(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2023 年 3 月 23 日 (23.03.2023)

WIPO | PCT

(10) 国际公布号
WO 2023/040102 A1

(51) 国际专利分类号:
*G16H 50/20* (2018.01)

(21) 国际申请号: PCT/CN2021/139502

(22) 国际申请日: 2021 年 12 月 20 日 (20.12.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202111089547.X 2021年9月16日 (16.09.2021) CN

(71) 申请人: 浙江大学(ZHEJIANG UNIVERSITY) [CN/CN]; 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。

(72) 发明人: 徐俊杰(XU, Junjie); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。 蔡秀军(CAI, Xiujun); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。 茅棋江(MAO, Qijiang); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。 潘浩奇(PAN, Haoqi); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。 梁霄(LIANG, Xiao); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。

(74) 代理人: 杭州求是专利事务所有限公司(HANGZHOU QIUSHI PATENT OFFICE CO., LTD.); 中国浙江省杭州市西湖区古墩路701号紫金广场C座1506室邱启旺, Zhejiang 310030 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,

(54) **Title:** GENE MODEL FOR DETERMINING PROGNOSIS OF PATIENT WITH HEPATIC CELL CARCINOMA, AND CONSTRUCTION METHOD THEREFOR AND USE THEREOF

(54) **发明名称:** 判断肝细胞肝癌患者预后的基因模型、构建方法和应用

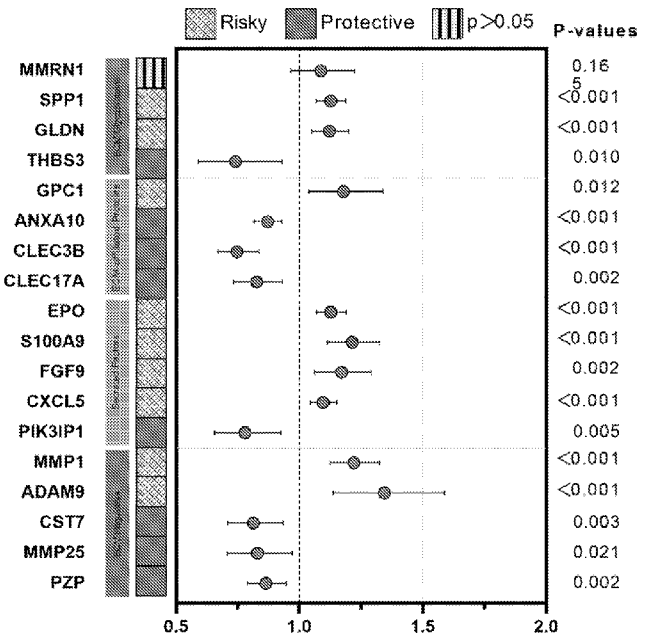


图 3

(57) **Abstract:** Disclosed in the present invention are a gene model for determining the prognosis of hepatic cell carcinoma, and a construction method therefor and the use thereof. In the present intention, the construction method comprises: comparing data of a sample of a patient with hepatic cell carcinoma with transcriptome data of a sample of a normal patient to obtain a gene with differential expression, integrating the gene with an extracellular matrix gene set, and then shrinking same by means of a LASSO-COX regression

WO 2023/040102 A1

[见续页]

BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

**(84) 指定国**(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

**本国际公布**：

— 包括国际检索报告(条约第21条(3))。

— 包括说明书序列表部分(细则5.2(a))。

model to obtain a model of 18 genes. The model of the present invention can evaluate the prognosis of the patient with hepatic cell carcinoma, and distinguish and select a patient with hepatic cell carcinoma having a poor prognosis, thereby guiding a clinician to provide a more positive treatment plan, and overtreatment on the patient with hepatic cell carcinoma at a low risk can also be avoided. The gene model helps to construct a tissue chip on the basis of an extracellular matrix gene, such that the prognosis of a patient after hepatic cell carcinoma surgery can be quickly evaluated, and clinical transformation is achieved.

**(57) 摘要**：本发明公开了一种判断肝细胞肝癌预后的基因模型及其构建方法和应用。本发明通过比较肝细胞肝癌患者样本的数据和正常患者样本的转录组数据，得到具有差异表达的基因，与细胞外基质基因集整合后通过LASSO-COX回归模型缩小得到一个18个基因的模型。本发明模型可以对肝细胞肝癌患者的预后进行评估，区分并甄选出预后差的肝细胞肝癌患者，从而指导临床医生提供更积极的治疗方案，同时也能避免对低风险的肝细胞肝癌患者进行过度治疗。通过该基因模型有助于构建一款基于细胞外基质基因的组织芯片，能够对肝细胞肝癌术后患者快速进行预后进行评估，实现临床转化。

# 判断肝细胞肝癌患者预后的基因模型、构建方法和应用

## 技术领域

**[0001]** 本发明属于生物医学技术领域，具体涉及一种用于判断肝细胞肝癌患者预后的基因模型及应用。

## 背景技术

**[0002]** 肝癌是世界范围内最常见的十大恶性肿瘤之一。全球每年约有 50 万新增病例，其中肝细胞肝癌占 85%。随着肿瘤标志物与影像学检查的推广、外科手术水平和多种新型治疗方式如动脉内化疗栓塞等的发展，原发性肝癌（hepatic cell carcinoma，HCC）的 5 年生存率有所改善。但总体来说，肝细胞肝癌的预后仍然不尽如人意。其中主要原因之一就是缺少有效的预测肝细胞肝癌患者预后的标志物，从而无法将肝细胞肝癌患者进行风险分层，无法指导临床医生对高风险肝细胞肝癌患者进行早干预早治疗。当前研究表明，肿瘤微环境，尤其是细胞外基质，能够促进肿瘤的生长、侵袭和转移，对肿瘤患者的预后具有较大的影响。

## 发明内容

**[0003]** 针对当前临床缺少有效的判断肝细胞肝癌患者预后的标志物，无法对肝细胞性肝癌患者预后进行判断，本发明从基因层面构建基因组合模型来评估肝细胞肝癌患者的预后，通过对肝细胞肝癌患者的细胞外基质相关基因进行整合分析，从而构建相关基因组合模型，构建一款基于细胞外基质基因的组织芯片，可以实现通过风险评分对肝细胞肝癌患者的预后进行评估。评估得到的结果有助于临床医生对肝癌患者进行分层，为肝细胞肝癌患者的精准治疗提供了可能性。

**[0004]** 本发明采用的方案具体如下：

一种判断肝细胞肝癌患者预后的基因模型的构建方法，包括如下步骤：

(1)、获取肝细胞肝癌和正常肝脏组织样本的转录组数据，比较肝细胞肝癌组织样本的数据和正常肝脏组织样本的数据中的差异基因，设定 P-value＜0.05 得到具有显著差异的基因，并将具有显著差异的基因与细胞外基质基因集（559 个细胞外基质相关基因）进行整合；

(2)、随后使用 LASSO 方法进行分析，基于 R 语言 glmnet 包使用 1000 次 Cox LASSO 回归迭代和 10 倍交叉验证，将种子基因缩小为 18 个与 HCC 预后相关的 ECM 基因集，包括：MMP1、EPO、MMRN1、S100A9、ADAM9、GPC1、SPP1、GLDN、FGF9、CXCL5、CST7、THBS3、ANXA10、PIK3IP1、MMP25、CLEC3B、PZP 和 CLEC17A 的 18 个基因组合（表 1），并以 18 个基因为标志物构建获得肝细胞肝癌预后预测风险评分模型。

