

论著·基础研究

转录组分析鉴定肺腺癌潜在的生物标志物

张伟然^{1,2}, 林雪峰³, 李鑫², 张浩², 王猛², 孙伟², 韩兴鹏², 孙大强^{1,4}

1. 天津医科大学研究生院, 天津 300070; 2. 天津市胸科医院胸外科, 天津 300222; 3. 天津医学高等专科学校护理系, 天津 300222; 4. 天津市南开医院胸外科, 天津 300100

[摘要] **目的**·通过转录组数据分析, 寻找相关的分子标志物, 用于肺腺癌的诊断和治疗。**方法**·通过差异表达分析, 对癌组织相对于正常组织的差异表达基因 (differentially expressed genes, DEGs)、差异表达 miRNAs (differentially expressed miRNAs, DEMs) 和差异表达 circRNAs (differentially expressed circRNAs, DECs) 进行筛选。对 DEGs 进行功能富集和通路分析, 并对 DEMs 进行靶向预测。构建 DEGs 和 DEMs 的调控网络, 并选取部分节点作为潜在的生物标志物。通过癌症基因组图谱 (the Cancer Genome Atlas, TCGA) 数据库和实时荧光定量 PCR (qRT-PCR) 对获得的生物标志物进行验证, 并分析其表达水平与总生存期和肿瘤分期的相关性。**结果**·在 3 个基因表达谱的 3 组 DEGs 中有 61 个重叠 DEGs, 富集在 32 个基因本体 (gene ontology, GO) 和 10 个通路中。共鉴定出 24 个 DEMs, 并筛选出 612 个 miRNA-DEGs 靶向关系对。在 circRNAs 表达谱中得到 92 个 DECs。ADRA1A、hsa-miR-141-5p 和 hsa-miR-191-3p 是调控网络中重要的节点。TCGA 和 qRT-PCR 验证结果与分析结果一致, hsa-miR-191-3p 与肿瘤分期间存在相关性。**结论**·ADRA1A、hsa-miR-141-5p、hsa-miR-191-3p、SFTPC、ITLN2 和 SLC644 可能是肺腺癌潜在的生物标志物, hsa-miR-191-3p 可能与肿瘤进展有关。

[关键词] 肺腺癌; 转录组分析; 生物信息学分析; 生物标志物

[DOI] 10.3969/j.issn.1674-8115.2020.12.005 **[中图分类号]** R734.2; R318.04 **[文献标志码]** A

Transcriptional identification of potential biomarkers of lung adenocarcinoma

ZHANG Wei-ran^{1,2}, LIN Xue-feng³, LI Xin², ZHANG Hao², WANG Meng², SUN Wei², HAN Xing-peng², SUN Da-qiang^{1,4}

1. Graduate School, Tianjin Medical University, Tianjin 300070, China; 2. Department of Thoracic Surgery, Tianjin Chest Hospital, Tianjin 300222, China; 3. Department of Nursing, Tianjin Medical College, Tianjin 300222, China; 4. Department of Thoracic Surgery, Tianjin Hospital of ITCWM, Nankai Hospital, Tianjin 300100, China

[Abstract] **Objective** · To identify some related molecular markers for the diagnosis and treatment of lung adenocarcinoma by transcriptome analysis. **Methods** · The differentially expressed analyses were performed to identify the differentially expressed genes (DEGs), the differentially expressed miRNAs (DEMs), and the differentially expressed circRNAs (DECs). Functional and pathway enrichment analyses were conducted for DEGs, and the targets prediction for DEMs. Regulated network of the DEGs and DEMs was constructed, and some candidates were selected. The biomarkers obtained were verified by the Cancer Genome Atlas (TCGA) database and the qRT-PCR, and the correlation between their expression levels and overall survival and tumor stage were analyzed. **Results** · Sixty-one overlaps were contained in the 3 sets of DEGs in 3 gene expression profiles, which were enriched in 32 gene ontology (GO) terms and 10 pathways. Twenty-four DEMs were identified, and 612 miRNA-target pairs were screened out that the target genes were DEGs. In the circRNA microarray, 92 DECs were obtained. ADRA1A, hsa-miR-141-5p and hsa-miR-191-3p were important nodes in the network. TCGA and qRT-PCR results were consistent with the microarray analysis results, in addition, hsa-miR-191-3p was significantly correlated with tumor stage. **Conclusion** · ADRA1A, hsa-miR-141-5p, hsa-miR-191-3p, SFTPC, ITLN2 and SLC644 might be potential biomarkers of lung adenocarcinoma, and hsa-miR-191-3p might be associated with tumor progression.

[Key words] lung adenocarcinoma; transcriptome analysis; bioinformatics analysis; potential biomarker

肺癌是世界上发病率和死亡率均排第一位的恶性肿瘤。据最新统计数据, 2020 年全世界肺癌新发患者 228 820 例, 死亡病例 135 720 例^[1]。非小细胞肺癌 (non-small-cell lung cancer, NSCLC) 约占所有肺癌病例的 85%^[2]。肺腺癌是一种常见的 NSCLC, 约占 NSCLC 的

50%^[3]。近年来, 肺腺癌机制的研究得到了广泛的开展, 研究人员发现遗传因素影响了肺腺癌的发生和发展。其中有研究^[4]发现, 肺腺癌中最常出现的是 *ALK*、*ROS1* 和 *RET* 重排, 另外还有 80 多种基因易位。此外, 还有很多常见的基因突变影响肺腺癌, 包括 *TP53* (46%)、*EGFR*

[基金项目] 天津市津南区科技计划项目 (201805006); 天津市科学技术委员会科技计划项目 (17YFZCSY00850)。

[作者简介] 张伟然 (1986—), 男, 博士生; 电子信箱: wrzhang1986@126.com。

[通信作者] 孙大强, 电子信箱: sdqmd@163.com。

[Funding Information] Science and Technology Project of Tianjin Jinnan District (201805006); Key Science and Technology Support Project of Tianjin Science and Technology Commission (17YFZCSY00850).

[Corresponding Author] SUN Da-qiang, E-mail: sdqmd@163.com.

