

本文引文格式:李永乐,韦万硕,蒋利和. SERINC1 表达在肺癌中的预后作用及与肿瘤免疫浸润的关系[J]. 右江民族医学院学报, 2024, 46(4): 486-496.

【论著与临床报道】

## SERINC1 表达在肺癌中的预后作用及与肿瘤免疫浸润的关系

李永乐<sup>1</sup>, 韦万硕<sup>2</sup>, 蒋利和<sup>1</sup>

(1. 右江民族医学院基础医学院, 广西 百色 533000;

2. 右江民族医学院临床医学院, 广西 百色 533000)

**摘要:** **目的** 本研究旨在探讨血清素转运蛋白 1(SERINC1)在肺癌预后中的作用。**方法** 从 GEO 数据库下载了肺癌的单细胞测序 scRNA 数据,提取上皮细胞显著表达基因。同时 Xena 数据库下载 TCGA 数据,并进行了差异分析。将以上两者差异显著表达基因取交集,得到了上皮细胞差异基因。使用 LASSO、随机森林、GBM、XGBoost 等多种机器学习算法取交集,得到了 FAM107A、SERINC1、MDK、GGCT、AVL9 和 FEZ1 6 个交集基因。由于 SERINC1 在肺癌中还未被研究,对 SERINC1 进行了一系列的临床相关性分析,包括诊断 ROC 分析、表达分析、富集分析、生存分析、蛋白表达、蛋白互作分析以及免疫浸润分析。**结果** 通过单细胞测序和机器学习筛选出了目标基因 SERINC1。ROC 分析 SERINC1 的 AUC 值高达 0.972,显示出很高的诊断效能。K-M 曲线显示,SERINC1 表达高的患者的预后明显好于表达量低的患者,其 HR 值为 0.71(95% CI :0.63~0.81)和 0.48(95% CI :0.39~0.61),这表明基因表达量越高的患者预后更好。免疫浸润分析揭示,SERINC1 与多种免疫细胞相关,且与多种 T 细胞的标志基因集相关。**结论** SERINC1 是肺癌的一个有前景的预后标志物,且与免疫浸润相关。

**关键词:** 肺肿瘤;单细胞测序;机器学习;血清素转运蛋白 1

**中图分类号:** R734.2 **文献标识码:** A **文章编号:** 1001-5817(2024)04-0486-11

doi:10.3969/j.issn.1001-5817.2024.04.006

### Prognostic effect of SERINC1 expression in lung cancer and its relationship with tumor immune infiltration

LI Yongle<sup>1</sup>, WEI Wanshuo<sup>2</sup>, JIANG Lihe<sup>1</sup>

(1. School of Basic Medicine, Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China; 2. School of Clinical Medicine, Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China)

**Abstract:** **Objective** This study aims to explore the role of SERINC1 (serotonin transporter 1) in the prognosis of lung cancer. **Methods** Single-cell RNA sequencing (scRNA) data of lung cancer were downloaded from the GEO database and genes significantly expressed in epithelial cells were extracted. Concurrently, TCGA data were downloaded from the Xena database and subjected to differential analysis. The intersection of significantly expressed genes from both sources identified differential genes in epithelial cells. Intersection of multiple machine learning algorithms, including Lasso, Random Forest, GBM, and XGBoost, yielded six intersecting genes: FAM107A, SERINC1, MDK, GGCT, AVL9, and FEZ1. As the effect of SERINC1 in lung cancer has not been studied, a series of clinical correlation analyses were conducted on SERINC1, including diagnostic ROC analysis, expression analysis, enrichment analysis, survival analysis, protein expression, pro-

**基金项目:**右江民族医学院高层次引进人才项目经费(yy2021sk002);广西壮族自治区大学生创新创业训练计划项目(S202310599056)

**第一作者:**李永乐,大学生,E-mail:1804426453@qq.com

**通讯作者:**蒋利和,博士,研究员,教授,硕士研究生导师,研究方向:肿瘤生物学,E-mail:Jianglihe@ymun.edu.cn

tein-protein interaction analysis, and immune infiltration analysis. **Results** The target gene SERINC1 was identified through single-cell sequencing and machine learning. The AUC value of ROC analysis for SERINC1 was as high as 0.972, indicating a high diagnostic efficacy. The Kaplan-Meier curve indicated that patients with high SERINC1 expression had significantly better prognosis than those with low expression, with HR values of 0.71 (95% CI :0.63~0.81) and 0.48 (95% CI :0.39~0.61), suggesting that higher gene expression is associated with better prognosis. Immune infiltration analysis revealed that SERINC1 is related to various immune cells and is associated with marker gene-sets of various T cells. **Conclusion** SERINC1 is a promising prognostic marker for lung cancer and is related to immune infiltration.

**Key words:** lung cancer; single-cell sequencing; machine learning; serotonin transporter 1

肺癌是全球癌症死亡率的主要原因之一<sup>[1]</sup>,其中非小细胞肺癌占有肺癌病例的大多数,非小细胞肺癌又分为腺癌、鳞癌以及大细胞癌。非小细胞肺癌的诊断和治疗同样面临巨大挑战,特别是在晚期,5年生存率仍然很低。肺癌的发展和转移与基因表达的异常调节有着密切的联系,基因表达的变化能够反映出肿瘤的生物特性和功能。因此,识别影响肺癌基因表达的关键因子对于提高诊断的准确性和治疗效果至关重要。

上皮细胞作为构成肺部组织的主要细胞类型,包括气道上皮细胞和肺泡上皮细胞,在肺部的结构和功能中发挥着核心作用,并且是吸入性致癌物质的主要靶点<sup>[2]</sup>。肺癌通常起源于上皮组织,特别是在长期暴露于吸烟等致癌因素下,上皮细胞可能发生基因突变,导致肺癌的发生<sup>[3]</sup>。并且上皮细胞通过上皮-间质转换(EMT)过程可以获得移动性和侵袭性,这是肺癌发展和转移的重要过程。EMT过程中,上皮细胞失去其特有的细胞连接和极性,转变为具有迁移能力的间质细胞<sup>[4]</sup>。因此,推测上皮细胞中的一些基因可能与肺癌的发生和发展相关。在本研究中,利用单细胞测序技术对细胞类型进行注释,并筛选出在上皮细胞中显著表达的基因。为了验证这一假设,将上皮细胞显著表达的基因与TCGA数据库中的差异基因进行交集分析,识别出上皮细胞显著表达的差异基因。接着,运用多种机器学习算法对这些基因进行筛选和评估,最终确定血清素转运蛋白1(SERINC1)为研究的目标基因。

## 1 材料与方法

1.1 数据获取 为了获取所需的数据,本课题组从多个公共数据库和在线数据库中下载了相关的数据集。首先,从GEO数据库<sup>[5]</sup>下载了肺癌的单细胞测序数据GSE198291。GEO是由美国国立卫生研究院(NIH)管理的公共数据库,专门用于存储基因表达数据和其他与基因功能相关的实验数据。其次,从Xena数据库下载了TCGA肺癌转录组数据。Xena是由UCSC Xena项目提供的在线数据集成和可视化平台,

