



(12)发明专利申请

(10)申请公布号 CN 106461663 A

(43)申请公布日 2017.02.22

(21)申请号 201480073216.9

保罗·奥沙利文 萨蒂什·库马尔

(22)申请日 2014.11.20

(74)专利代理机构 北京市磐华律师事务所

11336

(30)优先权数据

代理人 董巍 谢梅

61/907,013 2013.11.21 US

(85)PCT国际申请进入国家阶段日

(51)Int.Cl.

2016.07.14

G01N 33/574(2006.01)

(86)PCT国际申请的申请数据

PCT/US2014/066678 2014.11.20

(87)PCT国际申请的公布数据

W02015/077485 EN 2015.05.28

(71)申请人 环太平洋有限公司

地址 新西兰达尼丁

(72)发明人 马克·达尔芬 大卫·达林

詹姆斯·米勒·萨蒂

莱莫尼斯·卡瓦列尔瑞斯

权利要求书2页 说明书66页

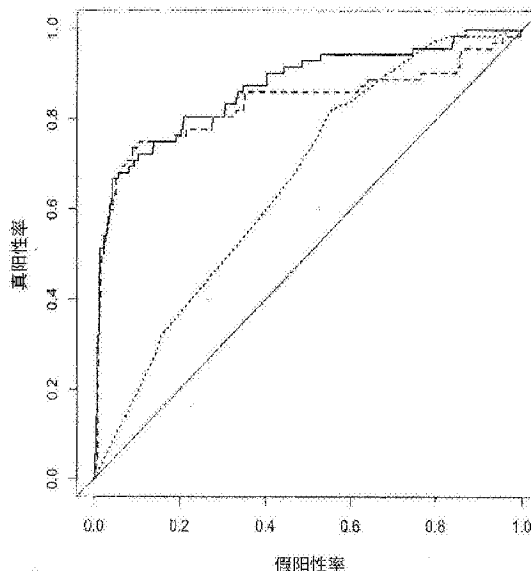
序列表4页 附图103页

(54)发明名称

使用基因型生物标志物和表型生物标志物
来归类无临床症状血尿患者

(57)摘要

用于识别低风险地具有尿路上皮癌(UC)的血尿患者的新方法,该方法包括将挑选的表型变量与基因型表达水平组合到新的指标,即“G+P指数”中。G+P指数将年龄、性别、吸烟史、血尿的存在和血尿的频率与遗传标志物,MDK、CDC2、HOXA13、IGFBP5和可选的IL8Rb的基因型表达相结合,然后确定由患者所获得的G+P指数值是在以下三组中的一组内:(1)UC高风险,(2)UC风险,或(3)UC低风险。



1. 一种用于确定呈现血尿的患者患有尿路上皮癌的风险水平的方法,其包括:
 - a. 提供来自所述患者的尿液样本;
 - b. 定量一个值,MI,其包括定量所述样本中人类MDK,CDC2,H0XA13和IGFBP5的表达水平;
 - c. 评估所述患者的表型变量HFREQ, AgeGT, 性别, SMK和RBC;
 - d. 根据公式 (i) 或公式 (ii) 计算G+P指数:
公式 (i), $G+P \text{ 指数} = (1 * HFREQ + 3 * Gender + 4 * SMK) + (5 * M1 + 2 * IL - 8)$, 或
公式 (ii), $G+P \text{ 指数} = (w1 * HFREQ + w2 * AgeGT50 + w3 * Gender + w4 * SMK + w5 * RBC) + (w6 * M1 + w7 * IL - 8)$, 或
 $G+P \text{ 指数} = -8.46 + 0.79IGF - 1.60H0XA + 2.10MDK + 0.95CDC - 0.38IL8R + 0.98SNS + 0.56Hfreq + 1.11Gender + 0.64Age$; 和
 - e. 确定G+P指数是否大于指示所述患者患有尿路上皮癌的风险水平的阈值。
2. 如权利要求1所述的方法,其中所述阈值是选自G+P指数值为0至5,或6至10,或11至15的组别,其中所述值为0至5表示低风险,6至10表示中等风险,并且11至15表示高风险。
3. 如权利要求2所述的方法,其中如果所述阈值是6至10的G+P指数值,那么所述患者经受额外的临床或实验室测试。
4. 如权利要求2所述的方法,其中如果所述阈值是11至15的G+P指数值,那么所述患者经受额外的临床或实验室测试。
5. 如权利要求2所述的方法,其中如果所述阈值是0至5的G+P指数值,那么所述患者被置于进一步临床和实验室测试的观察名单上。
6. 如权利要求1至5中任一项所述的方法,其中所述阈值是使用统计方法建立的。
7. 如权利要求1至6中任一项所述的方法,其中所述统计方法是线性判别分析 (LDA), 逻辑回归 (LogReg), 支持向量机 (SVM), K-5最近邻 (KN5N), 以及划分树分类器 (TREE) 中的任何一个。
8. 如权利要求1至7中任一项所述的方法,进一步包括定量从图6或图7中选择一个附加基因型标志物的表达。
9. 如权利要求1至8中任一项所述的方法,其中定量基因表达的所述步骤通过检测mRNA的水平进行。
10. 如权利要求1至8中任一项所述的方法,其中定量基因表达的所述步骤通过检测cDNA的水平进行。
11. 如权利要求1至8中任一项所述的方法,其中定量基因表达的所述步骤使用与所述cDNA的至少一部分互补的寡核苷酸进行。
12. 如权利要求1到9中任一项所述的方法,其中定量基因表达的所述步骤使用利用正向引物和反向引物的qRT-PCR法进行。
13. 如权利要求1至8中任一项所述的方法,其中定量基因表达的所述步骤通过检测蛋白的水平进行。
14. 如权利要求1至8中任一项所述的方法,其中定量基因表达的所述步骤通过检测肽的水平进行。
15. 如权利要求1至8或13至14中任一项所述的方法,其中定量基因表达的所述步骤使

用针对所述标志物的抗体进行。

16. 如权利要求1至8或13至15中任一项所述的方法,其中定量基因表达的所述步骤使用夹心型免疫测定方法,或使用抗体芯片进行。

17. 如权利要求1至8或13至16中任一项所述的方法,其中定量基因表达的所述步骤使用单克隆抗体进行。

18. 如权利要求1至8或13至17中任一项所述的方法,其中定量基因表达的所述步骤使用多克隆抗血清进行。

使用基因型生物标志物和表型生物标志物来归类无临床症状 血尿患者

[0001] 要求优先权

[0002] 本申请要求于2013年11月21日提交的、申请号为61/907,013的美国临时专利申请的优先权;发明人为David Darling、Satish Kumar、Mark Dalphin和Paul O'Sullivan。该临时申请通过引用完全并入本文中。

技术领域

[0003] 本发明涉及检测没有患有疾病的患者。具体地,本发明涉及使用遗传标志物(genetic marker)和表型标志物(phenotypic marker)来归类(triaging)患有血尿但没有癌症的患者。特别是,本发明涉及在归类肉眼可见血尿或显微可见血尿患者时对遗传标志物和表型标志物的分析。更特别地是,本发明涉及组合使用遗传标志物和表型标志物来归类无临床症状的肉眼可见或显微可见血尿患者以及预测患者的病情是否需要进一步的临床处理。

背景技术

[0004] 癌症的早期治疗大大提高癌症患者的存活率。在膀胱癌的情况下,相比于病变转移患者5年6%的存活率,诊断为局限于原发部位疾病的患者5年的存活率为73%(Altekruse等人)。因此,膀胱癌早期和准确诊断的发展可以导致对患者的改善的预后。为协助癌症的早期检测,大量癌症特异性标志物已经得到确认。然而使用这些标志物会导致患有炎性膀胱疾病,但没有膀胱癌的患者假阳性结果。

[0005] 无症状血尿(“AH”)是最常见的泌尿系统疾病的发现方式之一,根据人口数量其发病率在2%和30%之间(Schwartz G:Proper evaluation of asymptomatic microscopic hematuria in the era of evidence-based medicine-progress is being made.Mayo Clin Proc.2013,88(2);123-125;McDonald M,Swagerty D,Wetzel L:Assessment of Microscopic hematuria in adults.AFP 2006 73:10,Grossfield G,Wolf J,Litwan M,Hricak H,Shuler C,Agerter D等人.Asymptomatic microscopic hematuria in adults:summary of AUA best practice policy recommendations.AFP 2001:63:1145-54)。

[0006] 然而,AH表明泌尿系统恶性肿瘤发病率的病症在AH群体中的大范围为1.9-7%。对所有确诊AH患者进行的全面诊断工作对许多医疗保健系统是很大的负担。使用表型指标来分开高风险和低风险的患者已经在Loo等人一项最近的研究中被探索。(Loo R,Lieberman S,Slezak J,Landa H,Mariani A,Nicolaisen G,Aspera A and Jacobsen S:Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria.Mayo Clin Proc.2013,88(2);129-138)。

[0007] 上述研究中有4414例呈现确诊AH的患者,该研究表明其中73%的患者没有查明的原因,而26%的患者需要某种形式的泌尿科病情检查以查明原因。约2.5%呈现AH的患者被诊断为有尿路上皮肿瘤,其它疾病例如尿路感染(UTI)(2.3%),肾结石(16.2%),前列腺出

血(4%),和污染(0.4%)构成其他的诊断(Loo等人,同上)。

发明内容

[0008] 我们已经识别了该领域中的新问题,即如何辨别出无膀胱癌或有低膀胱癌风险但呈现血尿的患者。这解决了许多具有血尿但不是膀胱癌的患者可能经历并不需要的昂贵的和进一步侵入性的病情检查的问题。因此,本发明有用之处在于,当遗传信息和表型信息的组合提供辨别出无膀胱癌或有低膀胱癌风险的患者时,将这些个体从针对膀胱癌的全面病情检查的危险和费用中排除,并且本发明可以有效地将没患有癌症的患者从那些患有癌性疾病的患者,包括膀胱上皮癌,移行细胞癌(TCC)和非癌疾病的患者,包括炎症性疾病中归类出来。本发明提出了解决新问题的新方法,因为该方法对诊断非癌症,而不是诊断癌症是出乎意料地有用的。使用遗传标准和表型标准的组合出乎意料地提供了比单独的遗传变量或表型变量更好的辨别力。数据从541次观察中得出,这些观察来自于用于内部验证的、使用自举程序的在CLIA标准和CURT+North Shore下具有法律效力的产品试验。表型变量包括:(1)吸烟史、(2)血尿、(3)性别、和(4)年龄。遗传变量包括IGF、HOXA13、MDK、CDC、和IL8R的表达分析。遗传+表型模型(“G+P”)意外地表现得比单独的遗传变量或表型变量更好。

[0009] 肉眼可见血尿,或在尿液中发现可视觉识别的血液是膀胱癌患者中的普遍发现。对于那些患者,通常的标准做法是执行附加的诊断程序以诊断膀胱癌。然而,迅速地识别显微可见血尿患者并且理解显微可见血尿在尿路上皮癌方面的含义仍然是一个问题。

[0010] 在本文中,我们提供了改进的方法,用于确定如果呈现了肉眼可见血尿或显微可见血尿的患者患有膀胱癌的概率足够低,以保证不用进行额外的程序,那么这样的患者能否避免侵入性的和昂贵的进一步临床程序以检测尿路上皮癌(UC),包括膀胱癌。

[0011] 描述了归属于高概率尿路上皮癌(UC)的因素。除了环境因素例如吸烟史和职业暴露于芳香胺类之外,人口因素例如性别、种族和年龄明显有助于发展为UC的风险。基于这些因素进行医疗评估的患者的特性在特定的基础上常规使用。例如,广为接受的是呈现血尿的60岁有吸烟史的男性比呈现相同症状的35岁不吸烟的女性具有更高的UC阳性概率,然而尚未对这些差异对于患者患有UC的总概率的贡献进行定量。不同的基因型和表型因素赋予具体的权重并将这些与诊断测试结果组合可以显著增加非侵入性测试诊断能力的精度,并且在由临床和生物标志物的测试结果所限定的患者患有UC的概率基础上提供给临床医生更大的确定性来区分患者。

[0012] 尽管有一些可用于检测膀胱癌存在的方法,但是没有可靠和准确的方法以确定患者是否没有膀胱癌或患有膀胱癌的风险较低。为了解决这一需求,我们开发了新的分析方法,用于在呈现血尿(无论是肉眼可见血尿或显微可见血尿)患者中将癌症疾病和非癌疾病区别开。在本发明的一些方面中,我们将定量表型变量和定量遗传标志物表达进行组合,从而形成联合分离指数(“G+P指数”)以有效地从那些具有高概率UC的AH患者中归类出来具有低概率UC的AH患者。这种分离将那些不需要完整的泌尿病情检查的患者从那些确实需要完整病情检查的患者中做了界定,从而避免了对低概率UC患者进行不必要的病情检查。

[0013] 表型测定

[0014] 在G+P指数中被评价的表型变量包括血尿的频率(HFREQ)、年龄、性别、吸烟史、和红细胞计数(RBC)。这些术语在下文中进行定义。表型变量在本说明书中定义为包括临床发

现和观察。

[0015] 基因型测定

[0016] 通常,由Pacific Edge Ltd.开发的优选基因型测定包括遗传标志物CDC2、HOXA13、MDK和IGFBP5(“4标志物测定”)的表达的定量。在另一优选测定中,所述4个标志物和第5个标志物IL8R被定量(“5标志物”或Cxladder®测定;新西兰达尼丁的Pacific Edge Ltd.的商标)(Holyoake A, O’Sullivan P, Pollock R等人:Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder.Clin Cancer Res 2008;14:742,以及O’Sullivan P, Sharples K, Dalphin M等人:A Multigene Urine Test for the Detection and Stratification of Bladder Cancer in Patients Presenting with Hematuria.J Urol 2012, Vol.188 No 3;746),以及国际专利申请号PCT/NZ2011/000238,标题为“Novel Markers for Detection of Bladder Cancer”。这些出版物和专利申请中的每一个都通过引用完全包含在本说明书中,如同分别地通过引用包含在本说明书中。

[0017] 在优选的实施例中,4标志物测定可以在使用PCR扩增的未分级尿液上实施来定量4个mRNA标志物(CDC2、HOXA13、MDK和IGFBP5),其在尿路上皮癌中被过度表达。IL8R在嗜中性粒细胞被高度地过度表达,因此在非恶性炎症病症下被升高。包含该第5个mRNA标志物显著地降低了移行细胞癌(TCC)的假阳性检测的风险。从患者的角度来看,该测试是非侵入性的并且非常简单。取单一的尿液样本,往往是中段尿液但不排他,这往往在家中而不用进入诊所就可以完成。

[0018] Cxladder®测定在呈现肉眼可见血尿患者中已表现为比细胞学更加敏感。最值得注意的是,Cxladder®测定对所有大于Ta阶段的尿路上皮癌达到100%(预先指定特异性为85%)的灵敏度,对所有高等级肿瘤的灵敏度为97%。Cxladder®测定归于单一分值,该分值将在患者尿液中体现的5个基因的定量基因表达结合起来。基于患者患有尿路上皮癌的概率,该分值将患者分为三类。

[0019] 对于呈现血尿的患者(无论是肉眼可见血尿或显微可见血尿),本发明已经被证明通过加入在同一时间段内从患者收集的表型变量加强了这些基因型工具(无论是4标志物测定或Cxladder®测定),并且将这些组合成新的工具,可用于相对于患者患有尿路上皮癌(“UC”)的概率,将患者分开到三个定义的风险级别的指数。

[0020] 本发明的各方面

[0021] 本发明的各方面说明如下。可以理解的是,这些不是本发明的仅有的方面或实施例。普通技术人员可以将一个或多个方面结合在一起以产生另外的方面或实施例。

[0022] 一个方面包括用于确定呈现血尿的患者,或者患有尿路上皮癌的风险级别的方法,其包括:

[0023] 提供所述患者的尿液样本;

[0024] 定量一个值MI,包括定量在所述样本中的人类MDK、CDC2、HOXA13和IGFBP5的表达级别;

[0025] 评估所述患者的表型变量HFREQ、AgeGT、sex、SMK和RBC;

[0026] 根据下述公式中的任一个计算G+P指数:

[0027] 公式(i), G+P指数 = $(1*HFREQ+3*Gender+4*SMK) + (5*M1+2*IL-8)$,

[0028] 或者

[0029] 公式(ii), G+P指数 = $(w1*HFREQ+w2*AgeGT50+w3*Gender+w4*SMK+w5*RBC) + (w6*M1+w7*IL-8)$,

[0030] 或者

[0031] $G+P指数 = -8.46 + 0.79IGF - 1.60HOXA + 2.10MDK + 0.95CDC - 0.38IL8R + 0.98SNS + 0.56Hfreq + 1.11Gender + 0.64Age$; 和

[0032] 确定G+P指数是否大于指示所述患者患有尿路上皮癌的风险级别的阈值。

[0033] 另外的方面包括另一个方面的方法, 其中所述阈值选自G+P指数值为0至5、6至10、或11至15的组别, 其中所述值0到5指示低风险, 6至10指示中等风险, 并且11至15指示高风险。

[0034] 其它方面包括任何其它方面的方法, 其中, 如果所述阈值是G+P指数值为6至10, 那么所述患者要受到额外的临床测试或实验室测试。

[0035] 仍然进一步的方面包括任何现有方面的方法, 其中, 如果所述阈值是G+P指数值为11至15, 那么所述患者要受到额外的临床测试或实验室测试。

[0036] 仍然进一步的方面包括任何其它方面的方法, 其中, 如果所述阈值是G+P指数值为0至5, 那么所述患者被放在进一步临床测试或实验室测试的观察名单上。

[0037] 另外的方面包括任何其它方面的方法, 其中所述阈值是用统计方法建立的。

[0038] 仍然进一步的方面包括任何其它方面的方法, 其中所述统计方法是线性判别分析(LDA)、逻辑回归(LogReg)、支持向量机(SVM)、K-5最近邻(KN5N)和划分树分类器(TREE)中的任一项。

[0039] 另外的方面包括任何其它方面的方法, 还包括定量选自图6或图7的一个额外的基因型标志物的表达。

[0040] 其它方面包括任何先前方面的方法, 其中, 定量基因表达的所述步骤通过检测mRNA水平来进行。

[0041] 其它方面包括任何其它方面的方法, 其中, 定量基因表达的所述步骤通过检测cDNA水平来进行。

[0042] 仍然进一步的方面包括任何其它方面的方法, 其中, 定量基因表达的所述步骤使用与所述cDNA的至少一部分互补的寡核苷酸来进行。

[0043] 另外的方面包括任何其它方面的方法, 其中, 定量基因表达的所述步骤使用利用正向引物和反向引物的qRT-PCR法来进行。

[0044] 仍然另外的方面包括任何其它方面的方法, 其中, 定量基因表达的所述步骤通过检测蛋白水平来进行。

[0045] 仍然其它方面包括任何其它方面的方法, 其中, 定量基因表达的所述步骤通过检测肽的水平进行的。

[0046] 另外的方面包括任何其它方面的任一个的方法, 其中, 定量基因表达的所述步骤使用针对所述标志物的抗体来进行。

[0047] 仍然进一步另外的方面包括权利要求1至8或13至15的任一项的方法, 其中, 定量基因表达的所述步骤使用夹心型免疫测定方法, 或使用抗体芯片来进行。

[0048] 仍然进一步的方面包括任何其它方面的任一个的方法,其中,所述定量基因表达使用单克隆抗体来进行。

[0049] 其他方面包括任何其他方面的任一个的方法,其中所述定量基因表达使用多克隆抗血清来进行。

[0050] G+P指数

[0051] 上述表型变量和基因型变量根据以下关系组合成G+P指数:

[0052] $G+P \text{ 指数} = (w1*HFREQ + w2*AgeGT50 + w3*Gender + w4*SMK + w5*RBC) + (w6*M1 + w7*IL-8)$,

[0053] 其中, HFREQ指的是在6个月周期内每个高倍视野下找到3个或更多红细胞的频率;如果频率低则HFREQ=0,并且如果每个高倍视野下超过3个红细胞则HFREQ=1。AgeGT50指受试者的年龄,如果大于50岁则AgeGT50=1,如果不到50岁则AgeGT50=0。性别男则Gender=1,性别女则Gender=0。SMK指的是受试者是否当前或以前是吸烟者;如果是非吸烟者则SMK=0,如果是吸烟者,则SMK=1。RBC指红细胞计数;如果为25或更多则RBC被设置为1,如果小于25,则RBC为0。M1是遗传标志物MDK、CDC、IGFBP5和HOXA13的表达的组合;如果M1大于4.5,则将其设置为1,如果小于4.5,则M1为0。IL-8指的是IL-8的RNA的表达水平;如果IL-8>2.5则IL-8被设置为1,如果小于2.5,则IL-8为0。符号“*”表示乘法运算符,加权因子w1-w7分别是分配给上面在G+P指数中列出的每个变量的权重。

[0054] 在其它优选的实施例中,(AgeGT50和RBC)可以从模型中去掉,如下所示:

[0055] $G+P \text{ 指数} = (1*HFREQ + 3*Gender + 4*SMK) + (5*M1 + 2*IL-8)$

[0056] G+P指数的计算值在0至15之间。G+P指数值为11至15的患者被认为是膀胱癌“高风险”,并且指示了额外的膀胱癌病情检查的需求。G+P指数值为6至10的患者被认为是“中等风险”发展为膀胱癌,并且指示了需要额外的病情检查。G+P指数值为0至5的患者被认为是“低风险”发展为膀胱癌。“低风险”组的病人被放在观察等待名单上,如果额外的症状出现,或者显微可见血尿反复发生,则他们为可能的进一步病情检查被重新评估。

[0057] 如在示例3(图18和图19)中更充分地说明的,我们发现,单独的定量表型变量的ROC曲线生成适度水平的诊断能力。单独的定量基因型标志物的ROC曲线生成显著水平的诊断能力。当基因型变量和表型变量组合成G+P指数时,我们发现了出乎意料地更好的诊断能力。

[0058] 定量基因表达

[0059] 蛋白质或核酸(或者单独地或者彼此组合地)由细胞分泌或来自细胞分裂,或者通过细胞凋亡机制丧失,其具有作为血清或体液标志物来诊断疾病(包括膀胱的炎性疾病和/或膀胱癌)的功用,或者作为标志物监测确诊疾病的进展。蛋白质和细胞标志物的检测可以使用本领域中已知的方法来进行,并且包括使用RT-PCT, qRT-PCR, 单克隆抗体, 多克隆抗血清等。

[0060] 具体地,本发明提供了用于归类呈现血尿(无论是肉眼可见血尿或显微可见血尿)患者的方法,包括:(i) 提供生物样本;(ii) 检测所述样本中的一种或多种膀胱肿瘤标志物(BTM)。特别关注的膀胱肿瘤标志物包括MDK、CDC2、HOXA13和IGFBP5(“4标志物测定”)。可选地,也可以检测所述样本中的人类嗜中性粒细胞标志物白细胞介素8受体(IL8Rb)的水平(Cxbladder[®]测定)。癌症的存在可以通过与正常患者、患有早期膀胱癌的患者、和/或患有

炎症疾病的患者比较所述一个或多个BTM的水平来确认。例如,癌症的存在可通过对比BTM的表达和表达阈值来确认。阈值可以大约是表达水平至少是没患有癌症患者组的表达水平的1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、3、4、5、6、7、8、9或10、100、1000或高达10,000倍。在其它方面,IL8Rb的高表达,但没有膀胱肿瘤标志物的改变的表达可以指示炎症疾病,而不是癌症。

[0061] 本发明的方法可以与用于检测膀胱癌的任何合适的标志物联合使用。用于本发明的合适的标志物的示例在图6或图7中概述。本发明包括使用在图6或图7中概述的标志物中的任何一个或多个。

[0062] 可选地,在其它优选的实施例中,本发明可以包括IL8Rb与标志物MDK、CDC2、HOXA13和IGFBP5的一个或多个的任意组合,该组合还可以与适合于检测膀胱癌的一个或多个其他标志物(例如,在图6或图7中概述的任何一个或多个标志物)组合。尤其是,本发明包括标志物的任何一种或多种组合的表达的定量,这些组合为:IL8Rb/MDK,IL8Rb/CDC2,IL8Rb/HOXA13,IL8Rb/IGFBP5,IL8Rb/MDK/CDC2,IL8Rb/MDK/HOXA13,IL8Rb/MDK/IGFBP5,IL8Rb/CDC2/HOXA13,IL8Rb/CDC2/IGFBP5,IL8Rb/HOXA13/IGFBP5,IL8Rb/MDK/CDC2/HOXA13,IL8Rb/MDK/CDC2/IGFBP5,IL8Rb/CDC2/HOXA13/IGFBP5、和IL8Rb/MDK/CDC2/HOXA13/IGFBP5。这些组合可选地包括适合于检测膀胱癌的一个或多个其他标志物,例如在图6或图7中概述的标志物中的任何一个或多个。

[0063] 本发明还提供了用于检测膀胱的炎症病症的方法,包括:(i)提供来自患者的生物样本;和(ii)检测所述样本中的人类嗜中性粒细胞标志物白细胞介素8受体(IL8Rb)的水平。膀胱的炎症病症的存在通过与正常患者、呈现血尿患者和患有炎症膀胱病症的患者的IL8Rb的水平进行比较来确认。例如,膀胱的炎症病症的存在可通过比较标志物IL8Rb的表达和阈值来确认。阈值可以大约是表达水平至少是另一组患者的表达水平的1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、3、4、5、6、7、8、9、10、100、1000或高达10,000倍。

[0064] 本发明的优选的基因型方法可通过检测基因表达的任何合适的标志物来进行,例如利用任何合适的方法确定mRNA、cDNA、蛋白质或肽的水平。

[0065] 诊断的确认可以通过使用分类器系统来建立,例如线性判别分析(LDA)、逻辑回归(LogReg)、支持向量机(SVM)、K-5最近邻(KN5N),和划分树分类器(TREE)。

附图说明

[0066] 本发明是参考其具体实施例,并且参照附图进行说明的,其中:

[0067] 图1描绘了IL8Rb的蛋白和mRNA序列(也称为CXCR2)。

[0068] 图2描绘了散点图,其示出IL8Rb对从非恶性疾病(前列腺疾病,膀胱炎,尿路感染和尿石症)中分开TCC的效果。IL8Rb已替代了图2c和图2f中的不同的膀胱癌RNA标志物。(a) MDK/IGFBP5; (b) MDK/HOXA13; (c) MDK/IL8Rb; (d) CDC2/IGFBP5; (e) CDC2/HOXA13; (f) CDC2/IL8Rb。

[0069] 图3描绘了ROC曲线分析(灵敏度VS特异性),其示出将IL8Rb包括在使用线性判别分析(LD)和线性回归(LR)导出的诊断算法中的效果。ROC曲线是从患有TCC和上尿路癌症的患者(n=61),和患有非恶性疾病膀胱炎、尿路感染和尿路结石的患者(n=61)导出的。(a) LD1(实线)和LD2(虚线)。(b) LR1(实线)和LR2(虚线)。IL8Rb被包括在LD2和LR2中。

[0070] 图4描绘了延长的ROC曲线分析,其示出将IL8Rb包括在使用线性判别分析(LD)和线性回归(LR)导出的诊断算法中的效果。该ROC曲线是从TCC患者(n=56)和,不同于图3,队列中的任何非恶性疾病患者(n=386)导出的。(a) LD1(实线)和LD2(虚线)。(b) LR1(实线)和LR2(虚线)。IL8Rb被包括在LD2和LR2中。

[0071] 图5描绘了箱形图,其示出非恶性泌尿疾病患者的尿中IL8Rb mRNA的累积。RNA已经通过使用delta-Ct方法的qRT-PCR被定量(Holyoake等人,2008)。用此方法,较低的Ct反映较高的RNA水平。BPH:良性前列腺增生症;UTI:尿路感染;NS prostate:非特异性前列腺疾病;Vasc.Prostate:前列腺血管;warfarin:血尿继发于华法林的使用。对膀胱炎/UTI患者的观察是与所示的其它非恶性情况显著不同的($p=0.001$)。

[0072] 图6描绘了已知在膀胱癌中被过度表达的标志物,并且该标志物是适合于在本发明中使用的。

[0073] 图7描绘了已知在膀胱癌中被低估表达的标志物,并且该标志物是适合于在本发明中使用的。

[0074] 图8描绘了实施例2的患者招募程序和数量的流程图。

[0075] 图9描绘了患病3个月患者的基线临床和人口统计特征。

[0076] 图10描绘了尿液测试的总体灵敏度和特异性。

[0077] 图11描绘了各种ROC曲线;图11a描绘了NMP22 ELISA和uRNA-D的ROC曲线(测试包括四个标志物MDK+CDC2+IGFBP5+HOXA13);图11b描绘了五个标志物MDK、CDC2、HOXA13、IGFBP5和IL8Rb的ROC曲线。

[0078] 图12描绘了尿检对阶段、等级、肿瘤部位、肿瘤的多重性、血尿状态、尿液样本的肌酸酐和性别的敏感度。表格显示了TCC患者中尿检呈阳性的数目和百分比。

[0079] 图13描绘了尿检对诊断、肉眼可见血尿或肌酸酐和性别的特异性。表格显示了无TCC患者中尿检呈阴性的数目和百分比。

[0080] 图14描绘了采用五种不同的分类器模型(i)线性判别分析(LDA),(ii)逻辑回归(LogReg),(iii)支持向量机(SVM),(iv)K-5最近邻(KN5N),以及(v)划分树分类器(TREE)的下述标志物组合的ROC曲线:(a)MDK、(b)CDC、(c)IGFBP5、(d)HOXA13、(e)MDK+CDC2、(f)MDK+IGFBP5、(g)MDK+HOXA13、(h)CDC2+IGFBP5、(i)CDC+HOXA13、(j)IGF+HOXA13、(k)MDK+CDC2+IGFBP5、(l)MDK+CDC2+HOXA13、(m)MDK+IGFBP5+HOXA13、(n)CDC2+IGFBP5+HOXA13、(o)MDK+CDC2+IGFBP5+HOXA13,加上或减去IL8Rb。

[0081] 图15描绘了灵敏度选择性研究的结果。图15a描绘了图14中ROC曲线的高达20%的假阳性率(80%的特异性)的“曲线下面积”(AUC),且图15b示出了源自于加入IL8Rb的AUC的差异。

[0082] 图16描绘了采用五种不同的分类器模型(i)线性判别分析(LDA),(ii)逻辑回归(LogReg),(iii)支持向量机(SVM),(iv)K-5最近邻(KN5N),以及(v)划分树分类器(TREE)、不同组特异性(a)80%,(b)85%,(c)90%,(d)95%,(e)98%的四个标志物MDK、CDC2、IGFBP5和HOXA13,加上或减去IL8Rb的組合的灵敏度的图表。

[0083] 图17描绘了在不同组特异性(a)80%,(b)85%,(c)90%,(d)95%,(e)98%增加IL8Rb时灵敏度的增益,并且在不同组特异性(f)80%,(g)85%,(h)90%,(i)95%,(j)98%增加IL8Rb时灵敏度的所得增益。

[0084] 图18描绘了用于测试使用单独的基因测试、单独的表型评价、和/或同时使用基因测试和表型评价来研究的具有血尿患者的ROC曲线。

[0085] 图19描述了本发明的变量,性别、吸烟史和HFREQNEW的比值比(横轴)的图表。

[0086] 图20A和图20B描绘了报告诊断准确性的标准的流程图。图20A描绘本研究中所有三个队列中肉眼可见血尿患者的流程图。

[0087] 图20B描绘了报告包括在本研究中的显微可见血尿患者的诊断准确性的流程图。

[0088] 图21描绘了代表三个分类模型的ROC曲线。P指数(点线),G指数(虚线)和G+P指数(实线)。

[0089] 图22描绘了根据每个模型NPV对血尿检测阴性的患者的比例。P指数(点线),G指数(虚线),和G+P指数(实线)。

[0090] 图23描绘了检测结果(横轴)对归类结果(纵轴)之间的关系的图表。

[0091] 图24描绘了G2指数(横轴)对G1+P指数(纵轴)的图表。

具体实施方式

[0092] 定义

[0093] 在详细说明本发明的实施例之前,提供本说明书中所使用的一些术语的定义是有用的。

[0094] 术语“标志物”指与生物现象的存在定量地或定性地相关联的分子。“标志物”的示例包括多核苷酸,例如编码或非编码的基因或基因片段,编码或非编码的DNA或DNA片段、RNA或RNA片段;或基因产物,包括多肽,例如肽,寡肽,蛋白或蛋白片段;或与位于所述现象之下的机制直接地或间接地相关的任何相关的代谢产物,副产物,或任何其他识别分子,例如抗体或抗体片段。本发明的标志物包括如本文所公开的核苷酸序列(例如,基因库序列),特别是,全长序列,任何编码序列,任何片段,任何可能的探针(例如,跨越外显子—外显子边界创建的),其中包括捕获基序,发夹结构或荧光团,或其任何补充物,以及如上述所定义的任何可测量的标志物。

[0095] 如本说明书所用的“抗体”和类似术语是指免疫球蛋白分子和免疫球蛋白(Ig)分子的免疫活性部分,即含有与抗原特异性结合(免疫发生反应)的抗原结合位点的分子。这些包括,但不限于,多克隆、单克隆、嵌合、单链、Fc片段、Fab片段、Fab'片段和Fab₂片段以及Fab表达库。抗体分子涉及IgG、IgM、IgA、IgE和IgD类中的任何一种,它们通过存在于分子中的重链的性质彼此不相同。这些也包括子类,例如IgG1、IgG2等等。轻链可以是kappa链或lambda链。本说明书中提及的抗体包括涉及所有类、子类和类型。还包括嵌合抗体,例如,特定于多个源的其单克隆抗体或其片段,例如,老鼠或人类序列。此外还包括骆驼抗体,鲨鱼抗体或纳米抗体。

[0096] 术语“癌症”和“癌的”指或描述哺乳动物的生理疾病,其特征典型地为异常或失调的细胞生长。例如,癌症和癌症病理可以与以下状况相关,例如,具有转移邻近细胞的正常功能,与邻近细胞的正常功能相干扰,细胞激素或其他分泌产物的释放水平异常,炎性或免疫反应的抑制或加重,瘤样病变,恶性肿瘤前照,恶性肿瘤,周围或远处组织或器官的侵入,例如淋巴结,等等。

[0097] 术语“肿瘤”是指所有肿瘤细胞生长和增殖,无论是恶性的或良性的,以及所有的

癌前和癌变细胞和组织。

[0098] 术语“膀胱癌”是指产生于膀胱的肿瘤。这些肿瘤能够转移到任何器官。

[0099] 术语“BTM”或“膀胱肿瘤标志物”或“BTM家族成员”是指与尿路上皮癌、膀胱癌、膀胱移行细胞癌(TCC)、鳞状细胞癌和膀胱腺癌相关联的肿瘤标志物(TM)。术语BTM还包括单个标志物的组合,其组合提高了检测膀胱癌的敏感性和特异性。但是应当理解,术语BTM并不要求所述标志物仅对于膀胱肿瘤是特异性的。相反,BTM的表达可以在其它类型的细胞、患病细胞、肿瘤,包括恶性肿瘤中被改变。

[0100] 术语“低估表达BTM”是指标志物,其在膀胱肿瘤中展现的表达比在非恶性膀胱组织中低。

[0101] 术语“过度表达BTM”是指标志物,其在膀胱肿瘤中展现的表达比在非恶性膀胱组织中高。

[0102] 术语“差异表达的”,“差别表达”和类似的短语,指的是相对于其在控制受试者(例如对照样本)中的表达,其表达在患病(特别是癌症,例如黑色素瘤)的受试者(例如测试样本)中被激活至较高或较低的水平水平的基因标志物。该术语还包括其表达在相同疾病的不同阶段;在具有良好或不良的预后的疾病;或在有较高或较低水平的增殖的细胞中被激活至较高或较低的水平水平的标志物。差异表达的标志物可能在多核苷酸水平或多肽水平被激活或被抑制,或者可能受到选择性剪接以产生不同的多肽产物。例如,此类差异可以通过mRNA水平的变化、表面表达、分泌或多肽的其他分区来证明。

[0103] 差异表达可以包括两个或更多个标志物(例如,基因或其基因产物)之间的表达的对比;或两个或更多个标志物之间的表达的比例的对比(例如,基因或其基因产物);或在相同标志物的两种不同的加工产物(例如,转录物或多肽)之间的对比,该相同标志物在正常受试者和患病受试者之间;或在同一疾病的不同阶段之间;或在具有良好或不良预后的疾病之间;或在具有较高和较低水平的增殖的细胞之间;或在正常组织和病变组织(特别是癌症,或黑色素瘤)之间的不同。差异表达包括例如正常和患病细胞,或已经历不同疾病事件或疾病阶段的细胞中,或具有不同程度增殖的细胞中的基因或其表达产物中的暂存的表达模式或细胞表达模式中定量及定性的差异。

[0104] 术语“表达”包括生成多核苷酸和多肽,特别是,由基因或基因的一部分生成RNA(例如,mRNA),并且包括生成由RNA或基因或基因的一部分编码的多肽,以及与表达相关联的可检测材料的出现。例如,从多肽—多肽相互作用,多肽—核苷酸相互作用,或类似物形成复合物,包括在术语“表达”的范围内。另一个示例是配基(例如杂交探针或抗体)与基因或其他多核苷酸或寡核苷酸、多肽或蛋白片段的结合,以及配基的可视化。因此,微阵列上、杂交印记(例如Northern印迹)上,或在免疫印迹(如Western印迹)上,或珠阵列上,或通过PCR分析的点的强度,包括在构成其基础的生物分子的术语“表达”内。

[0105] 术语“基因表达阈值”,和“限定的表达阈值”可互换使用并且是指所讨论的标志物的水平,在该水平外,多核苷酸或多肽的表达水平用作患者疾病的预测标志物。例如,高于某个阈值的IL8Rb的表达是诊断患者患有炎症性病症的。使用膀胱癌标志物检测疑似膀胱癌患者时,也可以使用阈值。高于阈值的表达水平表示该患者患有炎症性膀胱疾病,容易造成癌症的假阳性测试,而低于阈值的IL8Rb的表达水平预示该患者没患有炎症性膀胱疾病。通过包含IL8Rb的测量,如果IL8Rb的水平低于阈值,那么来自膀胱肿瘤标志物的表达的任何结果

都可被信赖(即阳性结果很可能对患有癌症的病人是阳性的,而不是实际上从发炎粘膜的非恶性细胞的剥落造成的膀胱癌标志物的水平增加)。

[0106] 术语“诊断阈值”指的是一个阈值,根据这个阈值,患者可以说已经被诊断患有或没患有指定的疾病,例如膀胱癌。诊断阈值根据多种因素,例如人口、患病率以及可能的临床结果,通常被设定为达到期望的灵敏度和特异性。通常,诊断阈值可以使用算法,和/或计算机化的数据分析来计算和/或确定。

[0107] 确切的阈值将取决于人口以及任何被用来预测疾病的模型(预测模型)。阈值是从如在下面示例中所述的临床研究中用试验方法确定的。取决于所使用的预测模型,表达阈值可以被设定为达到最大灵敏度,或针对最大特异性设定,或者针对最小误差(最大分类率)设定。例如较高的阈值可以被设定为达到最小误差,但是这可能会导致较低的灵敏度。因此,对于任何给定的预测模型,临床研究将被用于设定表达阈值,其通常达到最高灵敏度同时具有最小误差率。通常,该阈值可能大约是在…的表达(注:此处表达不完全,说明书第16页倒数第二段末尾)

[0108] 术语“灵敏度”指的是患有某种疾病的个体(由模型)测试为阳性的比例。因此,提高的灵敏度意味着更少的假阴性测试结果。

[0109] 术语“特异性”是指没患有某种疾病的个体(由模型)测试为阴性的比例。因此,提高的特异性意味着更少的假阳性测试结果。

[0110] 术语“接受者操作特性”(“ROC曲线”)指对于特定标志物或测试不同切断点的真阳性率(灵敏度)对假阳性率(特异性)的曲线图。ROC曲线上的每个点代表特定的灵敏度/特异性点,其将对应给定的阈值。ROC曲线对于建立阈值以得到所期望的结果是重要的。ROC曲线下的面积表示(表达为曲线下面积(AUC)分析,可以是给定的标志物或由多个标志物组成的测试如何的度量,其可以区分两个或更多诊断结果。ROC曲线也可以用来比较两个不同测试的准确性。

[0111] 术语“寡核苷酸”是指多核苷酸,通常是探针或引物,包括但不限于单链脱氧核糖核苷酸,单链或双链核糖核苷酸, RNA:DNA 杂交体,和双链DNA。寡核苷酸,如单链DNA探针寡核苷酸,通常通过化学方法(例如使用可商购的自动寡核苷酸合成仪),或通过各种其他方法合成,包括体外表达系统,重组技术,以及在细胞和有机体中的表达。

[0112] 术语“过度表达”或“过度表达的”是指高于在正常组织中可见的患者的基因或标志物的表达水平。如果该表达是正常组织中或另一组患者的组织中的表达的1.1,1.2,1.3,1.4,1.5,1.6,1.7,1.8,1.9,2,10,100,1000,或高达10,000倍,那么该表达可被认为是过度表达的。

[0113] 术语“多核苷酸”,在单数或复数使用时,一般是指任何多核糖核苷酸或多聚脱氧核糖核苷酸,其可以是未修饰的RNA或DNA或修饰的RNA或DNA。这包括,但不限于,单链和双链DNA,包括单链和双链区域的DNA,单链和双链RNA,和包括单链和双链区域的RNA,包含可以是单链或,更典型地,双链或包括单链和双链区域的DNA和RNA的杂交分子。还包括的是包含RNA或DNA或RNA和DNA两者的三链区域。特别包括的是mRNA,cDNA,和基因组DNA,和任何它们的片段。该术语包括含有一个或多个修饰的碱基,例如氘化碱基,或不寻常的碱基,如肌苷的DNA和RNA。本发明的多核苷酸可以包括编码或非编码序列,或者有义或反义序列。应当理解的是,在本说明书中每个引用“多核苷酸”或类似的术语,将包括全长序列以及其任何

片段,衍生物或变体。

[0114] 术语“表型”是指在临床环境中,或在临床交谈中,或在病人的病史中可观察的特征。当用于计算G+P指数的公式中,“表型”或“P”是指患者的年龄,性别,血尿的发生率,和吸烟史。

[0115] “多肽”,如在本说明书中所使用,是指寡肽,肽,或蛋白序列,或其片段;以及天然存在的,重组的,合成的或半合成的分子。当“多肽”在本说明书中列举以指天然存在的蛋白分子的氨基酸序列时,“多肽”和类似的术语并不意味着限制氨基酸序列为全长分子的完整天然氨基酸序列。应当理解的是,在本说明书中每个引用“多肽”或类似的术语,将包括全长序列,以及其任何片段,衍生物或变体。

[0116] 术语“qPCR”或“QPCR”是指例如在PCR Technique:Quantitative PCR, J.W.Larrick,ed.,Eaton Publishing,1997,和A-Z of Quantitative PCR,S.Bustin,ed., IUL Press,2004中描述的定量聚合酶链反应。

[0117] 术语“逆转录”是指一个过程,在其中,包括信使RNA(“mRNA”)的寡核糖核苷酸被用作使用酶(“逆转录酶”)的互补寡脱氧核糖核苷酸(“cDNA”)的生化合成的模板,该酶结合到模板RNA,并且催化一系列加成反应,该加成反应依次附接脱氧核苷酸碱基,以形成与RNA模板互补的寡脱氧核糖核苷酸链。

[0118] 术语“血尿”被定义为尿液中存在血液。它可以表现为尿液中的肉眼可见血尿(血细胞的可视痕迹),或显微可见血尿(血液的微观痕迹)。显微可见血尿的确认指示被定义为在至少3个妥善收集的尿样上每个显微镜高倍视野(HPF)下呈现3个或更多个血细胞。显微可见血尿也可以在诊所通过尿液试纸(比色对比估计)进行检测。血尿(显微的或肉眼的)可以无症状(没有伴有血尿的其他症状)或有症状。其他症状包括尿痛(排尿疼痛),膀胱排空不完全的感觉或增加的频率或排尿。

[0119] 杂交反应的“严格性”是本领域的普通技术人员容易确定的,并且通常是取决于探针长度,洗涤温度,和盐浓度的经验计算。通常,较长的探针需要较高的温度进行适当的退火,而较短的探针需要较低的温度。当互补链存在于低于其解链温度的环境中时,杂交通常取决于变性DNA重新退火的能力。所述探针和可杂交序列之间的预期同源性程度越高,可使用的相对温度越高。其结果是,随之而来的较高的相对温度将趋向于使反应条件更为严格,而较低温度将趋向于使反应条件较为不严格。额外的细节和杂交反应的严格性的解释可在例如Ausubel等人,Current Protocols in Molecular Biology,Wiley Interscience Publishers,(1995)中找到。

[0120] “严格条件”或“高严格条件”,如本说明书中所定义的,典型地:(1)采用低离子强度和高温用于洗涤。例如在50℃的0.015M氯化钠/0.0015M柠檬酸钠/0.1%十二烷基硫酸钠;(2)在杂交过程中采用变性剂,如甲酰胺,例如,50%(体积/体积)甲酰胺与0.1%牛血清白蛋白/0.1%细胞分层液/0.1%聚乙烯吡咯烷酮/50mM磷酸钠,pH值为6.5的缓冲液与在42℃的750mM氯化钠,75mM柠檬酸钠;或(3)采用50%甲酰胺,5X SSC(0.75M氯化钠,0.075M柠檬酸钠),50mM磷酸钠(pH值为6.8),0.1%焦磷酸钠,5X,邓波特溶液,超声处理的鲑鱼精子DNA(50ug/ml),0.1%SDS(十二烷基硫酸钠)和在42℃的10%硫酸葡聚糖,与在42℃的0.2X SSC(氯化钠/柠檬酸钠)和在55℃的50%甲酰胺中洗涤,随后用包括在55℃的含有EDTA(乙二胺四乙酸)的0.1X SSC(氯化钠/柠檬酸钠)中高严格性洗涤。

[0121] “中等严格条件”可以被识别为如在Sambrook等人, Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989中所描述的, 并且包括对比上文描述的, 较为不严格地使用洗涤溶液和杂交条件(例如, 温度, 离子强度和% SDS(十二烷基硫酸钠))。中等严格条件的一个示例是于37℃在溶液中的过夜培养, 溶液包括: 20% 甲酰胺, 5X SSC (150mM氯化钠, 15mM柠檬酸三钠), 50mM磷酸钠 (pH值为7.6), 5X邓波特溶液, 10% 硫酸葡聚糖, 和20mg/ml变性剪切的鲑鱼精子DNA, 接着在约37-50℃的1X SSC (氯化钠/柠檬酸钠) 中清洗过滤器。本领域技术人员将认识到如何根据需要调整温度, 离子强度等以适应如探针长度等因素。

[0122] 术语“IL8Rb”是指嗜中性粒细胞标志物白介素8受体B (也称为趋化因子 (C-X-C 基序) 受体2 [CXCR2]) (图1; 序列号1和2), 并且包括标志物IL8Rb。该术语包括多核苷酸, 例如基因或基因片段, RNA或RNA片段; 或基因产物, 包括多肽, 如肽, 寡肽, 蛋白或蛋白片段; 或任何相关的代谢产物, 副产物, 或任何其他识别分子, 如抗体或抗体片段。

[0123] 术语“可靠性”包括假阳性和/或假阴性的低发生率。因此, 标志物的可靠性越高, 与使用该标志物作出诊断相关联的假阳性和/或假阴性越少。

[0124] “精度”是真实的结果 (真阳性加真阴性) 除以群体中的总病例数的比例, 根据公式:

$$\frac{\text{真阳性} + \text{真阴性}}{\text{测量总数}}$$

[0126] 术语“归类 (triage)”是指从那些具有合理概率患有膀胱癌以及需要进一步的临床病情检查, 包括膀胱镜检查或其他临床程序的血尿患者中区分出具有低概率患有膀胱癌的血尿患者。

[0127] 实施例

[0128] 因此, 在某些优选的实施例中, 提供了遗传标志物和表型标志物的组合, 其允许从那些具有足够风险患有膀胱癌以授权进一步的临床病情检查, 可能包括膀胱镜检查或其他程序的患者中区分出具有低概率患有膀胱癌的患者。在其他实施例中, 提供了标志物, 其有大于约70%的可靠性; 在其他实施例中, 大于约73%的可靠性, 仍然在其他实施例中, 大于约80%的可靠性, 又在进一步的实施例中, 大于约90%的可靠性, 在还有一些实施例中, 大于约95%的可靠性, 又在进一步的实施例中, 大于约98%的可靠性, 以及在某些实施例中, 约100%的可靠性。

[0129] 对于遗传分析, 除非另有指明, 本发明的实践将采用分子生物学 (包括重组技术), 微生物学, 细胞生物学和生物化学的常规技术, 这些技术是在本领域的技术范围内的。这些技术在以下文献中有充分解释, 例如, Molecular Cloning: A Laboratory Manual, 2nd edition, Sambrook等人, 1989; Oligonucleotide Synthesis, MJ Gait, ed., 1984; Animal Cell Culture, R.I. Freshney, ed., 1987; Methods in Enzymology, Academic Press, Inc.; Handbook of Experimental Immunology, 4th edition, D.M. Weir & C.C. Blackwell, eds., Blackwell Science Inc., 1987; Gene Transfer Vectors for Mammalian Cells, J.M. Miller & M.P. Calos, eds., 1987; Current Protocols in Molecular Biology, F.M. Ausubel等人, eds., 1987; 以及PCR: The Polymerase Chain Reaction, Mullis等人, eds., 1994。

[0130] 应当理解的是,上述术语可以指蛋白,DNA序列和/或RNA序列。也可以理解的是,上述术语也指具有如本说明书中所描述的同源序列的非人类蛋白质,DNA和/或RNA。

[0131] 本发明的实施例

[0132] 通常泌尿科医生所指的血尿患者预约在专门血尿门诊就诊。肉眼可见血尿的患者往往优先于那些显微可见血尿的患者。来自于初级保健提供者转诊时提供的信息可以是高度可变的,这使得依据患有尿路上皮癌的概率来精确分层患者是困难的。通常在患者就诊前,尿细胞学是常规地被要求的,并且如果为阳性,将增加患者接受完整的临床病情检查的优先权,但是虽然尿细胞学有时具有高度特异性,但是却具有非常低的敏感性,因此凭借其高假阴性率⁽⁶⁾,具有很低的实用价值。

[0133] 没患有膀胱癌的血尿患者的表型和基因型分析

[0134] G+P指数

[0135] 上述表型和基因型变量是根据以下关系组合成G+P指数的:

[0136] $G+P \text{ 指数} = (w1*HFREQ + w2*AgeGT50 + w3*Gender + w4*SMK + w5*RBC) + (w6*M1 + w7*IL-8)$,

[0137] 其中,HFREQ代表在6个月周期内每个高倍视野下找到3个或更多个红细胞的频率;如果频率低,则HFREQ=0,并且如果每个高倍视野下超过3个红细胞,则HFREQ=1, AgeGT50是指受试者的年龄,如果大于50岁,则AgeGT50=1,如果不到50岁,那么AgeGT50=0。性别男性为1,性别女性为0。SMK指受试者是否当前或以前是吸烟者;如果为非吸烟者则SMK=0,如果为吸烟者,则SMK=1。RBC指红细胞计数;如果为25个或更多,则RBC设置为1,而如果小于25,则RBC为0。M1是遗传标志物MDK,CDC,IGFBP5和HOXA13的表达的组合;如果M1>4.5,则将其设置为1,如果小于4.5,则将其设置为0。IL-8指IL-8的RNA的表达水平;如果IL-8>2.5,则IL-8被设置为1,如果IL-8<2.5,则IL-8=0。符号“*”表示乘法运算符,并且加权因子w1—w7分别为分配给上面在G+P指数中列出的每个变量的权重。

[0138] 在其它优选的实施例中,(AgeGT50和RBC)可以从模型中去掉,如下所示:

[0139] $G+P \text{ 指数} = (1*HFREQ + 3*Gender + 4*SMK) + (5*M1 + 2*IL-8)$

[0140] G+P指数的计算值在0至15之间。G+P指数值为11至15的患者被认为是膀胱癌“高风险”的,并且指出了额外的膀胱癌病情检查的需求。G+P指数值为6至10的患者被认为是“中等风险”发展为膀胱癌并且指示了需要额外的病情检查。G+P指数值为0至5的患者被认为是“低风险”发展为膀胱癌。“低风险”组的患者被放置在观察等待名单上,如果其他症状出现,或者如果显微可见血尿反复发生,则他们为可能的进一步病情检查被重新评估。

[0141] 对区分没患有和患有膀胱癌的患者有用的基因型变量包括RNA标志物“M1”的表达,“M1”是MDK,CDC,IGFBP5 (IGBP5),和HOXA13的组合。另一个基因型变量是IL8R的RNA的表达。这些基因型变量的系数示于下面的表7中。4.5和2.5分别为M1和IL8R的阈值,并且分别分配了系数5和2。

[0142] G+P指数的计算结果为0至15之间的一个值。G+P指数值为11至15被认为是膀胱癌“高风险”的,并且指出了额外的膀胱癌病情检查的需求。G+P指数值为6至10被认为是“中等风险”发展为膀胱癌并且指示了需要额外的病情检查。G+P指数值为0至5被认为是“低风险”发展为膀胱癌,并且这些患者被放置在等待名单上,如果其他症状出现,或者如果显微可见血尿反复发生,则他们为可能的全面病情检查被重新评估。

[0143] 如在示例3(图18和19)中更充分地描述的,我们发现,单独的表型数据的ROC曲线生成适度水平的诊断能力。单独的基因型数据的ROC曲线生成显著水平的诊断能力。当基因型数据和表型数据组合成G+P指数时,我们发现了出乎意料地更好的诊断能力。

[0144] 没患有膀胱癌的患者遗传分析

[0145] 在一些优选的实施例中,本发明结合使用4标志物测定或Cxbladder[®]测定(基因型变量)和五个主要风险因素(表型变量)中的一个或多个,以生成可用于归类显微镜可见血尿或肉眼可见血尿患者在患有尿路上皮癌的潜在风险方面的选择指数。虽然不排除灵活的膀胱镜检查的必要性,但基于表型变量和基因型变量被认定为具有尿路上皮癌高风险的患者可以更早就诊,可能全面改善患者的结果。

[0146] 基因型标志物可以作为工具用于检测无癌患者或选择患有疾病的低,中或高风险的患者组别。标志物可以,例如,在疾病组织和相应的非疾病组织之间差异表达。在这种情况下,差异表达的检测与疾病的存在相关联。可替代地,标志物可以直接与出现在疾病组织的变化,或疾病造成的变化相关联。炎性疾病与嗜中性粒细胞的增加相关联。已经发现的是,嗜中性粒细胞标志物白介素8受体B(IL8Rb;图1;序列号1和2),可以为样本中存在嗜中性粒细胞提供良好的标志物,并且因此可用作诊断标志物用于检测样本中的炎性疾病,特别是,检测炎性膀胱疾病。

[0147] 如图5所示,尿液中IL8Rb的累积指示炎性膀胱疾病的存在。具体而言,图5示出患有疾病的患者的尿液中IL8Rb的累积;良性前列腺增生,泌尿道感染,非特异性前列腺疾病,血管前列腺和继发于华法林的使用。然而,可以理解的是,IL8Rb的使用并不限于只检测这些疾病,但是这些示例表明,IL8Rb确实在患有膀胱炎性疾病的患者的样本中增加。也就是说,IL8Rb可以用作与膀胱疾病相关联的炎症的标志物,因此适合于用来检测与炎症相关联的任何疾病。因此,检测IL8Rb的量可以用作膀胱炎性疾病的标志物。更特别地是,IL8Rb可用于检测与嗜中性粒细胞的累积相关联的炎性膀胱疾病。

[0148] 对TCC的尿检很大程度上依赖于尿液中存在从脱落的肿瘤细胞获得的标志物。检测这些细胞的能力可以被大量的污染细胞,如血液和炎性细胞的存在所掩盖。此外,膀胱内膜的炎症可以导致非恶性细胞从黏膜脱落的增加。结果是,使用从膀胱移行细胞获得的标志物的尿检具有较高的可能性对取自患有膀胱炎,尿路感染或其他疾病导致的泌尿道炎症或移行细胞剥落,如尿路结石,的患者的尿样给出假阳性结果(Sanchez-Carbayo等人)。

[0149] 避免这样的假阳性结果的一种方式是在血液或炎症细胞中表达相对较低的标志物。使用这样的标志物的结果是在呈现非恶性,炎症性疾病的TCC患者中较少的假阳性。然而,在血液学来源的细胞中低表达的标志物未能弥补非恶性移行细胞脱落的增加率。