(20%)、*HER2* (2%)、*KRAS* (1%)、*ALK* (1%)、*BRAF* (1%) 和 *PIK3CA* (1%) 等^[5]。这些发现对肺腺癌的靶向治疗具有重要意义。例如, 克唑替尼对 *ALK* 或 *ROS1* 融合的肿瘤有效, 而吉非替尼、厄洛替尼和阿法替尼可用于 *EGFR* 突变的肿瘤患者^[6]。但目前临床应用的药物很少, 肺腺癌患者的死亡率很高, 患者很难获得长期生存^[7]。因此, 肺腺癌的分子机制亟待进一步探索。

目前, 很多研究通过基因芯片数据分析揭示了一些与肺腺癌相关的基因。Tian 等^[8]鉴定出 9 个 lncRNA 和 6 个 mRNA, 它们可能是诊断或预测人类肺腺癌细胞对紫杉醇耐药性的有效标志物。Liu 等^[9]通过分析 miRNA 基因芯片表达谱, 发现 miR-136、miR-376a 和 miR-31 在小鼠肺癌中均显著过表达, 同时 miR-31 作为致癌 miRNA 在肺癌中靶向抑制特异性的肿瘤抑制因子。Zhu 等^[10]报道了 hsa_circ_0013958 (一种环状 RNA) 可以作为一种潜在的非侵入性生物标志物用于肺腺癌的早期检测和筛查。另一项微阵列图谱研究^[11]表明, hsa_circ_0007385 在 NSCLC 肿瘤的形成中起重要作用, 为 NSCLC 的治疗提供了一个潜在的靶点。本研究综合分析了肺腺癌组织的基因、miRNA 和 circRNA 表达谱, 以探索肺腺癌的潜在生物标志物, 并通过癌症基因组图谱 (the Cancer Genome Atlas, TCGA) 对关键基因进行验证。

1 资料与方法

1.1 基因芯片数据

芯片数据 GSE104854^[12]、GSE102511、GSE85716、GSE83527 和 GSE101586^[13] 从基因表达数据库 (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) 中下载。GSE104854、GSE102511 和 GSE85716 是基因表达谱, 分别包括 3、16 和 6 个肺腺癌组织及其癌旁组织, 测序平台分别是 Ion Torrent Proton (Homo sapiens)、Ion Torrent Proton (Homo sapiens) 和 Agilent-062918 OE Human lncRNA Microarray V4.0 028004。GSE83527 是一个基于 Illumina Hiseq 2000 (Homo sapiens) 平台的 miRNA 表达谱, 包含 36 个肺腺癌样本及其癌旁样本。circRNA 表达谱芯片编号为 GSE101586, 有 10 例样本, 包括 5 例正常样本及 5 例癌症样本, 芯片平台为 Agilent-069978 Arraystar Human CircRNA microarray V1。

1.2 差异表达分析

对所有的 GEO 数据进行处理后, 通过 Limma V3.34.8 函数包 (<http://www.bioconductor.org/packages/3.6/bioc/html/>

limma.html) 进行差异表达基因 (differentially expressed genes, DEGs)、差异表达 miRNAs (differentially expressed miRNAs, DEMs) 和差异表达 circRNAs (differentially expressed circRNAs, DECs) 分析。GSE104854、GSE102511 和 GSE85716 的分析参数分别为 $|\text{Log}_2\text{FC}| > 3$ 、 $|\text{Log}_2\text{FC}| > 2$ 、 $|\text{Log}_2\text{FC}| > 2$, 均 $P < 0.05$ 。进一步筛选出上述 3 组表达谱的重叠 DEGs。此外, DEMs 和 DECs 的纳入标准均为 $|\text{Log}_2\text{FC}| > 1$ 和 $P < 0.05$ 。采用 circBase (<http://www.circbase.org/>) 进一步对 DEGs 与 DECs 的关系进行分析。

1.3 功能富集分析

用注释、可视化和综合发现数据库 (David, <https://david.ncifcrf.gov/summary.jsp>, 6.7 版)、京都基因和基因组百科全书 (KEGG PATHWAY, <http://www.genome.jp/kegg>) 和 Reactome (<http://www.reactome.org>) 对重叠 DEGs 进行功能富集分析, 得到 $P < 0.05$ 的基因本体 (gene ontology, GO) 和重叠 DEGs 所涉及的重要生物过程和通路。

1.4 miRNA 靶向基因的筛选

为了探索肺腺癌样本中 DEMs 的调控机制, 利用 miRWall v3.0 数据库 (<http://129.206.7.150/>) 获得 DEMs 的分子靶标和 miRNA-基因关系对, 进一步筛选出负调控的 miRNA-基因关系对。

1.5 综合调控网络构建

利用 String V10.5 (<https://string-db.org>) 数据库获得重叠 DEGs 的高置信 (> 400 分) 的蛋白-蛋白对, 基于蛋白-蛋白对和上述负调控的 miRNA-基因关系对建立调控网络, 并用 CytoCope V3.5.1 软件 (<http://www.CytoCAPE.org/Sudio.php>) 进行可视化。

1.6 关键基因的验证

从 TCGA (<https://cancer.genome.nih.gov/>) 中下载 15 例肺腺癌患者的癌组织样本及其癌旁组织进行 DEGs 筛选。TCGA 数据筛选参照中国临床肿瘤学会肿瘤标志物专家委员会和中国肿瘤驱动基因协会的专家共识及相关文献, 由 2 位病理学教授对每例样本进行病理定量和质量控制分析。DEGs 以 $|\text{Log}_2\text{FC}| > 2$ 、 $P < 0.05$ 为阈值, 记为 DEGs-TCGA。将 DEGs-TCGA 与上述基因表达谱的 3 组 DEGs 进行比较。此外, 采用聚类算法对癌组织样本和癌旁组织进行分离, 比较关键基因表达与患者总生存期和肿瘤分期期间的相关性。

收集 8 例 2018 年 1 月—10 月在天津市胸科医院就诊的肺腺癌患者的癌组织和癌旁组织, 提取总 RNA。使用 SYBR® Premix Ex Taq™ Kit 试剂盒 (TaKaRa, 日本) 进行 *SFTPC*、*ITLN2*、*SLC6A4*、*ADRA1A* 以及 *EGFR* 的实时荧光定量 PCR (qRT-PCR) 检测, hsa-mir-141-5p 和 hsa-mir-191-3p 采用 All-in-One™ miRNA qPCR Kit 试剂盒 (GeneCopoeia Inc., 美国) 检测。所有引物均由宝日医生物技术 (北京) 有限公司设计合成, 如表 1 所示。 β -actin 和 U6 为内参。收集患者的外周血样本, 采用酶联免疫吸附实验 (ELISA) 检测血清中肿瘤标志物癌胚抗原 (CEA) 和神经元特异性烯醇化酶 (NSE) 的表达水平。

表 1 *SFTPC*、*ITLN2*、*SLC6A4*、*ADRA1A*、hsa-mir-141-5p 和 hsa-mir-191-3p 的引物序列
Tab 1 Primer sequences of *SFTPC*, *ITLN2*, *SLC6A4*, *ADRA1A*, hsa-mir-141-5p and hsa-mir-191-3p

