International Research Journal of Biological Sciences _____ ISSN 2278-3202
Vol. **3(2),** 40-47, February **(2014)** *Int. Res. J. Biological Sci.*

Inferring Quercetin mediated miRNA-TF-gene Regulatory circuit using meta-analysis of Gene expression data

Dehury Budheswar and Sahu Jagajjit*
Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam-785013, INDIA

Available online at: www.isca.in, www.isca.me
Received 3rd October 2013, revised 30th November 2013, accepted 28th December 2013

Abstract

Both transcription factors (TFs) and microRNAs (miRNAs) are the key regulatory elements of genes at transcriptional and post-transcriptional levels. Though the mode of action these two elements vary significantly from each other, studies have shown that they can act upon genes in a combinatory network. In this study, we report an interaction map representing the combined regulatory effect of human TFs and miRNAs on the differentially expressed genes (DEGs) obtained from the meta-analysis of gene expression data based on quercetin mediated effect in human. Bioconductor packages were used with the help of R programming language to normalize and process three datasets: GSE7259, GSE15162 and GSE13899. Wilcoxon's t-test was performed with the help of JAVA Multi-Experiment Viewer program (MeV) and the results revealed a total of 605 unique DEGs at $p < 0.05$. Custom perl scripts searched 8 most common and frequent DEGs. BiNGO plug-in in Cytoscape analyzed the 8 DEGs and found 26 unique GO terms associated with 8 biological processes, 14 cellular components and 4 molecular functions. Bioinformatics tools identified 18 TFs for 7 DEGs except ATRX. The tools, miRanda and TargetScan identified 200 and 249 unique miRNAs targeting 8 DEGs and 9 TFs. The miRNA-TF-gene interaction map was constructed with the help of Cytoscape. The information obtained in this study provides insights into the dynamic characteristics of quercetin mediated miRNA-TF-gene circuit.

Keywords: Quercetin, bio conductor, wilcoxon's t-test, miRNA-TF-gene circuit, multi-experiment viewer, cytoscape.

Introduction

The mechanism of gene regulation is not understood fully till date as it involves numerous complex regulatory networks. However, transcription factors (TFs) and miRNAs have been found to act as the main regulatory elements in the transcriptional and post-transcriptional levels of gene expression respectively[1]. The global gene expression is regulated by TFs by binding to the cis-regulatory elements in the promoter regions[2,3]. In addition, TFs also participate in activation or repression of their target sequences by binding to specific region. The small non-coding endogenous miRNAs are ~22 nucleotides long RNA sequences generated from precursors that fold into hairpin structures[4]. A number of genes in both plant and animals are regulated by miRNAs. Both TF and miRNA regulatory networks were studied individually. Recent studies have found that the combinatory effect of TFs and miRNAs on the gene regulation is of immense importance[5,6].

Quercetin serves as important dietary component in many vegetables, foods, etc. and found to have a wide range of beneficial effects in several organisms, plants and animals including humans[7]. To know more about the effects of quercetin on different cells, numerous microarray studies have been carried out[8,9]. Microarray analysis is an exciting tool in molecular biology[10] which has played an indispensable role in the monitoring of gene expression profiling by the analysis of large datasets of thousands of genes in one run. The statistical analysis of the raw data produced plays a crucial role in these kinds of experiments. With the increased use of improved microarray technology huge amount of data is being generated and stored in different databases[11]. NCBI has launched Gene Expression Omnibus (GEO)[12] database to support the public use and dissemination of gene expression data from any organism or artificial source. It contains raw data for all types of microarray experiments which are available free for downloading. But these datasets are still needed to be more generalized to find out most important information regarding expressions of common genes. This can be done with the help of meta-analysis of these data using statistical methods[13]. Several computational algorithms and programs are available which provides the researchers an easy way to accomplish this statistical analysis. R and Bio-conductor provide the platform to use various statistical methods for microarray data analysis[14].

The present study was undertaken to uncover the latent information regarding the regulatory network of genes which are differentially expressed due to an over-supplement of quercetin. Meta-analysis was carried out based on the fact that letting all the available quercetin microarray experiment data processed and analyzed simultaneously with the help of bioinformatics approaches can put light into the most important genes, the pathways involved with these genes, gene networks and many other information. The combinatory network of TFs and miRNAs in this study can put us closer towards the understanding of complex gene regulatory systems.

