



(12)发明专利申请

(10)申请公布号 CN 107782901 A

(43)申请公布日 2018.03.09

(21)申请号 201710403116.3

(22)申请日 2013.04.10

(30)优先权数据

1206323.6 2012.04.10 GB

(62)分案原申请数据

201380029458.3 2013.04.10

(71)申请人 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 曹立莉

(51)Int.Cl.

G01N 33/68(2006.01)

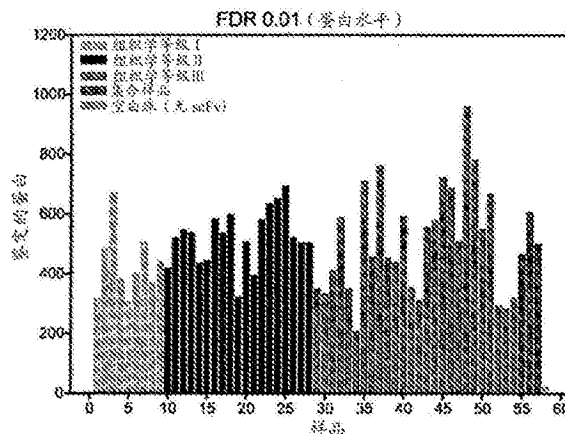
权利要求书1页 说明书50页 附图60页

(54)发明名称

用于确定乳癌相关疾病状态的方法和用于该方法中的阵列

(57)摘要

本发明提供了一种用于确定乳癌相关疾病状态的方法,其包括步骤:a)提供待测试的样品;和b)通过测量测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量来确定测试样品的生物标记物特征;其中测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量表示乳癌相关的疾病状态。本发明进一步提供了用于所述方法中的阵列和试剂盒。



1. 一种用于确定乳癌相关疾病状态的方法,其包括步骤:
  - a) 提供待测试的样品;和
  - b) 通过测量测试样品中一种或多种选自表1A、表1B和/或表1C中限定的生物标记物的存在和/或含量来确定测试样品的生物标记物特征;  
其中测试样品中一种或多种选自表1A、表1B和/或表1C中限定的生物标记物的存在和/或含量表示乳癌相关的疾病状态。
2. 根据权利要求1的方法,其中乳癌相关疾病状态是组织学等级和/或无转移生存时间。
3. 根据权利要求1或2的方法,其中乳癌相关疾病状态是乳癌细胞的组织学等级。
4. 根据权利要求3的方法,其中该方法进一步包括步骤:
  - c) 提供一种或多种包括组织学等级1乳癌细胞、组织学等级2乳癌细胞和/或组织学等级3乳癌细胞或由组织学等级1乳癌细胞、组织学等级2乳癌细胞和/或组织学等级3乳癌细胞组成的对照样品;和
  - d) 通过测量对照样品中步骤(b)中测量的一种或多种生物标记物的存在和/或含量来确定对照样品的生物标记物特征;  
其中测试样品中步骤(b)中测量的一种或多种生物标记物的存在和/或含量符合以下  
的情况下鉴定了乳癌细胞的存在:
    - i) 对应于包括第一个组织学等级的乳癌细胞或由第一个组织学等级的乳癌细胞组成的对照样品中的存在和/或含量(在存在的情况下);
    - ii) 不同于包括第二个组织学等级的乳癌细胞或由第二个组织学等级的乳癌细胞组成的对照样品中的存在和/或含量(在存在的情况下);和/或
    - iii) 不同于包括第三个组织学等级的乳癌细胞或由第三个组织学等级的乳癌细胞组成的对照样品中的存在和/或含量(在存在的情况下)。
5. 根据前述任一项权利要求的方法,其中使用一种或多种结合部分进行步骤(b)中测量一种或多种生物标记物的表达,每个所述结合部分能够选择性地结合编码表1中鉴定的生物标记物之一的核酸分子。
6. 一种用于根据前述任一项权利要求的方法中的阵列,该阵列包括如权利要求5中限定的一种或多种第一种结合剂。
7. 一种或多种选自表1A、表1B和/或表1C中限定的生物标记物用于确定乳癌相关疾病状态的用途。
8. 一种用于根据权利要求1至5中任一项的方法中的分析试剂盒,其包括:
  - C) 根据权利要求6的阵列;和
  - D) 用于进行如权利要求1至5任一项中限定的方法的说明书(任选)。
9. 基本上如本申请中限定的方法或用途。
10. 基本上如本申请中限定的阵列或试剂盒。

## 用于确定乳癌相关疾病状态的方法和用于该方法中的阵列

[0001] 本申请是中国发明专利申请(申请日:2013年4月10日;申请号:201380029458.3(国际申请号:PCT/IB2013/052858);发明名称:用于确定乳癌相关疾病状态的方法和用于该方法中的阵列)的分案申请。

### 发明领域

[0002] 本发明提供了用于确定乳癌相关疾病状态的方法,以及用于这些方法中的阵列和试剂盒。

### [0003] 发明背景

[0004] 乳癌(breast cancer)是最常诊断出来的癌症并且是女性中癌症死亡的主要诱因,占据了总癌症病例的23%和癌症相关死亡的14%(Jemal等,2011)。传统的临床病理学参数,如组织学等级、肿瘤大小、年龄、淋巴结牵涉和激素受体状况,用来决定治疗和估计预后(Ciocca和Elledge,2000;Elston和Ellis,1991;Hondermarck等,2008;Hudis,2007;Slamon等,2001)。组织学等级,其为最常用的预后因素之一,是组合评分,基于肿瘤细胞的形态学和细胞学特征的显微评价,反映出肿瘤的侵略性。然后将这种组合评分用于将乳癌分级成:级别1-缓慢生长和充分分化,级别2-适度分化,和级别3-高度增殖和差的分化(Elston和Ellis,1991)。然而,用于患者预后的组织学等级的临床值有争论,主要反映出目前与肿瘤分级相关的挑战(Frierson等,1995;Robbins等,1995)。此外,30-60%的肿瘤归类为组织学等级2,其已经变成代表非常不均匀的患者群并且证明了对于作出临床决定提供的信息较少(Sotiriou等,2006)。显然,传统临床实验室参数对于适当的预后和风险组辨别以及对于预测给予的治疗是否将成功仍然是不够的。因此,一些患者将被过度治疗、治疗不足或甚至用将不会给予任何益处的疗法进行治疗。因此,为了进一步推进预后,并且因此推进治疗结果,需要更深一步的乳癌生物学和肿瘤进展的分子理解与改进的个性化预后和治疗决定的方式的结合。

### 发明内容

[0005] 迄今为止,一组基因组成果已经产生了用于将乳癌类型分成亚组(Ivshina等,2006;Perou等,2000;Sorlie等,2001)以及用于乳癌预后和风险分级(Paik等,2004;van't Veer等,2002;van de Vijver等,2002)的分子特征(signature)。另一方面,已经预期蛋白质组学发现促进了关键的发现转变成临床实践(Hanash,2003)。关于这点,典型的基于质谱(MS)的蛋白质组学已经产生了有价值的乳癌蛋白质组库,主要靶向细胞系和少数的组织样品(Bouchal等,2009;Geiger等,2010;Geiger等,2012;Gong等,2008;Kang等,2010;Strande等,2009;Sutton等,2010),并且最近,亲和性蛋白质组学成果提出了用于乳癌诊断和用于预测复发风险的第一个多重血清谱(Carlsson等,2008;Carlsson等,2011)。尽管最近技术有进展,但是使用典型的蛋白质组学技术(Aebersold和Mann,2003)或亲和力蛋白质组学(Borrebaeck和Wingren,2011),以灵敏且可再现方式产生详细的大群粗蛋白组(例如,组织提取物)的蛋白质表达谱仍然是一种挑战。

[0006] 为了解决这些问题,发明人最近研发了整体蛋白质组检测(global proteome survey) (GPS) 技术平台(Wingren等,2009),其结合亲和力蛋白质组学和MS的最佳特征。GPS适用于以灵敏和定量方式发现尝试、再现地解密粗蛋白质组(Olsson等,2012;Olsson等,2011)。

[0007] 在这个研究中,发明人使用GPS深入地描绘了组织分级的乳癌组织的分子组织谱,反映出肿瘤进展。为此,分析了52个乳癌组织蛋白质组,以我们的知识来看,其代表了最大的无标记的基于LC-MS/MS的乳癌组织研究之一。使用正交方法成功证实了蛋白质表达谱。在长期运行中,这些组织生物标记物谱将为提高的分类和预后做好准备。

[0008] 因此,本发明的第一个方面提供了一种用于确定乳癌相关疾病状态的方法,其包括步骤:

[0009] a) 提供待测试的样品;和

[0010] b) 通过测量测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量来确定测试样品的生物标记物特征;

[0011] 其中测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量表示乳癌相关的疾病状态。因此,实际上,步骤(b)包括另外的步骤(b)(i):使用或基于测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量确定乳癌相关疾病状态的步骤。

[0012] “乳癌相关疾病状态”表示乳癌细胞的组织学等级和/或包括乳癌细胞的个体的无转移生存时间。

[0013] 乳癌相关疾病状态可以是(乳癌细胞的)组织学等级和/或(个体的)无转移生存时间。

[0014] “生物标记物”表示天然产生的生物分子,或其组分或片段,其测量可以提供用于乳癌预后的信息。例如,生物标记物可以是天然产生的蛋白质或碳水化合物部分,或其抗原组分或片段。

[0015] 优选,从哺乳动物提供待测试样品。哺乳动物可以是任何家畜或农场动物。优选,哺乳动物是大鼠、小鼠、豚鼠、猫、狗、马或灵长类。最优选,哺乳动物是人。优选,样品是包括乳癌细胞或由乳癌细胞组成的细胞或组织样品(或其衍生物),或同样优选,是源自包括乳癌细胞或由乳癌细胞组成的细胞或组织样品的蛋白质或核酸。优选,测试和对照样品源自相同物种。

[0016] 在乳癌相关疾病状态是或包括乳癌细胞的组织学分级的情况中,该方法可以进一步包括步骤:

[0017] c) 提供一种或多种包括组织学等级1乳癌细胞、组织学等级2乳癌细胞和/或组织学等级3乳癌细胞或由组织学等级1乳癌细胞、组织学等级2乳癌细胞和/或组织学等级3乳癌细胞组成的对照样品;和

[0018] d) 通过测量对照样品中步骤(b)中测量的一种或多种生物标记物的存在和/或含量来确定对照样品的生物标记物特征;

[0019] 其中在步骤(b)中测量的测试样品中一种或多种生物标记物的存在和/或含量符合以下的情况下鉴定了乳癌细胞的存在:

[0020] i) 对应于包括第一个组织学等级的乳癌细胞或由第一个组织学等级的乳癌细胞

组成的对照样品中的存在和/或含量(在存在的情况下)；

[0021] ii) 不同于包括第二个组织学等级的乳腺癌细胞或由第二个组织学等级的乳腺癌细胞组成的对照样品中的存在和/或含量(在存在的情况下)；

[0022] iii) 不同于包括第三个组织学等级的乳腺癌细胞或由第三个组织学等级的乳腺癌细胞组成的对照样品中的存在和/或含量(在存在的情况下)。

[0023] 因此,如果第一个组织学等级是Elston等级1,则第二个和第三个组织学等级(在存在的情况下)将是Elston等级2和Elston等级3(或反之亦然)。在第一个组织学等级是Elston等级2的情况下,第二个和第三个组织学等级(在存在的情况下)将是Elston等级1和Elston等级3(或反之亦然)。在第一个组织学等级是Elston等级3的情况下,第二个和第三个组织学等级(在存在的情况下)将是Elston等级1和Elston等级2(反之亦然)。

[0024] “对应于包括第一个组织学等级的乳腺癌细胞或由第一个组织学等级的乳腺癌细胞组成的对照样品中的存在和/或含量”表示存在和或含量等于包括第一个组织学等级的乳腺癌细胞或由第一个组织学等级的乳腺癌细胞组成的对照样品的存在和或含量;或与包括第二个组织学等级的乳腺癌细胞或由第二个组织学等级的乳腺癌细胞组成的对照样品和/或包括第三个组织学等级的乳腺癌细胞或由第三个组织学等级的乳腺癌细胞组成的对照样品相比,更接近于包括第一个组织学等级的乳腺癌细胞或由第一个组织学等级的乳腺癌细胞组成的对照样品的存在和或含量(或更接近于代表其的预定参照值)。优选,存在和/或含量为包括第一个组织学等级的乳腺癌细胞或由第一个组织学等级的乳腺癌细胞组成的对照样品的存在和或含量的至少60%,例如,至少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%。

[0025] “不同于包括第三个组织学等级的乳腺癌细胞或由第三个组织学等级的乳腺癌细胞组成的对照样品中的存在和/或含量”表示存在和或含量不同于包括第一个组织学等级的乳腺癌细胞或由第一个组织学等级的乳腺癌细胞组成的对照样品的存在和或含量;或不同于包括第二个组织学等级的乳腺癌细胞或由第二个组织学等级的乳腺癌细胞组成的对照样品的存在和或含量和/或包括第三个组织学等级的乳腺癌细胞或由第三个组织学等级的乳腺癌细胞组成的对照样品的存在和或含量(或不同于代表其的预定参照值)。优选,存在和/或含量不超过包括第二个组织学等级的乳腺癌细胞或由第二个组织学等级的乳腺癌细胞组成的对照样品和/或包括第三个组织学等级的乳腺癌细胞或由第三个组织学等级的乳腺癌细胞组成的对照样品的存在和或含量的40%,例如,不超过39%、38%、37%、36%、35%、34%、33%、32%、31%、30%、29%、28%、27%、26%、25%、24%、23%、22%、21%、20%、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%或0%。

[0026] 优选,组织学等级对照样品包括单个组织学等级的乳腺癌细胞或由单个组织学等级的乳腺癌细胞组成。优选,步骤(c)包括以下或由以下组成:

[0027] i) 提供一种或多种包括组织学等级1乳腺癌细胞或由组织学等级1乳腺癌细胞组成的对照样品;提供一种或多种包括组织学等级2乳腺癌细胞或由组织学等级2乳腺癌细胞组成的对照样品;和提供一种或多种包括组织学等级3乳腺癌细胞或由组织学等级3乳腺癌细胞组成的对照样品;

[0028] ii) 提供一种或多种包括组织学等级1乳癌细胞或由组织学等级1乳癌细胞组成的对照样品;和提供一种或多种包括组织学等级2乳癌细胞或由组织学等级2乳癌细胞组成的对照样品;

[0029] iii) 提供一种或多种包括组织学等级1乳癌细胞或由组织学等级1乳癌细胞组成的对照样品;和提供一种或多种包括组织学等级3乳癌细胞或由组织学等级3乳癌细胞组成的对照样品;

[0030] iv) 提供一种或多种包括组织学等级2乳癌细胞或由组织学等级2乳癌细胞组成的对照样品;和提供一种或多种包括组织学等级3乳癌细胞或由组织学等级3乳癌细胞组成的对照样品;

[0031] v) 提供一种或多种包括组织学等级1乳癌细胞或由组织学等级1乳癌细胞组成的对照样品;

[0032] vi) 提供一种或多种包括组织学等级2乳癌细胞或由组织学等级2乳癌细胞组成的对照样品;或

[0033] vii) 提供一种或多种包括组织学等级3乳癌细胞或由组织学等级3乳癌细胞组成的对照样品。

[0034] 在乳癌相关疾病状态是或包括个体的无转移生存时间的情况中,该方法可以进一步包括步骤:

[0035] c) 提供一种或多种包括来自具有低于10年无转移生存的个体的乳癌细胞或由来自具有低于10年无转移生存个体的乳癌细胞组成的第一个对照样品;和/或一种或多种包括来自具有10年或更长时间的无转移生存个体的乳癌细胞或由来自具有10年或更长时间的无转移生存个体的乳癌细胞组成的第二个对照样品;和

[0036] d) 通过测量对照样品中步骤(b)中测量的一种或多种生物标记物的存在和/或含量来确定对照样品的生物标记物特征;

[0037] 其中个体的无转移生存时间在步骤(b)中测量的一种或多种生物标记物的存在和/或含量对应于第一个对照样品的存在和/或含量(在存在的情况下)和/或不同于第二个对照样品的存在和/或含量(在存在的情况下)的情况下鉴定为少于10年;

[0038] 和其中个体的无转移生存时间在步骤(b)中测量的一种或多种生物标记物的存在和/或含量不同于第一个对照样品的存在和/或含量(在存在的情况下)和/或对应于第二个对照样品的存在和/或含量(在存在的情况下)的情况下鉴定为超过10年。

[0039] “对应于一种或多种第一个对照样品的存在和/或含量”表示存在和或含量与一种或多种第一个对照样品的存在和或含量相同;或与一种或多种第二个对照样品相比,更接近于第一个对照样品的存在和或含量(或代表其的预定参照值)。优选,存在和/或含量为第一个对照样品的存在和或含量的至少 60%,例如,至少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%。

[0040] “不同于一种或多种第二个对照样品的存在和/或含量”表示存在和或含量不同于第二个对照样品的存在和或含量(或代表其的约定参照值)。优选,存在和/或含量不超过第二个对照样品的存在和或含量的40%,例如,不超过39%、38%、37%、36%、35%、34%、

33%、32%、31%、30%、29%、28%、27%、26%、25%、24%、23%、22%、21%、20%、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%或0%。

[0041] 优选,一种或多种第一个和/或第二个无转移生存时间对照样品与待测试样品具有相同的组织学等级。

[0042] 优选,一种或多种对照样品是与待测试个体的年龄-和/或性别-相匹配的。换句话说,健康个体与待测试个体大致是相同年龄的(例如,5年内)和相同性别的。

[0043] 优选,将步骤(b)中测量的测试样品中的一种或多种生物标志物的存在和/或含量相对于预定的参照值进行比较。

[0044] 因此,优选步骤(b)中测量的测试样品中的一种或多种生物标记物的存在和/或含量显著不同于(即,统计学上不同于)步骤(d)中测量的一种或多种生物标记物的存在和/或含量或预定的参照值。例如,如所附实施例中讨论的,测试和对照样品中特定生物标记物的存在和/或含量之间的显著差异可以归类为其中 $p < 0.05$ 的那些(例如,其中 $p < 0.04$ , $p < 0.03$ , $p < 0.02$ 或其中 $p < 0.01$ )。

[0045] 因此,本发明的第一个方面的方法可以包括确定乳癌细胞的组织学等级和个体的无转移生存时间(同时或连续地)或由确定乳癌细胞的组织学等级和个体的无转移生存时间(同时或连续地)组成。

[0046] 优选,步骤(b)包括测量测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1中限定的生物标记物例如,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或至少79种选自表1中限定的生物标记物。

[0047] 因此,本发明的第一个方面可以包括用于确定乳癌细胞组织学等级的方法或由用于确定乳癌细胞组织学等级的方法组成(即,对乳癌样品分级,以确定组织学等级),所述方法包括步骤:

[0048] a) 提供待测试样品;

[0049] b) 通过测量测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量来确定测试样品的生物标记物特征;

[0050] 其中测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量表示乳癌细胞的组织学等级。

[0051] “确定乳癌细胞的组织学等级”表示将样品的乳癌细胞归类为组织学等级1(即,Elston等级1)、组织学等级2(即,Elston等级2)或组织学等级3(即,Elston等级3),如Elston,C.W.,和Ellis,I.O.(1991).Pathological prognostic factors in breast cancer.I.The value of histological grade in breast cancer: experience from a large study with long-term follow-up.Histopathology 19, 403-410中限定的,将其通过引用并入本申请中。

[0052] 在方法包括确定乳癌细胞组织学等级或由确定乳癌细胞组织学等级组成的情况

中,步骤(b)可以包括测量测试样品中一种或多种选自表1A中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1A中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1A中限定的生物标记物例如,至少2种选自表1A中限定的生物标记物。优选,步骤(b)包括测量测试样品中一种或多种选自表1B中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1B中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1B中限定的生物标记物例如,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或至少30种选自表1B中限定的生物标记物。优选,步骤(b)包括测量测试样品中一种或多种选自表1C中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1C中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1C中限定的生物标记物例如,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或至少28种选自表1C中限定的生物标记物。不太优选,步骤(b)包括测量测试样品中一种或多种选自表1D中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1D中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1D中限定的生物标记物例如,至少2、3、4、5、6、7、8、9或至少10种选自表1D中限定的生物标记物。还是不太优选,步骤(b)包括测量测试样品中一种或多种选自表1E中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1E中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1E中限定的生物标记物例如,至少2、3、4、5、6、7、8或至少9种选自表1E中限定的生物标记物。因此,步骤(b)可以包括测量测试样品中表1中限定的全部生物标记物的存在和/或含量或由测量测试样品中表1中限定的全部生物标记物的存在和/或含量组成。

[0053] 因此,本发明的第一个方面可以包括用于确定个体无转移生存时间的方法或由用于确定个体无转移生存时间的方法组成,所述方法包括步骤:

[0054] a) 提供待测试样品;

[0055] b) 通过测量测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量来确定测试样品的生物标记物特征;

[0056] 其中测试样品中一种或多种选自表1中限定的生物标记物的存在和/或含量表示个体的无转移生存时间。

[0057] “确定个体的无转移生存时间”表示预后从其获得测试样品的个体从初次诊断开始具有少于10年或超过10年的无转移生存时间(无远端转移的生存/DMFS)。

[0058] 在方法包括确定个体无转移生存或由确定个体无转移生存组成的情况中,步骤(b)可以包括测量测试样品中一种或多种选自表1A中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1A中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1A中限定的生物标记物例如,至少2种选自表1A中限定的生物标记物。优选,步骤(b)包括测量测试样品中一种或多种选自表1B中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1B中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1B中限定的生物标记物例如,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或至少30种选自表1B中限定的生物标记物。优选,步骤(b)包括测量测试样品中一种或多种选自表1D中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1D中限定的生物标记物的



存在和/或含量组成,所述一种或多种选自表1D中限定的生物标记物例如,至少2、3、4、5、6、7、8、9或至少10种选自表 1D中限定的生物标记物。优选,步骤(b)包括测量测试样品中表1A、表1B 和表1D中限定的全部生物标记物的存在和/或含量。

[0059] 在方法包括确定个体无转移生存或由确定个体无转移生存组成的情况中,尽管不太优选,但是步骤(b)可以包括测量测试样品中一种或多种选自表 1C中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1C中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表 1C中限定的生物标记物例如,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或至少 28种选自表1C中限定的生物标记物。还不太优选,步骤(b)包括测量测试样品中一种或多种选自表1E中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1E中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1E中限定的生物标记物例如,至少2、3、4、5、6、7、8或至少9种选自表1E中限定的生物标记物。还不太优选,步骤(b)可以包括测量测试样品中表1C和表1E中限定的全部生物标记物的存在和/或含量。

[0060] 因此,尽管不太优选,本发明第一个方面的方法可以包括确定个体无转移生存时间或由确定个体无转移生存时间组成,其中步骤(b)包括测量测试样品中标1中限定的全部生物标记物的存在和/或含量或由测量测试样品中标1 中限定的全部生物标记物的存在和/或含量组成。

[0061] 因此,根据本发明的第一个方面的方法可以包括测量SPON1表达。该方法可以包括测量KERA表达。该方法可以包括测量APCS表达。该方法可以包括测量ATP6V1G1表达。该方法可以包括测量RPS27L表达。该方法可以包括测量DPYSL3表达。该方法可以包括测量ERP44表达。该方法可以包括测量RAPGEF1表达。该方法可以包括测量ACLY表达。该方法可以包括测量CMA1表达。该方法可以包括测量MCM3表达。该方法可以包括测量ANGPTL2表达。该方法可以包括测量AEBP1表达。该方法可以包括测量UBE2V2表达。该方法可以包括测量MIS18BP1表达。该方法可以包括测量CLCF1表达。该方法可以包括测量ABAT表达。该方法可以包括测量SLC25A5表达。该方法可以包括测量STIP1表达。该方法可以包括测量 OLFL3表达。该方法可以包括测量CD3G表达。该方法可以包括测量MCM7 表达。该方法可以包括测量SLC25A11表达。该方法可以包括测量NOP56 表达。该方法可以包括测量RRP8表达。该方法可以包括测量SLTM表达。该方法可以包括测量TSN表达。该方法可以包括测量ECH1表达。该方法可以包括测量PRELP表达。该方法可以包括测量SARS表达。该方法可以包括测量RPS25表达。该方法可以包括测量ESYT1表达。该方法可以包括测量PODN表达。该方法可以包括测量RPRD1B表达。该方法可以包括测量RPLP0P6表达。该方法可以包括测量CD300LG表达。该方法可以包括测量SUGT1表达。该方法可以包括测量POTEF表达。该方法可以包括测量 KARS表达。该方法可以包括测量NDUFS2表达。该方法可以包括测量 HNRNPH2表达。该方法可以包括测量CALU表达。该方法可以包括测量 EIF3B表达。该方法可以包括测量SLC4A1AP表达。该方法可以包括测量 RPS5表达。该方法可以包括测量PLXDC2表达。该方法可以包括测量 KIAA1324表达。该方法可以包括测量MRC1表达。该方法可以包括测量 RPRD1A表达。该方法可以包括测量SHMT2表达。该方法可以包括测量CCT4表达。该方法可以包括测量TSSC1表达。该方法可以包括测量IKZF3 表达。该方法可以包括测量UBE2Q1表达。该方法可以包括测量PSMD9表达。该方法可以包括测量SNRNP70表达。该方法可以包括测量RALB表达。该方法可以包括测量AC02

表达。该方法可以包括测量MYO18A表达。该方法可以包括测量QARS表达。该方法可以包括测量PABPC4表达。该方法可以包括测量SCGB1D2表达。该方法可以包括测量PFKP表达。该方法可以包括测量SLC3A2表达。该方法可以包括测量ASPN表达。该方法可以包括测量CD38表达。该方法可以包括测量MXRA5表达。该方法可以包括测量CDK1表达。该方法可以包括测量STC2表达。该方法可以包括测量CTSC 表达。该方法可以包括测量NOP58表达。该方法可以包括测量PGK1表达。该方法可以包括测量FKBP3表达。该方法可以包括测量GSTM3表达。该方法可以包括测量CALML5表达。该方法可以包括测量PML表达。该方法可以包括测量ADAMTS4表达。该方法可以包括测量THBS1表达。该方法可以包括测量FN1表达。

[0062] 在一个优选的实施方案中,步骤(b)包括测量测试样品中一种或多种选自MCM7、NOP56、MCM3、RABPC4、MXRA5、STC2、SCGB1D2和ANGPTL2 的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自 MCM7、NOP56、MCM3、RABPC4、MXRA5、STC2、SCGB1D2和ANGPTL2 的生物标记物的存在和/或含量组成。例如,步骤(b)可以包括测量测试样品中的这些生物标记物中的2、3、4、5、6、7或8种的存在和/或含量或由测量测试样品中的这些生物标记物中的2、3、4、5、6、7或8种的存在和/或含量组成。优选,在这个实施方案中,乳癌相关疾病状态是组织学等级;然而,不太优选,乳癌相关疾病状态是或还包括无转移生存时间。

[0063] 在另一个优选实施方案中,步骤(b)包括测量测试样品中一种或多种选自OLFML3、SPON1、PODN和ASPN的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自OLFML3、SPON1、PODN和ASPN的生物标记物的存在和/或含量组成。例如,步骤(b)可以包括测量测试样品中的这些生物标记物中的2、3或4种的存在和/或含量或由测量测试样品中的这些生物标记物中的2、3或4种的存在和/或含量组成。优选,在这个实施方案中,乳癌相关疾病状态是组织学等级;然而,不太优选,乳癌相关疾病状态是或还包括无转移生存时间。

[0064] “表达”表示基因产物如mRNA或蛋白质的水平或含量。

[0065] 检测和/或测量蛋白质和/或核酸的浓度的方法是本领域技术人员公知的,参见,例如Sambrook and Russell,2001,Cold Spring Harbor Laboratory Press。

[0066] 用于检测和/或测量蛋白质的优选方法包括Western印迹、North-Western 印迹、免疫吸附测定(ELISA)、抗体微阵列、组织微阵列(TMA)、免疫沉淀、原位杂交和其他免疫组织化学技术、放射性免疫测定(RIA)、免疫放射性测定(IRMA)和免疫酶测定(IEMA),包括使用单克隆和/或多克隆抗体的夹层测定(sandwich assay)。示例性夹层测定描述于David等的美国专利4,376,110 和4,486,530中,通过引用并入本申请。载玻片上的细胞的抗体染色可以用于细胞学实验室诊断测试中公知的以及本领域技术人员已知的方法中。

[0067] 通常,ELISA涉及使用产生有色反应产物的酶,其通常在固相试验中。如辣根过氧化物酶和磷酸酶这样的酶已经被广泛使用。放大磷酸酶反应的一种方式是使用NADP作为底物,来产生NAD,其现在作为用于第二个酶系统的辅酶。因为组织中不存在酶,因此来自大肠杆菌的焦磷酸酶提供了良好的缀合物,其是稳定的并且产生了良好的反应颜色。还可以使用基于酶(如,萤光素酶)的化学发光系统。

[0068] 常常使用与维生素生物素的缀合,因为这可以通过其与酶联抗生物素蛋白(avidin)或抗生物素蛋白链菌素(streptavidin)的反应容易地检测出来,其以很高的特异性和亲和力与酶联抗生物素蛋白或抗生物素蛋白链菌素结合。

[0069] 用于核酸(例如,mRNA)的检测和/或测量的优选方法包括southern印迹、

northern印迹、聚合酶链式反应 (PCR)、逆转录酶PCR (RT-PCR)、定量实时 PCR (qRT-PCR)、纳米阵列、微阵列、大阵列 (macroarray)、放射自显影术和原位杂交。

[0070] 在本发明的第一个方面的一个实施方案中,步骤 (b) 包括测量编码一种或多种生物标记物的核酸分子的表达。核酸分子可以是cDNA分子或mRNA 分子。优选,核酸分子是mRNA分子。还优选,核酸分子是cDNA分子。

[0071] 因此,可以使用选自Southern印迹、Northern印迹、聚合酶链式反应 (PCR)、逆转录PCR (RT-PCR)、定量实时PCR (qRT-PCR)、纳米阵列、微阵列、大阵列、放射自显影术和原位杂交的方法来进行步骤 (b) 中一种或多种生物标记物的表达的测量。优选,使用DNA微阵列来确定步骤 (b) 中测量一种或多种生物标记物的表达。因此,该方法可以包括使用一种或多种结合部分 (moiety) 来测量步骤 (b) 中一种或多种生物标记物的表达或由使用一种或多种结合部分来测量步骤 (b) 中一种或多种生物标记物的表达组成,每个所述结合部分能够选择性地结合编码表1中鉴定的生物标记物之一的核酸分子。

[0072] 优选,一种或多种结合部分各自包含核酸分子或由核酸分子组成,所述核酸分子如DNA、RNA、PNA、LNA、GNA、TNA或PMO (优选DNA)。优选,一种或多种结合部分为5至100个核苷酸长。更优选,一种或多种核酸分子为15至35个核苷酸长。结合部分可以包括可检测部分。

[0073] 如以下讨论的,基于其结合给定的核酸、蛋白质或氨基酸基序的能力,从文库选择或筛选合适的结合剂 (也称为结合分子)。

[0074] 在本发明的第一个方面的另一个实施方案中,步骤 (b) 包括测量一种或多种生物标记物的蛋白质或多肽或其片段或衍生物的表达。优选,使用一种或多种结合部分进行步骤 (b) 中测量一种或多种生物标记物的表达,每个所述结合部分能够选择性地结合表1中鉴定的生物标记物之一。

[0075] 一种或多种结合部分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组成。

[0076] 术语“抗体”包括任何合成的抗体、重组抗体或抗体杂合物,如但不限于,由免疫球蛋白轻链和/或重链可变区和/或恒定区的噬菌体展示产生的单链抗体分子,或在本领域技术人员已知的免疫测定形式中能够结合抗原的其他免疫相互作用分子。本发明还包括使用抗体样结合剂,如亲和体 (affibody) 和适体。

[0077] 涉及保留了其特异性结合位点的抗体片段的合成的技术的综述可以在 Winter & Milstein (1991) Nature 349, 293-299中找到。

[0078] 另外地,或替换地,第一个结合分子中的一个或多个可以是适体 (参见 Collett 等, 2005, Methods 37:4-15)。

[0079] 分子文库,如抗体文库 (Clackson等, 1991, Nature 352, 624-628; Marks 等, 1991, J Mol Biol 222 (3) : 581-97)、肽文库 (Smith, 1985, Science 228 (4705) : 1315-7)、表达的cDNA文库 (Santi等, (2000) J Mol Biol 296 (2) : 497-508)、抗体框架以外的其他支架 (如亲和体) 上的文库 (Gunneriusson等, 1999, Appl Environ Microbiol 65 (9) : 4134-40) 或基于适体的文库 (Kenan等, 1999, Methods Mol Biol 118, 217-31), 可以用作从其选择用于本发明方法中的给定基序特异性的结合分子的来源。

[0080] 分子文库可以在原核细胞 (Clackson等, 1991, op.cit.; Marks等, 1991, op.cit.) 或真核细胞 (Kiecke等, 1999, Proc Natl Acad Sci USA, 96 (10) : 5651-6) 中体内表达或可

以在体外表达,而没有细胞的涉及(Hanes&Pluckthun,1997, Proc Natl Acad Sci USA 94 (10):4937-42;He&Taussig,1997,Nucleic Acids Res 25(24):5132-4;Nemoto等,1997, FEBS Lett,414(2):405-8)。

[0081] 在使用基于蛋白质的文库时的情况中,常常将编码潜在结合分子的文库的基因包装在病毒中,并且在病毒表面展示潜在的结合分子(Clackson等,1991, supra; Marks等,1991, supra; Smith,1985, supra)。

[0082] 可能最常使用的展示系统是其表面展示抗体片段的丝状噬菌体,所述抗体片段表达为与噬菌体的次要外壳蛋白的融合(Clackson等,1991, supra; Marks等,1991, supra)。然而,用于展示的其他合适系统包括使用其他病毒(EP 39678)、细菌(Gunneriusson等,1999, supra; Daugherty等,1998, Protein Eng 11(9):825-32; Daugherty等,1999, Protein Eng 12(7):613-21)和酵母(Shusta等,1999, J Mol Biol 292(5):949-56)。

[0083] 此外,已经利用所谓核糖体展示系统中的多肽产物与其编码mRNA的连接(Hanes& Pluckthun,1997, supra; He&Taussig,1997, supra; Nemoto 等,1997, supra),或替换地,多肽产物与编码DNA的连接(参见美国专利 5,856,090和WO 98/37186),研发了展示系统。

[0084] 抗体的可变重链( $V_H$ )和可变轻链( $V_L$ )结构域涉及抗原识别,这是早先首先在蛋白酶消化实验中认识到的事实。通过啮齿动物抗体的“人源化”发现了更多证实。啮齿动物来源的可变结构域可以与人来源的恒定结构域融合,使得所得到的抗体保留啮齿动物亲本抗体的抗原特异性(Morrison等,(1984)Proc.Natl.Acad.Sci.USA 81,6851-6855)。

[0085] 从涉及所有含有一个或多个可变结构域的抗体片段的细菌表达的实验已知,通过可变结构域给予抗原特异性并且与恒定结构域无关。这些分子包括Fab-样分子(Better等,(1988)Science 240,1041);Fv分子(Skerra等,(1988)Science 240,1038);单链Fv(ScFv)分子,其中 $V_H$ 和 $V_L$ 亲本结构域经由弹性寡核苷酸连接(Bird等(1988)Science 242,423; Huston等(1988)Proc. Natl.Acad.Sci.USA 85,5879)和包含分离的V结构域的单结构域抗体(dAb)(Ward等(1989)Nature 341,544)。涉及保留其特异性结合位点的抗体片段的合成的技术的综述可以在Winter&Milstein(1991)Nature 349,293-299 中找到。

[0086] 抗体或抗原结合片段可以选自完整抗体、Fv片段(例如,单链Fv和二硫化物键合的Fv)、Fab-样片段(例如,Fab片段,Fab'片段和 $F(ab)_2$ 片段)、单个可变结构域(例如, $V_H$ 和 $V_L$ 结构域)和结构域抗体(dAb,包括单一和双重形式[即,dAb-连接物-dAb])。优选,抗体或抗原结合片段是单链Fv(scFv)。

[0087] 一个或多个结合部分可以替换地包含抗体样结合剂或由抗体样结合剂组成,所述抗体样结合剂例如,亲和体或适体。

[0088] “scFv分子”表示其中 $V_H$ 和 $V_L$ 配偶体结构域通过弹性寡肽来连接。

[0089] 使用抗体片段,相比于完整抗体的优点是几倍的。片段的较小尺寸可以导致提高的药物学特性,如更好地渗透实体组织。除去了完整抗体的效应子功能,如补体结合。Fab、Fv、ScFv和dAb抗体片段全部可以在大肠杆菌中表达并分泌出来,因此允许容易地生产大量的所述片段。

[0090] 完整抗体和 $F(ab')_2$ 片段是“二价的”。“二价的”表示所述抗体和 $F(ab')_2$ 片段具有两个抗原结合位点。相反,Fab、Fv、ScFv和dAb片段是单价的,只具有一个抗原结合位点。

[0091] 抗体可以是单克隆或多克隆的。合适的单克隆抗体可以通过已知技术来制备,例

如,“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)中公开的那些,通过引用将这两篇并入本申请。

[0092] 当潜在结合分子选自文库时,通常使用具有限定基序的一个或多个选择剂肽。在用于选择剂肽的基序的设计中,可以使用提供结构从而降低肽中灵活性的氨基酸残基或允许与结合分子相互作用的带电的、极性或疏水性侧链。例如:

[0093] (i) 脯氨酸可以稳定肽结构,因为其侧链对 $\alpha$ 碳以及氮都能结合;

[0094] (ii) 苯丙氨酸、酪氨酸和色氨酸具有芳香族侧链并且是高度疏水性的,而亮氨酸和异亮氨酸具有脂肪族侧链并且也是疏水性的;

[0095] (iii) 赖氨酸、精氨酸和组氨酸具有碱性侧链,并且在中性pH下带正电,而天冬氨酸盐和谷氨酸盐具有酸性侧链并且在中性pH下是带负电的;

[0096] (iv) 天冬氨酸和谷氨酸在中性pH下是中性的,但含有参与氢键的酰胺基团;

[0097] (v) 丝氨酸、苏氨酸和酪氨酸侧链含有羟基基团,其可以参与氢键。

[0098] 通常,结合分子的选择可以涉及阵列技术和系统的使用,以分析与对应于结合分子类型的点的结合。

[0099] 因此,优选抗体或其片段是单克隆抗体或其片段。优选,抗体或抗原结合片段选自完整抗体、Fv片段(例如,单链Fv和二硫化物键合的Fv)、Fab-样片段(例如,Fab片段、Fab'片段和F(ab)<sub>2</sub>片段)、单个可变结构域(例如, V<sub>H</sub>和V<sub>L</sub>结构域)和结构域抗体(dAb,包括单一和双重形式[即,dAb-连接物-dAb])。因此,抗体或抗原结合片段可以是单链Fv(scFv)。或者,一个或多个结合部分包括抗体样结合剂或由抗体样结合剂组成,所述抗体样结合剂例如,亲和体或适体。一个或多个结合部分包括可检测部分。

[0100] “可检测部分”包括允许直接或间接测定其存在和/或相对含量和/或位置(例如,在阵列上的位置)的部分。

[0101] 合适的可检测部分是本领域公知的。

[0102] 例如,可检测部分可以是荧光和/或发光和/或化学发光部分,其暴露于特定条件时,可以检测到。这样的荧光部分可能需要暴露于特定波长和强度的照射(即,光),以引起荧光部分的激发,由此使其发射出可以检测到的特定波长的可检测荧光。

[0103] 或者,可检测部分可以是酶,其能够将(优选不可检测的)底物转化成可观察和/或检测到的可检测产物。合适的酶的实例在以下涉及例如ELISA测定中更详细地讨论。

[0104] 因此,可检测部分可以选自: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);或酶部分。优选,可检测部分包括放射性原子或由放射性原子组成。放射性原子可以选自镅-99m、碘-123、碘-125、碘-131、铟-111、氟-19、碳-13、氮-15、氧-17、磷-32、硫-35、氘、氚、铯-137、铯-138和钷-147。

[0105] 明显地,待检测物质(如,例如,本申请中所述的测试样品和/或对照样品中的一种或多种生物标记物和/或用于检测所选蛋白的抗体分子)必须具有足够的合适的原子同位素,使得能容易地检测到可检测部分。

[0106] 在可替换的优选实施方案中,结合部分的可检测部分是荧光部分。

[0107] 可以以已知的方式将放射性或其他标记结合至本发明方法的样品中存在的生物标记物中和/或本发明的结合部分中。例如,如果结合剂是多肽,其可以是生物合成的或可

以使用适当的氨基酸前体通过化学氨基酸合成来合成,其中涉及例如用氟-19替代氢。例如,可以经由结合部分中半胱氨酸残基连接标记物,如<sup>99m</sup>Tc、<sup>123</sup>I、<sup>186</sup>Rh、<sup>188</sup>Rh和<sup>111</sup>In。可以经由赖氨酸残基连接钇-90。IODOGEN方法(Fraker等(1978) *Biochem. Biophys. Res. Comm.* 80, 49-57)可以用于结合<sup>123</sup>I。参考文献(“*Monoclonal Antibodies in Immunoscintigraphy*”, J-F Chatal, CRC Press, 1989)详细描述了其他方法。用于将其他可检测部分(如酶、荧光、发光、化学发光或放射性部分)缀合至蛋白质的方法是本领域公知的。

[0108] 本领域技术人员将认识到可以用间接帮助确定所述蛋白质的存在、含量和/或位置的部分来标记待测试样品中的生物标记物。因此,该部分可以构成多组分可检测部分中的一种组分。例如,可以用生物素标记待测试样品中的生物标记物,这允许随后使用与可检测标记融合或以其它方式结合的抗生物素蛋白链菌素来检测。

[0109] 可检测部分可以选自荧光部分、发光部分、化学发光部分、放射性部分和酶部分。

[0110] 因此,可检测部分可以包括放射性原子或由放射性原子组成。放射性原子可以选自钆-99m、碘-123、碘-125、碘-131、铟-111、氟-19、碳-13、氮-15、氧-17、磷-32、硫-35、氩、氙、铯-137、铯-138和钇-90。

[0111] 或者,结合部分的可检测部分可以是荧光部分。

[0112] 在根据本发明的第一个方面的方法中,分别在步骤(b)和/或步骤(d)之前,处理步骤(a)和/或步骤(c)中提供的样品,使得用生物素标记样品中存在的任何生物标记物。可以使用包括抗生物素蛋白链菌素和可检测部分(如荧光部分)的检测剂来进行步骤(b)和/或步骤(d)。

[0113] 因此,可以首先使用第一种结合剂来分离和/或固定待测试样品中的目标蛋白质,此后,可以使用第二种结合剂来测定所述生物标记物的存在和/或相对含量。

[0114] 在一个实施方案中,第二种结合剂是抗体或其抗原结合片段;通常是重组抗体或其片段。方便地,抗体或其片段选自:scFv; Fab; 免疫球蛋白分子的结合结构域。以上详细描述了合适的抗体和片段,及其制备方法。

[0115] 或者,第二种结合剂可以是抗体样结合剂,如亲和体或适体。

[0116] 或者,在待测试样品中的蛋白质上的可检测部分包括特异性结合对的成员(例如,生物素)或由特异性结合对的成员(例如,生物素)组成的情况中,第二种结合剂可以包括特异性结合对的互补成员(例如,抗生物素蛋白链菌素)或由特异性结合对的互补成员(例如,抗生物素蛋白链菌素)组成。

[0117] 在使用检测测定的情况中,优选可检测部分选自:荧光部分;发光部分;化学发光部分;放射性部分;酶部分。以上描述了用于本发明方法中的合适的可检测部分的实例。

[0118] 用于检测血清或血浆蛋白的优选测定包括酶联免疫吸附测定(ELISA)、放射性免疫测定(RIA)、免疫放射性测定(IRMA)和免疫酶测定(IEMA),包括使用单克隆和/或多克隆抗体的夹层测定。示例性夹层测定描述于David等的美国专利4,376,110和4,486,530中,通过引用并入本申请。载玻片上的细胞的抗体染色可以用于细胞学实验室诊断测试中公知的以及本领域技术人员已知的方法中。

[0119] 因此,在一个实施方案中,测定是ELISA(酶联免疫吸附测定),其通常涉及使用产生有色反应产物的酶,其通常在固相测定中。如辣根过氧化物酶和磷酸酶这样的酶已经被广泛使用。放大磷酸酶反应的一种方式是使用NADP作为底物,来产生NAD,其现在作为用于

第二个酶系统的辅酶。因为组织中不存在酶,因此来自大肠杆菌的焦磷酸酶提供了良好的缀合物,其是稳定的并且产生了良好的反应颜色。还可以使用基于酶(如,荧光素酶)的化学发光系统。