Gene	Primer
<i>SFTPC</i>	
Forward	5'-TCAGATGCCATGCTGTATGCA-3'
Reverse	5'-GCTTTGAAGCCTGTTTGTTCCT-3'
<i>ITLN2</i>	
Forward	5'-GCAGGGCAACAAAGCAGACTA-3'
Reverse	5'-CAGGGCGCTGTTTCTCCAA-3'
<i>SLC6A4</i>	
Forward	5'-TTTGGGGAYGGAGAGGAATTAGATAAGGG-3'
Reverse	5'-AACRAAAAATCCTAACTTCTACTCTTAACTTTAC-3'
<i>ADRA1A</i>	
Forward	5'-CAGTGAGAACGGGGAGGA-3'
Reverse	5'-AGAAGGAAGTGGGGTGGG-3'
<i>EGFR</i>	
Forward	5'-ATGAGATGGAGGAAGACGG-3'
Reverse	5'-CGGCAGGATGTGGAGAT-3'
β -actin	
Forward	5'-CATCTCTTGCTCGAAGTCCA-3'
Reverse	5'-ATCATGTTTGAGACCTTCAACA-3'
hsa-mir-141-5p	
Forward	5'-CTGGGTCCATCTTCCAG-3'
Reverse	5'-GACCTTCTACCTGGGTC-3'
hsa-mir-191-3p	
Forward	5'-CACAGTCAACGGAATCCC-3'
Reverse	5'-CCAGTGCAGGGTCCGAGGTA-3'
U6	
Forward	5'-CAGTGCAGGGTCCGAGGTAT-3'
Reverse	5'-GCAGGGGCCATGGCTAATCTTCTGTATCG-3'

1.7 统计学分析

采用 SPSS 21.0 软件进行统计分析, 数据用 $\bar{x} \pm s$ 表示。采用 Student's *t* 检验进行差异分析, 采用 Pearson 相关性检验进行相关性分析, $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 DEGs、DEMs 和 DECs 筛选

在 GSE104854、GSE102511 和 GSE85716 中, 与癌旁组织相比, 肺腺癌组织分别检测到 412 个 (180 个上调和 232 个下调)、201 个 (71 个上调和 130 个下调) 和 709 个 (166 个上调和 543 个下调) DEGs, 图 1 显示了 3 组 DEGs 的 Venn 图。此外, 3 组 DEG 中包含 61 个重叠 DEGs (表 2), 其中 *P* 值较小的重叠 DEGs 包括 *SFTPC*、*SLC6A4*、*CA4*、*TMEM100* 和 *ITLN2*。这些重叠 DEGs 具有相同的调节趋势, 即同时上调或下调。

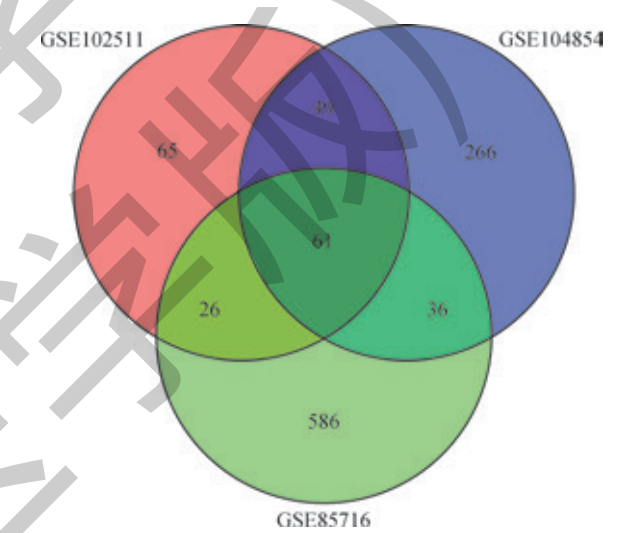


图 1 GSE104854、GSE102511 和 GSE85716 基因表达谱中 3 组 DEGs 的 Venn 图
Fig 1 Venn diagram of 3 sets of DEGs in gene expression profiles of GSE104854, GSE102511 and GSE85716

表 2 GSE104854、GSE102511 和 GSE85716 基因表达谱中的重叠 DEGs
Tab 2 Overlapped DEGs in gene expression profiles of GSE104854, GSE102511 and GSE85716

Gene	GSE104854		GSE102511		GSE85716	
	Regulation	<i>P</i> value	Regulation	<i>P</i> value	Regulation	<i>P</i> value
<i>SFTPC</i>	Down	0.002	Down	0.000	Down	0.005
<i>SLC6A4</i>	Down	0.002	Down	0.000	Down	0.000
<i>CA4</i>	Down	0.003	Down	0.000	Down	0.000
<i>TMEM100</i>	Down	0.000	Down	0.000	Down	0.000
<i>ITLN2</i>	Down	0.000	Down	0.000	Down	0.000
<i>AGER</i>	Down	0.000	Down	0.000	Down	0.000

Continued Tab

Gene	GSE104854		GSE102511		GSE85716	
	Regulation	P value	Regulation	P value	Regulation	P value
FAM107A	Down	0.000	Down	0.000	Down	0.000
SOSTDC1	Down	0.003	Down	0.000	Down	0.000
CLEC3B	Down	0.001	Down	0.000	Down	0.000
FGFBP2	Down	0.001	Down	0.000	Down	0.001
MYZAP	Down	0.000	Down	0.000	Down	0.000
NDRG4	Down	0.000	Down	0.000	Down	0.000
CLDN18	Down	0.036	Down	0.000	Down	0.000
VIPR1	Down	0.013	Down	0.000	Down	0.000
RTKN2	Down	0.000	Down	0.000	Down	0.000
CLDN5	Down	0.000	Down	0.000	Down	0.000
STXBP6	Down	0.000	Down	0.000	Down	0.000
BTNL9	Down	0.000	Down	0.000	Down	0.000
LIMS2	Down	0.000	Down	0.000	Down	0.000
EDNRB	Down	0.000	Down	0.000	Down	0.000
CCBE1	Down	0.008	Down	0.000	Down	0.000
SOX7	Down	0.001	Down	0.000	Down	0.000
EPAS1	Down	0.000	Down	0.000	Down	0.000
EDN1	Down	0.000	Down	0.000	Down	0.000
ADRA1A	Down	0.002	Down	0.000	Down	0.000
UPK3B	Down	0.000	Down	0.009	Down	0.000
SIPR1	Down	0.001	Down	0.000	Down	0.000
TEK	Down	0.000	Down	0.000	Down	0.000
RAMP2	Down	0.000	Down	0.000	Down	0.000
RAMP3	Down	0.000	Down	0.000	Down	0.000
LYVE1	Down	0.001	Down	0.000	Down	0.000
SEMA3G	Down	0.001	Down	0.000	Down	0.001
GRIA1	Down	0.006	Down	0.000	Down	0.000
ADRB1	Down	0.001	Down	0.000	Down	0.000
SGCG	Down	0.000	Down	0.000	Down	0.000
CLEC14A	Down	0.000	Down	0.000	Down	0.000
HSPA12B	Down	0.000	Down	0.000	Down	0.000
ROBO4	Down	0.000	Down	0.000	Down	0.000
HHIP	Down	0.002	Down	0.000	Down	0.011
ACVRL1	Down	0.000	Down	0.000	Down	0.000
HIGD1B	Down	0.000	Down	0.000	Down	0.000
SEMA6A	Down	0.001	Down	0.000	Down	0.000
NCKAP5	Down	0.000	Down	0.001	Down	0.000
GDF10	Down	0.001	Down	0.000	Down	0.000
DKK2	Down	0.006	Down	0.000	Down	0.001
DACH1	Down	0.003	Down	0.000	Down	0.000
EFCC1	Down	0.003	Down	0.000	Down	0.000

