[0278] 代表性固定剂是50%无水乙醇、2mM聚乙二醇(PEG)、1.85%甲醛。这种制剂的变化包括乙醇(50%-95%)、甲醇(20%-50%)以及仅福尔马林(甲醛)。另一种常用的固定剂是2%PEG 1500、50%乙醇以及3%甲醇。将玻片在室温下置于固定剂中约10-15分钟,然后取出并干燥。一旦玻片被固定,可以用诸如PBS的缓冲溶液对其进行漂洗。

[0279] 许多染料可以用于差异地突出和反差或“染色”细胞、亚细胞和组织特征或形态学结构。苏木精(hematoxylin)用于将核染色为蓝色或黑色。橘黄G-6和天青伊红(Eosin Azure)均将细胞质染色。橘黄G将含有角蛋白和糖原的细胞染成黄色。伊红Y用于将核仁、纤毛、红细胞和表面上皮扁平细胞染色。罗曼诺夫斯基(Romanowsky)染色用于空气干燥的玻

片,并且可以用于增强复型及区分细胞外与细胞质内材料。

[0280] 染色方法可以包括增加细胞对染色的通透性的处理。用去污剂处理细胞可以用于增加通透性。为了增加细胞和组织通透性,可以将固定的样品用溶剂、皂苷类或者非离子型去污剂进一步处理。酶促消化还可以改进组织样品中特异性靶标的可接近性。

[0281] 染色后,使用渐增的醇浓度进行连续醇漂洗将样品脱水。最终的洗涤使用二甲苯或诸如柑桔萜的二甲苯取代物,其具有接近在载玻片上应用的盖玻片的折射率。这个最后的步骤称作透明。一旦使样品脱水及透明,应用封固剂。所选的封固剂具有接近玻璃的折射率,并且能够使盖玻片与载玻片粘合。其还抑制细胞样品另外的干燥、收缩或褪色。

[0282] 无论使用的染色或处理,对肺细胞学样本的最后评价通过一些类型显微镜检查进行以允许通过肉眼观察形态学并确定标记的存在与否。示例性显微镜检查方法包括明视野显微镜、相差显微镜、荧光显微镜和微分干涉相差显微镜方法。

[0283] 如果在检查后需要对样品进行次级测试,则可以除去盖玻片并对载玻片进行脱色。脱色包括使用用于染色该载玻片的最初未加入染料的原始溶剂系统,并以与原始染色程序相反顺序进行。脱色还可以通过将该载玻片浸泡在酸醇中直至细胞无色来完成。一旦无色,则将载玻片用水浴充分漂洗并进行第二染色程序。

[0284] 此外,通过使用特异性分子试剂,如抗体或者核酸探针或适配体,可以将特异性分子区分与细胞形态学分析组合。这改进了诊断细胞学的精确性。显微切割可以用于分离细胞的子集以进行另外的评价,特别是用于遗传学评价异常染色体、基因表达或突变。

[0285] 制备用于组织学评价的组织样品包括固定、脱水、浸润(infiltration)、包埋和切片。用于组织学的固定试剂与用于细胞学的固定试剂非常相似或相同,并且在以诸如个体蛋白的分子为代价的情况中具有相同的保持形态学特征的问题。如果组织样品不进行固定和脱水而是代之以冷冻然后在冷冻时切片可以节省时间。这是更温和的处理程序,并且可以保留更多的个体标记。然而,冷冻对于组织样品的长期保存不可接受,因为由于冰晶体的引入引起亚细胞信息丧失。冷冻组织样品中的冰也妨碍切片过程产生极薄的切片,并且因此可以丧失一些显微镜分辨力和亚细胞结构的图像。除了福尔马林固定之外,四氧化锇也用于固定和染色磷脂(膜)。

[0286] 组织的脱水是通过用渐增浓度的醇连续洗涤来完成。透明使用可以与醇和包埋材料混溶的材料,并且包括从50:50醇:澄清试剂开始至100%澄清试剂(二甲苯或二甲苯取代物)的逐步处理过程。浸润包括将组织与液体形式的包埋剂(温热的蜡,硝化纤维溶液)一起温育,首先是50:50包埋剂:澄清剂,随后是100%包埋剂。包埋通过将组织置于模具或盒中并充填熔化的包埋剂如蜡、琼脂或明胶来完成。使包埋剂硬化。然后将硬化的组织样品切成薄片以用于染色和随后的检查。

[0287] 在染色之前,将组织切片脱蜡并再水合。用二甲苯使切片脱蜡,可以更换一次或多次二甲苯,并通过在递减浓度的醇中连续洗涤而再水合。在脱蜡之前,可以将组织切片于约80℃下在载玻片上热固定约20分钟。

[0288] 激光捕获显微切割允许从组织切片分离细胞的子集以进行进一步分析。

[0289] 在细胞学中,为了增强显微特征的观察,可以将组织切片或薄片用各种染色方法染色。许多可商购的染色方法可以用于增强或鉴定特定的特征。

[0290] 为了进一步增加分子试剂与细胞学/组织学样品的相互作用,已经开发了许多“分

析物检索(analyte retrieval)”技术。第一种这样的技术使用高温加热固定的样品。这种方法也称作热诱导的表位检索或HIER。已经使用了许多加热技术,包括蒸汽加热、微波、高压蒸汽、水浴以及加压蒸煮或这些加热方法的组合。分析物检索溶液包括例如水、柠檬酸盐和普通盐水缓冲液。分析物检索的关键是高温的时间,但是较低温度进行较长时间也已经成功使用。分析物检索的另一关键是加热溶液的pH。据发现低pH提供最佳的免疫染色,但是也产生经常需要使用第二组织切片作为阴性对照的背景。无论缓冲液组成,使用高pH溶液通常获得最一致的益处(增加免疫染色而不增加背景)。对特异性靶标的分析物检索方法根据经验对使用加热的靶标、时间、pH和缓冲液组成的变量加以优化。使用微波分析物检索方法允许用抗体试剂顺序染色不同的靶标。但是在染色步骤之间获得抗体与酶复合物所需的时间也证实使细胞膜分析物降解。微波加热方法也改进原位杂交方法。

[0291] 为了开始分析物检索过程,首先将切片脱蜡并水合。然后将玻片置于平皿或罐子中的10mM柠檬酸钠缓冲液pH 6.0中。代表性程序使用1100W微波,以100%功率对玻片微波处理2分钟,随后在确保玻片保留覆盖于液体中之后使用20%功率对玻片微波处理18分钟。然后使玻片在敞口容器中冷却,随后用蒸馏水漂洗。HIER可以与酶促消化组合使用以改进靶标对免疫化学试剂的反应性。

[0292] 一种这样的酶促消化方案使用蛋白酶K。20 μ g/ml浓度的蛋白酶K在50mM Tris碱、1mM EDTA、0.5%Triton X-100、pH 8.0缓冲液中制备。该方法首先包括将切片在更换2次的二甲苯中脱蜡,每次5分钟。然后将样品在更换2次的100%乙醇中水合,每次3分钟,在95%和80%乙醇中水合,每次1分钟,然后在蒸馏水中漂洗。将切片用蛋白酶K工作溶液覆盖,于37 $^{\circ}$ C下在加湿室中温育10-20分钟(最佳温育时间可以根据组织类型和固定程度而变化)。将切片在室温下冷却10分钟,然后在PBS吐温(Tween) 20中漂洗2次2分钟。如果需要,可以将切片封闭以消除来自内源化合物和酶的潜在干扰。然后将切片用在一抗稀释缓冲液中适当稀释的一抗在室温下温育1小时或者在4 $^{\circ}$ C下温育过夜。然后将该切片用PBS吐温20漂洗2次2分钟。如果需要特定的应用,可以进行另外的封闭,随后用PBS吐温20再漂洗3次2分钟,然后最后完成免疫染色方案。

[0293] 在室温下用1%SDS简单处理也已经证实改进了免疫组织化学染色。分析物检索方法已经应用于玻片固定切片(slide mounted section)以及自由浮动切片(free floating section)。另一处理选择是将玻片置于pH 6.0的含有柠檬酸和0.1诺纳德(Nonident) P40的罐子中,并加热至95 $^{\circ}$ C。然后将该玻片用诸如PBS的缓冲溶液洗涤。

[0294] 对于组织的免疫学染色,可以通过将切片浸入诸如血清或脱脂奶粉的蛋白溶液中来封闭抗体与组织蛋白的非特异性结合。

[0295] 封闭反应可包括需要降低内源生物素的水平;消除内源电荷作用;失活内源核酸酶;和/或失活内源酶如过氧化物酶和碱性磷酸酶。内源核酸酶可以通过以下方式失活:用蛋白酶K降解;热处理;使用螯合剂,如EDTA或EGTA;引入载体DNA或RNA;用离液剂处理,如尿素、硫脲、盐酸胍、硫氰酸胍、高氯酸锂等或焦碳酸二乙酯。碱性磷酸酶可以通过用0.1N HCl在室温下处理5分钟或用1mM左旋咪唑处理而失活。过氧化物酶活性可以通过用0.03%过氧化氢处理来消除。内源生物素可以通过将玻片或切片在室温下浸入抗生物素蛋白(链霉抗生物素蛋白,可以取代中性链亲和素(neutravidin))溶液中至少15分钟来封闭。然后将玻片或切片在缓冲液中洗涤至少10分钟。这个步骤可以重复至少3次。然后将玻片或切片浸入

生物素溶液中10分钟。这个步骤可以重复至少3次,每次使用新鲜的生物素溶液。重复缓冲液洗涤程序。应当减少封闭方案以防止破坏所关注的细胞或组织结构或者靶标或多个靶标,但是可以组合一种或多种这样的方案以“封闭”玻片或切片,然后与一种或多种慢解离速率适配体反应。参见Basic Medical Histology:the Biology of Cells,Tissues and Organs,authored by Richard G.Kessel,Oxford University Press,1998。

[0296] 使用质谱方法确定生物标记值

[0297] 许多质谱仪的配制(configuration)可以用于检测生物标记值。一些类型的质谱仪可以获得或可以用各种配制生产。通常,质谱仪具有如下主要部件:样品入口、离子源、质量分析仪、检测仪、真空系统以及设备控制系统和数据系统。样品入口、离子源和质量分析仪的差异通常限定设备的类型及其能力。例如,入口可以是毛细管柱液体层析源,或者可以是直接探针或镜台(stage)如用于基质辅助激光解吸电离中。常用的离子源是例如电喷射,包括纳米喷射(nanospray)和微喷射(microspray);或者基质辅助激光解吸电离。常用的质量分析仪包括四极滤质器(quadrupole mass filter)、离子阱质量分析仪和飞行时间质量分析仪。其他质谱方法为本领域熟知(参见Burlingame et al.Anal.Chem.70:647R-716R (1998);Kinter and Sherman.Protein sequencing and identification using tandem mass spectrometry.New York:Wiley-Interscience(2000))。

[0298] 蛋白生物标记和生物标记值可以通过任何如下方式检测和测量:电喷射离子化质谱(ESI-MS)、ESI-MS/MS、ESI-MS/(MS)ⁿ、基质辅助激光解吸离子化飞行时间质谱(MALDI-TOF-MS)、表面增强激光解吸/离子化飞行时间质谱分析(SELDI-TOF-MS)、硅表面解吸/离子化(DIOS)、二次离子质谱(SIMS)、四极飞行时间(Q-TOF)、称作ultraflex III TOF/TOF的串联式飞行时间(TOF/TOF)技术、大气压化学离子化质谱(APCI-MS)、APCI-MS/MS、APCI-(MS)^N、大气压光电离质谱(APPI-MS)、APPI-MS/MS和APPI-(MS)^N、四极质谱、傅里叶变换质谱(FTMS)、定量质谱以及离子阱质谱。

[0299] 样品制备策略用于在对蛋白生物标记进行质谱表征及确定生物标记值之前标记和富集样品。标记方法包括但不限于用于相对和绝对定量的等量异位标签(iTRAQ)和在细胞培养中用氨基酸稳定同位素标记(SILAC)。在质谱分析之前用于选择性富集候选生物标记蛋白样品的捕获试剂包括但不限于适配体、抗体、核酸探针、嵌合物、小分子、F(ab')₂片段、单链抗体片段、Fv片段、单链Fv片段、核酸、凝集素、配体-结合受体、affybodies、纳米抗体、锚蛋白、结构域抗体、可变抗体支架(例如双抗体等)印刷的聚合物、高亲合性多聚体、肽模拟物、拟肽、肽核酸、苏糖核酸、激素受体、细胞因子受体及合成的受体以及这些物质的修饰和片段。

[0300] 前述测定允许检测可用于诊断肺癌的方法中的生物标记值,其中所述方法包括在来自个体的生物学样品中检测至少N个生物标记值,所述至少N个生物标记值每个对应于选自表18、20或21提供的生物标记的组的生物标记,其中如下文详述,利用生物标记值的分类指示所述个体是否患有肺癌。虽然某些所述肺癌生物标记可以单独用于检测和诊断肺癌,但是本文所述的方法还用于分组肺癌生物标记的多个子集,其各自用作三个或更多个生物标记的组。因此,本申请的各个实施方案提供了包含N个生物标记的组合,其中N是至少三个生物标记。在其他实施方案中,N选自2-86个生物标记中的任意数。应当理解N可以选自任何上述范围以及相似但更高级范围中的任意数。根据本文所述的任何方法,可以单独检测和

分类生物标记值,或者可以共同检测和分类生物标记值,例如以多重测定形式。

[0301] 在另一方面,本发明提供了检测肺癌不存在的方法,所述方法包括在来自个体的生物学样品中检测至少N个生物标记值,所述至少N个生物标记值每个对应于选自表18、20或21提供的生物标记的组的生物标记,其中如下文详述,生物标记值的分类指示所述个体中不存在肺癌。虽然某些所述肺癌生物标记可以单独用于检测和诊断不存在肺癌,但是本文所述的方法还用于分组肺癌生物标记的多个子集,其各自用作三个或更多个生物标记的组。因此,本申请的各个实施方案提供了包含N个生物标记的组合,其中N是至少三个生物标记。在其他实施方案中,N选自2-86个生物标记中的任意数。应当理解N可以选自任何上述范围以及相似但更高级范围中的任意数。根据本文所述的任何方法,可以单独检测和分类生物标记值,或者可以共同检测和分类生物标记值,例如以多重测定形式。

[0302] 生物标记分类及疾病评分计算

[0303] 给定诊断测试的生物标记“特征”含有标记的集合,每个标记在所关注群体中具有不同水平。在此方面,不同水平可以指针对两个或更多个组中个体的标记水平的不同平均值(mean),或者两个或更多个组中的不同的方差,或者这两者的组合。对于最简单形式的诊断测试,这些标记可以用于将来自个体的未知样品分配到两组中的一组中,疾病组或非疾病组。将样品分配于两个或更多个组中的一组称为分类,用于实现这种分配的程序称为分类器或分类方法。分类方法也可以称为评分方法。有许多分类方法可以用于从生物标记值的集合构建诊断分类器。通常,分类方法最容易用监督学习技术进行,其中用获得自希望区分的两个(或更多个,对于多个分类状态)不同组内的个体的样品收集数据集。因为每个样品所属的类别(组或群体)事先对于每个样品均是已知的,所以可以训练分类方法以获得期望的分类应答。还可以使用无监督学习技术来产生诊断分类器。

[0304] 开发诊断分类器的常用方法包括决策树;bagging+boosting+forests;基于规则推论的学习(rule inference based learning);Parzen窗方法(Parzen Windows);线性模型;逻辑;神经网络方法;无监督聚类;K-means;分级上升/下降(hierarchical ascending/descending);半监督学习;原型方法;近邻取样(nearest neighbor);核密度估计(kernel density estimation);支持向量机(support vector machine);隐马尔可夫模型(hidden Markov model);玻尔兹曼学习(Boltzmann Learning);并且分类器可以简单组合或者以最小化特定目标函数的方式组合。综述参见例如Pattern Classification,R.O.Duda,et al.,editors,John Wiley&Sons,2nd edition,2001;还参见The Elements of Statistical Learning-Data Mining,Inference,and Prediction,T.Hastie,et al.,editors,Springer Science+Business Media,LLC,2nd edition,2009;它们均整体援引加入本文。

[0305] 为了用监督学习技术产生分类器,获得称为训练数据的样品集合。在诊断测试的情况下,训练数据包括来自未知样品稍后会被分配的不同组(类别)的样品。例如,收集自对照群体的个体和特定疾病群体的个体的样品可以组成训练数据以开发可以分类未知样品(或者,更特别地,样品所来自的个体)为患有该疾病或无该疾病的分类器。从训练数据开发分类器已知为训练该分类器。分类器训练的具体细节取决于监督学习技术的性质。作为示例,训练朴素贝叶斯(naïve Bayesian)分类器的实例在下文进行描述(参见例如Pattern Classification,R.O.Duda,et al.,editors,John Wiley&Sons,2nd edition,2001;还参

见The Elements of Statistical Learning-Data Mining, Inference, and Prediction, T.Hastie, et al., editors, Springer Science+Business Media, LLC, 2nd edition, 2009)。

[0306] 因为通常在训练集中存在比样品多得多的潜在生物标记值,所以必须小心避免过拟合。当统计学模型描述随机误差或噪声而非潜在关系时发生过拟合。过拟合可以由各种方式避免,这包括例如限制开发分类器中使用的标记数目,假设标记应答互相独立,限制采用的潜在统计学模型的复杂性,以及保证潜在统计学模型符合数据。

[0307] 使用生物标记的集合开发诊断测试的说明性实例包括应用朴素贝叶斯分类器,这是一种基于贝叶斯(Bayes)定理的简单或然性分类器,具有生物标记的严格独立处理。每个生物标记由针对每种类别中测量的RFU值或log RFU(相对荧光单位)值的类别依赖性概率密度函数(pdf)描述。一个类别中的标记的集合的共同pdf(joint pdf)假定为每个生物标记的个体类别依赖性pdf的积。在此情况下训练朴素贝叶斯分类器意味着分配参数(“参数化”)以表征类别依赖性pdf。类别依赖性pdf的任何潜在模型均可以使用,但是模型应该通常符合在训练集中观察的数据。

[0308] 具体地,测量疾病类别中生物标记i的值 x_i 的类别依赖性概率写作 $p(x_i|d)$,并且观

察具有值 $\underline{x}=(x_1, x_2, \dots, x_n)$ 的n个标记的整体朴素贝叶斯概率写作 $p(\underline{x}|\underline{d})=\prod_{i=1}^n p(x_i|d)$ 其

中各个 x_i 是以RFU或log RFU表示的测量的生物标记水平。对于未知的分类分配通过以下方法来促进:对于相同测量值,计算与不患病(对照)的概率 $p(c|\underline{x})$ 相比的具有测量的 $\underline{x}$ 的患病概率 $p(d|\underline{x})$ 。这些概率的比率通过应用贝叶斯定理从类别依赖性pdf计算,即

$\frac{p(c|\underline{x})}{p(d|\underline{x})} = \frac{p(\underline{x}|c)(1-P(d))}{p(\underline{x}|d)P(d)}$ 其中P(d)为对测试合适的群体中疾病的发病率。对这一比率的两边

取对数并从以上代入朴素贝叶斯类别依赖性概率,获得 $\ln \frac{p(c|\underline{x})}{p(d|\underline{x})} = \sum_{i=1}^n \ln \frac{p(x_i|c)}{p(x_i|d)} + \ln \frac{(1-P(d))}{P(d)}$ 。

这种形式已知为log似然比,并简单地表示不患有特定疾病比患有所述疾病的log似然性,并且主要由n个单独生物标记的单独log似然比的总和组成。在其最简单的形式中,未知样品(或者,更特别地,样品所来自的个体)分类为不患有疾病,如果上述比率大于0;而分类为患有疾病,如果所述比率小于0。

[0309] 在一示例性实施方案中,类别依赖性生物标记pdf $p(x_i|c)$ 和 $p(x_i|d)$ 假定为在测

量的RFU值 x_i 中为正态分布或log正态分布,即 $p(x_i|c) = \frac{1}{\sqrt{2\pi}\sigma_{c,i}} e^{-\frac{(x_i-\mu_{c,i})^2}{2\sigma_{c,i}^2}}$ 并且对于 $p(x_i|d)$,

具有 $\mu_{d,i}$ 和 $\sigma_{d,i}^2$ 的相似表达。模型的参数化要求估计来自训练数据的每个类别依赖性pdf的两个参数,平均值 μ 和方差 δ^2 。这可以通过各种方式实现,包括例如最大似然估计、最小平方以及本领域技术人员已知的任何其他方法。将 $p(x_i|c)$ 和 $p(x_i|d)$ 的正态分布代入上文所定义的

log-似然比, 获得以下表达式: $\ln \frac{p(c|x)}{p(d|x)} = \sum_{i=1}^n \ln \frac{\sigma_{d,i}}{\sigma_{c,i}} - \frac{1}{2} \sum_{i=1}^n \left[\left(\frac{x_i - \mu_{c,i}}{\sigma_{c,i}} \right)^2 - \left(\frac{x_i - \mu_{d,i}}{\sigma_{d,i}} \right)^2 \right] + \ln \frac{(1-P(d))}{P(d)}$ 。

一旦 μ 和 σ^2 的集合已对来自训练数据的每个类别的每个pdf定义, 并且确定群体中的疾病发病率, 则贝叶斯分类器完全确定并且可以用于分类具有测量值 x 的未知样品。

[0310] 朴素贝叶斯分类器的性能取决于用于构建和训练分类器的生物标记的数目和质量。如以下实施例3所定义, 单个生物标记根据其KS-距离(柯尔莫可洛夫-斯米洛夫(Kolmogorov-Smirnov))运行。如果分类器性能度量(metric)定义为灵敏性(真阳性的分数, f_{TP})和特异性(1减假阳性的分数, $1-f_{FP}$)的总和, 则完美分类器的评分为2, 随机分类器平均评分为1。利用KS-距离的定义, 最大化cdf函数的差异的值 x^* 可以通过对 x 求解 $\frac{\partial KS}{\partial x} = \frac{\partial(cdf_c(x) - cdf_d(x))}{\partial x} = 0$ 来发现, 这获得 $p(x^*|c) = p(x^*|d)$, 即当类别依赖性pdf正交(cross)时产生KS距离。将 x^* 的这个值代入KS-距离的表达式, 获得以下关于KS的定义

$$KS = cdf_c(x^*) - cdf_d(x^*) = \int_{-\infty}^{x^*} p(x|c) dx - \int_{-\infty}^{x^*} p(x|d) dx = 1 - \int_{x^*}^{\infty} p(x|c) dx - \int_{-\infty}^{x^*} p(x|d) dx$$

$= 1 - f_{FP} - f_{FN}$, 利用具有 x^* 的截断值(cut-off)的检验, KS距离为1减去总误差分数, 基本上为单个分析物贝叶斯(Bayesian)分类器。因为我们定义灵敏性+特异性 $= 2 - f_{FP} - f_{FN}$ 的分数, 组合KS-距离的上述定义, 我们可以得到灵敏性+特异性 $= 1 + KS$ 。我们选择具有固有地适合构建朴素贝叶斯分类器的统计量的生物标记。

[0311] 具有良好KS距离(例如 >0.3)的后续标记的加入通常会改善分类性能, 如果后续加入的标记独立于第一标记。使用灵敏性加特异性作为分类器评分, 用贪婪算法的变体会直接产生许多高分分类器。(贪婪算法是那些应用了解决问题的元启发式(metaheuristic)算法, 这些算法使得本地优化选择在每个阶段都可以找到全方位的最优化。)

[0312] 此处使用的算法在实施例4中详细描述。简言之, 从潜在生物标记表中产生所有单分析物分类器并加入到列表中。接下来, 向每个存储的单分析物分类器进行所有可能的第二分析物的加入, 在一个新列表上储存预定数目的最佳评分配对, 例如一千个。用这个最佳二-标记分类器的新列表开发所有可能的三标记分类器, 再次储存其中最佳的一千个。继续这个过程直至评分进入平稳状态或开始随着额外标记的加入而变差。可以评价那些在收敛后仍保持的高评分分类器对于目标用途的期望性能。例如, 在一种诊断应用中, 具有高灵敏性和中等特异性的分类器可以比中等灵敏性和高特异性是更期望的。在另一诊断应用中, 具有高特异性和中等灵敏性的分类器可以是更期望的。期望的性能水平通常基于在假阳性和假阴性数之间必须进行的交换(trade-off)而选择, 所述假阳性和假阴性可以针对特定诊断应用而各自容忍。这种交换通常取决于假阳性或假阴性误差的医学后果。

[0313] 现有技术已知各种其他技术, 并且可以用来使用朴素贝叶斯分类器从生物标记列表产生许多潜在分类器。在一实施方案中, 所谓的遗传算法可以使用上述定义的适合度评分用于组合不同的标记。遗传算法特别适于开发潜在分类器的大且多样的群体。在另一实施方案中, 所谓的蚁群优化(ant colony optimization)可以用于产生分类器的集合。现有

技术已知的其他策略也可以采用,包括例如其他进化策略以及模拟退火和其他随机检索方法。还可以采用元启发式方法如和声搜索(harmony search)。

[0314] 示例性实施方案使用表18、20或21所列的任意数目的肺癌生物标记以各种组合来产生用于检测肺癌的诊断测试(如何鉴定这些生物标记的详细描述见实施例2和6)。在一实施方案中,诊断肺癌的方法使用朴素贝叶斯分类方法联合表18、20或21所列的任意数目的肺癌生物标记。在示例性实施例(实施例3)中,用于从无症状吸烟者群体检测肺癌的最简单测试可以用诸如SCFsR的单个生物标记来构建,SCFsR在肺癌中下调,KS-距离为0.37 ($1+KS=1.37$)。使用来自表15的针对SCFsR的参数 $\mu_{c,i}$ 、 $\sigma_{c,i}$ 、 $\mu_{d,i}$ 和 $\sigma_{d,i}$ 以及上述log-似然性的等式,可以产生具有63%灵敏性和73%特异性的诊断测试(灵敏性+特异性=1.36),见表14。这个测试的ROC曲线示于图2并且AUC为0.75。

[0315] 加入例如KS-距离为0.5的生物标记HSP90a,显著改进分类器性能至灵敏性为76%及特异性为0.75% (灵敏性+特异性=1.51) 及AUC=0.84。

[0316] 注意,由两个生物标记构建的分类器的评分不是KS-距离的简单加和;当组合生物标记时KS-距离不是加合性时,要使用许多较弱标记来实现与强标记相同的性能水平。加入第三标记ERBB1例如增加分类器性能至78%灵敏性和83%特异性及AUC=0.87。加入额外的生物标记例如PTN、BTK、CD30、激肽释放酶7、LRIG3、LDH-H1和PARC产生了一系列肺癌测试,总结在表14中,并在图3中显示为一系列ROC曲线。作为分类器构建中使用的分析物数目的函数的分类器评分示于图4。这个示例性10标记分类器的灵敏性和特异性>87%,AUC为0.91。

[0317] 表18、20或21所列的标记可以许多方式组合以产生用于诊断肺癌的分类器。在一些实施方案中,根据所选的特定诊断性能标准,生物标记的组由不同数目的分析物组成。例如,生物标记的某些组合会产生比其他组合更灵敏(或更特异性)的测试。

[0318] 一旦组限定为包括来自表18、20或21的生物标记的特定集合,并且从训练数据集构建分类器,则诊断测试的定义完整。在一实施方案中,用于分类未知样品的程序示于图1A。在另一实施方案中,用于分类未知样品的程序示于图1B。将生物学样品适当稀释,然后进行一个或多个测定以产生用于分类的相关定量生物标记水平。测量的生物标记水平用作分类方法的输入,该方法输出所述样品的分类和任选的评分,这反映了类别分配的置信度。

[0319] 表21鉴定了可用于诊断组织和血液样品中的肺癌的86种生物标记。表20鉴定了在组织样品中鉴定的25种生物标记,但是所述生物标记也可用于血清和血浆样品。当与在生物标记发现的尝试中通常发现的相比时,这比预期令人惊讶地高,可能归因于所述研究的规模,其涵盖在几百个个体样品中测量的超过800种蛋白,在一些情况下浓度为低毫微微摩尔范围。据推测,发现的大量生物标记反映了在肿瘤生物学和身体对肿瘤存在的应答中涉及的不同生物化学途径;每种途径和过程均涉及许多蛋白。结果显示没有一小组蛋白的单个蛋白对这样复杂的过程独特提供信息;相反,多种蛋白包括在相关过程中,如凋亡或细胞外基质修复。

[0320] 鉴于在所述研究中鉴定的许多生物标记,可以预期能够衍生大量高性能分类器,它们可以用于各种诊断方法。为测试这个观点,用表1的生物标记评价了几万个分类器。如实施例4所述,表1所示的生物标记的许多子集可以组合以产生有用的分类器。作为示例,提供了含有1、2和3个生物标记的分类器用于如下两种用途的每一种的描述:吸烟者的高风险

肺癌筛查以及具有可由CT检测到的肺结节的个体的诊断。如实施例4所述,用表1的生物标记构建的所有分类器比用“非标记”构建的分类器明显更好。

[0321] 还测试了通过随机排除表1的一些标记获得的分类器的性能,随机排除产生了较小的子集,从这些子集来构建分类器。如实施例4部分3所述,从表1中标记的随机子集构建的分类器的表现类似于用表1中标记的完全列表构建的最佳分类器。

[0322] 还测试了通过从10-标记聚集中排除“最佳”单个标记而获得的10-标记分类器的性能。如实施例4部分3所述,未用表1的“最佳”标记构建的分类器也表现良好。表1所列生物标记的许多子集表现接近最佳,甚至在除去表中所列最好的15个标记之后也如此。这暗示任何特定分类器的性能特征可能不是由于生物标记的一些小核心组所致,并且疾病过程可能影响许多生物化学途径,其改变许多蛋白的表达水平。

[0323] 实施例4的结果提示一些可能的结论:首先,大量生物标记的鉴定允许它们聚集成巨大数目的分类器,其提供相似的高性能。第二,分类器可以这样构建:特定生物标记可以以反映冗余的方式取代其他生物标记,所述冗余无疑遍及潜在疾病过程的复杂性。也就是说,由表1中鉴定的任何单个生物标记贡献的有关疾病的信息与由其他生物标记贡献的信息重叠,由此表1中特定生物标记或生物标记的小组无需包括在任何分类器中。

[0324] 示例性实施方案使用从表38和39的数据构建的朴素贝叶斯分类器以分类未知样品。图1A和B中示出了程序。在一实施方案中,将生物学样品任选地稀释并且进行多重适配体测定。将来自所述测定的数据如实施例3所述归一化及校准,将所得生物标记水平用作贝叶斯分类方案的输入。对于每个测量的生物标记单独计算log-似然比,然后求和以产生最终分类评分,其也称为诊断评分。可以报道所得分配及总体分类评分。任选地,还可以报道对每个生物标记水平计算的个体log-似然性风险因子。分类评分计算的细节示于实施例3。

[0325] 为了证实本文所述的基于适配体的蛋白质组技术用于发现来自组织的疾病相关的生物标记的用途,如实施例6所述分析获得自8位非小细胞肺癌(NSCLC)患者的来自手术切除的均质组织样品。所有NSCLC患者均为吸烟者,年龄范围为47-75岁,并且覆盖NSCLC 1A期至3B期(表17)。3个样品获得自每个切除:肿瘤组织样品、邻近肺肿瘤组织。在用于蛋白质组谱的每个匀浆中调整并归一化总蛋白浓度,然后分析DNA微阵列平台以测量超过800种人蛋白的浓度(参见Gold et al., Nature Precedings, <http://precedings.nature.com/documents/4538/version/1> (2010))。

[0326] 表示为相对荧光单位(RFU)的蛋白浓度测量允许样品中蛋白特征的大规模比较(参见图21)。参考图21,首先对每个患者样品比较对照邻近与远端组织之间的蛋白表达水平(图21A)。在这个比较中,仅一种分析物(纤维蛋白原)在样品之间表现出超过2倍的差异。总的来说,大多数分析物产生的信号在邻近和远端组织中是相似的。

[0327] 相比之下,肿瘤组织与非肿瘤组织(邻近或远端)的比较鉴定了11种(1.3%)蛋白具有大于4倍的差异,并且53种(6.5%)具有大于2倍的差异(参见图21B和21C)。剩余的767种(93.5%)蛋白在肿瘤与非肿瘤组织之间表现出相对较小的差异。与邻近或远端组织相比,在肿瘤组织中一些蛋白基本上被抑制,而其他蛋白升高。邻近与肿瘤组织之间或者远端与肿瘤组织之间蛋白的差异表达总的来说是相似的。远端组织中的变化一般较大(图21),这证实大多数蛋白变化是局部肿瘤环境特异性的。

[0328] 为了鉴定NSCLC组织生物标记,利用如Ostroff et al. Nature Precedings,

<http://precedings.nature.com/documents/4537/version/1> (2010) 所述的Mann-Whitney 测试比较肿瘤、邻近和远端组织样品之间的蛋白表达水平。鉴定了具有最大倍数变化并在肿瘤与非肿瘤组织之间具有统计学显著差异的36种蛋白,显著性的错误发现率界限为 $q < 0.05$ (图23和24,以及表18)。在肿瘤组织中,这些蛋白中的20种上调,而16种下调。虽然用于这个研究的样品数目相对较少,采用有力的基于个体的研究设计,其中每个肿瘤样品具有其自己的健康组织对照。这消除与基于群体的研究设计相关的群体方差。越来越认识到适当选择的参考样品的可用性是生物标记发现研究中极为重要的组分 (Bossuyt (2011) *J. Am. Med. Assoc.* 305:2229-30; Ioannidis and Panagiotou (2011) *J. Am. Med. Assoc.* 305:2200-10; Diamandis (2010) *J. Natl. Cancer Inst.* 102:1462-7)。

[0329] 据信约三分之一 (13/36) 的本文鉴定的NSCLC组织生物标记是新的。剩余的三分之二 (23/36) 以前已报道为NSCLC肿瘤组织中差异表达的蛋白或基因 (表18)。

[0330] 如表19所示,根据实施例6的方法鉴定的生物标记可以大体上分类为与肿瘤生物学的重要标志 (Hanahan&Weinberg (2011) *Cell* 144:646-74) 相关的4个生物学过程:1) 血管发生,2) 生长和代谢,3) 炎症和凋亡,以及4) 侵入和转移。公认地,这些是方便但不准确的分类,其近似于高度复杂和动态的系统,其中这些分子常起多重和细微差异的作用。因此,给定系统的具体状态最终影响任何特定分子的表达和功能。在这些系统中,生物学理解远远没有完成。通过SOMAscan平台,使得各种组织和疾病过程中的大量蛋白的定量表达变得可能。这些数据提供新坐标以帮助作图这些系统的动力学,这反过来会提供肺癌以及其他疾病的生物学的更完整理解。来自目前研究的结果提供了关于NSCLC肿瘤生物学的新视角,具有熟悉的和新的元件。

[0331] 血管发生

[0332] 血管发生驱动新血管的生长以支持肿瘤生长和代谢。血管发生的调节是由正信号和负信号控制的复杂的生物学现象 (Hanahan&Weinberg, (2011) *Cell* 144:646-74)。本研究鉴定的NSCLC组织生物标记中有公知的正血管发生调节剂和负血管发生调节剂 (图23和24以及表19),以前全部在NSCLC肿瘤组织中已观察到 (Fontanini et al. (1999) *British Journal of Cancer* 79 (2):363-369; Imoto et al. (1998) *J. Thorac. Carciovasc. Surg* 115:1007-1011; Ohta et al. (2006) *Ann. Thorac. Surg.* 82:1180-1184; Iizasa et al. (2004) *Clinical Cancer Research* 10:5361-5366)。这些包括原型血管发生诱导剂VEGF以及抑制剂内皮抑制素和血小板反应蛋白-1 (TSP-1)。VEGF是强大的生长因子,其促进新血管生长,并且在NSCLC肿瘤组织中强烈上调,这与以前的观察一致 (Imoto et al. (1998) *J. Thorac. Carciovasc. Surg* 115:1007-1011),并且包括我们对来自NSCLC患者的血清样品的研究 (Ostroff et al. *Nature Precedings*, <http://precedings.nature.com/documents/4537/version/1> (2010))。内皮抑制素是胶原XVIII的蛋白水解产物,并且是内皮细胞增殖和血管发生的抑制剂 (Iizasa et al. (Aug. 2004) *Clinical Cancer Research* 10:5361-5366)。TSP-1和相关的血小板反应蛋白-2 (TSP-2) 在NSCLC肿瘤组织中基本上调。TSP-1和TSP-2为胞外基质蛋白,其具有通过与细胞表面受体、生长因子、细胞因子、基质金属蛋白酶和其他分子的各种相互作用调节的复杂的环境依赖性效应。在模型系统中典型地,TSP-1和TSP-2通过CD47受体抑制内皮细胞增殖和CD36受体诱导内皮细胞凋亡来抑制血管发生。还有TSP-1和TSP-2的促进血管发生影响的证据 (Bornstein (2009) *J. Cell*

