



电子、语音版

· 论 著 ·

## 细胞外基质相关基因预后模型在脑胶质瘤中 预测预后能力分析

贾牧原<sup>1,2</sup>, 刘羽阳<sup>3</sup>, 胡文涛<sup>2</sup>, 陈俊燚<sup>2</sup>, 张洪俊<sup>1</sup>, 龚欢欢<sup>4</sup>, 吴剑慧<sup>2</sup>, 任博文<sup>2</sup>, 刘嘉霖<sup>5</sup>, 陈凌<sup>5</sup>

1. 北京美中爱瑞肿瘤医院, 北京 100070
2. 解放军医学院, 北京 100853
3. 解放军联勤保障部队九二〇医院, 云南昆明 650032
4. 中南大学湘雅三医院重症医学科, 湖南长沙 410013
5. 解放军总医院第一医学中心神经外科医学部, 北京 100853

**摘要:**目的 构建细胞外基质(ECM)相关基因预后模型,评价其预测胶质瘤患者预后的能力,探索基于该模型的胶质瘤免疫微环境特征。方法 基于肿瘤基因组图谱(TCGA)和基因型-组织表达(GTEX)数据库中胶质瘤以及正常脑组织数据,筛选获得差异基因(DEGs);基于基因本体论(GO)数据库获取ECM相关基因,基于单因素Cox回归分析获取胶质瘤预后基因。将上述三部分交集获得重叠候选基因,再经由Lasso分析获取最佳的4基因预后模型,并于TCGA以及中国脑胶质瘤图谱数据库(CGGA)中胶质瘤队列中进行生存分析和Cox回归分析。基于4基因预后模型以及TCGA患者预后数据构建预后列线图,并在CGGA胶质瘤队列中进行验证。最后,基于富集分析、免疫检查点分析以及免疫浸润分析探索4基因预后模型相关的免疫微环境特征。结果 22个重叠候选基因经由Lasso分析后获得最佳4基因预后模型,该模型的风险评分能够较好地预测TCGA以及CGGA胶质瘤患者的预后,并且是危险因素。细胞系验证实验中提示U251细胞系(人源胶质瘤细胞)最佳4基因表达均高于HA1800(人源星形胶质细胞),符合TCGA以及CGGA数据库分析结果。基于4基因预后模型构建的预后列线图同样具有较好地预测患者预后的能力。高风险组患者肿瘤组织内具有较高水平的M2型巨噬细胞浸润且免疫检查点相关分子(PD-L1, B7-H3, CTLA4, PD1, TIM3以及LAG3)高于低风险组。结论 ECM相关基因模型以及预后列线图均能够较好地预测胶质瘤患者的预后,高风险组患者具有抑制性免疫微环境特征,免疫检查点抑制剂可能是该类患者的潜在治疗方式。 [国际神经病学神经外科学杂志, 2024, 51(2): 35-47]

**关键词:**细胞外基质;胶质瘤;基因富集分析;免疫细胞浸润;预后

中图分类号:R739.41

DOI:10.16636/j.cnki.jinn.1673-2642.2024.02.007

### Ability of a prognostic model based on extracellular matrix-associated genes in predicting the prognosis of glioma

JIA MUYUAN<sup>1,2</sup>, LIU YUYANG<sup>3</sup>, HU WENTAO<sup>2</sup>, CHEN JUNYI<sup>2</sup>, ZHANG HONGJUN<sup>1</sup>, GONG HUANHUAN<sup>4</sup>, WU JIANHUI<sup>2</sup>, REN BOWEN<sup>2</sup>, LIU JIALIN<sup>5</sup>, CHEN LING<sup>5</sup>

1. Arion Cancer Center, Beijing 100070, China

2. Medical School of Chinese PLA, Beijing 100853, China

3. Department of Neurosurgery, The 920th Hospital of Joint Logistics Support Force, Kunming, Yunnan 650032, China

4. Department of Critical Care Medicine, The Third Xiangya Hospital of Central South University, Changsha, Hunan 410013, China

5. Department of Neurosurgery, The First Medical Center of Chinese PLA General Hospital, Beijing 100853, China

基金项目:国家自然科学基金(U20A20380, 81672824, 82172680)。

收稿日期:2023-08-29;修回日期:2024-03-27

作者简介:贾牧原(1997—),男,医学硕士,主要从事胶质瘤创新综合治疗研究。Email:gravesjia@163.com。

通信作者:陈凌,Email:chen\_ling301@163.com。

Corresponding author: CHEN Ling, Email: chen\_ling301@163.com

**Abstract: Objective** To construct a prognostic model based on extracellular matrix (ECM)-associated genes, to evaluate its ability in predicting the prognosis of glioma patients, and to investigate the characteristics of the immune microenvironment of glioma based on this model. **Methods** The TCGA and GTEx databases were used to obtain the data of glioma and normal brain tissue, and then differentially expressed genes were obtained. The Gene Ontology (GO) database was used to obtain ECM-related genes, and a univariate Cox regression analysis was used to obtain the genes associated with the prognosis of glioma. The above three groups of genes were intersected to obtain overlapping candidate genes, and a LASSO analysis was used to obtain the optimal four-gene prognostic model. The glioma cohorts in the TCGA and CGGA databases were used for survival analysis and the Cox regression analysis. A prognostic nomogram was constructed based on the four-gene prognostic model and the prognostic data of patients in TCGA, which was validated in the glioma cohort in CGGA. Finally, enrichment analysis, immune checkpoint analysis, and immune infiltration analysis were used to investigate the characteristics of immune microenvironment associated with the four-gene prognostic model. **Results** The LASSO analysis was performed for the 22 overlapping candidate genes to obtain the optimal four-gene prognostic model, and the risk score of this model could better predict the prognosis of glioma patients in TCGA and CGGA and was a risk factor for the poor prognosis of patients. Cell line validation experiments showed that the U251 cell line (human-derived glioma cells) had higher expression levels of all four optimal genes than the HA1800 cell line (human-derived astrocytes cells), which was consistent with the results of data analysis for TCGA and CGGA databases. The prognostic nomogram constructed based on the four-gene prognostic model also had a good ability to predict patient prognosis. Compared with the low-risk group, the high-risk group had higher levels of M2 macrophage infiltration and immune checkpoint-related molecules (PD-L1, B7-H3, CTLA4, PD1, TIM3, and LAG3) in tumor tissue. **Conclusion** Both the model based on ECM-associated genes and the prognostic nomogram can well predict the prognosis of glioma patients. Patients in the high-risk group have the characteristics of suppressive immune microenvironment, and immune checkpoint inhibitors may be a potential treatment modality for such patients.

[Journal of International Neurology and Neurosurgery, 2024, 51(2): 35-47]

**Keywords:** extracellular matrix; glioma; gene set enrichment analysis; immune cell infiltration; prognosis

细胞外基质(extracellular matrix, ECM)主要由5类物质组成,包括胶原蛋白、非胶原蛋白、弹性蛋白、蛋白聚糖和氨基聚糖。ECM各组均积极参与并调节细胞生长、形态、迁移和代谢等活动,其中纤粘蛋白是细胞外基质重要组成部分,它促进细胞与细胞之间的黏附,与正常细胞相比,肿瘤细胞表面纤维蛋白明显减少,细胞形态变圆,细胞呈现细胞密集和重叠生长的生长特征,导致其侵袭性增强<sup>[1]</sup>。如今,越来越多的研究提示ECM与肿瘤细胞生物学功能具有相关性,例如:ECM重塑可影响胶质瘤的增殖、血管生成和侵袭等<sup>[2-5]</sup>。其次ECM有助于发现与癌症相关的疾病以及新的诊断工具和治疗靶点<sup>[6]</sup>。

胶质瘤是最常见的原发性颅内恶性肿瘤,发病率每年约5~8/100 000人,占颅内恶性肿瘤81%<sup>[7-8]</sup>,胶质瘤根据其生长模式和组织学可分为两种。一种是非弥漫性胶质瘤,如毛细胞型星形细胞瘤和室管膜瘤,通常表现为缓慢且更局限的生长模式。另一种为弥漫性胶质瘤,代表了大部分无法治愈的胶质瘤,其特征是在中枢神经系统实质内高度浸润性生长,弥漫性胶质瘤代表了绝大多数胶质瘤,可进一步细分为少突胶质瘤、星形细胞瘤和胶质

母细胞瘤(Glioblastoma, GBM),后者是弥漫性胶质瘤中最具侵袭性的亚型<sup>[9]</sup>。弥漫性胶质瘤的侵袭特征主要为沿白质束以及神经元周围和血管周围间隙的途径浸润脑实质,这导致肿瘤和周围脑实质之间通常没有明确界限,无法完整手术切除肿瘤,即使手术切除配合放疗,但肿瘤复发率近100%<sup>[10]</sup>,5年生存率低于5%<sup>[11-12]</sup>。近期研究表明,肿瘤相关的ECM被认为是胶质瘤侵袭功能的关键因素<sup>[13]</sup>,并且ECM的组成成分影响了胶质瘤细胞侵袭功能,其中糖苷酶和蛋白酶不仅调节肿瘤细胞黏附而且还影响其细胞骨架释放各种生长因子和趋化因子,进而激活胶质瘤细胞生成与ECM相关结合蛋白,例如透明质酸(hyaluronic acid, HA)受体:CD44和RHAMM(透明质酸介导的运动受体),整合素:a3b1、avb2、avb3和avb5等,以及黏合素<sup>[3]</sup>。这些ECM相关结合蛋白可能是胶质瘤侵袭功能的必要因素。

