

基于加权基因共表达网络分析 (WGCNA) 探讨龙葵抗肺腺癌功能基因模块及生物标记识别研究

刘燕玲¹, 吴美玲¹, 胡莹¹, 王旭景¹, 陈卓², 张爱琴^{2*}

1. 金华市中医医院, 浙江 金华 321017

2. 浙江中医药大学附属肿瘤医院, 浙江 杭州 310022

摘要: **目的** 采用加权基因共表达网络分析 (WGCNA) 法探讨龙葵抗肺腺癌的潜在生物靶标。**方法** 基于中药系统药理学技术平台 (TCMSP) 及文献检索, 以口服利用度 (OB)、类药性 (DL) 构建龙葵活性成分数据库, 采用 DRAR-CPI 服务器反向模拟分子-靶蛋白对接龙葵预测成分可能的作用靶标, 并结合 WGCNA 挖掘在美国国立生物技术信息中心 (NCBI) 的 GEO 数据库中的 GSE10072 数据集得到共表达基因模块, 并与龙葵预测靶标匹配映射, 得到龙葵潜在抗肺腺癌靶点。将预测靶标与抗肺腺癌基因分别利用生物学信息注释数据库 (Metascape) 进行 GO 生物学过程富集分析和 KEGG 通路富集分析, 并用 STRING 数据库结合 Cytoscape 软件可视化龙葵潜在抗肺腺癌靶点蛋白相互作用网络并进行网络拓扑学分析, 同时构建龙葵活性成分-抗癌靶点-通路网络, 探讨龙葵抗癌作用的机制。并通过 UALCAN 及 Kaplan Meier plotter 数据库分析关键基因在肺腺癌组织中转录水平的变化, 通过 KM plotter 分析关键基因与肺腺癌患者预后的关系。**结果** 共收集龙葵活性成分 9 个, 包括皮树脂醇、谷甾醇、薯蓣皂苷元、辣茄碱、槲皮素、 α -茄碱、澳洲茄碱、澳洲茄边碱、澳洲茄胺, 预测成分作用靶标 271 个, 筛选龙葵潜在抗癌靶点 41 个, 其潜在调控通路包括癌症通路、PI3K-Akt 信号通路、化学物致癌作用、癌症的中心碳代谢等通路。由蛋白互作网络分析可得关键基因为 EGFR、CASP8、HPGDS、FYN, 且证实高表达的 EGFR 和 CASP8、低表达的 HPGDS 及 FYN 与肺腺癌患者不良预后紧密相关。**结论** 龙葵抗肺腺癌具有多成分、多靶点、多途径作用的特点, 为其抗癌物质基础及作用机制的阐明提供了科学依据。

关键词: 加权基因共表达网络分析; 龙葵; 肺腺癌; GEO 数据库; 作用机制

中图分类号: R285.5 **文献标志码:** A **文章编号:** 0253-2670(2019)24-6073-11

DOI: 10.7501/j.issn.0253-2670.2019.24.023

Identification of possible molecular markers and functional gene modules associated with lung adenocarcinoma based on weighted co-expression network analysis method

LIU Yan-ling¹, WU Mei-ling¹, HU Ying¹, WANG Xu-jing¹, CHEN Zhuo², ZHANG Ai-qin²

1. Jinhua Hospital of TCM, Jinhua 321017, China

2. The affiliated Zhejiang Cancer Hospital of Zhejiang Chinese Medical University, Hangzhou 310022, China

Abstract: **Objective** The potential biological targets for anti-lung adenocarcinoma of *Solanum nigrum* were scored using the weighted co-expression network analysis (WGCNA) method. **Methods** A database of chemical components of *S. nigrum* was established through oral bioavailability (OB), drug-likeness (DL) based on Traditional Chinese Medicine Systems Pharmacology (TCMSP) and literature retrieval. The targets of active ingredients of *S. nigrum* were predicted based on reverse docking with DRAR-CPI server, and combined with WGCNA to mine GSE10072 dataset in Gene Expression Omnibus (GEO) database to obtain coexpression gene module. Furthermore, the potential anti-lung adenocarcinoma targets of *S. nigrum* were confirmed under intersected with predicted targets and coexpression genes. The GO terms of biological processes and KEGG pathway enrichment analysis of predicted targets and anti-lung adenocarcinoma targets were performed by Metascape database, respectively. Using the targets-pathways networks to study the mechanisms of *S. nigrum* in the fight against cancer. The transcriptional level expression of key

收稿日期: 2019-06-03

基金项目: 浙江省自然科学基金项目 (LY18H290002); 浙江省中医药管理局基金重点项目 (2019ZZ002)

作者简介: 刘燕玲, 女, 本科, 研究方向为中药临床药理。Tel: (0579)82136640 E-mail: Liuy10007@163.com

*通信作者 张爱琴, 女, 博士, 研究方向为中医药防治肿瘤。Tel: (0571)88122245 E-mail: zhanghaojianbb@163.com

String database combined with Cytoscape software to draw the proteins-proteins interactions (PPI), and active ingredients-targets-pathways networks to study the mechanisms of *S. nigrum* in the fight against cancer. The transcriptional level expression of key genes in lung adenocarcinoma cancer tissues and normal lung tissues was assessed based on UALCAN dataset. And the correlation between key genes and prognosis of lung cancer patients was calculated by KM plotter analysis. **Results** This study collected nine active components of *S. nigrum*, including medioresinol, sitosterol, diosgenin, solanocapsine, quercetin, α -chaconine, solasonin, solamargine, and solasodine. Totally 271 targets were predicted, and 41 potential anticancer targets were confirmed. The potential regulatory pathways included pathway in cancer, PI3K-Akt signaling pathway, chemical carcinogenesis, central carbon metabolism in cancer and so on. From the PPI network, we found that hub genes EGFR, CASP8, HPGDS, FYN, and high expression of EGFR and CASP8 were related to the poor overall survival in patient with lung adenocarcinoma. Oncontrary, lower expression of HPGDS and FYN were also associated with poor overall survival. **Conclusion** This study reflects the multi-component, multi-target and multi-pathway features of *S. nigrum*, and provides a scientific basis for anticancer substance and elucidating the mechanisms of action of *S. nigrum*, as well as a reference for the study of mechanisms.

Key words: weighted co-expression network analysis; *Solanum nigrum* L.; lung adenocarcinoma; GEO database; mechanisms

肺癌是目前最为常见的恶性肿瘤之一，其发病率呈现逐渐上升趋势。我国每年有超过 70 万例新的肺癌病例被确诊和近 60 万例患者死亡^[1]，位居恶性肿瘤发病和死亡率的首位，严重威胁人类的健康与生命。肺癌早期症状隐匿，不易察觉，大多患者确诊时已属晚期，丧失最佳治疗时机，其主要来源于气管、支气管黏膜或腺体的病变，非小细胞肺癌在肺癌患者中约占 85%，最常见的病理类型为腺癌^[2-3]。吸烟是肺癌发生、发展的重要作用因素，导致吸烟者发生 C→A 碱基颠换，并且引起表观遗传改变，使正常细胞癌变^[4-5]。然而吸烟导致细胞癌变进展为肿瘤的形成，其潜在的治疗分子靶点及机制仍需进一步阐明。并且对于肺腺癌的治疗，包括手术、放疗和化疗在内的传统治疗策略未能得到令人满意的显著生存效益，患者整体生存率较低^[6]。因此，进一步发现具有抗癌效应的中药活性物质及其潜在的分子生物标记物仍然具有重要的临床意义，以期肺癌早期诊断及治疗提供物质基础与分子靶点信息。

