论 著

脓毒症关键基因的筛选与生物信息学分析

马　兰[1]，胡迎春[1]，张　川[2]，胡　林[3]，梁思雪[1]，彭　淼[1]，何　亚[1]，陈睦虎[1]
1.西南医科大学附属医院 急诊医学部(泸州 646000)；2.雅安职业技术学院附属医院 内科(雅安 625100)；3.芦山县人民医院 儿科(雅安 625600)

【摘 要】目的 利用生物信息学研究脓毒症的潜在关键基因。**方法** 从基因表达数据库GEO中下载脓毒症基因表达数据集GSE28750：包括10名脓毒症患者为实验组，20名正常人为对照组。通过GEO2R软件数据进行均一化处理以及差异基因筛选（$P<0.01$，倍数>2倍）。差异基因被提交到DAVID进行GO分析和Pathway分析，再将筛选的差异基因提交到STRING中构建蛋白-蛋白互作（PPI）网络寻找关键基因。最后，通过GraphPad Prism 7做筛选关键基因的生存曲线，进而预测关键基因对脓毒血症预后的影响。**结果** 295个差异基因被筛选出来，包括188个基因上调和107个基因下调。差异基因主要富集的GO是免疫反应，小分子代谢过程，信号转导等；激活的信号通路包括：代谢途径，T细胞受体信号通路，MAPK信号通路等。PPI网络显示出关键基因*CD247*，*LCK*位于网络核心，且*CD247*，*LCK*高表达组生存率越高（$P<0.05$）。**结论** 关键基因*CD247*，*LCK*与脓毒症患者生存率呈正相关，可能会成为新的脓毒症生物标志物或药物靶标，但其具体功能还需进一步研究证实。

【关键词】基因芯片；脓毒症；生物信息学；*CD247*；*LCK*
【中图分类号】 R459.7　　**文献标志码** A　　DOI：10.3969/j.issn.2096-3351.2022.03.005

Screening of key genes in sepsis based on bioinformatics

MA Lan[1]，HU Yingchun[1]，ZHANG Chuan[2]，HU Lin[3]，LIANG Sixue[1]，PEBG Miao[1]，HE Ya[1]，CHEN Muhu[1]
1.Department of Emergency Medicine，the Affiliated Hospital of Southwest Medical University，Luzhou 646000，China；2.Department of Internal medicine Ya'an Uocational and Technical College，Ya'an 625100，China；3. Department of Pediatrics，Lushan People's Hospital，Ya'an 625600，China

【Abstract】Objective To investigate the potential key genes in sepsis using bioinformatics. **Methods** The sepsis-related gene expression data set GSE28750 was downloaded from the Gene Expression Omnibus，including 10 patients with sepsis as the experimental group and 20 healthy people as the control group. Data homogenization and differential gene screening were performed by GEO2R software（$P<0.01$，fold change > 2 times）. The differentially expressed genes（DEGs）were submitted to DAVID for GO analysis and Pathway analysis，and then the identified DEGs were submitted to STRING to construct a protein-protein interaction（PPI）network to search for key genes. The survival curves of the identified key genes were plotted by GraphPad Prism 7.0 to predict the impact of key genes on the prognosis of sepsis. **Results** A total of 295 DEGs were screened out，including 188 up-regulated genes and 107 down-regulated genes. The GO where DEGs were enriched mainly contained immune response，small molecule metabolic process，and signal transduction. The activated signaling pathways included metabolic pathway，T cell receptor signaling pathway，and MAPK signaling pathway. The PPI network showed that the key genes *CD247* and *LCK* were located in the core，and that the high expression groups of *CD247* and *LCK* had higher survival rates（$P<0.05$）. **Conclusion** The key genes *CD247* and *LCK* are positively correlated with the survival rate of patients with sepsis. They may become new biomarkers or drug targets for sepsis，but their specific functions need to be confirmed by further studies.