Material and Methods

Data Collection, Normalization: The raw data were collected for three datasets: GSE7259, GSE15162 and GSE13899 from the parental database GEO at NCBI which are the Affymetrix Human Genome U133 Plus 2.0 Array CEL files resenting the effect of Quercetin on the gene expression in human. The raw affymetrix CEL files were normalized using R programming codes in the R Console. Bioconductor project provides the 'affy' package[15] which implements routines for several summarization algorithms, including the Robust Multiarray Average (RMA)[16].

Expression Profiling: The normalized data were loaded onto a JAVA Multi-Experiment Viewer program (MeV) with several features for reading expression data and perform statistical analysis[17]. Welch's t-test was carried out individually for the three normalized datasets. The differentially expressed genes (DEGs) detected from the t-test were compared using an in-house perl script and most frequent DEGs were identified. The bio-molecular interaction networks were studied for the DEGs with the help of is open source software, Cytoscape[18]. Cytoscape provides numerous plug-ins to study the functions, pathways, etc. vividly. In this study, BiNGO plug-in was used primarily to determine the GO terms and pathways associated with the DEGs[19].

Construction of miRNA-TF-gene network: The Fasta sequences for all the available miRNAs of *Oryza sativa* were collected from miRBase. For the prediction of miRNAs targeting the DEGs, two dynamic programs *i.e.*, miRanda and TargetScan were used. Top 10 microRNAs were selected from different families based on the number of DEGs, they target. Using bioinformatics approaches, the TF binding sites for all the DEGs were identified. Again the best 10 miRNAs were targeted against the identified TFs. All the miRNA-TFs, miRNAs-genes and TF-genes were integrated using Cytoscape to draw the network.

Results and Discussion

Meta-analysis of gene expression data: The normalization process resulted with reduction of noise in the effect of systematic error that occurs while retaining full biological variation. The t-test on the normalized data identified a total of 605 unique differentially expressed genes at $p < 0.05$ which were dramatically up- or down-regulated with the over supplement of Quercetin (table-1). Highest number of DEGs *i.e.*, 422 was observed in GSE7259 followed by GSE13899 and GSE15162 with 139 and 44 number of DEGs. The volcano plots (figure-1) represents the up- and down-regulation of significant genes obtained from t-test. A comparison between all the DEGs found from the 3 dataset helped to identify 8 commonly expressed most frequent DEGs. In figure-2, a Venn diagram representing the common genes in between 3 datasets have been shown. BiNGO plug-in in Cytoscape revealed a total of 26 unique GO terms associated with the 8 DEGs. The GO terms found are associated with 8 biological processes, 14 cellular components and 4 molecular functions (table-2). Various statistical parameters including GO terms, description, GO classification, p-value, corrected p-value, cluster frequency, total frequency and genes in the respective clusters have been shown in the table-2.

Table-1
T-test result summary of three datasets

GEO Datasets	No of Samples	No of Control Samples	No of Treated Samples	No of Unique DEGs
GSE7259	8	4	4	422
GSE13899	6	3	3	139
GSE15162	8	4	4	44

Table-2
Clustering and GO annotation of the DEGs

GO-ID	Description	GO classification	p-value	Corr. p-value	Cluster freq	Total freq	Genes
GO:0005720	Nuclear hetero chromatin	CC	9.85E-05	3.09E-02	2/8 25.0%	34/17791 0.1%	ATRX H3F3B
GO:0000*792*	hetero chromatin	CC	2.14E-04	3.09E-02	2/8 25.0%	50/17791 0.2%	ATRX H3F3B
GO:0000*790*	Nuclear chromatin	CC	2.99E-04	3.09E-02	2/8 25.0%	59/17791 0.3%	ATRX H3F3B
GO:0015266	Protein channel activity	MF	4.50E-04	3.09E-02	1/8 12.5%	1/17791 0.0%	MCL1
GO:0014895	Smooth muscle hypertrophy	BP	4.50E-04	3.09E-02	1/8 12.5%	1/17791 0.0%	CYBA
GO:0019904	protein domain-specific binding	MF	5.31E-04	3.09E-02	3/8 37.5%	388/17791 2.1%	ATRX CYBA SH3BGRL

