

[文章编号] 1671-587X(2024)02-0379-13

DOI:10.13481/j.1671-587X.20240211

基于剪接因子 SRSF9 在头颈部鳞状细胞癌中的表达及临床意义的生物信息学分析

刘玉婷¹, 喻莹¹, 黎桂珍¹, 史沁雪², 李彬彬¹

(1. 广东医科大学基础医学院病理生理学教研室, 广东 东莞 523808; 2. 广东医科大学第二临床医学院, 广东 东莞 523808)

[摘要] **目的:** 基于生物信息学方法分析富含丝氨酸/精氨酸的剪接因子9 (SRSF9) 在头颈部鳞状细胞癌 (HNSCC) 中的表达情况、临床意义以及与肿瘤免疫浸润的关系, 并探讨其作用机制。**方法:** 利用癌症基因组图谱 (TCGA) 数据库、基因表达综合 (GEO) 数据库中 GSE30784 和 GSE13601 数据集、临床蛋白质组学肿瘤分析联盟 (CPTAC) 数据库、基因表达交互分析 (GEPIA) 数据库和 Kaplan-Meier plotter 数据库分析 SRSF9 在 HNSCC 中的表达及与患者临床病理特征和预后的关系; 利用 cBioPortal 数据库分析 SRSF9 基因变异情况; 应用 ESTIMATE 算法和肿瘤免疫估计资源 (TIMER) 数据库评估 SRSF9 表达与肿瘤微环境和肿瘤免疫细胞浸润的关系; 利用 LinkedOmics 数据库分析 SRSF9 在 HNSCC 中的共表达基因及其调控通路; 基于 TCGA SpliceSeq 数据库分析 HNSCC 中 SRSF9 调控的可变剪接事件, 并对靶基因进行基因本体 (GO) 功能富集分析和京都基因与基因组百科全书 (KEGG) 信号通路富集分析。**结果:** TCGA 数据库、GEO 数据库中 GSE30784 和 GSE13601 数据集分析, HNSCC 组织中 SRSF9 mRNA 表达水平较癌旁正常组织明显升高 ($P < 0.01$); CPTAC 数据库分析, 与正常组织比较, HNSCC 组织中 SRSF9 蛋白表达水平明显升高 ($P < 0.001$)。TCGA 数据库分析, SRSF9 mRNA 表达与 HNSCC 患者病理分级 ($P = 0.004$) 和 HPV 感染 ($P = 0.031$) 有关联。GEPIA 和 Kaplan-Meier plotter 数据库分析, HNSCC 组织中 SRSF9 mRNA 高表达均与患者总生存期 (OS) 较差相关 [风险比 (HR) = 1.40, $P = 0.019$; HR = 1.55, $P = 0.003$]。cBioPortal 数据库分析, HNSCC 中 SRSF9 基因拷贝数变异 (CNV) 占 26.85%, 其 CNV 与 SRSF9 mRNA 表达水平呈正相关关系 ($r = 0.44$, $P < 0.001$)。ESTIMATE 算法分析, SRSF9 mRNA 高表达组基质评分和免疫评分均较 SRSF9 mRNA 低表达组低 ($P < 0.001$), 肿瘤纯度则高于 SRSF9 mRNA 低表达组 ($P < 0.001$)。TIMER 数据库分析, SRSF9 mRNA 表达水平与 CD4+T 淋巴细胞浸润呈正相关关系 ($r = 0.186$, $P < 0.001$), 而与 B 淋巴细胞、CD8+T 淋巴细胞和树突状细胞浸润呈负相关关系 ($r = -0.269$, $P < 0.001$; $r = -0.353$, $P < 0.001$; $r = -0.304$, $P < 0.001$)。SRSF9 共表达基因富集分析, 核糖体、剪接体和代谢通路等相关基因上调, 而黏着斑、细胞因子和细胞黏附分子等相关基因下调。SRSF9 相关可变剪接靶基因主要涉及脂类代谢、胰高血糖素和紧密连接等通路。**结论:** SRSF9 在 HNSCC 中呈高表达, 且与患者预后不良和肿瘤微环境免疫细胞浸润相关, 可成为 HNSCC 诊断、预后评估和治疗的潜在分子靶标。

[关键词] 头颈部; 鳞状细胞癌; 富含丝氨酸/精氨酸剪接因子9; 可变剪接; 肿瘤免疫; 生物信息学

[中图分类号] R739.91 **[文献标志码]** A

[收稿日期] 2023-04-21

[基金项目] 国家自然科学基金项目 (81502411); 广东省科技厅自然科学基金项目 (2023A1515010448); 广东省东莞市科技局社会发展科技重点项目 (20231800936042); 广东医科大学大学生创新实验项目 (ZYDS004)

[作者简介] 刘玉婷 (1997-), 女, 浙江省丽水市人, 在读硕士研究生, 主要从事肿瘤发生机制方面的研究。

[通信作者] 李彬彬, 教授, 硕士研究生导师 (E-mail: libinbin1022@126.com)

Bioinformatics analysis based on expression of splicing factor SRSF9 in head and neck squamous cell carcinoma and clinical significance

LIU Yuting¹, YU Ying¹, LI Guizhen¹, SHI Qinxue², LI Binbin¹

(1. Department of Pathophysiology, School of Basic Medical Sciences, Guangdong Medical University, Dongguan 523808, China; 2. School of Second Clinical Medical Sciences, Guangdong Medical University, Dongguan 523808, China)

ABSTRACT Objective: To analyze the expression, clinical significance, and relationship with tumor immune infiltration of serine/arginine-rich splicing factor 9 (SRSF9) in head and neck squamous cell carcinoma (HNSCC) by bioinformatics methods, and to discuss its mechanism. **Methods:** The expression of SRSF9 in HNSCC and its relationship with clinical pathologic characteristics and prognosis of the patients were analyzed by The Cancer Genome Atlas (TCGA) Database, GSE30784 and GSE13601 datasets in Gene Expression Omnibus (GEO) Database, Clinical Proteomic Tumor Analysis Consortium (CPTAC) Database, Gene Expression Profiling Interactive Analysis (GEPIA) Database, and Kaplan-Meier plotter Database; the variations of SRSF9 gene were examined through cBioPortal Database; ESTIMATE algorithm and The Tumor Immune Estimation Resource (TIMER) Database were used to assess the correlation between SRSF9 expression and tumor microenvironment, as well as tumor immune cell infiltration; LinkedOmics Database was used to analyze the SRSF9 co-expressed genes and their regulatory pathways in the HNSCC; the TCGA SpliceSeq Database was used to analyze the variable splicing events regulated by SRSF9 in the HNSCC; Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway enrichment analysis were conducted on the target genes. **Results:** The analysis results from TCGA Database and GSE30784 and GSE13601 datasets in GEO Database showed that the expression level of SRSF9 mRNA in the HNSCC tissue was significantly higher than that in adjacent normal tissue ($P < 0.01$). The CPTAC Database analysis results showed that compared with normal tissue, the expression level of SRSF9 protein in the HNSCC tissue was significantly increased ($P < 0.001$). The SRSF9 mRNA expression was associated with pathological grading ($P = 0.004$) and HPV infection ($P = 0.031$) in the patients with HNSCC according to TCGA Database analysis. The GEPIA and Kaplan-Meier plotter Database analysis results showed that high expression of SRSF9 mRNA in the HNSCC tissue was correlated with poorer overall survival (OS) of the patients [hazard ratio (HR) = 1.40, $P = 0.019$; HR = 1.55, $P = 0.003$]. The cBioPortal Database analysis results showed that copy number variation (CNV) of SRSF9 gene occurred in 26.85% in HNSCC, and CNV was positively correlated with SRSF9 mRNA expression levels ($r = 0.44$, $P < 0.001$). The ESTIMATE algorithm analysis results showed that high expression of SRSF9 mRNA group had lower stromal and immune score, and higher tumor purity than those in low expression of SRSF9 mRNA group ($P < 0.001$). The TIMER Database analysis results showed there was a positive correlation between the expression of SRSF9 mRNA and CD4⁺ T lymphocyte infiltration ($r = 0.186$, $P < 0.001$), and there were negative correlations between the expression level SRSF9 mRNA and B lymphocyte, CD8⁺ T lymphocyte, and dendritic cell infiltrations ($r = -0.269$, $P < 0.001$; $r = -0.353$, $P < 0.001$; $r = -0.304$, $P < 0.001$). The co-expressed gene enrichment analysis results showed upregulation of genes related to ribosomes, spliceosomes, and metabolic pathways, and downregulation of genes associated with focal adhesions, cytokines, and cell adhesion molecules. The main pathways involved in SRSF9-related

variable splicing target genes were lipid metabolism, glucagon, and tight junctions. **Conclusion:** SRSF9 is highly expressed in HNSCC and is associated with poor prognosis and tumor microenvironment immune cell infiltration, suggesting its potential as a molecular target for the diagnosis, prognosis assessment, and treatment of HNSCC.

