Microbial Pathogenesis 107 (2017) 48–53

Contents lists available at ScienceDirect

Microbial Pathogenesis

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

Systematic expression profiling analysis mines dys-regulated modules in active tuberculosis based on re-weighted protein-protein interaction network and *attract* algorithm

Ying Sun [a], Yan Weng [b, *], Ying Zhang [c], Xiang Yan [d], Lei Guo [a], Jia Wang [a], Xin Song [a], Ying Yuan [a], Fu-Ye Chang [a], Chun-Ling Wang [a]

[a] Department of Cadres' Ward, China Meitan General Hospital, Beijing 100028, China
[b] Department of Gastroenterology, China Meitan General Hospital, Beijing 100028, China
[c] Central Supply Service Department, Jilin Hospital of Integrated Traditional Chinese and Western Medicine, Jilin 132400, Jilin Province, China
[d] Department of Anesthesiology, No 65334 Hospital of PLA, Yanji 133000, Jilin Province, China

ARTICLE INFO

Article history:
Received 30 December 2016
Received in revised form
9 March 2017
Accepted 16 March 2017
Available online 18 March 2017

Keywords:
Active tuberculosis
Protein-protein interaction network
Attractors
Dys-regulated modules

ABSTRACT

About 90% of tuberculosis (TB) patients latently infected with *Mycobacterium tuberculosis* (Mtb) show no symptoms, yet have a 10% chance in lifetime to progress active TB. Nevertheless, current diagnosis approaches need improvement in efficiency and sensitivity. The objective of this work was to detect potential signatures for active TB to further improve the understanding of the biological roles of functional modules involved in this disease. First, targeted networks of active TB and control groups were established via re-weighting protein-protein interaction (PPI) networks using Pearson's correlation coefficient (PCC). Candidate modules were detected from the targeted networks, and the modules with Jaccard score >0.7 were defined as attractors. After that, identification of dys-regulated modules was conducted from the attractors using *attract* method, Subsequently, gene oncology (GO) enrichment analyses were implemented for genes in the dys-regulated modules. We obtained 33 and 65 candidate modules from the targeted networks of control and active TB groups, respectively. Overall, 13 attractors were identified. Using the cut-off criteria of false discovery rate <0.05, there were 4 dys-regulated modules (Module 1, 2, 3, and 4). Based on the GO annotation results, genes in Modules 1, 2 and 4 were only involved in translation. Most genes in Module 1, 2 and 4 were associated with ribosomes. Accordingly, these dys-regulated modules might serve as potential biomarkers of active TB, facilitating the development for a more efficient, and sensitive diagnostic assay for active TB.

1. Introduction

Tuberculosis (TB) is an infectious disease and is caused by mycobacterium tuberculosis (Mtb) which is the most common causative agent [1]. As reported, about 5%–10% of Mtb-infected individuals will progress to active TB at some stage in their lifetime [2]. However, most people infected with Mtb remain asymptomatic, presumably because of protective immune response of host [3,4]. Little is known about the exact potential mechanisms of the transition to active TB [5,6]. Indeed, a bacteriologically confirmed detection of TB is challenging, because direct microscopy has low sensitivity, bacterial culture takes longer time, and sputum samples can be false negative. Thus, developing a sensitive and available approach for early diagnosis of active TB is urgently needed.

Researches about the host transcriptome in response to active TB are increasing through consulting literature [7–10]. For example, *CXCL10*, *ATP10A* and *TLR6* have been demonstrated to be effective at distinguishing active TB from latent TB infection [11]. In another study, *IL-8*, *FOXP3*, and *IL-12β* have been suggested to be the best discriminating signatures for TB [12]. These studies not only detected important genetic signatures indicative of active TB, but also extracted transcriptionally regulated bio-markers which were diverse in functions. Worriedly, many of gene biomarkers on the same disease are typically inconsistent due to varied genetic background of the study subjects or the diversity of the study

* Corresponding author. Department of Gastroenterology, China Meitan General Hospital, No. 29 Xibahe Nanli, Chaoyang District, Beijing 100028, China.
E-mail address: wengyansnow@126.com (Y. Weng).

http://dx.doi.org/10.1016/j.micpath.2017.03.013
0882-4010/

design [13]. Moreover, a list of significant genes identified in a given condition will not offer answers to why an output results. With the goal of further obtaining the mechanistic insights into the causality of these differential regulation, the interactions among the various molecular components in a system ought to be captured and were analyzed as a whole [14]. Networks of protein-protein interactions (PPI) having the ability of extracting functional (genetic) and structural (complex formation) associations provide a useful framework to analyze biological systems [15,16]. Moreover, interaction network has been indicated to be effective to mine subnetworks (modules) [17]. It is noteworthy that protein interactions generated by high-throughput technique are often associated with high false positive and false negative rates. Evaluating the reliability of protein interactions are necessary. Hence, a systems method is essential to capture modules behaviors between different conditions.

