

Identification and Analysis of Disease Target Network of Human MicroRNA and Predicting Promising Leads for ZNF439, a Potential Target for Breast Cancer

Vaijayanthi Raghavan and D. B. Manasa

Abstract—It has become increasingly evident that many diseases are linked to microRNAs and the role these biomolecules play in the development of numerous cancers. Thus, Biological network analysis offers novel approach in understanding basic mechanisms controlling normal biological processes and disease pathologies. Computational study of the disease-associated networks of human microRNAs is one of the means that provide complementary information. In the present study we predicted the disease association of all human microRNA predicted targets and the most promising leads were identified for the most potential target in breast cancer. The comprehensive list of human microRNAs was downloaded and the target genes of individual microRNA were retrieved based on miTG score using target prediction program. The resultant network was investigated based on the association of nodes with a relevant pathway, disease and pathological event. Further all the targets in the network for subjected for phylogenetic analysis and a comparative study was performed for these targets. The pattern of the conserved region was derived by Prosite analysis. Structure-based virtual screening involves docking of screened compounds and the protein target and the leads were identified based on the affinity and free energy of binding values. After extensive research, breast cancer was identified as the most significant disease and ZNF439 was identified as the most promising drug target in the network that was under study and the compounds such as Indolocarbozole, Camptothecin, Lucidenic Acid, Quercetin and Staurosporine were predicted as promising leads for ZNF439.

Index Terms—MicroRNA (miRNA), ZNF439, phylogenetic analysis, structure-based virtual screening.

I. INTRODUCTION

MicroRNAs (miRNAs) are single-stranded RNA molecules of approximately 22 nucleotides in length and the level of translation is regulated by gene expression. microRNAs binds to the 3' untranslated region (3'UTR) of target mRNAs by mediating posttranscriptional regulation of protein-coding genes and their main function is to downregulate gene expression including translational repression, mRNA cleavage and deadenylation. Thus miRNAs play an important role in fine-tuning the diverse

cellular functions such as development, differentiation, proliferation, apoptosis and metabolism [1].

Zinc finger proteins are encoded with zinc mediated nucleic acid binding proteins. The ZNFs are abundantly found in eukaryotic genomes and their main functions are deoxyribonucleic acid (DNA) recognition, protein folding and assembly, transcriptional activation, lipid binding, regulation of apoptosis and ribonucleic acid (RNA) packaging, the structure of ZNF439 is as shown in Fig. 1.

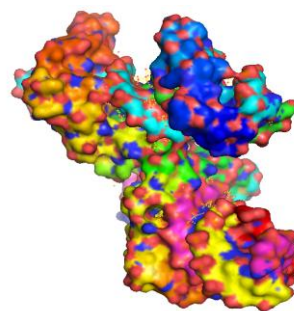


Fig. 1. Structure of ZNF439.

In general, zinc finger protein consists of anti-parallel hairpin motif. It consists of one alpha helix, two beta strands and a hairpin structure. The first Zn binds to Cys-3 of 1st beta strand and second Zn binds to Cys-6 that separates the two beta strands. The remaining two Zn is coordinated to Cys 19 & 23 which are present in C-terminal half of the alpha helix and the hairpin is followed by alpha helix of about 3.5 turns. Zn is present in the interior of the protein to form a stable finger structure in aqueous solution. The zinc does not interact with DNA.

II. MATERIAL AND METHODS

A. Network Analysis

The human miRNAs was extracted from miRBASE Release 17 [2]. The target genes of individual miRNA were explored using Diana-microT 3.0 target prediction program [3]. Highly reliable targets were selected based on the miTG score ≥ 20 , because the targets exhibited significantly a lower precision score with the miTG score < 20 . The gene names of miRNA target genes were uploaded onto Cytoscape [4]. Reactome FIs and Nemo are the Cytoscape plugins. Reactome FIs plugin constructs human miRNA gene interaction network based on the gene set analysis and Nemo plugin analyze the biological network of miRNA and reveals the ranking based on the clustered network.

