·研究论文·

DOI:10.16605/j.cnki.1007-7847.2020.02.006

基于生物信息学分析的非小细胞肺癌诊断预后相关基因的筛选

杨燕霞[1], 金 莲[2], 王 欣[1], 张 洁[1], 柳小平[1*]
(1. 甘肃省第二人民医院 呼吸科, 中国甘肃 兰州 730030; 2. 重庆市北碚区中医院, 中国重庆 400700)

摘 要: 为了从基因层面探讨非小细胞肺癌(non-small cell lung cancer, NSCLC)发生发展的内在机制, 筛选与 NSCLC 诊断、预后相关的基因, 为 NSCLC 分子机制的进一步研究提供生物信息学依据, 利用生物信息学方法对 GEO 数据库和 TCGA 数据库的数据集进行合并分析, 筛选 NSCLC 组织与正常肺组织之间的差异表达基因(differentially expressed genes, DEGs), 并对所取交集的 DEGs 进行基因集富集分析(gene set enrichment analysis, GSEA)、基因本体论(gene ontology, GO)分析、KEGG (kyoto encyclopedia of genes and genomes)通路富集分析、蛋白质相互作用(protein-protein interaction, PPI)分析、ROC 曲线诊断效能分析及 LASSO 生存分析。文中共筛选出 240 个 DEGs, 主要涉及核分裂、染色体分离等生物学过程。GSEA 分析结果显示, 富集的通路主要涉及 DNA 修复和细胞周期。从 PPI 网络中筛选出 20 个 hub 基因, ROC 结果显示, *UBE2C* (AUC=0.939)、*TOP2A* (AUC=0.927)、*RRM2* (AUC=0.927)、*CCNB1* (AUC=0.928)、*MKI67* (AUC=0.930)、*AURKA* (AUC=0.931)、*MELK* (AUC=0.950)相对具有较高的诊断价值, LASSO COX 回归结果则显示 *IL6*、*KIAA0101*、*MKI67*、*TPX2*、*AURKA*、*CDKN3* 及 *CDCA5* 与 NSCLC 患者的预后强相关。本研究结果表明, *ZWINT*、*KIF2C*、*MELK*、*CDCA5* 可能在 NSCLC 中发挥着重要的作用, 为阐明 NSCLC 的分子机制提供了新思路。

关键词: 非小细胞肺癌(NSCLC); 生物信息分析; 差异表达基因(DEGs); 预后相关基因
中图分类号: Q786, Q811.4, R734.2 **文献标识码**: A **文章编号**: 1007-7847(2020)02-0127-09

Screening of Genes Related to Diagnosis and Prognosis of Non-small Cell Lung Cancer Based on Bioinformatics Analysis

YANG Yan-xia[1], JIN Lian[2], WANG Xin[1], ZHANG Jie[1], LIU Xiao-ping[1*]
(1. *Department of Respiration, Second Provincial People's Hospital of Gansu, Lanzhou* 730030, *Gansu, China*; 2. *Beibei Traditional Chinese Medical Hospital, Chongqing* 400700, *China*)

Abstract: In order to explore the internal mechanism of non-small cell lung cancer (NSCLC) from the genetic level, select genes related to the diagnosis and prognosis of NSCLC, and provide bioinformatics basis for further studing of the molecular mechanism of NSCLC, the data sets of GEO and TCGA databases were combined and analyzed by bioinformatics method to screen differentially expressed genes (DEGs) between NSCLC and normal lung tissues. Then, the Gene Set Enrichment Analysis (GSEA), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), protein-protein interaction (PPI) analysis, ROC curve diagnosis efficiency analysis and LASSO survival analysis were carried out in the intersection of DEGs. A total of 240 DEGs were screened out, which were mainly involved in the biological processes of nuclear division and chromosome separation. GSEA analysis showed that the enrichment pathways were mainly linked to DNA repair and cell cycle. Twenty hub genes were screened out from PPI network. ROC results showed that *UBE2C*

收稿日期: 2019-07-12; **修回日期**: 2019-10-31
作者简介: 杨燕霞(1986—), 女, 甘肃兰州人, 主治医师, 主要从事肺癌方面的研究, E-mail: 497534365@qq.com; * 通信作者: 柳小平(1982—), 男, 甘肃兰州人, 主治医师, 主要从事肺癌方面的研究, E-mail: 630274618@qq.com。

(AUC=0.939), *TOP2A* (AUC=0.927), *RRM2* (AUC=0.927), *CCNB1* (AUC=0.928), *MKI67* (AUC=0.930), *AURKA* (AUC=0.931) and *MELK* (AUC=0.950) had relatively high diagnostic values. LASSO Cox regression showed that *IL6*, *KIAA0101*, *MKI67*, *TPX2*, *AURKA*, *CDKN3* and *CDCA5* were related to the prognosis of NSCLC patients. The results showed that *ZWINT*, *KIF2C*, *MELK* and *CDCA5* may play an important role in NSCLC. This provides a new way to elucidate the molecular mechanism of NSCLC.

Key words: non-small cell lung cancer (NSCLC); bioinformatics analysis; differentially expressed genes (DEGs); prognostic related genes

(*Life Science Research*, 2020, 24(2): 127~135)

作为肺癌的主要类型, 非小细胞肺癌(non-small cell lung cancer, NSCLC)约占肺癌的 85%[1]。尽管近年来在 NSCLC 的筛查、诊断及治疗等方面已经取得了一些进展, 但患者的 5 年生存率仍然很低, 只有 17%[2]。随着基因组时代的开启及分子生物学研究的发展, 在分子水平研究生命现象和疾病的发生机制引起了人们的极大关注, 近年来 NSCLC 靶向新药开发的研究也越来越受到重视[3]。然而, 多数患者会对靶向药物产生抗药性, 导致靶向药物治疗效果欠佳。因此, 发现新的 NSCLC 治疗靶点对提高 NSCLC 患者的生存率有着重要的意义。综合生物信息分析是将不同数据集整合到一起, 获得更多的临床样本, 以实现可信度更高的分析, 从而为实验研究提供可行的思路[4-6]。本研究旨在利用生物信息学方法, 从 GEO 及 TCGA 数据库中分别筛选与 NSCLC 相关的差异表达基因(differentially expressed genes, DEGs), 然后取交集, 对所获得的 DEGs 展开功能相关性分析, 紧接着通过蛋白质网络互作选取 hub 基因, 并对 hub 基因同时进行 ROC 诊断分析及 LASSO 生存分析, 选取与预后密切相关的诊断标志物, 以期为 NSCLC 提供预后生物标志物及治疗靶点, 并为进一步研究 NSCLC 的分子机制提供新的思路。

1 材料与方法

1.1 数据下载与预处理

从 GEO 数据库(http://www.ncbi.nlm.nih.gov/geo)[7]中筛选样本来源~~可靠的 NSCLC 表达谱数据~~集, 使用 R 语言(version 3.6.1; http://r-project.org/) GEOquery 包[8]下载并分析数据集~~GSE18842[9]、GSE101929~~[10], 两者均是基于 GPL570 平台的人肺组织表达谱数据集。其中, GSE18842 数据集包括 45 例正常肺组织和 46 例 NSCLC 组织, GSE101929 数据集包括 34 例正常肺组织和 32 例 NSCLC 组织。同样, 使用 R 语言 RTCGAToolbox 包[11]在TCGA 数据库(https://www.cancer.gov/tcga)[12]下载 NSCLC mRNA ~~基因表达数据以及临床数据~~, 使用 R 语言软件整理相关表达及表型数据。

