ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 5, October 2014**

# Computational Method for Reconstruction of Gene Regulatory Network Using Microarray Data

Shanthi Mahesh[1], Dr. Neha Mangla[2], Suhas A Bhyratae[3], J. Sally Josephine[4], Chaithra L[5], Archana B G[6]

Department of ISE, Atria Institute of Technology, Bengaluru, Karnataka, India[1,2,3,4,5]

**ABSTRACT:** The DNA microarray has been established as a tool for efficient collection of mRNA expression data for a large number of genes simultaneously.Mapping function approach maps pairs of genes that present similar positive and/or negative interactions and also specifies how the range of each gene is going to be segmented (labels). From all the label combinations a function transforms each pair of labels into another one, which identifies the type of interaction.

**KEYWORDS:** regulatory network; time series gene expression; gene interaction; topological analysis

## I. INTRODUCTION

The ultimate goal of the genomic revolution understands the genetic causes behind phenotypic characteristics of organisms. Such an understanding would mean having a blueprint which specifies the exact ways in which genetic components, like genes and proteins, interact to make a complex living system. The availability of genome-wide gene expression technologies has made at least a part of this goal closer, that of identifying the interactions between genes in a living system, or gene networks.

Biological networks are the representation of multiple interactions within a cell, a global view intended to help understand how relationships between molecules dictate cellular behaviour. Recent advances in molecular and computational biology have made possible the study of intricate transcriptional regulatory networks that describe gene expression as a function of regulatory inputs specified by interactions between proteins and DNA. Here we review the properties of transcriptional regulatory networks and the rapidly evolving approaches that will enable the elucidation of their structure and dynamic behaviour. Cancer of the esophagus, also known as esophageal cancer, is an uncommon but serious type of cancer that affects the esophagus (gullet). The esophagus is the medical name for the gullet, which is part of the digestive system. The esophagus is the long tube that carries food from the throat to the stomach. The top part of the esophagus lies behind the windpipe (trachea). The bottom part runs down through the chest between the spine and the heart.

**Types of esophageal cancer**
There are two main types of esophageal cancer:

- **Squamous cell carcinoma** forms in the upper part of the esophagus. It occurs when cells on the inside lining of the esophagus multiply abnormally.
- **Adenocarcinoma of the esophagus** forms in the lower part of the esophagus. It occurs when cells inside the mucous glands that line the esophagus multiply abnormally. The mucous glands produce a slimy substance to help food slide down the esophagus more easily.

## II. LITERATURE SURVEY

The present work is accomplished by the researchers in the field of Bioinformatics, SAGE [3] for (i). measuring gene expression, [4,5] (ii).to store this gene expression in structure denominated microarray, make possible the simultaneous study of numerous genes under different conditions. In [6] graph theoretic and statistical techniques

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

were used to identify tight groups of highly similar elements. In [7] a memetic algorithm is presented that is a genetic algorithm combined with local search-based on a tree representation of the data. In [8], a novel type of various genes under a set of samples during a series of time points is explored. Even evolutionary algorithm [9], have been used to discover clusters in gene expression data. In [10], genome-wide gene expression patterns are measured in human monocyte derived dendritic cells (DCs) infected in vitro with seasonal H1N1 influenza A/New Caledonia/20/1999. In [11], the authors have proposed a conditional mutual information estimator based on adaptive partitioning which allows condition and both discrete and continuous random variables. The proposed regulatory network inference algorithm provides better performance when the target network contains co-regulated and interactively regulated genes. In [12], the authors have shown an application of GTR Network on E.Coli gene transcriptome data and the results give a set of potential regulatory links with promising biological significance for isobutanol stress and other conditions. In[13], the authors have found that the network reconstruction accuracy and the values of the hyper parameters inferred with the proposed scheme were found to close to optimal with respect to the minimum reconstruction error dataset description. Evaluation of the performance of our approach which is experimentally tested on the esophageal dataset. The full data set can be downloaded from the Gene Expression Omnibus website: http://www.ncbi.nlm.nih.gov/geo/GSE2144. The dataset has information on 12706 genes under 10 different experimental conditions.

