ISSN 2395-8618

Comparative Study of Computational Tools for Hub Gene Selection from Genetic Network using Microarray Data

Bijeta Mandal, Saswati Mahapatra, and Tripti Swarnkar

***Abstract*—Selection of genes associated to complex diseases has been a challenging task in the field of bioinformatics. Through various studies it has been concluded that selection of highly connected intramodular hub genes in a co-expression network analysis approach leads to more biologically relevant gene lists. In this paper, we have assess the empirical performance of three existing network reconstruction methods Weighted Gene Correlation Network Analysis (WGCNA), Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE), Graphical Lasso (GLASSO). The study compares the extracted hub genes from estimated networks on the prostate cancer dataset based on two criteria: the first criterion evaluates the biological enrichment and the second criterion evaluates the statistical validation, prediction accuracy. The result suggests, though there is considerable amount of heterogeneity, randomness and variability in structures of networks estimated using different reconstruction methods, our findings provides evidence for similarity in hub genes selection. These findings after network analysis can provide an intuitive insight into selection of network estimation methods for specific range of gene expression in microarray datasets. *Index Terms*—Gene Selection, Intramodular hub gene, Co-expression network, Genetic Network, Network reconstruction, Network analysis, Microarray**

I. Introduction

Understanding the relationship among genes, is extremely fundamental with a specific end goal to analyze genomic data. Gene expression data can be productively dissected with network methods characterizing clusters of interconnected genes [1], with edges capturing interactions at different levels. Genetic interactions hypothesizes activities of biological pathway, cellular response [2], acknowledging elements of genes from their reliance on different genes [3], distinguishing novel biomarkers [4] and more precise classification methods [5]. The degree of interactions in the clusters are significantly higher than an irregular network exhibiting indistinguishable degree distribution [6]. Diverse statistical and bioinformatics techniques can be applied directly to microarray data to estimate networks of genetic interactions in different cellular states or disease stages with a common motive to glean an edge among a pair of genes by considering a cue of association, which is pivotal in different network reconstruction method [7]. Associations in network can be classified into two: Marginal associations that ignores the nearness of other genes while estimating an edge between genes and conditional associations that considers impact of nearness of other genes while concluding an edge between genes. Focusing on intramodular hubs instead of whole network hubs for co-expression network applications leads to better results of clinical significance [8] a key factors in a network architecture [9], and are often strongly enriched in specific functional categories or cell markers [10]. Empirical evidence shows gene selection based on intramodular connectivity leads to biologically more informative gene lists focusing on the relationship between modules and the sample trait [1], [11], [12], that prompts gene connectivity can be used for identifying hubs and differentially connected genes [11], [13], [14] for finding biological information embedded in microarray data [13], [14]. Our comparative study includes a comparison of three computational methods with publicly available software, Weighted Gene Correlation Network Analysis (WGCNA) [15], Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) [16], Graphical Lasso (GLASSO) [17] for reconstruction of genetic networks with undirected edges as it is not possible to estimate directed edges with observational data alone [18]. Different computational tools are implemented to a benchmark dataset to analyze similarities and differences in estimated networks and their performances in terms of intramodular hub genes. Finally, the presence of cancer related genes and their influence in specific cancer type using NCBI database and DAVID [19] has been studied. The result provides a insight into the presence of cancer-related genes in the hub gene modules found in known biological networks [20] and also helps in selection of most efficient network estimation method. The rest of the paper includes detail of methods for network reconstruction, proposed model, results analysis and discussion of our findings for reconstructed genetic networks, and future research scope.

Manuscript received on September 12, 2016, accepted for publication on December 17, 2016, published on June 30, 2017.

The authors are with the Department of Computer Application, Siksha 'O'Anusandhan University, Bhubaneswar 751 030, India (e-mail: bijetamandal@gmail.com, saswatimohapatra@soauniversity.ac.in, triptiswarnakar@soauniversity.ac.in).