【Key words】Gene chip；Sepsis；Bioinformatics；*CD247*；*LCK*

脓毒症是世界范围内的主要健康问题[1]，是宿主对感染的反应失调，继而导致器官功能障碍。病原体触发了最初的炎症-免疫反应，导致多种内皮、激素、代谢、免疫和其他途径的激活或抑制，继而产生循环和代谢紊乱，从而导致器官功能障碍[2]。脓毒症病情凶险，病死率高，大约有9%的脓毒症患者会发生脓毒性休克和多器官功能不全，重症监护室中一半以上的死亡是由脓毒性休克和多器官功能不全引起的，脓毒症成为重症监护病房内非心脏病人死亡的主要原因[3]。每年有超过75万人患上败血症，其中21.5万人死于该疾病[4]。目前，

基金项目：四川省科技厅项目（2019JDPT0003；2020YFS0517）
第一作者简介：马　兰，本科。
通信作者：陈睦虎，硕士。E-mail：cmh6186@126.com
引用本文：马兰，胡迎春，张川，等.脓毒症关键基因的筛选与生物信息学分析[J].西南医科大学学报，2022，45(3)：206-210.DOI：10.3969/j.issn.2096-3351.2022.03.005

对于脓毒症的治疗，主要给予抗感染、液体复苏、多器官的功能维持等综合治疗[5-6]。然而，即使有体外膜肺氧合（ECMO）这样的高级支持技术，脓毒症的预后仍然欠佳[7]。尽管近几年来脓毒症的研究已经广泛，但是与脓毒症相关的发病率和死亡率仍在增加。早期诊断和早期治疗对脓毒症患者很重要[8]。因此筛选潜在的生物标志物显得尤为重要。

RNA测序技术的进步引发了广泛的大规模基因表达研究和海量信使RNA（mRNA）数据[9]。准确、高效的读取比对是RNA序列分析的基本挑战之一[10]。通过RNA测序（RNA-Seq）进行全转录组测序，并适当结合生物信息学方法，可提供一种强大的工具对基因进行鉴定，以促进基因组疾病的诊断[11]。本研究通过数据集GSE28750找到了在脓毒症患者与正常人之间的差异基因，对差异基因进行生物信息学分析并结合临床资料，找到了关键基因，并采用生存曲线预测关键基因对疾病的影响。本研究有助于从基因水平深入了解脓毒症，这些找到的关键基因，将可能成为新的药物靶点或生物标志物。

1 材料与方法

1.1 数据来源

基因表达数据集GSE28750是从基因表达数据库GEO（http://www.ncbi.nlm.nih.gov/geo/）中下载的，它包含10个脓毒症患者外周血样本和20个正常人外周血样本。将此数据集提交到OmicShare在线数据分析平台中，采用小提琴图来判断数据的均一化水平。

1.2 数据预处理与差异基因筛选

使用在线软件GEO2R对数据集GSE28750进行差异基因分析，得到脓毒症样本和正常样本之间差异基因，设置$P < 0.01$，Fold Change > 2为截取值筛选差异基因，并运用Omic Share Tools的火山图可视化处理差异基因。

1.3 GO和Pathway分析

基因本体论（gene ontology，GO）涉及的基因和基因产物词汇分为三大类，涵盖生物学的生物过程、细胞组分、分子功能，通过GO富集分析以了解差异基因主要富集在哪些生物学功能、途径和细胞定位。通过Pathway分析以了解实验条件下显著改变的代谢通路。KEGG（Kyoto Encyclopedia of Genes and Genomes）是系统分析基因（及其编码产物）间关系、基因功能、基因组信息的数据库，把基因及表达信息作为一个整体网络进行研究。本研究中，我们把差异基因提交到DAVID（https://david.ncifcrf.gov/summary.jsp）中进行GO以及信号通路富集，并设置$P < 0.05$且FDR < 0.05为临界值，筛选出基因富集最显著10个功能和通路，其结果可视化采用OmicShare工具中制作气泡图。

1.4 PPI网络的构建

STRING（https://string-db.org/）数据库用于搜索已知蛋白质之间和预测蛋白质之间相互作用，该数据库可应用于2031个物种，包含960万种蛋白和1380万种蛋白质之间的相互作用。我们将差异基因提交到STRING中，筛选了所有经过实验验证的结果，并去除在网络中没有联系的点，设置最低要求的互动可信度为0.4，得到蛋白-蛋白互作网络（protein-protein interaction network，PPI network）。通过蛋白质之间的相互作用网络，预测得到关键基因。PPI中核心基因绘制热图，很好地展示了所有基因在不同样本的变化规律。在热图中，基因的表达量做了归一化的处理，并且表达模式相似的基因会被归为一类。

