

T 细胞参与慢性阻塞性肺疾病潜在机制的 单细胞转录组生物信息学研究

冯倩¹, 孙岩²

(1. 淄博市张店区中医院 重症医学科, 山东 淄博 255000; 2. 山东第一医科大学附属省立医院
血管外科, 山东 济南 250021)

摘要: **目的** 利用单细胞转录组生物信息学方法探讨 T 细胞参与慢性阻塞性肺疾病发病的潜在机制。**方法** 从高通量基因表达数据库下载单细胞转录组数据集 GSE173896, 利用 R 语言的 Seurat 数据包对数据进行过滤和标准化; 将具有相似特征基因的细胞进行聚类 and 注释, 筛选与慢性阻塞性肺疾病发病相关的细胞亚群, 提取 T 细胞亚群进行二次分析; 利用 R 语言 CellChat 数据包分析 T 细胞与其他细胞亚群之间细胞通讯; 筛选 T 细胞差异表达基因并进行富集分析, 构建蛋白质互作网络筛选核心基因; 下载 GSE69818、GSE37768 和 GSE106986 三个数据集, 验证并筛选慢性阻塞性肺疾病发病的关键核心基因, 绘制受试者操作特征曲线评价关键核心基因预测慢性阻塞性肺疾病发病的价值。**结果** 过滤获得慢性阻塞性肺疾病患者肺组织细胞 9 313 个, 健康对照者肺组织细胞 11 716 个; 慢性阻塞性肺疾病肺组织 T 细胞占比高, 其中辅助性 T 细胞数量高于健康对照组, 调节性 T 细胞数量少于健康对照组; T 细胞与单核细胞、B 细胞、内皮细胞和自然杀伤细胞等细胞的交互强度明显上调; 筛选获得 T 细胞相关的差异表达上调基因 46 个, 差异表达下调基因 32 个; 差异表达基因显著富集于白细胞介素-17、肿瘤坏死因子- α 、核转录因子 Kappa B 和 Toll 样受体等信号通路上。TIMP1、CXCL5 是慢性阻塞性肺疾病发病的关键核心基因。**结论** 慢性阻塞性肺疾病与健康肺组织细胞分布存在较大差异, T 细胞与慢性阻塞性肺疾病发病密切相关, 是预防和治疗慢性阻塞性肺疾病的潜在靶细胞, 慢性阻塞性肺疾病发病涉及多种炎性信号通路, TIMP1、CXCL5 是慢性阻塞性肺疾病发病的关键核心基因。

关键词: 慢性阻塞性肺疾病; 单细胞转录组; T 细胞; 生物信息学

中图分类号: R563.9

Single-cell transcriptomic bioinformatics study of potential mechanisms of T cells involvement in chronic obstructive pulmonary disease

FENG Qian¹, SUN Yan²

(1. Department of Intensive Care Unit, Zhangdian District Hospital of Traditional Chinese Medicine, Zibo, Shandong 255000, China; 2. Department of Vascular Surgery, Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong 250021, China)

Abstract: **[Objective]** To explore the potential mechanism of T cell involvement in the pathogenesis of chronic obstructive pulmonary disease (COPD) using single-cell transcriptome bioinformatics methods. **[Methods]** Single-cell transcriptome dataset GSE173896 was downloaded from the Gene Expression Omnibus database, the data were filtered and normalized using the Seurat package in R. Cells with similar characteristic genes were clustered and annotated to screen for cell subtypes associated with the development of COPD. T cell subtypes were extracted for secondary analysis. Cellchat package in R was used to analyze cell communication between T cells and other cell subtypes. Differential expression genes in T cells were screened and subjected to enrichment analysis to construct a protein interaction network to identify core genes. Three datasets (GSE69818, GSE37768, and GSE106986) were downloaded to verify and screen key core genes related to the development of COPD, and receiver operating

characteristic (ROC) curves were plotted to evaluate the value of predicting the occurrence of COPD. **【Results】** A total of 9,313 lung tissue cells from patients with COPD and 11,716 lung tissue cells from healthy controls were filtered. T cells accounted for a higher proportion of lung tissue cells in patients with COPD, and the number of helper T cells was higher than that in the healthy control group, while the number of regulatory T cells was lower. The interaction intensity between T cells and monocytes, B cells, endothelial cells, and natural killer cells was significantly upregulated. A total of 46 upregulated and 32 downregulated differential expression genes were obtained, which were significantly enriched in signaling pathways such as interleukin-17, tumor necrosis factor- α , nuclear factor kappa B, and Toll-like receptor. *TIMP1* and *CXCL5* were identified as key core genes related to the development of COPD. **【Conclusion】** There are significant differences in the distribution of lung tissue cells between COPD and healthy individuals. T cells are closely related to the pathogenesis of COPD and may be potential target cells for prevention and treatment of COPD. The development of COPD involves multiple inflammatory signaling pathways. *TIMP1* and *CXCL5* are key core genes related to the development of COPD.

