JOURNAL OF COMPLEMENTARY MEDICINE RESEARCH, 2020
VOL 11, NO. 3
10.5455/jcmr.2020.11.03.19

RESEARCH ARTICLE

Open Access

Computational Analysis of Oral Cancer Gene Expression Profile and Identification of MiRNAs and their Regulatory Hub Genes

Ariya S.S[1], Akhila Rachel James[1] and Baby Joseph[1, 2*]

[1]Department of Biotechnology, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India. 603103
[2] Department of Research, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India. 603103

Abstarct

Oral cancer is one of the major cancers causing death worldwide. Micro RNAs (miRNAs) are small RNA molecules, each capable of coding several genes. In oral cancer, many miRNAs were proven to be expressed in higher level in tumor samples when compared to normal. Iin this work, the miRNAs commonly expressed In oral cancer patients were analysed. The miRNAs hsa-miR-6514-3p', 'hsa-miR-34b-3p', 'hsa-miR-142-3p', 'hsa-miR-146b-5p', 'hsa-miR-451b', 'hsa-miR-519e' and 'hsa-miR-3591-3p were up-regulated in oral cancer patients. They coded for the genes AMN1, CUL1 and CDC14. The expression of these genes were related to oral cancer growth, progression and metastasis. Hence identifying drugs that could target these genes could help in reducing the oral cancer mortality by improving progression of patients.

ARTICLE HISTORY

Received April 27 2020,
Accepted May 28, 2020
Published October 07, 2020

KEYWORDS

miRNA, oral cancer, up-regulation, protein interaction network, master regulators

INTRODUCTION

Oral cancer is the tenth most common cancers in the world and is predominantly seen in tongue. It may also occur on gingiva, palate, lip and floor of the mouth. In India, oral cancer is a serious health issue, accounting for about 30% - 40% of overall cancer cases. Detecting the cancer in the primary phase is one of the most efficient way to decrease the mortality rate by timely treatment [1]. The Gene Expression Omnibus (GEO) database has been successfully used as a powerful tool to determine crucial genetic mutations in carcinogenesis and is largely used to discover promising therapeutic targets in the recent years [2].

MicroRNAs are small noncoding RNAs that regulate protein expression at posttranscriptional level through mRNA instability or translation inhibition by binding to the 3′ un-translated regions (3′UTR) of target genes. They are involved in many kinds of diseases including infections, tumors and autoimmune diseases [3].

In order to overcome the varying results attributed to the application of different microarray platforms, integrated bioinformatics approaches are being used widely in cancer research to identify the hub genes and transcripts [4]. By studying the differential expression of genes using microarray platform, the diseases' molecular nature can be identified [5]. Assembling the transcript networks from large expression datasets by computing the expression under several condition via matrix of all the expressed genes pairwise correlation and further bioinformatical approaches pave way to determine the vital genes and their allied proteins [6,7].

In this study, a co-expression network of various Differently Expressed Genes (DEGs) was constructed and the most significant modules in the network were segregated to determine the tumorigenic genes and transcripts and that control and regulate other genes and proteins involved in oral cancer progression and metastasis and hence, targeting them will give a better prognosis rate.

MATERIALS AND METHODS

Data acquisition from gene expression profile

The entire datasets relevant to oral cancer available at Gene Expression Omnibus (GEO), a functional genomic database available at National Center for Biotechnology Information (NCBI) was retrieved [8]. Initially, the key word "oral cancer" and "patients" and Homo sapiens" [porgn:__txid9606] revealed 370 hits. Data assortment was then done by eliminating datasets other than miRNAs. Finally,

* **Contact** Prof. Dr. Baby Joseph, Hindustan Institute of Technology and Science Kancheepuram Dist, Tamilnadu, scientistpetercmi@gmail.com, deanresearch@hindustanuniv.ac.in

a total number of 54 samples of tumour profiles and 68 from non-tumours profiles were analysed from GSE124566, GSE94863, GSE69002, GSE28100, GSE52811, GSE33299 and GSE37472 datasets respectively. The detailed information of datasets is given in table 1.

Gene expression analysis

In order to analyse the samples to discover the differently expressed genes, a matrix table was constructed from sample table using simple data extraction methods from the value columns. Based on sophisticated R based analysis using established Bioconductor [9] and R package [10, 11], the DEGs across determined. Benjamini Hochberg method was adapted for the estimation of the false discovery rate and p value [12]. By the analysis of variance (ANOVA) and t test, all the DEGs were recognized. The samples were compared with original submitter supplied processed microarray data using GEO Query and limma packages. Further, the DEGs were screened based on FC value ≥ 2.0 and p-value ≤ 0.05 from the top 250 DEGs to find the most significant upregulated ones. The downregulated genes were segregated based on p-value ≤ 0.05 and FC value ≤ -2.0. Table 2 gives detail about the gene expression analysis for the miRNAs.