KEYWORDS Head and neck; Squamous cell carcinoma; Serine/arginine-rich splicing factor 9; Alternative splicing; Tumor immunity; Bioinformatics

头颈部鳞状细胞癌 (head and neck squamous cell carcinoma, HNSCC) 是指原发于口腔、咽部 (鼻咽、口咽和下咽) 及喉部的鳞状上皮恶性肿瘤, 占头颈部肿瘤的 90% 以上。近年来, HNSCC 在世界范围内的发病率为 5%~6%, 已成为全球第八大常见恶性肿瘤^[1], 严重危害人类的健康。HNSCC 起病隐匿, 不易早期诊断, 多数患者首诊时已是晚期, 常伴有侵袭或转移, 其 5 年生存率低于 50%。近年来, 联合手术切除和放化疗等治疗手段在一定程度上提高了 HNSCC 患者的生存率, 但由于患者术后局部复发、远处转移和化疗耐药等, 整体预后仍然较差。因此, 开发 HNSCC 新的诊断和预后标志物及治疗靶点, 对指导个体化治疗并改善预后具有重要意义。RNA 可变剪接 (alternative splicing, AS) 是基因表达调控和产生蛋白质多样性的重要生物学过程。AS 异常与人类多种疾病包括肿瘤的发生密切相关, 与正常组织比较, 癌组织中普遍存在差异 AS 事件, 有望成为肿瘤的新型生物标志物和治疗靶点^[2-3]。富含丝氨酸/精氨酸的剪接因子 (serine/arginine-rich splicing factors, SRSFs) 家族是真核生物中一类进化十分保守的剪接调控因子, 包括 12 个家族成员 (SRSF1~12), 在 RNA 剪接体的组装和 AS 调控中发挥重要作用。在人类多种肿瘤中均发现 SRSFs 蛋白的异常表达, 通过调控肿瘤相关基因的异常剪接, 进而促进肿瘤的发生发展。SRSF9 作为该家族的成员之一, 在结直肠癌、神经胶质母细胞瘤、肺癌和肝癌等肿瘤组织中呈高表达, 并可促进肿瘤细胞的增殖、迁移及侵袭^[4-6]。但目前有关 SRSF9 在 HNSCC 中的表达及其生物学功能尚未见相关报道。本研究基于多个数据库的信息挖掘, 分析 SRSF9 在 HNSCC 中的表达情况、临床意义以及与肿瘤微环境 (tumor microenvironment, TME) 的关系, 并探讨其相关作用机制, 以期 HNSCC 早期诊断、预后评估以及治疗策略提供新的靶点。

1 资料与方法

1.1 数据的获取 从癌症基因图谱 (The Cancer Genome Atlas, TCGA) 数据库 (<https://portal.gdc.cancer.gov/>) 中下载 HNSCC 转录组数据及相应的临床信息, 并采用 TCGA biolinks R 包 (2.26.0) 进行数据整理; 从美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) (www.ncbi.nlm.nih.gov/geo) 的基因表达综合 (Gene Expression Omnibus, GEO) 数据库中下载 HNSCC 基因表达谱 GSE30784 和 GSE13601 数据集, 采用 Mann-Whitney *U* 检验分析在非配对肿瘤组织与正常组织中 SRSF9 mRNA 表达水平的差异; 采用配对 *t* 检验或 Wilcoxon 符号秩检验分析配对肿瘤组织与正常组织中的 SRSF9 mRNA 表达水平的差异, 以 $P < 0.05$ 为差异有统计学意义。采用 R 语言 pROC 包 (1.18.2) 绘制受试者工作特征 (receiver operating characteristic, ROC) 曲线分析指标的诊断性能, 计算 ROC 曲线下面积 (area under ROC curve, AUC)。利用 UALCAN 平台分析临床蛋白质组学肿瘤分析联盟 (Clinical Proteomic Tumor Analysis Consortium, CPTAC) 数据库中 SRSF9 蛋白在 HNSCC 肿瘤组织与正常组织中的表达水平差异。

1.2 SRSF9 mRNA 表达水平与患者临床病理特征和预后的关系 分析 TCGA-HNSCC 中 SRSF9 mRNA 表达水平与患者临床病理特征 [年龄、性别、肿瘤大小、淋巴结转移、临床分期、病理分级和人乳头状瘤病毒 (human papilloma virus, HPV) 感染] 的关系。2 组间 SRSF9 mRNA 表达水平比较采用 Mann-Whitney *U* 检验, 多组间 SRSF9 mRNA 表达水平比较采用 Kruskal-Wallis *H* 检验, 组间两两比较采用 Dunn's 检验。以 $P < 0.05$ 为差异有统计学意义。

利用基因表达谱交互式分析 (Gene Expression Profiling Interactive Analysis, GEPIA) 数据库 (<http://gepia.cancer-pku.cn/>) 和 Kaplan-Meier plotter 数据库 (<https://kmplot.com/>)

analysis/) 分析 SRSF9 mRNA 表达水平与 HNSCC 患者预后的相关性。根据 SRSF9 mRNA 表达水平将患者分为低表达组和高表达组, 利用 Kaplan-Meier 生存曲线分析患者总生存期 (overall survival, OS), 计算具有 95% 置信区间 (95% confidence interval, CI) 的风险比 (hazard ratio, HR), 使用 Log-rank 检验分析 2 组患者生存曲线的差异, 以 $P < 0.05$ 为差异有统计学意义。

1.3 HNSCC 中 SRSF9 基因拷贝数 利用 cBioPortal (<https://www.cbioportal.org/>) 数据库下载 HNSCC (TCGA, PanCancer Atlas) 数据集中 SRSF9 基因变异类型和频数, 采用两独立样本 t 检验分析 2 组间 SRSF9 mRNA 表达水平的差异, 以 $P < 0.05$ 为差异有统计学意义。利用 GSCA 数据库 (<http://bioinfo.life.hust.edu.cn/GSCA/#/>) 分析 SRSF9 基因拷贝数变异 (copy number variation, CNV) 对其 mRNA 表达水平的影响, 采用 Spearman 相关性分析并计算相关系数 (r) 和错误发现率 (false discovery rate, FDR)。采用 χ^2 检验或 Fisher 确切概率法分析 SRSF9 基因 CNV 在不同患者临床分期和病理分级间的差异, 以 $P < 0.05$ 为差异有统计学意义。

1.4 SRSF9 mRNA 表达与 TME 组成和免疫细胞浸润的关系 根据 TCGA-HNSCC 样本的表达数据, 使用 ESTIMATE 算法分析 TME 的组成, 获得基质分数、免疫分数和估计分数, 并采用 R 语言 pheatmap 包 (1.0.12) 绘制热图。根据 SRSF9 mRNA 表达水平的中位数将 HNSCC 患者样本分为低表达组和高表达组, 采用 Mann-Whitney U 检验分析 SRSF9 mRNA 表达水平与 TME 组成的关系, 以 $P < 0.05$ 为差异有统计学意义。使用 TIMER 数据库 (<http://timer.cistrome.org/>) 中 “Immune Association” 模块分析 SRSF9 mRNA 表达水平与肿瘤免疫细胞浸润的关系, 包括 B 细胞、CD8+ T 淋巴细胞、CD4+ T 淋巴细胞、巨噬细胞、中性粒细胞和树突状细胞, 并分析其对患者 OS 的影响, 以 $P < 0.05$ 为差异有统计学意义。

1.5 HNSCC 组织中 SRSF9 共表达基因的筛选和富集分析 利用 LinkedOmics 数据库 (<http://www.linkedomics.org>) 中的 “LinkFinder” 模块分析 TCGA-HNSCC 中 SRSF9 共表达基因。基于 Pearson 相关性分析, 以 $FDR < 0.01$ 筛选与 SRSF9 表达相关的上调基因和下调基因, 采用 R 语言

ggplot2 包 (3.4.2) 绘制火山图。利用 “LinkInterpreter” 模块中的基因富集分析 (Gene Set Enrichment Analysis, GSEA) 工具对差异相关基因进行京都基因和基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 通路富集分析, 设置随机组合次数为 1 000 次, 并采用柱状图表示最显著相关的前 10 个通路。