**[0005]** 一种上述构建方法构建获得的基因模型，具体为：

肝细胞肝癌患者风险评分=(0.069× MMP1 表达水平)+(0.049×EPO 表达水平)+(0.042×MMRN1 表达水平)+(0.036×S100A9 表达水平)+(0.027×ADAM9 表达水平)+(0.024×GPC1 表达水平)+(0.021×SPP1 表达水平)+(0.014×GLDN 表达水平)+(0.007×FGF9 表达水平)+(0.001×CXCL5 表达水平)-(0.024×CST7 表达水平)-(0.027×THBS3 表达水平)-(0.042×ANXA10 表达水平)-(0.049×PIK3IP1 表达水平)-(0.051×MMP25 表达水平)-(0.054×CLEC3B 表达水平)-(0.062×PZP 表达水平)-(0.069×CLEC17A 表达水平)。

**[0006]** 进一步地，将 TCGA 数据库作为训练集，GEO 数据库和 ICGC 数据库作为验证集，对所述基因模型的风险评分进行分析并通过 CLIP 分期和 TMN 分期对基因模型进行验证，表明所述肝细胞肝癌患者风险评分与生存期相关，风险评分高的患者生存期短、预后差。

**[0007]** 一种上述基因模型在评估肝细胞肝癌预后中的应用。

**[0008]** 一种基于细胞外基质基因的组织芯片，所述组织芯片包含检测 MMP1、EPO、MMRN1、S100A9、ADAM9、GPC1、SPP1、GLDN、FGF9、CXCL5、CST7、THBS3、ANXA10、PIK3IP1、MMP25、CLEC3B、PZP 和 CLEC17A 的探针。为肝细胞肝癌患者提供精准治疗提供了可能性。能够对肝细胞肝癌术后患者快速进行预后评估，实现临床转化。

**[0009]** 本发明的有益效果是：本发明构建了 18 个基因的基因组合模型，通过该基因模型可以构建一款基于细胞外基质基因的组织芯片，对肝细胞肝癌患者的预后进行评估，能够区分并甄选出预后差的肝细胞肝癌患者，即将肝细胞肝癌患者进行分层，筛选出高风险、预后差的肝细胞肝癌患者，指导临床医生对高风险患者提供更积极的治疗方案，同时也能避免对低风险的肝细胞肝癌患者进行过度治疗。

**附图说明**

**[0010]** 下面结合附图和实施例对本发明进一步说明；

图 1 为本发明的基因模型的差异基因火山图；

图 2 为本发明的基因模型的 LASSO-Cox 回归模型构建图；

图 3 为本发明的 18 个基因的组合基因模型图；

图 4 为训练集 TCGA 中肝细胞肝癌患者的风险评分分布图；其中横坐标为根据风险评分递增的患者序号，虚线为截断值；

图 5 为训练集 TCGA 中肝细胞肝癌患者的生存期分布图；其中横坐标为根据风险评分递增的患者序号，190 附近虚线为截断值，120 附近和 250 附近为死亡和存活差异明显的分界线；

图 6 为验证集 GEO 中肝细胞肝癌患者的生存期分布图；其中横坐标为根据风险评分递增的患者序号，虚线为截断值；

图 7 为验证集 ICGC 中肝细胞肝癌患者的生存期分布图；其中横坐标为根据风险评分递增的患者序号，虚线为截断值；

图 8 为训练集 TCGA 中肝细胞肝癌患者不同 CLIP 分期的风险评分结果图；

图 9 为训练集 TCGA 中肝细胞肝癌患者不同 TMN 分期的风险评分结果图；

图 10 为训练集 TCGA 中基于本发明基因模型分组后肝细胞肝癌患者预后与生存期的关系图；

图 11 为训练集 TCGA 中基于本发明基因模型分组的肝细胞肝癌患者预后的敏感性和特异性结果图；

图 12 为验证集 GEO 中基于本发明基因模型分组的肝细胞肝癌患者预后与生存期的关系图；

图 13 为验证集 ICGA 中基于本发明基因模型分组的肝细胞肝癌患者预后与生存期的关系图。

**具体实施方式**

**[0011]** 本发明提供了一种基于细胞外基质基因预测肝细胞肝癌预后的基因模型及应用。即针对肝细胞肝癌患者细胞外基质的差异基因，利用数据库中的肝细胞肝癌组织样本及正常肝脏组织样本数据和统计学分析建立肝细胞肝癌预后的风险模型，能够作为预测肝细胞肝癌患者预后的基因模型，从而构建一款基于细胞外基质基因的组织芯片，有助于对肝细胞肝癌术后患者预后进行评估。其中肝细胞肝癌组织样本纳入及排除的标准为：

（1）术前未曾接受过其他癌症治疗；

（2）无其他恶性肿瘤病史；

（3）具有完善的临床病理资料和随访信息。

**[0012]** 下面结合具体的实施例对本发明的效果作进一步说明。

**[0013]** 实施例 1：构建用于判断肝细胞肝癌患者预后的基因模型

本发明的用于判断肝细胞肝癌患者预后的基因模型，通过如下步骤构建获得：

(1)、先从 TCGA 数据库(https://portal.gdc.cancer.gov/)中下载 371 例肝细胞肝癌组织样本和 50 例正常肝脏组织样本的转录组数据及对应患者的临床信息（包括性别、总生存时间、生存状态等），比较 TCGA 数据库中肝细胞肝癌组织样本的数据与正常肝脏组织样本中的差异基因，设定 P-value＜0.05，得到具有显著差异的基因，并将具有显著差异的基因与 559 个细胞外基质（ECM）相关基因进行整合(见图 1)。

**[0014]** (2)、随后使用 LASSO 方法进行分析，基于 R 语言 glmnet 包使用 1000 次 Cox LASSO 回归迭代和 10 倍交叉验证，筛选出有统计意义的 18 个与 ECM 相关的候选基因及这些基因的预后 AUC、HR 值(见表 1 和图 2)。将 Cox LASSO 回归模型的系数作为权重，构建出一个基于包含 MMP1、EPO、MMRN1、S100A9、ADAM9、GPC1、SPP1、GLDN、FGF9、CXCL5、CST7、THBS3、ANXA10、PIK3IP1、MMP25、CLEC3B、PZP 和 CLEC17A 18 个基因为标志物的肝细胞肝癌预后预测风险评分模型(见图 3)。

**[0015]** 所述肝细胞肝癌预后预测风险评分模型具体为:肝细胞肝癌患者风险评分=(0.069×MMP1 表达水平)+(0.049×EPO 表达水平)+(0.042×MMRN1 表达水平)+(0.036×S100A9 表达水平)+(0.027×ADAM9 表达水平)+(0.024×GPC1 表达水平)+(0.021×SPP1 表达水平)+(0.014×GLDN 表达水平)+(0.007×FGF9 表达水平)+(0.001×CXCL5 表达水平)-(0.024×CST7 表达水平)-(0.027×THBS3 表达水平)-(0.042×ANXA10 表达水平)-(0.049×PIK3IP1 表达水平)-(0.051×MMP25 表达水平)-(0.054×CLEC3B 表达水平)-(0.062×PZP 表达水平)-(0.069×CLEC17A 表达水平)。

**[0016] 表 1 LASSO 回归模型后得到的 18 个 ECM 基因**

| 基因名称 | 疾病 | AUC | HR |
|---|---|---|---|
| MMP1 | 肝癌 | 0.628 | 1.220 |
| EPO | 肝癌 | 0.607 | 1.127 |
| MMRN1 | 肝癌 | 0.542 | 1.088 |
| S100A9 | 肝癌 | 0.589 | 1.213 |
| ADAM9 | 肝癌 | 0.586 | 1.344 |
| GPC1 | 肝癌 | 0.639 | 1.178 |
| SPP1 | 肝癌 | 0.614 | 1.127 |
| GLDN | 肝癌 | 0.603 | 1.122 |
| FGF9 | 肝癌 | 0.555 | 1.171 |
| CXCL5 | 肝癌 | 0.575 | 1.096 |
| CST7 | 肝癌 | 0.564 | 0.813 |
| THBS3 | 肝癌 | 0.623 | 0.741 |
| ANXA10 | 肝癌 | 0.629 | 0.870 |
| PIK3IP1 | 肝癌 | 0.559 | 0.779 |
| MMP25 | 肝癌 | 0.549 | 0.829 |
| CLEC3B | 肝癌 | 0.610 | 0.746 |
| PZP | 肝癌 | 0.608 | 0.863 |
| CLEC17A | 肝癌 | 0.593 | 0.826 |