2020年世界卫生组织(WHO)首次将分子生物标志物纳入胶质瘤亚型分类重要因素,其中包括:异柠檬酸脱氢酶(isocitrate dehydrogenase, IDH)突变状态、染色体臂1p/19q编码共同缺失、染色体重塑蛋白ATRAX的编码状态

和抑癌基因 p53 的突变等<sup>[14]</sup>。然而, IDH 突变和 1p/19q 编码缺失可以作为预测胶质瘤预后和发展的生物标志物, 但大部分患者均不适用于这些生物标志物<sup>[15]</sup>。所以, 可预测胶质瘤预后和发展的生物标志物研究在很大程度上是一个未被充分开发的领域。

目前, 浸润性胶质瘤的标准治疗包括最大限度的手术切除、放射治疗和替莫唑胺疗法, 然而由于血脑屏障、肿瘤微环境 (tumor microenvironment, TME) 和手术切除不完全等因素的存在阻碍了化疗药物的转运, 导致总体预后仍很差<sup>[16]</sup>。除标准治疗外免疫治疗是治疗胶质瘤的研究热点。免疫治疗主要是抑制肿瘤免疫抑制功能, 重新激活 T 细胞功能<sup>[17]</sup>。其次是通过向 T 细胞呈递肿瘤相关抗原激活抗肿瘤免疫反应<sup>[18]</sup>, 然而它的肿瘤治疗效果不如预期。随着研究的不断深入, 发现在胶质瘤进展过程中, 肿瘤周围由星形细胞、内皮细胞、趋化因子和细胞外蛋白组成, 这些物质能够促进肿瘤免疫逃逸<sup>[19-20]</sup>。长期以来, ECM 被认为是一种惰性构架而被忽视, 并且其状态已被鉴定可随免疫系统一同动态变化。尽管目前胶质瘤免疫治疗的结果令人失望, 但联合 ECM 或逆转肿瘤免疫抑制微环境疗法仍值得深入治疗, 未来 ECM 研究可能成为推动肿瘤免疫细胞疗法的助力器。

## 1 材料和方法

### 1.1 获得细胞外基质相关的基因及胶质瘤和正常大脑的数据收集

为了获得来自基因本体数据库 (Gene Ontology, <http://geneontology.org/>) 细胞外基质相关基因列表 (extracellular matrix-related gene, ECMRG), 检索词设定为 “Extracellular matrix”, 其余过滤参数设定为 “Human sapiens” 和 “genes and genes products”。最终, 共获得 951 组人源 ECMRG 数据。

为了收集胶质瘤和正常大脑的数据从基因型-组织表达 (genotype-tissue expression project, GTEx) 数据库中下载 1 152 个正常脑组织数据。利用 UCSC Xena 数据库 (<http://xena.ucsc.edu/>), 下载来自 TCGA 中 694 个胶质瘤样本的临床信息和相应基因表达数据。下载来自 CGGA 中 693 个中国胶质瘤患者相对信息作为验证集。搜集的所有数据中缺乏临床病理或预后信息的样本被排除, 最终共获得来自 TCGA 数据库的 616 名患者信息作为发现集和来自 CGGA 数据库的 542 名患者信息作为验证集, 除了差异基因获取使用 counts 数据格式外, 其余分析中转录组原始数据被标准化为 TPM 格式, 并进行  $\text{Log}_2(\text{TPM}+1)$  转换, 以用于后续分析。

### 1.2 获得差异表达的基因及筛选重叠的候选基因

使用 “limma” R 包 (3.6.3), 分析了胶质瘤 (TCGA 数据库) 和正常脑组织之间的差异表达基因 (differentially expressed genes, DEGs)。差异基因的筛选标准为:  $P.\text{adj}$

$< 0.05$ ,  $|\text{llog}_2(\text{Fold change})| > 2$ 。

基于发现集使用 “survival” R 包通过单因素 Cox 回归分析获取预后基因, 风险比 (hazard ratio, HR)  $\neq 1$  且  $P < 0.05$  条件的基因被选做进一步分析。基于韦恩图 (Venn) 分析, 将 ECM 相关基因, 预后基因和 DEGs 三部分取交集, 满足  $HR > 1$  且上调 DEGs 以及  $HR < 1$  且下调 DEGs 的 ECM 相关基因作为重叠候选基因, 以备后续分析。

### 1.3 建立细胞外基质相关的风险评分

使用 “glmnet” 和 “survival” R 包, 应用 Lasso 算法、重叠候选基因以及 TCGA 胶质瘤数据库构建 ECM 相关风险评分。其中选取十倍交叉验证以确定最佳  $\lambda$  值, 以提高风险评分的可靠性和客观性。同时, 考虑到在临床工作中的实用性, 风险评分中的变量数越少, 临床上实用性就越大。最终, 我们选择了 4 个预后基因: 基质金属蛋白酶 14 (matrix metalloproteinase 14, MMP14)、丝氨酸蛋白酶抑制剂 3 (serpin family a member 3, SERPINA3)、分泌型磷蛋白 1 (secreted phosphoprotein-1, SPP1)、基质金属蛋白酶 2 (matrix metalloproteinase 2, MMP2) 构建风险评分, 最终获得风险评分 (Risk Score) 公式:  $\text{Risk Score} = \sum_{i=1}^n \beta_i * \text{Exp}_i + \alpha$

其中 Exp 为基因表达量,  $\beta$  为基因所对应的回归系数,  $\alpha$  为常数项。

### 1.4 风险评分评价

利用人来源蛋白图谱 (human protein atlas, HPA) 数据库评估这 4 个基因的蛋白质表达水平。基于 TCGA 和 GTEx 数据库的数据评估 4 个基因在胶质瘤中的相对表达水平。此外, Kaplan-Meier (KM) 生存曲线和时间依赖性受试者工作特征曲线 (time-dependent receiver operating characteristic curve, tROC) 被用于评估这些基因的预后价值。根据风险评分的中位数将胶质瘤患者分为低风险组 (low risk) 和高风险组 (high risk)。

其次, 使用生存分析来评估低风险和高风险组之间的差异。通过 tROC 曲线评估风险评分的预测准确性。此外, 使用 “ggplot2” R 包对风险图进行了说明。单因素 Cox 比例风险回归分析用于评估临床参数 (年龄、性别、肿瘤的 WHO 分级、IDH 基因状态以及 1p/19q 共缺失状态) 和风险评分的预测胶质瘤患者预后的能力, 多因素 Cox 比例风险回归分析被用于评估风险评分是否能作为预后危险因素。

### 1.5 建立列线图 (Nomogram)

基于在发现集中分析所构建的风险评分以及经多因素 Cox 比例风险回归分析所得的预后因素 (年龄、性别、WHO 分级、IDH 状态、1p/19q 编码状态和风险评分) 使用 “rms” 和 “survival” R 包, 构建一个包含相关临床信息和预后因素的 Nomogram, 用来预测确诊胶质瘤超过 1、3 及 5 年患者生存概率。此外采用 tROC 评估模型预测胶质瘤

患者预后的准确性。tROC的曲线下面积(area under curve, AUC)被用于不同模型或预后因素间的比较,并采用校准曲线(calibration curve)评估Nomogram的精准度。

### 1.6 在CGGA数据库中对基因签名的验证

使用CGGA数据库对细胞外基质相关的风险评分进行了验证。根据在TCGA数据库中构建的公式,CGGA数据库的患者根据风险评分的中位数被分为低、高风险组,KM生存曲线被用来检测组间的差异,并使用tROC评估预后基因的预测预后性能。选择校准曲线和tROC来评估Nomogram的预测预后性能及精准度。

### 1.7 基因集富集分析及免疫微环境分析

使用“cluster Profiler”R软件包,进行基因集富集分析(gene set enrichment analysis, GSEA),以进行来自发现集(TCGA)的DEGs功能注释,并探索高风险组与低风险组相比的特征。评估时相关参数:错误发现率(false discovery rate, FDR)<0.25和 $P_{adj}$ <0.05。

为了说明风险分数与胶质瘤免疫微环境之间的关系,使用一种用于评估复杂混合组织样本中不同细胞类型相对丰度的计算方法(cell-type identification by estimating relative subsets of RNA transcripts, CIBERSORT)和肿瘤浸润免疫细胞定量分析法quanTIseq来评估TCGA数据库中高风险组和低风险组的免疫浸润水平。通过相关性分析来说明风险评分和免疫浸润状态之间的关系。此外,为了探索TCGA和CGGA数据库中ECMRGs特征和免疫疗法之间的关系,测量风险评分和免疫检查点之间的相关性。

### 1.8 胶质母细胞瘤细胞系培养

U251、HA1800细胞系均购自中国科学院上海细胞生物研究所,在含有10%胎牛血清(fetal bovine serum, FBS)的高糖型杜氏改良Eagle培养基(high-glucose Dulbecco's modified Eagle's, DMEM-HG)中,并置于37℃和5%二氧化碳CO<sub>2</sub>的培养箱中。