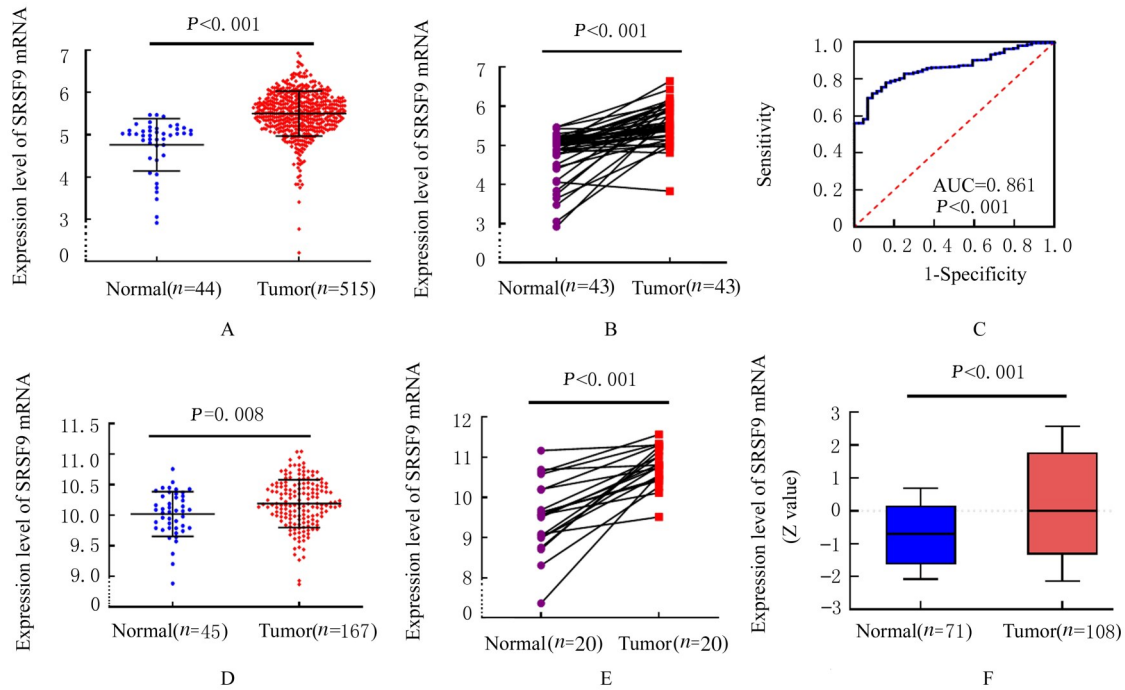
1.6 HNSCC 组织中 SRSF9 相关 AS 事件分析

从 TCGA SpliceSeq 数据库 (<http://bioinformatics.mdanderson.org/TCGASpliceSeq/>) 下载 HNSCC 的 AS 事件及相应的剪接百分比 (percent-spliced in, PSI) 矩阵, 将具有 PSI 值的样本占比 $\geq 75\%$ 的 AS 事件纳入本研究中。根据 TCGA RNA-seq 数据中 SRSF9 mRNA 表达水平, 选取 60 例 HNSCC 样本分为 mRNA 低表达组和 mRNA 高表达组, 分别计算可变剪接 PSI 值, 以 $|\Delta PSI| \geq 0.1$ 筛选差异可变剪接事件 (differential alternative splicing events, DASE)。利用 DAVID 数据库 (<https://david.ncicrf.gov/>) 对 SRSF9 相关 AS 基因进行基因本体 (Gene Ontology, GO) 功能富集分析和 KEGG 通路富集分析, 以 $P < 0.05$ 为差异有统计学意义。对显著相关的前 10 个过程或通路利用 R 语言 ggplot2 包 (3.4.2) 绘制气泡图。

2 结果

2.1 HNSCC 组织和癌旁正常组织中 SRSF9 mRNA 及蛋白的表达情况 利用 TCGA 数据库获得 515 例 HNSCC 样本和 44 例癌旁正常组织的 RNA-seq 数据, 分析结果显示: 与正常组织比较, 在 HNSCC 组织中 SRSF9 mRNA 表达水平明显升高 ($P < 0.001$), 见图 1A; 在 HNSCC 肿瘤组织中 SRSF9 mRNA 表达水平也明显高于癌旁正常组织 ($P < 0.001$), 见图 1B。ROC 曲线分析结果显示: AUC 为 0.861 ($P < 0.001$, 图 1C), 提示 SRSF9 对 HNSCC 具有较好的辅助诊断价值。GEO 数据库中 GSE30784 和 GSE13061 数据集分析结果显示: 与非配对或配对正常组织比较, 在 HNSCC 肿瘤组织中 SRSF9 mRNA 表达水平均明显升高 ($P < 0.01$), 见图 1D 和 1E。CPTAC 数据库分析结果显示: HNSCC 组织中 SRSF9 蛋白表达水平明显高于癌旁正常组织 ($P < 0.001$, 图 1F), 与 mRNA 水平的改变趋势一致。

2.2 SRSF9 mRNA 表达水平与患者临床病理特征和预后的相关性 分析 TCGA-HNSCC 中 515 例患



A—B: TCGA Database; C: ROC curve; D: GSE30784; E: GSE13601; F: CPTAC Database.

图1 HNSCC和癌旁正常组织中SRSF9 mRNA和蛋白的表达

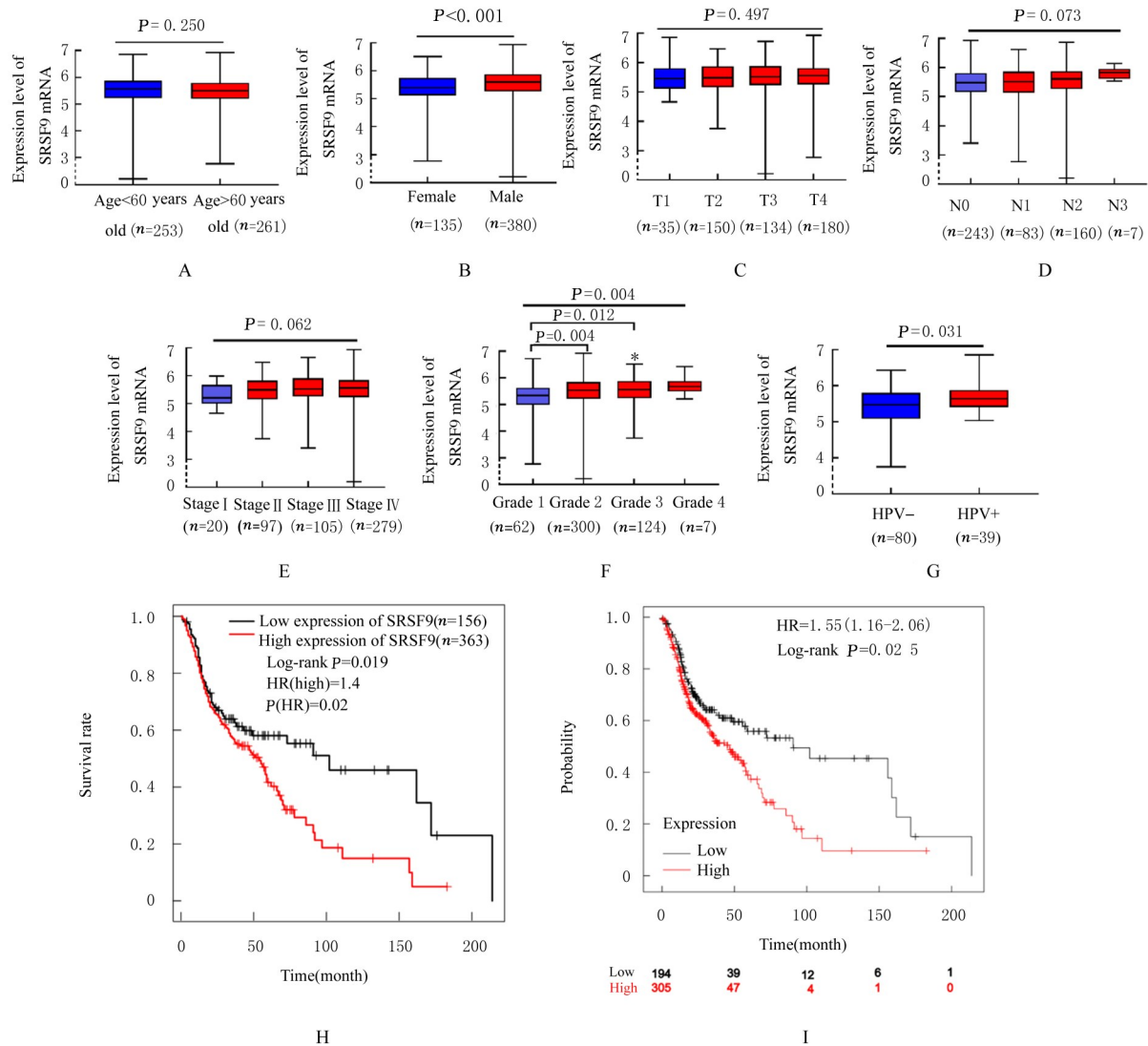
Fig. 1 Expressions of SRSF9 mRNA and protein in HNSCC and adjacent normal tissues

