

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

C12Q 1/68 (2006.01)

G06F 19/00 (2006.01)

G01N 33/574 (2006.01)



# [12] 发明专利说明书

专利号 ZL 01823658.8

[45] 授权公告日 2008 年 7 月 2 日

[11] 授权公告号 CN 100398663C

[22] 申请日 2001.7.23 [21] 申请号 01823658.8

[86] 国际申请 PCT/JP2001/006330 2001.7.23

[87] 国际公布 WO2003/010337 英 2003.2.6

[85] 进入国家阶段日期 2004.3.23

[73] 专利权人 霍夫曼 - 拉罗奇有限公司

地址 瑞士巴塞尔

[72] 发明人 冈正朗 浜本义彦 冈部尚文

[56] 参考文献

DE19840671A 2000.3.2

WO0100083A 2001.1.4

WO9709925A 1997.3.20

WO0140517A 2001.6.7

WO9856953A 1998.12.17

WO0133228A 2001.5.10

US5862304A 1999.1.19

WO0118542A 2001.3.15

STATISTICAL PATTERN RECOGNITION: A REVIEW. JAIN A K ET AL.; IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE INC. NEW YORK, US,, Vol. 22 No. 1. 2000

用于统计模式识别的一个神经网络模型. 刘秀虹等. 黑龙江大学(自然科学学报), 第 13 卷第 3 期. 1996

审查员 周 洋

[74] 专利代理机构 北京市柳沈律师事务所

代理人 巫肖南 封新琴

权利要求书 1 页 说明书 15 页 附图 5 页

[54] 发明名称

用于预测癌症复发的评定体系

[57] 摘要

本发明涉及预测癌症复发的评定体系。更具体地，本发明涉及选择基因和/或蛋白，并通过测量人肿瘤组织中基因和/或蛋白的表达，以及对它们的表达模式与具有癌症复发以及没有癌症复发的患者的人原发性肿瘤中基因和/或蛋白的表达模式进行比较，利用所选的基因和/或蛋白产生用于预测癌症复发的公式。本发明还涉及一种用于实施本发明的方法的试剂盒，包含 DNA 芯片、寡核苷酸芯片、蛋白芯片、肽、抗体、探针和引物，这些是实现 DNA 微阵列、寡核苷酸微阵列、蛋白阵列、NORTHERN 印迹、原位杂交、核糖核酸酶保护分析、蛋白质印迹、反转录聚合酶链式反应(以下称之为 RT-PCR)，以检测指示癌症复发的至少 2 个或多个基因和/或蛋白，优选 4 个或多个基因和/或蛋白，

个或多个基因和/或蛋白的表达所必需的。

1. 选择用于预测癌症复发的基因和 / 或蛋白子集的方法, 包括以下步骤:

(a) 检测有癌症复发和没有癌症复发的患者癌组织中基因和 / 或蛋白的表达水平或模式; 和,

(b) 通过基于所述基因和 / 或蛋白的表达水平或模式的统计分析, 选择至少 2 个基因和 / 或蛋白。

2. 权利要求 1 的方法, 其中所述统计分析是 Fisher 判据。

3. 权利要求 1 或 2 的方法, 其中所述基因和 / 或蛋白选自: 血小板衍生生长因子受体 $\alpha$  (PDGFRA)、肿瘤坏死因子 $\alpha$ 诱导蛋白 A20、溶酶体相关多跨膜蛋白(LAPTm5)、HLA-DR  $\alpha$ 重链 (II类抗原)、rel 原癌基因、Stat50、公认的丝氨酸/苏氨酸蛋白激酶、MADS/MEF2-家族转录因子(MEF2C)、来自 3p21.3 的 HUMLUCA19 人粘粒克隆 LUCA19、DEAD-盒蛋白 p72、波形蛋白和 KIAK0002。

## 用于预测癌症复发的评定体系

本发明涉及预测癌症复发的评定体系。更具体地，本发明涉及选择基因和/或蛋白，通过测量人肿瘤组织中基因和/或蛋白的表达，并将它们的表达模式与具有癌症复发以及没有癌症复发的患者的人原发性肿瘤中该基因和/或蛋白的表达模式进行比较，利用所选的基因和/或蛋白产生用于预测癌症复发的公式。

本发明还涉及一种用于实施本发明的方法的试剂盒，包含 DNA 芯片、寡核苷酸芯片、蛋白芯片、肽、抗体、探针和引物，这些是实现 DNA 微阵列、寡核苷酸微阵列、蛋白阵列、Northern 印迹、原位杂交、核糖核酸酶保护分析、Western 印迹、ELISA、反转录聚合酶链式反应（以下称之为 RT-PCR），以检测指示癌症复发的至少 2 个或多个基因和/或蛋白，优选 4 个或多个基因和/或蛋白，更优选 6 个或多个基因和/或蛋白，并且最优选 12 个或多个基因和/或蛋白的表达所必需的。

### 发明背景

癌症是世界上引起死亡的主要原因之一。癌症总体流行率大约占人口的 1%，并且年发病率大约为 0.5%。大约十分之一从医院遣返的患者被初步诊断为癌症。主要的现有治疗模式是外科手术切除术、放射疗法、化学疗法以及包括激素疗法在内的生物疗法。此外，新近发展的生物技术提供了新的治疗模式，例如基因疗法。尽管如此，癌症是可怕的疾病，因为在大多数情况下实际上不存在有效的治疗。癌症治疗的主要困难之一是癌细胞变得对药物具有抗性的能力，并且扩散到组织其它部位，在那里它们可以产生新的肿瘤，这常常导致复发。如果癌症复发在发生复发前可预测，则这种癌症通过用外科手术局部治疗是可医治的。

在各种肿瘤中，肝细胞癌(以下称之为 HCC)是世界上最为常见的致命癌症之一，并且发病数目在许多国家包括美国、日本、中国和欧洲国家是渐增的。乙型肝炎病毒(以下称之为 HBV)和丙型肝炎病毒(以下称之为 HCV)感染都可以是引起 HCC 的原因。事实上，HCC 患者的增加与慢性 HCV 感染的