### 1.9 实时定量PCR

实时定量PCR分析使用CFX96实时PCR检测系统(Bio-Rad)。GAPDH的表达被用于标准化。本研究使用的引物序列列于表1。采用2<sup>-ΔΔCt</sup>法检测相对定量。

### 1.10 统计学方法

本工作中的所有统计分析均使用R软件(3.6.3)进行。两组的统计学差异由Wilcoxon秩和检验确定,多组比较时采用Kruskal-Wallis检验。分类变量的比较采用 $\chi^2$ 检验。生存分析采用对数秩检验来评价两组之间的差异。 $P<0.05$ 为差异有统计学意义。

## 2 结果

### 2.1 重叠基因的获取

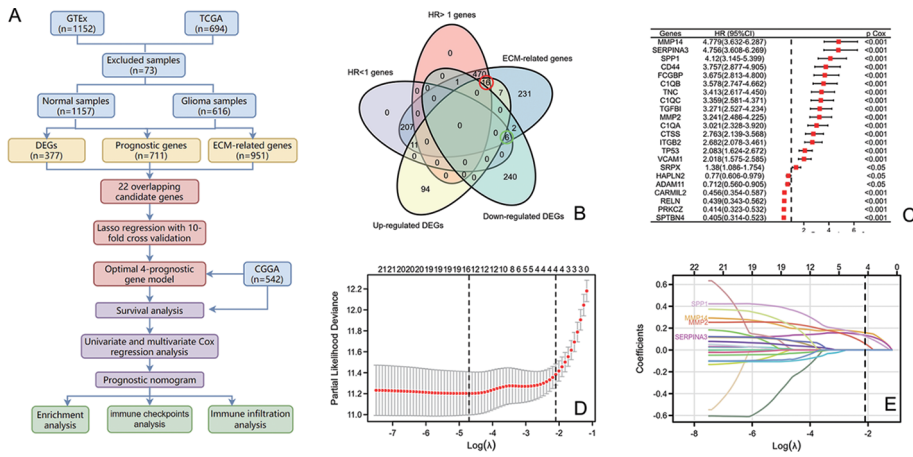
本部分整个分析为收集正常脑组织和胶质瘤中RNA-seq数据样本,从GTEX数据库中下载1152个样本,

TCGA数据库中下载694个样本(图1A)。为筛选出与ECM相关的候选差异基因。我们在差异基因(377个)中结合相应的临床信息,将预后基因( $HR>1$ )与上调的DEGs相交,获得16个候选基因。同样地,预后基因( $HR<1$ )与下调的DEGs相交,获得6个候选基因。共有22个候选基因用于后续分析,通过对作为训练数据库的TCGA数据库使用Cox回归模型来进行单因素生存分析。在22个与预后相关的基因,其中MMP14、SERPINA3、SPP1等基因为风险因素( $HR>1$ ),HAPLN2、ADAM11、CARML2、RELN、PRKCZ、SPTBN4在胶质瘤中起保护作用( $HR<1$ )(图1B, 1C)。为了提高22个候选基因的鲁棒性,使用这些基因在TCGA、GTEX数据库中通过Lasso算法进行额外分析(图1D),之后在构建风险评分特征分析中,筛选出4个预后基因(MMP14、SERPINA3、SPP1和MMP2),结果提示4个预后基因均为促癌基因(图1E)。

为了评估4个基因对构建风险评分特征的重要意义,从TCGA数据库中分析4个候选基因表达,结果提示4个基因在肿瘤组织中表达与正常脑组织相比较表达增高(图2A)。为验证TCGA数据库中差异基因筛选候选基因的准确性,对比4个预后基因在胶质瘤细胞系(U251)和正常星形胶质细胞(HA1800)中表达差异,实时定量PCR实验提示:MMP14、SERPINA3、SPP1和MMP2均在U251中表达增高,表达量具有统计学差异( $t_{MMP14}=15.520$ ,  $P_{MMP14}=0.0001$ ;  $t_{SERPINA3}=2.990$ ,  $P_{SERPINA3}=0.0403$ ;  $t_{SPP1}=4.609$ ,  $P_{SPP1}=0.010$ ;  $t_{MMP2}=26.390$ ,  $P_{MMP2}=0.000$ )(图2B)。HPA网站获得免疫组化病理标本数据。4个候选基因蛋白在正常脑组织、低级别和高级别胶质瘤组织中表达有统计学差异(图2C)。此外,我们利用Kaplan-Meier曲线和ROC曲线探讨了这些基因的预后价值,MMP14、SERPINA3、SPP1和MMP2为风险因素( $HR>1$ )(图2D),在ROC曲线中,该4个基因对于胶质瘤患者1、3及5年预后均具有一定的预测价值(图2E)。

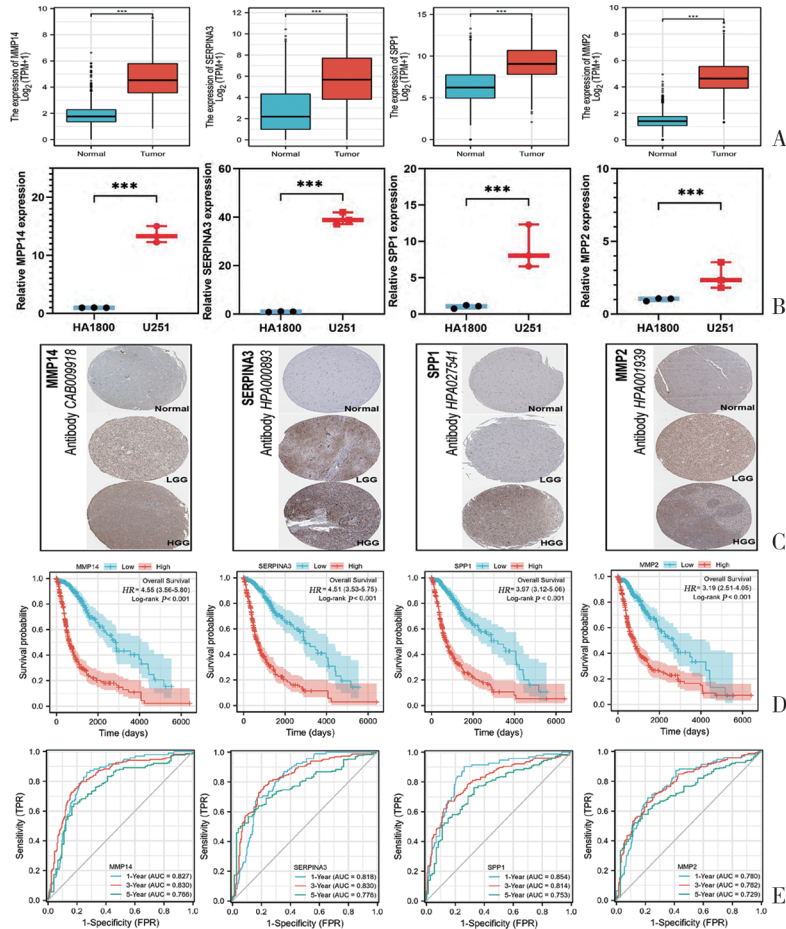
### 2.2 ECM相关基因风险评分标记的预后价值

为了获得风险评分特征对胶质瘤患者临床结果的预测效力,以所有患者风险评分的中位数作为标准,在TCGA和CGGA数据库中将样本分为高风险组和低风险组。生存分析表明高风险组胶质瘤患者与低风险组差异具有统计学意义(图3A, 4A),然后对1、3和5年的预后时间进行ROC曲线分析,以检验风险特征的预测效力。结果表明,风险评分在识别胶质瘤总生存期(overall survival, OS)上具有较高的准确性(图3B, 4B)。最后,分析表明随着风险评分增加,死亡患者的数量显著增加,患者生存期呈越来越短的趋势(图3C, 4C)。为了评估风险评分是否能成为预测预后的因素,使用CGGA数据库进行单变量和多变量Cox回归分析,通过单因素分析,观察到风险评分、年龄、WHO级别、IDH状态、1p/19q表达情况



A: 重叠基因分析流程图; B: 韦恩图筛选重叠基因 (筛选条件为  $\text{Log}_2(\text{Fold change}) > 2$  且  $HR > 1$  以及  $\text{Log}_2(\text{Fold change}) < -2$  且  $HR < 1$  的基因, 绿圈中即为满足条件的重叠基因); C: 森林图展示 22 个重叠基因基因预后相关指标; D: LASSO 回归系数筛选图; E: LASSO 回归变量轨迹图。