Manuscript received August 10, 2012; revised September 20, 2012. This work was partially carried out at Institute of Computational Biology (IOCB), Bangalore, India.

V. Raghavan is with the Peoples Education Society Institute of Technology, Department of Biotechnology, 100 ft ring road, Banashankari 3rd stage, Bangalore, 560085 India (e-mail: vr_94@pes.edu).

D. B. Manasa was with the Peoples Education Society Institute of Technology, Department of Biotechnology, 100 ft ring road, Banashankari 3rd stage, Bangalore, 560085 India (e-mail: manasadb@gmail.com).

B. Target Identification

After finding the pathway, disease and pathological event for an individual gene ID, the number of occurrences of pathway, disease and pathological event was analyzed. The pathway of a particular Gene ID was analyzed through Superpathways, a Cytoscape plugin and some through literature review.

The pathological event for each gene ID was analyzed and compared with the published literature review and also the number of times it occurred during the analysis of different rankings of pathway, disease and pathological event. The results of biological networking revealed that the rank one pathway as p53 signaling pathway, the rank one disease and the rank one pathological event as breast cancer [5].

The gene names responsible for Breast cancer were ZNF287, ZNF445, ZNF44, ZNF248, ZNF264, ZNF587, ZNF275, ZNF583, ZNF780B, ZNF805, ZNF439, ZIM2, ZNF283 and ZNF594. The network analysis of 14 breast cancer genes was performed using VisANT 3.93[6]. The breast cancer genes were loaded and the biomolecular interactions of the genes were analysed.

C. Multiple Sequence and Phylogenetic Analysis

The complete sequences of all the Zinc finger proteins was retrieved from UniprotKB/Swissprot database and submitted to Clustal W program. The multiple alignment obtained in Phylip format was then subjected to PROTDIST program and NJ plot of Phylip software. The evolutionary tree was generated using DRAWTREE program and the relationship between these target proteins of miRNAs was analyzed.

D. Detection of Pattern Corresponding to the Conserved Regions in the Group of MiRNA Targets

The set of protein sequences were then submitted to PROSITE database and conserved regions were analysed. These 14 genes were submitted as two groups and the results analysed.

E. Virtual Screening

A small molecule library of anti-cancer agents for breast cancer containing 191 compounds was created. Lipinski's rule of five was chosen as compound filter for virtual screening of this library using ZNF439 as the protein target and hence 122 screened compounds were qualified. The successful compounds were subjected to energy minimization using Marvin Sketch. Structure-based virtual screening using AutoDock Vina was then performed on these minimum energy conformers [7]. The compounds were ranked based on their docking scores. Top ten compounds were redocked with Hex 6.3 and obtained free energy of binding results (TABLE I) [8]. The Binding sites of these ten compounds were predicted using Q-SiteFinder and hydrogen bond interaction was utilized as a filter for the post processing of docking results, out of which five compounds showed Hydrogen bond interactions using PyMOL [9] as shown in the below in TABLE II and Fig. 8 a, b, c, d & e. Hence, these five compounds were proposed to be the best Lead molecules for ZNF439 with their predicted half maximal inhibitory concentration (IC₅₀) value with EcoliTox being reported [10].

III. RESULTS

The score of clustered network was consistent from rank 1 to rank 7 (i.e., *score*= 49.872). Considering the gene IDs from rank1 to rank7 for finding the pathway, disease and pathological event, the biological networking of human miRNA (Fig. 2) is an undirected graph consisted of nodes and edges where each node represents one miRNA and an edge represents connection between two nodes (miRNAs).