Commun.Signal.3 (3-4):189-200)。最后,报道的NSCLC组织中的TSP-1和TSP-2相对和绝对表达水平不同(Chijiwa et al.(2009)Oncology Reports 22:279-283;Chen et al.(2009)J Int Med Res 37:551-556;Oshika 1998,Fontanini et al.(1999)British Journal of Cancer 79(2):363-369),可能是由于它们复杂的功能。在这个研究中,据发现CD36在NSCLC肿瘤组织中下调,这可以表明肿瘤细胞的适应降低对TSP-1和TSP-2介导的凋亡的敏感性。

[0333] 生长和代谢

[0334] 鉴定的NSCLC生物标记中的10个与生长和代谢功能有关。这些生物标记中的一半参与细胞生长和能量代谢的复杂激素调节。调节胰岛素样生长因子(IGF)的活性的3种胰岛素样生长因子结合蛋白(IGFBP)在NSCLC肿瘤中上调(IGFBP-2、-5和-7)。几个报道已定性评价NSCLC中的IGFBP-2、-5和-7(表18),并且表明在NSCLC组织中比在正常组织中表达更高。胰岛素和IGF是强烈影响细胞生长和代谢的激素,并且癌细胞的生长和增殖常依赖于这些分子(Robert et al.(Aug.1999)Clinical Cancer Research5:2094-2102;Liu et al.(June 2007)Lung Cancer 56(3):307-317;Singhal et al.(2008)Lung Cancer 60:313-324)。这些激素反过来被我们发现在NSCLC肿瘤组织中上调的insulysin降解。激素脂连蛋白控制脂质代谢和胰岛素敏感性,并且我们发现脂连蛋白在NSCLC肿瘤中下调。剩余的5个生物标记碳酸酐酶III、NAGK、TrATPase、类胰蛋白酶 β -2和MAPK13是在细胞代谢中起作用的酶(表17)。

[0335] 炎症和凋亡

[0336] 炎症和凋亡是癌症细胞学的标志,并且许多潜在的生物标记与这些过程相关,这些过程以前已与NSCLC相关(表19)。据发现与转移相关的胱天蛋白酶-3在NSCLC肿瘤组织中上调(Chen et al.(2010)Lung Cancer(doi:10.16/j.lungcan.2010.10.015)。另一值得注意的实例是RAGE,据报道其在NSCLC组织中显著下调(Jing et al.(2010)Neoplasma.57:55-61,Bartling et al.(2005)Carcinogenesis 26:293-301)。这个发现与本文公开的测量一致,其中对于在肿瘤中比在非恶性组织中低的蛋白,sRAGE具有观察到的最大变化。虽然未受理论限制,但是一种假设是RAGE在上皮组织(epithelial organization)中起作用,并且肺肿瘤中降低水平的RAGE可以归因于丧失上皮组织结构,可能导致恶性转化(Bartling et al.(2005)Carcinogenesis26(2):293-301)。几种趋化因子如BCA-1、CXCL16、IL-8和NAP-2改变了(表18),这与这样的假设一致,具有来自免疫系统的先天和适应性装备(arm)的细胞的肿瘤的侵入提供影响增殖和血管发生信号的生物活性分子(Hanahan&Weinberg(2011)Cell 144:646-74)。

[0337] 侵入和转移

[0338] 潜在的生物标记的最大组包含在细胞-细胞和细胞-基质相互作用中发挥作用并参与侵入和转移的蛋白。许多蛋白以前已报道与NSCLC相关。最值得注意的是基质金属蛋白酶中的两种,MMP-7和MMP-12,其有助于胞外基质组分的蛋白降解以及底物如生长因子的加工(参见例如Su et al.(2004)Chinese Journal of Clinical Oncology 1(2):126-130;Wegmann et al.(1993)Eur.J.Cancer 29A(11):1578-1584)。这类过程公知在产生肿瘤微环境中起作用。据发现MMP-7和MMP-12均在NSCLC组织中上调(表18),这与使用基于抗体的测量的相似研究一致(Shah et al.(2010)The Journal of Thoracic and

Cardiovascular Surgery 139 (4):984-990)。MMP-7和MMP-12的过表达与NSCLC中的不良预后有关(Shah et al.(2010) The Journal of Thoracic and Cardiovascular Surgery 139 (4):984-990)。MMP-12水平与局部复发和转移疾病相关(Hofmann et al.(2005) Clin.Cancer Res.11:1086-92,Hoffman et al.(2006) Oncol.Rep.16:587-95))。研究的8个对象中的2个具有正常水平的MMP-12,而其他6个在肿瘤组织中与非肿瘤组织相比具有15-50x升高的MMP-12。

[0339] NSCLC生物标记作为组织化学探针的性能

[0340] 肿瘤与非肿瘤组织之间蛋白表达的差异的理解可以用来鉴定新的组织化学探针。这类探针可以允许更精确地分子表征肿瘤以及它们对周围基质的影响。图25证实鉴定的SOMAmer中的2种染色新鲜冷冻组织的能力,所述新鲜冷冻组织获得自用于发现这些生物标记的相同肿瘤切除。据发现血小板反应蛋白-2(TSP2)在肿瘤组织匀浆中增加,而巨噬细胞甘露糖受体(MRC1)减少。用这些SOMAmer组织染色与分析结果(profiling result)一致。其他实例以及染色模式的抗体确认如图27所示。

[0341] NSCLC组织和血清生物标记的比较

[0342] NSCLC患者的血清相对于无癌症对照的蛋白差异表达与NSCLC组织样品的蛋白差异表达相比产生有益启示(图26)。最引人注目的观察是蛋白表达的相对变化在组织中比在血清中大。这个结果可以预期,因为肿瘤组织是蛋白表达变化的来源,即使然后蛋白表达完全释放入循环,也稀释入血液的总体积许多倍。这个趋势在沿着图26中x-轴的数据点的延长分布中很明显,其中在相同规模上绘制轴以说明这个点。在肿瘤组织中改变的图23和24所示的分析物中的12种在来自NSCLC患者vs.对照的血清中也差异表达(图26中填充红色的圆圈)。大多数定向变化在组织与血清之间是相同的,但是一些不是。组织匀浆中蛋白的局部浓度显然不需要与蛋白的循环水平相关,而逆相关可以提供关于在疾病对正常组织中某些生物标记的重新分布的线索。

[0343] 最近几年中,具有可证实的诊断或临床用途的新生物标记的发现是相当大的挑战(Diamandis(2010) J.Natl.Cancer Inst.102:1462-7)。其原因包括分析前和分析人工制品的无处不在,合适的健康状态对照和简单的研究设计的不可用性,以及难以检测非常低浓度的蛋白水平的小变化。这个挑战对于癌症生物标记尤其突出,其中目的常是在早期鉴定相对大的人体中的微小恶性肿瘤。关于后点,提高发现真正的癌症生物标记的机会的一种方式是从疾病的来源如肿瘤组织以及循环获得蛋白表达数据。组合的结果可以部分证实潜在的生物标记的正确性。本申请证实关于公开的高度复用和灵敏的蛋白质组测定这是可能的。据证实组织如血浆或血清也适合SOMAscan,并且所得的NSCLC肿瘤组织与周围健康肺组织的蛋白表达的比较分析为从血清样品鉴定的潜在的NSCLC生物标记的现有数据集提供补充(参见U.S.Pub.No.2010/0070191)。在本案中,本文报道的36个组织生物标记中的三分之一或12个((BCA-1(BCL)、钙粘着蛋白-1(钙粘着蛋白-E)、过氧化氢酶、内皮抑制素、IGFBP-2、MRC1(巨噬细胞甘露糖受体)、MAPK-13(MK13)、MMP-7、MMP-12、NAGK、VEGF和YES)以前已在血清中鉴定。合在一起,这些数据有助于进一步理解伴随着NSCLC的变化的复杂性,并且为这种致命疾病的早期检测提供额外的潜在生物标记。

[0344] 试剂盒

[0345] 表20的生物标记(以及额外的生物医学信息)的任何组合可以通过使用合适的试

剂盒检测,如用于进行本文公开的方法的试剂盒。此外,任何试剂盒可以含有本文所述的一种或多种可检测标记,如荧光部分等。

[0346] 如本文进一步描述,在一实施方案中,试剂盒包含:(a)一种或多种捕获试剂(如至少一种适配体或抗体),以检测生物学样品中的一种或多种生物标记,其中所述生物标记包括表18、20或21所列的任何生物标记,及任选存在的(b)一种或多种软件或计算机程序产品,以将从其中获得所述生物学样品的个体分类为患有或不患有肺癌,或者确定所述个体患有肺癌的似然性。或者,除了一种或多种计算机程序产品之外,可以提供人工进行上述步骤的一种或多种操作指南。

[0347] 固体支持物与相应捕获试剂及信号产生材料的组合在本文中称作“检测装置”或“试剂盒”。试剂盒还可以包含使用该装置和试剂、处理样品以及分析数据的使用说明书。此外,试剂盒可以与计算机系统或软件一起使用以分析和报道生物学样品的分析结果。

[0348] 试剂盒还可以含有一种或多种试剂(如增溶缓冲液、去污剂、洗涤剂或缓冲液),以对生物学样品进行处理。本文所述的任何试剂盒还可以包含例如缓冲液、封闭剂、质谱基质材料、抗体捕获剂、阳性对照样品、阴性对照样品、软件以及信息如方案、指导和参考数据。

[0349] 在一方面,本发明提供了分析肺癌状态的试剂盒。试剂盒包含用于选自表18、20或21的一个或多个生物标记的PCR引物。试剂盒还可以包含使用及生物标记与肺癌相关性的指导。试剂盒还可以包含DNA阵列,其含有选自表20的一个或多个生物标记的补体、用于扩增或分离样品DNA的试剂和/或酶。试剂盒可以包含用于实时PCR的试剂,例如TaqMan探针和/或引物,以及酶。

[0350] 例如,试剂盒可以包含:(a)试剂,其至少包含用于定量测试样品中的一种或多种生物标记的捕获试剂,其中所述生物标记包括表18、20或21所列的生物标记或本文所述的任何其他生物标记或生物标记组;以及任选存在的(b)一种或多种算法或者计算机程序,以进行以下步骤:比较所述测试样品中定量的每个生物标记的量与一个或多个预定截断值,基于所述比较分配定量的每个生物标记评分,组合定量的每个生物标记的分配评分以获得总评分,比较该总评分与预定评分,以及使用所述比较确定个体是否患有肺癌。或者,除了一种或多种算法或计算机程序之外,可以提供人工进行上述步骤的一种或多种操作指南。

[0351] 计算机方法和软件

[0352] 一旦选择生物标记或生物标记组,则诊断个体的方法可包括如下步骤:1)收集或以其他方式获得生物学样品;2)进行分析方法以检测和测量所述生物学样品中的生物标记或生物标记组;3)进行用于收集生物标记值的方法所需的任何数据归一化或标准化;4)计算标记评分;5)组合所述标记评分以获得总诊断评分;以及6)报告所述个体的诊断评分。在这种方法中,诊断评分可以是所有标记计算的总和确定的单一数值,将该数值与指示疾病存在与否的预设阈值比较。或者,诊断评分可以是一系列带(bar),其各自代表生物标记值,并且可以将应答模式与预设模式比较以确定疾病的存在与否。

[0353] 本文所述方法的至少一些实施方案可以使用计算机实施。计算机系统100的实例在图6中示出。参考图6,系统100示出包含通过总线108电耦合的硬件元件,包括处理器101、输入设备102、输出设备103、存储设备104、计算机可读的存储介质读取器105a、通讯系统106、处理加速(如DSP或特定用途处理器)107和存储器109。计算机可读存储介质读取器105a与计算机可读的存储介质105b进一步耦合,该组合全面地代表远程、局域、固定和/

或可移动的存储装置加上存储介质、存储器等,以暂时和/或更长久地含有计算机可读取的信息,这可以包括存储设备104、存储器109和/或任何其他这样的可存储系统100资源。系统100还包含软件元件(目前位于工作内存191中),包括操作系统192及其他编码193,如程序、数据等。

[0354] 关于图6,系统100具有广泛的灵活性和可配置性。因此,例如单一计算机结构(single architecture)可以用于完成一个或多个服务器,其可以根据目前期望的方案、方案变化、扩展等进一步配置。然而,本领域技术人员应当了解可以根据更具体的应用要求更好地利用实施方案。例如,一个或多个系统元件可以在系统100部件内作为子元件执行(如在通讯系统106中)。也可以使用定制的硬件和/或特定元件可以在硬件、软件或者硬件和软件中执行。此外,虽然可以使用与其他计算机设备如网络输入/输出设备(未示出)连接,但是应当理解也可以利用有线、无线、调制解调器和/或与其他计算机设备的其他连接或多个连接。

[0355] 在一方面,所述系统可以包含含有肺癌特征性生物标记的特征的数据库。生物标记数据(或生物标记信息)可以用作计算机的输入以用作计算机执行方法的一部分。所述生物标记数据可以包括本文所述的数据。

[0356] 在一方面,所述系统还包含一个或多个装置以将输入数据提供给一个或多个处理器。

[0357] 所述系统还包含用于存储分等级的数据元件的数据集合的存储器。

[0358] 在另一方面,用于提供输入数据的装置包括检测仪以检测数据元件的特征,如质谱仪或者基因芯片读取器。

[0359] 所述系统还可以包含数据库管理系统。用户请求或询问可以通过数据库管理系统理解的适当语言格式化,该数据库管理系统处理所述询问以从训练集合的数据库中提取相关信息。

[0360] 所述系统可以与网络连接,所述网络连接网络服务器及一个或多个客户端。网络可以是本领域已知的局域网(LAN)或广域网(WAN)。优选地,服务器包括运行计算机程序产品(如软件)所需的硬件,以进入数据库数据来处理用户请求。

[0361] 所述系统可以包含操作系统(如UNIX或Linux),以执行来自数据库管理系统的命令。在一方面,操作系统可以在全球通讯网络上运行,如在国际互联网上运行,及利用全球通讯网络服务器来连接这样的网络。

[0362] 所述系统可以包括这样的—个或多个装置,其包含图形显示界面,该界面包括界面元件如按钮、下拉菜单、滚动条、输入文本的信息栏等,这些是本领域已知的图形用户界面的常见元件。用户界面上登录的请求可以传给系统中的应用程序以格式化从而在一个或多个系统数据库中搜寻相关信息。用户登录的请求或询问可以任何合适的数据库语言建立。

[0363] 图形用户界面可以通过作为操作系统一部分的图形用户界面编码产生,并且可以用于输入数据和/或显示输入的数据。处理的数据的结果可以在界面上显示,在与该系统通讯的打印机上打印,存储在存储装置中,和/或上传至网络或者可以计算机可读取介质的形式提供。

[0364] 所述系统可以与输入装置通讯,以将关于数据元件的数据提供给系统(如表达

值)。在一方面,输入装置可以包括基因表达谱系统,包括如质谱仪、基因芯片或阵列读取器等。

[0365] 根据各个实施方案的分析肺癌生物标记信息的方法和设备可以通过任何合适的方式执行,例如使用在计算机系统上运行的计算机程序。可以使用常规的计算机系统,其包含处理器和随机存取存储器,如可远程登录的应用服务器、网络服务器、个人电脑或工作站。其他计算机系统部件可以包括存储装置或信息存储系统,如大量存储系统和用户界面,例如常规的监视器、键盘和跟踪装置。计算机系统可以是单机系统,或者是包含服务器以及一个或多个数据库的计算机网络的一部分。

[0366] 肺癌生物标记分析系统可以提供完整数据分析的运算(function)和运行,如数据收集、处理、分析、报告和/或诊断。例如,在一实施方案中,计算机系统可以执行计算机程序,该程序可以接收、储存、搜寻、分析和报告关于肺癌生物标记的信息。计算机程序可以包含进行各种运算或运行的多个模块,如处理原始数据和产生补充数据的处理模块,以及分析原始数据和补充数据以产生肺癌状态和/或诊断的分析模块。诊断肺癌状态可以包括产生或收集任何其他信息,包括额外的生物学信息、关于个体相对于疾病的状况,鉴定是否需要进一步检测,或者另外评价个体的健康状态。

[0367] 关于图7,可以看出根据公开的实施方案的原理利用计算机的方法的一个实例。在图7中,示出了流程图3000。在模块(block) 3004中,可以检索个体的生物标记信息。所述生物标记信息可以从计算机数据库中检索,例如在测试个体的生物学样品之后。生物标记信息可以包含生物标记值,所述生物标记值每个对应于由选自表18(其中 $N=2-36$)、表20(其中 $N=2-25$)或表21(其中 $N=2-86$)提供的生物标记组成的组的至少 N 个生物标记之一。在模块3008中,计算机可以用于将每个生物标记值分类。在模块3012中,基于多个分类可以确定个体患有肺癌的似然性。可以将该指示输出至显示器或其他显示装置,以供人观察。因此,例如该指示可以在计算机的显示器屏幕或其他输出装置上显示。

[0368] 关于图8,通过流程图3200举例说明了根据另一实施方案的利用计算机的可选方法。在模块3204中,可以利用计算机检索个体的生物标记信息。生物标记信息包含生物标记值,所述生物标记值对应于选自表18、20或21提供的生物标记的组的生物标记。在模块3208中,可以用计算机将生物标记值进行分类。并且,在模块3212中,基于所述分类可以对个体患有肺癌的似然性作出指示。可以将该指示输出至显示器或其他显示装置,以供人观察。因此,例如该指示可以在计算机的显示器屏幕或其他输出装置上显示。

[0369] 本文所述的一些实施方案可以执行以包含计算程序产品。计算机程序产品可以包括具有包含于介质中的计算机可读取程序编码的计算机可读取介质,以使得应用程序可以在具有数据库的计算机上执行。

[0370] 如本文所用,“计算机程序产品”指自然或程序设计语言语句形式的组织化的指令集合,其包含于任何性质的物理介质上(如书写、电子、磁性、光学或其他性质),并且可以与计算机或其他自动化数据处理系统一起使用。这样的程序设计语言语句在由计算机或数据处理系统执行时,使得所述计算机或数据处理系统根据语句的特定内容起作用。计算机程序产品包括但不限于:包含于计算机可读取介质中的源代码和目标码和/或测试或数据库中的程序。此外,允许计算系统或数据处理设备以预选方式起作用的计算机程序产品可以以多种形式提供,包括但不限于源代码(original source code)、汇编码(assembly

code)、目标码、机器语言、前述代码的加密或压缩形式以及任何和所有等价物。

[0371] 在一方面,本发明提供了一种指示肺癌的似然性的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质,所述程序代码可由计算装置或系统的处理器执行,所述程序代码包括:对归因于来自个体的生物学样品的数据进行检索的代码,其中所述数据包括生物标记值,所述生物标记值每个对应于所述生物学样品中选自表18(其中 $N=2-36$)、表20(其中 $N=2-25$)或表21(其中 $N=2-86$)提供的生物标记的组的至少 N 个生物标记之一;以及执行分类方法的代码,所述分类方法指示作为所述生物标记值的函数的所述个体的肺疾病状态。

[0372] 在另一方面,本发明提供了一种指示肺癌的似然性的计算机程序产品。所述计算机程序产品包括包含程序代码的计算机可读取介质,所述程序代码可由计算装置或系统的处理器执行,所述程序代码包括:对归因于来自个体的生物学样品的数据进行检索的代码,其中所述数据包括生物标记值,所述生物标记值对应于所述生物学样品中选自表18(其中 $N=2-36$)、表20(其中 $N=2-25$)或表21(其中 $N=2-86$)提供的生物标记的组的生物标记;以及执行分类方法的代码,所述分类方法指示作为所述生物标记值的函数的所述个体的肺疾病状态。

[0373] 虽然已经描述了本发明的方法或设备的各种实施方案,但是应当理解所述实施方案可以通过与计算机耦合的代码执行,如在计算机上或可由计算机登录的代码。例如,软件和数据库可以用于执行上述许多方法。因此,除了由硬件完成的实施方案之外,还应当注意到这些实施方案可以通过使用这样的产品实现,所述产品包含具有在其中包含计算机可读取程序代码的计算机可用介质,其允许行使本文公开的功能。因此,期望所述实施方案也可以被视为由本专利以其程序代码方式等保护。此外,所述实施方案可以体现为存储在实际任何类型的计算机可读取存储器中的代码,包括但不限于RAM、ROM、磁性介质、光学介质或磁-光学介质。更通常地,所述实施方案可以在软件或硬件或者它们的任何组合中实施,包括但不限于在通用处理器、微代码、PLA或ASIC上运行的软件。

[0374] 还期望所述实施方案可以作为包含于载波中的计算机信号以及通过传送介质传送的信号(如电信号和光信号)实现。因此,上述各种类型的信息均可以在结构中格式化,如数据结构,并且作为电信号通过传送介质传送,或者存储在计算机可读取介质中。

[0375] 还应当注意,本文所列举的许多结构、材料和条文可以列举为用于行使功能或行使功能的步骤的方式。因此,应当理解这样的语言有权覆盖本说明书所公开的所有的这些结构、材料或条文或它们的等价物,包括通过引用并入本文的内容。

实施例

[0376] 下述实施例仅用于示例目的而不意图限制所附权利要求限定的本申请的范围。本文描述的所有实施例用本领域技术人员公知的常规标准技术进行。下述实施例描述的常规分子生物学技术可以如标准实验室手册所述进行,如Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3rd.ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (2001)。

[0377] 实施例1. 用于肺癌生物标记选择的样品的多重适配体分析

[0378] 本实施例描述用于分析样品和对照以鉴定表1第2列所列的生物标记的多重适配

体测定(见图9)。在本例中,多重分析使用820个适配体,每个对于一个特定靶标均是独特的。

[0379] 在这个方法中,每次添加溶液均更换移液器尖头。

[0380] 另外,除非另有说明,大多数溶液转移和洗涤添加使用Beckman Biomek Fx^P的96孔头(head)。除非另有说明,手工移液的方法步骤使用12通道P200Pipettman (Rainin Instruments, LLC, Oakland, CA)。称为SB17的定制缓冲液由内部(in-house)制备,其包含40mM HEPES、100mM NaCl、5mM KCl、5mM MgCl₂、1mM EDTA, pH7.5。除非另有说明,所有步骤均在室温下进行。

[0381] 1. 适配体储液的制备

[0382] 对于不具有光可裂解生物素接头的适配体,用合适的光可裂解生物素酰化引物在1x SB17, 0.05%吐温-20中以8x浓度制备用于10%、1%和0.03%血清的定制适配体储液,其中所得引物浓度是相关适配体浓度的3倍。引物与全部或部分相应适配体杂交。

[0383] 将3种8x适配体溶液的每一种分别以1:4稀释入1xSB17, 0.05%吐温-20 (1500μL的8x储液稀释入4500μL的1xSB17, 0.05%吐温-20) 以实现2x浓度。然后将每种稀释的适配体主混合物(master mix) 分配到4个2mL螺旋盖试管中,每个试管1500μL,并且加热至95℃保持5分钟,随后在37℃下温育15分钟。温育后,将对应于特定适配体主混合物的4个2mL试管组合入试剂槽(trough),将55μL的2x适配体混合物(对于所有三种混合物)手工移液入96孔Hybaid平板并用箔将平板密封。最终结果是3个96孔箔密封的Hybaid平板。如表2所示,各个适配体浓度在0.5-4nM的范围。

[0384] 2. 测定样品制备

[0385] 将储存在-80℃的100%血清的冷冻等份置于25℃水浴中保持10分钟。将融化的样品置于冰上,温和地涡漩(设为4)8秒钟,然后再置于冰上。

[0386] 在4℃下,用50μL 8通道多通道移液器(spanning pipettor)将16μL样品转移至96孔Hybaid平板中以制备20%样品溶液,每个孔含有64μL合适的样品稀释物(对于血清,0.8x SB17、0.05%吐温-20、2μM Z-block_2、0.6mM MgCl₂)。将这个平板保存在冰上直至开始下一个样品稀释步骤。

[0387] 为了开始样品和适配体平衡,将20%样品平板短暂离心并置于Beckman FX上,在此将其用96孔移液器上下移液来混合。然后通过将10μL的20%样品稀释至90μL的1xSB17, 0.05%吐温-20中来制备2%样品。然后,将6μL所得的2%样品稀释至194μL的1xSB17, 0.05%吐温-20中以制备0.06%样品平板。在Beckman Biomek Fx^P上进行稀释。每次转移后,通过上下移液来将溶液混合。然后,通过将55μL样品加入到55μL合适的2x适配体混合物中,将3个样品稀释平板转移至它们各自的适配体溶液。样品和适配体溶液通过上下移液在机器人(robot)上混合。

[0388] 3. 样品平衡结合

[0389] 将样品/适配体平板用箔密封,并且置于37℃培养箱中保持3.5小时,然后进行Catch 1步骤。

[0390] 4. 制备Catch 2珠平板

[0391] 将11mL MyOne (Invitrogen Corp., Carlsbad, CA) 链霉抗生物素蛋白C1珠等份用等体积的20mM NaOH洗涤2次(每次洗涤温育5分钟),用等体积的1x SB17, 0.05%吐温-20洗

涤3次,并重悬浮于11mL的1x SB17,0.05%吐温-20中。用12-通道(span)多通道移液器将50 μ L该溶液手工移液至96孔Hybaid平板的每个孔中。然后将平板用箔覆盖,并保存在4℃用于测定。

[0392] 5.制备Catch 1珠平板

[0393] 将3个0.45 μ m Millipore HV平板(Durapore membrane,Cat#MAHVN4550)用100 μ L的1x SB17,0.05%吐温-20平衡至少10分钟。然后通过平板过滤平衡缓冲液,向每孔中加入133.3 μ L的7.5%链霉抗生物素蛋白-琼脂糖珠浆液(于1x SB17,0.05%吐温-20中)。为了在将链霉抗生物素蛋白-琼脂糖珠转移入滤板时保持它们悬浮,将珠溶液用200 μ L,12-通道移液器手工混合15次。在将珠分配于3个滤板中之后,施加真空以除去珠上清。最后,将珠在滤板中用200 μ L的1x SB17,0.05%吐温-20洗涤,然后重悬于200 μ L的1x SB17,0.05%吐温-20中。将滤板的底部吸干(blot),并将平板储存以用于测定中。

[0394] 6.装载Cytomat

[0395] 将cytomat用所有尖头、平板、槽中的所有试剂(除了NHS-生物素试剂,其在即将加入到平板之前制备)、3个制备的catch 1滤板和1个制备的MyOne平板装载。

[0396] 7.Catch 1

[0397] 3.5小时平衡时间之后,将样品/适配体平板从培养箱中移出,离心约1分钟,除去箔,置于Beckman Biomek Fx^P平台上。开始Beckman Biomek Fx^P程序。除非另有说明,Catch 1中的所有后续步骤均由Beckman Biomek Fx^P机器人完成。在该程序内,向Catch 1滤板施加真空以除去珠上清。将100微升10%、1%和0.03%平衡结合反应的每一种加入到它们各自的Catch 1滤板,每个板用即用(on-deck)轨道摇床(orbital shaker)以800rpm混合10分钟。

[0398] 通过真空过滤除去未结合的溶液。通过分配溶液并立即施加真空以使溶液通过平板过滤,将Catch 1珠用190 μ L的1x SB17,0.05%吐温-20中的100 μ M生物素洗涤,然后用190 μ L的1x SB17,0.05%吐温-20洗涤。

[0399] 然后,将190 μ L的1x SB17,0.05%吐温-20加入Catch 1平板中。用即用吸干装置(blot station)将平板吸干以除去液滴,然后用轨道摇床在800rpm、25℃下温育10分钟。

[0400] 机器人通过真空过滤除去这个洗涤液,并用即用吸干装置吸干滤板底部以除去液滴。

[0401] 8.标记(Tagging)

[0402] 将NHS-PE04-生物素等份在37℃下融化6分钟,然后用标记缓冲液(pH=7.25的SB17,0.05%吐温-20)以1:100稀释。将NHS-PE04-生物素试剂在无水DMSO中溶解为100mM的浓度,并冷冻保存于-20℃。在机器人辅助下,将稀释的NHS-PE04-生物素试剂手工加入即用槽中,手工重新启动机器人程序以将100 μ L的NHS-PE04-生物素分配到每个Catch 1滤板的每个孔中。将这个溶液用Catch 1珠在轨道摇床上以800rpm振荡温育5分钟。

[0403] 9.动力学攻击(Kinetic Challenge)和光裂解

[0404] 通过向Catch 1平板加入150 μ L的1x SB17,0.05%吐温-20中的20mM甘氨酸来中止标记反应,同时仍含有NHS标签。然后将平板在轨道摇床上以800rpm温育1分钟。通过真空过滤除去NHS-标签/甘氨酸溶液。然后,将190 μ L的20mM甘氨酸(1x SB17,0.05%吐温-20)加入到每个平板,并在轨道摇床上以800rpm温育1分钟,之后通过真空过滤除去。

[0405] 将190μL的1x SB17,0.05%吐温-20加入到每个平板,并通过真空过滤除去。

[0406] 然后,通过加入190μL的1x SB17,0.05%吐温-20,将平板置于轨道摇床上以800rpm保持1分钟随后真空过滤来将Catch 1平板的孔洗涤3次。最后一次洗涤后,将平板置于1mL深孔平板的顶部并从平台上移开。将Catch 1平板在1000rpm下离心1分钟以在洗脱之前从琼脂糖珠上尽可能除去无关的体积。

[0407] 将平板放回Beckman Biomek Fx^P上,并向滤板的每孔中加入85μL的1x SB17,0.05%吐温-20中的10mM D₂SO₄。

[0408] 将滤板从平台上移开,置于Variomag Thermoshaker (Thermo Fisher Scientific, Inc., Waltham, MA) 上,在BlackRay (Ted Pella, Inc., Redding, CA) 光源下,以800rpm振荡同时照射10分钟。

[0409] 通过首先将10% Catch 1滤板置于1mL深孔平板的顶部并以1000rpm离心1分钟来将光裂解的溶液依次从每个Catch 1平板洗脱入同一个深孔平板。然后,将1%和0.03% catch 1平板依次离心入相同的深孔平板。

[0410] 10. Catch 2珠捕获

[0411] 将含有合并的catch 1洗脱物的1mL深孔块(block)置于用于catch 2的Beckman Biomek Fx^P的平台上。

[0412] 机器人将所有光裂解洗脱物从1mL深孔平板转移到含有先前制备的catch 2 MyOne磁珠的Hybaid平板上(通过磁性分离除去MyOne缓冲液后)。

[0413] 将溶液在Variomag Thermoshaker (Thermo Fisher Scientific, Inc., Waltham, MA) 上于25℃以1350rpm振荡温育5分钟。

[0414] 机器人将平板转移至即用磁性分离器。将平板在磁体上温育90秒,然后除去并弃去上清。

[0415] 11. 37℃ 30%甘油洗涤

[0416] 将catch 2平板移到即用保温摇床上,将75μL的1x SB17,0.05%吐温-20转移到每孔中。将平板在37℃下以1350rpm混合1分钟以重悬并加热珠。在37℃下,向catch 2平板的每个孔中转移75μL的60%甘油,并将平板继续在37℃下以1350rpm再混合1分钟。机器人将平板转移到37℃磁性分离器上,在此将其在磁体上温育2分钟,然后机器人除去并弃去上清。将这些洗涤再重复2次。

[0417] 将第三30%甘油洗涤液从catch 2珠除去后,将150μL的1x SB17,0.05%吐温-20加入每孔中,并在37℃下以1350rpm振荡温育1分钟,然后在37℃磁体上通过磁性分离除去。

[0418] 将catch 2珠用150μL的1x SB19,0.05%吐温-20以1350rpm振荡温育1分钟来进行最后洗涤一次,然后进行磁性分离。

[0419] 12. Catch 2珠洗脱及中和

[0420] 通过向每孔加入105μL含1M NaCl,0.05%吐温-20的100mM CAPSO来将适配体从catch 2珠洗脱。将珠用这种溶液以1300rpm振荡温育5分钟。

[0421] 然后,将catch 2平板置于磁性分离器上保持90秒,然后将90μL洗脱物转移至每孔含有10μL的500mM HCl、500mM HEPES、0.05%吐温-20的新96孔平板中。转移后,通过移液90μL上下5次来机器混合溶液。

[0422] 13. 杂交

[0423] Beckman Biomek Fx^P转移20μL中和的catch 2洗脱物至新鲜的Hybaid平板,向每孔加入5μL含有10x峰值(spike)杂交对照的10x Agilent Block。然后,将25μL的2x Agilent Hybridization缓冲液手工移液入含有中和的样品和封闭缓冲液的平板的每孔中,并且通过手工缓慢移液25μL上下15次以避免大量泡沫形成而将溶液混合。将平板以1000rpm离心1分钟。

[0424] 将gasket slide置于Agilent杂交室中,将40μL每种含有杂交和封闭溶液的样品手工移液至每个gasket中。以减少气泡形成的方式使用8-通道可调(variable)多通道移液器。将定制Agilent微阵列玻片(Agilent Technologies, Inc., Santa Clara, CA)以条形码朝上缓慢下降到gasket slide上(参见Agilent手册的详细描述)。

[0425] 将杂交室上部置于玻片/背衬(backing)夹心结构上,将夹紧支架(clamping bracket)盖在整个装置上。通过旋紧螺旋来夹紧这些装置。

[0426] 目视检查每个玻片/背衬玻片夹心结构以确保溶液气泡可以在样品内自由活动。如果气泡不自由活动,则轻拍杂交室以释放靠近垫片(gasket)的气泡。

[0427] 将组装的杂交室在Agilent杂交炉中在60℃下以20rpm旋转温育19小时。

[0428] 14. 杂交后洗涤

[0429] 将约400mL的Agilent Wash Buffer 1置于两个单独的玻璃染色皿的每一个中。将一个染色皿置于磁力搅拌板上,将玻片架和搅拌棒置于缓冲液中。

[0430] 通过将搅拌棒置于空玻璃染色皿中来制备用于Agilent Wash 2的染色皿。

[0431] 将第四玻璃染色皿置于一旁以用于最终乙腈洗涤。

[0432] 分解6个杂交室中的每一个。逐个将玻片/背衬夹心结构从其杂交室移开并浸入含有Wash 1的染色皿中。用一对镊子将玻片/背衬夹心结构撬开,同时仍浸没微阵列玻片。将玻片快速转移至磁力搅拌板上的Wash 1染色皿中的玻片架中。

[0433] 将玻片架缓慢上升和降低5次。将磁力搅拌器以低设定开启,并且将玻片温育5分钟。

[0434] 当Wash 1剩余1分钟时,将在培养箱中预热至37℃的Wash Buffer 2加入第二制备的染色皿中。将玻片架快速转移至Wash Buffer 2中,并且通过将其在染色皿顶部刮擦来除去玻片架底部的任何过量缓冲液。将玻片架缓慢上升和降低5次。将磁力搅拌器以低设定开启,并且将玻片温育5分钟。

[0435] 将玻片架从Wash 2中缓慢取出,将玻片从溶液中取出需要约15秒。

[0436] 当在Wash 2中剩余1分钟时,将乙腈(ACN)加入第四染色皿中。将玻片架转移至乙腈染色皿中。将玻片架缓慢上升和降低5次。将磁力搅拌器以低设定开启,并且将玻片温育5分钟。

[0437] 将玻片架缓慢从ACN染色皿中取出并置于吸水纸上。将玻片底部边缘快速干燥,并且将玻片置于干净的玻片盒中。

[0438] 15. 微阵列成像

[0439] 将微阵列玻片置于Agilent扫描仪玻片容器中,并根据厂商指导装载至Agilent微阵列扫描仪中。

[0440] 将玻片在Cy3通道中以5μm分辨率在100%PMT设定及XRD选项为0.05的条件下成像。将所得的tiff图像用Agilent特征抽取软件版本10.5处理。

[0441] 实施例2.生物标记鉴定

[0442] 进行针对3种不同诊断应用的潜在肺癌生物标记的鉴定,即诊断来自CT扫描的可疑小结,筛查无症状吸烟者的肺癌及诊断患有肺癌的个体。血清样品从4个不同位点收集以支持这3种应用,并且包括48例NSCLC病例、由重度吸烟者和具有良性小结的患者组成的218例高风险对照。用实施例1所述的多重适配体亲和测定测量和报道这些264个样品中每一个的820个分析物的RFU值。然后对每个分析物应用KS-检验。来自样品的两个集合的值之间的KS-距离(柯尔莫可洛夫-斯米洛夫统计量)是如下程度的非参数测量,即来自一个集合(集合A)的值的经验分布与来自另一集合(集合B)的值的分布的差异程度。对于阈值T的任何值,来自集合A的某些比例的值小于T,来自集合B的某些比例的值小于T。KS-距离测量对于任何T选择来自两个集合的值的比例之间的最大(无符号)差异。

