Identification of Over and Under Expressed Genes Mediating Allergic Asthma

Rajat K. De[1] and Anindya Bhattacharya[2]

[1] Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India
rajat@isical.ac.in
[2] Department of Computer Science & Engineering, Netaji Subhash Engineering College, Techno City, Garia, Kolkata 700152, India

Abstract. The present article focuses on identifying some of the genes mediating the development of asthma. Here we apply a pattern recognition based approach to identify the genes those are severely over or under expressed in the allergen samples. The methodology involves clustering on gene expression and fold values followed by determining similarity/dissimilarity among various clusters, and measuring the extent of over/under expression of genes. From this analysis we have identified several genes those have significantly changed their expression values for asthmatic condition, and have reported in the present article. Some of these observations are supported by some earlier investigations. Others have been stayed unnoticed so far, but may play crucial role in mediating the development of asthma.

Keywords: mouse, clustering, fold value, Jaccard score, DB-index.

1 Introduction

Modern genome research, being aided by high-throughput technologies, is producing enormous amount of data that has to be organized, classified and interpreted. In functional genomics, it is increasingly appreciated that various cellular processes are rooted from the dynamic interaction among its many constituents, such as, DNA, RNA, proteins and small molecules. This leads to the emerging of some challenging problems including determination of functions of genes/proteins, interaction among genes/proteins and proteins/proteins, and pathway analysis. Identification of genes whose expression patterns mediate a particular disease is also of great interest.

Asthma is an inflammatory disease that remains poorly understood and hard to control. This disease is characterized by airway hyper reactivity (AHR, defined by exaggerated airflow obstruction in response to bronchoconstrictors), mucus overproduction and chronic eosinophilic inflammation. AHR and mucus overproduction are consistently linked to asthma symptoms and morbidity. Asthma is thought to be mediated by Th2 lymphocytes, which produce a limited repertoire of cytokines, including interleukin-13 (IL-13) [1, 2, 3, 4].

M. Ali and R. Dapoigny (Eds.): IEA/AIE 2006, LNAI 4031, pp. 943–952, 2006.

In this article, we report a new methodology based on clustering for identifying a set of genes mediating the development of asthma. These genes are either over expressed or under expressed in the allergen samples as compared to those in normal ones, and are selected from a large set of genes using the methodology involving Partitioning Around Mediod (PAM) [5] and Fuzzy c-means (FCM) [6] clustering algorithms. The methodology of this selection considers oligonucleotide microarray gene expression data GDS958 [7] dealing with expression patterns of as many as 22690 genes of both normal and allergen samples. Clustering algorithms have been applied on gene expression as well as fold values (ratio of expression values of genes in allergen and normal samples). This is followed by determining similarity/dissimilarity among various clusters and also measuring the extent of over/under expression of genes. A set of genes mediating the development of asthma has been newly identified along with those already reported by some earlier investigations [1, 3, 8].

2 Data Description

The oligonucleotide microarray gene expression data (GDS958) that were used here were obtained from lung tissue of mouse [7]. The data set contains samples which had undergone either of the two types of strain, and are termed as Wild Type mouse and IL-13 Knocked Out mouse samples. Each of these Wild Type and IL-13 Knocked Out mouse samples can be either an Allergen sample or a Control sample. Two different types of treatments were used for both Wild Type strain and IL-13 Knocked Out strain: (i) House Dust Mite (HDM) and (ii) Phosphate Buffered Saline (PBS). Allergen samples were obtained from HDM treated Wild Type strain and IL-13 Knocked Out strain, whereas control samples from PBS treated Wild Type strain and IL-13 Knocked Out strain.

Thus we have four different types of sample: (1) HDM treated Wild Type allergen samples, (2) HDM treated IL 13 Knocked Out allergen samples, (3) PBS treated Wild Type control samples and (4) PBS treated IL 13 Knocked Out control samples. In the present investigation, we considered samples undergone only Wild Type strain, *i.e*, (1) HDM treated Wild Type allergen samples and (2) PBS treated Wild Type control samples.

The data set GDS958 contains expression pattern for 22690 genes obtained from six Wild Type samples. Three samples GSM21415, GSM21418 and GSM 21420 are HDM treated Wild Type allergen samples, whereas GSM21422, GSM 21424 and GSM21426 are PBS treated Wild Type control samples. Further information on this data is available at [7]. It is to be mentioned here that the data set contains expression profiles of cytokines including IL-13, IL-4, IL-5 and their receptors known as key mediators for allergen induced immediate development of asthma. This is due to hyperreactivity of the airway and mucus overproduction in the lung as immediate response for house dust mite allergen. The data set is last updated on December 12, 2004, so it is expected to contain recent information.