

·信息研究·

急性肺损伤差异表达基因的生物信息学分析

李金河¹, 周芹², 刘友坦³, 梁青春^{1,3}

(1. 南方医科大学第三附属医院麻醉科, 广东广州 510630; 2. 中山大学附属第一医院麻醉科, 广东广州 510080;
3. 南方医科大学深圳医院麻醉科, 广东深圳 518000)

摘要:【目的】应用生物信息学方法探讨脂多糖(LPS)诱导小鼠急性肺损伤(ALI)的相关基因,为阐明ALI的发病机制提供新思路。【方法】从NCBI(美国国立生物技术信息中心)公共数据平台GEO(Gene Expression Omnibus)下载基因芯片数据集GSE2411、GSE17355和GSE18341,使用在线分析工具GEO2R筛选出正常肺组织与急性肺损伤组织的差异表达基因。筛选出的差异表达基因利用在线数据库DAVID进行功能注释(GO)和通路分析(KEGG),利用STRING在线数据库和Cytoscape软件进行蛋白互作网络分析并计算网络及各个节点的拓扑特性。【结果】初筛出98个差异表达基因,其中上调的93个,下调的5个。GO分析发现它们主要富集在胞外区,主要参与炎症反应、免疫反应、脂多糖反应等生物学过程,主要参与细胞因子活性、趋化因子活性、CXCR趋化因子受体结合等细胞功能;KEGG分析发现它们参与了TNF信号通路,军团病信号通路和NF- κ B信号通路等。STRING在线数据库和Cytoscape软件分析发现*Il6*、*Tlr2*、*Cxcl2*、*Ccl3*、*Myd88*、*Timp1*、*Tnfaip3*、*Fpr2*、*Nfkb1a*、*Cd274*等为LPS诱导小鼠ALI相关的关键基因。【结论】采用生物信息学方法能够有效分析LPS诱导小鼠ALI差异表达的基因,其中*Cxcl2*、*Timp1*和*Fpr2*三个基因极少或尚未研究,表明其可能是急性肺损伤发病机制的研究新靶点。

关键词:脂多糖;急性肺损伤;差异表达基因;生物信息学

中图分类号:R563.9 文献标志码:A 文章编号:1672-3554(2018)06-0921-07

Bioinformatics Analysis of Differentially Expressed Genes in Acute Lung Injury

LI Jin-he¹, ZHOU Qin², LIU You-tan³, LIANG Qing-chun^{1,3}

(1. Department of Anesthesiology, The Third Affiliated Hospital of Southern Medical University, Guangzhou, 510630; 2.

Department of Anesthesiology, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, 510080; 3.

Department of Anesthesiology, The Affiliated Shenzhen Hospital of Southern Medical University, Shenzhen, 518000)

Corresponding to: Liang Qing-chun, E-mail:liangachun305@163.com

Abstract:【Objective】Bioinformatics analysis was used to screen the related genes of Lipopolysaccharide (LPS) induced acute lung injury (ALI) in mice and find new clues for the mechanism of ALI. 【Methods】We downloaded the gene data sets GSE2411, GSE17355 and GSE18341 from the GEO database, and screened the differentially expressed genes between normal tissues and ALI by using GEO2R analysis tools. Then, we did GO analysis and KEGG pathway analysis by the DAVID online database. Lastly, we analyzed the protein-protein interaction network by the STRING online database, and computed the network topology by the Cytoscape software. 【Results】We found that 98 differentially expressed genes, of which 93 genes were up-regulated and 5 were down-regulated. GO analysis found that these genes were mainly concentrated in extracellular space, and extracellular region. Biological processes were involved in inflammation, immune response, lipopolysaccharide reaction, cellular chemotaxis, immune system processes. Molecular function was involved in cytokine activity, chemokine activity and CXCR chemokine receptor binding. KEGG pathway analysis showed that these genes were mainly involved in the TNF signal pathway, Legionnaires disease signal pathway and NF- κ B signal pathway. Finally, *Il6*, *Tlr2*, *Cxcl2*, *Ccl3*, *Myd88*, *imp1*, *Tnfaip3*, *Fpr2*, *Nfkb1a*, *Cd274* were identified as the key genes associated

收稿日期:2018-08-21

基金项目:国家自然科学基金面上项目(8167010880)

作者简介:李金河,医师,E-mail:jinhemz@163.com;梁青春,通信作者,主治医师,E-mail:liangachun305@163.com

with ALI induced by LPS. 【Conclusions】 The bioinformatics method can be used to analyze the differentially expressed genes of ALI induced by LPS in mice. *Cxcl2*, *Timp1* and *Fpr2* genes which have not been studied, may become new research targets for the pathogenesis of acute lung injury.