However, a limitation of network modeling is that network is static at a given time and it does not readily obtain the dynamics in the cell, and can not by itself interpret cellular responses to active TB. The gap can be addressed by combining gene expression profiles to re-weighted PPI networks and establishing targeted networks for each condition. For example, Magger et al. [18] integrated PPI information and gene expression profile to build condition-specific PPI networks which were utilized to prioritize disease-related genes.

Herein, in an attempt to model the flow of information in response to active TB infection, we identified the dys-regulated modules based on re-weighted PPI networks and gene expression data. We applied Pearson's correlation coefficient (PCC) to re-weight PPI networks to further construct targeted networks of active TB and controls. Next, candidate modules were detected from the targeted networks, and these candidate modules were compared with each other to discover the similar modules using Jaccard similarity model. After that, identification of dys-regulated modules was conducted using *attract* method, following by gene oncology (GO) enrichment analyses for genes in dys-regulated modules. Our study further refined biomarker associated with active TB.

2. Material and methods

In our analysis, using PPI network of homo sapiens as a backbone, two condition-specific PPI networks were inferred, one for active TB and one for healthy control, by integrating microarray expression profiles in these two conditions. Subsequently, fast depth-first method was adopted to extract candidate modules from the targeted networks of two conditions. Afterwards, candidate modules in active TB and control were matched to further extract key ones displaying similar changes in gene composition. Ultimately, identification of dys-regulated modules were done using *attract* method, and genes involved therein were isolated. With the goal of detecting novel genes involved in active TB, we evaluated these dys-regulated module genes for their potential roles in active TB based on GO annotation. The specific description of each step was provided in Fig. 1.

2.1. Microarray acquisition and pre-treatment

Raw data for active TB were downloaded from the ArrayExpress server database through the accession number of E-GEOD-56153 which was provided by Ottenhoff et al. [19]. In the microarray profile of E-GEOD-56153, there were 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. All the TB patients were above 15 years of age, and randomly selected control

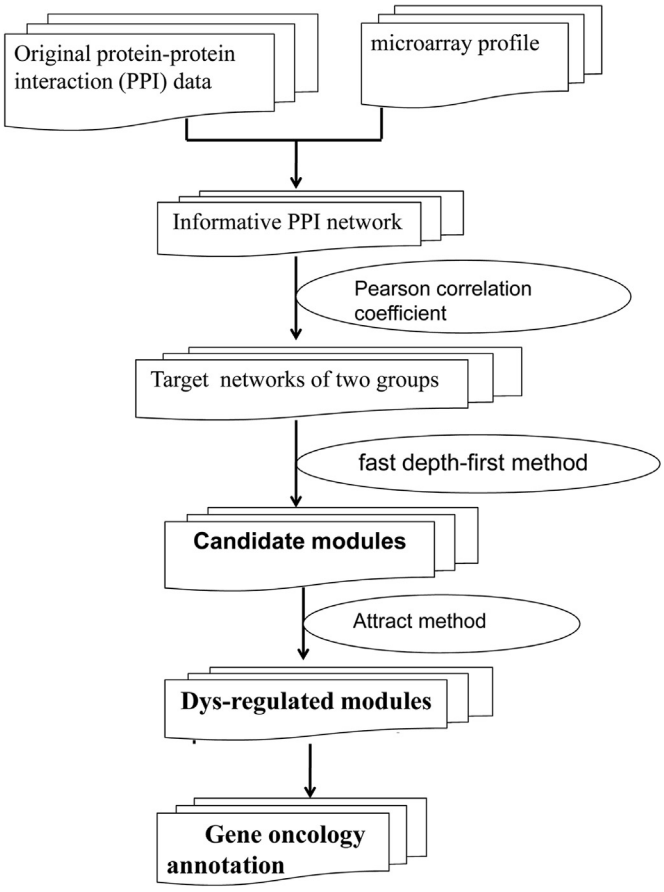

Fig. 1. The flow diagram of our analysis.

subjects had the same sex and age (±10%). In the current work, to further reveal the pathogenic mechanisms of active TB, only active TB patients (n = 18) and 18 healthy controls were selected for following analysis.

Raw microarray data were processed based on Affy package [20] from Bioconductor. In detail, the intensities of the spots were background adjustment through robust multi-array average (RMA) algorithm [21]. Background-adjustment values were normalized using quantiles method to obtain the unbiased data [22]. Then, MicroArray Suite (MAS) Version 5.0 software (Affymetrix) was applied to correct perfectmatch/mismatch values [23], following by expression values summarization using Medianpolish [24]. After that, probe data were mapped to genomics to obtain human gene symbols. Ultimately, a total of 17,638 genes were obtained.

2.2. Construction of informative PPI network

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, http://string-db.org/) database provides a comprehensive, but quality-controlled information between protein-protein associations [25]. STRING database ensembles 3 types protein interactions that were obtained from: mining of databases and the literature; high-throughput experimental data; and predictions based on genomic context analysis. Therefore, the protein interactions provided by STRING offered an integrated scoring scheme with higher confidence.

Herein, all PPIs of homo sapiens with combine-scores (covering 1,048,576 interactions) were derived from STRING database to construct the global PPI network. After discarding the proteins