



(12)发明专利

(10)授权公告号 CN 105917230 B

(45)授权公告日 2020.06.16

(21)申请号 201480071589.2

(22)申请日 2014.11.11

(65)同一申请的已公布的文献号

申请公布号 CN 105917230 A

(43)申请公布日 2016.08.31

(30)优先权数据

1319878.3 2013.11.11 GB

(85)PCT国际申请进入国家阶段日

2016.06.29

(86)PCT国际申请的申请数据

PCT/GB2014/053340 2014.11.11

(87)PCT国际申请的公布数据

W02015/067969 EN 2015.05.14

(73)专利权人 伊缪诺维亚公司

地址 瑞典伦德

(72)发明人 C.A.K.博雷贝克 C.L.B.温格伦

(74)专利代理机构 北京市柳沈律师事务所

11105

代理人 张文辉

(51)Int.Cl.

G01N 33/574(2006.01)

C12Q 1/6886(2018.01)

C40B 40/10(2006.01)

(56)对比文件

CN 102186994 A, 2011.09.14,

审查员 贾静

权利要求书2页 说明书51页

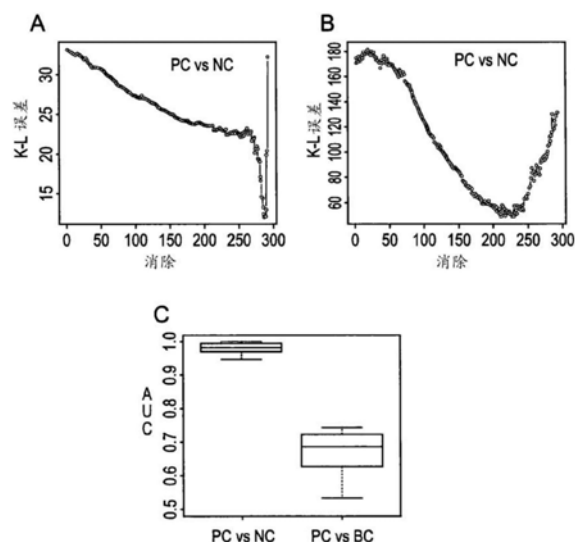
序列表2页 附图6页

(54)发明名称

用于确定胰腺癌的方法、阵列和其用途

(57)摘要

本发明涉及用于确定个体中胰腺癌的存在的方法,包括以下步骤,或由以下步骤组成:(a)提供来自所述个体的待测试的样品,和(b)通过测量所述测试样品中选自表A中确定的组的一种或多种生物标志物的表达确定所述测试样品的生物标志物签名,其中所述测试样品中选自表A中确定的组的一种或多种生物标志物的表达对于所述个体患有胰腺癌是指示性的。本发明也包括用于本发明的方法中的阵列和组件试剂盒。核心生物标志物是HDAH2和TNFRSF3。



1. 能够结合选自表A中确定的组的一种或多种生物标志物的一种或多种结合试剂在制备用于确定个体中胰腺癌的存在的试剂盒/阵列中的用途,

其中所述结合试剂是抗体或其抗原结合片段,

且其中所述一种或多种生物标志物选自下组:HADH2,白介素-4,白介素-6,lewis x,淋巴毒素-alpha,血管内皮生长因子,视觉系统同源盒2,载脂蛋白A1,补体C3,补体C4,补体C5,细胞周期素依赖的激酶2,细胞间粘附分子1,干扰素gamma,白介素-13,肌间蛋白-2,血浆蛋白酶C1抑制物,备解素和唾液酸化的Lewis x,且其中所述一种或多种生物标志物是HADH2或包括HADH2。

2. 根据权利要求1的用途,其中所述一种或多种生物标志物包括HADH2,白介素-4,白介素-6,lewis x,淋巴毒素-alpha,血管内皮生长因子,视觉系统同源盒2,载脂蛋白A1,补体C3,补体C4,补体C5,细胞周期素依赖的激酶2,细胞间粘附分子1,干扰素gamma,白介素-13,肌间蛋白-2,血浆蛋白酶C1抑制物,备解素和唾液酸化的Lewis x。

3. 根据权利要求2的用途,其中所述一种或多种生物标志物包含表A中列出的所有生物标志物。

4. 根据权利要求1的用途,其中所述胰腺癌选自下组:腺癌、腺鳞癌、印戒细胞癌(signet ring cell carcinoma)、肝样癌(hepatoid carcinoma)、胶样癌(colloid carcinoma)、未分化癌和伴有破骨细胞样巨细胞的未分化癌。

5. 根据权利要求1的用途,其中所述胰腺癌是腺癌。

6. 根据权利要求1的用途,其中所述抗体或其抗原结合片段选自下组:scFv;Fab;免疫球蛋白分子的结合域。

7. 根据权利要求1的用途,其中所述一种或多种结合试剂固定化在表面上。

8. 根据权利要求1的用途,其中所述用途是在制备用于确定个体中胰腺癌的存在的阵列中的用途。

9. 根据权利要求8的用途,其中所述阵列是基于珠的阵列。

10. 根据权利要求8的用途,其中所述阵列是基于表面的阵列。

11. 根据权利要求8的用途,其中所述阵列选自下组:宏阵列;微阵列;纳米阵列。

12. 一种用于确定个体中胰腺癌的存在的阵列,包含能够结合选自表A中确定的组的两种或更多种生物标志物的两种或更多种结合试剂

其中所述结合试剂是抗体或其抗原结合片段,

且其中所述两种或更多种生物标志物选自下组:HADH2,白介素-4,白介素-6,lewis x,淋巴毒素-alpha,血管内皮生长因子,视觉系统同源盒2,载脂蛋白A1,补体C3,补体C4,补体C5,细胞周期素依赖的激酶2,细胞间粘附分子1,干扰素gamma,白介素-13,肌间蛋白-2,血浆蛋白酶C1抑制物,备解素和唾液酸化的Lewis x,且其中所述两种或更多种生物标志物包括HADH2。

13. 根据权利要求12的阵列,包含能够结合下组中所有蛋白的结合试剂:HADH2,白介素-4,白介素-6,lewis x,淋巴毒素-alpha,血管内皮生长因子,视觉系统同源盒2,载脂蛋白A1,补体C3,补体C4,补体C5,细胞周期素依赖的激酶2,细胞间粘附分子1,干扰素gamma,白介素-13,肌间蛋白-2,血浆蛋白酶C1抑制物,备解素和唾液酸化的Lewis x。

14. 一种用于确定胰腺癌的存在的试剂盒,包含:

A) 能够结合选自表A中确定的组的两种或更多种生物标志物的两种或更多种结合试剂;

B) 用于进行在个体中确定胰腺癌的存在的方法的说明书,
其中所述结合试剂是抗体或其抗原结合片段,

且其中所述两种或更多种生物标志物选自下组:HADH2,白介素-4,白介素-6,lewis x,淋巴毒素-alpha,血管内皮生长因子,视觉系统同源盒2,载脂蛋白A1,补体C3,补体C4,补体C5,细胞周期素依赖的激酶2,细胞间粘附分子1,干扰素gamma,白介素-13,肌间蛋白-2,血浆蛋白酶C1抑制物,备解素和唾液酸化的Lewis x,且其中所述两种或更多种生物标志物包括HADH2。

15. 根据权利要求14的试剂盒,包含能够结合下组中所有蛋白的结合试剂:HADH2,白介素-4,白介素-6,lewis x,淋巴毒素-alpha,血管内皮生长因子,视觉系统同源盒2,载脂蛋白A1,补体C3,补体C4,补体C5,细胞周期素依赖的激酶2,细胞间粘附分子1,干扰素gamma,白介素-13,肌间蛋白-2,血浆蛋白酶C1抑制物,备解素和唾液酸化的Lewis x。

用于确定胰腺癌的方法、阵列和其用途

发明领域

[0001] 本发明涉及用于检测胰腺癌的方法,及用于该方法的生物标志物和阵列。

[0002] 背景

[0003] 在美国,胰腺导管腺癌,或胰腺癌(PC)是第4普遍的癌症相关死亡的原因,导致几乎与乳腺癌一样多的死亡,尽管发病率低10倍[1]。预后差主要是由于不能在早期检测PC,尽管数据证明从肿瘤起始到获得转移能力耗费多于五年[2],清楚的证明了如果有可用的标志物,则有早期检测的机会窗口。如今该癌症通常在晚期疾病进展检测到,其中肿瘤不可动手术且已经转移[1,3]。然而,尽管大肿瘤切除的五年存活仅为10-20%[4,5],如果可以切除<20mm的肿瘤,其增加到30-60%,且如果切除的大小为<10mm增加到75%[8]。较晚检测是由于非特异性的临床症状以及缺乏灵敏的技术和标志物用于早期诊断。有意思的是,研究提示胰腺肿瘤可能在早至临床诊断之前的六个月时的无症状期切除[9,10]。

[0004] 迄今评价得最多的PC标志物,CA19-9,缺点在于较差的特异性,其中在胰腺炎和其它癌症中也有升高的水平,并在Lewis a和b阴性肿瘤中完全缺失。结果,已经不鼓励使用CA19-9用于胰腺癌筛选[11]。尽管做出了许多努力,没有其它的单个生物标志物表现出胜过CA19-9,且近年来本领域已经朝着多重标志物组用于增加的灵敏度和特异性的方向前进[12]。组主要由高丰度的血液蛋白,频繁的急性反应物(例如CRP和SAA),已知的肿瘤标志物(例如,CA242,CA125和CEA),粘附分子(例如ICAM-1和ADCAM),参与细胞外基质降解的蛋白(例如,MMP和TIMP),脂蛋白(例如,Apo-C1,Apo-A2),以及几种其它分子组成,通常与CA19-9组合[15-20]。尽管报道了对于PC对健康对照或良性胰腺状况的高灵敏度,这些组中没有一个是已经在临床用使用中验证。

[0005] 除了已经提及的标志物外,几种免疫调节蛋白作为肿瘤生物标志物可能是感兴趣的,因为癌症和炎症之间的关联得到不断阐明[21]。考虑到许多免疫调节蛋白的功能的系统性效果以及多重性,通常认为2至5个标志物的小组将不足以特异性用于胰腺癌,特别是当试图从胰腺炎和呈现出相似症状的其他疾患区分胰腺癌时。然而,先前的研究已经显示了这些分析物数量的增加(≤ 25)可能产生高度疾病特异性的免疫签名,反映对疾病的系统性反应[22-24]。

[0006] 然而,对于免疫调节蛋白组的分析伴随着几项挑战。首先,感兴趣的蛋白的血清浓度表现出广阔的动态范围,从高微克到低皮克每mL,其使得使用常规蛋白组方法学对它们的同时检测变得复杂[25,26]。第二,似乎存在以下共识,有希望的癌症标志物更可能在最低丰度的,通常为低分子量的蛋白中发现[27,28]。第三,预期这些低丰度分析物的血清水平的疾病相关的变化是小的,因此对于充分的统计学分析将需要大量的临床样品[29]。为了这些目的,我们设计了高度多重性的重组抗体微阵列,具有靶向主要的免疫调节蛋白的接近300个scFv抗体[30]。使用这些阵列,可以以高度可重复和高通量的方式测量成百份样品中的蛋白表达。

[0007] 发明简述

[0008] 在本多中心研究中,来自患有胰腺癌、良性胰腺疾病的患者,以及正常对照的338

份个体血清样品在我们内部设计的抗体微阵列上分析。为了确定最有识别力的标志物签名,我们应用了迭代反向消除程序(iterative backward elimination procedure),基于支持向量机器分析(support vector machine analysis)和设计以预测抗体的最优组合[31]。为了这一目的,在训练组中鉴定了用于从良性和健康对照区分PC的25重免疫签名,在独立的测试组中预验证,并与来源于差异蛋白表达分析的签名比较。另外,基于肿瘤在胰腺中的原始位置可以将蛋白谱(protein profiling)应用于分层血清样品,这根据我们所知先前尚没有使用蛋白组学完成。总的来说,这些发现向胰腺癌的蛋白质组拼图添加了重要的信息,最后其可能导致对于成千上万的患者提供益处的多重(multiplexed)生物标志物。

[0009] 相应的,本发明的第一个方面提供在个体中检测胰腺癌的方法,包括以下步骤或由以下步骤组成:

[0010] a) 提供来自所述个体的待测试的样品;

[0011] b) 通过测量所述测试样品中选自表A(i),(ii)或(iii)中确定的组的至少一个生物标志物的表达确定所述测试样品的生物标志物签名;

[0012] 其中所述测试样品中选自表A(i),(ii)或(iii)中确定的组的一个或多个生物标志物的表达对于胰腺癌的存在是指示性的。

[0013] 通过“待测试的样品”,“测试样品”或“对照样品”,我们包括取自或来源于个体的组织或流体样品。优选所述待测试的样品从哺乳动物提供。所述哺乳动物可以是任意的驯养动物或农场动物。优选的,所述哺乳动物是大鼠,小鼠,豚鼠,猫,狗,马或灵长类。最优选的,所述哺乳动物是人。优选的样品是细胞或组织样品(或其衍生物),其包含以下物质,或由以下物质组成:血浆,血浆细胞,血清,组织细胞或同样优选的,来源于细胞或组织样品的蛋白或核酸。优选的测试和对照样品来源于相同的物种。

[0014] 在可替换或另外的实施方案中,所述组织样品是胰腺组织。在可替换或另外的实施方案中,所述细胞样品是胰腺细胞的样品。

[0015] 通过“表达(expression)”我们表示基因产物诸如mRNA或蛋白质的水平或量。

[0016] 检测和/或测定蛋白质和/或核酸的浓度的方法对于本领域技术人员而言是熟知的,参见例如Sambrook and Russell,2001,Cold Spring Harbor Laboratory Press。

[0017] 通过“生物标志物”我们表示自然存在的生物分子或其组分或其片段,对其的测量可提供有助于胰腺癌预后的信息。例如,所述生物标志物可以为自然存在的蛋白质或碳水化合物部分,或其抗原组分或其片段。

[0018] 在可替换或另外的实施方案中,所述方法包括步骤(a)和(b)以及下述另外的步骤,或由步骤(a)和(b),以及下述另外的步骤构成:

[0019] c) 提供来自未患胰腺癌的个体的一个或多个对照样品;

[0020] d) 通过测量在步骤(b)中测量的一种或多种生物标志物在对照样品中的表达来确定对照样品的生物标志物签名;

[0021] 其中,胰腺癌存在通过如下事件鉴定,该事件为在步骤(b)中测量的一种或多种生物标志物在测试样品中的表达不同于步骤(d)中测量的一种或多种生物标志物在对照样品中的表达。

[0022] 所述一个或多个对照样品可以来自健康个体(即未患有任意疾病或状况的个体),患有非胰腺疾病或状况的个体或患有良性胰腺疾病或状况的个体(例如,急性或慢性胰腺

炎)。

[0023] 在另一个实施方案中,所述方法包括步骤(a)和(b)以及下述另外的步骤,或由步骤(a)和(b),以及下述另外的步骤构成:

[0024] e) 提供来自患有胰腺癌的个体的一个或多个对照样品;

[0025] f) 通过检测在步骤(b)中测量的一种或多种生物标志物在对照样品中的表达来确定对照样品的生物标志物签名;

[0026] 其中,胰腺癌存在通过如下事件来鉴定,该事件为在步骤(b)中测量的一种或多种生物标志物在测试样品中的表达对应于步骤(f)中测量的一种或多种生物标志物在对照样品中的表达。

[0027] 在可替换或另外的实施方案中,使用标准或参考值来代替所述一个或多个阳性或阴性对照,或除所述一个或多个阳性或阴性对照外还使用标准或参考值。因此,可以在分开的程序中从测试值确定所述标准或参考值。

[0028] 在可替换或另外的实施方案中,所述方法用于确定胰腺癌的存在,所述胰腺癌起源于(i)胰头或(ii)胰体或胰尾。在这一实施方案中,步骤(e)可以包括(i)提供来自患有起源于胰头的胰腺癌的个体的一个或多个对照样品和/或(ii)提供来自患有起源于胰体或胰尾的胰腺癌的个体的一个或多个对照样品。

[0029] 在这一实施方案中,起源于胰头的胰腺癌存在通过如下事件来鉴定,该事件为在步骤(b)中测量的一种或多种生物标志物在(e)(i)测试样品(在存在的情况下)中的表达对应于步骤(f)中测量的一种或多种生物标志物在对照样品中的表达和/或在步骤(b)中测量的一种或多种生物标志物在(e)(ii)测试样品(在存在的情况下)中的表达不同于步骤(f)中测量的一种或多种生物标志物在对照样品中的表达。

[0030] 在这一实施方案中,起源于胰体或胰尾的胰腺癌存在通过如下事件来鉴定,该事件为在步骤(b)中测量的一种或多种生物标志物在(e)(i)测试样品(在存在的情况下)中的表达不同于步骤(f)中测量的一种或多种生物标志物在对照样品中的表达和/或在步骤(b)中测量的一种或多种生物标志物在(e)(ii)测试样品(在存在的情况下)中的表达对应于步骤(f)中测量的一种或多种生物标志物在对照样品中的表达。

[0031] 通过“对应于在对照样品中的表达”我们包括一种或多种生物标志物在待测样品中的表达与其在阳性对照样品中的表达相同或相似。优选一种或多种生物标志物在待测样品中的表达与其在阳性对照样品中的表达相同。

[0032] 生物标志物的差异化表达(上调或下调),或表达的缺乏可以通过本领域技术人员已知的任何合适的方式来确定。差异化表达通过至少小于0.05的p值($p < 0.05$)来确定,例如至少 < 0.04 、 < 0.03 、 < 0.02 、 < 0.01 、 < 0.009 、 < 0.005 、 < 0.001 、 < 0.0001 、 < 0.00001 或至少 < 0.000001 。优选,使用支持向量机(SVM)来确定差异化表达。优选,SVM为如下所述的SVM。最优选,SVM是下述表B中所述的SVM。

[0033] 本领域技术人员应当理解差异化表达可与单个生物标志物相关或与组合考虑的多个生物标志物(即作为生物标志物签名)相关。因此,p值可以与单个生物标志物关联或与一组生物标志物关联。实际上,当单独考虑时具有p值大于0.05差异化表达的蛋白质,在将它们的表达水平与一种或多种其他生物标志物结合考虑时,其仍然可用作根据本发明的生物标志物。

[0034] 如在所附的实例中所例证的,在组织、血、血清或血浆检测样品中的特定蛋白质的表达可以是个体中胰腺癌的指示。例如,在单个测试样品中特定血清蛋白质的相对表达可以是个体中胰腺癌存在的指示。

[0035] 步骤(b)可以包括测量表A(i)中列出的一种或多种生物标志物的表达或步骤(b)可以由其组成。步骤(b)可以包括测量表A(i)中列出的一种或多种生物标志物的表达或步骤(b)可以由其组成,例如表A(i)中列出的至少两种生物标志物。因此,步骤(b)可以包括测量表A(i)中列出的所有生物标志物的表达或步骤(b)可以由其组成。

[0036] 步骤(b)可以包括测量表A(ii)中列出的一种或多种生物标志物的表达或步骤(b)可以由其组成,例如表A(ii)中列出的至少2,3,4,5,6,7,8,或9种生物标志物。因此,步骤(b)可以包括测量表A(ii)中列出的所有生物标志物的表达或步骤(b)可以由其组成。

[0037] 步骤(b)可以包括测量表A(iii)中列出的一种或多种生物标志物的表达或步骤(b)可以由其组成,例如表A(iii)中列出的至少2,3,4,5,6,7,8,9,10,11,12,13,14,15,16或17种生物标志物。因此,步骤(b)可以包括测量表A(iii)中列出的所有生物标志物的表达或步骤(b)可以由其组成。

[0038] 步骤(b)可以包括测量表A(iv)中列出的一种或多种生物标志物的表达或步骤(b)可以由其组成,例如表A(iv)中列出的至少2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21或22,23或24种生物标志物。因此,步骤(b)可以包括测量表A(iv)中列出的所有生物标志物的表达或步骤(b)可以由其组成。

[0039] 步骤(b)可以包括测量表A(v)中列出的一种或多种生物标志物的表达或步骤(b)可以由其组成,例如表A(v)中列出的至少2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55或56种生物标志物。因此,步骤(b)可以包括测量表A(v)中列出的所有生物标志物的表达或步骤(b)可以由其组成。

[0040] 步骤(b)可以包括测量表A(vi)中列出的一种或多种生物标志物的表达或步骤(b)可以由其组成。因此步骤(b)可以包括测量表A(vi)中列出的所有生物标志物的表达或步骤(b)可以由其组成。

