



(12)发明专利

(10)授权公告号 CN 105102636 B

(45)授权公告日 2019.08.13

(21)申请号 201480015039.9

(72)发明人 M·阿尔比塔

(22)申请日 2014.03.13

(74)专利代理机构 上海专利商标事务所有限公司 31100

(65)同一申请的已公布的文献号

申请公布号 CN 105102636 A

代理人 余颖 陶家蓉

(43)申请公布日 2015.11.25

(51)Int.Cl.

C12Q 1/6886(2018.01)

(30)优先权数据

61/785,375 2013.03.14 US

(56)对比文件

US 2013058925 A1,2013.03.07,

WO 2012135008 A1,2012.10.04,

CN 102912030 A,2013.02.06,

WO 2010080933 A1,2010.07.15,

审查员 王胜佳

(85)PCT国际申请进入国家阶段日

2015.09.14

(86)PCT国际申请的申请数据

PCT/US2014/025860 2014.03.13

(87)PCT国际申请的公布数据

W02014/160120 EN 2014.10.02

(73)专利权人 尼欧基因组学实验室股份有限公

权利要求书1页 说明书21页

司

地址 美国佛罗里达州

序列表3页 附图6页

(54)发明名称

用于检测和测定前列腺癌预后的组合物和方法

(57)摘要

本公开提供检测和测定前列腺癌侵袭性的方法。这些方法可以用于确定患者是否需要活检以及指导治疗选择。

1. 一种有形计算机可读介质,其包括计算机执行时使计算机进行包括以下操作的计算机可读代码:

a) 接收对应于来自受试者的样本中的UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR基因的表达水平的信息;和

b) 接收对应于来自健康受试者的样本中至少PSA、ERG、GAPDH、B2M、PTEN、PCA3和PDLIM5基因的表达参考水平的信息,并测定至少PSA、ERG、GAPDH、B2M、PTEN、PCA3和PDLIM5基因的相对表达水平,其中所述基因与所述参考水平相比改变的表达表明所述受试者处于患有前列腺癌或侵袭性前列腺癌的风险,

其中计算机执行时所述计算机可读代码使计算机进行还包括以下的操作:

c) 计算所述样本的诊断评分,其中所述诊断评分指示所述样本来自患有前列腺癌或侵袭性前列腺癌的可能性。

2. 如权利要求1所述的有形计算机可读介质,其还包括接收对应于来自健康受试者的样本中的UAP1、IMPDH2、HSPD1、TMPRSS2或AR的表达参考水平的信息。

3. 如权利要求1所述的有形计算机可读介质,其中所述参考水平储存在所述有形计算机可读介质中。

4. 如权利要求1所述的有形计算机可读介质,其中所述接收信息包括从有形数据存储装置接收对应于来自受试者的样本中的一种或多种所述基因的表达水平的信息。

5. 如权利要求1所述的有形计算机可读介质,其还包括计算机执行时使计算机进行一种或多种包括以下的其他操作的计算机可读代码:向有形数据存储装置发送对应于一种或多种所述基因的相对表达水平的信息。

6. 如权利要求1所述的有形计算机可读介质,其中所述接收信息还包括接收对应于来自受试者的样本中的至少2种所述基因的表达水平的信息。

7. 如权利要求1所述的有形计算机可读介质,其中计算所述样本的诊断评分包括使用SVM算法、逻辑回归算法、lasso算法、boosting算法、bagging算法、随机森林算法、CART算法或MATT算法。

## 用于检测和测定前列腺癌预后的组合物和方法

[0001] 本申请要求2013年3月14日提交的美国临时专利申请号61/785,375的权益,其全部以引用方式并入本文。

[0002] 序列表的并入

[0003] 命名为“NGNLP0002W0\_ST25.txt”的文件中含有的序列表通过电子方式与本文一起提交并且以引用的方式并入本文,所述序列表的大小是4KB(在Microsoft Windows®中计算)并且创建于2014年3月13日。

[0004] 发明背景

### 1. 发明领域

[0005] 本发明总体上涉及癌症生物学的领域。更具体地讲,本发明涉及用于检测前列腺癌的存在以及确定前列腺癌侵袭性的方法。

[0006] 2. 相关技术描述

[0007] 前列腺癌是男性中继肺癌之后的第二最常见的癌症,并且其发病率由于人口老龄化日益升高。它也是男性中癌症相关死亡的第二主要原因。前列腺癌目前的筛选方法基于测定血清前列腺特异性抗原(PSA)。PSA水平 $\geq 4.0\text{ng每毫升}$ 是活检参照的一般阈值。已知PSA水平升高不实地表明可能存在前列腺癌,因为PSA水平升高也是良性前列腺增生(BPH)的特征,这是由于PSA水平和前列腺大小之间有相关性。依赖PSA水平导致75%假阳性以及太多不必要的活检。更重要地,即使检测到前列腺癌,这种癌症的临床行为显著不同,并且这种疾病在一些患者中可能是致死的而在其他患者中进展缓慢。目前的数据表明,通过依赖血清PSA,一些患者被过度治疗,因此,已经表明PSA测试可能导致更多伤害,这是由于不必要的前列腺切除术可能产生副作用。前列腺癌Gleason组织分级仍然是其临床表现最可靠的预测指标。有确信的数据证明当其肿瘤的Gleason评分是6时无论患者治疗与否都有相似的结果。

[0008] 为了改进血清PSA在其临床中的实用性已经进行许多尝试。游离和复合物PSA以及PSA的同种型已经作为PSA的辅助物使用,并且它们在灵敏度和特异性方面显示出一些改善,特别是在患者被认为处在“灰色地带”中的病例中,但是所有这些仍然不足以改善患有BPH的患者的癌症预测。还使用了PSA速度和倍增时间并且显示出一些改善,但是这种改善仍然有限制。不仅在预测癌症存在中需要改善PSA水平筛选从而避免不必要的活检,而且还需要开发也能够预测前列腺癌的临床表现的测试。

[0009] 发明概述

[0010] 本发明的实施方案提供一组可以用于前列腺癌的高精确度检测以及前列腺癌侵袭性测定的血液和尿液标记物。例如在一些方面,提供一种方法,所述方法用于根据受试者尿液样本中的至少一种mRNA和来自所述患者的血液样本中的至少一种mRNA的测定的表达水平鉴定受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险。在一些方面,这种方法还包括测定受试者血液中的至少一种蛋白质的水平。在其他方面,方法包括根据受试者尿液样本中的至少2或3种mRNA以及来自患者的血液样本中的至少2或3种mRNA的测定的表

达水平(以及任选受试者血液中的至少一种蛋白质的水平)鉴定受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险。

[0011] 因此,在一个实施方案中,提供检测受试者是否处于前列腺癌或侵袭性前列腺癌的风险的方法,其包括(a)从所述受试者获取生物样本;(b)测定所述样本中的至少3种基因的表达水平,所述至少3种基因选自由UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH和B2M组成的组;和(c)根据所述基因的表达水平鉴定所述受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险。在又一个方面,实施方案的方法包括(a)从所述受试者获取生物样本;(b)测定所述样本中的至少3种基因的表达水平,所述至少3种基因选自由UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR组成的组;和(c)根据所述基因的表达水平鉴定所述受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险。在一个方面,所述方法还包括鉴定受试者处于前列腺癌的风险。在另一个方面,所述方法还包括测定至少4、5、6、7、8、9、10、11或12种所述基因的表达水平。在又一个方面,所述方法还包括测定UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH和B2M基因的表达水平。在又一个方面,所述方法还包括测定UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR基因的表达水平。

[0012] 在实施方案的某些方面,受试者患有或被诊断患有前列腺癌。因此,方法可以包括鉴定患有癌症的受试者处于或没有处于侵袭性前列腺癌的风险。在某些方面,受试者先前已经进行前列腺切除术。在其他方面,受试者患有或被诊断患有前列腺肥大或良性前列腺增生(BPH)。

[0013] 在实施方案的一些方面,鉴定受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险是根据测定的基因的表达水平和受试者的年龄。在一个方面,鉴定受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险还包括使所述基因的表达水平与前列腺癌或侵袭性前列腺癌的风险关联。在一些情况下,这种关联步骤通过计算机进行。在一些方面,使用加权指示基因的测定的表达水平的相对预测值的算法。本文中提供这些算法的实例。在一些情况下,鉴定受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险还包括使用SVM算法、逻辑回归算法、lasso算法、boosting算法、bagging算法、随机森林算法、CART算法或MATT算法分析所述基因的表达水平。在一些情况下,这种分析可以通过计算机进行。

[0014] 在一些方面,根据实施方案使用的样本是血液样本、尿液样本或(在一些情况下)血液和尿液样本两者。在这些方面,所述方法还包括从受试者获取(直接或从第三方)血液样本或尿液样本。在又一个方面,所述方法还包括测定选自由血液或尿液样本中的UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR组成的组的至少3、4、5或更多种基因的表达水平。在又一个方面,所述方法还包括测定尿液样本中的UAP1、PDLIM5、IMPDH2、PCA3、TMPRSS2和/或HSPD1的表达水平。在又一个方面,所述方法还包括测定血液样本中的UAP1、IMPDH2、HSPD1、PSA和/或ERG的表达水平。

[0015] 在另一个方面,实施方案的方法包括(i)测定所述尿液样本中的HSPD1、IMPDH2和PDLIM5的表达水平和所述血液样本中的ERG的表达水平;(ii)测定尿液样本中的MPDH2、HSPD1、PCA3和PDLIM5的表达水平以及血液样本中的ERG和PSA的表达水平;或(iii)测定尿液样本中的MPDH2、HSPD1、PCA3和PDLIM5的表达水平以及血液样本中的UAP1、ERG和PSA的表达水平。

[0016] 在其他方面,实施方案的方法包括测定受试者的(i)尿液样本中的PCA3、PTEN和B2M的表达水平(例如,mRNA表达水平)和(ii)血液样本中的ERG、AR、B2M和GAPDH的表达水平(例如,mRNA表达水平)并且根据所述基因的表达水平鉴定受试者处于或没有处于前列腺癌(对比BPH)的风险。在一些方面,这种方法还包括测定受试者血液中的PSA蛋白的水平。因此,在实施方案的特定方面,方法包括测定(i)血液样本中的PSA的蛋白质表达水平;受试者的(ii)尿液样本中的PCA3、PTEN和B2M的mRNA表达水平和(iii)血液样本中的ERG、AR、B2M和GAPDH的mRNA表达水平并且根据表达水平鉴定受试者处于或没有处于前列腺癌(对比BPH)的风险。

[0017] 在又一些其他方面,实施方案的方法包括测定受试者的(i)尿液样本中的PSA、GAPDH、B2M、PTEN、PCA3和PDLIM5的表达水平(例如,mRNA表达水平)和(ii)血液样本中的ERG的表达水平(例如,mRNA表达水平)并且根据所述基因的表达水平鉴定受试者处于或没有处于侵袭性前列腺癌的风险。例如,在一些特定的方面,实施方案的方法包括测定受试者的(i)尿液样本中的PSA、GAPDH、B2M、PTEN、PCA3和PDLIM5的表达水平(例如,mRNA表达水平)和(ii)血液样本中的ERG、PCA3、B2M和HSPD1的表达水平(例如,mRNA表达水平)并且根据所述基因的表达水平鉴定受试者处于或没有处于侵袭性前列腺癌的风险。在一些方面,这种方法还包括测定受试者血液中的PSA蛋白的水平和/或确定受试者的年龄。因此,在实施方案的一些特定的方面,方法包括测定受试者的(i)血液样本中的PSA的蛋白质表达水平;(ii)尿液样本中的PSA、GAPDH、B2M、PTEN、PCA3和PDLIM5的mRNA表达水平和(iii)血液样本中的ERG、PCA3、B2M和HSPD1的mRNA表达水平并且根据表达水平鉴定受试者处于或没有处于侵袭性前列腺癌的风险。