[0120] 常常使用与维生素生物素的缀合,因为这可以通过其与酶联抗生物素蛋白或抗生物素蛋白链菌素的反应容易地检测出来,其以很高的特异性和亲和力与酶联抗生物素蛋白或抗生物素蛋白链菌素结合。

[0121] 在可替换的实施方案中,用于蛋白质检测的测定方便地是荧光测定。因此,第二种结合剂的可检测部分可以是荧光部分,如Alexa荧光团(例如,Alexa-647)。

[0122] 优选,如通过ROC AUC值测定的方法的预测准确性为至少0.50,例如,至少0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95、0.96、0.97、0.98或至少0.99。更优选,如通过ROC AUC值测定的方法的预测准确性为至少0.80(最优选1)。

[0123] 在本发明的第一个方面的方法中,可以使用如基于珠子的阵列或基于表面的阵列这样的阵列来进行。优选,阵列选自:大阵列;微阵列;纳米阵列。

[0124] 可以使用支持向量仪(SVM),如可从<http://cran.r-project.org/web/packages/e1071/index.html>(例如,e1071 1.5-24)获得的那些,来进行用于测定乳癌相关疾病状态的方法。然而,也可以使用任何其他合适的方式。SVM也可以用于测定包括如本申请中限定的一种或多种表1生物标记物或由本申请中限定的一种或多种表1生物标记物组成的生物标记物特征的ROC AUC。

[0125] 支持向量仪(SVM)是用于分类和回归的一组相关管理学习方法。给定一组训练实例,每个标记为属于两类中的一类,SVM训练算法构建预测新的实例是否落入一类或另一类的模型。直观地,SVM模型是作为空间中的点的实例表示,作图以使得分开类别的实例通过明显的尽可能宽的间隙分开。然后将新的实例作图至相同的空间中,并且基于它们落入间隙的哪一侧来预测属于哪一类。

[0126] 更正式地,支持向量仪在高或无穷的三维空间中构建了超平面或超平面组,其可以用于分类、回归或其他任务。直观地,通过对任何类别的最接近训练数据点具有最大距离的超平面来实现良好的分离(称为功能边缘),因为通常边缘越大,分类器的泛化误差越低。关于SVM的更多信息,参见,例如,Burges,1998,Data Mining and Knowledge Discovery, 2:121-167。

[0127] 在本发明的一个实施方案中,在使用已知物质(即,已知组织学等级的乳癌细胞或来自具有已知无远端转移生存的乳癌患者的乳癌细胞的生物标记物谱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能够知道与特定特征相关的生物标记物谱。一旦完成训练过程,然后SVM能够确定测试的生物标记物样品是否来自特定的乳癌样品类型(即,特定的乳癌相关疾病状态)。

[0128] 然而,可以通过用必需的训练参数对SVM进行预先编程来绕过这种训练程序。例如,使用其中详述的值和/或调整模式,基于表1中所列生物标记物的测量,使用表4中详述的SVM算法,根据已知的SVM参数,来鉴定属于特定乳癌相关疾病状态的细胞。

[0129] 本领域技术人员将认识到可以通过用合适的的数据选择训练SVM仪器,针对表1中所列的生物标记物的任意组合,来确定合适的SVM参数(即,来自已知生物学等级的细胞和/或来自具有已知无转移生存时间的个体的细胞的生物标记物测量)。

[0130] 或者,根据本领域已知的任何其他合适的统计学方法,如主成分分析 (PCA) 和其他多变量统计学分析(例如,向后逐步逻辑回归模型),表1数据可以用于确定特定的乳癌相关疾病状态。对于多变量统计学分析的综述,参见,例如,Schervish,Mark J. (1987年11月),“A Review of Multivariate Analysis”. *Statistical Science* 2(4):396-413,将其通过引用并入本申请中。

[0131] 优选,本发明的方法具有至少65%的准确性,例如,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%的准确性。

[0132] 优选,本发明的方法具有至少65%的灵敏度,例如,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。

[0133] 优选,本发明的方法具有至少65%的特异性,例如,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。

[0134] “准确性”表示方法的正确结果的比例,“灵敏度”表示正确归类为阳性的所有阳性化学物质的比例,和“特异性”表示正确归类为阴性的所有阴性化学物质的比例。

[0135] 本发明的第一个方面的方法可以进一步包括步骤:

[0136] e) 基于之前步骤中确定的乳癌相关疾病状态,给测试的个体提供治疗。

[0137] 因此,该方法包括根据其乳癌的组织学等级和/或根据其预测的无转移生存时间来治疗患者。例如,对于较高等级的乳癌和/或其中预测无转移生存时间相对低(例如,少于10年)/相对高(例如,超过10年),提供侵略性更高的治疗。可以根据当时的普遍指导,例如,参见,NICE Clinical Guideline 80“Early and locally advanced breast cancer: Diagnosis and treatment”(可从 <http://www.nice.org.uk/nicemedia/pdf/CG80NICEGuideline.pdf>获得),由本领域技术人员确定合适的治疗方法,将其通过引用并入本申请中。

[0138] 因此,本发明包括用于治疗乳癌的抗肿瘤剂,其中基于本发明第一个方面的方法的结果来确定剂量方案。

[0139] 本发明包括抗肿瘤剂在治疗乳癌中的用途,其中基于本发明第一个方面的方法的结果来确定剂量方案。

[0140] 本发明包括抗肿瘤剂在制备用于治疗乳癌的药物中的用途,其中基于本发明第一个方面的方法的结果来确定剂量方案。

[0141] 本发明包括治疗乳癌的方法,包括提供足量的抗肿瘤剂,其中基于本发明的第一个方面的方法的结果来确定足以治疗乳癌的抗肿瘤剂的含量。

[0142] 在一个实施方案中,抗肿瘤剂是烷化剂(ATC编码L01a)、抗代谢物(ATC 编码L01b)、植物生物碱或其他天然产物(ATC编码L01c)、细胞毒性抗生素或相关物质(ATC编码L01d)或其他抗肿瘤剂(ATC编码L01x)。



[0143] 因此,在一个实施方案中,抗肿瘤剂是选自以下的烷化剂:氮芥类似物(例如,环磷酰胺、苯丁酸氮芥、美法仑、氮芥、异环磷酰胺、氯乙环磷酰胺、泼尼氮芥或苯达莫司汀)、烷基磺酸盐(例如,白消安、苏消安或甘露舒凡)、乙撑亚胺(例如,噻替派、三亚胺醌或卡波醌)、亚硝基脲(例如,卡莫司汀、洛莫司汀、司莫司汀、链脲霉素、福莫司汀、尼莫司汀或雷莫司汀)、环氧化物(例如,依托格鲁)或另一种烷化剂(ATC编码L01ax,例如,二溴甘露醇、哌泊溴烷、替莫唑胺或达卡巴嗪)。

[0144] 在另一个实施方案中,抗肿瘤剂是选自以下的抗代谢物:叶酸类似物(例如,甲氨蝶呤、雷替曲塞、培美曲塞或普拉曲塞)、嘌呤类似物(例如,巯嘌呤、硫鸟嘌呤、克拉屈滨、氟达拉滨、克罗拉滨或奈拉滨)或嘧啶类似物(例如,阿糖胞苷、氟尿嘧啶、替加氟、卡莫氟、吉西他滨、卡培他滨、阿扎胞苷或地西他滨)。

[0145] 在再进一步的实施方案中,抗肿瘤剂是选自以下的植物碱或其它天然产物:长春花属生物碱或长春花属生物碱类似物(例如,长春花碱、长春新碱、长春地辛、长春瑞滨或长春氟宁)、鬼臼毒素衍生物(例如,依托泊苷或替尼泊苷)、秋水仙碱衍生物(例如,秋水仙胺)、紫杉烷(例如,紫杉醇、多西他赛或聚谷氨酸紫杉醇),或另一种植物碱或天然产物(ATC编码L01cx,例如,曲贝替丁)。

[0146] 在一个实施方案中,抗肿瘤剂是选自以下的细胞毒性抗生素或相关物质:放线菌素(例如,放线菌素D)、蒽环或相关物质(例如,阿霉素、柔红霉素、表柔比星、阿柔比星、佐柔比星、伊达比星、米托蒽醌、吡柔比星、戊柔比星、氨柔比星或匹杉琼(pixantrone))或另一种(ATC编号L01dc,例如,博来霉素、普卡霉素、丝裂霉素或伊沙匹隆)。

[0147] 在进一步的实施方案中,抗肿瘤剂是选自以下的其他抗肿瘤剂:铂化合物(例如,顺铂、卡铂、奥沙利铂、沙铂或polyplattin)、甲胍(例如,丙卡巴胍)、单克隆抗体(例如,依决洛单抗、利妥昔单抗、曲妥珠单抗、阿伦单抗、吉姆单抗、西妥昔单抗、贝伐单抗、帕尼单抗、卡妥索单抗或奥法木单抗)、用于光动力学疗法/放疗中的增敏剂(例如,卟吩姆钠、氨基乙酰丙酸甲酯、氨基乙酰丙酸、替莫泊芬或乙丙昔罗(efaproxiral))或蛋白激酶抑制剂(例如,伊马替尼、吉非替尼、厄洛替尼、舒尼替尼、索拉替尼、达沙替尼、拉帕替尼、尼洛替尼、替西罗莫司、依维莫司、帕唑帕尼、凡德他尼、阿法替尼、马赛替尼或toceranib)。

[0148] 在再进一步的实施方案中,抗肿瘤剂是选自以下的其他肿瘤剂:安吡啶、天冬酰胺酶、六甲蜜胺、羟基脲、氯尼达明、喷司他丁、米替福新、马索罗酚、雌氮芥、维A酸、米托胍脲、托泊替康、噻唑呋林、伊立替康、阿利维A酸、米托坦、培加帕酶、贝沙罗汀、三氧化二砷、地尼白介素(denileukin diftitox)、硼替佐米、塞来昔布、阿那格雷、奥利默森(oblimersen)、腺病毒载体定位码基因注射剂(sitimagene ceradenovec)、伏立诺他、罗米地辛、高三尖杉酯碱注射液(omacetaxine mepesuccinate)或艾日布林。

[0149] 因此,本发明的第二个方面提供了用于根据本发明第一个方面的方法中的阵列,该阵列包括如上关于本发明第一个方面限定的一种或多种的第一种结合剂。

[0150] 阵列结合剂可以包括共同地能够结合一种或多种选自表1A中限定的生物标记物的结合剂或由共同地能够结合一种或多种选自表1A中限定的生物标记物的结合剂组成,所述一种或多种选自表1A中限定的生物标记物例如,至少2种选自表1A中限定的生物标记物。优选,阵列结合剂可以包括共同地能够结合一种或多种选自表1B中限定的生物标记物的结合剂或由共同地能够结合一种或多种选自表1B中限定的生物标记物的结合剂组成,所述一

种或多种选自表1B中限定的生物标记物例如,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或至少30种选自表1B中限定的生物标记物。优选,阵列结合剂可以包括共同地能够结合一种或多种选自表1C中限定的生物标记物的结合剂或由共同地能够结合一种或多种选自表1C中限定的生物标记物的结合剂组成,所述一种或多种选自表1C中限定的生物标记物例如,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或至少28种选自表1C中限定的生物标记物。优选,阵列结合剂可以包括共同地能够结合一种或多种选自表1D中限定的生物标记物的结合剂或由共同地能够结合一种或多种选自表1D中限定的生物标记物的结合剂组成,所述一种或多种选自表1D中限定的生物标记物例如,至少2、3、4、5、6、7、8、9或至少10种选自表1D中限定的生物标记物。优选,阵列结合剂可以包括共同地能够结合一种或多种选自表1E中限定的生物标记物的结合剂或由共同地能够结合一种或多种选自表1E中限定的生物标记物的结合剂组成,所述一种或多种选自表1E中限定的生物标记物例如,至少2、3、4、5、6、7、8或至少9种选自表1E中限定的生物标记物。

[0151] 因此,阵列结合剂可以包括共同地能够结合表1A中限定的全部生物标记物的结合剂或由共同地能够结合表1A中限定的全部生物标记物的结合剂组成。阵列结合剂可以包括共同地能够结合表1B中限定的全部生物标记物的结合剂或由共同地能够结合表1B中限定的全部生物标记物的结合剂组成。阵列结合剂可以包括共同地能够结合表1C中限定的全部生物标记物的结合剂或由共同地能够结合表1C中限定的全部生物标记物的结合剂组成。阵列结合剂可以包括共同地能够结合表1D中限定的全部生物标记物的结合剂或由共同地能够结合表1D中限定的全部生物标记物的结合剂组成。阵列结合剂可以包括共同地能够结合表1E中限定的全部生物标记物的结合剂或由共同地能够结合表1E中限定的全部生物标记物的结合剂组成。优选,阵列结合剂包括共同地能够结合表1中限定的全部生物标记物的结合剂或由共同地能够结合表1中限定的全部生物标记物的结合剂组成。

[0152] 阵列中的第一种结合剂可以是固定的。

[0153] 阵列本身是本领域公知的。通常,它们由具有隔开(即,分离)区域(“点”)的线性或二维结构形成,每个具有有限的面积,在固体支持物的表面上形成。阵列还可以是珠子结构,其中每个珠子通过分子编码或颜色编码来识别或在连续流动中识别。在样品通过一系列点的情况中,还可以按序进行分析,每个点从溶液中吸附所述类别的分子。固体支持物通常是玻璃或聚合物,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。固体支持物可以是管子、珠子、圆盘、硅芯片、微板、聚乙二烯二氟化物(PVDF)膜、硝基纤维素膜、尼龙膜、其他多孔膜、无孔膜(例如,塑料、聚合物、塑胶、硅等)、多个聚合针,或多个微滴定孔,或任何其他适用于固定蛋白质、多核苷酸和其他合适分子和/或进行免疫测定的表面的形式。结合过程也是本领域公知的,并且通常由交联共价结合或物理吸附蛋白质分子、多核苷酸等至固体支持物上组成。或者,可以使用经由亲和力标签(tag)或相似构建体的探针的亲和力偶联。通过使用公知技术,如接触或非接触印刷、掩蔽或照相平版印刷术,可以限定每个点的位置。对于综述,参见,Jenkins,R.E., Pennington,S.R. (2001, Proteomics,2,13-29)和Lal等(2002,Drug Discov Today 15;7(18Suppl):S143-9)。

[0154] 通常,阵列是微阵列。“微阵列”包括具有至少约 $100/\text{cm}^2$ 的离散区域密度的区域的阵列的含义,并且优选至少 $1000/\text{cm}^2$ 。微阵列中的区域具有典型的尺寸,例如,直径,在约

10-250 $\mu$ m的范围中,并且与阵列中的其他区域隔开约相同的距离。阵列可替换地是大阵列或纳米阵列。

[0155] 一旦已经鉴定并分离合适的结合分子(以上讨论的),本领域技术人员可以使用分子生物学领域公知的方法来制造阵列;参见以下的实施例。

[0156] 本发明的第三个方面提供了一种或多种选自表1A、表1B、表1C、表1D和/或表1D中限定的生物标记物用于确定乳癌相关疾病状态的用途。

[0157] 在一个实施方案中,表1A、表1B、表1C、表1D和表1D中限定的全部生物标记物共同地用于确定乳癌相关疾病状态。

[0158] 本发明的第四个方面提供了用于根据本发明的第一个方面的方法中的分析试剂盒,其包括:

[0159] A) 根据本发明第二个方面的或如本发明第一个方面中限定的阵列;和

[0160] B) 用于进行如本发明第一个方面中限定的方法的说明书(任选)。

[0161] 分析试剂盒可以包括如本申请发明的第一个方面中限定的一种或多种对照样品。

[0162] 优选,现在将参照以下附图描述具体表现为本发明特定方面的非限制性实施例。

## 附图说明

[0163] 图1.肽和蛋白质统计学。(A) 每个样品鉴定的独特肽序列的总数(FDR 0.01,使用 Mascot+X!Tandem)。(B) 每个样品鉴定的装配蛋白质群组(protein group)的总数(FDR 0.01,设定在蛋白质水平,使用Mascot+X!Tandem)。(C) 每个蛋白质群组的独特肽的数量(FDR 0.01,设定在蛋白质水平,使用 Mascot+X!Tandem),在整个研究中形成2140个蛋白质群组的总蛋白质覆盖。(数据基于全部样品和运行,包括复制、集合样品运行和具有缺少的临床参数的样品)。(D) 相对于PeptideAtlas(版本2011-08Ens62,人)定量肽的评价(Progenesis LC-MS软件,限于使用FDR 0.01的Mascot评分的肽)。此外,对于PeptideAtlas中不存在的肽,为了评价是否已经报道了相应的蛋白质,进行第二个比较。在多个蛋白质就位的情况下,评价全部。(E) 肽长度的比较。(F) 在PeptideAtlas中观察到的肽频率。

[0164] 图2.完整GPS设定的再现性(即,捕获+LC-MS/MS)说明了针对代表性样品(样品7267)和参照(集合样品)样品的组合数据和单独的混合物。为了包括(绘图)数据点,必须在全部重复三次的运行中,将蛋白质定量(标准化丰度 >0)。这样的要求用于图A-E中绘制的所有数据。(A) 针对所有组合的数据进行了说明(基于1264个蛋白质)。(B) 针对CIMS-混合物1进行了说明(基于315个蛋白质)。(C) 针对CIMS-混合物2进行了说明(基于661个蛋白质)。(D) 针对CIMS-混合物3进行了说明(基于452个蛋白质)。(E) 针对CIMS-混合物4进行了说明(基于370个蛋白质)。

[0165] 图3.基于组织学等级、雌激素受体状况和HER2状况的显著差异表达的蛋白质。差异表达的分析物显示于热图(heatmap)中(红色-上调,绿色-下调)。(A) 组织学等级1、等级2和等级3样品的PCA-图和相关热图(在方差 0.2、p-值<0.01、q-值<0.25上过滤的数据)。此外,使用SVM的来自弃一交叉验证(leave-one out cross validation)方法的结果由ROC-面积值证实。(B) ER-阳性和ER-阴性样品的PCA-图和相关热图(在方差0.2、p-值<0.01、q-值 <0.32上过滤的数据)。此外,用ROC-曲线说明了来自弃一交叉验证方法的结果。(C) HER2-阳性和阴性样品的PCA-图和相关热图(在方差0.2、p-值 <0.01、q-值<0.9上过滤的数

据)。用ROC-曲线说明了来自弃一交叉验证方法的结果。

[0166] 图4.使用IPA的三个组织学等级的肿瘤类型之间的差异表达的分析物的生物学关联性。(A)将三个肿瘤群之间的鉴定为显著差异表达的蛋白质的49个蛋白质作图于它们的细胞位置。着色的 $\log_2$ -比例(中值等级3/中值等级1),其中红色说明上调,而绿色说明下调。已经指出具有已知的与肿瘤发生关联的蛋白质。(B)发现顶端报道的网络与DNA复制、重组、细胞周期和自由基清除相关。(C)发现第二个报道的网络与基因表达、传染病和癌症相关。

[0167] 图5.使用正交方法证实蛋白质表达谱。为此,使用了基于来自1411个组织学等级的肿瘤样品的mRNA表达谱。在组织学等级1、2和3中49个差异表达蛋白质中的42个成功作图(使用Gene Entrez ID)于GOBO-数据库中。(A)针对发现在组织学等级3肿瘤中展示出降低的蛋白质表达的蛋白质的mRNA表达谱(与组织学等级1相比的中值比例),其中15个(总共16个)分析物可以用GOBO-工具作图。此外,指出15个基因与不同基因群组模块(gene set module)表达谱的关联。灰色点表示实际的关联值。(B)针对发现在组织学等级3肿瘤中展示出提高的表达的蛋白质的mRNA表达谱(与组织学等级1相比),其中27个(总共33个)可以用GOBO-工具作图。此外,指出27个基因与不同基因群组模块表达谱的关联。灰色点表示实际的关联值。

[0168] 图6.表1生物标记物的不同长度和组合的示例性生物标记物特征的卡普兰-迈耶分析。

[0169] 图7.研究中使用的工作流的图示。(A)肿瘤样品制备和(B)使用CIMS-抗体的肽捕获,在LC-MS/MS上的运行时间表和数据分析。在移动至下一个CIMS-粘合剂混合物衍生的洗脱物之前,结束所有源自一个CIMS-粘合剂混合物的洗脱物的分析。因此,所有CIMS-粘合剂混合物分析从来自集合样品的洗脱物的分析开始,接着是根据组织学等级的随机顺序的一半个体样品的分析。通过混合物达到一半时,注入另一个集合样品,然后注入剩余样品,并且在结束时,用第三个集合样品结束。完成后,开始来自下一个CIMS-粘合剂混合物的洗脱物的分析。在粘合剂混合物2和粘合剂混合物3之间,运行两次空白珠的注入。空白珠不含连接的抗体,因此应当只洗脱结合磁珠的背景肽。在Proteios SE和Progenesis中分析数据,以获得肽和蛋白质的鉴定和定量。

[0170] 图8.作为Venn图说明的完整GPS设定(即,捕获+LC-MS/MS)的鉴定再现性。(A)对于样品7267的重复捕获运行之间的肽的重叠(所有独特序列)。显示了对于样品7267(上图)和单独混合物(较小的四个Venn图)的总覆盖的统计。从Proteios SE产生的数据(即,Mascot和X!Tandem评分的肽)。(B)对于集合样品的重复捕获运行之间的肽的重叠(所有独特序列)。显示了总(上图)和单独混合物的统计。从Proteios SE产生的数据(即,Mascot和X!Tandem评分的肽)。

[0171] 图9.对于定量蛋白质的 $\log_2$ MS强度的分布。(A)对于1364个蛋白质(排除了24个具有0的中值 $\log_2$ 强度值的蛋白质)绘制的中值标准化丰度(基于具有临床记录的50个样品)。条块根据MS强度着色,从淡黄色(低MS强度)至深红色(高MS强度)。(B)基于针对选定的蛋白质类别的GO生物学过程的 $\log_2$ MS强度的分布。使用Generic Gene Ontology(GO)Term Mapper工具(<http://go.princeton.edu/cgi-bin/GOTermMapper>),通过主要生物过程将分析物分组。

[0172] 图10. 在三个组织学等级之间的差异表达蛋白质中的8个的单独强度箱线图, 证明了组织学等级3肿瘤中的最高表达。

[0173] 图11. 在三个组织学等级之间的差异表达蛋白质中的8个的单独强度箱线图, 证明了组织学等级1肿瘤中的最高表达。

[0174] 图12. 组织学等级之间的扩充比较。(A) 组织学等级2 (H2) 和组织学等级 1 (H1) 之间, 组织学等级3 (H3) 和组织学等级1 (H1) 之间以及组织学等级3 (H3) 和组织学等级2 (H2) 之间的log<sub>2</sub>-倍数变化。顶部49个说明的分析物是鉴定为在三个等级之间差异表达的蛋白质特征。因此, 计算和显示了所有比较。较下面的47个分析物源自两个等级之间的SVM-计算, 同时不考虑第三个等级。对所有三个比较进行了所述计算, 并且因此汇编了显著性分析物的列表。使用Matrix2png (Pavlidis和Noble, 2003) 产生了矩阵色图。(B) 源自两组比较的SVM计算的ROC AUC值。列出了来自未过滤 (完整数据集) 以及过滤数据 (方差0.2和p-值<0.01) 的ROC AUC值。(C) 组织学等级1和3的热图 (在方差0.2, p-值<0.01, q-值<0.25上过滤的数据)。差异表达的分析物显示于热图中, 其中红色说明了上调, 而绿色说明了下调。(D) 使用在等级1和等级3之间差异表达的50个蛋白质的组织学等级1、等级2和等级3的PCA-图 (图12C)。

[0175] 图13. 针对在ER-状况比较或HER2/neu-状况比较中鉴定为差异表达的蛋白质子集举例说明的单独强度箱线图。(A) ER-阳性和ER-阴性肿瘤之间的差异表达的分析物。(B) HER2-阴性和HER2-阳性肿瘤之间的差异表达的分析物。

[0176] 图14. Ki67-阳性 (25% 截留) 和Ki67-阴性分级肿瘤的评价。差异表达的分析物显示于热图中, 其中红色说明上调, 而绿色说明下调。(A) Ki67-阳性和Ki67-阴性分级肿瘤的PCA-图。相应分析物和样品的热图。(在方差0.2, p-值<0.01, q-值<0.27上过滤的数据)。(B) 用ROC-曲线说明了来自使用SVM 的弃-交叉验证方法的结果。

[0177] 图15. 针对反映出组织学等级或ER-状况的差异表达分析物, 使用IPA 的转录因子关联网络分析。连接分子的线表示分子相关性, 而箭头的样式表示特定的分子相关性和相互作用的方向性。(A) 将多组组织学等级比较中鉴定的49个蛋白质用作输入。为了色码测量的分析物, 使用了针对组织学等级3/组织学等级1的中值的log<sub>2</sub>比例。红色说明上调。绿色说明下调。(B) 将ER-状况比较中鉴定的39个蛋白质用作输入。为了色码测量的分析物, 使用了中值的log<sub>2</sub>比例。在ER-阴性样品中, 红色说明上调, 而绿色说明下调。

[0178] 图16. 发现基于来自针对分析物子集举例说明的1411个组织学等级肿瘤样品的数据的单独mRNA表达谱展示了组织学等级之间显著差异的蛋白质表达。(A-E) 发现五个蛋白质的mRNA表达水平在组织学等级3肿瘤中呈现出提高的表达。(F-J) 发现五个蛋白质的mRNA表达水平在组织学等级3 肿瘤中呈现出降低的表达。

[0179] 图17. 基于来自1620个ER-状况限定的乳腺肿瘤样品的数据的mRNA 表达谱。使用gene entrez ID, 将39个差异表达的蛋白质中的32个成功作图于GOBO-数据库中。(A) 发现10个蛋白质的mRNA表达谱在ER-阳性肿瘤中呈现出提高的蛋白质表达。此外, 说明了10个基因与不同基因群组模块表达模式的关联。灰色点表示实际的关联值。(B) 发现蛋白质的mRNA表达谱在ER-阳性肿瘤中呈现出降低的表达。此外, 可以看到22个基因与不同基因群组模块表达模式的关联。灰色点表示实际的关联值。(C-D) 发现针对两个蛋白质举例说明的单独mRNA表达谱在ER-阳性肿瘤中呈现出提高的表达。(E-F) 发现针对两个蛋白质举例说明的单

独mRNA表达谱在ER-阳性肿瘤中呈现出降低的表达。

[0180] 图18.使用GOBO-数据库工具,对单独的mRNA表达谱作图,针对三个分析物举例说明,发现了在HER2/neu比较中呈现出显著差异的蛋白质表达。数据基于1881个可用的肿瘤样品。(A) HER2/neu, (B) S100A9, (C) GRB7。除了以上三个分析物,测试了第四个蛋白(登录号P22392),以进行作图。然而,由于使用Gene Entrez ID 4831或654364,来自GOBO-数据库工具的错误信息,因此丢失了该数据。

[0181] 图19.卡普兰-迈耶分析,使用DMFS作为10-年终点。使用Entrez Gene ID(转换swissprot ID后),将区分了组织学等级的49个蛋白质中的42个成功作图于基因表达数据库。将分析物分成两组(基于上调或下调,使用针对观察的蛋白质表达水平的组织学等级3和等级1之间的比例),形成15个下调分析物和27个上调分析物。然后使用基因表达数据库,将这两个组用于评价无远端转移的生存(DMFS)的潜在风险。通过将基因表达数据分层成三个分位数,使用DMFS作为用于组织学等级肿瘤的10-年终点( $n=1379$ ),来进行卡普兰-迈耶分析。此外,使用DMFS作为10-年终点,基于单个基因,产生了四个单独的卡普兰-迈耶分析(2个下调和两个上调),并且使用GOBO-工具,以相似的方式呈现出来。

[0182] 图20.乳癌组织样品选自52个样品的相同初始群,在此包括6个等级1样品,9个等级2样品和6个等级9样品。将样品在溶液中消化(胰蛋白酶化)并且使用选择性监控反应(SRM)设定分析(一种确定的基于质谱的方法)。靶向对应于来自生物标记物列表的8个蛋白质的9个肽并且定量。样品重复三份运行。使用Anubis分析了数据,接着进行P-值过滤( $p<0.01$ )和q-值过滤( $q<0.11$ )。使用截短的标记物列表,数据显示出乳癌组织样品可以根据等级区分。

[0183] 图21.乳癌组织样品选自52个样品的相同初始群,在此包括分布在等级1、2和3中的47个样品(具有技术重复性)。将样品在凝胶中消化(胰蛋白酶化),并且使用选择性监控反应(SRM)设定分析(一种确定的基于质谱的方法)。靶向对应于来自生物标记物列表的4个蛋白质的8个肽并且定量。样品重复三份运行。使用Anubis分析了数据,接着进行P-值过滤( $p<0.01$ )和q-值过滤( $q<0.009$ )。使用截短的标记物列表,数据显示出乳癌组织样品可以根据等级区分。

## 实施例

### [0184] 介绍

[0185] 使用目前的临床和实验室参数,乳癌患者中的肿瘤进展和预后是难以评价的,并且不存在候选多重组织生物标记物特征。在解决这一临床未满足需求的尝试中,发明人应用了最近研发的蛋白质组发现工具,意味着整体蛋白质组检测。因此,通过结合亲和力蛋白质组学,只基于9个抗体,和无标记 LC-MS/MS,发明人剖析了52个乳癌组织样品,代表了最大的乳癌组织蛋白质组研究之一,并且成功地产生了表示1388个蛋白质的详细定量蛋白质组图谱。结果表明发明人已经深入地破译了组织学等级的乳癌肿瘤的分子谱,其反映出肿瘤进展。更详细地,定义了49-丛(plex)组织生物标记物特征(其中 $p<0.01$ )和79-丛组织生物标记物(其中 $p<0.02$ )特征,以高准确性区分组织学等级1至3乳癌肿瘤。鉴定了生物学上高度相关的蛋白质,并且差异表达的蛋白质支持目前关于肿瘤微环境对肿瘤进展重塑的假设。此外,还证明了使用标记物来评估无远端转移的生存的风险。此外,分别描绘了反映出

ER-、HER2-和Ki67-状况的乳癌相关生物标记物特征。分别使用独立方法 (mRNA谱测定) 和患者群,确认了生物标记物特征。总之,这些分子谱提供了提高的乳癌分类和预后。

#### [0186] 实验程序

#### [0187] 临床样品

[0188] 该研究通过Lund,瑞典的区域伦理审查委员会 (regional ethics review board) 的批准。从Department of Oncology (SUS,Lund) 招募五十二名乳癌患者。获得了50个组织样品的全部临床记录。基于组织学等级1 (n=9)、等级 2 (n=17) 和等级3 (n=24),将样品进行细分。

#### [0189] 胰蛋白酶消化的人乳癌组织样品的制备

[0190] 从52个乳癌组织块提取蛋白质并且随后还原、烷基化、胰蛋白酶消化和最终储存在-80°C,直至进一步的使用。此外,通过合并来自所有消化样品的5 $\mu$ l等份试样来产生用作参照样品的集合样品,并且储存在-80°C,直至进一步的使用。在补充实验程序中提供了关于样品制备的详细内容。

#### [0191] CIMS-scFv抗体的产生和与磁珠的偶联

[0192] 在大肠杆菌培养物中产生了对抗六个短C-末端氨基酸肽基序的九个 CIMS scFv 抗体 (表S2),并且使用Ni<sup>2+</sup>-NTA亲和力和色谱纯化。接着,将纯化的抗体与磁珠偶联。在补充实验程序中提供了关于scFv生产和偶联的详细内容。

#### [0193] 无标记定量GPS实验

[0194] 通过混合等量的两种或三种不同粘合剂 (表S2),制得了四个不同集合 (称为CIMS-粘合剂混合物1至4) 的抗体-缀合的珠子。将抗体混合物暴露于胰蛋白酶样品,洗涤,并且最终用乙酸孵育,以洗脱捕获的肽。然后将洗脱液直接用于MS-分析,而没有任何其他清洗。使用26天MS-仪器时间,运行了完整的研究,分成6.5天的四个块 (一种CIMS-粘合剂混合物/块)。每个CIMS-粘合剂混合物单独分析全部样品一次。此外,在每个块内进行了选定样品的重复三次捕获,作为背对背 (back-to-back) LC-MS/MS运行。在4个块内和4个块之间随着时间推移重复分析了参照样品 (图7)。总共进行了238 个LC-MS/MS运行,并且在补充实验程序中提供了关于肽捕获和相关质谱分析的所有详细内容。

#### [0195] 蛋白质鉴定和定量

[0196] 通过两个软件包,即Proteios SE (Hakkinen等,2009) 和Progenesis LC-MS (Nonlinear Dynamics,UK),分析了所产生的数据。进行了相对正向和反向组合的数据库的研究 (Homo Sapiens Swiss-Prot,2011年8月,获得了总共 71324个数据库条目),基于为了产生肽鉴定而鉴定的反向目标的数量,评估了0.01的假发现率 (FDR)。Progenesis-LC-MS软件 (v4.0) 用于比对特征、鉴定 (Mascot) 和产生定量值。补充实验程序中提供了关于研究参数和数据加工的详细内容。

#### [0197] 统计学和生物信息学分析

[0198] 使用one-way ANOVA,将Qlucore Omics Explorer v (2.2) (Qlucore AB, Lund,瑞典) 用于鉴定显著上调或下调的蛋白质 (p<0.01)。基于Benjamini和 Hochberg方法 (Benjamini和Hochberg,1995),产生了q-值。在Qlucore中产生了主成分分析 (PCA) 图和热图。支持向量仪 (SVM) 是使用弃一交叉验证程序用于将样品归类的一种学习方法 (Cortes和 Vapnik,1995),并且对未过滤的和p-值过滤的数据进行了分析。使用SVM判定值和曲线下面

积 (AUC) 构建了接收者操作特征 (ROC) 曲线 (Lasko 等, 2005), 其用作分类器性能的测量。此外, 将 Ingenuity Systems Pathway Analysis (IPA) (v 11904312, www.ingenuity.com) 用于显著差异表达的蛋白质, 以提取信息, 如蛋白质定位、潜在的网络相互作用、转录因子关联以及与肿瘤发生的关联。使用 GOBO 搜寻工具 (Ringner 等, 2011), 相对大的针对具有临床参数 (如, 组织学等级 1、2 和 3, ER-状况或 HER2-状况) 的乳腺癌组织的公开的基因表达数据的群, 在 mRNA 水平最终证实了实验上产生的蛋白质特征。

#### [0199] 补充实验程序

#### [0200] 胰蛋白酶消化的人乳腺癌组织样品的制备

[0201] 从乳腺癌组织块提取蛋白质, 并且储存在  $-80^{\circ}\text{C}$ , 直至使用。简而言之, 将组织块 (约 50mg/样品) 在 Teflon 容器中均质化, 在液氮中预冷, 其通过将钢瓶固定在振荡器中  $2 \times 30$  秒同时在两轮振荡之间在液氮中快速冷却来进行。将均质的组织粉收集在含有 8M 脲、30mM Tris、5mM 乙酸镁和 4% (w/v) CHAPS (pH 8.5) 的裂解缓冲液中 (2mg 组织/30 $\mu$ l 缓冲液)。将试管简短涡旋并且在冰上孵育 40 分钟, 每 5 分钟将样品简短涡旋。孵育后, 将样品在 13000 rpm 涡旋, 并且将上清液转移至新试管中, 接着第二次离心。使用 Zeba 脱盐旋转柱 (Pierce, Rockford, IL, USA) 将缓冲液交换成 0.15M HEPES, 0.5M 脲 (pH 8.0), 之后使用 Total Protein Kit, Micro Lowry (Sigma, St. Louis, MO, USA) 测定蛋白质浓度。最后, 将样品等分, 并且储存在  $-80^{\circ}\text{C}$ , 直至进一步使用。将蛋白质提取物解冻, 还原, 烷基化和胰蛋白酶酰化。首先, 分别加入 SDS 和 TECP-HCl (Thermo Scientific, Rockford, IL, USA) 至 0.02% (w/v) 和 5mM, 并且将样品在  $56^{\circ}\text{C}$  还原 60 分钟。将样品冷却至室温, 接着加入碘乙酰胺至 10mM, 然后在室温烷基化 30 分钟。接着, 加入 20 $\mu$ g/mg 蛋白质的测序等级改性的胰蛋白酶 (Promega, Madison, Wisconsin, USA), 在  $37^{\circ}\text{C}$  持续 16 小时。为了确保完成消化, 加入第二个等份的胰蛋白酶 (10 $\mu$ g/mg 蛋白质), 并且将试管在  $37^{\circ}\text{C}$  再孵育 3 小时。最后, 将消化的样品等分, 并且储存在  $-80^{\circ}\text{C}$ , 直至进一步使用。此外, 制备分开的集合样品, 其通过将来自所有消化样品的 5 $\mu$ l 等份试样混合产生, 并且储存在  $-80^{\circ}\text{C}$ , 直至进一步使用。为了提高潜在的实验性蛋白质组覆盖, 仍然单独分析了其有限的临床数据在如下表 S1 中的两个样品, 并且包括在集合样品中。

#### [0202] CIMS-scFv 抗体的产生和与磁珠的偶联

[0203] 从 n-CoDeR (Soderlind 等, 2000) 文库选择了九个 CIMS scFv 抗体 (相对六个短 C-末端氨基酸基序 (称为 M-1、M-15、M-31、M-32、M-33 和 M-34) 的克隆 1-B03、15-A06、17-C08、17-E02、31-001-D01、32-3A-G03、33-3C-A09、33-3D-F06 和 34-3A-D10), 并且由 BioInvent International AB, Lund, 瑞典友好地提供 (表 S2)。最近已经测定了六个 CIMS 抗体的特异性和解离常数 (低  $\mu\text{M}$  范围) (Olsson 等, 2011)。在 100ml 大肠杆菌培养物中产生抗体, 并且使用  $\text{Ni}^{2+}$ -NTA 琼脂糖上的亲和色谱 (Qiagen, Hilden, 德国) 纯化。用 250mM 咪唑洗脱结合的分子, 相对 PBS (pH 7.4) 渗析 72 小时, 然后储存在  $+4^{\circ}\text{C}$ , 直至使用。通过测量 280nm 的吸光值来确定蛋白质浓度。通过在 Agilent Bioanalyzer (Agilent, Waldbronn, 德国) 上运行 Protein 80 芯片, 证实了 scFv 抗体的完整性和纯度。按照之前所述的 (Olsson 等, 2011), 将纯化的 scFv 与磁珠 (M-270 羧酸活化的, Invitrogen Dynal, Oslo) 单独偶联。简而言之, 将一批 180-250 $\mu$ g 纯化的 scFv 与  $\sim 9\text{mg}$  (300 $\mu$ l) 磁珠共价偶联 (EDC-NHS 化学), 并且储存在  $4^{\circ}\text{C}$  在 PBS 中的 0.005% (v/v) 吐温-20 中, 直至进一步使用。此外, 产生了一批空白珠 (即, 用偶联实验方案



产生的但没有添加scFv的珠子)。

[0204] 无标记定量GPS实验

[0205] 通过混合等量的两种或三种不同粘合剂制得了四种不同的缀合珠子集合(称为CIMS-粘合剂混合物1至4),所述混合物根据以下:混合物1(CIMS-33-3D-F06和CIMS-33-3C-A09)、混合物2(CIMS-17-C08和CIMS-17-E02)、混合物3(CIMS-15-A06和CIMS-34-3A-D10)和混合物4(CIMS-1-B03、CIMS-32-3A-G03和CIMS-31-001-D01)(表S2)。对于每次捕获,使用50 $\mu$ l集合珠子溶液,并且scFv-珠子从未重复使用。用350 $\mu$ l PBS预先洗涤珠子,接着暴露于35 $\mu$ l终体积的胰蛋白酶样品消化物(用PBS稀释并加入苯甲基磺酰氟(PMSF)至1mM的终浓度),然后边温和搅拌边将珠子孵育20分钟。接着,将试管置于磁铁上,除去上清液,并且分别用100和90 $\mu$ l PBS洗涤珠子(在每个洗涤步骤之间,将珠子转移至新的试管中,并且总的洗涤时间为5分钟)。最后,用9.5 $\mu$ l 5%(v/v)乙酸溶液将珠子孵育2分钟,以洗脱捕获的肽。然后将洗脱液直接用于质谱分析,而没有任何另外的清洗。

[0206] 将与Eksigent nanoLC 2DTM加HPLC系统(Eksigent technologies, Dublin,CA, USA)连接的ESI-LTQ-Orbitrap XL质谱仪(Thermo Electron, Bremen,德国)用于所有样品。自动取样机注入6 $\mu$ l GPS-产生的洗脱液。在每个分析样品之间使用了空白LC-MS/MS运行。将肽以15 $\mu$ l/min的恒定流速加样于预备柱上(PepMap 100, C18, 5 $\mu$ m, 5mm $\times$ 0.3mm, LC Packings, Amsterdam, Netherlands)。随后将肽在10 $\mu$ m熔融石英发射器, 75 $\mu$ m $\times$ 16cm (PicoTip<sup>TM</sup> Emitter, New Objective, Inc. Woburn, MA, USA)上分离,其内部用Reprosil-Pur C18-AQ树脂(3 $\mu$ m Dr. Maisch, GmbH, 德国)装填。用含有0.1%(v/v)甲酸的水中3至35%(v/v)乙腈的35分钟线性梯度洗脱肽,使用300nl/min的流速。以数据依赖性模式操作LTQ-Orbitrap,以在Orbitrap-MS(从m/z 400至2000)和LTQ-MS/MS获取之间自动转换。每次FT-MS扫描,在线性离子捕获器中获得了四个MS/MS谱,其使用针对内部校准的锁定质量选项(m/z 445.120025),在60,000FWHM标称分辨率设定下获取。使用两个重复计数,将动态排除列表限于500个条目,其具有20秒的重复持续时间,并且具有120秒的最大停留期。前体离子电荷状态筛选能够选择具有至少两个电荷的离子以及具有不确定电荷状态的排斥离子。将标准化碰撞能力设定为35%,并且针对每个光谱获取一个微扫描。每个CIMS-粘合剂混合物单独分析全部样品一次。此外,针对每种CIMS-粘合剂混合物,进行了集合样品(基于研究中的所有样品)的重复三次捕获,并且在较长时间段内针对MS-分析分布(每种粘合剂混合物LC-MS序列运行的开始、中间和结束)(图7)。这对于CIMS-粘合剂混合物1和4是可能的。然而,对于CIMS-粘合剂-混合物2和3,序列运行中超过一半,需要替换分析LC-柱(两次),并且确定在新替换的柱上直接运行预定的最后集合运行,导致在集合运行后分析了少数样品(9个样品中分别11个)。此外,针对每种CIMS-粘合剂混合物,对样品(7267, 8613)进行了重复三次捕获。将空白珠,即,没有任何缀合抗体的珠子,暴露于集合的消化物,以评价潜在的珠子背景结合肽。基于从两次空白珠“捕获”鉴定的背景结合肽的低数量,所有产生的数据均未过滤,除非另外指出。

[0207] 蛋白质鉴定和定量

[0208] 首先使用Proteios SE分析了所产生的数据,以用于使用Mascot和X!Tandem来产生鉴定。简而言之,使用Proteios(v 2.17)平台,将所有文件进行处理并转化成mzML和mgf格式,并且将以下搜寻参数用于Mascot和X!Tandem:酶:胰蛋白酶;错失的分裂1;固定的修

饰:脲基甲基(C);可变的修饰:甲硫氨酸氧化(O)。此外,可变N-乙酰基允许在X!Tandem中进行搜寻([www.thegpm.org/tandem/](http://www.thegpm.org/tandem/))。使用3ppm的肽质量公差和0.5Da的片段质量公差,并且相对正向和反向组合的数据库进行搜寻(Homo Sapiens Swiss-Prot,2011年8月,得到总共71324个数据库条目)。使用了Mascot 和X!Tandem中的自动化数据库搜寻和随后的组合(具有0.01的假发现率(FDR))(基于鉴定的反向目标的数量来评估),以用于产生肽鉴定。当使用 Proteios SE针对每个样品产生肽鉴定时,应用了蛋白质水平上的0.01的FDR。所有原始数据储存在Proteios SE内。

[0209] 由于分析时的Proteios SE没有给予定量的无标记插入式分析模块(正在研发中),将Progenesis-LC-MS软件(v 4.0)用于产生所有定量值。简而言之,在使用Progenesis-LC-MS软件之前,使用ProteoWizard软件包,将原始数据文件转化成mzXML。使用了具有默认设置和最小输入的内置功能查找工具、Mascot搜寻工具和组合的分级工具(CIMS-粘合剂-混合物1、2、3和4)。为了最佳特征比对,将针对各个CIMS-粘合剂混合物的集合样品的第一个注射运行(图7)用作参照比对文件,除了CIMS-混合物3运行,其中将一半集合运行用作参照比对文件。包括了对于CIMS-混合物1和2的10-50分钟停留时间之间比对和检测的特征,以及对于CIMS-混合物3和4的10-49分钟停留时间之间比对和检测的特征,以用于定量。提取所产生的标准化丰度值,并且用于统计学和生物信息学分析。由于使用Progenesis软件的限制,鉴定仅限于Mascot搜寻,意味着没有包括来自Proteios SE的X!Tandem产生的鉴定以用于下游定量分析。使用了上述相同的数据库(Homo Sapiens Swiss-Prot,2011年8月,正向和反向组合数据库)和搜寻参数,并且应用了 0.01的截留FDR值。