《圣济总录》曰：“肺积息贲气胀满咳嗽，涕唾脓血”。肺癌属中医学“肺积”“息贲”“积聚”“癌病”等范畴。其病因病机为正气盛衰和邪毒入侵。肺为娇脏，吸之则满，呼之则虚，为脏腑之华盖、百脉之朝会，病邪易从口鼻而入，肺气肃降失司，郁滞不宣，导致血瘀不行、痰凝、毒聚，日久形成肿块，因虚而病^[7]。中药具有低毒、扶正祛邪、多成分-多靶点以整体观调节机体运行的特点，并且联合化疗药物可发挥增效减毒的作用，为肺癌治疗增加多向选择。龙葵 *Solanum nigrum* L. 为双子叶茄科植物，又名苦葵、老鸦眼、天茄子、苦菜、野葡萄。《本草纲目》记载其“苦微甘、滑、寒、无毒”，具

有清热解毒、活血消肿、散结、利水通淋的功效^[8-10]。龙葵主要成分为生物碱，现代研究表明其对乳腺癌、肝癌、胰腺癌、黑色素瘤、脑胶质瘤、骨髓瘤等具有抑制细胞增殖、诱导细胞凋亡、阻滞细胞周期的作用^[9]。但是，龙葵成分复杂多样，其潜在作用靶点及机制不清晰，有待进一步开发研究。

加权基因共表达网络分析 (weighted correlation network analysis, WGCNA) 是一种能够从高通量数据中挖掘模块 (module) 信息，描述不同样品之间基因关联模式的系统生物学方法，如果某些基因在一个生理过程或不同组织中具有类似的表达变化模式，那么这些基因在一定程度上与功能是相关的，并将其定义为 1 个模块^[11]，根据模块的内连性和模块与表型之间的关联鉴定候补生物标记物或治疗靶标。邝浩发等^[12]应用 WGCNA 挖掘不同前列腺特异抗原 (PSA) 水平下的前列腺癌发展枢纽基因，发现 Snai2、trim29、lamb3、cyp3a5 和 slc14a1 基因很可能影响不同 PSA 水平下前列腺癌的发展。Qi 等^[13]通过 WGCNA 分析发现 9 个长链非编码 RNA (UCA1、LINC00857、PVT1、PCAT6、LINC00460、LINC00319、AP000553.1、AP000439.2、AP005233.2) 可能与非小细胞肺癌通路相关，为促进对人类癌症的理解提供了有价值的理论信息。An 等^[14]利用 WGCNA 技术筛选肺腺癌关键核心基因，结果提示 miR-21 靶向调节核心靶标 KIBRA，并通过靶向抑制 KIBRA 及 Hippo 信号通路，表达上调的 miR-21 增强腺癌的进程。由此可见，WGCNA 广泛应用于疾病标记物或基因靶点筛选当中，适用于中药活性成分结合高通量数据阐述疾病潜在作用基因群及多成分-多靶点-多通路的作用机制。

本研究基于中药系统药理学技术平台 (TCMSP), 采用吸收、分布、代谢、排泄 (ADME) 计算方法并结合文献检索收集龙葵多个活性成分, 通过 DRAR-CPI 服务器对活性成分进行模拟分子-靶蛋白对接, 筛选靶点; 对预测靶点构建蛋白互作、成分-靶点-通路网络以及生物信息注释; 进一步采用 WGCNA 挖掘从 NCBI 中 GEO 数据库下载的数据集获得与吸烟史相关的基因模块, 并与成分预测靶点相交, 得到龙葵成分核心靶点, 并进一步运用 UALCAN 及 Kaplan Meier Plotter 数据库初步探讨核心靶点在肺腺癌中的作用, 以期为进一步深入研究龙葵治疗肺癌提供参考。

1 材料与方法

1.1 龙葵活性成分的筛选与预处理

通过检索 TCMSP 数据库, 在 ADME 模块中设置口服生物利用度 (OB) $\geq 30\%$ 及类药性 (DL) ≥ 0.18 为筛选条件, 并检索查阅有关龙葵药效学文献, 收集符合条件的化合物。从 PubChem 数据库中获取相关成分 SDF 格式化学结构, 所有化学结构均采用 ChemBio 3D Ultra 软件储存为 Mol2 格式。进一步采用 Open Babel GUI 格式转换器将 Mol2 格式文件转换为 SMILES 格式, 以建立分子库备用。

1.2 分子对接及成分靶点筛选

将上述 SMILES 格式文件上传至 DRAR-CPI 服务器进行模拟分子-靶蛋白对接。DRAR-CPI 服务器利用双向 Z 转换功能将对接分数转换为 Z'-score, 从而提高精度, Z'-score 越小, 表示成分与预测靶点蛋白结合越牢固^[15]。当 Z'-score < -0.5 时化合物与预测靶点具有潜在的结合活性, 因此将 Z'-score < -0.5 的靶点确定为龙葵潜在靶点, 并由 UniProt 数据库将其蛋白靶点 PDB ID 转换为官方 UniProt ID, 整理归纳基因信息。

1.3 成分预测靶点功能富集分析

基因本体 (GO) 及 KEGG 通路富集广泛应用于高通量基因数据功能分析。Metascape 数据库为研究人员提供了一套全面的功能注释工具, 以了解大量基因背后的生物学意义。本研究采用 Metascape 数据库对龙葵预测靶点进行 GO 富集注释分析, 其中包含生物过程 (biological process) 以及 KEGG 通路富集分析。

1.4 临床数据来源

为了进一步探讨龙葵预测靶点的临床意义, 在 Pubmed 的 GEO DataSets 中以 “lung cancer” 为关

键词检索, 选择病理对照样本大于 15 例的数据集。本研究选用 GSE10072, 其发表于 2008 年 1 月 4 日, 最后数据更新时间为 2018 年 8 月 10 日。GSE10072 数据集包含样本为 107 个, 其中 58 例肿瘤样本和 49 例正常对照非肿瘤样本, 采用 Affymetrix Human Genome U133A Array 芯片检测全基因组基因。本研究应用 GEO 数据库中的 GEO2R 分析工具筛选数据集的差异基因。芯片归一化的原始强度数据由箱线图可视化, 用以考察样本数据之间是否具有可比性。样本临床信息包含患者性别、年龄、吸烟史及肿瘤临床分期。

1.5 WGCNA 共表达模块与临床信息模块基因相关的成分核心靶点确定

采用 R 软件中的 WGCNA 包对 GSE10072 所有基因间的相关模式进行了研究, 获得与临床信息相关的基因模块。计算各基因间 Pearson 相关系数, 构建加权邻接矩阵 $a_{mn} = |C_{mn}|^\beta$, 其中 a_{mn} 代表基因 m 与基因 n 的邻接系数; C_{mn} 代表基因 m 与基因 n 的皮尔逊相关系数, 通过选择软阈值 $\beta=9$ 确保无尺度网络的构建, 设定 Deepsplit 值为 2 和最小模块尺寸为 30 个基因, 进行模块识别从而生成树状图。通过聚类识别高度相似的模块, 并以高度剪切临界值为 0.25 合并相似的模块。此外, 通过皮尔逊相关性分析计算基因模块的特征基因与外部各临床信息的相关系数及 P 值, 并通过 labeled heatmap 函数以热图形式生动展示。以相关性最高的基因模块与待研究的临床特征为研究对象, 从而寻找有生物学意义的基因模块。