GO:0017124	SH3 domain binding	MF	8.57E-04	3.14E-02	2/8 25.0%	100/17791 0.5%	CYBA SH3BGRL
GO:0002268	Follicular dendritic cell differentiation	BP	8.99E-04	3.14E-02	1/8 12.5%	2/17791 0.0%	NFKB2
GO:0002266	Follicular dendritic cell activation	BP	8.99E-04	3.14E-02	1/8 12.5%	2/17791 0.0%	NFKB2
GO:0033257	Bcl3-p52 complex	CC	8.99E-04	3.14E-02	1/8 12.5%	2/17791 0.0%	NFKB2
GO:0017004	Cytochrome complex assembly	BP	1.35E-03	4.09E-02	1/8 12.5%	3/17791 0.0%	CYBA
GO:0044454	Nuclear chromosome part	CC	1.49E-03	4.09E-02	2/8 25.0%	132/17791 0.7%	ATRX H3F3B
GO:0014805	smooth muscle plasticity	BP	1.80E-03	4.09E-02	1/8 12.5%	4/17791 0.0%	CYBA
GO:0001740	Barr body	CC	1.80E-03	4.09E-02	1/8 12.5%	4/17791 0.0%	H3F3B
GO:0033256	I-kappaB/NF-kappaB complex	CC	1.80E-03	4.09E-02	1/8 12.5%	4/17791 0.0%	NFKB2
GO:0002467	germinal center formation	BP	2.25E-03	4.09E-02	1/8 12.5%	5/17791 0.0%	NFKB2
GO:0014896	muscle hypertrophy	BP	2.25E-03	4.09E-02	1/8 12.5%	5/17791 0.0%	CYBA
GO:0043231	intracellular membrane-bounded organelle	CC	2.35E-03	4.09E-02	8/8 100.0%	8349/17791 46.9%	ATRX CYBA SH3BGRL MCL1 ATP5C1 H3F3B NFKB2 RBBP6
GO:0000228	nuclear chromosome	CC	2.36E-03	4.09E-02	2/8 25.0%	167/17791 0.9%	ATRX H3F3B
GO:0043227	membrane-bounded organelle	CC	2.36E-03	4.09E-02	8/8 100.0%	8356/17791 46.9%	ATRX CYBA SH3BGRL MCL1 ATP5C1 H3F3B NFKB2 RBBP6
GO:0070087	chromo shadow domain binding	MF	2.70E-03	4.09E-02	1/8 12.5%	6/17791 0.0%	ATRX
GO:0000275	mitochondrial proton-transporting ATP synthase complex, catalytic core F(1)	CC	2.70E-03	4.09E-02	1/8 12.5%	6/17791 0.0%	ATP5C1
GO:0000805	X chromosome	CC	2.70E-03	4.09E-02	1/8 12.5%	6/17791 0.0%	H3F3B
GO:0045261	hydrogen-transporting ATP synthase, F1 sector	CC	3.14E-03	4.57E-02	1/8 12.5%	7/17791 0.0%	ATP5C1
GO:0000785	Chromatin	CC	3.43E-03	4.79E-02	2/8 25.0%	202/17791 1.1%	ATRX H3F3B
GO:0050665	hydrogen peroxide biosynthetic process	BP	3.59E-03	4.82E-02	1/8 12.5%	8/17791 0.0%	CYBA