增加是对应的 (El-Serag, H.B.& Mason, A.C. Rising incidence of hepatocellular carcinoma in the United States, *N.Engl.J.Med.*340,745-750 (1999) 以及 Okuda, K. Hepatocellular carcinoma, *J.Hepatol.*32,225-237 (2000) )。尽管 HCC 发病率提高, 对于这种疾病没有有前景的治疗。治疗 HCC 的主要问题是肝内转移。在 30 到 50%接受肝脏切除术的 HCC 患者中观察到复发 (Iizuka, N. 等. NM23-H1 and NM23-H2 messenger RNA abundance in human hepatocellular carcinoma, *Cancer Res.*55,652-657 (1995), Yamamoto, J. 等 Recurrence of hepatocellular carcinoma after surgery, *Br.J.Surg.*83, 1219-1222 (1996), 以及 Poon, R.T. 等 Different risk factors and prognosis for early and late intrahepatic recurrence after resection of hepatocellular carcinoma, *Cancer* 89,500-507 (2000) )。尽管病理学 TNM 分类系统已用于 HCC 的治疗, 这个系统只能拙劣地预测经受肝脏切除术患者的复发 (Izumi, R. 等 Prognostic factors of hepatocellular carcinoma in patient undergoing hepatic resection, *Gastroenterology* 106,720-727 (1994) )。也已经提出了许多分子作为 HCC 的预测性标志, 没有任何一个被证明临床上可用 (Iizuka, N. 等 NM23-H1 and NM23-H2 messenger RNA abundance in human hepatocellular carcinoma, *Cancer Res.*55,652-657 (1995), Hsu, H.C. 等 Expression of p53 gene in 184 unifocal hepatocellular carcinomas: association with tumor growth and invasiveness, *Cancer Res.*53,4691-4694 (1993), 以及 Mathew, J. 等 CD44 is expressed in hepatocellular carcinomas showing vascular invasion, *J.Pathol.* 179,74-79 (1996))。因此, 任何预测复发的方法对于理解癌症机制以及对于建立新的癌症疗法将是非常有价值的。然而, 由于存在着传统方法对预测复发的技术局限, 而且更多的局限可归于高度的患者间肿瘤异质性, 有必要设计新的方法描述肿瘤特征并预测癌症复发。

最近发展的允许人们对大量基因进行平行的表达分析的微阵列技术, 在医学领域开拓了一个新的纪元 (Schena, M. 等 Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* 270,467-470 (1995), 以及 DeRisi, J. 等 Use of a cDNA microarray to analyze gene expression patterns in human cancer, *Nature Genet.*14,457-460 (1996) )。尤其是, cDNA 微阵列对肿瘤基因表达的研究已经对恶性肿瘤的性质例如预后和药物敏感性提供了重要的深刻见解 (Alizadeh, A.A. 等 Distinct types of diffuse

large B-cell lymphoma identified by gene expression profiling, *Nature* 403,503-511 (2000), 以及 Scherf, U.等 A gene expression database for the molecular pharmacology of cancer, *Nature Genet.*24,236-244 (2000) )。

最近, 监控记忆被引入基因表达分析中 (Brazma, A.& Vilo, J. Gene expression data analysis, *FEBS Lett.*480,17-24 (2000) 以及 Kell, D.B.& King, R.D. On the optimization of classes for the assignment of unidentified reading frames in functional genomics programs: the need for machine learning, *Trends Biotechnol.*18,93-98 (2000) )。利用分类的样品, 监控记忆具有很好的关于数据性质的演绎知识的决定性优势 (Duda, R.O.等 *Pattern classification*, John Wiley & Sons (2001), 以及 Jain, A.K.等 *Statistical pattern recognition: A review*, *IEEE Trans. Pattern Analysis and Machine Intelligence.*22,4-37 (2000) )。然而, 没有任何一种以前公开的监控记忆方法直接地评价基因的组合从而能够利用有关统计特征的信息, 即基因分布结构 (Golub, T.R.等 *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring*, *Science* 286,531-537 (1999), 以及 Brown, M.P. 等 *Knowledge-based analysis of microarray gene expression data by using support vector machines*, *Proc. Natl. Acad. Sci. USA* 97,262-267 (2000))。

预测癌症复发的评定体系是用统计模式识别中的监控记忆(supervised learning in statistical pattern recognition)通过分析 DNA 微阵列数据创建的 (Duda, R.O.等 *Pattern classification*, John Wiley & Sons (2001))。

统计模式识别的监控记忆已经被成功地用于解决诸如文献分类、语音识别、生物统计分析、以及遥感等大量问题 (Jain, A.K.等 *Statistical pattern recognition : A review*, *IEEE Trans, Pattern Analysis and Machine Intelligence.*22,4-37 (2000) )。

在本发明中, 发明人提供了一种通过分析人原发性肿瘤基因和/或蛋白的表达而预测癌症复发的评定体系。也就是说, 本发明涉及一种预测癌症复发的方法, 其包括检测人原发性肿瘤组织基因和/或蛋白的表达, 并将它与具有癌症复发以及没有癌症复发的患者的人原发性肿瘤基因和/或蛋白的表达进行比较。

附图说明

图 1 图示了基因选择的程序(步骤 1-7)以及用最佳的基因子集对评定体系的评价(步骤 8-10)。

图 2 图示了最佳的基因数目。

图 3 图示了被选择用于早期肝内复发预测的基因的 mRNA 的平均差异。比较了 A 组(显示为 A)和 B 组(显示为 B)之间 12 个基因的 mRNA 的平均差异。

图 4 图示了用于早期肝内复发预测的病毒类型、TNM 阶段、以及分值(T 值)之间的关系。利用 12 个基因的最佳子集,用 3 个测试样品评价通过 30 个训练样品创建的评定体系。这一操作独立重复 10 次。计算所有测试样品的 T 值。在 T 值低于 0 时预测到早期肝内复发。不管阶段和病毒类型, T 为负值的所有 HCC 具有早期肝内复发,而 T 为正值的所有 HCC 没有复发。填充的, A 组(具有早期肝内复发的患者);空白的, B 组(没有早期肝内复发的患者);○, I 期;◇, II 期;△, III 期;□, IVA 期;B; HBV-阳性, C; HCV-阳性, N; HBV-HCV-双阴性。

图 5 图示了评定体系。

#### 发明的详细说明

在本发明中,利用了来自包括脑、肺、乳房、胃、肝、胰腺、胆囊、结肠、直肠、肾脏、膀胱、卵巢、子宫、前列腺以及皮肤的那些肿瘤的人组织。在外科手术期间切除人组织后,优选它们立即在液氮或含有干冰的丙酮中冷冻,并贮存于-70 和-80℃之间备用,包埋或不包埋在 O.C.T.化合物中(Sakura-Seiki, Tokyo, Japan, 目录号 4583)。

