International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 2 | Issue – 3

MicroRNA-Disease Predictions Based On Genomic Data

Ajitha. C, DivyaLakshmi. K, Jothi Jayashree. M
Department of Computer Science and Engineering,
Sri Muthukumaran Institute of Technology, Chennai, Tamil Nadu,India

ABSTRACT

Gene Ontology is a structured library of concepts related with one or more gene products through a process called annotation. Association Rules that discovers biologically relevant and corresponding associations. In the existing system, they used Gene Ontology-based Weighted Association Rules for extracting annotated datasets. We here adapt the MOAL algorithm to mine cross-ontology association rules. Cross ontology rules to manipulate the Protein values from three sub ontology's for identifying the gene attacked disease. It focused on intrinsic and extrinsic values. The Co-Regulatory modules between microRNA, Transcription Factor and gene on function level with multiple genomic data. The regulations are compared with the help of integration technique. Iterative Multiplicative Updating Algorithm is used in our project to solve the optimization module function for the above interactions. Comparing the regulatory modules and protein value for gene and generating Bayesian rose tree for the efficiency of our result.

Keywords: *microRNA(miRNA), transcription factor, co-regulatory module, genomic data, mining algorithm.*

1. INTRODUCTION

MicroRNAs (miRNAs) and transcription factors (TFs), as two vital gene regulatory molecules in multicellular organisms, share a common regulatory logic. MicroRNAs are a family of small, non-coding RNAs that regulate gene expression in a sequence-specific manner, which participate in the regulation of numerous cellular process at the posttranscriptional level, such as cancer progression. TFs are proteins that control gene regulation by binding to coregulatory elements in the gene promoter region at the transcriptional level. By activating or repressing their target genes, TFs can regulate the global gene expression program of a living cell, and form transcriptional regulatory networks. However, it's still a challenge to elucidate coregulation mechanisms between miRNAs and TFs.

Recently, researchers studied the co-regulation of miRNAs and TFs by finding out their shared downstream targets. The method adopts probabilistic models and statistical tests to measure the significance of the shared targets between the regulators, and to remove the insignificant co-regulating interactions that occurred by chance. Gene enrichment analysis was used in to identify significant coregulation between the transcriptional and posttranscriptional layers. They found that some biological processes emerged only in co-regulation and that the disruption of co-regulation may be closely related to cancers, suggesting the importance of the co-regulation of miRNAs and TFs which proposed a rule based method to discover the gene regulatory modules and their target genes based on the available predicted target binding information. These work provides a good resource for exploring the regulatory relationships or identifying the network motifs.

However, target prediction basedon sequences have high rate of false discoveries, which affect the quality of the discoveries of the above mentioned methods. It would be ideal if expression data can be used to refine the discoveries. Identification of modular structure of biological networks has greatly advanced our understanding of However, little is known about the modules that exist in miRNA-TF-gene regulation systems, and even less is known about these modules

role in specific biological processes and key regulation assemblies. Several studies have made efforts to uncover miRNAs and mRNA modules on extent, it is impossible to detect highly credible miRNAs modules.

Data mining is the process of analyzing hidden patterns of data in order to different perspectives for categorization into useful and effective information. The data information's are collected and gathered in common areas, such as data warehouses, for efficient analysis, data mining algorithms, facilitating business decision making and other information requirements to ultimately cut costs and increase revenue. Data mining is otherwise known as data discovery and knowledge discovery.

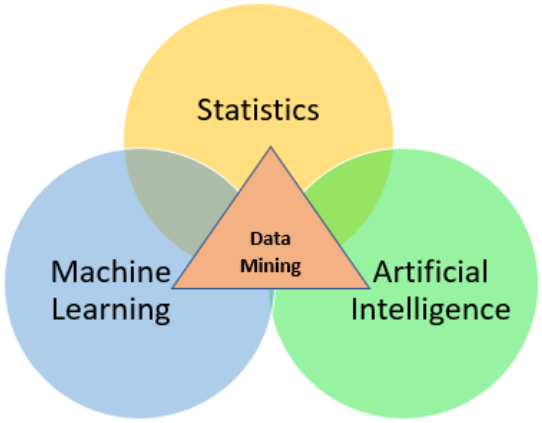

Fig. 1. Data Mining

2 WORKING PRINCIPLE

2.1 MOAL Algorithm

MOAL (Multi ontology data mining at all levels) algorithm for mines the cross ontology relationship between the ontologies. MOAL algorithm to mine cross-ontology association rules, i.e. rules that involve GO terms present in the three sub-ontologies of GO. By using collaborative filtering, user get the details about the gene id for cross ontology technique we have to compare the protein value and getting BP& MF value, or MF&CC value or CC&BP value getting the gene disease and symptoms for user requirements.

