doi:10.7659/j.issn.1005-6947.2021.03.005
http://dx.doi.org/10.7659/j.issn.1005-6947.2021.03.005
Chinese Journal of General Surgery, 2021, 30(3):276-285.

·专题研究·

基于生物信息学胰腺腺癌关键基因的筛选及支持向量机诊断模型的构建

张波[1]，徐涛[1]，徐浩[2]，夏雨[1]，周文策[1, 2]

（1. 兰州大学第一临床医学院，甘肃 兰州 730000；2. 兰州大学第一医院 普通外科，甘肃 兰州 730000）

摘　要

背景与目的：胰腺癌是一种常见的消化道恶性肿瘤，其主要病理类型为胰腺腺癌（PAAD），因早期诊断困难且缺乏有效的治疗措施，故预后极差。因此，寻找 PAAD 的诊治新靶标具有重要意义。本研究通过生物信息学方法筛选与 PAAD 诊断和预后相关的关键基因，构建分类 PAAD 样本和正常样本的支持向量机（SVM）模型，以期为 PAAD 的诊治及机制研究提供依据。

方法：从基因表达数据库（GEO）中下载 3 个芯片数据（GSE28735、GSE62165、GSE62452），应用 R 语言的 Limma 包筛选出 PAAD 组织和正常组织间的差异表达基因（DEGs）。利用 STRING 数据库对 DEGs 进行 GO 和 KEGG 通路富集分析。再以 STRING 数据库构建 DEGs 的蛋白互作网络（PPI），利用 Cytoscape 软件进行可视化编辑，并通过 MCODE 插件进行关键子网络分析。使用 R 语言的 survival 包筛选 PPI 和关键子网络中与预后相关的关键节点，将其上传至 Metascape 进行功能富集分析。利用 R 语言 caret 包中递归式特征消除（RFE）算法筛选关键节点中的最优特征基因，在 GEPIA 数据库中验证最优特征基因的表达差异，随后通过 R 语言的 e1071 包构建最优特征基因的 SVM 模型，并在 3 个芯片数据中借助 R 语言的 pROC 包对该模型进行验证。在 TCGA 数据库中，用 R 语言的 survminer 包筛选出最优特征基因中与 PAAD 预后相关的基因作为关键基因。

结果：共筛选出 257 个 DEGs，包括 168 个上调基因和 89 个下调基因。GO 分析结果表明 DEGs 主要参与细胞外基质的组成、细胞黏附、丝氨酸肽酶活性等生物学过程。KEGG 分析显示，DEGs 主要富集于蛋白质的消化和吸收、胰腺的分泌、黏着斑、PI3K-Akt 信号通路。生存分析筛选出 14 个关键节点同时在 GSE28735 和 GSE62452 中与预后相关（均 $P<0.05$），这些基因在肿瘤侵犯和肿瘤发生中发挥一定作用。RFE 筛选出 8 个最优特征基因：LAMA3、FN1、ITGA3、MET、PLAU、CENPF、MMP14、OAS2；GEPIA 数据库验证发现这 8 个最优特征基因在 PAAD 组织中明显上调（均 $P<0.01$）；这些基因构建的 SVM 模型在 3 个芯片数据中 ROC 曲线的 AUC 依次为 0.898、1.000、0.905。TCGA 数据库验证发现 LAMA3、ITGA3、MET、PLAU、CENPF 及 OAS2 的上调与 PAAD 预后不良有关（均 $P<0.05$）。

结论：关键基因 LAMA3、ITGA3、MET、PLAU、CENPF 及 OAS2 可能成为 PAAD 诊治的新靶点；基于 8 个最优特征基因构建的 SVM 模型可有效诊断 PAAD。

关键词　胰腺肿瘤；基因表达谱；支持向量机；计算生物学

中图分类号：R735.9

基金项目：甘肃省重点研发计划基金资助项目（17YF1FA128）；甘肃省兰州市人才创新创业基金资助项目（2017-RC-37）。

收稿日期：2020-10-29；**修订日期：**2021-02-12。

作者简介：张波，兰州大学第一临床医学院住院医师，主要从事肝胆胰疾病的临床和基础方面的研究。

通信作者：周文策，Email: zhouwc@lzu.edu.cn

Identification of hub genes in pancreatic adenocarcinoma and construction of a support vector machine diagnostic classifier based on bioinformatics approaches

ZHANG Bo[1], XU Tao[1], XU Hao[2], XIA Yu[1], ZHOU Wence[1, 2]

(1. The First Clinical Medical College, Lanzhou University, Lanzhou 730000, China; 2. Department of General Surgery, the First Hospital of Lanzhou University, Lanzhou 730000, China)

Abstract

Background and Aims: Pancreatic cancer is a common malignant tumor of the digestive tract. Its main pathological type is pancreatic adenocarcinoma (PAAD). Due to the difficulty of early diagnosis and lack of effective treatment measures, the prognosis of PAAD is extremely poor. Therefore, defining new targets for the diagnosis and treatment of PAAD is of great significance. This study was conducted to screen the hub genes related to the diagnosis and prognosis of PAAD by bioinformatics analysis, and then construct a support vector machine (SVM) model to classify PAAD and normal pancreatic samples, so as to provide a useful resource for researches in terms of diagnosis, treatment and mechanism of PAAD.

Methods: Three microarray datasets (GSE28735, GSE62165, GSE62452) were downloaded from the Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) between PAAD tissue and normal pancreatic tissue were screened using Limma package of R language. GO and KEGG pathway enrichment analysis of the DEGs were performed using STRING database. Then, protein-protein interaction networks (PPI) of the DEGs were generated using the STRING server and visualized by Cytoscape software. Key subnetwork module analyses were performed through MCODE plug-in. R language survival package was used to screen the key nodes related to prognosis in PPI and key subnetworks, and then, the key nodes were uploaded to Metascape for function enrichment analysis. The recursive feature elimination (RFE) algorithm in caret package of R language was used to select the optimal feature genes in key nodes, and the expression differences of the optimal feature genes were verified in GEPIA database. A SVM classifier based on the optimal feature genes was constructed using the R language e1071 package, and the R language pROC package was used to verify the model in the 3 microarray datasets. In the TCGA database, the R package survminer was used to select the genes related to the prognosis of PAAD among the optimal feature genes as the hub genes.

