



doi:10.7659/j.issn.1005-6947.2022.09.009
http://dx.doi.org/10.7659/j.issn.1005-6947.2022.09.009
Chinese Journal of General Surgery, 2022, 31(9):1203-1209.

· 基础研究 ·

基于机器学习的胰腺癌特征基因筛选初步研究

魏伟¹, 欧政林¹, 窦晓淋¹, 张帅¹, 唐翎^{2,3}

(中南大学湘雅医院 1. 普通外科 2. 药学部 3. 国家老年疾病临床医学研究中心, 湖南 长沙 410008)

摘要

背景与目的: 胰腺癌是一种难治的癌症, 90% 以上的患者在诊断后 1 年内死亡。胰腺癌病变组织和正常组织之间存在差异表达基因 (DEGs) 可能与胰腺癌的发生和发展密切相关。本研究运用机器学习方法对胰腺癌 DEGs 进行筛选, 以期研究该病的发生机制提供依据。

方法: 从公共基因 GEO 数据库中筛选胰腺癌基因表达谱, 使用线性回归模型软件包 Limma 对不同组的芯片进行差异性计算, 归一化; 使用 R 语言获得 DEGs, 对筛选出来的 DEGs 特征选择方法进一步进行筛选; 基于获得的核心 DEGs, 采用 AdaBoost 和 Bagging 算法分别构建胰腺癌预测模型。用 DAVID 网站对核心 DEGs 进行 GO 功能分析和 KEGG 通路富集分析, 再用 STRING 网站及 Cytoscape 软件对核心 DEGs 进行蛋白-蛋白相互作用 (PPI) 网络分析, 最后用 GEPIA 网站对预后相关的核心 DEGs 行生存分析。

结果: 通过特征筛选, 得到了 18 个关键的 DEGs; 以该 18 个 DEGs 建立特征子集, 结合 AdaBoost 算法建立了预测模型, 预报准确率可以达到 92.3%。通过对 DEGs 的 GO 和 KEGG 分析, 发现 CDK1、CCNA2 和 CCNB1 有间接作用, 对胰腺癌的形成和发展有一定的作用。生存分析显示, CDK1 ($P=0.0008$)、CCNB1 ($P=0.012$)、CSK2 ($P=0.023$)、CKS1B ($P=0.0013$) 的表达量与患者总生存期 (OS) 有相关性, 它们的表达量越高, 患者 OS 越短。

结论: 机器学习方法可较好地胰腺癌特征基因进行筛选, 对胰腺癌的诊治及相关的药物开发具有一定意义。

关键词

胰腺肿瘤; 基因表达谱; 机器学习; 计算生物学
中图分类号: R736.7

Machine learning-based feature gene screening of pancreatic cancer

WEI Wei¹, OU Zhenglin¹, DOU Xiaolin¹, ZHANG Shuai¹, TANG Ling^{2,3}

(1. Department of General Surgery 2. Department of Pharmacy 3. National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha 410008, China)

Abstract

Background and Aims: Pancreatic cancer is a difficult-to-treat disease and over 90% of the patients will die within one year of diagnosis. The presence of differentially expressed genes (DEGs) between diseased and normal pancreatic cancer tissues may closely associated with the development and progression of pancreatic cancer. This study was conducted to screen the DEGs in pancreatic cancer using a machine learning approach, so as to provide a basis for studying the pathogenetic mechanism of

基金项目: 湖南省自然科学基金资助项目 (2019JJ40489)。

收稿日期: 2022-06-26; **修订日期:** 2022-08-25。

作者简介: 魏伟, 中南大学湘雅医院主治医师, 主要从事胰腺疾病基础与临床方面的研究。

通信作者: 唐翎, Email: tangling82@126.com

this disease.

Methods: Pancreatic cancer gene expression profiles were screened from the public gene GEO database, differential calculations and normalizations were performed using the linear regression model package Limma for different groups of microarrays. The DEGs were obtained using the R language, and the selected DEGs were further screened by correlation-based feature selection method. Based on the hub DEGs obtained, AdaBoost and Bagging algorithms were used to construct pancreatic cancer prediction models respectively. The GO function analysis and KEGG enrichment analysis of the hub DEGs were performed through the DAVID website, and protein-protein interaction (PPI) network of the hub DEGs was analyzed using STRING database and Cytoscape software. Finally, survival analysis was performed on the relevant hub DEGs through the GEPIA website.

