29th NATIONAL RADIO SCIENCE CONFERENCE
(NRSC 2012)
April 10 - 12, 2012, Faculty of Engineering/Cairo University, Egypt

K4. Gene Network Construction and Pathways Analysis for High Throughput Microarrays

Nagwan M. Abdel Samee[1], Nahed H. Solouma [2], Yasser M. Kadah [3]
[1]Computer Engineering Department, Misr University for Science and Technology, Egypt
[2]Engineering Applications Department, NILES, Cairo University, Giza, Egypt
[3]Systems and Biomedical Engineering Department, Cairo University, Giza, Egypt

ABSTRACT

The key idea discussed in this paper is to infer gene regulatory network from high throughput microarray data for Hepatocellular Carcinoma (HCC). Working with such huge number of genes is a complex process. So, our framework for inferring gene interactions from large scale microarrays is based on a selected set of informative genes. We applied two measures of dependencies between genes: Correlation and mutual information. Therefore, two types of networks were constructed: Co-expression network and Mutual information network. Some Mutual information network inference algorithms: Context Likelihood of Relatedness (CLR), Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE), and Minimum Redundancy Network (MRNET) were applied. A proposed method for simplifying the complex structure of the inferred network is introduced using the Minimum Spanning Tree (MST) which provides a better visual interpretation of the constructed networks. From the constructed networks we were able to identify a set of functional gene modules. These modules were validated using the Gene Ontology (GO) enrichment. The GO enrichment analysis has proven the strength of the ARACNE inference algorithm over all other employed algorithms. Moreover, a comparison was carried out between the Mutual information network inference and the well known Bayesian inference. To establish this comparison, specific pathways in HCC were rather chosen. These pathways were tested for their significance using singular value decomposition. According to this comparison, again the ARACNE showed better results.

Keywords: *Gene Co-expression Network, Mutual Information Network, Bayesian Network, Pathways Analysis, and Singular Value Decomposition.*

I. INTRODUCTION

One of the hot topics in the area of bioinformatics is functional genomics. This topic focuses on the interactions and functions of each gene and its products (mRNA, protein) through the whole genome. In order to identify the functions of certain gene, we should be able to capture the associated gene expressions. Functional genomics uses microarray technology to measure the genes expressions levels under certain conditions. Using microarrays, expression levels of thousands or tens of thousands of genes can be measured simultaneously[1]. Consequently, microarray technology offers a powerful tool for studying gene profiles under certain conditions and constructing the functional relationships among genes. The gene network inference allows us to model the dependencies between genes of a dataset in a graph. The exact semantics of an arc in the graph usually represent a regulator/regulated gene interaction where the genes are represented by nodes in the graph. The latter graph is called a Gene Regulatory Network. Various network inference methods have been proposed in the literature such as, Bayesian network that represents the probabilistic dependencies between variables [2]-[3], Correlation network (Co-expression network) that sets an arc between two genes if it exhibits a high score based on a correlation measure [4], and Mutual information network (Relevance network) that sets an arc between two genes if it exhibits a high score based on pair wise mutual information [5]-[6].

Bayesian network theory connects causality and probability theory. While a variety of computational methods have been considered for reconstructing gene networks from observational gene expression data, Bayesian network (BN) based approaches have shown great promise to infer causal relationships between genes [7]. Bayesian network structure learning algorithms can be grouped in two categories:

- Constraint-based algorithms: These algorithms learn the network structure by analyzing the probabilistic relations entailed by the Markov property of Bayesian networks with conditional independence tests and then constructing a graph
- Score-based algorithms: These algorithms assign a score to each candidate Bayesian network and try to maximize it with some heuristic search algorithm.

The expression profiles of genes across several microarray samples can be highly correlated [4] and it is natural to represent the pair wise relationships between genes by a network connecting them. Genes with similar profiles may come from the same complex or pathway and so they may have a similar function [8]-[10]. Gene Co-expression network is suitable for representing such types of relationships. Gene Co-expression has been used in several publications [11]-[12] as a method of representing the transcriptome in many organisms: yeast, flies, worms, plants, and humans. Gene Co-expression network is an undirected graph, where the graph nodes correspond to genes and edges between nodes represent co-expression relationships. Compared to GRN, a Gene Co-expression network does not attempt to provide causal relationships between genes however a Gene Co-expression network contains gene neighborhood relationships[1]. Clustering of genes according to their expression similarities should yield modules of genes with similar profiles. However, clustering procedures are often subjective, and usually ignore the detailed relationships among genes; the biological insight obtained from clustering results is often limited [4]. Gene Co-expression network is based on the pair wise correlation between genes. An edge is drawn between two nodes, genes, if the corresponding pair wise correlation exceeds a specific threshold. In this manner, we will have an adjacency matrix, which is a measure of the connection strength between all pairs of genes. The entries of this matrix are either 0 or 1. If the value is 1, this indicates the existence of a connection between a pair of genes. Otherwise, no connection exists.

