



## (12) 发明专利申请

(10) 申请公布号 CN 101971026 A

(43) 申请公布日 2011.02.09

(21) 申请号 200880125970.7

(51) Int. Cl.

(22) 申请日 2008.12.15

*G01N 33/53* (2006.01)

(30) 优先权数据

*G01N 33/574* (2006.01)

07301680.0 2007.12.14 EP

(85) PCT申请进入国家阶段日

2010.07.30

(86) PCT申请的申请数据

PCT/IB2008/003836 2008.12.15

(87) PCT申请的公布数据

W02009/077864 EN 2009.06.25

(71) 申请人 特兰斯麦迪公司

地址 法国旺德夫尔莱南锡

(72) 发明人 伯纳德·毕海因 维尔日妮·欧希尔

玛丽·布鲁利亚德

(74) 专利代理机构 中原信达知识产权代理有限

责任公司 11219

代理人 张颖 樊卫民

权利要求书 2 页 说明书 29 页 附图 14 页

(54) 发明名称

检测 TIABS 的组合物和方法

(57) 摘要

本发明涉及用于评估对象的生理状态的新方法和产品。更具体来说,本发明涉及通过在来自对象的样品中测定针对特定异常蛋白结构域的抗体的水平、或带有对这样的异常蛋白结构域特异性的 TCR 的免疫细胞的存在或数量,来评估对象中癌症的存在、风险或阶段的方法。本发明也适用于评估对象对治疗的响应性,以及筛选候选药物和设计新疗法。本发明可用于任何哺乳动物对象,特别是用于人类对象。

1. 用于在对象中检测癌症的存在、风险或发展阶段的方法,所述方法包括将来自对象的样品在体外与包含由转录失真性产生的异常蛋白结构域的序列的多肽接触,其中所述多肽与所述样品中存在的抗体 (TIAB) 或带有 TCR 的细胞之间的复合物的形成是癌症的存在、风险或发展阶段的指示。

2. 权利要求 1 的方法,其中所述样品中所述 TIAB 或免疫细胞的水平与参照值进行比较,并且其中与参照值相比的变化是癌症的存在、风险或发展阶段的指示。

3. 权利要求 2 的方法,其中所述与参照值相比的变化是 10%、20%、30%、40%、50% 或以上。

4. 权利要求 1 至 3 任一项的方法,其中多肽包含选自 SEQ ID NOs :1 至 3334 的序列或其含有表位的片段。

5. 权利要求 4 的方法,其中多肽包含表 5 的 45 种 PAP 多肽中任一种的序列。

6. 权利要求 4 或 5 的方法,其中在所述样品中 TIAB 或免疫细胞的水平与参照水平相比增加,是癌症的存在、风险或发展阶段的指示。

7. 权利要求 1 至 6 任一项的方法,其中癌症是实体肿瘤或液体肿瘤,优选为结肠癌、肺癌、乳腺癌、卵巢癌、子宫癌、头颈部癌或黑素瘤。

8. 权利要求 1 至 7 任一项的方法,其中癌症是肺癌,并且多肽包含选自 SEQ ID NO 1、2、4、6、7、1413、581、1457、1769、1270、1276、1396、80、1810、1864、847、2119 或 1451 的序列或其含有表位的片段,特别是包含选自表 5 中显示的 PAP 1、2、4、6、7、24、25、28、29、44、48、66、70、68、69、74、94、82 的序列。

9. 权利要求 1 至 8 任一项的方法,其中接触步骤包括将样品同时与几种、优选为 2 到 10 种含有由转录失真性产生的独特异常蛋白结构域的序列的多肽接触。

10. 评估患者的生理状态的方法,所述方法包括在来自对象的样品中测量对由转录失真性产生的异常蛋白结构域特异性的抗体 (TIAB)、或与这种结构域结合的带有 TCR 的免疫细胞的存在或水平的步骤,其中所述样品中所述 TIAB 或免疫细胞与参照值相比的变化水平,是生理失调的指示。

11. 确定癌症治疗功效的方法,所述方法包括 (i) 在来自对象的样品中,测定至少一种包含由转录失真性产生的异常蛋白结构域的序列的多肽的水平、或 TIAB 或相应的带有 TCR 的细胞的水平,以及 (ii) 将所述水平与治疗之前或治疗较早期阶段从所述对象获取的样品中的水平进行比较。

12. 监测对象中癌症的进展或扩散的方法,所述方法包括 (i) 将从所述对象获得的样品与至少一种包含由转录失真性产生的异常蛋白结构域的序列的多肽接触, (ii) 测定所述样品中 TIAB 或相应的带有 TCR 的细胞的水平,以及 (iii) 将所述水平与参照值进行比较,变化是癌症的进展或扩散的指示。

13. 权利要求 11 至 12 任一项的方法,其中多肽包含选自 SEQ IDNOs :1 至 3334 的序列或其含有表位的片段。

14. 权利要求 12 的方法,其中多肽包含表 5 的 45 种 PAP 多肽中任一种的序列。

15. 权利要求 12 或 13 的方法,其中接触步骤包括将样品同时与几种、优选为 2 至 10 种所述多肽接触。

16. 确定个体是否正在产生含有选自 SEQ ID NOs :1-3334 的序列的多肽的方法,所述

方法包括将从所述个体获得的样品与指示所述多肽存在的药剂接触,以及确定所述药剂是否与所述样品结合。

17. 权利要求 16 的方法,其中所述药剂是与抗体结合的多肽,所述抗体与含有选自 SEQ ID N0s :1-3334 的序列的多肽特异性结合。

18. 权利要求 16 的方法,其中所述药剂是与免疫细胞结合的多肽,所述免疫细胞包含对含有选自 SEQ ID N0s :1-3334 的序列的多肽特异性的 TCR。

19. 权利要求 16 的方法,其中所述药剂是与含有选自 SEQ ID N0s :1-3334 的序列的多肽特异性结合的抗体或其部分。

20. 权利要求 16 的方法,其中所述药剂是免疫细胞,所述免疫细胞包含对含有选自 SEQ ID N0s :1-3334 的序列的多肽特异性的 TCR。

21. 权利要求 1 至 18 任一项的方法,其中多肽被固定化在支持物上。

22. 选择、表征、筛选或优化生物活性化合物的方法,所述方法包括将测试化合物在体外与基因接触,以及确定所述测试化合物调节从所述基因生产含有转录失真性缺口的 RNA 分子的能力。

23. 生产对转录失真性特异性的肽的方法,所述方法包括:

- a) 鉴定由转录失真性缺口产生的蛋白结构域;
- b) 合成含有 a) 的所述蛋白结构域的序列的肽;以及
- c) 任选在来自哺乳动物对象的生物学样品中证实肽与抗体结合。

24. 含有选自 SEQ ID N0s :1-3334 的序列或其含有表位的片段的多肽。

25. 分离的多核苷酸,其编码权利要求 24 的多肽。

26. 权利要求 25 的分离的多核苷酸,其中所述多核苷酸被标记。

27. 分离的核酸,其包含编码选自 SEQ ID N0s :1-3334 的多肽的第一个核苷酸序列或与其互补的序列,以及长度为 100 个或以下核苷酸的第二个核苷酸序列,其中所述第二个核苷酸序列在天然存在的核酸中与所述第一个核苷酸序列相邻。

28. 克隆或表达载体,其包含权利要求 25 或 27 的多核苷酸。

29. 细胞,其用权利要求 28 的载体转化或转染。

30. 分离的抗体或抗体的部分,其与选自 SEQ ID N0s :1-3334 的多肽特异性结合。

31. 分离的细胞,其与选自 SEQ ID N0s :1-3334 的多肽特异性结合。

32. 权利要求 31 的分离的细胞,其中所述细胞是包含对选自 SEQ ID N0s :1-3334 的所述多肽特异性的 TCR 的免疫细胞。

33. 固相支持物,其包含至少一种含有选自 SEQ ID N0s :1 至 3334 的序列或其含有表位的片段的多肽。

34. 固相支持物,其包含至少一种与选自 SEQ ID N0s :1-3334 的多肽特异性结合的抗体或抗体部分。

35. 装置或产品,其包含固定化在支持物上的至少一种含有由转录失真性缺口产生的异常蛋白结构域的序列的多肽。

36. 权利要求 35 的装置,其包含权利要求 23 的一种或几种多肽。

37. 试剂盒,其至少包含权利要求 35 或 36 的装置或产品,以及进行免疫反应的试剂。

## 检测 TIABS 的组合物和方法

### 技术领域

[0001] 本发明涉及用于评估对象的生理状态的新方法和产品。更具体来说,本发明涉及通过测定来自对象的样品中针对特定的异常蛋白结构域的抗体的水平、或带有对这样的异常蛋白结构域特异性的 TCR 的免疫细胞的存在或数量,来评估对象中癌症的存在、风险或阶段的方法。本发明也适用于评估对象对治疗的响应性,以及筛选候选药物和设计新疗法。本发明可用于任何哺乳动物对象,特别是用于人类对象。

### 背景技术

[0002] 在西方国家,癌症正逐渐变成死亡的主要原因,强有力的治疗益处、即患者的无病长期存活率 > 90% 主要在早期诊断的个体中获得<sup>1</sup>。癌症是癌基因和肿瘤抑制基因中突变积累的遗传疾病<sup>2</sup>。遗传检验对于鉴定个体的结肠、肺、乳腺、卵巢和神经内分泌癌的风险来说是有用的<sup>3,4</sup>。然而,因为对于给定的个体发生癌症的时间缺乏精确性,有遗传风险的患者的临床管理是复杂的<sup>5</sup>。

[0003] 癌细胞中的基因表达变化广泛,分析转录模式的差异导致理清了与良好或不良预后相关的分子标签 (molecular signature)<sup>6</sup>。这样的标签可以指导和优化治疗策略,但先决条件仍然是肿瘤的先期鉴定。

[0004] 通过三个主要方向进行了大量努力,以鉴定在易获得的人类体液中可检测的可靠的早期癌症分子标志物。首先,使用各种不同的分离和鉴定程序,分析在正常和癌症患者之间蛋白浓度和 / 或同工型的变化<sup>7,8</sup>。某些这样的方法适合于系统性筛选,但是该技术目前面临着丰富的血浆蛋白浓度 (>>> mg/ml) 与低水平组织泄漏所释放的蛋白浓度 (<< pg/ml) 的巨大差异。备选的方法由鉴定在癌症中差异表达的几个特定靶的表达情况,以及开发灵敏的分析方法以监测它们在血浆中的浓度所构成<sup>9</sup>。这两种策略都已成功地实施,但是均还没有提供强得足以用于系统性临床筛选的标志物<sup>10,11</sup>。研究的第二个中心是基于血浆中循环 DNA 的鉴定和表征。早期报道表明,血清中循环 DNA 的简单存在是诊断性的<sup>12</sup>;但是,这一点现在有疑问,因为健康个体的循环游离 DNA 浓度与癌症患者在同样的范围内<sup>13,14</sup>。维持正常的 DNA 甲基化模式对于适合的细胞功能来说是关键的,它的丧失是癌症发生过程中最早的分子变化之一<sup>15,16</sup>。几个团队已报道了血清中肿瘤相关甲基化模式的检测,但是在使用同样生物标志物和技术不同队伍之间成功率变动极大<sup>17-20</sup>。这是由于肿瘤 DNA 甲基化模式的多样性和肿瘤 DNA 的低丰度两方面最多代表正常体细胞单倍体基因组的 0.12%<sup>21,22</sup>。因此,即使是考虑到最近测序技术的进展<sup>23,24</sup>,但以少量循环癌症 DNA 检测癌症体细胞突变也是太接近于背景水平而不能提供强有力的评价。第三个中心由通过系统性搜索自身抗体来探测免疫系统对癌症的应答而构成<sup>25,26</sup>。这些抗体的存在已被确立,但是自体分子变成免疫原性的过程尚未被了解<sup>27</sup>。筛选从癌细胞 mRNA 构建的表达文库,导致鉴定到大量低灵敏度的抗体,当使用 > 20 种的组合时,在前列腺癌患者中获得了高达 82% 的灵敏度<sup>28,29</sup>。一种备选方法依赖于通过 2D 凝胶电泳的免疫印迹来鉴定自身抗体标签<sup>30</sup>。这导致随后鉴定到主要通过自身抗体鉴定到的不依赖于癌症状态的蛋白,以及有限数量的优先

与癌症血清反应的蛋白<sup>30</sup>。

[0005] 我们的策略是基于不同的基本原理,该原理直接源自于通过大规模癌症 DNA 测序项目获得的结果<sup>31,32</sup>。该重要工作产生的结论是,癌症体细胞突变的发生率比预期的高,但仍然是罕见的事件:估计的比率是每 10<sup>6</sup> 个碱基中 3.1 个,导致在给定肿瘤中平均 90 个氨基酸取代<sup>31</sup>。事实上,所有生物化学、生物学和临床属性在同样组织学亚型的癌症中是非均质的<sup>33</sup>。因此,我们搜寻了造成癌细胞异质性的其它机制。最近,我们指出,癌细胞 mRNA 序列与正常细胞相比,含有更多的碱基取代<sup>34</sup>。发生癌症 mRNA 碱基取代的位点比带有体细胞突变的位点被遇到的几率高 10<sup>4</sup> 倍,并且不与单核苷酸多态性 (SNP) 相符。因此,从同一患者的正常和癌细胞分离到的 mRNA 异质性的差异,不能用基因组水平上出现的差异来解释。癌症 mRNA 中的碱基取代,由对应于活性 RNA 聚合酶 II (PoI II) 所解链部分的 DNA 邻近序列的组成所决定<sup>34,35</sup>。最常见情况是,被取代碱基与紧邻着事件的之前或之后碱基一致。体外数据证明了 PoI II 在特定 DNA 邻近序列中的向前滑动<sup>36,37</sup>,因此,我们提出,转录失真性 (TI) 解释了一部分癌症 mRNA 不是基因组 DNA 的忠实拷贝。

[0006] 我们将这种分析扩展到整个基因组和所有可获得的人类转录本,证实了在癌症中 mRNA 碱基取代显著增加 (2.5 倍)。最重要的是,我们发现在癌症 mRNA 中,单个碱基缺失非常急剧地增加 (38 倍)。mRNA 中的缺口导致下游基因组信息的损失,并能够产生可能引发免疫应答的异常蛋白。我们已经在癌症患者中寻找并发现了针对从含有单个碱基缺口的癌症 mRNA 翻译的预测异常肽 (PAP) 的特异性 IgG。低丰度的多样化 IgG 的检测,为最常见形式的人类实体肿瘤的诊断提供了新方法。一组 IgG 有效地将患有非小细胞肺癌 (NSCLC) 的患者与未患癌症的对象区分开来。

[0007] 因此,本发明显示了这样的缺口 (和插入) 在癌症患者中急剧增加并产生了异常但是可以预测的免疫原性蛋白,这些蛋白代表了非常有效的生物标志物。

## 发明内容

[0008] 本发明的目的涉及一种用于在对象中检测癌症的存在、风险或发展阶段的方法,所述方法包括将来自对象的样品在体外与包含由转录失真性产生的异常蛋白结构域的序列的多肽接触,其中所述多肽与所述样品中存在的抗体 (TIAB) 或带有 TCR 的细胞之间形成复合物,是癌症的存在、风险或发展阶段的指示。

[0009] 本发明的另一个目的涉及评估对象的生理状态的方法,所述方法包括测定来自对象的样品中特异性针对由转录失真性产生的异常蛋白结构域的抗体 (TIAB) 或与这种结构域结合的带有 TCR 的免疫细胞的存在或水平的步骤,其中所述样品中所述 TIAB 或免疫细胞与参照值相比的水平变化,是生理失调的指示。

[0010] 本发明的另一个目的是确定癌症治疗的功效的方法,所述方法包括 (i) 在来自对象的样品中,测定至少一种包含由转录失真性产生的异常蛋白结构域的序列的多肽的水平,或 TIAB 或相应的带有 TCR 的细胞的水平,以及 (ii) 将所述水平与治疗之前或治疗较早期阶段从所述对象获取的样品中的所述水平进行比较。

[0011] 本发明的另一个目的是监测对象中癌症的进展或扩散的方法,所述方法包括 (i) 将从所述对象获得的样品与至少一种包含由转录失真性产生的异常蛋白结构域的序列的多肽接触, (ii) 测定所述样品中 TIAB 或相应的带有 TCR 的细胞的水平,以及 (iii) 将所述

水平与参照水平进行比较。参照值可以是从未患癌症或疾病的个体测定的平均值或中值、从对照患者获得的参照水平、从癌症发作之前的患者或使用对照多肽获得的参照水平。

[0012] 本发明的另一个目的涉及确定个体是否正在产生含有选自 SEQ ID N0s :1-3334 的序列的多肽的方法,所述方法包括将从所述个体获得的样品与指示所述多肽存在的药剂相接触,以及确定所述药剂是否与所述样品结合。

[0013] 本发明的另一个目的是选择、表征、筛选或优化生物活性化合物的方法,所述方法包括将测试化合物在体外与基因接触,以及确定所述测试化合物调节从所述基因生产含有转录失真性缺口和插入的 RNA 分子的能力。

[0014] 本发明的另一个目的在于生产对转录失真性特异的肽的方法,所述方法包括:

[0015] a) 鉴定由转录失真性缺口或插入产生的蛋白结构域;

[0016] b) 合成含有 a) 的所述蛋白结构域的序列的肽;以及

[0017] c) 任选在来自哺乳动物对象的生物学样品中证实该肽与抗体结合。

[0018] 本发明还涉及任何含有由缺口或插入转录失真性产生的异常蛋白结构域的序列或其含有表位的片段的多肽,特别是含有选自 SEQ ID N0s :1 到 3334 的序列或其含有表位的片段的多肽。

[0019] 本发明的另一个目的是分离的核酸,所述核酸编码上面描述的多肽、或包含编码选自 SEQ ID N0s :1-3334 的多肽的第一个核苷酸序列或与其互补的序列以及长度为 100 个或以下核苷酸的第二个核苷酸序列,其中所述第二个核苷酸序列在天然存在的核酸中与所述第一个核苷酸序列相邻。

[0020] 本发明的另一个目的是包含上述多核苷酸的克隆或表达载体,以及用该载体转化或转染的宿主细胞。

[0021] 本发明的另一个目的是分离的抗体或抗体的部分,它们与含有由缺口或插入转录失真性产生的异常蛋白结构域的序列的任何多肽、特别是与含有选自 SEQ ID N0s :1-3334 的序列或其含有表位的片段特异性的多肽结合。

[0022] 本发明的另一个目的是含有 TCR 的免疫细胞,所述 TCR 特异性针对含有由缺口或插入转录失真性产生的异常蛋白结构域的序列的任何多肽,特别是特异性针对含有选自 SEQ ID N0s :1-3334 的序列的多肽。

[0023] 本发明还涉及固相支持物,它包含至少一种含有由缺口或插入转录失真性产生的异常蛋白结构域的序列的多肽,特别是至少一种含有选自 SEQ ID N0s :1-3334 的序列或其含有表位的片段的多肽。

[0024] 本发明还涉及装置或产品,它们包含固定化在支持物上的至少一种含有由缺口或插入转录失真性产生的异常蛋白结构域的序列的多肽,特别是至少一种含有选自 SEQ ID N0s :1-3334 的序列或其含有表位的片段的多肽。

[0025] 本发明的另一个目的是试剂盒,其至少包含上面定义的装置或产品,以及进行免疫反应的试剂。

[0026] 本发明还涉及在对象中调节免疫应答的方法,所述方法包括处理对象以贫化表达特异性针对上面定义的多肽的 TCR 的免疫细胞。这样的免疫细胞典型地包括 B 细胞、树突状细胞或 T 细胞。贫化可以通过本技术领域已知的方法、例如使用特异性配体的离体 (ex vivo) 贫化来完成。

## 附图说明

[0027] 图 1. 位于 ORF 中的 K 缺口的图示。

[0028] 显示了在 Y 轴的癌症组和 X 轴的正常组中测量到的 ORF 中所有统计学显著的 K 缺口的 % 偏差。红色菱形表示选择用于生物学评估的 45K 缺口 (表 5)。插入的图显示受到位于 ORF 中指定数量的统计学显著的 K 缺口影响的转录本数量。