Results: A total of 257 DEGs were screened, including 168 up-regulated genes and 89 down-regulated genes. GO analysis showed that DEGs were mainly involved in biological processes such as the extracellular matrix organization, cell adhesion, serine-type peptidase activity. KEGG analysis showed that DEGs were mainly enriched in protein digestion and absorption, pancreatic secretion, focal adhesion and PI3K-Akt signaling pathway. Survival analysis showed that 14 key nodes were associated with the prognosis in both GSE28735 and GSE62452 (all $P<0.05$), and these genes played a certain role in neoplasm invasiveness and oncogenesis. RFE screened out 8 optimal feature genes: LAMA3, FN1, ITGA3, MET, PLAU, CENPF, MMP14, and OAS2; GEPIA database validation found that the 8 optimal feature genes were significantly up-regulated in PAAD tissues (all $P<0.01$). The AUC of ROC curve of the SVM model constructed by these genes in the 3 microarray datasets were 0.898, 1.000 and 0.905, respectively. TCGA database verification found that the up-regulations of LAMA3, ITGA3, MET, PLAU, CENPF and OAS2 were associated with poor prognosis of PAAD (all $P<0.05$).

Conclusion: The hub genes LAMA3, ITGA3, MET, PLAU, CENPF and OAS2 may be new targets for diagnosis or treatment of PAAD. The SVM model based on 8 optimal feature genes offers an effective tool for diagnosing PAAD.

Key words Pancreatic Neoplasms; Gene Expression Profiling; Support Vector Machine; Computational Biology

CLC number: R735.9

胰腺腺癌（pancreatic adenocarcinoma，PAAD）是最常见的胰腺肿瘤，占所有胰腺癌的绝大多数；由于其早期诊断困难，超过52%的患者发现时已有远处转移，大多数患者确诊在晚期，丧失了手术机会，对放化疗敏感度又差，5年生存率不足10%[1-3]。因此，寻找PAAD早期诊断和治疗的新靶点是近年来的研究热点。

基因改变对PAAD的发病至关重要，研究PAAD发生的分子机制是延长患者总生存期的关键。然而，目前尚未发现对胰腺癌敏感度强、特异度高的肿瘤标记物，也没有找到有效的治疗靶点[4]。单基因对肿瘤的诊断效果多不理想，局限性较大，多基因联合检测有望在诊断领域开辟新方向。支持向量机（support vector machine，SVM）是机器学习领域中一种重要的分类算法，它通过估计每个样本的内部联系和规则来判别样本类型，具有较高的分类精度[5]。随着生物信息技术的发展，基因芯片可快捷、高效地分析多个基因联合检测对肿瘤的诊断效能。为了探究PAAD的发病机理，本研究从基因表达数据库（Gene Expression Omnibus，GEO）中下载了3个芯片，筛选PAAD组织和正常组织之间的差异表达基因（differentially expression genes，DEGs），对这些DEGs进行基因本体论（Gene Ontology，GO）富集和京都基因与基因组百科全书（Kyoto Encyclopedia of Genes and Genomes，KEGG）信号通路富集分析，并通过蛋白互作网络（protein-protein interaction network，PPI）分析和关键子网络分析筛选出候选基因，探讨可能与PAAD诊断和预后相关的关键基因，并构建分类PAAD和正常样本的SVM模型，为以后研究PAAD的诊治和分子机制提供理论依据。

1 资料与方法

1.1 原始数据下载及预处理

从GEO数据库中下载3个PAAD基因表达谱芯片（GSE28735、GSE62165、GSE62452）；数据集GSE28735和GSE62452基于GPL6244 Affymetrix Human Gene 1.0 ST Array，而GSE62165芯片数据在GPL13667 Affymetrix Human Genome U219 Array平台上注释。其中GSE28735包括45例PAAD样本和45例正常组织样本；GSE62452含有69例PAAD样本和61例正常组织样本；GSE62165包括118例PAAD样本和13例正常组织样本。整理数据集中的样本分组及临床相关信息，并对表达谱按芯片注释信息进行重注释。从UCSC Xena下载得到癌症基因组图谱（The Cancer Genome Atlas，TCGA）数据库中胰腺癌的生存数据和基因表达谱，并对表达谱进行Log2（count+1）的标准化处理，以消除量纲差异。

1.2 DEGs的筛选

芯片数据预处理后，采用R语言Limma包[6]进行PAAD组织和正常组织的差异基因分析，筛选标准为|log FC|>1和P<0.05；通过Venn图得到3个芯片交集的DEGs，其中在3个芯片中均上调的基因作为上调DEGs，均下调的基因作为下调DEGs。

1.3 DEGs的GO分析和KEGG分析

将得到的DEGs上传至STRING数据库（http://www.string-db.org）中进行GO分析（包括生物学过程BP、细胞组分CC、分子功能MF）和KEGG通路富集分析，分析参数采用数据库默认，并将TOP10的显著富集结果以气泡图进行可视化。

1.4 DEGs的PPI分析及关键子网络分析

利用STRING数据库对DEGs进行PPI分析[7]，蛋白互作的combined score>0.4作为阈值条件，再通过Cytoscape软件（http://cytoscape.org）对该网络进行可视化[8]。利用Cytoscape软件中的MCODE插件识别该PPI网络中的关键子网络，参数设置为：Degree Cutoff=2，Node Score Cutoff=0.2，Max. Depth=100，K-Core=2，MCODE score≥5。

1.5 关键节点的筛选及其功能富集分析

数据集GSE28735和GSE62452的临床信息中包含生存信息，将所有关键子网络中包含的基因以及PPI分析结果中度值（degree）≥15的基因，分别在两个数据集中使用R语言的survival包进行生存分析，将在两个数据集中都具有显著生存意义（P<0.05）的基因作为关键节点。将得到的关键节点上传至Metascape数据库[9]进行功能富集分析。

