本文引文格式:冯浩然,粟小平,穆婷,等.基于生物信息学对头颈鳞癌的关键基因进行分析[J].右江民族医学院学报,2020,42(5):568-573.

【论著与临床报道】

基于生物信息学对头颈鳞癌的关键基因进行分析

冯浩然,粟小平,穆婷,王爽,黄旋平,李翠萍,何雨亮

[广西医科大学附属口腔医院,广西口腔颌面修复与重建研究自治区级重点实验室,
广西颅颌面畸形临床医学研究中心,颌面外科疾病诊治研究重点实验室
(广西高校重点实验室),广西 南宁 530021]

摘 要:目的 探讨头颈部鳞状细胞癌(head and neck squamous cell carcinoma,HNSCC)潜在的关键基因。**方法** 从 GEO 数据库中下载 GSE6631 和 GSE13398 芯片数据集,通过 GEO2R 工具筛选出 HNSCC 癌组织与正常组织的差异表达基因(differentially expressed genes,DEGs)。HNSCC 高通量测序数据从 TCGA 数据库中下载,通过 R 包 edgR 筛选 HNSCC 癌组织与正常组织的 DEGs。利用 STRING 数据库构建蛋白互作网络(protein-protein interaction network,PPI network)。关键基因通过 Cytoscape 软件及 GEPIA 在线工具进行筛选,并进行 GO(Gene Ontology)功能注释分析和 KEGG (Kyoto Encyclopedia of Genes and Genomes)通路富集分析。最后通过 QRT-PCR 在临床标本上对关键基因表达进行验证。**结果** 在本研究中,共筛选出 70 个 DEGs,其中 34 个基因表达上调,36 个基因表达下调。PPI 网络中候选关键基因的功能分析显示大部分基因富集于细胞外基质组织、细胞外基质受体相互作用、黏着斑和 PI3K-Akt 信号通路等。MCC 算法结合生存分析,最终确定 7 个关键基因:PLAU、FAP、LAMC2、SERPINH1、ITGA6、SPP1 和 MMP1。**结论** 7 个关键基因 PLAU、FAP、LAMC2、SERPINH1、ITGA6、SPP1、MMP1 可为 HNSCC 的诊断及治疗提供潜在靶点。

关键词:头颈部肿瘤;癌,鳞状细胞;差异表达基因;生物信息学

中图分类号:R739.91 **文献标识码:** A **文章编号:** 1001-5817(2020)05-0568-06

doi:10.3969/j.issn.1001-5817.2020.05.007

Analysis of key genes in head and neck squamous cell carcinoma based on bioinformatics

Feng Haoran, Su Xiaoping, Mu Ting, Wang Shuang, Huang Xuanping, Li Cuiping, He Yuliang

(*Affiliated Stomatology Hospital of Guangxi Medical University, Guangxi Key Laboratory of Oral and Maxillofacial Rehabilitation and Reconstruction, Guangxi Clinical Medical Research Center for Craniomaxillofacial Deformity, Key Laboratory for Diagnosis and Treatment of Maxillofacial Diseases/Key Laboratory of Guangxi Universities, Nanning* 530021, *Guangxi, China*)

Abstract: Objective To explore the potential key genes for head and neck squamous cell carcinoma (HNSCC). **Methods** The GSE6631 and GSE13398 microarrays were downloaded from the GEO database, and differentially expressed genes (DEGs) of HNSCC tissues and normal tissues were screened out by GEO2R. HNSCC high-throughput sequencing data were downloaded from TCGA database. DEGs of HNSCC cancer tissues and normal tissues were screened by R package edgR. STRING database was used to construct the pro-

基金项目:广西自然科学基金项目(2018GXNSFAA138003);广西科技计划项目(桂科 AD17129004);广西医疗卫生适宜技术开发与推广应用项目(S201687)