测试癌症复发可能性的患者的肿瘤组织中基因和/或蛋白的表达是通过测量 RNA 和/或蛋白的水平来分析的。在许多情况下, RNA 和/或蛋白的水平是通过测量包括荧光素和罗丹明在内的底物的荧光、发光氨(lumenole)的化学发光、放射性物质包括  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{35}\text{S}$ ,  $^{33}\text{p}$ ,  $^{32}\text{p}$  和  $^{125}\text{I}$  的放射活性、以及光学密度确定的。RNA 和/或蛋白的表达是通过已知的方法包括 DNA 微阵列(Schena, M.等 Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science 270,467-470 (1995), 以及 Lipshutz, R.J.等 High density synthetic oligonucleotide arrays, Nature Genet.21,20-24 (1999) )、RT-PCR (Weis, J.H.等 Detection of rare mRNAs via quantitative

RT-PCR, Trends Genetics 8,263-264 (1992), 以及 Bustin, S.A.Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, J.Mol.Endocrinol.25,169-193 (2000) )、NORTHERN 印迹和原位杂交 (Parker, R.M.& Barnes, N.M.mRNA : detection in situ and northern hybridization, Methods Mol.Biol.106, 247-283 (1999) )、核糖核酸酶保护试验 (Hod, Y.A.Simplified ribonuclease protection assay, Biotechniques 13,852-854 (1992), Saccomanno, C.F.等 A faster ribonuclease protection assay, Biotechniques 13,846-850 (1992) )、Western 印迹 (Towbin, H.等 Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets, Proc.Natl.Acad.Sci.USA 76,43504354 (1979), Burnette, W.N.Western blotting: Electrophoretic transfer of proteins form sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radioiodinated protein A, Anal.Biochem.112,195-203 (1981) )、ELISA 分析(Engvall, E.& Perlman, P.Enzyme-linked immunosorbent assay (ELISA): Quantitative assay of immunoglobulin G, Immunochemistry 8: 871-879 (1971) )、以及蛋白阵列 (Merchant, M.& Weinberger, S.R.Review: Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry, Electrophoresis 21,1164-1177 (2000), Paweletz, C.P.等 Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip, Drug Development Research 49,34-42 (2000) )确定的。

已有和没有早期复发的癌症患者肿瘤中基因和/或蛋白的表达以与测试复发可能性的患者相同的方式确定。

虽然癌症的早期复发在不同癌症类型中有差异,但通常是在切除后一到两年内发生。所以在切除后一到两年内已有复发的癌症患者的肿瘤可以被用作早期复发患者的肿瘤,而在切除后一到两年内没有复发的癌症患者的肿瘤可以用作没有早期复发的患者的肿瘤。

有早期复发和没有早期复发的癌症患者之间,肿瘤基因和/或蛋白的表达水平或模式的差别,可通过已知的统计学分析方法予以分析和检测。统计模式识别中的监控记忆可用于肿瘤基因和/或蛋白表达模式的统计学分析。通过统计模式识别中的监控记忆,从所检测的基因和/或蛋白中挑选出其表达可指示癌症复发的 2 个或多个基因和/或蛋白。

首先通过一维的判据(criteria)挑选一些可指示癌症复发的基因和/或蛋白。然后,用能够将基因和/或蛋白所有可能的组合考虑在内的“排除一个”的方法通过详尽的研究,从这些基因和/或蛋白中挑选出基因和/或蛋白的最佳子集。

预测癌症复发的公式是通过利用可指示癌症复发的至少2个或更多个基因和/或蛋白,优选4个或更多个基因和/或蛋白,更优选6个或更多个基因和/或蛋白,最优选12个或更多个基因和/或蛋白的最佳子集创建的。即使是在样品数目比基因和/或蛋白数目少时也能较好地工作的简单分类程序例如线性分类程序被用来创建公式(Duda, R.O.等 Pattern classification, John Wiley & Sons (2001), 及 Jain, A.K.等 Statistical pattern recognition: A review, IEEE Trans.Pattern Analysis and Machine Intelligence.22,4-37 (2000))。

本发明还涉及实施本发明的方法的试剂盒。检测可指示癌症复发的2个或多个基因和/或蛋白的表达模式的试剂盒由包括RNA提取试剂、用于合成cDNA和cRNA的酶、DNA芯片、寡核苷酸芯片、蛋白芯片、分析用的探针和引物、对照基因的DNA片断、以及各种蛋白的抗体在内的组分组成。试剂盒的组分很容易从市场上获得。例如,寡核苷酸芯片、胍-酚试剂、逆转录酶、T7 RNA聚合酶以及taq聚合酶可以购买并组装为本发明的试剂盒。

下列实施例仅仅阐释了本发明优选的用于预测癌症复发的方法,而不应当解释为局限于此。

## 实施例

### 实施例 1.选择病人用于早期肝内复发分析

据报道外科手术后的早期肝内复发(在一年之内)主要是因肝内转移引起的,而晚期复更可能是多发源点发生(Poon, R.T.等 Different risk factors and prognosis for early and late intrahepatic recurrence after resection of hepatocellular carcinoma, Cancer 89,500-507 (2000))。而且,公知肝内复发的患者比多发源点发生的患者的结果更严重(Yamamoto, J.等 Recurrence of hepatocellular carcinoma after surgery, Br.J.Surg.83,1219-1222 (1996), and Poon, R.T.等 Different risk factors and prognosis for early and late intrahepatic recurrence after resection of hepatocellular carcinoma, Cancer 89,500507 (2000))。因此研究了外科手术后的一年内与早期肝内复发关联的基因表达模