者 SRSF9 mRNA 表达水平与其临床病理特征的关系 (图 2A~2G), 结果显示: 随着肿瘤大小和淋巴结转移数的增加及临床分期以及病理分级的加重, SRSF9 表达呈现上升趋势, 不同性别、病理分级和 HPV 感染状态患者 SRSF9 mRNA 表达水平差异有统计学意义 ($P < 0.05$)。与病理分级 G1 比较, 在病理分级 G2 和 G3 患者的 SRSF9 mRNA 表达水平明显升高 ($P < 0.05$); 与 HPV 阴性比较, 在 HPV 阳性组织中 SRSF9 mRNA 表达水平明显升高 ($P < 0.05$)。利用 GEPIA 数据库分析 TCGA-HNSCC 中 SRSF9 mRNA 表达水平与患者 OS 的关联, 结果显示: SRSF9 mRNA 高表达组患者 OS 显著低于低表达组 ($HR=1.40$, $\text{Log-rank } P=0.019$, 图 2H); Kaplan-Meier plotter 数据库生存分析结果显示: SRSF9 mRNA 高表达组患者的 OS 较短 ($HR=1.55$, $\text{Log-rank } P=0.003$, 图 2I)。

2.3 HNSCC 组织中 SRSF9 基因拷贝数 对 488 例 HNSCC 组织 SRSF9 基因拷贝数变异的类型和频数进行分析, 结果显示: 有 131 例 (26.85%) 发生 SRSF9 基因 CNV, 包括拷贝数缺失 52 例 (10.66%) 和拷贝数扩增 79 例 (16.19%)。与二倍体比较, 基因拷贝数扩增的患者 SRSF9 mRNA 表达水平明显升高 ($P < 0.001$, 图 3A)。

Spearman 相关性分析结果显示: SRSF9 基因 CNV 与其 mRNA 表达水平呈显著正相关关系 ($r=0.44$, $FDR < 0.001$, 图 3B)。SRSF9 拷贝数扩增患者在临床分期 I 期、II 期、III 期和 IV 期中的比例分别为 16.0%、18.6%、18.4% 和 15.4%, 组间比较差异无统计学意义 ($P > 0.05$), 见图 3C; SRSF9 拷贝数扩增患者在病理分级 G1、G2 和 G3 中的比例分别为 10.0%、17.0% 和 15.5%, 各组间比较差异无统计学意义 ($P > 0.05$), 见图 3D)。

2.4 SRSF9 mRNA 表达与 TME 和免疫细胞浸润的关系 使用 ESTIMATE 算法估计 HNSCC TME 的组成, 获得各肿瘤样本的基质评分、免疫评分和估计评分 (图 4A)。与 SRSF9 低表达组比较, SRSF9 高表达组患者的基质评分、免疫评分和估计评分均明显降低 ($P < 0.001$, 见图 4B~4D), 提示肿瘤纯度增高。进一步评估 SRSF9 mRNA 表达水平与 TME 中免疫细胞浸润的关系, 结果见图 5A~5G, SRSF9 mRNA 表达水平与肿瘤纯度和 CD4+T 淋巴细胞浸润呈正相关关系 ($r=0.250$, $P < 0.001$; $r=0.186$, $P < 0.001$); 而与 B 细胞、CD8+T 淋巴细胞和树突状细胞浸润呈负相关关系 ($r=-0.269$, $P < 0.001$; $r=-0.353$, $P < 0.001$; $r=-0.304$, $P < 0.001$)。评估免疫细胞浸润、基



A: Age; B: Gender; C: Tumor diameter (T stage); D: Nodal metastasis (N stage); E: Clinical stage; F: Pathological grade; G: HPV infection; H: GEPIA Database; I: Kaplan-Meier plotter Database.

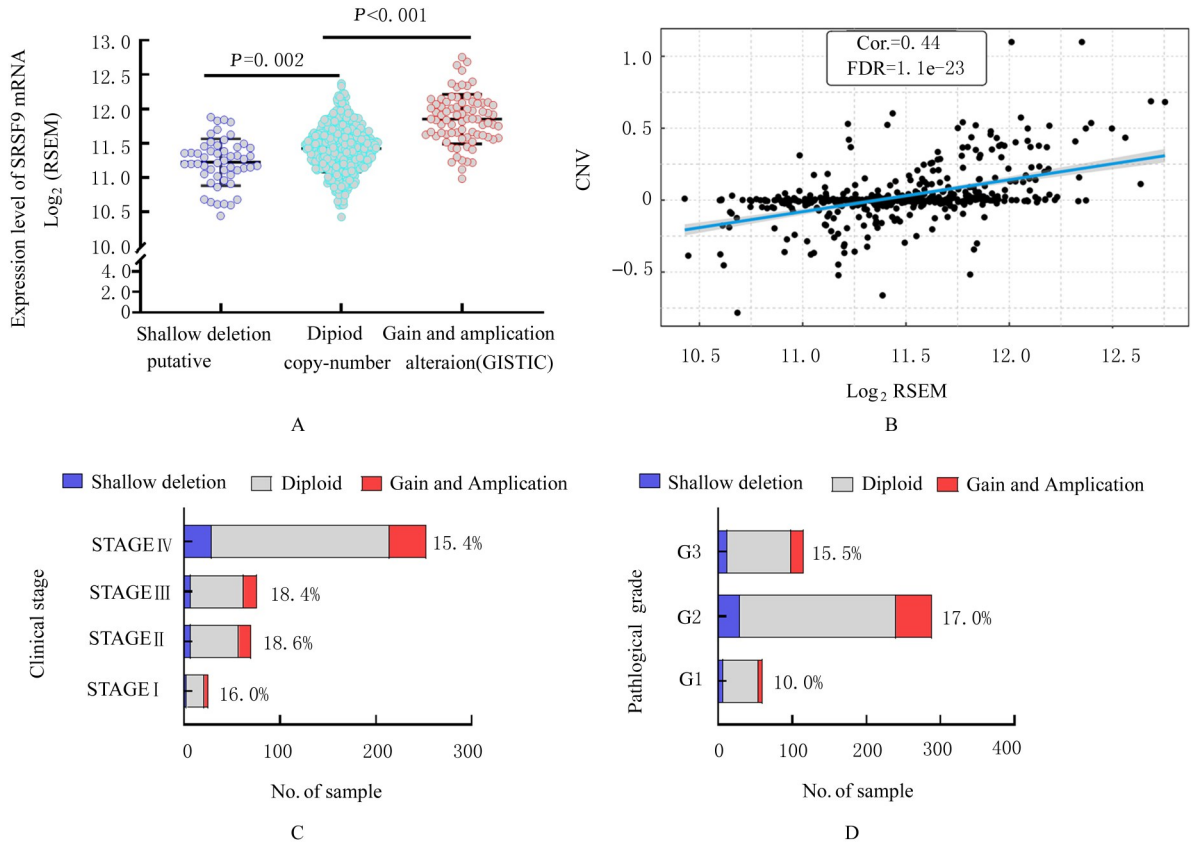
图2 SRSF9 mRNA 表达与HNSCC患者临床病理特征和预后的关系

Fig. 2 Relationships between SRSF9 mRNA expression and clinicopathological characteristics and prognosis of HNSCC patients

因表达与临床预后的关系, 结果显示: 在SRSF9 mRNA高表达组中, 患者较差的OS与B细胞浸润水平降低(HR=0.508, $P=0.019$)和CD8+T淋巴细胞浸润水平降低(HR=0.509, $P=0.028$)相关; 而CD4+T细胞的低浸润量则与患者OS较长相关(HR=2.36, $P=0.016$), 见图6A~6C。