第一作者简介:冯浩然(1994-),男,在读硕士研究生,研究方向:口腔颌面外科基础与临床研究,E-mail:450639276@qq.com

通讯作者简介:黄旋平(1973-),男,医学博士,教授,主任医师,研究方向:口腔颌面外科基础与临床研究,E-mail:hxp120@126.com

tein-protein interaction network (PPI network). Key genes were screened by Cytoscape software and GEPIA online tools, and GO (Gene Ontology) functional annotation analysis and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis were performed. Finally, QRT-PCR was used to verify the expression of key genes in clinical specimens. **Results** In this study, a total of 70 DEGs were selected, among which 34 genes were up-regulated and 36 genes were down-regulated. Functional analysis of key candidate genes in PPI network showed that most of the genes were enriched in extracellular matrix tissues, extracellular matrix receptor interaction, adhesion spots and PI3K-Akt signal transduction pathway. MCC algorithm combined with survival analysis finally identified 7 key genes: PLAU, FAP, LAMC2, SERPINH1, ITGA6, SPP1 and MMP1. **Conclusion** Seven key genes, PLAU, FAP, LAMC2, SERPINH1, ITGA6, SPP1 and MMP1, can provide potential targets for the diagnosis and treatment of HNSCC.

Key words: head and neck neoplasm; carcinoma, squamous cell; differentially expressed genes; bioinformatics

头颈部鳞状细胞癌(head and neck squamous cell carcinoma,HNSCC)是世界第六大常见癌症[1]。其特点是死亡率高,复发率高,异质性高,5 年生存率低及预后较差[2-4]。近些年来,虽然在 HNSCC 相关手术和化疗等方面有一定研究进展,但目前尚无有效的诊断和治疗方法及特效药物[5]。因此探究 HNSCC 发病相关的关键基因对于发现更有效的诊断和治疗方法具有重要意义。

近年来,生物信息学分析被广泛应用于各种癌症诊断和治疗的生物标志物预测中。在前列腺癌中,通过对 GEO 数据库的分析,发现 HSPA8、PPP2R1A、CTNNB1、ADCY5、ANXA1、COL9A2 为关键基因[6]。通过分析 GEO 及 TCGA 数据库,发现 CASR、CXCL12、SST 为与胃癌相关的关键基因[7]。许多与上述类似的研究也可见于多种与肿瘤发生相关的关键基因筛选中[8-11],生物信息学分析已成为寻找肿瘤潜在诊断和治疗靶点的有效手段。本研究通过对 GEO 及 TCGA 数据库的挖掘,筛选出与 HNSCC 相关的差异表达关键基因,这将有助于对进一步了解 HNSCC 的发生发展过程,并为探索其诊断与治疗靶点提供依据。

1　材料与方法

1.1　数据下载　从 GEO 数据库下载基因芯片数据 GSE6631 和 GSE13398,其中 GSE6631 包括 22 例 HNSCC 肿瘤组织和 22 例正常组织,GSE13398 包括 8 例 HNSCC 肿瘤组织和 8 例正常组织。RNA-seq 数据从 TCGA 数据库下载,其中包括 502 例 HNSCC 肿瘤组织和 44 例正常组织。

1.2　差异表达基因(differentially expressed genes,DEGs)的筛选　利用在线工具 GEO2R 对 GSE6631 和 GSE13398 的 DEGs 进行分析,用 R 包 edgR 提取 TCGA 数据库的 DEGs(以 $P<0.05$,$|\log2FC|\geq1$ 作为截止标准)。利用在线工具 Draw Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/)绘制韦恩图,对 GSE6631、GSE13398 和 TCGA 在 HNSCC 中的 DEGs 取交集。

1.3　功能富集分析　使用在线软件 DAVID(S6.8 版本,http://david.abcc.ncifcrf.gov/)和 KOBAS(3.0 版本,http://kobas.cbi.pku.edu.cn/)对重叠的 DEGs 分别进行 GO 功能注释分析($P<0.05$ 设为阈值)和 KEGG 通路富集分析($P<0.05$ 设为阈值)[12-13]。使用 R 包 ggplot2 绘制柱状图。

1.4　PPI 网络构建与分析　首先将重叠 DEGs 导入 STRING 在线数据库(http://string-db.org)进行分析。之后使用 Cytoscape 软件构建 PPI 网络并分析 DEGs 相互作用关系。最后利用 Cytohubba 插件 MCC 算法构建模块,筛选出候选关键基因,最后对其进行功能富集分析。

1.5　总体生存分析　使用 GEPIA(http://gepia.pku.cn/index.html)对筛选出的候选关键基因进行总体生存分析[14],截止标准设为 $P<0.05$。