[0018] 在实施方案的另一个方面,方法包括(a)测定受试者的(i)血液样本中的PSA的蛋白质表达水平;(ii)尿液样本中的PCA3、PTEN和B2M的mRNA表达水平和(iii)血液样本中的ERG、AR、B2M和GAPDH的mRNA表达水平并且根据表达水平确定受试者的第一前列腺癌风险因素;(b)测定受试者的(i)血液样本中的PSA的蛋白质表达水平;(ii)尿液样本中的PSA、GAPDH、B2M、PTEN、PCA3和PDLIM5的mRNA表达水平和(iii)血液样本中的ERG、PCA3、B2M和HSPD1的mRNA表达水平并且根据表达水平确定受试者的第二个前列腺癌风险因素;和(c)根据所述第一个和第二个前列腺癌风险因素鉴定受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险。在一些方面,这种方法可以用于选择供活检或抗癌疗法的受试者。

[0019] 在又一个方面,方法还包括测定样本中的基因的表达水平并且测定参考样本中的基因的表达水平;和通过比较来自所述受试者的所述样本中的所述基因的表达水平与所述参考样本中的所述基因的表达水平鉴定所述受试者处于或没有处于前列腺癌或侵袭性前列腺癌的风险。

[0020] 在一些方面,测定所述基因的表达包括测定你。测定蛋白质表达水平可以包括例如,进行ELISA、蛋白质印迹或与抗体阵列结合。在另一个方面,测定所述基因的表达包括测定RNA表达水平。测定RNA表达水平可以包括进行RT-PCR、RNA印迹或阵列杂交。优选地,测定基因的表达水平包括进行RT-PCR(例如,实时RT-PCR)。

[0021] 在一些方面,方法还包括报告受试者是否患有前列腺癌或患有侵袭性前列腺癌。报告可以包括准备口头、书面或电子报告。因此,提供报告可以包括向患者、医生、医院或保险公司提供报告。

[0022] 在另一个实施方案中,本公开提供治疗受试者的方法,所述方法包括根据实施方案选择鉴定为处于前列腺癌或侵袭性前列腺癌的风险的受试者并且对所述受试者施用抗癌疗法。例如,方法可以包括(a)获取来自受试者的样本中的至少3种基因的表达水平,所述至少3种基因选自由UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR组成的组;(b)根据所述基因的表达水平选择患有前列腺癌或患有侵袭性前列腺癌的受试者;和(c)用抗癌疗法治疗选择的受试者。在某些方面,抗癌疗法是化学疗法、放射疗法、激素疗法、靶向疗法、免疫疗法或手术疗法(例如,前列腺切除术)。

[0023] 在另一个实施方案中,本公开提供选择供诊断程序的受试者的方法,所述方法包括(a)获取来自受试者的样本中的至少3种基因的表达水平,所述至少3种基因选自由UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR组成的组;(b)根据所述基因的表达水平选择有患有前列腺癌或患有侵袭性前列腺癌风险的受试者;和(c)对受试者进行诊断程序。例如,诊断程序可以是活检。

[0024] 在又一个实施方案中,本公开提供确定患有前列腺癌的受试者的预后的方法,所述方法包括(a)从所述受试者获取生物样本;(b)测定所述样本中的至少3种基因的表达水平,所述至少3种基因选自由UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR组成的组;和(c)根据所述基因的表达水平鉴定所述受试者患有或不患有侵袭性前列腺癌。

[0025] 在又一个实施方案中,本公开提供有形计算机可读介质,其包括计算机执行时使计算机进行包括以下操作的计算机可读代码:(a)接收对应于来自受试者的样本中的UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR基因的表达水平的信息;和(b)测定一种或多种所述基因与参考水平相比的相对表达水平,其中一种或多种所述基因与参考水平相比改变的表达表明所述受试者处于患有前列腺癌或侵袭性前列腺癌的风险。

[0026] 在一个方面,有形计算机可读介质还包括接收对应于来自健康受试者的样本中的UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、PTEN和AR的参考表达水平的信息。在另一个方面,有形计算机可读介质还包括计算机执行时使计算机进行一种或多种包括以下的其他操作的计算机可读代码:向有形数据存储装置发送对应于一种或多种所述基因的相对表达水平的信息。在又一个方面,计算机可读代码是计算机执行时使计算机进行还包括以下操作的代码:(c)计算样本的诊断评分,其中所述诊断评分指示所述样本来自患有前列腺癌或侵袭性前列腺癌的受试者的可能性。在一个方面,计算样本的诊断评分包括使用SVM算法、逻辑回归算法、lasso算法、boosting算法、bagging算法、随机森林算法、CART算法或MATT算法。

[0027] 在又一个其他方面,参考水平储存在所述有形计算机可读介质中。在另一个方面,接收信息包括从有形数据存储装置接收对应于来自受试者的样本中的一种或多种所述基因的表达水平的信息。在又一个方面,接收信息还包括接收对应于来自受试者的样本中的至少2、3、4、5、6、7、8、9、10、11或12种所述基因的表达水平的信息。

[0028] 如本文在说明书中所用,“一个”或“一种”可意指一个(种)或多个(种)。如本文在权利要求中所用,当与词语“包括/包含(comprising)”结合使用时,词语“一个”或“一种”可意指一个(种)或多于一个(种)。

[0029] 除非明确指明仅仅指代替物,或代替物相互排斥,否则权利要求书中所用的术语“或”用于意指“和/或”,尽管本公开支持仅仅指代替物和“和/或”的定义。如本文所用的“另一”可意指至少第二个(种)或更多个(种)。

[0030] 在整个本申请中,术语“约”用于指明值包括装置、用以测定所述值的方法的误差的固有变化,或研究受试者中存在的变化。

[0031] 通过以下详细描述,本发明的其他目的、特征和优势将变得显而易见。然而,应理解的是,尽管指示本发明的优选实施方案,但是详细描述和特定实施例仅通过说明的方式给出,因为从此详细描述中,本发明的精神和范围内的各种改变和修改对于本领域技术人员来说将变得显而易见。

[0032] 附图简述

[0033] 以下附图组成本说明书的一部分,并且被包括以进一步说明本发明的某些方面。本发明可通过参考与本文提出的特定实施方案的详细描述结合的这些附图中的一个或多个来更好地理解。

[0034] 图1.在训练组中使用各种算法的AUC(图1A)和误差率(图1B)。还显示了算法中包括的六个变量各自的贡献(图1C)。

[0035] 图2.使用样本的测试组,用多种算法显示AUC(图2A)和误差率(图2B)。

[0036] 图3.确定从BPH区分癌症患者的分界点。中间的虚线是0.565,左侧和右侧虚线分别是0.55和0.58。

[0037] 图4.在训练组中使用各种算法的AUC(图4A)和误差率(图4B)。还显示了算法中包括的四个变量各自的贡献(图4C)。

[0038] 图5.从BPH/Gleason<7区分侵袭性前列腺癌的ROC曲线。

[0039] 图6.使用两个模型(癌症对比非癌症以及侵袭性癌症对比BPH/进展缓慢癌症)来预测的组合评分系统。每个方形表示一个患者。患者的分布在上两行中显示。75%具有一致结果(灵敏度=68%,特异性=99%)。25%Pog/Neg:混合:阴性/阳性<7/阳性≥7。

[0040] 图7.用于将PCa与BPH区分的试验数据的ROC曲线。分析中使用的标记物是(1)血清PSA蛋白水平;(2)血浆ERG mRNA水平;(3)血浆AR mRNA水平;(4)尿液PCA3mRNA水平;(5)尿液PTEN水平;(6)尿液B2M mRNA水平;(7)血浆B2M mRNA水平;和(8)血浆GAPDH mRNA水平。

[0041] 图8.用于从BPH/Gleason<7区分侵袭性前列腺癌的测定数据的ROC曲线。曲线显示使用不同数目标记物时的结果的曲线(即,梯级0是1种标记物;梯级1是两种标记物;梯级2是三种标记物等等...)。梯级8曲线中使用的标记物(达到0.79777的AUROC)是(1)血清PSA蛋白水平;(2)年龄;(3)尿液PSA;(4)血浆ERG mRNA水平;(5)尿液GAPDH mRNA水平;(6)尿液B2M mRNA水平;(7)尿液PTEN mRNA水平;(8)尿液PCA3mRNA水平;和(9)尿液PDLIM5mRNA水平。

[0042] 说明性实施方案的描述

[0043] 本文公开两种算法,一种用于预测患有良性前列腺增生(BPH)的患者中的前列腺癌的存在,并且第二种用于预测侵袭性前列腺癌(Gleason≥7)的存在。这些算法是通过经由实时PCR测定从尿液和血浆两者分离的生物标记物的组合开发的,这些生物标记物包括UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH和B2M。因此,本公开提供利用两种算法的评分系统来检测侵袭性前列腺癌。这种评分系统在75%的患者中提供存在侵袭

性前列腺癌的高精确预测(99%特异性和68%灵敏度)。在25%的患者中,只能够以88%特异性和67%灵敏度预测癌症的存在,但是不能预测疾病的侵袭性。这种手段可以用于确定患者是否需要活检以及什么时候存在所述活检可能是非代表性的疑问。

[0044] 第一种算法预测训练组中的AUC为0.77以及测试组中的AUC为0.78的癌症。总体的特异性和灵敏度分别是88%和67%。第二种算法预测Gleason $\geq 7$ 的患者,在训练组中显著更好的0.87的AUC和在测试组中0.88的AUC(99%特异性和47%灵敏度)。通过在评分系统中并入所述两个模型,75%的患者在两个模型之间显示一致性。在经由两个模型的一致的患者中,Gleason $\geq 7$ 的预测特异性为99%并且灵敏度为68%。在两个模型之间显示不一致的患者中,疾病侵袭性的预测不准确并且只可以使用第一个模型预测癌症对比没有癌症。

[0045] 所述测定随后通过并入两种额外的标记物(AR和PTEN mRNA水平)进一步开发。此外开发了用于(I)测定PCa对比BPH;和(II)高风险PCa(GS $\geq 7$ )对比低风险癌症(GS $< 7$ )或BPH的测定。对于这些分析的第一种(将PCa与BPH区分)使用的标记物是(1)血清PSA蛋白水平;(2)血浆ERG mRNA水平;(3)血浆AR mRNA水平;(4)尿液PCA3mRNA水平;(5)尿液PTEN水平;(6)尿液B2M mRNA水平;(7)血浆B2M mRNA水平;和(8)血浆GAPDH mRNA水平。使用这些标记物可以用0.87的AUROC将PCa与BPH区别开来。在使用0.64的分界点时,这个模型的测试组显示76%的灵敏度和71%的特异性(参见,例如,图7和表5)。使用以下标记物开发第二种分析(区分高风险PCa(GS $\geq 7$ )对比GS $< 7$ 癌症或BPH):(1)血清PSA蛋白水平;(2)年龄;(3)尿液PSA;(4)血浆ERG mRNA水平;(5)尿液GAPDH mRNA水平;(6)尿液B2M mRNA水平;(7)尿液PTEN mRNA水平;(8)尿液PCA3mRNA水平;(9)尿液PDLIM5mRNA水平;和任选地(10)血浆PCA3mRNA水平;(11)血浆B2M mRNA水平和(12)血浆HSPD1mRNA水平。使用这些标记物,可以用0.80的AUROC将高风险PCa与低等级癌症(GS $< 7$ )或BPH区别开来。