Analysis of Genes targeted by miRNA

The published and annotated sequences of miRNA was retrieved from the miRBase database (http://www.mirbase.org/) through which the Target scan database (http://www.targetscan.org/vert_72/) was accessed and the genes targeted by the miRNAs were predicted. Those genes with scores > 50 were chosen for further analysis. From the targeted genes, the common genes were identified to find the significant genes of oral cancer.

Protein interaction network and topology analysis

The protein interaction network was constructed by using Cytoscape [13] based on the information from BioGrid and Reactome database. The topological interactions occurring on the network were analysed by applying mathematical models for network topology analysis by applying graph theory. Thus the hub genes were identified.

Gene Ontology analysis

The gene ontology information helps to computationally analyse and attain knowledge about the different functions of genes determined by large scale molecular biology approaches and genetic experiments. It also shows a slight path by showing light on the pathways associated by the genes. The ontology terms for the genes were analysed using the BiNGO app of Cytoscape.

Table 1: Detailed information about datasets

Accession number	Title	Contributors	Updated on	Country
GSE124566	MicroRNA-204-5p is a tumor suppressor and potential therapeutic target in head and neck squamous cell carcinoma	Zhuang Z, Yu P, Xie N, Huang H, Wang C	2020	China
GSE94863	Identification of the RB loss-induced transcriptome in prostate cancer [RNA]	Knudsen KE, McNair C	2019	USA
GSE69002	MicroRNA Expression Profiling of Saliva in Tongue Squamous Cell Carcinoma Patients	Creighton C, Ozen M, Zhang Y	2017	USA
GSE28100	Differential expression of microRNAs between oral squamous cell carcinoma and healthy control tongues	Jung HM, Patel RS, Cheng JQ, Chan EK	2015	USA
GSE52811	microRNA expression in oral keratinocytes	Hunter KD, Murdoch C, Lambert DW	2017	UK
GSE33299	MiRNA array based, miRNA expression profiles of oral leukoplakia (OLK) and malignant transformed oral leukoplakia (mtOLK)	Jiang W, Xiao W, Bao Z	2012	China
GSE37472	Serum miRNA profiles of patients with oral cancer and pre-cancer	MacLellan SA, Garnis C, Poh CF	2013	Canada

RESULT AND DISCUSSION

Identification of DEGs from microarray data analysis

Comparison of the expression level of miRNAs in tumour and normal samples from different samples with microarray expression profile analysis revealed the most significant DEGs from each of the 7 datasets. Based on the fold change and the p value of the DEGs, the up-regulated and the down-regulated gene in each were determined. Among all the differently expressed ones, top 250 were considered as the most significant ones. As a result of this, 162 up-regulated transcripts and 211 down-regulated transcripts were segregated. Further information regarding the DEGs from each dataset is given in Table 2.

Seven different datasets of differently expressed transcripts in Homo sapiens with respect to oral cancer were found in NCBI GEO and were analysed. The normal and tumour samples were obtained for further analysis. The bioinformatical approaches lead to the identification of the commonly up-regulated transcripts in oral cancer patients. The up-regulated transcripts found in different datasets shows that these transcripts are found to be highly expressed in oral cancer around the world.

Screening of differentially expressed oral cancer transcripts

Illumination of important miRNA among all the DEGs was a crucial task. To sort of the top 250 genes, mining of data from each dataset was done based on the Benjamini and Hochberg method. The transcripts that were found to be duplicated were eliminated manually. The 7 datasets were evaluated further to recognize the common miRNAs present in all the 7 datasets. As a result, the transcripts 'hsa-miR-6514-3p', 'hsa-miR-34b-3p', 'hsa-miR-142-3p', 'hsa-miR-146b-5p', 'hsa-miR-451b', 'hsa-miR-519e' and 'hsa-miR-3591-3p' were commonly found in more than one dataset and were found to be up-regulated.

Regulationary interactions between miRNAs and Genes

The genes targeted by the miRNAs were identified by data analysis approach. This revealed a total number of 3575 genes among which the has-miR-519e targeted the maximum number of genes. The genes ARHGAP12, CASK, ENTPD1, FBXO32, FBXW2, MSI2, QKI and ZNF367 were the commonly targeted genes by the significant miRNAs. Table 3 details the number of genes targeted by miRNAs along with their sequences.

