International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 05, 2020
ISSN: 1475-7192

Comprehensive Gene Expression Analysis and Identification of Stat-1 as Master Regulator in Oral Squamous Cell Carcinoma

S.S. Ariya, Akhila Rachel James and Dr. Baby Joseph*

Abstract--- *Gene expression profiling and analysis helps in identification of master regulators inducing disease. The GEO dataset GSE74530 comprising of samples from 6 normal and 6 oral cancer patients were analyzed using R package of Bio Conductor program GEO2R. The differentially expressed genes among the samples were identified using Benjamini Hochberg method. The up-regulated and down-regulated genes were identified based on fc value > 2.0 and p-value < 0.5. Protein interaction network was constructed for the up-regulated and down-regulated genes with data from Bio Grid and Reactome database. The network was further explored by applying mathematical models for topological interaction analysis to determine the hub genes. Signal transducer and activator of transcription 1 (STAT1) was identified as the master regulator of Oral Squamous Cell Carcinoma (OSCC). Cross validation with The Cancer genome Atlas (TCGA) followed by gene ontology and functional enrichment analysis also revealed the role of STAT1 in progression of OSCC. Survival analysis with respect to expression of gene also showed the decrease in rate of patient prognosis with increase in expression of STAT1.*

Keywords--- *Oral Squamous Cell Carcinoma (OSCC), The Cancer Genome Atlas***.**

I. INTRODUCTION

Oral Squamous Cell Carcinoma (OSCC) is a common oral malignancy representing about 80-90% of all the cases (1). This is one among the top ten common cancers worldwide. Asia has large number of oral cancer patient and India is considered as the oral cancer capital of the world as the number of OSCC patients in this country is more (2). Major etiological and predisposing factors that contribute to the development of OSCC is smoking, alcohol abuse and use of tobacco and related products (3). In India, where practices like smoking and use of tobacco are culturally accepted, the rate of incidence and mortality due to OSCC is in its peak. The gene that occupy the top position in the regulatory hierarchy is called as the "master regulatory genes" or "master regulators". They work without the regulatory influence of other genes. Instead they regulate the expression and functioning of other genes. The term was coined by Susmu Ohno (4).

Gene expression analysis using micro array is a major tool to identify the differentially expressed genes. In this study, the gene expression profile GSE74530 comprising of data from twelve different samples among which 6 are normal and the remaining are of oral squamous cell carcinoma were analyzed (5). The differentially expressed genes from the set of samples were identified and protein interaction network was constructed to elucidate their role in metabolic pathways. The genes with high degree of connectivity in the protein interaction network was elucidated

S.S. Ariya, Department of Biotechnology, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India.
Akhila Rachel James, Department of Biotechnology, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India.
Dr. Baby Joseph, Prof., Dean Research, Department of Research, Hindustan Institute of Technology and Science, No: 1., Rajeev Gandhi Saalai (OMR), Padur, Kelambakkam, Chennai, Kancheepuram Dist, Tamilnadu, India.*
E-mail: scientistpetercmi@gmail.com, deanresearch@hindustanuniv.ac.in

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

and are considered as hub genes. Thus, from the hubs, the master regulator of OSCC were identified. The gene ontology terms of each genes along with the functional enrichment was also carried out.

II. MATERIALS AND METHODS

Gene expression Data Profile

The gene expression profile data of GSE74530 was retrieved from Gene Expression Omnibus Database (GEO) (5 and 6). It consists of twelve samples among which 6 are from normal and the remaining are from oral cancer patients. Expression profiling by array was the experimental type used. The tumor sample from oral cavity and adjacent non tumor tissue were obtained from patients enrolled in Phase 0 clinical trial. Affymetrix platform was used for determining the genes and the mRNA expression was analyzed to compare the differential gene expression between the normal and tumor samples.

Analysis of Gene Expression Data

The tumor and normal samples were loaded separately into the GEO2R analysis (7) platform and the differential expression of genes were analyzed by applying established BioConductor and R packages using limma query (8). Benjamini Hochberg method (9) was used for determining the differentially expressed genes (DEGs). On the basis of p-value (p-value ≤ 0.05) and fold change (fc ≥ 1.0), the up-regulated and down-regulated genes were identified. The top 250 DEGs were pre-processed and selected from the set of all genes expressed.