Results: Through feature screening, 18 key DEGs were obtained. A prediction model was built by using AdaBoost algorithm based on the feature subset containing the 18 DEGs, and the prediction accuracy reached 92.3%. The GO and KEGG analysis of the DEGs revealed an indirect role for CDK1, CCNA2 and CCNB1 in the formation and development of pancreatic cancer. Survival analysis showed that the expressions of CDK1 ($P=0.0008$), CCNB1 ($P=0.012$), CSK2 ($P=0.023$) and CKS1B ($P=0.0013$) were correlated with the overall survival (OS) of patients, and higher expressions of them were associated with shorter OS of patients.

Conclusion: Machine learning methods can be efficiently applied for hub genes screening in pancreatic cancer, and have certain significance for the diagnosis and treatment of pancreatic cancer and related drug development.

Key words

Pancreatic Neoplasms; Gene Expression Profiling; Machine Learning; Computational Biology

CLC number: R736.7

胰腺癌是一种恶性程度很高的肿瘤，在全世界的发病率和病死率都很高，全球范围内呈快速上升趋势，我国发病率逐年上升并呈年轻化的趋势，严重危害着人类的健康^[1-5]。80%的胰腺癌是胰腺导管腺癌，以高侵袭性和早期转移为特点，临床症状出现晚，发现时已为晚期，虽然放疗和化疗对延长患者生存期起到了一定的作用，但是患者中位生存期仍小于2年^[6-8]。

研究^[9-13]表明，胰腺癌病变组织和正常组织之间存在差异表达基因 (different expression genes, DEGs)，这些 DEGs 很可能导致胰腺癌的发生和发展。高通量基因芯片和测序技术作为基因表达分析的工具，已被广泛应用于识别肿瘤发生过程中遗传信息的改变^[14-15]。随着基因芯片技术的发展，已经产生了大量核酸数据，这些数据需要经过挖掘加工才能够被有效应用。芯片数据挖掘涉及很多方面，如图像处理、数据标准化、DEGs 筛选等。GEO (Gene Expression Omnibus) 数据库中有

丰富的肿瘤相关基因组及基因表达谱，为研究细胞癌基因表达情况和发现关键基因的变化规律提供基础^[16]。基于相关性的特征选择 (correlation-based feature selection, CFS) 变量筛选方法计算特征变量对于目标的整体贡献来判断特征变量集的分类能力大小，将与目标的相关性很低的变量，以及变量之间高度相关的冗余变量去掉，这种变量筛选方法在生命科学，药物设计等领域被广泛用于高维数据及特征筛选^[17]。

本研究基于 GEO 数据库中获得的基因芯片数据，通过特征筛选获得差异表达基因，建立相应的胰腺癌判别模型，结合 GO 功能富集分析和 KEGG 通路富集分析，蛋白质交互作用网络以及生存分析来研究部分导致胰腺癌的关键基因，初步探索胰腺癌潜在的治疗靶点，为研究胰腺癌分子机制提供基础，为胰腺癌的治疗诊断提供参考依据。

1 资料与方法

1.1 数据收集

数据库资料下载本研究中,从GEO数据库下载了GSE16515 mRNA表达谱,该表达谱采用GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array平台,包含胰腺组织样本52例,其中36例胰腺癌样本,16例正常组织样本。使用线性回归模型软件包Limma对不同组的芯片进行差异性计算,归一化。

1.2 筛选DEGs

利用R语言以 $|\log_2FC| > 2$, $P < 0.05$ 为标准,筛选出DEGs^[18]。

1.3 GO功能富集分析和KEGG通路富集分析

GO功能富集分析一般包括生物过程(biological process, BP)、分子功能(molecular function, MF)和细胞成分(cellular component, CC)^[19]。将这些DEGs导入到在线工具DAVID数据库中,分别进行GO和KEGG富集分析^[20-21],并利用得到的数据绘制气泡图。