它包含了来自TCGA、ICGC和GTEx等项目的大量基因组数据,以及来自临床试验和疾病预后研究的病人临床信息。所有数据获取时间均为2024年2月4日。

1.2 单细胞测序分析 使用Seurat软件包对GSE198291数据进行了质量控制和标准化处理,去除了低质量的细胞和基因,以及可能的污染物和双重计数<sup>[6]</sup>。对标准化后的数据进行了PCA,以降低数据的维度,并揭示数据中的主要变异模式<sup>[7]</sup>。根据PCA的方差解释率和细胞聚类效果,选择了合适的主成分个数,作为后续分析的输入。使用t-SNE方法对PCA的结果进行了非线性降维和可视化,以识别数据中的不同细胞群体<sup>[8]</sup>。使用FindClusters函数对细胞进行了基于密度的聚类,根据模块化系数和轮廓系数,选择了合适的分辨率参数,得到了最优的细胞聚类结果<sup>[9-11]</sup>。随后使用SingleR包进行细胞类型注释,其中参考数据集为celldex:Human Primary Cell Atlas Data()提供的人类主要细胞图谱数据,通过SingleR分析,将每个细胞与最相似的参考细胞类型进行匹配,得到每个细胞的预测标签,使用FindAllMarkers函数识别在不同细胞类型之间表达差异显著的基因,为了识别在上皮细胞中显著表达的基因,设置了1个最小百分比阈值(min. pct),要求1个基因在至少25%的上皮细胞中被检测到表达,并且设置对数倍数变化阈值(logfc. threshold)为1,只考虑在上皮细胞中相对于其他细胞类型至少有2倍表达量变化的基因。然后,从差异表达分析的结果中,提取与上皮细胞相关的显著差异表达基因。使用Monocle软件包对上皮细胞进行了细胞轨迹分析,以揭示细胞状态的动态变化和发育过程<sup>[12-13]</sup>。使用了基于反向图的伪时序重构方法,根据细胞的表达谱,构建了细胞的分支轨迹,并使用了基于差异表达的分支点检测方法,确定了细胞的分支点和分支方向。

1.3 TCGA转录组数据差异分析 使用DESeq2包进行差异分析。使用median ratio方法对数据进行了标准化<sup>[14]</sup>,使用maximum likelihood方法对数据进行

了离散度估计,使用 Wald test 方法对数据进行了差异检验。设定  $|\log_2FC| > 1$  和  $P < 0.05$  作为差异表达基因的筛选阈值。

1.4 上皮细胞基因与差异基因取交集并提取其表达矩阵 差异基因与上皮细胞显著基因取交集,得到上皮细胞差异基因,提取其在 TCGA 样本中的 Bulk 转录组数据。

1.5 4 种机器学习方法筛选特征基因 使用了 4 种机器学习算法来筛选出对疾病预测最具影响力的基因。其中,LASSO(Least Absolute Shrinkage and Selection Operator,LASSO)回归是一种基于线性回归的变量选择方法<sup>[15]</sup>,随机森林算法(Random Forest Algorithm,RFA)是一种基于多个决策树的集成学习方法<sup>[16]</sup>,GBM (Gradient Boosting Machine,GBM)是一种基于梯度提升的迭代决策树算法<sup>[17]</sup>,XGBoost (eXtreme Gradient Boosting,XGB)是一种针对 GBM 进行优化的分布式梯度提升库<sup>[18]</sup>。利用 LASSO、随机森林、GBM、XGBoost 这 4 种机器学习算法分别从上皮细胞差异基因中筛选特征基因。4 种机器学习方法具体筛选过程如下,LASSO:通过 glmnet 函数构建了正则化路径,并使用 cv.glmnet 函数进行了 10 折交叉验证,以确定最佳的正则化参数 lambda,随后选择了在交叉验证中偏差最小的 lambda 值作为最优值,并从中提取了具有非零系数的特征基因。随机森林:首先训练了一个包含 1 000 棵决策树的随机森林模型。然后,通过模型的误差率来确定最优的树的数量。接下来,计算了每个基因的 MeanDecreaseGini 重要性评分,该评分反映了基因在随机森林模型中的贡献度。最后,选择了重要性评分  $> 2$  的基因作为疾病特征基因。GBM:首先设置了随机种子以确保结果的可重复性。接着,使用 5 折交叉验证来控制模型拟合过程。通过 GBM 模型,计算了每个基因的重要性,并选择了重要性不为零的基因作为特征基因。XGBoost:首先设置了随机种子以确保结果的可重复性。接着,使用重复的 10 折交叉验证来控制模型拟合过程。通过 XGBoost 模型,计算了每个基因的重要性,并选择了重要性不为零的基因作为特征基因。随后分析这 4 种机器学习筛选的基因集之间的特有和共有的部分,并且用 ggplot2 包和 VennDiagram 包对结果进行可视化并绘制韦恩图,以获得最终的候选基因。

1.6 受试者工作特征曲线 (ROC) 曲线判断 SERINC1 对肺癌的诊断价值使用了 pROC<sup>[19]</sup>包<sup>[19]</sup>,对 598 个 TCGA 样本(其中正常样本 59 个、肿瘤样本 539 个)进行了 ROC 分析,并用 ggplot2 包<sup>[20]</sup>绘制了 ROC 曲线。

1.7 表达量分析 使用 ggplot2 包分析 SERINC1 基

因在 The Cancer Genome Atlas (TCGA) 数据库中 598 个样本(其中正常样本 59 个、肿瘤样本 539 个)的表达情况。利用 TIMER 2.0 数据库<sup>[21]</sup>进一步探索 SERINC1 在泛癌中的表达情况。

1.8 单基因差异分析 为了探索与 SERINC1 表达水平有显著相关性的基因,对 SERINC1 基因进行了单基因差异分析。首先计算基因 SERINC1 在所有样本中的表达水平的中位数,并根据 SERINC1 表达水平高于或低于中位数将样本分为两组(high 和 low),使用 DESeq Data Set From Matrix 函数创建 DESeq2 分析所需的数据集对象,运行 DESeq2 函数进行差异表达分析,最后提取对数 2 倍变化( $\log_2$ Fold Change)绝对值  $> 1$  且  $P < 0.05$  的基因,这些基因在统计上与 SERINC1 表达水平的高低显著相关。

1.9 富集分析 对与 SERINC1 表达相关的基因进行了富集分析<sup>[22]</sup>,以检测这些基因是否在特定的生物过程或通路中有过多或过少的表达。富集分析是一种统计方法,可以评估一个基因集合(本研究中为 SERINC1 相关基因)与某些功能注释的关联性。将上一步单基因差异分析筛选的基因纳入富集分析,首先利用 org.Hs.eg.db 包进行 ID 转换,然后用 clusterProfiler 包进行富集分析。

1.10 生存分析 为了探索 SERINC1 基因表达水平对肿瘤患者生存的影响,使用了 Kaplan-Meier Plotter 这一在线数据库<sup>[23]</sup>,输入 SERINC1 的基因符号,并根据不同的癌症类型和生存指标进行了分析。

1.11 PPI 蛋白互作网络分析 GeneMANIA 数据库(<http://genemania.org>)能够预测基因功能。利用 GeneMANIA 数据库丰富的基因组学和蛋白质组学数据构建了与 SERINC1 相互作用的基因网络。STRING 数据库<sup>[24]</sup>(<https://string-db.org/>)是分析蛋白质相互作用的数据库。利用该数据库构建了 SERINC1 的蛋白质-蛋白质相互作用(PPI)网络。