Keywords: chronic obstructive pulmonary disease; single-cell transcriptome; T cells; bioinformatics

慢性阻塞性肺病 (chronic obstructive pulmonary disease, COPD) 是一种以不完全可逆性气道阻塞为特征的慢性呼吸系统疾病, 具有较高的死亡率和致残率^[1]。据统计, 全球约有 4 亿人罹患 COPD, COPD 已成为第三大致死疾病^[2]。气道慢性炎症是 COPD 的核心发病机制之一, 是 COPD 发生和发展的重要因素^[3]。既往研究证实, T 细胞, 如 CD4⁺和 CD8⁺ T 细胞的活化、增殖等表型能通过调控炎症反应等途径, 参与 COPD 发病^[4], 但目前 T 淋巴细胞参与 COPD 发病的具体机制未完全阐明。近年来, 单细胞转录组测序技术日臻成熟, 其高分辨率的数据分析可帮助科研工作者更加深入地探索疾病的内在机制。本研究拟利用生物信息学技术挖掘 COPD 样本的单细胞测序数据, 探索 T 细胞在 COPD 发病中的作用, 为 COPD 的机制研究提供参考。

1 资料与方法

1.1 数据处理

从美国公共功能基因组学数据存储库 GEO 中检索 COPD 样本的单细胞测序数据集 GSE173896, 该数据集中包含 3 例 COPD 患者和 3 例健康对照者肺组织的单细胞测序数据。利用 R 语言的 Seurat 数据包对数据进行标准化处理, 并过滤基因异常表达的数据^[5]。

1.2 细胞聚类和注释

利用 Seurat 数据包筛选具有相同特征的细胞, 进行降维分析, 聚类为不同的细胞亚群。使用 t-分布邻域嵌入算法 (t-distributed stochastic neighbor embedding, tSNE) 进行降维、整合、分群。利用 SingleR 数据包注释细胞名称, 使用 SingleR 数据

包中的 “scores within cells” 评分验证细胞注释的可信度。提取 T 细胞亚群进行二次聚类分析^[6]。

1.3 细胞间通讯分析

利用 CellChat 数据包进行细胞通讯分析, 将细胞亚群之间的交互数量和强度进行可视化。

1.4 差异基因筛选和功能富集分析

使用 limma 数据包和 Seurat 数据包筛选存在显著差异分布的细胞亚群中的差异表达基因, 进行基因本体功能注释 (gene ontology, GO) 富集和京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 信号通路富集分析。

1.5 蛋白互作网络构建和核心基因筛选

利用 STRING 数据库构建差异基因蛋白互作 (protein-protein interaction, PPI) 网络, 导入 Cytoscape 软件 (3.9.1) 筛选核心基因并进行可视化。

1.6 核心基因验证

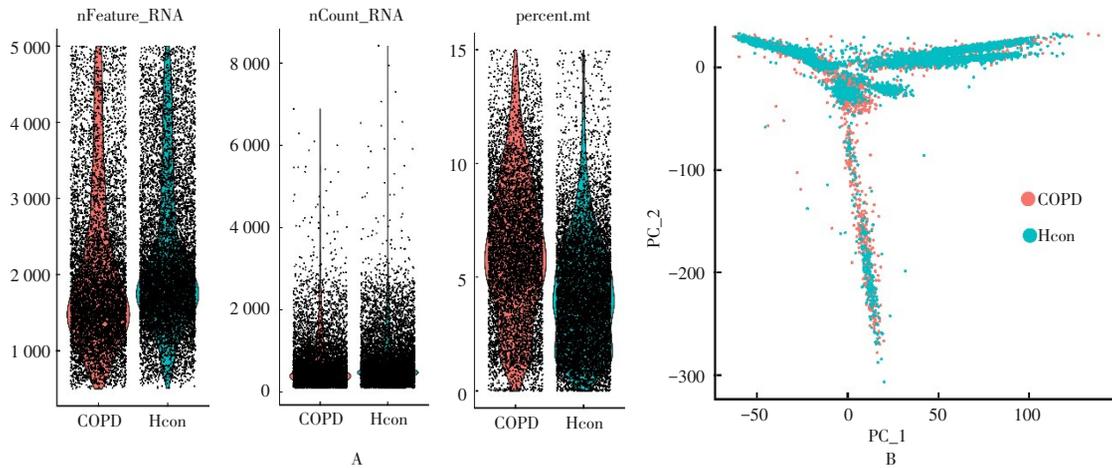
登录 GEO 数据库, 下载 GSE69818、GSE37768 和 GSE106986 三个数据集作为验证集, 验证集包含 102 例 COPD 患者和 14 例健康对照者肺组织的基因表达数据。使用 R 语言 “limma” 数据包对验证集数据进行标准化, 提取核心基因在该数据集中的表达水平并进行差异比较, 采用受试者操作特征 (receiver operating characteristic, ROC) 曲线下面积 (area under the ROC curve, AUC) 评估核心基因预测 COPD 发病的可信度。

2 结果

2.1 单细胞数据清洗与质量控制

共获得 21 029 个细胞的测序数据, 其中 COPD 组细胞 9 313 个, 健康对照 (healthy control,

Hcon) 组细胞 11 716 个, 见图 1。



A: 小提琴图显示筛选得到的总基因数>200、<5 000 和线粒体基因比例<15%的单细胞样本; B: 主成分分析 (principal component analysis, PCA) 去批次效应后显示两组患者的数据重叠良好, 具有可比性。