Function Enrichment and Ontology

Functional enrichment analysis was carried out using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (10). To determine the genes involved in cancer related pathways, Kyoto Encyclopedia for Genes and Genomes (KEGG) was used (11). Similarly, Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GORILLA) was used to predict gene ontology terms of the genes involved in OSCC (12). Only those with significant p-value < 0.5 were considered for progressive analysis.

Network Construction and Analysis

The metabolic and signaling pathways can be mapped and characterized using gene-protein interaction network. Additionally the protein interaction network construction paved way for the identification of function of unknown proteins. Cytoscape tool (13) was used for predicting the PPI network by integrating the experimental data from BioGrid and Reactome FIS databases using The Proteomics Standard Initiative Common QUery InterfaCe (PSICQUIC) (14). The topological interaction analysis was carried out using mathematical models of scoring functions provided by Cytohubba (15). Finally, on the basis of the impact, rank and occurrence of the genes, the master regulators were identified.

TCGA – Cross-Validation and Survival Analysis

The master regulator identified was further cross checked with The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/) using the Genomics Data Commons (GDC) portal to confirm the role of the gene in OSCC growth, progression and metastasis (16 and 17). Furthermore, the role of genes with respect to survival rate

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

of the patients were also determined.

III. RESULT AND DISCUSSION

Identification of DEGs from Microarray Data Analysis

The expression analysis of the micro array data profile from OSCC patients paved way for the identification of a total number of 54675 genes. From the top 250 differentially expressed gene, the analysis of up-regulated and down-regulated genes using statistical method revealed 133 genes as highly expressed and the expression level of 39 genes were low.

Functional Enrichment Analysis of DEGs

On the basis of connotation, the up-regulated genes were narrow downed with the help of functional and gene enrichment analysis. Furthermore, the gene ontology for the up-regulated genes were also identified. This put light on the involvement of the genes up-regulated in different metabolic processes and functions which enhances the progression of OSCC was clear. All the ontologies with significant p-values were considered. Table 1 details the different ontology terms along with their functions and significant p-values.

Table 1: Gene Ontology Terms for the Up-Regulated Genes

S.No.:	Gene Ontology	Description	*p*-value
Process			
1.	GO:0022617	Extracellular matrix disassembly	1.94E-6
2.	GO:0022411	Cellular component disassembly	4.1E-6
3.	GO:0030574	Collagen catabolic process	1.7E-5
4.	GO:0032963	Collagen metabolic process	4.66E-4
5.	GO:0008544	Epidermis development	8.43-E-4
Function			
1.	GO:0046914	Transition metal ion binding	3.99E-6
2.	GO:0004222	Metalloendopeptidase activity	5.01E-5
3.	GO:0008270	Zinc ion binding	5.01E-5
4.	GO:0008237	Metallopeptidase activity	5.01E-5
5.	GO:0004175	Endopeptidase activity	4.66E-4
6.	GO:0008233	Peptidase activity	6.67E-4
7.	GO:0070011	Peptidase activity, acting on L-amino acid peptides	6.67E-4
8.	GO:0048081	Receptor ligand activity	9.04E-4
Component			
1.	GO:0005576	Extracellular region	7.36 E-5
2.	GO:0005615	Extracellular space	3.23E-4

The up-regulated genes were involved in several cellular processes like extracellular matrix disassembly and cellular component disassembly. This make it evident that the highly expressed genes in the patient will have significant effect in inducing metastasis. This is because it is important for the cells to undergo such processes before metastasis. For example, the disassembly efficiently aids in detachment of the tumor cell at primary site which further enhances the transportation of the tumor cell to secondary site thus promoting invasion and finally metastasis of the cancer. Furthermore their presence in extracellular region and space further proves the role of these proteins outside the cells which reduces the attaching efficiency of the tumor cells at specific site and mutually enhancing invasion and metastasis. This also shows the role of certain gene in angiogenesis process as well.