2.5 SRSF9共表达基因及其调控的信号通路 利用LinkedOmics数据库分析TCGA-HNSCC样本中SRSF9的共表达基因, 以FDR<0.01进行筛选结果显示: 4 599个基因与SRSF9表达呈正相关关系, 5 125个基因与SRSF9表达呈负相关关系, 其

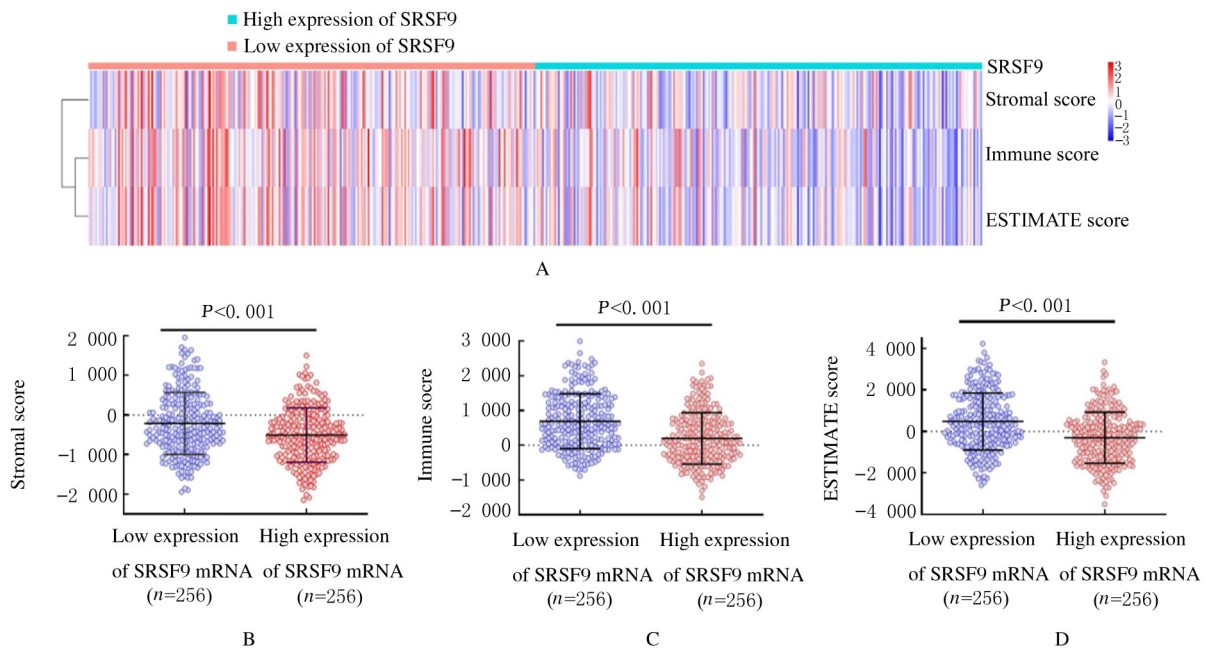
中正相关性最高的前3个基因分别是小核糖核蛋白多肽F (small nuclear ribonucleoprotein polypeptide F, SNRPF)、细胞色素C氧化酶蛋白6A1 (cytochrome c oxidase subunit 6A1, COX6A1)和卷曲螺旋结构域蛋白59 (coiled-coil domain-containing protein 59, CCDC59), 负相关性最高的前3个基因分别是Janus激酶1 (Janus kinase 1, JAK1)、包含盘状同源区域蛋白2 (PDZ domain-containing protein 2, PDZD2)和人免疫缺陷病毒I型增强子结合蛋白3 (human immunodeficiency virus type I enhancer binding protein 3, HIVEP3),



A: CNV of SRSF9; B: Correlation between SRSF9 CNV and its expression; C: Relationship between SRSF9 CNV and clinical stage; D: Relationship between SRSF9 CNV and pathological grade.

图3 HNSCC 组织中 SRSF9 基因的拷贝数

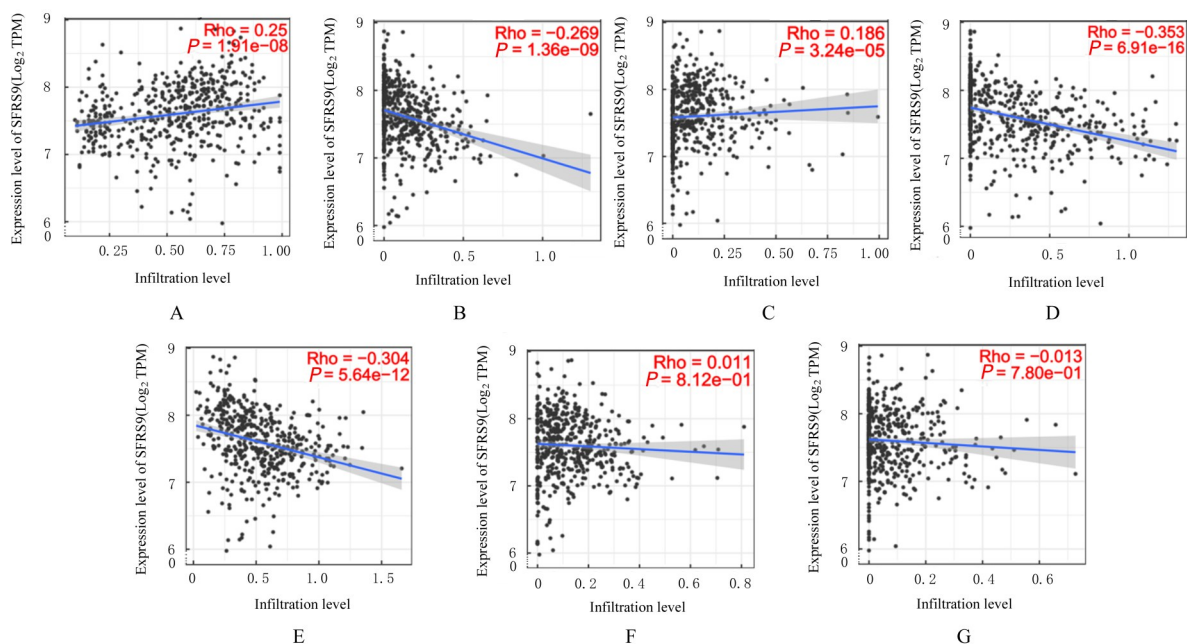
Fig. 3 Copy number of SRSF9 gene in HNSCC tissue



A: Composition of tumor microenvironment; B: Stromal score; C: Immune score; D: ESTIMATE score.

图4 SRSF9 mRNA 表达与 HNSCC TME 组成的关系

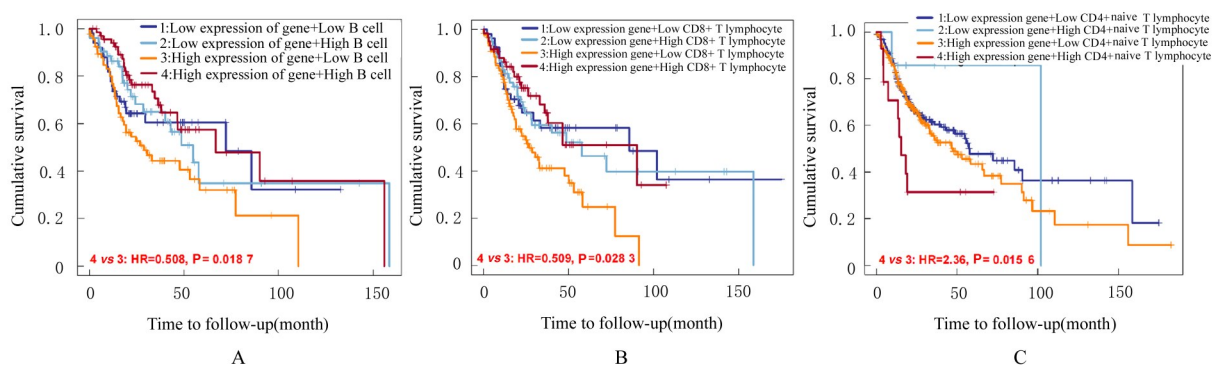
Fig. 4 Relationships between SRSF9 expression and composition of HNSCC TME



A: Purity; B: B lymphocyte; C: CD4+ T lymphocyte; D: CD8+ T lymphocyte; E: Myeloid dendritic cell; F: Neutrophil; G: Macrophage.