1.2 差异表达基因的筛选

通过 affy 包[13]将原始的 CEL 文件进行背景校正及均一化处理, 校正效果使用密度图进行可视化, 并转化为探针表达矩阵; 根据 Bioconductor 平台对应的 GPL 平台注释文件, 并对探针进行基因注释。通过 limma 软件包[14]筛选出 DEGs, 以$|\log_2$ fold change ($\log_2$FC)$|>2$ 且 $P<0.05$ 为差异截取标准。为了展示 DGEs 的差异表达情况, 以 ggplot2 软件包[15]绘制火山图。对于 TCGA 数据库中获取的数据, 则使用 edgeR 包[16]进行 DEGs 的筛选, 条件同样满足$|\log_2$FC$|>2$ 且 $P<0.05$。最后, 使用 VennDiagram 包[17]对三者取交集并以韦恩图进行可视化。

1.3 GSEA 分析、差异表达基因的基因本体论和通路富集分析

基因集富集分析(gene set enrichment analysis, GSEA)[18]通过评估一个预先定义的基因集的基因在与表型相关度排序的基因表中的分布趋势, 从而判断其对表型的贡献。基因本体论(gene ontology, GO)[19]是用来注释基因及其产物的常用方法, 大规模基因的注释经常使用该分析方法。文中使用 R 语言 clusterProfiler 包[20]对 DEGs 进行 GO 和 KEGG (kyoto encyclopedia of genes and genomes) 通路富集分析, 同时以 c2.cp.kegg.v6.0.symbols.gmt 作为参考基因集进行 GSEA 分析, $P<0.05$ 认为具有统计学意义。

1.4 蛋白质互作网络分析

STRING (version 11.0; http://string-db.org)[21]是用于评估蛋白质-蛋白质~~相互作用~~(protein-protein interaction, PPI)信息的在线工具。Cytoscape 常用于复杂网络的可视化[22], 其插件 cytoHubba 可用于计算基因所得度值, 常用来筛选 hub 基因[23]。首先将 240 个 DEGs 导入 STRING 中, 得到它们的互

作关系, 再将所得互作关系导入 Cytoscape 软件, 并用 cytoHubba 以 Degree 算法为标准筛选 hub 基因, 定义得分排名前 20 的基因为所得 hub 基因。

1.5　Hub 基因 ROC 诊断分析及 LASSO 生存分析

受试者操作特征曲线(receiver operating characteristic, ROC)能够直观地鉴别各诊断指标的诊断效能, ROC 曲线越靠近左上角, 曲线下面积(area under curve, AUC)越大, 诊断价值越高[24]。文中使用 R 语言 pROC 包[25]对所获得的 hub 基因矩阵进行 ROC 诊断分析, 以筛选具有诊断价值的 hub 基因; 使用软件包 glmnet[26]对 hub 基因进行 LASSO COX 回归分析, 从 hub 基因中筛选出跟预后强相关的基因。

2　结果

2.1　差异表达基因筛选结果

GEO 数据库来源数据集 GSE18842、GSE101929 的标准化处理结果如图 1A 所示, 两组样本密度图曲线基本重合, 可见两组样本来源可靠。GSE18842、GSE101929 数据集 DEGs 火山图展示结果如图 1B 所示, GSE18842 数据集中共筛选出 735 个 DEGs, GSE101929 数据集中共筛选出 858 个 DEGs。此外, TCGA 数据库中共筛选出 951 个 DEGs。对 GSE18842 数据集、GSE101929 数据集和 TCGA 数据库所得 DEGs 取交集, 共筛选出 240 个 DEGs (图 2)。

2.2　GSEA 分析及 GO 和 KEGG 通路富集分析

GSEA 分析不需要对基因进行表达差异的筛选, 能保留表达变化不大但功能重要的基因, 因此相比于 GO 和 KEGG 富集分析, 该方法保留了更多的信息。本文的 GSEA 分析结果显示, NSCLC 组富集的通路主要涉及 DNA 修复和细胞周期, 其中 *MCM* 基因家族以及 *BUB* 基因家族在其中作用突出, 结果如图 3 所示。进一步采用 R 语言对 240 个 DEGs 进行 GO 和 KEGG 通路富集分析, 结果如图 4 所示。DEGs 主要涉及核分裂、染色体

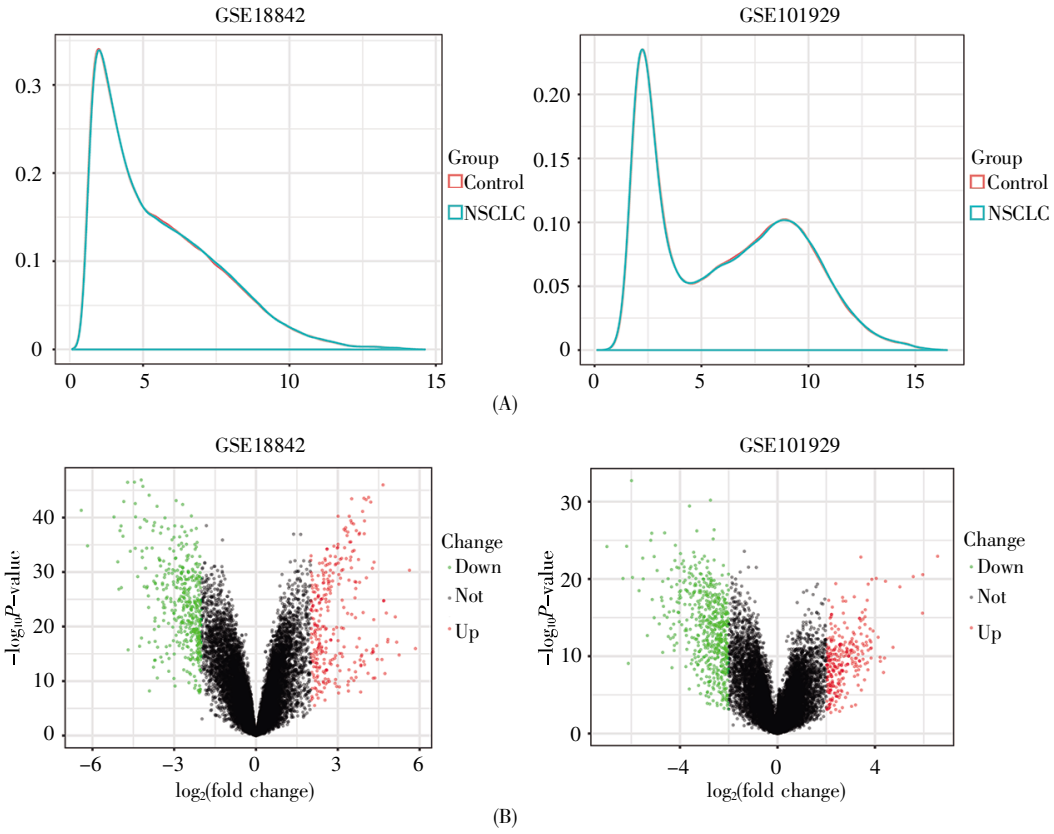


图 1　数据集预处理及 DEGs 火山图

(A) GSE18842 数据集和 GSE101929 数据集标准化处理之后的密度图; (B) GSE18842 数据集和 GSE101929 数据集中所获 DEGs 的火山图。

Fig.1　Dataset preprocessing and volcano map of DEGs

(A) The density map after standardization of GSE18842 and GSE101929 datasets; (B) The DEGs volcano map in GSE18842 and GSE101929 datasets.