PPI Network Analysis

Protein interaction network is an efficient way for identifying the genes and their interactions as well as the

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

protein – protein interactions occurring within or outside the cells collectively. Here, the protein interaction network was constructed for both up-regulated and the down-regulated genes by obtaining experimentally validated interaction data from BioGrid and Reactome functional Interactome databases. The network thus created is given in figure 1A and 1B respectively. The nodes and the edges in the network represents the proteins and the interactions they make between each other. The up-regulated genes created 7990 nodes and 4121 edges with data from BioGrid database while as the interactions information from Reactome databases created 3662 nodes and 1568 edges. Totally, the merged network consists of 11652 nodes and 5689 edges. Similarly, the network constructed for the down-regulated genes from the BioGrid database accounts for 900 nodes and 749 edges while as there were 135 nodes and 151 edges for the network created from the Reactome database. Collectively the merged network for down-regulated genes showed 1035 nodes and 900 edges respectively.

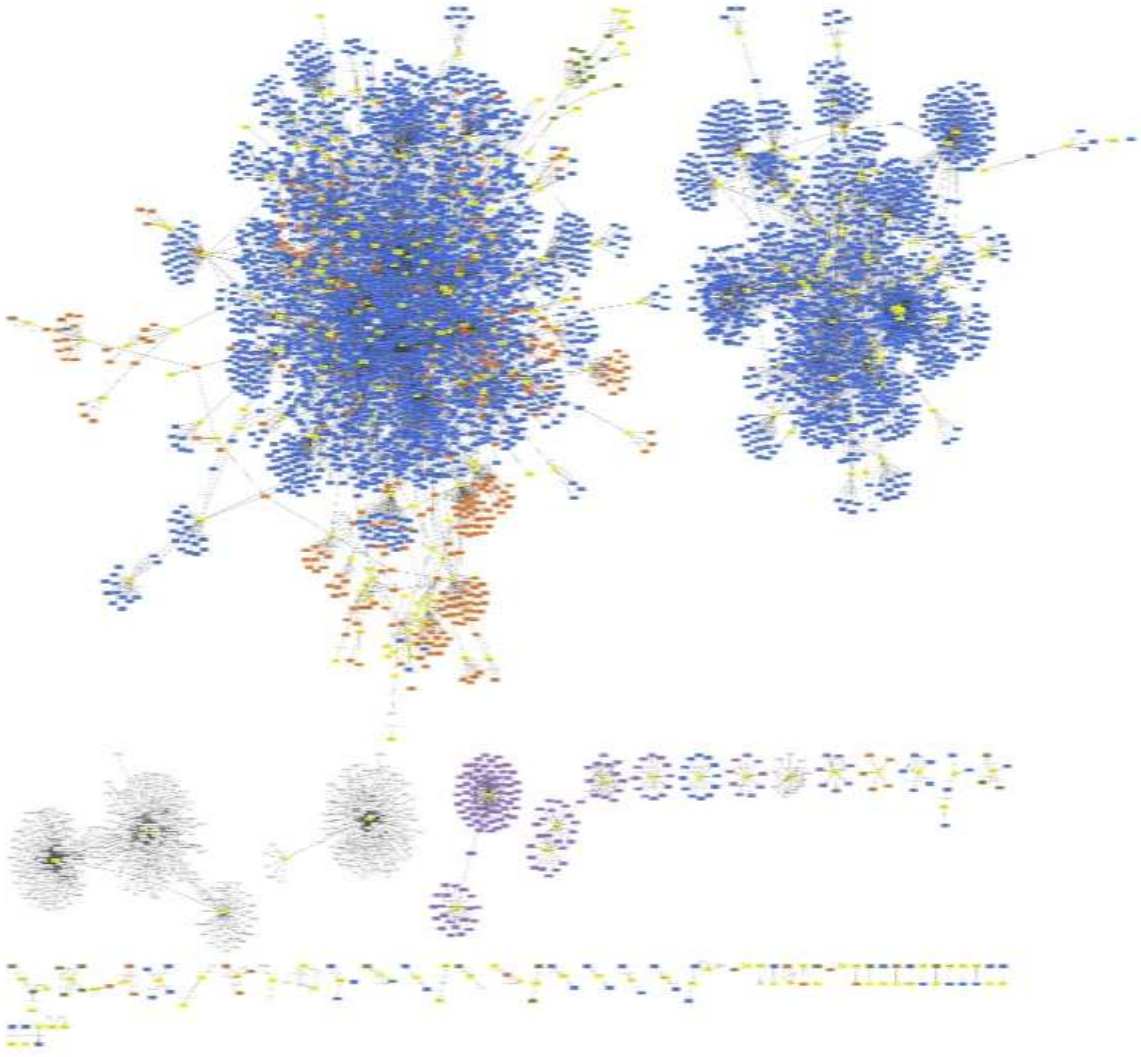

Figure 1A: Protein Interaction Network of the Up-Regulated Genes

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

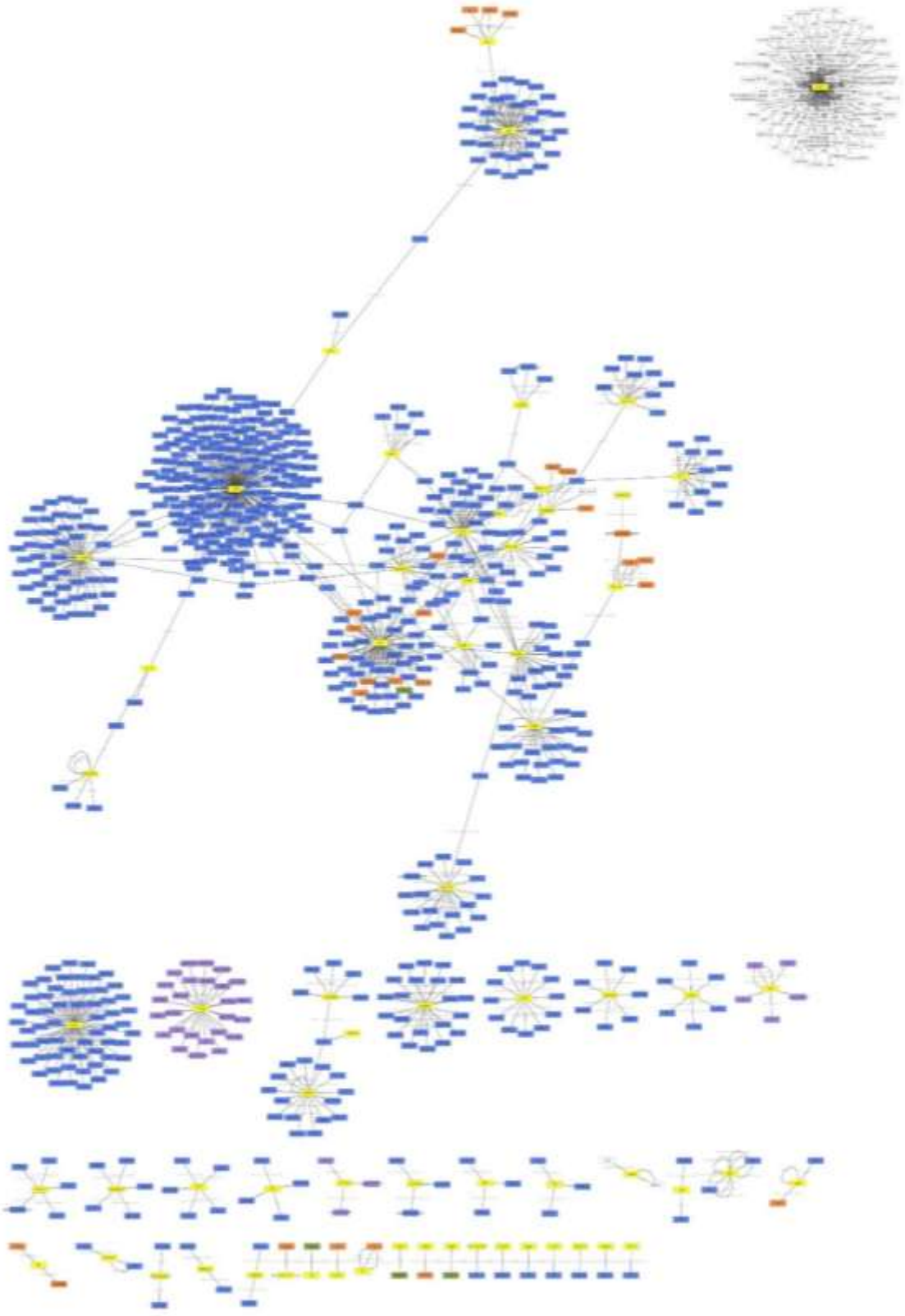

Figure 1B: Protein Interaction Network of the Down-Regulated Genes

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

Signal Transducer and Activator of Transcription 1 as a Master Regulators in OSCC