1.5 生存曲线

数据集GSE65682有800余例血液样本数据，其中包括400余例脓毒血症患者外周血的样本，每个样本都有具体的生存时间与预后状态。为了进一步研究上述关键的基因对脓毒血症休克患者预后的影响，我们应用GraphPad Prism 7.0制作生存曲线，进而推测该基因对疾病的影响。

2 结果

2.1 差异基因筛选

获得预处理的GSE28750数据集后，此数据集中各个样本的中位数基本位于一条直线上，即样本之间有可比性（图1）。根据$P < 0.01$和|logFC| > 2，获得295个差异基因，包括188个上调基因和107个下调基因（图2）。

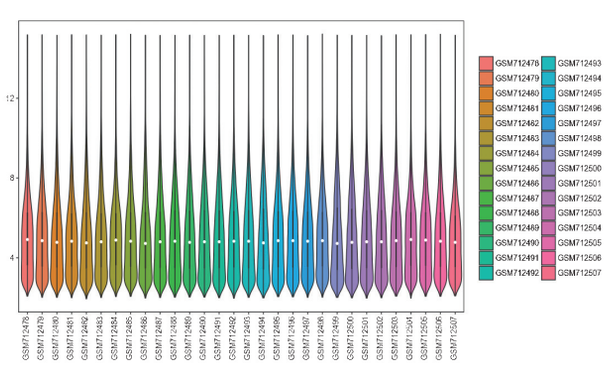


图1 样本基因表达丰度小提琴图

Figure 1 Violin map of gene expression abundance in samples

2.2 GO及Pathway分析

筛选出的差异基因富集最显著20个功能是：免疫反应、小分子代谢过程、细胞表面受体信号通路、信号转导、先天免疫反应、T细胞受体信号通路、凝血（图3）等。最显著的20个信号通路包括：造血细胞谱系、HTLV-I感染、代谢途径、核糖体、T细胞受体信号通路、MAPK信号通路、抗原加工和呈递、金黄色葡萄球菌感染（图4）等。

2.3 PPI网络的构建

PPI网络图是根据既往研究蛋白与蛋白之间的作用，从而推断哪些基因可能更重要。网络节点是蛋白质，蛋白质间的连线表示预测的功能关联。理论上说，连线越多越靠近中央区域，这些基因就越可能更重要，即所谓的核心基因。同时我们结合不同的GO分析（不同颜色代表不同功能），还能初步了解这些核心基因的功能。该网络中*CD247*、*LCK*连线多且颜色多，说明相比之下，它们是具有更多的相互作用且更多功能的蛋白质，被我们筛选为核心基因（图5）。我们通过热图进一步查看这些核心区域的基因表达情况（图6）。

