

基于TCGA数据库构建和评估口腔癌的免疫相关预后模型

巴颖,张核子,余晨笛,卢晓萍,操利超

(深圳市核子基因科技有限公司,广东 深圳 518071)

摘要:目的 构建和评估口腔癌免疫相关的预后模型,探索口腔癌肿瘤免疫微环境的分子特性。方法 基于TCGA数据库中HNSC队列的mRNA表达数据、临床信息和ImmPort数据库中的免疫相关基因列表,利用R包limma分析得到免疫相关的差异表达基因列表,通过单因子回归和多因子回归分析构建口腔癌的免疫相关预后模型,并进一步结合临床特征构建列线图模型,综合评估预后模型的性能。结果 共得到1533个差异表达基因,其中73个基因与免疫相关,通过单因子回归和多因子回归分析得到6个免疫相关的差异基因与预后相关;构建的预后模型ROC曲线下的面积在3年时为0.678,4年时为0.671,5年时为0.683;构建的列线图模型的C-index从0.63增加至0.67,基于建立的预后风险模型,高风险评分组的预后差于低风险评分组,差异有统计学意义($P<0.05$);预后风险评分与N分期相关($P<0.05$),与病理分期和T分期无关($P>0.05$);高风险评分组的基因MUC16突变率偏高,而NOTCH1则相反($P<0.05$),且高风险评分组的免疫治疗相关靶基因的表达水平偏低($P<0.05$);较高的TMB与较差的预后相关($P<0.05$)。结论 本次构建的预后模型和对肿瘤免疫微环境的分子特性的探索可能有助于口腔癌患者的预后风险预测和免疫治疗。

关键词:口腔癌;预后模型;肿瘤微环境;TCGA数据库;免疫治疗

中图分类号:R446

文献标识码:A

DOI:10.3969/j.issn.1006-1959.2022.06.004

文章编号:1006-1959(2022)06-0014-07

Construction and Evaluation of Immune Related Prognosis Model of Oral Cancer Based on TCGA Database

BA Ying,ZHANG He-zi,YU Chen-di,LU Xiao-ping,CAO Li-chao

(Shenzhen Nucleus Gene Technology Co., Ltd.,Shenzhen 518071,Guangdong,China)

Abstract: Objective To construct and evaluate an immune-related prognostic model for oral cancer and explore the molecular characteristics of tumor immune microenvironment. **Methods** Based on the mRNA expression data and clinical information of HNSC cohort in TCGA database and the list of immune-related genes in ImmPort database, the list of immune-related differentially expressed genes was obtained by R-package limma analysis. The immune-related prognosis model of oral cancer was constructed by single factor regression and multi-factor regression analysis, and the line graph model was further constructed by combining the clinical characteristics to comprehensively evaluate the performance of the prognosis model. **Results** A total of 1533 DEGs were obtained, of which 73 genes were immune related. Six immune-related DEGs were significantly correlated with prognosis by univariable Cox regression and multivariate Cox regression analysis ($P<0.05$). The area under the ROC curve of the constructed prognostic model was 0.678 at 3 years, 0.671 at 4 years and 0.683 at 5 years. The C-index of the constructed nomogram model increased from 0.63 to 0.67. Based on the established prognostic risk model, the prognosis of the high-risk score group was worse than that of the low-risk score group ($P<0.05$). The prognostic risk score was significantly correlated with N stage ($P<0.05$), but not correlated with pathological stage and T stage ($P>0.05$). In addition, the mutation rate of gene MUC16 in the high-risk score group was significantly higher, while NOTCH1 was the opposite ($P<0.05$), and the expression level of immunotherapy related target genes in the high-risk score group was significantly lower ($P<0.05$), and higher TMB was significantly associated with poor prognosis ($P<0.05$). **Conclusion** The constructed prognostic model and the exploration of the molecular characteristics of tumor immune microenvironment will help to predict the prognostic risk and immunotherapy of oral cancer patients.

Key words: Oral cancer; Prognostic model; Tumor microenvironment; TCGA database; Immunotherapy

口腔癌(oral cancer)是头颈部较常见的恶性肿瘤之一,已引起全球关注^[1]。在国际分类定义中,口腔癌是包括嘴唇、舌头、牙龈、口腔底部、颊粘膜、口腔前庭或磨牙后区等部位的癌症^[2]。尽管目前对口腔癌的诊断和治疗方面已取得了一定的进展,但口腔癌患者的死亡率依然很高,主要原因之一是缺乏有效的预后生物标志物^[3]。因此,探索口腔癌潜在的预后标志物和治疗靶点是当前亟需解决的问题。免疫检查点抑制剂(immune checkpoint inhibitors, ICIs)是一种很有前景的免疫疗法,通过抑制负性调

节受体发挥作用,如程序性细胞死亡受体1(programmed cell death protein 1, PD-1)和细胞毒性T淋巴细胞抗原4(cytotoxic T lymphocyte antigen 4, CTLA4),从而激活抗肿瘤免疫^[4]。免疫治疗在口腔癌中具有更好的疗效,有关口腔癌免疫学的研究不断增加^[5,6]。然而,由于肿瘤免疫微环境的异质性和复杂性,只有一小部分患者受益于免疫治疗,持续缓解的患者仍很少。许多免疫相关生物标志物被认为是口腔癌患者有用的预测指标,但目前尚缺乏一种理想的生物标志物用于临床^[7-10]。为了预测和改善口腔癌患者的预后,本研究基于癌症基因组图谱-头颈部鳞状细胞癌(the cancer genome atlas-head and neck squamous cell carcinoma, TCGA-HNSC)队列构建和评估了口腔癌免疫相关预后模型,分析该预后模型与口腔癌患者临床病理特征之间的关系,探索

