

〔文章编号〕 1007-0893(2021)12-0009-04

DOI: 10.16458/j.cnki.1007-0893.2021.12.003

生物信息学筛选活动性肺结核的潜在生物标志物的临床研究

冶学燕^{1,2} 仝雪薇^{1,2} 刘春燕¹ 张新^{1*}

(1. 石河子大学医学院第二附属医院 新疆生产建设兵团医院, 新疆 乌鲁木齐 830002; 2. 石河子大学医学院, 新疆 石河子 832000)

〔摘要〕 **目的:** 利用生物信息学技术筛选活动性肺结核 (PTB) 诊断和治疗监测的潜在生物标志物。**方法:** 通过 GEO 数据库下载 GSE19444 和 GSE19435 基因芯片数据集。利用 GEO2R 工具筛选活动性 PTB 患者与健康对照间的差异表达基因 (DEG), 使用 CentiScape 插件确定中枢基因 (Hub 基因), 并通过 GSE56153 数据集验证 Hub 基因在活动性 PTB 治疗期间的变化。**结果:** 两个数据集中共筛选 505 个 DEG, 上调 251 个, 下调 254 个。DEG 网络的 9 个子模块中模块 1 评分最高, 包含的 25 个基因主要参与病毒防御、干扰素信号通路等反应。在 25 个基因中筛选了 13 个 Hub 基因, 通过验证, 提示上调的 IFIT3、IFI35、STAT1 和 GBP1 在初诊活动性 PTB、治疗 8 周、治疗 28 周和健康对照组的表达水平有显著降低趋势。**结论:** 筛选出的 IFIT3、IFI35、STAT1 和 GBP1 与活动性 PTB 的发生和治疗密切相关, 有望成为 PTB 诊断和治疗监测的潜在生物标志物。

〔关键词〕 肺结核; 结核分枝杆菌; 生物信息学; 差异表达基因

〔中图分类号〕 R 521 〔文献标识码〕 A

Screening Potential Biomarkers of Active Pulmonary Tuberculosis by Bioinformatics

YE Xue-yan^{1,2}, TONG Xue-wei^{1,2}, LIU Chun-yan¹, ZHANG Xin^{1*}

(1. The Second Affiliated Hospital of Medical College of Shihezi University, Xinjiang Production and Construction Corps Hospital, Xinjiang Urumqi 830002; 2. Medicine School of Shihezi University, Xinjiang Shihezi 832000)

〔Abstract〕 **Objective** To screen potential biomarkers for diagnosis and treatment monitoring of active pulmonary tuberculosis (PTB) by bioinformatics. **Methods** GSE19444 and GSE19435 gene chip data sets were downloaded from GEO database. GEO2R tool was used to screen differentially expressed genes between active PTB patients and healthy controls. CentiScape plug-in was used to determine the hub gene (Hub gene), and GSE56153 dataset was used to verify the changes of Hub gene during active PTB therapy. **Results** A total of 505 DEG, were selected from the two data sets, 251 were up-regulated and 254 were down-regulated. The score of module 1 is the highest among the 9 sub-modules of DEG network, and 25 genes are mainly involved in virus defense, interferon signal pathway and other responses. Among the 25 genes, 13 Hub genes were screened. The results showed that the expression levels of up-regulated IFIT3, IFI35, STAT1 and GBP1 decreased significantly in the first diagnosis of active PTB, 8 weeks of treatment, 28 weeks of treatment and the control group. **Conclusion** The screened IFIT3, IFI35, STAT1 and GBP1 are closely related to the occurrence and treatment of active PTB and are expected to be potential biomarkers for the diagnosis and treatment monitoring of PTB.

〔Key Words〕 Pulmonary tuberculosis; Mycobacterium tuberculosis; Bioinformatics; Differentially expressed genes

肺结核 (pulmonary tuberculosis, PTB) 是由结核分枝杆菌 (mycobacterium tuberculosis, Mtb) 引起的一种广泛传播、危害生命的全球性传染病^[1]。世界卫生组织 (World health organization, WHO) 报告, 2018 年全球新发 PTB 约 1000 万例, 导致 150 万人死亡, 造成沉重的社会负担^[1-2]。快速准确的诊断以及对患者进行治疗效果的有效监测是控制结核病传播和减少患者死亡率的重要手段^[3]。目前, PTB 的诊断方法主