Mutual information network is a subcategory of inference approaches which is based on the mutual information. The manner of inferring the Mutual information network is similar to the Co-expression network inference. Both produce an undirected graph and are easy to be implemented. However, in the former network, a link between two nodes is set if it exhibits a high score based on pair wise mutual information instead of the correlation measurement. Mutual information is an information-theoretic measure of dependencies.

To conclude, there are basically three gene network inference methods: Bayesian network, Gene Co-expression network and Mutual information network. The Bayesian network inference is a powerful approach of estimating the causal relationships between genes, but it is computationally too expensive for dealing with large number of genes and it may yield inaccurate results in case of having insufficient number of samples[4] which is the nature of microarrays data sets. Therefore, in this work the Bayesian inference has not been used in inferring the whole GRN. Instead, it has been used in the inference of a part of predefined pathways of HCC to measure its performance compared to the other inference methods. On the other side, there are some advantages of Mutual information over the correlation as a measure of similarity between two variables. The Mutual information captures the non linear dependencies, an interesting feature in biology where biological interactions are believed to be non linear [13]-[14]. However, the correlation is a measure of linear relationship between two variables. The mutual information is rather fast to compute [15]. This motivated us to use the Mutual information network inference in estimating large networks from high density microarrays. The retrieved network can reveal modules with a better inner relationship than those using the Co-expression network inference. This will be illustrated through a fair comparison between such types of network inference.

II. Materials and Methods

The aim of this work is to study microarrays of Hepatocellular Carcinoma to estimate and analyze its gene networks and pathways. Hepatocellular Carcinoma is a major type of liver cancer. It rises as a consequence of underlying liver diseases such as viral hepatitis and liver cirrhosis. Hepatitis B virus (HBV), hepatitis C virus (HCV) and intakes of alcohol are widely recognized as the three major etiological factors of HCC. But, HCV is a predominant cause of HCC. RNA expression data for liver samples from subjects with HCC as a complication of HCV cirrhosis were used in this study. Thirty five microarray samples were downloaded from Gene Expression Omnibus (GEO). Nineteen of theses samples are taken from normal subjects. The remaining sixteen samples are for subjects with HCC as a complication of HCV cirrhosis. This data were collected on the Affymetrix HG-U133A 2.0 platform. The raw data in ".CEL" format was collected from GEO and an up-to-date probe set definition (.CDF file) based on ENTREZ Gene sequence was used in place of the Affymetrix original probe set definition. The raw data were pre-processed using the RMA algorithm which is implemented in the Affy package provided by Bioconductor. When the RMA preprocessing algorithm is applied to this data set, we get a gene expression matrix which contains the activity level of each gene across the thirty five samples. The number of rows of this matrix corresponds to the number of genes which is 22277 genes. The dimension of this matrix was reduced through a filtration of insignificant genes[16]. This paves the way to infer the gene network on significant genes only.

The proposed frame work of inferring gene network from high throughput microarrays is shown in figure 1. There are five basic steps in this frame work: preprocessing of microarrays, selection of informative genes, determining the dependencies between genes, network inference, and visualization. The preprocessing and gene filtration have been discussed in[16] and it yielded only172 significant genes. Therefore, in this work we constructed the network on these genes. Therefore, the carried out research was conducted to cover the other three modules in this framework. There are two common methods for determining the dependencies between genes: Correlation coefficients and Mutual information.

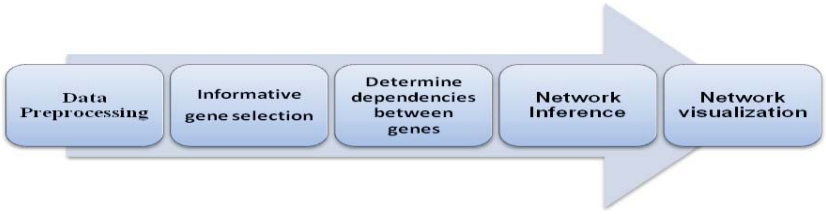

Fig. 1: The proposed frame work of inferring a gene network from high throughput microarrays.

II.1.1 Pearson's Correlation Coefficient

The correlation coefficient measures the strength of a linear relationship between two variables therefore; it is used as a similarity measure between two genes expression profiles. The correlation coefficient is always between -1 and +1. The closer the correlation is to +/-1, the closer to a perfect linear relationship. A strong association between genes should have Pearson correlation coefficient values from ±0.7 to ±1. The Pearson's correlation coefficient between two expression profiles X and Y is defined as in equation 1.

$$r = \frac{\sum_{i=1}^{Q}(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{Q}(Y_i - \bar{Y})^2}\sqrt{\sum_{i=1}^{Q}(X_i - \bar{X})^2}} \tag{1}$$

II.1.2 Mutual Information

The Mutual information measures the amount of information that can be obtained about one random variable by observing another one[17]. Formally, the mutual information of two discrete random variables X and Y can be defined by equation 2. Intuitively, mutual information measures the information that X and Y share: it measures how much knowing one of these variables reduces our uncertainty about the other. For example, if X and Y are independent, then knowing X does not give any information about Y and vice versa, so their mutual information is zero. On the other hand, if X and Y are identical then all information conveyed by X is shared with Y: knowing X determines the value of Y and vice versa. Therefore, Mutual information can be equivalently expressed by the equation 3.