2.2 Sub Ontologies

2.2.1 Molecular Function

Molecular function activities that occur in molecular level, "catalytic activity" or "binding activity". GO molecular function that perform the actions which specify where, when, or in what context the action takes place.The activities are performed by assembled complexes of gene products corresponding to the activities that performs individual gene products.It is easy to confuse a gene product name with its molecular function. Accurately infer miRNAs functional regulation. Meanwhile, these methods have not considered TFs regulation and the modules only contain miRNAs and genes.

2.2.2 Cellular Component

Cellular Component is the one which describes a location that are related to cellular compartments along with structures, which are occupied by a macro-molecular machine when it carries a molecular function. Gene products which are described by biologists in two major ways and they are :

1. Cellular Structures and
2. Stable macromolecular complex.

2.2.3 Biological Process

A biological process is a series of events which describes one or more organized assemblies of molecular functions.Distinguition of Biological process and Molecular function.

2.3 Problem formulation

To identify miRNA-TF-gene co-regulatory modules, we design an objective function with three components (redFigure As mentioned above, the optimization function consists of a joint NMF, a regularized term for three prior networks and a sparse penalized term. Here we provide the final optimization function: *min*

$$W,H1,H2,H3 = \Sigma||Xi - WHi||F2 - \lambda 1Tr(H2AH2T)\text{---} \quad (1)$$
$$i=1$$
$$-\lambda 2Tr(H1BH2T) - \lambda 3Tr(H3CH2T) + \gamma 1||W||2F$$
$$+ \gamma 2(\Sigma||hj||12 + \Sigma||hj'||21 + \Sigma||hj''||12), j\, j'\, j''$$
$$s.t.\ W \geq 0, Hi \geq 0, i = 1,...,3$$

Where expression matrix $X1,2,3$ is decomposed by basic matrix W with size of $N \times K$ and coefficient matrix $H1,2,3$ with size of $K \times M$. The following subsections will provide the details as well as the solution of objective function.

3 OPTIMUM SOLUTION

We applied the MOAL algorithm to identify miRNA-TF-gene modules by integrating multiple independent data sources.

3.1 Intrinsic

Normal protein value of human is compared to the cross ontology value.Cross ontology values are BP&MF, CC&BP, CC&MF. If the protein value is lesser than the cross ontology value then the condition is said to be intrinsic.

3.2 Extrinsic

Normal protein value of human is compared to the cross ontology value.Cross ontology values is the interactions of GO terms. If the protein value is higher than the cross ontology value then the condition is said to be extrinsic.

3.3 Choose of parameters

The proposed SNCoNMF algorithm requires setting of several parameters as described in the pseudo code. Here it's important to decide the value of the reduced dimension of matrix factorization K. According to researches, a miRNA cluster analysis which required miRNA cluster data from the miRBase articles. As a result, we obtained about 20 clusters containing miRNAs range from 2 to 50. So in this paper we set the K to 20, approximately equals to the number of miRNA clusters represented in our data. Meanwhile, we set parameters λ1,λ2,λ3,γ1 and γ2 to 0.01, 0.01, 0.01, 20, 10, respectively. Due to the lack of TFs, we set the threshold T to 2 for TFs by conducting a series of tests, while T is set to 3 for miRNAs and genes.

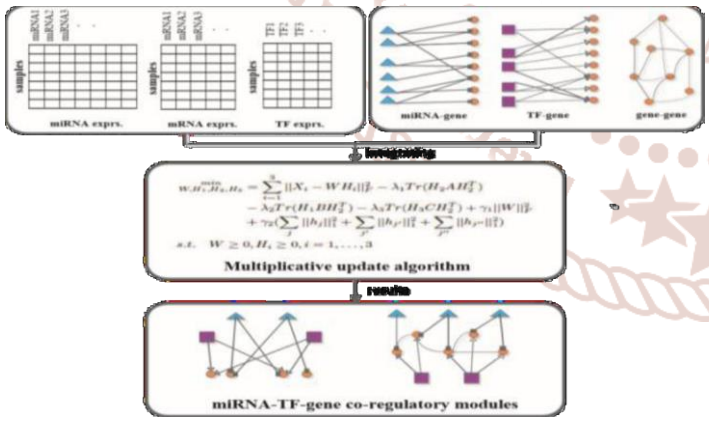

Fig. 2.Flow chart of Regulatory modules.