## III. METHODOLOGY

The basic idea of the gene regulatory network involves the processes to
(1) identify the expression level of interacting nodes,
(2) evaluating how interactions change with time (e.g., through a cell cycle or during differentiation),
(3) the phenotypic impact of key nodes.

This way many expressed genes can be extracted from the microarray data of any disease of any organism. The algorithm presented in this approach can be divided into the following (see Figure 1):

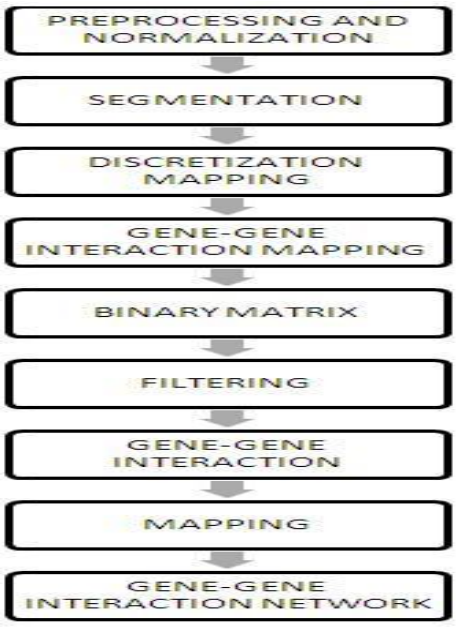


**Figure 1: Design**

The algorithm implements the following functionality:

**Pre-processing and Normalization**

During pre-processing, we try to remove information corresponding to genes that do not show any interesting changes during the experiment, which reduces the size of the data set.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 5, October 2014**

Example: EMPTY, NaN values.

**B. Encoding of each gene expression, *Segmentation:***

Due to the fact that the expression levels are represented by numerical values, the segmentation is done by discretizing the range of values.

**C. Discretization mapping**

Discretization is carried out by defining an symbol Ω, which is used to provide labels for the mapping, and a mapping function α, which is used to assign labels from Ω to the numerical values. The definition of Ω and α is defined by the user, characters for Ω and a discretization mapping table for α, in which the user can also make use of symbol ∞(infinite), μ(mean) and σ(standard deviation).

**D. Representation of the interaction of every two genes, *Gene-Gene interaction Matrix (GGIM)***

Once gene expression level has been labelled, next step is to focus on the interaction between every pair of genes. Table 1 shows the discretization mapping and Table 2 shows GGIM. Initially, another symbol Π is required to assign a label to any possible combination of gene pairs. Π={Z,S,P,N,Q}, S(Similar), P(Positive), N(Negative) and Q(both expressed). The interaction mapping function β is defined as, β: ΩX Ω--□ Π.

**E. Build a binary-valued matrix**

The alphabets P, Q, N are assigned values 1 and S, Z with 0. The alphabets are compared and the values 0/1 are stored in a new matrix. The sum of the P+Q+N of every gene is calculated.

**F. Filtering of most representative interaction, *Filtering:***

The fact that two genes are inhibited under most or all of the experimental conditions has no biological importance. Therefore, this situation can be easily ignored. When two genes are both expressed under most or all of the experimental conditions that might have biological meaning. In fact, this studies focuses on this aspect: the interaction expressed–expressed.

**G. Generate Gene-Gene interaction, *GGI:***

The indices of the most highly interactive pairs of genes is inferred from the filtered expression value array.

**H. Index mapping**

The gene names are mapped from the interaction pairs.

**I. Construct Gene Regulatory Network, GRN**

The result of Gene-Gene interaction matrix is imported into the network visualization and analysis tool, Cytoscape. Cytoscape's roots are in Systems Biology, where it is used for integrating bio molecular interaction networks with high-throughput expression data and other molecular state information.