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\frac{p(x,y)}{p_1(x)p_2(y)} \tag{2}$$

$$\left.\begin{aligned} I(X;Y) &= H(X) - H(X\mid Y) \\ &= H(Y) - H(Y\mid X) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}\right\} \tag{3}$$

Intuitively, if entropy H(X) is regarded as a measure of uncertainty about a random variable then $H(X \mid Y)$ is a measure of what Y does not say about X. This is "the amount of uncertainty remaining about X after Y is known", and thus the right side of the first of these equalities can be read as "the amount of uncertainty in X, minus the amount of uncertainty in X which remains after Y is known", which is equivalent to "the amount of uncertainty in X which is removed by knowing Y". This corroborates the intuitive meaning of mutual information as the amount of information (that is, reduction in uncertainty) that knowing either variable provides about the other.

II.1.3 Gene Co-expression Network

In figure 2, we show the basic steps of inferring a gene Co-expression network for the selected set of significant genes. The prerequisite step of constructing a Gene Co-expression network is the estimation of a pair wise correlation between the selected set of informative genes according to their expression values. This similarity matrix (correlation coefficient matrix) is examined to produce a graph matrix (adjacency matrix). An edge is drawn between two nodes, genes, if the corresponding Pearson's correlation coefficient between exceeds a specific threshold (±0.7 to ±1 for a strong association between genes). This check will yield an adjacency matrix, which has entries of either 0 or 1 (a value of 1 indicates the existence of a connection between a pair of genes and the value of zero means there is no connection exists). By this structure of the adjacency matrix, it can be feed to a graph drawing algorithm which will produce a graph representing the relationships between genes. The igraph library is R package that provides data types and functions for implementing different algorithms of graph drawing. The package also provides simple function for graph plotting and visualization.

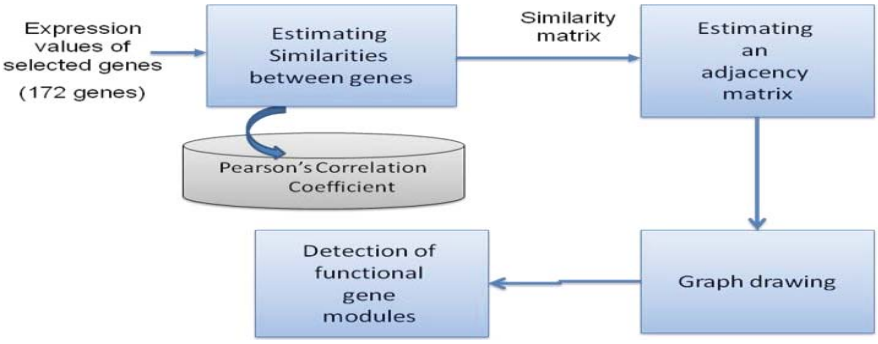

Fig. 2: Gene co-expression network construction and visualization.

II.1.4 Mutual Information Network

Mutual information network inference proceeds in two steps. The first step is the computation of the mutual information matrix (MIM). The second step is the computation of an edge score for each pair of nodes by an inference algorithm that takes the MIM matrix as input. Three inference methods [17]: CLR, ARACNE and MRNET will be reviewed in this section as a most commonly used inference methods.

II.1.4.1 CLR

This algorithm [6] extends the relevance network (RELNET) algorithm[18] by computing a score for each pair of genes according to the empirical distribution of the mutual information values as shown in figure 3. Instead of considering the score as the mutual information $I(X_i, X_j)$ between genes X_i and X_j , it will be $W_{ij} = \sqrt{z_i^2 + z_j^2}$ where z_i and z_j are the mapped scores of the mutual information values and can be calculated using equation 4.

$$z_i = \max(0, \frac{I(X_i, X_j) - \mu_i}{\sigma_i}) \tag{4}$$

In equation 4, μ_i and σ_j are respectively the mean and standard deviation of the empirical distribution of the mutual information values $I(X_i, X_k)$ where k =1 to n (number of genes).

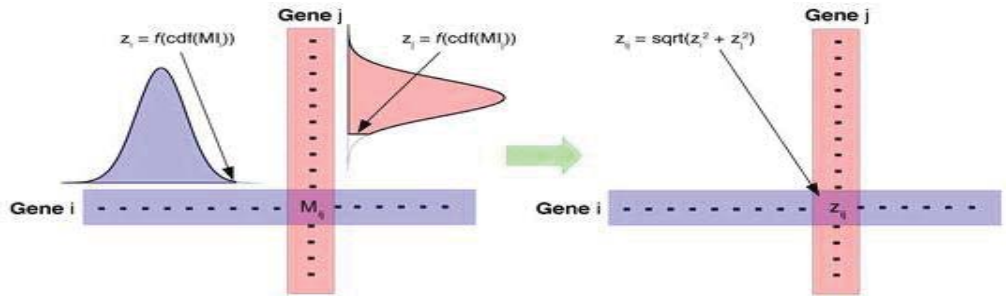

Fig 3: The basic principal of the CLR algorithm.