Table 2: Information related to the genes differentially expressed from each dataset

Accession no:	Normal	Tumor	Up-regulated miRNAs	Down-regulated miRNAs
GSE124566	10	10	41	26
GSE94863	8	8	0	0
GSE69002	4	3	43	39
GSE28100	2	17	67	5
GSE52811	2	6	2	126
GSE33299	5	5	5	5
GSE37472	26	29	4	10

Table 3: Information related to the genes differentially expressed from each dataset

miRNAs	Sequence	No. of gene targets
hsa-miR-6514-3p	CUGCCUGUUCUUCCACUCCAG	358
hsa-miR-34b-3p	CAAUCACUAACUCCACUGCCAU	528
hsa-miR-142-3p	UGUAGUGUUUCCUACUUUAUGGA	418
hsa-miR-146b-5p	UGAGAACUGAAUUCCAUAGGCUG	487
hsa-miR-451b	UAGCAAGAGAACCAUUACCAUU	673
hsa-miR-519e*	UUCUCCAAAAGGGAGCACUUUC	1111

miR-142 are short non-coding RNAs that are involved in port transcriptional of gene expression in multicellular organisms. It is located on the 17th chromosome at 17q22 [14]. miR-142-3p was first identified for its presence in development in the lymphoid system and was subsequently in leukaemia. Deregulation of miR-142-3p constituted a miR signature specific to malignant oral neoplasms. It was also found to be associated with OSCC [3].

Protein interaction network construction and analysis

The protein – protein interaction network was constructed for the common genes targeted by miRNAs. The Biogrid and Reactome database was used as reference databases to create the protein-protein interaction network. The nodes and edges which represent the proteins and their relationship was constructed computationally. The PPInetwork created for the common genes is shown in the figure 1. Furthermore, by local and global scoring methods of topological interaction analysis, the genes AMN1, CDC14 and CUL1 was identified to play most significant role in oral cancer progression in patients. Additionally this result clarifies that these genes can be considered as hub genes targeted by the common miRNAs expressed in oral cancer.