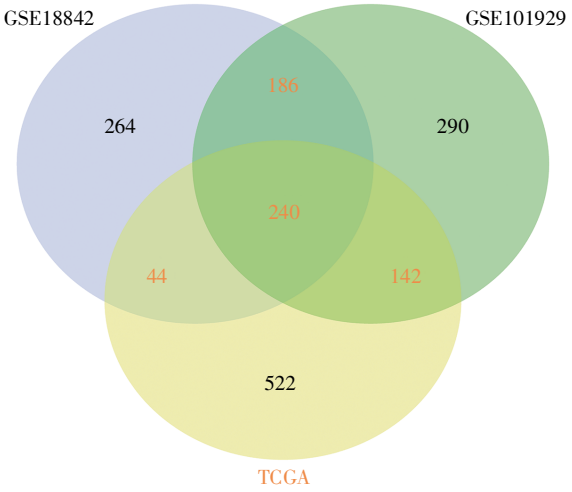


图 2 数据集中 DEGs 交集的 Venn 图
Fig.2 Venn map of differential gene intersection of datasets

分离的调控、核分裂调控等生物过程, 主要与细胞外基质(extracellular matrix, ECM)受体相互作用信号通路以及细胞黏附分子、细胞周期等通路相关。

2.3 DEGs 所编码蛋白质之间的相互作用分析

将 240 个 DEGs 输入 STRING 工具, 然后将所得互作数据(图 5A)导入 Cytoscape 中, 使用插件 cytoHubba 找出 hub 基因, 结果如图 5B 所示。*IL6*、*UBE2C*、*KIAA0101*、*TOP2A*、*MAD2L1*、*CDC20*、*CEP55*、*RRM2*、*MKI67*、*CDC6*、*CCNB1*、*KIF2C*、*TPX2*、*AURKA*、*CDKN3*、*MELK*、*CDCA5*、*CENPF*、*NUF2*、*ZWINT* 为所得 hub 基因。

2.4 ROC 诊断分析及 LASSO 生存分析

AUC>0.5 的情况下, AUC 值越接近 1, 表明诊断标志物的诊断效果越好。基于 GEO 数据集, 我们利用 R 语言绘制了 20 个 hub 基因的 ROC 曲线。结果如图 6 所示, hub 基因的 AUC 基本位于 0.7~0.9, 其中 *UBE2C* (AUC=0.939)、*TOP2A* (AUC=0.927)、*RRM2* (AUC=0.927)、*CCNB1* (AUC=0.928)、*MKI67* (AUC=0.930)、*AURKA* (AUC=0.931)和 *MELK* (AUC=0.950)相对具有较高的诊断价值。利用 LASSO 回归分析 hub 基因对 NSCLC 预后的影响, 结果如图 7 所示, 基因 *IL6*、*KIAA0101*、*MKI67*、*TPX2*、*AURKA*、*CDKN3*、*CDCA5* 均与 NSCLC 患者的生存预后显著相关。

3 讨论

NSCLC 是导致全球癌症相关死亡的主要原因之一, 其死亡率目前呈上升趋势[27]。虽然得益于免疫靶向治疗, 不少患者的生活质量得到改善, 但是晚期 NSCLC 患者的预后依旧很差。而且, 尽