#### [0210] 结果

[0211] 在这个研究中,使用GPS破译了52个粗制乳腺癌组织提取物的半-整体蛋白质表达谱(鉴定和定量)。描绘了反映出组织学等级的组织生物标记物特征以及其他关键临床实验室参数,如雌激素受体(ER)、HER2和Ki-67。描述了实验设计的整体工作流显示于图7中。

#### [0212] 蛋白质覆盖、动态范围和测定性能

[0213] 使用GPS,鉴定了总共2,140个蛋白质群组(图1A-C)。鉴定再现性高,形成了54.7%肽重叠(图8A)。比较起来,在整个项目中重复分析的参照样品显示出43.9%肽鉴定重叠(图8B)。在鉴定的蛋白质中,总共成功定量了1388个蛋白质(图9),并且随后用于搜寻疾病相关标记物。发现针对7267-样品定量的总中值CV值为10.8%(图2A),而参照样品的相应总中值CV值为22.8%(图2A)。特别地,对应于61个蛋白质的约38%(833个肽)的定量的肽,之前在PeptideAtlas中没有被报道过(图1D),这表明是基本上新的覆盖。通过相当大的一部分的检测肽较短的事实进一步强调了这一情况,与之前报道的那些相比,具有9对11个氨基酸的中值长(图1E)。

[0214] 评价了针对所有定量的蛋白质测量的 $\log_2$ -MS强度标准化丰度的分布,并且显示出 $\sim 10^6$ 的动态范围(图9A)。通过在由PeptideAtlas中常常报道至很少报道的范围内肽容易检测的事实进一步说明了通过GPS测定产生的深入覆盖(图1F)。然后通过主要的生物学过程,将检测的蛋白质分组,并且发现在几个组中分布(图9B)。令人感兴趣地,用过程如翻译进行分组的蛋白质(例如,60S核糖体蛋白),与预期的一样,发现其相比于例如有丝分裂中涉及的其他蛋白质(例如,CDK1)而言呈现出更高的整体丰度。总之,数据表明了GPS以可再现的方式提供新的和深入的覆盖的能力。

[0215] 反映出组织学等级的蛋白质表达谱

[0216] 首先,发明人检查了是否可以破译反映出组织学等级的组织生物标记物特征。使用多变量分析(3组比较),在等级1、等级2和等级3群组之间鉴定出49个显著( $p < 0.01$ ,  $q$ -值  $< 0.25$ )差异表达的蛋白质。基于这种特征,PCA-图显示出组织学等级1和等级3肿瘤可以很好地分离,而组织学等级2肿瘤似乎异质性更高,并且在其他两组中散布(图3A)。观察到随着组织学等级提高上调和下调分析物的模式。例如,细胞周期蛋白-依赖性激酶1(CDK1)、迷你染色体维持复合物组分3(MCM3)、DNA复制许可因子MCM7、ATP-柠檬酸合成酶(ACLY)、多腺苷酸-结合蛋白4(PABPC4)和6-磷酸果糖激酶类型C(PFKP)是上调组织标记物(图3A和图10)。相反,如keratocan(KERA)、spondin(SPON1)、asporin(ASPB)、脂肪细胞增强子-结合蛋白1(AEBP1)、胰凝乳蛋白酶(CMA1)和olfactomedin-样蛋白3(OLFML3)的分析物是下调分析物,即,在组织等级1肿瘤中呈现出较高的表达水平(图3A和图11)。

[0217] 随后发明人检查了49个 $p$ -值过滤的( $p < 0.01$ )生物标记物列表是否可以用于基于组织学等级将组织归类。为此,发明人运行了使用SVM的弃一交叉验证,并且收集所有样品的判定值。然后将预测值用于构建ROC曲线,并且计算AUC值(图3A)。结果表明组织学等级肿瘤亚组可以很好地分离(AUC=0.75-0.93),尽管等级2再次显示出异质性更高。

[0218] 接着,发明人研究了使用两组比较替代多变量方法对限定差异表达标记物的影响(图12)。如预期的,数据显示出单独组织学亚组的分类提高了,如通过AUC-值判断的(AUC=0.91-0.92)。聚焦于组织学等级1对等级3,描绘了50个显著( $p < 0.01$ )差异表达的分析物,其中31个与之前的49-生物标记物特征重叠(cfs.图3和12C)。当将组织学等级2作图于等级1对等级1的50个冷冻生物标记物比较时,再次呈现出异质特征并且在两个群中散布(cfs.图3A和12D)。

[0219] ER-状况的影响

[0220] 由于24个组织学等级3肿瘤中的14个归类为ER-阴性,并且17个阴性样品中的14个实际上是等级3肿瘤,发明人研究了ER-状况对表达谱的直接影响。为了测试这种假设,只使用ER-阳性样品( $n = 33$ )重新检查了肿瘤。采取了多变量方法,结果表明18个显著差异表达的蛋白质( $p < 0.01$ ,  $q$ -值  $< 0.51$ )是确定的,并且组织学等级1对等级3组织可以很好地归类(0.9的AUC-值,数据未显示)。特别地,发现18个分析物中的16个(例如,ASPN、SPON1、KERA、ACLY、APCS和PABPC4)与最初破译的49个生物标记物特征重叠(图3A)。因此,数据进一步支持49个生物标记物特征反映出组织学等级的观察。

[0221] 此外,发明人还检查了是否可以解释ER-相关组织生物标记物特征。结果表明ER-阳性和ER-阴性乳癌组织可以很好地归类(AUC=0.82)(图3B),并且鉴定出39个差异表达的分析物( $p < 0.01$ ,  $q$ -值  $< 0.32$ )(例如GREB1,图3B和13A)。因此,数据显示出已经检测到ER-相关组织生物标记物特征。

[0222] 反映出HER2/neu-状况和Ki67-状况的蛋白质表达谱

[0223] 当使用弃一交叉验证基于HER2/neu-状况比较52个乳癌组织提取物时,数据显示出可以区分2个群(AUC=0.98),并且鉴定出五个差异表达的标记物( $p < 0.01$ ,  $q$ -值  $< 0.9$ )(图3C)。最重要地,发现受体酪氨酸-蛋白质激酶erbB-2(HER2)是上调蛋白质(图3C和图13B)。

[0224] 此外,以相似的方式,也可以破译反映出Ki67-状况的组织蛋白特征(其中将25%

的Ki67-阳性癌核用作截留)。总体上,发现45个蛋白质是差异表达的 ( $p < 0.01$ ,  $q$ -值  $< 0.27$ ) (图14A)。数据证明了可以分离Ki67-阳性对Ki67- 阴性肿瘤 ( $AUC = 0.84$ ) (图14B)。因此,结果表明已经确定了反映出HER2/neu 状况和Ki67-状况的蛋白质表达谱。

#### [0225] 生物学相关性

[0226] 然后检查了49个区分组织学等级1至3的组织生物标记物特征的生物学相关性。为此,使用IPA软件,将每个单独的蛋白质的细胞位置作图(图 4A),并且研究了网络相关功能和潜在的关联(图4B-4C)。显示了反映出细胞位置的主要的下调蛋白质(胞外基质(ECM))和上调分析物(质膜、细胞质和核)的图。更重要地,发现顶级网络与DNA复制、重组和修复、细胞周期和自由基清除相关,而第二高的等级网络与基因表达、传染病和癌症相关。值得注意的是,第1级网络内的几种蛋白质与NF- $\kappa$ B和VEGF直接或间接相关(图 4B)。此外,第二个网络内的大部分ECM蛋白质是确定的,并且几种与转化生长因子- $\beta$  (TGF $\beta$ 1) 直接或间接相关(图4C)。因此,结果表明了已经鉴定出反映了组织学等级的生物学上高度相关的组织生物标记物。

[0227] 此外,还使用IPA评价了49个组织生物标记物特征和转录因子网络之间的相关性(图15)。值得注意的是,发现了Rb和E2F2在顶级相关的转录调节剂中(图15A)。比较起来,当组织生物标记物特征区分ER-阳性对ER- 阴性肿瘤时,发现雌激素受体2(ESR2)和孕酮受体(PGR)在顶级相关的调节剂中(图15B中)。

#### [0228] 候选乳癌进展特征的证实

[0229] 在确认区分组织学等级1至3的49个组织生物标记物特征的尝试中,将数据与公众可获得的正交乳癌mRNA谱数据集相比。确认群由1,881个样品组成,其中1,411个具有指定的组织学等级,包括等级1 ( $n = 239$ ),等级 2 ( $n = 667$ )和等级3 ( $n = 495$ )。使用基因entrez ID,49个组织生物标记物中的四十二个可以作图于基因表达数据库中,并且随后用于确认测试中。

[0230] 然后基于针对级别3对级别1观察到的下调(15个分析物)或上调(27个分析物)的蛋白质表达谱,将42个组织标记物分成两组,并且与相应的mRNA 表达谱进行比较(图5)。用mRNA表达水平充分确认了下调(例如,SPON1 和KERA)(图5A、11、16I和17J)和上调蛋白质(例如,CDK1和MCM3)(图 5B、10、16A和16B)的蛋白质表达谱。令人感兴趣地,发现了上调标记物呈现出与检查点和M-期基因模块具有高相关性的mRNA谱(图5A),而下调标记物的组呈现出与基质基因群组模块具有高相关性的mRNA谱(图5B)。

#### [0231] ER-和HER2-相关组织生物标记物特征的确认

[0232] 然后以相似的方式,使用如上相同的公众可利用的正交乳癌mRNA谱数据库,尝试确认反映出ER-状况(图3B)和HER2-状况(图3C)的组织生物标记物特征。

[0233] 在ER的情况中,确认组由具有指定的ER-状况的1,620个样品组成,包括395个ER-阴性和1225个ER-阳性样品。39个组织生物标记物中的三十二个可以作图于基因表达数据库,并且随后用于确认中。然后基于观察到的蛋白质表达谱,将32个标记物分成两组(10个上调和22个下调),并且与相应的mRNA表达谱进行比较(图17)。除了少数例外(例如,补体C3(图17F 和图13A)),用相应的mRNA表达谱充分确认了观察到的蛋白质表达谱(cfs. 图3B、13A和17)。在这一内容中,令人感兴趣地注意到发现ER-阳性肿瘤中的上调蛋白质组呈现出与类固醇应答基因模块具有高相关性的mRNA谱,而发现下调蛋白质组呈现出与免疫应答和

基底基因群组模块具有高相关性的mRNA谱(图17A-17B)。

[0234] 用于HER2的确认组由1,881个样品组成,分成HER2-阳性(n=152)、基底(n=357)、内腔-A(n=483)、内腔-B(n=289)、正常样(n=257)和未分类(n=344)。5个组织标记物中的三个可以作图于确认数据集,并且用于随后的评价中(图18)。结果表明蛋白质表达谱和基因表达谱充分相关(cfs.图3C、13B和18),进一步确认了观察结果。

[0235] 评价无远端转移的生存

[0236] 最后,发明人再次使用相同的公众可利用的基因表达数据集,检查了49个反映出组织学等级的组织生物标记物特征是否也可以用于评价无远端转移的生存(DMFS)的风险。49个组织生物标记物中的四十二个可以作图于具有10年终点生存数据的1379个样品。将标记物分成两组,反映出等级3对等级1中的下调(n=15)和上调(n=27)的标记物,然后通过基于这些分析物的表达水平将基因表达数据分级成三个分位数(低、中等和高),使用具有10-年终点的DMFS进行卡普兰-迈耶分析(图19)。特别地,数据表明了下调的分析物(主要是ECM相关的分析物)的群预测了DMFS的风险。实际上,这可以通过靶向单个下调(例如,KERA和OLFM3)或上调(例如,CDK1)的生物标记物来完成。

[0237] 讨论

[0238] 在这个研究中,发明人已经第一次深入地破译了反映出乳腺癌中肿瘤进展的多重组织生物标记物特征,下一步针对乳腺癌中的个性化用药进行研究。使用发明人最近内部研发的GPS技术(Olsson等,2012;Olsson等,2011;Wingren等,2009)来实现这一成果。因此,通过结合亲和性蛋白质组学,只基于9个抗体和无标记LC-MS/MS,发明人剖析了52个乳腺癌组织样品,其代表了最大的乳腺癌组织蛋白质组研究之一,并且成功地产生了反应出1388个蛋白质的详细定量蛋白质组图谱。

[0239] 更详细地,描绘了以高特异性和灵敏度区分组织学等级1至3乳腺癌肿瘤的最先49-从组织生物标记物特征。这一列表可以扩充至79种差异表达的标记物,将p-值标准设定至 $p < 0.02$ ,但在此将讨论聚焦于头49种分析物( $p < 0.01$ )。分子谱,或蛋白质指纹,支持了目前的观点:等级1和等级3肿瘤区别更大,而等级2肿瘤异质性更高(Sotiriou等,2006)。当剖析特征时,鉴定了之前已知的与乳腺癌相关的生物标记物以及新的候选生物标记物。从技术观点看,这种新的覆盖是由大比例的(~38%)定量的肽之前在PeptideAtlas数据库中无报道的事实而反映出来的(Deutsch等,2008)。在相对Human Protein Atlas项目(Uhlen等,2010)搜寻这49种分析物时,这种由GPS设定提供的新的覆盖也变得明显。尽管Human Protein Atlas项目目前覆盖了超过50%的非-多余人蛋白质组,没有任何抗体或任何组织学染色报道了49种差异表达蛋白质中的13种。

[0240] 已经表明单独的ES-状况影响了乳腺肿瘤中超过10%的基因的表达,并且通常认为对生存具有影响。由于ER-阴性乳腺癌通常侵略性更大,并且基于抗雌激素的治疗是无效的,因此迫切地需要其他靶向的治疗(Rochefort等,2003)。发明人鉴定了能够以适当的特异性和灵敏度区分ER-阳性和ER-阴性肿瘤的39种蛋白质特征。值得注意的是,39种标记物中的11种目前尚未被Human Protein Atlas项目覆盖,再次概述了通过GPS技术提供的新覆盖(Uhlen等,2010)。39种标记物之一,GREB1,已经表明作为对内分泌治疗应答的候选临床标记物以及潜在的治疗靶标(Hnatyszyn等,2010;Rae等,2005)。GREB1是介导雌激素刺激的细胞增殖的雌激素调节的基因,并且最近报道了在ER-阳性乳腺癌细胞和正常乳房组织

中表达,但在ER-阴性样品中不表达,这概括了其作为ER替代标记物的潜能(Hnatyszyn等,2010)。用GPS产生的蛋白质谱进一步支持这一概念(图13A)。

[0241] 此外,破译了能够区分临床上限定的HER2-阳性和HER2-阴性样品的5种蛋白质特征(图3C)。实际上,鉴定了低丰度受体酪氨酸-蛋白激酶 erbB-2(HER2-蛋白),定量,并且发现其是差异表达标记物之一。因此,将临床情况中使用GPS测量HER2的可能设想为目前使用的基于传统免疫组织化学或荧光原位杂交(FISH)的检测系统的补充。最近的研究表明五个基于HER2的测试中的一个可能产生了不正确的结果(Phillips等,2009)。此外,还发现S100-A9和生长因子受体结合的蛋白7(Grb7)在大部分HER2-阳性限定样品中呈现出提高的表达(图13B)。最近报道了高Grb7表达与高HER2-表达相关,并且用于限定具有生存降低的乳癌患者的亚群(Nadler等,2010)。S100基因家族编码低分子量钙结合蛋白,并且特异性的S100成员与癌症进展、转移相关,并且具有作为乳癌患者中抗药性预测标记物的可能(McKieman等,2011;Yang等,2011)。

[0242] 最重要地,使用GOBO工具(Ringner等,2011),使用独立的数据集和正交方法(mRNA表达水平),不仅确认了反映出组织学等级的生物标记物特征,而且确认了反映出ER-状况和HER2-状况的生物标记物特征。由于其常常是通过基因特征捕获的功能性过程,而且不是重要的单独基因,因此基于与已知基因群组模块的关联,评价了上调和下调的蛋白质组(Wirapati等,2008)。与针对基质、检查点和类固醇应答的基因群组模块的显著关联特别值得注意(图5和图17)。此外,当使用组织学衍生的蛋白质分析物评价DMFS作为终点时,数据清楚地表明更差的临床结果,特别是使用下调的ECM蛋白质时的结果。因此,独立的mRNA确认,增加了针对报道的候选生物标记物特征及其在未来乳房组织肿瘤分类中的潜能的强烈支持。

[0243] 总之,发明人已经证明了发明人最近研发的GPS技术平台用于临床蛋白质组发现谱测定的适用性。在这个研究中已经报道了反映出组织学等级的组织生物标记物特征,即,肿瘤进展,以及其他关键的临床实验室参数,如ER-、HER2-和Ki67-状况;这些新的组织生物标记物谱允许提高的乳癌的分类和预后。

[0244] 本发明涉及以下实施方式:

[0245] 1.一种用于确定乳癌相关疾病状态的方法,其包括步骤:

[0246] a) 提供待测试的样品;和

[0247] b) 通过测量测试样品中一种或多种选自表1A、表1B和/或表1C中限定的生物标记物的存在和/或含量来确定测试样品的生物标记物特征;

[0248] 其中测试样品中一种或多种选自表1A、表1B和/或表1C中限定的生物标记物的存在和/或含量表示乳癌相关的疾病状态。

[0249] 2.根据实施方式1的方法,其中乳癌相关疾病状态是组织学等级和/或无转移生存时间。

[0250] 3.根据实施方式1或2的方法,其中乳癌相关疾病状态是乳癌细胞的组织学等级。

[0251] 4.根据实施方式3的方法,其中该方法进一步包括步骤:

[0252] c) 提供一种或多种包括组织学等级1乳癌细胞、组织学等级2乳癌细胞和/或组织学等级3乳癌细胞或由组织学等级1乳癌细胞、组织学等级2乳癌细胞和/或组织学等级3乳癌细胞组成的对照样品;和

[0253] d) 通过测量对照样品中步骤(b)中测量的一种或多种生物标记物的存在和/或含量来确定对照样品的生物标记物特征;

[0254] 其中测试样品中步骤(b)中测量的一种或多种生物标记物的存在和/或含量符合以下的情况下鉴定了乳癌细胞的存在:

[0255] i) 对应于包括第一个组织学等级的乳癌细胞或由第一个组织学等级的乳癌细胞组成的对照样品中的存在和/或含量(在存在的情况下);

[0256] ii) 不同于包括第二个组织学等级的乳癌细胞或由第二个组织学等级的乳癌细胞组成的对照样品中的存在和/或含量(在存在的情况下);和/或

[0257] iii) 不同于包括第三个组织学等级的乳癌细胞或由第三个组织学等级的乳癌细胞组成的对照样品中的存在和/或含量(在存在的情况下)。

[0258] 5. 根据实施方式4的方法,其中每个对照样品包括单个组织学等级的乳癌细胞或由单个组织学等级的乳癌细胞组成。

[0259] 6. 根据实施方式4或5的方法,其中步骤(c)包括以下或由以下组成:

[0260] i) 提供一种或多种包括组织学等级1乳癌细胞或由组织学等级1乳癌细胞组成的对照样品;提供一种或多种包括组织学等级2乳癌细胞或由组织学等级2乳癌细胞组成的对照样品;和提供一种或多种包括组织学等级3乳癌细胞或由组织学等级3乳癌细胞组成的对照样品;

[0261] ii) 提供一种或多种包括组织学等级1乳癌细胞或由组织学等级1乳癌细胞组成的对照样品;和提供一种或多种包括组织学等级2乳癌细胞或由组织学等级2乳癌细胞组成的对照样品;

[0262] iii) 提供一种或多种包括组织学等级1乳癌细胞或由组织学等级1乳癌细胞组成的对照样品;和提供一种或多种包括组织学等级3乳癌细胞或由组织学等级3乳癌细胞组成的对照样品;

[0263] iv) 提供一种或多种包括组织学等级2乳癌细胞或由组织学等级2乳癌细胞组成的对照样品;和提供一种或多种包括组织学等级3乳癌细胞或由组织学等级3乳癌细胞组成的对照样品;

[0264] v) 提供一种或多种包括组织学等级1乳癌细胞或由组织学等级1乳癌细胞组成的对照样品;

[0265] vi) 提供一种或多种包括组织学等级2乳癌细胞或由组织学等级2乳癌细胞组成的对照样品;或

[0266] vii) 提供一种或多种包括组织学等级3乳癌细胞或由组织学等级3乳癌细胞组成的对照样品。

[0267] 7. 根据实施方式1或2的方法,其中乳癌相关疾病状态是个体的无转移生存时间。

[0268] 8. 根据实施方式7的方法,其中该方法进一步包括步骤:

[0269] c) 提供一种或多种包括来自具有低于10年无转移生存的个体的乳癌细胞或由来自具有低于10年无转移生存个体的乳癌细胞组成的第一个对照样品;和/或一种或多种包括来自具有10年或更长时间的无转移生存个体的乳癌细胞或由来自具有10年或更长时间的无转移生存个体的乳癌细胞组成的第二个对照样品;和

[0270] d) 通过测量对照样品中步骤(b)中测量的一种或多种生物标记物的存在和/或含

量来确定对照样品的生物标记物特征；

[0271] 其中个体的无转移生存时间在步骤 (b) 中测量的一种或多种生物标记物的存在和/或含量对应于第一个对照样品的存在和/或含量(在存在的情况下) 和/或不同于第二个对照样品的存在和/或含量(在存在的情况下)的情况下鉴定为少于10年；

[0272] 和其中个体的无转移生存时间在步骤 (b) 中测量的一种或多种生物标记物的存在和/或含量不同于第一个对照样品的存在和/或含量(在存在的情况下) 和/或对应于第二个对照样品的存在和/或含量(在存在的情况下)的情况下鉴定为超过10年。

[0273] 9. 根据实施方式8的方法,其中一种或多种第一个和/或第二个对照样品与待测试样品具有相同的组织学等级。

[0274] 10. 根据实施方式3至6任一项的方法,其中步骤 (b) 包括测量测试样品中一种或多种选自表1A中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1A中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1A中限定的生物标记物例如,至少2种选自表 1A中限定的生物标记物。

[0275] 11. 根据实施方式3至6和10任一项的方法,其中步骤 (b) 包括测量测试样品中一种或多种选自表1B中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1B中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1B中限定的生物标记物例如,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或至少30种选自表1B中限定的生物标记物。

[0276] 12. 根据实施方式3至6、10和11任一项的方法,其中步骤 (b) 包括测量测试样品中一种或多种选自表1C中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1C中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1C中限定的生物标记物例如,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或至少28种选自表1C中限定的生物标记物。

[0277] 13. 根据实施方式3至6和10至12任一项的方法,其中步骤 (b) 包括测量测试样品中一种或多种选自表1D中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1D中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1D中限定的生物标记物例如,至少2、3、4、5、6、7、8、9或至少10种选自表1D中限定的生物标记物。

[0278] 14. 根据实施方式3至6和10和13任一项的方法,其中步骤 (b) 包括测量测试样品中一种或多种选自表1E中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1E中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1E中限定的生物标记物例如,至少2、3、4、5、6、7、8或至少9种选自表1E中限定的生物标记物。

[0279] 15. 根据实施方式3至6和10至14任一项的方法,其中步骤 (b) 包括测量测试样品中表1中限定的全部生物标记物的存在和/或含量或由测量测试样品中表1中限定的全部生物标记物的存在和/或含量组成。

[0280] 16. 根据实施方式7至9任一项的方法,其中步骤 (b) 包括测量测试样品中一种或多种选自表1A中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1A中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1A中限定的生物标记物例如,至少2种选自表 1A中限定的生物标记物。

[0281] 17. 根据实施方式7至9和16任一项的方法,其中步骤 (b) 包括测量测试样品中一种



或多种选自表1B中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1B中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1B中限定的生物标记物例如,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或至少30种选自表1B中限定的生物标记物。

[0282] 18. 根据实施方式7至9、16和17任一项的方法,其中步骤(b)包括测量测试样品中一种或多种选自表1D中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1D中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1D中限定的生物标记物例如,至少2、3、4、5、6、7、8、9或至少10种选自表1D中限定的生物标记物。

[0283] 19. 根据实施方式7至9、16至18任一项的方法,其中步骤(b)包括测量测试样品中表1A、表1B和表1D中限定的全部生物标记物的存在和/或含量或由测量测试样品中表1A、表1B和表1D中限定的全部生物标记物的存在和/或含量组成。

[0284] 20. 根据实施方式7至9、16至19任一项的方法,其中步骤(b)包括测量测试样品中一种或多种选自表1C中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1C中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1C中限定的生物标记物例如,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或至少28种选自表1C中限定的生物标记物。

[0285] 21. 根据实施方式7至9、16至20任一项的方法,其中步骤(b)包括测量测试样品中一种或多种选自表1E中限定的生物标记物的存在和/或含量或由测量测试样品中一种或多种选自表1E中限定的生物标记物的存在和/或含量组成,所述一种或多种选自表1E中限定的生物标记物例如,至少2、3、4、5、6、7、8或至少9种选自表1E中限定的生物标记物。

[0286] 22. 根据实施方式7至9、16至21任一项的方法,其中步骤(b)包括测量测试样品中表1C和表1E中限定的全部生物标记物的存在和/或含量或由测量测试样品中表1C和表1E中限定的全部生物标记物的存在和/或含量组成。

[0287] 23. 根据实施方式7至9、16至22任一项的方法,其中步骤(b)包括测量测试样品中表1中限定的全部生物标记物的存在和/或含量或由测量测试样品中表1中限定的全部生物标记物的存在和/或含量组成。

[0288] 24. 根据前述实施方式任一项的方法,其中步骤(b)包括测量编码一种或多种生物标记物的核酸分子的表达。

[0289] 25. 根据实施方式24的方法,其中核酸分子是cDNA分子或mRNA分子。

[0290] 26. 根据实施方式25的方法,其中核酸分子是mRNA分子。

[0291] 27. 根据实施方式25或26的方法,其中使用选自Southern杂交、Northern杂交、聚合酶链式反应(PCR)、逆转录酶PCR(RT-PCR)、定量实时PCR(qRT-PCR)、纳米阵列、微阵列、大阵列、放射自显影术和原位杂交的方法来进行步骤(b)中测量一种或多种生物标记物的表达。

[0292] 28. 根据实施方式25至27任一项的方法,其中使用DNA微阵列确定步骤(b)中一种或多种生物标记物表达的测量。

[0293] 29. 根据前述任一项实施方式的方法,其中使用一种或多种结合部分进行步骤(b)中测量一种或多种生物标记物的表达,每个所述结合部分能够选择性地结合编码表1中鉴定的生物标记物之一的核酸分子。

- [0294] 30. 根据实施方式29的方法,其中一种或多种结合部分各自包括核酸分子或由核酸分子组成。
- [0295] 31. 根据实施方式30的方法,其中一种或多种结合部分各自包括DNA、RNA、PNA、LNA、GNA、TNA或PMO或由DNA、RNA、PNA、LNA、GNA、TNA或PMO组成。
- [0296] 32. 根据实施方式30或31的方法,其中一种或多种结合部分各自包括 DNA或由DNA组成。
- [0297] 33. 根据实施方式30至32任一项的方法,其中一种或多种结合部分是 5至100个核苷酸长。
- [0298] 34. 根据实施方式30至33任一项的方法,其中一种或多种核酸分子是 15至35个核苷酸长。
- [0299] 35. 根据实施方式30至34任一项的方法,其中结合部分包括可检测部分。
- [0300] 36. 根据实施方式1至23任一项的方法,其中步骤(b)包括测量一种或多种生物标记物的蛋白质或多肽的表达。
- [0301] 37. 根据实施方式36的方法,其中使用一种或多种结合部分进行步骤(b)中测量一种或多种生物标记物的表达,每个所述结合部分能够选择性地结合表1中鉴定的生物标记物之一。
- [0302] 38. 根据实施方式37的方法,其中一种或多种结合部分包括抗体或其抗原结合片段或由抗体或其抗原结合片段组成。
- [0303] 39. 根据实施方式38的方法,其中抗体或其片段是单克隆抗体或其片段。
- [0304] 40. 根据实施方式38或39的方法,其中抗体或其抗原结合片段选自完整抗体、Fv片段(例如,单链Fv和二硫化物键合的Fv)、Fab-样片段(例如,Fab片段、Fab'片段和F(ab)<sub>2</sub>片段)、单个可变结构域(例如,V<sub>H</sub>和V<sub>L</sub>结构域)和结构域抗体(dAb,包括单一和双重形式[即,dAb-连接物-dAb])。
- [0305] 41. 根据实施方式40的方法,其中抗体或抗原结合片段是单链Fv(scFv)。
- [0306] 42. 根据实施方式41的方法,其中一种或多种结合部分包括抗体样结合剂或由抗体样结合剂组成,所述抗体样结合剂例如亲和体或适体。
- [0307] 43. 根据实施方式37至42任一项的方法,其中一种或多种结合部分包括可检测部分。
- [0308] 44. 根据实施方式35或43的方法,其中可检测部分选自荧光部分、发光部分、化学发光部分、放射性部分和酶部分。
- [0309] 45. 根据实施方式44的方法,其中可检测部分包括放射性原子或由放射性原子组成。
- [0310] 46. 根据实施方式45的方法,其中放射性原子选自钷-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和铪-90。
- [0311] 47. 根据实施方式45的方法,其中结合部分的可检测部分是荧光部分。
- [0312] 48. 根据前述任一项实施方式的方法,其中分别在步骤(b)和/或步骤(d)之前,处理步骤(a)和/或步骤(c)中提供的样品,使得用生物素标记样品中存在的任何生物标记物,并且使用包括荧光可检测部分和抗生物素蛋白链菌素的检测剂进行步骤(b)和/或步骤(d)。

[0313] 49. 根据前述任一项实施方式的方法, 其中如通过ROC AUC值测定的方法的预测准确性为至少0.50, 例如, 至少0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95、0.96、0.97、0.98或至少0.99。

[0314] 50. 根据实施方式49的方法, 其中如通过ROC AUC值测定的方法的预测准确性为至少0.80。

[0315] 51. 根据前述任一项实施方式的方法, 其中使用阵列进行步骤(b)。

[0316] 52. 根据实施方式51的方法, 其中阵列是基于珠子的阵列。

[0317] 53. 根据实施方式51的方法, 其中阵列是基于表面的阵列。

[0318] 54. 根据实施方式51至53任一项的方法, 其中阵列选自: 大阵列; 微阵列; 纳米阵列。

[0319] 55. 一种用于根据前述任一项实施方式的方法中的阵列, 该阵列包括如实施方式29至35和37至48任一项中限定的一种或多种第一种结合剂。

[0320] 56. 根据实施方式55的阵列, 包括能够共同地结合表1A中限定的全部生物标记物的结合剂。

[0321] 57. 根据实施方式55或56的阵列, 包括能够共同地结合表1B中限定的全部生物标记物的结合剂。

[0322] 58. 根据实施方式55至57任一项的阵列, 包括能够共同地结合表1C中限定的全部生物标记物的结合剂。

[0323] 59. 根据实施方式55至58任一项的阵列, 包括能够共同地结合表1D中限定的全部生物标记物的结合剂。

[0324] 60. 根据实施方式55至59任一项的阵列, 包括能够共同地结合表1E中限定的全部生物标记物的结合剂。

[0325] 61. 根据实施方式55至60任一项的阵列, 包括能够共同地结合表1中限定的全部生物标记物的结合剂。

[0326] 62. 根据实施方式55至61任一项的阵列, 其中第一种结合剂是固定的。

[0327] 63. 一种或多种选自表1A、表1B和/或表1C中限定的生物标记物用于确定乳癌相关疾病状态的用途。

[0328] 64. 根据实施方式63的用途, 其中共同地使用表1A、表1B、表1C、表1D和表1E中限定的全部生物标记物, 用于确定乳癌相关疾病状态。

[0329] 65. 一种用于根据实施方式1至54任一项的方法中的分析试剂盒, 其包括:

[0330] C) 根据实施方式55至62任一项的阵列; 和

[0331] D) 用于进行如实施方式1至55任一项中限定的方法的说明书(任选)。

[0332] 66. 根据实施方式65的分析试剂盒, 进一步包括一种或多种对照样品。

[0333] 67. 基本上如本申请中限定的方法或用途。

[0334] 68. 基本上如本申请中限定的阵列或试剂盒。

[0335] 参考文献

[0336] Aebersold, R. 和 Mann, M. (2003). Mass spectrometry-based proteomics. Nature 422, 198-207.

[0337] Benjamini, Y. 和 Hochberg, Y. (1995). Controlling the false discovery rate:

a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57,289–300.

[0338] Bergamaschi, A., Tagliabue, E., Sorlie, T., Naume, B., Triulzi, T., Orlandi, R., Russnes, H.G., Nesland, J.M., Tammi, R., Auvinen, P. 等. (2008). Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *The Journal of pathology* 214,357–367.

[0339] Bierie, B. 和 Moses, H.L. (2006). Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. *Nature reviews Cancer* 6,506–520.

[0340] Borrebaeck, C.A. 和 Wingren, C. (2011). Recombinant antibodies for the generation of antibody arrays. *Methods Mol Biol* 785,247–262.

[0341] Bouchal, P., Roumeliotis, T., Hrstka, R., Nenutil, R., Vojtesek, B. 和 Garbis, S.D. (2009). Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography–tandem mass spectrometry (iTRAQ–2DLC–MS/MS) based quantitative proteomic analysis. *Journal of proteome research* 8,362–373.

[0342] Carlsson, A., Wingren, C., Ingvarsson, J., Ellmark, P., Baldertorp, B., Ferno, M., Olsson, H. 和 Borrebaeck, C.A. (2008). Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur J Cancer* 44,472–480.

[0343] Carlsson, A., Wingren, C., Kristensson, M., Rose, C., Ferno, M., Olsson, H., Jernstrom, H., Ek, S., Gustavsson, E., Ingvar, C. 等 (2011). 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* 108,14252–14257.

[0344] Ciocca, D.R. 和 Elledge, R. (2000). Molecular markers for predicting response to tamoxifen in breast cancer patients. *Endocrine* 13,1–10.

[0345] Cortes, C. 和 Vapnik, V. (1995). Support–Vector Networks. *Machine Learning* 20,273–297.

[0346] Cortez, D., Glick, G. 和 Elledge, S.J. (2004). Minichromosome maintenance proteins are direct targets of the ATM and ATR checkpoint kinases. *Proceedings of the National Academy of Sciences of the United States of America* 101,10078–10083.

[0347] Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., Delorenzi, M., Piccart, M. 和 Sotiriou, C. (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clinical cancer research: an official journal of the American Association for Cancer Research* 14,5158–5165.

[0348] Deutsch, E.W., Lam, H. 和 Aebersold, R. (2008). PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO reports* 9,429–434.

- [0349] Dowsett, M., Goldhirsch, A., Hayes, D.F., Senn, H.J., Wood, W. 和 Viale, G. (2007). International Web-based consultation on priorities for translational breast cancer research. *Breast cancer research:BCR* 9, R81.
- [0350] Elston, C.W. 和 Ellis, I.O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 19, 403-410.
- [0351] Fata, J.E., Werb, Z. 和 Bissell, M.J. (2004). Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. *Breast cancer research:BCR* 6, 1-11.
- [0352] Frierson, H.F., Jr., Wolber, R.A., Berean, K.W., Franquemont, D.W., Gaffey, M.J., Boyd, J.C. 和 Wilbur, D.C. (1995). Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *American journal of clinical pathology* 103, 195-198.
- [0353] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R. 和 Mann, M. (2010). Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature methods* 7, 383-385.
- [0354] Geiger, T., Madden, S.F., Gallagher, W.M., Cox, J. 和 Mann, M. (2012). Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer research*.
- [0355] Gong, Y., Wang, N., Wu, F., Cass, C.E., Damaraju, S., Mackey, J.R. 和 Li, L. (2008). Proteome profile of human breast cancer tissue generated by LC-ESI-MS/MS combined with sequential protein precipitation and solubilization. *Journal of proteome research* 7, 3583-3590.
- [0356] Ha, S.A., Shin, S.M., Namkoong, H., Lee, H., Cho, G.W., Hur, S.Y., Kim, T.E. 和 Kim, J.W. (2004). Cancer-associated expression of minichromosome maintenance 3 gene in several human cancers and its involvement in tumorigenesis. *Clinical cancer research: an official journal of the American Association for Cancer Research* 10, 8386-8395.
- [0357] Hakkinen, J., Vincic, G., Mansson, O., Warell, K. 和 Levander, F. (2009). The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data. *Journal of proteome research* 8, 3037-3043.
- [0358] Hanahan, D. 和 Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57-70.
- [0359] Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646-674.
- [0360] Hanash, S. (2003). Disease proteomics. *Nature* 422, 226-232.
- [0361] Hnatyszyn, H.J., Liu, M., Hilger, A., Herbert, L., Gomez-Fernandez, C.R., Jorda, M., Thomas, D., Rae, J.M., El-Ashry, D. 和 Lippman, M.E. (2010). Correlation of

GREB1 mRNA with protein expression in breast cancer: validation of a novel GREB1 monoclonal antibody. *Breast cancer research and treatment* 122, 371-380.

[0362] Hondermarck, H., Tastet, C., El Yazidi-Belkoura, I., Toillon, R.A. 和 Le Bourhis, X. (2008). Proteomics of breast cancer: the quest for markers and therapeutic targets. *Journal of proteome research* 7, 1403-1411.

[0363] Hudis, C.A. (2007). Trastuzumab—mechanism of action and use in clinical practice. *The New England journal of medicine* 357, 39-51.

[0364] Ivshina, A.V., George, J., Senko, O., Mow, B., Putti, T.C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H. 等 (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer research* 66, 10292-10301.

[0365] Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E. 和 Forman, D. (2011). Global cancer statistics. *CA: a cancer journal for clinicians* 61, 69-90.

[0366] Johnson, N., Li, Y.C., Walton, Z.E., Cheng, K.A., Li, D., Rodig, S.J., Moreau, L.A., Unitt, C., Bronson, R.T., Thomas, H.D. 等 (2011). Compromised CDK1 activity sensitizes BRCA-proficient cancers to PARP inhibition. *Nature medicine* 17, 875-882.

[0367] Kang, S., Kim, M.J., An, H., Kim, B.G., Choi, Y.P., Kang, K.S., Gao, M. Q., Park, H., Na, H.J., Kim, H.K. 等 (2010). Proteomic molecular portrait of interface zone in breast cancer. *Journal of proteome research* 9, 5638-5645.

[0368] Kuhn, U. 和 Wahle, E. (2004). Structure and function of poly(A) binding proteins. *Biochimica et biophysica acta* 1678, 67-84.

[0369] Lasko, T.A., Bhagwat, J.G., Zou, K.H. 和 Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of biomedical informatics* 38, 404-415.

[0370] Malumbres, M. 和 Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews Cancer* 9, 153-166.

[0371] Mangus, D.A., Evans, M.C. 和 Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome biology* 4, 223.

[0372] McKiernan, E., McDermott, E.W., Evoy, D., Crown, J. 和 Duffy, M.J. (2011). The role of S100 genes in breast cancer progression. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* 32, 441-450.

[0373] Moon, J.S., Kim, H.E., Koh, E., Park, S.H., Jin, W.J., Park, B.W., Park, S. W. 和 Kim, K.S. (2011). Kruppel-like factor 4 (KLF4) activates the transcription of the gene for the platelet isoform of phosphofructokinase (PFKP) in breast cancer. *The Journal of biological chemistry* 286, 23808-23816.

[0374] Nadler, Y., Gonzalez, A.M., Camp, R.L., Rimm, D.L., Kluger, H.M. 和 Kluger, Y.

(2010). Growth factor receptor-bound protein-7 (Grb7) as a prognostic marker and therapeutic target in breast cancer. *Annals of oncology: official journal of the European Society for Medical Oncology/ESMO* 21,466-473.

[0375] Olsson, N., James, P., Borrebaeck, C.A. 和 Wingren, C. (2012). Quantitative proteomics targeting classes of motif-containing peptides using immunoaffinity-based mass spectrometry. Submitted.

[0376] Olsson, N., Wingren, C., Mattsson, M., James, P., D, O.C., Nilsson, F., Cahill, D.J. 和 Borrebaeck, C.A. (2011). Proteomic analysis and discovery using affinity proteomics and mass spectrometry. *Molecular & cellular proteomics: MCP* 10, M110 003962.

[0377] Olsson, N., Wingren, C., Mattsson, M., James, P., D, O.C., Nilsson, F., Cahill, D.J., 和 Borrebaeck, C.A. (2011). Proteomic analysis and discovery using affinity proteomics and mass spectrometry. *Molecular & cellular proteomics: MCP* 10, M110 003962.

[0378] Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T. 等 (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine* 351, 2817-2826.

[0379] Pavlidis, P. 和 Noble, W.S. (2003). Matrix2png: a utility for visualizing matrix data. *Bioinformatics* 19, 295-296.

[0380] Pei, D.S., Qian, G.W., Tian, H., Mou, J., Li, W. 和 Zheng, J.N. (2012). Analysis of human Ki-67 gene promoter and identification of the Sp1 binding sites for Ki-67 transcription. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* 33, 257-266.

[0381] Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A. 等 (2000). Molecular portraits of human breast tumours. *Nature* 406, 747-752.

[0382] Phillips, K.A., Marshall, D.A., Haas, J.S., Elkin, E.B., Liang, S.Y., Hassett, M.J., Ferrusi, I., Brock, J.E. 和 Van Bebbber, S.L. (2009). Clinical practice patterns and cost effectiveness of human epidermal growth receptor 2 testing strategies in breast cancer patients. *Cancer* 115, 5166-5174.

[0383] Place, A.E., Jin Huh, S. 和 Polyak, K. (2011). The microenvironment in breast cancer progression: biology and implications for treatment. *Breast cancer research: BCR* 13, 227.

[0384] Rae, J.M., Johnson, M.D., Scheys, J.O., Cordero, K.E., Larios, J.M. 和 Lippman, M.E. (2005). GREB1 is a critical regulator of hormone dependent breast cancer growth. *Breast cancer research and treatment* 92, 141-149.

[0385] Ringner, M., Fredlund, E., Hakkinen, J., Borg, A. 和 Staaf, J. (2011). GOBO: gene expression-based outcome for breast cancer online. *PloS one* 6, e17911.

- [0386] Robbins,P.,Pinder,S.,de Klerk,N.,Dawkins,H.,Harvey,J.,Sterrett,G.,Ellis,I.和Elston,C.(1995).Histological grading of breast carcinomas:a study of interobserver agreement.Human pathology 26,873-879.
- [0387] Rochefort,H.,Glondou,M.,Sahla,M.E.,Platet,N.和Garcia,M.(2003). How to target estrogen receptor-negative breast cancer?Endocrine-related cancer 10, 261-266.
- [0388] Slamon,D.J.,Leyland-Jones,B.,Shak,S.,Fuchs,H.,Paton,V.,Bajamonde,A.,Fleming,T.,Eiermann,W.,Wolter,J.,Pegram,M.等(2001).Use of chemotherapy plus a monoclonal antibody against HER2for metastatic breast cancer that overexpresses HER2.The New England journal of medicine 344, 783-792.
- [0389] Soderlind,E.,Strandberg,L.,Jirholt,P.,Kobayashi,N.,Alexeiva,V.,Aberg,A.M.,Nilsson,A.,Jansson,B.,Ohlin,M.,Wingren,C.等(2000).Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries.Nature biotechnology 18,852-856.
- [0390] Sorlie,T.,Perou,C.M.,Tibshirani,R.,Aas,T.,Geisler,S.,Johnsen,H.,Hastie,T.,Eisen,M.B.,van de Rijn,M.,Jeffrey,S.S.等(2001).Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.Proceedings of the National Academy of Sciences of the United States of America 98,10869-10874.
- [0391] Sotiriou,C.,Wirapati,P.,Loi,S.,Harris,A.,Fox,S.,Smeds,J.,Nordgren,H.,Farmer,P.,Praz,V.,Haibe-Kains,B.等(2006).Gene expression profiling in breast cancer:understanding the molecular basis of histologic grade to improve prognosis.Journal of the National Cancer Institute 98,262-272.
- [0392] Strande,V.,Canelle,L.,Tastet,C.,Burlet-Schiltz,O.,Monsarrat,B.和Hondermarck,H.(2009).The proteome of the human breast cancer cell line MDA-MB-231:Analysis by LTQ-Orbitrap mass spectrometry.Proteomics Clinical applications 3,41-50.
- [0393] Sutton,C.W.,Rustogi,N.,Gurkan,C.,Scally,A.,Loizidou,M.A.,Hadjisavvas,A.和Kyriacou,K.(2010).Quantitative proteomic profiling of matched normal and tumor breast tissues.Journal of proteome research 9, 3891-3902.
- [0394] Turashvili,G.,Bouchal,J.,Baumforth,K.,Wei,W.,Dziechciarkova,M.,Ehrmann,J.,Klein,J.,Fridman,E.,Skarda,J.,Srovnal,J.等(2007).Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis.BMC cancer 7,55.
- [0395] Uhlen,M.,Oksvold,P.,Fagerberg,L.,Lundberg,E.,Jonasson,K.,Forsberg,M.,Zwahlen,M.,Kampf,C.,Wester,K.,Hober,S.等(2010).Towards a knowledge-based Human Protein Atlas.Nature biotechnology 28,1248-1250.
- [0396] van de Vijver,M.J.,He,Y.D.,van't Veer,L.J.,Dai,H.,Hart,A.A.,Voskuil,D.W.,Schreiber,G.J.,Peterse,J.L.,Roberts,C.,Marton,M.J.等(2002).A gene-



expression signature as a predictor of survival in breast cancer. *The New England journal of medicine* 347,1999–2009.