[0046] 此外,通过组合上文描述的两种分析的结果可以获得高特异性和灵敏度的诊断(不需要活检)。在两种分析都是阴性的情况下,有较高没有癌症的可能性,并且在任何情况下,有非常低的高风险癌症的可能性。这些受试者因此可以放弃更具创伤性的诊断,如活检,并且会需要更少的监测。另一方面,当两种分析都是阳性时,有较高可能性所述受试者患有癌症并且癌症是侵袭性的。这些受试者会经受活检和/或(侵袭性)抗癌疗法,如手术切除。同样地,如果测定表明受试者是“PCa阴性”但是高风险癌症阳性,那么所述受试者患有癌症的可能性较高并且所述癌症是高风险的。此外,这些受试者将会经受活检和/或(侵袭性)抗癌疗法。在指示为“PCa阳性”但是高风险PCa阴性的患者的情况下,所述患者患有癌症的可能性高,但是所述癌症不太可能是高风险的。这些受试者可以经受活检,但是可能不会需要立即的侵袭性疗法或监测。

[0047] 因此,新开发的测定和分析特别有助于确定需要进行前列腺活检并且可以有助于在主动监视上监测患者以及预测发展。然而,PCa的存在和侵袭性的这种预测是以活检结果为基础的。

[0048] 具体来讲,本文中鉴定的尿液和血浆表达标记物包括:

[0049] -PDZ和LIM结构域5(PDLIM5)参见,例如NCBI登记号NM\_006457.4、NM\_001011513.3、NM\_001011515.2、NM\_001011516.2、NM\_001256425.1、NM\_001256426.1、NM\_001256427.1、NM\_001256428.1、NR\_046186.1和NM\_001256429.1,以引用方式并入本文。

[0050] -跨膜蛋白酶,丝氨酸2(TMPRSS2),参见例如,NCBI登记号NM\_001135099.1和NM\_

005656.3,以引用方式并入本文。

[0051] -UDP-N-乙酰氨基葡萄糖焦磷酸化酶1 (UAP1) ,参见例如,NCBI登记号NM\_003115.4,以引用方式并入本文。

[0052] -IMP(肌昔5'-单磷酸)脱氢酶2 (IMPDH2) ,参见例如,NCBI登记号NM\_000884.2,以引用方式并入本文。

[0053] -热休克60kDa蛋白1(分子伴侣) (HSPD1) ,参见例如,NCBI登记号NM\_002156.4和NM\_199440.1,以引用方式并入本文。

[0054] -前列腺癌抗原3 (PCA3) ,参见例如,NCBI登记号NR\_015342.1,以引用方式并入本文。

[0055] -PSA或激肽释放酶相关肽酶3 (KLK3) ,参见例如,NCBI登记号NM\_001030047.1、NM\_001030048.1和NM\_001648.2,以引用方式并入本文。

[0056] -v-ets骨髓成红血细胞增多症病毒E26致癌基因同系物 (ERG) ,参见例如,NCBI登记号NM\_001136154.1、NM\_001136155.1、NM\_001243428.1、NM\_001243429.1、NM\_001243432.1、NM\_004449.4和NM\_182918.3,以引用方式并入本文。

[0057] -PTEN或磷酸酶和张力蛋白同系物,参见例如,NCBI登记号NM\_000314.4,以引用方式并入本文。

[0058] -AR或雄激素受体,参见例如,NCBI登记号NM\_000044.3和NM\_001011645.2,以引用方式并入本文。

[0059] -甘油醛-3-磷酸脱氢酶 (GAPDH) ,参见例如,NCBI登记号NM\_001256799.1和NM\_002046.4,以引用方式并入本文。

[0060] - $\beta$ -2-微球蛋白 (B2M) ,参见例如,NCBI登记号NM\_004048.2,以引用方式并入本文。

#### [0061] I. 生物标记物检测

[0062] 生物标记物或基因的表达可以通过本领域熟知的多种技术测定。定量生物标记物的信使RNA (mRNA) 的水平可以用于测定生物标记物的表达。另外,定量生物标记物的蛋白产物的水平可以用于测定生物标记物的表达。关于下面讨论的方法的附加信息可以在Ausubel等 (2003) 或Sambrook等 (1989) 中找到。本领域技术人员将会了解可以操作哪些参数来优化目标mRNA或蛋白的检测。

[0063] 在一些实施方案中,所述获取表达信息可以包括RNA定量,例如,cDNA微阵列、定量RT-PCR、原位杂交、RNA印迹或核酸酶保护。所述获取表达信息可以包括蛋白质定量,例如,包括免疫组织化学、ELISA、放射免疫测定 (RIA) 、免疫放射测定、荧光免疫测定、化学发光测定、生物发光测定、凝胶电泳、蛋白质印迹分析、质谱分析或蛋白质微阵列的蛋白定量。

[0064] 核酸微阵列可以用于定量多个生物标记物的差异表达。微阵列分析可以按照制造商方案使用市售设备进行,如使用Affymetrix GeneChip®技术 (Santa Clara, CA) 或来自Incyte的微阵列系统 (Fremont, CA) 。例如,可以将单链核酸(例如,cDNA或寡核苷酸) 平铺或排列在微芯片基底上。随后将排列的序列与来自目标细胞的特异性核酸探针杂交。荧光标记的cDNA探针可以通过借助于从目标细胞提取的RNA的逆转录并入荧光标记的脱氧核苷酸来产生。另外, RNA可以通过体外转录扩增并且用标记物(如生物素) 标记。随后在非常严格的条件下将标记的探针与微芯片上的固定核酸杂交。在严格的洗涤以移除非特异性结合的探针之后,通过共焦激光显微镜或另一种检测方法(如CCD摄像机) 扫描芯片。杂交文件中的

原始荧光强度数据通常用加强多芯片平均(RMA)算法预处理从而产生表达值。

[0065] 定量实时PCR(qRT-PCR)还可以用于测定多个生物标记物的差异表达。在qRT-PCR中, RNA模板通常被逆转录成cDNA, 它随后通过PCR反应扩增。实时逐循环追踪PCR产物的量, 这允许测定mRNA初始浓度。为了测定PCR产物的量, 反应可以在荧光染料存在下进行, 如SYBR绿, 它结合双链DNA。所述反应还可以用对扩增的DNA特异性的荧光报告分子探针进行。

[0066] 荧光报告分子探针的非限制性实例是TaqMan®探针(Applied Biosystems, Foster City, CA)。当在PCR延伸循环期间移除淬灭团时荧光报告分子探针发荧光。多重qRT-PCR可以通过使用多个基因特异性报告分子探针(每个含有不同的荧光团)进行。在每个循环期间记录荧光值并且荧光值代表扩增反应中至那个点的扩增的产物的量。为了使误差最小化并且减少任何两个样本之间的差异, 可以使用参考标准进行qRT-PCR。理想的参考标准在不同组织之间以恒定水平表达, 并且不受实验处理影响。合适的参考标准包括但不限于管家基因甘油醛-3-磷酸-脱氢酶(GAPDH)及β-肌动蛋白的mRNA。在原始样本中的mRNA水平或每种生物标记物的表达的倍数变化可以使用本领域熟知的计算确定。

[0067] 还可以使用免疫组织化学染色来测定多个生物标记物的差异表达。这种方法能够通过蛋白质与特异性抗体的相互作用实现组织切片的细胞中的蛋白质的定位。为此, 可以将组织固定在甲醛或另一种合适的固定剂中, 包埋在蜡或塑料中, 并且使用切片机切割成薄切片(约0.1mm至几mm厚)。另外, 可以将组织冷冻并且使用恒冷切片机切割成薄切片。可以将组织切片排列并且固定在固体表面上(即, 组织微阵列)。将组织切片与针对目标抗原的一抗孵育, 随后冲洗以除去未结合的抗体。所述一抗可以与检测系统偶合, 或者可以用与检测系统偶合的二抗检测所述一抗。检测系统可以是荧光团或者它可以是酶(如辣根过氧化酶或碱性磷酸酶), 它可以将底物转化成比色、荧光或化学发光产物。通常在显微镜下扫描染色的组织切片。因为来自癌症受试者的组织样本可能是异质的, 即, 一些细胞可能是正常的而其他细胞可能是癌性的, 所以可以测定组织中的阳性染色细胞的百分比。这个测量结果以及染色强度的定量可以用于产生生物标记物的表达值。

[0068] 酶联免疫吸附测定或ELISA可以用于测定多个生物标记物的差异表达。有许多ELISA测定的变型。所有都是基于在固体表面(通常为微量滴定板)上固定抗原或抗体。最初的ELISA方法包括制备含有目标生物标记物蛋白的样本, 用样本涂覆微量滴定板的孔, 将每个孔与识别特异性抗原的一抗孵育, 冲洗未结合的抗体, 并且随后检测抗体-抗原复合物。抗体-抗体复合物可以直接检测。为此, 将一抗与检测系统(如产生可检测产物的酶)缀合。抗体-抗体复合物可以间接检测。为此, 如上所述, 通过与检测系统缀合的二抗检测一抗。随后扫描微量滴定板并且可以使用本领域已知的方法将原始强度数据转化成表达值。

[0069] 还可以使用抗体微阵列来测定多个生物标记物的差异表达。为此, 将多个抗体排列并且共价连接至微阵列或生物芯片的表面。通常用荧光染料或生物素标记含有目标生物标记物蛋白质的蛋白质提取物。将标记的生物标记物蛋白质与抗体微阵列孵育。冲洗以除去未结合的蛋白之后, 扫描微阵列。可以使用本领域已知的方法将原始荧光强度数据转化成表达值。

[0070] 还可以使用Luminex多重微球来测定多个生物标记物的差异表达。这些微小的聚苯乙烯珠粒内部用荧光染料标记颜色, 从而每个珠粒都具有独特的光谱特征(其中最多有100种)。用将会结合目标靶标(即, 分别是生物标记物mRNA或蛋白质)的特异性寡核苷酸或

特异性抗体标记具有相同特征的珠粒。反过来,靶标也用荧光报告分子标记。因此,有两种颜色来源,一种来自珠粒而另一种来自靶标上的报告分子。随后将珠粒与含有靶标的样本孵育,其中在一个孔中可以检测到最多100种。珠粒的小尺寸/表面积以及珠粒对靶标的三维接触几乎允许结合反应期间的溶液相动力学。捕获的靶标通过基于流式细胞术的高科技射流技术检测,其中激光激发标识每个珠粒的内部染料并且还激发测定期间捕获的任何报告分子染料。来自采集文件的数据可以使用本领域已知的方法转化成表达值。

[0071] 还可以使用原位杂交来测定多个生物标记物的差异表达。这种方法允许组织切片的细胞中的目标mRNA的定位。对于这种方法,可以将组织冷冻或固定并且包埋,并且随后切割成薄切片,将薄切片排列并且固定在固体表面上。将组织切片与将会与目标mRNA杂交的标记的反义探针孵育。杂交和冲洗步骤通常在非常严格的条件下进行。探针可以用可被另一种蛋白质或抗体检测的荧光团或小标记物(如生物素或洋地黄毒昔)标记,从而使标记的杂交物可以在显微镜下检测和观察。可以同时检测多个mRNA,前提是每个反义探针都具有可识别的标记。通常在显微镜下扫描杂交的组织阵列。因为来自癌症受试者的组织样本可能是异质的,即,一些细胞可能是正常的而其他细胞可能是癌性的,所以可以测定组织中的阳性染色细胞的百分比。这个测量结果以及染色强度的定量可以用于产生每个生物标记物的表达值。

