

DOI:10.12113/202104020

COPD 差异表达基因的生物信息学分析及在 LUSC 样本中的表达

刘艳玲*, 刘静, 童明琼, 范娜, 王晓玥, 孙婉

(德州学院医药与护理学院, 山东德州 253023)

摘要:为确定慢性阻塞性肺病(COPD)的分子标记物及 COPD 与肺鳞状细胞癌(LUSC)共存的差异表达基因,探寻 COPD 合并肺癌的预测因子,发现新的治疗靶点。本研究采用生物信息学方法,从 GEO 数据库中筛选 3 套基因芯片数据集,挖掘 COPD 患者小气道上皮细胞(SAEC)的差异表达基因(DEG)以及潜在的生物标记物,并通过基因本体(GO)、京都基因与基因组百科全书(KEGG)富集分析预测 DEGs 的功能及参与的代谢途径。继而对 DEGs 构建 PPI 网络,使用 Cytoscape 软件筛选子模块和 Hub 基因,并将 Hub 基因通过 TCGA 数据库分析其在 LUSC 中的差异表达情况及差异基因间的相关性。结果共获得 52 个上调基因和 24 个下调基因,代谢通路主要集中在细胞色素 P450 对外源物质的代谢、化学致癌、花生四烯酸代谢及甲状腺激素合成四条途径上,通过 Cytoscape 软件从 PPI 网络中筛选得到 2 个功能模块和 10 个 Hub 基因,进一步验证发现其中 5 个基因在 TCGA 数据库中的 LUSC 样本中同样差异表达。由此推测 *SPP1*、*ALDH3A1*、*SPRR3*、*KRT6A* 和 *SPRR1B* 可能为 COPD 分子标记物及 COPD 与 LUSC 共存的 DEGs,从而为研究 COPD 和 LUSC 的发病机制及二者潜在关系奠定良好的基础。

关键词:COPD; 差异表达基因; 生物信息学分析; LUSC

中图分类号:Q343.1 文献标志码:A 文章编号:1672-5565(2022)04-294-09

Bioinformatics analysis of differentially expressed genes in COPD and their differential expression in LUSC

LIU Yanling*, LIU Jing, TONG Mingqiong, FAN Na, WANG Xiaoyue, SUN Wan

(School of Medicine and Nursing, Dezhou University, Dezhou 253023, Shandong, China)

Abstract: In order to determine the molecular markers of chronic obstructive pulmonary disease (COPD) and the differential expression genes (DEGs) of COPD and lung squamous cell carcinoma (LUSC), explore the predictors of COPD and lung cancer, and find new therapeutic targets, three gene expression datasets were selected from GEO database, and DEGs and potential biomarkers of small airway epithelial cells (SAECs) of COPD were extracted based on bioinformatics method. The functions and metabolic pathways of DEGs were predicted by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomic (KEGG). Then, the modules and hub genes were screened by Cytoscape from PPI networks. Lastly, the difference expression of hub genes in LUSC and the correlation of the differential genes were analyzed using TCGA database. Results showed that 52 up-regulated and 24 down-regulated genes were obtained. These 76 DEGs were mainly related to metabolism of xenobiotics by cytochrome P450, chemical carcinogenesis, arachidonic acid metabolism, and thyroid hormone synthesis. Then two functional modules and ten hub genes were screened from PPI network by Cytoscape. Further analysis showed that five of the ten hub genes were DEGs in LUSC samples in TCGA database. So *SPP1*, *ALDH3A1*, *SPRR3*, *KRT6A*, and *SPRR1B* may be molecular markers in COPD and DEGs in both COPD and LUSC, which lays a foundation for the research of pathogenesis of COPD and LUSC and corresponding relations.

Keywords: COPD; Differentially expressed genes; Bioinformatics analysis; LUSC

慢性阻塞性肺病(Chronic Obstructive Pulmonary Disease, COPD)是一种典型的具有进行性、不可逆

的气流限制和反复呼吸道症状的慢性呼吸系统疾病,致残率和病死率很高,全球 40 岁以上发病率已

收稿日期:2021-04-28;修回日期:2021-10-20.

基金项目德州学院博士科研基金项目(No.2019xgrc25).

*通信作者:刘艳玲,博士,讲师,研究方向:生物信息学.E-mail:lyling9030@163.com.

高达 9–10%^[1-2]。COPD 并非单一疾病,它不仅损伤肺组织,同时还会损伤包括心脏、骨骼、免疫系统、消化系统和中枢神经系统在内的肺外器官和组织, COPD 目前已成为全球第三大死因^[3-4]。

COPD 患者肺部的慢性炎症反应可导致肺实质组织受损,并破坏正常的修复功能和防御机制,导致小气道纤维化,最终产生气体陷闭和进行性气流受限,成为导致气道阻力增加的主要因素^[5-6]。由此可见, COPD 与小气道病变息息相关,尽早干预轻中度 COPD 患者的小气道病变,对于延缓病程具有重要意义。研究表明,被诊断为肺癌的患者中 40%–70% 患有 COPD^[7],而 COPD 患者的肺癌发病率为 16.7%,且最长见的病理类型是肺鳞状细胞癌(Lung Squamous Cell Carcinoma, LUSC)^[8]。对于吸烟者来说, COPD 患者患肺癌的几率是非 COPD 患者的 4~6 倍^[9],并且轻中度 COPD 患者患肺癌的几率较重度 COPD 更高,这可能与轻中度 COPD 患者存在炎症和氧化应激增高有关^[10]。目前, COPD 与肺癌共存机制尚不清楚,如果能够探寻 COPD 合并肺癌的预测因子,研究两者共同的发病机制,发现新的治疗靶点,可为 COPD 的预防、治疗和肺癌的早期诊断、降低死亡率提供理论依据。

目前,关于 COPD 的生物信息学研究还较少,王小乐等通过对 COPD 患者肺泡巨噬细胞的芯片数据集进行生物信息学分析,共获得 43 个差异表达基因,且与免疫和炎症的发生和发展有关^[11]。谷雨等对 COPD 患者肺组织的芯片数据进行了生物信息学分析,找到了 FGG, FGA, IL-6, SERPINE1 和 SPP1 5 个 Hub 基因,根据调控途径的进一步分析发现差异基因表达可能与代谢及细胞生物功能异常有关^[12]。

