

• 论著 •

文章编号: 1007-8738(2023)11-0988-08

基于机器学习鼻咽癌关键特征基因筛选及其与免疫细胞相关性分析

张浩轩^{1,2}, 马俊杰³, 安韶光³, 荀丽雪¹, 陆进^{1,2}, 蒋成义^{4*} (蚌埠医学院: ¹人体解剖学教研室, ²数字医学与智慧健康安徽省重点实验室, ³临床医学院 2020 级, ⁴第一附属医院耳鼻喉科, 安徽 蚌埠 233030)

[摘要] 目的 利用机器学习法筛选鼻咽癌(NPC)关键特征基因并分析其与免疫细胞相关性。方法 从基因表达数据集(GEO)下载 NPC 训练集数据 GSE12452 与 GSE13597 以及验证训练集数据 GSE53819。首先,对训练集数据进行合并,并筛选差异表达基因(DEG);其次,对 DEG 进行基因本体论(GO)、京都基因与基因组百科全书(KEGG)、基因集富集分析(GSEA)以及免疫细胞浸润分析;再次,采用最小绝对收缩选择(LASSO)算法和支持向量机(SVM)算法对训练集数据中 NPC 相关特征基因进行识别并在验证集中检验,同时利用受试者工作特征(ROC)曲线下面积(AUC)确定关键特征基因;最后,分析关键特征基因与免疫细胞的相关性。结果 共得到 55 个 DEG,43 个下调基因,12 个上调基因;其 GO 功能主要富集在体液免疫反应、细胞分化、中性粒细胞激活以及趋化因子受体结合等方面;而 KEGG 主要富集在细胞介素 17(IL-17)信号通路;GSEA 富集在细胞周期、细胞外基质受体相互作用、癌症通路以及 DNA 复制。免疫细胞浸润分析显示,初始 B 细胞、记忆 B 细胞以及 CD4⁺ 静息记忆细胞在 NPC 显著降低,而 CD8⁺ T 细胞、CD4⁺ 初始 T 细胞、活化 CD4⁺ 记忆 T 细胞、滤泡辅助 T 细胞、M0 和 M1 巨噬细胞在 NPC 显著增加。通过 LASSO 和 SVM 筛选的特征基因中,仅卷曲螺旋结构域 19(CCDC19)、层连蛋白 $\beta 1$ 亚基(LAMB1)、精子相关抗原 6(SPAG6)和 RAD51 相关蛋白 1(RAD51API)四个关键特征基因 ROC 的 AUC 在训练集与验证集均大于 0.9,且与免疫细胞浸润密切相关。结论 通过机器学习算法筛选出 NPC 发生过程中的关键特征基因 CCDC19、LAMB1、SPAG6 以及 RAD51API,并与免疫细胞浸润密切相关。

[关键词] 鼻咽癌;机器学习;免疫细胞浸润;支持向量机

[中图分类号] R766.3, R392.2 **[文献标志码]** A

DOI:10.13423/j.cnki.cjcmi.009669

Screen of key characteristic genes of nasopharyngeal carcinoma (NPC) base on machine learning and analysis of their correlation with immune cells

ZHANG Haoxuan^{1,2}, MA Junjie³, AN Shaoguang³, XU Lixue¹, LU Jin^{1,2}, JIANG Chengyi^{4*}

¹Department of Human Anatomy, ²Anhui Key Laboratory of Computational Medicine and Intelligent Health, ³Grade 2020 of Clinical Medical College, ⁴Otolaryngology department of First affiliated hospital, Bengbu Medical College, Bengbu 233030, China

* Corresponding author, E-mail: jiangchengyi1975@163.com

[Abstract] **Objective** Machine learning was used to screen the key characteristic genes of nasopharyngeal carcinoma (NPC) and analyze their correlation with immune cells. **Methods** Download the NPC training datasets (GSE12452 and GSE13597) and the validation dataset (GSE53819) from the Gene Expression Omnibus (GEO). Firstly, the training data sets were merged and screened for differentially expressed genes (DEGs); Secondly, the DEGs were analyzed by gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), gene set enrichment analysis (GSEA), and immune cell infiltration analysis. Next, the least absolute shrinkage and selection operator (LASSO) and support vector machine (SVM) algorithms were used to identify NPC-related genes in the training datasets and examined in the validation dataset, to further identify key genes using the area under curve (AUC) of receiver operating characteristic curve (ROC); Finally, the correlation between the key genes and immune cells was analyzed. **Results** A total of 55 DEGs were obtained, including 43 down-regulated genes and 12 up-regulated genes. The GO functions were enriched in humoral immune response, cell differentiation, neutrophil

收稿日期: 2022-05-31; 接受日期: 2023-07-07

基金项目: 安徽省教育厅自然科学研究重点项目(KJ2020A0553); 2022 年国家级大学生创新训练项目(202210367002); 安徽省重点研究与开发计划人口健康专项(202104j00005)

作者简介: 张浩轩(1980-),男,江苏淮安人,副教授,硕士

Tel: 13965285216; E-mail: zhxww@foxmail.com

* 通讯作者,蒋成义, E-mail: jiangchengyi1975@163.com

activation and chemokine receptor binding. The KEGG were mainly enriched in the IL-17 signaling pathway. The GSEA was enriched in cell cycle , extracellular matrix receptor interactions , cancer pathways and DNA replication. Immune infiltration analysis showed that the expression of naive B cells , memory B cells , and resting memory CD4⁺ T cells was significantly lower in NPC , while CD8⁺ T cells , naive CD4⁺ T cells , activated memory CD4⁺ T cells , follicular helper T cells , M0 macrophages and M1 macrophages were highly expressed in NPC. Among the feature genes screened by LASSO and SVM , only CCDC19 , LAMB1 , SPAG6 and RAD51AP1 genes' AUC were greater than 0.9 in both the training and validation datasets and were closely associated with immune cell infiltration. **Conclusion** The key genes CCDC19 , LAMB1 , SPAG6 and RAD51AP1 in NPC development are screened by machine learning algorithms , and are closely associated with immune cell infiltration.