1.6 最优特征基因的筛选和及其表达验证

递归特征消除（recursive feature elimination，RFE）是一种贪婪的算法，通过重复构建模型筛选出分类的最佳基因组合[10]。为了进

一步缩小候选基因的范围，准确识别最优特征基因，将数据集GSE28735作为训练集，其他两个数据集作为验证集。在训练集中，利用R语言caret包中的RFE算法从关键节点里筛选出最优特征基因组合。在10倍交叉验证中，以均方根误差（RMSE）最小、准确率最高的基因组合为最佳基因组合。由于TCGA数据库中正常胰腺样本较少，故在GEPIA[11]数据库对最优特征基因的差异表达进行验证，筛选条件为$|\log_2 FC|>1$和$P<0.01$。

1.7 SVM 模型的构建及验证

为探索8个最优特征基因联合检测在分类PAAD和正常样本中的作用，使用 R 语言 e1071 包[12]构建基于这些基因的SVM模型，最终选择参数为SVM-Type：eps-regression；SVM-Kernel：radial；cost：1；gamma：0.125；epsilon：0.1，并在训练集GSE28735和验证集（GSE62165和GSE62452）中采用R 语言 pROC包[13]绘制受试者工作特征（receiver operating characteristic，ROC）曲线对该模型进行验证。

1.8 最优特征基因的预后分析

将TCGA数据库中的PAAD的样本挑选出来，剔除正常样本，对缺乏生存信息的样本删除，使用R语言survminer包[14]计算每个最优特征基因的最佳截断值（optimal cutoff），对于mRNA表达值>optimal cutoff视为高表达，≤optimal cutoff作为低表达，结合生存时间和生存状态进行生存分析，以$P<0.05$为差异有统计学意义。

2 结　果

2.1 DEGs 的筛选

通过对GSE28735、GSE62452、GSE62165数据集的分析，分别鉴定出388个（243个上调，145个下调）、283个（185个上调，98个下调）和3063个（1 993个上调，1 070个下调）差异基因。韦恩图分析发现257个DEGs均在3个数据集中差异表达，其中168个为上调DEGs，89个是下调DEGs（图1）。

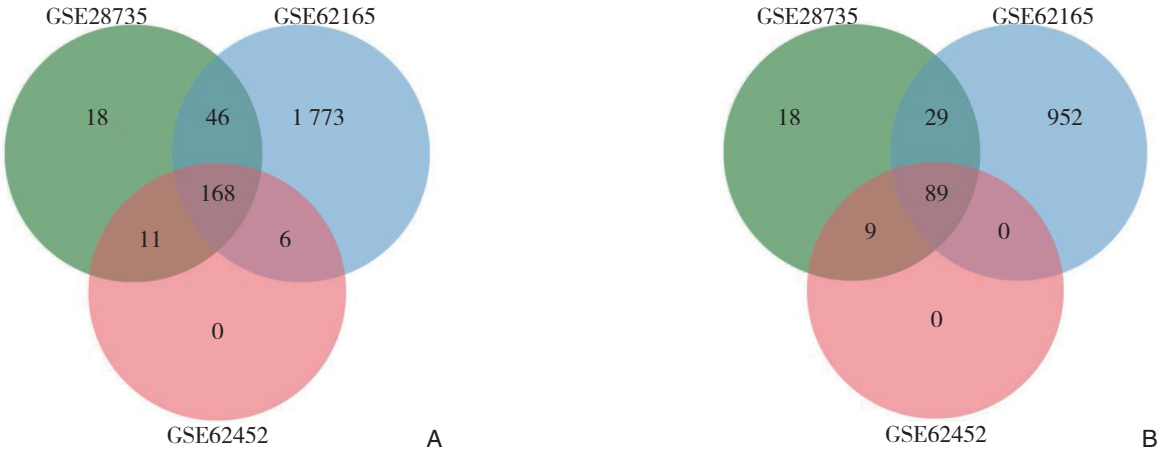

图 1　DEGs 的 Venn 图　　A：上调 DEGs；B：下调 DEGs

Figure 1　Venn diagram of DEGs　　A: Up-regulated DEGs; B: Down-regulated DEGs

2.2 DEGs 的功能和通路富集分析

GO富集分析结果显示：257个DEGs主要参与BP的细胞外基质的组成、细胞外结构组成、细胞黏附、组织发育等过程；CC主要聚集于细胞外区、细胞外区域部分、细胞外间隙等；MF主要与丝氨酸肽酶活性、丝氨酸内肽酶活性、肽酶活性等有关。KEGG通路分析发现，DEGs主要参与蛋白质的消化和吸收、胰腺的分泌、细胞外基质受体相互作用、黏着斑、PI3K-Akt信号通路等（均$P<0.05$）（图2）。

2.3 PPI 及关键子网络分析

通过STRING数据库构建257个DEGs的PPI网络，使用Cytoscape进行可视化，该PPI网络共有210个节点和822条边（图3）。在PPI 网络中筛选出29个度值≥15的重要基因（表1）。利用MCODE插件筛选出4个关键子网络（图4）。结合关键子网络，发现一些新的在PAAD的进展中起调控作用的基因，如COL5A2、OAS2、DDX60、CELA2A，这为以后研究PAAD的分子机制提供更多的依据。