[0150] 已经被发现的是,来自发炎组织的所剥离的移行细胞对膀胱癌尿检精度的负面影响,可以通过改进泌尿道炎性病症患者的鉴定来最小化。这里,已经令人惊奇地发现,使用标志物IL8Rb与一种或多种膀胱肿瘤标志物(BTM's)相结合提供了对于膀胱癌的更精确的检测。特别是,对于膀胱癌的基于标志物的检测,其包括标志物IL8Rb,是假阳性结果的可能性更低,这会导致患者患有炎性非癌病症的结果。

[0151] 一般来说,炎性病症存在或不存在由具有基因表达的阈值来确认,高于该阈值的IL8Rb的表达指示炎性病症。例如,IL8Rb的表达高于某个阈值是诊断患者患有炎性病症(见上文所描述的阈值)。

[0152] 当IL8Rb与预测膀胱癌存在的一个或多个标志物结合使用时,膀胱肿瘤标志物的升高的表达的存在和IL8Rb的表达高于某一阈值,可以预测患者患有炎症病症而不是癌症。此外,如果测试是对于患者的尿液进行的,那么该结果可以预测患者患有炎症膀胱疾病。膀胱肿瘤标志物的高水平最有可能是作为炎症结果的来自黏膜的非恶性细胞的结果。也就是说,患者虽然具有高水平的膀胱肿瘤标志物,但实际上没患有膀胱癌——即假阳性。

[0153] 可替代地,如果患者具有一种或多种膀胱肿瘤标志物的异常高水平或诊断水平,但是IL8Rb的水平低于阈值,则患者可能患有癌症,尤其是膀胱癌或尿路上皮癌。如果测试是对于来自患者的尿液进行的,则尤其是这样。这个结果对于健康服务提供者是显著有利的,因为他们可以肯定患者确实患有癌症,并且可以立即开始治疗,而不必担心该结果实际上是由给出假阳性结果的炎症疾病导致的。

[0154] 已经令人惊讶地表明,来自编码嗜中性粒细胞标志物白介素8受体B(IL8Rb)的基因的RNA的定量提高了使用公知的TCC或BTM标志物检测TCC患者的整体表现。IL8Rb的参考序列示出于图1和序列号1和2)。除了其在TCC检测中的作用,还探索了IL8Rb是否可以用作尿液标志物以帮助炎症疾病的诊断(图5)。

[0155] 使用IL8Rb标志物可以利用用于检测基因表达水平的已知方法,单独用于检测膀胱的炎症病症。用于检测基因表达的方法的示例概述如下。

[0156] 可替代地,IL8Rb可以与多个BTM中的一个相结合,以检测膀胱癌。已被示出,通过利用炎症疾病标志物IL8Rb作为测试膀胱癌的一部分,发炎组织对创建假阳性结果的影响被最小化。标志物IL8Rb可以与任何膀胱癌标志物联合使用,或者可替代地可以与两个或更多个标志物一起使用,作为特征的一部分,用于检测膀胱癌。

[0157] 减少假阳性结果的数量是指更少的没患有膀胱癌的患者经受可能不必要的程序,包括膀胱镜检查,该检查本身就有风险。减少假阴性结果的数量意味着膀胱癌患者更有可能被检测出,并因此对癌症进行进一步评估。

[0158] IL8Rb改善膀胱癌检测结果的作用来自将患有非恶性疾病和膀胱癌的患者分开的能力。做到这一点,因为IL8Rb的增加指示样本中增加的嗜中性粒细胞的存在。因此,IL8Rb的能力不依赖于所使用的膀胱肿瘤标志物。如图2和图12至图15所示,当与多种膀胱肿瘤标志物和膀胱肿瘤标志物的组合相结合时,IL8Rb有增加标志物的能力的特异性的总体效果,以检测受试者的癌症。

[0159] 根据本发明的一个特征的示例是将IL8Rb与MDK,CDC2,IGFBP5和HOXA13结合使用,其也可以与适于检测膀胱癌的一个或多个其它标志物结合,例如图6或图7中列出的多个标志物中的任一项。如在图14和图15中所示,IL8Rb可以用在标志物的任意组合中,特别是组合IL8Rb/MDK,IL8Rb/CDC2,IL8Rb/HOXA13,IL8Rb/IGFBP5,IL8Rb/MDK/CDC2,IL8Rb/MDK/HOXA13,IL8Rb/MDK/IGFBP5,IL8Rb/CDC2/HOXA13,IL8Rb/CDC2/IGFBP5,IL8Rb/HOXA13/IGFBP5,IL8Rb/MDK/CDC2/HOXA13,IL8Rb/MDK/CDC2/IGFBP5,IL8Rb/CDC2/HOXA13/IGFBP5和IL8Rb/MDK/CDC2/HOXA13/IGFBP5。如图14和图15所示,IL8Rb的加入增加了标志物或标志物组合准确地诊断受试者的膀胱癌的能力。本发明不限于这些特定的组合,但是可任选地包括适于检测膀胱癌的一种或多种其它标志物,例如图6或图7中列出的多个标志物中的任一项。下面表1显示了特异性标志物MDK,CDC2,IGFBP5和HOXA13和IL8Rb的标识。

[0160] 表1:膀胱肿瘤标志物的标识

[0161]

PE 基 因名称	HGNC 基 因 名 称 (官方)	NCBI 参照序列	NCBI Entrez 基 因 ID	HGNC 链接
MDK	MDK	NM_002391	4192	http:// www.genenames.org/data/hgnc_data.php?hgnc_id=6972
CDC	CDK1	NM_001170406	983	http:// www.genenames.org/data/hgnc_data.php?hgnc_id=1722
IGF	IGFBP5	NM_000599	3488	http:// www.genenames.org/data/hgnc_data.php?hgnc_id=5474
HOXA	HOXA13	NM_000522	3209	http:// www.genenames.org/data/hgnc_data.php?hgnc_id=5102
IL8Rb	CXCR2	NM_001168298	3579	http:// www.genenames.org/data/hgnc_data.php?hgnc_id=6027

[0162] 图2至图4和图12至图17示出了将IL8Rb与四个已知的有代表性的膀胱癌的标志物;MDK,CDC2,IGFBP5和HOXA13组合使用的效果。结果表明,通过组合使用,如果IL8Rb单独与每个标志物(图2,图14和图15至图17)组合使用,而且还有当与四个BTM标志物的所有可能组合一起用作特征时,将TCC患者的样本和那些非恶性疾病患者的样本区分开的能力得到了提高。

[0163] 如图10至图13所示,IL8Rb加入到四个标志物MDK,CDC2,IGFBP5和HOXA13(uRNA-D)中不仅相比单独的四个标志物增加了测试的整体表现,而且与其他已知的测试相比该测试也是非常有利的,其他已知的测试包括:埃利萨的NMP22[®]“美国马萨诸塞州Matriotech, Inc.的注册商标”,NMP22BladderChek[®](美国马萨诸塞州Matriotech, Inc.的注册商标),和细胞学。图14至图17也示出了IL8Rb在四个标志物MDK,CDC2,IGFBP5和HOXA13的各种组合中的效果。

[0164] 图14示出了使用五个不同的分类器模型(i)线性判别分析(LDA), (ii)逻辑回归(LogReg), (iii)支持向量机(SVM), (iv)K-5最近邻(KN5N),以及(v)划分树分类器(TREE)计算的四个标志物MDK,CDC2,IGFBP5和HOXA13加上和不加IL8Rb的所有组合的ROC曲线图。图15列表显示了所有5种分类器及4个生物标志物加上和不加IL8Rb的所有15种组合的曲线下

面积 (AUC)。此AUC计算被限制在假阳性率0至20%的面积,覆盖特异性的有效范围 (80–100%)。AUC定量了图14的ROC曲线上的可见差异。图16示出了在具有和不具有IL8Rb时所测量的4个标志物的所有组合在特异性为(a) 80%, (b) 85%, (c) 90%, (d) 95%, 和 (e) 98%的敏感性。图17列表显示了在固定特异性分别为(a, f) 80%, (b, g) 85%, (c, h) 90%, (d, i) 95%, 和 (e, j) 98%时,灵敏度 (ROC曲线上的垂直方向;较好的是“向上”)或特异性 (ROC曲线上的水平方向;较好的是向左)的变化。

[0165] 这些结果表明,通常IL8Rb提高了生物标志物 (MDK, CDC, IGFBP5和HOXA13), 单独地或组合地,从正常的样本中分类出肿瘤的能力。

[0166] 这些结果通常表明IL8Rb能够增加测试能够检测出的膀胱癌或尿路上皮癌的精度。最大的收获是可看到标志物未加入IL8Rb则表现不好,或者分类器表现不好。较小的收获是可看到在加入IL8Rb之前标志物和/或分类器表现很好,因此其改进的空间较小。重要的是注意到,结果显示基于群体的分析以及当诊断个体患者,尤其是那些对BTM标志物的表达的诊断可能不清楚的患者,加入IL8Rb的益处可能更大。

[0167] 这些结果表明,IL8Rb不仅可以用于检测膀胱炎性疾病,在与标志物结合时还可以用于检测膀胱癌,导致减少“假阳性”结果减少引起的改进的膀胱癌检测。

[0168] 这些结果还表明IL8Rb的效用在于,其影响到各种标志物组合的整体表现,并且证实了IL8Rb改善一个或多个膀胱癌标志物精确检测患者中癌症的表现的能力。此外,图14和图15表明,使用一系列分类器模型可以实现相同的结果,以及表明该结果不依赖于分类器模型或算法,而是依赖于所用的标志物的组合。这些结果证实,任何适当的分类器模型或算法可以在本发明中使用。特别是,图14和图15表明,IL8Rb在较高特异性上的效果更大,特别是在临床上最适用的范围内。

[0169] 因此,使用本发明的G+P指数,我们能够基于表型变量和遗传变量准确地归类糖尿病患者到不同组:“高风险”组,患有尿路上皮癌组并且需要立即进一步的病情检查,“风险”组,并且需要立即的病情检查,“低风险”组,患者可以置于观察名单上,用于以后的评估。

[0170] 在身体样本中检测遗传标志物

[0171] 在若干优选的实施例,对癌症的测定可期望在从血液,血浆,血清,例如使用腹腔冲洗液获得的腹膜液,或其它体液,例如尿,淋巴液,脑脊髓液,胃液或粪便样本上获得的样本上进行。对于检测膀胱的炎症病症或膀胱癌,理想地是,测试在尿样上进行。

[0172] 具体而言,用于检测炎症性膀胱疾病或膀胱癌的本方法可以在来自身体的任何合适的样本上执行,该样本将指示尿液的,但是理想地是IL8RB的,以及任何直接从尿样中确认的其他癌症标志物的水平。

[0173] 检测可以对尿样直接进行,或者样本可以通过加入本领域中已知的任何合适的化合物或缓冲剂被稳定,从而稳定并防止样本中RNA和/或蛋白的分解,使得其可以在稍后进行分析,甚至确保在分析过程中RNA和/或蛋白是稳定的。

[0174] 确定受试者尿液中的蛋白和/或RNA水平可以直接对尿液进行,或者可以对尿液进行处理,以进一步纯化和/或浓缩RNA和/或蛋白质。用于提取和/或浓缩蛋白和/或RNA的许多方法在本领域中是众所周知的,并且可以在本发明中使用。

[0175] 可以理解的是,许多方法是本领域公知的用于确认特定基因的表达水平,RNA和/或蛋白的水平,并且可以在本发明中使用任何合适的方法。一些常用的方法概述如下,然

而,本发明并不限于这些方法,并且用于定量蛋白和/或RNA水平的任何方法都是适用于本发明的。

[0176] 使用基因型标志物检测疾病和癌症的一般方法

[0177] 用于确定表达水平的一般方法概述如下,但是可以理解的是,用于确定表达水平的任何方法都是合适的。

[0178] 定量聚合酶链反应(qPCR)

[0179] 定量PCR(qPCR)可以使用特定的引物和探针对肿瘤样本,血清和血浆进行。在受控制的反应中,在PCR反应中形成的产物的量(Sambrook,J.,E Fritsch,E.和T Maniatis, Molecular Cloning:A Laboratory Manual 3rd.Cold Spring Harbor Laboratory Press: Cold Spring Harbor(2001))与初始模板的量相关联。PCR产物的定量可以在反应物变得有限之前通过在对数期时停止PCR反应来进行。然后PCR产物在琼脂糖或聚丙烯酰胺凝胶中电泳,用溴化乙锭或类似的DNA染色剂染色,并且染色的强度通过光密度测定法测量。可替代地,PCR反应的进展可以使用PCR机器例如Applied Biosystems' Prism 7000TM(美国,康涅狄格州,Applied Biosystems Corporation的商标)或Roche LightCyclerTM(美国,加利福尼亚,Roche Molecular Systems,Inc.的商标)测量,其实时测量产物的积累。实时PCR测量DNA嵌入染料例如Sybr Green进入合成的PCR产物的荧光,或当猝灭剂分子裂解时报告物分子释放的荧光;报告物分子和猝灭剂分子被掺入到寡核苷酸探针中,该寡核苷酸探针杂交到在延伸自引物寡核苷酸的DNA链之后的目标DNA分子。寡核苷酸探针被在下个PCR循环中的Taq聚合酶的酶促作用移位并降解,从猝灭剂分子中释放报告物。在一个变型中,众所周知的Scorpion,该探针共价连接至引物。

[0180] 逆转录PCR(RT-PCR)

[0181] RT-PCR可以用于比较不同样本群体,正常和肿瘤组织,有或没有药物治疗的RNA水平,以表征表达模式,在密切相关的RNA之间做区分,并且分析RNA结构。

[0182] 对于RT-PCR,第一步是从目标样本中分离RNA。初始材料典型地是分别从人类肿瘤或肿瘤细胞系,和相应的正常组织或细胞系中分离出的所有RNA。RNA可从各种样本中分离,例如来自乳房,肺,结肠(例如,大肠或小肠),结肠直肠,胃,食道,肛门,直肠,前列腺,脑,肝,肾,胰腺,脾脏,胸腺,睾丸,卵巢,子宫,膀胱等组织,原发性肿瘤或肿瘤细胞系的组织,以及来自健康供体的合并样本的肿瘤样本。如果RNA的来源是肿瘤,RNA可以从,例如,冷冻的或存档的石蜡包埋并且固定(例如,福尔马林固定的)的组织样本中提取。

[0183] 通过RT-PCR的基因表达分析的第一步是将RNA模板逆转录成cDNA,接着在PCR反应中其指数扩增。两种最常用的逆转录酶是禽成髓细胞瘤病毒逆转录酶(AMV-RT)和莫洛尼鼠白血病病毒逆转录酶(MMLV-RT)。逆转录步骤典型地依赖于不同的情况和表达谱的目的,使用特定的引物,随机六聚体,或寡-dT引物做指引。例如,提取的RNA可以按照制造商的说明使用GeneAmp[®] RNA PCR试剂盒(美国,加利福尼亚,Perkin Elmer)进行逆转录。然后获得的cDNA可以在随后的PCR反应中用作模板。

[0184] 尽管PCR步骤可以使用多种热稳定的依赖DNA的DNA聚合酶,但其通常采用Taq DNA聚合酶,其具有5'-3'核酸酶活性但缺乏3'-5'校对核酸内切酶活性。因此,TagMan[®] qPCR(美国加利福尼亚州,Roche Molecular Systems,Inc.的注册商标)通常利用Taq或Tth聚合

酶的5'核酸酶活性来水解结合到其目标扩增子的杂交探针,但是具有等效的5'核酸酶活性的任何酶都可以使用。

[0185] 两个寡核苷酸引物被用于产生PCR反应的典型特征的扩增子。第三个寡核苷酸或探针,被设计成用于检测位于两个PCR引物之间的核苷酸序列。探针是不可被Taq DNA聚合酶延伸的,并用报告物荧光染料和淬灭剂荧光染料示踪。当两种染料在探针上的位置彼此靠近时,来自报告物染料的任何激光诱发的发射由淬灭剂染料淬灭。在扩增反应过程中,Taq DNA聚合酶以依赖模板的方式裂解探针。所得探针片段在溶液中解离,并且来自释放的报告物染料的信号是免于第二荧光团的淬灭效应的。报告物染料的一个分子被释放用于每个合成的新分子,并且非淬灭的报告物染料的检测为数据的定量解释提供基础。

[0186] **TaqMan[®]** RT-PCR (美国,加利福尼亚州,Roche Molecular Systems, Inc.的注册商标)可以使用商售设备实施,例如,ABI PRISM 7700[™]序列检测系统(美国,康涅狄格,Applied Biosystems的商标)(美国,加利福尼亚州,福斯特城,Perkin-Elmer-Applied Biosystems),或Lightcycler[™](美国,加利福尼亚州,Roche Molecular Systems, Inc.的注册商标)(德国,曼海姆,Roche Molecular Biochemicals)。在优选的实施例中,5'核酸酶程序在实时定量PCR装置如ABI PRISM 7700[™]序列检测系统上运行。该系统由热循环仪,激光器,电荷耦合器件(CCD),摄像头和计算机组成。该系统以96孔格式放大热循环仪上的样本。在扩增过程中,激光诱导的荧光信号通过所有96孔的光纤电缆实时收集,并且在CCD处检测到。该系统包括运行仪器以及分析数据的软件。

[0187] 5'核酸酶测定数据最初表达为C_p,或循环阈值。如上所讨论的,荧光值在每个循环期间被记录并且代表产品的量扩增到扩增反应中的某一点。当荧光信号第一次被记录为统计显著时,该点就是循环阈值C_p。

[0188] 实时定量PCR (qRT-PCR)

[0189] RT-PCR技术最近的变化是实时定量PCR,其通过双示踪的荧光生成试剂探针(即**TagMan[®]**探针)测量PCR产物积累。实时PCR与定量竞争PCR和定量比较PCR均兼容,前者为每个目标序列的正常化使用内部竞争者,而后者使用包含在样本内的标准化基因,或RT-PCR的管家基因。进一步的细节由,例如,Held等人,Genome Research 6:986-994 (1996)提供。

[0190] 可以使用固定的石蜡包埋组织作为RNA源来确定表达水平。根据本发明的一个方面,PCR引物被设计为存在于要被扩增的基因中的侧翼内含子序列。在本实施例中,引物/探针设计的第一步是划定基因内的内含子序列。这可以通过公开可用软件来完成,例如由Kent, W. J., Genome Res. 12 (4): 656-64 (2002)开发的DNA BLAT软件,或由BLAST软件及其演变。后续步骤按照完好建立的PCR引物和探针方法进行。

[0191] 为了避免非特异性信号,在设计引物和探针时,掩盖内含子内的重复序列是有用的。这可以通过使用贝勒医学院在线可用的重复掩蔽程序很容易实现,其对重复元素库掩藏DNA序列,并返回在其中重复元素被掩蔽的查询序列。然后可以使用任何商业上或其他公开可用的引物/探针设计包,例如Primer Express (Applied Biosystems); MGB assay-by-design (Applied Biosystems); Primer3 (Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the VIMNV for general users and for biologist programmers in:

Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386), 掩蔽序列可用于设计引物和探针序列。

[0192] PCR引物设计中需考虑的最重要因素包括引物长度, 熔化温度(T_m), 和G/C含量, 特异性, 互补引物序列, 和3'末端序列。一般情况下, 最佳的PCR引物通常是长度为17-30的碱基, 并含有约20-80%, 例如, 约50-60%G+C碱基。熔化温度在50至80℃之间, 例如, 约50至70℃通常是优选的。有关PCR引物和探针设计的进一步的指导方针请见, 例如, Dieffenbach, C.W.等人, *General Concepts for PCR Primer Design in: PCR Primer, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1995, pp.133-155; Innis and Gelfand, *Optimization of PCRs in: PCR Protocols, A Guide to Methods and Applications*, CRC Press, London, 1994, pp.5-11; 以及Plasterer, T.N. *Primerselect: Primer and probe design. Methods Mol. Biol.* 70:520-527 (1997), 其整个公开都明确通过引用并入在本说明书中。

[0193] 微阵列分析

[0194] 差异表达也可以使用微阵列技术来识别, 或者证实。因此, 疾病特异性标志物的表达谱可以使用微阵列技术在新鲜或石蜡包埋的肿瘤组织中测定。在该方法中, 所关注的多核苷酸序列(包括cDNA及寡核苷酸)在微芯片底物上铺排, 或排列。然后排列的序列(即捕获探针)与来自所关注的(即, 目标)细胞或组织的特定多核苷酸杂交。正如RT-PCR法, RNA的来源通常是分离自人类肿瘤或肿瘤细胞系的全部RNA, 以及相应的正常组织或细胞系。因此RNA可以从多种原发性肿瘤或肿瘤细胞系中分离。如果RNA的来源是原发性肿瘤, 那么RNA可以, 例如, 从冷冻或存档的福尔马林固定的石蜡包埋(FFPE)组织样本和固定的(如福尔马林固定的)组织样本中提取, 在日常临床实践中, 这些样本是常规制备和保存的。

[0195] 在微阵列技术的具体实施例中, cDNA克隆的PCR扩增嵌入物施加到底物上。底物可以包括最多1, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 或75个核苷酸序列。在其它方面, 底物可以包括至少10, 000个核苷酸序列。固定在微芯片上的微阵列序列, 适合于在严格条件下的杂交。如其它实施例, 微阵列的目标可以是至少长度为50, 100, 200, 400, 500, 1000, 或2000个碱基; 或长度为50-100, 100-200, 100-500, 100-1000, 100-2000或500-5000个碱基。如其他实施例, 微阵列的捕获探针可以是至少长度为10, 15, 20, 25, 50, 75, 80, 或100个碱基; 或长度为10-15, 10-20, 10-25, 10-50, 10-75, 10-80, 或20-80个碱基。

[0196] 荧光标记的cDNA探针可通过从所关注的组织中提取的RNA的逆转录掺入荧光核苷酸产生。施加到芯片的标记的cDNA探针具有特异性地杂交到阵列上的DNA的各点。经过严格洗涤以除去非特异性结合的探针后, 该芯片通过激光共聚焦显微镜或另一种检测方法, 例如CCD照相机, 扫描。每个阵列单元的杂交定量允许评估相应的mRNA丰度。用双色荧光, 由两个RNA来源产生的分别标记的cDNA探针成对杂交到阵列。因此对应于每个指定基因的两种来源的转录物的相对丰度被同时确定。

[0197] 杂交的小型化规模为大量基因提供对表达模式方便和快速的评价。此类方法已经显示出具有检测稀有转录所需的灵敏度, 其在每个细胞中的几个拷贝上被表达, 并且可再现地检测在表达水平上至少约两倍的差异(Schena等人, *Proc. Natl. Acad. Sci. USA* 93(2): 106-149 (1996))。微阵列分析可按照生产商的协议, 通过商售设备来进行, 例如使用

Affymetrix GenChip[®] 技术, Illumina 微阵列技术或 Incyte 的微阵列技术。用于基因表达的大规模分析的微阵列方法的发展使得有可能系统地搜索多种肿瘤类型的癌症分类的分子标志物和结果预测。

[0198] RNA 分离, 纯化和扩增

[0199] mRNA 提取的一般方法在本领域中是众所周知的, 并且在分子生物学的标准教科书, 包括 Ausubel 等人, Current Protocols of Molecular Biology, John Wiley and Sons (1997) 中公开。从石蜡包埋组织中提取 RNA 的方法公开在, 例如, Rupp and Locker, Lab Invest. 56:A67 (1987), 和 De Sandres 等人, BioTechniques 18:42044 (1995)。特别地, 可以使用来自商家, 例如 Qiagen 的纯化试剂盒, 缓冲液组和蛋白酶, 根据制造商的说明进行 RNA 分离。例如, 来自体外培养的细胞的全部 RNA 可以用 Qiagen RNeasy[®] “德国, 希尔登, Qiagen GmbH 的注册商标” 微型列被分离出来。其他商售的 RNA 分离试剂盒包括 MasterPure[™] Complete DNA and RNA Purification Kit (EPICENTRE (D, Madison, WI), 和 Paraffin Block RNA Isolation Kit (Ambion, Inc.)。来自组织样本的全部 RNA 可以使用 RNA Stat-60 (Tel-Test) 进行分离。制备自肿瘤的 RNA 可, 例如, 通过氯化铯密度梯度离心被分离。

[0200] 使用固定的、石蜡包埋的组织作为 RNA 来源, 包括 mRNA 分离, 纯化, 引物延伸和扩增用于分析基因表达的有代表性的协议的步骤在各种出版的期刊文章中给出 (例如: T.E. Godfrey 等人 J. Molec. Diagnostics 2:84-91 (2000); K. Specht 等人, Am. J. Pathol. 158:419-29 (2001))。简要地说, 有代表性的过程开始于切割约 10 微米厚的石蜡包埋的肿瘤组织样本的切片。然后提取 RNA, 以及去除蛋白和 DNA。在分析 RNA 浓度后, 如果需要的话, 可以包括 RNA 修复和/或扩增步骤, 并且使用 RT-PCR 跟随的基因特异性启动子逆转录 RNA。最后, 分析数据, 以在检查肿瘤样本中鉴别的特征基因表达模式的基础上确定对患者最佳的治疗选择。

[0201] 免疫组织化学技术和蛋白质组学

[0202] 免疫组织化学方法也适用于检测本发明的增殖标志物的表达水平。因此, 抗体或抗血清, 优选多克隆抗血清, 及最优选的特定于每个标志物的单克隆抗体, 用于检测表达。抗体可通过直接标记抗体本身被检测, 例如, 用放射性标记, 荧光标记, 半抗原标记例如生物素, 或酶例如辣根过氧化物酶或碱性磷酸酶。可替代地, 未标记的初级抗体与经标记的二次抗体 (包括抗血清, 多克隆抗血清或特异于初级抗体的单克隆抗体) 结合使用。免疫组织化学协议和试剂盒在本领域中是众所周知的, 并且是商售的。

[0203] 蛋白质组学可用于在某一时间点对存在于样本 (例如, 组织, 有机体, 或细胞培养物) 中的多肽进行分析。特别是, 蛋白质组技术可用于评估样本中多肽表达的全局变化 (也称为表达蛋白质组学)。蛋白质组学分析典型地包括: (1) 通过 2-D 聚丙烯酰胺凝胶电泳 (2-D PAGE) 分离样本中的单个多肽; (2) 鉴定, 例如, 通过质谱法或 N-末端测序法从凝胶中回收的单个多肽, 和 (3) 使用生物信息学分析数据。蛋白质组学方法对于补充基因表达谱的其他方法是有价值的, 并且可以单独地或与其他方法组合地用于检测本发明的增殖标志物的产物。

[0204] 使用为标志物选择的核酸探针的杂交方法

[0205] 这些方法涉及将核酸探针结合于载体, 并且在合适的条件下与从测试样本获得的

RNA或cDNA杂交 (Sambrook, J., E Fritsch, E. and T Maniatis, Molecular Cloning: A Laboratory Manual 3rd Cold Spring Harbor Laboratory Press: Cold Spring Harbor (2001))。这些方法可以应用于来源于肿瘤组织或流体样本的标志物。所述RNA或cDNA标本通常用荧光或放射性分子标记,以使能检测和定量。在一些应用中,杂交的DNA可以用支链,荧光标记的结构进行标记,以提高信号强度 (Nolte, F.S., Branched DNA signal amplification for direct quantization of nucleic acid sequences in clinical specimens. Adv. Clin. Chem. 33, 201-35 (1998))。在通过荧光检测或凝胶图像的光密度法定量杂交的量之前,未杂交的标记通过在低盐溶液如0.1x SSC, 0.5% SDS中全面洗涤去除。该载体可以是固体,如尼龙或硝酸纤维素膜,或由在悬浮液体中时杂交的微球体或珠组成。为允许洗涤和纯化,珠子可以是磁性的 (Haukanes, B-I and Kvam, C., Application of magnetic beads in bioassays. Bio/Technology 11, 60-63 (1993))或荧光标记的,以能够使用流式细胞仪 (例如参见: Spiro, A., Lowe, M. and Brown, D., A Bead-Based Method for Multiplexed Identification and Quantitation of DNA Sequences Using Flow Cytometry. Appl. Env. Micro. 66, 4258-4265 (2000))。

[0206] 杂交技术的一个变种是QuantiGenePlex[®]测定 (美国, 加利福尼亚州, Panomics的注册商标) (Genospectra, Fremont), 其结合了荧光珠载体和分支DNA信号扩增。杂交技术的另一个变种是Quantikine[®] mRNA测定 (明尼阿波利斯, R&D System)。方法如制造商的使用说明中所描述的。简要地说, 测定使用结合到地高辛的寡核苷酸杂交探针。在颜色测定法中, 用连接到碱性磷酸酶的抗地高辛抗体检测杂交。

[0207] 其他方法在本领域是公知的, 不需要在本说明书中进一步说明。

[0208] 酶联免疫吸附测定 (ELISA)

[0209] 简言之, 在夹心ELISA测定法中, 针对标志物的多克隆或单克隆抗体结合到固体载体 (Crowther, J.R. The ELISA guidebook. Humana Press: New Jersey (2000); Harlow, E. and Lane, D., Using antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (1999))或悬浮液珠上。其他方法是本领域已知的, 不需要在本说明书中进一步说明。单克隆抗体可以是衍生于或选自噬菌体抗体库的杂交瘤 (Hust M. and Dubel S., Phage display vectors for the in vitro generation of human antibody fragments. Methods Mol Biol. 295: 71-96 (2005))。非特异性结合位点被非目标蛋白制剂和洗涤剂阻挡。然后, 捕获抗体用来自带有抗原的患者的样本或组织的制剂培养。在抗体/抗原复合物用检测目标标志物的第二抗体培养之前, 洗涤该混合物。所述第二抗体通常结合到荧光分子或其他报告物分子上, 其可以在酶反应中被检测到或用结合到报告物的第三抗体检测到 (Crowther, 同上)。或者, 在直接ELISA测定中, 含有标志物的制剂可以结合到载体或珠, 并直接用抗体-报告物结合物检测目标抗原 (Crowther, 同上)。

[0210] 用于制备单克隆抗体和多克隆抗血清的方法是本领域中是众所周知的, 不需要在本说明书中进一步说明。

[0211] 免疫检测

[0212] 该方法也可以用于来自切除肿瘤术前和术后的膀胱癌患者的血清或血浆中标志物家族成员的免疫检测, 其它癌症患者的标志物家族成员的免疫检测, 包括但不限于, 结肠

直肠癌,胰癌,卵巢癌,黑色素瘤,肝癌,食道癌,胃癌,子宫内膜癌,和脑癌,以及来自膀胱癌患者的尿液和粪便中的标志物家族成员的免疫检测。

[0213] 也可使用其他标准免疫检测技术,例如免疫印迹法或免疫沉淀法(Harlow,E.and Lane,D.,Using antibodies:a laboratory manual.Cold Spring Harbor Laboratory Press:Cold Spring Harbor(1999))检测组织或样本中的疾病标志物。在免疫印迹法中,来自组织或含有标志物的液体的蛋白制剂,在变性或非变性条件下,通过聚丙烯酰胺凝胶被电泳。然后蛋白被转移到膜载体例如尼龙。如在免疫组织化学技术中所说明的,然后该标志物直接或间接地与单克隆或多克隆抗体反应。或者,在一些制剂中,蛋白可以直接点样到膜上,而不用事先电泳分离。信号可以用光密度法进行定量。

[0214] 在免疫沉淀法中,含有标志物的可溶性制剂在针对标志物的单克隆或多克隆抗体中培养。然后该反应用由带有共价连接的蛋白A或蛋白G的琼脂糖或聚丙烯酰胺制成的惰性珠来培养。蛋白A或蛋白G珠特别地与形成结合到珠的抗体—标志物—抗原的固定复合物的抗体相互作用。洗涤后,结合的标志物可以通过免疫印迹法或ELISA检测和定量。

[0215] 基于基因型分析建立诊断

[0216] 一旦已经获得IL8Rb和可选的一个或多个进一步的癌症标志物的表达水平,那么该受试者的诊断可以被建立。如果IL8Rb的表达是在没患有炎性膀胱疾病的受试者中可见的表达之上,和/或与已知患有炎性膀胱疾病的受试者的表达水平相一致,那么受试者将被诊断为患有炎性膀胱疾病。或者,如果IL8Rb的表达不是在没患有炎性膀胱疾病的受试者中可见的表达之上,和/或在已知患有炎性膀胱疾病的受试者的表达水平之下,那么受试者将被诊断为没患有炎性膀胱疾病。

[0217] 在IL8Rb与膀胱癌的一个或多个标志物结合使用的情况下,则IL8Rb的表达水平将与没患有炎性膀胱疾病的受试者,和/或已知患有炎性膀胱疾病的受试者的表达水平进行比较。该一个或多个癌症标志物与没患有膀胱癌的受试者和/或已知患有膀胱癌的受试者的表达水平进行比较。如果IL8Rb的表达水平与没患有炎性膀胱疾病的受试者一致(小于患有炎性膀胱疾病的受试者)并且一个或多个膀胱癌标志物的表达水平与患有膀胱癌的受试者一致(差别于没患有膀胱癌的受试者),那么受试者被诊断为患有膀胱癌。如果IL8Rb的表达水平比没患有炎性膀胱疾病的受试者高(与患有炎性膀胱疾病的受试者一致)并且所述一个或多个膀胱癌标志物的表达水平与患有膀胱癌的受试者一致(差别于没患有膀胱癌的受试者),那么受试者被诊断为具有炎性膀胱疾病。如果IL8Rb的表达水平与没患有炎性膀胱疾病的受试者一致(小于患有炎性膀胱疾病的受试者)并且所述一个或多个膀胱癌标志物的表达水平与没患有膀胱癌的受试者一致(差别于确实患有膀胱癌的受试者),那么受试者被诊断为既没患有膀胱癌也没患有炎性膀胱疾病。

[0218] 因为诊断用标志物的正常和疾病的表达之间经常有表达水平的重叠,为给受试者建立诊断,典型的是建立分类阈值。分类阈值是一个值或阈值,其区分受试者到疾病类别或非疾病类别中。通常使用接受者操作特性(ROC)曲线来评估阈值,其绘制对于所有可能的阈值的灵敏度对特异性。

[0219] 诊断阈值的确定

[0220] 对于使用疾病标志物的测试,可以得出诊断阈值,其使能样本被称为该疾病的阳性或阴性,例如,膀胱癌。这些诊断阈值通过分析进行了膀胱癌或炎性膀胱疾病的存在调

查的一批患者来确定。诊断阈值可以针对不同的测试应用而变化；例如，对于在群体筛查中使用测试的诊断阈值使用基本上没有泌尿症状的一批患者确定，并且这些诊断阈值可以不同于那些用在测试处于膀胱癌复发监视中的患者的阈值。诊断阈值可以被选择以提供在所需的临床环境中的实践水平的测试特异性；即，允许合理的灵敏度而没有数目过多的接收假阳性结果的患者的特异性。这种特异性可以是在80-100%的范围内。

[0221] 诊断阈值通过将每个标志物的基因型表达水平结合的算法应用到来自前瞻性临床试验的每个样本上来确定。

[0222] 使用的样本来自膀胱癌患者和一系列非恶性泌尿紊乱患者。诊断阈值通过确定导致所期望特异性的算法的得分来选择。例如，在一些应用中，85%的特异性是所期望的。然后诊断阈值通过选择导致85%的没有膀胱癌的患者被正确归类为癌症阴性的算法得分来设置。在其他应用中（例如群体筛查），倾向于更高的特异性，例如90%。为给此应用设置阈值，导致90%的没有膀胱癌的患者被正确归类为癌症阴性的算法得分被选择。使用算法的示例在示例中概述。

[0223] 作为单一阈值的替代，测试可以使用测试间隔，其提供不同程度的疾病存在的可能性，并且具有与它们相关联的不同的临床结果。例如，测试可能有三个间隔；一个与膀胱癌存在的高风险（例如90%）相关联，第二个与膀胱癌的低风险相关联，第三个视为疾病疑似。“疑似”间隔可以与建议在所界定的时间周期内做重复测试相关联。

[0224] 数据分析

[0225] 一旦完成测试RNA和/或蛋白的量的方法，则为了确定与肿瘤和非肿瘤样本相关联的生物标志物值的分布，必须分析数据。这通常包含正规化原始数据，即，去除背景“噪音”等并计算任何双重（或以上）数据的平均值，与标准比较，并且建立界点值或阈值以最佳分离两类样本。已知许多方法能做到这一点，并且确切方法将取决于所使用的确定RNA和/或蛋白的量的具体方法。

[0226] 下面是当使用qRT-PCR时，数据分析如何被实施的示例。然而，可以理解的是，一般过程可适应于用于建立RNA和/或蛋白含量的其它方法，或者其它方法可以通过本领域的技术人员来建立以实现相同的结果。

[0227] 数据

[0228] 荧光测量在PCR的每个循环的波长 ω_i $i=1, 2$ 上进行。因此，对于每个孔，我们观察到一对荧光曲线，由 $f_t(\omega_i)$ 表示，其中 $t=1, \dots, k$ 表示循环数，并且 $i=1, 2$ 索引波长。

[0229] 荧光曲线有一个S形的形状，其开始于邻近的水平基线并且平滑地增加至上渐近线。荧光曲线与直线基线分开的点 C_p 的位置将被用于表征目标基因的浓度。 C_p 的准确定义如下。

[0230] 下面是处理这些数据的方案的示例。

- [0231] • 补偿频段之间的荧光重叠，
- [0232] • 为每个荧光曲线评估平滑模型以估算 C_p
- [0233] • 组合来自复制的孔的数据。
- [0234] • 估算标准曲线
- [0235] • 计算相对于标准的浓度。

[0236] 每个生物样本产生5个基因的相对浓度，它们是判别函数的输入数据。

[0237] 色彩补偿

[0238] 用 $W_{tj}(\omega)$ 表示在循环 t 与频率 ω 的染料 j 的荧光水平。在复用测定中,在任何频率 ω 的测量响应是在该频率的所有染料的贡献的总和,对于每个循环均是如此。

[0239] $f_t(\omega) = W_{t1}(\omega) + W_{t2}(\omega) + \dots$

[0240] 色彩补偿的目的是从所观察到的混合物 $f_t(\omega)$ 中提取个体贡献 $W_{tj}(\omega)$ 。

[0241] 在理想的情况下,由于在频率 ω 的染料 j 的荧光 $W_{tj}(\omega_0)$ 与其在参考频率 ω_0 的荧光 $W_{tj}(\omega_0)$ 成比例,而不管 $W_{tj}(\omega_0)$ 的水平。这表明对于某些待确定的比例常数 A_{12} 和 A_{21} 是线性关系。

$$[0242] \begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} W_{t1}(\omega_1) + W_{t2}(\omega_1) \\ W_{t1}(\omega_2) + W_{t2}(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix}$$

[0243] 实际上有通过引入线性项到该系统有效地建模的额外效果,所以

$$[0244] \begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix} + \begin{bmatrix} a_1 + b_1 t \\ a_2 + b_2 t \end{bmatrix}$$

[0245] 尽管被线性基线扭曲,在估计“色彩补偿”参数 A_{12} 和 A_{21} 后,我们可以通过矩阵乘法恢复 $W_{t1}(\omega_1)$ 和 $W_{t2}(\omega_2)$:

$$[0246] \begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} + \begin{bmatrix} a_1^* + b_1^* t \\ a_2^* + b_2^* t \end{bmatrix}$$

[0247] $W_{t1}(\omega_1)$ 和 $W_{t2}(\omega_2)$ 被称为“颜色补偿”数据。当评估低于2的色彩补偿数据的模型时,此表达式的最后一项的线性失真 $a_i^* + b_i^* t$ 将被容纳在基准估算内。它对于 C_p 估算没有影响。

[0248] 色彩补偿系数的估算需要使用单一(与双重相对)探针的单独测定。则 $W_{t2}(\omega_2) = 0$,得出:

$$[0249] \begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 + b_1 t \\ a_2 + b_2 t \end{bmatrix}$$

[0250] 因此,

$$[0251] f_t(\omega_2) = A_{21} f_t(\omega_1) + a_2^* + b_2^* t$$

[0252] 系数 A_{21} 可以通过 $f_t(\omega_2)$ 关于 $f_t(\omega_1)$ 和PCR循环 t ,其中 $t=1, \dots, k$ 的普通线性回归来估算。

[0253] 模型估计

[0254] 在这一部分中,让 y_t $t=1, \dots, k$ 表示色彩补偿的荧光曲线。

[0255] 扩增

[0256] 模型只是为荧光曲线而估计的,该荧光曲线显示不平凡的扩增。我们定义术语“扩增”为从色彩补偿的荧光曲线的线性基线的非平凡偏差。使用信噪比(SNR)来定量扩增。这里的SNR被定义为信号方差与噪声方差的比值。噪声方差设定为测定程序校准的一部分,并且保持不变:为此目的,使用来自对于没有扩增的孔,即没有RNA的孔的基线的线性模型的剩余方差,。对于每一个荧光曲线,估计信号方差作为来自最佳拟合直线的剩余方差(“最佳”是指在最小平方意义上)。

[0257] • 如果SNR低于指定的阈值,则荧光曲线接近线性和无扩增存在。然后没有从基线

离开的点并且样本中的浓度可被定为零。

[0258] • 如果SNR高于阈值,则扩增存在并且浓度可以估算。

[0259] 选择(无量纲) SNR的阈值以提供“扩增”和“非扩增”曲线之间的明确区别。例如,阈值的以下范围是对标志物是有效的。

[0260]	荧光	基因	范围
	JOE	MDK	40-120
	JOE	CDC	35-70
	JOE	IL8R	30-60 5
	FAM	IGF	50-80
	FAM	HOXA	50-150
	FAM	XENO	50-80

[0262] 模型

[0263] 为每个荧光曲线估计S形模型。模型的任何合适的参数形式都可以使用,但必须能够将以下特征模型化:

[0264] • 可以具有非零斜率的线性基线,

[0265] • 关于中点不对称。

[0266] • 在下部和上部水平的渐近线

[0267] • 从基线到上渐近线的平滑增加

[0268] 实现这些要求的模型的一个示例是

$$[0269] \quad g_t(\theta) = A + A_s t + \frac{D}{\left(1 + \left(\frac{t}{B}\right)^E\right)^F}$$

[0270] 我们称此为“6PL模型”。参数向量 $\theta = [A, A_s, D, B, E, F]$ 是受以下限制的,以确保 $g_t(\theta)$ 为 t 的递增函数并且具有荧光曲线的实证性质。

[0271] $D > 0, B > 0, E < 0, F < 0$

[0272] 其他两个参数确定基准线 $A + A_s t$,并且这些参数并不需要明确的约束,但是 A 始终为正并且坡度参数 A_s 总是很小。参数 D 确定基线以上的扩增水平。剩余的参数 B, E, F 本身没有固有的解释而是控制曲线的形状。这些参数也是仅有的影响 C_p 的估算的参数。当 $A_s = 0$,这是众所周知的5参数逻辑函数(5PL),如果,另外, $F = 1$,这个模型降低为4参数逻辑模型(4PL),Gottschalk and Dunn(2005),Spiess等人(2008)。

[0273] 初始化

[0274] 对于非线性估计初始值设定为

[0275] • $A_s = 0, F = 1$

[0276] • $A = \text{mean}(y, \dots, y_5)$

[0277] • $D = \text{range}(y_1, \dots, y_k)$

[0278] • B=对应于半高的循环

[0279] • E通过转换 $g_t(\theta)$ 为线性形式进行初始化,该线性形式已设置其余参

[0280] 数的值为上述定义的它们的初始值。线性化获得

$$[0281] \quad E \log \left(1 + \frac{t}{B} \right) = \log \left(\frac{D}{y_t - A} \right)$$

[0282] 现在,对于所选择的 t ,通过对 $\log \left(1 + \frac{t}{B} \right)$ 回归 $\log \left(\frac{D}{y_t - A} \right)$ 估算 E ,使得

$$[0283] \quad A + \frac{D}{10} < y_t < A + \frac{9D}{10}$$

[0284] 导致(与其自身的初始化)几乎相同的分析的该模型的另一种形式是:

$$[0285] \quad A + A_s t + \frac{D}{\left(1 + \exp \left(-\frac{t - B}{E} \right) \right)^F}$$

[0286] 当 $A_s=0$ 时,有时这被称为Richards方程,Richards (1959)。

[0287] 估算准则

[0288] 估算参数,以最小化惩罚平方和的准则:

$$[0289] \quad \sum_t (y_t - g_t(\theta))^2 + \lambda(\theta)$$

[0290] 这里 $\lambda(\theta)$ 是惩罚 θ 中一些(或所有)参数的大值的非负函数。此方法被称为正则化回归或岭回归(Hoerl, 1962),并且可以通过设置参数向量 θ 的合适的先验分布从贝叶斯观点导出。惩罚的满意选择是:

$$[0291] \quad \lambda(\theta) = \lambda (B^2 + D^2 + E^2 + F^2)$$

[0292] 大值的 λ 使参数估计值倾向于零,并降低了参数估计值的变化。相反,小的(或零) λ 导致不稳定的参数估计值和最小化算法收敛困难。 λ 的选择是偏差和方差或稳定性之间的折衷。经验证据表明,如果 λ 在以下范围内选择,可以实现偏差和方差之间满意的折衷:

$$[0293] \quad 0.01 > \lambda > 0.0001$$

[0294] 这个选择也保证了优化算法的收敛性。

[0295] 算法的选择

[0296] 对于 λ 在上述范围内的任何选择,在前面的段落中的描述完全定义了参数估计。基于传统的高斯-牛顿法(如在More中实现的Levenberg-Marquardt算法,1978)的非线性最小平方法已成功地被使用并且是合适的方法。优化算法的一般目的,例如Nelder and Mead, 1965,或Byrd等实现的Broyden-Fletcher算法,1995)也已成功地在本说明书中被试验。

[0297] C_p 估计

[0298] C_p 是在时间 t 的点,其最大化了 $g_t(\theta)$ 的二阶导数。每个荧光曲线产生一个 C_p ,其表征目标基因的浓度。每组技术复制的所估计的 C_p 的平均值被计算并在随后的分析中使用。

[0299] 标准曲线

[0300] 绝对浓度或相对浓度从与同一PCR板的标准曲线进行比较而得出。使用线性模型对稀释系列建模:

[0301] $C_p=R+S \log_{10}Conc$

[0302] 其中,Conc是该标准的绝对浓度或相对浓度。截距和斜率参数是针对具体板的。通过设置群体模型对截距和斜率参数中板之间的变异建模:

[0303] $R \sim N(\mu_R, \sigma_R^2)$

[0304] $S \sim N(\mu_S, \sigma_S^2)$

[0305] 其中,参数 $\mu_R, \sigma_R^2, \mu_S, \sigma_S^2$ 如下所述在现有数据的基础上设置。然后给定的板R和S可以被解释为来自这些群体的观察。

[0306] 对于在浓度Conc_j的标准的仿样i,以下模型可使用:

[0307] $C_p(i, j) = R + S \log_{10}Conc_j + \epsilon_{ij}$

[0308] 其中, $\epsilon_{ij} \sim N(0, \sigma_{\epsilon}^2)$ 。需要注意的是残差的方差取决于C_p。Var(ϵ_{ij})的经验估计见表2。通过最大化似然函数估计参数R和S。通过以下表达式根据PCR过程的效率解释斜率参数:

[0309]
$$S = - \frac{1}{\log_{10} Efficiency}$$

[0310] 此模型有贝叶斯解释:给定参数 $\mu_R, \sigma_R^2, \mu_S, \sigma_S^2$ 模糊(无信息)的先验分布。然后R和S和C_p(i, j)的群体模型完全确定了现有数据的概率模型。马尔科夫链蒙特卡罗(MCMC)算法(Lunn等人,2009)允许对 $\mu_R, \sigma_R^2, \mu_S, \sigma_S^2$ 的估算。如果省略了先验分布,会得到传统的频率论解释结果。依照此估计程序,能够获得表3中的依赖基因的群体参数估计。

[0311] 表2:残差方差

[0312]

C _p	σ ²	C _p	σ ²
12	0.0100	21	0.0466
13	0.0108	22	0.0625
14	0.0119	23	0.0860
15	0.0134	24	0.1212
16	0.0155	25	0.1750
17	0.0184	26	0.2591
18	0.0224	27	0.3931
19	0.0279	28	0.6112
20	0.0356	29	0.9741

[0313] 表3:标准曲线的斜率和截距的总体参数

[0314]		μ_R	σ_R^2	μ_S	σ_S^2
	MDK	19.49	0.5112	-3.426	0.0481
	CDC	18.91	0.2343	-3.414	0.0198
	IL8R	31.43	0.0919	-3.192	0.0017
	IGF	20.63	0.3835	-3.275	0.0247
	HOXA	22.51	0.1544	-3.270	0.0037

[0315] 标准曲线的截距和斜率的估计由 $\hat{R}$ 和 $\hat{S}$ 表示。

[0316] 相对浓度 ΔC_p

[0317] 使用标准曲线由表达式 $C_{p(REF)} = \hat{R} + \hat{S} \log_{10} Conc_{REF}$ 来计算浓度 $Conc_{REF}$ 下的 $C_{p(REF)}$ 。样本的相对浓度由以下表达式给出：

$$[0318] \quad \Delta C_p = \frac{C_p - C_{p(REF)}}{\hat{S}} = \log_{10} \frac{Conc_{SAMPLE}}{Conc_{REF}}$$

[0319] 或者, $\hat{S}$ 可在对应于PCR效率为2的固定水平来近似。则 $\hat{S} = -1/\log_{10}(2) = -3.32$ 。为任何一种选择使用相同的符号 ΔC_p 。所得的 ΔC_p 估计, 每个基因一个, 作为下一步中的判别函数的输入值。

[0320] 判别函数

[0321] 所述 ΔC_p 值对应于去除板到板变化的相对生物标志物值。5个 ΔC_p 值的彼此对照检查 (例如, 参见图2), 显示了肿瘤样本通常是如何与肺肿瘤样本具有不同的生物标志物值的。此外, 虽然肿瘤和正常组织有区域重叠, 但是大量样本被有效分离。在这些情况下, 许多不同的统计分类器可用于从肿瘤样本中分离出正常样本来。我们在这里展示了几个分类器的示范运行以分离这些样本。我们使用5种不同的分类方法: 1) 线性判别分析 (LDA); 2) 逻辑回归 (LogReg); 3) 支持向量机 (SVM); 4) 基于5个近邻的K最近邻 (KN5N); 5) 递归划分树 (TREE) (引用: Venables&Ripley and Dalgaard)。

[0322] 分类器的创建需要包含大量样本的生物标志物值的数据集, 该数据集应代表由分类器测试的最终群体。例如, 如果分类器是用于筛选高危人群的 (如50岁和更老, 吸烟者), 则创建分类器所需的数据集 (被称为“训练集”) 应反映该群体并仅包含来自大于50岁的吸烟者的样本。通常, 为得到像灵敏度和特异性这样的参数的误差小于10%的测量精度, 训练集需要大于300个样本。

[0323] 分类器的有效性的估计可使用交叉验证进行。在交叉验证中 (维基百科: 交叉验证), 数据集被划分成小数目的相等大小的分区 (通常为3至10个)。一部分留出来, 剩下的部分用来建立分类器; 然后留出的部分由新的分类器测试并记录下它的预测。对每一个部分都依次这样做, 并且所有的预测组合并进行分析, 以计算出分类器的特征: 灵敏度, 特异性, 等等。如果交叉验证是通过将数据划分成10份进行的, 它被称为10倍交叉验证; 同样地, 3份将是3倍交叉验证。如果有多少样本数据就被分成多少份, 那么这就是所谓的“留一交叉验证”。通过测试不用于构建分类器的数据, 此方法提供了在不存在附加样本的情况下, 对分类器性能的估计。

[0324] 我们已经使用4个生物标志物,MDK,IGFBP5,CDC2,和HOXA13,加上和不加生物标志物IL8Rb的所有15种组合,利用本说明书中其他部分描述的临床试验数据集(示例1)建立了分类器,并且用10倍交叉验证测试了这30个分类器。这针对上面列出的5个分类器类型的每一个完成并计算出ROC曲线。所有工作使用R统计编程环境(CITE)进行改善。这些结果(图14)表明,在大多数情况下,加上IL8Rb的分类器对诊断有用的特异性的值(假阳性率为0至20%;特异性为100至80%)更为敏感。诊断有用的特异性区域的曲线下面积(AUC)用于定量分类器如何与大的值良好运行的,其表示更好的分类器性能。图15a列表示出了每个分类器和生物标志物组合的AUC,而图15b示出了当加入IL8Rb时,每种条件下AUC的增加量。在大多数情况下,加入IL8Rb提高了做出准确诊断的能力。在图16中用表格列出了所有分类器的对诊断有用的特异性值的具体的灵敏度值。此外,图17以表格列出了加入IL8Rb后,灵敏度或特异性的增益量。

[0325] 分类器的效用被创建,当创建完分类器并对其测试完,它用来测试新的样本。为简化结果的解释,建立临界分数或阈值;在临界分数一侧的样本被认为是肿瘤阳性,而在另一侧的为肿瘤阴性。附加的临界分数可以建立,例如,以指示结果的确定性水平不断提高。在这种情况下,我们已经建立了临界分数,其在我们的训练集中给出了15%的假阳性率。使用我们的交叉验证过的ROC曲线,则我们可以估计我们的敏感性。通常情况下,发明人还在75%的阳性预测值处建立临界分数。为使用这些临界分数,我们为小于由85%的特异性建立的临界分数的分数建立了“阴性”结果。得分高于75%PPV的被称为“阳性”,得分在两者之间的被称为“不确定”或“疑似”。

[0326] IL8Rb的抗体

[0327] 在另外的方面,本发明包括针对IL8Rb的抗体制造。标志物IL8Rb可以被足量制备以适合于引发免疫反应。在一些情况下,全长IL8Rb可以使用,而在其他情况下,IL8Rb的肽片段可以是足够作为免疫原的。免疫原可以注入合适的宿主(例如,小鼠,兔等),如果想要的话,免疫佐剂,如弗氏完全佐剂或弗氏不完全佐剂可以注射,以增加免疫反应。可以理解的是,制备抗体在免疫学领域中是常规的并且不需要在本说明书中进一步描述。其结果是,可以产生针对IL8Rb的抗体,包括单克隆抗体或噬菌体抗体。

[0328] 仍然在进一步的实施例中,抗体可以针对该蛋白或本说明书中确认的肿瘤标志物的蛋白核或IL8Rb独特的寡核苷酸序列制造。虽然某些蛋白可以被糖基化,但是糖基化模式的变化可以,在某些情况下,导致误检测缺乏寻常糖基化模式的IL8Rb的形式。因此,在本发明的某些方面,IL8Rb免疫原可以包括去糖基化IL8Rb或去糖基化IL8Rb片段。可以使用本领域中已知的一种或多种糖苷酶来实现去糖基化。可替代地,IL8Rb cDNA可在糖基化缺陷型细胞系,例如原核细胞系,包括大肠杆菌等中表达。

[0329] 可制成在其中具有IL8Rb编码的寡核苷酸的表达载体。许多这样的载体可基于本领域中已知的标准载体。载体可用于转染各种细胞系,以产生IL8Rb-产生的细胞系,其可以被用于生产期望量的IL8Rb以发展特定抗体或其它检测IL8Rb的其他试剂或用于标准化已有的IL8Rb测定法的其他试剂。

[0330] 试剂盒

[0331] 基于本发明的发现,几种类型的检测试剂盒可以设想和制备。首先,可以制成具有预加载有检测分子(或“捕获试剂”)的检测装置的试剂盒。在检测IL8Rb mRNA的实施例中,

这种设备包括底物(例如,玻璃,硅,石英,金属等),在其上与待检测的mRNA杂交的作为捕获试剂的寡核苷酸被绑定。在一些实施例中,mRNA的直接检测可以通过将mRNA(用cy3,cy5,放射性或其它标记来标记)杂交到底物上的寡核苷酸来实现。在其他实施例中,mRNA的检测可以通过首先制备与所需的mRNA互补的DNA(cDNA)来完成。然后,标记的cDNA可以杂交到底物上的寡核苷酸,并被检测。

[0332] 抗体也可以在试剂盒中作为捕获试剂使用。在一些实施例中,底物(例如,多孔板)可以具有附接到其上的特定IL8Rb和BTM捕获试剂。在一些实施例中,试剂盒可包括封闭试剂。封闭试剂可用于减少非特异性结合。例如,非特异性的寡核苷酸的结合可使用来自不包含IL8Rb和BTM寡核苷酸的任何方便来源的过量的DNA被减少,例如鲑鱼精子DNA。非特异性抗体结合可以使用过量的阻断蛋白,例如血清白蛋白来降低。可以理解的是,用于检测寡核苷酸和蛋白的许多方法在本领域中是已知的,并且可使用可特定检测标志物相关联分子的任何策略并且其被认为落入本发明的范围内。