Key words: lipopolysaccharide; acute lung injury; differentially expressed gene; bioinformatics analysis

[J SUN Yat-sen Univ (Med Sci), 2018, 39(6):921-927]

急性肺损伤(acute lung injury, ALI)是临床上的常见急重症,主要为肺泡上皮和毛细血管内皮细胞损伤,表现为弥漫性肺间质及肺泡水肿,发病时可导致急性呼吸功能障碍。目前,虽然提出一些新的治疗方案,但其病死率仍很高,可达35%-40%^[1]。ALI的发病机制至今尚未明确,可能是一个涉及多基因以及多信号通路的疾病。现阶段,部分研究围绕ALI的发病机制进行单基因或多基因探讨,虽然它们能够揭示部分基因在ALI的发生发展的作用,但一般较单一且无法全面系统地阐明ALI发生过程中多基因和多信号通路之间的相互作用和关系^[2]。近年来,随着现代生物技术的迅速发展如基因芯片、高通量测序等,为全面研究ALI发病的分子机制提供了新的手段^[3]。因此,本研究希望通过应用生物信息学方法探讨分析脂多糖(LPS)诱导小鼠ALI的致病相关基因,为进一步阐明ALI发病的分子机制提供新的研究线索。

1 材料与方法

1.1 材料

从NCBI公共数据平台GEO(Gene Expression Omnibus)搜索“急性肺损伤”,选择种属为小鼠系列,下载基因芯片数据集GSE2411(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2411>)、GSE17355(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17355>)和GSE18341(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18341>)。GSE2411包含6个LPS标本和6个正常肺组织标本,GSE17355包含LPS 1d、LPS 4d、LPS 10d标本,本项目选取LPS 1d的3个标本和3个正常肺组织标本,GSE18341包含4个LPS标本和4个正常肺组织标本。

1.2 研究方法

1.2.1 差异基因筛选及数据处理 使用GEO2R(<https://www.ncbi.nlm.nih.gov/geo/geo2r/>)对数据集进行处理,GSE2411下载的基因数目为22690个,GSE17355下载的基因数目为13029个,GSE18341下载的基因数目为45037个。对这些基因进行筛选,条件为adj.P.Val < 0.01和 $-\log_{10}FCI > 1$,得出差异表达基因(differentially expressed genes, DEG)进行韦恩图(<http://bioinformatics.psb.ugent.be/webtools/Venn/>)分析。

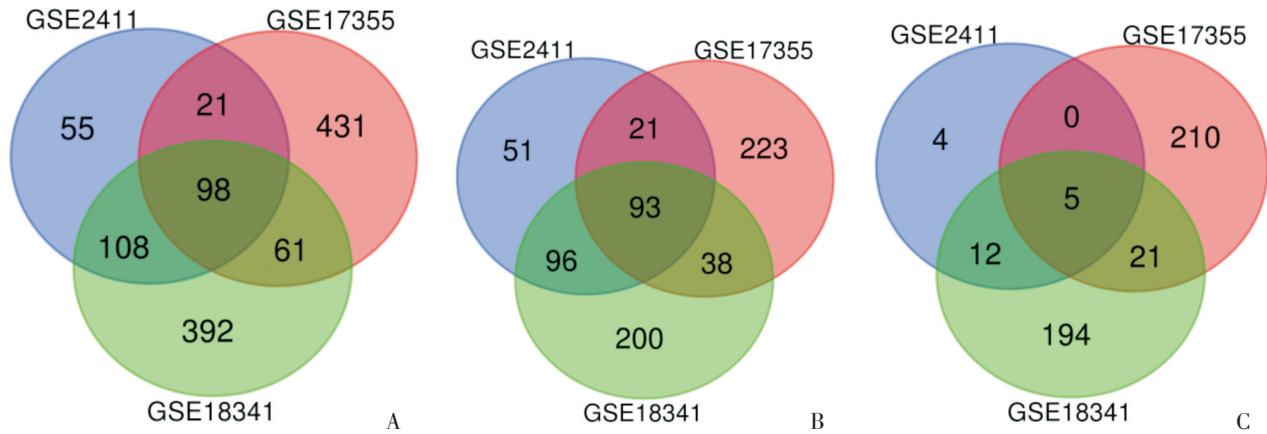
1.2.2 GO富集分析和KEGG通路分析 生物信息学软件DAVID(<http://david.abcc.ncifcrf.gov/conversion.jsp>)是一个在线生物信息数据库,它包含了生物数据和分析工具,其中GO分析用于注释基因生物学过程,KEGG富集分析用于了解其相关信号通路。