II.1.4.2 ARACNE

The ARACNE stands for Algorithm for the Reconstruction of Accurate Cellular Networks [7]. Its theorem states that if gene X_1 interacts with gene X_3 through X_2 then $I(X_1, X_3) \leq \min(I(X_1, X_2), I(X_2, X_3))$. The link between X_1 and X_3 is considered a weak edge for this triplet. ARACNE consists of three steps: it begins by assigning to each pair of nodes a weight equal to their mutual information. Then, as in RELNET, all edges that have mutual information $I(X_i, X_j)$ less than a given threshold are removed. Finally, the weakest edge of each triplet is interpreted as an indirect interaction and is removed. This algorithm can yield optimum performance if the network is a tree including only pair wise interactions, then this method should guarantee the reconstruction of the original network, once it is provided with the exact MIM[7]. However, it has a higher complexity than the first two methods: RELNET and CLR since it considers all triplets of genes.

II.1.4.3 MRNET

The network is inferred using the maximum relevance/minimum redundancy (MRMR) feature selection method. The MRMR method has been introduced in [19] as a feature filtration in a supervised learning problems. The basic idea of this algorithm is that direct interactions (i.e. the most informative variables to the target gene) should be well ranked whereas indirect interactions should be badly ranked. Consider a supervised learning task where the output is denoted by Y and V is the set of input variables. The method assigns each input variable of V a score which is the difference between the mutual information with the output variable Y (*maximum relevance*) and the average mutual information with the previously ranked variables (*minimum redundancy*). The method starts by selecting the variable X_i having the highest mutual information to the target Y. The second selected variable X_j will be the one that has the highest mutual information $I(X_i, Y)$ to the target and at the same time it has the lowest mutual information $I(X_i, X_j)$ to the previously selected variable.

II.1.5 Gene Pathway Analysis

To establish a comparison between all of the network inference methods mentioned above, we need to compare the inferred network with the true structure. However, this test is not possible for real microarrays as the true network structure is unknown. This motivated us for developing a new method of validation rather than the traditional validation method which uses synthetic data sets. We used a predefined set of genes (pathway) that has a predefined structure in testing those methods of network inference. A pathway is a predefined set of genes that are known in advance to be related in regulating a specific phenotype. Our biological interpretation of the top ten selected genes [16] yielded one significant KEGG pathway for Hepatocellular carcinoma: TGF-beta signalling pathway (TGF). We used another significant pathway for HCC from the biological literature, P53 signalling pathway (TP53). We analyzed these pathways to assess their significance in Hepatocellular carcinoma before being used in our comparison. A pathway analysis focuses on the changes in expression of group of genes in a pathway. A pathway activity level is the level of activity of a pathway in different samples. We used the Singular Value Decomposition (SVD) to measure the pathway activity level [20] as shown in figure 4. The SVD is applied on genes in a pathway. The activity level of a pathway in a given sample could be defined as the 'level' of expression of a certain metagene in that sample, i.e., the level of the first metagene from the SVD of the matrix of expression levels. As shown in figure 4, the columns of the matrix W are the orthonormal eigenvectors or metagenes of Y, D is a diagonal matrix containing the associated eigenvalues, and each column of C (each row of C')is a vector of coefficients for one of the samples indicating the level of each metagene in the sample.

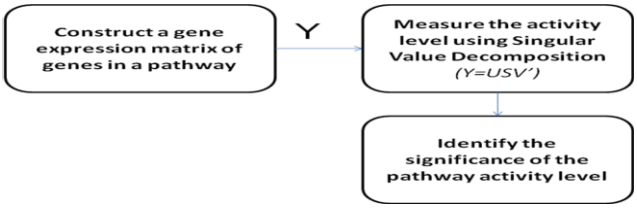

Fig. 4(a): Gene pathway analysis using Singular value decomposition.

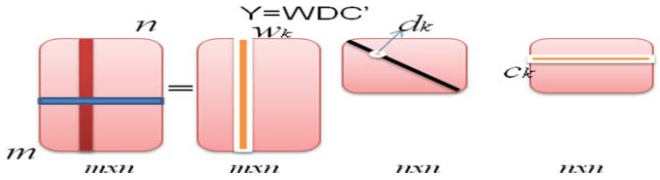

Fig. 4(b): Applying the Singular value decomposition on a gene expression matrix.

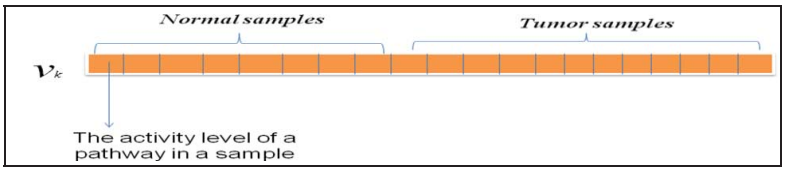

Fig. 4(c): The activity level of a pathway in a given sample j can be taken as the coefficient cj for the first metagene.