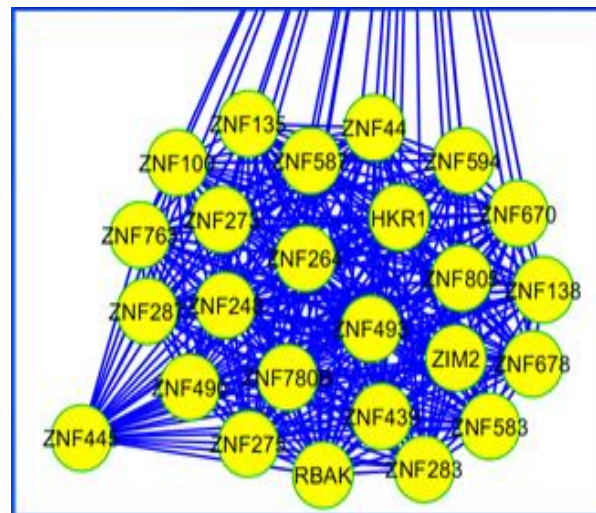


Fig. 2. Biological networking of human miRNA.

The results of biological networking revealed that the rank one pathway as p53 signaling pathway, the rank one disease and the rank one pathological event as breast cancer. The network analysis of breast cancer genes was performed using VisANT and the most interacting genes which is predominantly influencing for breast cancer was ZNF439 (PDBID: 2I13) [6]. The analysis results of pathway, disease and pathological event are as represented graphically below in the Fig. 3, Fig. 4, and Fig. 5.

It was found that ZNF439 was very closely related to ZIM2 and the most distantly related protein was ZNF594. However these sequences were further subjected to Prosite analysis for detecting the pattern in order to identify the conserved amino acid residues. (Fig. 6)

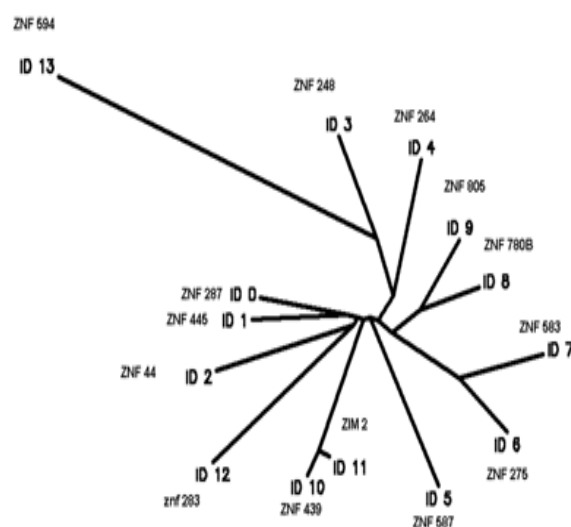


Fig. 6. Evolutionary tree of all the miRNA targets in the pathway.

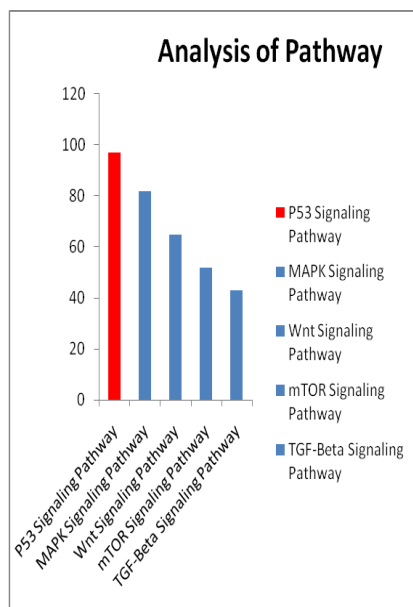


Fig. 3. Analysis of pathway.

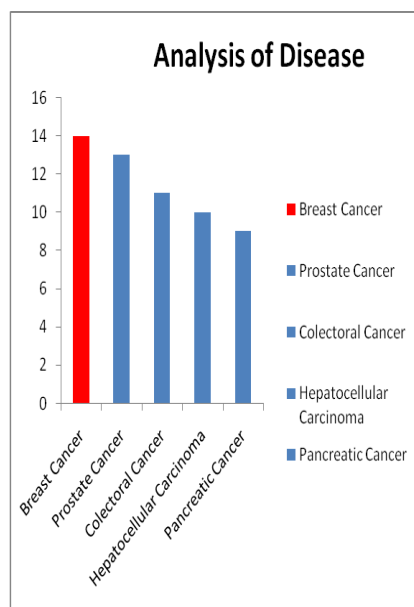


Fig. 4. Analysis of disease.

