ACGLM: A Hybrid Approach to Select and Combine Gene Expression Regulation in Cancer Datasets

Hesham Abdulatef Mohammed Al-bukhaiti¹, a and Jiawei Luo², *

¹, ² College of Information Science and Engineering, Hunan University, Changsha, Hunan, China 410082.

* Corresponding author e-mail: luojiawei@hnu.edu.cn, aheshamibb@yahoo.com

Abstract. Cancer is one of the causes of death in the world and many genes are involved in it. Transcription factors (TFs) and microRNAs (miRNAs) are primary gene regulators and regulatory mechanisms for cells to define their targets. The study of the regulatory mechanisms of the two main regulators is complex, but this leads to a deeper interpretation of biological processes. In order to avoid exhaustive search and unnecessary genes, firstly, mRNA expression and miRNA expression are clustered by K-means cluster, then, applied ANOVA test to select significant genes. We proposed a gene regulatory network (GRN) estimation method, using Directed networks with generalized linear regression to predict and explain the relationships between regulators and their targets. Where through GO TERM and KEGG PATHWAY for target genes we got many processes such as cell communication, regulation of the biologic process, biological regulation and cell cycle, DNA replication, and cell cycle, these processes are considered significant to the cancer diseases. By comparing with other methodologies Our approach was better, as well as the results were consistent with the medical literature, where the important regulators in our gene regulatory network have a major role in cancer this explains the efficiency of this approach.

1. Introduction

Cancer is a deadly disease, and one of the causes of death in the world and many genes are involved in several cancers, such as lung cancer, breast cancer, stomach, and others. Many studies indicate the involvement of both TFs and microRNA in the cancer disease [1-4]. The major transcriptional regulator is transcription factors (TFs), which bind to cis-regulatory elements (CREs) in the promoter regions of genes and control the downstream transcriptional activities [6]. MicroRNAs, a small ~ 22 nucleotide noncoding RNA species, have been shown to play a predominant role in post-transcriptional and translational regulation [6]. Directly or indirectly the regulators can control their expression and the expression of their mutual targets in number of form like feedback and feed-forward loops [7-9].

The integration of gene expression, TF regulatory network and miRNA regulatory network analyses helps to study key genomic factors and, reveal the important interactions that lead the scientists and researchers to probe the depths of biological processes. The study of the Regulatory
mechanisms of the two main regulators is complex, but this lead to a deeper interpretation of biological processes [10].

A number of mathematical methods have been used to reveal the structure of the genetic network [11-13], here genes are the nodes and the interaction between genes represents the edges. Linear regression also used to infer the GRN [4, 13]. In [14] a novel approach of combine data from the TCGA project and ENCOD. this approach has ability to infer the relationships between miRNA, TF and gene using the activities of TF using the Encyclopedia of DNA Elements (ENCODE) TF binding data and estimated Regulatory activities . [15] Proposed regression method to search frequent miRNA-mRNA interactions using the TCGA data with different cancer types.

In this study, we combined three techniques, K-means, ANOVA test and linear regression. clustering and ANOVA used to filter and select significant genes from mRNA expression and miRNA expression datasets. The purpose of clustering is to form perfect clusters, which implies grouping genes based on their similarity [16-18].

Linear regression is a model to explain the relationships between co-regulators and the target genes in the GRNs and reveal the complex interplay in the gene network [19].

2. Datasets

the Breast Cancer (BRCA) dataset collected from the TCGA project. miRNA and mRNA expression profiles have a ratio of 331/711 and 331/13306 between samples and genes respectively. BRCA datasets were downloaded from http://cancergenome.nih.gov/ [20].

The data libraries TF2gene, TF2miRNA and mir2gene downloaded from http://www.scbit.org/cgrnb/download.htm, the two libraries of “TF2gene” and “TF2miRNA” from file ‘tfbsConsSites. txt’ and ‘tfbsConsFactors.txt’ obtained from UCSC hg19, http://genome.ucsc.edu/cgi-in/hgTables. The mir2gene file, obtained from http://starbase.sysu.edu.cn/, in addition to putative “mir2gene” regulation relationships [21, 22].

3. Methods

In this research, we combine three techniques to pre-processing datasets then integrating whole data to get Gene Regulatory network. Figure 1 shows the Structure of methodology.