作者简介:巴颖(1972.11-),女,辽宁辽阳人,硕士,主治医师,主要从事基因组学研究及其在人类疾病早筛早诊领域的应用

通讯作者:操利超(1984.4-),湖北蕲春县人,硕士,工程师,主要从事基于多组学测序数据和临床指标的结直肠癌预测诊断和预后模型的研究

肿瘤免疫微环境的特征,包括肿瘤浸润细胞构成、肿瘤突变负荷(tumor mutational burden,TMB)、以及 PD-1/PD-L1/CLTA4 的 mRNA 表达水平,现报道如下。

1 材料与方法

1.1 数据获取 癌症基因组图谱 (the cancer genome atlas,TCGA)中 mRNA 表达数据集和对应的临床信息从 UCSC Xena 平台([\[pages/\]\(https://xenabrowser.net/data-pages/\)\)下载,选择队列为癌症基因组数据共享系统 \(genomic data commons,GDC\)TCGA-HNSC,剔除非口腔癌相关部位的样本,样本信息见表 1。从免疫学数据库和分析平台 \(the immunology database and analysis portal,ImmPort\) \(<https://immport.niaid.nih.gov/>\)下载免疫相关基因列表,总计 1509 个。](https://xenabrowser.net/data-</p></div><div data-bbox=)

表 1 TCGA 数据集中样本信息[n(%)]

Item	Tumor samples (n=337)	Normal samples (n=32)	All samples (n=369)	Item	Tumor samples (n=337)	Normal samples (n=32)	All samples (n=369)
Sex				T Stage			
Femal	103(30.56)	12(37.50)	115(31.17)	T ₀	1(0.30)	0	1(0.27)
Male	234(69.43)	20(62.50)	254(68.83)	T ₁	31(9.20)	3(9.38)	34(9.21)
Age(years)				T ₂	102(30.26)	14(43.75)	116(31.43)
≤60	162(48.07)	14(43.75)	176(47.70)	T ₃	65(19.29)	7(21.87)	72(19.51)
>60	175(51.93)	18(56.25)	193(52.30)	T ₄	111(32.94)	8(25.00)	119(32.25)
Tumor stage				N/A	27(8.01)	0	27(7.32)
I	19(5.64)	2(6.25)	21(5.69)	N Stage			
II	55(16.32)	10(31.25)	65(17.62)	N ₀	121(35.91)	12(37.50)	133(36.04)
III	60(17.80)	8(25.00)	68(18.43)	N ₁	49(14.54)	6(18.75)	55(14.91)
IV	168(49.85)	11(34.38)	179(48.51)	N ₂	108(32.05)	6(18.75)	114(30.90)
N/A	35(10.39)	1(3.12)	36(9.75)	N ₃	3(0.89)	0	3(0.81)
M Stage				N/A	56(16.61)	8(25.00)	64(17.34)
M ₀	122(36.20)	1(3.12)	123(33.33)				
MX	42(12.46)	3(9.38)	45(12.20)				
N/A	173(51.34)	28(87.50)	201(54.47)				

1.2 构建免疫相关的预后风险模型 首先,利用 R 包对 TCGA 数据集进行差异基因分析,过滤标准为 adjusted P -value<0.05 和差异倍数大于 1.5 倍 ($|\log_2FC|>0.585$),然后与免疫基因列表取交集,得到免疫相关的差异基因。接着,使用 R 包 Survival 进行单变量 Cox 比例风险回归模型,选取与生存预后相关的差异基因。通过多因子回归分析确定每个预后因子的回归系数,建立预后风险评估模型,预测患者生存率。公式为:

风险分数= \sum 差异基因的回归系数 $\chi_i \times$ 归一化处理后的基因表达量 β_i

1.3 绘制生存曲线和 ROC 曲线 为了评估构建的风险预后模型的性能,利用 R 包 survivalROC 绘制受试者工作特征曲线 (receiver operating characteristic curve,ROC),在 ROC 曲线的转折点选择最佳风险评分临界值,转折点处真阳性和假阳性之间的差异最大。高于临界值的患者属于高风险评分组,低于临界值的患者属于低风险评分组,并使用 R 包 Survminer 绘制两组的生存曲线。

1.4 构建和验证列线图 为了提高预后模型的性能,

绘制列线图,使用 R 包 rms,通过整合风险评分模型和临床信息,包括年龄、性别和肿瘤分期,可视化不同患者特征的预后价值。此外,基于单因子回归分析的森林图分析临床信息与总生存期 (overall survival,OS)间的关系。其中,一致性指数 (concordance index,C-index)可反映列线图的预测准确性。

1.5 肿瘤免疫浸润细胞类型的构成比较 基于构建的预后风险模型,将肿瘤样本分为高风险评分组和低风险评分组,并采用 CIBERSORT 算法^[11],获取 22 种肿瘤浸润免疫细胞的比例。随后,通过非配对 t 检验比较高风险评分组和低风险评分组之间的免疫图谱。