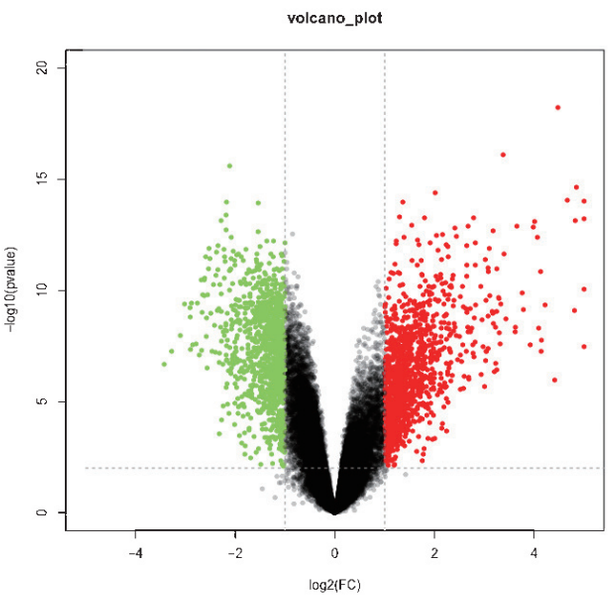


图2　差异表达基因的火山图

Figure 2　Volcanic map of DEGs

注：每个点为一个基因，横坐标为差异倍数，纵坐标为差异显著性P值的负对数（底数为10）。绿色为下调基因，红色为上调基因

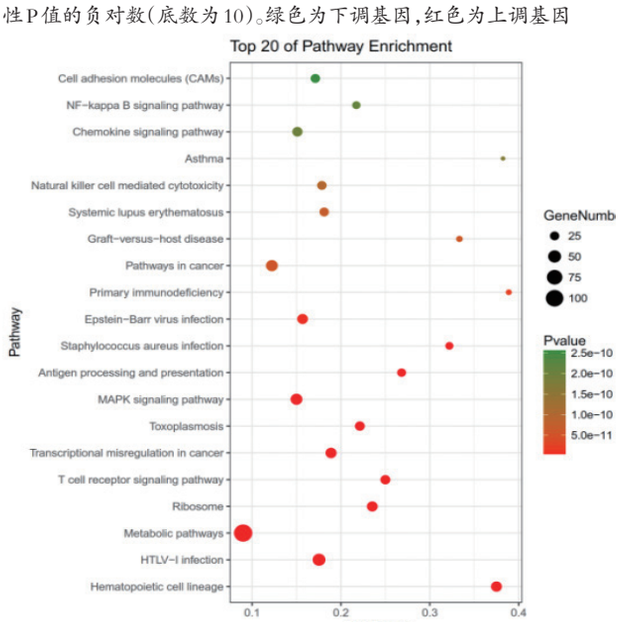


图3　GO分析气泡图

Figure 3　GO analysis

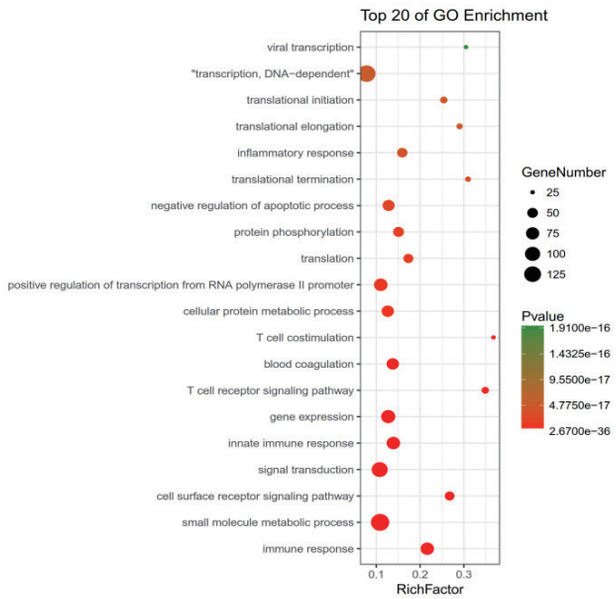


图4　信号通路富集气泡图

Figure 4　Signal pathway enrichment

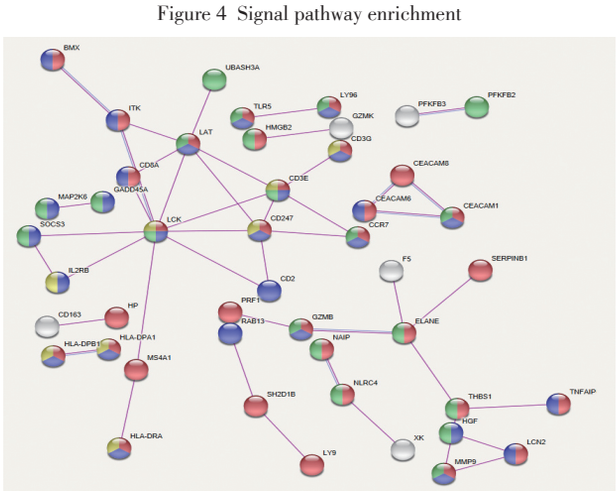


图5　PPI网络图

Figure 5　PPI Network Diagram

注：免疫反应；信号传导；细胞联系的调节；Th1与Th2细胞分化

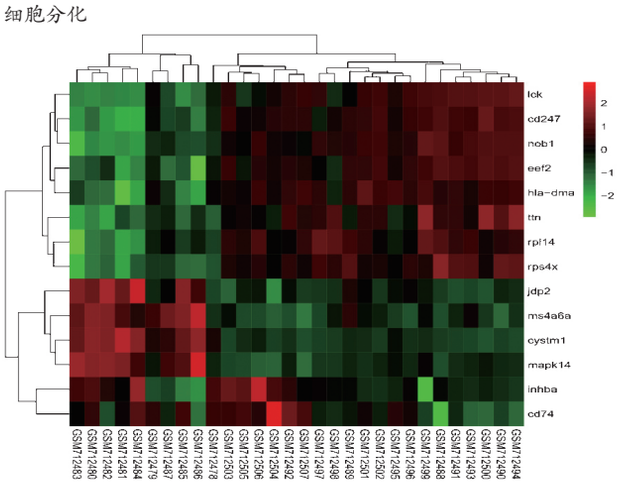


图6　核心基因热图

Figure 6　Heat map of core genes

注：每列为一个样本，每行为一个基因。红色为表达上调，绿色为表达下调。右侧显示的是核心基因的名称

2.4 生存曲线

本研究以 $P < 0.05$ 为截取值，对关键基因 *CD247*、*LCK* 进行生存曲线分析，得出它们均与患者预后呈现正相关，即其中两者表达值越高，生存的时间越长（图7A–B），我们推测这些核心基因可能促进了脓毒症患者的免疫功能。

3 讨论

脓毒症是感染引起的炎症反应失调进而导致的危及生命的器官功能障碍[12]，其导致死亡并消耗大量医疗资源。尽管技术上取得了进步，病理生理学的知识不断提高，但对疾病的机制仍知之甚少[13]。临床上，尽管进行了强有力的抗感染治疗，但仍有一些脓毒症患者死亡，尤其是在合并脓毒症休克时。脓毒症是病原体与宿主免疫系统之间的一场死亡竞赛，促炎症系统与抗炎症系统之间的平衡决定了患者的命运，脓毒症死亡患者T细胞、B细胞、DC细胞以及KCs细胞出现功能衰竭[14]。