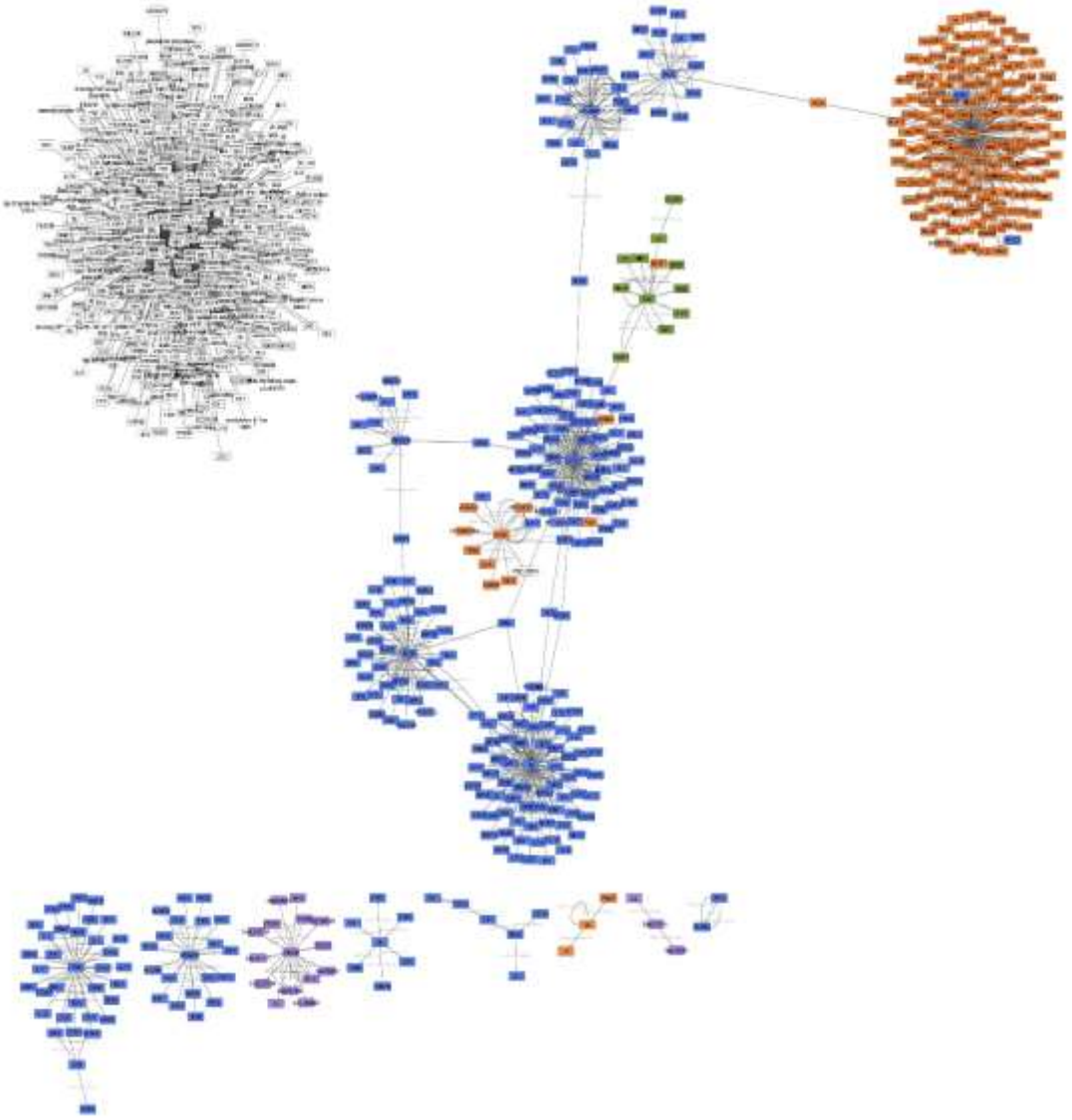

Figure 1: Protein-protein interaction network of the common gene targets of miRNAs

Gene function analysis

The gene ontology terms for the constructed PPI network was also analysed which put light on the important functions of the genes present in the network. These functions include most of the gene ontology terms which are prominent for cancer progression and metastasis. Figure 2 depicts the gene ontology network created by the genes of interest.