1.2.3 蛋白质相互作用网络建设与模块分析 将上述分析得到的DEG导入String在线分析网站(<http://www.string-db.org/>)(version10.5),设置可信度为0.4(confidence:0.4)则认为具有统计学意义,得出蛋白质相互作用网络图。Cytoscape(版本3.6.1)是一个开源性的生物信息学软件平台,用于可视化分子相互作用网络。运用插件MCODE(参数为:DegreeCutoff:2,Node Score Cutoff:0.2,K-Core:2,Max.Depth:100)对关键基因的模块进行可视化图形展示。使用在线Uniprot数据库(<https://www.uniprot.org/>)对互作的节点度均大于7的10个基因进行功能分析。

2 结果

2.1 确定DEG

GEO数据库平台中获得符合标准的基因芯片数据集3个,GSE2411、GSE17355和GSE18341。对它们进行DEG分析,分别筛选出282、611和659个基因。利用韦恩图(图1)可知重叠3数据



A: Differentially expressed genes; B: Up-regulated genes; C: Down-regulated genes

图1 GSE2411、GSE17355、GSE18341 中差异表达基因的表达

Fig.1 Identifications of DEG in mRNA expression profile of GSE2411, GSE17355 and GSE18341

表1 差异表达基因

Table 1 Differentially expressed genes

Up- or Down-regulated	number	Differentially expressed genes
Up-regulated	93	<i>Ctps Cd274 Gpx2 Samsn1 Sod2 Tubb6 Tlr2 Ccl12 Nfkbia Rhou Slc7a2 Timp1 Irgm2 Clec4e D17H6S56E-5 Litaf Map3k8 Cxcl2 Plaur Mt1 Gch1 Ccl3 Csf1 Serpina3n Noct Myd88 Ifit3 Il1r2 Lcn2 Adamts4 Akr1b8 Ccl9 Rsad2 Slc15a3 Gadd45g Sphk1 Ms4a6d Clec4d Asprv1 Oasl1 Mt2 Fpr2 Saa3 Il6 Bcl3 Tnfaip3 St3gal1 Cxcl10 Rgs16 Tnfrsf9 Snx10 Slfn2 Osmr Slc26a4 Cfb Cebpd Slpi Ccl4 Tnfaip2 Ccl17 Parp14 Socs3 Gcle Cd14 Arg2 Plek Upp1 Vcam1 Csf2 Clip4 Icam1 Irgm1 Cxcl5 Casp4 Selp Birc3 Ch25h Thn1 Serpina3m Ifit2 Orm1 Adm Igtf Ifi47 Tifa Nfil3 Nfkib2 Il1b Slfn1 Gbp3 Nfkibz Mefv Cxcl1</i>
Down-regulated	5	<i>Elf5, Ppp1r14c, Tnfrsf19, Meis1, Etv5</i>