1.6 肿瘤免疫微环境的特征分析 使用 R 包 maftools 可视化高风险评分组和低风险评分组的突变谱^[12]。计算 TMB 值,并通过非配对 t 检验统计高风险评分组和低风险评分组间 TMB 的差异;使用 Kaplan-Meier 方法比较高风险评分组和低风险评分组间 OS 的差异;采用 Wilcoxon 检验比较高风险评分组和低风险评分组间免疫检查点及其配体的 mRNA 表达水平。

2 结果

2.1 构建口腔癌的免疫相关预后模型 共得到 1533 个差异表达基因,其中上调差异基因数量为 726 个,下调基因数量为 807 个,见图 1A;与免疫基因列表取交集后,共得到 73 个免疫相关基因,其中 51 个基因下调,22 个基因上调。单因子回归和多因子回归分析表明,有 6 个免疫相关的差异基因与 OS 有关,见表 2。其中,回归系数见图 1B~图 1D。根据逐步回归模型,Akaike 信息标准(Akaike information criterion,AIC)为 1536.21,C-index 为 0.63,见图 1E。

2.2 预后模型的性能评估 基于已构建的预后风险模型,将口腔癌患者分为高风险评分组和低风险评分组,其中,cutoff 值设为 0.02。可以看出,随着风险得分

的增加,生存时间呈现缩短的趋势,并且高危组的死亡比例比低危组高,见图 2A、图 2B;基因 CD40LG 和 CMA1 在低风险评分组表达量高,在高风险评分组表达量低,而其余 3 个基因趋势相反,见图 2C;高风险评分患者的 OS 比低评分患者预后较差,见图 2D。

2.3 预后模型的统计分析 绘制 ROC 曲线和肿瘤分层分析进一步评估风险模型的性能,结果显示,ROC 曲线下面积(area under the curve,AUC)在 3 年时为 0.678,4 年时为 0.671,5 年时为 0.683,见图 3A。Wilcoxon 检验表明,预后风险评分与 N 分期(区域淋巴结转移)密切相关($P=0.034$),但与病理分期和 T 分期无相关性($P>0.05$),见图 3B、图 3C、图 3D。

表 2 与口腔癌预后相关的基因信息

Gene name	HR	HR.95L	HR.95H	P-value
ADIPOQ	1.07483216	1.010341229	1.143439602	0.022263319
ESRRG	1.12179433	1.045575551	1.20356919	0.001367562
CMA1	0.894853224	0.822032095	0.974125337	0.010308677
CD40LG	0.919577325	0.846045612	0.999499844	0.048640516
FGF9	1.100937584	1.006118666	1.204692453	0.036374619