Continued Tab

Gene	GSE104854		GSE102511		GSE85716	
	Regulation	P value	Regulation	P value	Regulation	P value
SMAD6	Down	0.000	Down	0.000	Down	0.000
PYCR1	Up	0.000	Up	0.000	Up	0.000
KIF26B	Up	0.000	Up	0.000	Up	0.009
SPDEF	Up	0.004	Up	0.000	Up	0.000
COMP	Up	0.002	Up	0.000	Up	0.019
TOX3	Up	0.001	Up	0.000	Up	0.012
CP	Up	0.002	Up	0.000	Up	0.015
ABCC3	Up	0.000	Up	0.000	Up	0.000
FAM83A	Up	0.001	Up	0.000	Up	0.000
EEF1A2	Up	0.000	Up	0.002	Up	0.019
GLIL3	Up	0.000	Up	0.001	Up	0.001
TMPRSS4	Up	0.000	Up	0.000	Up	0.005
COL10A1	Up	0.000	Up	0.000	Up	0.002
SPP1	Up	0.000	Up	0.000	Up	0.000

在 GSE83527 中, 与癌旁组织相比, 肺腺癌组织仅发现 24 个 DEMs, 其中 11 个上调, 13 个下调 (表 3)。所有 DEMs 在 2 组间差异均有统计学意义 ($P<0.05$)。

表 3 肺腺癌组织相对于癌旁组织的 DEMs
Tab 3 DEMs in lung adenocarcinoma tissues compared with their paracancerous tissues

miRNA	logFC	AveExpr	P value
hsa-mir-451a	-3.571	11.008	0.000
hsa-mir-30a	-3.050	15.191	0.000
hsa-mir-30d	-2.316	12.873	0.000
hsa-mir-143	-1.931	16.833	0.000
hsa-mir-101	-1.891	14.337	0.000
hsa-mir-140	-1.462	10.504	0.000
hsa-mir-100	-1.416	13.178	0.000
hsa-let-7c	-1.364	10.749	0.000
hsa-mir-30b	-1.305	10.272	0.000
hsa-mir-375	-1.204	12.172	0.001
hsa-mir-30c	-1.121	9.969	0.000
hsa-mir-126	-1.048	13.079	0.000
hsa-mir-29c	-1.009	12.382	0.000
hsa-mir-93	1.175	12.174	0.000
hsa-mir-148a	1.197	15.262	0.000
hsa-mir-191	1.254	10.025	0.000
hsa-mir-200c	1.360	12.977	0.000
hsa-mir-21	1.408	16.872	0.000

Continued Tab

miRNA	logFC	AveExpr	P value
hsa-mir-1307	1.500	10.472	0.000
hsa-mir-182	1.990	13.744	0.000
hsa-mir-141	2.197	10.743	0.000
hsa-mir-183	2.274	12.408	0.000
hsa-mir-9	3.573	9.618	0.000
hsa-mir-205	5.838	11.200	0.000

Note: logFC—log (fold-change); AveExpr—average expression.

表 4 根据 *P* 值排列的前 30 个最显著的 DECs 及其位置

Tab 4 Top 30 most significant DECs according to *P* value and their position and corresponding genes

circRNA	logFC	P value	Intragenic	Position	Gene
hsa_circRNA_102231	1.402	0.000	Exonic	chr17:79813017-79813462	P4HB
hsa_circRNA_104513	-2.350	0.000	Exonic	chr7:140402665-140404763	NDUFB2
hsa_circRNA_104736	-1.174	0.000	Exonic	chr9:16435552-16437522	BNC2
hsa_circRNA_100395	-2.458	0.000	Exonic	chr1:173726114-173744981	KLHL20
hsa_circRNA_104342	-1.910	0.000	Exonic	chr7:33185853-33217203	BBS9
hsa_circRNA_104126	-2.349	0.001	Exonic	chr6:53365044-53365148	GCLC
hsa_circRNA_103749	-1.127	0.001	Exonic	chr4:151023629-151023964	DCLK2
hsa_circRNA_101282	-1.873	0.001	Exonic	chr13:95686858-95735544	ABCC4
hsa_circRNA_104200	-1.150	0.002	Exonic	chr6:137015277-137041727	MAP3K5
hsa_circRNA_103211	-1.169	0.002	Exonic	chr22:36872779-36872903	TXN2
hsa_circRNA_000963	-1.719	0.002	Intronic	chr6:136610762-136610850	BCLAF1
hsa_circRNA_102442	-2.410	0.003	Exonic	chr19:10610070-10610756	KEAP1
hsa_circRNA_000864	-1.728	0.003	Intronic	chr16:398402-398484	AXIN1
hsa_circRNA_100833	1.658	0.003	Exonic	chr11:61605249-61615756	FADS2
hsa_circRNA_103137	-2.306	0.003	Exonic	chr21:43334706-43342194	C2CD2
hsa_circRNA_104121	-1.111	0.003	Exonic	chr6:52380803-52381535	TRAM2
hsa_circRNA_101278	-1.114	0.003	Exonic	chr13:77791974-77818086	MYCBP2
hsa_circRNA_001350	-1.914	0.003	Intronic	chr10:97999787-97999925	BLNK
hsa_circRNA_101737	-1.058	0.004	Exonic	chr16:18852886-18856973	SMG1
hsa_circRNA_104099	-1.697	0.004	Exonic	chr6:35195356-35201078	SCUBE3
hsa_circRNA_100660	-1.823	0.004	Exonic	chr10:99967857-99969656	C10orf28
hsa_circRNA_102602	-1.386	0.004	Exonic	chr19:58436990-58439415	ZNF418
hsa_circRNA_101712	-1.059	0.004	Exonic	chr16:14687157-14698083	PARN
hsa_circRNA_101373	-1.699	0.004	Exonic	chr14:68156928-68159769	RDH11
hsa_circRNA_101969	-1.350	0.005	Exonic	chr17:7225183-7225329	NEURL4
hsa_circRNA_105013	-1.903	0.005	Internal	chrX:79962925-79975155	BRWD3
hsa_circRNA_101706	-1.557	0.005	Exonic	chr16:11868091-11876244	ZC3H7A
hsa_circRNA_104551	-1.512	0.005	Exonic	chr8:9437668-9437893	TNKS
hsa_circRNA_101213	-1.882	0.006	Exonic	chr12:131357380-131357465	RAN
hsa_circRNA_000046	-1.133	0.006	Intronic	chr1:40506473-40506637	CAP1