* BP: Biological Processes; CC: Cellular Components; MF: Molecular Functions

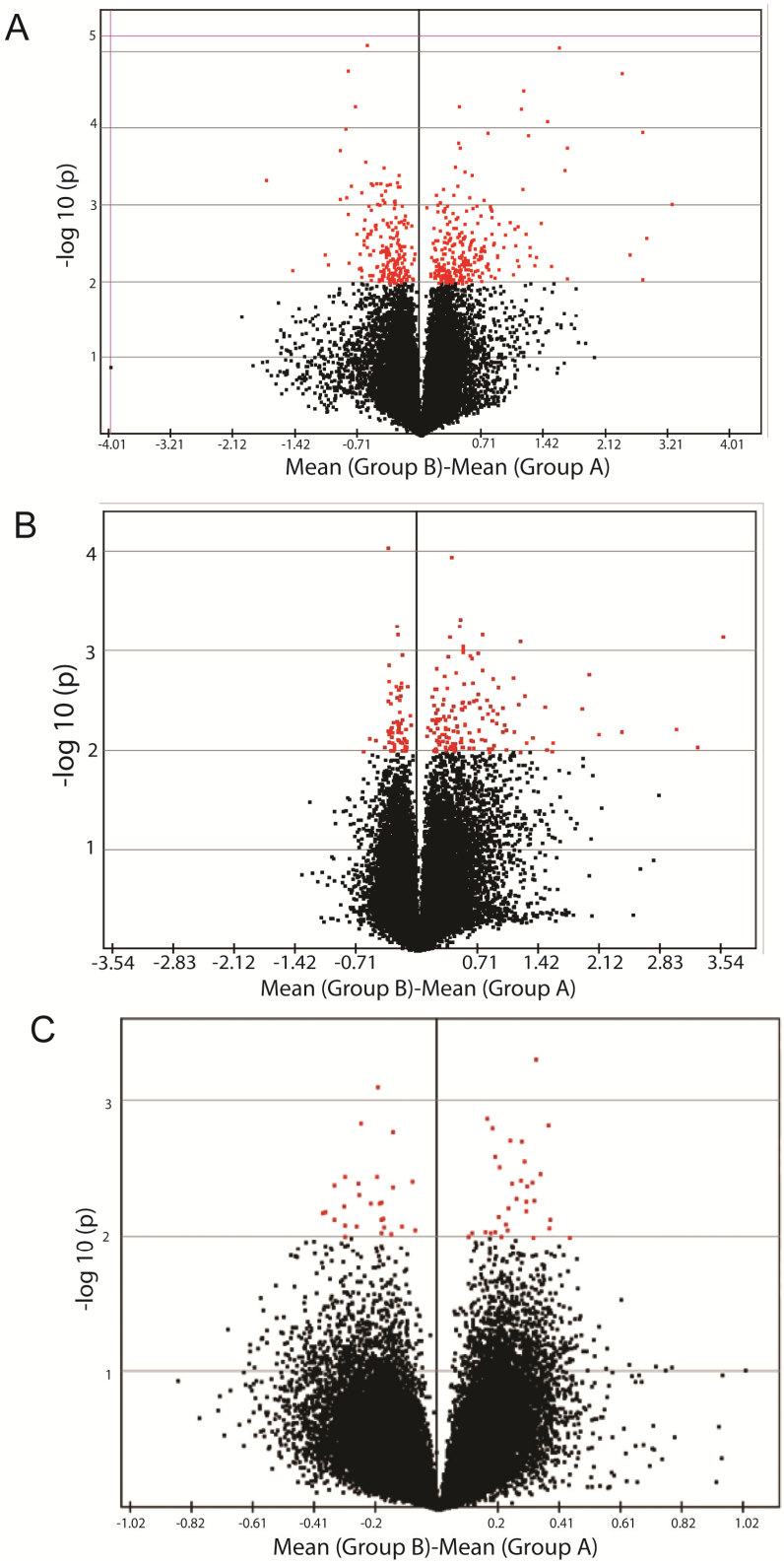

Figure-1

Volcano plot for the differentially expressed genes detected using t-test from the 3 microarray datasets: a) GSE7259, b) GSE13899 and c) GSE15162