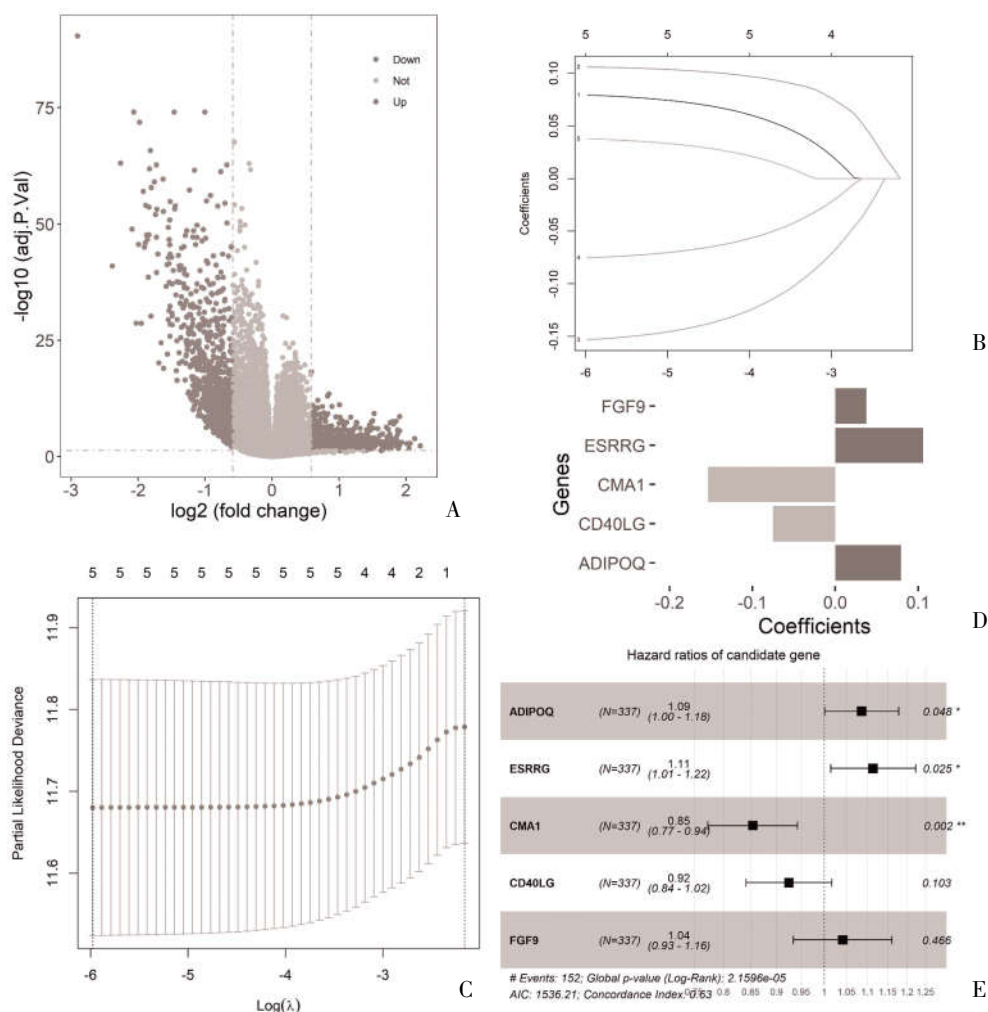


图 1 鉴定预后相关的差异表达基因

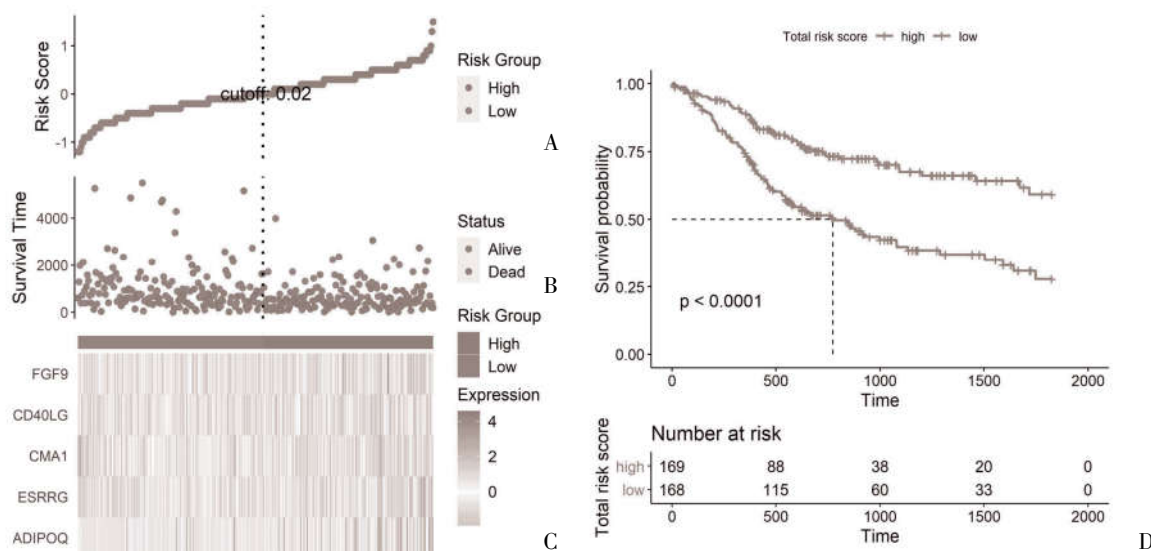


图 2 预后风险评分分组和评估

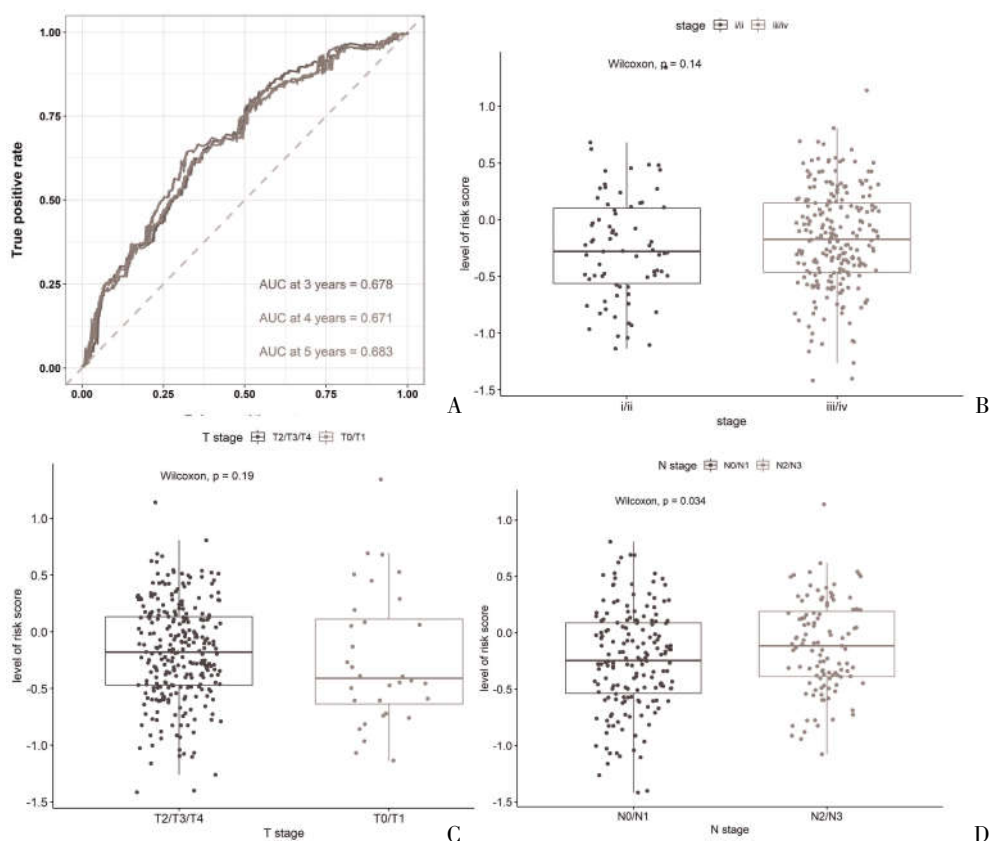


图 3 预后风险评估模型性能评价

2.4 构建和评估列线图模型 在列线图中,每个变量的得分可以在分数表上找到,然后通过计算总分来估计 3、4 和 5 年的生存概率,见图 4A;森林图显示,患者年龄(>60 岁)、肿瘤Ⅳ期和风险评分与 OS 相关 ($P < 0.05$),见图 4B。另外,绘制校准曲线来验证列线图的性能,结果显示,构建的列线图模型性能良好,预测曲线接近理想曲线,见图 4C~图 4E。此外,该列线图(C-index:0.67)的预测准确性高于风险评分模型(C-index:0.63)。