1.6 共同基因靶点蛋白-蛋白互作网络构建及 Hub 基因识别

通过韦恩图进行匹配映射识别成分预测靶点与吸烟史正负相关枢纽模块的共同基因, 即得龙葵活性成分潜在的抗肺腺癌作用靶点。STRING 数据库 ([https://string-db.org/Version 10.5](https://string-db.org/Version%2010.5)) 是用于预测蛋白质-蛋白质相互作用 (protein-protein interaction, PPI) 的生物学数据库, 可用于检索经实验验证或基于文献预测的蛋白/基因相互作用。为了进一步探讨龙葵抗癌共同靶点基因可能的协同作用, 本研究应用 STRING 数据库构建靶点 PPI 网络。在数据库中将蛋白种类设置为 “Homo sapiens” (人类) 进行操作, 最低相互作用阈值设为高置信度 “high confidence ≥ 0.4 ”, 其余参数均保持默认。进一步将获得的 PPI 数据导入 Cytoscape (Version 3.6.1, <http://www.cytoscape.org>)。

cytoscape.org) 进行可视化分析, 并应用 Network Analyzer 分析模块进行网络拓扑结构计算, 依据节点度 (degree) 值识别网络 Hub 基因, 其值越大, 表明与该节点相连边越多, 在维持网络整体结构上可能占据重要地位, 并再次采用 Metascape 数据库对龙葵抗癌作用靶点进行生物信息学分析。

1.7 龙葵活性成分-抗癌靶点-通路网络的构建

应用 Cytoscape 软件构建活性成分-抗肺腺癌靶点基因-基因通路”关系网络。网络中节点 (node) 代表龙葵、药物活性成分、关键靶点基因、基因富集通路; 网络中边 (edge) 用来连接龙葵与活性成分、活性成分与关键靶点基因、关键靶点基因与富集通路。整个网络展示龙葵-活性成分-靶点-通路之间的联系, 通过构建这一网络探析龙葵抗癌的作用机制。

1.8 成分核心靶点表达量及临床预后价值的评估

UALCAN 数据库 (<http://ualcan.path.uab.edu/index.html>) 是用于分析癌症转录组数据线上信息门户, 有利于用户识别生物标记物或在计算机上验证感兴趣的潜在基因并提供基于基因表达和患者生存信息的高质量图表^[16]。Kaplan Meier plotter ([\[kmplot.com/analysis/index.php?p=background\]\(http://kmplot.com/analysis/index.php?p=background\)\) 有效地利用 18 674 个癌症样本, 其中包括 5 143 例乳腺癌、1 816 例卵巢癌、2 467 例肺癌、364 例肝癌、1 065 例胃癌患者, 评估了 154 675 个基因相关的无复发和整体生存效益, 广泛用于评估潜在靶点的预后价值。本研究应用 UALCAN 及 Kaplan Meier plotter 数据库分别分析成分核心靶点在肺腺癌中相对正常癌旁组织的表达量及其对患者整体生存率的影响。](http://</p>
</div>
<div data-bbox=)

2 结果

2.1 龙葵活性成分的筛选

龙葵主要成分包括糖苷生物碱、糖蛋白、多糖以及多酚等成分, 其中以生物碱为主要活性成分。本研究基于 TCMSP 数据库通过 $OB \geq 30\%$ 和 $DL \geq 0.18$ 筛选条件获得皮树脂醇、谷甾醇、薯蓣皂苷元、辣椒碱和槲皮素 5 个化合物, 进一步查阅文献筛选龙葵生物碱类中药化学成分以 α -茄碱、澳洲茄碱、澳洲茄边碱、澳洲茄胺作为候选研究对象, 并在 TCMSP 数据库查询相应 OB 及 DL 属性 (表 1)。虽生物碱类成分未完全满足上述筛选条件, 鉴于其具有广泛的药理作用, 故仍然纳入本研究范畴。

表 1 龙葵 9 个化学活性成分信息

Table 1 Detail information of nine compounds of *Solanum nigrum*

编号	化合物	CAS 号	分子 ID	OB/%	DL
1	α -茄碱	20562-02-1	MOL008735	21.63	0.05
2	澳洲茄碱	19121-58-5	MOL006853	25.94	0.06
3	澳洲茄边碱	20311-51-7	MOL006855	31.36	0.06
4	澳洲茄胺	107484-55-9	MOL007351	11.72	0.76
5	皮树脂醇	40957-99-1	MOL002058	57.20	0.62
6	谷甾醇	83-46-5	MOL000359	36.91	0.75
7	薯蓣皂苷元	512-04-9	MOL000546	80.88	0.81
8	辣椒碱	639-86-1	MOL007356	52.94	0.67
9	槲皮素	74893-81-5	MOL000098	46.43	0.28

2.2 龙葵活性成分潜在靶点预测

依据 Z' -score < -0.5 的标准, 从 DRAR-CPI 分子-靶蛋白对接中可得知, α -茄碱可能与 PDPK1、CSNK1G2、TK1 结合度较高; 澳洲茄碱可能与 C1S、ST14、SENPI、APRT、CRP 结合度较高; 澳洲茄边碱可能与 CD2、PAK6、PADI4、TXN、CA2、CLC、ALB、MGMT、CA3、BLVRB、MAPK1、AKR1B1、GLA、NTRK3、HINT1、NEU2 结合度较高; 澳洲茄胺可能与 HSD11B1、CSK、MTHFD1、CBR3、G6PD、GSTP1 等 133 个靶点结合度较高; 皮树脂醇可能与 CDK2、SYK、SDS、HINT1、PLA2G2A

等 101 个靶点结合度较高; 谷甾醇与 HSD11B1、ASS1、PAK5、PREP、HSD17B1 等 108 个靶点结合度较高; 薯蓣皂苷元与 ADH1A、F13A1、SORD、NQO2、G6PD 等 149 个靶点结合度较高; 辣椒碱与 CSK、G6PD、GSK3B、CTSB、SORD 等 142 个靶点结合度较高; 槲皮素与 HINT1、CSNK2B、CSNK2A1、HSD17B10、ANXA3 等 100 个靶点具有潜在结合的可能性。上述各成分靶点合并去重后最终获得成分预测靶点为 271 个。

2.3 龙葵预测靶点功能富集分析

将龙葵 271 个靶基因导入 Metascape 数据库分

别在 GO 生物过程及 KEGG 信号通路中富集。根据 P 及 $Q < 0.05$ 筛选出 249 条 GO 生物过程, 并将其归纳为 16 种类型, 其中 small molecule biosynthetic process、drug metabolic process、response to toxic substance 的 P 及 Q 值最低, 涉及基因数也较多。其余 GO 生物过程见表 2。肺腺癌的发生发展涉及多方面的生物过程, 如细胞增殖凋亡的失衡、毒物反应、氧化还原、药物代谢异常蛋白自磷酸化等, 龙葵可能通过调节上述多个生物过程干预癌症的进程。由表 3 可知, 预测靶点与 7 条信号通路密切相关, 分别为流体剪切应力与动脉粥样硬化、化学物致癌作用、细胞色素 P450 对异物代谢的影响, 药物代谢-细胞色素氧化酶 P450、铂耐药、谷胱甘肽代谢和类固醇激素生物合成信号通路。由此推测, 龙葵可能主要通过调节以上信号通路发挥抗癌作用。