[Key words] nasopharyngeal carcinoma(NPC) ; machine learning; immune cell infiltration; support vector machine(SVM)

鼻咽癌(nasopharyngeal carcinoma , NPC) 是常见的头颈部恶性肿瘤,好发于鼻咽顶与侧壁,发病率位居耳鼻咽喉部恶性肿瘤的首位^[1]。中国是 NPC 发病率最高的国家,年发病率占全球的 60% 以上,且以华南地区为主^[2]。目前,NPC 早期临床症状不典型,约 10% 的患者首次诊断已发生远处转移,导致其治疗失败^[3]。然而,随着医学诊疗技术的进步,NPC 患者的 5 年生存率在逐步提高,但由于肿瘤细胞的增殖与扩散,导致部分患者预后不良^[4]。NPC 患者死亡的主要原因是肿瘤细胞复发和转移,且局部肿瘤组织切除术移除的 NPC 患者仍然具有较高的转移和复发风险,并对放化疗不敏感^[5-6]。因此,研究 NPC 发生发展过程的关键基因有助于其早期诊断与治疗。

目前,基于生物信息学与机器学习算法的疾病风险预测成为研究热点^[7-8]。生物信息学的快速发展

有助于利用公共数据筛选疾病发生发展的特征基因,而机器学习方法在医学大数据领域的深入应用,使得疾病风险预测能力得到大幅提高。然而,目前尚没有研究同时使用最小绝对收缩选择算法(least absolute shrinkage and selection operator ,LASSO) 和支持向量机(support vector machine ,SVM) 识别 NPC 的生物标志物。因此,通过两种机器学习算法识别 NPC 诊断性生物标志物,对 NPC 患者的早期预防、早期诊断和早期治疗以及提高生存率具有重要临床意义。

1 材料和方法

1.1 材料

从基因表达数据集(Gene Expression Omnibus , GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) 下载 3 个 NPC 相关芯片数据,其中 GSE12452 和 GSE13597 用作训练集,GSE53819 用作验证集(表 1) 。

表 1 NPC 芯片数据信息

芯片名称	正常样本	肿瘤样本	平台信息
GSE12452	10	31	
GPL570			HG-U133_Plus_2]Affymetrix Human Genome U133 Plus 2.0 Array
GSE13597	3	25	GPL96 [HG-U133A]Affymetrix Human Genome U133A Array
GSE53819	18	18	GPL6480 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F

1.2 方法

1.2.1 数据分析方法 对 3 个 NPC 相关芯片数据集进行深入分析。首先,对数据进行合并处理,进行基因集富集分析(gene set enrichment analysis ,GSEA) 以及免疫细胞浸润分析; 其次,筛选差异表达基因(differentially expressed genes ,DEG) ,对 DEG 进行基因本体论(gene ontology ,GO) 、京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes ,KEGG) 分析; 再次,采用 LASSO 和 SVM 算法对训练集中 NPC 相关特征基因进行识别并在验证集中检验; 最后,分析关键特征基因与免疫细胞的相关性(图 1) 。

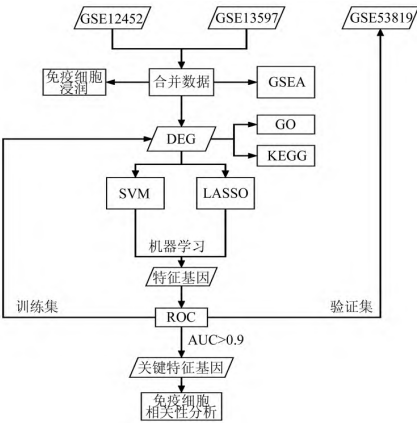


图 1 数据分析流程图

1.2.2 数据合并 利用 R 语言“limma”包和“sva”包与 Perl 语言对 GSE12452 与 GSE13597 数据进行合并处理,同时对合并后的数据进行 DEG 分析 ($|\log FC| > 2$ 且 $P < 0.05$) 以及热图与火山图绘制。

1.2.3 GO、KEGG 与 GSEA 分析 利用 R 语言相关数据分析包,对 DEG 进行 GO 和 KEGG 分析,并对合并后的数据进行 GSEA 分析(“limma”包、“DOSE”包、“clusterProfiler”包和“enrichplot”包)。

1.2.4 免疫细胞浸润分析 运用 R 语言免疫相关软件包“CIBERSORT”、“corrpilot”包、“vioplot”包)分析 GSE12452 与 GSE13597 合并后数据中正常鼻咽组织与 NPC 组织中的免疫细胞浸润情况及其相关性。

1.2.5 机器学习算法筛选特征基因 运用 R 语言相关软件包,通过 LASSO (“glmnet”包)和 SVM (“caret”包)算法从 DEG 表达谱筛选特征基因,并进行韦恩图绘制,得到两者共同的特征基因并进行深入分析。

1.2.6 特征基因分析 运用 R 语言相关软件包,对筛选得到的特征基因在训练集中进行受试者工作特征(receiver operating characteristic, ROC)曲线分析,且利用测试集数据对特征基因的 ROC 曲线“pROC”包)验证并进行表达水平分析(“ggpubr”包),选择曲线下面积(area under curve, AUC)值大于 0.9 的基因作为 NPC 的关键基因。同时,对关键特征基因在