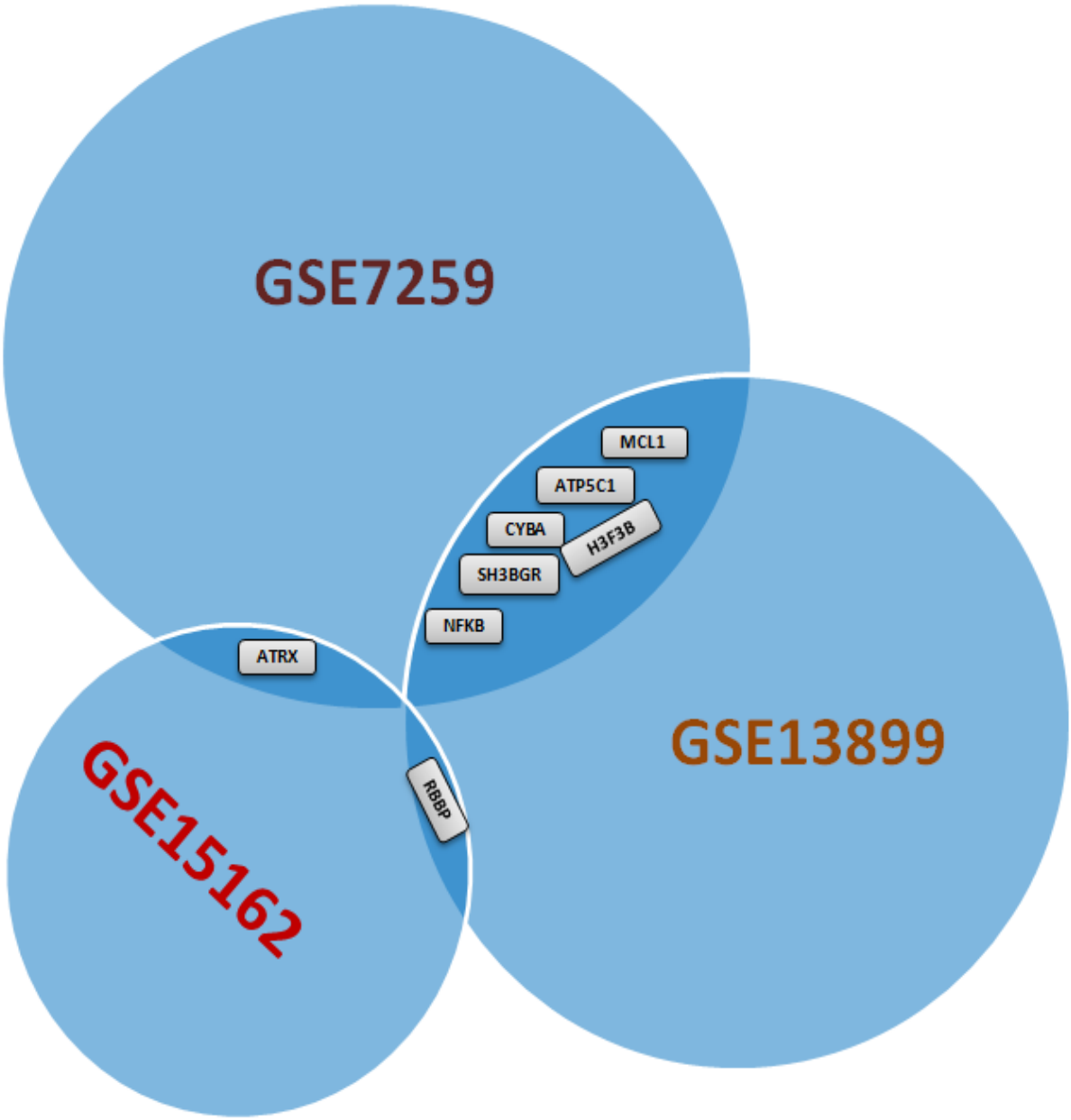

Figure-2
Venn diagram representations of DEGs in three datasets

miRNA-TF-gene Network construction: A combined use of miRanda and TargetScan lead to the detection of 200 unique miRNAs targeting 8 DEGs. A total of 18 unique TFs were identified for all the 7 DEGs except ATRX. All the 18 TFs were searched for miRNA target sites which resulted 249 unique miRNAs targeting 9 TFs. The 9 TFs having miRNA target sites are: CTCF, E2F1, EGR1, SP1, YY1, MAX, POU2F2, SRF and USF1. Table-3 shows the information regarding the top 10 miRNAs selected, 9 TFs and 8 genes detected. With the help of Cytoscape tool, an interaction map was constructed which is presented in figure-3. None of the 9 TFs was found to regulate ATRX and ATP5C1. Though NFKB2 is most frequently targeted gene by 5 TFs out of 9, no miRNA out of the 10 selected was found to target this gene. RBP6 is found to be targeted by all the 10 miRNAs. The miRNA; hsa-miR-590-3p had taken part in most number of regulations in case of genes as well as TFs. A total of 7 TFs and 4 genes were detected which can be targeted by hsa-miR-590-3p. The gene CYBA was found to be targeted only by hsa-miR-106a.

Table-3
List of miRNAs (from the list of top 10) and transcription factors regulating DEGs

Frequently common DEGs	miRNA IDs targeting the DEGs	TFs targeting the DEGs
ATRX	hsa-miR-32, hsa-miR-320a, hsa-miR-144, hsa-miR-410, hsa-miR-26b, hsa-miR-590-3p, hsa-miR-495, hsa-miR-1297	No hits
ATP5C1	hsa-miR-32, hsa-miR-26b, hsa-miR-590-3p, hsa-miR-495, hsa-miR-1297	Cjun, Cfos, Gabp
CYBA	hsa-miR-106a	SP1, Ap2gamma, Max, Gabp, Ap2alpha, Cmyc
RBBP6	hsa-miR-32, hsa-miR-320a, hsa-miR-144, hsa-miR-410, hsa-miR-26b, hsa-miR-590-3p, hsa-miR-495, hsa-miR-1297, hsa-miR-106a, hsa-miR-128	Max, USF1
H3F3B	hsa-miR-320a, hsa-miR-410, hsa-miR-128	SP1, Yy1, Cjun, Egr1
SH3BGRL	hsa-miR-144, hsa-miR-590-3p	SP1, Egr1
MCL1	hsa-miR-32, hsa-miR-320a, hsa-miR-144, hsa-miR-410, hsa-miR-26b, hsa-miR-495, hsa-miR-1297, hsa-miR-106a, hsa-miR-128	SP1, Yy1, Cjun, Cfos, PU1, Srf
NFKB2	No hits	SP1, Yy1, Egr1, Ap2gamma, Ap2alpha, PU1, CTCF, EBF, E2F1NFKB