1.4 蛋白质相互作用网络分析

利用在线数据库STRING构建DEGs之间的蛋白-蛋白相互作用(protein-protein interaction, PPI)网络图,分析蛋白质功能之间的相互作用。随后

应用Cytoscape软件(3.8.0)在线工具使PPI的网络模块可视化^[22]。

1.5 生存分析

利用PPI网络分析筛选出其中的靶基因之后,应用在线工具GEPIA数据库对TCGA数据库中有详细临床资料的胰腺癌患者,采用Kaplan-Meier生存分析研究靶基因表达水平与胰腺癌患者总生存期(overall survival, OS)之间的关系^[23],验证其在胰腺癌发生过程中的参与情况。

2 结果

2.1 DEGs初步筛选

“差异”在生物学数据分析时有两层含义,一是统计学上的差异,另一个则是生物学上的差异,所以DEGs的选取通常设置至少两个阈值:统计显著性量度 P 值和基因表达变化量(fold change, FC)。在本研究中,通过计算发现,胰腺癌与对照的基因表达差异比较,满足 $P < 0.05$ 及 $|\log_2FC| > 2$ 的有1121个基因(图1),其中上调基因834个,下调基因287个。可以看出,有较多基因的差异倍数较高,推测胰腺肿瘤亦可能是从良性肿瘤向恶性肿瘤转变的过程,所以才存在很大的差异性。

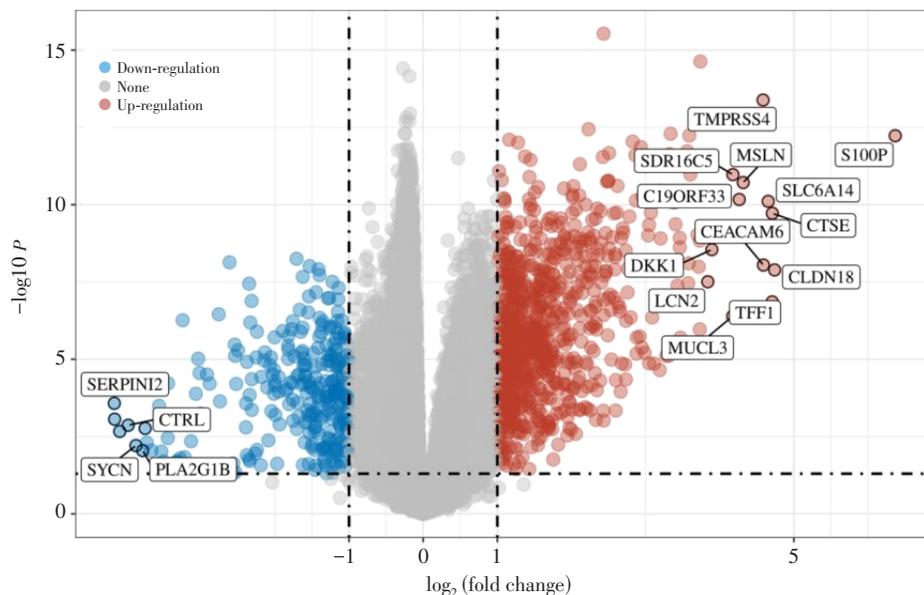


图1 DEGs火山图(蓝色点表示满足阈值的下调DEGs,红色点表示满足阈值的上调DEGs,灰色点表示不满足阈值的DEGs)

Figure 1 Volcano plots of DEGs (Blue dots indicating the down-regulated DEGs that meet the threshold, red dots indicating the up-regulated DEGs that meet the threshold, and gray dots indicating DEGs that do not meet the threshold)

2.2 基于CFS特征过滤方法的DEGs筛选

使用CFS算法对2.1部分中筛选出的DEGs进行进一步筛选,通过胰腺癌样本与正常样本比较,共筛选出18个DEGs,其中包括了16个上调基因(BUB1B、CCNA2、CCNB1、CDC20、CDC6、CDK1、CKS1B、CKS2、EPHA4、MAD1L1、MAD2L1、MCM2、NDC80、RACGAP1、TTK、ZWINT)和2个下调基因(ADHFE1、PSMD6)。以该18个DEGs为变量,使用Adaboost算法和Bagging算法,并用4种弱分类器作为基本分类器,分别构建胰腺癌判别预测模型(表1)。结果显示,两种算法,使用不同分类器时,预测准确率都高于80%,说明我们筛选得到的18个DEGs,能够很好地识别肿瘤患者。其中,以RandomForest为弱分类器,采用Adaboost方法所得到的判别模型的预报准确率最高,可以达到92.3%。从表中可以看出,8种算法建立的预测模型中,筛选的18个DEGs,无论选择何种算法,都能够较好地地区分胰腺癌与正常样本。