[0029] 图 2. 在肺癌患者 c-DNA 文库中丝切蛋白 (Cofilin) 基因上预测的缺失的异常 mRNA 检测。

[0030] 图 2a. 所选缺口的生物信息学预测和性质。

[0031] 图 2b. 克隆策略。

[0032] 图 2c. 对正常和变体进行的 qPCR。

[0033] 图 2d. cDNA 变体序列。

[0034] 图 2e. 基因组 DNA 序列。

[0035] 图 3. 在对照和带有指定形式实体肿瘤的癌症患者的血浆中检测识别 15 种 PAP (表 5 的肽 1-15) 的 IgG。

[0036] 图 3a. 从空白的链亲和素包被的孔中记录到的荧光强度信号中, 减去与生物素化 PAP 温育的每个单独样品记录的荧光强度信号。如果该差值的强度对应于对照中记录到的值的下一半, 用浅蓝显示。深蓝对应于处于对照值的上一半中的信号。高于对照、但是在阳性信号的下一半, 用浅红表示。深红表示处于阳性信号的上一半中的患者。由于样品不足, 空白细胞显示测试不能进行。对照是 26 位健康个体, 患有指定的各种形式癌症的患者对于每种癌症来说按照阶段从顶 (早期) 至底 (晚期) 排序。在图的底部显示了 Wilcoxon 检验的显著性 p- 值。

[0037] 图 3b. 在对照和带有指定形式实体肿瘤的癌症患者的血浆中检测针对 13 种预测异常肽的 IgG。

[0038] 从使用 PAP 14 和 15、或当 PAP 14 的信息不能获得时单独使用 PAP15 记录到的荧光强度信号平均值中, 减去与生物素化 PAP 温育的每个单独样品记录到的荧光强度信号。原始数据与图 3A 的相同。计算出的信号强度如果对应于对照中记录到的值的下一半, 用浅蓝显示。深蓝对应于处于对照值的上一半中的信号。高于对照、但是在阳性信号的下一半, 用浅红表示。深红表示阳性信号的上一半。

[0039] 图 4. 在 NSCLC 相对于未患癌症的对照的血清中, 针对 PAP 的 IgG 的检测。

[0040] 图 4A. 测量了针对 37 种 PAP 的 IgG (表 5 的前 37 种肽)。产生了具有 37 种 PAP 对应氨基酸序列的 N- 末端生物素化肽, 并用作诱饵在链亲和素孔中结合推定的免疫球蛋白。样品以 1/100 稀释度进行测试, 在清洗后, 与 PAP 结合的 IgG 通过抗人类 IgG Fc 结构域的第二抗体显示。对照患者包括 25 位健康个体和 12 位患有 COPD 的患者 (浅蓝和深蓝组) 和 49 位 NSCLC 患者 (研究 II, 表 4) (红色组)。显示了病例相对于所有对照的 Wilcoxon 检验的 p- 值。

[0041] 图 4B. 37 个对照与 49 个肺癌之间的统计学分析。对于每个 TIAB 给出了非参数 Wilcoxon 检验的 p- 值。

[0042] 图 5. 没有检测到针对规范肽 (CP) 的 IgG

[0043] 测量了针对基因组的规范阅读所对应的肽、即没有缺口的 RefSeq 的相应 mRNA 翻译的相应肽的 IgG 的荧光强度。规范肽的序列在表 6 中给出。

[0044] 图 5A. 测量了针对 PAP 1 和在同样基因上选择的规范肽 (I<sub>64</sub> 到 Q<sub>93</sub>) 的 IgG。还测量了针对来自白蛋白的规范肽 (Q<sub>128</sub> 到 P<sub>143</sub> 之间) 的 IgG。所有来自图 4 的患者均包括在内。

[0045] 图 5B. 规范的和预测的异常肽序列的示意图。

[0046] 图 5C. 显示了 11 个对照 (蓝色组) 和 16 个 NSCLC (红色组) 的针对 PAP 7、24 和 28 及其规范肽的 IgG。

[0047] 图 6. 在小鼠 (*Mus musculus*) 血清中检测针对 PAP 的 IgG

[0048] 图 6A. 人类 (*Homo sapiens*) 与小鼠序列之间的同源性的生物信息学分析。显示了 PAP 7、48 和 62 的 mRNA 和 PAP 比对, 并显示出这些序列是保守的。阴性对照 CP 7 的蛋白序列在人类和小鼠之间也是保守的。PAP 2 和 9 在对照和人类 NSCLC 之间是可区别的, 但是在小鼠中不保守。

[0049] 图 6B. 对 12 只正常小鼠 (C57B1/6) 皮下注射  $5 \times 10^5$  个 LLC1 细胞。在注射当天和 1-2-3 周后, 测量针对 PAP 48、62、7 以及相应 CP 的 TIAB 滴度。在注射当天和 3 周后, 在 4 只小鼠中测量针对 PAP 2 和 9 的 TIAB 滴度。显示了平均值  $\pm$  SEM。

[0050] 图 7. 在肺癌与未患癌症的对照的血清中, 针对 PAP 的 IgG 滴度的组合。

[0051] 图 7A. 对照患者包括 161 位健康个体 (蓝色组) 和 140 位肺癌患者 (研究 III, 表 4), 肺癌中包括腺癌 ADK (红色组)、鳞状癌 (橙色组) 和其他 (黄色组)。支持向量机 (Support Vector Machine) 可以区分对照与具有 6 个 PAP (7、29、48、66、68、70) 的肺癌。显示了与超平面的距离; 通过 SVM 模型显示为负值的患者被分类为无癌。显示出正值的患者被分类为有癌。

[0052] 图 7B. 显示了被分类为阳性的肺癌患者根据其年龄的百分率。

[0053] 图 7C. 显示了被分类为阳性的肺癌患者根据其疾病的组织病理学的百分率。

[0054] 图 7D. 显示了在手术后 3 年无病的肺癌患者与死亡或活着但癌症复发的患者之间, 距 SVM 超平面的距离的差异。

[0055] 图 8. 在肺癌与乳腺癌的血清中, 针对 PAP 的 IgG 滴度的组合。对照患者包括 20 位健康个体 (蓝色组)、20 位肺癌患者 (红色组) 和 20 位乳腺癌患者 (紫色组) (研究 IV, 表 4)。

[0056] 图 8A. 显示了辨别肺癌与未患癌症的对照的 PAP 组合。

[0057] 图 8B. 显示了辨别肺癌与乳腺癌的几种 PAP 组合。

[0058] 图 8C. 显示了辨别乳腺癌与未患癌症的对照的 PAP 组合。

[0059] 发明详述

[0060] 本发明涉及通过测量 TIAB 水平来评估对象的生理状态的新方法和产品。更具体来说, 本发明涉及通过测量来自对象的样品中的 TIAB 水平, 来评估对象中癌症的存在、风险或阶段的方法。本发明也适合于评估对象对治疗的响应性、监测癌症的进展或扩散以及筛选候选药物。

[0061] 转录失真性是指一种新的机制, 通过该机制在细胞中从单一转录本序列产生几种不同的 RNA 分子。这种新鉴定到的机制可能影响任何基因, 是非随机的, 并服从特定的规则, 正如在本文中引为参考的共同待决申请 n° PCT/EP07/057541 中公开的。

[0062] 本申请显示,转录失真性能够在 RNA 分子中导入缺口或插入,从而从单个基因产生多种可检测的异常蛋白序列(TI 多肽序列)。这些 TI 多肽序列特别引人关注,因为它们的长度足以包含表位,哺乳动物可以产生针对该表位的抗体。结果,这种异常蛋白在对象中的表达,可以通过测量来自对象的样品中相应抗体或带有 TCR 的细胞的存在来评估。

[0063] 现在,本发明提供了用于预测和/或鉴定由缺口或插入转录失真性事件从任何基因产生的这种异常蛋白结构域的序列的方法,以及生产含有这样的 TI 序列的多肽的方法。本发明还公开了超过 2000 种缺口 TI(gTI)多肽和超过 1000 种插入 TI(iTI)多肽,并证明在人类样品中针对这些多肽序列的抗体的存在与对象中癌症的存在之间的显著关联性。更具体来说,在大多数(>75%)具有常见形式实体肿瘤患者的血清中,检测到了超过正常对象的针对预测的异常肽(PAP)的特异性 IgG 水平的增加。所有 7 种常见形式的实体肿瘤(结肠、肺、乳腺、卵巢、子宫、头颈部和黑素瘤)都引起针对异常蛋白的 IgG 的产生。在大多数患有早期疾病、即淋巴结阴性并且没有转移的对象中,观察到了特异性 IgG 水平的增加。

[0064] 因此,测量这种被称为 TIAB(转录失真性抗体)的针对 TI 多肽的抗体、或带有特异性针对这种异常结构域的 TCR 受体的相应免疫细胞,代表了用于检测和监测疾病、以及用于药物开发的新途径。

#### [0065] TIAB

[0066] 在本发明的上下文中,术语 TIAB(“转录失真性抗体”)是指与包含在由 TI、特别是由 gTI 或 iTI 产生的蛋白序列中的表位特异性结合的抗体。更具体来说,TIAB 是指由哺乳动物天然产生的、针对包含在由 TI、特别是由 gTI 或 iTI(缺口和插入转录失真性)产生的蛋白序列中的表位的抗体。TIAB 可以是任何类型,包括 IgG、IgM、IgA、IgE、IgD 等。当抗体与特定表位或序列的结合能够可靠地与非特异性结合(即与另一个抗原、特别是不包含所述结构域的天然蛋白)区分开时,抗体对于所述表位或序列是“特异的”。

[0067] 一方面,TIAB 或 TIAB 的部分可以附着到固相支持物上。这种附着将 TIAB 维持在适合的构象中,允许当与含有特异性 gTI 或 iTI 多肽的样品接触时与所述特异性 gTI 或 iTI 多肽结合。附着可以是共价或非共价的,直接附着到支持物上或通过间隔基团。在本领域中已经报道了将抗体固定化到支持物(聚合物、陶瓷、塑料、玻璃、氧化硅等)上的各种技术。支持物可以是磁性的,例如磁性珠,以便于例如分离。

[0068] 带有对这样的 TI 多肽特异的 TCR 的免疫细胞,包括免疫系统的任何含有 TCR 的细胞,例如,T 细胞例如 CTL、CD4+ 淋巴细胞、CD8+ 淋巴细胞和/或 Treg 细胞,以及抗原呈递细胞:B 细胞、树突状细胞或巨噬细胞。具体来说,该术语包括任何产生 TIAB 的免疫细胞。这样的细胞可以在常规条件下培养,并使用本发明的 TI 多肽作为(共)刺激因子在体外或离体扩增。

#### [0069] TI 多肽及其生产

[0070] 正如将在下面公开的,现在,本发明公开各种不同的 TI 多肽的序列,并允许从实际上任何基因预测 TI 序列。

[0071] 在第一个实施方案中,本发明涉及含有 gTI 序列、即由缺口转录失真性产生的异常蛋白结构域的序列的分离多肽。本发明的多肽的具体例子包括选自 SEQ ID NOs :1 到 2206 的序列(参见表 3a)或其含有表位的片段。

[0072] 在第二个实施方案中,本发明涉及含有插入 TI 序列、即由插入转录失真性产生

的异常蛋白结构域的序列的分离多肽。本发明的多肽的具体例子包括选自 SEQ ID NOs : 2207-3334 的序列 (参见表 3b) 或其含有表位的片段。

[0073] 术语“含有表位的片段”是指含有形成抗体或 TCR 表达细胞的免疫学表位的至少 6 个、优选至少 8 个、更优选至少 10 个、最优选至少 12 个连续的氨基酸残基的任何片段。这样的表位可以是线性的或构象的,并对 B 或 T 细胞特异。

[0074] 本发明的 TI 多肽典型地包含 8 到 100 个氨基酸,优选 8 到 50 个、更优选 10 到 40 个氨基酸。本发明的多肽可以通过任何常规技术、例如人工多肽合成或重组技术来产生。

[0075] 本发明的多肽可以任选地包含附加的残基或功能,例如但不限于附加的氨基酸残基、化学或生物学基团,包括标记物、标签、稳定剂、定向部分、纯化标签、分泌肽、官能化反应性基团等。这样的附加残基或功能可以被化学衍生、作为融合蛋白的氨基酸序列区添加、复合或以其它方式共价或非共价结合。它们也可以包含天然或非天然氨基酸。多肽可以采用溶解形式,或附着于(或复合或包埋于)支持物、例如基质、柱、珠子、板、膜、载玻片、细胞、脂质、孔等。

[0076] 在具体实施方案中,多肽被生物素化以与链亲和素形成复合物。

[0077] 本发明的多肽可以作为单体或多体存在。此外,它们可以是线性构象或特别是空间构象。就此而言,多肽可以包含在特定支架中以显示特定构象。

[0078] 本发明的多肽可以在疫苗组合物中用作免疫原或用于生产特异性抗体。它们也可用于将药物或其他分子(例如标记物)靶向生物体中的特定位点。它们也可用作特异性试剂,检测或定量来自任何样品的特异性抗体或带有 TCR 的免疫细胞。

[0079] 就此而言,本发明的一个具体目的在于包含如上定义附着到固相支持物上的多肽的装置或产品。附着优选为末端附着,从而将多肽维持在适合的构象中,以允许当与含有特异性抗体的样品接触时与该抗体结合。附着可以是共价或非共价的,直接附着到支持物上或通过间隔基团。在本领域中已经报道了将肽固定化到支持物(聚合物、陶瓷、塑料、玻璃、氧化硅等)上的各种技术,例如在 Hall 等, Mechanisms of ageing and development 128(2007)161 中公开的。支持物可以是磁性的,例如磁性珠,以便于例如分离。

[0080] 装置优选包含例如以预定的次序排列的多种本发明的多肽,以便可以使用同一种装置检测或测量几种 TIAB。

[0081] 装置典型地由任何固体或半固体支持物制成,例如滴定板、培养皿、载玻片、孔、膜、珠子、柱等。支持物典型地包含至少两种选自 SEQ ID NO :1 到 3334 的多肽或其含有表位的片段,更优选选自表 5 的 45 种 PAP 多肽(包括在 SEQ ID NO 1-3334 中)。

[0082] 在最优选实施方案中,本发明的方法或支持物使用含有表 5 的不同 PAP 多肽序列的至少 2 种、优选至少 3 种多肽的组合。

[0083] 在具体实施方案中,装置或方法使用选自 PAP 1、2、4、6、7、24、25、28、29、44 或 48(表 5)的至少一种、两种或三种多肽。

[0084] 在另一个具体实施方案中,本发明的方法或支持物使用表 5 的不同 PAP 多肽的组合,所述组合选自下列:

[0085] - 多肽 PAP7、PAP66、PAP70、PAP29、PAP68 和 PAP48 ;

[0086] - 多肽 PAP7、PAP48、PAP70 和 PAP29 ;

[0087] - 多肽 PAP6、PAP29、PAP70 和 PAP82 ;

[0088] - 多肽 PAP6、PAP7、PAP29、PAP48、PAP70 和 PAP82；

[0089] - 多肽 PAP6、PAP29、PAP70 和 PAP69；

[0090] - 多肽 PAP7、PAP48、PAP70、PAP74 和 PAP29；或

[0091] - 多肽 PAP7、PAP29 和 PAP94。

[0092] 在具体实施方案中，装置包含 2 到 10 种多肽。

[0093] 支持物可以包含附加的物体或生物学元件，例如对照多肽和 / 或具有不同免疫反应性的多肽。

[0094] 多肽与 TIAB 之间的免疫复合物的形成可以通过已知技术进行评估，例如使用特异性针对人类抗体的第二标记抗体，或通过竞争反应等。

[0095] 本发明的另一方面在于试剂盒，其包含上面公开的装置，以及一种或数种执行免疫反应、即形成和检测免疫复合物的试剂。

[0096] TI 多核苷酸

[0097] 本发明的另一个实施方案涉及含有编码上面定义的多肽的核苷酸序列的多核苷酸或其互补链。具体来说，该多核苷酸包含编码选自 SEQID NOs :1-3334 的多肽的第一种核苷酸序列或与其互补的序列，以及长度为 100 个或以下核苷酸的第二种核苷酸序列，其中所述第二种核苷酸序列在天然存在的核酸中与所述第一种核苷酸序列相邻。与第一种核苷酸序列相邻的第二种核苷酸序列的长度可以是例如 75、50、25、10 或 0。

[0098] 本发明的多核苷酸可以是 DNA 或 RNA，例如互补 DNA、合成的 DNA、mRNA 或它们的类似物，这些类似物含有例如修饰的核苷酸例如 3' 烷氧基核糖核苷酸、甲基膦酸等和肽核酸 (PNA) 等。多核苷酸可以被标记。多核苷酸可以使用本申请中包含的序列信息，按照本身在本技术领域中公知的技术来生产，例如通过化学合成方法、体外转录或通过重组 DNA 方法来生产。具体来说，多核苷酸可以通过化学寡核苷酸合成、文库筛选、扩增、连接、重组技术及其组合来生产。

[0099] 本发明的具体实施方案在于编码含有选自 SEQ ID 2207-3334 的序列或其含有表位的片段的多肽的多核苷酸。

[0100] 本发明的多核苷酸可以包含附加的核苷酸序列，例如调控区，即可用于引起或调控多肽表达的启动子、增强子、沉默子、终止子等。

[0101] 本发明的多核苷酸可用于产生本发明的重组多肽。它们也可用于设计特异性试剂，例如引物、探针或反义分子（包括反义 RNA、iRNA、适配体、核酶等），所述试剂特异性检测、结合或影响编码如上定义的多肽的多核苷酸的表达。它们也可用作治疗分子（例如作为基因疗法程序中工程化病毒、例如但不限于工程化腺病毒或腺相关病毒载体的一部分），或用于产生重组细胞或遗传修饰的非人类动物，它们可用于例如在化合物文库中筛选调节如上定义的多肽的活性的药剂。

[0102] 在本发明的上下文中，核酸“探针”是指具有能够与转录失真性结构域或其互补体选择性杂交、并适合于检测含有所述结构域或互补体的样品的存在（或其量）的多核苷酸序列的核酸或寡核苷酸。探针优选与转录失真性结构域完全互补，但是，可以容忍一定的错配。探针典型包含长度为 8 到 1500 个核苷酸、例如 10 到 1000 个、更优选 10 到 800 个、典型 20 到 700 个核苷酸的单链核酸。应该理解，较长的探针也可以使用。本发明的优选探针是长度在 8 到 400 个核苷酸之间、能够与转录失真性结构域特异性杂交的单链核酸分子。

[0103] 术语“引物”是指核酸或寡核苷酸,其所具有的多核苷酸序列能够与转录失真性结构域或其互补体、或与位于转录失真性结构域侧翼的核酸区域选择性杂交,并适合于在含有所述结构域或互补体的样品中扩增全部或部分所述转录失真性结构域。本发明的典型引物是长度为大约 5 到 60 个核苷酸、更优选长度为大约 8 到大约 50 个核苷酸、更进一步优选长度为大约 10 到 40、35、30 或 25 个核苷酸的单链核酸分子。优选完全互补性,以确保高特异性。但是,正如上面对于探针所讨论的,可以容忍一定的错配。

[0104] 本发明的另一方面在于载体,例如包含如上定义的多核苷酸的表达或克隆载体。这样的载体可以选自质粒、重组病毒、噬菌体、附加体 (episome)、人工染色体等。许多这样的载体是可商购的,并可以按照本技术领域公知的重组技术来生产,例如在手册例如在此以其全文引为参考的 Sambrook 等,《分子克隆》(第二版)(Molecular Cloning(2d ed.)) (Cold Spring Harbor Press 1989) 中提到的方法。

[0105] 本发明的另一方面在于用如上定义的多核苷酸或载体转化或转染的宿主细胞。宿主细胞可以是能够被遗传修饰并优选被培养的任何细胞。细胞可以是真核的或原核的,例如哺乳动物细胞、昆虫细胞、植物细胞、酵母、真菌、细菌细胞等。典型的例子包括哺乳动物的原代或确立的细胞 (3T3、CHO、Vero、Hela 等) 以及酵母细胞 (例如酵母 (Saccharomyces) 菌种、克鲁维酵母菌 (Kluyveromyces) 等) 和细菌 (例如大肠杆菌 (E. coli))。应该理解,本发明不限于任何具体细胞类型,并且遵循常用的通用知识,能够适用于任何种类的细胞。