要依赖于患者样本中检测到 Mtb 以及影像学评估, 但存在耗时长、阳性率低等缺点^[4-5]。抗结核药物治疗过程中患者的临床症状消退以及 Mtb 培养转阴性是 PTB 治疗监测的主要指标, 但这些评价指标存在很大程度的不确定性^[6]。因此, 挖掘与 PTB 发生和发展相关的潜在生物标志物对于 PTB 的诊断和治疗监测至关重要。近年高通量技术的基因组学、转录组学等快速发展, 产生了大量基因表达谱数据, 利用生物

〔收稿日期〕 2021 - 04 - 28

〔基金项目〕 新疆生产建设兵团医院科研项目资助课题 (2018010); 兵团科技攻关项目资助课题 (2018AB024)

〔作者简介〕 冶学燕, 女, 在读医学硕士, 主要研究方向是肺结核潜在生物标志物。

〔* 通信作者〕 张新 (E-mail: xjzhangxin108@126.com)

信息技术整合这些数据被广泛用于人类疾病中差异表达基因 (differentially expressed gene, DEG) 的筛选, 为探索疾病的特异性生物标志物提供支持^[7-8]。本研究通过基因表达综合数据库 (gene expression omnibus database, GEO) 选取了 3 个数据集, 利用生物信息学技术分析活动性 PTB 患者和健康对照全血转录组数据, 期望找到在 PTB 中起关键作用的潜在中枢基因, 在分子水平上揭示结核病的发病机理并寻找 PTB 诊断和治疗监测的潜在生物标志物。

1 资料与方法

1.1 数据来源

通过美国国立生物技术信息中心 NCBI 数据库 GEO (<http://www.ncbi.nlm.nih.gov/geo/>), 选取了 3 个与结核病相关的基因表达数据集。GSE19444 和 GSE19435 是 GPL6947 平台 Illumina HumanHT-12 v3.0 芯片^[9], 分析 GSE19444 中 21 例活动性 PTB 未治疗样本和 12 例健康对照样本数据及 GSE19435 中 7 例活动性 PTB 未治疗样本和 12 例健康对照样本数据, 用于中枢基因的筛选。GSE56153 是 GPL6883 平台 Illumina HumanRef-8 v3.0 芯片^[10], 分析活动组 (初诊活动性 PTB)、治疗组 (治疗 8 周) 和恢复组 (治疗 28 周) 3 个时间点 53 例样本和 18 例健康对照样本数据, 用于筛选结果的验证。

1.2 DEG 筛选

利用 GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r>) 在线工具分析筛选组两个芯片基因表达情况, 设置 $P < 0.05$ 和 $|\log FC| > 1$ 为阈值筛选 DEG。使用 R 软件的 ggplot2 包绘制各芯片火山图。

1.3 关键模块基因筛选及分析

使用 STRING 10.0 (<http://string-db.org>) 探索基因/蛋白质的功能及相互作用关系, 设置最低交互得分 > 0.4 , 得到共表达 DEG 的蛋白-蛋白相互作用 (protein-protein interaction, PPI) 网络信息。使用 Cytoscape 3.6.0 及插件 MCODE 识别 PPI 中密集连接的子网络模块。将评分最高的模块基因作为关键模块基因, 使用 STRING 进行基因本体论 (gene ontology, GO) 功能注释、京都基因和基因组百科全书 (kyoto encyclopedia of genes and genomes, KEGG) 通路和 Reactome 通路等分析, 设置显著富集参数错误发生率 (false discovery rate, FDR) < 0.05 , 以了解模块基因参与的生物过程、细胞组成、分子功能及 KEGG 和 Reactome 通路等信息。

1.4 中枢基因 (Hub 基因) 筛选及验证

利用插件 CentiScape^[11] 计算最高评分模块基因的节点中心度, 默认分值为平均值, 设置 Degree Centrality、Betweenness Centrality 和 Closeness Centrality 为 “more”, 用

“or” 连接筛选 Hub 基因。在 GSE56153 数据集中验证 Hub 基因在活动组、治疗组、恢复组和健康对照组的表达趋势变化。

2 结果

2.1 DEG 筛选结果

GSE19444 数据集中 DEG 为 1126 个, 其中上调 587 个, 下调 539 个, 见封三图 1a。GSE19435 数据集中 DEG 为 2514 个, 其中上调 1033 个, 下调 1481 个, 见封三图 1b。在两个数据集中共表达 DEG 为 505 个, 上调 251 个, 下调 254 个。