II. METHODS AND MATERIALS

A. Methods

WGCNA [15] is a genetic network reconstruction tool based on marginal measure of correlation patterns among

ISSN 2395-8618

genes that incorporates functions for finding modules of highly correlated genes. In WGCNA, gene significance and module eigengene or intramodular hub gene based connectivity among genes facilitate gene screening methods to identify candidate biomarkers, and can be used to generate testable hypotheses for validation in independent datasets [21]. WGCNA is implemented using R software. ARACNE [16] is based on removal of non-linear similarities among expression levels for a pair of genes. The algorithm computes pair wise mutual information MI_{ij} for each pair of genes *i* and *j*, and applies DPI (Data Processing Inequality) as a pruning step for removal of the false positive edges corresponding to indirect interactions in the network. ARACNE is classified as a method based on a blend of marginal and conditional associations and is implemented using package minet [22] in the Bioconductor. Graphical lasso (GLASSO) [17] is a network estimation tool based on sparsity inducing penalties i.e. lasso penalty assuming multivariate normality for random variables [7]. GLASSO estimates inverse covariance matrix by maximizing the l_1- penalized log likelihood function to construct a sparse graphs of conditional independence relations among the genes. The tuning parameter ρ is a positive number controlling the degree of sparsity. It is implemented with package glasso in R.

B. Datasets

The prostate cancer microarray dataset for homo sapiens consisting 104 samples and 20000 genes with 6 variants in samples [20], is utilized in our study to show the effectiveness of our proposed model, obtained from NCBIs Gene Expression Omnibus (GEO). For the purpose of statistical analysis the samples are categorized into two, 70 diseased samples and 34 normal samples. Biological dataset adopted for validation collected from the NCBI gene database includes 7238 cancer-related genes and 2202 prostate cancer genes.

C. Proposed Model

Block diagram in Fig 1, represents the schematic work flow of the proposed hub genes selection model.

1) Data Preprocessing.: The methods are being implemented in R software using different R packages. The preprocessing of data includes cleaning of data by removal of genes with large number of missing values. Hierarchical clustering is performed for finding sample outliers in the samples. Missing values of a gene are replaced with the mean value of observed data. The genes are filtered based on their variances across diseased and normal samples producing 100 samples and 14689 probes. Due to technical limitations regarding memory allocation during GLASSO implementation (System specification:12 GB RAM) we had to confined the number of probes not more than 10000 for different computational tool implementation.

Computational Tools Implementation.:

a) WGCNA.: Pearson correlation S_{ij} is calculated for the gene expression profile and are then transformed into adjacency matrix by applying a power adjacency function $|S_{ij}|\,\beta$, where the exponent β is the power estimate to obtain a scale-free topology [23]. Further co-expression values are converted to the topology overlap measure (TOM), that facilitates the identification of gene modules. The output of the implementation showed 19 modules. Based on high module membership and intra modular connectivity hub gene modules are selected.

b) ARACNE.: Mutual information (MI) is evaluated between each pair of genes and is taken as input to the aracne() function for network estimation. The number of the edges are controlled by thresholding the value of MI for each pair of genes in the network. The output of the implementation showed 1 module. For analysis and comparison with network estimation from other tools, connectivity of each node is considered.

c) GLASSO.: Covariance matrix is calculated between each pair of genes and taken as input to glasso() function to calculate an inverse covariance matrix for network estimation. In our implementation we have opted for two variations of GLASSO i.e. defining diagonal of inverse covariance to be penalized or not. The output of the implementation showed 1 module each for both the variations of GLASSO implementation considering the module constraint of minimum 25 genes, taken as standard in WGCNA.

2) Extraction of Hub Gene Modules.: A hub gene module with high intramodular connectivity can be considered as a gene module with strongly interacting genes. Study shows genes with higher module membership show higher intramodular connectivity and are more biologically significant [15]. A set of twenty top ranked genes are extracted from each module to create hub gene modules for further analysis. Integrated modules shows improved classification performance in gene selection [20], so we have selected five top ranked genes from individual hub gene module to construct a integrated hub module.

3) Performance evaluation of selected hub genes.:

a) Statistical Analysis.: Predictive accuracy of the hub genes are measured in terms of Matthews coefficient correlation (MCC), as it is a measure of quality of binary classification. [24], [25]. MCC, overall accuracy, sensitivity, specificity, precision and f-measure are adopted for statistical analysis in comparison to the known true classes [25].