[0443] 生物标记的集合可以用于构建分类器,该分类器将样品分配为对照或疾病组。事实上,许多这样的分类器是从这些生物标记的集合产生,并且确定了任何生物标记用于良好评分分类器的频率。在顶级评分分类器中最常出现的那些生物标记对于产生诊断测试是最有用的。在本实施例中,贝叶斯分类器用于开发分类空间,但是许多其他监督学习技术可以为此目的而采用。任何单独分类器的评分适合度使用贝叶斯表面分类器的接受者操作特征曲线下面积(AUC of ROC)而判断,假定疾病发病率为0.5。这个评分度量从0到1变化,1是无误差分类器。从生物标记群体测量构建贝叶斯分类器的细节如实施例3所述。

[0444] 实施例3.肺癌的朴素贝叶斯分类

[0445] 从鉴定为可用于区分NSCLC与高风险对照组的生物标记列表,选择一组5个生物标记并构建朴素贝叶斯分类器,见表14。将类别依赖性概率密度函数(pdf) $p(x_i|c)$ 和 $p(x_i|d)$ 建模为正态分布函数,其中 x_i 为生物标记 i 的测量的RFU值的log, c 和 d 指对照和疾病群体,该函数的特征在于平均值 μ 和方差 σ^2 。5个生物标记的pdf的参数列于表15,拟合为正态pdf的原始数据与模型的实例示于图5。如图5所示,潜在的假设看起来拟合数据非常好。

[0446] 这样的模型的朴素贝叶斯分类在下述等式中给出,

$$\ln \frac{p(c|x)}{p(d|x)} = \sum_{i=1}^n \left(\ln \frac{\sigma_{d,i}}{\sigma_{c,i}} - \frac{1}{2} \left[\left(\frac{x_i - \mu_{c,i}}{\sigma_{c,i}} \right)^2 - \left(\frac{x_i - \mu_{d,i}}{\sigma_{d,i}} \right)^2 \right] \right) + \ln \frac{(1-P(d))}{P(d)} \quad \text{其中 } P(d) \text{ 为适合测试的}$$

群体中的疾病发病率,此处 $n=5$ 。求和中的每一项是单个标记的log似然比,不患有所关注的疾病(即在这种情况下是NSCLC)比患有该疾病的样品的总log似然比。这是这些单独项加上表示疾病发病率的项的简单加和。为简便起见,我们假设 $P(d) = 0.5$,从而

$$\ln \frac{(1-P(d))}{P(d)} = 0$$

[0447] 鉴于10个生物标记中每一个的log(RFU)的未知样品测量。将包含对照比疾病类别的log似然比的单独成分列表,并且可以从表15的参数和 x 的值计算。单个log似然比之和是3.47,或者无疾病比患病的似然性是32:1,其中似然性 $= e^{3.47} = 32$ 。所有5个生物标记均一致地发现偏好(favor)对照组。将似然性相乘给出上述相同结果;未知样品不患有疾病的似然性为32:1。事实上,这个样品来自训练集合中的对照群体。虽然这个实施例证实利用表15中的生物标记分类血清样品,但是相同方法可以用于任何组织类型,使用来自表21的生

物标记的任何集合。

[0448] 实施例4.选择用于分类器的生物标记组的贪婪算法

[0449] 部分1

[0450] 本实施例描述了从表21选择生物标记以形成可以在本文所述的任何方法中用作分类器的组。选择包含MMP-12的生物标记的组以及表21的生物标记的子集用于构建具有良好性能的分类器。这个方法还用于确定哪些潜在标记包括为实施例2中的生物标记。

[0451] 此处所用的分类器性能的测量是ROC曲线下面积(AUC);0.5的性能是随机(抛硬币(coin toss))分类器的基线预期值,比随机差的分类器的评分在0.0至0.5之间,比随机性能好的分类器的评分在0.5至1.0之间。无误差的完美分类器具有1.0的灵敏性、1.0的特异性和1.0的AUC。可以将实施例4的方法用于性能的其他常规测量,如F-测量、灵敏性和特异性之和或者灵敏性和特异性之积。特别地,可能想要用不同加权处理特异性和特异性,从而选择具有较高特异性而损失一些灵敏性的那些分类器,或者选择具有较高灵敏性而损失一些特异性的那些分类器。因为本文所述的方法仅涉及“性能”的一个测量,因此可以使用导致单个性能测量的任何加权方案。不同应用对于真阳性和真阴性发现会有不同益处,以及与假阳性发现和假阴性发现相关的不同成本。例如,筛查无症状吸烟者以及在CT中发现的良性小结的差异诊断通常不具有相同的特异性与灵敏性之间的最佳平衡(trade-off)。两种测试的不同需求通常需要对阳性和阴性误分类设定不同的加权,反映在性能测量中。改变性能测量通常会改变对于给定数据集合的选自表21的确切标记子集。

[0452] 对于实施例3所述的区分肺癌样品与对照样品的贝叶斯方法,通过生物标记在疾病和良性训练样品中的分布来将分类器完全参数化,并且生物标记列表选自表21;即,给定训练数据集合,选择包括的标记子集以一对一的方式确定分类器。

[0453] 此处采用的贪婪方法用于从表21检索最佳标记子集。对于小数目标标记或具有较少标记的分类器,列举每个可能的标记子集并评价用该特定标记集合构建的分类器的性能(见实施例4,部分2)。(这种方法在统计学领域公知为“最佳子集选择”;参见例如The Elements of Statistical Learning-Data Mining, Inference, and Prediction, T.Hastie, et al., editors, Springer Science+Business Media, LLC, 2nd edition, 2009)。但是,对于本文所述分类器,多个标记的组合的数目可能非常大,评价例如来自86个标记列表(表21)的5个标记的每个可能集合(即34,826,302种组合)是不可行的。因为通过每个标记子集检索不切实际,所以单个最佳子集可能不能发现;但是,通过使用这种方法,发现了许多优异的子集,并且在许多情况下,任何这些子集可以代表最佳的子集。

[0454] 代替评价每个可能的标记集合,可以进行“贪婪”正向逐步方法(“greedy”forward stepwise approach)(参见例如Dabney AR, Storey JD (2007) Optimality Driven Nearest Centroid Classification from Genomic Data. PLoS ONE 2(10):e1002.doi:10.1371/journal.pone.0001002)。使用这种方法,分类器以最佳的单个标记(基于各个标记的KS-距离)开始,并且在每步通过依次尝试标记列表中目前不是分类器中的标记集合的成员的每个成员而成长。将与现有分类器组合评分最佳的一个标记加入到分类器中。重复直至不再实现性能的进一步改善。不幸地,这种方法可能错过有价值的标记组合,一些单独的标记在方法终止前没有全部被选择。

[0455] 此处使用的贪婪方法是前述正向逐步方法的加工,为了扩大检索,而不是在每个

步骤中保持仅单个候选分类器(标记子集),保持候选分类器列表。该列表用每个单个标记子集(使用表中每个标记自身)播种。通过从目前在列表中的分类器衍生新的分类器(标记子集)并将它们加入列表中来扩大列表。通过加入不是该分类器的部分的来自表1的任何标记(其加入到子集不会复制现有子集)(这些称为“允许标记”)延伸目前在列表上的每个标记子集。每个现有标记子集通过来自列表的每个允许标记延伸。很明显,这种方法最终会产生每个可能的子集,并且该列表会用完空间。因此,所有产生的分类器仅保持到列表小于某些预定大小时(通常足以保持所有三个标记子集)。一旦列表达到预定大小限制,其变成精英(elitist);即,仅显示一定水平性能的那些分类器保持在列表上,其他分类器跌落到列表末尾并被丢弃。这通过保持按照分类器性能顺序分选列表来实现;插入至少与目前列表上最差分类器一样好的新的分类器,使得强制排除当前末尾的后进者。一种进一步执行细节是列表在每个产生步骤被完全置换;因此,列表上的每个分类器具有相同数目的标记,在每个步骤每个分类器的标记数目增加一个。

[0456] 因为这个方法使用不同标记组合产生候选分类器列表,所以人们会询问是否可以组合分类器以避免可能由最佳单个分类器或由最佳分类器的少数组产生的误差。这类“总体”和“专家委员会(committee of experts)”方法是统计学和机器学习领域公知的,并且包括例如“averaging”、“voting”、“stacking”、“bagging”和“boosting”(参见例如The Elements of Statistical Learning-Data Mining, Inference, and Prediction, T.Hastie, et al., editors, Springer Science+Business Media, LLC, 2nd edition, 2009)。简单分类器的这些组合提供了通过包括一些不同分类器及因此来自生物标记表的更大的标记集合的信息来降低分类中由于任何特定标记集合中的噪声所导致的方差的方法,这有效地在分类器之间平均。这种方法的有用性的实例是其可以防止单个标记中的异常值(outlier)负面影响单个样品的分类。测量更大数目信号的需求在常规的“一次一个标记”抗体测定中可能是不切实际的,但是对于完全多重适配体测定没有下降趋势(downside)。这些技术获益于更广泛的生物标记表,并且使用有关疾病过程的多种信息来源以提供更稳健的分类。

[0457] 部分2

[0458] 表1中选择的生物标记给出的分类器表现好于用“非标记”(即具有不满足包括在表1中的标准的信号的蛋白(如实施例2所述))构建的分类器。

[0459] 对于仅含有一个、两个和三个标记的分类器,用表1的生物标记获得的所有可能的分类器被列举并检查性能分布,与从随机选择的非标记信号的相似表构建的分类器进行比较。

[0460] 在图17和图18中,灵敏性和特异性之和用作性能的测量;1.0的性能是随机(抛硬币)分类器的基线预期值。分类器性能的柱状图与来自从40个非标记信号的“非标记”表构建的分类器的相似穷举的性能的柱状图进行比较;所述40个信号随机选自400个未证实的对照与疾病群体之间的差异信号的适配体(KS-距离<1.4)。

[0461] 图17显示从表13的生物标记参数构建的所有可能的单标记、二标记和三标记分类器对于可以区分良性小结和NSCLC的生物标记的性能的柱状图,并且将这些分类器与用40个“非标记”适配体RFU信号构建的所有可能的单标记、二标记和三标记分类器进行比较。图17A显示单标记分类器性能的柱状图,图17B显示二标记分类器性能的柱状图,而图17C显示

三标记分类器性能的柱状图。

[0462] 在图17中,实线表示使用表13中良性小结和NSCLC的生物标记数据的所有单标记、二标记和三标记分类器的分类器性能的柱状图。虚线是使用良性小结和NSCLC的数据但是使用随机非标记信号集合的所有单标记、二标记和三标记分类器的分类器性能的柱状图。

[0463] 图18显示从表12的生物标记参数构建的所有可能的单标记、二标记和三标记分类器对于可以区分无症状吸烟者和NSCLC的生物标记的性能的柱状图,并且将它们与用40个“非标记”适配体RFU信号构建的所有可能的单标记、二标记和三标记分类器进行比较。图18A显示单标记分类器性能的柱状图,图18B显示二标记分类器性能的柱状图,而图18C显示三标记分类器性能的柱状图。

[0464] 在图18中,实线表示使用表12中无症状吸烟者和NSCLC的生物标记参数的所有单标记、二标记和三标记分类器的分类器性能的柱状图。虚线是使用无症状吸烟者和NSCLC的数据但是使用随机非标记信号集合的所有单标记、二标记和三标记分类器的分类器性能的柱状图。

[0465] 从表1所列的标记构建的分类器形成了独特的柱状图,对于所有单标记、二标记和三标记比较,与用来自“非标记”的信号构建的分类器分离良好。从表1的生物标记构建的分类器的性能和AUC评分随标记数增加也比从非标记构建的分类器更快,随着每个分类器标记数增加,标记和非标记分类器之间的分离增加。用表38和39中所列的生物标记构建的所有分类器比用“非标记”构建的分类器表现更好。

[0466] 部分3

[0467] 为了测试是否标记的核心子集导致分类器的良好性能,将一半标记随机从表38和39的生物标记列表中放弃。如通过灵敏性加特异性所测量的,用于区分良性小结与恶性小结的分类器的性能略微下降0.07(从1.74到1.67),并且用于区分患有癌症的吸烟者与不患癌症的吸烟者的分类器的性能也略微降低0.06(从1.76到1.70)。生物标记表的子集的性能特征的影响是所列生物标记的多个子集有效地构建诊断测试,没有特定的核心子集标记决定分类器的性能。

[0468] 根据这些结果,测试了排除表12和13的最佳标记的分类器。图19比较了用表12和13中的生物标记完全列表构建的分类器的性能和用排除顶级标记的来自表38和39的生物标记集合构建的分类器的性能。

[0469] 图19证实未用最佳标记构建的分类器表现良好,这暗示分类器的性能不是由一些小的核心组标记所决定,并且与疾病相关的潜在过程中的变化反映在许多蛋白的活性中。表1中的许多生物标记子集表现接近最佳,甚至在除去表1的40个标记的最好的15个之后也是如此。

[0470] 图19A显示对用2-10个标记构建的用于区分良性小结与NSCLC的分类器的作用。甚至在放弃表13的15个顶级标记(由KS-距离排名)之后,良性小结比NSCLC性能也随着选自该表的标记数增加,达到1.65以上(灵敏性+特异性)。

[0471] 图19B显示对用2-10个标记构建的用于区分无症状吸烟者与NSCLC的分类器的作用。甚至在放弃表12的15个顶级标记(由KS-距离排名)之后,无症状吸烟者比NSCLC性能也随着选自该表的标记数增加,达到1.7以上(灵敏性+特异性),并且紧密接近选自表12的生物标记完全列表的最佳分类器的性能。

[0472] 最后,图20示出根据实施例3从表12和13中的参数列表构建的典型分类器的ROC性能。图20A示出来自如实施例3中假设标记的独立性的模型性能,图20B示出使用用于产生表12和13中的参数的测定数据集合的实际ROC曲线。可以看出给定数目的所选标记的性能在定性上是一致的,定量一致性随着标记数增加而降低。(这与如下观点一致,即有关疾病过程的任何特定生物标记贡献的信息与表12和13提供的其他生物标记贡献的信息是冗余的)。图20由此证实表12和13联合实施例3所述的方法允许构建和评价可用于区分NSCLC与良性小结及区分具有NSCLC的无症状吸烟者与不具有NSCLC的无症状吸烟者的非常多的分类器。

[0473] 实施例5.在Pull-down测定中证实适配体特异性

[0474] 多重测定中的最终读取基于测定中的连续捕获步骤后回收的适配体的量。多重测定基于这样的前提,即在测定结束时回收的适配体的量与原始复杂混合物(例如血浆)中的蛋白的量成比例。为了证实这个信号的确来自目的分析物而非血浆中非特异性结合的蛋白,我们开发了血浆中的基于凝胶的pull-down测定。这个测定可以用于目视证实用适配体平衡后期望的蛋白的确从血浆中拉出(pull out),以及证实与其目的蛋白靶标结合的适配体可以作为复合物幸免于测定中的动力学攻击步骤。在本实施例描述的实验中,在这个pull-down测定结束时蛋白的回收需要在平衡后蛋白保持非共价结合适配体几乎2小时。重要的是,在这个实施例中我们还提供了证据以表明在这些步骤中非特异性结合的蛋白解离并对最终信号无显著贡献。应当注意本实施例描述的pull-down程序包括上述多重测定中的所有关键步骤。

[0475] 血浆Pull-down测定

[0476] 通过将50 μ L EDTA-血浆在含0.05%吐温-20(SB18T)和2 μ M Z-Block的SB18中稀释成100 μ L来制备血浆样品。将血浆溶液用10皮摩尔(pmol) PBDC-适配体在150 μ L的终体积中于37 $^{\circ}$ C下平衡2小时。平衡后,将复合物和未结合的适配体通过在RT下于Durapore滤板中振荡温育5分钟来用133 μ L的7.5%链霉抗生物素蛋白-琼脂糖珠浆液捕获。将与珠结合的样品如实施例1所述用生物素及用缓冲液在真空中洗涤。洗涤后,将结合的蛋白用生物素稀释剂中的0.5mM NHS-S-S-生物素、0.25mM NHS-Alexa647在RT下振荡5分钟来进行标记。这个染色步骤允许生物素酰化用于将蛋白捕获在链霉抗生物素蛋白珠上,以及高灵敏染色用于在凝胶上检测。如实施例1所述,将样品用甘氨酸及用缓冲液洗涤。用Black Ray光源在RT下振荡10分钟,通过光裂解将适配体从珠上释放。在此时,生物素酰化的蛋白通过在RT下振荡5分钟而捕获在0.5mg MyOne链霉抗生物素蛋白珠上。这个步骤捕获与适配体结合的蛋白以及自从最初平衡可能与适配体解离的蛋白。如实施例1所述洗涤珠。通过用SB17T中的50mM DTT在37 $^{\circ}$ C下振荡温育25分钟来从MyOne链霉抗生物素蛋白珠洗脱蛋白。然后,将洗脱物转移至用与适配体的3'固定区互补的序列包被的MyOne珠,并且在37 $^{\circ}$ C下振荡温育25分钟。这个步骤捕获所有剩余的适配体。将珠用100 μ L SB17T洗涤1分钟,进行2次;并且用100 μ L SB19T洗涤1分钟,进行1次。通过用45 μ L的20mM NaOH振荡温育2分钟以破坏杂交的链来从这些最终珠上洗脱适配体。将40 μ L这种洗脱物用10 μ L含有0.05%吐温-20的80mM HCl中和。将相当于5%的来自珠的第一集合的洗脱物(代表结合适配体的所有血浆蛋白)和20%的来自珠的最后集合的洗脱物(代表在我们的临床测定结束时保持结合的所有血浆蛋白)的等份于还原和变性条件下在NuPAGE 4-12%Bis-Tris凝胶(Invitrogen)上进行电泳。将凝胶于

Alpha Innotech FluorChem Q扫描仪上在Cy5通道中成像以成像蛋白。

[0477] 分别选择针对LBP ($\sim 1 \times 10^{-7}$ M于血浆中, 多肽MW ~ 60 kDa)、C9 ($\sim 1 \times 10^{-6}$ M于血浆中, 多肽MW ~ 60 kDa) 和IgM ($\sim 9 \times 10^{-6}$ M于血浆中, MW ~ 70 kDa和23 kDa) 的适配体的Pull-down凝胶。(参见图16)。

[0478] 对于每个凝胶, 泳道1是来自链霉抗生物素蛋白-琼脂糖珠的洗脱物, 泳道2是最终洗脱物, 泳道3是MW标记泳道(主要条带从上至下为110、50、30、15和3.5 kDa)。从这些凝胶可以看出在最初平衡中有少量血浆蛋白的非特异性结合, 但是在进行测定的捕获步骤后仅留下靶标。很明显单个适配体试剂足以捕获其目的分析物, 无需预先耗竭或者分级血浆。然后, 在这些步骤后剩余的适配体的量与最初样品中的分析物的量成比例。

[0479] 实施例6.NSCLC手术切除的分析

[0480] 为了证实本文所述基于平台的技术鉴定来自组织的疾病相关生物标记的用途, 分析获得自8位NSCLC患者的来自手术切除的匀浆组织样品。所有NSCLC患者均为吸烟者, 年龄范围为47-75岁, 并且覆盖NSCLC 1A期至3B期(表17)。所有组织样品均通过手术期间切除5-10分钟内冷冻组织然后将组织置于OCT介质(10.24%聚乙烯醇、4.26%聚乙二醇和85.5%非反应性成分)中来获得。从每个切除获得3个样品: 肿瘤组织、邻近健康组织(在肿瘤的1 cm内)和远端无关肺组织。在保持样品恒定冷冻的同时, 切出5个10 μ m厚的切片, 从组织周围除去过量的OCT, 并且放置在冷冻的1.5 mL微量离心管中。加入200 μ L匀浆缓冲液(SB18缓冲液加上PI混合物(Pierce HALT无镁蛋白酶抑制剂混合物)后, 将样品在微量离心管中于冰上用旋转杵(rotary pestle)匀浆30秒, 直至没有组织片段可见。然后将样品在离心机中以21,000 g离心10分钟, 然后通过0.2 μ m多孔平板滤器过滤入无菌的多孔板。将5 μ L等份用于BCA蛋白测定, 并且将剩余的样品在-70 $^{\circ}$ C下冷冻保存并密封于96孔板中。

[0481] 将样品总蛋白在SB17T缓冲液(包含0.05%吐温20的SB17缓冲液)中调整至16 μ g/mL用于蛋白质组谱。将以这种方式制备的样品在多重适配体测定中运行, 如以前所述, 所述多重适配体测定测量超过800种蛋白(Ostroff et al., Nature Precedings, <http://precedings.nature.com/documents/4537/version/1> (2010))。在测量的分析物中, 大多数在肿瘤、邻近组织和远端组织之间未改变。然而, 与邻近和远端组织相比, 一些蛋白显然被抑制(图24), 而其他蛋白在肿瘤组织中显著升高(图23)。

[0482] 上述实施方案和实施例仅作为实例示出。没有特定实施方案、实施例或者特定实施方案或实施例的要素被认为是任何权利要求关键的、需要的或必需的要素或特征。另外, 本文描述的要素不是实施所附权利要求所需的, 除非明确描述为“必需”或“关键的”。可以对所公开的实施方案进行各种改变、修饰、取代和其他变化而不背离所附权利要求限定的本申请的范围。包括附图和实施例的说明书是示例性而非限制性的, 所有这样的修饰和取代包括在本申请范围内。因此, 本申请的范围应当由所附的权利要求及其法律等价物限定而非由上述实施例限定。例如, 任何方法或过程权利要求中列举的步骤可以任何可行顺序执行而限于任何实施方案、实施例或权利要求中的顺序。另外, 在任何上述方法中, 表18、表20或表21的一个或多个生物标记可以作为单个生物标记或作为来自任何组的生物标记特别地排除。

[0483]

表1 肺癌生物标记

第1列	第2列	第3列	第4列	第5列	第6列
生物标记	生物标记名称	可选蛋白名称	基因名称 (Entrez Gene Link)	良性小结比 NSCLC	吸烟者比 NSCLC
1	AMPM2	甲硫氨酸氨肽酶2 p67eIF2 p67 起始因子2-相关67 kDa 糖蛋白肽酶M2 MetAP 2 MAP 2	METAP2		X
2	Apo A-I	脱脂载脂蛋白A-1 脱脂载脂蛋白A-1	APOA1	X	
3	b-ECGF	酸性FGF FGF1 β-ECGF β-内皮细胞生长因子	FGF1	X	

[0484]

表1 (续)

4	BLC	BLC B 淋巴细胞化学吸引剂 小诱导型细胞因子B13 CXCL13 BCA-1	CXCL13	X	X
5	BMP-1	骨形态发生蛋白1 原胶原C-蛋白酶 PCP 哺乳动物tollloid蛋白 mTld	BMP1	X	X
6	BTK	酪氨酸蛋白激酶BTK 酪氨酸蛋白激酶 丙种球蛋白缺乏症酪氨酸激酶 ATK B细胞祖先激酶	BTK		X
7	C1s	补体C1s亚成分	C1S		X
8	C9	C1s, 激活的, 双链形式 补体成分C9	C9	X	X
9	钙粘着蛋白E	钙粘着蛋白-1 上皮钙粘着蛋白 E-钙粘着蛋白 桑椹胚粘着蛋白 (uvomorulin) CAM 120/80 CD 抗原=CD324	CDH1	X	
10	钙粘着蛋白-6	肾-钙粘着蛋白 K-钙粘着蛋白	CDH6	X	

[0485]

表1 (续)

11	需钙蛋白酶I	需钙蛋白酶I (需钙蛋白酶-1 催化亚基与需钙蛋白酶小亚基I 的二聚体) 催化亚基的同义词 包括需钙蛋白酶-1大亚基: 钙激活中性蛋白酶I 微摩尔需钙蛋白酶 细胞增殖诱导基因30蛋白 小亚基的同义词 包括: 钙依赖性蛋白酶小亚基I 钙激活中性蛋白酶 小亚基CAMP小亚基	CAPN1 CAPNS1	X	
12	过氧化氢酶	过氧化氢酶	CAT	X	
13	CATC	二肽基肽酶I前体 二肽基肽酶I DPP-I DPPI 组织蛋白酶C 组织蛋白酶J 二肽基转移酶	CTSC	X	
14	组织蛋白酶H	组织蛋白酶H	CTSH	X	

[0486]

表1 (续)

15	CD30 配体	肿瘤坏死因子配体超家族成员8 CD30-L CD153 抗原	TNFSF8	X	X
16	CDK5-p35	CDK5/p35是细胞分裂蛋白激酶5与 细胞周期蛋白依赖性激酶5激活蛋白1 的p35链的二聚体 细胞分裂蛋白激酶5又称为： 细胞周期蛋白依赖性激酶5 Tau蛋白激酶II催化亚基 丝氨酸/苏氨酸蛋白激酶 PSSALRE 细胞周期蛋白依赖性激酶5 激活剂1的p35链又称为： 细胞周期蛋白依赖性激酶5调节亚基1 CDK5激活剂1 细胞周期蛋白依赖性激酶5调节亚基1 Tau蛋白激酶II调节亚基	CDK5 CDK5R1		X

[0487]

表1 (续)

17	CK-MB	肌酸磷酸激酶-MB同工酶, 其是以下成分的二聚体 肌酸激酶M型与B型 肌酸激酶M与B链 M-CK与B-CK CKM与CKB	CKB CKM	X	X
18	CNDP1	β -Ala-His二肽酶 肌肽二肽酶1 CNDP二肽酶1 血清肌肽酶 谷氨酸羧肽酶样蛋白2	CNDP1	X	X
19	接触蛋白-5	神经识别分子NB-2 hNB-2	CNTN5		X
20	CSK	酪氨酸蛋白激酶CSK C-SRC激酶 蛋白酪氨酸激酶CYL	CSK	X	X
21	亲环蛋白A	亲环蛋白A 肽基脯氨酸顺反异构酶 PPIase 肽基脯氨酸异构酶 环孢菌素A结合蛋白 旋转异构酶A PPIase A	PPIA		X
22	内皮抑制素	内皮抑制素, 其切割自胶原蛋白 α -1 (VXIII)链	COL18A1		X

[0488]

表1 (续)

23	ERBB1	表皮生长因子受体 受体酪氨酸蛋白激酶 ErbB-1 EGFR HER1	EGFR	X	X
24	FGF-17	成纤维细胞生长因子-17	FGF17	X	X
25	FYN	原癌基因酪氨酸蛋白激酶Fyn 原癌基因Syn p59-Fyn	FYN		X
26	GAPDH, 肝	甘油醛-3-磷酸脱氢酶	GAPDH	X	X
27	HMG-1	高速泳动族蛋白B1 两性霉素 (amphotericin) 神经突生长促进蛋白	HMG1	X	
28	HSP 90a	热休克蛋白HSP90- α HSP 86 肾癌抗原NY-REN-38	HSP90AA1	X	X
29	HSP 90b	热休克蛋白HSP90- β HSP 90 HSP 84	HSP90AB1	X	
30	IGFBP-2	胰岛素样生长因子结合蛋白2 (IGF-结合蛋白2; IGFBP-2; IBP-2; BP2)	IGFBP2	X	X
31	IL-15 Ra	白介素-15受体亚基 α	IL15RA		X

[0489]

表1 (续)

32	IL-17B	白介素-17B 神经元白介素-17相关因子 白介素-20 细胞因子样蛋白ZCYT07	IL17B	X	
33	IMB1	输入蛋白亚基 $\beta-1$ 核周蛋白亚基 $\beta-1$ 核因子P97 输入蛋白-90	KPNB1	X	
34	激肽释放酶7	激肽释放酶-7 hK7 角质层胰凝乳蛋白酶 hSCCE 丝氨酸蛋白酶6	KLK7		X
35	KPCI	蛋白激酶C γ 型 nPKC- γ 非典型型蛋白激酶C- $\lambda/1$ aPKC- $\lambda/1$ PRKC- $\lambda/1$	PRKCI	X	X
36	LDH-H1	L-乳酸脱氢酶B链 LDH-B LDH α 亚基(LDH heart subunit) LDH-H 肾癌抗原NY-REN-46	LDHB		X

[0490]

表1 (续)

37	LGMN	豆荚蛋白 (legumain) 蛋白酶, 半胱氨酸1 天冬酰胺酰内肽酶	LGMN	X	
38	LRIG3	富亮氨酸重复序列和 免疫球蛋白样结构域蛋白3	LRIG3	X	X
39	巨噬细胞 甘露糖受体	巨噬细胞甘露糖受体1 MMR C型凝集素结构域家族13 成员D CD 抗原=CD206	MRC1	X	
40	MEK1	双特异性促分裂原活化 蛋白激酶激酶1 MAPK/ERK 激酶1 ERK激活蛋白激酶1	MAP2K1	X	X
41	METAP1	甲硫氨酸氨肽酶1 MetAP 1 MAP 1 肽酶 M1	METAP1	X	
42	中期因子	神经突增生促进蛋白 神经突增生促进因子2 妊娠中期 (midgestation) 和肾蛋白 双调蛋白相关蛋白 ARAP	MDK		X

[0491]

表1 (续)

43	MIP-5	C-C基序趋化因子15 小诱导型细胞因子A15 巨噬细胞炎症蛋白5 趋化因子CC-2 HCC-2 NCC-3 MIP-1 δ 白细胞趋素(leukotactin)-1 LKN-1 Mip-2b	MIP5		X
44	MK13	促分裂原活化蛋白激酶13 MAP激酶p38 δ 促分裂原活化蛋白激酶p38 δ 应激激活蛋白激酶4	MAPK13	X	
45	MMP-7	基质溶解因子(matrilysin) Pump-1蛋白酶 子富金属蛋白酶 基质金属蛋白酶-7 MMP-7 基质蛋白	MMP7	X	
46	NACA	新生多肽相关复合物亚基 α NAC- α α -NAC 变应原-Hom s 2	NACA	X	
47	NAGK	N-乙酰葡萄糖胺酶 GlcNAc 酶	NAGK	X	

[0492]

表1 (续)

48	PARC	C-C基序趋化因子18 小诱导型细胞因子A18 巨噬细胞炎症蛋白4 MIP-4 肺和活化调节趋化因子 CC趋化因子PARC 交替巨噬细胞活化相关CC 趋化因子1 AMAC-1 树突细胞趋化因子1 DC-CK1	CCL18	X	X
49	蛋白酶-3	蛋白酶-3 PR-3 AGP7 P29 成髓细胞蛋白酶 白细胞蛋白酶3 韦格纳自身抗原(Wegener's autoantigen) 嗜中性粒细胞蛋白酶4 NP4 C-ANCA 抗原	PRTN3	X	X
50	凝血酶原	凝血酶原 (凝血因子II)	F2	X	X

[0493]

表1 (续)

51	PTN	多营养因子 肝素结合的生长相关分子 HB-GAM 肝素结合性生长因子8 HBGF-8 成骨细胞特异性因子1 OSF-1 肝素结合神经突增生促进 因子1 HBNF-1 肝素结合脑促分裂原 HBBM	PTN	X	
52	RAC1	Ras相关C3肉毒杆菌毒素底物1 p21-Rac1 Ras样蛋白TC25 细胞迁移诱导基因5蛋白	RAC1	X	
53	肾素	肾素 血管紧张肽原酶	REN	X	
54	RGM-C	血幼素 (hemojuvelin) 血色素沉着病2型蛋白 RGM结构域家族成员C	HFE2	X	
55	SCF sR	肥大/干细胞生长因子受体 (SCFR; 原癌基因酪氨酸 蛋白激酶Kit; c-kit CD ₃₄ 抗原=CD117)	KIT	X	

[0494]

表1 (续)

56	sL-选择蛋白 (sL-选择素)	sL-选择蛋白 白细胞粘附分子-1 淋巴结归巢受体 LAM-1 L-选择蛋白 L-选择蛋白, 可溶性 白细胞表面抗原Leu-8 TQ1 gp90-MEL 白细胞-内皮细胞粘附分子1 LECAM1 CD62抗原样家族成员L	SELL	X
57	TCTP	翻译控制肿瘤蛋白 p23 组胺释放因子 HRF Fortilin	TPT1	X
58	UBE2N	泛素缀合酶E2 N 泛素蛋白连接酶N 泛素载体蛋白N Ubc13 Bendless-样泛素缀合酶	UBE2N	X
59	泛素+1	泛素	RPS27A	X

[0495]

表1 (续)

60	VEGF	血管内皮生长因子A VEGF-A 血管通透因子	VEGFA	X	
61	YES	原癌基因酪氨酸蛋白 激酶YES c-Yes p61-Yes	YES	X	

[0496] 表2 配适体浓度

靶标	最终适配体浓度 (nM)	靶标	最终适配体浓度 (nM)
AMPM2	0.5	IMB1	1
Apo A-1	0.25	激肽释放酶7	0.5
b-ECGF	2	KPCI	0.25
BLC	0.25	LDH-H 1	0.5
BMP-1	1	LGMN	0.5
ETK	0.25	LRIG3	0.25
Cl _s	0.25	巨噬细胞甘露糖受体	2
C9	1	MEK1	0.5
钙粘着蛋白E	0.25	METAP1	0.25
钙粘着蛋白-8	0.5	中期因子	0.5
需钙蛋白酶I	0.5	MIP-5	1
过氧化氢酶	0.5	MK13	1
CATC	0.5	MMP-7	0.25
组织蛋白酶H	0.5	NACA	0.5
CD30 配体	0.5	NAGK	0.5
CDK5/p35	0.5	PARC	0.5
CK-MB	1	蛋白酶-3	1
CNDP1	0.5	凝血酶原	0.5
接触蛋白-5	1	PTN	0.25
CSK		RAC1	0.5
亲环蛋白A	0.5	肾素	0.25
内皮抑制素	1	RGM-C	0.5
ERBB1	0.5	SCF sR	1
FYN	0.25	sL-选择蛋白	0.5
GAPDH,肝	0.25	TCTP	0.5
HMG-1	0.5	UBE2N	0.5
HSP 90a	0.5	泛素+1	0.5
HSP 90b	0.5	VEGF	1
IGFBP-2	1	YES	0.5
IL-15 Ra	0.5		
IL-17B	0.5		

[0497]

[0498] 表3

[0499]

位点	NSCLC	良性结节	无症状吸烟者
1	32	0	47
2	63	176	128
3	70	195	94
4	54	49	83
总和	213	420	352
男性	51%	46%	49%
女性	49%	54%	51%
中值 年龄	68	60	57
中值包年	40	42	34
中值 FEV1	1.94	2.43	2.58
中值 FEV 1%	74	88	90
中值 FEV1/FVC	70	72	73

[0500] 表4 聚集数据的良性结节-NSCLC中所鉴定的生物标记

[0501]

SCF sR	CNDP1	应激诱导的磷蛋白1
RGM-C	MEK1	LRIG3
ERBB1	MDHC	ERK-1
钙粘着蛋白E	过氧化氢酶	亲环蛋白A
CK-MB	BMP-1	胱天蛋白酶-3
METAP1	ART	UFM1
HSP90a	C9	RAC1
IGFBP-2	TCPTP	过氧化物氧还蛋白-1
需钙蛋白酶I	RPS6KA3	PAFAH β 亚基
KPCI	IMB1	MK01
MMP-7	UBC9	整合蛋白 α 1b1
β -ECGF	泛素+1	IDE
HSP90b	组织蛋白酶H	CAMK2A
NAGK	CSK21	BLC
FGF-17	BTK	BARK1
巨噬细胞甘露糖受体	凝血酶	eIF-5
MK13	LYN	UFC1
NACA	HSP70	RS7

GAPDH	UBE2N	PRKACA
CSK	TCTP	AMPM2
激活蛋白A	RabGDP解离抑制剂 β	应激诱导的磷蛋白1
凝血酶原	MAPKAPK3	

[0502] 表5 聚集数据的吸烟者-NSCLC中所鉴定的生物标记

[0503]

SCF sR	肾素	胱天蛋白酶-3
PTN	CSK	AMPM2
HSP90a	接触蛋白-5	RS7
激肽释放酶7	UBE2N	OCAD1
LRIG3	MPIF-1	HSP70
IGFBP-2	PRKACA	GSK-3 α
PARC	粒酶A	FSTL3
CD30配体	泛素+1	PAFAH β 亚基
凝血酶原	NAGK	整合蛋白a1b1
ERBB1	组织蛋白酶S	ERK-1
KPCI	TCTP	CSK21
BTK	UBC9	CATC
GAPDH, 肝	MK13	MK01
CK-MB	半胱氨酸蛋白酶抑制剂C	pTEN
LDH-H1	RPS6KA3	b2-微球蛋白
CNDP1	IL-15Ra	UFM1
RAC1	需钙蛋白酶I	UFC1
C9	MAPKAPK3	过氧化物氧还蛋白-1
FGF-17	IMB1	PKB
内皮抑制素	BARK1	IDE
亲环蛋白A	组织蛋白酶H	HSP90b
C1s	巨噬细胞甘露糖受体	BGH3
CD30	Dtk	BLC
BMP-1	NACA	XPNPEP1
SBDS	RabGDP解离抑制剂 β	TNFsR-I
MIP-5	LYN	DUS3
CCL28	METAP1	
MMP-7	MK12	