集包含了98个基因,其中93个表达上调,5个表达下调(表1)。

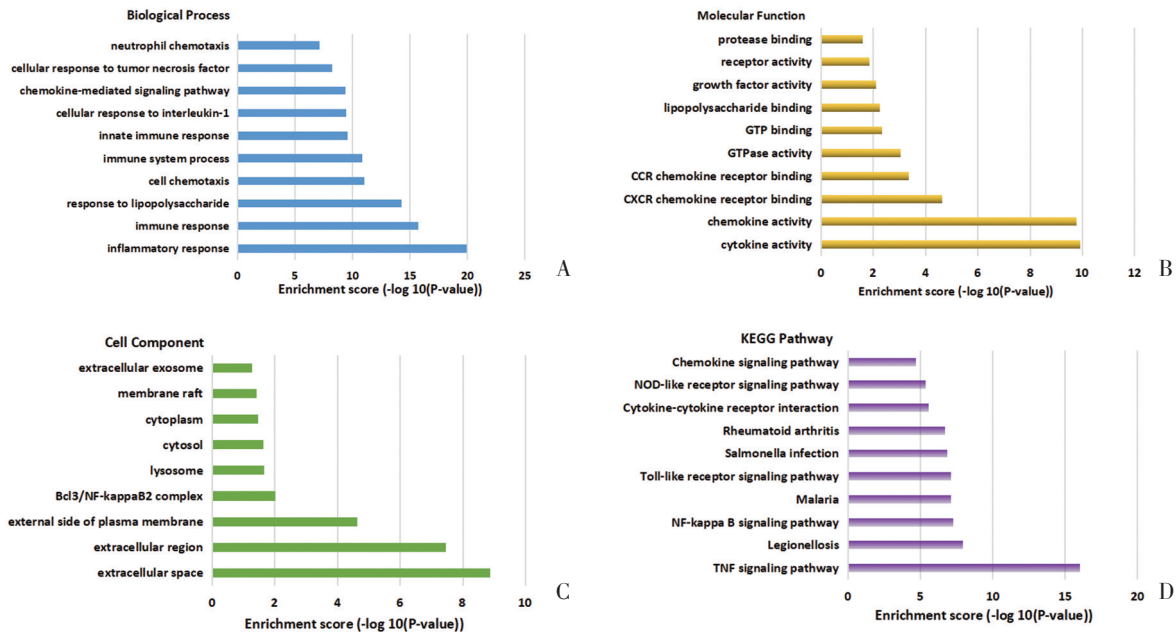
2.2 DEG的GO分析和KEGG通路分析

通过生物信息学软件DAVID对DEG进行分析。通过GO富集分析这些DEG的主要生物学过程(biological process)、分子功能(molecular function)以及细胞组件(cellular component),发现它们主要参与了炎症反应、免疫反应、脂多糖反应、细胞趋化作用、免疫系统过程等生物学过程(图2A);分子功能包括细胞因子活性、趋化因子活性、CXCR趋化因子受体结合、CCR趋化因子受体结合、GTP酶活性等的功能(图2B);而细胞组成

分析显示这些基因大多集中在细胞外区(图2C)。KEGG通路分析显示主要集中的信号通路有3条,包括TNF信号通路(mmu04668:TNF signaling pathway),军团病信号通路(mmu05134: Legionellosis)和NF-κB信号通路(mmu04064: NF-κB signaling pathway)(图2D)。

2.3 DEG的蛋白互作网络分析

在机体内,基因或蛋白的功能发挥通常需要彼此之间进行相互作用或联系。为了查找LPS诱导小鼠ALI相关的关键基因,从蛋白水平探究这些差异表达基因在机体内全方面的相互作用和关系,我们构建了DEG编码蛋白的互作网络。对上



A: Biological Process; B: Molecular Function; C: Cellular Component; D: KEGG Pathway

图2 GO富集分析和KEGG通路分析

Fig.2 GO and KEGG pathway enrichment analysis of DEG

述98个DEG行STRING在线数据库蛋白质相互作用网络分析,结果显示,大部分蛋白间都存在着相互作用,且多个蛋白位于网络的中心,与周围蛋白关系密切(图3A)。插件MCODE分析后发现满足其参数要求的有3个子模块,第一个子模块由8个节点和16条边组成(图3B),第二个子模块由4个节点和6条边组成(图3C),第三个子模块由7个节点和10条边组成(图3D)。