## IV. EXPERIMENTS AND RESULTS

Regulatory network for 60 genes and 484 interactions is shown in Figure 2. Statistical Analysis of Gene regulatory network has been shown for 10 genes with 18 interactions in Table 1 and Figure 3. The resultant gene-gene interaction matrix is imported into the analyzing tool. The resultant analysis is shown below.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

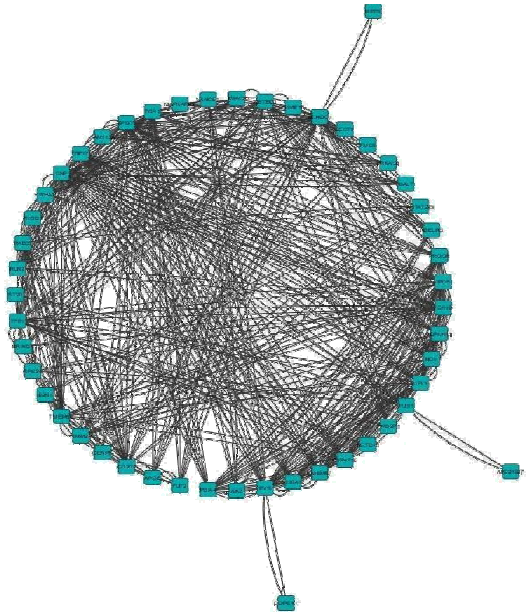

**Figure 2: Regulatory Networks having 60 genes and 484 interactions**

| | |
|---|---|
| Clustering coefficient | 0.9 |
| Connected components | 1 |
| Network diameter | 2 |
| Network radius | 1 |
| Network centralization | 0.600 |
| Shortest paths | 30(100%) |
| Characteristic path length | 1.4 |
| Average no of neighbors | 30 |
| Number of nodes | 6 |
| Number of interactions | 18 |
| Network density | 0.6 |
| Network heterogeneity | 0.33 |
| Isolated nodes | 0 |
| Number of self-loops | 0 |
| Analysis time(sec) | 0.33 |