[0333] 当结合到固体载体上时,例如抗体芯片,其允许采用单个芯片检测多个标志物,也可以使用抗体。

[0334] 除了底物,测试试剂盒可以包含捕获试剂(如探针),洗涤溶液(例如,SSC,其它盐,缓冲剂,去污剂等),以及检测基团(例如,cy3,cy5,放射性标记等)。试剂盒还可以包括使用说明和包装。

[0335] 可以使用任何合适的技术来实施检测样本中的IL8Rb和BTM,并且可以包括,但不限于,寡核苷酸探针,qPCR或针对癌症标志物产生的抗体。

[0336] 应该理解的是,待测试的样本并不限于疑似为炎性疾病或肿瘤的组织样本。该标志物可被分泌到血清或其他体液中。因此,样本可以包括任何身体的样本,并且包括活检,血液,血清,腹腔冲洗液,脑脊髓液,尿液和粪便的样本。

[0337] 还应当理解,本发明不限于检测人类的癌症,还适用于检测任何动物的癌症,包括但不限于狗,猫,马,牛,羊,鹿,猪和已知患有癌症的任何其他动物。

[0338] 体液中炎性疾病或肿瘤标志物的常规测试

[0339] 一般而言,在这些体液中测定寡核苷酸,蛋白和肽的方法在本领域中是公知的。寡核苷酸的检测可以使用杂交的方法来进行,例如Northern印迹,Southern印迹或微阵列法,或qPCR。用于检测蛋白的方法包括如酶联免疫吸附测定(ELISA),具有抗体的蛋白芯片,悬浮珠粒放射免疫分析法(RIA),免疫印迹和凝集素结合。然而,为了说明的目的,疾病标志物的液体水平可以使用夹心型酶联免疫吸附测定(ELISA)定量。对于血浆测定法,适当稀释的样本的或连续稀释的标准标志物的5微升等分试样和75微升过氧化物酶抗人标志物抗体添加至微量滴定板的孔中。在30℃下30分钟培养时段后,这些孔用0.05%吐温20的磷酸盐缓冲液(PBS)冲洗,以除去未结合的抗体。然后标志物和抗标志物抗体的结合复合物在30℃下用含有H₂O₂的邻苯二胺培养15分钟。该反应通过加入1M硫酸终止,并且用微量滴定板读数器测量492nm处的吸光率。

[0340] 可以理解的是,抗IL8Rb抗体可以是单克隆抗体或多克隆抗血清。也可以理解的是,任何其他体液可以适当地进行研究。

[0341] 在生理意义上,被分泌的标志物不必要是有用的。相反,标志物蛋白或基因进入血清的任何机制,对生产所述可检测,可量化水平的标志物都是有效的。因此,细胞中的可溶

性蛋白的正常分泌物,从质膜脱落的膜蛋白,以其形式表达的分泌的可变剪接形式的mRNA或蛋白,细胞死亡(或凋亡)可以产生足够水平的有用标志物。

[0342] 对于使用血清标志物作为工具来诊断和/或评估各种类型癌症的治疗功效的支持越来越多。

[0343] 示例

[0344] 本说明书中描述的示例是为了说明本发明实施例的目的。分析的其他实施例,方法和类型都在分子诊断领域的普通技术人员的范围之内,并不需要在此详细描述。本领域范围内的其他实施例被认为是本发明的一部分。

[0345] 示例1:膀胱癌的基因型分析

[0346] 方法

[0347] 患者:2008年4月至2009年9月期间,485例具有肉眼可见血尿,但没有尿路恶性肿瘤史的患者,在新西兰和澳大利亚的11个泌尿科诊所被招募。紧邻接受膀胱镜检查 and 任何附加的诊断程序前,每个患者提供一份尿样。进入研究三个月后,做出了诊断。这些485例患者中,使用下述方法成功地获得442例患者的所有五个研究基因的基因表达数据。这些患者的特征示于表4。

[0348] 表4:研究人群I的特点

[0349]

诊断	数量
良性前列腺增生症	18
膀胱炎	18
运动性血尿	3
非特异性肾脏疾病	3
非特异性瘤	3

[0350]

非特异性前列腺疾病	63
前列腺血管	49
其它泌尿癌 (非 TCC)	5
浅表血管	3
尿道狭窄	6
尿路感染	18
尿石症	25
华法林的使用	10
病因不明	155
混杂的	7
TCC	56
总计	442

[0351] 表4示出在患者初次出现显见血尿后的三个月时,每个主要诊断类别中患者的数量。

[0352] 尿液分析:尿样用中央审查细胞学检查(新西兰,达尼丁,Southern Community Laboratories)进行分析。诊断测试NMP22BladderChek[®] (Matritech)和NMP22 ELISA (Matritech)按照制造商的说明在临床现场(BladderChek[®])或由Southern Community Laboratories (NMP22 ELISA)实行。

[0353] RNA定量:每个患者的2mls或尿液用含5.64M异硫氰酸胍,0.5%十二烷基肌氨酸钠和50mM pH6.5醋酸钠的RNA提取缓冲液混合。然后,如前所述,所有RNA由Trizol萃取(Invitrogen)和RNeasy程序(Qiagen)提取。RNA从35ul水的柱中洗脱,并在每个后续单工或双工定量逆转录聚合酶链式反应(qRT-PCR)测定中使用3ul。每16ulqRT-PCR反应含有0.3U RNase-OUT (Invitrogen),0.225uM每个Taqman探针,1.25U Superscript III (Invitrogen),0.275uM每种引物,1.5U Fast Start Taq聚合酶(Roche),10mM DTT,0.375mM dNTP,4.5mM硫酸镁,1.6ul 10x Fast Start PCR缓冲液(Roche)和2.6ul GC Rich 溶液(Roche)。针对五个研究基因MDK,CDC2,HXA 13,IGFBP5和IL8Rb中的每一个的引物和荧光双重标记的探针是从Integrated DNA Technologies(美国,Coralville)获得的。引物/探针序列示于表2。反应设置在96孔板上并且在Roche LightCycler[®]480上做如下循环:50℃,15分钟;95℃,8分钟;95℃,15秒的10个循环,60℃,2分钟和95℃,15秒的30个循环,60℃,1分钟。参照RNA(从汇集的细胞系衍生的RNA)的1/16系列稀释的标准曲线包括在每块板上以产生0.3pg/μl至20ng/μl的范围。在最后30个热循环的扩展相位上收集数据并作为原始文本文件输出。下面的表5示出了用于五个RNA标志物的qRT-PCR定量的引物和探针序列。

[0354] 表5

[0355]

标志物	前向序列	反向序列	探针
MDK	TGC ACC CCC AAG ACC AAA (序列号 3)	TGA TTA AAG CTAC ACG AGC AGA CAG AA (序列号 4)	CCT TCC CTT TCT TGG CU TGG CCT TT (序列号 5)
IGFBP5	CGT TGT ACC TGC CCA ATT GTG A (序列号 6)	GGG ACG CAT CAC TCA ACG TT (序列号 7)	AAG AGA AAG CAG TGC AAA CCT TCC CGT (序列号 8)
CDC2	GCC GCC GCG GAA TAA T (序列号 9)	TGT CTA CCC TTA TAC ACA ACT CCA TAG G (序列号 10)	AGC CGG GAT CTA CCA TAC CCA TTG ACT AAC T (序列号 11)
HOXA13	TGG AAC GGC CAA ATG TAC TG (序列号 12)	TGG CGT ATT CCC GTT CAA GT (序列号 13)	ACT CTG CCC GAC GTG GTC TCC CA (序列号 14)
IL8Rb	CCT TGA GGC ACA GTG AAG ACA TC (序列号 15)	CCT GTA GGA CAC CTC CAG AAG AG (序列号 16)	TGG CCA CTC CAA TAA CAG CAG GTC ACA (序列号 17)

[0356] qRT-PCR数据分析

[0357] 原始荧光数据从RocheLightCycler[®] 480导出为制表符分隔文件,其包含循环次数对板上所有孔的荧光数据的两个通道。该数据用施加色彩补偿 ([Bernard1999]) 到数据以校正从一个荧光通道渗色到另一个荧光通道的R程序进行处理。然后,它拟合了百分之五的逻辑模型以使用最大二阶导数法 ([Spiess2008]) 来估计C_p。

[0358] 所有样本和对照组复制地施加到PCR板上。复制孔中的C_p值在使用前取平均值。如果两个C_p值之间的差超过了3个单位,重复该样本。为提供跨越PCR板的标准化,C_p值表示为相对于20ng/μl参照RNA的ΔC_p值(从汇集的细胞系衍生的RNA):

[0359] $\Delta C_p = C_p(\text{样本}) - C_p(\text{参照RNA})$

[0360] 统计分析

[0361] MDK, CDC2, HOXA13, IGFBP5和IL8Rb的qRT-PCR ΔC_p值用于生成分类器,以基于线性

判别分析或逻辑回归 ([Venables2002]) 分离含有TCC的样本和不含有TCC的样本。在这两种情况下,基因之间的相互作用在分类器模型中被允许。LDA的生成遵循标准程序,例如在“Modern Applied Statistics with S,4th edition”by W.N.Venables and B.D.Ripley (2002),Springer中所述。任何不完整的数据被清理由研究而来的数据集,然后R统计环境(R Development Core Team(2009))和来自MASS包的函数“lda”(Venables and Ripley (2002))用于生成和测试临床试验数据的线性判别。

[0362] 逻辑回归分类器的生成以类似于生成LDA的方式实施。同样,不完整的数据被清理出研究数据。逻辑回归分类器使用R创建;不需要额外的软件包。逻辑回归依照如Dalgaard (2008)所述实施。分类器间的比较采用ROC曲线,使用R包,ROCR(Sing等人,2009)进行。使用Macskassy等人([Macskassy2005])的方法生成ROC曲线的置信区间。生成以下算法:

[0363] 线性判别分类器

[0364] 第一个分类器,线性判别,(称为LDA-3),是基于五个基因值(通过减去参考值归一化到参考值),其允许基因之间的多路相互作用。该分类器是使用从名为“MASS”包中的'lda()'函数内置在R中(R version 2.9.1;MASS version 7.2-49)。该分类器是使用下面的方程式建立的:

[0365] `lda3<-lda(TCC.YN~MDK*IGF*CDC*HOXA*IL8R,data=uRNA.Trial)`

[0366] 其中lda3是创建的模型;TCC.YN为由膀胱镜测定的“尿液中存在TCC”的真值(是或否);MDK,IGF,CDC,HOXA和IL8R是归一化的基因Cp值;uRNA Trial是包含来自临床试验的5个基因和TCC.YN(是或否)中的每一个的Cp值的数据文件。公式中星号,'*',的使用表示乘法。分类器得分的估计过程为输入含有五个基因值以及分类器,lda3,的新数据帧,以输出分类器得分(score):

[0367] `Score<-c(predict(lda3,new.data)$x)`

[0368] 其中,“score”是分类器用于预测TCC存在的输出;“lda3”是上面创建的分类器并且“new.data”是包含五个基因的测定值的数据文件,这五个基因用与在分类器创建中所用的相同的名字。该句法,`\$x'和“c(...)”的存在是为了从由预测函数返回的大量信息中特定提取得分。设置得分的临界分数为0.112及以上,为尿样中存在TCCs设置我们的特异性为85%。LDA-3的系数示出在表6:

[0369] 表6

	MDK.d.R100	5.333639e+00
	IGF.d.R100	3.905978e+00
	CDC.d.R100	6.877143e-01
	HOXA.d.R100	6.073742e+00
	IL8R.d.R100	-1.229466e+00
	MDK.d.R100:IGF.d.R100	-7.420480e-01
	MDK.d.R100:CDC.d.R100	-2.611158e-01
	IGF.d.R100:CDC.d.R100	-1.965410e-01
	MDK.d.R100:HOXA.d.R100	-8.491556e-01
	IGF.d.R100:HOXA.d.R100	-4.037102e-01
	CDC.d.R100:HOXA.d.R100	-3.429627e-01
	MDK.d.R100:IL8R.d.R100	1.903118e-01
	IGF.d.R100:IL8R.d.R100	2.684005e-01
	CDC.d.R100:IL8R.d.R100	-1.229809e-01
	HOXA.d.R100:IL8R.d.R100	2.909062e-01
[0370]	MDK.d.R100:IGF.d.R100:CDC.d.R100	4.108895e-02
	MDK.d.R100:IGF.d.R100:HOXA.d.R100	7.664999e-02
	MDK.d.R100:CDC.d.R100:HOXA.d.R100	4.832034e-02
	IGF.d.R100:CDC.d.R100:HOXA.d.R100	2.116340e-02
	MDK.d.R100:IGF.d.R100:IL8R.d.R100	-3.750854e-02
	MDK.d.R100:CDC.d.R100:IL8R.d.R100	1.664612e-02
	IGF.d.R100:CDC.d.R100:IL8R.d.R100	2.089442e-03
	MDK.d.R100:HOXA.d.R100:IL8R.d.R100	-1.539486e-02
	IGF.d.R100:HOXA.d.R100:IL8R.d.R100	-3.894153e-02
	CDC.d.R100:HOXA.d.R100:IL8R.d.R100	6.295032e-03
	MDK.d.R100:IGF.d.R100:CDC.d.R100:HOXA.d.R100	-4.359738e-03
	MDK.d.R100:IGF.d.R100:CDC.d.R100:IL8R.d.R100	-2.019317e-04
	MDK.d.R100:IGF.d.R100:HOXA.d.R100:IL8R.d.R100	3.746882e-03
	MDK.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	-2.902150e-03
	IGF.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	4.799489e-04
	MDK.d.R100:IGF.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	7.512308e-05

[0371] 逻辑回归分类器

[0372] 基于逻辑回归的第二个分类器来源于和LDA-3相同的清理过的数据集。然而代替使用lda()函数,我们使用统计包中的glm()函数(包括带有R的基础安装),如下所示:

[0373] `lr1<-glm(TCC.YN~CDC*IGF*HOXA*IL8R*MDK,`

[0374] `family=binomial("logit"),data=uRNA.Trial),`

[0375] 其中,“lr1”是所创建的分类器,其它参数是如针对线性判别所描述的。再次,使用‘*’运算符指定全面互动。分类是以非常类似于对LDA-3的方式进行:

[0376] `score<-predict(lr1,new.data,type='response'),`

[0377] 其中,“score”是用于基于“new.data”中五个基因的测量值分类尿样的值,如上。lr1的临界分数被设定为0.102,以达到85%的特异性;临界分数的值被认为是TCC阳性。分类器的系数是:

[0378] -103.0818143+

[0379] 3.9043769*CDC.d.R100+

[0380] 13.1120675*IGF.d.R100+

[0381] 17.4771819*HOXA.d.R100+

[0382] -10.7711519*IL8R.d.R100+

[0383] 21.1027595*MDK.d.R100+

[0384] -0.5938881*CDC.d.R100*IGF.d.R100+

[0385] -1.0736184*CDC.d.R100*HOXA.d.R100+
[0386] -1.3340189*IGF.d.R100*HOXA.d.R100+
[0387] 0.3126461*CDC.d.R100*IL8R.d.R100+
[0388] 1.4597355*IGF.d.R100*IL8R.d.R100+
[0389] 1.8739459*HOXA.d.R100*IL8R.d.R100+
[0390] -1.035054*CDC.d.R100*MDK.d.R100+
[0391] -2.5885156*IGF.d.R100*MDK.d.R100+
[0392] -2.7013483*HOXA.d.R100*MDK.d.R100+
[0393] 1.4546134*IL8R.d.R100*MDK.d.R100+
[0394] 0.0767503*CDC.d.R100*IGF.d.R100*HOXA.d.R100+
[0395] -0.0663361*CDC.d.R100*IGF.d.R100*IL8R.d.R100+
[0396] -0.1015552*CDC.d.R100*HOXA.d.R100*IL8R.d.R100+
[0397] -0.2110656*IGF.d.R100*HOXA.d.R100*IL8R.d.R100+
[0398] 0.1361215*CDC.d.R100*IGF.d.R100*MDK.d.R100+
[0399] 0.1601118*CDC.d.R100*HOXA.d.R100*MDK.d.R100+
[0400] 0.259745*IGF.d.R100*HOXA.d.R100*MDK.d.R100+
[0401] -0.0106468*CDC.d.R100*IL8R.d.R100*MDK.d.R100+
[0402] -0.1947899*IGF.d.R100*IL8R.d.R100*MDK.d.R100+
[0403] -0.185286*HOXA.d.R100*IL8R.d.R100*MDK.d.R100+
[0404] 0.0136603*CDC.d.R100*IGF.d.R100*HOXA.d.R100*IL8R.d.R100+
[0405] -0.0151368*CDC.d.R100*IGF.d.R100*HOXA.d.R100*MDK.d.R100+
[0406] 0.0056651*CDC.d.R100*IGF.d.R100*IL8R.d.R100*MDK.d.R100+
[0407] 0.0030538*CDC.d.R100*HOXA.d.R100*IL8R.d.R100*MDK.d.R100+
[0408] 0.0232556*IGF.d.R100*HOXA.d.R100*IL8R.d.R100*MDK.d.R100+
[0409] -0.000867*CDC.d.R100*IGF.d.R100*HOXA.d.R100*IL8R.d.R100*MDK.d.R100

[0410] 结果

[0411] 尿样的qRT-PCR分析

[0412] 为获得IL8Rb对TCC检测的影响的综览,使用从患有TCC (n=56) 或非恶性疾病尿石症 (n=25), 尿路感染 (n=18) 或膀胱炎 (n=18) 的患者的尿液获得的qRT-PCR数据, 构建二维散点图。使用成对基因由四个基因标志物 (MDK, CDC2, HOXA13, IGFBP5) 构建散点图。然后IL8Rb代替每对中的一个基因, 并且重新绘制数据。这些图在图2a-f中示出。在MDK散点图 (图2a-c) 中用IL8Rb代替IGFBP5和HOXA13示出了TCC患者和那些非恶性疾病患者的样本之间改进的分离。在CDC2散点图中用IL8Rb代替IGFBP5和HOXA13 (图2d-f) 观察到相同的趋势。

[0413] 然后IL8Rb对显见血尿患者的TCC正确诊断的贡献通过ROC曲线分析进行定量。标记物 (MDK, CDC2, IGFBP5和HOXA13) 和IL8Rb中的每个基因的qRT-PCR数据被用于开发线性判别算法, 其将TCC患者和那些非TCC患者之间的判别最大化。使用整批442个样本: LD1, 其使用来自MDK, CDC2, HOXA13和IGFBP5的qRT-PCR数据和LD2, 其使用MDK, CDC2, HOXA13, IGFBP5和IL8Rb, 来开发两个线性判别算法。然后LD1和LD2被用来生成ROC曲线, 该曲线示出确诊患有TCC (n=56) 或非恶性疾病尿石症 (n=25), 尿路感染 (n=18) 或膀胱炎 (n=18) 的患者组

的TCC检测的灵敏度和特异性。图3a示出了LD1和LD2的ROC曲线。LD1的ROC曲线下的面积为78%，而LD2的ROC曲线下面积为84%。

[0414] 作为线性判别分析的替代，逻辑回归被用作独立的方法来开发用于判别TCC患者和那些非恶性疾病患者的算法。至于线性判别分析，逻辑回归算法采用整批442个样本开发。使用逻辑回归获得ROC曲线，上述56个TCC样本和61个非恶性样本示于图3b。LR1的ROC曲线下面积(使用来自MDK, CDC2, HOXA13和IGFBP5的qRT-PCR数据获得的)为80%，而LR2的ROC曲线下面积(使用来自MDK, CDC2, HOXA13, IGFBP5和IL8Rb的qRT-PCR数据获得的)为86%。该数据清楚地例示了将IL8Rb加入到利用尿样检测TCC的方法中可导致TCC患者和非恶性疾病，如膀胱炎，尿路感染和尿石症患者之间改善的判别。

[0415] 为确认由IL8Rb提供的、用于TCC患者和尿石症，尿路感染或膀胱炎患者之间的判别的改善的精度，这些患者被保持在一批未经选择的患者中，其包括更多的和更多样的非恶性患者，用表1中的整批442个样本重复ROC曲线分析。在此分析中，LD1和LD2的曲线下面积分别为86%和89% (图4a)。同样，LR1的曲线下面积为87%，LR2的曲线下面积为91% (图4b)。这个结果证实，IL8Rb导致使用尿样检测TCC中改善的精度。

[0416] 由于包含IL8Rb导致癌症检测中的此改善进一步通过在442个患者组中应用LD1/LD2和LR1/LR2进行例示，然后单独确定TCC的Ta阶段的检测的灵敏度。Ta阶段的肿瘤是更小、更分化的肿瘤，其通常比更高阶段的肿瘤更难检测。在85%特异性上，LD1检测到18/31 (58%) 的Ta肿瘤，而LD2检测到19/31 (61%) 的Ta肿瘤。LR1检测到21/31 (68%)，而LR2检测到24/31 (77%) (85%特异性)。该数据表明，包含IL8Rb到LD和LR算法中提高了Ta阶段肿瘤的检测灵敏度高达9%。相比于这些RNA测试，这项研究中的其他三个膀胱癌试验示出了，检测Ta肿瘤的显著降低的精度：尿细胞学 (39%灵敏度, 94%特异性)，NMP22 ELISA (35%灵敏度, 88%特异性) 和NMP22 (BladderChek® “美国，马萨诸塞州，MatriTech, Inc. 的注册商标”) (39%灵敏度, 96%特异性)。

[0417] IL8Rb作为尿道炎症诊断的辅助

[0418] 为了确定IL8Rb用于诊断由于如膀胱炎或尿路感染等原因患有尿道炎症的患者的能力，诊断为良性前列腺增生，非特异性前列腺病，血管性前列腺，继发于华法林使用的血尿，和膀胱炎/尿路感染的血尿患者的IL8Rb mRNA尿液水平通过qRT-PCR确定。这些疾病中的每一个的平均IL8Rb Δ Ct水平分别是-3.12, -3.10, -2.84, -1.98和-5.27。使用Wilcoxon秩和检验，确定患者膀胱炎/尿路感染和其它非恶性状态的组合病症的患者的IL8Rb水平的平均值之间的差异为显著 ($p=0.001$)。描绘该数据的箱型图示于图5。该数据表明相比所检查的其他非恶性疾病的患者，IL8Rb水平在大多数诊断患有膀胱炎或尿路感染的患者中有升高。图表之间的重叠可能由三个因素的组合来解释：(i) 标准临床实践无法正确地诊断每个疾病，(ii) 合并症 (例如，感染和良性前列腺增生)，和 (iii) 良性前列腺增生，非特异性前列腺疾病，血管前列腺或继发于华法林使用的血尿的患者子集中高尿嗜中性粒细胞计数的正常关联。无论如何，鉴于炎症和嗜中性粒细胞数之间的严格关联，尿液中IL8Rb的定量提供了检测尿路炎症的精确方法，无论是作为感染的后果或与其它非恶性疾病的关联。

[0419] 示例2

[0420] 方法

[0421] 研究群体

[0422] 在2008年4月28日至2009年8月11日之间在新西兰的9个泌尿外科诊所和澳大利亚的2个泌尿外科诊所,前瞻性地招募了一系列连续无TCC病史的患者。患者组包括示例1中所用的患者,但是还包括额外的46例患者,其数据没有用于第一个分析。进一步的研究还包括所获得的结果的进一步分析。样本的收集和RNA的收集和测试如示例1所述。

[0423] RNA测试开发

[0424] **uRNA[®]**由四个mRNA标志物CDC2,HOXA13,MDK和IGFBP5组成。这些标志物是基于它们在血液和炎性细胞中的低表达以及在TCC中的过度表达被选择。在这个队列研究中,我们前瞻性指定线性判别算法(uRNA-D),将其四个标志物组合为单一的分數。uRNA-D是独立的,是在先前的数据集上开发出来的。然而,它不是使用代表测试针对的目标人口的严格特征患者组获得的。作为结果,该研究协议还定义了新的算法(分类器-D)的开发,用于使用从招募到当前队列研究的患者处获得的数据来使用五个标志物CDC2,HOXA13,MDK,IGFBP5和IL8Rb。

[0425] 除了分类器-D,第二种算法(分类器-S)是使用队列研究数据获得的,以使用能识别进展期($\geq$ 阶段1)或高度(WHO/ISUP 1998年分类)的肿瘤。算法-S包括所有五个标志物,包括CDC2和HOXA13,其之前已经示出为在Ta阶段肿瘤和那些 $\geq$ 阶段1肿瘤之间被差异表达。

[0426] 分类器开发

[0427] 根据在本说明书中概述的方法,为使用五个标志物CDC2,HOXA13,MDK,IGFBP5和IL8RB开发两个分类器(分类器-D和分类器-S),是基于在本研究中获得的数据。简单地说,逻辑回归模型是使用统计编程环境R建立的(R Development Core Team(2011).R:A language and environment for statistical computing.R Foundation for Statistical Computing,Vienna,Austria.ISBN 3-900051-07-0,URL <http://www.R-project.org/>)。使用五个标志物的每一个的 Δ Cp值和它们的双向相互作用(例如,MDK x CDC2,MDK x IGFBP5等)建立的模型被评估其分类的能力;用留一法交叉验证程序估计具有最低AIC值的那些模型在特异度设为85%时的灵敏度。对分类器-D和分类器-S的每一个,几个模型展示了与选择最少数量的参数的模型相媲美的性能。

[0428] 统计方法

[0429] 诊断测试是在协议中规定的,为灵敏度和特异性计算出比例和95%置信区间。使用Stata roctab和roccomp命令(Statacorp和DeLong)画出并比较接受者操作特性(ROC)曲线。对于分类器-D,置信区间不合适,但Fisher精确检验或卡方检验(当样本大小允许时)被用来测试TCC或患者的特性及真阳性或假阳性结果可能性之间的关联。逻辑回归模型用于探索与假阳性和假阴性结果相关的因素。所有分析均在Stata 11.2版中实行。

[0430] 结果

[0431] 本研究最初共招募517例患者;4%的患者被排除,因为发现他们是不合格的($n=10$),没有经过膀胱镜检查的($n=9$),TCC状态不合规定($n=2$),或者他们没有提供可接受的尿样($n=2$)(图8)。另外10例患者从分析中排除,因为他们没有一个或多个尿测试的结果。剩余的485例患者的基线人口统计学和临床特征示于图9。

[0432] 队列中TCC的患病率为13.6%。两例错过复查阶段(两例都是当地复查为Ta),两例都没有给定复查级别(一例为当地病理学家诊断1级,另一例为低级)。66例肿瘤中,55例是浅表的(Ta,T1或Tis阶段),11例为肌层浸润性(T2)。没有患者具有可检出的区域淋巴结转

移或加入。使用1973分级系统,24例被分类为3级,38例为2级,3例为1级和1例未知。使用WHO98系统,29例被分类为高级,4例混合,32例低级和1例未知。除了TCC,两例患者被诊断为乳头状瘤,七例诊断为其他肿瘤(其中五例是泌尿外科的)。

[0433] uRNA-D测试的临界值是在特异性设定在85%的研究队列上确定的。根据此临界值,相比于NMP22™ ELISA (50%), **Bladderchek**® (38%) 和细胞学 (56%), uRNA-D检测到66个TCC病例中的41个 (62%灵敏度)。在队列数据分类器-D上开发的RNA测试在85%的特异性检测到TCC病例中的54个 (82%), 在90%的特异性检测到48个 (73%)。uRNA-D和NMP22™ ELISA值可以直接比较,因为这两个测试在本研究之前得到充分说明。图21示出ROC曲线;曲线下面积 (AUC) 分别为0.81和0.73 ($p=0.03$)。分类器-D的ROC曲线为0.87 (图21), 并相对于uRNA-D的性能改善似乎主要是在临床上相关的特异性范围内 (80%以上)。

[0434] 总体上,分类器-D检测出的高级/3级肿瘤的97%, 相比于uRNA-D (83%), 细胞学 (83%), NMP22 ELISA (69%) 和 **Bladderchek**® (38%)。相比于其他测试的范围为28-41% (图12), 分类器-D也对检测低级肿瘤 (69%) 更灵敏。分类器-D对所有阶段 ≥ 1 加两个Tis的TCC病例为阳性,但是对Ta阶段灵敏度为68% ($p=0.016$, 图12)。这仍然大大高于其他测试, uRNA-D是第二高的,为41%。相比那些没有肉眼可见血尿或显微可见血尿 ($p<0.0005$) 的TCC患者,尿样中明显具有肉眼可见血尿或显微可见血尿的TCC患者更可能通过加入IL8Rb检测到其TCC,尽管这可能至少部分是肉眼可见血尿或显微可见血尿患者中高阶段和高级TCC的比例较高的结果。数目不足以在回归分析中做进一步探讨。

[0435] 分类器-D错过的12个病例中,所有均为Ta阶段的,并且除了1个外,其余均是低级的 (WHO ISUP 1998)。12个中只有2个 (均为低级, Ta阶段的TCC) 是被另一个测试检测出的 (1个由NMP22™ ELISA和 **BladderChek**®, 一个由uRNA-D)。分类器-D错过的12个病例中,所有均为Ta阶段的,并且除了1个外,其余均是低级的 (WHO ISUP 1998)。12个中只有2个 (均为低级, Ta阶段的TCC) 是被另一个测试检测出的 (1个由NMP22™ ELISA和 **BladderChek**®, 一个由uRNA-D)。细胞学没有检测到分类器-D错过的任何TCC。

[0436] 患者A:高级肾盂肿瘤T2,没有并发Tis,没有给定大小。

[0437] 患者B:高级膀胱肿瘤T3a,没有并发Tis, 2x3cm

[0438] 患者C:高级肿瘤,测量4.8x5.6cm,具有大面积的基质和肌层浸润,延伸至膀胱周的脂肪,无转移的证据。

[0439] 那些采用替代诊断的患者中,根据尿样特性的尿液测试的特异性示于图13。肉眼可见血尿或显微可见血尿的控制组患者比没有肉眼可见血尿或显微可见血尿 ($p=0.002$) 的患者更可能有假阳性测试,并且有一个趋势,即患者可能还有结石,虽然通过诊断特异性的差异整体上统计上不显著 ($p=0.12$)。有5例具有其他泌尿癌症的患者;其中只有一个给出了阳性的分类器-D测试结果。拟合逻辑回归模型的结果是相似的。在诊断和肉眼可见血尿或显微可见血尿的逻辑回归模型中,与肉眼可见血尿或显微可见血尿状况的相关性仍然显著 ($p=0.006$), 并且当直接相比于无诊断时,结石患者具有2.7倍增加几率的假阳性测试 (95%CI (1.1至6.4), $p=0.03$)。年龄不影响测试的特异性。

[0440] 尿样中检测出的肉眼可见血尿或显微可见血尿是与测试灵敏度明显相关的唯一因素。在这个队列中那些肉眼可见血尿或显微可见血尿患者的阳性测试的预测值是63%,

那些没有肉眼可见血尿或显微可见血尿患者的阳性测试的预测值是24%，这主要反映了肉眼可见血尿或显微可见血尿患者中更大的TCC患病率(39%对6%)。

[0441] 有54例TCC患者的分类器-D测试为阳性。这些患者采用分类器-S分类为严重TCC和不严重TCC。严重TCC定义为阶段 ≥ 1 或任何阶段的3级。在90%的特异性，分类器-S正确分类32/35(91%)的严重TCC病例。

[0442] 示例3:血尿I期患者的基因型和表型组合分析

[0443] 本研究着重于呈现确诊无症状镜下血尿(AH)的患者,其正在经历对可能的尿路上皮癌(UC)的全面临床病情检查。约500例患者被招收到参与了这项研究。

[0444] 如本说明书中所用,术语在额外的示例中定义如下。

[0445] 目标

[0446] 目标用于确定:(1)在计划进行全面泌尿外科临床病情检查的呈现显微可见血尿的患者中基因型和表型算法的疗效,(2)用于在呈现确诊镜下血尿患者中检测初级UC的基因型和表型算法G+P指数的性能特点(灵敏度,特异性,ROC曲线下面积,阳性和阴性预测值);(3)由基因型和表型工具G+P指数正确诊断为UC阴性的患者的数量,他们因此不需要调查性的膀胱镜检查。

[0447] 研究群体

[0448] 研究群体包括呈现确诊显微可见血尿的患者,其满足研究的要求。患者是从由全科诊所转介到泌尿外科诊所的患者中招募的。

[0449] 知情同意

[0450] 计划进行调查性的膀胱镜检查的患者被接触,以讨论可能的参与。患者被告知研究的性质并且得到同意。患者提供人口统计,职业和吸烟史的信息,并在提供尿液样本之前,确保他们充分了解患者信息和知情同意书。研究协调员完成详细介绍基因型和表型指数的相关输入信息的CRF页并传输数据进行分析。

[0451] 纳入标准

[0452] 随后正在进行尿路上皮癌膀胱镜调查的患者被确诊在2或3个妥善收集的尿液样品上临床发现镜下血尿(每高倍视野(HPF)至少3个RBC)。

[0453] 病人愿意遵守研究要求。

[0454] 患者已年满18岁。

[0455] 排除标准

[0456] 尿路上皮癌(UC)病史。

[0457] 当前呈现肉眼可见血尿。

[0458] 确诊(恶性或其他)的肉眼可见血尿发作期的病史(过去12个月)。

[0459] G+P指数

[0460] 我们开发了新的指数,“G+P指数”,其包括基因型和表型数据两者的组合。基因型(“G”)组成部分利用RNA生物标志物表达信息,与收集自相同时间窗口中的患者的五个临床因素(表型数据“P”)一起使用,以确定AH患者的UC风险。

[0461] 所有患者接受标准的临床病情检查以确定真实的临床结果,并且该研究的输出是以所收集的临床数据和收集自患者尿样的基因型数据为基础的模拟输出。这样,患者护理不会因研究输出的结果而改变。患者提供尿样,其被送去用于遗传分析。

[0462] 患者归类

[0463] 参与本研究的患者的总体护理标准没有变化。根据当前护理标准,所有计划进行全面泌尿科病情检查的患者接受适当的调查。

[0464] 研究数据

[0465] 人口统计学和风险因素信息是基因型和表型指数是输入数据。最终的疾病状态(由灵活的膀胱镜检查 and 后续检查确定)用(Φ)结果和人口统计信息进行校对,并进行统计分析。

[0466] G+P指数的确定

[0467] 使用从收集自大量新西兰和澳大利亚站点收集的约500个患者的样本获得的数据集,预测“TCC=yes”可能性的训练模型被开发。

[0468] 收集用于训练和验证群体的变量的数据

[0469] 表型变量

[0470] 临床研究结果是:性别,年龄,吸烟史,和HFREQNEW(≤ 1 表示低; >1 表示高)。

[0471] 基因型变量

[0472] 基因型变量包括RNA标志物:M1(=MDK+CDC+IGBP5-HOXA13)和IL8R的表达。下面的表7示出验证G+P指数的因素中的每一个的系数的估算值:

[0473] 表7

[0474]

最大似然估计的分析							
参数		DF	估计值	标准误差	Wald 卡方	Pr> 卡方	Exp(Est)
截距		1	-4.8445	0.6532	55.0092	<.0001	0.008
性别	2	1	-1.8544	0.6750	7.5484	0.0060	0.157
吸烟	2	1	-0.9049	0.4537	3.9775	0.0461	0.405
吸烟	3	1	-1.3407	0.4936	7.3756	0.0066	0.262
HFRE-QNEW	低	1	-0.7015	0.3478	4.0681	0.0437	0.496
M1		1	1.2221	0.1553	61.9328	<.0001	3.394
IL8R		1	-0.2408	0.1789	1.8112	0.1784	0.786

[0475] 图18描绘了G+P指数的ROC曲线。图18是根据本发明的实施例的结果的灵敏度(纵轴)对1-特异性(横轴)的图。为了比较,对角线描绘了模型。单独基于表型信息的结果被示为点划线,单独基于基因型信息的结果被示为虚线,基于G+P指数的结果被示为实线。这些结果表明,基因型和表型信息的组合提供了对预测结果的出人意料的、显著的改善。

[0476] 探索模型考虑7个表型变量,但AgeGT50表现影响不大,而RBC数据不足,所以这两个变量从最终模型中去掉。在剩余的五个表型变量和两个RNA标志物的显著性水平基础上,构建一个指数。使用训练数据集中M1,IL8R和TCC(=yes)之间的相互关系,阈值4.5和2.5分别用于M1和IL8R。得分5,4,3,2和1被分配到M1,吸烟者,男,IL8R和HFREQ—这导致指数得分

范围从0到15。基于系数的合成算法如下,所组合的G+P指数为:

[0477] $G+P \text{ 指数} = (1 * HFREQ + 3 * Gender + 4 * SMK) + (5 * M1 + 2 * IL - 8)$

[0478] 被保留在最终模型中的不同临床因素的比值比示于图19。比值比可以依赖于它偏离1(即,没有影响)多远被解释为具有对受试者有害或保护的作用。比值比,其置信界限不包括1,是统计显著的。通常,相比于具有小比值比的因素(例如HFREQ),具有较高比数比的因素(例如SMK,Gender)被分配较大的权重。

[0479] 完整模型的分类表示于下面表8中。

[0480] 表8

[0481]

分类表									
概率水平	正确		不正确		百分比				
	事件	非事件	事件	非事件	正确	灵敏度	特异性	假阳性	假阴性
0.000	66	0	415	0	13.7	100.0	0.0	86.3	0.0
0.050	61	245	170	5	63.6	92.4	39.0	73.6	2.0
0.100	54	319	96	12	77.5	81.8	76.9	64.0	3.6
0.150	47	358	57	19	84.2	71.2	86.3	54.8	5.0
0.200	45	376	39	21	87.5	68.2	90.6	46.4	5.3
0.250	40	381	34	26	87.5	60.6	91.8	45.9	6.4
0.300	37	391	24	29	89.0	56.1	94.2	39.3	6.9
0.350	36	398	17	30	90.2	54.5	95.9	32.1	7.0
0.400	35	402	13	31	90.9	53.0	96.9	27.1	7.2
0.450	35	403	12	31	91.1	53.0	97.1	25.5	7.1
0.500	31	407	8	35	91.1	47.0	98.1	20.5	7.9
0.550	29	408	7	37	90.9	43.9	98.3	19.4	8.3
0.600	29	408	7	37	90.9	43.9	98.3	19.4	8.3
0.650	28	409	6	38	90.9	42.4	98.6	17.6	8.5
0.700	24	409	6	42	90.0	36.4	98.6	20.0	9.3
0.750	20	409	6	46	89.2	30.3	98.6	23.1	10.1
0.800	12	410	5	54	87.7	18.2	98.8	29.4	11.6
0.850	11	411	4	55	87.7	16.7	99.0	26.7	11.8
0.900	6	412	3	60	86.9	9.1	99.3	33.3	12.7
0.950	1	413	2	65	86.1	1.5	99.5	66.7	13.6
1.000	0	415	0	66	86.3	0.0	100.0	0.0	13.7

[0482] G+P指数的初步验证研究

[0483] 为了进一步测试G+P指数的使用,我们进行了另一项研究。基于各种临床和RNA标志物的统计显著性,构建一个指数。共有98名受试者,其TCC状态(yes或no)以及G+P指数的变量是可用的。

[0484] 得分5,4,3,2和1被分配到M1(基因测试),吸烟者,男,IL8R和HFREQ—这导致指数得分范围从0到15。真阳性和真阴性的数量分别为6和84。同样,假阳性和假阴性的数量分别为5和3。因此,所提出的指数的总体精度为0.92。

[0485] 基于G+P指数的含义及跟进

[0486] 如果G+P指数结果表明UC“高风险”定义为得分11-15或以上,患者具有接受临床指示的灵活的膀胱镜检查 and 腹部超声的优先权。

[0487] 如果G+P指数结果表明UC“中风险”定义为得分6-10,患者要接受复查和临床实践的随访。可以考虑使用细胞学,输尿管镜和/或CT扫描。

[0488] 如果G+P指数结果表明UC“低风险定义为得分0-5的”,患者将接受正常标准护理,并被放置在适当的等候名单上。

[0489] 参考文献

[0490] 下列参考文献涉及到上面的公开内容。

[0491] Agresti,A.and Coull,B.A. (1998) Approximate is better than “ex-act” for interval estimation of binomial proportions. The American Statistician, 52, 119-126.

[0492] Altekruse SF, Kosary CL, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds). SEER Cancer Statistics Review, 1975-2007, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2007/, based on November 2009 SEER data submission, posted to the SEER web site, 2010.

[0493] Brown, L.D. Cai, T.T. and DasGupta, A. (2001) Interval estimation for a binomial proportion. Statistical Science, 16, 101-133.

[0494] Buckhaults等人, (Cancer Research 61:6996-7001 (2002) describes certain secreted and cell surface genes expressed in colorectal tumours.

[0495] Byrd, R.H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. SIAM J. Scientific Computing, 16, 1190-1208.

[0496] Dalgaard (2008). “Introductory Statistics with R, 2nd edition”; Peter Dalgaard, Chapter 13 (2008), Springer, ISBN 978-0-387-79053-4.

[0497] DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44:837-845.

[0498] Gottschalk, P.G. and Dunn, J.R. (2005) The five-parameter logistic: A characterization and comparison with the four-parameter logistic. Analytic Biochemistry, 343, 54-65. Doi:10.1016/j.ab.2005.04.035

- [0499] Grossfield G, Wolf J, Litwan M, Hricak H, Shuler C, Agerter D 等人. Asymptomatic microscopic hematuria in adults: summary of AUA best practice policy recommendations. *AFP* 2001;63:1145-54.
- [0500] Hall et al (*Laryngoscope* 113(1):77-81 (2003) (PMID:12679418) (Abstract) described predictive value of serum thyroglobulin in thyroid cancer.
- [0501] Hoerl, A.E. (1962) Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58, 54-59.
- [0502] Holyoake A, O'Sullivan P, Pollock R, Best T, Watanabe J, Kajita Y 等人. Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder. *Clin Cancer Res.* 2008 Feb 1;14(3):742-9.
- [0503] Hotte 等人, (*AJ. American Cancer Society* 95(3):507-512 (2002) describes plasma osteopontin as a protein detectable in human body fluids and is associated with certain malignancies.
- [0504] Kim 等人, (*JAMA* 287(13):1671-1679 (2002) describes osteopontin as a potential diagnostic biomarker for ovarian cancer.
- [0505] Koopman 等人, (*Cancer Epidemiol. Biomarkers Prev* 13(3):487-491 (2004) (Abstract) describes osteopontin as a biomarker for pancreatic adenocarcinoma.
- [0506] Kuo 等人 (*Clin. Chim. Acta.* 294(1-2):157-168 (2000) (Abstract) describes serum matrix metalloproteinase-2 and -9 in HCV- and HBV-infected patients.
- [0507] Leman 等人, (*Urology*, 69(4) 714-20 (2007) (Abstract) describes EPCA-2 as a serum marker for prostate cancer.
- [0508] Loo R, Lieberman S, Slezak J, Landa H, Mariani A, Nicolaisen G, Aspera A and Jacobsen S: Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria. *Mayo Clin Proc.* 2013, 88(2):129-138.
- [0509] Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine* 28:3049-3082.
- [0510] Marchi 等人, (*Cancer* 112, 1313-1324 (2008) (Abstract) describes ProApolipoprotein A1 as a serum marker of brain metastases in lung cancer patients.
- [0511] Martin 等人, (*Prostate Cancer Prostatic Dis.* March 9, 2004 (PMID:15007379) (Abstract) described use of human kallikrein 2, prostate-specific antigen (PSA) and free PSA as markers for detection of prostate cancer.
- [0512] Mazzaferri 等人, (*J. Clin. Endocrinol. Metab.* 88(4):1433-1441 (2003) (Abstract) describes thyroglobulin as a potential monitoring method for patients with thyroid carcinoma.
- [0513] McDonald M, Swagerty D, Wetzel L: Assessment of Microscopic hematuria in adults. *AFP* 2006 73:10.

- [0514] Melle等人, (Clin.Chem.53(4) ,629-635 (2007) (Abstract)describes HSP27 as a serum marker for pancreatic adenocarcinoma.
- [0515] More,J.J.(1978) The Levenberg-Marquardt algorithm:implementation and theory,in Lecture Notes in Mathematics 630:Numerical Analysis,G.A.Watson (Ed.),Springer-Verlag:Berlin,pp.105-116.
- [0516] Nelder,J.A.and Mead,R.(1965) A simplex algorithm for function minimization.Computer Journal 7,308-31.
- [0517] O'Sullivan P,Sharples K,Dalphin M等人:A Multigene Urine Test for the Detection and Stratification of Bladder Cancer in Patients Presenting with Hematuria.J Urol 2012,Vol 188 No 3;746
- [0518] Pellegrini等人, (Cancer Immunol.Immunother.49 (7) :388-394 (2000) (Abstract)describes measurement of soluble carcinoembryonic antigen and TIMP 1 as markers for pre-invasive colorectal cancer.
- [0519] (R Development Core Team(2009) .R:A language and environment for statistical computing.R Foundation for Statistical Computing,Vienna, Austria.ISBN 3-900051-07-0,URL <http://www.R-project.org/>)
- [0520] Richards,F.J.(1959) A flexible growth function for empirical use.Journal of Experimental Botany.10,290-300.
- [0521] Rudland等人, (Cancer Rese arch 62:3417-3427 (2002)describes osteopontin as metastasis associated protein in human breast cancer.
- [0522] Sing等人(2009) .(Tobias Sing,Oliver Sander,Niko Beerenwinkel and Thomas Lengauer(2009) .ROCR:Visualizing the performance of scoring classifiers..R package version 1.0-4.<http://CRAN.R-project.org/package=ROCR>) .
- [0523] Speiss,A.-N.,Feig,C.and Ritz,C.(2008) Highly accurate sigmoidal fitting of real time RCR data by introducing a parameter for asymmetry.BMC Bioinformatics,9,221.Doi:10.1186/1471-2105-9-211.
- [0524] Sprenger H,Lloyd AR,Lautens LL,Bonner TI,Kelvin DJ.Structure,genomic organization,and expression of the human interleukin-8 receptor B gene.J Biol Chem.1994 Apr 15;269(15):11065-72.
- [0525] StataCorp.2009.Stata Statistical Software:Release 11.College Station, TX:StataCorp LP.
- [0526] Schwartz G:Proper evaluation of asymptomatic microscopic hematuria in the era of evidence-based medicine-progress is being made.Mayo Clin Proc.2013,88(2) ;123-125.
- [0527] Trivedi C and Messing EM:Commentary:the role of cytologic analysis of voided urine in the work-up of asymptomatic microhematuria.BMC Urology 2009, 9:13
- [0528] Tsigkou等人, (I Clin Endocrinol Metab,92(7) 2526-31 (2007) (Abstract)

describes total inhibin as a potential serum marker for ovarian cancer.

[0529] Venables, W.N. & Ripley, B.D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0).

[0530] Whitley 等人, (Dim Lab. Med. 24 (1): 29-47 (2004) (Abstract) describes thyroglobulin as a serum marker for thyroid carcinoma.

[0531] Yoshikawa 等人, (Cancer Letters, 151: 81-86 (2000) describes tissue inhibitor of 30matrix metalloproteinase-1 in plasma of patients with gastric cancer.

[0532] 示例4: 使用G+P指数II归类呈现血尿的患者

[0533] 当G+P指数值在11至15的范围时, G+P指数指示阳性。

[0534] 定义

[0535] 如在本说明书中使用的, 下列定义用在此处和下面的示例中。

[0536] “AMH”是指无症状显微镜可见血尿;

[0537] “AUA”是指美国泌尿学协会;

[0538] “AUC”是指曲线下面积;

[0539] “CI”是指置信区间;

[0540] “CT”是指计算机断层扫描;

[0541] “ELISA”是指酶联免疫吸附测定;

[0542] “FISH”是指荧光原位杂交;

[0543] “HPF”是指高倍视野;

[0544] “logOR”是指对数比值比;

[0545] “Hfreq”是指最近血尿发作时血尿的日均频率;

[0546] “ISUP”是指泌尿病理学国际协会;

[0547] “MRI”指核磁共振成像;

[0548] “NPV”指阴性预测值;

[0549] “OR”表示比值比;

[0550] “QC”是指质量控制;

[0551] “QoL”是指生活质量;

[0552] “表型 (Phenotypic)”是用来定义临床预测的特征并将它们与已被广泛地定义为“基因型”变量的基于基因表达的生物标志物区分开。

[0553] “RBC”指红细胞;

[0554] “ROC”是指接受者操作曲线;

[0555] “RT-qPCR”是指定量逆转录聚合酶链反应;

[0556] “STARD”是指报告诊断精度的标准;

[0557] “UC”是指尿路上皮癌;

[0558] “WHO”是指世界卫生组织。

[0559] 简介

[0560] 血尿, 其与例如良性前列腺肿大, 感染或尿路结石这些原因最经常相关联, 但也是尿路上皮癌 (UC) 的症状, 估计一般人群中在1%和22%之间的患者会发生血尿 [1, 2]。肉眼

可见血尿的特征在于患者尿中可见的颜色变化,而显微可见血尿更精确地定义为在3个同时收集的尿样中每高倍视野中存在 ≥ 3 个红细胞(RBCs/HPF) [2]。显微可见血尿患者中UC的总体发病率已经被报道为约4%,而一些研究一致表明,UC的发病率在肉眼可见血尿的患者中要高得多,范围约为12%-23% [2-6],在泌尿评估中肉眼可见血尿患者的UC发病率达到显微可见血尿患者UC发病率的近四倍 [7]。值得注意的是,鉴于美国泌尿学协会(AUA)准则的最近变化 [2],已经看到无症状血尿(AMH)的阈值降低至单个样本中 ≥ 3 RBCs/HPF,并且甚至更低的阈值(≥ 1 RBC/HPF)已经被提出 [8],因此将接受泌尿病情检查以调查潜在UC的患者的数量相应增加,并且可预期这些患者对医疗保健系统的整体临床和经济负担相应增加。

[0561] 这种血尿相关的转诊给泌尿科带来了显著的临床负担,因为所有患者必须进行全面的病情检查以提供通常无确定结果的诊断。此外,现有的诊断测试—其中许多是侵入性或具有能够对患者生活质量(QoL)产生不利影响的高辐射负载,特别是如果依照当前指南授权的,患者接受重复的膀胱镜检查 [2]。据报道,对于没有预防性抗生素实施膀胱镜检查,22%的患者患有无症状菌,1.9%的患者在30天内发展成发热性尿路感染(UTI) [9]。其他研究也报道了膀胱镜检查之后,肉眼可见血尿,排尿疼痛和男性的短暂勃起功能障碍的患病率较高 [10,11]。

[0562] 作为血尿患者接受全面的泌尿病情检查的结果,医疗保健系统也遭受显著的经济负担 [12,13],而且已得出结论,尿细胞学检查增加了成本,但没有提供任何显著的诊断益处 [14-16]。因此,融合精确的、非侵入性的测试到呈现血尿患者的初级临床病情检查允许医生有效地归类血尿患者,从而降低为检测UC而接受全面泌尿病情检查和调查性膀胱镜检查的患者数目,并提供给患者和医疗保健系统显著的好处 [15-19]。

[0563] 一些临床预测的特征,包括年龄,性别,吸烟史及血尿程度,是公认的血尿患者的UC危险因素 [3,20-22]。最近,一些团体试图基于临床预测特征来开发预测血尿患者UC风险的模型 [20-22],但关键是,这些模型提供了有限的精度且大都集中在检测UC患者而不是排除没有疾病的患者。因此,即使与尿细胞学组合使用,这些关注检测的模型已经不足以在初期评估期间可靠地确定患病的患者 [20-22]。

[0564] 尽管呈现肉眼可见血尿患者中UC发病率较高,但是多项研究显示与那些呈现肉眼可见血尿的患者相比,呈现显微可见血尿的患者在UC的分级和分期的分布上无显著差异 [5,23-25]。因此,AUA建议所有肉眼可见血尿患者或AMH患者均转诊到泌尿科接受全面的泌尿病情检查,因为血尿的严重程度不足以预测UC的存在 [2]。然而,因为在初期评估中,血尿患者可能仅经历有限的尿液分析,包括细胞学和一些情况下成像研究,例如超声波,所以全面的泌尿病情检查通常是必要的,以决定性地检测到或排除UC。虽然尿细胞学检查被目前的指南指定并且常规地用于疑似UC患者,但是细胞学检查结果往往是非典型或可疑的尚无定论的结果,并且饱受UC相关血尿患者假阳性结果的相对较高风险带来的低诊断率的苦果 [2,26,27]。因此,可能很难在初期评估期间排除血尿的良性原因,无论是肉眼可见血尿或AMH,尤其是如果UC相关血尿是间歇性的并在针对良性原因的治疗后似乎痊愈的 [12]。

[0565] 许多基于基因的研究已经开始描述UC患者的尿液生物标志物,并且这些生物标志物可以凭自身的能力对检测疾病是有用的 [28,29]。也存在机会基于患者的临床特征和基因表达谱对患者归类。与单独的临床特征相比,将NMP22酶联免疫吸附测定(ELISA)测试或

一组基因标志物与临床特征组合已经显示出可以提高诊断的准确性,但是这些组合模型尚未在整体诊断准确性上带来显著的进步,特别是当试图识别低风险患者时[30,31]。尽管如此,将临床因素和特异性基因表达结合为组合的算法被认为可能为诊断和管理血尿或UC患者提供最好的指导[32]。

[0566] Cxbladder™ Detect (新西兰,达尼丁,Pacific Edge Ltd.),在未分馏的尿液上进行多基因测试之前已经显示出比尿细胞学和NMP22在肉眼可见血尿患者中检测UC更灵敏[33]并且在比较分析中,比尿细胞学,NMP22和荧光原位杂交(FISH)更准确(Kasabov, Darling, Breen等人,未发表的言论)。Cxbladder Detect使用定量逆转录聚合酶链反应(RT-qPCR)技术来定量5个mRNA标志物,四个标志物在UC中过度表达,同时第五个标志物在非恶性炎症病症中升高,并且当用于检测呈现血尿的患者中的UC时,Cxbladder Detect提供高水平的特异性和灵敏度[33]。据推测,结合收集自同一患者的高性能的遗传标志物和表型变量的综合模型将使用高灵敏度(即UC患者接收假阴性结果的低可能性),高阴性预测值(即所有阴性结果是真实的比例高)和高测试阴性率提供卓越的临床解决方案,以使能精确归类具有UC低概率的患者。当将这些基因型和表型变量组合成新的偏析模型时,能够使具有UC低概率的血尿患者被识别和归类,而不是经历全面的泌尿病情检查。

[0567] 方法

[0568] 患者的选择

[0569] 695例患者的前瞻性样本用于分析,其中采用传统的临床评估确定真实的临床结果。研究样本包括血尿患者的初始队列如前所述得到同意并取样[33],其中一系列连续的517例近期肉眼可见血尿病史,年龄 ≥ 45 岁,没有UC病史的患者从澳大利亚和新西兰的9个泌尿诊所前瞻性招募。这些患者被随访三个月用于确定UC状态或在尿液样本多基因分析后给出替代诊断,阳性的UC诊断是基于膀胱镜外观和组织病理学检查。采用1998年世界卫生组织(WHO)/泌尿病理学国际协会(ISUP)的共识分类,疾病的阶段根据病理和影像诊断调查确定的TNM分期标准分类,肿瘤分级是按照当地的病理学实践进行划分的[34]。

[0570] 肉眼可见血尿事件后,正在接受泌尿外科调查的另外的94例患者和84例患者的队列随后在2012年3月和2013年4月之间从新西兰的两个中心被招募并包括到模型开发中。在他们之前参加最初的研究的经验及在个人临床环境中评估Cxbladder Detect产品的意愿基础上,所述中心被选择参与其中。

[0571] 另外的代表性测试集的45例呈现显微可见血尿的患者被前瞻性地集合并用于对G+P指数的进一步验证,具体如下。

[0572] 合格标准是与[33]相似的,除了患者年龄 ≥ 18 岁外,那些曾经历了膀胱镜检查以调查UC,并被证明是阴性的,也有资格入选。此外,如在[33]中,显示出指示为UTI或膀胱或肾结石症状的患者,被排除在外。

[0573] 这项研究的伦理批准是由所有参与的中心通过的,并且提供样本的所有患者签署了知情同意书。

[0574] 尿样收集和评估

[0575] 为了提供基因表达数据,使用来自Pacific Edge的尿液取样系统收集参与者的单一中段尿样。所有研究样本的多基因分析是根据标准操作程序进行的,如同用于商售的Cxbladder Detect多基因测试。来自最初的队列的所有尿样(4.5mL)是膀胱镜检查之前在

诊所被收集的,并且经由真空驱动抽吸转为稳定液体,然后在48小时内送到Pacific Edge。之后将样本储存于-80℃,直到需要进行批量分析。之后队列的样本以同样的方式进行收集,但是根据Pacific Edge诊断实验室实行的修订的质量控制(QC)限制和容限测试,样本在环境温度下运送到Pacific Edge并在收集的7天内进行处理。

[0576] 统计分析

[0577] 单变量逻辑回归用来估计与UC相关的4个二进制表型变量的未调整(原始)对数比值比(logOR)系数:性别,吸烟史和患者最近的血尿发作期间血尿的日均频率(Hfreq;参见表9)。