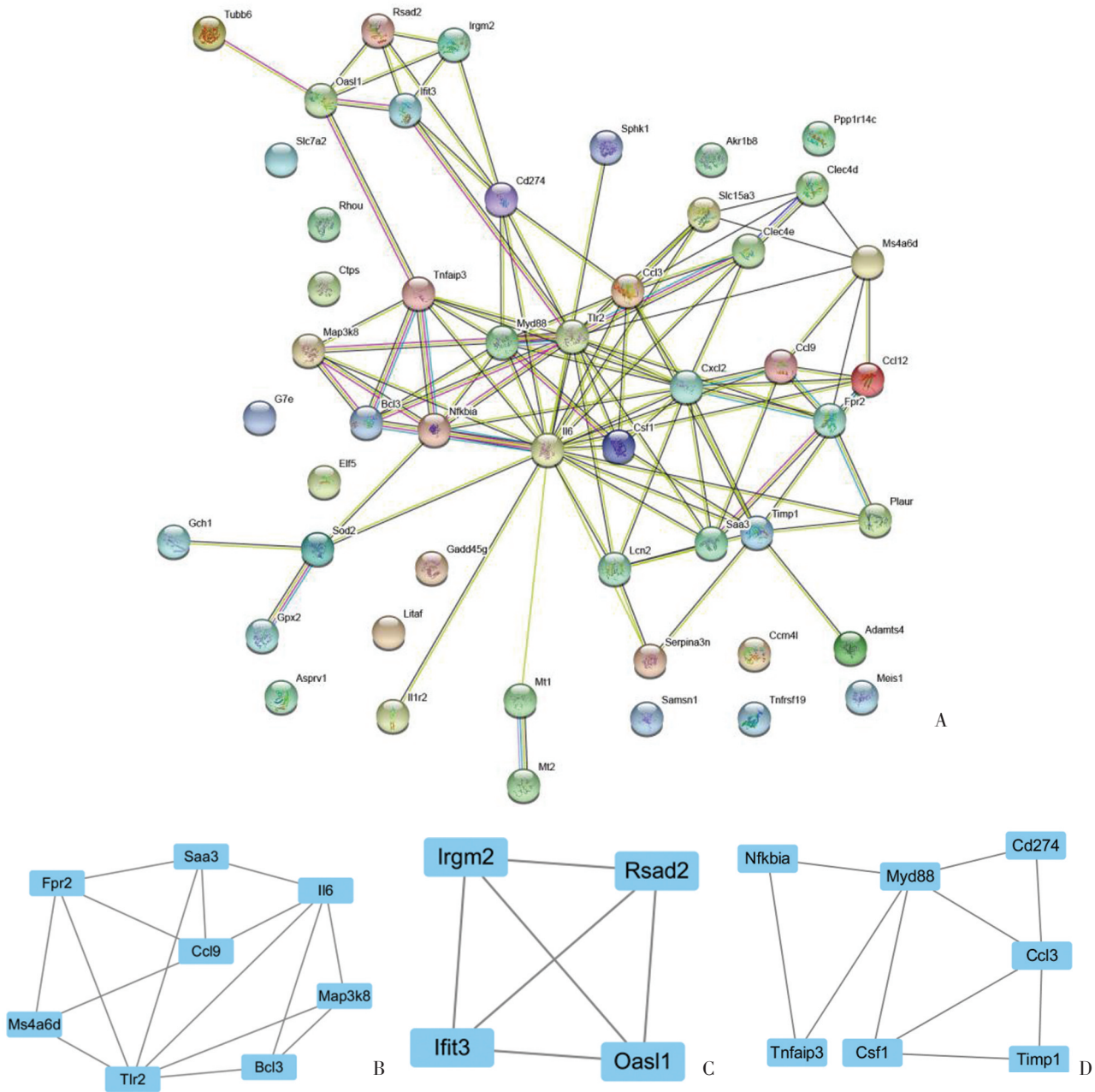
按照互作关系从多到少的原则,初步确定了*Il6*、*Tlr2*、*Cxcl2*、*Ccl3*、*Myd88*、*Timp1*、*Tnfaip3*、*Fpr2*、*Nfkb1a*、*Cd274*这10个基因,它们的基因名称、缩写和功能如表2所示。这10个基因与其它基因存在较强的相互作用关系,互作的节点度均大于7,可能为LPS诱导小鼠ALI相关的关键基因。