表1 不同弱分类器对胰腺癌的预测结果

Table 1 Prediction result of different weak classifiers for pancreatic cancer

集成算法	基本算法	预报准确率(%)		
		总预报准确率	正常	疾病
AdaBoost	KNN	90.4	81.3	94.4
	C4.5	90.4	87.5	91.7
	RF	92.3	87.5	94.4
	RT	88.5	87.5	88.9
Bagging	KNN	88.5	81.3	91.7
	C4.5	84.6	75.0	88.9
	RF	86.5	81.3	88.9
	RT	86.5	81.3	88.9

2.3 DEGs的生物学功能分析

GO功能富集分析结果显示,在BP方面中RNA聚合酶II启动子的正/负性转录调控(positive/negative regulation of transcription from RNA polymerase II promoter)和DNA模板的转录正调控(positive regulation of transcription, DNA-templated)富集的基因数量较多;在CC方面,DEGs主要与胞质(cytosol)、细胞核(nucleus)和核质体(nucleoplasm)富集的基因数量较多;而DEGs的MF主要集中在蛋白结合(protein binding)和同样蛋白结合(identical protein binding)(图2)。KEGG通路富集分析表明,它们主要参与癌症通路

(pathways in cancer), Wnt信号通路(Wnt signaling pathway), HIF-1信号通路(HIF-1 signaling pathway)和甲状腺激素信号通路(thyroid hormone signaling pathway)等途径(图3)。

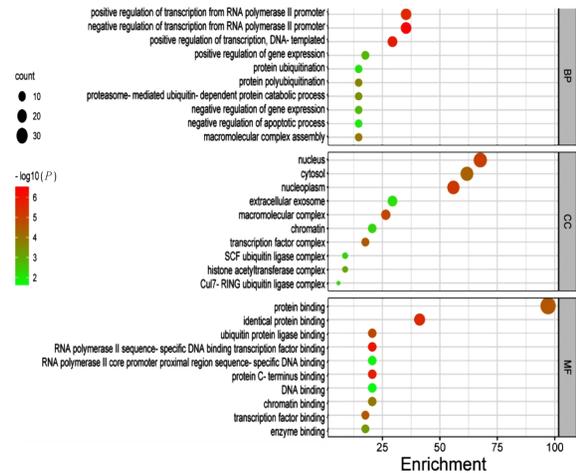


图2 差异表达基因的GO功能富集分析

Figure 2 Functional enrichment analysis of GO for differentially expressed genes

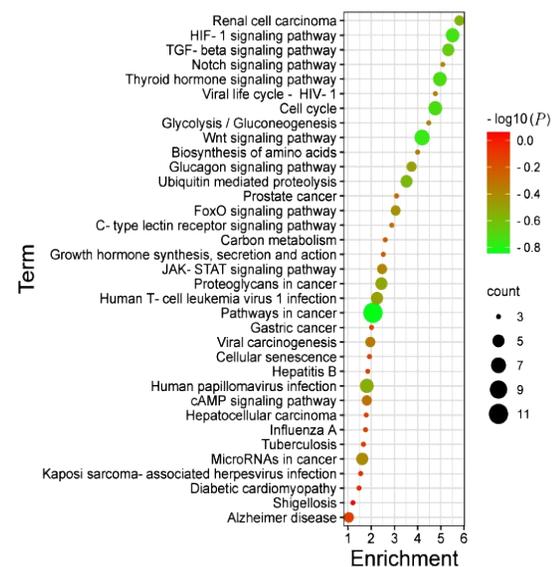


图3 DEGs的KEGG功能富集分析

Figure 3 KEGG functional enrichment analysis of DEGs

2.4 PPI网络分析

使用在线数据库STRING构建DEGs之间PPI网络。由图4可知,细胞分裂周期蛋白20(CDC20)结构度最高,和较多其他节点有互作。其次是细胞周期蛋白A2(CCNA2)、细胞周期蛋白B1(CCNB1)和周期蛋白依赖性激酶1(CDK1)。这些基因可能在胰腺癌的发生和发展中起到核心作用。