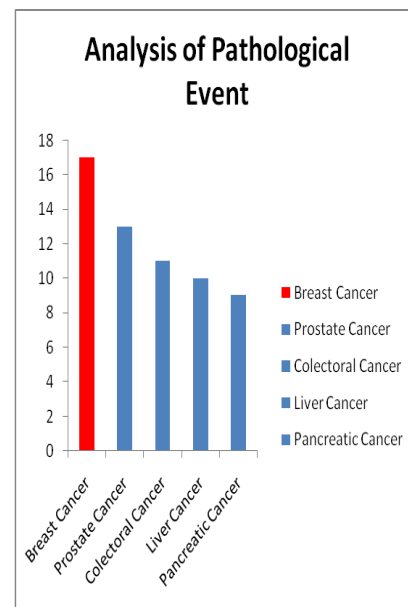


Fig. 5. Analysis of pathological event.

PROSITE results showed two important domains viz. SCAN_BOX domain and C2H2 domain that were highly conserved among all members of the miRNA targets in the pathway. These regions corresponded to the conserved sequence having the following pattern of C2H2 domain of Zinc Finger proteins that are potential sites for DNA/RNA binding. The results were compared with the multiple sequence alignment of all the miRNA target sequences and a pattern was obtained that was specifically for Class C2H2 domain of Zinc Finger proteins. (Fig. 7). It was observed from the multiple sequence alignment results that spacing of their zinc-binding residues and number of Zinc fingers differed in various classes of Zinc finger proteins.

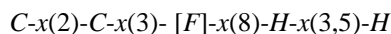


Fig. 7. Conserved pattern derived from multiple sequence alignment

Virtual screening results showed the following five leads as the most promising compounds that can be considered on

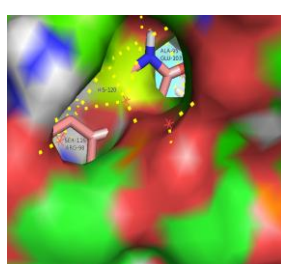
prioritization for synthesis and testing.(TABLE I) and (Fig. 8). The 2D structures of these top scoring compounds are given in Fig. 9.

TABLE I: TOP TEN COMPOUNDS WITH FREE ENERGY OF BINDING VALUES OBTAINED BY HEX 6.3.

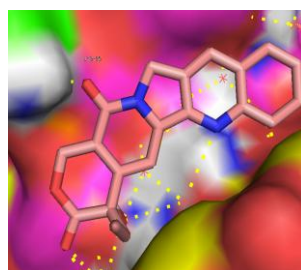
Sl No	MOLECULE	FREE ENERGY OF BINDING (kcal/mol)
1	INDOLOCARBOZOLE	-272.39
2	WITHAERIN_A	-312.16
3	NIMBOLIDE	-263.13
4	SILIBININ	-332.59
5	CAMPTOTHECIN	-255.77
6	SILYMARIN	-309.59
7	LUCIDENIC ACID	-255.35
8	QUERCETIN	-274.02
9	STAUROSPORINE	-268.06
10	DEGUELIN	-246.51

TABLE II: ANALYSIS OF INTERACTION WITH ZNF439, AUTODOCK VINA, HEX 6.3B RESULTS.