III. RESULTS

III.1.1 GENE NETWORK

For each gene network inference method, a gene network was constructed for the 172 significant genes. One Gene Co-expression network that is based on the estimation of the Pearson's correlation coefficient between genes is shown in figure 5 and three Mutual Information networks that were constructed using CLR, ARACNE and MRNET are shown in figure 6. Both of the inferred Gene Co-expression network and Mutual Information network should reveal functional modules of HCC. However, as shown in these figures, we cannot visually divide the network into modules due to the complex structure of these networks (9715 edges was detected using CLR, 9628 edges was detected using MRNET and 349 edges was detected using ARACNE). This complex structure was simplified by a divisive hierarchal clustering: Minimum Spanning Tree (MST). The MST has no loops and yields a set of divisible clusters. In figures 7 and 8, we drew MST for the constructed networks in figures 5 and 6. Functional modules in these trees were selected visually and marked by red ovals. The enrichment of each gene set (module) was tested using the GO term enrichment. There have been quite many tools to find enriched GO terms. Gene Ontology Enrichment Analysis Software Toolkit (GOEAST), a web-based software toolkit for discovering statistically enriched GO terms among a given gene list, was used to assess the enrichment of the retrieved modules from each inference method. The GO enrichment analysis lists the ontology (Biological function, Cellular component, or molecular function) of each Go term along with its p-value and the numbers of Affy IDs that have this GO term.

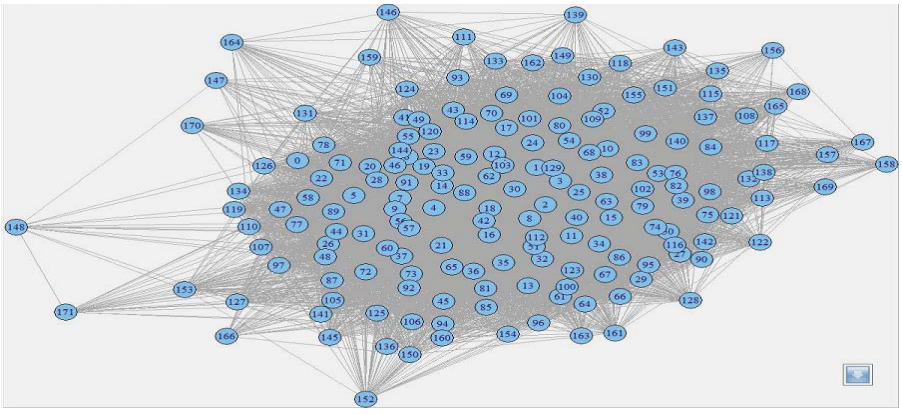

Fig. 5: Gene Co-expression network for 172 significant genes.

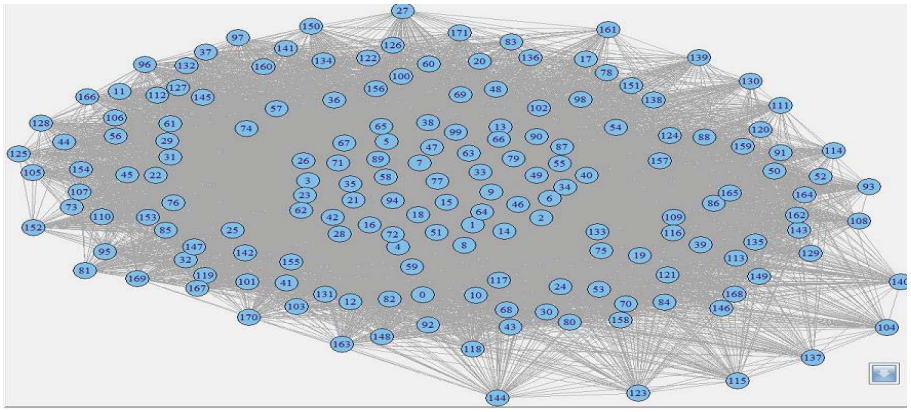

Fig. 6(a): The Mutual information network that constructed using CLR inference algorithm.

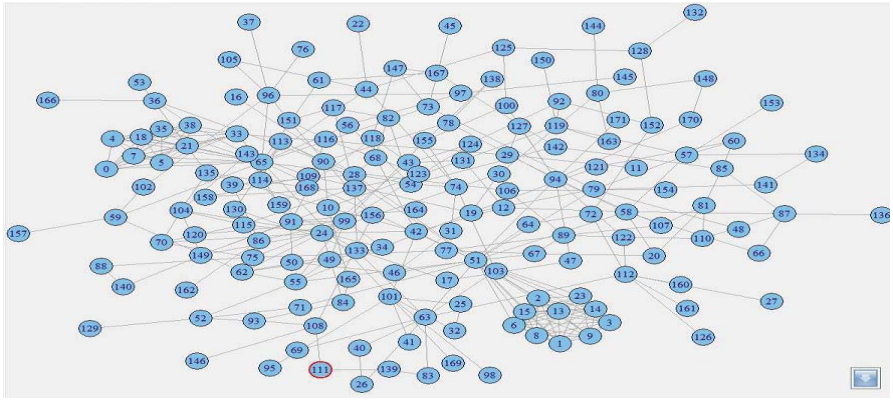

Fig. 6(b): The Mutual information network that constructed using ARACNE inference algorithm.

