



(12) 发明专利申请

(10) 申请公布号 CN 116987790 A

(43) 申请公布日 2023.11.03

(21) 申请号 202310993898.6

D·N·海耶斯

(22) 申请日 2017.05.17

(74) 专利代理机构 北京坤瑞律师事务所 11494

(30) 优先权数据

专利代理人 封新琴

62/337,591 2016.05.17 US

(51) Int.CI.

62/337,645 2016.05.17 US

C12Q 1/6886 (2018.01)

62/396,587 2016.09.19 US

62/420,836 2016.11.11 US

62/425,717 2016.11.23 US

(62) 分案原申请数据

201780044229.7 2017.05.17

(71) 申请人 基因中心治疗公司

地址 美国北卡罗来纳州

申请人 北卡罗来纳大学教堂山分校

(72) 发明人 H·法鲁基 M·莱-高德曼

G·梅休 J·赛罗德 C·佩鲁

权利要求书1页 说明书87页

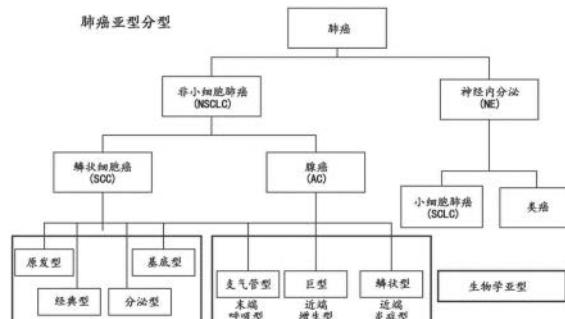
序列表(电子公布) 附图34页

(54) 发明名称

对肺腺癌亚型分型的方法

(57) 摘要

提供了用于通过检测选自一组肺腺癌基因标签的至少一种分类器生物标记物的表达水平来确定个体的肺腺癌(AD)亚型的方法和组合物。本文还提供了用于确定具有一种腺癌亚型的个体对诸如免疫疗法等疗法的反应的方法和组合物。



1. 一种用于确定从患者获得的肺组织样品的腺癌(AD)亚型的方法,该方法包括检测表1的至少一种分类器生物标记物的表达水平,其中对该分类器生物标记物的表达水平的检测特异性地鉴定末端呼吸单元(TRU)、近端增生(PP)或近端炎症(PI)AD亚型。

2. 权利要求1的方法,其中该方法还包括将检测到的表1的该至少一种分类器生物标记物的表达水平与至少一个样品训练集中的表1的该至少一种分类器生物标记物的表达进行比较,其中该至少一个样品训练集包括来自参考AD TRU样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PP样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PI样品的表1的该至少一种分类器生物标记物的表达数据或其组合;并且基于该比较步骤的结果将该样品分类为TRU、PP或PI亚型。

3. 权利要求2的方法,其中该比较步骤包括应用统计算法,该统计算法包括确定从该样品获得的该表达数据与来自该至少一个训练集的表达数据之间的相关性;并且基于该统计算法的结果将该样品分类为TRU、PP或PI亚型。

4. 上述权利要求中任一项的方法,其中在核酸水平上检测该分类器生物标记物的该表达水平。

5. 权利要求4的方法,其中该核酸水平是RNA或cDNA。

6. 权利要求4或5的方法,其中对表达水平的该检测包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。

7. 权利要求6的方法,其中通过进行qRT-PCR检测该表达水平。

8. 权利要求7的方法,其中对该表达水平的该检测包括使用对表1的至少一种分类器生物标记物有特异性的至少一对寡核苷酸引物。

9. 上述权利要求中任一项的方法,其中该样品是从该患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。

10. 权利要求9的方法,其中该体液是血液或其部分、尿液、唾液或痰液。

对肺腺癌亚型分型的方法

本申请是申请日为2017年5月17日、申请号为201780044229.7、发明名称为“对肺腺癌亚型分型的方法”的发明专利申请的分案申请。

交叉引用

本申请要求2016年5月17日提交的美国临时申请号62/337,591、2016年5月17日提交的美国临时申请号62/337,645、2016年9月19日提交的美国临时申请号62/396,587、2016年11月11日提交的美国临时申请号62/420,836和2016年11月23日提交的美国临时申请号62/425,717的优先权,出于所有目的将每个临时申请通过引用以其整体并入。

技术领域

本发明涉及用于确定肺样品的腺癌亚型和用于预测患有特定肺癌亚型的患者对治疗的反应的方法。

关于序列表的声明

与本申请相关的序列表以文本格式提供以代替纸质拷贝,并且通过引用特此并入本说明书中。包含该序列表的文本文件的名称是GNCN_009_01W0_SeqList_ST25.txt。文本文件为194KB,并且创建于2017年5月16日,并通过EFS-Web以电子方式提交。

发明背景

肺癌是美国以及全世界中癌症死亡的主要原因。2005年诊断出大约172,000例肺部肿瘤,估计有163,000例死亡,超过结肠、乳腺和前列腺癌症的总和。至少75%的患者出现局部晚期疾病。尽管已经作出很多努力来使用诸如高分辨率CT等技术改进筛查,但这些方法通常会产生假阳性结果,并且通常不会改变结果。因此,即使是早期检测到的小肿瘤也对于如下患者构成重大威胁,该患者的I期肺癌术后5年存活率估计在47%至63%之间。对于晚期疾病患者,预后较差,中值存活期低于一年。一般而言,姑息治疗有效但不可持续,并且对总体存活期的平均影响为大约3个月。

在群体水平上,肺癌的根本原因显然是烟草使用,其中全部肺癌中的90%直接归因于吸烟。吸烟与肺癌如此紧密相关,以致于它击败与大多数其他危险因素的决定性联系;虽然石棉、氡和一些肺部刺激物通常被认为是肺癌的危险因素。强烈怀疑遗传联系,然而,确切的机制仍然被确定在罕见的孟德尔癌症综合征选定组之外。尽管有许多分类方案和正在进行的临床试验,但在临床诊断和治疗领域总体上取得了令人失望的进展。

大多数肺癌被归类为非小细胞肺癌(NSCLC)(>85%),其是一个多样化的组,在整个呼吸道中都有亚型。腺癌(AD)和鳞状细胞癌(SCC或SQ)是NSCLC的两个主要亚型,其诊断频率几乎相等,但常常在不同位置被发现,SCC出现在更中心的位置。由世界卫生组织(WHO)开发的第6版肺癌共识分类(consensus classification of lung cancers)描述了不少于90种恶性形态类别和变体。通常可能存在异质性,尤其是在>1.5cm的较大肿瘤中,使得形态学分类更加困难并且导致诸如腺癌-鳞状细胞癌等命名。此外,组织学诊断再现性的研究已经表明病理学家内一致性和病理学家间一致性有限。形态学的变化性、有限的组织样品以及对不断增加的治疗靶向标记物列表评估的需要对当前的诊断标准提出了挑战。如下观点

进一步突出了这一点：肺癌的各种形态学亚型之间的区分在指导患者管理方面是必不可少的并且可以使用另外的分子测试来鉴定特定的治疗靶标记物。

目前，基于基因表达的肺腺癌(AD)亚型分型主要限于如下研究方案，该研究方案涉及：从新鲜冷冻肺肿瘤中提取RNA，随后使用超过500种基因的定量基因表达来应用最近质心(nearest centroid)预测因子。已显示基于基因表达的腺癌亚型分型将腺癌肿瘤分类为3个生物学上不同的亚型(末端呼吸单元型(TRU；以前称为支气管型(Bronchioid))、近端炎症型(PI；以前称为鳞状型(Squamoid))和近端增生型(PP；以前称为巨型(Magnoid)))，其可以在基因组谱(包括基因表达、突变谱和拷贝数变化)方面不同。此外，这三种亚型可以在其预后方面，在其吸烟者与非吸烟者的分布方面，在其EGFR改变、ALK重排、TP53突变的发生率方面，在其血管生成特征方面以及在其免疫原性反应特征方面各不相同。尽管有AD亚型分型的预后和预测益处的证据，但是需要>500种基因的基因表达结合复杂的生物信息学分析，阻碍了AD亚型分型在药物开发和/或临床中的应用。

癌症免疫监视的原则是免疫系统可以鉴定癌前细胞和癌性细胞，并在这些细胞变成临床相关性之前杀死它们，这已经在免疫缺陷小鼠模型中得到证明。先天性和适应性免疫反应可以一起工作以促进或抑制癌症生长，并且逃避免疫破坏是癌症的新兴标志。在历史上，免疫刺激的方法对于临床上的肺癌患者无效。肿瘤抗原表达的缺陷和抗原呈递细胞(APC)的呈现，免疫抑制细胞和细胞因子的浸润以及无效的T细胞活化可以导致在肿瘤部位的免疫抑制。对癌症和免疫系统的理解的进展已经导致有效的疗法，该疗法激活抗肿瘤反应，甚至在已经高度发展免疫逃避方法的肿瘤(例如肺癌)中。然而，由于患者中免疫激活和免疫抑制之间的微妙平衡，由肺部肿瘤引起的高免疫抑制作用限制了这些进展的有益效果。例如，在NSCLC中，阻碍免疫激活的免疫抑制细胞的作用很高，这被认为与肿瘤的类型、疾病的进展期和肿瘤负荷有关。

因此，开发一种有效区分固有的肺腺癌亚型的方法对临床诊断和疾病管理至关重要。因此，需要新的方法来进一步确定可能对免疫疗法有反应的群体。本发明部分地基于患者的腺癌亚型(末端呼吸单元型(TRU)、近端炎症型(PI)、近端增生型(PP))来解决该领域中的这些和其他需求，以便确定腺癌患者群体的预后或疾病结果。与本领域已知的诊断方法相比，本发明的方法提供了用于确定肺癌的细胞和分子起源(例如，对AD亚型分型)的手段，并且可以提供更准确的诊断和可应用的治疗。

发明概述

在一个方面，本文提供了用于确定从患者获得的肺组织样品的腺癌(AD)亚型的方法，该方法包括检测表1的至少一种分类器生物标记物的表达水平，其中该分类器生物标记物的表达水平的检测特异地鉴定末端呼吸单元(TRU)、近端增生(PP)或近端炎症(PI)AD亚型。在一些情况下，该方法还包括将检测到的表1的该至少一种分类器生物标记物的表达水平与至少一个样品训练集中的表1的该至少一种分类器生物标记物的表达进行比较，其中该至少一个样品训练集包括来自参考AD TRU样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PP样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PI样品的表1的该至少一种分类器生物标记物的表达数据或其组合；并且基于比较步骤的结果将样品分类为TRU、PP或PI亚型。在一些情况下，比较步骤包括应用统计算法，该统计算法包括确定从样品获得的表达数据与来自该至少一个训练集的表达数据之间的相

关性；并且基于统计算法的结果将样品分类为TRU、PP或PI亚型。在一些情况下，在核酸水平检测分类器生物标记物的表达水平。在一些情况下，核酸水平是RNA或cDNA。在一些情况下，对表达水平的检测包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。在一些情况下，通过进行qRT-PCR检测表达水平。在一些情况下，表达水平的检测包括使用对表1的至少一种分类器生物标记物有特异性的至少一对寡核苷酸引物。在一些情况下，样品是从患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。在一些情况下，体液是血液或其部分、尿液、唾液或痰液。在一些情况下，该至少一种分类器生物标记物包含多种分类器生物标记物。在一些情况下，该多种分类器生物标记物包含表1的至少两种分类器生物标记物、至少8种分类器生物标记物、至少16种分类器生物标记物、至少24种分类器生物标记物、至少32种分类器生物标记物、至少40种分类器生物标记物、或至少48种分类器生物标记物。在一些情况下，该至少一种分类器生物标记物包含表1的所有分类器生物标记物。

在另一个方面，本文提供了用于确定从患者获得的肺组织样品的腺癌(AD)亚型的方法，该方法包括检测编码在肺癌细胞中具有特定表达模式的分类器生物标记物的至少一种核酸分子的表达水平，其中该分类器生物标记物选自下组，该组由以下组成：表1中列出的分类器基因，该方法包括：(a) 从来自患者的肺组织样品中分离核酸材料；(b) 将该核酸材料与基本上与该分类器生物标记物的核酸分子的部分互补的寡核苷酸混合；并且(c) 检测该分类器生物标记物的表达。在一些情况下，该方法还包括将表1的该至少一种分类器生物标记物的检测到的表达水平与至少一个样品训练集中的表1的该至少一种分类器生物标记物的表达进行比较，其中该至少一个样品训练集包括来自参考AD TRU样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PP样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PI样品的表1的该至少一种分类器生物标记物的表达数据或其组合；并且基于比较步骤的结果将样品分类为TRU、PP或PI亚型。在一些情况下，比较步骤包括应用统计算法，该统计算法包括确定从样品获得的表达数据与来自该至少一个训练集的表达数据之间的相关性；并且基于统计算法的结果将样品分类为TRU、PP或PI亚型。在一些情况下，对表达水平的检测包括进行qRT-PCR或任何基于杂交的基因测定。在一些情况下，通过进行qRT-PCR检测表达水平。在一些情况下，表达水平的检测包括使用对表1的至少一种分类器生物标记物有特异性的至少一对寡核苷酸引物。在一些情况下，该方法还包括基于检测到的分类器生物标记物的表达水平预测对用于治疗肺腺癌(AD)亚型的疗法的反应。在一些情况下，该疗法是化学疗法、血管生成抑制剂和/或免疫疗法。在一些情况下，该肺AD亚型是TRU，并且疗法是化学疗法或血管生成抑制剂。在一些情况下，该肺AD亚型是PP，并且疗法是化学疗法。在一些情况下，该肺AD亚型是PI，并且疗法是免疫疗法。在一些情况下，样品是从患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。在一些情况下，体液是血液或其部分、尿液、唾液或痰液。在一些情况下，编码分类器生物标记物的该至少一种核酸分子包含编码多种分类器生物标记物的多种核酸分子。在一些情况下，该多种分类器生物标记物包含表1的至少两种分类器生物标记物、至少5种分类器生物标记物、至少10种分类器生物标记物、至少

20种分类器生物标记物、或至少30种分类器生物标记物。在一些情况下,该至少一种分类器生物标记物包含表1的所有分类器生物标记物。

在又另一个方面,本文提供了检测从患者获得的肺组织样品中的生物标记物的方法,该方法包括使用扩增、杂交和/或测序测定来测量选自表1的多种生物标记物核酸的表达水平。在一些情况下,肺组织样品先前被诊断为腺癌。在一些情况下,先前的诊断是通过组织学检查。在一些情况下,扩增、杂交和/或测序测定包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。在一些情况下,通过进行qRT-PCR检测表达水平。在一些情况下,对该表达水平的该检测包括对于选自表1的该多种生物标记物核酸中的每一种使用至少一对寡核苷酸引物。在一些情况下,样品是从患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。在一些情况下,体液是血液或其部分、尿液、唾液或痰液。在一些情况下,该多种生物标记物核酸包含表1的至少两种生物标记物核酸、至少10种生物标记物核酸、至少20种生物标记物核酸、至少30种生物标记物核酸、至少40种生物标记物核酸、至少50种生物标记物核酸、至少60种生物标记物核酸、或至少70种生物标记物核酸,基本上由其组成或由其组成。在一些情况下,该多种生物标记物核酸包含表1的所有分类器生物标记物核酸,基本上由其组成或由其组成。

在一个方面,本文提供了检测从患者获得的肺组织样品中的生物标记物的方法,该方法基本上由以下组成:使用扩增、杂交和/或测序测定来测量选自表1的多种生物标记物核酸的表达水平。在一些情况下,肺组织样品先前被诊断为腺癌。在一些情况下,先前的诊断是通过组织学检查。在一些情况下,扩增、杂交和/或测序测定包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。在一些情况下,通过进行qRT-PCR检测表达水平。在一些情况下,对该表达水平的该检测包括对于选自表1的该多种生物标记物核酸中的每一种使用至少一对寡核苷酸引物。在一些情况下,样品是从患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。在一些情况下,体液是血液或其部分、尿液、唾液或痰液。在一些情况下,该多种生物标记物核酸包含表1的至少两种生物标记物核酸、至少10种生物标记物核酸、至少20种生物标记物核酸、至少30种生物标记物核酸、至少40种生物标记物核酸、至少50种生物标记物核酸、至少60种生物标记物核酸、或至少70种生物标记物核酸,基本上由其组成或由其组成。在一些情况下,该多种生物标记物核酸包含表1的所有分类器生物标记物核酸,基本上由其组成或由其组成。

在另一个方面,本文提供了检测从患者获得的肺组织样品中的生物标记物的方法,该方法由以下组成:使用扩增、杂交和/或测序测定来测量选自表1的多种生物标记物核酸的表达水平。在一些情况下,肺组织样品先前被诊断为腺癌。在一些情况下,先前的诊断是通过组织学检查。在一些情况下,扩增、杂交和/或测序测定包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。在一些情况下,通过进行qRT-PCR检测表达水平。在一些情况下,对该

表达水平的该检测包括对于选自表1的该多种生物标记物核酸中的每一种使用至少一对寡核苷酸引物。在一些情况下,样品是从患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。在一些情况下,体液是血液或其部分、尿液、唾液或痰液。在一些情况下,该多种生物标记物核酸包含表1的至少两种生物标记物核酸、至少10种生物标记物核酸、至少20种生物标记物核酸、至少30种生物标记物核酸、至少40种生物标记物核酸、至少50种生物标记物核酸、至少60种生物标记物核酸、或至少70种生物标记物核酸,基本上由其组成或由其组成。在一些情况下,该多种生物标记物核酸包含表1的所有分类器生物标记物核酸,基本上由其组成或由其组成。

在另一个方面,本文提供了确定腺癌患者是否可能对免疫疗法有反应的方法,该方法包括确定来自患者的肺组织样品的腺癌亚型,其中该腺癌亚型选自下组,该组由以下组成:鳞状型(近端炎症型)、支气管型(末端呼吸单元型)和巨型(近端增生型);并且基于亚型,评估患者是否可能对免疫疗法有反应。在一些情况下,免疫疗法包括检查点抑制剂疗法。在一些情况下,检查点抑制剂靶向PD-1或PD-L1。在一些情况下,检查点抑制剂靶向CTLA-4。在一些情况下,检查点抑制剂是派姆单抗(Pembrolizumab)、纳武单抗(Nivolumab)或其抗原片段结合片段。在一些情况下,检查点抑制剂是伊匹单抗(Ipilimumab)或其抗原结合片段。在一些情况下,最初经由样品的组织学分析确定患者患有腺癌。在一些情况下,患者的腺癌分子亚型选自鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型),并且经由对从患者获得的样品的组织学分析来确定。在一些情况下,样品是从患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体或体液。在一些情况下,体液是血液或其部分、尿液、唾液或痰液。在一些情况下,对腺癌亚型的确定包括确定多种分类器生物标记物的表达水平。在一些情况下,对该多种分类器生物标记物的表达水平的确定是在核酸水平上通过进行RNA测序、逆转录酶聚合酶链反应(RT-PCR)或基于杂交的分析来进行。在一些情况下,用于确定腺癌亚型的该多种分类器生物标记物选自可公开获得的肺腺癌数据集。在一些情况下,该可公开获得的肺腺癌数据集是TCGA肺AD RNAseq数据集。在一些情况下,用于确定腺癌亚型的该多种分类器生物标记物选自表1。在一些情况下,RT-PCR是定量实时逆转录酶聚合酶链反应(qRT-PCR)。在一些情况下,使用对表1的该多种分类器生物标记物有特异性的引物进行RT-PCR。在一些情况下,该方法还包括将检测到的表1的该多种分类器生物标记物的表达水平与至少一个样品训练集中的表1的该多种分类器生物标记物的表达进行比较,其中该至少一个样品训练集包括来自参考腺癌TRU样品的表1的该多种分类器生物标记物的表达数据、来自参考腺癌PP样品的表1的该多种分类器生物标记物的表达数据、来自参考腺癌PI样品的表1的该多种分类器生物标记物的表达数据、或其组合;并且基于比较步骤的结果将第一个样品分类为TRU、PP或PI。在一些情况下,比较步骤包括应用统计算法,该统计算法包括确定从样品获得的表达数据与来自该至少一个训练集的表达数据之间的相关性;并且基于统计算法的结果将样品分类为TRU、PP或PI亚型。在一些情况下,该多种分类器生物标记物包含表1中列出的分类器生物标记物中的每一种。

在又另一个方面,本文提供了用于选择腺癌患者进行免疫疗法的方法,该方法包括确定来自该患者的肺组织样品的腺癌亚型;并且基于该亚型选择患者进行免疫疗法。在一些情况下,免疫疗法包括检查点抑制剂疗法。在一些情况下,检查点抑制剂靶向PD-1或

PD-L1。在一些情况下,检查点抑制剂靶向CTLA-4。在一些情况下,检查点抑制剂是派姆单抗(Pembrolizumab)、纳武单抗(Nivolumab)或其抗原片段结合片段。在一些情况下,检查点抑制剂是伊匹单抗(Ipilimumab)或其抗原结合片段。在一些情况下,最初经由样品的组织学分析确定患者患有腺癌。在一些情况下,患者的腺癌分子亚型选自鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型),并且经由对从患者获得的样品的组织学分析来确定。在一些情况下,样品是从患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体或体液。在一些情况下,体液是血液或其部分、尿液、唾液或痰液。在一些情况下,对腺癌亚型的确定包括确定多种分类器生物标记物的表达水平。在一些情况下,对该多种分类器生物标记物的表达水平的确定是在核酸水平上通过进行RNA测序、逆转录酶聚合酶链反应(RT-PCR)或基于杂交的分析来进行。在一些情况下,用于确定腺癌亚型的该多种分类器生物标记物选自可公开获得的肺腺癌数据集。在一些情况下,可公开获得的肺腺癌数据集是TCGA肺AD RNAseq数据集。在一些情况下,用于确定腺癌亚型的该多种分类器生物标记物选自表1。在一些情况下,RT-PCR是定量实时逆转录酶聚合酶链反应(qRT-PCR)。在一些情况下,使用对表1的该多种分类器生物标记物有特异性的引物进行RT-PCR。在一些情况下,该方法还包括将检测到的表1的该多种分类器生物标记物的表达水平与至少一个样品训练集中的表1的该多种分类器生物标记物的表达进行比较,其中该至少一个样品训练集包括来自参考腺癌TRU样品的表1的该多种分类器生物标记物的表达数据、来自参考腺癌PP样品的表1的该多种分类器生物标记物的表达数据、来自参考腺癌PI样品的表1的该多种分类器生物标记物的表达数据、或其组合;并且基于比较步骤的结果将第一个样品分类为TRU、PP或PI。在一些情况下,比较步骤包括应用统计算法,该统计算法包括确定从样品获得的表达数据与来自该至少一个训练集的表达数据之间的相关性;并且基于统计算法的结果将样品分类为TRU、PP或PI亚型。在一些情况下,该多种分类器生物标记物包含表1中列出的分类器生物标记物中的每一种。

在一个方面,本文提供了治疗受试者的肺癌的方法,该方法包括:测量从受试者获得的肺癌样品中至少一种生物标记物核酸的表达水平,其中该至少一种生物标记物核酸选自表1中列出的一组生物标记物,其中该至少一种生物标记物的存在、不存在和/或水平指示肺癌的亚型;并且基于肺癌的亚型给予免疫治疗剂。在一些情况下,肺癌样品是腺癌肺癌样品,在一些情况下,选自该组生物标记物的该至少一种生物标记物核酸包含表1的至少两种生物标记物核酸、至少8种生物标记物核酸、至少16种生物标记物核酸、至少32种生物标记物核酸、或所有48种生物标记物核酸,基本上由其组成或由其组成。在一些情况下,肺组织样品先前被诊断为腺癌。在一些情况下,先前的诊断是通过组织学检查。在一些情况下,该方法还包括测量来自另外一组生物标记物的至少一种生物标记物的表达。在一些情况下,另外一组生物标记物包括先天性免疫细胞(IIC)、适应性免疫细胞(AIC)的基因表达标签(gene expression signature),一种或多种个体免疫生物标记物,一种或多种干扰素(IFN)基因,一种或多种主要组织相容性复合物II类(MHCII)基因或其组合。在一些情况下,另外一组生物标记物包含选自表4A、表4B、表5、表6、表7或其组合的基因。在一些情况下,AIC的基因表达标签选自表4A。在一些情况下,IIC的基因表达标签选自表4B。在一些情况下,该一种或多种个体免疫生物标记物选自表5。在一些情况下,该一种或多种IFN基因选自表6。在一些情况下,该一种或多种MHCII基因选自表7。在一些情况下,测量表达水平使用扩

增、杂交和/或测序测定来进行。在一些情况下，扩增、杂交和/或测序测定包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。在一些情况下，通过进行qRT-PCR检测表达水平。在一些情况下，样品是从患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。在一些情况下，体液是血液或其部分、尿液、唾液或痰液。在一些情况下，受试者的腺癌亚型选自鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型)。在一些情况下，肺癌亚型是近端炎症型的，并且其中该免疫治疗剂包含检查点抑制剂。在一些情况下，检查点抑制剂靶向PD-1或PD-L1。在一些情况下，检查点抑制剂靶向CTLA-4。在一些情况下，检查点抑制剂是派姆单抗(Pembrolizumab)、纳武单抗(Nivolumab)或其抗原片段结合片段。在一些情况下，检查点抑制剂是伊匹单抗(Ipilimumab)或其抗原结合片段。在一些情况下，该至少一种生物标记物核酸是多种生物标记物核酸，其中该多种生物标记物核酸包含表1中所列出的至少一种生物标记物核酸与来自可公开获得的肺腺癌数据集的一种或多种生物标记物核酸的组合，其中该多种生物标记物核酸的存在、不存在和/或水平指示肺癌的亚型。在一些情况下，该至少一种生物标记物核酸是多种生物标记物核酸，其中该多种生物标记物核酸包含表1中所列出的所有生物标记物核酸与来自可公开获得的肺腺癌数据集的一种或多种生物标记物核酸的组合，其中该多种生物标记物核酸的存在、不存在和/或水平指示肺癌的亚型。在一些情况下，可公开获得的肺腺癌数据集是TCGA肺AD RNAseq数据集。

附图简述

图1示出了肺癌亚型分型和鳞状细胞癌(SCC或SQ)和腺癌(AC或AD)的生物学亚型。

图2示出了在实施例1中描述的研究中使用的肺AD数据集。

图3示出了癌症基因组图集(Cancer Genome Atlas, TCGA)肺AD数据集中的免疫细胞标签表达(即，来自实施例1的Bindea等人的参考文献)、其他免疫标记物和个体免疫标记物的热图。TRU=末端呼吸单元型，PP=近端增生型，PI=近端炎症型。

图4示出了TCGA AD数据集中的免疫细胞标签的相关矩阵，其中通过分层聚类来排列标签。深红色表示一对基因更呈正相关。白色表示没有相关。深蓝色表示负相关。

图5示出了如实施例1中所述的跨多个AD数据集的T细胞标签基因表达亚型模式的再现性。TRU=末端呼吸单元型，PP=近端增生型，PI=近端炎症型。RNAseq(Illumina, San Diego, CA)以及来自Affymetrix(Santa Clara, CA)和Agilent(Santa Clara, CA)两者的微阵列。

图6示出了在TCGA数据集的腺癌(AD)评估中在CD274(PD-L1)表达与适应性免疫细胞(AIC)标签之间的关联(经调整的R平方)。亚型的关联始终大于PD-L1的关联。Tcm=中枢记忆T细胞，Tem=效应记忆T细胞，Th1=1型T辅助细胞，Th2=2型T辅助细胞，TFH=T滤泡辅助细胞，Th17=T辅助17细胞，Treg=T调节细胞，Tgd=γδT细胞。

图7A-图7B示出了如实施例1中所述的整体和按照亚型的标签-存活关联。从分层的cox模型计算的风险比(HR)和置信区间(CI)对应于标准化免疫标记物的单位增加，并针对病理阶段进行调整。针对阶段调整亚型特异性HR(整体通过阶段和亚型进行调整)，并且仅显示了对于至少一种亚型具有显著关联(标称p<0.05)的免疫特征。AD=腺癌，TRU=末端

呼吸单元型,PP=近端增生型,PI=近端炎症型,MHC II=主要组织相容性II类基因标签,Th1=1型T辅助细胞,Th2=2型T辅助细胞,TFH=T滤泡辅助细胞,Th17=T辅助17细胞,Treg=T调节细胞,DC=树突细胞,iDC=未成熟树突细胞。图7A-图7B显示了TCGA群组中(图7A)或者TGCA、Shedden和Tomida群组中(图7B)免疫细胞标签和标记物按照AD亚型的存活关联。对于图7A,亚型特异性免疫标记物风险比和95%置信区间是针对TCGA群组($n=515$ AD)中的5年总体存活期。

图8示出了在RNASeq肺腺癌(AD)数据集上从癌症基因组图集(TCGA)进行的五折交叉验证研究。用于确定包括以用于对AD亚型分型的最佳基因数量。末端呼吸单元(TRU)以前称为支气管型。近端增生型(PP)以前称为巨型。近端炎症型(PI)以前称为鳞状型。

图9示出了通过本文所述的AD预测因子的基因标签训练的轮廓(silhouette)得分来选择原型样品。

图10示出了来自选择用于区分支气管型样品(末端呼吸单元型)的48个基因分类器的16基因子集的中值基因表达。

图11示出了来自选择用于区分巨型样品(近端增生型)的48个基因分类器的16基因子集的中值基因表达。

图12示出了来自选择用于区分鳞状型样品(近端炎症型)的48个基因分类器的16基因子集的中值基因表达。

图13示出了本文提供的48个基因标签与定义多个验证数据集的金标准亚型的506个基因分类器的AD亚型预测的一致性。与金标准(TCGA)的一致性为87%。与Shedden、Tomida、UNC和FFPE的一致性分别为87%、79%、92%和84%。

图14示出了癌症基因组图集(TCGA)肺AD数据集中的免疫细胞标签(即,来自实施例3的Bindea等人的参考文献)、和其他免疫标记物的热图。

图15示出了如实施例3中所述的跨多个AD数据集的T细胞标签基因表达亚型模式的再现性。

图16示出了如实施例3中所述的在CD274(PD-L1)表达与适应性免疫细胞(AIC)标签之间相对于在亚型与AIC标签之间的关联(经调整的R平方)。Tcm=中枢记忆T细胞,Tem=效应记忆T细胞,Th1=1型T辅助细胞,Th2=2型T辅助细胞,TFH=T滤泡辅助细胞,Th17=T辅助17细胞,Treg=T调节细胞,Tgd= γ δT细胞。

图17示出了如实施例3中所述的整体和按照亚型的AD标签-存活关联。根据分层cox模型计算风险比(HR)和置信区间。针对阶段调整亚型特异性HR(整体通过阶段和亚型进行调整),并且仅显示了具有显著关联($p<0.05$)的免疫特征。

图18示出了所有免疫细胞和免疫标志物(即,IFN基因、MHCII基因和个体免疫标志物PDL1、PDL2、PDCD1和CTLA4)相对于AD的箱形图。TRU=末端呼吸单元型,PP=近端增生型,PI=近端炎症型,AC=腺癌。

图19示出了腺癌(AD)亚型非沉默突变负荷、AD中的STK11失活(突变和/或缺失)、和MHC II类标签,连同Kruskal-Wallis关联检验p值。TRU=末端呼吸单元型,PP=近端增生型,PI=近端炎症型,MHC II=主要组织相容性II类基因标签。

图20示出了在增生、非沉默突变负荷、和关键药物靶标(CD274(PD-L1)、PDCD1(PD-1)和CTLA4)方面的显著腺癌(AD)亚型差异。如实施例4中所述来测定AD。

图21示出了针对临床实体瘤突变小组(表8中公开的322种基因)中的大多数基因,AD亚型的显著药物靶基因表达差异。在AD亚型中,65%的小组基因显示出显著可变的表达(KW Bonferroni阈值p<0.000155)。如实施例4中所述,测定AD亚型分型。

图22示出了针对临床实体瘤突变小组(表8中公开的322种基因)中的大多数基因,AD的显著药物靶基因表达差异。在AD亚型中,63%的小组基因显示出显著可变的表达(KW Bonferroni阈值p<0.000155)。如实施例5中所述,测定AD亚型分型。

图23示出了在增生方面的显著腺癌(AD)亚型差异。如实施例5中所述,测定AD亚型分型。

发明详述

概述

本发明提供了用于鉴定或诊断肺癌的试剂盒、组合物和方法。也就是说,该方法可以用于肺癌(特别是肺腺癌(AD))的在分子上定义的亚组。该方法提供了肺癌的分类,其可以针对治疗反应进行预后和预测。虽然是用于流行病学目的的有用术语,但“肺癌”可能不是指特定疾病,而是可以代表肺部、支气管和胸膜的肿瘤的异质集合。出于实际目的,肺癌通常可以分为两种组织学亚型-小细胞肺癌(SCLC)和非小细胞肺癌(NSCLC)。这些主要肿瘤类型可以以不同的频率存在,可以具有不同的解剖学位置,可以具有不同的转移偏好,可以对疗法有不同的反应,并且可能源自不同的细胞祖细胞。

“确定腺癌亚型”可以包括,例如,诊断或检测肺腺癌的存在和类型,监测疾病的进展,以及鉴定或检测指示亚型的细胞或样品。

在一个实施方案中,通过评价一个或多个受试者样品中多种分类器基因或生物标记物的表达模式或表达谱来评估肺癌状态。出于讨论的目的,术语受试者或受试者样品是指个体而不论健康和/或疾病状态。受试者可以是受试者、研究参与者、对照受试者、筛查受试者、或在本发明的上下文中从其获得和评估样品的任何其他类别的个体。因此,受试者可以被诊断患有肺腺癌(包括亚型或其等级),可以出现肺AD癌的一种或多种症状、或肺癌的易感因素(例如家族(遗传)或病史(医学)因子),可以正在经历肺癌治疗或疗法等。可替代地,关于任何上述因子或标准,受试者可以是健康的。应当理解,如本文所用的术语“健康的”与肺癌状态有关,因为术语“健康的”不能被定义为对应于任何绝对评价或状态。因此,参考任何特定疾病或疾病标准定义为健康的个体实际上可以被诊断患有任何其他一种或多种疾病,或者表现出任何其他一种或多种疾病标准(包括一种或多种其他癌症)。

如本文所用,“表达谱”或“生物标记物谱”或“基因标签”包含对应于鉴别性或分类器基因或生物标记物的表达的相对丰度、水平、存在或不存在的测量值的一个或多个值。表达谱可以来源于在肺癌诊断之前或之后的受试者,可以来源于在治疗或疗法之前或之后的一个或多个时间点从受试者收集的生物样品,可以来源于在其间没有进行治疗或疗法(例如,以监测疾病的进展或评估被诊断患有肺癌或有患肺癌风险的受试者的疾病的发展)的一个或多个时间点从受试者收集的生物样品,或者可以从健康的受试者收集。术语受试者可以与患者互换使用。患者可以是人类患者。本文提供的生物标记物谱的该一种或多种生物标记物选自表1的一种或多种生物标记物。

如本文所用,术语“确定表达水平”或“确定表达谱”或“检测表达水平”或“检测表达谱”如关于生物标记物或分类器所用,意指向样品(例如受试者或患者的样品和/或对照

样品)应用生物标记物特异性试剂(例如探针、引物或抗体)和/或方法,用于定量、半定量或定性地确定或测量一种或多种生物标记物的量(例如生物标记物多肽或mRNA(或由其衍生的cDNA)的量)。例如,生物标记物的水平可以通过许多方法确定,包括例如免疫测定,包括例如免疫组织化学、ELISA、Western印迹、免疫沉淀等,其中生物标记物检测试剂例如抗体(例如标记的抗体)特异性结合生物标记物并允许例如相对或绝对确定多肽生物标记物的量;杂交和PCR方案,其中使用探针或引物或引物组来确定核酸生物标记物的量,包括例如基于探针和基于扩增的方法,包括例如微阵列分析、RT-PCR(如定量RT-PCR(qRT-PCR))、基因表达系列分析(SAGE)、Northern印迹、数字分子条形码技术(例如纳米串(Nanostring)计数器分析)和TaqMan定量PCR试验。可以应用其他mRNA检测和定量方法,例如在福尔马林固定的石蜡包埋(FFPE)组织样品或细胞中的mRNA原位杂交。该技术目前由QuantiGene ViewRNA(Affymetrix)提供,其使用每个mRNA的探针组,该探针组特异性结合扩增系统以扩增杂交信号;可以使用标准荧光显微镜或成像系统使这些放大的信号可视化。例如,该系统可以检测和测量异质样品中的转录物水平;例如,如果样品具有存在于相同的组织切片中的正常和肿瘤细胞。如上所述,基于TaqMan探针的基因表达分析(基于PCR)也可以用于测量组织样品中的基因表达水平,并且该技术已显示可用于测量FFPE样品中的mRNA水平。简言之,基于TaqMan探针的测定利用与mRNA靶标特异性杂交的探针。该探针含有附着于各末端的淬灭剂染料和报告染料(荧光分子),并且仅在与mRNA靶标特异性杂交时才发出荧光。在扩增步骤期间,聚合酶的核酸外切酶活性导致淬灭剂和报告染料从探针上分离,并且可以发生荧光发射。记录该荧光发射并通过检测系统测量信号;将这些信号强度用于计算样品中给定转录物(基因表达)的丰度。

在一个实施方案中,与本文(例如,表1和表2)所述的基因盒或分类器基因相关的“表达谱”或“生物标记物谱”或“基因标签”可以用于在正常样品与肿瘤样品之间作出区分。在另一个实施方案中,肿瘤样品是肺腺癌(AD)。在另一个实施方案中,基于使用本文提供的方法确定的表达谱,AD可以进一步被分类为支气管型、鳞状型和巨型。使用肿瘤活组织检查组织对支气管型、鳞状型和巨型腺癌的表征已经在Hayes等人(2006)J.Clin.Oncol.24(31):5079-90中进行了描述。使用本文(例如,表1)公开的分类器基因的表达谱可以为特异性鉴定肺腺癌亚型以及为评价治疗肺腺癌的治疗功效提供有价值的分子工具。因此,本发明提供了用于对受试者的分子AD亚型进行筛查和分类的方法和用于监测肺AD的某些治疗性治疗的功效的方法。

在一些情况下,本文提供的单一分类器基因能够鉴定肺腺癌的亚型,其预测成功率为至少约70%、至少约71%、至少约72%、约73%、约74%、约75%、约76%、约77%、约78%、约79%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、高达100%。

在一些情况下,如本文提供的单一分类器基因能够确定肺腺癌亚型,其灵敏度或特异性为至少约70%、至少约71%、至少约72%、约73%、约74%、约75%、约76%、约77%、约78%、约79%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、高达100%。

本发明还涵盖能够区分使用现有方法无法检测的各种肺腺癌亚型的系统。该系统能够处理大量受试者和受试者变量例如表达谱和其他诊断标准。本文所述的方法也可以用于类似于药物基因组学(pharmacogenomics)的“药物代谢组学”(pharmacometabonomics)，例如，预测对疗法的反应。在该实施方案中，可以使用表达谱作为“反应”的证据将受试者分成“反应者”和“非反应者”，然后可以使用表达谱的特征来靶向将可能对特定治疗过程有反应的未来受试者。

表达谱可以与其他诊断方法组合使用，包括组织化学、免疫组织化学、细胞学、免疫细胞学、和视觉诊断方法，包括肺组织的组织学或形态学评价。

在本发明的各种实施方案中，将来源于受试者的表达谱与参考表达谱进行比较。“参考表达谱”可以是从在治疗或疗法之前的受试者来源的谱；可以是在特定时间点(通常在治疗或疗法之前或之后，但也可以包括在诊断肺癌之前或之后的特定时间点)从受试者样品产生的谱；或者可以来源于健康个体或来自健康个体的合并参考。参考表达谱可以是肺癌通用的，或者可以对肺腺癌的不同亚型有特异性。

可以将参考表达谱与测试表达谱进行比较。“测试表达谱”可以源自于与参考表达谱相同的受试者(除了在随后的时间点(例如，在收集参考表达谱之后的一天或多天、一周或几周、或者一个月或几个月)之外)，或者可以源自于不同的受试者。总之，可以将受试者的任何测试表达谱与来自具有TRU、PP或PI亚型的受试者的先前收集的谱进行比较。

本发明的分类器生物标记物可以包括核酸(RNA、cDNA和DNA)和蛋白质、及其变体和片段。此类生物标记物可以包括含有编码生物标记物的核酸序列的全部或部分序列的DNA，或这样的序列的互补序列。本文所述的生物标记物可以包括含有任何感兴趣核酸序列的全部或部分序列的RNA，或其在逆转录反应中体外合成获得的非天然cDNA产物。生物标记物核酸还可以包括感兴趣核酸序列的任何表达产物或其部分。生物标记物蛋白质可以是由本发明的DNA生物标记物编码或与其对应的蛋白质。生物标记物蛋白质可以包含任何生物标记物蛋白质或多肽的全部或部分氨基酸序列。生物标记物核酸可以从细胞中提取，或者可以是无细胞的或从细胞外囊泡实体(如外泌体)中提取。

“分类器生物标记物”或“生物标记物”或“分类器基因”可以是与正常或健康细胞或组织相比其在组织或细胞中表达的水平改变的任何基因或蛋白质。例如，“分类器生物标记物”或“生物标记物”或“分类器基因”可以是在特定肺腺癌亚型中其在组织或细胞中表达的水平改变的任何基因或蛋白质。检测本发明的生物标记物可以允许对特定亚型的确定。“分类器生物标记物”或“生物标记物”或“分类器基因”可以是相对于如本文提供的参考或对照被上调(例如表达增加)或被下调(例如表达降低)的那种基因或蛋白质。参考或对照可以是如本文提供的任何参考或对照。在一些实施方案中，可以将在肺腺癌的特定亚型中上调或下调的基因的表达值合并到一个基因盒中。在每个基因盒中的总体表达水平在本文中称为“表达谱”，并且用于根据肺腺癌的亚型对测试样品进行分类。然而，应理解，对本文公开的每种基因的表达的独立评价可以用于对肿瘤亚型进行分类，而无需将被上调和下调的基因分组成一个或多个基因盒。在一些情况下，如表2中所示，可以将总共48种生物标记物用于AD亚型确定。对于每种AD亚型，16种生物标记物中的8种可以是负相关基因，而8种可以是正相关基因，其可以被选择作为特定AD亚型的基因标签。

本发明的分类器生物标记物包括在肺腺癌中选择性表达的任何基因或蛋白质，如

本文上面所定义。样品生物标记物基因列于下表1或表2中。在表2中,该表的第一列表示选择用于区分末端呼吸单元型 (TRU) 的生物标记物列表。该表的中间列表示选择用于区分近端增生型 (PP) 的生物标记物列表。该表的最后一列表示选择用于区分近端炎症型 (PI) 的生物标记物列表。

表1中显示了用于肺AD亚型分型的分类器生物标记物的如本文所述的由tsat表示的相对基因表达水平。在一个实施方案中,表1中显示了用于肺腺癌亚型分型的分类器生物标记物的基因表达水平。在一个实施方案中,所有48种基因都可以用于分类AD的亚型。在一个实施方案中,前16种基因是末端呼吸单元型的选定基因标签生物标记物,其中与非TRU样品相比,基因编号1-8被上调且基因编号9-16被下调。在另一个实施方案中,基因编号17-32是对近端增生型 (PP) 有特异性的选定基因标签生物标记物,其中与非PP样品相比,基因编号17-24被上调且基因编号25-32被下调。在又另一个实施方案中,基因编号33-48是对近端炎症型 (PI) 有特异性的选定基因标签生物标记物,其中与非PI样品相比,基因编号33-40被上调且基因编号41-48被下调。

表1.肺腺癌(AD)亚型的48种分类器生物标记物的基因质心(Gene Centroid)

基因 #	基因 符号	基因 名称	末端呼吸单元 型 (TRU)	近端增生型 (PP)	近端炎症型 (PI)	GenBank登 录号*	SEQ ID NO:
1	FIGF	C-fos 诱 导 的 生 长 因 子	2.129901586	-1.173222174	-1.545843019	AY874421. 1	1
2	CTSH	组织 蛋白 酶H	1.099895637	-0.797376345	-0.531006607	NM_004390 .4	2
3	SCTR	分 泌 素 受 体	2.043898366	-1.911062476	-1.836386831	NM_002980 .2	3
4	CYP4 B1	细 胞 色 素 P450 家 族 4 亚	2.462733828	-1.447070454	-1.481195844	NM_001319 161.1	4