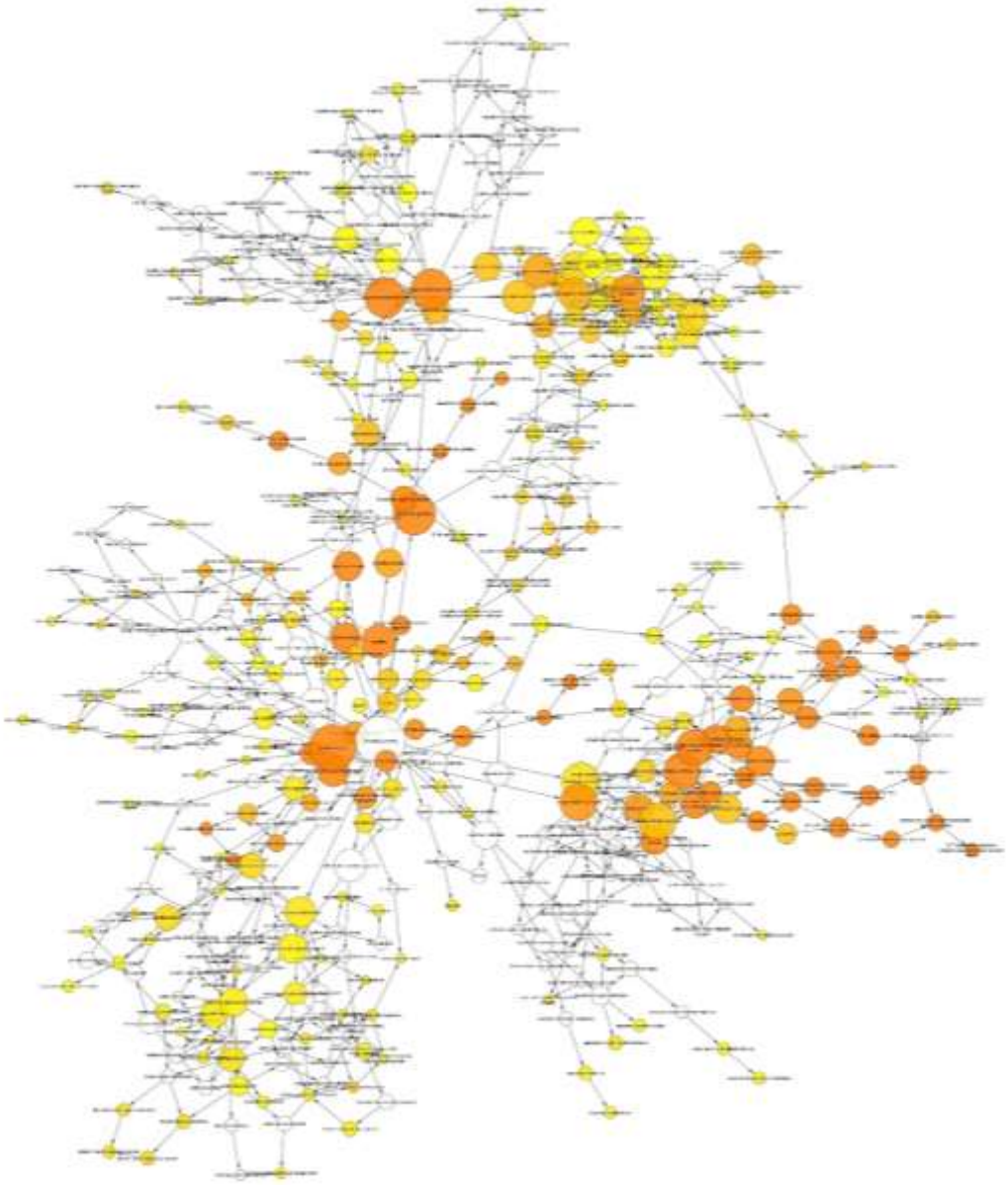

Figure 2: The gene ontology network of the genes

CONCLUSION

The miRNAs hsa-miR-6514-3p', 'hsa-miR-34b-3p', 'hsa-miR-142-3p', 'hsa-miR-146b-5p', 'hsa-miR-451b', 'hsa-miR-519e' and 'hsa-miR-3591-3p were identified as the highly expressed miRNAs in oral cancer. The genes AMN1, CUL1 and CDC14 were identified as the hub genes that are targeted by the miRNAs expressed highly in oral cancer. Hence studies concentrating more in identifying the drugs that could target genes that have direct link with the miRNAs could help in reducing the oral cancer mortality by improving progression of patients.

REFERENCES

1. Z.S. Jairajpuri, S. Rana, A. Hajela and Jetley S, "Toward early diagnosis of oral cancer: Diagnostic utility of cytomorphological features, a pilot study." National journal of maxillofacial surgery vol. 10,1 (2019): 20-26. doi:10.4103/njms.NJMS_12_17
2. P. Yan, Y. He, K. Xie, S. Kong and W. Zhao, "In silico analyses for potential key genes associated with gastric cancer". PeerJ, 6, e6092. doi.org/10.7717/peerj.6092
3. Y. Wang, J. Liang and H. Qin. "Elevated expression of miR-142-3p is related to the pro-inflammatory function of monocyte-derived dendritic cells in SLE." *Arthritis Res Ther* **18,** 263 (2016). https://doi.org/10.1186/s13075-016-1158-z
4. L. Cao, Y. Chen, M. Zhang, D.Q. Xu, Y. Liu, S, X, Liu and P. Wang, " Identification of hub genes and potential molecular mechanisms in gastric cancer by integrated bioinformatics analysis". PeerJ. 2018 Jul 2;6:e5180. doi: 10.7717/peerj.5180

5. W.W. George and R.M. Simon, "A random variance model for detection of differential gene expression in small microarray Experiments." *Bioinformatics* 2003; 19(18): 2448-55.
6. S. Bruckner, F. H. Fener, M. K. Richard, R. Shamir and R. Sharan, "Topology-Free Querying of Protein Interaction Networks." Journal of Computational Biology 2010; 17 (3) 237-52.
7. A. K. Agrahari and P. D. C. George, "A computational approach to identify a potential alternative drug with its positive impact towards PMP22." *J Cell Biochem.* 2017; 118(11):3730-43.
8. R. Edgar, M. Domrachev and A. E. Lash, "Gene Expression omnibus: NCBI gene expression and hybridization array data repository." Nucleic Acid Research; 2002; 30(1):207-10.
9. R. C. Gentleman, V. J. Carey, M. B. Douglas, B. Bolstad, M. Dettling and S. Dudoit, "Bioconductor: open software development for computational biology and bioinformatics." Genome Biol. 2004;5(10):R80.
10. T. Barrett, E. Stephen, Wilhite, L. Pierre, E. Carlos, F. Irene, Kim and T. Maxim, "NCBI GEO: archive for functional genomics data sets—update." *Nucleic Acids Res.* 2013; 41:D991-5.
11. Y. Benjamini, Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing." Journal of the Royal Statistical Society. 1995 **57** (1): 289–300.
12. M. Manikandan, D.M/ Rao, G. Arunkumar, M. Manickavasagam, K.S.Rajkumar, R. Rajaraman, A. K. Munirajan, "Oral squamous cell carcinoma: microRNA expression profiling and integrative analyses for elucidation of tumourigenesis mechanism." Mol Cancer. 2016 Apr 7;15:28. doi: 10.1186/s12943-016-0512-8. PMID: 27056547; PMCID: PMC4823852.
13. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MIR142