3 讨论

近年来,随着现代生物技术的高速发展如生物芯片、高通量测序等的出现,使生物信息学越来越备受关注。通过大数据分析,挖掘与疾病发展过程中起主导作用的相关基因,为疾病的机制研究提供思路是生物信息学的一个重要应用。如邢

静等^[4]利用生物信息学方法筛选特发性肺纤维相关基因,揭示特发性肺纤维的发生是由多种基因表达改变所致。ALI是临床常见病,近年来虽然有多种药物进行了基础研究和临床试验,但其治疗仍没取得令人满意效果^[5]。目前ALI发病机制尚未被完全阐明^[6],因此本研究应用生物信息学方法,通过分析筛选与该疾病相关的基因,为进一步揭示ALI发病的分子机制提供新的研究线索,同时为疾病治疗新靶点的寻找提供理论依据。

在本研究中,我们通过对三个基因芯片数据集进行分析,初步筛选出具有表达差异的基因98个,其中93个表达上调,它们为*Il6*、*Tlr2*、*Cxcl2*、*Ccl3*、*Myd88*等,5个表达下调,即*Elf5*、*Ppp1r14c*、*Tnfrsf19*、*Meis1*、*Etv5*。*Il6*是炎症和免疫中一个重要细胞因子,它的主要表达激活因子是IL1b和肿瘤坏死因子,而且也是TNF信号通路的参与者。在一些动物模型中,通过敲除*Il6*或施用*Il6*或*Il6R*的抗体可以明显预防或抑制疾病进展。如敲除*Il6*可以明显改善大鼠肠道缺血再灌注引起的急性肺损伤^[7]。*TLR2*是TOLL样受体家族(Toll like receptors, TLRs)成员,它们一般与先天性免疫应答相关,能感受不同相关分子模式的病原体



A: PPI network; B, C, D: Significant modules

图3 差异表达基因的蛋白-蛋白相互作用图和差异表达蛋白的子模块相互作用图

Fig.3 The PPI network of DEG were constructed using Cytoscape and Significant modules

分子,从而活化下游信号通路,启动免疫应答。其中的信号通路涉及两种模式,即髓样分化因子(Myd88)依赖型和非依赖型^[8]。前者以TLR3为代表,后者包括TLR2在内的其他TLR家族成员。Tianzhu Z等研究发现,通过调节基因*Tlr2*、*Tlr4*、*Myd88*、*NF-κB*的表达能改善LPS诱导的小鼠急性肺损伤^[9]。CCL3(亦称MIP1)属于趋化因子家

属,它们是机体内炎症反应和免疫应答的重要参与者。研究证明CCL3可以在机械通气引起的急性肺损伤中作为早期的炎症指标^[10]。TNFAIP3(亦称A20)是NF-κB的抑制物,它可以使酶活性具有双重泛素化,在NF-κB以及TNF介导的细胞凋亡信号通路中发挥负调控作用^[11],从而缓慢或抑制炎症发生。研究发现,TNFAIP3在过敏性哮

