Hans Journal of Biomedicine 生物医学, 2026, 16(1), 67-81
Published Online January 2026 in Hans. https://www.hanspub.org/journal/hjbm
https://doi.org/10.12677/hjbm.2026.161008

Hans汉斯

基于生物信息学发掘与鉴定食管癌潜在生物标志物

韦　佳[1*]，韩子璠[1*]，陈伊蕾[1*]，程向前[1]，邢浩宽[2]，张艺潇[2]，张义炫[2]，陈健新[2#]

[1]河北北方学院第一临床医学院，河北　张家口
[2]河北北方学院基础医学院，河北　张家口

收稿日期：2025年11月15日；录用日期：2026年1月8日；发布日期：2026年1月16日

摘　要

目的：基于生物信息学，对食管癌(Esophageal cancer, ESCA)潜在的生物标志物在食管癌发生发展中的作用机制进行探究。方法：通过GEO数据库(https://www.ncbi.nlm.nih.gov/geo/)和TCGA数据库(https://portal.gdc.cancer.gov)获取食管癌数据集后，按照$p < 0.05$的标准进行差异基因的筛选。将初步筛选后的基因通过四种机器学习方法(Boruta、Lasso、XGBoost、随机森林和SVM-RFE算法)对筛选出的关键基因在食管癌中的重要性进行评估后，对其进行通路分析用以揭示其作用机制。进行受试者工作特征曲线(receiver operating characteristic curves analysis, ROC)分析，评估基因对于食管癌的诊断效能。最后对基因进行生存分析与单细胞分析，分别评估基因对ESCA预后的影响与基因在肿瘤组织中单个细胞水平的表达情况。绘制了ADAM12、CTHRC1、IBSP和OLR1基因在食管癌中的多组学图谱，揭示了其对于食管癌的发生、发展与预后的高度相关性。结果：ADAM12、CTHR1、IBSP、OLR1为食管癌的危险因素，其表达的高低对于预测食管癌的发生和预后具有很高的准确性，并且低表达组相比于高表达组预后更加良好。结论：得到了与食管癌的发生和预后相关的四个基因：ADAM12、CTHR1、IBSP、OLR1，可能成为食管癌早期诊断的生物标志物。

关键词

食管癌，生物标志物，单细胞组学，空间转录组学

Identification of Potential Biomarkers for Esophageal Cancer Based on Bioinformatics

Jia Wei[1*], Zifan Han[1*], Yilei Chen[1*], Xiangqian Cheng[1], Haokuan Xing[2], Yixiao Zhang[2], Yixuan Zhang[2], Jianxin Chen[2#]

[1]The First Clinical Medical College, Hebei North University, Zhangjiakou Hebei

[*]共同第一作者。
[#]通讯作者。

文章引用: 韦佳, 韩子璠, 陈伊蕾, 程向前, 邢浩宽, 张艺潇, 张义炫, 陈健新. 基于生物信息学发掘与鉴定食管癌潜在生物标志物[J]. 生物医学, 2026, 16(1): 67-81. DOI: 10.12677/hjbm.2026.161008

[2]School of Basic Medical Sciences, Hebei North University, Zhangjiakou Hebei

Received: November 15, 2025; accepted: January 8, 2026; published: January 16, 2026

Abstract

Objective: To investigate the mechanism of potential biomarkers in the occurrence and development of esophageal cancer (ESCA) based on bioinformatics methods. Methods: Esophageal cancer-related datasets were retrieved from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database and The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov) database. Differentially expressed genes were screened with the criterion of $p < 0.05$. The importance of the initially screened genes in esophageal cancer was evaluated by four machine learning methods, namely Boruta algorithm, Lasso algorithm, extreme gradient boosting (XGBoost) algorithm, random forest algorithm, and support vector machine-recursive feature elimination (SVM-RFE) algorithm. Pathway analysis was subsequently performed to elucidate their underlying mechanisms of action. Receiver operating characteristic (ROC) curve analysis was conducted to assess the diagnostic efficacy of these genes for esophageal cancer. Finally, survival analysis and single-cell analysis were carried out to determine the impact of the target genes on the prognosis of ESCA patients and their expression characteristics at the single-cell level in tumor tissues, respectively. Multi-omics profiles of ADAM12, CTHRC1, IBSP and OLR1 genes in esophageal cancer were plotted, and their high correlation with the occurrence, development and prognosis of esophageal cancer was analyzed. Results: ADAM12, CTHRC1, IBSP and OLR1 were identified as risk factors for esophageal cancer. The expression levels of these genes showed high accuracy in predicting the occurrence and prognosis of esophageal cancer, and patients in the low-expression group had significantly better prognosis than those in the high-expression group. Conclusion: Four genes (ADAM12, CTHRC1, IBSP, OLR1) closely associated with the occurrence and prognosis of esophageal cancer were screened out in this study. These genes are expected to serve as potential biomarkers for the early diagnosis of esophageal cancer.

Keywords

Esophageal Cancer, Biomarker, Single-Cell Omics, Spatial Transcriptomics

1. 引言

食管癌(Esophageal cancer, ESCA)是一种常见的消化系统恶性肿瘤，在中国因癌症死亡的病例中标化死亡率排第五位，在我国超过全球死亡总例数 30%的癌种中位列第一[1]。由于食管癌的早期临床症状并不明显，大于 50%的患者确诊时已经处于疾病发展的中晚期，导致其治疗效果和预后水平较差。此外，食管癌患者的死亡率逐年升高，患者的 5 年生存率不足 10%，明显低于其他恶性肿瘤。目前根治性切除和纵隔淋巴结清扫术依旧是该病临床治疗的首选方法，但部分患者术后淋巴结病理仍呈阳性，需进行术后辅助化疗，预后效果并不理想。对于食管癌患者，早期确诊和及时手术利于预后，而淋巴结转移不利于预后，有研究表明，当食管癌发展至粘膜下浸润 >1000 μm 时，淋巴结转移的风险会显著增加。因此，早期诊断对于提高食管癌患者的生存率十分重要[2]-[4]。本研究基于基因表达数据库(Gene Expression

Omnibus, GEO)和癌症基因组图谱(The Cancer Genome Atlas, TCGA)数据库，查找出与食管癌发病相关的基因后，进行多种机器学习算法(Boruta、Lasso、XGBoost、随机森林和支持向量机)后进行多组学分析，评估了 ADAM12、CTHR1、IBSP、OLR1 四个基因对食管癌发生发展的机制，旨在寻找食管癌的免疫学标志物，为其早期治疗提供帮助，为后续的研究提供思路。

2. 资料与方法

2.1. 数据集的处理和差异表达基因分析

2.1.1. GEO、TCGA 数据的获取

从 GEO 数据库(https://www.ncbi.nlm.nih.gov/geo/)和 TCGA 数据库(https://portal.gdc.cancer.gov)中以“Esophageal cancer”为关键词检索，获得食管癌组织基因表达数据和正常食管组织基因表达数据。

2.1.2. 差异基因筛选

选用基因芯片 GSE149609，使用 R 语言再次标准化数据，并采用 limma 包分别对数据集进行差异分析，按照 $p < 0.05$ 的标准筛选上调和下调的差异基因，最后进行可视化作图展示差异基因。

2.2. 机器学习进行基因进阶筛选

2.2.1. Boruta 机器学习

(1) 创建差异基因的副本后，打乱重复副本的值，用以消除其与目标变量的相关性，即为影子特征。

(2) 将原副本与洗过的副本合并后，使用随机森林方法，比较差异基因每个变量的重要性，进一步消除不相关的特征。其中，重要性显著高于影子属性的属性会被标记为“重要”的，显著低于影子属性的属性会被标记为“不重要”。

2.2.2. Lasso 机器学习

采用 glmnet 包来执行 Lasso 回归，指定 family 参数为 binomial 后，设置弹性网络中的混合参数 alpha 参数是 1，即在损失函数中加入一个 L1 正则化项。选用 cv.glmnet 函数，deviance 为度量指标执行十折交叉验证，从而实现特征选择。