		家 族 B 成 员 1					
5	GPR1 16	G 蛋 白 偶 联 受 体 116	1.289460077	-0.972597916	-0.731487829	AY140958. 1	5
6	ADH1 B	醇 脱 氢 酶 1B (I 类)	2.013525076	-1.580299515	-1.094580574	NM_001286 650.1	6
7	CBX7	色 素 框 (Ch romo box) 7	0.728027298	-0.698222051	-0.243583657	NM_175709 .3	7
8	HLF	肝 白 血 病 因 子	1.479193357	-1.28826965	-1.018563422	M95585.1	8
9	CEP5 5	中 心 体 蛋 白 55	- 1.524932169	0.5743319	0.580921528	NM_018131 .4	9
10	TPX2	Tpx2 , 微 管 相 关	- 1.704080763	0.587761579	0.583674937	NM_012112 .4	10
11	BUB1 B	BUB 1 有 丝 分	- 1.531514951	0.769199954	0.543731288	AF107297.1	11

		裂 检 查 点 丝 氨 酸/苏 氨 酸 激 酶 B				
12	KIF4 A	驱动 蛋白 家族 成员 4A	- 1.794045266	0.570328759	0.599399471	NM_012310 .4 12
13	CCNB 2	细 胞 周 期 素B2	- 1.442466223	0.602807712	0.526093335	NM_004701 .3 13
14	KIF14	驱动 蛋白 家族 成员 14	- 1.66445145	0.762295222	0.543132477	NM_014875 .2 14
15	MEL K	母 体 胚 胎 亮 氨 酸 拉 链 激 酶	- 1.685012297	0.584181432	0.694064307	NM_014791 .3 15
16	KIF11	驱动 蛋白 家族 成员 11	- 1.183768087	0.693181955	0.481955763	NM_004523 .3 16

17	FGL1	纤 维 蛋 白 原 样 1	- 0.978882607	4.89751413	-1.958269455	NM_004467 .3	17
18	PBK	PDZ 结 合 激 酶	- 1.407694417	1.278522857	0.404652088	NM_018492 .3	18
19	HSPD1	热 休 克 蛋 白 家 族 D (Hs p60) 成 员 1	热 休 克 蛋 白 家 族 D (Hs p60) 成 员 - 0.469703958	0.624572377	0.111400174	NM_002156 .4	19
20	TDG	胸 腺 嘧 啶 DNA 糖 基 化 酶	- 0.351189471	0.60348929	0.076442589	NM_003211 .4	20
21	PRC1	胞 质 分 裂 蛋 白 调 节 剂1	- 1.159074285	0.797575854	0.461100041	NM_003981 .3	21
22	DUSP4	双 特 异 性 磷 酸 酶4	- 0.704273045	1.933259798	-0.283343923	NM_001394 .6	22
23	GTPBP4	GTP 结 合	- 0.467281005	0.543583167	0.038904486	NM_012341 .2	23

		蛋白 4					
24	ZWIN T	ZW1 0 相 互作 用动 粒蛋 白	-			NM_007057 .3	24
25	TLR2	Toll 样受 体2	0.672774085	-1.389004155	0.098176794	NM_001318 787.1	25
26	CD74	CD74 分子	0.689011729	-1.365243826	0.239872217	NM_001025 159.2	26
27	HLA- DPB1	主要 组织 相容 性复 合 物, II 类, DP β 1	0.70548523	-1.431001558	0.157288388	M83664.1	27
28	HLA- DPA1	主要 组织 相容 性复 合 物, II 类, DP α 1	0.620746458	-1.622212879	0.206805676	NM_033554 .3	28
29	HLA- DRA	主要 组织	0.47615106	-1.517000712	0.209882138	NM_019111 .4	29

		相容性复合物, II类, DR α				
30	ITGB2	整合素亚基 β 2	0.227015125	-1.489015066	0.473986644	NM_000211.4
31	FAS	Fas细胞表面死亡受体	0.120924174	-1.244937359	0.608312102	KM114217.1
32	HLA-DRB1	主要组织相容性复合物, II类, DR β 1	0.561088415	-1.639812592	0.272965507	NM_002124.3
33	PLAU	纤溶酶原激活物, 尿激酶	0.723116671	-0.71054832	1.628730403	NM_002658.4
34	GBP1	鸟苷酸结合蛋	0.302372654	-0.688857626	1.204326606	NM_002053.2

		白1					
35	DSE	硫酸 皮肤 素差 向异 构酶	-	0.101374419	-0.602077696	0.748133278	NM_013352 .3 35
36	CCDC 109B	含 109B 的卷 曲螺 旋结 构域	-	0.13855818	-0.703783616	0.7964386	BC002633.2 36
37	TGFB I	转化 生长 因子 β 诱 导的	-	0.328357044	-0.746331889	1.164873128	NM_000358 .2 37
38	CXCL 10	C-X- C 基 序趋 化因 子配 体10	-	0.434345777	-0.62067894	1.70756508	NM_001565 .3 38
39	LGAL S1	凝集 素， 半乳 糖昔 结合 可溶 蛋白 1	-	0.291230377	-0.549722715	0.957730776	NM_002305 .3 39

40	TUBB 6	微管蛋白 β 6 V类	- 0.153163739	-0.328431543	0.781293298	NM_032525 .2	40
41	GJB1	间隙连接蛋白 β 1	1.567852415	0.672938467	-3.61601989	NM_001097 642.2	41
42	RAP1 GAP	RAP1 GTP 酶活化蛋白	1.019990653	0.138302482	-1.426817837	NM_001145 658.1	42
43	CACN A2D2	电压门控钙通道辅助亚基 α 2 δ 2	1.610819757	-0.126189977	-2.357279793	NM_001005 505.2	43
44	SELE NBP1	硒结合蛋白1	1.0475958	-0.331350331	-1.209058454	NM_003944 .3	44
45	TFCP 2L1	转录因子CP2样1	0.218606218	0.952552471	-1.320932951	NM_014553 .2	45
46	SORB S2	含Sorbin和SH3	0.603086366	0.462888705	-1.412139816	NM_001270 771.1	46

		结 构 域 的 蛋 白 2					
47	UNC1 3B	Unc- 13 同 源 物 B	0.293706669	0.418115853	-0.978505828	NM_006377 .3	47
48	TACC 2	含 转 化 酸 性 卷 曲 螺 旋 的 蛋 白 2	0.206302979	0.928437713	-0.822332116	AF220152.2	48

*每个GenBank登录号是所列基因的代表性或示例性GenBank登录号,并且出于所有目的将其通过引用以其整体并入本文。此外,每个列出的代表性或示例性登录号不应解释为将权利要求限制为特定登录号。

表2.针对末端呼吸单元型、近端增生型和近端炎症型选择的分类器生物标记物

末端呼吸 单元型 (TRU)	近端增生 型 (PP)	近端炎症型 (PI)
CEP55	TLR2	GJB1
TPX2	CD74	RAP1GAP
BUB1B	HLA-DPB1	CACNA2D2
KIF4A	HLA-DPA1	SELENBP1
CCNB2	HLA-DRA	TFCP2L1
KIF14	ITGB2	SORBS2
MELK	FAS	UNC13B
KIF11	HLA-DRB1	TACC2

HLF	ZWINT	TUBB6
CBX7	GTPBP4	LGALS1
ADH1B	DUSP4	CXCL10
GPR116	PRC1	TGFBI
CYP4B1	TDG	CCDC109B
SCTR	HSPD1	DSE
CTSH	PBK	GBP1
FIGF	FGL1	PLAU

诊断用途

在一个实施方案中,本文提供的方法和组合物允许用比本领域已知的分子AD亚型分型方法所需的基因更少的基因来区分腺癌的三种亚型:(1)末端呼吸单元型(TRU),以前称为支气管型;(2)近端增生型(PP),以前称为巨型;以及(3)近端炎症型(PI),以前称为鳞状型。

通常,将本文提供的方法用于将肺癌样品分类为特定的肺癌亚型(例如腺癌的亚型)。在一个实施方案中,该方法包括测量、检测或确定任何可公开获得的肺AD表达数据集的至少一种分类器生物标记物的表达水平。在一个实施方案中,该方法包括检测或确定从患者或受试者获得的肺癌样品中表1的至少一种分类器生物标记物的表达水平。用于本文所述的检测或区分方法的肺癌样品可以是先前确定或诊断为腺癌样品的样品。先前的诊断可以是基于组织学分析。组织学分析可以由一名或多名病理学家进行。

在一个实施方案中,测量或检测步骤是在核酸水平上,其通过在适合于RNA-seq、RT-PCR或杂交的条件下,用与该至少一种分类器生物标记物(例如表1的分类器生物标记物)的cDNA分子的部分基本上互补的寡核苷酸进行RNA-seq、逆转录酶聚合酶链反应(RT-PCR)或杂交测定,并且基于检测步骤获得该至少一种分类器生物标记物的表达水平。然后将该至少一种分类器生物标记物的表达水平与来自至少一个样品训练集的该至少一种分类器生物标记物(例如表1的分类器生物标记物)的参考表达水平进行比较。该至少一个样品训练集可以包含,(i)来自过表达该至少一种生物标记物的样品的至少一种生物标记物的表达水平,(ii)来自参考鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型)样品的表达水平,或(iii)来自无腺癌的肺部样品的表达水平,并且将肺组织样品分类为鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型)亚型。然后可以基于比较步骤的结果将肺癌样品分类为支气管型、鳞状型或巨型腺癌亚型。在一个实施方案中,比较步骤可以包括应用统计算法,该统计算法包括确定从肺部组织或癌症样品获得的表达数据与来自该至少一个训练集的表达数据之间的相关性;并且基于该统计算法的结果将肺部组织或癌症样品分类为鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型)亚型。

在一个实施方案中,该方法包括在从患者获得的肺癌样品中在核酸水平上探测本文提供的至少一种分类器生物标记物(例如表1的分类器生物标记物)的水平。肺癌样品可

以是先前确定或诊断为腺癌样品的样品。先前的诊断可以是基于组织学分析。组织学分析可以由一名或多名病理学家进行。在一个实施方案中,探测步骤包括在适合于一种或多种寡核苷酸与其互补序列或基本互补序列杂交的条件下将样品与该一种或多种寡核苷酸混合,该寡核苷酸与本文提供的该至少一种分类器生物标记物(例如表1的分类器生物标记物)的cDNA分子的部分基本上互补;检测该一种或多种寡核苷酸与其互补序列或基本互补序列之间是否发生杂交;并且基于检测步骤获得该至少一种分类器生物标记物的杂交值。然后将该至少一种分类器生物标记物的杂交值与来自至少一个样品训练集的一个或多个参考杂交值进行比较。例如,该至少一个样品训练集包含来自参考TRU腺癌、PP腺癌和/或PI腺癌样品的杂交值。基于比较步骤的结果,将肺癌样品分类为例如TRU、PP或PI。

肺组织样品可以是从人受试者或患者分离的任何样品。例如,在一个实施方案中,对包埋在石蜡中的肺活组织检查上进行分析。在一个实施方案中,样品可以是新鲜冷冻的肺组织样品。在另一个实施方案中,样品可以是从患者获得的体液。体液可以是血液或其部分(即,血清、血浆)、尿液、唾液、痰液或脑脊髓液(CSF)。样品可以含有用于在本文提供的方法中使用的核酸的细胞来源以及细胞外来源。细胞外来源可以是无细胞的DNA和/或外泌体。在一个实施方案中,样品可以是细胞沉淀物或洗涤液。本发明的这个方面提供了一种通过精确鉴定主要组织学类型(甚至从小的活组织检查中)来改善当前诊断的手段。本发明的方法(包括RT-PCR方法)是灵敏的、精确的并且具有多分析物能力以便用于石蜡包埋的样品。参见例如,Cronin等人(2004)Am.J Pathol.164(1):35-42,通过引用并入本文。

在光学显微镜评价之前,福尔马林固定和石蜡中的组织包埋是组织处理的通用方法。福尔马林固定的石蜡包埋的(FFPE)样本提供的主要优点是保留组织切片中的细胞和结构学形态细节。(Fox等人(1985)J Histochem Cytochem 33:845-853)。在其中处理活组织检查样本的标准缓冲福尔马林固定剂典型地是含有37%甲醛和10%-15%甲醇的水性溶液。甲醛是一种高反应性偶极化合物,其导致在体外形成蛋白质-核酸和蛋白质-蛋白质交联(Clark等人(1986)J Histochem Cytochem 34:1509-1512;McGhee和von Hippel(1975)Biochemistry 14:1281-1296,各自通过引用并入本文)。

在一个实施方案中,本文使用的样品获自个体,并且包含福尔马林固定的石蜡包埋的(FFPE)组织。然而,其他组织和样品类型适合于在本文中使用。在一个实施方案中,其他组织和样品类型可以是新鲜冷冻的组织、洗涤液体或细胞沉淀物等。在一个实施方案中,样品可以是从个体获得的体液。体液可以是血液或其部分(例如,血清、血浆)、尿液、痰液、唾液或脑脊髓液(CSF)。如本文提供的生物标记物核酸可以从细胞中提取,或者可以是无细胞的或从细胞外囊泡实体(如外泌体)中提取。

用于从FFPE组织中分离RNA的方法在本领域中是已知的。在一个实施方案中,可以从FFPE组织中分离总RNA,如由Bibikova等人(2004)American Journal of Pathology 165:1799-1807所述,通过引用并入本文。同样,可以使用高纯的RNA石蜡试剂盒(High Pure RNA Paraffin Kit)(罗氏公司(Roche))。通过二甲苯萃取去除石蜡,随后用乙醇洗涤。可以使用MasterPure纯化试剂盒(Epicenter, Madison, Wis.)从切片组织块中分离RNA;包括DNA酶I处理步骤。可以根据供应商的说明书(Invitrogen Life Technologies, Carlsbad, Calif.)使用Trizol试剂从冷冻样品中提取RNA。可以将具有可测量的残留基因组DNA的样品重新进行DNaseI处理并测定DNA污染。可以根据制造商的方案进行所有纯化、DNA酶处理

和其他步骤。在总RNA分离之后,可以将样品储存在-80°C下直至使用。

用于mRNA提取的一般方法是本领域熟知的,并且公开于分子生物学的标准教科书中,包括Ausubel等人,编辑,Current Protocols in Molecular Biology,John Wiley& Sons, New York 1987-1999。从石蜡包埋的组织中提取RNA的方法公开于例如Rupp和Locker (Lab Invest. 56:A67, 1987) 以及De Andres等人(Biotechniques 18:42-44, 1995) 中。具体地,可以使用来自商业制造商例如Qiagen(Valencia, Calif.)的纯化试剂盒、缓冲液组和蛋白酶根据制造商的说明书进行RNA分离。例如,可以使用Qiagen RNeasy微型柱分离来自培养物中的细胞的总RNA。其他可商购的RNA分离试剂盒包括MasterPureTM。完整的DNA和RNA纯化试剂盒(Complete DNA and RNA Purification Kit) (Epicentre, Madison, Wis.) 以及石蜡块RNA分离试剂盒(Paraffin Block RNA Isolation Kit) (Ambion, Austin, Tex.)。可以例如使用RNA Stat-60(Tel-Test, Friendswood, Tex.)分离来自组织样品的总RNA。可以例如通过氯化铯密度梯度离心分离从肿瘤制备的RNA。另外,可以使用本领域技术人员熟知的技术(例如Chomczynski的单步RNA分离方法(美国专利号4,843,155,出于所有目的将该文献通过引用以其整体并入)容易地处理大量的组织样品。

在一个实施方案中,样品包含从肺组织样品(例如腺癌样品)收获的细胞。可以使用本领域已知的标准技术从生物样品中收获细胞。例如,在一个实施方案中,通过离心细胞样品并重悬浮沉淀的细胞来收获细胞。可以将细胞重悬浮于缓冲溶液如磷酸盐缓冲盐水(PBS)中。在将细胞悬浮液离心获得细胞沉淀之后,可以将细胞裂解以提取核酸,例如信使RNA。所有获得自受试者的样品,包括经受任何种类的进一步处理的那些,均应认为是获得自该受试者。

在一个实施方案中,在检测本文所列出的生物标记物组合的生物标记物水平之前,将样品进一步处理。例如,在细胞或组织样品中的mRNA可以与该样品的其他组分分开。可以将样品浓缩和/或纯化以分离处于其非天然状态的mRNA,因为mRNA不是处于其天然环境中。例如,研究已经表明体内mRNA的较高级结构不同于相同序列的体外结构(参见例如, Rouskin等人(2014) .Nature 505, 第701-705页,出于所有目的以其整体并入本文)。

在一个实施方案中,来自样品的mRNA与合成DNA探针杂交,该探针在一些实施方案中包括检测部分(例如,可检测标记、捕获序列、条形码报告序列)。因此,在这些实施方案中,最终制得非天然mRNA-cDNA复合物并将其用于检测生物标记物。在另一个实施方案中,将来自样品的mRNA直接用可检测标记(例如荧光团)进行标记。在另外的实施方案中,非天然标记的mRNA分子杂交至cDNA探针,并且检测复合物。

在一个实施方案中,一旦从样品获得mRNA,则将其在杂交反应中转化为互补DNA(cDNA),或者与一种或多种cDNA探针一起用于杂交反应中。cDNA在体内不存在,并因此是非天然分子。另外,cDNA-mRNA杂交体是合成的并且在体内不存在。除了cDNA在体内不存在之外,cDNA必然不同于mRNA,因为它包括脱氧核糖核酸而非核糖核酸。然后例如通过聚合酶链反应(PCR)或本领域普通技术人员已知的其他扩增方法将cDNA扩增。例如,可以采用的其他扩增方法包括连接酶链反应(LCR) (Wu和Wallace, Genomics, 4:560(1989) , Landegren等人, Science, 241:1077 (1988) ,出于所有目的将该文献通过引用以其整体并入) ;转录扩增(Kwoh等人, Proc.Natl.Acad.Sci.USA, 86:1173(1989) ,出于所有目的将该文献通过引用以其整体并入) ;自持序列复制(self-sustained sequence replication) (Guatelli等人,

Proc. Nat. Acad. Sci. USA, 87: 1874 (1990), 出于所有目的将该文献通过引用以其整体并入), 出于所有目的将该文献通过引用以其整体并入; 以及基于核酸的序列扩增(NASBA)。用于选择PCR扩增引物的指南对于本领域普通技术人员而言是已知的。参见例如, McPherson 等人, PCR Basics: From Background to Bench, Springer-Verlag, 2000, 出于所有目的将该文献通过引用以其整体并入。这种扩增反应的产物(即扩增的cDNA)必然也是非天然产物。首先, 如上所述, cDNA是非天然分子。其次, 在PCR的情况下, 扩增过程用于针对起始材料的每个单独cDNA分子产生数以亿计的cDNA拷贝。生成的拷贝数目与体内存在的mRNA的拷贝数目相差甚远。

在一个实施方案中, 将cDNA用将另外的DNA序列(例如, 衔接子、报告子、捕获序列或部分、条形码)引入片段上的引物(例如, 使用衔接子特异性引物)扩增, 或者将mRNA或cDNA生物标记物序列直接与包含该另外的序列(例如, 衔接子、报告子、捕获序列或部分、条形码)的cDNA探针杂交。因此扩增和/或mRNA与cDNA探针的杂交用于通过引入另外的序列并形成非天然杂交体而从非天然单链cDNA或mRNA产生非天然双链分子。另外, 如本领域普通技术人员已知的, 扩增程序具有与其相关的错误率。因此, 扩增将另外的修饰引入cDNA分子中。在一个实施方案中, 在用衔接子特异性引物扩增期间, 可检测标记(例如荧光团)被添加到单链cDNA分子上。因此, 扩增还用于产生天然不存在的DNA复合物, 至少因为(i) cDNA在体内不存在, (i) 衔接子序列被添加到cDNA分子的末端以得到体内不存在的DNA序列, (ii) 与扩增相关的错误率进一步产生体内不存在的DNA序列, (iii) 与天然存在的那些相比cDNA分子的结构不同, 以及(iv) 将可检测标记化学添加到cDNA分子上。

在一些实施方案中, 感兴趣的生物标记物的表达是经由检测非天然cDNA分子在核酸水平上检测的。

在一些实施方案中, 用于肺癌AD亚型分型的方法包括检测分类器生物标记物组的表达水平。在一些实施方案中, 检测包括在核酸水平或蛋白质水平上的表1的所有分类器生物标记物。在另一个实施方案中, 检测表1的分类器生物标记物的单个或子集(例如, 从约8种至约16种)。例如, 在一个实施方案中, 在确定肺癌AD亚型的方法中检测了表1中的从约5种至约10种、从约5种至约15种、从约5种至约20种、从约5种至约25种、从约5种至约30种、从约5种至约35种、从约5种至约40种、从约5种至约45种、从约5种至约48种生物标记物。在另一个实施方案中, 在确定肺癌亚型的方法中检测来自表1的每种生物标记物。在另一个实施方案中, 选择来自表1的16种生物标记物作为特定肺癌AD亚型的基因标签。

例如使用对分类器生物标记物有特异性的引物和/或探针、和/或类似物, 可以通过任何合适的技术(包括但不限于RNA-seq、逆转录酶聚合酶链反应(RT-PCR)、微阵列杂交测定、或另一种杂交测定(例如纳米串测定))来进行检测。在一些情况下, 用于扩增方法(例如, RT-PCR或qRT-PCR)的引物是适用于结合本文提供的分类器基因(例如表1中列出的分类器生物标记物)的任何正向和反向引物。

本文所述的生物标记物包括含有任何感兴趣核酸序列的全部或部分序列的RNA, 或其在逆转录反应中体外合成获得的非天然cDNA产物。术语“片段”旨在指代通常包含至少10个、15个、20个、50个、75个、100个、150个、200个、250个、300个、350个、400个、450个、500个、550个、600个、650个、700个、800个、900个、1,000个、1,200个、或1,500个连续核苷酸或高达存在于本文公开的全长生物标记物多核苷酸中的核苷酸数目的多核苷酸的一部分。生

物标记物多核苷酸的片段将通常编码至少15个、25个、30个、50个、100个、150个、200个、或250个连续氨基酸,或高达存在于本发明的全长生物标记物蛋白中的氨基酸总数。

在一些实施方案中,过表达(例如RNA转录物或其表达产物的过表达)通过归一化至参考RNA转录物或其表达产物的水平来确定,该参考RNA转录物或其表达产物可以是样品中的所有测量的转录物(或其产物)或特定的RNA转录物(或其非天然cDNA产物)的参考集。进行归一化以校正所测定的RNA或cDNA的量的差异以及所使用的RNA或cDNA的质量的变化两者,或使这两者归一化。因此,测定典型地测量并结合某些归一化基因(包括众所周知的管家基因,如例如GAPDH和/或β-肌动蛋白)的表达。可替代地,归一化可以是基于所有所测定的生物标记物或其大的子集的平均或中值信号(全局归一化方法)。

可以将分离的mRNA用于杂交或扩增测定,其包括但不限于Southern或Northern分析、PCR分析和探针阵列、纳米串测定。检测mRNA水平的一种方法涉及使分离的mRNA或合成的cDNA与核酸分子(探针)接触,该核酸分子可以与被检测基因编码的mRNA杂交。核酸探针可以是例如cDNA或其部分,例如长度为至少7个、15个、30个、50个、100个、250个、或500个核苷酸的寡核苷酸并且足以在严格条件下与本发明的非天然cDNA或mRNA生物标记物特异性地杂交。

如上所解释,在一个实施方案中,一旦从样品中获得mRNA,就在杂交反应中将其转化为互补DNA(cDNA)。可以用包含与特定mRNA的一部分互补的序列的寡核苷酸或引物进行mRNA向cDNA的转化。可以用包含随机序列的寡核苷酸或引物进行mRNA向cDNA的转化。可以用包含与mRNA的poly(A)尾互补的序列的寡核苷酸或引物进行mRNA向cDNA的转化。cDNA在体内不存在,并因此是非天然分子。在另外的实施方案中,然后例如通过聚合酶链反应(PCR)或本领域普通技术人员已知的其他扩增方法将cDNA扩增。可以用包含与本文提供的分类器基因(例如表1中的分类器生物标记物)的至少一部分互补的序列的正向和/或反向引物进行PCR。这种扩增反应的产物(即扩增的cDNA)必然是非天然产物。如上所述,cDNA是非天然分子。其次,在PCR的情况下,扩增过程用于针对起始材料的每个单独cDNA分子产生数以亿计的cDNA拷贝。生成的拷贝数目与体内存在的mRNA的拷贝数目相差甚远。

在一个实施方案中,用将另外的DNA序列(衔接子序列)引入片段上的引物(使用衔接子特异性引物)扩增cDNA。衔接子序列可以是尾部,其中该尾部序列不与cDNA互补。例如,包含与本文提供的分类器基因(例如来自表1的分类器生物标记物)的至少一部分互补的序列的正向和/或反向引物可以包含尾部序列。因此,通过将条形码、衔接子和/或报告序列引入已经非天然的cDNA上,扩增用于从非天然单链cDNA产生非天然双链分子。在一个实施方案中,在用衔接子特异性引物扩增期间,可检测标记(例如荧光团)被添加到单链cDNA分子上。因此,扩增还用于产生天然不存在的DNA复合物,至少因为(i) cDNA在体内不存在,(ii)衔接子序列被添加到cDNA分子的末端以得到体内不存在的DNA序列,(iii)与扩增相关的错误率进一步产生体内不存在的DNA序列,(iv)与天然存在的那些相比cDNA分子的结构不同,以及(v)将可检测标记化学添加到cDNA分子上。

在一个实施方案中,经由与探针杂交,例如经由微阵列,将合成的cDNA(例如,扩增的cDNA)固定在固体表面上。在另一个实施方案中,经由引入与cDNA产物杂交的荧光探针,经由实时聚合酶链反应(PCR)检测cDNA产物。例如,在一个实施方案中,通过定量荧光RT-PCR(例如,使用TaqMan®探针)评估生物标记物检测。对于PCR分析,本领域中可获得用于

确定用于在分析中使用的引物序列的众所周知的方法。

在一个实施方案中,本文提供的生物标记物通过使用捕获探针和/或报告探针的杂交反应来检测。例如,杂交探针是衍生到固体表面(例如珠粒、玻璃或硅基底)上的探针。在另一个实施方案中,捕获探针存在于溶液中并与患者样品混合,随后例如通过生物素-亲和素相互作用将杂交产物附着于表面(例如,其中生物素是捕获探针的一部分,并且亲和素在表面上)。在一个实施方案中,杂交测定使用捕获探针和报告探针两者。报告探针可以与捕获探针或生物标记物核酸中的任一者杂交。然后对报告探针进行计数和检测,以确定样品中一种或多种生物标记物的水平。在一个实施方案中,捕获探针和/或报告探针含有可检测标记,和/或允许官能化到表面上的基团。

例如,nCounter基因分析系统(参见例如,Geiss等人(2008)Nat.Biotechnol.26,第317-325页,出于所有目的,通过引用以其整体并入)适用于与本文提供的方法一起使用。

在美国专利号7,473,767和8,492,094(出于所有目的将其公开内容通过引用以其整体并入)中描述的杂交测定适用于与本文提供的方法一起使用,即,适用于检测本文所述的生物标记物和生物标记物组合。

可以使用膜印迹(例如用于杂交分析,如Northern印迹、Southern印迹、斑点印迹等)、或微孔、样品管、凝胶、珠粒或纤维(或包含结合核酸的任何固体支持物)来监测生物标记物水平。参见例如,美国专利号5,770,722、5,874,219、5,744,305、5,677,195和5,445,934,各自通过引用以其整体并入。

在一个实施方案中,将微阵列用于检测生物标记物水平。由于不同实验之间的再现性,微阵列特别适合于此目的。DNA微阵列提供了一种同时测量大量基因的表达水平的方法。每个阵列由附着于固体支持物上的可再现模式的捕获探针组成。标记的RNA或DNA与阵列上的互补探针杂交,并然后通过激光扫描阵列上每个探针的杂交强度来检测,并将其转化为代表相对基因表达水平的定量值。参见例如,美国专利号6,040,138、5,800,992和6,020,135、6,033,860和6,344,316各自通过引用以其整体并入。高密度寡核苷酸阵列特别适用于确定样品中大量RNA的基因表达谱。

使用机械合成方法合成这些阵列的技术描述于例如美国专利号5,384,261中。尽管通常使用平面阵列表面,但是阵列可以制造在几乎任何形状的表面或甚至多个表面上。阵列可以是在珠粒、凝胶、聚合物表面、纤维(例如光纤)、玻璃或任何其他适当的基底上的核酸(或肽)。参见例如,美国专利号5,770,358、5,789,162、5,708,153、6,040,193和5,800,992,各自通过引用以其整体并入。阵列可以以这样一种方式包装,以致于允许全包式设备的诊断或其他操作。参见例如,美国专利号5,856,174和5,922,591,各自通过引用以其整体并入。

在一个实施方案中,将基因表达系列分析(SAGE)用于本文所述的方法中。SAGE是允许同时和定量分析大量基因转录物的方法,无需为每种转录物提供单独的杂交探针。首先,生成短序列标签(约10-14bp),其中含有足够的信息来独特地鉴定转录物,条件是该标签是从每个转录物内的独特位置获得的。然后,许多转录物连接在一起形成长系列分子,可以对其进行测序,同时揭示多个标签的身份。通过确定单个标签的丰度并鉴定与每个标签对应的基因,可以定量评价任何转录物群体的表达模式。参见Velculescu等人Science 270:484-87,1995;Cell 88:243-51,1997,将该文献通过引用以其整体并入。

在核酸水平上的另外的生物标记物水平分析方法是使用测序方法,例如RNAseq、下一代测序和大规模平行标签测序(MPSS),如Brenner等人(Nat.Biotech.18:630-34,2000,将该文献通过引用以其整体并入)所描述的。这是在单独的5μm直径微珠上将基于非凝胶的标签测序与数百万个模板的体外克隆组合的一种测序方法。首先,通过体外克隆构建DNA模板的微珠文库。随后是在流动池中以高密度(典型地大于3.0X 10⁶个微珠/cm²)组装含有模板的微珠的平面阵列。使用不需要DNA片段分离的基于荧光的标签测序方法,同时分析每个微珠上的所克隆模板的自由端。已经显示该方法在一次操作中同时且准确地提供来自酵母cDNA文库的数十万个基因标签序列。

在核酸水平上的另一种生物标记物水平分析方法是使用扩增方法,如例如RT-PCR或定量RT-PCR(qRT-PCR)。用于确定样品中生物标记物mRNA水平的方法可以包括:核酸扩增的过程,例如通过RT-PCR(Mullis,1987,美国专利号4,683,202中所述的实验性实施方案)、连接酶链反应(Barany(1991)Proc.Natl.Acad.Sci.USA 88:189-193)、自持序列复制(Guatelli等人(1990)Proc.Natl.Acad.Sci.USA 87:1874-1878)、转录扩增系统(Kwoh等人(1989)Proc.Natl.Acad.Sci.USA 86:1173-1177)、Q_B复制酶(Lizardi等人(1988)Bio/Technology 6:1197)、滚环复制(Lizardi等人,美国专利号5,854,033)或任何其他核酸扩增方法;随后使用本领域技术人员熟知的技术检测所扩增的分子。许多不同的PCR或qRT-PCR方案是本领域已知的,并且可以使用目前描述的组合物直接应用于或适用于检测和/或定量样品中区别基因的表达。参见例如,Fan等人(2004)Genome Res.14:878-885,通过引用并入本文。通常,在PCR中,通过与至少一种寡核苷酸引物或一对寡核苷酸引物反应来扩增靶多核苷酸序列。一种或多种引物与靶核酸的互补区杂交,并且DNA聚合酶延伸该一种或多种引物以扩增靶序列。在足以提供基于聚合酶的核酸扩增产物的条件下,一种大小的核酸片段为主反应产物(作为扩增产物的靶多核苷酸序列)。重复扩增循环以增加单个靶多核苷酸序列的浓度。反应可以在通常用于PCR的任何热循环仪中进行。

在一些情况下定量RT-PCR(qRT-PCR)(也称为实时RT-PCR)是优选的,因为它不仅提供定量测量,还提供减少的时间和污染。如本文所用,“定量PCR”(或“实时qRT-PCR”)是指当PCR扩增发生时在不需要对反应产物重复取样的情况下直接监测PCR扩增的进程。在定量PCR中,可以在产生反应产物时通过信号传导机制(例如荧光)监测反应产物,并在信号上升到背景水平之后但在反应达到平稳之前跟踪反应产物。达到可检测或“阈值”荧光水平所需的循环次数直接随着PCR过程开始时可扩增靶标的浓度而变化,从而能够测量信号强度以实时提供样品中靶核酸的量的量度。可以将DNA结合染料(例如,SYBR green)或标记探针用于检测通过PCR扩增产生的延伸产物。可以使用利用包含本发明序列的标记探针的任何探针形式。

免疫组织化学方法也适用于检测本发明的生物标记物的水平。可以将样品冷冻用于后续制备或立即置于固定剂溶液中。可以组织样品通过用试剂如福尔马林、戊二醛、甲醇等处理来固定,并包埋在石蜡中。从福尔马林固定的、石蜡包埋的组织样品制备用于免疫组织化学分析的切片的方法是本领域熟知的。

在一个实施方案中,将本文提供的生物标记物(例如表1的分类器生物标记物(或其子集,例如8种至16种、16种至32种、或32种至48种生物标记物))的水平针对所有RNA转录物或其非天然cDNA表达产物、或样品中蛋白质产物的表达水平、或者RNA转录物的参考集或

其非天然cDNA表达产物的参考集、或其样品中蛋白质产物的参考集的表达水平进行归一化。

在一个实施方案中,可以使用本文提供的分类器基因(例如表1中列出的分类器生物标记物)中的一种或多种的蛋白质表达水平来评价肺腺癌亚型。可以使用免疫学检测方法来测量蛋白质表达水平。可以用于本文中的免疫学检测方法包括但不限于使用如下技术的竞争性和非竞争性测定系统,例如Western印迹、放射免疫测定、ELISA(酶联免疫吸附测定)、“三明治”免疫测定、免疫沉淀测定、沉淀素反应、凝胶扩散沉淀素反应、免疫扩散测定、凝集测定、补体固定测定、免疫放射测定、荧光免疫测定、蛋白A免疫测定等。此类测定是常规的并且是本领域熟知的(参见例如,Ausubel等人,编辑,1994,Current Protocols in Molecular Biology,第I卷,John Wiley&Sons,Inc.,New York,将其通过引用以其整体并入)。

在一个实施方案中,利用对生物标记物蛋白有特异性的抗体来检测身体样品中生物标记物蛋白的表达。该方法包括从患者或受试者获得身体样品,使身体样品与针对在肺癌细胞中选择性表达的生物标记物的至少一种抗体接触,并且检测抗体结合以确定生物标记物是否在患者样品中表达。本发明的优选方面提供了用于诊断肺癌亚型的免疫细胞化学技术。本领域技术人员将认识到,本文下面描述的免疫细胞化学方法可以手动进行或以自动方式进行。

如贯穿全文所提供的,本文所述的方法提供了用于确定患者的肺癌AD亚型的方法。一旦例如通过测量非天然cDNA生物标记物水平或非天然mRNA-cDNA生物标记物复合物确定了生物标记物水平,就例如通过使用统计学方法或直接比较检测水平将生物标记物水平与参考值或参考样品进行比较,以确定肺癌分子AD亚型。基于该比较,将患者的肺癌样品进行AD分类,例如分类为TRU、PP或PI。

在一个实施方案中,将本文提供的该至少一种分类器生物标记物(例如表1的分类器生物标记物)的表达水平值与来自至少一个样品训练集的一个或多个参考表达水平值进行比较,其中该至少一个样品训练集包含来自一个或多个参考样品的表达水平值。在另外的实施方案中,该至少一个样品训练集包含来自近端炎症型(鳞状型)、近端增生型(巨型)、末端呼吸单元型(支气管型)样品或其组合的本文提供的至少一种分类器生物标记物(例如表1的分类器生物标记物)的表达水平值。

在单独的实施方案中,将本文提供的该至少一种分类器生物标记物(例如表1的分类器生物标记物)的杂交值与来自至少一个样品训练集的一个或多个参考杂交值进行比较,其中该至少一个样品训练集包含来自一个或多个参考样品的杂交值。在另外的实施方案中,该至少一个样品训练集包含来自近端炎症型(鳞状型)、近端增生型(巨型)、末端呼吸单元型(支气管型)样品或其组合的本文提供的至少一种分类器生物标记物(例如表1的分类器生物标记物)的杂交值。本文提供了用于将检测到的生物标记物水平与参考值和/或参考样品进行比较的方法。基于该比较,在一个实施方案中,获得从受试者样品获得的生物标记物水平与参考值之间的相关性。然后进行肺癌AD亚型的评估。

可以使用各种统计方法来辅助从患者获得的生物标记物水平与例如来自至少一个样品训练集的参考生物标记物水平的比较。

在一个实施方案中,采用监督模式识别方法。监督模式识别方法的例子可以包括

但不限于最近质心方法(Dabney (2005) *Bioinformatics* 21 (22) :4148-4154以及Tibshirani等人 (2002) *Proc. Natl. Acad. Sci. USA* 99 (10) :6576-6572) ;类别分析的软独立建模(SIMCA) (参见例如,Wold,1976) ;偏最小二乘分析(PLS) (参见例如,Wold,1966;Joreskog,1982;Frank,1984;Bro,R.,1997) ;线性判别分析(LDA) (参见例如,Nillson,1965) ;K-最近邻分析(KNN) (参见例如,Brown等人,1996) ;人工神经网络(ANN) (参见例如,Wasserman,1989;Anker等人,1992;Hare,1994) ;概率神经网络(PNN) (参见例如,Parzen,1962;Bishop,1995;Speckt,1990;Broomhead等人,1988;Patterson,1996) ;规则归纳法(RI) (参见例如,Quinlan,1986) ;以及贝叶斯方法(参见例如,Brethorst,1990a,1990b,1988)。在一个实施方案中,基于基因表达数据用于鉴定肿瘤亚型的分类器是基于质心的方法,其描述于Mullins等人(2007) *Clin Chem.* 53 (7) :1273-9(将其各自通过引用以其整体并入本文)中。

在其他实施方案中,采用无监督训练方法,并因此,不使用训练集。

再次参考用于监督学习方法的样品训练集,在一些实施方案中,一个或多个样品训练集可以包括来自腺癌样品的多种或所有分类器生物标记物(例如,表1的所有分类器生物标记物)的表达数据。该多种分类器生物标记物可以包含表1的至少两种分类器生物标记物、至少8种分类器生物标记物、至少16种分类器生物标记物、至少24种分类器生物标记物、至少32种分类器生物标记物、至少40种分类器生物标记物、或至少48种分类器生物标记物。在一些实施方案中,对一个或多个样品训练集进行归一化以去除样品与样品之间的变化。

在一些实施方案中,比较可以包括应用统计算法(如例如,任何合适的多变量统计分析模型),该统计算法可以是参数的或非参数的。在一些实施方案中,应用统计算法可以包括确定从人肺组织样品获得的表达数据与来自一个或多个腺癌训练集的表达数据之间的相关性。在一些实施方案中,进行交叉验证,如(例如)留一交叉验证(leave-one-out cross-validation, (LOOCV))。在一些实施方案中,进行整合关联(integrative correlation)。在一些实施方案中,进行斯皮尔曼关联(Spearman correlation)。在一些实施方案中,基于基因表达数据,将基于质心的方法用于统计算法,如Mullins等人(2007) *Clin Chem.* 53 (7) :1273-9(将该文献通过引用以其整体并入本文)所描述的。

可以将来自受试者的样品(测试样品)进行的基因表达的结果与一个或多个已知或怀疑是正常的生物样品(“参考样品”或“正常样品”,例如,非腺癌样品)或来自其的数据进行比较。在一些实施方案中,参考样品或参考基因表达数据获自或源自于已知具有腺癌的特定分子亚型(即,鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型))的个体。

可以同时或在与测试样品不同的时间对参考样品进行测定。可替代地,来自参考样品的生物标记物水平信息可以存储在数据库或其他装置中以供在以后的日期访问。

可以将对测试样品的测定的生物标记物水平结果与对参考样品的相同测定的结果进行比较。在一些情况下,对参考样品的测定的结果来自数据库或是一个或多个参考值。在一些情况下,对参考样品的测定的结果是本领域技术人员已知的或通常接受的值或值的范围。在一些情况下,比较是定性的。在其他情况下,比较是定量的。在一些情况下,定性比较或定量比较可能涉及但不限于以下项中的一项或多项:比较荧光值、光斑强度、吸光度值、化学发光信号、直方图、临界阈值、统计显著性值、本文所述基因的表达水平、mRNA拷贝数目。

在一个实施方案中,针对每个生物标记物水平小组测量计算优势比 (OR)。这里,OR是患者的测量生物标记物值与结果(例如,肺腺癌亚型)之间的关联的量度。例如,参见, J.Can.Acad.Child Adolesc.Psychiatry 2010;19(3):227-229,出于所有目的将该文献通过引用以其整体并入。

在一个实施方案中,可以确定指定的统计置信水平,以便提供关于肺癌亚型的置信水平。例如,可以确定大于90%的置信水平可以是肺癌亚型的有用预测因子。在其他实施方案中,可以选择更严格或不太严格的置信水平。例如,可以选择约或至少约50%、60%、70%、75%、80%、85%、90%、95%、97.5%、99%、99.5%、或99.9%的置信水平。在一些情况下,所提供的置信水平可以与样品的质量、数据的质量、分析的质量、所使用的方法、和/或所分析基因表达值的数目(即,基因的数目)相关。可以基于预期的假阳性或假阴性数目来选择用于提供反应的可能性的指定置信水平。选择参数以便达到指定置信水平或以便鉴定具有诊断能力的标记物的方法包括但不限于受试者工作特征(ROC)曲线分析、副法线ROC、主成分分析、优势比分析、偏最小二乘分析、奇异值分解、最小绝对收缩与选择算子分析、最小角回归、以及阈值梯度定向正则化方法。

在一些情况下,确定肺腺癌亚型可以通过应用被设计为使基因表达数据可靠性归一化和/或改进的算法加以改进。在本发明的一些实施方案中,由于大量的待处理的单独数据点,数据分析利用了计算机或其他设备、机器或装置来应用本文所述的各种算法。“机器学习算法”是指基于计算的预测方法,本领域技术人员也将其称为“分类器”,用于表征一个或多个基因表达谱,例如,以便确定肺腺癌亚型。在一个实施方案中,使通过例如基于微阵列的杂交测定、测序测定、纳米串测定等确定的生物标记物水平经受算法以便对谱进行分类。监督学习通常涉及“训练”分类器以识别亚型(例如鳞状型(近端炎症型)阳性、支气管型(末端呼吸单元型)阳性或巨型(近端增生型)阳性)之间的区别,并然后在独立的测试集上“测试”分类器的准确性。因此,对于新的未知样品,分类器可以用于预测例如样品所属的类别(例如,鳞状型与支气管型与巨型)。

在一些情况下,稳健多阵列平均(RMA)方法可以用于使原始数据归一化。RMA方法通过计算许多微阵列上各匹配细胞的背景校正强度开始。在一个实施方案中,背景校正值被限制为正值,如Irizarry等人(2003).Biostatistics April 4(2):249-64(出于所有目的将该文献通过引用以其整体并入)所述。在背景校正之后,然后获得各背景校正的匹配细胞强度的以2为底的对数。然后使用分位数归一化方法(其中对于各输入阵列和各探针值,阵列百分位探针值被所有阵列百分位点的平均值替换,该方法在Bolstad等人Bioinformatics 2003(将该文献通过引用以其整体并入)中有更完整的描述)使各微阵列上背景校正的对数转化的匹配强度归一化。在分位数归一化之后,然后可以将归一化的数据拟合到线性模型中,以获得各微阵列上各探针的强度测量值。然后可以使用Tukey中值平滑算法(Tukey, J.W., Exploratory Data Analysis. 1977(出于所有目的将该文献通过引用以其整体并入))来确定归一化探针集数据的对数尺度强度水平。

可以执行各种其他软件程序。在某些方法中,可以使用glmnet通过lasso惩罚的逻辑回归来进行特征选择和模型估计(Friedman等人(2010).Journal of statistical software 33(1):1-22,将该文献通过引用以其整体并入)。可以使用TopHat来比对原始读数(Trapnell等人(2009).Bioinformatics 25(9):1105-11,将该文献通过引用以其整体并

入)。在方法中,使用e1071库(Meyer D. Support vector machines:the interface to libsvm package e1071.2014,将该文献通过引用以其整体并入),将高维特征(top feature)(N的范围为从10至200)用于训练线性支持向量机(SVM)(Suykens JAK, Vandewalle J.Least Squares Support Vector Machine Classifiers.Neural Processing Letters 1999;9(3):293-300,将该文献通过引用以其整体并入)。在一个实施方案中,使用pROC包(Robin X,Turck N,Hainard A,等人pROC:an open-source package for R and S+to analyze and compare ROC curves.BMC bioinformatics 2011;12:77(将该文献通过引用以其整体并入))来计算置信区间。