在 GSE101586 中, 肺腺癌组织和癌旁组织发现 92 个 DECs, 包括 10 个上调 DECs 和 82 个下调 DECs。进一步确定这些 DECs 的位置, 其中 70 个位于编码区的外显子区, 18 个位于编码区的内含子区, 3 个位于内部转录间隔区, 1 个位于非编码区的 5' 非翻译区 (5'UTR)。此外, 筛选出与 DECs 有关系的基因中, 在上述 3 组 DEGs 中出现过的基因只有 *LDB2* 和 *PREX1*, 而重叠 DEGs 中没有筛选出与 DECs 相关的基因。前 30 个差异最显著的 DECs 的位置和相应的基因见表 4。

2.2 功能通路富集

对 DEGs 进行功能和通路富集分析, 共得到 32 个通路, 包括 23 个生物过程、6 个细胞组分、1 个分子功能、1 个 KEGG 通路和 1 个 RACTOME 通路 (图 2)。从结果中可以看出大部分 DEGs 显著富集于系统发育、迁

移调控、心率调控、血压调控等生物学过程, 主要富集于 GPCR 及刺激神经组织的受体配体相互作用信号通路。此外, 富集到通路中最多的基因分别为 *SIPRI*、*EDN1* 和 *ACVRL1*, 分别富集在 24、24 和 23 个通路中。

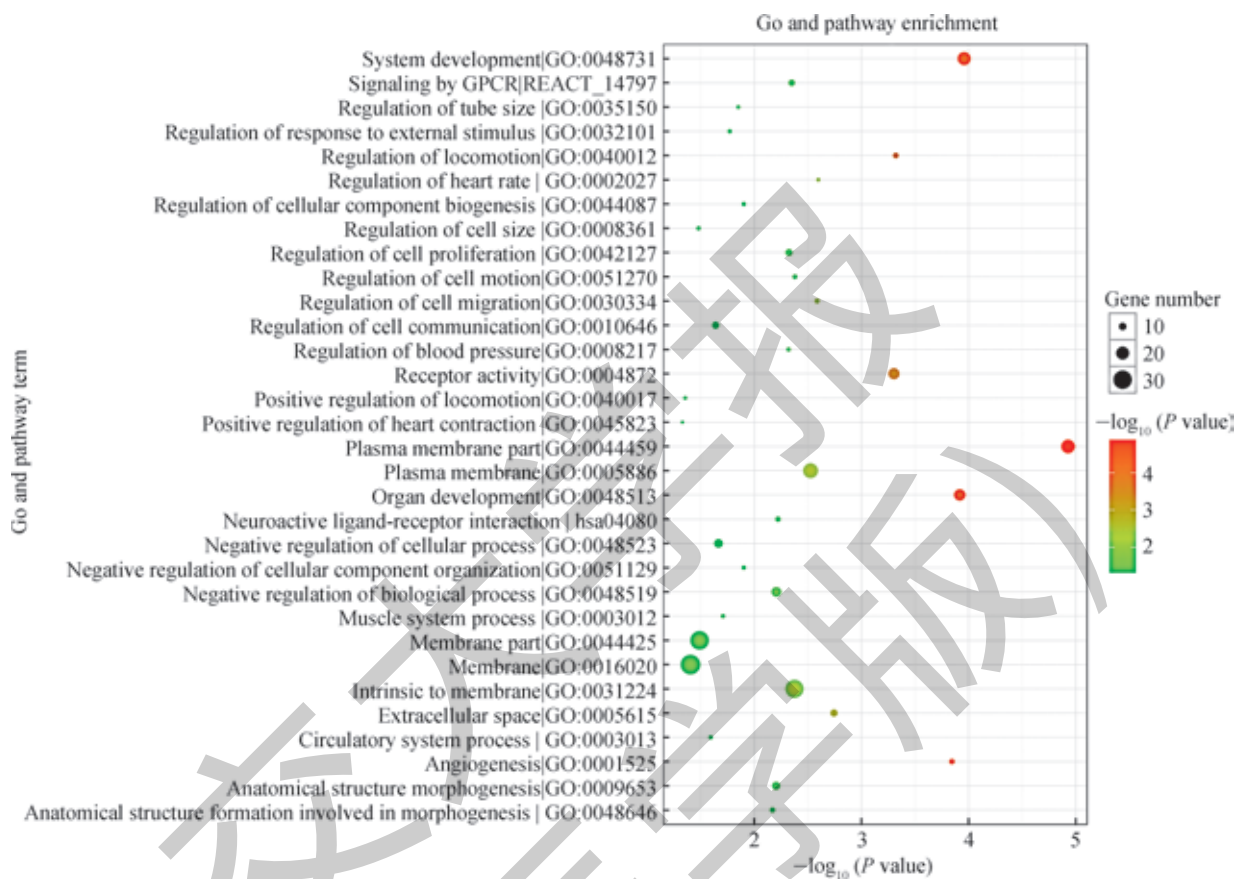


图2 重叠 DEGs 富集的 GO 和通路

Fig 2 Enriched GO and pathway terms of the overlapped DEGs

2.3 DEMs 的靶标预测

根据芯片数据得到了 24 个 DEMs, 并筛选出 DEMs 的靶基因, 仅保留为 DEGs 关系对的靶向基因。由此筛选出 612 对 miRNA- 基因关系对, 其中 303 对为负调控。

2.4 重叠 DEGs 与 DEMs 的调控网络

经过 String 数据库分析, 共鉴定出得分在 400 分以上的 19 对蛋白 - 蛋白相互作用对, 涉及 19 个重叠 DEGs, *EDNRB-EDN1*、*RAMP2-VIPRI* 和 *RAMP3-VIPRI* 是得分最高的 3 个关系对 (分别为 997、947 和 947 分)。基于 19 个蛋白 - 蛋白对和以上 303 个负调控的 miRNA- 基因对构建了调节网络 (图 3), 其中, hsa-mir-141-5p、hsa-mir-191-3p 和 *ADRA1A* 的节点度最高, 被认为是潜在的生物标志物。

2.5 DEGs 的验证

在 TCGA 数据库中共筛选出 1 175 个 DEGs, 并与癌旁组织进行比较, 其中 412 例上调, 763 例下调。这些基因中有 55 个基因 (10 个上调, 45 个下调) 与上述的重叠 DEGs (表 2) 相同并具有同样的表达趋势。其余 6 个未鉴定的基因分别为 *TOX3*、*EDN1*、*SPDEF*、*FAM83A*、*FGFBP2* 和 *COMP*, 说明本研究所鉴定的 DEGs 是可靠的。此外, *SFTPC*、*ITLN2* 和 *SLC6A4* 的表达量较高, 且在 2 组 DEGs 中均具有显著差异。因此, 选择 *SFTPC*、*ITLN2* 和 *SLC6A4* 作为潜在的标志物。

关键基因 *SFTPC*、*ITLN2*、*SLC6A4*、hsa-mir-141-5p、hsa-mir-191-3p 和 *ADRA1A* 的表达水平与患者总生存期和肿瘤分期期间的相关性见表 5。相关系数用来衡量 2 个数据集合的线性关系, 相关系数的绝对值越大, 相关性越强。