2.2 关键模块基因筛选及分析结果

在 505 个共表达 DEG 的 PPI 网络中, 识别了 9 个密集连接的模块。评分最高的模块 1 作为关键模块, 包含 25 个表达均上调的潜在中枢基因, 见封三图 2a, 基因表达热图分别见封三图 2b、c。GO 功能分析显示 25 个基因与病毒的防御反应、I 型干扰素信号通路和固有免疫应答等生物过程显著相关。KEGG 和 Reactome 通路表明这些基因与丙型肝炎、NOD 样受体信号通路和细胞因子在免疫系统中的信号等通路显著相关, 见表 1 (表中每一个分析类别只列出了前 3 项)。

表 1 模块 1 中 25 个基因的功能分析

条目	描述	计数	错误发生率
生物过程 (GO)			
GO:0051607	对病毒的防御反应	18	2.38E-28
GO:0060337	I 型干扰素信号通路	13	3.20E-23
GO:0006952	防御反应	21	9.09E-20
分子功能 (GO)			
GO:0001730	2'-5' 寡腺苷酸合成酶活性	3	5.46E-06
GO:0003725	双链 RNA 结合	4	0.00017
GO:0042296	ISG15 转移酶活性	2	0.00087
细胞组成 (GO)			
GO:0005829	细胞内液	16	0.0062
GO:0005737	细胞质	23	0.0105
GO:0044444	部分细胞质	20	0.0415
KEGG 通路			
hsa05160	丙型肝炎	5	1.10E-05
hsa04621	NOD 样受体信号通路	5	2.28E-05
hsa05168	单纯疱疹病毒感染	5	2.58E-05
Reactome 通路			
HSA-913531	干扰素信号传导	17	8.44E-27
HSA-1280215	免疫系统中的细胞因子信号转导	18	3.10E-20
HSA-1169410	IFN 刺激基因的抗病毒机制	9	1.33E-14

2.3 Hub 基因筛选

在 25 个潜在中枢基因中筛选了 GBP1、IFI35、IFI44、IFI44L、IFI6、IFIT1、IFIT3、ISG15、OAS1、OAS2、RSAD2、STAT1 和 XAF1 共 13 个基因作为本研究的 Hub 基因, 其功能主要与巨噬细胞反应和干扰素诱导的蛋白相关, 见表 2。

表 2 13 个 Hub 基因及其功能

基 因	功 能
GBP1	充当上游调节剂, 调节巨噬细胞中多种形式的细胞死亡, 并促进激活微生物特异性下游途径
IFI35	I 型干扰素依赖性转录物, 上调炎症信号
IFI44	该蛋白质聚集形成微管结构
IFI44L	参与干扰素抗病毒反应
IFI6	是定位于内质网的抗凋亡蛋白
IFIT1	干扰素诱导的抗病毒 RNA 结合蛋白, 可特异性结合带有 5'-三磷酸基团的单链 RNA, 从而充当病毒单链 RNA 的传感器抑制病毒信使 RNA 的表达
IFIT3	干扰素诱导的抗病毒蛋白, 可抑制某些病毒信使 RNA 的表达
ISG15	泛素样蛋白, 可诱发病毒和细菌感染或某些遗传毒性应激
OAS1	干扰素诱导蛋白, 这种蛋白质激活潜在的核糖核酸酶 L, 导致病毒 RNA 降解并抑制病毒复制
OAS2	编码 2-5A 合成酶家族的成员, 是涉及病毒感染固有免疫反应的必需蛋白
RSAD2	可以抑制 DNA 和 RNA 病毒, 抗病毒反应和先天免疫中起作用
STAT1	该基因编码的蛋白质可以被各种配体激活, 包括 α -干扰素, γ -干扰素, EGF, PDGF 和 IL-6
XAF1	XAF1 与 IRF-1 形成正反馈回路, 以驱动细胞凋亡应激反应

2.4 活动性 PTB 治疗期间 Hub 基因表达趋势变化

活动性 PTB 初诊到治疗结束期间, 在筛选的 13 个表达水平上调的 Hub 基因中, IFIT3、IFI35、STAT1 和 GBP1 在活动组(初诊活动性 PTB)、治疗组(治疗 8 周)和恢复组(治疗 28 周)和健康对照组的表达水平有依次显著下降趋势见插图六图 3。