此外,可以过滤数据以除去可能被认为可疑的数据。在一个实施方案中,从具有少于约4个、5个、6个、7个或8个鸟苷+胞嘧啶核苷酸的微阵列探针衍生出的数据可以被认为不可靠的,因为它们有异常杂交倾向或二级结构问题。类似地,从具有多于约12个、13个、14个、15个、16个、17个、18个、19个、20个、21个或22个鸟苷+胞嘧啶核苷酸的微阵列探针衍生出的数据可以被认为不可靠的,因为它们有异常杂交倾向或二级结构问题。

在本发明的一些实施方案中,如果不以可检测的水平(高于背景)鉴定出来自探针集的数据,则它们可以从分析中排除。

在本公开文本的一些实施方案中,没有表现出变异或表现出低变异的探针集可以从进一步分析中排除。通过卡方检验从分析中排除低变异的探针集。在一个实施方案中,如果转化的变异在具有(N-1)自由度的卡方分布的99%置信区间的左侧,则探针集被认为是低变异的。 $(N-1) * \text{探针集变异} / (\text{基因探针集变异}) \cdot \text{Chi-Sq}(N-1)$,其中N是输入CEL文件的数目,(N-1)是卡方分布的自由度,且“基因的探针集变异”是整个基因的探针集变异的平均值。在本发明的一些实施方案中,如果给定mRNA或mRNA集的探针集包含少于通过前述用于GC含量、可靠性、变异等的过滤器步骤的最低数目的探针,则可以从进一步分析中排除它们。例如,在一些实施方案中,如果给定基因或转录物簇的探针集包含少于约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15个或少于约20个探针,则从进一步分析中排除它们。

在一个实施方案中,生物标记物水平数据分析的方法还包括使用如本文提供的特征选择算法。在本发明的一些实施方案中,特征选择通过使用LIMMA软件包(Smyth, G.K. (2005). Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber(编辑), Springer, New York, 第397-420页(出于所有目的将该文献通过引用以其整体并入))提供。

在一个实施方案中,生物标记物水平数据分析的方法包括使用预分类器算法。例如,算法可以使用特异性分子指纹来根据它们的组成对样品进行预分类,然后再应用校正/归一化因子。然后可以将该数据/信息输入到最终分类算法中,其将整合该信息来辅助最终的诊断。

在一个实施方案中,生物标记物水平数据分析的方法还包括使用如本文提供的分类器算法。在本发明的一个实施方案中,提供了对角线线性判别分析、k-近邻算法、支持向量机(SVM)算法、线性支持向量机、随机森林算法、或基于概率模型的方法、或它们的组合用于微阵列数据的分类。在一些实施方案中,基于在感兴趣的类别之间生物标记物水平的差异的统计学显著性,选择区分样品的鉴定标记物(例如,不同生物标记物水平谱、和/或腺癌

的不同分子亚型(例如,鳞状型、支气管型、巨型)的鉴定标记物)。在一些情况下,通过针对错误发现率(FDR)应用Benjamin Hochberg校正或另一种校正来调节统计学显著性。

在一些情况下,分类器算法可以用荟萃分析方法补充,如Fisher和Kaufman等人2007 *Bioinformatics* 23(13):1599-606(出于所有目的将该文献通过引用以其整体并入)描述的方法。在一些情况下,分类器算法可以用荟萃分析方法补充,如可重复性分析。

用于导出后验概率并将后验概率应用于生物标记物水平数据分析的方法在本领域中是已知的,并且已经描述于例如Smyth, G.K. 2004 *Stat. App. Genet. Mol. Biol.* 3: Article 3(出于所有目的将该文献通过引用以其整体并入)中。在一些情况下,可以将后验概率用于本发明的方法中,以对由分类器算法提供的标记物进行分级。

生物标记物水平谱分析结果的统计学评价可以提供指示以下中的一种或多种的一个或多个定量值:腺癌的分子亚型(鳞状型、支气管型或巨型);特定治疗干预(例如,血管生成抑制剂治疗、化学疗法或免疫疗法)成功的可能性。在一个实施方案中,数据以其最有用的形式直接呈现给医师以指导患者护理,或者用于在临床试验中定义患者群体或针对给定药物定义患者群体。分子谱分析的结果可使用本领域已知的多种方法进行统计学评估,该方法包括但不限于:学生T检验、双侧T检验、皮尔逊秩和分析、隐马尔可夫模型分析、q-q作图分析、主成分分析、单因素ANOVA、双因素ANOVA、LIMMA等。

在一些情况下,可通过随时间追踪受试者以确定原始诊断的准确性而确定准确性。在其他情况下,准确性可以按确定性的方式或使用统计学方法确立。例如,受试者工作特征(ROC)分析可用于确定最佳测定参数,以实现特定水平的准确性、特异性、阳性预测值、阴性预测值和/或错误发现率。

在一些情况下,将生物标记物水平谱分析测定的结果输入数据库中以供分子谱分析企业、个人、医疗供应商、或保险供应商的代表或代理访问。在一些情况下,测定结果包括企业的代表、代理或咨询人员(如医疗专业人员)进行的样品分类、鉴定、或诊断。在其他情况下,自动提供数据的计算机或算法分析。在一些情况下,分子谱分析企业可以向个人、保险供应商、医疗供应商、研究人员或政府实体针对以下项中的一种或多种收费:进行的分子谱分析测定、咨询服务、数据分析、报告结果或数据库访问。

在本发明的一些实施方案中,生物标记物水平谱分析测定的结果作为计算机屏幕上的报告或纸件记录呈现。在一些实施方案中,报告可以包括但不限于如以下中的一种或多种信息:与参考样品或一个或多个参考值相比生物标记物的水平(例如,如通过拷贝数目或荧光强度等报告的);基于生物标记物水平值和肺腺癌亚型和建议疗法,受试者将对特定疗法有反应的可能性。

在一个实施方案中,基因表达谱分析的结果可以分为以下中的一种或多种:鳞状型(近端炎症型)阳性、支气管型(末端呼吸单元型)阳性、巨型(近端增生型)阳性、鳞状型(近端炎症型)阴性、支气管型(末端呼吸单元型)阴性、巨型(近端增生型)阴性;可能对血管生成抑制剂、免疫疗法或化学疗法有反应;不太可能对血管生成抑制剂、免疫疗法或化学疗法有反应;或其组合。

在本发明的一些实施方案中,使用经训练的算法对结果进行分类。本发明的经训练的算法包括已使用已知基因表达值和/或正常样品(例如,来自诊断患有特定分子亚型的腺癌的个体的样品)的参考集开发的算法。在一些情况下,已知基因表达值的参考集从已被

诊断患有特定分子亚型的腺癌并且还已知对血管生成抑制剂疗法有反应(或不反应)的个体获得。在一些情况下,已知基因表达值的参考集从已被诊断患有特定分子亚型的腺癌并且还已知对免疫疗法有反应(或不反应)的个体获得。在一些情况下,已知基因表达值的参考集从已被诊断患有特定分子亚型的腺癌并且还已知对化学疗法有反应(或不反应)的个体获得。

适用于对样品分类的算法包括但不限于k-近邻算法、支持向量机、线性判别分析、对角线线性判别分析、updown、朴素贝叶斯算法、神经网络算法、隐马尔可夫模型算法、基因算法、或其任何组合。

当将二元分类器与实际真值(例如,来自生物样品的值)进行比较时,通常存在四种可能的结果。如果预测的结果是p(其中“p”是正分类器输出,例如存在删除或重复综合症)并且实际值也是p,则将其称作真阳性(TP);但是,如果实际值是n,则将其称为是假阳性(FP)。相反,当预测结果和实际值两者均为n(其中“n”是负分类器输出,例如没有删除或重复综合症)时,出现真阴性,而当预测结果为n同时实际值为p时,出现假阴性。在一个实施方案中,考虑进行试图确定人是否可能或不太可能对血管生成抑制剂疗法有反应的测试。当这个人测试为阳性但实际上确实有反应时,在这种情况下出现假阳性。另一方面,当这个人测试为阴性时(表明他们不太可能有反应),当他们实际上可能有反应时,出现假阴性。对于对肺癌亚型的分类而言也是如此。

疾病的阳性预测值(PPV)或精确率或后验概率是具有阳性测试结果的被正确诊断为可能反应或不太可能反应的、或者被诊断为具有正确的肺癌亚型、或者它们的组合的受试者的比例。它反映了阳性测试反映出测试的基础病状的概率。但是,它的值确实依赖于疾病的流行度,其可发生变化。在一个实施例中,提供以下特征:FP(假阳性);TN(真阴性);TP(真阳性);FN(假阴性)。假阳性率(α) = $FP / (FP + TN)$ - 特异性;假阴性率(β) = $FN / (TP + FN)$ - 灵敏度;能力 = 灵敏度 = $1 - \beta$;阳性似然比 = 灵敏度 / (1 - 特异性);阴性似然比 = $(1 - \text{灵敏度}) / \text{特异性}$ 。阴性预测值(NPV)是被正确诊断的具有阴性测试结果的受试者的比例。

在一些实施方案中,主题方法的生物标记物水平分析的结果提供了给出的诊断正确的统计学置信水平。在一些实施方案中,这种统计学置信水平是至少约或多于约85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.5%、或更多。

在一些实施方案中,该方法还包括基于样品中的生物标记物水平与(例如存在于至少一个训练集中的)参考生物标记物水平的比较将肺组织样品分类为特定肺癌亚型。在一些实施方案中,如果比较的结果满足一个或多个标准,如例如,最小百分比一致性,基于百分比一致性计算的统计量的值,例如(例如,卡帕(kappa)统计量,最小相关性(例如,皮尔逊相关性)和/或类似物,则肺组织样品被分类为特定亚型。

据预期,本文描述的方法可以通过软件(存储在存储器中和/或在硬件上执行)、硬件或其组合来进行。硬件模块可以包括通用处理器、现场可编程门阵列(FPGA)、和/或专用集成电路(ASIC)。软件模块(在硬件上执行)可以用各种软件语言(例如,计算机代码)表达,包括Unix实用工具、C、C++、JavaTM、Ruby、SQL、SAS[®]、R编程语言/软件环境、Visual BasicTM、以及其他面向对象的程序性或其他编程语言和开发工具。计算机代码的例子包括但不限于微代码或微指令、机器指令(如由编译器产生的)、用于产生网络服务的代码、以及包含使用解释器由计算机执行的高水平指令的文件。计算机代码的另外的例子包括但不限

于控制信号、加密代码、和压缩代码。

本文描述的一些实施方案涉及具有非暂时性计算机可读介质(还可以称作非暂时性处理器可读介质或存储器)的设备,该介质上具有用于执行各种计算机实施的操作的指令或计算机代码和/或本文公开的方法。计算机可读介质(或处理器可读介质)在它本身不包括暂时性传播信号(例如,在传输介质如空间或缆线上传播携带信息的电磁波)的意义上是非暂时性的。该介质和计算机代码(还可以称作代码)可以是被设计并构建为用于特定目的或多个目的那些。非暂时性计算机可读介质的例子包括但不限于:磁存储介质,如硬盘、软盘和磁带;光存储介质,如光碟/数字视频光盘(CD/DVD)、光碟只读存储器(CD-ROM)和全息设备;磁光存储介质,如光碟;载波信号处理模块;以及专门配置成存储和执行程序代码的硬件设备,例如专用集成电路(ASIC)、可编程逻辑设备(PLD)、只读存储器(ROM)和随机存取存储器(RAM)设备。在此描述的其他实施方案涉及计算机程序产品,其可以包括例如在此讨论的指令和/或计算机代码。

在一些实施方案中,单一生物标记物、或从约5种至约10种、从约8种至约16种、从约5种至约15种、从约5种至约20种、从约5种至约25种、从约5种至约30种、从约5种至约35种、从约5种至约40种、从约5种至约45种、从约5种至约48种生物标记物(例如,如表1中公开的)能够对肺腺癌的亚型进行分类,其中预测成功率为至少约70%、至少约71%、至少约72%、约73%、约74%、约75%、约76%、约77%、约78%、约79%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、高达100%、以及介于两者之间的所有值。在一些实施方案中,本文(例如,在表1中)公开的生物标记物的任何组合可以用于获得至少约70%、至少约71%、至少约72%、约73%、约74%、约75%、约76%、约77%、约78%、约79%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、高达100%、以及介于两者之间的所有值的预测成功率。

在一些实施方案中,单一生物标记物、或从约5种至约10种、从约8种至约16种、从约5种至约15种、从约5种至约20种、从约5种至约25种、从约5种至约30种、从约5种至约35种、从约5种至约40种、从约5种至约45种、从约5种至约48种生物标记物(例如,如表1中所公开的)能够对肺腺癌亚型进行分类,其中灵敏度或特异性为至少约70%、至少约71%、至少约72%、约73%、约74%、约75%、约76%、约77%、约78%、约79%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、高达100%、以及介于两者之间的所有值。在一些实施方案中,本文公开的生物标记物的任何组合可以用于获得至少约70%、至少约71%、至少约72%、约73%、约74%、约75%、约76%、约77%、约78%、约79%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、高达100%、以及介于两者之间的所有值的灵敏度或特异性。

分类器基因选择

在一个实施方案中,本文提供的方法和组合物可用于通过分析一组生物标记物的表达来确定来自患者的样品(例如,肺组织样品)的AD亚型,其中与本领域已知的用于对肺

AD亚型进行分子分类的方法中的生物标记物相比,该组生物标记物包含较少数量的生物标记物。在一些情况下,该组生物标记物是少于250种、240种、230种、220种、210种、200种、150种、100种、95种或90种生物标记物。在一些情况下,该组生物标记物是少于50种生物标记物。在一些情况下,该组生物标记物是表1中列出的48种生物标记物的组。在一些情况下,该组生物标记物是表1中列出的生物标记物的子集。可用于本文提供的方法和组合物中的生物标记物或分类器基因可以选自来自一个或多个数据库的一个或多个肺腺癌数据集。该数据库可以是公共数据库。在一个实施方案中,可用于本文提供的用于检测或诊断肺腺癌亚型的方法和组合物中的分类器基因(例如,表1和表2中列出的一种或多种基因)选自来自癌症基因组图集(TCGA)的肺腺癌RNAseq数据集。在一个实施方案中,通过对分类器基因的大集合进行基于计算机模拟的过程以便确定其表达谱可以用于确定从受试者获得的样品的AD亚型的最小基因数目来选择可用于本文提供的方法和组合物的分类器基因,例如表1中的那些。在一些情况下,分类器基因的大集合可以是肺AD RNAseq数据集,如例如来自TCGA。在一些情况下,分类器基因的大集合可以是本文所述的506种基因分类器,其中该506种基因分类器可以用于定义金标准亚型。用于选择如本文提供的用于确定来自患者的样品的肺AD亚型的基因盒的计算机模拟过程可以包括对标准506种分类器基因应用或使用进行修改的至最近质心阵列分类(CLaNC)算法,以针对每种亚型选择相同数量的负相关基因和正相关基因。为了确定包括在标签中的基因的最佳数目(例如,如表1中所示的每种亚型16个),该过程可以还包括使用如本文提供的TCGA肺腺癌数据集进行5折交叉验证以产生交叉-验证曲线,如图8所示。为了获得基因分类器的最终列表,该方法可以进一步包括将至最近质心阵列分类(CLaNC)应用至整个TCGA数据集减去20%的具有最低金标准亚型预测强度的样品,并从每种亚型中去除相等数量,例如图9中所示。

在一个实施方案中,该方法还包括验证基因分类器。验证可以包括在几个新鲜冷冻的可公开获得的阵列和RNAseq数据集中测试分类器的表达,并基于所述表达水平称呼亚型,并随后将该表达与如由先前公布的506种基因标签定义的金标准亚型称谓进行比较。然后可以在归档的福尔马林固定的石蜡包埋(FFPE)腺癌样品的新收集的RNAseq数据集中进行基因标签(例如,表1)的最终验证,以确保FFPE样品中的可比性能。在一个实施方案中,基于本文所述的计算机模拟CLaNC过程选择表1的分类器生物标记物。将基因符号和官方基因名称分别列于第2列和第3列。

在一个实施方案中,本发明的方法需要检测在从患者获得的肺癌细胞样品中的其表达被改变的至少1种、至少2种、至少3种、至少4种、至少5种、至少6种、至少7种、至少8种、至少9种、至少10种、至少11种、至少12种、至少13种、至少14种、至少15种、或多达16种分类器生物标记物,以鉴定TRU、PP或PI肺腺癌亚型。这同样适用于如本文提供的其他分类器基因表达数据集。

在另一个实施方案中,本发明的方法需要在从患者获得的肺癌细胞样品(例如,肺AD样品)中检测出表1的48种基因生物标记物中的总共至少1种、至少2种、至少5种、至少8种、至少10种、至少16种、至少20种、至少30种、至少32种、或多达48种分类器生物标记物,以便鉴定TRU、PP或PI肺腺癌亚型。这同样适用于如本文提供的其他分类器基因表达数据集。

在一个实施方案中,表1中的至少1种、至少2种、至少3种、至少4种、至少5种、至少6种、至少7种、或多达8种生物标记物在肺腺癌的特定亚型中被“上调”。在另一个实施方案

中,表1中的至少1种、至少2种、至少3种、至少4种、至少5种、至少6种、至少7种、或多达8种生物标记物在肺腺癌的特定亚型中被“下调”。这同样适用于如本文提供的其他分类器基因表达数据集。

在一个实施方案中,如本文提供的“上调的”生物标记物的表达水平增加了约0.5倍、约1倍、约1.5倍、约2倍、约2.5倍、约3倍、约3.5倍、约4倍、约4.5倍、约5倍、以及介于两者之间的任何值。在另一个实施方案中,如本文提供的“下调的”生物标记物的表达水平降低了约0.8倍、约1.4倍、约2倍、约2.6倍、约3.2倍、约3.6倍、约4倍、以及介于两者之间的任何值。

认识到在本发明的实践中可以使用另外的基因或蛋白质。例如,波形蛋白(中间丝状蛋白家族的成员)可以用于鉴定腺癌亚型近端增生型(巨型)亚型,并且SMA可以用于鉴定近端炎症型(鳞状型)亚型。通常,可用于对肺腺癌亚型进行分类的基因包括能够独立地在正常与肿瘤之间、或肺癌的不同类别或等级之间作出区分的那些基因。如果受试者工作特征(receiver operator characteristic,ROC)曲线下面积约为1,则认为基因能够可靠地区分亚型。

临床/治疗用途

在一个实施方案中,本文提供了一种用于确定患有癌症的患者的疾病结果或预后的方法。在一些情况下,该癌症是肺癌。可以通过检查一段时间或间隔(例如,0至36个月或者0至60个月)的总体存活期来测量疾病结果或预后。在一个实施方案中,根据亚型(例如,对于肺癌,腺癌(TRU、PI和PP))对存活期进行分析。可以使用标准卡普兰-迈耶(Kaplan-Meier)图以及Cox比例风险模型来评估无复发和总体存活期。

在一个实施方案中,在确定患者的肺癌亚型后,选择患者以接受合适的疗法,例如化学疗法或用血管生成抑制剂的药物疗法或免疫疗法。在一个实施方案中,在确定患者的肺癌亚型后,向患者给予合适的治疗剂,例如一种或多种化学治疗剂、或血管生成抑制剂、或一种或多种免疫治疗剂。在一个实施方案中,该疗法是免疫疗法,并且该免疫治疗剂是检查点抑制剂、单克隆抗体、生物反应调节剂、治疗性疫苗或细胞免疫疗法。

本发明的方法还可用于评价对疗法的临床反应,以及对于临床试验的终点用于评价新疗法的功效。可以将顺序诊断表达谱向正常移动的程度用作候选疗法的功效的一种量度。

在一个实施方案中,本发明的方法还可用于基于肺腺癌(AD)亚型预测对不同线的疗法的反应。例如,可以通过更准确地分配肿瘤亚型来改进化学治疗反应。同样,可以基于肿瘤亚型来配制治疗方案。例如,临床试验已显示令人信服的证据证明VEGF抑制剂(贝伐单抗)可以有效治疗NSCLC。

在一个实施方案中,末端呼吸单元(TRU)亚型对EGFR抑制剂和培美曲塞可以具有增强的反应。在另一个实施方案中,近端增生型(PP)对化学疗法可以具有增强的反应。在另一个实施方案中,近端炎症型(PI)对免疫疗法可以具有增强的反应。在另一个实施方案中,所有亚型对化学疗法、血管生成抑制剂治疗和免疫疗法可以具有增强的反应。

血管生成抑制剂

在一个实施方案中,在确定患者的肺腺癌亚型后,选择患者以接受用血管生成抑制剂的药物疗法。

在一个实施方案中,血管生成抑制剂是血管内皮生长因子(VEGF)抑制剂、VEGF受体抑制剂、血小板衍生生长因子(PDGF)抑制剂或PDGF受体抑制剂。

每个生物标记物小组可以包括分类器可用的一种、两种、三种、四种、五种、六种、七种、八种或更多种生物标记物(也称为“分类器生物标记物”),以评估腺癌患者是否可能对血管生成抑制剂疗法有反应;选择腺癌患者以接受血管生成抑制剂疗法;确定“缺氧得分”和/或将腺癌样品亚型分型为鳞状型(也称为近端炎症型)、支气管型(也称为末端呼吸单元型)或巨型(也称为近端增生型)分子亚型。如本文所用,术语“分类器”可以指用于统计分类的任何算法,并且可以在硬件、软件或其组合中实施。分类器可以具有2级、3级、4级或更高级的分类,并且可以依赖于被分类的实体的性质。可以采用一种或多种分类器来实现本文公开的方面。

通常,本文提供了确定腺癌患者是否可能对血管生成抑制剂疗法有反应的方法、或选择腺癌患者以接受血管生成抑制剂疗法的方法。在一个实施方案中,该方法包括使用本文所述的方法评估患者的腺癌亚型是鳞状型(近端炎症型)、支气管型(末端呼吸单元型)还是巨型(近端增生型)(例如,评估表1中的一种或多种分类器生物标记物的表达),并且在核酸水平上探测来自患者的腺癌样品中选自下组的至少五种生物标记物的水平,该组由以下组成:RRAGD、FABP5、UCHL1、GAL、PLOD、DDIT4、VEGF、ADM、ANGPTL4、NDRG1、NP、SLC16A3、和C140RF58(参见表3)。在另外的实施方案中,探测步骤包括将样品与五种或更多种寡核苷酸混合,该五种或更多种寡核苷酸在适合于该五种或更多种寡核苷酸与其互补序列或基本互补序列杂交的条件下与该至少五种生物标记物的核酸分子的部分基本上互补,检测在该五种或更多种寡核苷酸与其互补序列或基本互补序列之间是否发生杂交;并且基于该检测步骤获得样品的杂交值。然后将样品的杂交值与来自该至少一个样品训练集的一个或多个参考杂交值进行比较,其中该至少一个样品训练集包括(i)来自过表达该至少五种生物标记物或过表达该至少五种生物标记物的子集的样品的该至少五种生物标记物的一个或多个杂交值,(ii)来自参考鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型)样品的该至少五种生物标记物的杂交值,或(iii)来自无腺癌肺样品的该至少五种生物标记物的杂交值。然后基于(i)患者的腺癌亚型和(ii)比较的结果来确定患者是否可能对血管生成抑制剂疗法有反应、或选择患者以接受血管生成抑制剂。

表3.缺氧谱的生物标记物

名称	缩写	GenBank登录号
RRAGD	Ras相关的GTP结合D	BC003088
FABP5	脂肪酸结合蛋白5	M94856
UCHL1	泛素羧基末端酯酶L1	NM_004181
GAL	甘丙肽	BC030241
PLOD	前胶原-赖氨酸, 2-氧化戊二酸 5-双加氧酶赖氨酸羟化酶	M98252
DDIT4	DNA损伤诱导的转录物4	NM_019058

表3.缺氧谱的生物标记物

名称	缩写	GenBank登录号
VEGF	血管内皮生长因子	M32977
ADM	肾上腺髓质素	NM_001124
ANGPTL4	血管生成素样4	AF202636
NDRG1	N-myc下游调节基因1	NM_006096
NP	核苷磷酸化酶	NM 000270
SLC16A3	溶质载体家族16单羧酸转运蛋白成员3	NM_004207
C14ORF58	染色体14开放阅读框58	AK000378

上述一组十三种生物标记物或其子集在本文中也称为“缺氧谱”。

在一个实施方案中,本文提供的方法包括在从受试者获得的腺癌样品中确定选自RRAGD、FABP5、UCHL1、GAL、PLOD、DDIT4、VEGF、ADM、ANGPTL4、NDRG1、NP、SLC16A3、和C14ORF58的至少五种生物标记物、至少六种生物标记物、至少七种生物标记物、至少八种生物标记物、至少九种生物标记物、或至少十种生物标记物、或者五种至十三种、六种至十三种、七种至十三种、八种至十三种、九种至十三种或十种至十三种生物标记物的水平。在一些情况下,生物标记物表达可以针对样品中所有RNA转录物或其表达产物的表达水平或针对RNA转录物或其表达产物的参考集进行归一化。如全文所解释的参考集可以是与腺癌样品平行测试的实际样品,或者可以是来自数据库或存储的数据集的值的参考集。在一个实施方案中,以拷贝数目、相对荧光值或检测的荧光值报告表达水平。如使用本文提供的方法确定的缺氧谱连同腺癌亚型的生物标记物的表达水平可以用于本文所述的方法中,以确定患者是否可能对血管生成抑制剂疗法有反应。

在一个实施方案中,针对样品中的所有RNA转录物或其非天然cDNA表达产物、或蛋白质产物的表达水平,或者样品中的RNA转录物的参考集或其非天然cDNA表达产物的参考集、或其蛋白质产物的参考集的表达水平,对十三种生物标记物(或其子集,如上所述,例如,五种或更多种、从约五种至约13种)的表达水平进行归一化。

在一个实施方案中,血管生成抑制剂治疗包括但不限于整合素拮抗剂;选择素拮抗剂;粘附分子拮抗剂(细胞间粘附分子(ICAM)-1、ICAM-2、ICAM-3、血小板内皮粘附分子(PCAM)、血管细胞粘附分子(VCAM)的拮抗剂);淋巴细胞功能相关抗原1(LFA-1);碱性成纤维细胞生长因子拮抗剂;血管内皮生长因子(VEGF)调节剂;血小板衍生生长因子(PDGF)调节剂(例如,PDGF拮抗剂)。

在确定受试者是否可能对整合素拮抗剂有反应的一个实施方案中,该整合素拮抗剂是小分子整合素拮抗剂,例如,由Paolillo等人(Mini Rev Med Chem, 2009, 第12卷,第1439-1446页,将该文献通过引用以其整体并入)描述的拮抗剂、或白细胞粘附诱导细胞因子或生长因子拮抗剂(例如,肿瘤坏死因子- α (TNF- α)、白细胞介素-1 β (IL-1 β)、单核细胞趋

化蛋白-1(MCP-1)和血管内皮生长因子(VEGF))，如美国专利号6,524,581(将该文献通过引用以其整体并入本文)中所述。

本文提供的方法还可用于确定受试者是否可能对以下血管生成抑制剂中的一种或多种有反应：干扰素 γ 1 β 、干扰素 γ 1 β (Actimmune \circledR)与吡非尼酮、ACUHTR028、 α V β 5、氨基苯甲酸钾、淀粉样蛋白P、ANG1122、ANG1170、ANG3062、ANG3281、ANG3298、ANG4011、抗CTGF RNAi、Aplidin、黄芪提取物与丹参和五味子、动脉粥样硬化斑块阻断剂、Azol、AZX100、BB3、结缔组织生长因子抗体、CT140、达那唑、艾思瑞(Esbriet)、EXC001、EXC002、EXC003、EXC004、EXC005、F647、FG3019、Fibrocorin、卵泡抑素、FT011、半乳凝素-3抑制剂、GKT137831、GMCT01、GMCT02、GRMD01、GRMD02、GRN510、Heberon Alfa R、干扰素 α -2 β 、ITMN520、JKB119、JKB121、JKB122、KRX168、LPA1受体拮抗剂、MGN4220、MIA2、微小RNA29a寡核苷酸、MMI0100、诺斯卡品、PBI4050、PBI4419、PDGFR抑制剂、PF-06473871、PGN0052、Pirespa、Pirfenex、吡非尼酮、普利提环肽(plitidepsin)、PRM151、Px102、PYN17、PYN22与PYN17、Relivergen、rhPTX2融合蛋白、RXI109、胰泌素、STX100、TGF- β 抑制剂、转化生长因子、 β -受体2寡核苷酸、VA999260、XV615或其组合。

在另一个实施方案中，提供了一种用于确定受试者是否可能对一种或多种内源性血管生成抑制剂有反应的方法。在另外的实施方案中，内源性血管生成抑制剂是内皮抑素、源自XVIII型胶原的20kDa C-末端片段、血管抑素(胞浆素的38kDa片段)、血小板反应蛋白(TSP)蛋白家族的成员。在另外的实施方案中，该血管生成抑制剂是TSP-1、TSP-2、TSP-3、TSP-4和TSP-5。还提供了用于确定对以下血管生成抑制剂中的一种或多种有反应的可能性的方法：可溶性VEGF受体，例如可溶性VEGFR-1和神经毡蛋白1(NPR1)、血管生成素-1、血管生成素-2、血管形成抑制素、钙网蛋白、血小板因子-4；金属蛋白酶组织抑制剂(TIMP)(例如，TIMP1、TIMP2、TIMP3、TIMP4)；软骨衍生的血管生成抑制剂(例如，肌钙蛋白肽I和chondromodulin I)；具有血小板反应蛋白基序1的解整合素和金属蛋白酶；干扰素(IFN)(例如，IFN- α 、IFN- β 、IFN- γ)；趋化因子，例如具有C-X-C基序的趋化因子(例如，CXCL10，也称为干扰素 γ 诱导的蛋白质10或小的诱导型细胞因子B10)；白细胞介素细胞因子(例如，IL-4、IL-12、IL-18)；凝血素；抗凝血酶III片段；催乳素；由TNFSF15基因编码的蛋白质；骨桥蛋白；乳腺丝抑蛋白；血管能抑素；增殖素相关蛋白。

在一个实施方案中，提供了用于确定对以下血管生成抑制剂中的一种或多种有反应的可能性的方法：血管生成素-1、血管生成素-2、血管抑素、内皮抑素、血管形成抑制素、血小板反应蛋白、钙网蛋白、血小板因子-4、TIMP、CDAI、干扰素 α 、干扰素 β 、血管内皮生长因子抑制剂(VEGI)meth-1、meth-2、催乳素、VEGI、SPARC、骨桥蛋白、乳腺丝抑蛋白、血管能抑素、增殖素相关蛋白(PRIP)、网状内皮系统刺激素(restin)、TSP-1、TSP-2、干扰素 γ 1 β 、ACUHTR028、 α V β 5、氨基苯甲酸钾、淀粉样蛋白P、ANG1122、ANG1170、ANG3062、ANG3281、ANG3298、ANG4011、抗CTGF RNAi、Aplidin、黄芪提取物与丹参和五味子、动脉粥样硬化斑块阻断剂、Azol、AZX100、BB3、结缔组织生长因子抗体、CT140、达那唑、艾思瑞(Esbriet)、EXC001、EXC002、EXC003、EXC004、EXC005、F647、FG3019、Fibrocorin、卵泡抑素、FT011、半乳凝素-3抑制剂、GKT137831、GMCT01、GMCT02、GRMD01、GRMD02、GRN510、Heberon Alfa R、干扰素 α -2 β 、ITMN520、JKB119、JKB121、JKB122、KRX168、LPA1受体拮抗剂、MGN4220、MIA2、微小RNA29a寡核苷酸、MMI0100、诺斯卡品、PBI4050、PBI4419、PDGFR抑制剂、PF-06473871、

PGN0052、Pirespa、Pirfenex、吡非尼酮、普利提环肽、PRM151、Px102、PYN17、PYN22与PYN17、Relivergen、rhPTX2融合蛋白、RXI109、胰泌素、STX100、TGF- β 抑制剂、转化生长因子、 β -受体2寡核苷酸、VA999260、XV615或其组合。

在又另一个实施方案中,血管生成抑制剂可以包括帕唑帕尼(Votrient)、舒尼替尼(Sutent)、索拉非尼(Nexavar)、阿西替尼(Inlyta)、帕纳替尼(Iclusig)、凡德他尼(vandetanib)(Caprelsa)、卡博替尼(Cometriq)、雷莫芦单抗(Cyramza)、瑞格拉非尼(Stivarga)、阿柏西普(ziv-aflibercept)(Zaltrap)、莫特塞尼(motesanib)或其组合。在另一个实施方案中,血管生成抑制剂是VEGF抑制剂。在另外的实施方案中,该VEGF抑制剂是阿西替尼、卡博替尼、阿柏西普(aflibercept)、布立尼布、tivozanib、雷莫芦单抗或莫特塞尼。在又另外的实施方案中,血管生成抑制剂是莫特塞尼。

在一个实施方案中,本文提供的方法涉及确定受试者对血小板衍生生长因子(PDGF)家族成员的拮抗剂(例如,抑制、减少或调节PDGF受体(PDGFR)的信号传导和/或活性的药物)有反应的可能性。例如,在一个实施方案中,PDGF拮抗剂是抗PDGF适体、抗PDGF抗体或其片段、抗PDGFR抗体或其片段、或小分子拮抗剂。在一个实施方案中,PDGF拮抗剂是PDGFR- α 或PDGFR- β 的拮抗剂。在一个实施方案中,该PDGF拮抗剂是抗PDGF- β 适体E10030、舒尼替尼、阿西替尼、索拉菲尼、伊马替尼、甲磺酸伊马替尼、尼达尼布、盐酸帕唑帕尼、帕纳替尼、MK-2461、多韦替尼、帕唑帕尼、crenolanib、PP-121、替拉替尼、伊马替尼、KRN 633、CP 673451、TSU-68、Ki8751、amuvatinib、tivozanib、马赛替尼、二磷酸莫替沙尼、、多韦替尼二乳酸、利尼伐尼(linifanib)(ABT-869)。

在确定患者是否可能对血管生成抑制剂疗法有反应、或选择患者以接受血管生成抑制剂疗法时,在一个实施方案中,向患者给予血管生成抑制剂。血管生成抑制剂可以是本文所述的任何血管生成抑制剂。

免疫疗法

在一个实施方案中,本文提供了用于通过以下方式确定腺癌(AD)肺癌患者是否可能对免疫疗法有反应的方法:确定从患者获得的样品的AD亚型、以及基于该AD肺癌亚型评估患者是否可能对免疫疗法有反应。在另一个实施方案中,本文提供了通过以下方式选择患有AD的患者进行免疫疗法的方法:确定来自患者的样品的AD亚型、以及基于该AD亚型选择患者进行免疫疗法。可以使用本领域已知的对AD亚型分型的任何方法来进行从患者获得的样品的AD亚型的确定。在一个实施方案中,从患者获得的样品先前已被诊断为AD,并且将本文提供的方法用于确定该样品的AD亚型。先前的诊断可以是基于组织学分析。组织学分析可以由一名或多名病理学家进行。在一个实施方案中,通过一组或一小组生物标记物或其子集的基因表达分析以便产生表达谱来进行AD亚型分型。可以对从患者获得的肺癌样品(例如,肺癌AD样品)进行基因表达分析,以便确定选自本文所述的可公开获得的肺癌数据库和/或本文提供的表1的一种或多种生物标记物的存在、不存在或表达水平。AD亚型可以选自下组,该组由以下组成:鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型)。免疫疗法可以是本文提供的任何免疫疗法。在一个实施方案中,免疫疗法包括给予一种或多种检查点抑制剂。该检查点抑制剂可以是本文提供的任何检查点抑制剂,如例如靶向PD-1、PD-L1或CTLA4的检查点抑制剂。

如本文所公开,生物标记物小组或其子集可以是在任何一个或多个可公开获得的

AD基因表达数据集中公开的那些。在一个实施方案中,该肺癌是AD,并且该生物标记物小组或其子集是例如癌症基因组图集(TCGA)肺AD RNAseq基因表达数据集(n=515)。在一个实施方案中,该肺癌是AD,并且该生物标记物小组或其子集是例如Shedden等人(Nat Med 2008;14 (8):822-827,将其内容通过引用以其整体并入本文)中公开的AD基因表达数据集(n=442)。在一个实施方案中,该肺癌是AD,并且该生物标记物小组或其子集是例如Tomida等人(J Clin Oncol 2009;27 (17):2793-2799,将其内容通过引用以其整体并入本文)中公开的AD基因表达数据集(n=117)。在一个实施方案中,该肺癌是AD,并且该生物标记物小组或其子集是例如Wilkerson等人(PLoS One 2012;7 (5):e36530,将其内容通过引用以其整体并入本文)中公开的AD基因表达数据集(n=116)。在一个实施方案中,该肺癌是AD,并且该生物标记物小组或其子集是例如表1中公开的AD基因表达数据集。在一个实施方案中,该肺癌是AD,并且该生物标记物小组或其子集是例如表1中公开的AD基因表达数据集与来自可公开获得的AD表达数据集的一种或多种生物标记物的组合。在表2中,该表的第一列代表用于区分末端呼吸单元(TRU)的生物标记物列表。该表的中间列代表用于区分近端增生型(PP)的生物标记物列表。该表的最后一列代表用于区分近端炎症型(PI)的生物标记物列表。在一些情况下,如表2中所示,可以将总共48种生物标记物用于AD亚型确定。对于每种AD亚型,16种生物标记物中的8种可以是负相关基因,而8种可以是可以被选择作为特定AD亚型的基因标签的正相关基因。

[0164] 在一些实施方案中,用于肺癌亚型分型(例如,AD亚型分型)的方法包括检测分类器生物标记物组的表达水平。分类器生物标记物组可以是来自可公开获得的数据库的一组生物标记物,如例如一个或多个TCGA肺AD RNASeq基因表达数据集或本文提供的任何其他数据集。在一些实施方案中,检测包括在核酸水平或蛋白质水平上的表1的所有分类器生物标记物或本文提供的任何其他数据集。在另一个实施方案中,检测表1的单个分类器生物标记物或表1的分类器生物标记物的子集或本文提供的任何其他数据集,例如,从约五种至约二十种。在另一个实施方案中,检测表1的单个分类器生物标记物或表1的分类器生物标记物的子集和/或本文提供的任何其他数据集,例如,从约16种至约48种。在另一个实施方案中,检测表1的所有分类器生物标记物或本文提供的任何其他数据集。在另一个实施方案中,检测表1的至少一种或所有分类器生物标记物与本文提供的任何其他AD数据集的一种或多种分类器生物标记物的组合。例如使用对分类器生物标记物有特异性的引物和/或探针、和/或类似物,可以通过任何合适的技术(包括但不限于RNA-seq、逆转录酶聚合酶链反应(RT-PCR)、微阵列杂交测定、或另一种杂交测定(例如纳米串测定))来进行检测。在一些情况下,可用于扩增方法(例如,RT-PCR或qRT-PCR)的引物是适合用于单独地或组合地与来自本文提供的数据集的分类器基因结合的任何正向和反向引物。

在一个实施方案中,在如本文提供的确定肺癌亚型的方法中检测了在本文提供的任何AD基因表达数据集(包括例如,针对AD肺样品的表1)中的从约1种至约5种、约5种至约10种、从约5种至约15种、从约5种至约20种、从约5种至约25种、从约5种至约30种、从约5种至约35种、从约5种至约40种、从约5种至约45种、从约5种至约50种、从约5种至约55种、从约5种至约60种、从约5种至约65种、从约5种至约70种、从约5种至约75种、或从约5种至约80种生物标记物。在另一个实施方案中,在如本文提供的确定肺癌亚型的方法中检测了来自本文提供的AD基因表达数据集中的任何一个(包括例如针对AD肺样品的表1)的每种生物标记

物。

在一个实施方案中,本文提供的方法还包括确定AD亚型中免疫激活的存在、不存在或水平。可以通过创建表达谱或检测与从患者获得的样品(例如,肺癌样品)中的每种AD亚型相关的先天性免疫细胞和/或适应性免疫细胞相关的一种或多种生物标记物的表达来确定免疫细胞活化的存在或水平。在一个实施方案中,通过监测Bindea等人(Immunity 2013;39(4):782-795,将其内容通过引用以其整体并入本文)的免疫细胞标签来确定与AD亚型相关的免疫细胞活化。在一个实施方案中,该方法还包括测量单基因免疫生物标记物,如例如CTLA4、PDCD1和CD274 (PD-L1)、PDCDLG2 (PD-L2) 和/或IFN基因标签。与AD亚型相关的免疫激活(先天性和/或适应性)的存在或可检测水平可以指示或预测具有所述AD亚型的患者可以修改为免疫疗法。免疫疗法可以是用如本文提供的检查点抑制剂进行的治疗。在一个实施方案中,AD的PI亚型具有免疫表达。在一个实施方案中,本文提供了一种用于检测从患者获得的样品(例如,肺癌AD样品)中的本文提供的至少一种分类器生物标记物的表达的方法,该方法还包括在所述样品中检测到如本文提供的免疫激活后给予免疫治疗剂。

在一个实施方案中,该方法包括确定肺癌AD样品的亚型并随后确定所述亚型的免疫细胞活化的水平。在一个实施方案中,通过使用如本文所述的测序(例如,RNASeq)、扩增(例如,qRT-PCR)或杂交测定(例如,微阵列分析)确定一种或多种分类器生物标记物的表达水平来确定亚型。该一种或多种生物标记物可以选自可公开获得的数据库(例如,TCGA肺AD RNASeq基因表达数据集或本文提供的任何其他可公开获得的AD基因表达数据集)。在一些实施方案中,可以将表1的生物标记物用于特异性确定从患者获得的AD肺样品的亚型。在一个实施方案中,通过测量免疫标记物的基因表达标签来确定免疫细胞活化的水平。可以在用于对如本文所述的肺癌样品进行亚型分型的相同和/或不同样品中测量免疫标记物。可以测量的免疫标记物可以包含先天性免疫细胞(IIC)和/或适应性免疫细胞(AIC)基因标签、干扰素(IFN)基因标签、个体免疫标记物、主要组织相容性复合物II类(MHC II类)基因或其组合,由其组成或基本上由其组成。IIC和AIC两者的基因表达标签可以是本领域已知的所述细胞类型的任何已知基因标签。例如,免疫基因标签可以是来自Bindea等人(Immunity 2013;39(4):782-795)的那些。在一个实施方案中,用于在本文提供的方法中使用的免疫标记物选自表4A和/或表4B。个体免疫标记物可以是CTLA4、PDCD1和CD274 (PD-L1)。在一个实施方案中,用于在本文提供的方法中使用的个体免疫标记物选自表5。免疫标记物可以是一种或多种干扰素(INF)基因。在一个实施方案中,用于在本文提供的方法中使用的免疫标记物选自表6。免疫标记物可以是一种或多种MHCII基因。在一个实施方案中,用于在本文提供的方法中使用的免疫标记物选自表7。在又另一个实施方案中,用于在本文提供的方法中使用的免疫标记物选自表4A、表4B、表5、表6、表7、或其组合。