表2 10个关键基因的功能
Table 2 Functional roles of 10 genes

NO.	Gene	Full name	Function
1	<i>Il6</i>	Interleukin-6	<i>Il6</i> is a potent inducer of the acute phase response.
2	<i>Tlr2</i>	Toll-like receptor 2	<i>Tlr2</i> can cooperate with TLR1 or TLR6 to mediate the innate immune response to bacterial lipoproteins or lipopeptides.
3	<i>Cxcl2</i>	C-X-C motif chemokine 2	<i>Cxcl2</i> can induce the Chemotactic of human polymorphonuclear leukocytes.
4	<i>Ccl3</i>	C-C motif chemokine 3	<i>Ccl3</i> can regulate the inflammatory, pyrogenic and chemokinetic properties.
5	<i>Myd88</i>	Myeloid differentiation primary response protein MyD88	<i>Myd88</i> is the adapter protein involved in the Toll-like receptor and IL-1 receptor signaling pathway in the innate immune response.
6	<i>Timp1</i>	Metalloproteinase inhibitor 1	<i>Timp1</i> inhibits cell migration and proliferation.
7	<i>Tnfaip3</i>	Tumor necrosis factor alpha-induced protein 3	<i>Tnfaip3</i> is involved in immune and inflammatory responses signaled by cytokines.
8	<i>Fpr2</i>	Formyl peptide receptor 2	<i>Fpr2</i> can stimulate or inhibit inflammatory responses.
9	<i>Nfkbia</i>	NF-kappa-B inhibitor alpha	<i>Nfkbia</i> can regulate immune and pro-inflammatory responses on cellular stimulation.
10	<i>Cd274</i>	Cluster of differentiation 274	<i>Cd274</i> plays a critical role in induction and maintenance of immune tolerance to self.

喘、类风湿关节炎、炎症性肠炎和系统性红斑狼疮等炎症性疾病中确实起关键的炎症和免疫调控作用^[12]。另外,在LPS诱导的鼠急性肺损伤模型中,TNFAIP3可以抑制NF- κ B信号通路和调节巨噬细胞的极化来减轻炎症情况^[13]。CD274(cluster of differentiation 274)也称细胞程式死亡-配体1(Programmed cell death 1 ligand 1, PDL1),由*Cd274*基因编码,与其配体结合,可以传导抑制性的信号,减低淋巴结CD8⁺T细胞的增生。有研究发现,CD274可以通过抑制血细胞凝酸酶的激活减轻由脓毒血症导致的急性肺损伤^[14]。

CXCL2属于CXC家族的一个趋化因子,通过与细胞表面趋化因子CXCR2的互动而发挥作用。和其他趋化因子类似,CXCL2也是一种强大的嗜中性粒细胞趋化剂和许多免疫反应的参与者。有研究表明在哮喘的发病过程中,CXCL2对气道平滑肌细胞迁移起调节作用^[15]。虽然目前极少有研究证明CXCL2与ALI相关,但CXCL2能通过促进炎症细胞在损伤和感染组织的浸润和活化,提示其参与炎症过程,可能与ALI的发病形成相关。TIMP1是基质金属蛋白酶的组织抑制剂,能抑制细胞迁移和增殖。研究表明TIMP1能通过CD63和ITGB1激活细胞信号级联,调节细胞分

化、迁移和细胞死亡^[16]。这种机制是否参与了ALI的发病尚待进一步的研究。FPR2是G蛋白偶联受体的经典化学引诱物受体,它的功能复杂,通过识别不同起源和结构多样性的配体,可以刺激或抑制炎症反应。如其结合配体脂氧素-A4(LXA4)和膜联蛋白-A1(ANXA1)则起到抗炎作用;而结合血清淀粉样蛋白-A(SAA)和抗菌肽(LL-37)配体则起促炎作用。同样,ALI的发病是否与FPR2及其相应配体结合产生促炎或抗炎反应有关,仍有待实验验证。

综上,本研究通过应用生物信息学方法对GEO数据库中LPS诱导ALI相关基因的数据集进行挖掘,成功筛选98个差异表达基因,然后对这些基因行GO分析和KEGG通路分析,初步明确了它们的功能和相关的信号通路。此外,通过构建蛋白质互作网络图,从中选出了*Il6*、*Tlr2*、*Cxcl2*、*Ccl3*、*Myd88*、*Timp1*、*Tnfaip3*、*Fpr2*、*Nfkbia*、*Cd274*这10个基因,其中7个关键基因被验证在ALI中的发生中占重要地位,而*Cxcl2*、*Timp1*和*Fpr2*三个基因极少或尚未在ALI中有研究,提示其可能是ALI发病机制新的研究靶点,可能需要进一步体内外实验及临床资料验证。

参考文献:

- [1] Zamboni M, Vincent JL. Mortality rates for patients with acute lung injury/ARDS have decreased over time [J]. *Chest*, 2008, 133(5): 1120-1127.
- [2] Mannam P, Shinn AS, Srivastava A, et al. Mkk3 regulates mitochondrial biogenesis and mitophagy in sepsis-induced lung injury [J]. *Am J Physiol Lung Cellular Mol Physiol*, 2014, 306(7): L604-619.
- [3] Liu D, Mao P, Huang Y, et al. Proteomic analysis of lung tissue in a rat acute lung injury model: Identification of prdx1 as a promoter of inflammation [J]. *Mediators Inflamm*, 2014, 2014(4): 69358-69365.
- [4] 邢静, 黄鑫炎, 郭禹标. 特发性肺纤维化相关基因的筛选和生物信息学分析[J]. *中山大学学报(医学科学版)*, 2017, 38(6): 926-930.
Xing J, Huang XY, Guo YB. Screening and bioinformatics analysis of idiopathic pulmonary fibrosis related genes [J]. *J SUN Yat-sen Univ (Med Sci)*, 2017, 38(6): 926-930.
- [5] Wang G, Huang X, Li Y, et al. Parp-1 inhibitor, dpq, attenuates lps-induced acute lung injury through inhibiting nf-kappab-mediated inflammatory response [J]. *PloS one*, 2013, 8(11): e79757.
- [6] Price LC, McAuley DF, Marino PS, et al. Pathophysiology of pulmonary hypertension in acute lung injury [J]. *Am J Physiol Lung Cellular Mol Physiol*, 2012, 302(9): L803-815.
- [7] Yuan B, Xiong LL, Wen MD, et al. Interleukin6 mRNA knockdown ameliorates acute lung injury induced by intestinal ischemia reperfusion in rats by upregulating interleukin10 expression [J]. *Mol Med Rep*, 2017, 16(3): 2529-2537.
- [8] Qiu Y, Shen Y, Li X, et al. Molecular cloning and functional characterization of a novel isoform of chicken myeloid differentiation factor 88 (myd88) [J]. *Dev Compar Immunol*, 2008, 32(12): 1522-1530.
- [9] Tianzhu Z, Shumin W. Esculin inhibits the inflammation of lps-induced acute lung injury in mice via regulation of tlr/nf-kappab pathways [J]. *Inflammation*, 2015, 38(4): 1529-1536.
- [10] Wu Y, Wang L, Meng L, et al. Mip-1alpha and nf-kappab as indicators of acute kidney injury secondary to acute lung injury in mechanically ventilated patients [J]. *Eu Re Med Pharmac Sci*, 2016, 20(18): 3830-3834.
- [11] Verstrepen L, Verhelst K, van Loo G, et al. Expression, biological activities and mechanisms of action of a20 (tnfaip3) [J]. *Biochem Pharmacol*, 2010, 80(12): 2009-2020.
- [12] Vereecke L, Beyaert R, van Loo G. The ubiquitin-editing enzyme a20 (tnfaip3) is a central regulator of immunopathology [J]. *Trends in Immunol*, 2009, 30(8): 383-391.
- [13] Wang Y, Song Z, Bi J, et al. A20 protein regulates lipopolysaccharide-induced acute lung injury by downregulation of nf-kappab and macrophage polarization in rats [J]. *Mol Med Rep*, 2017, 16(4): 4964-4972.
- [14] Tang L, Bai J, Chung CS, et al. Programmed cell death receptor ligand 1 modulates the regulatory T cells' capacity to repress shock/sepsis-induced indirect acute lung injury by recruiting phosphatase Src homology region 2 domain-containing phosphatase 1 [J]. *Shock*, 2015, 43(1): 47-54.
- [15] Al-Alwan LA, Chang Y, Mogas A, et al. Differential roles of cxcl2 and cxcl3 and their receptors in regulating normal and asthmatic airway smooth muscle cell migration [J]. *J Immunol*, 2013, 191(5): 2731-2741.
- [16] Toricelli M, Melo FH, Peres GB, et al. Timp1 interacts with beta-1 integrin and cd63 along melanoma genesis and confers anoikis resistance by activating pi3-k signaling pathway independently of akt phosphorylation [J]. *Mol Cancer*, 2013, 12(1): 1095-1104.

(编辑 孙慧兰)