#### [0106] 诊断

[0107] 本发明允许执行检测或诊断分析,所述分析可用于例如检测源自于对象的样品中疾病的存在、不存在、易感性、风险或严重性。在具体实施方案中,疾病是癌症。术语“诊断学”应该被解释为包括了药物基因组学、预后等的方法。

[0108] 在一个具体方面,本发明涉及在体外或离体检测对象、优选为人类对象中疾病的存在、不存在、易感性、风险或严重性的方法,所述方法包括将来自对象的样品与如上定义的多肽相接触,并确定免疫复合物的形成。最优选,将多肽固定化在支持物上。在优选实施方案中,方法包括将样品与如上公开的装置接触,并测定免疫复合物的形成。优选,多肽选自 SEQ ID NO :1-3334 或其含有表位的片段,最优选,选自表 5 的 45 种 PAP (包含在 SEQ ID NO 1-3334)。

[0109] 在另一方面,本发明涉及在体外或离体检测对象、优选为人类对象中疾病的存在、不存在、易感性、风险或严重性的方法,所述方法包括将来自对象的样品与如上定义的 TIAB 或 TIAB 的一部分或相应的带有 TCR 的细胞相接触,并确定免疫复合物的形成。最优选,将 TIAB 或相应的带有 TCR 的细胞固定化在支持物上。在优选实施方案中,方法包括将样品与如上公开的装置接触,并测定免疫复合物的形成。在另一个优选实施方案中, TIAB 或相应的带有 TCR 的细胞对选自 SEQ IDNO :1-3334 或其含有表位的片段、优选为选自表 5 的 45 种 PAP 的多肽具有特异性。

[0110] 本发明的具体目的在于在对象中检测癌症的存在、不存在、易感性、风险或严重性的方法,所述方法包括在体外或离体将来自对象的样品与如上定义的多肽相接触,并确定免疫复合物的形成。更优选,多肽被固定化在支持物上,并选自 SEQ ID NOs :1-3334 或其含有表位的片段,最优选选自表 5 的 45 种 PAP。

[0111] 本发明的另一个目的在于体外或离体检测生物学样品、优选人类生物学样品中疾

病的存在、不存在、易感性、风险或严重性的方法,所述方法包括将所述样品与如上定义的多肽接触,并确定表达对这样的多肽特异性的 TCR 的免疫细胞的存在。优选,多肽选自 SEQ ID NOs :1-3334 或其含有表位的片段,最优选选自表 5 的 45 种 PAP。

[0112] 本发明的另一方面在于体外或离体评估对象、优选为人类对象中转录失真性水平的方法,所述方法包括将来自对象的样品与如上定义的多肽接触,并确定免疫复合物的形成。最优选,多肽被固定化在支持物上。在优选实施方案中,方法包括将样品与如上定义的装置接触,并确定免疫复合物的形成。

[0113] 本发明的另一方面在于体外或离体评估对象、优选为人类对象中转录失真性水平的方法,所述方法包括将来自对象的样品与如上定义的多肽相接触,并确定表达对这种多肽特异性的 TCR 的免疫细胞的存在。

[0114] 本发明的另一个实施方案涉及测定癌症治疗功效的方法,所述方法包括 (i) 在来自对象中样品中,测定至少一种包含由转录失真性产生的异常蛋白结构域的序列的多肽的水平,或 TIAB 或相应的带有 TCR 的细胞的水平,以及 (ii) 将所述水平与治疗之前或治疗较早阶段从所述对象获取的样品中的水平进行比较。优选下,多肽选自 SEQ ID NOs :1-3334 或其含有表位的片段,更优选选自表 5 的 45 种 PAP。

[0115] 本发明的另一方面涉及确定个体是否制造下述多肽的方法,所述多肽含有由转录失真性产生的异常蛋白结构域的序列、特别是含有选自 SEQ ID NOs :1-3334 的序列或其含有表位的片段,所述方法包括将从所述个体获得的样品与指示所述多肽存在的药剂相接触,并确定所述药剂是否与所述样品结合。

[0116] 在第一个实施方案中,将从对象获得的样品与多肽相接触,所述多肽与下述多肽的抗体结合,所述多肽特异性针对含有由转录失真性产生的异常蛋白结构域的序列、特别是含有选自 SEQ ID NOs :1-3334 的序列或其含有表位的片段。

[0117] 在另一种实施方案中,将从对象获得的样品与结合免疫细胞的多肽相接触,所述免疫细胞包含特异性针对下述多肽的 TCR,所述多肽含有由转录失真性产生的异常蛋白结构域的序列、特别是含有选自 SEQ ID NOs :1-3334 的序列或其含有表位的片段。

[0118] 在另一个实施方案中,将从对象获得的样品与特异性针对下述多肽的抗体或其部分相接触,所述多肽含有由转录失真性产生的异常蛋白结构域的序列,特别是含有选自 SEQ ID NOs :1-3334 或其含有表位的片段的序列。

[0119] 在另一个实施方案中,将从对象获得的样品与含有特异性针对下述多肽的 TCR 的免疫细胞相接触,所述多肽含有由转录失真性产生的异常蛋白结构域的序列,特别是含有选自 SEQ ID NOs :1-3334 或其含有表位的片段的序列。

[0120] 本发明还涉及监测对象中癌症的进展或扩散的方法,所述方法包括 (i) 将从所述对象获得的样品与至少一种包含由转录失真性产生的异常蛋白结构域的序列的多肽相接触,(ii) 测定所述样品中 TIAB 或相应的带有 TCR 的细胞的水平,以及 (iii) 将所述水平与参照水平进行比较,所述参照水平是来自未患癌症的对象的平均值或中值、或来自癌症发作前对象的对照值。优选多肽包含 PAP 12 的序列,针对它的抗体滴度在不可手术的患者中与可手术患者相比明显增加,从而提供了与疾病扩散相关的指示。

[0121] 样品中 TIAB 或相应的带有 TCR 的免疫细胞的存在(或增加),指示了癌症疾病的存在、易感性或发展阶段。因此,本发明允许设计更有效且定制的适当治疗性干预。此外,

这种在症状出现之前的水平的测定,允许实行预防性方案。

[0122] 本发明的诊断方法可以在体外、离体或体内进行,优选在体外或离体进行。样品可以是源自于对象的含有适合的抗体或免疫细胞的任何生物学样品。这样的样品的例子包括体液、组织、细胞样品、器官、活检样品等。最优选的样品是血液、血浆、血清、唾液、精液等。在执行方法之前,可以对样品进行处理,以便赋予或改进抗体的测试可用性。处理可以包括例如下列一种或多种:细胞裂解(例如机械、物理、化学等)、离心、萃取、柱层析等。

[0123] 在优选实施方案中,测试在血清或血浆上进行。

[0124] 此外,在最优选实施方案中,样品在测试前按照 EP08305293.6 中所公开的进行处理。事实上,申请人显示,当样品在  $-20^{\circ}\text{C}$  下或  $-80^{\circ}\text{C}$  下冷冻保持新鲜时,最适测试条件是不同的。更优选,待测试样品进行了适合于显露出抗体结合位点、改变结合位点的构象或暴露出抗体结合位点的化学或物理处理。更优选,处理包括将样品在至少  $36^{\circ}\text{C}$  的温度下加热一段足以活化抗体的时间。优选的温度是  $36^{\circ}\text{C}$  到  $70^{\circ}\text{C}$  之间,优选在  $36^{\circ}\text{C}$  到  $60^{\circ}\text{C}$  之间。

[0125] 确定样品中 TIAB 或特异性免疫细胞的存在、不存在或相对丰度,可以通过在本技术领域本身已知的多种技术来进行。这样的技术包括但不限于用于检测免疫复合物的方法,例如但不限于 ELISA、放射免疫分析(RIA)、荧光免疫分析、微阵列、微芯片、斑点杂交(dot-blot)、western 印迹、EIA、IEMA、IRMA 或 IFMA(也参见 JP Gosling 主编的《免疫分析实用方法》(Immunoassays, a practical approach), Oxford University Press)。在具体实施方案中,方法包括将样品与多肽在允许免疫复合物形成的条件下相接触,以及使用第二标记试剂揭示所述形成。

[0126] 在典型的实施方案中,方法包括将 TIAB 或免疫细胞的测量水平与参照水平进行比较,其中差异指示了对象中的功能障碍,例如癌症。与参照值相比,变化典型为变动 10%、20%、30%、40%、50% 或以上。更具体来说,与参照值相比,水平的变化是增加,这表明存在癌症。

[0127] 参照值可以是从未患癌症或疾病的个体测定的平均值或中值、从对照患者获得的参照水平、从癌症发作前或具有对照多肽的对象获得的参照水平。

[0128] 在优选实施方案中,与参照水平相比,所述样品中 TIAB 或免疫细胞的水平的变化(例如增加)指示了癌症的存在、风险或发展阶段。

[0129] 接触可以在任何适合的装置中进行,例如板、微量滴定皿、试管、孔、玻璃、柱等。在具体实施方案中,接触在包被有多肽的基质上进行。基质可以是固体或半固体基质,例如任何含有玻璃、塑料、尼龙、纸、金属、聚合物等的适合的支持物。基质可以具有各种不同的形式或尺寸,例如载玻片、膜、珠子、柱、凝胶等。接触可以在适合于要在样品的多肽与抗体之间形成可检测的抗体-抗原复合物的任何条件下进行。

[0130] 在具体实施方案中,方法包括将来自对象的样品与多种如上定义的多肽(包被的支持物)相接触,并确定免疫复合物的存在。

[0131] 在具体实施方案中,方法包括将样品与多组珠子接触,每组珠子包被有如上定义的不同多肽。

[0132] 在另一个具体实施方案中,方法包括将样品与其上排列有几种如上定义的多肽的载玻片或膜相接触。

[0133] 在另一个具体实施方案中,方法包括将样品与多孔滴定板接触,其中至少部分孔

用如上定义的不同多肽包被。

[0134] 本发明可用于确定对象中任何癌症的存在、风险或阶段。这包括实体肿瘤,例如但不限于结肠、肺、乳腺、卵巢、子宫、肝或头颈部癌症,以及黑素瘤、脑肿瘤等。本发明也可用于液体肿瘤,例如白血病。本发明可用于在具有发生癌症的风险的对象中甚至在其早期阶段进行第一次筛查来检测所述癌症。在第二次筛查中,依赖于用于检测的多肽,本发明可用于更精确地鉴定癌症的类型。就此而言,正如在图 3 中公开的,包含 PAP 1、2、3、4 或 7(表 5) 序列、或其含有表位的片段的多肽,允许鉴定具有各种不同类型癌症的患者。

[0135] 在具体实施方案中,本发明用于确定肺癌的存在、风险或阶段,多肽包含选自 PAP 1、2、4、6、7、24、25、28、29、44 和 48(表 5)(图 4) 的序列或其含有表位的片段。

#### [0136] 药物筛选

[0137] 本发明还允许通过评估候选分子调节 TIAB 水平或相应免疫细胞的能力,来设计(或筛选)新药物。

[0138] 本发明的具体目的在于选择、表征、筛选或优化生物活性化合物的方法,所述方法包括确定测试化合物是否调节 TIAB 水平。TIAB 水平的调节可以针对特定蛋白、或针对预定的一组蛋白或全面进行评估。

[0139] 本发明的另一个实施方案在于选择、表征、筛选或优化生物活性化合物的方法,所述方法包括在体外将测试化合物与基因相接触,并确定所述测试化合物调节从所述基因生产含有转录失真性缺口和插入的 RNA 分子的能力。

[0140] 本发明的另一个实施方案在于选择、表征、筛选或优化生物活性化合物的方法,所述方法包括在体外将测试化合物与表达特异性针对上面定义的多肽的 TCR 受体的免疫细胞相接触,并确定所述测试化合物调节所述细胞的活性或生长的能力。

[0141] 上述的筛选分析可以在任何适合的装置中进行,例如板、管、平皿、烧瓶等。典型地,在多孔微量滴定皿中进行分析。使用本发明,可以平行地分析几种测试化合物。此外,测试化合物可以是各种不同来源的,天然和组成的。它可以是任何有机或无机物质,例如脂质、肽、多肽、核酸、小分子,分离的或与其它物质形成混合物。化合物可以是例如化合物组合文库的全部或一部分。

[0142] 本发明的其他方面和优点将在下面的实验部分中公开,实验部分应该被当作说明性的,不对本申请的范围构成限制。

#### [0143] 实施例

##### [0144] A- 材料与方法

##### [0145] 血浆样品

[0146] 从出于与癌症无关的生物学测试的目的在当地大学医院(Nancy University Hospital) 看病的正常人类对象(n = 26) 抽取血液样品。临床记录由训练有素的医生检查,他们确定这些对象没有急性病、不被怀疑患有活动的癌症、过敏或自体免疫病症。该组包括具有癌症风险因子、例如吸烟和肥胖的患者,一名对照在取血样之前 10 年通过外科手术除去了子宫癌,一名在取血样时是怀孕的。患有慢性阻塞性肺病的患者(n = 12) 被召集到同一个科(n = 6) 或召集到同一所医院的核医学科(n = 6)。所有患有 COPD 的患者在取血样时没有恶化性事件。患有各种形式实体肿瘤的患者(n = 46) 在治疗前的 PET-CT 癌症扩散评估时取样,通过分析病理学样品完成分级(Nancy University Hospital)。患有活动

性 NSCLC 的患者 (n = 49) 被召集到 Strasbourg University Hospital, 他们是肺癌纵向研究的一部分。在分级时抽取血样。所有在这些医学研究机构看病的患者通过签署同意书, 同意将他们的样品出于研究目的匿名进行测试。这些样品的收集和分析, 按照法国法律向法国卫生部 (French ministry of Health) 和研究部 (Ministry of Research) 申报。

#### [0147] 生物信息学程序

[0148] 按照以前的描述进行分析, 但是做出了下列修改<sup>34</sup>。使用数据库中可以获得的组织来源, 对每个 EST 进行检索并指派为癌症组或正常组序列。然后使用 MegaBlast 2.2.16<sup>38</sup> 将每个序列对来自 NCBI 的人类 RNA RefSeq 进行一次比对。保留单个最好比对分值。每个没有在其长度的 70% 以上进行比对的 EST 不纳入考虑。只有当上游 10 个碱基和下游 10 个碱基与 RefSeq 完全匹配时, 具有单个碱基序列变异的位置才被纳入考虑。将每个比对末端的最前和最后 50 个碱基删除。如果需要, 缺口和插入位于任何元组 (n-uplet) 的最后一个核苷酸上。

#### [0149] 生物化学分析

[0150] N-末端生物素化的、具有将鉴定到的 K 缺口考虑在内通过 RefSeq 的计算机翻译而确定的氨基酸序列的肽, 以及对应于白蛋白和 MRPL12 基因的规范序列的肽, 从不同制造商购买。将样品稀释 100 倍, 使用可商购试剂并按照制造商的说明书, 在 ImmunoCAP 100 (Phadia, Uppsala, Sweden) 上进行 IgG 检测。除了在图 3 中由于样品缺少的几个例外之外, 所有样品分析双份。癌症和对照以及 PAP 的次序或测试是随机的。在不存在内标的情况下, 结果表示成荧光单位 (FU)。

#### [0151] 统计学方法

[0152] EST 碱基组成的统计学显著性的检验和由于多次测试而产生的假阳性的估算, 按照以前描述的<sup>34</sup>进行。使用非参数秩比较 Wilcoxon 检验来检验 IgG 滴度的差异。

#### [0153] B- 结果

##### [0154] 从 TI 缺口产生的蛋白结构域的鉴定

[0155] 为了在基因组范围的尺度上分析 EST 异质性, 我们从未过滤过的 (noncurated) dbEST<sup>39</sup> 检索了在 2000 年 1 月到 2007 年 7 月之间公布的所有序列。根据它们在数据库中标明的正常或癌症起源将这些序列分开。然后将每个 EST 对来自 NCBI 的所有人类 RefSeq RNA 序列 (2007 年 7 月) 进行一次比对<sup>40</sup>。首先, 我们检验了正常与癌症矩阵之间在任何给定 RefSeq 位置处发生的差异的统计学显著性, 然后估算由多次测试产生的假阳性<sup>41</sup>。具有统计学显著性的序列差异的位置, 如果变异在癌症中过多被称为 K, 相反, 当在正常中过多时被称为 N。

[0156] 从表 1 的结果中得出的最重要的观察是 K 缺口的发生甚至比 K 碱基取代更加常见。定义这些缺口的生物信息学限制是严格的: 具有单个碱基缺口的给定 EST 位置, 只有当其侧翼上游和下游存在 10 个与 RefSeq 完全匹配的碱基时, 才被纳入考虑。对于位于 ORF 内的 2191 个 K 缺口来说, 正常和癌症的带缺口 EST 百分率显示在图 1 中。引人注目的是, K 插入比 N 插入更常见 5 倍, K 缺口比 N 缺口更常见 ~ 13 倍。减去统计学假阳性的估算值, 将比率增加到 38, 这是因为 K 缺口的 p- 值比 N 缺口的低得多。与 K 不同, N 缺口很少并明显包含大比例的假阳性 (表 1)。因此, 我们集中于对位于 ORF 中的 K 缺口位置进行进一步分析。

[0157] 表 2 概述了我们分析的条目。显然,没有由于癌症或正常组中显示的 EST 或转录本数量的差异而导致的明显偏差。从表 1 和 2 中,人们可以估算出,在癌细胞中 mRNA 序列的变异、即取代、插入和缺口,以每一千个碱基 1-2 个的比率发生。这大大超过了体细胞突变的比率:每一百万个碱基 1-2 个<sup>31</sup>。

[0158] mRNA 中的 SBG (单碱基缺口) 可以由体细胞或生殖细胞突变、或由忽略了读取单个 DNA 碱基并继续转录的 RNAP 所引起。我们也考虑到了下述假说,即剪接机制的向前或向后滑动导致 SBG 位于第一个或最后一个外显子碱基上。但是发现后一种机制是不可能的,因为 99.2% 鉴定到的 SBG 不在紧邻外显子-内含子边界内。

[0159] 丢失的 mRNA 碱基的组成具有下列次序:U(47%) > C(39%) >> G(10%) > A(4%)。这种分布主要是由于随机而偏离的(拟合优度卡方检验,双尾,  $\alpha = 0.05$ ,  $P = 10^{-248}$ )。此外,分别有 99%、76%、98% 和 95% 的 U、C、G 和 A 缺口,出现在一个或多个相同碱基的重复区内。最后,对于 97% 的 U 缺口来说,在紧接下游发现了 G。因此,基因组 DNA 的邻近序列部分决定了癌细胞中 mRNA 缺口的可能出现。DNA 邻近序列对发生 TI 事件的影响的详细分析,将在别处报告。

#### [0160] 异常 mRNA 检测

[0161] 为了验证这些生物信息学结论,我们从肺癌患者的 c-DNA 文库克隆了质粒,该质粒在 qPCR 和测序后显示出在预测的位置出现 SBG (图 2A-E)。我们分析了从同一个个体的正常组织获得的同样数量的克隆,没有发现任何序列变异(数据未显示)。从同一患者的癌症、邻近和正常组织获得的基因组 DNA 的直接测序,明确地证明了在该位置缺乏体细胞或生殖细胞突变。我们不能排除鉴定到的 mRNA 缺口是克隆过程中人工产生的。但是,这种事件与通过生物信息学预测的位置精确一致,是不可能的。因此,可以合理地假设,在癌细胞中,少量但是可检测比例的 mRNA 不是基因组 DNA 的忠实拷贝。

#### [0162] 材料与amp;方法

##### [0163] 质粒制备:

[0164] 使用了 pBAD 质粒 (Invitrogen) 以便在克隆到的序列上游带有诱导型启动子。将从 pBS-SK+ 质粒扩增到的  $\alpha$  肽的序列,与 pBAD 质粒的 CCATGG 克隆位点中存在的 ATG 序列异相克隆 (cloned out of phase),产生了 pBAD- $\alpha$  质粒。在不存在被克隆序列的情况下,不产生  $\alpha$  肽,大肠杆菌菌落为白色。