2.2.3. XGBoost 机器学习

借助 xgb.DMatrix 函数将数据转换为 XGBoost 的内部数据结构。在进行模型训练时，选定 xgboost 函数，将目标函数确定为“binary:logistic”，用于二分类问题并采用逻辑回归损失，同时将迭代次数设置为 25。随后，运用 xgb.importance 函数进行特征选择。最后，选用 xgb.ggplot.importance 函数绘制特征重要性图。

2.2.4. 随机森林

(1) 对原始数据进行预处理后，通过自主采样的方式抽出多个数据集后进行特征随机选择，进一步增加多样性。

(2) 使用数据集构建决策树，Out-of-Bag (OOB)误差率进行交叉验证

(3) 选择用佳树的数量训练随机森林模型后，使用 importance 函数计算特征重要性，得出特征在模型中的重要性得分，得分越高该特征对模型的影响越大。选取前 20%的特征基因。

2.2.5. 向量机递归特征消除(SVM-RFE)算法

基于支持向量机进行机器学习，设定交叉验证的折数为 10，对数据集进行 10 折划分后，对每一折应用 SVM-RFE 算法，获取特征的重要性排名，对所有折的排名结果进行整合后，计算平均排名，根据平均

排名对特征进行排序。

2.3. 通路分析

采用 limma 包执行差异分析，clusterProfiler 包执行基因集富集分析后，选取分别在高/低表达组显著富集的 top5 通路进行可视化。

2.4. 受试者工作特征曲线(Receiver Operating Characteristic Curves Analysis, ROC)分析

使用 pROC 包对差异基因进行 ROC 分析，计算 95%置信区间，获得曲线下总面积，使用 pROC 的 ci 函数评估 AUC，绘制平滑 ROC 曲线。

2.5. 生存分析

2.5.1. 生存风险比 Meta 分析

采用逆方差法，选择对数 HR 的值为主要指标，对单因素 cox 生存分析的结果进行 Meta 分析，标准误差使用置信区间 95% CI 计算后，使用 R (4.3.2 版本 0)的“Meta”包进行统计分析和可视化。

2.5.2. Kaplan-Meier 生存分析

采用 survival 包进行 Kaplan-Meier 生存分析，$P < 0.05$ 为有统计学意义。使用 survfit 函数进行 log rank test 评估高、低表达组的显著性。

2.6. 单细胞测序

对食管癌的单细胞测序数据进行整合，分析目标基因在食管癌中单个细胞水平层面的表达。

3. 结果

3.1. 基因的差异分析

对 GSE149609 中的基因进行分析，其中红色为上调，蓝色为下调(图 1)。以 20 个基因为一组进行筛选，最终选定一基因组中的四个基因：ADAM12、CTHRC1、IBSP 和 OLR1。

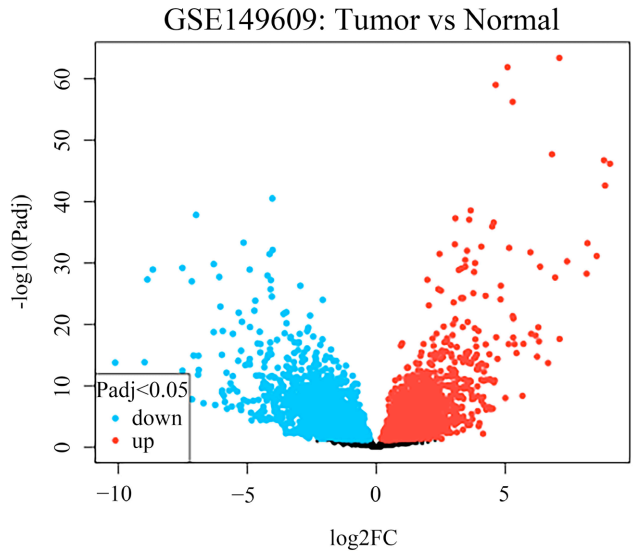


Figure 1. Differential expression analysis of the GSE149609 dataset
图 1. GSE149609 数据集的差异表达分析

使用 GTEx 的正常样本 TPM 表达量与 ADAM12、CTHRC1、IBSP 和 OLR1 四个基因的 TCGA 肿瘤 TPM 表达量进行配对后，通过$(x - \mu)/\sigma$ 将数据转化为无单位的 Z-Score 分值，使得数据标准统一化。当 z-score 大于 3.0 或小于−3.0 时，可归类为离群值予以去除。采用 Wilcoxon Rank Sum Tests 比较四个 DGEs 与 GTEx 正常组织之间的表达量统计学差异，并绘制表达箱线图(图 2)，其中四个基因的差异均具有统计学意义($p < 0.05$)。

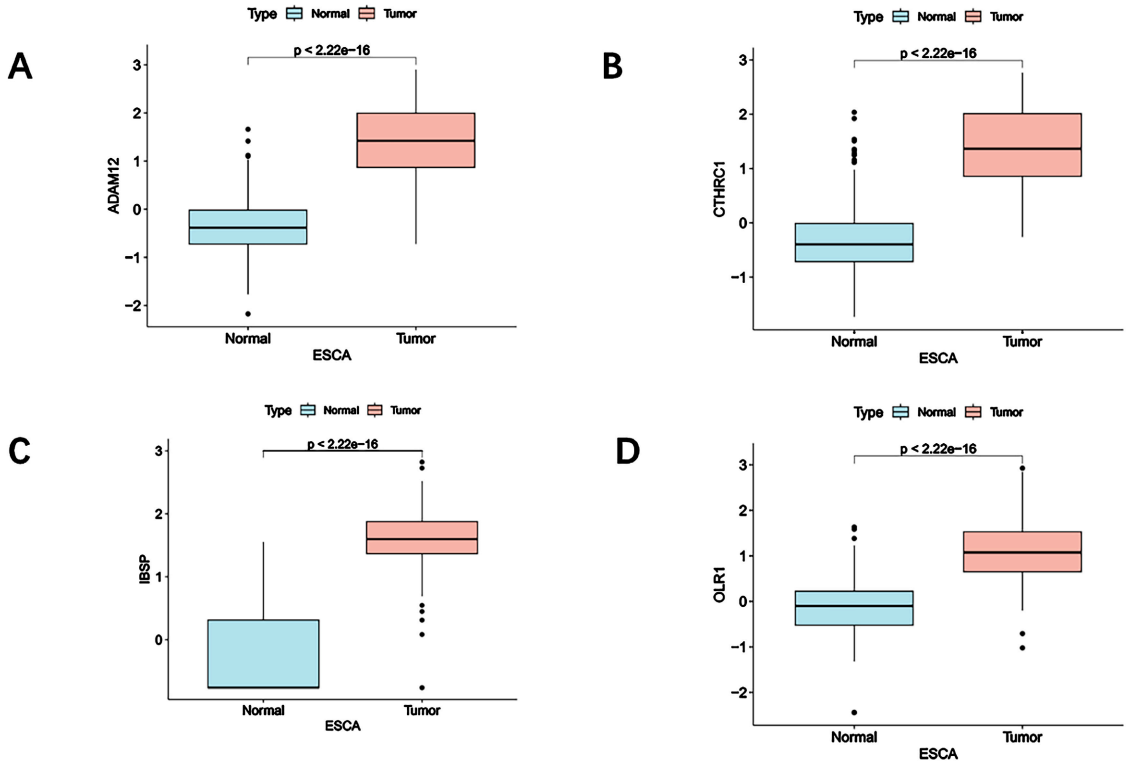


Figure 2. Box plots of differential analysis between DGEs and normal tissues. A: Box plot of ADAM12; B: Box plot of CTHRC1; C: Box plot of IBSP; D: Box plot of OLR1
图 2. DGEs 与正常组织的差异分析箱线图。A：ADAM12 的箱线图；B：CTHRC1 的箱线图 C：IBSP 的箱线图；D：OLR1 的箱线图

