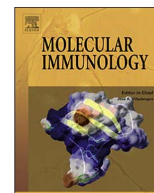
Molecular Immunology 92 (2017) 38–44

Contents lists available at ScienceDirect

Molecular Immunology

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



Identification of potential transcriptomic markers in developing asthma: An integrative analysis of gene expression profiles

Fang Fang*, Jian Pan, Yanhong Li, Yiping Li, Xing Feng, Jian Wang*

Institute of Pediatric Research, Children's Hospital of Soochow University, 92 Zhongnan Street, Suzhou 215025, China

ARTICLE INFO

Keywords:
Transcriptomic markers
Co-expression network
Asthma
Diagnosis
SVM model

ABSTRACT

The goal of this study was to identify potential transcriptomic markers in developing asthma by an integrative analysis of multiple public microarray data sets. Using the R software and bioconductor packages, we performed a statistical analysis to identify differentially expressed (DE) genes in asthma, and further performed functional interpretation (enrichment analysis and co-expression network construction) and classification quality evaluation of the DE genes identified. 3 microarray datasets (192 cases and 91 controls in total) were collected for this analysis. 62 DE genes were identified in asthma, among which 43 genes were up-regulated and 19 genes were down-regulated. The up-regulated gene with the highest Log2 Fold Change (LFC) was *CLCA1* (LFC = 2.81). The down-regulated gene with the highest absolute LFC was *BPIFA1* (LFC = −1.45). Enrichment analysis revealed that those DE genes strongly associated with proteolysis, retina homeostasis, humoral immune response, and salivary secretion. A support vector machine classifier (asthma versus healthy control) was also trained based on DE genes. In conclusion, the consistently DE genes identified in this study are suggested as candidate transcriptomic markers for asthma diagnosis, and provide novel insights into the pathogenesis of asthma.

1. Introduction

Asthma represents a chronic respiratory disorder characterized by airway inflammation, airflow obstruction, and bronchial hyperresponsiveness to stimuli (Mason et al., 2005). Millions of people suffer from asthma all over the world (Subbarao et al., 2009). Airway epithelial cell (AEC) dysfunction plays an important role in asthma (Holgate, 2007; Loxham et al., 2014), hence systematic screening is required to identify AEC abnormalities and improve the diagnosis and treatment of asthma.

Rapid growth of high-throughput transcriptomic data largely enables gene expression profiling and diagnostic targets identification in disease nowadays. In the past decade, several studies have focused on the transcriptional profiling of asthma using microarrays to identify candidate genes involved in asthma (Woodruff et al., 2007; Kicic et al., 2010; Williams-DeVane et al., 2013; George et al., 2015; Voraphani et al., 2014). Analysis of multiple transcriptomic datasets has the likelihood of discovering robust candidates for diagnosis and treatment. Therefore, we investigated gene expression patterns in AEC between asthma patients and healthy controls in this study based on public microarray datasets. The differently expressed (DE) genes identified in this study were further interpreted by enrichment analysis, co-expression network construction and receiver operating characteristic (ROC) curve analysis.

To carry out these analyses, we used the R software (http://www.r-project.org/) and bioconductor packages (Gentleman et al., 2004) for data pre-procession, DE gene identification, ROC analysis, and support vector machine (SVM) model training. Enrichment analysis and co-expression network construction were also performed using DAVID (Huang da et al., 2009a; Huang da et al., 2009b) and Cytoscape (Shannon et al., 2003) software, respectively.

2. Materials and methods

Flow chart of the study is shown in Fig. 1.

2.1. Microarray datasets search and selection

In this study, we searched public microarray datasets till Apr 7, 2016 according to the keywords "asthma" in Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2013). The datasets obtained were further selected for subsequent analysis and our selection criteria were: (a) case-control dataset; (b)

Abbrevations: DE, differentially expressed; LFC, log2 fold change; AEC, airway epithelial cell; ROC, receiver operating characteristic; SVM, support vector machine; GEO, gene expression omnibus; RMA, robust multichip average; FDR, false discovery rate; GO, gene ontology; AUC, area under the curve; CLCA1, chloride channel accessory 1; BPIFA1, BPI fold containing family A member 1

* Corresponding authors.
E-mail addresses: baseff77@hotmail.com (F. Fang), wj196312@vip.163.com (J. Wang).

http://dx.doi.org/10.1016/j.molimm.2017.09.021
Received 25 March 2017; Received in revised form 24 September 2017; Accepted 30 September 2017
0161-5890/ © 2017 Elsevier Ltd. All rights reserved.