2.4 加权共表达网络构建及枢纽模块的确定

如图 1 所示, 共 58 例肿瘤样本和 49 例正常对照非肿瘤样本芯片原始信号强度数据归一化后, 中

位数值相对处于同一水平线, 提示所选择样品较为合理, 具有可比较性。由 GEO2R 分析获得 22 260 个基因用于构建共表达网络分析。本研究选择 $\beta = 9$ ($r^2 = 0.89$) 作为软阈值 (图 2-A), 网络中基因间的链接符合无尺度网络分布, 结果见图 2-B。动态树切割可以识别模块, 模块中的基因表达值具有相似性。高度相似的模块被合并, 一共有 32 个共表达模块被鉴定, 其大小为 32~6 245 个基因, 每个模块以 1 种颜色作为标记, 而模块“灰色”保留的为没有共表达的基因 (图 3-A)。在模块-特征关系图中 (图 3-B) 中 brown ($r = -0.2$, $P = 0.04$) 与 grey60 ($r = -0.27$, $P = 0.006$) 模块与患者吸烟史特征具有负相关性, 而 salmon ($r = 0.2$, $P = 0.04$) 模块与其具有最高的正相关性。对于吸烟的临床信息, 同样显示着 brown、grey60 和 salmon 模块的基因显著性 (gene significance, GS) 相关性最大 (图 3-C)。因此, 在随后分析中综合 brown、grey60 和 salmon 模块和吸烟史作为待研究临床特征。

表 2 龙葵预测靶点 GO 生物过程富集分析

Table 2 GO terms of biological processes enrichment analysis of predicted targets in *S. nigrum*

GO 生物过程	基因数	占比/%	P 值	Q 值
small molecule biosynthetic process	60	22.14	$4.861\ 12 \times 10^{-34}$	$1.103\ 44 \times 10^{-29}$
drug metabolic process	61	22.51	$9.672\ 12 \times 10^{-34}$	$1.103\ 44 \times 10^{-29}$
response to toxic substance	47	17.34	$4.957\ 30 \times 10^{-29}$	$2.827\ 77 \times 10^{-25}$
organic hydroxy compound metabolic process	46	16.97	$5.661\ 42 \times 10^{-28}$	$2.583\ 53 \times 10^{-24}$
cellular response to hormone stimulus	50	18.45	$4.837\ 51 \times 10^{-26}$	$1.576\ 82 \times 10^{-22}$
small molecule catabolic process	41	15.13	$7.068\ 53 \times 10^{-26}$	$2.016\ 03 \times 10^{-22}$
response to nutrient levels	42	15.50	$3.822\ 46 \times 10^{-25}$	$8.721\ 71 \times 10^{-22}$
response to steroid hormone	33	12.18	$1.798\ 31 \times 10^{-23}$	$2.735\ 46 \times 10^{-20}$
response to molecule of bacterial origin	33	12.18	$1.905\ 71 \times 10^{-21}$	$2.174\ 13 \times 10^{-18}$
reactive oxygen species metabolic process	30	11.07	$6.989\ 65 \times 10^{-21}$	$7.249\ 22 \times 10^{-18}$
response to acid chemical	32	11.81	$1.420\ 63 \times 10^{-20}$	$1.397\ 71 \times 10^{-17}$
response to oxidative stress	36	13.28	$1.470\ 18 \times 10^{-20}$	$1.397\ 71 \times 10^{-17}$
protein autophosphorylation	27	9.96	$6.389\ 55 \times 10^{-20}$	$4.555\ 95 \times 10^{-17}$
inflammatory response	45	16.61	$1.619\ 72 \times 10^{-19}$	$9.988\ 40 \times 10^{-17}$
apoptotic signaling pathway	39	14.39	$3.703\ 24 \times 10^{-19}$	$2.060\ 90 \times 10^{-16}$
gland development	34	12.55	$5.801\ 46 \times 10^{-19}$	$3.127\ 13 \times 10^{-16}$

表 3 龙葵预测靶点 KEGG 信号通路富集分析

Table 3 KEGG pathways enrichment analysis of predicted targets in *S. nigrum*

KEGG 信号通路	基因数	占比/%	P 值	Q 值
fluid shear stress and atherosclerosis	26	9.59	$3.753\ 83 \times 10^{-24}$	$6.190\ 04 \times 10^{-21}$
chemical carcinogenesis	19	7.01	$5.015\ 47 \times 10^{-20}$	$3.814\ 60 \times 10^{-17}$
metabolism of xenobiotics by cytochrome P450	17	6.27	$5.605\ 84 \times 10^{-18}$	$2.610\ 38 \times 10^{-15}$
drug metabolism-cytochrome P450	15	5.54	$1.602\ 27 \times 10^{-15}$	$4.154\ 43 \times 10^{-13}$
platinum drug resistance	15	5.54	$3.134\ 23 \times 10^{-15}$	$7.607\ 84 \times 10^{-13}$
glutathione metabolism	12	4.43	$7.103\ 34 \times 10^{-13}$	$9.314\ 76 \times 10^{-11}$
steroid hormone biosynthesis	7	2.58	$3.840\ 40 \times 10^{-6}$	$1.146\ 94 \times 10^{-4}$

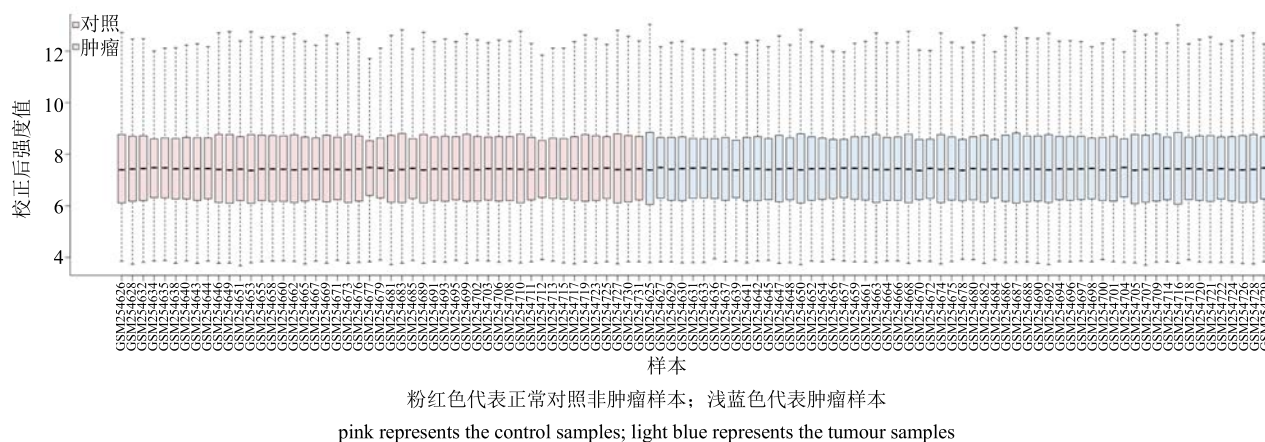
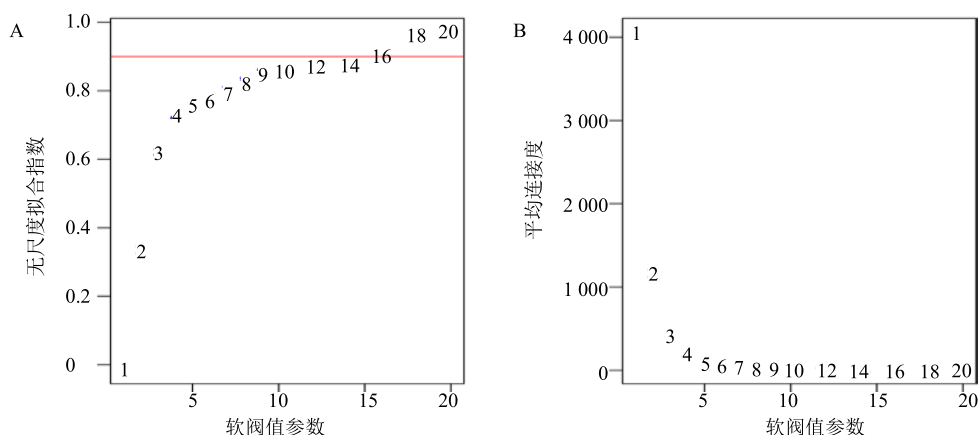