3 讨 论

PTB 至今仍是全球死亡率较高的传染病之一。传统的 PTB 诊断标准及 Xpert[®]MTB/RIF 等新技术存在缺陷, 目前仍缺乏 PTB 诊断和治疗监测较理想的方法^[12-13]。近年, WHO 提示应使用非痰标志物检测方法用于结核病的快速诊断, 血液基因表达标志物作为痰液中 Mtb 检查的替代方法引起广泛关注^[14], 一些生物标志物也被证实可作为 PTB 诊断或预后指标。例如, Suzuki Y 等^[15]发现 PTB 患者血清中 sCD206 升高并与预后密切相关, 被认为是 PTB 的潜在生物标志物。洪佳等^[16]通过生物信息学确定了 13 个能够反映 Mtb 活动的核心基因, 这些标志物有望提高临床活动性结核的诊断效率。然而受到宿主免疫反应的复杂性和样本来源异质性的限制, 这些标志物的诊断效率仍欠佳, 因此, 准确的找到 PTB 的潜在生物标志物仍需不断探索。

在本研究中, 利用生物信息学整合了两个来自同一平台的 GEO 基因芯片数据集 GSE19444 和 GSE19435, 筛选出 505 个 DEG。在 DEG 的 PPI 网络中模块 1 评分最高, 故将模块 1 中 25 个基因作为潜在中枢基因进行进一步筛选。随后筛选出了 13 个 Hub 基因, 主要参与巨噬细胞反应和干扰

素诱导的蛋白。通过数据集 GSE56153 验证 Hub 基因在活动性 PTB 治疗期间表达水平, FI44、IFI44L、IFI6、IFIT1、ISG15、OAS1、OAS2、RSAD2 和 XAF1 在 4 个不同分组中的表达水平无差异或仅在某两组间差异显著, 未显示出在治疗过程中明显的逐步降低趋势, 故本研究认为这 9 个基因可能不是活动性 PTB 密切相关基因。而 IFIT3、IFI35、STAT1 和 GBP1 的表达水平在治疗的不同时间点逐步显著降低, 主要参与了 NOD 样受体信号通路、干扰素信号、结核及 Th1 和 Th2 细胞分化通路途径^[17], 它们有望成为活动性 PTB 的诊断和治疗监测的潜在生物靶点。

IFIT3 和 IFI35 是 I 型干扰素诱导的抗病毒蛋白。在结核病发病机理中, 干扰素诱导的蛋白是宿主免疫防御反应的关键介质^[18]。Hare NJ 等^[19]研究表明, 宿主感染 Mtb 后, ISG15、IFIT1、IFIT2 和 IFIT3 表达显著增加并参与促炎症细胞因子的释放。STAT1 是细胞因子和生长因子的响应蛋白, 已在结核病中得到广泛研究^[20-21]。STAT1 的磷酸化可以促进下游凋亡因子激活, 而非磷酸化的转换可以引起 Mtb 免疫逃逸, 这些均取决于宿主 Mtb 感染的时间等特性^[22]。人 GBP1 可以有效调节巨噬细胞凋亡并对感染的病原体产生特异性死亡^[23], 然而巨噬细胞作为 Mtb 感染的主要宿主细胞, 其感染后反应与 GBP1 间的免疫机制鲜有报道。上述研究表明, 本研究利用生物信息学挖掘的 IFIT3、IFI35、STAT1 和 GBP1 与活动性 PTB 发生发展密切相关。这些基因可能充当 Mtb 感染的免疫调节剂、治疗靶标和潜在的生物标记物而值得进一步研究。

本研究利用生物信息学技术构建了活动性 PTB 转录组基因网络, 确定了 IFIT3、IFI35、STAT1 和 GBP1 与 PTB 的发生及治疗密切相关, 这些基因有望作为 PTB 诊断和治疗监测的潜在生物标志物, 但生物信息学结果需要分子生物学实验验证其可靠性。

[参考文献]

- (1) Macneil A, Glaziou P, Sismanidis C, et al. Global Epidemiology of Tuberculosis and Progress Toward Meeting Global Targets-Worldwide, 2018 (J) . MMWR. Morbidity and mortality weekly report, 2020, 69(11): 281-285.
- (2) Long B, Liang SY, Koyfman A, et al. Tuberculosis: a focused review for the emergency medicine clinician (J) . American Journal of Emergency Medicine, 2019, 38(5): 1014-1022.
- (3) Abu-Raddad LJ, Sabatelli L, Achterberg JT, et al. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics (J) . Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(33): 13980-13985.
- (4) Siddiqi K, Lambert ML, Walley JD. Clinical diagnosis of smear-negative pulmonary tuberculosis in low-income countries: The current evidence (J) . Lancet Infectious