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \quad (1)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

ISSN 2395-8618

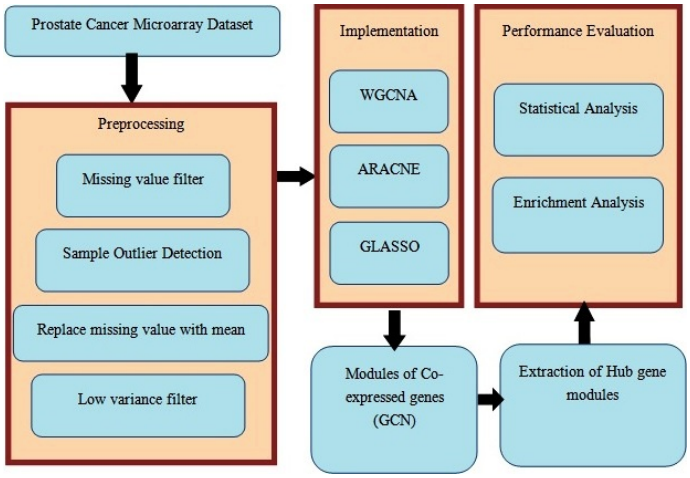

Fig. 1. Steps for gene selection in proposed model

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$F - measure = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{6}$$

where TP is number of true-positive samples, TN is count for true-negative samples, FP is number of false-positive samples and FN is number of false-negative samples.

b) Enrichment Analysis.: The biological significance of the selected hub genes are firstly validated with the percentage of disease-related genes in them and secondly the results are validated by summarizing the genes belonging to an enriched functional category measured in terms of p-value [26] and fold enrichment, of enriched attributes (EA) using DAVID [19].

III. RESULT AND DISCUSSION

In the paper we have performed three step evaluation of the selected modules for the hub genes selection.
(i) Comparison of modules based on the graph density of hub gene modules.
(ii) Effectiveness of selected hub gene modules are analyzed in terms of prediction accuracy.
(iii) Biological significance is analyzed involving presence of disease related genes and enriched attributes.
After applying the progression (ii) of our proposed model for GCN construction utilizing distinctive computational strategies, brought about 19 modules in WGCNA ,their genes are ranked predicated on their intra modular connectivity. An arrangement of twenty top positioned genes are extracted from every module to extracted 19 hub gene modules for further analysis, that tallies to cull of 439 genes. Assuming integrated modules shows amended relegation performance in gene glean [20], We have sorted out five top ranked genes from individual hub gene module to contrive a integrated hub module Hub5 with 125 genes. Hub genes and subset of hub gene modules are constructed from the modules estimated using ARACNE(A1-603 genes, A2-179 genes) and GLASSO(for penalized diagonal false: F1-497 genes, F2-134 genes, for penalized diagonal true: T1-497 genes, T2-140 genes implementation, by considering the heterogeneity in degree distribution for network estimates utilizing distinctive computational tools and number of genes selected as hub genes in WGCNA for individual and integrated modules as standard. The distinct co-expressed gene modules and integrated modules constructed using distinctive computational strategies of our approach are designated as following: Bl Black, B Blue, Br Brown, C Cyan, G Green, GY Green Yellow, G60 Grey 60, LC Light Cyan, LG Light Green, LY Light Yellow, M Magenta, MB Midnight Blue, P Pink, Pu Purple, R Red, S Salmon, Tn Tan, T Turquoise, Y Yellow, Hub5 are the co-expressed hub gene modules obtained using WGCNA based network construction approach. A1, A2 are the co-expressed hub gene modules obtained using ARACNE based network construction approach and F1, F2, T1, T2 are the co-expressed hub gene modules obtained using GLASSO based network construction approach.