[0578] 表9:与UC相关的二进制表型变量的定义和其相应得分

表型变量	得分	
	0	1
Gender (性别)	女性	男性
Age (年龄)	<60 岁	≥60 岁
Smoking history (吸烟史)	从不吸烟	当前或过去吸烟
Hfreq (血尿频率)	≤1 次/天	>1 次/天

[0580] 对所有四个表型变量的多变量逻辑回归被用来生成表型模式(P指数)中的调整logOR系数。

[0581] 使用逻辑回归来开发G指数,以确定UC和尿样中五个Cxbladder® Detect基因(IGFBP5,H0XA13,MDK,CDK1和CXCR2)的mRNA浓度之间的相关性。使用来自G指数和P指数的所有九个变量的组合生成多变量基因型—表型模型(G+P指数)。这些线性模型确定logOR,患者患有UC的概率可以从logOR导出。

[0582] 每个模型的相对表现在绘出如每个模型确定的测试UC时的假阳性率对真阳性率的接受者操作曲线(ROCs)中说明。曲线下面积(AUC)用来将每个模型的相对效率与被认为接近1是最佳的AUC做对比。

[0583] 当对相同的数据集执行模型估计和预测时,为了减少潜在的偏差,使用自举重复采样[35]为三个逻辑回归模型中的每一个计算偏差修正的AUC。来自原始样本的名义AUC和来自自举样本的平均AUC之间的差值是样本偏差的估计值,并且名义AUC进行相应调整。也得到偏差修正置信区间(CIs)自举估计值[36]。

[0584] 此外,每个模型的性能特点必须超过0.97的NPV阈值,其为此临床测试的设计标准,采用具有高测试阴性率的进一步附加说明而具有尽可能高的灵敏度。当归类(triaging)出具有UC低概率的呈现血尿的患者时,选择测试阴性率以提供较高的临床解决方案。在G指数,P指数和G+P指数之间作比较,并且根据灵敏度和具有足够高的测试阴性率

的NPV确定每个模型的性能,以提供用于归类(triaging)出具有UC低概率的血尿患者的有效工具。

[0585] 结果

[0586] 样本人口

[0587] 跨三个队列登记的695例肉眼可见血尿例患者中,由于缺乏足够的数据或者样本而未能达到QC标准,23例被认为不合格,并且来自另外85例患者的样本在注册后被排除(参见图20A)。总计,来自587例患者的样本可用于为包括72例UC阳性和515例UC阴性的样本建模。

[0588] 在提供的来自显微可见血尿患者的45个样本中,40个适合于分析,5例患者被认为不合格,并从分析中排除(参见图20B)。所有45例患者接受了全面的泌尿评估而且临床事实被确认为UC阴性。来自两个样本群体的全部人口统计数据列于表10。

[0589] 表10:肉眼可见血尿和显微可见血尿患者的采用完整数据的样本群体人口统计特征

[0590]

参数		肉眼可见 血 尿 患 者 (N=587) , n (%)	显微可见 血 尿 患 者 (N=40) , n (%)
Age (年 龄) , 岁	0-49	65 (11.1)	21 (52.5)
	50-59	111 (18.9)	
	60-69	145 (24.7)	19 (47.5)
	70-79	175 (29.8)	
	80-100	91 (15.5)	
Gender (性别)	女	113 (19.3)	25 (62.5)
	男	474 (80.7)	15 (37.5)
Smoking history (吸烟 史)	从不吸烟	246 (41.9)	25 (62.5)
	当前或过去 吸烟	341 (58.1)	15 (37.5)
Hfeq (血 尿频率) (次/ 天)	≤ 1	332 (56.6)	40 (100)
	>1	255 (43.4)	—
肿瘤阶段	正常	515 (87.7)	40 (100)
	T1	16 (2.7)	—
	T2	11 (1.9)	—
	T3	2 (0.3)	—
	Ta	40 (6.8)	—
	Tis	3 (0.5)	—

- [0591] 肉眼可见血尿患者的表型变量和UC风险的关系
- [0592] 四个二进制表型变量的每一个的未调整的单变量逻辑回归分析表明,年龄≥60岁,男性,吸烟史和肉眼可见血尿的高频率均与UC的风险增加相关(表11)。
- [0593] 表11:血尿患者通过表型和基因型因素UC的未调整和调整的OR
- [0594]

	未调整 OR	调整的 P 变	调整的 G+P
--	--------	---------	---------

[0595]

				(95% CI)	量 OR (95% CI)	变量 OR (95% CI)
表型变量		对照组	UC			
Age (年龄), 岁	<60	151	11	2.30	2.24	1.89
	≥ 60	364	61	(1.22-4.73)	(1.18-4.65)	(0.85-4.64)
Gender (性别)	女	105	8	2.05	1.58	3.03
	男	410	64	(1.01-4.75)	(0.76-3.72)	(1.12-9.36)
Smoking history (吸烟史)	从不吸烟	227	19	2.20	2.19	2.67
	当前或过去吸烟	288	53	(1.29-3.91)	(1.27-3.92)	(1.34-5.64)
Hfreq(平均次数/天)	≤ 1	300	32	1.74	1.80	1.76
	>1	215	40	(1.06-2.88)	(1.08-3.00)	(0.93-3.35)
基因型变量				未调整 OR (95% CI)	调整的 G 变量 OR (95% CI)	调整的 G+P 变量 OR (95% CI)
<i>IGFBP5</i>				7.34 (4.59-12.33)	2.15 (1.03-4.58)	2.21 (1.03-4.83)
<i>HOXA13</i>				6.27 (3.92-10.34)	0.33 (0.13-0.83)	0.20 (0.07-0.56)
<i>MDK</i>				7.10 (4.73-11.10)	4.76 (1.74-13.62)	8.14 (2.64-26.60)
<i>CDKI</i>				7.80 (5.11-12.39)	3.47 (1.39-9.13)	2.59 (0.98-7.18)
<i>CXCR2</i>				1.69 (1.36-2.10)	0.65 (0.45-0.92)	0.69 (0.47-0.98)

[0596] 调整的P指数,G指数和G+P指数变量OR分别是P指数,G指数和G+P指数的取幂系数。

[0597] 用多变量逻辑回归模型计算调整的logOR系数。

[0598] $P\text{指数} = -3.78 + 0.81 \times \text{Age} + 0.46 \times \text{Gender} + 0.78 \times \text{Smoking history} + 0.59 \times \text{Hfeq}$

[0599] 如表9指定的,其中每个表型变量被分配给二进制得分0或1,并且系数的置信区间在表11给出。P指数的AUC的偏差修正估计值为0.66 (95%CI:0.55-0.67;图21)。

[0600] 肉眼可见血尿患者的基因型变量与UC风险之间的关系

[0601] G指数使用尿样中的五个基因IGFBP5,HOXA13,MDK,CDK1和CXCR2的mRNA浓度的对数通过逻辑回归进行估算以预测UC的发生。

[0602] $G\text{指数} = -6.22 + 0.77 \times \text{IGFBP5} - 1.11 \times \text{HOXA13} + 1.56 \times \text{MDK} + 1.24 \times \text{CDK1} - 0.43 \times \text{CXCR2}$

[0603] G指数给出0.83的偏差修正AUC (95%CI:0.74-0.89;图21)。

[0604] 肉眼可见血尿患者的基因型和表型变量与UC风险之间的关系

[0605] 然后五个连续基因型变量与四个二进制表型变量组合,以使用多变量逻辑回归估算G+P指数。

[0606] $G+P\text{指数} = -8.46 + (0.79 \times \text{IGF} - 1.60 \times \text{HOXA} + 2.10 \times \text{MDK} + 0.95 \times \text{CDC} - 0.38 \times \text{IL8R}) + (0.64 \times \text{Age} + 1.11 \times \text{Gender} + 0.98 \times \text{Smoking history} + 0.56 \times \text{Hfeq})$

[0607] G+P指数给出0.86的偏差修正AUC (95%CI:0.80-0.91)。

[0608] G指数和G+P指数的比较

[0609] G指数和G+P指数的置信区间之间有重叠,所以配对测试的自举版本是通过确定每个自举样本的G指数和G+P指数的AUC的差构建的。通过从可用于分析的原始587个样本随机抽样替代原始样本来生成样本大小 $n=587$ 的一万个自举样本。模型之间的差异所产生的95%CI为0.01-0.08。因此,两个AUC之间的真实差小于0.01的概率 <0.025 ,这表明G+P指数的AUC显著大于G指数的AUC的可能性高。

[0610] 模型的NPV和灵敏度

[0611] G+P指数在测试阴性率从0.2到0.7的范围内产生 $\text{NPV} > 0.97$,并且几乎总是高于G指数模型的NPV (图22)。当测试阴性率为0.4时,G+P指数提供敏感性为0.95和NPV为0.98的性能特征 (表12;图22)。与此相反,当测试阴性率为0.4时,G指数只达到0.86的灵敏度和NPV 0.96 (表12)。

[0612] 表12:当为改变测试阴性率设置阈值时,每个模型的性能特征

[0613]

阈值 (logOR)	测试阴性率 (95% CI)	NPV (95% CI)	灵敏度 (95% CI)	特异性 (95% CI)
P 指数				
-2.54	0.25(0.21-0.28)	0.97(0.92-0.99)	0.93(0.85-0.98)	0.27(0.23-0.31)
-2.52	0.38(0.34-0.42)	0.95(0.91-0.97)	0.83(0.74-0.91)	0.41(0.37-0.45)
-2.39	0.42(0.37-0.45)	0.95(0.91-0.97)	0.82(0.72-0.90)	0.45(0.40-0.49)
-1.95	0.51(0.47-0.54)	0.92(0.89-0.95)	0.68(0.56-0.78)	0.53(0.49-0.57)
-1.93	0.51(0.46-0.55)	0.92(0.89-0.95)	0.68(0.56-0.78)	0.53(0.49-0.58)
-1.73	0.82(0.79-0.85)	0.90(0.87-0.92)	0.32(0.22-0.43)	0.84(0.81-0.87)
G 指数				
-3.46	0.20(0.17-0.23)	0.94(0.88-0.97)	0.90(0.80-0.95)	0.22(0.18-0.25)
-3.23	0.30(0.26-0.34)	0.95(0.91-0.98)	0.89(0.80-0.95)	0.33(0.28-0.37)
-3.04	0.40(0.36-0.44)	0.96(0.92-0.98)	0.86(0.77-0.93)	0.44(0.40-0.48)
-2.86	0.50(0.46-0.54)	0.97(0.94-0.98)	0.86(0.77-0.93)	0.55(0.51-0.59)
-2.63	0.60(0.56-0.63)	0.96(0.94-0.98)	0.82(0.71-0.90)	0.66(0.62-0.69)
-2.41	0.70(0.66-0.73)	0.96(0.94-0.98)	0.78(0.65-0.86)	0.77(0.73-0.80)
G+P 指数				
-4.02	0.20(0.17-0.23)	0.97(0.93-0.99)	0.96(0.88-0.99)	0.22(0.19-0.26)
-3.67	0.30(0.26-0.33)	0.98(0.94-0.99)	0.94(0.87-0.99)	0.33(0.29-0.37)
-3.33	0.40(0.36-0.44)	0.98(0.95-1.00)	0.95(0.86-0.98)	0.45(0.40-0.49)
-2.99	0.50(0.46-0.54)	0.98(0.96-0.99)	0.92(0.83-0.97)	0.56(0.52-0.60)

[0614]

-2.71	0.60(0.56-0.64)	0.97(0.95-0.99)	0.86(0.76-0.93)	0.67(0.63-0.71)
-2.37	0.70(0.66-0.73)	0.97(0.94-0.98)	0.80(0.70-0.88)	0.77(0.73-0.80)

[0615] G+P指数在显微可见血尿患者中的应用

[0616] 虽然G+P指数是使用来自肉眼可见血尿患者的数据开发的,但它的稳健性在来自显微可见血尿(Hfreq=0)患者的另外40个样本中被测试。预期在显微可见血尿群体中测试阴性率更高,因为UC的发病率在该群体中较低,并且使用0.4的测试阴性率,32(80%)例患者测试为阴性并将被正确地归类出来,因此不需要全面的泌尿外科病情检查以确定UC。

[0617] 讨论

[0618] 本研究定义了一个临床工具,其提供给临床医生和医师为了UC检测有效地从需要进行全面的泌尿病情检查中归类出呈现血尿患者的能力。该研究提出了采用基于自举的CI估计值的内部验证的基因型—表型模式,G+P指数,其提供了高灵敏度和高NPV(即个别UC患者的低概率提供了假阴性结果,而所有阴性结果的高比例为真)的组合,该组合不是排他地单独由基因型或表型数据得到的模型所提供的。这为临床医生和医师提供了一个独特的机会,来归类出肉眼可见血尿和显微可见血尿患者,特别是通过识别不需要全面的泌尿外科病情检查的具有UC低风险的患者。

[0619] 在高灵敏度情况下的高测试阴性率是有效的归类测试的重要考虑因素,归类测试旨在引导UC低概率患者远离全面的临床病情检查[37]。因此,在0.4的测试阴性率时,这里提出的G+P指数的灵敏度最大化了灵敏度和NPV两者(分别为0.95和0.98)。这可以与选自[33]中公开的基因型模型中的最佳拟合(灵敏度=0.82;NPV=0.97)作比较,并且也是与膀胱镜检查(灵敏度=0.89-0.98;NPV=0.99)和使用计算机断层(CT)扫描或磁共振成像(MRI)的虚拟膀胱镜(灵敏度分别为0.94和0.91)两者的灵敏度和NPV可比较的[38-40]。

[0620] 应当承认,在这种情况下用来推导G指数,P指数和G+P指数的样本群体由肉眼可见血尿患者组成。然而,在0.4的测试阴性率下,这个对肉眼可见血尿患者测试的高灵敏度允许G+P指数被应用在肉眼可见血尿和显微可见血尿群体两者上。假定有UC和没有UC的患者类似地分布在显微可见血尿和肉眼可见血尿患者群体中,因为在显微可见血尿群体中为4%的预期UC发病率,也可以在显微可见血尿患者人群中预期高NPV。

[0621] 通过将G+P指数应用到没有UC的显微可见血尿患者的样本群体上,已表明基于结果,80%的患者会被归类出来。只有20%的会被转诊做全面的泌尿科病情检查。这与传统指南相比,目前看到所有不能归因于良性原因的显微可见血尿患者(100%)均要经历全面的泌尿科病情检查,招致不必要的重大成本并负面影响患者的QoL。

[0622] 血尿的严重程度与患者患有UC的概率相关,但与任何肿瘤的阶段或等级无关,估计转诊到泌尿科的分别为96%的显微可见血尿患者和77-88%的肉眼可见血尿患者没有UC[2-6]。因此,避免血尿患者的潜在地不必要的泌尿外科病情检查有几个好处。膀胱镜检查可与副作用,例如排尿疼痛,出血,尿路感染,男性性功能障碍和伴随不确定或未经确认的UC诊断[9-11]的焦虑相关联。最值得注意的是,这种新方法具有减少资源负担和减少与UC阴性患者的全面泌尿病情检查相关的经济成本的潜力。例如,在英国,最初的细胞学和/或肿瘤标志物测试为阴性的血尿患者避免做膀胱镜检查已经被评估为,每名被评估的患者节省约770美元(每名患者483英镑)[13]。当用在初期评估环境时,这里所描述的G+P指数提供了对使用尿细胞学的有效的替代。这在初级评估是由初级保健医生进行的环境中尤为重要。

[0623] 在此基础上,如果我们指定每次全面泌尿病情检查任意的“名义成本”为4,500美元,1000名显微可见血尿患者的病情检查总成本为接近450万美元。相反,如果任意的名义成本为2,500美元、80%的显微可见血尿患者利用G+P指数被归类出来,那么剩余20%患者接受测试和全面泌尿科病情检查的总直接成本为340万美元。这提供了每1000例显微可见血尿患者理论净节省约110万美元的直接成本。

[0624] 虽然由O'Sullivan等人开发的基因型算法[33]由与呈现在这里的G+P指数相同的

基因型成分组成,但是敏感性和特异性之间的平衡为了正在经历全面的泌尿科病情检查的有症状患者(即呈现血尿)的最佳初级UC检测而进行校准。相反,本研究中的G+P指数还纳入了表型变量,并且已经被优化为高灵敏度和高NPV,以分离出那些不需要针对疑似UC的全面的泌尿科病情检查的血尿患者。没有试图定义或选择UC患者。取而代之的目的是有把握地排除那些没有UC的患者,因此,所有没分离出来的患者将进行全面的泌尿病情检查。

[0625] 虽然当评估呈现血尿患者的UC风险时,一些研究之前已经寻求开发考虑表型的预测模型,但是单独的依赖表型的模型的准确性好像是有限的。例如,Loo等人[21]前瞻性地调查了表型参数是否可以用来确定可能不需要泌尿外科转诊和全面病情检查的显微可见血尿患者,并且得出结论,年龄、男性和近期诊断为肉眼可见血尿是UC的显著预测指标。吸烟史和在最近的尿检 $>25\text{RBCs/HPF}$ 孤立地不是UC的统计显著预测指标,但是即使当被包括在他们的‘血尿风险指数’中以提高预测的准确性时,该指数得到0.809的AUC[21]。有趣的是,对于吸烟史和性别,这项研究中的表型的OR和那些由Loo等人识别的表型参数可与重叠的95%CI相比,并且虽然年龄、性别和吸烟史在每个模型中有类似的加权,呈现在这里的G+P指数的基因型组分的影响可能引起更高的AUC[21]。

[0626] 同样,Cha等人[20]报道,年龄、吸烟史和血尿的程度,而不是性别,是与无症状性血尿患者的UC存在显著相关的,并使用了多变量模型来开发由预测UC的表型和尿细胞学检查数据组成的列线图。如同Loo等人[21]的研究,所报道的表型OR与这里报道的那些表型OR是可比的,但是即使在将尿细胞学列入到列线图之后,在[20]中报道的0.831的AUC低于G+P指数的AUC。

[0627] 在另一项研究中,Tan等人[22]使用由患者的年龄,性别,吸烟史和血尿程度得到的列线图,回顾性地分层被转诊到专门的泌尿外科诊所的血尿患者到高风险组和低风险组。虽然由于缺乏数据(405例患者中的80例)大部分的患者被排除,与本研究的比较必须谨慎,0.804的AUC,0.900的敏感性和0.953的NPV均低于这里描述的G+P指数。

[0628] 也已经进行了多种尝试通过用尿生物标志物测试的结果补充它们来改善表型模型的精确度。当护理蛋白质组学测定的核基质蛋白NMP22点被孤立地用于检测UC,它具有0.557的灵敏度和0.968的NPV[17]。Lotan等人[41]发表了多变量算法,包括表型因素,NMP22和尿细胞学和用0.802[31]的AUC进行前瞻性验证的预测UC的0.826的AUC。但是,要注意的是,这个模型试图在有和没有UC的高风险患者之间作区分,而不是最大化敏感性和NPV以归类出UC低概率的患者。

[0629] 特别是,用包含基因型和表型数据的算法得到的改进的精度以前已经在乳腺癌中证明了[42-45]。同样,Mitra等人[30]使用分子标志物和吸烟强度的组合来计算多变量模型,其优于在预测UC患者幸存的常规临床病理参数。然而,本研究首次证明表型风险因素可以与基因型数据相结合,以增加模型的准确性,该模型分离血尿患者到需要不同水平的泌尿外科跟进和临床护理的类别,而不是预测幸存的。

[0630] 当表型数据与模型中的基因型数据相结合,数据的解析可能影响模型的准确度。例如,吸烟是很好理解的UC风险因素,并且包括在用于检测UC的大多数表型模型中。在Cha等人[20],Tan等人[22],Lotan等人[31,41]和当前的研究中,使用二进制判别式从不吸烟者和当前/以前吸烟者,而Mitra等人[30]基于吸烟年限和每天吸烟数量计算吸烟强度,Loo等人[21]分类吸烟者为从不吸烟,被动吸烟者,已戒烟的吸烟者和当前吸烟者。虽然已知的

是暴露于吸烟显著增加UC风险[46],随意地定义的表型变量可以限制表型模型的总体精度和效用。与此相反,患者的基因型和表型变量之间的相互作用就不会无法预期。然而,结合表型因素和遗传变量的影响到单个工具中改进了这项研究中所描述的模型的准确度。类似的原则也适用于描述血尿表型。呈现显微可见血尿或肉眼可见血尿的患者基本上是在生物连续体上,并且有患有UC[2-6,21]的不同可能性。因此,尽管本研究中的所有显微可见血尿患者具有0的Hfreq得分,他们血尿的严重程度,与其他表型因素组合,可能是间接地体现在G+P指数的基因型成分中。

[0631] 参考文献

[0632] 紧接下面列举的文章参考示例4,并且通过充分引用全部并入本说明书中。

[0633] 1.Kelly JD,Fawcett DP,Goldberg LC:Assessment and management of non-visible haematuria in primary care.BMJ2009,338:a3021.

[0634] 2.Davis R,Jones JS,Barocas DA,Castle EP,Lang EK,Leveillee RJ,Messing EM,Miller SD,Peterson AC,Turk TMT,Weitzel W:Diagnosis,evaluation and follow-up of asymptomatic microhematuria (AMH) in adults:AUA guideline.J Urol 2012,188(6Suppl):2473-2481.

[0635] 3.Sutton JM,Evaluation of hematuria in adults.JAMA 1990,263:2475-2480.

[0636] 4.Khadra MH,Pickard RS,Charlton M,Powell PH,Neal DE:A prospective analysis of 1,930 patients with hematuria to evaluate clinical practice.J Urol 2000,163:524-527.

[0637] 5.Davidson P:Re-design of a haematuria clinic:Assessment of 2346 haematuria patients.J Urol 2011,185(4S):e495.

[0638] 6.Price SJ,Shephard EA,Stapley SA,Barraclough K,Hamilton WT:Non-visible versus visible haematuria and bladder cancer risk:A study of electronic records in primary care.Br J Gen Pract 2014,64:e584-e589.

[0639] 7.Buteau A,Seideman CA,Svatek RS,Youssef RF,Chakrabati G,Reed G,Bhat D,Lotan Y:What is evaluation of hematuria by primary care physicians?Use of electronic medical records to assess practice patterns with intermediate follow-up.Urol Oncol 2014,32:128-134.

[0640] 8.Jimbo M:Evaluation and management of hematuria.Prim Care 2010,373:461-472.

[0641] 9.Herr HW:The risk of urinary tract infection after flexible cystoscopy in bladder tumor patients who did not receive prophylactic antibiotics.J Urol 2014 Jul 18.pii:S00225347(14)03963-9.

[0642] 10.Burke DM,Shackley DC,O'Reilly PH:The community-based morbidity of flexible cystoscopy.BJU Int 2002,89:347-349.

[0643] 11.Stav K,Leibovici D,Goren E,Livshitz A,Siegel YI,Lindner A,Zisman A:Adverse effects of cystoscopy and its impact on patients'quality of life and sexual performance.Isr Med Assoc J 2004,6:474-478.

- [0644] 12.Rao PK,Jones JS:How to evaluate'dipstick hematuria':What to do before you refer.Cleve Clin J Med 2008,75:227-233.
- [0645] 13.Rodgers M,Nixon J,Hempel S,Aho T,Kelly J,Neal D,Duffy S,Ritchie G,Kleijnen J,Westwood M:Diagnostic tests and algorithms used in the investigation of haematuria:systematic reviews and economic evaluation.Health Technol Assess 2006,10:iii-iv,xi-259.
- [0646] 14.Falebita OA,Lee G,Sweeney P:Urine cytology in the evaluation of urological malignancy revisited:is it still necessary?Urol Int 2010,84:45-49.
- [0647] 15.Feifer AH,Steinberg J,Tanguay S,Aprikian AG,Brimo F,Kassouf W:Utility of urine cytology in the workup of asymptomatic microscopic hematuria in low-risk patients.Urology 2010,75:1278-1282.
- [0648] 16.Svatek RS,Hollenbeck BK,HolmängS, Lee R,Kim S,Stenzl A,Lotan Y.The economics of bladder cancer:Costs and considerations of caring for this disease.Eur Urol 2014,66:253-262.
- [0649] 17.Grossman HB,Messing E,Soloway M,Tomera K,Katz G,Berger Y,Shen Y:Detection of bladder cancer using a point-of-care proteomic assay.JAMA 2005,293:810-816.
- [0650] 18.Friedlander DF,Resnick MJ,You C,Bassett J,Yarlagadda V,Penson DF,Barocas DA:Variation in the intensity of hematuria evaluation:a target for primary care quality improvement.Am J Med 2014,127:633-640.
- [0651] 19.Shinagare AB,Silverman SG,Gershanik EF,Chang SL,Khorasani R:Evaluating hematuria:impact of guideline adherence on urologic cancer diagnosis.Am J Med 2014,127:625-632.
- [0652] 20.Cha EK,Tirsar LA,Schwentner C,Hennenlotter J,Christos PJ,Stenzl A,Mian C,Martini T,Pycha A,Shariat SF,Schmitz-Dräger BJ:Accurate risk assessment of patients with asymptomatic hematuria for the presence of bladder cancer.World J Urol 2012,30:847-852.
- [0653] 21.Loo RK,Lieberman SF,Slezak JM,Landa LM,Mariani AJ,Nicolaisen G,Aspera AM,Jacobsen SJ:Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria.Mayo Clin Proc 2013,88:129-138.
- [0654] 22.Tan GH,Shah SA,Ann HS,Hemdan SN,Shen LC,Galib NAFA,Singam P,Kong CHC,Hong GE,Bahadzor B,Zainuddin ZM:Stratifying patients with haematuria into high or low risk groups for bladder cancer:a novel clinical scoring system.Asian Pac J Cancer Prev 2013,14:6327-6330.
- [0655] 23.Sultana S,Goodman C,Bryne D,Baxby K:Microscopic haematuria:urological investigation using a standard protocol.Br J Urol 1996,78:691-698.
- [0656] 24.Sugimura K,Ikemoto S-I,Kawashima H,Nishisaka N,Kishimoto T:Microscopic hematuria as a screening marker for urinary tract

malignancies. *Int J Urol* 2001,8:1-5.

[0657] 25. Viswanath S, Zelhof B, Ho E, Sethia K, Mills R: Is routine urine cytology useful in the haematuria clinic? *Ann R Coll Surg Engl* 2008,90:153-155.

[0658] 26. Steiner H, Bergmeister M, Verdorfer I, Granig T, Mikuz G, Bartsch G, Stoehr B, Brunner A: Early results of bladder-cancer screening in a high-risk population of heavy smokers. *BJU Int* 2008,102:291-296.

[0659] 27. Yeung C, Dinh T, Lee J: The health economics of bladder cancer: An updated review of the published literature. *Pharmacoeconomics* 2014,32:1093-1104.

[0660] 28. Holyoake A, O'Sullivan P, Pollock R, Best T, Watanabe J, Kajita Y, Matsui Y, Ito M, Nishiyama H, Kerr N, da Silva Tatley F, Cambridge L, Toro T, Ogawa O, Guilford P: Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder. *Clin Cancer Res* 2008,14:742-749.

[0661] 29. Sapre N, Anderson PD, Costello AJ, Hovens CM, Corcoran NM: Gene-based urinary biomarkers for bladder cancer: an unfulfilled promise? *Urol Oncol* 2014,32:48.e9-e17.

[0662] 30. Mitra AP, Castelao JE, Hawes D, Tsao-Wei DD, Jiang X, Shi SR, Datar RH, Skinner EC, Stein JP, Groshen S, Yu MC, Ross RK, Skinner DG, Cortessis VK, Cote RJ: Combination of molecular alterations and smoking intensity predicts bladder cancer outcome. *Cancer* 2013,119:756-765.

[0663] 31. Lotan Y, Svatek RS, Krabbe LM, Xylinas E, Klatte T, Shariat SF: Prospective external validation of model for bladder cancer detection. *J Urol* 2014,192:1343-1348.

[0664] 32. Abogunrin F, O'Kane HF, Ruddock MW, Stevenson M, Reid CN, O'Sullivan JM, Anderson NH, O'Rourke D, Duggan B, Lamont JV, Boyd RE, Hamilton P, Nambirajan T, Williamson KE: The impact of biomarkers in multi-variate algorithms for bladder cancer diagnosis in patients with hematuria. *Cancer* 2012,118:2641-2650.

[0665] 33. O'Sullivan P, Sharples K, Dalphin M, Davidson P, Gilling P, Cambridge L, Harvey J, Toro T, Giles N, Luxmanan C, Felipe Alves C, Yoon HS, Hinder V, Masters J, Kennedy-Smith A, Beaven T, Guilford P: A multigene urine test for the detection and stratification of bladder cancer in patients presenting with hematuria. *J Urol* 2012,188:741-747.

[0666] 34. Epstein JI, Amin MB, Reuter VR, Mostofi FK: The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of the urinary bladder. Bladder Consensus Conference Committee. *Am J Surg Pathol* 1998,22:1435-

1448.

[0667] 35.Efron B,Tibshirani RJ:An introduction to the bootstrap.New York: Chapman Hill;1993.

[0668] 36.DiCiccio TJ,Tibshirani RJ:Bootstrap confidence intervals.Statist Sci 1996,11:189-228.

[0669] 37.Van't Hoog AH,Cobelens F,Vassall A, van Kampen S,Dorman SE,Alland D,Ellner J:Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis:A decision analysis.PLoS ONE 2013,8:e82786.

[0670] 38.Qu X,Huang X,Wu L,Huang G,Ping X,Yan W:Comparison of virtual cystoscopy and ultrasonography for bladder cancer detection:A meta-analysis.Eur J Radiol 2010,80:188-197.

[0671] 39.Mowatt G,N'Dow J,Vale L,Nabi G,Boachie C,Cook JA,Fraser C,Griffiths TR;Aberdeen Technology Assessment Review (TAR) Group:Photodynamic diagnosis of bladder cancer compared with white light cystoscopy:Systematic review and meta-analysis.Int J Technol Assess Health Care 2011,27:3-10.

[0672] 40.Buick CG,Nazir SA,Mallett S,Turney BW,Onwu NN,Roberts IS,Crew JP,Cowan NC:Evaluation of diagnostic strategies for bladder cancer using computed tomography (CT) urography,flexible cystoscopy and voided urine cytology:results for 778 patients from a hospital haematuria clinic.BJU Int 2012,110:84-94.

[0673] 41.Lotan Y,Capitanio U,Shariat SF,Hutterer GC,Karakiewicz PI:Impact of clinical factors,including a point-of-care nuclear matrix protein-22 assay and cytology,on bladder cancer detection.BJU Int 2009,103:1368-1374.

[0674] 42.Tyrer J,Duffy SW,Cuzick J:A breast cancer prediction model incorporating familial and personal risk factors.Stat Med 2004,23:1111-1130.

[0675] 43.Dubsky P,Brase JC,Jakesz R,Rudas M,Singer CF,Greil R,Dietze O,Luissner I,Klug E,Sedivy R,Bachner M,Mayr D,Schmidt M,Gehrmann MC,Petry C,Weber KE,Fisch K,Kronenwett R,Gnant M,Filipits M,on behalf the Austrian Breast and Colorectal Cancer Study Group (ABCSG):The EndoPredict score provides prognostic information on late distant metastases in ER+THER2—breast cancer patients.Br J Cancer 2013,109:2959-2964.

[0676] 44.Brentnall AR,Evans DG,Cuzick J:Distribution of breast cancer risk from SNPs and classical risk factors in women of routine screening age in the UK.Br J Cancer 2014,110:827-828.

[0677] 45.Filipits M,Nielsen TO,Rudas M,Greil R,Stöger H,Jakesz R,Bago-Horvath Z,Dietze O,Regitnig P,Gruber-Rossipal C,Müller-Holzner E,Singer CF,Mlineritsch B,Dubsky P,Bauernhofer T,Hubalek M,Knauer M,Trapl H,Fesl C,Schaper C,Ferree S,Liu S,Cowens JW,Gnant M,Austrian Breast and Colorectal

Cancer Study Group:The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer.Clin Cancer Res 2014,20:1298-1305.

[0678] 46.Krabbe LM,Svatek RS,Shariat SF,Messing E,Lotan Y:Bladder Cancer Risk:Use of the PLCO and NLST to Identify as Suitable Screening Cohort.Urol Oncol.20149 July 16):pii:S1078-1439(14)00215-4.

[0679] 结论

[0680] 总之,这里报告的G+P指数表明了改变临床实用的重大机会。G+P指数能够准确地归类出患者,其呈现血尿给他们的临床医生或医师,并具有高整体测试阴性率、高水平灵敏度和高NPV的UC概率很低。这个模型适合于在血尿患者的初期评估过程中使用以归类出不需要全面的泌尿科病情检查的患者,从而潜在地减少了为了针对UC的专业泌尿外科评估而需要转诊的血尿患者的数量,并有助于保持患者的QoL以及减少诊断相关的费用。

[0681] 示例5:使用G+P指数III归类血尿患者(“G2”)

[0682] G+P指数进一步用于根据UC风险对患者进行分类。然后风险类别用于为患者划分优先次序用于继续调查。

[0683] 本示例提供了用于使用指数归类(triaging)患者的替代性方法,在本说明书中称为“G2+P指数”。它与上述示例4描述的G+P指数相似。

[0684] G2的定义

[0685] 这个分类器计算下面的公式:

[0686] $M1 = [IGF] - [HOXA] + [MDK] + [CDC]$;

[0687] 然后,

[0688] $R = -6.9802$

[0689] $+0.2007 * M1$

[0690] $+1.7893 * [IL8R]$

[0691] $+0.1552 * M1^2$

[0692] $-0.2882 * [IL8R]^2$

[0693] $-0.0720 * M1 * [IL8R]$

[0694] 获得的得分:

[0695] $SCORE = e^R / (1 + e^R)$, 其中 [IGF], [HOXA], [MDK], [CDC] 和 [IL8R] 分别是基因 IGF, HOXA, MDK, CDC 和 IL8R 的样本浓度的对数; “*” 是普通的乘法, $e = 2.718282 \dots$ 是自然 (Napierian) 对数的底数。

[0696] 高得分表示存在UC的可能性较高。作为示例,我们可以设定0.12的阈值并宣布 $SCORE \geq 0.12$ 作为患有UC的可能性高,低于0.12的得分为患有UC的可能性低。

[0697] 示例6:同时使用 [G1=P] 一指数和G2归类出低概率尿路上皮癌的呈现血尿的患者

[0698] A. 显微可见血尿患者

[0699] 患者数据集

[0700] • 45例患者样本被测定。

[0701] • 5例患者缺乏吸烟状况并从本分析中被排除。

[0702] • 所有患者都完成了全面的泌尿科病情检查,并且都没有尿路上皮癌。

[0703] • 表型变量: 性别, 年龄, 吸烟状况

[0704] • 病人的人口统计示于下表13。

[0705] 表13

[0706]

性别	年龄	不是吸烟者	前任吸烟者	吸烟者
----	----	-------	-------	-----

[0707]

女性	年龄<60	10	1	2
	年龄≥60	9	1	2
男性	年龄<60	3	2	3
	年龄≥60	3	2	2

[0708] 观察到的测试阴性率:

[0709] 使用测试阴性率40%作为阈值, 数据示于下表14。

[0710] 表14

[0711]

测试阴性率40%	LR<0.12	0.12≤LR<0.23	0.23≤LR
归类阴性	28	3	1
归类阳性	4	2	2

[0712] 使用测试阴性率50%作为阈值, 数据示于下表15。

[0713] 表15

[0714]

测试阴性率50%	LR<0.12	0.12≤LR<0.23	0.23≤LR
归类阴性	28	3	3
归类阳性	4	2	0

[0715] 使用测试阴性率60%作为阈值给出了相同的结果/

[0716] 对此归类分类, 风险最高组 (男, 当前或之前吸烟者, 年龄≥60) 均为阳性。

[0717] 总结和结论

[0718] 1. 根据临床指导原则的约定, 所有41例患者将通常会接受全面泌尿病情检查。

[0719] 2. 所有41例患者确实接受了全面的病情检查并且被确定没有尿路上皮癌。

[0720] 3. 在测试阴性率 (TNR) 为50%, 85%的显微可见血尿患者将被筛选出, 因此, 结果是不用接受全面的泌尿病情检查。

[0721] 4. 如果在50%的TNR使用Cxbldadder-归类 (归类指数-3.33), 则85.4%的患者将被归类出来, 因此, 正确地将不会接受全面的泌尿病情检查。

[0722] 5. 如果在40%的TNR使用Cxbldadder-归类 (归类指数-3.0), 则80.5%的患者将被

归类出来,因此,结果是正确地将不会接受全面的泌尿病情检查。

[0723] B.肉眼可见血尿患者

[0724] 患者数据集

[0725] 使用来自临床试验数据和North Shore and CURT产品试用的587个样本。这个数据集是由包含年龄,性别,吸烟状况和血尿频率,以及IGF,H0XA,MDK,CDC,IL8R的基因浓度的完整数据构成的子集。

[0726] 我们使用了用于开发下面的Cxb bladder-归类模型的相同数据;

[0727] $G+P \text{ 指数} = -8.46 + 0.79 \text{IGF} - 1.60 \text{H0XA} + 2.10 \text{MDK} + 0.95 \text{CDC} - 0.38 \text{IL8R} + 0.98 \text{SNS} + 0.56 \text{Hfreq} + 1.11 \text{Gender} + 0.64 \text{Age}$

[0728] 我们在图24中绘制了针对G2诊断得分的归类得分。归类阈值-3.33,-2.99和-2.71分别对应于测试阴性率40%,50%和60%。垂直线是0.12和0.23的Cxb bladder阈值。实心圆对应肿瘤;绿色是Ta,红色是所有其他更高阶段的肿瘤。表16示出了临床研究结果。

[0729] 表16

[0730]

阶段	无癌症	T1	T2	T2a	T3	Ta	Tis
数目	515	16	10	1	2	40	3

[0731] 我们考虑了图24的4个象限的计数并且通过各个临界值确定归类指数和G2。

[0732] -3的Cxb bladder-归类(归类指数)的阈值对应于50%的测试阴性率。表17示出了这些结果。

[0733] 表17

[0734]

象限	阈值	对照组	TCC
----	----	-----	-----

[0735]

左下	归类<-3 并且 G2<0.12	272	6
左上	归类>-3 并且 G2<0.12	155	7
右下	归类<-3 并且 G2>=0.12	16	0
右上	归类>-3 并且 G2>=0.12	72	59

[0736] 使用-3.33的Cxb bladder-归类阈值(40%的试验阴性率)和相同的0.12的G2阈值,我们观察到的数据示于表18。

[0737] 表18

[0738]

象限	阈值	对照组	TCC
左下	归类<-3.33 并且 G2<0.12	222	4
左上	归类>-3.33 并且 G2<0.12	205	9
右下	归类<-3.33 并且 G2>=0.12	9	0
右上	归类>-3.33 并且 G2>=0.12	79	59

[0739] 总结和结论

[0740] 1. 在相同的时间间隔对相同患者使用Cxb bladder归类 (G+P) 和Cxb bladder检测 (G2) 的系列组合把患者全面分离到四个主要临床分组中。

[0741] 2. 对于呈现显微可见血尿和肉眼可见血尿的患者的组合群体, 并使用40%的试验阴性率, 587例患者中的235例 (40.0%) 被归类出来。发明人的结论是这些患者将不需要检测UC的全面的病情检查。

[0742] 3. 对于同一群体, 相应的剩余组352 (60%) 例患者将接受全面的泌尿科病情检查。

[0743] 4. 这个剩余组包含了所有的高级和晚期肿瘤。这些患者没有被归类出来, 因此结果是将正确地接受全面的泌尿科病情检查。

[0744] 5. 共有4例低级Ta' s (肿瘤总数的5.6%) 将被归类出来, 并且不会接受全面的病情检查。

[0745] 6. 如果归类规则允许归类出所有低于40%的测试阴性率并且G2得分<0.12的患者, 则共有13例低级Ta' s被归类出来, 其将不会接受全面的病情检查, 并且所有高级晚期肿瘤会被找出并接受全面的泌尿科病情检查。

[0746] 7. 这些修改的归类规则也导致在所有587例患者中共有440例患者被归类出来 (75%)。

[0747] 优势和总体结论

[0748] 结论是, 这里报告的G+P指数表明了改变临床实用的显著机会。G+P指数能够准确地归类出患者, 其呈现血尿给他们的临床医生或医师, 其具有高整体测试阴性率、高水平灵敏度和高NPV的UC低概率。这个模型适合于由初级护理医生用于归类不需要全面的泌尿科病情检查的患者, 从而减少了需要转诊以接受专业泌尿外科UC评估的血尿患者的数量, 并有助于保持患者的生活质量以及减少了诊断相关的费用。本公开的方法提供了对那些血尿患者的继续调查缺乏需求的出人意料地准确的评估。这些代表不使用本说明书中公开的方法已经不能取得出色的效果。

[0749] 通过引用并入

[0750] 所有专利、专利申请和非专利文献引文全部通过引用并入本说明书中,如同分别以如此方式并入。

[0751] 行业实用性

[0752] 本发明的实施例是在医疗保健和医药领域是有用的。

技术领域

[0753]

[0754] 本发明的实施例提供了高度准确、灵敏和特异性的计算机实现的方法,用于归类(triaging)患者,以确定哪些患者不需要真正的短期程序或后续处理。该方法基于来自糖尿病患者的特定遗传和表型信息分析通过提供新的和非显而易见的计算机操作来改进计算机操作,其产生真实的,有用的和具体的成果,以提高医疗质量和降低成本。

序列表

<110> 环太平洋有限公司
 <120> 使用基因型生物标志物和表型生物标志物来归类无临床症状糖尿病患者
 <130> PEBL 1028 W01
 <160> 17
 <170> 专利3.5版
 <210> 1
 <211> 360
 <212> PRT
 <213> 人

<400> 1
 Met Glu Asp Phe Asn Met Glu Ser Asp Ser Phe Glu Asp Phe Trp Lys
 1 5 10 15
 Gly Glu Asp Leu Ser Asn Tyr Ser Tyr Ser Ser Thr Leu Pro Pro Phe
 20 25 30
 Leu Leu Asp Ala Ala Pro Cys Glu Pro Glu Ser Leu Glu Ile Asn Lys
 35 40 45
 Tyr Phe Val Val Ile Ile Tyr Ala Leu Val Phe Leu Leu Ser Leu Leu
 50 55 60
 Gly Asn Ser Leu Val Met Leu Val Ile Leu Tyr Ser Arg Val Gly Arg
 65 70 75 80
 Ser Val Thr Asp Val Tyr Leu Leu Asn Leu Ala Leu Ala Asp Leu Leu
 85 90 95
 Phe Ala Leu Thr Leu Pro Ile Trp Ala Ala Ser Lys Val Asn Gly Trp
 100 105 110
 Ile Phe Gly Thr Phe Leu Cys Lys Val Val Ser Leu Leu Lys Glu Val
 115 120 125
 Asn Phe Tyr Ser Gly Ile Leu Leu Ala Cys Ile Ser Val Asp Arg
 130 135 140
 Tyr Leu Ala Ile Val His Ala Thr Arg Thr Leu Thr Gln Lys Arg Tyr
 145 150 155 160
 Leu Val Lys Phe Ile Cys Leu Ser Ile Trp Gly Leu Ser Leu Leu Leu
 165 170 175
 Ala Leu Pro Val Leu Leu Phe Arg Arg Thr Val Tyr Ser Ser Asn Val
 180 185 190
 Ser Pro Ala Cys Tyr Glu Asp Met Gly Asn Asn Thr Ala Asn Trp Arg
 195 200 205
 Met Leu Leu Arg Ile Leu Pro Gln Ser Phe Gly Phe Ile Val Pro Leu
 210 215 220
 Leu Ile Met Leu Phe Cys Tyr Gly Phe Thr Leu Arg Thr Leu Phe Lys
 225 230 235 240
 Ala His Met Gly Gln Lys His Arg Ala Met Arg Val Ile Phe Ala Val
 245 250 255
 Val Leu Ile Phe Leu Leu Cys Trp Leu Pro Tyr Asn Leu Val Leu Leu
 260 265 270
 Ala Asp Thr Leu Met Arg Thr Gln Val Ile Gln Glu Thr Cys Glu Arg
 275 280 285
 Arg Asn His Ile Asp Arg Ala Leu Asp Ala Thr Glu Ile Leu Gly Ile
 290 295 300
 Leu His Ser Cys Leu Asn Pro Leu Ile Tyr Ala Phe Ile Gly Gln Lys
 305 310 315 320
 Phe Arg His Gly Leu Leu Lys Ile Leu Ala Ile His Gly Leu Ile Ser
 325 330 335
 Lys Asp Ser Leu Pro Lys Asp Ser Arg Pro Ser Phe Val Gly Ser Ser
 340 345 350
 Ser Gly His Thr Ser Thr Thr Leu
 355 360

[0001]

<210> 2
 <211> 2880
 <212> DNA
 <213> 人

<400> 2
 aggttcaaaa cattcagaga cagaaggtag atagacaaat ctccaccttc agactgtag 60
 gctctccag aagccatcag acaggaagat gtgaaaatec ccagcactca tcccagaatc 120
 actaagtgc acctgtctg ggcacaaagc ccaggacaga cctcattgtt cctctgtggg 180
 aataacctcc caggaggga tcttgattt ccccttgca aeccaggtca gaagtttcat 240
 cgtcaagggtt gtttcatctt ttttttctg tctaacagct ctgactacca cccaaccttg 300
 aggcacagtg aagacatcgg tggecactec aataacagea ggtaacagct getcttcagg 360
 aggtgtccta caggtagaaa gccagcgac ccagtcagga ttttaagtta cctcaaaaat 420
 ggaagatttt aacatggaga gtgacagctt tgaagatttc tggaaagggt aagatcttag 480
 taattacagt tacagctcta cctgcccc ttttctacta gatgccgcc catgtgaacc 540

	agaatccctg gaaatcaaca agtattttgt ggctattatc tatgcccctg taticctget	600
	gagcctgctg ggaactcecc tcgtgatgct ggctatctta tacagcaggg tcggccgctc	660
	cgtcactgat gtctaccfgc tgaacctagc cttggccgac ctactctttg ccttgacctt	720
	gcccatctgg gccgacctca aggtgaatgg ctggattttt ggcacattcc tgtgcaaggt	780
	ggctcactc ctgaaggaag tcaacttcta tagtggcacc ctgetactgg cctgcatcag	840
	tgtggaccgt taactggcca ttgtccatgc cacacgcaca ctgaccacaga agcgtacttt	900
	ggtaaaattc atatgtctca geatctgggg tctgtccctg ctcciggecc tgcctgtctt	960
	acttttccga aggaccgtct actcatccaa tgttagccca gcctgctatg aggacatggg	1020
	caacaatata gcaaacctggc ggatgctgtt accgatactg ccccgactct ttggtttcat	1080
	cgtgccactg ctgatcatgc tgttctgcta cggattcacc ctgcgtacgc tgtttaaggg	1140
	ccacatgggg cagaagcacc gggecatgcg ggctatcttt gctgtcgtcc tcactttcct	1200
	gctctgtctg ctgacctaca aectggctct gctggcagac accctcatga ggaccaggt	1260
	gattccaggag acctgtgagc gccgcaatca cctgaccggc gctctggatg ccaccgagat	1320
	tctgggcacc ctccacagct gcccaacccc cctcatctac gccttcattg gccagaagtt	1380
	tcgccatgga ctctctcaaga ttctagctat acatggcttg atcagcaagg actcctgccc	1440
	caaaagacagc aggccttccct ttgttgctc ttcttcaggc cacacttcca ctactctcta	1500
	agacctcctg cetaagtcca gccccgtggg gttcctccct tctcttcaca gtcacattcc	1560
	aagcctcatg tccactgggt ctctctggct ccagtgctca tgcagccccc attgtgtgca	1620
	caggaaatag aggaggccac gttcttacta gtttcccttg catggtttag aaagcttgcc	1680
	ctggtgctc accccttgcc ataattacta tgcattttgc tggagctctg cccatcctgc	1740
	ccctgagccc atggcactct atgttctaa agtgaaaaat ctacactcca gtgagacagc	1800
	tctgcatact cattaggatg gctagiatca aaagaaagaa aatcaggctg gccaacgggg	1860
	tgaaacctg tctctactaa aaatacaaaa aaaaaaaata attagccggg cgtggtgtg	1920
	agtgcctgta atcacagcta cttgggagge tgagatggga gaatcacttg aaccggggag	1980
	gcagaggttg cagtagcccg agattgtgcc cctgcactcc agcctgagcg acagtagagc	2040
	tctgtctcag tccatgaaga tctagaggag aaactggaac tctcagcgt tgcctggggg	2100
	gattgtaaaa tgggtgacc actgcagaag acagtatggc agctttcttc aaaacttcag	2160
	acatagaatt aacacatgat cctgcaattc cacttatagg aattgaccca caagaaatga	2220
	aagcagggac ttgaacccat atttgiacac caatattcat agcagcttat tcacaagacc	2280
	caaaaggcag aagcaaccca aatgttcate aatgaatgaa tgaatggcta agcaaaatgt	2340
	gatattgaac taacgaagta tecttcagcc tgaagagga atgaagtact catacatgtt	2400
	acaacacgga cgaaccttga aaactttatg ctaagtgaan taagccagac atcaacagat	2460
	aaatagttaa tgattccacc tacatgaggf actgagagtg aacaaattta cagagacaga	2520
	aagcagaaca gtgattacca gggactgagg ggaggggagc atgggaagtg acggtttaat	2580
	gggcacaggg ttatgtttaa ggatgttgaa aaagtcttgc agataaacag tagtgatagt	2640
	tgtaccgcaa tgtgacttaa tgccactaaa ttgacactta aanaatgttt aaatgggtcaa	2700
	ttttgttatg tatattttat atcaatttaa aaaaaaacct gagccccaaa aggtatttta	2760
	atcaccaggg ctgattaaac caaggctaga accacctgce tatatttttt gttaaatgat	2820
	ttcattcaat atcttttttt taataaacca tttttacttg ggtgtttata aaaaaaaaaa	2880

[0002]

<210> 3
 <211> 18
 <212> DNA
 <213> 合成

<400> 3
 tgcaccccc aagacccaaa 18

<210> 4
 <211> 26
 <212> DNA
 <213> 合成

<400> 4
 tgattaaagc taacgagcag acagaa 26

<210> 5
 <211> 26
 <212> DNA
 <213> 合成

<400> 5
 ccttcccttt ctggctttg gccttt 26

<210> 6
 <211> 22
 <212> DNA
 <213> 合成

<400> 6
 cgttgtacct gcccaattgt ga 22

<210> 7
 <211> 20
 <212> DNA

	<213>	合成	
	<400>	7	
		gggacgcatc actcaacgtt	20
	<210>	8	
	<211>	27	
	<212>	DNA	
	<213>	合成	
	<400>	8	
		aagagaaagc agtgcaaacc ttcccgt	27
	<210>	9	
	<211>	16	
	<212>	DNA	
	<213>	合成	
	<400>	9	
		gccgccgccc aataat	16
	<210>	10	
	<211>	28	
	<212>	DNA	
	<213>	合成	
	<400>	10	
		tgtctacctt tatcacaaac tccatagg	28
	<210>	11	
	<211>	31	
	<212>	DNA	
	<213>	合成	
	<400>	11	
		agccgggatac taccataccc attgactaac t	31
[0003]	<210>	12	
	<211>	20	
	<212>	DNA	
	<213>	合成	
	<400>	12	
		tggaacggcc aaatgtactg	20
	<210>	13	
	<211>	20	
	<212>	DNA	
	<213>	合成	
	<400>	13	
		tggcgtattc ccgttcaagt	20
	<210>	14	
	<211>	23	
	<212>	DNA	
	<213>	合成	
	<400>	14	
		aetctgcccg acgtggtctc cca	23
	<210>	15	
	<211>	23	
	<212>	DNA	
	<213>	合成	
	<400>	15	
		ccttgaggca cagtgaagac atc	23
	<210>	16	
	<211>	23	
	<212>	DNA	
	<213>	合成	
	<400>	16	

	cctgtaggac acctccagaa gag	23
[0004]	<210> 17	
	<211> 27	
	<212> DNA	
	<213> 合成	
	<400> 17	
	tggccactcc aataacagca ggtcaca	27

位点 NM_001557 2880 bp mRNA
 定义: 智人趋化因子 (C-X-C基序) 受体2 (CXCR2), 转录变体1, mRNA

登录号 NM_001557
 版本 NM_001557.3 GI:269973857

蛋白质序列: "MEDFNMESDSFEDFWKGEDLSNYSYSTLPFPFLDDAAPCEPESL
 EINKYFVVIYALVFLSLGNSLVLILYSRVGRSVTDVYLLNLALADLLFALTLP
 IWAASKVNGWIFGTFELCKVVSLLKEVNFYSGILLACISVDRYLAIVHATRTLTKRY
 LVKFICLSIWGLSLLALPVLFRRTVYSSNVSPACYEDMGNNNTANWRMLLRILPQSF
 GPIVPLLIMLFCYGETLRTLFKAHMGQKHRAMRVIFAVVLIPLLCWLPYNLVLLADTLL
 MRTQVIQETCERRNHIDRALDATEILGILHSCINPLIYAFIGQKFRHGLLKILAIHGL
 ISKDSLPLKDSRPSFVCGSSSGHTSTTL"

起源

```

1 aggttcaaaa cattcagaga cagaaggtgg atagacaaaat ctccaccttc agactggtag
61 gctcctccag aagccatcag acaggaagat gtgaaaaatcc ccagcaactca tcccagaatc
121 actaagtggc acctgtcctg ggccaaagtc ccaggacaga cctcattggt cctctgtggg
181 aataacctcc caggagggca tcttggattt ccccttgca accagggtca gaagtttcat
241 cgtcaagggt gtttcatctt ttttttcttg tctaaccagt ctgactacca cccaaccttg
301 aggcacagtg aagacatcgg tggccactcc aataacagca ggtcacagct gotcttctgg
361 aggtgtccta caggtgaaaa gccccagcgac ccagtcaggga tttaagttta cctcaaaaat
421 ggaagatttt aacatggaga gtgacagctt tgaagatttc tggaaaagggt aagatcttag
481 taattacagt tacagctcta cctgcccccc ttttctacta gatgccgccc catgtgaacc
541 agaatccctg gaaatcaaca agtattttgt ggtcattatc tatgccctgg tattcctgct

601 gagcctgctg ggaactccc tcgtgatgct ggtcatctta tacagcaggg toggccgctc

```

图1

```

661 cgtcaactgat gtctaacctgc tgaacctagc cttggccgac ctactctttg ccctgacctt
721 gcccatctgg gccgcctcca aggtgaatgg ctggattttt ggcacattcc tgtgcaagggt
781 ggtctcactc ctgaagggaag tcaacttcta tagtggcatc ctgctactgg cctgcatcag
841 tgtggaccctg tacctggcca ttgtccatgc cacacgcaca ctgaccaga agcgtacttt
901 ggtcaaatc atatgtctca gcatctgggg tctgtccttg ctctggccc tgcctgtctt
961 acttttccga aggaccgtct actcatccaa tcttagccca gcctgctatg aggacatggg
1021 caacaataca gcaaaactggc ggalgtgtt acggatcctg cccagtcct ttggcttcat
1081 cgtgccactg ctgacatgc tgttctgcta cggattcacc ctgcglacgc tgtttaaggc
1141 ccacatgggg cagaagcac gggccatgcg ggtcatcttt gctgtcgtcc tcatcttctt
1201 gctctgctgg ctgccctaca acctgggtcct gctggcagac acctcatga ggacccagggt
1261 gatccaggag acctgtgagc gccgcaatca catcgaccgg gctctggatg ccaccgagat
1321 tctggggcatc cttcacagct gcccaaccc cctcatctac gccttcattg gccagaagtt
1381 tcgccatgga ctccctcaaga ttctagctat acatggcttg atcagcaagg actccctgcc
1441 caaagacagc aggccttctt ttgttggtc ttcttcaggg cacacttcca ctactctcta
1501 agacctcctg cctaagtgca gccctgggtc gttctcctct tctcttcaaa gtcacattcc
1561 aagcctcatg tccactgggt cttcttggtc tcagtggtcaa tgcagccccc atttgtgtca
1621 caggaagtag aggaggccac gttcttacta gtttcccttg catggtttag aaagcttggc
1681 ctggtgcttc acctctgccc ataattacta tgtcatltgc tggagctctg cccatcctgc
1741 cctgagccc atggcactct atgttctaag aagtgaataa ctacactcca gtgagacagc
1801 tetgcatact cattaggatg gctagtatca aaagaaagaa aatcaggctg gccaacgggg
1861 tgaaccctg tctctactaa aaatacaaaa cctgactcca attagccggg cgtggtggtg
1921 agtgcctgta atcacagcta cttgggagggc tgagatggga gaatcaacttg aaccggggag
1981 gcagagggttg cagtgagccg agattgtgcc cctgcactcc agcctgagcg acagtggagc
2041 tctgtctcag tccatgaaga tgtagaggag aaactggaac tctcgagcgt tgcctggggg
2101 gattgtaaaa tgggtgtgacc actgcagaag acagtatggc agctttcctc aaaacttcag
2161 acatagaatt aacacatgat cctgcaattc cacttatagg aattgaccca caagaaatga
2221 aagcagggac ttgaacccat atttgtacac caatatctcat agcagcttat tcacaagacc
2281 caaaggccag aagcaaccca eatgttcato aatgaatgaa tgaatggcta agcaaaatgt
2341 gatattgacc taacgaagta tccctcagcc tccctcagcc atgaagtact catacatgtt

2401 acaaacacgga cgaaccttga aaactttatg ctaagtgaaa taagccagac atcaacagat