图5 SRSF9 mRNA 表达与HNSCC中免疫细胞浸润的关系

Fig. 5 Relationships between SRSF9 expression and immune cell infiltration in HNSCC



A: B lymphocyte; B: CD8+ T lymphocyte; C: CD4+ T lymphocyte.

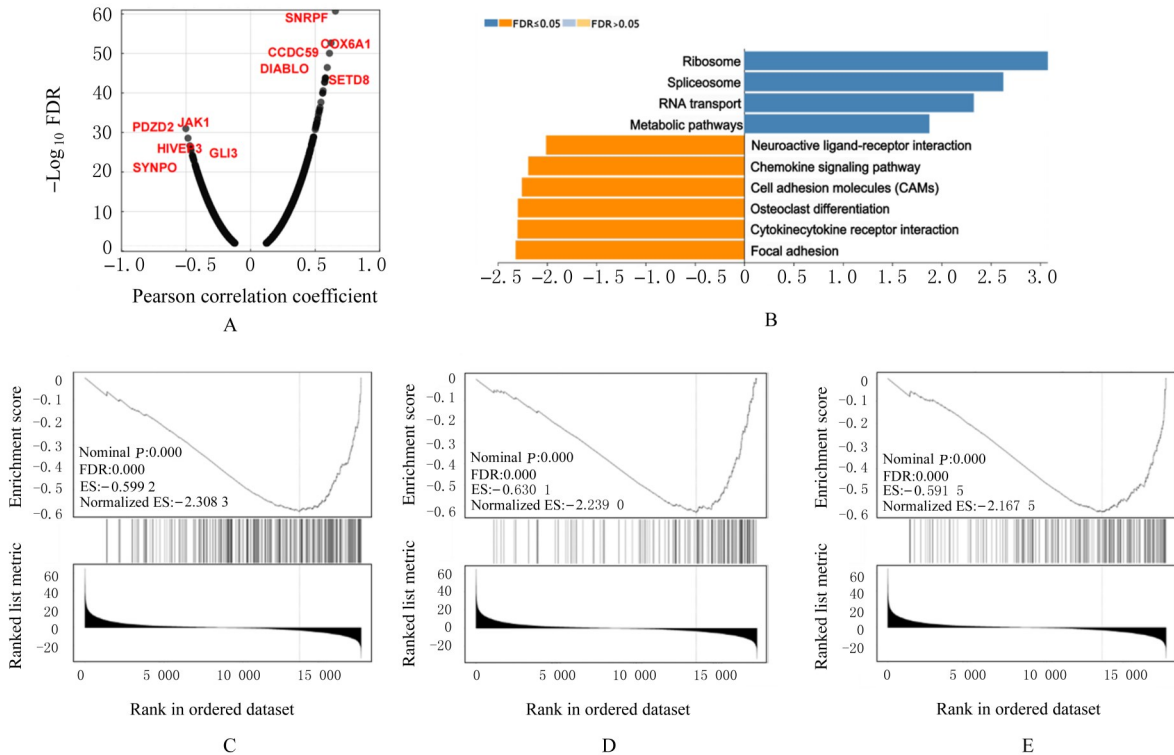
图6 SRSF9 mRNA 表达和免疫细胞浸润与HNSCC患者预后的关系

Fig. 6 Relationships between SRSF9 expression and immune cell infiltration and prognosis of HNSCC patients

见图7A。通过GSEA进行KEGG通路富集分析,结果显示:正相关基因主要参与核糖体、剪接体、RNA转运和代谢通路等;而负相关基因主要富集于黏着斑、细胞因子-细胞因子受体相互作用、破骨细胞分化、细胞黏附分子和趋化因子信号通路等,见图7B。其中,细胞因子-细胞因子受体相互作用、细胞黏附分子和趋化因子信号通路主要参与抗肿瘤免疫调控(图7C~7E)。

2.6 SRSF9调控的AS事件和靶基因 提取SpliceSeq的初始矩阵,共包含42 848个AS、501例HNSCC组织样本和43例配对癌旁正常组织

样本。按照有效值占比>90%且PSI标准差>0.1的标准进一步对AS事件进行过滤,同时排除有效值占比<80%的样本18例。经过滤,共获得485例肿瘤样本和6 712个AS事件纳入后续分析。分别计算SRSF9低表达组和高表达组(图8A)的PSI值,按 $|\Delta\text{PSI}| \geq 0.1$ 筛选共得到1 133个DASE和715个差异剪接基因,其中促进AS事件为606个(53.5%),抑制AS为527个(46.5%),均包含可变受体位点(alternate acceptor site, AA)、可变供体位点(alternate donor site, AD)、外显子跳跃(exon skipping, ES)、可变启动子(alternate



A: Co-expressed genes with SRSF9; B: KEGG pathway analysis; C: Cytokine-cytokine receptor interaction; D: Cell adhesion molecules; E: Chemokine signaling pathway.

图7 HNSCC组织中SRSF9共表达基因的富集分析

Fig. 7 Enrichment analysis on SRSF9 co-expressed genes in HNSCC tissue

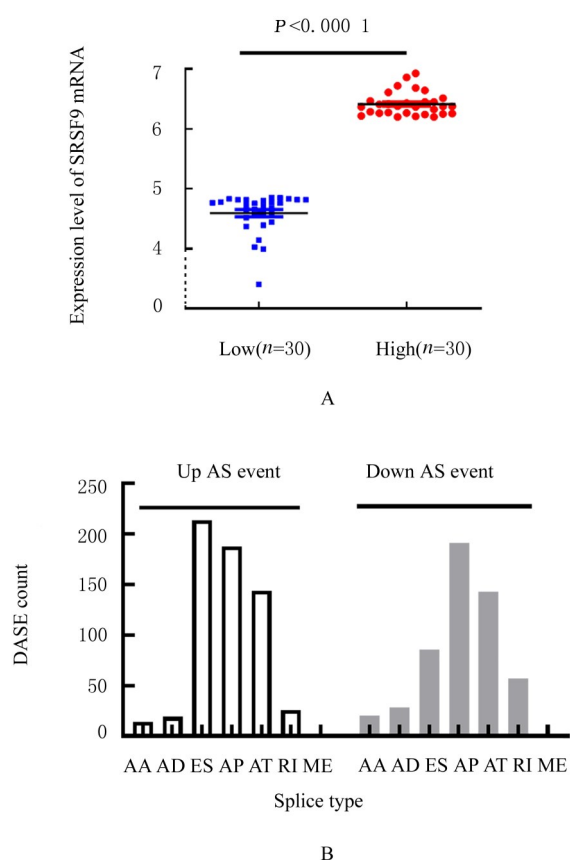
promoter, AP)、可变终止子 (alternate terminator, AT)、内含子保留 (retained intron, RI) 和外显子互斥 (mutually exclusive exons, ME) 7种类型的AS事件。在促进AS事件中, 以ES和AP最常见, 而在抑制AS事件中, 则以AP和AT最常见, 见图8B。对SRSF9调控的AS基因进行GO功能注释和KEGG富集分析, 发现其生物学过程 (biological process, BP) 主要涉及RNA剪接和加工、翻译、DNA修复、蛋白稳定性、催化活性调节、凋亡和细胞迁移等 (图9A); KEGG通路富集分析显示: 主要参与丙酸、二羧酸和脂肪酸代谢, 胰高血糖素信号通路, 紧密连接, 人类免疫缺陷病毒1感染, 剪接体, 碳代谢和氨基酸代谢等 (图9B)。

3 讨论

超过95%的人类基因存在AS, 其调控与机体的生物学功能相适应, AS异常则可能导致生理功能的紊乱和疾病的发生^[7-8]。研究^[9-10]发现: 在肿瘤中存在广泛的AS异常, 并通过调控多种生物学行为 (增殖、抗凋亡、侵袭、血管形成、免疫逃逸

和能量代谢等) 促进肿瘤发生和发展。与正常组织比较, 在肿瘤组织中存在大量肿瘤特异的异常剪接异构体, 这些异常剪接产生的新抗原, 也可能被主要组织相容性复合体 (major histocompatibility complex, MHC) 递呈到肿瘤细胞表面, 并被T淋巴细胞识别, 有望成为肿瘤免疫治疗的潜在新靶点^[11-12]。

AS过程主要受到剪接元件、剪接因子和RNA表观遗传修饰等的调控。SRSFs蛋白是一类重要的剪接因子家族, 主要参与pre-mRNA组成性剪接和可变剪接、mRNA出核、稳定性以及翻译等多个过程^[13]。SRSFs多个成员在不同类型肿瘤组织中均存在异常表达, 其导致肿瘤相关基因的AS失调, 可能是肿瘤发生发展的重要机制^[14-16]。SRSF1在乳腺癌、肺癌和结直肠癌等多种肿瘤组织中的表达水平均有所上升, 其通过影响MYB、乳腺癌易感基因1 (breast cancer susceptibility gene 1, BRCA1) 和血管内皮生长因子 (vascular endothelial growth factor, VEGF) 等基因的AS异常, 进而促进细胞增殖、抗凋亡和迁移等生物学过



A: SRSF9 high-expression group and low-expression group;
B: DASE.