本研究采用生物信息学方法,从 GEO (Gene Expression Omnibus) 数据库中筛选 3 个 COPD 芯片数据集,挖掘 COPD 患者小气道上皮细胞 (Small airway epithelium, SAEC) 的差异表达基因 (Differentially Expressed Genes, DEGs) 以及潜在的生物标记物,并通过基因本体 (Gene Ontology, GO) 和京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 富集分析预测 DEGs 的功能及参与的代谢途径。通过对 DEGs 进行 COPD 的蛋白质网络互作分析,构建 PPI (Protein-Protein Interaction) 网络,并使用 Cytoscape 软件中的 CytoHubba 和 MCODE 筛选子模块和 Hub 基因。最后将 10 个 Hub 基因通过 TCGA 和 GTEx 数据库验证其在 LUSC 样本中的差异表达情况及差异基因间的相关性,为确定 COPD 的分子标记物及 COPD 与

LUSC 共存的 DEGs 奠定一定的基础。

1 材料与方法

1.1 材料

使用“COPD”作为关键词,在 GEO 数据库 (<http://www.ncbi.nlm.nih.gov/geo/>) 中搜索并选择了 3 个芯片数据集: GSE11784、GSE11906 和 GSE20257,均来自于 GPL570 平台 (Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix, Inc)。由于芯片数据都出自威尔·康奈尔医学院遗传医学系实验室,其中的部分样本在不同芯片间存在重复,去除重复样本后,最终 GSE11784 纳入 53 例正常对照和 14 例 COPD 患者, GSE11906 纳入 46 例正常对照和 15 例 COPD 患者, GSE20257 纳入 22 例正常对照和 9 例 COPD 患者 (见表 1), 3 个芯片共纳入 159 例样本,其中包含 121 例正常对照和 38 例 COPD 患者。

表 1 GEO 数据库 COPD 芯片数据集

Table 1 COPD microarray datasets from GEO database

芯片数据集	样本	正常样本/个	COPD 样本数/个	平台
GSE11784	SAG	53	14	GPL570
GSE11906	SAG	46	15	GPL570
GSE20257	SAG	22	9	GPL570
合计		121	38	

注:缩写: SAG, Small airway epithelium; 小气道上皮细胞。

1.2 方法

1.2.1 微阵列数据处理

分别对 3 个数据集探针的矩阵文件取 log₂, 然后使用 Perl 5.0 软件 (<http://www.perl.org>) 对数据进行注释、合并,再通过 R 软件 (4.0.3 版本) 的 sva 和 limma 包对合并后的数据进行批次校正。

1.2.2 DEGs 分析

使用 R 软件中的 Limma 包对正常样本和 COPD 样本进行比较,在校正后的 P 值 (*AdjP*) < 0.05 的条件下鉴定 DEGs。

1.2.3 Go 和 KEGG 富集分析

使用 R 软件中的 ClusterProfiler 和 org.Hs.eg.db 包进行 GO 和 KEGG 富集分析,以 *P* < 0.05 定义差异具有统计学意义。

1.2.4 PPI 网络分析

使用在线网站 STRING (<http://string-db.org>) 分析 DEGs 的相互作用关系,选择最小交互作用得分 > 0.4 的蛋白质对构建 PPI 网络。然后分别使用 Cytoscape 中的 MCODE (*degree cut* ≥ 2, *node score cut*

≥ 0.2 , $K\text{-core} \geq 2$, $\max\ depth = 100$) 和 CytoHubba 插件筛选重要的子模块和 Hub 基因。

1.2.5 Hub 基因在 TCGA 和 GTEx 数据库中的显著性验证

使用在线网站 GEPIA (Gene Expression Profiling Interactive Analysis) (<http://gepia.cancer-pku.cn/index.html>), 选择 TCGA 和 GTEx 数据库中的 LUSC 样本 (486 例) 和正常样本 (338 例) 对 Hub 基因进行差异性验证 ($|\log_2FC| > 2$, $p\text{-value} < 0.05$, $Jitter\ Size = 0.4$)。