[0397] van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T. 等 (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536.

[0398] Wang, T.H., Chao, A., Tsai, C.L., Chang, C.L., Chen, S.H., Lee, Y.S., Chen, J.K., Lin, Y.J., Chang, P.Y., Wang, C.J. 等 (2010). Stress-induced phosphoprotein Ias a secreted biomarker for human ovarian cancer promotes cancer cell proliferation. *Molecular & cellular proteomics: MCP* 9, 1873–1884.

[0399] Wingren, C., James, P. 和 Borrebaeck, C.A. (2009). Strategy for surveying the proteome using affinity proteomics and mass spectrometry. *Proteomics* 9, 1511–1517.

[0400] Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schutz, F. 等 (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast cancer research: BCR* 10, R65.

[0401] Yang, W.S., Moon, H.G., Kim, H.S., Choi, E.J., Yu, M.H., Noh, D.Y. 和 Lee, C. (2011). Proteomic Approach Reveals FKBP4 and S100A9 as Potential Prediction Markers of Therapeutic Response to Neoadjuvant Chemotherapy in Patients with Breast Cancer. *Journal of proteome research*.

[0402] Zhang, L., Reidy, S.P., Bogachev, O., Hall, B.K., Majdalawieh, A. 和 Ro, H. S. (2011). Lactation defect with impaired secretory activation in AEBP1-null mice. *PloS one* 6, e27795.

[0403]

表 1: 用于测定乳腺癌相关疾病状态的生物标记物

A) 核心生物标记物 (H-级和 DMSF)

Prot acc.	p-值	q-值	F-统计值	名称	H-级	DMSF	H1 vs H2	H1 vs H3	H2 vs H3	DMSF*
1 O60938	3.54E-06	0.001483776	16.58840179	KERA	是	是	下调 H2	下调 H3	下调 H3	下调 H3
2 Q9HCB6	3.27E-06	0.001483776	16.72493744	SPON1	是	是	下调 H2	下调 H3	下调 H3	下调 H3

个体比较中的上调或下调

B) 优选的生物标记物(H-级和 DMSF)

Prot acc.	p-值	q-值	F-统计值	名称	H-级	DMSF	H1 vs H2	H1 vs H3	H2 vs H3	DMSF
3 P02743	1.74E-06	0.001483776	17.82355118	APCS	是	是	下调 H2	下调 H3	下调 H3	下调 H3
4 O75348	0.000799949	0.123381185	8.331157684	ATP6V1G1	是	是	下调 H2	上调 H3	上调 H3	下调 H3
5 Q7IUM5	0.00098264	0.123381185	8.053757668	RPS27L	是	是	上调 H2	上调 H3	上调 H3	上调 H3
6 Q14195	0.001159707	0.123381185	7.832079411	DPYSL3	是	是	下调 H2	下调 H3	下调 H3	下调 H3
7 Q9BS26	0.001177863	0.123381185	7.811373711	ERP44	是	是	下调 H2	上调 H3	上调 H3	上调 H3
8 Q13905	0.001910644	0.184744553	7.173429966	RAPGEF1	是	是	下调 H2	上调 H3	上调 H3	上调 H3
9 P33396	0.002351787	0.194665471	6.903478146	ACLY	是	是	上调 H2	上调 H3	上调 H3	上调 H3
10 P23946	0.002436705	0.194665471	6.857621193	CMA1	是	是	下调 H2	下调 H3	下调 H3	下调 H3
11 P25205	0.002640618	0.194665471	6.75398016	MCM3	是	是	上调 H2	上调 H3	上调 H3	上调 H3
12 Q9UKU9	0.002787572	0.194665471	6.684337139	ANGPTL2	是	是	下调 H2	下调 H3	下调 H3	下调 H3
13 Q8IUX7	0.00329872	0.210392612	6.468857288	AEBP1	是	是	下调 H2	下调 H3	下调 H3	下调 H3
14 Q15819	0.003347536	0.210392612	6.450129509	UBE2V2	是	是	上调 H2	上调 H3	下调 H3	上调 H3
15 Q6PON0	0.003670423	0.215540903	6.333002567	MIS18BP1	是	是	上调 H2	上调 H3	上调 H3	上调 H3
16 Q9UBD9	0.003821762	0.215540903	6.281753063	CLCF1	是	是	上调 H2	上调 H3	上调 H3	上调 H3
17 P80404	0.004283228	0.220097415	6.137635708	ABAT	是	是	上调 H2	下调 H3	下调 H3	下调 H3
18 P05141	0.004800725	0.220097415	5.994134426	SLC25A5	是	是	上调 H2	上调 H3	上调 H3	上调 H3
19 P31948	0.005012118	0.220097415	5.940101147	STIP1	是	是	上调 H2	上调 H3	上调 H3	上调 H3
20 Q9NRN5	0.00549968	0.220097415	5.824034214	OLFML3	是	是	下调 H2	下调 H3	下调 H3	下调 H3
21 P09693	0.006353439	0.220097415	5.644515991	CD3G	是	是	上调 H2	上调 H3	上调 H3	上调 H3
22 P33993	0.006506666	0.220097415	5.614975929	MCM7	是	是	上调 H2	上调 H3	上调 H3	上调 H3

个体比较中的上调或下调

[0404]

23	Q02978	0.006755395	0.220097415	5.568533328	SLC25A11	是	是	下调 H2	上调 H3	上调 H3	上调 H3	上调 H3
24	O00567	0.006943766	0.220097415	5.534535408	NOP56	是	是	上调 H2	上调 H3	上调 H3	上调 H3	上调 H3
25	O43159	0.006985712	0.220097415	5.527095318	RRP8	是	是	上调 H2	上调 H3	下调 H3	上调 H3	上调 H3
26	Q9NWH9	0.007683607	0.220097415	5.409715176	SLTM	是	是	上调 H2	上调 H3	上调 H3	上调 H3	上调 H3
27	Q15631	0.007749403	0.220097415	5.399227619	TSN	是	是	上调 H2	上调 H3	上调 H3	上调 H3	上调 H3
28	Q13011	0.007879382	0.220097415	5.378779411	ECHL	是	是	下调 H2	上调 H3	上调 H3	上调 H3	上调 H3
29	P51888	0.008461666	0.229086405	5.291296959	PRELP	是	是	下调 H2	下调 H3	下调 H3	下调 H3	下调 H3
30	P49591	0.008565681	0.229086405	5.276332378	SARS	是	是	上调 H2	上调 H3	下调 H3	上调 H3	上调 H3
31	P62851	0.009544854	0.249955868	5.144096375	RPS25	是	是	上调 H2	上调 H3	上调 H3	上调 H3	上调 H3
32	Q9BSJ8	0.009871082	0.253223467	5.103161812	ESYT1	是	是	下调 H2	上调 H3	上调 H3	上调 H3	上调 H3

C) 优选的生物标记物 (H-级)

个体比较中的上调或下调

Prot. acc.	p-值	q-值	F-统计值	名称	H-级	DMSF	H1 vs H2	H1 vs H3	H2 vs H3	DMSF*	
33	Q7Z5L7	0.000518761	0.123381185	8.92324543	PODN	是	上调 H2	下调 H3	下调 H3	T.B.D.	T.B.D.
34	Q9NQG5	0.00488782	0.220097415	5.971577644	RPRD1B	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
35	Q8NHW5	0.005050767	0.220097415	5.930479527	RPLP0P6	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
36	Q6UXG3	0.005269477	0.220097415	5.877438545	CD300LG	是	下调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
37	Q9Y2Z0	0.005865416	0.220097415	5.743804455	SUGT1	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
38	A5A3E0	0.00721476	0.220097415	5.487272739	POTEF	是	上调 H2	上调 H3	下调 H3	T.B.D.	T.B.D.
39	Q15046	0.010250654	0.257701434	5.057272911	KARS	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
40	O75306	0.010613548	0.261592736	5.015027523	NDUFS2	是	上调 H2	上调 H3	下调 H3	T.B.D.	T.B.D.
41	P55795	0.01129597	0.267906306	4.939514637	HNRNPH2	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
42	O43852-2	0.01160955	0.26933089	4.906396389	CALU	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
43	P55884	0.012088996	0.26933089	4.857522011	EIF3B	是	下调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
44	Q9BWU0	0.012345209	0.26933089	4.8322258	SLC4A1AP	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
45	P46782	0.01242736	0.26933089	4.824230671	RPS5	是	上调 H2	上调 H3	下调 H3	T.B.D.	T.B.D.
46	Q6UX71	0.012772194	0.272112683	4.791261196	PLXDC2	是	上调 H2	下调 H3	下调 H3	T.B.D.	T.B.D.
47	Q6UXG2	0.01324416	0.277465145	4.747610569	KIAA1324	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
48	P22897	0.014702935	0.299546134	4.622289181	MRC1	是	上调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.
49	Q96P16	0.014796831	0.299546134	4.614672184	RPRD1A	是	下调 H2	上调 H3	上调 H3	T.B.D.	T.B.D.

[0405]

50	P34897	0.015248769	0.299546134	4.578701496	SHMT2	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.
51	P50991	0.015384493	0.299546134	4.568115711	CCT4	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.
52	Q53HC9	0.016831151	0.299546134	4.460979462	TSSC1	是	T.B.D.	下调 H2	上调 H3	上调 H3	T.B.D.
53	Q9UKT9	0.016933656	0.299546134	4.452352047	IKZF3	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.
54	Q7Z7E8	0.017060978	0.299546134	4.444847107	UBE2Q1	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.
55	O00233	0.017963635	0.305783473	4.383607388	PSMD9	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.
56	P08621	0.018244837	0.305783473	4.365183353	SNRNP70	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.
57	P11234	0.018597743	0.307596876	4.342475891	RALB	是	T.B.D.	上调 H2	上调 H3	下调 H3	T.B.D.
58	Q99798	0.018888468	0.307987276	4.324104309	ACO2	是	T.B.D.	下调 H2	上调 H3	上调 H3	T.B.D.
59	Q92614	0.019111382	0.307987276	4.310216427	MYO18A	是	T.B.D.	上调 H2	上调 H3	下调 H3	T.B.D.
60	P47897	0.019857326	0.308754485	4.264941692	QARS	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.

个体比较中的上调或下调

D) 任选的生物标记物 (H-级和 DMFS)

Prot. acc.	p-值	q-值	F-统计值	名称	H-级	DMSF	H1 vs H2	H1 vs H3	H2 vs H3	DMFS
61	Q13310	0.000177815	0.055878409	10.43467236	PABPC4	是	上调 H2	上调 H3	下调 H3	上调 H3
62	Q95969	0.000669988	0.123381185	8.572206497	SCGB1D2	是	下调 H2	下调 H3	下调 H3	下调 H3
63	Q01813	0.00080514	0.123381185	8.322399139	PFKP	是	上调 H2	上调 H3	上调 H3	上调 H3
64	P08195	0.001042488	0.123381185	7.974472523	SLC3A2	是	上调 H2	上调 H3	上调 H3	上调 H3
65	Q9BXN1	0.002542201	0.194665471	6.802918911	ASPN	是	下调 H2	下调 H3	下调 H3	下调 H3
66	P28907	0.003943867	0.215540903	6.241922855	CD38	是	上调 H2	上调 H3	上调 H3	上调 H3
67	Q9NR99	0.00573962	0.220097415	5.770796299	MXRA5	是	下调 H2	下调 H3	下调 H3	下调 H3
68	P06493	0.006362452	0.220097415	5.642757893	CDK1	是	上调 H2	上调 H3	上调 H3	上调 H3
69	O76061	0.006825774	0.220097415	5.555717945	STC2	是	下调 H2	下调 H3	下调 H3	下调 H3
70	P53634	0.007255011	0.220097415	5.480411053	CTSC	是	上调 H2	上调 H3	下调 H3	上调 H3

个体比较中的上调或下调

E) 任选的生物标记物 (H-级)

Prot. acc.	p-值	q-值	F-统计值	名称	H-级	DMSF	H1 vs H2	H1 vs H3	H2 vs H3	DMFS*
71	Q9Y2X3	0.0078144	0.220097415	5.388957977	NOP58	是	上调 H2	上调 H3	上调 H3	T.B.D.
72	P00558	0.011192822	0.267906306	4.950618267	PGK1	是	上调 H2	上调 H3	上调 H3	T.B.D.
73	Q00688	0.011943688	0.26933089	4.872117996	FKBP3	是	上调 H2	上调 H3	上调 H3	T.B.D.

[0406]

74	P21266	0.016301906	0.299546134	4.499019623	GSTM3	是	T.B.D.	下调 H2	下调 H3	下调 H3	T.B.D.
75	Q9NZT1	0.016391	0.299546134	4.492526531	CALML5	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.
76	P29590	0.016807432	0.299546134	4.462657452	PML	是	T.B.D.	上调 H2	上调 H3	上调 H3	T.B.D.
77	O75173	0.016811191	0.299546134	4.462391376	ADAMTS4	是	T.B.D.	下调 H2	下调 H3	下调 H3	T.B.D.
78	P07996	0.017157774	0.299546134	4.438120365	THBS1	是	T.B.D.	下调 H2	下调 H3	下调 H3	T.B.D.
79	P02751	0.018241382	0.305783473	4.365407944	FN1	是	T.B.D.	下调 H2	下调 H3	下调 H3	T.B.D.

\*基于中值 H1/H3 比  
T.B.D. = 待测定

[0407]

表2:用于确定乳癌相关疾病状态的生物标记物的推荐名称

Prot acc.	名称	推荐名称
Q9HCB6	SPON1	Spondin-1
O60938	KERA	Keratocan
P02743	APCS	血清淀粉状蛋白 P-组分
Q7Z5L7	PODN	Podocan
O75348	ATP6V1G1	V-型质子 ATP 酶亚基 G 1
Q7IUM5	RPS27L	40S 核糖体蛋白 S27-样
Q14195	DPYSL3	二氢嘧啶酶-相关蛋白 3
Q9BS26	ERP44	内质网驻留蛋白 44
Q13905	RAPGEF1	Rap 鸟嘌呤核苷酸交换因子 1
P53396	ACLY	ATP-柠檬酸合成酶
P23946	CMA1	胰凝乳蛋白酶
P25205	MCM3	DNA 复制许可因子 MCM3
Q9UKU9	ANGPTL2	血管生成素-相关蛋白 2
Q8IUX7	AEBP1	脂肪细胞增强子-结合蛋白 1
Q15819	UBE2V2	泛素-缀合酶 E2 变体 2
Q6P0N0	MIS18BP1	Mis18-结合蛋白 1
Q9UBD9	CLCF1	心脏营养素(Cardiotrophin)-样细胞因子 1
P80404	ABAT	4-氨基丁酸盐氨基转移酶, 线粒体
P05141	SLC25A5	ADP/ATP 转位酶 2
Q9NQG5	RPRD1B	含核前-mRNA 结构域的蛋白质 1B 的调节
P31948	STIP1	应激-诱导的-磷酸蛋白 1
Q8NHW5	RPLP0P6	60S 酸性核糖体蛋白 P0-样
Q6UXG3	CD300LG	CMRF35-样分子 9
Q9NRN5	OLFML3	Olfactomedin-样蛋白 3
Q9Y2Z0	SUGT1	SKP1 同系物的 G2 等位基因的抑制剂
P09693	CD3G	T-细胞表面糖蛋白 CD3 $\gamma$ 链
P33993	MCM7	DNA 复制许可因子 MCM7
Q02978	SLC25A11	线粒体 2-酮戊二酸/苹果酸载体蛋白
O00567	NOP56	核仁蛋白 56
O43159	RRP8	核糖体 RNA-加工蛋白 8
A5A3E0	POTEF	POTE 锚蛋白结构域家族成员 F
Q9NWH9	SLTM	SAFB-样转录调节剂
Q15631	TSN	Translin
Q13011	ECH1	$\delta(3,5)$ - $\delta(2,4)$ -二烯酰基-CoA 异构酶, 线粒体
P51888	PRELP	Prolargin
P49591	SARS	丝氨酸-tRNA 连接酶, 细胞质
P62851	RPS25	40S 核糖体蛋白 S25
Q9BSJ8	ESYT1	延伸突触蛋白-1
Q15046	KARS	赖氨酸-tRNA 连接酶
O75306	NDUFS2	NADH 脱氢酶[泛醌] 铁-硫蛋白 2, 线粒体
P55795	HNRNPH2	异源核核糖核蛋白 H2
O43852-2	CALU	钙腔蛋白(Calumenin)
P55884	EIF3B	真核生物翻译启动因子 3 亚基 B
Q9BWU0	SLC4A1AP	Kanadaptin

[0408]

[0409]

P46782	RPS5	40S 核糖体蛋白 S5
Q6UX71	PLXDC2	含 Plexin 结构域蛋白 2
Q6UXG2	KIAA1324	UPF0577 蛋白 KIAA1324
P22897	MRC1	巨噬细胞甘露糖受体 1
Q96P16	RPRD1A	含核前-mRNA 结构域的蛋白 1A 的调节
P34897	SHMT2	丝氨酸羟甲基转移酶, 线粒体
P50991	CCT4	T-复合蛋白 1 亚基 $\delta$
Q53HC9	TSSC1	蛋白 TSSC1
Q9UKT9	IKZF3	锌指蛋白 Aiolos
Q7Z7E8	UBE2Q1	泛素-缀合酶 E2 Q1
O00233	PSMD9	26S 蛋白酶体非-ATP 酶调节亚基 9
P08621	SNRNP70	U1 小核核糖核蛋白 70 kDa
P11234	RALB	Ras-相关蛋白 Ral-B
Q99798	ACO2	乌头酸水合酶, 线粒体
Q92614	MYO18A	非常规肌球蛋白-XVIIIa
P47897	QARS	谷氨酰胺-tRNA 连接酶
Q13310	PABPC4	聚腺苷酸-结合蛋白 4
O95969	SCGB1D2	乳腺珠蛋白(Secretoglobin)家族 1D 成员 2
Q01813	PFKP	C 型 6-磷酸果糖激酶
P08195	SLC3A2	4F2 细胞表面抗原重链
Q9BXN1	ASPN	Asporin
P28907	CD38	ADP-核糖基环化酶 1
Q9NR99	MXRA5	基质-重塑相关蛋白 5
P06493	CDK1	细胞周期蛋白-依赖性激酶 1
O76061	STC2	斯钙素(Stanniocalcin)-2
P53634	CTSC	二肽酰基酶 1
Q9Y2X3	NOP58	核蛋白 58
P00558	PGK1	磷酸甘油酸激酶 1
Q00688	FKBP3	肽酰-脯氨酸-反异构酶 FKBP3
P21266	GSTM3	谷胱甘肽 S-转移酶 Mu 3
Q9NZT1	CALML5	钙调蛋白-样蛋白 5
P29590	PML	蛋白 PML
O75173	ADAMTS4	具有凝血栓蛋白基序 4 的解聚素和金属蛋白酶
P07996	THBS1	凝血栓蛋白-1
P02751	FN1	纤连蛋白

[0410] 表3: 示例性生物标记物组合的ROC AUC值

	ROC AUC 值		
	H1 vs H3	H2 vs H3	H1 vs H2
生物标记物特征			
核心-1	0.94	0.82	0.69
核心-2	0.82	0.82	0.56
核心-1+核心-2	0.94	0.83	0.50
核心-1+核心-2+(标记物 3-12)	0.90	0.85	0.69
[0411] 核心-1+核心-2+(标记物 3-22)	0.88	0.76	0.88
核心-1+核心-2+(标记物 3-32)	0.81	0.73	0.94
核心-1+核心-2+(标记物 3-42)	0.81	0.74	0.87
核心-1+核心-2+(标记物 3-52)	0.80	0.77	0.76
核心-1+核心-2+(标记物 3-60)	0.80	0.77	0.76
核心-1+核心-2+(标记物 3-62)	0.93	0.86	0.76
核心-1+核心-2+(标记物 3-72)	0.94	0.83	0.82
核心-1+核心-2+(标记物 3-79)	0.94	0.79	0.86

[0412] 表4:组织学等级SVM脚本

[0413]

```

filnamn<-"Input.txt"
# 1.1 Change FILNAME to datafile -----
# Läser in och logaritmerar datan
rawfile <- read.delim(filnamn)
samplenames <- as.character(rawfile[,1])
Diagnosis <- rawfile[,2]
Morphology<- rawfile[,3]
Treatment<-rawfile[,4]
data <- t(rawfile[,-c(1:4)])
ProteinNames <- read.delim(filnamn,header=FALSE)
ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
ProteinNames <- ProteinNames[-(1:4)]
rownames(data) <- ProteinNames
colnames(data) <- samplenames
logdata <- log(data)/log(3)

# Tar reda på vilka gruppjämförelser som ska göras
PairWiseGroups <-
as.matrix(read.delim("Comparisons_to_do.txt",header=FALSE)) # 1.2 Change
filename and use criteria file -----

# Definierar Wilcoxon-testet
wilcoxtest <- function(prot,subset1,subset2){
  res <- wilcox.test(prot[subset1],prot[subset2])
  res$p.value
}

# Definierar foldchange
foldchange <- function(prot,subset1,subset2){
  2^(mean(prot[subset1]) - mean(prot[subset2]))
}

# Definierar q-värdesberäkningen
BenjaminiHochberg <- function(pvalues){
  # This function takes a vector of p-values as input and outputs
  # their q-values. No reordering of the values is performed
  NAindices <- is.na(pvalues)
  Aindices <- !NAindices
  Apvalues <- pvalues[Aindices]
  N <- length(Apvalues)
  orderedindices <- order(Apvalues)
  OrdValues <- Apvalues[orderedindices]
}

```



[0414]

```

CorrectedValues <- OrdValues * N / (1:N)
MinValues <- CorrectedValues
for (i in 1:N){MinValues[i] <- min(CorrectedValues[i:N])}
Aqvalues <- numeric(N)
Aqvalues[orderedindices] <- MinValues
Qvalues <- pvalues
Qvalues[Aindices] <- Aqvalues
return(Qvalues)
}

# Laddar in två bibliotek
library(MASS)
library(gplots)

# Definierar färger till heatmapen
redgreen <- function(n)
{
  c(
    hsv(h=0/6, v=c( rep( seq(1,0.3,length=5) , c(13,10,8,6,4) ) , 0 ) ) ,
    hsv(h=2/6, v=c( 0 , rep( seq(0.3,1,length=5) , c(3,5,7,9,11) ) ) )
  )
}
pal <- rev(redgreen(100));

#Laddar in fler bibliotek och funktioner
library(e1071)
source("NaiveBayesian")

#Definierar SVM med Leave One Out
svmLOOvalues <- function(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="")
  res <- numeric(nsamples)
  sign <- numeric(nsamples)
  for (i in 1:nsamples){
    svmtrain <- svm(t(data[,-i]) , fac[-i] , kernel="linear" )
    pred <- predict(svmtrain , t(data[,i]) , decision.values=TRUE)
    res[i] <- as.numeric(attributes(pred)$decision.values)
    facnames <- colnames(attributes(pred)$decision.values)[1]
    if (facnames == paste(levels(fac)[1],"/",levels(fac)[2],sep="")){sign[i] <- 1}
  }
  if (facnames == paste(levels(fac)[2],"/",levels(fac)[1],sep="")){sign[i] <- -1}
  }
  if (length(unique(sign)) >1){print("error")}
  res <- sign * res
  names <- colnames(data , do.NULL=FALSE)
  orden <- order(res , decreasing=TRUE)
  Samples <- data.frame(names[orden],res[orden],fac[orden])
  ROCdata <- myROC(res,fac)
  SenSpe <- SensitivitySpecificity(res,fac)

return(list(SampleInformation=SampleInformation,ROCArea=ROCdata[1],p.value=ROCAdata[2],SenSpe <- SenSpe,samples=Samples))
}

# Definierar hur analysen ska köras om man INTE ANVÄNDER apriorianalyser
Analysera<- function(group1 ,group2){
  outputfiletxt <- paste(group1," versus ",group2,".txt" , sep="")
  outputfilepdf <- paste(group1," versus ",group2,".pdf" , sep="")
  #outputfilejpeg <- paste(group1," versus ",group2,".jpg" , sep="")
  subset1 <- is.element(Diagnosis , strsplit(group1,",")[1])
  subset2 <- is.element(Diagnosis , strsplit(group2,",")[1])
  wilcoxvalues <- apply(logdata , 1 , wilcoxtest , subset1 , subset2)
}

```

[0415]

```

foldchange <- apply(logdata , 1 , foldchange , subset1 , subset2)
QvaluesAll <- BenjaminiHochberg(wilcoxvalues)
HugeTable <- cbind(ProteinNames,foldchange,wilcoxvalues,QvaluesAll)
write.table(HugeTable, file=outputfiletxt , quote=FALSE,
sep="\t",row.names=FALSE)
color <- rep('black' , length(subset1))
color[subset1] <- 'red'
color[subset2] <- 'blue'
pdf(outputfilepdf)
#jpeg(outputfilejpeg, quality=100, width=600, height=600)
Sam <- sammon(dist(t(logdata[,subset1|subset2])), k=2)
plot(Sam$points , type="n" , xlab = NA , ylab=NA, main="All proteins" ,asp=1)
text(Sam$point , labels = colnames(logdata[,subset1|subset2]),
col=color[subset1|subset2])
heatmap.2(logdata[,subset1|subset2] , labRow = row.names(logdata),
trace="none" , labCol ="" , ColSideColors= color[subset1|subset2],col.pal ,
ns.color= "grey", key=FALSE , symkey =FALSE , tracecol = "black" , main ="" ,
dendrogram= 'both' , scale = "row" ,cexRow=0.2)
svmfac <- factor(rep('rest',ncol(logdata)),levels=c(group1,group2,'rest'))
svmfac[subset1] <- group1
svmfac[subset2] <- group2
svmResAll <- svmLCOvalues(logdata[,subset1|subset2] ,
factor(as.character(svmfac[subset1|subset2]),levels=c(group1,group2)))
ROCplot(svmResAll , sensspecnumber=4)

# N <- length(ProteinNames)
# par(mfrow=c(1,2))
# for(k in 1:N){
#   boxplot(data[k,subset1],data[k,subset2] , names=c(group1, group2),
main=c(ProteinNames[k]," test"))
# }

write("", file=outputfiletxt , append=TRUE)
write("All proteins" , file=outputfiletxt , append=TRUE)
write("", file=outputfiletxt , append=TRUE)
for (i in 1:5){write.table(svmResAll[[i]], file=outputfiletxt , append=TRUE,
sep="\t" , quote=FALSE)
write( "" , file=outputfiletxt , append=TRUE)
}
dev.off()
}

Analysera("X","Y") # 1.3 Select comparisons to do-----

```

[0416] 补充表1

样品 ID	组织学等级	年龄	肿瘤大小(mm)	ER / PgR / HER2 / ki67_gt_25	Lymph_pos	Nr of Lymph pos.
6616	1	37,62	22	+/+/-/-	是	1
6617	1	66,30	20	+/+/-/na	是	5
7149	1	74,49	31	+/+/-/-	否	0
7454	1	47,82	22	+/+/-/-	是	5
7940	1	53,94	30	+/+/-/-	否	0
8415	1	66,61	31	+/+/-/-	是	4
[0417] 9317	1	47,42	18	+/+/-/-	否	0
9795	1	43,26	15	+/+/-/-	是	2
10524	1	64,34	30	+/+/-/-	否	0
4404	2	49,92	25	+/+/-/+	是	1
5614	2	45,48	37	+/+/-/-	是	8
5096	2	37,35	6	+/+/-/+	是	8
5572	2	43,55	18	-/-/-	是	2
6096	2	36,92	12	+/+/-/-	是	1
6627	2	43,63	15	+/+/-/-	是	2
7015	2	46,77	22	+/+/-/na	是	1
7267	2	48,39	22	+/-/-/-	是	1
7296	2	46,38	14	+/+/-/na	是	4
8173	2	47,03	25	+/+/-/na	是	10
9257	2	43,78	7	+/+/-/+	否	0
9340	2	52,10	29	+/+/-/-	否	0
5402	2	44,26	50	-/+/-/-	是	5
6514	2	49,18	30	-/-/na/na	否	0
7424	2	47,98	25	+/-/-/+	是	1
8278	2	47,54	10	+/+/-/-	是	1
8504	2	49,66	25	+/+/-/-	是	1
5706	3	41,19	50	-/-/+	是	5
4239	3	40,66	33	-/-/+	否	0
5744	3	44,04	21	+/+/na/-	否	0
5811	3	49,75	45	-/-/+	是	1
5997	3	46,37	20	-/-/-/na	否	0
6009	3	49,57	20	+/+/-/-	否	0
6029	3	42,80	25	-/-/na/+	是	4
[0418] 6158	3	55,81	20	+/+/-/+	是	2
6191	3	45,04	25	-/-/-	否	0
6276	3	52,30	32	-/-/+	是	6
4723	3	48,89	40	-/-/+	是	3
5198	3	46,66	32	-/+/+/na	是	4
5203	3	33,22	30	-/-/-	是	1
5634	3	44,33	25	+/+/-/na	是	2
5996	3	50,94	22	-/-/+	是	2
6013	3	41,60	50	+/+/-/na	是	3
6176	3	50,62	35	+/+/-/+	否	0
6503	3	43,39	28	-/-/+	否	0
6877	3	34,39	27	+/+/-/+	是	8
7694	3	47,66	18	-/-/+	是	1
7722	3	46,61	27	+/+/na/-	否	0
8613	3	44,04	35	+/+/-/-	是	6
9322	3	50,33	30	-/-/+	否	0
9460	3	49,01	17	+/+/-/+	否	0
5784	na	na	na	na/na/na/na	na	na
4917	na	na	na	na/na/na/na	na	na

[0419] na = 不可获得

[0420] 补充表2

	CIMS 抗体*	选择肽	亲和性 (K <sub>D</sub> ) (μM)	混合
	CIMS-33-3C-A09	生物素-SGSGLSADHR	1.6	1
	CIMS-33-3D-F06	生物素-SGSGLSADHR	5.1	1
	CIMS-17-C08	生物素-SGSGSSAYSR	0.2	2
[0421]	CIMS-17-E02	生物素-SGSGSSAYSR	0.4	2
	CIMS-15-A06	生物素-SGSGLTEFAK	2.2	3
	CIMS-34-3A-D10	生物素-SGSGSEHLR	2.5	3
	CIMS-1-B03	生物素-SGSGEDFR	3.5	4
	CIMS-31-001-D01	生物素-SGSGLNWVGK	NA	4
	CIMS-32-3A-G03	生物素-SGSGQEASEFK	11.5	4
[0422]	*关于结合剂特征的详细内容,参见Olsson等(2011)MCP M110.003962.			