3.4 Module character and size distribution

We performed the proposed SNCoNMF algorithm on breast cancer dataset and obtained 20 miRNA-TF-gene coregulatory module which are composed of by a set of miRNAs, TFs and genes that are denoted as miRNA modules, TF modules and gene modules, respectively.TF-gene modules identified in this paper have an average of 5.5 miRNAs, 2.5 TFs and 28.15 genes per module. The average density of all module is 0.0176. Meanwhile, we calculated the average miRNA-gene expression correlation and TF-gene expression correlation among all modules as described in section.

In addition, to verify feasibility of our method, we run the SNMNMF algorithm on our datasets which TFs are treated as genes. In SNMNMF, the average miRNAs, genes number and TFs are 5.6 26.4 and 0.55, respectively, which genes and TFs are less than ours, especially the TTs. It demonstrates that SNCoNMF can effectively discover miRNA-TF-gene co-regulatory modules. Meanwhile, due to more genes in per module, SNCoNMF bears a less average module density.

In addition, experimental results show that all of modules from SNCoNMF are enriched in at least one GO-BP term and only one module Gene Ontology biological process term clusters for genes in each modules having 5 genes at least. We found that 45% (9/20) modules are enriched at least 5 genes in top 3 enriched GO-BP terms. Further, these GO-BP terms are enriched with similar functions or similar genes. Taking module 20 for example, the top 3 GO terms are GO:0002376, GO:0006955 and GO:0002682 which all related to immune system or immune response. Meanwhile, they are enriched some similar genes, like CD7,CD38.In order to analyze pathways enrichment, similar to functional enrichment, we calculated the top pathways that enriched by more than 3 genes. There are total 121 pathways enriched by all modules which 24.8% (30/121) have genes more than 3.

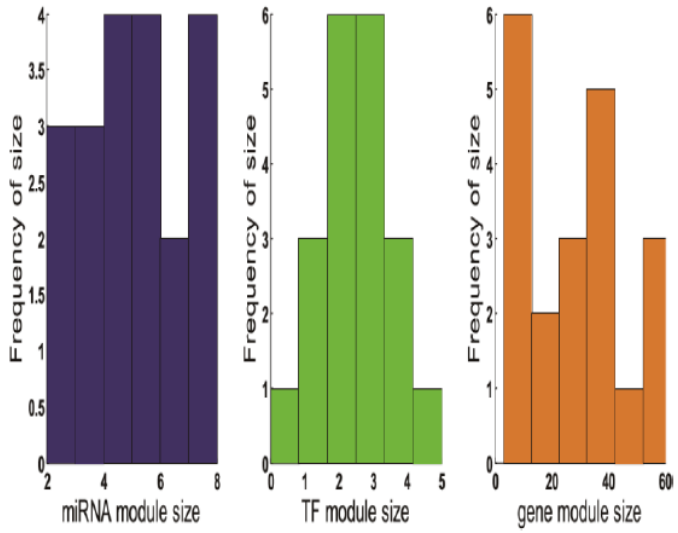

Fig. 3. Module size distribution for miRNA modules, TF modules and gene modules, respectively.

4 SYSTEM ARCHITECTURE

The cell's activity is organized as a network of interacting modules which are the set of genes co-regulated modules that respond to different conditions. We adapt a probabilistic method for identifying the regulatory modules from genomic data. Our procedure identifies modules of coregulated genes, their regulators and the conditions under which regulation occurs, generating testable hypotheses in the form 'regulator X regulates module Y under conditions W'. Admin will able to Update the Gene and maintain the website using specified modules. User can Search, Select the gene and View the Gene and precaution details based on Gene ID.Comparing the regulations between miRNA-TF interaction, TF-gene interactions and gene-miRNA interaction with the help of Integration Technique. Relation among Co-Regulatory modules is identified. Protein value for gene is generated using Bayesian rose tree for efficiency of the result.