[0504] 表6 通过位点在良性结节-NSCLC中所鉴定的生物标记

[0505]

ERBB1	FGF-17
LRIG3	CD30配体
HMG-1	LGMN

YES	蛋白酶-3
C9	MEK1
MK13	BLC
巨噬细胞甘露糖受体	IL-17B
ApoA-I	CATC
CNDP1	钙粘着蛋白-6
BMP-1	

[0506] 表7 通过位点在吸烟者-NSCLC中所鉴定的生物标记

[0507]

激肽释放酶7	CSK	天青杀素
SCF sR	FYN	b2-微球蛋白
ERBB1	BLC	OCAD1
C9	TCTP	LGMN
LRIG3	中期因子	PKB
AMPM2	FGF-17	XPNPEP1
HSP90a	MEK1	钙粘着蛋白-6
sL-选择蛋白	BMP-1	pTEN
BTK	LYN	LYNB
CNDP1	整合蛋白a1b1	DUS3
CDK5-p35	PKB γ	碳酸酐酶XIII

[0508] 表8 混合数据集的良性结节-NSCLC中所鉴定的生物标记

[0509]

YES	过氧化氢酶	PAFAH β 亚基	eIF-5
MKT3	凝血酶原	AMPM2	TNFR-1
LRIG3	BTK	TCPTP	BLC
HMG-1	DRG-1	BGH3	MAPKAPK3
ERBB1	UBE2N	泛素+1	b2- 微球蛋白
钙粘着蛋白E	激活蛋白A	BARK1	SOD
CK-MB	TCTP	LYN	GSK-3 α
C9	UBC9	PRKACA	纤维蛋白原
SCFR	NAGK	LGMN	ERK-1
CNDP1	需钙蛋白酶I	整合蛋白a1b1	钙粘着蛋白-6
RGM-C	GAPDH	HSP70	IDE
METAP1	UFM1	XPNPEP1	UFC1
巨噬细胞甘露糖受体	胱天蛋白酶-3	应激诱导的磷蛋白1	PSA-ACT
BMP-1	b-ECGF	RPS6KA3	CATC
KPC1	RAC1	SHP-2	pTEN
IGFBP-2	MDHC	CEA	PSA
CSK	蛋白酶-3	OCAD1	CATE
NACA	MK01	亲环蛋白A	过氧化物氢还蛋白-1
		RabGDP 解离抑制剂 β	
IMB1	MEK1		SBDS
组织蛋白酶H	HSP90a	DUS3	RS7
			碳酸酐酶 XIII
MMP-7	凝血酶	CAMK2A	
VEGF	FGF-17	CaMKK α	
HSP90b	ART	CSK21	

[0510] 表9 混合数据集合的吸烟者-NSCLC中所鉴定的生物标记

[0511]

SCFsR	UBE2N	半胱氨酸蛋白酶抑制剂c	GSK-3 α
LRIG3	MIP-5	LYN	CATC
HSP90a	接触蛋白-5	MPII-1	SBDS
ERBB1	泛素+1	GCP-2	PAFAH β 亚基
C9	巨噬细胞甘露糖受体	KPCI	IMB1
AMPM2	PRKACA	MK12	CSK21
激肽释放酶7	组织蛋白酶S	MAPKAPK3	PKB
PTN	BMP-1	整合蛋白a1b1	Dik
PARC	亲环蛋白A	HSP70	DUS3
CD30 配体	CCL28	RPS6KA3	需钙蛋白酶I
凝血酶原	内皮抑制素	NACA	TNFR-1
CSK	组织蛋白酶H	RS7	PTP-1B
CK-MB	粒酶A	过氧化物氢还蛋白-1	IDE
BTK	GAPDH, 肝	MMP-7	HSP90b
C1s	EGF-17	pTEN	纤维蛋白原
IGFBP-2	BARK1	UFM1	胱天蛋白酶-3
LDH-H1	BLC	UBC9	PSA-ACT
RAC1	RabGDP 解离抑制剂 β	FSTL3	OCAD1
肾素	CD30	BGH3	SOD
CNDP1	MK13	UFC1	METAP1
TCTP	NAGK	MK01	PSA
IL-15Ra	b2- 微球蛋白	ERK-1	

[0512] 表10

[0513]

肺癌的生物标记	良性结节	吸烟者
AMPM2	YES	SCFsR
BMP-1	MK13	LRIG3
BTK	LRIG3	HSP90a
C1s	HMG-1	ERBB1
C9	ERBB1	C9
钙粘着蛋白E	钙粘着蛋白E	AMPM2
过氧化氢酶	CK-MB	激肽释放酶7
组织蛋白酶H	C9	PTN
CD30配体	SCFsR	PARC
CK-MB	CNDP1	CD30配体
CNDP1	RGM-C	凝血酶原
接触蛋白-5	METAP1	CSK
CSK	巨噬细胞甘露糖受体	CK-MB
ERBB1	BMP-1	BTK
HMG-1	KPCI	C1s

[0514] 表10 (续)

[0515]

HSP90a	IGFBP-2	IGFBP-2
HSP90b	CSK	LDH-H1

IGFBP-2	NACA	RAC1
IL-15Ra	IMB1	肾素
IMB1	组织蛋白酶H	CNDP1
激肽释放酶7	MMP-7	TCTP
KPCI	VEGF	IL-15Ra
LDH-H1	HSP90b	UBE2N
LRIG3	过氧化氢酶	MIP-5
巨噬细胞甘露糖受体	凝血酶原	接触蛋白-5
METAP1	ApoA-I	泛素+1
MIP-5	b-ECGF	BLC
MK13	BLC	BMP-1
MMP-7	钙粘着蛋白-6	CDK5-p35
NACA	需钙蛋白酶I	亲环蛋白A
PARC	CATC	内皮抑制素
凝血酶原	CD30配体	FGF-17
PTN	FGF-17	FYN
RAC1	GAPDH	GAPDH
肾素	HSP90a	KPCI
RGM-C	IL-17B	MEK1
SCF sR	LGMN	中期因子
TCTP	MEK1	sL-选择蛋白
UBE2N	NAGK	
泛素+1	蛋白酶-3	
VEGF		
YES		
ApoA-I		
b-ECGF		
BLC		
钙粘着蛋白-6		
需钙蛋白酶I		
CATC		
CDK5-p35		
亲环蛋白A		
内皮抑制素		
FYN		
FGf-17		
GAPDH		
IL-17B		
LGMN		
MEK1		

[0516] 表10 (续)

[0517]

中期因子		
NAGK		
蛋白酶-3		
sL-选择蛋白		

[0518] 表11

指定生物标记 的适配体	溶液 K_d (M)	测定 LLOQ (M)	上调或下调
AMPM2	3×10^{-10}	NM	上
Apo A-I	9×10^{-09}	2×10^{-11}	下
β -ECGF	1×10^{-10} (混合)	NM	上
BLC	5×10^{-10} (混合)	7×10^{-14}	上
BMP-1	2×10^{-10}	9×10^{-13}	下
BTk	8×10^{-10} (混合)	2×10^{-13}	上
Cl ₈	8×10^{-09}	7×10^{-12}	上
C9	1×10^{-11}	1×10^{-14}	下
钙粘着蛋白E	3×10^{-10}	2×10^{-12}	下
钙粘着蛋白-6	2×10^{-09}	2×10^{-12}	上
需钙蛋白酶I	2×10^{-11}	7×10^{-14}	上
过氧化氢酶	7×10^{-10} (混合)	8×10^{-14}	上
CATC	8×10^{-08}	NM	上
组织蛋白酶H	1×10^{-09} (混合)	8×10^{-13}	上
CD30 配体	2×10^{-09} (混合)	7×10^{-13}	上
CDK5/p35	2×10^{-10}	NM	上
CK-MB	1×10^{-08} (混合)	NM	下
CNDP1	3×10^{-08}	NM	下
接触蛋白-5	3×10^{-11}	NM	下
CSK	3×10^{-10}	5×10^{-13}	上
亲环蛋白A	1×10^{-09} (混合)	2×10^{-13}	上
内皮抑制素	5×10^{-10}	1×10^{-13}	上

[0520] 表11 (续)

[0521]

ERBB1	1×10^{-10}	4×10^{-13}	下
FGF-17	5×10^{-10} (混合)	NM	上
FYN	3×10^{-09} (混合)	NM	上
GAPDH	8×10^{-12}	4×10^{-13}	上
HMG-1	2×10^{-10}	1×10^{-12}	上
HSP 90 α	1×10^{-10}	1×10^{-12}	上
HSP90 β	2×10^{-10}	4×10^{-13}	上
IGFBP-2	6×10^{-10}	9×10^{-13}	上
IL-15 R α	4×10^{-11} (混合)	1×10^{-13}	上
IL-17B	3×10^{-11} (混合)	4×10^{-13}	上
IMB1	8×10^{-08} (混合)	NM	上
激肽释放酶7	6×10^{-11}	2×10^{-12}	下
KPC1	9×10^{-09}	NM	上
LDH-H1	1×10^{-09}	8×10^{-13}	上
LGMN	7×10^{-09}	NM	上
LRIG3	3×10^{-11}	8×10^{-13}	下
巨噬细胞 甘露糖受体	1×10^{-09}	1×10^{-11}	上
MEK1	6×10^{-10}	NM	上
METAP1	7×10^{-11}	9×10^{-13}	上
中期因子	2×10^{-10}	4×10^{-11}	上
MIP-5	9×10^{-09} (混合)	2×10^{-13}	上
MK13	2×10^{-09}	NM	上
MMP-7	7×10^{-11}	3×10^{-13}	上
NACA	2×10^{-11}	NM	上
NAGK	2×10^{-09} (混合)	NM	上
PARC	9×10^{-11}	1×10^{-13}	上
蛋白酶-3	5×10^{-09} (混合)	4×10^{-12}	上
凝血酶原	5×10^{-09}	1×10^{-12}	下
PTN	4×10^{-11}	5×10^{-12}	上
RAC1	7×10^{-11}	NM	上
肾素	3×10^{-11}	3×10^{-13}	上
RGM-C	3×10^{-11}	NM	下

[0522] 表11 (续)

[0523]

SCF sR	5×10^{-11}	3×10^{-12}	下
sL- 选择蛋白	2×10^{-10} (混合)	2×10^{-13}	下
TCTP	2×10^{-11} (混合)	NM	上
UBE2N	6×10^{-11} (混合)	NM	上
泛素+1	2×10^{-10}	1×10^{-12}	上
VEGF	4×10^{-10}	9×10^{-14}	上
YES	2×10^{-09}	NM	上

[0524] 表12 吸烟者对照组的参数

[0525]

表1的生物 标记编号	生物标记	μ_c	σ_c^2	μ_d	σ_d^2	KS	p-值	AUC
1	AMPM2	3.05	1.07E-02	3.20	3.62E-02	0.45	5.55E-24	0.75
4	BLC	2.58	1.23E-02	2.72	3.97E-02	0.37	8.72E-17	0.74
5	BMP-1	4.13	1.32E-02	4.00	2.01E-02	0.38	1.21E-17	0.75
6	BTK	3.12	2.44E-01	3.51	2.45E-01	0.35	3.25E-15	0.72
7	C1s	4.01	3.47E-03	4.06	4.23E-03	0.31	4.68E-12	0.69
8	C9	5.31	3.54E-03	5.38	5.37E-03	0.43	3.49E-22	0.75
15	CD30 配体	3.21	2.86E-03	3.26	4.42E-03	0.31	1.08E-11	0.70
16	CDK5-p35	2.98	3.48E-03	3.02	4.75E-03	0.25	1.63E-07	0.67
17	CK-MB	3.25	5.18E-02	3.07	4.89E-02	0.33	1.42E-13	0.71
18	CNDP1	3.65	1.97E-02	3.52	3.07E-02	0.36	4.14E-16	0.73
19	接触蛋白-5	3.66	9.35E-03	3.59	1.33E-02	0.31	1.67E-11	0.68
20	CSK	3.25	6.59E-02	3.54	1.10E-01	0.41	1.33E-20	0.76
21	亲环蛋白A	4.42	6.04E-02	4.65	6.80E-02	0.38	2.17E-17	0.73
22	内皮抑制素	4.61	4.29E-03	4.67	1.07E-02	0.32	1.42E-12	0.69
23	ERBB1	4.17	2.25E-03	4.10	5.18E-03	0.47	9.39E-27	0.78
24	FGF-17	3.08	1.12E-03	3.11	1.31E-03	0.32	1.07E-12	0.71
25	FYN	3.18	6.88E-02	3.24	7.99E-02	0.13	1.53E-02	0.58
26	GAPDH	3.26	7.32E-02	3.51	1.62E-01	0.40	2.02E-19	0.68
28	HSP90a	4.45	1.86E-02	4.61	1.86E-02	0.50	3.09E-30	0.80
30	IGFBP-2	4.30	3.42E-02	4.48	4.17E-02	0.37	5.40E-17	0.74
31	IL-15 Ra	3.03	9.74E-03	3.12	2.10E-02	0.31	7.31E-12	0.69
34	激肽释放酶7	3.52	8.67E-03	3.44	1.21E-02	0.36	2.47E-15	0.70
35	KPC1	2.58	2.92E-03	2.66	1.01E-02	0.40	2.30E-19	0.74
36	LDH-H1	3.60	8.03E-03	3.67	1.45E-02	0.32	3.70E-12	0.68
38	LRIG3	3.55	3.10E-03	3.50	3.60E-03	0.36	1.39E-15	0.72
40	MEK1	2.81	1.54E-03	2.84	2.75E-03	0.28	1.96E-09	0.67
42	中期因子	3.21	3.13E-02	3.24	5.58E-02	0.13	1.90E-02	0.56
43	MIP-5	3.60	3.65E-02	3.77	5.88E-02	0.34	8.40E-14	0.70
48	PARC	4.90	1.94E-02	5.01	2.13E-02	0.34	7.01E-14	0.71
50	凝血酶原	4.68	5.37E-02	4.53	4.31E-02	0.32	1.09E-12	0.68
51	PTN	3.73	7.08E-03	3.80	7.36E-03	0.34	3.97E-14	0.72
52	RAC1	3.85	6.13E-02	4.09	7.31E-02	0.40	4.60E-19	0.72
53	肾素	3.25	2.52E-02	3.39	6.36E-02	0.30	4.23E-11	0.68
55	SCF sR	3.79	1.11E-02	3.68	1.48E-02	0.37	9.90E-17	0.75
56	sL- 选择蛋白	4.46	5.63E-03	4.40	9.30E-03	0.30	6.24E-11	0.69
57	TCTP	4.19	4.69E-02	4.44	7.43E-02	0.43	9.69E-22	0.76
58	UBE2N	4.42	9.30E-02	4.67	9.53E-02	0.34	6.56E-14	0.72
59	泛素+1	4.25	1.75E-02	4.34	1.43E-02	0.31	1.55E-11	0.68

[0526] 表13 良性结节对照组参数

[0527]

表1的生物 标记编号	生物标记	μ_c	σ_c^2	μ_d	σ_d^2	KS	p-值	AUC
2	ApoA-I	3.83	1.04E-02	3.77	1.56E-02	0.24	1.67E-07	0.65
3	b-ECGF	3.03	1.27E-03	3.06	1.53E-03	0.30	7.50E-12	0.68
4	BLC	2.60	1.50E-02	2.72	3.97E-02	0.31	1.77E-12	0.70
5	BMP-1	4.11	1.39E-02	4.00	2.01E-02	0.32	2.00E-13	0.72
8	C9	5.31	4.84E-03	5.38	5.37E-03	0.39	9.42E-20	0.75
9	钙粘着蛋白E	4.51	5.91E-03	4.43	9.86E-03	0.37	1.93E-17	0.74
10	钙粘着蛋白-6	2.91	3.79E-03	2.98	1.12E-02	0.36	1.42E-16	0.72
11	需钙蛋白酶I	4.37	1.33E-02	4.50	2.32E-02	0.40	7.63E-21	0.75
12	过氧化氢酶	4.27	2.09E-02	4.37	1.30E-02	0.34	4.30E-15	0.72
13	CATC	2.80	5.83E-03	2.86	7.63E-03	0.31	8.55E-13	0.69
14	组织蛋白酶H	4.59	3.24E-03	4.63	7.54E-03	0.30	4.29E-12	0.66
15	CD30 配体	3.21	4.19E-03	3.26	4.42E-03	0.26	4.70E-09	0.68
17	CK-MB	3.23	4.47E-02	3.07	4.88E-02	0.32	2.76E-13	0.70
18	CNDP1	3.65	2.03E-02	3.52	3.07E-02	0.35	2.04E-15	0.72
20	CSK	3.25	7.98E-02	3.54	1.10E-01	0.41	2.35E-21	0.76
23	ERBB1	4.17	2.76E-03	4.10	5.18E-03	0.46	1.22E-26	0.77
24	FGF-17	3.08	1.26E-03	3.11	1.31E-03	0.31	9.59E-19	0.71
26	GAPDH	3.22	7.96E-02	3.51	1.62E-01	0.40	7.88E-21	0.69
27	HMG-1	4.01	4.57E-02	4.19	7.55E-02	0.30	1.99E-11	0.70
28	HSP90a	4.43	2.23E-02	4.61	1.86E-02	0.51	1.26E-33	0.81
29	HSP90b	3.06	3.70E-03	3.14	9.67E-03	0.42	2.73E-22	0.75
30	IGFBP-2	4.32	3.57E-02	4.48	4.17E-02	0.35	2.30E-15	0.73
32	IL-17B	2.19	3.73E-03	2.23	4.16E-03	0.28	3.65E-10	0.68
33	IMB1	3.47	2.21E-02	3.67	5.45E-02	0.42	2.04E-22	0.75
35	KPC1	2.57	3.26E-03	2.66	1.01E-02	0.43	3.57E-23	0.75
37	LGMN	3.13	2.03E-03	3.17	4.15E-03	0.30	1.15E-11	0.69
38	LRIQ3	3.55	3.59E-03	3.50	3.60E-03	0.33	9.00E-14	0.71
39	巨噬细胞 甘露糖受体	4.10	1.51E-02	4.22	2.48E-02	0.36	7.24E-17	0.72
40	MEK1	2.81	1.77E-03	2.84	2.75E-03	0.31	3.79E-12	0.69
41	METAP1	2.67	2.45E-02	2.99	5.83E-02	0.44	2.99E-24	0.75
44	MK13	2.79	3.38E-03	2.85	4.88E-03	0.36	6.16E-17	0.74
45	MMP-7	3.64	3.24E-02	3.82	4.85E-02	0.37	1.89E-17	0.73
46	NACA	3.11	8.28E-03	3.21	2.63E-02	0.34	4.91E-15	0.70
47	NAGK	3.71	2.04E-02	3.84	2.63E-02	0.38	7.50E-19	0.73
49	蛋白酶-3	3.95	9.09E-02	4.18	1.23E-01	0.30	2.22E-11	0.69
50	凝血酶原	4.67	4.19E-02	4.53	4.31E-02	0.32	2.17E-13	0.68
54	RGM-G	4.44	4.85E-03	4.38	8.13E-03	0.30	1.00E-11	0.69
55	SCF sR	3.77	9.71E-03	3.68	1.48E-02	0.35	1.96E-15	0.72
60	VEGF	3.55	8.80E-03	3.62	1.14E-02	0.30	1.27E-11	0.69
61	YES	2.97	9.54E-04	3.00	1.73E-03	0.29	7.59E-11	0.67

[0529]

表15 源自朴素贝叶斯分类器的训练组的参数

生物标记	μ_c	μ_d	σ_c	σ_d	$\bar{x}$	$p(c \bar{x})$	$p(d \bar{x})$	$\ln(p(c \bar{x})/p(d \bar{x}))$
C9	11.713	11.934	0.199	0.210	11.667	1.946	0.843	0.836
LRIG3	7.409	7.307	0.090	0.084	7.372	4.058	3.511	0.145
GAPDH	9.027	9.385	0.511	0.230	9.000	0.780	0.428	0.599
MMP12	6.139	6.346	0.096	0.255	6.129	4.115	1.087	1.332
KLK7	8.130	7.979	0.230	0.298	8.419	0.789	0.450	0.562

[0530] 表16 表21中所有标记的组织和血清的朴素贝叶斯参数

[0531]

生物标记	组织 μ_c	组织 μ_d	组织 σ_c	组织 σ_d	血清 μ_c	血清 μ_d	血清 σ_c	血清 σ_d
活化素A	5.927	6.713	0.124	0.816	6.990	7.060	0.089	0.120
脂联素	9.357	8.560	0.456	0.154	8.986	9.141	0.406	0.391
AMPM2	9.352	9.918	0.393	0.313	7.067	7.079	0.091	0.107
Apo A-I	6.554	6.573	0.312	0.271	8.699	8.593	0.139	0.130
b-ECGF	6.731	7.256	0.606	0.739	6.205	6.160	0.056	0.072
BGN	7.989	6.821	0.653	0.168	7.140	7.067	0.125	0.077
BLC	6.283	7.776	0.253	0.971	7.065	7.058	0.066	0.072
BMP-1	5.149	5.377	0.071	0.239	8.766	8.548	0.213	0.234
BTX	8.782	7.757	0.547	1.303	7.567	7.856	0.464	0.304
C1s	7.989	7.973	0.206	0.298	8.532	8.540	0.106	0.121
C9	11.488	11.417	0.325	0.380	11.715	11.936	0.189	0.223
钙粘着蛋白-6	7.129	7.502	0.034	0.265	7.971	7.959	0.087	0.067
钙粘着蛋白E	7.370	7.916	0.458	0.349	9.252	9.050	0.200	0.181
需钙蛋白酶I	9.641	9.962	0.503	0.553	10.358	10.466	0.132	0.143
碳酸酐酶III	9.096	7.504	0.288	0.890	8.552	8.687	0.474	0.351
半胱天冬酶-3	6.758	7.426	0.248	0.110	7.097	7.136	0.338	0.367
过氧化氢酶	11.253	10.581	0.127	0.633	10.051	10.243	0.392	0.276
CATC	7.839	7.783	0.457	0.385	7.248	7.229	0.088	0.088
组织蛋白酶H	12.452	11.758	0.158	0.410	9.485	9.585	0.124	0.210
CD30配体	6.655	6.911	0.069	0.168	7.622	7.605	0.038	0.035
CD36抗原	7.026	6.262	0.478	0.175	8.252	8.224	0.114	0.141
CDK5/p35	6.581	6.741	0.210	0.215	6.986	7.044	0.083	0.075
CX-M8	8.912	8.564	0.703	0.611	7.515	7.230	0.317	0.307
CNDP1	7.293	7.292	0.062	0.189	9.995	9.754	0.295	0.375
接触蛋白-5	5.793	5.777	0.063	0.146	6.749	6.689	0.109	0.141
CSK	8.526	8.181	0.370	0.736	6.809	7.186	0.388	0.245
CXCL16, 可溶	8.216	7.559	0.517	0.449	9.660	9.744	0.185	0.230
亲环素A	11.751	11.668	0.159	0.123	8.586	8.784	0.323	0.233
内皮他丁	8.669	8.096	0.233	0.374	8.763	8.876	0.125	0.162
ERBB1	7.041	7.263	0.134	0.336	10.578	10.428	0.119	0.135
ESAM	8.659	7.451	0.376	0.473	9.022	9.033	0.151	0.142
FGF-17	6.111	5.998	0.036	0.066	6.897	6.902	0.062	0.069
纤连蛋白	9.681	10.795	0.452	1.097	11.288	11.105	0.253	0.269
FYN	8.003	7.834	0.149	0.262	8.002	8.033	0.123	0.086
GAPDH, 肝	12.703	12.713	0.123	0.152	9.033	9.410	0.536	0.194
HMG-1	11.639	11.541	0.545	0.615	8.430	8.546	0.133	0.096
HSP 90a	11.569	11.820	0.479	0.279	9.165	9.343	0.226	0.182
HSP 90b	8.509	9.422	0.974	0.960	7.635	7.653	0.053	0.059

[0532] 表16 (续)

[0533]

IDE	8.426	9.023	0.362	0.302	7.670	7.728	0.096	0.106
IGFBP-2	7.715	9.591	0.416	1.413	8.514	9.006	0.417	0.448
IGFBP-5	7.619	9.347	0.282	1.263	9.705	9.675	0.126	0.138
IGFBP-7	8.999	9.843	0.717	0.307	9.251	9.156	0.148	0.172
IL-15 Ra	6.088	6.577	0.123	0.318	7.058	7.066	0.056	0.071
IL-178	5.441	5.531	0.051	0.139	6.267	6.257	0.052	0.066
IL-8	7.037	8.206	0.145	0.631	7.114	7.109	0.052	0.066
IMB1	5.867	6.218	1.300	1.010	7.326	7.390	0.150	0.152
激肽释放酶7	5.990	6.152	0.146	0.447	8.132	7.964	0.221	0.295
KPC1	6.589	6.821	0.244	0.420	6.195	6.194	0.053	0.046
LDH-H 1	12.527	12.640	0.135	0.169	7.221	7.261	0.140	0.198
LGMN	7.964	8.124	0.084	0.101	8.404	8.377	0.074	0.070
LRIG3	6.198	6.213	0.383	0.336	7.411	7.301	0.090	0.092
巨噬细胞 甘露糖受体	6.738	5.654	0.394	0.440	8.132	8.233	0.203	0.253
MEK1	6.543	6.657	0.305	0.505	5.979	5.966	0.039	0.048
METAP1	9.004	9.807	0.540	0.412	7.955	7.982	0.095	0.081
Midkine	6.619	7.223	0.770	1.112	7.714	7.714	0.298	0.193
MIP-5	5.582	5.657	0.041	0.090	8.560	8.689	0.262	0.233
MK13	7.195	7.793	0.260	0.491	NA	NA	NA	NA
MMP-12	5.822	8.677	0.182	1.045	6.129	6.323	0.100	0.260
MMP-7	6.800	8.224	0.440	0.215	8.881	9.232	0.235	0.182
NACA	6.480	6.738	0.207	0.183	7.774	7.791	0.111	0.108
NAGK	9.469	9.986	0.328	0.457	7.385	7.476	0.203	0.216
NAP-2	10.672	9.447	0.357	0.842	7.765	7.775	0.286	0.342
PARC	9.519	9.315	0.537	0.169	10.087	10.291	0.424	0.369
蛋白水解酶-3	7.667	6.963	0.789	0.850	8.340	8.394	0.461	0.504
凝血酶原	7.245	7.400	0.443	0.390	NA	NA	NA	NA
P-选择素	7.947	6.593	0.263	0.508	9.937	9.944	0.278	0.199
PTN	7.363	7.301	0.492	0.531	8.149	8.250	0.116	0.152
RAC1	11.522	11.299	0.109	0.220	8.408	8.697	0.378	0.323
肾素	5.964	5.894	0.039	0.080	7.675	7.797	0.338	0.506
RGM-C	6.677	6.646	0.049	0.084	9.765	9.700	0.164	0.180
SCF sR	6.607	6.639	0.163	0.175	9.603	9.503	0.139	0.141
SLPI	10.635	9.435	0.676	0.476	NA	NA	NA	NA
sL-选择素	6.524	6.827	0.149	0.166	NA	NA	NA	NA
sRAGE	11.154	7.304	0.619	0.912	7.001	6.845	0.333	0.297
TCTP	10.524	10.395	0.087	0.127	8.847	9.137	0.290	0.224
凝血酶敏感蛋白-1	9.012	10.305	0.520	1.093	9.187	8.950	0.558	0.349
TPS82	10.798	9.138	0.668	1.055	7.714	7.435	0.346	0.441

[0534] 表16 (续)

[0535]

TrATPase	11.031	8.887	0.993	0.703	9.099	9.168	0.204	0.148
TSP2	6.569	7.837	0.085	0.627	7.468	7.562	0.162	0.218
UBE2N	10.654	10.725	0.166	0.140	9.234	9.487	0.521	0.288
泛素+1	10.948	10.860	0.249	0.275	9.218	9.284	0.249	0.171
uPA	5.747	6.564	0.119	0.445	6.868	6.874	0.104	0.126

URB	7.180	8.539	0.283	0.699	8.689	8.756	0.173	0.202
VEGF	6.313	7.593	0.088	1.074	7.699	7.769	0.096	0.145
vWF	7.927	7.193	0.263	0.139	10.531	10.684	0.236	0.200
YES	7.086	7.723	0.386	0.314	6.593	6.605	0.065	0.067

[0536] 表17 8个NSCLC样品的患者人口统计学、切除位置和肿瘤类型

[0537]

年龄	性别	吸烟史	位置	阶段	组织Dx
47	F	吸烟者	左上叶	pT3pN1pMx 阶段IIIA	有病灶鳞状分化的分化不足的非小细胞CA
73	F	吸烟者	左下叶	pT2pN0pMx 阶段IB	分化不足的鳞状细胞癌
48	M	吸烟者	右上叶	pT2pN1pMx 阶段IIIA	分化不足的鳞状细胞癌
60	F	吸烟者	左上叶	T4 N1 M0 阶段IIIB - 注意基于临床肺衰退的T4差别; 肿瘤根据大小标准为pT2	分化不足的鳞状细胞癌
51	F	吸烟者	右上叶	pT2pN0pMx 阶段IB	适度分化的腺癌
71	F	吸烟者	右上叶	pT2pN0pMx 阶段IB	充分分化的腺癌
75	F	吸烟者	右下叶	pT1N0Mx 阶段IA	充分分化的腺癌
73	M	吸烟者	左上叶	pT1bN0Mx 阶段IA	非典型类癌肿瘤(即神经内分泌,嗜铬粒蛋白IHC阳性)

[0538] 表18 肿瘤和正常组织间差异表达的生物标记

[0539]

生物标记#	生物标记命名	可选蛋白名称	基因	上调/下调
1	活化素A	抑制素β-A链 活化素β-A链 红细胞分化蛋白 EDF	INHBA	上调
2	脂连素	30 kDa脂肪细胞补体相关蛋白 脂肪细胞补体相关30 kDa蛋白 ACRP30 脂肪细胞, 含C1q和胶原结构域的蛋白 脂肪最丰富的基因转录物1蛋白 apM-1 明胶结合蛋白 Adipolemi	ADIPOQ	下调
3	BCA-1*	C-X-C基序趋化因子13 Angie B细胞-吸引趋化因子1 B淋巴细胞化学吸引剂 CXC趋化因子BLC 小-可诱导趋化因子B13 BLC	CXCL13	上调
4	双糖链蛋白聚糖	骨/软骨蛋白聚糖I PG-S1	BGN	下调

[0540] 表18 (续)

[0541]

5	钙粘着蛋白-1*	CAM 120/80 上皮钙粘着蛋白 E-钙粘着蛋白 桑格粘着蛋白 CD234	CDH1	上调
6	碳酸酐酶III	碳酸酐酶3 碳酸酐酶III CA-III	CA3	下调
7	半胱天冬酶-3	CASP-3 APOPA1N 半胱氨酸蛋白酶CPP32 CPP-32 蛋白Yama蛋白酶 SRBEP酶切活性1 SCAS-1	CASP3	上调
8	过氧化氢酶*		CAT	下调
9	CD36抗原	血小板糖蛋白4 脂肪酸移位酶 FAT 糖蛋白IIb GPIIb 白细胞分化抗原CD36 PAS-IV PAS-4 血小板胶原受体 血小板糖蛋白IV GPIV 凝血酶敏感蛋白受体	CD36	下调
10	CXC16, 可溶	C-X-C基序趋化因子16 磷脂酰丝氨酸和氧化的低密度脂蛋白 的净化剂受体 SR-P3OX 小-可诱导趋化因子E16 跨膜趋化因子CXCL16	CXCL16	下调
11	内皮他丁*		COL18A1	下调
12	ESAM	内皮细胞-选择性粘附分子	ESAM	下调
13	纤连蛋白	FN 冷-可溶球蛋白 CIG FNT	FN1	上调
14	胰岛素溶酶	胰岛素降解酶 A- β -降解蛋白酶 胰岛素蛋白酶 胰岛素酶	IDE	上调
15	IGFBP-2*	类胰岛素生长因子-结合蛋白2 IBF-2 IGF-结合蛋白2	IGFBP2	上调
16	IGFBP-5	类胰岛素生长因子-结合蛋白5 IBF-5 IGF-结合蛋白5	IGFBP5	上调

[0542] 表18 (续)

[0543]

17	IGFBP-7	类胰岛素生长因子-结合蛋白7 IBP-7 IGFBP-rP1 MAC25蛋白 FG12-刺激因子 前列腺素-刺激因子 源自肿瘤的粘附因子 TAP	IGFBP7	上调
18	IL-8	白介素-8 C-X-C基序趋化因子8 依其白介素 粒细胞趋化蛋白1 GCP-1 源自单核细胞的嗜中性粒细胞趋化因子 MUNCT 源自单核细胞的嗜中性粒细胞活化肽 MONAP 嗜中性粒细胞活化蛋白 NAP-1 蛋白3-10C T细胞趋化因子	IL8	上调
19	MRC1*	巨噬细胞甘露糖受体1 MNR C型凝集素结构域家族13成员D 类-C型凝集素结构域家族13成员D 类-巨噬细胞甘露糖受体1蛋白1 CD206	MRC1	下调
20	MAPK13*	促细胞分裂原-活化的蛋白激酶13 MAP激酶13 促细胞分裂原-活化的蛋白激酶p38 δ MAP激酶p38 δ 应激-激活的蛋白激酶4	MAPK13	上调
21	MMP-7*	基质溶解因子 基质蛋白 基质金属蛋白酶-7 Pump-1蛋白酶 子宫金属蛋白酶	MMP7	上调
22	MMP-12*	巨噬细胞金属弹性蛋白酶 MME 巨噬细胞弹性蛋白酶 ME hME 基质金属蛋白酶-12	MMP12	上调
23	NAGK*	N-乙酰基-D-葡萄糖胺酶 N-乙酰基葡萄糖胺酶 GlcNac酶	NAGK	上调
24	NAP-2	嗜中性粒细胞-活化肽2	PPRP	下调
25	P-选择素	类-CD62抗原家族成员F 颗粒膜蛋白140 GMP-140 白细胞-内皮细胞粘附分子3 LECAM-3 血小板活化依赖性颗粒-外膜蛋白 PADGEM CD62P	SELP	下调

[0544] 表18 (续)

[0545]

26	SLPI	抗白细胞蛋白酶 ALP BLP1 HUS1-1 粘膜蛋白酶抑制剂 MPI 蛋白酶抑制剂WAP4 分泌性白细胞蛋白酶抑制剂 精液蛋白酶抑制剂 WAP四-二硫核结构域蛋白4	SLPI	下调
27	sRAGE	高度糖基化终产物特异性受体 用于高度糖基化终产物的受体	AGER	下调
28	凝血酶敏感蛋白-1	TSP-1	THBS1	上调
29	凝血酶敏感蛋白-2	TSP-2	THBS1	上调
30	TrATPase	5-型酒石酸盐抗性酸性磷酸酶 TR-AP 酒石酸盐抗性酸性ATPase 5-型酸性磷酸酶	ACP5	下调
31	类胰蛋白酶-2	类胰蛋白酶-2 类胰蛋白酶-2 类胰蛋白酶II TRYB2	TPSB2	下调
32	uPA	尿激酶类纤溶酶原激活物 U-纤溶酶原激活物 尿激酶	PLAU	上调
33	URB	含卷曲螺旋结构域的蛋白90 由癌基因下调的蛋白1 在BRS-3缺陷小鼠中上调的同系物	CCDC80	上调
34	VEGF*	血管内皮生长因子A VEGF-A 血管渗透因子 VFP	VEGFA	上调
35	vWF	von Willebrand因子	VWF	下调
36	YES*	酪氨酸-蛋白酶Yes 原癌基因c-Yes PE1-Yes	YES1	上调

[0546] *在血清和肿瘤组织中都表达的生物标记的重叠

[0547] 表19 NSCLC组织生物标记依照生物学过程分类

[0548]

血管发生	生长和代谢	炎症&凋亡	侵入、转移 (ECM)
VEGF	脂连素*	活化素A	双糖链蛋白聚糖*
内皮他丁	碳酸酐酶III*	BCA-1*	钙粘着蛋白-1
凝血酶敏感蛋白-1	IGFBP-2	过氧化氢酶	CD36抗原
凝血酶敏感蛋白-2	IGFBP-5	CXCL16, 可溶*	ESAM
	IGFBP-7	IL-8	纤连蛋白*
	胰岛素溶酶*	MRC1*	MMP-7
	NAGK*	NAP-2	MMP-12
	TrATPase*	sRAGE	P-选择素*
	类胰蛋白酶-b2	SLPI	URB*
	MAPK13*	uPA	vWF
		半胱天冬酶-3	凝血酶敏感蛋白-1
			凝血酶敏感蛋白-2
			YES

[0549] *新NSCLC生物标记

[0550] 表20 NSCLC组织中鉴别出的生物标记*

[0551]

