Bioinformatics and Biosystems 2007, Vol. 2, No. 2, pp. 71-74

Inferring genetic regulatory networks of the inflammatory bowel disease in human peripheral blood mononuclear cells

Jinki Kim[1], Doheon Lee[2], Gwan-Su Yi[1]

[1]School of Engineering, Information and Communication University

[2]Department of Bio and Brain Engineering, KAIST

Abstract

Cell phenotypes are determined by groups of functionally related genes. Microarray profiling of gene expression provides us response of cellular state to its perturbation. Several methods for uncovering a cellular network show reliable network reconstruction. In this study, we present reconstruction of genetic regulatory network of inflammation bowel disease in human peripheral blood mononuclear cell. The microarray based on Affymetrix Gene Chip Human Genome U133 Array Set HG-U133A is processed and applied network reconstruction algorithm, ARACNe. As a result, we will show that inferred network composed of 450 nodes and 2017 edges is roughly scale-free network and hierarchical organization. The major hub, CCNL2 (cyclin A2), in inferred network is shown to be associated with inflammatory function as well as apoptotic function

Keywords: Gene regulatory network, Microarray, ARACNe, Inflammation bowel disease

Introduction

The cellular immune system is a physiological protection mechanism for maintaining cellular stable state against infection or irritation. Inflammation or cell suicide called apoptosis occurs as a response of immune system depending on the cell type and environmental condition. (Seige. et al., 2006) These cell phenotypes are determined by the expression of thousand of genes and interaction between gene products. These activities are accomplished by well-organized complex network that regulate expression of genes. Thus, to reveal the underlying mechanism in the cellular network, understanding of its network organization is first step to elucidate cell physiology and disease.

Microarray profiling of gene expression gives us large amount of gene expression data. Each sample in microarray is the response of cellular state to environmental or cellular perturbation. It indicates that microarray profiling of gene expression can be a good source to uncover the underlying cellular network structure. Several methodologies were suggested to make a more reliable network based on microarray profiling. For example, bayesian network is a statistical approach to make a genetic regulatory network based on bayes's rule. (Friedman N. et al., 2000) And ARACNe is also statistical approach based on information theory. (Margolin. et al., 2006, Basso. et al.,2005)

In this study, we present reconstruction of genetic regulatory network of inflammation bowel disease in human peripheral blood mononuclear cell. The 127 microarray samples consisting of three subsets such as normal, ulcerative colitis, and Crohn's disease were used as input. (Burczynski et al., 2006) Each sample based on Affymetrix Gene Chip Human Genome U133 Array Set HG-U133A consists of different 22283 genes. In order to eliminate the number of gene providing less significant information, the genes having a less than 1.2 deviation in expression level are selected and removed. Then, ARACNe is applied to these processed microarray data to infer a genetic regulatory network. As a result, we obtained gene regulatory network composed of 450 nodes and 2017 edges. Especially, the inferred network reveals that hub gene CCNL2 taking largest interactions is cyclin A2, which is implicated with tumor cell growth inhibition and apoptosis.

Corresponding Author: Gwan-Su Yi (Email: gsyi@icu.ac. kr)

JK and GSY were supported by Korea Research Foundation Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund) (KRF-2005-003-C00157) and grant No. R01-2005-000-10824-0 from the Basic Research Program of the Korea Science & Engineering Foundation. DL was supported by the Korea Science and Engineering Foundation(KOSEF) through the National Research Lab. Program (No. 2005-01450), and the Korean Systems Biology Research Grant (2005-00343).

Data and Methods

The schematic diagram of a network reconstruction procedure is depicted in Figure 1. It consists of three step processes. In data-preprocessing step, the microarray data sets are first normalized and filtered to remove the uninformative genes. This process includes reduction of dimension by selection of differentially expressed genes. In network reconstruction step, network reconstruction algorithm is applied to processed microarray data sets to make a genetic regulatory network. Finally, in data post-processing step, the reconstructed network is visualized and interpreted to uncover the biological meaning