3.1 Prepressing

The idea of pre-processing in order to avoid redundant data, exhaustive search, and yield unnecessary candidate genes. Firstly, mRNA expression and miRNA expression are clustered by K-means cluster, then, applied ANOVA test to select significant genes from clusters.
3.1.1 k-means

In the clustering process the choice of features is of extreme often is a goal for researchers, since different features might lead to different clustering results. Here, we are using k-means in the pre-processing phase of gene expression with ANOVA Test [23].

In this study, we are applying the K-means cluster with Euclidean distance (ED) when K = 3 and iteration = 100. Firstly, the initial clusters is constructed, and a center point assign to each cluster. The cluster center is constantly replaced by the mean point on the respective cluster until it reaches convergence.

Cluster objects based on minimum distance were used to determine the centroid base on the cluster average. The next step is to associate each point of the selected data set to the nearest centroid. Usually, the process of determining the appropriate number of clusters is ambiguous and confusing. Using “mclust” package to define the suitable number of clusters according to the Bayesian Information Criterion for expectation-maximization. Then, based on the “mclust” the package and evaluation models. Assigning three random points from that each dataset as initial centers. Based on the centers and datasets, the ED was computed from each point to the its center. The k-means aimed at minimizing an objective function, and the squared error function is as follows:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

$$\left\| x_i^{(j)} - c_j \right\|^2$$ is a measure the distance a points, $$c_j$$ is an indicator of the distance of the n, $$x_i^{(j)}$$ points from the centers.

3.1.2 ANOVA Test.

The idea in the analysis of variance (ANOVA) is, variability in quantity being measured (gene expression) can be partitioned into various identifiable sources like the experimental factors, and random noise [24].

In ANOVA can test whether the variance for factor, or more of factors, is statistically significant compared to the variability due to random sources. The clusters defined as a factor having three levels, cluster1, cluster2, and cluster3. Each cluster replicates according to the number of genes in it. Afterward, the ANOVA test was designed as a single factor and multi-level to define each gene. The significant genes were analyzed with p-value=0.01 in R programming. The model formula $$X \sim \text{groups}$$ equivalent the One Way ANOVA test, where x is content all of the groups.

3.2 Network modeling.

This work proposed approach uses the stepwise linear regression to uncover the gene network from gene expression [25, 26]. The GRN based on the forward prediction regulation relationship include miRNAs and TFs. Expression values contain 331 samples, 522 miRNAs and 11512 genes for breast cancer datasets. Firstly, we filtered out the regulation relationships that are not significant to the Intended target with a simple linear regression as in Equation 2, with TF and miRNA regulators and thier putative targets. $$Y_t$$ Modeled as a response variable t is a target (an mRNA or a miRNA), $$r$$ is the regulators (a TF or a MiRNA) for the targets, $$B_r$$ and $$A_r$$, are the values of expression level and regulatory efficacy of a regulator $$r$$. $$\alpha_0$$ and $$\varepsilon$$ is the Intercept and the residual error respectively, with normal distribution and mean equal zero.

$$Y_t = \alpha_0 + B_r A_r + \varepsilon$$

After a simple linear regression, the miRNAs were selected to the target regulation, with negative regulation efficacy $$A_r$$. Thereby, the putative relationships were dropped at $$\alpha = 0.05$$ for the TFs and miRNAs to reduce the number of putative regulatory relationships to a far large. Depending on the previous step, the remaining regulators for especial targets, we modeled it in the formal multivariate linear regression Equation 2. In Equation 3, the variables are independent, where $$Y_t$$, is the log2
Expression value of mRNA target, intercept, is a constant across all genes, $A_{\text{trg}}$ is the transcription factors (TFs) that regulate genes, $B_{\text{tg}}$ is the mRNA gene expression levels for the TFs. $A_{\text{mreg}}$, is the miRNA regulators that regulate genes, $B_{\text{mg}}$, is the mRNA gene expression levels for the miRNAs, and $\epsilon$ is the residual error.

$$Y_t = \alpha_0 + B_{\text{tg}} A_{\text{trg}} + B_{\text{mg}} A_{\text{mreg}} + \epsilon \quad (3)$$

The generalized regression function was used to fit a linear model of an outcome to one or more predictor variables. The expression levels in multi-regression formed by the regulators with $P \leq \alpha$ that came from a simple regression in Equation 3.