3.2. 机器学习

3.2.1. Boruta

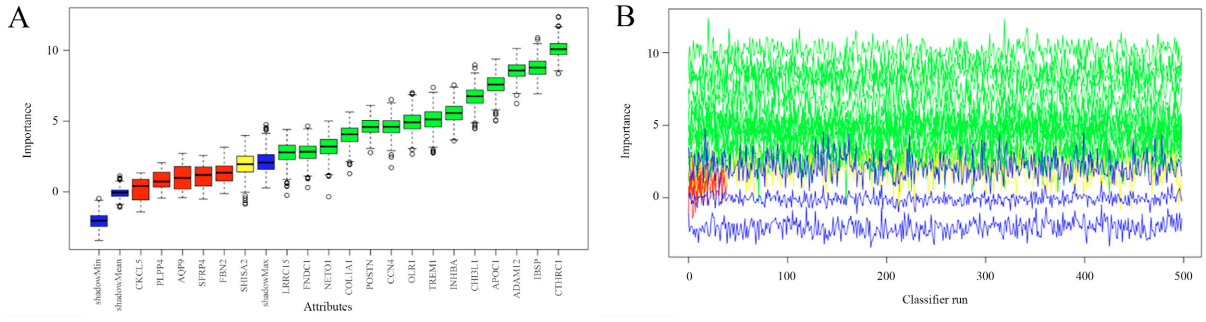


Figure 3. Boruta machine learning
图 3. Boruta 机器学习

图中 y 轴表示每个特征的重要性分数，分数越高表明该特征在模型中的重要性就越高。其中 ADAM12、CTHRC1、IBSP 和 OLR1 均为青色，表示已被 Boruta 算法确认为与预测变量显著相关的“已确认”特征(图 3)。

3.2.2. Lasso

将选定基因组中的 20 个 DGEs 纳入 Lasso 回归模型后，使用 plot 函数绘制十折交叉验证结果，其中第一条线对应于交叉验证误差最小的 λ 值，为最优的 λ 值(图 4(A))。在最优的 λ 值下，ADAM12、CTHRC1、IBSP 和 OLR1 在模型中有较大影响(图 4(B))。

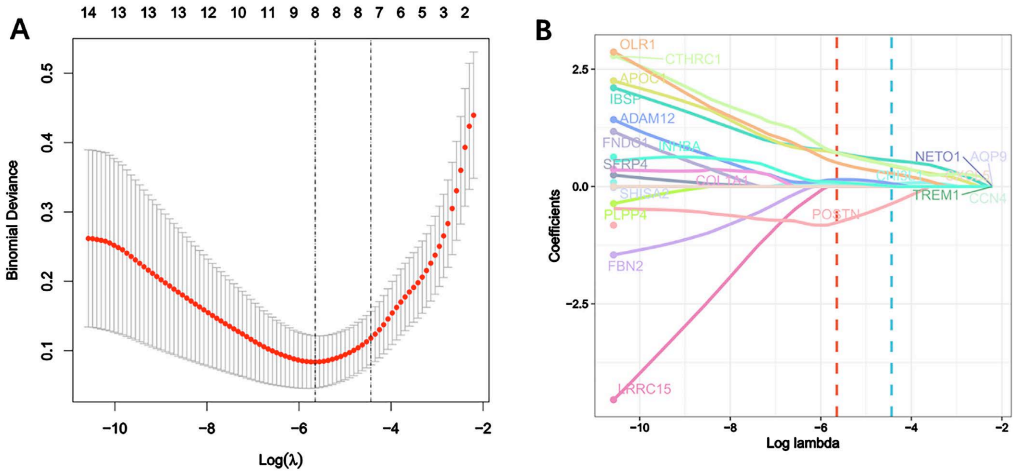


Figure 4. A: Cross-validation curve of Lasso regression; B: Coefficient path of Lasso regression
图 4. A：Lasso 回归交叉验证曲线；B：Lasso 回归系数路径

3.2.3. XGBoost

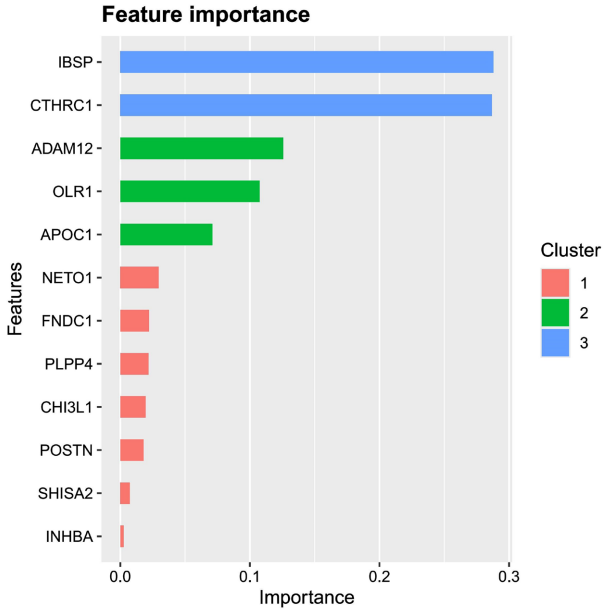


Figure 5. Evaluation of XGBoost variable importance
图 5. XGBoost 变量重要性评价

将基因组中的基因纳入 xgboost 模型，使用 xgb.importance 函数获取特征的重要性评估，xgb.ggplot.importance 函数绘制 XGBoost 模型的特征重要性图(图 5)，获得了 12 个重要特征，包括：IBSP、CTHRC1、ADAM12、OLR1、APOC1、NETO1、FNDC1、PLPP4、CHI3L1、POSTN、SHISA2、INHBA，其中 IBSP 得分最高。

3.2.4. 随机森林

绘制随机森林的 Out-of-Bag 误差率随树数量变化的图形，直观地展示模型误差率的变化趋势(图 6(A))。选取最优随机树量后，对影响因素的重要性进行排序，鉴定了重要性前 20%的特征，包括 ADAM12、CTHRC1、IBSP 和 OLR1 (图 6(B))。

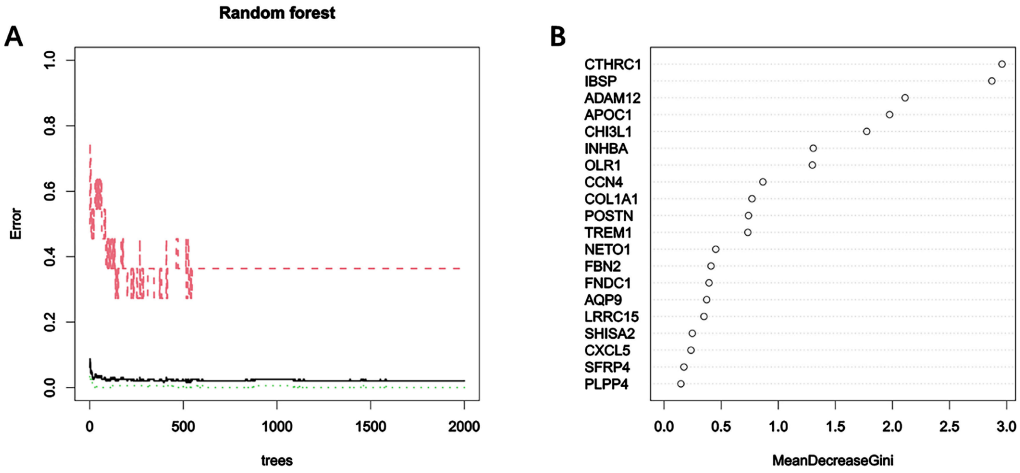


Figure 6. A: Variation plot of random forest error rate; B: Importance ranking plot
图 6. A：随机森林误差率变化图；B：重要性排序图