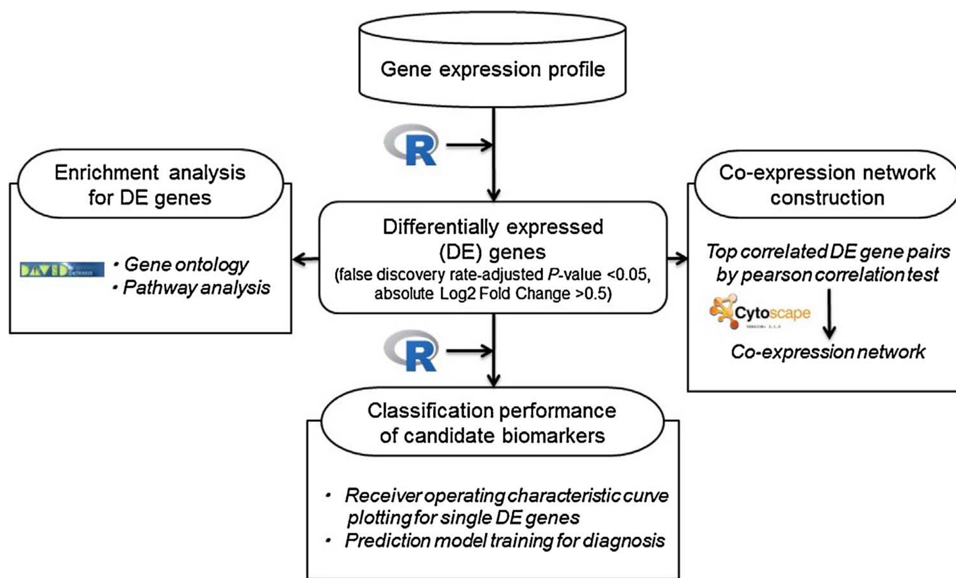


Fig. 1. Flow chart of the study.

dataset using airway epithelial cells for gene expression analysis; (c) dataset providing detailed gene expression data; (d) dataset with sample size larger than 50. Animal studies and studies performing investigations using blood samples were excluded.

Two investigators independently collected data from each eligible dataset. The data were composed of GEO accession, sample size, sample source, platform, and raw gene expression data. Through checking between the two investigators, a final data collection was determined.

2.2. Data analysis methods

According to the data collected from each eligible microarray study, 2 Affymetrix datasets and 1 Agilent dataset were included in this analysis. For the 2 Affymetrix datasets, raw data saved in CEL files of the two microarray datasets downloaded from GEO database were merged, and then pre-processed (background correction, quantile normalization, log2 transformed) using the Robust Multichip Average (RMA) method of the R package "affy" (Gautier et al., 2004). Next, the hybridization probes were mapped to genes (Entrez IDs) according to the platform table. Probes mapping to multiple genes and probes not mapping to genes were excluded. When multiple probes mapped to the same gene, arithmetic mean of probe values was calculated to represent gene expression. The DE genes in asthma compared with healthy controls were identified using the R package "limma" (Smyth, 2005; Ritchie et al., 2015) according to the criteria: (a) false discovery rate (FDR)-adjusted P-value < 0.05 from Wilcoxon test; (b) absolute Log2 Fold Change (LFC) > 0.5.

Functional interpretation (gene ontology (GO) biological process analysis and KEGG pathway analysis) of the DE genes identified according to the 2 Affymetrix datasets was further performed using DAVID 6.8 (Huang da et al., 2009a; Huang da et al., 2009b). In GO analysis, a P value threshold of 0.05 was used to identify significantly enriched GO biological process terms (Falcon and Gentleman, 2007). In pathway analysis, enrichment analysis was carried out using the hypergeometric test with a P value threshold of 0.05 based on the KEGG database (Kanehisa and Goto, 2000). Pearson correlation coefficients were then calculated between DE genes according to their expression levels. The correlation threshold was set to > 0.8 and the eligible correlated gene pairs were selected for a co-expression network construction using Cytoscape 3.1.0 (Shannon et al., 2003). In addition, classification performance of single DE genes were evaluated by ROC curve plotting and area under the curve (AUC) values calculation using the R package "pROC" (Robin et al., 2011). Next, recursive feature selection was performed based on DE genes using the R package "caret" (Kuhn, 2008) and a SVM model for asthma diagnosis was trained based on the selected features using the R package "kernlab" (Karatzoglou et al., 2004) with Gaussian RBF kernel and 10-fold cross-validation.

For the Agilent dataset which was used for cross-platform validation, the pre-processed series matrix file was downloaded from GEO database, and mapped to genes (Entrez IDs) according to the platform table. Features not mapping to genes were excluded. When multiple features mapped to the same gene, arithmetic mean was calculated to represent gene expression. The DE genes were identified using the same criteria aforementioned for the Affymetrix datasets.

3. Results

3.1. Differentially expressed genes in asthma

Original search identified 5 studies in total. Then, 2 studies were excluded among which one study performed investigation with sample size smaller than 50, and another study used blood samples other than airway epithelial cells. Through searching and selection, a final list of 3 microarray datasets (2 datasets from Affymetrix microarrays and 1 dataset from Agilent microarray) (Woodruff et al., 2007; Voraphani et al., 2014) was collected for subsequent analysis. The detailed information of these 3 datasets is presented in Table 1.

For the 2 Affymetrix datasets, pre-processing resulted in expression data of 20,283 genes in 104 asthma samples and 71 healthy controls. According to the criteria (FDR-adjusted P-value < 0.05 from Wilcoxon test, absolute LFC > 0.5), 62 genes in total were identified to be differentially expressed between asthma patients and healthy controls across microarray datasets (Supplementary Table 1). Heat maps of expression data for the 62 DE genes in the 2 datasets used are shown in Fig. 2A and B respectively. Among the 62 DE genes, 43 genes were up-regulated and 19 genes were down-regulated. The top 10 most significantly up-regulated genes and top 10 most significantly down-regulated genes are shown in Table 2. The up-regulated gene with the highest LFC was *CLCA1* (LFC = 2.81). CLCA1, chloride channel accessory 1, is a secreted protein which functions in the activation of endogenous calcium-dependent chloride channel (Yurtsever et al., 2012; Hamann et al., 2009) and is implicated in asthma (Nakanishi et al., 2001). The down-regulated gene with the highest absolute LFC was *BPIFA1* (LFC = −1.45). BPIFA1 (BPI fold containing family A

Download English Version:

https://daneshyari.com/en/article/5591829

Download Persian Version:

https://daneshyari.com/article/5591829

Daneshyari.com