A. Graph Density Analysis

We surmise that the more precise and dense the gene module, the higher the quality measure [20]. In Fig. 2 and Fig. 3, we have summarized the results for all hub gene modules from different computational tools in terms of graph density (the ratio between number of edges and number of nodes/genes) for prostate cancer dataset. After obtaining graph density measure for different co-expressed hub gene module from different computational tools, we filtered 6 different hub gene modules in WGCNA (five individual modules and one integrated hub gene module) for prostate dataset. As number of modules in ARACNE and GLASSO implementation is very less so all the hub gene modules are considered for the study. The selected hub gene modules show comparatively high graph density with respect to the intramodular connectivity. Thus, these selected hub gene modules, are further considered for statistical and biological analysis.

ISSN 2395-8618

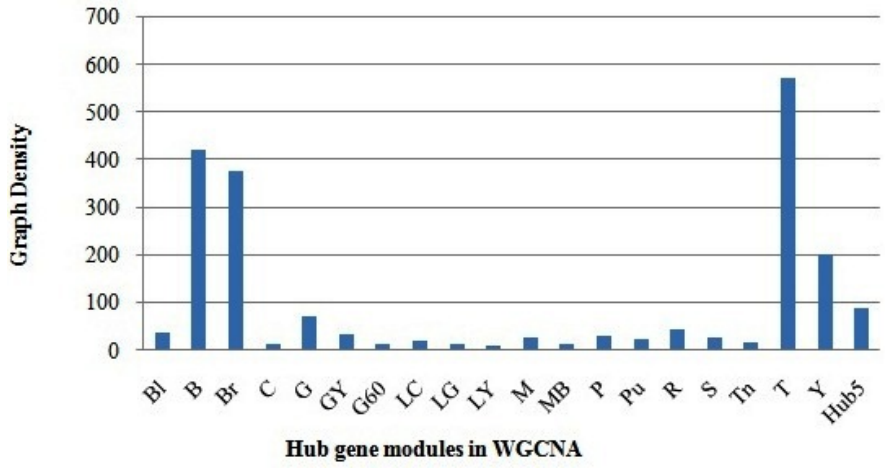

Fig. 2. Graph density of Hub gene module for WGCNA

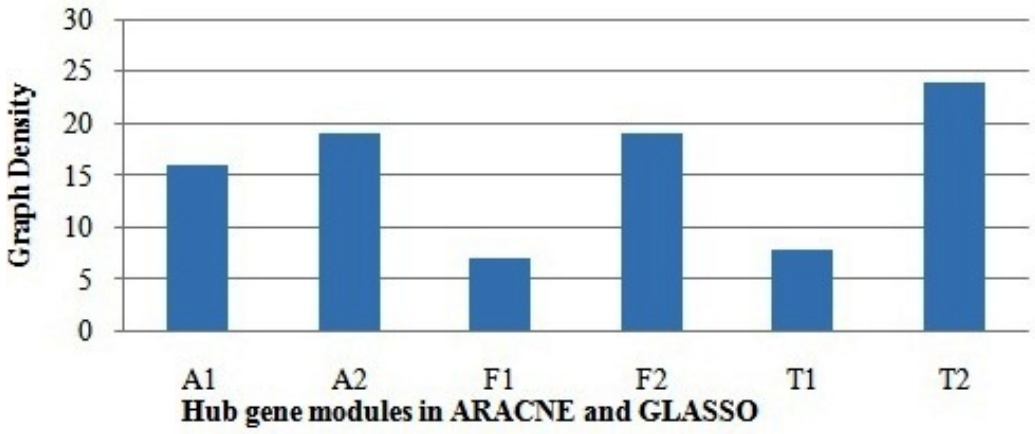

Fig. 3. Graph density of Hub gene module for ARACNE and GLASSO

B. Classification performance

Hub genes are high degree nodes that incline to play a consequential role in the functional modules [27]. The performance of the hub gene modules is evaluated in terms of predictive accuracy as listed in Table 1 for prostate cancer dataset . The kNN (k=3), Random Forest and SVM with tenfold cross validation are applied as classifiers [20]. From Table 1, we observed few hub gene modules in WGCNA (Blue,Hub5) shows better results than of the individual hub gene modules, for ARACNE (A1,A2) and for GLASSO (F1, F2, T1, F1) the results are good.