图 1 单细胞测序数据清洗与质量控制

2.2 单细胞亚群注释

两组样本细胞均包含 T 细胞、B 细胞、单核细胞、巨噬细胞、自然杀伤细胞、上皮细胞、内皮细胞等, 但类别存在显著差异, COPD 组样本中 T 细胞数量占比显著高于 Hcon 组 (图 2)。提取 T 细胞亚群进行聚类、注释, 结果表明, COPD 组样本中辅助性 T 细胞 (T helper cells, Th) 数量显著多于 Hcon 组, 调节性 T 细胞 (regulatory T cells, Treg) 数量显著少于 Hcon 组 (图 3)。

2.3 细胞通讯

T 细胞与单核细胞、自然杀伤细胞、巨噬细胞、B 细胞、内皮细胞等交互作用强度明显高于其他细胞间的交互强度, 提示 COPD 炎症细胞间通讯作用明显上调。上皮细胞交互作用和数量下降最为明显, 而自然杀伤细胞、T 细胞与其他细胞亚群交互作用数量较大。T 细胞与单核细胞、B 细胞、内皮细胞和自然杀伤细胞等细胞的交互强度明显上调, 其中 T 细胞与单核细胞的交互作用上调最为明显 (图 4)。

2.4 差异基因富集分析

使用 R 语言 “limma” 数据包和 Seurat 数据包的 “FindAllMarkers” 函数, 以 “|Log2FC| ≥ 1” 和校正后 $P < 0.05$ 为条件, 筛选各细胞亚群差异表达基因, 提取 T 细胞差异表达基因并进行富集分析 (图 5)。筛选获得 T 细胞相关差异表达上调基因 46 个, 下调基因各 32 个。富集分析结果表明, 与 COPD 发病相关的信号通路以上调表达为主, 涉及

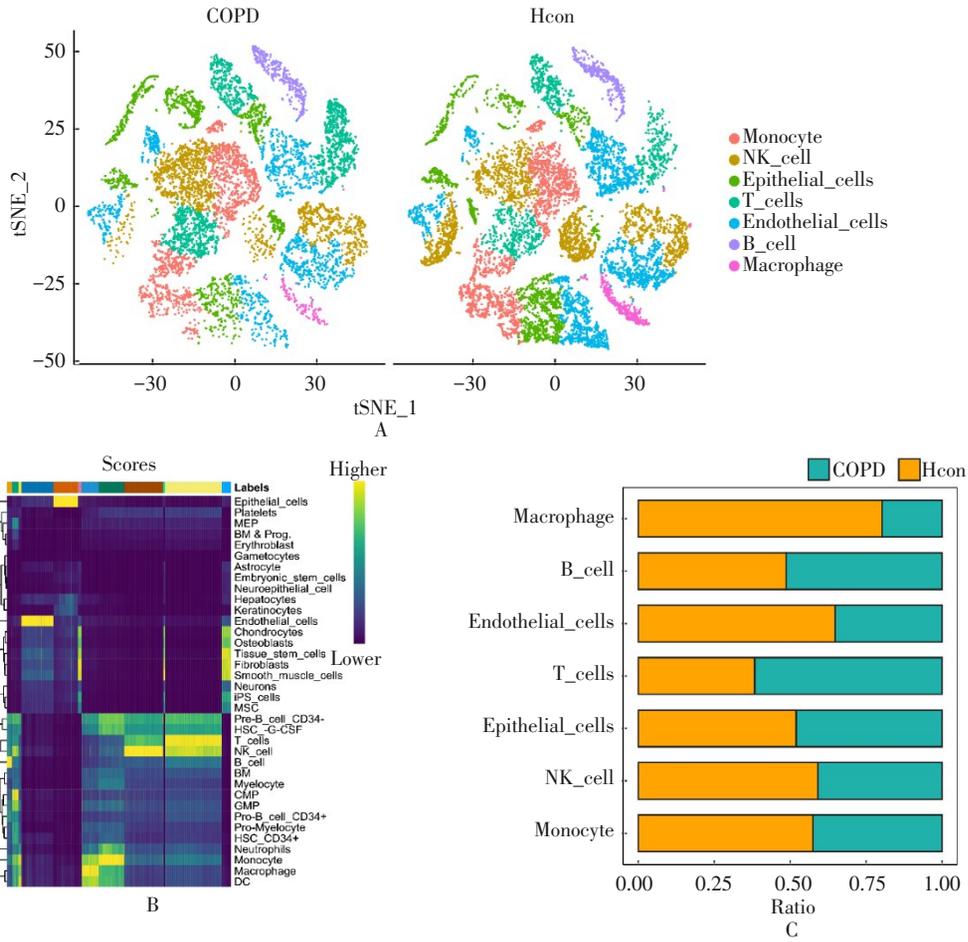
白细胞介素 (interleukin, IL) 17、肿瘤坏死因子 (tumor necrosis factor, TNF) α 、核转录因子 Kappa B (nuclear factor kappa B, NF- κ B) 和 Toll 样受体等信号通路; 涉及中性粒细胞趋化等生物学过程、细胞外空间等细胞组分以及趋化因子活化等分子功能 (图 6)。