4. Results and Discussion

We got 522 MicroRNA expression and 11512 genes as a significant pattern by k-means cluster and ANOVA test under $P$-value=0.01. We observed the number of significant gene expression from the pre-processing process is not little; this was due to using the clustering process before the ANOVA test, where gene expression clusters from the clustering process are significantly enriched for particular functional. In clustering of Gene expression can exploration of the data, avoid getting lost between the thousands of genes [23]. During the ANOVA test to evaluate the significant genes from the clusters, the small p-value for most genes yielded several significant genes.

Moreover, when the number of clusters is more than three clusters, the significant genes will also be more. This is due to similar genes falling into the same cluster, thus tend to be more significantly enriched, and thus, the significant genes will be more. ANOVA had an impact on the selection of the significant genes and construction gene network to discover the relationship between the regulators and targets. We pruned the regulatory relationships with hope more smooth to predicted regulatory relationships. Then we made up the combinatorial network of transcriptional regulations and miRNA regulations in BRCA with the remaining regulator-target relationships. The GRN include a number of Edges and vertices table 1. Figure 2 shows subnetwork with regulation factors in vertices and the interactive in Edges.

Table 1. Vertices and Edges in Gene Regulatory Network

| Dataset         | Number Of Objects |
|-----------------|-------------------|
| **Vertices**    |                   |
| miRNAs          | 463               |
| Non-TF          | 9084              |
| TFs             | 9836              |
| Total           | 19383             |
| **Edges**       |                   |
| miRNA-Gene      | 323               |
| TF-microRNA     | 3109              |
| TF-Gene         | 35402             |
| Total           | 31970             |
We represented the network of the output of the linear regression process, where the linear regression process got the file represents all nodes and edges. Figure 2 presents 98 nodes as subnetwork from the whole network that contains 154 nodes.

The large sizes of nodes in GRN indicate that values of out-degree are large with the color like in CRBE1, E2F1, and MIA3. The small nodes indicate that values of out-degree are small with bright color. The arrows indicate the target gene, while the circle in the bottom of the edge indicates the source gene.

Figure 2. Sub-network

to estimated the Overall False Discovery (FDR), we randomly transposing values of expression data. We considered the result of regulation edges from randomizing data sets as false discovery. As we see in Table 2, the FDR rate of the gene regulatory network is 19.4%.

Table 2. FDR rate

| Datasets | Regulation type | Number of Edges | FRD % |
|----------|----------------|-----------------|-------|
| miRNA→gene | 323 | 30.1(± 0.8) |
| TF→gene | 3109 | 13.2(± -0.7) |
| TF→miRNA | 35402 | 20.6.1(+ -3.9) |
| Overall | 31970 | 19.1(+ - 0.7) |

5. Network Inference

The network inference discusses and displays the main Players in Breast cancer related to the gene regulatory network. Gene regulatory network produces a large number of interactions. We selected the common regulators, according to number of out-degrees, Betweenness, and closeness. Table 2 shows five regulators without-degrees and closeness.

CREB1 has 1010 with out-degree as the biggest number for regulating targets. In addition to Betweenness as the biggest number also (109084.8) and closeness is 7.85E-08 In the literature, CREB1 has a significantly high level in tumor tissues of breast cancer. Compared to benign breast tumors [27]. CREB is also involved in tumor initiation, progression, and metastasis, so it supporting as a proto-oncogene [28]. In prostate cancer noticed the CREB in status of over-expression and over-activation in the cancer tissues. In many cancer cell lines such as lung cancer and leukemia, CREB is
found as an inhibitor of cell proliferation and apoptosis, this makes CREB a promising cancer treatment [29, 30].

we found CREB1 has 15 miRNAs as targets, hsa-let-7d, hsa-mir-10b, hsa-mir-1281, hsa-mir-141, hsa-mir-196b, hsa-mir-200c, hsa-mir-202, hsa-mir-76b, hsa-mir-379, hsa-mir-10, hsa-mir-296, hsa-mir-44b, hsa-mir-629, hsa-mir-645 and hsa-mir-758. Researches indicate to hsa-miR-106 family has It has a share in the formation of some cancerous diseases such as stomach and breast cancer. [31, 32], also, hsa-let-7d, hsa-mir-200c, and hsa-mir-196b were found Implicated in breast cancer [33-37]. E2F1 also is important results in this study, E2F1 shown after CREB1, without-degree, and Betweenness, and it is regulating 986 of genes. Also, MIA3, ARNT, and AHR are important results in the network of BRCA based on the Human Protein Atlas (http://www.proteinatlas.org/), classified every one of the participants in breast cancer to varying degrees, MIA3 is classified more involvement than ARNT and AHR.