C. Biological Significance analysis

The biological analysis of co-expressed hub gene modules are based on the fol-lowing criteria:

1) Disease-related genes analysis.: Fig 4, illustrates the efficacy of hub gene selection in terms of identifying disease-related genes represented as the percentage of studied cancer (prostate) related genes in each significant hub gene module. It is been observed that the hub gene modules with enhanced prediction accuracy have high fraction of co-expressed cancer-related genes. They are being considered as significant for further study for genes mostly related with disease.

2) Analysis of enriched attributes associated with prostate cancer hub gene module.: The biological significance is evaluated in terms of percentage of genes related with specific relevant biological process in each hub gene module and are shown in Table 2 for prostate cancer data. The biological significance of the genes belonging to an enriched functional category can be measured in terms of p-value [26]. The results are validated using p-value value cut-off of 5×10^2 and fold enrichment (FE) 1.5 [6], of enriched attributes/functions (EA) in our study. Since DAVID gene ID is unique per gene, it is more accurate to use DAVID ID to present the gene-annotation association by removing any redundancy in user gene list. Interestingly, Hub5, A1, F1, F2, T1, T2 shows relatively large number of EAs satisfying the p-value and FE cut-off.

ISSN 2395-8618

TABLE I
BIOLOGICAL FUNCTIONAL ANALYSIS OF GENES IN HUB GENE MODULES IN TERMS OF ENRICHED ATTRIBUTE COUNT

M	NG	CL	3NN						RF						SVM					
			Sen	Spec	Prec	Fm	Mcc	Acc	Sen	Spec	Prec	Fm	Mcc	Acc	Sen	Spec	Prec	Fm	Mcc	Acc
B	20	N	0.62	0.86	0.07	0.66	S0.50	0.78	0.56	0.97	0.91	0.69	0.62	0.83	0.35	1.00	1.00	0.52	0.51	0.78
		P	0.86	0.62	0.81	0.84	0.50		0.97	0.56	0.81	0.88	0.62		1.00	0.35	0.75	0.86	0.51	
Br	21	N	0.77	0.77	0.63	0.69	0.52	0.77	0.65	0.85	0.69	0.67	0.50	0.78	0.12	0.88	0.33	0.17	-0.01	0.62
		P	0.77	0.77	0.86	0.82	0.52		0.85	0.65	0.82	0.84	0.50		0.88	0.12	0.66	0.75	-0.01	
G	24	N	0.65	0.79	0.61	0.63	0.43	0.74	0.41	0.88	0.64	0.50	0.33	0.72	0.38	0.94	0.77	0.51	0.41	0.75
		P	0.79	0.65	0.81	0.80	0.43		0.88	0.41	0.74	0.81	0.33		0.94	0.38	0.75	0.51	0.41	
T	21	N	0.53	0.74	0.51	0.52	0.27	0.67	0.47	0.86	0.64	0.54	0.37	0.73	0.00	1.00	0.00	0.00	0.00	0.66
		P	0.74	0.53	0.75	0.75	0.27		0.86	0.47	0.76	0.81	0.37		1.00	0.00	0.66	0.80	0.00	
Y	22	N	0.47	0.79	0.53	0.50	0.27	0.68	0.35	0.94	0.75	0.48	0.38	0.74	0.03	0.99	0.50	0.06	0.05	0.66
		P	0.79	0.47	0.74	0.77	0.27		0.94	0.35	0.74	0.83	0.38		0.99	0.03	0.66	0.79	0.05	
Hub5	125	N	0.79	0.96	0.90	0.84	0.77	0.90	0.68	0.96	0.89	0.77	0.68	0.86	0.71	0.94	0.86	0.77	0.68	0.86
		P	0.96	0.79	0.90	0.93	0.77		0.96	0.68	0.85	0.90	0.68		0.94	0.71	0.86	0.90	0.68	
A1	603	N	0.91	0.86	0.78	0.84	0.75	0.88	0.68	0.99	0.96	0.79	0.73	0.88	0.59	0.97	0.91	0.71	0.64	0.84
		P	0.86	0.91	0.95	0.91	0.75		0.99	0.68	0.86	0.92	0.73		0.97	0.59	0.82	0.89	0.64	
A2	179	N	0.85	0.85	0.74	0.80	0.68	0.85	0.62	0.94	0.84	0.71	0.61	0.83	0.53	0.97	0.90	0.67	0.59	0.82
		P	0.85	0.85	0.92	0.88	0.68		0.94	0.62	0.83	0.88	0.61		0.97	0.53	0.80	0.88	0.59	
F1	497	N	0.91	0.92	0.86	0.89	0.83	0.92	0.74	0.97	0.93	0.82	0.75	0.89	0.71	0.97	0.92	0.80	0.73	0.88
		P	0.92	0.91	0.95	0.94	0.83		0.97	0.74	0.88	0.92	0.75		0.71	0.71	0.87	0.91	0.73	
F2	134	N	0.91	0.91	0.84	0.87	0.81	0.91	0.79	0.97	0.93	0.86	0.80	0.91	0.82	0.96	0.90	0.86	0.80	0.91
		P	0.91	0.91	0.95	0.93	0.81		0.97	0.79	0.90	0.93	0.80		0.96	0.82	0.91	0.93	0.80	
T1	497	N	0.91	0.92	0.86	0.89	0.83	0.92	0.74	0.97	0.93	0.82	0.75	0.89	0.71	0.97	0.92	0.80	0.73	0.88
		P	0.92	0.91	0.95	0.94	0.83		0.97	0.74	0.88	0.92	0.75		0.97	0.71	0.87	0.91	0.73	
T2	140	N	0.91	0.91	0.84	0.87	0.81	0.91	0.82	0.97	0.93	0.88	0.82	0.92	0.79	0.97	0.93	0.86	0.80	0.91
		P	0.91	0.91	0.95	0.93	0.81		0.97	0.82	0.91	0.94	0.82		0.97	0.79	0.90	0.93	0.80	