有大量证据表明脓毒症导致中性粒细胞功能改变，而中性粒细胞通过其产生 H_2O_2 是先天免疫的第一道防线[15]。但目前没有理想的抗菌疗法和药物治疗，治疗脓毒症已经达到了一个瓶颈。在临床实践中，决定患者预后的关键因素及其潜在机制值得我们进行深入研究。我们从基因水平出发，利用生物信息学方法，通过基因芯片分析，找到了关于脓毒症的两个关键基因，这两个关键基因有望成为脓毒症新的生物标志物和治疗靶点。我们从基因表达数据库GEO中下载表达数据集GSE28750，并通过生物信息学技术筛选出多个核心基因[16]，同时结合临床预后数据，最终确定 *CD247*、*LCK* 与脓毒症患者的预后呈现正相关。因此，我们推测这两个基因为脓毒症潜在的核心基因，为后期的研究靶点。

CD247 是T细胞抗原受体（TCR）复合物的重要接头和信号转导分子，它与NK细胞激活受体如NKp46、NKp30和CD16A（FcγRIII）偶联。其表达的非编码序列的多态性和变异是一个严格调控的过程，与多种健康状况下的免疫反应改变有关[17]。目前，已经有研究证明，*CD247* 表达量与器官衰竭负相关[18]，胸腺瘤患者获得的胸腺细胞群中观察到 *CD247* 的存在减少[19]。有研究表明系统红斑狼疮与 *CD247* 的缺失有关[20]，而它是T细胞受体（TCR）CD3复合体的一个组成部分，进而影响免疫应答[21]。

LCK是一种Src相关蛋白酪氨酸激酶，与CD4和CD8分子结合，对T细胞发育和T细胞活化至关重要，已经成为治疗癌症和神经元疾病的新型药物靶向分子[22-23]。LCK作为一种56KDa的蛋白质，可磷酸化各种蛋白质（例如ZAP–70，ITK和蛋白激酶C）的酪氨酸残基，其表达和活性的改变可能导致各种疾病，例如癌症，哮喘，糖尿病，类风湿性关节炎，牛皮癣，炎症性肠病（例如克罗恩氏病和溃疡性结肠炎）[24]。LCK是一种通过T细胞抗原受体进行信号转导所需的非受体蛋白酪氨酸激酶[25]，在T细胞发育的几个阶段需要基于TCR的信号传导，每个基于T细胞抗原受体的信号都需要Src家族激酶，而LCK起主要作用。既往研究发现LCK与小鼠脓毒症预后呈正相关[26]。而我们的GO分析结果显示T细胞受体通路确实被激活，这也再次支持了我们的推测。