```

图1 (续)

2461 aaatagttta tgattccacc tacatgaggt actgagagtg aacaaattta cagagacaga
 2521 aagcagaaca gtgattacca gggactgagg ggaggaggagc atgggaagtg acggttttaa
 2581 gggcacagggt tttatgttta ggtggttgaa aaagtcttgc agataaacag tagtgatagt
 2641 tgtaccgcaa tglacttaa tgccactaaa ttgacactta aatatgggtt aaatgggtcaa
 2701 tttgttatg tatattttat atcaatttaa aaaaaaacct gagcccaaaa aggtatttta
 2761 atcaccagg ctgattaaac caaggctaga accacctgcc tataattttt gttaaatgat
 2821 ttcattcaat atottttttt taataaacca tttttacttg ggtgttttata aaaaaaaaaa

图1 (续)

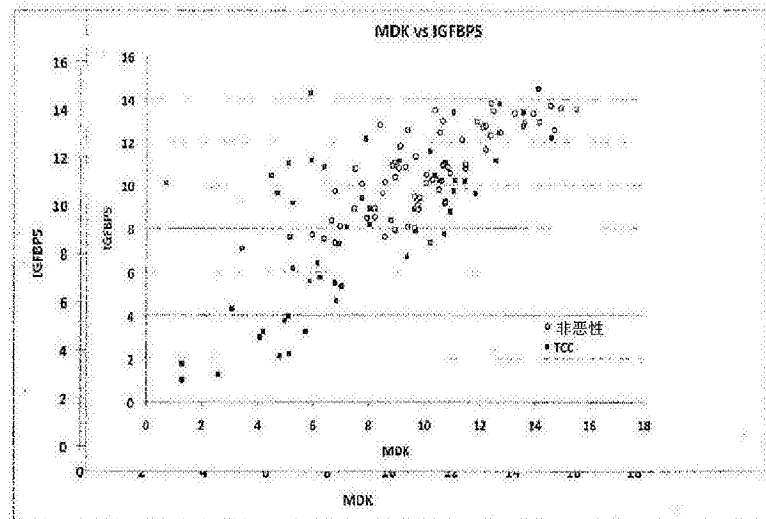


图2a

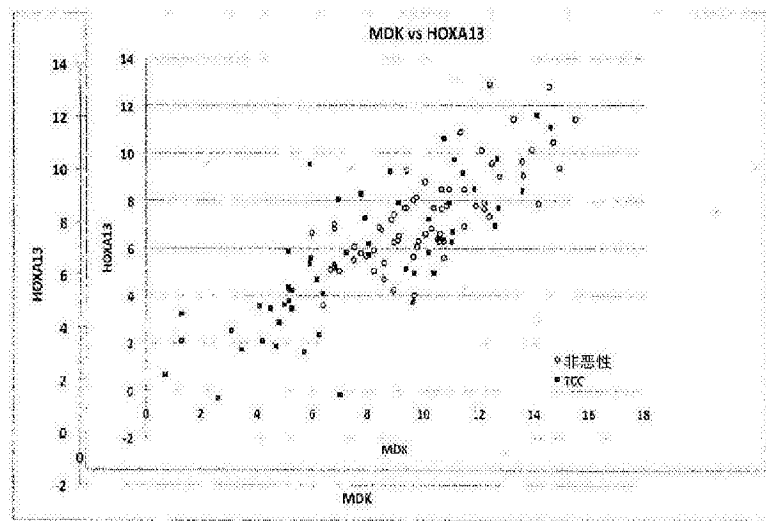


图2b

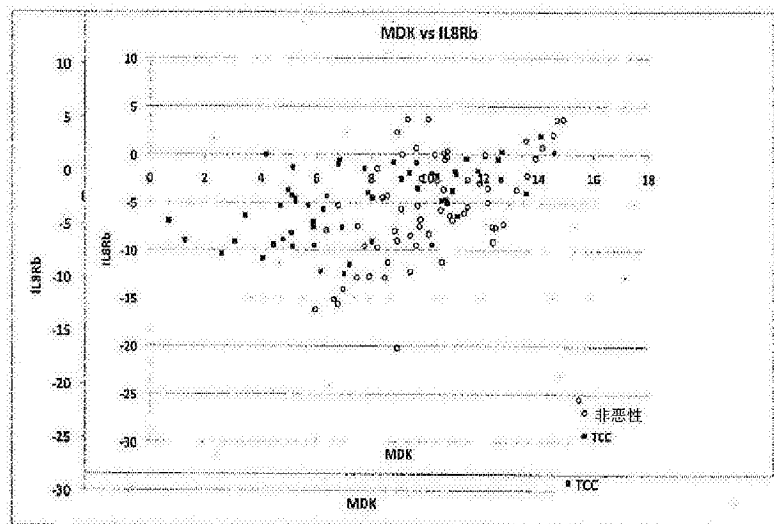


图2c

图2

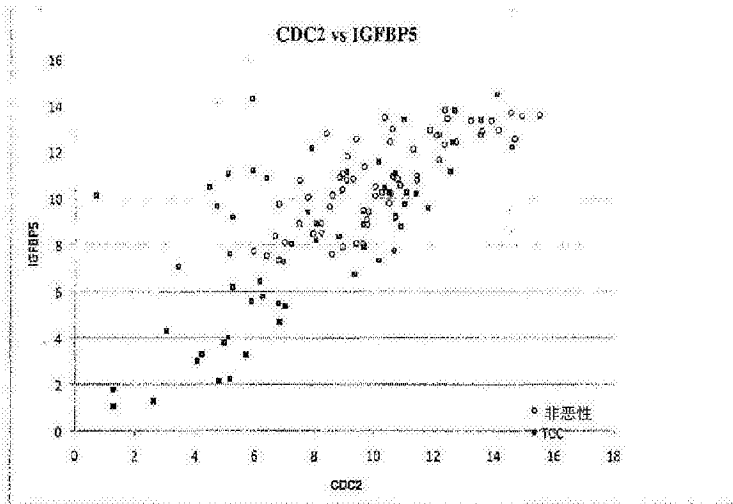


图2d

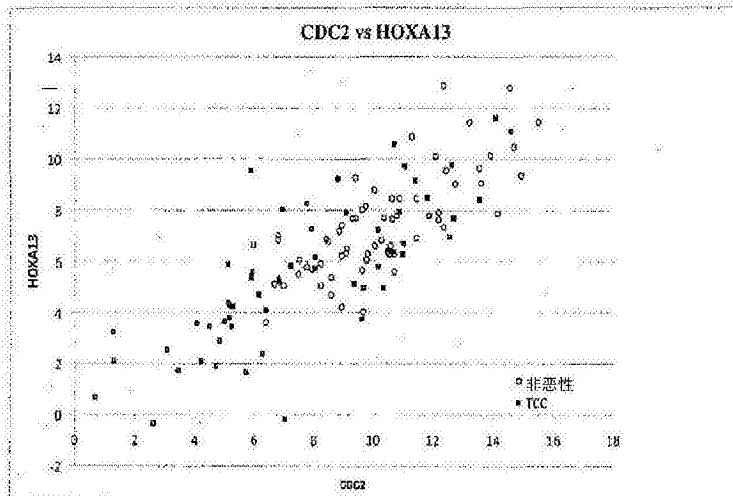


图2e

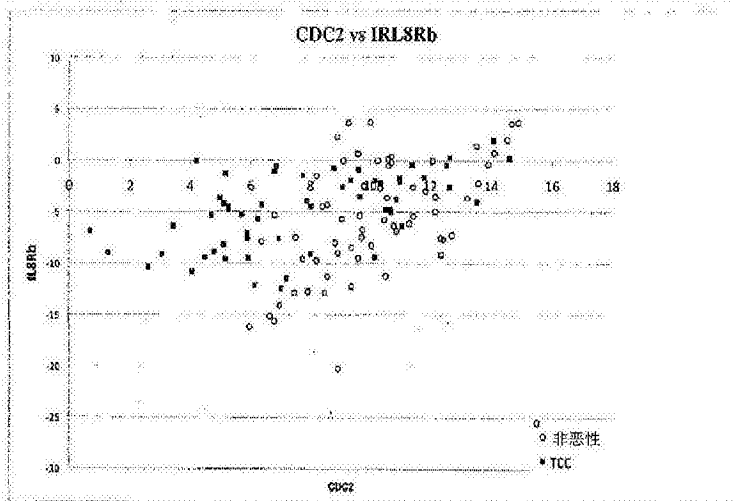


图2f

图2(续)

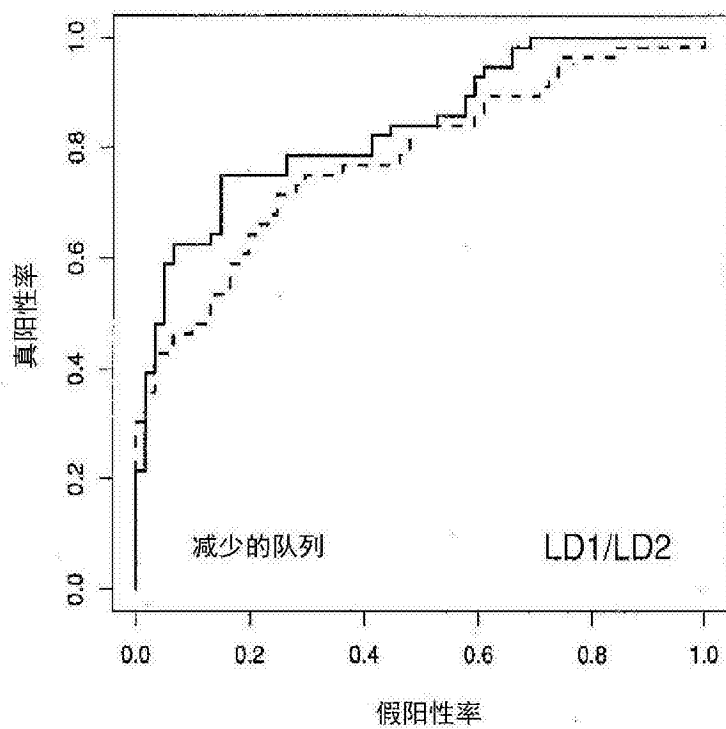


图3a

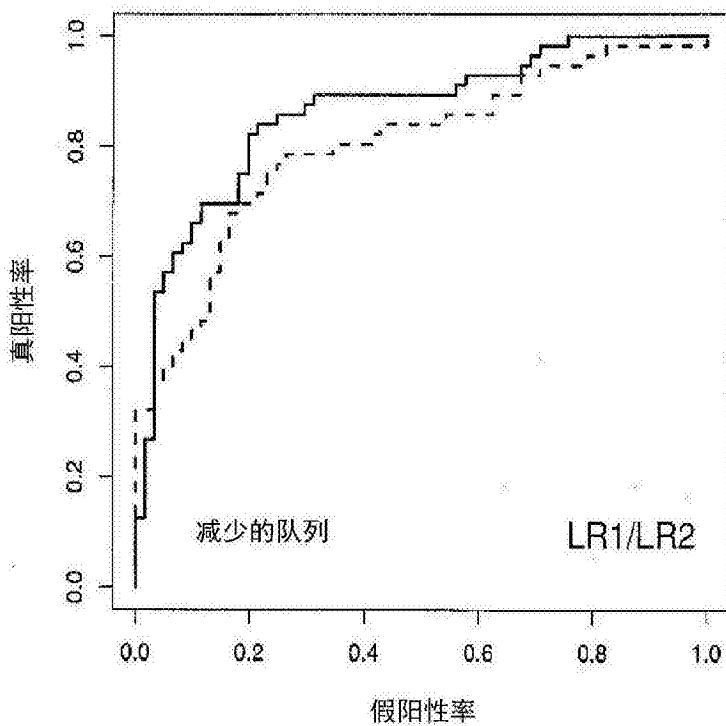


图3b

图3a-b

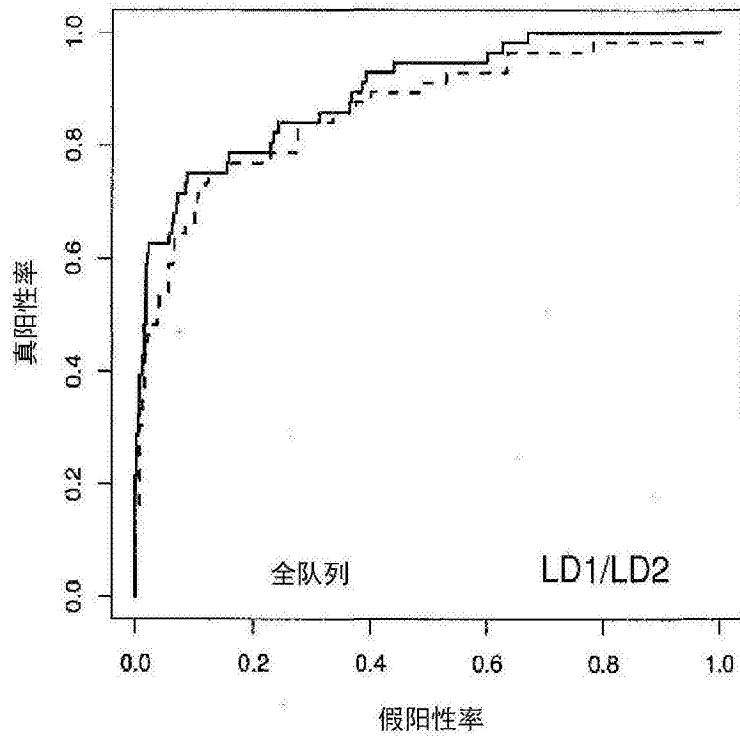


图4a

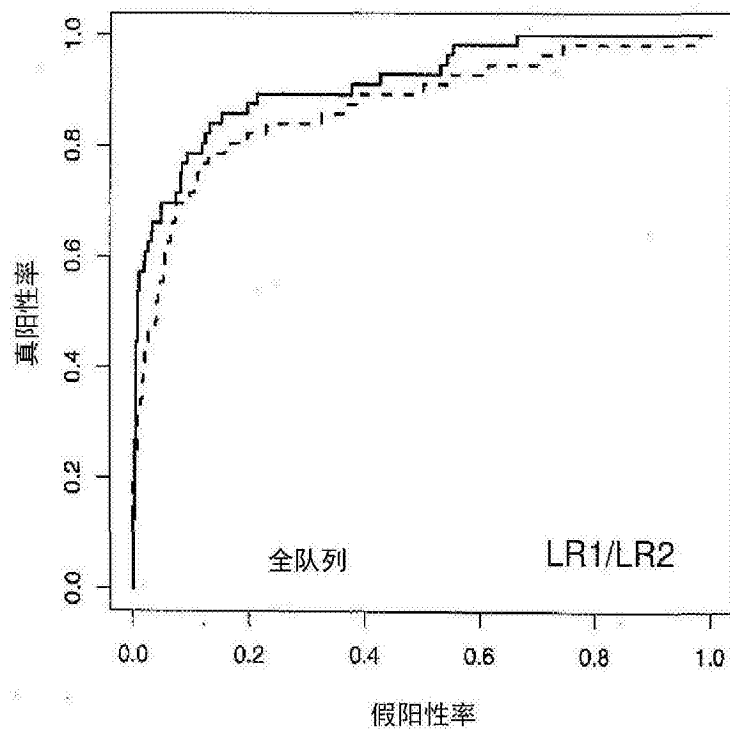


图4b

图4a-b

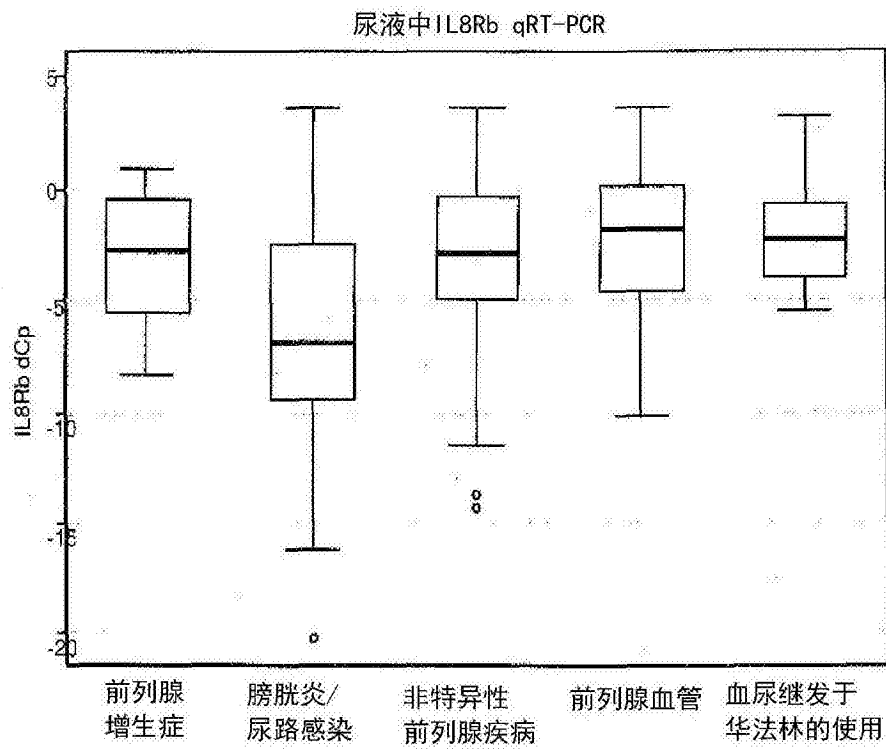


图5

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	获得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
ras同源 基因家族, 成员B (RHOB), mRNA	NP_004031.1	NM_004040.2	mwghuman3 OK#A:08114	-2.684592	2.13794E- 10	2.25758E- 09	-10.53678	38	8789.5	4176	23
假定蛋白 FLJ21511 (FLJ21511), mRNA	NP_079363.1	NM_025087.1	mwghuman3 OK#A:00410	-2.080908	2.54894E- 11	3.22908E- 08	-6.925773	68	14455	9272.5	0
配对框基因 8 (PAX8), 转录变异体 PAX8A, mRNA	NP_003457.1	NM_003466.3	mwghuman3 OK#B:3648	-1.961896	4.71674E- 11	4.07014E- 08	-6.806281	82	19419	12999.5	0
无义转录 同源A的 UPF3调节剂 (酵母) (UPF3A), 转录变异体 1, mRNA	NP_075387.1	NM_023011.2	mwghuman3 OK#A:06295	-3.84007	1.03703E- 13	3.34027E- 08	-7.811653	96.5	5663.5	1825	0
白三烯B412- 羟基脱氢酶 (LTB4DH), mRNA	NP_036344.1	NM_012212.2	mwghuman3 OK#B:8686	-5.034613	3.08873E- 10	1.41836E- 07	-6.436167	122.5	11340.5	3138	0

图6

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
RAB11家族 相互作用 蛋白2 (1类) (RAB11FIP2), mRNA	NP_055719.1	NM_014904.1	mwghuman3 OK#A:10479	-1.993629	4.57825E- 09	5.83951E- 08	-7.711624	156	19636.5	12951	0
TSC22域 家族, 成员1 (TSC22D1), 转录变异体 2, mRNA	NP_008013.1	NM_006022.2	mwghuman3 OK#A:00577	-2.133065	1.12033E- 07	1.8543E- 07	-7.698568	156.5	8415	6130	1
亚硫酸盐氧化酶 (SUOX), 核基因编码 线粒体蛋白, 转录变异体 1, mRNA	NP_000447.2	NM_000456.2	mwghuman3 OK#A:05512	-1.622803	4.52404E- 10	5.29156E- 09	-6.728892	223	11312.5	8789.5	0
三角形四肽 重链21A (TTC21A), mRNA	NP_665698.1	NM_145755.1	mwghuman3 OK#B:6976	-2.976809	2.11314E- 07	4.28806E- 07	-6.286432	225	8375.5	3870.5	1

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	获得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
RAS p21 蛋白催化剂 (GTPase 催化蛋白) 1 (RAS1), 转录变异体 2, mRNA	NP_072179.1	NM_022650.1	mwghuman3 OK#A:08292	-1.947454	3.74564E- 09	2.68048E -06	-5.99299	235	6895.5	5277.5	7
集团 genscan 预测		AC024384.3.1 07975.119435 .1	mwghuman3 OK#C:6641	-2.471322	4.47058E- 08	1.54585E -06	-6.085076	255.5	3327.5	1537	
集团 genscan 预测		AL031669.28. 1.94224.1	mwghuman3 OK#C:0930	-1.774123	6.77812E- 09	1.13272E -06	-5.56116	288.5	4367	3218	
双能蛋白, DNA复制 抑制剂 (GMNN), mRNA	NP_056979.1	NM_015895.3	mwghuman3 OK#A:03435	-2.046753	1.94846E- 07	4.07898E -06	-5.99032	311	11844	8413.75	2
脂肪酸结合 蛋白1, 肝脏 (FABP1), mRNA	NP_001434.1	NM_001443.1	mwghuman3 OK#A:09506	-1.985247	2.20148E- 06	6.83453E -07	-6.750782	312.5	25595	19500.75	0
v-jun肉瘤病毒 17致癌基因 同源物 (鸟的) (avian) (JUN), mRNA	NP_002219.1	NM_002228.3	mwghuman3 OK#A:04848	-3.222311	6.2171E- 07	4.59274E -06	-6.383849	318.5	12139	5321.5	11

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
穿膜域和 卷曲螺旋域4 (TMCO4), mRNA 集团 genscan 预测	NP_859070.2	NM_181719.2	mwghuman3 OK#B:3094	-1.799143	8.41415E- 09	3.21175E- 06	-5.702577	322.5	4584.5	3158	5
突触囊泡蛋白样 蛋白2 (SYTL2), 转录变异体 b,	NP_115755.2	AP002080.2.3 7398.74587.1	mwghuman3 OK#C:4035	-1.860121	5.75299E- 08	1.01121E- 07	-7.300326	332.5	4972.5	2643	
mRNA 细胞周期蛋白G1 (CCNG1), 转录变异体 1,	NP_004051.1	NM_032379.3	mwghuman3 OK#B:9411	-2.343357	4.07526E- 10	9.87494E- 09	-7.993821	342	19920	12053.5	0
mRNA F-box蛋白 34 (FBXO34), mRNA	NP_060413.2	NM_017943.2	mwghuman3 OK#B:5261	-1.83737	2.79204E- 08	4.64998E- 07	-6.214776	365	9269.5	4453	7
假定蛋白 xp_097916 loc150582	NP_060413.2	NM_017943.2	mwghuman3 OK#B:4885	-1.801846	1.15594E- 06	1.13272E- 06	-7.165506	365.5	17957.5	13192.75	5
含RNA 假尿苷酸 合成酶域 (RPUSD4), mRNA	NP_116184.1	NM_032795.1	mwghuman3 OK#B:6228	-2.291304	2.69775E- 08	3.84322E- 06	-5.296383	375	3591	1276	
			mwghuman3 OK#B:6888	-1.588438	2.0145E- 10	1.14227E- 08	-8.404434	407	15747.5	11247	5

图6(续)

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
10D重域 ³³ (WDR33), 转录变异体 ₂ , mRNA	NP_0010066 23.1	NM_0010066 22.1	mwghuman3 OK#A:08115	-1.765322	1.47232E- 07	1.45307E- 06	-6.190057	413	14876	13787.5	8
FLJ20859基因 (FLJ20859), 转录变异体 ₁ , mRNA	NP_0010251 62.1	NM_0010299 91.1	mwghuman3 OK#A:10470	-1.708985	8.133E-08	1.13272E- 06	-6.182853	420.5	3781	2693.75	
催化转录因素 ₃ (ATF3), 转录变异体 ₂ , mRNA	NP_004015.3	NM_004024.3	mwghuman3 OK#A:00568	-3.134291	8.40437E- 06	1.51698E- 07	-7.446028	421	13832.5	5301.25	3
细胞色素P450, 家族3, 子 家族A, 多肽5 (CYP3A5), mRNA	NP_000768.1	NM_000777.2	mwghuman3 OK#A:04559	-2.482162	7.0225E- 08	1.9772E- 06	-1.813425	451.5	20350.5	12190.5	2
集团 genscan 预测		AL161725.13, 1.181179.3	mwghuman3 OK#C:3443	-1.955495	1.19555E- 09	9.24336E- 06	-5.681512	455	2738.5	1799	
KIT配体 (KITLG), 转录变异体 _b , mRNA	NP_000890.1	NM_000899.3	mwghuman3 OK#A:07355	-4.28251	3.8867E- 06	7.23184E- 07	-5.884133	470	19497	10599.75	2

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
补体衰变 加速因子 (CD55, Crome 血型 系统)(DAF), mRNA	NP_000585.1	NM_000574.2	mwghuman3 OK#A.06760	-2.148033	1.44112E- 06	6.53282E -06	-5.668164	471.5	9806.5	4753.25	10
溶质载体 家族 25 (线粒体载体, 磷酸盐载体), 成员 23 (SLC25A23), mRNA	NP_077008.2	NM_024103.2	mwghuman3 OK#A.05530	-1.465915	7.79426E- 09	8.80903E -07	-5.741772	497	16285.5	12418.75	4
无翅型MMTV 整合位点 家族, 成员2B (WNT2B), 转录变异体 WNT- 2B1, mRNA	NP_004176.2	NM_004185.2	mwghuman3 OK#A.05696	-10.745275	2.92382E- 12	3.19726E -07	-6.175471	501	10716	2016.25	0
原钙黏蛋白 γ 子家族A, 12 (PCDHGA12), 转录变异体 1, mRNA	NP_003726.1	NM_003735.2	mwghuman3 OK#A.03393	-2.070447	2.89557E- 07	6.92439E -06	-4.982742	511	2537	1931.5	0

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
集团 genscan 预测		AC010828.3, 125573.1581 93.2	mwghuman3 OK#C:4737	-1.679465	2.34945E- 07	6.18322E- 07	-5.341693	511	26962	25287.75	
嗅觉感受器 家族1, 子 家族D, 成员5 (OR1D5), mRNA	NP_055381.1	NM_014566.1	mwghuman3 OK#A:08766	-1.640751	5.38884E- 08	4.32856E- 06	-5.134756	539.5	1644	1134	0
G蛋白耦 联受体126 (GPR126), 转录变异体 bt, mRNA	NP_940971.1	NM_198569.1	mwghuman3 OK#B:2989	-2.181735	3.9002E- 07	1.05021E- 05	-5.047901	549	15955.5	8040.75	0
类似于ba408e- 5.4新型蛋白 dmpk样cdc42 结合激酶β cdc42bpbloc 144850		XM_090553	mwghuman3 OK#B:9158	-4.265957	2.3863E- 08	5.29006E- 07	-6.259229	573	1784	632	
范素特异性 肽酶9, Y连 锁型(脂肪 小平面样, 果蝇属) (USP9Y), mRNA	NP_004645.2	NM_004654.3	mwghuman3 OK#A:10655	-1.584065	3.40815E- 07	8.23838E- 06	-5.267284	587	10035.5	7858.25	1

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
集团 genscan 预测		AC022554.2.1 1725.14597.1	mwghuman3 OK#C:7523	-2.800527	1.74508E- 07	6.41199E -07	-5.902891	597	3330	1296	
溶质载体家族35 (OMP硅铝质 酸输运者), 成员A1 (SLC35A1), mRNA	NP_006407. 1	NM_006416.2	mwghuman3 OK#B:6203	-1.879113	1.20877E- 06	5.48118E -06	-5.297694	606	20639	15436.75	1
离核苷酸位置 的细胞系sc30 t受体α链v-j 交界区tcr v 29 1 j 11 编码序列 报告跨距		U14083	mwghuman3 OK#B:9790	-2.442854	9.94216E- 08	1.14277E -08	-7.365379	617	6673	2661.5	
染色体2 可读框33 (C2orf33), mRNA	NP_064579. 3	NM_020194.4	mwghuman3 OK#A:06988	-1.615311	3.24982E- 07	9.24336E -06	-5.392632	624	11555.5	8868.25	4
集团 genscan 预测		AC073108.7.1 14033.192530 1	mwghuman3 OK#C:4991	-2.137812	1.453E-05	1.06394E -06	-6.345675	650.5	9963.5	5676.75	

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
UDP-N-乙酰基- α -D-半乳糖胺:多肽 N-乙酰氨基半乳糖转移酶7 (GALNT7), mRNA	NP_059119.1	NM_017423.1	mwghuman3 OK#A:01420	-1.401046	7.53042E-08	1.13272E-06	-6.143981	656.5	10573	8997.75	2
染色体11可读框1 (C11orf1), mRNA	NP_073598.1	NM_022761.1	mwghuman3 OK#A:01834	-1.81726	1.38032E-06	1.22937E-05	-5.104229	673.5	12025.5	9585.25	0
核内异质核糖核蛋白H2 (H1) (HNRPH2), 转录变体1, mRNA	NP_062543.1	NM_019597.3	mwghuman3 OK#A:08547	-2.03901	2.40055E-06	6.16248E-06	-5.465662	684	777	472	9
沃纳综合征 (WRN), mRNA	NP_000544.1	NM_000553.2	mwghuman3 OK#A:03163	-1.605462	7.49706E-08	6.16248E-06	-4.681771	696	4585	4351.25	3
集团 gensecan 预测		AC040160.3.6 1245.138821.5	mwghuman3 OK#C:7317	-4.953847	7.54477E-12	1.43666E-08	-6.602968	706.5	4523.5	744.5	

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
内收蛋白3 (γ) (ADD3) 转录变异性 mRNA	NP_058432.1	NM_016824.3	mwghuman3 OK#B:5211	-2.092475	1.34331E- 05	1.16156E- 05	-5.883251	716	7668.5	4808.75	8
羟类固醇 脱氢酶2 (HSDL2), mRNA	NP_115679.2	NM_032303.2	mwghuman3 OK#B:3888	-2.1317	9.39999E- 07	5.81228E- 06	-5.58267	732.5	7276.5	5055	2
DNA-破坏- 可诱导转录4 (DDIT4), mRNA	NP_061931.1	NM_019058.2	mwghuman3 OK#B:2115	-2.660775	6.58834E- 05	5.16818E- 06	-5.969252	750	10960.5	5062.75	12
RNA结合蛋白, 自身抗原 (hnRNP-关联 于致死黄同 系物(老鼠)) (BALY), 转录变异性 mRNA	NP_057951.1	NM_016732.1	mwghuman3 OK#A:06376	-1.901576	1.7076E- 07	4.32856E- 06	-4.392163	750.5	977	386.5	26

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWIG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
酰基辅酶A 合成酶中 链家族成员3 (ACSM3), 转录变异体 1, mRNA	NP_005613.2	NM_005622.3	mwghuman3 OK#B:9076	-1.967123	1.7751E- 05	7.33836E -06	-5.050721	755.5	18906.5	12252.25	3
BCL2相关的 永生基因 (BAG1), mRNA	NP_004314.3	NM_004323.3	mwghuman3 OK#A:09581	-2.525417	2.13163E- 08	5.29006E -07	-5.934904	776	9031	4503.5	1
神经前体细胞 表达的, 显影 减量调节的 4	NP_006145.1	NM_006154.1	mwghuman3 OK#B:7862	-1.900818	3.7763E- 06	8.23334E -06	-5.249282	781	15497	12075	2
(NEDD4), 转录变异体 1, mRNA	NP_116220.1	NM_032831.1	mwghuman3 OK#B:2324	-1.780735	4.63759E- 06	0.000016 624	-5.43532	783	11586.5	6704.75	16
染色体7 可读框19 (C7orf19), mRNA	NP_653176.2	NM_144575.2	mwghuman3 OK#B:6754	-1.542431	4.41704E- 06	5.81228E -06	-4.843249	808.5	23318	19953.5	0
钙激活蛋白酶13 (CAPN13), mRNA	NP_002136.1	NM_002145.2	mwghuman3 OK#A:01639	-2.702649	2.94498E- 11	5.08294E -08	-6.774876	814.5	11537.5	6408	3

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
包含28A的 卷曲螺旋域 (CCDC28A), mRNA	NP_056254.1	NM_015439.2	mwghuman3 OK#B:3128	-1.950806	1.97345E- 06	5.48118E -06	-5.330597	826	19022	14223.75	4
肌原纤维形 成调节因子1 (MR-1), 转录变异体 1, mRNA	NP_056303.2	NM_015488.3	mwghuman3 OK#B:3179	-1.526799	1.22924E- 05	6.92439E -06	-5.874478	832	13007	10755.75	6
鸟嘌呤核苷酸 结合蛋白 (G蛋白) β 多肽2-样1 (GNB2L1), mRNA	NP_006089.1	NM_006098.4	mwghuman3 OK#C:0841	-1.789085	2.16356E- 06	4.35419E -05	-4.900058	844.5	1710	1178	106
包含核膜1的 血影蛋白重复 (SYNE1), 转录变异体 β, mRNA	NP_056108.1	NM_015293.1	mwghuman3 OK#B:9104	-1.414534	7.52733E- 07	3.02433E -06	-5.983107	851	20159.5	17233.75	64
集团 genscan 预测		AL021155.1.1 .107603.5	mwghuman3 OK#C:7334	-1.766409	2.12121E- 06	2.01355E -05	-4.468876	853	1264	847.5	

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
骨成型蛋白 受体, 类型 1A (BMPRI1A) , mRNA	NP_004320.2	NM_004329.2	mwghuman3 OK#A:02484	-1.558632	2.22237E- 07	1.22937E- -05	-4.618211	874.5	5116	5037.5	4
包含B1的IQ 基序 (IQCB1) , 转录变异体1 , mRNA	NP_0010188 64.1	NM_0010235 70.1	mwghuman3 OK#B:4020	-1.646698	2.79495E- 05	1.62908E- -05	-5.169855	882.5	10922	8773.5	2
集团 genscan 预测		AC069502.9.1 .18390.1	mwghuman3 OK#C:3663	-1.464573	6.94397E- 07	2.53657E- -05	-4.951029	901	21078	19035	
类似于死亡 相关的蛋白 (LOC92196) mRNA	NP_0010179 20.1	NM_0010179 20.1	mwghuman3 OK#B:6718	-2.553467	5.76345E- 07	7.76152E- -07	-5.221107	907	23255	15868	0
集团 genscan 预测		AC012203.4.1 05191.123571 .1	mwghuman3 OK#C:4578	-1.438718	3.21936E- 05	4.85862E- -07	-5.989899	907	27071.5	25867.5	
集团genscan 预测mwg匹配 这些参考号码 NM_000986		AL158153.8.4 5139.113162. 3	mwghuman3 OK#C:2173	-2.089238	4.80586E- 07	4.32856E- -06	-5.395407	909.5	1290	1010.25	

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
生长抑制和 DNA-破坏- 可诱导型, α (GADD45A) , mRNA	NP_001915.1	NM_001924.2	mwghuman3 OK#A.04346	-1.68216	7.53906E- 06	2.67916E- 05	-4.924921	916.5	11634.5	8596.75	3
应力70蛋白 伴侣, 微粒体 相关的60kDa (STCH) , mRNA	NP_008879.3	NM_006948.4	mwghuman3 OK#A.05064	-1.761334	4.58163E- 05	1.72268E- 05	-5.784774	916.5	10794	6231.25	5
集团 genscan 预测		Z84474.1.1.1 07526.2	mwghuman3 OK#C.0311	-1.884499	1.62444E- 08	2.23413E- 06	-5.434325	933.5	1974	1488.5	
具有序列相 似性的家族 44, 成员B (FAM44B) , mRNA	NP_612378.1	NM_138369.1	mwghuman3 OK#B.4192	-1.357114	7.91633E- 07	1.64431E- 06	-6.047548	942.5	15308.5	13004.75	1
集团 genscan 预测		AC078789.15. 1.55074.2	mwghuman3 OK#C.6283	-1.786484	1.80836E- 06	3.70831E- 05	-4.861731	948	25533.5	22990.75	
柠檬酸裂合 酶β 柱 (CLYBL) , 转录变异体1 , mRNA	NP_612124.3	NM_138280.3	mwghuman3 OK#B.8855	-1.444822	1.17336E- 06	5.16818E- 06	-5.775319	953	21231	19330.25	1

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
集团 genscan 预测		AP001015.3.6 6054.115655. 1	mwghuman3 OK#C:3998	-1.709242	1.48259E- 07	7.36585E -05	-4.504002	972.5	3374.5	2739.5	
集团 genscan 预测		AL159169.14. 1.86155.2	mwghuman3 OK#C:4465	-1.474035	8.96434E- 07	3.25437E -06	-5.020854	998.5	28575	28030	
染色体6 可读框130 (C6orf130), mRNA	NP_659500.1	NM_145063.2	mwghuman3 OK#B:4652	-1.674296	1.77778E- 05	3.32944E -05	-4.846371	1011. 5	14804	10754.75	3
转甲基酶样1 (METTL1), 转录变异体1 , mRNA	NP_005362.1	NM_005371.3	mwghuman3 OK#A:00607	-1.464425	1.73108E- 07	5.48118E -06	-5.447819	1026	943	645.5	2
核小体组装 蛋白1-样4 (NAP1L4) , mRNA	NP_005960.1	NM_005969.3	mwghuman3 OK#A:00706	-1.902267	1.44614E- 05	6.99321E -05	-4.962357	1042. 5	10336.5	8978.5	12
肿瘤蛋白D52 (TPD52), 转录变异体3 , mRNA	NP_006070.1	NM_005079.2	mwghuman3 OK#A:05683	-1.55806	1.79544E- 06	3.15409E -05	-4.599069	1070	19033	14646.75	4
包含4的EH域 (EHD4), mRNA	NP_644670.1	NM_139265.2	mwghuman3 OK#B:5911	-1.64935	1.97896E- 05	3.70831E -05	-4.903865	1077	12134	9187.5	4

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
成纤维细胞 生长因子4 (FGFR4), 转录变异体2 , mRNA	NP_075252.2	NM_022963.2	mwghuman3 OK#B:8174	-2.638426	2.02402E- 06	3.21175E- 06	-6.251766	1087. 5	17096	9762.5	0
KIAA0674 (KIAA0674), mRNA	XP_376903.2	XM_376903.2	mwghuman3 OK#B:0240	-1.362276	3.68739E- 06	3.62055E- 06	-5.432041	1092. 5	10744	9097.25	4
染色体16 bac 克隆cit987 sk-a-363e6		U91321	mwghuman3 OK#B:5375	-1.568388	5.91844E- 05	6.53282E- 06	-6.349455	1114. 5	10576.5	6083.25	
集团 genscan 预测		AC073610.11. 63494.102478 3	mwghuman3 OK#C:1337	-1.828108	1.68631E- 06	1.30086E- 05	-5.253574	1114. 5	1434.5	737	
雄性激素受体 (AR), 转录 变异体1, mRNA	NP_000035.2	NM_000044.2	mwghuman3 OK#A:01003	-1.575215	1.18973E- 07	1.54035E- 05	-4.291628	1120. 5	18067.5	15252.5	0
甲状腺转录 因子1 (TTF1) , mRNA	NP_003308.1	NM_003317.3	mwghuman3 OK#A:07197	-1.466987	8.57665E- 07	1.54585E- 06	-6.023277	1124. 5	2139	1026.25	0

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
包含核膜1的 血影蛋白重复 (SYNE1), 最长的转录 变体, mRNA	NP_892006.1	NM_182961.1	mwghuman3 OK#C:2340	-1.440164	6.37872E- 05	2.15085E- 05	-5.805222	1124. 5	26092.5	21681.5	64
集团 genscan 预测		AC019084.7.9 3615.134992. 2	mwghuman3 OK#C:5482	-1.55364	1.48892E- 06	7.77593E- 06	-5.043927	1126. 5	8785.5	8789.5	
集团 genscan 预测		AC026235.12. 9210.157122. 4	mwghuman3 OK#C:3969	-1.693854	8.74991E- 08	0.000122 803	-4.393663	1136. 5	804	493.5	
集团 genscan 预测		AP002840.1.9 9623.121675. 1	mwghuman3 OK#C:5950	-1.784079	2.96992E- 05	7.75722E- 05	-4.69759	1156	16693.5	14033.25	
CDC样激酶1 (CLK1), 转录变体1 , mRNA	NP_004062.2	NM_004071.2	mwghuman3 OK#A:07840	2.034707	6.15281E- 05	6.99321E- 05	-5.146308	1157. 5	12199	7934.75	4
类脂 (LPIN1) , mRNA	NP_663731.1	NM_145693.1	mwghuman3 OK#B:8172	-1.851742	1.40757E- 07	2.98753E- 05	-4.301729	1159. 5	3757	2609	2
核纤层蛋白 A/C (LMNA), 转录变体2 , mRNA	NP_005563.1	NM_005572.2	mwghuman3 OK#A:08943	-1.724223	9.63727E- 05	5.67358E- 05	-5.305029	1162. 5	4447	3295	10
集团 genscan 预测		AC079456.15. 119608.15149 0.1	mwghuman3 OK#C:3309	-1.39662	6.39081E- 06	4.32856E- 06	-6.439308	1169. 5	26755.5	26493.5	

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
类似于riken cdna 1500002b03 克隆mgc:12928 图像: 4298414		BC007768	mwghuman3 OK#B:4242	-1.518302	5.21322E- 06	3.51403E- 05	-4.463332	1171	21159	20375	
集团 genscan 预测 透明质酸 葡萄糖胺酶3 (HYAL3) mRNA		AC027086.2.2 3033.31927.1	mwghuman3 OK#C:6982	-1.659003	1.16113E- 11	3.82785E- 07	-6.295271	1172. 5	2757	2062.75	
集团 genscan 预测 VprBP蛋白 (VprBP) mRNA	NP_003540.2	NM_003549.2	mwghuman3 OK#A:09529	-2.01277	6.70905E- 08	1.16156E- 05	-4.684969	1175	4326	3464.5	4
集团 genscan 预测 VprBP蛋白 (VprBP) mRNA		AC083822.10. 126464.12912 2.1	mwghuman3 OK#C:9101	-1.479602	6.01319E- 06	2.53657E- 05	-4.526959	1177	22091.5	20578.5	
核RNA输出因子 1 (NXF1), mRNA	NP_055518.1	NM_014703.1	mwghuman3 OK#B:0332	-1.720326	3.28082E- 05	2.40122E- 05	-5.246863	1178	17312.5	14031	0
核RNA输出因子 1 (NXF1), mRNA	NP_006353.2	NM_006362.3	mwghuman3 OK#B:1047	-1.485619	7.48559E- 05	1.62908E- 05	-5.176445	1195. 5	15210	14049.75	20
空气孔同源 物2 (果蝇属) (EMX2), mRNA	NP_004089.1	NM_004098.2	mwghuman3 OK#B:9279	-1.745086	9.4996E- 09	5.6417E- 07	-5.709572	1199. 5	19380	10749.25	0
同源框D8 (HOXD8) mRNA	NP_062458.1	NM_019558.2	mwghuman3 OK#B:3465	-1.41555	6.70306E- 08	2.1019E- 06	-5.674131	1206	9672.5	7512.25	0

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
hq2591的 pro2591 预测蛋白		AF119886	mwghuman3 OK#B:1001	-1.487408	7.83971E- 05	1.30297E- -05	-4.929888	1212	27543.5	26699	0
集团 genscan 预测		AC016432.3.1 6946.19371.1	mwghuman3 OK#C:6058	-3.528975	4.33482E- 13	5.08294E- -08	-6.731611	1216. 5	3962.5	1303	
集团 genscan 预测		AF286885.1.7 6372.115954. 3	mwghuman3 OK#C:7800	-1.83051	1.1443E- 05	5.11795E- -05	-4.359199	1218	21163.5	21224.5	
锌指蛋白25 (KOX 19) (ZNF25), mRNA	NP_659448.1	NM_145011.2	mwghuman3 OK#C:1648	-1.473477	6.43814E- 09	4.35863E- -07	-5.84264	1253. 5	21026.5	19171	1
锌指蛋白626 (ZNF626), mRNA	NP_660340.1	NM_145297.2	mwghuman3 OK#B:4148	-1.773183	2.31455E- 06	6.63845E- -05	-4.294361	1266	14806.5	10341	1
生长抑制及 DNA-破坏- 可诱导, β (GADD45B) , mRNA	NP_056490.1	NM_015675.1	mwghuman3 OK#A:00537	-1.950707	0.002000 32	2.03519E- -05	-5.558029	1270. 5	13309	8615.5	5
星云状小体 (NEBL), 转录变异体1 , mRNA	NP_006384.1	NM_006393.1	mwghuman3 OK#A:10459	-1.513391	1.95755E- 05	3.15409E- -05	-4.789714	1276	22936	20052.5	1

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
假定蛋白 pro2832 pro2832		NM_018541	mwghuman3 OK#A:00371	-1.457038	3.06355E- 05	6.99321E- 05	-5.164495	1289. 5	16692.5	14266.5	
环指蛋白44 (RNF44), mRNA	NP_055716.1	NM_014901.4	mwghuman3 OK#A:10680	-1.353366	7.76079E- 07	2.37433E- 06	-5.842265	1292	6288.5	4664.5	1
REX1, RNA 外切核酶1 同源物 (S. 啤酒) (REX01), mRNA	NP_065746.2	NM_020695.2	mwghuman3 OK#B:8218	-1.36867	1.73225E- 05	2.38069E- 05	-5.111364	1305. 5	4933	3309.5	6
噬噬受体P2Y, G蛋白耦合, 4 (P2RY4), mRNA	NP_002556.1	NM_002565.3	mwghuman3 OK#A:00433	-1.732075	5.0193E- 08	2.68048E- 06	-5.510614	1309	18307	12139.5	0
集团 genscan 预测		AL023804.2.1 96460.1	mwghuman3 OK#C:5405	-1.533348	1.03637E- 06	2.806E- 05	-4.693772	1313	28610	27980.5	
集团 genscan 预测		AL138823.3.4 9422.51911.1	mwghuman3 OK#C:8800	-1.894452	4.41409E- 05	0.000129 14	-4.650026	1320	11214.5	7548	
类似于 假定蛋白 (LOC440804), mRNA	XP_036936.3	XM_036936.3	mwghuman3 OK#B:4832	-4.049648	5.38047E- 11	1.51698E- 07	-6.579174	1332	3617	996.5	0

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
染色体X 可读框41 (CXorf41), mRNA	NP_775765.1	NM_173494.1	mwghuman3 OK#C:1526	-1.448331	1.77872E- 05	2.53657E- 05	-5.140193	1334	27414	24062.75	0
染色体20 可读框152 (C20orf152), mRNA	NP_543024.1	NM_080834.1	mwghuman3 OK#B:3251	-1.948612	6.84261E- 06	1.82138E- 05	-4.987586	1349	10775	8652.5	0
集团 genscan 预测		AL390766.11, 35686.142382 .1	mwghuman3 OK#C:4546	-1.516487	5.60386E- 05	4.81914E- 05	-4.642986	1352	28469.5	27078.25	
糖原磷酸化酶; 肝脏(她的病, 糖原储存疾病 类型VI) (PYGL) mRNA	NP_002854.3	NM_002863.3	mwghuman3 OK#A:04342	-2.190281	5.14475E- 06	0.000269 936	-4.316788	1388. 5	10289.5	7222	21
集团 genscan 预测		AC013764.3.1 17501.157349 .1	mwghuman3 OK#C:3627	-1.671949	8.77886E- 06	7.36585E- 05	-4.28185	1395. 5	1205	826.5	
卵泡抑制素-样4 (FSTL4), mRNA	NP_055897.1	NM_015082.1	mwghuman3 OK#B:0510	-1.544042	4.79088E- 05	4.35419E- 05	-4.607654	1405. 5	17051	12319	0

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
染色体14 可读框168 (C14orf168), mRNA	NP_113615.1	NM_031427.1	mwghuman3 OK#B:4011	-1.512002	1.56095E- 05	5.97939E- 05	-4.695177	1414	15738	14705.75	1
醛脱氢酶7家族, 成员A1 (ALDH 7A1), mRNA	NP_001173.1	NM_001182.2	mwghuman3 OK#A:08361	-1.837545	4.80787E- 07	0.000105 506	-3.903403	1433	20103.5	14345.5	2
集团 genscan 预测		AC010904.9.3 83.204857.4	mwghuman3 OK#C:7433	-1.40403	4.74755E- 05	3.07694E- 05	-4.852328	1444	27423.5	25979.25	
激肽释放素8 (丝氨酸蛋 白酶/卵巢) (KLK8), 转录变体2, mRNA	NP_653088.1	NM_144505.1	mwghuman3 OK#B:0129	-2.091505	3.56009E- 06	2.61655E- 06	-5.991679	1449. 5	8391	6488	0
亮氨酸链 在癌症1中 下调-样 (LDOC1L), mRNA	NP_115663.2	NM_032287.2	mwghuman3 OK#B:8710	-1.501021	2.6566E- 06	1.22937E- 05	-4.280727	1454. 5	18509.5	14157	0
两孔区段通道1 (TPCN1), mRNA	NP_060371.2	NM_017901.3	mwghuman3 OK#B:7759	-1.497443	4.32588E- 06	3.32944E- 05	-4.567011	1460. 5	17545	11635.75	8

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	获得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
假定蛋白 MGC11242 (MGC11242) , mRNA	NP_077296.1	NM_024320.2	mwghuman3 OK#A:05130	-1.527839	2.70218E- 05	1.72268E- 05	-4.004461	1463. 5	17565	13134.5	0
磷酸弗林蛋白酶 酸性氨基酸族 分泌蛋白1 (PA CS1), mRNA	NP_060496.2	NM_018026.2	mwghuman3 OK#B:7703	-1.494259	5.23361E- 05	7.36585E- 05	-4.899266	1482. 5	9352	7757.75	19
假定蛋白 xp_ 036406 loc91138		XM_036406	mwghuman3 OK#B:8080	-1.533764	1.02161E- 05	4.59223E- 05	-4.93538	1486. 5	15372.5	14652	
KIAA1274 (KIAA1274), mRNA	NP_055246.1	NM_014431.1	mwghuman3 OK#B:8047	-1.41942	7.87044E- 05	2.27276E- 05	-5.659345	1488. 5	8059	5065	0
集团 genscan 预测		AC090651.1.9 901.36661.2	mwghuman3 OK#C:6027	-1.525691	2.39913E- 05	7.36585E- 05	-4.620984	1489. 5	21092	20253	
假定蛋白 FLJ13111 (FLJ13111) , mRNA	NP_079358.1	NM_025082.1	mwghuman3 OK#A:02690	-1.496988	3.69096E- 05	7.08197E- 05	-4.489468	1494	4861	7579.25	10
血型糖蛋白C (Gerbich血型) (GYPC), 转录变异体1, mRNA	NP_002092.1	NM_002101.3	mwghuman3 OK#A:01278	-1.628829	9.49623E- 06	0.000135 784	-4.110107	1502	8042	6771	2

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MMWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
假定LOC401510 (LOC401510) , mRNA	XP_376843.2	XM_376843.2	mwghuman3 OK#B:6807	-1.402922	3.03035E- 06	0.000105 506	-4.471835	1504	13882.5	11233.25	0
集团 genscan 预测		AC068282.4.1 68870.201224 .1	mwghuman3 OK#C:1421	-1.273265	1.4683E- 06	4.32856E -06	-5.673395	1505. 5	4650	3232.25	
磷脂酶A2, 组IVA (胞 质的, 钙 依赖的) (PLA2G4A) , mRNA	NP_077734.1	NIM_024420.1	mwghuman3 OK#C:2726	-1.656839	1.32347E- 05	2.82936E -05	-4.522341	1508. 5	17687.5	14140.5	5
丝束蛋白1 (同 源异构体) (P LS1), mRNA	NP_002661.1	NIM_002670.1	mwghuman3 OK#A:06551	-1.626206	5.67628E- 05	2.98753E -05	-4.952263	1509. 5	20943	14040.5	1
蛋白激酶 锚定蛋白-样1 (NBEAL1), mRNA	NP_945183.1	NIM_198945.1	mwghuman3 OK#C:8646	-1.528665	7.51176E- 06	2.82936E -05	-5.005478	1511. 5	23554.5	20060	3
乙醇脱氢酶4 (11类), 脾 多肽 (ADH4) , mRNA	NP_000661.2	NIM_000670.3	mwghuman3 OK#A:04002	-1.485933	6.40488E- 05	7.36585E -05	-4.608649	1520. 5	22288	20240.75	0

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	获得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
染色体1p34.1-36.11上的克隆rp5-94718包含ests gss和1stss。部分新型cub sush1 scr 重复域包含p		AL355178	mwghuman3 OK#B:9614	-1.466888	8.43097E-06	3.15409E-05	-4.397922	1529	4947.5	4650.25	
细胞色素b5还原酶3 (CYB5R3), 转录变异体M, mRAN	NP_000389.1	NM_000398.4	mwghuman3 OK#A:01423	-1.497796	0.000121606	9.05266E-05	-5.28459	1531.5	9574.5	6353.5	9
Hq1477的克隆flb5634 pro 1477预测的蛋白		AF130059	mwghuman3 OK#B:1074	-1.699661	1.17568E-05	3.32944E-05	-4.802532	1535	5685	3946.75	10
锌指, DHHC型包含2 (ZDHHC2), mRAN	NP_057437.1	NM_016353.2	mwghuman3 OK#A:03494	-1.588633	8.03858E-06	0.000105506	-4.741236	1536	13684.5	10087.5	0
视锥杆同源框(CRX), mRAN	NP_000545.1	NM_000554.2	mwghuman3 OK#A:09219	-3.022299	2.51326E-05	9.24336E-06	-5.849282	1548	16588.5	7500.5	0

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
染色体14可读 框154 (C14orf154), 转录 变体1, mRNA	NP_115609.2	NM_032233.2	mwghuman3 OK#B:2665	-1.704609	8.00302E- 05	0.000166 293	-4.427372	1561. 5	16629	15080.75	5
假定蛋白 xp_098665 loc155085		XM_098665	mwghuman3 OK#B:6056	-1.672814	4.27782E- 05	0.000142 749	-4.312941	1566. 5	8795	6847.25	
克隆 hc0117 pro0117		AF090895	mwghuman3 OK#B:0914	-1.888405	1.18377E- 05	9.05266E -05	-4.735763	1570. 5	20158	14933.5	4
前列腺癌和 乳腺癌过度 表达1 (PB0 V1), mRNA	NP_067648.1	NM_021635.1	mwghuman3 OK#A:05114	-1.610345	5.40979E- 06	8.1682E- 05	-4.437567	1571. 5	22494.5	19845.75	0
酯酶D/甲酰谷 胱甘肽水解酶 (ESD), mRNA	NP_001975.1	NM_001984.1	mwghuman3 OK#C:2841	-1.535083	6.75303E- 06	4.84259E -05	-4.777065	1574	8049	4838.75	6
ATG4自我吞噬 相关的4同源 物B (S. 啤酒) (ATG4B), 转录变体1 , mRNA	NP_037457.3	NM_013325.4	mwghuman3 OK#B:2748	-1.343856	8.59727E- 06	1.6224E- 05	-4.71932	1575	15166	16712.25	13

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWIG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
NADPH氧化酶1 (NOX1), 转录 异变体NOH-1Lv , mRNA	NP_039249.1	NM_013955.1	mwghuman3 OK#B:1370	-1.903809	5.04367E- 06	2.65189E- 05	-3.799677	1585. 5	18807.5	17371.5	0
集团 genscan 预测		AL445675.9.1 .171985.3	mwghuman3 OK#C:8741	-1.774793	2.60881E- 05	0.000182 916	-4.133137	1598	22483.5	18071.5	
推定核蛋白 ORF1-FL49 (ORF1-FL49) , mRNA	NP_115788.1	NM_032412.2	mwghuman3 OK#B:7958	-1.499659	6.10749E- 05	3.32944E- 05	-4.457256	1610. 5	8598	4986	8
THUMP域包含1 (THUMP1), mRNA	NP_060206.2	NM_017736.3	mwghuman3 OK#B:2088	-1.313257	4.11592E- 07	6.53282E- 06	-5.325518	1651. 5	14877	15313.5	0
谷氨酰胺酰基 肽环化转移酶 (谷氨酰胺酰 基环化酶) (QPCT), mRNA	NP_036545.1	NM_012413.3	mwghuman3 OK#A:09422	-1.769571	4.69509E- 07	0.000122 803	-3.882851	1651. 5	11960	8357.75	2
PHD指蛋白23 (PHF23), mRNA	NP_077273.1	NM_024297.1	mwghuman3 OK#A:02678	-1.398461	1.43961E- 05	0.000109 884	-4.356774	1657. 5	10656.5	9355.75	0