图 1 样品原始强度归一化数值分布

Fig. 1 Normalized raw intensity distributions of samples



A-无尺度的拟合指数 (Y轴) 作为一种软阈值参数 (X轴) 的函数 B-平均连接度 (Y轴) 作为一种软阈值参数 (X轴) 的函数
A-scale-free fit index (Y-axis) as a function of the soft-thresholding power (X-axis) B-mean connectivity (Y-axis) as a function of the soft-thresholding power (X-axis)

图 2 软阈值参数网络拓扑结构的分析

Fig. 2 Analysis of network topology for soft-thresholding parameters

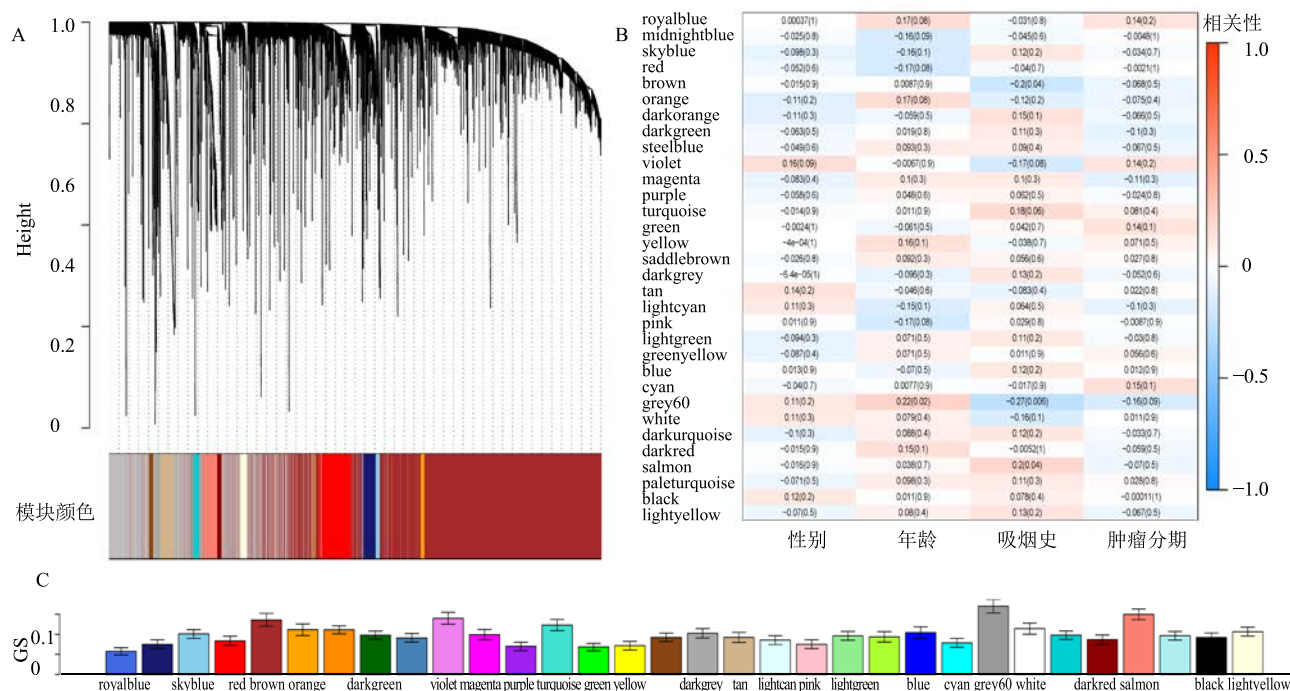
2.5 龙葵活性成分抗癌核心靶点的确定

将龙葵的活性成分所对应的靶点基因和 brown、grey60、salmon 模块相关的靶标基因进行匹配, 获得两者的共同基因即为龙葵可能的抗肺腺癌关键靶点。如图 4-A 所示, 共识别到 46 个龙葵潜在抗肺腺癌基因 (CA2、CA3、HSD11B1、BCHE、EGFR、B4GALT1、PLK1、ICAM2、NTRK2、CASP8、CD44、CYB5R3、CLIC4、DHFR、DAPK1、PDK2、ANXA3、ITGA1、CX3CL1、CASP1、IL1R1、CASP7、HDAC7、FYN、CDK2、NR1H2、ADH1B、CLK1、BMP1R1A、QDPR、PDE4A、NR3C1、PPARD、RARA、JAK2、HPGDS、FGFR2、CD81、PRDX2、HSD17B4、RBP4、MAOA、ADH5、NCOA1、LDHA、B2M), 且主要集中于 brown 模块与吸烟史负相关基因当

中, 可能提示吸烟对人体抗癌基因具有下调效应。经 STRING 数据库构建抗癌基因蛋白互作网络及 Cytoscape 软件计算, 并以节点 $\text{degree} \geq 10$ 为参考条件识别网络中关键核心靶点。如图 4-B 所示, 鉴定 EGFR ($\text{degree} = 16$)、CASP8 ($\text{degree} = 12$)、HPGDS ($\text{degree} = 11$) 和 FYN ($\text{degree} = 10$) 作为龙葵抗肺腺癌潜在核心靶点。

2.6 龙葵抗肺腺癌潜在靶点生物功能及通路分析

GO 注释结果显示, 龙葵的 46 个潜在抗肺腺癌靶点基因在生物过程中的富集主要集中于细菌来源性反应, 对无机物、营养物、氧化应激、有毒物质、有机羟基化合物代谢过程、细胞增殖负调控等生物过程 (图 5-A)。表明龙葵潜在抗肺腺癌靶点可能通过参与上述多个生物学过程而发挥影响癌症进程的



A-于邻接值差异的层次聚类获得基因聚类树 B-模块-特征关系，每一行对应一个模块特征基因，每列对应临床特征，每个格子里都包含相应的第 1 行的相关性值和第 2 行的 P 值；颜色表示相关性大小 C-共表达模块基因显著性
A-gene clustering tree obtained by hierarchical of adjacency-based dissimilarity B-relationship of module and feature. Each row represents a module eigengene and each column to a clinical feature. Each rectangle contains the corresponding correlation in the first line and the P-value in the second line. The table uses color to indicate the significance of the correlation C-gene significance of various co-expression module