1.2.6 Hub 基因的相关性分析

使用在线网站 GEPIA 网站, 选择 TCGA 数据库中的 LUSC 样本 (486 例) 和正常样本 (50 例) 对 Hub 基因进行差异分析。

2 结果与分析

2.1 COPD 患者的 DEGs

通过对 GSE11784、GSE11906 和 GSE20257 进行 RNA 水平上正常样本和 COPD 样本分析, 共获

得 76 个 DEGs, 包括 52 个上调基因和 24 个下调基因 (见表 2), DEGs 的火山图和热图 (见图 1 和图 2)。

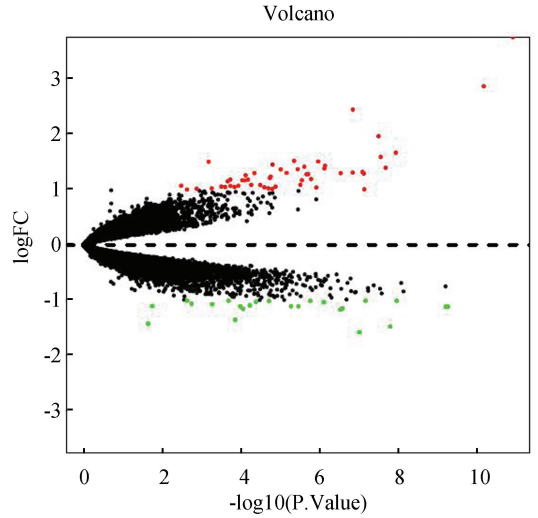


图 1 DEGs 的火山图

Fig.1 Volcano plot for DEGs

注: 图中红点代表显著上调基因, 绿点代表显著下调基因, 黑点代表非显著表达基因, 所有差异表达基因符合校正后 p 值 < 0.05 。

表 2 三个 GEO 数据集中的 76 个 DEGs

Table 2 A total of 76 DEGs in three GEO datasets

差异表达基因 (DEGs)	基因
上调	<i>CYP1A1</i> 、 <i>CYP1B1</i> 、 <i>SLC7A11</i> 、 <i>LINC00942</i> 、 <i>CABYR</i> 、 <i>CCL2</i> 、 <i>SPRR1A</i> 、 <i>GAD1</i> 、 <i>MMP12</i> 、 <i>AKR1B10</i> 、 <i>AHRR</i> 、 <i>TCN1</i> 、 <i>ST3GAL4-AS1</i> 、 <i>LOC284825</i> 、 <i>TPRXL</i> 、 <i>PHOX</i> 、 <i>GPX2</i> 、 <i>DUOX2</i> 、 <i>CYP4F2</i> 、 <i>SPRR3</i> 、 <i>CGREF1</i> 、 <i>NROB1</i> 、 <i>SFRP2</i> 、 <i>CLEC5A</i> 、 <i>UCHL1</i> 、 <i>MUCL1</i> 、 <i>FCGR2B</i> 、 <i>SCG3</i> 、 <i>CBR1</i> 、 <i>MEP1A</i> 、 <i>RP11-203B7.1</i> 、 <i>PRKG1-AS1</i> 、 <i>EGF</i> 、 <i>LINC101082</i> 、 <i>BPIFB2</i> 、 <i>PTPRH</i> 、 <i>KRT6A</i> 、 <i>KRT13</i> 、 <i>SPP1</i> 、 <i>STATH</i> 、 <i>LOC344887</i> 、 <i>SPRR1B</i> 、 <i>PRSS50</i> 、 <i>JAKMIP3</i> 、 <i>ALDH3A1</i> 、 <i>DUOX2</i> 、 <i>LOC101928386</i> 、 <i>ERVMER61-1</i> 、 <i>BPIFA1</i> 、 <i>HS3ST3A1</i> 、 <i>BMP4</i> 、 <i>PROK2</i>
下调	<i>THSD7A</i> 、 <i>PCDH17</i> 、 <i>KCNA1</i> 、 <i>C8orf12</i> 、 <i>LOC283177</i> 、 <i>LTF</i> 、 <i>TMEM45A</i> 、 <i>LIPF</i> 、 <i>NOC4L</i> 、 <i>LAG3</i> 、 <i>VGLL3</i> 、 <i>LOC100507560</i> 、 <i>WIF1</i> 、 <i>LINC00277</i> 、 <i>SEC14L3</i> 、 <i>BC045791</i> 、 <i>LOC100130744</i> 、 <i>ITLN1</i> 、 <i>PF4</i> 、 <i>MT1M</i> 、 <i>APELA</i> 、 <i>FLJ37786</i> 、 <i>HBM</i> 、 <i>SCGB3A2</i>

2.2 GO 富集分析

通过 R 软件的 ClusterProfiler 和 org.Hs.eg.db 包进行 DEGs 的 GO 富集分析 (见表 3), DEGs 主要富

集在受体-配体活性、信号受体激活剂活性、四吡咯结合、内肽酶活性、血红素结合和 G 蛋白偶联受体结合等功能上。

表 3 COPD 差异表达基因的 GO 富集分析

Table 3 GO enrichment analysis of DEGs in COPD samples

GO 富集通路编号	生物学过程	节点	P 值
GO:0048018	受体-配体活性	7	7.45×10^{-4}
GO:0030546	信号受体激活剂活性	7	7.91×10^{-4}
GO:0046906	t 四吡咯结合	6	6.43×10^{-6}
GO:0004175	内肽酶活性	6	2.37×10^{-3}
GO:0020037	血红素结合	5	6.84×10^{-5}
GO:0001664	G 蛋白偶联受体结合	5	2.13×10^{-3}
GO:0016614	氧化还原酶活性	4	7.1×10^{-4}
GO:0005506	铁离子结合	4	1.24×10^{-3}
GO:0005125	细胞因子活性	4	6.09×10^{-3}
GO:0070330	芳香化酶活性	3	6.22×10^{-05}

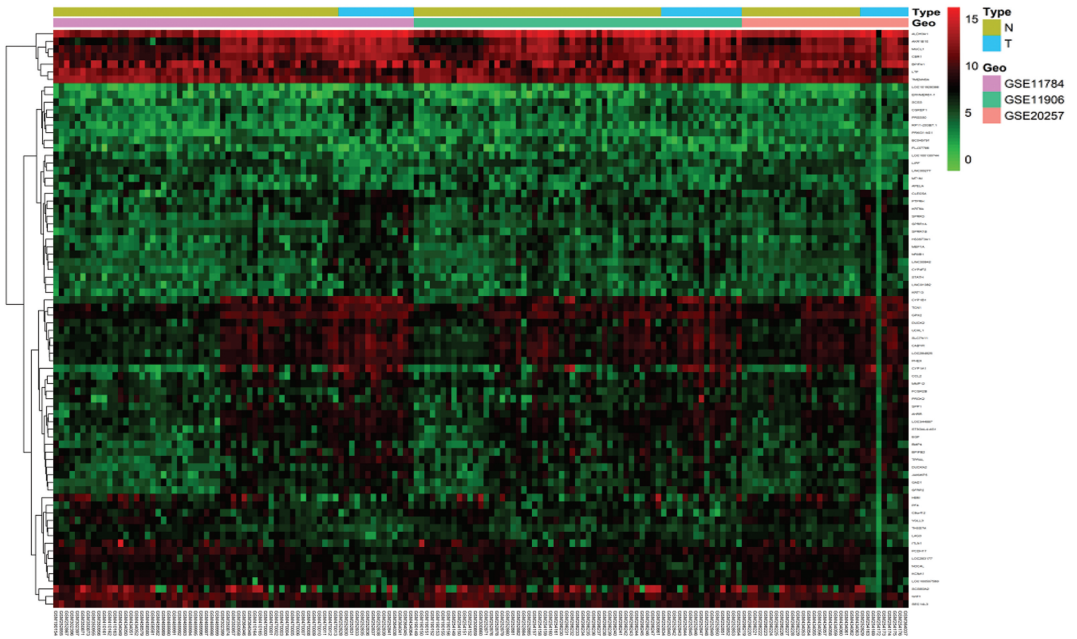


图 2 COPD 和正常样本 DEGs 热图

Fig.2 Heat map for DEGs between COPD and normal samples

注:图中从红色到绿色,基因表达水平逐渐下降,所有差异表达基因符合校正后 p 值 < 0.05 ; N 代表正常样本, T 代表肿瘤样本。

2.3 KEGG 代谢通路分析

通过 R 软件的 ClusterProfiler 包进行 DEGs 的 KEGG 代谢通路分析,结果(见如图 3 和表 4), DEGs 主要富集在细胞色素 P450 对外源物质的代谢、化学致癌、花生四烯酸代谢和甲状腺激素合成 4 条代谢途径中。