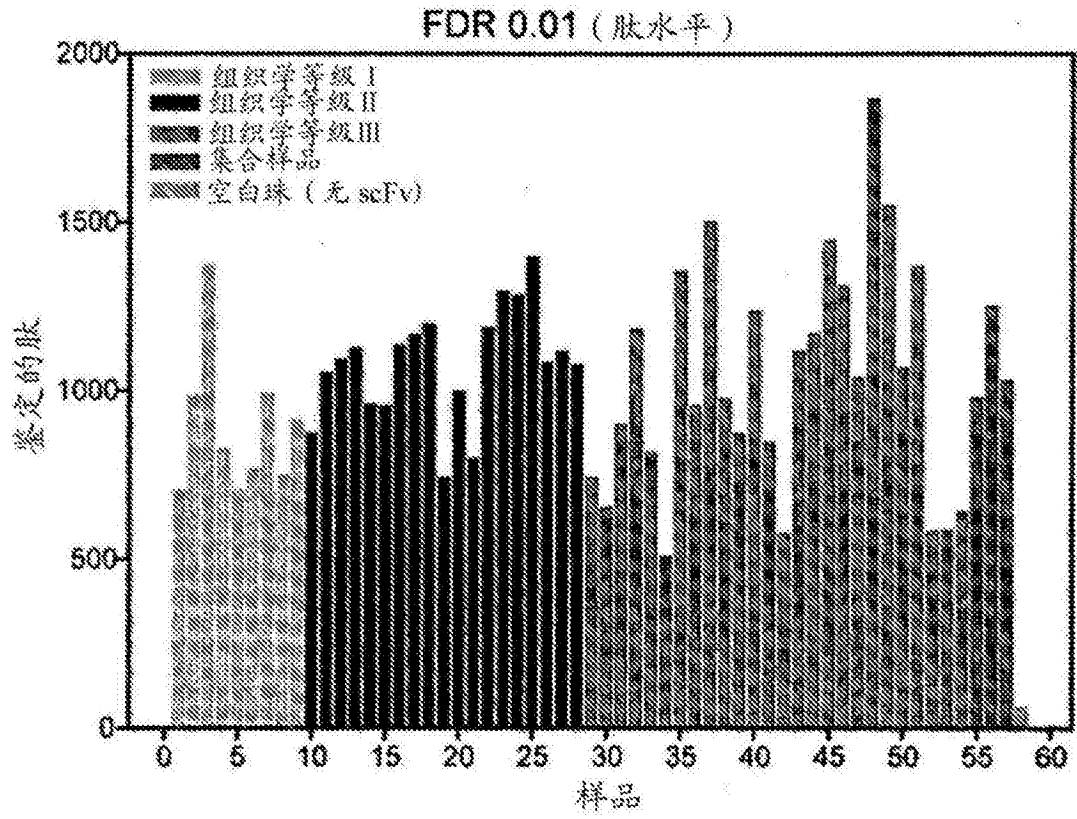


图1A

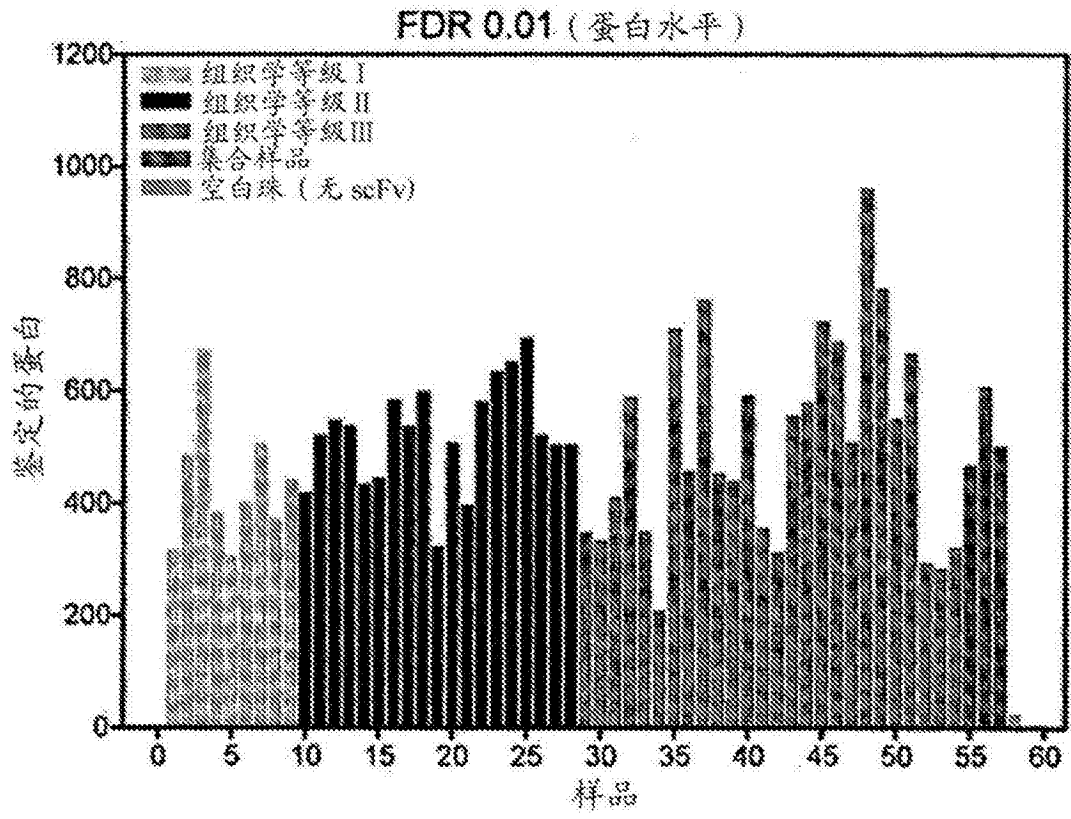


图1B

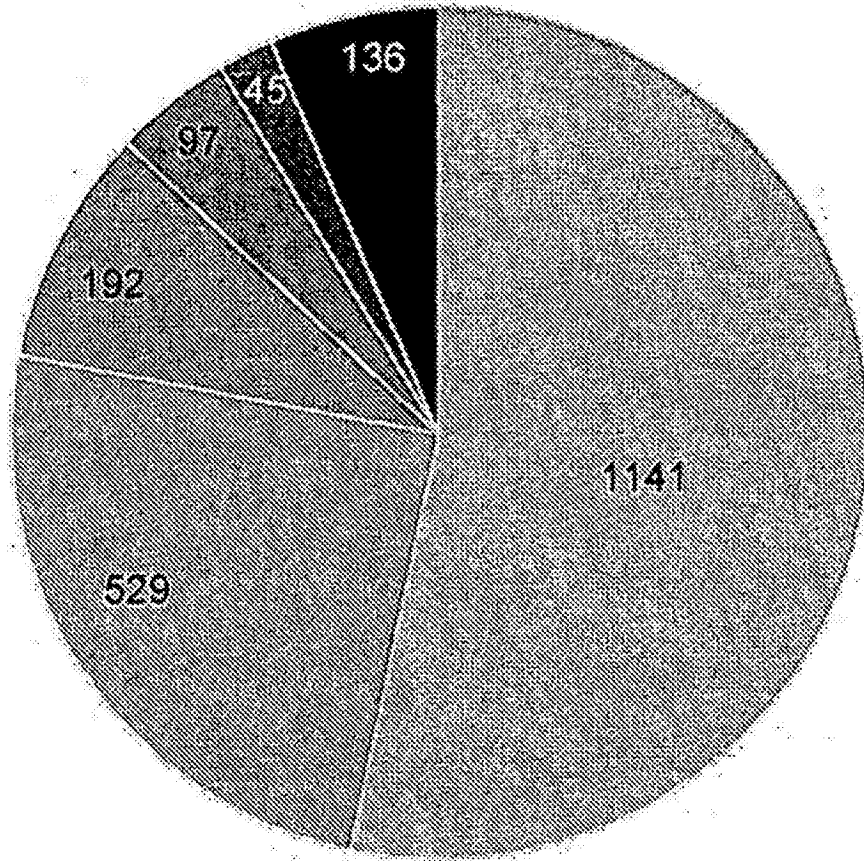


图1C

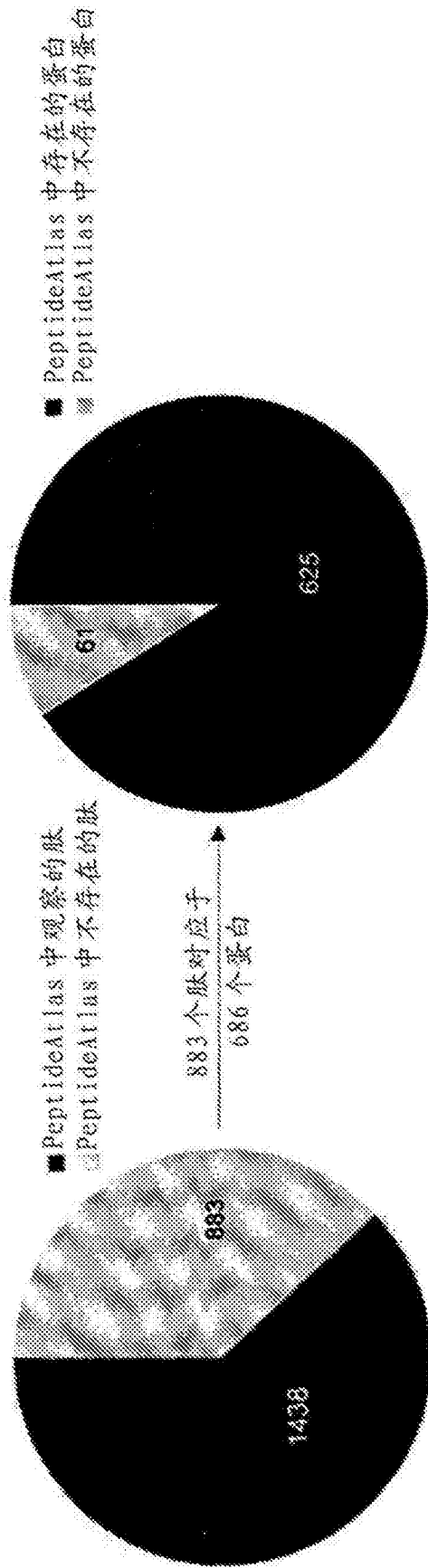


图1D



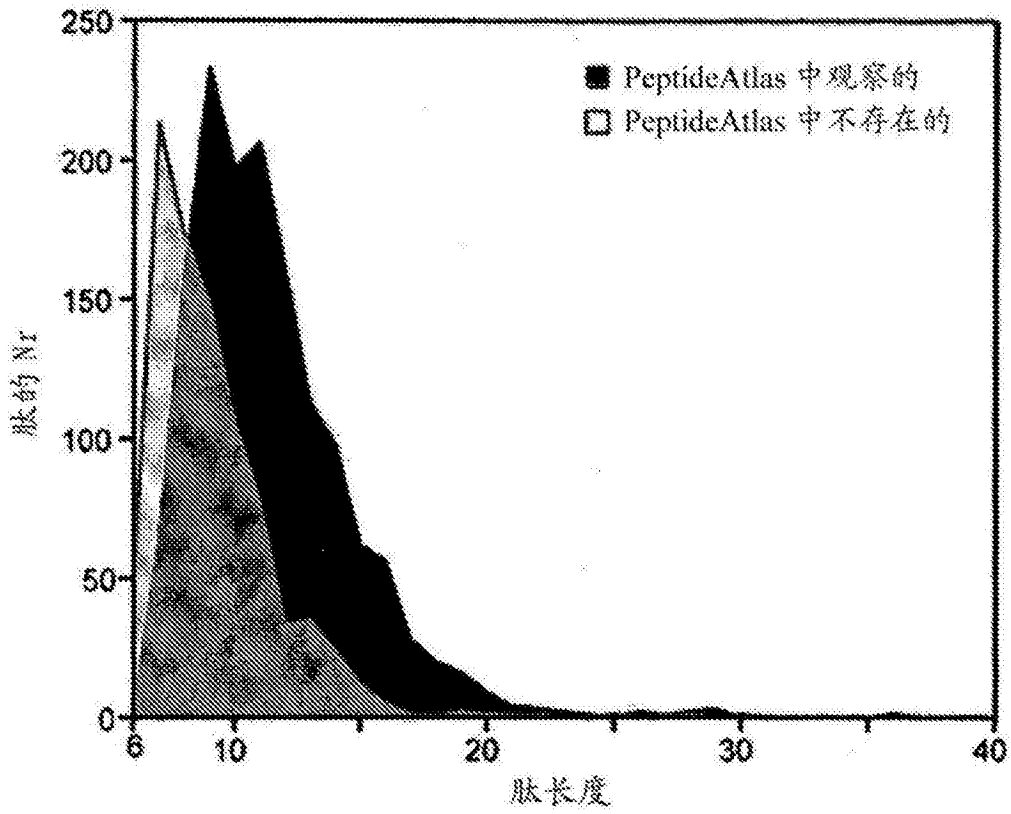


图1E

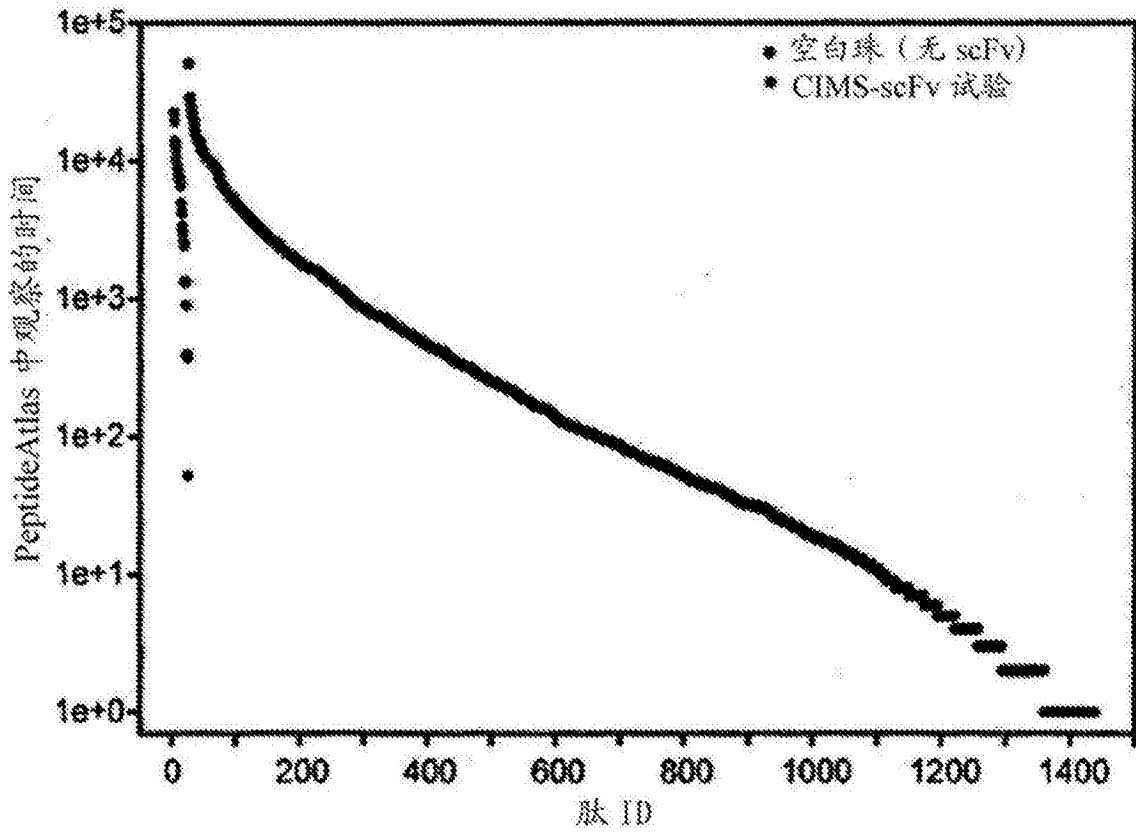


图1F

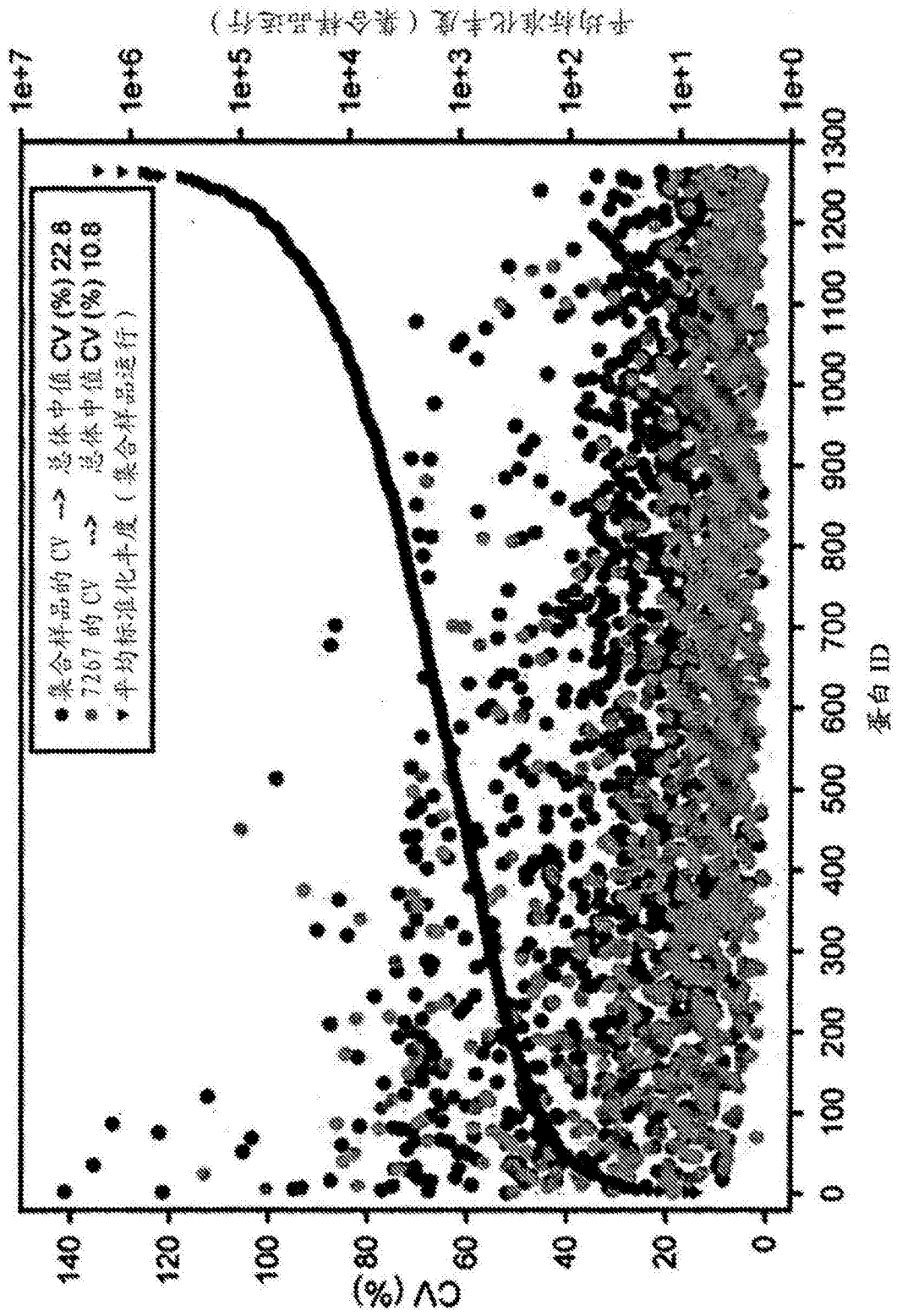


图2A

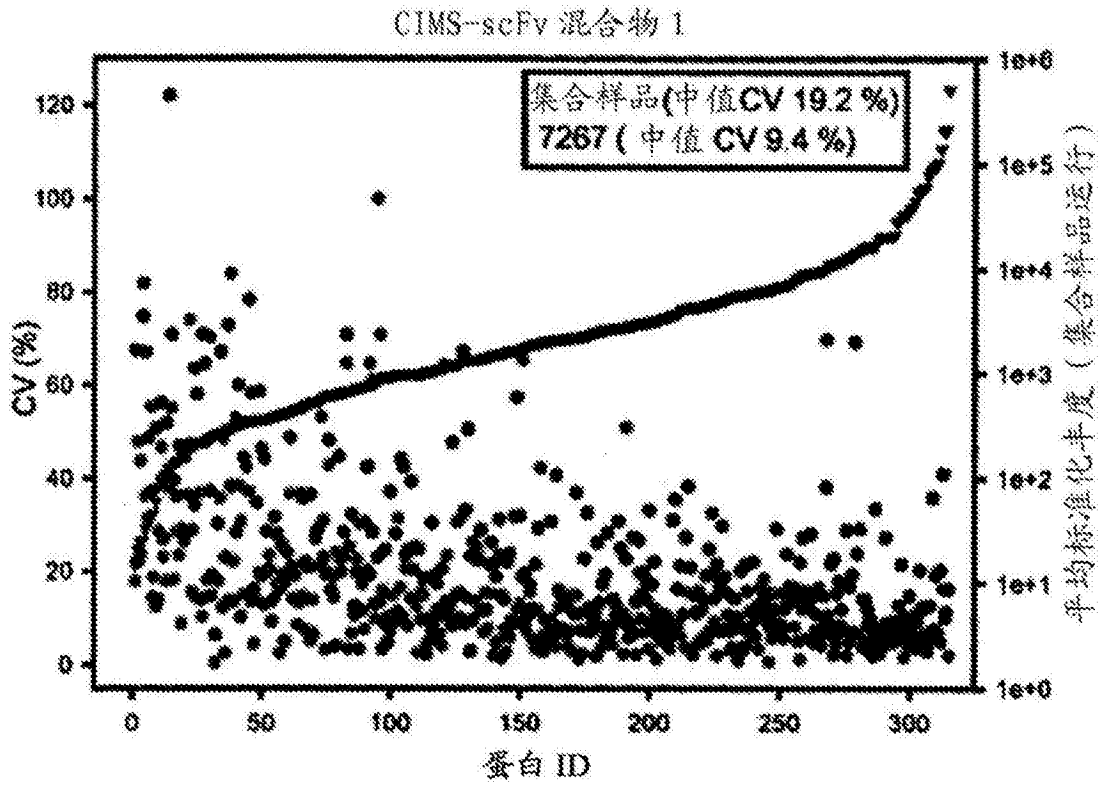


图2B

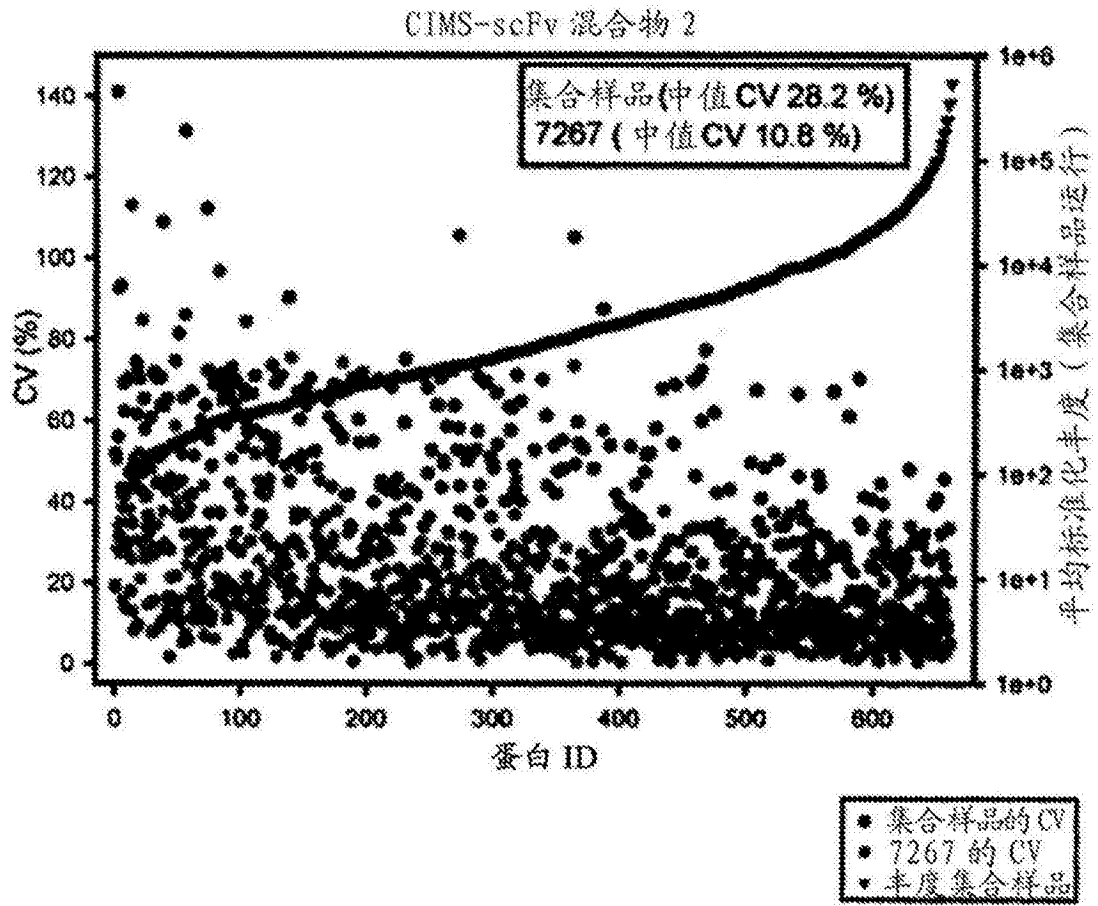


图2C

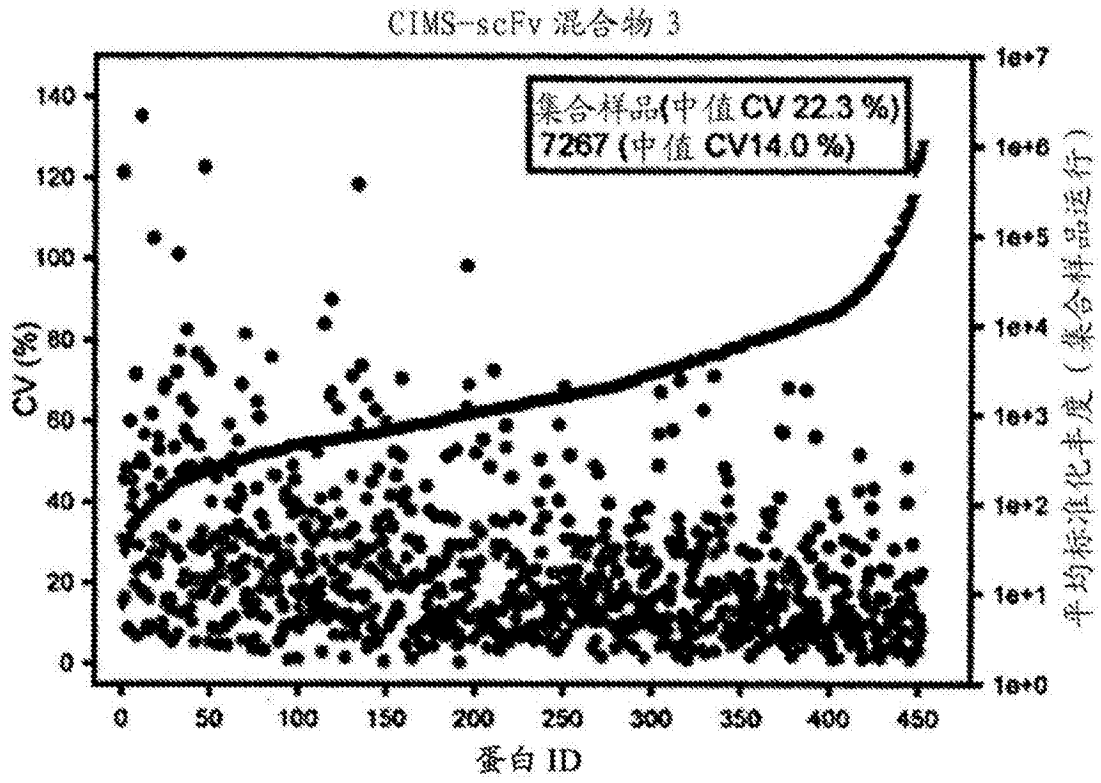


图2D

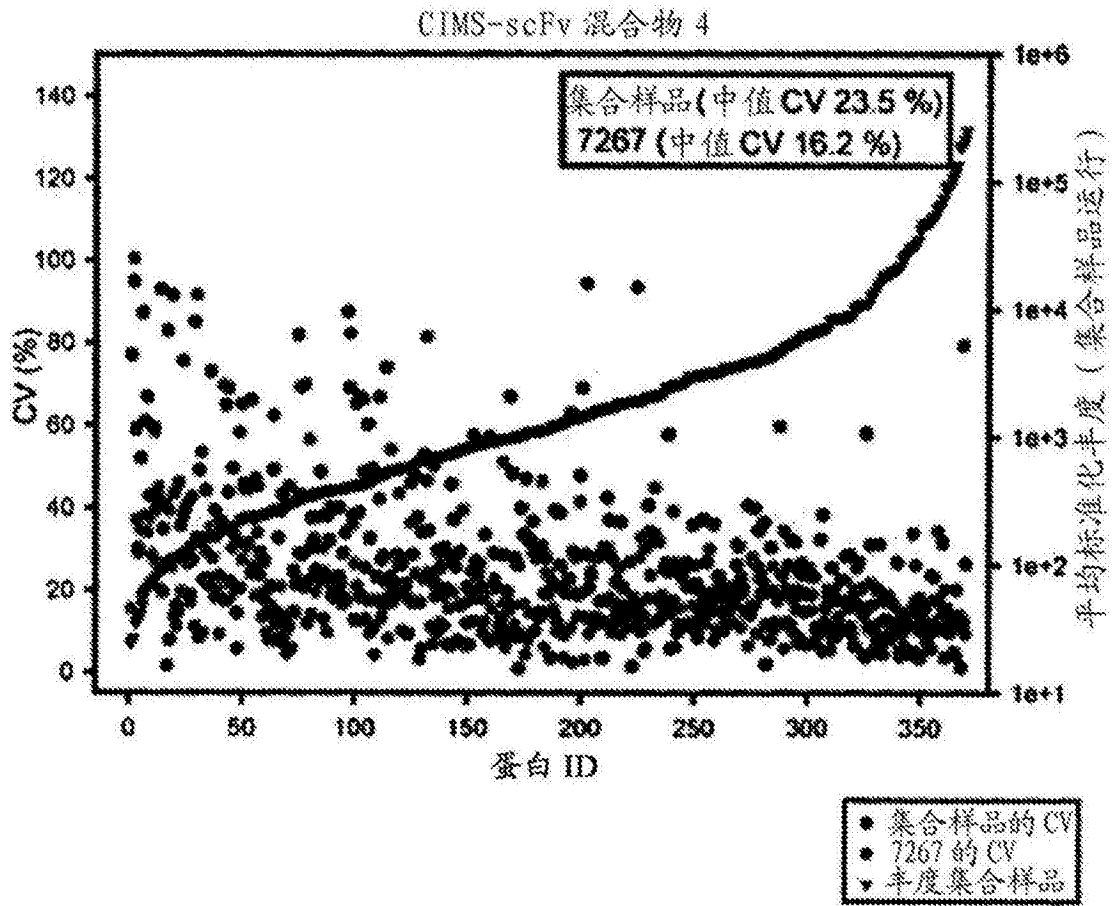


图2E

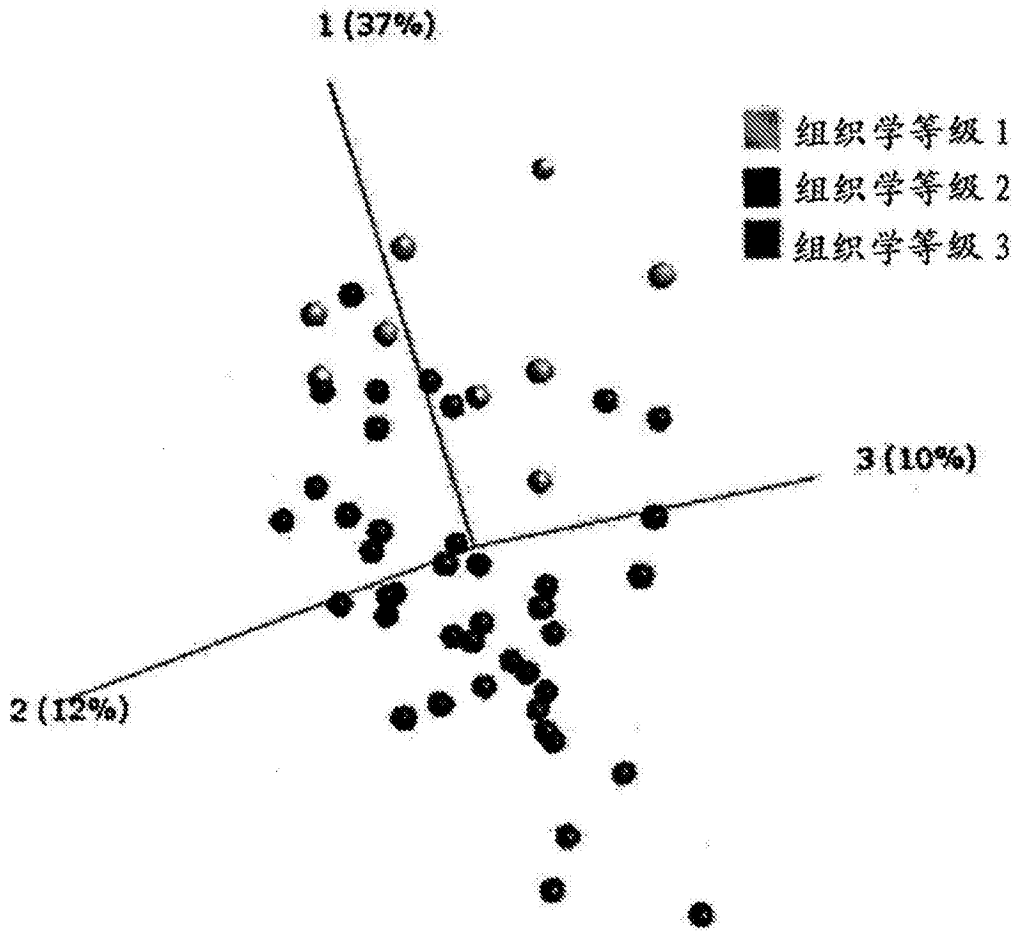
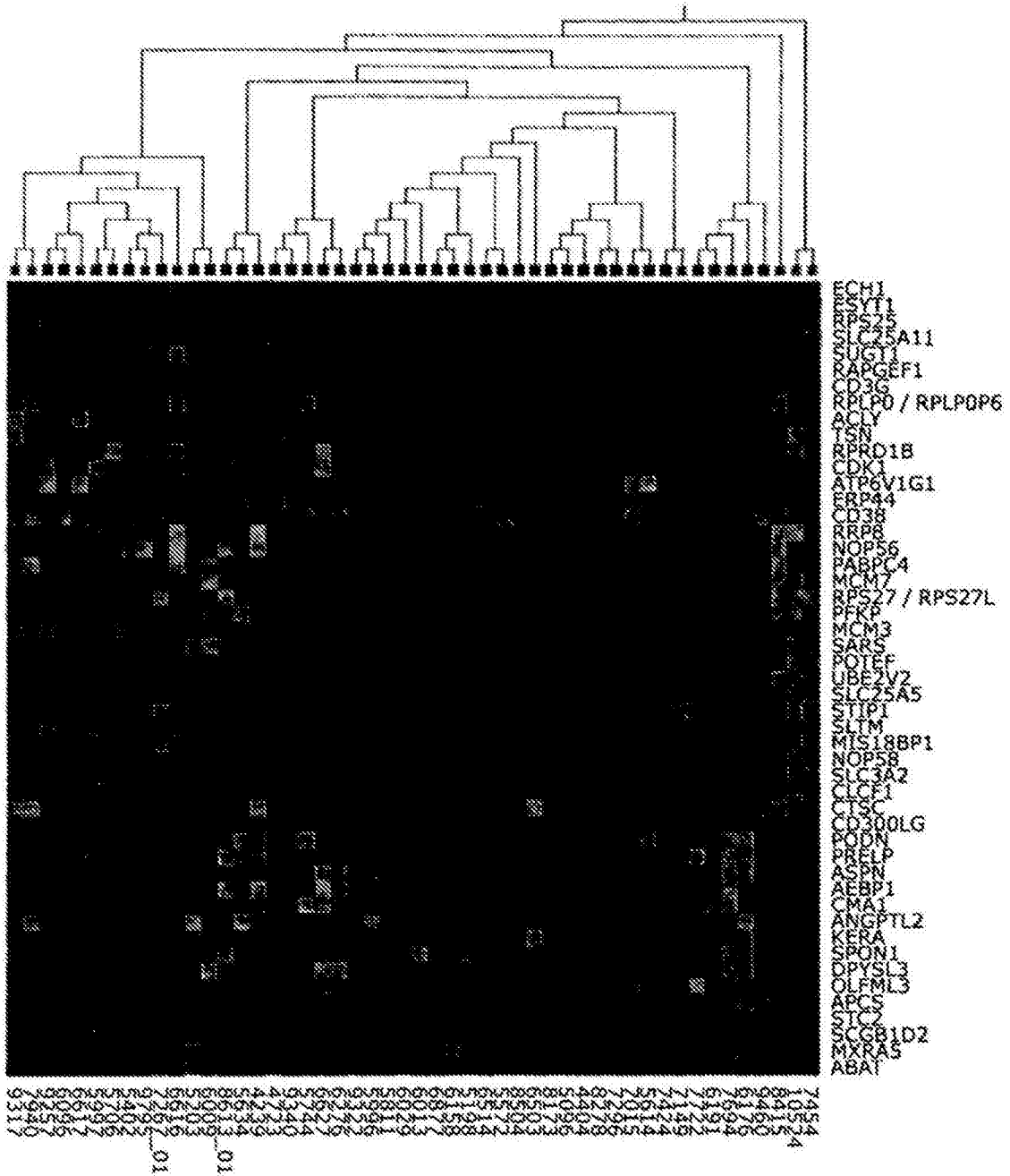


图3A





比较	ROC 面积
H1 vs. H2	0.91
H1 vs. H3	0.93
H2 vs. H3	0.75
H1, H2 vs. H3	0.77
H1 vs. H2, H3	0.88
H2 vs. H1, H3	0.7

图3A续

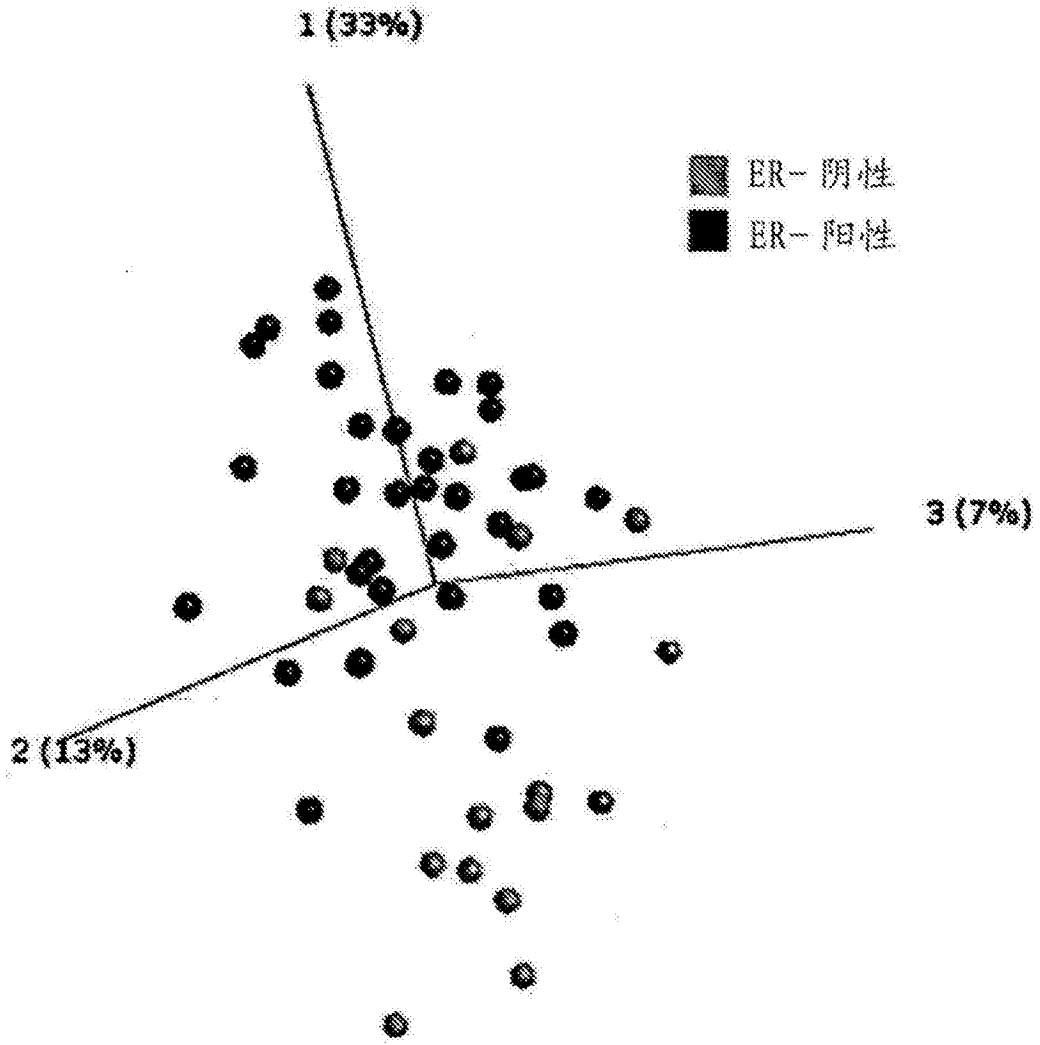


图3B



ROC 面积=0.82

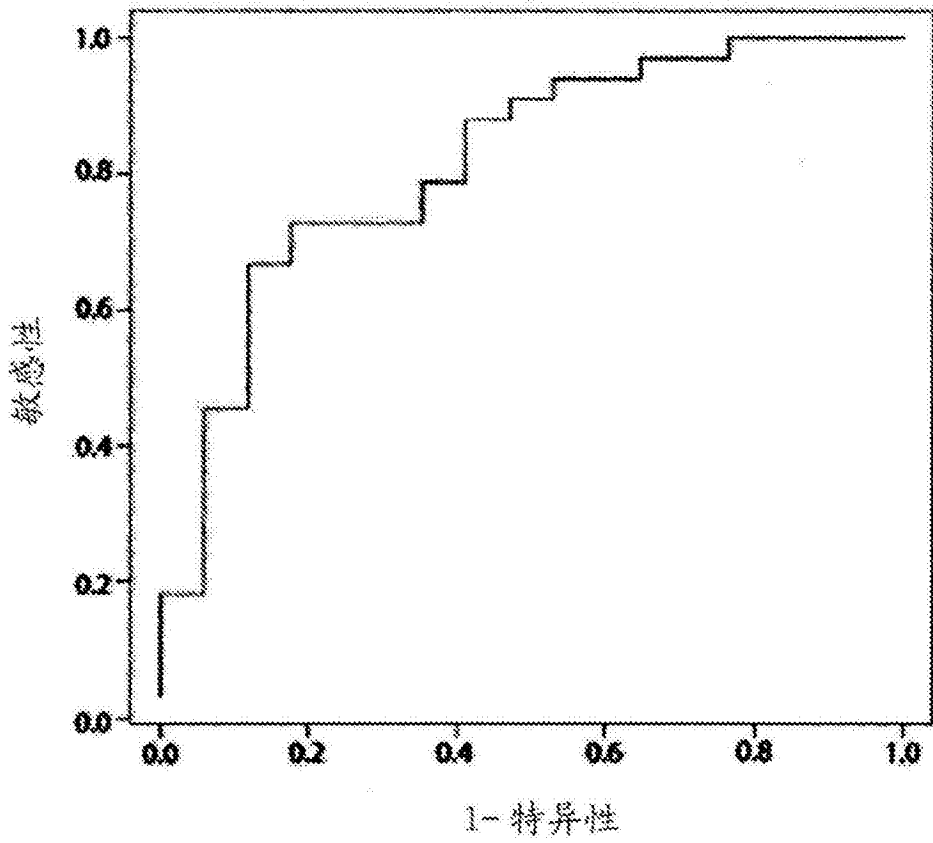


图3B续

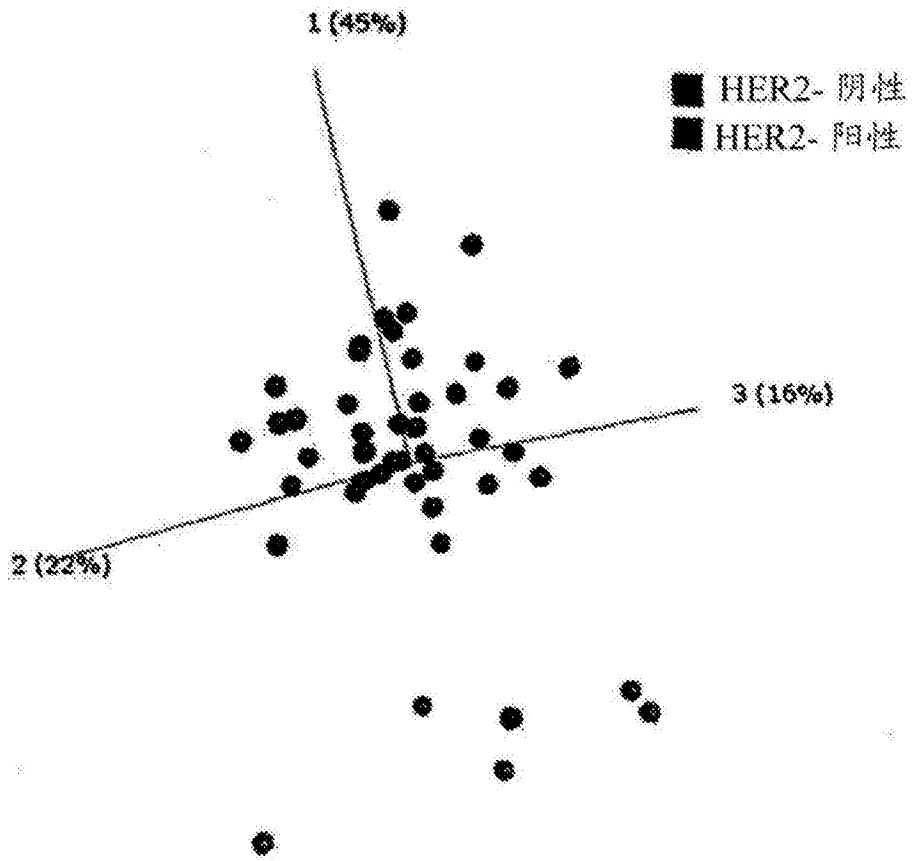


图3C

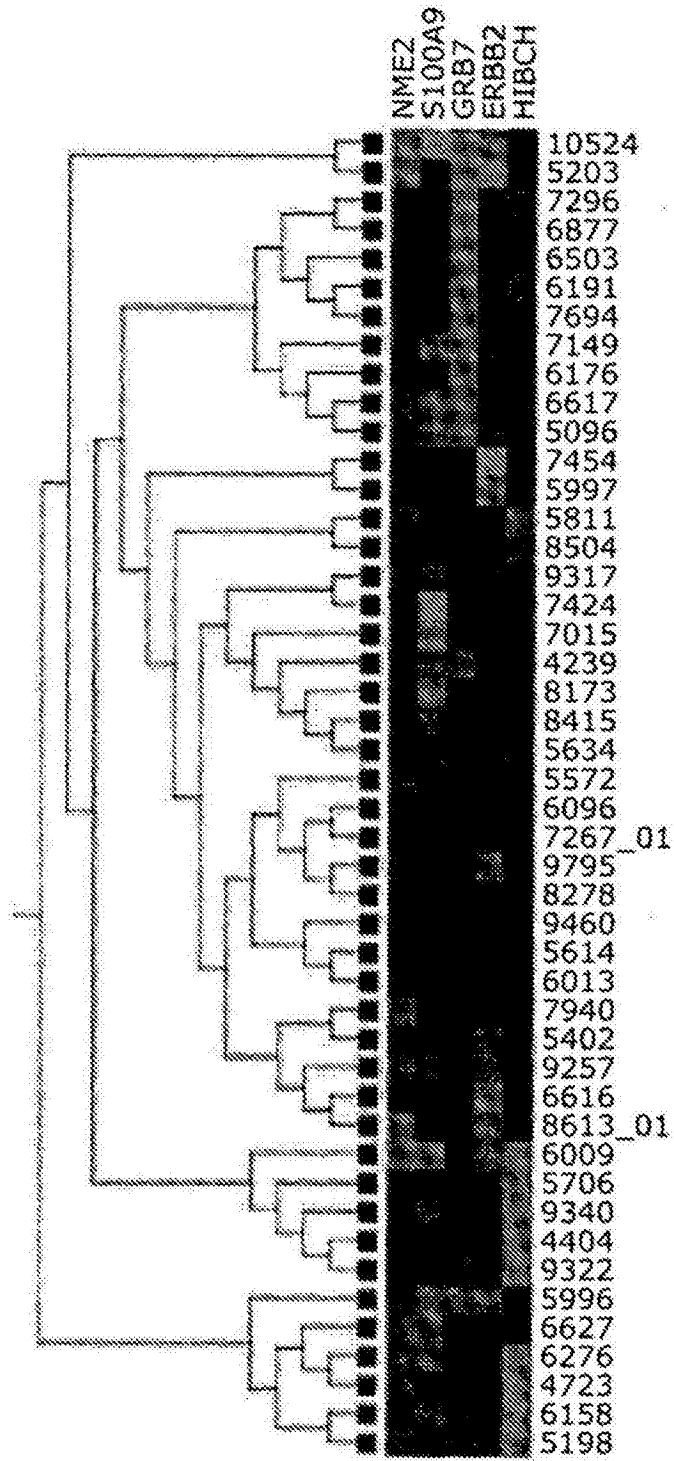


图3C续

ROC 面积 = 0.98

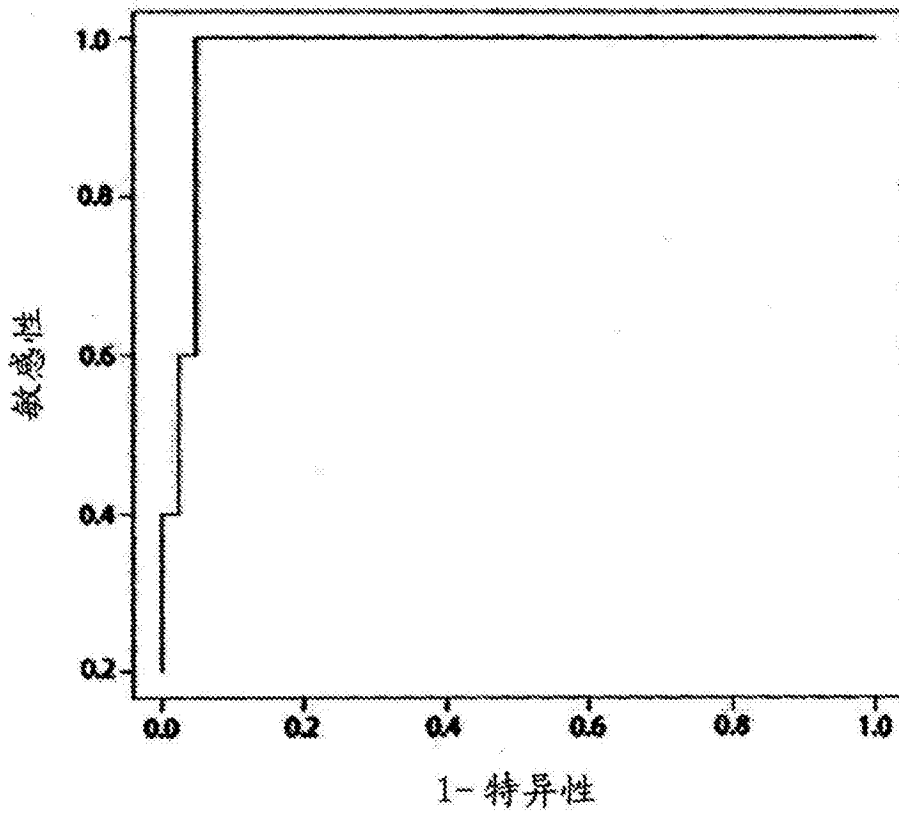


图3C续

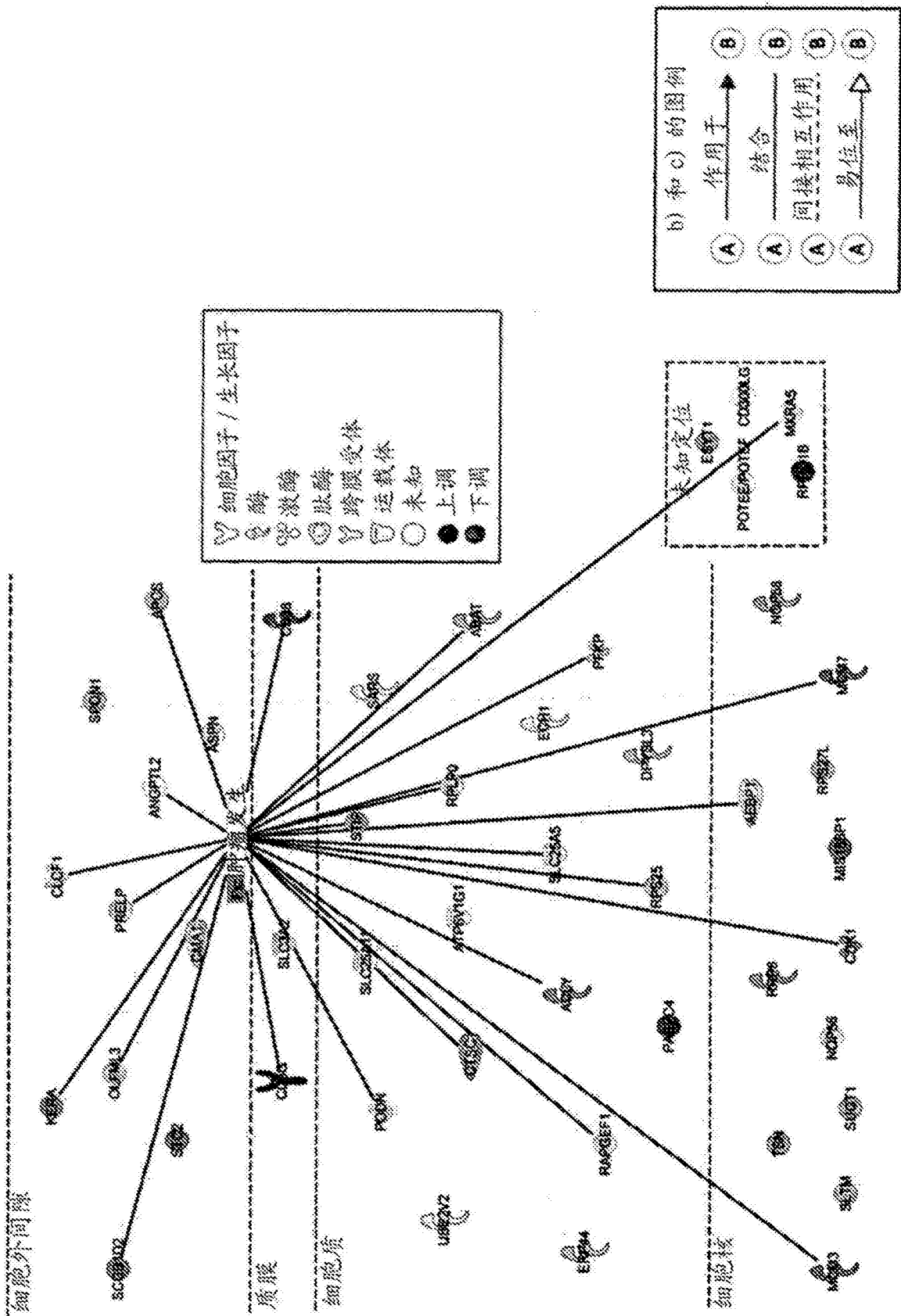
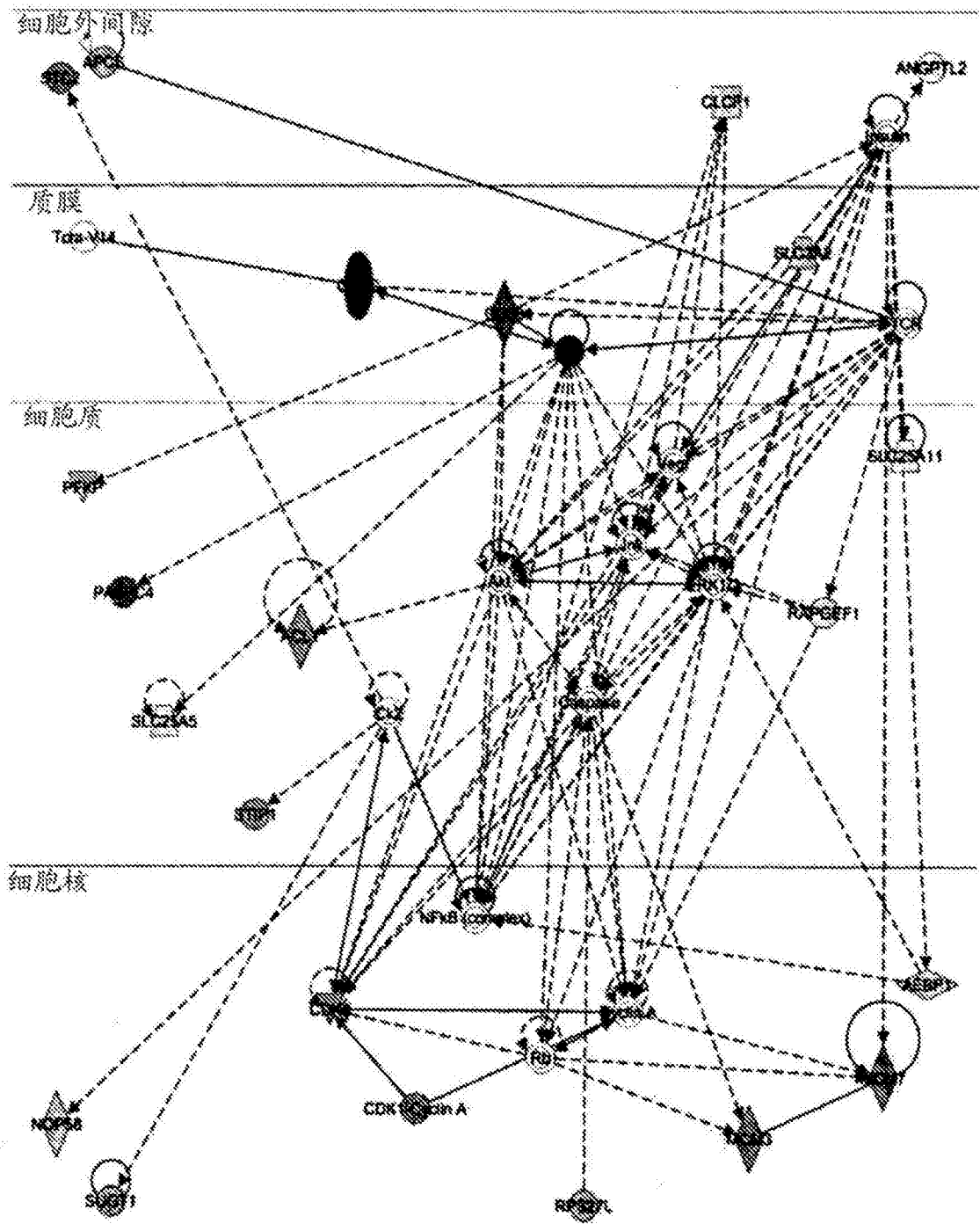