[0041] 步骤(b)可以包括测量表A中列出的一种或多种生物标志物的表达或步骤(b)可以由其组成,例如表A中列出的至少2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109或110种生物标志物。因此,步骤(b)可以包括测量表A中列出的所有生物标志物的表达或步骤(b)可以由其组成。

[0042] 在所述方法用于确定胰腺癌的存在的情况下步骤(b)优选包括测量以下表达或步骤(b)由其组成:

[0043] 来自表A(i)中列出的生物标志物的一种或多种生物标志物,例如表A(i)中列出的至少2种生物标志物;

[0044] 来自表A(ii)中列出的生物标志物的一种或多种生物标志物,例如表A(ii)中列出的至少2,3,4,5,6,7,8或9种生物标志物;

[0045] 来自表A (iii) 中列出的生物标志物的一种或多种生物标志物,例如表A (iii) 中列出的至少2,3,4,5,6,7,8,9,10,11,12,13,14,15,16或17种生物标志物;

[0046] 来自表A (iv) 中列出的生物标志物的一种或多种生物标志物,例如表A (iv) 中列出的至少2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20, 21,22,23或24种生物标志物;

[0047] 来自表A (v) 中列出的生物标志物的一种或多种生物标志物,例如表A (v) 中列出的至少2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28, 29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53, 54,55或56种生物标志物;

[0048] 来自表A (vi) 中列出的生物标志物的一种或多种生物标志物;因此步骤 (b) 可以包括测量表A (vi) 中列出的所有生物标志物的表达或步骤 (b) 可以由其组成。

[0049] 在所述方法用于确定起源于 (i) 胰头或 (ii) 胰体或胰尾的胰腺腺癌的存在的情况下步骤 (b) 优选包括测量以下表达或步骤 (b) 由其组成:

[0050] 来自表A (ii) 中列出的生物标志物的一种或多种生物标志物,例如表A (ii) 中列出的至少2,3,4,5,6,7,8或9种生物标志物;

[0051] 来自表A (iv) 中列出的生物标志物的一种或多种生物标志物,例如表A (iv) 中列出的至少2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23或24种生物标志物;和/或

[0052] 表A (vi) 中列出的一种或多种生物标志物。

[0053] 当涉及“正常”疾病状态时,我们包括未患慢性胰腺炎 (ChP) 或急性炎症性胰腺炎 (AIP) 的个体。优选未患任何胰腺疾病或病况的个体。最优选,个体为健康的个体,即他们未患任何疾病或病况的。

[0054] 在可替换或另外的实施方案中,步骤 (b) 包括测量转录因子SOX-11的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量整联蛋白alpha-10的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量EDFR的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量EPFR的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量LSADHR的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量SEAHLR的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量AQHQWDGLLSYQDSLS的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量WTRNSNMNYWLIIRL的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量WDSR的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量DFAEDK的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量FASN蛋白的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量GAK蛋白的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量HADH2蛋白的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量LNVWGK的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量LTEFAK的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量LYEIAR的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量巨核细胞相关的酪氨酸蛋白激酶的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包括测量氧固醇结合蛋白相关蛋白3的表达或由其组成。在可替换或另外的实施方案中,步骤 (b) 包

括测量QEASFK的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量SSAYS R的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量QEASFK的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量TEEQLK的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量TLYVGK的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量FLLMQYGGMDEHAR的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量GIVKLYEDEG的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量GIVKLYEDEG的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量肿瘤坏死因子受体超家族成员3的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量酪氨酸蛋白激酶SYK的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量血管动蛋白(Angiomotin)的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量C-C基序趋化因子2的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量C-C基序趋化因子5的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量CD40配体的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量胰高血糖素样肽-1的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量免疫球蛋白M的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-1 α 的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-1受体拮抗蛋白的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-11的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-12 的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-16的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-18的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-2的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-3的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-4的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-6的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-7的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-9的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量Lewis x的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量淋巴毒素- α 的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量转化生长因子 β -1的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量血管内皮生长因子的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量视觉系统同源盒2的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量HLA-DR/DP的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量载脂蛋白A1的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量载脂蛋白A4的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量载脂蛋白B-100的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量线粒体的ATP合酶 β 亚基的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量 β -半乳糖苷酶的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量组织蛋白酶(Cathepsin) W的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量C-C基序趋化因子13的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量C-C基

序趋化因子7的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量CD40蛋白的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量补体C1q的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量补体C1s的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量补体C3的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量补体C4的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量补体C5的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量补体因子B的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量细胞周期素依赖的激酶2的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量胱抑素-C(Cystatin-C)的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量嗜酸细胞活化趋化因子(Eotaxin)的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量表皮生长因子受体的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量胰高血糖素样肽1受体的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量粒细胞-巨噬细胞集落刺激因子的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量整联蛋白alpha-11的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量细胞间粘附分子1的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量干扰素 γ 的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-1beta的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-10的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-13的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-5的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量白细胞介素-8的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量角蛋白I型细胞骨架19的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量瘦素的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量基膜聚糖(Lumican)的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量丝裂原活化的蛋白激酶1的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量丝裂原活化的蛋白激酶8的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量粘蛋白-1的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量肌间蛋白(Myomesin)-2的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量骨桥蛋白的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量磷脂酰肌醇3-激酶调节性alpha亚基的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量血浆蛋白酶C1抑制物的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量备解素(Properdin)的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量前列腺特异性抗原的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量受体酪氨酸蛋白激酶erbB-2的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量无义转录物调控蛋白3B(Regulator of nonsense transcripts 3B)的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量核糖体蛋白S6激酶alpha-2的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量唾液酸化的Lewis x的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量信号转导接头蛋白2的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量SUMO缀合的酶UBC9的表达或由其组

成。在可替换或另外的实施方案中,步骤(b)包括测量TBC1域家族成员9的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量跨膜肽的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量肿瘤坏死因子alpha的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量肿瘤坏死因子受体超家族成员14的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量酪氨酸蛋白激酶BTK的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量酪氨酸蛋白激酶JAK3的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量酪氨酸蛋白磷酸酶非受体I型(Tyrosine-protein phosphatase non-receptor type 1)的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量泛素羧基末端水解酶同工酶L5的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量泛素缀合的酶E2C的表达或由其组成。在可替换或另外的实施方案中,步骤(b)包括测量FIQTDK的表达或由其组成。

[0055] 在可替换或另外的实施方案中,步骤(b)不包括测量转录因子SOX-11的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量整联蛋白alpha-10的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量EDFR的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量EPFR的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量LSADHR的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量SEAHLR的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量AQHQWDGLLSYQDSLS的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量WTRNSNMNYWLIIRL的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量WDSR的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量DFAEDK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量FASN蛋白的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量GAK蛋白的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量HADH2蛋白的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量LNVWGK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量LTEFAK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量LYEIAR的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量巨核细胞相关的酪氨酸蛋白激酶的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量氧固醇结合蛋白相关蛋白3的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量QEASFK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量SSAYSR的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量QEASFK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量TEEQLK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量TLYVGK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量FLLMQYGGMDEHAR的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量GIVKYLYEDEG的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量GIVKYLYEDEG的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量肿瘤坏死因子受体超家族成员3的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量酪氨酸蛋白激酶SYK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量血管动蛋白的表达或不由其

组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量C-C基序趋化因子2的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量C-C基序趋化因子5的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量CD40配体的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量胰高血糖素样肽-1的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量免疫球蛋白M的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-1 α 的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-1受体拮抗蛋白的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-11的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-12的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-16的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-18的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-2的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-3的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-4的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-6的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-7的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量白细胞介素-9的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量Lewis x的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量淋巴毒素- α 的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量转化生长因子beta-1的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量血管内皮生长因子的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量视觉系统同源盒2的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量HLA-DR/DP的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量载脂蛋白A1的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量载脂蛋白A4的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量载脂蛋白B-100的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量线粒体的ATP合酶beta亚基的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量beta-半乳糖苷酶的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量组织蛋白酶 (Cathepsin) W的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量C-C基序趋化因子13的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量C-C基序趋化因子7的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量CD40蛋白的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量补体C1q的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量补体C1s的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量补体C3的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量补体C4的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量补体C5的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量补体因子B的表达或不由其组成。在可替换或另外的实施方案中，步骤 (b) 不包括测量细胞周期素依赖的激酶2的表达或不由其组成。在可替换

或另外的实施方案中,步骤(b)不包括测量胱抑素-C (Cystatin-C) 的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量嗜酸细胞活化趋化因子(Eotaxin)的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量表皮生长因子受体的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量胰高血糖素样肽1受体的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量粒细胞-巨噬细胞集落刺激因子的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量整联蛋白alpha-11的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量细胞间粘附分子1的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量干扰素 γ 的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量白细胞介素-1beta的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量白细胞介素-10的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量白细胞介素-13的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量白细胞介素-5的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量白细胞介素-8的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量角蛋白I型细胞骨架19的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量瘦素的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量基膜聚糖(Lumican)的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量丝裂原活化的蛋白激酶1的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量丝裂原活化的蛋白激酶8的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量粘蛋白-1的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量肌间蛋白-2的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量骨桥蛋白的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量磷脂酰肌醇3-激酶调节性alpha亚基的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量血浆蛋白酶C1抑制物的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量备解素(Properdin)的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量前列腺特异性抗原的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量受体酪氨酸蛋白激酶erbB-2的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量无义转录物调控蛋白3B(Regulator of nonsense transcripts 3B)的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量核糖体蛋白S6激酶alpha-2的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量唾液酸化的Lewis x的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量信号转导接头蛋白2的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量SUMO缀合的酶UBC9的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量TBC1域家族成员9的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量跨膜肽的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量肿瘤坏死因子alpha的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量肿瘤坏死因子受体超家族成员14的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量酪氨酸蛋白激酶BTK的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量酪氨酸蛋白激酶JAK3的表达或不由其组成。

其组成。在可替换或另外的实施方案中,步骤(b)不包括测量酪氨酸蛋白磷酸酶非受体I型(Tyrosine-protein phosphatase non-receptor type 1)的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量泛素羧基末端水解酶同工酶L5的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量泛素缀合的酶E2 C的表达或不由其组成。在可替换或另外的实施方案中,步骤(b)不包括测量FIQTDK的表达或不由其组成。

[0056] 通过“跨膜肽”或“TM肽”我们表示来源于10TM蛋白的肽,以下SEQ ID NO:1的scFV抗体构建物对其具有特异性(其中CDR序列通过加粗,斜体示出):

[0057] MAEVQLLESGLLVQPGGSLRLSCAASGFT***FSSYGFHWVRQAPG***KGLEWV

SLISWDGGSTYYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAVYYC***ARGTWFD***PWGQGT
LVTSSGGGGSGGGSGGGGSQSVLTQPPSASGTPGQRVTISCS***GSSSNIGNNAVN***WYQQLPGTAPKLLIY
RNNQRPSGVPDRFSGSKSGTSASLAISGLRSEDEADYY***CAAWDDSLSW***VFGGGTKLTVLG

[0058] [SEQ ID NO:1]

[0059] 因此,可以使用该scFV或与该scFV竞争结合10TM蛋白的任意抗体或其抗原结合片段。例如,所述抗体,或其抗原结合片段,可以包含与SEQ ID NO:1中所示的相同的CDR。

[0060] 本领域的技术人员将理解为了纯化的目的,可以使用亲和标签(例如,在C-末端)产生这样的抗体。例如,可以利用以下SEQ ID NO:2的亲和标签:

[0061] DYKDHDGDYKDHDIDYKDDDDKAAAHHHHHH

[0062] [SEQ ID NO:2]

[0063] 在可替换或另外的实施方案中,步骤(b)中测量的一个或多个生物标志物包含选自下组的一个或多个生物标志物,或由其组成:

[0064] 血管动蛋白,Apo-A1,Apo-A4,ATP-5B,BTK,C1 inh,C1q,C3,C5,CD40,CD40L,胱抑素C,嗜酸细胞活化趋化因子,因子B,GAK,GM-CSF,HADH2,IL-11,IL-13,IL-3,IL-4,IL-6,IL-8,IL-9,KSYK-1,LDL,MAPK1,MCP-1,PTP-1B,唾液酸化的Lewis x,Sox11A,TGF-beta1,TNF-alpha,TNFRSF3,UCLH5和UPF3B。

[0065] 在这一实施方案中,所述方法可以用于区分胰腺癌(PC)和非胰腺癌 症(NC)和/或良性胰腺状况(BC)。

[0066] 在可替换或另外的实施方案中,步骤(b)中测量的一个或多个生物标志物包含选自下组的一个或多个生物标志物,或由其组成:

[0067] 血管动蛋白,ATP-5B,C1inh,C1q,C3,C5,CD40,胱抑素C,嗜酸细胞活化趋化因子,因子B,GAK,HADH2,IL-11,IL-13,IL-6,IL-8,LDL和TNF-alpha。

[0068] 在这一实施方案中,所述方法可以用于区分胰腺癌(PC)和非胰腺癌症(NC)。

[0069] 在可替换或另外的实施方案中,步骤(b)中测量的一个或多个生物标志物包含选自下组的一个或多个生物标志物,或由其组成:

[0070] Apo-A1,Apo-A4,BTK,C1inh,C5,CD40L,CIMS,因子B,GM-CSF,HADH2,IL-3,IL-4,IL-4,IL-9,KSYK-1,MAPK1,MCP-1,PTP-1B,唾液酸化的Lewis x,Sox11A,TGF-beta1,TNFRSF3,UCLH5和UPF3B。

[0071] 在这一实施方案中,所述方法可以用于区分胰腺癌(PC)和良性胰腺状况(BC)。

[0072] 在可替换或另外的实施方案中,所述方法包含以下步骤,或由以下步骤组成:步骤(a)和(b),可选的,步骤(c)和(d),可选的,步骤(e)和(f),以及另外的步骤:

[0073] g) 基于步骤(b)中测量的一种或多种生物标志物的表达确定胰腺癌的存在。

[0074] 在可替换或另外的实施方案中基于表2中鉴定的趋势(上或下调)。

[0075] 在可替换或另外的实施方案中步骤(b)和,在存在的情况下,步骤(d)和(f)通过使所述待测试的样品与步骤(b)中测量的一种或多种生物标志物的结合部分接触来进行,例如,作为以下确定的第一结合部分。

[0076] 在可替换或另外的实施方案中,所述方法包括使用表3中列出的生物标志物或由其组成。

[0077] 一般来说,诊断在具有至少为0.55的ROC AUC时进行,例如具有至少0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95、0.96、0.97、0.98、0.99的ROC AUC或1.00的ROC AUC。优选,诊断在具有至少0.85的ROC AUC时进行,以及最优选ROC AUC为1。

[0078] 典型地,使用支持向量机(SVM),例如可得自 <http://cran.r-project.org/web/packages/e1071/index.html>的那些(例如,e10711.5-24)来进行诊断。但是也可以使用任何其他合适的手段。

[0079] 支持向量机(SVM)是用于分类和回归的一组相关的经监督的学习方法。给定一组训练实例,每一个被标记为属于两类中的一类,SVM训练算法构建了用于预测新的实例是否落入一类或另一类的模型。直观地,SVM模型将实例呈现为空间中的点,被绘制成使得不同分类的实例被尽量宽的明显的间隙分开。然后,新的实例被绘制到相同的空间中并基于它们落入间隙的那一侧来预测其属于哪一类。

[0080] 更正式的是,支持向量机在高维或无穷维空间构建超平面或一组超平面,其可被用于分类、回归或其他任务。直观地,良好的分离通过与任何类别的最近训练数据点具有最大距离(被称为功能边缘)的超平面来实现,因为通常边缘越大分类机的泛化错误越低。关于SVM的更多信息,可以参考例如Burges,1998,Data Mining and Knowledge Discovery, 2:121-167。

[0081] 在本发明的可替换或另外的实施方案中,在使用来自已知疾病状态的个体(例如,已知个体具有胰腺癌,已知个体具有急性炎症性胰腺炎,已知个体具有慢性胰腺炎或已知个体是健康的)的生物标志物概貌进行本发明的方法之前,对SVM进行‘训练’。通过运行这样的训练样品,SVM能够学习什么样的生物标志物概貌与胰腺癌相关。一旦训练过程结束,然后SVM能够确认是否测试的样品的生物标志物是否来自具有胰腺癌的个体。

[0082] 但是,通过在SVM中用必须的训练参数进行预编程可以省略该训练步骤。例如,可以根据使用在表B中详细描述SVM算法,基于在表A中列出的任意或全部生物标志物的测定值,根据已知的SVM参数进行诊断。

[0083] 本领域技术人员应当理解通过利用适当选择的数据(即,来自具有已知胰腺癌状态的个体的生物标志物的测定值)来训练SVM机,可以针对表A中列出的生物标志物的任何组合来确定适当的SVM参数。或者,表2和3数据可以用于根据本领域已知的任何其它适合的统计方法确定具体的胰腺癌相关的疾病状态。

[0084] 优选地,本发明的方法具有至少60%的准确度,例如61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、

81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或100%的准确度。

[0085] 优选地,本发明的方法具有至少60%的灵敏度,例如61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的灵敏度。

[0086] 优选地,本发明的方法具有至少60%的特异性,例如61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的特异性。

[0087] 通过“准确度”我们表示方法的正确结果的比例,通过“灵敏度”我们表示被正确分类为阳性的所有PaC阳性样品的比例,以及通过“特异性”我们表示被正确分类为阴性的所有PaC阴性样品的比例。

[0088] 在可替换或另外的实施方式中,未患胰腺癌的个体未患胰腺癌 (PaC)、慢性胰腺炎 (ChP) 或急性炎症性胰腺炎 (AIP)。更优选地,健康个体未患任何胰腺疾病或病况。甚至更优选地,未患胰腺癌的个体未患任何疾病或病况。最优选的,未患胰腺癌的个体是健康个体。通过“健康个体”我们包括被本领域技术人员认为身体强壮以及没有身体疾病的个体。

[0089] 但在另外的实施方式中,未患胰腺癌的个体是患慢性胰腺炎的。在仍另外的实施方式中,未患胰腺癌的个体是患急性炎症性胰腺炎的。

[0090] 如上所述,本发明的方法用于检测个体中胰腺癌的存在。在一个备选或另外的实施方式中,胰腺癌选自下组:腺癌 (adenocarcinoma)、腺鳞癌 (adenosquamous carcinoma)、印戒细胞癌 (signet ring cell carcinoma)、肝样癌 (hepatoid carcinoma)、胶样癌 (colloid carcinoma)、未分化癌和伴有破骨细胞样巨细胞 (osteoclast-like giant cells) 的未分化癌。优选的胰腺癌是胰腺腺癌 (pancreatic adenocarcinoma)。更优选的胰腺癌是胰腺导管腺癌 (pancreatic ductal adenocarcinoma),也称为胰腺外分泌肿瘤 (exocrine pancreatic cancer)。

[0091] 在进一步的实施方式中,使用能够与一种或多种生物标志物结合的第一结合试剂 (即使用一种或多种第一结合试剂,其中每种结合试剂能够特异性结合一种或多种生物标志物之一) 来进行步骤 (b)、(d) 和/或步骤 (f)。本领域技术人员应当理解第一结合试剂可包括下述物质或由下述物质构成:针对蛋白质生物标志物中的一种具有特异性的单一类型或大量不同种类,每种对于不同的蛋白质生物标志物具有特异性。

[0092] 基于它们结合给定的基序 (motif) 的能力,合适的结合试剂 (也被称为结合分子) 可以选自如下所述的库。

[0093] 至少一种类型的结合试剂,以及更通常地全部类型可包括下述物质或由下述物质构成:抗体或其抗原结合片段,或其变体。

[0094] 抗体的制备和利用的方法在本领域中是已知的,例如参见Antibodies: A Laboratory Manual, 1988, Harlow & Lane, Cold Spring Harbor Press, ISBN-13: 978-0879693145, Using Antibodies: A Laboratory Manual, 1998, Harlow & Lane, Cold Spring Harbor Press, ISBN-13: 978-0879695446 和 Making and Using Antibodies: A Practical

Handbook, 2006, Howard&Kaser, CRC Press, ISBN-13: 978-0849335280 (通过参考将其公开内容并入本文)。

[0095] 因此, 片段可以包括一个或多个重链可变域 (V_H) 或轻链可变域 (V_L)。例如, 术语抗体片段包括 Fab-类分子 (Better 等人 (1988) Science 240, 1041); Fv 分子 (Skerra 等人 (1988) Science 240, 1038); 单链 Fv (ScFv) 分子, 其中 V_H 和 V_L 伴侣域 (partner domain) 通过柔性的寡肽连接 (Bird 等人 (1988) Science 242, 423; Huston 等人 (1988) Proc. Natl. Acad. Sci. USA 85, 5879) 以及包括分离的 V 域的单域抗体 (dAbs) (Ward 等人 (1989) Nature 341, 544)。