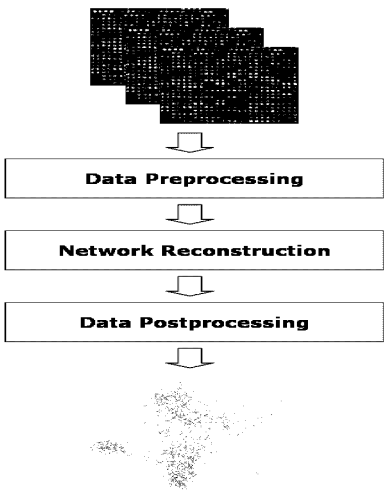

Figure 1. Network Reconstruction Procedure

1. Microarray Data Set

There are several of microarray chip standards used in biological research area. We decided to use Affymetrix Gene Chip Human Genome U133 Array Set HG-U133A and select publically available microarray data set from NCBI's GEO databases. (Barrett. et al., 2005) In order to select biologically and statistically meaningful collection of data, GEO DataSet(GDS) which is curated set of sample data is only retrieved and selected. The microarray data set used in this study is originally aimed at diagnosis of Ulcerative colitis (UC) and Crohn's disease (CD). These two diseases are not distinguished in some patient through standard diagnosis. Thus these data sets are designed to distinguish differentially expressed genes between UC and CD. It is composed of three subsets, which is normal, ulcerative disease, and Crohn's disease. We used these data sets to infer the genetic regulatory network of inflammation bowel disease in human peripheral blood mononuclear cell.

2. Data Preprocessing

Microarray data sets we used in this study are composed of 127 samples and 22283 genes. In order to apply network reconstruction algorithm to microarray data sets and obtain robust result, we should reduce the number of genes through selection of differentially expressed genes. We first apply quantile normalization to raw microarray data to enforce the chips to have identical intensity distribution. The missing values are replaced with mean value of the marker across all microarrays. And then log2 transformation is applied to all the microarray data sets. Finally, we select differentially expressed genes which have more than 1.2 deviations in expression level. It leads to the reduction of dimension from 22283 genes to 1599 genes.

3. Network Reconstruction

We used ARACNe(Algorithm for the Reconstruction of Accurate Cellular Networks) to make a reliable genetic regulatory network from micorarray data sets. It has been reported that ARACNe generates a robust genetic regulatory network if there is more than 100 microarray samples. The ARACNe program in geWorkBench software was used, because this software makes us to process micorarray data more easily and convenience. The parameters in ARACNe program was used as like tutorial recommendation. The P value of 1e-7 for MI estimation and a DPI tolerance of 15% were used. It generates a genetic regulatory network of Inflammation bowel disease in human peripheral blood mononuclear cell, which is composed of 450 nodes and 2017 edges.

4. Data Postprocessing

As a data post-processing methods, we first made a visualization of inferred genetic regulatory network. It provides us to get insight into organization of complex network. The visual representation of the network is performed through Cytoscape software package, which is incorporated into geWorkBench Software. And secondly, the gene taking largest interactions was analyzed in terms of its biological function. Even though the gene is not a transcription factor, we can predict the function of uncharacterized gene in cellular context of a particular microarray data set.

Results

1. Construction of Inflammation bowel disease network

We used ARACNe to reconstruct cellular network from a set of 126 expression profiles of inflammation bowel disease in human peripheral blood mononuclear cell. ARACNe inferred a network with approximately 2000 interactions, which may include both normal and inflammatory cells. We summarized the global connectivity properties in figure 2. The diamond shape in figure 2 represents the experimental plot from inferred network. In order to check the scale free property in inferred cellular network, we plot the theoretical value in square shape according to the power's law, in which parameters are arbitrarily selected to represent the similar behavior of experimental plot. We can regard the inferred network approximately as scale free network with some large difference. The figure 3 depicts visual representation of inferred cellular network through Cytoscape.

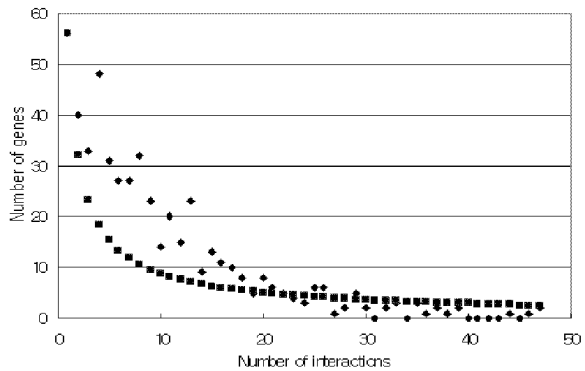

Figure 2. Experimental and hypothetical relation between number of interactions and genes