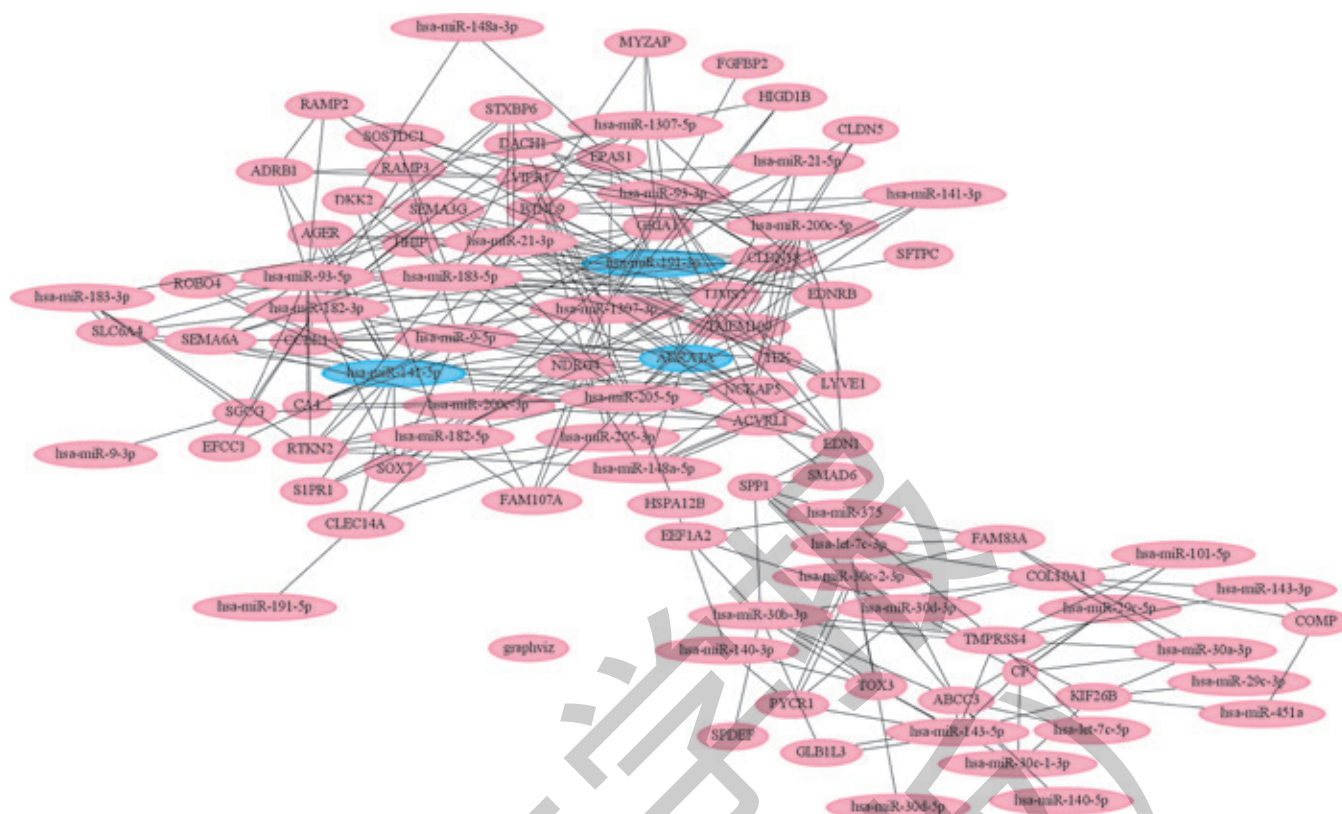


Fig 3 Regulated network of overlapped DEGs and DEMs

异,但两者间差异无统计学意义 ($P>0.05$)。EGFR 在癌组织中的表达水平高于癌旁组织,血清中 CEA 和 NSE 的表达水平分别为 5.36 $\mu\text{g/L}$ 和 14.69 U/mL。

Tab 5 Correlation of *SFTPC*, *ITLN2*, *SLC6A4*, *ADRA1A*, hsa-mir-141-5p and hsa-mir-191-3p expression with overall survival and tumor stage

Gene	Survival		Staging	
	Coefficient	P value	Coefficient	P value
<i>SFTPC</i>	0.232	0.405	−0.329	0.231
<i>ITLN2</i>	0.113	0.689	−0.314	0.254
<i>SLC6A4</i>	−0.043	0.879	−0.147	0.601
<i>ADRA1A</i>	0.184	0.512	−0.345	0.208
hsa-mir-141-5p	0.012	0.967	0.340	0.215
hsa-mir-191-3p	−0.255	0.359	0.680	0.005

Gene	Cancer tissue	Paracancer tissue
SFTPC	~0.7	~1.1
ITLN2	~1.1	~1.0
SLC6A4	~0.7	~1.1
ADRA1A	~1.0	~1.1
miR-141-5p	~2.7	~1.0
miR-191-3p	~1.6	~1.0
EGFR	~1.4	~1.0

图4 肺腺癌组织和癌旁组织中 *SFTPC*、*ITLN2*、*SLC6A4*、*ADRA1A*、*hsa-mir-141-5p*、*hsa-mir-191-3p* 和 *EGFR* 的相对表达水平

Fig 4 Relative expression levels of *SFTPC*, *ITLN2*, *SLC6A4*, *ADRA1A*, hsa-mir-141-5p, hsa-mir-191-3p and *EGFR* in lung adenocarcinoma tissues and paracancer tissues

3 讨论

GEO 数据库是由美国国立生物技术信息中心 (National Center for Biotechnology Information, NCBI) 创建并维护的基因表达数据库, 其中收录了世界各国研究机构提交的高通量基因表达数据。GEO 数据库的设计初衷是为了收集整理各种基因表达芯片数据, 后来陆续加入了甲基化芯片、lncRNA、miRNA、CNV 芯片等各种芯片, 甚至高通量测序数据。在本研究中, 从 GEO 数据库中下载了 3 个基因表达谱、1 个 miRNA 表达谱和 1 个 circRNA 表达谱, 通过差异表达分析、功能富集分析、miRNA- 基因关系对筛选以及综合调控网络构建, 最终筛选出 hsa-miR-141-5p、hsa-miR-191-3p、*ADRA1A*、*SFTPC*、*ITLN2* 和 *SLC6A4* 为肺腺癌潜在的标志物。