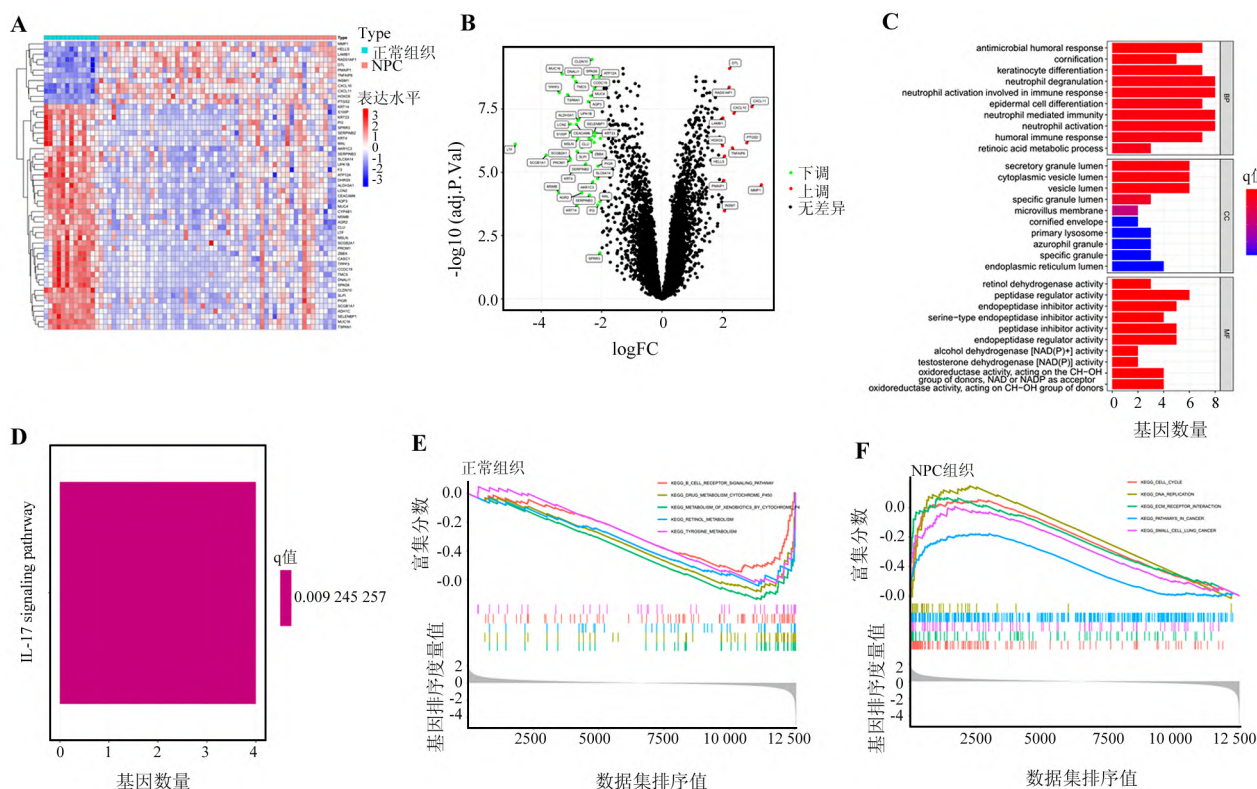
NPC 中与免疫细胞的相关性进行分析(“reshape2”包、“ggpubr”包以及“ggExtra”包)。

1.2.7 统计学分析 利用 R4.1.1 版本对所下载的数据进行分析处理, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 DEG 的筛选与分析

通过 GSE12452 与 GSE13597 数据合并及 DEG 分析,绘制 DEG 热图(图 2A)与火山图(图 2B),共得到 55 个差异基因,其中 43 个下调基因,12 个上调基因。富集分析显示,DEG 的 GO 功能中生物过程(biological process, BP)主要富集在体液反应、细胞分化以及参与免疫反应的中性粒细胞激活等;细胞组分(cellular component, CC)主要位于分泌颗粒内腔、胞质囊腔、特定颗粒内腔以及微绒毛膜;分子功能(molecular function, MF)主要富集在视黄醇脱氢酶活性、肽链内切酶抑制剂活性、醇脱氢酶 [NAD(P)⁺]活性、酶抑制剂的活动、趋化因子受体结合以及铁离子结合等(图 2C);而 DEG 的 KEGG 主要富集在白细胞介素 17(interleukin 17, IL-17)信号通路(图 2D)。GSEA 富集在细胞周期、细胞外基质(extracellular matrix, ECM)受体相互作用、癌症通路以及 DNA 复制上(图 2E、F)。



A: DEG 的热图; B: DEG 的火山图; C: DEG 的 GO 功能分析条形图; D: DEG 的 KEGG 分析条形图; E、F: GSEA 富集分析。