1.6　验证关键基因在临床样本中的表达　在采集样本前获得患者知情同意后,取来自 HNSCC 患者的 10 例口腔肿瘤样本与相匹配的正常样本。用 TRIZOL 试剂从肿瘤组织和正常组织中提取总 RNA。用 Nanodrop 2000 和琼脂糖凝胶电泳检测总 RNA 的质量和浓度。cDNA 使用 PrimeScript™RT 试剂盒(Takara,RR047A,日本)获得。QRT-PCR 使用 TB Green® Premix Ex Taq™ II (Tli RNaseH Plus)(Takara,RR820A,日本)和 QuantStudio™ 5 Real-Time PCR 系统(Applied Biosystems,美国)进行。相关表达用 $2^{-\Delta\Delta CT}$ 进行计算分析。$P<0.05$ 被认为差异具有统计学意义。

2　结果

2.1　差异表达基因　在芯片数据集 GSE6631 筛选出

187 个 DEGs，在芯片数据集 GSE13398 筛选出 4259 个 DEGs。从 TCGA 数据库中筛选出 19844 个 DEGs。在两个基因芯片数据集和 TCGA 数据中共发现 70 个 DEGs 存在重叠（见图 1、表 1）。

2.2　DEGs 的 GO 功能注释分析　对 HNSCC 中 70 个重叠 DEGs 进行 GO 分析。结果显示，DEGs 主要富集于细胞外基质分解、细胞黏附和胶原分解代谢等生物学过程。细胞组成分析表明，DEGs 在细胞外基质、细胞外外泌体中富集。分子功能分析表明，DEGs 主要富集于细胞外基质结构成分，血小板来源的生长因子结合等（见图 2A）。

2.3　DEGs 的 KEGG 通路富集分析　KEGG 信号通路富集分析结果显示，大多数 DEGs 富集于细胞外基质受体相互作用、黏着斑和 PI3K-Akt 信号通路（见图 2B）。

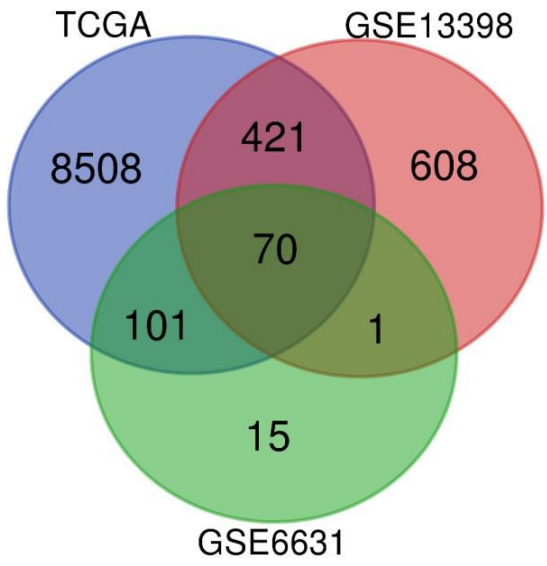


图 1　GSE6631、GSE13398 和 TCGA 中 DEGs 维恩图

表 1　重叠 DEGs 的上调和下调

DEGs	Genes Name
Up-regulated	COL1A1，HTRA1，SULF1，COL11A1，PDPN，COL3A1，FN1，SPP1，NEFL，MMP12，MMP11，COL4A2，LUM，LAMC2，SERPINH1，FAP，THBS2，TGFBI，CDH11，MMP1，MMP3，COL1A2，ITGB4，COL6A3，SPARC，COL5A2，ITGA6，POSTN，COL4A1，MMP10，PLAU，LAMB1，SEMA3C，CTSC
Down-regulated	FHL1，CYP3A5，CLDN10，KRT4，KRT13，MAL，SERPINB2，NUCB2，COBL，TFF3，CEACAM6，PPL，SASH1，IL1RN，CEACAM1，PSCA，ECM1，PRSS3，HPGD，MGLL，CD24，KLK11，DUSP5，CEACAM5，CSTB，SLURP1，PRR4，CRISP3，CLU，ZNF185，BLNK，LPIN1，ANXA1，SCEL，EXPH5，PPP1R3C