图 3 模块基因关联临床特征的确定

Fig. 3 Identification of clinical features combined with modular genes

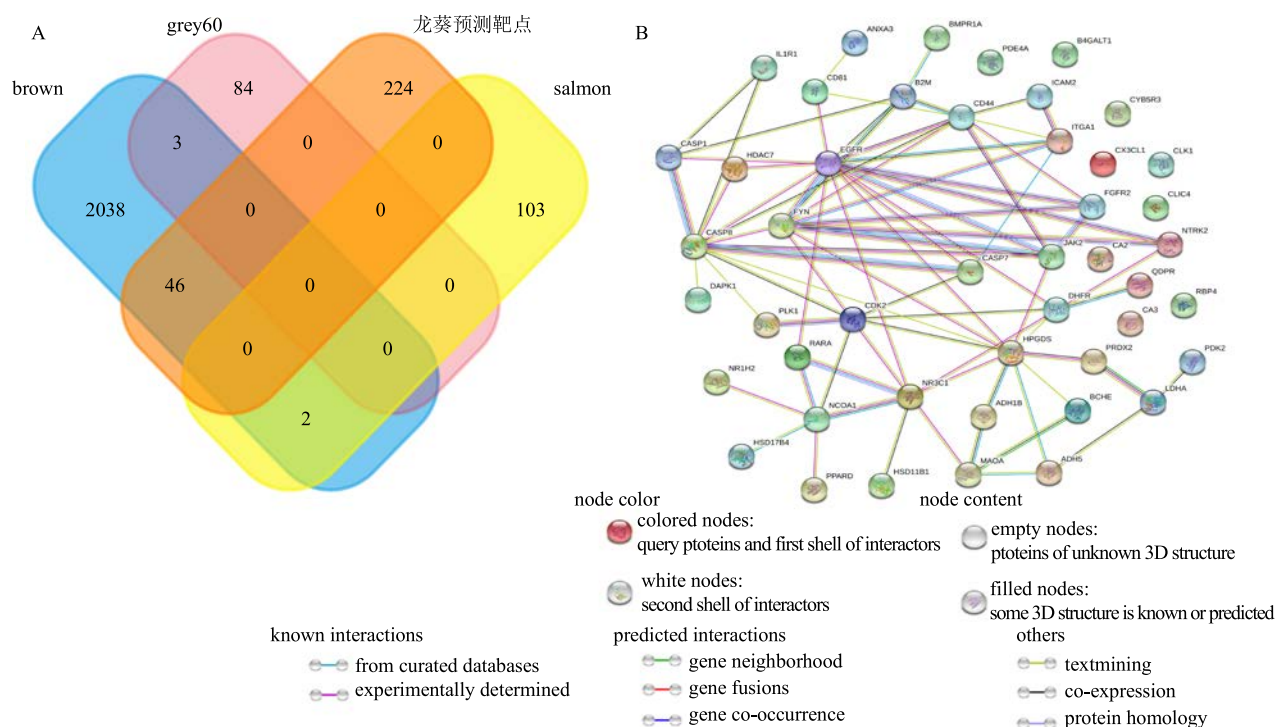


图 4 龙葵活性成分潜在抗肺腺癌靶点的确定 (A) 及抗癌基因蛋白互作网络 (B)

Fig. 4 Identification of potential targets of active components of *S. nigrum* against lung adenocarcinoma (A) and PPI network of anti-oncogenes (B)

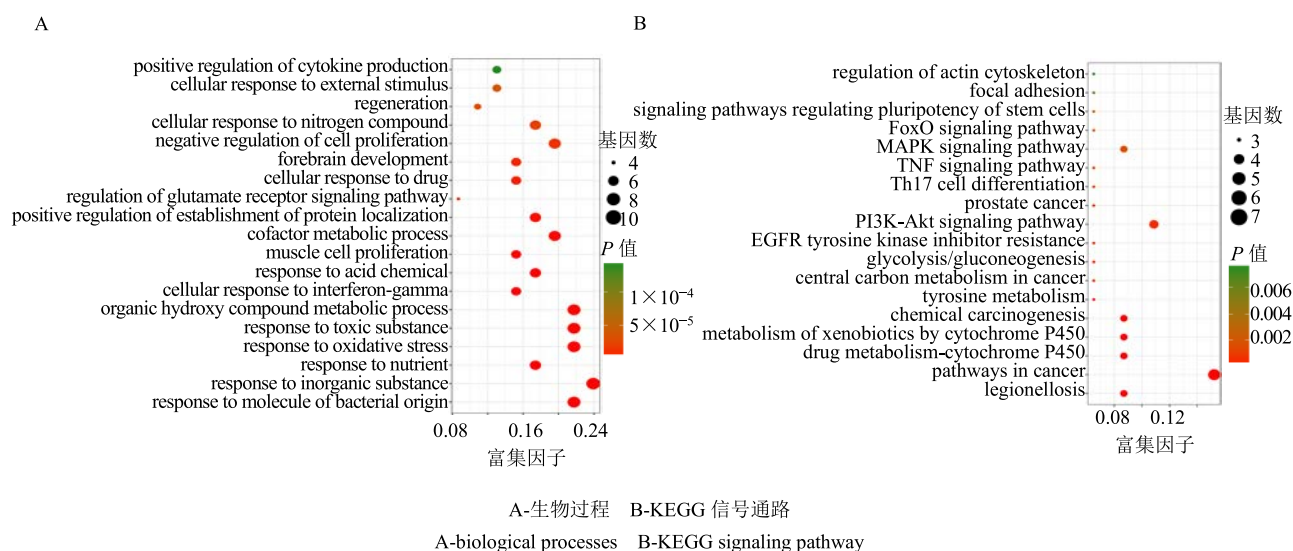


图 5 龙葵活性成分抗癌关键靶点的生物功能富集分析

Fig. 5 Enrichment analysis on biological function of core targets from active ingredients of *S. nigrum*

作用。KEGG 通路富集分析显示, 龙葵活性成分的抗癌基因靶点主要富集的通路为癌症通路、PI3K-Akt 信号通路、MAPK 信号通路、化学物致癌作用、细胞色素 P450 对异物代谢的影响、药物代谢-细胞色素氧化酶 P450、癌症的中心碳代谢等, 并与前列腺癌密切相关, 表明龙葵的活性成分潜在抗肺癌靶点分布于不同通路中, 提示各通路协调发挥抗癌作用(图 5-B)。同时, 潜在靶点基因富集于前列腺癌种上, 提示此类型的癌症可作为未来研究龙葵抗癌疗效的重要方向之一。

2.7 龙葵活性成分-抗癌靶点-通路网络构建与分析

将龙葵活性成分、潜在抗癌基因以及 KEGG 富集信号通路导入 Cytoscape 软件进行网络构建可视化(图 6)。结果表明, 龙葵多活性成分作用于多靶点、多通路。与此同时, 网络提示上述核心靶点 EGFR 可能与薯蓣皂苷元、皮树脂醇、辣椒碱、槲皮素和澳洲茄碱相作用, CASP8 可能与澳洲茄胺、薯蓣皂苷元、辣椒碱相作用, HPGDS 可能与薯蓣皂苷元、辣椒碱和皮树脂醇具有较好结合性, FYN 可能与薯蓣皂苷元、辣椒碱、澳洲茄胺具有较好结合能力, 体现了龙葵多成分-多靶点相互协调共同参与调节多通路的作用特点。

2.8 龙葵活性成分关键靶点临床意义

为了进一步探讨识别得到的龙葵活性成分潜在抗癌靶点的临床意义, 通过 UALCAN 线上工具挖掘 TCGA 数据库关键靶点基因在肺癌的表达水平, 并采用 Kaplan Meier plotter 数据库探讨异常表

达靶点对患者生存率的影响, 从而评估关键靶点的治疗及预后价值。由图 7-A 所示, 靶点基因 EGFR 在肺癌原发性肿瘤当中与正常癌旁组织相比显著降低 ($P=1.4119 \times 10^{-4}$); 然而 CASP8 在原发肿瘤组织中显著升高 ($P=1.6244 \times 10^{-12}$)。此外, 靶点基因 HPGDS ($P=1.6243 \times 10^{-12}$) 和 FYN ($P=2.3227 \times 10^{-7}$) 与正常组织相比其表达水平明显地下降。通过对靶基因的整体生存率分析可知, 靶点基因 EGFR ($HR=1.51$, $\lg P=0.0091$, 图 7-B), CASP8 ($HR=1.36$, $\lg P=0.0450$) 处于高表达时与不良的生存预后显著相关。而 HPGDS ($HR=0.53$, $\lg P=0.0003$) 和 FYN ($HR=0.64$, $\lg P=0.0037$) 处于低表达时与不良的生存预后显著相关。