图4A



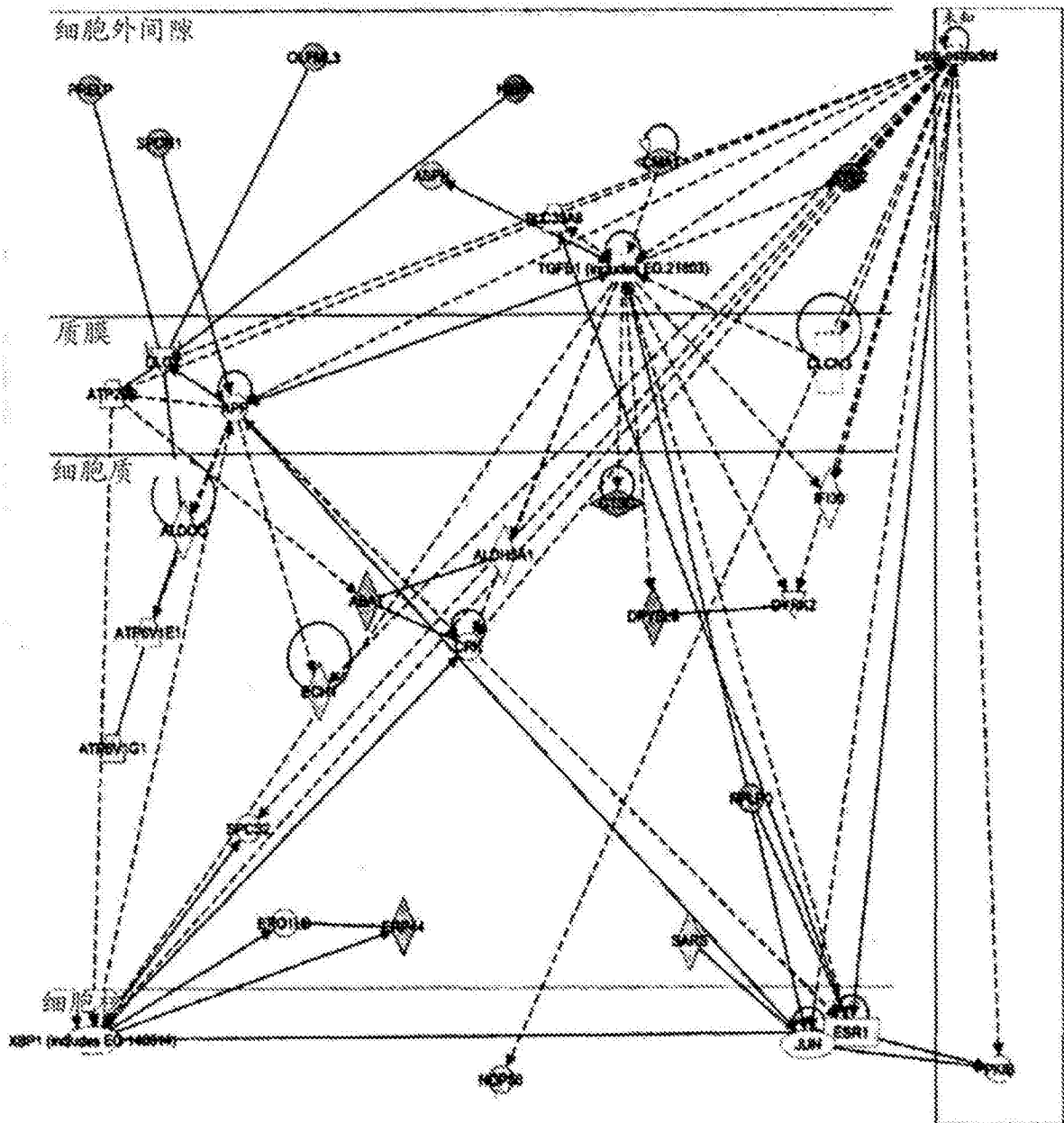
网络 1



DNA 复制、重组和修复，细胞周期和自由基清除

图4B

网络 2



基因表达、传染病和癌症

图4C

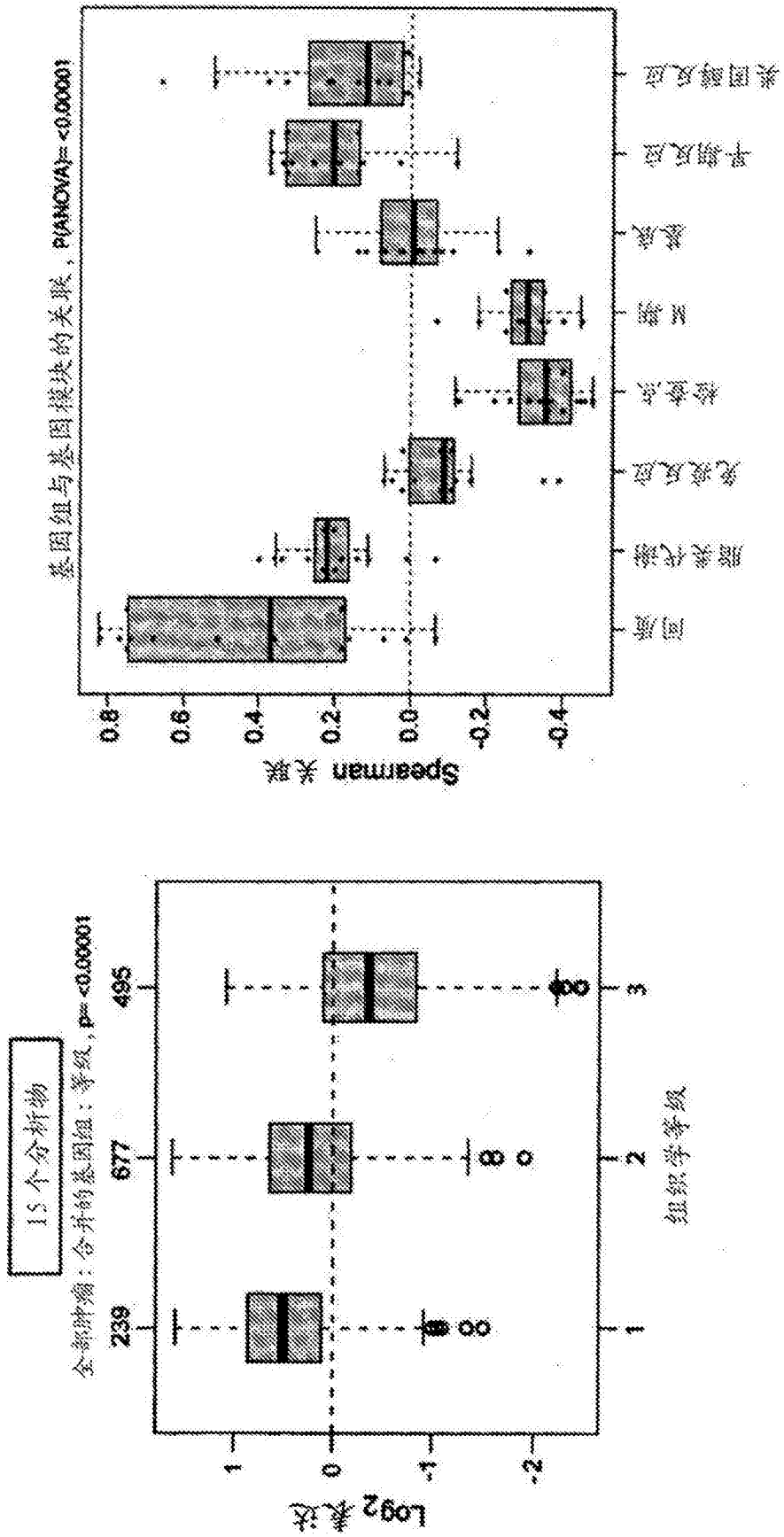


图5A

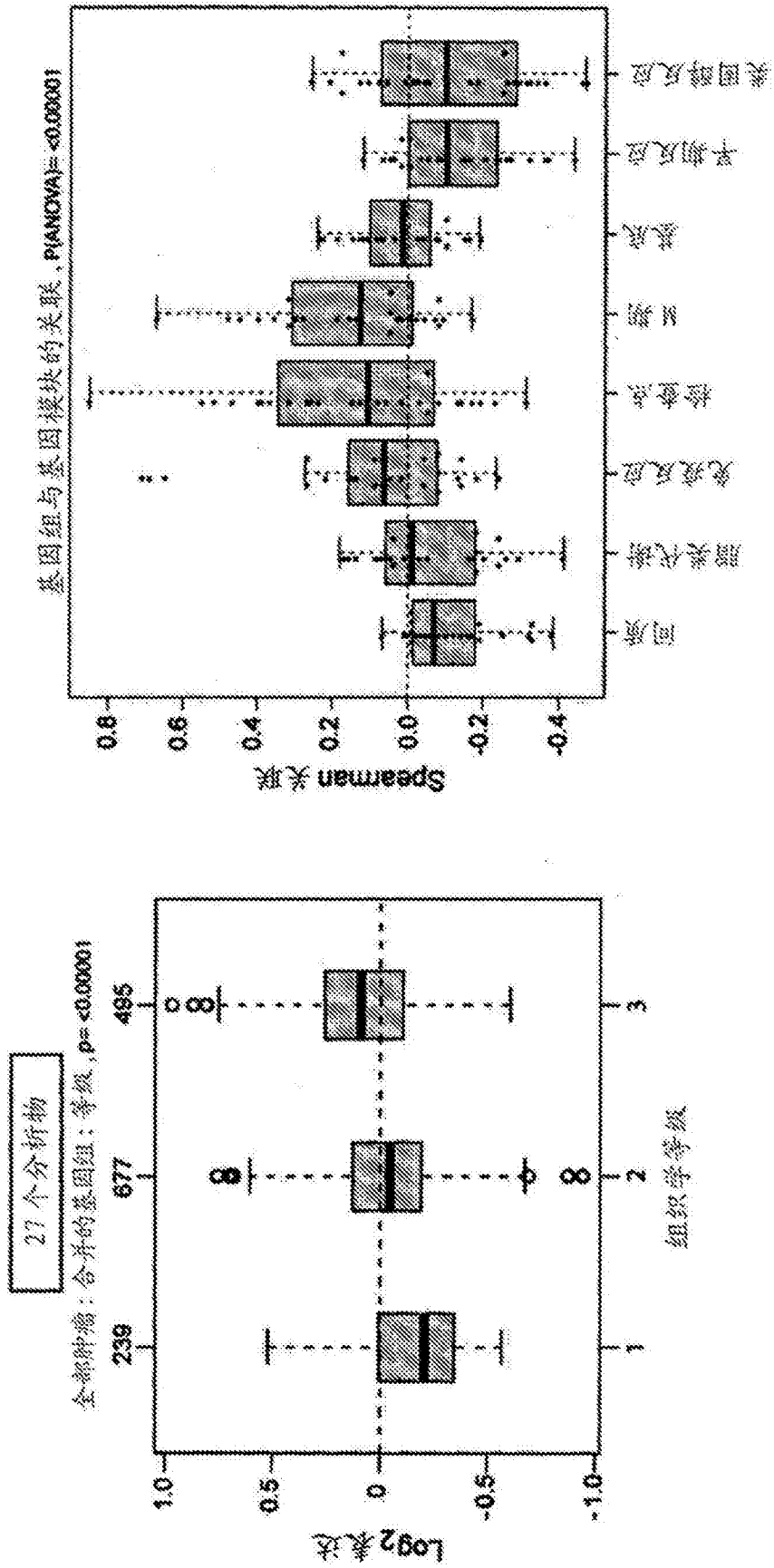
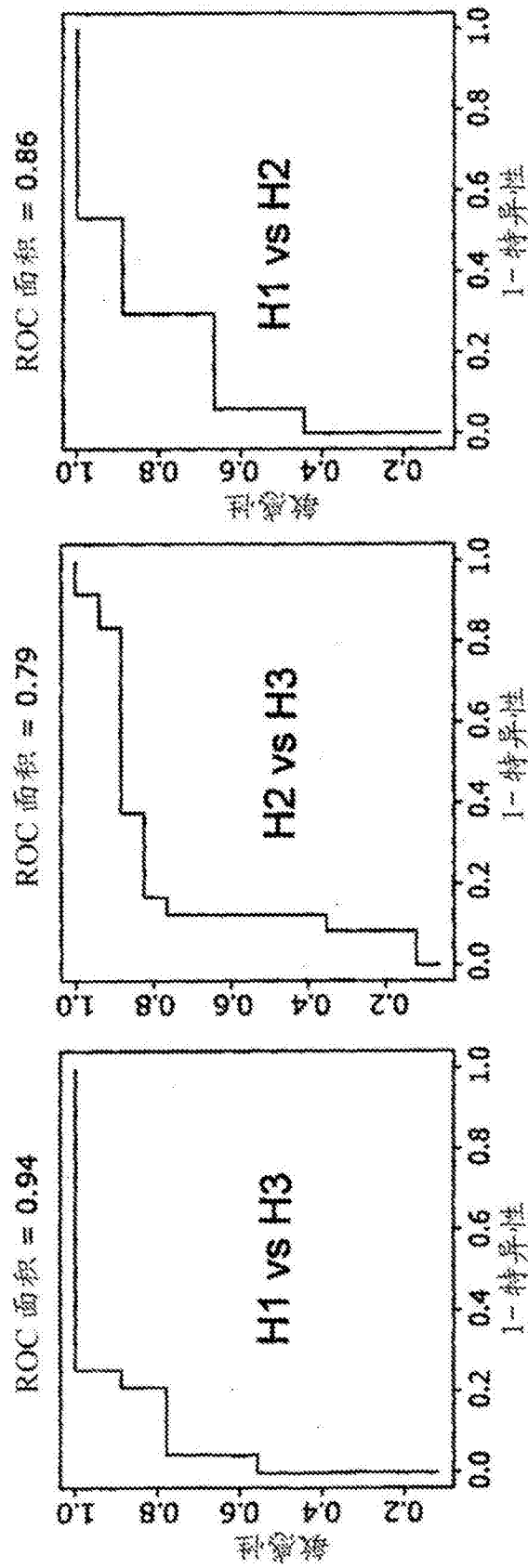


图5B



分别区分 H1 vs H2、H2 vs H3 和 H1 vs H3 的针对 79- 丛生物标记物特性的 ROC AUC 值

图6

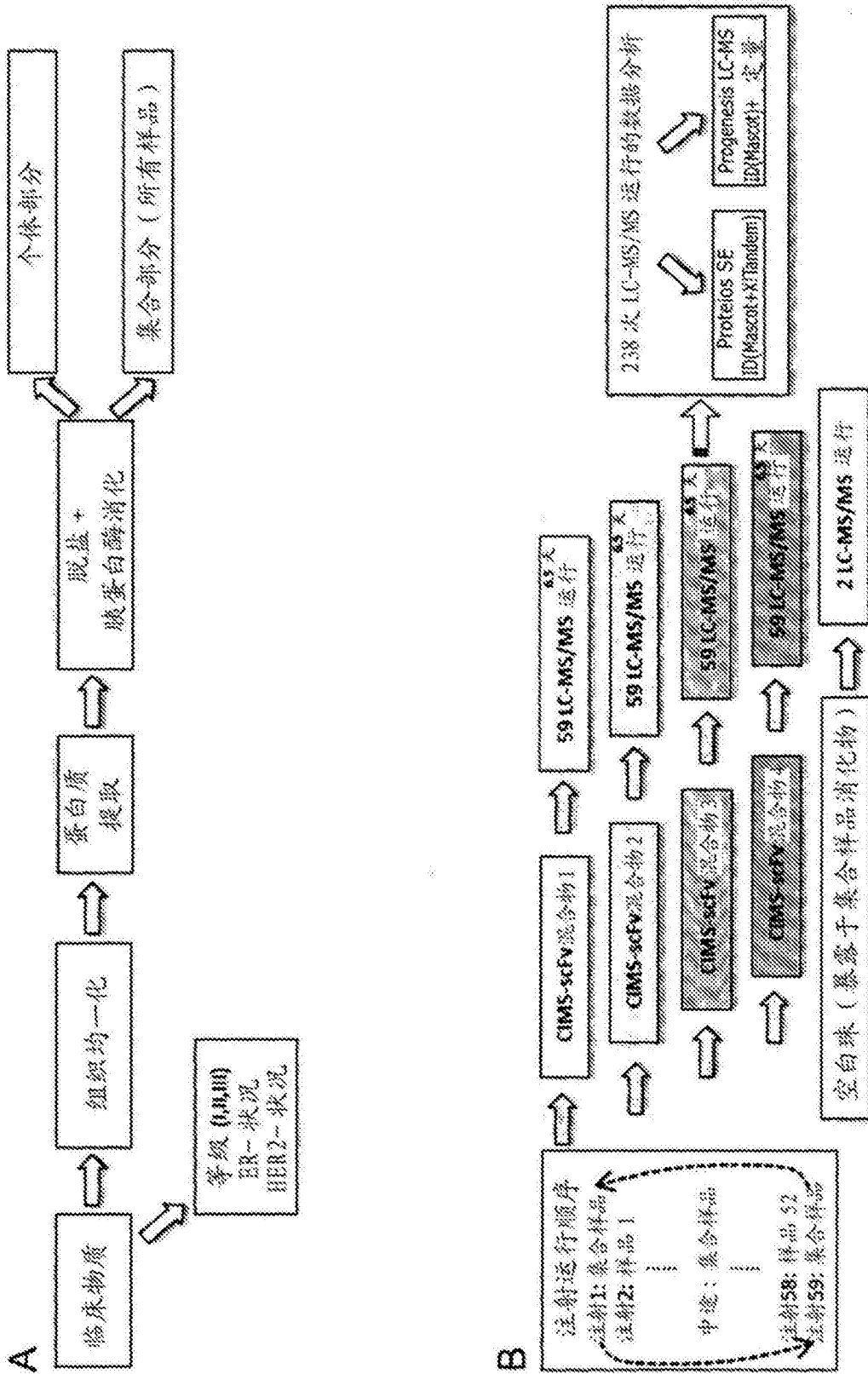


图7

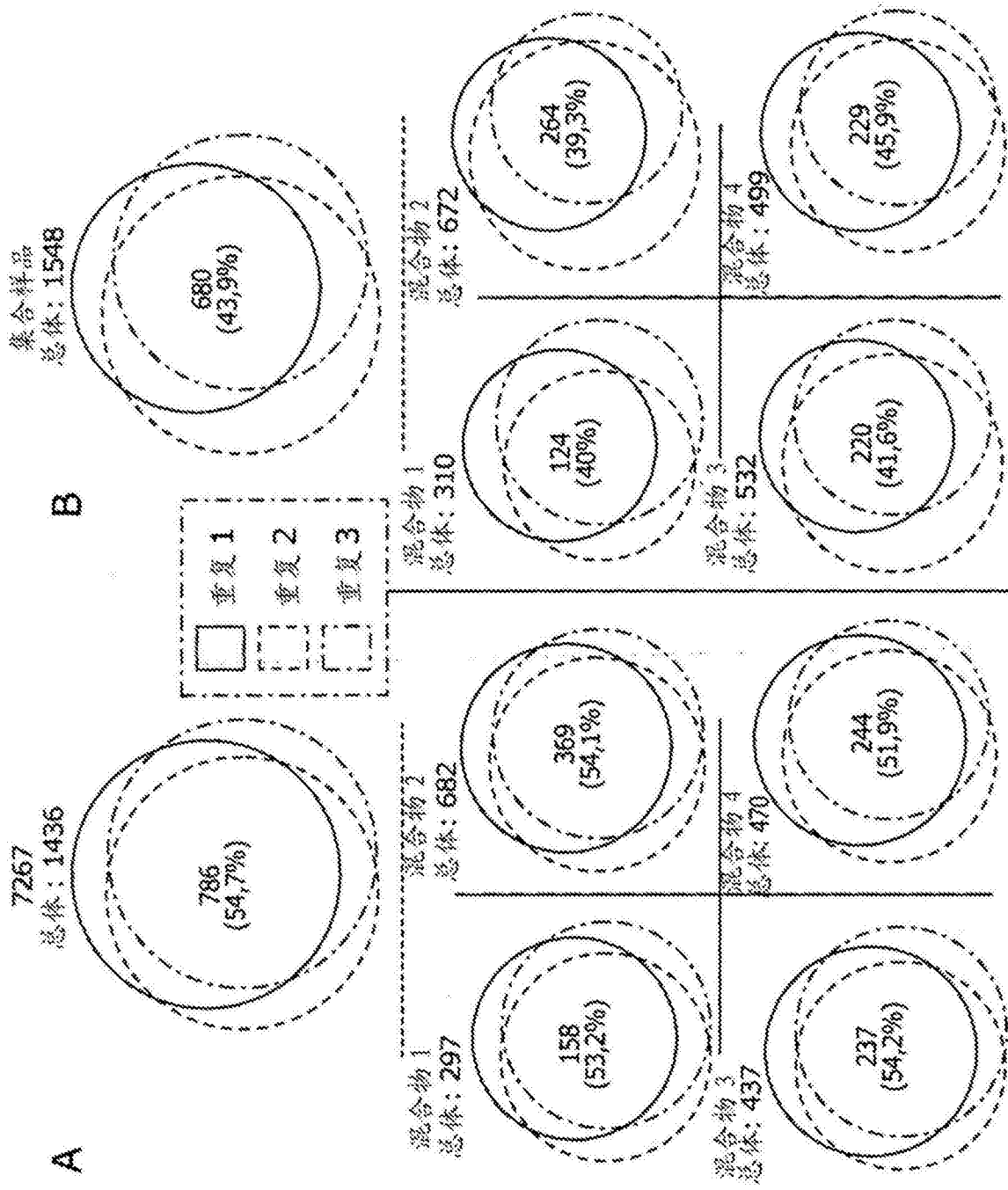


图8

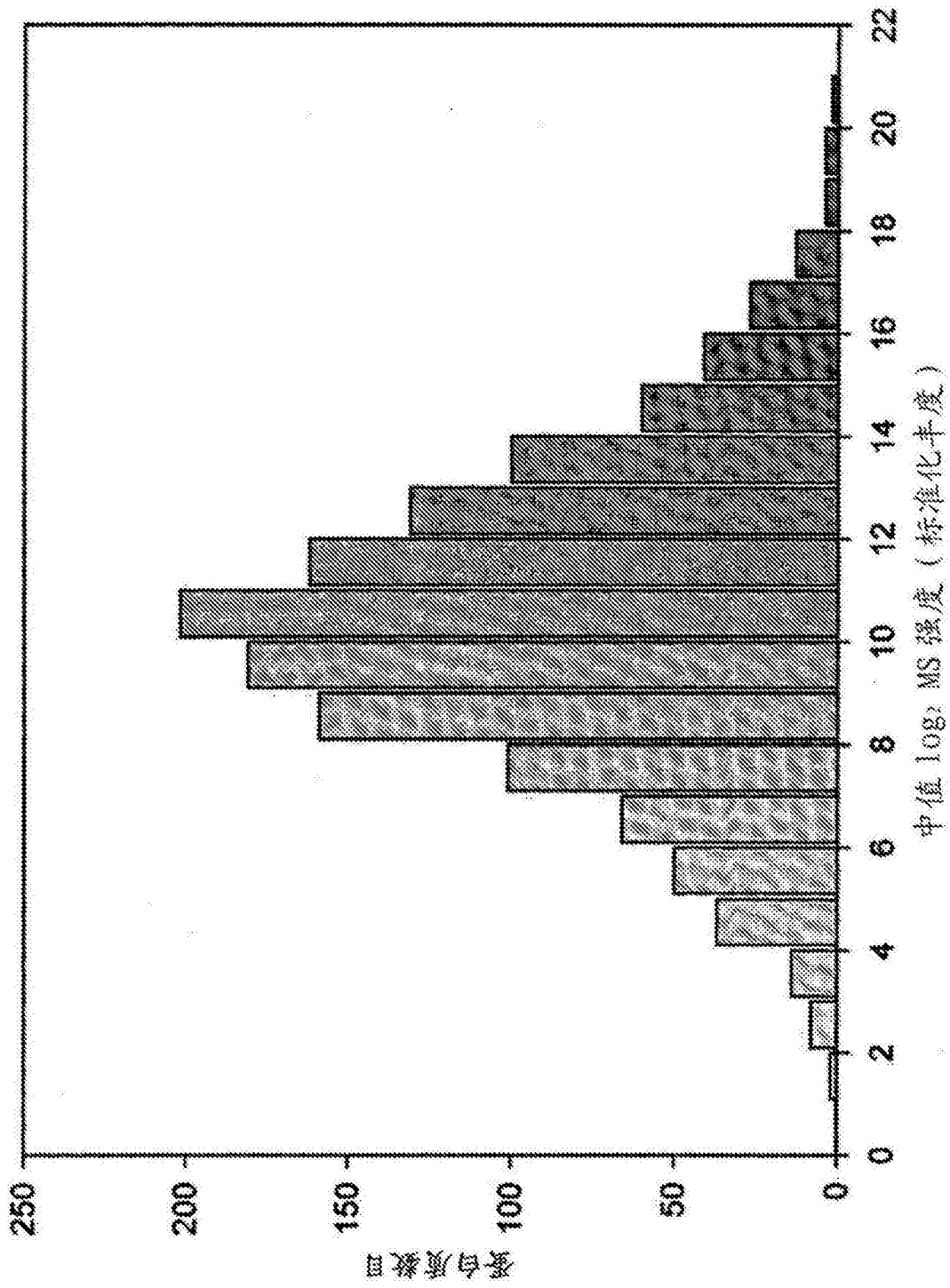


图9A



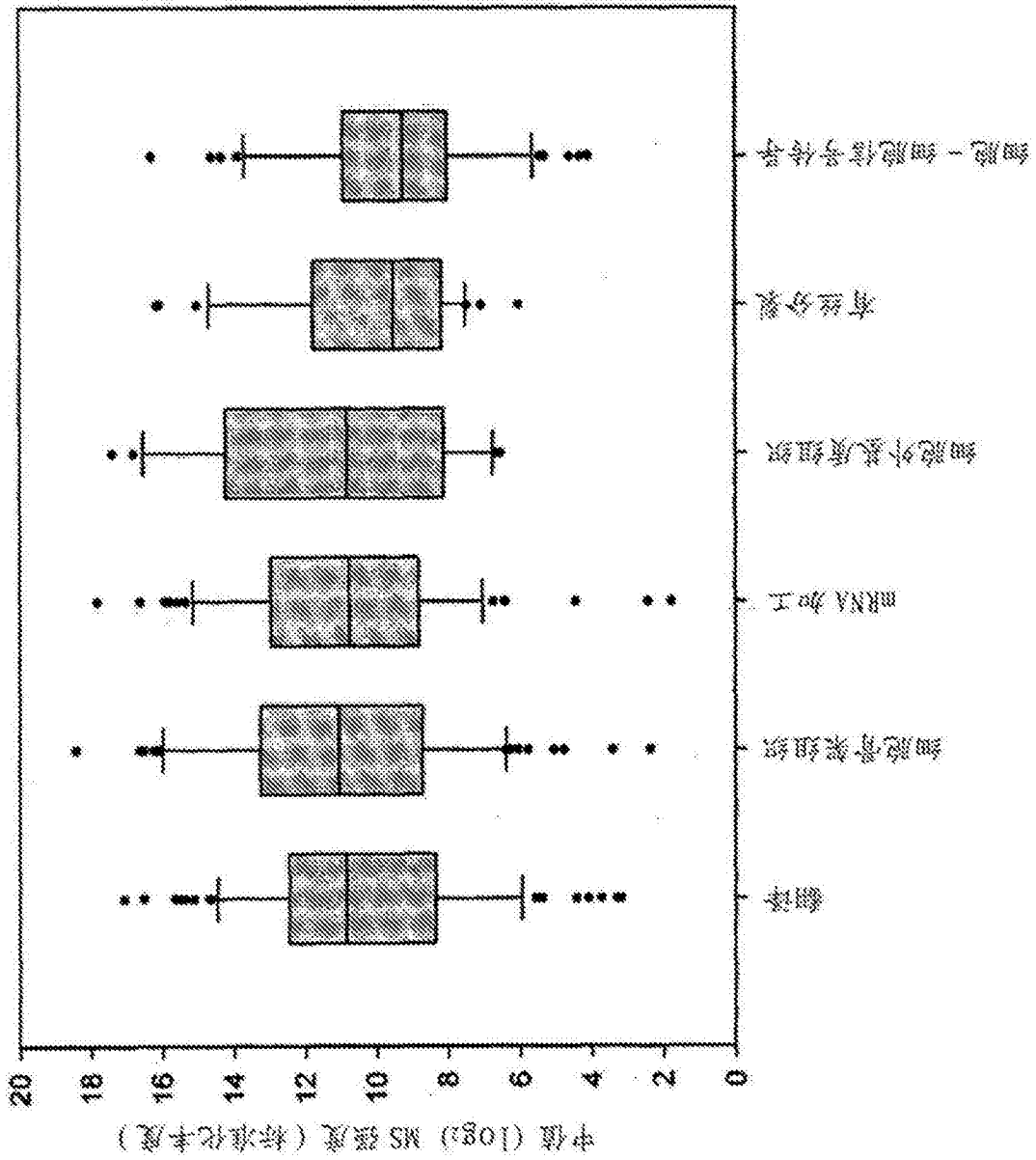


图9B

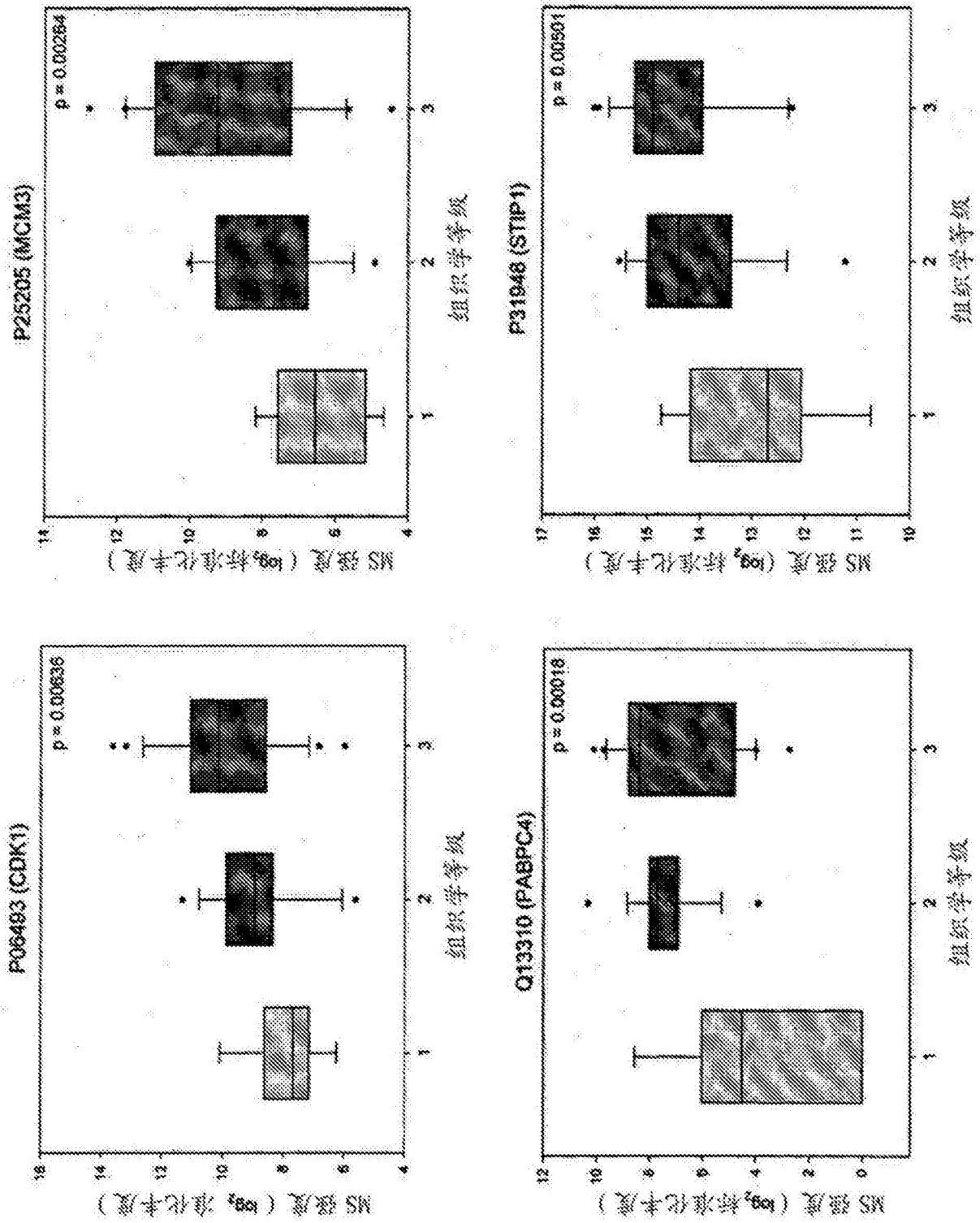


图10

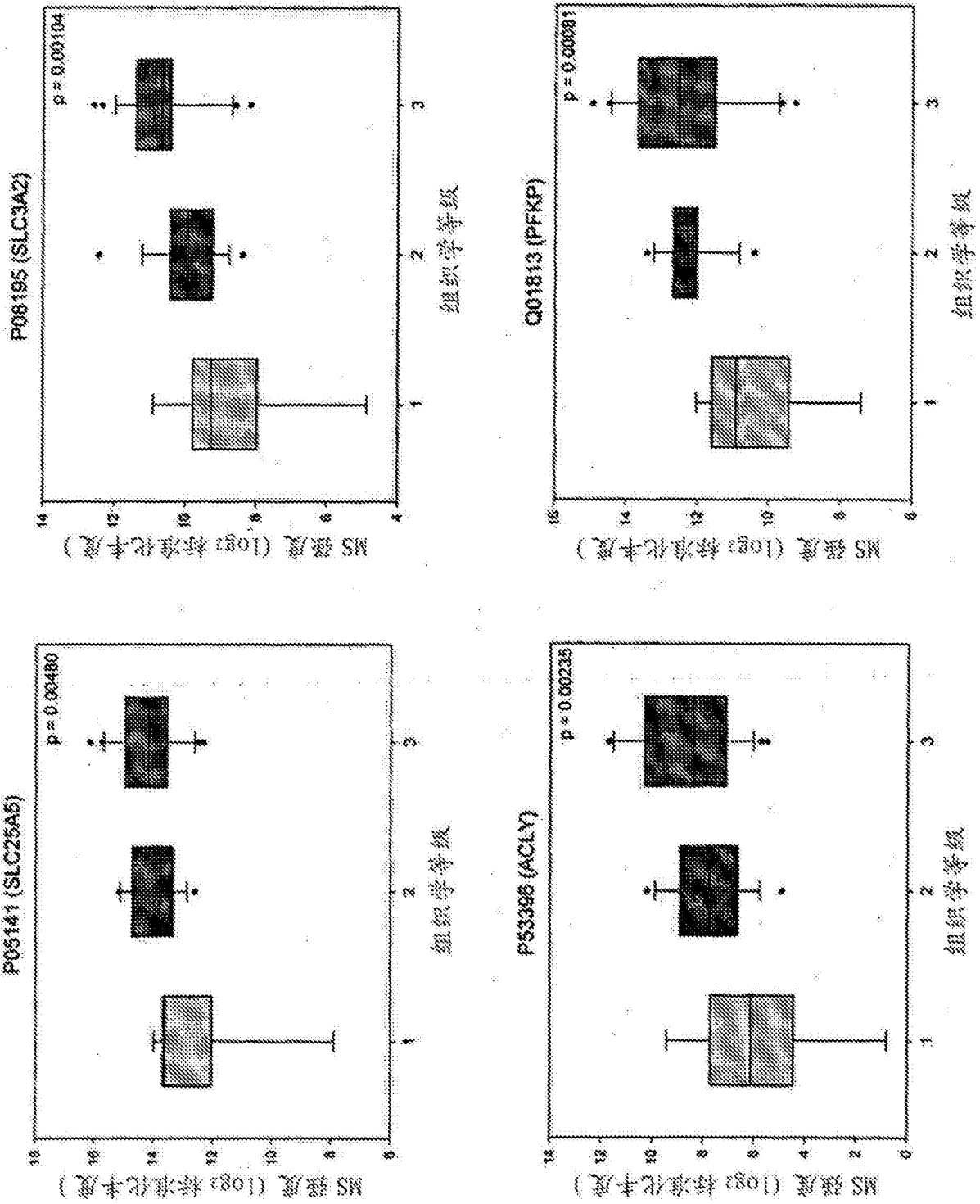


图10(续)

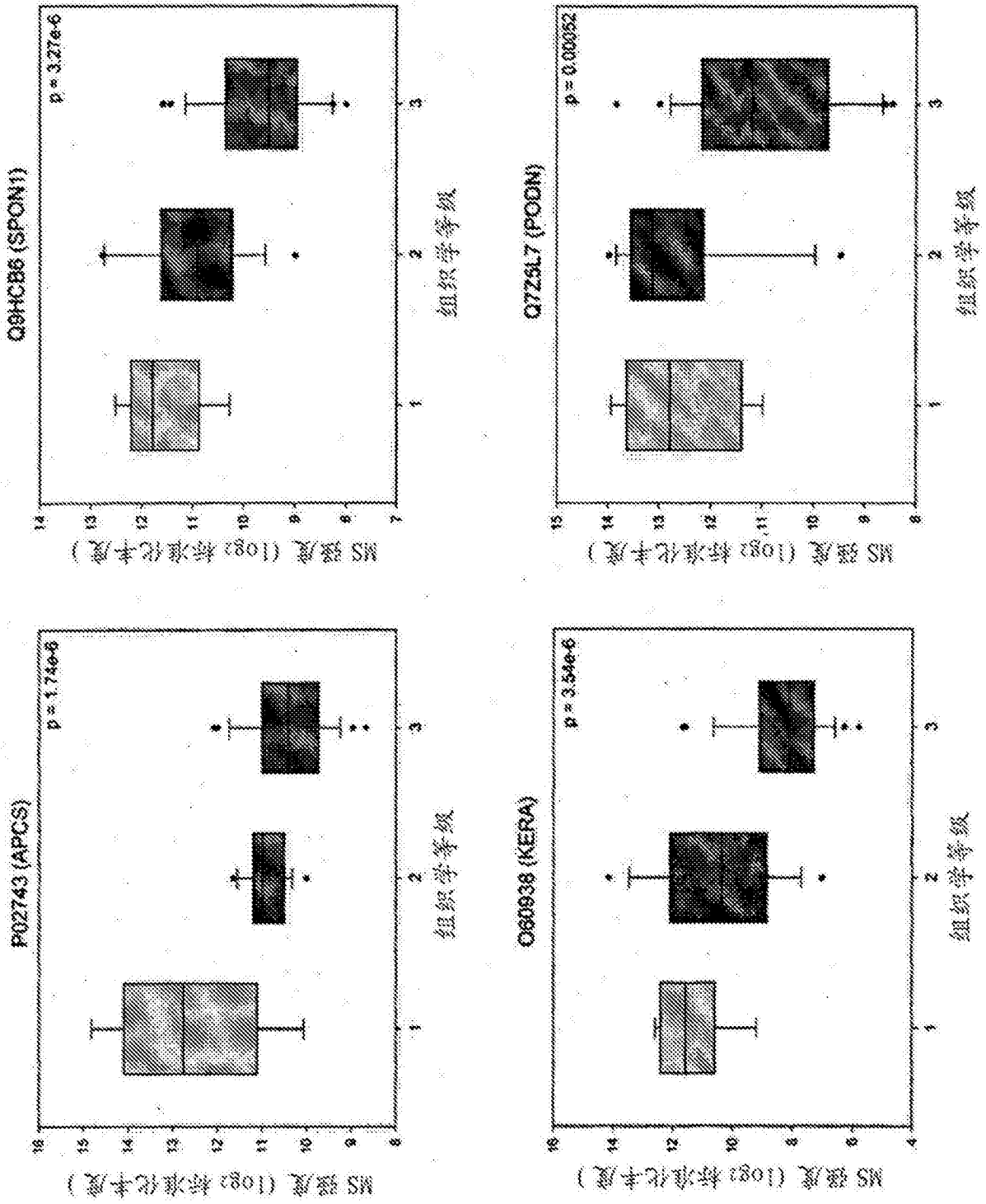


图11

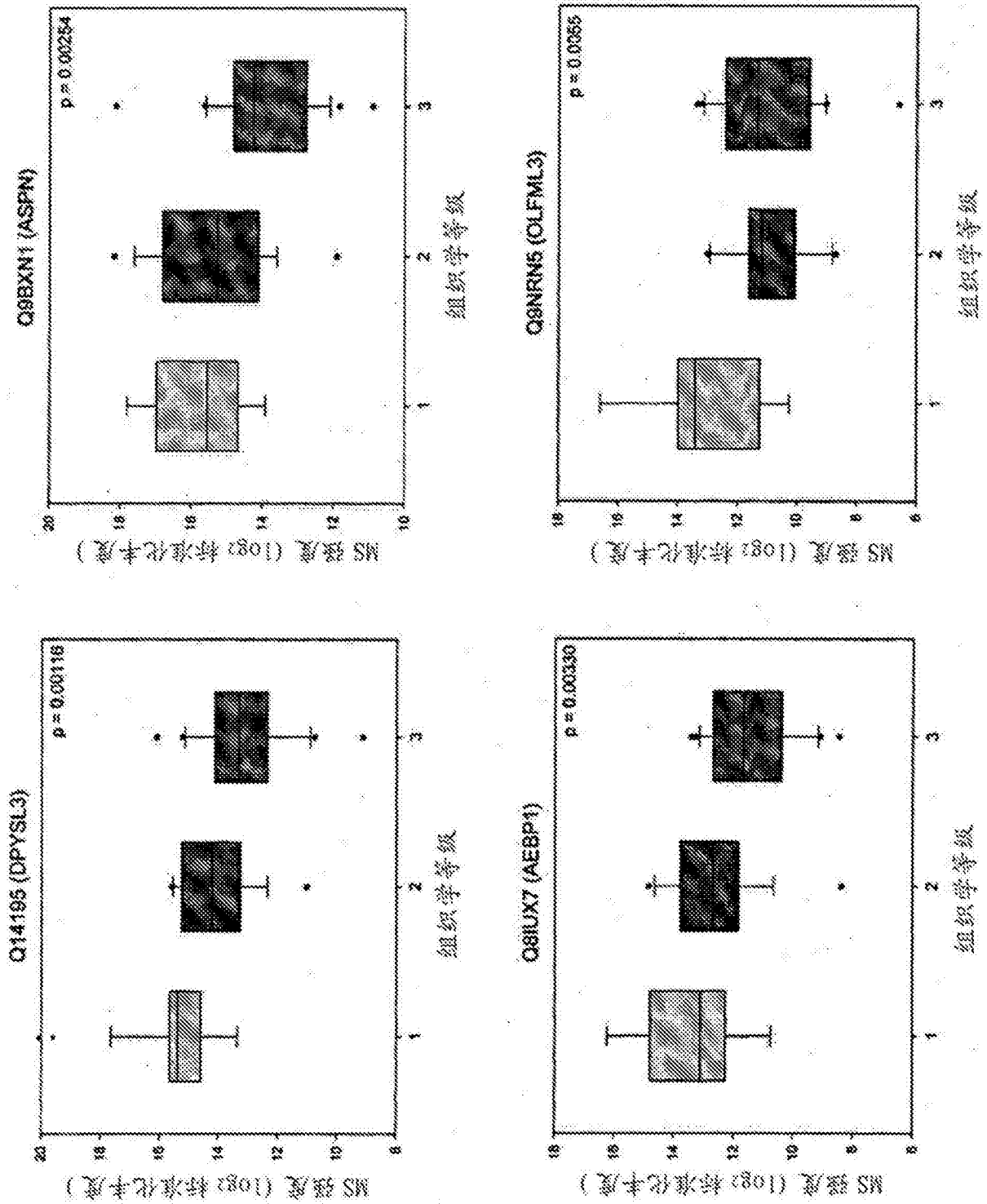


图11(续)