表 4A. 用于在本文提供的方法中使用的适应性免疫细胞(AIC)基因标签免疫标记物。

细胞类型	B 细胞	T 细胞	T 辅助细胞	Tem	Tem	Th1 细胞
	ABCB34(ATP 结合金亚家族 B 成员 4; NM_000443)	BCL11B(B-细胞淋巴瘤/白血病 11B; AJ404614.1)	ANP32B(酸性核磷蛋白 32 家族成员 B; NM_006401.2)	AQP3(水通道蛋白 3; NM_004925.4)	AKT3(AKT 丝氨酸/苏氨酸激酶 3; NM_005465.4)	APBB2(淀粉样前体蛋白结合家族 B 成员 2; NM_001166054.1)
BACH2(BTB 结构域和 CNC 同系物 2; NM_021813.3)	CD2(CD2 分子; NM_001328609.1)	ASF1A(抗沉默转录因子伴侣; NM_014034.2)	ATF7IP(激活转录因子 7 相互作用蛋白; NM_181352.1)	C7orf54(含葡萄糖核磷酸酶和 tudor 结构域的蛋白 1; NM_051199.1)	APOD(载脂蛋白 D; NM_001647.3)	
人类基因 (基因名称; GenBank 登录号*)	BCL11A(B-细胞 CLL/淋巴瘤 11A; NM_022893.3)	CD28(CD28 分子; NM_001243078.1)	ATF2(激活转录因子 2; NM_001256093.1)	ATM(ATM 丝氨酸/苏氨酸激酶; NM_000051.3)	CCR2(C-C 基序 2; NM_001123396.1)	ATP9A(ATP 酶, 脂转运蛋白 9A; NM_006045.2)
	BLK(BLK 原癌基因, Src 家族酪氨酸激酶; NM_001715.2)	CD3D(CD3d 分子; NM_000732.4)	BATF(碱性亮氨酸拉链 ATF 样转录因子; NM_006399.3)	CASP8(半胱天冬酶 8; NM_001228.4)	DDX17(DEAD 核解旋酶 17; NM_006386.4)	BST2(骨髓基质细胞抗原 2; NM_004335.3)
	BLNK(B-细胞连接蛋白; NM_013314.3)	CD3E(CD3e 分子; NM_000733.3)	C13orf34(北极光蛋白(aurora borealis); EU834129.1)	CDC14A(细胞分裂周期 14A; NM_003672.3)	EWSR1(EWS RNA 结合蛋白 1; NM_013986.3)	BTG3(BTG 抗增殖因子 3; NM_001130914.1)
	CCR9(C-C 基序趋化因子受体 9; NM_031200.2)	CD3G(CD3g 分子; NM_000733.2)	CD28(CD28 分子; NM_006139.3)	CEP68(中性体蛋白 68; NM_015147.2)	FLJ1(Flt-1 原癌基因, ETS 转录因子; NM_002017.4)	CC14(C-C 基序趋化因子配体 4; NM_002984.3)

CD19(CD19 分子; NM_001178098.1)	CD6(CD6 分子; NM_006725.4)	DDX50(DEAD 盒 解旋酶 50; NM_024045.1)	DDX50(BRCA2 区 域, mRNA 序列 CG030; U50531.1)	GDPD5(含甘油磷 酸二酯酶二酯 酶结构域的蛋白 5; NM_030792.6)	CD38(CD38 分 子; NM_001775.3)
CD72(CD72 分 子; NM_001782.2)	CD96(CD96 分 子; NM_198196.2)	FAM111A(具有序 列相似性的家族 111 成员 A; NM_022074.3)	CLUAPI(聚集素 相关蛋白 1; NM_015041.2)	LTK(白细胞受体 酪氨酸激酶; NM_002344.5)	CD70(CD70 分 子; NM_001252.4)
COCH(cochlin; NM_001135058.1)	GIMAP5(GTP 酶, IMAP 家族成 员 5; NM_018384.4)	FRYLF(YF4 样转录 共激活物; NM_015030.1)	CREBZF(CREB/A TF bZIP 转录因 子; NM_001039618.2)	MEFV(地中海 热; NM_000243.2)	CMAH(胞苷单磷 酸-N-乙酰神经氨 酸羟化酶, 假基 因; NR_002174.2)
CR2(补体 C3d 受 体 2; NM_001006658.2)	ITM2A(整合膜蛋 白 2A; NM_004867.4)	FUSIP1(富含丝氨 酸和精氨酸的剪接 因子 10; NM_0066625.5)	CYLD(CYLD 赖氨 酸 63 去泛素化 酶; NM_015247.2)	NFATC4(活化 T 细胞核因子 4; NM_001136022.2)	CSF2(集落刺激因 子 2; NM_000758.3)
DTNBB(小鼠营 养蛋白 β; NM_021907.4)	LCK(LCK 原癌基 因, Src 家族酪氨 酸激酶;	GOLGA8A(高尔基 体蛋白 A8 家族成 员 A; NM_001042771.2)	CYorf15B(滑行蛋 白 γ 假基因, Y 染 色体连锁的; NR_045128.1)	PRKY(蛋白激 酶, Y 染色体连 锁的, 假基因; NR_028062.1)	CTLA4(细胞毒性 T 淋巴细胞相关 蛋白 4; NM_005214.4)
FAM30A(具有序 列相似性的家族 30 成员 A; NR_026890.2)	NCALD(神经钙蛋 白 δ; NM_001040624.1)	ICOS(诱导型 T 细 胞共刺激因子; NM_012092.3)	DOCK9(胞质分裂 作用因子 9; NM_015296.2)	TBC1D5(TBC1 结 构域家族成员 5; NM_001134381.1)	DGK(二酰甘油 激酶 5; NM_004717.3)
FCRL2(Fc 受体样 2; NM_030764.3)	PRKCQ(蛋白激酶 Cθ; NM_006257.4)	ITM2A(整合膜蛋 白 2A; NM_004867.4)	FOXP1(叉头框 P1; NM_032682.5)	TBCD(微管蛋白 折叠辅因子 D; NM_005993.4)	DOK5(停靠蛋白 5; NM_018431.4)

GLDC(甘氨酸脱羧酶; NM_000170.2)	SH2D1A(含 SH2 结构域的蛋白 1A; NM_002351.4)	LRBA/ILPS 反应性 米色样错蛋白; NM_00119282.2)	FYB(FYN 结合蛋 白; NM_001465.4)	TRA(T 细胞受体 αδ 基因座; NG_001332.3)	DPP4(二肽基肽酶 4; NM_001935.3)
GNG7(G 蛋白亚 基 77; NM_052847.2)	SKAP1(src 激酶相 关磷蛋白 1; NM_001075099.1)	NAPIL4(核小体组 装蛋白 1 样 4; NM_005969.3)	HNRPH1(核不均 一核糖核蛋白 H1(HD); NM_001257293.1)	VIL2(埃及蛋白; NM_003379.4)	DUSP5(双特异性 磷酸酶 5; NM_004419.3)
HLA-DOB(主 要 组织相容性复合 物, II 类, DOB; NM_002120.3)	TRA(T 细胞受体 αδ 基因座; NG_001332.3)	NUP107(核孔蛋白 107; NM_020401.3)	INPP4B(肌醇多磷 酸-4-磷酸酶 II 型 B; NM_003866.3)	EGFL6(多 EGF 样 结构蛋白 6; NM_015507.3)	
HLA-DQAI(主要 组织相容性复合 物, II 类, DQa1; NM_002122.3)	TRAC(核受体共阻 遏物 2; NM_006312.5)	PHF10(PHD 指蛋 白 10; NM_018288.3)	KLF12(Kruppel 样 因子 12; NM_007249.4)		GGT1(γ-谷氨酰转 移酶 1; NM_013421.2)
IGHAI(免疫球蛋白 重链基因座; NG_001019.6)	TRAT1(T 细胞受 体相关跨膜衔接子 1; NM_016388.3)	PPP2R5C(蛋白磷 酸酶 2 调节亚基 B', γ; NM_001161725.1)	LOC20134(具有 序列相似性的家族 153 成员 B; NM_001265615.1)		HBEGF(肝素结合 EGF 样生长因 子; NM_001945.2)
IGHGI(免疫球蛋白 重链基因座; NG_001019.6)	TRBC1(T 细胞受 体 β 基因座; NG_001333.2)	RPA1(复制蛋白 AI; NM_002945.3)	MAP3K1(丝裂原 活化蛋白激酶激酶 1, E3 泛素蛋白连接酶; NM_005921.1)		IFNG(干扰素 γ; NM_000619.2)

IGHM(免渡球蛋白重链基因座; NG_001019.6)	SEC24C(SEC24 同系物 C, COPII 外壳蛋白复合物组分; NM_004922.3)	ML1(赖氨酸(K)特异性甲基转移酶 2A; NM_005933.3)	IL12RB2(白细胞介素 12 受体亚基 β2; NM_001319233.1)
IGKC(免渡球蛋白 κ 基因座, 邻近 V-簇和 J-C 簇; NG_000834.1)	SLC25A12(溶质载体家族 25 成员 12; NM_003705.4)	NEFL(神经丝轻链多肽; NM_006158.4)	IL22(白细胞介素 22; NM_020525.4)
IGL(免渡球蛋白 λ 基因座; NG_000002.1)	TRA(T 细胞受体 αδ 基因座; NG_001332.3)	NFATC3(活化 T 细胞核因子 3; NM_173165.2)	LRP8(LDL 受体相关蛋白 8; NM_017522.4)
KIAA0125(具有序列相似性的家族 30 成员 A; NR_026800.2)	UBE2L3(泛素缀合酶 E2 L3; NM_003347.3)	PCM1(中心粒外周物质 1; NM_001315507.1)	LRRN3(富含亮氨酸的重复神经元 3; NM_018334.4)
MEF2C(肌细胞增强因子 2C; NM_001308002.1)	YME1L1(YME1 样 1 ATP 酶; NM_001253866.1)	PCNX(山核桃素同系物 1; NM_014982.2)	LTA(淋巴毒素 α; NM_000595.3)
MICAL3(含微管相关单氧酶、钙调理蛋白和 LIM 结构域的蛋白 3; NM_001136004.3)		PDXDC2(含吡哆醛依賴性脱羧酶结构域的蛋白 2, 假基因; NR_003610.1)	SGCB(肌聚糖, β(43kDa 肌营养不良蛋白相关糖蛋白); NM_000232.4)
MS4A1(跨膜 4-结构域 A1; NM_021950.3)		PHC3(多同源异型同系物 3; NM_001308116.1)	SYNGR3(完触蛋白 3; NM_004209.5)

OSBPL10(氧化酶结合蛋白样蛋白 10; NM_017784.4)		POLR2J2(RNA 聚合酶 II 亚基 J2; NM_032959.5)	ZBTB32(含锌指和 BTB 结构域的蛋白 32; NM_014383.2)
PNOC(弧形肽前体; NM_001284244.1)		PSPC1(旁链组分 1; NM_001042414.2)	
QRSL1(谷氨酰胺酰基-tRNA 合酶(谷氨酰胺水解)-样蛋白 1; NM_018292.4)		REPS1(含 RALBP1 相关 Eps 结构域的蛋白 1; NM_001128617.2)	
SCN3A(电压门控钠通道 α 亚基 3; NM_001081677.1)		RPL11-74E24.2(含锌指 CCCH 型结构域的蛋白; NM_001271675.1)	
SLC15A2(溶质载体家族 15 成员 2; NM_017007074.1)		RPP38(核糖核酸酶 PMRP 亚基 P38; NM_001265601.1)	
SPIB(Spi-B 转录因子; NM_001244000.1)		SLC7A6(溶质载体家族 7 成员 6; NM_003983.5)	
TCL1A(T 细胞白血病/淋巴瘤 1A; NM_001098725.1)		SNRPN(小核核糖核蛋白多肽 N; NM_0222807.3)	
TNFRSF17(TNF 受体超家族成员		ST3GAL1(ST3 β -半乳糖苷 α -2,3-唾	

17; NM_001192.2)			液酸转移酶 1; NM_173344.2)	
		STX16(突触融合蛋白 16; NM_001204868.1)		
		TIMM8A(线粒体 内膜转位酶 8 同系 物 A; NM_001145951.1)		
		TRAF3IP3(TRAF3 相互作用蛋白 3; NM_001320144.1)		
		TXK(TXK 磷酸 激酶; NM_003328.2)		
		USP9Y(泛素特异 性肽酶 9, Y 染色 体连锁的; NG_008311.1)		

*每个 GenBank 登录号是所列基因的代表性或示例性 GenBank 登录号，并且出于所有目的将其通过引用以其整体并入本文。此外，每个列出的代表性或示例性登录号不应解释为将权利要求限制为特定登录号。

表 4A 续

细胞类型	Th2 细胞	TFH	Th17 细胞	TReg	CD8 T 细胞	Tgd	细胞毒性细胞
人类基 因(基因 名称; GenBank)	ADCY1(腺苷酸环化 酶 1; NM_001281768.1)	B3GAT1(β-1,3-葡萄糖 酸转移酶 1; NM_018644.3)	IL17A(白细胞介 素 17A; NM_002190.2)	FOXP3(叉头框 P3; NM_014009.3)	ABT1(基端转录激活 物 1; NM_013375.3)	Clorf61(染色体 1 开放阅读框 61; NM_006365.2)	APBA2(淀粉样 β 前体蛋白结合 族 A 成员 2; NM_005503.3)

登录号 *)	AH1(Abelson 辅助 整合位点 1; NM_001134831.1)	BLR1(c-N-c 酪化因子受 体类型 5; EF444957.1)	IL17RA(白细胞介 素 17 受体 A; NM_014339.6)	AES(分裂相关氨基 末端增强子(amino- terminal enhancer of split); NM_198969.1)	CD160(CD160 分 子; NM_007053.3)	APOL3(载脂蛋白 L3; NM_014349.2)
	A1582773(n17d08_x1 NCI CGAP_Bm25 智人 cDNA 克隆; A1582773.1)	C18orf1(含低密度脂蛋白 受体 A 类结构域的蛋 白 4; NM_181481.4)	RORC(RAR 相关 孤儿受体 C; NM_001001523.1)	APBA2(淀粉样 β 前 蛋白结合家族 A 成员 2; NM_001130414.1)	FEZ1(成员和延伸 蛋白 ζ ; AF123659.1)	CTSW(组织蛋白 酶 W; NM_001335.3)
	ANK1(端蛋白 1; NM_020476.2)	CDK5R1(细胞周期素依赖 性激酶 5 调节亚基 1; NM_003885.2)	CHGB(嗜铬粒蛋白 B; NM_001819.2)	ARHGAP8(Rho GTP 酶激活蛋白 8; NM_001198726.1)	TARP/TCR γ 变异 阅读框蛋白; NM_001003806.1)	DUSP2(双特异性 磷酸酶 2; NM_004418.3)
	BIRC5(含杆状病毒 IAP 重复的蛋白 5; NM_001012271.1)			C12orf47(MAPKAPK5 反义 RNA 1; NR_015404.1)	TRD/T 细胞受体 $\alpha\delta$ 基因座; NG_001332.3)	GNLY(颗粒溶 素; NM_012483.3)
	CDC25C(细胞分裂 周期 25C; NM_001318098.1)	CH3L2(几丁质酶 3 样 蛋白 2; NM_001025199.1)	CXCL13(C-X-C 基序趋 化因子配体 13; 7; NM_001134420.1)	C19orf6(跨膜蛋白 259; NM_001033026.1)	TRGV9/T 细胞受 体 YV 区域 9; X69385.1)	GZMA(颗粒酶 A; NM_006144.3)
	CENPF(着丝粒蛋白 F; NM_016343.3)	HEY1(具有 YRPW 基序 1 的 hes 相关家族 bHLH 转录因子; NM_001282851.1)	CXCL13(C-X-C 基序趋 化因子配体 13; 7; NM_001134420.1)	C4orf15(IAU5 augmin 样复合物亚基 3; NM_001303143.1)	C4orf15(IAU5 augmin 样复合物亚基 3; NM_001303143.1)	GZMH(颗粒酶 H; NM_0012270781.1)
	CXCR6(杀伤细胞 巢素样受体 B1; NM_002258.2)	HIST1H4K(组蛋白簇 1 H4 家族成员 k; NM_003541.2)		CAMLG(钙调节配 体; NM_001745.3)	CAMLG(钙调节配 体; NM_001745.3)	KLRB1(杀伤细胞 巢素样受体 B1; NM_002258.2)
	DHFR(二氢叶酸还 原酶; NM_001290354.1)	ICA1(胰岛细胞自 身抗 原 1; NM_001136020.2)		CD8A(CD8a 分子; NM_001768.6)	CD8A(CD8a 分子; NM_001768.6)	KLRD1(杀伤细胞 巢素样受体 D1; NM_001114396.1)
				CD8B(CD8b 分子; NM_001178100.1)	CD8B(CD8b 分子; NM_001178100.1)	KLRF1(杀伤细胞 巢素样受体 F1; NM_001291822.1)

EVIS(脊髓灰质炎病毒结合位点 5; NM_001308248.1)	KCNK5(双孔域钾通道 亚家族 K 成员 5; NM_003740.3)	CDKN2AIP(CDKN2A 相互作用蛋白; NM_001317343.1)	KLRK1(杀伤细胞 凝集素样受体 K1; NM_007560.3)
GATA3(GATA 端合 蛋白 3; NM_001002295.1)	KIAA1324(KIAA1324; NM_001284353.1)	DNAIB1(Dna I 核珠壳 蛋白家族(Hsp40)成员 B1; NM_001313964.1)	NKG7(自然杀伤 细胞颗粒蛋白 7; NM_005601.3)
GSTA4(谷胱甘肽 S- 转移酶 aa4; NM_001512.3)	MAF(MAF bZIP 转录因 子; NM_001031804.2)	FLT3L(Gfms 相关酶 氨酸激酶 3 配体; NM_001278538.1)	RORA(RAR 相关 孤儿受体 A; NM_134262.2)
HELLS(解旋酶, 淋 巴细胞特异性; NM_001289074.1)	MAGEH1(MAGE 家族 成员 H1; NM_014061.4)	GADPH5A(生长停滞 和 DNA 损伤诱导型 蛋白 q; NM_001199742.1)	RUNX3(runt 相关 转录因子 3; NM_004350.2)
IL26(白细胞介素 26; NM_018402.1)	MKL2(MKL1)心肌素样 2; NM_014048.4)	GZMM1(颗粒酶 M; NM_001258351.1)	SIGIRR(含单 Ig 和 TIR 结构域的 蛋白; NM_001135054.1)
LAR2(白细胞相关 免疫球蛋白样受体 2; NM_021270.4)	MYO6(肌球蛋白 VI; NM_001300899.1)	KLF9(Kruppel 样因子 9; NM_001206.2)	WHDC1L1(与肌 动蛋白相关的 WAS 蛋白质同系 物, 高尔基体膜 和微管蛋白基 3; NR_003521.1)
LIMA1(LIM 结构域 和肌动蛋白结合蛋 白 1; NM_001243775.1)	MYO7A(肌球蛋白 VIIA; NM_001127179.2)	LEPROTIL1(激素受体 重叠转录物样蛋白 1; NM_001128208.1)	ZBTB16(含锌指 和 BTB 结构域的 蛋白 16; NM_001018011.1)
MB(肌红蛋白; NM_203377.1)	PASK(含 PAS 结构域的 丝氨酸/苏氨酸激酶; NM_001252119.1)	LIM41(Lck 相互作用 跨膜衔接子 1; NM_017806.3)	

MICAL2(含微管相关单氨基、钙调理蛋白和LIM结构域的蛋白2; NM_001282663.1)	PDCD1(程序性细胞死亡 1; NM_005018.2)	MYST3(MYST 组蛋白乙酰转移酶(单核细胞白血病)3; NM_006766.4)
NEIL3(nei样 DNA 糖基化酶 3; NM_018248.2)	POMT1(蛋白质 O-甘露糖转糖酶 1; NM_001136114.1)	PF4(血小板因子 4; NM_002619.3)
PHEX(磷酸盐调节内肽酶同系物, X; 染色体连锁的; NM_000444.5)	PTPN13(蛋白酪氨酸磷酸酶, 非受体类型 13; NM_080685.2)	PP1PR2(蛋白磷酸酶 1 调节抑制剂亚基 2; NM_001291504.1)
PMCH(原黑色素浓集激素; NM_002674.3)	PVALB(小清蛋白; NM_001315532.1)	PRF1(穿孔素 1; NM_005041.4)
PTGIS(2'合酶; NM_000961.3)	SH3TC1(SH3 结构域和三角形四肽重复蛋白 1; NM_018896.4)	PRRS5(富脯氨酸蛋白 5; NM_181333.3)
SLC39A14(溶质载体家族 39 成员 14; NM_001135153.1)	SIRPG(信号调节蛋白; NM_018556.3)	RBM3(tRNA 结合基序 (RNPI, RRM) 蛋白 3; NM_006743.4)
SMAD2(SMAD 家族成员 2; NM_001135937.2)	SLC7A10(溶质载体家族 7 成员 10; NM_019849.2)	SFI1(剪接因子 1; NM_004630.3)
SNRPD1(小核核糖核蛋白 D1 多肽; NM_001291916.1)	SMAD1(SMAD 家族成员 1; NM_001003668.1)	SFRS7(富含丝氨酸和精氨酸的剪接因子 7; NM_001031684.2)
WDHD1(WD 重复和 HMG-盒 DNA 结合蛋白 1; NM_001008396.2)	ST8SIA1(ST8α-N-乙酰神经氨酸苷 α-2,8-唾液酸转糖酶 1; NM_001304450.1)	SLC16A7(溶质载体家族 16 成员 7; NM_001270622.1)

STK39(丝氨酸/苏氨酸激酶 39; NM_013233.2)		TBCC(微管蛋白折量 辅因子 C; NM_003192.2)
THADA(THADA, 含纯 丝重复的蛋白; NM_001271644.1)		THUMPD1(含 THUMPD 结构域的蛋 白 1; NM_017736.4)
TOX(胸腺细胞选择相 关高迁移率族蛋白; NM_014729.2)		TMC6(跨膜通道样蛋 白 6; NM_001321185.1)
TSHR(促甲状腺激素受 体; NM_000369.2)		TSC22D3(TSC22 结构 域家族成员 3; NM_001318470.1)
ZNF764(锌指蛋白 764; NM_001172679.1)		VAMP2(囊泡相关膜 蛋白 2; NM_014232.2)
		ZEB1(锌指 E 合成合 同酶 1; NM_001128128.2)
		ZFP36L2(ZFP36 环指 蛋白样蛋白 2; NM_006887.4)
		ZNF22(锌指蛋白 22; NM_006963.4)
		ZNF609(锌指蛋白 609; NM_015042.1)
		ZNF91(锌指蛋白 91; NM_001300951.1)

*每个 GenBank 登录号是所列基因的代表性或示例性 GenBank 登录号，并且出于所有目的将其通过引用以其整体并入本文。此外，每个列出的代表性或示例性登录号不应解释为将权利要求限制为特定登录号。

表 4B 用于在本文提供的方法中使用的先天性免疫细胞(IIC)基因标签免疫标记物。

细胞类型

人类基因 (基因名 称; GenBank 登录号*)	NK 细胞	NK CD56 阳性细胞	NK CD56 明细细胞	DC	iDC
				CCL13(C-C 基序趋化因子配体 13; NM_005408.2)	ABCG2(ATP-结合盒, 亚家族 G(WHITE)成员 2(初级血型); NM_001257386.1)
	ADARB1(腺苷脱氨酶, RNA 特异性 B1; NM_001112)	EDG8(鞘氨醇-1-磷酸 盐受体 5; NM_00116215.1)	BG255923(溶血磷脂 酰胆碱酰基转移酶 4; NM_153613.2)	CCL13(C-C 基序趋化 因子配体 13; NM_005408.2)	ABCG2(ATP-结合盒, 亚家族 G(WHITE)成员 2(初级血型); NM_001257386.1)
	AF107846(神经内分泌 特异性高亲基体蛋白 P55; AF107846.1)	FLJ20699(cDNA FLJ20699 fis, 克隆 KAIA2372; AK000706.1)	DUSP4(双特异性磷酸 酶 4; NM_057158.3)	CCL17(C-C 基序趋化 因子配体 17; NM_002987.2)	BLVRB(胆绿素还原酶 B; NM_000713.2)
55	AL080130(cDNA DKFZp434E033(来自克 隆 DKFZp434E033); AL080130.1)	GTF3C1(通用转录 因子 IIIC 亚基 1; NM_001286242.1)	FOXJ1(叉头框 J1; NM_001454.3)	CCL22(C-C 基序趋化 因子配体 22; NM_002990.4)	CARD9(半胱天冬酶 家族成员 9; NM_052814.3)
	ALDH1B1(醛脱氨酶 1 家族成员 B1; NM_000692.4)	GZMB(颗粒酶 B; NM_004131.4)	MADD(MAP 激酶激 活死亡结构域; NM_001135944.1)	CD209(CD209 分子; NM_001144899.1)	CD1A(CD1a 分子; NM_001763.2)
	ARL6IP2(atlastin GTP 酶 2; NM_001330461.1)	IL21R(白细胞介素 21 受体; NM_181079.4)	MPPE1(含金属磷酸 酯酶结构域的蛋白 1, mRNA; NM_001044370.1)	HSD11B1(羟基类固醇 11-β 脱氢酶 1; NM_001206741.1)	CD1B(CD1b 分子; NM_001764.2)
	BCL2(细胞凋亡调节物 (BCL2); NM_000633.2)	KIR2DL3(杀伤细胞 免疫球蛋白样受 体, 两个 Ig 结构域 和长胞质尾 3; NM_015868.2)	MUC3B(粘蛋白 3B 胞表面相关; JQ511939.1)	NPR1(利钠肽受体 1; NM_000906.3)	CD1C(CD1c 分子; NM_001765.2)
	CDC5L(细胞分裂周期 5 样; NM_001253.3)	KIR2DS1(杀伤细胞 免免疫球蛋白样受 体, 两个 Ig 结构域	PPFIBP2(PPFIA 结合 蛋白 2; AY630619.1)	PPFIBP2(PPFIA 结合 蛋白 2; XR_930917.2)	CD1E(CD1e 分子; NM_001185115.1)

FGF18(成纤维细胞生长因子 18; NM_003862.2)	和短胞质尾 1; NM_014512.1)	KIR2DS2(杀伤细胞免渡球蛋白样受体, 两个 Ig 结构域和短胞质尾 2; NM_001291700.1)	PLA2G6(磷脂酶 A2 组 VI; NM_001004426.1)	CH25H(胱氨酸 25-羟化酶; NM_003956.3)
FUT5(岩藻糖转移酶 5; NM_002034.2)	KIR2DS5(杀伤细胞免渡球蛋白样受体, 两个 Ig 结构域和短胞质尾 5; NM_014513.2)	RRAD(Ras 相关糖酵解抑制剂和钙通道调节剂; NM_001128850.1)		CLEC10A(C 型凝集素结构域家族 10 成员 A; NM_001330070.1)
FZRI(起泡/细胞分裂周期 20 相关蛋白 1; XM_005259573.4)	KIR3DL1(杀伤细胞免渡球蛋白样受体, 三个 Ig 结构域和长胞质尾 1; NM_013289.2)	SEPT6(隔蛋白 6; NM_145802.3)		CSF1R(集落刺激因子 1 受体; NM_001288705.1)
GAGE2(G 抗原 2; NM_001127212.1)	KIR3DL2(杀伤细胞免渡球蛋白样受体, 三个 Ig 结构域和长胞质尾 2; NM_006737.3)	XCL1(N-C 基序趋化因子配体 1; NM_002995.2)		CTNS(胱氨酸转运蛋白 (cystinosin), 溶酶体胱氨酸转运蛋白; NM_001031681.2)
IGFBP5(胰岛素样生长因子结合蛋白 5; NM_000599.3)	KIR3DL3(杀伤细胞免渡球蛋白样受体, 三个 Ig 结构域和长胞质尾 3; NM_153443.4)			F13A1(因子 XIII 亚基; AH002691.2)

LDB3(LIM 结构域结合蛋白 3; NM_001171611.)	KIR3DS1(杀伤细胞免免疫球蛋白样受体，三个 Ig 结构域和短胞质尾 1; NM_001083539.2)	FABP4(脂肪酸结合蛋白 4; NM_001442.2)
LOC643313(类似于假设蛋白 LOC284701; XM_933043.1)	SPON2(脊椎蛋白 2; NM_001199021.1)	FZD2(卷曲类受体 2; NM_001466.3)
LOC730096(假设蛋白 LOC730096; NC_000022.9)	TMEPAI(前列腺跨膜雄激素诱导蛋白 1; NM_199169.2)	GSTT1(谷胱甘肽 S-转移酶 θ1; NM_001293814.1)
MAPRE3(微管相关蛋白 RP/EB 家族成员 3; NM_001303050.1)		GUCA1A(鸟苷环化酶激活物 1A; NM_001319062.1)
MCM3AP(微小染色体维持复合物组分 3 相关蛋白; NM_003906.4)		HS3ST2(硫酸乙酰肝素(葡萄糖胺)3-O-磺基转移酶 2; NM_006043.1)
MRC2(甘露糖受体 C 类型 2; NM_006039.4)		LMAN2L(凝集素甘露糖结合 2 样; NM_001322355.1)
NCR1(自然细胞毒性触发受体 1; NM_001242357.2)		MMP12(基质金属蛋白酶 12; NM_002426.5)
NM_014114(Pro00097 蛋白质; NM_014114.1)		MS4A6A(跨膜结构域 4 亚家族 A6A; NM_001330275.1)
NM_014274(瞬时受体电位阳离子通道亚家族		NM_021941(染色体 21 开放阅读框 97; NM_021941.1)

V 成员 6; NM_014274.3)		NUD19(nudix 水解酶 9; NM_001248011.1)
NM_017616(KN 基序和 锚蛋白重复结构域 2; NM_015493.6)		PPARG(过氧化物酶体 增殖物激活受体 γ ; NM_005037.5)
PDLIM4(PDZ 和 LIM 结构域 4; NM_003687.3)		PREP(脯氨酸内肽酶; NM_002726.4)
PRX(periaxin; NM_020956.2)		RAP1GAP(RAP1 GTP 酶激活蛋白; NM_001330383.1)
PSMD4(蛋白酶体 26S 亚基, 非 ATP 酶 4; NM_001330692.1)		SLC26A6(溶质载体家 族 26 成员 6; NM_001281733.1)
RP5-886K2.1(神经丝蛋 白 AD7c-NTP; AF010144.1)		SLC7A8(溶质载体家 族 7 成员 8; NR_049767.1)
SLC30A5(溶质载体家族 30 成员 5; NM_001251969.1)		SYT17(突触结合蛋白 17; NM_001330509.1)
SMEK1(蛋白磷酸酶 4 调节亚基 3A; NM_001284280.1)		TACSTD2(肿瘤相关钙 信号传感器 2; NM_002353.2)
SPN(速福林; NM_003123.4)		TM7SF4(树突细胞表达 的七种跨膜蛋白; NM_001257317.1)
TBXA2R(血栓素 A2 受 体; NM_001060.5)		

TCTN2(结构蛋白家族成员(tectonic family member)2; NM_001143850.2)			VASH1(血管生成抑制蛋白 1; NM_014909.4)
TINAGL1(肾小管间质性肾炎抗原样蛋白 1; NM_001204415.1)			
XCL1(X-C 基序趋化因子配体 1; NM_002995.2)			
XCL2(X-C 基序趋化因子配体 2; NM_003175.3)			
ZNF205(锌指蛋白 205; NM_001278158.1)			
ZNF528(锌指蛋白 528; NM_032423.2)			
ZNF747(锌指蛋白 747; NM_023931.3)			

*每个 GenBank 登录号是所列基因的代表性或示例性 GenBank 登录号，并且出于所有目的将其通过引用以其整体并入本文。此外，每个列出的代表性或示例性登录号不应解释为将权利要求限制为特定登录号。

表 4B 续

细胞类型	aDC	pDC	嗜酸性粒细胞	巨噬细胞	肥大细胞	嗜中性粒细胞
人基因 因 (基因名 称; NM_002981)	CCL1 (趋化因子(C-C 基序)配体 1; NM_002981)	IL3RA(白细胞介 素 3 受体亚基 α; NM_001267713.1)	ABHD2(含 α/β 水 解酶(abhydrolase) 结构域的蛋白 2; NM_007011.7)	APOE(载脂蛋白 E; NM_001302691.1)	ABCC4(ATP 结合盒 亚家族 C 成员 4; NM_001301829.1)	ALPL(碱性磷酸 酶, 肝脏/骨/肾 脏; NM_001127501.3)

GenBank 登录号 EBI3(Epstein-Barr 病毒诱导蛋白 3; NM_005755.2)*	ACACB(乙酰-CoA 羧化酶 β; NM_001093.3)	ATG7(自体吞噬相关蛋白 7; NM_001144912.1)	ADCYAPI(腺苷酸环化酶激活多肽 1; NM_0011174)	BST1(骨髓基质细胞抗原 1; NM_004334.2)
INDO(吲哚胺-吡咯 2,3 双加氧酶; AY221100.1)	C9orf156(tRNA 甲基转移酶 O; NM_001330725.1)	BCA11(支链氨基酸转氨酶 1; NM_001178094.1)	CALB2(钙结合蛋白 2; NM_001740.4)	CD93(CD93 分子; NM_012072.3)
LAMP3(溶酶体相关膜蛋白 3; NM_014398.3)	CAT(过氧化氢酶; NM_001752.3)	CCL7(C-C 基序趋化因子配体 7; NM_006273.3)	CEACAM8(癌胚抗原相关细胞粘附分子 8; NM_001816.3)	CEACAM3(癌胚抗原相关细胞粘附分子 3; NM_001277163.2)
OAS3(2'-5'-寡腺苷酸合酶 3; NM_006187.3)	CCR3(C-C 基序趋化因子受体 3; NM_178329.2)	CD163(CD163 分子; NM_203416.3)	CMA1(糜酶 1, 肥大细胞; NM_001308083.1)	CREB5(cAMP 反应元件结合蛋白 5; NM_001011666.2)
	CLC(夏-莱-二氏晶体质(Charcot-Leyden crystal)半乳糖凝集素; NM_001828.5)	CD68(CD68 分子; NM_001040059.1)	CPA3(羧肽酶 A3; NM_001870.3)	CRISPLD2(含 LCCL 结构域的富含半胱氨酸的分泌蛋白 2; NM_031476.3)
	CYSL1R2(半胱氨酸白三烯受体 2; NM_001308471.1)	CD84(CD84 分子; NM_001184881.1)	CTSG(组织蛋白酶 G; NM_001911.2)	CSF3R(集落刺激因子 3 受体; NM_172313.2)
	EMR1(含 EGF 样模块的粘蛋白样激素受体样蛋白 1; DQ217942.1)	CT3L1(几丁质酶 3 样蛋白 1; NM_001276.2)	ELA2(嗜中性粒细胞弹性蛋白酶; EU617980.1)	CYP4F3(细胞色素 P450 家族 4 亚家族 F 成员 3; NM_001199209.1)

说 明 书

	EPN2(epsin 2; NM_001102664.1)	CHT1(几丁质酶 1; NM_001270509.1)	GATA2(GATA 结合 蛋白 2; NM_001145661.1)	DYSF(dysferlin; NM_001130455.1)
	GALC(半乳糖神 经酰胺酶; NM_000153.3)	CLEC5A(C 型凝集 素结构域家族 5 成 员 A; NM_001301167.1)	HDC(C型凝集素; NM_002112.3)	FCAR(IgA 受体的 Fc 片段; NM_133278.3)
	GPR44(孤儿 G 蛋 白偶联受体; AF118265.1)	COL8A2(胶原类型 VIIIa 2 链; NM_001294347.1)	HPGD(羟基前列腺 素脱氢酶 15- (NAD); NM_001256307.1)	FCGR3B(IgG 受体 IIIb 的 Fc 片段; NM_001271035.1)
	HES1(hes 家族 bHLH 转录因子 1; NM_005524.3)	COLEC12(胶原凝 集素亚家族成员 12; NM_130386.2)	KIT(KIT 原癌基因 受体酪氨酸激酶; NM_000222.2)	FLJ11151(假设蛋白 FLJ11151; BC006289.2)
	HIST1H1C(组蛋白 H1 家族成员 c; NM_005319.3)	CTSK(组织蛋白酶 K; NM_000396.3)	LOC339524(长基因 间非蛋白编码 RNA 1140; NR_026985.1)	FPR1(甲酰肽受体 1; NM_001193306.1)
	HRII4(组胺受体 H4; NM_001143828.1)	CXCL5(C-X-C 基 序趋化因子配体 5; NM_002994.4)	LOH11ICR2A(BCSC- 1 同种型; AY366508.1)	FPR1(甲酰肽受体 样受体; M84562.1)
	IGSF2(免疫球蛋白 超家族成员 2; BC130327.1)	CYTBB(细胞色素 b-245 β 链; NM_000397.3)	MAOB(单胺氧化酶 B; NM_000898.4)	G0S2(G0/G1 期开 关调节蛋白 2; NM_015714.3)
	IL5RA(白细胞介 素 5 受体亚基 α; NM_001243099.1)	DNASE2B(脱氧核 糖核酸酶 2β; NM_058248.1)	MIPI(黑素亲和 素; NM_001042467.2)	ITST1H2BC(组蛋白 白簇 1 H2B 家族成 员 c; NM_003526.2)

KBTBD11(含kelch 重复和 BTB 结构域的蛋白 1; NM_014867.2)	EMPI(上皮细胞膜蛋白 1; NM_001423.2)	MPO(髓过氧化物酶; NM_000250.1)	HPSE(乙酰肝素酶; NM_001098540.2)
KCNH2(电压门控钾通道亚家族 H(eag 相关)成员 2; NM_000238.3)	FDX1(铁氧化还原蛋白 1; NM_004109.4)	MS4A2(跨膜结构域 4 亚家族 A2; NM_001256916.1)	IL8RA(白细胞介素 8 受体 α; L19591.1)
LRPS1(LDL _r 受体相关蛋白 5 样蛋白; NM_001135772.1)	FN1(纤连蛋白 1; NM_001306131.1)	NM_003293(类胰蛋白酶 α/β1; NM_003294.3)	IL8RB(白细胞介素 8 受体类型 B; U11878.1)
MYO15B(肌球蛋白 XVIB; NM_001309242.1)	GM2A(GM2 神经节苷脂激活物; NM_000405.4)	NR0B1(核受体亚家族 0 组 B 成员 1; NM_000475.4)	KCNJ15(电压门控钾通道亚家族成员 15; NM_001276438.1)
RCOR3(REST 共阻遏物 3; NM_001136224.2)	GPC4(磷脂酰肌醇蛋白聚糖 4; NM_001448.2)	PGDS(造血前列腺素 D 合酶; NM_014485.2)	KIAA0329(含结构蛋白 β 螺旋桨重复的蛋白 2(tectonin beta-propeller repeat containing 2); NM_014844.4)
RNASE2(核糖核酸酶 A 家族成员 2; NM_002934.2)	KAL1(anosmin 1; NM_000216.3)	PPM1H(蛋白磷酸酶, Mg ²⁺ /Mn ²⁺ 依赖性 1H; NM_020700.1)	LILRB2(白细胞免疫球蛋白样受体 B2; NR_103521.2)

RNU2(U2 snRNA; U57614.1)	MARCO(具有胶原 结构的巨噬细胞受 体; NM_006770.3)	PRG2(蛋白聚糖 2, 原嗜酸性粒细胞主 要碱性蛋白; NM_001302927.1)	MGAM(麦芽糖酶- 葡萄糖淀粉酶; NM_004668.2)
RRP12(核糖体 RNA 加工蛋白 12 同系物; NM_001284337.1)	ME1(苹果酸酶 1; NM_002395.5)	PTGS1(前列腺素内 过氧化物合酶 1; NM_000962.3)	MME(膜金属内肽 酶; NM_007289.2)
SLAHI(siah E3 泛 素蛋白连接酶 1; NM_003031.3)	MS4A4A(跨膜结 构域 4 亚家族 A4A; NM_001243266.1)	SCG2(分泌颗粒蛋白 II; NM_003469.4)	PDE4B(磷酸二酯酶 4B; NM_001297440.1)
SMPD3(鞘磷脂酶- 酸二酯酶 3; NM_018667.3)	MSR1(巨噬细胞清 道夫受体 1; NM_138716.2)	SIGLEC6(唾液酸结 合性 Ig 样凝集素 6; NM_19845.5)	S100A12(S100 钙结 合蛋白 A12; NM_005621.1)
SYNJ1(突触小泡 磷酸酶 1; NM_001160302.1)	PCOLCE2(前胶原 C-内肽酶增强子 2; NM_013363.3)	SLC18A2(溶质载体 家族 18 成员 A2; NM_003054.4)	SIGLEC5(唾液酸结 合性 Ig 样凝集素 5; NM_003830.3)
TGFBI(TGFB 诱 导因子同源框 1; NM_174886.2)	PTGDS(前列腺素 D2 合酶; NM_000954.5)	SLC24A3(溶质载体 家族 24 成员 3; NM_020689.3)	SLC22A4(溶质载体 家族 22 成员 4; NM_003059.2)
THBS1(血小板反 应蛋白 1; NM_003246.3)	RAII4(视黄酸诱导 蛋白 14; NM_001145525.1)	TALL1(T 细胞急性淋 巴细胞白血病 1 蛋 白; X51990.1)	SLC25A37(溶质载体 家族 25 成员 37; NM_001317812.1)
THBS4(血小板反 应蛋白 4; NM_001306213.1)	SCARB2(清道夫 受体 B 类成员 2; NM_001204255.1)	TPSAB1(类胰蛋白 酶 α/β ; NM_003294.3)	TNFRSF10C(TNFT 受体超家族成员 10c; NM_003841.3)

TIPARP(TCDD 诱导型聚(ADP-核 糖)聚合酶; NM_001184718.1)	SCG5(分泌粒蛋白 V; NM_001144757.2)	TPSB2(类胰蛋白酶 β2; NM_024164.5)	VNN3(血管非炎症 蛋白 3(vanin 3); NM_001291703.1)
TKILL(转氨酶酶 样蛋白 1; NM_001145934.1)	SGMS1(鞘磷脂合 酶 1; NM_147156.3)	SULT1C2(磺基转 移酶家族 1C 成员 2; NM_176825.2)	

* 每个 GenBank 登录号是所列基因的代表性或示例性 GenBank 登录号，并且出于所有目的将其通过引用以其整体并入本文。此外，每个列出的代表性或示例性登录号不应解释为将权利要求限制为特定登录号。

表5. 用于在本文提供的方法中使用的个体免疫标记物。

基因名称	缩写	GenBank登录号*
------	----	-------------

程序性死亡配体1	PDL1	NM_014143
程序性死亡配体2	PDL2	AY254343
程序性细胞死亡1	PDCD1	NM_005018
细胞毒性T淋巴细胞相关蛋白4	CTLA4	NM_005214

*每个GenBank登录号是所列基因的代表性或示例性GenBank登录号,并且出于所有目的将其通过引用以其整体并入本文。此外,每个列出的代表性或示例性登录号不应解释为将权利要求限制为特定登录号。

表6. 用于在本文提供的方法中使用的干扰素(IFN)基因。

基因名称	缩写	GenBank登录号*
趋化因子(C-X-C基序)配体10	CXCL10	NM_001565
C-X-C基序趋化因子配体9	CXCL9	NM_002416
干扰素 α 诱导蛋白27	IFI27	NM_001130080
三角形四肽重复干扰素诱导蛋白1	IFIT1	NM_001548
三角形四肽重复干扰素诱导蛋白2	IFIT2	NM_001547
三角形四肽重复干扰素诱导蛋白3	IFIT3	NM_001549
MX发动蛋白样GTP酶1	MX1	NM_001144925
MX发动蛋白样GTP酶2	MX2	XM_005260983
2'-5'-寡腺苷酸合成酶1	OAS1	NM_016816
2'-5'-寡腺苷酸合成酶2	OAS2	NM_016817
信号转导子和转录激活因子1	STAT1	NM_007315
信号转导子和转录激活因子2	STAT2	NM_005419

*每个GenBank登录号是所列基因的代表性或示例性GenBank登录号,并且出于所有目的将其通过引用以其整体并入本文。此外,每个列出的代表性或示例性登录号不应解释为将权利要求限制为特定登录号。

表7用于在本文提供的方法中使用的MHC II类基因。

名称	缩写	GenBank登录号*
CD74	智人 (Homo sapiens) CD74分子	NM_001025159
名称	缩写	GenBank登录号*
	(CD74)	
CIITA	II类主要组织相容性复合物 反式激活因子	NM_001286402
CTSH	组织蛋白酶H	NM_004390
HLA-DMA	智人主要组织相容性复合物, II 类, DM α	NM_006120
HLA-DPA1	智人主要组织相容性复合物, II 类, DP α 1	NM_033554
HLA-DPB1	人类 MHC II类淋巴细胞抗原 (HLA-DP) β 链	M83664
HLA-DQA1	智人主要组织相容性复合物, II 类, DQ α 1	NM_002122
HLA-DRB1	智人主要组织相容性复合物, II 类, DR β 1	NM_002124
HLA-DRB5	智人主要组织相容性复合物, II 类, DR β 5	NM_002125
HLA-DRB6	智人主要组织相容性复合物, II 类, DR β 6	NR_001298
NCOA1	智人核受体共激活物1	NM_003743

*每个GenBank登录号是所列基因的代表性或示例性GenBank登录号,并且出于所有目的将其通过引用以其整体并入本文。此外,每个列出的代表性或示例性登录号不应解释为将权利要求限制为特定登录号。

在一个实施方案中,在使用如本文提供的任何方法和分类器生物标记物小组或其子集确定患者的AD肺癌亚型(单独或者组合确定如本文提供的一种或多种免疫细胞标记物的表达)时,选择患者以用免疫治疗剂进行治疗或给予免疫治疗剂。该免疫治疗剂可以是检查点抑制剂、单克隆抗体、生物反应调节剂、治疗性疫苗或细胞免疫疗法。