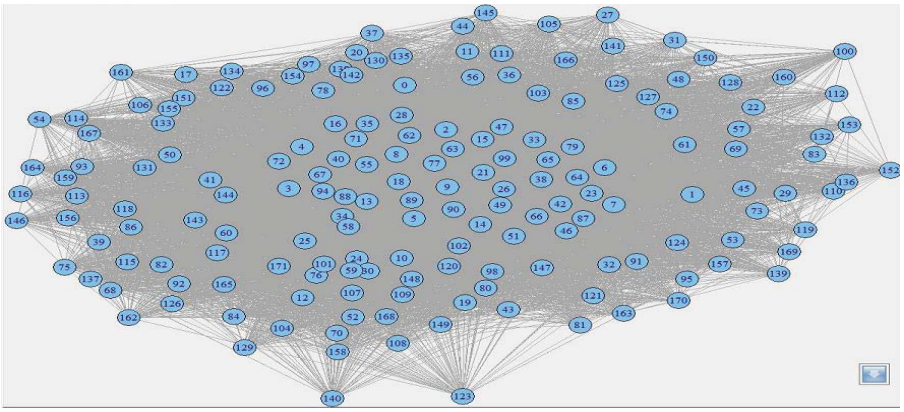

Fig. 6(c): The Mutual information network that constructed using MRNET inference algorithm.

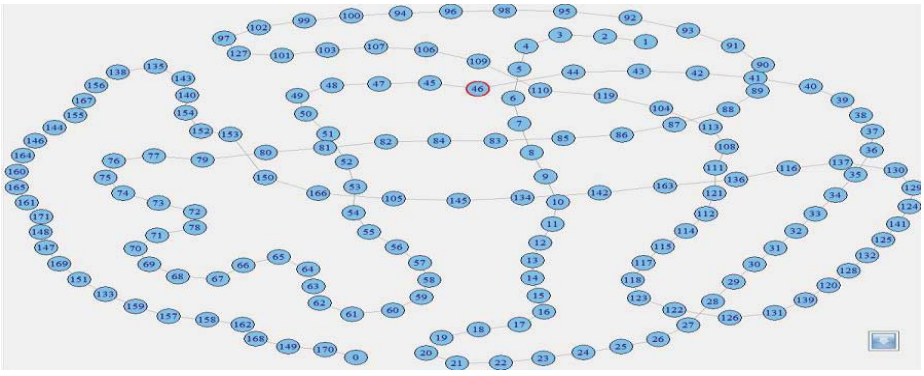

Fig. 7: Minimum Spanning tree for Gene Co-expression network.

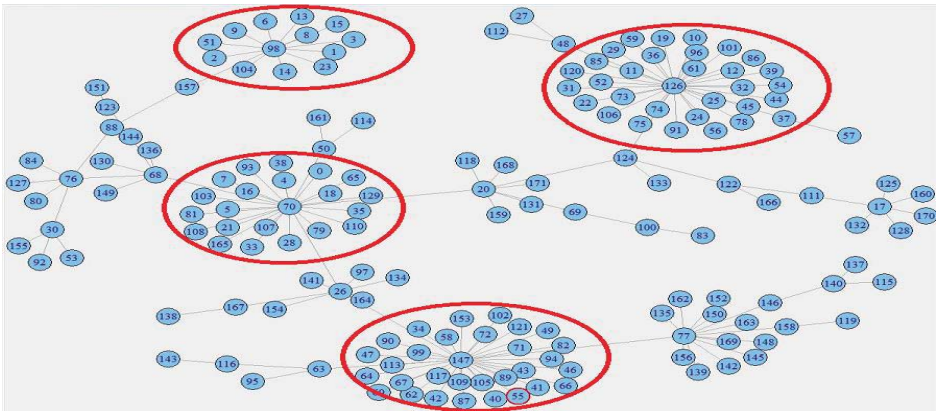

Fig. 8(a): Minimum Spanning tree for the Gene network inferred using CLR.

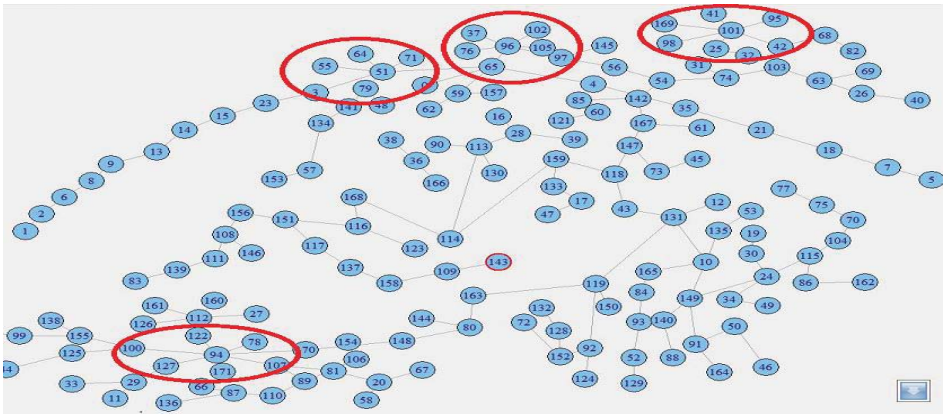

Fig. 8(b): Minimum Spanning tree for the Gene network inferred using ARACNE.