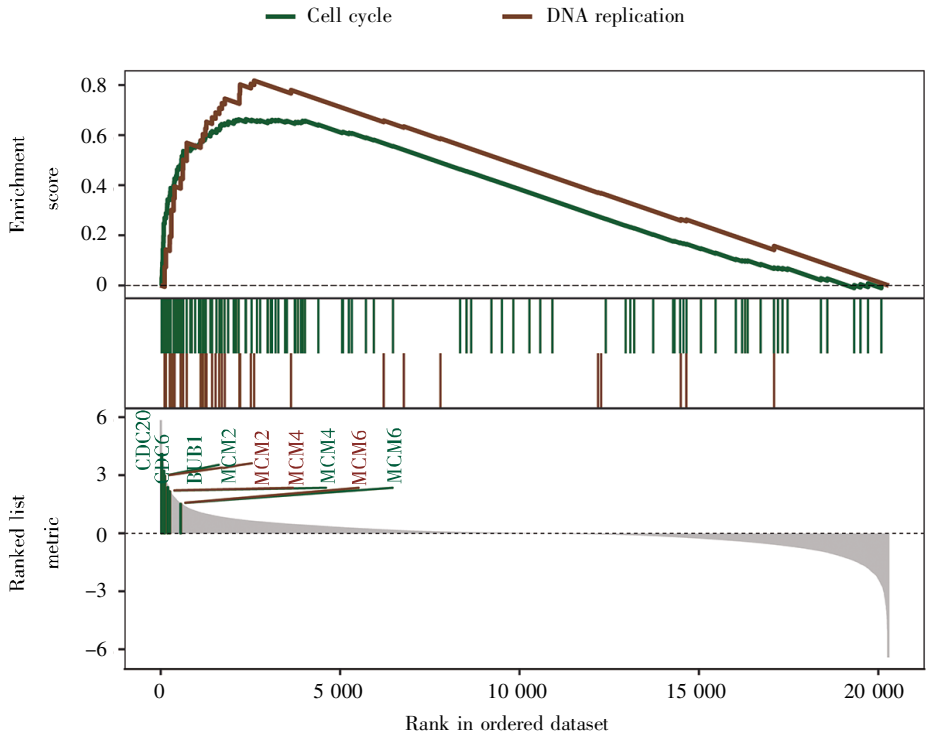


图 3 GSEA 通路富集分析
Fig.3 Enrichment analysis of GSEA pathway

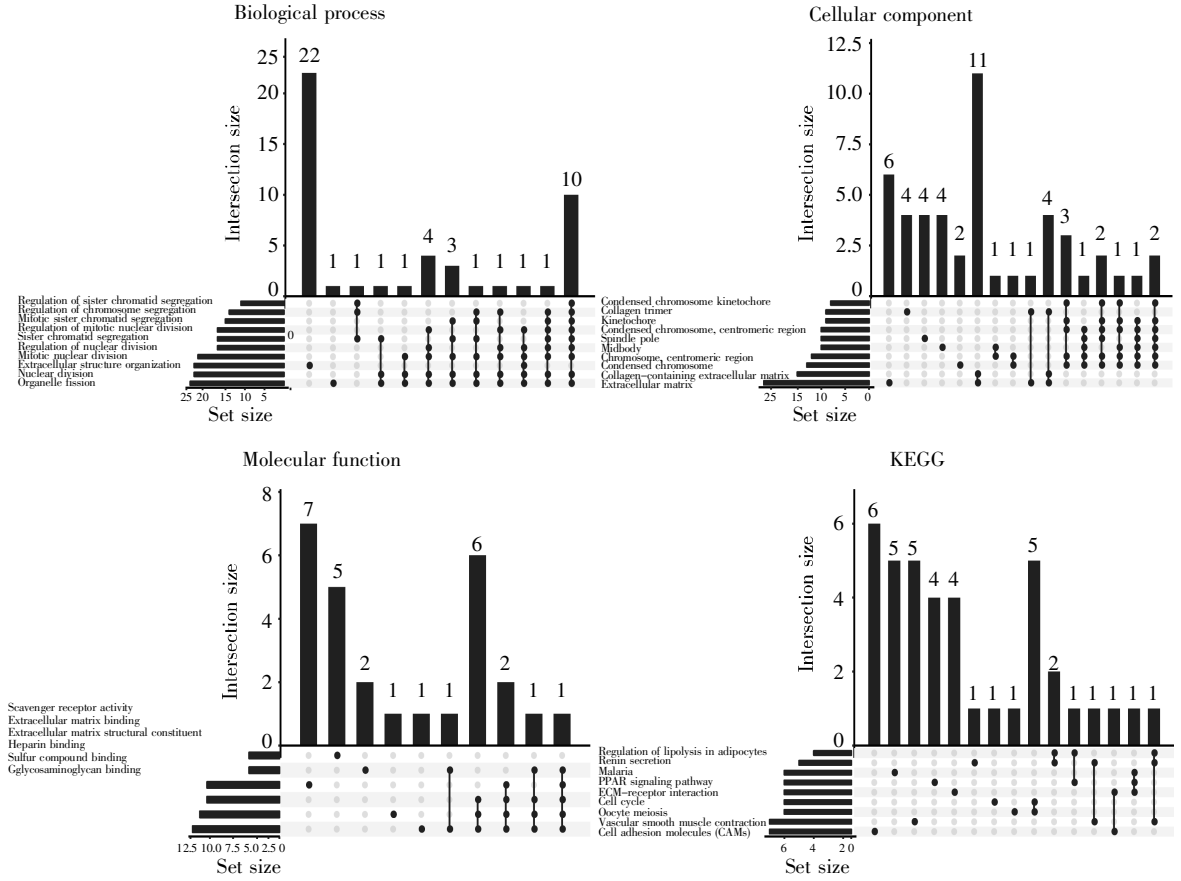


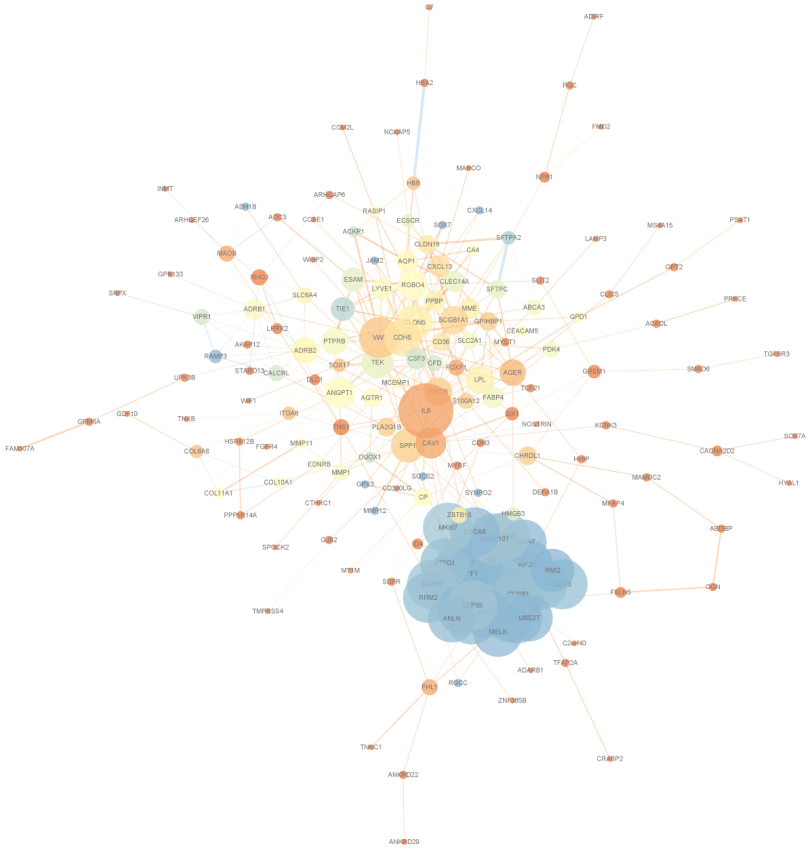
图 4　差异表达基因的 GO 分析及 KEGG 通路富集分析

Fig.4　GO analysis and KEGG pathway enrichment analysis of DEGs

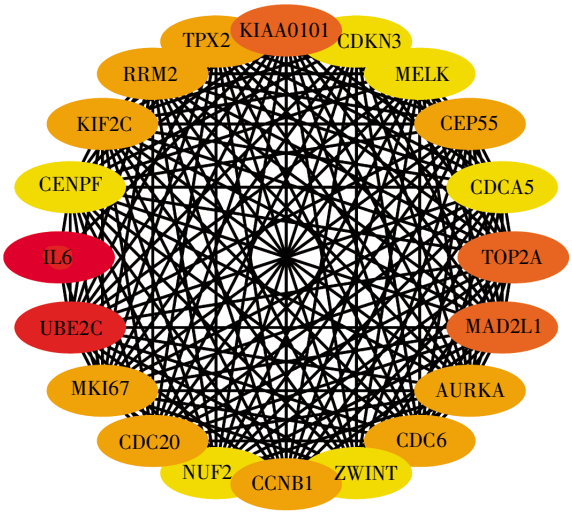
管人们在 NSCLC 样本的基因组学、转录组学、蛋白质组学以及代谢组学等领域的研究中发现了不少有望成为诊断和预后的候选生物标志物[28-29], 但是仍然没有改变当前 NSCLC 患者预后较差的事实。为提高肺癌患者的存活率, 给更多的 NSCLC 患者带来希望, 现阶段仍迫切需要发现新的有效的诊断和预后标记物。因此, 进一步阐明 NSCLC 的发生发展机制, 寻找有效的预后生物标志物, 对 NSCLC 患者的预后意义重大。