2.4 COPD 差异表达基因蛋白质网络互作分析

使用在线网站 STRING (<http://string-db.org>) 对 52 个上调基因和 24 个下调基因进行 COPD 的蛋白质互作分析,构建 PPI 网络(见图 4),每个基因代表一个节点,每 2 个蛋白质之间的连线代表一个关系对,总共有 36 个节点和 51 个关系对。

使用 Cytoscape 软件中的 MCODE 从 PPI 网络中筛选得到 2 个功能模块(见图 5),进一步使用 CytoHubba 来确定 Hub 基因(见图 6),评分前十位的基因分别是 *SPP1*、*EGF*、*CCL2*、*ALDH3A1*、*BMP4*、*SPRR3*、*KRT6A*、*BPIFB2*、*CYP1A1* 和 *SPRR1B*,且都属于上调基因。与图 5 比较发现,这 10 个基因有 7 个存在于两个功能模块中,因此将这 10 个基因作为 Hub 基因做进一步分析对于 COPD 患病机理及分子

机制具有一定的研究意义。

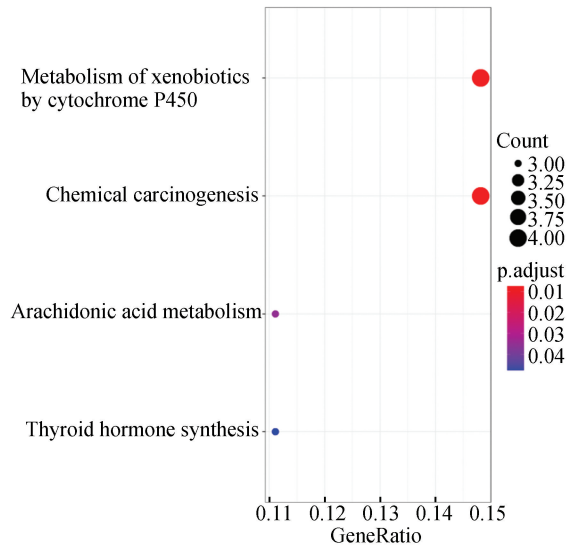


图 3 COPD 差异表达基因的 KEGG 代谢通路富集分析

Fig.3 KEGG enrichment analysis of DEGs in COPD

注: X 轴代表差异基因中与该通路相关的基因数与整个差异基因总数的比值, Y 轴代表 KEGG 富集通路。

表 4 COPD 差异表达基因的 KEGG 代谢通路富集分析

Table 4 KEGG pathway analysis of DEGs in COPD samples

代谢途径	ID	节点	p 值	校正后 p 值	基因
细胞色素 P450 对外源物质的代谢	hsa00980	4	1.18×10^{-4}	7.502×10^{-3}	<i>CYP1A1</i> 、 <i>CYP1B1</i> 、 <i>CBR1</i> 、 <i>ALDH3A1</i>
化学致癌	hsa05204	4	1.50×10^{-4}	7.502×10^{-3}	<i>CYP1A1</i> 、 <i>CYP1B1</i> 、 <i>CBR1</i> 、 <i>ALDH3A1</i>
花生四烯酸代谢	hsa00590	3	1.04×10^{-3}	3.479×10^{-2}	<i>GPX2</i> 、 <i>CYP4F2</i> 、 <i>CBR1</i>
甲状腺激素合成	hsa04918	3	1.90×10^{-3}	4.746×10^{-2}	<i>GPX2</i> 、 <i>DUOXA2</i> 、 <i>DUOX2</i>