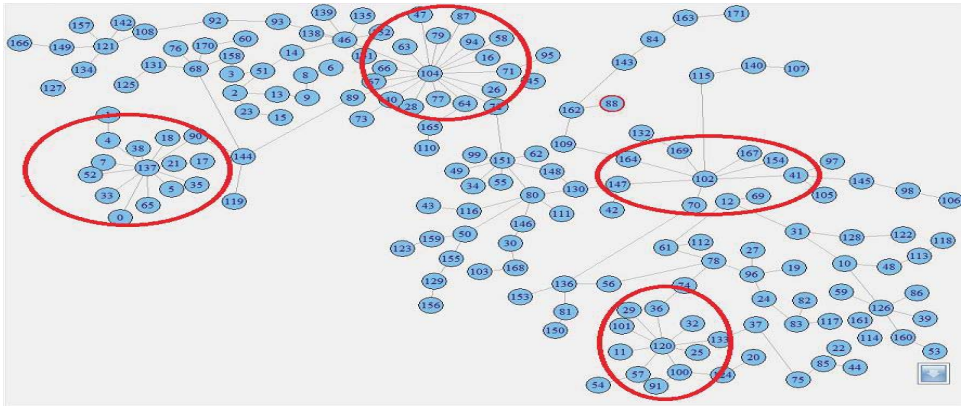

Fig. 8(c): Minimum Spanning tree for the Gene network inferred using MRNET.

III.1.2 GENE PATHWAY ANALYSIS

The signal profiles of the activity levels of the TGF and P53 pathways in our data set, HCC microarrays, are shown in figure 9. These figures reveal that genes in these pathways are differentially expressed in different groups of samples (normal and tumour samples of HCC). The significance of these pathways is determined using t-test (t = -14.5422 and p-value = 1.013e-13 for TGF pathway, t = -10.0829 and p-value = 6.371e-11 for P53 pathway).

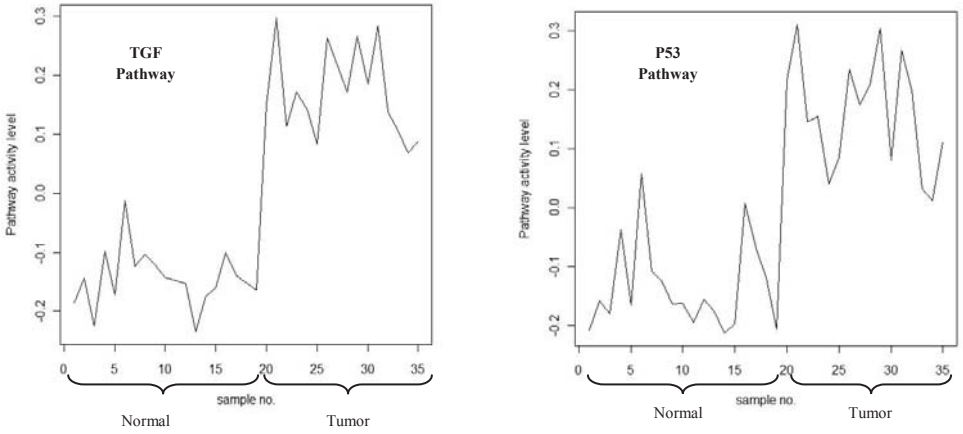

Fig. 9: The activity levels of TGF and P53 pathways.

III.1.3 Comparison between Bayesian Network and Mutual Information Network

A subset of genes in the TGF pathway, nine genes, was submitted to a Bayesian and an information based inference algorithms (CLR, ARACNE, and MRNET). The selected genes were mapped into 145 Affy IDs due to the redundancy of the same gene with different Affy IDs. Therefore, the expression of a gene in a sample was considered as the mean value of the expression values of its corresponding Affy IDs. The precision, the fraction of true edges among the inferred edges, of the inferred networks using a constraint based Bayesian algorithm, a score based Bayesian algorithm, CLR, ARACNE and MRNET was 80%, 40%, 29%, 60% and 40%. In the same manner, seven genes from the P53 pathway were submitted to these inference algorithms. However, the retrieved precision of most of these algorithms was too poor on the selected subset of genes except the ARACNE and MRNET algorithms.