[0072] 在又一个实施方案中,可将标记物水平与来自对照的标记物水平比较,其中对照可以包括取自一个或多个患者的一个或多个肿瘤样本,所述患者被确定为患有某种转移性肿瘤或不患有某种转移性肿瘤,或两种情况。

[0073] 对照可以包括作为患者个体数据的同时获取的数据(例如,在相同杂交实验中),或可以是例如储存在计算机或计算机可读介质上的储存值或数值集合。如果使用后者,那么从初始或后续样本获取的用于选择标记物的新患者数据就可以与相同标记物的存储数据比较而不需要额外的对照实验。

[0074] 标记物表达的统计分析

[0075] 如在本文中进一步详述,一旦已经获取样本的表达水平的测量结果,就可以对测量结果应用算法从而计算所述样本的诊断评分。一般来讲,确定样本诊断评分中使用的算法可以包括使用SVM算法、逻辑回归算法、lasso算法、boosting算法、bagging算法、随机森林算法、CART算法或MATT算法。可以应用于对本文中公开的标记物的测量结果的实例特异性算法包括但不限于以下(u-表示尿液标记物,p-表示血浆标记物):

[0076] 公式#1:: $\log_{10} \text{odds} = 1.1459 + 0.1776 * \text{sPSA} - 0.00004505 * \text{uPCA3} - 0.001314 * \text{pHSPD1} + 0.0001012 * \text{pIMPDH2} + 0.0006353 * \text{pPDLIM5} - 0.9314 * \text{pERG}$

[0077]  $\text{odds} = \exp(\log_{10} \text{odds})$

[0078]  $\text{prob} = \text{odds} / (1 + \text{odds})$

[0079] 公式#2:: $\log_{10} \text{odds} = -0.1303 + 0.786 * \text{sPSA} + 0.0000440 * \text{uPCA3} - 0.0013 * \text{pHSPD1} + 0.00000102 * \text{pIMPDH2} + 0.00000072856 * \text{pPDLIM5} - 0.00002379 * \text{pERG}$

[0080]  $\text{odds} = \exp(\log_{10} \text{odds})$

[0081]  $\text{prob} = \text{odds} / (1 + \text{odds})$

[0082] 公式#3: $\log_{10} \text{odds} = 0.1569 + 0.2786 * \text{sPSA} - 0.00004405 * \text{uPCA3} - 0.0001114 * \text{pHSPD1} + 0.0001052 * \text{pIMPDH2} + 0.0000006253 * \text{pPDLIM5} - 0.0009314 * \text{pERG}$

[0083] odds=exp(log\_odds)

[0084] prob=odds/(1+odds)

[0085] 公式#5:log\_odds=1.340e+00+1.999e-01\*sPSA+1.237e-04\*pERG-2.367e-05\*uPDLIM5+1.613e-04\*pUAP1

[0086] odds=exp(log\_odds)

[0087] prob=odds/(1+odds)

[0088] 公式#5:log\_odds=-2.670e+00+2.955e-01\*sPSA-2.288e-04\*pERG-7.885e-05\*uPDLIM5+2.623e-04\*pUAP1

[0089] odds=exp(log\_odds)

[0090] prob=odds/(1+odds)

[0091] 在一些情况下,在确定适当的函数形式之后,以其适当的函数形式的所有表达标记物都可以汇合在逻辑回归方程中。除了测定一致性指数之外,可以检查模型的灵敏度和特异性。对ROC(接受者操作特性)曲线绘图以检查模型的预测能力。ROC曲线只是模型灵敏度对比假阳性率的图形。ROC曲线下的面积(AUC)越大,模型一致性指数越好并且模型以高灵敏度和特异性预测重现性越好。AUC只是ROC曲线下的面积;AUC为1表明完美的预测能力-100%灵敏度并且0%假阳性。AUC为0.5表明所述模型的预测结果准确性就像随机可能性一样。AUC越接近1,模型的预测能力越好。一致性指数是模型区分风险能力的度量,换句话说预测低风险观察结果具有低可能性并且预测所述事件处于高风险的观察结果发生可能性高。灵敏度是实际上稍后复现的检测阳性重现的患者的比例。特异性是实际上没有复现的检测阴性重现的患者的比例。假阳性率是1减去特异性,换句话说,它是检测阳性重现但是实际上没有复现的患者的比例。

[0092] II. 定义

[0093] 如本文使用的“获取生物样本”或“获取血液样本”是指例如直接或间接地接收生物或血液样本。本文使用的生物样本基本上包括无细胞体液,如血浆、血清和尿液。例如,在一些实施方案中,生物样本如血液样本或含有外周血液单核细胞(PBMC)的样本是直接从处在或靠近分析生物样本的实验室或地点的受试者获得的。在其他实施方案中,生物样本可以由第三方抽取或提取并且随后转移到例如独立单位或地点用于分析。在其他实施方案中,可以使用床旁测试技术(point-of care test)在相同地点获取并且测试样本。在这些实施方案中,所述获取是指例如从患者、从实验室、从医生办公室、从邮件、快递员或邮局等接收样本。在一些其他方面,所述方法还可以包括向受试者、保健支付者、主治医师、药剂师、药房权益经理或所述测定可能有关的任何人报告所述测定。

[0094] “受试者”或“患者”表示想要疗法或诊断测试的任何单一受试者。在这种情况下,受试者或患者通常是指人。还预期作为受试者包括的是没有显示任何临床病征的临床研究试验中涉及的任何受试者,或流行病学研究中涉及的受试者,或用作对照的受试者。

[0095] 如本文使用的“表达升高”是指癌症样本中的表达水平相对于合适的对照(例如,非癌性组织或细胞样本,参考标准)提高或升高,其中基因表达水平的提高或升高是统计上显著的( $p < 0.05$ )。癌症样本中的基因表达相对于对照的升高是否是统计上显著的可以使用适当的t检验(例如,单样本t检验、双样本t检验、韦尔奇t检验(Welch's t-test))或本领域技术人员已知的其他统计检验确定。癌症中过表达的基因可以是例如在癌症中过表达的

已知的或已经先前确定的基因。

[0096] 如本文使用的“表达降低”是指癌症样本中的表达水平相对于合适的对照(例如,非癌性组织或细胞样本,参考标准)减少或降低,其中基因表达水平的减少或降低是统计上显著的( $p < 0.05$ )。在一些实施方案中,基因表达的减少或降低水平可以是完全没有基因表达或表达水平为零。癌症样本中的基因表达相对于对照的降低是否是统计上显著的可以使用适当的t检验(例如,单样本t检验、双样本t检验、韦尔奇t检验)或本领域技术人员已知的其他统计检验确定。癌症中低表达的基因可以是例如已知或已经先前确定在癌症中低表达的基因。

[0097] 本文中的术语“抗原结合片段”是广义上使用的并且特定地涵盖完整的单克隆抗体、多克隆抗体、由至少两种完整的抗体形成的多特异性抗体(例如,双特异性抗体)和抗体片段。

[0098] 本文使用的术语“引物”意欲涵盖能够在模板依赖性过程中引发新生核酸合成的任何核酸。引物可以是长度为十至二十和/或三十个碱基对的寡核苷酸,但是可以使用更长的序列。引物可以以双链和/或单链形式提供,不过单链形式是优选的。

### III. 实施例

[0099] 包括以下实施例以示范本发明的优选实施方案。本领域技术人员应理解以下实施例中公开的技术表示由本发明人发现的,在本发明实践中发挥良好作用的技术,并且因此可以被认为构成本发明实践的优选模式。然而,根据本公开,本领域技术人员应理解,在不脱离本发明的精神和范围的情况下可以在已公开并仍获得类似或相似结果的特定实施方案中做出许多改变。

#### [0100] 实施例1-患者和方法

[0101] 患者和样本。尿液和血液样本是从分为三个组的141名男性收集。组1包括61名活检之后前列腺癌阳性的患者。组2包括60名活检之后前列腺癌阴性的患者。组3包括20名最近进行前列腺切除术的患者。为组1和组3中的患者提供按照Gleason评分的肿瘤组织分级。测定并记录每个患者的血清PSA水平。从没有DRE的每个患者收集尿液,立即运送并且次日处理。收集的尿液的体积在30mL至110mL范围内。每个患者提供一个不含防腐剂的具有不同尿量的收集杯并且所有患者提供大约9mL在EDTA中保存的外周血液。所有工作都用具有许可形式的IRB批准的流程(Western IRP)进行并且所有样本都从社区实习泌尿科收集。

[0102] 尿液和血浆处理。从每个患者收集的尿液通过使用具有3kDa膜的Amcion Ultra-15离心过滤机单元离心来浓缩(Millipore, Billerica, Massachusetts)。使用浮桶式转头在4,000x g下离心尿液直到只剩下1mL浓缩尿液。将血浆从外周血液样本分离并且用于提取总核酸。使用NucliSens (BioMerieux, Durham, NC) 提取试剂盒从患者尿液和血浆提取总核酸。

[0103] 定量RT-PCR。使用RNA超敏一步定量RT-PCR系统(Applied Biosystems, Foster City, California)进行定量RT-PCR,所述系统使用ViiA 7实时PCR系统(Applied Biosystems),其中热循环条件如下:50°C持续15min的保持阶段,95°C持续2min,随后95°C持续15秒和60°C持续30秒的45个循环。用于PDLIM5、PCA3、TMPRSS2:ERG和ERG的引物探针组作为分别具有测定号Hs00935062\_m1、Hs01371939\_g1、Hs03063375和Hs01554629\_m1的

TaqMan®基因表达测定(Applied Biosystems)购买。用于UAP1的引物探针组产生70bp的PCR产物:5'-TTGCATTCAGAAAGGAGCAGACT-3'(正向;SEQ ID NO:1);5'-CAACTGGTTCTGTAGGGTTCGTT-3'(反向;SEQ ID NO:2)和5'-VIC®-TGGAGCAAAGGTGGTAGA-小沟结合剂非荧光淬灭团(MGBNNFQ)-3'(探针;SEQ ID NO:3)。用于HSPD1的引物探针组产生64bp的PCR产物:5'-AACCTGTGACCACCCCTGAA-3'(正向;SEQ ID NO:4);5'-TCTTGTCCTCGTTGCAGAAA-3'(反向;SEQ ID NO:5);5'-VIC®ATTGCACAGGTTGCTAC-MGBNFQ-3'(探针;SEQ ID NO:6)。用于IMPDH2的引物探针组被设计成包含外显子10和11并且产生74bp的PCR产物:5'-CCACAGTCATGATGGGCTCTC-3'(正向;SEQ ID NO:7);5'-GGATCCCATCGGAAAAGAAGTA(反向;SEQ ID NO:8);5'-6FAM™-ACCACTGAGGCCCT-MGBNFQ-3'(探针SEQ ID NO:9)。用于PSA的引物探针组产生67bp的PCR产物:5'-CCACTGCATCAGGAACAAAAG-3'(正向;SEQ ID NO:10);5'-TGTGTCTTCAGGATGAAACAGG-3'(反向;SEQ ID NO:11);5'-VIC®-CGTGATCTGCTGGGT-MGBNNFQ(探针;SEQ ID NO:12)。B2M和GAPDH mRNA转录产物作为对照测定并且作为预开发TaqMan®测定试剂(Applied Biosystems)购买。人前列腺癌细胞(CRL-2505)用于提供阳性对照的RNA(ATCC)并且用QIAamp RNA血液微型试剂盒(Qiagen, Hilden, Germany)提取。阴性对照从FirstChoice®人前列腺总RNA(Applied Biosystems)获得。