本研究采用生物信息学的方法分析 NSCLC 和正常对照组之间的差异表达基因。分别从GEO、TCGA 数据库下载 NSCLC 数据集, 经预处理后, 将数据集合并取交集, 共获取 240 个 DEGs。为了解 NSCLC 所涉及的通路, 对 DEGs 进行了 GSEA 分析、GO 功能注释和 KEGG 通路富集分析。GSEA 结果显示 DEGs 主要富集到细胞周期及 DNA 修复通路, 之前已有大量研究表明两者在 NSCLC 的发生发展及预后中起着重要的作用[30-34], 这也进一步证实我们的数据整合分析结果是可靠的。GO 分析结果显示, DEGs 主要参与的生物过程为核分裂、染色体分离的调控、核分裂调控等; 主要富集到细胞外基质; 主要富集于细胞外基质结构成分以及胞外基质结合等。相关研究表明, 核分裂、染色体分离的调控、核分裂调控与肿瘤的发生发展及转移有着密切关系[35-36]; 细胞外基质结构成分以及胞外基质与肿瘤转移和侵袭有关[37]。此外, KEGG 通路富集分析显示, ECM 受体相互作用信号通路在 NSCLC 中起着一定作用, 细胞与 ECM 之间的特异性相互作用由主要成分为整合素的跨膜分子介导, 这些相互作用可以控制细胞黏附、迁移及周期[38-39]; 而大量研究表明, 整合素在 NSCLC 发生发展中起着重要的桥梁作用[40-42]。因此, ECM-受体相互作用信号通路有望成为 NSCLC 潜在的药物治疗靶点。由此可见, 我们的研究结果与既往研究发现相一致。

为进一步筛选与 NSCLC 诊断及预后密切相关的基因, 我们利用 STRING 得到 DEGs 的互作网络关系, 再利用 Cytoscape 插件 cytoHubba 分析得到了 20 个 hub 基因 *IL6*、*UBE2C*、*KIAA0101*、*TOP2A*、*MAD2L1*、*CDC20*、*CEP55*、*RRM2*、*MKI67*、



(A)



(B)

图 5 差异表达基因所编码蛋白质的 PPI 分析图

(A) PPI 网络分析图。节点的大小表示聚类系数，节点越大，聚类系数越大，说明该基因在网络中占据比重就越大。节点颜色表示度，度越大，说明该节点连线就越多，度从大到小分别为橙、黄、蓝。线的粗细代表综合得分，得分越高线越粗。线的颜色代表共表达，同一种颜色说明两蛋白间存在互作关系；(B) Hub 基因所编码蛋白质的互作示意图。颜色越红越深，富集分数越高。

Fig.5 Protein-protein interaction analysis of differential gene-encoded proteins

(A) The PPI network analysis diagram. The size of a node represents the clustering coefficient, the larger the node, the larger the clustering coefficient, indicating that the gene occupies a larger proportion in the network. The node color indicates the degree, the greater the degree, the more connected the node. Degrees from big to small are orange, yellow, and blue. The thickness of the line represents the comprehensive score, the higher the score, the thicker the line. The color of the line represents co-expression, the same color indicates that there is an interaction between the two proteins; (B) The interaction diagram of hub gene-encoded proteins. The darker the color is, the higher the concentration is.

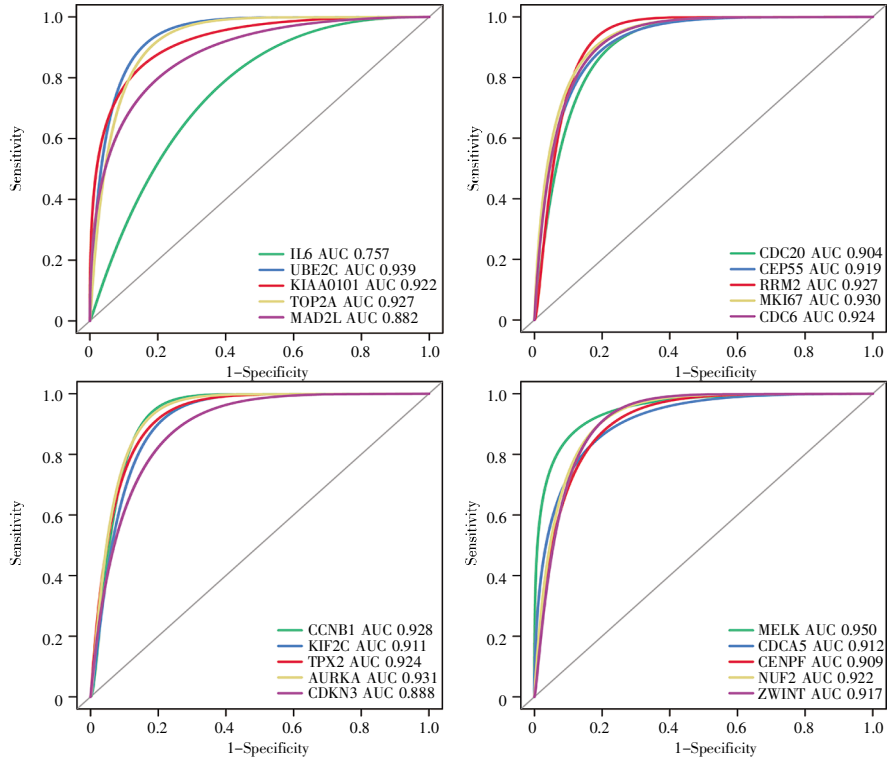


图 6　Hub 基因 ROC 诊断曲线图

横坐标为假阳性率, 越接近零准确率越高; 纵坐标为敏感度, 也称为真阳性率, 越大代表准确率越好。

Fig.6　ROC diagnostic curves of hub genes

The abscissa shows the false positive rate; the closer it to zero, the higher the accuracy. The ordinate represents sensitivity, also known as the true positive rate. The higher the rate, the greater the accuracy.

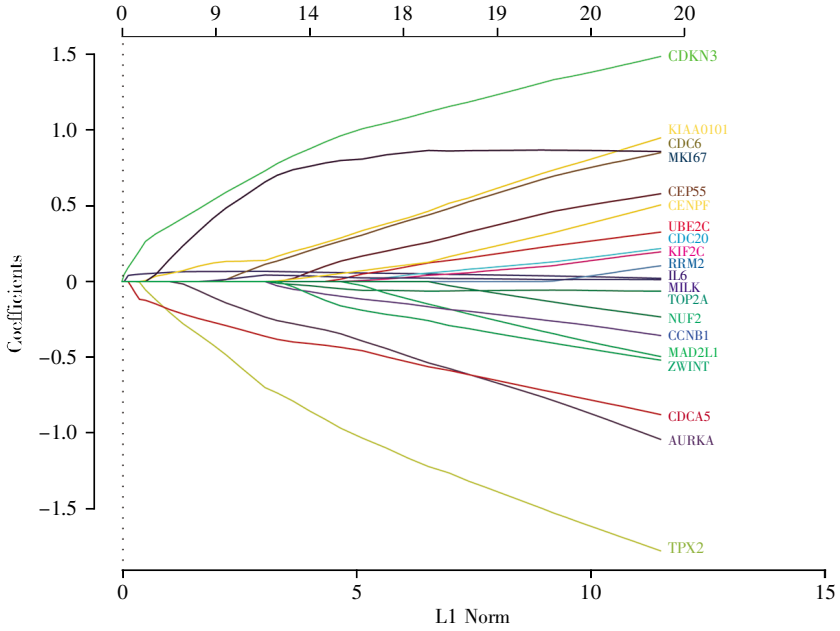


图 7　Hub 基因 LASSO 回归图

每一条曲线代表了一个自变量系数的变化轨迹, 纵坐标是系数的值, 下横坐标是 log(lambda), 上横坐标是此时模型中非零系数的个数。

Fig.7　Hub gene LASSO regression map

Each curve represents the change track of the coefficient of each independent variable. The ordinate is the value of the coefficient. The lower abscissa is log(lambda), and the upper abscissa is the number of non−zero coefficients in the model at the time.