图 1 预后模型重叠基因的筛选及候选基因鉴定

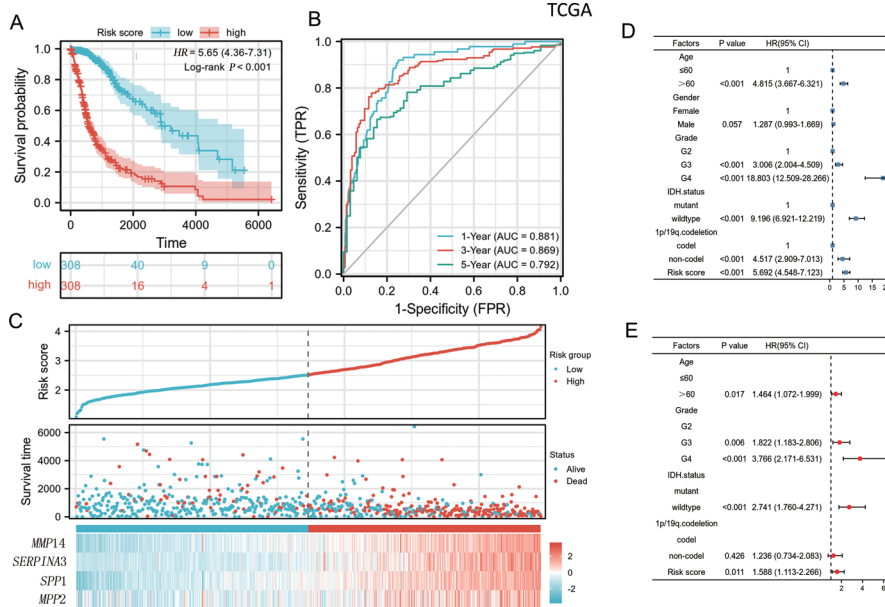


A: TCGA 数据库中分析 4 个候选基因转录水平表达规律; B: PCR 实时定量实验提示 U251 中 4 个候选基因表达水平均高于 HA1800 的表达水平(均  $P < 0.05$ ); C: 基于 HPA 数据验证 4 个候选基因在低级别以及高级别胶质瘤组织中的表达情况; D: TCGA 数据库胶质瘤患者生存分析, 4 个候选基因表达高的胶质瘤患者预后不佳,  $HR_{MMP14} (95\%CI) = 4.55 (3.56 \sim 5.80)$ ,  $\text{Log-rank } P < 0.001$ ;  $HR_{SERPINA3} (95\%CI) = 4.51 (3.53 \sim 5.75)$ ,  $\text{Log-rank } P < 0.001$ ;  $HR_{SPP1} (95\%CI) = 4.55 (3.56 \sim 5.80)$ ,  $\text{Log-rank } P < 0.001$ ;  $HR_{MMP2} (95\%CI) = 3.19 (2.51 \sim 4.05)$ ,  $\text{Log-rank } P < 0.001$ ; E: TCGA 数据库中 4 个候选基因表达增高的胶质瘤患者时间依赖 ROC 曲线; \*\*: 组间比较, 差异有统计学意义 ( $P < 0.05$ )。

图 2 预后模型中候选基因的验证

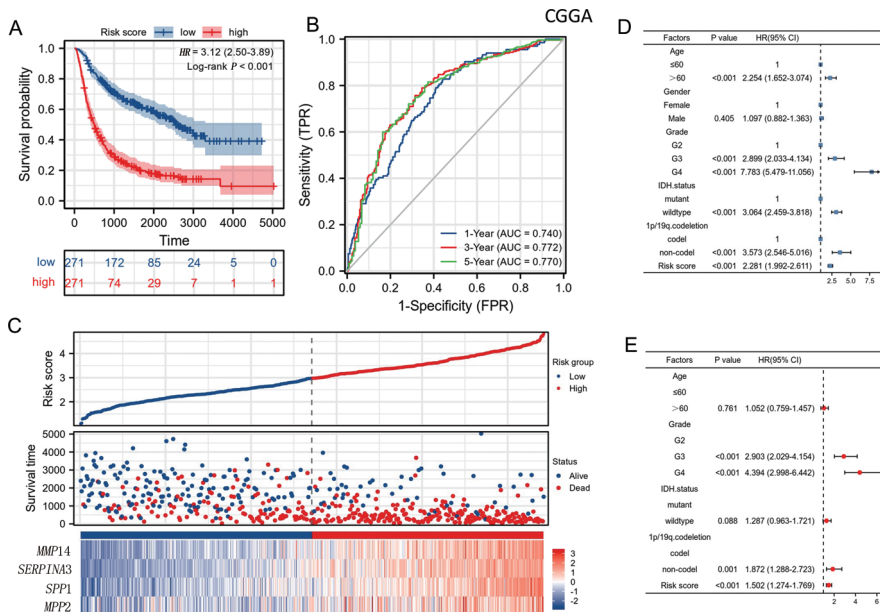
与预后相关(图3D)。基于上述因素进行多因素分析,风险评分仍然与OS密切相关(图3E)。在验证数据库(TCGA)中也得到了相似的结果,在多因素分析中包含了

相同的因素;风险评分也与OS密切相关(图4D, 4E)。一致的结果表明,通过与ECM相关基因构建的风险评分是胶质瘤的一个强有力的预后因素。



A:TCGA数据库胶质瘤患者生存分析,风险评分高的胶质瘤患者预后不佳,HR(95%CI)=5.65(4.36~7.31),Log-rank P<0.001; B:TCGA数据库胶质瘤患者时间依赖ROC曲线,风险评分是较好的预后指标,1年AUC=0.881,3年AUC=0.869,5年AUC=0.792;C:基于风险评分绘制TCGA数据库胶质瘤风险因子图;D:森林图展示TCGA数据库胶质瘤患者单因素风险回归分析结果,风险评分是预后不佳指标;E:森林图展示TCGA数据库胶质瘤患者多因素风险回归分析结果,风险评分是预后不佳指标。

图3 风险评分是TCGA胶质瘤患者的预后指标



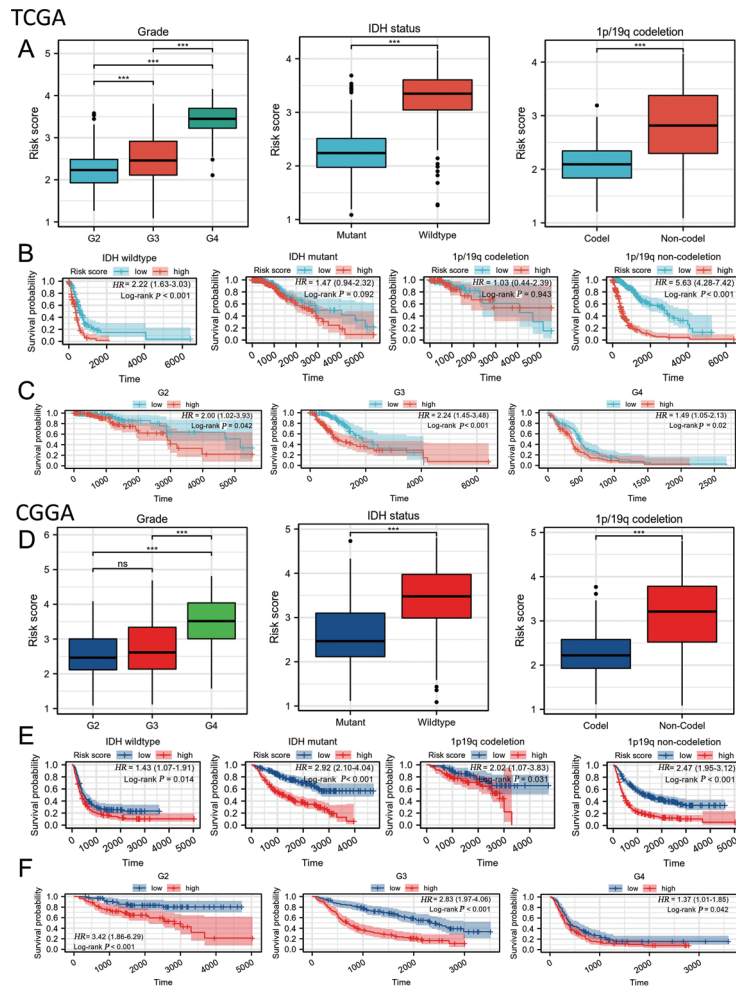
A:CGGA数据库胶质瘤患者生存分析,风险评分高的胶质瘤患者预后不佳,HR(95%CI)=3.12(2.50~3.89),Log-rank P<0.001; B:CGGA数据库胶质瘤患者时间依赖ROC曲线,风险评分是较好的预后指标,1年AUC=0.740,3年AUC=0.772,5年AUC=0.770; C:基于风险评分绘制CGGA数据库胶质瘤风险因子图;D:森林图展示CGGA数据库胶质瘤患者单因素风险回归分析结果,风险评分是预后不佳指标;E:森林图展示CGGA数据库胶质瘤患者多因素风险回归分析结果,风险评分是预后不佳指标。

图4 风险评分是CGGA胶质瘤患者的预后指标

2.3 ECM相关基因风险评分与胶质瘤临床病理特征的关联

为了进一步评估风险评分的预后价值,对TCGA数据库的胶质瘤患者进行了亚组分析,包括等级、IDH状态和1p/19q编码状态,根据相应的病理特征,使用箱式图分析确定了风险评分和病理特征之间具有统计学差异( $t_{G2-3}=4.808, P_{G2-3}=0.000; t_{G2-4}=16.762, P_{G2-4}=0.000; t_{G3-4}=12.740, P_{G3-4}=0.000$ )( $P_{IDH1 \text{ 突变型-野生型}}=0.000$ ,数据不符合正态分布采用U检验验证,  $P_{1p/19q \text{ 联合缺失-非联合缺失}}=0.000$ ,数据不符合正态分布采用U检验验证)(图5A),并且在CGGA数