[0104] 实施例2-结果

[0105] 患者特征。患有活检证实的前列腺癌和BPH的患者具有相似的年龄(中值分别是66对比63)( $p=0.21$ )(表1)。人种分布也是相似的,大多数患者是白种人(表1)。然而,正如所料在两个组的血清PSA之间有显著差异( $p<0.001$ ),其中BPH组的中值为4.4ng/ml,癌症组的中值为5.7ng/ml(表1)。因为对照数据和样本是在20名前列腺癌前列腺切除术之后的患者上收集的。如表1所示,这组患者具有相似的年龄和种族背景,但是PSA同样显著地比BPH和癌症组更低(中值为0.01ng/ml)。癌症患者和前列腺切除术后患者之间的Gleason组织分级相似。Gleason分级是根据基于2005共识会议的新改良的系统完成的。

[0106] 前列腺切除术后以及癌症和BPH患者之间的显著差异。在单变量分析中,在前列腺切除术后患者以及癌症组之间,尿液中的PDLIM5( $p=0.005$ )、UAP1( $p=0.001$ )、PCA3( $p<0.0001$ )、TMPRSS( $p=0.009$ )和血浆中的HSPD( $p=0.01$ )、IMPDH2( $p=0.003$ )、UAP1( $p=0.02$ )和ERG( $p=0.02$ )有显著的( $p<0.05$ )差异。

[0107] 前列腺切除术后和BPH的尿液中的HSPD1( $p=0.004$ )、IMPDH2( $p=0.002$ )、PDLIM5( $p=0.0003$ )、UAP1( $p=0.0003$ )、PCA3( $p<0.0001$ )、TMPRSS和( $p=0.0006$ )以及血浆中的HSPD( $p=0.006$ )、IMPDH2( $p=0.002$ )、UAP1( $p=0.03$ )之间有显著差异。这清楚地显示大部分这些标记物是前列腺特异性的并且这反映在血浆样本以及尿液样本中。

[0108] 使用单变量比较的BPH和前列腺癌之间的边际差异。在单变量分析中,BPH和前列腺癌之间只在尿液中的HSPD1( $p=0.05$ )、IMPDH2( $p=0.01$ )、PDLIM5( $p=0.05$ )和血浆中的Erg( $p=0.0003$ )中有显著差异。

[0109] 除了血浆ERG表达之外,BPH和癌症之间的差异极小,这反映出辨别两种病状中的困难并且很可能是因为患有癌症的大部分患者还患有BPH的事实。

[0110] 多变量分析以及区分BPH与癌症的算法的开发。为了能够使患有前列腺癌的患者与BPH患者区分并且同时利用尽可能多的变量,但是还除去不相关的变量,发明人研究了数

学算法的价值。发明人首先将样本分到学习(训练)组和测试组中,前者包括70名患者(35名癌症患者和35名BPH患者),后者包括51名患者(26名癌症患者和25名BPH患者)。此外,在使用测试51名患者组验证模型之前,还使用了训练组,其中大约三分之二用于建立模型,三分之一用于测试。开发算法中包括的变量是UAP1、PDLIM5、IMPDH2、HSPD1、PCA3、PSA、TMPRSS2、ERG、GAPDH、B2M、年龄和血清PSA。

[0111] 发明人使用了多种数学算法来选择特征并且比较了各种算法之间的平均AUC和平均误差率。所有使用的算法都基于机器学习并且包括逻辑回归、SVM(支持向量机(Support vector machine))、Lasso(最小绝对收缩和选择算法(least absolute shrinkage and selection operator))、boosting、bagging、随机森林、CART(分类和回归树(classification and regression tree))、matt和ctree(条件干涉树(conditional interference tree))。如表2和图1所示,来自所有算法的最好的AUC和最小的误差率是通过逻辑回归获得的。在这个训练组的算法测试中显示出0.77的AUC和0.27的平均误差率。在这个模型中,包括六个变量并且每个变量的贡献显示在图1中。特征消除用于消除不利于改善模型的变量。这个模型中包括的六个变量是血浆ERG、血清PSA、尿液PCA3、尿液MPDH2、尿液PDLIM5和尿液HSPD1。

[0112] 当对测试组应用相同模型时,获得相似的结果(图2)。对于逻辑回归,发明人获得这个组的平均AUC值为0.78。当考虑所有121个样本并且每个组测试100次每次选择随机样本时,发明人获得的AUC在0.70和0.85之间变化。逻辑回归算法显示分界点为0.565(图3)最小误差率为0.25。在这个分界点,特异性和灵敏度分别在88%和67%。

[0113] 在这组患者中单独使用血清PSA并且分界点为4,特异性为62%并且灵敏度为56%。使用sPSA分界点 $>14.1$ ,我们获得100%特异性但是灵敏度为18%。

[0114] 多变量分析和区分侵袭性前列腺癌的算法的开发。已经表明在改进的Gleason评分系统中评分 $<7$ 是进展缓慢癌症并且所述癌症的死亡风险非常小。在前列腺癌Gleason评分 $<6$ 的患者中,无论是否治疗,诊断后10至15年内死亡的风险是相同的(Carter等,JC0,2012年12月10日)。因此,我们集中了Gleason $<7$ 的患者以及患有BPH的患者并且研究了我们的生物标记物在从其余的患者(Gleason $<7$ 和BPH)(89名患者)中预测Gleason $\geq 7$ (32名患者)的前列腺癌患者的潜力。

[0115] 将全部数据集随机分配到训练(69名患者)中,包括18名患有侵袭性癌症的患者和51名患有BPH/Gleason $<7$ 的患者。测试组(52名患者)包括14名患有侵袭性癌症的患者和38名患有BPH/Gleason $<7$ 的患者。

[0116] 数学模型是使用训练组以如上所述的相同方式建立并且比较了AUC和误差率。图4显示每种算法的平均AUC和误差率。此外,基于随机选择之后100次测试,逻辑回归在训练组中以0.87的平均AUC证明是提供信息最多的模型。测试组显示0.88的AUC。当组合并且测试全部样本时,AUC是0.88。在这个模型中,四个变量足以开发这个算法,并且这包括血清PSA、血浆UPA1、血浆ERG和尿液PDCIM5,如图4所示。这些变量每个的贡献在图4C中显示。

[0117] 根据AUC,我们选择0.61作为分界点,它给出0.99的特异性和0.47的灵敏度(表3)。

[0118] 晚期癌症的数目相对较小(32名患者),然而,0.87的AUC值在一个标准偏差之内。基于50个迭代测试,平均值 $\pm 1SD$ 是0.73至0.92。

[0119] 从患有进展缓慢癌症或BPH患者检测患有侵袭性癌症患者的组合模型。上述两种

模型使用不同变量和不同算法,它们完全独立。当使用两个模型评价个体患者时,通过两个模型获得一致的结果很可能表示更强的预测。为了研究这个,发明人使用全部121名患者比较了两个模型之间的结果。在121名患者中,91名(75%)具有一致的结果。在这组患者中,预测侵袭性癌症对比进展缓慢癌症或BPH中的特异性和灵敏度分别是99%和68% (表4,图6)。其余患者(总数的25%)具有不一致的结果,并且由于实际原因应该只在特异性和灵敏度为88%和67%的存在或不存在前列腺癌的预测中考虑,但是不能针对癌症的侵袭性可靠地分类。

[0120] 表1.研究中使用的患者的特征。

	癌症	BPH	前列腺切除术后	P-值
[0121]	年龄[中值(范围)]	66 (45-84)	63 (45-84)	67 (50-77)
	种族	82%W, 5%B, 10%H, 2%A	78%W, 3%B, 17%H, 0%A	85%W, 5%B, 10%H, 0%A
	组织分级	47%(1), 23%(2), 15%(3), 15%(4)		21%(1), 53%(2), 16%(3), 10%(4)
	PSA(ng/ml)	5.7 (1.5-283)	4.4 (0.5-14.1)	0.01 (0-6.0)
				<0.001

[0122] 表2.使用训练组通过用来区分癌症和BPH的各种数学算法获取的AUC和误差率。

[0123]

方法	平均-AUROC	标准-AUROC	平均-误差	标准-误差
逻辑回归	0.773	0.067	0.269	0.01
lasso	0.726	0.072	0.322	0.01
svm	0.672	0.082	0.365	0.012
boosting	0.667	0.084	0.387	0.01
bagging	0.643	0.089	0.392	0.012
随机森林	0.642	0.079	0.397	0.011
cart	0.609	0.081	0.397	0.01
matt	0.586	0.061	0.415	0.008
ctree	0.54	0.049	0.444	0.006

[0124] 表3.用于区分侵袭性前列腺癌与BPH/进展缓慢的前列腺癌平均AUC和标准偏差。

[0125]

方法	平均_AUROC	标准_AUROC
逻辑回归	0.828	0.094
Lasso	0.824	0.094
boosting	0.797	0.093
随机森林	0.738	0.107

Matt	0.725	0.089
[0126]		
Bagging	0.713	0.113
Svm	0.699	0.105
Cart	0.649	0.084
Ctree	0.617	0.128

[0127] 表4.所述三种算法的灵敏度、特异性、阳性预测值 (PPV) 和阴性预测值 (NPV)。

		估计值	95%置信度	
			下限	上限
[0128]	在分界点=0.565 的癌症对比 BPH	灵敏度	0.67	0.54 0.78
		特异性	0.88	0.77 0.95
		PPV	0.85	0.72 0.93
		NPV	0.73	0.61 0.82
	在分界点=0.61 的侵袭性癌症对比 BPH/Gleason <7	灵敏度	0.47	0.30 0.65
		特异性	0.99	0.93 1.00
		PPV	0.94	0.68 1.00
		NPV	0.84	0.75 0.90
	用于预测侵袭性癌症对比 BPH/Gleason <7 的组合模型	灵敏度	0.68	0.45 0.85
		特异性	0.99	0.91 1.00
		PPV	0.94	0.68 1.00
		NPV	0.91	0.81 0.96

[0129] 实施例3-使用额外的标记物的测定

[0130] 材料和方法

[0131] 研究设计和患者

[0132] 从来自四个泌尿科实践的呈现前列腺肿大并且预定前列腺活检的287名男性收集尿液和血液样本。用于活检的肿瘤的组织GS证实PCa是由每个患者的部位提供的。Gleason分级根据基于2005共识会议的新修改的系统进行 (Epstein等, 2006, 以引用方式并入本文。) 活检显示103名(36%)患者患有BPH并且184名(64%)患者患有PCa。107名PCa患者在高风险组 (PCa的58%以及总数的37%) 中。排除接收BPH或PCa的任何疗法的患者并且为了参与所述研究需要患者是新确诊的。尿液样本在没有数字化直肠检查 (DRE) 的情况下收集的并且在收集48小时内处理。所有患者提供9mL在乙二胺四乙酸 (EDTA) 中的外周血液。没有其他选择标准, 样本表示普通患者。所有实验室工作都在IRB批准的流程 (Western IRP) 下进行。