2.5 体细胞突变统计及其与口腔癌生存预后的关系 基于建立的免疫预后模型,将肿瘤样本分为高风险评分组和低风险评分组,分别统计体细胞突变类型、突变模式和基因突变率。结果显示,突变以错义突变为主,主要突变类型为单核苷酸变异(single nucleotide variants, SNV),突变模式主要为 C>T。在高风险评分组中,突变率排名前 10 的基因为 TP53、TTN、MUC16、FAT1、CDKN2A、KMT2D、LRP1B、NOTCH1、DNAH5 和 SYNE1,在低风险评分组中,突

变率排名前 10 的基因为 TP53、TTN、FAT1、NOTCH1、CDKN2A、PIK3CA、CASP8、SYNE1、CSMD3 和 LRP1B, 见图 5A, 图 5B; 其中基因 MUC16 和 NOTCH1 在高风险评分组和低风险评分组中的突变率比较, 差异有统计学意义($P<0.05$), 见图 5C; 对基因 MUC16 和 NOTCH1 分别进行生存曲

线分析, 两个基因均不能作为独立的预后因子($P>0.05$), 见图 5D, 图 5E; 高风险评分组和低风险评分组的 TMB 比较, 差异无统计学意义($P>0.05$), 见图 5F; 但生存曲线分析显示, TMB 值可作为独立的预后因子, 见图 5G。

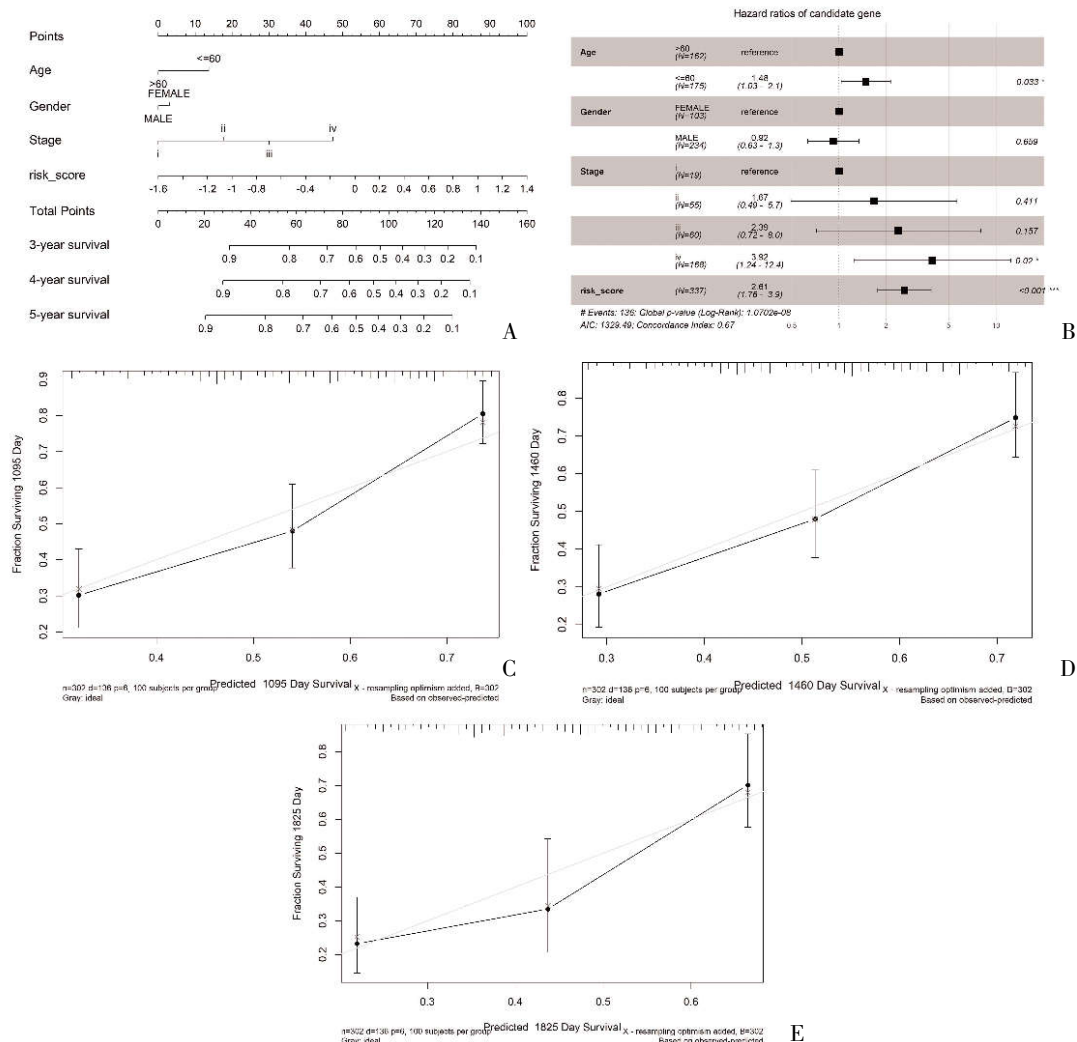
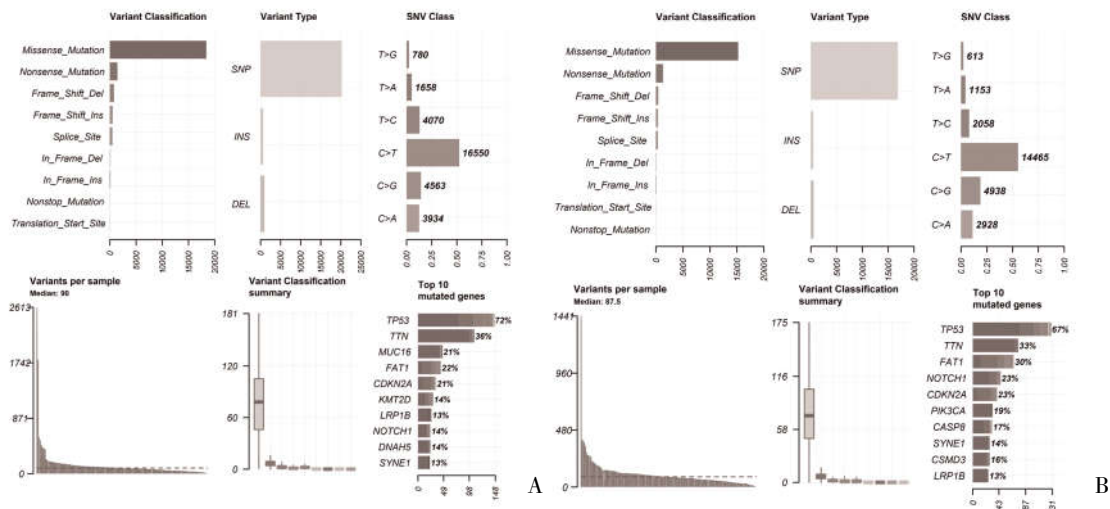