生物标记#	生物标记命名
1	活化素A
2	脂连素
3	双糖链蛋白聚糖
4	碳酸酐酶III
5	半胱天冬酶-3
6	CD36抗原
7	CXCL16,可溶
8	ESAM
9	纤连蛋白
10	胰岛素溶酶
11	IGFBP-5
12	IGFBP-7
13	IL-8
14	MMP-12
15	NAP-2
16	P-选择素
17	SLPI
18	sRAGE
19	凝血酶敏感蛋白-1
20	凝血酶敏感蛋白-2
21	TrATPase
22	类胰蛋白酶 β -2
23	uPA
24	URB
25	vWF

[0552] *此列表不包括在组织和血清样品中均鉴别出的生物标记

[0553] 表21血清和组织中鉴别出的生物标记

[0554]

生物标记	生物标记命名
1	活化素A
2	脂联素
3	AMPM2
4	Apo A-I
5	双糖链蛋白聚糖
6	b-ECGF
7	BLC*

8	BMP-1
9	BTK
10	C1s
11	C9
12	钙粘着蛋白E*
13	钙粘着蛋白-6
14	需钙蛋白酶I
15	碳酸酐酶III
16	半胱天冬酶-3
17	过氧化氢酶*
18	CATC
19	组织蛋白酶H
20	CD30配体
21	CD36抗原
22	CDK5-p35
23	CK-MB
24	CNDP1
25	接触蛋白-5
26	CSK
27	CXCL 16,可溶
28	亲环素A
29	内皮他丁*
30	ERBB1
31	ESAM
32	FGF-17
33	纤连蛋白
34	FYN
35	GAPDH,肝
36	HMG-1
37	HSP90a
38	HSP90b
39	IGFBP-2*
40	IGFBP-5
41	IGFBP-7
42	IL-8
43	IL-15Ra
44	IL-17B
45	IMB1
46	胰岛素溶酶
47	激肽释放酶7

48	KPCI
49	LDH-H1

[0555] 表21 (续)

[0556]

50	LGMN
51	LRIG3
52	巨噬细胞甘露糖受体*
53	MEK1
54	METAP1
55	中期因子
56	MIP-5
57	MK13*
58	MMP-7*
59	MMP-12*
60	NACA
61	NAGK*
62	NAP-2
63	PARC
64	P-选择素
65	蛋白水解酶-3
66	凝血酶原
67	PTN
68	RAC1
69	肾素
70	RGM-C
71	SCF sR
72	SLPI
73	sL-选择素
74	sRAGE
75	TCTP
76	凝血酶敏感蛋白-1
77	凝血酶敏感蛋白-2
78	TrATPase
79	类胰蛋白酶 β -2
80	UBE2N
81	泛素+1
82	uPA
83	URB
84	VEGF*
85	vWF

86	YES*
----	------

[0557] *在血清组织中均鉴别出的生物标记

[0558] 表22 81组包括MMP-12的双生物标记

[0559]

标记			AUC
1	CSK	MMP-12	0.848
2	GAPDH 肝	MMP-12	0.842
3	亲环素A	MMP-12	0.832
4	ICTP	MMP-12	0.831
5	C9	MMP-12	0.828
6	LRIG3	MMP-12	0.826
7	MMP-7	MMP-12	0.824
8	BMP-1	MMP-12	0.823
9	SCF sR	MMP-12	0.823
10	ERBB1	MMP-12	0.822
11	RAC1	MMP-12	0.822
12	激肽释放酶7	MMP-12	0.822
13	HSP 90a	MMP-12	0.817
14	CDK5p35	MMP-12	0.815
15	IGFBP-2	MMP-12	0.812
16	HMG-1	MMP-12	0.809
17	钙粘着蛋白E	MMP-12	0.808
18	S-ECGF	MMP-12	0.807
19	基质蛋白酶I	MMP-12	0.805
20	RGM-C	MMP-12	0.804
21	IMB1	MMP-12	0.802
22	UBE2N	MMP-12	0.802
23	LGMN	MMP-12	0.801
24	过氧化氢酶	MMP-12	0.801
25	CK-MB	MMP-12	0.800
26	DTK	MMP-12	0.799
27	内皮他丁	MMP-12	0.791
28	BCN	MMP-12	0.791
29	PTN	MMP-12	0.790
30	CD30 配体	MMP-12	0.789
31	活化素A	MMP-12	0.785
32	vWF	MMP-12	0.784
33	TSP2	MMP-12	0.784
34	IL-8	MMP-12	0.782
35	脂联素	MMP-12	0.781
36	凝血酶敏感蛋白-1	MMP-12	0.779
37	NAGK	MMP-12	0.777
38	MIP-5	MMP-12	0.776
39	VEGF	MMP-12	0.776
40	NACA	MMP-12	0.773
41	LDH-H 1	MMP-12	0.771
42	CNDP1	MMP-12	0.770
43	IGFBP-7	MMP-12	0.770
44	蛋白水解酶-3	MMP-12	0.769
45	TPSB2	MMP-12	0.769
46	Apo A-1	MMP-12	0.768
47	巨噬细胞甘露糖受体	MMP-12	0.768
48	泛素+1	MMP-12	0.767
49	IDE	MMP-12	0.767
50	组织蛋白酶H	MMP-12	0.766
51	CXCL16, 可溶	MMP-12	0.763
52	TrATPase	MMP-12	0.762

[0560] 表22 (续)

[0561]

53	半胱天冬酶-3	MMP-12	0.757
54	钙粘着蛋白-6	MMP-12	0.757
55	接触蛋白-5	MMP-12	0.756
56	BLC	MMP-12	0.756
57	FGF-17	MMP-12	0.755
58	纤连蛋白	MMP-12	0.754
59	NAP-2	MMP-12	0.754
60	HSP 90b	MMP-12	0.754
61	C1s	MMP-12	0.753
62	AMPM2	MMP-12	0.752
63	IL-17B	MMP-12	0.752
64	IL-15Ra	MMP-12	0.751
65	uPA	MMP-12	0.750
66	PARC	MMP-12	0.749
67	IGFBP-5	MMP-12	0.748
68	肾素	MMP-12	0.745
69	KPCI	MMP-12	0.742
70	METAP1	MMP-12	0.742
71	碳酸酐酶III	MMP-12	0.740
72	CATC	MMP-12	0.740
73	MEK1	MMP-12	0.740
74	URB	MMP-12	0.736
75	CD36抗原	MMP-12	0.735
76	中期因子	MMP-12	0.735
77	sRAGE	MMP-12	0.731
78	ESAM	MMP-12	0.729
79	YES	MMP-12	0.728
80	P-选择素	MMP-12	0.723
81	FYN	MMP-12	0.707

[0562] 表23 100组包括MMP-12的三标记物

[0563]

标记				AUC
1	C9	GAPDH, 肝	MMP-12	0.879
2	MMP-7	GAPDH, 肝	MMP-12	0.876
3	C9	CSK	MMP-12	0.875
4	BMP-1	CSK	MMP-12	0.869
5	BMP-1	GAPDH, 肝	MMP-12	0.868
6	LRIG3	GAPDH, 肝	MMP-12	0.867
7	激肽释放酶7	GAPDH, 肝	MMP-12	0.867
8	MMP-7	CSK	MMP-12	0.867
9	RAC1	C9	MMP-12	0.865
10	CSK	激肽释放酶7	MMP-12	0.865
11	IGFBP-2	GAPDH, 肝	MMP-12	0.864
12	CDK5/p35	GAPDH, 肝	MMP-12	0.862
13	C9	TCTP	MMP-12	0.862
14	C9	亲环素A	MMP-12	0.862
15	LRIG3	CSK	MMP-12	0.862
16	SCF sR	CSK	MMP-12	0.861
17	SCF sR	GAPDH, 肝	MMP-12	0.861
18	IGFBP-2	CSK	MMP-12	0.860
19	MMP-7	TCTP	MMP-12	0.860
20	b-ECGF	GAPDH, 肝	MMP-12	0.860
21	ERBB1	CSK	MMP-12	0.859
22	RAC1	BMP-1	MMP-12	0.858
23	LRIG3	TCTP	MMP-12	0.858
24	ERBB1	GAPDH, 肝	MMP-12	0.857
25	BMP-1	TCTP	MMP-12	0.857
26	RGM-C	CSK	MMP-12	0.857
27	CSK	b-ECGF	MMP-12	0.856
28	RAC1	激肽释放酶7	MMP-12	0.856
29	CDK5/p35	CSK	MMP-12	0.856
30	HMG-1	MMP-7	MMP-12	0.856
31	CK-MB	GAPDH, 肝	MMP-12	0.855
32	RAC1	CDK5/p35	MMP-12	0.855
33	CSK	凝血酶敏感蛋白-1	MMP-12	0.854
34	C9	BTB	MMP-12	0.854
35	RAC1	LRIG3	MMP-12	0.854
36	HSP 90 α	C9	MMP-12	0.854
37	活化素A	GAPDH, 肝	MMP-12	0.854
38	HSP 90 α	BMP-1	MMP-12	0.854
39	内皮他丁	CSK	MMP-12	0.853
40	CSK	GAPDH, 肝	MMP-12	0.853
41	BMP-1	亲环素A	MMP-12	0.853
42	ERBB1	TCTP	MMP-12	0.853
43	GAPDH, 肝	TCTP	MMP-12	0.853
44	LCMN	GAPDH, 肝	MMP-12	0.853
45	HSP 90 α	LRIG3	MMP-12	0.853
46	C9	激肽释放酶7	MMP-12	0.852
47	SCF sR	LRIG3	MMP-12	0.852
48	需钙蛋白酶I	C9	MMP-12	0.852
49	C9	过氧化氢酶	MMP-12	0.852
50	HMG-1	C9	MMP-12	0.852
51	C9	LRIG3	MMP-12	0.852
52	LRIG3	激肽释放酶7	MMP-12	0.852

[0564] 表23 (续)

[0565]

53	SCF sR	TCTP	MMP-12	0.851
54	SCF sR	亲环素A	MMP-12	0.851
55	BMP-1	UBE2N	MMP-12	0.851

56	激肽释放酶7	TCTP	MMP-12	0.851
57	MMP-7	UBE2N	MMP-12	0.851
58	MMP-7	RAC1	MMP-12	0.851
59	激肽释放酶7	亲环素A	MMP-12	0.850
60	MIP-5	GAPDH, 肝	MMP-12	0.850
61	CSK	CK-MB	MMP-12	0.850
62	MMP-7	亲环素A	MMP-12	0.850
63	CSK	LGMN	MMP-12	0.850
64	RGM-C	GAPDH, 肝	MMP-12	0.850
65	CDK5/p35	TCTP	MMP-12	0.850
66	PTN	GAPDH, 肝	MMP-12	0.850
67	脂联素	GAPDH, 肝	MMP-12	0.850
68	LRIG3	UBE2N	MMP-12	0.849
69	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.849
70	SCF sR	C9	MMP-12	0.849
71	CSK	过氧化氢酶	MMP-12	0.849
72	内皮他丁	GAPDH, 肝	MMP-12	0.849
73	SCF sR	RAC1	MMP-12	0.849
74	RAC1	b-ECGF	MMP-12	0.849
75	TPSB2	GAPDH, 肝	MMP-12	0.849
76	C9	UBE2N	MMP-12	0.849
77	b-ECGF	TCTP	MMP-12	0.849
78	C9	IMB1	MMP-12	0.849
79	需钙蛋白酶I	GAPDH, 肝	MMP-12	0.848
80	CSK	IL-8	MMP-12	0.848
81	CSK	脂联素	MMP-12	0.848
82	激肽释放酶7	IMB1	MMP-12	0.848
83	需钙蛋白酶I	CSK	MMP-12	0.848
84	巨噬细胞甘露糖受体	GAPDH, 肝	MMP-12	0.848
85	SCF sR	CDK5/p35	MMP-12	0.848
86	IGFBP-2	亲环素A	MMP-12	0.848
87	CDK5/p35	BTK	MMP-12	0.848
88	巨噬细胞甘露糖受体	CSK	MMP-12	0.847
89	钙粘着蛋白E	IGFBP-2	MMP-12	0.847
90	凝血酶敏感蛋白-1	TCTP	MMP-12	0.847
91	ERBB1	C9	MMP-12	0.847
92	RAC1	RGM-C	MMP-12	0.847
93	ERBB1	亲环素A	MMP-12	0.847
94	CXCL16, 可溶	GAPDH, 肝	MMP-12	0.847
95	RGM-C	亲环素A	MMP-12	0.847
96	LRIG3	亲环素A	MMP-12	0.847

97	ERBB1	RAC1	MMP-12	0.847
98	激肽释放酶7	UBE2N	MMP-12	0.847
99	MMP-7	LRIG3	MMP-12	0.847
100	BMP-1	BTK	MMP-12	0.847

[0566] 表24 100组包括MMP-12的四生物标记

	标记				AUC
1	MMP-7	C9	GAPDH, 肝	MMP-12	0.892
2	C9	LRIG3	GAPDH, 肝	MMP-12	0.890
3	MMP-7	LRIG3	GAPDH, 肝	MMP-12	0.889
4	SCF α R	C9	GAPDH, 肝	MMP-12	0.889
5	IGFBP-2	C9	GAPDH, 肝	MMP-12	0.889
6	C9	LGMN	GAPDH, 肝	MMP-12	0.889
7	IGFBP-2	MMP-7	GAPDH, 肝	MMP-12	0.889
8	MMP-7	BMP-1	GAPDH, 肝	MMP-12	0.889
9	C9	激肽释放酶7	GAPDH, 肝	MMP-12	0.889
10	C9	BMP-1	GAPDH, 肝	MMP-12	0.888
11	ERBB1	C9	GAPDH, 肝	MMP-12	0.887
12	MMP-7	CDK5/p35	GAPDH, 肝	MMP-12	0.886
13	C9	b-ECGF	GAPDH, 肝	MMP-12	0.886
14	钙粘着蛋白E	MMP-7	GAPDH, 肝	MMP-12	0.885
15	MMP-7	TPSB2	GAPDH, 肝	MMP-12	0.885
16	巨噬细胞甘露糖受体	C9	GAPDH, 肝	MMP-12	0.885
17	MMP-7	IGFBP-7	GAPDH, 肝	MMP-12	0.885
18	MMP-7	CSK	GAPDH, 肝	MMP-12	0.884
19	MMP-7	b-ECGF	GAPDH, 肝	MMP-12	0.884
20	C9	脂联素	GAPDH, 肝	MMP-12	0.884
21	C9	TPSB2	GAPDH, 肝	MMP-12	0.884
22	HMG-1	MMP-7	CSK	MMP-12	0.884
23	SCF α R	MMP-7	GAPDH, 肝	MMP-12	0.884
24	MMP-7	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.883
25	CXCL16, 可溶	C9	GAPDH, 肝	MMP-12	0.883
26	MMP-7	激肽释放酶7	GAPDH, 肝	MMP-12	0.883
27	C9	MIP-3	GAPDH, 肝	MMP-12	0.883
28	SCF α R	BMP-1	GAPDH, 肝	MMP-12	0.883
29	C9	CSK	LGMN	MMP-12	0.883
30	C9	GAPDH, 肝	LDH-H1	MMP-12	0.882
31	C9	RGM-C	GAPDH, 肝	MMP-12	0.882
32	巨噬细胞甘露糖受体	MMP-7	GAPDH, 肝	MMP-12	0.882
33	内皮他丁	C9	GAPDH, 肝	MMP-12	0.882
34	MMP-7	RGM-C	GAPDH, 肝	MMP-12	0.882
35	LRIG3	激肽释放酶7	GAPDH, 肝	MMP-12	0.882
36	C9	钙粘着蛋白-6	GAPDH, 肝	MMP-12	0.882
37	MMP-7	LRIG3	CSK	MMP-12	0.882
38	MMP-7	GAPDH, 肝	ICYP	MMP-12	0.882
39	C9	CSK	激肽释放酶7	MMP-12	0.881
40	MMP-7	GAPDH, 肝	LDH-H1	MMP-12	0.881
41	MMP-7	LGMN	GAPDH, 肝	MMP-12	0.881
42	ERBB1	MMP-7	GAPDH, 肝	MMP-12	0.881
43	HMG-1	MMP-7	GAPDH, 肝	MMP-12	0.881
44	IGFBP-2	激肽释放酶7	GAPDH, 肝	MMP-12	0.881
45	C9	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.881
46	C9	CDK5/p35	GAPDH, 肝	MMP-12	0.881
47	ERBB1	C9	CSK	MMP-12	0.881
48	MMP-7	BMP-1	CSK	MMP-12	0.881
49	钙粘着蛋白E	MMP-7	CSK	MMP-12	0.881
50	SCF α R	CDK5/p35	GAPDH, 肝	MMP-12	0.881

[0567]

[0568] 表24 (续)

[0569]

51	C9	RGM C	CSK	MMP 12	0.881
52	C9	GAPDH,肝	NACA	MMP 12	0.880
53	C9	LRIG3	CSK	MMP 12	0.880
54	MMP 7	脂联素	GAPDH,肝	MMP 12	0.880
55	C9	CSK	GAPDH,肝	MMP 12	0.880
56	LRIG3	BMP 1	GAPDH,肝	MMP 12	0.880
57	MMP 7	PTN	GAPDH,肝	MMP 12	0.880
58	C9	CSK	凝血酶敏感蛋白 1	MMP 12	0.880
59	活化素A	C9	GAPDH,肝	MMP 12	0.880
60	内皮他丁	C9	CSK	MMP 12	0.880
61	IGFBP 2	BMP 1	GAPDH,肝	MMP 12	0.880
62	IGFBP 2	LRIG3	GAPDH,肝	MMP 12	0.880
63	SCF sR	MMP 7	CSK	MMP 12	0.880
64	钙粘着蛋白E	C9	GAPDH,肝	MMP 12	0.880
65	SCF sR	LRIG3	GAPDH,肝	MMP 12	0.880
66	需钙蛋白酶I	C9	GAPDH,肝	MMP 12	0.879
67	RAC1	C9	激肽释放酶7	MMP 12	0.879
68	C9	LRIG3	TCTP	MMP 12	0.879
69	C9	CDK5/p35	CSK	MMP 12	0.879
70	C9	CSK	LDH H1	MMP 12	0.879
71	ERBB1	MMP 7	CSK	MMP 12	0.879
72	活化素A	MMP 7	GAPDH,肝	MMP 12	0.879
73	IGFBP 2	凝血酶敏感蛋白 1	GAPDH,肝	MMP 12	0.879
74	C9	蛋白水解酶 3	GAPDH,肝	MMP 12	0.878
75	vWF	C9	GAPDH,肝	MMP 12	0.878
76	MMP 7	CNDP1	GAPDH,肝	MMP 12	0.878
77	C9	BMP 1	CSK	MMP 12	0.878
78	C9	CK MB	GAPDH,肝	MMP 12	0.878
79	IGFBP 2	MMP 7	CSK	MMP 12	0.878
80	MMP 7	GAPDH,肝	纤连蛋白	MMP 12	0.878
81	MMP 7	CD30配体	GAPDH,肝	MMP 12	0.878
82	C9	CDK5/p35	TCTP	MMP 12	0.878
83	C9	CNDP1	GAPDH,肝	MMP 12	0.878
84	需钙蛋白酶I	C9	CSK	MMP 12	0.878
85	MMP 7	C9	CSK	MMP 12	0.877
86	MMP 7	CK MB	GAPDH,肝	MMP 12	0.877
87	需钙蛋白酶I	MMP 7	GAPDH,肝	MMP 12	0.877
88	SCF sR	C9	CSK	MMP 12	0.877
89	MMP 7	钙粘着蛋白 6	GAPDH,肝	MMP 12	0.877
90	MMP 7	过氧化氢酶	GAPDH,肝	MMP 12	0.877
91	MMP 7	CDK5/p35	CSK	MMP 12	0.877
92	MMP 7	RAC1	GAPDH,肝	MMP 12	0.877
93	SCF sR	激肽释放酶7	GAPDH,肝	MMP 12	0.877
94	C9	过氧化氢酶	GAPDH,肝	MMP 12	0.877
95	C9	FGF 17	GAPDH,肝	MMP 12	0.877

96	HMG 1	MMP 7	TCTP	MMP 12	0.877
97	ERBB1	C9	TCTP	MMP 12	0.877
98	MMP 7	GAPDH, 肝	NACA	MMP 12	0.877
99	ERBB1	BMP 1	GAPDH, 肝	MMP 12	0.877
100	HSP 90a	C9	LRIG3	MMP 12	0.877

[0570] 表25 100组包括MMP-12的五生物标记

标记					AUC	
1	C9	LRIG3	激肽释放酶7	GAPDH, 肝	MMP-12	0.900
2	MMP-7	C9	LRIG3	GAPDH, 肝	MMP-12	0.900
3	SCF sR	C9	LRIG3	GAPDH, 肝	MMP-12	0.900
4	SCF sR	C9	BMP-1	GAPDH, 肝	MMP-12	0.898
5	IGFBP-2	MMP-7	LRIG3	GAPDH, 肝	MMP-12	0.898
6	IGFBP-2	C9	激肽释放酶7	GAPDH, 肝	MMP-12	0.897
7	MMP-7	C9	BMP-1	GAPDH, 肝	MMP-12	0.897
8	MMP-7	C9	TPSB2	GAPDH, 肝	MMP-12	0.897
9	IGFBP-2	MMP-7	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.897
10	MMP-7	C9	RGMC	GAPDH, 肝	MMP-12	0.897
11	HMG-1	MMP-7	C9	GAPDH, 肝	MMP-12	0.897
12	巨噬细胞甘露糖受体	C9	LRIG3	GAPDH, 肝	MMP-12	0.897
13	C9	LRIG3	LGMN	GAPDH, 肝	MMP-12	0.897
14	C9	激肽释放酶7	LGMN	GAPDH, 肝	MMP-12	0.897
15	钙粘着蛋白E	MMP-7	BMP-1	GAPDH, 肝	MMP-12	0.897
16	巨噬细胞甘露糖受体	C9	激肽释放酶7	GAPDH, 肝	MMP-12	0.897
17	MMP-7	C9	激肽释放酶7	GAPDH, 肝	MMP-12	0.896
18	钙粘着蛋白E	IGFBP-2	MMP-7	GAPDH, 肝	MMP-12	0.896
19	MMP-7	C9	b-ECGF	GAPDH, 肝	MMP-12	0.896
20	IGFBP-2	C9	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.896
21	SCF sR	C9	CDK5/p35	GAPDH, 肝	MMP-12	0.896
22	C9	BMP-1	LGMN	GAPDH, 肝	MMP-12	0.896
23	MMP-7	LRIG3	BMP-1	GAPDH, 肝	MMP-12	0.896
24	MMP-7	C9	CDK5/p35	GAPDH, 肝	MMP-12	0.896
25	SCF sR	MMP-7	CDK5/p35	GAPDH, 肝	MMP-12	0.896
26	SCF sR	MMP-7	BMP-1	GAPDH, 肝	MMP-12	0.896
27	IGFBP-2	C9	LRIG3	GAPDH, 肝	MMP-12	0.896
28	SCF sR	MMP-7	C9	GAPDH, 肝	MMP-12	0.895
29	巨噬细胞甘露糖受体	ERBB1	C9	GAPDH, 肝	MMP-12	0.895
30	IGFBP-2	MMP-7	激肽释放酶7	GAPDH, 肝	MMP-12	0.895
31	IGFBP-2	MMP-7	CDK5/p35	GAPDH, 肝	MMP-12	0.895
32	MMP-7	C9	GAPDH, 肝	LDH-H1	MMP-12	0.895
33	IGFBP-2	MMP-7	C9	GAPDH, 肝	MMP-12	0.895
34	巨噬细胞甘露糖受体	MMP-7	C9	GAPDH, 肝	MMP-12	0.895
35	MMP-7	IGFBP-7	BMP-1	GAPDH, 肝	MMP-12	0.895
36	C9	BMP-1	激肽释放酶7	GAPDH, 肝	MMP-12	0.895
37	SCF sR	MMP-7	LRIG3	GAPDH, 肝	MMP-12	0.895
38	SCF sR	IGFBP-2	MMP-7	GAPDH, 肝	MMP-12	0.895
39	C9	b-ECGF	LGMN	GAPDH, 肝	MMP-12	0.895
40	MMP-7	C9	LGMN	GAPDH, 肝	MMP-12	0.895
41	SCF sR	HMG-1	MMP-7	CSK	MMP-12	0.895
42	IGFBP-2	MMP-7	IGFBP-7	GAPDH, 肝	MMP-12	0.895
43	钙粘着蛋白E	MMP-7	CDK5/p35	GAPDH, 肝	MMP-12	0.895
44	MMP-7	C9	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.895
45	巨噬细胞甘露糖受体	MMP-7	LRIG3	GAPDH, 肝	MMP-12	0.895
46	MMP-7	IGFBP-7	LRIG3	GAPDH, 肝	MMP-12	0.895
47	IGFBP-2	MMP-7	BMP-1	GAPDH, 肝	MMP-12	0.895
48	MMP-7	BMP-1	CSK	GAPDH, 肝	MMP-12	0.895

[0571]

[0572] 表25 (续)

49	MMP-7	LRIG3	激肽释放酶7	GAPDH, 肝	MMP-12	0.895
50	IGFBP-2	MMP-7	TPSB2	GAPDH, 肝	MMP-12	0.895
51	MMP-7	BMP-1	GAPDH, 肝	LDH-H1	MMP-12	0.894
52	C9	激肽释放酶7	TPSB2	GAPDH, 肝	MMP-12	0.894
53	钙粘着蛋白E	MMP-7	s-ECGF	GAPDH, 肝	MMP-12	0.894
54	SCF sR	C9	激肽释放酶7	GAPDH, 肝	MMP-12	0.894
55	IGFBP-2	MMP-7	GAPDH, 肝	LDH-H1	MMP-12	0.894
56	SCF sR	C9	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.894
57	MMP-7	BMP-1	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.894
58	MMP-7	C9	CSK	GAPDH, 肝	MMP-12	0.894
59	内皮他丁	C9	LRIG3	GAPDH, 肝	MMP-12	0.894
60	C9	LRIG3	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.894
61	C9	LRIG3	BMP-1	GAPDH, 肝	MMP-12	0.894
62	IGFBP-2	MMP-7	s-ECGF	GAPDH, 肝	MMP-12	0.894
63	IGFBP-2	MMP-7	LGMN	GAPDH, 肝	MMP-12	0.894
64	MMP-7	C9	钙粘着蛋白-6	GAPDH, 肝	MMP-12	0.894
65	钙粘着蛋白E	MMP-7	C9	GAPDH, 肝	MMP-12	0.894
66	HMG-1	MMP-7	C9	CSK	MMP-12	0.894
67	SCF sR	C9	脂联素	GAPDH, 肝	MMP-12	0.894
68	CXCL16, 可溶	C9	LRIG3	GAPDH, 肝	MMP-12	0.894
69	IGFBP-2	MMP-7	CSK	GAPDH, 肝	MMP-12	0.894
70	SCF sR	巨噬细胞甘露糖受体	C9	GAPDH, 肝	MMP-12	0.894
71	MMP-7	C9	脂联素	GAPDH, 肝	MMP-12	0.894
72	钙粘着蛋白E	MMP-7	TPSB2	GAPDH, 肝	MMP-12	0.894
73	C9	LRIG3	钙粘着蛋白-6	GAPDH, 肝	MMP-12	0.894
74	MMP-7	LRIG3	CSK	GAPDH, 肝	MMP-12	0.894
75	MMP-7	LRIG3	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.894
76	ERBB1	MMP-7	C9	GAPDH, 肝	MMP-12	0.894
77	ERBB1	C9	BMP-1	GAPDH, 肝	MMP-12	0.894
78	ERBB1	C9	LGMN	GAPDH, 肝	MMP-12	0.894
79	C9	脂联素	LGMN	GAPDH, 肝	MMP-12	0.894
80	MMP-7	BMP-1	CDK5p35	GAPDH, 肝	MMP-12	0.893
81	MMP-7	BMP-1	LGMN	GAPDH, 肝	MMP-12	0.893
82	C9	LRIG3	GAPDH, 肝	NACA	MMP-12	0.893
83	C9	激肽释放酶7	GAPDH, 肝	LDH-H1	MMP-12	0.893
84	C9	激肽释放酶7	脂联素	GAPDH, 肝	MMP-12	0.893
85	IGFBP-2	MMP-7	钙粘着蛋白-6	GAPDH, 肝	MMP-12	0.893
86	IGFBP-2	C9	LGMN	GAPDH, 肝	MMP-12	0.893
87	MMP-7	BMP-1	TPSB2	GAPDH, 肝	MMP-12	0.893
88	C9	LRIG3	脂联素	GAPDH, 肝	MMP-12	0.893
89	C9	TPSB2	LGMN	GAPDH, 肝	MMP-12	0.893
90	MMP-7	C9	CD30 配体	GAPDH, 肝	MMP-12	0.893
91	钙粘着蛋白E	MMP-7	LRIG3	GAPDH, 肝	MMP-12	0.893
92	SCF sR	IGFBP-2	C9	GAPDH, 肝	MMP-12	0.893
93	HMG-1	IGFBP-2	MMP-7	GAPDH, 肝	MMP-12	0.893
94	SCF sR	钙粘着蛋白E	MMP-7	GAPDH, 肝	MMP-12	0.893
95	SCF sR	MMP-7	凝血酶敏感蛋白-1	GAPDH, 肝	MMP-12	0.893
96	IGFBP-2	C9	BMP-1	GAPDH, 肝	MMP-12	0.893
97	MMP-7	LRIG3	GAPDH, 肝	TCTP	MMP-12	0.893
98	C9	LRIG3	GAPDH, 肝	LDH-H1	MMP-12	0.893
99	SCF sR	C9	TPSB2	GAPDH, 肝	MMP-12	0.893
100	IGFBP-2	巨噬细胞甘露糖受体	MMP-7	GAPDH, 肝	MMP-12	0.893

[0573]