hsa-mir-141-5p 和 hsa-mir-191-3p 在肺癌中的作用尚不清楚, 只有少数研究揭示了它们在其他癌症中的潜在作用。Wang 等^[14] 确定了 10 个 mRNA (hsa-mir-141-5p、hsa-mir-1271-5p、hsa-mir-574-3p 等), 对人类卵巢癌组织和正常组织的敏感度和特异度分别为 97% 和 92%。Plieskatt 等^[15] 报道了 hsa-mir-141-5p 和其他 miRNA 结合与 viverrini 感染的胆管癌亚型有关。Wu 等^[16] 发现与非侵袭性垂体腺瘤相比, hsa-mir-191-3p 在侵袭性垂体腺瘤中的表达上调。*ADRA1A* 在精神分裂症患者中受到广泛关注, 它是导致更严重代谢异常的危险因素^[17]。在肿瘤中, *ADRA1A* 是一种抑癌基因; 与老年肝癌患者相比, 年轻肝癌患者中 *ADRA1A* 含量较少^[18]。此外, *ADRA1A* 在子宫内膜癌患者外周血中高表达, *ADRA1A* 的表达与国际妇产科协会 (FIGO) 分期和淋巴结转移有关。因此, 血清 *ADRA1A* 的表达可用于评估生存率, 并可能参与子宫内膜癌的发病和转移。本研究发现 hsa-mir-141-5p、hsa-mir-191-3p 和 *ADRA1A* 可能在肺腺癌中发挥一定的作用, 并且调控网络进一步提供了可能的调控对, 例如 hsa-mir-141-5p 和 hsa-mir-191-3p 可能通过靶向 *ADRA1A* 发挥作用。

表面活性蛋白 C (surfactant protein C, *SFTPC*) 在肺癌中的研究已经取得了一定的成果。它是肺表面活性物质蛋白之一, 与肺细胞的分化有关^[19]。Saintigny 等^[20] 认为 *SFTPC* 在肺腺癌组织中下调, 与本研究结果一致。2014 年的一项研究^[21] 表明, *SFTPC* 的表达与 NSCLC 较短的总生存期显著相关。*SLC6A4* 是 5-羟色胺转运体基因, 其表达或甲基化与抑郁症的发生、表型及预后密切相关^[22-23]。同时, *SLC6A4* 变异与结直肠癌患者的低生存率相关^[24]。此外, *SLC6A4* 变异是慢性阻塞性肺病合并肺癌的风险因素^[25]。然而, 并没有发现另一潜在标志物 *ITLN2* 在肺癌中

的作用。在本研究中, 与正常组织相比, *SFTPC*、*ITLN2* 和 *SLC6A4* 在肺组织中的表达有显著变化, 提示它们可能在肺腺癌中起一定作用, 但尚需实验室和临床证实, 这也是我们下一步要开展的工作。

我们也试图从 circRNA 表达水平揭示肺腺癌的相关机制, 并获得了一些 DECs。然而, 只鉴定了 3 组 DECs 中 2 个相应的基因 *LDB2* 和 *PREX1*, 而没有发现重叠的 DECs。因此, 推测 circRNAs 的表达可能具时空特异性, 可能与肺腺癌无关, 具体原因尚需进一步研究。

在本研究中, 对于从 GEO 数据库中发现的重叠 DEGs, 通过 TCGA 数据库进行了验证。TCGA 数据库是由美国国家癌症研究所和美国人类基因组研究所联合启动的一个项目。作为目前最大的癌症基因信息数据库, TCGA 收录了各种人类癌症的临床数据、基因组变异、mRNA 和 miRNA 表达以及甲基化等数据, 是肿瘤研究中资源最丰富、数据最权威的数据库^[26]。本研究得到的 61 个重叠 DEGs 中, 大多数都在 TCGA 数据库中得到验证, 因此被认为是准确和可信的。只有 6 个基因, 即 *TOX3*、*EDN1*、*SPDEF*、*FAM83A*、*FGFBP2* 和 *COMP* 未被证实, 它们在调控网络中没有起到关键作用。此外, 我们对 TCGA 数据中 *SFTPC*、*ITLN2*、*SLC6A4*、hsa-mir-141-5p、hsa-mir-191-3p 和 *ADRA1A* 的表达水平与患者总生存期和肿瘤分期进行了相关性研究, 其中 hsa-mir-191-3p 与患者的分期间存在显著相关性。这种相关性很可能反映了 hsa-mir-191-3p 对肿瘤进展和细胞迁移的影响, 有必要进行更多的研究, 以进一步阐释 hsa-mir-191-3p 作为肿瘤分期标志物的有效性及其在肿瘤细胞行为中的作用。其他基因的表达水平与患者的总生存期和肿瘤分期间均无相关性, 但在癌组织和癌旁组织中的表达存在显著差异, 说明这些基因在肺腺癌中的差异表达与患者的发病情况相关, 与生存和预后无关。因此, 这些基因在肺腺癌的临床诊断和治疗中可能具有潜在的应用价值。

在对选定的分子标志物 *SFTPC*、*ITLN2*、*SLC6A4*、hsa-mir-141-5p、hsa-mir-191-3p 和 *ADRA1A* 进行验证时, 癌组织中 *SFTPC* 和 *SLC6A4* 的表达水平明显高于癌旁组织, hsa-mir-141-5p 和 hsa-mir-191-3p 的表达水平明显降低, 这些结果与上述生物信息学分析结果一致。而 *ITLN2* 和 *ADRA1A* 在癌组织中的表达水平虽然低于癌旁组织, 但两者间差异无统计学意义。芯片中的样本来自各个国家和地区, 包括白种人、亚洲人和非裔人群等, 而本研究收集的病例大部分为本市患者, 这种个体差异可能导致不同人群中基因表达水平的差异。因此, 后续需要进一步进行临床试验, 对所选的标志物进行验证。此外, 癌组

织中 *EGFR* 的表达水平明显高于癌旁组织, 这与以往的研究结果一致^[27]。*EGFR* 是表皮生长因子受体家族成员之一, *EGFR* 信号通路对细胞的生长、增殖和分化等生理过程发挥重要的作用, *EGFR* 阳性肺癌患者的生存期明显短于 *EGFR* 阴性患者^[28]。*CEA* 和 *NSE* 是肺癌中常用的肿瘤标志物, 在正常人体中的水平分别低于 5.0 $\mu\text{g/L}$ 和 12.5 U/mL , 在临床中经常可以发现肺腺癌患者血清 *CEA* 及 *NSE* 水平升高^[29]。本研究中检测结果显示 *CEA* 和 *NSE* 在肺腺癌患者中的表达水平分别为 5.36 $\mu\text{g/L}$ 和 14.69 U/mL , 这

与以往的研究结果一致。

在本研究中, 转录组测序和分析应用于识别重要的 miRNA 和与转录组测序相关的基因。*ADRA1A*、*hsa-miR-141-5p*、*hsa-miR-191-3p*、*SFTPC*、*ITLN2* 和 *SLC6A4* 可能是肺腺癌的潜在标志物。此外, 以上 6 个基因组合在临床检测中可能具有更高的准确度、敏感度及特异度, 为肺腺癌的早期诊断和靶向治疗提供了理论依据和潜在的研究方向。然而, 这些生物标志物的筛选仅仅是通过生物信息学分析, 还需要更多的临床试验和更直接的证据进行验证。