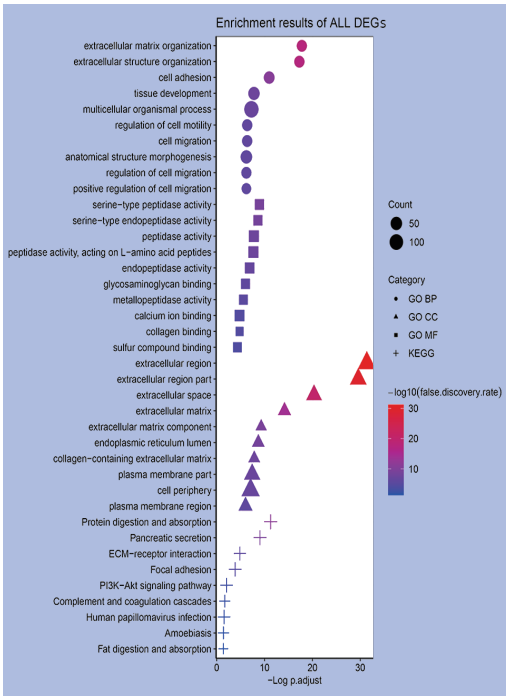

图 2 257 个 DEGs 的功能和通路富集气泡图

Figure 2 Function and pathway enrichment bubble graph of 257 DEGs

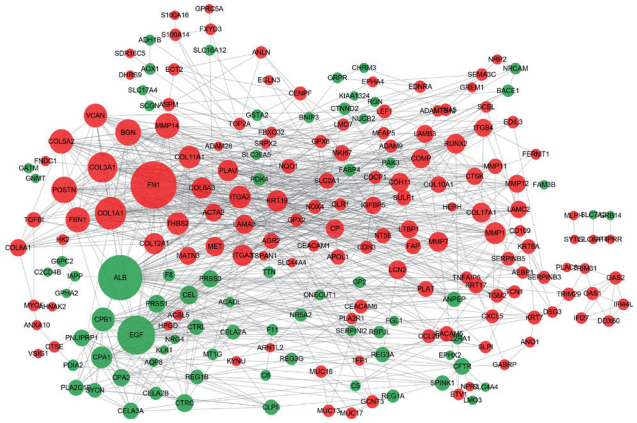

图 3 PPI 图（红色为上调基因，绿色为下调基因）

Figure 3 PPI (red color showing the up-regulated genes, and green color showing the down-regulated genes)

表 1 筛选出度值≥ 15 的 29 个基因

Table 1 29 screened genes with a degree ≥ 15

基因	度值	表达	基因	度值	表达
FN1	58	上调	CPA1	19	下调
ALB	55	下调	MMP1	19	上调
EGF	45	下调	COL6A3	18	上调
COL1A1	35	上调	ITGA3	18	上调
COL3A1	32	上调	COL12A1	18	上调
BGN	26	上调	PLAU	18	上调
POSTEN	26	上调	KRT19	18	上调
FBN1	24	上调	COL17A1	16	上调
MMP14	24	上调	MMP7	16	上调
COL5A2	23	上调	RUNX2	16	上调
VCAN	22	上调	COMP	15	上调
CPB1	22	下调	LAMA3	15	上调
ITGA2	20	上调	MET	15	上调
COL11A1	20	上调	CDH11	15	上调
THBS2	20	上调			

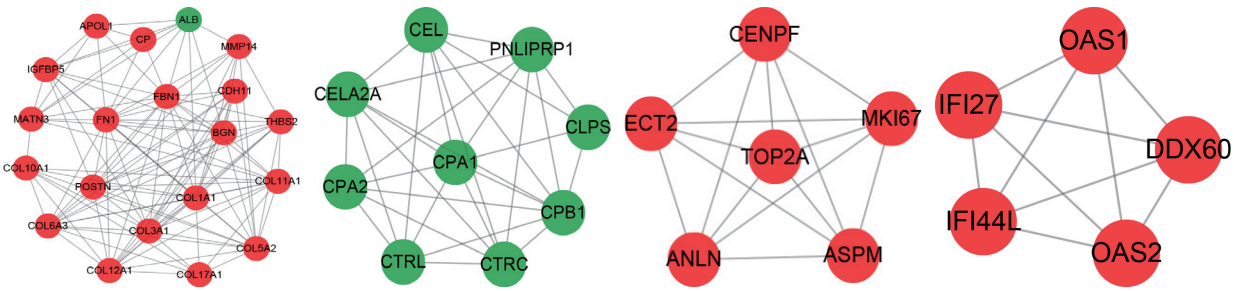

图 4 关键子网络分析（包括 4 个模块）

Figure 4 Key sub-module analysis of the PPI (including 4 modules)

2.4 关键节点及其功能富集分析

4个关键子网络包含的所有基因及PPI网络中degree≥15的基因共有52个。将这52 个基因分别在数据集 GSE62452和GSE28735中进行生存分析，分别得到26个和22个与PAAD预后相关的基因（P<0.05），其中14个基因（ANLN、BGN、CENPF、DDX60、FN1、IFI44L、ITGA3、LAMA3、MET、MKI67、MMP14、OAS2、PLAU、TOP2A）在两个数据集中均与预后相关（图5A）。利用Metascape对这14个关键节点进行功能富集分析，发现这些基因在肿瘤侵犯和肿瘤发生中发挥一定作用（图5B）。

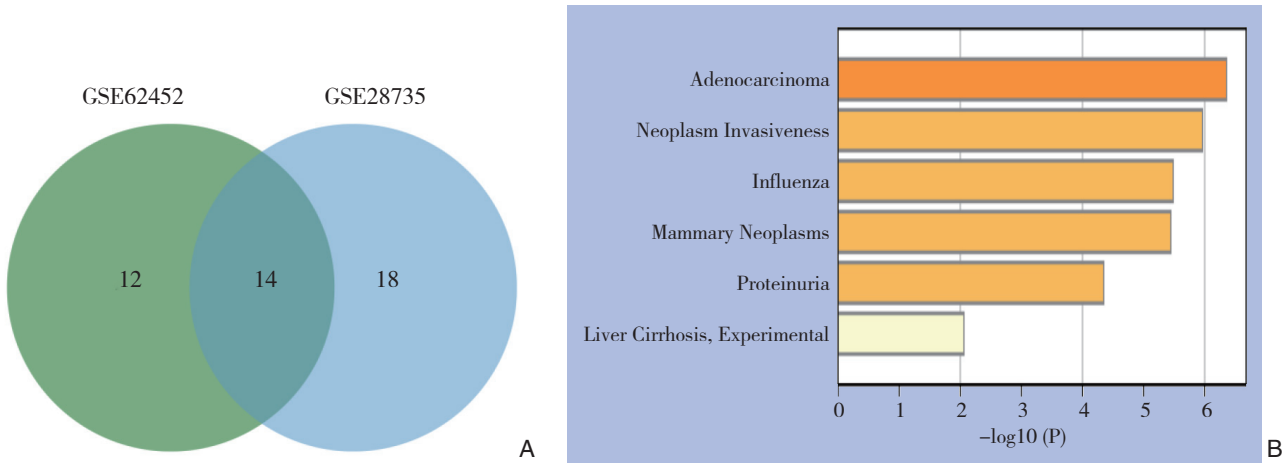

图 5　关键节点的筛选及其功能富集分析　A：Venn 示 GSE62452 和 GSE28735 数据集中与预后相关基因的交叉验证结果；B：Metascape 对 14 个关键节点进行的功能和通路富集分析结果