[0133] 尿液和血浆处理

[0134] 通过在浮桶式转头中在4,000x g下使用具有3KDa膜的Amcion Ultra-15离心过滤机单元 (Millipore, Billerica, MA) 离心将来自每个患者的净排尿液浓缩到1ml的体积。使用标准离心将血浆与外周血液分离。使用NucliSENS®提取试剂盒 (BioMerieux, Durham, NC) 从浓缩尿液或血浆中提取总核酸。

[0135] 定量逆转录-聚合酶链式反应 (qRT-PCR)

[0136] 定量逆转录-实时聚合酶链式反应(qRT-PCR)在ViiA<sup>TM</sup> 7实时PCR系统(Applied Biosystems)上使用RNA超敏一步定量RT-PCR系统(Applied Biosystems,Foster City,CA)进行,其中热循环条件如下:50℃持续15min的保持阶段,95℃持续2min,随后95℃持续15秒和60℃持续30秒的45个循环。从First Choice<sup>®</sup>人前列腺总RNA(Applied Biosystems)获得六点连续稀释标准物。PDLIM5、PCA3、TMPRSS2、ERG和PTEN引物和探针作为分别具有测定号Hs00935062\_m1、Hs01371939\_g1、Hs01120965\_m1、Hs01554629\_m1和Hs01920652\_s1的TaqMan基因表达测定(Applied Biosystems)购买。用于UAP1的引物探针组产生70bp的PCR产物:5' - TTG CATT CAG AAAG GAG CAG ACT -3' (正向; SEQ ID NO:1) ; 5' - CAACTGGTTCTGTAGGGTTCGTT-3' (反向; SEQ ID NO:2) 和VIC<sup>®</sup>-TGGAGCAAAGGTGGTAGA-MGBNFQ (探针; SEQ ID NO:3)。用于HSPD1的引物探针组产生64bp的PCR产物:5' - AACCTGTGACCACCCCTGAA-3' (正向; SEQ ID NO:4) ; 5' - TCTTTGTCTCCGTTGCAGAAA-3' (反向; SEQ ID NO:5) ; VIC<sup>®</sup>-ATTGCACAGGTTGCTAC-MGBNFQ (探针; SEQ ID NO:6)。用于IMPDH2的引物探针组被设计成包含外显子10和11并且产生74bp的PCR产物:5' - CCACAGTCATGATGGGCTCTC-3' (正向; SEQ ID NO:7) ; 5' - GGATCCCATCGGAAAAGAAGTA (反向; SEQ ID NO:8) ; 6FAM<sup>TM</sup>-ACCACTGAGGCCCT-MGBNFQ (探针; SEQ ID NO:9)。用于PSA的引物探针组产生67bp的PCR产物:5' - CCACTGCATCAGGAACAAAG-3' (正向; SEQ ID NO:10) ; 5' - TGTGTCTTCAGGATGAAACAGG-3' (反向; SEQ ID NO:11) ; VIC<sup>®</sup>-CGTGATCTGCTGGGT-MGBNNFQ (探针; SEQ ID NO:12)。用于AR的引物探针组被设计成包含外显子6和7并且产生91bp的PCR产物:5' - GGA ATT CCT GTGCATGAAAGC-3' (正向; SEQ ID NO:13) ; 5' - CATTGAAGTTCATCAAAGAATT-3' (反向; SEQ ID NO:14) ; VIC<sup>®</sup>-CTTCAGCATTATTCCAGTG-MGBNFQ (探针; SEQ ID NO:15)。购买用于B2M和GAPDH的预开发TaqMan<sup>®</sup>分析试剂(Applied Biosystems)以便测定其mRNA转录产物作为对照。在所有测定中,使用等量血浆提取RNA,将RNA洗脱到等量冲洗缓冲液中,并且在每个测定中使用等量RNA溶液。类似地,对于尿液来说,从1ml总浓缩尿液中提取RNA,洗脱到等量冲洗缓冲液中,并且在每个测定中使用等量RNA溶液。

### [0137] 结果

[0138] 活检结果显示287名患者中的103名(36%)患有BPH以及184名(64%)患者患有PCa,其中107名(占PCa的58%以及总数的37%)患有高风险PCa。使用训练组开发用于将PCa与BPH区分的算法。对于这个评价使用的标记物是(1)血清PSA蛋白水平;(2)血浆ERG mRNA水平;(3)血浆AR mRNA水平;(4)尿液PCA3mRNA水平;(5)尿液PTEN水平;(6)尿液B2M mRNA水平;(7)血浆B2M mRNA水平;和(8)血浆GAPDH mRNA水平。使用这些标记物可以用0.87的接收工作特性曲线下的面积(AUROC)将PCa与BPH区别开来。在使用0.64的分界点时,这个模型的测试组显示76%的灵敏度和71%的特异性(参见,例如,图7和表5)。

[0139] 表5:在0.64分界点预测PCa的测试组的结果

估计值	95%置信区间		
		下限	
发病率	0.65	0.55	0.74
灵敏度	0.76	0.64	0.86
特异性	0.71	0.52	0.84
对于任何特定测试结果, 其可能性将是:			
[0140] 阳性	0.60	0.49	0.69
阴性	0.40	0.31	0.51
对于任何特定阳性测试结果, 其可能性将是:			
真阳性	0.83	0.70	0.91
假阳性	0.17	0.09	0.30
对于任何特定阴性测试结果, 其可能性将是:			
真阴性	0.62	0.45	0.76
假阴性	0.38	0.24	0.55

[0141] 开发了其他算法来预测患有高风险PCa (GS $\geq 7$ ) 对比GS<7癌症或BPH的患者。对于这个评价使用的标记物是(1) 血清PSA蛋白水平; (2) 年龄; (3) 尿液PSA mRNA水平; (4) 血浆ERG mRNA水平; (5) 尿液GAPDH mRNA水平; (6) 尿液B2M mRNA水平; (7) 尿液PTEN mRNA水平; (8) 尿液PCA3mRNA水平; 和 (9) 尿液PDLIM5mRNA水平。用这些标记物, 高风险PCa可以用0.80的AUROC与低等级癌症(GS<7) 或BPH区别开来(参见, 例如, 图8和表6)。在一些其他计算中, 使用额外的三种标记物((10) PCA3mRNA水平; (11) 血浆B2M mRNA水平和(12) 血浆HSPD1mRNA水平), 获得了0.8487的AUROC。

[0142] 表6:在0.27分界点预测高风险PCa的测试组的结果

估计值	95%置信区间		
		下限	
发病率	0.35	0.26	0.44
灵敏度	0.44	0.28	0.60

特异性	0.76	0.64	0.85
对于任何特定测试结果, 其可能性将是:			
阳性	0.31	0.23	0.40
阴性	0.69	0.60	0.77
对于任何特定阳性测试结果, 其可能性将是:			
真阳性	0.49	0.32	0.66
假阳性	0.51	0.34	0.68
对于任何特定阴性测试结果, 其可能性将是:			
真阴性	0.72	0.60	0.81
假阴性	0.28	0.19	0.40

[0144] [0145] 进一步分析显示在两种分析之间具有一致的结果的患者对于患有高等级侵袭性PCa显示89%的特异性和59%的灵敏度(表7),对于患有PCa并且没有BPH的患者显示94%的特异性和81%的灵敏度(表8),但是容忍低风险PCa的非检测。因此,组合两种分析并且接受如果两个之一对于癌症是阳性的PCa的诊断,不管侵袭性,分别显示82%和92%的特异性和灵敏度(表9),具有遗漏低风险癌症的可能性(PPV=86%并且NPV=90%)。除了sPSA之外,在两种算法中贡献最大的生物标记物是血浆和尿液ERG、PTEN、AR和PCA3mRNA,其次是血浆和尿液中的PDLIM5和PSA mRNA。

[0146] 表7:用于检测高等级侵袭性PCa的组合分析(两种分析是阳性或阴性:287名中的184名,64%)

估计值	95%置信区间		
		下限	上限
发病率	0.35	0.28	0.42
灵敏度	0.59	0.46	0.71
特异性	0.89	0.82	0.94
对于任何特定测试结果, 其可能性将是:			
阳性	0.28	0.22	0.35

[0148]

阴性	0.72	0.65	0.78
对于任何特定阳性测试结果, 其可能性将是:			
真阳性	0.75	0.60	0.85
假阳性	0.25	0.15	0.40
对于任何特定阴性测试结果, 其可能性将是:			
真阴性	0.80	0.72	0.87
假阴性	0.20	0.13	0.28

[0149] 表8:检测PCa和非BPH的一致结果,接受癌症,如果容许GS<7,如果遗漏或检测(287名中的184名,64%)。

[0150]

估计值	95%置信区间		
		下限	上限
发病率	0.38	0.31	0.46
灵敏度	0.81	0.70	0.89
特异性	0.94	0.87	0.97
对于任何特定测试结果, 其可能性将是:			
阳性	0.35	0.28	0.42
阴性	0.65	0.58	0.72
对于任何特定阳性测试结果, 其可能性将是:			
真阳性	0.89	0.78	0.95
假阳性	0.11	0.05	0.22
对于任何特定阴性测试结果, 其可能性将是:			
真阴性	0.89	0.82	0.94
假阴性	0.11	0.06	0.18

[0151] 表9:结果,如果任一分析是PCa阳性(不考虑侵袭性),并且假设容许GS<7,如果确定为阴性。

[0152]

估计值	95%置信区间		
		下限	上限
发病率	0.54	0.48	0.60
灵敏度	0.92	0.86	0.95
特异性	0.82	0.74	0.88

[0153]	对于任何特定测试结果, 其可能性将是:			
	阳性	0.58	0.52	0.64
	阴性	0.42	0.36	0.48
	对于任何特定阳性测试结果, 其可能性将是:			
[0154]	真阳性	0.86	0.79	0.90
	假阳性	0.14	0.10	0.21
[0155]	对于任何特定阴性测试结果, 其可能性将是:			
	真阴性	0.89	0.82	0.94
	假阴性	0.11	0.06	0.18
	因此, 通过组合上文描述的两种分析的结果(即, 用于将PCa与BPH区分的标记物的测定以及用于将高风险PCa与低风险PCa(GS<7)或BPH区分的标记物的分析), 可以实现高特异性和灵敏的诊断。用这里详述的研究获得的特异性诊断结果表明:			

[0155] 1) 两种分析都是阴性:

[0156] -无任何前列腺癌的证据(灵敏度=59%, 特异性=89%)

[0157] -无高风险侵袭性(Gleason $\geq 7$ )的证据, 但是不能完全排除低等级癌症(Gleason<7) (灵敏度=81%, 特异性=94%)

[0158] 2) 两种分析都是阳性:

[0159] -患有侵袭性癌症(Gleason $\geq 7$ )的高可能性(灵敏度=59%, 特异性=89%)

[0160] -患有任何前列腺癌(任何等级)的高可能性(灵敏度=81%, 特异性=94%)

[0161] 3) PCa阳性和高等级阴性:

[0162] -患有任何癌症的高可能性(灵敏度=92%, 特异性=82), 但是大不可能是高等级的(特异性=76%, 灵敏度=44%)

[0163] 4) PCa阴性和高等级阳性

[0164] -患有任何癌症的高可能性(灵敏度=92%, 特异性=82), 但是可能是高等级的(特异性=76%, 灵敏度=44%)

[0165] \*\*\*

[0166] 根据本公开, 本文公开的并且要求保护的所有方法可在无需过度实验的情况下进行和实施。尽管本发明的组合物和方法已经根据优选实施方案加以描述, 但对本领域技术人员显而易见的是可使本文所述的方法和本文所述方法的步骤或步骤的顺序发生变化, 而不偏离本发明的概念、精神和范围。更具体说来, 显而易见的是在化学上和生理学上相关的某些试剂可取代本文所述的试剂, 同时达到相同或相似结果。对本领域技术人员来说显而易见的是所有此类相似的替代和修改被认为在如由随附权利要求限定的本发明的精神、范围和概念内。

[0167] 参考文献

[0168] 在以下参考文献对本文阐述的那些内容提供示例性程序或其他细节补充的程度上, 将所述参考文件明确地以引用方式并入本文中。

[0169] Ausubel et al., Current protocols in molecular biology, John Wiley&Sons Ltd, Wiley Interscience, 2003.