据库中进行验证,除Grade2-3之间无统计学差异,其余都得到了相同的结果( $t_{G2-3}=2.075, P_{G2-3}=0.114; t_{G2-4}=10.010, P_{G2-4}=0.000; t_{G3-4}=8.752, P_{G3-4}=0.000$ )( $P_{IDH1 \text{ 突变型-野生型}}=0.000$ ,数据不符合正态分布采用U检验验证,  $P_{1p/19q \text{ 联合缺失-非联合缺失}}=0.000$ ,数据不符合正态分布采用U检验验证)(图5D)。此外,利用Kaplan-Meier曲线和tROC曲线来评估风险评分在胶质瘤亚组中的预后价值,分析提示风险评估在不同胶质瘤亚组中均为风险因素( $HR>1$ )(图5B, 5C),在CGGA数据库验证中也得出相同结论(图5E, 5F)。



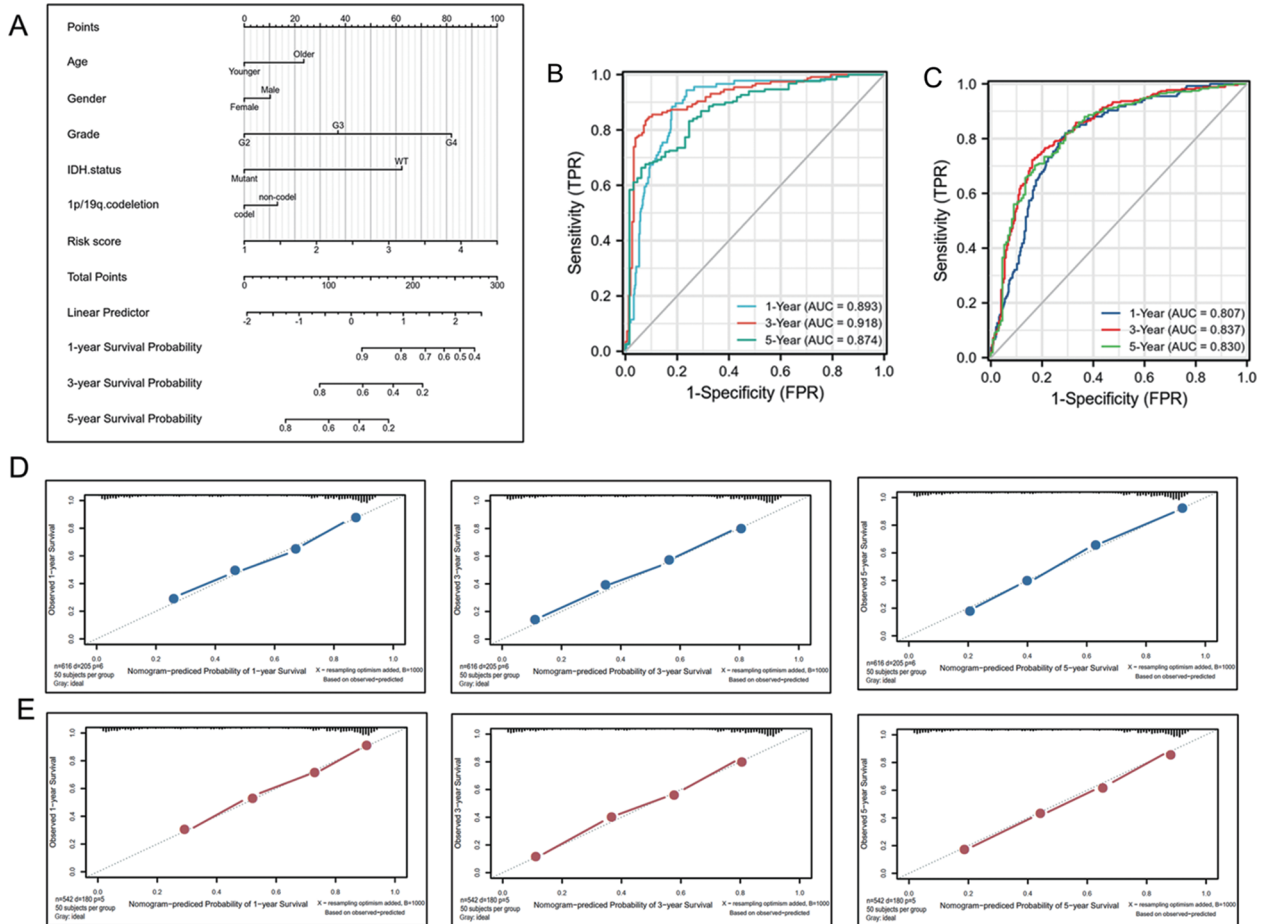
A:TCGA数据库中风险评分在不同类型胶质瘤中的分布,风险评分随肿瘤级别、IDH野生型和1p/19q共缺失,呈正相关性; B:TCGA数据库中IDH突变型胶质瘤生存分析,HR(95%CI)=2.22(1.63~3.03),Log-rank P<0.001;IDH野生型胶质瘤生存分析,HR(95%CI)=1.47(0.94~2.32),Log-rank P=0.092;1p/19q共缺失型,HR(95%CI)=1.03(0.44~2.39),Log-rank P=0.943;非1p/19q共缺失型,HR(95%CI)=5.63(4.28~7.42),Log-rank P<0.001; C:TCGA数据库中WHO2级胶质瘤生存分析,HR(95%CI)=2.00(1.02~3.93),Log-rank P<0.042;WHO3级胶质瘤生存分析HR(95%CI)=2.24(1.45~3.48),Log-rank P<0.001;WHO4级胶质瘤生存分析HR(95%CI)=1.49(1.05~2.13),Log-rank P<0.02; D:CGGA数据库中风险评分在不同类型胶质瘤中的分布,风险评分随肿瘤级别、IDH野生型和1p/19q共缺失,呈正相关性; E:CGGA数据库中IDH突变型胶质瘤生存分析,HR(95%CI)=1.43(1.07~1.91),Log-rank P=0.014;IDH野生型胶质瘤生存分析,HR(95%CI)=2.92(2.10~4.04),Log-rank P<0.001;1p/19q共缺失型,HR(95%CI)=2.02(1.07~3.83),Log-rank P=0.031;非1p/19q共缺失型,HR(95%CI)=2.47(1.95~3.12),Log-rank P<0.001; F:CGGA数据库中WHO2级胶质瘤生存分析,HR(95%CI)=3.42(1.86~6.29),Log-rank P<0.001;WHO3级胶质瘤生存分析HR(95%CI)=2.83(1.97~4.06),Log-rank P<0.001;WHO4级胶质瘤生存分析HR(95%CI)=1.37(1.01~1.85),Log-rank P=0.042; \*\*\*:组间比较,差异有统计学意义(P<0.05)。

图5 TCGA和CGGA数据库中风险评分与胶质瘤临床病理特征的关联

### 2.4 ECM 相关基因风险评估构建 Nomogram 的预后价值

为了进一步加强该模型的临床应用,我们构建了一个包括相关临床特征和预后因素的Nomogram,以预测确诊胶质瘤超过3及5年患者的生存概率(图6A)。为评估Nomogram的预测性能,时间相关的ROC和校准曲线的AUC被用来评估Nomogram的性能。在TCGA和CGGA数

据库中,对胶质瘤患者1,3和5年预后ROC曲线分析提示Nomogram在识别胶质瘤患者OS中均具有较高的准确性(图6B、6C),通过TCGA和CGGA数据进行校准曲线分析,结果表明Nomogram在预测1、3和5年患者生存期中,实际曲线接近理想曲线,提示该Nomogram模型预测效果尚佳(图6D、6E)。



A: 基于TCGA胶质瘤数据库构建风险评分相关的预后列线图; B: 基于TCGA胶质瘤数据库构建校准曲线评价预后列线图; C: 基于CGGA胶质瘤数据库构建校准曲线验证预后列线图; D: 基于TCGA胶质瘤患者数据构建时间依赖ROC曲线,评价预后相关指标; E: 基于CGGA胶质瘤患者数据构建时间依赖ROC曲线,评价预后相关指标。