Figure 5　Screening of key nodes and function enrichment analysis　A: The Venn diagram showing the result of cross-validation of prognostic related genes in the GSE62452 and GSE28735 datasets; B: Function and pathway enrichment analysis of 14 key nodes by using Metascape

2.5 最优特征基因及其表达验证

在最优参数（最小RMSE=0.3429，最大准确度=0.8694）下，用RFE算法筛选出8个最优特征基因：LAMA3、FN1、ITGA3、MET、PLAU、CENPF、MMP14、OAS2（图6）。GEPIA数据库验证发现这8个最优特征基因在PAAD组织中的表达均高于正常组织，差异有统计学意义（均P<0.01）。

2.6 SVM 模型及其验证

通过R语言e1071包构建8个最优特征基因诊断PAAD的SVM建模如图7所示，该模型在训练集GSE28735中的曲线下面积（AUC）为0.898，灵敏度和特异度均为0.911；在验证集GSE62452中的AUC为0.905，敏感度为0.855，特异度为0.918；在验证集GSE62165中的AUC为1，敏感度和特异度均为1；表明该SVM模型可以较好地区分PAAD和正常样本。

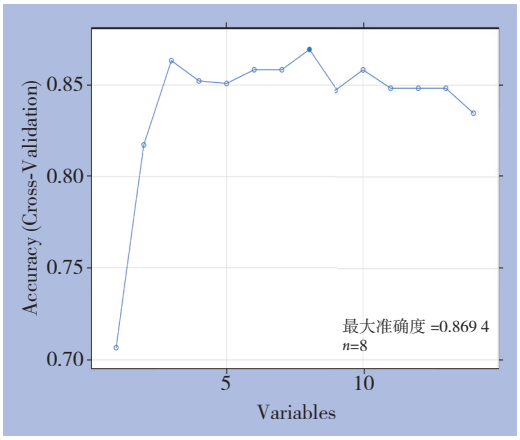

图 6　最优特征基因组合的准确度曲线（横轴代表基因变量的数量，纵轴代表交叉验证的准确性，点标记是最优特征基因组合对应的基因数）

Figure 6　Accuracy curve for screening the optimal feature gene combination (the horizontal axis representing the number of gene variables, the vertical axis representing the cross-validation accuracy, and the marked content standing for the number of genes corresponding to the optimal gene combination)

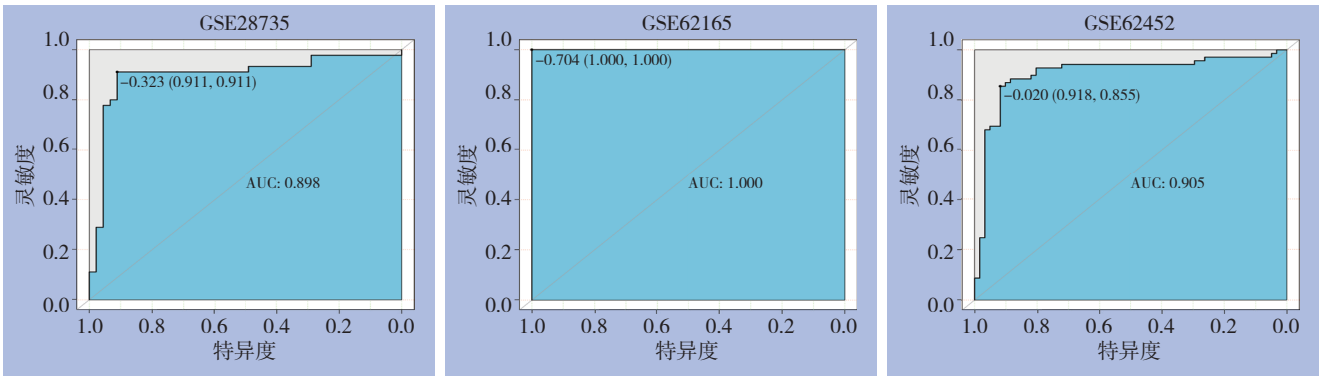

图 7　训练集（GSE28735）及验证集（GSE62165 和 GSE62452）对 SVM 模型的 ROC 曲线

Figure 7　ROC curves of the training set (GSE28735) and the validation sets (GSE62165 and GSE62452) on the SVM classifier

2.7　最优特征基因的生存分析

从TCGA数据库中筛选出178例PAAD样本，Kaplan-Meier生存曲线发现，与高表达LAMA3、ITGA3、MET、PLAU、CENPF及OAS2组的PAAD患者相比，这些基因的低表达组总生存率更长（均P<0.05）（图8），而FN1、MMP14上调对PAAD的生存率无明显影响（均P>0.05）。