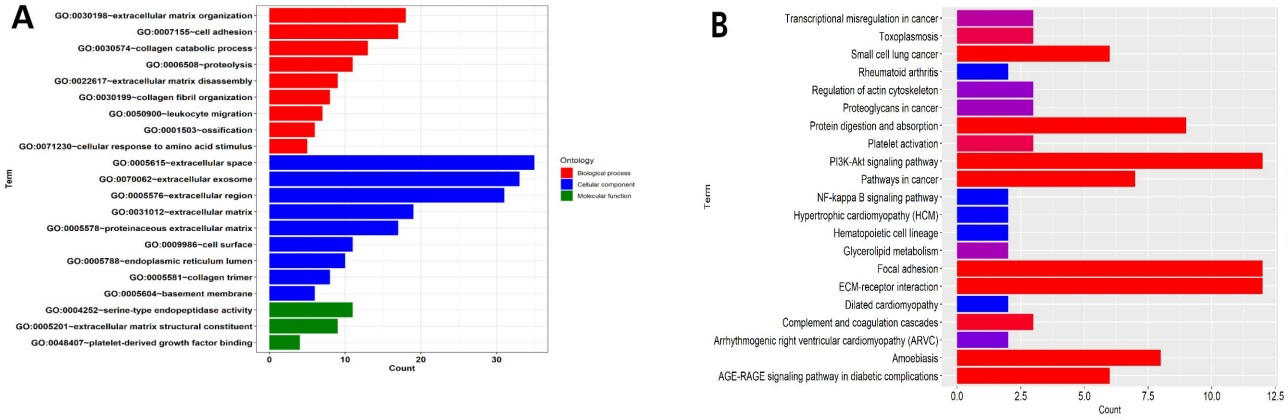
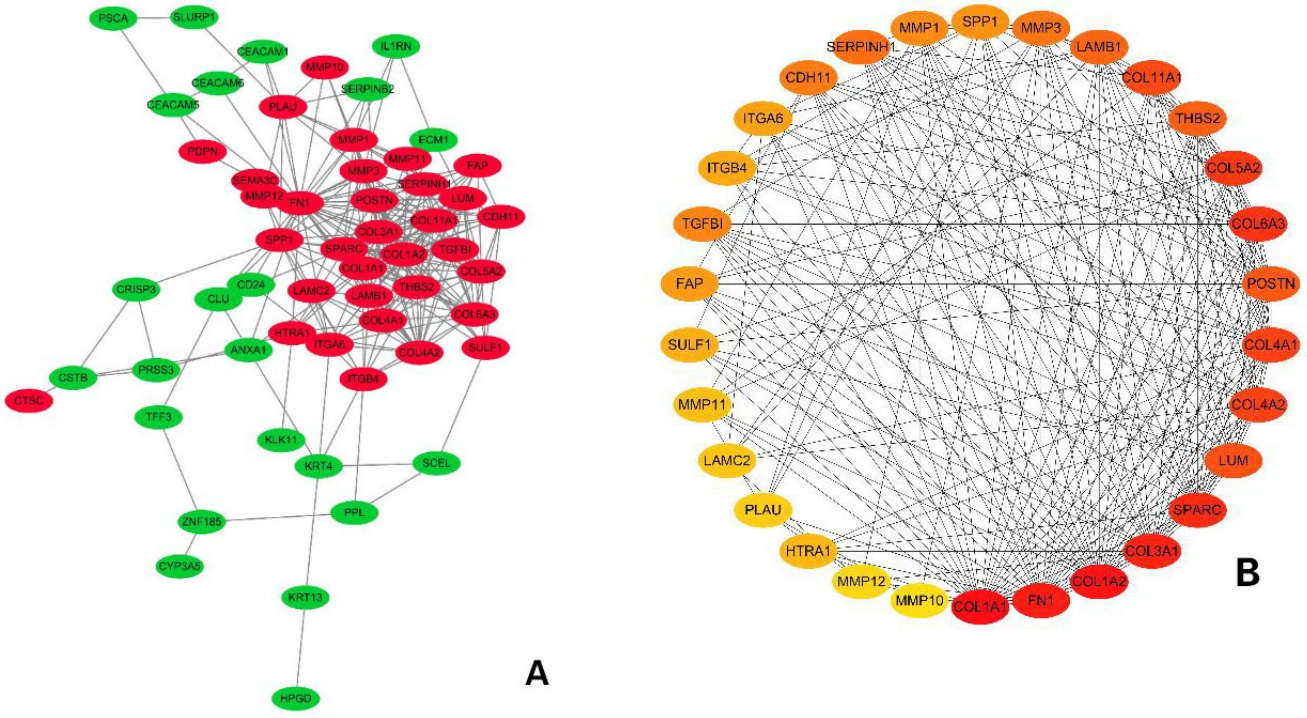