图6 构建并验证ECM相关基因风险评分的预后列线图

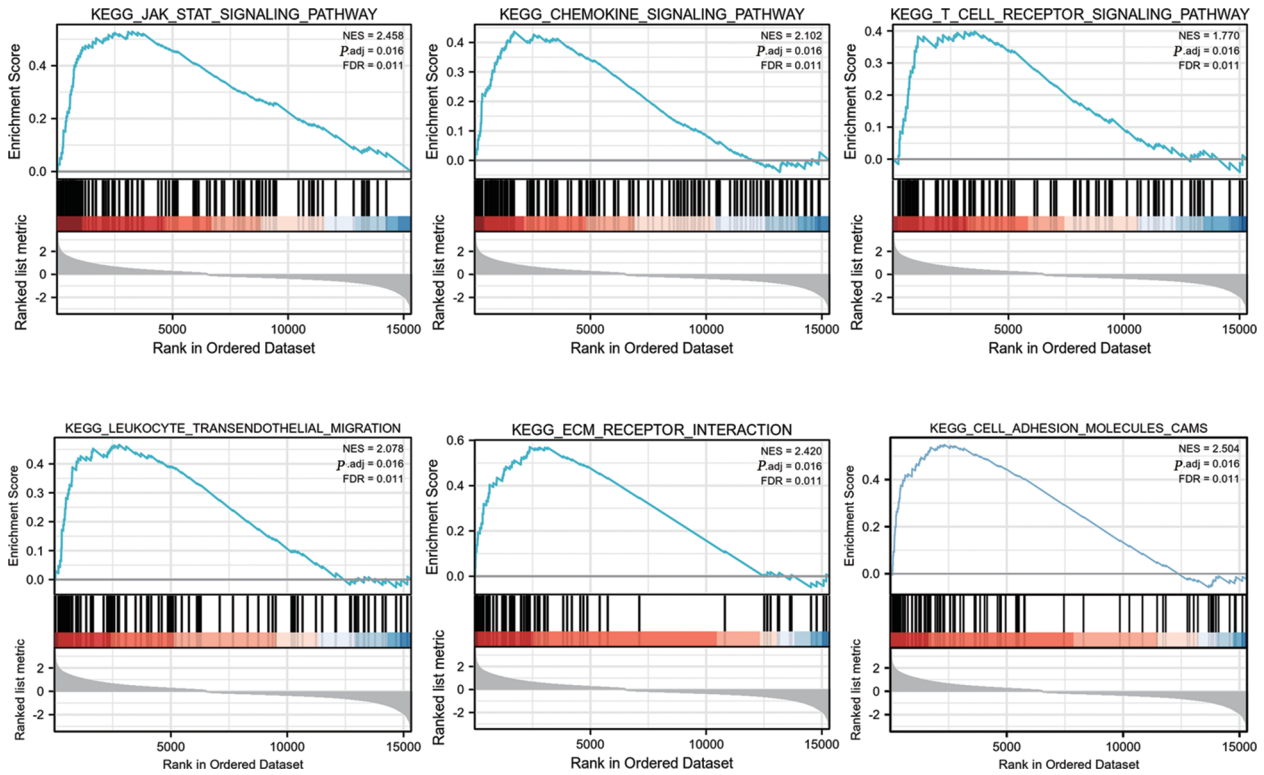
### 2.5 富集分析

为了进一步明确两组间所具有的不同生物学功能及通路活性,将两组所有表达基因采用GSEA分析,分析结果提示:JAK-STAT信号通路、趋化因子信号通路、T细胞抗原受体信号通路、白细胞外周迁移、细胞外基质受体作用和细胞黏附分子功能中富集(图7)。

### 2.6 风险评分与免疫相关性

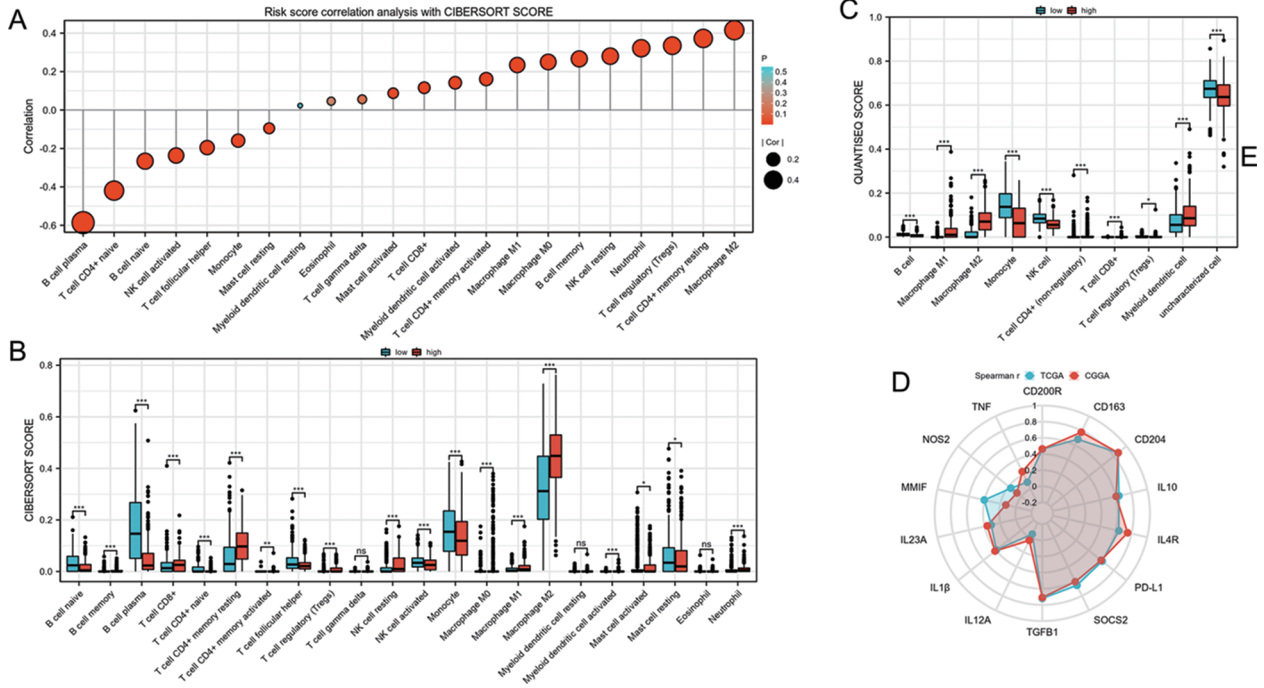
为了说明风险评分与胶质瘤免疫微环境之间的关系,使用CIBERSORT和QUANTISEQ数据库分析免疫浸润水平。在22种免疫细胞亚型的基因表达特征集与风险

评分的相关性(图8A)。此外,为评估风险评分和免疫浸润状态的关系,将高风险组与低风险分别进行免疫细胞亚型基因表达相关性分析,结果提示低风险组免疫细胞浸润性更高,高风险组巨噬细胞M2型分化比例更高(图8B),同时在QUANTISEQ数据库中进行了验证,结果一致(图8C)。此外,在TCGA和CGGA中分析了风险评分与免疫检查点的相关性,结果提示CD163、CD204、IL10、IL4R等免疫检查点具有高正相关性(Spearman  $R > 0.6$ ) IL12A、TNF、NOS2正相关性较弱(Spearman  $R \leq 0.3$ ) (图8D)。