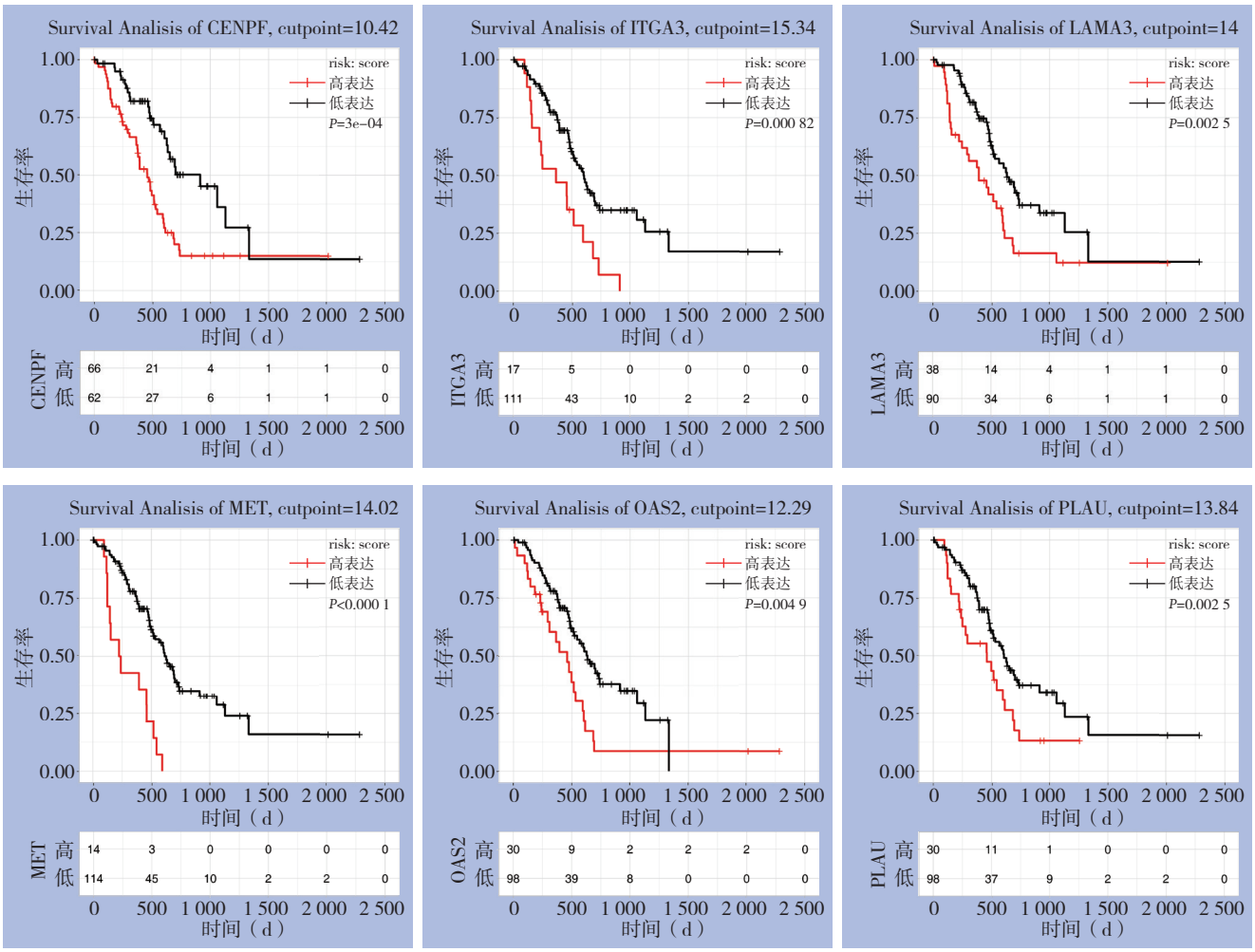

图 8　基于最佳截断值进行的 TCGA 中最优特征基因与 PAAD 关系的生存分析

Figure 8　Survival analysis of the relationship between optimal feature genes and PAAD in TCGA database based on the optimal cutoff

3　讨　论

尽管PAAD的治疗已经取得了一些进展，但由于缺乏准确诊断PAAD的特异度生物标志物及个体化治疗的有效靶点，PAAD仍然是一种难治性癌症[15]。探索PAAD的分子机制，找出潜在的诊断及预后的生物标志物，具有重要意义。在本研究中，从GSE28735、GSE62165及GSE62452数据集中筛选出257个在胰腺癌和正常组织之间的DEGs，其中168个上调DEGs和89个下调DEGs。通过GO和KEGG通路分析，发现这些DEGs主要参与的细胞外基质[16]、细胞黏附[17]、细胞迁移[18]、胰腺的分泌[19]、PI3K-Akt信号通路[20]已被证明与PAAD的发生和发展密切相关，表明本研究筛选的DEGs可靠性高。PPI预测和关键子网络分析发现一些新的在PAAD进展中起调控作用的基因，如COL5A2、OAS2、DDX60、CELA2A。既往研究发现这些基因与疾病的发生或治疗密切相关[21-24]，如DDX60的表达可调节乳腺癌患者对放疗的敏感度[23]，CELA2A突变与代谢综合征的发生相关[24]等；但他们在PAAD中尚无相关研究，这为以后研究PAAD的分子机制提供了参考。数据集中的生存分析筛选到14个与生存相关的关键节点，这些基因参与肿瘤的发生和侵犯，值得进一步研究。由递归特征消除算法筛选到8个最优特征基因，经GEPIA数据库证实这8个最优特征基因在PAAD组织中的表达高于正常组织，与芯片结果一致。鉴于单个靶点对肿瘤诊断价值有限，本研究通过R语言的e1071包对8个最优特征基因构建SVM模型，ROC曲线验证发现3个数据集的AUC均在0.85以上，敏感度和特异度高。据我们所知，用于PAAD诊断的SVM模型以前很少有报道，本研究构建的SVM模型可有效区分PAAD样本和正常样本，有望用于临床实践。