[0096] 术语“抗体变体”包括任何合成抗体、重组抗体或抗体杂合物, 例如但不限于: 通过免疫球蛋白轻链和/或重链可变区和/或恒定区的噬菌体展示制备的单链抗体分子, 或在本领域人员已知的免疫测定方法中能够与抗原结合的其他免疫相互作用分子。

[0097] 可以在 Winter&Milstein (1991) Nature 349, 293-299 中找到涉及抗体片段合成的技术的一般性综述, 其保留它们的特异性结合位点。

[0098] 分子库例如抗体库 (Clackson 等人, 1991, Nature 352, 624-628; Marks 等人, 1991, J Mol Biol 222 (3): 581-97)、肽库 (Smith, 1985, Science 228 (4705): 1315-7)、表达的 cDNA 库 (Santi 等人 (2000) J Mol Biol 296 (2): 497-508)、在抗体框架之外的其他支架例如亲和体 (affibody) 上的库 (Gunneriusson 等人, 1999, Appl Environ Microbiol 65 (9): 4134-40) 或基于适体 (aptamer) 的库 (Kenan 等人, 1999, Methods Mol Biol 118, 217-31) 可以用作从中选择用于本发明方法的对于给定的基序具有特异性的结合分子的源。

[0099] 分子库可以在原核生物 (Clackson 等人, 1991, 同上; Marks 等人, 1991, 同上) 或真核生物细胞 (Kieckhefer 等人, 1999, Proc Natl Acad Sci USA, 96 (10): 5651-6) 中体内表达, 或可以不涉及细胞体外表达 (Hanes&Pluckthun, 1997, Proc Natl Acad Sci USA 94 (10): 4937-42; He&Taussig, 1997, Nucleic Acids Res 25 (24): 5132-4; Nemoto 等人, 1997, FEBS Lett, 414 (2): 405-8)。

[0100] 在使用基于蛋白质的库的情况下, 通常编码潜在结合分子的库的基因被包装到病毒中并且在病毒的表面展示潜在结合分子 (Clackson et al, 1991, 同上; Marks 等人, 1991, 同上; Smith, 1985, 同上)。

[0101] 目前最常见的这样的系统是丝状噬菌体, 在它们的表面展示抗体片段, 抗体片段表达为融合至噬菌体的次要衣壳蛋白 (Clackson 等人, 1991, 同上; Marks 等人, 1991, 同上)。但是, 还已经使用了利用其他病毒 (EP 39578)、细菌 (Gunneriusson 等人, 1999, 同上; Daugherty 等人, 1998, Protein Eng 11 (9): 825-32; Daugherty 等人, 1999, Protein Eng 12 (7): 613-21)、以及酵母 (Shusta 等人, 1999, J Mol Biol 292 (5): 949-56) 展示的其他系统。

[0102] 另外, 还开发了利用多肽产物连接至其编码 mRNA 的展示系统, 称为核糖体展示系统 (Hanes&Pluckthun, 1997, 同上; He&Taussig, 1997, 同上; Nemoto 等人, 1997, 同上), 或将多肽产物连接至编码 DNA (参见美国专利 No. 5,856,090 和 WO 98/37186) 的展示系统。

[0103] 当从库中选择了潜在结合分子时, 通常使用具有限定的基序的一个或几个筛选肽 (selector peptide)。可以使用提供结构、降低肽的柔性或能够与结合分子相互作用的带电、极性的或疏水侧链的氨基酸残基用于设计筛选肽的基序。

[0104] 例如:

[0105] (i) 脯氨酸由于其侧链结合至 α 碳以及氮两者可以稳定肽结构;

[0106] (ii) 苯丙氨酸、酪氨酸和色氨酸具有芳香侧链并且是非常疏水的,而亮氨酸和异亮氨酸具有脂肪侧链并且也是非常疏水的;

[0107] (iii) 赖氨酸、精氨酸和组氨酸具有碱性侧链并且在中性pH下会带正电,而天冬氨酸和谷氨酸具有酸性侧链并且在中性pH下会带负电;

[0108] (iv) 天冬酰胺和谷氨酰胺在中性pH时为中性的,但它们含有可以形成氢键的酰胺基团;

[0109] (v) 丝氨酸、苏氨酸和酪氨酸侧链含有羟基,其可以形成氢键。

[0110] 通常,结合试剂的选择可以涉及利用阵列技术和系统来根据结合分子的类型分析结合点。

[0111] 在一个备选或另外的实施方式中,第一结合试剂被固定在表面上(例如,在多孔板或阵列上)。

[0112] 抗体的重链可变域(V_H)和轻链可变域(V_L)参与抗原识别,这是由早期的蛋白酶消化实验首先认识到的事实。后来在啮齿类抗体的“人源化”工作中也得到确认。啮齿类来源的可变域可以融合至人类来源的恒定域,使得得到的抗体保留对啮齿类亲本抗体的抗原特异性(Morrison等人(1984)Proc.Natl.Acad.Sci.USA 81,6851-6855)。

[0113] 关于抗原特异性由多个可变域赋予且其独立于恒定域的事实是从多个涉及抗体片段的细菌表达的实验知晓的,这些实验都包括一个或多个可变域。这些分子包括Fab-类分子(Better等人,(1988)Science 240,1041);Fv分子(Skerra 等人(1988)Science 240,1038);单链Fv(ScFv)分子,其中通过柔性的寡肽连接 V_H 和 V_L 伴侣区域(Bird等人(1988)Science 242,423;Huston等人(1988)Proc.Natl.Acad.Sci.USA 85,5879)以及包括分离的V域的单域抗体(dAbs)(Ward等人(1989)Nature 341,544)。可以在Winter&Milstein(1991)Nature 349,293-299中找到涉及保留它们的特异性结合位点的抗体片段的合成的技术的一般性综述。

[0114] 通过“ScFv分子”我们表示指如下分子,其中 V_H 和 V_L 伴侣域通过柔性的寡肽连接。

[0115] 使用抗体片段而非整个抗体的优点是数倍的。片段的较小尺寸可导致改善的药理性质,例如更好的实体组织渗透性。除去了完整抗体的效应器功能,例如补体结合。Fab、Fv、ScFv和dAb抗体片段可以全部在大肠杆菌中表达并由大肠杆菌中分泌,从而使得容易进行所述片段的大量生产。

[0116] 完整抗体以及 $F(ab')_2$ 片段为“二价的”。“二价的”是指所述抗体和 $F(ab')_2$ 片段具有两个抗原结合位点。相对地,Fab、Fv、ScFv和dAb片段是单价的,仅具有一个抗原结合位点。

[0117] 抗体可以是单克隆或多克隆的。可以通过已知的技术来制备合适的单克隆抗体,例如在“Monoclonal Antibodies:A manual of techniques”,H Zola(CRC Press,1988)和“Monoclonal Hybridoma Antibodies:Techniques and applications”,J G R Hurrell(CRC Press,1982)中描述的那些,在此通过参考将它们并入本文。

[0118] 在可替换的或另外的实施方案中,第一结合试剂被固定在表面上(例如在多孔板或阵列上)。

[0119] 使用抗体片段而非整个抗体的优点是数倍的。片段的较小尺寸可导致改善的药理

性质,例如更好的实体组织渗透性。除去了完整抗体的效应器功能,例如补体结合。Fab、Fv、ScFv和dAb抗体片段可以全部在大肠杆菌中表达并由大肠杆菌中分泌,从而使得容易进行所述片段的大量生产。

[0120] 完整抗体以及F(ab')₂片段为“二价的”。“二价的”是指所述抗体和F(ab')₂片段具有两个抗原结合位点。相对地,Fab、Fv、ScFv和dAb片段是单价的,仅具有一个抗原结合位点。

[0121] 抗体可以是单克隆或多克隆的。可以通过已知的技术来制备合适的单克隆抗体,例如在“Monoclonal Antibodies:A manual of techniques”,H Zola (CRC Press,1988)和“Monoclonal Hybridoma Antibodies:Techniques and applications”,J G R Hurrell (CRC Press,1982)中描述的那些,在此通过参考将它们并入本文。

[0122] 因此,第一结合试剂可以包括抗体或其抗原结合片段或由抗体或其抗原结合片段构成。优选抗体或其抗原结合片段是重组抗体或其抗原结合片段。抗体或其抗原结合片段可以选自下组:scFv、Fab、和免疫球蛋白分子的结合域。

[0123] 第一结合试剂可以固定在表面上。

[0124] 可以用可检测部分标记检测样品中的一种或多种生物标志物。

[0125] 通过“可检测部分(detectable moiety)”我们包括如下含义,该部分可以被检测以及可以确定该部分的相对量和/或位置(例如,在阵列上的位置)。

[0126] 合适的可检测部分是本领域公知的。

[0127] 因此,该可检测部分可以是当暴露于特定的条件下时,可以被检测的荧光和/或发光和/或化学发光部分。例如,荧光部分可需要被暴露于在特定波长和强度的放射(例如光)下以引起荧光部分的激发,从而使得其发射出可以被检测的在特定波长的可检测的荧光。

[0128] 或者,可检测部分可以是酶,该酶能够将(优选检测不到的)底物变换成可视的和/或可检测的可检测产物。合适的酶的实例可以参考例如ELISA分析在下文中详细讨论。

[0129] 或者,可检测部分可以是用于成像的放射性原子。合适的放射性原子包括用于闪烁显像研究的^{99m}Tc和¹²³I。其他容易的可检测部分包括,例如用于磁共振成像(MRI)的自旋标志物,诸如又是¹²³I、¹³¹I、¹¹¹In、¹⁹F、¹³C、¹⁵N、¹⁷O、钆、锰或铁。显然,待检测的试剂(例如,在本文中描述的测试样品和/或对照样品中的一种或多种生物标志物和/或用于检测选择的蛋白质的抗体分子)必须具有足够的合适的原子同位素以使得可以容易地检测可检测部分。

[0130] 可以按照已知的方式将放射-或其他标记组合到本发明的试剂(即,本发明方法的样品中存在的蛋白质和/或本发明的结合试剂)中。例如,如果结合部分是多肽,其可以通过生物合成或可以通过使用合适的氨基酸前体(涉及,例如用氟-19代替氢)通过化学氨基酸合成来合成。可以例如通过半胱氨酸残基将标记物,诸如^{99m}Tc、¹²³I、¹⁸⁶Rh、¹⁸⁸Rh和¹¹¹In附着到结合部分中。可以经由赖氨酸残基附着钇-90。可以使用IODOGEN方法(Fraker等人(1978) Biochem.Biophys.Res.Comm.80,49-57)来掺入¹²³I。参考文献(“Monoclonal Antibodies in Immunoscintigraphy”,J-F Chatal,CRC Press,1989)详细地描述了其他方法。用于将其他可检测部分(诸如酶、荧光、发光、化学发光或放射性部分)偶联至蛋白质的方法是本领域所熟知的。

[0131] 优选的,用可检测部分标记对照样品(一个或多个)中的一种或多种生物标志物。

可检测部分可以选自下组:荧光部分;发光部分(luminescent moiety);化学发光部分(chemiluminescent moiety);放射性部分;酶部分。但是,优选可检测部分为生物素。

[0132] 在另外的实施方式中,使用包括能够与一种或多种生物标志物结合的第二结合试剂的测定法进行步骤(b)、(d)和/或步骤(f),该第二结合试剂包括可检测部分。优选地,第二结合试剂包括或由以下物质构成:抗体或其抗原结合片段。优选地,抗体或其抗原结合片段为重组抗体或其抗原结合片段。最优选地,抗体或其抗原结合片段选自下组:scFv、Fab、以及免疫球蛋白分子的结合域。在一个备选的或另外的实施方式中,可检测部分选自下组:荧光部分;发光部分;化学发光部分;放射性部分和酶部分。优选地,可检测部分为荧光部分(例如,Alexa Fluor染料如Alexa647)。

[0133] 在一个备选的或另外的实施方式中,本发明第一方面的方法包括 ELISA(酶联免疫吸附测定)或由ELISA构成。

[0134] 用于检测血清或血浆蛋白质的优选的分析包括酶联免疫吸附测定(ELISA)、放射免疫分析(RIA)、免疫放射分析(IRMA)以及免疫酶分析(IEMA),包括使用单克隆和/或多克隆抗体的三明治分析。David等人在美国专利No.4,376,110和No.4,486,530中描述了示例性的三明治分析,通过参考并入本文。在载玻片(slide)上的细胞抗体染色可以用于在细胞学实验室诊断测试中已知的方法中,如本领域技术人员所述熟知的。

[0135] 通常,该分析为ELISA(酶联免疫吸附测定),其通常涉及使用产生有色的反应产物的酶,通常在固相分析中。酶诸如辣根过氧化物酶和磷酸酶已经被广泛使用。放大磷酸酶反应的方式是使用NADP作为底物产生NAD,其此时将作为用于第二酶系统的辅酶。来自大肠杆菌的焦磷酸酶提供了良好的偶联物(conjugate),因为该酶不存在于组织中,其是稳定的并且提供良好的反应颜色。还可以使用基于酶诸如萤光素酶的化学-发光系统。

[0136] ELISA方法是本领域公知的,例如参见The ELISA Guidebook(Methods in Molecular Biology),2000,Crowther,Humana Press,ISBN-13:978-0896037281(通过参考将其公开内容并入)。

[0137] 经常使用与维生素生物素的缀合,因为生物素可以通过其与酶连接的抗生素蛋白或链霉亲和素(生物素以大的特异性和亲和性与其结合)的反应容易地被检测。

[0138] 然而,备选地,使用阵列进行步骤(b)、(d)和/或步骤(f)。阵列本身是本领域所熟知的。通常它们是由具有分开的间距(即离散的)区域(“点”)的线性的或二维结构形成的,每个均具有有限的面积并形成在固体支撑体的表面上。阵列还可以是微珠结构的,其中每个微珠可以通过分子密码或颜色密码识别,或在连续流体中被识别。还可以序贯实施分析,其中样品通过一系列点,每个点从溶液中吸附该类分子。固体支撑物通常为玻璃或聚合物,最常用的聚合物是纤维素、聚丙烯酰胺、尼龙、聚苯乙烯、聚氯乙烯或聚丙烯。固体支撑体可以为管、珠子、盘、硅芯片、微孔板、聚偏氟乙烯(PVDF)膜、硝基纤维素膜、尼龙膜、其他多孔膜、非多孔膜(例如,塑料、聚合物、有机玻璃(perspex)、硅等)、大量聚合物针(polymeric pin)、或大量微滴定孔、或适合用于固定蛋白质、多核苷酸和其他合适的分子和/或进行免疫分析的任何其他表面形式。结合方法是本领域所熟知的并且通常包括交联共价结合或物理吸附蛋白质分子、多核苷酸或等至固体支撑体。通过使用熟知的技术,例如接触式或非接触式印刷、掩膜或光刻,可以确定每个点的位置。其综述参见Jenkins,R.E.,Pennington,S.R.(2001,Proteomics,2,13-29)和Lal等人(2002,Drug Discov Today 15;7(18Suppl):

S143-9)。

[0139] 通常阵列称为微阵列。“微阵列”是指如下含义：具有至少约 $100/\text{cm}^2$ ，优选至少约 $1000/\text{cm}^2$ 离散区域密度的区域的阵列。微阵列中的区域具有通常的维度，例如直径在约 $10\sim 250\mu\text{m}$ 的范围，并通过大致相同的距离与阵列中的其他区域分开。阵列还可以是宏阵列(macroarray)或纳米阵列。

[0140] 一旦适合的结合分子(如上所述)被鉴定和分离出，本领域技术人员可以使用分子生物学领域中熟知的方法来制造阵列。

[0141] 因此，阵列可以是基于珠的阵列或基于表面的阵列。优选地，阵列选自下组：宏阵列、微阵列和纳米阵列。

[0142] 在可替换或另外的实施方式中，根据本发明的第一方面的方法包括：

[0143] (i) 利用生物素标记样品中存在的生物标志物；

[0144] (ii) 使生物素-标记的蛋白质与阵列接触，该阵列包括固定在其表面上的离散区域的大量scFv，该scFv对于表III中的一种或多种蛋白质具有特异性；

[0145] (iii) 使已固定的scFv与包括荧光染料的链霉亲和素偶联物接触；以及

[0146] (iv) 检测在阵列表面上的离散区域的染料存在，

[0147] 其中阵列表面上的染料的出现是来自表A的生物标志物在样品中表达的指示。

[0148] 在可替换的实施方式中，步骤(b)、(d)和/或(f)包括检测编码一种或多种生物标志物的核酸分子的表达。优选核酸分子为cDNA分子或mRNA分子。最优选核酸分子为mRNA分子。

[0149] 因此，在步骤(b)、(d)和/或(f)中检测一种或多种生物标志物的表达可使用选自下组的方法进行，该组由Southern杂交、Northern杂交、聚合酶链式反应(PCR)、反转录PCR(RT-PCR)、定量实时PCR(qRT-PCR)、纳米阵列、微阵列、宏阵列、放射自显影和原位杂交构成。优选地，使用DNA微阵列确定步骤(b)中一种或多种生物标志物的表达。

[0150] 在一个备选或另外的实施方式中，在步骤(b)、(d)和/或(f)中检测一种或多种生物标志物的表达通过使用一种或多种结合部分来进行，每个结合部分各自能够选择性地结合编码在表A中鉴定的一种生物标志物的核酸分子。

[0151] 在另外的实施方式中，一种或多种结合部分各自包括核酸分子或各自由核酸分子分子构成。因此，一种或多种结合部分可各自包括下述物质或由下述物质构成：DNA、RNA、PNA、LNA、GNA、TNA或PMO。但优选一种或多种结合分子各自包括DNA或由DNA构成。

[0152] 优选地，一种或多种结合部分的长度为 $5\sim 100$ 个核苷酸。更优选地，一种或多种结合部分的长度为 $15\sim 35$ 个核苷酸。还更优选地，结合部分包括可检测部分。

[0153] 在另外的实施方式中，可检测部分选自下组：荧光部分；发光部分；化学发光部分；放射性部分(例如放射性原子)；以及酶部分。优选地，可检测部分包括放射性原子或由放射性原子构成。放射性原子可选自下组：锝-99m、碘-123、碘-125、碘-131、铟-111、氟-19、碳-13、氮-15、氧-17、磷-32、硫-35、氘、氚、镱-186、镱-188和钇-90。

[0154] 但是，结合部分的可检测部分可以是荧光部分(例如，Alexa Fluor染料，例如Alexa647)。

[0155] 在可替换或另外的实施方式中，步骤(b)、(d)和/或(f)中提供的样品选自下组：未分级的血、血浆、血清、组织液、胰组织、胰汁、胆汁和尿液。优选地，步骤(b)、(d)和/或(f)中

提供的样品选自下组：未分级的血、血浆和血清。更优选地，步骤(b)、(d)和/或(f)中提供的样品为血浆。在另外优选的实施方式中，步骤(b)、(d)和/或(f)中提供的样品为血清。

[0156] 在可替换的实施方案中，本发明的第一个方面的方法是用于从胰头癌区分胰体癌和/或胰尾癌，包含以下步骤，或由以下步骤组成：

[0157] a) 提供来自个体的待测试的样品；

[0158] b) 通过测量测试样品中选自下组的至少一种生物标志物确定所述测试样品的生物标志物签名：表A(i)，(ii)，(iv)和/或(vi)；

[0159] 其中所述测试样品中选自表A(i)，(ii)和/或(iii)的至少一种生物标志物的表达对于胰体癌和/或胰尾癌，或胰头癌的存在是指示性的。在这一实施方案中，所述测试样品可以是已经鉴定为胰腺癌样品(例如，使用上文所述的方法)。或者，所述测试样品可以没有鉴定为胰腺癌样品。可以适当地选择阳性和阴性对照样品。

[0160] 在可替换或另外的实施方案中，在胰腺癌的鉴定后适当地治疗所述患者。例如，肿瘤可以手术去除(切除)，使用化学疗法(即，抗肿瘤药物)和/或使用放射疗法治疗。因此，本发明包括治疗患有胰腺癌的人的方法，其中所述患者使用本发明的第一个方面的方法鉴定为患有胰腺癌。

[0161] 本发明的第二方面提供用于检测个体中胰腺癌存在的阵列，其包括如本发明第一方面所限定的一种或多种结合试剂。

[0162] 在上文中讨论了适合用于本发明的方法中的阵列。

[0163] 优选所述一种或多种结合试剂能够与表A中定义的所有蛋白质结合。

[0164] 本发明的第三方面提供选自本发明第一方面定义的组的一种或多种生物标志物的用途，用作确定个体中胰腺癌存在的诊断标志物。优选地，表III中定义的所有蛋白用作确定个体中胰腺癌存在的诊断标志物。