富集分析展示KEGG前6名词条。

图7 高、低风险组胶质瘤患者间富集分析结果

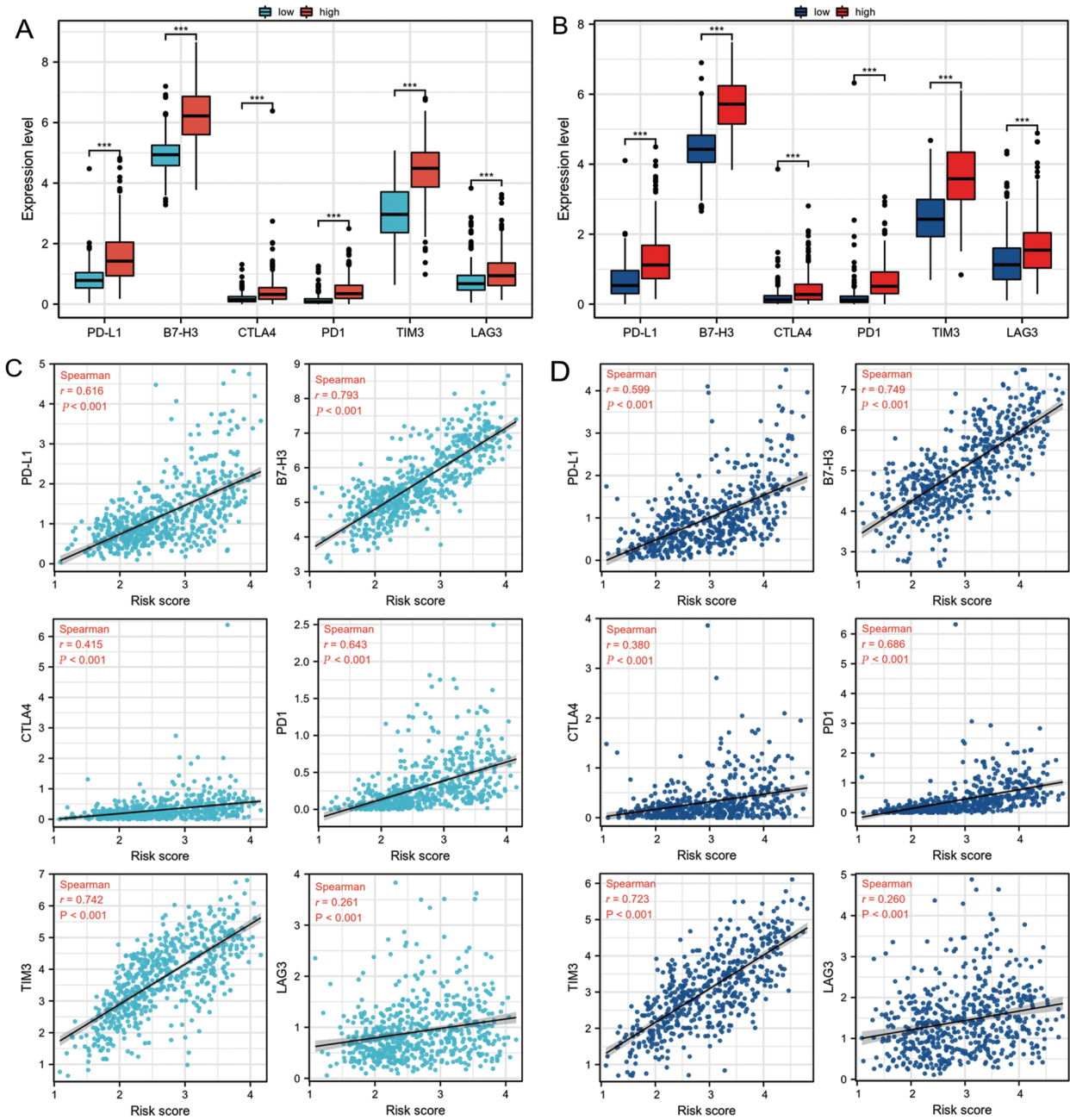


A: 基于CIBERSORT评分评估高、低风险组胶质瘤患者间相关性分析; B: 基于CIBERSORT评分评估高、低风险组胶质瘤患者间免疫浸润; C: 基于QUANTISEQ评分评估高、低风险组胶质瘤患者间免疫浸润; D: 基于TCGA和CGGA数据库评估高、低风险组胶质瘤患者间免疫检查点的相关性; \*\*\*: 组间比较, 差异有统计学意义( $P < 0.05$ )。

图8 高、低风险组胶质瘤患者间肿瘤免疫微环境分析

为讨论风险评分与免疫治疗结合的可能性, 分别在TCGA和CGGA数据库中, 将高低风险组与免疫治疗靶点基因(PD-L1、B7-H3、PD1等)表达进行相关性分析, 结果提示高风险组免疫治疗靶点基因表达增加(图9A和9B), 随后分别在TCGA和CGGA数据库中进行Spearman相关

性评分分析, 结果提示风险评分与PD-L1、B7-H3、PD1和TIM3具有强正相关性(Spearman  $R > 0$  且  $> 0.6$ ), 与CTLA4、LAG3表达具有较弱正相关性(Spearman  $R > 0$  且  $\leq 0.3$ )(图9C和9D)。



A:TCGA数据库中,风险评分与胶质瘤患者免疫检查点标志物表达均为正相关; B:CGGA数据库中,风险评分与胶质瘤患者免疫检查点标志物表达均为正相关; C:基于TCGA数据评估高、低风险组胶质瘤患者间免疫检查点标志物表达的相关性; D:基于CGGA数据评估高、低风险组胶质瘤患者间免疫检查点标志物表达的相关性。\*\*\*:组间比较,差异有统计学意义( $P < 0.05$ )。

图9 风险评分与肿瘤免疫检查点相关性分析

### 3 讨论

本研究基于开源数据库(TCGA、CGGA和GTEX)构建了ECM相关基因预后模型,并对该模型进行了验证。此外,基于该模型,笔者对胶质瘤患者相关特征进行了深入地探索,并基于该特征鉴定了潜在的治疗靶点和药物。笔者同时兼顾了准确性和易用性,尤其针对模型的易用性,通过重叠基因,核心基因和候选基因的逐步筛选流程,缩减了模型中变量的数量,使得其临床应用更加简便易行。与此同时,基于MMP14、SERPINA3、SPP1和MMP2的4个ECM相关基因风险评分能较好地预测患者预后,在TCGA和CGGA数据库中高风险组患者预后差于低风险组,且风险评分是患者预后的预后危险因素,iROC表明该评分预测患者预后效能良好,在2个数据库中1、3和5年的AUC值均大于0.7。

在筛选变量的过程中,本研究创新地引入了HPA数据库蛋白表达验证步骤。HPA数据库内置了多瘤种的转录水平以及蛋白水平的数据集。尤其针对蛋白水平,针对同一蛋白该数据库至少选择两种独立抗体进行免疫组化染色,以确保获得准确的染色结果。因此,笔者针对MMP14、SERPINA3、SPP1和MMP2等变量进行了转录水平和蛋白水平的验证,以确保其在胶质瘤中表达的准确性。基于CGGA数据库转录水平验证发现,MMP14、SERPINA3、SPP1和MMP2在胶质瘤中4个基因均表达增高。因此转录水平的验证基本符合预期(图2A)。在蛋白水平的验证中,笔者发现MMP14、SERPINA3、SPP1和MMP2在HCG中均表达增高(图2B)。因此蛋白水平的验证基本符合预期。

笔者根据4个候选基因计算的风险评分,经分析表明风险评分导致胶质瘤预后不良的一个因素。研究表明,肿瘤微环境中特有的表达SPP1的巨噬细胞,该类细胞具有独特的代谢特征、泡沫细胞(foam cell)样特征<sup>[21]</sup>以及活跃的细胞外基质重塑功能<sup>[22]</sup>。SPP1的过度表达可能诱导巨噬细胞介导MES样细胞活化,从而达到肿瘤免疫豁免<sup>[23-24]</sup>。胶质瘤中MMP14表达下降后体内荷瘤小鼠降低了肿瘤异种移植概率,并提高了小鼠的存活率,同时在体外实验中提示MMP14过度表达将诱导胶质瘤细胞更强的耐药性和耐放射性<sup>[25]</sup>。MMP2的表达降低影响了胶质瘤细胞的增殖能力,同时入侵及迁徙能力也受到了影响<sup>[26]</sup>,这些选定的ECM相关基因作为胶质瘤治疗的靶点值得进一步研究。

KEGG分析发现,DEGs主要富集在JAK-STAT信号通路<sup>[27]</sup>、趋化因子信号通路、T细胞抗原受体信号通路<sup>[28]</sup>、白细胞外周迁移<sup>[29]</sup>、细胞外基质受体作用和细胞黏附等词条中(图7A),高、低风险组间差异主要富集于免疫相关生物活动有关。CIBERSORT的结果表明,免疫细胞浸润水平与风险评分相关。分析指出风险评分与巨噬

细胞M2分型呈正相关性,众所周知巨噬细胞M2型主要功能为抑制炎症,从而在肿瘤微环境中起到抑制肿瘤的免疫作用,同时巨噬细胞的M1/M2比例失调也是肿瘤发生发展<sup>[30]</sup>,免疫逃逸及耐药重要原因<sup>[31-32]</sup>,风险评分与所有M2型标志物呈正相关,进一步佐证了多重免疫浸润分析的结果。上述结果表明,M2型巨噬细胞在高风险组中增多,可能是其抑制性免疫微环境的重要构成因素。

近年来对肿瘤细胞ECM在实体肿瘤中的研究和认识呈指数增长,认为肿瘤周围细胞外基质在生物化学、生物力学、结构组成和分布上发生改变,从而影响肿瘤细胞迁移、增殖、分化等重要生物学功能<sup>[33]</sup>。近期研究指出,ECM在实体肿瘤的侵袭性中起到重要影响作用<sup>[34]</sup>,并且ECM中的信号传导可以驱动肿瘤细胞的增殖、分化和免疫逃逸<sup>[4, 35]</sup>。

胶质瘤作为高发原发性颅脑恶性肿瘤,具有较高的发病率和死亡率,患者确诊罹患脑胶质瘤5年生存率不足10%。近年来,在胶质瘤治疗方案研究中,未见十分有效改善患者生存质量的方案<sup>[36]</sup>。肿瘤免疫治疗的发展,特别是免疫检查点阻断疗法在癌症治疗方面取得了一定进展,然而仅有少部分患者能够通过肿瘤免疫治疗手段获得显著治疗效果。近十年内,越来越多研究表明,ECM作为肿瘤微环境的重要一员在肿瘤免疫抑制和免疫逃逸中起主要作用<sup>[37-38]</sup>,肿瘤相关ECM相关研究将会是提升肿瘤免疫治疗效果的金钥匙。ECM长期以来一直被定义为一种惰性结构而被忽略,但如今已被认为免疫系统高动态合作伙伴。ECM不仅可以为细胞的正常生理活动提供结构支撑,其丰富的蛋白成分以及免疫活性分子在机体稳态和病理状态下,均发挥着不可或缺的免疫调节作用。反之,免疫系统又可维持基质微环境稳态、修复损伤后ECM的完整性<sup>[37, 39-41]</sup>。

综上所述,肿瘤ECM的状态与肿瘤发生发展和免疫活动息息相关,这对我们更加深入了解肿瘤机制具有很大的帮助。此外风险评分与免疫检查点的表达呈正相关,这说明高风险组患者具有更高优先级免疫治疗的临床意义,极大地增加了笔者所构建的风险评分的临床意义,在今后的研究中,笔者会在此基础上深入探索风险评分与免疫功能关系,拓宽临床应用范围。随着对肿瘤ECM研究深入,笔者相信ECM相关基因将成为胶质瘤治疗的推进器。

### 参 考 文 献

- [1] BONNANS C, CHOU J, WERB Z. Remodelling the extracellular matrix in development and disease[J]. Nat Rev Mol Cell Biol, 2014, 15(12): 786-801.
- [2] CUI X, MORALES RTT, QIAN WY, et al. Hacking macrophage-associated immunosuppression for regulating glioblastoma angiogenesis[J]. Biomaterials, 2018, 161: 164-178.