在另一个实施方案中,免疫治疗剂是检查点抑制剂。在一些情况下,提供了一种用于确定对一种或多种检查点抑制剂有反应的可能性的方法。在一个实施方案中,该检查点

抑制剂是PD-1/PD-L1检查点抑制剂。该PD-1/PD-L1检查点抑制剂可以是纳武单抗、派姆单抗、阿特珠单抗(atezolizumab)、度伐单抗、拉立珠单抗(lambrolizumab)或阿维鲁单抗(avelumab)。在一个实施方案中,该检查点抑制剂是CTLA-4检查点抑制剂。该CTLA-4检查点抑制剂可以是伊匹单抗(ipilimumab)或tremelimumab。在一个实施方案中,该检查点抑制剂是检查点抑制剂的组合,如例如,与一种或多种CTLA-4检查点抑制剂组合使用的一种或多种PD-1/PD-L1检查点抑制剂的组合。

在一个实施方案中,免疫治疗剂是单克隆抗体。在一些情况下,提供了一种用于确定对一种或多种单克隆抗体有反应的可能性的方法。单克隆抗体可以针对肿瘤细胞或针对肿瘤产物。单克隆抗体可以是帕尼单抗、马妥珠单抗、necitumunab、曲妥珠单抗、阿麦妥昔单抗(amatuximab)、贝伐单抗、雷莫芦单抗、巴维昔单抗、patritumab、rilotumumab、西妥昔单抗、immu-132或登西珠单抗(demcizumab)。

在又另一个实施方案中,免疫治疗剂是治疗性疫苗。在一些情况下,提供了一种用于确定对一种或多种治疗性疫苗有反应的可能性的方法。该治疗性疫苗可以是肽或肿瘤细胞疫苗。该疫苗可以靶向MAGE-3抗原、NY-ESO-1抗原、p53抗原、存活素抗原或MUC1抗原。治疗性癌症疫苗可以是GVAX (GM-CSF基因转染的肿瘤细胞疫苗)、belagenpumatucel-L(用TGF- β 2反义质粒修饰的四个经辐射的NSCLC细胞系制备的同种异体肿瘤细胞疫苗)、MAGE-A3疫苗(由MAGE-A3蛋白和佐剂AS15构成)、(1)-BLP-25抗MUC-1(靶向肿瘤细胞上表达的MUC-1)、CimaVax EGF(由与载体蛋白质偶联的人重组表皮生长因子(EGF)构成的疫苗)、WT1肽疫苗(由四个维尔姆斯(Wilms)肿瘤抑制基因类似物肽构成)、CRS-207(编码人间皮素的活减毒单核细胞增多性李斯特菌载体)、Bec2/BCG(诱导抗GD3抗体)、GV1001(靶向人端粒酶逆转录酶)、tergenpumatucel-L(由经遗传修饰以包括免疫系统对其强烈反应的小鼠基因的人肺癌细胞构成)、TG4010(靶向MUC1抗原)、racotumomab(模仿在多种人类癌症上表达的NGcGM3神经节苷脂的抗独特型抗体)、tecemotide(脂质体BLP25;由MUC1的串联重复区域制成的基于脂质体的疫苗)或DRibbles(由九种癌症抗原加TLR佐剂制成的疫苗)。

在一个实施方案中,免疫治疗剂是生物反应调节剂。在一些情况下,提供了一种用于确定对一种或多种生物反应调节剂有反应的可能性的方法。该生物反应调节剂可能引发炎症,如例如PF-3512676(CpG 7909)(toll样受体9激动剂)、CpG-ODN 2006(下调Treg)、卡介苗(Bacillus Calmette-Guerin,BCG)、母牛分枝杆菌(SRL172)(现在经常作为佐剂进行测试的非特异性免疫刺激剂)。该生物反应调节剂可以是细胞因子疗法,如例如IL-2+肿瘤坏死因子 α (TNF- α)或干扰素 α (诱导T细胞增殖)、干扰素 γ (诱导肿瘤细胞凋亡)或Mda-7(IL-24)(Mda-7/IL-24诱导肿瘤细胞凋亡并抑制肿瘤血管生成)。该生物反应调节剂可以是集落刺激因子,如例如粒细胞集落刺激因子。该生物反应调节剂可以是多模式效应物,如例如多靶VEGFR:沙利度胺及类似物如来那度胺和泊马度胺、环磷酰胺、环孢菌素、地尼白介素(denileukin diftitox)、talactoferrin、trabectedin或全反式视黄酸。

在一个实施方案中,免疫疗法是细胞免疫疗法。在一些情况下,一种用于确定对一种或多种细胞治疗剂有反应的可能性的方法。细胞免疫治疗剂可以是树突细胞(DC)(离体产生的负载肿瘤抗原的DC疫苗)、T细胞(离体产生的淋巴因子激活的杀伤细胞;细胞因子诱导的杀伤细胞;活化的T细胞; $\gamma\delta$ T细胞)或自然杀伤细胞。

在一些情况下,AD的特定亚型具有不同水平的免疫激活(例如,先天性免疫力和/

或适应性免疫力),使得具有升高的或可检测的免疫激活(例如,先天性免疫力和/或适应性免疫力)的亚型被选择以接受用本文所述的一种或多种免疫治疗剂进行的治疗。在一个实施方案中,与其他AD亚型或肺癌亚型相比,AD的PP亚型具有低的免疫激活(例如,先天性免疫力和/或适应性免疫力)。在一些情况下,AD的特定亚型具有高水平或升高水平的免疫激活。在一些情况下,与其他AD亚型或肺癌亚型相比,AD的PI亚型具有升高水平的免疫激活(例如,先天性免疫力和/或适应性免疫力)。在一个实施方案中,不选择具有低水平免疫激活或无免疫激活(例如,先天性免疫力和/或适应性免疫力)的AD亚型用于接受用本文所述的一种或多种免疫治疗剂进行的治疗。

检测方法

在一个实施方案中,本文提供的方法和组合物允许检测从受试者获得的肺癌样品(例如,腺癌肺癌样品)中的至少一种核酸。该至少一种核酸可以是本文提供的分类器生物标记物。在一个实施方案中,使用本文提供的方法和组合物检测的该至少一种核酸选自表1。在一个实施方案中,检测从受试者获得的肺癌样品中的一种或多种核酸(例如,分类器生物标记物)的方法包括使用本文提供的任何方法测量至少一种或多种生物标记物的表达水平,基本上由其组成或由其组成。该生物标记物可以选自表1。在一些情况下,该多种生物标记物核酸包含表1中的至少两种生物标记物核酸、至少8种生物标记物核酸、至少16种生物标记物核酸、至少24种生物标记物核酸、至少32种生物标记物核酸或所有48种生物标记物核酸,基本上由其组成或由其组成。该检测可以是在核酸水平。该检测可以通过使用本文公开的任何扩增、杂交和/或测序测定。

在另一个实施方案中,本文提供的方法和组合物允许检测从受试者获得的肺癌样品(例如,腺癌肺癌样品)中的至少一种核酸或多种核酸,使得该至少一种核酸或该多种核酸选自表1中列出的生物标记物;并且允许从一组生物标记物中检测至少一种生物标记物,该生物标记物的存在、不存在和/或表达水平指示免疫激活。用于指示免疫激活的一组生物标记物可以是适应性免疫细胞(AIC)(例如,表4A)和/或先天性免疫细胞(IIC)(例如,表4B)的基因表达标签、个体免疫生物标记物(例如,表5)、干扰素基因(例如,表6)、主要组织相容性复合物II类(MHC II)基因(例如,表7)或其组合。IIC和AIC两者的基因表达标签可以是本领域已知的任何基因标签,如例如Bindea等人(Immunity 2013;39(4):782-795)中列出的基因标签。该检测可以是在核酸水平。该检测可以通过使用本文公开的任何扩增、杂交和/或测序测定。

试剂盒

可以进一步提供用于实践本发明的方法的试剂盒。“试剂盒”可以涵盖包含用于特异性检测本发明的生物标记物的表达的至少一种试剂(例如,抗体、核酸探针或引物等)的任何制品(例如,包装或容器)。该试剂盒可以作为用于执行本发明方法的单元进行促销、分发或销售。另外,该试剂盒可以包含描述该试剂盒及其使用方法的包装说明书。

在一个实施方案中,提供了用于实践本发明方法的试剂盒。此类试剂盒与手动和自动免疫细胞化学技术(例如,细胞染色)兼容。这些试剂盒包含针对感兴趣的生物标记物的至少一种抗体、用于检测与生物标记物结合的抗体的化学品、复染剂和任选地上蓝剂(以促进阳性染色细胞的鉴定)。检测抗原-抗体结合的任何化学品都可以用于本发明的实践中。该试剂盒可以包含至少2种、至少3种、至少4种、至少5种、至少6种、至少7种、至少8种、至

少9种、至少10种、或更多种抗体,用于在本发明的方法中使用。

实施例

通过参考以下实施例来进一步展示本发明。但是,应注意的是这些实施例,像以上所述的实施方案一样,是说明性的并且不以任何方式构成对本发明范围的限制。

实施例1-肺腺癌内在亚型之间的免疫细胞活化差异和与CD274 (PD-L1) 表达的可变相关性。

介绍

基于基因表达的肺腺癌(AD) 亚型分型将AD肿瘤分类为具有不同生物学和临床特征的不同亚型。基于基因表达的亚型分型一致地鉴定了3种不同的肺AD生物学类型:末端呼吸单元型(TRU) (以前为支气管型)、近端增生型(PP) (以前为巨型) 和近端炎症型(PI) (以前为鳞状型) (1,2) (参见图1)。AD亚型展示了基因组改变、肿瘤驱动因素、预后以及可能对各种疗法的反应的关键差异(1-2)。

方法

使用以前公布的Bindea等人(3)免疫细胞基因标签(总共24种)和AD亚型分型基因表达标签(1-2),针对与AD亚型相关的免疫细胞特征检测了几个可公开获得的肺AD数据集(2,4和5)和1个最近收集的基因表达数据集(参见图2)。对按照亚型的免疫差异的这种研究使用了来自Bindea等人[3]的24种免疫细胞基因标签,该24种免疫细胞基因标签的每种标签都有不同数量的基因并被分类为适应性或先天性免疫细胞标签(参见表4A-表4B)。适应性免疫细胞(AIC)标签(表4A)包括T细胞、中枢记忆T细胞(Tcm)、效应记忆T细胞(Tem)、T辅助细胞(Th)、1型T辅助细胞(Th1)、2型T辅助细胞(Th2)、T滤泡辅助细胞(Tfh)、T辅助17细胞(Th17)、T调节细胞(Treg)、 γ δ T细胞(Tgd)、CD8 T细胞、细胞毒性T细胞、B细胞,并且先天性免疫细胞(IIC)标签(表4B)包括自然杀伤(NK)、NK CD56暗细胞、NK CD56亮细胞、树突细胞(DC)、未成熟树突细胞(iDC)、树突细胞(pDC)、活化树突细胞(aDC)、肥大细胞、嗜酸性粒细胞、巨噬细胞和嗜中性粒细胞。除了先天性免疫细胞(IIC)和适应性免疫细胞(AIC)两者的基因表达标签以外,在3种AD亚型(TRU、PP和PI)中还检测了13种基因IFN标签(IFN;表6)、13种基因MHC II类标签得分(Forero[6];表7)以及表5中的单基因免疫生物标记物(CTLA4、PDCD1、CD274 (PD-L1)、和PDCDLG2 (PD-L2))。

AD数据集包括如上所述的几种可公开获得的肺癌基因表达数据集和新收集的福尔马林固定的石蜡包埋(FFPE)肺肿瘤样品(n=88)的腺癌数据集。新收集的AD数据集的88个福尔马林固定的石蜡包埋(FFPE)样品是在北卡罗来纳大学教堂山分校(University of North Carolina at Chapel Hill, UNC-CH)根据批准的IRB方案收集的存档残留肺肿瘤样品。在RNA提取之前,宏观切开FFPE样品切片(3个10 μ m切片)。使用Illumina's的RNA-Access试剂盒(San Diego, CA)以100ng/样品的输入进行转录组富集的RNAseq。使用hg19作为参考比对序列数据,并使用cufflinks构建转录组(Trapnell2010)。使用cuff compare来注释转录组并计算基因表达计数。

对于AD,使用4个已公布的和1个最近收集的肺腺癌样品的基因表达数据集(即GeneCentric表达数据集),总共1278个患者样品。已公布的数据集包括TCGA[2]、Shedden等人[4]、Tomida等人[5]、和Wilkerson等人[1],源自新鲜冷冻样本。GeneCentric表达数据集

源自于福尔马林固定的石蜡包埋(FFPE)样本。对于TCGA,从Firehose下载上四分位数归一化RSEM数据并进行log2变换。来自Shedden等人[4]的Affymetrix Cel文件从caIntegrator网站下载,并使用R中的Affy包生成强大的多阵列平均表达量度。对于Tomida等人[5](GSE13213)和Wilkerson等人[1](GSE26939),从Gene Expression Omnibus(GEO)网站下载归一化的Agilent阵列数据。

为了确定腺癌亚型(TRU、PP和PI),使用如先前在Wilkerson等人[1]中所述的公布的506种基因最近质心分类器。在标签中的基因中值中心之后,为每个样品分配对应于与其最大相关的质心的亚型。(皮尔逊)

使用腺癌的TCGA数据,通过绘制成对Spearman秩相关系数的矩阵来评估30种标记物之间的相关性,其中标记物通过分级聚类排序(参见图4)。为了按照亚型研究总体免疫标记物趋势,绘制表达热图,其中样品按亚型排列并且标记物根据Bindea等人[3]中的排序进行分组(参见图3)。为了评价亚型之间免疫标记物差异的再现性,针对每个数据集按照亚型绘制归一化T细胞标签(参见图5)。

使用线性回归评价免疫细胞标签与肿瘤亚型和与CD274表达的关联。更具体地,为了评估亚型作为免疫标记物预测因子相对于PD-L1的预测因子的预测强度,在TCGA数据集中拟合每个标签用亚型作为唯一预测因子和再用PD-L1作为唯一预测因子的线性回归模型。PD-L1表达被视为低/中/高分类变量,其在每组中具有相等比例。检查当亚型是预测因子时调整的R平方相对于当PD-L1是预测因子时调整的R平方的散点图的总体趋势(参见图6)。

使用来自TCGA腺癌的补充信息(Lawrence 2013)中可获得的每Mb数据的非沉默突变负荷,分别使用Kruskal-Wallis检验和Spearman相关系数研究突变负荷-T细胞表达的关联。对于TCGA腺癌,从Firehose下载STK11 CNV和突变状态,并使用Fisher精确检验评价STK11失活-亚型的关联。这里,当一个样品被报告为已缺失和/或已突变时,它被称作失活的。为了测试AD中的STK11在针对亚型调整后是否显示关联的证据,将T细胞表达的线性模型与AD中的失活的STK11作为唯一预测因子拟合,并且在针对亚型调整后再次拟合。

使用Kruskal-Wallis检验研究亚型和免疫标签与13种基因MHC II类标签的关联[Forero[6];表7](计算为列表(表7)中的所有基因的平均值)。对于免疫标签-MHC II类关联,计算Spearman相关系数。

还分析了免疫标签的分级聚类和成对标签相关性。使用分层cox比例风险模型评价第I-III阶段样品的存活标签关联,从而允许每个数据集中的不同基线风险。更具体地,使用Cox比例风险模型,在每种亚型内对TCGA数据集中的免疫标记物-存活关联进行总体和单独测试。将免疫标记物集中并按比例缩放以具有平均值0和方差1,并且排除第IV阶段患者。针对阶段调整的特定亚型内的评价,并针对阶段和亚型调整总体评价。制作显示每种标签的风险比和置信区间的森林图(参见图7A和图7B)。使用R 3.2.0软件(<http://www.R-project.org>)进行所有统计分析。

结果

免疫细胞基因标签的热图分析和无监督分级聚类提供了AD的内在亚型的分离(参见图3和图4)。对免疫细胞基因标签(AIC和IIC两者)以及个体免疫基因标记物的检测揭示了AD亚型之间的明显差异(参见图3)。在AD中,对于所检测的大多数细胞类型,PP亚型中的

免疫表达始终较低。大多数T细胞的表达在TRU和PI中相似,但是可以通过在TRU亚型中一些先天性免疫细胞(树突细胞、NK CD56亮细胞、肥大细胞、嗜酸性粒细胞)和几种适应性免疫细胞(B细胞、TFH、Tcm、Th17、CD8 T细胞)的更高表达在TRU与PI之间作出区分,然而PI亚型显示出Th1和Th2、Treg、细胞毒性T细胞和NKCD56暗细胞的更高表达(可以在图18中找到按照AD亚型的所有免疫细胞和标记物的箱形图)。免疫疗法靶向CTLA4和CD274 (PD-L1),证明在多个数据集中在PI亚型中的表达始终较高(箱形图补充图18)。在PP肿瘤中,相对于其他AD,适应性免疫细胞和先天性免疫细胞表达以及免疫疗法靶标表达都被抑制。(图18)。

总体而言,免疫激活在AD的PI亚型中最为突出,证明了先天性免疫细胞和适应性免疫细胞的激活。相反,AD的PP亚型表现出较低的免疫激活。

使用分级聚类,相关性矩阵揭示了适应性免疫细胞和先天性免疫细胞的聚类(参见图4)。在AD中,适应性免疫特征例如T细胞、细胞毒性细胞、CD8细胞、Th1细胞、PDCD1、CTLA4和Treg具有高的成对相关性,并且类似地,对于先天性免疫细胞(包括iDC、DC、巨噬细胞、嗜中性粒细胞、肥大细胞和嗜酸性粒细胞)是相关的(图4)。

与AD亚型相比,进行了CD274 (PD-L1) 表达与适应性免疫细胞标签的关联强度分析。如图6中所示,对于AD亚型,关联强度(调整的R平方)是混合的,表明对于一些细胞(B细胞、T细胞、Th1、Treg、细胞毒性细胞、T辅助细胞、Tem、Tgd)CD274关联更大,而对于其他细胞(TFH、Th2、CD8、Th17和Tcm)AD亚型关联更大。

免疫细胞标签主要在TCGA数据集中评价,然而发现如通过免疫细胞标签测量的,AD亚型免疫差异在多个数据集中是非常可再现的(参见图5)。AD亚型中的T细胞免疫细胞标签表达亚型差异在多个基因表达数据集中具有显著的可再现性,该基因表达数据集源自冷冻样品和FFPE样品两者并且涉及多个基因表达平台包括RNAseq (Illumina, San Diego, CA) 和来自Affymetrix (Santa Clara, CA) 和Agilent (Santa Clara, CA) 两者的微阵列。总体而言,免疫细胞标签基因表达模式在多个AD(参见图5)数据集中是一致的。

TCGA AD数据中的非沉默突变负荷按照亚型而不同,PI显示负荷最高,而TRU负荷最低(图19)。富含TP53突变的PI亚型与免疫细胞表达升高相关联,然而,尽管具有相对高的免疫表达,TRU具有最低的突变负荷。在AD数据集中,突变负荷与T细胞免疫细胞表达不强烈相关(在AD中Spearman相关性=-0.07)。

几个其他基因组特征例如AD中STK11的缺失(Cao[7], Shabath[8], Koyama[9])已被认为是NSCLC中免疫应答降低的可能原因。在低免疫应答腺癌PP亚型中富含STK11失活。AD中的STK11失活与较低的免疫细胞表达相关,然而在使用线性回归针对亚型调整之后,STK11不是显著的预测因子(在针对亚型调整后AD中的STK11 $p=0.0007$ 至 $p=0.43$)。

使用公布的13种基因MHC II类标签(Forero[6])研究AD肺癌中的免疫细胞表达与MHC II类基因的关联。MHC II类基因表达与AD中的几种免疫细胞强烈相关,包括T细胞表达(AD中的Spearman相关性=0.66)、B细胞表达(AD中的Spearman相关性=0.5)和DC表达(AD中的Spearman相关性=0.69)。在肿瘤邻近正常肺组织中的MHC II类基因表达与肿瘤相比显著更高,并且在肿瘤亚型中差异表达(图19)。在作为T细胞免疫细胞表达预测因子的MHC II类标签的线性模型中,在针对AD亚型调整后MHC II类仍然显著(对于MHC II, $p<1E-50$)。

使用cox比例风险模型,计算增加一个表达单位的亚型特异性风险比(HR)。针对病理阶段调整亚型特异性HR,并计算置信区间(CI)。在图7A-图7B中显示了对于至少一种亚型

显著(标称p值<0.05)的标记物的风险比和置信区间。在图7A-图7B中显示了针对一种或多种亚型显示显著存活关联的细胞标签或基因的HR和CI。对于AD亚型,许多先天性和适应性免疫细胞、CD274 (PD-L1) MHC II类标签和CTLA4的表达的单位增加与AD的PI亚型中的存活改善显著相关,但在其他亚型中没有显著相关(图7A-图7B)。总体而言,免疫细胞标签的存活分析表明T辅助17和T滤泡辅助免疫细胞预测AD中的存活改善($p<0.001$) (参见图7A-图7B)。

结论

肺AD基因表达亚型在其免疫景观方面不同。AD的内在生物学亚型揭示了免疫细胞活化的关键差异,其不总是与CD274表达相关并且展现出与存活的可变关联。AD PP亚型显示出最小的免疫浸润(抑制的免疫细胞表达),表明对immunoRX的应答降低。AD PI亚型显示出与改善的存活相关的免疫特征表达。此外,非沉默突变负荷与亚型之间的免疫细胞表达不相关;然而,MHC II类基因表达高度相关。增加的免疫和MHC II基因表达与AD的TRU和PI亚型的存活改善相关。

通过引用并入

出于所有目的,以下参考文献通过引用以其整体并入。

- 1.) Wilkerson MD, et al. PLoS One 2012;7(5):e36530. PMID 22590557
- 2.) TCGA Lung AdenoC. Nature 2014;511(7511):543-550. PMID 25079552
- 3.) Bindea et al., Immunity 2013;39(4):782-95. PMID 24138885
- 4.) Shedden K. et al. Nat Med 2008;14(8):822-827. PMID 18641660
- 5.) Tomida S, et al. J Clin Oncol 2009;27(17):2793-99. PMID 19414676
- 6.) Forero A, Li Y, Dongquan C, et al. Expression of the MHC class II pathway in triple negative breast cancer tumor cells is associated with a good prognosis and infiltrating lymphocytes. Cancer Immunol Res 2016;4(5):390-399.
- 7.) Cao C, Gao R, Zhang M, et al. Role of LKB1-CRTC1 on glycosylated COX-2 and response to COX-2 inhibition in lung cancer. J Natl Cancer Inst. 2015;107(1):1-11.
- 8.) Shabath MB, Welsh EA, Fultp WJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation, and immune surveillance in lung adenocarcinoma. Oncogene. 2015;1-8.
- 9.) Koyama S, Akbey EA, Li Y, et al. STK11/LKB1 deficiency promotes neutrophil recruitment and proinflammatory cytokine production to suppress T-cell activity in the lung tumor microenvironment. Cancer Res 2016;76(5):999-1008.

实施例2-肺腺癌亚型分型标签的开发和验证

背景

几项基因组研究已经证明了在其基因组谱(包括基因表达、突变谱和拷贝数目变化)方面可能有差异的三种不同的内在肺腺癌亚型[1-3]。三种生物学AD亚型TRU、PP和PI不仅在其基因组特征上不同,而且在临床特征方面也显示出潜在的重要差异[1-4]。AD的基因

表达亚型可以显示出在肿瘤分化、远端复发的可能性、阶段特异性存活、潜在的肿瘤驱动因子和炎症反应方面的显著差异[1-4]，并且可能不容易通过基于标准形态学的技术(显微镜和免疫组织化学)区分。还提出了对化学疗法[2]、培美曲塞[5]和/或EGFR抑制剂疗法的潜在反应差异[2]。在末端呼吸单元(TRU)亚型中证实了EGFR过表达的富集[2,3]。与LKB1/STK11缺失结合的更高的KRAS突变频率，更可能发生在近端增生(PP)亚型中[2,3]。TP53突变和免疫基因激活是近端炎症(PI)亚型的标志[2-4]。初步数据可能证明TRU亚型中对EGFR抑制剂的反应增强、PP亚型中对化学疗法的反应增强、TRU亚型中的培美曲塞反应增强、以及PI亚型中对免疫疗法的潜在反应的可能性[2-6]。新兴数据表明，按照基因表达亚型的AD分类可以提供有价值的信息，从而补充药物靶标突变检测并为肺癌患者管理提供信息。

目的

肺腺癌(AD)亚型分型主要限于如下研究方案，该研究方案涉及：从新鲜冷冻肺肿瘤中提取RNA，随后使用超过500种基因的定量基因表达来应用最近质心(nearest centroid)预测因子。尽管有腺癌亚型分型的预后和预测益处的证据，但是对新鲜冷冻组织的需求，需要>500种基因的基因表达结合复杂的生物信息学分析，阻碍了AD亚型分型在药物开发和/或临床中的应用。这项研究的目的是开发用于区分腺癌的三种亚型(末端呼吸单元型(TRU)，以前称为支气管型；近端增生型(PP)，以前称为巨型；和近端炎症型(PI)，以前称为鳞状型)的稳健而有效的基因标签(需要较少的基因)。新的有效基因标签可以用于从新鲜冷冻或FFPE肿瘤样品中可靠地对AD亚型分型，使其适用于使用任何可获得的定量RNA平台(qRT-PCR、RNAseq、Affymetrix或Agilent阵列)进行诊断应用和/或药物开发。用于区分腺癌亚型的48种基因标签的开发如本文的方法部分中所述进行。

方法

使用515个肺腺癌癌症基因组图集(TCGA)RNAseq数据集用于训练和506种基因分类器用于定义金标准亚型，开发了48种基因标签，其在应用于几个独立测试集时保持低错误分类率。从标准的506种分类器基因开始，使用至最近质心阵列分类(CLaNC)[7]算法加以修改用于对每种亚型选择相等数量的负相关基因和正相关基因。使用TCGA肺腺癌数据集进行基于5折交叉验证曲线选择包括在标签中的基因的最佳数目(每种亚型16个)(参见图8)。图9中显示了用于训练预测因子的原型样品的选择，其中为了获得48种基因的最终列表，将CLaNC应用于整个TCGA数据集减去20%的具有最低金标准亚型预测强度的样品，从每种亚型中去除相等数量(图9)。然后在几个新鲜冷冻公开可获得的阵列和RNAseq数据集[2,8,9]中测试48种基因标签，并将结果与如由先前公布的506种基因标签[2]定义的金标准亚型称谓进行比较。然后在存档的FFPE腺癌样品的新收集的RNAseq数据集中进行48种基因标签(表1)的最终验证，以确保在FFPE样品中具有可比的性能。

为了验证所选的48种基因标签的一致性能，新收集的FFPE样品是肺腺癌(AD)残留存档样品(主要是手术样品)，这些样品是在北卡罗来纳大学教堂山分校(NC)的IRB批准方案下收集的。由病理学家审查样品的肿瘤细胞，并且在提取之前宏观切开三个10 μ m组织切片以富集肿瘤细胞。将RNA定量并且每个样品输入100ng。使用Illumina RNA-Access试剂盒构建富集转录组的测序文库。通过使用BA分析仪对测序文库进行质量控制，并使用qPCR进行定量。在Illumina RNA-Access平台(50bp PE, 20-30百万读数)上生成序列数据，并且通过使用fastQC进行质量控制。使用STAR比对器将序列结果与hg19参考序列比对，并使用

Cufflinks构建转录组[10]。使用cuff compare来注释转录组并计算各种表达基因的计数。根据癌症基因组图集肺腺癌分析中使用的方法,将RSEM表达计数估计值进行上四分位数归一化和log2转化[3,11]。

结果

表2中显示了该研究中开发的48种基因标签基因列表,而每个AD亚型的48种基因标签基因列表的T统计量可以在表1中找到。分别在图10、图11和图12中显示了为每种AD亚型(支气管型、巨型、鳞状型)选择的16种基因的中值基因表达。在图13中显示了使用48种基因标签的亚型称谓与在几个不同的测试数据集中公布的506种基因标签亚型称谓的一致性。新开发的48种基因标签表明在新收集的FFPE数据集中的一致性为0.84,并且在其他3个新鲜冷冻的测试数据集中为0.79-0.92的范围。以下是测试数据集、RNA平台类型和所使用的腺癌样品数目的汇总。

参考文献	RNA平台	腺癌样品
TCGA腺癌	RNAseq	515
Shedden等人	Affymetrix阵列	442
Tomida等人	Agilent阵列	117
新收集的UNC FFPE样品	RNAseq	88

结论

描述了用于AD亚型分型的有效的48种基因标签的开发和验证。当应用于几个独立的测试集时,所得的48种基因标签保持低错误分类率。因此,新标签从新鲜冷冻或FFPE肿瘤样品中可靠地对AD进行亚型分型,并且可以使用从包括RNAseq和阵列在内的各种平台生成的基因表达数据可靠地执行。

通过引用并入

出于所有目的,以下参考文献通过引用以其整体并入。

1.) Hayes DN, Monti S, Parmigiani G, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J Clin Oncol* 2006;24(31):5079-5090.

2.) Wilkerson M, Yin X, Walter V, et al. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS ONE* 2012;7(5) e36530. Doi:10.1371/journal.pone.0036530.

3.) Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511.7511(2014):543-550.

4.) Ringner M, Jonsson G, Staaf J. Prognostic and Chemotherapy Predictive Value of Gene-Expression Phenotypes in Primary Lung Adenocarcinoma. *Clin Cancer Research* 2016;22(1):218-29.

5.) Fennell DA, Myrand SP, Nguyen TS, Ferry D, Kerr KM, et al. Association between Gene Expression Profile and Clinical Outcome of Pemetrexed-Based Treatment in Patients with Advanced Non-Small Cell Lung Cancer: Exploratory Results from a Phase II study. *PLOS one* 2014;September14 9(9):e107455.

6.) Skoulidis F, Byers LA, Diao L, Papadimitrakopoulou VA, Tong P, et al. Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. *Cancer Discov* 2015; Aug 5(8):860-77.

7.) Dabney AR. ClaNc: Point-and-click software for classifying microarrays to nearest centroids. *Bioinformatics*. 2006; 22:122-123. doi:10.1093/bioinformatics/bti756

8.) Shedden K, Taylor JMG, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study: director's challenge consortium for the molecular classification of lung adenocarcinoma. *Nat Med* 2008; 14(8):822-827. doi:10.1038/nm.1790.

9.) Tomida S, Takeuchi T, Shimada Y, Arima C, Maatsuo K, et al. Relapse-Related Molecular Signature Identifies Patients With Dismal Prognosis. *J Clin Oncol* 2009; 27(17):2793-99.

10.) Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 2010; 28(5):511-5.

11.) Li B, and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011, 12: 323 doi:10.1186/1471-2105-12-323

实施例3-如使用来自实施例2的肺腺癌亚型分型48种基因标签所确定的肺腺癌内在亚型之间的免疫细胞活化差异。

方法

使用以前公布的Bindea等人(3)免疫细胞基因标签(总共24种)和实施例2中描述的用于对AD亚型分型的肺AD亚型分型基因标签,针对与如使用实施例2中描述的肺AD基因标签所确定的AD亚型相关的免疫细胞特征检测了几个可公开获得的肺AD数据集(1-2,4-5;参见图2)。在3种AD亚型(TRU、PP和PI)中检测了先天性免疫细胞(IIC)和适应性免疫细胞(AIC)两者的基因表达标签、13种基因IFN标签(IFN)、以及单基因免疫生物标记物(CTLA4、PDCD1、以及CD274(PD-L1)、PDCDLG2(PD-L2))。使用线性回归评价免疫细胞标签与肿瘤亚型和与CD274表达的关联。还分析了免疫标签的分级聚类和成对标签相关性。使用分层cox比例风险模型评价第I-III阶段样品的存活标签关联,从而允许每个数据集中的不同基线风险。

结果

使用实施例2的TCGA AD数据集和48种基因AD亚型分型标签,免疫细胞基因标签的热图分析和无监督分级聚类以与实施例1中观察到的相似的方式提供AD的内在亚型的分离(参见图3和图14)。此外,与在实施例1中观察到的相似,免疫细胞标签基因表达模式在多个AD(参见图15)数据集中是一致的(参见图5)。与AD亚型相比,进行了CD274(PD-L1)表达与适应性免疫细胞标签的关联强度分析。如图16中所示(像实施例1的图6中一样),对于AD亚型,

关联强度(调整的R平方)是混合的,表明对于一些细胞(B细胞、T细胞、Th1、Treg、细胞毒性细胞、T辅助细胞、Tem、Tgd)CD274关联更大,而对于其他细胞(TFH、Th2、CD8、Th17和Tcm)AD亚型关联更大。如在实施例1中,免疫激活在AD的PI亚型中最突出,而AD的PP亚型表现出较低的免疫激活,并且AD亚型和CD274表达类似地预测AIC表达(参见图6和图16)。

使用cox比例风险模型,计算增加一个表达单位的亚型特异性风险比。针对病理阶段调整亚型特异性HR,并计算置信区间。在图17中显示了针对一种或多种亚型显示显著存活关联的细胞标签或基因的HR和CI。对于AD亚型,像在实施例1中一样,许多先天性和适应性免疫细胞、CD274 (PD-L1) 和CTLA4的表达的单位增加与AD的PI亚型中的存活改善显著相关,但在其他亚型中没有显著相关(参见图7A-图7B和图17)。总体而言,像实施例1中一样,免疫细胞标签的存活分析表明T辅助17和T滤泡辅助免疫细胞预测AD中的存活改善($p < 0.001$) (参见图7A-图7B和图17)。

结论

在显示肺AD亚型在其免疫景观中如何有差异方面,实施例2中描述的AD亚型分型的48种基因标签显示出与实施例1中使用的一种或多种AD亚型分型基因标签相似的结果。与实施例1的AD亚型分型基因标签一致,该实施例中使用的AD亚型分型基因标签显示肺AD基因表达亚型在其免疫景观中有差异。AD的内在生物学亚型揭示了免疫细胞活化的关键差异,其不总是与CD274表达相关并且展示与存活的可变关联。AD PP亚型显示出最小的免疫浸润,表明对immunoRX的应答降低。AD PI亚型显示出与改善的存活相关的免疫特征表达。

通过引用并入

出于所有目的,以下参考文献通过引用以其整体并入。

- 1.) Wilkerson MD, et al. PLoS One 2012;7(5):e36530. PMID 22590557
- 2.) TCGA Lung AdenoC. Nature 2014;511(7511):543-550. PMID 25079552
- 3.) Bindea et al., Immunity 2013;39(4):782-95. PMID 24138885
- 4.) Shedden K, et al. Nat Med 2008;14(8):822-827. PMID 18641660
- 5.) Tomida S, et al. J Clin Oncol 2009;27(17):2793-99. PMID 19414676
- 6.) Lee ES, et al. Cancer Res 2008;14(22):7397-7404. PMID 19010856
- 7.) Raponi M, et al. Cancer Res 2006;66(7):466-72. PMID 16885343

实施例4-肺腺癌的表达亚型揭示了不同的免疫景观和独特的体细胞遗传特征,表明对多种药物靶标的差异性反应

介绍:基于基因表达的肺腺癌(AD)亚型分型将AD肿瘤分类为具有可变结果和对疗法有潜在反应的不同亚型。基于基因表达的亚型分型一致地鉴定了3种不同的肺AD生物学类型:末端呼吸单元型(TRU)(以前为支气管型)、近端增生型(PP)(以前为巨型)和近端炎症型(PI)(以前为鳞状型)(1,2)(参见图1)。AD亚型展示了基因组改变、肿瘤驱动因素、预后以及可能对各种疗法的反应的关键差异(1-2)。

方法:作为实施例1中进行的实验的后续,在使用图2中所示的TCGA肺癌基因表达数据集(AD n=515)²确定的来自实施例1的肺AD亚型中评价差异性药物靶基因表达。使用基因表达模式在实施例1中定义了先前公布的AD亚型(TRU、PP和PI)。在该实施例中,作为对免疫细胞基因标签(Bindea等人24种免疫细胞类型)⁴、单一免疫基因生物标记物(CTLA4、PDCD1 (PD-1) 和CD274 (PD-L1))的表达、增殖(11种基因标签;参见表9)⁵、和在实施例1中完

成的非沉默突变负荷的检测的补充,根据来自实施例1的AD亚型检测出来自临床肿瘤学实体瘤突变小组(322种基因,参见表8)³的基因的可变表达。使用具有Bonferroni校正Kruskal-Wallis(KW)检验评估差异性基因表达,而将线性回归和Spearman相关性用于评价非沉默突变负荷、肿瘤亚型和CD274(PD-L1)表达与免疫细胞表达的关联。

结果:如图21中所示,观察到AD亚型中208个/322个肿瘤小组基因(65%)的可变表达(KW Bonferroni阈值p<0.000155)。大多数药物靶基因(包括但不限于AD中的AURKA、CHEK1、ROS1、CD274(PD-L1)、CSF1R和ERBB4)在各亚型中表现出强烈的差异性表达(p<1E-28)。此外,从表10可以看出,来自临床肿瘤学实体瘤突变小组的322种基因中的前25种基因在AD亚型之间显示出差异性基因表达。免疫细胞表达在亚型之间也是高度可变的(参见图3)。AD的PI亚型表现出最大的免疫细胞表达,而AD的PP亚型表现出低的免疫细胞表达(参见图3)。非沉默突变负荷与免疫细胞表达没有强烈的相关性(在AD中,Spearman相关性=-0.07),然而,AD的PI亚型(其富含TP53突变)与升高的免疫细胞表达和高的突变负荷相关(参见图20)。总体而言,如图20中所示,在增殖、非沉默突变负荷和关键药物靶标CD274(PD-L1)、PDCD1(PD-1)、和CTLA4方面存在显著的AD亚型差异。

结论:肺AD的分子亚型在临床实体瘤测序小组中包括的大多数关键药物靶基因的表达方面有差异。肺AD的分子亚型揭示了宿主免疫应答和免疫靶标的差异性表达。应该单独地并且与免疫细胞特征和关键突变靶标组合地研究亚型作为潜在生物标记物的药物敏感性的评价。

通过引用并入

出于所有目的,以下参考文献通过引用以其整体并入。

- 1.) Wilkerson MD, et al. PLoS One 2012;7(5):e36530. PMID 22590557
- 2.) TCGA Lung AdenoC. Nature 2014;511(7511):543-550. PMID 25079552
- 3.) Foundation Medicine Solid Tumor Mutation Panel accessed October 2014.
- 4.) Bindea et al., Immunity 2013;39(4):782-95. PMID 24138885
- 5.) Neilson T0, et al. Clin Cancer Res 2010;16(21):522-5232. PMID 20837693.