1.12 CIBERSRT 分析 对样本进行 CIBERSORT 分析<sup>[25]</sup>,并且计算基因表达与这些免疫细胞丰度之间的相关性。

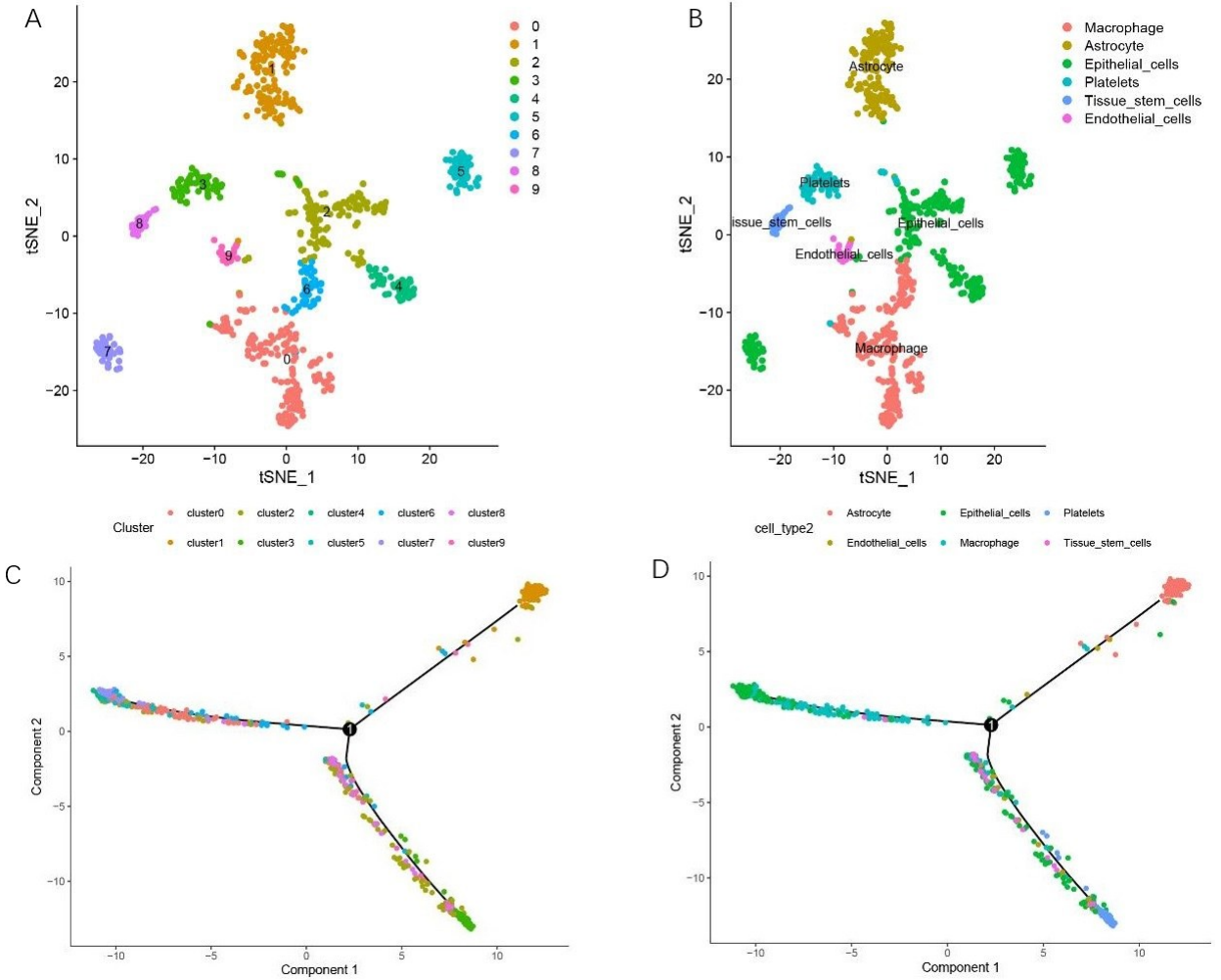
1.13 蛋白表达水平 利用 HPA 数据库<sup>[26]</sup>,分析了 SERINC1 基因在肺癌样本和正常样本中的蛋白质表达水平。HPA 数据库是一个公开的资源,收集了大量的人类蛋白质表达数据,基于免疫组化、免疫荧光和 RNA 测序等技术。

## 2 结果

2.1 单细胞测序分析 在单细胞测序分析中,对细胞进行了类群注释,成功地标注出了 6 类细胞,包括巨噬细胞(macrophages)、星形胶质细胞(astrocytes)、上皮细胞(epithelial cells)、血小板(platelets)、组织干细胞

(tissue stem cells), 以及内皮细胞(endothelial cells) (见图 1A、图 1B)。通过单细胞拟时序分析, 发现星形胶质细胞是最早出现的细胞类型, 而上皮细胞则是最后出现的(见图 1C、图 1D)。上皮细胞在肺部的结构

和功能中起着关键作用, 也是吸入性致癌物质的主要靶点<sup>[27-28]</sup>。因此, 本课题组选择了在上皮细胞中显著表达的 206 个基因进行进一步的研究。



注: A. 细胞分群; B. 细胞分群后注释; C、D. 拟时序分析结果。

图 1 单细胞测序分析

2.2 TCGA 差异分析 使用了  $|\log_2FC| > 1$  和  $P < 0.05$  作为差异表达基因的筛选阈值, 从而得到了 5395 个 DEGs, 见图 2。

2.3 机器学习筛选特征基因 使用了 4 种机器学习算法来筛选出与疾病预测相关的特征基因。其中, LASSO 回归选出了 21 个基因(见图 3A), 随机森林选出了 13 个基因(见图 3B、图 3C), GBM 选出了 44 个基因(见图 3D), XGBOOST 选出了 21 个基因(见图 3E), 其中, GBM 和 XGBoost 的基因筛选标准均为重要性  $> 0$ 。将这 4 种算法得到的特征基因进行交集分析, 最终确定了 6 个候选基因(见图 3F), 分别为 FAM107A、SERINC1、MDK、GGCT、AVL9 和 FEZ1。

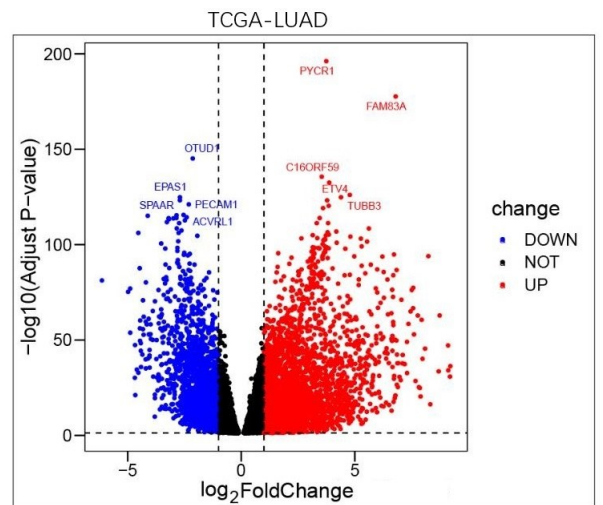
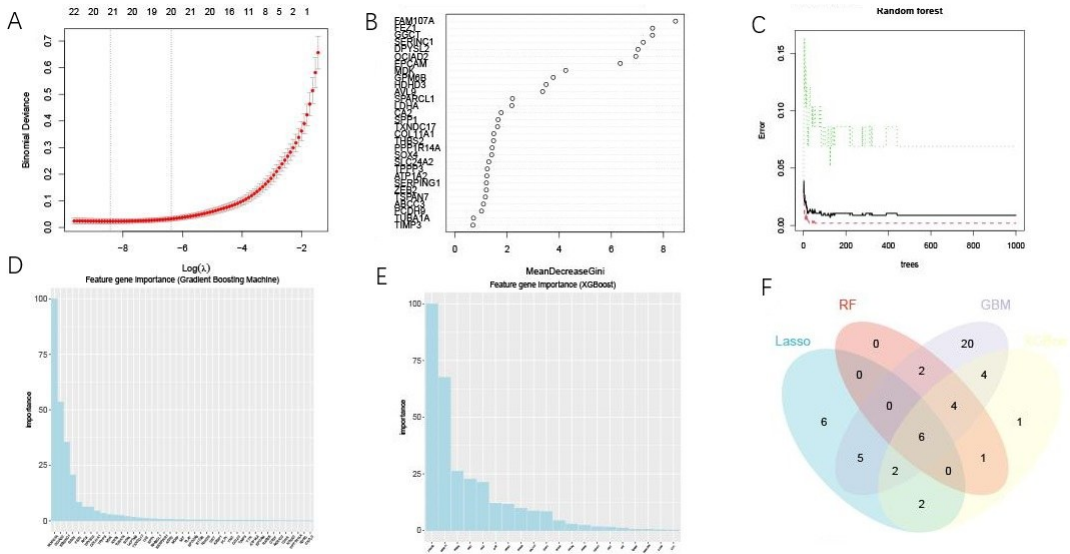


图 2 差异分析火山图



注: A. LASSO 回归; B、C. 随机森林; D. GBM; E. XGBoost; F. 交集韦恩图。

图 3 核心基因的筛选

2.4 诊断 ROC 曲线分析 为了探索 SERINC1 在肺癌诊断中的潜力,对其进行了 ROC 曲线分析。ROC 曲线的 AUC 反映了模型的预测能力,AUC 越接近 1.0,预测能力越强。结果显示,AUC 高达 0.972,说明 SERINC1 具有极高的诊断价值。这一发现突显了 SERINC1 在肺癌诊断中的重要作用,并为其在临床上的应用奠定了坚实的基础,见图 4。