##### [0165] 插入物制备和克隆:

[0166] 来自从同样个体获得的癌性肺和邻近正常组织的 cDNA (Biochain Inc.), 使用与 CFL1 基因互补的寡核苷酸和高保真性 Phusion 聚合酶 (Finnzyme), 按照制造商的推荐方法通过 PCR 进行扩增。然后将 cDNA 在 Nucleospin Extract II 柱 (Macherey Nagel) 上进行纯化,在琼脂糖凝胶上可视化,并用 NcoI 和 NheI 限制性酶 (Biolabs) 消化。然后将产物连接到用同样酶消化并被脱磷酸化的 pBAD- $\alpha$  质粒中。使用连接混合物转化大肠杆菌 TOP10 (Invitrogen) 细胞,并铺于 LB 氨苄青霉素 (100mg/L) 阿拉伯糖 (0.5%) X-Gal (80  $\mu$ g/mL) 平板上。

##### [0167] 菌落筛选:

[0168] 当克隆没有缺口的 CFL1 序列时, $\alpha$  肽不与 ATG 同相:菌落为白色。如果克隆到带有缺口的 CFL1 序列,产生了  $\alpha$  肽,无活性的  $\beta$ -半乳糖苷酶(存在于细菌的基因组上)得

到互补,大肠杆菌菌落变为蓝色(图 2B)。

[0169] 将蓝色菌落生长在补充有氨苄青霉素(100mg/L)的 LB 培养基中,取 1  $\mu$ L 培养物通过实时 PCR、使用 CFL1 特异性寡核苷酸和 Syber-绿 I(Sigma) 进行筛选(图 2C)。第一个寡核苷酸(图中的绿色)对这两个序列均有特异性,显示出 2 个样品之间质粒拷贝数没有差别。第二个寡核苷酸对 CFL1 参照序列(Refseq) 有特异性,当序列不与 Refseq 一致时(图中的红色)显示出 Ct 之间的差异。

[0170] 提取显示出 Ct 之间差异的质粒 DNA 克隆,并使用质粒(GATC biotech)上存在的寡核苷酸进行测序(图 2C)。

[0171] 将序列与 CFL1 参照序列比对。

[0172] TIAB 检测

[0173] 图 1 显示了在癌症组(Y 轴)和正常组(X 轴)二者中对每个 K 位置记录到的偏差的百分率。K 缺口分布在为数有限(532 或 1.4%)的转录本上(图 1 中的插图)。这些 mRNA 已经失去了它们的规范阅读框,将被翻译成异常的、可能免疫原性的蛋白。为了检验这种假说,我们选择了代表 2206 个 ORF 缺口位置的一组 15 个 K 缺口(表 5, PAP1 到 15),它们分布在 8 条不同的人类染色体上。这 15 个位置分别来自 8、5 和 2 个 U、C 和 G 的缺失。我们证实了预测由具有单个碱基缺口的 mRNA 翻译产生的氨基酸序列 1) 编码长于 12 的氨基酸序列, 2) 不与任何已知的人类蛋白在超过 7 个连续的氨基酸上匹配(Swiss-Prot<sup>42</sup>), 3) 彼此之间没有氨基酸序列同源性。我们还确立,所选择的 K 缺口还未被 Sanger 研究所体细胞突变目录(Sanger Institute Catalogue Of Somatic Mutations(<http://www.sanger.ac.uk/cosmic>)<sup>43</sup>) 或被包含了本筛选中涉及的 15 个基因中的 11 个基因的两项最近的癌细胞基因组深度测序工作<sup>31,32</sup>, 鉴定为癌症体细胞突变。K 缺口不对应于生物学验证的或推导的 SNP。最后,并根据 dbSNP 数据库的最近更新(2007 年 9 月 21 日),没有 SNP 导入的移码与所定义位置上游的单个缺口所产生的移码相同<sup>44</sup>。

[0174] 将从人类对象抽取的血样分成两组或多组。所有样品是残余血清。至少一个组包括患有活动癌症的患者。收集与所有组、包括对照和活性癌症组相关的临床数据,并由训练有素的医生进行确认。关于对照的数据可以包括癌症风险因子。关于癌症患者的数据可以包括分级和对治疗的响应。对组进行设计,以评估一组 TIAB 以及它们对具体诊断指征例如早期癌症检测、癌症类型的鉴定、疾病严重性和进程的预测或对治疗的响应的特异性和灵敏度。

[0175] 产生了对应于这些预测异常肽(PAP)的合成的 N-末端生物素化肽,并各自包被在链亲和素 Elia 孔(Phadia, Uppsala Sweden)上。将来自 46 位癌症患者(研究 I)和 26 位对照对象(表 4)的血清,与空白(没有肽包被的孔)或肽包被的孔温育。在清洗后,与孔结合的 IgG 用产生荧光的商购的抗人类 IgG 不可变结构域第二抗体进行显示。在第一次分析中,将在任何给定对象中使用任何给定 PAP 测量到的荧光强度,从在同样对象中使用没有用肽包被的链亲和素孔(空白)时测量到的荧光强度中减去。结果显示了在癌症患者中,与对照相比,针对 PAP 1、2、4、7 的 IgG 有统计学显著的增加(Wilcoxon 检验分别为  $P < 2E^{-8}$ ;  $P < 2E^{-3}$ ;  $P < 5E^{-2}$ ;  $P < 3E^{-2}$ ) (图 3A)。在年轻(< 50 岁)对较年老(> 50 岁)正常对象中检测到的任何 IgG 的水平没有统计学显著的差异(Wilcoxon 检验 = NS),并且没有由于对照的性别引起的显著差异(Wilcoxon 检验 = NS)。接下来,我们测试了针对 PAP 的 IgG 的

检测是否在癌症和对照对象之间具有差异。当 PAP 包被的孔与空白孔之间的荧光强度差异比在对照组中测量到的最高值荧光强度更高时,测试被认为是阳性的(图 3 浅红和深红)。因此,将特异性任意地设定为 100%。在这些条件下,除了一个之外,所有的 PAP 都在至少一个癌症患者上检测到;15 个 PAP 中的 6 个在 10% 以上的患者中鉴定到了超过对照的 IgG 水平。合在一起考虑,患有 7 种最常见实体肿瘤形式的 46 位患者中的 35 位(76%),具有高于对照组所限定的阈值的 IgG 水平。图 3 显示了结肠、肺和头颈部癌患者具有更多样化的阳性信号组,并分别对 11、10 和 8 种 PAP 阳性。乳腺癌患者的 IgG 只结合 6 种 PAP。这种多样性在患有卵巢、皮肤和子宫癌的患者中进一步减少(图 3)。

[0176] 因此,使用该第一组 15 种 PAP,覆盖度和特异性对于肺癌来说是最适的。灵敏的早期肺癌化验不存在,因此在该阶段的诊断是稀少的,5 年存活率仅为 14% (46)。

#### [0177] 用于早期肺癌诊断的 TIAB 检测

[0178] 本发明的重要暗示是,基于简单的血液化验,有可能使早期肺癌诊断成为可能。数据分析不需要复杂的统计学方法,因此过度拟合的风险降到最低<sup>10</sup>。此外,我们的试验不是系统性搜索生物标志物,而是根据生物信息学预测驱动的假说,因此由于多次测试导致偏差的风险低<sup>10</sup>。但是,由于这种发现的临床暗示,我们在独立的研究中寻求重复。

[0179] 从不同制造商购买了具有选自表 5 的 45 种 PAP 的氨基酸序列的合成的 N-末端生物素化肽,构成 TIAB 诱饵组,并各自包被到 Reacti-Bind 链亲和素包被的板(Pierce Biotechnology, Rockford, Illinois)上。将样品稀释 100 倍,分析以双份进行。与肽结合的血清 IgG 使用商购的结合有酶、特别是磷酸酶的抗人类 IgG 不可变结构域第二抗体进行显示。使用可商购试剂进行与荧光底物的反应。荧光读取在 FLUOstar Galaxy 微孔板读板器(BMG Labtech, Offenburg, Germany)上、按照制造商的说明书进行。

[0180] 如果可以建立阈值来分开至少两组为该指征所设计的人类对象的话,选择组中的 TIAB 用于具体的诊断指征。对于选择的 TIAB 来说,在一个组中的绝对荧光强度高于阈值,以同样方式在另一个组中测量到的荧光低于阈值。

#### [0181] 选择 PAP 以监测疾病进展或扩散

[0182] 我们陈述了在前述实施例的方法(选择用于具体诊断指征的 PAP/TIAB)后,在一组代表了疾病的各种进展或扩散阶段的患者的 PAP 组中选择 PAP。

[0183] 接下来,我们陈述了测试该组 PAP 在代表非小细胞肺癌(NSCLC)的各种阶段的第一组 49 位患者中的功效(图 4)。

[0184] 从 25 位健康对照、12 位患有非癌症肺病,以及在诊断时从 49 位患有不同阶段 NSCLC(表 4,研究 II)的患者获取血样。这些 NSCLC 患者代表了该疾病在法国的当前状态。由于缺乏可靠的早期检测方法,大多数 NSCLC 在诊断时已是晚期疾病,只有 20%处于早期。每位患者和对照对表 5 的前 37 种 PAP 记录到的荧光强度(FI)结果显示为图 4A。组之间的差异的统计学显著性通过 Wilcoxon 检验确定,结果在每张小图上标出。可以看到,对于 37 种 PAP 中的 33 种来说,在肺癌患者中的 FI 与对照相比明显增加。对于 10 种最具辨别力的 PAP 来说,Wilcoxon 检验的 p 值为  $10^{-15}$  到  $10^{-10}$ (图 4B)。单独的 PAP1 允许完美地区分对照和 NSCLC(特异性 = 100%,灵敏度 = 100%)。

[0185] 因此,带有 SBG 的异常 mRNA 翻译的副产物有助于调节对 NSCLC 的体液免疫应答这个假说,似乎是根据的。通过测试与 PAP 结合的 IgG 对其氨基酸序列有特异性,我们验证了

这一点。因此,我们测量了针对白蛋白肽以及对基因组的基因 1 和 3 规范阅读所对应的肽的 IgG 的水平,所述基因 1 和 3 是对那些针对它们相应规范肽 (CP) 的 IgG 最具鉴别力的 PAP 的基因。这些 CP 由编码 PAP (7、24、28) 的相同基因和区段编码,但是其氨基酸序列源自于人类基因组的规范阅读,即没有移码。数据显示,在肺癌患者中针对 CPs 的 Ig 的滴度比针对 PAP 的低得多 (图 5),并且 CP 不能辨别病例与对照 (Wilcoxon NS)。

#### [0186] 在小鼠中检测 TIAB

[0187] 为了将 TIAB 的概念扩展到其它哺乳动物并证实 TIAB 检测由癌症引起,我们寻求将该观察结果转换到小鼠模型。首先,我们选择了能够有效辨别 NSCLC 患者与对照的 5 种 PAP。如图 6A 中所示,它们中的三种源自于在小鼠和人类之间在基因组水平上高度保守的基因。最重要的是,可能受影响的碱基在两个物种中是相同的,对于上游的 4 个碱基和下游的 2 个碱基来说情况也是这样。以前,我们已显示了这种短的 DNA 邻近序列允许 TI 事件发生的重要性<sup>34</sup>。另外 2 种所选的 PAP 源自于在小鼠和人类之间不保守的基因。可以产生 PAP 9 的基因在人类中存在,但是在鼠类基因组中不存在。在小鼠中,产生 PAP 2 翻译的基因在预定位置处发生 SBG,在编码 7 个氨基酸后引入了终止密码子。还包含了对应于 PAP 7 的 CP 的附加阴性对照。将具有免疫活性 (C57B16) 的小鼠 (n = 12) 用小鼠 Lewis 肺癌 (LLC1) 皮下接种<sup>45,46</sup>。在 LLC1 移植之前以及以一周的间隔直到 21 天,测量与 5 种 PAP 和一种 CP 结合的 IgG。这时,平均肿瘤尺寸是 3.15+/-0.4cm<sup>3</sup>。如图 6B 中所示,在肿瘤植入后 2 周,针对 PAP 7、PAP 48 和 PAP 62 的 IgG 显著增加。对于 PAP 7、PAP 48 和 PAP 62 来说,成对 t-检验的 p 值分别为 1\*10<sup>-4</sup>、3\*10<sup>-4</sup>、9\*10<sup>-4</sup>。我们没有观察到针对 CP7 的 IgG 水平的显著增加。

#### [0188] 材料与方法

##### [0189] 细胞培养

[0190] 鼠类 Lewis 肺癌细胞系 (LLC1) 从美国典型培养物保藏中心 (American Type Culture Collection (ATCC)) 获得。将细胞在含有补充了 10% FBS、链霉素 (0.1mg/ml) 和青霉素 (100 单位 /ml) 的 RPMI1640 培养基 (Invitrogen, France) 的 75cm<sup>2</sup> 摇瓶中培养,维持在 37°C 下、含有 5% CO<sub>2</sub> 的潮湿气氛中。

##### [0191] 肿瘤移植

[0192] 将 LLC1 肿瘤细胞 (5\*10<sup>5</sup> 个细胞在终体积为 0.1ml 的 RPMI 1640 中) 皮下 (s. c.) 注射到 7 周龄 C57b1/6 雌性小鼠 (Janvier, France) 的右后臀腿区域中。在 s. c. 注射 LLC1 细胞后 21 天,通过测量每个肿瘤的二等分直径,并使用公式  $V = a^2 * b * 0.5236$  计算,其中“a”是较大直径,“b”是较小直径,来测量肿瘤体积。在 s. c. 注射肿瘤细胞之前,在异氟烷麻醉下取 100 μl 血样,作为 T0。每周一次在异氟烷麻醉下取 100 μl 血样到 EDTA 管中,作为 T7、14 和 21 天。

##### [0193] 临床验证

[0194] 为了验证本发明的 PAP 组合能够产生强有力的肺癌诊断这个事实,我们进行了大规模的回顾性病例对照研究,该研究包括 161 位对照对象,他们是健康的献血者,年龄为 18 到 65 岁,没有排除吸烟者。患者是 140 位患有早期非小细胞肺癌的个体。在诊断时从这些患者收集血液。该组中所有患者都符合外科干预标准,因此对于绝大多数来说是早期阶段。该组中所有患者经历了外科手术,因此对所有患者都获得了手术后分级和病理分类。必须强调,该组中的患者没有接受手术前化疗或放疗。对照和患者的临床特征由表 4 的研究 III

显示。测试了患者和对照针对 6 种 PAP 的 TIAB, 测量在特定实验条件下进行。这种标记物组合的诊断值的统计学分析, 使用支持向量机进行确定。在分析了可选分类方法的性能后, 保留 SVM。SVM 确定了 6 维超平面, 为对照和患者到该超平面的个体距离提供了度量。因此, 数据显示成每个对象到超平面的相对距离。可以看到, 癌症患者和对照这两个群体分离良好 (图 7A)。总体测试性能为灵敏度 86%, 特异性 97%。只有 5 个对照在超平面的错误一侧。19 位患有肺癌的患者在超平面的错误一侧。在反复交叉验证后, 灵敏度和特异性分别为 82% 和 95%。在较年轻和较年老患者之间, 灵敏度没有差异 (图 7B)。从数据的检查可以看出, 当前测试的灵敏度对于腺癌患者来说较低, 它们的诊断灵敏度是 76%, 相形之下, 对于其它类型的非小细胞癌来说, 测试性能是灵敏度 > 90%。测试性能的这些差异是统计学显著的 (图 7C)。

[0195] 外科手术对肺癌的好处已经充分确立。这种好处不仅表现为预期寿命增加, 而且表现为确定的治愈, 这可以通过诊断数量与死亡数量的分析来确定。在法国, 这些数字是  $\approx 33000$  例新诊断和  $\approx 28000$  例死亡。因此, 可以可靠地确定 5000 位患者从肺癌治愈, 即 15%。所有的肺癌存活者都经历了外科手术程序, 但不是经历了外科手术的所有肺癌患者都被治愈。目前, 没有方法能够把将会得益于单独的外科手术的个体与不会从中得益的个体区分开。因此, 我们寻求就诊断时预测的严重性方面评估该基于 6 种 PAP 的测试的性能。为了实现这一点, 我们将研究群体细分成 2 个组。在第一个组中是我们有证据记录到无病存活超过 36 个月的患者。在第二个组中, 患者在外科手术后 36 个月内发生了死亡或疾病复发。我们没有包括对其跟踪的时间短于 36 个月的患者, 仅仅因为这些个体的跟踪时间太短而不能确定结果。然后, 我们比较了这两个规模相似的患者组到超平面的距离。从图 7D 的数据可以明显看出, 强烈得益于外科手术并且 3 年无病的患者, 与从外科手术得益较少的患者相比, 具有明显更长的到超平面的距离 ( $P = 0.005$ )。该发现带来了 2 个即刻应用。首先, 有可能该测试将鉴定患有早期肺癌并将从外科手术获得最有利的益处的患者。其次, 对于有阳性测试诊断但到超平面的距离短的患者, 应该适当设置合适的替代治疗干预。这样的替代措施可以是常规的放疗或化疗。然而, 我们目前对数据的解释是对肺癌的存在具有低免疫应答的患者, 不太可能得益于单独的外科手术。就此而言, 在外科手术前或之后的短时间内, 使用药物增强它们的免疫应答, 可能是有用的。因此, 我们在这里示例了发现 PAP 的重要性, 作为将来创新性治疗策略的指导。

[0196] 接下来, 我们通过特定实验条件下测试各种不同的 PAP 组合, 试图发展肺癌特异性的测试。一种新的 PAP 组合对肺癌与对照相比, 获得了 100% 的特异性和 90% 的灵敏度。最重要的是, 当应用于患有乳腺癌的患者时, 在 20 位患有乳腺癌的患者中只有 3 位显示出阳性测试。因此, 在肺癌对乳腺癌的比较中, 分别有 10% 和 15% 的患者被错误分类。我们认为, 该误差率将导致不必要的和昂贵的下游诊断程序。对肺癌的跟踪诊断是胸部 CT 扫描, 而对乳腺癌是乳房造影术和回声图像术。PET 扫描可用于评估疾病的扩散。因此, 我们开发了能够有效区分肺癌与乳腺癌的特异性测试。发现有 PAP 的 4 种组合可以实现该目的。这些标志物的组合示例在图 8B 中。在所有 4 种情况下, 肺癌患者与乳腺癌患者被区分开, 其中有一位到三位患者重叠。因为这些标记物组合在疾病严重性方面的预测价值需要进一步评估, 因此介绍了它们的临床重要性。我们的确预测到, 与对肺所示例的相似, 特定的 PAP 组合将在大规模研究中显示出对乳腺癌的临床结果的预测。在能够实现这一点之

前,需要特异性乳腺癌测试。目前我们已经鉴定到 3 种标志物的组合,所述组合在特定条件下以 60% 的灵敏度和 95% 的特异性从对照中鉴定出乳腺癌(图 8C)。我们现在的观点是,鉴定到 PAP 的组合,其能够从简单的血液测试指示在最常见癌症的早期阶段存在,将是可能的。其次,PAP 的组合将为疾病的精确定位提供准确的指示,从而允许选择适合的次级诊断程序,例如 CT 扫描、乳房造影术、超声扫描术、纤维镜检查术、内窥镜检查法、活组织检查。PAP 测试的第三个方面将为每个个体对外科手术治疗的响应提供预后,从而为需要其它治疗措施提供指示。PAP 的第四个方面将为监测疾病的复发和 / 或其对治疗的有利响应提供工具。

#### [0197] C- 讨论

[0198] 我们已经鉴定了触发虽然弱但是多样化的 IgG 生产的人类癌细胞蛋白异质性的新的、可预测的来源。这些低滴度的特异性 IgG 的精确测定为早期癌症诊断提供了有前景的机会,并能够提供与疾病扩散有关的信息。该发现源自于生物信息学预测与针对异常肽的特异性 IgG 的免疫学检测的趋同 (convergence)。含有 PAP 序列的蛋白不能从正常人类 mRNA 翻译、而是只能从由于单碱基缺口而失去其规范基因组信息的 mRNA 翻译。

[0199] 我们的数据支持了下述结论,即癌细胞中存在预测的带有单个缺口的 mRNA,并且它们至少部分翻译。鉴定到的 EST 缺口的出现太常见而不会是通过癌症体细胞突变产生的<sup>31,32</sup>。目前对癌症体细胞突变率的估算导致预测了 12 种缺失,其中有 4 种是单个缺口。相反,我们观察了 2206 个统计学显著的事件。此外,没有一个所选的缺口对应于推断的或生物学证实的 SNPs<sup>44</sup>。因此,mRNA 缺口不可能在基因组水平产生,因此一定发生在下游,即转录期间或转录后不久。