3.2.5. SVM-RFE

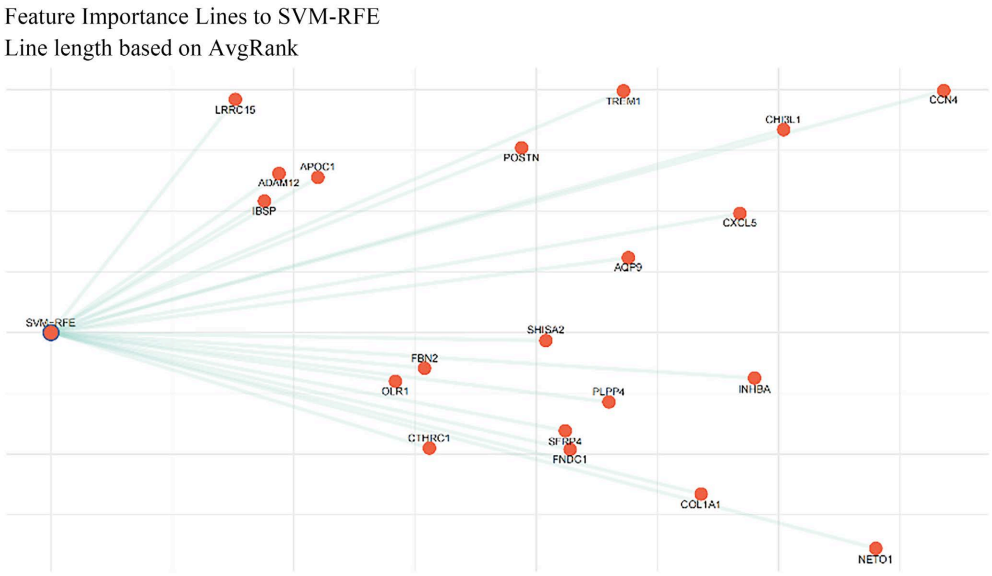


Figure 7. Average ranking plot
图 7. 平均排序图

采用支持向量机递归特征消除(SVM-RFE)算法在基因表达数据中识别并排序最重要的特征(图 7)，重要性 top10 的特征基因如下：LRRC15、IBSP、ADAM12、APOC1、OLR1、FBN2、CTHRC1、POSTN、SHISA2、SFRP4。

3.3. 通路分析

将表达最高的 30%的样本定义为高表达组，表达最低的 30%的样本定义为低表达组，对选中的四个基因进行通路分析(图 8~9)。不同的颜色代表不同的基因集，柱状图方向为左，意味着在低表达组显著富集，向右则意味在高表达组显著富集(图 10)。结果提示 ADAM12 与 CTHRC1 的 Angiogenesis、EMT、Invasion 通路，IBSP 的 EMT、Invasion 通路，OLR1 的 Angiogenesis、Apoptosis、Differentiation、EMT、Hypoxia、Inflammation、Invasion、Metastasis、Proliferation、Quiescence、Stemness 通路呈显著高表达富集趋势。

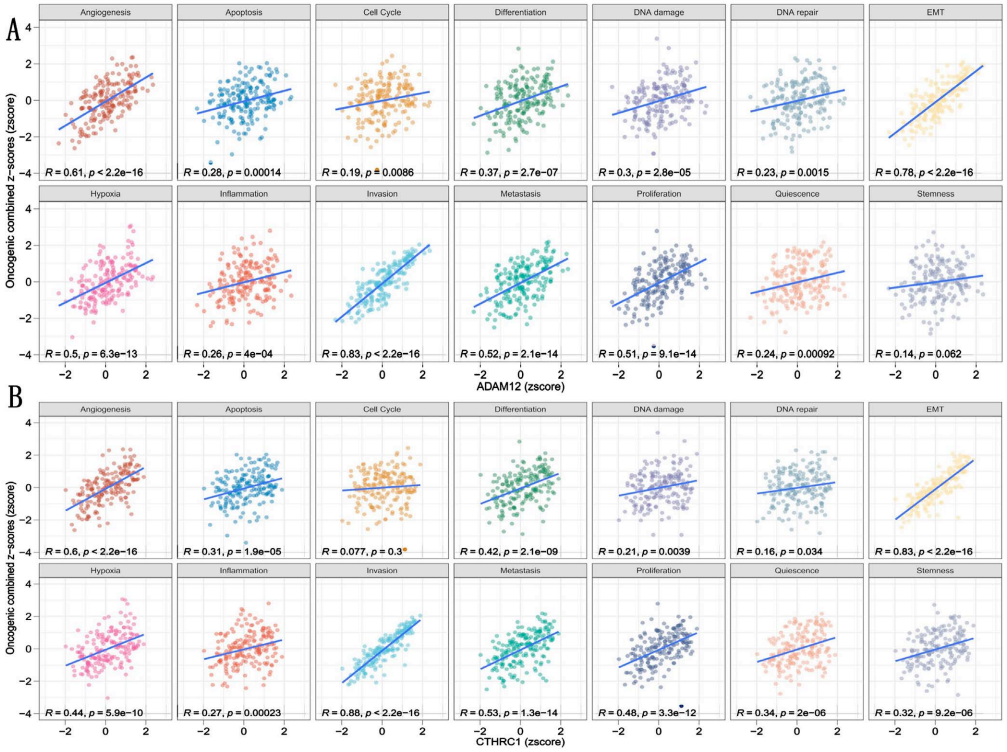
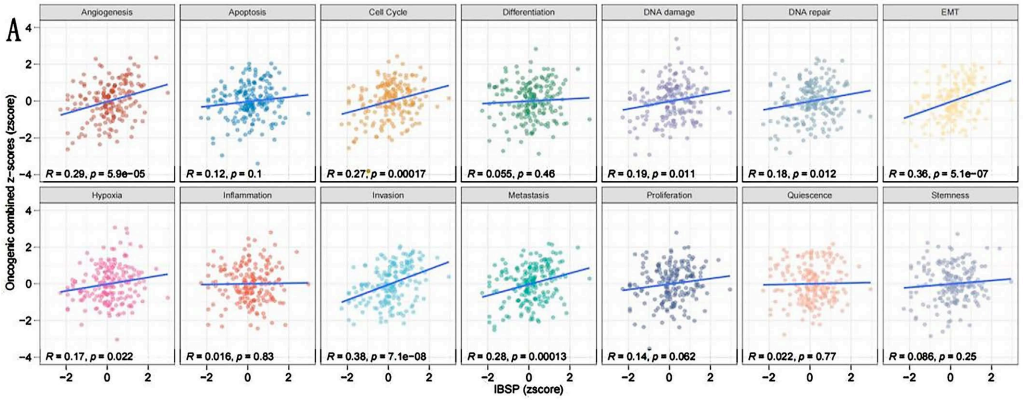


Figure 8. Pathway analysis plots of ADAM12 and CTHRC1
图 8. ADAM12 通路分析图与 CTHRC1 通路分析图