2.5 SERINC1 表达量分析 TIMER2 数据库 SERINC1 泛癌表达分析显示,SERINC1 在含肺癌的 18 种癌症中具有差异表达(见图 5A)。在 598 个 TC-GA 样本中的表达量分析显示 SERINC1 在正常组织表达较肿瘤组织更高 ( $P < 0.01$ ),见图 5B、图 5C。

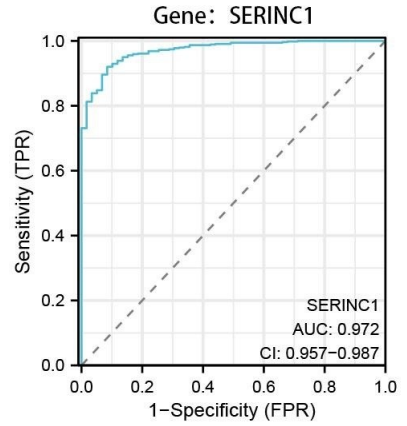
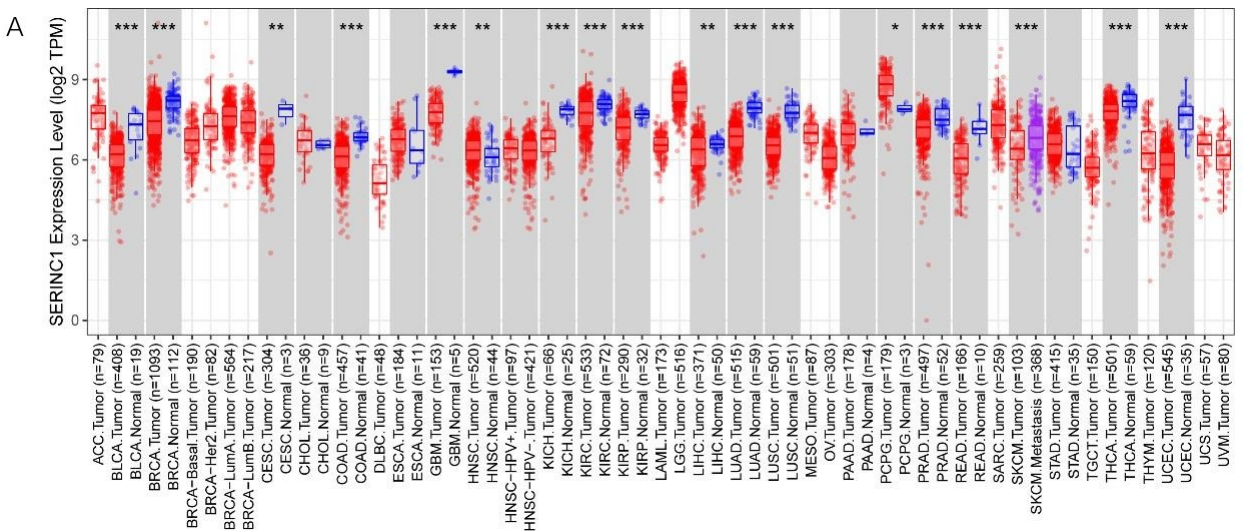
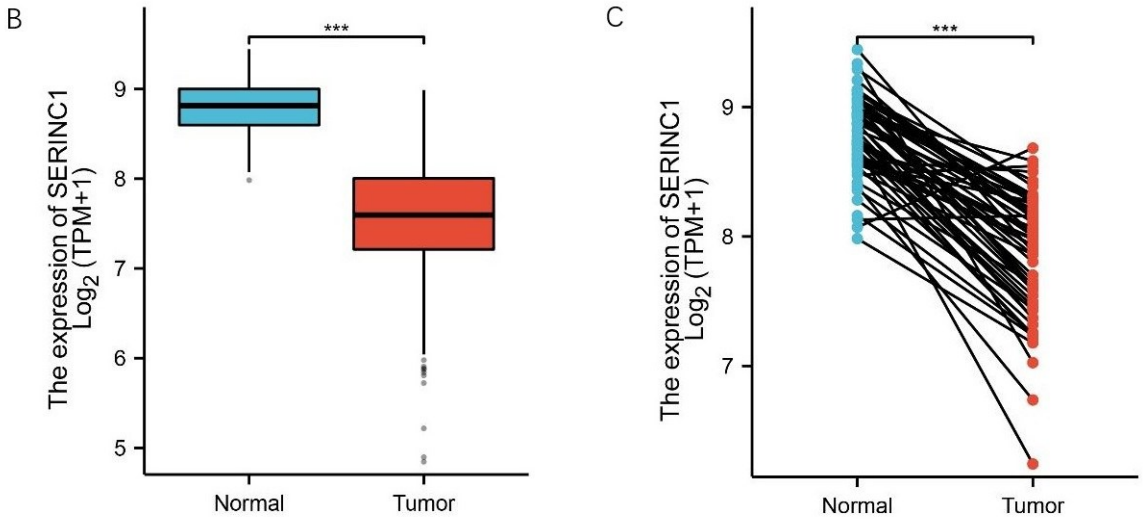


图 4 SERINC1 的 ROC 诊断曲线





注:A. SERINC1 在泛癌中的表达;B. SERINC1 非配对样本表达量分析;C. SERINC1 配对样本表达量分析。

图 5 SERINC1 表达量分析

2.6 对 TCGA 样本作单基因差异分析并进行富集分析 首先利用 DESeq2 包对肺癌样本和正常肺组织样本的 TCGA 基因表达矩阵数据进行了差异分析,筛选出了与 SERINC1 基因相关性较高的基因,并根据它们的表达水平将样本分为高表达组和低表达组。然后,对这两组样本的基因表达进行了统计检验,找出了显著差异表达的基因。随后将与 SERINC1 显著相关的基因纳入富集分析,富集分析的结果显示它们主要涉及 3 个方面的功能:离子通道、细胞外基质 (ECM) 和补体凝血系统。离子通道在细胞信号传导和细胞周期调控中起着关键作用。它们的异常活动可能导致肺癌细胞的增殖、迁移和侵袭。ECM 是由细胞分泌到细胞

外间质中的大分子物质构成的复杂网架动态结构。它不仅提供动态组织完整性,而且作为信号分子参与并驱动许多生物反应。ECM 的失调可能是肺癌等慢性疾病的直接或间接原因。ECM 还包含各种分泌蛋白,包括细胞因子、趋化因子和生长因子,这些分泌蛋白潜在参与免疫细胞调节。补体系统是一种体液免疫系统,通过一系列复杂的反应,可以引起溶菌作用、吞噬作用、趋化作用和天然免疫的杀菌作用等。凝血系统则是防止血管破损出血的重要过程。有研究表明,补体和凝血级联反应在肿瘤微环境中可能通过激活凝血级联反应来促进肿瘤细胞侵袭和转移,同时也可以通过激活补体系统来抑制免疫细胞的杀伤作用,见图 6。