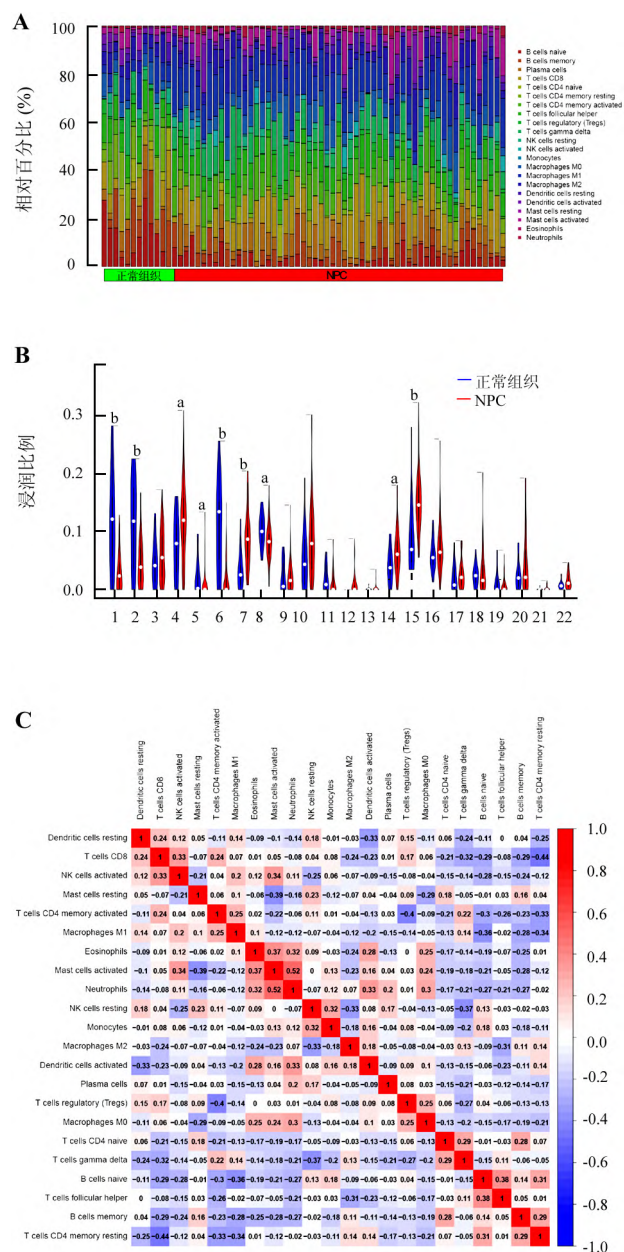
图 2 DEG 的筛选与生物信息学分析

2.3 免疫细胞浸润分析

通过对训练集数据进行免疫细胞浸润分析发现, 初始 B 细胞、记忆 B 细胞以及 $CD4^+$ 静息记忆细胞浸润水平在 NPC 显著降低 ($P < 0.05$), 而 $CD8^+$ T 细胞、 $CD4^+$ 初始 T 细胞、活化 $CD4^+$ 记忆 T 细胞、滤泡辅助 T 细胞、M0 和 M1 巨噬细胞在 NPC 的浸润水平显著增加 ($P < 0.05$ 或 $P < 0.01$, 图 3A、B)。免疫细胞相关性分析发现, $CD4^+$ 静息记忆细胞与 $CD8^+$ T 细胞呈负相关 ($r = -0.44$), 活化 $CD4^+$ 记忆 T 细胞与调节性 T 细胞 (T cells regulatory, Treg) 负相关 ($r = -0.4$); 而中性粒细胞与活化的肥大细胞正相关性最强 ($r = 0.52$, 图 3C)。