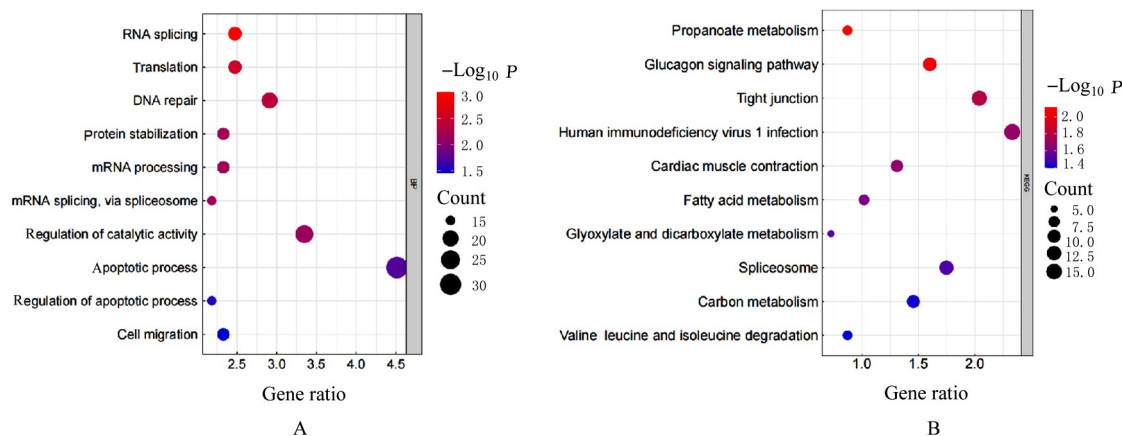
图8 HNSCC组织中SRSF9调控的AS事件

Fig. 8 AS events regulated by SRSF9 in HNSCC tissue

程^[16-18]。FU等^[4]发现在NIH3T3细胞中过表达SRSF9可促进软琼脂集落形成,并在裸鼠体内诱导肿瘤形成,表明SRSF9具有致瘤功能。SRSF9在肝癌组织中呈明显高表达,并且与其进展和预后

不良有关^[6];LIU等^[19]证实长链非编码RNA牛磺酸上调基因1(long non-coding RNA taurine up-regulated gene 1, lncRNA TUG1)/微小RNA-328-3p(microRNA-328-3p)/SRSF9轴的异常调控促进肝癌的发生和发展。研究^[20]发现:SRSF9通过调控谷胱甘肽过氧化物酶4(glutathione peroxidase 4, GPX4)的表达驱动结直肠癌的发生,并抵抗Erastin诱导的铁死亡,为靶向铁死亡治疗结直肠癌提供新策略。

本研究通过对TCGA、GEO和CPTAC数据库的挖掘,发现在多组HNSCC的研究中,在HNSCC肿瘤组织中SRSF9 mRNA和蛋白表达水平有明显升高,SRSF9 mRNA表达水平对患者具有较好的诊断价值;随着肿瘤大小、淋巴结转移以及临床分期和病理分级的进展,SRSF9 mRNA表达呈上升趋势,且其高表达与患者预后不良显著相关。高危型HPV16/18亚型感染被认为是HNSCC,尤其是口咽部肿瘤的独立致病因素^[21]。HPV16感染可通过上调SRSF蛋白激酶1(SRSF protein kinase 1, SRPK1)导致磷酸化SRSF1水平升高,调控角质形成细胞的分化^[22]。SOMBERG等^[23]研究显示:SRSF9可直接引起HPV16基因组剪接位点改变,诱导HPV16晚期衣壳蛋白L1 mRNA的产生。与HPV阴性患者比较,HPV阳性HNSCC患者肿瘤组织中SRSF9 mRNA表达水平明显升高,提示SRSF9可能参与HPV病毒的组装和增殖,但其具体机制有待进一步探讨。上述研究结果提示:SRSF9可能作为促癌基因参与HNSCC的发生发展,其表达水平在一定程度上



A: GO analysis on biological process; B: KEGG signaling pathway analysis.

图9 HNSCC中SRSF9调控AS基因的功能富集分析

Fig. 9 Functional enrichment analysis on AS genes regulated by SRSF9 in HNSCC tissue

可影响患者预后。

CNV 是癌基因组改变的重要特点, 通过引起基因剂量、活性或结构的改变影响基因的表达、表型变异和适应性, 是肿瘤遗传变异的关键因素^[24]。HNSCC 具有高度 CNV, ESSERS 等^[25] 发现 7 种癌源性 CNV 特征对 HNSCC 患者顺铂治疗反应和预后具有一定的预测价值。在 HNSCC 中, SRSF9 基因具有较高的变异率 (26.85%), 主要表现为基因的扩增 (16.19%); 相关性分析结果显示: SRSF9 基因 CNV 与其 mRNA 表达水平呈正相关关系, 表明 HNSCC 中 SRSF9 mRNA 表达上调可能与其拷贝数扩增有关, 但未发现其 CNV 与临床分期和病理分级相关。

TME 是肿瘤细胞赖以生存和发展的内环境, 主要由细胞外基质、成纤维细胞、血管/淋巴管和免疫细胞组成。免疫细胞和基质细胞作为 TME 中 2 种主要类型的非肿瘤组分, 对肿瘤进展、预后和免疫治疗的有效性有显著的影响^[26]。本研究结果显示: 在 HNSCC 中, SRSF9 表达水平与 TME 中免疫细胞和基质细胞的浸润呈负相关关系, 与肿瘤纯度呈正相关关系, 提示 SRSF9 可能参与 HNSCC 免疫抑制微环境的调控。进一步分析结果显示: SRSF9 mRNA 表达与 CD4+T 淋巴细胞浸润呈正相关关系; 与 B 细胞、CD8+T 淋巴细胞和树突状细胞浸润呈负相关关系。CD8+T 淋巴细胞是肿瘤免疫的主要效应细胞, 肿瘤抗原特异性 CD8+T 淋巴细胞激活不足以及肿瘤中免疫浸润不良是肿瘤重要的免疫逃逸机制。SRSF9 mRNA 高表达导致 CD8+T 细胞低浸润, 并与患者预后不良相关, 其可能参与 HNSCC 肿瘤免疫逃逸的发生。HNSCC 组织中 SRSF9 共表达基因的 GSEA 聚类分析显示: SRSF9 共表达基因与细胞因子-细胞因子受体相互作用、细胞黏附分子和趋化因子信号通路均呈负相关关系, 上述信号通路在肿瘤免疫功能维持和免疫耐受产生等过程中发挥重要作用。与 SRSF9 呈显著负相关关系的 JAK1 是非受体酪氨酸激酶 Janus 激酶家族的关键成员, 研究^[27] 显示: 抑制 JAK1/STAT3 可通过作用于 TME 引起促肿瘤炎症因子的产生, 从而导致乳腺癌患者治疗抵抗。上述结果提示: SRSF9 可能通过调控 HNSCC 中各种肿瘤浸润免疫细胞的浸润水平及其活化状态, 从而影响免疫抑制微环境的形成, 促进肿瘤发生发展, 并影响患者预后, SRSF9 可能成为 HNSCC 肿瘤免疫治疗