2.5 PPI 网络构建

利用 Cytoscape 软件构建 PPI 网络, 筛选核心基因并进行可视化, 结果提示, *CXCL8*、*IL1B*、*CCL2*、*IL1RN*、*TIMP1*、*NFKBIA*、*CXCL2*、*CCL3*、*CCL4* 和 *CXCL5* 是 T 细胞差异表达上调基因的核心基因; *LGALS3BP*、*HLA-DRB1*、*HLA-C*、*HLA-DQB1*、*HLA-DQA2*、*IFI27*、*CD9*、*LY6E*、*THBS1* 和 *SERPING1* 是 T 细胞差异表达下调基因的核心基因 (图 7)。

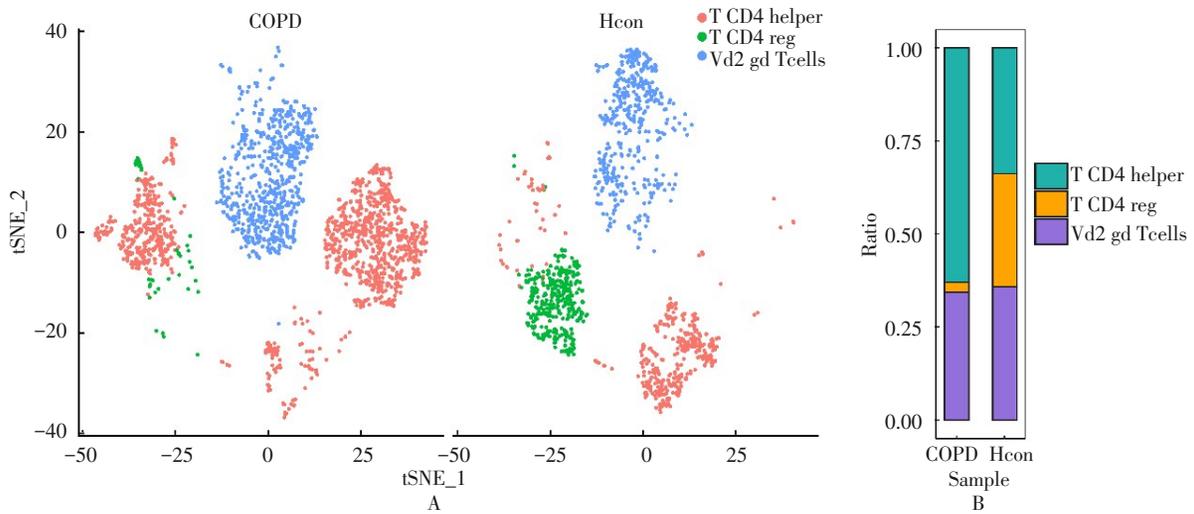
2.6 表达水平验证

提取核心基因在验证集中的表达水平并进行差异比较, 结果表明, T 细胞差异表达上调的 *CXCL5*、*TIMP1* 在验证集中 COPD 样本表达量也显著高于 Hcon 组样本, *HLA-DQA2* 在验证集中未见表达, 其他核心基因表达差异无统计学意义。结果提示, *CXCL5*、*TIMP1* 是 COPD 发病的关键核心基因。ROC 结果表明, *CXCL5*、*TIMP1* 预测 COPD 发病的 AUC 分别为 0.759 (0.631~0.888)、0.681 (0.544~0.818), 二者联合预测 COPD 发病的 AUC 为 0.765 (0.638~0.893), 可信度较好 (图 8)。



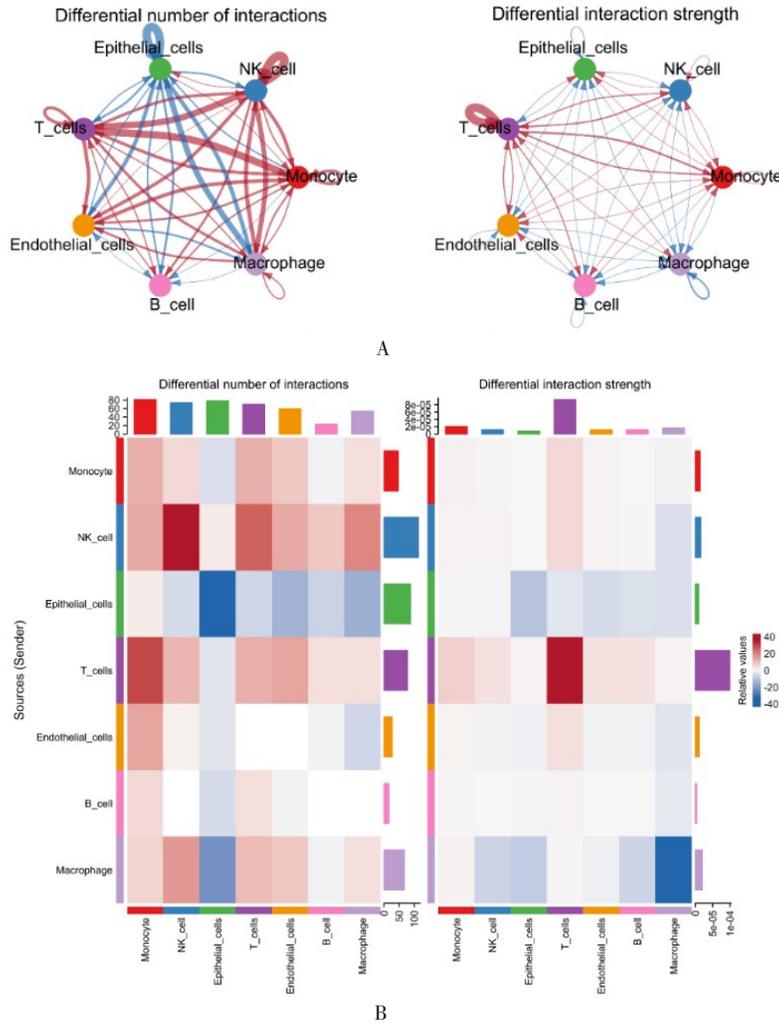
A: 两组样本单细胞注释结果和分布差异; B: 细胞注释评分 (scores within cells); C: 两组样本单细胞分布差异柱状图。