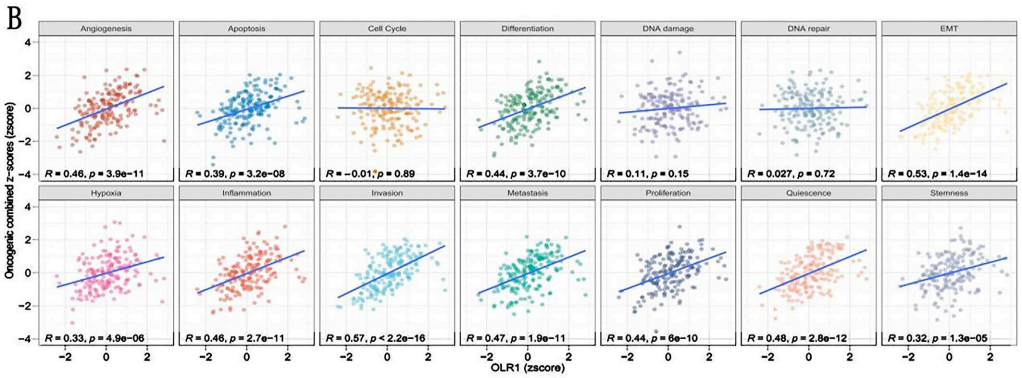


Figure 9. Pathway analysis plots of IBSP and OLR1
图 9. IBSP 通路分析图与 OLR1 通路分析图

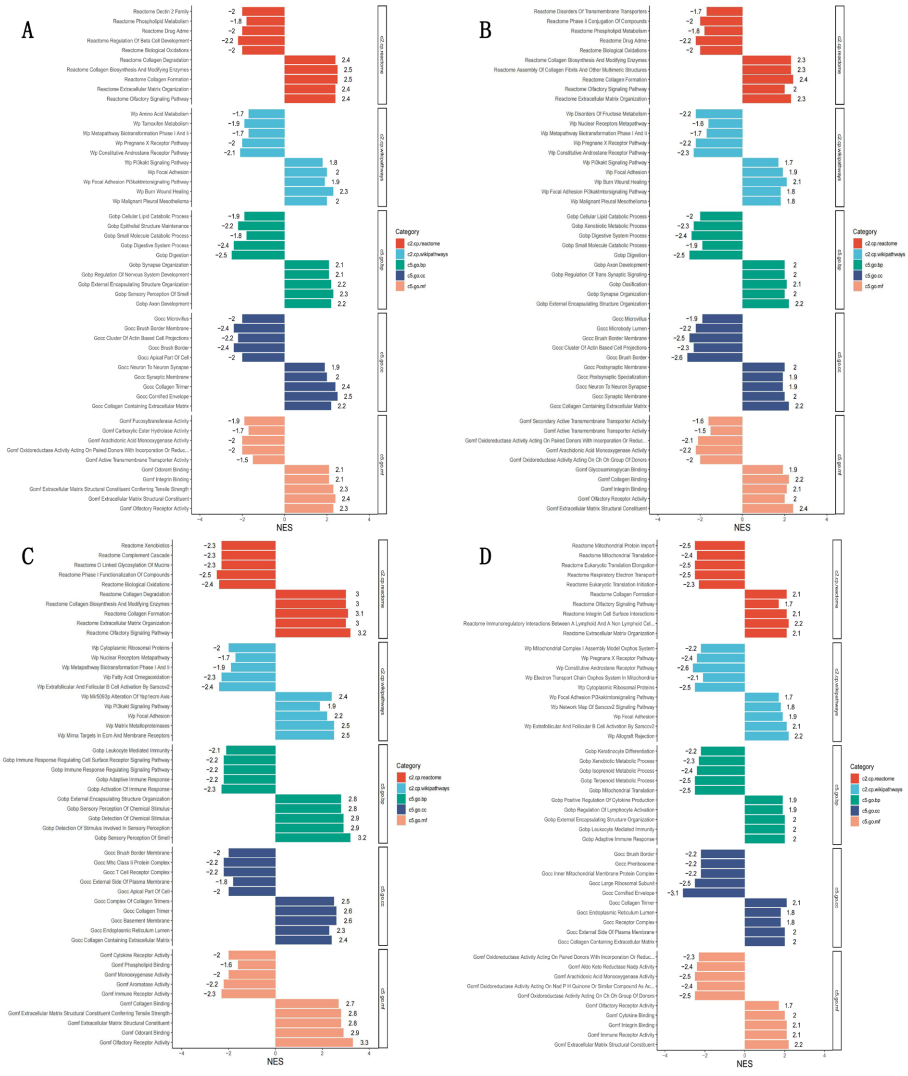


Figure 10. Multiple gene set enrichment analysis of high/low expression groups performed by clusterProfiler package. A. Enrichment analysis of ADAM12; B. Enrichment analysis of CTHRC1; C. Enrichment analysis of IBSP; D. Enrichment analysis of OLR1
图 10. clusterProfiler 包执行高/低表达组的多个基因集富集分析。A. ADAM12 富集分析；B. CTHRC1 富集分析；C. IBSP 富集分析；D. OLR1 富集分析

3.4. ROC 分析

ROC 曲线分析提示，采用 ADAM12、CTHR1、IBSP、OLR1 的表达预测食管癌疾病组与正常组均具有很高的准确性，AUC 值分别为 0.971、0.974、0.938、0.967，95%置信区间分别为 0.958~0.983、0.963~0.983、0.916~0.957、0.947~0.982 (图 11)。

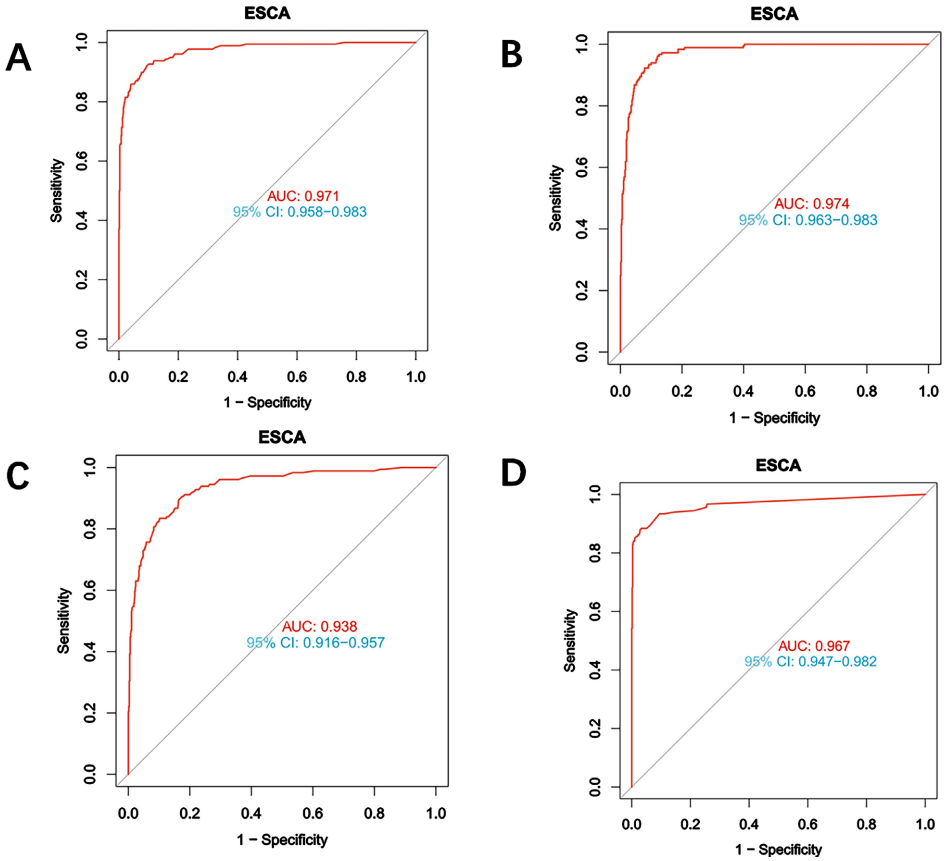


Figure 11. ROC analysis evaluating the diagnostic efficacy of gene expression for distinguishing tumor groups from normal groups
图 11. ROC 评估基因表达对肿瘤组与正常组的诊断效能

3.5. 生存分析

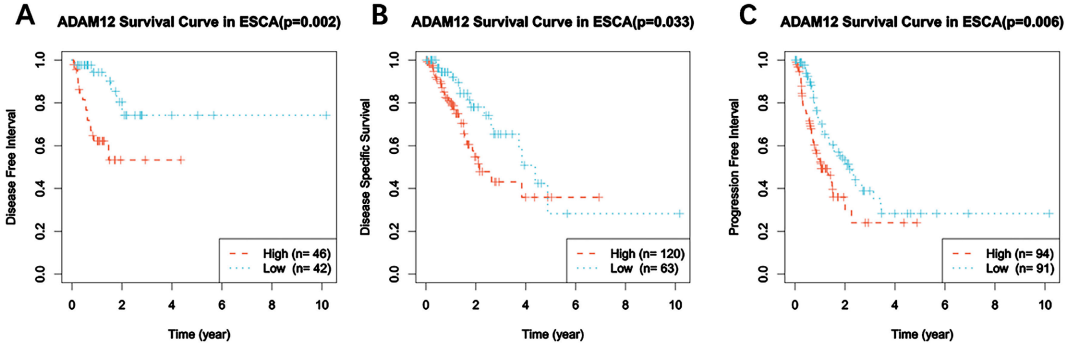


Figure 12. Kaplan-Meier survival analysis plot of ADAM12
图 12. ADAM12Kaplan-Meier 生存分析图