Bold hub gene module specifies the hub gene modules showing comparable good predictive accuracy measures. NG number of genes, Cl Class label, 3NN 3 nearest neighbors, RF random forest, SVM support vector machine, Sen sensitivity, Spec specificity, Prec precision, Fm F-measure, Mcc Matthews correlation coefficient; Acc prediction accuracy, N negative (normal) sample, P positive (prostate cancer) sample

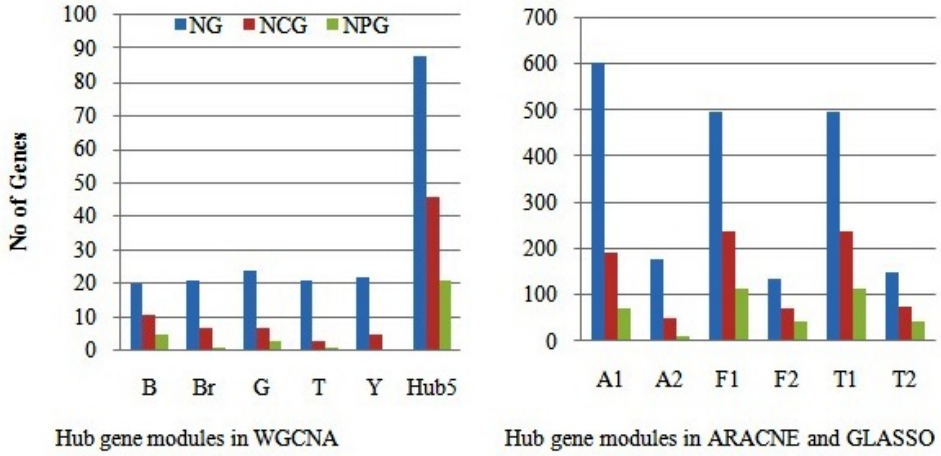

Fig. 4. Biological significance study of the hub gene modules in terms of the presence of disease-related genes for prostate cancer dataset in WGCNA, ARACNE and GLASSO. NG number of genes in hub gene module, NCG number of Cancer genes in hub gene module, NPG number of Prostate Cancer genes in hub gene module

ISSN 2395-8618

Few of the biological functions more related to the disease are also found enriched in the modules. These processes mainly include transcription, translation and RNA binding that plays an important role in protein regulation. Acetylation and phosphoproteins are known to play a vital role in genetics modification that occurs in cancer.The dysregulation of cell cycle, spliceosome and focal adhesion plays pivotal role in cancer metastasis. Regulation of apoptosis, UBL conjugation are important parts of programmed cell death and have significant change in cancer progression [20].