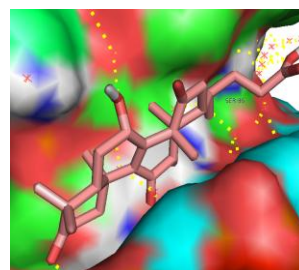
Serial No.	POTENTIAL LEAD COMPOUND	DOCKING SCORE ^a (kcal/mol)	FREE ENERGY OF BINDING ^b	INTERACTING RESIDUES
1)	INDOLOCARBOZOLE	- 9.2	- 272.39	GLU -103, HIS- 120, ALA-95, SER-116, ARG-98, GLN-117, THR-99
2)	CAMPTOTHECIN	- 8.8	- 255.77	LYS-85
3)	LUCIDENIC ACID	- 8.5	- 255.35	SER-86
4)	QUERCETIN	- 8.5	- 274.02	LYS-85
5)	STAUROSPORINE	- 8.5	- 268.06	PRO-77, TYR-78, LYS-104, SER-114



(a)



(b)



(c)

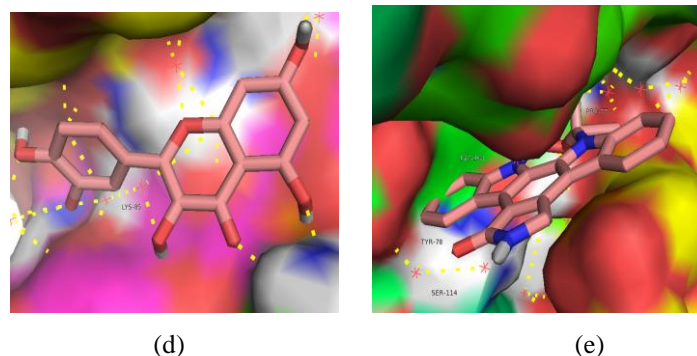


Fig. 8. 3D structures of the potential lead compounds and its interactions with ZNF439 (a) indolocarbazole (b) camptothecin (c) lucidenic Acid (d) quercetin (e) Staurosporine. The dotted line represents H bond interactions between the compound and ZNF439.