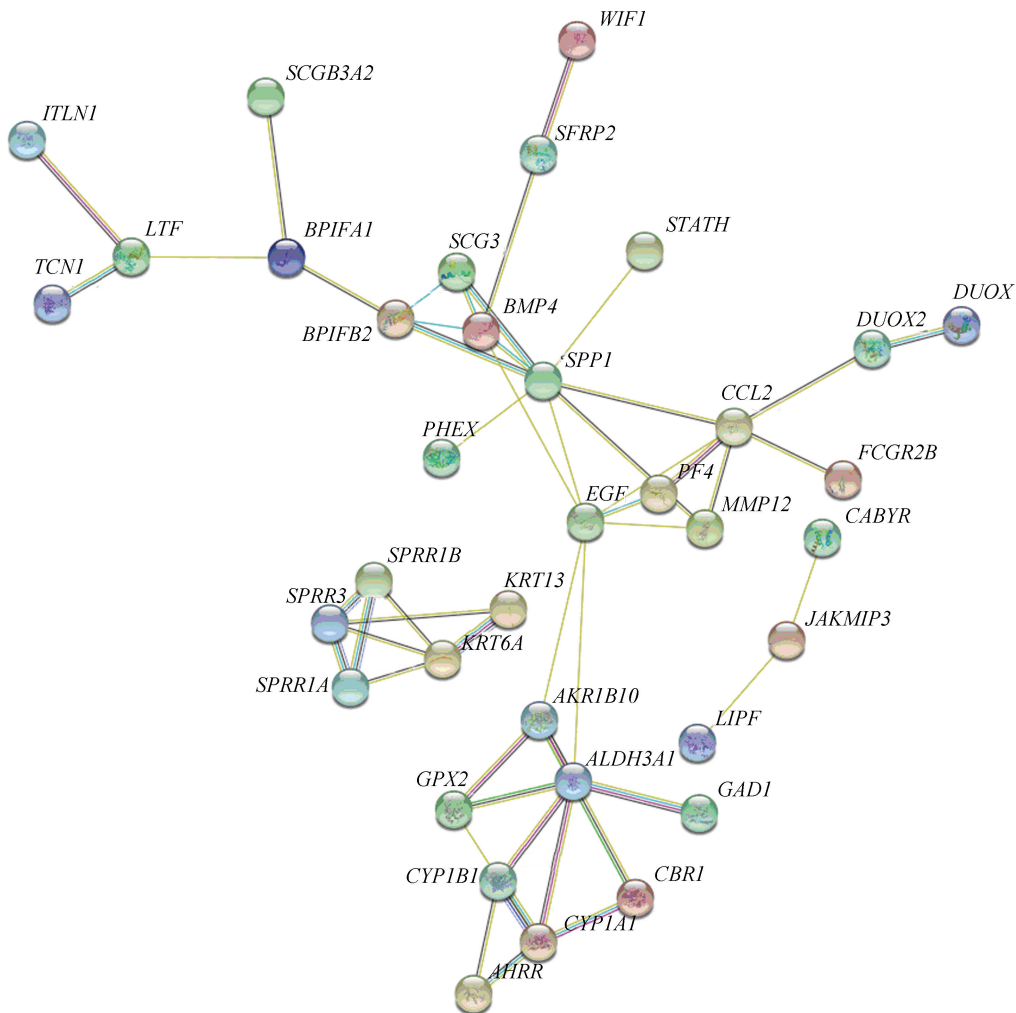


图4 COPD 差异基因的 PPI 网络

Fig.4 PPI network of DEGs in COPD samples

注:图中圆圈代表基因,连线代表基因间蛋白质的相互作用。

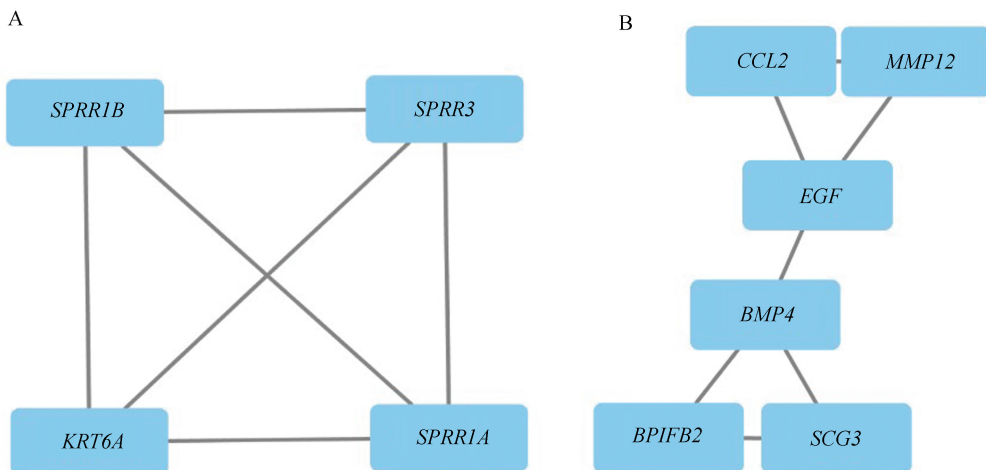


图5 用 MCODE 插件从 PPI 网络中筛选获得两个关键功能模块(A 和 B)

Fig.5 PPI network of two key modules obtained by MCODE in cytoscape(A and B)

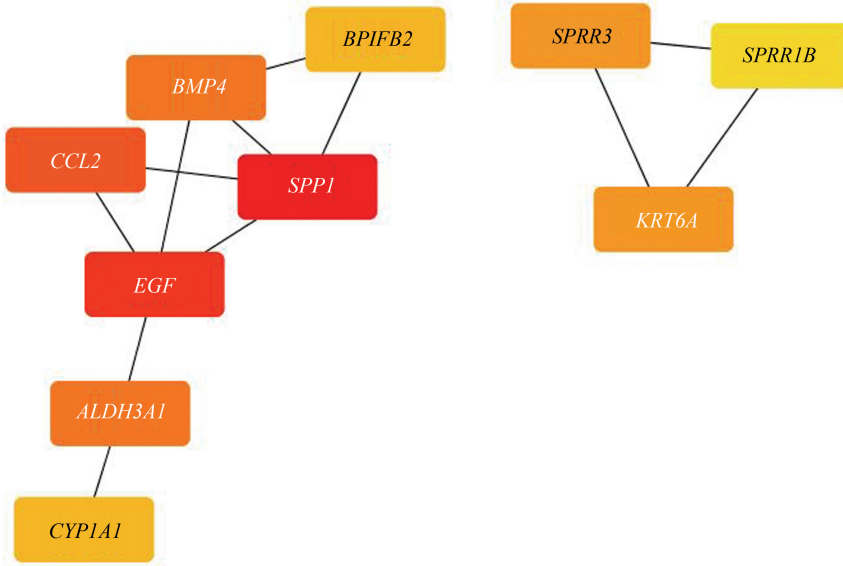


图 6 通过 PPI 网络筛选 10 个 Hub 基因
 Fig.6 Ten hub genes screened by PPI network

注:图中颜色越深,基因评分越高。

2.5 Hub 基因在 TCGA 和 GTEx 数据库中的差异性验证

TCGA (The Cancer Genome Atlas) 是一个涵盖 33 种癌症及对应正常样本的数据库, GTEx (Genotype Tissue Expression) 数据库仅包含正常人的样本, 通常与 TCGA 数据库结合使用, 以弥补 TCGA 正常样本的不足。本研究中, 为探寻 COPD 与 LUSC 共同的生物标记

物, 将 GEO 数据库中获得 COPD 的 10 个 Hub 基因在 TCGA 和 GTEx 数据库的 LUSC 样本中进行差异性验证。基因 *SPP1*、*ALDH3A1*、*SPRR3*、*KRT6A* 和 *SPRR1B* 均具有显著性差异(见图 7), 表明这 5 个基因是 COPD 和 LUSC 共同的 DEGs。进一步分析发现, *SPP1* 和 *ALDH3A1*、*SPRR3* 和 *KRT6A* 呈显著正相关(见图 8), 其它基因间无显著性关系。

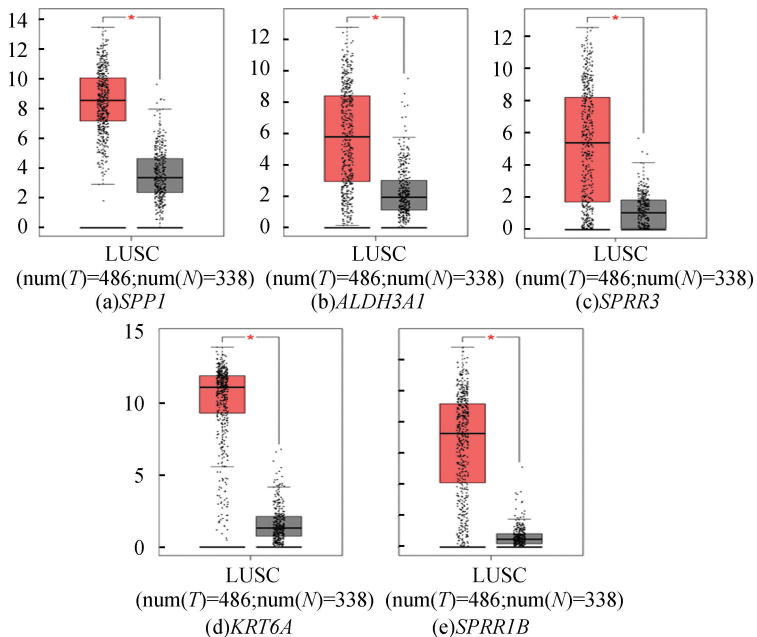


图 7 Hub 基因 *SPP1*(a)、*ALDH3A1*(b)、*SPRR3*(c)、*KRT6A*(d)、*SPRR1B*(e) 在 TCGA 和 GTEx 数据库中 LUSC 样本中的差异性验证