- [0170] Carter et al., J.Clin.Oncol., 30:4294-4296, 2012.
- [0171] Epstein et al., "Update on the Gleason grading system for prostate cancer: results of an international consensus conference of urologic pathologists," Adv.Anat.Puthol., 13(1):57-9, 2006.
- [0172] Sambrook et al., Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, 1989.

## 序列表

<110> M·阿尔伯塔 (ALBITAR, Maher)	
<120> 用于检测和测定前列腺癌预后的组合物和方法	
<130> NGNL.P0002W0	
<140> 未知	
<141> 2014-03-13	
<150> 61/785, 375	
<151> 2013-03-14	
<160> 15	
<170> PatentIn version 3.5	
<210> 1	
<211> 23	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 1	
ttgcattcag aaaggaggcag act	23
<210> 2	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
[0001]	
<400> 2	
caactgggttc tgttagggttc gttt	24
<210> 3	
<211> 18	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 3	
tggagcaaag gtggtaga	18
<210> 4	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 4	
aacctgtgac caccctgaa	20
<210> 5	
<211> 22	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 5	
tctttgtctc cgtttgaga aa	22

<210> 6		
<211> 17		
<212> DNA		
<213> 人工序列		
<220>		
<223> 合成引物		
<400> 6		
attgcacagg ttgctac		17
<210> 7		
<211> 21		
<212> DNA		
<213> 人工序列		
<220>		
<223> 合成引物		
<400> 7		
ccacagtcat gatgggctct c		21
<210> 8		
<211> 22		
<212> DNA		
<213> 人工序列		
<220>		
<223> 合成引物		
<400> 8		
ggatcccatc ggaaaagaag ta		22
[0002] <210> 9		
<211> 15		
<212> DNA		
<213> 人工序列		
<220>		
<223> 合成引物		
<400> 9		
accactgagg cccct		15
<210> 10		
<211> 21		
<212> DNA		
<213> 人工序列		
<220>		
<223> 合成引物		
<400> 10		
ccactgcata aggaacaaaa g		21
<210> 11		
<211> 22		
<212> DNA		
<213> 人工序列		
<220>		
<223> 合成引物		
<400> 11		
tgtgtttca ggatgaaaca gg		22
<210> 12		
<211> 16		

---

<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 12	
cgtgatcttg ctgggt	16
<210> 13	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 13	
ggaattcctg tgcatgaaag c	21
[0003]	
<210> 14	
<211> 23	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 14	
cattcgaagt tcatcaaaga att	23
<210> 15	
<211> 19	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成引物	
<400> 15	
cttcagcatt attccagtg	19