式。

33个患者在1997年3月和2000年1月之间在Yamaguchi University医院经受过HCC外科手术治疗。在外科手术前从所有的病例中都得到了书面的公告同意。研究方案在1996年5月得到了Yamaguchi大学医学院人类用途制度审查委员会的批准。在外科手术后在所有患者中进行了HCC组织病理学诊断。组织病理学检查同样揭示了在所有33个HCC样品中没有残余肿瘤(R0)。表1显示了基于Union Internationale Contre le Cancer (UICC) TNM分类的33个患者的临床病理学特征(Sobin, L.H.& Wittekind, C. TNM classification of Malignant Tumors, 5th ed., UICC, Wiley-Liss, 74-77 (1997))。在血清学上, 7个患者是乙型肝炎表面抗原阳性, 22个患者是抗-HCV抗体阳性, 而其余4个患者两者均为阴性。这33个患者继肝切除之后每3个月通过超声照相法、计算机X线断层摄影术、和甲胎蛋白水平来追踪癌症复发。必要时, 增加磁共振成像和肝血管造影术。这33个HCC患者患者中, 有12人(36%)发现早期肝内复发。在这12个患者的11人中, 复发的HCC在残余的肝脏内检测为多小结或者弥漫性扩散。在一个患者中, 在外科手术后9个月在邻近切除的原发性病变的部位检测到作为单个小结的新肿瘤。然后, 观察到多发性肺转移。其余的21个患者没有一个在外科手术后一年之内发生肝内复发及其它远距离的转移。这些患者被分成两组; 在一年之内具有肝内复发的患者在A组 (n = 12)而那些没有复发的在B组 (n = 21) (表1)。使用 $\chi^2$ 检验和Fisher's exact 检验阐明这2组之间临床病理学因素方面的差异。

## 实施例 2. 自组织中提取 RNA

组织碎片 (大约 125mm<sup>3</sup>) 被悬浮在 TRIZOL (Life Technologies, Gaithersburg, USA, 目录号 15596-018) 或 Sepasol-RNAI (Nacalai tesque, Kyoto, Japan, 目录号 306-55) 中, 并用 Polytron 匀浆两次 (Kinematica, Littau, Switzerland) (以最大速度匀浆 5 秒)。加入氯仿之后, 将组织匀浆物以 15,000 × g 离心 10 分钟, 并回收含 RNA 的水相。用异丙醇沉淀细胞总 RNA, 用 70% 乙醇洗涤 1 次, 并悬浮在 DEPC-处理的水中 (Life Technologies, Gaithersburg, USA, 目录号 10813-012)。RNA 在用 1.5 单位 DNase I (Life Technologies, Gaithersburg, USA, 目录号 18068015) 处理之后, 用 TRIZOL/氯仿再抽提, 用乙醇沉淀, 并溶在 DEPC-处理的水中。之后, 根据厂家的操作说明书, 利用

RNeasy Mini 试剂盒 (QIAGEN, Hilden, Germany, 目录号 74104) 除去小分子量核苷酸。总 RNA 的质量依据琼脂糖凝胶电泳后 28S 与 18S 核糖体 RNA 的比率判断。纯化的总 RNA 在 70%乙醇溶液中贮存于-80℃备用。

### 实施例 3.cDNA 和标记的 cRNA 探针的合成

利用逆向 SuperScript 选择系统 (Life Technologies, Gaithersburg, USA, 目录号 18090-019) 根据厂家的操作说明书合成 cDNA。5 μg 纯化的总 RNA 与含有 T7 启动子序列的寡聚-dT 引物 (Sawady Technology, Tokyo, Japan) 和 200 单位 SuperScriptII 反转录酶杂交, 并在 42℃温育 1 小时。产生的 cDNA 用酚/氯仿提取并用 Phase Lock Gel Light (Eppendorf, Hamburg, Germany, 目录号 0032 005.101) 纯化。

还利用 MEGAscript T7 试剂盒 (Ambion, Austin, USA, 目录号 1334), 并以所述 cDNA 作为模板, 根据厂家的操作说明书合成 cRNA。大约 5 μg cDNA 与 2 μl 含有 T7 聚合酶、腺苷三磷酸(ATP)和鸟苷三磷酸(GTP)各 7.5 mM、胞苷三磷酸(CTP)和尿苷三磷酸(UTP)各 5.625 mM、Bio-11-CTP 和 Bio-16-UTP (ENZO Diagnostics, Farmingdale, USA, 目录号分别为 42818 和 42814)各 1.875 mM 的酶混合物在 37℃温育 6 小时。单核苷酸和短寡核苷酸通过 CHROMA SPIN +STE-100 柱(Clontech, Palo Alto, USA, 目录号 K1302-2)柱层析除去, 洗脱物中的 cRNA 通过添加乙醇而沉淀。cRNA 的质量依据琼脂糖凝胶电泳后 cRNA 的长度判断。纯化的 cRNA 在 70%乙醇溶液中贮存于-80℃备用。

### 实施例 4.有复发和没有复发的患者的肿瘤基因表达分析

来自肝癌患者的人原发性肿瘤的基因表达通过高密度寡核苷酸微阵列检测 (HuGeneFL array, Affymetrix, Santa Clara, USA, 目录号 510137) (Lipshutz, R.L. 等 High density synthetic oligonucleotide arrays, Nature Genet.21, 20-24 (1999))。为与芯片上的寡核苷酸杂交, cRNA 于 95℃在含有 40 mM Tris (Sigma, St.Louis, USA, 目录号 T1503) -乙酸 (Wako, Osaka, Japan, 目录号 017-00256) (pH8.1)、100 mM 乙酸钾 (Wako, Osaka, Japan, 目录号 160-03175)、和 30mM 乙酸镁 (Wako, Osaka, Japan, 目录号 130-00095) 的缓冲液中片断化 35 分钟。杂交在 200 μl 含有 0.1M 2-(N-吗啉代)乙磺酸(MES)

(Sigma, St. Louis, USA, 目录号 M-3885) (pH6.7)、1M NaCl (Nacalai tescque, Tokyo, Japan, 目录号 313-20)、0.01% 聚氧烯(polyoxylyene) (10)辛基酚醚 (Wako, Osaka, Japan, 目录号 168-11805)、20  $\mu\text{g}$  鲑鱼精 DNA (Promega, Madison, USA, 目录号 D181B)、100  $\mu\text{g}$  乙酰化牛血清白蛋白 (Sigma, St.Louis, USA, 目录号 B-8894)、10  $\mu\text{g}$  片断化的 cRNA、以及生物素标记的对照寡核苷酸, 生物素-5'-CTGAACGGTAGCATCTTGAC-3' (Sawady technology, Tokyo, Japan) 的缓冲液中于 45 $^{\circ}\text{C}$  进行 12 小时。在用含有 0.01M MES (pH6.7)、0.1M NaCl、0.001%聚氧烯(10)辛基酚醚缓冲液的缓冲溶液洗涤芯片后, 芯片与生物素标记的抗-链抗生物素蛋白抗体 (Funakoshi, Tokyo, Japan, 目录号 BA0500)温育, 并用链抗生物素蛋白 R-藻红蛋白着色 (Molecular Probes, Eugene, USA, 目录号 S-866), 以便提高杂交信号, 正如操作说明书中所述的那样 (Affymetrix, Santa Clara, USA)。用激光扫描仪(Affymetrix, Santa Clara, USA) 收集每个像素水平, 并且用 Affymetrix GeneChip ver.3.3 和 Affymetrix Microarray Suite ver.4.0 软件计算每种 cDNA 的表达水平和可信度 (Present/Absent call)。从这个实验中, 测定了肝癌患者的人原发性肿瘤中 6000 个基因的表达。