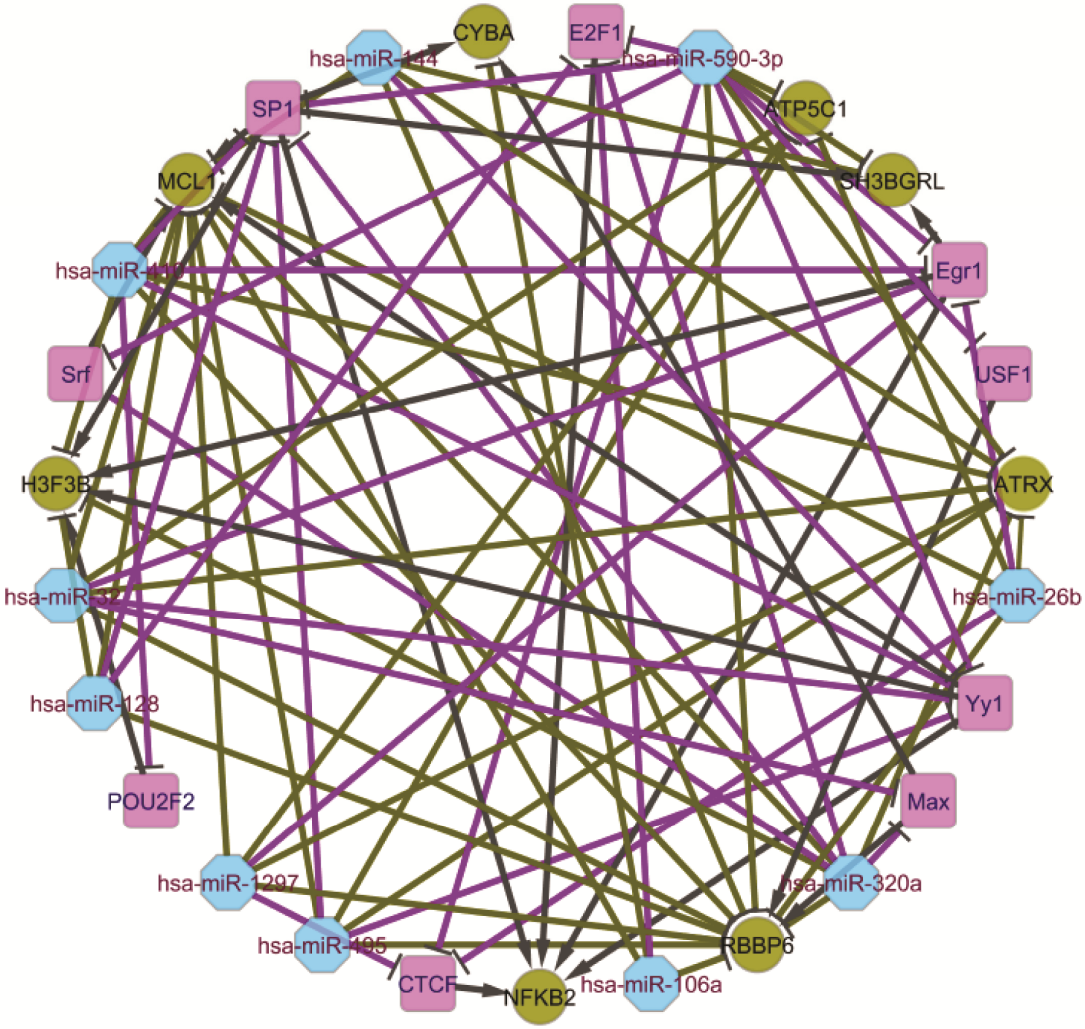

Figure-3
miRNA-TF-DEG combinatory circuit developed using Cytoscape ver. 2.8.2

During the past few years, scientists have made efforts to explore the gene regulations with the help of combinatory regulation of miRNAs and TFs[5,6,20-23]. These regulatory networks are believed to be of immense importance in extracting information in case of differentially expressed genes. As of now, several studies involving meta-analysis have been carried out with the publicly available microarray data to uncover the significant genes which are dramatically up- and down-regulated[24,25,26]. The datasets chosen for this study represents the gene expression profiling with the over-supplement of quercetin[8,9]. Quercetin has already been proved to prevent cancer in the previous studies[7] and has many other health effects upon intake it as a supplement. The present study was an attempt to study most frequent DEGs and their regulatory network. The miRNA hsa-miR-590-3p is found to be the most important which participated in regulation of maximum number of DEGs and TFs. The location of hsa-miR-590-3p is o the chromosome-7 at the position 73,603,528-73,607,624 bps.