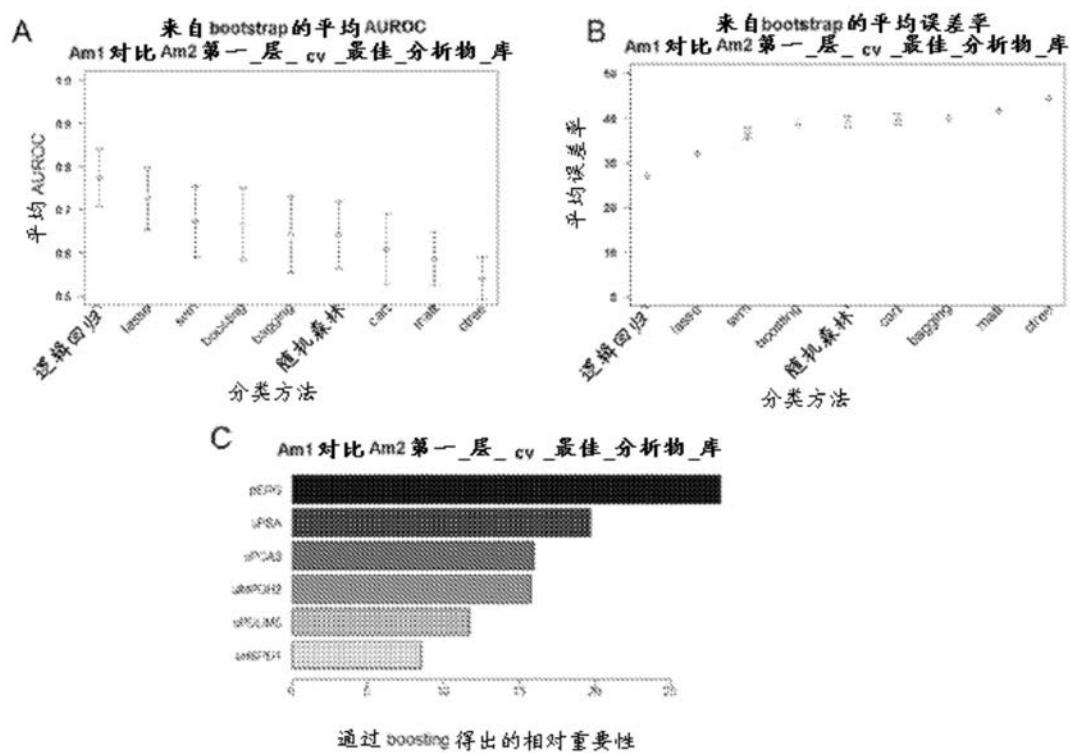


图1

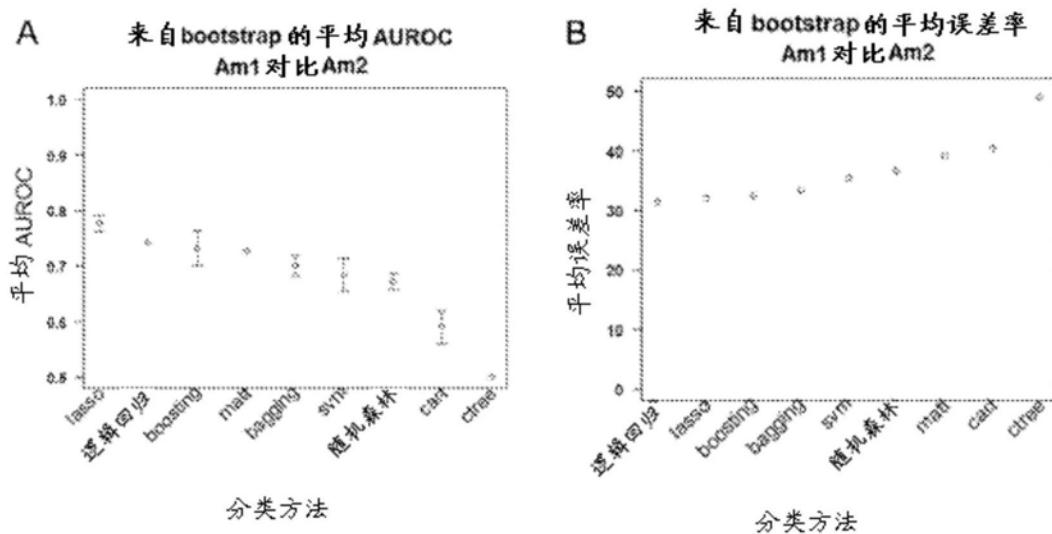


图2

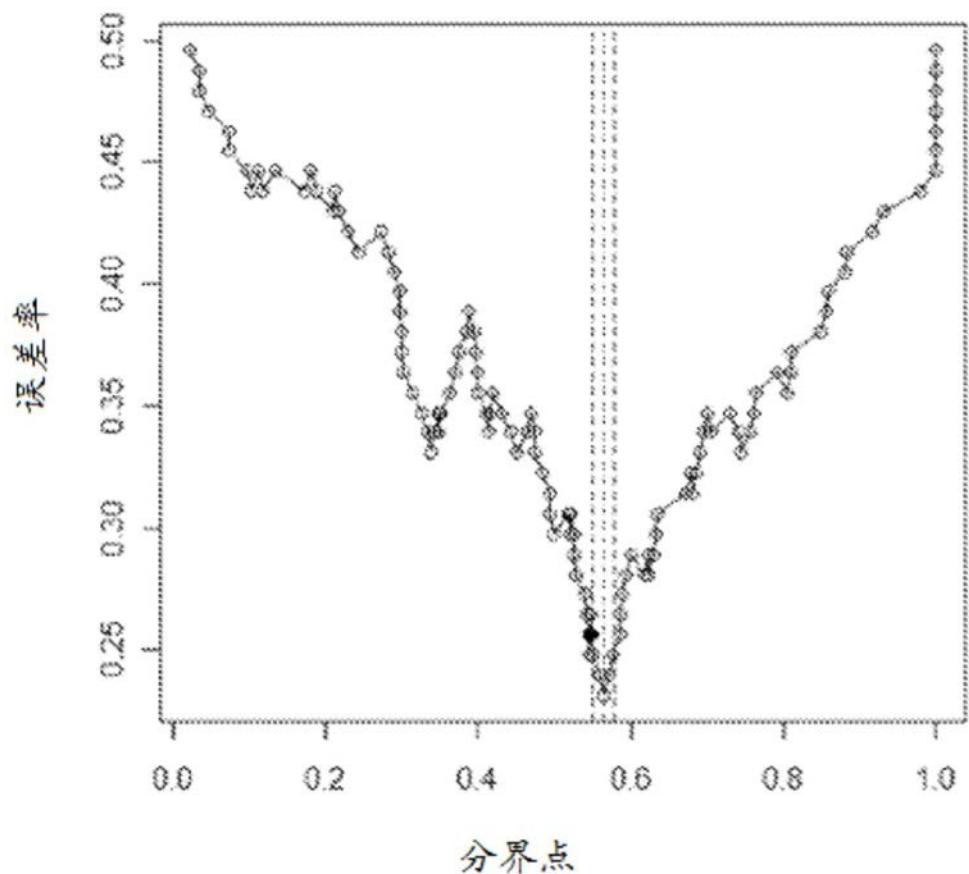


图3

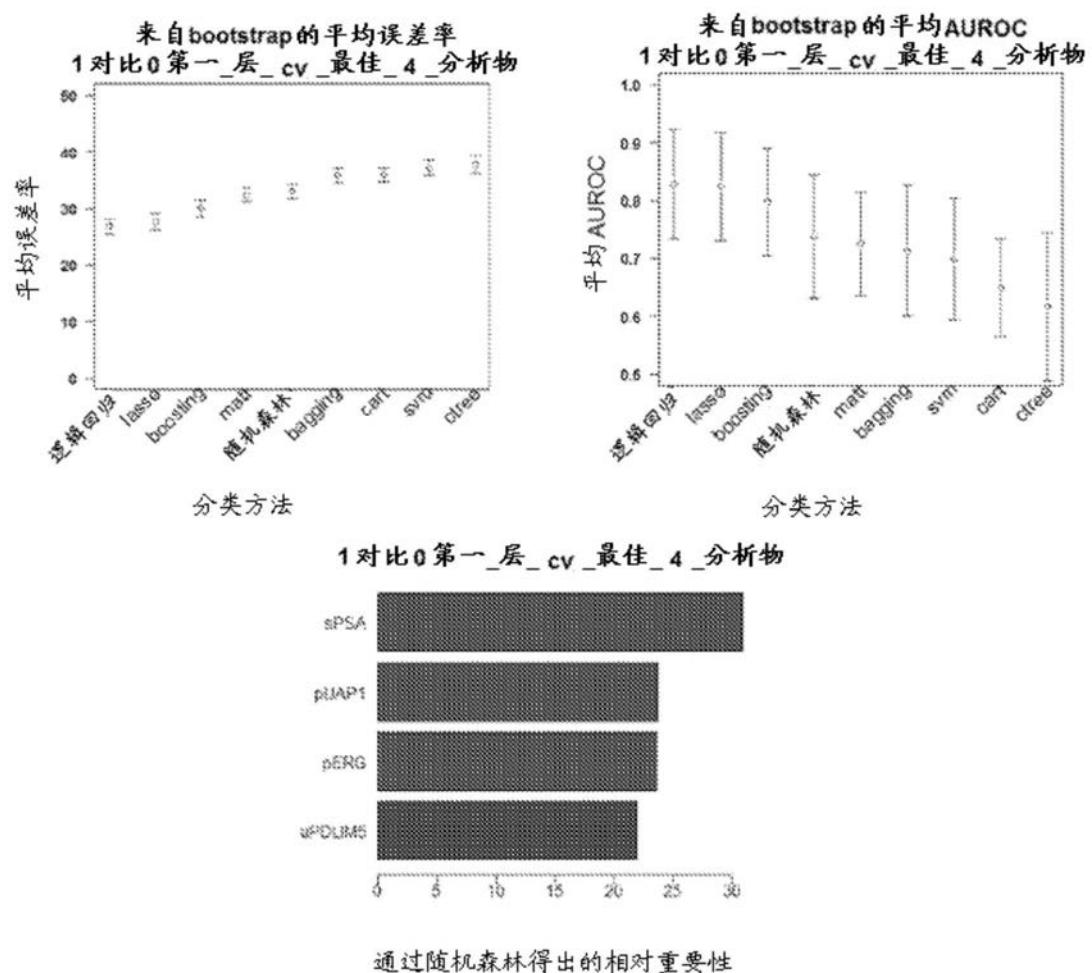


图4

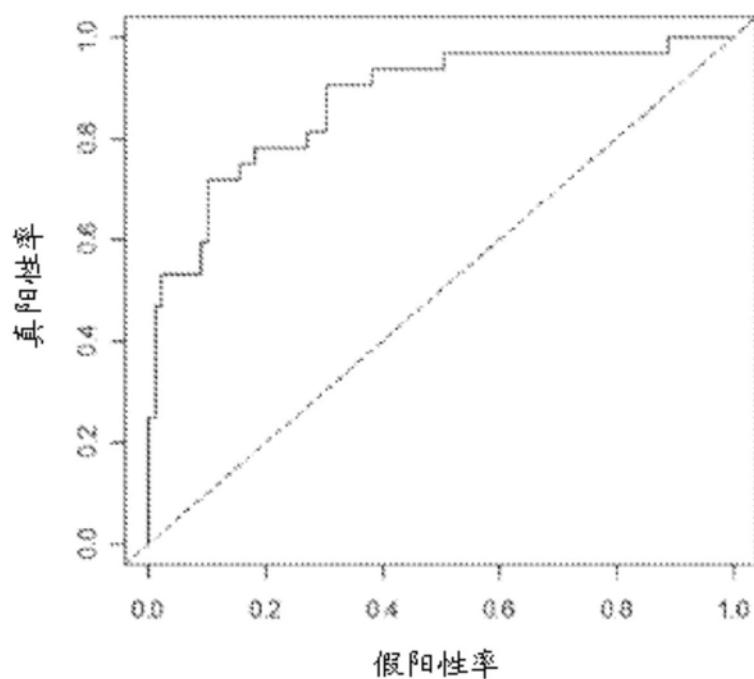


图5

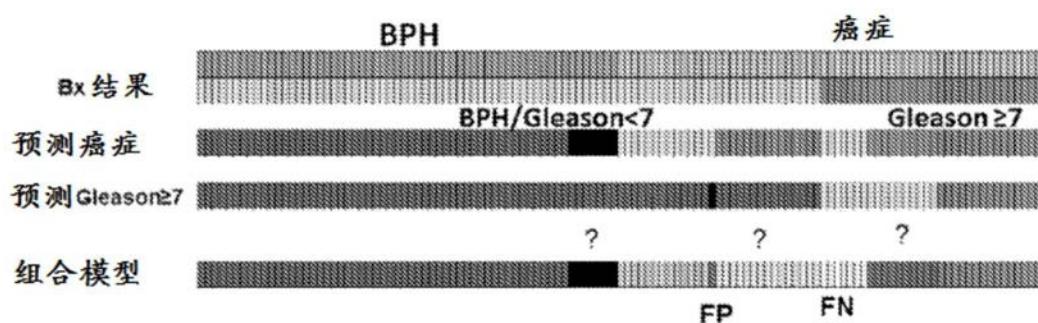


图6

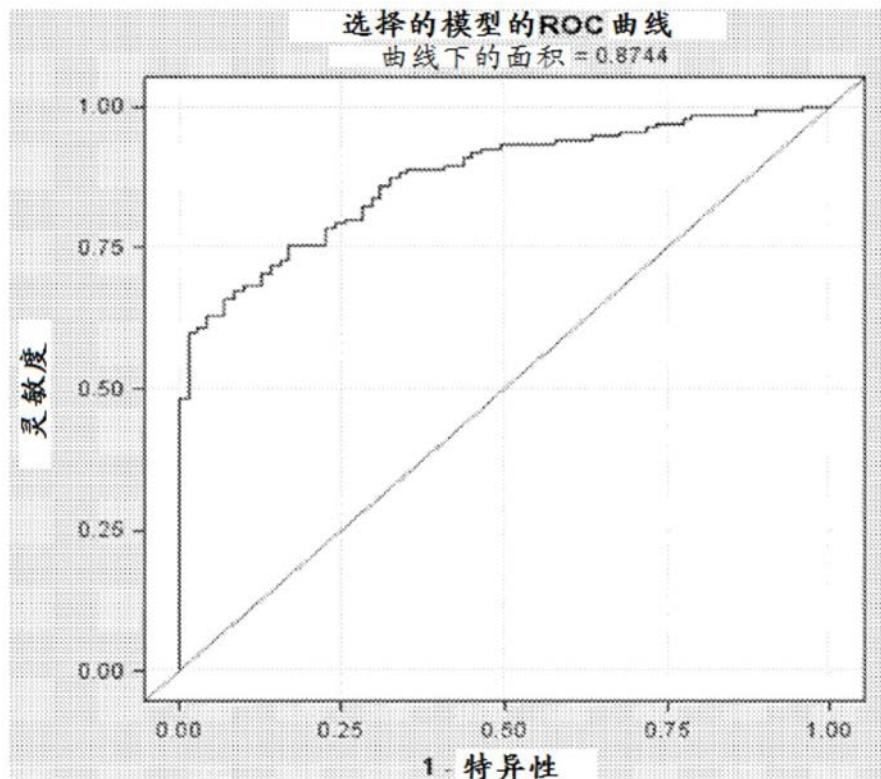


图7

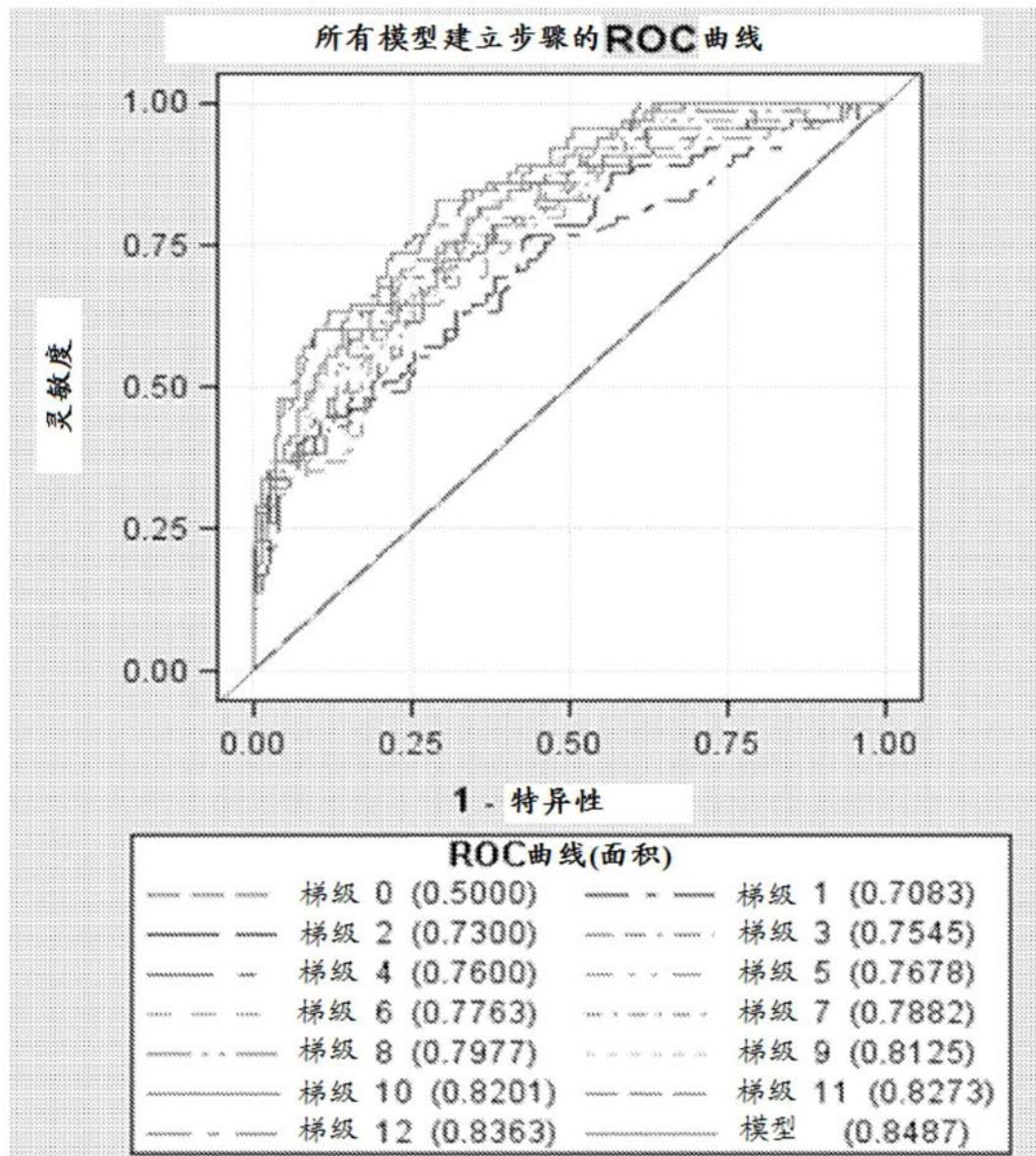


图8