实施例 2：肝细胞肝癌预后预测风险评分模型在评估肝细胞肝癌预后中的应用

将 TCGA 数据库 371 例肝细胞肝癌组织样本转录组数据作为训练集，GEO 数据库(https://www.ncbi.nlm.nih.gov/geo/)GSE140520 中 247 例肝细胞肝癌组织和 ICGC 数据库

（https://daco.icgc.org/）的 203 例肝细胞肝癌组织的数据作为验证集，根据风险评分模型分别计算训练集每个肝细胞肝癌患者的评分，取评分的中位数（0.044954）作为截断值将其分为高风险分值组和低风险分值组，绘制两组患者的风险评分和生存期、CLIP 分期和 TMN 分期的关系图（图 4-8），验证肝细胞肝癌预后预测风险评分模型的效果。其中，图 4 和图 5 为训练集 TCGA 中肝细胞肝癌患者根据截断值的风险评分分布图和生存期分布图，图 6 和图 7 为验证集 GEO 和 ICGC 中肝细胞肝癌患者根据截断值的生存期分布图，图 8 和图 9 为训练集 TCGA 中肝细胞肝癌患者不同 CLIP 分期和 TMN 分期的风险评分结果，可以看到风险评分越高，患者存活率越高，CLIP 分期与 TMN 分期越高，表明该模型具有良好的肝细胞肝癌分型效果。

**[0017]** 进一步地，通过 ROC 曲线评估模型的预测性能：图 10 为训练集 TCGA 数据库中肝细胞肝癌患者的预后与生存期的关系图，其中高风险分值组的肝细胞肝癌患者的生存期短，预后较低风险分值组患者差（见图 10），图 11、表 2 为该模型验证 HCC 预后的敏感性和特异性结果，风险模型的 3 年 AUC 为 0.81，敏感性为 73.7%，特异性为 75%；5 年 AUC 为 0.79，敏感性为 77.3%，特异性为 71.7%；用 GEO 数据库(https://www.ncbi.nlm.nih.gov/geo/)GSE140520 中 247 例肝细胞肝癌组织和 ICGC 数据库中的 203 例肝细胞肝癌组织的数据作为验证集进行验证（见图 12 和图 13），结果与 TCGA 数据库中结果一致，表 3 为该模型在 GEO 数据库中验证 HCC 预后的敏感性和特异性结果，风险模型的 3 年 AUC 为 0.626，敏感性为 68.8%，特异性为 55.8%；5 年 AUC 为 0.625，敏感性为 60.0%，特异性为 34.7%；表 4 为该模型在 ICGC 数据库钟验证 HCC 预后的敏感性和特异性结果，风险模型的 3 年 AUC 为 0.723，敏感性为 93.3%，特异性为 52.7%；5 年 AUC 为 0.717，敏感性为 88.9%，特异性为 52.3%；风险评分高的患者生存期短、预后差。表明本发明的肝细胞肝癌预后预测风险评分模型可以用于评估肝细胞肝癌预后。

**[0018] 表 2 TCGA 数据库风险模型的敏感性和特异性检验结果**

| 风险模型 | AUC | 敏感性 | 特异性 |
|---|---|---|---|
| 3 年 ROC | 0.81 | 73.7% | 75.0% |
| 5 年 ROC | 0.79 | 77.3% | 71.7% |

**表 3 GEO 数据库风险模型的敏感性和特异性检验结果**

| 风险模型 | AUC | 敏感性 | 特异性 |
|---|---|---|---|
| 3 年 ROC | 0.626 | 68.8% | 55.8% |
| 5 年 ROC | 0.625 | 60.0% | 34.7% |

**表 4 ICGC 数据库风险模型的敏感性和特异性检验结果**

| 风险模型 | AUC | 敏感性 | 特异性 |
|---|---|---|---|
| 3 年 ROC | 0.723 | 93.3% | 52.7% |
| 5 年 ROC | 0.717 | 88.9% | 52.3% |

本发明还提供了一种基因芯片，即：将检测 MMP1、EPO、MMRN1、S100A9、ADAM9、GPC1、SPP1、GLDN、FGF9、CXCL5、CST7、THBS3、ANXA10、PIK3IP1、MMP25、CLEC3B、PZP 和 CLEC17A 18 个基因的探针按上述模型构建成基因芯片，便于在临床中应用，其中，各基因探针序列优选如表 5 所示，针对一个基因的多个探针，可以选择探针测试结果的平均值作为该基因最终的表达水平。

**[0019] 表 5 基因芯片各基因探针序列**

| 基因名称 | 基因版本 | 基因序列 |
|---|---|---|
| MMP1 | HG19 | GGATCTGGAGATGGGCCGATAAAGTCAGTACGCAAAAGAAGAGTACGAAAGGACTAAACT（SEQ ID NO.1） |
| EPO | HG19 | GAACCATGAAGACAGGATGGGGGCTGGCCTCTGGCTCTCATGGGGTCCAAGTTTTGTGTA（SEQ ID NO.2） |
| MMRN1 | HG19 | GCTTGCATTTGAGTCTGAAAATATTAACAGTGAAATACACTGTGATAGGGTTTTAACTGG（SEQ ID NO.3） |
| S100A9 | HG19 | CACAAATGCAGACAAGCAGCTGAGCTTCGAGGAGTTCATCATGCTGATGGCGAGGCTAAC（SEQ ID NO.4） |
| ADAM9 | HG19 | ATTTCCGTTTCCATCATTGAATAAGTCTTATTCAGTCATCGGTGAGGTTAATGCACTAAT（SEQ ID NO.5） |
| GPC1 | HG19 | AGGTCCCCGGTTGCTGGTCAGGTCCCCATGGCTTGTTCTCTGGAACCTGACTTTAGATGT（SEQ ID NO.6） |
| SPP1 | HG19 | TTTTGAAGATAAACCGAAACCTTCCAAACAGTCACTTCAGTCTTACCAAGAGGCTTTGCA（SEQ ID NO.7） |
| GLDN | HG19 | TCTTGGCCCCTGTGTGAATTCCTGCCTTTCCCAGAAATGAGTCCAGGGTGTCTGACCTCA（SEQ ID NO.8）<br>GCACGTTGTTTACAACAACTCTCTCTACTACCACAAAGGGGGTTCTAATACCCTAGTGAG（SEQ ID NO.9） |
| FGF9 | HG19 | CAAAAGGACTGCGGCCTGATGCATGCTGGAAAAAGACACGCTTTTCATTTCTGATCAGTT（SEQ ID NO.10） |
| CXCL5 | HG19 | TTTGCTGTTATTTTATCTGCTATGCTATTGAAGTTTTGGCAATTGACTATAGTGTGAGCC（SEQ ID NO.11） |
| CST7 | HG19 | CATCACAAGGGCCCTAGTTCAGATAGTGAAAGGCCTGAAATATATGCTGGAGGTGGAAAT（SEQ ID NO.12） |
| THBS3 | HG19 | GGGGCGTCTTGGTGTATTCTGCTTCTCCCAAGAAAACATAATTTGGTCCAATCTCCAGTA（SEQ ID NO.13） |
| ANXA10 | HG19 | TGGTGATGCTGAGGACTACTAAAATGAAGAGGACTTGGAGTACTGTGCACTCCTCTTTCT（SEQ ID NO.14） |