[0165] 本发明的第四方面提供用于诊断胰腺癌存在的试剂盒，其包括：

[0166] A) 一个或多个根据本发明第一方面的第一结合试剂或根据本发明第二方面的阵列；以及

[0167] B) 用于进行根据本发明第一方面的方法的说明。

[0168] 本发明的第五方面提供选自本发明第一方面定义的组的一种或多种结合剂/部分的用途，用于确定个体中胰腺癌的存在。优选地，表III中定义的所有蛋白的结合试剂/部分用作诊断标志物用于确定个体中胰腺癌的存在。

[0169] 优选的，将参考下述表和图来描述体现本发明的某些方面的优选的、非限制性示例：

[0170] 图1. 根据诊断给样品着色的PCA(红色=PC, 黄色=BC, 蓝色=NC)。该数据已过滤至 $p=1e-10$ (63种抗体)。

[0171] 图2. 反向消除分析(Backward elimination analysis)。第一训练组中每轮抗体消除后的Kullback-Leibler(K-L)误差标绘为(A)PC vs NC和(B)PC vs BC。(C)在十个不同的训练/测试组的对中从25-抗体SVM模型生成的AUC值的盒形图。

[0172] 图3. 根据肿瘤的位点给样品着色的PCA(红色=头肿瘤, 黄色=体和尾肿瘤, 蓝色=正常对照)。该数据已过滤至 $p=1e-10$ (46种抗体)。

[0173] 图4. 在样品杂交和扫描后，具有命名为A1-6和B1-7的13个亚阵列的微阵列的代表

性图像。放大了亚阵列B4,示出了阵列布局设计,具有33x31个点。该阵列由三个区段组成,所述三个区段由BSA标记的印刷行分开(行1、11、21和31)。每种抗体被印在三个重复的点,每个区段中一个,分布于每个区段中的不同位置。

[0174] 图5.作为PCA绘图示出的数据的预处理。A)根据五轮(天)的分析着色的原始数据,命名为R1-R5。B)标准化的数据表明基于几轮分析的集群(cluster)已经消除。

[0175] 图6.通过生物标志物鉴定为疾病的通路分析,基于来自完整组抗体的数据,其中有对于PC vs NC的对应的p值和倍数变化。

实施例

[0176] 材料和方法

[0177] 样品

[0178] 知情同意后,在西班牙的五个不同地点收集了338份血清样品,来自患有胰腺癌的患者(PC, n=156),良性对照(BC, n=152),和来自正常对照(NC, n=30)(表1)。整个样品组在单个场合标记,使用先前优化的方案[32,33]。简短的说,将粗制样品以1:45稀释到PBS中,产生约2mg/ml的蛋白浓度,并使用15:1的摩尔过剩的生物素标记蛋白,使用0.6mM EZ-Link Sulfo-NHS-LC-Biotin(Pierce, Rockford, IL, USA)。通过相对PBS透析72小时去除未结合的生物素。将标记的样品等分并储存在-20℃。

[0179] 抗体

[0180] 针对98个已知血清抗原和31个4-6个氨基酸基序(这里命名为CIMS 1-31)[34]的总计293种人类重组scFv抗体用作抗体微阵列的内容物(见表4使用的抗体的完整列表)。所述抗体在15mL大肠杆菌培养物中产生,并使用MagneHis蛋白纯化系统(Promega, Madison, WI, USA)和KingFisher96机械装置(Thermo Fisher Scientific, Waltham, MA, USA)从外周胞质中以300μL纯化。使用Zeba 96孔脱盐旋板(Pierce)将洗脱缓冲液交换为PBS。使用NanoDrop(Thermo Scientific, Wilmington, DE, USA)测量蛋白产出并使用10%SDS-PAGE(Invitrogen, Carlsbad, CA, USA)检查纯度。

[0181] 抗体微阵列

[0182] 在黑色MaxiSorp载玻片(NUNC, Roskilde, Denmark)上使用非接触性打印机(SciFlexarrayer S11, Scienion, Berlin, Germany)产生了抗体微阵列。将十三个相同的亚阵列印制在每个载玻片上,每个阵列由31x33个点组成,点的直径130μm,具有200μm的点-到-点中心距离。每个亚阵列由三区段组成,由标记的BSA的印制行分开(图4)。每种抗体以三个重复印制,每个亚阵列区段中一个,放置在不同的位置以确保充分的重复性。每天印制八个载玻片,即104个亚阵列,持续五天。印制过夜进行,且载玻片在第二天立即用于阵列分析。

[0183] 将每个载玻片安装到杂交垫圈中(Schott, Jena, Germany)并使用PBSMT(1% (w/v) 牛奶, 1% (v/v) Tween-20溶于PBS)封闭一小时。同时,标记的样品的等分试样在冰上解冻并以1:10稀释于PBSMT,在96孔板中。在从稀释板添加样品之前,载玻片使用PBST(0.05% (v/v) Tween-20溶于PBS)洗涤四次。将样品在摇床上孵育两小时,使用PBST洗涤四次,在摇床上与溶于PBST的1μg/mL链霉亲和素-Alexa孵育一小时,接着再使用PBST洗涤四次。最后,将载玻片从杂交室卸下,快速浸没在dH₂O中,并在N₂流下干燥。立即使用共聚焦微阵列扫描器

(PerkinElmer Life and Analytical Sciences, Wellesley, MA, USA) 以10 μ m分辨率扫描载玻片, 使用60% PMT增益和90% 激光功率。使用ScanArray Express软件版本4.0 (PerkinElmer Life and Analytical Sciences) 量化信号强度, 使用固定的循环选项(fixed circle option)。减去局部背景的信号强度值用于数据分析。

[0184] 数据预处理

[0185] 使用三个重复点的平均值, 除非任意重复CV超过平均值的15%, 在这种情况下将其排除, 并转而使用两个剩下重复的平均值。平均重复CV为8.3% ($\pm 5.5\%$)。应用15%的截止CV, 70%的数据值从三个重复点计算, 且剩下的30%从两个重复。

[0186] 为了对标准化策略评价和对方差初步分析, 使用3D主成分分析(Qlucore, Lund, Sweden) 使数据可视化。认为一个样品(慢性胰腺炎) 是异常值(几乎没有获得任何信号) 并从数据中排除。对log10原始数据的主成分分析表明在以下项之间没有能够发现差异i) 载玻片上的样品亚阵列定位, ii) 患者性别, iii) 患者年龄, 和iv) 血清样品收集中心。轻微的系统差异仅能够在分析的天数之间观察到(1-5轮, 可能是由于阵列印制期间湿度的小差异, 特别是对于第1天, 见图4), 其通过标准化中和。该数据在两个步骤中标准化。首先, 不同轮(天) 的分析之间的差异使用减去组平均值的策略消除。在每轮分析中计算每个抗体的平均强度, 并从单个值减去, 从而零中心化(zero centering) 数据。将来自每个抗体的总体平均信号加到每个相应的数据点以避免负值。第二, 阵列-到-阵列的差异(例如, 内在样品背景荧光差异(见图4)) 通过计算每个亚阵列的比例因子处理, 基于具有最低CV的20%的抗体, 如先前已经描述的[23, 35]。数据的标准化在PCA图中可视化(图5)。

[0187] 数据分析

[0188] 分析基于PC vs. NC和PC vs. BC的两组比较, 使用PCA, 分级群聚(hierarchical clustering), Student's t检验和倍数变化, 以及多组ANOVA(PC, NC和CP) (Qlucore)。使用线性内核将约束成本(cost of constraints) 设置为一(默认值) 在R中(www.r-project.org) 进行了支持向量机(SVM) 分析。将数据随机分成为训练和测试组, 其中来自每个组的三分之二的样品在训练组中而剩下的三分之一的样品在测试组中。使用训练组数据应用反向消除算法(backward elimination algorithm), 每次排除一种抗体并迭代消除抗体, 当在分类分析中获得最小Kullback-Leibler差异(divergence) 时将其去除, 如先前所描述的[31]。待消除的最后25种抗体用于在训练组中创建分类模型, 其直接应用于对应的测试组中。ROC-曲线下的面积(AUC) 用作测试组中签名的准确性的测量。在十个不同的, 随机生成的训练和测试组对中将该程序重复十次。最终, 基于十个训练组中消除的顺序, 对每种抗体给予分数。该分数从消除程序的平均持续时间计算(待消除的第一种抗体=1, 待消除的最后一种抗体=293)。灵敏度(SN) 和特异性(SP) 从SVM预测值的最优阈值计算, 这里定义为SN+SP的最高总和。最后, 使用MetaCore进行通路分析(Thomson Reuters, New York, NY, USA)。

[0189] 结果

[0190] 差异蛋白表达分析

[0191] 对变化的分析揭示了血清样品中在癌症和对照之间具有强差异信号传到模式的大量抗体。实际上, 多组ANOVA表明75%的抗体在血清数据中表现出 $p < 0.001$ 的显著水平。主成分分析证明癌症样品明显比良性对照从正常更加分化(图1)。

[0192] 对于每个亚组比较,表现出最低p值的25种抗体示于表2中。PC vs.NC比较揭示了强上调和下调分析物,而PC vs.BC分析鉴定了具有较为温和倍数变化的分析物,且其几乎都在PC相对BC中上调(表2)。对于PC vs.NC,表现出最高差异表达水平的分析物是GAK, IL-6, LDL, 和MAPK8 ($p \leq 4 \cdot 10^{-20}$),而对于PC vs.BC鉴定了蛋白显著不同组,其中胱抑素C, IL-13和IL-1 α 表现出最低的p值 ($p \leq 3 \cdot 10^{-10}$)。

[0193] 基于整组抗体 ($n=293$) 和对于PC vs.NC的对应p值和倍数变化的通路分析表明,当通过生物标志物搜索疾病时,涉及胰岛素产生(胰岛功能亢进和胰岛素抗性)的疾病作为排名靠前(图6)。与代谢相关的症状,例如葡萄糖代谢病症,肥胖,和超重也显著相关,以及糖尿病I型和II型,克罗恩氏病(Crohn's disease),肝炎,自身免疫和感染状况,以及各种类型的赘生物(包括胰腺)的核心生物标志物网络。因此通过途径分析鉴定的大多数疾病已经与PC相关,或与PC症状性相关。

[0194] 用于分类的签名

[0195] 接着,将数据随机分为训练和测试组。训练组用于鉴定抗体的最区分性的组合,通过应用基于SVM分析的反向消除算法,一个接一个排除抗体。对PC和NC的分类是高度精确的,如通过整个消除过程中较小的Kullback-Leibler (K-L) 差异 (≤ 33.2) 所暗示的(图2A)。当在消除过程中只剩下七种抗体时,达到了突出的K-L最小值(12.0)。该7丛蛋白组,包括IL-6,胱抑素C, IL-8, IL-11, C1抑制物,嗜酸细胞活化趋化因子和HADH2,在先前没有看到的样品的对应测试组中表现出分别为100%和96%的灵敏度(SN)和特异性(SP),其证明可以将少量的分析物组合用于对PC和NC的接近完美的分类。相反,当PC与BC比较时,K-L值显著较高 (≤ 181.3)。这里,最小K-L值(50.0)是不那么突出的,并暗示需要大得多的抗体组用于PC和BC的最优区分(图2B)。对于每个亚组比较,重复程序直到十个不同的,随机生成的训练组已经用于反向消除。所得的K-L曲线与图2中所示的那些高度相似,表明需要平均67种抗体用于PC vs.BC的最优分类(图2B)。

[0196] 为了生成并评价可行数量的抗体的签名,我们从每个消除程序选择了前25种抗体并使用这些在训练组中构建SVM分类模型。通过该模型在对应的测试组中生成的AUC值用作分类准确性的测量(图2C)。每个生成的25-抗体签名可以以高准确性从NC预测PC(平均AUC 0.98)。十个签名的灵敏度和特异性从90%SN和85%SP到100%SN和SP,具有平均SN和SP为95%。相反,较难从BC预测PC(平均AUC 0.67),对于最佳表现签名具有76%SN和67%SP。

[0197] 最后,每种抗体给予分数,对应其在消除程序中的平均持续时间(表3)。在PC vs.NC分析中,十个签名高度相似。例如,首位的抗体,靶向IL-11,在293个消除中具有291.4的分数,在10次中有4次是最后被消除的抗体。总共,PC vs.NC的25种分数最高的抗体代表了20种非冗余的分析物,范围从细胞因子和趋化因子(IL-11, IL-6, IL-13, IL-8, TNF- α , 和嗜酸细胞活化趋化因子),补体组分(C1抑制物, C1q, C5, 和因子B),酶(HADH2, GAK和ATP-5B)到更多。对于PC vs.BC,作为最高的标志物出现的是一种蛋白的高度不同组,具有MAPK1, TNFRSF3, UCHL5, IL-4, Apo-A1, Apo-A4, CD40配体, KSYK和其它在分数最高的分析物中(表3)。

[0198] 即便基于抗体分数(表3)的签名与那些得自p值(表2)的不同,观察到了明显的重叠,特别是对于PC vs.NC签名,其中25种抗体中的8种(IL-11, IL-6, IL-13, HADH2, LDL, GAK,

C1q, 和TNF- α) 出现在两种签名中。对于PC vs. NC所有的分数最高的25种抗体实际上显著地差异表达, 即便不是所有的出现在前25的p值列表中。

[0199] 肿瘤定位

[0200] 基于血清蛋白概貌, 癌症样品也可以根据胰腺中的肿瘤起源分层。主成分分析表明来自定位于胰体和胰尾的肿瘤的样品比来自定位于胰头的肿瘤的样品在一定程度上集群更靠近正常对照, 且可以观察到基于肿瘤定位(头vs. 体/尾) 的癌症样品的分离(图3)。差异蛋白表达分析揭示广泛的分析物列表, 其在胰头和胰体/尾肿瘤间具有血清样品中显著不同的水平, 其中39%的抗体表现出p值<0.001, 且与胰体和胰尾肿瘤相比, 在胰头肿瘤中几乎只表现出上调的水平(表5)。

[0201] 讨论

[0202] 本研究代表了目前已经进行的对用于预测胰腺癌的生物标志物组的最 大的多中心分析之一。据我们所知, 这是使用亲和蛋白质组在这么大的胰腺癌群体(>300个样品) 中对血清分析物的最多重分析。使用内部的重组抗体微阵列进行分析, 这是开发了几年的平台[32, 36], 且其现在包括接近300种抗体, 针对一定范围的靶标严格地选择, 所述靶标主要是免疫调节蛋白质组。采用新的方案用于高通量产生和纯化, 这些抗体可以在少于一周内容易地生成, 并以微微升的规模以三个重复的点快速地印制在平面微阵列上。在目前的设置中, 每天每工作站可以平行分析超过一百份样品, 仅使用微小体积(<1 μ L) 的未稀释血清, 使得能够在几天的过程中分析大样品群体。

[0203] 在本研究中, 在这些阵列上分析了338个血清样品的概貌, 将胰腺癌与正常和良性对照比较。使用高度多重测定例如现在的这个, 很可能出现一定程度的相关性, 特别是当靶向高度相互联系的蛋白时, 例如免疫系统的那些。即便单个抗体的辨别力(由单一p值代表) 也可能是感兴趣的, 将其它统计学方法用于鉴定抗体的最优组合可能是更准确的[37]。这里, 使用了基于支持向量机分析的监督模型。将数据分为训练组, 通过迭代反向消除算法从其鉴定生物标志物签名, 以及互补的验证组, 在其中测试所述签名的分类能力。

[0204] 通过采用这一方法, 从差异表达模式(由p值表示) 以及从反向消除分析两者导出了感兴趣的蛋白的列表。后者表明10种抗体中的4种将足够用于接近完美的区分PC vs. NC, 结果令人鼓舞并确认了我们先前的观察, 即可以通过测定免疫调节蛋白组容易地将PC从健康对照区分[23, 24]。然而, 当将PC与BC(主要是慢性胰腺炎) 比较时, 似乎需要多达67种抗体用于最优分类能力。这可以通过测量蛋白的类型部分解释; 免疫调节分析物可能不会表现出单独的疾病特异性, 因为免疫系统在任何状况下是高度影响的。确实, 本研究中进行的通路分析, 确认了PC和一系列其他状况高度相似的系统性效应, 例如胰岛功能亢进, 胰岛素抗性和代谢疾病, 以及自身免疫和感染, 再次证明了从症状相关的良性状况区分PC的挑战。因此, 鉴定PC的相关免疫签名是平衡的行为(act of balance)。由于每个结合物增加了测定成本和复杂程度, 小的组是合意的, 但签名仍然需要足够大以构成对疾病的敏感并特异的指纹。在发现的这一阶段, 我们推断25种分析物将是可行的起始点, 而后续研究将辨别这些签名是否可以浓缩到更小的生物标志物组, 而仍然表现出诊断免疫测定所需的灵敏度和特异性。

[0205] 虽然采用了两种高度不同的策略用于签名鉴定, 在来源于不同分析的签名之间仍然存在大的重叠, 即获得最高的总体反向消除分数的标志物一般也高度差异表达。例如,

IL-11, IL-6, C1抑制剂, IL-13, HADH2, LDL, GAK, C1q, 和TNF- α 出现在两种签名中, 因此对于PC vs. NC表现出低p值和高的反向消除分数。对于PC vs. BC, 两种签名通过C5, Apo-A4, BTK, TGF- β 1, MCP-1, 和UPF3B重叠。值得注意的是, 当PC与NC相比和BC相比时, 一些标志物都表现出潜力。这些中的几个, 包括C1抑制物, C5, 因子B, IL-13, MCP-1和TNF- α , 已经在我们[23, 24]和其他人[38-40]先前的研究与PC相关, 而HADH2, 一种乙酰CoA脱氢酶, 据我们所知目前尚未对于PC报道。另外, 我们先前没有测量的其它蛋白也表现出对于PC区分的潜力。例如, GAK, 一种丝氨酸/苏氨酸激酶, 在PC vs. NC中大量下调, 也出现在PC vs. NC的反向消除签名中。另外TNFRSF3 (TNF- β 受体), 和UPF3B (mRNA调节蛋白), 也包括在PC vs. NC的p值和抗体分数签名两者中。最后, MAPK1 (ERK2), 一种MAPK/ERK信号通路的激酶, 其已经证明在PC中下调[41], 在PC vs. BC的反向消除签名中是最高分数的蛋白。

[0206] 血清样品也基于胰腺中的肿瘤位置分层, 据我们所知其先前尚未使用蛋白质组进行。表明样品在某种程度上可以基于肿瘤来源分离, 且来自患有胰尾和胰体肿瘤的患者样品的部分相比来自患有胰头癌的患者血清略为较接近正常对照聚集。因此, 这些结果证明了基于其在胰腺中的起源, 癌症的系统性影响的差异。相比来自胰尾/胰体肿瘤的样品, 来自胰头肿瘤的样品中发现大量的血清免疫调节蛋白上调, 可能反映了定位于胰头中的肿瘤更深远的系统性影响, 通常侵入连接胰腺和十二指肠的周围肠系膜血管[3]。因此, 这些发现提示抗体微阵列上的血清概貌可以应用于区分头部和体部/尾部胰腺肿瘤, 其能够潜在地辅助肿瘤治疗的决策。

[0207] 尽管是第4致命的癌症, 但PC的发病率低, 在美国每100000个体约11例[42]。从健康-经济学的角度, 低发病率使得难以在普通人群中促进PC的筛选。基于我们先前报道的灵敏度和特异性[24], 然而最近的研究确认了筛选高风险PC个体的成本高效性[43]。PC的风险因素不仅包括胰腺炎和良性赘生物, 也包括例如家族性胰腺癌, 遗传性胰腺炎, BRCA突变, Peutz-Jeghers症状, 糖尿病和幽门螺杆菌感染[44]。下一步将是探索本研究中鉴定的签名, 用于增加的PC风险的个体中的PC诊断, 所述个体代表了诊断性PC免疫测定的相关目标群体。

[0208] 总的来说, 这一广泛多重的, 多中心的研究, 其揭示了与胰腺癌相关的免疫签名, 表现出90-100%范围内的灵敏度和特异性, 清楚地证明了用于PC诊断的适用性, 也表明重组抗体微阵列用于基于胰腺中肿瘤位置对血清样品分级的潜力。

[0209] 参考文献

[0210] [1] Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA: a cancer journal for clinicians. 2012; 62: 10-29.

[0211] [2] Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010; 467: 1114-7.

[0212] [3] Hidalgo M. Pancreatic cancer. The New England journal of medicine. 2010; 362: 1605-17.