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
鼻咽癌相关 的基因蛋白8 (NAG8), mRNA	NP_055226.1	NM_014411.2	mwghuman3 OK#B:1482	-1.388055	2.38325E- 05	5.10584E- 05	-4.601658	1660	15731.5	12161	0
基因 genscan 预测		AC013487.5.1 02379.117540 1	mwghuman3 OK#C:3348	-1.465932	0.0001448 57	2.37433E- 06	-6.225756	1662	14711	9423	
基因 genscan 预测		AC079757.5.5 4037.62672.1	mwghuman3 OK#C:3844	-1.486502	2.69625E- 05	4.12788E- 05	-4.765229	1684. 5	25678.5	23575.5	
假定蛋白 FLJ22313 (FLJ22313), mRNA	NP_071768.2	NM_022373.3	mwghuman3 OK#B:3974	-1.47429	0.0001047 25	5.67358E- 05	-4.784258	1688. 5	8360.5	9165.25	2
全细胞色素C合酶 (细胞色素 C 铁原卅啉裂解酶) (HCCS), mRNA	NP_005324.2	NM_005333.2	mwghuman3 OK#A:04510	-2.084412	0.0003326 32	0.000211 886	-5.979649	1692. 5	10130.5	5488.25	4
双特异性磷酸 酯酶 (DUSP5), mRNA	NP_004410.3	NM_004419.3	mwghuman3 OK#A:04028	-1.866767	5.52516E- 05	2.27276E- 05	-5.708882	1694. 5	16243.5	8885	2
假定蛋白 FLJ20245 (FLJ20245), mRNA	NP_060193.1	NM_017723.1	mwghuman3 OK#B:2083	-1.525601	6.09068E- 05	0.000116 76	-4.325617	1697	16122	13101	3

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	获得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
蛋白激酶, AMP 催化型, γ 2非 催化亚基 (PRK AG2), mRNA	NP_057287.2	NM_016203.2	mwghuman3 OK#A:00771	-1.408092	1.41299E- 05	8.59969E- 05	-4.435082	1697. 5	16315.5	13437.25	5
染色体20上的 克隆rp11-314a4 包含部分eva2 无睡眠同源物2 ests an sts和 gsss starts i		AL359434	mwghuman3 OK#B:9638	-1.591958	1.67257E- 05	6.56055E- 05	-4.203762	1702	25348.5	23840.75	
假定蛋白 pro1598 pro1598		NM_018503	mwghuman3 OK#B:0983	-1.446989	3.89533E- 06	1.30096E- 05	-5.339446	1716. 5	24112.5	22618.75	
NIMA (在有丝 分裂基因a中 没有) 相关的激 酶9 (NEK9), mRNA	NP_149107.3	NM_033116.3	mwghuman3 OK#C:0238	-1.359292	3.922E-05	4.35419E- 05	-4.78725	1722. 5	10134.5	7455.5	5
甘氨酸受体, α 3 (GLRA3) mRNA	NP_006520.1	NM_006529.1	mwghuman3 OK#A:03690	-1.412224	0.0001037 63	8.1682E- 05	-4.59741	1726	9472.5	8083.25	2

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
集团 genscan 预测		AC004408.1.1 .113252.2	mwghuman3 OK#C:9512	-1.54073	2.55537E- 06	1.54035E- 05	-3.618774	1727	2977	2622.75	
集团 genscan 预测		AP001655.2.1 48977.153525 .1	mwghuman3 OK#C:7868	-1.285997	2.64478E- 05	1.96829E- 05	-4.601991	1746	28566	28245.5	
HERV-H 关联 LTR 3 (HHLA3) ，转录变异体3， mRNA	NP_0010317 22.1	NM_0010366 45.1	mwghuman3 OK#B:1046	-2.043346	8.67677E- 05	9.52067E- 05	-3.941702	1759	13925.5	11798	0
集团 genscan 预测		AL355497.14. 1.215397.9	mwghuman3 OK#C:7911	-3.056765	1.10766E- 06	3.82785E- 07	-6.916789	1760	5672.5	2481.25	
纤连蛋白富含 亮氨酸的跨膜 蛋白3 (FLRT3) ，转录变异体1， mRNA	NP_037413.1	NM_013281.2	mwghuman3 OK#A:05822	-4.204569	1.4941E- 08	9.99201E- 07	-6.042443	1770	18156	9231.25	0
cyp3a5等位基因 cyp3a5*3 选择性剪接蛋白		AF355802	mwghuman3 OK#B:1830	-1.371576	7.33293E- 06	1.91497E- 05	-4.362299	1784	26198.5	24518.25	

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
微原纤维蛋白- 相关的蛋白3-样 (MFAP3L), 转录变异体1, mRNA	NP_067679.5	NM_021647.5	mwghuman3 OK#B:0217	-1.987417	0.0001373 75	1.98715E -06	-4.959479	1785	22262	18475.25	3
细胞分裂周期 相关蛋白7 (CD CA7) 转录变异 体1, mRNA	NP_114148.2	NM_031942.3	mwghuman3 OK#B:4810	-1.959598	9.26637E- 05	0.000150 05	-4.246093	1788	14826	12577.25	1
WD重复域61 (WDR61), mRNA	NP_079510.1	NM_025234.1	mwghuman3 OK#B:1624	-1.531798	0.0001544 91	0.000105 506	-4.38785	1807. 5	9756.5	8041.75	3
蛋白激酶, AMP 催化型 B1非催 化亚基 (PRKAB 1), mRNA	NP_006244.2	NM_006253.4	mwghuman3 OK#A:04996	-1.514877	9.29449E- 05	0.000105 506	-4.4366	1805. 5	16011.5	13954	3
锌指, A20域, 包含2 (ZA20D 2), mRNA	NP_005998.1	NM_006007.1	mwghuman3 OK#A:10213	-1.857923	2.78106E- 06	2.67916E -05	-4.781221	1808. 5	6856.5	4563.25	13
类似于 pro2852 loc93374		XM_050978	mwghuman3 OK#B:7206	-1.359998	8.48914E- 05	6.30077E -05	-5.206744	1817	19166	16438	

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	获得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
过氧化氢酶 (CAT), mRNA	NP_001743.1	NM_001752.2	mwghuman3 OK#C:3530	-2.225675	8.1938E- 07	6.92439E- 06	-4.777325	1820	7219.5	4208.75	31
肽基脯氨酰异 构酶 (羧环蛋 白) -样2 (PP IL2), 转录 变异体2, MRNA	NP_680480.1	NM_148175.1	mwghuman3 OK#B:6262	-1.384263	4.77437E- 07	4.59274E- 06	-5.396075	1834. 5	3806	3514	5
表胶原蛋白 (PRX), 转录 变异体2, mRNA	NP_870998.1	NM_181882.1	mwghuman3 OK#A:01781	-1.399169	1.03356E- 05	6.63845E- 05	-4.549994	1838	19643.5	17122.5	0
乳腺癌放大 序列3 (BCA S3), mRNA	NP_060149.2	NM_017679.2	mwghuman3 OK#A:01576	-1.41274	2.55423E- 07	3.64015E- 06	-4.03611	1839	15754.5	17960.25	2
HtrA丝氨酸 肽酶2_HIRA2	AAF66598.1	AF141307.1	mwghuman3 OK#B:1170	-1.284282	2.27894E- 05	1.86362E- 05	-4.071282	1850	22760.5	20044.75	
集团 genscan 预测		AC012095.13. 12848.17419. 1	mwghuman3 OK#C:7012	-1.517696	0.0001343 14	0.000172 435	-4.407333	1854	26139.5	25010.5	
形成素结合 蛋白4 (FNB P4), mRNA	NP_056123.1	NM_015308.1	mwghuman3 OK#B:8897	-1.360042	0.0001194 22	2.15085E- 05	-5.532889	1855	17142	14814.25	11

图6(续)

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
醛基脱氢酶1 家族, 成员L1 (ALDH1L1), mRNA	NP_036322.2	NM_012190.2	mwghuman3 OK#A:07479	-2.396851	2.32809E- 05	0.000157 701	-3.740555	1856. 5	13469	8189.5	3
集团 genscan 预测		AP001487.3.8 5282.114407. 1	mwghuman3 OK#C:8733	-1.603991	4.09016E- 06	5.97939E -05	-4.353485	1870	3882	3994.75	
集团 genscan 预测		AC015651.18. 1.191583.4	mwghuman3 OK#C:2191	-1.392136	5.55423E- 07	1.72268E -05	-4.562358	1873. 5	2463	1928.25	
OUT域包含5 (OTUD5), mRNA	NP_060072.1	NM_017602.2	mwghuman3 OK#B:7999	-1.721318	3.42749E- 06	1.37651E -05	-5.356883	1894. 5	1453	1011	14
集团 genscan 预测		AC026668.5.8 4589.93793.1	mwghuman3 OK#C:5546	-1.668723	3.74924E- 06	2.40122E -05	-5.209366	1899	3201	1676.5	
染色体20可读框 121 (c20orf121) , mRNA	NP_077307.1	NM_024331.2	mwghuman3 OK#A:03031	-1.520517	5.93503E- 08	1.82138E -05	-4.163768	1906	8788	6894	2
丙酮酸脱氢酶 激酶, 同工酶4 (PDK4), mRNA	NP_002603.1	NM_002612.2	mwghuman3 OK#A:07564	-1.735687	1.72294E- 05	1.22937E -05	-5.445156	1911. 5	18067	13997.25	1

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
KIAA1049蛋白 (KIAA1049) , mRNA	NP_055787.1	NM_014972.1	mwghuman3 OK#B:2555	-1.400381	1.30483E- 05	7.36585E- 05	-4.350759	1913. 5	10972	8868.5	12
DnaJ (Hsp40) 同源物, 家族B, 成员6 (DNAJB6), 转录变异体1, mRNA	NP_490647.1	NM_058246.3	mwghuman3 OK#A:04664	-1.455502	9.60314E- 05	0.000150 05	-4.286094	1917	3609	2895.5	14
外源凝集素, 半乳糖苷结合, 可溶型, 3 (半 乳糖凝集素3) (LG ALS3), mRNA	NP_002297.1	NM_002306.1	mwghuman3 OK#A:10306	-1.93733	6.69527E- 06	2.55121E- 05	-4.690215	1923	2753	1803.5	6
PTPRF交互蛋白, 结合蛋白2 (lip rin 3.2) (PPFI BP2), mRNA	NP_003612.1	NM_003621.1	mwghuman3 OK#B:0552	-1.881545	1.28489E- 06	2.98753E- 05	-4.14602	1937. 5	6372.5	4594.5	2
聚合酶 (DNA 导向的), 入 (POL), mRNA	NP_037406.1	NM_013274.2	mwghuman3 OK#B:1279	-1.399567	3.5582E- 05	2.98753E- 05	-4.793654	1938	15108.5	11461.25	8

图6(续)

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
生长差异因子15 (GDF15), mRNA	NP_004855.1	NM_004864.1	mwghuman3 OK#A:07217	-2.483648	0.0002241 34	0.000257 25	-4.296268	1944	6377	3127.5	2
假定蛋白 xp_089809 loc159765		XM_089809	mwghuman3 OK#B:6474	-1.381507	9.90157E- 07	2.40122E -05	-4.093057	1949	11626.5	11522.5	
富亮氨酸重复 序列以及无因 α 基序包含1 (LRSAM1), 转录变异体1, mRNA	NP_612370.3	NM_138361.3	mwghuman3 OK#B:4416	-2.034574	0.0001840 31	0.000233 537	-4.13512	1959. 5	7671.5	4421.75	1
运输蛋白粒子 复合物6A (TR APP6A), mRNA	NP_077013.1	NM_024108.1	mwghuman3 OK#A:02117	-1.452954	2.44741E- 06	2.53657E -05	-3.509555	1985	1279	1294	8
集团 genscan 预测		AL357146.10. 1.82494.1	mwghuman3 OK#C:3608	-1.287782	1.49736E- 05	6.05474E -05	-4.826559	1989	28444.5	26083.25	
假定蛋白 MGC14327 (MGC14327), mRNA	NP_444273.1	NM_053045.1	mwghuman3 OK#B:4427	-1.504495	0.0001502 58	0.000174 117	-4.521635	1992	16805	15440.5	4

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
类似于驱动 蛋白家族成 员5b克隆 mgc: 1565 图像: 429779		BC009353	mwghuman3 OK#B:4438	-1.51682	0.0001882 14	0.000192 131	-4.531075	1992	23071.5	21809.75	
含钾通道 四聚化结 构域包含3 (KCTD3), mRNA	NP_057205.2	NM_016121.3	mwghuman3 OK#A:04386	-1.358006	1.34593E- 05	5.38262E- 05	-4.810824	1994. 5	17479.5	13203.75	3
集团 genscan 预测		AC018425.3.1 53134.186975 1	mwghuman3 OK#C:3925	-1.496534	0.0001495 25	0.000192 131	-4.315259	1997	25336	25414.75	
鸟嘌呤核苷酸 结合蛋白 (G 蛋白) α 抑制 活性多肽1 (GNAI1). mRNA	NP_002060.4	NM_002069.4	mwghuman3 OK#A:10201	-1.631361	5.34686E- 06	0.000433 377	-3.885346	2003	16303.5	14082	1
微管相关蛋 白4 (MAP4), 转录变异体1, mRNA	NP_002366.2	NM_002375.3	mwghuman3 OK#B:4355	-1.468752	0.0001551 87	0.000257 25	-4.423897	2004	10052	9080	11

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
核仁蛋白家族6 (RNA相关的) (NOL6), 转录 变异体 α, mRNA	NP_075068.2	NM_022917.4	mwghuman3 OK#A:07182	-1.30257	1.12604E- 06	3.70831E- 05	-4.767639	2009	10513	8821	7
脑血管内皮 细胞粘附分子1 (CEECAM1), mRNA	NP_057258.2	NM_016174.3	mwghuman3 OK#A:10231	-1.516413	4.80848E- 05	0.000109 198	-3.667581	2014	16662.5	13757.75	0
泛醌-细胞色素c 还原酶, 复合物 III, 子组VII, 9.5kDa (UOCCR), 核基因编码线粒 体蛋白, mRNA	NP_055217.2	NM_014402.3	mwghuman3 OK#B:8060	-1.71221	4.79681E- 05	4.35419E- 05	-5.049156	2014	4199.5	2985.25	14
假定蛋白 xp_092745 loc164337		XM_092745	mwghuman3 OK#B:3255	-1.386871	3.26973E- 05	0.000165 718	-4.513374	2041	19510	15741.25	

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
假定蛋白 FLJ22965 (FLJ22965), mRNA	NP_071384.1	NM_022101.2	mwghuman3 OK#A:09776	-1.903469	7.16955E- 05	0.000475 592	-4.078844	2041. 5	2596	2337	0
染色体9可读框 102 (C9orf102) , 转录变异体1, mRNA	NP_064592.1	NM_020207.2	mwghuman3 OK#B:6956	-1.369819	3.05523E- 05	0.000105 506	-4.676244	2049	16457.5	13970	0
鞘氨醇-1-磷酸 盐磷酸酶1 (SG pp1), mRNA	NP_110418.1	NM_030791.2	mwghuman3 OK#B:1809	-1.477332	4.90035E- 05	0.000142 749	-4.328137	2062. 5	20094	17541.5	1
普列克底物 蛋白及Sec7 域包含3 (PS D3), 转录 变异体1, mRNA	NP_056125.2	NM_015310.2	mwghuman3 OK#B:9515	-1.395381	2.70551E- 05	0.000142 896	-3.953529	2072	14023	13223.25	1
假定蛋白 LOC283874 (LOC283874), mRNA	NP_0010127 49.1	NM_0010127 31.1	mwghuman3 OK#B:2101	-1.376341	1.45793E- 05	2.67916E- 05	-5.079903	2090. 5	23511.5	23000	1

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MMIG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
锌指, FYVE域 包含21 (ZFYE 21), mRNA	NP_076976.1	NM_024071.2	mwghuman3 OK#A:05642	-1.579562	0.0001142 53	0.000342 648	-4.238027	2094	2994.5	2740.25	0
叉头框E1 (甲 状腺转录因子2) (FOXE1), mRNA	NP_004464.2	NM_004473.3	mwghuman3 OK#A:07899	-1.678652	1.3104E- 05	0.000192 131	-3.837954	2107	7269	5293.75	0
集团 genscan 预测		AL158168.17. 82811.114103 1	mwghuman3 OK#C:5892	-1.54059	5.72345E- 05	8.88591E- 06	-4.136599	2114	28578	27753.75	
类似于每六聚 物重复蛋白5 loc149135		XM_089208	mwghuman3 OK#B:9694	-1.363387	0.0001055 64	3.77672E- 05	-3.9371	2131	27002.5	27249.5	
锌指, 基质蛋白 型1 (ZMAT1), 转录变体3, mRNA	NP_115817.1	NM_032441.1	mwghuman3 OK#C:0106	-1.423696	0.0001562 67	0.000211 896	-4.538065	2134. 5	20339.5	17190.25	2
核信号2的 内质网 (ER N2), mRNA	NP_150296.2	NM_033266.2	mwghuman3 OK#B:9444	-1.735187	1.98165E- 05	2.70994E- 05	-4.509363	2135	24789	18903	0

图6 (续)

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	获得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
环指蛋白38 (RNF38), 转录变异体1, mRNA	NP_073618.3	NM_022781.3	mwghuman3 OK#B:5588	-1.302965	2.50283E- 06	2.15085E- 05	-4.986001	2138	24546.5	22148.75	1
Myc诱导的线 粒体蛋白 (mi- mitin), mRNA	NP_777549.1	NM_174889.2	mwghuman3 OK#B:3716	-1.474873	8.439E-05	0.000359 234	-4.207401	2151. 5	9215	7030.25	1
TGF β 可诱导 核蛋白1 (T1 NP1), mRNA	NP_055701.1	NM_014886.2	mwghuman3 OK#A:06520	-1.655993	0.0003187 59	0.000571 767	-4.796546	2153	7530	5545.5	6
类似于多能发育 相关基因5胚胎 干细胞基因1 (LOC341912), mRNA	XP_292301.3	XM_292301.3	mwghuman3 OK#B:8134	-1.639786	7.64251E- 07	4.32856E- 06	-5.437632	2155	18581.5	12815.25	1
脱氢酶/还原酶 (SDR家族) 成 员3 (DHR3), mRNA	NP_004744.2	NM_004753.4	mwghuman3 OK#B:0738	-1.439285	4.39582E- 05	0.000222 464	-3.974197	2166. 5	4626	3570	3

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
黄素包单 氧化酶4 (F MO4), mRNA	NP_002013.1	NM_002022.1	mwghuman3 OK#B:2927	-1.804252	0.0001959 67	0.000222 464	-4.159771	2173. 5	21428	18241	1
集团 genscan 预测		AC022526.4.9 087.13827.1	mwghuman3 OK#C:9081	-1.518222	0.0001264 09	0.000174 117	-4.437239	2174	25713.5	23316	
血清学检测 结肠癌抗原8 (SDCCAG8), mRNA	NP_006633.1	NM_006642.1	mwghuman3 OK#B:4992	-1.359218	5.81032E- 06	1.22937E- 05	-4.628131	2177. 5	18648	16566	4
CDC14细胞分裂 周期14同源物B (S.啤酒) (CD C14B), 转录 变体2, mRNA	NP_201588.1	NM_033331.1	mwghuman3 OK#B:0753	-1.438486	0.0001500 76	0.000245 124	-4.702655	2178. 5	23322	22971	2
髓系/淋巴或 混合的世系白 血病 (三聚同 源物, 果蝇属) ; 易位给3 (ML LT3), mRNA	NP_004520.1	NM_004529.1	mwghuman3 OK#A:00725	-1.503791	3.87447E- 05	0.000311 603	-4.282874	2188	20164.5	17462	4

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
细胞色素P450, 家族27, 子家 族A, 多肽1 (CYP27A1), 核基因编码线 粒体蛋白, mRNA	NP_000775.1	NM_000784.2	mwghuman3 OK#A:07422	-1.399906	4.05019E- 05	3.91276E- 06	-4.065883	2197	20714	13781.5	10
Pre-tnk细胞 相关蛋白1f6 3' 末端		L17326	mwghuman3 OK#B:9025	-1.643879	5.92984E- 08	3.84322E- 06	-5.055927	2201	6449.5	4787.5	
Hepcidin 抗菌肽 (HA MP), mRNA	NP_066998.1	NM_021175.2	mwghuman3 OK#A:06397	-1.279742	2.53831E- 05	1.72288E- 05	-4.893297	2205. 5	3722.5	3385.25	0
集团 genscan 预测		AC078819.13. 1.41283.1	mwghuman3 OK#C:7792	-1.349603	2.94705E- 06	7.88194E- 05	-4.262185	2213	21381.5	17800.75	
假定蛋白 BC009862 (LOC90113), mRNA	XP_291077.3	XM_291077.3	mwghuman3 OK#B:3668	-1.769275	0.0003405 59	0.000454 027	-4.413577	2221. 5	6261.5	5176	0
伸长因子, RNA 聚合酶II, 2 (ELL2), mRNA	NP_036213.1	NM_012081.3	mwghuman3 OK#A:00356	-1.499648	5.40285E- 05	0.000157 701	-4.384409	2223. 5	14138	11702.5	4

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
EGFR1 癌基因 伴侣 (FGFR10P) , 转录变异体1, mRNA	NP_008976.1	NM_007045.2	mwghuman3 OK#B:4675	-1.513357	9.11409E- 05	0.000233 537	-4.085108	2228. 5	19894	19145.25	0
集团 genscan 预测 EH域结合蛋 白1 (EHBP1) , mRNA		AC025953.2.1 16948.127472 .1	mwghuman3 OK#C:6647	-2.043219	0.0001551 45	0.000283 207	-3.955032	2231	21041	17429.5	
	NP_056067.1	NM_015252.2	mwghuman3 OK#B:0395	-1.35833	2.50342E- 06	6.99321E -05	-4.628579	2249. 5	13151.5	12131	5
丙酮辅酶 A羧化酶, α 多肽 (POCA), mRNA	NP_000273.2	NM_000282.2	mwghuman3 OK#B:3551	-1.505038	6.411E-05	0.000110 999	-4.042034	2251	16323	13459	5
集团 genscan 预测		AL138499.4.1 .185713.1	mwghuman3 OK#C:4137	-1.655664	1.65577E- 05	8.59969E -05	-3.700791	2252. 5	6728	6542.5	
集团 genscan 预测		AC025937.3.2 8407.38498.1	mwghuman3 OK#C:9420	-5.821331	1.29566E- 11	3.14698E -07	-6.568115	2255. 5	3524	1068.5	

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
肌纤蛋白, 小梁网糖 皮质激素 诱导反应 蛋白 (MYOC), mRNA	NP_000252.1	NM_000261.1	mwghuman3 OK#A:09642	-1.364765	4.61814E- 05	2.61402E- 05	-4.638227	2263	23815.5	22357.25	0
hspc074		AF161337	mwghuman3 OK#B:1281	-1.61806	1.54965E- 06	1.54035E- 05	-4.426706	2264. 5	1792	1727	3
环氧水解酶2, 细胞质蛋白 (EPHX2), mRNA	NP_001970.2	NM_001979.4	mwghuman3 OK#A:08300	-1.416398	5.8666E- 06	5.81228E- 06	-4.895716	2267. 5	19752.5	16857	3
着丝粒蛋白C1 (CENPC1), mRNA	NP_001803.2	NM_001812.2	mwghuman3 OK#A:10535	-1.578112	0.0001833 56	0.000192 131	-4.258306	2275	19346.5	14107.25	4
多(A)聚合酶β (羧基特异性) (PAPOLB), mRNA	NP_064529.4	NM_020144.4	mwghuman3 OK#A:00034	-1.533242	0.0001793 94	0.000311 603	-4.111646	2284	24777	23772.75	1
寡核苷酸/寡糖 -结合折叠包含1 (OBFC1), mRNA	NP_079204.1	NM_024928.3	mwghuman3 OK#A:07366	-1.597262	0.0001519 06	0.000150 05	-4.339099	2292	15940.5	11655	4

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
三重基序-包含 10 (TRIM10), 转录变异体1, mRNA	NP_006769.1	NM_006778.2	mwghuman3 OK#A:10515	-1.750912	5.52205E- 05	0.000135 784	-4.621742	2296. 5	4089.5	1565.5	0
Rap2-结 合蛋白9 (RPIB9), mRNA	NP_612147.1	NM_138290.1	mwghuman3 OK#B:7696	-1.451227	1.03873E- 06	1.37651E -05	-5.066475	2297	24136.5	22036.5	1
fksg17		NM_032031	mwghuman3 OK#B:1656	-1.576144	0.0001418 34	0.000233 537	-4.037961	2320. 5	2050	1701.75	1
G蛋白耦 合受体83 (GPR83), mRNA	NP_057624.2	NM_016540.2	mwghuman3 OK#A:05359	-1.472466	6.54909E- 05	0.000182 916	-3.811779	2322. 5	18971.5	15879.25	0
集团 genscan 预测		AC078987.9.1 35718.144151 .1	mwghuman3 OK#C:4632	-1.300374	2.50018E- 05	1.72268E -05	-4.918105	2322. 5	25018.5	21332.5	
LIM同源框9 (LH X9), 转录变异 体1, mRNA	NP_064589.2	NM_020204.2	mwghuman3 OK#B:8634	-1.374971	4.6951E- 06	9.24336E -06	-5.312469	2333. 5	16158	13872.25	0
皮层同源物1 (果蝇属) (PEL1), mRNA	NP_065702.2	NM_020651.2	mwghuman3 OK#A:10770	-1.492094	8.8328E- 06	8.23838E -06	-5.23835	2336	22421.5	17524.5	7

图6 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
假定 LOC401500 (LOC401500), mRNA	XP_379629.1	XM_379629.1	mwghuman3 OK#B:6223	-1.325856	2.27785E- 05	4.12788E- 05	-4.402696	2336. 5	26315	25320.25	0
TSC22域家族, 成员3 (TSC22 D3), 转录变 异体2, mRNA	NP_004080.2	NM_004089.3	mwghuman3 OK#B:0462	-1.491298	5.41371E- 05	1.72268E- 05	-5.010311	2350. 5	18693.5	13200.5	43
G蛋白耦 合受体51 (GPR51), mRNA	NP_005449.5	NM_005458.5	mwghuman3 OK#A:04815	-1.389185	0.0001017 06	2.98753E- 05	-4.845168	2361	3420	2489.25	0
染色体10可读 框61 (C10or f61), 转录 变异体2, mRNA	NP_056446.1	NM_015631.2	mwghuman3 OK#B:3105	-1.555923	0.0001770 95	0.000297 088	-4.510054	2366. 5	10134	8091.5	3
十字形碱 蛋白包含2C (JMJ2C), mRNA	NP_055876.1	NM_015061.1	mwghuman3 OK#B:8433	-1.317474	6.41124E- 05	0.000142 749	-4.870495	2366. 5	24516.5	22272.75	11
锌指蛋白547 (ZNF547), mRNA	NP_775902.2	NM_173631.2	mwghuman3 OK#B:6716	-1.63256	0.0002511 74	0.000257 25	-3.984452	2370	19484	19537	2

图6(续)

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	MWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
假定蛋白 FLJ20032 (FLJ20032), mRNA	NP_060098.2	NM_017628.2	mwghuman3 OK#B:2068	-1.74414	6.61052E- 05	0.000105 506	-4.166213	2374	15233.5	9767.75	16
集团 genscan 预测		AC064677.3.1 21131.166993 .1	mwghuman3 OK#C:7466	-1.319284	4.28953E- 05	2.03519E -05	-4.802131	2374	27348	26740.25	
集团 genscan 预测		AC027755.2.8 1287.97090.1	mwghuman3 OK#C:3411	-1.499071	0.0001259 01	0.000331 474	-4.145868	2378. 5	14385.5	10781.5	
KIAA0690 (KIAA0690), mRNA	NP_055994.1	NM_015179.2	mwghuman3 OK#B:7202	-1.642414	0.0001739 36	0.000433 377	-4.08872	2380	14459	11844.25	10
SH3域结合 谷氨酸富 酸蛋白样2 (SH3BGR12), mRNA	NP_113657.1	NM_031469.1	mwghuman3 OK#B:1794	-1.995457	0.0002796 76	0.000376 568	-4.033794	2381	9893.5	6985.75	3
cdna: flj21394 fis 克隆co103 536未命名 蛋白产品		AK025047	mwghuman3 OK#B:2580	-2.258109	0.0004473 76	0.000359 234	-4.071961	2387. 5	17288	11926.75	0
假定蛋白 xp_039231 loc91565		XM_039231	mwghuman3 OK#B:6431	-1.337232	2.70835E- 05	0.000174 117	-4.141157	2398	24099.5	20885	

图6(续)

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
GPI 锚定 分子样蛋白	NP_002057.1	NM_002066.1	mwghuman3 OK#A:02108	-1.438045	0.0000724 93	0.000129 14	-4.162956	2405	16818	14265.75	0
小核RNA催化 复合物, 多肽5, 19kDa (SNAPC5), mRNA	NP_006040.1	NM_006049.1	mwghuman3 OK#B:0945	-1.323754	1.74702E- 05	0.000184 787	-4.006319	2444	17556.5	17969.25	6
集团 genscan 预测		AC027689.10. 1.180573.5	mwghuman3 OK#C:9758	-1.644263	0.0002550 65	2.98753E- 05	-3.674603	2451	16829.5	10684.25	
线粒体肿瘤 抑制因子1 (MTUS1), 核基因编码 线粒体蛋白, 转录变异体1, mRNA	NP_0010019 24.1	NM_0010019 24.1	mwghuman3 OK#B:7775	-1.715161	0.0003036 63	0.000269 936	-4.435392	2461. 5	17489	15873	0
集团 genscan 预测		AL049829.4.1 .196292.3	mwghuman3 OK#C:4548	-1.402914	1.70061E- 06	2.49927E- 06	-5.01844	2471	28131	26764	
集团 genscan 预测		AC084842.1.7 903.10639.1	mwghuman3 OK#C:7638	-1.400793	0.0008120 94	7.36585E- 05	-5.107784	2475. 5	14235.5	11061.25	

图6 (续)

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	MMWG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
SAFB-样, 转录调节器 (SLIM) 转录 变异体1, mRNA	NP_079031.2	NIM_024755.2	mwghuman3 OK#B:2154	-1.309539	2.00445E- 05	0.000122 803	-4.680439	2477. 5	12490	12280	12
鸟嘌呤核苷酸 结合蛋白 (G蛋 白) α 催化活性 多肽, 嗅觉型 (GNAL), 转录 变异体2, mRNA	NP_002062.1	NIM_002071.1	mwghuman3 OK#A:06286	-1.836112	0.0003002 6	0.000498 109	-4.10637	2478	19959	16690.75	0
类似于cg14182, 产品 loc146175		XM_085352	mwghuman3 OK#B:7365	-1.746066	0.0001949 23	0.000433 377	-3.812615	2487	14791.5	9814.5	
PH域及富亮氨酸 重复蛋白磷酸酶 (PHLPP), mRNA	NP_919431.1	NIM_194449.1	mwghuman3 OK#B:2242	-2.043619	0.0003079 45	0.000498 109	-4.006924	2500	15125	10009.75	7

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
自闭症 候选2 (AUTS2) mRNA	NP_056385.1	NM_015570.1	mwghuman3 OK#B:0110	-1.807633	0.0003692 99	0.000382 814	-4.213775	2503	18404	14412.75	8
假定蛋白 xp_097338 lcc147909		XM_097338	mwghuman3 OK#B:5652	-1.428873	1.32234E- 06	2.67916E- 05	-4.387947	2513	4388	2991.75	
富亮氨酸 重复包含G 蛋白耦合受体4 (LGR4), mRNA 集团 预测	NP_060960.1	NM_018490.1	mwghuman3 OK#A:06106	-1.519205	7.81416E- 05	0.000122 803	-3.810559	2514. 5	21417.5	19684	2
编程的 细胞死亡1 (PDCD1), mRNA		ENSG000001 13407	mwghuman3 OK#C:2560	-1.740805	1.93346E- 05	9.52811E- 05	-4.388894	2525	8797.5	6359.5	
钙粘着蛋白 20, 类型2 (CDH20), mRNA	NP_005009.1	NM_005018.1	mwghuman3 OK#B:9147	-1.519488	1.50412E- 06	3.41028E- 06	-5.713059	2537	8575	7118.5	0
	NP_114097.2	NM_031891.2	mwghuman3 OK#B:7584	-1.393019	1.91884E- 05	2.73302E- 05	-4.362445	2538	15588	12275.75	0

图6(续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	MMG 探针名称	中位倍 数变化	t-检测	Wilcoxon 检测	SAM	秩得分	肿瘤 组织 中位 强度秩	非恶性 组织 中位 强度秩	血液
甘氨酸脱氢酶 (脱羧的; 甘 氨酸脱羧酶, 甘氨酸分裂系 统蛋白) (GL DC), mRNA	NP_000161.1	NM_000170.1	mwghuman3 OK#B:8959	-1.405054	9.33312E- 05	0.000116 76	-4.398207	2542. 5	6039	4550.5	1
蛋白酶丝 氨酸, 35 (PRSS35), mRNA	NP_699193.1	NM_153362.1	mwghuman3 OK#B:3964	-1.451818	0.0002370 39	0.000257 25	-4.319803	2545	16909	14109	0
集团 gensean 预测		AC008134.3.1 175132.1	mwghuman3 OK#C:3643	-1.601839	0.0003788 4	0.000286 639	-4.042534	2545	27071.5	25825	
ATP-结合组件, 子家族A (ABC1), 成员12 (ABCA12), 转录变异体1, mRNA	NP_775099.2	NM_173076.2	mwghuman3 OK#B:3505	-1.653773	0.0002371 78	0.000475 592	-3.927921	2561. 5	19931	18593.5	0
多毛翅同 源物3 (果蝇属) (SUHW3), mRNA	NP_060136.1	NM_017666.2	mwghuman3 OK#B:2904	-1.678106	0.0001863 59	0.000546 156	-3.864729	2568	17527	15364.5	1

图6 (续)

基因名称/符号	蛋白登录号	mRNA 参考 序列登录号	中位倍数变化	秩得分	肿瘤组织 中位强度秩	非恶性组织 中位强度秩	血液EST号
假定蛋白 FLJ21511 (FLJ21511), mRNA 成对框基因 8 (PAX8), 转录变异体 PAX8A, mRNA UPF3无义转录同源A的UPF3调节剂 (yeast) (UPF3A), 转录变异体 1, mRNA	NP_079363.1 NP_003457.1 NP_075387.1	NM_025087.1 NM_003466.3 NM_023011.2	-2.080908 -1.961896 -3.84007	68 82 96.5	14455 19419 5663.5	9272.5 12999.5 1825	0 0 0
白三烯B412-羟基脱氢酶 (LTB4DH), mRNA RAB11家族相互作用蛋白2 (1类) (RAB11FIP2), mRNA TSC22域家族, 成员1 (TSC22D1), 转录变异体 2, mRNA 亚硫酸盐氧化酶 (SUOX), 核基因编码线粒体蛋白, 转录变异体1 1, mRNA 三角形四肽重复21A (TTC21A), mRNA 突触囊泡蛋白样蛋白2 (SYTL2), 转录变异体 b, mRNA 无翅型WNT整合位点家族, 成员2B (WNT2B), 转录变异体 WNT-2B1, mRNA 原钙黏蛋白 v 子家族A, 12 (PCDHGA12), 转录变异体 1, mRNA 嗅觉感受器家族1, 子家族D, 成员5 (OR1D5), mRNA G蛋白耦联受体126 (GPR126), 转录变异体 b1, mRNA	NP_036344.1 NP_055719.1 NP_006013.1 NP_000447.2 NP_665698.1 NP_115755.2 NP_004176.2 NP_003726.1 NP_055381.1 NP_940971.1	NM_012212.2 NM_014904.1 NM_006022.2 NM_000456.2 NM_145755.1 NM_032379.3 NM_004185.2 NM_003735.2 NM_014566.1 NM_198569.1	-5.034613 -1.993629 -2.133085 -1.622803 -2.976809 -2.343357 -10.745275 -2.070447 -1.640751 -2.181735	122.5 156 156.5 223 225 342 501 511 539.5 549	11340.5 19636.5 8415 11312.5 8375.5 19920 10716 2537 1644 15965.5	3138 12951 6130 8789.5 3870.5 12053.5 2016.25 1931.5 1134 8040.75	0 0 1 0 1 0 0 0 0 0
范素特异性肽酶9, Y连锁型 (脂肪小平面样, 果蝇属) (USP9Y), mRNA	NP_004645.2	NM_004654.3	-1.584085	587	10035.5	7858.25	1

图7

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	中位倍数变化	秩得分	肿瘤组织 中位强度秩	非恶性组织 中位强度秩	血液EST号
染色体11可读框1 (C11orf1), mRNA	NP_073598.1	NM_022761.1	-1.81726	673.5	12025.5	9585.25	0
BCL2相关的永生基因 (BAG1), mRNA	NP_004314.3	NM_004323.3	-2.525417	776	9031	4503.5	1
具有序列相似性的家族44, 成员B (FAM44B), mRNA	NP_612378.1	NM_138369.1	-1.357114	942.5	15308.5	13004.75	1
成纤维细胞生长因子4 (FGFR4), 转录变体2, mRNA	NP_075252.2	NM_022963.2	-2.638426	1087.5	17096	9762.5	0
甲状腺转录因子1 (TTF1), mRNA	NP_003308.1	NM_003317.3	-1.468987	1124.5	2139	1026.25	0
Vpr-BP蛋白 (Vpr-BP), mRNA	NP_065518.1	NM_014703.1	-1.720326	1178	17312.5	14031	0
空气孔同源物2 (果蝇属) (EMX2), mRNA	NP_004089.1	NM_004098.2	-1.745086	1199.5	19380	10749.25	0
同源框8 (HOXD8), mRNA	NP_062458.1	NM_019558.2	-1.415555	1206	9672.5	7512.25	0
转指蛋白626 (ZNF626), mRNA	NP_660340.1	NM_145297.2	-1.773183	1266	14806.5	10341	1
环指蛋白44 (RNF44), mRNA	NP_055716.1	NM_014901.4	-1.353366	1292	6288.5	4664.5	1
嘧啶受体P2Y ₄ G蛋白耦合, 4 (P2RY4), mRNA	NP_002556.1	NM_002565.3	-1.732075	1309	18307	12139.5	0
类似于假定蛋白 (LOC440804), mRNA	XP_036936.3	XM_036936.3	-4.049648	1332	3847	996.5	0
染色体20可读框152 (C20orf152), mRNA	NP_543024.1	NM_080834.1	-1.948612	1349	10775	8652.5	0
卵泡抑素-样4 (FSTL4), mRNA	NP_055897.1	NM_015082.1	-1.544042	1405.5	17051	12319	0
染色体14可读框168 (C14orf168), mRNA	NP_113615.1	NM_031427.1	-1.512002	1414	15738	14705.75	1
激活释放素8 (丝氨酸蛋白酶/卵素) (KLK8), 转录变体2, mRNA	NP_653088.1	NM_144505.1	-2.091505	1449.5	8391	6488	0
亮氨酸拉肽在癌症1中下调-样 (LD001L), mRNA	NP_115663.2	NM_032287.2	-1.501021	1454.5	16509.5	14157	0
假定蛋白MGC11242 (MGC11242), mRNA	NP_077296.1	NM_024320.2	-1.527839	1463.5	17565	13134.5	0
KIAA1274 (KIAA1274), mRNA	NP_055246.1	NM_014431.1	-1.41942	1488.5	8059	5065	0
假定LOC401510 (LOC401510), mRNA	XP_376843.2	XM_376843.2	-1.402922	1504	13882.5	11233.25	0

图7 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	中位倍数变化	秩得分	肿瘤组织 中位强度秩	非恶性组织 中位强度秩	血液EST号
丝束蛋白1 (同源异构体) (PLS1), mRNA 锌指, DHH型包含2 (ZDHC2), mRNA	NP_002661.1	NM_002670.1	-1.626206	1509.5	20943	14040.5	1
视锥杆同源框 (CRX), mRNA PHD指蛋白23 (PHF23), mRNA	NP_057437.1 NP_000545.1 NP_077273.1	NM_016353.2 NM_000554.2 NM_024297.1	-1.588333 -3.022299 -1.398461	1536 1548 1657.5	13684.5 16588.5 10656.5	10087.5 7500.5 9355.75	0 0 0
鼻咽癌相关的基因蛋白8 (NAG8), mRNA HERV-H 关联LTR 3 (HILA3), 转录变体3, mRNA	NP_055226.1	NM_014411.2 NM_001036645.1	-1.388055 -2.043346	1660 1759	15731.5 13925.5	12161 11798	0 0
纤连蛋白富含亮氨酸的跨膜蛋白3 (FLRT3), 转录变体1, mRNA	NP_001031722.1						
细胞分裂周期相关蛋白7 (CDCA7) 转录变体1, mRNA 丙酮酸脱氢酶激酶, 同工酶4 (PDK4), mRNA	NP_037413.1 NP_114148.2 NP_002603.1	NM_013281.2 NM_031942.3 NM_002612.2	-4.204569 -1.959598 -1.735687	1770 1788 1911.5	18156 14826 18067	9231.25 12577.25 13997.25	0 1 1
富含亮氨酸重复序列以及无α基序包含1 (LRSM1), 转录变体1, mRNA	NP_612370.3	NM_138361.3	-2.034574	1959.5	7671.5	4421.75	1
鸟嘌呤核苷酸结合蛋白 (G蛋白) α抑制活性多肽1(GNAI1), mRNA 脑血管内皮细胞粘附 分子1 (CEECAM1), mRNA 假定蛋白 FLJ22965 (FLJ22965), mRNA	NP_002060.4 NP_057258.2 NP_071384.1	NM_002069.4 NM_016174.3 NM_022101.2	-1.631361 -1.516413 -1.903469	2003 2014 2041.5	16803.5 16662.5 2596	14082 13757.75 2337	1 0 0
染色体9可读框102 (C9orf102), 转录变体1, mRNA 普列克底物蛋白及Sec7域包含3 (PSD3), 转录变体1, mRNA 锌指, FYVE域包含21 (ZFYVE21), mRNA 叉头框E1 (甲状腺转录因子2) (FOXE1), mRNA	NP_064592.1 NP_056125.2 NP_076976.1 NP_004464.2	NM_020207.2 NM_015310.2 NM_024071.2 NM_004473.3	-1.369819 -1.395381 -1.579562 -1.678652	2049 2072 2094 2107	16457.5 14023 2994.5 7269	13970 13233.25 2740.25 5293.75	0 1 0 0

图7 (续)

基因名称/符号	蛋白登录号	mRNA参考 序列登录号	中位倍数变化	秩得分	肿瘤组织 中位强度秩	非恶性组织 中位强度秩	血液EST号
Myo诱导的线粒体蛋白 (mitin), mRNA	NP_777549.1	NM_174889.2	-1.474873	2151.5	9215	7030.25	1
类似于多能发育相关基因5 胚胎干细胞基因1 (LOC341912), mRNA	XP_292301.3	XM_292301.3	-1.639786	2155	18581.5	12815.25	1
Hepcidin抗菌肽 (HAMP), mRNA	NP_066998.1	NM_021175.2	-1.279742	2205.5	3722.5	3385.25	0
假定蛋白 BC009862 (LOC90113), mRNA	XP_291077.3	XM_291077.3	-1.769275	2221.5	6261.5	5176	0
三重基序-包含10 (TRIM10), 转录变体1, mRNA	NP_006769.1	NM_006778.2 NM_032031	-1.750912 -1.576144	2296.5 2320.5	4069.5 2050	1565.5 1701.75	0 1
flksg17 flksg17 LIM同源框9 (LHX9), 转录变体1, mRNA	NP_064589.2	NM_020204.2	-1.374971	2333.5	16158	13872.25	0
G蛋白耦合受体51 (GPR51), mRNA	NP_005449.5	NM_005458.5	-1.389185	2361	3420	2489.25	0
cdna: flj21394.f1s 克隆cc103536未命名蛋白产品 GPI锚定分子样蛋白 (GML), mRNA		AK025047	-2.258109	2387.5	17288	11926.75	0
编程的细胞死亡1 (PDCD1), mRNA	NP_002057.1	NM_002066.1	-1.438045	2405	16818	14265.75	0
钙粘着蛋白20, 类型2 (CDH20), mRNA	NP_005009.1 NP_114097.2	NM_005018.1 NM_031891.2	-1.519488 -1.393019	2537 2538	8575 15588	7118.5 12275.75	0 0
甘氨酸胺酶(脱羧的; 甘氨酸脱羧酶, 甘氨酸分裂系统蛋白) (GLDC), mRNA	NP_000161.1	NM_000170.1	-1.405054	2542.5	6039	4550.5	1
蛋白脯氨酸, 35 (PRSS35), mRNA	NP_699193.1	NM_153362.1	-1.451818	2545	16909	14109	0

图7(续)

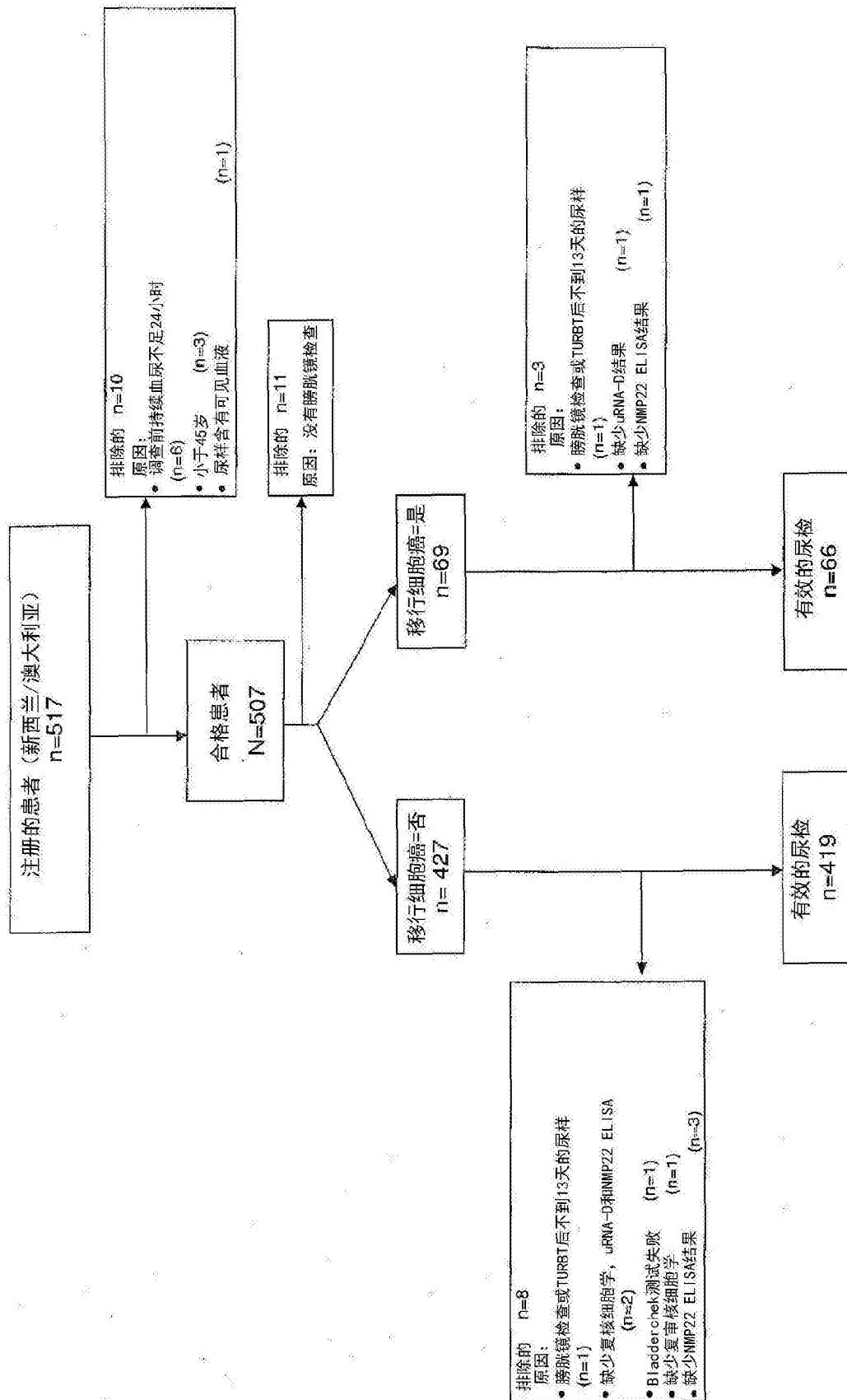


图8

	无具体诊断 (N=164)	非恶性疾病 (N=255)	移行细胞癌 (N=66)	总计 (N=485)
国家, n (%)				
新西兰	147 (89.6)	239 (93.7)	64 (97.0)	450 (92.8)
澳大利亚	17 (10.4)	16 (6.27)	2 (3.0)	35 (7.2)
年龄, 中位数 (IQ范围)	64 (55, 75)	71 (62, 78)	71 (63, 80)	69 (59, 77)
性别, n (%)				
男性	112 (68.3)	216 (84.7)	61 (92.4)	389 (80.2)
女性	52 (31.7)	39 (15.3)	5 (7.5)	96 (19.8)
种族, n (%)				
欧洲人	133 (81.1)	228 (89.4)	62 (93.9)	423 (87.2)
毛利人	15 (9.1)	17 (6.7)	1 (1.5)	33 (6.8)
其他	16 (9.8)	10 (3.9)	3 (4.5)	29 (6.0)
血尿, n (%)				
血尿年限				
<6个月	131 (79.9)	198 (77.6)	57 (86.4)	386 (79.6)
6个月-1年	11 (6.7)	25 (9.8)	6 (9.1)	42 (8.7)
1-2年	11 (6.7)	18 (7.1)	3 (4.5)	32 (6.6)
3-5年	5 (3.0)	5 (2.0)	0 (0.0)	10 (2.1)
>5年	6 (3.7)	9 (3.5)	0 (0.0)	15 (3.1)
上一次发作结束时间				
1-7天	9 (5.5)	26 (10.2)	20 (30.3)	55 (11.3)
8-28天	14 (8.5)	27 (10.6)	13 (19.7)	54 (11.1)
1-2个月	60 (36.6)	86 (33.7)	14 (21.2)	160 (33.0)
3-6个月	70 (42.7)	100 (39.2)	19 (28.8)	189 (39.0)
>6个月	11 (6.7)	15 (5.9)	0 (0.0)	26 (5.4)
不知道	0 (0.0)	1 (0.4)	0 (0.0)	1 (0.2)
频率				
≤ 1次每天	104 (63.4)	160 (62.7)	30 (45.5)	294 (60.6)
> 1次每天	59 (36.0)	95 (37.3)	36 (54.5)	190 (39.2)
不知道	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.2)
吸烟状况, n (%)				
现在吸烟	34 (20.7)	28 (11.0)	14 (21.2)	76 (15.7)
以前吸烟	66 (40.2)	114 (44.7)	35 (53.0)	215 (44.3)
从不吸烟	64 (39.0)	113 (44.3)	17 (25.8)	194 (40.0)

图9

测试	总灵敏度	总特异性
uRNA-D	62.1% (49.3, 73.8)	85.2% (固定的)
NMP22 Elisa	50.0% (37.4, 62.6)	88.0% (84.6, 91.0)
NMP22 BladderChek	37.9% (26.2, 50.7)	96.4% (94.2, 98.0)
细胞学 *	56.1% (43.3, 68.3)	94.5 (91.9, 96.5)
五基因 {包括IL8BRb} †	81.8%	85.1% (固定的)
五基因 {包括IL8BRb} † (90%特异性)	72.7%	89.9% (固定的)

*包括在确定一些患者的移行细胞癌中
†决定于研究数据，因此没有显示置信区间

图10

图11a

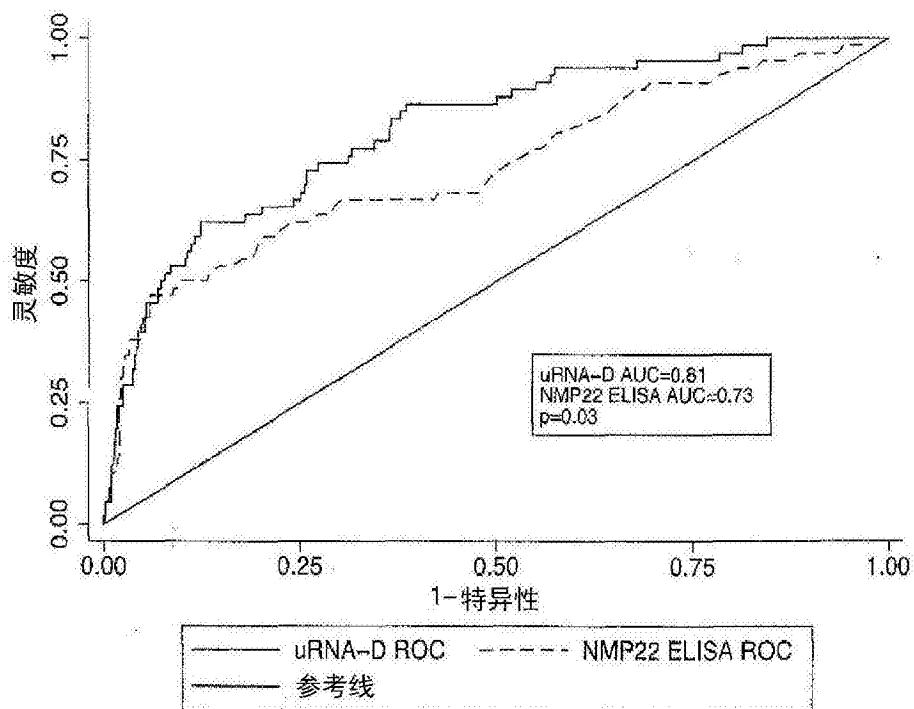


图11b

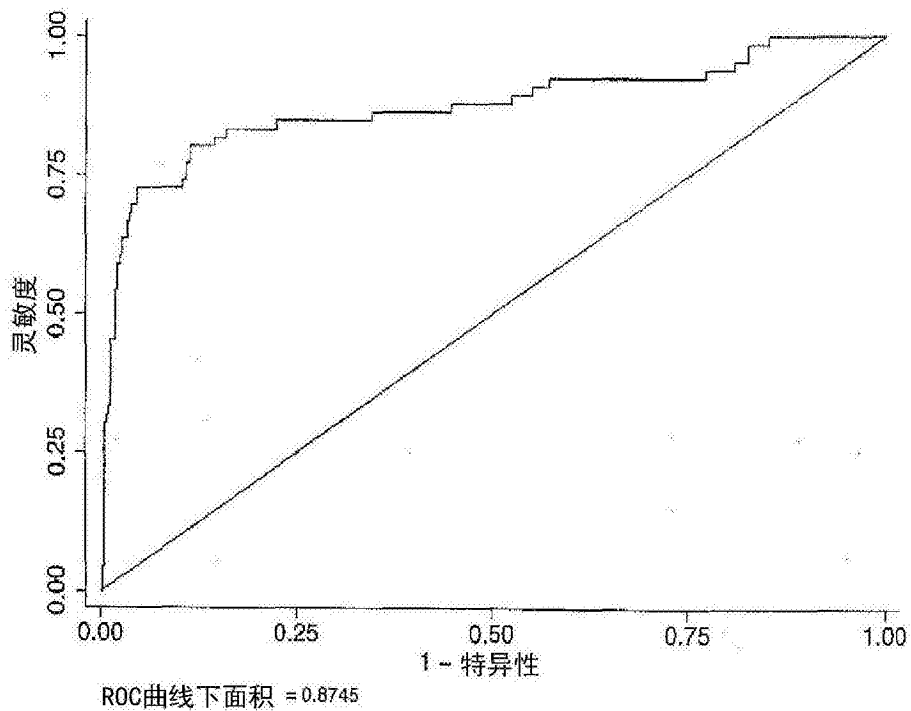


图11

	细胞学	NMP22 Elisa	BladderChek	uRNA-D	五个标志物
分期 (n=66)	n, % (95% CI)	n, % (95% CI)	n, % (95% CI)	n, % (95% CI)	n, %
Tis (n=2)	2, 100% (16, 100)	0, 0% (0, 84)	0, 0% (0, 84)	1, 50% (1, 99)	2, 100%
Ta (n=37)	13, 35% (20, 53)	13, 35% (20, 53)	14, 38% (22, 55)	15, 41% (25, 58)	25, 68%
T1 (n=16)	11, 69% (41, 89)	12, 75% (48, 93)	8, 50% (25, 75)	15, 94 (70, 100)	16, 100%
T2 (n=9)	9, 100% (66, 100)	6, 67% (30, 93)	2, 22% (3, 60)	9, 100 (66, 100)	9, 100%
≥T3 (n=2)	2, 100% (16, 100)	2, 100% (16, 100)	1, 50% (11, 99)	1, 50% (13, 99)	2, 100%
分级 WHO73 (n=65)					$p^*=0.019$
1 (n=3)	1, 33% (1, 91)	1, 33% (1, 91)	1, 33% (1, 91)	1, 33% (1, 91)	1, 33%
2 (n=38)	17, 44% (29, 62)	15, 39% (24, 57)	13, 34% (20, 51)	21, 55% (38, 71)	29, 76%
3 (n=24)	19, 79% (58, 93)	17, 71% (49, 87)	11, 46% (26, 67)	19, 79% (58, 93)	23, 96%
分级 WHOISUP1998 (n=65)					$p^*=0.016$
低 (n=32)	9, 28% (14, 47)	10, 31% (16, 50)	13, 41% (24, 59)	13, 41% (24, 59)	22, 69%
混合 (n=4)	4, 100% (40, 100)	3, 75% (19, 99)	1, 25% (1, 81)	4, 100 (40, 100)	4, 100%
高 (n=29)	24, 83% (64, 94)	20, 69% (49, 85)	11, 38% (21, 58)	24, 83% (64, 94)	28, 97%
肿瘤部位 (n=66)					$p^*=0.012$
膀胱 (n=62)	35, 56% (43, 69)	30, 48% (35, 61)	25, 40% (28, 54)	39, 63% (50, 75)	50, 81%
上, 尿道 (n=4)	2, 50% (7, 93)	3, 75% (19, 99)	0, 0% (0, 60)	2, 50% (7, 93)	4, 100%
多重性 (n=66)					$p^*=0.33$
单病灶 (n=52)	27, 52% (38, 66)	23, 44% (30, 59)	17, 33% (20, 47)	30, 58% (43, 71)	41, 79%
多病灶 (n=13)	10, 77% (46, 95)	10, 77% (46, 95)	8, 62% (32, 86)	11, 85% (55, 98)	12, 92%
显微可见血尿 (n=65)					$p^*=0.43$
是 (n=43)	27, 63% (47, 77)	27, 63% (47, 77)	20, 47% (31, 62)	31, 72% (56, 85)	41, 95%
否 (n=22)	9, 41% (21, 64)	5, 23% (8, 45)	4, 18% (5, 40)	9, 41% (21, 64)	12, 55%
肌酸酐 (n=55)					$p^*<0.0005$
0-4.3 (n=11)	5, 45 (17, 77)	3, 27% (6, 61)	4, 36% (11, 69)	4, 36% (11, 69)	7, 64%
4.3-7.1 (n=14)	9, 64% (35, 87)	6, 43% (18, 71)	4, 29% (8, 58)	8, 57% (29, 82)	12, 86%
7.1-11.2 (n=22)	11, 50% (28, 72)	14, 64% (41, 83)	11, 50% (28, 72)	16, 73% (50, 89)	18, 82%
≥11.2 (n=8)	6, 75% (35, 97)	3, 38% (9, 76)	2, 25% (3, 65)	6, 75% (35, 97)	8, 100%
性别 (n=66)					$p^*=0.29$
男性 (n=61)	35, 57% (44, 70)	29, 48% (35, 61)	23, 38% (26, 51)	37, 61% (47, 73)	49, 80%
女性 (n=5)	2, 40% (5, 85)	4, 80% (28, 99)	2, 40% (5, 85)	4, 80% (28, 99)	5, 100%
					$p^*=0.58$