The hsa-miR-590-3p had been found to be actively involved in other diseases mainly Alzheimer[27]. In all the TFs participating in the network, SP1 is the one which regulates highest *i.e.*, 5 number of DEGs and 5 out of 10 miRNAs targets SP1. The SP1 gene encodes zinc finger transcription factor which bind to GC-rich motifs of a number of promoters thereby involved in many cellular processes, including cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodelling. Studies found that an increase in expression level of SP1 TF can inhibit cell cycle progression leading to induce apoptosis[28].

Our study shows that hsa-miR-590-3p also targets SP1 transcription factor. SH3BGRL is found to be the only gene regulated by both hsa-miR-590-3p and SP1. Co-expression of SH3BGRL with v-rel oncogene can reduce v-Rel-expressing fibroblasts, lymphoid cells, and splenic tumour cells[29]. The combinatory effect of hsa-miR-590-3p and SP1 can be further studied extensively to uncover many other regulatory networks in several other diseases.

Conclusion

In the present study, we report a computational scaffold for construction of miRNA-TF-DEG interaction map for the DEGs obtained from the meta-analysis of gene expression data representing studies related to quercetin effect in human. The information obtained in this piece of work will help in getting into the combinatory make up of miRNA and TF in gene regulations. Again the quercetin effect shows that the whole network designed here is totally mediated by quercetin itself. The report also reveals several miRNA and TFs which are of equal importance though these are not a part of the network. The combinatory circuits of miRNA and TFs can be further studied to uncover valuable information which will lead to open ways for treatment of several diseases.

References

1. Arora S., Rana R., Chhabra A., Jaiswal A. and Rani V., miRNA-transcription factor interactions: a combinatorial regulation of gene expression, *Mol Genet Genomics.,* **288,** 77-87 **(2013)**

2. Vaquerizas J.M., Kummerfeld S.K., Teichmann S.A. and Luscombe N.M., A census of human transcription factors: function, expression and evolution, *Nat Rev Genet.,* **10,** 252-63 **(2009)**

3. Lee T.I., Rinaldi N.J., Robert F., Odom D.T., Bar-Joseph Z., Gerber G.K., Hannett N.M., Harbison C.T., Thompson C.M., Simon I., Zeitlinger J., Jennings E.G., Murray H.L., Gordon D.B., Ren B., Wyrick J.J., Tagne J.B., Volkert T.L., Fraenkel E., Gifford D.K. and Young RA, Transcriptional regulatory networks in *Saccharomyces cerevisiae*, *Science.,* **298,** 799-804 **(2002)**

4. Bartel D.P., MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell.,* **116,** 281-297 **(2004)**

5. Sun J., Gong X., Purow B. and Zhao Z., Uncovering MicroRNA and Transcription Factor Mediated Regulatory Networks in Glioblastoma, *PLoS Comput Biol.,* **8,** e1002488 **(2012)**

6. Le T.D., Liu L., Liu B., Tsykin A., Goodall G.J., Satou K. and Li J., Inferring microRNA and transcription factor regulatory networks in heterogeneous data, *BMC Bioinformatics.,* **14,** 92 **(2013)**

7. Sharmila G., Bhat F.A., Arunkumar R., Elumalai P., Raja S.P., Senthilkumar K. and Arunakaran J., Chemopreventive effect of quercetin, a natural dietary flavonoid on prostate cancer in *in vivo* model, *Clin Nutr.,* pii:S0261-5614, 00235-5 **(2013)**

8. Dihal A.A., Tilburgs C., van Erk M.J., Rietjens I.M., Woutersen R.A. and Stierum R.H., Pathway and single gene analyses of inhibited Caco-2 differentiation by ascorbate-stabilized quercetin suggest enhancement of cellular processes associated with development of colon cancer, *Mol Nutr Food Res.,* **51,** 1031-1045 **(2007)**

9. Boomgaarden I., Egert S., Rimbach G., Wolffram S., Müller M.J. and Döring F., Quercetin supplementation and its effect on human monocyte gene expression profiles *in vivo*, *Br J Nutr.,* **104,** 336-345 **(2010)**

10. Brown P.O. and Botstein D., Exploring the new world of the genome with DNA microarrays, *Nat. Genet.,* **21,** 33-37 **(1999)**

11. Moreau Y., Aerts S., De M.B., De S.B. and Dabrowski M., Comparison and meta-analysis of microarray data: from the bench to the computer desk, *Trends Genet.,* **19,** 570-577 **(2003)**