3 讨论

龙葵中多种活性物质具有丰富的药理效应, 尤以生物碱类成分为主, 现代应用龙葵主治牙痛、感冒、痢疾、慢性支气管炎、乳腺炎和癌症等。随着现代生活节奏的加快, 环境因素、饮食结构等的改变, 癌症的发生率亦是趋于增高。龙葵抗肿瘤作用近些年来成为研究的重点, 对骨髓瘤 U266 细胞、肝癌 HepG2 细胞、乳腺癌 MCF-7 细胞等肿瘤细胞及其荷瘤小鼠体内移植瘤具有抑制作用, 且存在多种作用机制。吴素娜等^[7]发现, 龙葵中的茄碱发挥其抗胰腺癌转移的作用, 可能与抑制胰腺癌细胞 Panc-1 中的基质金属蛋白酶-2 (MMP-2) 及 MMP-9 的表达, 调控 Sata3 信号通路有关。钟伟枫等^[18]研究发现龙葵素诱导前列腺癌 Du145 细胞产生活性氧

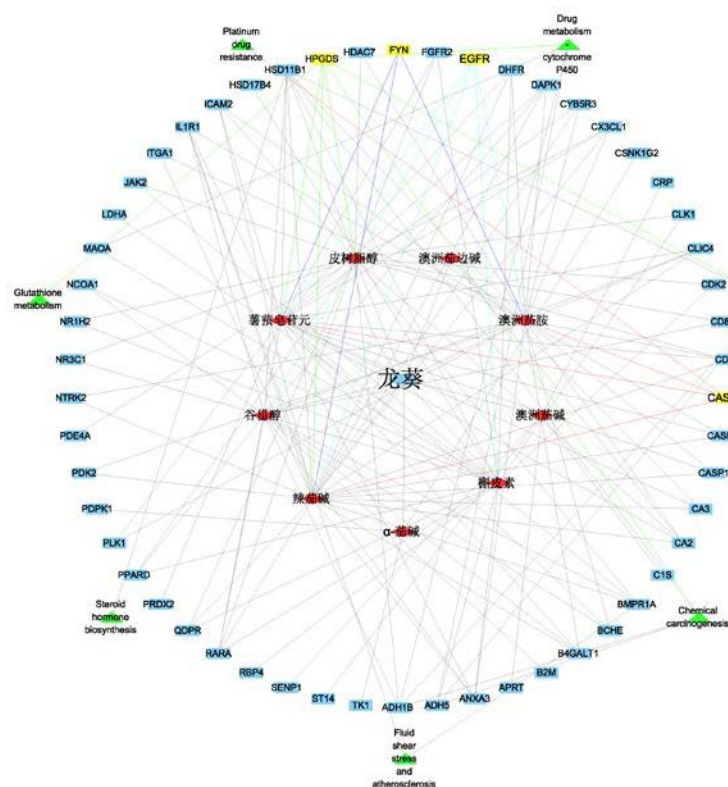


图 6 龙葵活性成分-潜在抗癌靶点-通路网络

Fig. 6 Potential anticancer targets-pathways network of active ingredients of *S. nigrum*

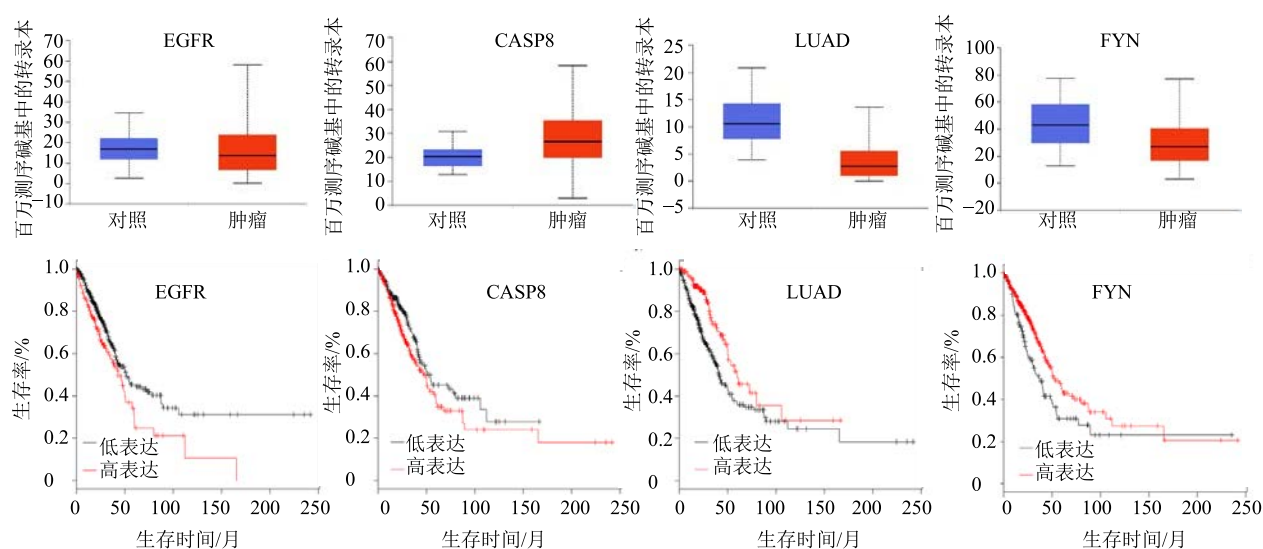


图 7 龙葵活性成分关键靶点在肺腺癌中的表达量及整体生存率的影响

Fig. 7 Effect of expression and overall survival of key targets associated with *S. nigrum*

(ROS), 激活 p38 信号通路促进细胞凋亡。肺癌是全球范围内死亡率最高的恶性肿瘤之一, 石芳等^[19]研究表明澳洲茄碱是含甾体结构的生物碱, 为龙葵抗肿瘤作用的主要活性成分, 对肺癌细胞及 MMPs

和基质金属蛋白酶组织抑制剂 (TIMPs) 表达的作用, 提示澳洲茄碱可抑制肺癌细胞侵袭并与下调 MMP-9、MMP-2 和上调 TIMP-1、TIMP-2 表达有关。发展安全有效的新治疗手段, 并进一步确定识

别与肺癌早期诊断、治疗和预后指标相关的靶点信息是当前肺癌临床治疗亟待完善的重要问题。

本研究应用系统生物学活性成分靶点预测及 WGCNA 深入挖掘与临床预后相关的潜在靶标。应用 TCMSP 数据库及结合文献检索,以皮树脂醇、谷甾醇、薯蓣皂苷元、辣茄碱、槲皮素、 α -茄碱、澳洲茄碱、澳洲茄边碱、澳洲茄胺为研究对象反向预测相应靶点,最终得到龙葵预测靶点共 271 个。进一步通过 WGCNA 方法再次分析了 GEO 数据库中 GSE10072 数据集,得到与吸烟史密切相关的基因共表达模块,通过与龙葵预测靶点相映射匹配,共识别到 46 个龙葵潜在抗肿瘤靶点。在进行 GO 生物过程及信号通路富集时发现,龙葵抗癌靶点主要富集在药物代谢过程、小分子生物合成过程、细胞对激素刺激的反应、对有毒物质的反应、蛋白磷酸化等生物过程。KEGG 信号通路富集结果发现癌症通路、PI3K-Akt 信号通路、细胞色素 P450 对外源物的代谢、药物代谢-细胞色素氧化酶 P450、癌症的中心碳代谢、谷胱甘肽代谢通路密切相关。朱仲玲等^[20]发现,氧化还原和谷胱甘肽代谢与肿瘤进程密切相关,还原型谷胱甘肽(GSH)是体内重要的抗氧化剂和凋亡抑制剂,其水平提高会使肿瘤细胞耐药、抗凋亡。高世勇等^[21]报道,龙葵碱有效减低 HepG2 细胞内的 GSH 水平,使得 ROS 未能及时清除而损伤线粒体结构,促进 HepG2 细胞凋亡,以上报道与本研究具有一定的吻合性。