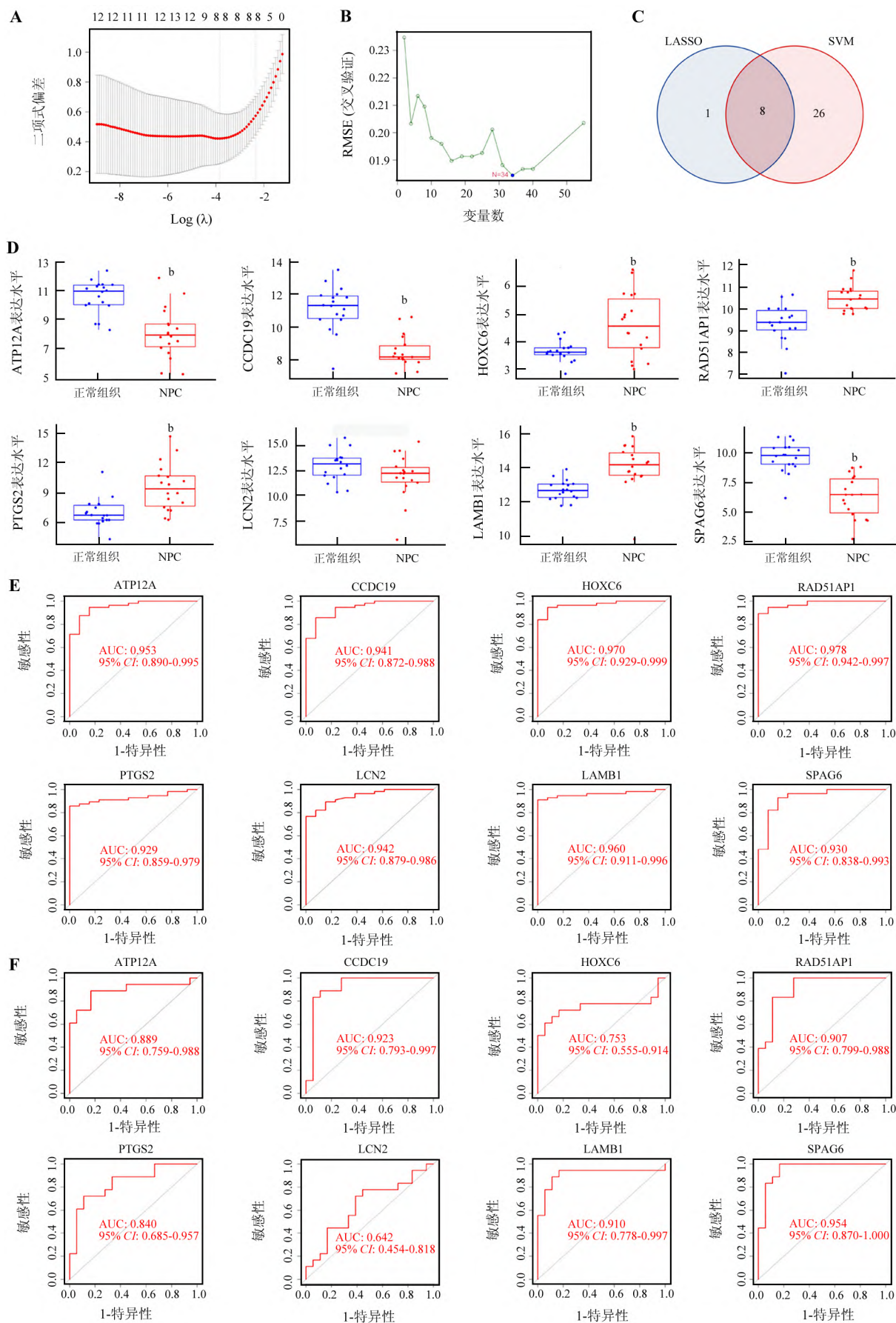
2.4 特征基因筛选和验证

通过 LASSO 算法筛选得到 9 个特征基因 (图 4A), 而 SVM 算法筛选共得到 34 个特征基因 (图 4B)。通过绘制韦恩图, 最后得到两者共同的特征基因 8 个, 分别为非胃型 H^+/K^+ ATP 酶 $\alpha 2$ 亚基 (ATPase H^+/K^+ non-gastric alpha2 subunit, ATP12A)、精子相关抗原 6 (sperm associated antigen 6, SPAG6)、RAD51 相关蛋白 1 (RAD51 associated protein 1, RAD51AP1)、卷曲螺旋结构域 19 (coiled-coil domain containing 19, CCDC19)、层连蛋白 $\beta 1$ 亚基 (laminin subunit beta 1, LAMB1)、脂质运载蛋白 2 (lipocalin 2, LCN2)、前列腺素内过氧化物合成酶 2 (prostaglandin-endoperoxide synthase 2, PTGS2) 和同源盒 C6 (homeobox C6, HOXC6) 基因 (图 4C)。利用测试集数据分析特征基因在 NPC 的表达水平, 结果显示 ATP12A、CCDC19 和 SPAG6 在 NPC 表达显著降低, 差异有统计学意义 ($P < 0.05$), 而 LCN2 在 NPC 虽然表达降低, 但差异无统计学意义 ($P > 0.05$); RAD51AP1、LAMB1、PTGS2 和 HOXC6 在 NPC 显著高表达 ($P < 0.05$ 或 $P < 0.01$, 图 4D)。特征基因在训练集的 ROC 曲线分析显示, AUC 均大于 0.9 (图 4E), 而在测试集中, 结果显示仅有 CCDC19、LAMB1、SPAG6 和 RAD51AP1 基因的 AUC 大于 0.9 (图 4F)。



A: 免疫细胞浸润比例直方图; B: 免疫细胞浸润比例小提琴图; C: 免疫细胞浸润相关性分析。1: 初始 B 细胞; 2: 记忆 B 细胞; 3: 浆细胞; 4: $CD8^+$ T 细胞; 5: 初始 $CD4^+$ T 细胞; 6: 静息记忆 $CD4^+$ T 细胞; 7: 激发的记忆 $CD4^+$ T 细胞; 8: 滤泡辅助性 T 细胞; 9: 调节 T 细胞; 10: $\gamma\delta$ T 细胞; 11: 静息 NK 细胞; 12: 激发的 NK 细胞; 13: 单核细胞; 14: M0 巨噬细胞; 15: M1 巨噬细胞; 16: M2 巨噬细胞; 17: 静息树突状细胞; 18: 激发的树突状细胞; 19: 静息肥大细胞; 20: 激发的肥大细胞; 21: 嗜酸性粒细胞; 22: 中性粒细胞。^a $P < 0.05$, ^b $P < 0.01$ vs 正常组织。

图 3 免疫细胞浸润和免疫细胞相关性分析



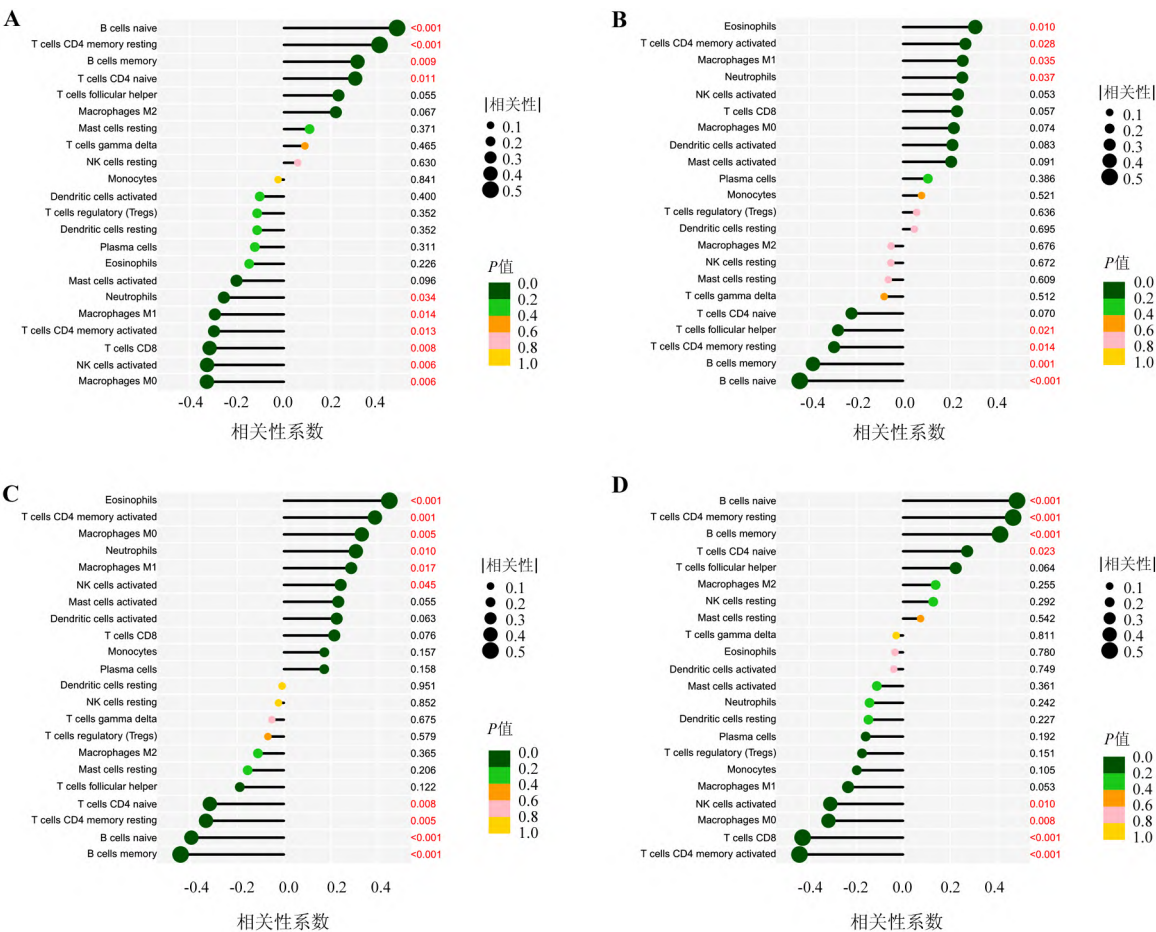
A: LASSO; B: SVM; C: 韦恩图; D: 特征基因在 NPC 组织的表达水平; E: 训练集; F: 验证集. ^b $P < 0.01$ vs 正常组织.