- Diseases, 2003, 3(5): 288-296.
- (5) Wei ZH, Li YT, Wei CJ, et al. The meta-analysis for ideal cytokines to distinguish the latent and active TB infection (J) . BMC Pulm Med, 2020, 20(1): 248.
- (6) Alao MA, Maroushek SR, Chan YH, et al. Treatment outcomes of Nigerian patients with tuberculosis: A retrospective 25-year review in a regional medical center (J) . Plos One, 2020, 15(10): e0239225.
- (7) Esterhuyse MM, Weiner J, Caron E, et al. Epigenetics and Proteomics Join Transcriptomics in the Quest for Tuberculosis Biomarkers (J) . mBio, 2015, 6(5): e01187-15.
- (8) Yi XH, Zhang B, Fu YR, et al. STAT1 and its related molecules as potential biomarkers in Mycobacterium tuberculosis infection (J) . J Cell Mol Med, 2020, 24(1): 2866-2878.
- (9) Berry Matthew PR, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis (J) . Nature, 2010, 466(7309): 973-977.
- (10) Ottenhoff T, Hari DR, Yang N, et al. Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis (J) . Plos one, 2012, 7(9): e45839.
- (11) Scardoni G, Petterlini M, Laudanna C. Analyzing biological network parameters with CentiScaPe (J) . Bioinformatics, 2009, 25(21): 2857-2859.
- (12) Kohli M, Schiller I, Dendukuri N, et al. Xpert[®]MTB/RIF assay for extrapulmonary tuberculosis and rifampicin resistance (J) . Cochrane Database Syst Rev, 2018, 8(8): CD0012768.
- (13) Zhang L, Shi X, Zhang Y, et al. Analysis of factors influencing diagnostic accuracy of T-spot. TB for active tuberculosis in clinical practice (J) . Sci Rep, 2017, 7(1): 7764-7770.
- (14) Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures (J) . PLoS Med, 2019, 16(4): e1002786.
- (15) Suzuki Y, Shirai M, Asada K, et al. Macrophage mannose receptor, CD206, predict prognosis in patients with pulmonary tuberculosis (J) . Sci Rep, 2018, 8(1): 13129.
- (16) 洪佳, 李汛. 基于生物信息学的活动性结核生物标志物的筛选 (J) . 国际检验医学杂志, 2020, 41(16): 1983-1986.
- (17) Giovannozzi S, Lemmens V, Hendrix J, et al. Live Cell Imaging Demonstrates Multiple Routes Toward a STAT1 Gain-of-Function Phenotype (J) . Front Immunol, 2020, 11(6): 1114-1122.
- (18) Banks DA, Ahlbrand SE, Hughitt VK, et al. Mycobacterium tuberculosis Inhibits Autocrine Type I IFN Signaling to Increase Intracellular Survival (J) . J Immunol, 2019, 202(8): 2348-2359.
- (19) Hare NJ, Chan B, Chan E, et al. Microparticles released from Mycobacterium tuberculosis-infected human macrophages Controltain increased levels of the type I interferon inducible proteins including ISG15 (J) . Proteomics, 2015, 15(17): 3020-3029.
- (20) Bénard A, Sakwa I, Schierloh P, et al. B Cells Producing Type I IFN Modulate Macrophage Polarization in Tuberculosis (J) . Am J Respir Crit Care Med, 2018, 197(6): 801-813.
- (21) Dupont M, Souriant S, Balboa L, et al. Tuberculosis-associated IFN-I induces Siglec-1 on tunneling nanotubes and favors HIV-1 spread in macrophages (J) . Elife, 2020, 9(3): undefined.
- (22) Yao KZ, Chen Q, Wu YY, et al. MycobacteriumUnphosphorylated STAT1 represses apoptosis in macrophages during infection (J) . J Cell Sci, 2017, 130(10): 1740-1751.
- (23) Kim BH, Shenoy AR, Kumar P, et al. A family of IFN- γ -inducible 65-kD GTPases protects against bacterial infection (J) . Science, 2011, 332(6030): 717-721.