对 ADAM12、CTHR1、IBSP 和 OLR1 进行 3 个生存期(DSS，PFI 和 DFI)的 Kaplan-Meier 生存分析(图 12~15)，红色为基因高表达组，蓝色为低表达组，$p < 0.05$ 即为显著。四组中 p 均小于 0.05。结果表明，食管癌患者中，ADAM12、CTHR1、IBSP 和 OLR1 低表达组相比于高表达组预后更加良好。

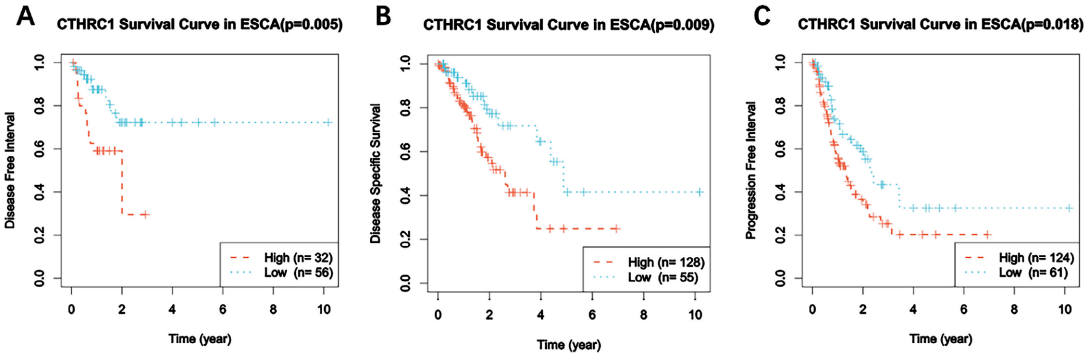


Figure 13. Kaplan-Meier survival analysis plot of CTHRC1
图 13. CTHRC1 Kaplan-Meier 生存分析图

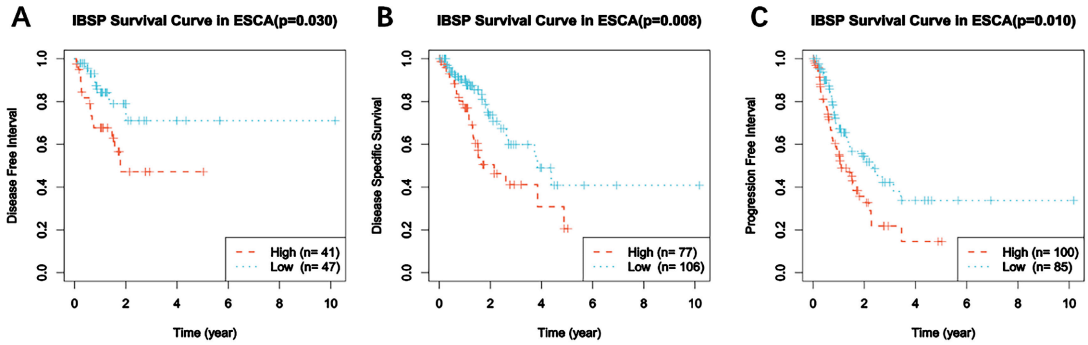


Figure 14. Kaplan-Meier survival analysis plot of IBSP
图 14. IBSP Kaplan-Meier 生存分析图

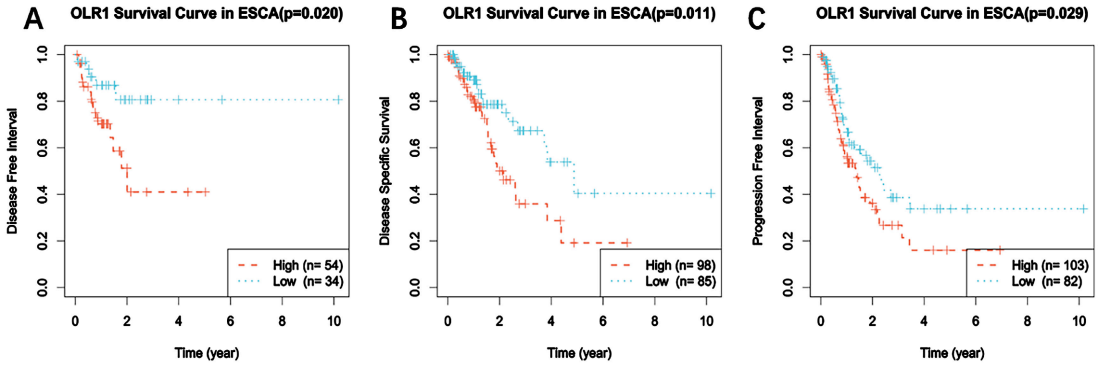


Figure 15. Kaplan-Meier survival analysis plot of OLR1
图 15. OLR1 Kaplan-Meier 生存分析图

采用逆方差法对上述生存分析的结果进行 Meta 分析，其中对数 HR 值为主要测量指标，标准误差使用 95% CI (置信区间)计算(图 16)。结果表明 ADAM12、CTHR1、IBSP、OLR1 的 OS、DSS 和 DFI 相关($p < 0.05$)，且 HR 均>1，均属危险因素，荟萃分析的结果与之一致，不同的生存期的结果存在较低的异质性。

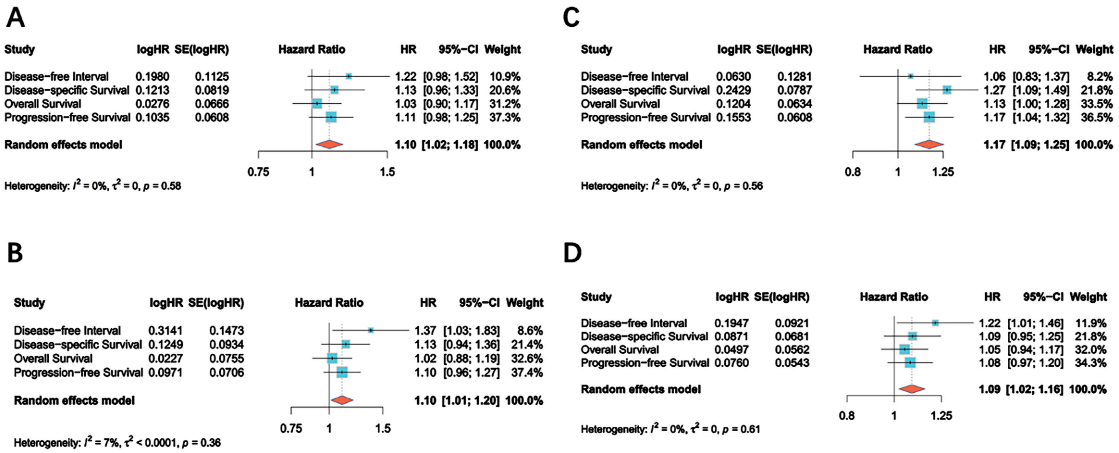


Figure 16. Meta-analysis plot of survival hazard ratios. A. ADAM12 analysis plot; B. CTHRC1 analysis plot; C. IBSP analysis plot; D. OLR1 analysis plot
图 16. 生存风险比 Meta 分析图。A. ADAM12 分析图；B. CTHR1 分析图；C. IBSP 分析图；D. OLR1 分析图