[0213] [4] Conlon KC, Klimstra DS, Brennan MF. Long-term survival after curative resection for pancreatic ductal adenocarcinoma. Clinicopathologic analysis of 5-year survivors. Annals of surgery. 1996; 223: 273-9.

- [0214] [5] Sohn TA, Yeo CJ, Cameron JL, Koniaris L, Kaushal S, Abrams RA, et al. Resected adenocarcinoma of the pancreas-616 patients: results, outcomes, and prognostic indicators. *Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract*. 2000;4:567-79.
- [0215] [6] Furukawa H, Okada S, Saisho H, Ariyama J, Karasawa E, Nakaizumi A, et al. Clinicopathologic features of small pancreatic adenocarcinoma. A collective study. *Cancer*. 1996;78:986-90.
- [0216] [7] Shimizu Y, Yasui K, Matsueda K, Yanagisawa A, Yamao K. Small carcinoma of the pancreas is curable: new computed tomography finding, pathological study and postoperative results from a single institute. *Journal of gastroenterology and hepatology*. 2005;20:1591-4.
- [0217] [8] Ishikawa O, Ohigashi H, Imaoka S, Nakaizumi A, Uehara H, Kitamura T, et al. Minute carcinoma of the pancreas measuring 1cm or less in diameter--collective review of Japanese case reports. *Hepato-gastroenterology*. 1999;46:8-15.
- [0218] [9] Gangi S, Fletcher JG, Nathan MA, Christensen JA, Harmsen WS, Crownhart BS, et al. Time interval between abnormalities seen on CT and the clinical diagnosis of pancreatic cancer: retrospective review of CT scans obtained before diagnosis. *AJR American journal of roentgenology*. 2004;182:897-903.
- [0219] [10] Pelaez-Luna M, Takahashi N, Fletcher JG, Chari ST. Resectability of presymptomatic pancreatic cancer and its relationship to onset of diabetes: a retrospective review of CT scans and fasting glucose values prior to diagnosis. *The American journal of gastroenterology*. 2007;102:2157-63.
- [0220] [11] Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2006;24:5313-27.
- [0221] [12] Koopmann J, Rosenzweig CN, Zhang Z, Canto MI, Brown DA, Hunter M, et al. Serum markers in patients with resectable pancreatic adenocarcinoma: macrophage inhibitory cytokine 1 versus CA19-9. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2006;12:442-6.
- [13] Brand RE, Nolen BM, Zeh HJ, Allen PJ, Eloubeidi MA, Goldberg M, et al. Serum biomarker panels for the detection of pancreatic cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2011;17:805-16.
- [0222] [14] Bunger S, Laubert T, Roblick UJ, Habermann JK. Serum biomarkers for improved diagnostic of pancreatic cancer: a current overview. *Journal of cancer research and clinical oncology*. 2011;137:375-89.
- [0223] [15] Jiang JT, Wu CP, Deng HF, Lu MY, Wu J, Zhang HY, et al. Serum level of

TSGF,CA242 and CA19-9 in pancreatic cancer.World journal of gastroenterology: WJG.2004;10:1675-7.

[0224] [16]Duraker N,Hot S,Polat Y,Hobek A,Gencler N,Urhan N.CEA,CA 19-9,and CA 125 in the differential diagnosis of benign and malignant pancreatic diseases with or without jaundice.Journal of surgical oncology.2007;95:142-7.

[0225] [17]Ni XG,Bai XF,Mao YL,Shao YF,Wu JX,Shan Y,et al.The clinical value of serum CEA,CA19-9,and CA242 in the diagnosis and prognosis of pancreatic cancer.European journal of surgical oncology:the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology.2005;31:164-9.

[0226] [18]Faca VM,Song KS,Wang H,Zhang Q,Krasnoselsky AL,Newcomb LF,et al.A mouse to human search for plasma proteome changes associated with pancreatic tumor development.PLoS medicine.2008;5:e123.

[0227] [19]Firpo MA,Gay DZ,Granger SR,Scaife CL,DiSario JA,Boucher KM,et al.Improved diagnosis of pancreatic adenocarcinoma using haptoglobin and serum amyloid A in a panel screen.World journal of surgery.2009;33:716-22.

[0228] [20]Orchekowski R,Hamelinck D,Li L,Gliwa E,vanBrocklin M,Marrero JA, et al.Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer.Cancer research.2005;65:11193-202.

[0229] [21]Coussens LM,Werb Z.Inflammation and cancer.Nature.2002;420:860-7.

[0230] [22]Carlsson A,Wuttge DM,Ingvarsson J,Bengtsson AA,Sturfelt G, Borrebaeck CA,et al.Serum protein profiling of systemic lupus erythematosus and systemic sclerosis using recombinant antibody microarrays.Molecular& cellular proteomics:MCP.2011;10:M110 005033.

[0231] [23]Ingvarsson J,Wingren C,Carlsson A,Ellmark P,Wahren B,Engstrom G, et al.Detection of pancreatic cancer using antibody microarray-based serum protein profiling.Proteomics.2008;8:2211-9.

[0232] [24]Wingren C,Sandstrom A,Segersvard R,Carlsson A,Andersson R,Lohr M, et al.Identification of serum biomarker signatures associated with pancreatic cancer.Cancer research.2012;72:2481-90.

[0233] [25]Anderson NL,Anderson NG.The human plasma proteome:history, character,and diagnostic prospects.Molecular&cellular proteomics:MCP.2002;1: 845-67.

[0234] [26]Zhang H,Liu AY,Loriaux P,Wollscheid B,Zhou Y,Watts JD,et al.Mass spectrometric detection of tissue proteins in plasma.Molecular&cellular proteomics:MCP.2007;6:64-71.

[0235] [27]Surinova S,Schiess R,Huttenhain R,Cerciello F,Wollscheid B, Aebersold R.On the development of plasma protein biomarkers.Journal of proteomeresearch.2011;10:5-16.

- [0236] [28]Haab BB,Geierstanger BH,Michailidis G,Vitzthum F,Forrester S,Okon R,et al.Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens:systematic variation between sample types and calibration of mass spectrometry data.*Proteomics*.2005;5:3278-91.
- [0237] [29]Alonzo TA,Pepe MS,Moskowitz CS.Sample size calculations for comparative studies of medical tests for detecting presence of disease.*Statistics in medicine*.2002;21:835-52.
- [0238] [30]Borrebaeck CA,Wingren C.Design of high-density antibody microarrays for disease proteomics:key technological issues.*Journal of proteomics*.2009;72:928-35.
- [0239] [31]Carlsson A,Wingren C,Kristensson M,Rose C,Ferno M,Olsson H,et al.Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases.*Proceedings of the National Academy of Sciences of the United States of America*.2011;108:14252-7.
- [0240] [32]Wingren C,Ingvarsson J,Dexlin L,Szul D,Borrebaeck CA.Design of recombinant antibody microarrays for complex proteome analysis:choice of sample labeling-tag and solid support.*Proteomics*.2007;7:3055-65.
- [0241] [33]Carlsson A,Persson O,Ingvarsson J,Widegren B,Salford L,Borrebaeck CA,et al.Plasma proteome profiling reveals biomarker patterns associated with prognosis and therapy selection in glioblastoma multiforme patients.*Proteomics Clinical applications*.2010;4:591-602.
- [0242] [34]Olsson N,Wallin S,James P,Borrebaeck CA,Wingren C.Epitope-specificity of recombinant antibodies reveals promiscuous peptide-binding properties.*Protein science:a publication of the Protein Society*.2012;21:1897-910.
- [0243] [35]Carlsson A,Wingren C,Ingvarsson J,Ellmark P,Baldertorp B,Ferno M,et al.Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays.*Eur J Cancer*.2008;44:472-80.
- [0244] [36]Ingvarsson J,Larsson A,Sjoholm AG,Truedsson L,Jansson B,Borrebaeck CA,et al.Design of recombinant antibody microarrays for serum protein profiling:targeting of complement proteins.*Journal of proteome research*.2007;6:3527-36.
- [0245] [37]Quackenbush J.Computational analysis of microarray data.*Nature reviews Genetics*.2001;2:418-27.
- [0246] [38]Ma Y,Hwang RF,Logsdon CD,Ullrich SE.Dynamic mast cell-stromal cell interactions promote growth of pancreatic cancer.*Cancer research*.2013;73:3927-37.
- [0247] [39]Kudo-Saito C,Shirako H,Ohike M,Tsukamoto N,Kawakami Y.CCL2 is critical for immunosuppression to promote cancer metastasis.*Clinical&*

experimental metastasis.2013;30:393-405.

[0248] [40]McDade TP,Perugini RA,Vittimberga FJ,Jr.,Carrigan RC,Callery MP.Salicylates inhibit NF-kappaB activation and enhance TNF-alpha-induced apoptosis in human pancreatic cancer cells.The Journal of surgical research.1999;83:56-61.

[0249] [41]Arlt A,Muerkoster SS,Schafer H.Targeting apoptosis pathways in pancreatic cancer.Cancer letters.2013;332:346-58.

[0250] [42]Shaib YH,Davila JA,El-Serag HB.The epidemiology of pancreatic cancer in the United States:changes below the surface.Alimentary pharmacology&therapeutics.2006;24:87-94.

[0251] [43]Ghatnekar O,Andersson R,Svensson M,Persson U,Ringdahl U,Zeilon P, et al.Modelling the benefits of early diagnosis of pancreatic cancer using a biomarker signature.International journal of cancer Journal international du cancer.2013.

[0252] [44]Konstantinou F,Syrigos KN,Saif MW.Pancreatic cancer:what about screening and detection?JOP:Journal of the pancreas.2013;14:312-5.

[0253] 表1.血清样品人口统计学

[0254]	诊断	亚组	性别	年龄 (平均	样品数量
			(M/F)	±sd)	

[0255]	胰腺癌 (PC)	92/64	66 ± 13	156
	胰体			16
	胰头			97
	胰尾 Tail			10
	其它 Other			16
	未指明			17
	良性对照 (BC)	117/35	52 ± 14	152
	急性胰腺炎			33
	慢性胰腺炎			110
	Langerhan 赘生 物			3
	胰腺赘生物			6
	正常对照 (NC)	20/10	62 ± 14	30
总数				338

[0256] 表2. 基于差异蛋白表达的显著性的前25种抗体。括号中示出的是个体抗体克隆后缀 (对于多个抗体克隆靶向的分析物)。FC=倍数变化

	PC-NC-BC ANOVA		PC-NC t 检验			PC-BC t 检验		
	抗体	p 值	抗体	p 值	FC	抗体	p 值	FC
[0257]	GAK (3)	4.14E-48	GAK (3)	3.44E-30	0.68	胱抑素 C (3)	3.80E-11	1.08
	IL-6 (7)	4.00E-43	IL-6 (7)	1.11E-27	0.71	IL-13 (3)	1.09E-10	1.13
	GAK (2)	3.25E-36	GAK (1)	3.25E-26	0.77	IL-1α (1)	3.08E-10	1.11
	IL-11 (2)	1.78E-32	LDL (2)	1.15E-22	0.62	表面 ag X	5.57E-09	1.10
	LDL (2)	6.09E-32	MAPK8 (1)	4.24E-20	0.81	BTK (2)	6.06E-09	1.05
	TNF-α (3)	1.89E-31	IL-11 (2)	4.56E-18	0.69	胱抑素 C (4)	6.60E-09	1.05

	组织蛋白原 W	4.88E-27	TNF- α (3)	2.93E-16	0.76	CIMS (26)	6.79E-09	1.10
	IL-13 (3)	8.15E-27	HADH2 (3)	2.49E-15	0.89	CD40 (1)	1.02E-08	1.09
	MAPK8 (1)	3.02E-22	组织蛋白原 W	3.21E-14	0.79	TNFRSF3 (2)	1.29E-08	1.05
	IL-1 α (1)	5.83E-20	TNFRSF3 (1)	5.08E-13	0.91	ORP-3 (2)	2.16E-08	1.07
	IL-13 (2)	3.36E-19	VEGF (3)	1.52E-12	1.24	Apo-A4 (3)	2.66E-08	1.07
	TNFRSF3 (1)	8.51E-16	IL-1ra (1)	3.24E-12	1.16	UPF3B (2)	2.84E-08	1.04
	IL-18 (2)	5.33E-15	C1 inh (1)	1.20E-11	1.22	MUC-1 (1)	3.49E-08	1.07
	IL-1ra (1)	9.67E-15	IL-13 (3)	2.73E-11	0.77	TNF- α (3)	4.31E-08	1.08
	HADH2 (3)	1.06E-14	C1q	4.12E-11	1.11	CIMS (16)	4.67E-08	1.08
[0258]	CD40 (1)	1.48E-14	IL-16 (3)	5.19E-11	1.15	ATP-5B (1)	5.68E-08	1.06
	胱抑素 C (4)	2.36E-14	VEGF (1)	5.43E-11	1.19	CIMS (12)	6.00E-08	1.06
	IL-4 (3)	3.12E-14	CD40L	6.21E-11	1.13	IL-13 (1)	6.31E-08	1.08
	CIMS (18)	6.91E-14	IL-4 (3)	8.82E-11	1.24	MCP-1 (4)	9.64E-08	1.03
	CIMS (16)	7.21E-14	Sialyl Lewis x	9.47E-11	1.10	CIMS (1)	1.14E-07	1.08
	VEGF (3)	8.52E-14	IL-18 (2)	1.12E-10	1.20	CD40 (3)	1.28E-07	1.06
	FASN (3)	1.18E-13	CIMS (18)	1.43E-10	1.16	组织蛋白原 W	1.43E-07	1.07
	TGF- β 1 (2)	2.59E-13	CIMS (23)	2.64E-10	1.13	Digoxin	1.55E-07	1.08
	CIMS (26)	2.61E-13	MCP-1 (2)	3.08E-10	1.16	TGF- β 1 (2)	1.88E-07	1.09
	CIMS (25)	6.47E-13	CIMS (25)	3.44E-10	1.24	CIMS (24)	3.32E-07	1.08

[0259] 表3. 基于来自十次反向消除迭代的抗体分数的前25种抗体。括号中示出的是个体抗体克隆后缀(对于多个抗体克隆靶向的分析物)。

	PC vs NC	分数	PC vs BC	分数
[0260]	IL-11 (2)	291.4	MAPK1 (3)	276.5

	IL-6 (7)	288.1	C5 (2)	273
	胱抑素 C (1)	286.9	TNFRSF3 (1)	265.5
	C1 inh (3)	279.2	TNFRSF3 (2)	260.9
	血管动蛋白 (1)	276	UCHL5	259.6
	IL-13 (2)	272.9	IL-4 (3)	258.7
	IL-13 (3)	270.7	因子 B(3)	258
	CD40 (1)	270.6	Apo-A4 (3)	257.5
	HADH2 (3)	270.4	KSYK-1	255.2
	HADH2 (4)	269.7	Sox11A	253.1
	C1 inh. (4)	269.4	CD40L	252.2
	C1 inh. (2)	269.2	Apo-A1 (1)	251.4
	LDL (2)	268.1	CIMS (13)	250.1
[0261]	GAK (3)	268	BTK (2)	246.1
	C3 (1)	266.1	GM-CSF (5)	245
	CIMS (5)	264.3	TGF- β 1 (2)	239.5
	C1q	261.1	PTP-1B (2)	237.2
	CD40 (4)	259.6	MCP-1 (7)	235.1
	IL-8 (2)	259.4	UPF3B (1)	232.5
	C5 (2)	258.5	C1 inh. (4)	228.3
	ATP-5B (3)	257.1	唾液酸化 Lewis x	227.6
	因子 B (4)	256.2	IL-3 (1)	225.8
	CIMS (10)	253.6	IL-9 (2)	224.2
	TNF- α (3)	253.5	HADH2 (2)	222.7
	嗜酸细胞活化趋	248.4	IL-4 (4)	222.4
[0262]	化因子 (3)			

[0263] 表4. 抗体微阵列上靶向的抗原

	蛋白	全名	抗体克隆
			数
	血管动蛋白	血管动蛋白	2
	Apo-A1	载脂蛋白 A1	3
	Apo-A4	载脂蛋白 A4	3
	ATP-5B	ATP 合酶 beta 亚基, 线粒体	3
	β -半乳糖苷酶	Beta-半乳糖苷酶	1
	BTK	酪氨酸蛋白激酶 BTK	4
	C1 抑制物	血浆蛋白酶 C1 抑制物	4
	C1q	补体 C1q	1
	C1s	补体 C1s	1
	C3	补体 C3	6
	C4	补体 C4	4
	C5	补体 C5	3
[0264]	CD40	CD40 蛋白	4
	CD40L	CD40 配体	1
	CDK-2	细胞周期素依赖的激酶 2	2
	CHX10	视觉系统同源盒 2	3
	CT17	霍乱毒素 B 亚基	1
	胱抑素 C	胱抑素 C	4
	Digoxin	Digoxin	1
	DUSP9	双特异性蛋白磷酸酶 9	1
	EGFR	表皮生长因子受体 (Cetuximab (Erbix) 抗体)	1
	嗜酸细胞活化趋化因子	嗜酸细胞活化趋化因子	3
	因子 B	补体因子 B	4
	FASN	FASN 蛋白	4
	GAK	GAK 蛋白	3
	GLP-1	胰高血糖素样肽 1	1

	GLP-1R	胰高血糖素样肽 1 受体	1
	GM-CSF	粒细胞-巨噬细胞集落刺激因子	6
	HADH2	HADH2 蛋白	4
	Her2/ErbB-2	受体酪氨酸激酶 erbB-2	4
	HLA-DR/DP	HLA-DR/DP	1
	ICAM-1	细胞间粘附分子 1	1
	IFN- γ	干扰素 gamma	3
	IgM	免疫球蛋白 M	5
	IL-10	白细胞介素-10	3
	IL-11	白细胞介素-11	3
	IL-12	白细胞介素-12	4
	IL-13	白细胞介素-13	3
	IL-16	白细胞介素-16	3
	IL-18	白细胞介素-18	3
	IL-1 α	白细胞介素-1 alpha	3
[0265]	IL-1 β	白细胞介素-1 beta	3
	IL-1ra	白细胞介素-1 受体拮抗蛋白	3
	IL-2	白细胞介素-2	3
	IL-3	白细胞介素-3	3
	IL-4	白细胞介素-4	4
	IL-5	白细胞介素-5	3
	IL-6	白细胞介素-6	8
	IL-7	白细胞介素-7	2
	IL-8	白细胞介素-8	3
	IL-9	白细胞介素-9	3
	整联蛋白 α -10	整联蛋白 alpha-10	1
	整联蛋白 α -11	整联蛋白 alpha-11	1
	JAK3	酪氨酸蛋白激酶 JAK3	1
	角蛋白 19	角蛋白, I 型细胞骨架 19	3
	KSYK	酪氨酸蛋白激酶 SYK	2
	LDL	载脂蛋白 B-100	2

	瘦素	瘦素	1
	Lewis x	Lewis x	2
	Lewis y	Lewis y	1
	基膜聚糖	Lumican	1
	MAPK1	丝裂原活化的蛋白激酶 1	4
	MAPK8	丝裂原活化的蛋白激酶 8	3
	MATK	巨核细胞相关的酪氨酸蛋白激酶	3
	MCP-1	C-C 基序趋化因子 2	9
	MCP-3	C-C 基序趋化因子 7	3
	MCP-4	C-C 基序趋化因子 13	3
	MUC-1	粘蛋白-1	6
	肌间蛋白-2	肌间蛋白-2	2
	ORP-3	氧固醇结合蛋白相关蛋白 3	2
	骨桥蛋白	骨桥蛋白	3
	P85A	磷脂酰肌醇 3-激酶调节 alpha 亚基	3
[0266]	PKB gamma	RAC-gamma 丝氨酸/苏氨酸蛋白激酶	2
	组织蛋白原 W	组织蛋白酶 W	1
	备解素	备解素	1
	PSA	前列腺特异性抗原	1
	PTK-6	蛋白酪氨酸激酶 6	1
	PTP-1B	酪氨酸蛋白磷酸酶非受体 I 型	3
	RANTES	C-C 基序趋化因子 5	3
	RPS6KA2	核糖体蛋白 S6 激酶 alpha-2	3
	唾液酸化 Lewis x	唾液酸化 Lewis x	1
	Sox11A	转录因子 SOX-11	1
	STAP2	信号转导接头蛋白 2	4
	STAT1	信号转导蛋白和转录激活子 1-alpha/beta	2
	表面 Ag X	表面 Ag X	1
	TBC1D9	TBC1 域家族成员 9	3
	TENS4	张力蛋白-4	1
	TGF-β1	转化生长因子 beta-1	3