[0200] 已经确立,成熟前和成熟 mRNA 的碱基可以被酶修饰,但是还没有已知的人类 mRNA 编辑酶显示出能够从单链人类 RNA 中除去单个碱基<sup>47</sup>。已经显示,锥虫 (Trypanosoma) 线粒体 mRNA 编辑体 (editosome) 能够进行特异性 U 缺失<sup>48</sup>。但是,没有在人类基因组中发现锥虫和利什曼原虫 (Leishmania) 编辑体蛋白的同源物。此外,这种编辑机制是 U 特异性的,不能解释观察到的 53% 的人类癌症非 U 缺口。我们已经考虑到了剪接机制的向前或向后滑动可能引入将会影响外显子的最后或第一个碱基的单碱基缺口的可能性<sup>49</sup>。没有被测试的缺口位于这样的位置上。此外,后一种机制被发现是不可能的,因为超过 2000 个被鉴定的 SBG 中 99.2% 不在紧邻外显子 - 内含子边界内。最后,显然,短的 DNA 邻近序列对癌症 EST 缺口的出现施加了强烈影响,这与对 EST 碱基取代所证明的相似<sup>34,36,37</sup>。因此,我们目前假设,由 Pol II 引起的遗漏单个碱基掺入、即 TI,导致了有缺口的 mRNA 的出现。

[0201] 我们进一步示例了在预定位置处出现具有简单缺口的 mRNA,在不存在体细胞或生殖细胞突变情况下在癌细胞中发生。因此,转录失真性的生物信息学概念在人类中得到生物学验证。

[0202] 本文报道的发现的第二个最重要的结果是人类基因组的规范阅读不足以解释癌细胞的异质性。因此我们提出在癌细胞中转录失真性增加。支持这种提议的证据如下。首先,我们以前已经通过 DHPLC 显示,癌细胞的 mRNA 与从正常细胞分离到的相比,更为异质<sup>34</sup>。癌症与正常 mRNA 相比序列变异增加,这由依靠 SAGE 实验的独立研究所证实。其次,对所有可用的人类 mRNA 衍生序列的分析显示出,相对于正常的文库,在癌症中碱基取代、插入和缺口 (SBG) 有统计学显著增加。这些事件的发生比癌症体细胞突变的发生更常见 10<sup>3</sup>

倍。如果出现在 DNA 水平上,这种突变率很可能是致死性的。因此,可以合理地假定,这些变异发生在转录期间或转录后不久,并只影响成熟前和成熟 mRNA,即瞬时分子。第三,目前还不知道 TI 之外的其它分子机制能够从 mRNA 除去或添加单个碱基并然后重新组装序列。因此,在 DNA 水平上不存在突变的情况下直接观察到在人类肺癌中出现的预测的 SBG,表明 RNAP 可以跳过阅读单个 DNA 碱基并仍然继续前进。最后,来自其他组的研究显示,即使不存在癌症, TI 也能在体内发生。具体来说,在 Brattleboro 大鼠中,在 GAGAG 序列中发生的 GA 缺失使血管加压素转录本恢复正常,从而抑制了尿崩症。影响  $\beta$ -淀粉样肽和泛素 B 的重复 A 序列的转录移码,产生了可以通过阿兹海默氏病斑的免疫染色检测到的蛋白。

[0203] 目前,我们倾向于下述假说,即癌症 mRNA 的异质性增加是癌发生的结果而不是原因。的确,我们正在发生间变性大细胞淋巴瘤并带有间变性淋巴瘤激酶 (ALK) 易位的儿童的血清中,检测针对我们现在的 PAP 组的特异性 IgG(G Delsol 和 B Bihain,未发表的结果)<sup>50</sup>。对啮齿动物的研究已经证实,引起该激酶的组成性表达的易位是最初的致癌事件,其单独就足以引起转化<sup>51</sup>。因此,检测到反映出编码功能不相关基因——即不是该特异性易位的一部分的基因——的异常 mRNA 的产生的阳性信号,表明该现象作为这种致癌病变的结果而发生。但是,有可能 TI 有助于加速癌发生。事实上,参与转录、翻译和 DNA 修复的调控的几个基因,因为其推断的基因功能不是 PAP 选择过程的一部分而没有包括在本研究中,它们通过生物信息学被鉴定为具有 K 缺口。因此,可能我们面临着多样性和强度随着疾病发展的严重性而增加的自身催化过程。

[0204] 生物信息学表明, mRNA 中 SBG 的出现是癌症的常见特征。但是,在淋巴瘤患者 (N = 27) 中也发现了 IgG 分布情况的差异。在 NSCLC 中通常为阳性的 PAP 1 和 2 在滤泡性和间变性淋巴瘤患者中都是阴性的。这与 PAP 4 和 7 相反,它们在间变性大细胞淋巴瘤中通常是阳性的,但是在滤泡性淋巴瘤中不是。因此,由于可用的 PAP 组的多样性 (> 2000 个候选物),我们提出了设计肿瘤特异性 PAP 组。通过证明 PAP 有效地将肺癌患者与乳腺癌患者分开的能力,我们已经示例了这种概念。

[0205] 我们的结论是,具有单碱基缺口的 mRNA 至少部分翻译成异常蛋白,表明在癌细胞中无义介导的 mRNA 降解可能是有缺陷的<sup>52</sup>。考虑到当前的蛋白质组学工作,这种高度多样化的异常蛋白组还迄今保持未被检测到,是令人吃惊的。解释有两方面。1) 通过质谱进行的蛋白质鉴定依赖于与已知或推断的氨基酸序列所定义的预测谱图所观察到的匹配<sup>53</sup>。从 mRNA 缺口产生的异常蛋白的氨基酸序列不在当前的蛋白数据库 (Swiss-Prot/TrEMBL<sup>42</sup>) 中,因此不能通过 MS/MS 分析进行鉴定。2) 蛋白酶体快速降解异常蛋白,产生了可能异常的免疫原性肽<sup>54</sup>

[0206] TI 在癌症中增加的概念,引起了对目前的癌症生物标志物发现策略的质询,并导致提出了新方法。系统性癌症蛋白质组学方法产生了冲突的结果,分歧被归因于分析前条件的变化。情况可能确实如此,但现在必须考虑可选的解释。如果人们接受癌细胞蛋白的异质性大大超过当前的估计,那么有可能样品量不足以完全探测蛋白变体的高度多样化的所有组成部分。正如前面提到的,目前的蛋白质组学的另一个限制是,质谱数据目前用人类基因组的规范阅读进行解释。因此,具有异常氨基酸序列的蛋白可能逃脱了适合的鉴定。因此,在癌症蛋白质组学成功之前,可能需要不仅是方法上、而且是概念上的改变。通过将本发明的转录失真性考虑在内,可以鉴定更可靠和相关的生物标志物。

[0207] 我们已经显示,在小鼠癌细胞中有三种异常蛋白由高度保守但功能不相关的基因编码。在 LLC1 中丰度最高的异常蛋白是源自于 Poly(A) 结合蛋白胞质 1 (PABPC1) 的异常蛋白 (PAP 62)。正常情况下, PABPC1 与 mRNA poly A 结合,并调节无义介导的降解 (NMD) 途径,该途径降解含有提前终止的 mRNA。将 PABPC1 束缚在提前终止密码子的下游废止了 NMD。在 LLC1 中丰度第二高的异常蛋白由波形蛋白基因 (VIM) 编码 (PAP 48)。波形蛋白是 III 型中间纤维蛋白,形成同源或异源聚合结构,有助于支持细胞膜、将核和细胞器保持在限定位置、以及与微管结合。丰度第三高的异常蛋白是由 IK 基因编码的异常蛋白 (PAP 7)。IK 的正常功能是抑制  $\gamma$  干扰素诱导的 II 类主要组织相容性复合物的表达的细胞因子的功能。IK 也被鉴定为软骨肉瘤相关蛋白 2。这些变体在癌细胞中存在的后果目前还不知道。但是,绝不能排除强烈干扰癌细胞生物学的可能性,它们对癌细胞代谢、形态学改变以及 mRNA 异质性的贡献将需要进一步研究。在目前阶段,我们已经能够确定,大多数 PAP 调节对人类中的 NSCL 癌和小鼠中的 LLC1 的体液免疫应答。我们预计,这些异常蛋白的癌细胞生产可能通过显性阴性或阳性效应明显改变细胞的功能。因此,这三种高度保守的基因可能提供了新的治疗靶。

[0208] 小鼠肺癌模型分析在 LLC1 的存在和抗 PAP IgG 检测之间建立了因果关系。因此,抗 PAP IgG 似乎是由癌症引发的正常的、适时的免疫应答的一部分。有趣的是,在小鼠中,抗 PAP IgG 水平比在患有肺癌的人类中测量到的高得多 (100 倍)。在小鼠中肺部肿瘤的相对尺寸也大得多,并且 LLC1 被异位植入到皮下组织中,这些事实为这些差异提供了可能的解释。

[0209] 因此,本发明描述了新的机制,癌症通过该机制调节体液免疫应答。在目前阶段,我们提出,新的机制 TI 造成了癌细胞 mRNA 的异质性的急剧增加,这些异常信息的一部分翻译成异常蛋白,一些异常蛋白在癌细胞中积累,大多数异常蛋白调节癌症体液免疫应答。因此,本发明提供的产品和方法允许正确地区分患有癌症的患者与未患活动癌症的患者。因此,有可能为处于风险的个体精心设计系统性生物化学筛查,对具有阳性测试的患者进行全身成像,并增加早期阶段诊断的对象的比例。

[0210] 表 1. 统计学检验的结果

[0211]

	K	LBE	N	LBE	K/N	(K-LBE)/(N-LBE)
缺口	2,761	11	216	144	12.78	38.19
ORF 内的缺口	2.191		162		13.52	
取代	1,894	92	928	186	2.04	2.43

[0212] LBE 是指基于假阳性率估算量的位置。

[0213] 表 2. 生物信息学分析的结果

	正常	癌症
从 NCBI 检索到的 EST	3,949,323	3,043,498
具有 EST 匹配的转录本的数量	34,974	34,788
具有 EST 匹配的转录本的数量		33,111
[0214] 分析的核苷酸		88,372,747
由 >70 EST 所定义的位置		2,829,135
匹配统计学限制的位置		
取代		5,784
缺口		3,790

[0215] 分析结果在检索了从 2000 年 1 月到 2007 年 7 月发布到未过滤公共数据库中的所有可用人类 ESTs 后得出。表中还显示了匹配第一（有效性 > 70）的第二次统计学检验标准的位置的数量。

[0216] 表 3. TI 肽和核酸序列

[0217] 表 3a :2206 个缺口 TI 肽的核酸和氨基酸序列。在本文件中指称的 SEQ ID NOs 1-2206 表示在表 3a 的第 6 列中描述的肽序列。

[0218] 表 3b :1128 个插入 TI 肽的核酸和氨基酸序列。在本文件中指称的 SEQ ID NOs 2207-3334 表示在表 3b 的第 6 列中描述的肽序列。

[0219] 表 4. 对照个体和癌症对象的临床数据。

[0220]

	数量	年龄 (岁)	女性	男性	阶段*
健康对照	26	55 ± 18	13	13	-
慢性阻塞性肺病	12	55 ± 11	2	10	-
所有癌症 研究 I	46	61 ± 11	26	20	
结肠	9	65 ± 13	1	8	T+N0M0 → T+N+M+
肺 (7 NSCLC + 2 SCLC)	9	67 ± 7	2	7	T+N0M0 → T+N+M+
乳腺	9	60 ± 11	9	0	T+N0M0 → T+N+M+
卵巢	4	58 ± 6	4	0	T+N0M0 → T+N+M+
子宫	5	50 ± 8	5	0	T+N0M0 → T+N+M0
头颈部	7	60 ± 11	3	4	T+N0M0 → T+N+M+
黑素瘤	3	58 ± 14	2	1	T+N+M0 → T+N+M+
肺癌 研究 II (NSCLC)	49	67 ± 13	10	39	
	10	66 ± 14	3	7	N0M0
	25	68 ± 14	4	21	N+M0
	14	67 ± 10	3	11	N+M1

[0221] ● 国际抗癌联合会 (International Union Against Cancer) (UICC):《恶性肿瘤的 TNM 分类》(TNM Classification of malignant tumours), 第四版, Hermanek P, Sobin LH 主编, Berlin, Heidelberg, New York:Springer Verlag;1987. 1992 年修订。

[0222] 研究 III

[0223]

	n	%	年龄 (y)	最小	最大
对照	161		42 ± 14	18	65
NSCLC	140		61 ± 11	38	86
T1-2-3-4N0M0	78	56	61	38	86
T+N+M0	43	31	69	44	86

T+N+M+	19	13	61	46	77
ADK	67	48	60	45	82
鳞状	40	33	65	48	86
其他	33	19	58	38	82

[0224] 研究 IV

[0225]

	n	%	年龄 (y)	最小	最大
对照	20		47±13	20	62
NSCLC	20		61±10	48	83
T1N0M0	9	45	60±7	49	70
T2N0M0	11	55	62±12	48	83
乳腺	20		56±10	36	68
1 级水平	4	20	60±6	53	67
2 级水平	9	45	56±10	40	68
3 级水平	7	35	54±11	36	68

[0226] 表 5. 45 种 PAP 多肽的特征

基因	登记号	位置	肽的长度	肽的偏差	%偏差	肽	PAP 编号	SEQ ID NO	在SEQ ID 上的PAP坐标
MRPL12	NM_002949.2	687	30	3.85	17.74	WRRWAAPWFSSLLQGLVFRGPPRARSR	1	1	1-30
DECR1	NM_001359.1	257	35	9.30	41.36	NLNSFHLEKRCYHLIVFKERWHSLLEVLALVRE	2	2	1-35
ALDOA	NM_184041.1	602	25	0.42	4.26	GWMMGLSAVPSRTRKTELTPSPGVVC	3	3	1-25
COX4I1	NM_001861.2	549	18	2.43	9.05	RKALTRKSGWPSRFRGCWT	4	4	1-18
TP11	NM_000365.4	149	13	1.59	6.59	LPLLPSTSPGRS	5	5	1-13
ENO1	NM_001428.2	546	17	1.13	6.23	SLTWLALTKSSCQSRRS	6	6	1-17
IK	NM_006083.3	614	19	0.00	15.36	LCFKRYELRLPAKRKRNRN	7	7	1-19
LY2	NM_000239.1	78	12	8.56	41.03	ARSLKGVSWPEL	8	8	1-12
PRR4	NM_007244.2	215	30	0.00	30.14	VIVVTKMIVLSDHQNEAITAILPHLLFK	9	177D	2-31
GCHN1	NM_031966.2	598	27	0.93	22.73	LILPLAQWKLHDLVPLQKTCVRLSLM	10	10	1-27
CRABP2	NM_001878.2	537	37	2.00	18.62	STSESEWQVEPRPKPTTGHAFRPAFLPPSHPLL	11	11	1-37
HSPA8	NM_153201.1	165	20	0.09	2.23	PMIRETEFLQAMSPRLTNG	12	12	1-20
LCP1	NM_002298.2	227	16	1.31	18.89	LPKLLMAMDTASMS	13	13	1-16
PSMD13	NM_002817.2	223	21	1.60	12.57	LPKEMVLSFMKTLNVNLTNG	14	14	1-21
IH	NM_000143.2	141	16	0.00	21.13	FGLRTRLEWQAKPFG	15	15	1-16
GPI	NM_000175.2	1255	16	5.56	33.04	PMASMLFTSSSTKAPR	20	44	1-16
ACQ2	NM_001098.2	623	12	5.17	20.21	MLWMSWLSPGSG	21	53Z	19-21
NDUFB5	NM_002492.2	194	13	4.84	15.08	LSSDLLDSMTGVF	22	1063	1-13
NDUFB3	NM_004551.1	195	23	4.41	17.95	LESMWLKSQPSMSNRFCPASMS	23	198D	1-23
NDUFA51	NM_005003.2	366	19	2.07	19.08	WTKWRLSWPWPNLGLKFL	24	1413	1-19
ECH1	NM_001398.2	871	30	1.59	14.61	RFPARAFWRCPARSTCCPATIRWPRAST	25	581	1-30
NEM1	NM_002520.5	589	11	8.33	27.59	LEVVARFERKK	26	1075	1-11
ECHS1	NM_004092.2	145	22	5.13	12.41	SPRLVLTSTSSOKKEGRITPWG	27	1390	1-22
CFL1	NM_005507.2	447	30	0.48	4.31	LPRCCQRTAAMPMDQMLRPRRARRRWC	28	1467	1-30
MRPL3	NM_007208.2	275	30	1.76	12.38	LLEVFMERVVHGGMSIFLKKMSHSLSSWSL	29	1769	1-30
CYCS	NM_018947.4	351	12	8.78	35.77	WRIPRSTSLQK	35	1972	1-12
E2M	NM_004048.2	193	15	2.58	10.85	AMCLGFIHTLKLTY	37	1323	1-15
ILF2	NM_004515.2	360	19	1.92	15.81	HLKCKLRKFDWRWPKRGG	38	1373	1-19
BCAP31	NM_005745.6	184	27	1.54	8.72	LCCFSAFPSFLKDGRRFSSPGWWSW	39	1499	1-27
PHB	NM_002634.2	703	12	11.69	20.41	SSLLRATPROLS	40	1098	1-12
EGR1	NM_000291.2	587	11	0.39	5.71	LALLTEPTAPW	41	57	1-11
UCRC1	NM_003265.2	323	30	1.44	10.55	WSTWLSRQRIQLAVPWRRWRWAWGPIIMP	44	1270	1-30
NDUFB1	NM_007163.2	657	28	2.47	11.84	LWCAGLQPTSVERRQRSSPLRASRASA	46	1767	1-28
ATF4	NM_001675.2	1387	11	5.50	17.22	SCPLQQLV	47	864	1-11
VIM	NM_003380.2	652	30	0.00	2.49	TWPRTSCASGRNCRRCFRERKPKTPCNLS	48	1276	2-31
FTTG1	NM_004219.2	296	11	1.82	19.77	WELSTELQSL	49	1336	1-11
FABP1	NM_002568.3	1293	19	1.16	4.17	VELRKRWNQRNLSANLNR	62	1079	1-19
CDKN1A	NM_000389.2	378	30	0	6.09	LAPHLCCRGQCQRRTMWTCHLVLPLCLAQG	70	80	1-30
BRDX6	NM_004905.2	322	30	0.31	3.66	VKSPQKSLYFPSSMIGSLPSCWACWQQ	66	1906	1-30
RPL13A	NM_012423.2	355	17	0.63	4.88	TTRKSGWVWLLPFSKCV	68	1810	1-17
APFX1	NM_001641.2	418	30	0.81	8.72	QKMKTRKQGERAQCMRTPQIRKPHVANL	74	847	1-30
MLF2	NM_005439.1	343	31	2.27	12.64	PGLFAAGCSRLLELSPPLGCVCKRVVSWTCLG	82	1451	1-31
TUBB	NM_178014.2	370	30	0.40	9.76	LARSLDQTLTYLVSLGOVITGPKATTQRAP	94	2119	1-30
LAPTM4A	NM_014713.3	758	21	2.81	7.08	MLCPLPLKHLSTFCQPMKWP	69	1854	1-21
CCT8	NM_006585.2	536	22	0.21	6.70	WYVVLQKTFELMKSHLYFVPP	86	1899	1-22

[0227]

[0228] “PAP 编号”是在本文本中指称的 PAP 的编号。

[0229] “SEQ ID NO”是序列表中的标识符。

[0230] “SEQ ID 上的 PAP 坐标”是指 PAP 多肽的氨基酸残基在引用的 SEQ ID 中的位置。

[0231] 表 6. 5 个阴性对照

[0232]

符号	基因名	序列
	ALB	QHKDDNPNL PRLVVP
基因 1 上的 CP	MRP L12	IQQLVQDIASLTLLLEISDLNELLKKTLLKIQ
CP7	IK	ALLQKVRAEIASKEKEEEE
CP28	CFL1	FVKMLPDKDCRYALYDATYETKESKKEDLV
CP24	NDUFB1	LDQVEIIMAMEDEFGFEIP

[0233] 参考文献

[0234] 1. Weinberg, R. in The biology of cancer 655-724(《癌症的生物学》), 655-724(Garland Science, Taylor and Francis Group, LLC, 2007).

[0235] 2. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. (癌症基因和它们控制的途径) Nat Med 10, 789-99 (2004).