| | | |
|---|---|---|
| PIK3IP1 | HG19 | TTTGTGTTCTGGTTAAAACCCTACCACTCCCCCGCTTTTTTGGCGAATCCTAGTAAGAGT（SEQ ID NO.15）<br>AAGGAAAGATTGTAATCTCACCTCCAGATCCATAGGCTTCTGCTAGGAGCATGTTGCTGA（SEQ ID NO.16）<br>AAATCTCACCATATCCATTTGACTTAGGCCTTTTGGAGTTAGGCAGAAGGGCCCTTCTTC（SEQ ID NO.17）<br>CCAACCTCAGTGGTTCAGTGGTGTCTGCAAAATGCATCCATCCTGCCATCTGAGTAGCAG（SEQ ID NO.18）<br>ATGAGAAGACATGGAATGTTGGAAGCAGCAACCGAAACAAAGCTGAAAACCTGTTGCGAG（SEQ ID NO.19）<br>TTTGTGTTCTGGTTAAAACCCTACCACTCCCCCGCTTTTTTGGCGAATCCTAGTAAGAGT（SEQ ID NO.20） |
| MMP25 | HG19 | CTGTGTGTTCTCTGGATCTTTTCAGCCCTGTGGTCCAGTGTCCATCACAGCCATGCTGAC（SEQ ID NO.21） |
| CLEC3B | HG19 | TCCTCTCCGTGCGCTTGGAGCCTCTTTTTGCAAATAAAGTTGGTGCAGCTTCGCGGAGAG（SEQ ID NO.22） |
| PZP | HG19 | CTGGTTTTATTCCCCTGAAACCAACAGTAAAAATGCTTGAAAGATCTAGCTCTGTGAGCC（SEQ ID NO.23） |
| CLEC17A | HG19 | ACATTGCCCGTGTAAGAGCTGACACCAACCAGTCCCTGGTGGAACTTTGGGGCTTATTAG（SEQ ID NO.24） |

上述对具体实施方式的描述是为了便于该技术领域的普通技术人员能理解和使用本发明。熟悉本领域技术人员显然可以容易的对这些具体实施方式做出各种修改，并把在此说明的一般原理应用到其他实施例中，而不必经过创造性的劳动。因此，本发明不限于上述具体实施方式。本领域技术人员根据本发明的原理，不脱离本发明的范畴所做出的改进和修改都应该在本发明的保护范围之内。

# 权 利 要 求 书

**1.** 一种基因组合在制备判断肝细胞肝癌预后的组织芯片中的应用，其特征在于，所述组织芯片包含检测 MMP1、EPO、MMRN1、S100A9、ADAM9、GPC1、SPP1、GLDN、FGF9、CXCL5、CST7、THBS3、ANXA10、PIK3IP1、MMP25、CLEC3B、PZP 和 CLEC17A 的探针，判断肝细胞肝癌预后的基因模型为：肝细胞肝癌患者风险评分=(0.069×MMP1 表达水平)+(0.049×EPO 表达水平)+(0.042×MMRN1 表达水平)+(0.036×S100A9 表达水平)+(0.027×ADAM9 表达水平)+(0.024×GPC1 表达水平)+(0.021×SPP1 表达水平)+(0.014×GLDN 表达水平)+(0.007×FGF9 表达水平)+(0.001×CXCL5 表达水平)-(0.024×CST7 表达水平)-(0.027×THBS3 表达水平)-(0.042×ANXA10 表达水平)-(0.049×PIK3IP1 表达水平)-(0.051×MMP25 表达水平)-(0.054×CLEC3B 表达水平)-(0.062×PZP 表达水平)-(0.069×CLEC17A 表达水平)。

2.根据权利要求 1 所述的应用，其特征在于，所述肝细胞肝癌患者风险评分与生存期相关，风险评分高的患者生存期短、预后差。

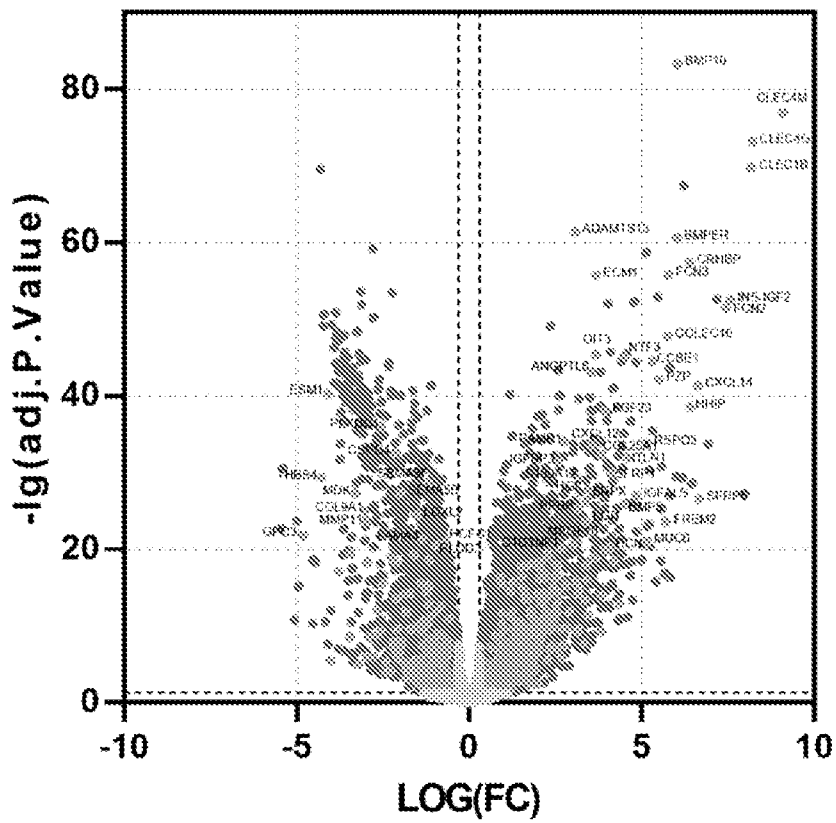
-lg(adj.P.Value)
80
60
40
20
0
-10
-5
0
5
10
LOG(FC)

图 1

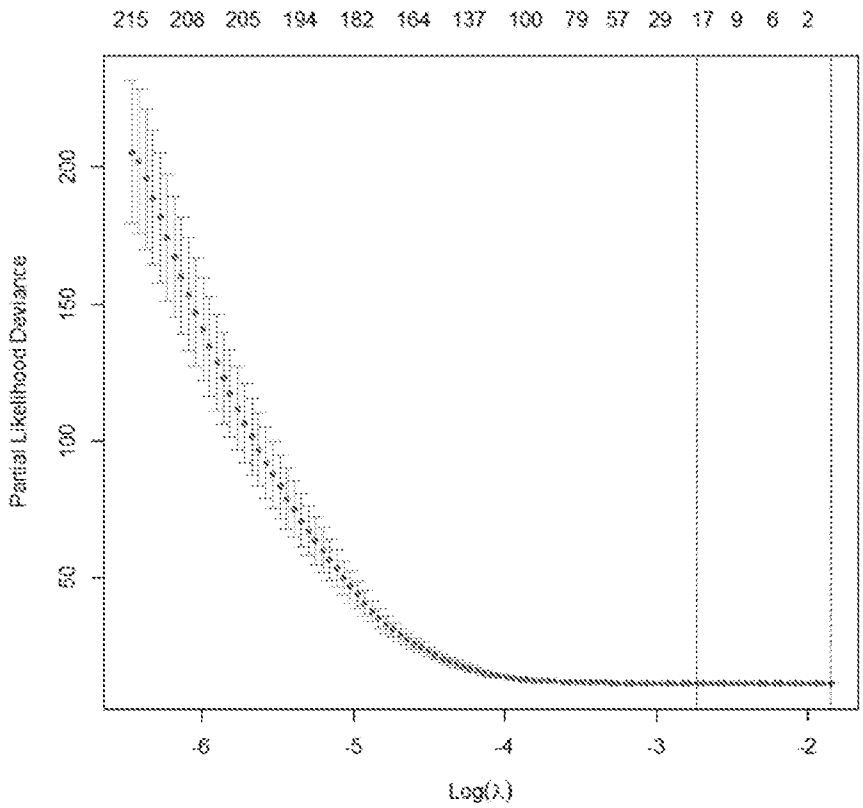
215 208 205 194 182 164 137 100 79 57 29 17 9 6 2
Partial Likelihood Deviance
200
150
100
50
-6
-5
-4
-3
-2
Log(λ)