实施例5:肺腺癌的表达亚型揭示了不同的免疫景观和独特的体细胞遗传特征,表明对多种药物靶标的差异性反应

介绍:就像实施例4中一样,该实施例的目的是评估临幊上重要基因在先前定义的腺癌(AD)基因表达亚型中的差异性表达。与使用如实施例1中描述的TCGA肺癌基因表达数据集(AD n=515)²进行基于基因表达的AD亚型分型的实施例4相反,使用实施例2中描述的48个基因集进行本实施例中基于基因表达的AD亚型分型。此外,临幊上重要的基因是322种基因(参见表8),其构成用于管理肿瘤学患者的临床实体瘤突变测序小组以鉴定影响治疗管理的基因组改变和/或确定靶向药物临床试验的合格性。就像实施例4中一样,还使用11种基因增殖标签(参见表9)评估在AD亚型之间的肿瘤增殖的差异。

方法:使用TCGA肺癌基因表达数据集(腺癌(AD) n=515)¹,在肺AD亚型中评价差异性药物靶基因表达。如前所述的(最近质心预测)²使用Clanc48 AD subtyper(参见实施例2和本文所述)来定义在AD中的亚型。检测了AD亚型末端呼吸单元型(TRU)、近端增生型(PP)

和近端炎症型 (PI)。根据AD, 检测来自临床肿瘤学实体瘤突变小组 (322种基因)³的基因的可变表达。使用具有Bonferroni校正的Kruskal-Wallis (KW) 检验评估差异性基因表达。此外, 增殖得分计算为11种基因PAM50增殖标签⁴中可获得基因的平均表达 ($\log_2(RSEM+1)$)。使用Kruskal-Wallis检验来测试亚型-增殖的关联。

结果:与图21相似,图22显示了观察到的AD亚型之间的203个/322个肿瘤小组基因 (63%) 的可变表达 (KW Bonferroni阈值 $p < 0.000155$)。此外,就像实施例4中的图20中一样,存在显著的AD亚型增殖差异(参见图23)。此外,从表11中所见的来自临床肿瘤学实体瘤突变小组的322种基因中的前25种基因在AD亚型之间显示出差异性基因表达与表10中发现的非常相似。

结论:就像实施例4中一样,肺AD的分子亚型在临床实体瘤测序小组中包括的大多数关键药物靶基因的表达方面有差异。肺AD的分子亚型揭示了宿主免疫应答和免疫靶标的差异性表达。

通过引用并入

出于所有目的,以下参考文献通过引用以其整体并入。

- 1.) TCGA Lung AD. *Nature* 2014;511(7511):543-550. PMID 25079552
- 2.) Wilkerson MD, et al. *PLoS One* 2012;7(5):e36530. PMID 22590557
- 3.) Foundation Medicine Solid Tumor Mutation Panel accessed October 6, 2014.
- 4.) Neilson T0, Parker JS, Leung S, et al. *Clin Cancer Res* 2010;16(21):5222-5232. PMID 20837693

表8.临床实体瘤突变测序小组的322种基因³

ABL1	C11orf30 (EMSY)	DDR2	FGFR4	IL7R	MET	PIK3CA	SDHD	TSHR
ABL2	CARD11	DICER1	FH	INHBA	MITF	PIK3CB	SETD2	U2AF 1

ACVR 1B	CBFB	DNMT3 A	FLCN	INPP4B	MLH1	PIK3CG	SF3B1	VEGF A
AKT1	CBL	DOT1L	FLT1	IRF2	MPL	PIK3R1	SLIT2	VHL
AKT2	CCND1	EGFR	FLT3	IRF4	MRE11A	PIK3R2	SMAD2	WISP 3
AKT3	CCND2	EP300	FLT4	IRS2	MSH2	PLCG2	SMAD3	WT1
ALK	CCND3	EPHA3	FOXL2	JAK1	MSH6	PMS2	SMAD4	XPO1
AMER 1 (FAM1 23B)	CCNE1	EPHA5	FOXP1	JAK2	MTOR	POLD1	SMARC A4	ZBTB 2
APC	CD274	EPHA7	FRS2	JAK3	MUTYH	POLE	SMARC B1	ZNF2 17
AR	CD79A	EPHB1	FUBP1	JUN	MYC	PPP2R1A	SMO	ZNF7 03
ARAF	CD79B	ERBB2	GABRA6	KAT6A (MYST3)	MYCL (MYCL1)	PRDM1	SNCAIP	ETV4
ARFR P1	CDC73	ERBB3	GATA1	KDM5A	MYCN	PREX2	SOCS1	ETV5
ARID1 A	CDH1	ERBB4	GATA2	KDM5C	MYD88	PRKAR1 A	SOX10	ETV6
ARID1 B	CDK12	ERG	GATA3	KDM6A	NF1	PRKCI	SOX2	ETV1
ARID2	CDK4	ERRFI1	GATA4	KDR	NF2	PRKDC	SOX9	NFKB IA
ASXL	CDK6	ESR1	GATA6	KEAP1	NFE2L2	PRSS8	SPEN	

1								
ATM	CDK8	EZH2	GID4 (C17orf3 9)	KEL	NFKBIA	PTCH1	SPOP	
ATR	CDKN1A	FAM46C	GLI1	KIT	NKX2-1	PTEN	SPTA1	
ATRX	CDKN1B	FANCA	GNA11	KLHL6	NOTCH1	PTPN11	SRC	
AURK A	CDKN2A	FANCC	GNA13	KMT2A (MLL)	NOTCH2	QKI	STAG2	
AURK B	CDKN2B	FANCD2	GNAQ	KMT2C (MLL3)	NOTCH3	RAC1	STAT3	
AXIN1	CDKN2C	FANCE	GNAS	KMT2D (MLL2)	NPM1	RAD50	STAT4	
AXL	CEBPA	FANCF	GPR124	KRAS	NRAS	RAD51	STK11	
BAP1	CHD2	FANCG	GRIN2A	LMO1	NSD1	RAF1	SUFU	
BARD 1	CHD4	FANCL	GRM3	LRP1B	NTRK1	RANBP2	SYK	
BCL2	CHEK1	FAS	GSK3B	LYN	NTRK2	RARA	TAF1	
BCL2 L1	CHEK2	FAT1	H3F3A	LZTR1	NTRK3	RB1	TBX3	
BCL2 L2	CIC	FBXW7	HGF	MAGI2	NUP93	RBM10	TERC	
BCOR	CREBBP	FGF10	HNF1A	MAP2K1	PAK3	RET	TERT (仅启动 子)	
BCOR L1	CRKL	FGF14	HRAS	MAP2K2	PALB2	RICTOR	TET2	

BLM	CRLF2	FGF19	HSD3B1	MAP2K4	PARK2	RNF43	TGFBR2	
BRAF	CSF1R	FGF23	HSP90A A1	MAP3K1	PAX5	ROS1	TNFAIP3	
BRCA 1	CTCF	FGF3	IDH1	MCL1	PBRM1	RPTOR	TNFRSF 14	
BRCA 2	CTNNA1	FGF4	IDH2	MDM2	PDCD1LG 2	RUNX1	TOP1	
BRD4	CTNNB1	FGF6	IGF1R	MDM4	PDGFRA	RUNX1T 1	TOP2A	
BRIP1	CUL3	FGFR1	IGF2	MED12	PDGFRB	SDHA	TP53	
BTG1	CYLD	FGFR2	IKBKE	MEF2B	PDK1	SDHB	TSC1	
BTK	DAXX	FGFR3	IKZF1	MEN1	PIK3C2B	SDHC	TSC2	

表9. 11种基因增殖基因标签

BIRC5	CDCA1 (NUF2)	MKI67	TYMS
CCNB1	CEP55	PTTG1	UBE2C
CDC20	KNTC2(NDC80)	RRM2	

表10. 如实施例4中所确定的AD表达亚型的322肿瘤小组3的前25种分化基因。

AD基因	KW p值
AURKA	1.40E-50
AURKB	1.06E-49
TOP2A	1.88E-46
RAD51	2.28E-46
CHEK1	3.40E-44
BLM	1.40E-43
TMPRSS2	6.34E-40

FAS	9.42E-39
ROS1	1.07E-37
EZH2	2.18E-37
BRCA1	1.16E-36
CD274	1.26E-35
CCNE1	4.95E-35
BRIP1	2.50E-34
ERBB4	2.16E-33
CSF1R	2.97E-33
PDCD1LG2	9.44E-33
FANCG	1.22E-32
BTK	3.48E-32
CHEK2	3.13E-30
CEBPA	4.87E-30
AXL	2.12E-29
FANCD2	3.93E-29
ETV1	1.66E-27
DNMT3A	5.53E-26

表11. 如实施例5中所确定的AD表达亚型的322肿瘤小组3的前25种分化基因。

AD基因	KW p值
AURKA	9.48E-57
AURKB	1.81E-56
TOP2A	1.74E-54
RAD51	6.87E-53
CHEK1	6.77E-49
BLM	2.08E-48
BRCA1	3.25E-44
CCNE1	7.10E-42
EZH2	2.19E-41
TMPRSS2	4.67E-41
BRIP1	4.52E-39
FANCG	1.34E-35
CHEK2	1.83E-35
FAS	3.16E-34
FANCD2	1.54E-33
ROS1	3.42E-32
CEBPA	6.55E-31
ERBB4	1.05E-30
FANCA	2.63E-29
MSH6	5.67E-29
BRCA2	4.75E-27
CD274	4.95E-27
TGFBR2	1.12E-26
POLE	2.82E-26
ETV1	2.32E-25

上述不同的实施方案可以组合以提供另外的实施方案。将在本说明书中引用的和/或在申请数据表中列举的所有美国专利申请公开案、美国专利申请、外国专利、外国专

利申请和非专利公开案通过引用以其整体结合在此。如果必要的话,可以修改实施方案的方面,以采用不同专利、申请和公开案的概念以提供另外的实施方案。

根据上文详细说明,可以对实施方案作出这些和其他改变。总体上,在以下权利要求书中,所使用的术语不应解读为将权利要求书限制为说明书和权利要求书中公开的具体实施方案,而应解读为包括所有可能的实施方案连同这些权利要求所享有的等效权利的全部范围。因此,权利要求书不受本公开文本的限制。

更具体地,本申请提供下列各项:

1. 一种用于确定从患者获得的肺组织样品的腺癌(AD)亚型的方法,该方法包括检测表1的至少一种分类器生物标记物的表达水平,其中对该分类器生物标记物的表达水平的检测特异性地鉴定末端呼吸单元(TRU)、近端增生(PP)或近端炎症(PI)AD亚型。

2. 项1的方法,其中该方法还包括将检测到的表1的该至少一种分类器生物标记物的表达水平与至少一个样品训练集中的表1的该至少一种分类器生物标记物的表达进行比较,其中该至少一个样品训练集包括来自参考AD TRU样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PP样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PI样品的表1的该至少一种分类器生物标记物的表达数据或其

组合;并且基于该比较步骤的结果将该样品分类为TRU、PP或PI亚型。

3. 项2的方法,其中该比较步骤包括应用统计算法,该统计算法包括确定从该样品获得的该表达数据与来自该至少一个训练集的表达数据之间的相关性;并且基于该统计算法的结果将该样品分类为TRU、PP或PI亚型。

4. 上述项中任一项的方法,其中在核酸水平上检测该分类器生物标记物的该表达水平。

5. 项4的方法,其中该核酸水平是RNA或cDNA。

6. 项4或5的方法,其中对表达水平的该检测包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。

7. 项6的方法,其中通过进行qRT-PCR检测该表达水平。

8. 项7的方法,其中对该表达水平的该检测包括使用对表1的至少一种分类器生物标记物有特异性的至少一对寡核苷酸引物。

9. 上述项中任一项的方法,其中该样品是从该患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。

10. 项9的方法,其中该体液是血液或其部分、尿液、唾液或痰液。

11. 上述项中任一项的方法,其中该至少一种分类器生物标记物包含多种分类器生物标记物。

12. 项11的方法,其中该多种分类器生物标记物包含表1的至少两种分类器生物标记物、至少8种分类器生物标记物、至少16种分类器生物标记物、至少24种分类器生物标记物、至少32种分类器生物标记物、至少40种分类器生物标记物、或至少48种分类器生物标记物。

13. 项1-10中任一项的方法,其中该至少一种分类器生物标记物包含表1的所有分

类器生物标记物。

14. 一种用于确定从患者获得的肺组织样品的腺癌(AD)亚型的方法,该方法包括检测编码在肺癌细胞中具有特定表达模式的分类器生物标记物的至少一种核酸分子的表达水平,其中该分类器生物标记物选自下组,该组由表1中列出的分类器基因组成,该方法包括:(a) 从来自患者的肺组织样品中分离核酸材料;(b) 将该核酸材料与基本上与该分类器生物标记物的核酸分子的部分互补的寡核苷酸混合;并且(c) 检测该分类器生物标记物的表达。

15. 项14的方法,其中该方法还包括将表1的该至少一种分类器生物标记物的检测到的表达水平与至少一个样品训练集中的表1的该至少一种分类器生物标记物的表达进行比较,其中该至少一个样品训练集包括来自参考AD TRU样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PP样品的表1的该至少一种分类器生物标记物的表达数据、来自参考AD PI样品的表1的该至少一种分类器生物标记物的表达数据或

其组合;并且基于该比较步骤的结果将该样品分类为TRU、PP或PI亚型。

16. 项15的方法,其中该比较步骤包括应用统计算法,该统计算法包括确定从该样品获得的该表达数据与来自该至少一个训练集的表达数据之间的相关性;并且基于该统计算法的结果将该样品分类为TRU、PP或PI亚型。

17. 项14-16中任一项的方法,其中对该表达水平的该检测包括进行qRT-PCR或任何基于杂交的基因测定。

18. 项17的方法,其中通过进行qRT-PCR检测该表达水平。

19. 项18的方法,其中对该表达水平的该检测包括使用对表1的至少一种分类器生物标记物有特异性的至少一对寡核苷酸引物。

20. 项14-19中任一项的方法,还包括基于检测到的该分类器生物标记物的表达水平预测对用于治疗肺腺癌(AD)亚型的疗法的反应。

21. 项20的方法,其中该疗法是化学疗法、血管生成抑制剂和/或免疫疗法。

22. 项21的方法,其中该肺AD亚型是TRU,并且该疗法是化学疗法或血管生成抑制剂。

23. 项21的方法,其中该肺AD亚型是PP,并且该疗法是化学疗法。

24. 项21的方法,其中该肺AD亚型是PI,并且该疗法是免疫疗法。

25. 项14-24中任一项的方法,其中该样品是从该患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。

26. 项25的方法,其中该体液是血液或其部分、尿液、唾液或痰液。

27. 项14-26中任一项的方法,其中编码分类器生物标记物的该至少一种核酸分子包含编码多种分类器生物标记物的多种核酸分子。

28. 项27的方法,其中该多种分类器生物标记物包含表1的至少两种分类器生物标记物、至少5种分类器生物标记物、至少10种分类器生物标记物、至少20种分类器生物标记物、或至少30种分类器生物标记物。

29. 项14-26中任一项的方法,其中该至少一种分类器生物标记物包含表1的所有分类器生物标记物。

30. 一种检测从患者获得的肺组织样品中的生物标记物的方法,该方法包括使用

扩增、杂交和/或测序测定来测量选自表1的多种生物标记物核酸的表达水平。

31. 项30的方法,其中该肺组织样品先前被诊断为是腺癌。

32. 项31的方法,其中该先前的诊断是通过组织学检查。

33. 项30-32中任一项的方法,其中该扩增、杂交和/或测序测定包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。

34. 项33的方法,其中通过进行qRT-PCR检测该表达水平。

35. 项34的方法,其中对该表达水平的该检测包括对于选自表1的该多种生物标记物核酸中的每一种使用至少一对寡核苷酸引物。

36. 项30-35中任一项的方法,其中该样品是从该患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。

37. 项36的方法,其中该体液是血液或其部分、尿液、唾液或痰液。

38. 项30-37中任一项的方法,其中该多种生物标记物核酸包含表1的至少两种生物标记物核酸、至少10种生物标记物核酸、至少20种生物标记物核酸、至少30种生物标记物核酸、至少40种生物标记物核酸、至少50种生物标记物核酸、至少60种生物标记物核酸、或至少70种生物标记物核酸,基本上由其组成或由其组成。

39. 项30-37中任一项的方法,其中该多种生物标记物核酸包含表1的所有分类器生物标记物核酸,基本上由其组成或由其组成。

40. 一种检测从患者获得的肺组织样品中的生物标记物的方法,该方法基本上由以下组成:使用扩增、杂交和/或测序测定来测量选自表1的多种生物标记物核酸的表达水平。

41. 项40的方法,其中该肺组织样品先前被诊断为是腺癌。

42. 项41的方法,其中该先前的诊断是通过组织学检查。

43. 项40-42中任一项的方法,其中该扩增、杂交和/或测序测定包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。

44. 项43的方法,其中通过进行qRT-PCR检测该表达水平。

45. 项44的方法,其中对该表达水平的该检测包括对于选自表1的该多种生物标记物核酸中的每一种使用至少一对寡核苷酸引物。

46. 项40-45中任一项的方法,其中该样品是从该患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。

47. 项46的方法,其中该体液是血液或其部分、尿液、唾液或痰液。

48. 项40-47中任一项的方法,其中该多种生物标记物核酸包含表1的至少两种生物标记物核酸、至少10种生物标记物核酸、至少20种生物标记物核酸、至少30种生物标记物核酸、至少40种生物标记物核酸、至少50种生物标记物核酸、至少60种生物标记物核酸、或至少70种生物标记物核酸,基本上由其组成或由其组成。

49. 项40-47中任一项的方法,其中该多种生物标记物核酸包含表1的所有分类器

生物标记物核酸,基本上由其组成或由其组成。

50. 一种检测从患者获得的肺组织样品中的生物标记物的方法,该方法由以下组成:使用扩增、杂交和/或测序测定来测量选自表1的多种生物标记物核酸的表达水平。

51. 项50的方法,其中该肺组织样品先前被诊断为是腺癌。

52. 项51的方法,其中该先前的诊断是通过组织学检查。

53. 项50-52中任一项的方法,其中该扩增、杂交和/或测序测定包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。

54. 项53的方法,其中通过进行qRT-PCR检测该表达水平。

55. 项54的方法,其中对该表达水平的该检测包括对于选自表1的该多种生物标记物核酸中的每一种使用至少一对寡核苷酸引物。

56. 项50-55中任一项的方法,其中该样品是从该患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。

57. 项56的方法,其中该体液是血液或其部分、尿液、唾液或痰液。

58. 项50-57中任一项的方法,其中该多种生物标记物核酸包含表1的至少两种生物标记物核酸、至少10种生物标记物核酸、至少20种生物标记物核酸、至少30种生物标记物核酸、至少40种生物标记物核酸、至少50种生物标记物核酸、至少60种生物标记物核酸、或至少70种生物标记物核酸,基本上由其组成或由其组成。

59. 项50-57中任一项的方法,其中该多种生物标记物核酸包含表1的所有分类器生物标记物核酸,基本上由其组成或由其组成。

60. 一种确定腺癌患者是否可能对免疫疗法有反应的方法,该方法包括,

确定来自该患者的肺组织样品的腺癌亚型,其中该腺癌亚型选自下组,该组由以下组成:鳞状型(近端炎症型)、支气管型(末端呼吸单元型)

和巨型(近端增生型);并且

基于该亚型,评估该患者是否可能对免疫疗法有反应。

61. 一种用于选择腺癌患者进行免疫疗法的方法,该方法包括确定来自该患者的肺组织样品的腺癌亚型;并且基于该亚型选择患者进行免疫疗法。

62. 项60或61的方法,其中,该免疫疗法包括检查点抑制剂疗法。

63. 项62的方法,其中该检查点抑制剂靶向PD-1或PD-L1。

64. 项62的方法,其中该检查点抑制剂靶向CTLA-4。

65. 项63的方法,其中该检查点抑制剂是派姆单抗、纳武单抗或其抗原片段结合片段。

66. 项64的方法,其中该检查点抑制剂是伊匹单抗或其抗原结合片段。

67. 项60-66中任一项的方法,其中最初该患者经由样品的组织学分析被确定患有腺癌。

68. 项60-67中任一项的方法,其中该患者的腺癌分子亚型选自鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型),并且

经由对从该患者获得的样品的组织学分析来确定。

69. 项67-68中任一项的方法,其中该样品是从该患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体或体液。

70. 项69的方法,其中该体液是血液或其部分、尿液、唾液或痰液。

71. 项60-70中任一项的方法,其中对该腺癌亚型的该确定包括确定多种分类器生物标记物的表达水平。

72. 项71的方法,其中对该多种分类器生物标记物的表达水平的该确定是在核酸水平上通过进行RNA测序、逆转录酶聚合酶链反应(RT-PCR)或基于杂交的分析来进行。

73. 项71或72的方法,其中,用于确定该腺癌亚型的该多种分类器生物标记物选自可公开获得的肺腺癌数据集。

74. 项73的方法,其中该可公开获得的肺腺癌数据集是TCGA肺AD RNAseq数据集。

75. 项72的方法,其中用于确定该腺癌亚型的该多种分类器生物标记物选自表1。

76. 项75的方法,其中该RT-PCR是定量实时逆转录酶聚合酶链反应(qRT-PCR)。

77. 项76的方法,其中使用对表1的该多种分类器生物标记物有特异性的引物进行该RT-PCR。

78. 项71-77中任一项的方法,还包括:将检测到的表1的该多种分类器生物标记物的表达水平与至少一个样品训练集中的表1的该多种分类器生物标记物的表达进行比较,其中该至少一个样品训练集包括来自参考腺癌TRU样品的表1的该多种分类器生物标记物的表达数据、来自参考腺癌PP样品的表1的该多种分类器生物标记物的表达数据、来自参考腺癌PI样品的表1的该多种分类器生物标记物的表达数据、或其组合;并且基于该比较步骤的结果将该第一样品分类为TRU、PP或PI。

79. 项78的方法,其中该比较步骤包括应用统计算法,该统计算法包括确定从该样品获得的该表达数据与来自该至少一个训练集的表达数据之间的相关性;并且基于该统计算法的结果将该样品分类为TRU、PP或PI亚型。

80. 项71-79中任一项的方法,其中该多种分类器生物标记物包含表1中列出的分类器生物标记物中的每一种。

81. 一种治疗受试者中的肺癌的方法,该方法包括:

测量从该受试者获得的肺癌样品中至少一种生物标记物核酸的表达水平,其中该至少一种生物标记物核酸选自表1中列出的一组生物标记物,其中该至少一种生物标记物的存在、不存在和/或水平指示肺癌的亚型;并且

基于肺癌的该亚型给予免疫治疗剂。

82. 项81的方法,其中该肺癌样品是腺癌肺癌样品。

83. 项82的方法,其中选自该组生物标记物的该至少一种生物标记物核酸包含表1的至少两种生物标记物核酸、至少8种生物标记物核酸、至少16种生物标记物核酸、至少32种生物标记物核酸、或所有48种生物标记物核酸,基本上由其组成或由其组成。

84. 项81-83中任一项的方法,其中该肺组织样品先前被诊断为是腺癌。

85. 项84的方法,其中该先前的诊断是通过组织学检查。

86. 项81-85中任一项的方法,还包括:测量来自另外一组生物标记物的至少一种生物标记物的表达。

87. 项86的方法,其中该另外一组生物标记物包括先天性免疫细胞(IIC)、适应性

免疫细胞(AIC)的基因表达标签,一种或多种个体免疫生物标记物,一种或多种干扰素(IFN)基因,一种或多种主要组织相容性复合物II类(MHCII)基因或其组合。

88. 项87的方法,其中该另外一组生物标记物包含选自表4A、表4B、表5、表6、表7或其组合的基因。

89. 项87的方法,其中该AIC的基因表达标签选自表4A。

90. 项87的方法,其中该IIC的基因表达标签选自表4B。

91. 项87的方法,其中该一种或多种个体免疫生物标记物选自表5。

92. 项87的方法,其中该一种或多种IFN基因选自表6。

93. 项87的方法,其中该一种或多种MHCII基因选自表7。

94. 项81-93中任一项的方法,其中对该表达水平的该测量是使用扩增、杂交和/或测序测定来进行。

95. 项94的方法,其中该扩增、杂交和/或测序测定包括进行定量实时逆转录酶聚合酶链反应(qRT-PCR)、RNAseq、微阵列、基因芯片、nCounter基因表达测定、基因表达系列分析(SAGE)、基因表达快速分析(RAGE)、核酸酶保护测定、Northern印迹或任何其他等同的基因表达检测技术。

96. 项95的方法,其中通过进行qRT-PCR检测该表达水平。

97. 项81-96中任一项的方法,其中该样品是从该患者获得的福尔马林固定的石蜡包埋的(FFPE)肺组织样品、新鲜或冷冻的组织样品、外泌体、洗涤液体、细胞沉淀物或体液。

98. 项97的方法,其中该体液是血液或其部分、尿液、唾液或痰液。

99. 项81-98中任一项的方法,其中该受试者的腺癌亚型选自鳞状型(近端炎症型)、支气管型(末端呼吸单元型)或巨型(近端增生型)。

100. 项99的方法,其中该肺癌亚型是近端炎症型的,并且其中该免疫治疗剂包含检查点抑制剂。

101. 项100的方法,其中该检查点抑制剂靶向PD-1或PD-L1。

102. 项100的方法,其中该检查点抑制剂靶向CTLA-4。

103. 项101的方法,其中该检查点抑制剂是派姆单抗、纳武单抗或其抗原片段结合片段。

104. 项102的方法,其中该检查点抑制剂是伊匹单抗或其抗原结合片段。

105. 项81的方法,其中该至少一种生物标记物核酸是多种生物标记物核酸,其中该多种生物标记物核酸包含表1中所列出的至少一种生物标记物核酸与来自可公开获得的肺腺癌数据集的一种或多种生物标记物核酸的组合,其中该多种生物标记物核酸的存在、不存在和/或水平指示肺癌的亚型。

106. 项81的方法,其中该至少一种生物标记物核酸是多种生物标记物核酸,其中该多种生物标记物核酸包含表1中所列出的所有生物标记物核酸与来自可公开获得的肺腺癌数据集的一种或多种生物标记物核酸的组合,

其中该多种生物标记物核酸的存在、不存在和/或水平指示肺癌的亚型。107. 项105或106的方法,其中,该可公开获得的肺腺癌数据集是TCGA肺AD RNAseq数据集。

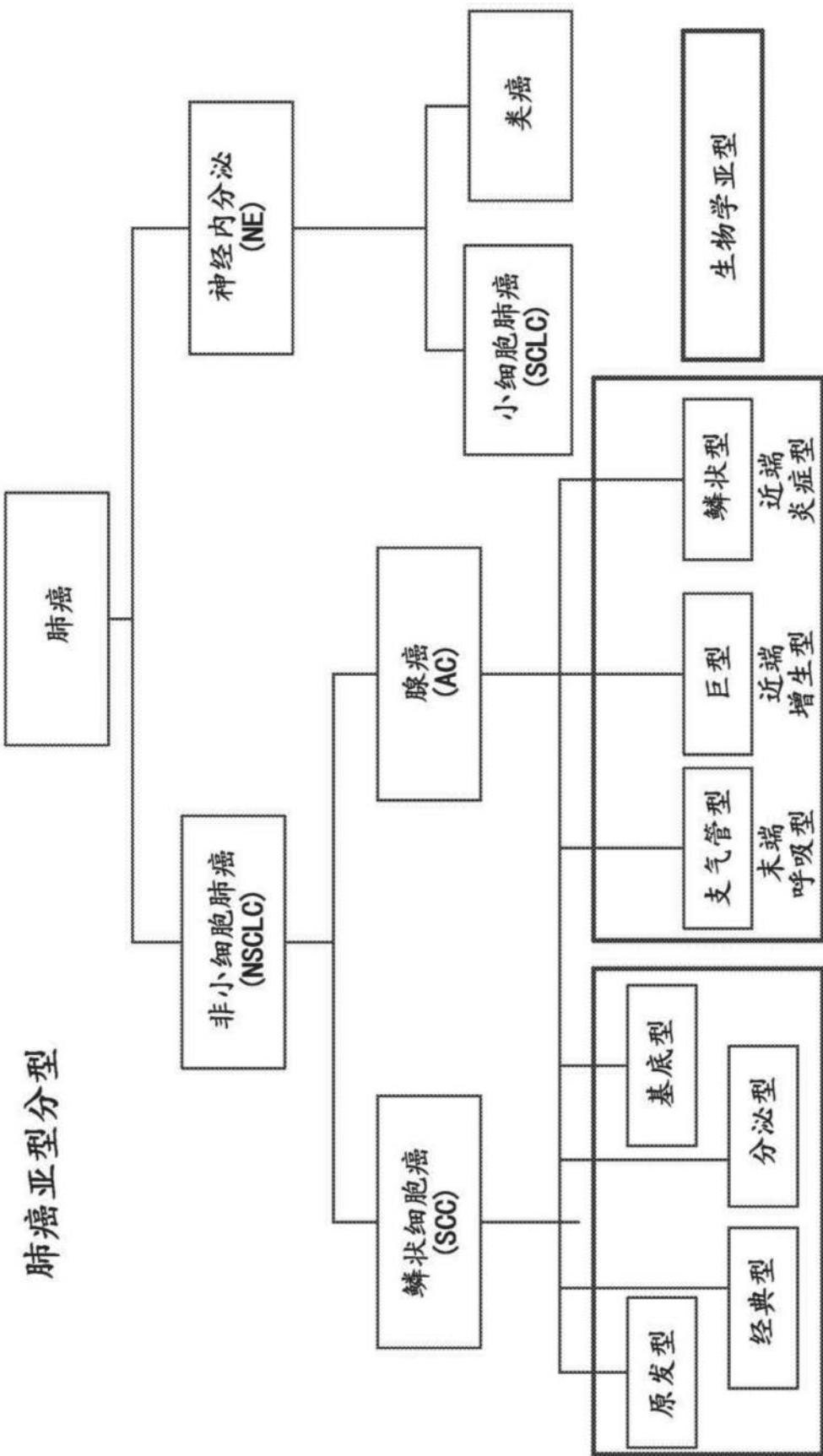


图1

特征	TCGA ²	Shedden ⁶	Tomida ⁷	UNC ¹	GeneCentric
样品总数	515	442	117	116	88
组织保存	新鲜冷冻	新鲜冷冻	新鲜冷冻	新鲜冷冻	FFPE
亚型					
支气管型	188	159	45	47	26
巨型	154	161	40	40	29
鳞状型	173	122	32	29	33
性别					
女性/男性/NA	274/237/4	219/223/0	57/60/0	63/53/0	
诊断年龄					
中值/ (范围)	66/(38-88)	65/(33-87)	61/(32-84)	65/(41-90)	
年龄未获得	35	0	0	0	
阶段					
I	276	276	79	73	42
II	123	95	13	21	7
III	84	68	25	19	15
IV	27	0	0	2	24
年龄未获得	5	3	0	1	0
是否曾吸烟					
是	346	300	61	82	
否	2	49	56	11	
吸烟状况未获得	167	93	0	23	

图2

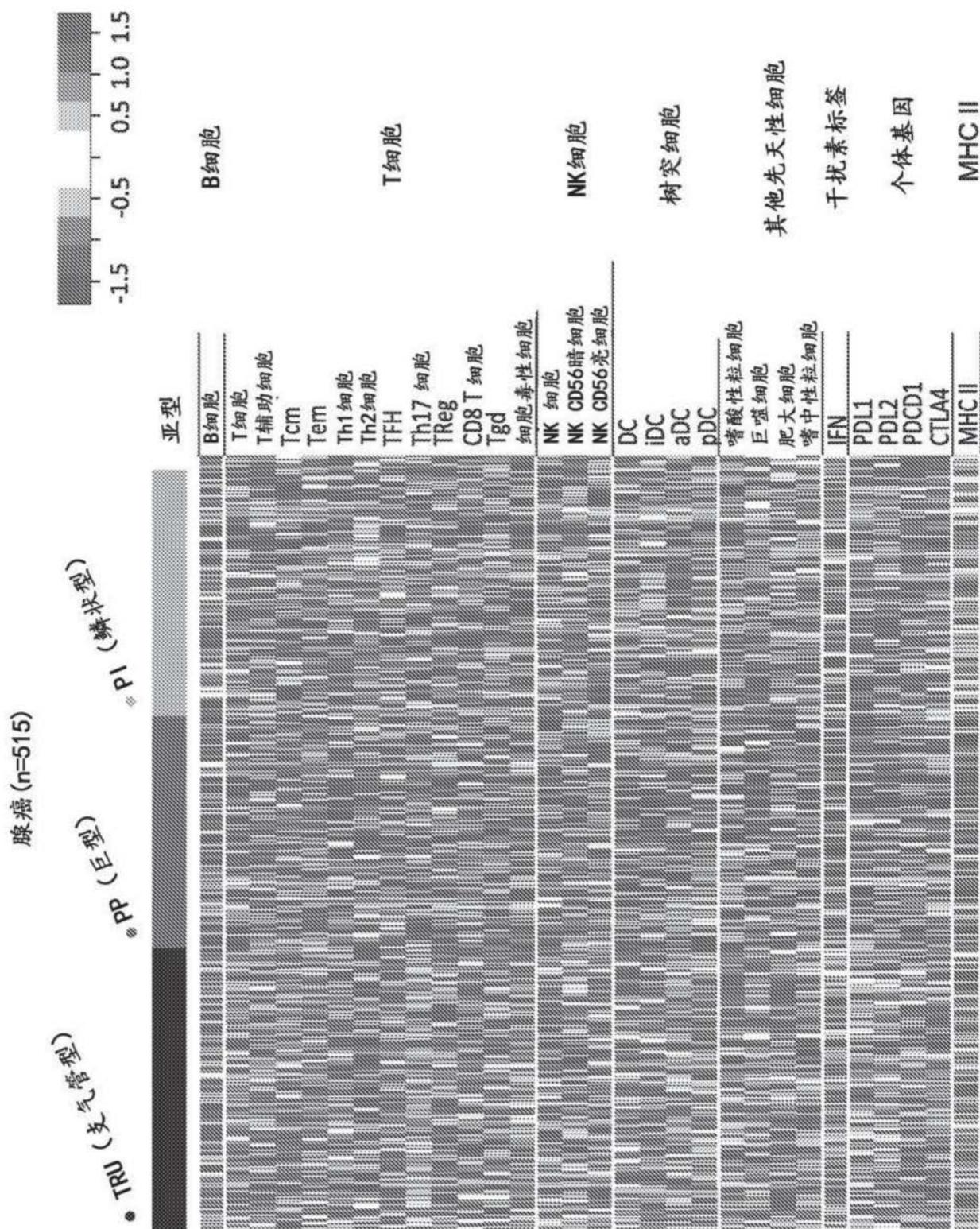


图3

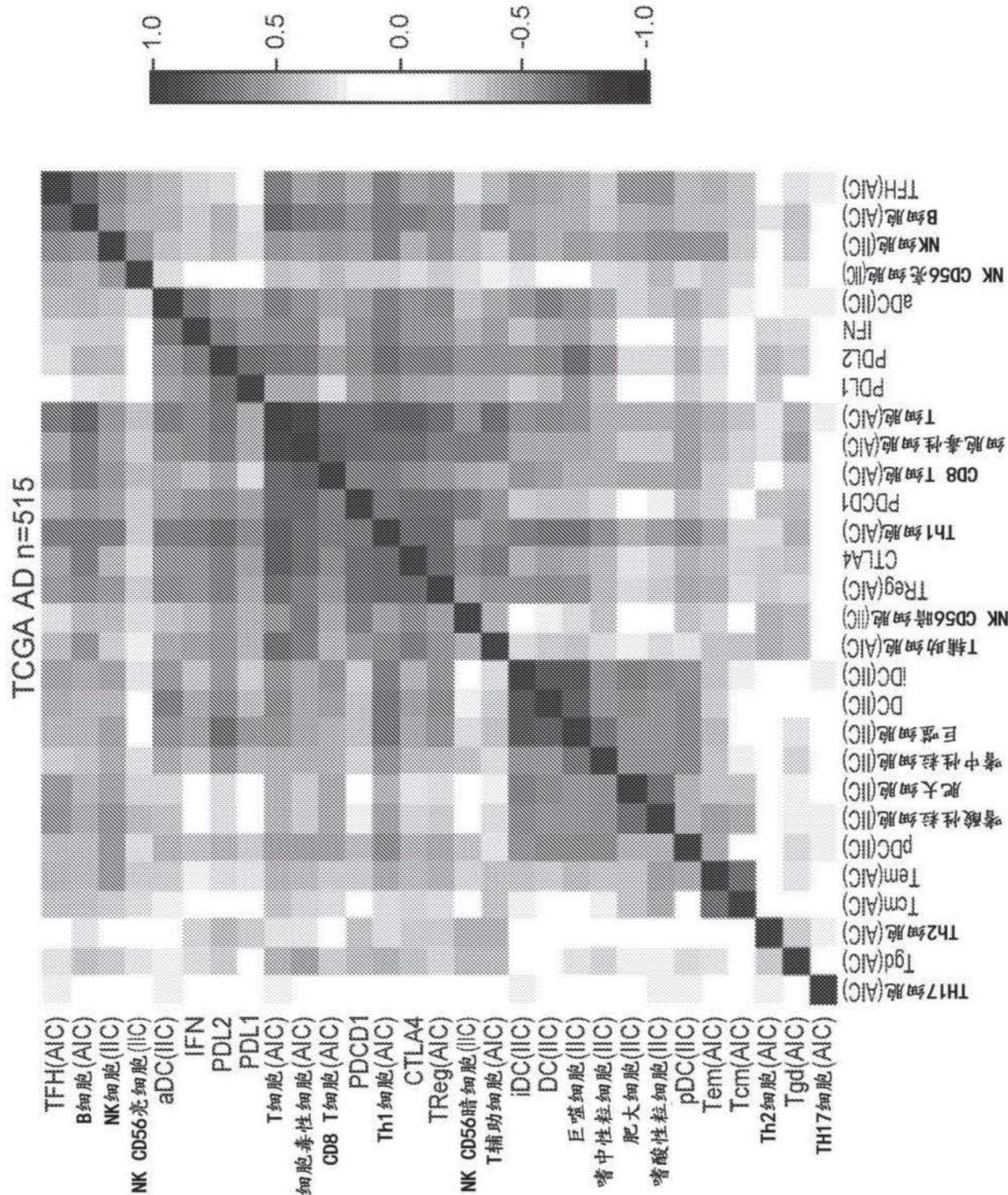


图4

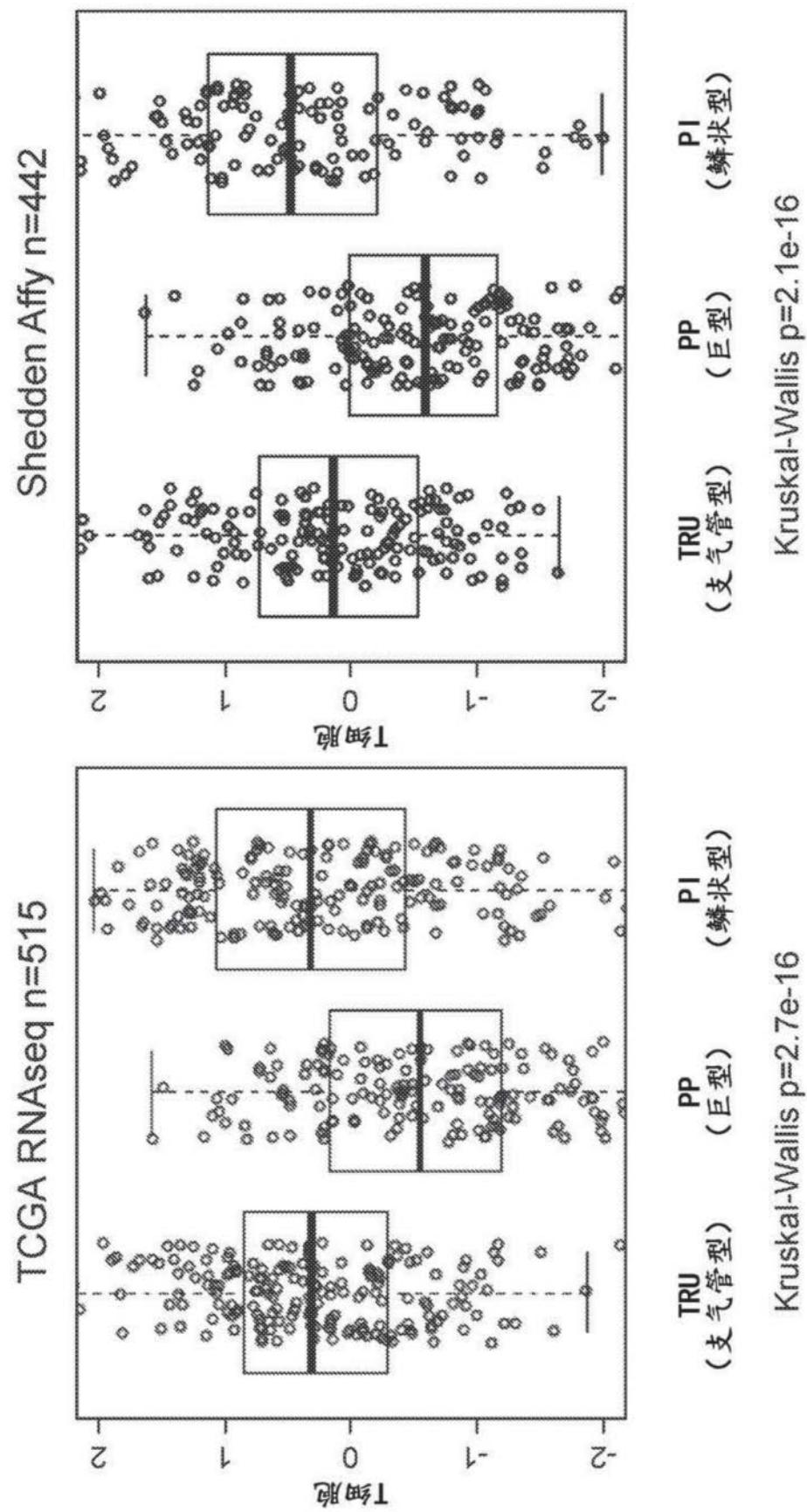


图5

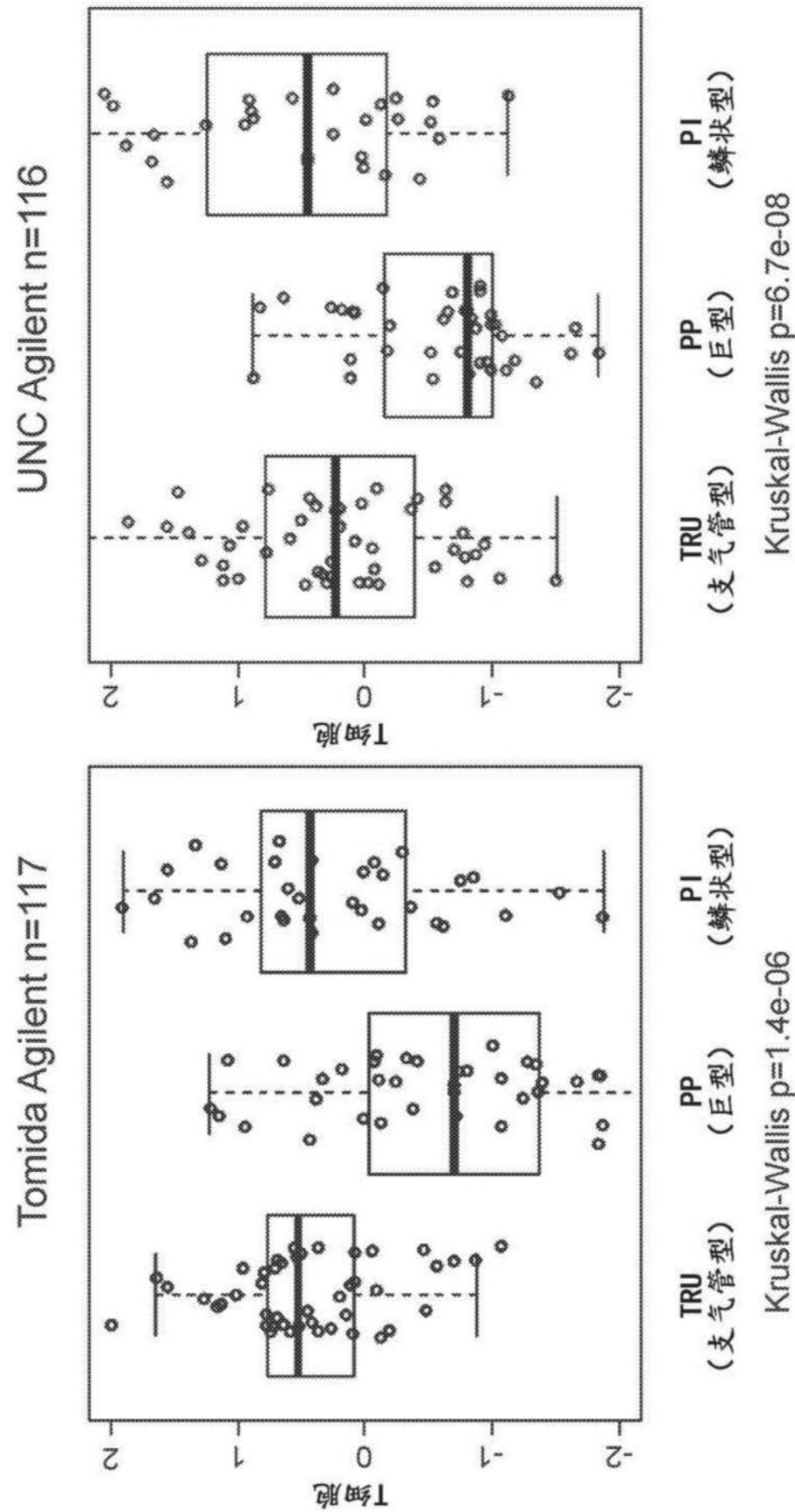


图5(续)

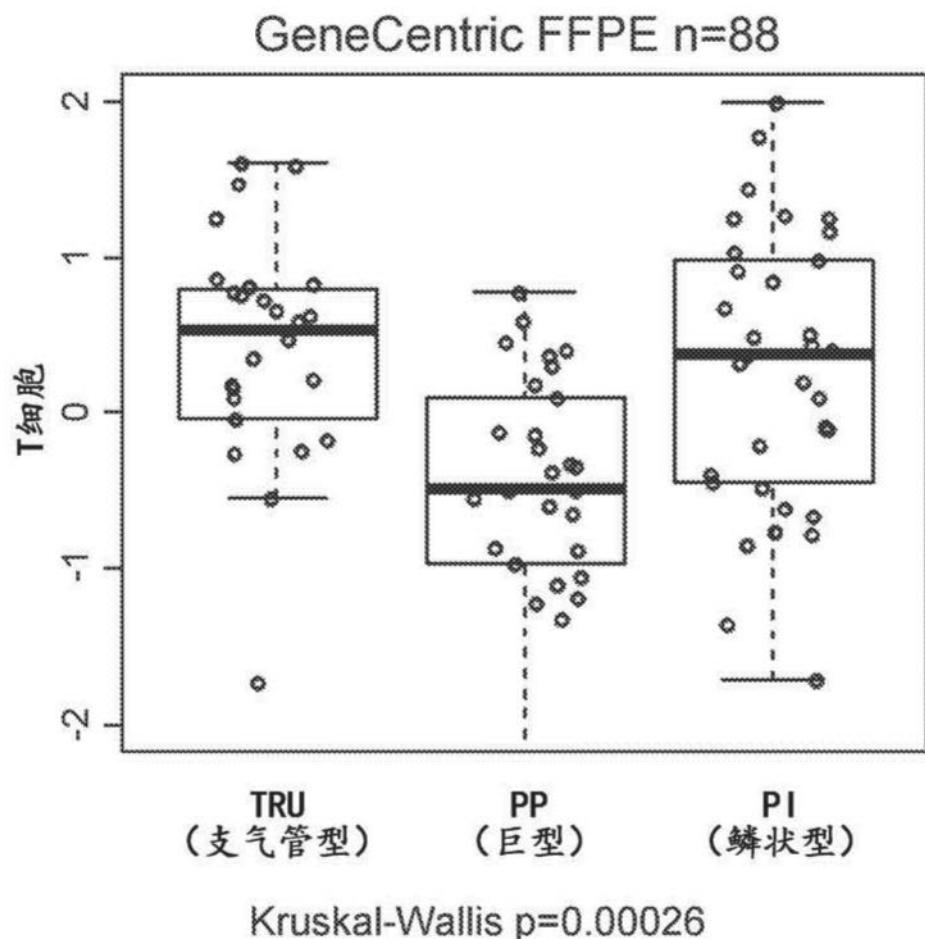


图5(续)