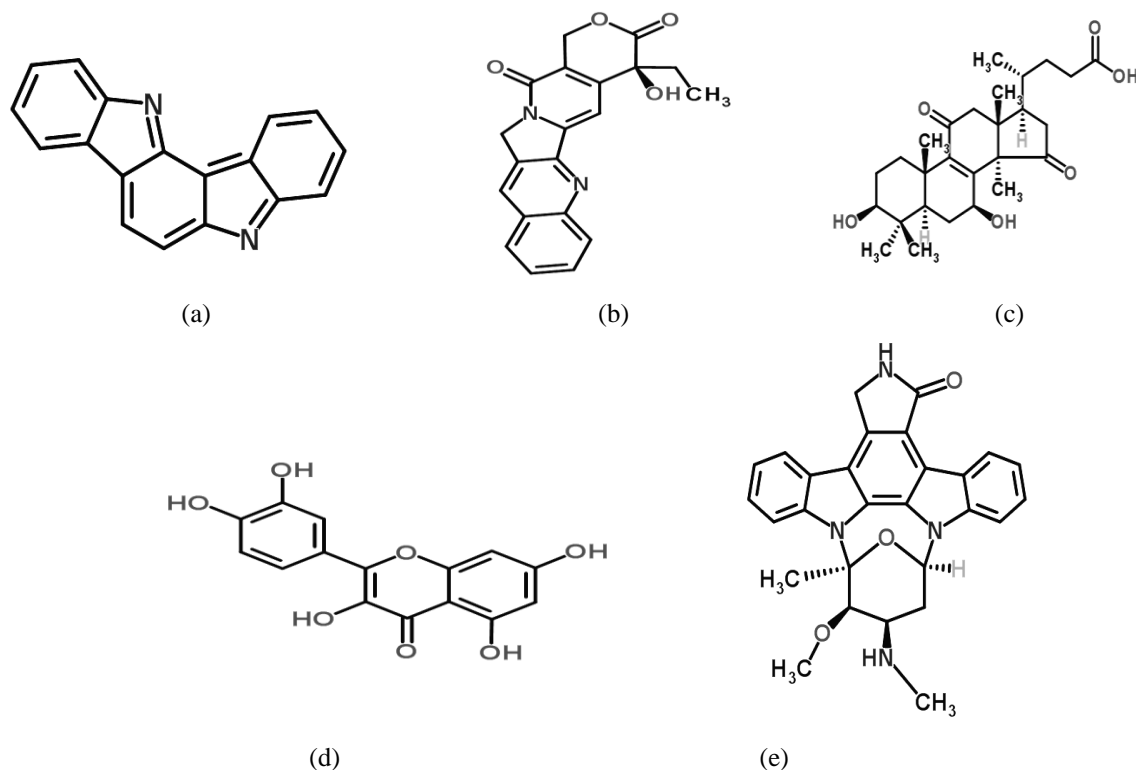


Fig. 9. 2D structures of the potential lead compounds (a) indolocarbazole (b) camptothecin (c) lucidenic Acid (d) quercetin (e) staurosporine.

IV. DISCUSSION

In the present study, the reliable predicted targets of all human miRNAs constructed biologically meaningful disease gene network of 1509 nodes and 7476 edges. In the human miRNA target network, the most relevant pathway predicted is p53 Signaling Pathway, the disease is Breast Cancer and the pathological event is Breast Cancer. The most suitable target for breast cancer has been identified as ZNF439. A small molecule library of anti-cancer agents was screened to identify best Leads. The compounds such as Indolocarbazole, Camptothecin, Lucidenic Acid, Quercetin and Staurosporine were considered as best leads for ZNF439 which shows favourable interactions computationally. As these are predicted to be promising lead compounds, they can be prioritized for synthesis and testing, through In-vitro assays and optimization of breast cancer and other related cancers.

The Lead compounds such as Indolocarbazole, Camptothecin, Lucidenic Acid, Quercetin and Staurosporine are the natural products and provides significant anti-cancer

activity. Staurosporine and Rebeccamycin are the natural products belong to the family of indolocarbazole alkaloids with antitumor properties. Currently an intense effort exists for the creation of indolocarbazole derivatives for the treatment of cancer and neurodegenerative disorders [11].

V. CONCLUSION

MicroRNAs were first described by Rosalind Lee, Victor Ambros and Rhonda Feinbaum in 1993[12]. The term microRNA was initiated in 2001. In general, hundreds of target mRNAs are concurrently downregulated by a single miRNA by pairing to the 3'UTR of mRNA from either perfect or imperfect sequence complementarity. miRNAs provide a wide range of cellular functions such as differentiation, metabolism, development, apoptosis and proliferation[13]. In the present study, analysis of miRNA target network showed that the most relevant pathological event is 'breast cancer'. Furthermore, the highly relevant diseases include 'breast cancer', 'prostate cancer', and

‘colectoral cancer’. The human microRNAome plays a more specialized role in regulation of tumorigenesis. Therefore, the miRNA might serve as the most effective approach to suppressing the tumorigenic potential of a wide range of cancers simultaneously miRNA-based therapy directed to targeting many cancer-associated pathways [14].

The different classes of ZnFs differ largely by their function, identity and spacing of their zinc-binding residues. The common feature shared by these classes is to mediate the interaction of proteins with other biomolecules including DNA, RNA, other proteins or lipids. Familial breast cancers are associated with defects in RING-finger proteins. Recently, the genome-wide sequencing studies in breast cancer have identified zinc finger protein 668 (ZNF668) as a breast tumor suppressor gene which mainly functions in regulating p53 stability [15].

The protocol used in our work can be applied to find targets for other diseases and propose leads for these targets.

ACKNOWLEDGMENT

It is with immense thankfulness that we acknowledge Institute of Computational Biology, Bangalore for providing the most modern laboratory facilities. Special thanks to our faculty and management members - Department of Biotechnology, PES Institute of Technology, Bangalore for their continuous support and cooperation.