图 2　DEGs 的 GO 功能注释和 KEGG 通路富集分析

2.4　DEGs 的 PPI 网络分析　通过 STRING 数据库对 DEGs 进行分析，共筛选出互作评分＞0.4 的 56 个 DEGs（33 个上调，23 个下调）进入 PPI 网络，该网络包括 56 个节点及 265 条边。利用 Cytoscape 软件构建可视化 PPI 网络（见图 3A）。利用插件 CytoHubba 的 MCC 算法进一步分析 PPI 网络，筛选出前 30 个候选关键基因（见图 3B）。对 30 个候选关键基因进行功能富集分析，结果显示候选关键基因主要与细胞外基质组织、胶原分解代谢过程、细胞外基质受体相互作用、黏着斑、PI3K-Akt 信号通路等相关（见图 4A、图 4B）。

2.5　候选关键基因的生存分析　使用基于 TCGA 数据库的在线工具 GEPIA 对候选关键基因进行生存分析。分析结果表明，以下 7 个基因：PLAU（P = 0.00049）、FAP（P = 0.0027）、LAMC2（P = 0.013）、SERPINH1（HSP47）（P = 0.022）、ITGA6（P = 0.035）、SPP1（P = 0.045）和 MMP1（P = 0.046）与 HNSCC 患者的总体生存期显著相关（见图 5），这些基因的高表达与较差的总体生存期有关。



注：A：差异上调基因（红色）和下调基因（绿色）的 PPI 网络；B：利用 Cytohubba 插件 MCC 算法在 PPI 网络中建立模块。

图 3　PPI 网络与模型分析

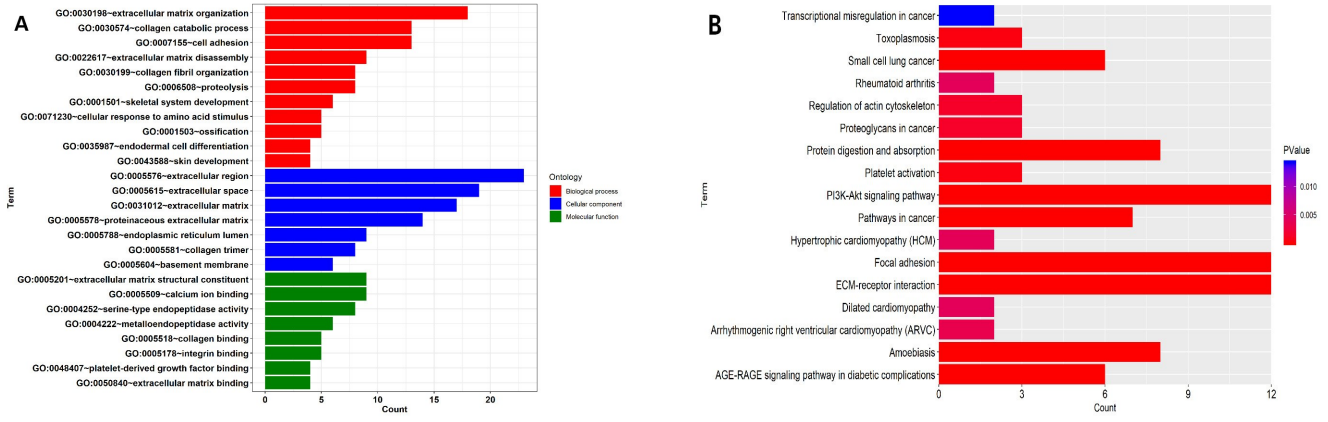
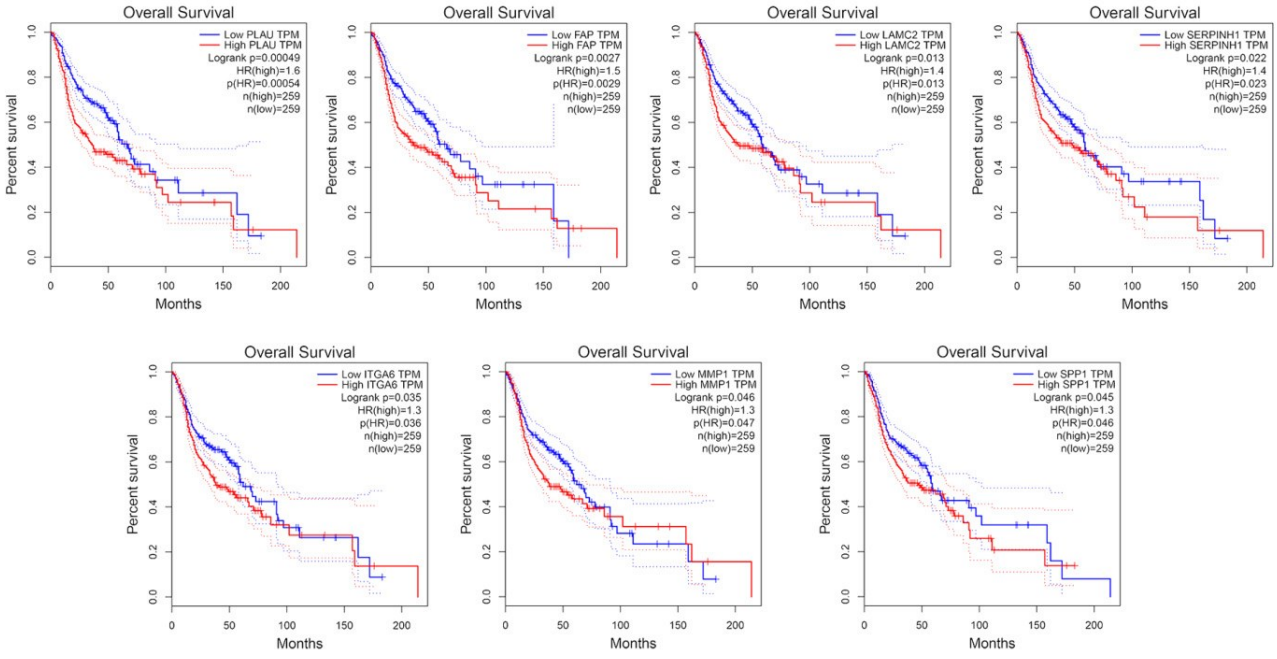