**Figure 4: Statistical Analysis of GRN**

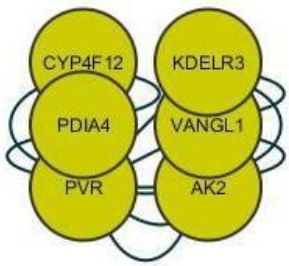


**Figure 3: Regulatory Networks having 10 genes and 18 interactions**

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 5, October 2014**

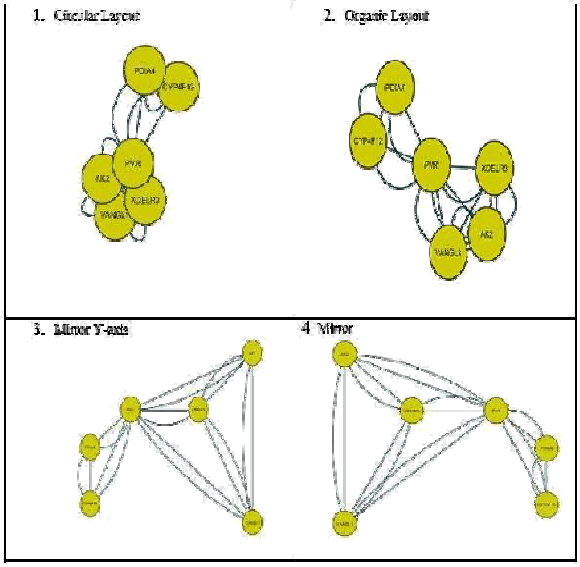


**Figure 4: Topological representation of the gene-gene interaction network for 10 genes**

## V. CONCLUSION

In this work a novel approach comprising the three main features viz, Discretization mapping function, gene-gene mapping function and filtering function has been used on the yeast data to compute regulatory relationships between gene-pairs and topological analysis of reconstructed network. The microarray data consider here consists of 12706 genes fewer than 10 different experimental conditions. Our study yields four major outcomes; first we identified differential expressed genes in dataset. Second, the interactions between differential expressed genes have been identified. Third, genes regulating most of the other genes (hubs) where identified. Result The utility and reliability of the result obtained are discussed and experimental validation can be done. Fourth, provides the topology analysis of reconstructed network revealed a large number of interactions in the oesophageal data. Finally the results of the present investigation provide and excelled understanding of the interaction mechanism of the oesophageal data and provide new insight into the biomedical world. This work is first of its kind in the literature.

## REFERENCES

[1] J.M. Bower and H. Bolouri, editors. Computational modeling of genetic and bio-chemical networks. MIT Press, 2001
[2] H. de Jong. Modeling and simulation of genetic regulatory systems: a literature review. J. Comp. Bio. , 9(1):67–103, 2002.
[3]Velculescu, V.E. Characterization of the yeast transcriptome. Cell, 88, 243-251,1997.
[3] Schena, M. (1996) Genome analysis with gene expression microarray. Bioessaya, 18, 427-431, 1996.
[4] Lipshutz,R.J High density synthesis oligonucleotide arrays, Nature Genetics Supplement, 21, 20-24, 2000.
[5] R Sharan and R Shamir CLICK: A clustering Algorithm for Gene Expression Analysis Proc Int Conf Intell Syst Mol Biol, 8, 307-316, 2000.
[6] N. Speer and P. Merz and C. Spieth and A. Zell Clustering Gene Expression Data with Memetic Algorithms based on Minimum Spanning Trees IEEE Press, 3, 1848-1855, 2003.
[7] Daxin Jiang, Jian Pei, Murali Ramanathan, Chung Tang, and Aidong Zhang Mingin Coherent Gene Clusters from Gene-Sample-Time Microarray Data KDD, 430-439, 2004.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 5, October 2014**

[9] Patrick C.H. Ma, and Keith C.C. Chan Discovering Clusters in Gene ExpressionData using Evolutionary Approach ICTAI, 459-466, 2003.
[10] Elena Zaslavsky1*, German Nudelman1, Susanna Marquez2,4, Uri Hershberg2,5, Boris M Hartmann1, Juilee Thakar2, Stuart C Sealfon1, Steven H Kleinstein2,3*. Reconstruction of regulatory networks through temporal enrichment profiling and its application to H1N1 influenza viral infection. 10th International Conference on Artificial Immune Systems (ICARIS) Cambridge, UK. 18-21 July 2011.
[11 ]Kuo-Ching Liang and Xiaodong Wang, Department of Electrical Engineering,
Columbia University, New York. "Gene Regulatory Network Reconstruction using Conditional Mutual Information", EURASIP Journal on Bioinformatics and System Biology, Vol. 2008, Article ID 253894, 14 pages.
[11] Yao Fu, Laura R Jarboe and Julie A Dickerson. "Reconstructing genome-wide regulatory network of E.coli using transcriptome data and predicted transcription factor activities", BMC Bioinformatics 2011, 12:233.
[12] Adriano V. Werhli and Dirk Husmeier. "Reconstructing Gene Regulatory Networks with Bayesian Networks by combining Expression Data with Multiple Sources of Prioi knowledge". Statistical Applications in Genetics and Molecular Biology, vol 6, Issue 1, 2007.
[13] Khalid Raza and Rafat Parveen, "Reconstruction of gene regulatory network of colon cancer using information theoretic approach", 4th INTERNATIONAL CONFERENCE, CONFLUENCE 2013.
[14] Khalid Raza and Rajni Jaiswal, " Reconstruction and Analysis of Cancer-specific Gene Regulatory Networks from Gene Expression Profiles", International Journal on Bioinformatics & Biosciences, Vol. 3, No. 2, pp. 25-34, 2013. DOI: 10.5121/ijbb.2013.3203.