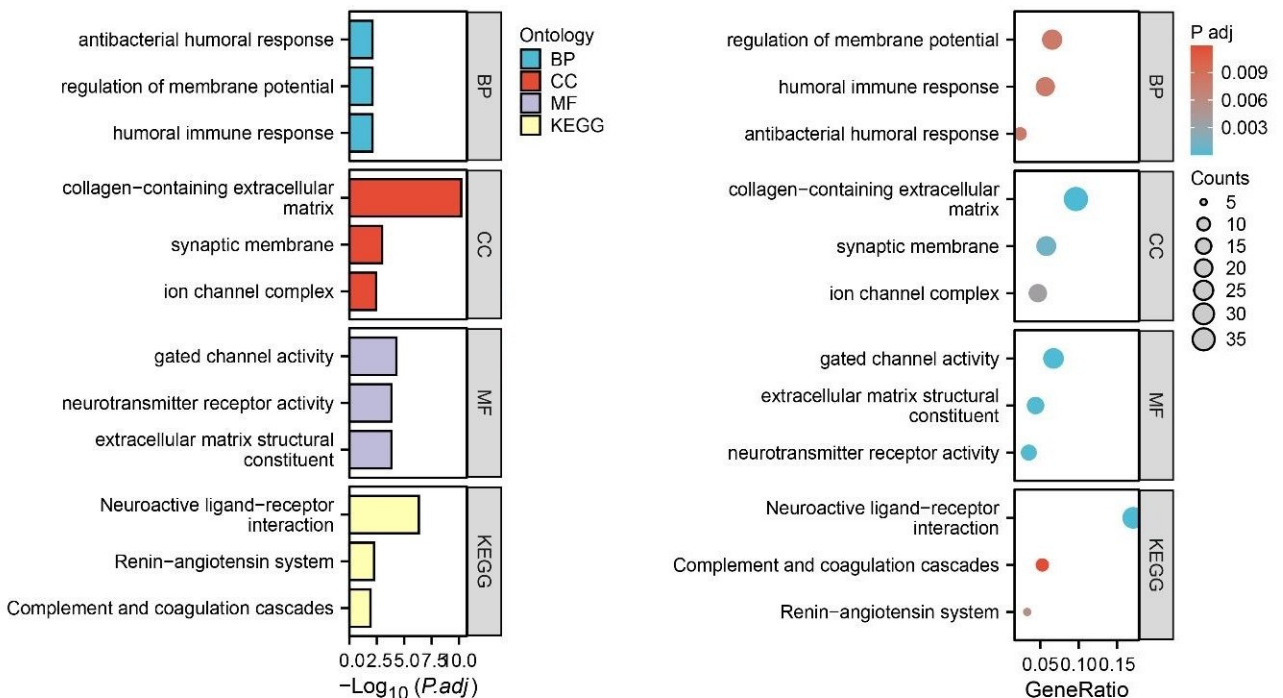
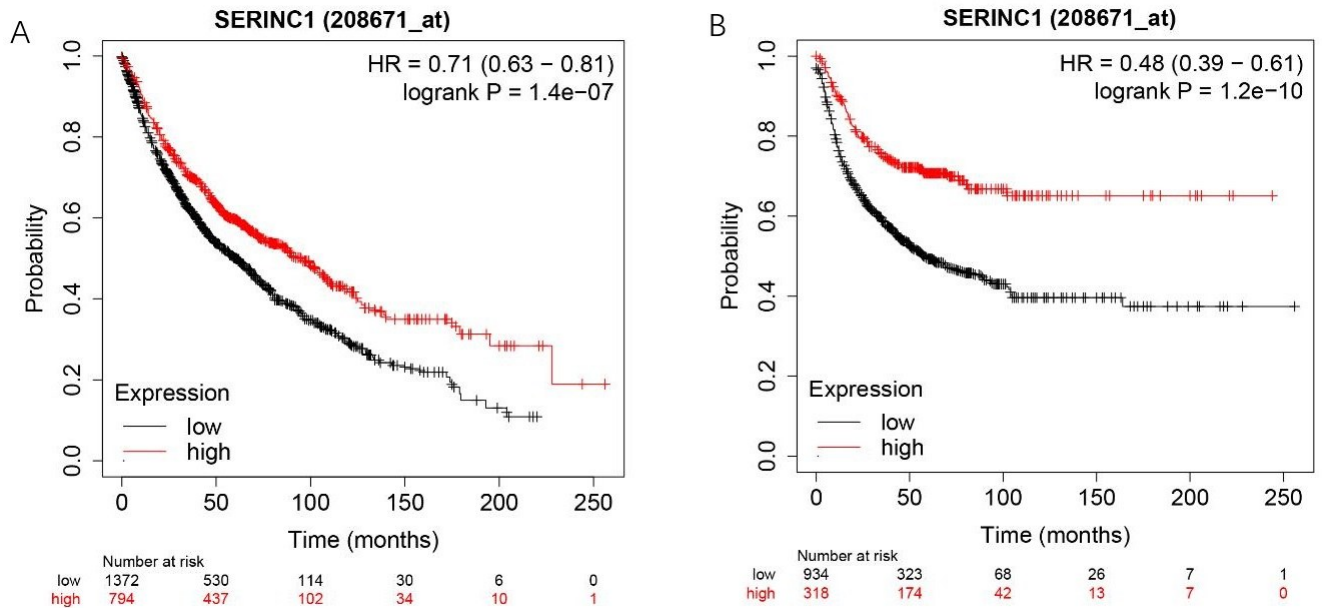


图 6 富集分析

2.7 生存分析 使用了在线工具 Kaplan-Meier Plotter 来分析 SERINC1 在肺癌中的生存曲线。选取了所有的肺癌样本,共计 2 852 个,其中 GSE102287 ( $n = 66$ )、GSE14814 ( $n = 90$ )、GSE157011 ( $n = 235$ )、GSE19188 ( $n = 156$ )、GSE29013 ( $n = 55$ )、GSE30219 ( $n = 307$ )、GSE31210 ( $n = 246$ )、GSE3141 ( $n = 111$ )、GSE31908 ( $n = 40$ )、GSE37745 ( $n = 196$ )、GSE43580 ( $n = 150$ )、GSE4573 ( $n = 130$ )、GSE50081 ( $n = 181$ )、GSE68465 ( $n = 462$ )、GSE77803 ( $n = 156$ )、GSE8894 ( $n = 138$ )、TCGA ( $n = 133$ )。分别绘

制了总生存期(OS)和无进展生存期(FP)的 Kaplan-Meier 曲线,在勾选 OS 选项时,样本数量显示为 2166,勾选 FP 时,样本数量显示为 1252。绘制结果显示,SERINC1 表达高的患者的预后明显优于表达量低的患者。单因素 COX 回归分析结果显示 SERINC1 的风险比(HR)值为 0.71 (95% CI : 0.63 ~ 0.81) 和 0.48 (95% CI : 0.39 ~ 0.61),这表明 SERINC1 是一个保护因素,即 SERINC1 的表达量越高,患者的预后越好。



注:A. OS 生存曲线;B. FP 生存曲线。

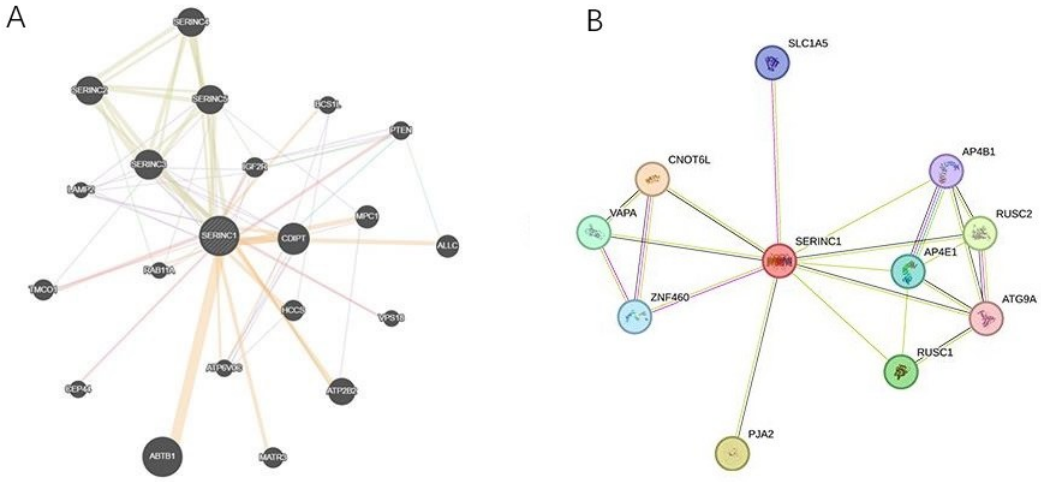
图 7 K-M 曲线

2.8 PPI 蛋白互作分析 GeneMania 数据库用于构建与 SERINC1 相互作用的基因-基因网络。结果显示,SERINC3、SERINC5 和 ABTB1 等 20 个基因与 SERINC1 的相关性较大(见图 8A)。使用 STRING 数据库获得了与智人 SERINC1 相互作用的蛋白质-蛋白质相互作用网络,与 SERINC1 相互作用的基因包括 SLC1A5、AP4E1 和 CNOT6L 等(见图 8B)。后续研究可以根据它们与 SERINC1 相互作用的具体机制进行进一步的研究。