图 4　候选关键基因的 GO 功能注释和 KEGG 通路富集分析

2.6　临床标本中关键基因的 mRNA 表达　应用 QRT-PCR 检测 10 对 HNSCC 组织及正常组织的关键基因 mRNA 表达水平。结果显示，肿瘤组织中以下 7 个基因 PLAU（$P=0.017$）、FAP（$P=0.025$）、LAMC2（$P=0.019$）、SERPINH1（HSP47）（$P=0.041$）、ITGA6（$P=0.011$）、SPP1（$P=0.011$）、MMP1（$P=0.044$）的 mRNA 水平均明显高于正常组织（见图 6）。

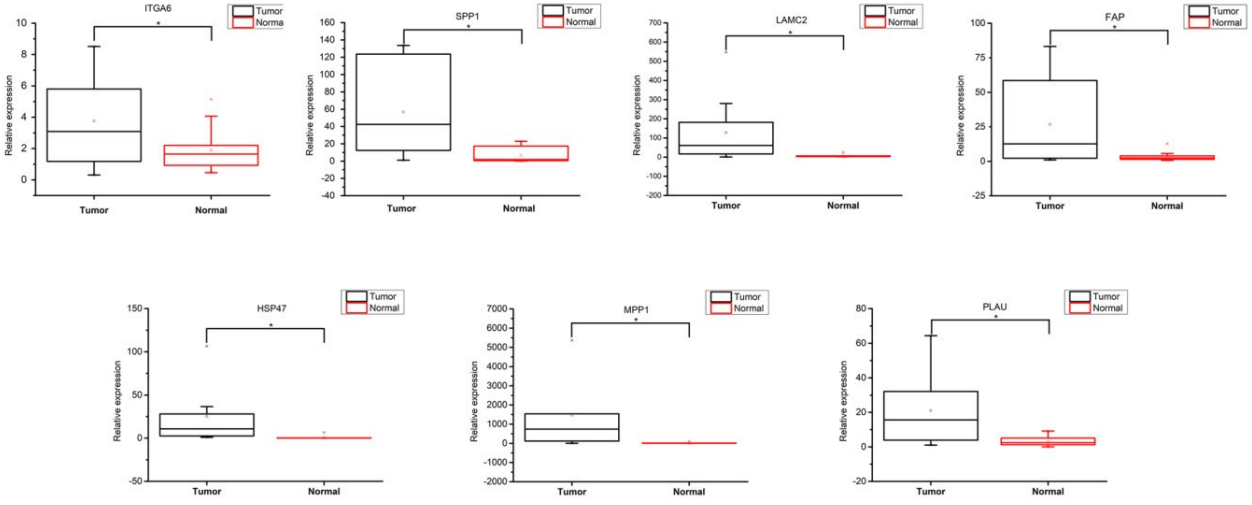
3　讨论

HNSCC 是第六大常见肿瘤，目前尚无有效的诊断和治疗方法来改善其预后。因此了解与 HNSCC 发生发展相关的分子机制是必要的。在过去的几十年里，大多数研究都集中在与 HNSCC 相关的单个基因上。本研究综合分析了与 HNSCC 相关的两个 GEO 数据集和 TCGA 测序数据。利用 R 包 edgeR，GEO2R 及 Draw Venn Diagram 在线工具，共筛选出 70 个重叠 DEGs，其中 34 个基因表达上调，36 个基因表达下调。将 DEGs 导入 STRING 数据库获取 PPI 网络数据，利用 Cytoscape 软件进行可视化，使用插件 CytoHubba 筛选出前 30 个候选关键基因。对候选关键基因 KEGG 通路富集分析结果显示，DEGs 在细胞外基质受体相互作用、黏着斑、PI3K-Akt 信号通路等



注：使用 GEPIA 对候选关键基因进行生存分析，其中 7 个基因与 HNSCC 患者总体生存期显著相关。

图 5　生存曲线图



注：* $P<0.05$。

图 6　在临床样本中关键基因的 mRNA 表达水平

方面显著富集。细胞外基质是由结构和功能大分子组成的复杂混合物，在组织和器官的形态形成以及细胞和组织结构功能的维持中起着重要的作用。同时细胞外基质也在各种肿瘤的发生发展中起着重要的作用。与胶原蛋白、层粘连蛋白和纤连蛋白相关的基因参与了细胞外基质受体的相互作用。许多研究表明，胶原蛋白家族在 HNSCC 中起重要作用。COLVI 被认为是一种潜在的诊断和治疗 HNSCC 的生物标志物。