图 4 列线图模型的构建与验证



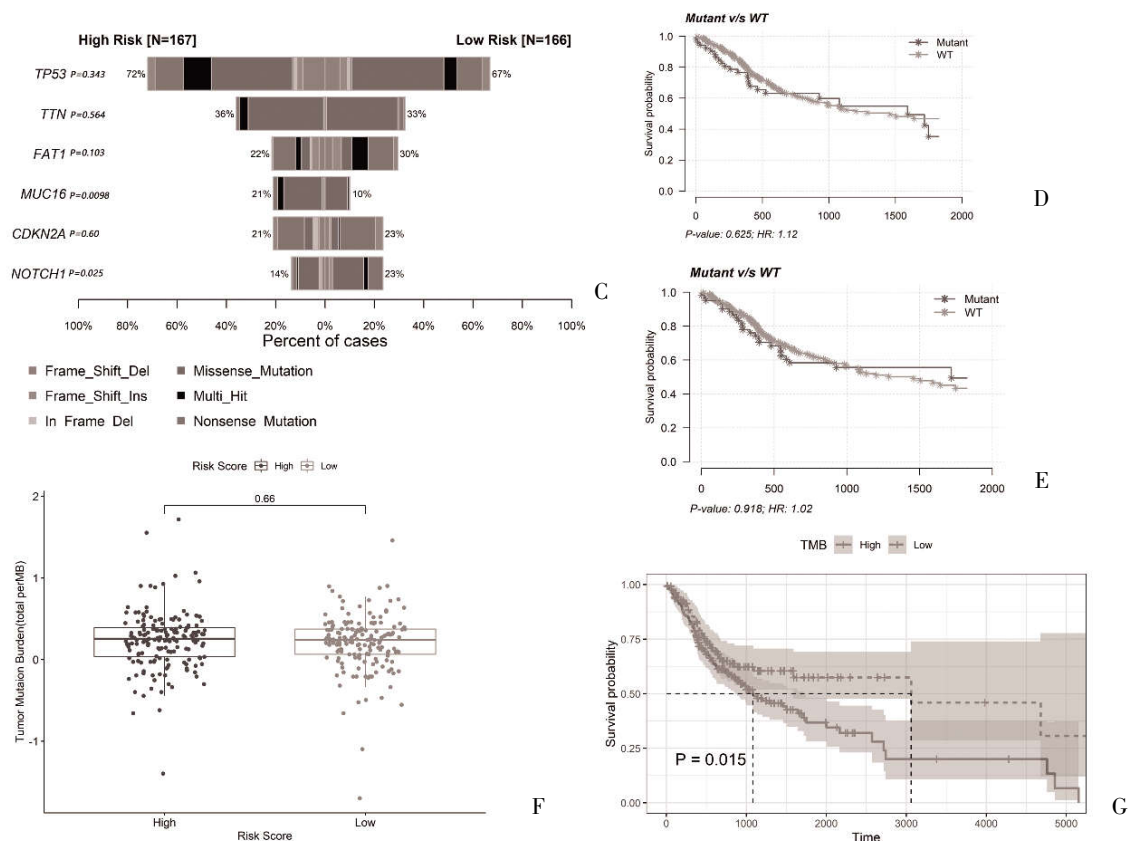


图 5 体细胞突变分析统计及其与 OS 的关系(续)