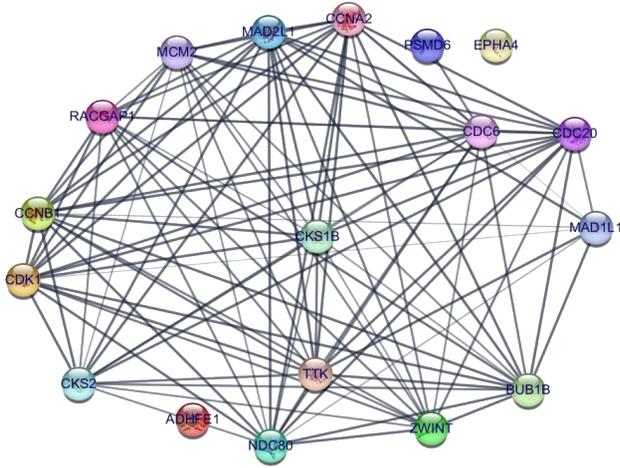


图4 DEGs蛋白交互作用网络图

Figure 4 Protein interaction network of DEGs

2.5 生存分析

为了进一步验证靶基因与OS的关系, 在利用PPI筛选出靶基因后, 使用在线分析工具GEPIA数据对TCGA数据库中有详细资料的胰腺癌患者进行Kaplan-Meier生存分析, 研究靶基因表达量与胰腺患者OS之间的关系。选取基因表达量的中位数为基线, 将其划分为高表达组和低表达组。其中CDK1 ($P=0.0008$)、CCNB1 ($P=0.012$)、CKS2 ($P=0.023$)、CKS1B ($P=0.0013$)的表达量与患者OS具有相关性, 这些基因表达量越高, 患者OS越短(图5)。