Fig.7 Significance verification of *SPP1*(a), *ALDH3A1*(b), *SPRR3*(c), *KRT6A*(d), and *SPRR1B*(e) in LUSC samples in TCGA and GTEx databases

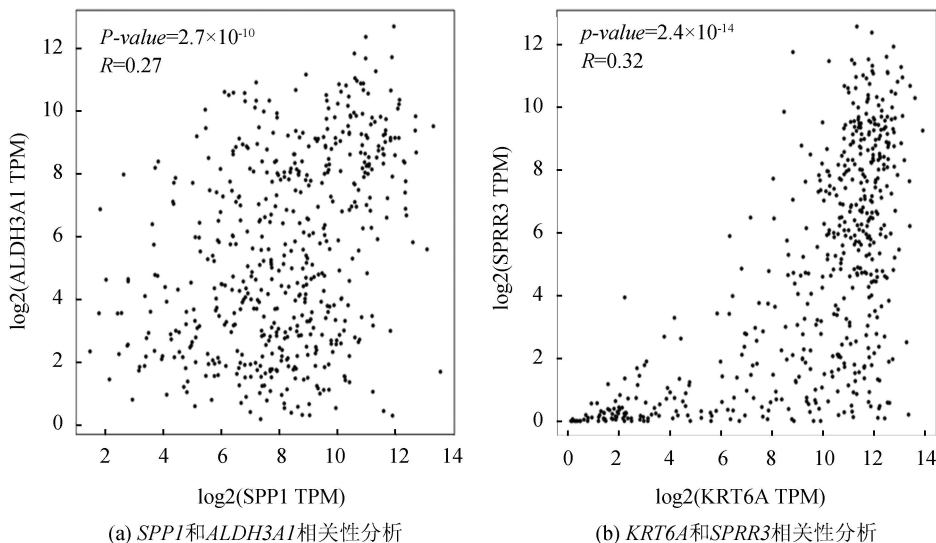


图8 Hub基因相关性分析

Fig.8 Correlation analysis of hub genes

注:A;SPP1和ALDH3A1相关性分析,B;KRT6A和SPRR3相关性分析.

3 讨论

COPD是一种常见的以持续气流受限为特征的疾病,气流受限进行性发展,与气道和肺脏对有毒颗粒或气体的慢性炎症反应增强有关^[13]。小气道管腔狭窄,空气阻力将成倍增加,周围气道阻力增高是小气道病变的重要病理生理学特征,故当小气道有炎症或痰液阻塞,很容易造成闭合、萎陷,COPD病变也多从小气道开始^[14]。

本研究以小气道上皮细胞为研究对象,筛选3个COPD的GEO芯片数据集进行生物信息学分析,获得了52个上调基因和24个下调基因。GO分析发现DEGs主要富集在受体-配体活性、信号受体激活剂活性、四吡咯结合、内肽酶活性、血红素结合和G蛋白偶联受体结合等功能上,KEGG代谢通路主要集中在细胞色素P450对外源物质的代谢、化学致癌、花生四烯酸代谢和甲状腺激素合成4条途径上。进一步分析发现参与GO和KEGG代谢通路的DEGs均为上调基因,这些功能和代谢通路均与COPD发病机理有密切关系。程序性细胞死亡受体1(PD-1)通过与其配体PD-L1结合,负调控T淋巴细胞活化及效应功能,参与免疫耐受的维持,从而促进了COPD的发生和发展过程^[15]。乙酰胆碱受体M₃是引起气道平滑肌收缩的主要受体,激活后发生构象改变,并进一步激活磷脂酶C,催化二磷酸磷脂酰肌醇水解为三磷酸肌醇,促进钙离子通道开放,引起气道平滑肌收缩,气流阻力增大^[16]。花生四烯酸是在炎症反应中由细胞膜磷脂在磷脂酶A₂催化下

产生的,并经过酶促反应生成多种代谢物,它们可以通过多种途径刺激呼吸道中的伤害性感受器,导致炎症细胞聚集而发生局部效应^[17]。由此可见筛选得到的差异表达基因在COPD的发病机制中占有重要地位,可为后续研究奠定良好的基础。

进一步取交互作用得分>0.4的蛋白质构建PPI网络,发现共有36个基因参与到网络构建中,为获得关键基因,使用Cytoscape软件中的MCODE和CytoHubba插件对PPI网络图进行分析,分别获得了两个重要的子模块和10个Hub基因,比较发现有7个Hub基因位于子模块中,其中ALDH3A1和CYP1A1为细胞色素P450对外源物质的代谢和化学致癌途径中的重要基因(见表4),因此可初步认为筛选得到的Hub基因对于进一步研究COPD患病机理及分子机制具有一定的研究意义。为寻求COPD与LUSC共存的DEGs,将10个Hub基因通过TCGA和GTEX数据库验证其在LUSC中的差异表达情况,从而确定SPP1、ALDH3A1、SPRR3、KRT6A和SPRR1B为COPD与LUSC共存的DEGs。进一步分析其相关性发现仅SPP1和ALDH3A1、SPRR3和KRT6A间呈显著正相关(见图8),目前尚未有该基因对间作用机制的相关报道,是否存在共同的调节因子还有待进一步研究。

早期研究并没有发现SPP1与COPD之间的联系^[18],血液分析结果显示癌症分期和SPP1表达水平相互矛盾^[19],而最新报道表明SPP1在COPD和肺癌患者中均高表达,并且与肺癌较低的生存率密切相关^[20],这与本研究结果一致,可以初步认为COPD中SPP1的高表达与患肺癌风险增加有关,