Table 3. Out-degrees, Betweenness, and closeness for each BRCA dataset

| Datasets | Regulator | Out-degree | Betweenness | Closeness   |
|----------|-----------|------------|-------------|-------------|
|          | CREB1     | 1010       | 109084.8    | 7.85E-08    |
|          | E2F1      | 986        | 67957.87    | 7.85E-08    |
| BRCA     | MIA3      | 980        | 6732.726    | 8.17E-08    |
|          | ARNT      | 518        | 18475.28    | 7.85E-08    |
|          | AHR       | 814        | 48707.82    | 7.85E-08    |

Figures 3 and 4 show the GO term and KEGG pathway for the BRCA dataset with the R package. The GO uses to perform enrichment analysis on gene sets. A set of genes that are up-regulated under certain conditions, an enrichment analysis will find which GO terms are over-represented (or under-represented) using annotations for that gene set. The KEGG pathway helps in this research help us to know towards potentially important genes, regulators, and discover complex patterns involving mutations, and gene expression data of various patient groups in the biological pathway context. Also, mutually exclusive genomic alteration patterns in a specific pathway can be generated.

The results show the most significant enrichments such as cell communication, regulation of the biologic process, biological regulation, cellular response and act as in figure 3. KEGG pathway enrichment analysis in figure 4 shows the most significant pathways in Pathways in cancer, p53 signaling pathway, Cell Cycle, leukemia cancer, in addition to other pathways By a different percentage. All of them were cancer-related signal pathways.

Figure 3. GO term BRCA module

Figure 4. KEGG Pathway BRCA module
We evaluated the performance of the method with the resulting of true positive and false-positive rates, which resulted in the corresponding ROC (Receiver operating characteristic) curves. We used the CytoNCA package in R [40], to calculate the values of the Area under Curve (AUC). Comparing the centrality measures, it includes true positive (TP), false positive (FP), f-measures (F) and accuracy (ACC). For the BRCA gene network, we got the 96% as true positive (TP), and 4% as a false positive (FP), also the F-measures was is 1, and the accuracy (ACC) is the 0.96.

6. Conclusion

This thesis proposed A Hybrid Approach based on, a statistical, data-driven way, using clustering, regression, and network inference techniques utilizing parallel expression datasets of miRNAs and mRNAs to build the gene regulatory network (GRN). Gene Regulatory Network (GRN) uncovered some of the regulators like AHR and ARNT that play significant roles in the regulation of cell proliferation, such as A549 cells, it is a known fact that A549 cells have been associated with cancerous diseases. The results also explained the relationship between regulators and their targets, several regulators found that involvement in cancer diseases such as CREB1, E2F1 also found the head-miR-106b and hsa-mir-200c as a target for CREB1 regulator. The combinatorial gene regulatory network included transcriptional regulations and miRNA regulations, permitting the investigation and study from the genes related to cancer. Through comparison our proposed method with others researches, in addition to medical literature that has been mentioned, we can say, the results we have obtained from the approach used in this work for Breast cancer represent a successful method and important tributary in the study of gene expression networks and investigation of genes related to cancer diseases. These results prove the effectiveness of this method for cancer-related genes identification and these cancer-related genes can be selected for further analysis.

7. Acknowledgments

This work is supported by Hunan Provincial Natural Science Foundation of China (Grant no.13JJ2017).