*CDC*6、*CCNB*1、*KIF*2*C*、*TPX*2、*AURKA*、*CDKN*3、*MELK*、*CDCA*5、*CENPF*、*NUF*2、*ZWINT*。其中，*IL*6、*UBE*2*C*、*KIAA*0101、*TOP*2*A*、*MAD*2*L*1、*RRM*2、*MKI*67、*CCNB*1、*TPX*2、*AURKA*、*CDKN*3 在 NSCLC 发生发展中的作用研究较为深入，机制相对较为明确[43-53]。同时，*CDC*6、*CEP*55、*MELK*、*CDCA*5、*CENPF*、*NUF*2、*CDC*20 在 NSCLC 的研究中也有见报道[54-57]，但是，对于其具体导致 NSCLC 发生的分子机制仍有待进一步研究。*ZWINT* 所编码蛋白质在间期细胞的细胞质中分布均匀，是动粒形成和纺锤体检查点活动所必需的[58]。2019 年 Peng 等[59]的研究表明其可能是肺癌治疗的下一个重要靶点，因此，关于其参与 NSCLC 发生发展的机制研究值得进一步深入。*KIF*2*C* 编码一种类似运动蛋白的蛋白质，该蛋白质作为一种依赖于微管的分子马达，能使正端微管解聚，从而促进有丝分裂染色体分离[60]，而染色体分离与肿瘤的发生发展存在密切的关系，虽然目前尚无研究表明 *KIF*2*C* 与 NSCLC 有关系，但以上结果均提示 *KIF*2*C* 可能参与了 NSCLC 的发生发展。

此外，我们对 20 个 hub 基因进行了 ROC 诊断分析及 LASSO 生存分析，ROC 诊断分析结果表明 *UBE*2*C*、*TOP*2*A*、*RRM*2、*CCNB*1、*MKI*67、*AURKA*、*MELK* 相对具有较高的诊断价值，LASSO 生存分析结果则显示 *IL*6、*KIAA*0101、*MKI*67、*TPX*2、*AURKA*、*CDKN*3、*CDCA*5 均与 NSCLC 患者生存预后显著相关，它们可能是 NSCLC 潜在的预后生物标志物。

综上所述，细胞周期和 DNA 修复对 NSCLC 发生发展起着关键作用；基于诊断分析、预后分析以及文献复习的结果，我们预测 *ZWINT*、*KIF*2*C*、*MELK*、*CDCA*5 可能在 NSCLC 中发挥着重要的作用；同时，ECM-受体相互作用信号通路与 NSCLC 密切相关，相关机制值得进一步深入研究。总之，这些结果为阐明 NSCLC 发生发展的分子机制提供了理论依据，并确定了 *ZWINT*、*KIF*2*C*、*MELK*、*CDCA*5 可能成为诊断生物标志物、潜在治疗靶点和预后指标的新关键基因，有助于开发诊断和治疗 NSCLC 的新策略。

参考文献(References):

[1] SIEGEL R, NAISHADHAM D, JEMAL A. Cancer statistics, 2012[J]. CA: A Cancer Journal for Clinicians, 2012, 62(1): 10-29.

[2] WEI S, TIAN J, SONG X, *et al*. Causes of death and competing risk analysis of the associated factors for non-small cell lung cancer using the Surveillance, Epidemiology, and End Results database[J]. Journal of Cancer Research and Clinical Oncology, 2018, 144(1): 145-155.

[3] ZHOU G, ZHANG F, GUO Y, *et al*. miR-200c enhances sensitivity of drug-resistant non-small cell lung cancer to gefitinib by suppression of PI3K/Akt signaling pathway and inhibites cell migration via targeting ZEB1[J]. Biomedicine & Pharmacotherapy, 2017, 85: 113-119.

[4] HUANG D W, SHERMAN B T, LEMPICKI R A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists[J]. Nucleic Acids Research, 2009, 37(1): 1-13.

[5] SAEYS Y, INZA I, LARRANAGA P. A review of feature selection techniques in bioinformatics[J]. Bioinformatics, 2007, 23(19): 2507-2517.

[6] 解琪琪, 李文洲, 史卫东, 等. SNI 大鼠模型诱发的早期病理性疼痛相关基因芯片数据的生物信息学分析[J]. 生命科学研究(XIE Qi-qi, LI Wen-zhou, SHI Wei-dong, *et al*. Bioinformatics analysis of gene chip data related to early pathological pain induced by SNI rat model[J]. Life Science Research), 2018, 22(5): 384-389.

[7] BARRETT T, WILHITE S E, LEDOUX P, *et al*. NCBI GEO: archive for functional genomics data sets-update[J]. Nucleic Acids Research, 2013, 41(D1): D991-D995.

[8] SEAN D, MELTZER P S. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor[J]. Bioinformatics, 2007, 23(14): 1846-1847.

[9] SANCHEZ-PALENCIA A, GOMEZ-MORALES M, GOMEZ-CAPILLA J A, *et al*. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer[J]. International Journal of Cancer, 2011, 129(2): 355-364.

[10] MITCHELL K A, ZINGONE A, TOULABI L, *et al*. Comparative transcriptome profiling reveals coding and noncoding RNA differences in NSCLC from African Americans and European Americans[J]. Clinical Cancer Research, 2017, 23(23): 7412-7425.

[11] SAMUR M K. RTCGAToolbox: a new tool for exporting TCGA Firehose data[J]. PLoS One, 2014, 9(9): e106397.

[12] WEINSTEIN J N, COLLISSON E A, MILLS G B, *et al*. The Cancer Genome Atlas Pan-Cancer analysis project[J]. Nature Genetics, 2013, 45(10): 1113-1120.

[13] GAUTIER L, COPE L, BOLSTAD B M, *et al*. affy-analysis of Affymetrix GeneChip data at the probe level[J]. Bioinformatics, 2004, 20(3): 307-315.

[14] RITCHIE M E, PHIPSON B, WU D, *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. Nucleic Acids Research, 2015, 43(7): e47.

[15] WICKHAM H. ggplot2: Elegant Graphics for Data Analysis[M]. New York: Springer, 2009.

[16] ROBINSON M D, MCCARTHY D J, SMYTH G K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data[J]. Bioinformatics, 2010, 26(1): 139-140.

[17] CHEN H, BOUTROS P C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R[J]. BioMed Central Bioinformatics, 2011, 12: 35.

[18] SUBRAMANIAN A, KUEHN H, GOULD J, *et al*. GSEA-P: a desktop application for Gene Set Enrichment Analysis[J]. Bioinformatics, 2007, 23(23): 3251-3253.