- [3] FERRER VP, MOURA NETO V, MENTLEIN R. Glioma infiltration and extracellular matrix: key players and modulators [J]. *Glia*, 2018, 66(8): 1542-1565.
- [4] KIYOKAWA J, KAWAMURA Y, GHOUSE SM, et al. Modification of extracellular matrix enhances oncolytic adenovirus immunotherapy in glioblastoma[J]. *Clin Cancer Res*, 2021, 27(3): 889-902.
- [5] TAGA T, TABU K. Glioma progression and recurrence involving maintenance and expansion strategies of glioma stem cells by organizing self-advantageous niche microenvironments[J]. *Inflamm Regen*, 2020, 40: 33.
- [6] THEOCHARIS AD, MANOU D, KARAMANOS NK. The extracellular matrix as a multitasking player in disease[J]. *FEBS J*, 2019, 286(15): 2830-2869.
- [7] OSTROM QT, GITTLEMAN H, TRUITT G, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2011-2015[J]. *Neuro Oncol*, 2018, 20(suppl\_4): iv1-iv86.
- [8] ALEXANDER BM, CLOUGHESY TF. Adult glioblastoma[J]. *J Clin Oncol*, 2017, 35(21): 2402-2409.
- [9] LOUIS DN, PERRY A, WESSELING P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary [J]. *Neuro Oncol*, 2021, 23(8): 1231-1251.
- [10] SCHAFF LR, MELLINGHOFF IK. Glioblastoma and other primary brain malignancies in adults: a review[J]. *JAMA*, 2023, 329(7): 574-587.
- [11] PAOLILLO M, BOSELLI C, SCHINELLI S. Glioblastoma under siege: an overview of current therapeutic strategies[J]. *Brain Sci*, 2018, 8(1): 15.
- [12] OMURO A, DEANGELIS LM. Glioblastoma and other malignant gliomas: a clinical review[J]. *JAMA*, 2013, 310(17): 1842-1850.
- [13] FABIAN C, HAN MZ, BJERKVIK R, et al. Novel facets of glioma invasion[J]. *Int Rev Cell Mol Biol*, 2021, 360: 33-64.
- [14] BAI J, VARGHESE J, JAIN R. Adult glioma WHO classification update, genomics, and imaging: what the radiologists need to know[J]. *Top Magn Reson Imaging*, 2020, 29(2): 71-82.
- [15] AIBAIDULA A, CHAN AKY, SHI ZF, et al. Adult IDH wild-type lower-grade gliomas should be further stratified[J]. *Neuro Oncol*, 2017, 19(10): 1327-1337.
- [16] ZARNETT OJ, SAHGAL A, GOSIO J, et al. Treatment of elderly patients with glioblastoma: a systematic evidence-based analysis [J]. *JAMA Neurol*, 2015, 72(5): 589-596.
- [17] REARDON DA, BRANDES AA, OMURO A, et al. Effect of nivolumab vs bevacizumab in patients with recurrent glioblastoma: the CheckMate 143 phase 3 randomized clinical trial[J]. *JAMA Oncol*, 2020, 6(7): 1003-1010.
- [18] DAUBON T, HEMADOU A, ROMERO GARMENDIA I, et al. Glioblastoma immune landscape and the potential of new immunotherapies[J]. *Front Immunol*, 2020, 11: 585616.
- [19] CHEN Q, HAN B, MENG XQ, et al. Immunogenomic analysis reveals LGALS1 contributes to the immune heterogeneity and immunosuppression in glioma[J]. *Int J Cancer*, 2019, 145(2): 517-530.
- [20] CHEN ZH, HAMBARDZUMYAN D. Immune microenvironment in glioblastoma subtypes[J]. *Front Immunol*, 2018, 9: 1004.
- [21] PANG XC, ZHANG JL, HE X, et al. SPP1 promotes enzalutamide resistance and epithelial-mesenchymal-transition activation in castration-resistant prostate cancer via PI3K/AKT and ERK1/2 pathways[J]. *Oxid Med Cell Longev*, 2021, 2021: 5806602.
- [22] LIU LL, ZHANG RY, DENG JW, et al. Construction of TME and identification of crosstalk between malignant cells and macrophages by SPP1 in hepatocellular carcinoma[J]. *Cancer Immunol Immunother*, 2022, 71(1): 121-136.
- [23] HE C, SHENG LY, PAN DS, et al. Single-Cell transcriptomic analysis revealed a critical role of SPP1/CD44-mediated crosstalk between macrophages and cancer cells in glioma[J]. *Front Cell Dev Biol*, 2021, 9: 779319.
- [24] DE SCHEPPER S, GE JZ, CROWLEY G, et al. Perivascular cells induce microglial phagocytic states and synaptic engulfment via SPP1 in mouse models of Alzheimer's disease[J]. *Nat Neurosci*, 2023, 26(3): 406-415.
- [25] ULASOV I, THACI B, SARVAIYA P, et al. Inhibition of MMP14 potentiates the therapeutic effect of temozolomide and radiation in gliomas[J]. *Cancer Med*, 2013, 2(4): 457-467.
- [26] WU H, LIU L, ZHU JM. MiR-93-5p inhibited proliferation and metastasis of glioma cells by targeting MMP2[J]. *Eur Rev Med Pharmacol Sci*, 2019, 23(21): 9517-9524.
- [27] HU XY, LI J, FU MR, et al. The JAK/STAT signaling pathway: from bench to clinic[J]. *Signal Transduct Target Ther*, 2021, 6(1): 402.
- [28] SCHUBERT ML, SCHMITT M, WANG L, et al. Side-effect management of chimeric antigen receptor (CAR) T-cell therapy [J]. *Ann Oncol*, 2021, 32(1): 34-48.
- [29] LIN YJ, WEI KC, CHEN PY, et al. Roles of neutrophils in glioma and brain metastases[J]. *Front Immunol*, 2021, 12: 701383.
- [30] GUNASSEKARAN GR, POONGKAVITHAI VADEVOO SM, BAEK MC, et al. M1 macrophage exosomes engineered to foster M1 polarization and target the IL-4 receptor inhibit tumor growth by reprogramming tumor-associated macrophages into M1-like macrophages[J]. *Biomaterials*, 2021, 278: 121137.
- [31] ZHANG M, HUTTER G, KAHN SA, et al. Anti-CD47 treatment stimulates phagocytosis of glioblastoma by M1 and M2 polarized macrophages and promotes M1 polarized macrophages *in vivo*[J]. *PLoS One*, 2016, 11(4): e0153550.
- [32] LOCATI M, CURTALE G, DiversityMANTOVANI A, mechanisms, and significance of macrophage plasticity[J]. *Annu Rev Pathol*, 2020, 15: 123-147.
- [33] COX TR. The matrix in cancer[J]. *Nat Rev Cancer*, 2021, 21(4):

- 217-238.
- [34] ELIA I, ROSSI M, STEGEN S, et al. Breast cancer cells rely on environmental pyruvate to shape the metastatic niche[J]. *Nature*, 2019, 568(7750): 117-121.
- [35] GOLDBRUNNER RH, BERNSTEIN JJ, TONN JC. ECM-mediated glioma cell invasion[J]. *Microsc Res Tech*, 1998, 43(3): 250-257.
- [36] GUSYATINER O, HEGI ME. Glioma epigenetics: from subclassification to novel treatment options[J]. *Semin Cancer Biol*, 2018, 51: 50-58.
- [37] SUTHERLAND TE, DYER DP, ALLEN JE. The extracellular matrix and the immune system: A mutually dependent relationship[J]. *Science*, 2023, 379(6633): eabp8964.
- [38] KOLESNIKOFF N, CHEN CH, SAMUEL MS. Interrelationships between the extracellular matrix and the immune microenvironment that govern epithelial tumour progression[J]. *Clin Sci (Lond)*, 2022, 136(5): 361-377.
- [39] WINKLER J, ABISOYE-OGUNNIYAN A, METCALF KJ, et al. Concepts of extracellular matrix remodelling in tumour progression and metastasis[J]. *Nat Commun*, 2020, 11(1): 5120.
- [40] GHORBANI S, YONG VW. The extracellular matrix as modifier of neuroinflammation and remyelination in multiple sclerosis[J]. *Brain*, 2021, 144(7): 1958-1973.
- [41] MAROZZI M, PARNIGONI A, NEGRI A, et al. Inflammation, extracellular matrix remodeling, and proteostasis in tumor microenvironment[J]. *Int J Mol Sci*, 2021, 22(15): 8102.

责任编辑:王荣兵