### 实施例 5. 动态 RT-PCR 分析

基因的表达还通过动态 RT-PCR 来测定。动态 RT-PCR 通过实时荧光 PCR 系统进行。利用 LightCycler 仪 (LightCycler system, Roche Diagnostics, Mannheim, Germany, 目录号 2011468), 在 20  $\mu\text{l}$  由主混合物和缓冲液 (LightCycler DNA Master hybridization probes, Roche Diagnostics, Mannheim, Germany, 目录号 2158825), 4 mM 氯化镁 (Nacalai tescque, Tokyo, Japan, 目录号 7791-18-6), 10 皮摩尔 PCR 引物 (Sawady Technology, Tokyo, Japan), 4 皮摩尔荧光杂交探针 (Nihon Genome Research Laboratories, Sendai, Japan) (其被设计成在扩增产物链上以头对尾的排列与靶序列杂交), 以及 2  $\mu\text{l}$  模板 cDNA 组成的反应混合物中、在 LightCycler 毛细管 (Roche Diagnostics, Mannheim, Germany, 目录号 1909339) 中进行 PCR 扩增。供体探针在 3'-端由荧光标记, 而受体探针在 5'-端由 LC-Red640 标记并且在 3'-端进行磷酸化修饰以阻断延伸。供体探针 3'-端与受体探针 5'-端间的缺口有 1-3 个碱基。扩增前, 向反应混合物中加入 0.16  $\mu\text{l}$  TaqStart 抗体 (Clontech, Palo Alto, USA,

目录号 5400-1), 接着室温下温育 10 分钟以阻断引物延伸。然后, 在 95°C 温育 90 秒使抗体失活, 扩增在 LightCycler 中按照 95°C 温育 0 秒变性, 在 57-60°C 温育 3-10 秒退火, 以及在 72°C 温育 90 秒延伸进行 40 个循环, 其中温度斜率为 20°C/秒。通过在每一扩增循环的退火期终止时测量荧光信号来实现实时 PCR 监测。为使分离的 RNA 的完整性合格并使靶序列的拷贝数标准化, 还利用杂交探针进行了甘油醛-3-磷酸脱氢酶(GAPDH)的动态 RT-PCR 分析。靶 mRNA 和 GAPDH mRNA 的外标物由质粒 DNA 的 10-倍系列稀释物( $10^3$  到  $10^8$ )制备。每个样品中 mRNA 的定量通过参考在每个时间点构建的标准曲线根据 LightCycler 软件自动进行 (LightCycler software version 3, Roche Diagnostics, Mannheim, Germany)。

#### 实施例 6. 鉴定基因集合, 其表达能够区分具有早期肝内复发的肝癌患者和不具有早期肝内复发的肝癌患者

早期肝内复发往往与原发性肿瘤的数目和 TNM 阶段相关联, p 值分别为 0.041 和 0.006, 但是与其它临床病理学因素无关(表 1)。外科手术时原发性肿瘤的数目仅在有限的灵敏度和特异性方面将 A 组和 B 组区分开来(分别为 62% 和 75%)。TNM 分期对于分开 A 组和 B 组同样具有有限的灵敏度(67%)和特异性(83%)。因此, 显然这些传统的分类法不能够预测早期肝内复发。

统计模式识别中的监控记忆被用来分析高密度寡核苷酸微阵列的数据。评定体系用训练样品设计并用测试样品验证其功效(图 1)。为保持训练和测试样品的独立性, 采取了交换训练和测试样品的交叉验证法。33 个可用样品通过交叉验证法被分成 30 个训练样品和 3 个测试样品(图 1, 步骤 1)。根据先验概率, 建立了由 A 组的 11 个样品和 B 组的 19 个样品组成的 10 组训练样品。结果, 建立了 10 组 3 个测试样品, 由 A 组的一个样品和 B 组的两个样品组成。

利用 Fisher 判据, 从所有的检测基因中选择在 A 组和 B 组之间具有大于两倍的中值均差的 50 个基因来建立预测评定体系(图 1, 步骤 2-3), 该判据由下列公式(I)给出。

$$F(i) = \frac{(\mu_A(i) - \mu_B(i))^2}{P(A)\sigma_A^2(i) + P(B)\sigma_B^2(i)}$$

其中,  $\mu_A(i)$ 是 A 组样品均值向量 $\mu_A$ 的第 i 个分量,  $\sigma_A^2(i)$ 是 A 组样品协方差矩阵 $\Sigma_A$ 的第 i 个对角元素, 而  $P(A)$ 是 A 组的先验概率。

然后, 如下所述鉴定用于评定体系的最佳基因子集。

Fisher 线性分类程序依下列公式(II)将测试样品 x 分配到 A 组。

如果 $F_A(x) < F_B(x)$

其中

$$F_A(x) = \frac{1}{2}(x - \mu_A)^T \Sigma_W^{-1} (x - \mu_A) - \ln P(A)$$

$$\Sigma_W = (P(A)\Sigma_A + P(B)\Sigma_B)$$

在排除一个的方法中, 样品均值向量、样品协方差矩阵、以及先验概率是通过利用 29 个样品作为训练样品估计的。然后, 将剩余的样品作为假测试样品对所生成的 Fisher 线性分类程序进行测试。重复这一操作 30 次。计算每一种可能的基因子集的错误率。例如, 当从 50 个基因中选择 5 个时, 待检测的子集数目为二百万。

