Microbial Pathogenesis 105 (2017) 68–73

Contents lists available at ScienceDirect

Microbial Pathogenesis

journal homepage: www.elsevier.com/locate/micpath

Microarray and network-based identification of functional modules and pathways of active tuberculosis

Zhong-Rui Bian [a], Juan Yin [b], Wen Sun [b], Dian-Jie Lin [c, *]

[a] Department of Cardiology, The Second Hospital of Shandong University, Jinan 250033, Shandong Province, China
[b] Beijing Spirallink Medical Research Institute, Beijing 100054, China
[c] Department of Respiratory Medicine, Shandong Provincial Hospital, Jinan 250021, Shandong Province, China

ARTICLE INFO

Article history:
Received 12 December 2016
Received in revised form
7 February 2017
Accepted 7 February 2017
Available online 8 February 2017

Keywords:
Active tuberculosis
Ego genes
Differential expression network
Differential modules
Reactome

ABSTRACT

Diagnose of active tuberculosis (TB) is challenging and treatment response is also difficult to efficiently monitor. The aim of this study was to use an integrated analysis of microarray and network-based method to the samples from publically available datasets to obtain a diagnostic module set and pathways in active TB. Towards this goal, background protein-protein interactions (PPI) network was generated based on global PPI information and gene expression data, following by identification of differential expression network (DEN) from the background PPI network. Then, ego genes were extracted according to the degree features in DEN. Next, module collection was conducted by ego gene expansion based on EgoNet algorithm. After that, differential expression of modules between active TB and controls was evaluated using random permutation test. Finally, biological significance of differential modules was detected by pathways enrichment analysis based on Reactome database, and Fisher's exact test was implemented to extract differential pathways for active TB. Totally, 47 ego genes and 47 candidate modules were identified from the DEN. By setting the cutoff-criteria of gene size >5 and classification accuracy $\geq$0.9, 7 ego modules (Module 4, Module 7, Module 9, Module 19, Module 25, Module 38 and Module 43) were extracted, and all of them had the statistical significance between active TB and controls. Then, Fisher's exact test was conducted to capture differential pathways for active TB. Interestingly, genes in Module 4, Module 25, Module 38, and Module 43 were enriched in the same pathway, formation of a pool of free 40S subunits. Significant pathway for Module 7 and Module 9 was eukaryotic translation termination, and for Module 19 was nonsense mediated decay enhanced by the exon junction complex (EJC). Accordingly, differential modules and pathways might be potential biomarkers for treating active TB, and provide valuable clues for better understanding of molecular mechanism of active TB.

1. Introduction

Tuberculosis (TB) is a threat to global health caused by the intracellular bacterium *Mycobacterium tuberculosis* (M.tb). As reported, there are an approximated 9 million new incident cases of TB annually, with 1.5 million deaths per year [1]. Moreover, about 5–10% of M. tb-infected people develop active TB at some stage in their life, but the other infected individuals remain asymptomatic [2]. Of note, effective drug treatment regimens are hampered by the increasing multidrug-resistant to TB [3]. Thus, early and accurate diagnostics would significantly control the spread of this disease by facilitating early intervention. Nevertheless, detecting TB is a major challenge because the carriers are frequently asymptomatic. Moreover, the exact mechanisms of active TB remain elusive. Thus, a better understanding of the biological processes involved in progression of active TB will contribute toward better intervention measures.

With the advances of high-throughput technologies, these technologies have been applied to explore diagnostic biomarkers and biological processes of human diseases [4], which shed novel insights into the potential mechanisms of active TB. Several studies of active TB on the basis of microarray profiles have indicated guiding principles of its molecular initiation and progression, and

* Corresponding author. Department of Respiratory Medicine, Shandong Provincial Hospital, No. 9677 Jingshi Road, Lixia District, Jinan 250021, Shandong Province, China.
E-mail address: linmountain_power@126.com (D.-J. Lin).

http://dx.doi.org/10.1016/j.micpath.2017.02.012
0882-4010/

these may provide guidance for the investigation of potential molecular biomarkers for the early detection of active TB [5–7]. For example, Guerra et al. [8] have indicated that IL26 is a new candidate biomarker for TB susceptibility. Moreover, Lee et al. [9] have demonstrated that ASUN, DHX29, and PTPRC might be potential biomarkers for the detection of active TB infections. Nevertheless, the overlap is very low for the most significant genes among multiple studies [10]. Moreover, those studies mostly focused on single gene related analysis. Remarkably, genes frequently interact with each other to execute cellular activities [11]. Multiple genes tightly regulated in a specific cellular process can be regarded as a process-specific module [12,13]. Hence, one can measure significant genes and cellular processes for disease using the network strategy, especially protein-protein interaction (PPI) network [14]. Additionally, PPI networks having the ability of identifying functional and structural associations offer a useful framework to analyze biological systems [15,16]. With the development of high throughput technique, the data of large-scale protein interaction is accumulated, but a great many of significant interactions are not tested [17]. Fortunately, to certain degree, this difficulty might be worked out by utilizing sub-networks or modules of the complex network [18]. Significantly, several studies have extracted functional modules from the PPI networks [19,20]. Therefore, it is crucial to find important modules for better understanding of biological events associated with TB.