References

[1] Xing-Ming Zhao1, Ke-Qin, Liu, Guanghui Zhu4, Feng He1, Be atrice Du al, Jean-Michel Richer, De-Shuang Huang1, Chang-Jun Jiang1, Jin-Kao Hao, and Luoman Chen, "Identifying cancer-related microRNAs based on gene expression data," Bioinformatics, 2015.
[2] Hui Yu, Kang Tu, Yi-Jie Wang, Jun-Zhe Mao, Lu Xie, Yuan-Yuan Li2 and YiXue Li, "Combinatorial network of transcriptional regulation and microRNA regulation in human cancer," BMC Systems Biology, vol. 6:61, 2012.
[3] Bino John, Anton J. Enright, Alexei Aravin, Thomas Tuschl, Chris Sander, Debora S. Marks, "Human MicroRNA Targets," PLoS Biol, vol. 2, no. 11, 2004.
[4] Ramakrishna Mitra, Mick D. Edmonds, Jingchun Sun, Min Zhao, Hui Yu, Christine M. Eischen, and Zhongming Zhao, Reproducible combinatorial regulatory networks elucidate novel oncogenic microRNAs in non-small cell lung cancer," Bioinformatics, vol. 20, p. 1356–1368, 2014.
[5] Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM, A census of human transcription factors: function, expression, and evolution, Nature Reviews Genetics, vol. 10, p. 252–263, April 2009.
[6] Bartel DP, "MicroRNAs: Target Recognition and Regulatory Functions, Cell, vol. 136(2), p. 215–233, Jan 2009.
[7] Pedro M. Costa and Maria C. Pedroso de Lima, MicroRNAs as Molecular Targets for Cancer Therapy: On the Modulation of MicroRNA Expression. Pharmaceuticals 2013, 6(10), 1195-1220
[8] Tsang J, Zhu J, van Oudenaarden A. MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. Mol Cell. Elsevier; 2015; 26: 753±767.
[9] Martinez NJ, Ow MC, Barrasa MI, Hammell M, Sequerra R, Doucette-Stamm L, et al. A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity. Genes Dev. 2008; 22: 2535±2549. doi: 10.1101/gad.1678608 PMID: 18794350
[10] Thuc D Le1, Lin Liu, Bing Liu, Anna Tsykin, Gregory J Goodall, Kenji Satou, and Jiuyong Li, "Inferring microRNA and transcription factor regulatory networks in heterogeneous data," BMC Bioinformatics, vol. 14:92, 2013.

[11] I. Chebil, R. Nicolle, G. Santini, C. Rouveiro, and M. Elati, "Hybrid Method Inference for the Construction of Cooperative Regulatory Network in HuMan," IEEE Transactions on Nanobioscience, vol. 13, pp. 97-103, Jun 2014.

[12] Edmundo Bonilla-Huerta, Alberto Hernandez-Montiel, Roberto Morales Caporal, and Marco Arjona-Lopez, "Hybrid Framework Using Multiple-Filters and an Embedded Approach for an Efficient Selection and Classification of Microarray Data," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 13, Feb 2016.

[13] Zijian Dong, Tiecheng Song, Chuang Yuan, "Inference of Gene Regulatory Networks from Genetic Perturbations with Linear Regression Model," Plos one, vol. 8, no. 12, Dec 2013.

[14] Yue Li, Minggao Liang, Zhaolei Zhang, "Regression Analysis of Combined Gene Expression Regulation in Acute Myeloid Leukemia," PLOS Computational Biology, vol. 10, no. 10, Oct 2014.

[15] Jacobsen A, Silber J, Harinath G, Huse JT, Schultz N, et al, "Analysis of microRNA-target interactions across diverse cancer types," Nature Structural & Molecular Biology, vol. 20, p. 1325–1332, 2013.

[16] Yue Li, Minggao Liang, Zhaolei Zhang, "Regression Analysis of Combined Gene Expression Regulation in Acute Myeloid Leukemia," PLOS Computational Biology, vol. 10, no. 10, Oct 2014.

[17] Qian Ren, Xinjian Zhou, "Application of an Improved K-means Algorithm in Gene Expression Data Analysis," IEEE International Conference on Systems Biology (ISB), pp. 87-91, 2011.

[18] D. Asir Antony Gnana Singh1, A. Escalin Fernando2, E. Jebamalar Leavline3, Performance Analysis on Clusterin Approaches for Gene Expression Data, Vol. 5, Issue 2, February 2016.

[19] Cheng Liang, Yue Li, and Jiawei Luo, A Novel Method to Detect Function MicroRNA Regulatory Modules by Biclques Merging, EEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 13, no. 3, pp. 549-556, May-June 2016.

[20] Huayong Xu, Hui Yu, Kang Tu, Qianqian Shi, Chaochun Wei, Yuan Li, Yi-Xue Li, "cGRNB: a web server for building combinatorial gene regulatory networks through integrated engineering of seed-matching sequence information and gene expression datasets," BMC Systems Biology, 2013.