REFERENCES

- [1] D. P. Bartel, “MicroRNAs: target recognition and regulatory functions,” *Cell*, 2009.
- [2] A. Kozomara and S. G. Jones, “MiRBase: Integrating microRNA annotation and deep-sequencing data,” *Nucleic Acids Research*, 2011.
- [3] M. Maragkakis, P. Alexiou, G. L. Papadopoulos, M. Reczko, T. Dalamagas, G. Giannopoulos, G. Goumas, E. Koukis, K. Kourtis, V. A. Simossis, P. Sethupathy, T. Vergoulis, N. Koziris, T. Sellis, P. Tsanakas, and A. G. Hatzigeorgiou, “Accurate microRNA target prediction correlates with protein repression levels,” *BMC Bioinformatics*, 2009.
- [4] M. E. Smoot, K. Ono, J. Ruscheinski, P. L. Wang, and T. Ideker, “Cytoscape 2.8: New features for data integration and network visualization,” *Bioinformatics*, 2011.
- [5] M. Gasco, S. Shami, and T. Crook, “The p53 Pathway in breast cancer,” *Breast Cancer Res*, vol. 4, pp. 70-76, 2002.

- [6] Z. Hu, J. Mellor, J. Wu, and C. DeLisi, “VisANT: An online visualization and analysis tool for biological interaction data,” *BMC Bioinformatics*, 2004.
- [7] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, “Autodock4 and AutoDockTools4: Automated docking with selective receptor flexibility,” *J. Computational Chemistry*, 2009.
- [8] D. W. Ritchie and V. Venkatraman, “Ultra-fast FFT protein docking on graphics processors,” *Bioinformatics*, 2010.
- [9] M. Sastry, J. F. Lowrie, S. L. Dixon, and W. Sherman, “Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments,” *J. Chem. Inf. Model.*, 2010.
- [10] A. G. Planson, P. Carbonell, E. Paillard, N. Pollet, and J. L. Faulon, “Compound toxicity screening and structure-activity relationship modeling in *Escherichia coli*,” *Biotechnol Bioeng*, 2012.
- [11] C. Sanchez, L. Zhu, A. F. Brana, A. P. Salas, J. Rohr, C. Mendez, and J. A. Salas, “Combinatorial biosynthesis of antitumor indolocarbazole compounds,” *PNAS*, 2005.
- [12] Y. Wang and R. Belloch, “Cell cycle regulation by microRNAs in stem cells,” *Cell Cycle in Development*, vol. 19, pp. 459-469, 2011.
- [13] M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, W. Gao, and Q. Cui, “An analysis of human microRNA and disease associations,” *PLoS One*, 2008.
- [14] J. Satoh and H. Tabunoki, “Comprehensive analysis’ of human miRNA target networks,” *BioData Mining*, 2011.
- [15] R. Hu, G. Peng, H. Dai, *et al.*, “ZNF668 functions as a tumor suppressor by regulating p53 stability and function in breast cancer,” *Cancer Res*, 2011.



Vaijayanthi Raghavan was born in Bangalore, India. Raghavan received her Bachelor degree in chemistry (1988) from Mumbai University and her Masters in Bioinformatics (2007) from Bharathidasan University, Tamilnadu, INDIA. Currently Raghavan is pursuing her PhD in the area of Cancer Biology from Visvesvaraya Technological University, Belgaum, INDIA under the guidance of Dr. Maulishree Agrahari. Her major study is on p53, a tumor suppressor protein and in the area of cancer research. She has worked in CIPLA, INDIA – a leading pharmaceutical company as *Quality Control executive* for a period of 10 years and then moved to Software Technology Group, Bangalore, INDIA and was heading the Bioinformatics division for a period of 5 years. Currently she is working as *Assistant Professor & Head Bioinformatics domain* in the Department of Biotechnology, PES Institute of Technology, Bangalore, INDIA since August 2007. She has to her credit 2 international and 2 national conference papers. Manasa D B is currently a student of MTech Bioinformatics in the Department of Biotechnology, PES Institute of Technology, Bangalore, INDIA. She has presented a paper in the International conference on Biological and Life Sciences, Singapore, 2012 on miRNA Related work.