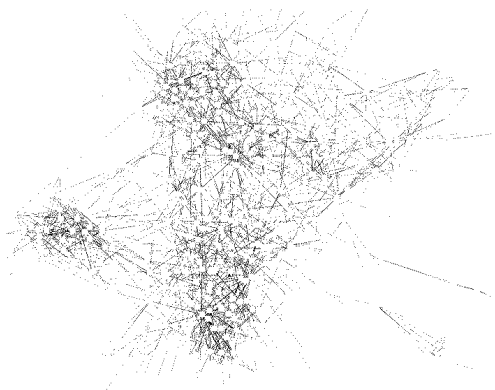

Figure 3. Inferred genetic regulatory network of inflammatory bowel disease

2. Construction of CCNL2(cyclin A2) subnetwork

As expectation of the scale-free nature of the inferred cellular network, a small number of genes explain most of the connections. Genes are prioritized based on number of links connected with it. The top prioritized gene in this inferred network is CCNL2 (cyclin A2) having 42 links with other genes. The visual representation of CCNL2 subnetwork is depicted following Figure 4. The cyclin protein is generally known to be closely associated with cell cycle progression. GO (Gene Ontology) describes that it has properties of a regulation of transcription in Process. It has been also reported that it is a new member of cyclin family, which might regulate the transcription and RNA processing of certain apoptosis-related factors, resulting in tumor cell growth inhibition and apoptosis. (Yang L. et al., 2004) It shows indirectly that inflammation occurs after cell growth inhibition as like apoptosis.

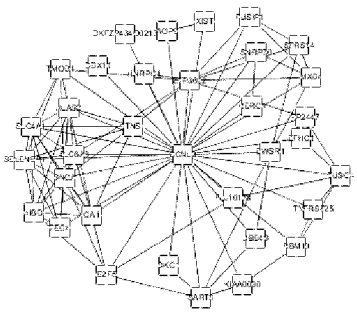

Figure 4. The CCNL2 subnetwork

Conclusion

Reconstruction of genetic regulatory network is the basis of understanding the cellular phenotype and disease. In this study, we presented reconstruction of genetic regulatory network of inflammation bowel disease in human peripheral blood mononuclear cell. The microarray samples based on Affymetrix Gene Chip Human Genome U133 Array Set HG-U133A is processed and applied network reconstruction algorithm. The algorithm ARACNe used in this study generates robust cellular network. As a result of this process, cellular network composed of 450 and 2017 edges is obtained. Overall node degree distribution roughly follows power's law, which has properties of hierarchical organization and scale-free network. The gene CCNL2 having largest adjacent neighbor

reveals that even though it is known to triggers tumor growth inhibition and apoptosis, it has a significant role in inflammation.

This study includes several limitations throughout the overall network reconstruction processes. First of all, this study is performed based on small number of microarray samples. In the ARACNe protocol, at least 100 samples are required to make a reliable cellular network. It means that as the number of sample increase, the reliability of inferred network is more guaranteed. Another weak point is that we made use of mean value to calculate the missing value in microarray sample. The imputation making a more realistic value in missing value calculation may guarantee more reliable cellular network construction. Finally, the parameters in ARACNe software should be adjusted to make a reliable algorithm performance. Because it reveals different algorithm performances depending on the nature of the microarray sample data. We should use the parameters to make a best performance on given microarray data through several of simulations.

References

[1] Seige. et al (2006) Caspase at the crossroads of immune-cell life and death. *Nature Reviews Immunology*, 6(4):308-17

[2] Friedman N. et al (2000) Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7:601~620

[3] Margolin. et al (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7 Suppl 1:S7.

[4] Basso. et al (2005) Reverse engineering of regulatory networks in human B cells. *Nature Genetics*. 37(4):382-390.

[5] Burczynski et al. (2006) Molecular Classification of Crohn's Disease and Ulcerative Colitis Patients Using transcriptional profiles in peripheral blood mononuclear cells. *The Journal of Molecular Diagnosis*, 8(1):51-61

[6] Barrett. et al (2005) NCBI GEO: mining millions of expression profiles--database and tools. *Nucleic Acids Research*. 33(Database issue):D562-6.

[7] Yang L. et al (2004) Cyclin L2, a novel RNA polymerase II-associated cyclin, is involved in pre-mRNA splicing and induces apoptosis of human hepatocellular carcinoma cells. *Journal of Biological Chemistry*, 279(12):11639-48