图 2

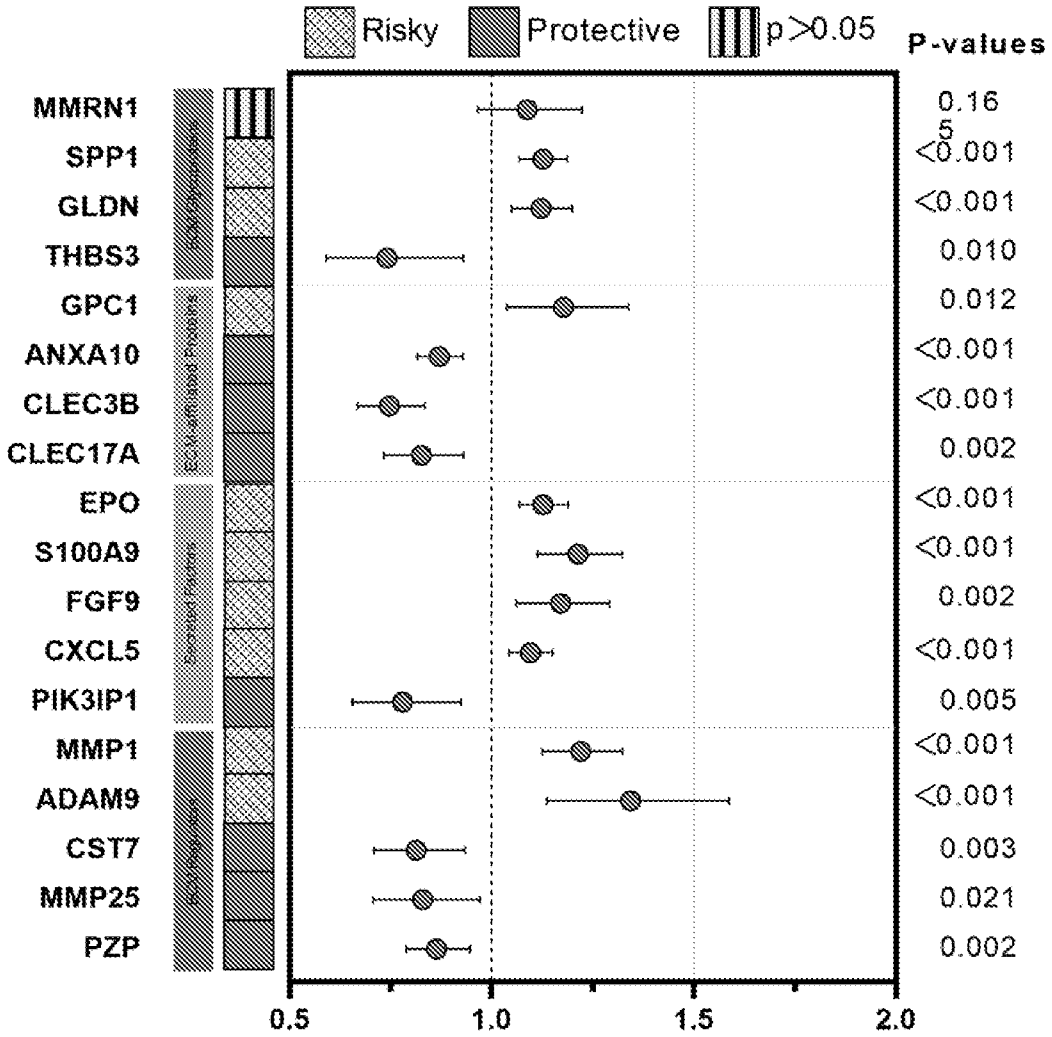
Risky
Protective
p>0.05
P-values
MMRN1
SPP1
GLDN
THBS3
GPC1
ANXA10
CLEC3B
CLEC17A
EPO
S100A9
FGF9
CXCL5
PIK3IP1
MMP1
ADAM9
CST7
MMP25
PZP
0.165
<0.001
<0.001
0.010
0.012
<0.001
<0.001
0.002
<0.001
<0.001
0.002
<0.001
0.005
<0.001
<0.001
0.003
0.021
0.002
0.5
1.0
1.5
2.0

图 3

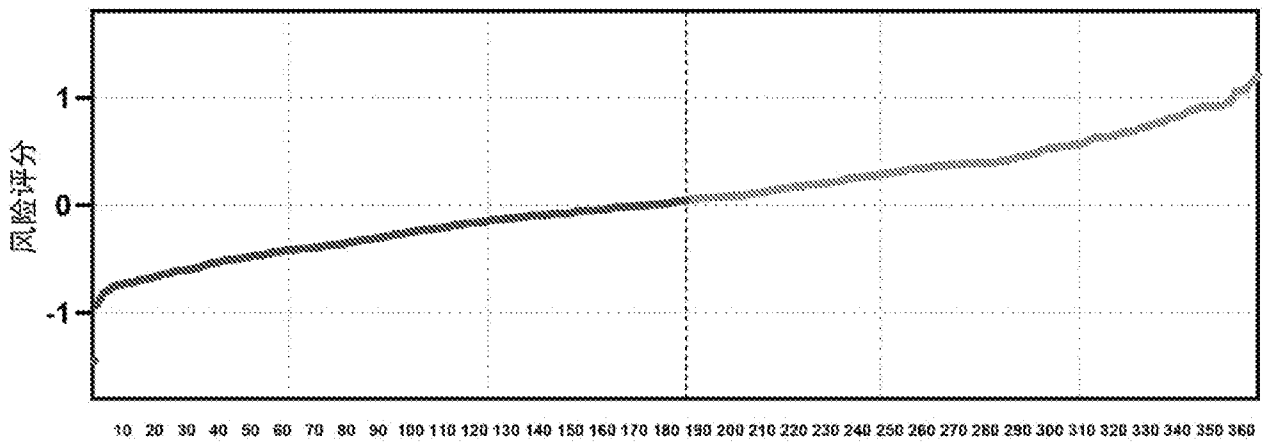
风险评分
1
0
-1
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360

图 4

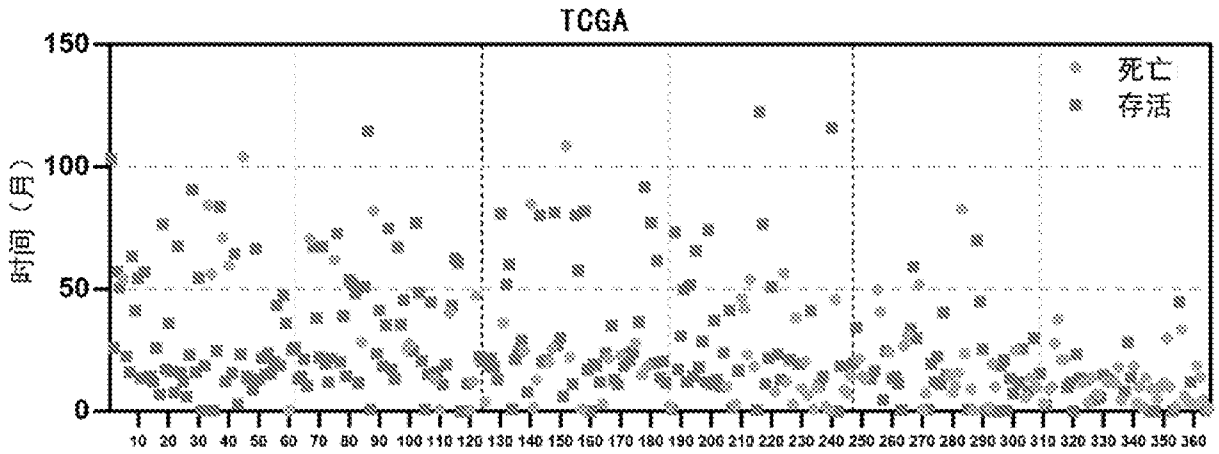
TCGA
死亡
存活
时间（月）
150
100
50
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360