SPP1 可能是一个治疗靶点,用于延缓 COPD 患者肺癌的发展,提高其生存时间。

ALDH3A1 属于乙醛脱氢酶家族 3 中的成员 A1,在人体的胃、食道和肺中含量较高。通过前列腺肿瘤异种移植的动物实验发现 *ALDH3A1* 可以促进前列腺肿瘤细胞肺转移的发生^[21]。*ALDH3A1* 基因表达还与大肝癌早期复发转移关系密切,并且该基因表达越高,复发转移时间越长^[22]。目前关于 *ALDH3A1* 对 COPD 和 LUSC 之间的关系还鲜有报道,通过 KEGG 代谢通路分析我们可以看到 *ALDH3A1* 参与了细胞色素 P450 对外源物质的代谢和化学致癌两个途径(见表 4)。细胞色素 P450 是人体内主要的多环芳烃代谢酶,参与内外源性物质的代谢,对于化学致癌物来说,可把无活性的前致癌物激活转变为电子化合物,与 DNA 或蛋白质结合,导致癌变。由于 *ALDH3A1* 在 COPD 和 LUSC 中均上调,可以认为高表达的 *ALDH3A1* 加速了细胞色素 P450 对外源物质尤其是化学致癌物质的代谢,从而促进了 COPD 和 LUSC 的发生和发展。

SPRRs 为一类富含脯氨酸蛋白的基因家族,该家族共有 4 个成员:*SPRR1*、*SPRR2*、*SPRR3* 和 *SPRR4*。它们蛋白产物的中央片段均由串联重复的氨基酸(XKXPEPXX(X 为任意氨基酸))序列组成:*SPRR1*、*SPRR3* 和 *SPRR4* 为 8 个,*SPRR2* 为 9 个。本研究涉及到的 Hub 基因为 *SPRR1B* 和 *SPRR3*。研究表明这两个基因参与多种癌症的发生和发展。*SPRR1B* 在口腔鳞癌组织中过表达,从而影响细胞增长^[23],它还参与黑色素瘤的转移^[24],并且是气道上皮细胞鳞状细胞分化的早期生物标志物,其表达下调伴随细胞的恶性转化^[25]。通过 meta 和 Kaplan-Meier 分析,发现 *SPRR1B* 在肺腺癌组织中表达增强,敲除该基因可抑制肺腺癌细胞增殖、迁移和侵袭;此外,通过火山图发现肺癌组织中 *SPRR3* 的表达水平也明显增高^[26]。同样,*SPRR1B* 和 *SPRR3* 在肺癌 H1299 细胞系中均过表达^[27]。本研究与前人结果一致,*SPRR1B* 和 *SPRR3* 在 COPD 和 LUSC 中均表达上调,提示这两个基因在 COPD 和 LUSC 发生机制的探讨中具有潜在的研究价值,建议 *SPRR* 家族基因与 COPD 和肺癌的关系应进一步深入研究。

4 结 论

综上所述,*SPP1*、*ALDH3A1*、*SPRR3*、*KRT6A* 和 *SPRR1B* 可作为 COPD 的分子标记物及 LUSC 的 DEGs,尤其是 *ALDH3A1* 和 *SPRR* 家族基因与 COPD 和 LUSC 间的关系及作用机制,有望做进一步深入

的研究,为找到 COPD 和 LUSC 新的治疗靶点提供理论依据。

参考文献(References)

- [1] WANG C, XU J Y, YANG L, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study[J]. *Lancet*, 2018, 391(10131): 1706-1717. DOI:10.1016/S0140-6736(18)30841-9.
- [2] ZHANG Q, SONG W, AYIDAERHAN N, et al. PTPLAD2 and USP49 involved in the pathogenesis of smoke-induced COPD by integrative bioinformatics analysis[J]. *International Journal of COPD*, 2020, 15: 2515-2526. DOI:10.2147/COPD.S250576.
- [3] SINGH D, AGUSTI A, ANZUETO A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD science committee report 2019[J]. *European Respiratory Journal*, 2019, 53(5): 1900164. DOI:10.1183/13993003.00164-2019.
- [4] LOZANO R, NAGHAVI M, FOREMAN K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010[J]. *Lancet*, 2016, 380(9859): 2095-2128. DOI:10.1016/S0140-6736(16)11728-0.
- [5] DEEPAK D. Recognition of small airways obstruction in asthma and COPD—the road less travelled[J]. *Journal of Clinical & Diagnostic Research*, 2017, 11(3): TE01-TE05. DOI:10.7860/JCDR/2017/19920.9478.
- [6] 王晓娟, 方向阳. 慢性阻塞性肺疾病全球倡议 2019: 慢性阻塞性肺疾病诊断、治疗与预防全球策略解读[J]. *中国全科医学*, 2019, 22(18): 2141-2149. DOI:10.12114/j.issn.1007-9572.2019.00.064. WANG Xiaojuan, FANG Xiangyang. Interpretation of the global strategy for the diagnosis, management, and prevention of COPD 2019 by GOLD[J]. *Chinese General Practice*, 2019, 22(18): 2141-2149. DOI:10.12114/j.issn.1007-9572.2019.00.064.
- [7] LOGANATHAN R S, STOVER D E, SHI W, et al. Prevalence of COPD in women compared to men around the time of diagnosis of primary lung cancer[J]. *Chest*, 2006, 129(5): 1305-1312. DOI:10.1378/chest.129.5.1305.
- [8] DE TORRES J, MARFN J M, CASANOVA C, et al. Lung cancer in patients with chronic obstructive pulmonary disease—incidence and predicting factors[J]. *American Journal of Respiratory and Critical Care Medicine*, 2011, 184(8): 913-919. DOI:10.1164/rccm.201103-0430OC.
- [9] YOUNG R P, HOPKINS R J, CHRISTMAS T, et al. COPD prevalence is increased in lung cancer, independent of age, sex and smoking history[J]. *European Respiratory Journal*, 2009, 34(2): 380-386. DOI:10.1183/09031936.00144208.
- [10] BRETT A S, AL-HASAN M N. COPD exacerbations—A target for antibiotic stewardship[J]. *New England Journal*