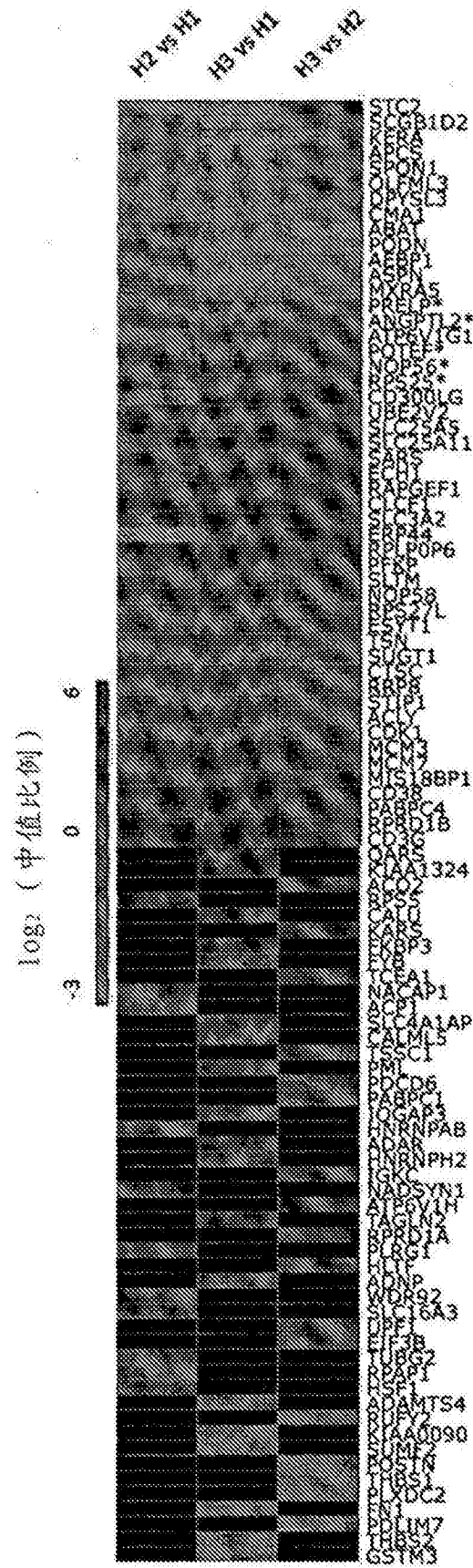


图12A

比较	ROC 面积 (未过滤)	ROC 面积 (两组比较)
H1 vs. H2	0.63	0.92
H1 vs. H3	0.71	0.92
H2 vs. H3	0.77	0.91
H1, H2 vs. H3	0.77	0.89
H1 vs. H2, H3	0.64	0.85
H2 vs. H1, H3	0.65	0.75

图12B

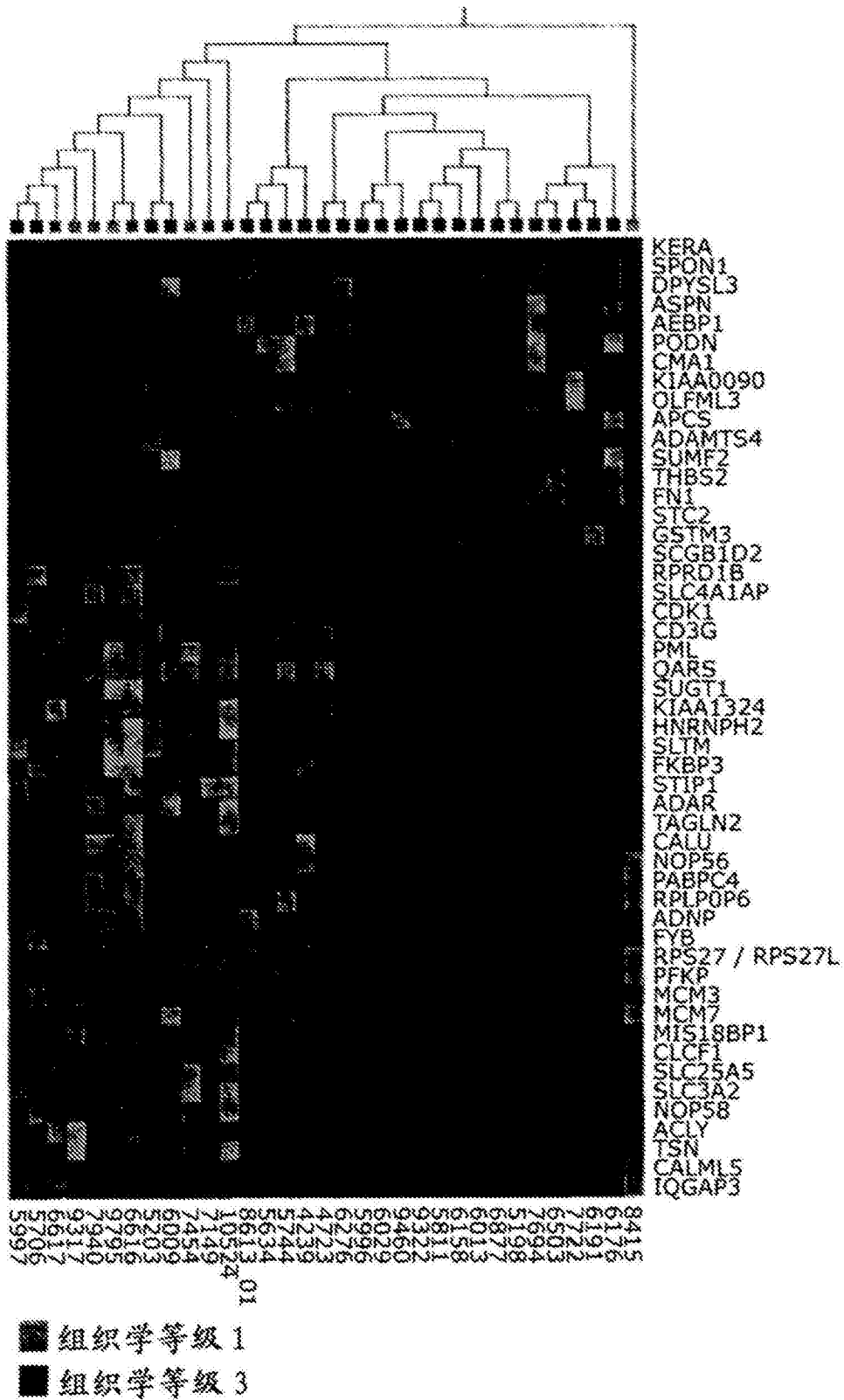


图12C



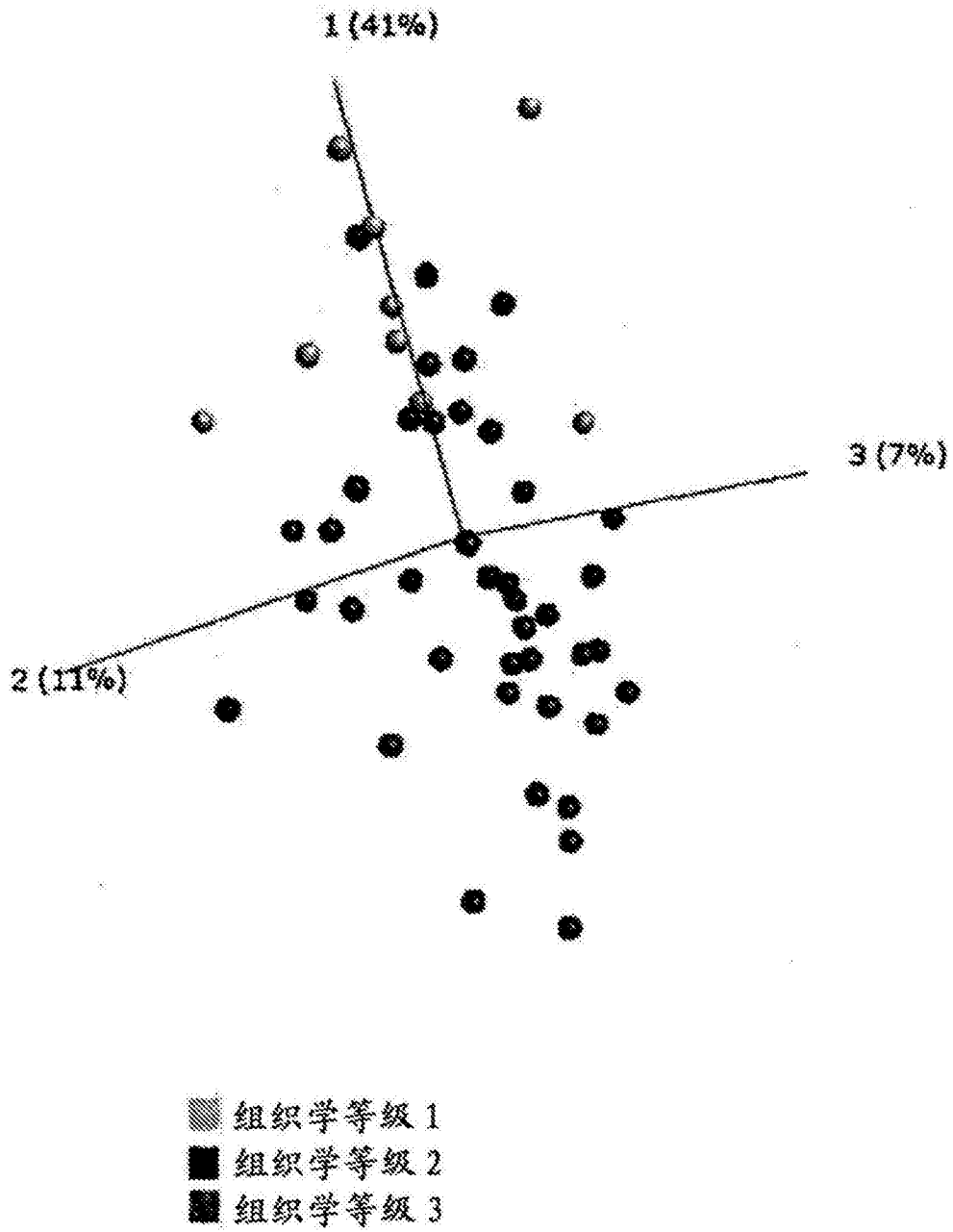


图12D

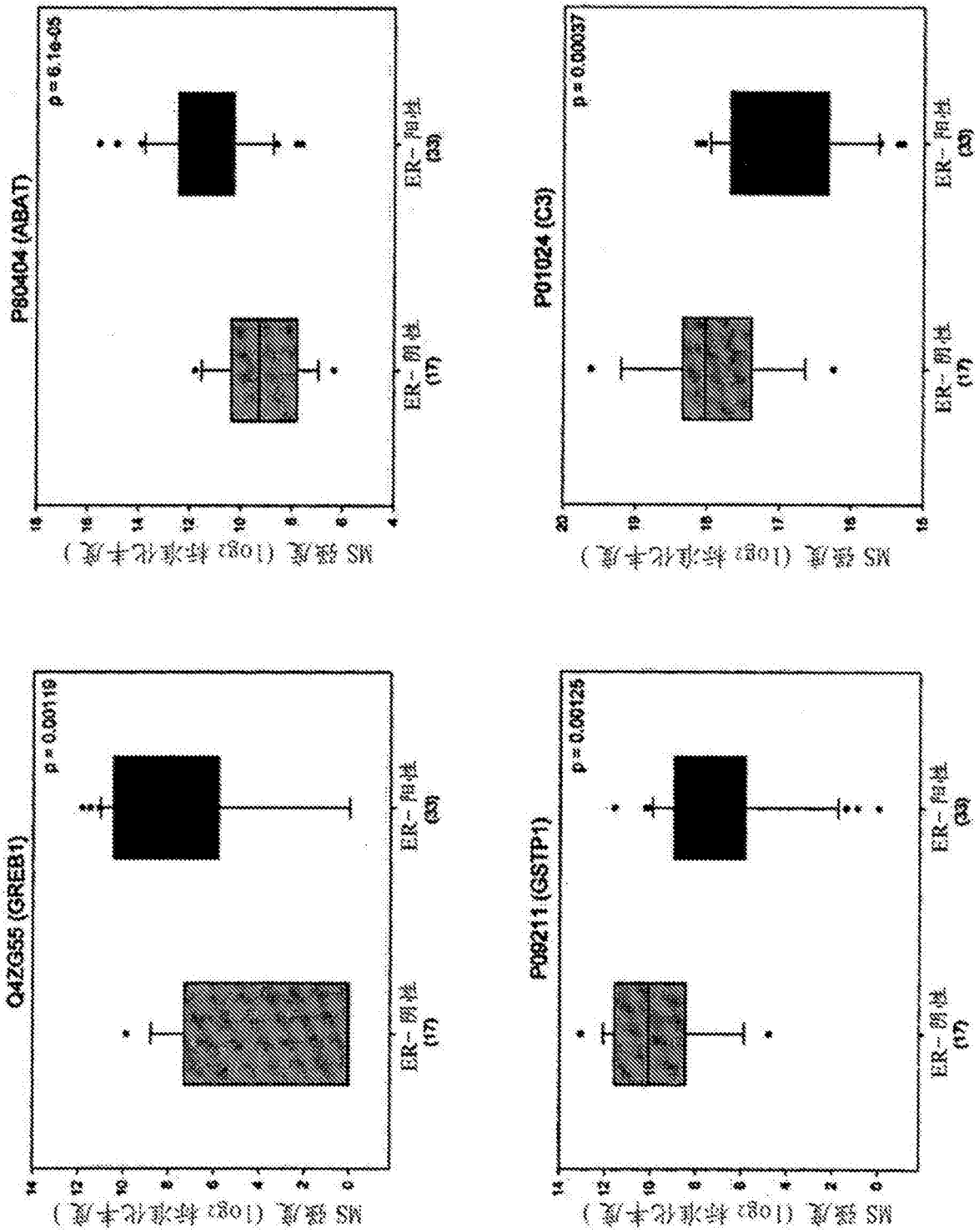


图13A

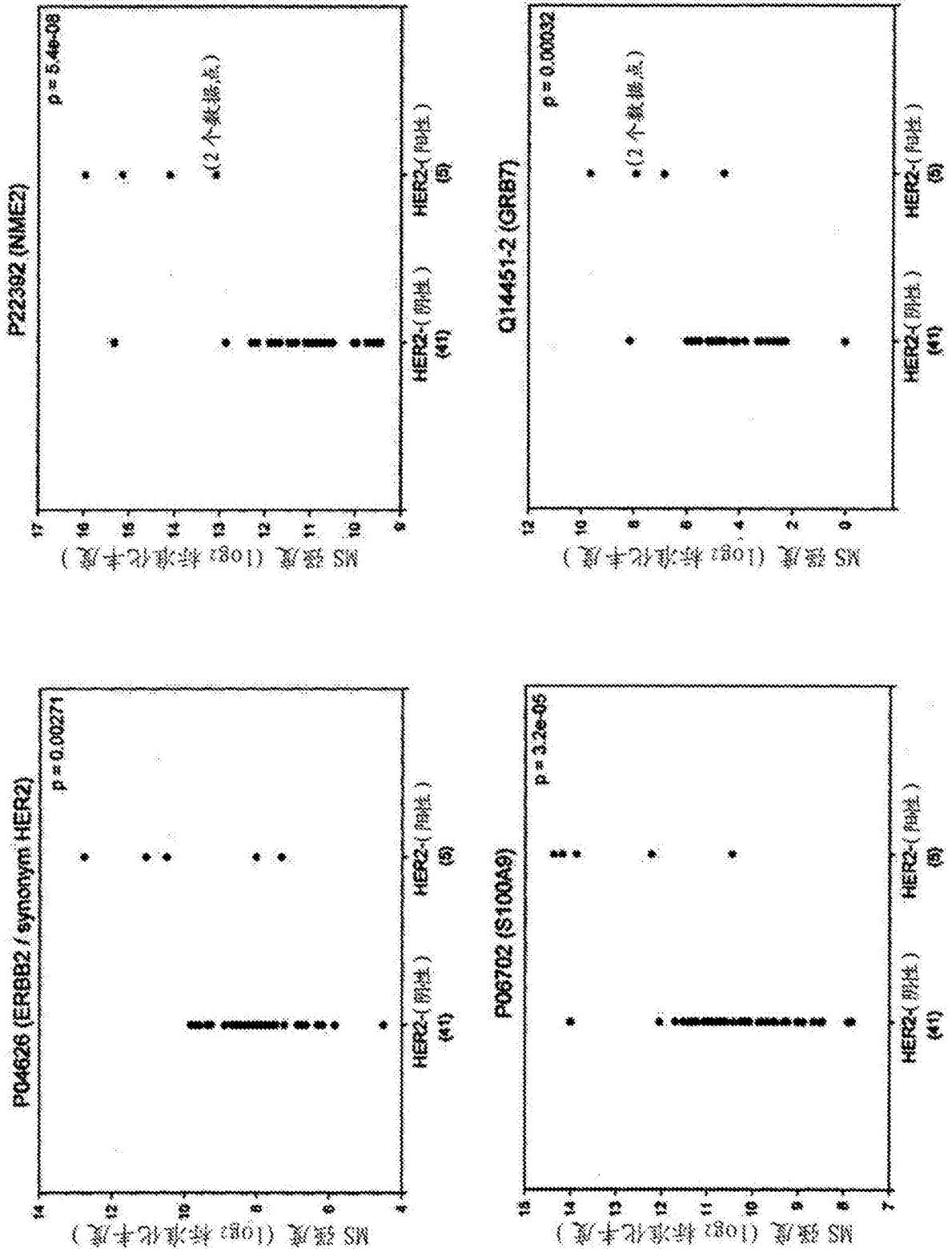


图13B

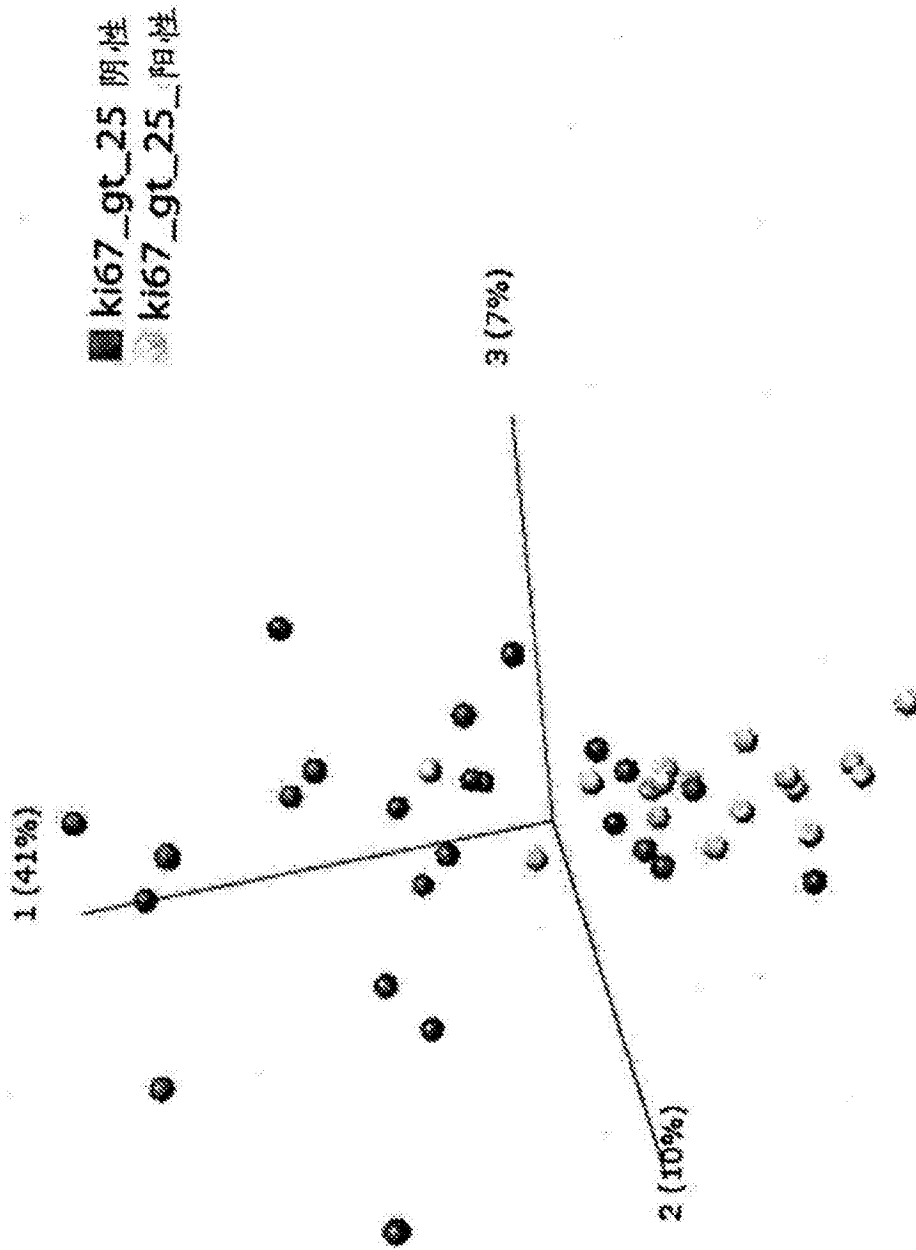


图14A

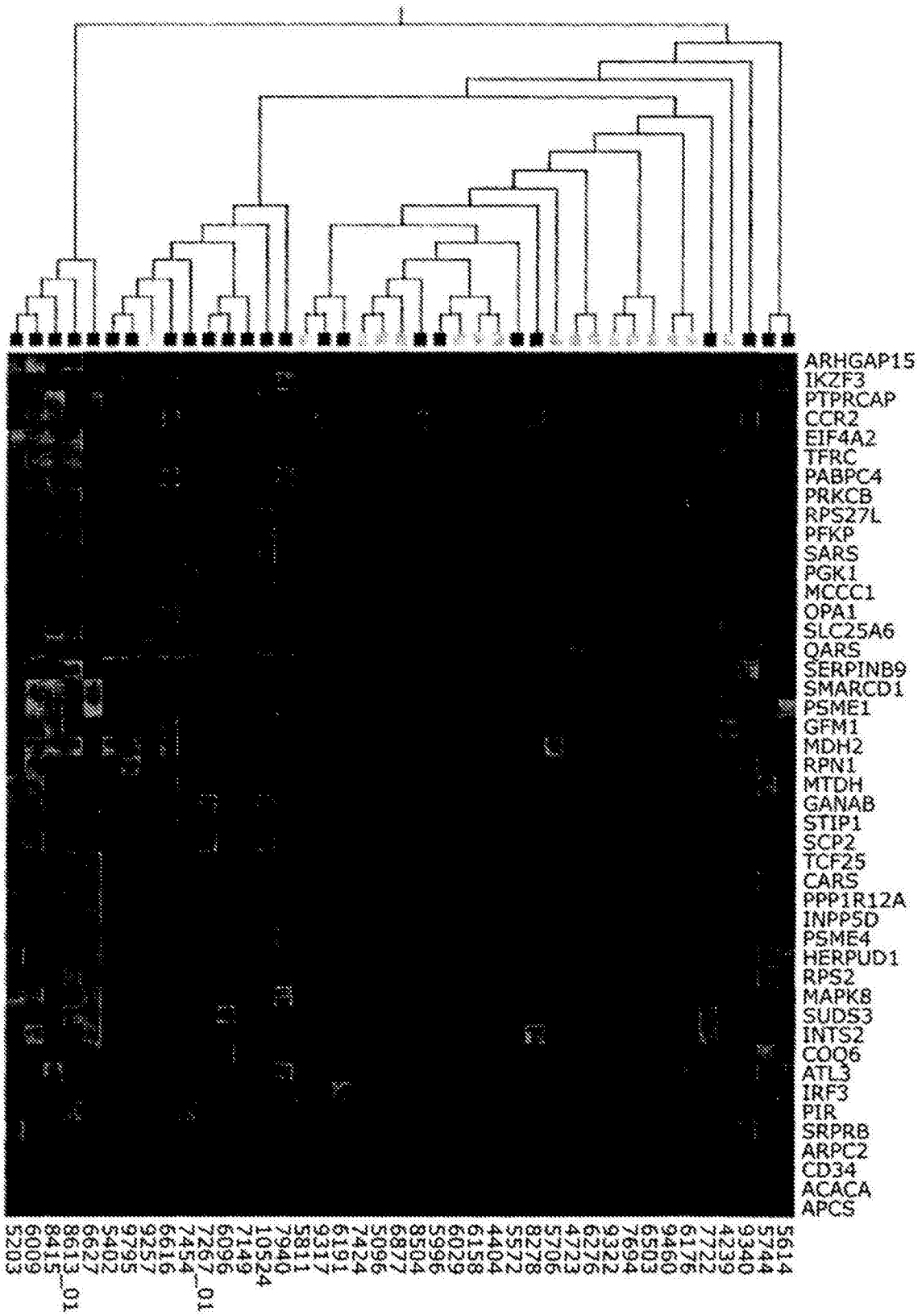


图14A(续)

ROC 面积 = 0.84

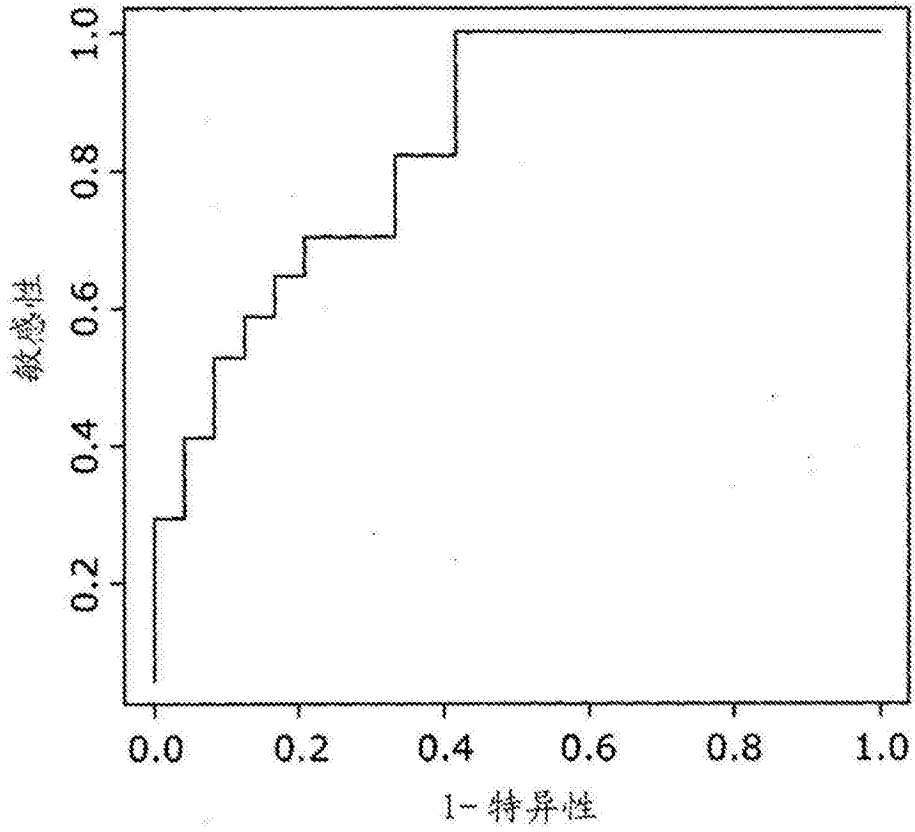
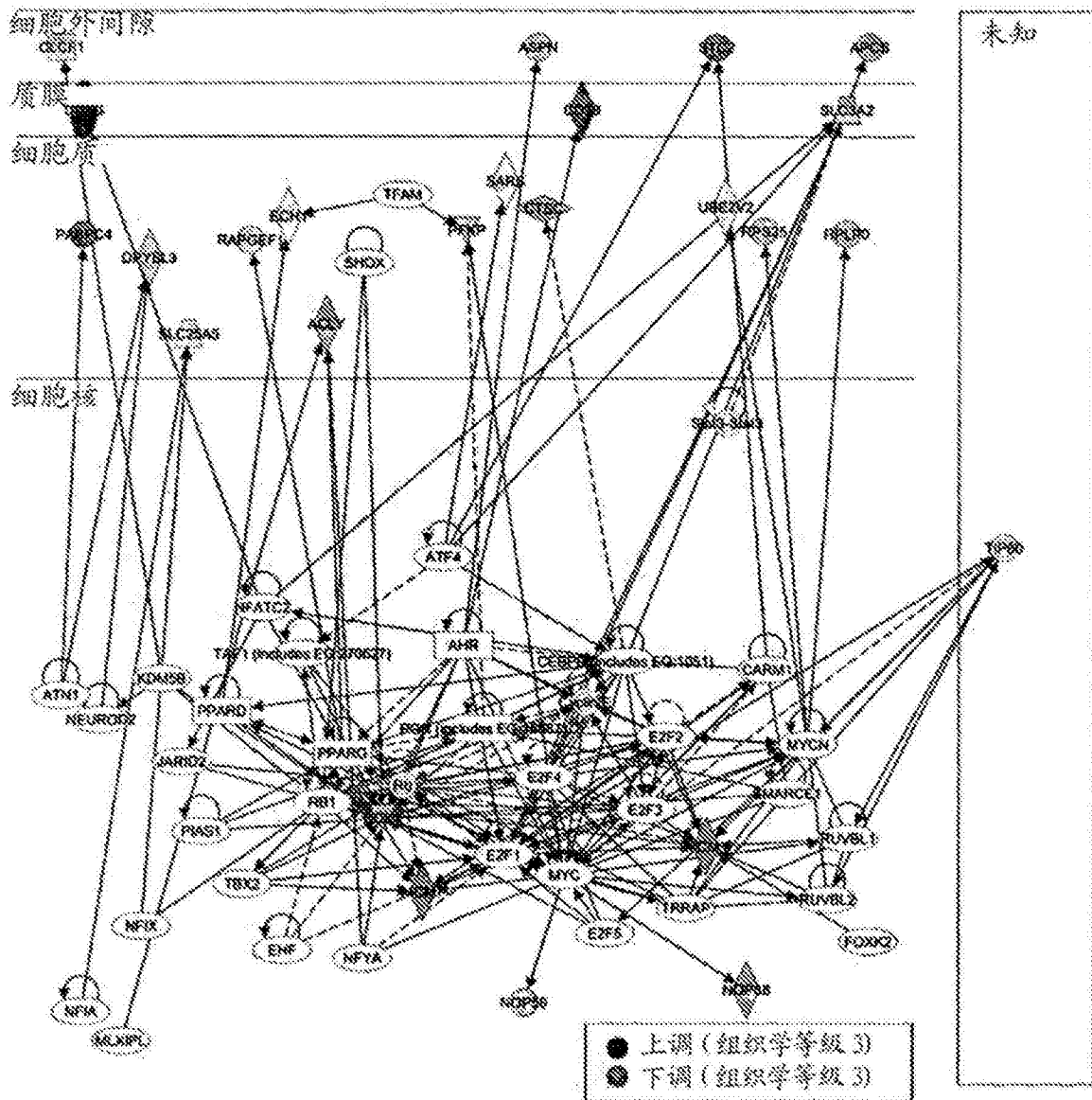