图 2 细胞聚类和注释



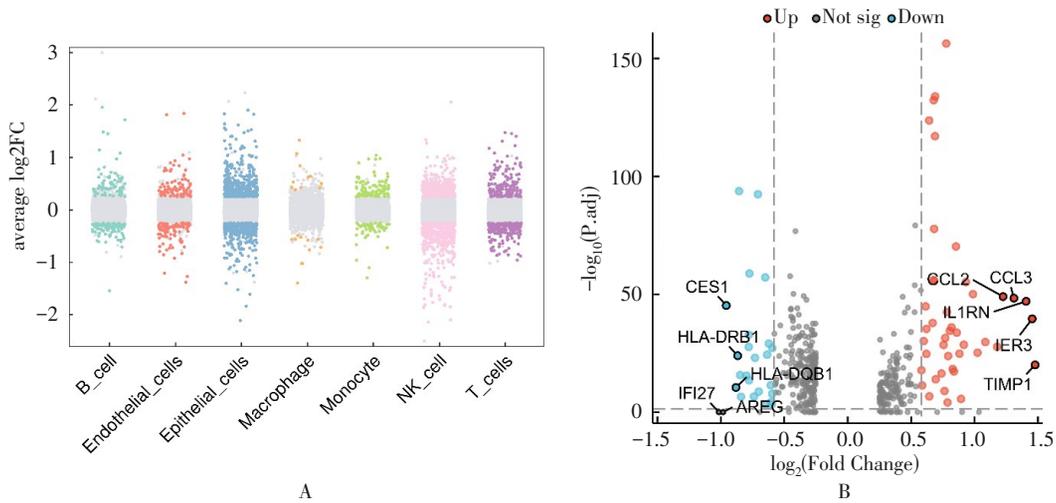
A: 两组样本T细胞亚群tSNE图; B: 两组样本T细胞亚群比例柱状图。

图 3 T 细胞亚群聚类



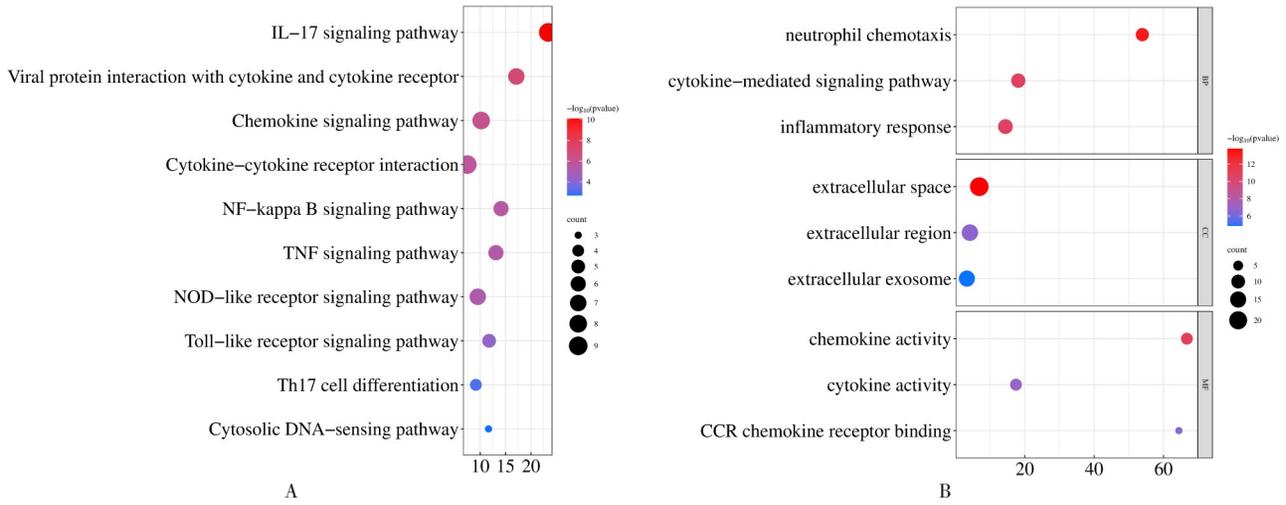
A: 交互作用数量和强度变化圆图, 红色线条表示上调, 蓝色线条表示下调; B: 交互数量热图, 红色表示上调信号, 蓝色表示下调信号。