[19] ASHBURNER M, BALL C A, BLAKE J A, *et al*. Gene Ontology: tool for the unification of biology[J]. Nature Genetics, 2000, 25(1): 25-29.

[20] YU G, WANG L, HAN Y, *et al*. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. Omics, 2012, 16(5): 284-287.

[21] SZKLARCZYK D, GABLE A L, LYON D, *et al*. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets[J]. Nucleic Acids Research, 2019, 47(D1): D607-D613.

[22] SHANNON P, MARKIEL A, OZIER O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. Genome Research, 2003, 13(11): 2498-2504.

[23] CHIN C, CHEN S, WU H, *et al*. cytoHubba: identifying hub objects and sub-networks from complex interactome[J]. BioMed Central Systems Biology, 2014, 8(Suppl. 4): S11.

[24] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. Radiology, 1982, 143(1): 29-36.

[25] ROBIN X, TURCK N, HAINARD A, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves[J]. BioMed Central Bioinformatics, 2011, 12: 77.

[26] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Regularization paths for generalized linear models via coordinate descent[J]. Journal of Statistical Software, 2010, 33(1): 1-22.

[27] SOCINSKI M A, OBASAJU C, GANDARA D, *et al*. Current and emergent therapy options for advanced squamous cell lung cancer[J]. Journal of Thoracic Oncology, 2018, 13(2): 165-183.

[28] GOCHEVA V, NABA A, BHUTKAR A, *et al*. Quantitative proteomics identify Tenascin-C as a promoter of lung cancer progression and contributor to a signature prognostic of patient survival[J]. Proceedings of the National Academy of Sciences USA, 2017, 114(28): E5625-E5634.

[29] MUSHARRAF S G, MAZHAR S, CHOUDHARY M I, *et al*. Plasma metabolite profiling and chemometric analyses of lung cancer along with three controls through gas chromatography-mass spectrometry[J]. Scientific Reports, 2015, 5: 8607.

[30] CAI J, WU J, ZHANG H, *et al*. MiR-186 downregulation correlates with poor survival in lung adenocarcinoma, where it interferes with cell-cycle regulation[J]. Cancer Research, 2013, 73 (2): 756-766.

[31] LOPEZ-GONZALEZ J S, PRADO-GARCIA H, AGUILAR-CAZARES D, *et al*. Apoptosis and cell cycle disturbances induced by coumarin and 7-hydroxycoumarin on human lung carcinoma cell lines[J]. Lung Cancer, 2004, 43(3): 275-283.

[32] WU Y F, OU C C, CHIEN P J, *et al*. Chidamide-induced ROS accumulation and miR-129-3p-dependent cell cycle arrest in non-small lung cancer cells[J]. Phytomedicine, 2018. DOI: https://doi.org/10.1016/j.phymed.2018.09.218.

[33] ROSELL R, PIFARRE A, MONZO M, *et al*. Reduced survival in patients with stage-I non-small-cell lung cancer associated with DNA-replication errors[J]. International Journal of Cancer, 1997, 74(3): 330-334.

[34] SAVIOZZI S, CEPPI P, NOVELLO S, *et al*. Non-small cell lung cancer exhibits transcript overexpression of genes associated with homologous recombination and DNA replication pathways[J]. Cancer Research, 2009, 69(8): 3390-3396.

[35] JALLEPALLI P V, LENGAUER C. Chromosome segregation and cancer: cutting through the mystery[J]. Nature Reviews Cancer, 2001, 1(2): 109-117.

[36] LOSADA A. Cohesin in cancer: chromosome segregation and beyond[J]. Nature Reviews Cancer, 2014, 14(6): 389-393.

[37] JONES J I, GOCKERMAN A, BUSBY W H, *et al*. Extracellular-matrix contains insulin-like growth-factor binding protein-5: potentiation of the effects of IGF-I[J]. Journal of Cell Biology, 1993, 121(3): 679-687.

[38] ZHANG H, TAO J, SHENG L, *et al*. Twist2 promotes kidney cancer cell proliferation and invasion by regulating ITGA6 and CD44 expression in the ECM-receptor interaction pathway[J]. OncoTargets and Therapy, 2016, 9: 1801-1812.

[39] FRULLANTI E, FALVELLA F S, NOCI S, *et al*. ECM-receptor interaction signature in normal tissue distinguishes lung adenocarcinoma patients from patients with lung metastasis[J]. European Journal of Cancer, 2012, 48(Suppl. 5): S159-S160.

[40] BARNEY L E, DANDLEY E C, JANSEN L E, *et al*. A cell-ECM screening method to predict breast cancer metastasis[J]. Integrative Biology, 2015, 7(2): 198-212.

[41] CHENG X, LU S, CUI Y. ECRG2 regulates ECM degradation and uPAR/FPRL1 pathway contributing cell invasion/migration[J]. Cancer Letters, 2010, 290(1): 87-95.

[42] KARMAKAR S, MUKHERJEE R. Integrin receptors and ECM proteins involved in preferential adhesion of colon carcinoma cells to lung cells[J]. Cancer Letters, 2003, 196(2): 217-227.

[43] CAETANO M S, ZHANG H, CUMPIAN A M, *et al*. IL6 blockade reprograms the lung tumor microenvironment to limit the development and progression of K-ras-mutant lung cancer[J]. Cancer Research, 2016, 76(11): 3189-3199.

[44] GUO J, WU Y, DU J, *et al*. Deregulation of UBE2C-mediated autophagy repression aggravates NSCLC progression[J]. Oncogenesis, 2018, 7: 49.

[45] KATO T, DAIGO Y, ARAGAKI M, *et al*. Overexpression of KIAA0101 predicts poor prognosis in primary lung cancer patients[J]. Lung Cancer, 2012, 75(1): 110-118.

[46] MIRSKI S E L, SPARKS K E, YU Q, *et al*. A truncated cytoplasmic topoisomerase II alpha in a drug-resistant lung cancer cell line is encoded by a *TOP2A* allele with a partial deletion of exon 34[J]. International Journal of Cancer, 2000, 85(4): 534-539.

[47] SHI Y X, ZHU T, ZOU T, *et al*. Prognostic and predictive values of CDK1 and MAD2L1 in lung adenocarcinoma[J]. Oncotarget, 2016, 7(51): 85235-85243.

[48] BOUKOVINAS I, PAPADAKI C, MENDEZ P, *et al*. Tumor BRCA1, RRM1 and RRM2 mRNA expression levels and clinical response to first-line gemcitabine plus docetaxel in non-small-cell lung cancer patients[J]. PLoS One, 2008, 3(11): e3695.

[49] LI Y, CHEN Y L, XIE Y T, *et al*. Association study of germline variants in CCNB1 and CDK1 with breast cancer susceptibility, progression, and survival among Chinese Han women[J]. PLoS One, 2013, 8(12): e84489.

[50] SCHNEIDER M A, CHRISTOPOULOS P, MULEY T, *et al*. AURKA, DLGAP5, TPX2, KIF11 and CKAP5: five specific mitosis-associated genes correlate with poor prognosis for non-small cell lung cancer patients[J]. International Journal of Oncology, 2017, 50(2): 365-372.