[0236] 3. ASCO. American Society of Clinical Oncology policy statement update: genetic testing for cancer susceptibility. (美国临床肿瘤学学会最新政策陈述: 用

于癌症易感性的遗传测试) *J Clin Oncol* 21, 2397-406 (2003).

[0237] 4. Fackenthal, J. D. & Olopade, O. I. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. (在多种多样的群体中与BRCA1和BRCA2相关的乳腺癌风险) *Nat Rev Cancer* 7, 937-48 (2007).

[0238] 5. Guillem, J. G. 等, ASCO/SSO review of current role of risk-reducing surgery in common hereditary cancer syndromes. (ASCO/SSO对当前降低风险的外科手术在常见遗传性癌症综合中的作用的评述) *J Clin Oncol* 24, 4642-60 (2006).

[0239] 6. van de Vijver, M. J. 等, A gene-expression signature as a predictor of survival in breast cancer. (基因表达特征作为乳腺癌存活性的预测因子) *N Engl J Med* 347, 1999-2009 (2002).

[0240] 7. Anderson, N. L. 等, The human plasma proteome: a nonredundant list developed by combination of four separate sources. (人类血浆蛋白质组: 通过四种独立来源的组合产生的非冗余名单) *MolCell Proteomics* 3, 311-26 (2004).

[0241] 8. Wulfschlegel, J. D., Liotta, L. A. & Petricoin, E. F. Proteomic applications for the early detection of cancer. (蛋白质组学用于癌症早期检测) *Nat Rev Cancer* 3, 267-75 (2003).

[0242] 9. Ishikawa, N. 等, ADAM8 as a novel serological and histochemical marker for lung cancer. (ADAM8作为肺癌的新的血清学和组织化学标志物) *Clin Cancer Res* 10, 8363-70 (2004).

[0243] 10. Ransohoff, D. F. Rules of evidence for cancer molecular-marker discovery and validation. (用于癌症分子标志物的发现和验证的证据的规则) *Nat Rev Cancer* 4, 309-14 (2004).

[0244] 11. Ransohoff, D. F. Bias as a threat to the validity of cancer molecular-marker research. (偏差是对癌症分子标志物研究的有效性的威胁) *Nat Rev Cancer* 5, 142-9 (2005).

[0245] 12. Stroun, M. 等, Neoplastic characteristics of the DNA found in the plasma of cancer patients. (在癌症患者血浆中发现的DNA的肿瘤特征) *Oncology* 46, 318-22 (1989).

[0246] 13. Boddy, J. L., Gal, S., Malone, P. R., Harris, A. L. & Wainscoat, J. S. Prospective study of quantitation of plasma DNA levels in the diagnosis of malignant versus benign prostate disease. (恶性对良性前列腺疾病诊断中血浆DNA水平定量的前瞻性研究) *Clin Cancer Res* 11, 1394-9 (2005).

[0247] 14. Boddy, J. L. 等, The role of cell-free DNA size distribution in the management of prostate cancer. (无细胞DNA尺寸分布在前列腺癌管理中的作用) *Oncol Res* 16, 35-41 (2006).

[0248] 15. Lund, A. H. & van Lohuizen, M. Epigenetics and cancer. (表观遗传学与癌症) *Genes Dev* 18, 2315-35 (2004).

[0249] 16. Ducasse, M. & Brown, M. A. Epigenetic aberrations and cancer. (表观遗传异常与癌症) *Mol Cancer* 5, 60 (2006).

[0250] 17. Goessl, C. 等, Fluorescent methylation-specific polymerase chain reaction for DNA-based detection of prostate cancer in bodily fluids. (用于在体液中基于 DNA 检测前列腺癌的荧光甲基化特异性聚合酶链反应) *Cancer Res* 60, 5941-5(2000).

[0251] 18. Jeronimo, C. 等, Quantitative GSTP1 hypermethylation in bodily fluids of patients with prostate cancer. (患有前列腺癌患者的体液中的定量 GSTP1 过甲基化) *Urology* 60, 1131-5(2002).

[0252] 19. Reibenwein, J. 等, Promoter hypermethylation of GSTP1, AR, and 14-3-3sigma in serum of prostate cancer patients and its clinical relevance. (前列腺癌患者血清中 GSTP1、AR 和 14-3-3 $\sigma$  的启动子过甲基化及其临床关联性) *Prostate* 67, 427-32(2007).

[0253] 20. Wang, Y. 等, Identification of epigenetic aberrant promoter methylation of RASSF1A in serum DNA and its clinicopathological significance in lung cancer. (在血清 DNA 中鉴定 RASSF1A 的表观遗传异常的启动子甲基化及其在肺癌中的临床病理学重要性) *Lung Cancer* 56, 289-94(2007).

[0254] 21. Diehl, F. 等, Detection and quantification of mutations in the plasma of patients with colorectal tumors. (在患有结肠直肠癌患者的血浆中检测和定量突变) *Proc Natl Acad Sci U S A* 102, 16368-73(2005).

[0255] 22. Korshunova, Y. 等, Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. (大规模平行亚硫酸氢盐焦磷酸测序揭示出从组织和血清 DNA 获得的乳腺癌相关性胞嘧啶甲基化样式的分子复杂性) *Genome Res*(2007).

[0256] 23. Bentley, D. R. Whole-genome re-sequencing. (全基因组重新测序) *Curr Opin Genet Dev* 16, 545-52(2006).

[0257] 24. Meyer, M., Stenzel, U., Myles, S., Prufer, K. & Hofreiter, M. Targeted high-throughput sequencing of tagged nucleic acid samples. (带标签核酸样品的定向高通量测序) *Nucleic Acids Res* 35, e97(2007).

[0258] 25. Tan, E. M. Autoantibodies as reporters identifying aberrant cellular mechanisms in tumorigenesis. (自身抗体作为鉴定肿瘤发生中的异常细胞机制的报告因子) *J Clin Invest* 108, 1411-5(2001).

[0259] 26. Finn, O. J. Immune response as a biomarker for cancer detection and a lot more. (免疫应答作为癌症检测以及许多其他的生物标志物) *N Engl J Med* 353, 1288-90(2005).

[0260] 27. Zinkernagel, R. M. What is missing in immunology to understand immunity? (在免疫学中对于理解免疫性来说丢失了什么?) *Nat Immunol* 1, 181-5(2000).

[0261] 28. Wang, X. 等, Autoantibody signatures in prostate cancer. (前列腺癌中的自身抗体特征) *N Engl J Med* 353, 1224-35(2005).

- [0262] 29. Somers, V. A. 等, A panel of candidate tumor antigens in colorectal cancer revealed by the serological selection of a phage displayed cDNA expression library. (通过血清学筛选噬菌体展示的 cDNA 表达文库所显示的结肠直肠癌中的一组候选肿瘤抗原) *J Immunol* 169, 2772-80 (2002).
- [0263] 30. Hardouin, J., Lasserre, J. P., Sylvius, L., Joubert-Caron, R. & Caron, M. Cancer immunomics: from serological proteome analysis to multiple affinity protein profiling. (癌症免疫组学: 从血清蛋白质组学分析到多重亲和性蛋白质情况分析) *Ann N Y Acad Sci* 1107, 223-30 (2007).
- [0264] 31. Sjoblom, T. 等, The consensus coding sequences of human breast and colorectal cancers. (人类乳腺和结肠直肠癌的保守编码序列) *Science* 314, 268-74 (2006).
- [0265] 32. Wood, L. D. 等, The genomic landscapes of human breast and colorectal cancers. (人类乳腺和结肠直肠癌的基因组学前景) *Science* 318, 1108-13 (2007).
- [0266] 33. Nelkin B, P. D., Robinson S, Small D, Vogelstein B. (Owens, A., Coffey, DS, Baylin, SB) 主编), 441-460 (Academic Press, New York, 1982).
- [0267] 34. Brulliard, M. 等, Nonrandom variations in human cancer ESTs indicate that mRNA heterogeneity increases during carcinogenesis. (人类癌症 EST 中的非随机变异表明癌发生过程中 mRNA 异质性的增加) *Proc Natl Acad Sci U S A* 104, 7522-7 (2007).
- [0268] 35. Armache, K. J., Kettenberger, H. & Cramer, P. The dynamic machinery of mRNA elongation. (mRNA 延伸的动态机械) *Curr Opin Struct Biol* 15, 197-203 (2005).
- [0269] 36. Kashkina, E. 等, Template misalignment in multisubunit RNA polymerases and transcription fidelity. (多亚基 RNA 聚合酶中的模板未对准和转录保真性) *Mol Cell* 24, 257-66 (2006).
- [0270] 37. Pomerantz, R. T., Temiakov, D., Anikin, M., Vassylyev, D. G. & McAllister, W. T. A mechanism of nucleotide misincorporation during transcription due to template-strand misalignment. (转录过程中由模板-链未对准引起的核苷酸错误掺入的机制) *Mol Cell* 24, 245-55 (2006).
- [0271] 38. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. (用于 DNA 序列比对的贪婪算法) *J Comput Biol* 7, 203-14 (2000).
- [0272] 39. Boguski, M. S., Lowe, T. M. & Tolstoshev, C. M. dbEST--database for "expressed sequence tags". (dbEST——用于“表达序列标签”的数据库) *Nat Genet* 4, 332-3 (1993).
- [0273] 40. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. (NCBI 参照序列 (RefSeq): 基因组、转录本和蛋白质的过滤过的非冗余序列数据库) *Nucleic Acids Res* 35, D61-5 (2007).
- [0274] 41. Dalmaso, C., Broet, P. Procédures d'estimation du falsediscovery

rate basées sur la distribution des degrés de signification. Journal de la Société Française de Statistiques 146(2005).

[0275] 42. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. (SWISS-PROT 蛋白质序列数据库及其在 1999 年的补充 TrEMBL) Nucleic Acids Res 27, 49-54(1999).

[0276] 43. Bamford, S. 等, The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. (COSMIC (癌症中体细胞突变的目录) 数据库和网站) Br J Cancer 91, 355-8(2004).

[0277] 44. Sherry, S. T. 等, dbSNP: the NCBI database of genetic variation. (dbSNP: NCBI 的遗传变异数据库) Nucleic Acids Res 29, 308-11(2001).

[0278] 45. Sharma, S. 等, T cell-derived IL-10 promotes lung cancer growth by suppressing both T cell and APC function. (T 细胞衍生的 IL-10 通过抑制 T 细胞和 APC 功能促进肺癌生长) J Immunol 163, 5020-8(1999).

[0279] 46. Bertram, J. S. & Janik, P. Establishment of a cloned line of Lewis Lung Carcinoma cells adapted to cell culture. (适用于细胞培养的 Lewis 肺癌细胞克隆系的建立) Cancer Lett 11, 63-73(1980).

[0280] 47. Gott, J. M. & Emeson, R. B. Functions and mechanisms of RNA editing. (RNA 编辑的功能和机制) Annu Rev Genet 34, 499-531(2000).

[0281] 48. Rogers, K., Gao, G. & Simpson, L. U-specific 3' -5' exoribonucleases involved in U-deletion RNA editing in trypanosomatid mitochondria. (U-特异性 3' -5' 外切核酸酶参与锥虫线粒体中 U 缺失的 RNA 编辑) J Biol Chem(2007).

[0282] 49. Lodish, H., Berk, A., Zipursky, L., Matsudaira, P., Baltimore, D., Darnell, J. in Molecular cell biology (《分子细胞生物学》) (Freeman 主编) 404-452(2000).

[0283] 50. Lamant, L. 等, Gene-expression profiling of systemic anaplastic large-cell lymphoma reveals differences based on ALK status and two distinct morphologic ALK+ subtypes. (系统性间变性大细胞淋巴瘤的基因表达情况分析显示出基于 ALK 状态的差异和两种不同形态的 ALK+ 亚型) Blood 109, 2156-64(2007).

[0284] 51. Chiarle, R. 等, NPM-ALK transgenic mice spontaneously develop T-cell lymphomas and plasma cell tumors. (NPM-ALK 转基因小鼠自发发生 T 细胞淋巴瘤和浆细胞肿瘤) Blood 101, 1919-27(2003).

[0285] 52. Wormington, M. Zero tolerance for nonsense: nonsense-mediated mRNA decay uses multiple degradation pathways. (对无义的零容忍: 无义介导的 mRNA 降解使用多种降解途径) Mol Cell 12, 536-8(2003).

[0286] 53. Biemann, K. Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. (通过串联质谱术和高能碰撞诱导解离对肽进行测序) Methods Enzymol 193, 455-79(1990).

[0287] 54. Glickman, M. H. & Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. (泛素-蛋白酶体蛋白水解途径: 为构建而破坏) Physiol Rev 82, 373-428(2002).