3.6. 单细胞分析基因的表达

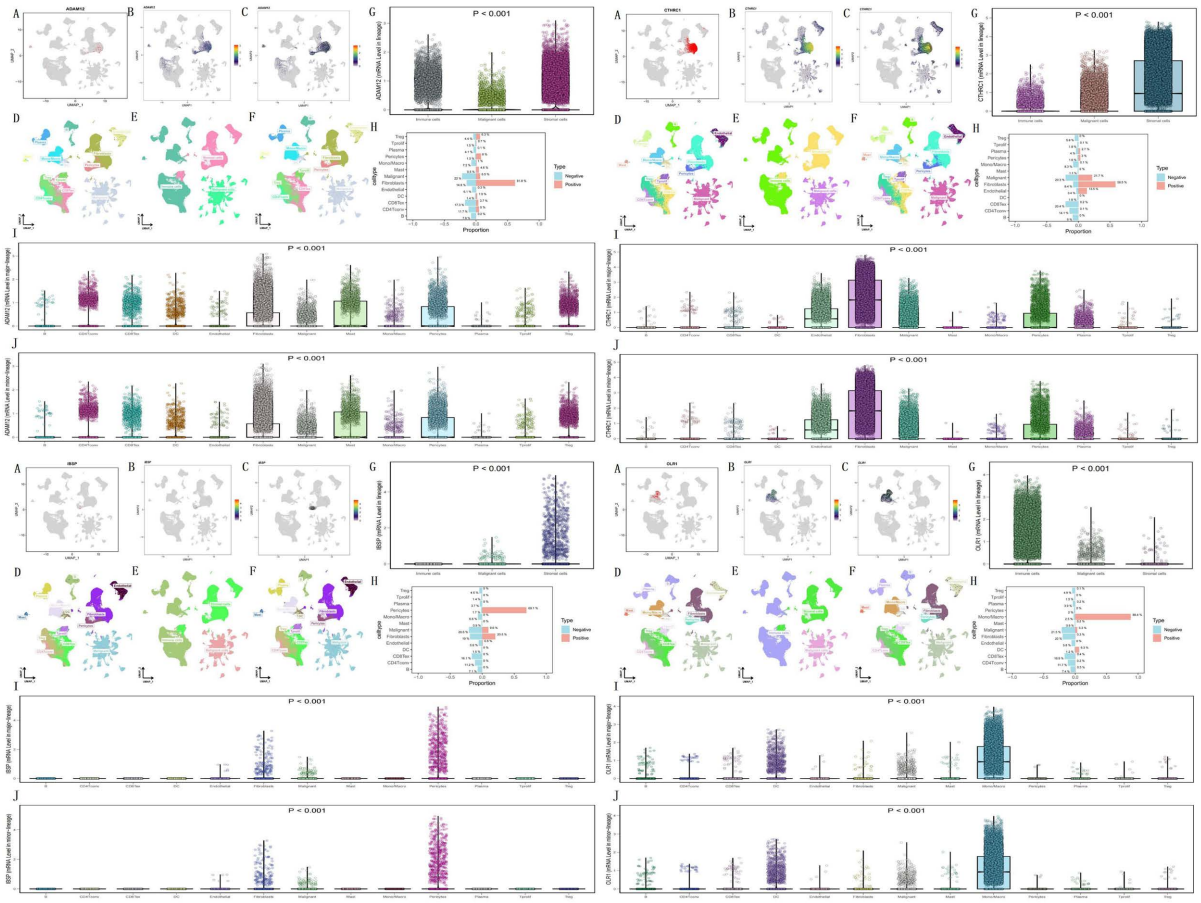


Figure 17. Single-Cell analysis plot of ADAM12 (Top Left), Single-Cell analysis plot of CTHRC1 (Top Right), Single-Cell analysis plot of IBSP (Bottom Left), Single-Cell analysis plot of OLR1 (Bottom Right)
图 17. ADAM12 的单细胞分析图(左上)，CTHR1 的单细胞分析图(右上)，IBSP 的单细胞分析图(左下)，OLR1 的单细胞分析图(右下)

对数据集 GSE149609 进行单细胞分析。在分析之前首先对细胞进行质量控制，如消除批次效应，以获得质量较好的细胞，便于后续的分析。三个细胞系免疫细胞、恶性细胞和基质细胞中包括 13 种细胞群：B 细胞、CD4 常规 T 细胞、CD8+T 细胞、树突状细胞、内皮细胞、胚胎成纤维细胞、恶性肿瘤细胞、肥大细胞、中性粒细胞、周细胞、浆细胞、T 细胞和调节性 T 细胞。

对 ADAM12、CTHR1、IBSP、OLR1 进行单基因表达 UMAP 定位后，对其在不同细胞中的表达和细胞比例进行差异分析(图 17)。结果显示，ADAM12、CTHR1、IBSP 和 OLR1 在基质细胞、恶性细胞和胚胎成纤维细胞中均表达显著，同时 ADAM12 在免疫细胞与调节性 T 细胞表达较为显著，CTHR1 在内皮细胞中表达较为显著，IBSP 与 OLR1 在周细胞中的表达最为显著。

结果表明四个基因均在恶性细胞与胚胎成纤维细胞中显著表达，当四个基因中任一基因的表达呈现升高趋势时，提示可能有食管癌的发生。

4. 讨论

食管癌(Esophageal cancer, ESCA)是最常见的恶性肿瘤之一，在 2020 年的相关调查中，ESCA 的发病率于所有癌症中位列第七，总死亡率居第六，这表明每 18 例因癌症死亡的病例中，就有 1 例是死于食管癌的，此外，发展中国家人群的发病率显著高于发达国家，特别是中国，根据 2018 年的统计，我国食管癌的发病率位居榜首，全世界半数的食管癌患者均在我国[5]。食管癌分为食管鳞状细胞癌(Esophageal squamous cell carcinoma, ESCC)和食管腺癌(Esophageal adenocarcinoma, EAC)两种亚型，食管鳞状细胞癌是我国食管癌的主要病理类型，占所有食管癌患者中的 90%，这可能与我国部分地区的饮食习惯如大量咀嚼槟榔、食用腌制食物和较烫的食物，遗传变异、基因和环境的相互作用等有关[6]。

食管癌的临床症状包括进食后梗噎感、胸骨后疼痛、进行性吞咽困难、背痛等，早期症状隐匿，通常并不明显，当出现明显的不适时病情往往已进展至中晚期，肿瘤狭窄引起的吞咽困难是晚期食管癌患者的主要症状，吞咽困难程度与食管癌根治术后患者的生存率息息相关，因此早期诊断与治疗尤为重要[7]。目前临床上根治性切除术和淋巴结清扫术是治疗食管癌的主要方式，在尚无用于临床早期诊断 ESCA 的生物标志物之时，准确评估患者的淋巴结状态对手术预后有重要意义[8] [9]。AJCC/UICC 发布的第 8 版 TNM 分期系统仍然是全球公认的食管癌分期标准，但是由于缺乏明确的食管癌手术指南，这种分期方法容易出现偏倚，对患者的生存期造成影响。其他的淋巴结分期系统还包括 N 分期、淋巴结率(LNR)和淋巴结阳性对数几率(LODDS)等，其中 LODDS 系统不仅关注受累淋巴结的数目，还可以分析阴性淋巴结的数目，在多种肿瘤中均显示出优秀的预测价值，但其作用主要是对于食管癌患者在淋巴结转移数低于 12 枚的情况下进行分期，在淋巴结并未受侵袭的早期患者身上并不适用，并且对其的研究还较为局限，没有确定的分类标准，在临床上的适用性仍存在争议[10]-[12]。在过去的临床食管癌诊断中常常用到 CT 和 X 线，它们不仅可以显示食管腔内的情况，还可以直接地观察到病变组织的黏膜转变情况和是否存在充盈和缺损的问题，但其诊断的准确性相对较低，日常实践中食管癌诊断的假阳性率和假阴性率高，并不可靠。早期食管癌浸润黏膜层或黏膜下层时，可通过内镜下黏膜切除术(endoscopic mucosal resection, EMR)和内镜黏膜下剥离术(endoscopic submucosal dissection, ESD)进行根治[13]，尤其是 ESD，可以内镜下完全剥离病灶，对周边组织和血管的损伤少，不仅能降低对机体造成的损伤，还能降低因为病灶的残留导致复发的可能性，有研究表明接受 ESD 治疗的患者中引发并发症和复发的患者仅为 5.36%，但当疾病进展到中晚期甚至发生远处转移时，单纯手术效果不佳，患者甚至需要进行食管切除术与全身放化疗，许多患者由于化疗药物的毒性作用，为了延长生存期只能减少化疗药物剂量或提前终止化疗，但同时可能导致治疗效果不佳，反而增加了风险[14] [15]。因此，探究可用于食管癌早期诊断的标志物成为当下研究的重点方向[16]。