图4 特征基因筛选及在 NPC 组织的表达及关键基因的 ROC 曲线

2.5 关键特征基因与免疫细胞相关性分析

对在训练集与验证集中 AUC 均大于 0.9 的特征基因作为关键特征基因进行深入分析,研究其与免疫细胞相关性。结果显示,仅有 4 个关键特征基因,分别是 CCDC19、LAMB1、RAD51AP1 以及 SPAG6。CCDC19 与初始 B 细胞、CD4⁺ 静息记忆 T 细胞、记忆 B 细胞以及初始 CD4⁺ T 细胞浸润水平显著正相关 ($P < 0.05$),而与中性粒细胞、M0 和 M1 巨噬细胞、活化的 CD4⁺ 记忆 T 细胞以及活化的自然杀伤 (natural killer, NK) 细胞浸润水平显著负相关 ($P < 0.05$); LAMB1 与嗜酸性粒细胞、活化的 CD4⁺ 记忆 T 细胞、M1 巨噬细胞以及中性粒细胞浸润水平

显著正相关 ($P < 0.05$),而与滤泡辅助 T 细胞、CD4⁺ 静息记忆 T 细胞以及记忆 B 细胞等浸润水平显著负相关 ($P < 0.05$); RAD51AP1 与嗜酸性粒细胞、活化的 CD4⁺ 记忆 T 细胞以及中性粒细胞等浸润水平显著正相关 ($P < 0.05$),而与初始 CD4⁺ T 细胞、CD4⁺ 静息记忆 T 细胞以及初始 B 细胞等浸润水平显著负相关 ($P < 0.05$); SPAG6 与初始 B 细胞、CD4⁺ 静息记忆 T 细胞以及记忆 B 细胞等浸润水平显著正相关 ($P < 0.05$),而与活化 NK 细胞、M0 巨噬细胞以及 CD8⁺ T 细胞等浸润水平显著负相关 ($P < 0.05$)。CCDC19 与 SPAG6 具有相似的免疫细胞浸润情况,而 LAMB1 与 RAD51AP1 具有相似的免疫细胞浸润情况(图 5)。



A: CCDC19; B: LAMB1; C: RAD51AP1; D: SPAG6.

图5 关键特征基因与免疫细胞相关性

3 讨论

由于鼻咽部的解剖位置较为隐匿,NPC 患者早期临床症状不典型^[9-10]。因此,早期的针对性预防与治疗,对提高 NPC 患者的总体生存率具有重要意义。虽然人类 EB(Epstein-Barr) 病毒在早期诊断 NPC 方面发挥重要作用,但缺乏特异性,且检测方法较为复杂,而晚期 NPC 易发生转移,且对抗肿瘤药物和放

化疗不敏感^[11-12]。有研究发现,NPC 发展过程中常伴有癌基因的异常表达^[13]。因此探索 NPC 发展过程中的关键基因具有重要的现实意义。

机器学习正在医学领域得到快速发展,且主要应用于影像组学与病理学图像分割^[14-15],以预测疾病发病风险为目的。LASSO 和 SVM 目前是在预测疾病发生风险领域应用最为广泛的两种机器学习算法。

LASSO 算法与传统回归分析方法相比,可通过参数估计,消除数据之间的共线性影响,从而明确不同特征变量之间的相关性,并且在处理具有高纬度数据的生物信息学方面有显著优势^[16]。SVM 算法可根据不同维度的数据实现最佳分类,消除数据过度拟合现象,达到最佳折衷^[17]。因此,综合利用两种机器学习算法的优势,可有效筛选 NPC 发生过程的关键特征基因。

研究发现,DEG 的 GO 功能主要富集在体液免疫反应、角化与细胞分化、中性粒细胞激活以及趋化因子受体结合等上。体液免疫反应在抗肿瘤过程中发挥关键性作用,且由 B 细胞介导^[18];而 B 细胞可通过转化成为浆细胞,进而产生抗肿瘤细胞的免疫球蛋白,从而进一步发挥抗肿瘤作用^[19]。通过分析 NPC 免疫微环境发现,幼稚 B 细胞和记忆 B 细胞在免疫微环境中的表达显著下降,而浆细胞的表达有所上升。这可能是由于 B 细胞在抗肿瘤细胞增殖过程中耗竭所致,且有学者通过研究肾细胞癌发现,肿瘤三级淋巴结构 (tertiary lymphoid structures, TLS) 在维持 B 细胞成熟和抗体生成过程中发挥重要作用,可起到直接抗肿瘤作用^[20]。此外,有研究发现肿瘤组织的中性粒细胞的标志物 CD54 表达增高^[21],而且局部浸润的中性粒细胞可抑制人体的细胞免疫力并导致肿瘤细胞免疫逃逸^[22]。另有文献报道,在实体瘤的中性粒细胞胞外诱捕网 (neutrophil extracellular traps, NET) 不利于患者的预后^[23];而本研究通过免疫相关分析发现 NPC 免疫微环境中性粒细胞表达有所增高。趋化因子及其受体与 NPC 的复发转移也密切相关^[24],而细胞分化与 NPC 分类密切相关,并影响患者的放化疗及预后^[25-26]。因此,DEG 的功能与肿瘤细胞增殖、转移、分化及患者预后密切相关。而 KEGG 通路分析发现,DEG 主要富集在 IL-17 信号通路上,而该信号通路在 NPC 的发生发展过程中发挥重要作用^[27]。此外,GSEA 富集分析结果显示,其与肿瘤的发生发展以及 NPC 的进展密切相关。因此,无论是 DEG 的 GO、KEGG 还是 GSEA 富集分析,结果均显示与肿瘤以及 NPC 的发生发展、转移、复发及预后密切相关。