图 5

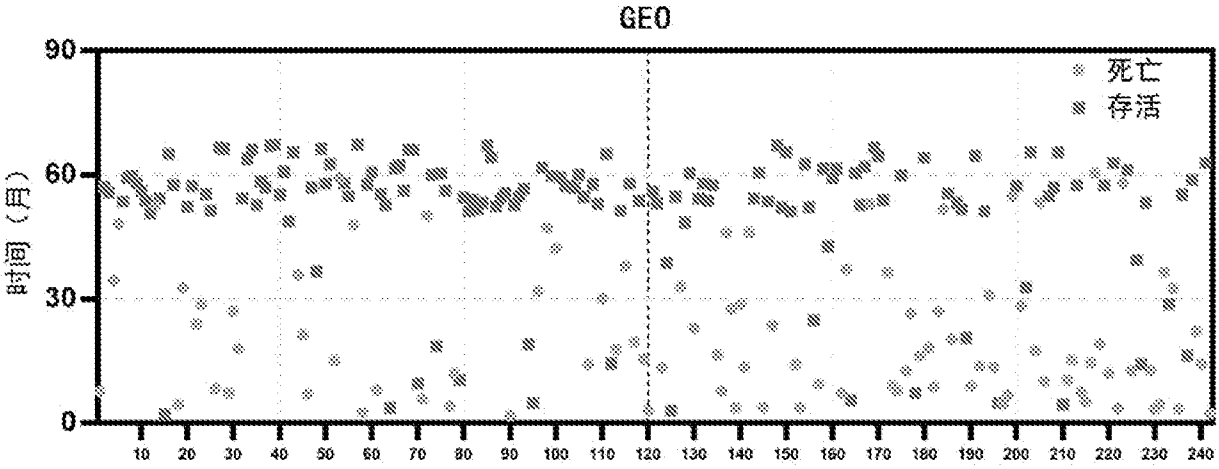
GEO
死亡
存活
时间（月）
90
60
30
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240

图 6

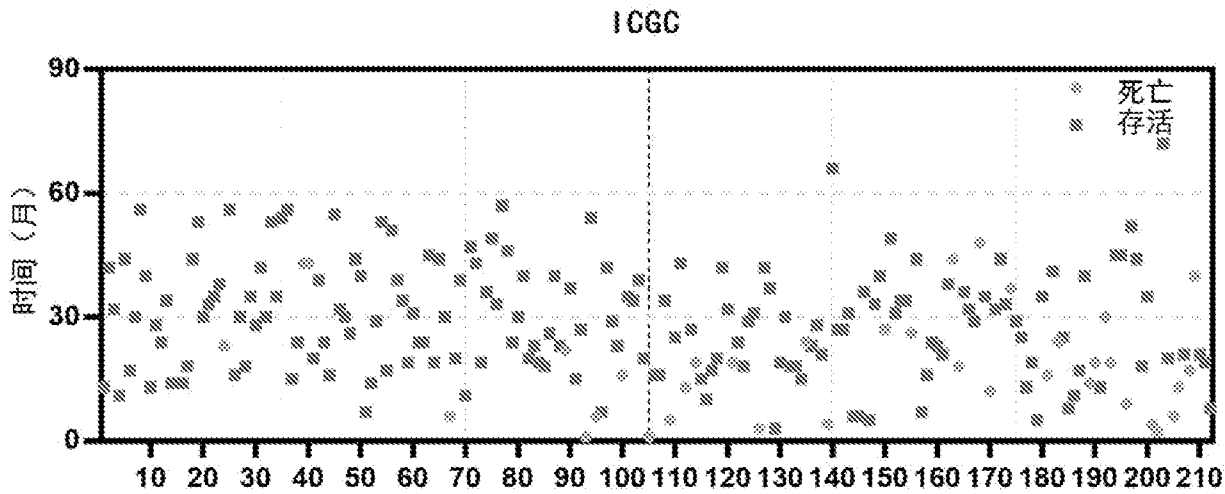
ICGC
死亡
存活
时间（月）
90
60
30
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210