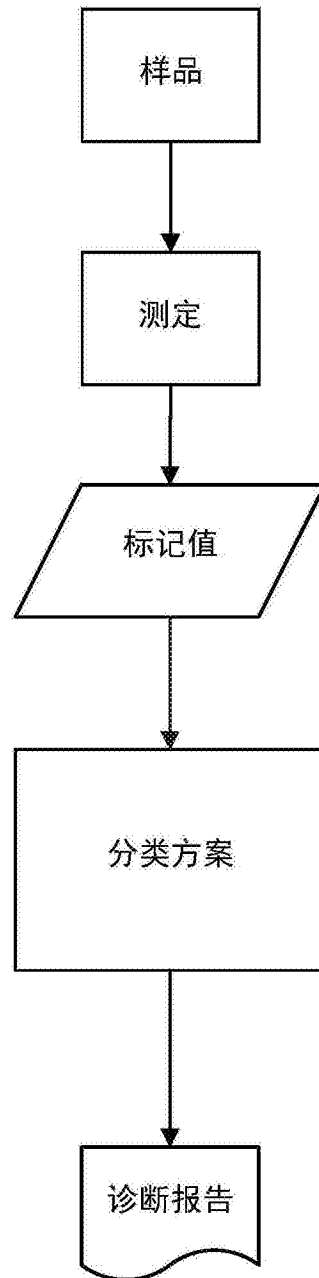


图1A

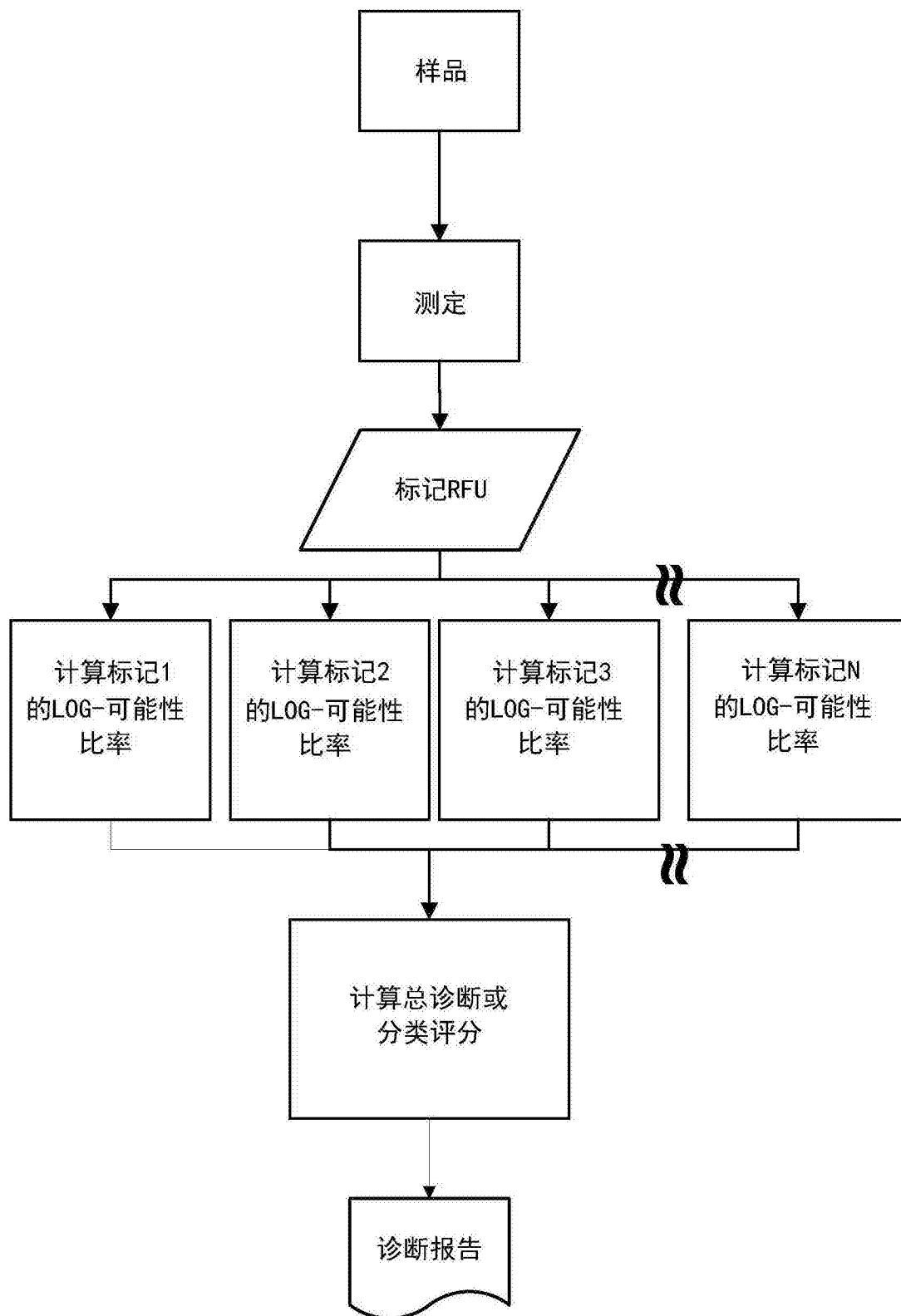


图1B

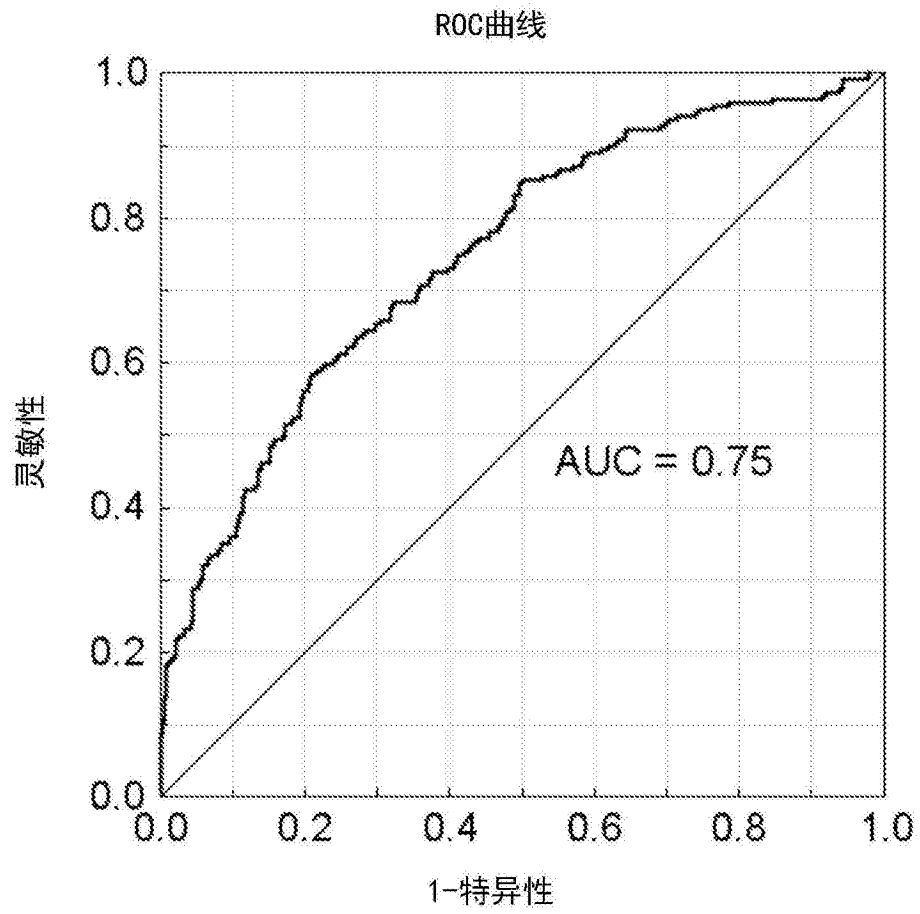


图2

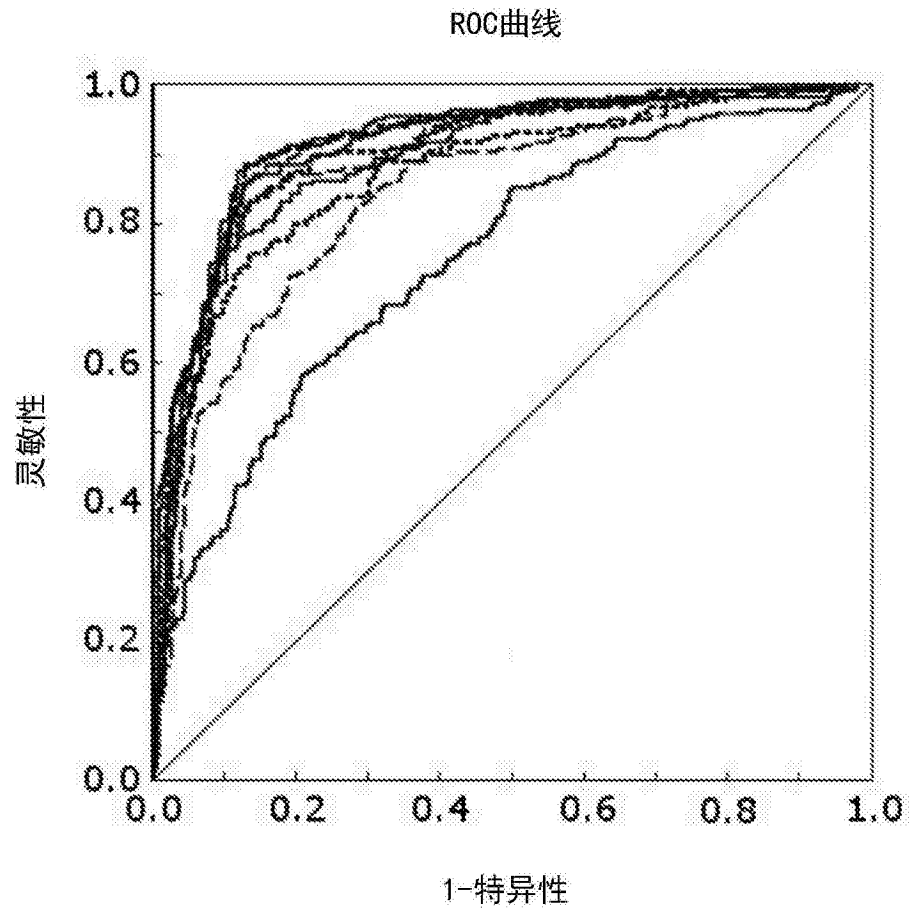


图3

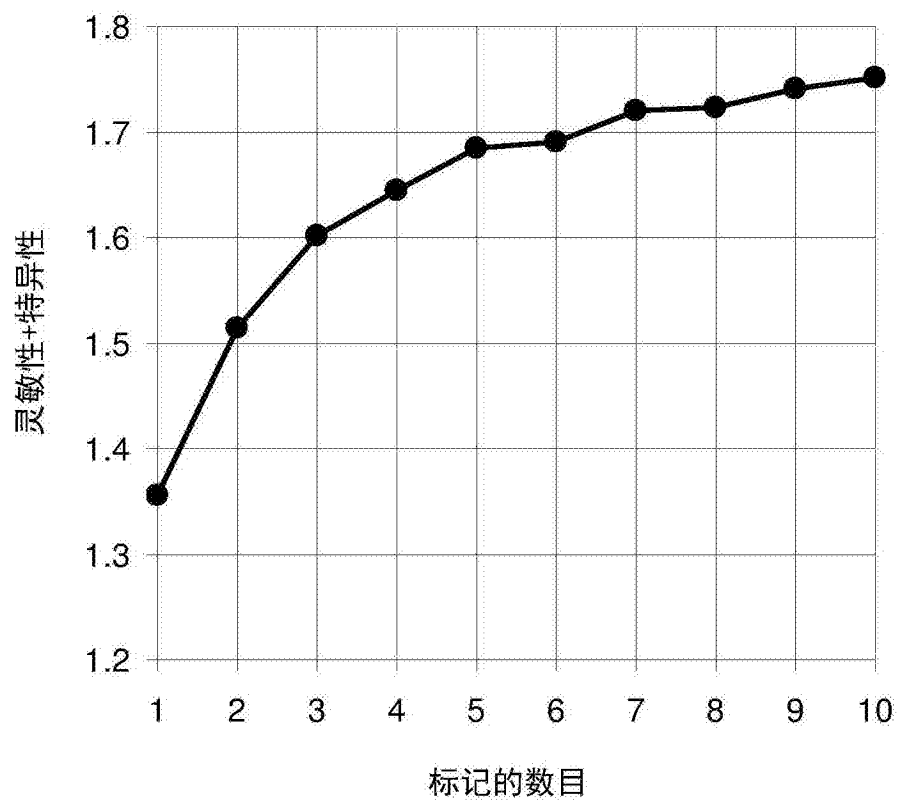


图4

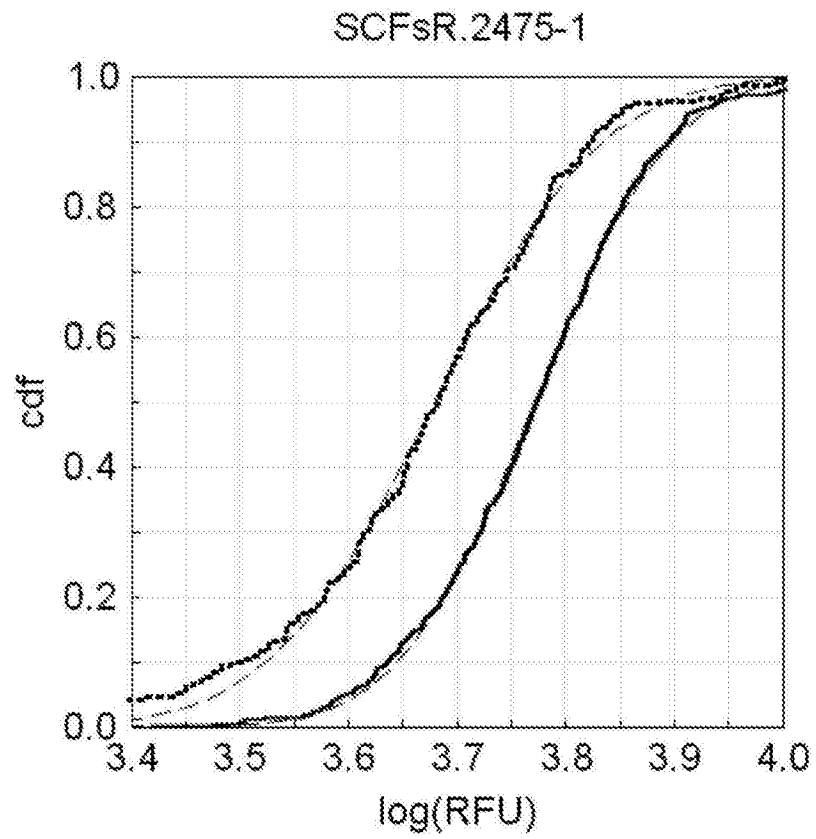


图5

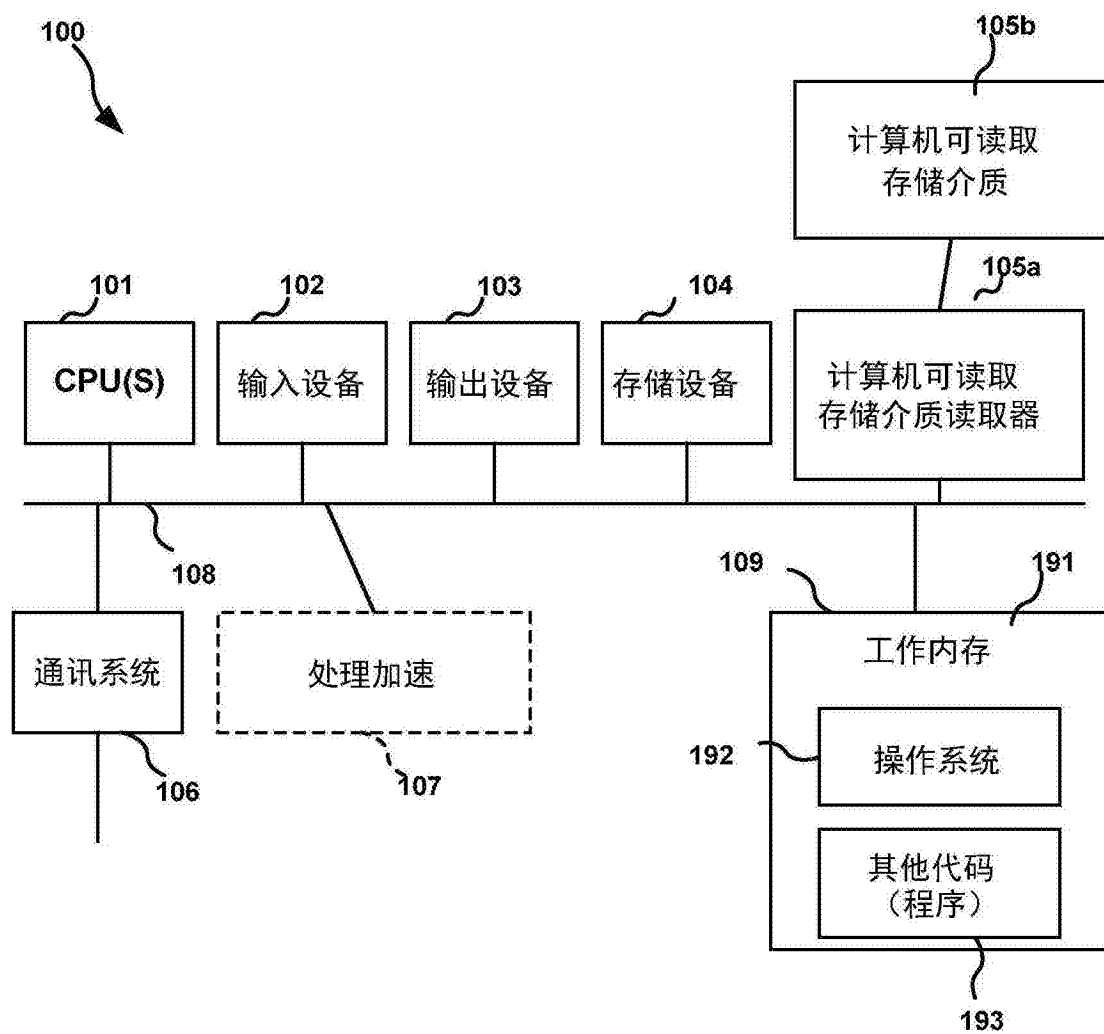


图6

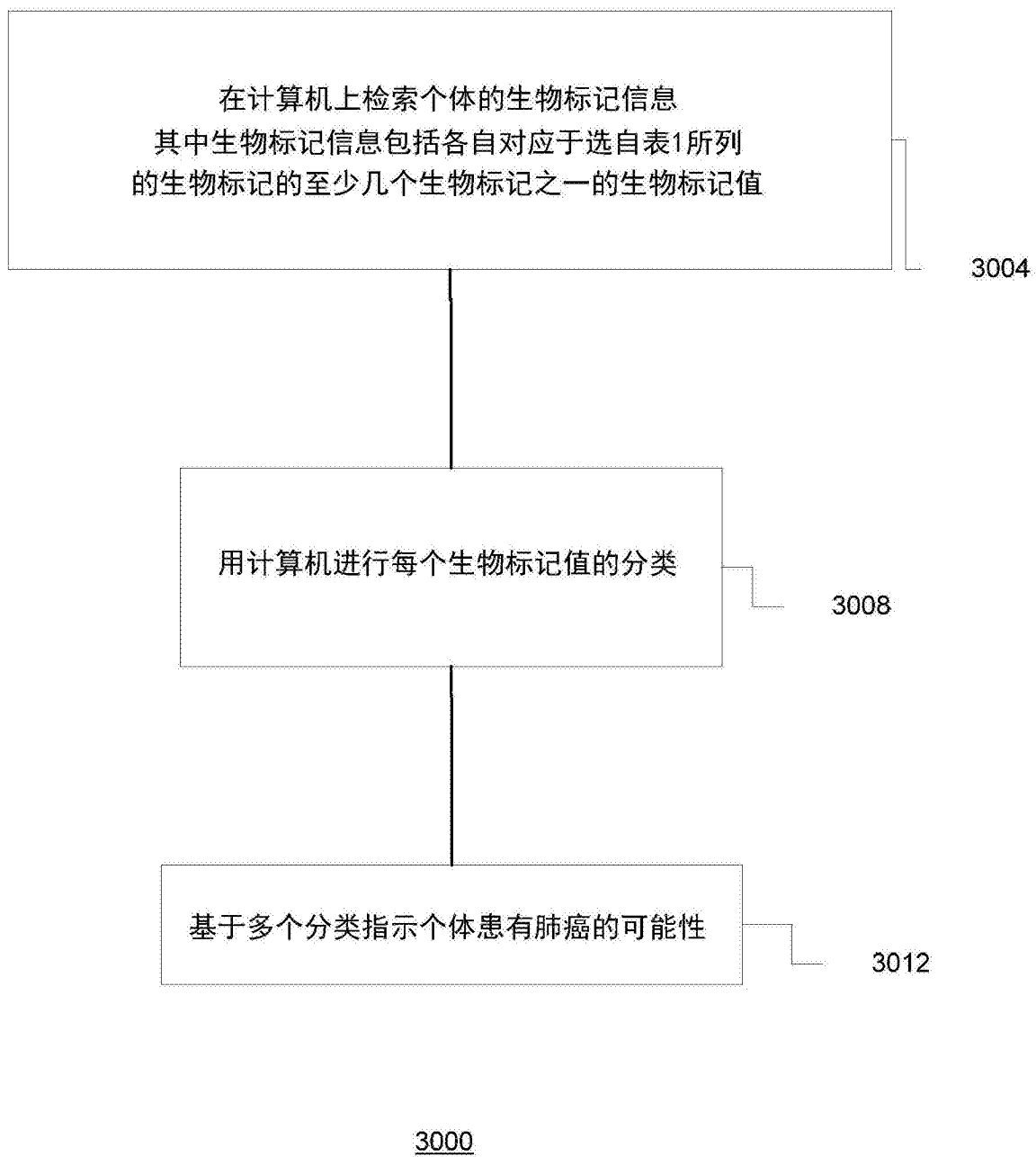


图7

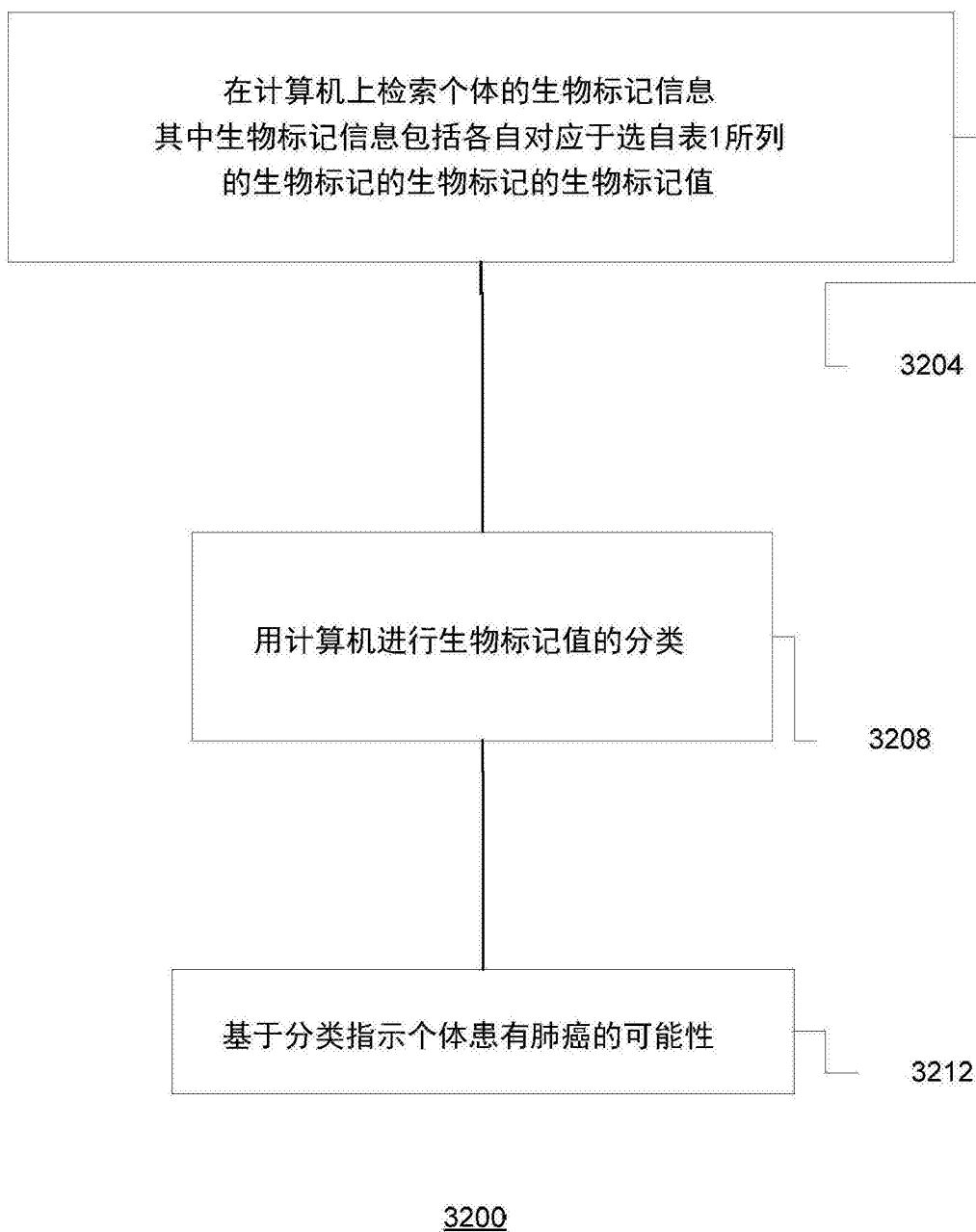


图8

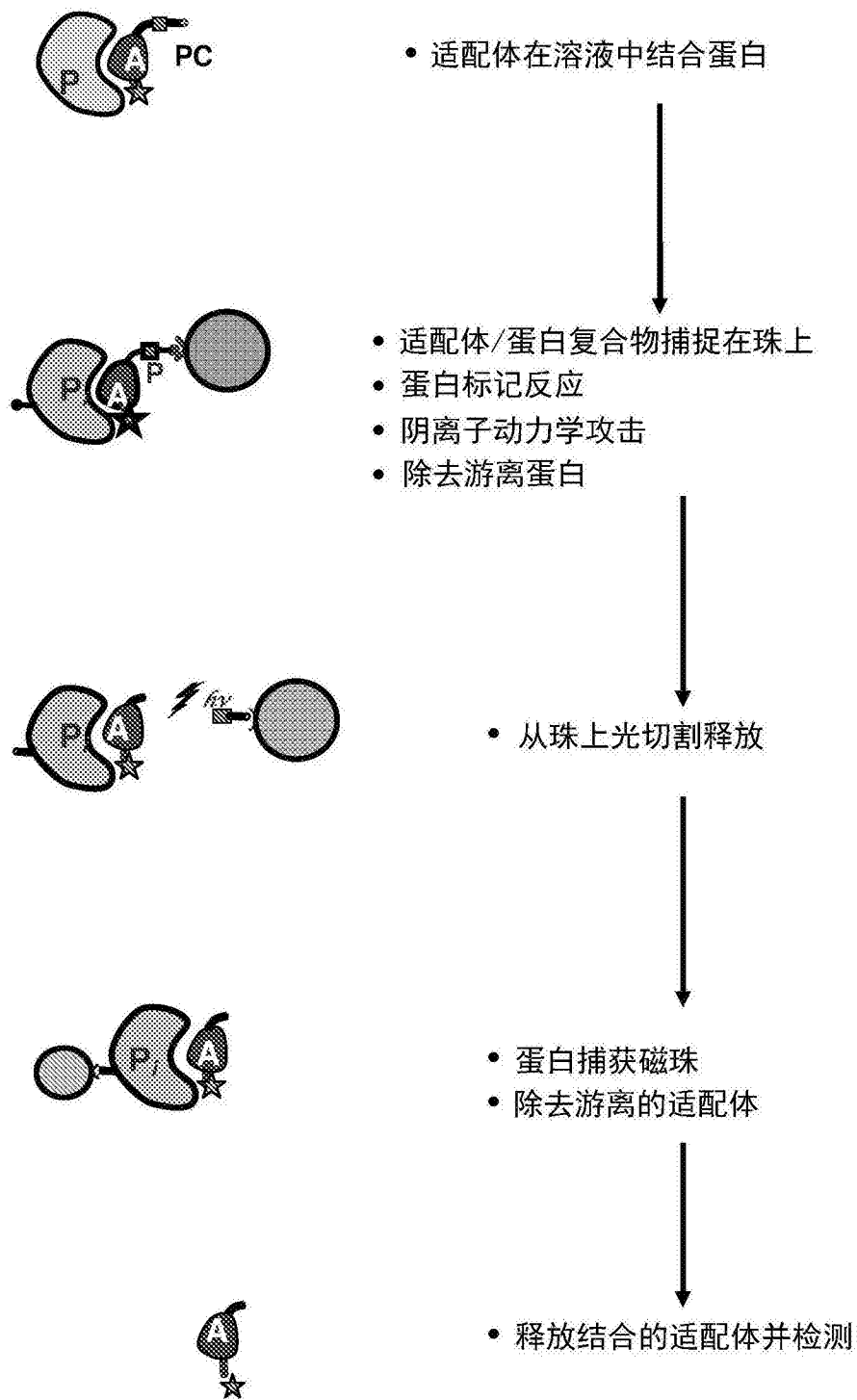


图9

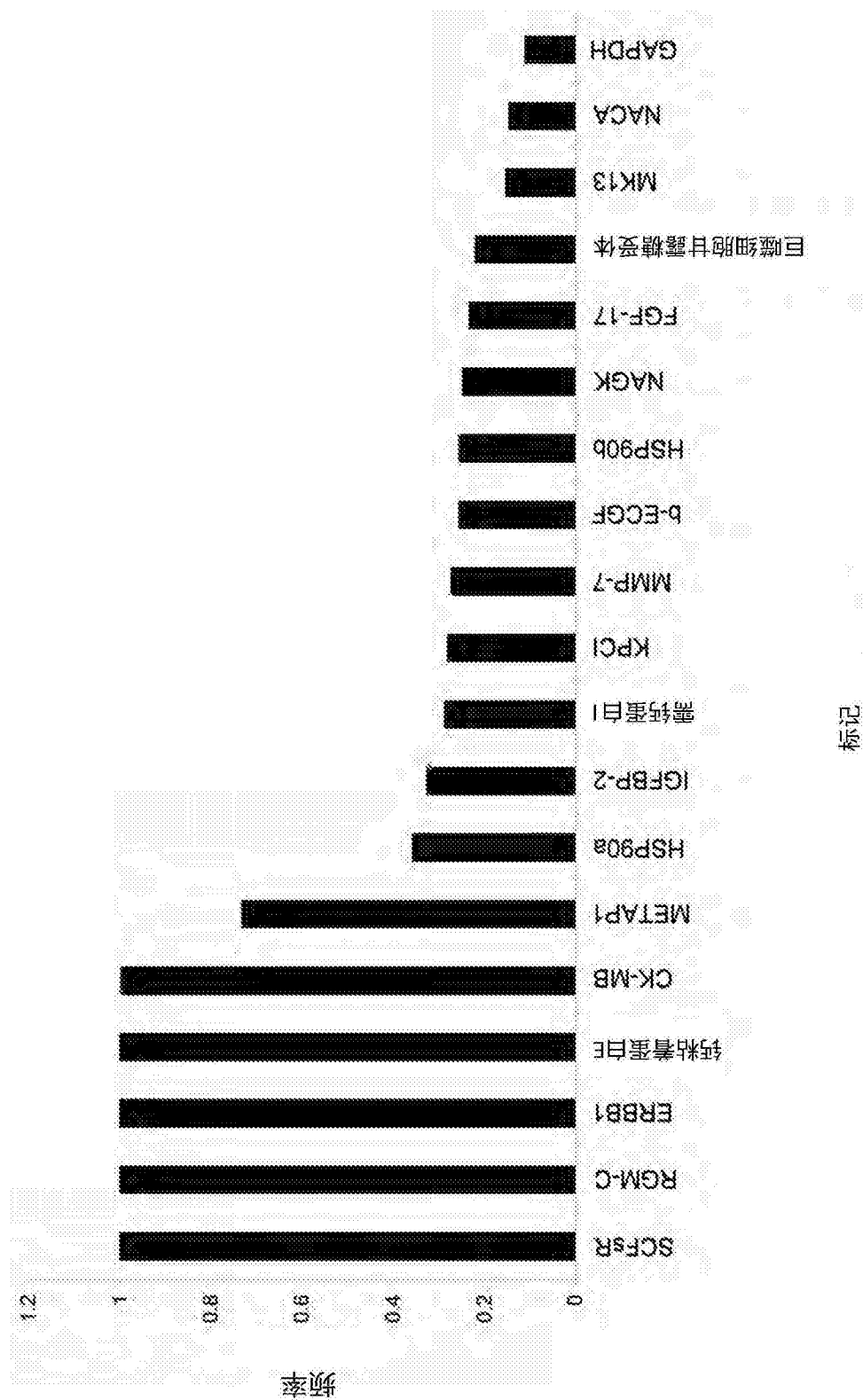


图10

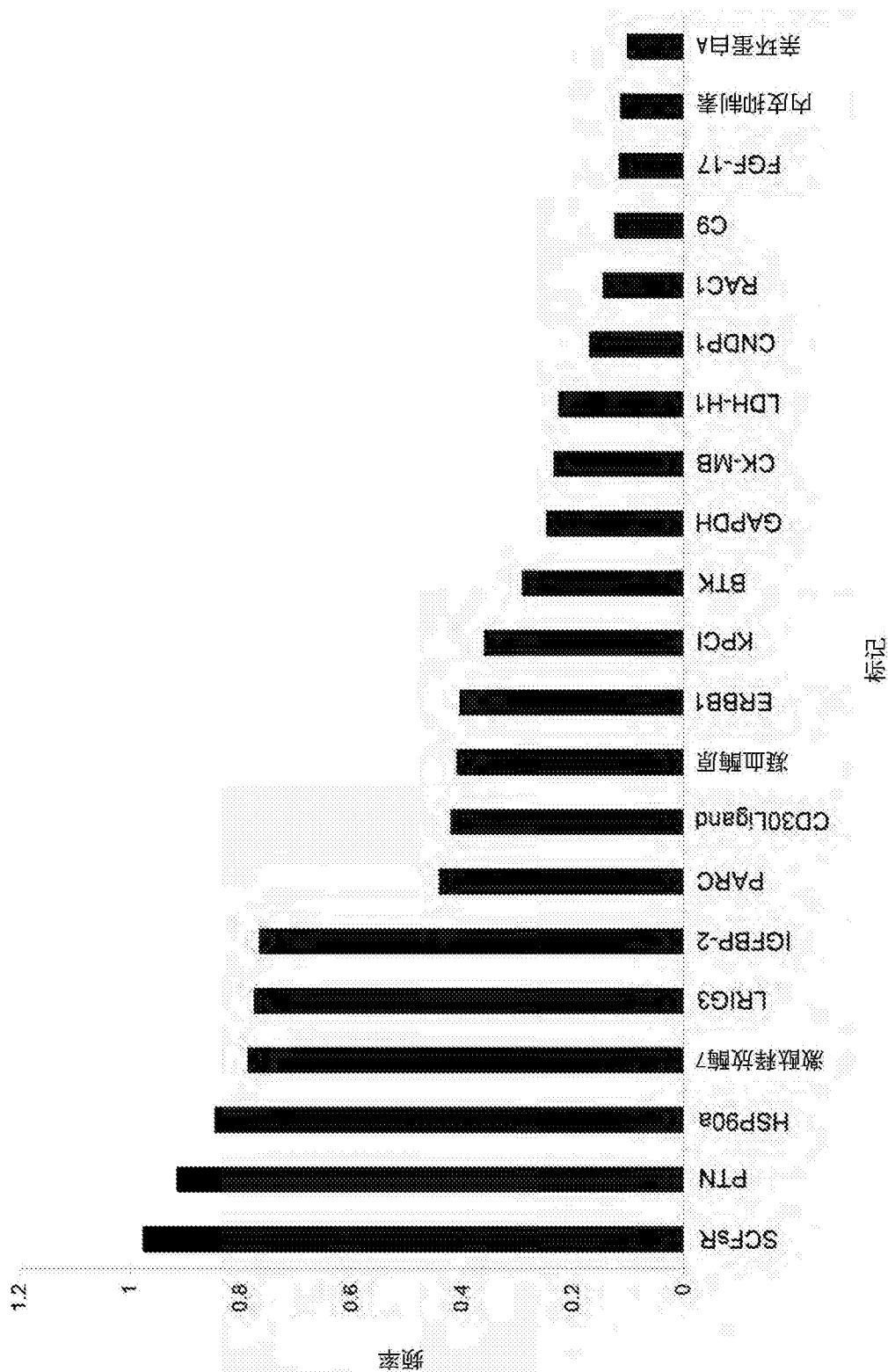


图11

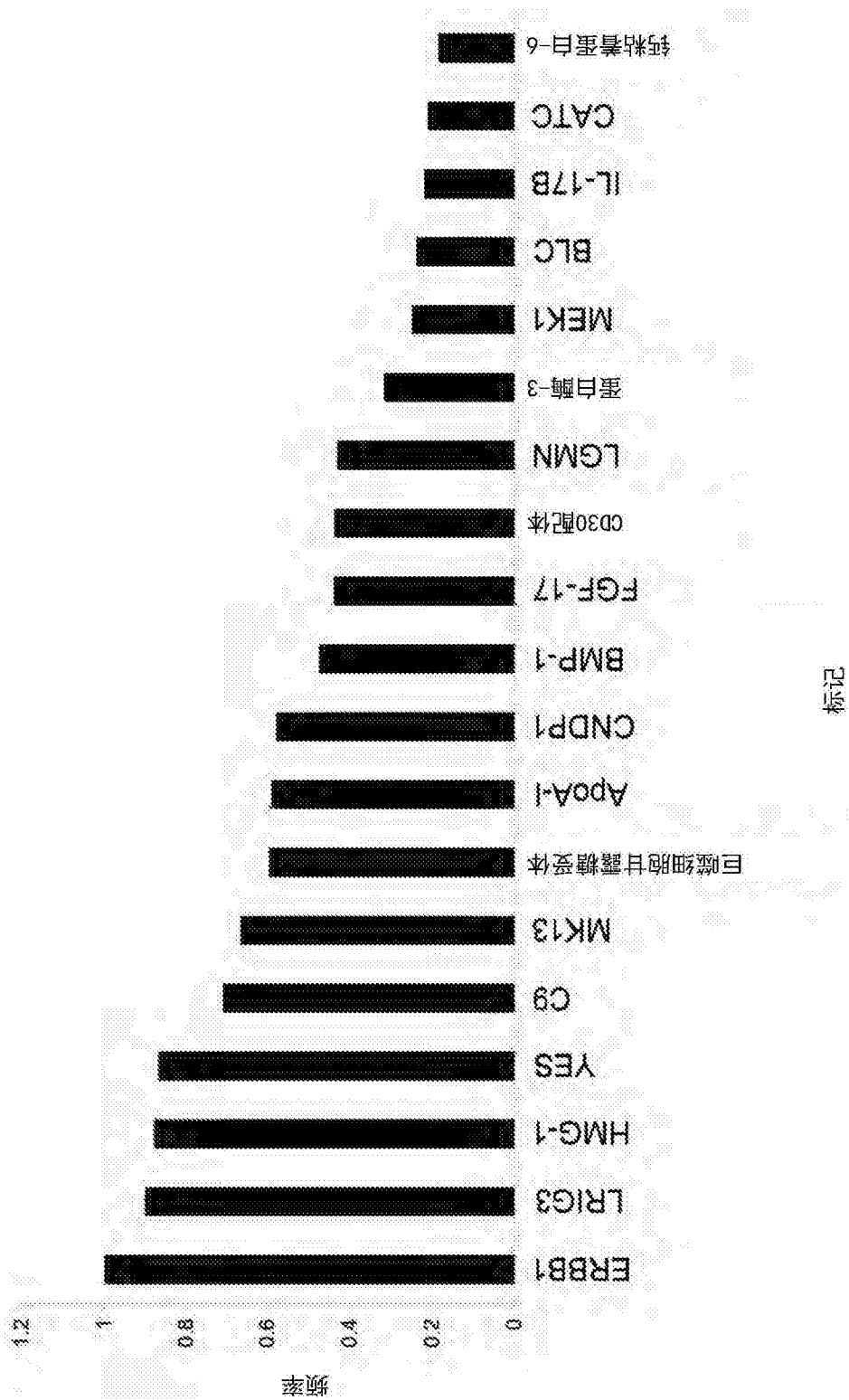


图12

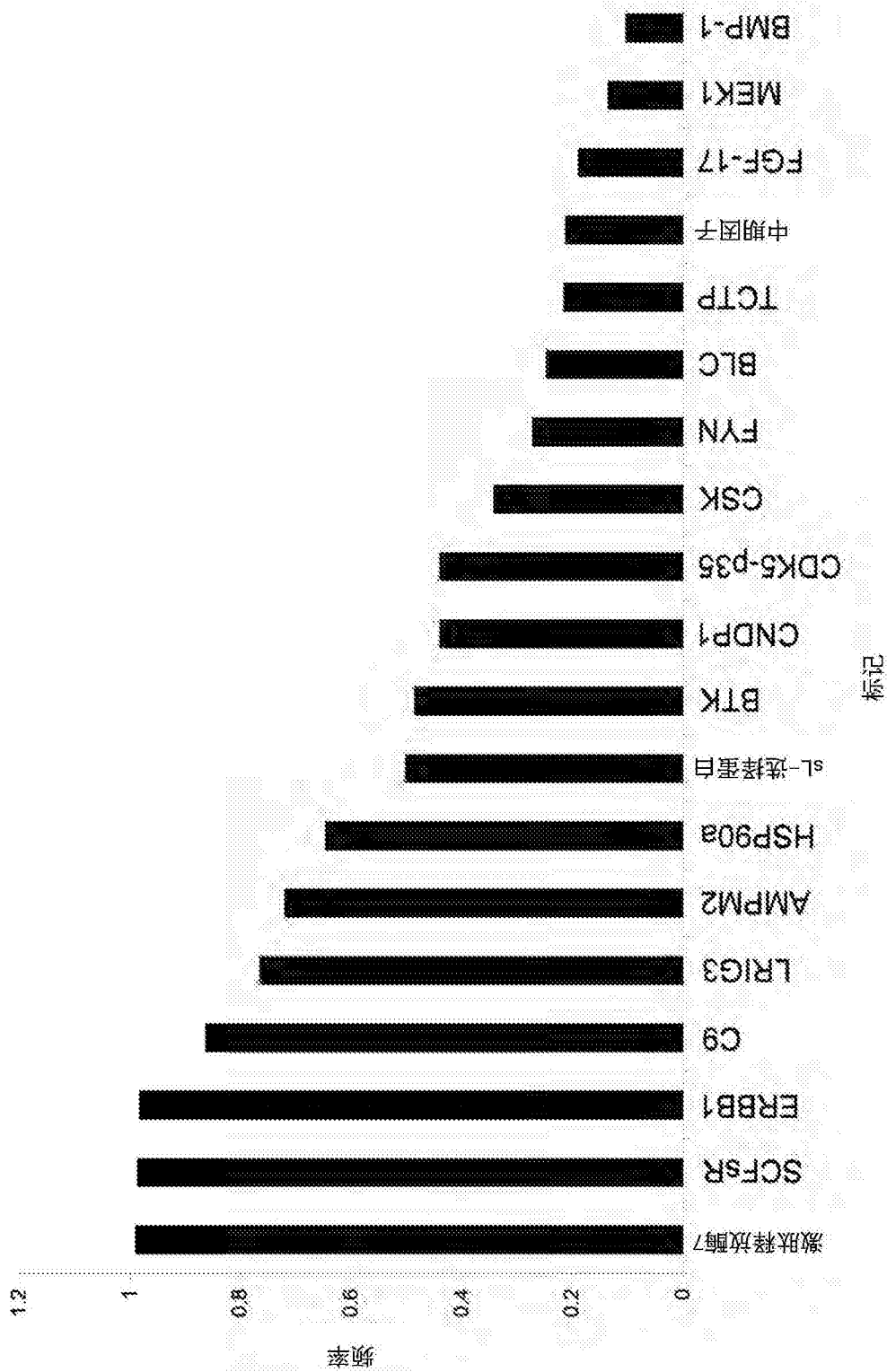


图13

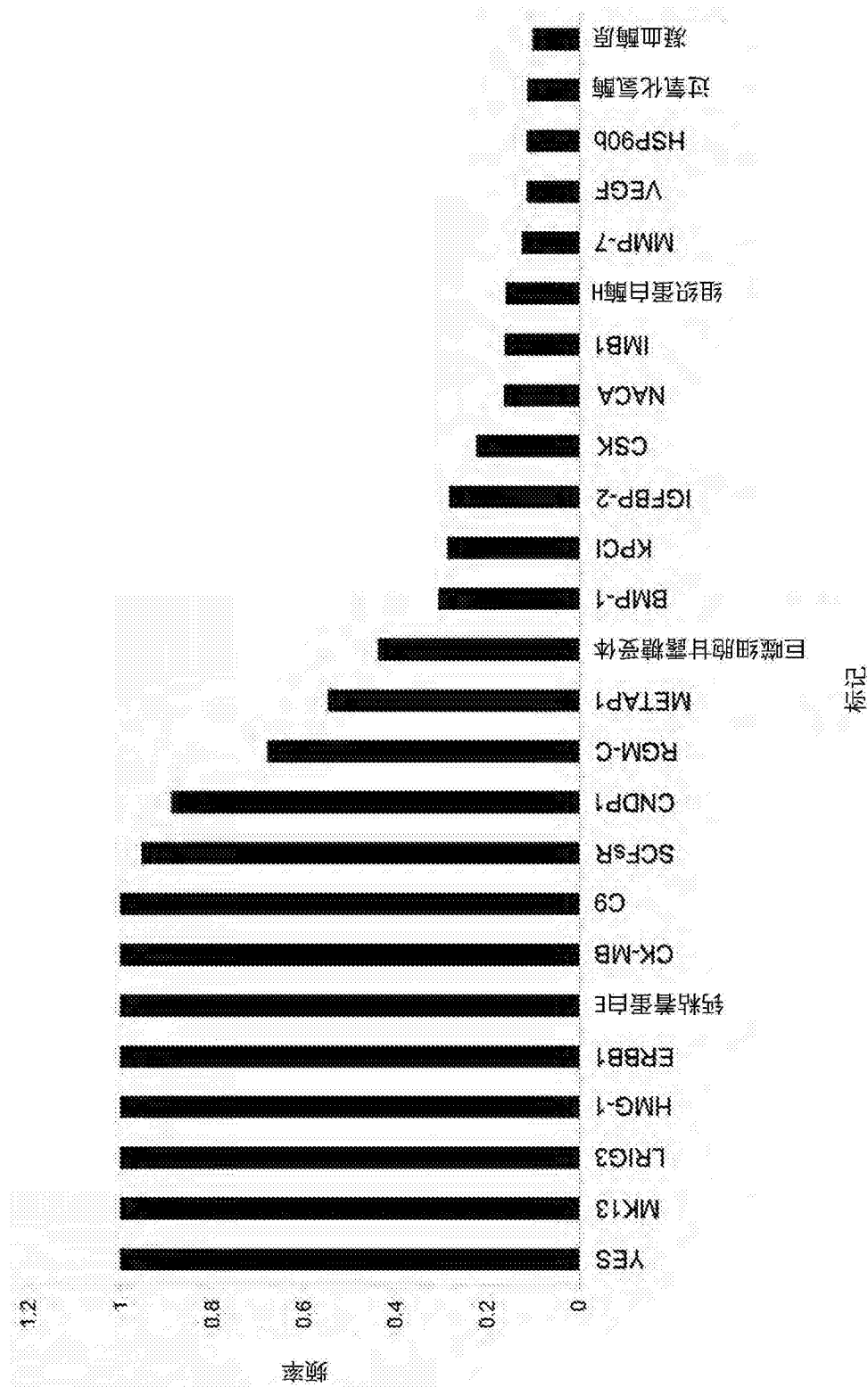


图14

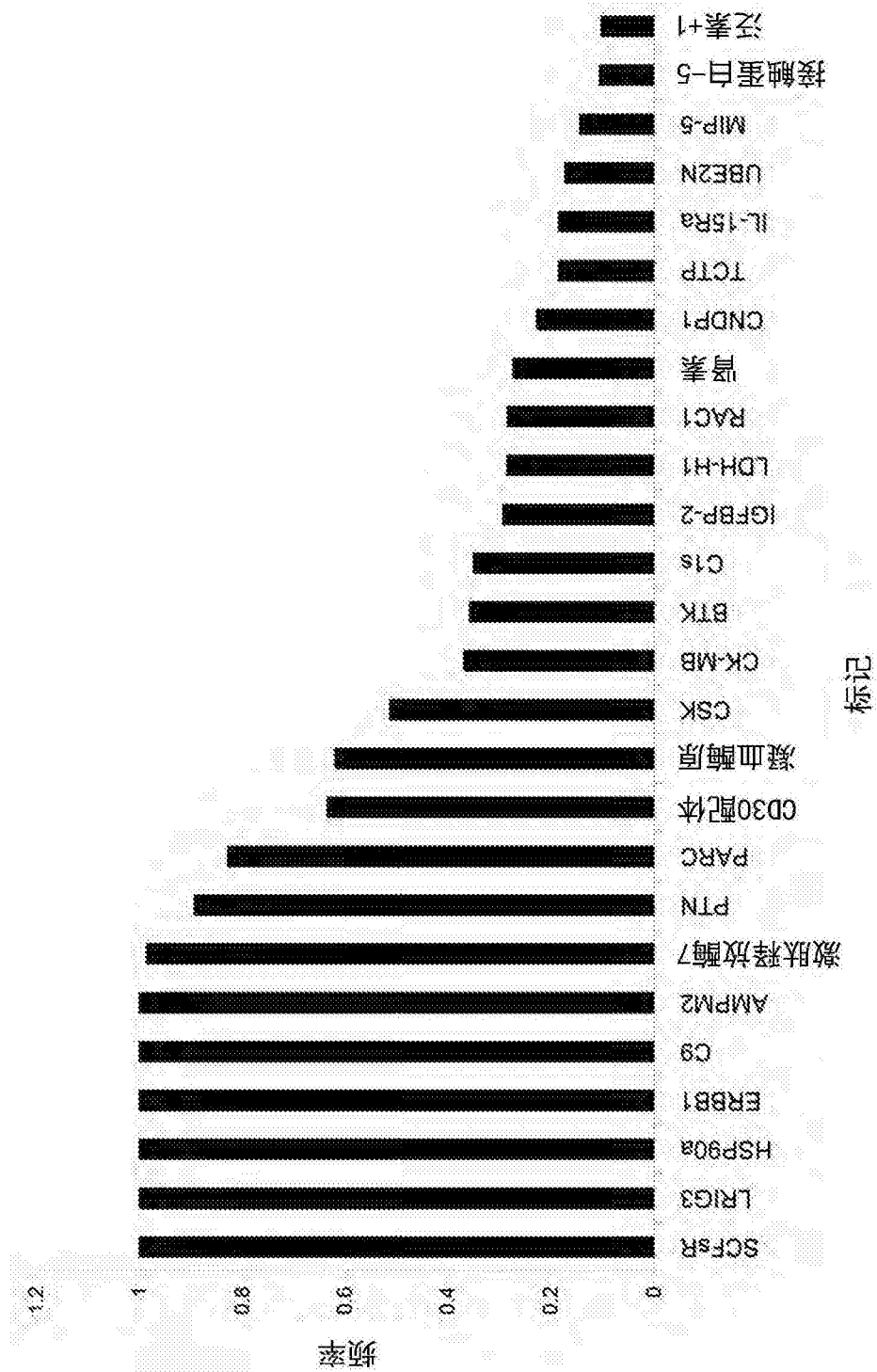


图15

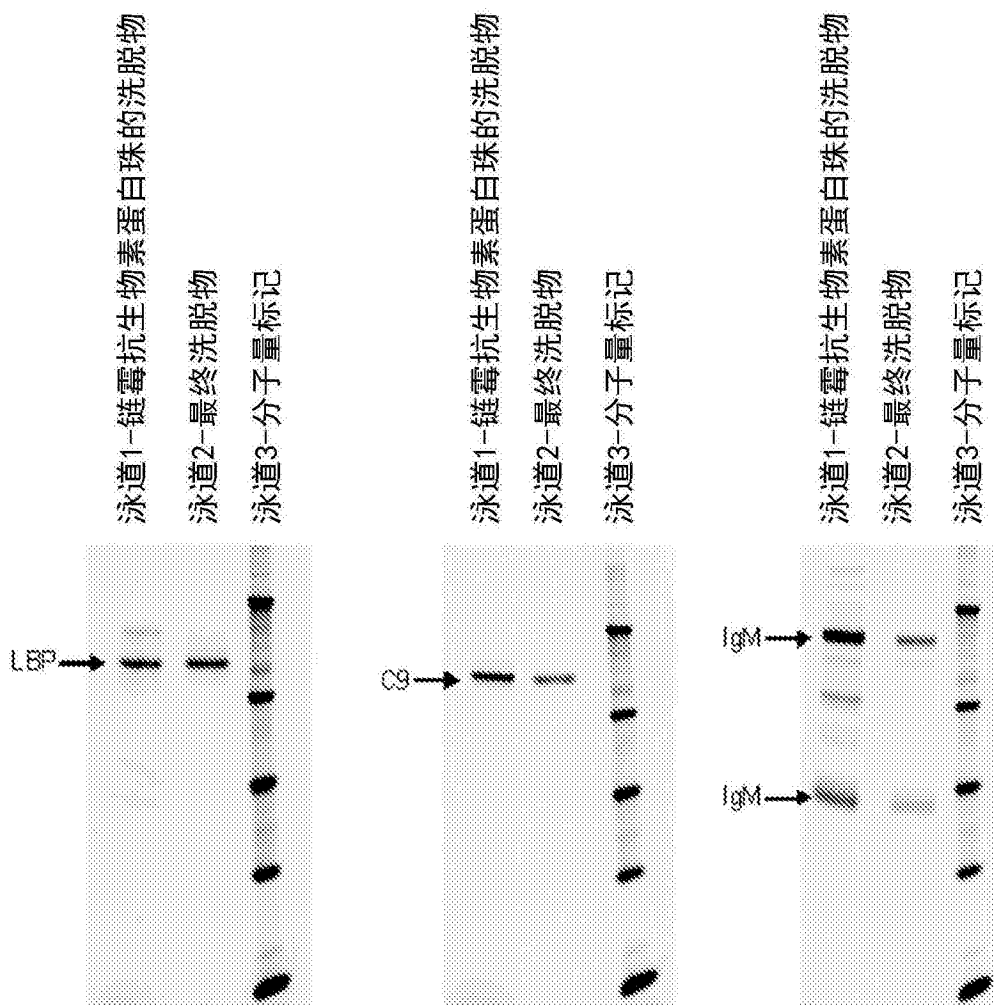


图16

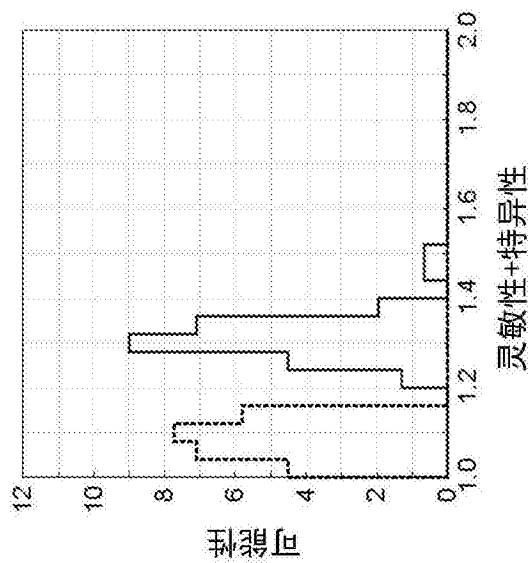


图17A

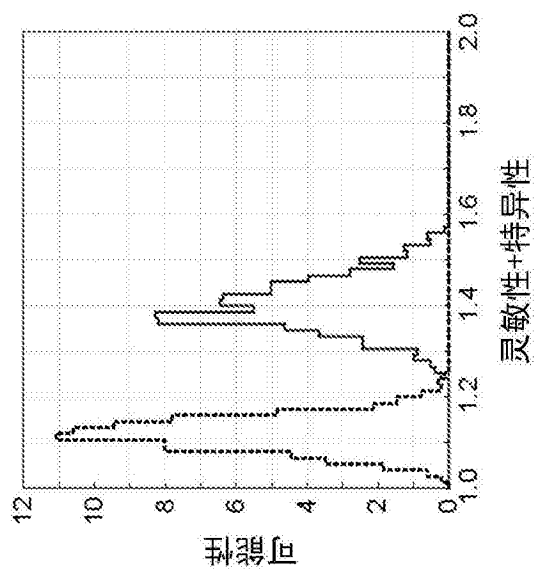


图17B

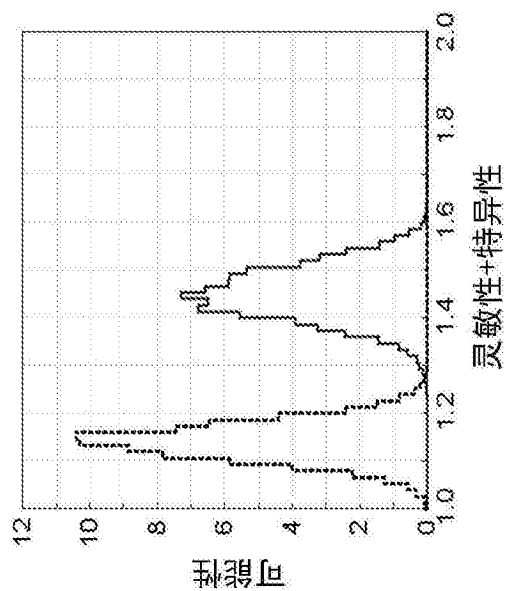


图17C

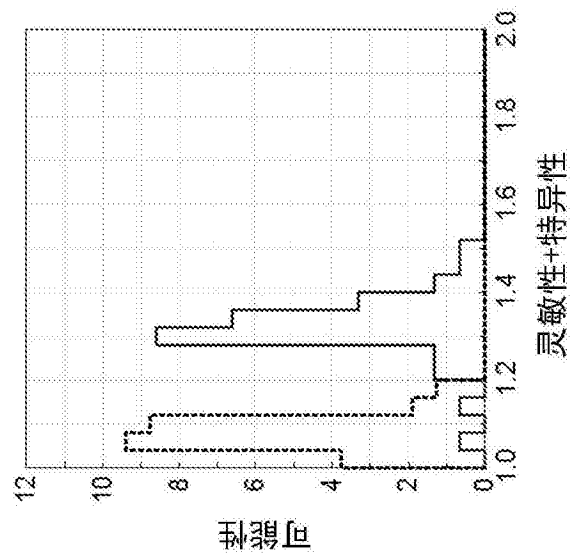


图18A

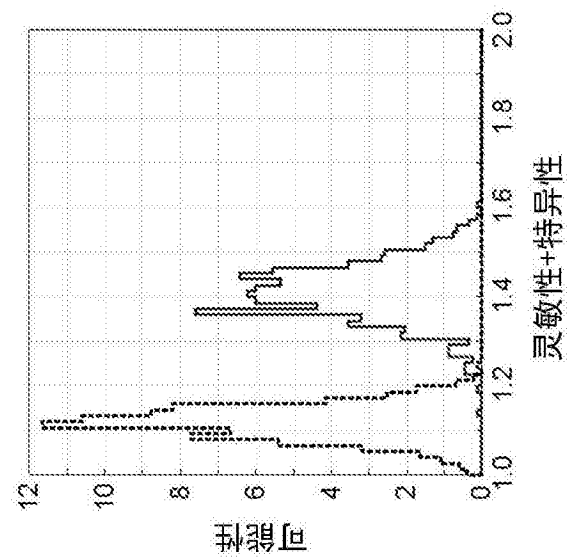


图18B

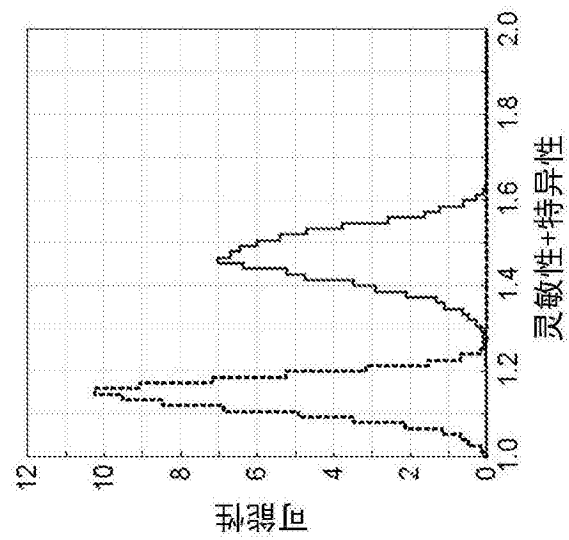


图18C

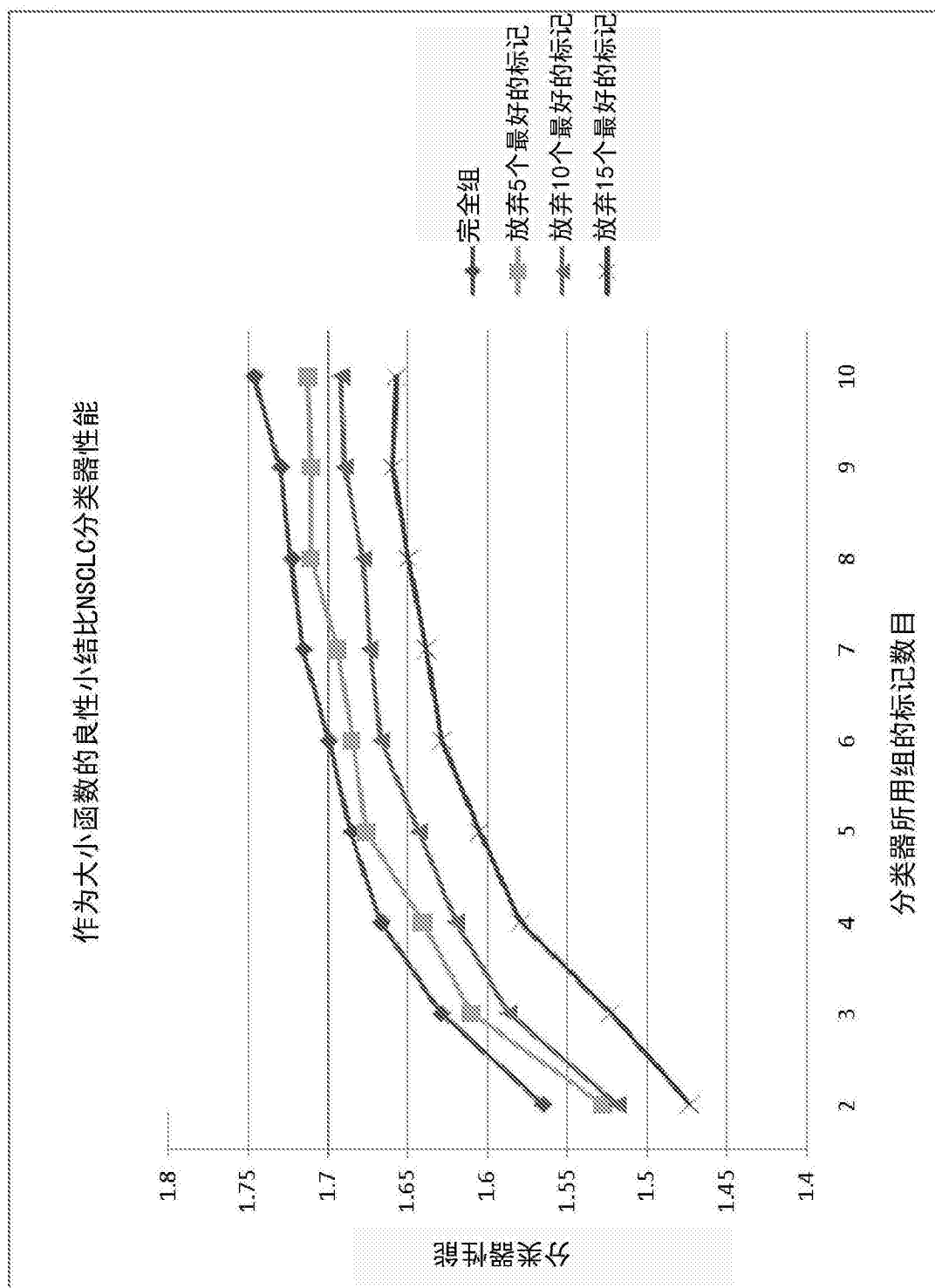


图19A

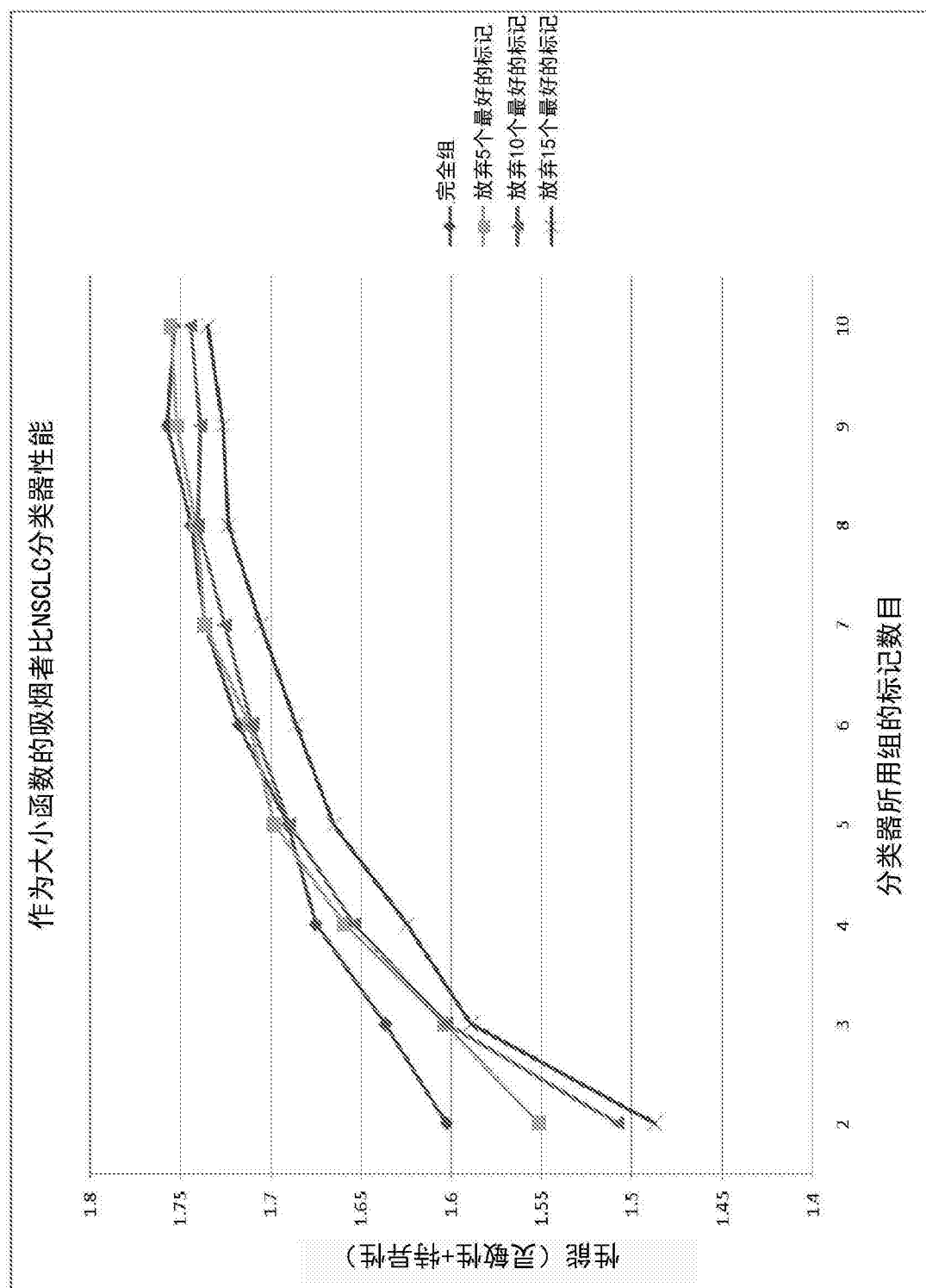


图19B