IV. CONCLUSION

The advantage of focusing on intramodular hub genes instead of whole network of co-expressed genes leads to better selection of biologically enriched and statically significant biomarkers. The study shows the comparison of gene selection using three widely used standard computational tools. We have evaluated the selected hub gene modules for three different benchmark methods based on their graph density, predic-tion accuracy and presence of enriched attributes. Considering graph density as meas-ure, modules formed in WGCNA are more dense than modules estimated from other tools. The statistical analysis of selected modules based on graph density shows, modules in ARACNE, GLASSO and integrated module in WGCNA have compara-tively similar class performance and outperforming the individual modules in WGCNA showing moderate accuracy. The modules in GLASSO are biologically more significant with respect to presence of enriched attributes than the modules in ARACNE and WGCNA. All the standard computational methods used in the study are showing similar performance, at the same time GLASSO and ARACNE are show-ing more computational complexity based on size of the modules created. The hub gene selected using different computational tools may further be provided to different known networks which may provide greater insights into the fundamental biology and pathogenesis of the disease.

REFERENCES

[1] P. Langfelder, P. S. Mischel, and S. Horvath, "When is hub gene selection better than standard meta-analysis?" *PloS one*, vol. 8, no. 4, p. e61505, 2013.

[2] A. B. Parsons, R. L. Brost, H. Ding, Z. Li, C. Zhang, B. Sheikh, G. W. Brown, P. M. Kane, T. R. Hughes, and C. Boone, "Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways," *Nature biotechnology*, vol. 22, no. 1, pp. 62–69, 2004.

[3] P. Ye, B. D. Peyser, X. Pan, J. D. Boeke, F. A. Spencer, and J. S. Bader, "Gene function prediction from congruent synthetic lethal interactions in yeast," *Molecular systems biology*, vol. 1, no. 1, 2005.

[4] L. Chen, J. Xuan, R. B. Riggins, R. Clarke, and Y. Wang, "Identifying cancer biomarkers by network-constrained support vector machines," *BMC systems biology*, vol. 5, no. 1, p. 1, 2011.

[5] P. Hu, S. B. Bull, and H. Jiang, "Gene network modular-based classification of microarray samples," *BMC bioinformatics*, vol. 13, no. 10, p. 1, 2012.

[6] T. Swarnkar, S. N. Simoes, D. C. Martins, A. Anurak, H. Brentani, R. F. Hashimoto, and P. Mitra, "Multiview clustering on ppi network for gene selection and enrichment from microarray data," in *Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on*. IEEE.

[7] N. Sedaghat, T. Saegusa, T. Randolph, and A. Shojaie, "Comparative study of computational methods for reconstructing genetic networks of cancer-related pathways," *Cancer Inform*, vol. 13, no. Suppl 2, 2014.

[8] S. Horvath, B. Zhang, M. Carlson, K. Lu, S. Zhu, R. Felciano, M. Laurance, W. Zhao, S. Qi, Z. Chen *et al.*, "Analysis of oncogenic signaling networks in glioblastoma identifies aspm as a molecular target," *Proceedings of the National Academy of Sciences*, vol. 103, no. 46, pp. 17 402–17 407, 2006.

[9] E. Almaas, "Biological impacts and context of network theory," *Journal of Experimental Biology*, vol. 210, no. 9, pp. 1548–1558, 2007.

[10] M. C. Oldham, G. Konopka, K. Iwamoto, P. Langfelder, T. Kato, S. Horvath, and D. H. Geschwind, "Functional organization of the transcriptome in human brain," *Nature neuroscience*, vol. 11, no. 11, pp. 1271–1282, 2008.

[11] T. F. Fuller, A. Ghazalpour, J. E. Aten, T. A. Drake, A. J. Lusis, and S. Horvath, "Weighted gene coexpression network analysis strategies applied to mouse weight," *Mammalian Genome*, vol. 18, no. 6-7, pp. 463–472, 2007.