下一步, 选择错误率最小化了的候选基因子集 (图 1, 步骤 4)。这个试验独立重复 10 次 (图 1, 步骤 5)。

在这些候选基因子集中, 选择在 10 次试验中自始至终出现最为频繁的基因子集作为区分两组的最佳基因子集(图 1, 步骤 6)。利用所选择的最佳基因子集, 分值 T 可通过下列公式(III)给出。

$$T(x) = F_A(x) - F_B(x)$$

在这个评定体系中, 具有负 T 值的所有 HCC 被归为 A 组(早期肝内复发组), 而具有正 T 值的所有 HCC 被归为 B 组(无复发组)。

最佳基因数目根据由下列公式(IV)给出的判据 J 确定(图 1, 步骤 7)。

$$J = \frac{1}{30} \left[ \sum_{x \in B} T(x) - \sum_{x \in A} T(x) \right]$$

判据 J 衡量 A 组和 B 组的可分性。针对不同基因数目计算 10 个不同训练集中 J 值的平均数和 95% 可信度间隔 (图 2)。可分性随着基因数目的增加而变得更好。当基因数目达到 10 和 12 时, 95% 可信度间隔变得几乎类似, 表明 12 对于两组的可分性而言是最为适当的基因数目(图 2)。

**实施例 7.12 个基因的最佳子集, 其表达可指示早期肝内复发**

根据上述算法, 鉴定了区分 A 组和 B 组的 12 个基因的最佳子集。最佳基因子集包括血小板衍生生长因子受体 $\alpha$  (PDGFRA)、肿瘤坏死因子 $\alpha$  (TNF- $\alpha$ ) 诱导蛋白 A20、溶酶体相关多跨膜蛋白(LAPTm5)、HLA-DR  $\alpha$ 重链 (II 类抗原)、rel 原癌基因、Staf50、公认的丝氨酸/苏氨酸蛋白激酶、MADS/MEF2-家族转录因子(MEF2C)、来自 3p21.3 的 HUMLUCA19 人粘粒克隆 LUCA19、DEAD-盒蛋白 p72、波形蛋白和 KIAK0002 的基因 (表 2)。在所选的 12 个基因中, 有 11 个的表达在 A 组中是下调的; A 组中这些基因的均差中值比 B 组的一半还少 (图 3)。相反, HUMLUCA19 基因的表达在 A 组中是上调的; A 组中 HUMLUCA19 基因的均差中值比 B 组增加了 3 倍多(图 3)。早期肝内预测分值的准确性是用每组 3 个测试样品的 10 组不同集合评估的 (图 4)。HCC 早期复发是通过计算来自 HCC 患者的 12 个基因的 T 值预测的。当 T 值低于 0 时, 在外科手术后一年之内复发是很有可能, 而当 T 值大于 0 时, 在外科手术后一年之内复发的可能性很小。这个评定体系在全部 10 个试验中可以完美地预测 3 个测试样品的早期肝内复发 (图 4)。该评定体系不依赖于病毒感染模式, 并且比 TNM 分期系统更加准确 (图 4)。基于全部 33 个 HCC 以及上述 12 个基因的评定体系(图 5)包括下列公式 (V)。

公式(V)

$$T(x) = 0.053862x_1 + 0.038848x_2 + 0.030176x_3 + 0.001824x_4 + 0.096997x_5 + 0.017259x_6 + 0.015908x_7 + 0.103081x_8 - 0.093746x_9 + 0.024031x_{10} - 0.005417x_{11} - 0.119177x_{12} - 11.046007,$$

其中  $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}$  为血小板衍生生长因子受体 $\alpha$ (PDGFRA)、肿瘤坏死因子 $\alpha$  (TNF- $\alpha$ )诱导蛋白 A20、溶酶体相关多跨膜蛋白(LAPTm5)、HLA-DR  $\alpha$ 重链、rel 原癌基因、Staf50、公认的丝氨酸/苏氨酸蛋白激酶、MADS/MEF2-家族转录因子(MEF2C)、来自 3p21.3 的 HUMLUCA19 人粘粒克隆 LUCA19、DEAD-盒蛋白 p72、波形蛋白和 KIAK0002 基因的 mRNAs 的校准均差(表 2)。

本发明所选择的 12 个基因参与广范围的生物学过程。其中, 免疫应答相关基因例如 HLA-DR  $\alpha$ 重链、肿瘤坏死因子 $\alpha$ 诱导蛋白 A20 和 Staf50, 在具有早期肝内复发的 HCC 中是下调的。由于 HLA-DR  $\alpha$ 重链被认为在巨噬细胞的抗原递呈中起着重要作用 (Tissot, C.& Mechti, N.Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus

type 1 long terminal repeat expression, *J.Biol.Chem.*270, 14891-14898 (1995) ), 它在肿瘤组织中的下调可能帮助肿瘤细胞逃脱宿主的免疫监视。Rel 原癌基因与 NF- $\kappa$ B 一样参与胞内信号传导途径, 它在具有早期肝内复发的 HCC 中也是下调的。此外, rel/NF- $\kappa$ B 的表达据报道与 T 细胞激活相关 (Mora, A 等 NF-kappa B/Rel participation in the lymphoidne-dependent proliferation of T lymphoid cells, *J.Immunol.*166, 2218-2227 (2000) )。因此, 看起来本发明所选择的用于预测早期肝内复发的许多基因参与削弱宿主抗高转移潜力 HCC 细胞的免疫应答。

还通过寡核苷酸微阵列分析了其他跟踪期最近达到 1 年的 HCC 患者的基因表达模式, 并根据上述公式计算了 12 个基因的表达分值。存活并且外科手术一年多以后没有复发的患者的 T 值为阳性(正值), 而其他在外科手术后一年内具有肝内复发的患者的 T 值为阴性(负数)。因此, 由从 6000 个基因中获得的 12 个基因的子集构成的评定体系能够准确地预测早期肝内复发。将统计学模式识别中的监控记忆应用到临床标本上, 可以预先提供预防、诊断、以及治疗其它疾病和 HCC 的关键信息。此外, 不仅仅 DNA 微阵列, 而且其它方法例如 RT-PCR 也可用来测定基因的最佳集合的表达。

表 1  
用于早期肝内复发的 HCC 临床病理学因素

因素	A 组(n=12)	B 组 (n=21)	P 值
性别			N.S.
男	8	16	
女	4	5	
年龄			N.S.
≤60	5	7	
>60	7	14	
病毒感染			N.S.
HBV	3	4	
HCV	8	14	
非-B 非-C	1	3	
原发病灶			0.041
单发肿瘤	3	13	
多发肿瘤	9	8	
肿瘤大小(cm)			N.S.
<2.0	0	5	
2.0-5.0	8	14	
>5.0	4	2	
分期*			0.006
I/II	2	14	
III/IV	10	7	
组织学分级*			N.S.
G1	0	2	
G2	9	17	
G3	3	2	
静脉入侵*			N.S.
(-)	7	18	
(+)	5	3	
非肿瘤性肝 非特异性改变			N.S.
慢性肝炎	1	1	
肝硬化	2	10	
肝硬化	9	10	