[0267]	TM 肽	跨膜肽	1
	TNF- α	肿瘤坏死因子	3
	TNF- β	淋巴毒素- α	4
	TNFRSF14	肿瘤坏死因子受体超家族成员 14	2
	TNFRSF3	肿瘤坏死因子受体超家族成员 3	3
	UBC9	SUMO-缀合酶 UBC9	3
	UBE2C	泛素-缀合酶 E2 C	2
	UCHL5	泛素羧基-末端水解同工酶 L5	1
	UPF3B	无义转录物调节蛋白 3B	2
	VEGF	内皮细胞生长因子	4

[0268] 表5. 从患有不同位置的胰腺肿瘤的患者抽取的血清样品的差异蛋白表达分析。结果表示为体部+尾部肿瘤vs. 头部肿瘤, 排名前40的抗体 ($p < 5 \cdot 10^{-5}$)。FC = 倍数变化。

	抗体	<i>p-val</i>	<i>FC</i>
[0269]	IL-1 α (2)	4.32E-07	0.874447
	CIMS (16)	1.28E-06	0.874676
	VEGF (1)	1.47E-06	0.870962
	TNF- β (2)	1.57E-06	0.872955
	IL-11 (2)	2.15E-06	0.833144
	CIMS (18)	3.09E-06	0.891206
	CD40L	3.89E-06	0.917468
	IL-3 (3)	4.19E-06	0.87497
	CIMS (30)	4.88E-06	0.919455
	IL-6 (1)	7.01E-06	0.873484
	HLA-DR/DP	8.10E-06	0.879348
	IL-2 (3)	8.34E-06	0.845869
	血管动蛋白 (2)	8.45E-06	0.892806
	整联蛋白 α -10	9.07E-06	0.864779
	IL-18 (3)	9.40E-06	0.899675
	Sox11A	1.05E-05	0.898637
	IL-7 (1)	1.18E-05	0.892816

[0270]	MCP-1 (3)	1.20E-05	0.885187
	表面 ag X	1.20E-05	0.865919
	IL-9 (1)	1.21E-05	0.872639
	CIMS (20)	1.36E-05	0.873286
	IL-12 (3)	1.67E-05	0.872587
	Lewis x (1)	1.77E-05	0.898973
	IgM (3)	1.91E-05	0.893885
	IL-7 (2)	1.98E-05	0.893886
	CIMS (25)	2.25E-05	0.851688
	CIMS (2)	2.30E-05	0.88583
	CIMS (6)	2.41E-05	0.858863
	IL-16 (3)	2.57E-05	0.905756
	GLP-1	2.78E-05	0.890237
	CHX10 (3)	2.83E-05	0.889864
	IL-4 (3)	3.02E-05	0.860565
	VEGF (4)	3.20E-05	0.848107
	IL-3 (2)	3.35E-05	0.871146
	IL-2 (2)	3.89E-05	0.843949
	IL-1ra (3)	3.93E-05	0.88975
	RANTES (3)	4.51E-05	0.933006
	CIMS (31)	4.80E-05	0.886723
	CIMS (27)	4.90E-05	0.915073
	TGF-β1 (3)	4.91E-05	0.889071

[0271] 表A

[0272]	诊断			
	PC (体部&尾部)			
	Uniprot entry ID	推荐的蛋白名称	缩写名	PC vs PC (头部)

(i) 核心生物标志物

	Q6IBS9	HADH2 蛋白	HADH2	X	
	P36941	肿瘤坏死因子受体超家族 3	TNFRSF3	X	
	(ii)优选的生物标志物(PC & BTvH)				
	P35716	转录因子 SOX-11	Sox11A	X	X
		整联蛋白			
	O75578	整联蛋白 alpha-10	a-10	X	X
	NA	EDFR		X	X
	NA	EPFR		X	X
	NA	LSADHR		X	X
	NA	SEHLR		X	X
	NA	AQQHQWDGLLSYQDSL		X	X
	NA	WTRNSNMNYWLIIRL		X	X
	NA	WDSR		X	X
	(iii)优选的生物标志物(PC)				
	NA	DFAEDK		X	
[0273]	Q6PJJ3	FASN 蛋白	FASN	X	
	Q5U4P5	GAK 蛋白	GAK	X	
	NA	LNVDGK		X	
	NA	LTEFAK		X	
	NA	LYEIR		X	
		巨核细胞相关的酪氨酸蛋			
	P42679	白激酶	MATK	X	
	Q9H4L5	氧固醇结合蛋白相关蛋白 3	ORP-3	X	
	NA	QEASFK		X	
	NA	SSAYSR		X	
	NA	QEASFK		X	
	NA	TEEQLK		X	
	NA	TLYVGK		X	
	NA	FLLMQYGGMDEHAR		X	
	NA	GIVKYLYEDEG		X	
	NA	GIVKYLYEDEG		X	

	P43405	酪氨酸蛋白激酶 SYK	KSYK	X	
	(iv)可选的生物标志物 (PC & BTvH)				
	Q4VCS5	血管动蛋白	血管动蛋白	X	X
	P13500	C-C 基序趋化因子 2	MCP-1	X	X
	P13501	C-C 基序趋化因子 5	RANTES	X	X
	P29965	CD40 配体	CD40L	X	X
	P01275	胰高血糖素样肽-1	GLP-1	X	X
	NA	免疫球蛋白 M	IgM	X	X
	P01583	白细胞介素-1 alpha	IL-1a	X	X
	P18510	白细胞介素-1 受体拮抗蛋白	IL-1ra	X	X
	P20809	白细胞介素-11	IL-11	X	X
	P29459/60	白细胞介素-12	IL-12	X	X
	Q14005	白细胞介素-16	IL-16	X	X
	Q14116	白细胞介素-18	IL-18	X	X
	P60568	白细胞介素-2	IL-2	X	X
[0274]	P08700	白细胞介素-3	IL-3	X	X
	P05112	白细胞介素-4	IL-4	X	X
	P05231	白细胞介素-6	IL-6	X	X
	P13232	白细胞介素-7	IL-7	X	X
	P15248	白细胞介素-9	IL-9	X	X
	NA	Lewis x	Lewis x	X	X
	P01374	淋巴毒素-alpha	TNF-b	X	X
	P01137	转化生长因子 beta-1	TGF-b1	X	X
	P15692	血管内皮生长因子	VEGF	X	X
	P58304	视觉系统同源盒 2	CHX10	X	X
	P01903/P01911/				
	P79483/P13762/				
	Q30154/P20036				
	/		HLA-DR/D		
	P04440		P	X	X
	(v) 可选的生物标志物 (PC)				

	P02647	载脂蛋白 A1	Apo-A1	X
	P06727	载脂蛋白 A4	Apo-A4	X
	P04114	载脂蛋白 B-100	LDL	X
	P06576	ATP 合酶 beta 亚基, 线粒体	ATP-5B	X
	P16278	Beta-半乳糖苷酶	B-半乳糖苷酶	X
	P56202	组织蛋白酶 W	组织蛋白原 W	X
	Q99616	C-C 基序趋化因子 13	MCP-4	X
	P80098	C-C 基序趋化因子 7	MCP-3	X
	Q6P2H9	CD40 蛋白	CD40	X
	P02745/6/7	补体 C1q	C1q	X
	P09871	补体 C1s	C1s	X
	P01024	补体 C3	C3	X
	P0COL4/5	补体 C4	C4	X
	P01031	补体 C5	C5	X
	P00751	补体因子 B	因子 B	X
[0275]	P24941	细胞周期素依赖的激酶	CDK-2	X
	P01034	胱抑素-C	胱抑素 C	X
		嗜酸细胞活化		
	P51671	嗜酸细胞活化趋化因子	趋化因子	X
	P00533	表皮细胞生长因子受体	EGFR	X
	P43220	胰高血糖素样肽 1 受体	GLP-1R	X
		粒细胞-巨噬细胞集落刺激		
	P04141	因子	GM-CSF	X
	Q9UKX5	整联蛋白 alpha-11	整联蛋白 a-11	X
	P05362	细胞间粘附分子 1	ICAM-1	X
	P01579	干扰素 gamma	IFN-g	X
	P01584	白介素-1 beta	IL-1b	X
	P22301	白介素-10	IL-10	X
	P35225	白介素-13	IL-13	X
	P05113	白介素-5	IL-5	X
	P10145	白介素-8	IL-8	X

	P08727	角蛋白 I 型细胞骨架 19	角蛋白 19	X
	P41159	瘦素	瘦素	X
	P51884	基膜聚糖	基膜聚糖	X
	P28482	丝裂原活化的蛋白激酶 1	MAPK1	X
	P45983	丝裂原活化的蛋白激酶 8	MAPK8	X
	P15941	粘蛋白-1	MUC-1	X
	P54296	肌间蛋白-2	肌间蛋白-2	X
	P10451	骨桥蛋白	骨桥蛋白	X
		磷脂酰肌醇 3-激酶调节		
	P27986	alpha 亚基	P85A	X
	P05155	血浆蛋白酶 C1 抑制物	C1 inh	X
	P27918	备解素	备解素	X
	P07288	前列腺特异性抗原	PSA	X
	P04626	受体酪氨酸蛋白激酶 erbB-2	Her2/ErbB-2	X
	Q9BZI7	无义转录物调节蛋白 3B	UPF3B	X
[0276]	Q15349	核糖体蛋白 S6 激酶 alpha-2	RPS6KA2	X
		唾液酸化		
	NA	唾液酸化 Lewis x	Lewis x	X
	Q9UGK3	信号转导接头蛋白 2	STAP2	X
	P63279	SUMO-缀合的酶 UBC9	UBC9	X
	Q6ZT07	TBC1 域家族成员 9	TBC1D9	X
	NA	跨膜肽	(Tm 肽)	X
	P01375	肿瘤坏死因子 alpha	TNF-a	X
		肿瘤坏死因子受体超家族		
	Q92956	成员 14	TNFRSF14	X
	Q06187	酪氨酸蛋白激酶 BTK	BTK	X
	P52333	酪氨酸蛋白激酶 JAK3	JAK3	X
		酪氨酸蛋白磷酸酶非受体 I		
	P18031	型	PTP-1B	X
		泛素羧基末端水解酶同工		
	Q9Y5K5	酶 L5	UCHL5	X
	O00762	泛素缀合的酶 E2 C	UBE2C	X

[0277]

(vi) 优选的生物标志物 (BTvH)

NA	FIQTDK	X
----	--------	---

[0278] 表B-训练的SVM程序

[0279] 使用e1071 1.5-24SVM获得了以下参数,可以从 <http://cran.r-project.org/web/packages/e1071/index.html> 获得

(i) Elimination function

```
# getWorstAb: Returns the name of the "worst performing" probe
getWorstAb <- function(errors, abNames)
{
  return(abNames[order(errors, decreasing = F)[1]])
}

# testModels: Tests all models with each probe eliminated once
testModels <- function(models, elimData, averages, svmfac)
{
  nsamples <- ncol(elimData)
  d <- as.numeric(svmfac)-1
  y <- numeric(nsamples)
  E <- numeric(nsamples)
  analytes <- nrow(elimData)
  errors <- numeric(nrow(elimData))

  for(k in 1:analytes)
  {
    # Replace probe "k" with the average, but store the original values
    backup <- elimData[k,]
    elimData[k,] <- averages[k]

    # Do a LOO loop over the dataset using the already calibrated models
    for (i in 1:nsamples)
    {
      pred <- predict(models[[i]], t(elimData[,i]), decision.values=TRUE)
      # Save the decision values
      y[i] <- as.numeric(attributes(pred)$decision.values)
    }

    # Calculate "likelihoods"
    y = 1/(1 + exp(-y))

    # Calculate the K-L error for the classification with one probe eliminated
    for (i in 1:nsamples)
    {
      E[i] <- -(d[i]*log(y[i])+(1-d[i])*log(1-y[i]))
    }

    # Save the error
    errors[k] <- sum(E)

    # Put the original probe data back
    elimData[k,] <- backup
  }
  return( errors )
}
```

```

# getNewElimData: Removes the correct antibody from training data
getNewElimData <- function(errors, elimData)
{
  # Position for smallest error
  tasBort <- order(errors, decreasing = F)[1]
  return(elimData[-tasBort,])
}

# getSmallestError: Returns the smallest error
getSmallestError <- function(errors)
{
  return(min(errors))
}

# getNewAverages: creates a new list of averages after elimination of a probe
getNewAverages <- function(errors, averages)
{
  # find the position of the smallest error
  tasBort <- order(errors, decreasing = F)[1]
  return(averages[-tasBort])
}

# getRemovedAb: Returns the name of the probe to be eliminated
getRemovedAb <- function(errors, abNames)
{
  return(abNames[order(errors, decreasing = T)[1]])
}

backElim <- function(filename, group1, group2, log = FALSE)
{
  # Include libraries "e1071" and "tcltk"
  library(e1071)
  library(tcltk)

  # Read data
  rawfile <- read.delim(filename)

  # Read groups
  groups <- rawfile[,2]

  # Read sample names
  samplenames <- as.character(rawfile[,1])

  # Create data set
  data <- t(rawfile[, -c(1,2)])

  # Log data if specified
  if (log) {data <- log(data)/log(2)}

  # Get probe names
  ProteinNames <- read.delim(filename, header=FALSE)
  ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
  ProteinNames <- ProteinNames[-(1:2)]

  # Get number of probes
  antal <- length(ProteinNames)

  # Name samples and probes
  rownames(data) <- ProteinNames
  colnames(data) <- samplenames

  # Create subsets (boolean)
  subset1 <- is.element(groups, strsplit(group1, ",")[1])
  subset2 <- is.element(groups, strsplit(group2, ",")[1])

```

[0281]

```

# Create group list as factors
svmfac <- factor(rep('rest',ncol(data)),levels=c(group1,group2,'rest'))
svmfac[subset1] <- group1
svmfac[subset2] <- group2
svmfac <- svmfac[subset1|subset2]

# Create vector for storage of K-L errors for each signature length
smallestErrorPerLength <- rep(NA,antal)

# Calculate the mean signal for each probe over all samples
averages <- apply(data, 1, mean)

# Create vector for storage of probe names. The prob that was
# eliminated when the error was at its minimum will be stored here
abOrder <- rep(NA,antal)

# Create a dataset to do the elimination in
elimData <- data[,subset1|subset2]

# Find number of samples
nsamples <- ncol(elimData)

# New subsets in elimData
subset1 <- svmfac==group1
subset2 <- svmfac==group2

print(paste(nsamples, "samples"),quote=F)
print(paste(" ",sum(subset1), "in", group1),quote=F)
print(paste(" ",sum(subset2), "in", group2),quote=F)
# List for storage of SVM models
models <- numeric(nsamples)

# Variable for storing the number of eliminated probes
borttagna <- 0

[0282]

# Create a progress bar
wrst <- 0
proc <- 0
m <- 0
for(i in 1:(antal-1)){
  m <- m+(antal-i)*sqrt(antal-i)
}
pb <- tkProgressBar(title = "Progress:", min = 0, max = m, width = 400)

# Update progress bar
setTkProgressBar(pb, proc, label=paste("Eliminating first ab\t", round((proc)/(m)*100, 0),"% done"))

# Check if groups are given in correct order
control <- as.numeric(svmfac)
if(sum(control[subset1]) > sum(control[subset2]))
{
  print("ERROR: Change the order of group1 and group2 in the data file!!!")
  break
}

# Run backward elimination until only two probes remain
for(j in 1:(antal-1))
{

  # Train a model for each N-1 combination of samples using all probes left
  # in the dataset elimData
  for (i in 1:nsamples)
  {
    # Store the models in a vector (models) containing the model lists.
    models[i] <- list(svm(t(elimData[,-i]), svmfac[-i], kernel="linear"))
  }

  # All models needed for the LOO cross-validation are now calibrated and stored in "models".

```

```

# The models are all tested on their corresponding LOO samples in elimData.
# Each probe in elimData is eliminated in one LOO round by replacing it with the
# previously calculated mean value.
#
# When all models are tested on elimData with one probe eliminated the K-L error is
# calculated and stored in "errors". This procedure is repeated until all probes
# have been eliminated once, resulting in a list of K-L errors with the same length
# as the number of probes left in the dataset

# Create a list with K-L errors for the current number of probes (antal + 1 - j)
# where each probe has been eliminated
errors <- testModels(models, elimData, averages, svmfac)

# Add the name of the probe with the worst effect on classification to "abOrder"
wrst<-getWorstAb(errors, row.names(elimData))
abOrder[j] <- wrst

# Add the smallest K-L error value (corresponding to the probe added to "abOrder" above")
smallestErrorPerLength[j] <- getSmallestError(errors)

# Remove the probe from "averages"
averages <- getNewAverages(errors, averages)

# Remove the probe from "elimData"
elimData <- getNewElimData(errors, elimData)

# Add 1 to the number of removed probes
borttagna <- borttagna + 1

# Inform user that a probe has been removed and what time it is
print(paste(j, "probes eliminated @", Sys.time()), sep="")

# Update progress bar
proc <- proc + (antal-j)*sqrt(antal-j)
setTkProgressBar(pb, proc, label=paste(wrst,"eliminated\t", round((proc/m)*100, 0), "% done"))
}

# Add the name of the last, "uneliminated", probe to "abOrder"
abOrder[length(abOrder)] <- setdiff(ProteinNames, abOrder)

# Save the result to file with a random number attached to the
# filename to avoid replacing an old run.
filename <- paste("Backward elimination result(", morm(1)+1, ").txt", sep="")
write.table(cbind(smallestErrorPerLength, abOrder), file=filename, sep="\t", quote = F, row.names = F)
plot(smallestErrorPerLength, type = "b", ylab = "K-L Error", xlab = "Eliminations")
}

```

[0283]

(ii) Naïve Bayesian

```

NBTrainer <- function(data, fac){
  MeanVariancePval <- function(vec, fac){
    vec1 <- vec[fac==levels(fac)[1]]
    vec2 <- vec[fac==levels(fac)[2]]
    if (sum(!is.na(vec1))<=2 | sum(!is.na(vec2))<=2){
      return(c(NA,NA,NA,NA,NA))
    }
    mean1 <- mean(vec1, na.rm=TRUE)
    var1 <- var(vec1, na.rm=TRUE)
    mean2 <- mean(vec2, na.rm=TRUE)
    var2 <- var(vec2, na.rm=TRUE)
    if (var1==0 | var2==0){return(c(NA,NA,NA,NA,NA))}
    pval <- t.test(vec1,vec2,var.equal=TRUE)$p.value
    return(c(mean1,var1,mean2,var2,pval))
  }
  return(t(apply(data, 1, MeanVariancePval, fac)))
}

```

```

NBpredicter <- function(testdata , NBtrained , topnumber=Inf , logfoldcut=0 , pcut=1){
  if (topnumber==Inf){
    indices <- !is.na(NBtrained[,5]) & NBtrained[,5]<=pcut & abs(NBtrained[,1]-NBtrained[,3])>=logfoldcut
  }else{
    preindices <- !is.na(NBtrained[,5]) & NBtrained[,5]<=pcut
    abs(NBtrained[preindices,1]-NBtrained[preindices,3]) -> foldchange
    cutfold <- sort(foldchange , decreasing=TRUE )[min(topnumber,length(foldchange))]
    indices <- preindices & (abs(NBtrained[,1]-NBtrained[,3]) >= cutfold)
  }
  NBtrainedred <- matrix(NBtrained[indices,],ncol=ncol(NBtrained))
  testdatared <- matrix(testdata[indices,], ncol=ncol(testdata))
  singlegene <- function( genepred){
    l1 <- -((genepred[6] - genepred[1])^2)/(2*genepred[2])-0.5*log(2*pi*genepred[2])
    l2 <- -((genepred[6] - genepred[3])^2)/(2*genepred[4])-0.5*log(2*pi*genepred[4])
    #print(genepred)
    return(l1-l2)
  }
  NBvectorpredicter <- function(vec){
    combined <- cbind(NBtrainedred , vec)
    combined <- matrix(combined[!is.na(vec),], ncol=6)
    return(sum(apply( combined , 1 , singlegene)))
  }
  return(apply(testdatared , 2 , NBvectorpredicter))
}

```

```

myROC <- function(numbers , fac){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  wilcoxresult <- wilcox.test(numbers~fac , alternative="greater")
  ROCarea <- as.numeric(wilcoxresult$statistic)/(n1*n2)
  pval <- wilcoxresult$p.value
  return(c(ROCarea,pval))
}

```

[0284]

```

SensitivitySpecificity <- function(numbers, fac){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  un <- sort(unique(numbers), decreasing=TRUE)
  SenSpe <- function(x){
    sen <- sum(numbers>=x & fac==levels(fac)[1])/n1
    spe <- 1 - sum(numbers>=x & fac==levels(fac)[2])/n2
    return(list(Sensitivity=sen,Specificity=spe))
  }
  return(t(sapply(un , SenSpe)))
}