With the goal of uncovering the underlying mechanisms of active TB, we sought to extract several differential modules via analyzing previously published microarray data from the public database containing the samples of active TB and control. The pathological mechanisms underlying active TB were assessed by means of the following steps: Firstly, differential expression network (DEN) was generated based on protein-protein interactions (PPI) network which was retrieved from the STRING database; Secondly, ego genes were extracted according to the degree feature in the DEN. Thirdly, module collection was conducted by ego gene expansion based on EgoNet algorithm; Fourthly, differential expression of modules between TB and controls was evaluated by random permutation test; and Finally, biological significance of differential modules was detected by pathways enrichment analysis based on Reactome database. The results of our paper might contribute to understanding the pathogenesis of active TB and provide potential bio-signatures for effective therapies of active TB.

2. Material and methods

The flowchart of our work was outlined in Fig. 1. The current analysis was comprised of the following components - data selection (microarray profile, PPI data), construction of DEN, finding differential modules, and module annotation with functional categories based on Reactome database. The detailed description of each component was provided as follows.

2.1. Microarray availability and pre-treatment

Raw data for active TB were recruited from the ArrayExpress server database (accession number: E-GEOD-56153). The microarray profile offered by Ottenhoff et al. [21], included 18 active TB patients, 18 healthy controls, 15 active TB patients after 8 weeks of treatment, and 20 recovered patients after 28 weeks of treatment. In our study, to further explore the molecular mechanisms underlying TB, we only chose 18 TB patients and 18 healthy controls for subsequent analysis. Raw data were pre-treated by MicroArray Suite (MAS) Version 5.0 software (Affymetrix) [22]. After the probe data were mapped to the gene symbols, a total of 17,638 genes were obtained.

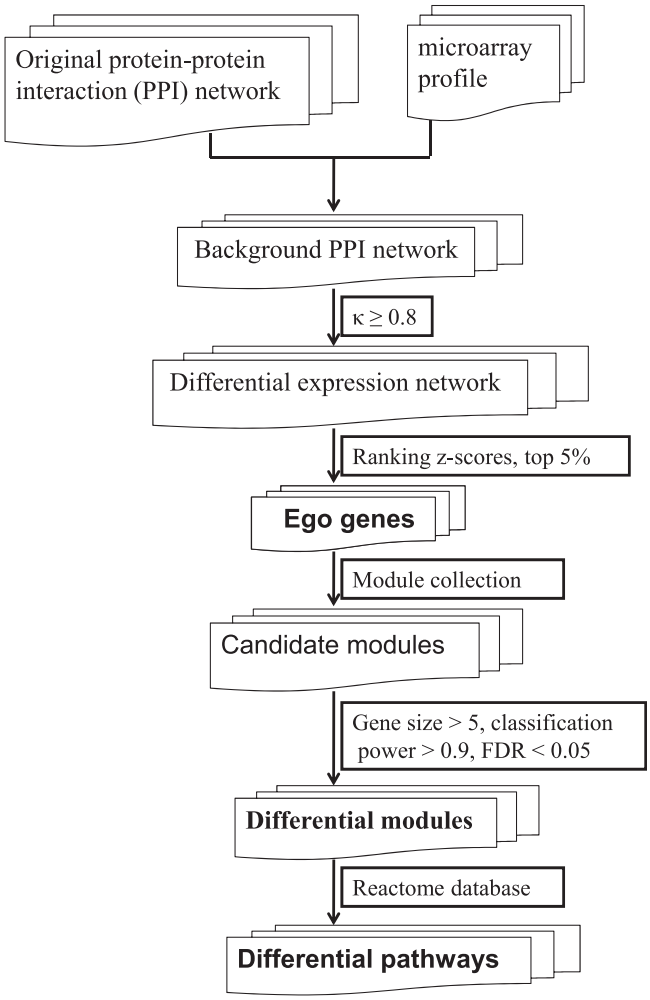

Fig. 1. The scheme flow of our study.

2.2. Construction of DEN and calculation of weight value of each interaction

Begin with, all global PPIs in human covering 787,896 interactions and 16,730 genes were retrieved from the STRING database [23]. Next, all the genes in microarray profile identified above were mapped to the global PPI network to filter the unnecessary interactions. Eventually, 50,355 interactions among 8157 genes were extracted to construct the background PPI network.

Subsequently, pearson correlated coefficient (PCC) was employed to evaluate the interactions in the background PPI network, which was an index to measure the probability of two co-expressed genes. In the current work, we determined the PCC absolute value of an interaction as the predefined threshold κ, and only edges with correlations higher than κ ($\kappa \geq 0.8$) were selected to construct the DEN. Afterwards, a weight value was assigned to each edge in the DEN, which was calculated using one-side *t*-test based on the P values of differential expression in the TB and control samples.

2.3. Finding differential modules

The ego algorithm was designed to detect modules which were