IV. Discussion

As observed from figures 7 and 8, it is clear that the visual identification of gene modules in the inferred Mutual information network is easier. Gene modules for the Gene co-expression network in figure 7 could not be visually identified. The GO enrichment analysis of the selected modules in the inferred network using the Mutual information shown in figure 8 yielded three significant modules for CLR, two significant modules for MRNET and four significant modules for ARACNE. Therefore, the ARACNE surpassed all other algorithms in retrieving enriched gene modules.

The signal profiles charted in figure 9 of the activity levels of the TGF and P53 pathways show the significance of these pathways in our data sets (Hepatocellular Carcinoma). The assessment of the significance of TGF and P53 signalling pathways in HCC is an essential step before using them as a true network in comparing different inference algorithms. Checking the significance of a pathway in a specific disease using the Singular Value Decomposition is a powerful method for discovering the associated pathways with diseases.

V. Conclusion

Measuring the dependencies between genes is a prerequisite step of the gene network inference. Two measures of dependencies (Correlation and mutual information) were used. Therefore, two types of networks (Co-expression network and Mutual information network) were constructed. The complex structure of the inferred network was simplified using the Minimum Spanning Trees. Both types of the inferred networks should reveal functional gene modules. However, the visual identification of gene modules was easier in the inferred Mutual information network than in the inferred gene Co-expression network. The functionality of the selected gene modules in both types of inferred network were analysed using Gene Ontology enrichment. This comparison yielded that the ARACNE as one of the Mutual information network inference algorithms surpassed all other algorithms in retrieving enriched gene modules.

An additional comparison was held to assess the performance of the Mutual information network. The Mutual information network inference algorithms were compared to the Bayesian inference on predefined significant pathways (TGF and P53) in HCC. The significance of the selected pathways was analyzed using singular value decomposition. This comparison yielded a better precision of the inferred network using the ARACNE.

REFERENCES

1. M. Zhu and Q. Wu, "Transcription network construction for large-scale microarray datasets using a high-performance computing approach," *BMC Genomics,* 2008, 9(Suppl 1):S5.
2. J. Pearl, "Causality: Models, Reasoning, and Inference," *Cambridge University Press,* 2000.
3. R.E. Neapolitan, "Learning Bayesian Networks," *Prentice Hall,* 2003.
4. S. Horvath and J. Dong, "Geometric Interpretation of Gene Coexpression Network Analysis," *PLoS Comput Biol,* 2008, 4(8):e1000117.
5. J. Faith, B. Hayete, J. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. Collins, and T. Gardner, "Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles," *PLoS Biology,* 2007.
6. A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky and R. Favera, "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," *BMC Bioinformatics,* 2006.
7. X. Chen, M. Chen and K. Ning, "BNArray: an R package for constructing gene regulatory networks from microarray data by using Bayesian network," *Bioinformatics,* 2006, 22(23):2952-2954.
8. M. SP. Eisen, P. Brown and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proc Natl Acad Sci U S A, 95: 14863–14868,* 1998.
9. T. OO. Ideker, B. Schwikowski, A. Siegel, "Discovering regulatory and signalling circuits in molecular interaction networks," *Bioinformatics,* 2002, 18(S233–S240).
10. Y. LH. Huang, H. Hu, X. Yan, M. Waterman and et al, "Systematic discovery of functional modules and context-specific functional annotation of human genome," *Bioinformatics,* 2007, 23(i222–i229).
11. JM. SE. Stuart, D. Koller and SK. Kim, "A gene-coexpression network for global discovery of conserved genetic modules," *Science,* 2003, 302(249–255).
12. S. BC. Carter, M. Griffin and A Bond, "Gene co-expression network topology provides a framework for molecular characterization of cellular state," *Bioinformatics,* 2004, 20(2242-2250).
13. PM. Bowers, S. J. Cokus, D. Eisenberg, and T. O. Yeates, "Use of logic relationships to decipher protein network organization," *Science,* 2004, 306(5705)(2246-2249).
14. E. W. Hartl DLaJ, "Genetics: Analysis of Genes and Genomes," *Jones and Bartlett Publishers,* 2001.
15. PE. Meyer, "Information-Theoretic Variable Selection and Network Inference from Microarray Data," *PhD thesis,* 2008, Département d'Informatique, Faculté des Sciences(de l'Université Libre de Bruxelles ,Bruxelles, Belgique).
16. N. M. Abdel Samee, N. H. Solouma and Y. M. Kadah, "Detection of Biomarkers for Hepatocellular Carcinoma using a hybrid univariate Gene Selection Methods," *Theoretical Biology and Medical Modelling,* 2011, In Press.
17. P. Meyer, F. Lafitte and G. Bontempi, "minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information," *BMC Bioinformatics,* 2008, 9(1):461.
18. I. S. Butte AJaK, "Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurments," *Pacific Symposium on Biocomputing,* 2000, 5(415–426.).
19. P. E. Meyer, F. Lafitte and G. Bontempi, "Information-Theoretic Inference of Large Transcriptional Regulatory Networks," *EURASIP J Bioinform Syst Biol,* 2007, (1): 79879.
20. J. Tomfohr and Th. B. Kepler, "Pathway level analysis of gene expression using singular value decomposition," *BMC Bioinformatics,* 2005, 6: 225.