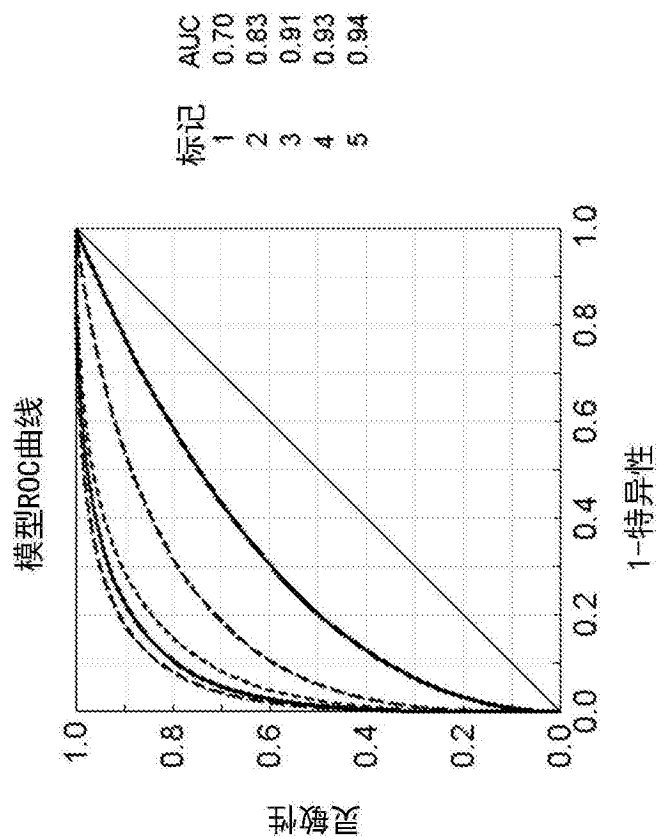


图20A

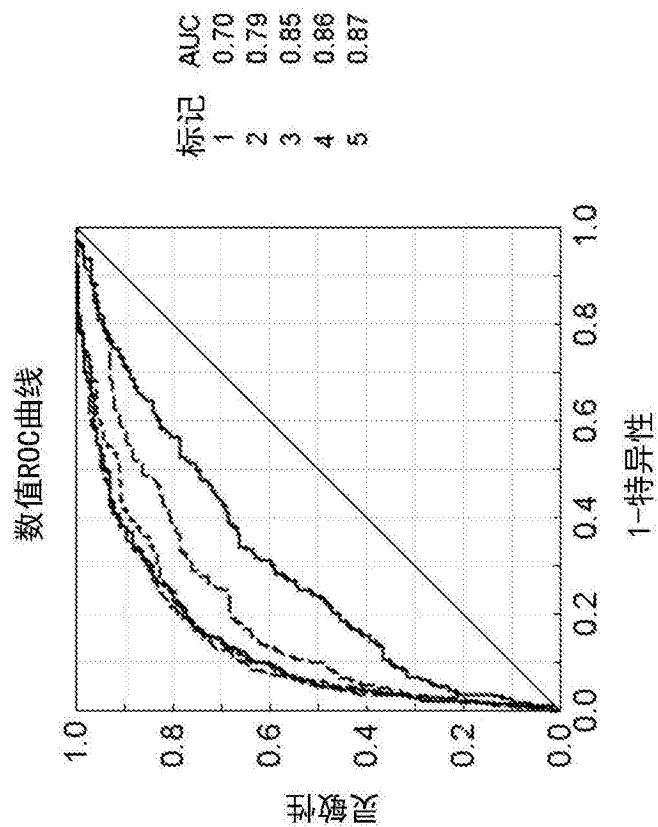


图20B

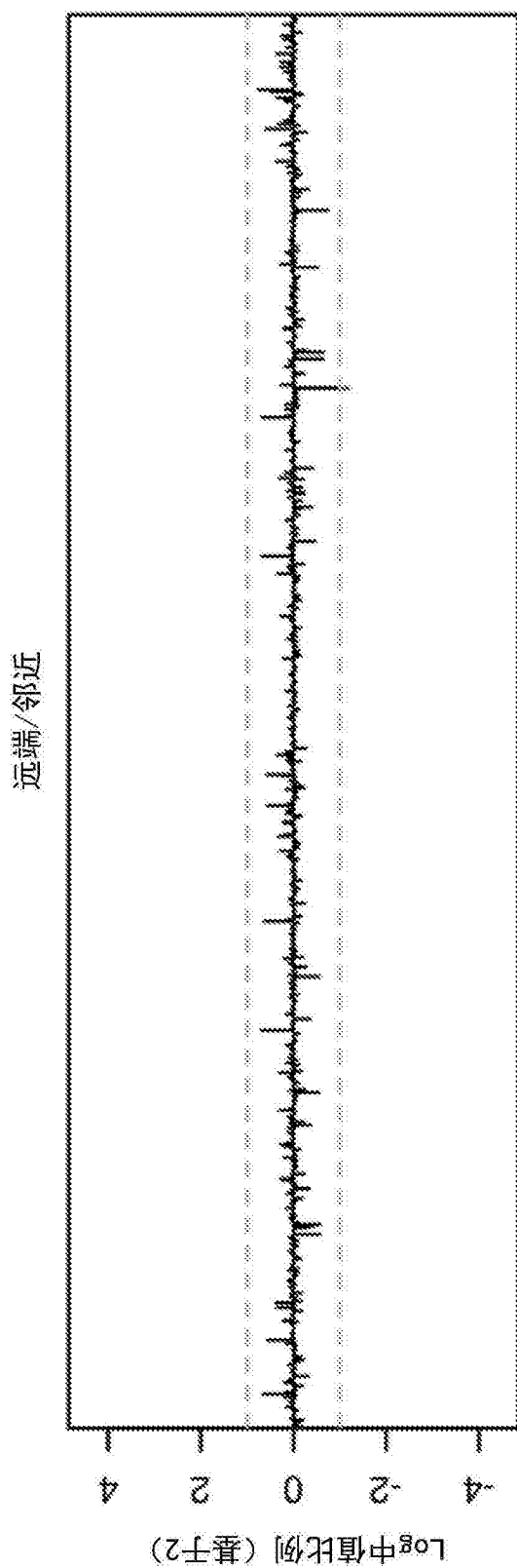


图21A

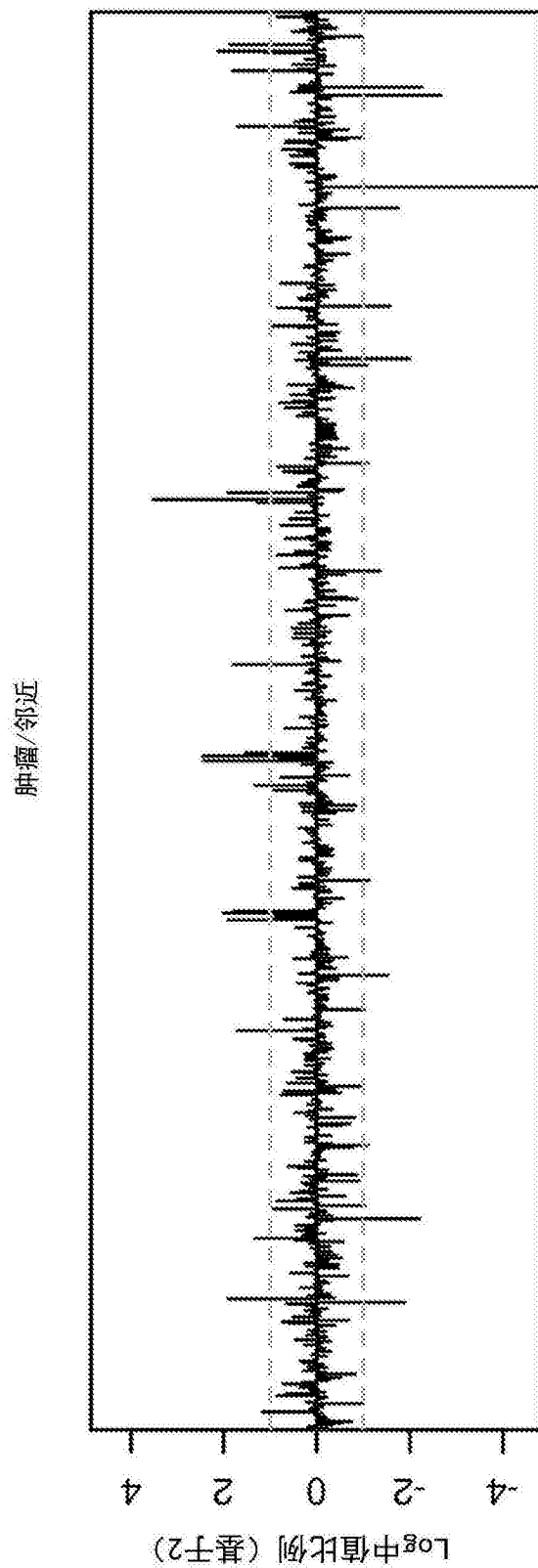


图21B

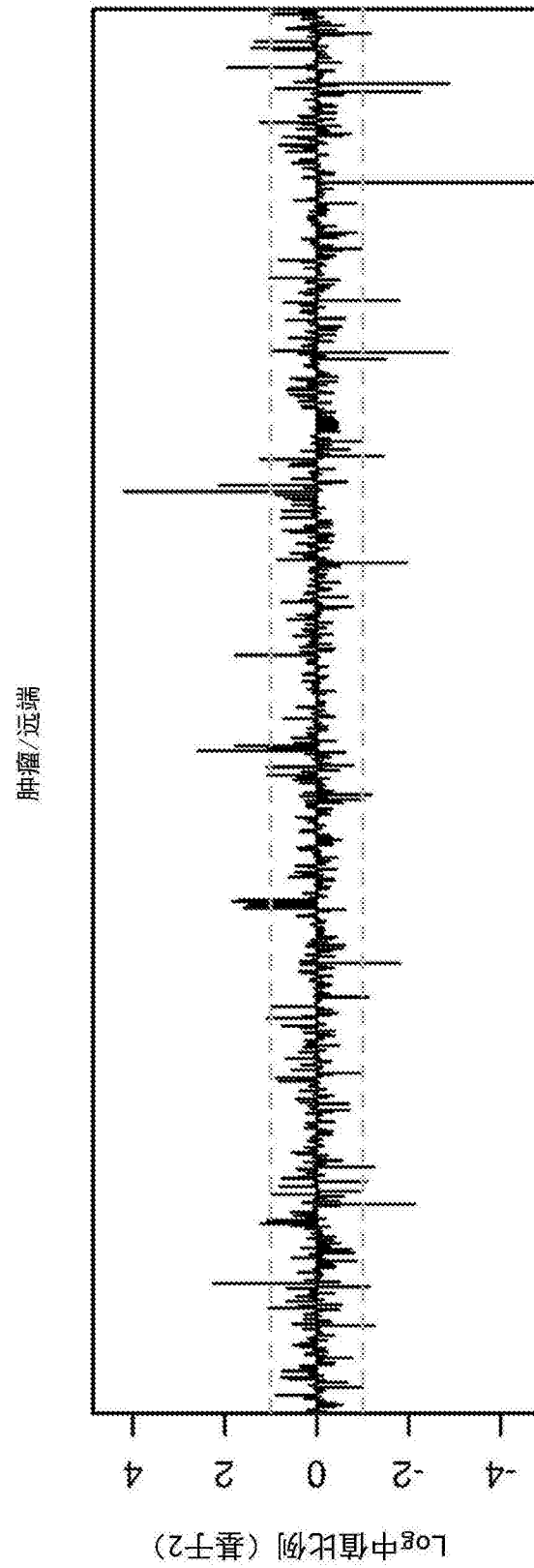


图21C

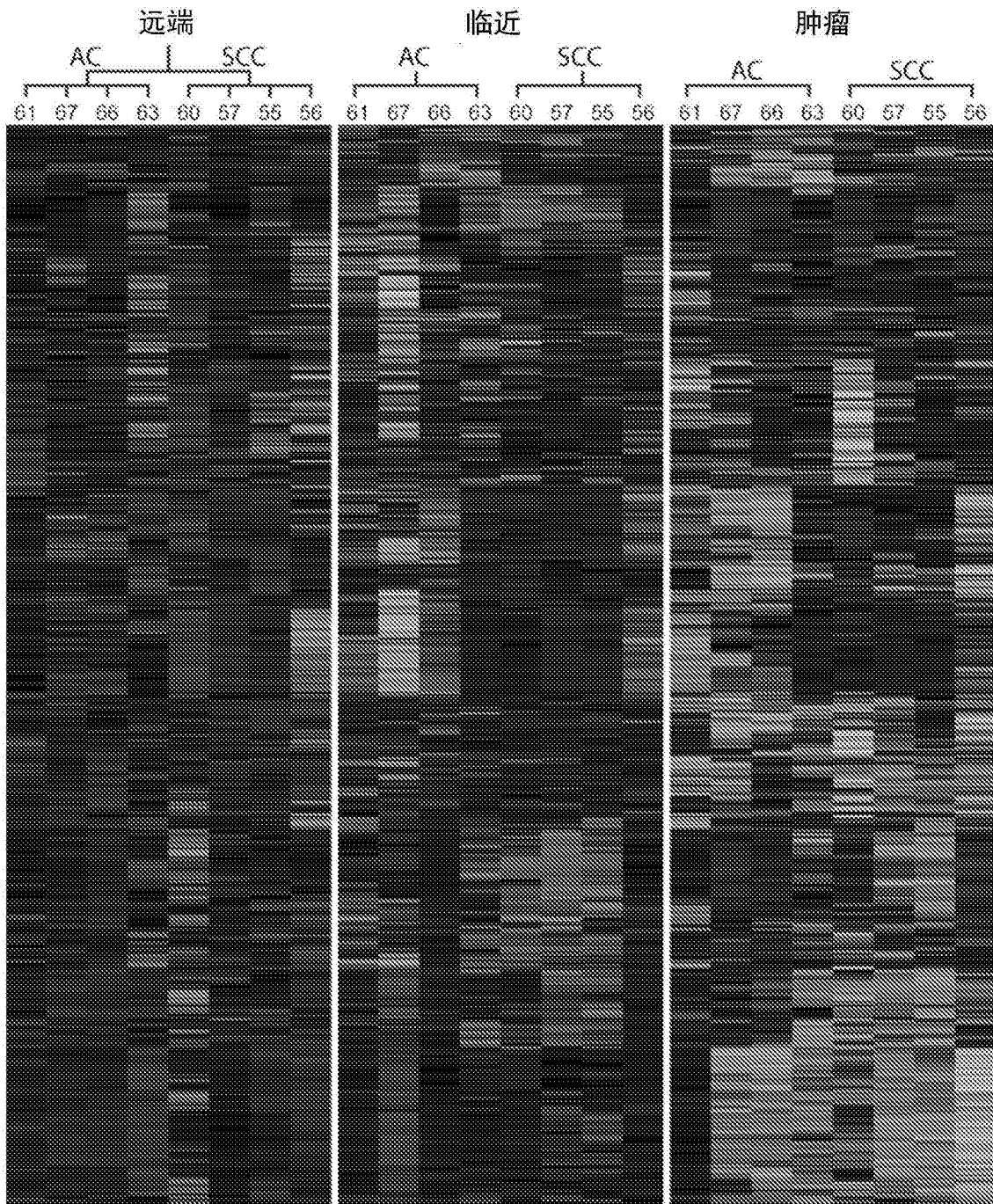


图22

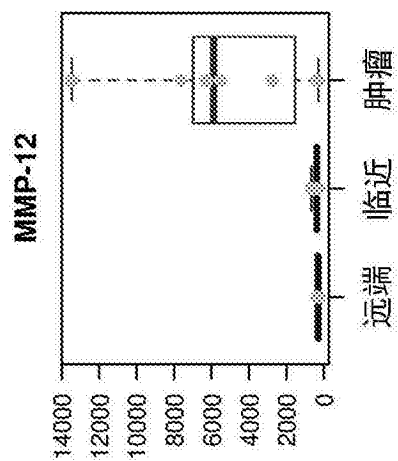


图23A

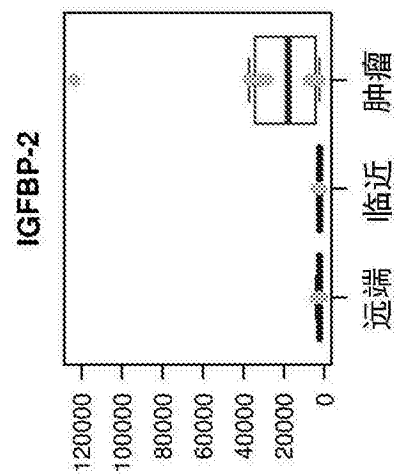


图23B

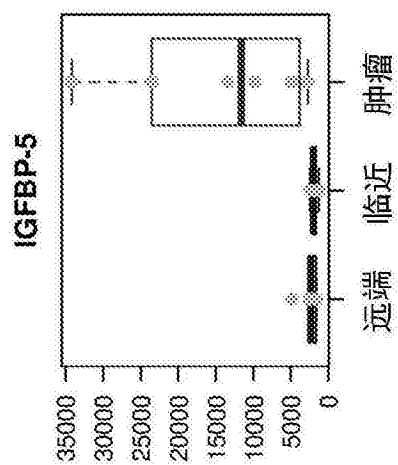


图23C

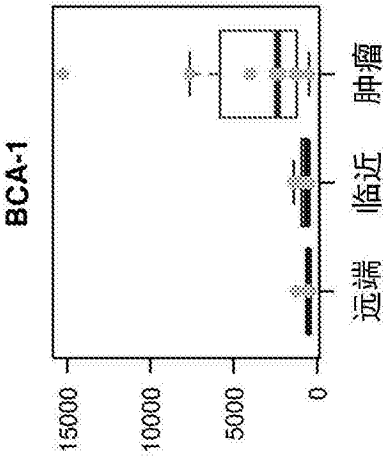


图23D

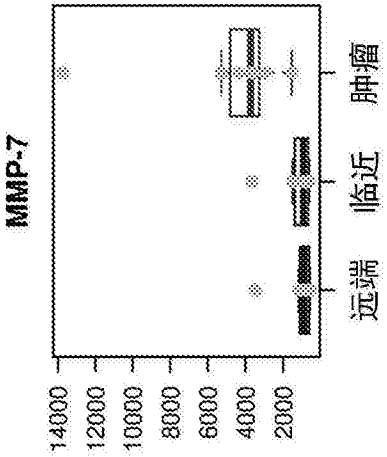


图23E

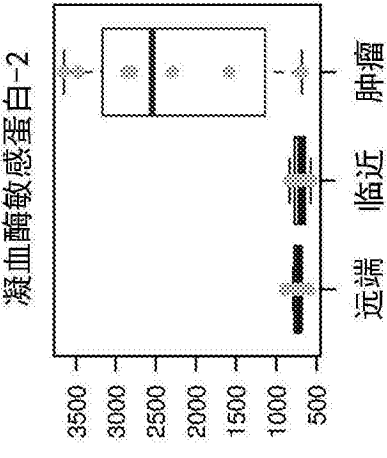


图23F

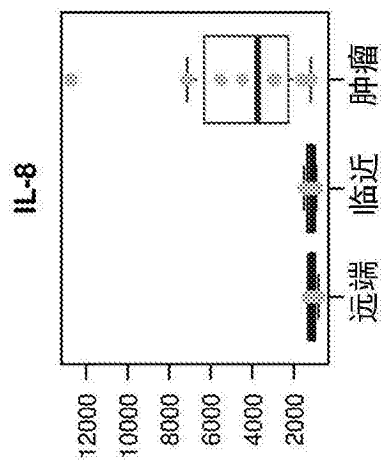


图23G

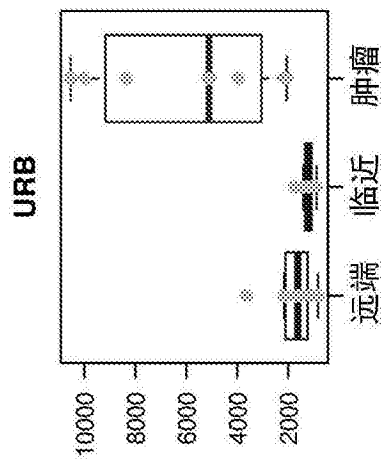


图23H

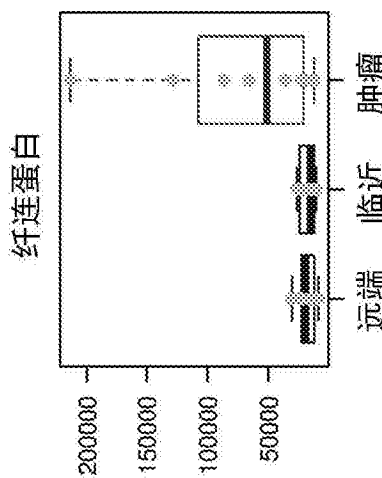


图23I

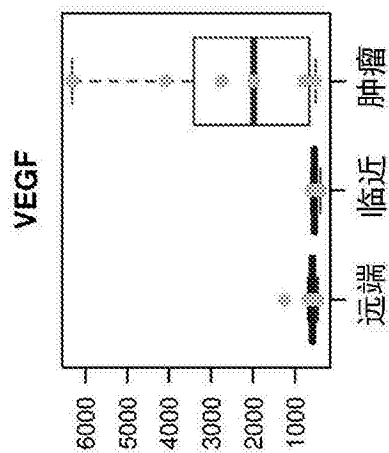


图23J

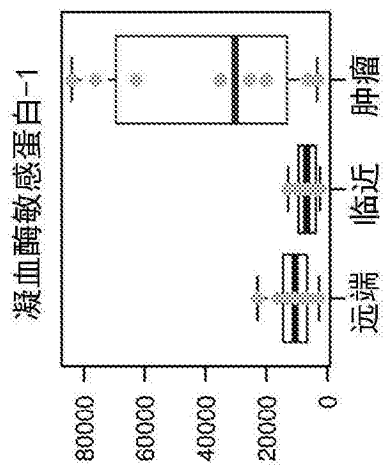


图23K

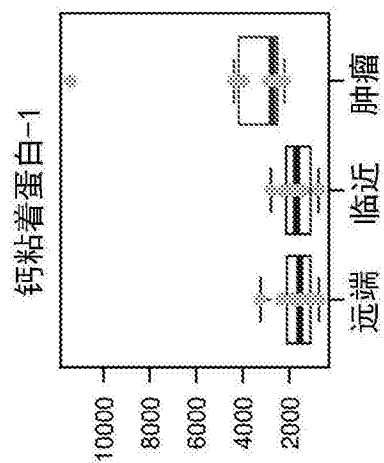


图23L

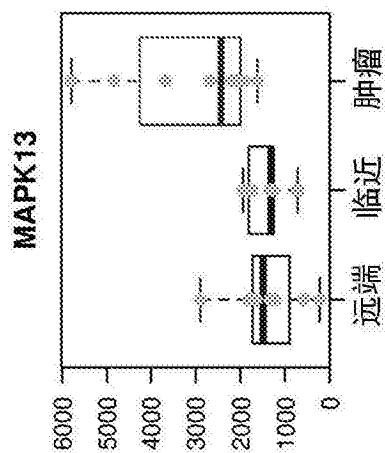


图23M

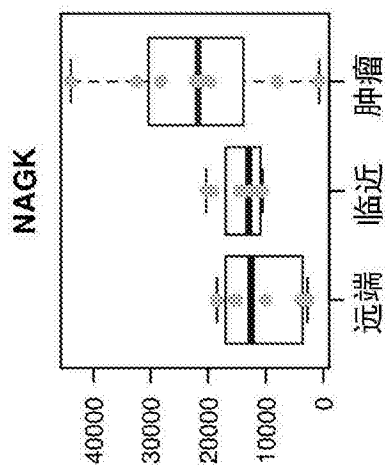


图23N

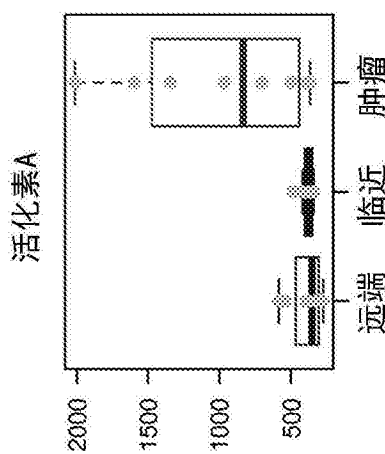


图23O

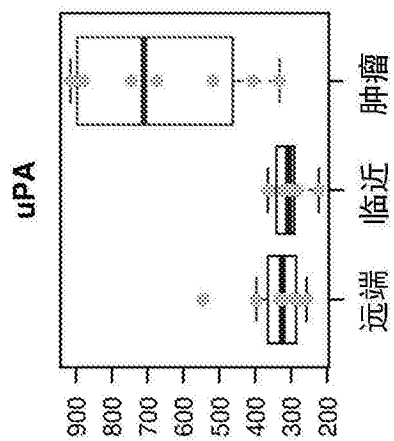


图23P

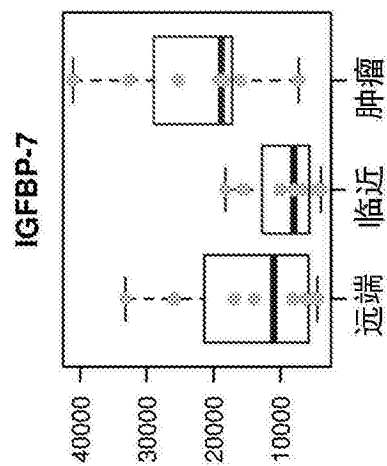


图23Q

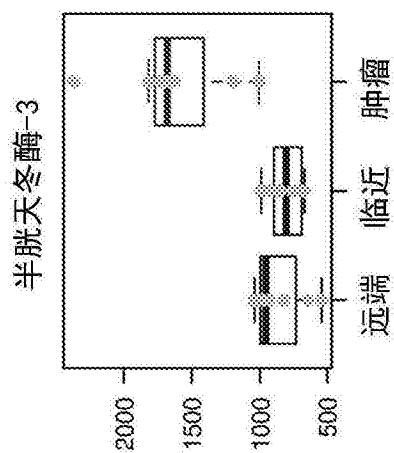


图23R

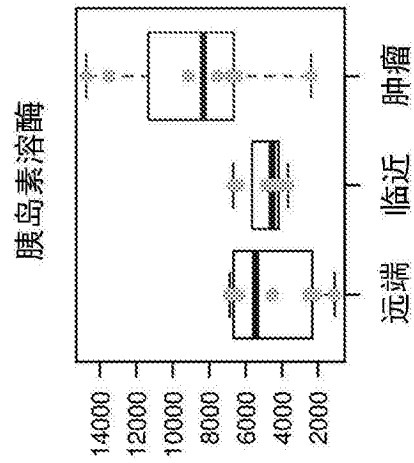


图23S