2.6 口腔癌肿瘤免疫微环境 基于 CIBERSORT 算法,估算每个口腔癌患者中 22 种免疫细胞的比例,并比较高风险评分组和低风险评分组之间的免疫细胞比例,结果显示,有 6 种免疫细胞类型存在差异 ($P<0.05$),见图 6A;另外,虽然免疫细胞 NK.cells.

activated 在高风险评分组和低风险评分组的免疫浸润程度不存在差异,但其与生存预后相关($P<0.05$),见图 6B。Wilcoxon 检验结果表明,高风险评分组 PD-L1 ($P=1.8\times 10^{-7}$)、PD-1 ($P=0.0079$) 和 CTLA-4 ($P=9.7\times 10^{-7}$)的表达水平较低,见图 6C~图 6E。

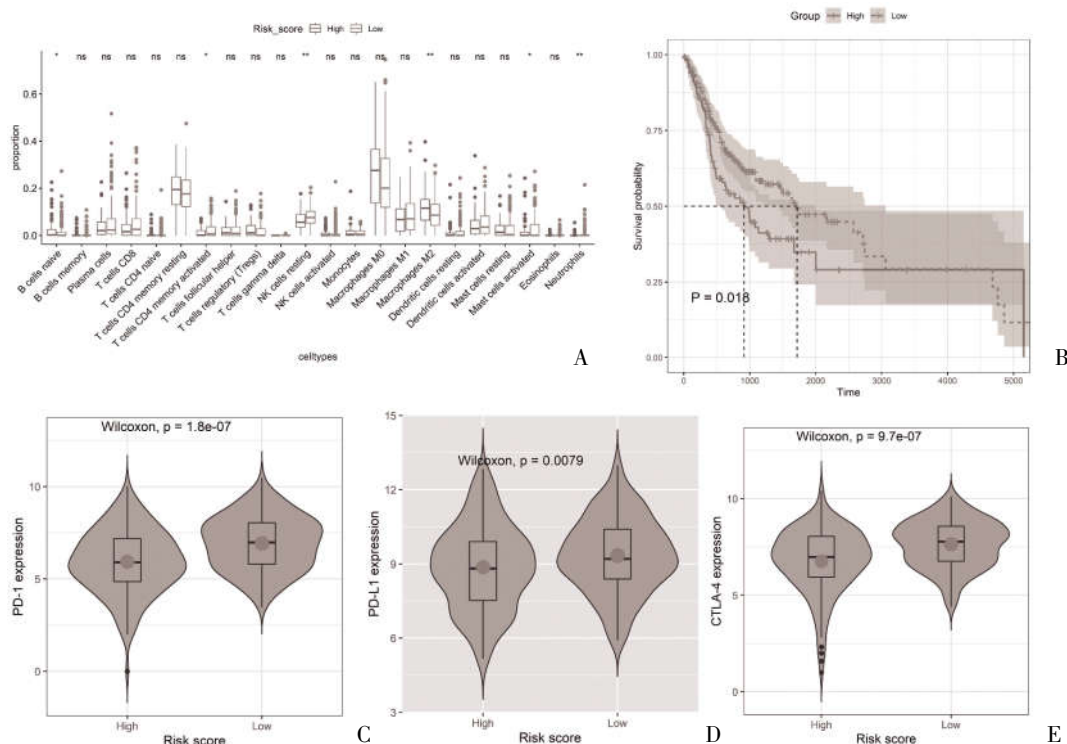


图 6 口腔癌肿瘤免疫微环境

3 讨论

免疫细胞在肿瘤的演变中起着重要作用^[4],目前 ICIs 已应用于中晚期实体瘤的治疗^[13]。对于口腔癌,纳武利尤单抗(nivolumab)和派姆单抗(pembrolizumab)均被批准用于治疗复发/转移性疾病^[14]。然而,由于对肿瘤微环境特性的理解不足,这在一定程度上阻碍了免疫治疗的广泛应用。近年来,大量研究鉴定出了与口腔癌诊断和预后相关的免疫相关生物标志物^[8,9,15,16]。但为了最大限度地发挥免疫治疗的作用,还需要探索更多可靠的生物标志物。

本研究基于 5 个免疫相关差异基因构建了口腔癌预后风险评估模型,这 5 个基因为 FGF9、CD40LG、CMA1、ESRRG、ADIPOQ。已有研究表明^[7,17,18],CD40LG、CMA1 和 ESRRG 与口腔癌的预后显著相关。FGF9 是一种生长因子,属于 FGFs 家族,FGF9 是许多癌症(包括肺腺癌、前列腺癌和肝癌)肿瘤演化的关键因子^[19]。CD40LG 是一种跨膜蛋白,属于肿瘤坏死因子超家族,存在于 CD4⁺T 细胞、B 细胞、单核细胞和自然杀伤细胞及癌细胞中,具有增强肿瘤细胞的凋亡和抗原呈递细胞的功能^[17]。CMA1 是一种乳糜蛋白酶,由肥大细胞分泌,在癌症中还没有明确的研究。有研究表明^[20],通过增强 ESRRG 的表达和功能,可在癌细胞中发挥抑癌活性。ADIPOQ 表达的脂联素可显著抑制乳腺癌生长并诱导细胞凋亡^[21],但其在口腔癌中的分子机理还未见报道。