图 4 COPD 和 Hcon 肺组织细胞通讯网络



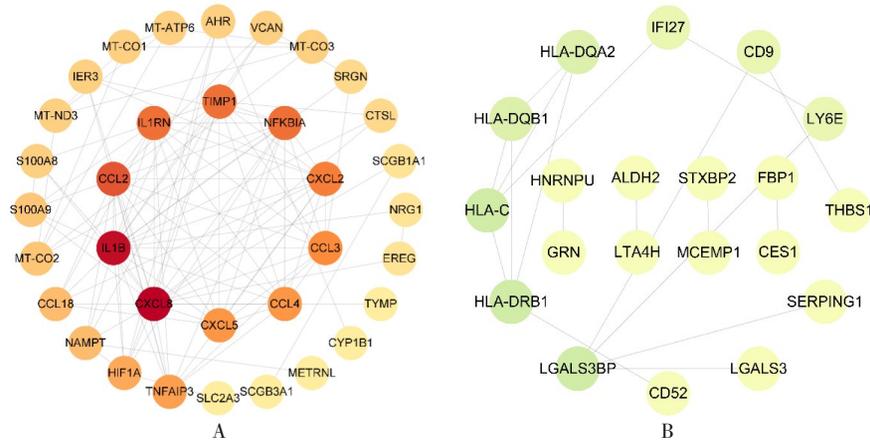
A: 各细胞亚群差异表达基因曼哈顿图; B: T细胞相关差异表达基因火山图。

图 5 差异基因火山图



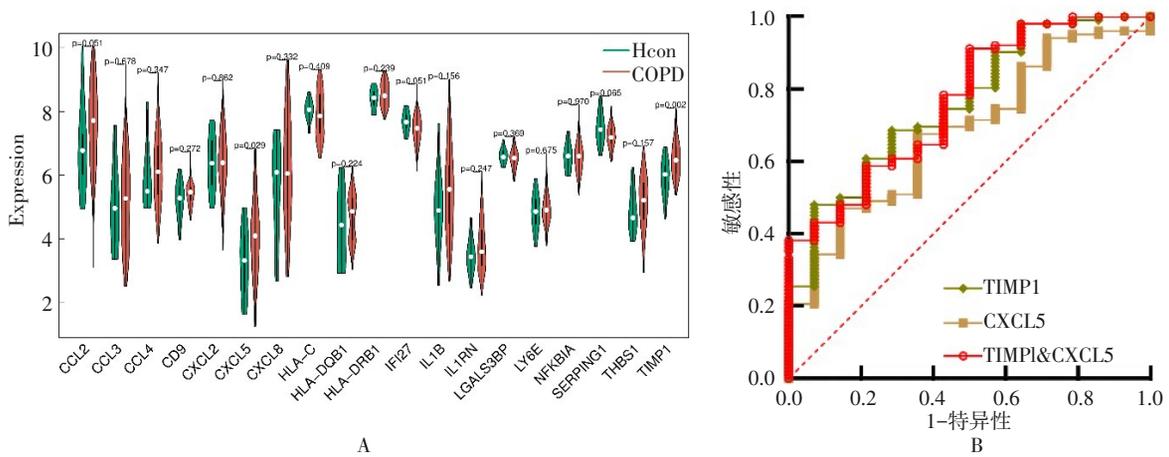
A: 信号通路富集; B: GO 富集。

图 6 T 细胞差异表达上调基因 GO 和 KEGG 信号通路富集分析



A: 差异表达上调基因 PPI 网络; B: 差异表达下调基因 PPI 网络。

图 7 T 细胞差异基因 PPI 网络图



A: 验证集中 T 细胞差异表达基因表达量差异比较; B: 关键核心基因预测 COPD 发病的 ROC 曲线。

图 8 核心基因验证

3 讨论

炎症反应是 COPD 发病的核心机制之一，T 细胞活化、增殖在炎症反应中发挥重要作用^[7]。T 细胞活化和增殖可诱导中性粒细胞释放炎症因子，导致 COPD 患者免疫功能下降，CD4⁺T 淋巴细胞数量减少，CD8⁺T 淋巴细胞释放细胞毒性物质，激活巨噬细胞，导致肺损伤和功能下降^[8]。本研究结果表明，COPD 组样本中 T 细胞数量占比显著高于 Hcon 组；细胞间通讯分析结果提示，T 细胞与其他细胞亚群之间的交互作用强度和数量明显高于其他细胞间的交互强度及数量；提示 T 细胞在 COPD 发病中发挥重要作用。T 细胞亚群聚类、注释结果表明，COPD 组样本中 Th 数量显著多于 Hcon 组，Treg 数量显著少于 Hcon 组。Th1 是 CD4⁺T 细胞亚型之一，主要通过释放干扰素- γ 、肿瘤坏死因子- α (TNF- α) 及白细胞介素-12 (IL-12) 等促炎因子介导细胞免疫^[9]。干扰素- γ 可与 TNF- α 协同，诱导 CXC 趋化因子高表达，促进中性粒细胞介导的炎症反应，介导肺组织破坏，参与 COPD 发病^[10]。Th2 可分泌白细胞介素-4 (IL-4)、白细胞介素-6 (IL-6) 和白细胞介素-9 (IL-9) 等抗炎因子，Th1/Th2 失衡是呼吸系统疾病常见的表型^[11]。既往研究证实，COPD 大鼠模型外周血 Th1 和 Th2 表达上调，且 Th1/Th2 比值升高^[12]。CD4⁺T 细胞还包括 Treg 和 Th17 亚群。研究报道，COPD 患者血浆 Th17 表达水平增加，Th17/Treg 比值升高，提示 Th17/Treg 失衡参与 COPD 发病^[4]。有研究证实，COPD 患者外周血 Treg 数量显著下降，Th17/Treg 比值显著升高^[13]。另有研究发现，COPD 患者外周血 Treg 数量减少，且 Treg 数量与 COPD 严重程度呈负相关^[14]。