*每个TCC特点和Cxbladder测试结果之间的关联的Fisher精确检验的P值

图12

	细胞学	NMP22 Elisa	NMP22 BladderChek	uRNA-D*	五个标志物*
	n, % 95% CI)	n, % 95% CI)	n, % 95% CI)	n, %	n, %
诊断					
无诊断 (n=164)					
	154, 94% (89, 97)	144, 88% (82, 92)	160, 98% (94, 99)	144, 88%	144, 88%
非恶性诊断					
良性前列腺肥大/前列腺炎 (n=130)					
	123, 95% (89, 98)	117, 90% (84, 95)	127, 98% (93, 100)	113, 87%	109/128=85%
膀胱炎/泌尿道的感染或炎症 (n=39)					
	36, 92% (79, 98)	34, 87% (73, 96)	34, 87% (73, 96)	28, 72%	32, 82%
结石 (n=28)					
	25, 89% (72, 98)	23, 82% (63, 94)	27, 96% (82, 100)	20, 71%	19, 68%
血尿继发于华法林 (n=10)					
	10, 100% (69, 100)	9, 90% (55, 100)	10, 100% (69, 100)	8, 80%	8, 80%
其他泌尿系肿瘤 (n=5)					
	5, 100% (48, 100)	4, 80% (28, 99)	5, 100% (48, 100)	5, 100%	4, 80%
显微镜可见血尿 (n=417)					$p^{\dagger}=0.12$
是 n=99	87, 88% (80, 94)	81, 82% (73, 89)	91, 92% (85, 96)	74, 75%	73/97=75%
否 n=318	307, 97% (94, 98)	287, 90% (86, 93)	311, 98% (96, 99)	281, 88%	279/317=88%
肌酐酞 (n=386)					$p^{\dagger}=0.002$
0-4.3, n=97	95, 98% (93, 100)	90, 93% (86, 97)	95, 98% (93, 100)	87, 90%	83/96=86%
4.3-7.1, n=94	92, 98% (93, 100)	81, 86% (78, 92)	90, 96% (89, 99)	81, 86%	84/93=90%
7.1-11.2, n=92	79, 86% (77, 92)	83, 90% (82, 95)	90, 98% (92, 100)	76, 83%	76/92=83%
≥11.2, n=103	101, 98% (93, 100)	91, 88% (81, 94)	101, 98% (93, 100)	88, 85%	82/102=80%
性别 (n=419)					$p^{\dagger}=0.32$
男性 n=328	311, 95% (92, 97)	290, 88% (84, 92)	316, 96% (94, 98)	291, 89%	280/325=86%
女性 n=91	85, 93% (86, 98)	79, 87% (78, 93)	88, 97% (91, 99)	65, 71%	74/91=81%
					$p^{\dagger}=0.25$

*总特异性固定为85%

†每个特点(诊断, 显微镜可见血尿, 肌酐酞, 性别)和Cxb ladder测试结果之间的关联的卡方检验的P值

图13

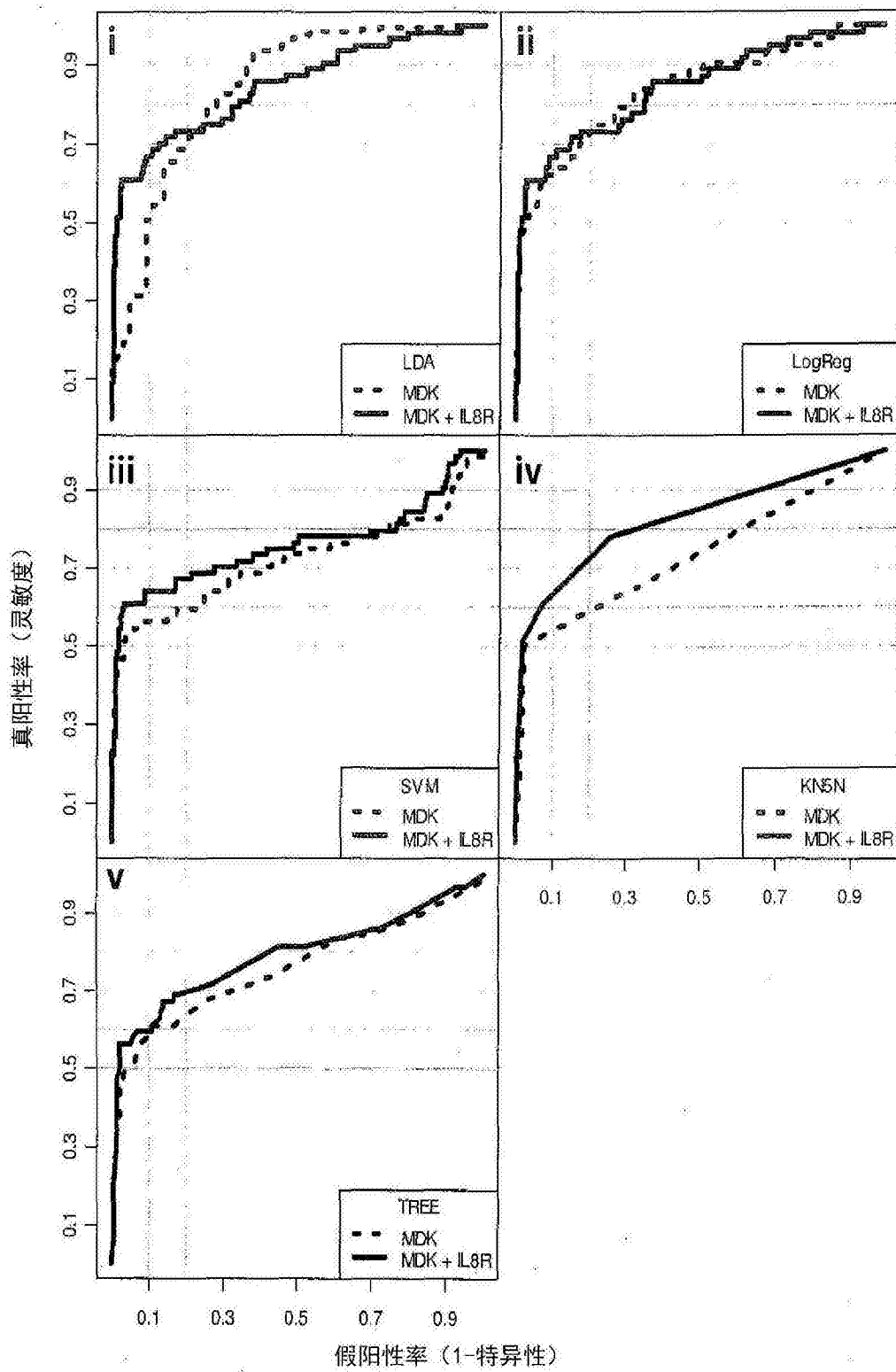
基因：MDK ($\pm$ IL8R)

图14a

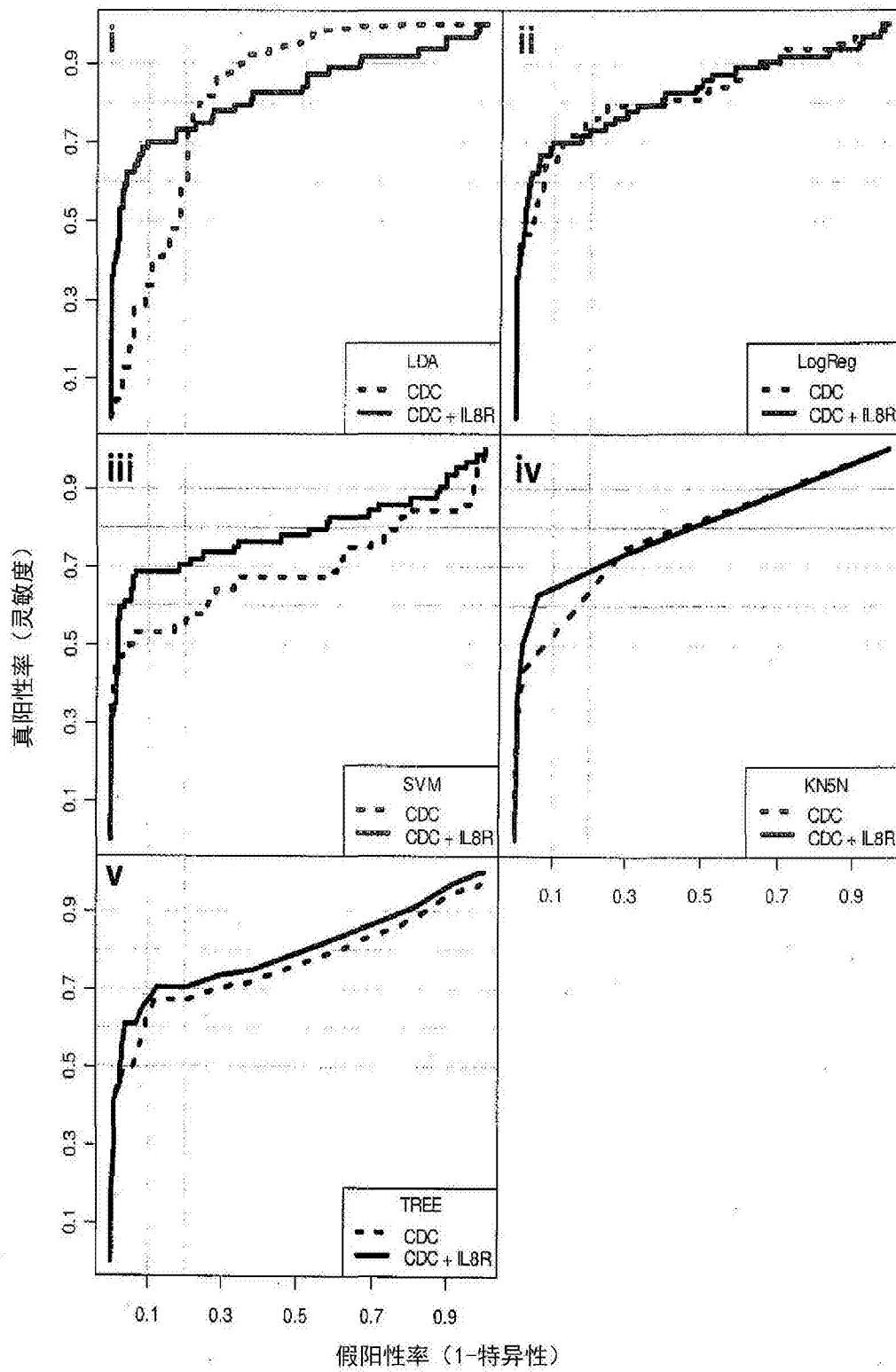
基因：CDC ($\pm$ IL8R)

图14b

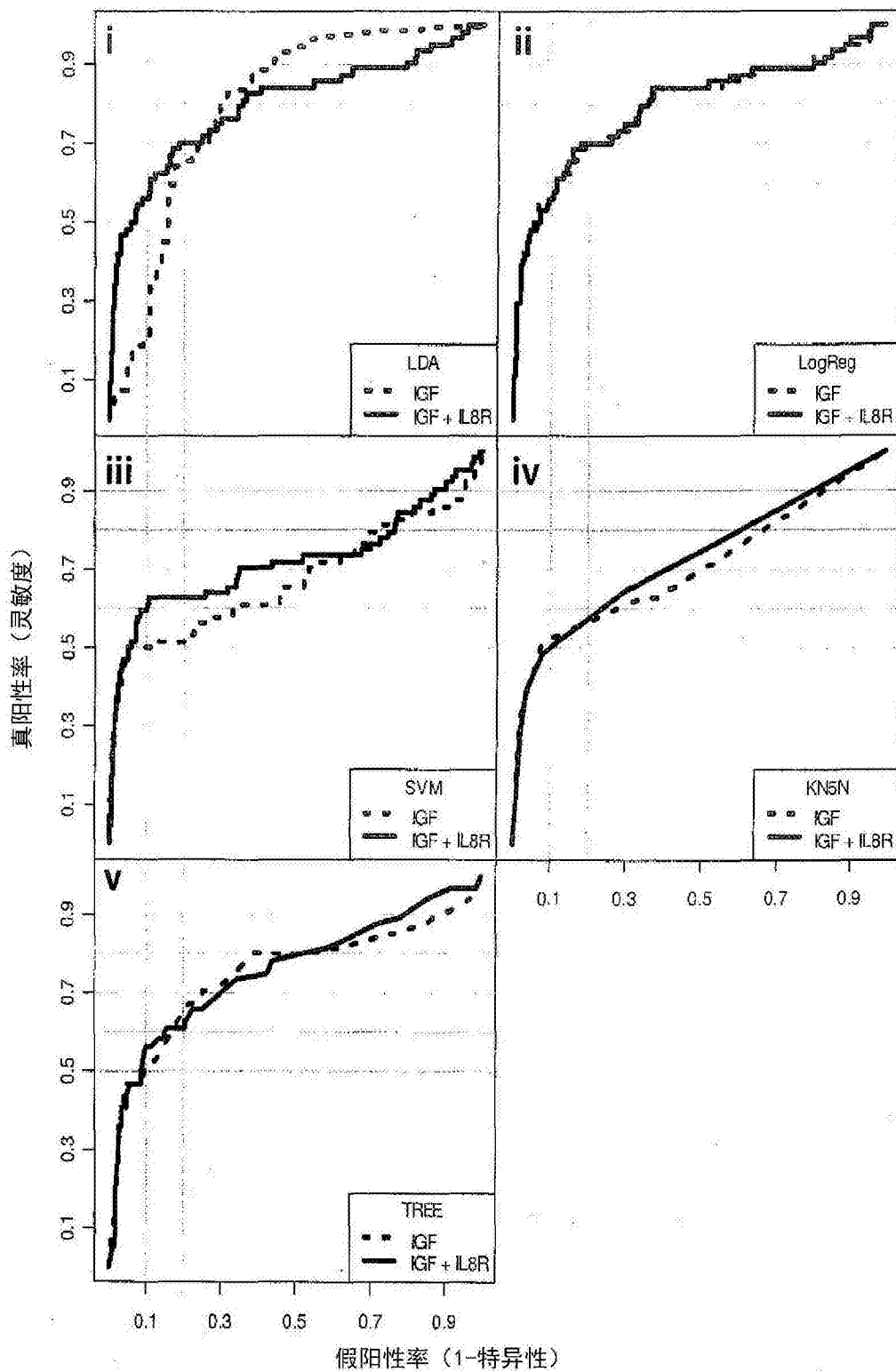
基因 : IGF ($\pm$ IL8R)

图14c

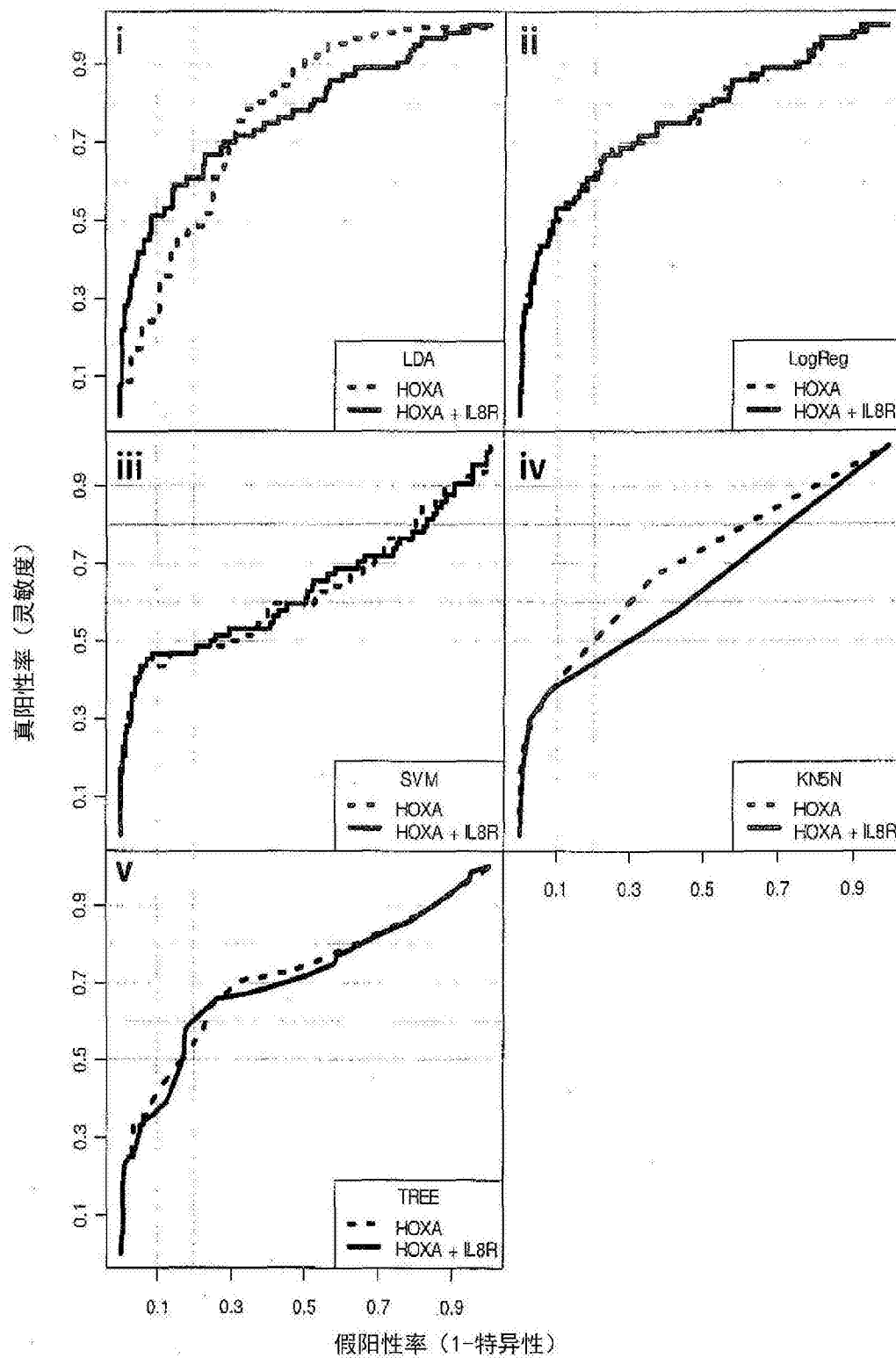
基因：HOXA ($\pm$ IL8R)

图14d

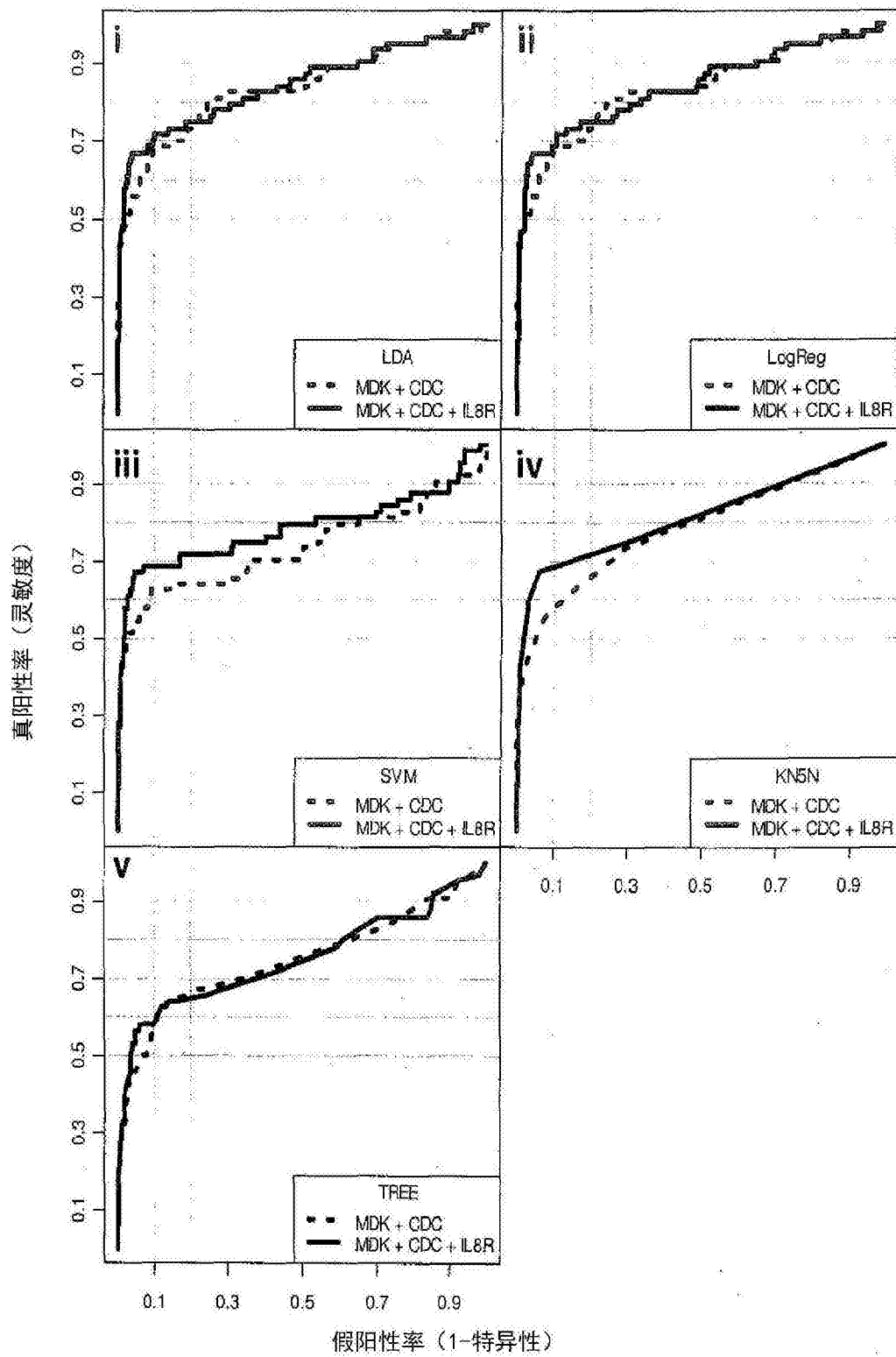
基因：MDK + CDC ($\pm$ IL8R)

图14e

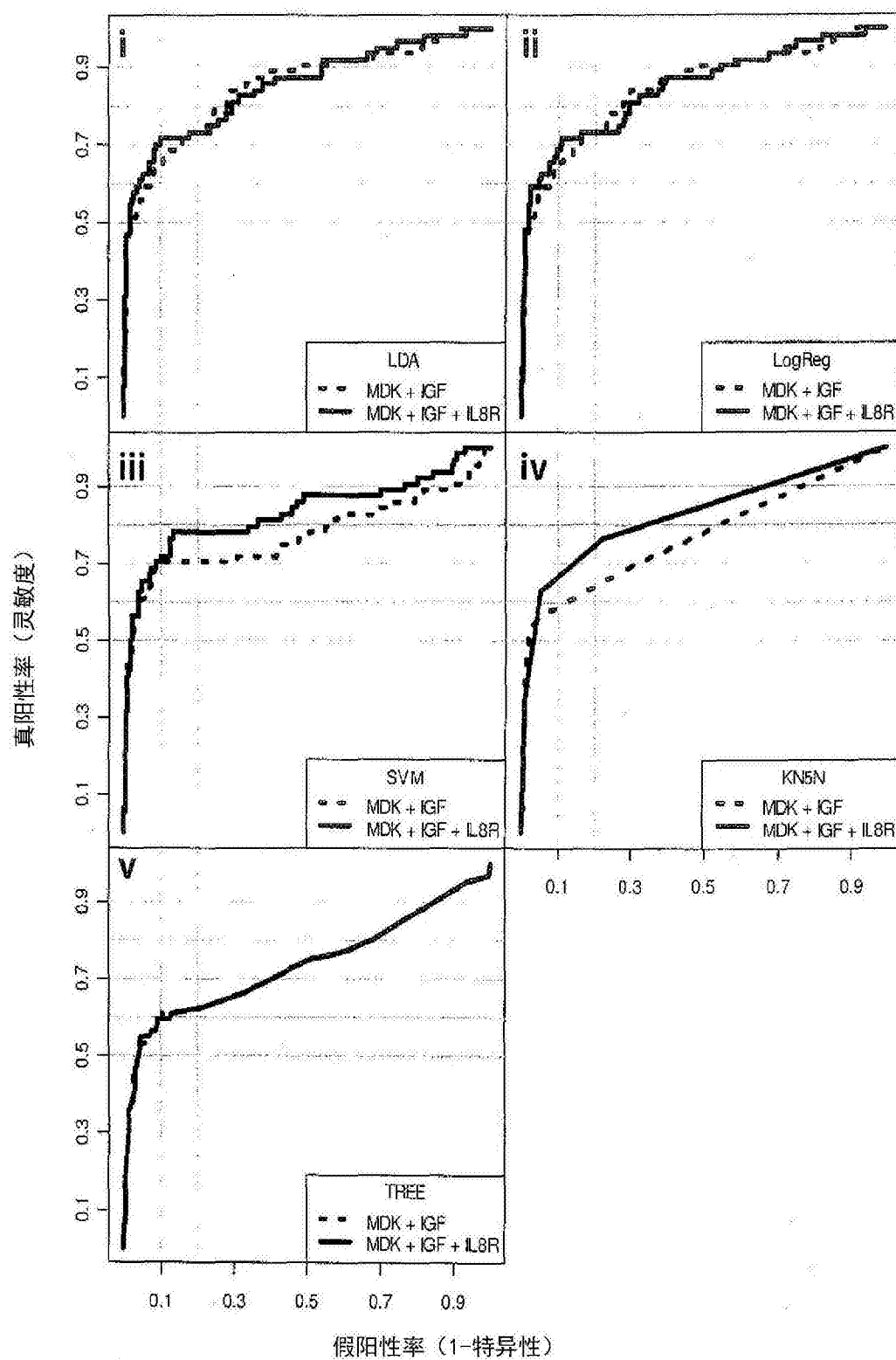
基因：MDK + IGF ($\pm$ IL8R)

图14f

基因：MDK + HOXA ($\pm$ IL8R)

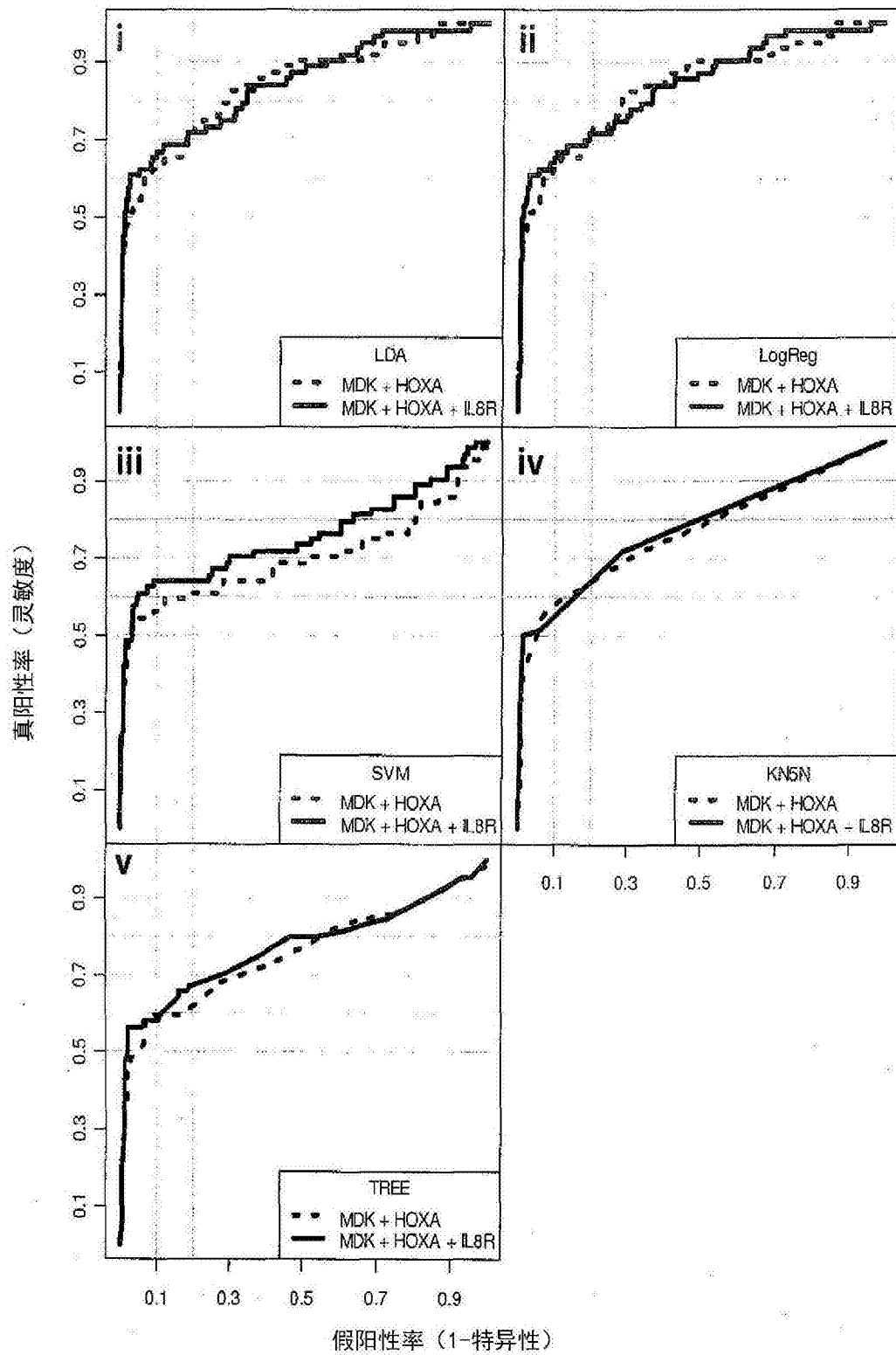


图14g

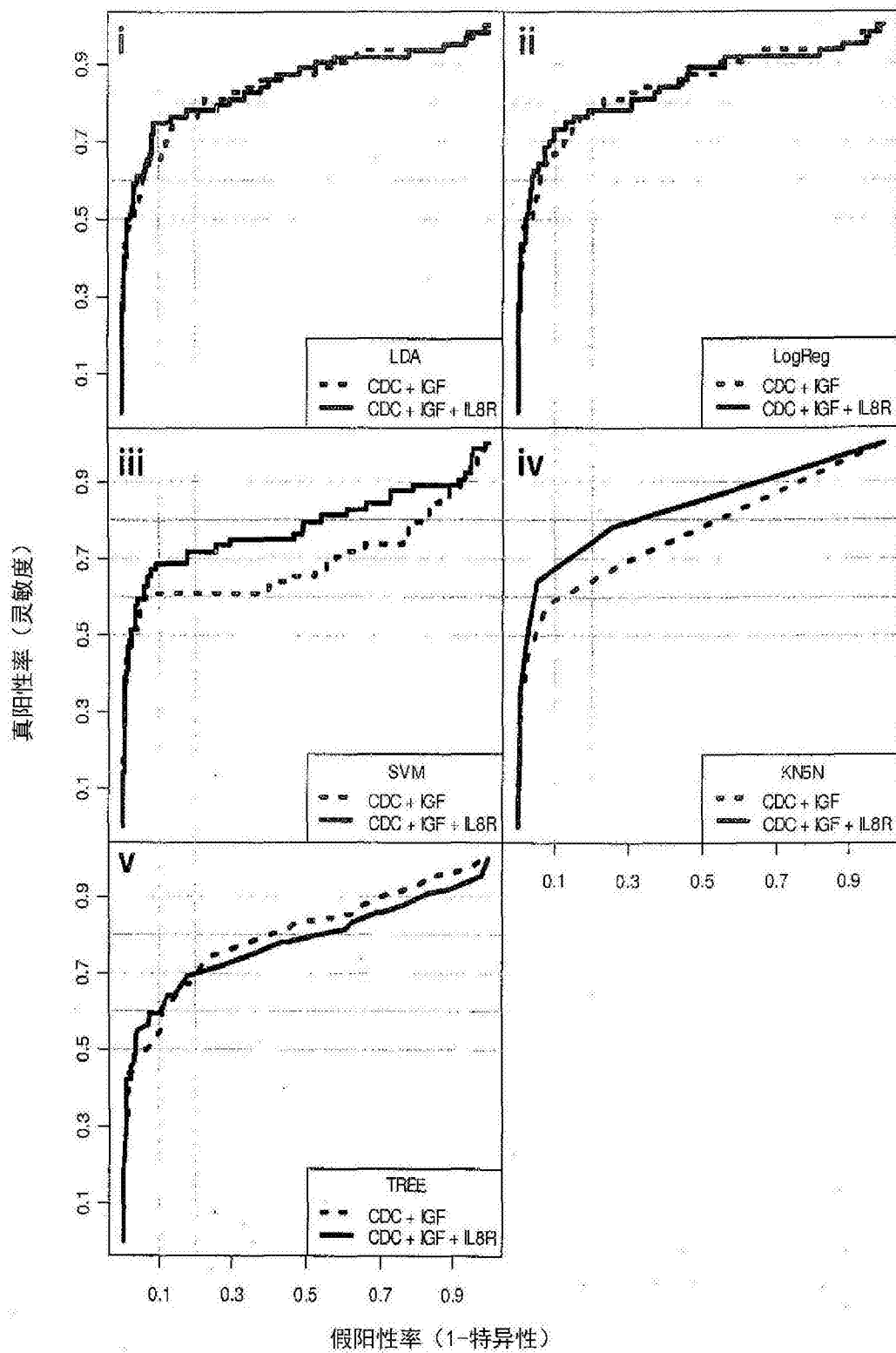
基因：CDC + IGF ($\pm$ IL8R)

图14h

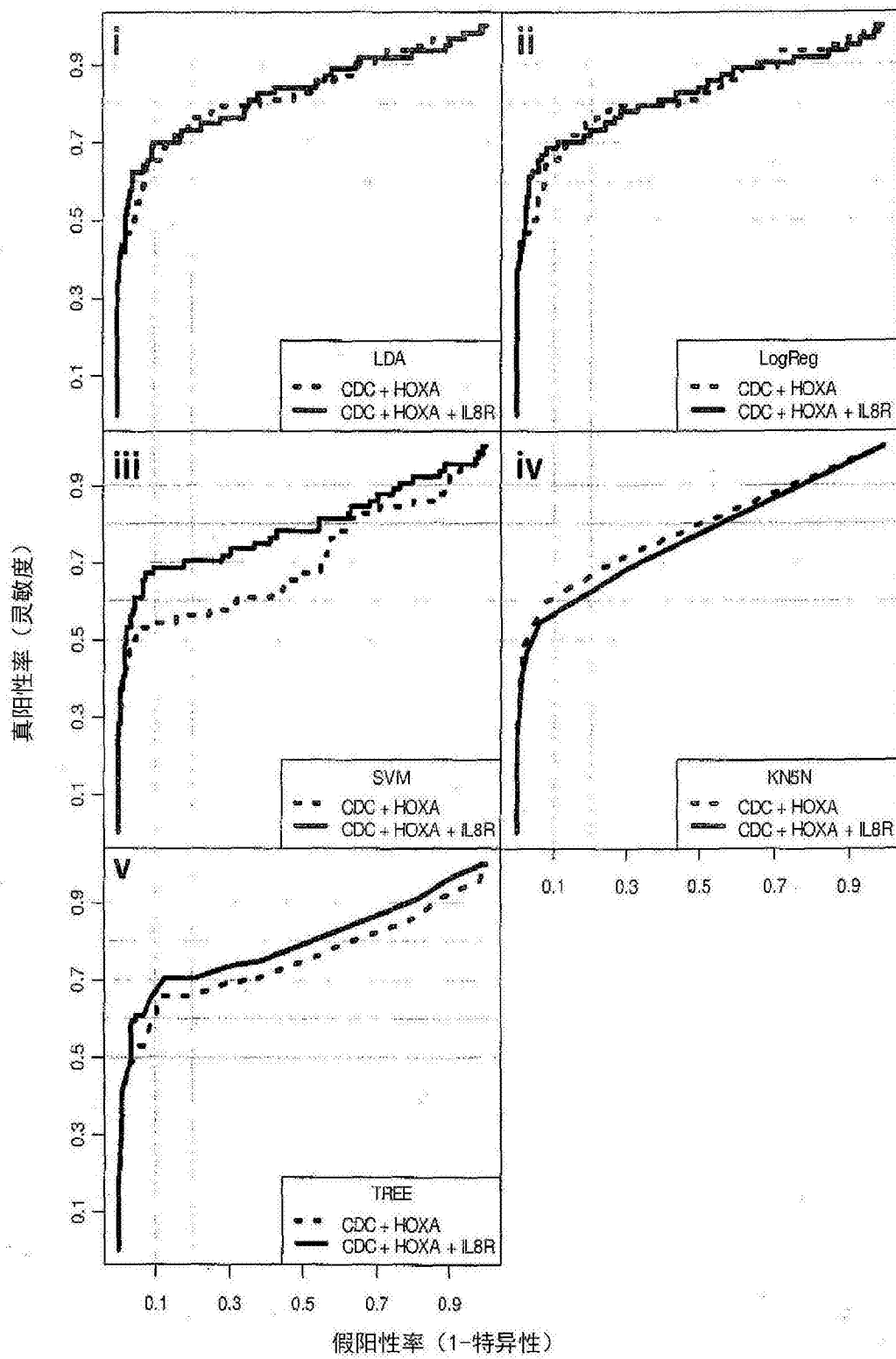
基因: CDC + HOXA ($\pm$ IL8R)

图14i

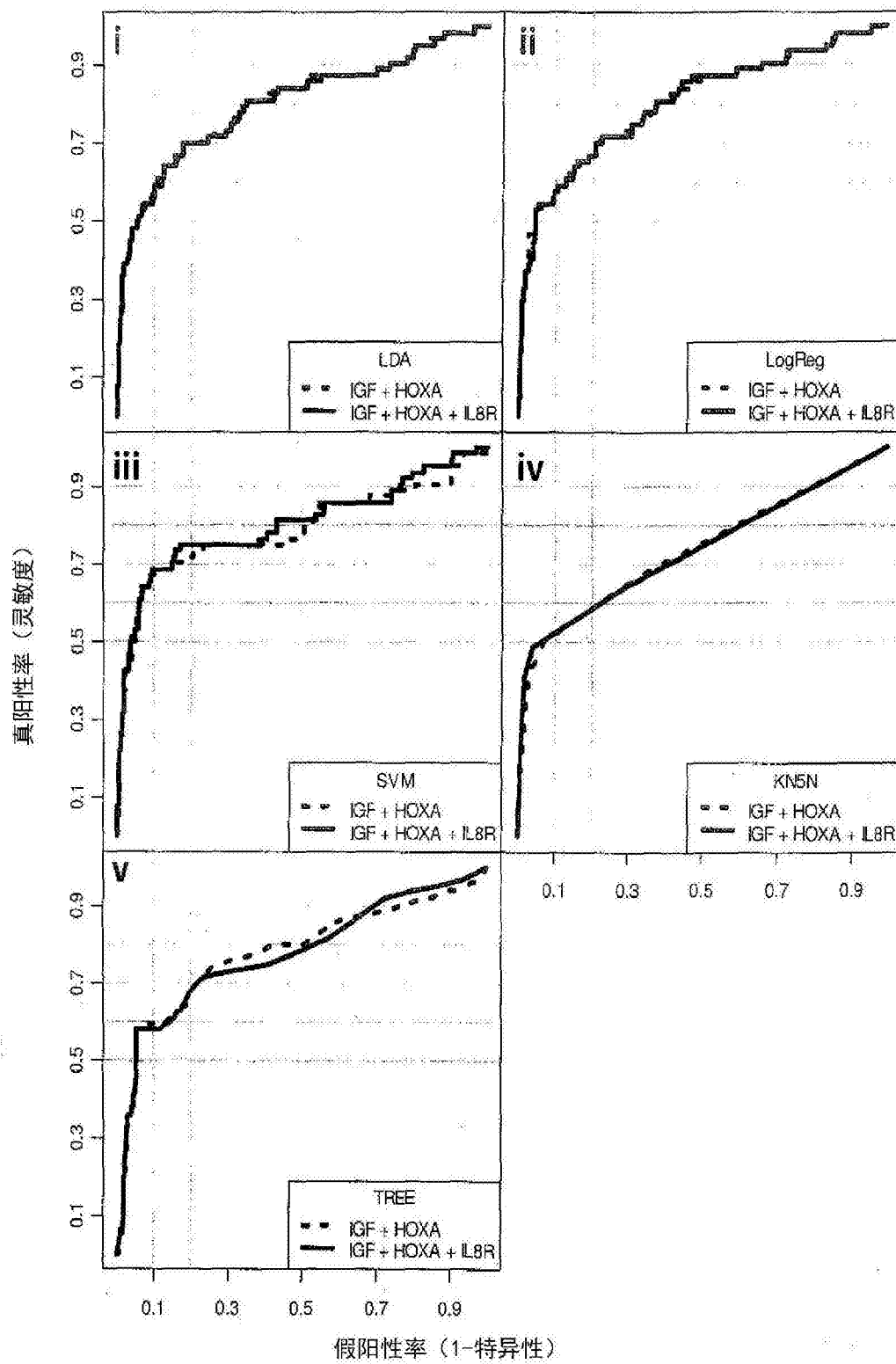
基因：IGF + HOXA ($\pm$ IL8R)

图14j

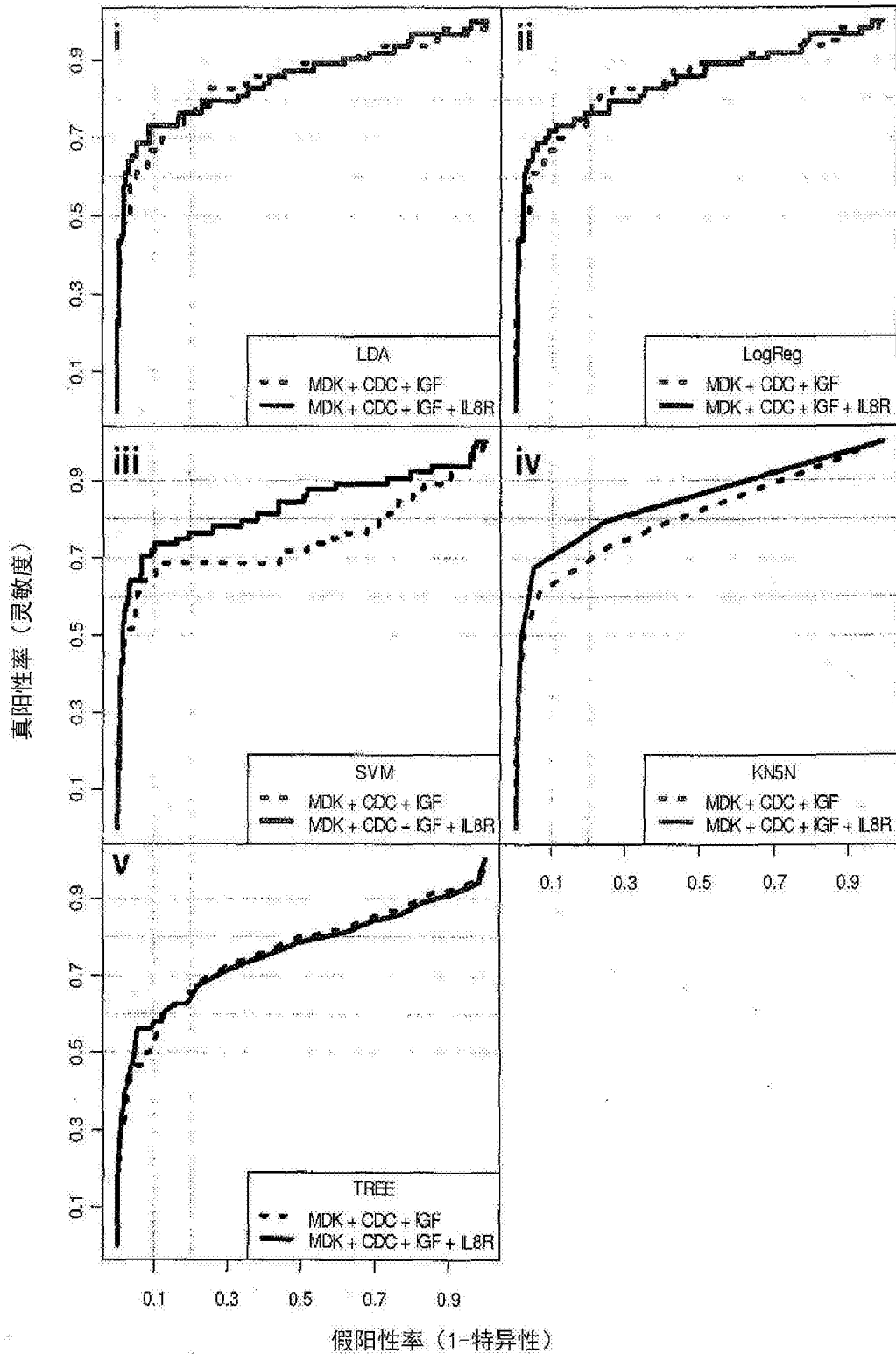
基因：MDK + CDC + IGF ($\pm$ IL8R)

图14k

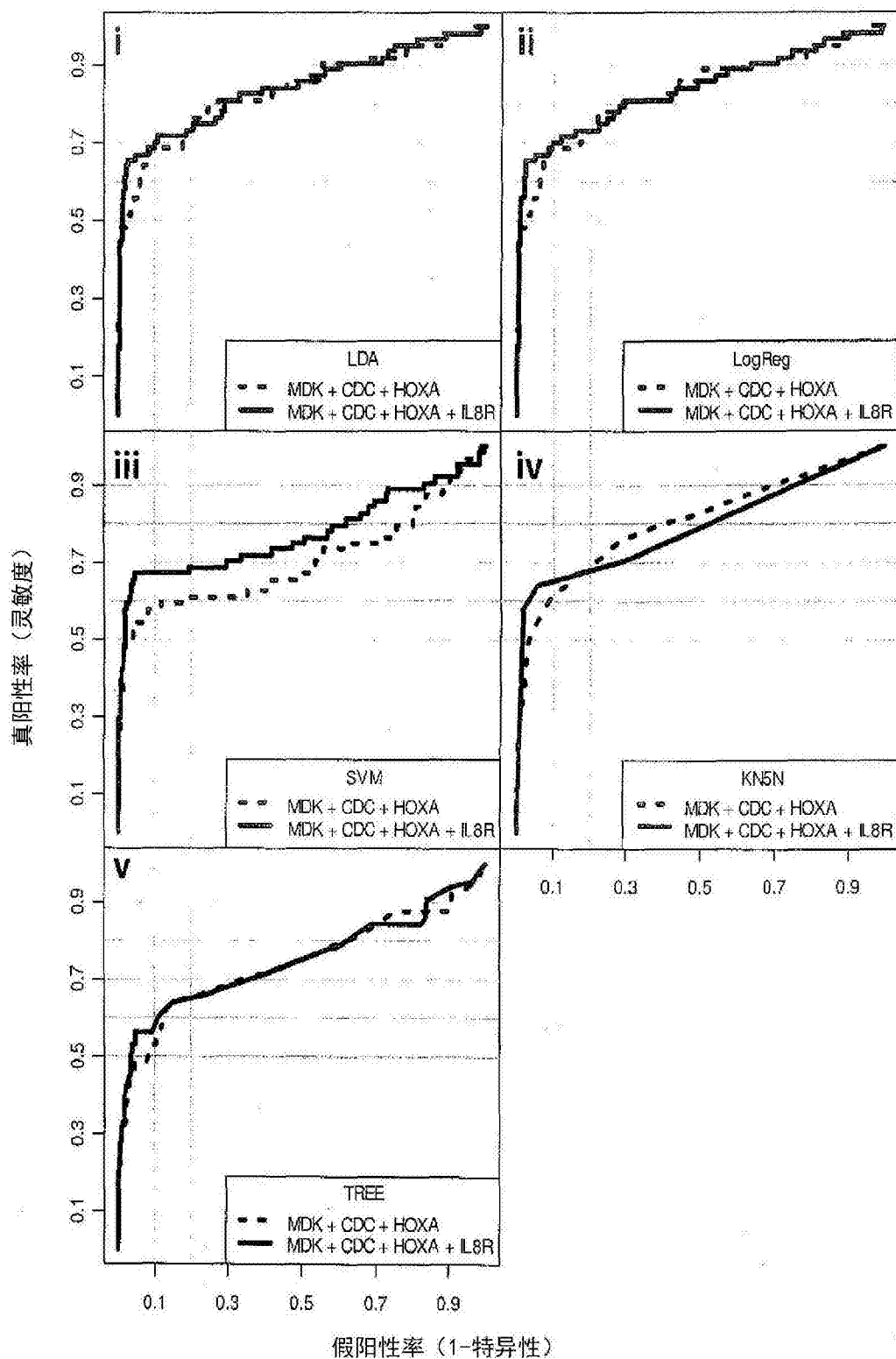
基因：MDK + CDC + HOXA ($\pm$ IL8R)

图141

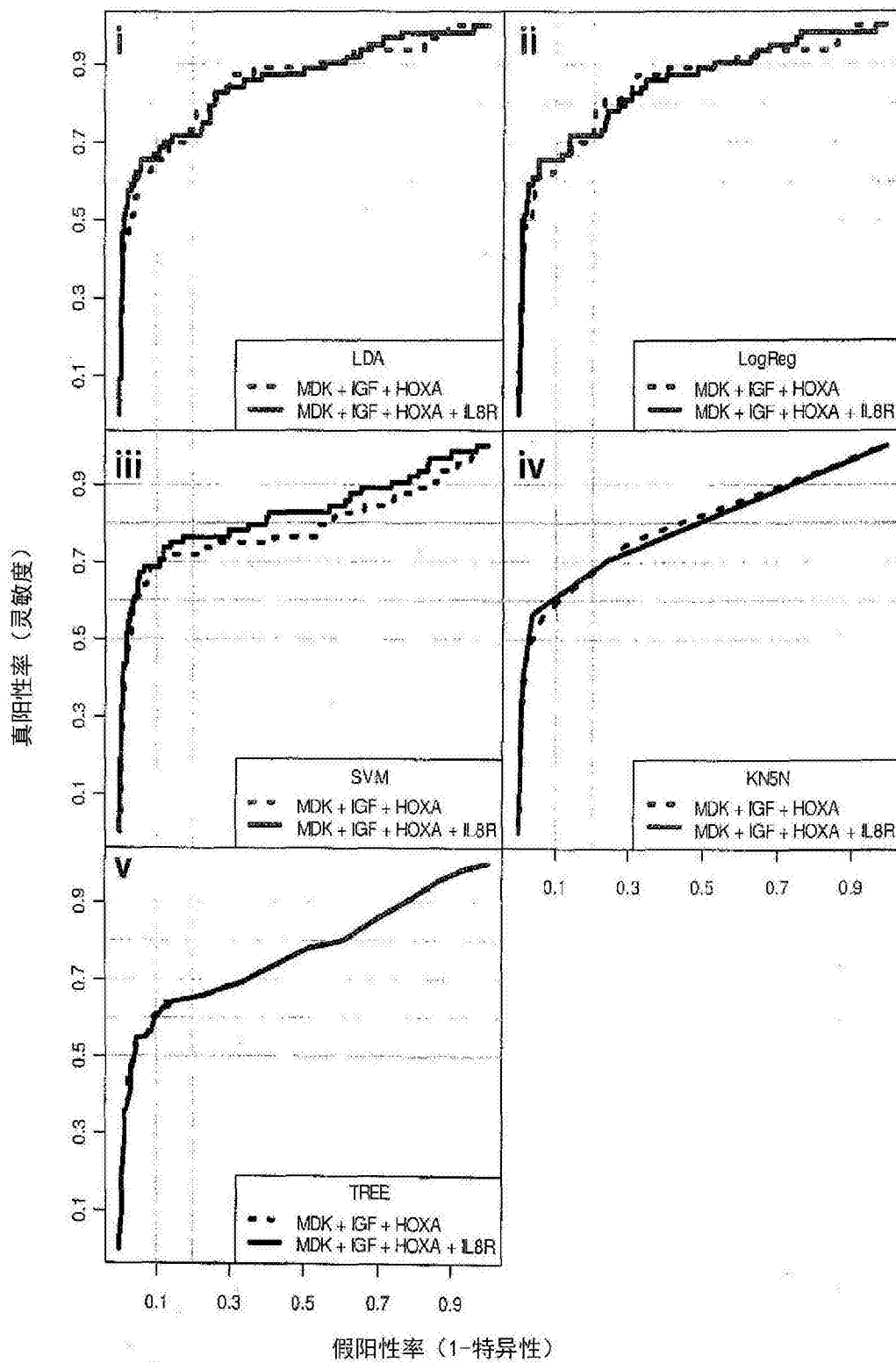
基因：MDK + IGF + HOXA ($\pm$ IL8R)

图14m

基因：CDC + IGF + HOXA ($\pm$ IL8R)