在TCGA数据库中验证8个最优基因的预后价值，其中CENPF、ITGA3、LAMA3、MET、OAS2、PLAU与PAAD患者的预后有关，即基因高表达者预后差、基因低表达者总生存率更长；该6个基因被视为关键基因。既往研究表明，ITGA3、LAMA3、CENPF在胰腺癌组织中高表达，且这些基因高表达的胰腺癌患者预后更差[25-27]，这与本研究结果一致。Jiao等[25]证实，ITGA3对胰腺癌的早期诊断有一定价值，其表达水平与组织学类型、生存状态以及复发相关。肿瘤间质相互作用是肿瘤治疗的重要靶点。研究[28]表明ITGA3是PAAD肿瘤间质相互作用的靶点之一，靶向ITGA3可能是PAAD治疗的潜在方法。Yang等[26]指出LAMA3不仅与PAAD患者的预后有关，还有诊断PAAD的潜力。研究[29]发现LAMA3的沉默可抑制胰腺癌细胞的增殖、迁移和侵袭，并促进细胞凋亡，是PAAD潜在的治疗靶点。Chen等[27]发现CENPF的表达可促进胰腺癌细胞的增殖、迁移、上皮间质转化（EMT）及有丝分裂G2/M的转化，这种调节与TNF信号通路有关。PLAU可促进细胞外基质降解，参与细胞的侵袭和迁移。雷公藤甲素通过下调PLAU抑制胰腺癌细胞增殖和迁移，在此过程中，EMT信号通路被激活[30]。MET是受体酪氨酸激酶家族成员，包含MET的9个基因组合可有效预测PAAD患者的预后[31]。然而，这些关键基因在PAAD中的具体作用机制尚不清楚，仍需进一步研究。OAS2是2'，5'-寡腺苷酸合成酶中一员，在口腔鳞状细胞癌中高表达，与患者预后呈负相关[32]；而在乳腺癌中，OAS2高表达的患者预后较好[22]，这与本研究OAS2对PAAD的预后存异。可见，OAS2在不同肿瘤中的作用并不一致，目前缺乏OAS2在PAAD中的研究，需要进一步实验来研究OAS2在PAAD中的具体作用。

本研究通过生物信息学分析寻找可能参与PAAD发病的DEGs，通过对这些DEGs进行分析，以更好地了解PAAD致癌的机制。研究结果显示，COL5A2、OAS2、DDX60、CELA2A为探究PAAD的分子机制提供新路经；关键基因LAMA3、ITGA3、MET、PLAU、CENPF、OAS2可能成为PAAD诊断或治疗的新靶点；基于8个最优特征基因构建的SVM模型可有效诊断PAAD。本研究为今后研究PAAD的分子机制、生物标志物及治疗靶点提供了新的思路。然而，该研究的主要局限性是缺乏基础实验来证实这些结果的真实性，分析结果的可靠性严重依托于本报告中涉及数据集的准确性。未来的研究应通过基础实验和临床实践来验证本研究的结果。

参考文献

[1]　Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021[J].

CA Cancer J Clin, 2021, 71(1):7–33. doi: 10.3322/caac.21654.

[2] Mizrahi JD, Surana R, Valle JW, et al. Pancreatic cancer[J]. Lancet, 2020, 395(10242):2008–2020. doi: 10.1016/S0140–6736(20)30974–0.

[3] 邹蔡峰, 傅德良. 胰腺癌的新辅助治疗[J]. 中国普通外科杂志, 2020, 29(3):260–267. doi:10.7659/j.issn. 1005–6947.2020.03.002.
Zou CF, Fu DL. Neoadjuvant therapy for pancreatic carcinoma[J]. Chinese Journal of General Surgery, 2020, 29(3):260–267. doi:10.7659/j.issn.1005–6947.2020.03.002.

[4] Sinha V, Shinde S, Saxena S, et al. A Comprehensive Review of Diagnostic and Therapeutic Strategies for the Management of Pancreatic Cancer[J]. Crit Rev Oncog, 2020, 25(4):381–404. doi: 10.1615/CritRevOncog.2020035971.

[5] Daliri MR. Feature selection using binary particle swarm optimization and support vector machines for medical diagnosis[J]. Biomed Tech (Berl), 2012, 57(5):395–402. doi: 10.1515/bmt-2012–0009.

[6] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. Nucleic Acids Res, 2015, 43(7):e47. doi: 10.1093/nar/gkv007.

[7] Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible[J]. Nucleic Acids Res, 2017, 45(D1):D362–D368. doi: 10.1093/nar/gkw937.

[8] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. Genome Res, 2003, 13(11):2498–2504. doi: 10.1101/gr.1239303.

[9] Zhou YY, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets[J]. Nat Commun, 2019, 10(1):1523. doi: 10.1038/s41467–019–09234–6.

[10] Lu X, Yang Y, Wu F, Discriminative analysis of schizophrenia using support vector machine and recursive feature elimination on structural MRI images[J]. Medicine (Baltimore), 2016, 95(30):e3973. doi: 10.1097/MD.0000000000003973.

[11] 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):W98–102. doi: 10.1093/nar/gkx247.

[12] Meyer D. Support Vector Machines. The Interface to libsvm in package e1071[CP]. Austria: Technische Universität Wien, 2010, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.161.737&rep=rep1&type=pdf.

[13] Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves[J]. BMC Bioinformatics, 2011, 12:77. doi: 10.1186/1471–2105–12–77.

[14] Kassambara A, Kosinski M, Biecek P. Survminer: Drawing Survival Curves Using ‘ggplot2’. 2019. R Package Version 0.4.6[CP]. Available online: https://CRAN.R-project.org/package=survminer (accessed on 20 April 2020).

[15] 范红星, 倪建勋, 薄彪, 等. 术前白蛋白-胆红素评分以及其与CA19–9联合作为胰腺癌患者预后指标的临床价值[J]. 中国普通外科杂志, 2020, 29(3):310–316. doi:10.7659/j.issn. 1005–6947.2020.03.008.
Fan HX, Ni JX, Bo B, et al. Clinical value of preoperative albumin-bilirubin score and its combination with CA19–9 as prognostic indicators for pancreatic cancer patients[J]. Chinese Journal of General Surgery, 2020, 29(3):310–316. doi:10.7659/j.issn.1005–6947.2020.03.008.

[16] Zeltz C, Primac I, Erusappan P, et al. Cancer-associated fibroblasts in desmoplastic tumors: emerging role of integrins[J]. Semin Cancer Biol, 2020, 62:166–181. doi: 10.1016/j.semcancer.2019.08.004.

[17] Osipov A, Blair AB, Liberto J, et al. Inhibition of focal adhesion kinase enhances antitumor response of radiation therapy in pancreatic cancer through CD8+ T cells[J]. Cancer Biol Med, 2021, 18(1):206–214. doi: 10.20892/j.issn.2095–3941.2020.0273.