免疫微环境的改变与肿瘤的发生及患者预后密切相关。有研究发现在 EB 病毒感染的 NPC 组织中,以 B 细胞受累为主,而且 EB 病毒可长期潜伏于记忆 B 细胞^[28]。在机体处于免疫系统受抑制状态下,潜伏于记忆 B 细胞的 EB 病毒得到激活,从而导致 NPC 的发生。此外,有研究者通过分析 NPC 患者静息状态下 CD4⁺ T 细胞发现,其表达水平显著降低,并认

为其可作为评估 NPC 患者免疫状态的依据^[29];而另有学者研究发现,CD4⁺ T 细胞/CD8⁺ T 细胞比值下降与 NPC 的局部复发和远处转移密切相关,并认为其可用于预测 NPC 患者的生存预后^[30]。针对 M0 和 M1 巨噬细胞在 NPC 免疫微环境中高表达,有学者研究发现其在 NPC 发生发展过程中发挥重要作用,并与肿瘤细胞侵袭力密切相关,并且影响患者生存预后^[31]。

总之,通过机器学习算法筛选 NPC 关键特征基因,并分析其与免疫细胞的相关性,在利用医疗大数据的基础上,初步实现了生物信息学与医学的有机融合,实现了学科交叉研究,可为其他研究领域的学者提供一种创新的研究方法。但由于研究所涉及的 NPC 样本量不够庞大,且缺乏相关的实验验证,故关键特征基因在 NPC 发生过程中的作用机制研究尚不够深入。因此,后续还需搜集 NPC 大样本数据,进行相关基础与临床研究,以明确关键基因在 NPC 发生过程中的作用机制。

参考文献:

- [1] 杨月,张培培,张浩轩,等. 鼻咽癌基因表达谱的免疫细胞浸润特征分析及临床意义[J]. 细胞与分子免疫学杂志, 2020, 36(12): 1069-1075.
Yang Y, Zhang P, Zhang H, et al. Analysis of immune cell infiltration characteristics of nasopharyngeal carcinoma Gene expression profiling and its clinical significance[J]. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi, 2020, 36(12): 1069-1075.
- [2] 张浩轩,陆进,蒋成义,等. miR-429-4-3p 通过靶向抑制 WEE1 蛋白激酶表达逆转 HNE1/CDDP 人鼻咽癌细胞顺铂耐药[J]. 细胞与分子免疫学杂志, 2019, 35(11): 1014-1019.
Zhang H, Lu J, Jiang C, et al. miR-429-4-3p reverses cisplatin resistance in HNE1/CDDP human nasopharyngeal cancer cells by targeting inhibition of WEE1 protein kinase expression[J]. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi, 2019, 35(11): 1014-1019.
- [3] Fei Z, Hong H, Xu T, et al. Analysis of risk characteristics for metachronous metastasis in different period of nasopharyngeal carcinoma[J]. BMC Cancer, 2023, 23(1): 165.
- [4] 潘玉峰,赵强,王汐,等. miR-634 调控趋化因子受体-4 蛋白对鼻咽癌细胞生物活性的作用机制[J]. 中国耳鼻咽喉头颈外科, 2022, 29(2): 79-84.
Pan Y, Zhao Q, Wang X, et al. Mechanism of miR-634 regulation of chemokine receptor-4 protein on biological activity of nasopharyngeal carcinoma cells[J]. Zhongguo Er Bi Yan Hou Tou Jing Wai Ke, 2022, 29(2): 79-84.
- [5] Liu G Y, Li W Z, Xie C B, et al. Trajectories of EBV DNA and identifying the potential long-term survivors in metastatic nasopharyngeal carcinoma[J]. Am J Cancer Res, 2021, 11(8): 3946-3955.
- [6] Huang Y, Yan D, Wang M, et al. Clinicopathological factors affecting the prognosis of massive hemorrhage after radiotherapy for patients having nasopharyngeal carcinoma[J]. J Cancer Res Ther, 2021,