本研究结果显示,免疫细胞 NK cells、activated 的免疫浸润程度和 TMB 是口腔癌潜在的独立预后标志。口腔癌高风险评分患者的免疫检查点及其配体(PD-L1、PD-1 和 CTLA-4)表达水平较低,这可能表明构建的免疫相关风险评估模型能够为免疫治疗提供一定的临床指导意义。

综上所述,本研究成功构建了免疫相关的口腔癌风险评估预后模型,并探索了肿瘤免疫微环境的特征,这可能有助于口腔癌患者的预后和免疫治疗。

参考文献:

[1]Wong T,Wiesenfeld D.Oral Cancer [J].Aust Dent J,2018,63(Suppl1):S91-S99.
[2]Bawaskar HS,Bawaskar PH.Oral diseases: a global public health challenge[J].Lancet,2020,395(10219):185-186.
[3]Khurshid Z,Zafar MS,Khan RS,et al.Role of Salivary Biomarkers in Oral Cancer Detection [J].Adv Clin Chem,2018(86):23-70.
[4]Lei X,Lei Y,Li JK,et al.Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy[J].Cancer Lett,2020(470):126-133.
[5]Kondoh N,Mizuno -Kamiya M,Umemura N,et al.Immunomodulatory aspects in the progression and treatment of oral malignancy[J].Jpn Dent Sci Rev,2019,55(1):113-120.
[6]Dorta -Estremera S,Hegde VL,Slay RB,et al.Targeting interferon signaling and CTLA-4 enhance the therapeutic efficacy of anti-PD-1 immunotherapy in preclinical model of HPV⁺ oral

cancer[J].J Immunother Cancer,2019,7(1):252.
[7]Huang GZ,Wu QQ,Zheng ZN,et al.Identification of Candidate Biomarkers and Analysis of Prognostic Values in Oral Squamous Cell Carcinoma[J].Front Oncol,2019(9):1054.
[8]Zhao XT,Zhu Y,Zhou JF,et al.Development of a novel 7 immune-related genes prognostic model for oral cancer: A study based on TCGA database[J].Oral Oncol,2021(112):105088.
[9]Sasahira T,Kirita T.Hallmarks of Cancer -Related Newly Prognostic Factors of Oral Squamous Cell Carcinoma [J].Int J Mol Sci,2018,19(8):2413.
[10]Wang J,Wang Y,Kong F,et al.Identification of a six-gene prognostic signature for oral squamous cell carcinoma [J].J Cell Physiol,2020,235(3):3056-3068.
[11]Chen B,Khodadoust MS,Liu CL,et al.Profiling Tumor Infiltrating Immune Cells with CIBERSORT[J].Methods Mol Biol,2018(1711):243-259.
[12]Mayakonda A,Lin DC,Assenov Y,et al.Maftools:efficient and comprehensive analysis of somatic variants in cancer[J].Genome Res,2018,28(11):1747-1756.
[13]Bagchi S,Yuan R,Engleman EG.Immune Checkpoint Inhibitors for the Treatment of Cancer: Clinical Impact and Mechanisms of Response and Resistance [J].Annu Rev Pathol,2021(16):223-249.
[14]Doumas S,Foukas PG,Economopoulou P,et al.Atypical patterns of responses in the era of immune checkpoint inhibitors in head and neck cancer[J].Oral Oncol,2020(100):104477.
[15]Yao Y,Yan Z,Lian S,et al.Prognostic value of novel immune-related genomic biomarkers identified in head and neck squamous cell carcinoma [J].J Immunother Cancer,2020,8(2):e000444.
[16]Hadler -Olsen E,Wirsing AM.Tissue -infiltrating immune cells as prognostic markers in oral squamous cell carcinoma: a systematic review and meta-analysis[J].Br J Cancer,2019,120(7):714-727.
[17]Rah YC,Ahn JC,Jeon EH,et al.Low expression of CD40L in tumor-free lymph node of oral cavity cancer related with poor prognosis[J].Int J Clin Oncol,2018,23(5):851-859.
[18]Azad AK,Bairati I,Qiu X,et al.A genome-wide association study of non-HPV-related head and neck squamous cell carcinoma identifies prognostic genetic sequence variants in the MAP-kinase and hormone pathways [J].Cancer Epidemiol,2016(42):173-180.
[19]Wang R,Sun Y,Yu W,et al.Downregulation of miRNA -214 in cancer-associated fibroblasts contributes to migration and invasion of gastric cancer cells through targeting FGF9 and inducing EMT[J].J Exp Clin Cancer Res,2019,38(1):20.
[20]Peng Z,Guan Q,Luo J,et al.Sophoridine exerts tumor-suppressive activities via promoting ESRRG-mediated β -catenin degradation in gastric cancer[J].BMC Cancer,2020,20(1):582.
[21]Chung SJ,Nagaraju GP,Nagalingam A,et al.ADIPOQ/adiponectin induces cytotoxic autophagy in breast cancer cells through STK11/LKB1-mediated activation of the AMPK -ULK1 axis[J].Autophagy,2017,13(8):1386-1403.

收稿日期:2021-11-22;修回日期:2021-12-02

编辑/成森