[18] Guo YY, Tong Y, Zhu HY, et al. Quercetin suppresses pancreatic ductal adenocarcinoma progression via inhibition of SHH and TGF-β/Smad signaling pathways[J]. Cell Biol Toxicol, 2020. doi: 10.1007/s10565–020–09562–0. [Online ahead of print]

[19] Nakamura S, Sadakari Y, Ohtsuka T, et al. Pancreatic Juice Exosomal MicroRNAs as Biomarkers for Detection of Pancreatic Ductal Adenocarcinoma[J]. Ann Surg Oncol, 2019, 26(7):2104–2111. doi: 10.1245/s10434–019–07269-z.

[20] Nweke E, Ntwasa M, Brand M, et al. Increased expression of plakoglobin is associated with upregulated MAPK and PI3K/AKT signalling pathways in early resectable pancreatic ductal adenocarcinoma[J]. Oncol Lett, 2020, 19(6):4133–4141. doi: 10.3892/ol.2020.11473.

[21] Kohrt SE, Awadallah WN, Phillips RA 3rd, et al. Identification of Genes Required for Enzalutamide Resistance in Castration-Resistant Prostate Cancer Cells[J]. Mol Cancer Ther, 2021, 20(2):398–409. doi: 10.1158/1535–7163.MCT-20–0244.

[22] Zhang YJ, Yu CR. Prognostic characterization of OAS1/OAS2/OAS3/OASL in breast cancer[J]. BMC Cancer, 2020, 20(1):575. doi: 10.1186/s12885–020–07034–6.

[23] Xin DG, Liu JF, Gu JC, et al. Low Expression of DDX60 Gene Might Associate with the Radiosensitivity for Patients with Breast Cancer[J]. J Oncol, 2020, 2020:8309492. doi: 10.1155/2020/8309492.

[24] Esteghamat F, Broughton JS, Smith E, et al. CELA2A mutations predispose to early-onset atherosclerosis and metabolic syndrome and affect plasma insulin and platelet activation[J]. Nat Genet, 2019, 51(8):1233–1243. doi: 10.1038/s41588–019–0470–3.

[25] Jiao Y, Li YQ, Liu SY, et al. ITGA3 serves as a diagnostic and prognostic biomarker for pancreatic cancer[J]. Onco Targets Ther, 2019, 12:4141–4152. doi: 10.2147/OTT.S201675.

[26] Yang CK, Liu ZQ, Zeng XM, et al. Evaluation of the diagnostic ability of laminin gene family for pancreatic ductal adenocarcinoma[J]. Aging (Albany NY), 2019, 11(11):3679–3703. doi: 10.18632/aging.102007.

[27] Chen HJ, Wang XM, Wu FB, et al. Centromere protein F is identified as a novel therapeutic target by genomics profile and contributing to the progression of pancreatic cancer[J]. Genomics, 2020, 113(1 Pt 2):1087–1095. doi: 10.1016/j.ygeno.2020.10.039.

[28] Hiroshima Y, Kasajima R, Kimura Y, et al. Novel targets identified by integrated cancer-stromal interactome analysis of pancreatic adenocarcinoma[J]. Cancer Lett, 2020, 469:217–227.doi: 10.1016/j.canlet.2019.10.031.

[29] Huang CQ, Chen J. Laminin-332 mediates proliferation, apoptosis, invasion, migration and epithelial-to-mesenchymal transition in pancreatic ductal adenocarcinoma[J]. Mol Med Rep, 2021, 23(1):1. doi: 10.3892/mmr.2020.11649.

[30] Zhao X, Liu Z, Ren ZY, et al. Triptolide inhibits pancreatic cancer cell proliferation and migration via down-regulating PLAU based on network pharmacology of Tripterygium wilfordii Hook F[J]. Eur J Pharmacol, 2020, 880:173225. doi: 10.1016/j.ejphar.2020.173225.

[31] Wu MM, Li XB, Zhang TP, et al. Identification of a Nine-Gene Signature and Establishment of a Prognostic Nomogram Predicting Overall Survival of Pancreatic Cancer[J]. Front Oncol, 2019, 9:996. doi: 10.3389/fonc.2019.00996.

[32] Wang JY, Wang YY, Kong FZ, et al. Identification of a six-gene prognostic signature for oral squamous cell carcinoma[J]. J Cell Physiol, 2020, 235(3):3056–3068. doi: 10.1002/jcp.29210.

（本文编辑　宋涛）

本文引用格式：张波, 徐涛, 徐浩, 等. 基于生物信息学胰腺腺癌关键基因的筛选及支持向量机诊断模型的构建[J]. 中国普通外科杂志, 2021, 30(3):276–285. doi:10.7659/j.issn.1005–6947.2021.03.005

Cite this article as: Zhang B, Xu T, Xu H, et al. Identification of hub genes in pancreatic adenocarcinoma and construction of a support vector machine diagnostic classifier based on bioinformatics approaches[J]. Chin J Gen Surg, 2021, 30(3):276–285. doi:10.7659/j.issn.1005–6947.2021.03.005

本刊 2021 年各期重点内容安排

本刊 2021 年各期重点内容安排如下，欢迎赐稿。

第 1 期　肝脏肿瘤外科治疗及相关实验研究

第 2 期　胆道肿瘤外科治疗及相关实验研究

第 3 期　胰腺肿瘤外科治疗及相关实验研究

第 4 期　胃肠肿瘤外科治疗及相关实验研究

第 5 期　甲状腺肿瘤外科治疗及相关实验研究

第 6 期　主动脉疾病外科治疗及相关实验研究

第 7 期　肝脏外科临床实践与基础研究

第 8 期　胆道外科临床实践与基础研究

第 9 期　胰腺外科临床实践与基础研究

第 10 期　胃肠外科临床实践与基础研究

第 11 期　乳腺、甲状腺外科临床实践与基础研究

第 12 期　血管外科临床实践与基础研究

中国普通外科杂志编辑部