4 结论

CD247，*LCK* 等核心基因可能通过T细胞受体通路促进脓毒症患者的免疫功能，从而达到杀灭病原微生物的作用，影响患者的预后。研究与脓毒症相关的 *CD247*，*LCK* 核心基因生物学功能和代谢途径，有利于从基因水平寻找早期诊断脓毒症的生物学标志物和治疗脓毒症的药物靶点，但这些核心基因在脓毒症中的相关作用，仍需进一步研究和证实。

（利益冲突：无）

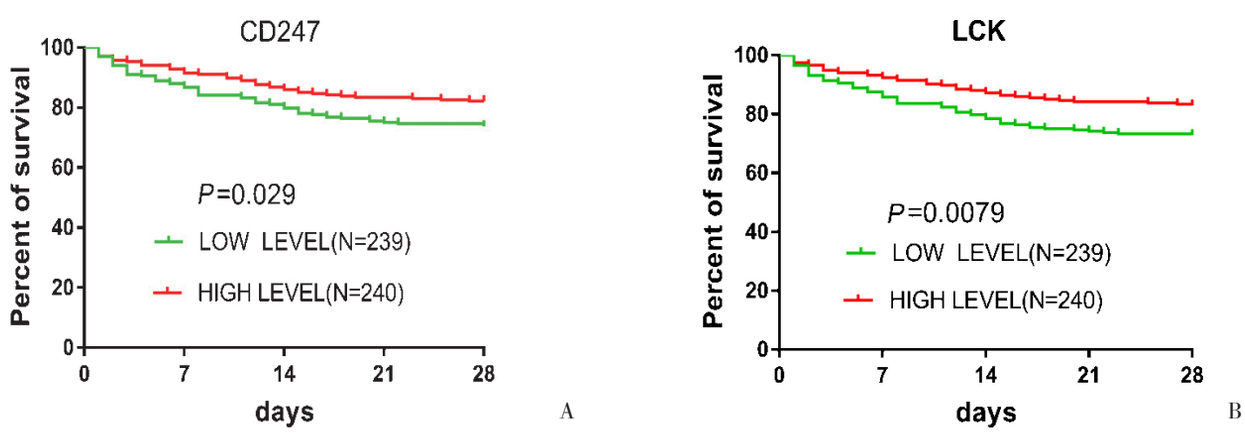


图7 A–B *CD247*（A）与 *LCK*（B）的生存曲线

Figure 7 A–B Survival Curve of *CD247* (A) and *LCK* (B)

注：红色代表高表达组（$n = 240$）；绿色代表低表达组（$n = 239$），$P < 0.05$

参 考 文 献

[1] VAN DER SLIKKE EC, AN AY, HANCOCK REW, *et al*. Exploring the pathophysiology of post-sepsis syndrome to identify therapeutic opportunities[J]. EBioMedicine, 2020, 61: 103044.

[2] ARINA P, SINGER M. Pathophysiology of sepsis[J]. Curr Opin Anaesthesiol, 2021, 34(2): 77-84.

[3] ROBINSON K, KRUGER P, PRINS J, *et al*. The metabolic syndrome in critically ill patients[J]. Best Pract Res Clin Endocrinol Metab, 2011, 25(5): 835-845.

[4] ZHUANG YG, PENG H, CHEN YZ, *et al*. Dynamic monitoring of monocyte HLA-DR expression for the diagnosis, prognosis, and prediction of sepsis[J]. Front Biosci (Landmark Ed), 2017, 22: 1344-1354.

[5] SEYMOUR CW, GESTEN F, PRESCOTT HC, *et al*. Time to treatment and mortality during mandated emergency care for sepsis[J]. N Engl J Med, 2017, 376(23): 2235-2244.

[6] BERGER RE, RIVERS E, LEVY MM. Management of septic shock[J]. N Engl J Med, 2017, 376(23): 2282-2285.

[7] CHANG CH, CHEN HC, CAFFREY JL, *et al*. Survival analysis after extracorporeal membrane oxygenation in critically ill adults: a nationwide cohort study[J]. Circulation, 2016, 133 (24): 2423-2433.