本研究基于生物信息学，通过 GEO 数据库(https://www.ncbi.nlm.nih.gov/geo/)和 TCGA 数据库(https://portal.gdc.cancer.gov)获得了食管癌组织基因表达数据和正常食管组织基因表达数据，筛选出癌组织与正常组织表达明显差异的基因组。进一步采用 Boruta 机器学习、Lasso 机器学习、极端梯度提升(eXtreme Gradient Boost)、随机森林(RF)和向量机递归特征消除(SVM-RFE)算法五种机器学习算法(ML)进行筛选，其中 XGBoost 模型是基于一组决策树的 ML 算法，使用梯度提升框架，通过梯度下降算法最大限度地减少错误，是硬件和软件优化技术的理想组合，可使用最少的计算资源获得卓越的结果；RF 模型可以减轻训练变化并增强模型泛化和集成，是一种 ML 分类器，可通过多个树进行训练和样本预测，得到最重要的影响因素；SVM 是分类和回归问题的线性模型，可以解决线性和非线性问题，并将数据分成几类，最终得到特征的重要性排名[17]。进一步筛选出了 ADAM12、CTHR1、IBSP、OLR1 四个基因，进行通路分析、受试者工作特征曲线(receiver operating characteristic curves analysis, ROC)分析、生存分析与单细胞分析来对基因进行验证，最终得到了运用 ADAM12、CTHR1、IBSP 和 OLR1 的表达预测食管癌疾病组与正常组均具有很高的准确性，当四个基因表达显著升高时，可能提示有食管癌的发生。

综上所述，本研究指出了 ADAM12、CTHR1、IBSP、OLR1 在食管癌的早期诊断和预后方面有重要意义，可能成为食管癌诊断的新型标志物。但本研究的结论还需要实验与检测临床样本来进行进一步的验证。

基金项目

河北北方学院大学生创新创业训练计划(项目编号：XJ10092027)。

参考文献

[1] 姚一菲, 孙可欣, 郑荣寿.《2022 全球癌症统计报告》解读: 中国与全球对比[J]. 中国普外基础与临床杂志, 2024, 31(7): 769-780.

[2] 鲍峰, 车云, 吴宗阳, 等. 以外科手术为主的局部晚期食管癌多学科治疗策略的研究进展[J]. 中国临床医学, 2023, 30(3): 547-552.

[3] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., *et al*. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA*: *A Cancer Journal for Clinicians*, **71**, 209-249. https://doi.org/10.3322/caac.21660

[4] Straum, S., Wollan, K., Rekstad, L.C. and Fossmark, R. (2024) Esophageal Cancers Missed at Upper Endoscopy in Central Norway 2004 to 2021—A Population-Based Study. *BMC Gastroenterology*, **24**, Article No. 279. https://doi.org/10.1186/s12876-024-03371-z

[5] 高艳. 食管鳞癌预后标志物的筛选及其生物信息学分析[D]: [硕士学位论文]. 乌鲁木齐: 新疆医科大学, 2020.

[6] Tu, R., Zhong, D., Li, P., Li, Y., Chen, Z., Hu, F., *et al*. (2024) Assessment of LINC-PINT Genetic Polymorphisms and Esophageal Squamous Cell Carcinoma Risk in the Hainan Han Population. *Annals of Medicine*, **56**, Article ID: 2397569. https://doi.org/10.1080/07853890.2024.2397569

[7] Fujita, T., Sato, K., Fujiwara, N., Kajiyama, D., Kubo, Y. and Daiko, H. (2024) Robot-Assisted Cervical Esophagectomy with Simultaneous Transhiatal Abdominal Procedure for Thoracic Esophageal Carcinoma. *Surgical Endoscopy*, **38**, 6413-6422. https://doi.org/10.1007/s00464-024-11214-x

[8] 李娟花, 张坤, 董亚丽. 超声内镜联合血清 SCC、CA19-9 检测对早期食管癌的诊断价值[J]. 实用癌症杂志, 2024, 39(8): 1253-1256.

[9] 吴小玲, 卢航超. 食管癌根治患者 LODDS 变化及 3 年预后状况评估[J]. 浙江创伤外科, 2023, 28(9): 1632-1635+1639.

[10] Deboever, N., Jones, C.M., Yamashita, K., Ajani, J.A. and Hofstetter, W.L. (2024) Advances in Diagnosis and Management of Cancer of the Esophagus. *BMJ*, **385**, e074962. https://doi.org/10.1136/bmj-2023-074962

[11] Wang, Z., Li, F., Zhu, M., Lu, T., Wen, L., Yang, S., *et al*. (2024) Prognostic Prediction and Comparison of Three Staging Programs for Patients with Advanced (T2-T4) Esophageal Squamous Carcinoma after Radical Resection. *Frontiers in Oncology*, **14**, Article ID: 1376527. https://doi.org/10.3389/fonc.2024.1376527

[12] 韩晖琼, 高亚萍, 王磊, 等. 食管癌根治术后患者预后模型的构建与验证: 一项基于 SEER 数据库的研究[J]. 河南医学研究, 2024, 33(16): 2892-2897.

[13] 杨轶, 赵春玲. 内镜微创手术对早期食管癌术后免疫功能及复发的影响[J]. 实用癌症杂志, 2024, 39(9): 1548-1550.

[14] Tsuji, T., Matsuda, S., Sato, Y., Tanaka, K., Sasaki, K., Watanabe, M., *et al*. (2024) Safety and Efficacy of Conversion Therapy after Systemic Chemotherapy in Advanced Esophageal Cancer with Distant Metastases: A Multicenter Retrospective Observational Study. *Annals of Surgical Oncology*, **32**, 274-283. https://doi.org/10.1245/s10434-024-16196-7

[15] Jiang, L., Zhu, J., Chen, X., Wang, Y., Wu, L., Wan, G., *et al*. (2024) Reduction in Chemotherapy Relative Dose Intensity Decreases Overall Survival of Neoadjuvant Chemoradiotherapy in Patients with Locally Advanced Esophageal Carcinoma. *BMC Cancer*, **24**, Article No. 945. https://doi.org/10.1186/s12885-024-12724-6

[16] Hao, J., Liu, W., Zhao, C. and Xia, T. (2021) The Diagnostic Significance of 64-Slice Spiral CT Combined with Serological CA19-9, Bcl-2, CYFRA21-1 Detection in Thoracic Esophageal Carcinoma. *Translational Cancer Research*, **10**, 5383-5389. https://doi.org/10.21037/tcr-21-2522

[17] Zhou, L., Wang, Y., Zhu, W., Zhao, Y., Yu, Y., Hu, Q., *et al*. (2024) A Retrospective Study Differentiating Nontuberculous Mycobacterial Pulmonary Disease from Pulmonary Tuberculosis on Computed Tomography Using Radiomics and Machine Learning Algorithms. *Annals of Medicine*, **56**, Article ID: 2401613. https://doi.org/10.1080/07853890.2024.2401613