12. Barrett T., Suzek T.O., Troup D.B., Wilhite S.E., Ngau W.C., Ledoux P., Rudnev D., Lash A.E., Fujibuchi W. and

Edgar R., NCBI GEO: mining millions of expression profiles--database and tools, *Nucleic Acids Res.,* **33,** D562-566 **(2005)**

13. de Magalhaes J.P., Curado J. and Church G.M., Meta-analysis of age-related gene expression profiles identifies common signatures of aging, *Bioinformatics.,* **25,** 875-881 **(2009)**

14. Gentleman R., Carey V.J., Huber W., Irizarray R. and Dudoit S. eds., Statistics for Biology and Health: Bioinformatics and Computational Biology Solutions Using R and Bioconductor, New York, New York, USA, Springer Publishing, 229-248, **(2005)**

15. Gautier L., Cope L., Bolstad B.M. and Irizarry R.A., affy–analysis of Affymetrix Gene Chip data at the probe level, *Bioinformatics.,* **20,** 307-315, **(2003)**

16. Irizarry R.A., Hobbs B., Collin F., Beazer-Barclay Y.D., Antonellis K.J., Scherf U. and Speed T.P., Exploration, normalization, and summaries of high density oligonucleotide array probe level data, *Biostatistics.,* **4,** 249-264, **(2003)**

17. Chu V.T., Gottardo R., Raftery A.E., Bumgarner R.E. and Yeung K.Y., MeV+R: using MeV as a graphical user interface for Bioconductor applications in microarray analysis, *Genome Biol.,* **9,** R118 **(2008)**

18. Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B. and Ideker T., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.,* **13,** 2498-2504, **(2003)**

19. Maere S., Heymans K. and Kuiper M., BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks, *Bioinformatics.,* **21,** 3448-3449, **(2005)**

20. Qiu C., Wang D., Wang E. and Cui Q., An upstream interacting context based framework for the computational inference of microRNA functions, *Mol Biosyst.,* **8,** 1492-1498, **(2012)**

21. Jiang W., Zhang Y., Meng F., Lian B., Chen X., Yu X., Dai E., Wang S., Liu X., Li X., Wang L. and Li X., Identification of active transcription factor and miRNA regulatory pathways in Alzheimer's disease, *Bioinformatics.,* **29 (20)**, 2596-602 **(2013)**

22. Iwama H., Coordinated networks of microRNAs and transcription factors with evolutionary perspectives, *Adv Exp Med Biol.,* **774,** 169-187, **(2013)**

23. Aguda B.D., Modeling microRNA-transcription factor networks in cancer, *Adv Exp Med Biol.,* **774,** 149-167 **(2013)**

24. Tseng G.C., Ghosh D. and Feingold E., Comprehensive literature review and statistical considerations for microarray meta-analysis, *Nucleic Acids Res.,* **40,** 3785-3799, **(2012)**

25. Moreau Y., Aerts S., De M.B., De S.B. and Dabrowski M., Comparison and meta-analysis of microarray data: from the bench to the computer desk, *Trends Genet.,* **19,** 570-577 **(2003)**

26. Meyer F.R., Grausgruber H., Binter C., Mair G.E., Guelly C., Vogl C. and Steinborn R., Cross-platform microarray meta-analysis for the mouse jejunum selects novel reference genes with highly uniform levels of expression, *PLoS One.,* **8,** e63125, **(2013)**

27. Villa C., Fenoglio C., De R.M., Clerici F., Marcone A., Benussi L., Ghidoni R., Gallone S., Cortini F., Serpente M., Cantoni C., Fumagalli G., Martinelli B.F., Cappa S., Binetti G., Franceschi M., Rainero I., Giordana M.T., Mariani C., Bresolin N., Scarpini E. and Galimberti D., Role of hnRNP-A1 and miR-590-3p in neuronal death: genetics and expression analysis in patients with Alzheimer disease and frontotemporal lobar degeneration, *Rejuvenation Res.*, **14,** 275-281, **(2011)**

28. Deniaud E., Baguet J., Chalard R., Blanquier B., Brinza L., Meunier J., Michallet M.C., Laugraud A., Ah-Soon C., Wierinckx A., Castellazzi M., Lachuer J., Gautier C., Marvel J. and Leverrier Y., Overexpression of transcription factor Sp1 leads to gene expression perturbations and cell cycle inhibition, *PLoS One*., **4,** e7035 **(2009)**

29. Majid S.M., Liss A.S., You M. and Bose H.R., The suppression of SH3BGRL is important for v-Rel-mediated transformation, *Oncogene*., **25,** 756-768, **(2006)**