图 7

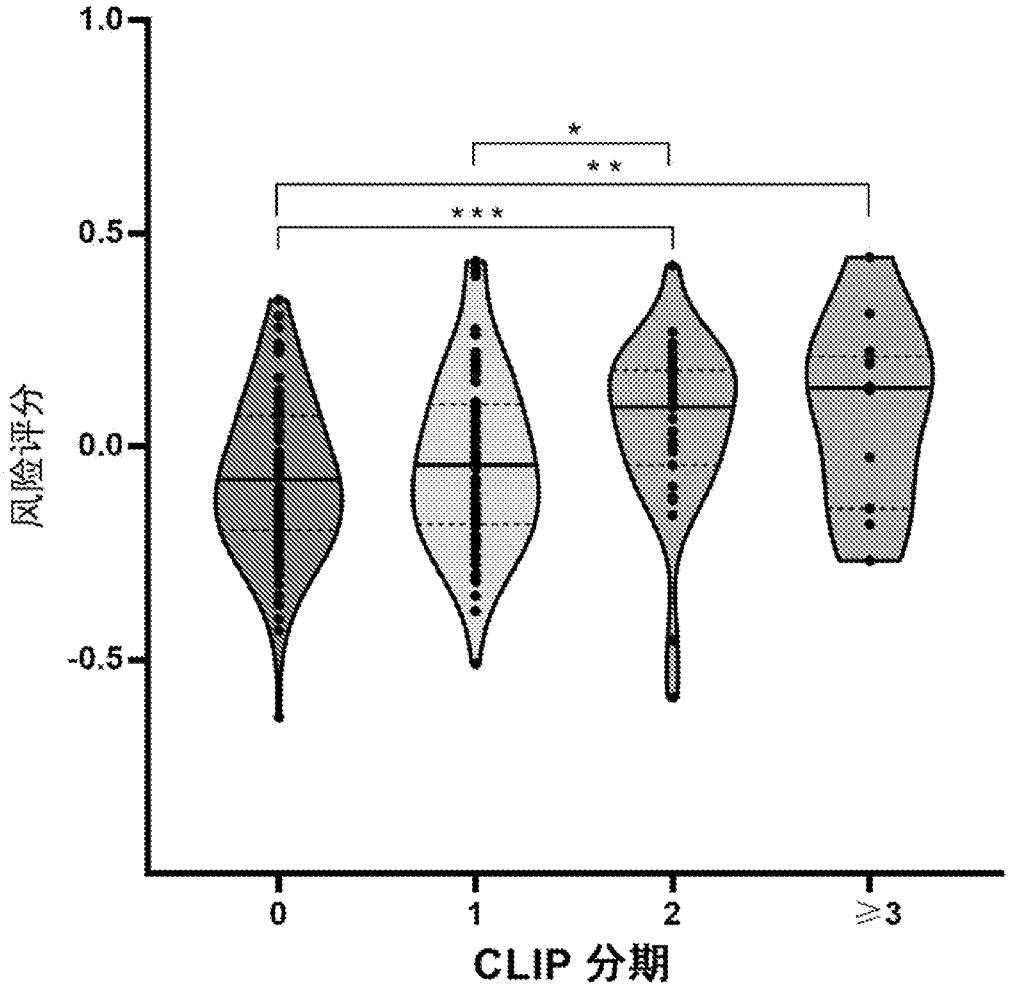


图 8

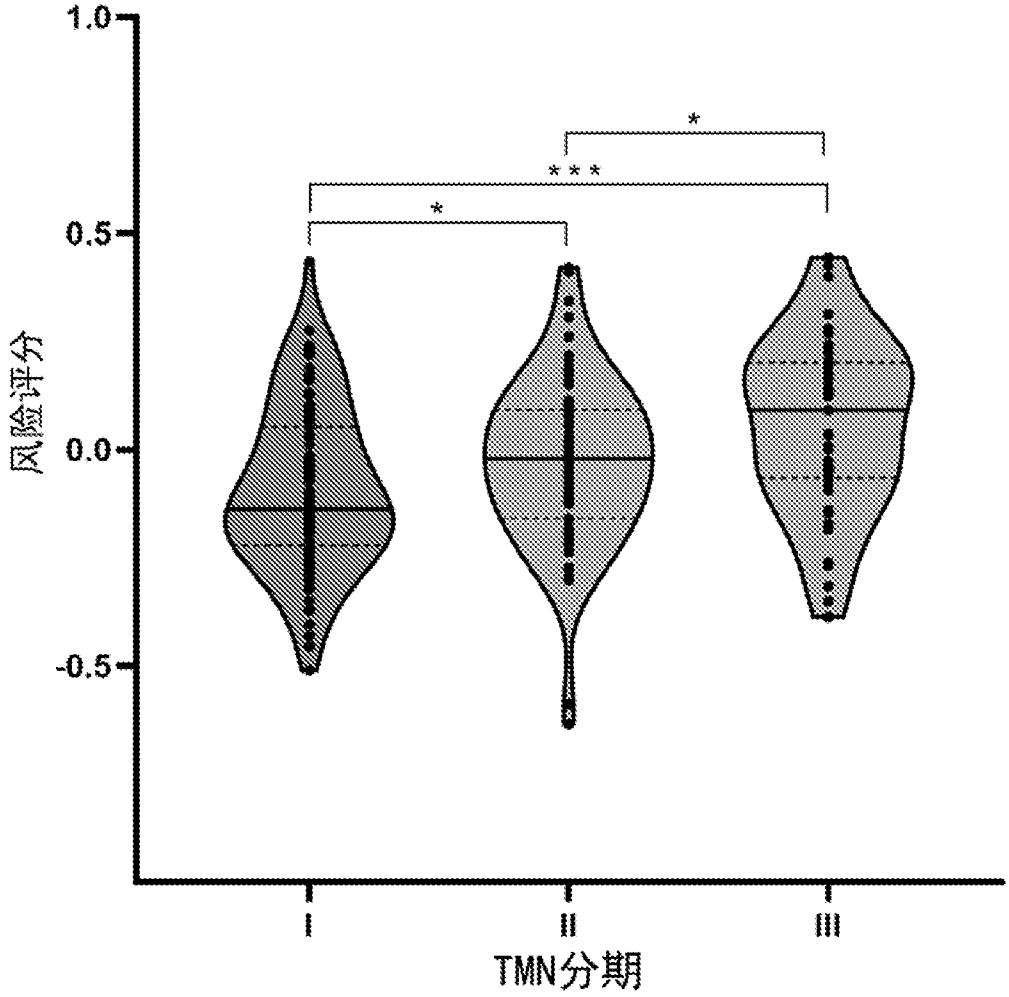
1.0
0.5
0.0
-0.5
风险评分
*
***
*
I
II
III
TMN分期

图 9

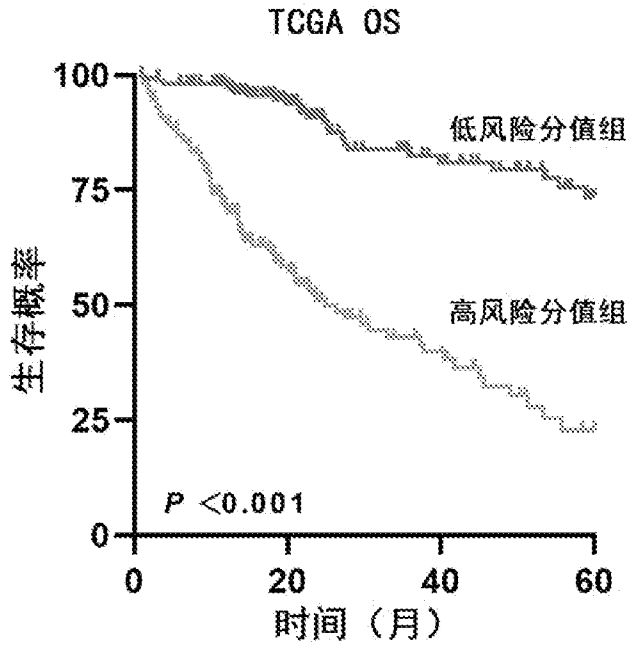
TCGA OS
100
75
50
25
0
生存概率
低风险分值组
高风险分值组
P <0.001
0
20
40
60
时间（月）

图 10

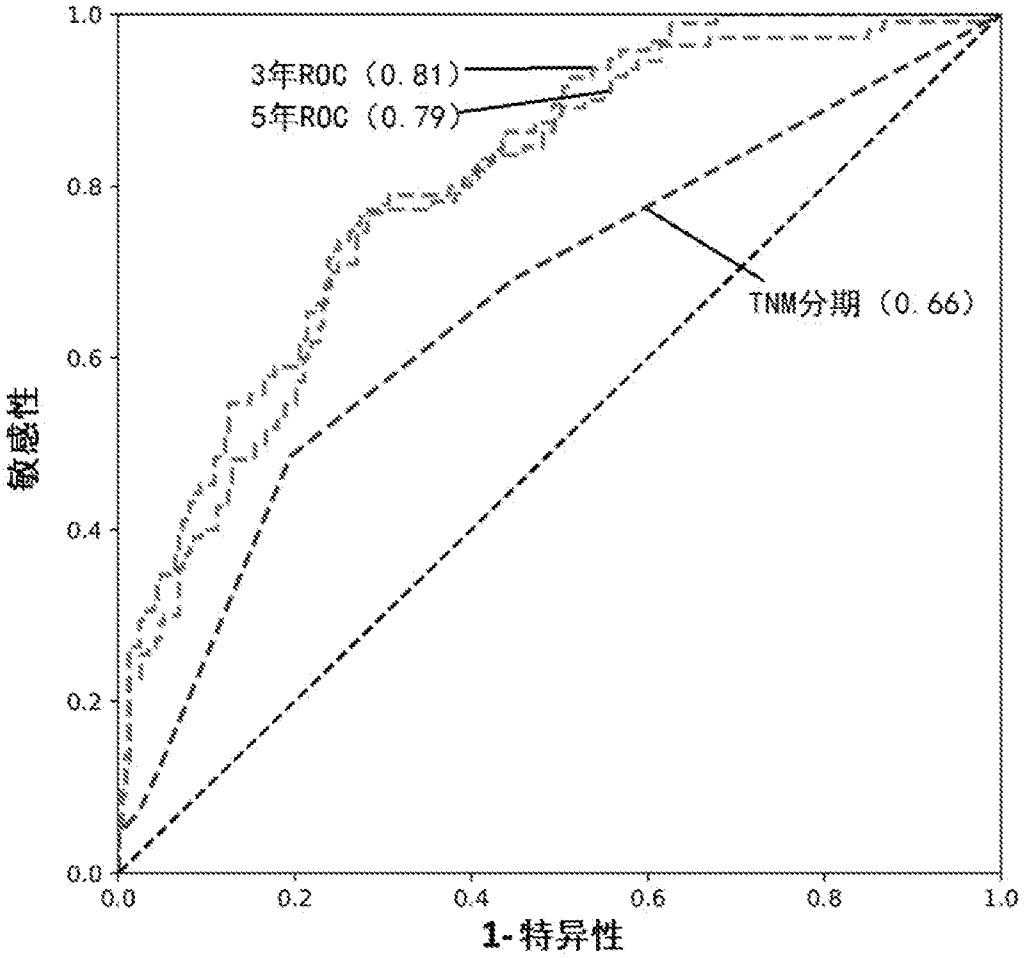
1.0
0.8
0.6
0.4
0.2
0.0
敏感性
3年ROC（0.81）
5年ROC（0.79）
TNM分期（0.66）
0.0
0.2
0.4
0.6
0.8
1.0
1-特异性