TCGA AD n=515

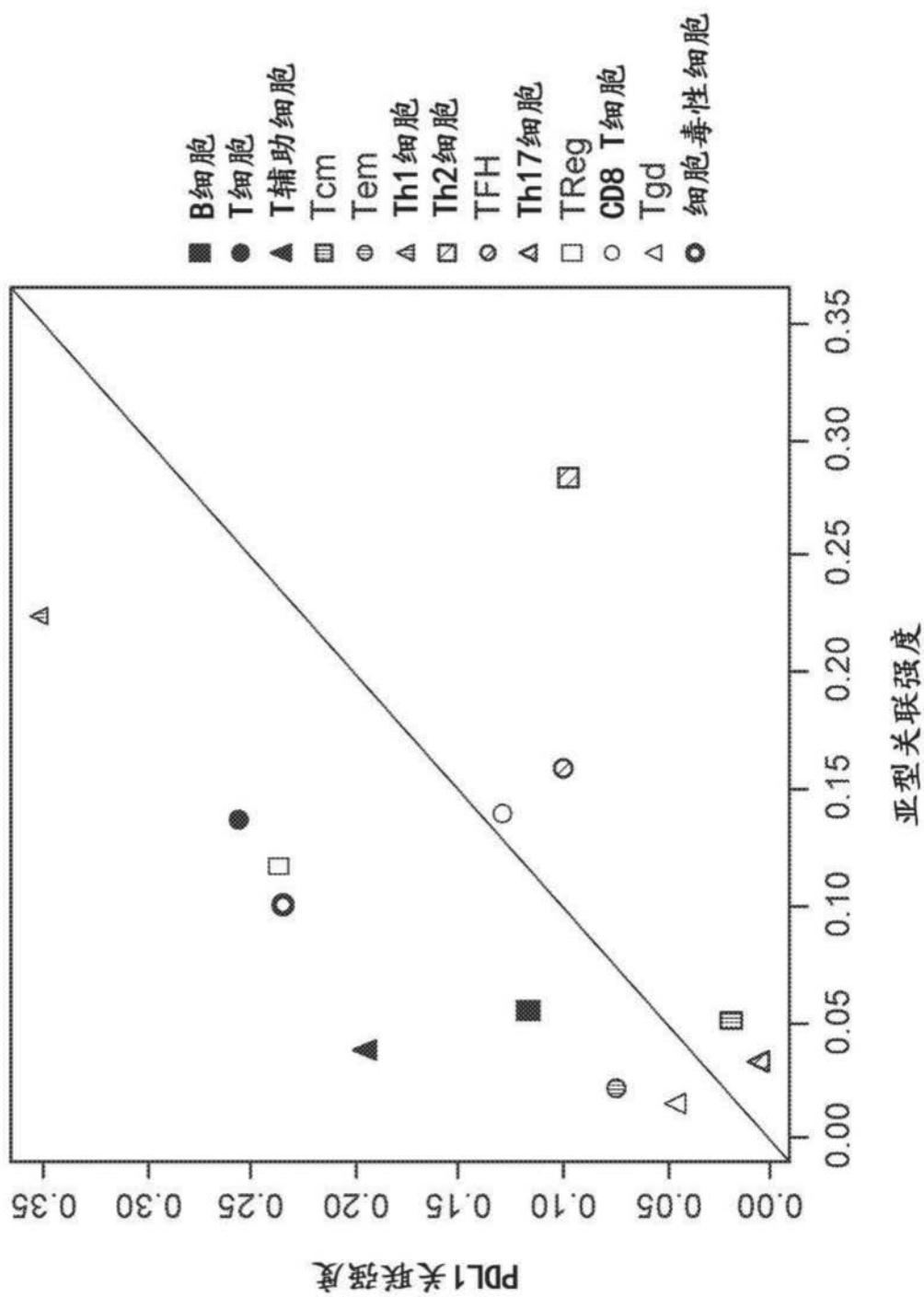


图6

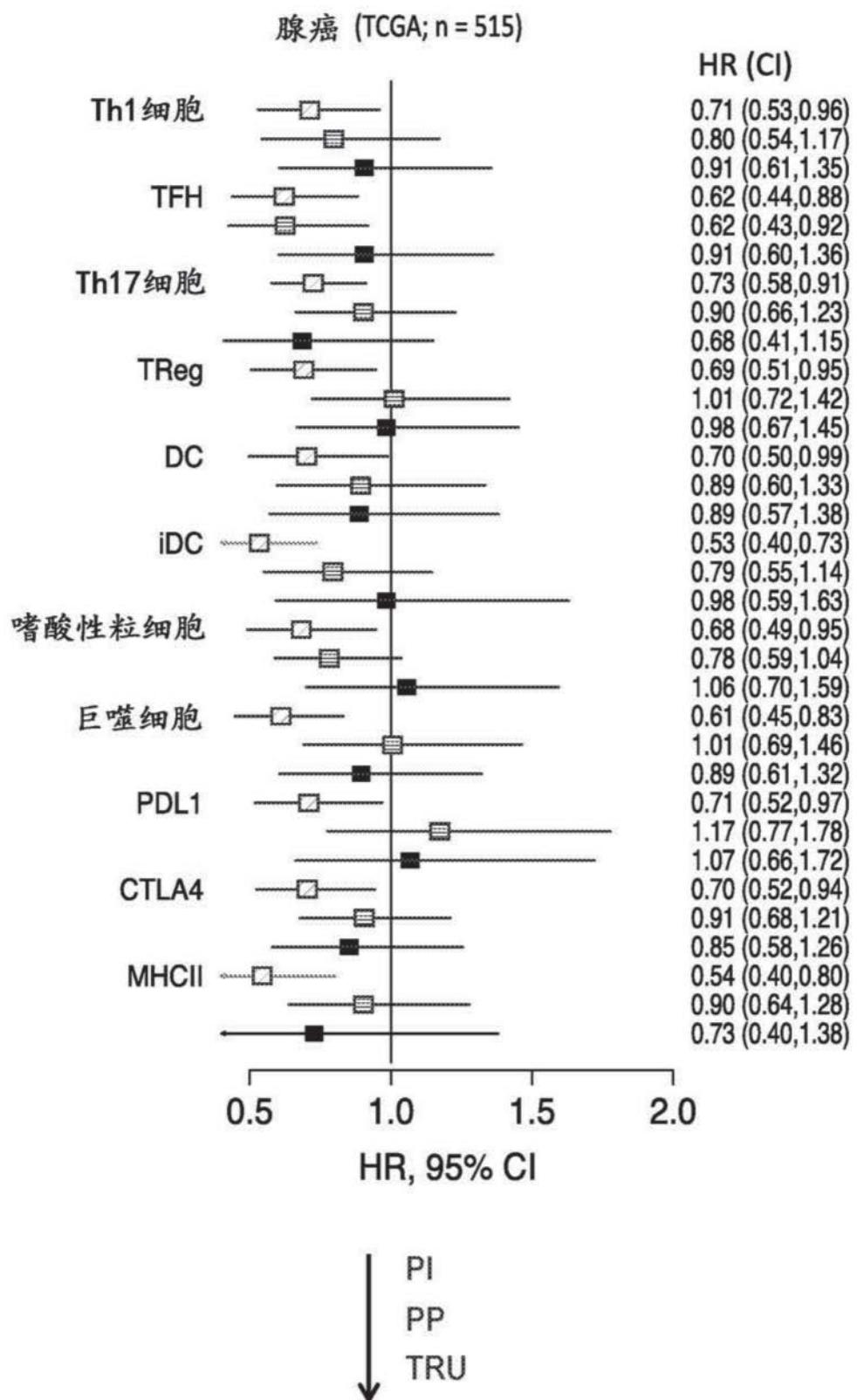


图7A

腺癌 (TCGA、Shedden、和Tomida)

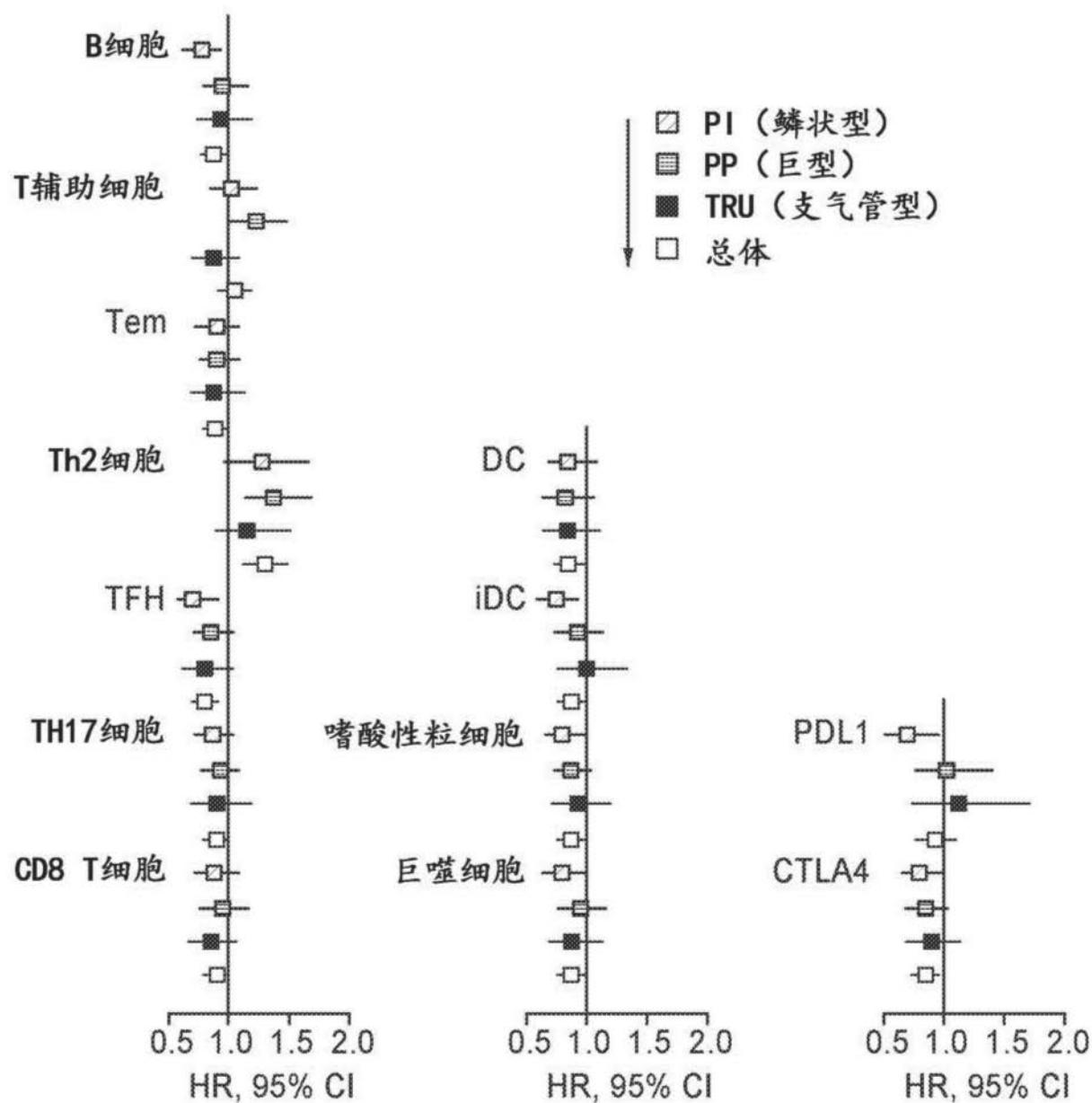


图7B

TCGA 5折交叉验证, 50-50 高-低方法

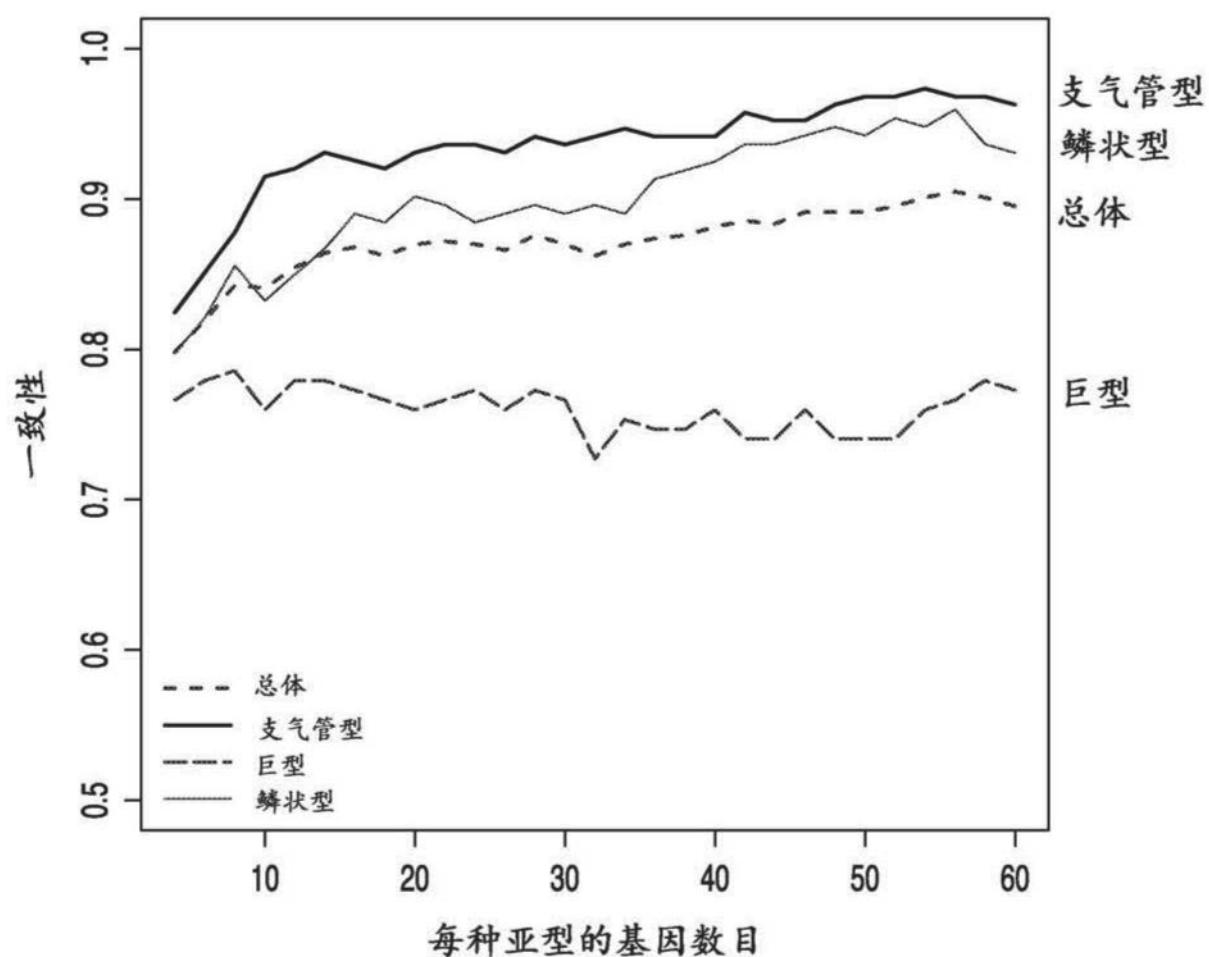


图8

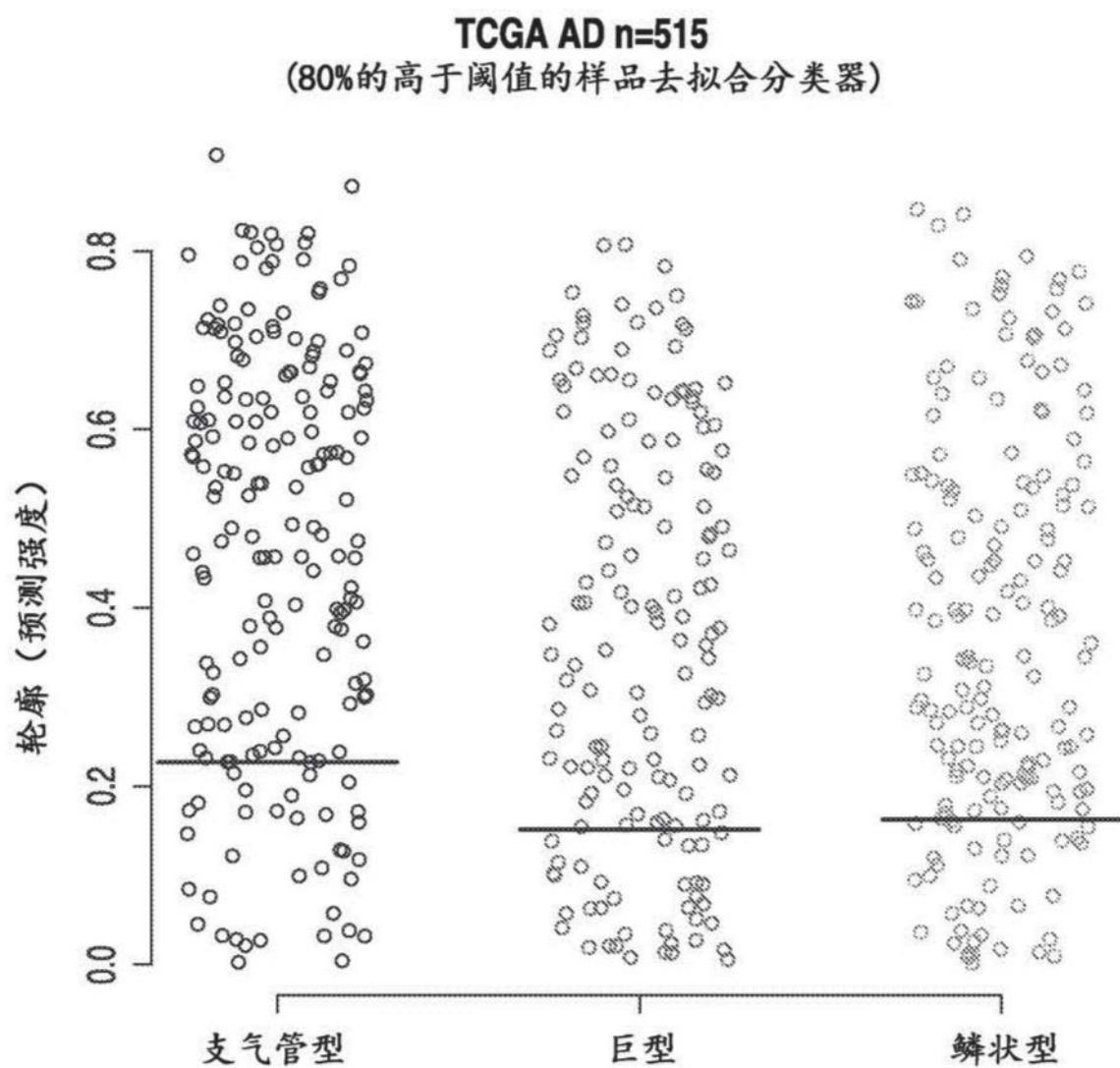


图9

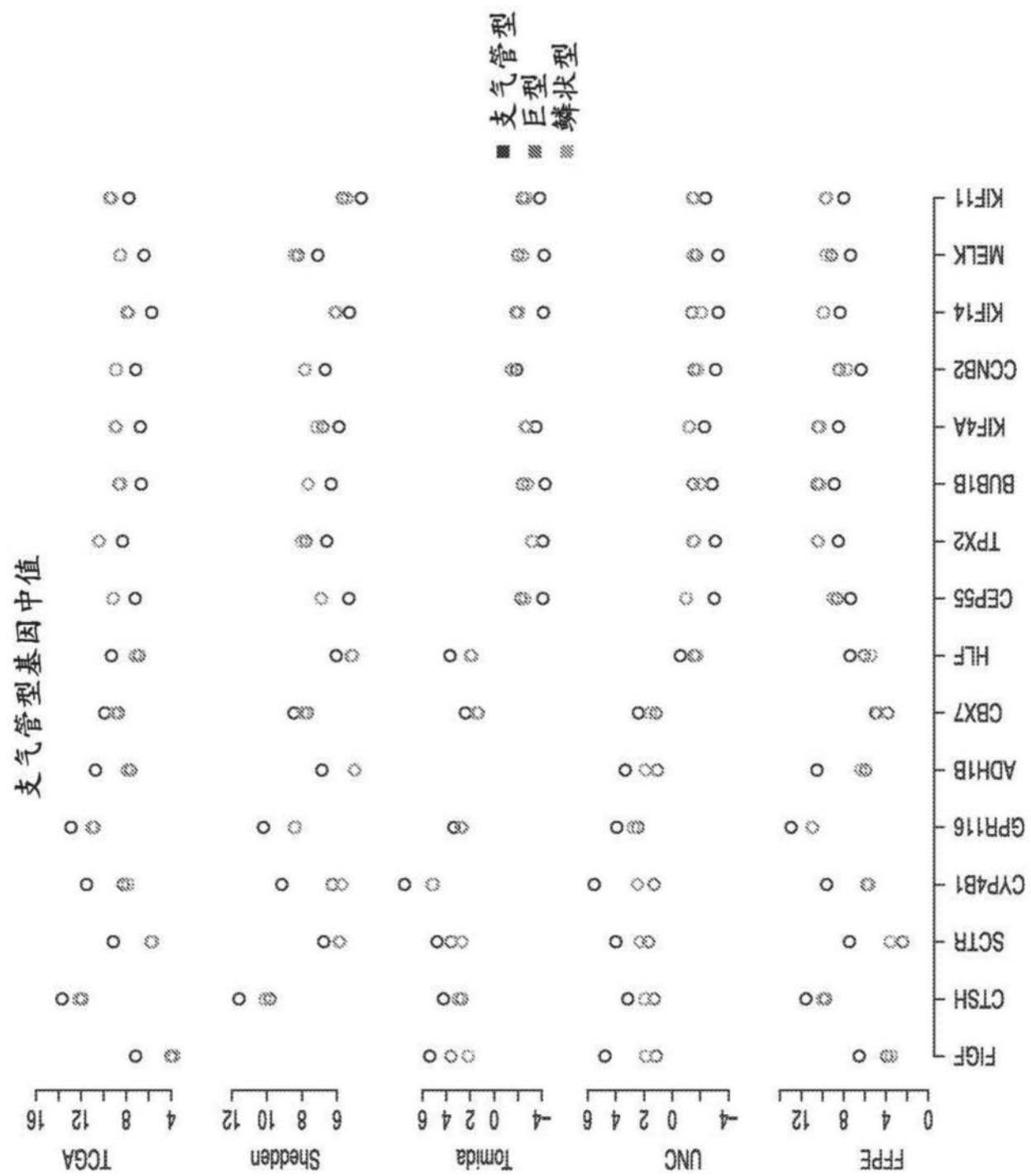


图10

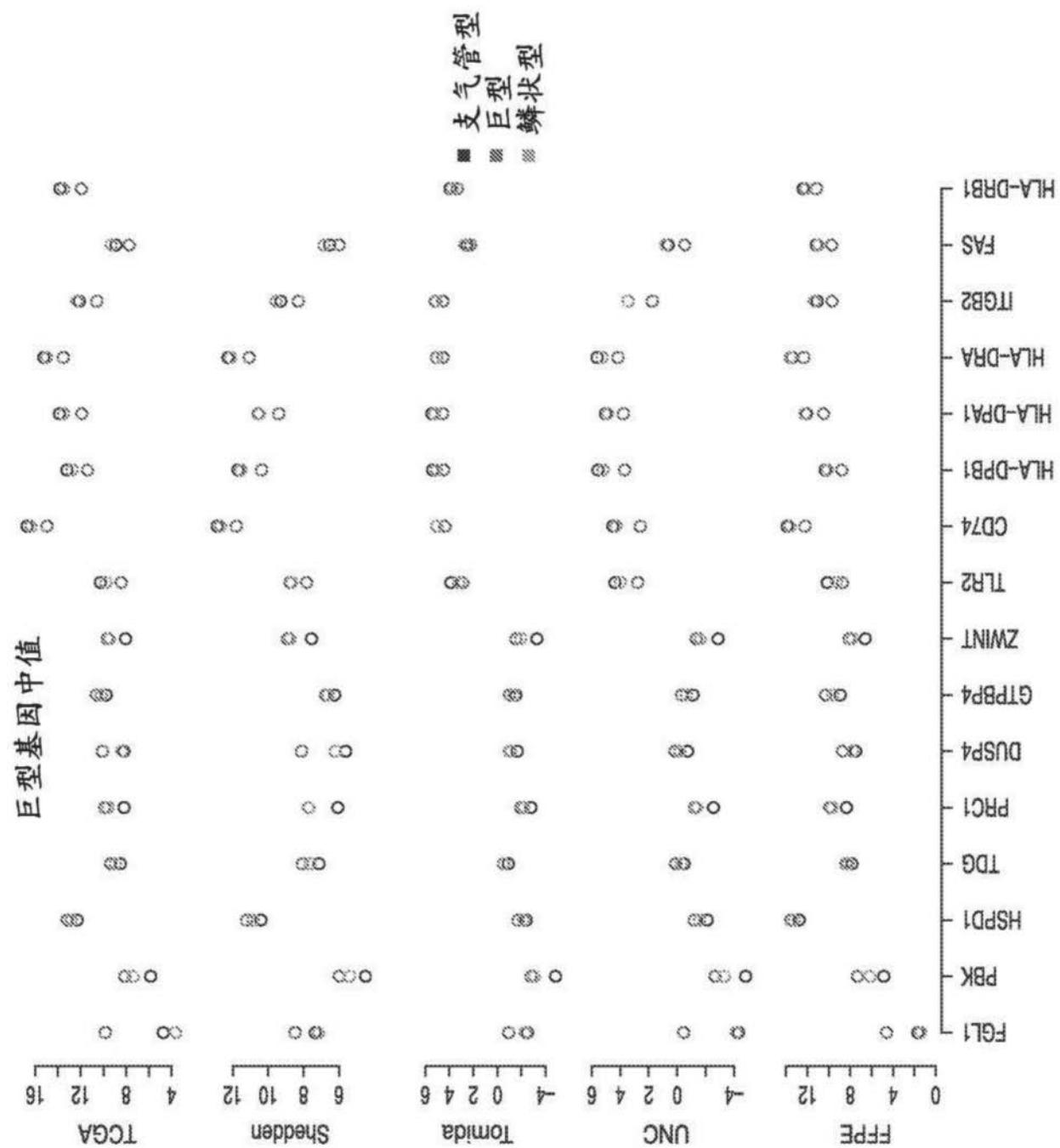


图11

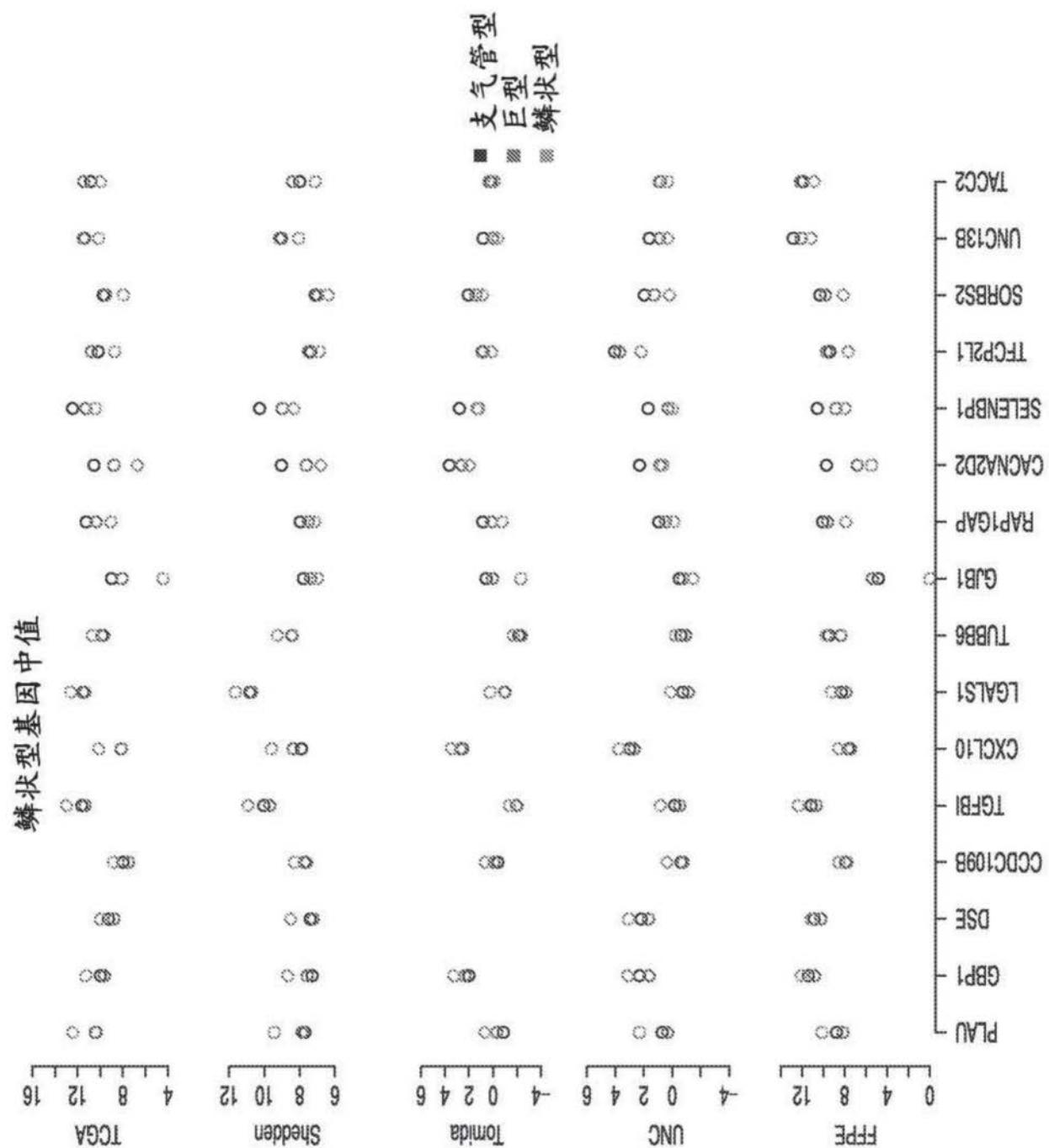


图12

金标准（行）与clanc48（列）

		支气管型	巨型	鳞状型
GeneCentric FFPE	支气管型	23	0	3
	巨型	1	23	5
	鳞状型	3	2	28
shedden	支气管型	151	5	3
	巨型	16	126	19
	鳞状型	7	6	109
tcga	支气管型	175	3	10
	巨型	12	120	22
	鳞状型	9	11	153
tomida	支气管型	43	2	0
	巨型	4	28	8
	鳞状型	2	8	22
unc	支气管型	44	3	0
	巨型	0	38	2
	鳞状型	1	3	25

一致性

	GeneCentric				
	FFPE	shedden	tcga	tomida	unc
一致性	0.84	0.87	0.87	0.79	0.92

图13

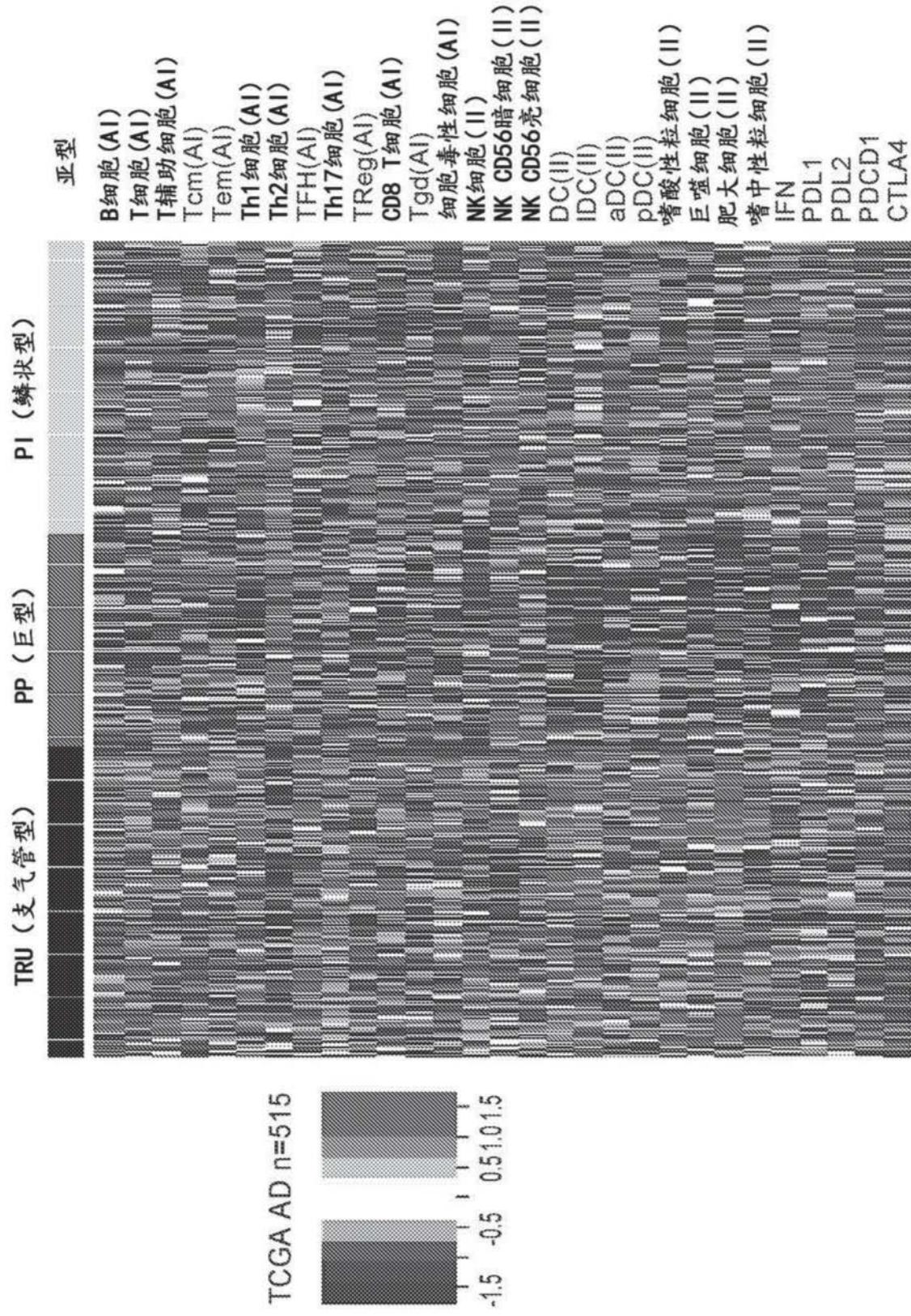


图14

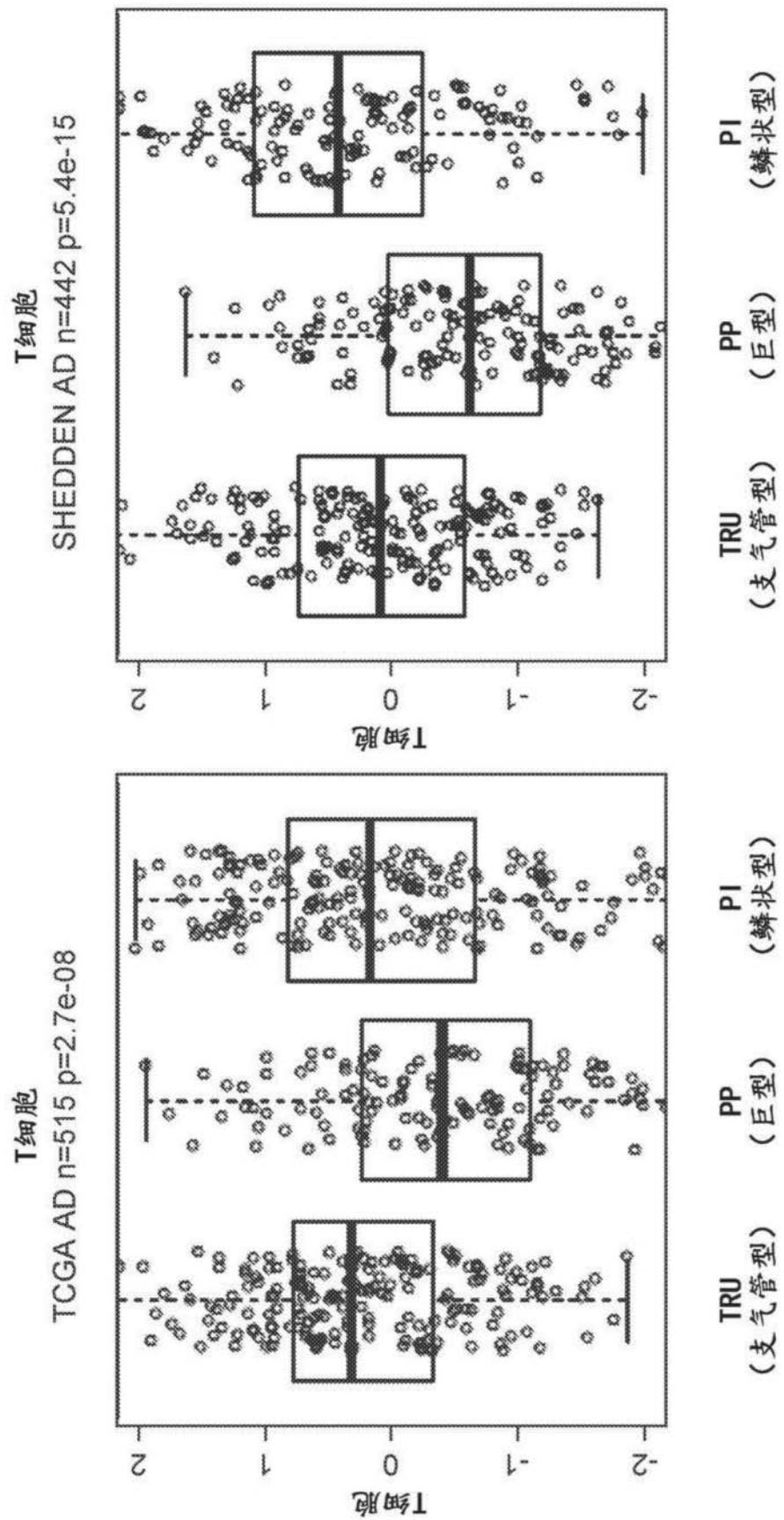


图15

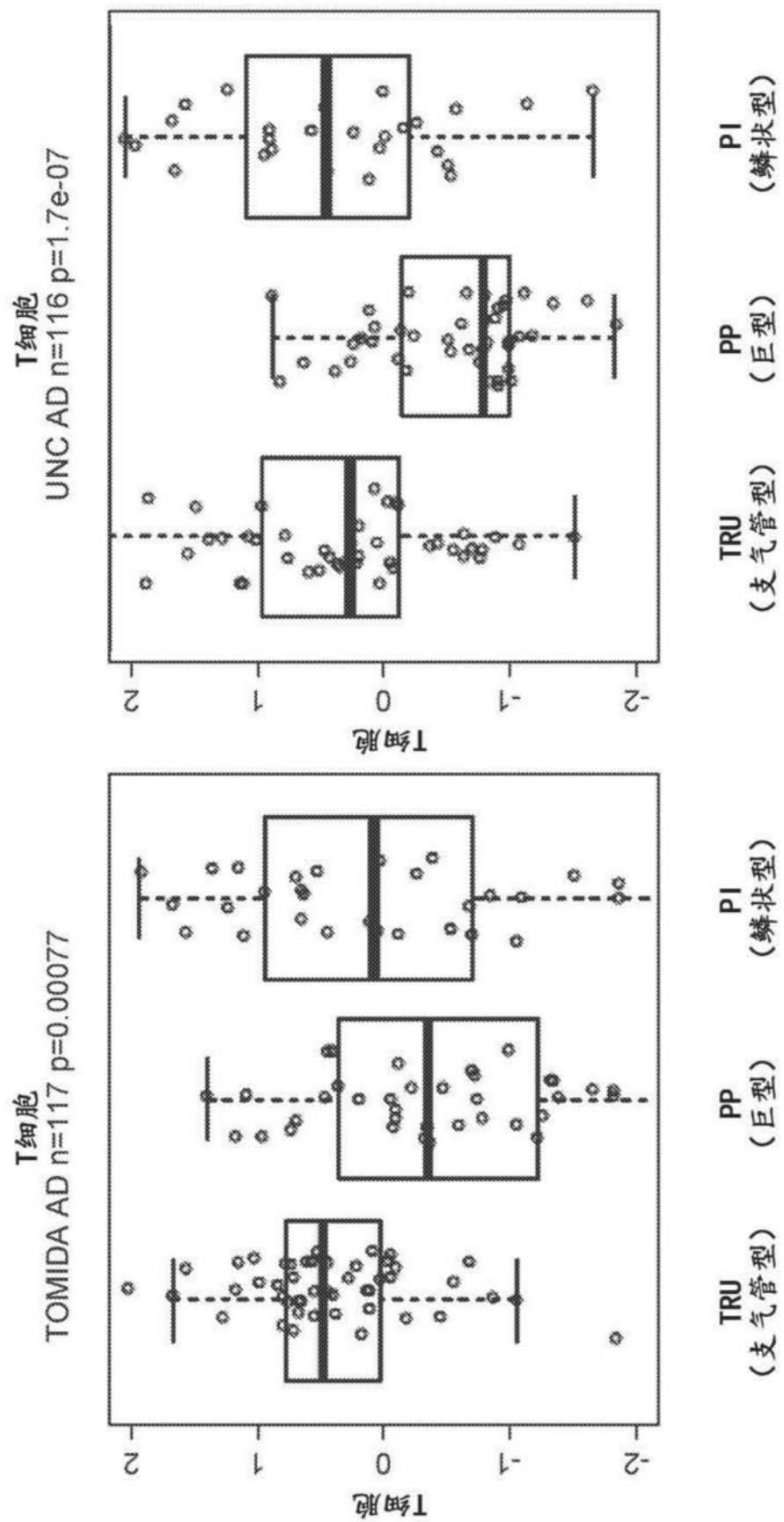


图15(续)

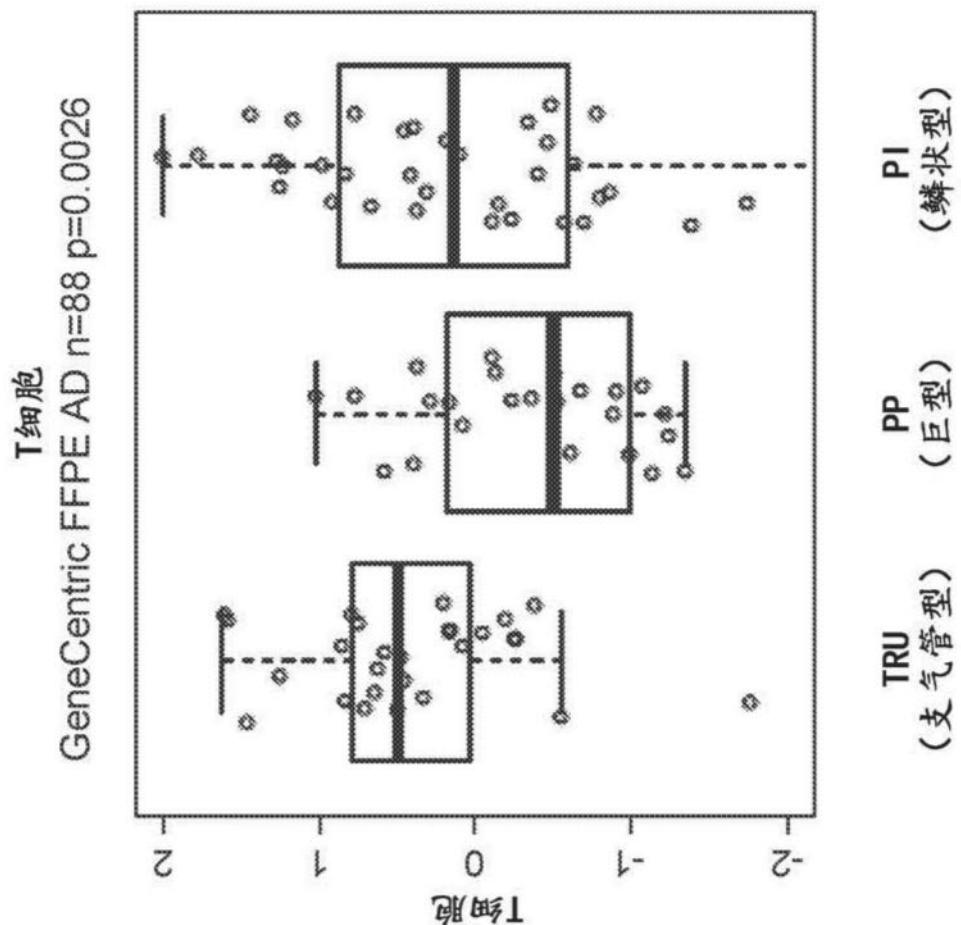


图15(续)