(下转第 159 页)

[29] WESLEY U V, HATCHER J F, AYVACI E R, *et al*. Regulation of dipeptidyl peptidase Ⅳ in the post-stroke rat brain and *in vitro* ischemia: implications for chemokine-mediated neural progenitor cell migration and angiogenesis[J]. Molecular Neurobiology, 2017, 54(7): 4973-4985.
[30] FOLESTAD E, KUNATH A, WAGSATER D. PDGF-C and PDGF-D signaling in vascular diseases and animal models[J]. Molecular Aspects of Medicine, 2018, 62: 1-11.
[31] CHEN J, ORTMEIER S B, SAVINOVA O V, *et al*. Thyroid hormone induces sprouting angiogenesis in adult heart of hypothyroid mice through the PDGF-Akt pathway[J]. Journal of Cellular and Molecular Medicine, 2012, 16(11): 2726-2735.
[32] MULLER H, HU J, POPP R, *et al*. Deleted in malignant brain tumors 1 is present in the vascular extracellular matrix and promotes angiogenesis[J]. Arteriosclerosis, Thrombosis, and Vascular Biology, 2012, 32(2): 442-448.
[33] HU Y, RAO S S, WANG Z X, *et al*. Exosomes from human umbilical cord blood accelerate cutaneous wound healing through miR-21-3p-mediated promotion of angiogenesis and fibroblast function[J]. Theranostics, 2018, 8(1): 169-184.
[34] NI J, LIU X, YIN Y, *et al*. Exosomes derived from *TIMP*2-modified human umbilical cord mesenchymal stem cells enhance the repair effect in rat model with myocardial infarction possibly by the Akt/Sfrp2 pathway[J]. Oxidative Medicine and Cellular Longevity, 2019, 2019: 1958941.
[35] WANG X, GU H, HUANG W, *et al*. Hsp20-mediated activation of exosome biogenesis in cardiomyocytes improves cardiac function and angiogenesis in diabetic mice[J]. Diabetes, 2016, 65(10): 3111-3128.
[36] HAN C, ZHOU J, LIANG C, *et al*. Human umbilical cord mesenchymal stem cell derived exosomes encapsulated in functional peptide hydrogels promote cardiac repair[J]. Biomaterials Science, 2019, 7(7): 2920-2933.
[37] VANDERGRIFF A, HUANG K, SHEN D, *et al*. Targeting regenerative exosomes to myocardial infarction using cardiac homing peptide[J]. Theranostics, 2018, 8(7): 1869-1878.
[38] WANG X, CHEN Y, ZHAO Z, *et al*. Engineered exosomes with ischemic myocardium-targeting peptide for targeted therapy in myocardial infarction[J]. Journal of the American Heart Association, 2018, 7(15): e008737.
[39] ZHU L P, TIAN T, WANG J Y, *et al*. Hypoxia-elicited mesenchymal stem cell-derived exosomes facilitates cardiac repair through miR-125b-mediated prevention of cell death in myocardial infarction[J]. Theranostics, 2018, 8(22): 6163-6177.
[40] WANG J, LI W, LU Z, *et al*. The use of RGD-engineered exosomes for enhanced targeting ability and synergistic therapy toward angiogenesis[J]. Nanoscale, 2017, 9(40): 15598-15605.
[41] BALA S, CSAK T, MOMEN-HERAVI F, *et al*. Biodistribution and function of extracellular miRNA-155 in mice[J]. Scientific Reports, 2015, 5: 10721.
[42] KANKI S, JAALOUK D E, LEE S, *et al*. Identification of targeting peptides for ischemic myocardium by *in vivo* phage display[J]. Journal of Molecular and Cellular Cardiology, 2011, 50(5): 841-848.
[43] TANG J, SU T. Targeted repair of heart injury by stem cells fused with platelet nanovesicles[J]. Nature Biomedical Engineering, 2018, 2: 17-26.
[44] LIU M, LI M, SUN S, *et al*. The use of antibody modified liposomes loaded with AMO-1 to deliver oligonucleotides to ischemic myocardium for arrhythmia therapy[J]. Biomaterials, 2014, 35(11): 3697-3707.
[45] NGUYEN C T, DAWKINS J, BI X, *et al*. Diffusion tensor cardiac magnetic resonance reveals exosomes from cardiosphere-derived cells preserve myocardial fiber architecture after myocardial infarction[J]. JACC Basic to Translational Science, 2018, 3(1): 97-109.

(上接第 135 页)

[51] MA Z, ZHANG B, WANG D, *et al*. Tanshinones suppress AURKA through up-regulation of miR-32 expression in non-small cell lung cancer[J]. Oncotarget, 2015, 6(24): 20111-20120.
[52] GAO L M, ZHENG Y, WANG P, *et al*. Tumor-suppressive effects of microRNA-181d-5p on non-small-cell lung cancer through the CDKN3-mediated Akt signaling pathway *in vivo* and *in vitro*[J]. American Journal of Physiology-Lung Cellular and Molecular Physiology, 2019, 316(5): L918-L933.
[53] WEI D M, CHEN W J, MENG R M, *et al*. Augmented expression of Ki-67 is correlated with clinicopathological characteristics and prognosis for lung cancer patients: an up-dated systematic review and meta-analysis with 108 studies and 14 732 patients[J]. Respiratory Research, 2018, 19: 150.
[54] ZHANG X, XIAO D, WANG Z, *et al*. MicroRNA-26a/b regulate DNA replication licensing, tumorigenesis, and prognosis by targeting CDC6 in lung cancer[J]. Molecular Cancer Research, 2014, 12(11): 1535-1546.
[55] MA X P, ZHANG W, WU B Q, *et al*. Correlations between mRNA levels of centrosomal protein 55 (CEP55) and clinical features of patients with lung cancer[J]. Medical Science Monitor, 2018, 24: 3093-3097.
[56] LI Y, TANG H, SUN Z, *et al*. Network-based approach identified cell cycle genes as predictor of overall survival in lung adenocarcinoma patients[J]. Lung Cancer, 2013, 80(1): 91-98.
[57] TAKAHASHI T, HARUKI N, NOMOTO S, *et al*. Identification of frequent impairment of the mitotic checkpoint and molecular analysis of the mitotic checkpoint genes, hsMAD2 and p55CDC, in human lung cancers[J]. Oncogene, 1999, 18(30): 4295-4300.
[58] YING H, XU Z, CHEN M, *et al*. Overexpression of ZWINT predicts poor prognosis and promotes the proliferation of hepatocellular carcinoma by regulating cell-cycle-related proteins[J]. OncoTargets and Therapy, 2018, 11: 689-702.
[59] PENG F, LI Q, NIU S Q, *et al*. ZWINT is the next potential target for lung cancer therapy[J]. Journal of Cancer Research and Clinical Oncology, 2019, 145(3): 661-673.
[60] BAKHOUM S F, THOMPSON S L, MANNING A L, *et al*. Genome stability is ensured by temporal control of kinetochore-microtubule dynamics[J]. Nature Cell Biology, 2009, 11(1): 27-35.