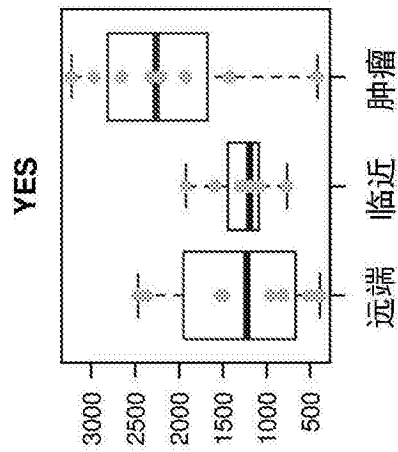


图23T

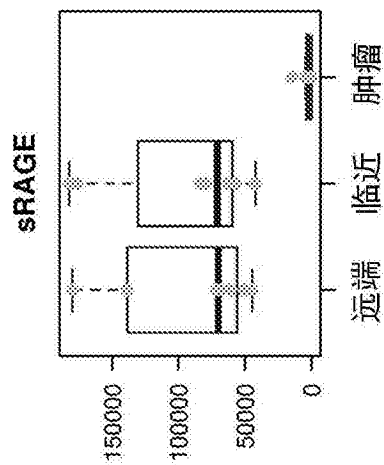


图24A

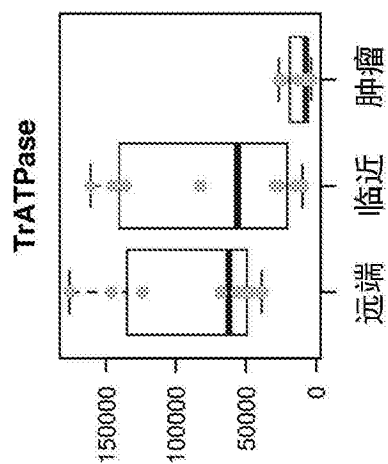


图24B

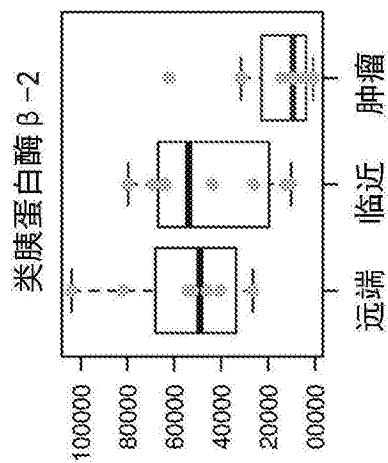


图24C

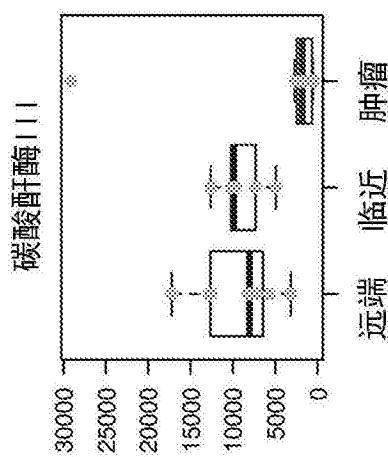


图24D

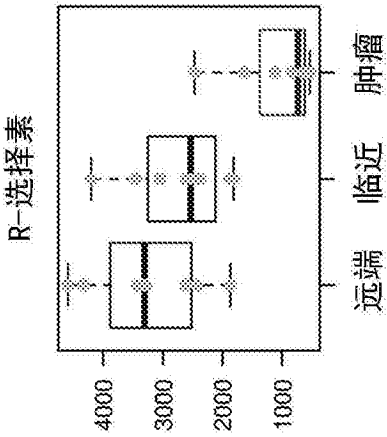


图24E

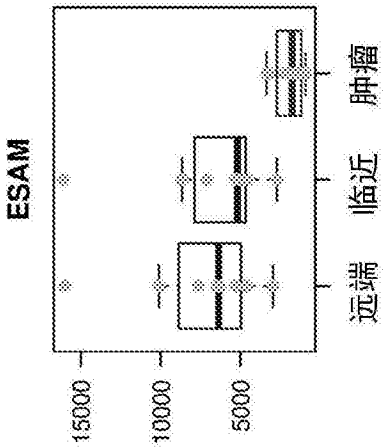


图24F

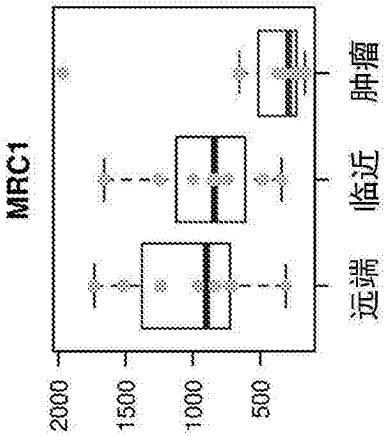


图24G

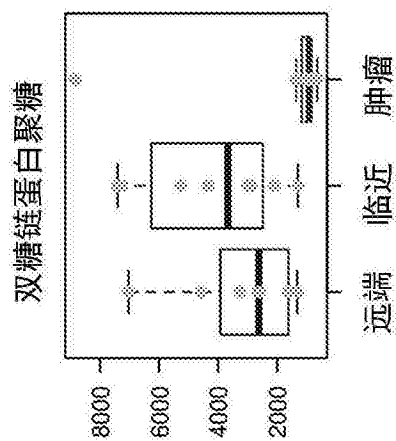


图24H

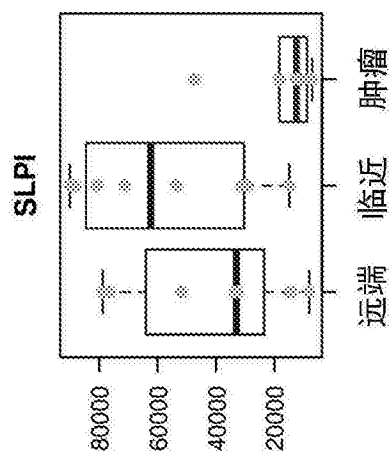


图24I

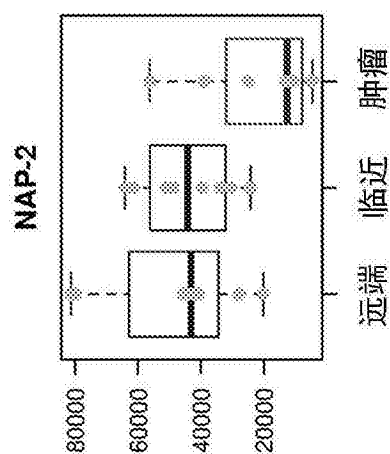


图24J

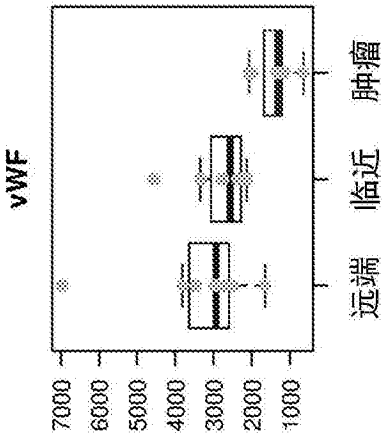


图24K

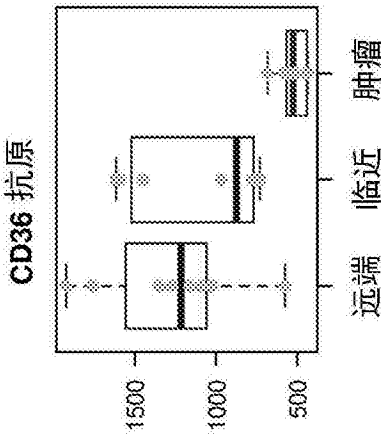


图24L

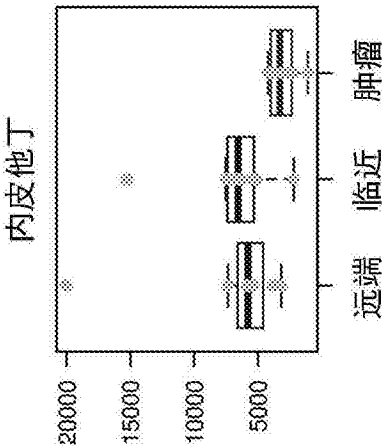


图24M

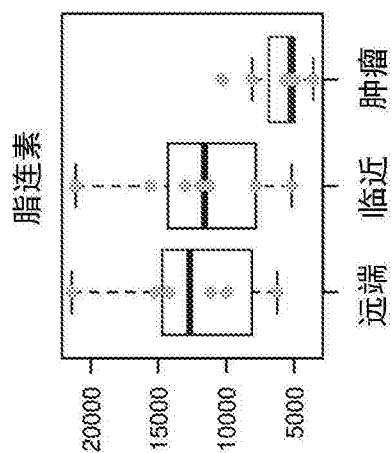


图24N

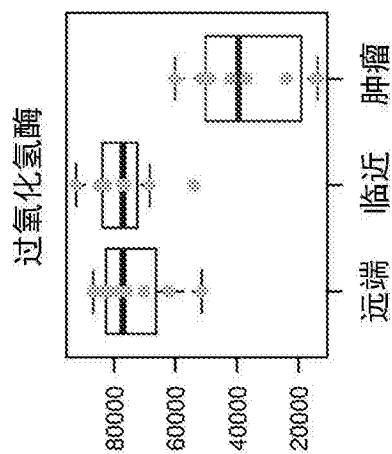


图24O

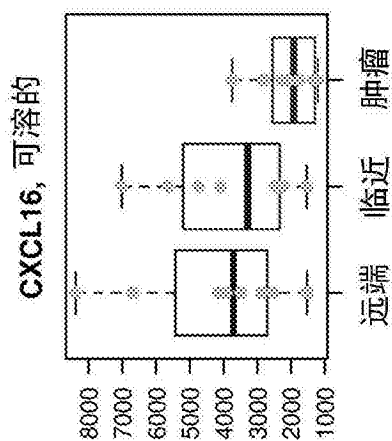


图24P

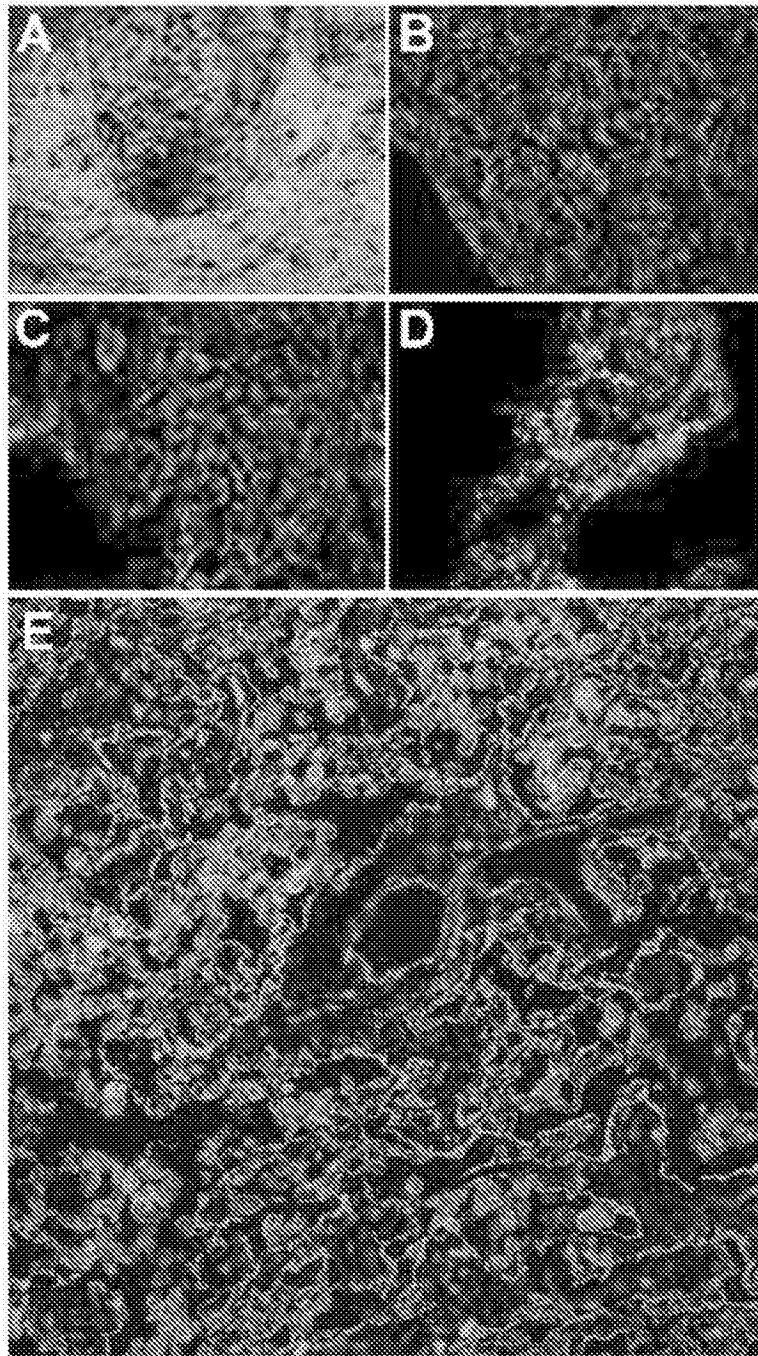


图25

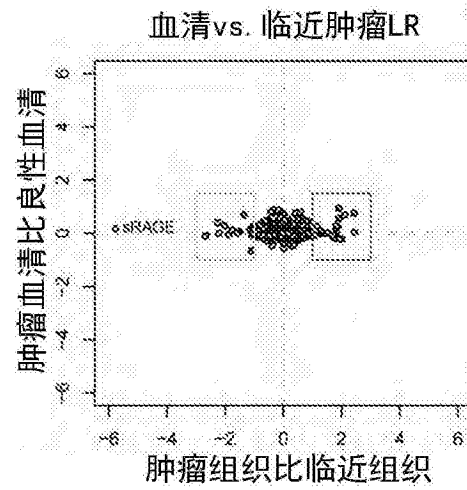


图26A

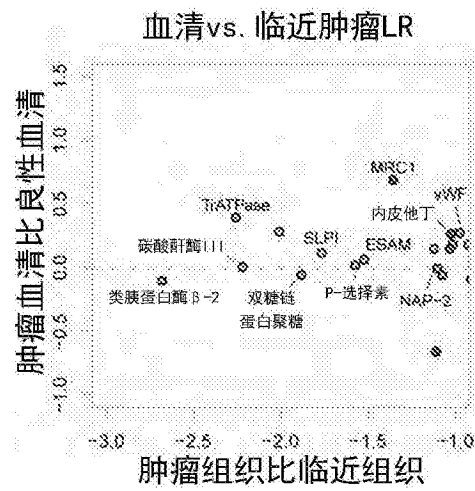


图26B

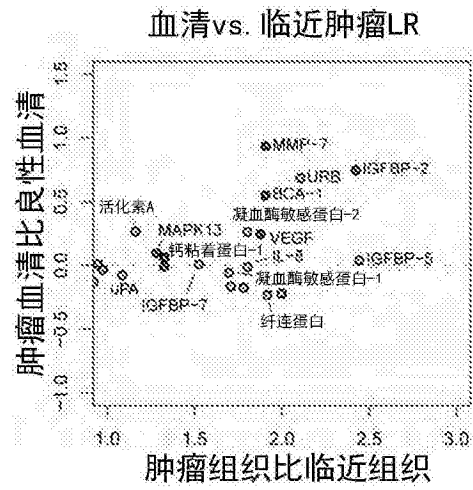


图26C

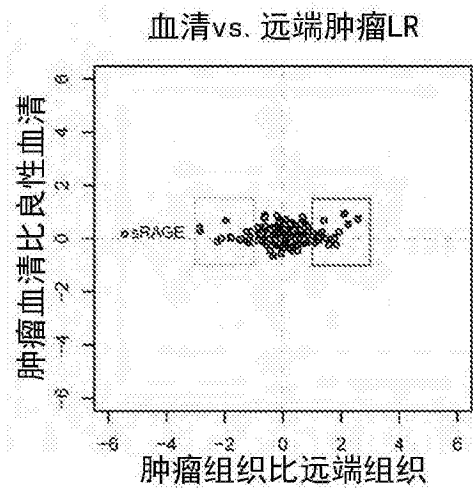


图26D

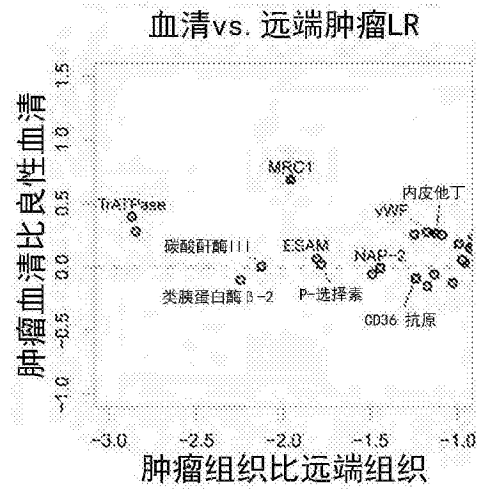


图26E

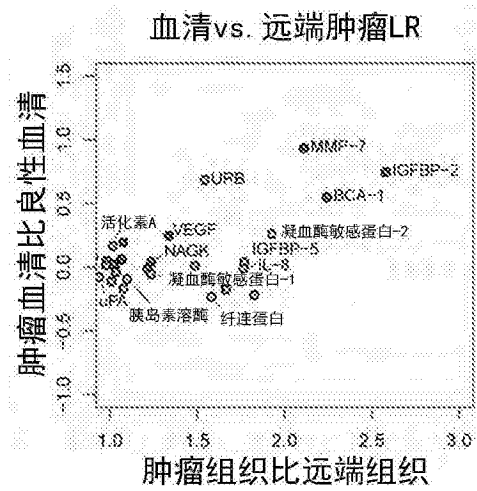


图26F

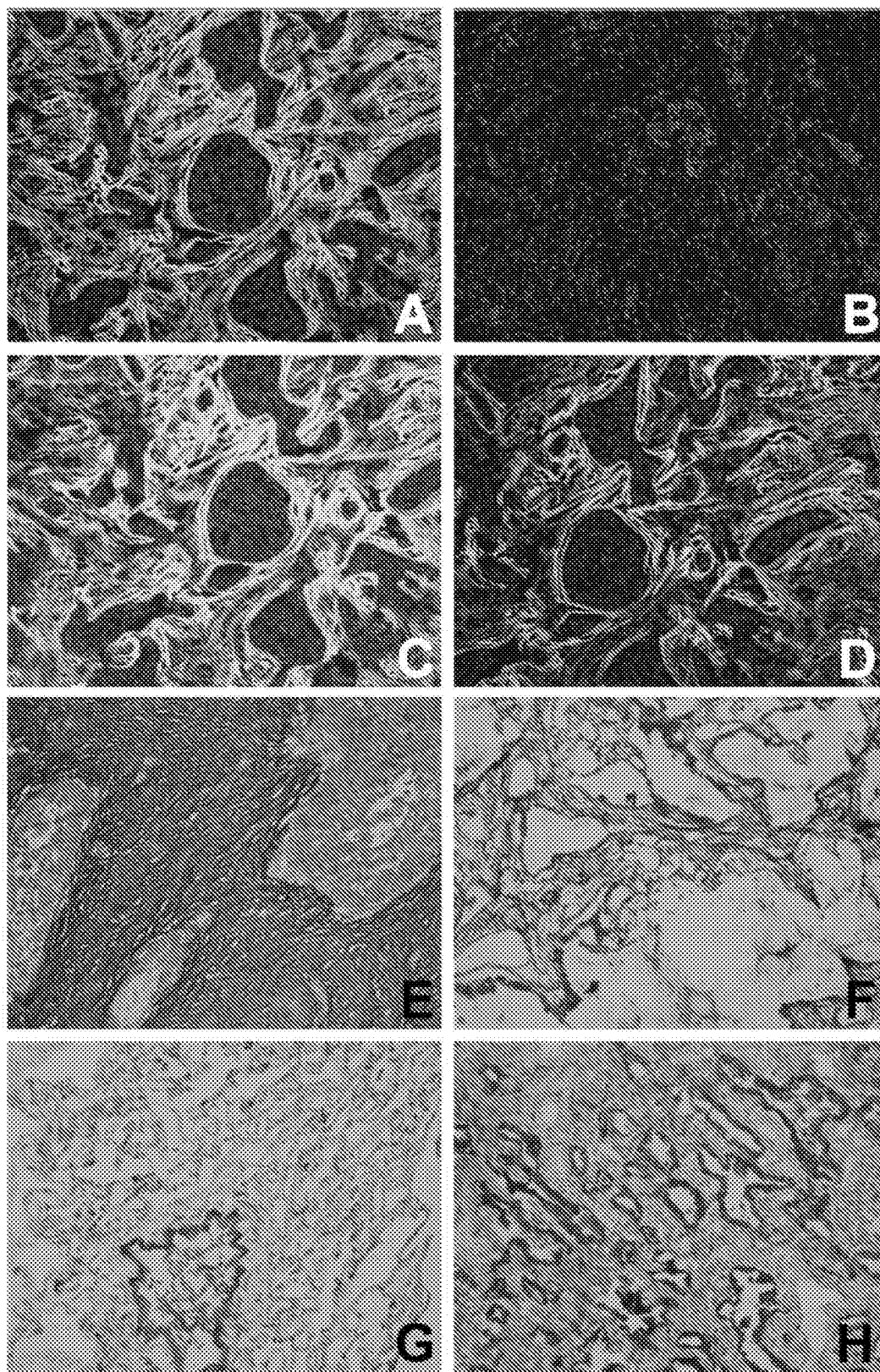


图27