2.9 免疫浸润分析 利用 CIBERSORT 对肺癌样本中的免疫细胞进行了分析(见图 9A)。发现 SERINC1 与多种免疫细胞的比例有显著的相关性( $P < 0.05$ ),如单核细胞、M2 型巨噬细胞、静息和活化的肥大细胞、静息和朴素的 CD4 T 细胞、活化的树突状细胞、中性粒细胞、滤泡辅助 T 细胞、调节 T 细胞、浆细胞和朴素

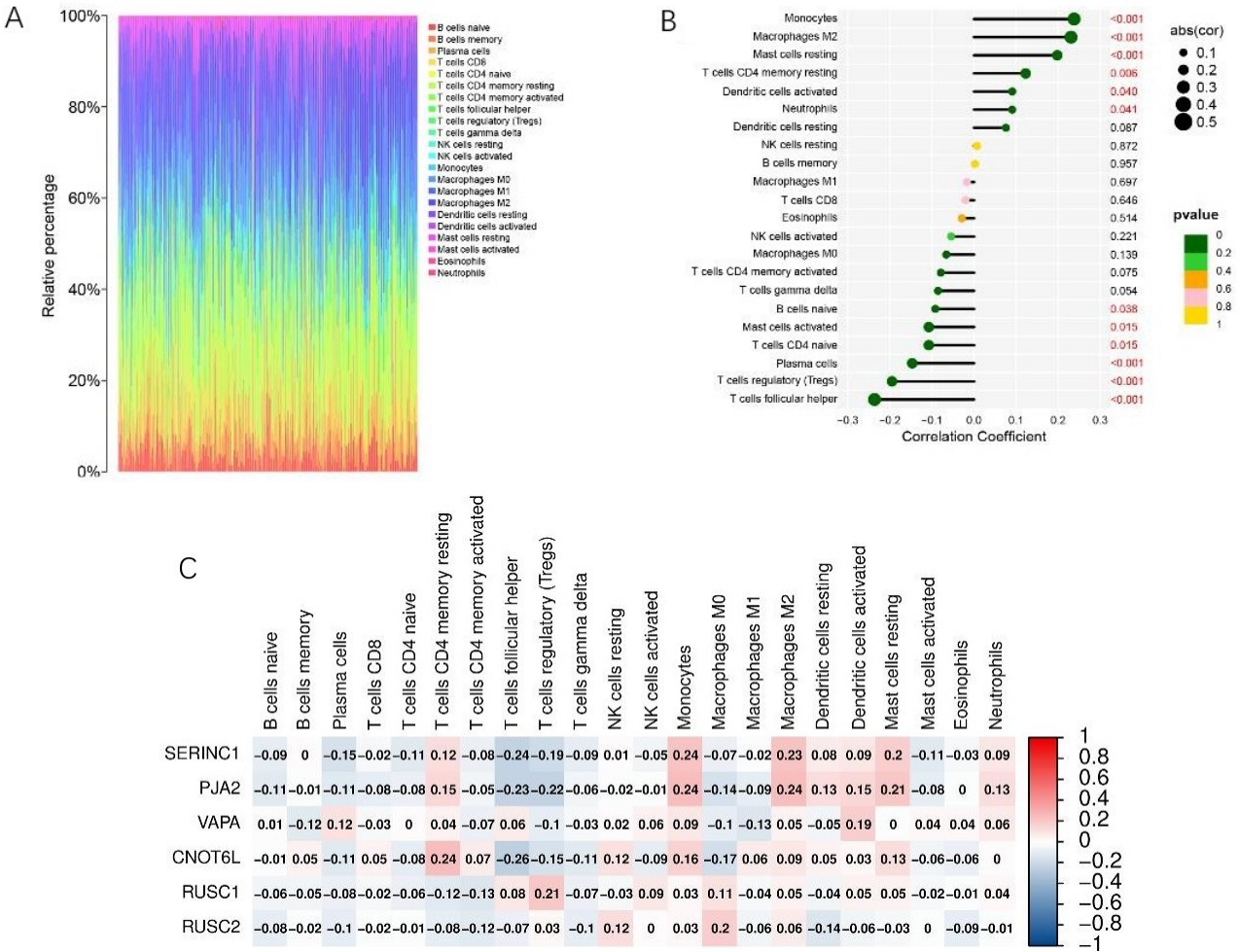
B 细胞等(见图 9B)。还绘制了 SERINC1 及其相关的 5 个基因(VAPA、CNOT6L、RUSC1、RUSC2 和 PJA2)与免疫细胞的相关性热图(见图 9C),结果显示这些基因与免疫细胞的比例也有显著的相关性。这些发现表明 SERINC1 及其相关基因在免疫反应中起着重要的作用,这对于理解肺癌的免疫微环境有重要的意义。

2.10 SERINC1 与多种 T 细胞标志基因的关系 使用了 GEPIA2 这一在线数据库,分析了 SERINC1 与多个 T 细胞免疫基因集的表达相关性。输入基因名称 SERINC1 后,在 Suggested Signatures 选择相应的 T 细胞,绘制 SERINC1 和 GEPIA2 显示的 T 细胞标志基因相关图。由此发现 SERINC1 与多个 T 细胞的标志基因有显著的相关性。这些发现表明 SERINC1 在肺癌免疫反应中可能扮演重要的角色,见表 1。



注:A. SERINC1 基因-基因相互作用网络;B. SERINC1 蛋白-蛋白互作网络。

图 8 蛋白互作分析



注:A. CIBERSORT 分析;B. SERINC1 与免疫细胞相关性;C. SERINC1 及其相关的 5 个基因与免疫细胞的相关性。

图 9 免疫浸润分析

2.11 SERINC1 表达与免疫细胞的相关性 在与 SERINC1 表达相关的 12 种免疫细胞中,有 6 种免疫细胞呈正相关,分别是中性粒细胞、激活态树突状细胞、静息态记忆 CD4 T 细胞、静息态肥大细胞、M2 型

巨噬细胞、单核细胞,有 6 种免疫细胞呈负相关,分别是滤泡辅助 T 细胞、调节 T 细胞(Tregs)、浆细胞、初级 CD4 T 细胞、激活态肥大细胞、初级 B 细胞,见图 10。

表 1 SERINC1 与多种 T 细胞标志基因的相关性

CELL	Makers	P	R
Naïve T cell	CCR7, LEF1, TCF7, SELL	<0.001	0.21
Effector T cell	CX3CR1, FGF2, FGF3, FGF4, FGF5, FGF6, FGF7, FGF8, FGF9, FGF10, FGF11, FGF12, FGF13, FGF14, FGF15, FGF16, FGF17, FGF18, FGF19, FGF20, FGF21, FGF22, FGF23, FGF24, FGF25, FGF26, FGF27, FGF28, FGF29, FGF30, FGF31, FGF32, FGF33, FGF34, FGF35, FGF36, FGF37, FGF38, FGF39, FGF40, FGF41, FGF42, FGF43, FGF44, FGF45, FGF46, FGF47, FGF48, FGF49, FGF50, FGF51, FGF52, FGF53, FGF54, FGF55, FGF56, FGF57, FGF58, FGF59, FGF60, FGF61, FGF62, FGF63, FGF64, FGF65, FGF66, FGF67, FGF68, FGF69, FGF70, FGF71, FGF72, FGF73, FGF74, FGF75, FGF76, FGF77, FGF78, FGF79, FGF80, FGF81, FGF82, FGF83, FGF84, FGF85, FGF86, FGF87, FGF88, FGF89, FGF90, FGF91, FGF92, FGF93, FGF94, FGF95, FGF96, FGF97, FGF98, FGF99, FGF100	<0.001	0.6
Effector memory	PDCD1, DUSP4, GZMK, GZMA, IFNG	0.170	-0.048
Central memory	CCR7, SELL, IL7R	<0.001	0.460
Resident memory	CD69, ITGAE, CXCR6, MYADM	<0.001	0.550
Exhausted T cell	HAVCR2, TIGIT, LAG3, PDCD1, CXCL13, LAYN	0.014	-0.085
Resting treg T cell	FOXP3, IL2RA	<0.001	-0.210
Effector treg T cell	FOXP3, CTLA4, CCR8, TNFRSF9	<0.001	-0.210
Th1-like	CXCL13, HAVCR2, IFNG, CXCR3, BHLHE40, CD4	0.610	-0.018