COL1A1 基因的下调抑制了 OSCC 细胞的增殖、侵袭和有丝分裂[15]。黏着斑是通过整合素将细胞骨架与细胞外基质连接起来的大型蛋白复合物。整合素是由 α 和 β 跨膜形成的异质二聚体，在细胞膜信号向细胞内部汇聚的过程中起主要作用。PI3K-Akt 信号通路影响多个靶蛋白的翻译和转录，这些靶蛋白涉及多种细胞特性，如增殖、转移等。

使用 GEPIA 进行生存分析，我们最终确定了 7 个关键基因 PLAU、FAP、LAMC2、SERPINH1、ITGA6、SPP1 和 MMP1 作为 HNSCC 的诊断及治疗潜在靶点和生物标志物。Teichgräber V 等[16]的研究结果表明，FAP 在上皮性肿瘤中过表达并促进肿瘤生长，而 FAP 的特异性抑制可在体外阻止肿瘤的进展。研究表明，FAP 可作为 HNSCC 潜在的 CAR-T 靶标[17]。

LAMC2 在口腔鳞状细胞癌(OSCC)组织中显著上调，且在黏膜白斑中 LAMC2 阳性比 LAMC2 阴性的恶变风险增加了约 11 倍[15]。MMP1 在 HNSCC 组织中过表达，抑制 MMP1 可降低 HNSCC 的侵袭能力[18]。OSCC 中 SPP1 的表达明显高于正常口腔黏膜，且 SPP1 可调节 OSCC 的增殖、迁移和侵袭[19]。PLAU 基因编码尿激酶纤溶酶原激活物(uPA)，其过表达增强 HNSCC 的增殖，迁移和侵袭能力[20]。ITGA6 是整合素家族的一员，已知它能介导与细胞外基质的相互作用，并增强多种细胞运动和信号输出能力。研究表明，ITGA6 的表达值与 HNSCC 患者的整体生存有关[21]。SERPINH1 属于丝氨酸蛋白酶抑制剂超家族，其在多种癌症中高度表达，并可驱动癌细胞的恶性行为。研究表明 SERPINH1 的上调与 HNSCC 患者较差的总体生存率有关[22]。

综上所述，本研究通过对多个数据集进行综合生物信息学分析，最后筛选出与 HNSCC 发生发展相关的关键基因。这将有助于提高我们对 HNSCC 病因及潜在分子机制的认识。为研发肿瘤靶向药物，开展个体化治疗提供依据。

参考文献：

[1] Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012[J]. Int J Cancer, 2015, 136(5): E359-386.