[12] P. Langfelder, L. W. Castellani, Z. Zhou, E. Paul, R. Davis, E. E. Schadt, A. J. Lusis, S. Horvath, and M. Mehrabian, "A systems genetic analysis of high density lipoprotein metabolism and network preservation across mouse models," *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, vol. 1821, no. 3, pp. 435–447, 2012.

[13] X. Xu and A. Zhang, "Selecting informative genes from microarray dataset by incorporating gene ontology," in *Fifth IEEE Symposium on Bioinformatics and Bioengineering (BIBE'05)*. IEEE, 2005, pp. 241–245.

[14] R. Díaz-Uriarte and S. A. De Andres, "Gene selection and classification of microarray data using random forest," *BMC bioinformatics*, vol. 7, no. 1, p. 1, 2006.

[15] P. Langfelder and S. Horvath, "Wgcna: an r package for weighted correlation network analysis," *BMC bioinformatics*, vol. 9, no. 1, p. 1, 2008.

[16] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, and A. Califano, "Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," *BMC bioinformatics*, vol. 7, no. Suppl 1, p. S7, 2006.

[17] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[18] A. Shojaie, A. Jauhiainen, M. Kallitsis, and G. Michailidis, "Inferring regulatory networks by combining perturbation screens and steady state gene expression profiles," *PloS one*, vol. 9, no. 2, p. e82393, 2014.

[19] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," *Nucleic acids research*, vol. 37, no. 1, pp. 1–13, 2009.

[20] T. Swarnkar, S. N. Simões, A. Anura, H. Brentani, J. Chatterjee, R. F. Hashimoto, D. C. Martins, and P. Mitra, "Identifying dense subgraphs in protein–protein interaction network for gene selection from microarray data," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 4, no. 1, pp. 1–18, 2015.

[21] S. Horvath and J. Dong, "Geometric interpretation of gene coexpression network analysis," *PLoS comput biol*, vol. 4, no. 8, p. e1000117, 2008.

[22] P. E. Meyer, F. Lafitte, and G. Bontempi, "minet: Ar/bioconductor package for inferring large transcriptional networks using mutual information," *BMC bioinformatics*, vol. 9, no. 1, p. 1, 2008.

[23] A. Li and S. Horvath, "Network neighborhood analysis with the multi-node topological overlap measure," *Bioinformatics*, vol. 23, no. 2, pp. 222–231, 2007.

[24] P. Dao, K. Wang, C. Collins, M. Ester, A. Lapuk, and S. C. Sahinalp, "Optimally discriminative subnetwork markers predict response to chemotherapy," *Bioinformatics*, vol. 27, no. 13, pp. i205–i213, 2011.

[25] J. Ahn, Y. Yoon, C. Park, E. Shin, and S. Park, "Integrative gene network construction for predicting a set of complementary prostate cancer genes," *Bioinformatics*, vol. 27, no. 13, pp. 1846–1853, 2011.

[26] A. Ghosh, B. C. Dhara, and R. K. De, "Selection of genes mediating certain cancers, using a neuro-fuzzy approach," *Neurocomputing*, vol. 133, pp. 122–140, 2014.

[27] I. W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, S. Bull, T. Pawson, Q. Morris, and J. L. Wrana, "Dynamic modularity in protein interaction networks predicts breast cancer outcome," *Nature biotechnology*, vol. 27, no. 2, pp. 199–204, 2009.

ISSN 2395-8618

TABLE II
BIOLOGICAL FUNCTIONAL ANALYSIS OF GENES IN HUB GENE MODULES IN TERMS OF ENRICHED ATTRIBUTE COUNT

CM	Modules	NG	DC	No of EA
WGCNA	B	20	15	14
	Br	21	18	1
	G	24	19	10
	T	21	13	1
	Y	22	9	1
	Hub5	125	88	120
ARACNE	A1	603	421	219
ARACNE	A2	179	126	53
GLASSO	F1	497	367	580
GLASSO	F2	134	103	369
GLASSO	T1	497	367	580
GLASSO	T2	140	108	340

Computational method, NG number of genes, DC DAVID ID count EA enriched attribute