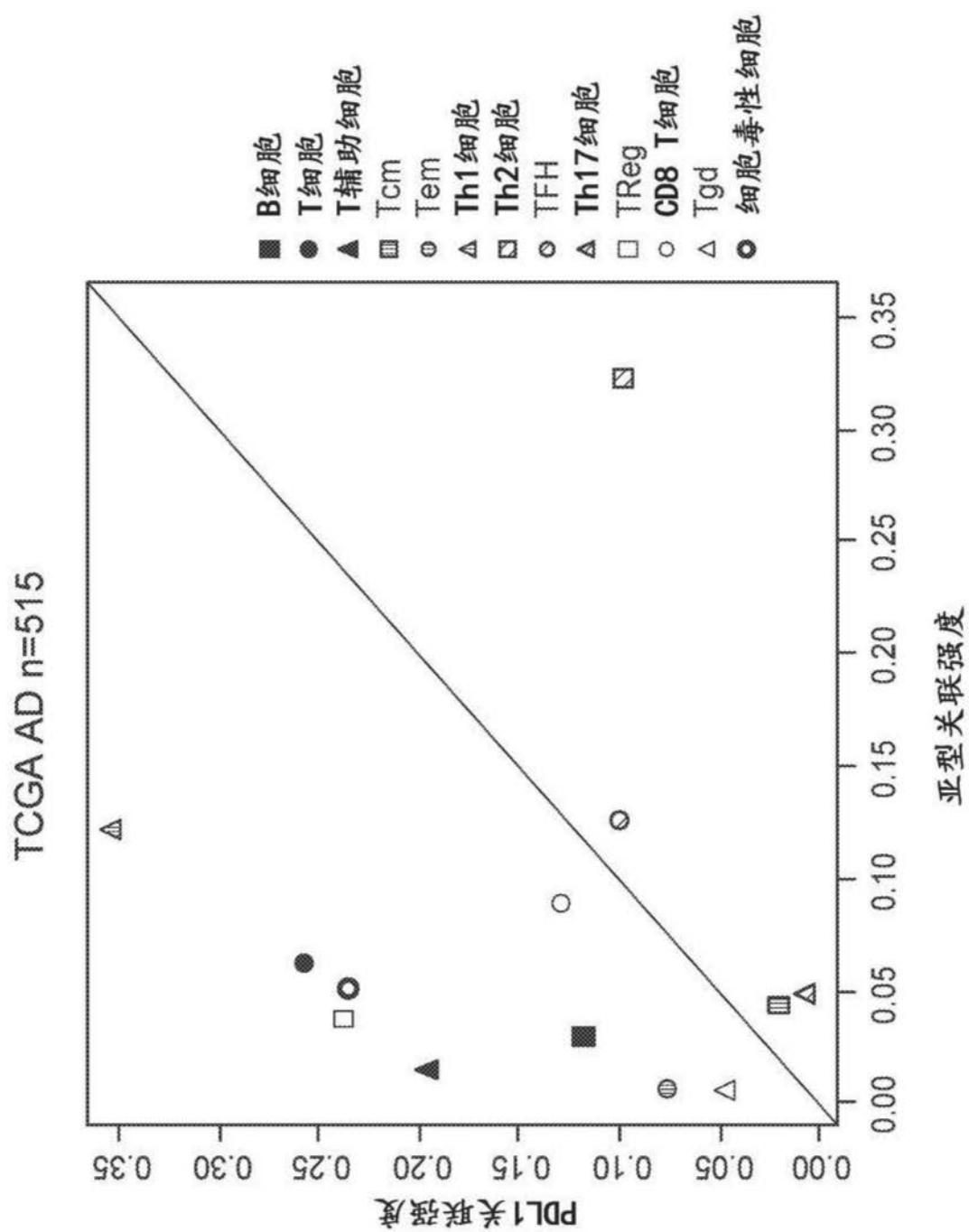


图16

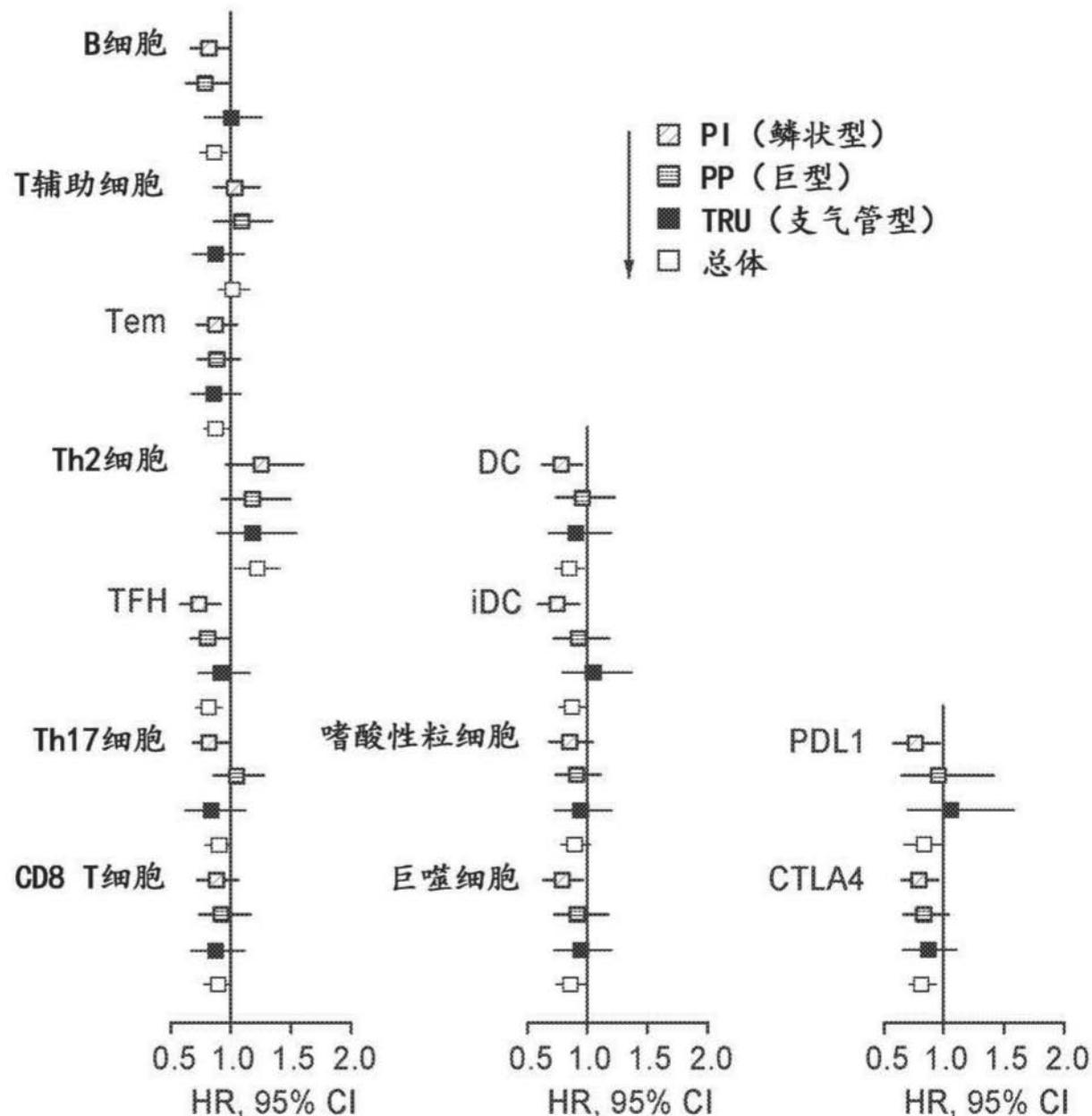


图17

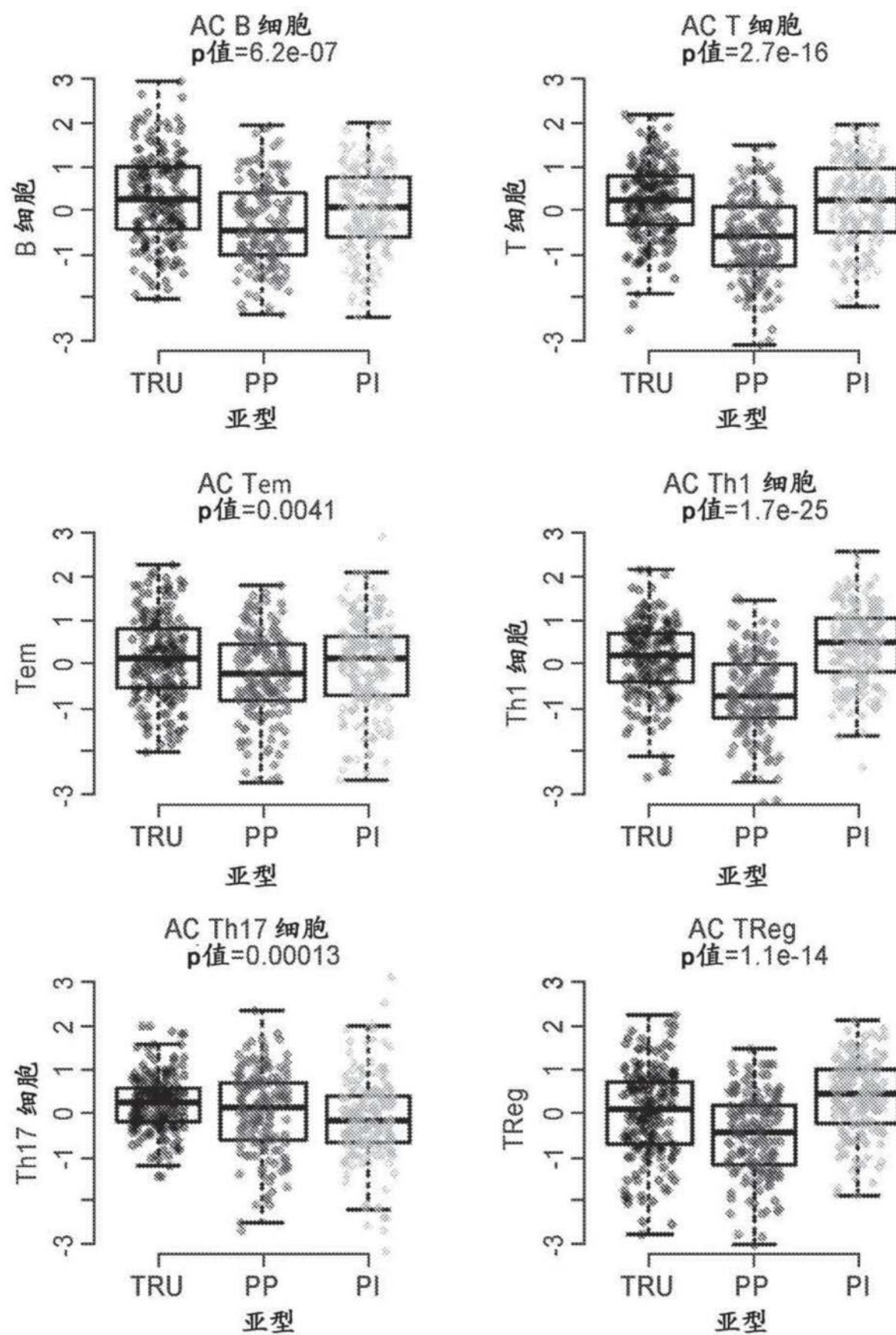


图18

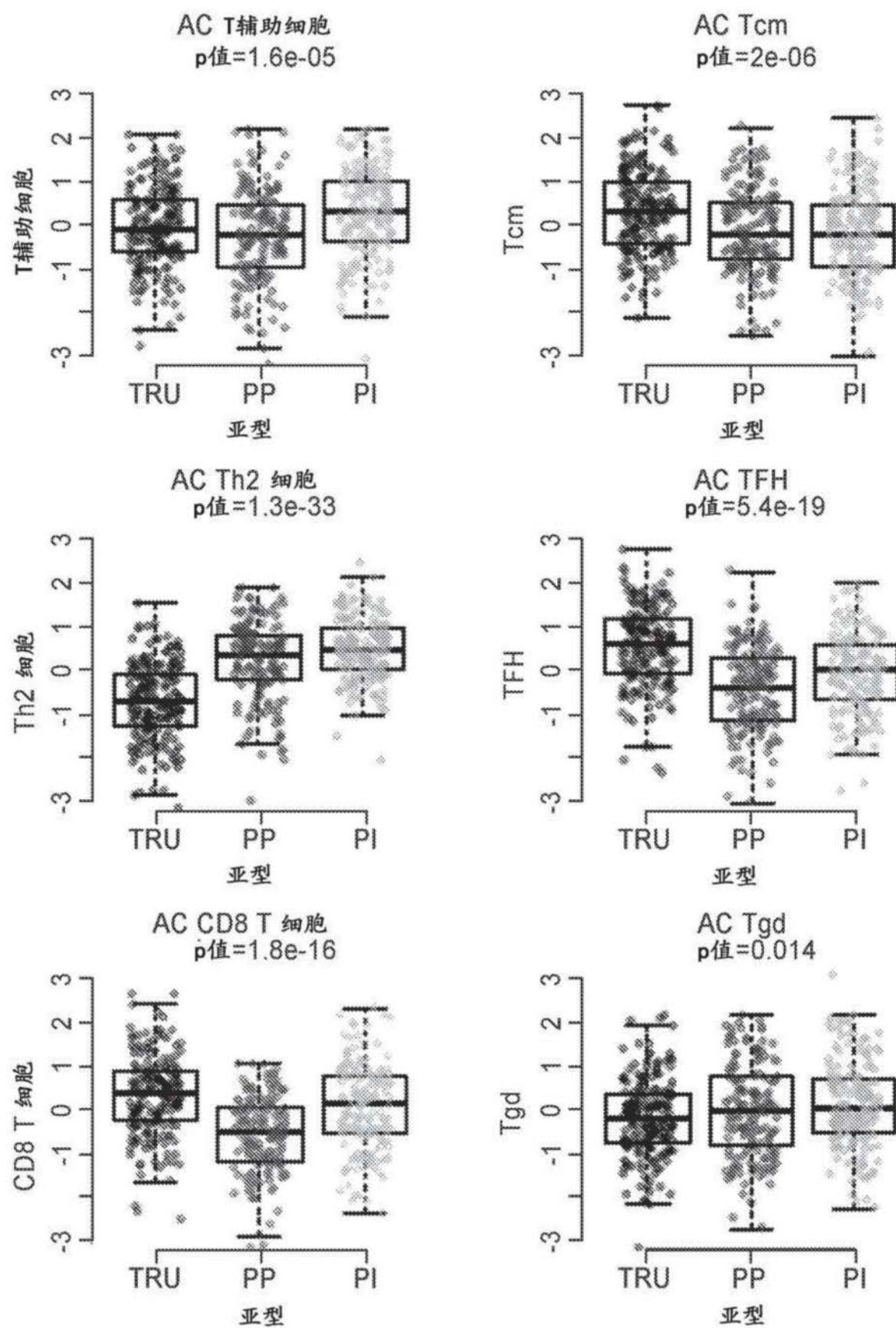


图18(续)

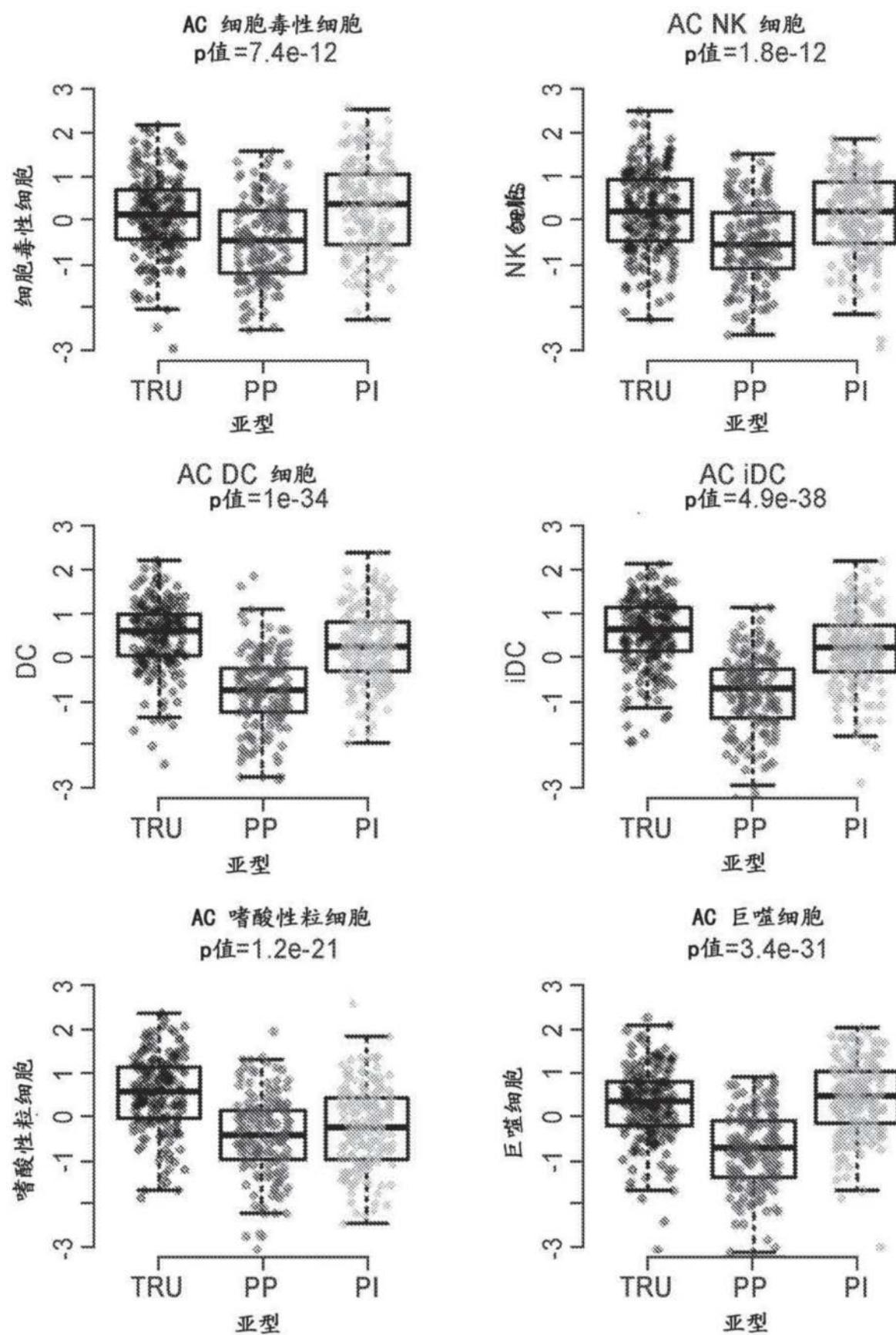


图18(续)

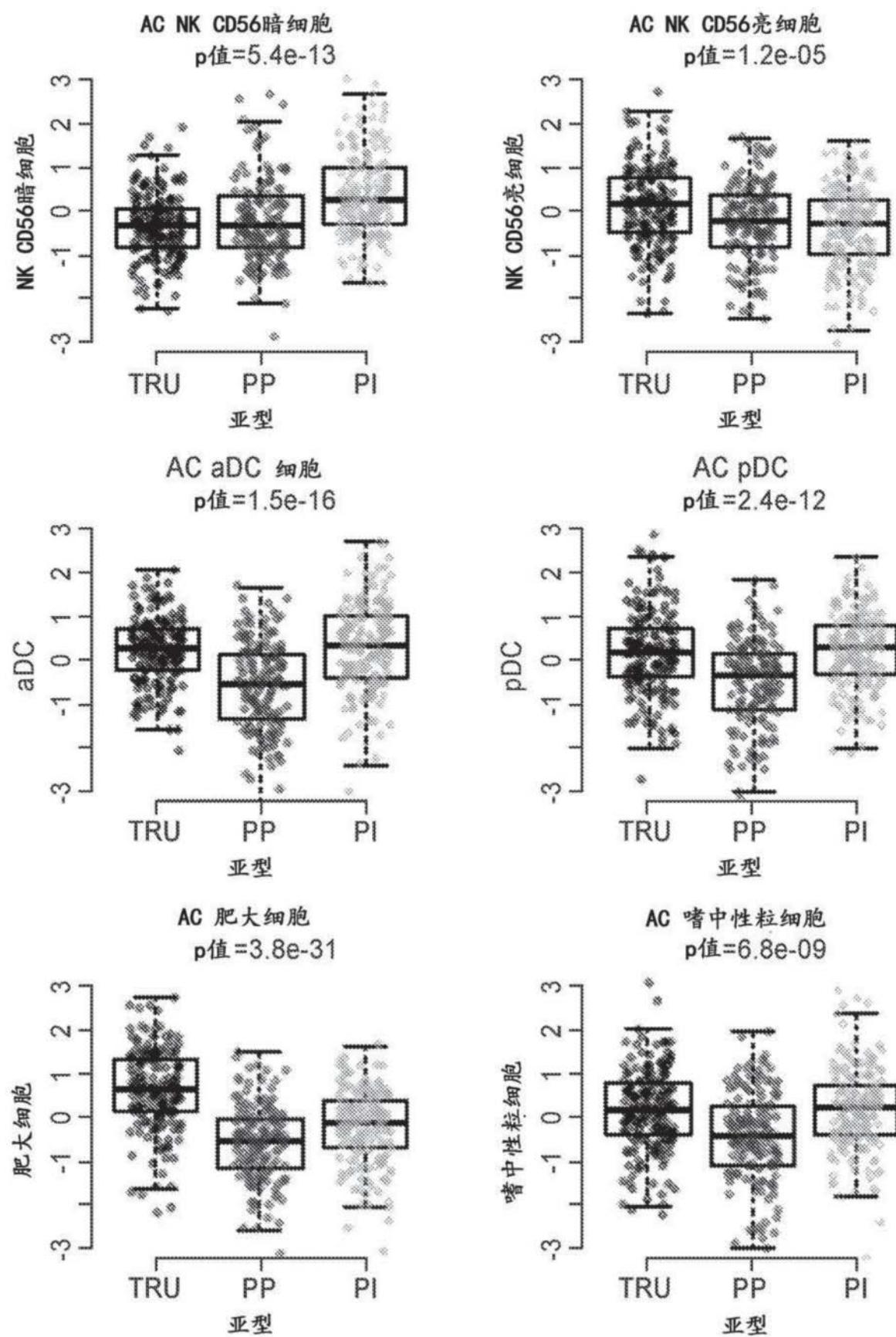


图18(续)

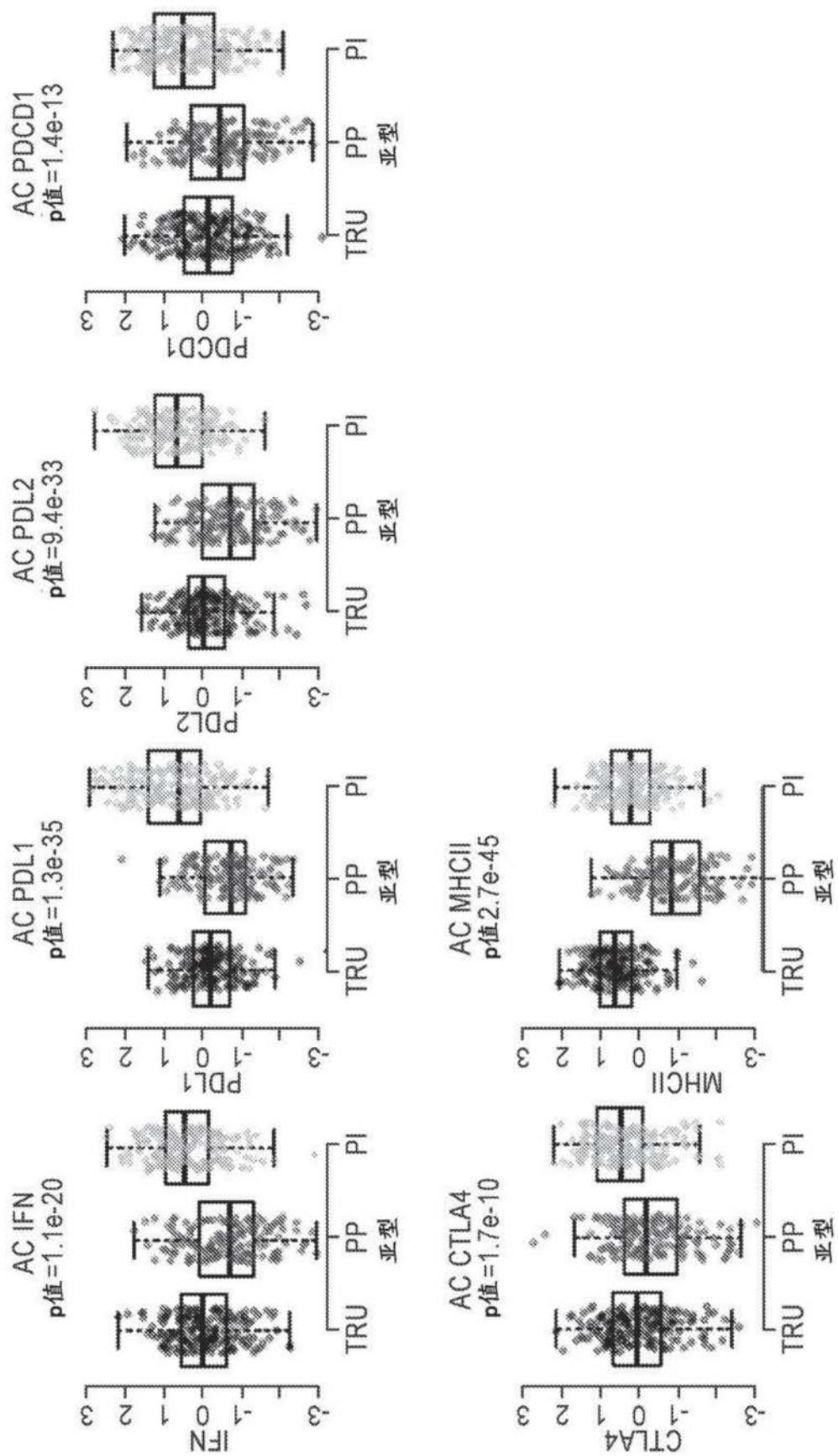


图18(续)

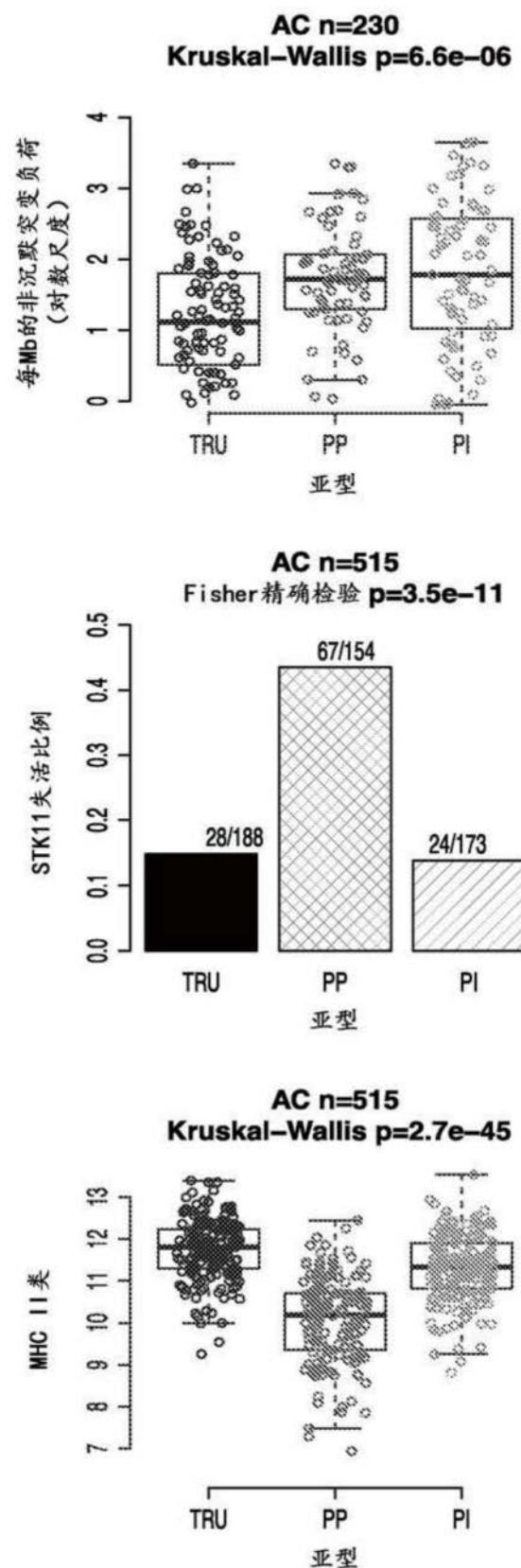


图19

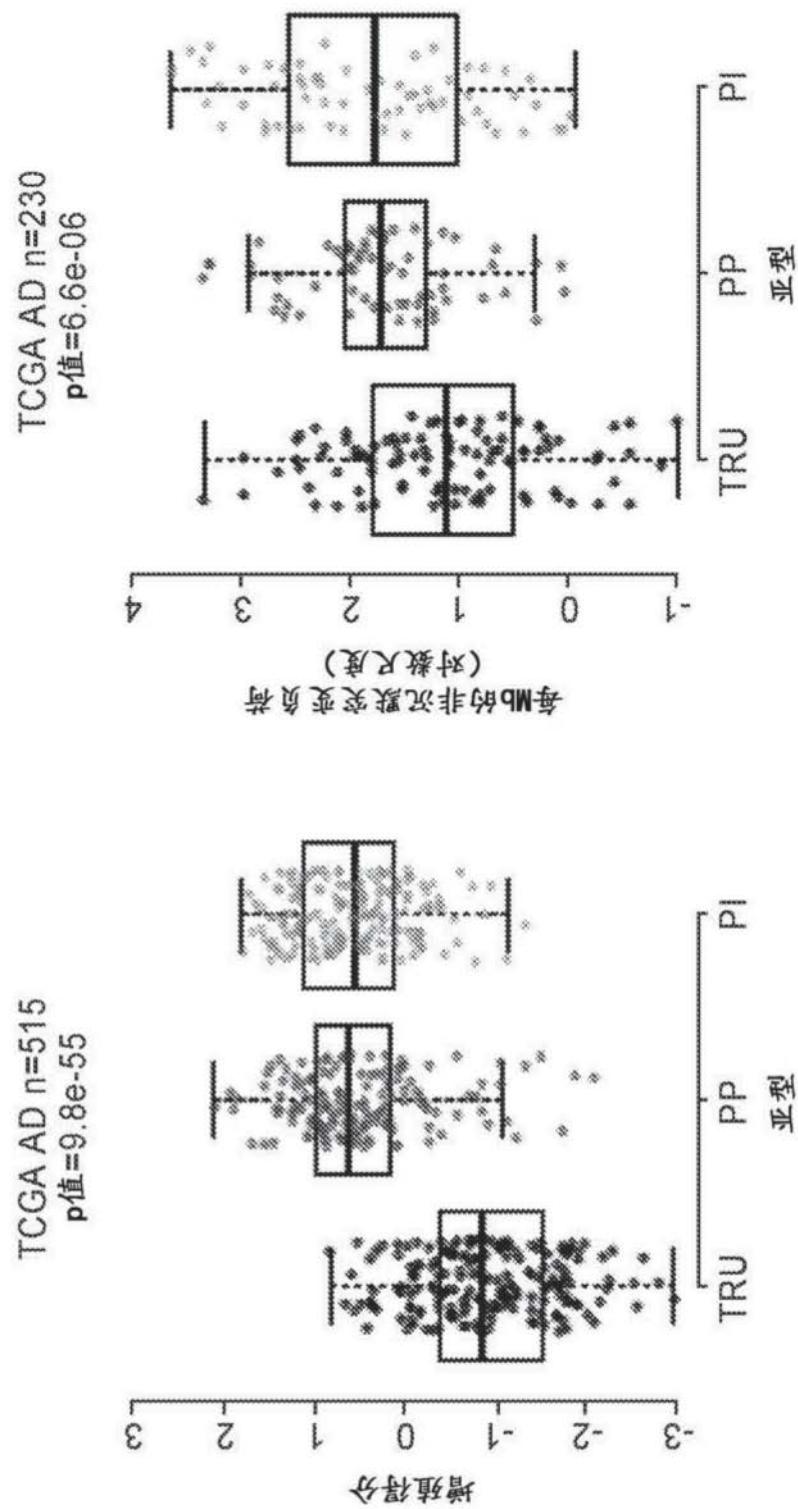


图20

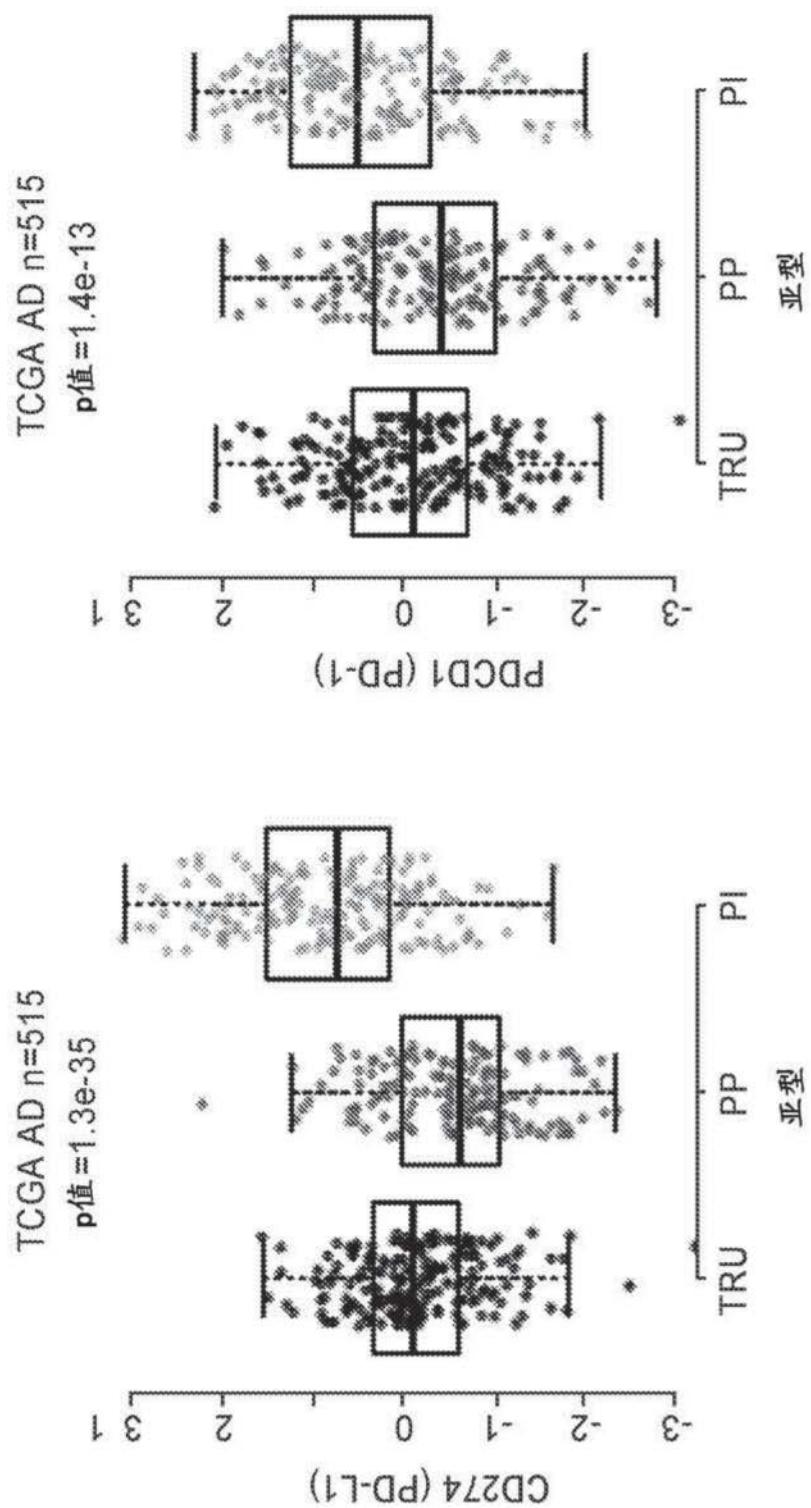


图20(续)

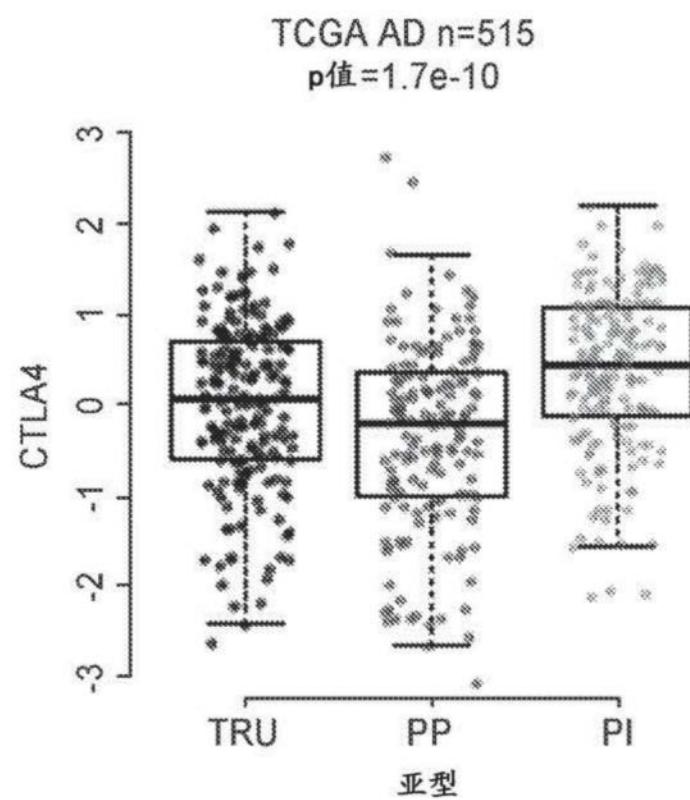
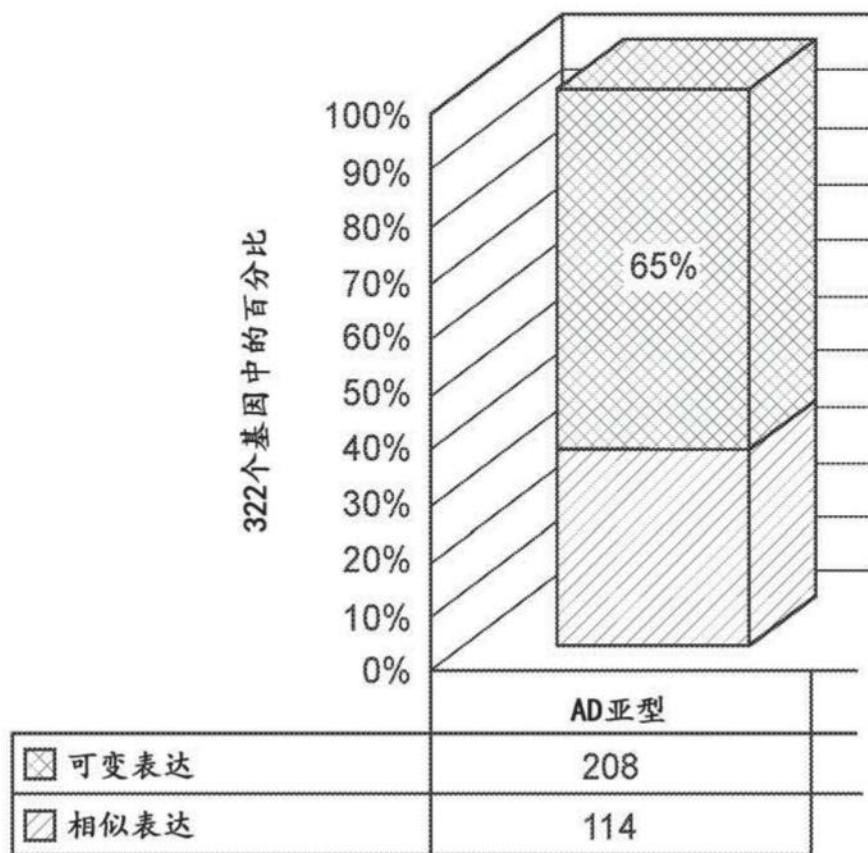


图20(续)



TCGA AD n=515

图21

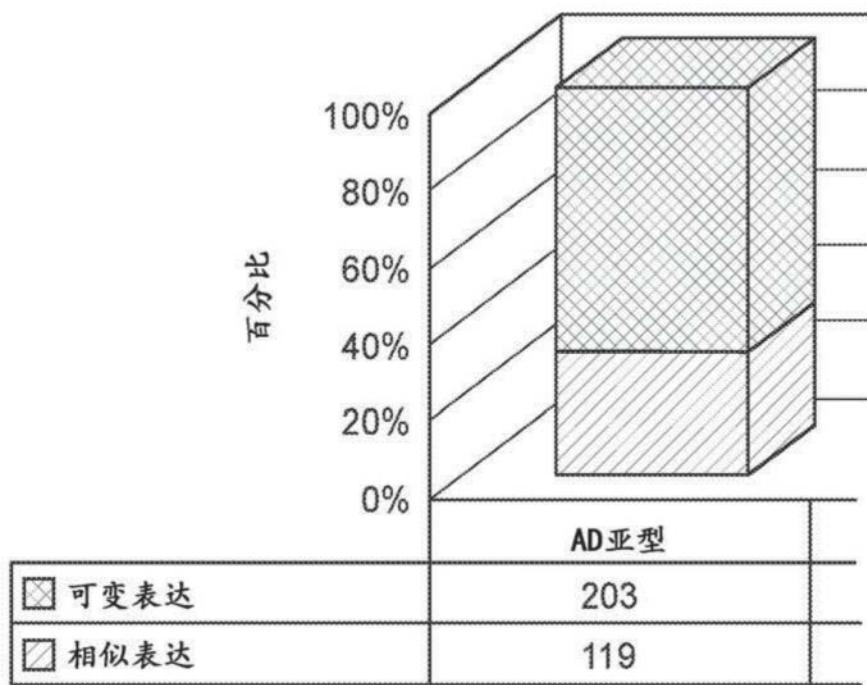


图22

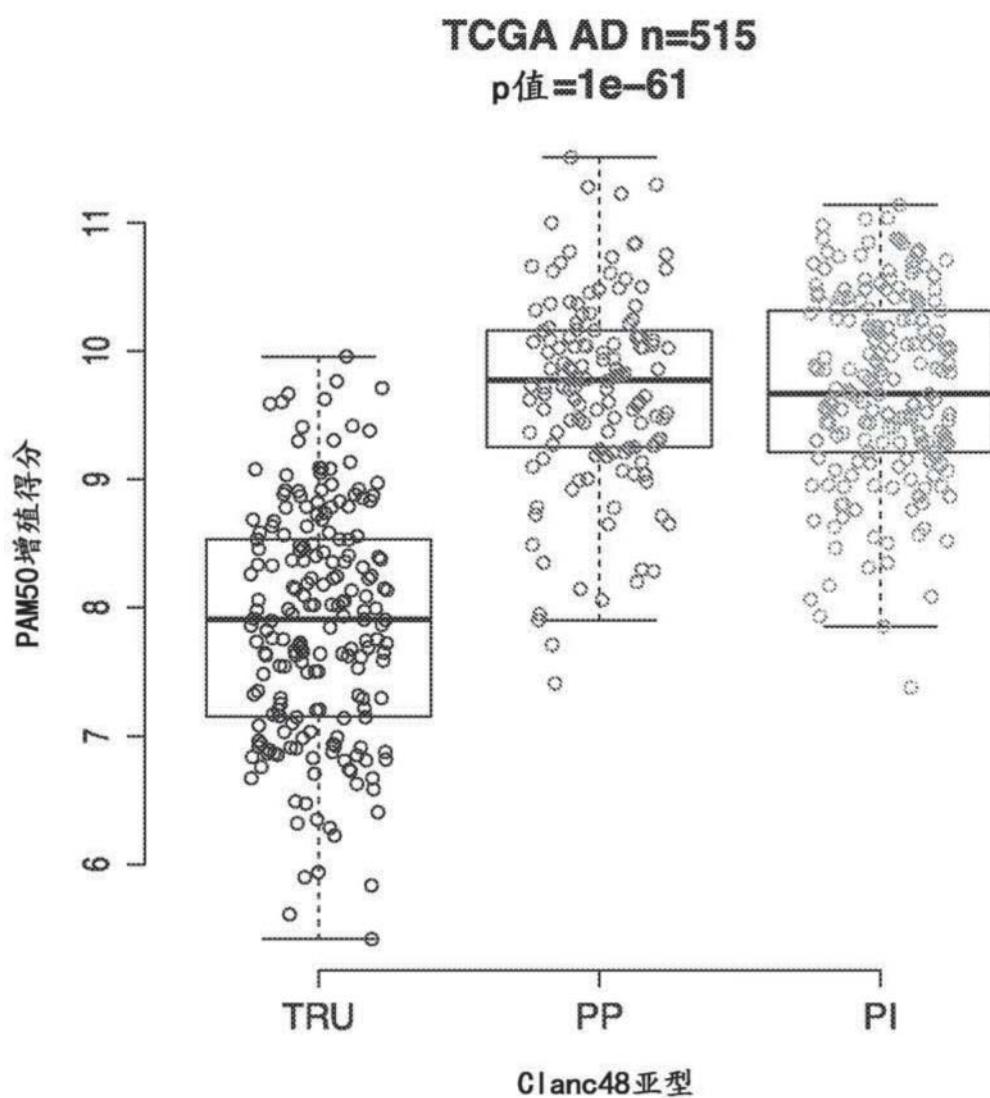


图23