We here adapt the method called Saccharomyces cerevisiae expression data set, which shows the ability to identify functionally coherent modules and their correct regulators. We present microarray related experiments for supporting three novel predictions, that suggest regulatory roles of previously uncharacterized proteins.

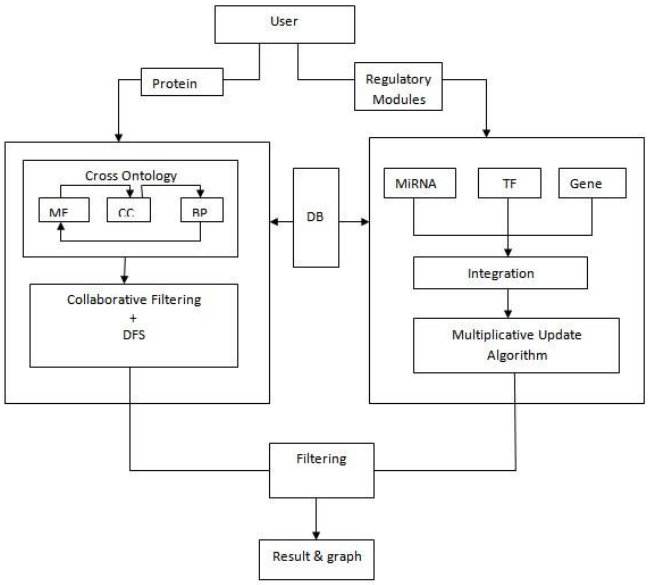

Fig.4.System Architecture of Disease associations.

In this system, we applied an integrated framework which will indicate the gene regulatory modules from the cells cycle. Incorporating multiple biological data sources, including gene expression profiles, gene ontology mechanism, and molecular interaction. In human body there exists 846 genes which plays some putative roles in the regulation of cell cycle, 46 transcription factors and 39 gene ontology groups are identified and data will be recorded. We reconstructed regulatory modules to ensure the underlying regulatory relationships. Four regulatory network motives that are identified from the interaction of gene modules.

5 MERITS AND DEMERITS

The impacts and drawbacks of existing system are,

5.1 Advantages

- Evaluates the annotation consistency.
- Avoids possible inconsistent or redundant annotations.
- Classical association rules mining algorithms used to identify victim affected by cancer.

5.2 Disadvantages

- CARM Algorithms are not able to deal with different sources of production of GO annotations.
- Candidate rules with low Information Content.
- A large amount of information is usually missed.

5.3 Merits

The main advantage of our proposed system are,

- Medical description for particular gene disease can be easily accessible.
- Quick retrieval of data.
- Identifying the data is less complex.
- Computed medical description for safeguarding the generations.

6 CONCLUSION

Relevant progresses in biotechnology and system biology are creating a remarkable amount of biomolecular data and semantic annotation. Biomolecular data increase in number and quality, but are dispersed and only partially connected. Integration and mining of these distributed and evolving data and information have the high potential of discovering hidden biomedical knowledge useful in understanding complex biological phenomena. Normal (or) pathological, and ultimately of enhancing diagnosis prognosis and treatment; but such integration poses

huge challenges. Our work has tackled them by developing a novel and generalized way to define and easily maintain. updated and extend an integration of many evolving and heterogeneous data sources. Our approach proved useful to extract biomedical knowledge about complex biological processes and diseases.