```

```

NBloopreparer <- function(data , fac){
  nsamples <- ncol(data)
  ngenes <- nrow(data)
  NBtrainedarray <- array(NA , dim=c(ngenes,5,nsamples))
  for (i in 1:nsamples){ print(i)
    NBtrainedarray[,i] <- NBtrainer(matrix(data[,i],ncol=nsamples-1),fac[-i])
  }
  return(NBtrainedarray)
}

```

```

NBleaveoneout <- function(NBtrainedarray , data , fac , topnumber=Inf , logfoldcut=0 , pcut=1){
  nsamples <- ncol(data)
  loglikelihoods <- rep(NA , nsamples)
  for (i in 1:nsamples){
    loglikelihoods[i]<- NBpredicter(matrix(data[,i],ncol=1),NBtrainedarray[,i],topnumber,logfoldcut,pcut)
  }
  return(loglikelihoods)
}

```

```

NBloocv <- function(NBtrainedarray , data , fac , topnumber=Inf , logfoldcut=0 , pcut=1){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])

```



```

SampleInformation <- paste(levels(fac)[1],"n1"," ",levels(fac)[2],"n2,sep="")
loglikelihoods <- NBleaveoneout(NBtrainedarray , data,fac,topnumber,logfoldcut,pcut)
names <- colnames(data , do.NULL=FALSE)
orden <- order(loglikelihoods , decreasing=TRUE)
Samples <- data.frame(names[orden],loglikelihoods[orden],fac[orden])
ROCdata <- myROC(loglikelihoods,fac)
SenSpe <- SensitivitySpecificity(loglikelihoods,fac)
return(list(SampleInformation=SampleInformation,ROCarea=ROCdata[1],p.value=ROCdata[2],
            topnumber=topnumber,pcut=pcut,SenSpe <- SenSpe,samples=Samples))
}

NBtwooutpreparer <- function(data , fac){
  nsamples <- ncol(data)
  ngenes <- nrow(data)
  NBdoublearray <- array(NA , dim=c(ngenes,5,nsamples*(nsamples-1)/2))
  for (i in 2:nsamples){
    for (j in 1:(i-1)){ print(paste(i," ",j));
      NBdoublearray[,,(i-1)*(i-2)/2+j] <- NBtrainer(matrix(data[,c(i,j)],ncol=nsamples-2),fac[-c(i,j)])
    }
  }
  return(NBdoublearray)
}

NBmaximizer <- function(NBtrainedarray , data , fac){
  functomaximize <- function(pcut , topnumber){
    NBloocv(NBtrainedarray , data , fac , topnumber=top , pcut=pcut)$ROCarea
  }
  rocmax <- 0
  pcutmax <- numeric(0)
  topmax <- numeric(0)
  pcutset <- c(1,0.05,0.01,0.005,0.001, 0.0003 , 0.0005,0.0001)
  topset <- c(1,2,5,10,20,50,100)
  for (pcut in pcutset){
    for (top in topset){
      currentroc <- functomaximize(pcut,top); # print(paste(pcut," ",top," ",currentroc))
      if (currentroc >= rocmax){
        rocmax <- currentroc
        pcutmax <- pcut
        topmax <- top
      }
    }
  }
  print(paste("Result ",pcutmax," ",topmax," ",rocmax))
  return(c(pcutmax,topmax))
}

NBtotalvalidation <- function(NBdoublearray , NBtrainedarray , data , fac){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  nsamples <- n1+n2
  ngenes <- nrow(data)
  SampleInformation <- paste(levels(fac)[1],"n1"," ",levels(fac)[2],"n2,sep="")
  maxarray <- matrix(NA , nrow=nsamples , ncol=2)
  colnames(maxarray) <- c('pcut','topnumber')
  NormScore <- numeric(nsamples)
  loglikelihoods <- numeric(nsamples)
  for (i in 1:nsamples){
    NBtemptrainedarray <- array(NA , dim=c(ngenes,5,nsamples-1))
    if (i > 1){
      for (j in 1:(i-1)){
        NBtemptrainedarray[,j] <- NBdoublearray[,,(i-1)*(i-2)/2+j]
      }
    }
    if (i < nsamples){
      for (j in (i+1):nsamples){
        NBtemptrainedarray[,j-1] <- NBdoublearray[,,(j-1)*(j-2)/2+i]
      }
    }
  }
}

```

[0285]

```

maxarray[i,] <- NBmaximizer( NBtemptrainedarray , data[,i] , fac[-i])
temploglikelihoods <- NBpredicter( data, NBtrainedarray[,i] , pcut = maxarray[i,1] , topnumber=maxarray[i,2])
loglikelihoods[i] <- temploglikelihoods[i]
meanll <- mean(temploglikelihoods[-i])
sdll <- sd(temploglikelihoods[-i])
NormScore[i] <- (temploglikelihoods[i] - meanll)/sdll
}
names <- colnames(data , do.NULL=FALSE)
orden <- order( NormScore , decreasing=TRUE)
Samples <- data.frame(names[orden],NormScore[orden],loglikelihoods[orden],fac[orden] ,maxarray[orden,])
ROCdata <- myROC(NormScore,fac)
SenSpe <- SensitivitySpecificity(NormScore,fac)
return(list(SampleInformation=SampleInformation,ROCarea=ROCdata[1],p.value=ROCdata[2],
            ,SenSpe <- SenSpe,samples=Samples))
}

ROCplot <- function(clasRes , sensspecnumber=6){
  Sensitivity <- as.numeric(clasRes[[sensspecnumber]][,1])
  Specificity <- as.numeric(clasRes[[sensspecnumber]][,2])
  OneMinusSpecificity <- 1- Specificity
  ROCarea <- round(clasRes$ROC,digits=2)
  plot(OneMinusSpecificity , Sensitivity , type="l" , xlab="1-specificity" , ylab="sensitivity")
  title(paste("ROC area = " ,ROCarea),font.main=1)
}

ROCplotReverse <- function(clasRes){
  Sensitivity <- rev(as.numeric(clasRes[[4]][,2]))
  Specificity <- rev(as.numeric(clasRes[[4]][,1]))
  OneMinusSpecificity <- 1- Specificity
  ROCarea <- round(clasRes$ROC,digits=2)
  plot(OneMinusSpecificity , Sensitivity , type="l" , xlab="1-specificity" , ylab="sensitivity")
  title(paste("ROC area = " ,ROCarea),font.main=1)
}

```

[0286]

(iii) ChP-PaC train1+test1 model

```

> str(svmtrain)
List of 29
 $ call      : language svm.default(x = t(training[aprioriBoolean, ]), y = facTr, kernel = "linear")
 $ type      : num 0
 $ kernel    : num 0
 $ cost      : num 1
 $ degree    : num 3
 $ gamma     : num 0.04
 $ coef0     : num 0
 $ nu        : num 0.5
 $ epsilon   : num 0.1
 $ sparse    : logi FALSE
 $ scaled    : logi [1:25] TRUE TRUE TRUE TRUE TRUE TRUE ...
 $ x.scale   :List of 2
 .. $ scaled:center: Named num [1:25] 3.14 3.42 3.28 3.33 2.66 ...
 .. .. attr(*, "names")= chr [1:25] "ApoA1_001_C08" "APOA4_5" "C1e_032_A11" "C5.12" ...
 .. $ scaled:scale : Named num [1:25] 0.117 0.157 0.183 0.109 0.112 ...
 .. .. attr(*, "names")= chr [1:25] "ApoA1_001_C08" "APOA4_5" "C1e_032_A11" "C5.12" ...
 $ y.scale   : NULL
 $ nclasses  : int 2
 $ levels    : chr [1:3] "ChP" "PaC" "rest"
 $ tot.nSV   : int 97
 $ nSV       : int [1:2] 49 48
 $ labels    : int [1:2] 1 2
 $ SV        : num [1:97, 1:25] -0.407 0.147 -0.912 -0.256 0.478 ...
 .. attr(*, "dimnames")=List of 2
 .. .. $ : chr [1:97] "R2S1B7" "R4S5A1" "R1S1B4" "R1S8A3" ...
 .. .. $ : chr [1:25] "ApoA1_001_C08" "APOA4_5" "C1e_032_A11" "C5.12" ...
 $ index     : int [1:97] 1 3 6 8 13 16 20 21 22 23 ...
 $ rho       : num 0.0805
 $ compprob  : logi FALSE

```

```

$ probA      : NULL
$ probB      : NULL
$ sigma      : NULL
$ coefs      : num [1:97, 1] 1 1 0.0466 1 1 ...
$ na.action   : NULL
$ fitted      : Factor w/ 3 levels "ChP","PaC","rest": 1 1 1 1 1 1 2 1 1 ...
..- attr(*, "names")= chr [1:204] "R2S1B7" "R1S2A6" "R4S5A1" "R3S5A2" ...
$ decision.values: num [1:204, 1] 0.935 2.207 0.329 3.534 1.665 ...
..- attr(*, "dimnames")=List of 2
.. ..$ : chr [1:204] "R2S1B7" "R1S2A6" "R4S5A1" "R3S5A2" ...
.. ..$ : chr "ChP/PaC"
- attr(*, "class")= chr "svm"
>

```

(iv) N-PaC train1+test1 model

```

[0287] > str(svmtrain)
List of 29
 $ call      : language svm.default(x = t(training[aprioriBoolean, ]), y = facTr, kernel = "linear")
 $ type      : num 0
 $ kernel    : num 0
 $ cost      : num 1
 $ degree    : num 3
 $ gamma     : num 0.04
 $ coef0     : num 0
 $ nu        : num 0.5
 $ epsilon   : num 0.1
 $ sparse    : logi FALSE
 $ scaled    : logi [1:25] TRUE TRUE TRUE TRUE TRUE TRUE ...
 $ x.scale   :List of 2
 ..$ scaled:center: Named num [1:25] 3.53 3.37 3.11 3.26 3.72 ...
 ..- attr(*, "names")= chr [1:25] "Angiomotin.1" "APOA4_2" "BITM8.001.B04" "C1e_032_A11" ...
 ..$ scaled:scale : Named num [1:25] 0.135 0.186 0.235 0.192 0.177 ...
 ..- attr(*, "names")= chr [1:25] "Angiomotin.1" "APOA4_2" "BITM8.001.B04" "C1e_032_A11" ...
 $ y.scale   : NULL
 $ nclasses  : int 2
 $ levels    : chr [1:3] "N" "PaC" "rest"
 $ tot.nSV   : int 15
 $ nSV       : int [1:2] 8 7
 $ labels    : int [1:2] 1 2
 $ SV        : num [1:15, 1:25] -0.392 -1.505 0.246 0.174 1.489 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:15] "R3S6A1" "R3S1A5" "R4S5A5" "R3S5A3" ...
 .. ..$ : chr [1:25] "Angiomotin.1" "APOA4_2" "BITM8.001.B04" "C1e_032_A11" ...
 $ index     : int [1:15] 1 9 14 15 16 17 19 20 27 69 ...
 $ rho       : num 1.4
 $ compprob  : logi FALSE
 $ probA     : NULL
 $ probB     : NULL
 $ sigma     : NULL
 $ coefs     : num [1:15, 1] 0.0642 0.0156 0.0414 0.0302 0.0184 ...
 $ na.action : NULL
 $ fitted    : Factor w/ 3 levels "N","PaC","rest": 1 1 1 1 1 1 1 1 1 ...
 ..- attr(*, "names")= chr [1:124] "R3S6A1" "R6S1A3" "R3S4B3" "R6S2B1" ...
 $ decision.values: num [1:124, 1] 1 1.88 1.2 1.67 2.61 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:124] "R3S6A1" "R6S1A3" "R3S4B3" "R6S2B1" ...
 .. ..$ : chr "N/PaC"
 - attr(*, "class")= chr "svm"
>

```

(v) Script ChP-PaC train1+test1

```

apri<-c("IL-1a-145 ",
" IL-4-55 ",
" TGF-b1-34 ",

```

```

" C-BTK-1 ",
" C-UBE2C-1 ",
" C-PTPN1-3 ",
" CT17",
" C5-12 ",
" Sox11A_A6 ",
" C-HADH2-8 ",
" MCP1_001_A04 ",
" C-TNFRSF3-1 ",
" C-UCHL5-1 ",
" TNF-a-89 ",
" C-TNFRSF3-2 ",
" Eotaxin-2 ",
" MCP1_005_A11 ",
" Cle_032_A11 ",
" APOA4_5 ",
" ApoA1_001_C08 ",
" C-MK08-1 ",
" C-FASN-6 ",
" IL-16-1 ",
" C-Keratin19-3 ",
" C-GAK-5 ")

library(MASS)
library(gplots)
library(e1071)
#source("C:/Program/R/R-2.8.1/library/NaiveBayesian")
source("NaiveBayesian")

filnamn<-"serum ChP-PaC train1+test1.txt"
rawfile <- read.delim(filnamn)
samplenames <- as.character(rawfile[,1])
groups <- rawfile[,2]
data <- t(rawfile[, -c(1,2)])
ProteinNames <- read.delim(filnamn,header=FALSE)
ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
ProteinNames <- ProteinNames[-(1:2)]
rownames(data) <- ProteinNames
colnames(data) <- samplenames

group1 <- "ChP"
group2 <- "PaC"

nTrainingSamples <- 204
nTestSamples <- 102

training <- data[,1:nTrainingSamples ]
test <- data[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)]

aprioriBoolean <- is.element(rownames(data) , apri)
facTr <- factor(rep("rest",ncol(training)),levels=c(group1, group2, "rest"))
subset1Tr <- is.element(groups[1:nTrainingSamples] , group1)
subset2Tr <- is.element(groups[1:nTrainingSamples] , group2)
facTr[subset1Tr] <- group1
facTr[subset2Tr] <- group2

facTe <- factor(rep("rest",ncol(test)),levels=c(group1, group2, "rest"))
subset1Te <- is.element(groups[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)] , strsplit(group1,"")[[1]])
subset2Te <- is.element(groups[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)] , strsplit(group2,"")[[1]])
facTe[subset1Te] <- group1
facTe[subset2Te] <- group2

svmtrain <- svm(t(training[aprioriBoolean,]) , facTr, kernel="linear")
pred <- predict(svmtrain, t(test[aprioriBoolean,]) , decision.values = TRUE, probability = T)
res <- as.numeric(attributes(pred)$decision.values, probability = T)

```

```

facnames <- colnames(attributes(pred)$decision.values)[1]
ROCdata <- myROC(res,facTe)
ROCdata[1]

SenSpe <- SensitivitySpecificity(res,facTe)
Sensitivity <- as.numeric(SenSpe[,1])
Specificity <- as.numeric(SenSpe[,2])
omSpecificity <- 1-Specificity
plot(omSpecificity, Sensitivity, ylab="Sensitivity", xlab="1-Specificity",type="l")
mtext(side=1, line = -1.1, paste("ROC AUC = ",signif(ROCdata[1], digits=2)))

ROCdata

Sensitivity+Specificity
max(Sensitivity+Specificity)

SenSpe

```

(vi) Script N-PaC train1+test1

[0289]

```

apri<-c("C-IL6-5 ",
" CystC_002_A01 ",
" IL-8-10 ",
" IL-11-42 ",
" Cle_032_A11 ",
" Eotaxin-2 ",
" C-HADH2-8 ",
" FB_009_B01 ",
" C5-9 ",
" CD40-44 ",
" C-CDK2-1 ",
" FN3-001-B04 ",
" Angiomotin-1 ",
" LDL-2 ",
" C5_029_C08 ",
" Clq-4 ",
" IL-1a-122 ",
" C-CDK2-2 ",
" Cle_032_E05 ",
" APOA4_2 ",
" FB_009_E05 ",
" C-P85A-4 ",
" C-BTK-1 ",
" BITM8-001-B04 ",
" IL-13-16 ")

library(MASS)
library(gplots)
library(e1071)
#source("C:/Program/R/R-2.8.1/library/NaiveBayesian")
source("NaiveBayesian")

filnamn<-"serum N-PaC train1+test1.txt"
rawfile <- read.delim(filnamn)
samplernames <- as.character(rawfile[,1])
groups <- rawfile[,2]
data <- t(rawfile[,-c(1,2)])
ProteinNames <- read.delim(filnamn,header=FALSE)
ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
ProteinNames <- ProteinNames[-(1:2)]
rownames(data) <- ProteinNames
colnames(data) <- samplernames

group1 <- "N"
group2 <- "PaC"

```

```

nTrainingSamples <- 124
nTestSamples <- 62

training <- data[,1:nTrainingSamples]
test <- data[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)]

aprioriBoolean <- is.element(rownames(data), apri)
facTr <- factor(rep("rest", ncol(training)), levels=c(group1, group2, "rest"))
subset1Tr <- is.element(groups[1:nTrainingSamples], group1)
subset2Tr <- is.element(groups[1:nTrainingSamples], group2)
facTr[subset1Tr] <- group1
facTr[subset2Tr] <- group2

facTe <- factor(rep("rest", ncol(test)), levels=c(group1, group2, "rest"))
subset1Te <- is.element(groups[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)], strsplit(group1, ",")[[1]])
subset2Te <- is.element(groups[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)], strsplit(group2, ",")[[1]])
facTe[subset1Te] <- group1
facTe[subset2Te] <- group2

```

[0290]

```

svmtrain <- svm(t(training[aprioriBoolean,]), facTr, kernel="linear")
pred <- predict(svmtrain, t(test[aprioriBoolean,]), decision.values = TRUE, probability = T)
res <- as.numeric(attributes(pred)$decision.values, probability = T)
facnames <- colnames(attributes(pred)$decision.values)[1]
ROCdata <- myROC(res, facTe)
ROCdata[1]

SenSpe <- SensitivitySpecificity(res, facTe)
Sensitivity <- as.numeric(SenSpe[,1])
Specificity <- as.numeric(SenSpe[,2])
omSpecificity <- 1-Specificity
plot(omSpecificity, Sensitivity, ylab="Sensitivity", xlab="1-Specificity", type="l")
mtext(side=1, line = -1.1, paste("ROC AUC = ", signif(ROCdata[1], digits=2)))

ROCdata

Sensitivity+Specificity
max(Sensitivity+Specificity)

SenSpe

```

[0291] 表C-生物标志物和生物标志物组合的示例性区分能力 (PC v NC)

ROC-AUC	生物标志物签名
0.9	HADH2
0.83	TNFRSF3
0.93	HADH2+TNFRSF3
0.96	HADH2+TNFRSF3+GAK
0.95	HADH2+TNFRSF3+GAK+UPF3B
0.95	HADH2+TNFRSF3+GAK+UPF3B+整联蛋白 α -10

[0292]

- <110> 伊缪诺维亚公司
<120> 用于确定胰腺癌的方法、阵列和其用途

<140> PCT/EP2014/056630
<141> 2014-04-02

<150> GB 1305940.7
<151> 2013-04-02

<160> 2

<170> SeqWin99

<210> 1
<211> 242
<212> PRT
<213> 人类

[0001] <400> 1
Met Ala Glu Val Gln Leu Leu Glu Ser Gly Gly Gly Leu Val Gln Pro
1 5 10 15
Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser
20 25 30
Ser Tyr Gly Phe His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu
35 40 45
Trp Val Ser Leu Ile Ser Trp Asp Gly Gly Ser Thr Tyr Tyr Ala Asp
50 55 60
Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr
65 70 75 80
Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr
85 90 95
Tyr Cys Ala Arg Gly Thr Trp Phe Asp Pro Trp Gly Gln Gly Thr Leu
100 105 110

Val Thr Val Ser Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
115 120 125

Gly Gly Gly Ser Gln Ser Val Leu Thr Gln Pro Pro Ser Ala Ser Gly
130 135 140

Thr Pro Gly Gln Arg Val Thr Ile Ser Cys Ser Gly Ser Ser Ser Asn
145 150 155 160

Ile Gly Asn Asn Ala Val Asn Trp Tyr Gln Gln Leu Pro Gly Thr Ala
165 170 175

Pro Lys Leu Leu Ile Tyr Arg Asn Asn Gln Arg Pro Ser Gly Val Pro
180 185 190

Asp Arg Phe Ser Gly Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile
195 200 205

Ser Gly Leu Arg Ser Glu Asp Glu Ala Asp Tyr Tyr Cys Ala Ala Trp
210 215 220

[0002]