本研究中，T 细胞相关差异表达基因富集结果提示，与 COPD 发病相关的信号通路涉及 IL-17、TNF- α 、NF- κ B 和 Toll 样受体等信号通路。IL-17 主要由 Th17 分泌，是经典的促炎因子，既可诱导单核细胞趋化因子在肺部聚集和活化，又可刺激 IL-6、一氧化氮等因子表达，加重 COPD 患者肺部炎症反应^[15]。NF- κ B 可诱导细胞因子和炎症介质表达^[16]。吴洁等^[17] 研究发现，COPD 患者血清 NF- κ B 水平显著高于健康对照组，且 NF- κ B 水平与 COPD 严重程度呈正相关。Toll 样受体是体内自然模式受体，负责连接先天免疫和获得性免疫，当 Toll 结合病原体有关的分子时，相应的信号途

径被激活，生物学效应得以发挥，如果该信号通路活化异常，可加重 COPD、呼吸道感染和鼻炎等疾病的进程^[18-19]。

本研究结果提示，*CXCL5*、*TIMP1* 是 COPD 发病的关键核心基因二者预测 COPD 发病的可信度较好。基质金属蛋白酶组织抑制剂-1 (tissue inhibitor matrix metalloproteinase 1, *TIMP1*) 是细胞外基质降解的重要调节因子，可以调节基质金属蛋白酶的转换效率，二者广泛参与肺损伤、气道重塑等病理过程^[20-21]。研究发现，COPD 患者肺组织 *TIMP1*、基质金属蛋白酶-2 和基质金属蛋白酶-9 的表达水平显著升高，*TIMP1* 和基质金属蛋白酶-2 的表达水平与 COPD 的严重程度呈正相关^[22]。CXC 趋化因子配体 5 (CXC chemokine ligand 5, *CXCL5*) 可诱导中性粒细胞聚集。CHEN 等^[23] 研究证实，COPD 患者和小鼠模型血浆和支气管肺泡灌洗液中 *CXCL5* 表达水平升高，且 *CXCL5* 水平升高与肺功能下降密切相关。WANG 等^[24] 研究发现，COPD 模型小鼠血浆 *CXCL5* 等趋化因子水平升高，发挥募集中性粒细胞，激活中性粒细胞胞外诱捕网，加重 COPD 病情。李玄丹等^[25] 研究证实，COPD 患者血清 *CXCL5* 表达水平升高，且与 COPD 严重程度密切相关。

综上所述，本研究通过生物信息学思路分析 COPD 患者和健康对照者肺组织单细胞转录组数据发现，T 细胞在 COPD 发病中发挥重要作用，COPD 发病涉及 IL-17、TNF- α 、NF- κ B 和 Toll 样受体等信号通路，*TIMP1*、*CXCL5* 等基因是 COPD 发病的关键核心基因。

参考文献

- [1] SUN SL, SHEN YH, WANG J, et al. Identification and validation of autophagy-related genes in chronic obstructive pulmonary disease[J]. Int J Chron Obstruct Pulmon Dis, 2021, 16: 67-78.
- [2] LABAKI WW, ROSENBERG SR. Chronic obstructive pulmonary disease[J]. Ann Intern Med, 2020, 173(3): ITC17-ITC32.
- [3] BRIGHTLING C, GREENING N. Airway inflammation in COPD: progress to precision medicine[J]. Eur Respir J, 2019, 54(2): 1900651.
- [4] ZHANG TS, SHANG FF, MA YH, et al. Caveolin-1 promotes the imbalance of Th17/treg in chronic obstructive pulmonary disease by regulating Hsp70 expression[J]. Int J Chron Obstruct Pulmon Dis, 2023, 18: 565-574.
- [5] MANGIOLA S, DOYLE MA, PAPPENFUSS AT. Interfacing Seurat with the R tidy universe[J]. Bioinformatics, 2021, 37(22): 4100-4107.