7 REFERENCES

1) O. Hobert, "Gene regulation by transcription factors and microrna," *Science*, vol. 319, no. 5871, pp. 1785–1786, 2008.
2) L. He and G. J. Hannon, "Microrna: small rna with a big role in gene regulation," *Nature Reviews Genetics*, vol. 5, no. 7, pp. 522–531, 2004.
3) J. LU, G. Getz, and E. A. Miska, "Microrna expression profiles classify human cancers," *nature*, vol. 435, no. 9, pp. 834–838, Jun 2005.
4) D. H. Tran, K. Satou, T. B. Ho, and T. H. Pham, "Computational discovery of mir-tf regulatory modules in human genome," *Bioinformation*, vol. 4, no. 8, pp. 371–377, 2010.
5) Q. Cui, Z. Yu, E. O. Purisima, and E. Wang, "Principles of microrna regulation of a human cellular signaling network," *molecular systems biology*, vol. 2, p. 46, September 2006.
6) X. Yang, M. Feng, X. Jiang, Z. Wu, Z. Li, M. Aau, and Q. Yu, "mir-449a and mir-449b are direct transcriptional targets of e2f1 and negatively regulate prbce2f1 activity through a feedback loop by targeting cdk6 and cdc25a," Genes Dev., vol. 23, no. 20, pp. 2388–2393, October 2009.
7) S. Yoon and G. D. Micheli, "Prediction of regulatory modules comprising microrna and target genes," Bioinformatics, vol. 21, no. 2, pp. ii93–ii100, 2005.
8) X. Peng, Y. Li, K.-A. Walters, E. R. Rosenzweig, S. L. Lederer, L. D. Aicher, S. Proll, and M. G. Katze, "Prediction of regulatory modules comprising microrna and target genes," BMC Genomics, vol. 10, p. 373, 2009.
9) W. E, C. X, H. R, K. H, L. I, M. V, M. T, P. M, R. I, and S. F, "Transfac: an integrated system for gene expression regulation," Nucleic Acids Research, vol. 28, no. 1, pp. 316– 319, January 2000.
10) J. C. Huang, T. Babak, T. W. Corson, G. Chua, S. khan, IEEE TRANSACTION ON NANOBIOSCIENCE , VOL.16 , NO.1 , JAN.2017.
11) B. L. Gallie, T. R. Hughes, B. J. Blencowe, B. J. Frey, and Q. D. Morris, "Using expression profiling data to identify human microrna targets," Nature Methods, vol. 4, pp. 1045– 1049, 2007.
12) C. Cheng and L. M.Li, "Inferring microrna activities by combining gene expression with microrna target prediction," PLoS one, vol. 3, no. 4, p. e1989, April 2008.
13) M. Lee, S. Y. Rha, H. C. Chung, D. M. Virshup, and P. Tan, "A densely interconnected genome-wide network of microrna and oncogenic pathways revealed using gene expression signatures," PLoS genetics, vol. 7, no. 12, p. e1002415, December 2011.
14) S. Zhang, Q. Li, J. Liu, and X. J. Zhou, "A novel computational framework for simultaneous integration of multiple types of genomic data to identify microrna-gene regulatory modules," Bioinformatics, vol. 27, no. 13, pp. i401–i409, 2011.
15) Y. Zhou, J. Ferguson, J. T. Chang, and Y. Kluger, "Inter-and intra-combinatorial regulation by transcription factors and microrna," BMC genomics, vol. 8, no. 1, p. 1, 2007.
16) C.-Y. Chen, S.-T. Chen, C.-S. Fuh, H.-F. Juan, and H.-C. Huang, "Coregulation of transcription factors and microrna in human transcriptional regulatory network," BMC bioinformatics, vol. 12, no. 1, p. 1, 2011.
17) D. H. Tran, K. Satou, T. B. Ho, and T. H. Pham, "Computational discovery of mir-tf regulatory modules in human genome," Bioinformation, vol. 4, no. 8, pp. 371–377, 2010.
18) T. H. Hwang, G. Atluri, R. Kuang, V. Kumar, T. Starr,K. A. Silverstein, P. M. Haverty, Z. Zhang, and J. Liu, "Large-scale integrative network-based analysis identifies common pathways disrupted by copy number alterations across cancers," BMC genomics, vol. 14, no. 1, p. 440, 2013.
19) J.-G. Joung, K.-B. Hwang, J.-W. Nam, S.-J. Kim, and B.-T. Zhang, "Discovery of micrornacmrna modules via population-based probabilistic learning," Bioinformatics, vol. 23, no. 9, pp. 1141–1147, 2007.
20) D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in Advances in neural information processing systems, 2001, Conference Proceedings, pp. 556– 562.

21) Y. Li, C. Qiu, J. Tu, B. Geng, J. Yang, T. Jiang, and Q. Cui, "HMDD v2. 0: a database for experimentally supported human microrna and disease associations," Nucleic acids research, p. gkt1023, 2013.
22) O. Hobert, "Gene regulation by transcription factors and microRNA's," Science, vol. 319, no. 5871, pp. 1785–1786, 2008.
23) L. He and G. J. Hannon, "Micrornas: small RNA's with a big role in gene regulation," Nature Reviews Genetics, vol. 5, no. 7, pp. 522–531, 2004