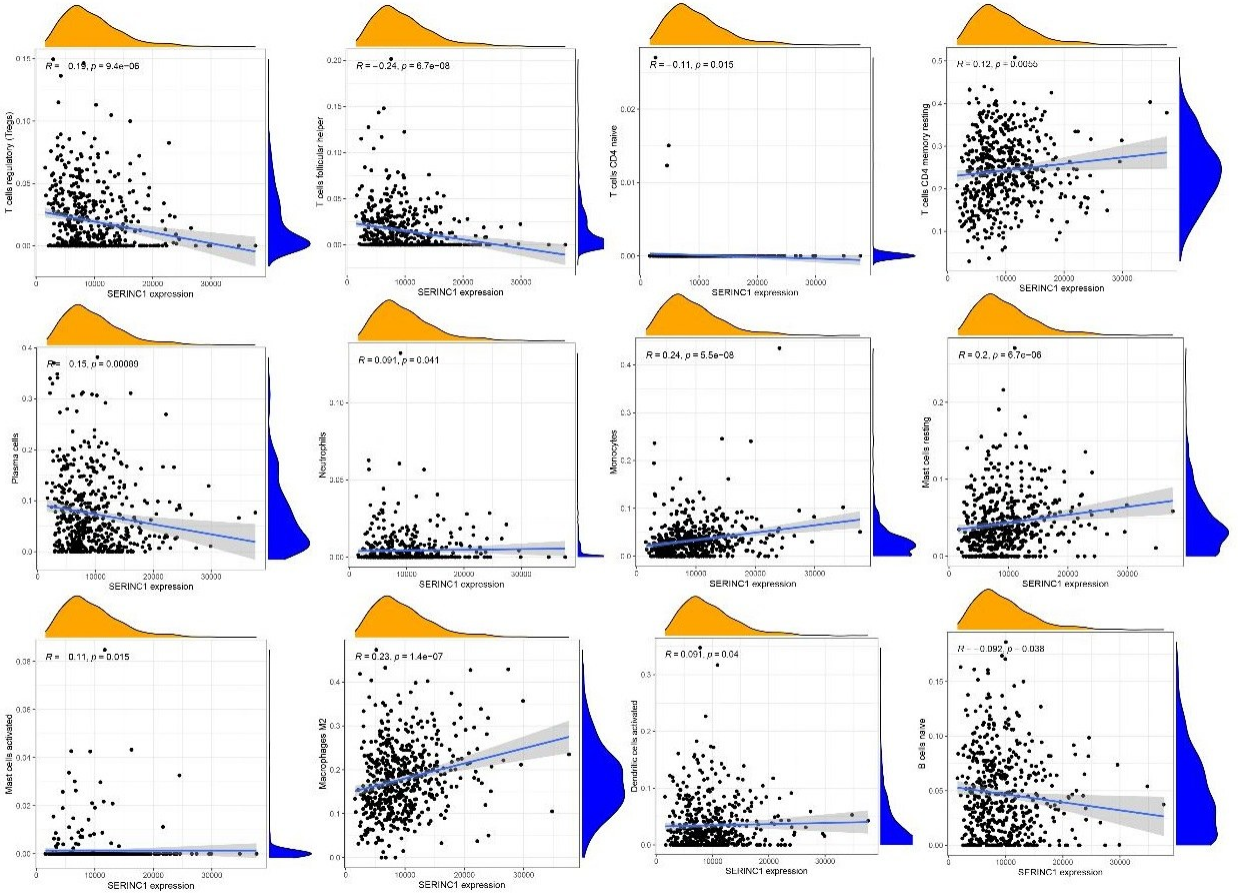
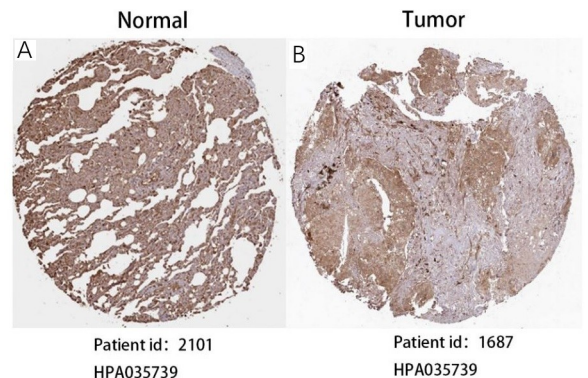


图 10 SERINC1 与免疫细胞的相关性

2.12 SERINC1 在正常组织和肿瘤组织中的蛋白表达水平 通过 HPA 数据库对 SERINC1 蛋白的表达水平进行了分析。如图 11 所示，正常样本中的 SERINC1 蛋白表达水平明显高于肿瘤样本的表达水平。这一结果是通过比较样本染色的深浅得出的，这进一步证实了 SERINC1 在正常细胞中的重要作用，而在肿瘤细胞中，其表达水平可能会降低。



注:A. SERINC1 在正常组织中的蛋白表达水平；  
B. SERINC1 在肺癌样本中的蛋白表达水平。

图 11 SERINC1 蛋白表达水平

## 3 讨论

有研究发现<sup>[29-31]</sup>,丝氨酸整合因子(SERINC)是一种膜蛋白,其家族共有 5 个成员,其中 SERINC5 和 SERINC3 是目前最新发现的抗艾滋病病毒的天然免疫分子,可阻断病毒感染新的细胞,从而降低病毒的感染力。已有研究表明<sup>[32]</sup>,SERINC1 是一种参与磷脂和鞘磷脂合成的蛋白,能够促进丝氨酸的嵌入。SERINC1 在肝癌细胞中高表达,可以通过阻止 p21 的转录激活调节细胞周期<sup>[33]</sup>。然而,其在肺癌中的作用还尚未明确。本研究的目的是探讨 SERINC1 在肺癌中的表达和功能,以及其与肺癌的免疫和预后的关系。

本研究发现 SERINC1 在肺癌组织中的表达水平显著下调,与正常组织相比,这种差异具有统计学意义。通过 Kaplan-Meier 曲线和单因素 COX 回归分析,进一步发现 SERINC1 的表达水平与患者的预后有关,且表达水平越高,患者的生存期越长。为了深入理解 SERINC1 的功能和作用机制,对 SERINC1 进行了功能富集分析,发现其主要涉及含胶原的细胞外基质、补体和凝血级联反应等生物过程。还利用 STRING 数据库构建了 SERINC1 的蛋白互作网络,发现其与多种肿瘤相关蛋白有显著的相互作用关系,如 CNOT6L、VAPA 等,如 VAPA 和 CNOT6L 的表达量下降则显著加速了肿瘤细胞的增殖<sup>[34]</sup>。此外,还研究了 SERINC1 与肺癌中的免疫细胞的关系。利用 CIBERSORT 分析了 SERINC1 与 22 种免疫细胞浸润量的相关性,发现 SERINC1 与 12 种免疫细胞有显著的相关性。这表明 SERINC1 可能是一种免疫刺激因子,能够促进肿瘤微环境中的抗肿瘤免疫反应<sup>[35]</sup>。还利用 GEPIA 数据库分析了 SERINC1 与多个 T 细胞标志基因的相关性,发现 SERINC1 与多个 T 细胞标志基因存在显著相关性<sup>[35]</sup>。这些结果进一步证实了 SERINC1 在肺癌免疫反应中的重要作用。最后,利用 HPA 数据库查看了 SERINC1 在正常和肿瘤样本中的蛋白表达水平,发现 SERINC1 在正常样本中的蛋白表达水平明显高于肿瘤样本。

综上所述,本研究表明,SERINC1 可能是一种具有潜力的肺癌预后标志物,且通过影响免疫微环境从而影响患者的预后。这些发现提供了新的视角来理解肺癌的发生、发展和预后,也为未来的肺癌治疗提供了新的靶点。

## 参考文献:

[1] LEITER A,VELUSWAMY R R,WISNIVESKY J P. The global burden of lung cancer: current status and future trends[J]. *Nat Rev Clin Oncol*,2023,20(9):624-639.  
[2] SHENOY A T,LYON DE ANA C,ARAFI E I, et al.