[8] QI L, YAN Y, CHEN BR, *et al*. Research progress of circRNA as a biomarker of sepsis: a narrative review[J]. Ann Transl Med, 2021, 9(8): 720.

[9] ZOABI Y, SHOMRON N. Processing and analysis of RNA-seq data from public resources[J].Methods Mol Biol, 2021, 2243: 81-94.

[10] YU XT, LIU XJ. Mapping RNA-seq reads to transcriptomes efficiently based on learning to hash method[J]. Comput Biol Med, 2020, 116: 103539.

[11] SAEIDIAN AH, YOUSSEFIAN L, VAHIDNEZHAD H, *et al*. Research techniques made simple: whole-transcriptome sequencing by RNA-Seq for diagnosis of monogenic disorders[J]. J Invest Dermatol, 2020, 140(6): 1117-1126.e1.

[12] 朱碧溱. 病毒性脓毒症研究进展[J]. 中华儿科杂志, 2021, 59(3): 250-253.

[13] SILOŞI CA, SILOŞI I, P-DUREANU V, *et al*. Sepsis and identification of reliable biomarkers for postoperative period prognosis [J]. Rom J Morphol Embryol, 2018, 59(1): 77-91.

[14] HOTCHKISS RS, MONNERET G, PAYEN D. Sepsis-induced immunosuppression: from cellular dysfunctions to immunotherapy[J]. Nat Rev Immunol, 2013, 13(12): 862-874.

[15] COHEN J. The immunopathogenesis of sepsis[J]. Nature, 2002, 420(6917): 885-891.

[16] 王海清, 左文, 陈睦虎, 等. 基因生物信息学的脓毒症潜在发病机制研究及生物标记物筛选[J]. 西南医科大学学报, 2018, 41(1): 27-31.

[17] AL-AKIOUI-SANZ K, MORARU M, VILCHES C. A simple genotyping method for CD247 3′-untranslated region polymorphism rs1052231 and characterization of a reference cell panel[J]. HLA, 2021, 98(3): 218-222.

[18] LI YX, ZHOU J, BURKOVSKIY I, *et al*. ATP in red blood cells as biomarker for sepsis in humans[J]. Med Hypotheses, 2019, 124: 84-86.

[19] ALHASHEM F, TIREN-VERBEET NL, ALP E, *et al*. Treatment of sepsis: what is the antibiotic choice in bacteremia due to carbapenem resistant Enterobacteriaceae?[J]. World J Clin Cases, 2017, 5(8): 324-332.

[20] ALMANSA R, HEREDIA-RODRÍGUEZ M, GOMEZ-SANCHEZ E, *et al*. Transcriptomic correlates of organ failure extent in sepsis[J]. J Infect, 2015, 70(5): 445-456.

[21] CHRISTOPOULOS P, DOPFER EP, MALKOVSKY M, *et al*. A novel thymoma-associated immunodeficiency with increased naive T cells and reduced CD247 expression[J]. J Immunol, 2015, 194(7): 3045-3053.

[22] BOZSO SJ, KANG JJH, NAGENDRAN J. The role of competing mechanisms on Lck regulation[J]. Immunol Res, 2020, 68(5): 289-295.

[23] BOMMHARDT U, SCHRAVEN B, SIMEONI L. Beyond TCR signaling: emerging functions of Lck in cancer and immunotherapy[J]. Int J Mol Sci, 2019, 20(14): 3500.

[24] KUMAR SINGH P, KASHYAP A, SILAKARI O. Exploration of the therapeutic aspects of Lck: a kinase target in inflammatory mediated pathological conditions[J]. Biomed Pharmacother, 2018, 108: 1565-1571.

[25] PUNT S, HOUWING-DUISTERMAAT JJ, SCHULKENS IA, *et al*. Correlations between immune response and vascularization qRT-PCR gene expression clusters in squamous cervical cancer[J]. Mol Cancer, 2015, 14: 71.

[26] CHEN MH, CHEN X, HU YC, *et al*. Screening of key genes related to the prognosis of mouse sepsis[J]. Biosci Rep, 2020, 40(10): BSR20202649.

(2021-08-26 收稿, 2022-03-07 修回)