- [6] ZHANG DN, LU WP, CUI SS, et al. Establishment of an ovarian cancer omentum metastasis-related prognostic model by integrated analysis of scRNA-seq and bulk RNA-seq[J]. *J Ovarian Res*, 2022, 15(1): 123.
- [7] 陈宇鸣, 于旭华, 梁紫尧, 等. 慢性阻塞性肺疾病大鼠模型 CD8⁺T 细胞自噬水平研究及其数量的相关性分析[J/OL]. *中国免疫学杂志*, 1-17. [2024-11-25]. <http://kns.cnki.net/kcms/detail/22.1126.R.20221205.1132.001.html>.
- [8] XUE WL, MA JY, LI Y, et al. Role of CD₄⁺ T and CD₈⁺ T lymphocytes-mediated cellular immunity in pathogenesis of chronic obstructive pulmonary disease[J]. *J Immunol Res*, 2022, 2022: 1429213.
- [9] XIONG XF, ZHU M, WU HX, et al. Immunophenotype in acute exacerbation of chronic obstructive pulmonary disease: a cross-sectional study[J]. *Respir Res*, 2022, 23(1): 137.
- [10] TANG WH, RONG Y, ZHANG HM, et al. The correlation between a Th1/Th2 cytokines imbalance and vitamin D level in patients with early chronic obstructive pulmonary disease (COPD), based on screening results[J]. *Front Physiol*, 2023, 14: 1032786.
- [11] CHEN ZX, YAN QQ, ZHANG ZM, et al. Immunomodulatory effects of hydrolyzed seawater pearl tablet (HSPT) on Th1/Th2 functionality in a mice model of chronic obstructive pulmonary disease (COPD) induced by cigarette smoke[J]. *Evid Based Complement Alternat Med*, 2020, 2020: 5931652.
- [12] CHI YM, DI QG, HAN GC, et al. MiR-29b mediates the regulation of Nrf2 on airway epithelial remodeling and Th1/Th2 differentiation in COPD rats[J]. *Saudi J Biol Sci*, 2019, 26(8): 1915-1921.
- [13] 邝向东, 刘畅, 王成存, 等. Th1/Th2 和 Th17/Treg 平衡与 COPD 合并下呼吸道感染的关系[J]. *中华医院感染学杂志*, 2022, 32(3): 394-398.
- [14] 陈兴怡, 符丹丹, 肖雪. 慢性阻塞性肺疾病患者外周血辅助性 T 细胞-17/调节性 T 细胞表达及意义[J]. *临床军医杂志*, 2019, 47(4): 436-437, 440.
- [15] 王辉, 叶彩虹, 马焕丽, 等. 吸烟介导的 COPD 呼吸道微生态失调对 Treg/Th17 失衡的影响[J]. *分子诊断与治疗杂志*, 2021, 13(3): 437-440, 444.
- [16] YU H, LIN LB, ZHANG ZQ, et al. Targeting NF- κ B pathway for the therapy of diseases: mechanism and clinical study[J]. *Signal Transduct Target Ther*, 2020, 5(1): 209.
- [17] 吴洁, 向英, 谢芳, 等. 老年 AECOPD 患者 IL-8 和 NF- κ B 及 ICAM-1 与肺功能的关系[J]. *贵州医科大学学报*, 2023, 48(1): 104-108.
- [18] 周勇, 余洋, 程文涛, 等. 虎杖苷对慢性阻塞性肺疾病大鼠气道炎症及 TLR4/NF- κ B 信号通路的影响[J]. *中药药理与临床*, 2019, 35(2): 35-40.
- [19] 何月贤, 宋文秀, 张丽梅, 等. 呼吸道合胞病毒感染致毛细支气管炎患儿血清 TLR4、TGF- β 1 水平变化观察[J]. *山东医药*, 2020, 60(12): 65-68.
- [20] LI XM, MA DH, ZHA XD, et al. Ilomastat, a synthetic inhibitor of MMPs, prevents lung injury induced by γ -ray irradiation in mice[J]. *Oncotarget*, 2017, 8(37): 60789-60808.
- [21] CHRISTOPOULOU ME, PAPAKONSTANTINO E, STOLZ D. Matrix metalloproteinases in chronic obstructive pulmonary disease[J]. *Int J Mol Sci*, 2023, 24(4): 3786.
- [22] ZHANG YX, LI YC, YE Z, et al. Expression of matrix metalloproteinase-2, matrix metalloproteinase-9, tissue inhibitor of metalloproteinase-1, and changes in alveolar *Septa* in patients with chronic obstructive pulmonary disease[J]. *Med Sci Monit*, 2020, 26: e925278.
- [23] CHEN J, DAI LQ, WANG T, et al. The elevated CXCL5 levels in circulation are associated with lung function decline in COPD patients and cigarette smoking-induced mouse model of COPD[J]. *Ann Med*, 2019, 51(5/6): 314-329.
- [24] WANG K, LIAO Y, LI XO, et al. Inhibition of neutrophil elastase prevents cigarette smoke exposure-induced formation of neutrophil extracellular traps and improves lung function in a mouse model of chronic obstructive pulmonary disease[J]. *Int Immunopharmacol*, 2023, 114: 109537.
- [25] 李玄丹, 李材忠, 唐咏婕, 等. CX3CL1、CXCL10 及 CXCL5 在慢性阻塞性肺疾病急性加重期患者中的表达及其预后评估价值[J]. *中国免疫学杂志*, 2022, 38(20): 2523-2527+2533.

(方丽蓉 编辑)