\*，基于 UICC 的 TNM 分类进行的评定

HBV: 乙型肝炎病毒, HCV: 丙型肝炎病毒, 非-B 非-C: 既不是 HBV 也不是 HCV

A 组: 早期肝内复发(+), B 组: 早期肝内复发(-)

N.S.: 不显著

表 2  
预测早期肝内复发的公式和 12 个基因

公式		
$T(x) = 0.053862x_1 + 0.038848x_2 + 0.030176x_3 + 0.001824x_4 + 0.096997x_5 + 0.017259x_6 + 0.015908x_7 + 0.103081x_8 - 0.093746x_9 + 0.024031x_{10} - 0.005417x_{11} - 0.119177x_{12} - 11.046007$		
GB*		描述
x1;	M21574	血小板衍生生长因子受体 $\alpha$ (PDGFRA)
x2;	M59465	肿瘤坏死因子 $\alpha$ 诱导蛋白 A20
X3;	U51240	溶酶体相关多跨膜蛋白 (LAPTm5)
x4;	X00274	HLA-DR $\alpha$ 重链 (II类抗原)
x5;	X75042	rel 原癌基因
x6;	X82200	Staf50
x7;	Y10032	公认的丝氨酸/苏氨酸蛋白激酶
x8;	L08895	MADS/MEF2-家族转录因子 (MEF2C)
x9;	AC000063	来自 3p21.3 的 HUMLUCA19 人粘粒克隆 LUCA19
x10;	U59321	DEAD-盒蛋白 p72
x11;	Z19554	波形蛋白
x12;	D13639	KIAK0002 基因

GB\*: 基因库登录号

- 步骤 1.通过交叉验证法将 33 个可用样品分成 30 个训练样品和 3 个测试样品。
- 步骤 2.利用 30 个训练样品计算 Fisher 判据。
- 步骤 3.以 Fisher 判据和倍数改变为基础选择区分 A 和 B 组的较好的 50 个基因。
- 步骤 4.用遗漏一个的方法通过详尽的研究从 50 个基因中选择候选基因子集。
- 步骤 5.独立重复步骤 1-4 (10 次)。
- 步骤 6.选择在 10 次试验中最频繁出现的最佳基因子集。
- 步骤 7.根据用 10 组不同的训练样品计算的 J 判据确定最佳的基因数目。
- 
- 步骤 8.利用步骤 1 中选择的 30 个训练样品设计评定体系。
- 步骤 9.由步骤 8 中设计的评定体系区分步骤 1 中选择的 3 个测试样品。
- 步骤 10.用 10 组不同的训练和测试样品独立重复步骤 8 和 9 共 10 次。