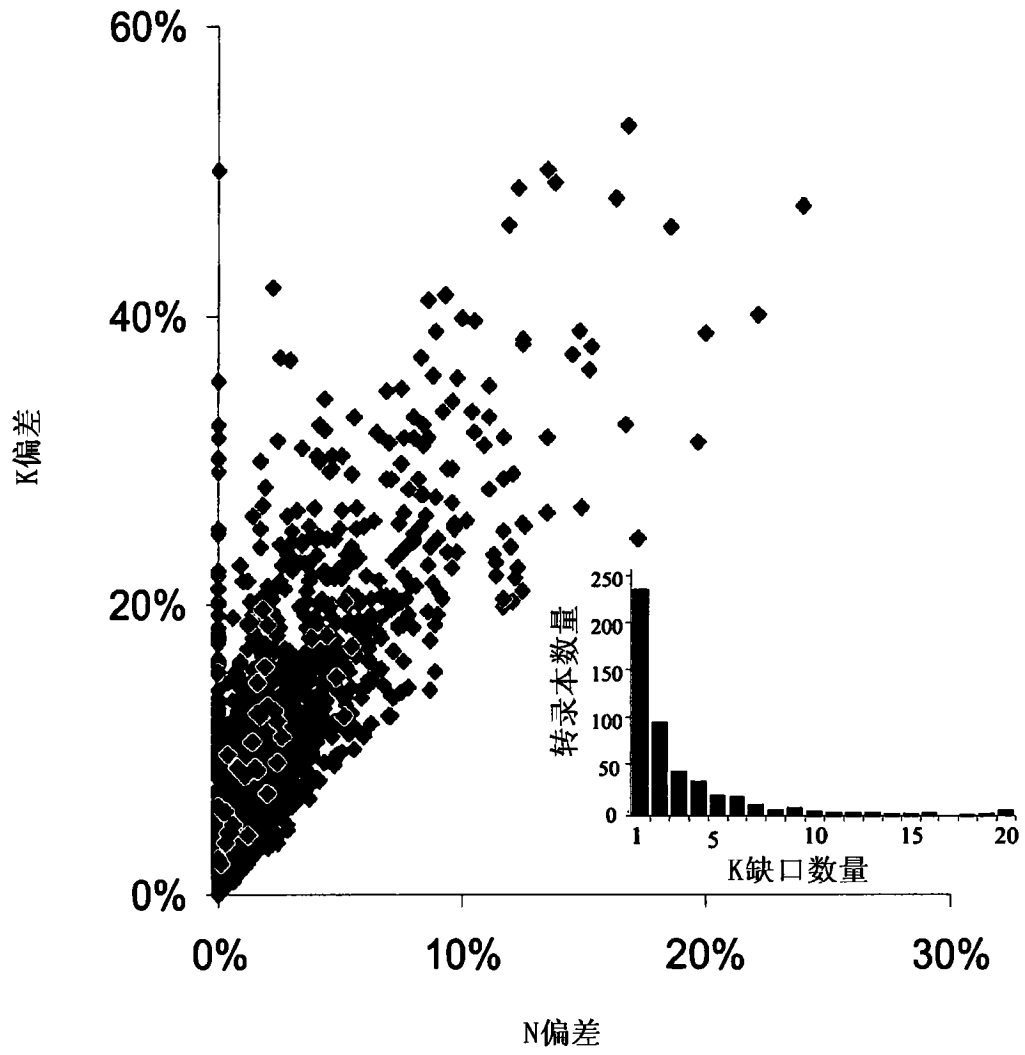


图 1

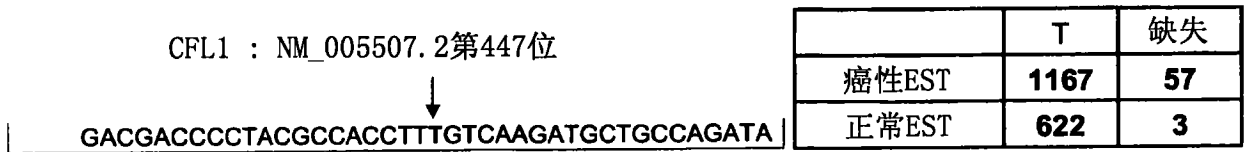


图 2A

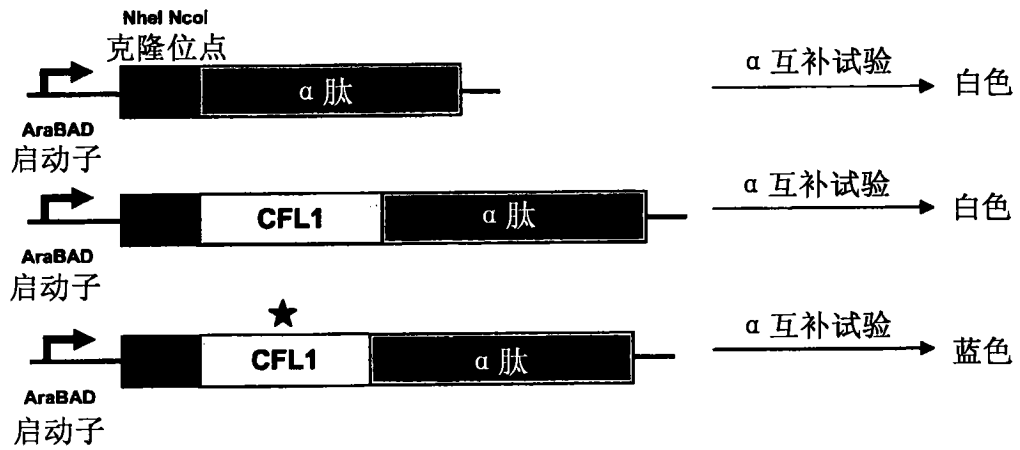


图 2B

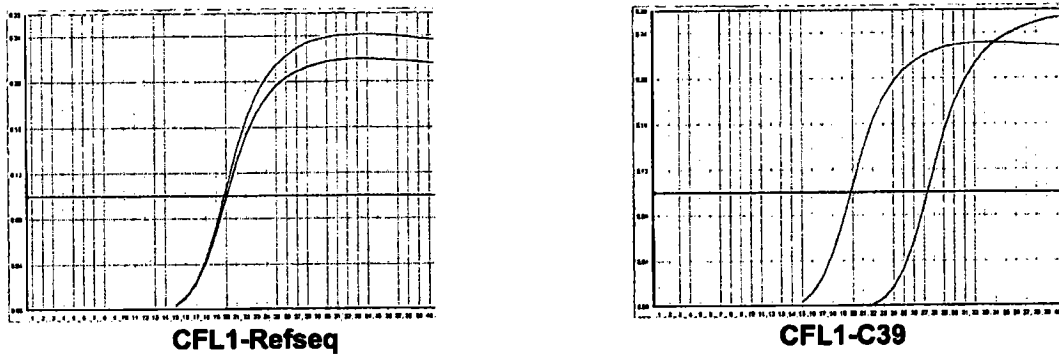


图 2C



图 2D

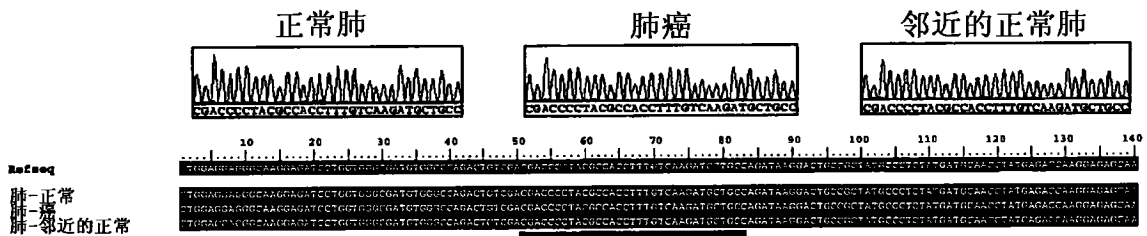


图 2E



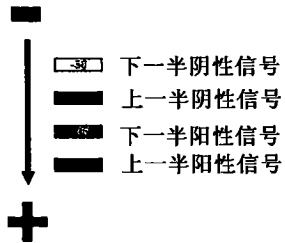
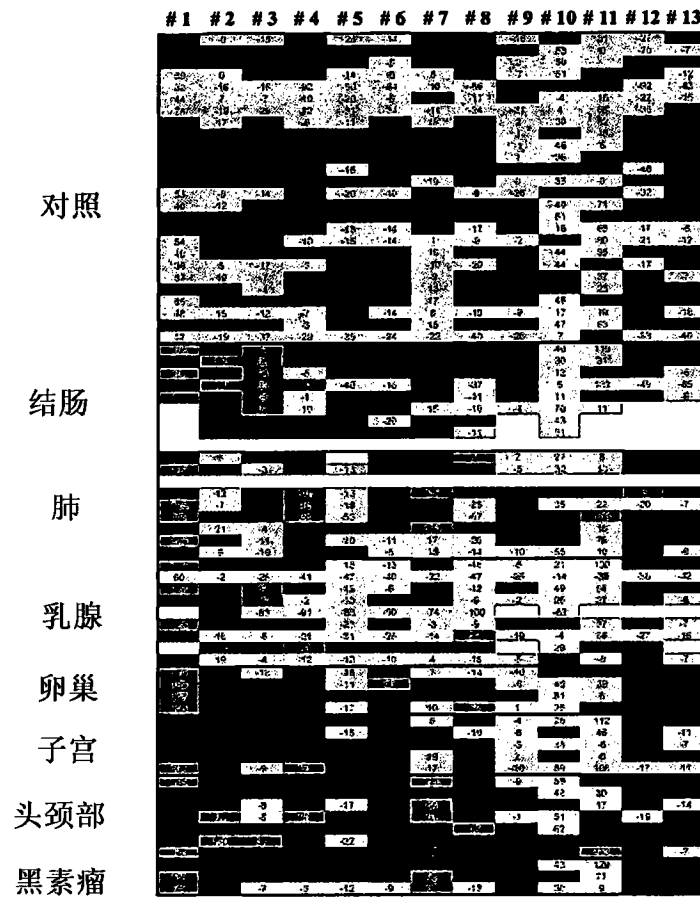
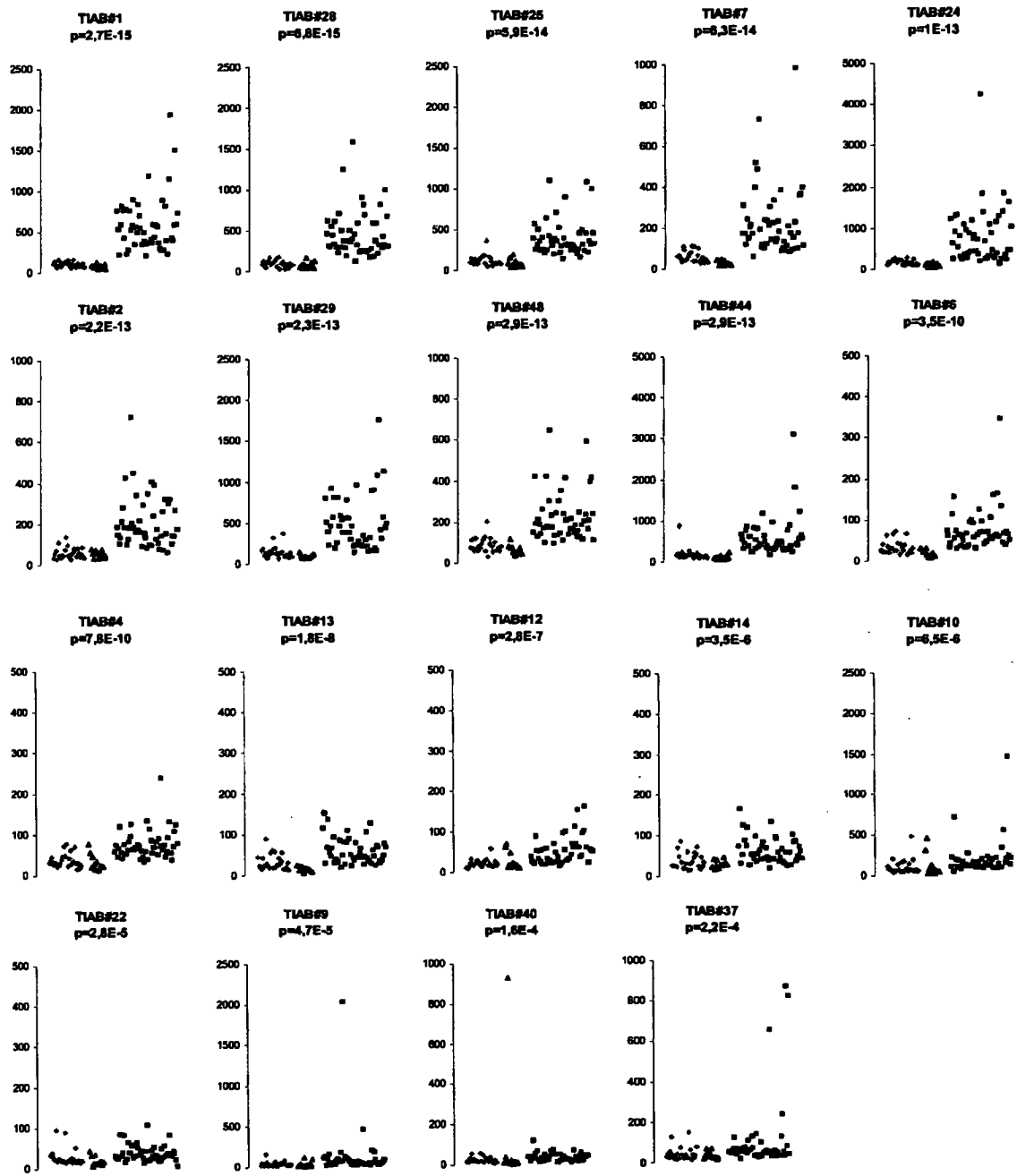


图 3B



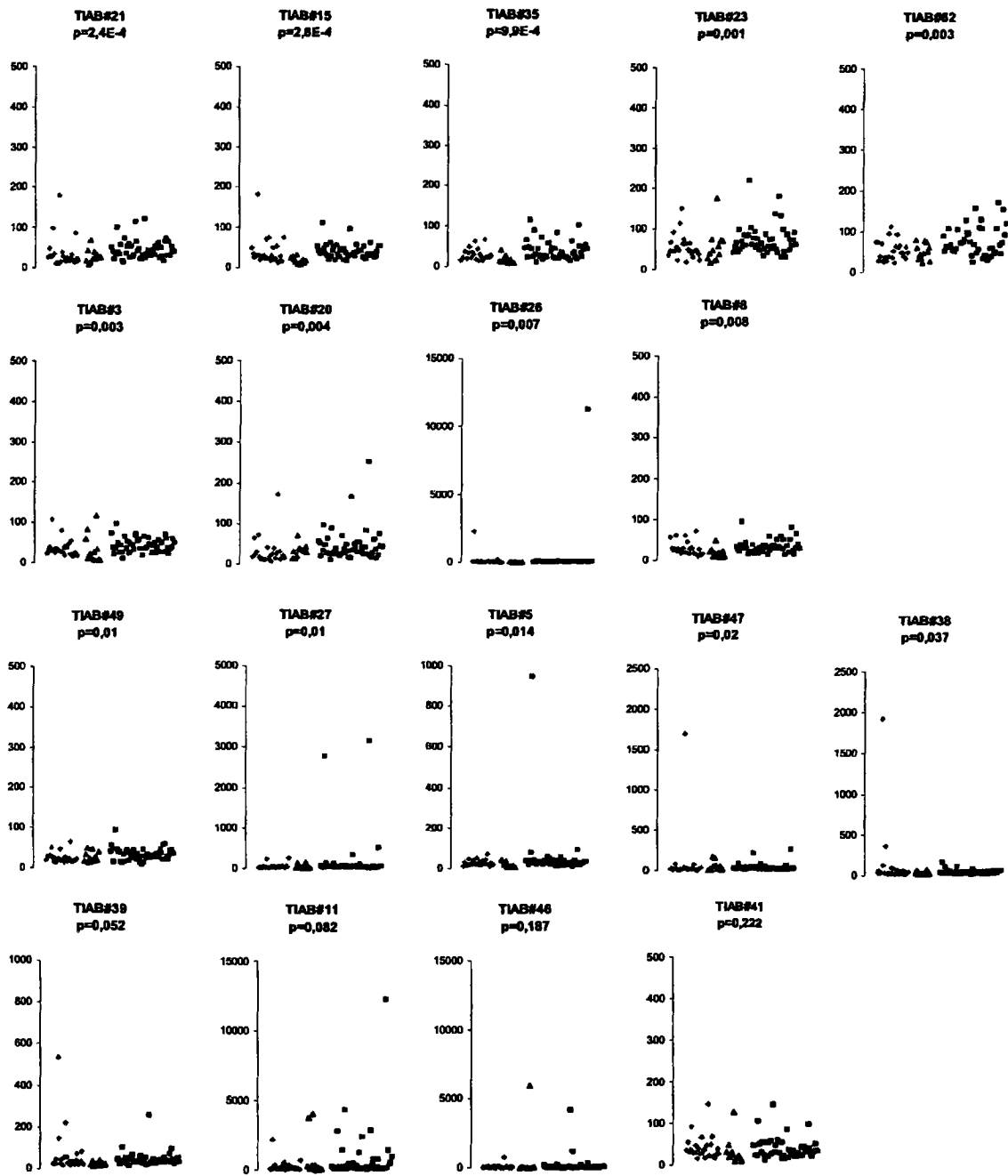


图 4A

变量	Wilcoxon p-值
TIAB_1	2,7E-15
TIAB_28	6,8E-15
TIAB_25	5,9E-14
TIAB_7	6,3E-14
TIAB_24	1,0E-13
TIAB_2	2,2E-13
TIAB_29	2,3E-13
TIAB_48	2,9E-13
TIAB_44	2,9E-13
TIAB_6	3,5E-10
TIAB_4	7,8E-10
TIAB_13	1,8E-08
TIAB_12	2,8E-07
TIAB_14	3,5E-06
TIAB_10	6,5E-06
TIAB_22	2,8E-05
TIAB_9	4,7E-05
TIAB_40	1,6E-04
TIAB_37	2,2E-04
TIAB_21	2,4E-04
TIAB_15	2,6E-04
TIAB_35	9,9E-04
TIAB_23	0,001
TIAB_62	0,003
TIAB_3	0,003
TIAB_20	0,004
TIAB_26	0,007
TIAB_8	0,008
TIAB_49	0,010
TIAB_27	0,010
TIAB_5	0,014
TIAB_47	0,020
TIAB_38	0,037
TIAB_39	0,052
TIAB_11	0,082
TIAB_46	0,187
TIAB_41	0,222

图 4B

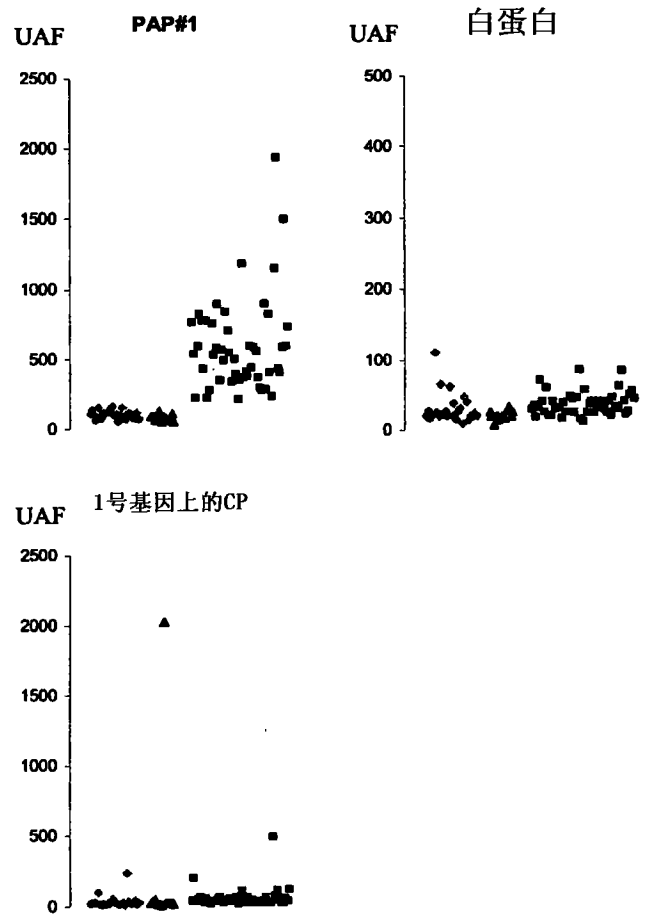


图 5A

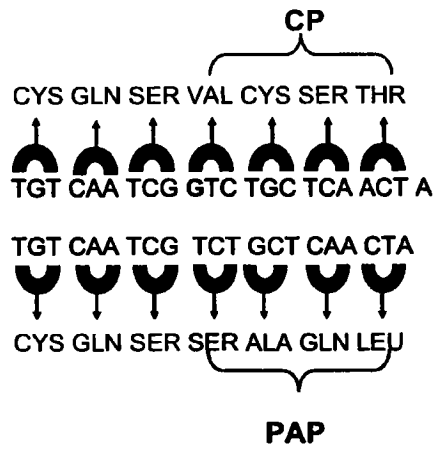


图 5B

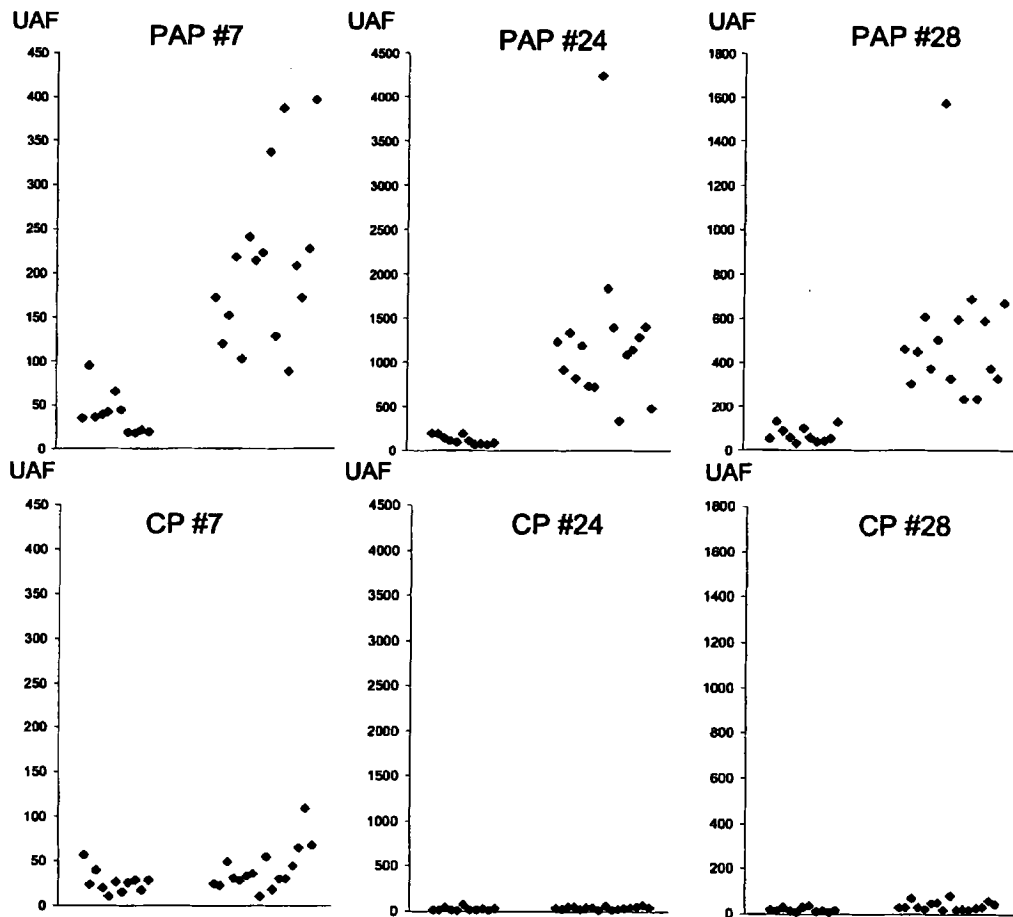


图 5C

# 48	人类 小鼠	<pre> gtggagcggcgaacaacctgcccaggacatccatgcg       gtggagcggcgaacaacctgcccaggacatccatgcg                     </pre>	TI →	<pre> INPRTSCASGRNCRRCFRERKPTPC INPRTSC  RNCRRC  RERKPK  PC INPRTSCGCRNCRRCRSRERKPKPC                     </pre>
# 62	人类 小鼠	<pre> aatggaaaacaatttatcttggctcgagctcagaaaa       aatggaaaacagatttatcttggctcgagctcagaaaa                     </pre>	TI →	<pre> ELRKNWGRNLSANLNR ELRR+WNGRNLSANL+R ELRKNWGRNLSANLSR                     </pre>
#7	人类 小鼠	<pre> ttggtgaaaggcttggcttcttggctctgcttcaaaagg       ttggtgaaaggcttggcttcttggctctgcttcaaaagg                     </pre>	TI →	<pre> LRCFRRYELRLPAKRRKRN L  CFR  LRLPAKR+RKRN LRCFRRCALRLPAKRRKRN                     </pre>
# 2	人类 小鼠	<pre> HLNSPHLFRKRCYHLIVFERQWHSLLGEVLALVKE +LRS SLNSSSP                     </pre>		
# 9	人类 小鼠	<pre> VIVVTRQWVLSRDEQNQEAITAILPHLLPK 无基因                     </pre>		
CP # 7	人类 小鼠	<pre> ALLQVRAEIASKEKEEEE ALLQVRAEIASKEKEEEE ALLQVRAEIASKEKEEEE                     </pre>		

图 6A

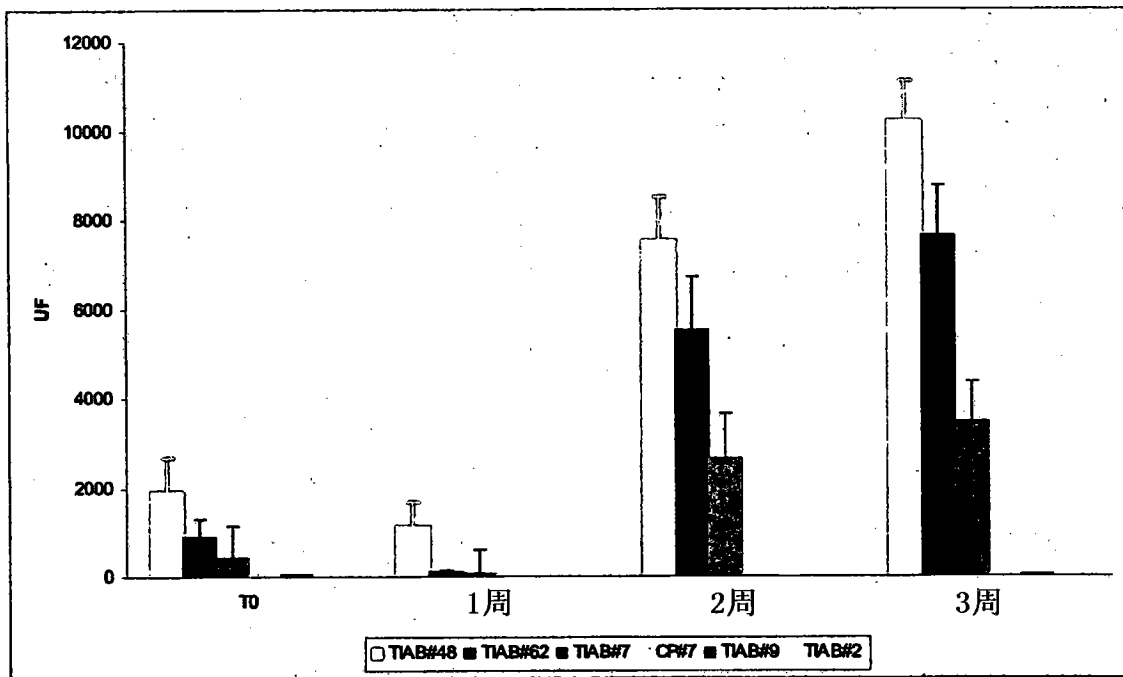


图 6B

PAP组合 : IK (PAP 7), PRDX6 (PAP 66), CDKN1A (PAP 70),  
MRPL3 (PAP 29), RPL13A (PAP 68), VIM (PAP 48)