图14B

组织学等级

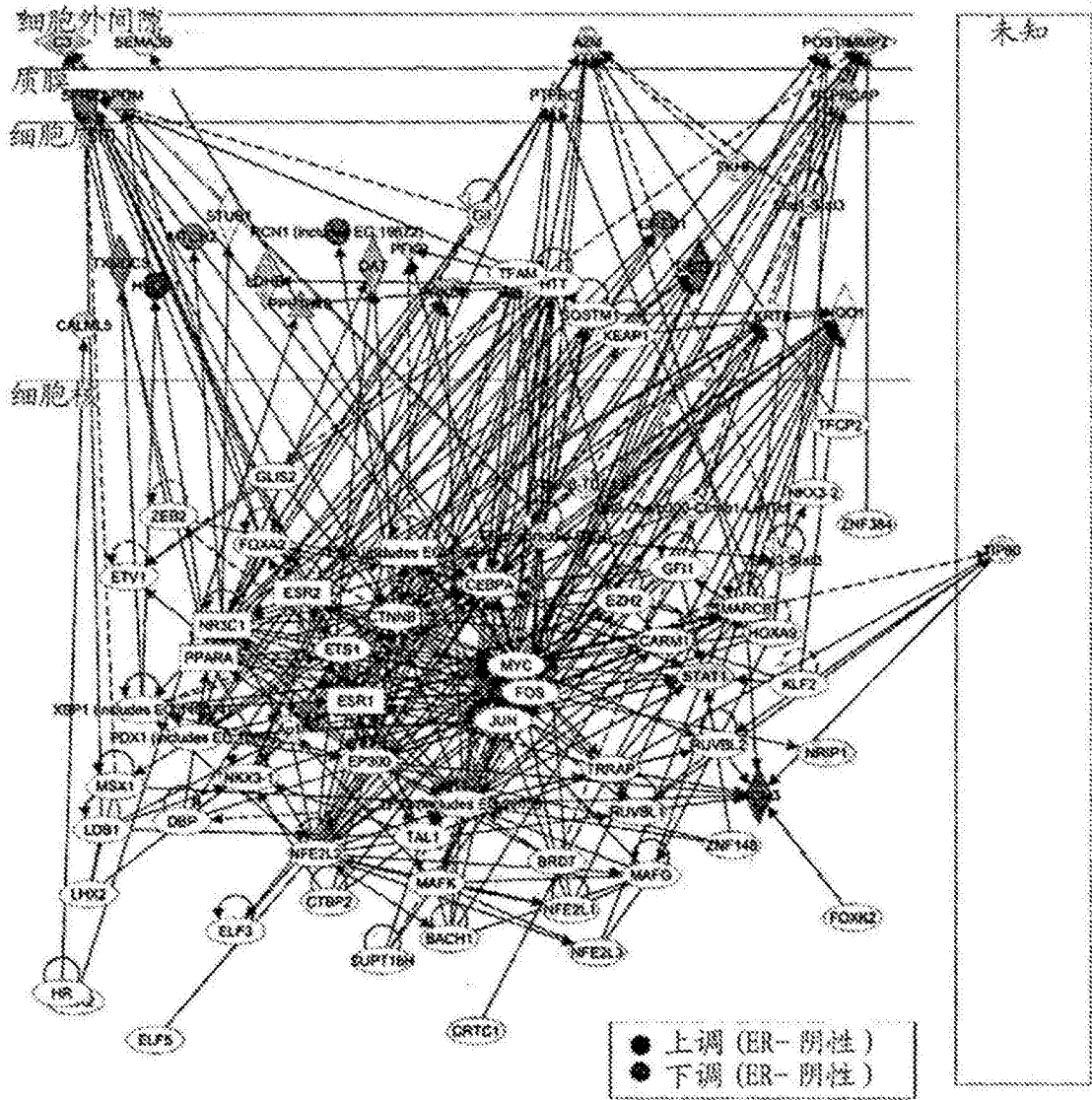


转录因子关联网络

顶部转录调节因子
Rb
E2F2
E2F3
TFAM
ATF4

图15A

雌激素受体状况



转录因子关联网络

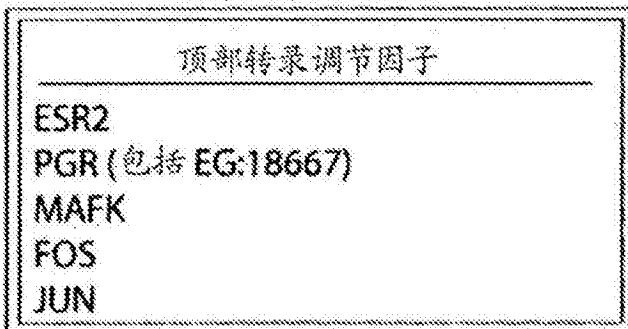


图15B



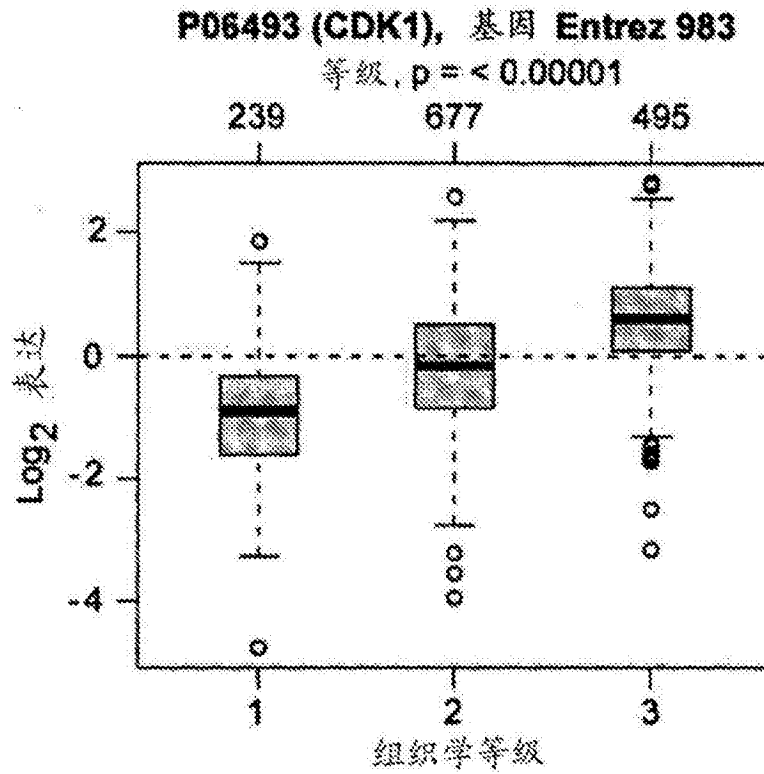


图16A

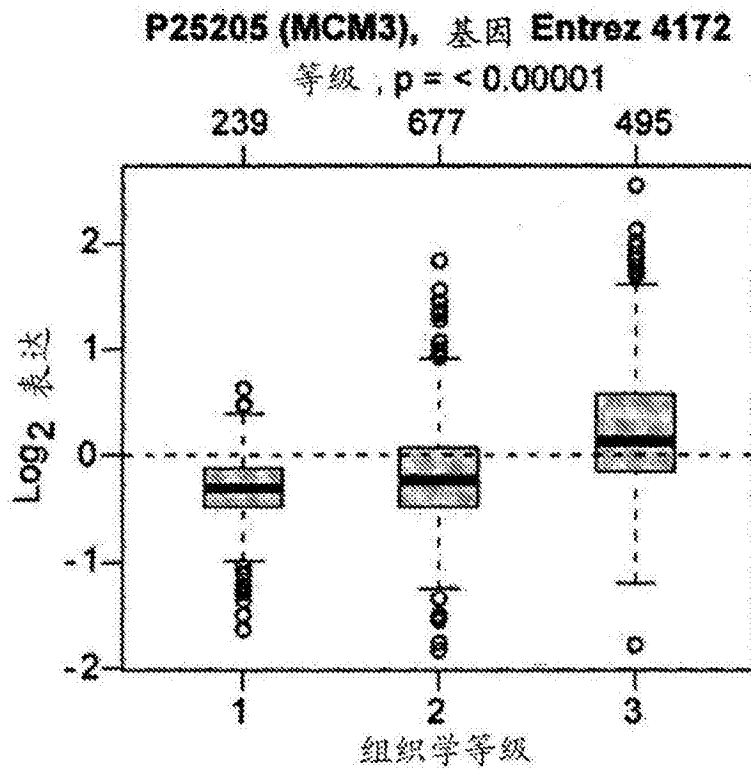


图16B

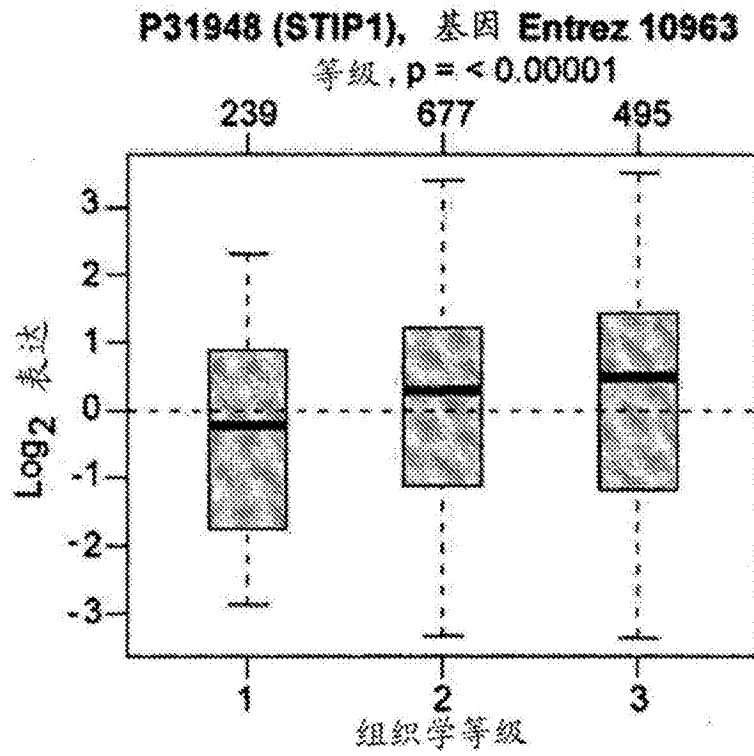


图16C

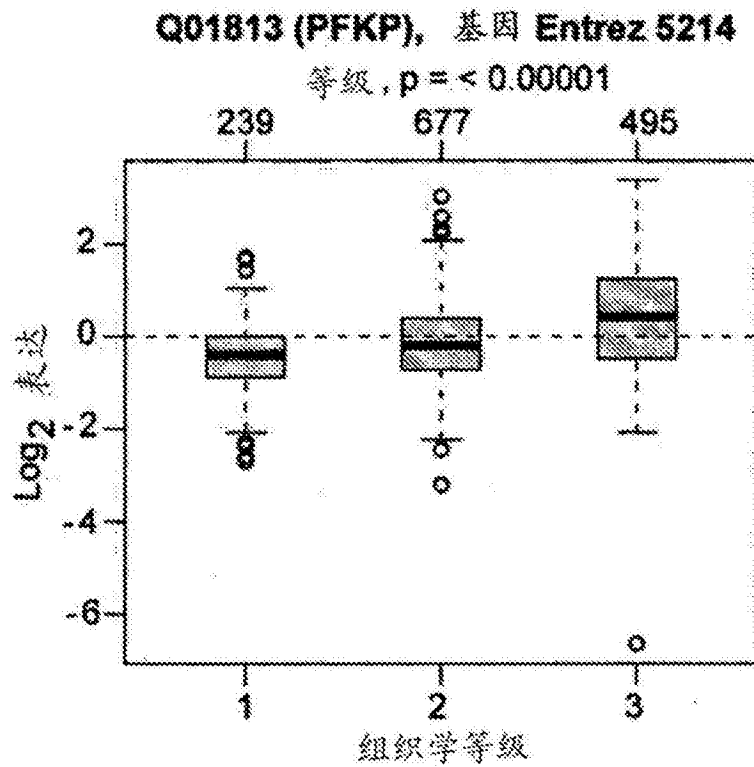


图16D

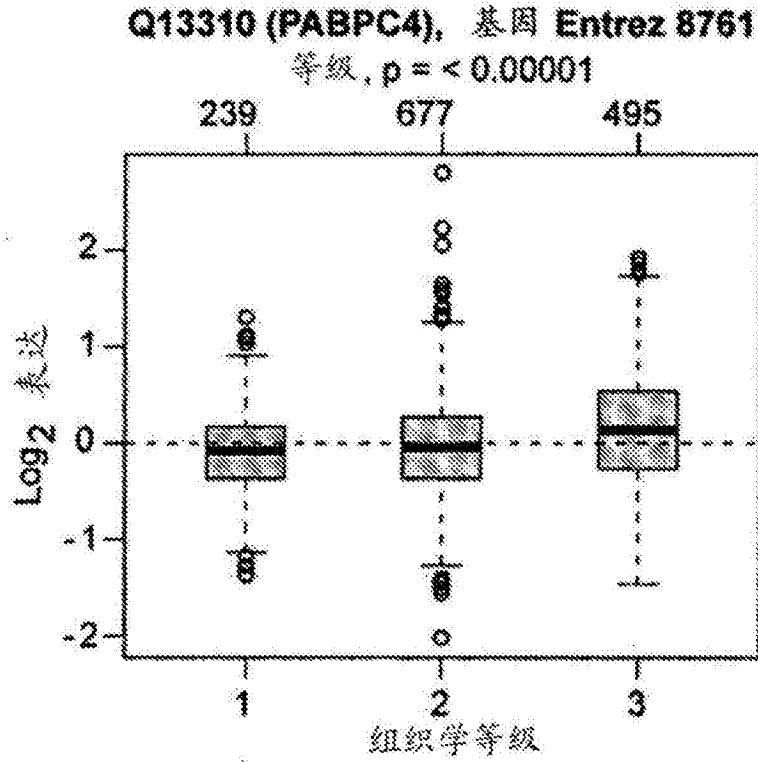


图16E

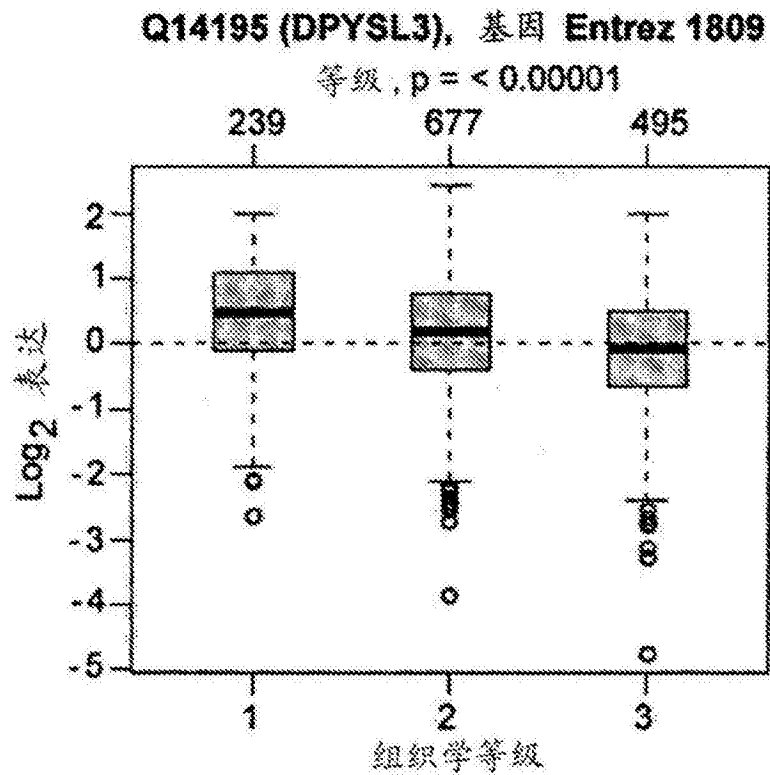


图16F

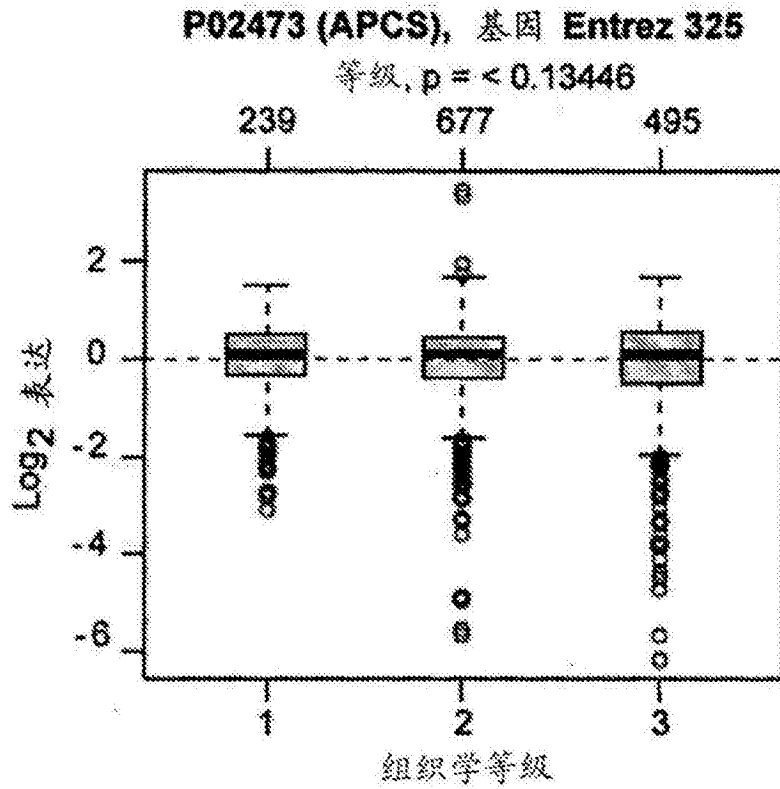


图16G

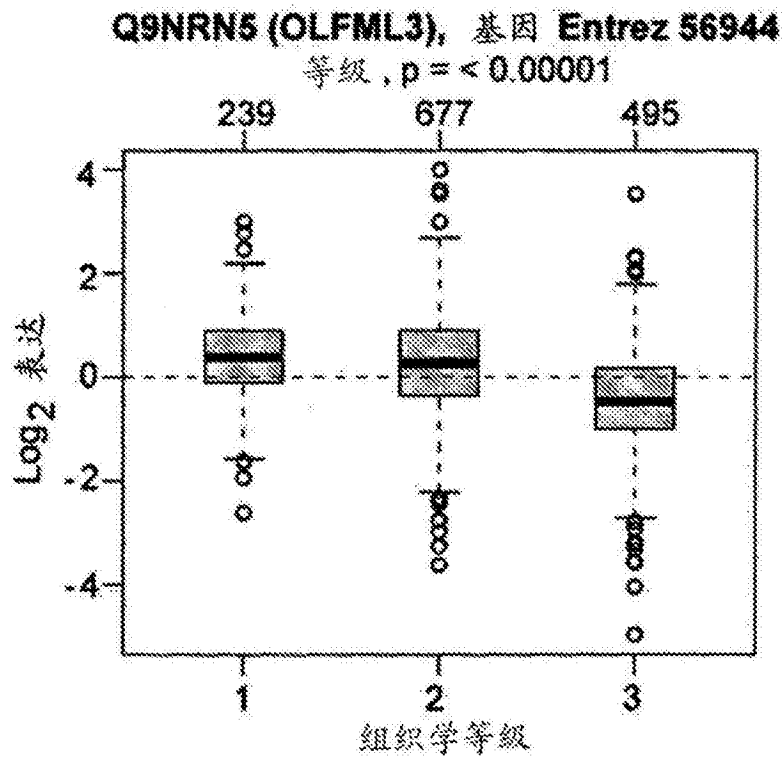


图16H

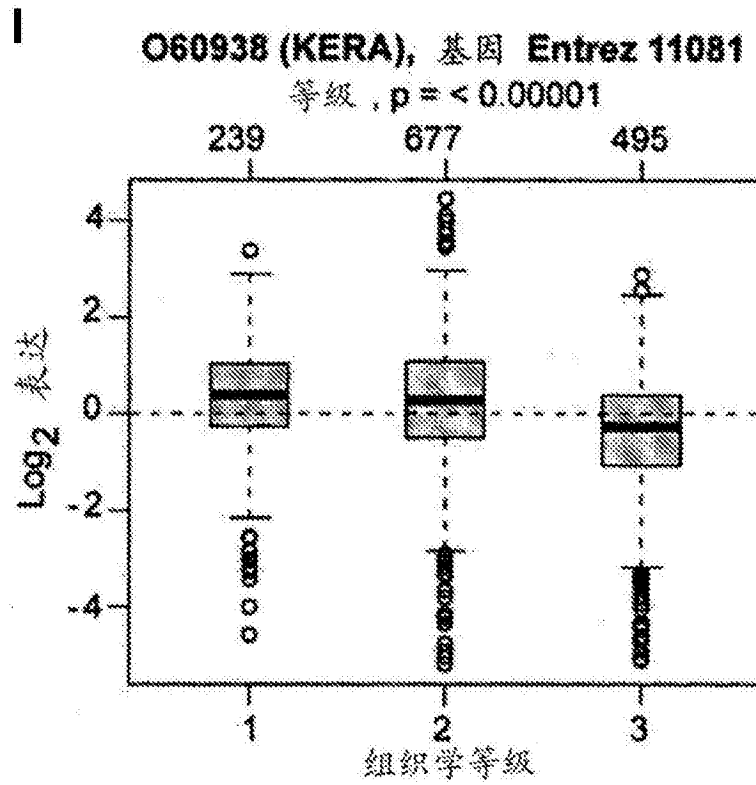


图16I

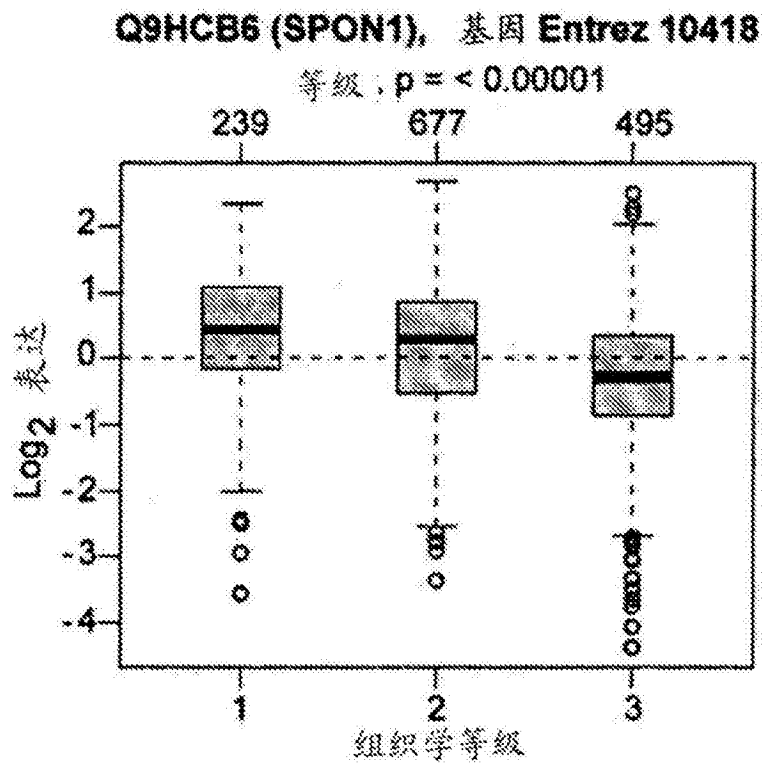


图16J

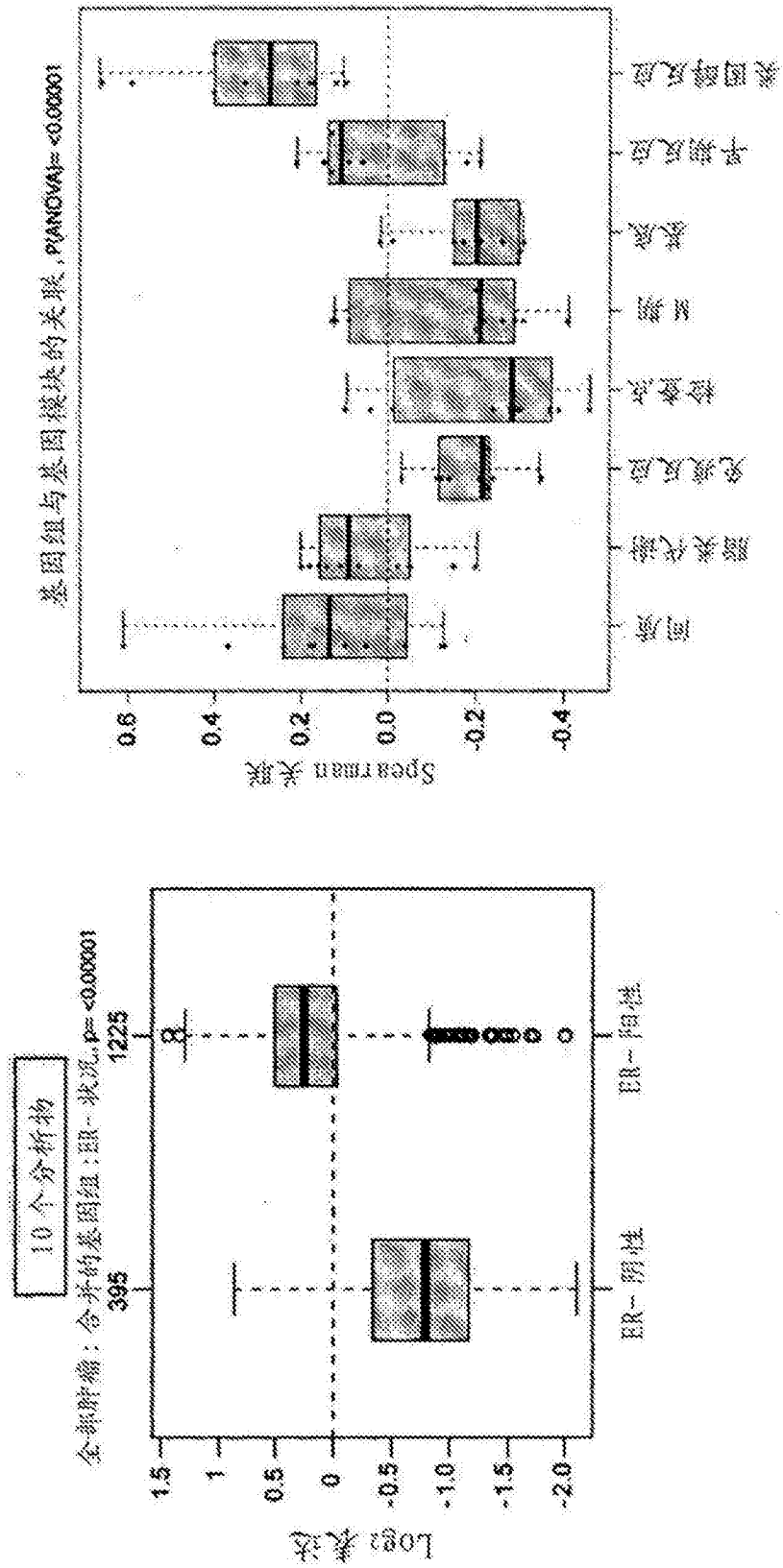


图17A

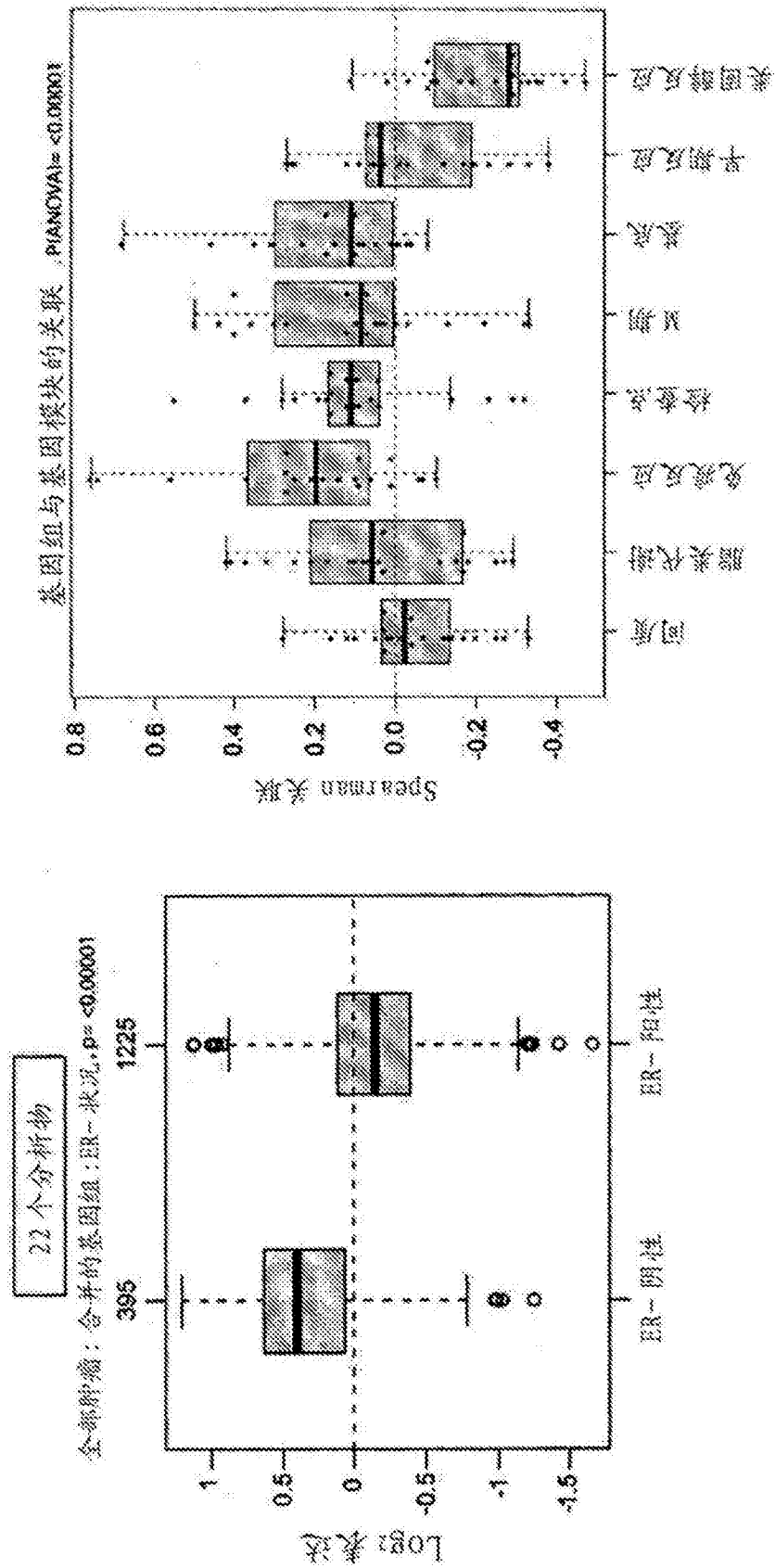


图17B

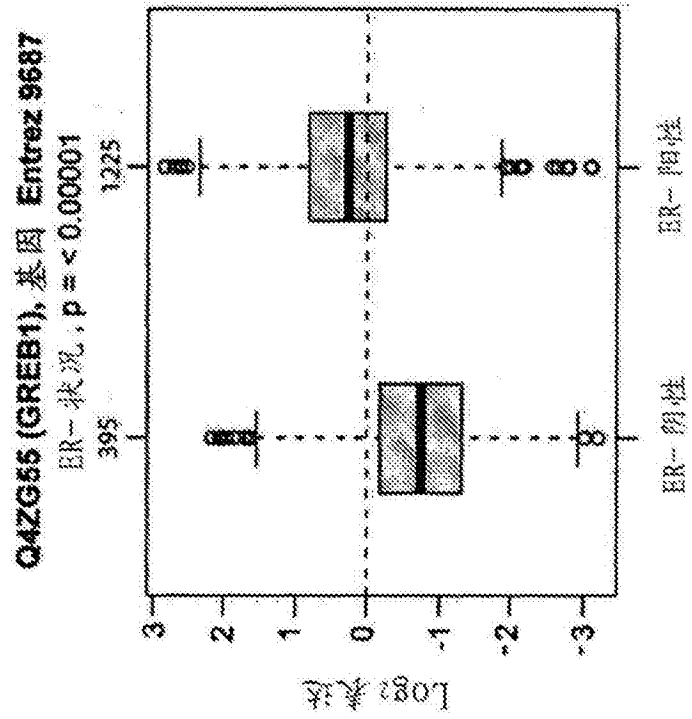


图17C

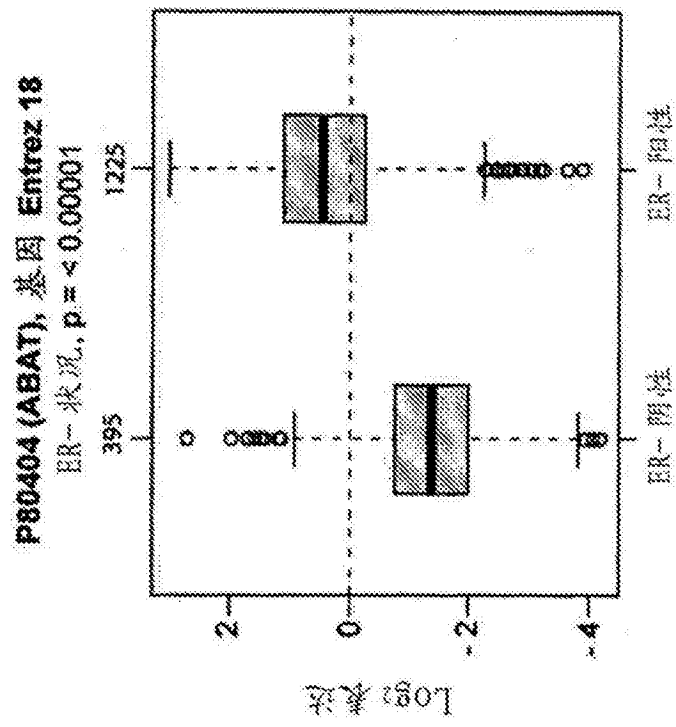


图17D



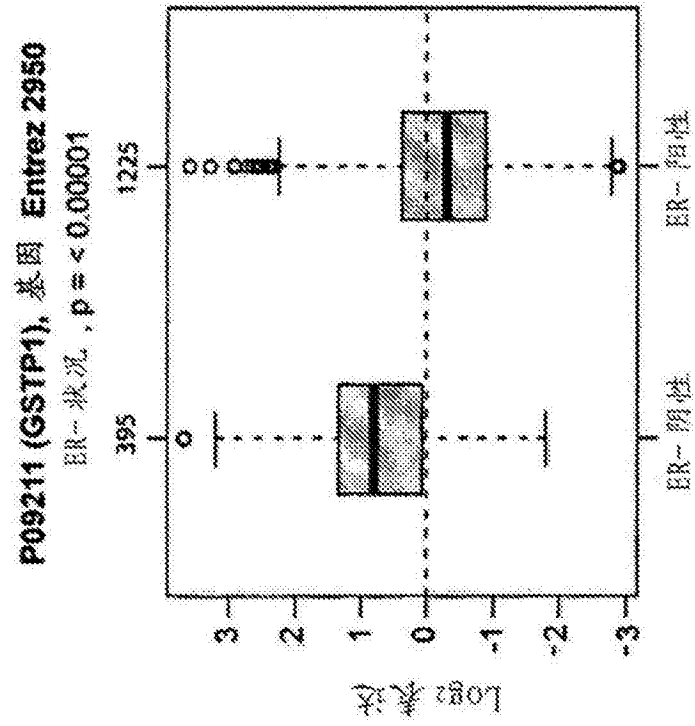


图17E

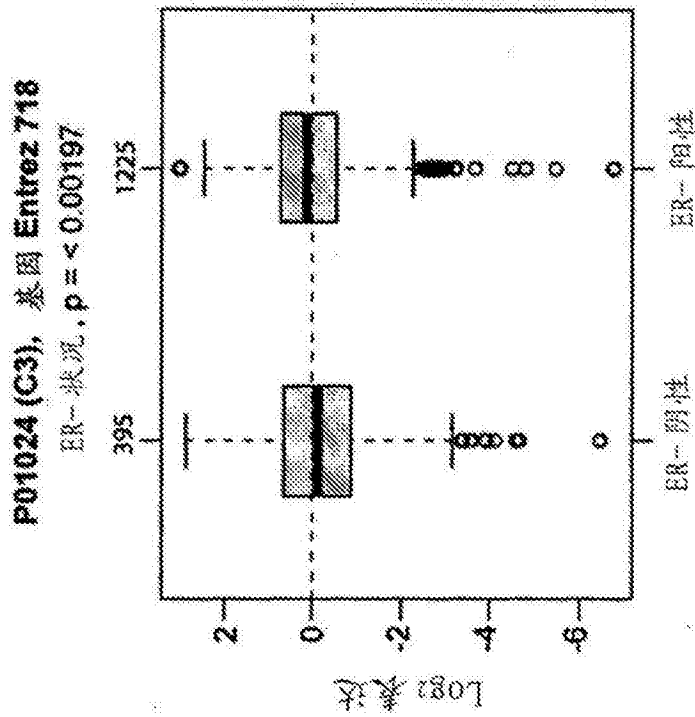
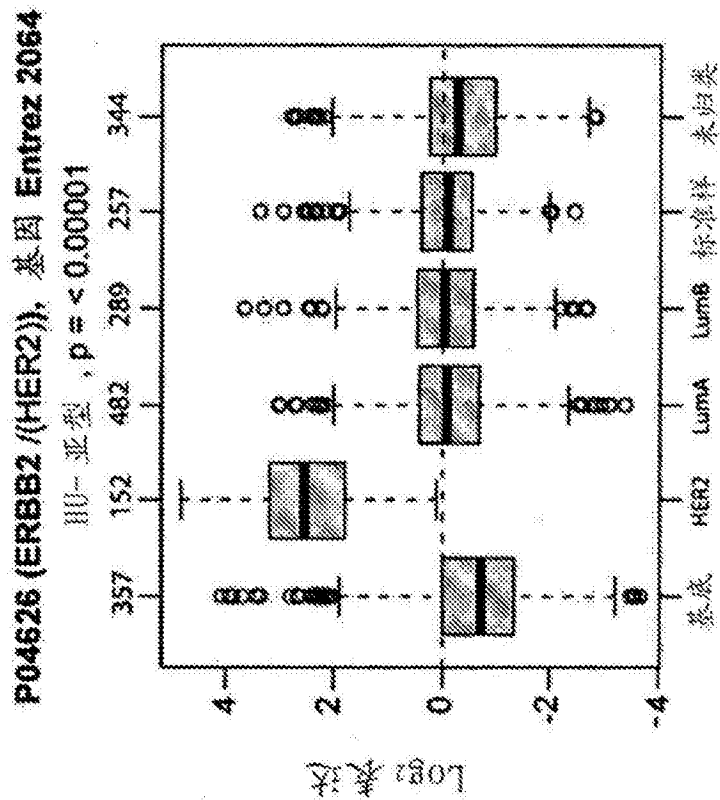


图17F



P22392 使用 Gene Entrez:  
4831 或 654364  
在基因表达数据库中不能作图

图18A

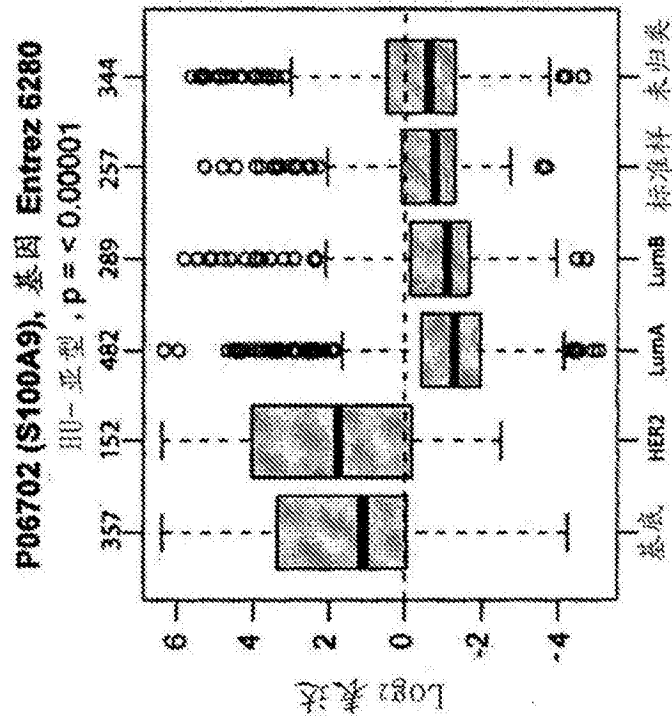


图18B

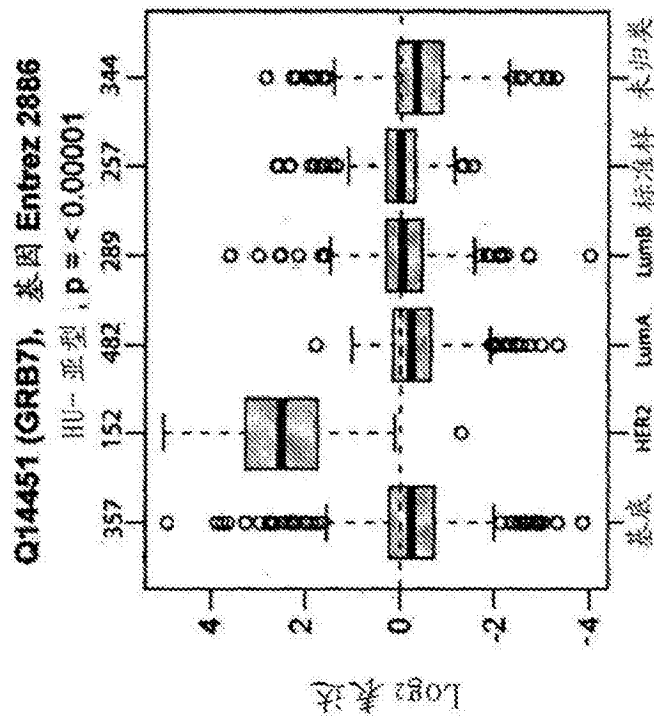


图18C

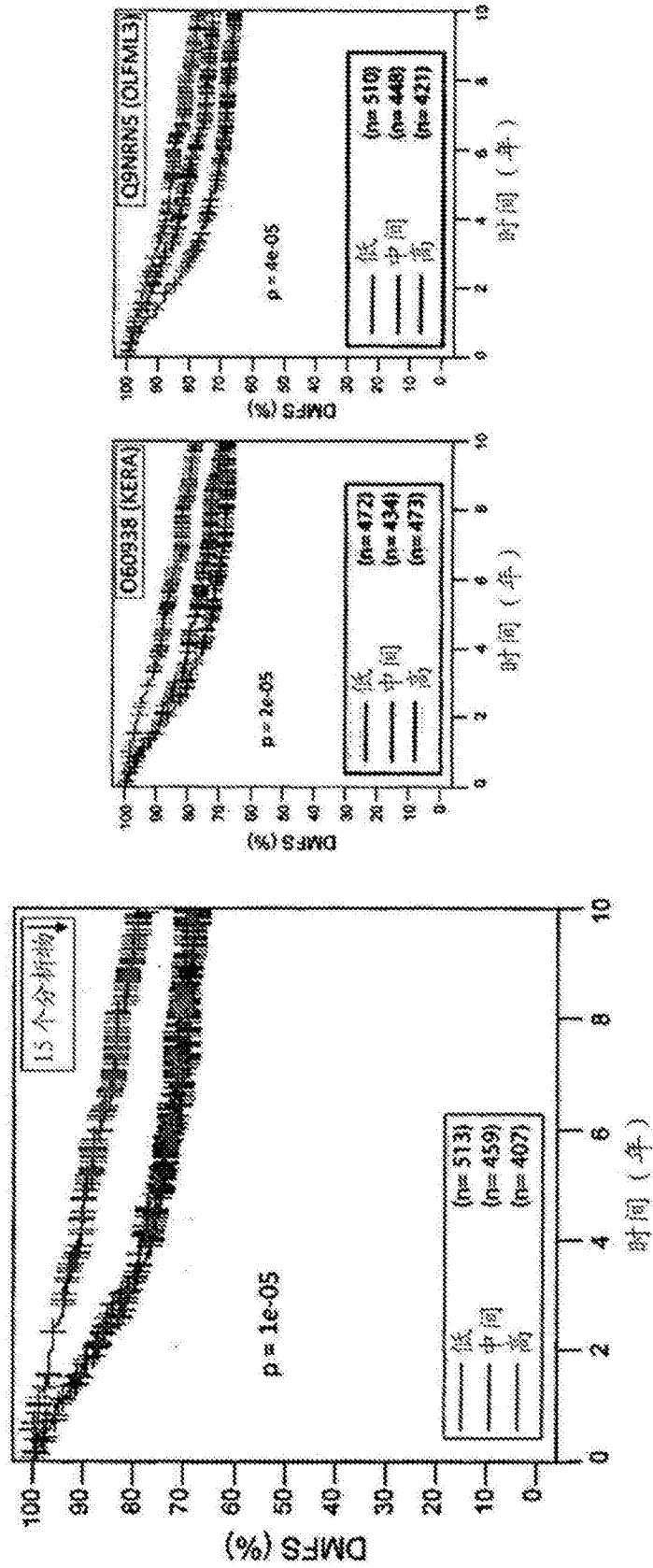


图19

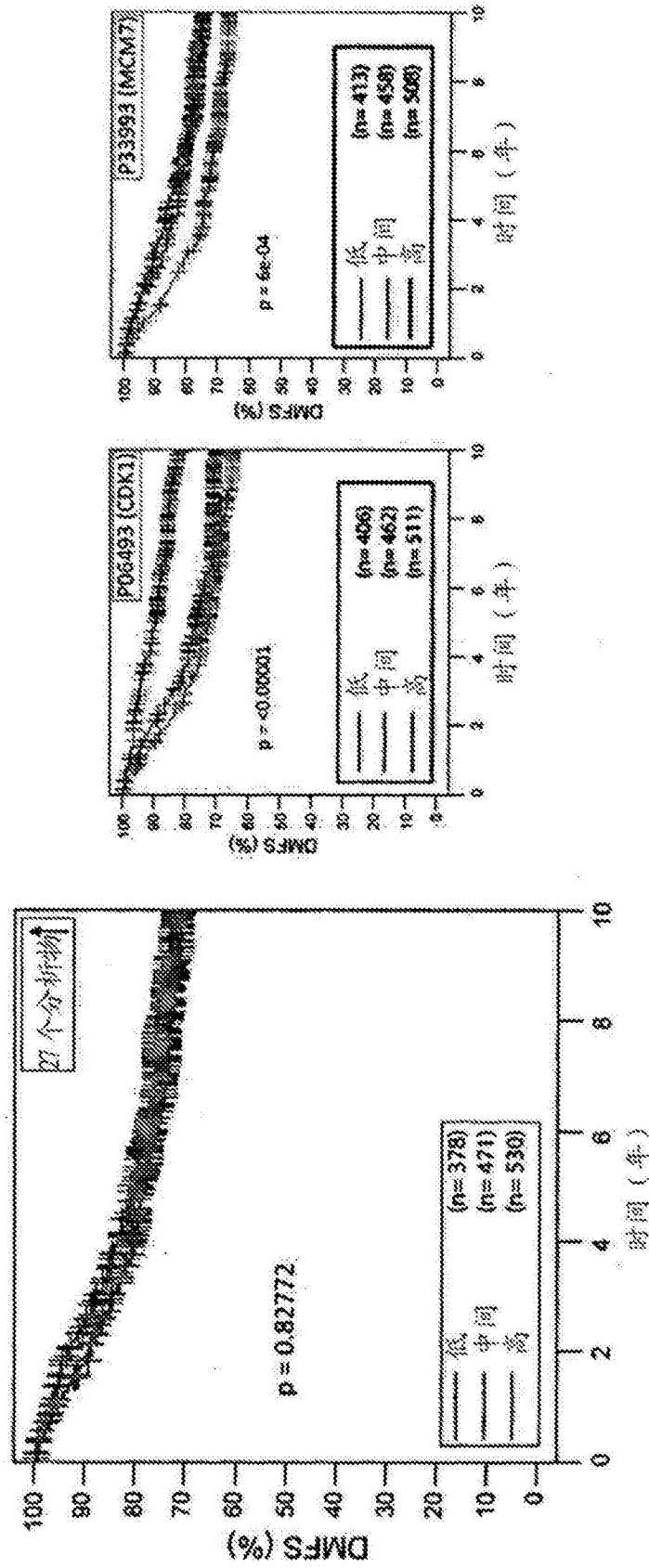


图19(续)

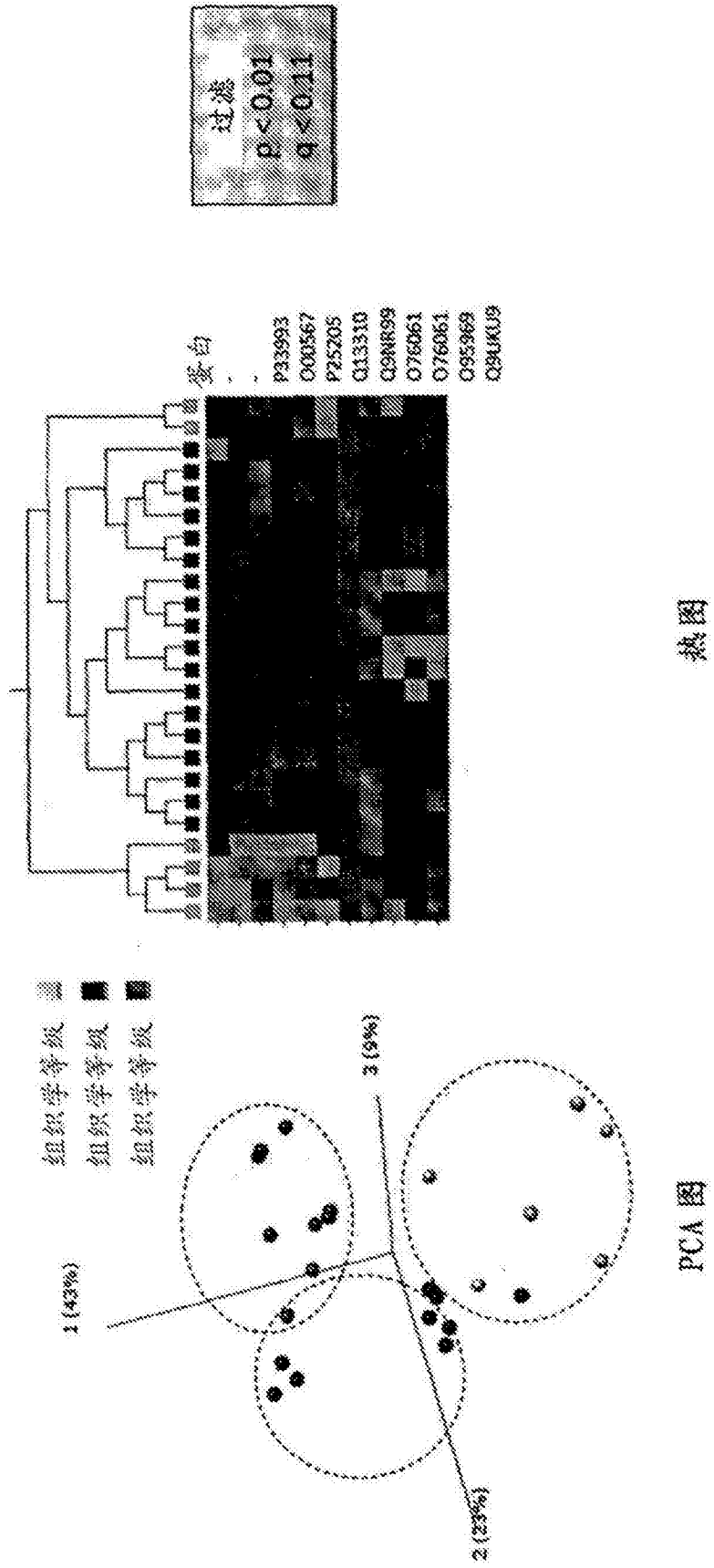


图20

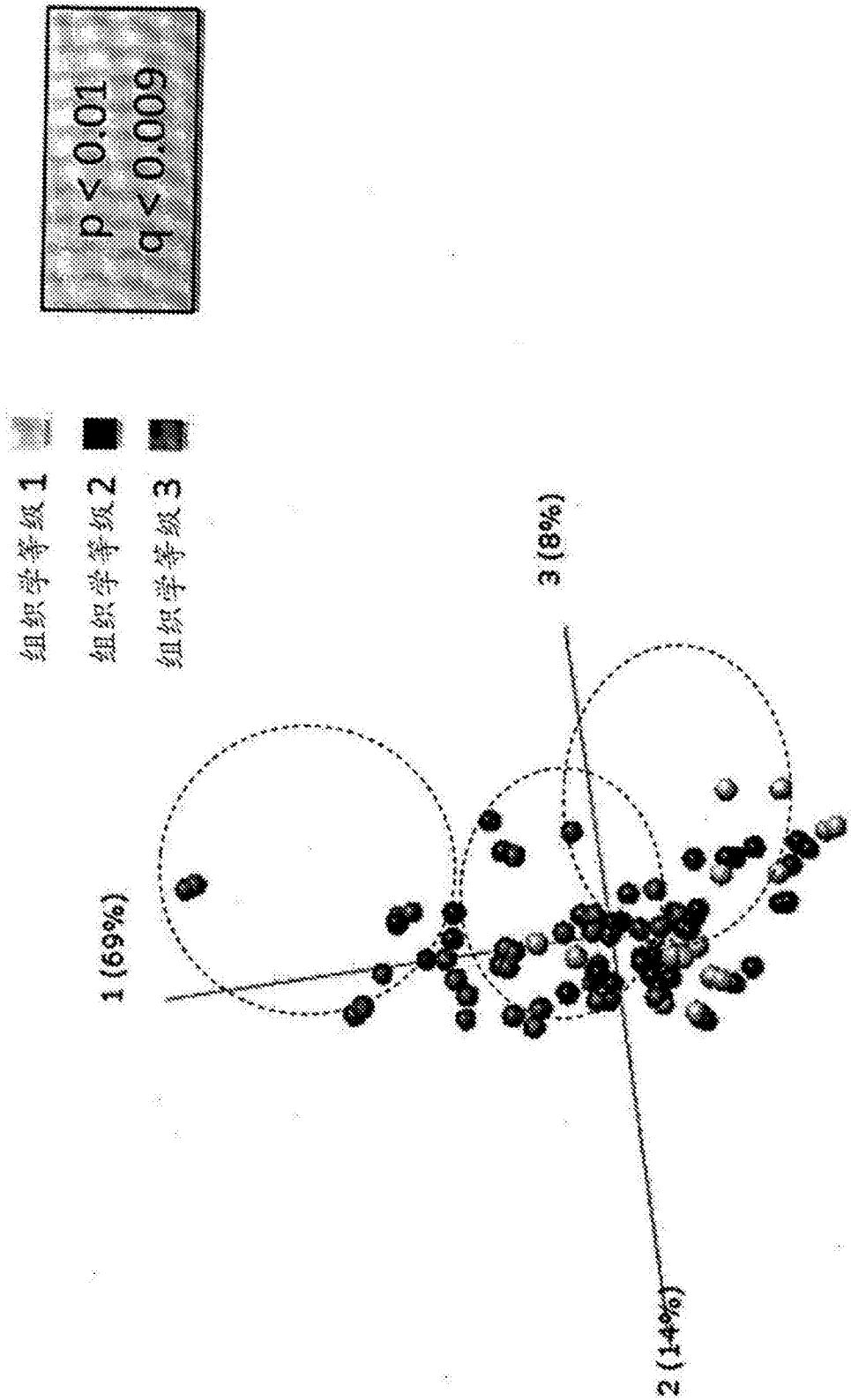


图21

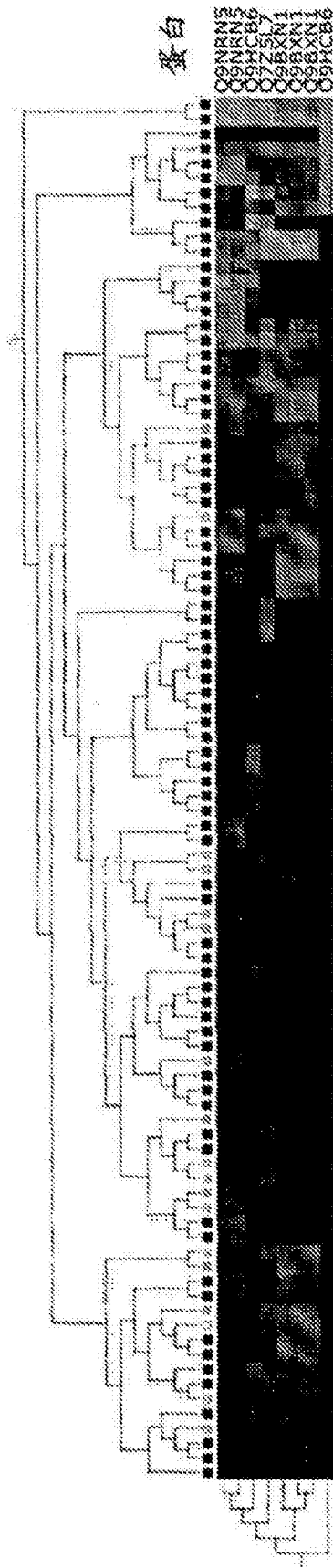


图21 (续)