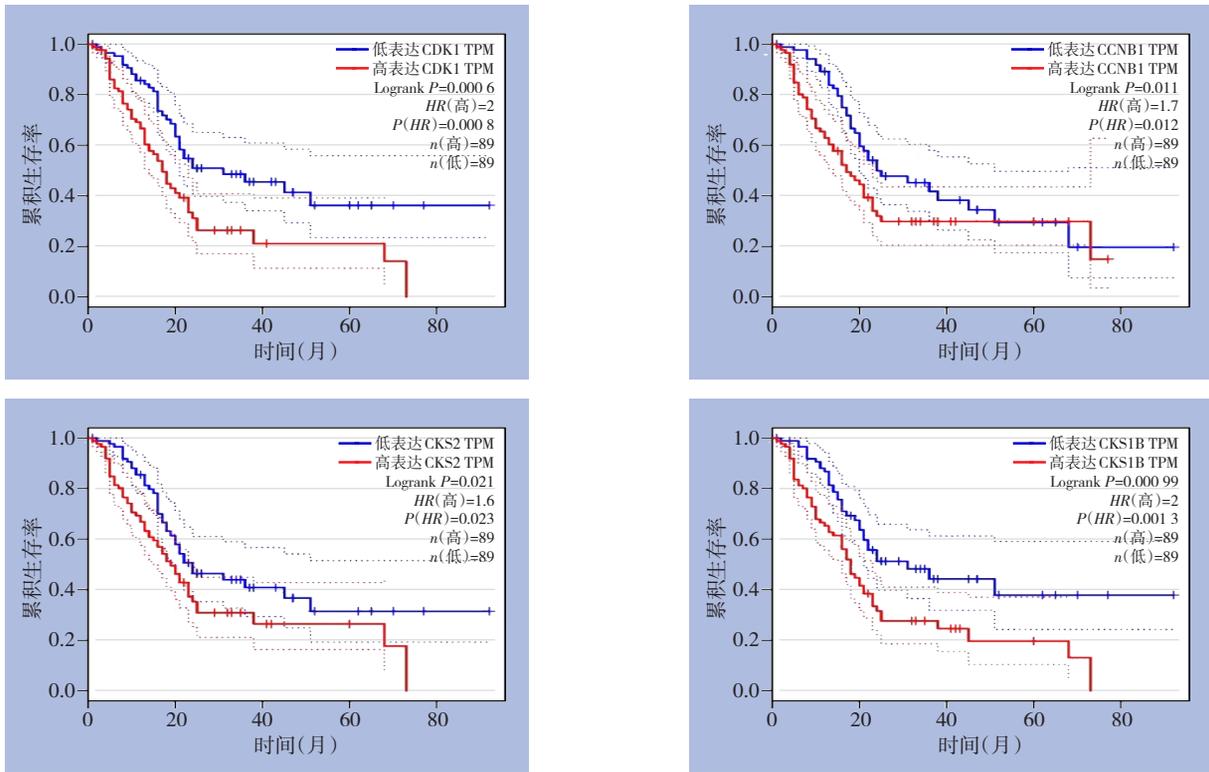


图5 关键基因表达与胰腺癌患者生存的关系

Figure 5 Relations of the expressions of the hub genes with the survival of pancreatic cancer patients

3 讨论

本研究一共筛选出18个DEGs, GO功能富集分析表明这些DEGs在RNA聚合酶II启动子的正/负性转录调控、DNA模板的转录正调控富集、CC方面, DEGs主要与胞质、细胞核, 蛋白结合和同样蛋白结合中起作用。KEGG通路富集分析表明它们主要参与癌症通路、Wnt信号通路、HIF-1信号

通路和甲状腺激素信号通路等途径中起作用。

CCNA2和CCNB1编码的蛋白都属于细胞周期蛋白家族, 其成员的特点是在细胞周期中蛋白丰度具有周期性^[24]。细胞周期蛋白作为CDK激酶的调节器发挥作用。不同的细胞周期蛋白表现出不同的表达和降解模式, 有助于每个有丝分裂事件的时间协调。目前已有有关CCNA2和CCNB1基因在癌组织中的表达、信号通路和预后关系相关的研究^[25-27]。

肿瘤细胞以持续分裂、增殖不受控制为特点,细胞周期蛋白依赖性激酶 CDK 对肿瘤细胞的生存具有重要意义。CDK1 能通过 BRCA1 的磷酸化促进 DNA 双链断裂的同源重组修复过程和细胞周期检查点激活。因此,CDK1 是许多生物学过程中调控的核心,包括细胞周期调控、DNA 复制、DNA 损伤修复等,并将这些生物学过程与细胞周期进程紧密联系起来^[28-29]。鉴于 CDK1 对细胞周期的调控、检查点的激活、DNA 损伤修复发挥着关键性的作用,CDK1 成为信号通路的重要位点,相关的抑制剂开发受到了广泛的关注,尤其是其临床研究更多指向胰腺癌^[30-33]。

在本研究中,采用 CFS 算法对胰腺癌的 DEGs 进行筛选,获得 18 个 DEGs,在此基础上使用机器学习方法构建了胰腺癌判别模型。通过 GO 和 KEGG 生物功能分析,PPI 网络分析和生存率分析,发现 CDK1、CCNA2 和 CCNB1 的可能与胰腺癌的发生和发展有关。