图 1

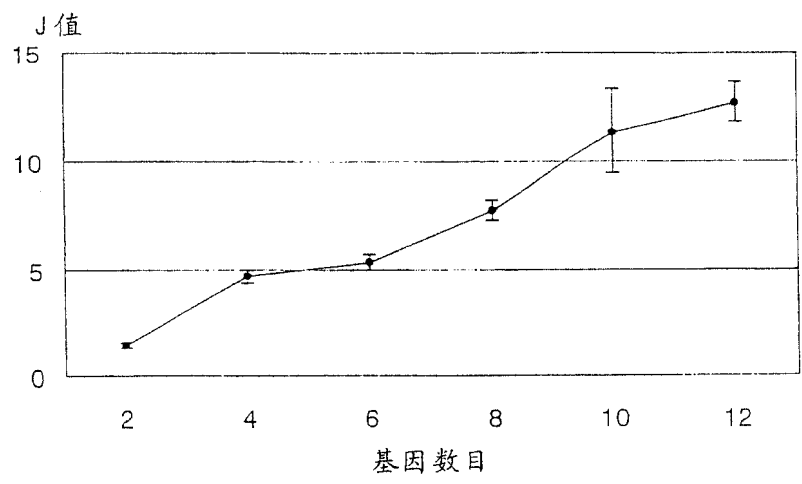


图 2

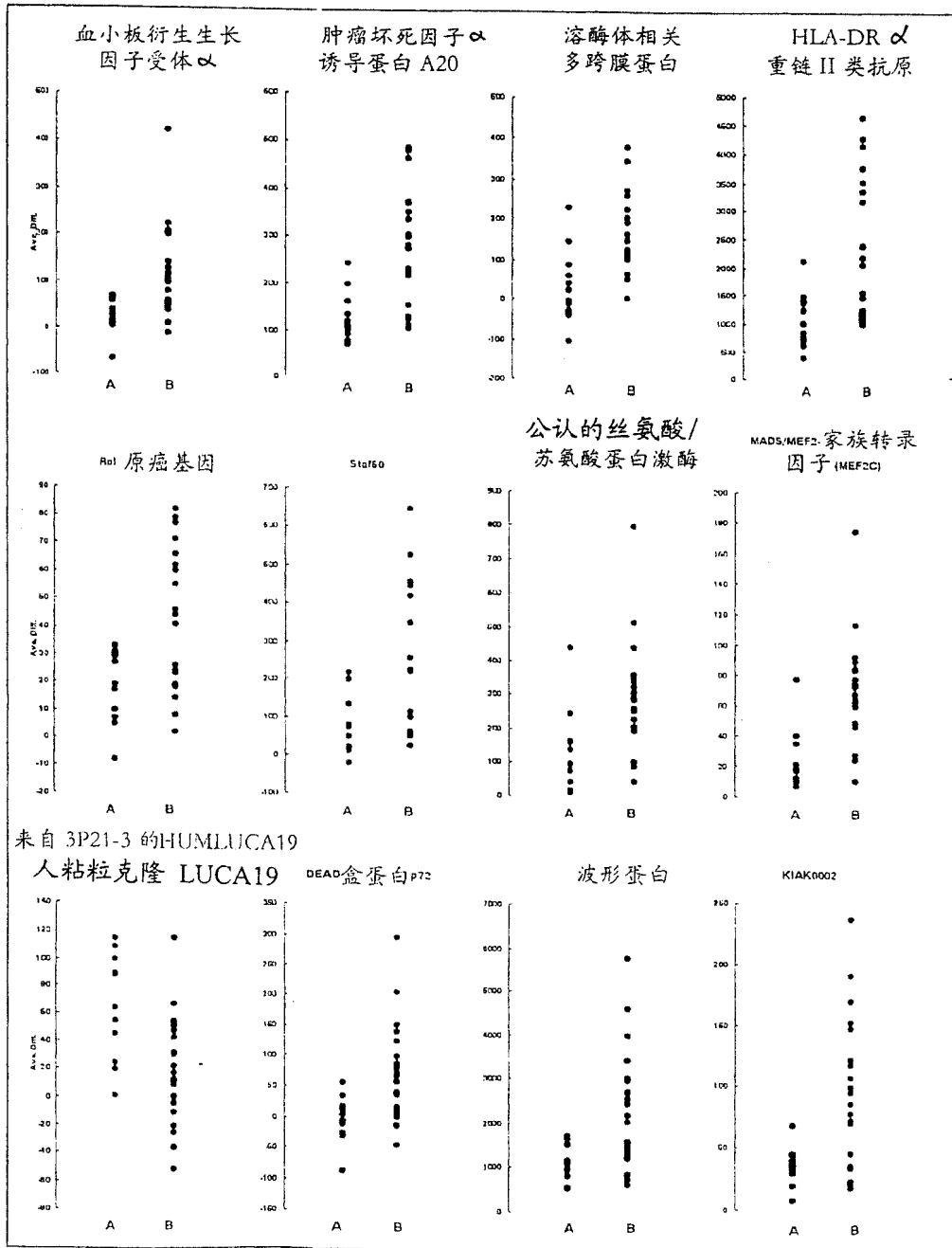


图 3

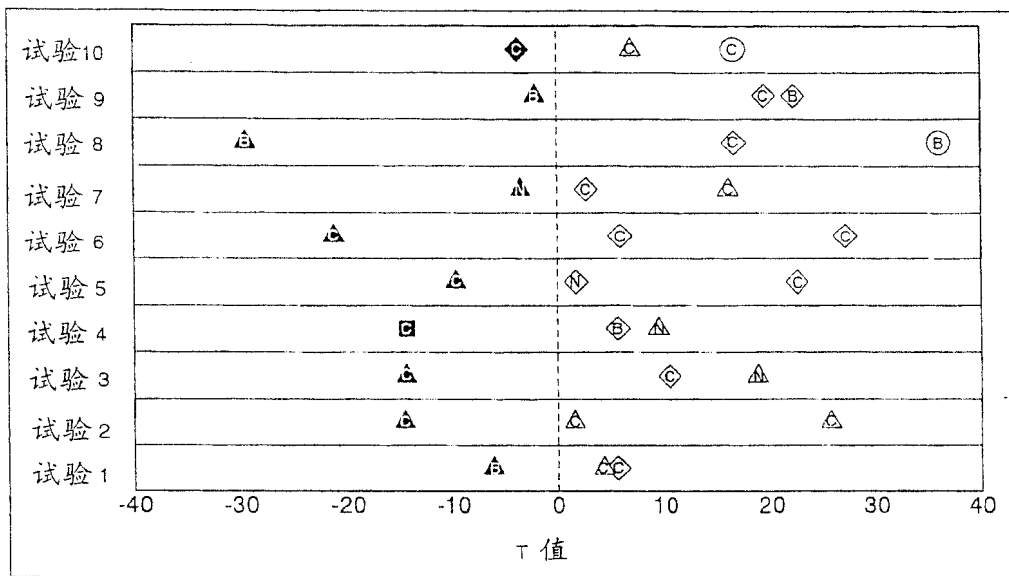


图 4

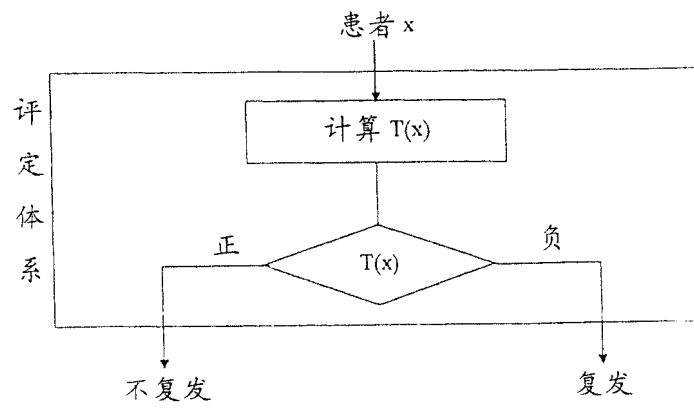


图 5

专利名称(译)	用于预测癌症复发的评定体系		
公开(公告)号	<a href="#">CN100398663C</a>	公开(公告)日	2008-07-02
申请号	CN01823658.8	申请日	2001-07-23
申请(专利权)人(译)	霍夫曼-拉罗奇有限公司		
当前申请(专利权)人(译)	霍夫曼-拉罗奇有限公司		
[标]发明人	冈正朗 浜本义彦 冈部尚文		
发明人	冈正朗 浜本义彦 冈部尚文		
IPC分类号	C12Q1/68 G06F19/00 G01N33/574 G01N33/53 C12N15/09 G01N37/00		
CPC分类号	C12Q1/6809 C12Q1/6886 G06F19/18 G01N33/574 G06F19/20 C12Q2600/118 C12Q1/6837 G16B20/00 G16B25/00		
审查员(译)	周洋		
其他公开文献	CN1549864A		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

#### 摘要(译)

本发明涉及预测癌症复发的评定体系。更具体地，本发明涉及选择基因和/或蛋白，并通过测量人肿瘤组织中基因和/或蛋白的表达，以及对它们的表达模式与具有癌症复发以及没有癌症复发的患者的人原发性肿瘤中基因和/或蛋白的表达模式进行比较，利用所选的基因和/或蛋白产生用于预测癌症复发的公式。本发明还涉及一种用于实施本发明的方法的试剂盒，包含DNA芯片、寡核苷酸芯片、蛋白芯片、肽、抗体、探针和引物，这些是实现DNA微阵列、寡核苷酸微阵列、蛋白阵列、NORTHERN印迹、原位杂交、核糖核酸酶保护分析、蛋白质印迹、反转录聚合酶链式反应(以下称之为RT-PCR)，以检测指示癌症复发的至少2个或更多个基因和/或蛋白，优选4个或更多个基因和/或蛋白，更优选6个或更多个基因和/或蛋白，并且最优选12个或更多个基因和/或蛋白的表达所必需的。