[21] Jian-Hua Yang, Jun-Hao Li, Peng Shao, Hui Zhou, Yue-Qin Chen, and Liang-Hu Qu, "Starbase: a database for exploring microRNA–mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data," Nucleic Acids Research, vol. 39, pp. 202-209, 2011.

[22] Patrik D’haeseleer, How does gene expression clustering work?, Nature Biotechnology, vol. 23, pp. 1499-1501, Dec 2005.

[23] Paul Pavlidis, Using ANOVA for gene selection from microarray studies of the nervous system, Elsevier, p. 282–289, April 2003

[24] Arsham Gahramani, Fiona M Watt, Nicholas M Luscombe, Generative adversarial networks simulate gene expression and predict perturbations in single cells, bioRXIV,2018.

[25] Benjamin J Lengerich Bryon Aragam Eric P Xing, Personalized regression enables sample-specific pan-cancer analysis, Bioinformatics, 2018, 34(13).

[26] Chhabra A1, Fernando H, Watkins G, Mansel RE, Jiang WG, "Expression of transcription factor CREB1 in human breast cancer and its correlation with prognosis," Oncology reports, vol. 18, no. 4, pp. 953-958, Oct 2007

[27] Xiangshu Xiao, Bingbing X. Li1, Bryan Mitton, Alan Ikeda, and Kathleen M Sakamoto, "Targeting CREB for cancer therapy: friend or foe," Curr Cancer Drug Targets, vol. 10(4), p. 384–391, June 2010.

[28] Deng X, Liu H, Huang J, Cheng L, Keller ET, Parsons SJ, Hu CD, "Ionizing radiation induces prostate cancer neuroendocrine differentiation through interplay of CREB and ATF2: implications for disease progression," Cancer Res, vol. 68 (23), p. 9663–9670, Dec 2008.
[29] Cekanova M, Majidy M, Masi T, Al-Wadei HA, Schuller HM, "Overexpressed Raf-1 and phosphorylated cyclic adenosine 3’-5’-monophosphate response element-binding protein are early markers for lung adenocarcinoma," Cancer, vol. 109(6), p. 1164–1173, 2007.
[30] Pan S, Yu F, Song E, "Tumor Invasion and Metastasis Initiated by mir-106b In Breast Cancer by Targeting BRMS1 and RB," Cancer Res, vol. 69, no. 24, Dec 2009.
[31] Petrocca F, Vecchione A, "Croce CM: Emerging role of miR-106b-25/miR-17-92 clusters in the control of transforming growth factor beta signaling," Cancer Res, vol. 68 (20), p. 8191–8194, 2008.
[32] Iorio MV1, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, Ménard S, Palazzo JP, Rosenberg A, Musiani P, Volinia S, Nenc I, Calin GA, Querzoli P, Negrini M, Croce CM, "MicroRNA gene expression deregulation in human breast cancer," Cancer Res, vol. 65(16), pp. 7065-7070, Aug 2005.
[33] Florian Kopp, Prajakta S. Oak, Ernst Wagner, Andreas Roidl, "miR-200c Sensitizes Breast Cancer Cells to Doxorubicin Treatment by Decreasing TrkB and Bmi1 Expression," Plos One, vol. 7, no. 11, Nov 2012.
[34] Ya-Ching Lu, Joseph T Chang, Err-Cheng Chan, Yin-Kai Cha o, TaSen Yeh, Jinn-Shiun Chen, Ann-Joy Cheng, "miR-196, an Emerging Cancer Biomarker for Digestive Tract Cancers," Journal of Cancer, vol. 7(6), pp. 650-655, 2016.
[35] Piyush B Madhamshettiwar, Stefan R Maetschke, Melissa J Davis, Antonio Reverter and Mark A Ragan, "Gene regulatory network inference: evaluation and application to ovarian cancer allow the prioritization of drug targets," Genome Medicine, vol. 4:41, May 2012.
[36] Abramova MV, Zatulovskiy EA, Svetlikova SB, Kukushkin AN, Pospelov VA, "e2f1 Gene is a new member of Wnt/beta-catenin/Tcf-regulated genes," Biochem Biophys Res Commun, vol. 391(1), pp. 142-146, 2010.
[37] Tang Y, Li M2, Wang J, Pan Y4, Wu FX5. CytoNCA: a Cytoscape plugin for centrality, analysis, and evaluation of protein interaction networks. Biosystems. 2015, 127:67-72.