的潜在靶点。

当 SRSFs 在肿瘤中表达异常时, 则可能导致肿瘤特异的 AS 事件失调, 促进细胞恶性生物学行为的发生。本研究利用 TCGA SpliceSeq 数据库分析 HNSCC 中 SRSF9 差异表达引起的 AS 事件的改变筛选 SRSF9 相关 AS 事件和基因, 探讨其发挥促癌功能的分子机制。对 SRSF9 相关 AS 基因的 KEGG 通路分析结果显示: 其主要参与脂质代谢、糖代谢和氨基酸代谢等物质代谢过程及调节, 对 SRSF9 共表达基因的分析也显示其正相关基因主要参与代谢通路的调控。肿瘤在发生发展过程中常伴随着代谢的改变, 对肿瘤代谢组学的研究将为肿瘤早期诊断、个性化治疗以及预后判断提供新的生物标志物^[28-29]。在 HNSCC 中发现有脂质合成的异常活跃和多种脂质代谢酶的上调, 通过脂质代谢重编程以应对 TME 和转移定植的需求, 靶向脂质代谢有望成为 HNSCC 治疗的新策略^[30-31]。本研究发现 SRSF9 相关可变剪接基因参与丙酸、二羧酸和脂肪酸等多个脂质代谢过程, 提示其调控的脂质代谢异常可能是 SRSF9 促进 HNSCC 发生发展的重要因素, 但具体作用靶点和机制还有待进一步探讨。

综上所述, 本研究发现剪接因子 SRSF9 在 HNSCC 中明显高表达, 且与病理分级和患者预后不良相关; SRSF9 表达水平可影响 TME 中免疫细胞浸润, 并通过介导可变剪接参与脂质代谢等通路的调控, 促进 HNSCC 发生发展。SRSF9 有望成为 HNSCC 诊断、预后生物标志物和潜在的免疫治疗靶点。

利益冲突声明:

所有作者声明不存在利益冲突。

作者贡献声明:

李彬彬、刘玉婷负责论文的整体设计, 刘玉婷和喻莹负责论文的撰写, 刘玉婷、喻莹、史沁雪和黎桂珍负责论文的数据收集和统计学分析, 李彬彬负责论文的修改和最后定稿。

[参考文献]

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA Cancer J Clin*, 2021, 71 (3): 209-249.
- [2] SCIARRILLO R, WOJTUSZKIEWICZ A,

- ASSARAF Y G, et al. The role of alternative splicing in cancer: from oncogenesis to drug resistance [J]. *Drug Resist Updat*, 2020, 53:100728.
- [3] FACKENTHAL J D. Alternative mRNA splicing and promising therapies in cancer [J]. *Biomolecules*, 2023, 13(3):561.
- [4] FU Y, HUANG B L, SHI Z, et al. SRSF1 and SRSF9 RNA binding proteins promote Wnt signalling-mediated tumorigenesis by enhancing β -catenin biosynthesis [J]. *EMBO Mol Med*, 2013, 5(5):737-750.
- [5] WANG X Y, LU X S, WANG P, et al. SRSF9 promotes colorectal cancer progression via stabilizing DSN1 mRNA in an m6A-related manner [J]. *J Transl Med*, 2022, 20(1):198.
- [6] ZHANG G S, LIU B, SHANG H, et al. High expression of serine and arginine-rich splicing factor 9 (SRSF9) is associated with hepatocellular carcinoma progression and a poor prognosis [J]. *BMC Med Genomics*, 2022, 15(1):180.
- [7] HUI J Y. Regulation of mammalian pre-mRNA splicing [J]. *Sci China C Life Sci*, 2009, 52(3): 253-260.
- [8] CHOI S, CHO N, KIM K K. The implications of alternative pre-mRNA splicing in cell signal transduction [J]. *Exp Mol Med*, 2023, 55(4):755-766.
- [9] KOZLOVSKI I, SIEGFRIED Z, AMAR-SCHWARTZ A, et al. The role of RNA alternative splicing in regulating cancer metabolism [J]. *Hum Genet*, 2017, 136(9):1113-1127.
- [10] BRADLEY R K, ANCZUKÓW O. RNA splicing dysregulation and the hallmarks of cancer [J]. *Nat Rev Cancer*, 2023, 23(3):135-155.
- [11] FRANKIW L, BALTIMORE D, LI G D. Alternative mRNA splicing in cancer immunotherapy [J]. *Nat Rev Immunol*, 2019, 19(11):675-687.
- [12] CHENG R, XU Z C, LUO M, et al. Identification of alternative splicing-derived cancer neoantigens for mRNA vaccine development [J]. *Brief Bioinform*, 2022, 23(2): bbab553.
- [13] KUMAR K, SINHA S K, MAITY U, et al. Insights into established and emerging roles of SR protein family in plants and animals [J]. *Wiley Interdiscip Rev RNA*, 2023, 14(3): e1763.
- [14] LUO C L, CHENG Y M, LIU Y G, et al. SRSF2 regulates alternative splicing to drive hepatocellular carcinoma development [J]. *Cancer Res*, 2017, 77(5): 1168-1178.
- [15] WAN L D, DENG M, ZHANG H H. SR splicing factors promote cancer via multiple regulatory mechanisms [J]. *Genes*, 2022, 13(9): 1659.
- [16] ZHENG X, PENG Q, WANG L J, et al. Serine/arginine-rich splicing factors: the bridge linking alternative splicing and cancer [J]. *Int J Biol Sci*, 2020, 16(13):2442-2453.
- [17] JIANG L Y, HUANG J Q, HIGGS B W, et al. Genomic landscape survey identifies SRSF1 as a key oncodriver in small cell lung cancer [J]. *PLoS Genet*, 2016, 12(4): e1005895.
- [18] DU J X, LUO Y H, ZHANG S J, et al. Splicing factor SRSF1 promotes breast cancer progression via oncogenic splice switching of PTPMT1 [J]. *J Exp Clin Cancer Res*, 2021, 40(1):171.
- [19] LIU Y D, MAO X, MA Z C, et al. Aberrant regulation of LncRNA TUG1-microRNA-328-3p-SRSF9 mRNA Axis in hepatocellular carcinoma: a promising target for prognosis and therapy [J]. *Mol Cancer*, 2022, 21(1):36.
- [20] WANG R, SU Q, YIN H Z, et al. Inhibition of SRSF9 enhances the sensitivity of colorectal cancer to erastin-induced ferroptosis by reducing glutathione peroxidase 4 expression [J]. *Int J Biochem Cell Biol*, 2021, 134: 105948.
- [21] TACHEZY R, KLOZAR J, RUBENSTEIN L, et al. Demographic and risk factors in patients with head and neck tumors [J]. *J Med Virol*, 2009, 81(5): 878-887.
- [22] MOLE S, FAIZO A A A, HERNANDEZ-LOPEZ H, et al. Human papillomavirus type 16 infection activates the host serine arginine protein kinase 1 (SRPK1) - splicing factor axis [J]. *J Gen Virol*, 2020, 101(5): 523-532.
- [23] SOMBERG M, LI X Z, JOHANSSON C, et al. Serine/arginine-rich protein 30c activates human papillomavirus type 16 L1 mRNA expression via a bimodal mechanism [J]. *J Gen Virol*, 2011, 92(Pt 10): 2411-2421.
- [24] TAN E S, KNEPPER T C, WANG X, et al. Copy number alterations as novel biomarkers and therapeutic targets in colorectal cancer [J]. *Cancers*, 2022, 14(9): 2223.
- [25] ESSERS P B M, VAN DER HEIJDEN M, VOSSSEN D, et al. Ovarian cancer-derived copy number alterations signatures are prognostic in chemoradiotherapy-treated head and neck squamous cell carcinoma [J]. *Int J Cancer*, 2020, 147(6):1732-1739.
- [26] TURLEY S J, CREMASCO V, ASTARITA J L. Immunological hallmarks of stromal cells in the tumour microenvironment [J]. *Nat Rev Immunol*, 2015, 15(11): 669-682.

- [27] CHEN B, LAI J G, DAI D N, et al. JAK1 as a prognostic marker and its correlation with immune infiltrates in breast cancer [J]. *Aging (Albany NY)*, 2019, 11(23):11124-11135.
- [28] SCHMIDT D R, PATEL R, KIRSCH D G, et al. Metabolomics in cancer research and emerging applications in clinical oncology [J]. *CA Cancer J Clin*, 2021, 71(4): 333-358.
- [29] BIAN X L, LIU R, MENG Y, et al. Lipid metabolism and cancer [J]. *J Exp Med*, 2021, 218(1):1-17.
- [30] MIAO X W, WANG B L, CHEN K L, et al. Perspectives of lipid metabolism reprogramming in head and neck squamous cell carcinoma: an overview [J]. *Front Oncol*, 2022, 12:1008361.
- [31] LIN X H, ZHOU W K, LIU Z Q, et al. Targeting cellular metabolism in head and neck cancer precision medicine era: a promising strategy to overcome therapy resistance [J]. *Oral Dis*, 2023, 29(8): 3101-3120.