Asp Asp Ser Leu Ser Trp Val Phe Gly Gly Gly Thr Lys Leu Thr Val
225 230 235 240

Leu Gly

<210> 2

<211> 31

<212> PRT

<213> 人工序列

<220>

<223> 6 x 组氨酸亲和标签

<400> 2

Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp Tyr
1 5 10 15

Lys Asp Asp Asp Asp Lys Ala Ala Ala His His His His His His
20 25 30

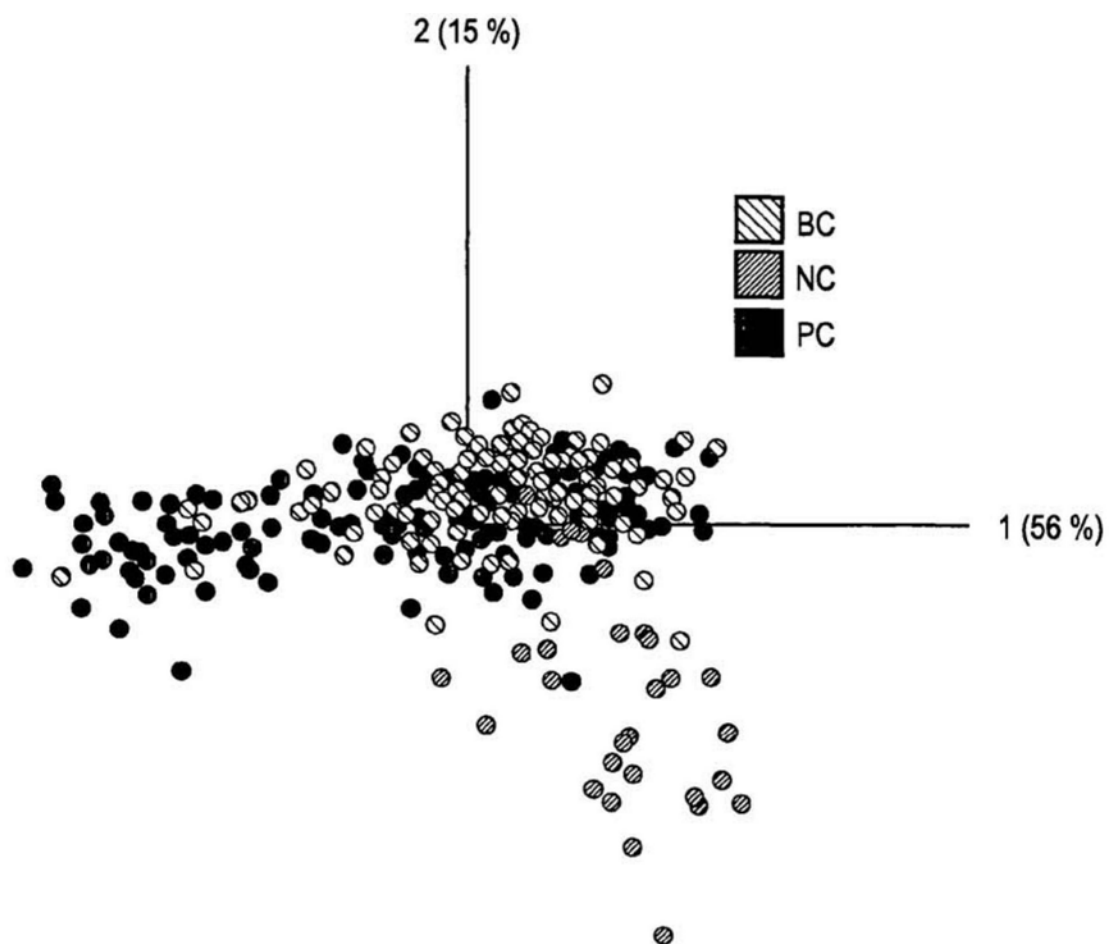


图1

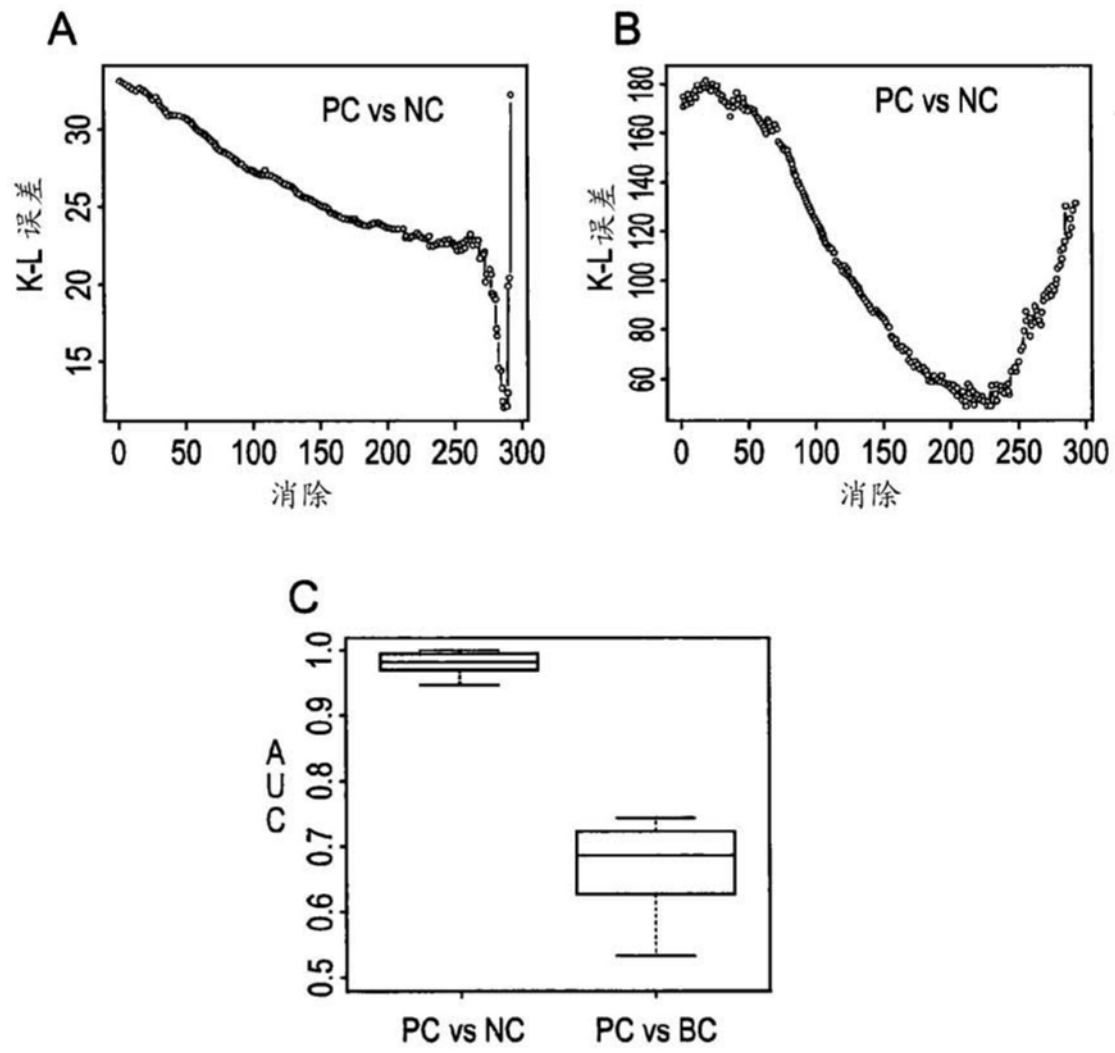


图2

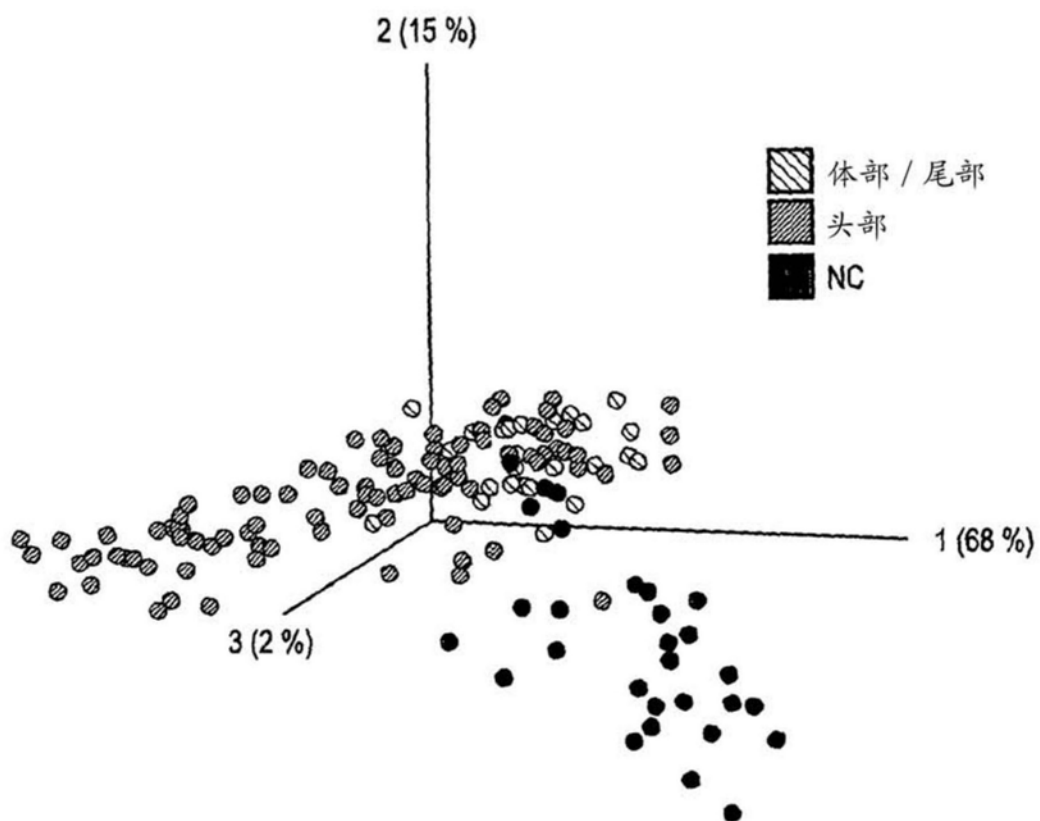


图3

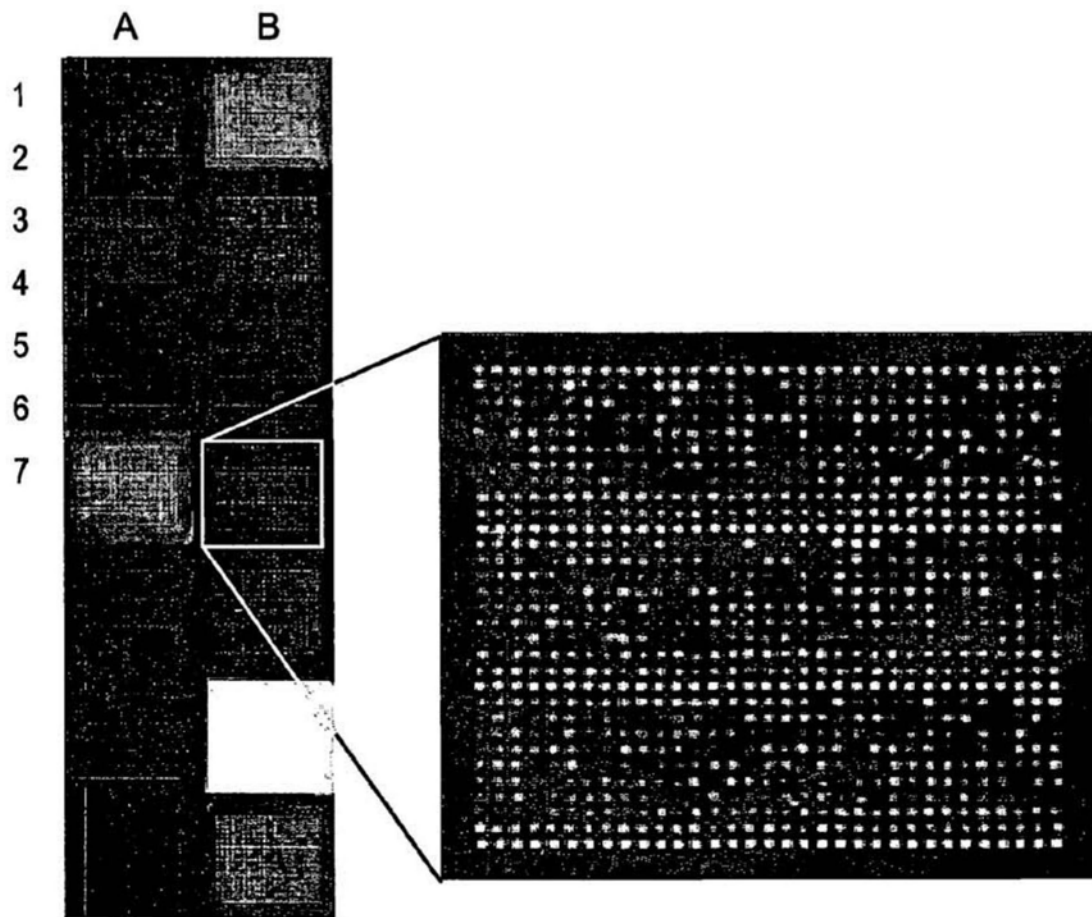


图4

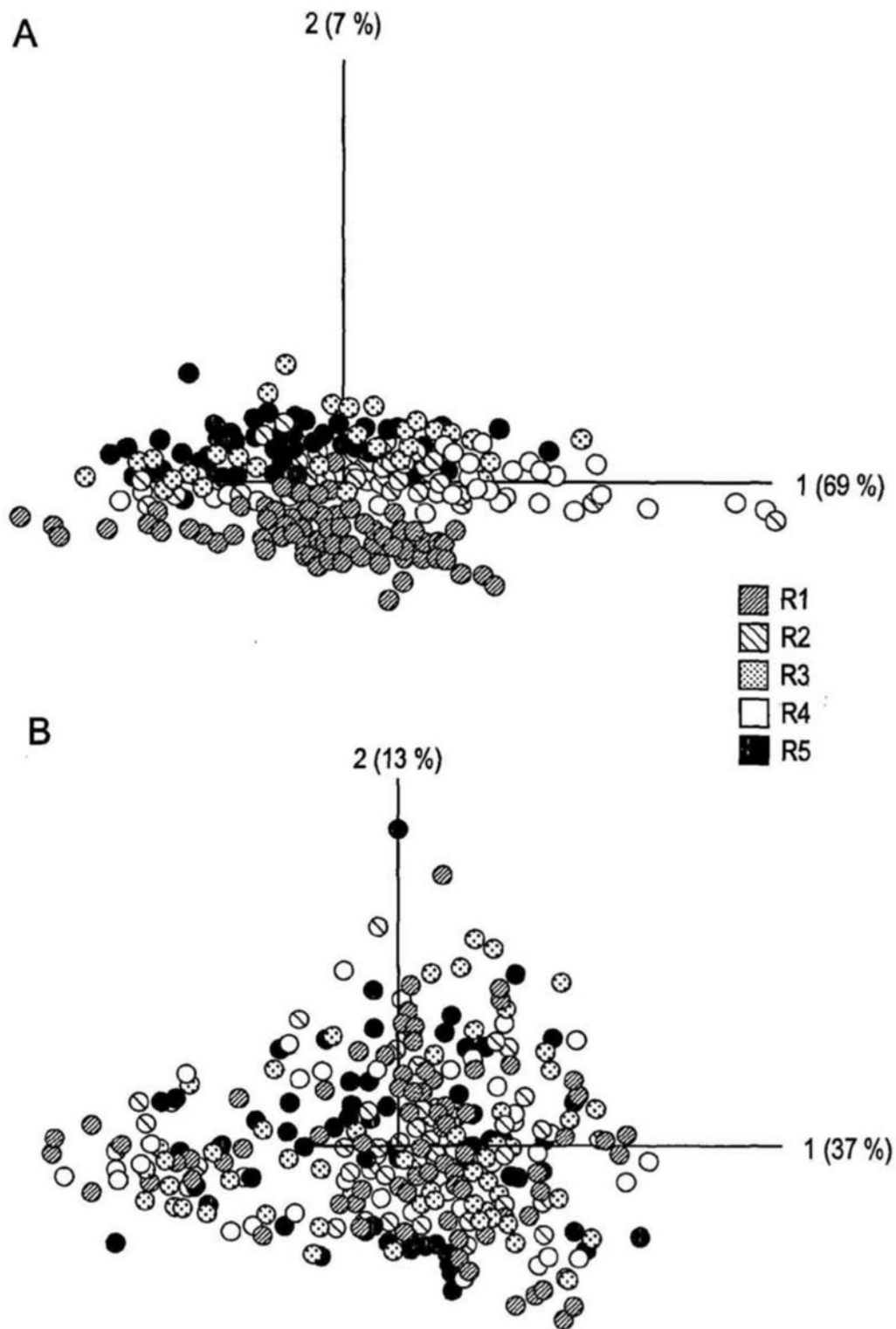


图5

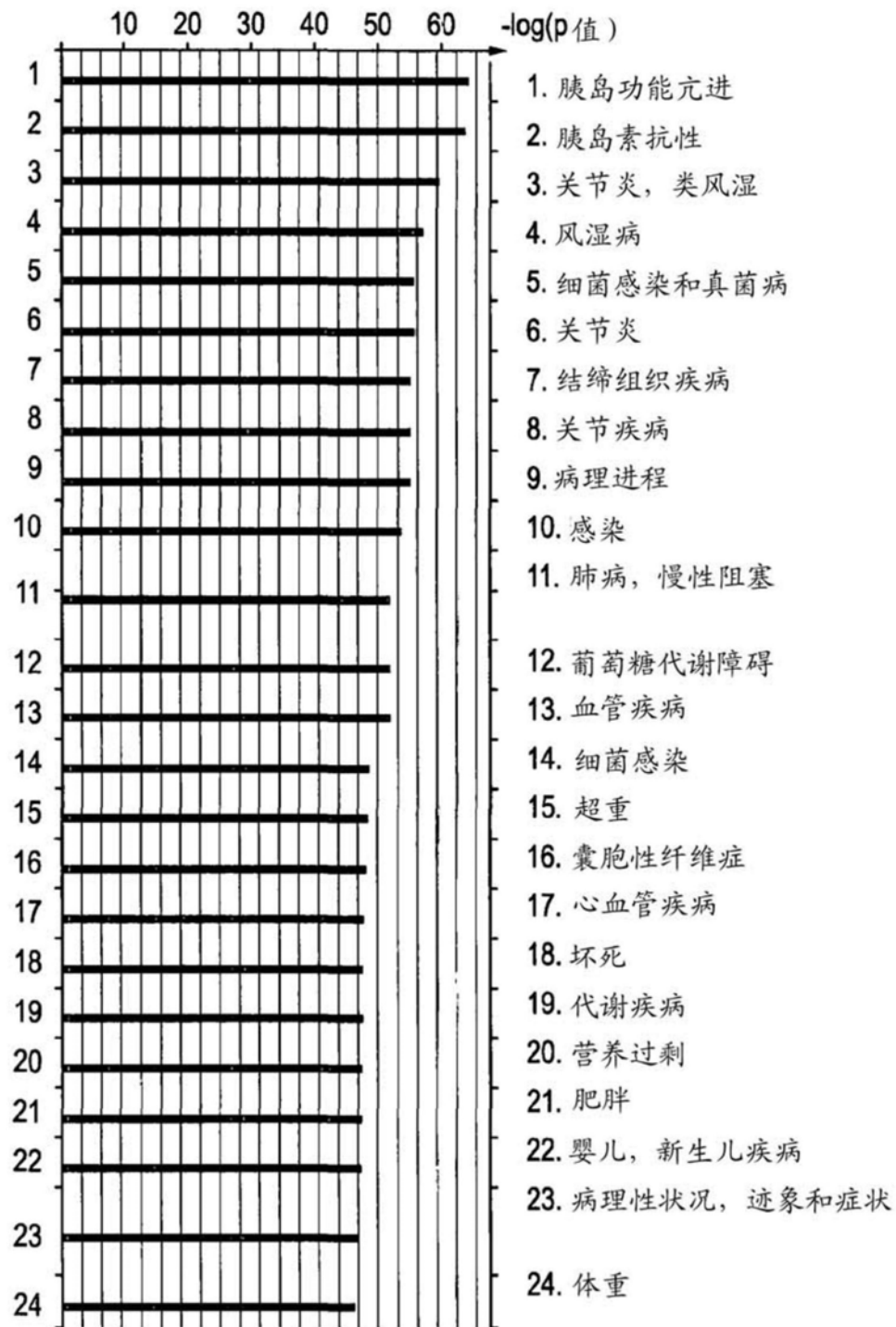


图6