图 11

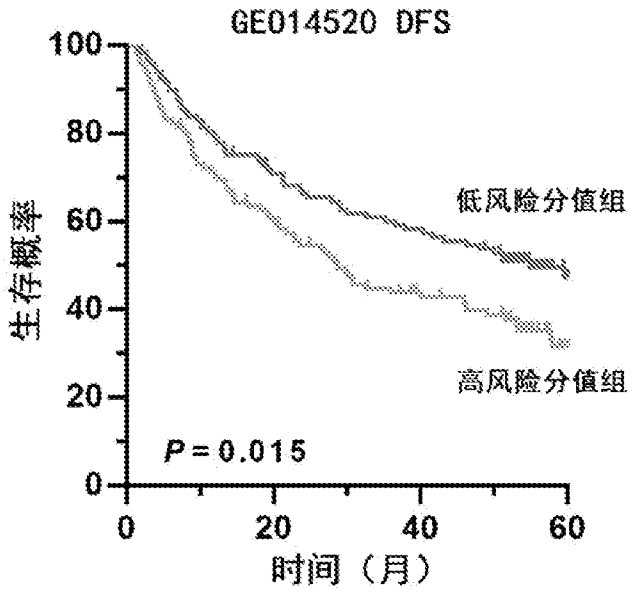
GE014520 DFS
100
80
60
40
20
0
生存概率
低风险分值组
高风险分值组
P = 0.015
0
20
40
60
时间（月）

图 12

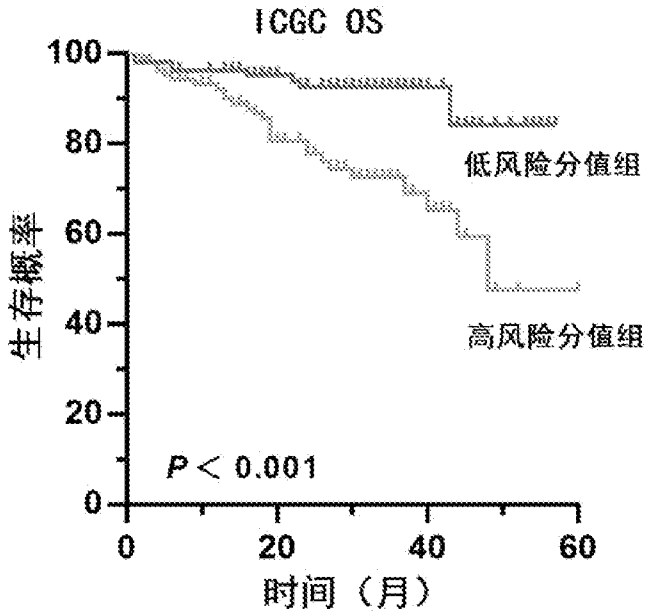


图 13

# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/CN2021/139502**

**A. CLASSIFICATION OF SUBJECT MATTER**

G16H 50/20(2018.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G16H; G16B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

cntxt, ven, entxt, cnki: 癌, 肝癌, 探针, 基因, 预测, 检测, 风险, 预后, 分析, 芯片, 模型, mmp1, epo, mmrn1, s100a9, adam9, gpc1, spp1, gldn, fgf9, cxcl5, cst7, thbs3, anxa10, pik3ip1, mmp25, clec3b, pzp, clec17a, chip, probe, gene, cancer, model

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| PX | CN 113539376 A (ZHEJIANG UNIVERSITY) 22 October 2021 (2021-10-22)<br>claims 1 and 2 | 1-2 |
| A | CN 102191323 A (ZHONGSHAN HOSPITAL, FUDAN UNIVESITY) 21 September 2011 (2011-09-21)<br>entire document | 1-2 |
| A | CN 105063049 A (SHANGHAI DIDA BIOLOG SCIENCE & TECHNOLOGY CO., LTD.) 18 November 2015 (2015-11-18)<br>entire document | 1-2 |
| A | CN 101560554 B (ZHONGSHAN HOSPITAL, FUDAN UNIVESITY) 04 May 2011 (2011-05-04)<br>entire document | 1, 2 |

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| **10 May 2022** | **19 May 2022** |
| Name and mailing address of the ISA/CN<br>**China National Intellectual Property Administration (ISA/CN)**<br>**No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China**<br>Facsimile No. **(86-10)62019451** | Authorized officer<br><br>Telephone No. |

**INTERNATIONAL SEARCH REPORT**
**Information on patent family members**

International application No.
**PCT/CN2021/139502**

| Patent document cited in search report | Publication date (day/month/year) | Patent family member(s) | Publication date (day/month/year) |
|---|---|---|---|
| CN 113539376 A | 22 October 2021 | None | |
| CN 102191323 A | 21 September 2011 | None | |
| CN 105063049 A | 18 November 2015 | None | |
| CN 101560554 B | 04 May 2011 | CN 101560554 A | 21 October 2009 |

# 国际检索报告

国际申请号 PCT/CN2021/139502

A. 主题的分类

G16H 50/20(2018.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G16H; G16B

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

cntxt,ven,entxt,cnki: 癌，肝癌，探针，基因，预测，检测，风险，预后，分析，芯片，模型，mmp1，epo，mmrn1，s100a9，adam9，gpc1，spp1，gldn，fgf9，cxcl5，cst7，thbs3，anxa10，pik3ip1，mmp25，clec3b，pzp，clec17a，chip，probe，gene，cancer，model

C. 相关文件

| 类 型* | 引用文件，必要时，指明相关段落 | 相关的权利要求 |
|---|---|---|
| PX | CN 113539376 A (浙江大学) 2021年10月22日 (2021-10-22)<br>权利要求1-2 | 1-2 |
| A | CN 102191323 A (复旦大学附属中山医院) 2011年9月21日 (2011-09-21)<br>全文 | 1-2 |
| A | CN 105063049 A (上海缔达生物科技有限公司) 2015年11月18日 (2015-11-18)<br>全文 | 1-2 |
| A | CN 101560554 B (复旦大学附属中山医院) 2011年5月4日 (2011-05-04)<br>全文 | 1-2 |

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

| 国际检索实际完成的日期 | 国际检索报告邮寄日期 |
|---|---|
| 2022年5月10日 | 2022年5月19日 |
| ISA/CN的名称和邮寄地址<br>中国国家知识产权局(ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br>传真号 (86-10)62019451 | 受权官员<br>汪宁<br>电话号码 62411681 |

# 国际检索报告
## 关于同族专利的信息

国际申请号
PCT/CN2021/139502

| 检索报告引用的专利文件 | 公布日（年/月/日） | 同族专利 | 公布日（年/月/日） |
|---|---|---|---|
| CN 113539376 A | 2021年10月22日 | 无 | |
| CN 102191323 A | 2011年9月21日 | 无 | |
| CN 105063049 A | 2015年11月18日 | 无 | |
| CN 101560554 B | 2011年5月4日 | CN 101560554 A | 2009年10月21日 |