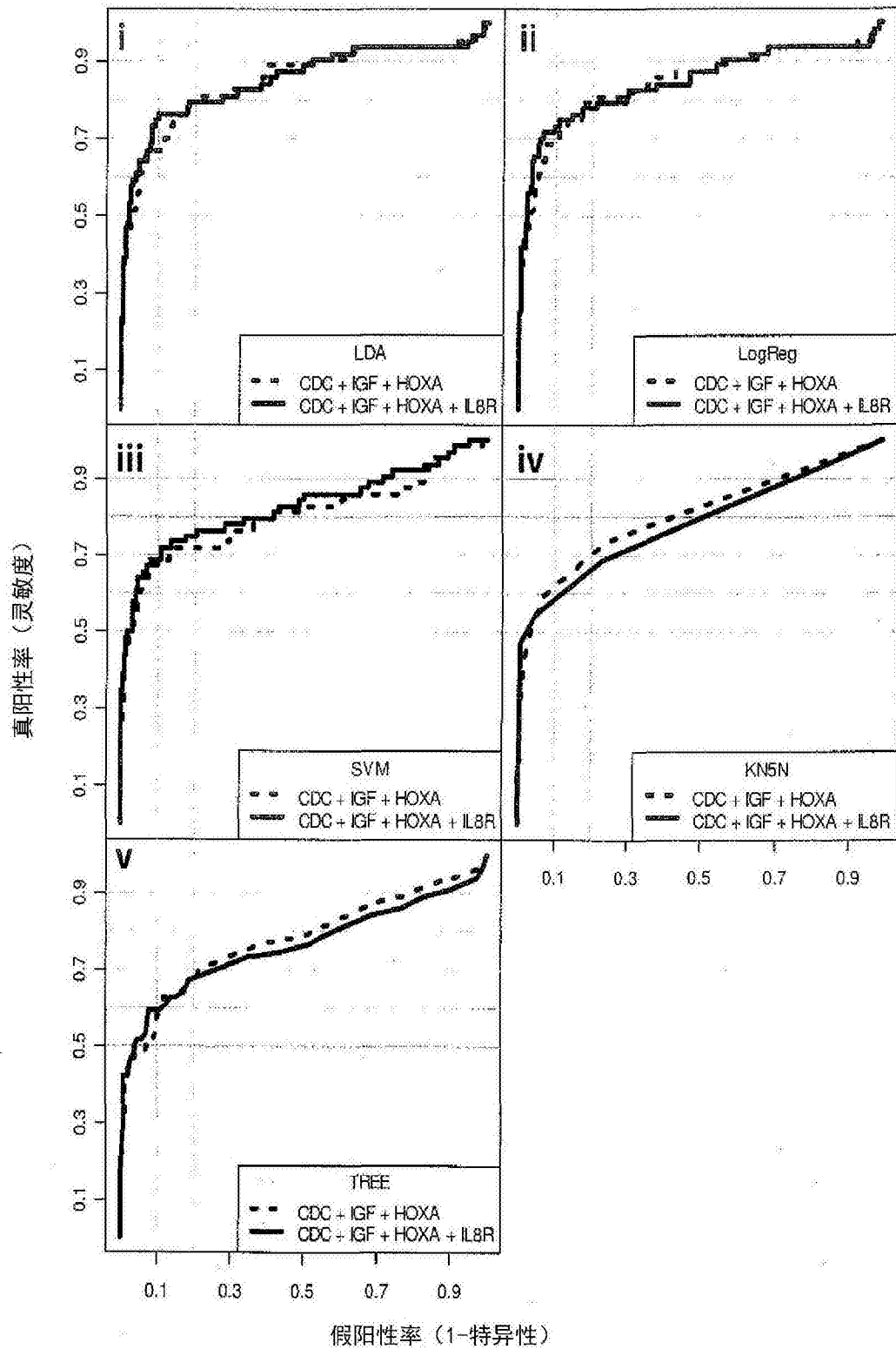


图14n

基因：MDK + CDC + IGF + HOXA ($\pm$ IL8R)

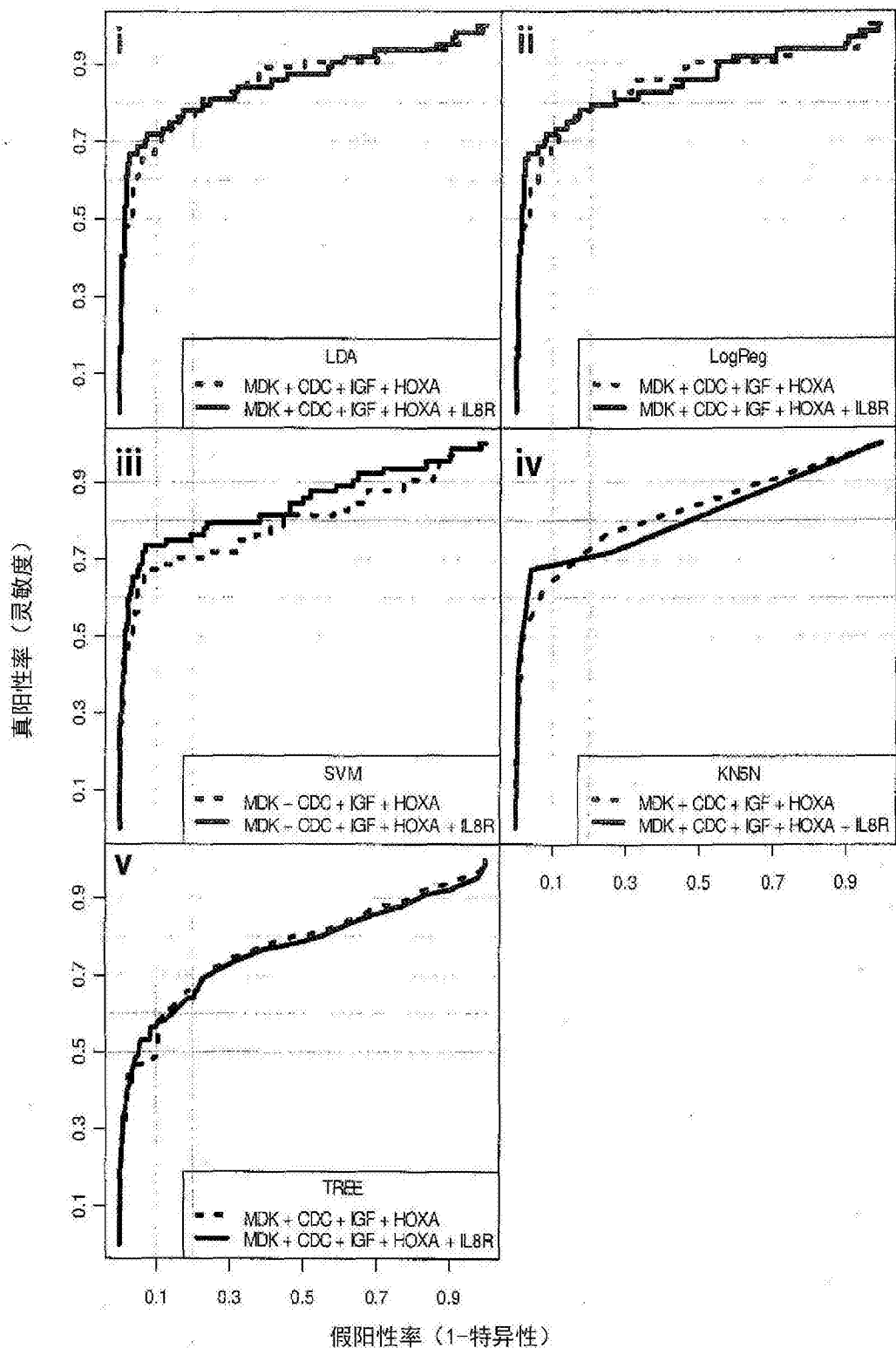


图14o

模型	LDA		LogReg		SVM		KNSN		TREE	
	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
MDK	0.087	0.127	0.119	0.127	0.107	0.121	0.100	0.120	0.107	0.117
CDC2	0.062	0.129	0.124	0.129	0.102	0.127	0.102	0.121	0.114	0.124
IGFBP5	0.060	0.108	0.109	0.108	0.094	0.107	0.093	0.090	0.093	0.096
HOXA13	0.058	0.094	0.094	0.094	0.082	0.084	0.075	0.071	0.078	0.075
MDK + CDC2	0.124	0.134	0.124	0.133	0.115	0.131	0.109	0.128	0.107	0.111
MDK + IGFBP5	0.122	0.129	0.121	0.129	0.125	0.133	0.111	0.122	0.107	0.107
MDK + HOXA13	0.119	0.126	0.118	0.125	0.108	0.118	0.107	0.107	0.107	0.114
CDC2 + IGFBP5	0.128	0.134	0.127	0.133	0.114	0.125	0.109	0.125	0.109	0.115
CDC + HOXA13	0.122	0.128	0.123	0.129	0.103	0.125	0.116	0.108	0.113	0.123
IGF + HOXA13	0.109	0.109	0.108	0.107	0.119	0.121	0.096	0.098	0.102	0.101
MDK + CDC2 + IGFBP5	0.127	0.135	0.127	0.135	0.123	0.134	0.119	0.132	0.103	0.108
MDK + CDC2 + HOXA13	0.125	0.133	0.125	0.134	0.109	0.127	0.113	0.122	0.104	0.111
MDK + IGFBP5 + HOXA13	0.122	0.128	0.121	0.127	0.126	0.132	0.111	0.115	0.110	0.109
CDC2 + IGFBP5 + HOXA13	0.129	0.136	0.131	0.136	0.126	0.131	0.117	0.114	0.108	0.111
MDK + CDC2 + IGFBP5 + HOXA13	0.130	0.137	0.130	0.138	0.124	0.137	0.122	0.129	0.103	0.106

假阳性率为0.00到0.20的ROC曲线下面积 (AUC)

图15a

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.040	0.008	0.014	0.019	0.010
CDC2	0.067	0.005	0.025	0.018	0.009
IGFBP5	0.048	-0.001	0.013	-0.003	0.002
HOXA13	0.036	0.000	0.001	-0.004	-0.003
MDK + CDC2	0.009	0.009	0.015	0.019	0.005
MDK + IGFBP5	0.007	0.007	0.008	0.011	-0.000
MDK + HOXA13	0.007	0.007	0.010	0.000	0.007
CDC2 + IGFBP5	0.007	0.006	0.012	0.016	0.006
CDC + HOXA13	0.006	0.006	0.022	-0.007	0.010
IGF + HOXA13	0.000	-0.001	0.002	0.002	-0.000
MDK + CDC2 + IGFBP5	0.009	0.008	0.012	0.013	0.004
MDK + CDC2 + HOXA13	0.008	0.008	0.018	0.009	0.006
MDK + IGFBP5 + HOXA13	0.006	0.006	0.006	0.004	-0.001
CDC2 + IGFBP5 + HOXA13	0.006	0.006	0.005	-0.003	0.004
MDK+CDC2+IGFBP5+HOXA13	0.007	0.007	0.013	0.007	0.002

AUC的增加 (在假阳性率为0.0至0.2上)

图15b

a	模型	LDA			LogReg			SVM			KN5N			TREE		
		-IL8Rb	+IL8Rb	-IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb
	MDK	0.721	0.734	0.734	0.734	0.734	0.734	0.594	0.672	0.531	0.781	0.625	0.688			
	CDC2	0.753	0.734	0.734	0.750	0.734	0.734	0.562	0.703	0.750	0.734	0.672	0.703			
	IGFBP5	0.657	0.703	0.688	0.688	0.703	0.703	0.516	0.625	0.516	0.656	0.656	0.609			
	HOXA13	0.486	0.609	0.625	0.625	0.609	0.609	0.484	0.469	0.406	0.375	0.547	0.578			
	MDK + CDC2	0.734	0.750	0.734	0.734	0.750	0.734	0.641	0.719	0.734	0.750	0.672	0.641			
	MDK + IGFBP5	0.734	0.734	0.734	0.734	0.734	0.734	0.703	0.781	0.688	0.766	0.625	0.625			
	MDK + HOXA13	0.719	0.719	0.734	0.734	0.719	0.719	0.609	0.641	0.703	0.719	0.609	0.672			
	CDC2 + IGFBP5	0.766	0.781	0.766	0.766	0.781	0.781	0.609	0.719	0.688	0.781	0.672	0.688			
	CDC + HOXA13	0.750	0.734	0.750	0.750	0.734	0.734	0.562	0.703	0.703	0.688	0.656	0.703			
	IGF + HOXA13	0.703	0.703	0.672	0.672	0.672	0.672	0.719	0.750	0.500	0.641	0.688	0.672			
	MDK + CDC2 + IGFBP5	0.766	0.766	0.766	0.766	0.766	0.766	0.688	0.766	0.734	0.797	0.656	0.625			
	MDK + CDC2 + HOXA13	0.734	0.734	0.734	0.734	0.734	0.734	0.609	0.688	0.766	0.703	0.656	0.656			
	MDK + IGFBP5 + HOXA13	0.734	0.719	0.734	0.734	0.719	0.719	0.719	0.766	0.750	0.703	0.656	0.656			
	CDC2 + IGFBP5 + HOXA13	0.797	0.797	0.812	0.812	0.781	0.781	0.719	0.750	0.734	0.688	0.672	0.672			
	MDK+CDC2+IGFBP5+HOXA13	0.766	0.781	0.781	0.781	0.797	0.797	0.703	0.766	0.766	0.719	0.656	0.641			

特异性80%的灵敏度

图16a

b	模型	LDA		LogReg		SVM		KN5N		TREE	
		-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
	MDK	0.657	0.719	0.672	0.719	0.578	0.641	0.531	0.609	0.609	0.672
	CDC2	0.481	0.703	0.719	0.703	0.531	0.688	0.500	0.625	0.672	0.703
	IGFBP5	0.595	0.641	0.625	0.656	0.516	0.625	0.516	0.484	0.562	0.578
	HOXA13	0.462	0.594	0.562	0.562	0.469	0.469	0.406	0.375	0.500	0.406
	MDK + CDC2	0.688	0.734	0.688	0.734	0.641	0.688	0.562	0.672	0.641	0.641
	MDK + IGFBP5	0.703	0.719	0.703	0.719	0.703	0.781	0.562	0.766	0.609	0.609
	MDK + HOXA13	0.656	0.688	0.656	0.688	0.594	0.641	0.562	0.516	0.594	0.625
	CDC2 + IGFBP5	0.750	0.766	0.750	0.766	0.609	0.688	0.578	0.641	0.625	0.641
	CDC + HOXA13	0.719	0.703	0.719	0.703	0.547	0.688	0.594	0.547	0.656	0.703
	IGF + HOXA13	0.641	0.641	0.641	0.641	0.688	0.719	0.500	0.484	0.609	0.609
	MDK + CDC2 + IGFBP5	0.734	0.734	0.734	0.734	0.688	0.734	0.609	0.797	0.609	0.625
	MDK + CDC2 + HOXA13	0.688	0.719	0.688	0.719	0.594	0.672	0.609	0.641	0.641	0.641
	MDK + IGFBP5 + HOXA13	0.703	0.719	0.703	0.719	0.719	0.750	0.562	0.703	0.641	0.641
	CDC2 + IGFBP5 + HOXA13	0.750	0.766	0.750	0.766	0.719	0.734	0.594	0.688	0.625	0.625
	MDK+CDC2+IGFBP5+HOXA13	0.766	0.750	0.750	0.750	0.688	0.750	0.625	0.719	0.609	0.594

特异性85%的灵敏度

图16b

C	模型	LDA		LogReg		SVM		KNN		TREE	
		-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
	MDK	0.511	0.672	0.625	0.672	0.562	0.641	0.531	0.609	0.609	0.594
	CDC2	0.338	0.703	0.656	0.703	0.531	0.688	0.500	0.625	0.625	0.672
	IGFBP5	0.212	0.562	0.562	0.562	0.500	0.594	0.516	0.484	0.500	0.562
	HOXA13	0.257	0.516	0.500	0.531	0.438	0.469	0.406	0.375	0.391	0.359
	MDK + CDC2	0.688	0.703	0.688	0.688	0.625	0.688	0.562	0.672	0.562	0.578
	MDK + IGFBP5	0.641	0.719	0.641	0.688	0.703	0.719	0.562	0.625	0.578	0.594
	MDK + HOXA13	0.641	0.672	0.625	0.656	0.562	0.641	0.562	0.516	0.594	0.578
	CDC2 + IGFBP5	0.656	0.750	0.672	0.734	0.609	0.688	0.578	0.641	0.547	0.594
	CDC2 + HOXA13	0.641	0.703	0.656	0.688	0.547	0.688	0.594	0.547	0.594	0.672
	IGF + HOXA13	0.594	0.594	0.578	0.578	0.672	0.688	0.500	0.484	0.594	0.578
	MDK + CDC2 + IGFBP5	0.672	0.734	0.672	0.719	0.656	0.734	0.609	0.672	0.531	0.578
	MDK + CDC2 + HOXA13	0.688	0.688	0.688	0.703	0.578	0.672	0.609	0.641	0.531	0.562
	MDK + IGFBP5 + HOXA13	0.641	0.672	0.625	0.656	0.688	0.688	0.562	0.562	0.609	0.594
	CDC2 + IGFBP5 + HOXA13	0.672	0.766	0.703	0.734	0.672	0.688	0.594	0.547	0.578	0.594
	MDK+CDC2+IGFBP5+HOXA13	0.688	0.719	0.703	0.719	0.672	0.734	0.625	0.672	0.500	0.562

特异性90%的灵敏度

图16c

d	模型	LDA		LogReg		SVM		KNSN		TREE	
		-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
	MDK	0.304	0.609	0.347	0.609	0.531	0.609	0.500	0.609	0.500	0.562
	CDC2	0.180	0.625	0.516	0.625	0.500	0.609	0.438	0.625	0.500	0.609
	IGFBP5	0.148	0.484	0.484	0.484	0.500	0.500	0.359	0.391	0.469	0.438
	HOXA13	0.173	0.422	0.438	0.422	0.422	0.406	0.297	0.297	0.344	0.297
	MDK + CDC2	0.562	0.672	0.562	0.672	0.547	0.672	0.453	0.672	0.453	0.562
	MDK + IGFBP5	0.562	0.609	0.578	0.609	0.609	0.656	0.562	0.625	0.531	0.547
	MDK + HOXA13	0.531	0.609	0.531	0.609	0.531	0.609	0.562	0.516	0.500	0.562
	CDC2 + IGFBP5	0.562	0.609	0.562	0.625	0.562	0.594	0.453	0.641	0.500	0.547
	CDC + HOXA13	0.516	0.625	0.500	0.625	0.516	0.609	0.594	0.547	0.516	0.609
	IGF + HOXA13	0.484	0.484	0.531	0.531	0.547	0.531	0.500	0.484	0.406	0.406
	MDK + CDC2 + IGFBP5	0.594	0.656	0.609	0.672	0.578	0.641	0.609	0.672	0.469	0.500
	MDK + CDC2 + HOXA13	0.562	0.672	0.562	0.672	0.547	0.672	0.516	0.641	0.484	0.562
	MDK + IGFBP5 + HOXA13	0.609	0.625	0.609	0.609	0.609	0.636	0.562	0.562	0.531	0.547
	CDC2 + IGFBP5 + HOXA13	0.609	0.641	0.578	0.636	0.594	0.641	0.594	0.547	0.484	0.516
	MDK+CDC2+IGFBP5+HOXA13	0.609	0.688	0.594	0.672	0.625	0.672	0.625	0.672	0.469	0.484

特异性95%的灵敏度

图16d

e	模型	LDA		LogReg		SVM		KNSN		TREE	
		-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
	MDK	0.163	0.516	0.484	0.516	0.469	0.547	0.312	0.516	0.359	0.562
	CDC2	0.047	0.453	0.469	0.469	0.453	0.562	0.438	0.500	0.469	0.422
	IGFBP5	0.054	0.391	0.391	0.391	0.359	0.375	0.234	0.266	0.219	0.266
	HOXA13	0.089	0.281	0.312	0.281	0.297	0.281	0.203	0.297	0.234	0.234
	MDK + CDC2	0.484	0.578	0.500	0.578	0.469	0.578	0.375	0.422	0.359	0.391
	MDK + IGFBP5	0.484	0.547	0.484	0.547	0.469	0.500	0.531	0.344	0.359	0.359
	MDK + HOXA13	0.484	0.547	0.469	0.547	0.469	0.484	0.422	0.500	0.359	0.562
	CDC2 + IGFBP5	0.484	0.500	0.484	0.500	0.469	0.484	0.453	0.484	0.438	0.422
	CDC + HOXA13	0.469	0.422	0.453	0.469	0.422	0.484	0.484	0.406	0.453	0.453
	IGF + HOXA13	0.391	0.391	0.375	0.375	0.375	0.422	0.234	0.406	0.219	0.250
	MDK + CDC2 + IGFBP5	0.484	0.578	0.484	0.562	0.500	0.562	0.500	0.484	0.312	0.391
	MDK + CDC2 + HOXA13	0.484	0.609	0.484	0.609	0.484	0.578	0.344	0.578	0.359	0.391
	MDK + IGFBP5 + HOXA13	0.469	0.531	0.500	0.547	0.469	0.516	0.469	0.406	0.359	0.359
	CDC2 + IGFBP5 + HOXA13	0.469	0.484	0.469	0.484	0.469	0.500	0.469	0.469	0.453	0.422
	MDK+ CDC2+IGFBP5+HOXA13	0.484	0.609	0.484	0.609	0.469	0.531	0.516	0.484	0.344	0.391

特异性98%的灵敏度

图16e

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.013	0.000	0.078	0.250	0.062
CDC2	-0.019	-0.016	0.141	-0.016	0.031
IGFBP5	0.046	0.016	0.109	0.141	-0.047
HOXA13	0.123	-0.016	-0.016	-0.031	0.031
MDK + CDC2	0.016	0.016	0.078	0.016	-0.031
MDK + IGFBP5	0.000	0.000	0.078	0.078	0.000
MDK + HOXA13	0.000	-0.016	0.031	0.016	0.062
CDC2 + IGFBP5	0.016	0.016	0.109	0.094	0.016
CDC + HOXA13	-0.016	-0.016	0.141	-0.016	0.047
IGF + HOXA13	0.000	0.000	0.031	0.141	-0.016
MDK + CDC2 + IGFBP5	0.000	0.000	0.078	0.062	-0.031
MDK + CDC2 + HOXA13	0.000	0.000	0.078	-0.062	0.000
MDK + IGFBP5 + HOXA13	-0.016	-0.016	0.047	-0.047	0.000
CDC2 + IGFBP5 + HOXA13	0.000	-0.031	0.031	-0.047	0.000
MDK+CDC2+IGFBP5+HOXA13	0.016	0.016	0.062	-0.047	-0.016

在特异性=0.80处开始,灵敏度的增加

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.062	0.047	0.047	0.078	0.062
CDC2	0.222	-0.016	-0.016	0.125	0.031
IGFBP5	0.046	0.031	0.031	-0.031	0.016
HOXA13	0.132	0.000	0.031	-0.031	-0.094
MDK + CDC2	0.047	0.047	0.016	0.109	0.000
MDK + IGFBP5	0.016	0.016	0.016	0.203	0.000
MDK + HOXA13	0.031	0.031	0.016	-0.047	0.031
CDC2 + IGFBP5	0.016	0.016	0.016	0.062	0.016
CDC + HOXA13	-0.016	-0.016	-0.016	-0.047	0.047
IGF + HOXA13	0.000	0.000	0.062	-0.016	0.000
MDK + CDC2 + IGFBP5	0.000	0.000	0.016	0.188	0.016
MDK + CDC2 + HOXA13	0.031	0.031	0.000	0.031	0.000
MDK + IGFBP5 + HOXA13	0.016	0.016	0.047	0.141	0.000
CDC2 + IGFBP5 + HOXA13	0.016	0.016	0.031	0.094	0.000
MDK+CDC2+IGFBP5+HOXA13	-0.016	0.000	-0.016	0.094	-0.016

在特异性=0.85处开始,灵敏度的增加

图17a

图17b

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.161	0.047	0.078	0.078	-0.016
CDC2	0.365	0.047	0.156	0.125	0.047
IGFBP5	0.350	0.000	0.094	-0.031	0.062
HOXA13	0.259	0.031	0.031	-0.031	-0.031
MDK + CDC2	0.016	0.000	0.062	0.109	0.016
MDK + IGFBP5	0.078	0.047	0.016	0.062	0.016
MDK + HOXA13	0.031	0.031	0.078	-0.047	-0.016
CDC2 + IGFBP5	0.094	0.062	0.078	0.062	0.047
CDC + HOXA13	0.062	0.031	0.141	-0.047	0.078
IGF + HOXA13	0.000	0.000	0.016	-0.016	-0.016
MDK + CDC2 + IGFBP5	0.062	0.047	0.078	0.062	0.047
MDK + CDC2 + HOXA13	0.000	0.016	0.094	0.031	0.031
MDK + IGFBP5 + HOXA13	0.031	0.031	0.000	0.000	-0.016
CDC2 + IGFBP5 + HOXA13	0.094	0.031	0.016	-0.047	0.016
MDK+CDC2+IGFBP5+HOXA13	0.031	0.016	0.062	0.047	0.062

在特异性=0.90处开始,灵敏度的增加

图17c

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.306	0.062	0.078	0.109	0.062
CDC2	0.445	0.109	0.109	0.188	0.109
IGFBP5	0.336	0.000	0.000	0.031	-0.031
HOXA13	0.249	-0.016	-0.016	0.000	-0.047
MDK + CDC2	0.109	0.109	0.125	0.219	0.109
MDK + IGFBP5	0.047	0.031	0.047	0.062	0.016
MDK + HOXA13	0.078	0.078	0.078	-0.047	0.062
CDC2 + IGFBP5	0.047	0.062	0.031	0.188	0.047
CDC + HOXA13	0.109	0.125	0.094	-0.047	0.094
IGF + HOXA13	0.000	0.000	-0.016	-0.016	0.000
MDK + CDC2 + IGFBP5	0.062	0.062	0.062	0.062	0.031
MDK + CDC2 + HOXA13	0.109	0.109	0.125	0.125	0.078
MDK + IGFBP5 + HOXA13	0.016	0.000	0.047	0.000	0.016
CDC2 + IGFBP5 + HOXA13	0.031	0.078	0.047	-0.047	0.031
MDK+CDC2+IGFBP5+HOXA13	0.078	0.078	0.047	0.047	0.016

在特异性=0.95处开始,灵敏度的增加

图17d

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.353	0.031	0.078	0.203	0.203
CDC2	0.406	0.000	0.109	0.062	-0.047
IGFBP5	0.336	0.000	0.016	0.031	0.047
HOXA13	0.192	-0.031	-0.016	0.094	0.000
MDK + CDC2	0.094	0.078	0.109	0.047	0.031
MDK + IGFBP5	0.062	0.062	0.031	-0.188	0.000
MDK + HOXA13	0.062	0.078	0.016	0.078	0.203
CDC2 + IGFBP5	0.016	0.016	0.016	0.031	-0.016
CDC + HOXA13	-0.047	0.016	0.062	-0.078	0.000
IGF + HOXA13	0.000	0.000	0.047	0.172	0.031
MDK + CDC2 + IGFBP5	0.094	0.078	0.062	-0.016	0.078
MDK + CDC2 + HOXA13	0.125	0.125	0.094	0.234	0.031
MDK + IGFBP5 + HOXA13	0.062	0.047	0.047	-0.062	0.000
CDC2 + IGFBP5 + HOXA13	0.016	0.016	0.031	0.000	-0.031
MDK+CDC2+IGFBP5+HOXA13	0.125	0.125	0.062	-0.031	0.047

在特异性=0.98处开始，灵敏度的增加

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.030	0.000	0.168	0.180	0.069
CDC2	-0.025	-0.040	0.178	-0.104	0.094
IGFBP5	0.037	0.022	0.131	0.119	-0.025
HOXA13	0.114	-0.007	-0.002	0.116	0.027
MDK + CDC2	0.012	0.030	0.158	-0.101	-0.049
MDK + IGFBP5	0.000	0.000	0.101	0.146	-0.022
MDK + HOXA13	0.000	-0.054	0.126	-0.089	0.072
CDC2 + IGFBP5	0.022	0.012	0.143	0.146	0.035
CDC + HOXA13	-0.025	-0.040	0.158	-0.109	0.106
IGF + HOXA13	0.000	0.000	0.042	0.158	0.005
MDK + CDC2 + IGFBP5	0.000	0.000	0.133	0.151	-0.012
MDK + CDC2 + HOXA13	0.000	0.000	0.165	-0.091	-0.035
MDK + IGFBP5 + HOXA13	-0.022	-0.022	0.081	-0.042	-0.032
CDC2 + IGFBP5 + HOXA13	0.000	-0.094	0.064	-0.032	0.015
MDK+CDC2+IGFBP5+HOXA13	0.027	0.000	0.138	-0.057	-0.010

在特异性=0.80处开始，特异性的增加

图17e

图17f

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.059	0.041	0.123	0.130	0.039
CDC2	0.125	-0.023	0.130	0.130	0.044
IGFBP5	0.041	0.009	0.081	0.069	0.036
HOXA13	0.064	-0.001	-0.001	0.066	-0.015
MDK + CDC2	0.054	0.046	0.108	0.118	-0.001
MDK + IGFBP5	0.051	0.041	0.051	0.096	0.019
MDK + HOXA13	0.051	0.044	0.103	0.086	0.041
CDC2 + IGFBP5	0.017	0.002	0.093	0.096	0.031
CDC + HOXA13	-0.018	-0.028	0.113	0.086	0.056
IGF + HOXA13	-0.001	-0.001	0.002	0.108	-0.008
MDK + CDC2 + IGFBP5	-0.001	-0.001	0.083	0.101	0.014
MDK + CDC2 + HOXA13	0.049	0.054	0.118	0.130	-0.001
MDK + IGFBP5 + HOXA13	0.007	0.014	0.031	0.113	0.009
CDC2 + IGFBP5 + HOXA13	0.051	0.007	0.014	0.101	-0.003
MDK+CDC2+IGFBP5+HOXA13	-0.015	-0.001	0.088	0.110	0.004

在特异性=0.85处开始，特异性的增加

图17g

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.078	0.021	0.075	0.080	-0.011
CDC2	0.098	0.038	0.080	0.080	0.028
IGFBP5	0.090	0.001	0.043	0.019	0.011
HOXA13	0.088	0.004	0.028	0.016	-0.026
MDK + CDC2	0.004	0.001	0.058	0.068	0.043
MDK + IGFBP5	0.028	0.026	0.001	0.046	0.009
MDK + HOXA13	0.011	0.016	0.063	0.036	-0.009
CDC2 + IGFBP5	0.021	0.028	0.043	0.046	0.058
CDC + HOXA13	0.019	0.033	0.063	0.036	0.053
IGF + HOXA13	0.001	0.001	0.004	0.058	-0.031
MDK + CDC2 + IGFBP5	0.048	0.043	0.036	0.051	0.048
MDK + CDC2 + HOXA13	0.001	0.004	0.075	0.080	0.053
MDK + IGFBP5 + HOXA13	0.041	0.046	0.001	0.063	0.004
CDC2 + IGFBP5 + HOXA13	0.021	0.033	0.021	0.051	-0.001
MDK+CDC2+IGFBP5+HOXA13	0.031	0.021	0.048	0.060	0.046

在特异性=0.90处开始，特异性的增加

图17h

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.043	0.025	0.030	0.030	0.028
CDC2	0.050	0.025	0.030	0.030	0.023
IGFBP5	0.043	0.001	0.001	0.015	-0.007
HOXA13	0.045	-0.002	-0.004	0.020	-0.012
MDK + CDC2	0.030	0.030	0.033	0.038	0.013
MDK + IGFBP5	0.023	0.025	0.010	-0.004	0.006
MDK + HOXA13	0.030	0.030	0.013	-0.014	0.028
CDC2 + IGFBP5	0.020	0.018	0.015	0.025	0.015
CDC + HOXA13	0.023	0.025	0.025	-0.014	0.015
IGF + HOXA13	0.001	0.001	-0.007	0.008	0.003
MDK + CDC2 + IGFBP5	0.025	0.023	0.023	0.001	0.010
MDK + CDC2 + HOXA13	0.035	0.033	0.033	0.030	0.013
MDK + IGFBP5 + HOXA13	0.006	0.001	0.003	0.013	0.006
CDC2 + IGFBP5 + HOXA13	0.003	0.015	0.008	0.001	0.008
MDK+CDC2+IGFBP5+HOXA13	0.028	0.030	0.015	0.010	0.010

在特异性=0.95处开始，特异性的增加

模型	LDA	LogReg	SVM	KN5N	TREE
MDK	0.015	0.005	0.008	0.005	0.008
CDC2	0.020	0.000	0.000	0.000	-0.007
IGFBP5	0.018	0.000	0.000	0.000	0.003
HOXA13	0.015	-0.007	-0.007	0.010	0.000
MDK + CDC2	0.003	0.003	0.005	0.008	0.003
MDK + IGFBP5	0.000	0.000	0.000	-0.010	0.000
MDK + HOXA13	0.008	0.008	0.008	0.010	0.008
CDC2 + IGFBP5	0.005	0.000	0.003	-0.005	0.000
CDC + HOXA13	-0.002	0.000	0.000	-0.010	-0.002
IGF + HOXA13	0.000	0.000	0.003	0.008	0.003
MDK + CDC2 + IGFBP5	0.003	0.003	0.005	0.005	0.010
MDK + CDC2 + HOXA13	0.008	0.008	0.003	0.005	0.003
MDK + IGFBP5 + HOXA13	0.008	0.005	0.000	0.005	0.000
CDC2 + IGFBP5 + HOXA13	0.000	0.000	0.008	0.013	-0.007
MDK+CDC2+IGFBP5+HOXA13	0.005	0.005	0.005	0.005	0.010

在特异性=0.98处开始，特异性的增加

图17i

图17j

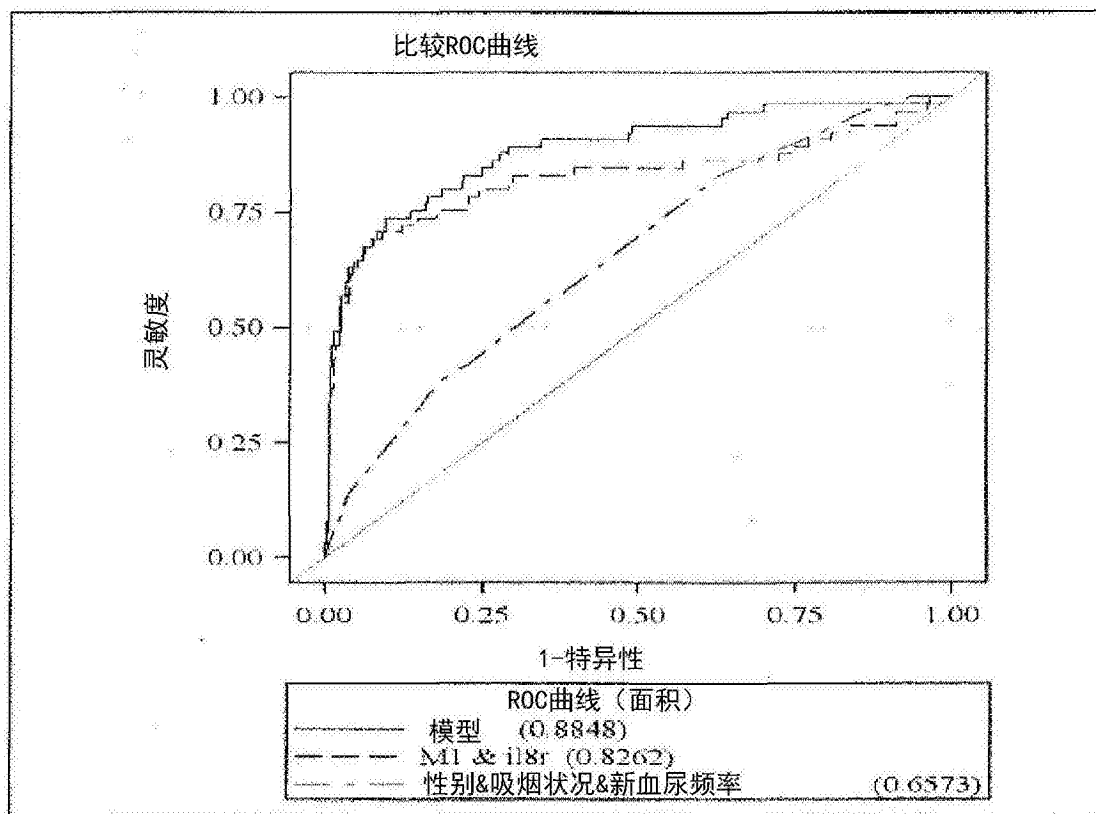


图18

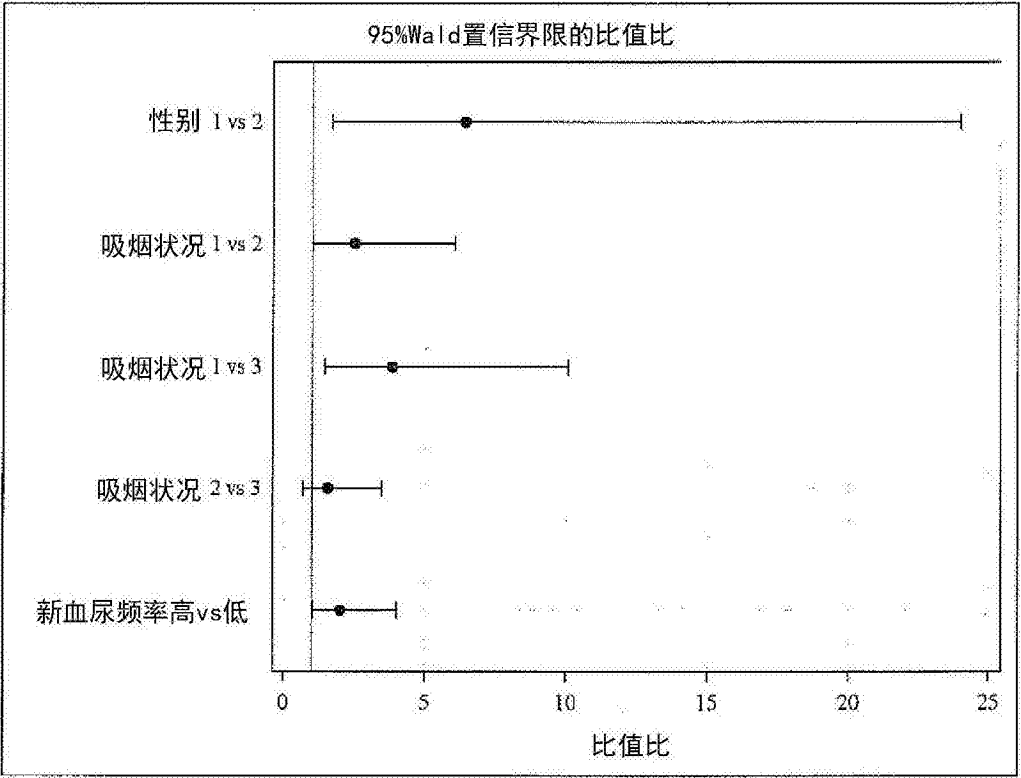


图19

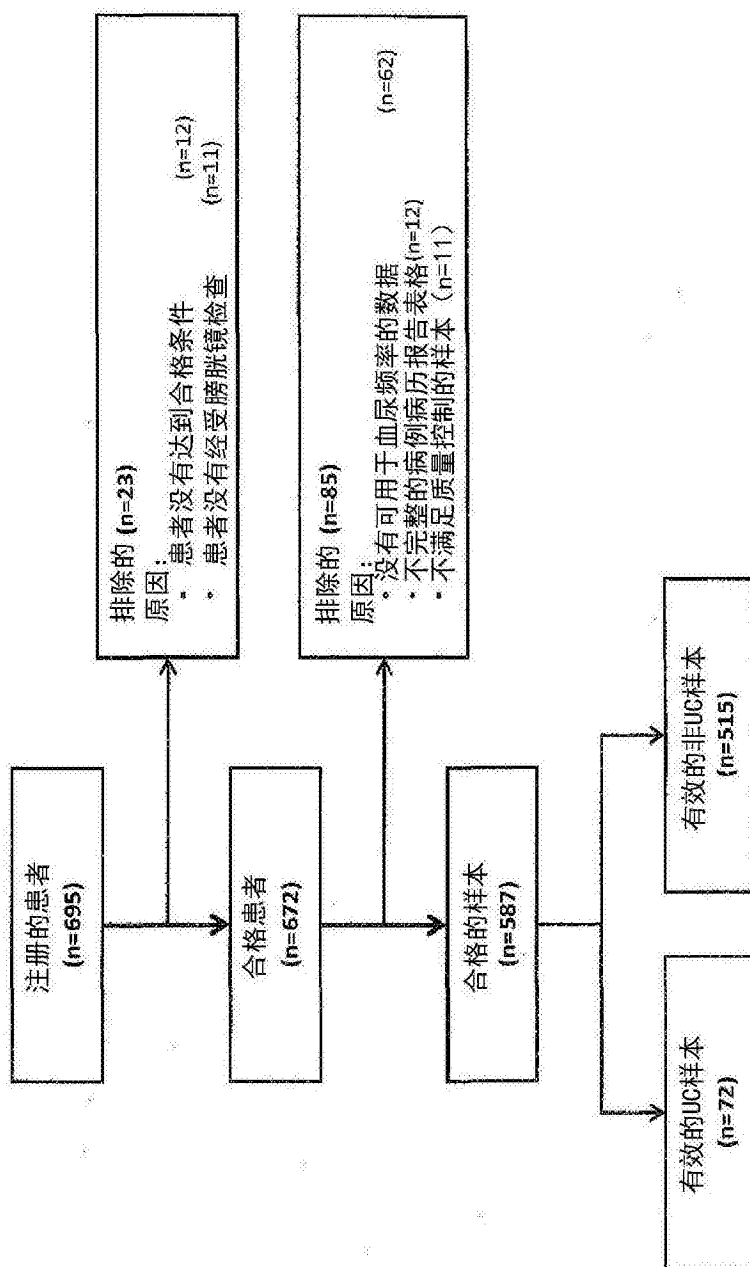


图20A

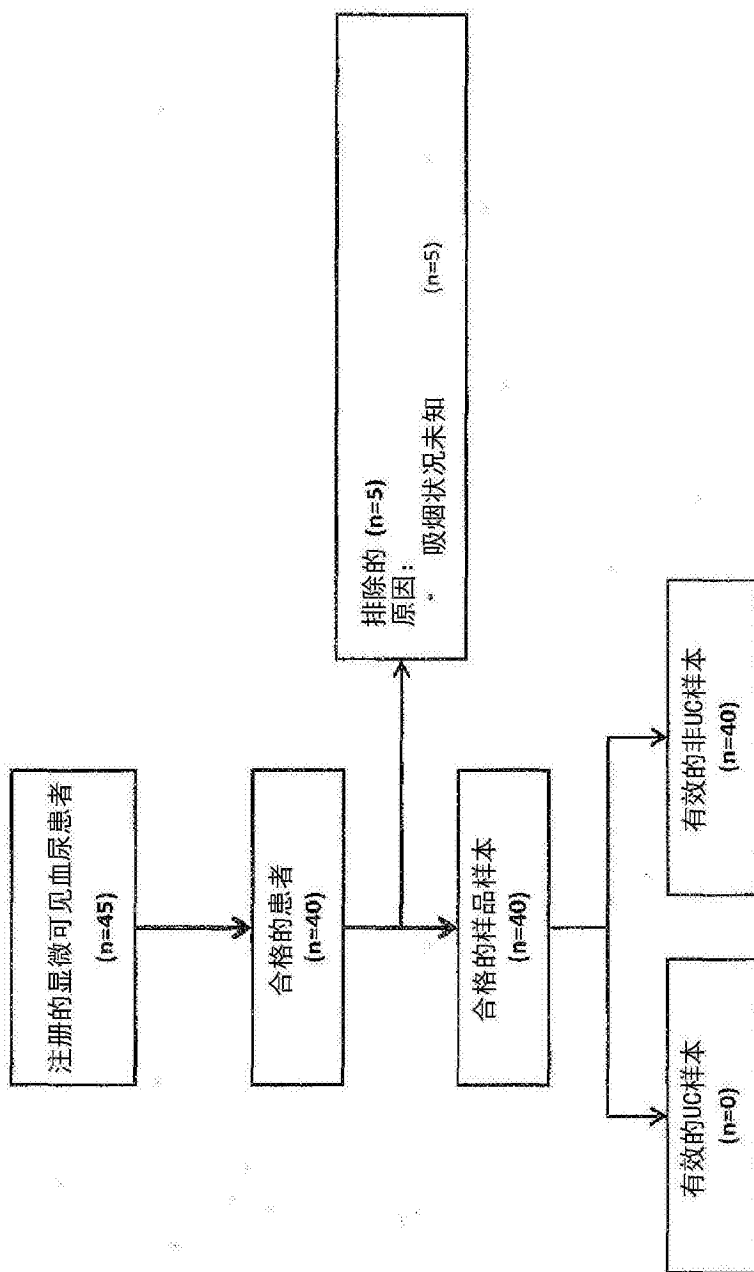


图20B

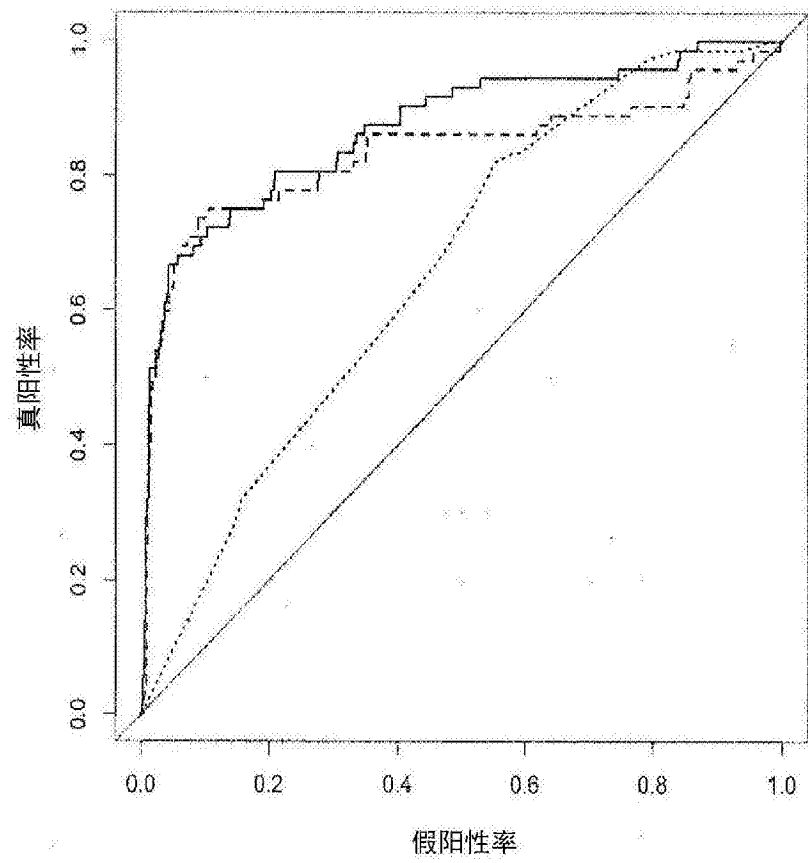


图21

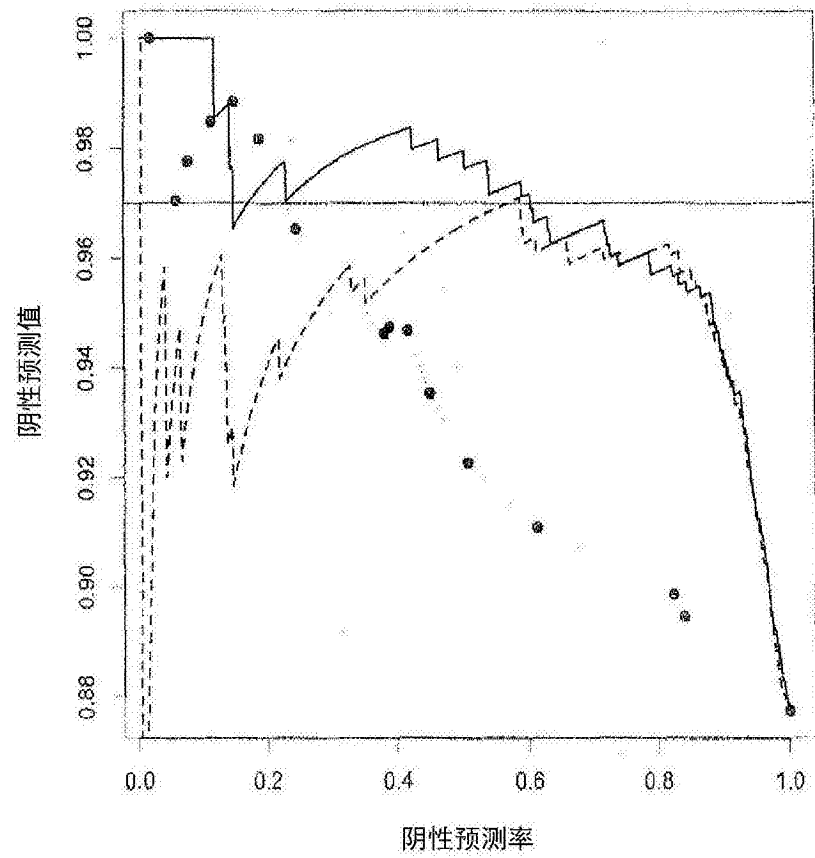


图22

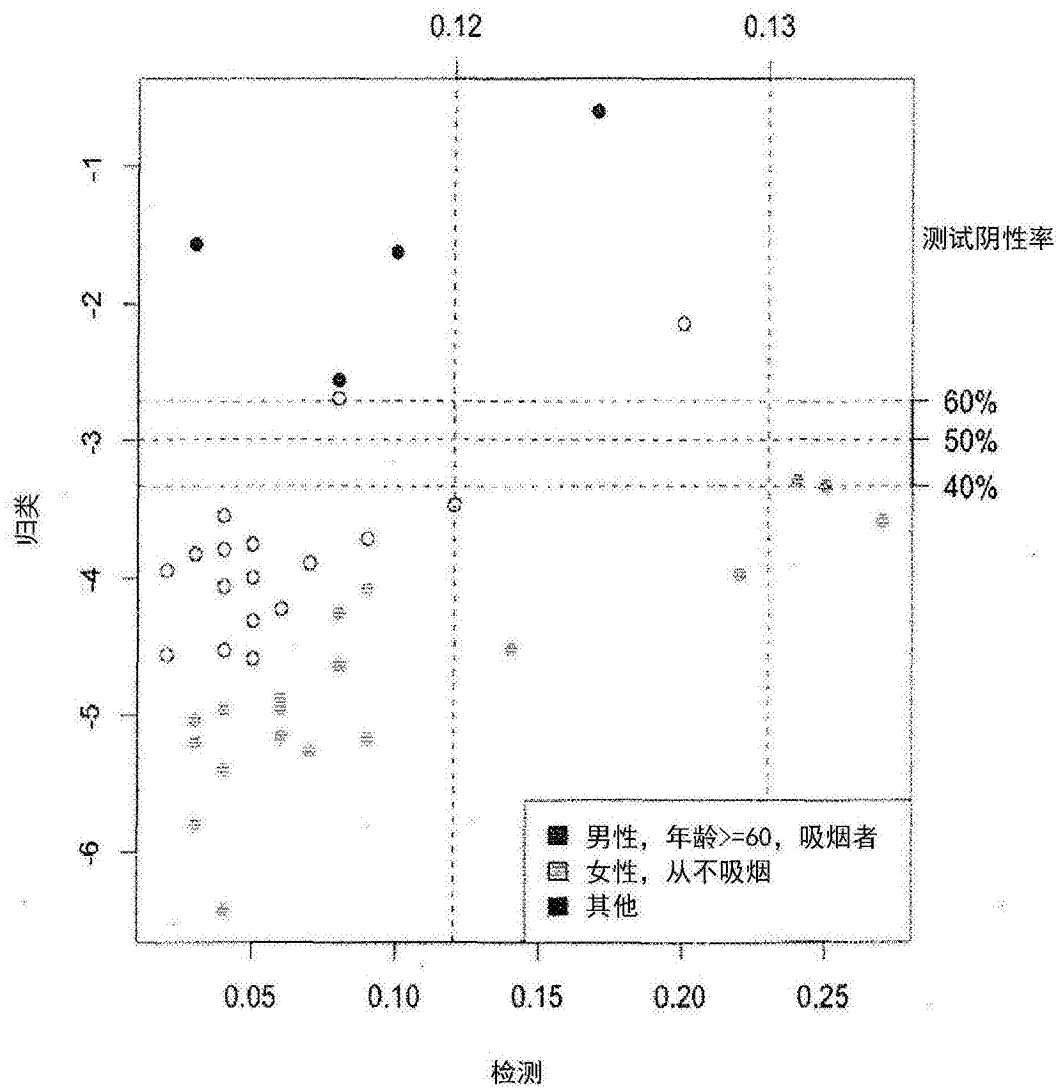


图23

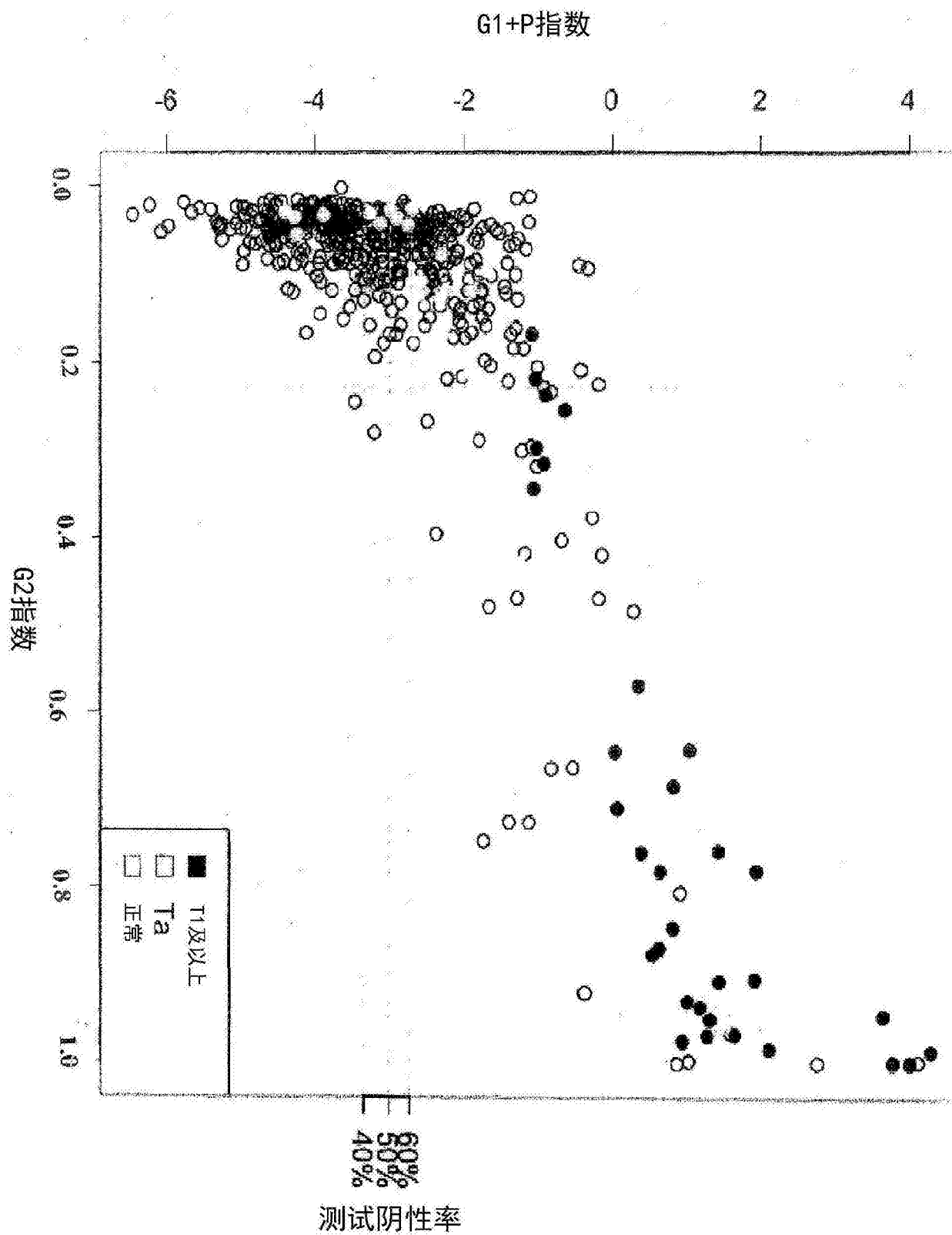


图24