- 17(5): 1219-1224.
- [7] Auwul M R, Rahman M R, Gov E, et al. Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19 [J]. *Brief Bioinform*, 2021, 22(5): bbab120.
- [8] Liu L P, Lu L, Zhao Q Q, et al. Identification and validation of the pyroptosis-related molecular subtypes of lung adenocarcinoma by bioinformatics and machine learning[J]. *Front Cell Dev Biol*, 2021, 9: 756340.
- [9] 贺巧,李贤兵,李罗娜,等. 非高发区人群中血浆 EB 病毒 DNA 对鼻咽癌的筛查及临床应用价值[J]. *中华检验医学杂志*, 2022, 45(4): 381-387.
- He Q, Li X, Li L, et al. Screening of plasma EBV DNA for nasopharyngeal carcinoma in a non-high prevalence area population and its clinical application value[J]. *Zhonghua Jian Yan Yi Xue Za Zhi*, 2022, 45(4): 381-387.
- [10] Mao Y P, Wang S X, Gao T S, et al. Medial retropharyngeal nodal region sparing radiotherapy versus standard radiotherapy in patients with nasopharyngeal carcinoma: open label, non-inferiority, multicentre, randomised, phase 3 trial[J]. *BMJ*, 2023, 380: e072133.
- [11] Nguyen S, Kong T, Berthelet E, et al. A unique case of primary EBV-positive, HPV-negative nasopharyngeal carcinoma located in the tonsil[J]. *Head Neck Pathol*, 2021, 15(3): 1017-1022.
- [12] Müller Coan B G, Cesarman E, Acencio M L, et al. Latent membrane protein 1 (LMP1) from Epstein-Barr virus (EBV) strains M81 and B95.8 modulate miRNA expression when expressed in immortalized human nasopharyngeal cells [J]. *Genes (Basel)*, 2022, 13(2): 353.
- [13] Liu Y H, Li X N, Zhang Y W, et al. NAP1L1 targeting suppresses the proliferation of nasopharyngeal carcinoma [J]. *Biomed Pharmacother*, 2021, 143: 112096.
- [14] Xu M, Calhoun V, Jiang R, et al. Brain imaging-based machine learning in autism spectrum disorder: methods and applications[J]. *J Neurosci Methods*, 2021, 361: 109271.
- [15] Punchoo R, Bhoora S, Pillay N. Applications of machine learning in the chemical pathology laboratory[J]. *J Clin Pathol*, 2021, 74(7): 435-442.
- [16] Shao W, Yang Z, Fu Y, et al. The pyroptosis-related signature predicts prognosis and indicates immune microenvironment infiltration in gastric cancer[J]. *Front Cell Dev Biol*, 2021, 9: 676485.
- [17] Shen S, Liu Z, Wang J, et al. Machine learning assisted Cameriere method for dental age estimation [J]. *BMC Oral Health*, 2021, 21(1): 641.
- [18] Kinker G S, Vitiello G A F, Ferreira W A S, et al. B cell orchestration of anti-tumor immune responses: A matter of cell localization and communication[J]. *Front Cell Dev Biol*, 2021, 9: 678127.
- [19] Cancro M P, Tomayko M M. Memory B cells and plasma cells: The differentiative continuum of humoral immunity [J]. *Immunol Rev*, 2021, 303(1): 72-82.
- [20] Meylan M, Petitprez F, Becht E, et al. Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer[J]. *Immunity*, 2022, 55(3): 527-541. e5.
- [21] 汤成婕,疏龙强,吴缘开,等. 中性粒细胞的免疫抑制亚群及功能和机制研究进展[J]. *细胞与分子免疫学杂志*, 2020, 36(9): 833-842.
- Tang C, Shu L, Wu Y, et al. Advances in the study of immunosuppressive subpopulations of neutrophils and their functions and mechanisms [J]. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi*, 2020, 36(9): 833-842.
- [22] Faget J, Peters S, Quantin X, et al. Neutrophils in the era of immune checkpoint blockade [J]. *J Immunother Cancer*, 2021, 9(7): e002242.
- [23] Yang C, Wang Z, Li L, et al. Aged neutrophils form mitochondria-dependent vital NETs to promote breast cancer lung metastasis[J]. *J Immunother Cancer*, 2021, 9(10): e002875.
- [24] 阮鹏,谭爱丽. 趋化因子 CXCL5 通过调控 ERK/MAPK 信号通路抑制肿瘤免疫促进鼻咽癌恶化的机制研究[J]. *中国免疫学杂志*, 2020, 36(21): 2613-2618.
- Ruan P, Tan A. Mechanisms by which chemokine CXCL5 suppresses tumor immunity to promote nasopharyngeal carcinoma progression by regulating ERK/MAPK signaling pathway [J]. *Zhongguo Mian Yi Xue Za Zhi*, 2020, 36(21): 2613-2618.
- [25] 孙海力,李绍霄,林海亮. 转录组-芯片数据联合分析鉴定鼻咽癌发生潜在关键信号通路和调节基因[J]. *中国耳鼻咽喉头颈外科*, 2020, 27(3): 119-122.
- Sun H, Li S, Lin H. Identification of potential key signaling pathways and regulatory genes for nasopharyngeal carcinogenesis by combined transcriptome-microarray data analysis[J]. *Zhongguo Er Bi Yan Hou Tou Jin Wai Ke*, 2020, 27(3): 119-122.
- [26] Kam N W, Laczka O, Li X, et al. ENOX2 inhibition enhances infiltration of effector memory T-cell and mediates response to chemotherapy in immune-quiescent nasopharyngeal carcinoma[J]. *J Adv Res*, 2023 Apr 13: S2090-1232(23)00093-0. DOI: 10.1016/j.jare.2023.04.001. Online ahead of print.
- [27] Tai J, Park J, Han M, et al. Screening key genes and biological pathways in nasopharyngeal carcinoma by integrated bioinformatics analysis[J]. *Int J Mol Sci*, 2022, 23(24): 15701.
- [28] Andrei G, Trompet E, Snoeck R. Novel therapeutics for Epstein-Barr virus[J]. *Molecules*, 2019, 24(5): 997.
- [29] 赵美淇,梁新强,刘海洲,等. 肿瘤患者外周血 T 淋巴细胞亚群及 NK 细胞比例检测的临床意义[J]. *中国实验诊断学*, 2019, 23(3): 381-384.
- Zhao M, Liang X, Liu H, et al. Clinical significance of peripheral blood T lymphocyte subpopulation and NK cell ratio detection in tumor patients [J]. *Zhongguo Shi Yan Zhen Duan Xue*, 2019, 23(3): 381-384.
- [30] Shen D S, Yan C, Liang Y, et al. Prognostic significance of circulating lymphocyte subsets before treatment in patients with nasopharyngeal carcinoma[J]. *Cancer Manag Res*, 2021, 13: 8109-8120.
- [31] Wang Y J, Sun Q, Ye Y, et al. FGF-2 signaling in nasopharyngeal carcinoma modulates pericyte-macrophage crosstalk and metastasis [J]. *JCI Insight*, 2022, 7(10): e157874.