利益冲突:所有作者均声明不存在利益冲突。

参考文献

- [1] Berrondo C, Flax J, Kuchero V, et al. Expression of the long non-coding RNA HOTAIR correlates with disease progression in bladder cancer and is contained in bladder cancer patient urinary exosomes[J]. PLoS One, 2016, 11(1): e0147236. doi: 10.1371/journal.pone.0147236.
- [2] Kleeff J, Korc M, Apte M, et al. Pancreatic cancer[J]. Nat Rev Dis Primers, 2016, 2:16022. doi: 10.1038/nrdp.2016.22.
- [3] Rivera JA, Graeme-Cook F, Werner J, et al. A rat model of pancreatic ductal adenocarcinoma: targeting chemical carcinogens[J]. Surgery, 1997, 122(1):82-90. doi: 10.1016/s0039-6060(97)90268-3.
- [4] Song SY. Etiology and carcinogenesis of pancreatic ductal adenocarcinoma[J]. Korean J Gastroenterol, 2008, 51(2):71-83.
- [5] Vassaux G, Angelova A, Baril P, et al. The promise of gene therapy for pancreatic cancer[J]. Hum Gene Ther, 2016, 27(2):127-133. doi: 10.1089/hum.2015.141.
- [6] Luo ZF, Li YN, Wang HM, et al. Hepatocyte nuclear factor 1A (HNF1A) as a possible tumor suppressor in pancreatic cancer[J]. PLoS One, 2015, 10(3): e0121082. doi: 10.1371/journal.pone.0121082.
- [7] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022[J]. CA Cancer J Clin, 2022, 72(1):7-33. doi: 10.3322/caac.21708.
- [8] 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.
- [9] 任天宇,周新童,党胜春. miR-486-5p 的靶基因预测及其在胰腺腺癌中作用的生物信息学分析[J]. 中国普通外科杂志, 2020, 29(6):715-722. doi:10.7659/j.issn.1005-6947.2020.06.012.
Ren TY, Zhon XT, Dang SC. Prediction of target gene of miR-486-5p and bioinformatics analysis of their roles in pancreatic adenocarcinoma[J]. Chinese Journal of General Surgery, 2020, 29(6):715-722. doi:10.7659/j.issn.1005-6947.2020.06.012.
- [10] 王欢,金钢. 胰腺癌精准治疗的现状和展望[J]. 中国普通外科杂志, 2021, 30(9): 997-1005. doi: 10.7659/j.issn.1005-6947.2021.09.001.
Wang H, Jin G. Current status and future perspective of precision medicine in pancreatic cancer treatment[J]. Chinese Journal of General Surgery, 2021, 30(9):997-1005. doi: 10.7659/j.issn.1005-6947.2021.09.001.
- [11] 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.
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]. Chinese Journal of General Surgery, 2021, 30(3): 276-285. doi:10.7659/j.issn.1005-6947.2021.03.005.
- [12] 张帆,李泽东,彭禹,等. 基于血清 miRNA 表达数据的胰腺癌诊断决策树构建[J]. 中国普通外科杂志, 2021, 30(2):211-218. doi: 10.7659/j.issn.1005-6947.2021.02.010.
Zhang F, Li ZD, Peng Y, et al. Construction of decision tree for diagnosis of pancreatic cancer based on serum miRNA expression data[J]. Chinese Journal of General Surgery, 2021, 30(2):211-218. doi: 10.7659/j.issn.1005-6947.2021.02.010.
- [13] 韩伟光,莘玮,苏水霞,等. 基于生物信息学的胆囊癌差异表达谱中关键蛋白调控基因分析[J]. 中国普通外科杂志, 2021, 30(2): 165-172. doi: 10.7659/j.issn.1005-6947.2021.02.006.
Han WG, Shen W, Su SX, et al. Analysis of key protein regulatory genes in differential expression profile of gallbladder cancer based on bioinformatics approaches[J]. Chinese Journal of General Surgery, 2021, 30(2): 165-172. doi: 10.7659/j.issn.1005-6947.2021.02.006.
- [14] Lu YQ, Li Z, Lin C, et al. Translation role of circRNAs in cancers[J]. J Clin Lab Anal, 2021, 35(7): e23866. doi: 10.1002/jcla.23866.
- [15] 李文波,孙成杰,周国俊,等. 肝细胞癌发生发展关键基因及其功能的生物信息学分析[J]. 中国普通外科杂志, 2021, 30(1):32-43. doi:10.7659/j.issn.1005-6947.2021.01.005.