龙葵预测靶点蛋白互作网络及成分-抗癌靶点-通路网络分析发现,靶点 EGFR 可能与成分薯蓣皂苷元、皮树脂醇、辣茄碱、槲皮素和澳洲茄碱相互作用,CASP8 可能与澳洲茄胺、薯蓣皂苷元、辣茄碱相互作用,HPGDS 可能与薯蓣皂苷元、辣茄碱和皮树脂醇和 FYN 可能与薯蓣皂苷元、辣茄碱、澳洲茄胺具有较好结合性。表皮生长因子受体(epidermal growth factor receptor, EGFR)是表皮生长因子受体家族成员之一,是一种跨膜糖蛋白,通常存在于上皮细胞表面,在恶性肿瘤中常常过表达,EGFR 信号通路在肺癌的发生发展中占有重要作用。EGFR 常在 19 外显子的氨基酸残基 747 至 750 缺失突变和位于 21 外显子的 L858R 点突变,突变后促使肿瘤细胞受体自身磷酸化,酪氨酸激酶活化,从而影响肿瘤细胞的增殖、生长、转移、侵袭和血管生成^[22]。Caspase (CASP) 家族基因在细胞凋亡,肿瘤的发生、发展,肿瘤的化疗疗效及化疗药物的

耐受中也发挥重要作用。CASP8 作为外源性起始 CASP 蛋白进一步激活下游效应 CASPs-CASP3CASP6、CASP7 因子而诱导细胞凋亡的发生发展。虽然目前尚未见到有关龙葵成分通过调节靶点 CASP8、HPGDS 和 FYN 影响肺腺癌进程的报道。但本研究利用公共数据库挖掘发现 EGFR 及 CASP8 高表达、HPGDS 及 FYN 低表达与肺腺癌患者不良预后密切相关。预测所得到龙葵成分和抗肺腺癌靶点与已有文献报道结果吻合,表明基于系统生物学活性成分筛选和靶点预测并结合 WGCNA 策略具有一定的准确性,初步揭示了龙葵抗肺腺癌药效活性成分多靶点、多通路多维调控的作用特点,并为下一步深入研究与龙葵活性成分具有良好结合能力潜在靶点和作用机制提供了方向。

参考文献

- [1] 李杰. MiR-382 对 $\gamma\delta$ T 细胞杀伤肺癌细胞活性的影响及机制研究 [J]. 中国现代应用药学, 2018, 35(1): 24-28.
- [2] Zheng Y Z, Ma R, Zhou J K, et al. ROR1 is a novel prognostic biomarker in patients with lung adenocarcinoma [J]. *Sci Rep*, 2016, doi: 10.1038/srep36447.
- [3] 张伟星, 张浩, 翁伟芳. 川楝素下调 ATF2 蛋白的表达增强顺铂对肺癌细胞的凋亡诱导活性研究 [J]. 中国现代应用药学, 2018, 35(4): 547-551.
- [4] Torre L A, Bray F, Siegel R L, et al. Global cancer statistics, 2012 [J]. *CA Cancer J Clin*, 2015, 65(2): 87-108.
- [5] Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma [J]. *Nature*, 2014, 511(7511): 543-550.
- [6] Wei X, Lai Y, Li J, et al. PSCA and MUC1 in non-small-cell lung cancer as targets of chimeric antigen receptor T cells [J]. *Oncoimmunology*, 2017, 6(3): e1284722.
- [7] 宋会颖, 张虹. 中医药治疗原发性支气管肺癌概况 [J]. 河北中医, 2018, 40(10): 1586-1590.
- [8] 唐朝辉, 张岩, 李娜, 等. 澳洲茄边碱提取纯化工艺及其抗肿瘤作用的研究 [J]. 中国中药杂志, 2011, 36(16): 2192-2195.
- [9] 武凤霞, 赵宁, 梅全喜, 等. 龙葵抗肿瘤机制及其研究进展 [J]. 中药材, 2017, 40(5): 1238-1241.
- [10] 石芳, 巫林, 王妍, 等. 龙葵提取物澳洲茄碱对肺癌细胞侵袭及 MMPs/TIMPs 表达的影响 [J]. 肿瘤药理学, 2018, 8(3): 337-341.
- [11] Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis [J]. *BMC*

- Bioinformatics*, 2008, doi: 10.1186/1471-2105-9-559.
- [12] 邝浩发, 胡艳玲. 基于共表达网络挖掘不同前列腺特异抗原水平下的前列腺癌发展相关基因 [J]. 蛇志, 2018, 30(2): 172-175.
- [13] Qi G, Kong W, Mou X *et al.* A new method for excavating feature lncRNA in lung adenocarcinoma based on pathway crosstalk analysis [J]. *J Cell Biochem*, 2018, doi: 10.1002/jcb.28177.
- [14] An Y, Zhang Q, Li X, *et al.* Upregulated microRNA miR-21 promotes the progression of lung adenocarcinoma through inhibition of KIBRA and the Hippo signaling pathway [J]. *Biomed Pharmacother*, 2018, doi: 10.1016/j.biopha.2018.09.125.
- [15] Yang L, Luo H, Chen J, *et al.* SePreSA: A server for the prediction of populations susceptible to serious adverse drug reactions implementing the methodology of a chemical-protein interactome [J]. *Nucl Acid Res*, 2009, doi: 10.1093/nar/gkp312.
- [16] Chandrashekar D S, Bashel B, Balasubramanya S A H, *et al.* UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses [J]. *Neoplasia*, 2017, 19(8): 649-658.
- [17] 吴素娜, 刘乐伟, 黄智铭. 茄碱通过调控 Stat3 信号通路影响胰腺癌细胞 Panc-1 中 MMP-2 和 MMP-9 表达 [J]. 肝胆胰外科杂志, 2016, 28(5): 401-404.
- [18] 钟伟枫, 刘思平, 姚史武, 等. 龙葵素通过 ROS 激活 p38 信号通路诱导人前列腺癌 Du145 细胞凋亡 [J]. 中药药理与临床, 2016, 32(6): 29-33.
- [19] 石 芳, 巫 林, 王 妍, 等. 龙葵提取物澳洲茄碱对肺癌细胞侵袭及 MMPs/TIMPs 表达的影响 [J]. 肿瘤药理学, 2018, 8(3): 337-341.
- [20] 朱仲玲, 阎 昭. 靶向谷胱甘肽抗氧化系统逆转肿瘤耐药的研究进展 [J]. 中国肿瘤临床, 2015, 42(23): 1143-1147.
- [21] 高世勇, 徐丽丽, 季宇彬. 龙葵碱调控还原型谷胱甘肽和活性氧氧化还原体系损伤线粒体超微结构诱导 HepG2 细胞凋亡 [J]. 中草药, 2009, 40(11): 1779-1784.
- [22] Yasuda H, Park E, Yun C H, *et al.* Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer [J]. *Sci Trans Med*, 2013, doi: 10.1126/scitranslmed.3007205.