- of Medicine, 2019, 381 (2): 174–185. DOI: 10.1056/NEJMe1905520.
- [11] 王小乐, 朱洁, 童佳兵, 等. 慢性阻塞性肺疾病肺泡巨噬细胞差异表达基因的生物信息学分析[J]. 细胞与分子免疫学杂志, 2020, 36 (11): 961–966. DOI: 10.13423/j.cnki.cjcmi.009091.
- WANG Xiaole, ZHU Jie, TONG Jiabing, et al. Bioinformatical analysis of differentially expressed genes in alveolar macrophages of chronic obstructive pulmonary disease (COPD) [J]. Chinese Journal of Cellular and Molecular Immunology, 2020, 36 (11): 961–966. DOI: 10.13423/j.cnki.cjcmi.009091.
- [12] 谷雨, 王文俊, 胡悦, 等. 基于芯片数据的慢性阻塞性肺疾病患者的生物信息学分析[J]. 国际呼吸杂志, 2021, 41 (10): 736–743. DOI: 10.3760/cma.j.cn131368-20210128-00078.
- GU Yu, WANG Wenjun, HU Yue, et al. Bioinformatics analysis in chronic obstructive pulmonary disease patients based on microarray data [J]. International Journal of Respiration, 2021, 41 (10): 736–743. DOI: 10.3760/cma.j.cn131368-20210128-00078.
- [13] 王志华, 谢俊刚. 慢性阻塞性肺疾病的表观遗传学机制研究进展[J]. 中华结核和呼吸杂志, 2017, 40 (6): 472–474. DOI: 10.3760/cma.j.issn.1001-0939.2017.06.015.
- WANG Zhihua, XIE Jungang. Research progress on epigenetic mechanism of COPD [J]. Chinese Journal of Tuberculosis and Respiratory Diseases, 2017, 40 (6): 472–474. DOI: 10.3760/cma.j.issn.1001-0939.2017.06.015.
- [14] 朱金源, 马莉琼, 张锦. 家族序列相似性 13A 基因干扰对人气道上皮细胞凋亡和增殖的影响及与慢性阻塞性肺疾病患者小气道重塑的关系[J]. 中华医学杂志, 2020, 100 (32): 2481–2487. DOI: 10.3760/cma.j.cn112137-20200326-00945.
- ZHU Jinyuan, MA Liqiong, ZHANG Jin. Effect of family with sequence similarity 13 member A gene interference on apoptosis and proliferation of human airway epithelial cells and its relationship with small airway remodeling in patients with chronic obstructive pulmonary disease [J]. National Medical Journal of China, 2020, 100 (32): 2481–2487. DOI: 10.3760/cma.j.cn112137-20200326-00945.
- [15] 周建鹏, 邓静敏, 段敏超. 程序性细胞死亡受体 1 与慢性阻塞性肺疾病中 T 淋巴细胞免疫[J]. 中华微生物学和免疫学杂志, 2021, 41 (3): 237–241. DOI: 10.3760/cma.j.cn112309-20200713-00352.
- ZHOU Jianpeng, DENG Jingmin, DUAN Minchao. Programmed cell death 1 and T lymphocyte immunity in chronic obstructive pulmonary disease [J]. Chinese Journal of Microbiology and Immunology, 2021, 41 (3): 237–241. DOI: 10.3760/cma.j.cn112309-20200713-00352.
- [16] GOSENS R, ZAAGSMA J, MEURS H, et al. Muscarinic receptor signaling in the pathophysiology of asthma and COPD [J]. Respiratory Research, 2006, 7: 73. DOI: 10.1186/1465-9921-7-73.
- [17] 林树新, 於峻. 花生四烯酸代谢物对呼吸道感受器的作用[J]. 生理学报, 2007 (2): 141–149. DOI: 10.13294/j.aps.2007.02.005.
- LIN Shuxin, YU Jun. Effects of arachidonic acid metabolites on airway sensors [J]. Acta Physiologica Sinica, 2007 (2): 141–149. DOI: 10.13294/j.aps.2007.02.005.
- [18] PAPANORFYRIOU A, LOUKIDES S, KOSTIKAS K, et al. Increased levels of osteopontin in sputum supernatant in patients with COPD [J]. Chest, 2014, 146 (4): 951–958. DOI: 10.1378/chest.13-2440.
- [19] RUD A K, BOYE K, OIJORDSBAKKEN M, et al. Osteopontin is a prognostic biomarker in non-small cell lung cancer [J]. BMC Cancer, 2013, 13 (1): 540. DOI: 10.1186/1471-2407-13-540.
- [20] MIAO T, XIAO W, DU L, et al. High expression of SPP1 in patients with chronic obstructive pulmonary disease (COPD) is correlated with increased risk of lung cancer [J]. FEBS Open Bio, 2021, 11 (4): 1237–1249. DOI: 10.1002/2211-5463.13127.
- [21] YAN J, MELO J D, CUTZ J C, et al. Aldehyde dehydrogenase 3A1 associates with prostate tumorigenesis [J]. British Journal of Cancer, 2014, 110 (10): 2593–2603. DOI: 10.1038/bjc.2014.201.
- [22] 魏复群. ALDH3A1 基因表达与大肝癌早期复发转移关系研究[D]. 福州: 福建医科大学, 2014.
- WEI Fuqun. Study on the relationship between the expression of ALDH3A1 gene and early recurrence and metastasis of large hepatocellular carcinoma [D]. Fuzhou: Fujian Medical University, 2014.
- [23] MICHIFURI Y, HIROHASHI Y, TORIGOE T, et al. Small proline-rich protein-1B is overexpressed in human oral squamous cell cancer stem-like cells and is related to their growth through activation of MAP kinase signal [J]. Biochemical and Biophysical Research Communications, 2013, 439 (1): 96–102. DOI: 10.1016/j.bbrc.2013.08.021.
- [24] SHENG Z, HAN W, HUANG B, et al. Screening and identification of potential prognostic biomarkers in metastatic skin cutaneous melanoma by bioinformatics analysis [J]. Journal of Cellular and Molecular Medicine, 2020, 24 (19): 11613–11618. DOI: 10.1111/jcmm.15822.
- [25] TESFAIGZI J, CARLSON D M. Expression, regulation, and function of the SPR family of proteins [J]. Cell Biochemistry & Biophysics, 1999, 30 (2): 243–265. DOI: 10.1007/BF02738069.
- [26] ZHANG Z, SHI R, XU S, et al. Identification of small proline-rich protein 1B (SPRR1B) as a prognostically predictive biomarker for lung adenocarcinoma by integrative bioinformatic analysis [J]. Thoracic Cancer, 2021, 2 (4): 796–806. DOI: 10.1111/1759-7714.13836.
- [27] XIONG Y, LI M, ZHANG P, et al. Study on genotype in lung squamous carcinoma by high-throughput of transcriptome sequence [J]. Chinese Journal of Lung Cancer, 2017, 20 (11): 727–731. DOI: 10.3779/j.issn.1009-3419.2017.11.01.