- Li WB, Sun CJ, Zhou GJ, et al. Bioinformatics analysis of key genes and their functions in occurrence and development hepatocellular carcinoma[J]. Chinese Journal of General Surgery, 2021, 30(1):32-43. doi:10.7659/j.issn.1005-6947.2021.01.005.
- [16] Wang ZC, Lachmann A, Ma'ayan A. Mining data and metadata from the gene expression omnibus[J]. Biophys Rev, 2019, 11(1): 103-110. doi: 10.1007/s12551-018-0490-8.
- [17] Jain I, Jain VK, Jain R. Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification[J]. Appl Soft Comput, 2018, 62:203-215. doi: 10.1016/j.asoc.2017.09.038.
- [18] Radoaca A. Simple Venn diagrams for multisets[C]//17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). Romania, Timisoara. 2015:181-184. doi: 10.1109/SYNASC.2015.36.
- [19] Consortium TGO. Gene ontology annotations and resources[J]. Nucleic Acids Res, 2012, 41(D1): D530-535. doi: 10.1093/nar/gks1050.
- [20] Huang DW, Sherman BT, Tan QN, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists[J]. Nucleic Acids Res, 2007, 35(Web Server issue):W169-175. doi: 10.1093/nar/gkm415.
- [21] Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment[J]. Nucleic Acids Res, 2008, 36(Database issue):D480-484. doi: 10.1093/nar/gkm882.
- [22] Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life[J]. Nucleic Acids Res, 2015, 43(Database issue): D447-452. doi: 10.1093/nar/gku1003.
- [23] 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.
- [24] Gan YQ, Li YM, Li T, et al. CCNA2 acts as a novel biomarker in regulating the growth and apoptosis of colorectal cancer[J]. Cancer Manag Res, 2018, 10:5113-5124. doi: 10.2147/CMAR.S176833.
- [25] Chen EB, Qin X, Peng K, et al. HnRNPR-CCNB1/CENPF axis contributes to gastric cancer proliferation and metastasis[J]. Aging, 2019, 11(18):7473-7491. doi: 10.18632/aging.102254.
- [26] Kim SK, Roh YG, Park K, et al. Expression signature defined by FOXM1-CCNB1 activation predicts disease recurrence in non-muscle-invasive bladder cancer[J]. Clin Cancer Res, 2014, 20(12): 3233-3243. doi: 10.1158/1078-0432.ccr-13-2761.
- [27] Li SF, Liu N, Piao JX, et al. CCNB1 expedites the progression of cervical squamous cell carcinoma via the regulation by FOXM1[J]. Onco Targets Ther, 2020, 13: 12383-12395. doi: 10.2147/OTT.S279951.
- [28] Izadi S, Nikkhoo A, Hojjat-Farsangi M, et al. CDK1 in breast cancer: implications for theranostic potential[J]. Anti Cancer Agents Med. Chem., 2020, 20(7): 758-767. doi: 10.2174/1871520620666200203125712.
- [29] Liao HW, Ji F, Ying SM. CDK1: beyond cell cycle regulation[J]. Aging, 2017, 9(12):2465-2466. doi: 10.18632/aging.101348.
- [30] Huang Z, Shen G, Gao J. CDK1 promotes the stemness of lung cancer cells through interacting with Sox2[J]. Clin Transl Oncol, 2021, 23(9):1743-1751. doi: 10.1007/s12094-021-02575-z.
- [31] Ødum N. CDK1 links to RAR γ in treatment response of cancer cells[J]. Cell Cycle, 2013, 12(11):1659. doi: 10.4161/cc.25069.
- [32] Yasukawa M, Ando Y, Yamashita T, et al. CDK1 dependent phosphorylation of hTERT contributes to cancer progression[J]. Nat. Commun., 2020, 11:1557. doi: 10.1038/s41467-020-15289-7.
- [33] Ying X, Che X, Wang JZ, et al. CDK1 serves as a novel therapeutic target for endometrioid endometrial cancer[J]. J Cancer, 2021, 12(8):2206-2215. doi: 10.7150/jca.51139.

(本文编辑 宋涛)

本文引用格式:魏伟, 欧政林, 窦晓淋, 等. 基于机器学习的胰腺癌特征基因筛选初步研究[J]. 中国普通外科杂志, 2022, 31(9):1203-1209. doi:10.7659/j.issn.1005-6947.2022.09.009

Cite this article as: Wei W, Ou ZL, Dou XL, et al. Machine learning-based feature gene screening of pancreatic cancer[J]. Chin J Gen Surg, 2022, 31(9):1203-1209. doi:10.7659/j.issn.1005-6947.2022.09.009