Antigen presentation by lung epithelial cells directs CD4<sup>+</sup> T<sub>RM</sub> cell function and regulates barrier immunity[J]. *Nat Commun*,2021,12(1):5834.  
[3] HAURA E B,CAMIDGE D R,RECKAMP K, et al. Molecular origins of lung cancer: prospects for personalized prevention and therapy[J]. *J Thorac Oncol*,2010,5(6 Suppl 3):S207-213.  
[4] MENJU T,DATE H. Lung cancer and epithelial-mesenchymal transition[J]. *Gen Thorac Cardiovasc Surg*,2021,69(5):781-789.  
[5] CLOUGH E,BARRETT T. The gene expression omnibus database[J]. *Methods Mol Biol*,2016,1418:93-110.  
[6] SLOVIN S,CARISSIMO A,PANARIELLO F, et al. Single-cell RNA sequencing analysis: a step-by-step overview[J]. *Methods Mol Biol*,2021,2284:343-365.  
[7] ABDELHAFEZ O H,OTHMAN E M,FAHIM J R, et al. Metabolomics analysis and biological investigation of three Malvaceae plants[J]. *Phytochem Anal*,2020,31(2):204-214.  
[8] KOBAK D,BERENS P. The art of using t-SNE for single-cell transcriptomics[J]. *Nat Commun*,2019,10(1):5416.  
[9] DI Z Y,ZHOU S C,XU G R, et al. Single-cell and WGCNA uncover a prognostic model and potential oncogenes in colorectal cancer[J]. *Biol Proced Online*,2022,24(1):13.  
[10] LIU C L,PU M,MA Y B, et al. Intra-tumor heterogeneity and prognostic risk signature for hepatocellular carcinoma based on single-cell analysis[J]. *Exp Biol Med (Maywood)*,2022,247(19):1741-1751.  
[11] KANG X,ZHANG K L,WANG Y K, et al. Single-cell RNA sequencing analysis of human chondrocytes reveals cell-cell communication alterations mediated by interactive signaling pathways in osteoarthritis[J]. *Front Cell Dev Biol*,2023,11:1099287.  
[12] SLOVIN S,CARISSIMO A,PANARIELLO F, et al. Single-cell RNA sequencing analysis: a step-by-step overview[J]. *Methods Mol Biol*,2021,2284:343-365.  
[13] CAO J Y,SPIELMANN M,QIU X J, et al. The single-cell transcriptional landscape of mammalian organogenesis[J]. *Nature*,2019,566(7745):496-502.  
[14] LOVE M I,HUBER W,ANDERS S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2[J]. *Genome Biol*,2014,15(12):550.  
[15] KANG J,CHOI Y J,KIM I K, et al. LASSO-based machine learning algorithm for prediction of lymph node metastasis in T1 colorectal cancer[J]. *Cancer Res Treat*,2021,53(3):773-783.  
[16] HUDSON J I,HUDSON Y,KANYAMA G, et al. Causal factors in childhood and adolescence leading to anabolic-androgenic steroid use: a machine learning approach[J]. *Drug Alcohol Depend Rep*,2023,10:100215.

- [17] LIN S W, LU W B, WANG T, et al. Predictive model of acute kidney injury in critically ill patients with acute pancreatitis; a machine learning approach using the MIMIC-IV database[J]. *Ren Fail*, 2024, 46(1):2303395.
- [18] HOU N Z, LI M Z, HE L, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3; a machine learning approach using XGboost [J]. *J Transl Med*, 2020, 18(1):462.
- [19] CHAI J L, LU B W, DU H T, et al. Pyroptosis-related potential diagnostic biomarkers in steroid-induced osteonecrosis of the femoral head [J]. *BMC Musculoskelet Disord*, 2023, 24(1):609.
- [20] BIAN J, XIONG W J, YANG Z, et al. Identification and prognostic biomarkers among ZDHHC4/12/18/24, and APT2 in lung adenocarcinoma [J]. *Sci Rep*, 2024, 14(1):522.
- [21] DING G Q, ZHANG T, YU H, et al. Prognostic value of CXCR6 and the correlation with immune infiltration in the microenvironment of skin cutaneous melanoma [J]. *Biochem Genet*, 2024. doi:10.1007/S10528-023-10647-8.
- [22] YI J Q, LUO X H, HUANG W J, et al. PGK1 is a potential biomarker for early diagnosis and prognosis of hepatocellular carcinoma [J]. *Oncol Lett*, 2024, 27(3):109.
- [23] XIANG N, CHEN T, ZHAO X L, et al. In vitro assessment of roles of PPP1R14B in cervical and endometrial cancer [J]. *Tissue Cell*, 2022, 77:101845.
- [24] SHI Y Y, CHEN Z Q, HUANG L X, et al. A network pharmacology approach to reveal the key ingredients in *Scrophulariae Radix* (SR) and their effects against Alzheimer's disease [J]. *Heliyon*, 2024, 10(3):e24785.
- [25] QIN L M, LIN H B, ZHANG G F, et al. Screening and validation of differentially expressed genes in polymyositis [J]. *Heliyon*, 2024, 10(3):e24537.
- [26] MA N, ZHOU T, LI C Y, et al. A pan-cancer analysis of the prognosis and immune infiltration of eEF1A2 and its potential function in thyroid carcinoma [J]. *Heliyon*, 2024, 10(3):e24455.
- [27] BAUER A K, SIEGRIST K J, WOLFF M, et al. The carcinogenic properties of overlooked yet prevalent polycyclic aromatic hydrocarbons in human lung epithelial cells [J]. *Toxics*, 2022, 10(1):28.
- [28] KUMAZAKI M, SHIMOMURA I, KIYONO T, et al. Cell-type specific tumorigenesis with Ras oncogenes in human lung epithelial cells [J]. *Biochem Biophys Res Commun*, 2020, 525(2):483-490.
- [29] LEONHARDT S A, PURDY M D, GROVER J R, et al. Antiviral HIV-1 SERINC restriction factors disrupt virus membrane asymmetry [J]. *Nat Commun*, 2023, 14(1):4368.
- [30] RAMDAS P, BHARDWAJ V, SINGH A, et al. Coelacanth SERINC2 inhibits HIV-1 infectivity and is counteracted by envelope glycoprotein from foamy virus [J]. *J Virol*, 2021, 95(13):e0022921.
- [31] JIN S W, MWIMANZI F M, MANN J K, et al. Variation in HIV-1 Nef function within and among viral subtypes reveals genetically separable antagonism of SERINC3 and SERINC5 [J]. *PLoS Pathog*, 2020, 16(9):e1008813.
- [32] CHU E P, ELSON C M, POLLOCK A H, et al. Disruption of Serinc1, which facilitates serine-derived lipid synthesis, fails to alter macrophage function, lymphocyte proliferation or autoimmune disease susceptibility [J]. *Mol Immunol*, 2017, 82:19-33.
- [33] REN W H, YANG C Y, YANG X M, et al. siRNA-mediated knockdown of hTDE2 retards cell cycle progression through transcriptional activation of p21 [J]. *Oncol Rep*, 2014, 31(3):1314-1322.
- [34] WANG H Y, RADOMSKA H S, PHELPS M et al. Replication study: coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs [J]. *Elife*, 2020, 9:e56651.
- [35] WANG J H, ZHAO Y J, LIAO P J, et al. Immune checkpoint expression patterns on T cell subsets in light-chain amyloidosis; VISTA, PD-1, and TIGIT as potential therapeutic targets [J]. *Blood Sci*, 2024, 6(1):e00181.

收稿日期:2024-02-28;修回日期:2024-03-09