参 · 考 · 文 · 献

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020[J]. CA Cancer J Clin, 2020, 70(1): 7-30.
- [2] Pitroda SP, Stack ME, Liu GF, et al. JAK2 inhibitor SAR302503 abrogates PD-L1 expression and targets therapy-resistant non-small cell lung cancers[J]. Mol Cancer Ther, 2018, 17(4): 732-739.
- [3] Denisenko TV, Budkevich IN, Zhivotovsky B. Cell death-based treatment of lung adenocarcinoma[J]. Cell Death Dis, 2018, 9(2): 117.
- [4] Huret JL. Lung: translocations in adenocarcinoma[J]. Atlas Genet Cytogenet Oncol Haematol, 2015. DOI: 10.4267/2042/56443.
- [5] McGuire S. World cancer report 2014. Geneva, Switzerland: world health organization, international agency for research on cancer, WHO press, 2015[J]. Adv Nutr, 2016, 7(2): 418-419.
- [6] Ettinger DS, Wood DE, Aisner DL, et al. Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology[J]. J Natl Compr Canc Netw, 2017, 15(4): 504-535.
- [7] Huang WW, Zhu WZ, Mu DL, et al. Perioperative management may improve long-term survival in patients after lung cancer surgery: a retrospective cohort study[J]. Anesth Analg, 2018, 126(5): 1666-1674.
- [8] Tian X, Zhang HY, Zhang BJ, et al. Microarray expression profile of long non-coding RNAs in paclitaxel-resistant human lung adenocarcinoma cells[J]. Oncol Rep, 2017, 38(1): 293-300.
- [9] Liu X, Sempere LF, Ouyang HX, et al. MicroRNA-31 functions as an oncogenic microRNA in mouse and human lung cancer cells by repressing specific tumor suppressors[J]. J Clin Invest, 2010, 120(4): 1298-1309.
- [10] Zhu XL, Wang XY, Wei SZ, et al. Hsa_circ_0013958: a circular RNA and potential novel biomarker for lung adenocarcinoma[J]. FEBS J, 2017, 284(14): 2170-2182.
- [11] Jiang MM, Mai ZT, Wan SZ, et al. Microarray profiles reveal that circular RNA hsa_circ_0007385 functions as an oncogene in non-small cell lung cancer tumorigenesis[J]. J Cancer Res Clin Oncol, 2018, 144(4): 667-674.
- [12] Ding XX, Zhang S, Li X, et al. Profiling expression of coding genes, long noncoding RNA, and circular RNA in lung adenocarcinoma by ribosomal RNA-depleted RNA sequencing[J]. FEBS Open Bio, 2018, 8(4): 544-555.
- [13] Zong L, Sun QC, Zhang HP, et al. Increased expression of circRNA_102231 in lung cancer and its clinical significance[J]. Biomedicine Pharmacother, 2018, 102: 639-644.
- [14] Wang L, Zhu MJ, Ren AM, et al. A ten-microRNA signature identified from a genome-wide microRNA expression profiling in human epithelial ovarian cancer[J]. PLoS One, 2014, 9(5): e96472. DOI: 10.1371/journal.pone.0096472.
- [15] Plieskatt JL, Rinaldi G, Feng YJ, et al. Distinct miRNA signatures associate with subtypes of cholangiocarcinoma from infection with the tumorigenic liver fluke *Opisthorchis viverrini*[J]. J Hepatol, 2014, 61(4): 850-858.
- [16] Wu SL, Gu Y, Huang YY, et al. Novel biomarkers for non-functioning invasive pituitary adenomas were identified by using analysis of microRNAs expression profile[J]. Biochem Genet, 2017, 55(3): 253-267.
- [17] Cheng C, Chiu HJ, Loh EI-W, et al. Association of the ADRA1A gene and the severity of metabolic abnormalities in patients with schizophrenia[J]. Prog Neuropsychopharmacol Biol Psychiatry, 2012, 36(1): 205-210.
- [18] Wang HW, Hsieh TH, Huang SY, et al. Forfeited hepatogenesis program and increased embryonic stem cell traits in young hepatocellular carcinoma (HCC) comparing to elderly HCC[J]. BMC Genom, 2013, 14(1): 736.
- [19] Setoguchi Y, Tamaoki F, Nakai M, et al. The frequency of mutations in the surfactant protein C gene (SFTPC) is much higher in pediatric interstitial lung disease (PILD) than either adult-onset familial or sporadic interstitial pneumonia (IP)[C]. San Diego: American Thoracic Society 2009 International Conference, 2009.
- [20] Saintigny P, Massarelli E, Lin S, et al. CXCR2 expression in tumor cells is a poor prognostic factor and promotes invasion and metastasis in lung adenocarcinoma[J]. Cancer Res, 2013, 73(2): 571-582.
- [21] Lindskog C, Fagerberg L, Hallström B, et al. The lung-specific proteome defined by integration of transcriptomics and antibody-based profiling[J]. FASEB J, 2014, 28(12): 5184-5196.
- [22] Dong C, Wong ML, Licinio J. Sequence variations of ABCB1, SLC6A2, SLC6A3, SLC6A4, CREB1, CRHR1 and NTRK2: association with major depression and antidepressant response in Mexican-Americans[J]. Mol Psychiatry, 2009, 14(12): 1105-1118.
- [23] Philibert RA, Sandhu H, Hollenbeck N, et al. The relationship of 5HTT (SLC6A4) methylation and genotype on mRNA expression and liability to major depression and alcohol dependence in subjects from the Iowa Adoption Studies[J]. Am J Med Genet B Neuropsychiatr Genet, 2008, 147B(5): 543-549.
- [24] Savas S, Hyde A, Stuckless SN, et al. Serotonin transporter gene (SLC6A4) variations are associated with poor survival in colorectal cancer patients[J]. PLoS One, 2012, 7(7): e38953. DOI: 10.1371/journal.pone.0038953.
- [25] de Andrade M, Li Y, Marks RS, et al. Genetic variants associated with the risk of chronic obstructive pulmonary disease with and without lung cancer[J]. Cancer Prev Res (Phila), 2012, 5(3): 365-373.
- [26] 谢龙祥, 闫中文, 党艺方, 等. TCGA 数据库: 海量癌症数据的源泉 [J]. 河南大学学报 (医学版), 2018, 37(3): 223-228.
- [27] 赵敏, 田峰, 邸鑫, 等. RBM5、EGFR 和 KRAS 在非小细胞肺癌中的表达及其意义 [J]. 中国实验诊断学, 2019, 23(2): 303-308.
- [28] Arfaoui A, Kriaa L, Znaïdi N, et al. Over-expression of EGFR is closely correlated to poor prognosis in Tunisian patients with non-small cell lung adenocarcinoma[J]. J Immunoassay Immunochem, 2014, 35(3): 256-268.
- [29] Peng Y, Wang Y, Hao XZ, et al. Utility of multiple increased lung cancer tumor markers in treatment of patients with advanced lung adenocarcinoma[J]. Chin J Lung Cancer, 2017, 20(10): 690-694.

[收稿日期] 2020-01-19

[本文编辑] 吴 洋