[2] Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma[J]. Oral Oncol, 2013, 49(3): 211-215.

[3] Chauhan SS, Kaur J, Kumar M, et al. Prediction of recurrence-free survival using a protein expression-based risk classifier for head and neck cancer[J]. Oncogenesis, 2015, 4(4): e147.

[4] Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer[J]. Nat Rev Cancer, 2011, 11(1): 9-22.

[5] Pulte D, Brenner H. Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis[J]. Oncologist, 2010, 15(9): 994-1001.

[6] Li DY, Hao XY, Song YS. Identification of the Key MicroRNAs and the miRNA-mRNA Regulatory Pathways in Prostate Cancer by Bioinformatics Methods[J]. Biomed Res Int, 2018, 2018: 6204128.

[7] Li H, Liu JW, Liu S, et al. Bioinformatics-Based Identification of Methylated-Differentially Expressed Genes and Related Pathways in Gastric Cancer[J]. Dig Dis Sci, 2017, 62(11): 3029-3039.

[8] Dong ZY, Wang JW, Zhang HQ, et al. Identification of potential key genes in esophageal adenocarcinoma using bioinformatics[J]. Exp Ther Med, 2019, 18(5): 3291-3298.

[9] Hui L, Yang N, Yang HJ, et al. Identification of biomarkers with a tumor stage-dependent expression and exploration of the mechanism involved in laryngeal squamous cell carcinoma[J]. Oncol Rep, 2015, 34(5): 2627-2635.

[10] Guo YC, Bao YH, Ma M, et al. Identification of Key Candidate Genes and Pathways in Colorectal Cancer by Integrated Bioinformatical Analysis[J]. Int J Mol Sci, 2017, 18(4): 722.

[11] Kumar R, Samal SK, Routray S, et al. Identification of oral cancer related candidate genes by integrating protein-protein interactions, gene ontology, pathway analysis and immunohistochemistry[J]. Sci Rep, 2017, 7(1): 3472.

[12] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists[J]. Nucleic Acids Res, 2009, 37(1): 1-13.

[13] Xie C, Mao XZ, Huang JJ, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases[J]. Nucleic Acids Res, 2011, 39: W316-322.

[14] Tang ZF, Li CW, Kang BX, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses[J]. Nucleic Acids Res, 2017, 45(W1): W98-W102.

[15] Nguyen CTK, Okamura T, Morita KI, et al. LAMC2 is a predictive marker for the malignant progression of leukoplakia[J]. J Oral Pathol Med, 2017, 46(3): 223-231.

[16] Teichgräber V, Monasterio C, Chaitanya K, et al. Specific inhibition of fibroblast activation protein (FAP)-alpha prevents tumor progression in vitro[J]. Adv Med Sci, 2015, 60(2): 264-272.

[17] Park YP, Jin LC, Bennett KB, et al. CD70 as a target for chimeric antigen receptor T cells in head and neck squamous cell carcinoma[J]. Oral Oncol, 2018, 78: 145-150.

[18] Kidacki M, Lehman HL, Green MV, et al. p120-Catenin Downregulation and PIK3CA Mutations Cooperate to Induce Invasion through MMP1 in HNSCC[J]. Mol Cancer Res, 2017, 15(10): 1398-1409.

[19] Zou B, Li J, Xu K, et al. Identification of key candidate genes and pathways in oral squamous cell carcinoma by integrated Bioinformatics analysis[J]. Exp Ther Med, 2019, 17(5): 4089-4099.

[20] Pavón MA, Arroyo-Solera I, Céspedes MV, et al. uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy[J]. Oncotarget, 2016, 7(35): 57351-57366.

[21] Zhao L, Chi WW, Cao H, et al. Screening and clinical significance of tumor markers in head and neck squamous cell carcinoma through bioinformatics analysis[J]. Mol Med Rep, 2019, 19(1): 143-154.

[22] Fan GY, Tu YQ, Wu N, et al. The expression profiles and prognostic values of HSPs family members in Head and neck cancer[J]. Cancer Cell Int, 2020, 20: 220.

收稿日期：2020-05-16；**修回日期：**2020-07-07