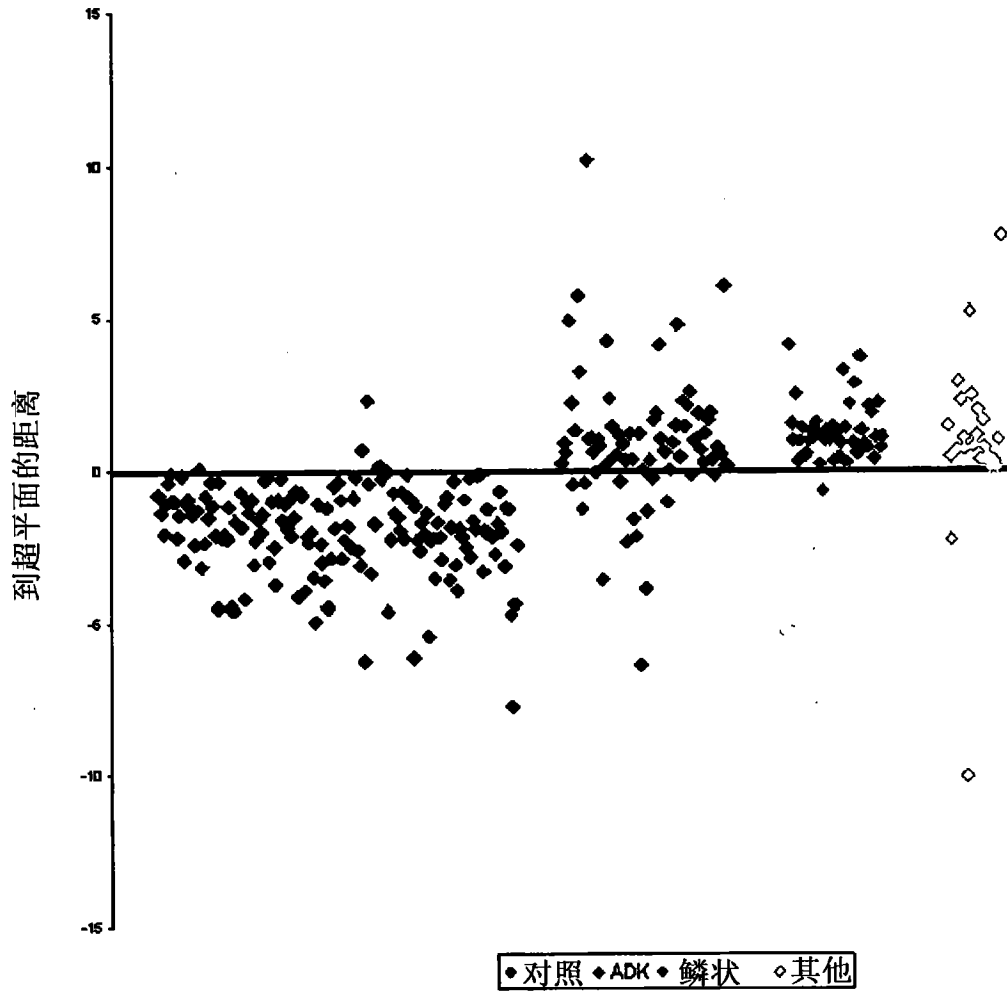


图 7A

年龄 (岁)	n	阳性测试	%
35 - 50	17	15	88
50 - 65	68	59	87
> 65	55	47	85

} NS

图 7B

组织病理学	n	阳性测试	%
ADK	67	51	76
鳞状	40	39	98
其他	33	31	93

X<sup>2</sup>  
P<0.001

图 7C

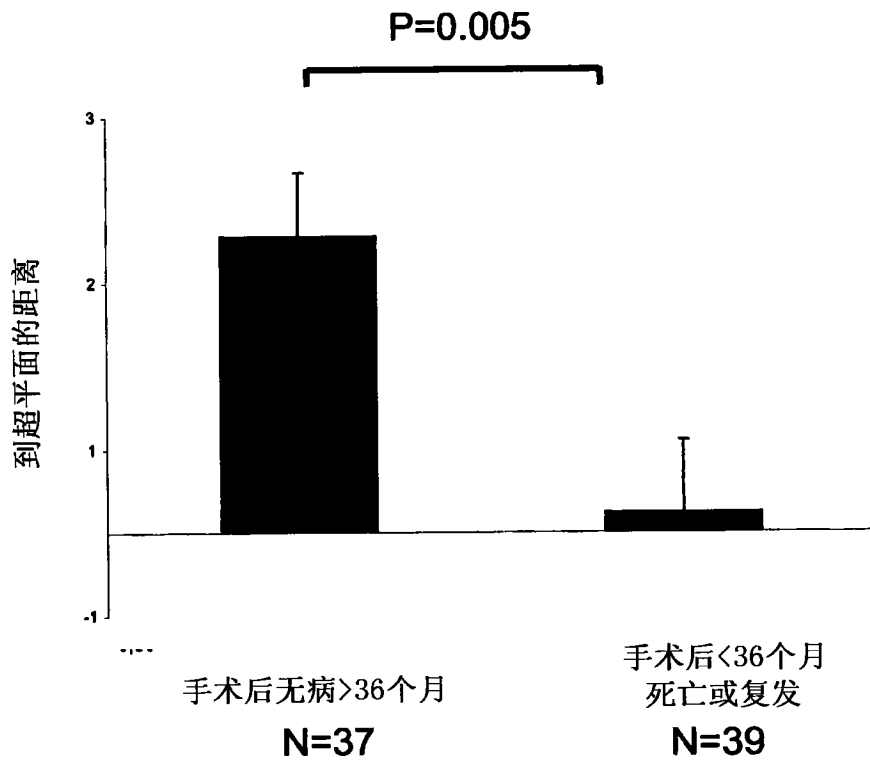


图 7D

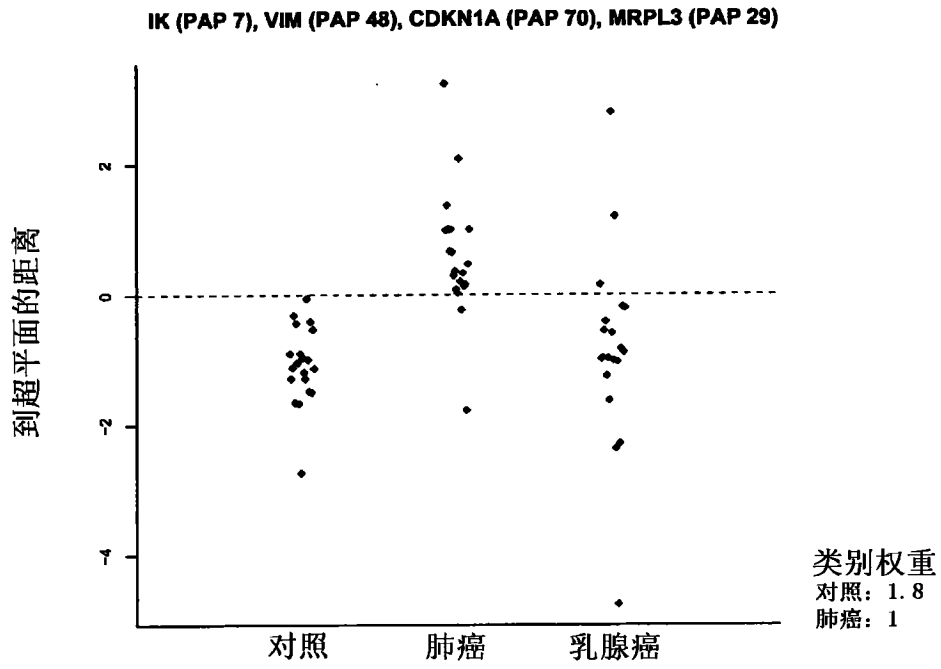
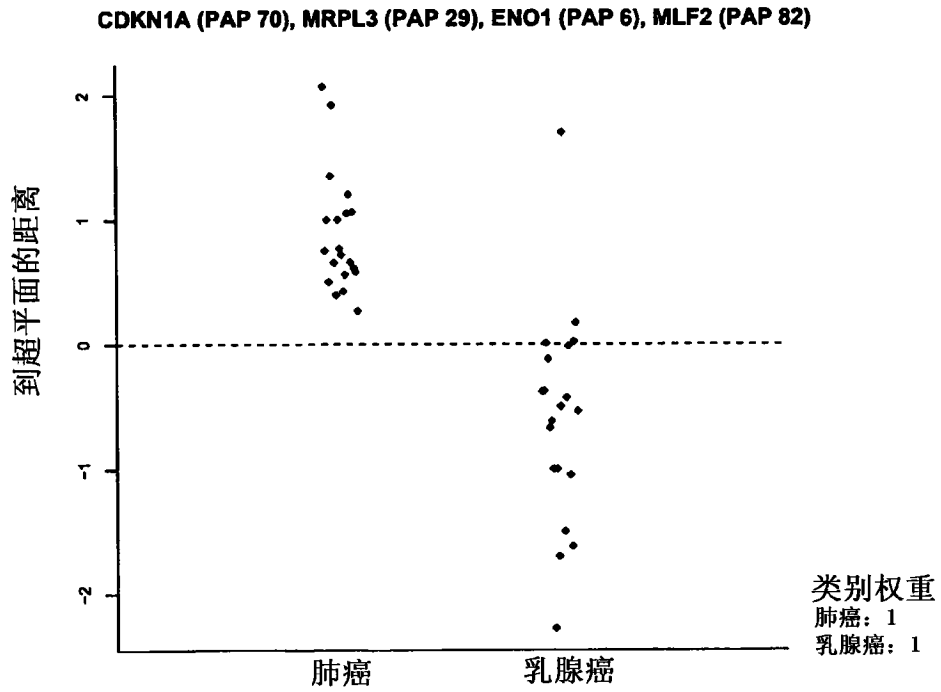
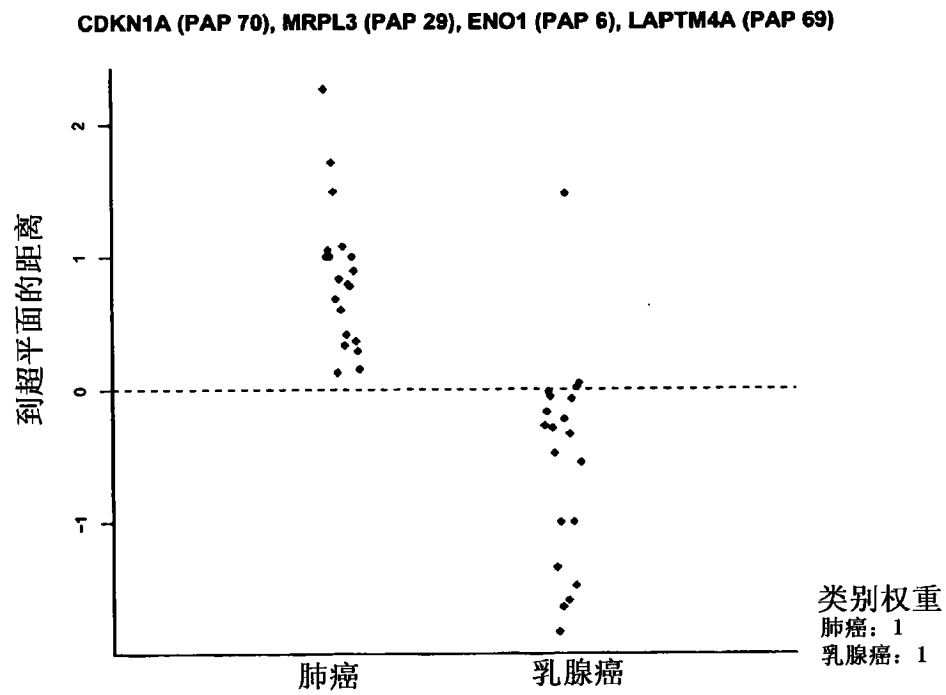
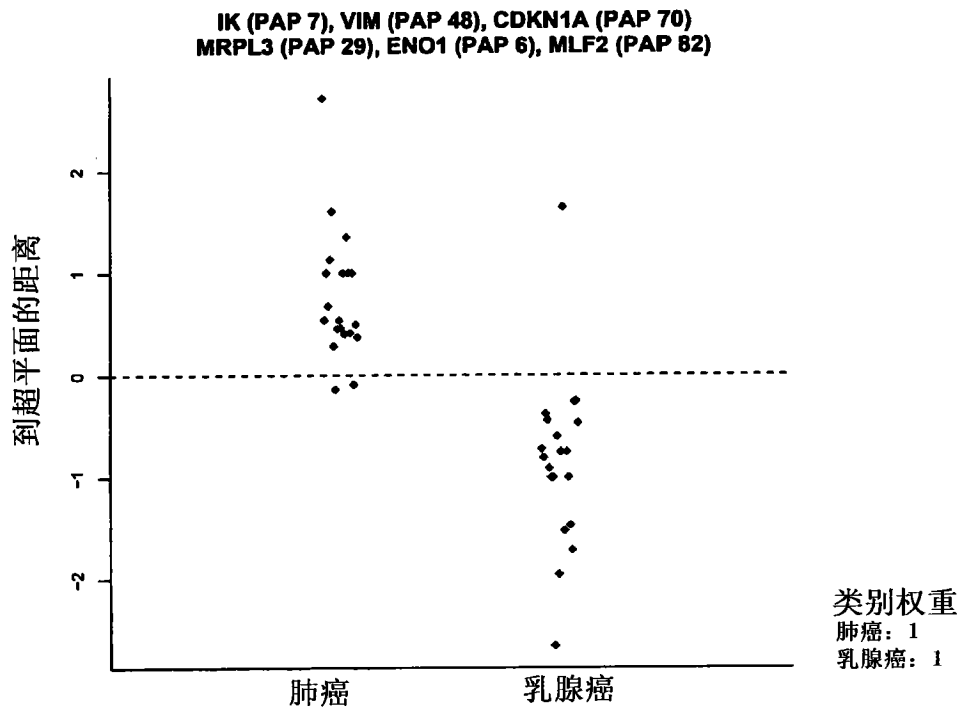


图 8A





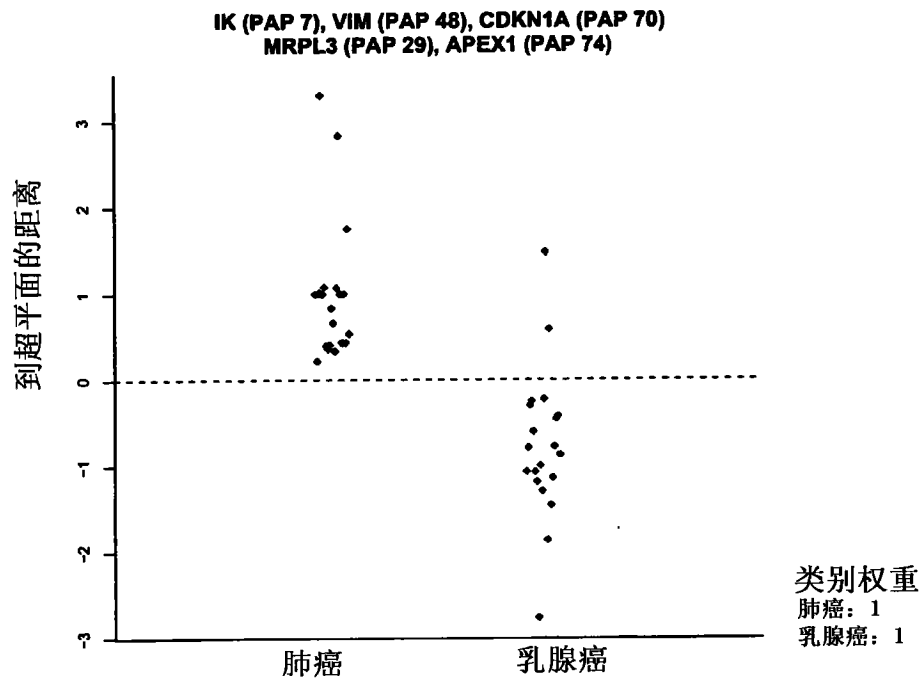


图 8B

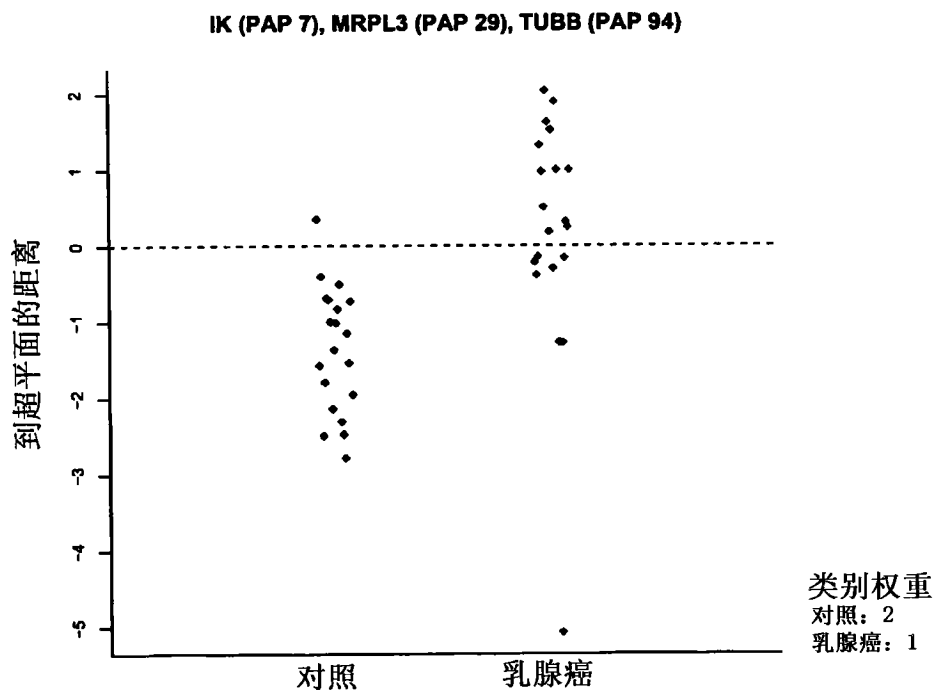


图 8C

专利名称(译)	检测TIABS的组合物和方法		
公开(公告)号	<a href="#">CN101971026A</a>	公开(公告)日	2011-02-09
申请号	CN200880125970.7	申请日	2008-12-15
[标]发明人	伯纳德毕海因 维尔日妮欧希尔 玛丽布鲁利亚德		
发明人	伯纳德·毕海因 维尔日妮·欧希尔 玛丽·布鲁利亚德		
IPC分类号	G01N33/574 G01N33/53		
CPC分类号	G01N2333/7051 G01N2800/56 G01N33/57484 G01N2800/52 C07K16/32 G01N33/53 G01N33/574 G01N33/68		
代理人(译)	张颖		
优先权	20070301680 2007-12-14 EP 2007301680 2007-12-14 EP		
其他公开文献	CN101971026B		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

#### 摘要(译)

本发明涉及用于评估对象的生理状态的新方法和产品。更具体来说，本发明涉及通过在来自对象的样品中测定针对特定异常蛋白结构域的抗体的水平、或带有对这样的异常蛋白结构域特异性的TCR的免疫细胞的存在或数量，来评估对象中癌症的存在、风险或阶段的方法。本发明也适用于评估对象对治疗的响应性，以及筛选候选药物和设计新疗法。本发明可用于任何哺乳动物对象，特别是用于人类对象。

	正常	癌症
从 NCBI 检索到的 EST	3,949,323	3,043,498
具有 EST 匹配的转录本的数量	34,974	34,788
具有 EST 匹配的转录本的数量		33,111
分析的核苷酸		88,372,747
由>70 EST 所定义的位置		2,829,135
匹配统计学限制的位置		
取代		5,784
缺口		3,790