The topological interactions occurring on the network was predicted by applying mathematical models for scoring function in networks to identify the hub genes among them and thus the master regulator. Twelve different mathematical scoring functions both local and global methods namely Maximal Clique Centrality (MCC), Density of Maximum Neighborhood Component (DMNC), Maximum Neighborhood Component (MNC), Degree, Edge Percolated Component (EPC), Betweenness, Bottle neck, Closeness, Clustering coefficient, EcCentricity, Radiality and stress based on their shortest path were applied for the network separately and the top 10 gene topologies were identified from each network separately. The networks thus created with the top 10 shortest paths are shown in figure 2A to 2L respectively. As a result, the Signal Transducer and Activator of Transcription 1 (STAT1) was found to be one of the major master regulators involved in inducing OSCC progression in patients. This gene was used as a prognostic marker for renal, ovarian and pancreatic cancer (Human Protein Atlas).

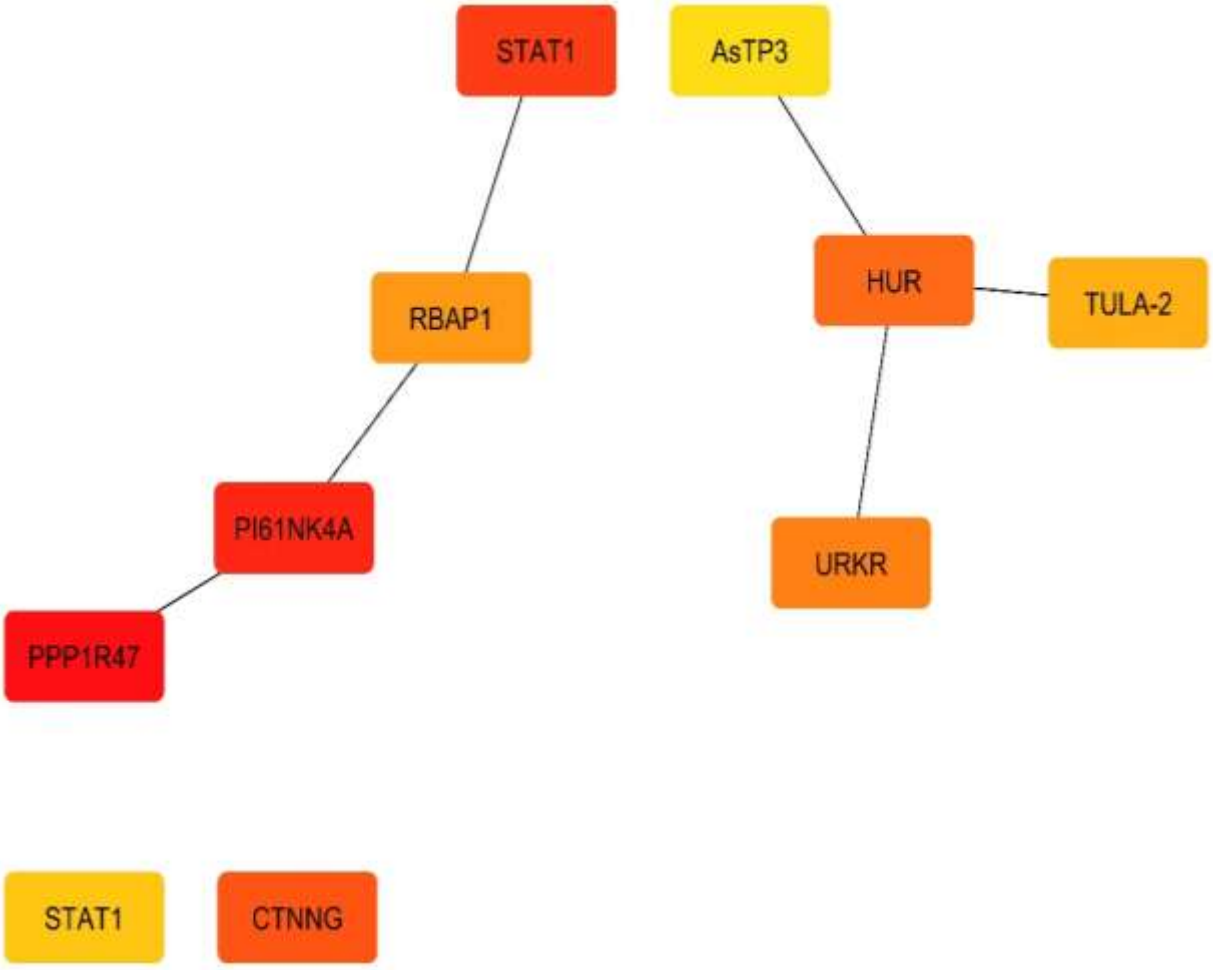

Figure 2A: Betweenness

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

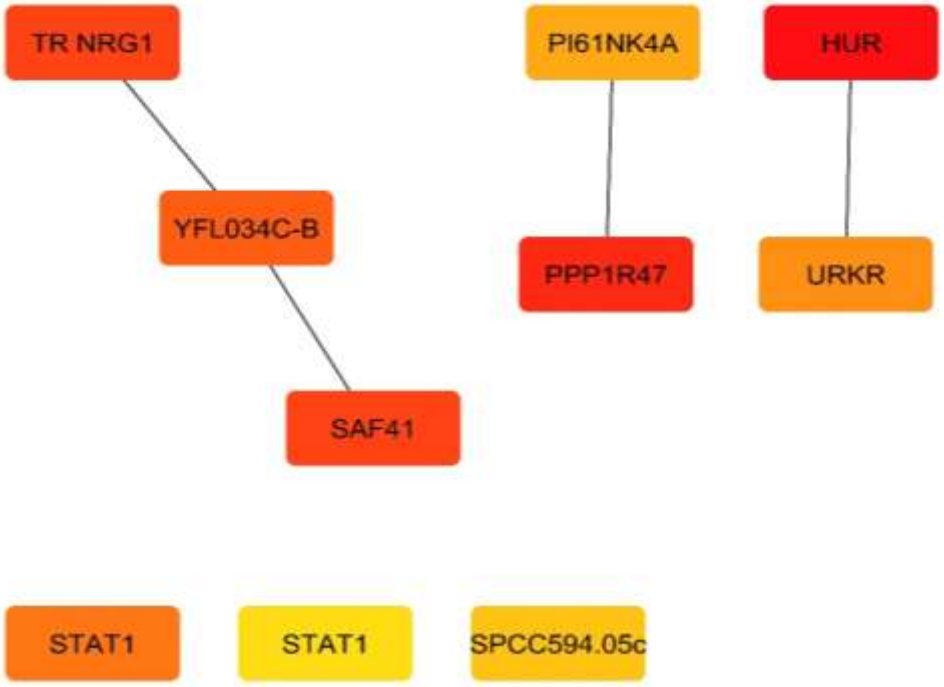

Figure 2B: Bottleneck

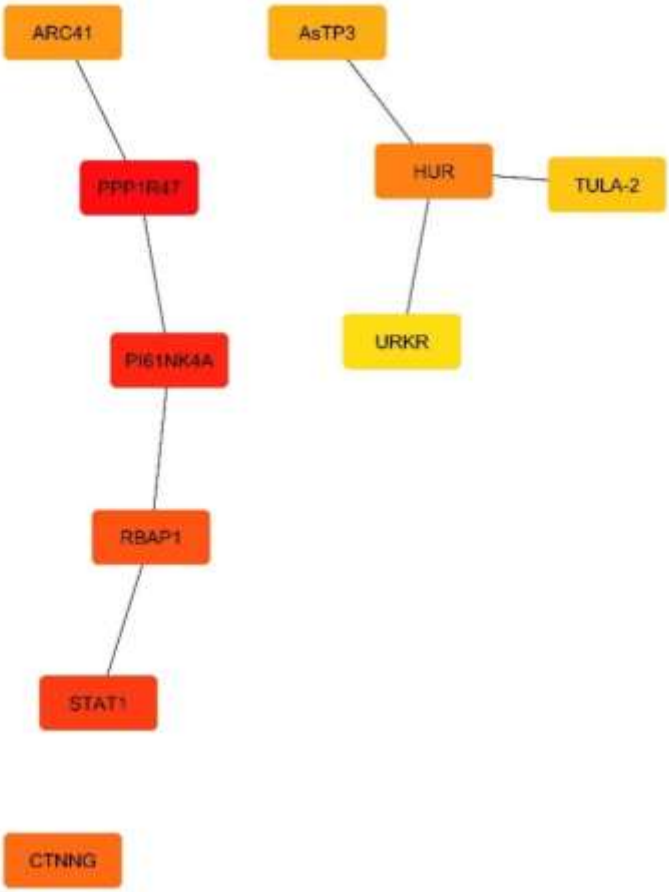

Figure 2C: Closeness

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

Figure 2D: Clustering Coefficient

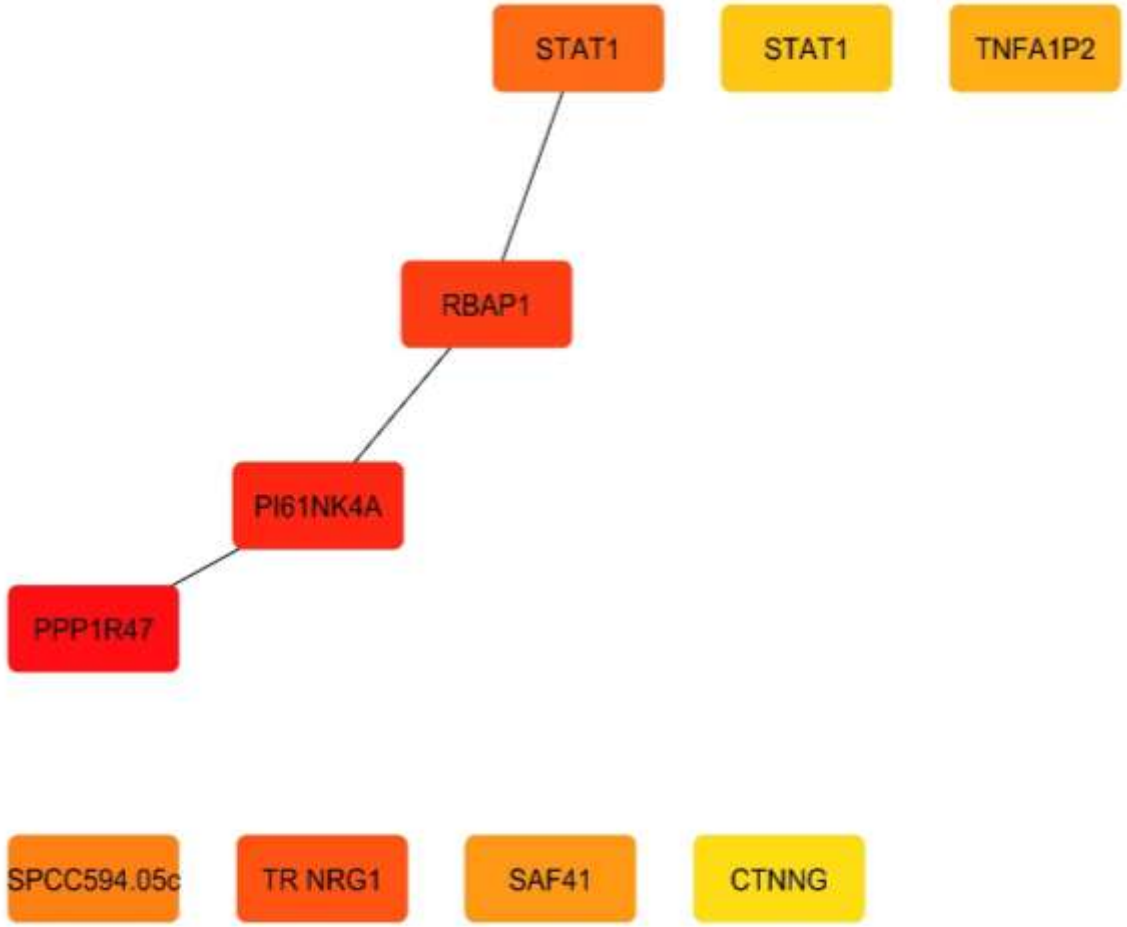

Figure 2E: Degree

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

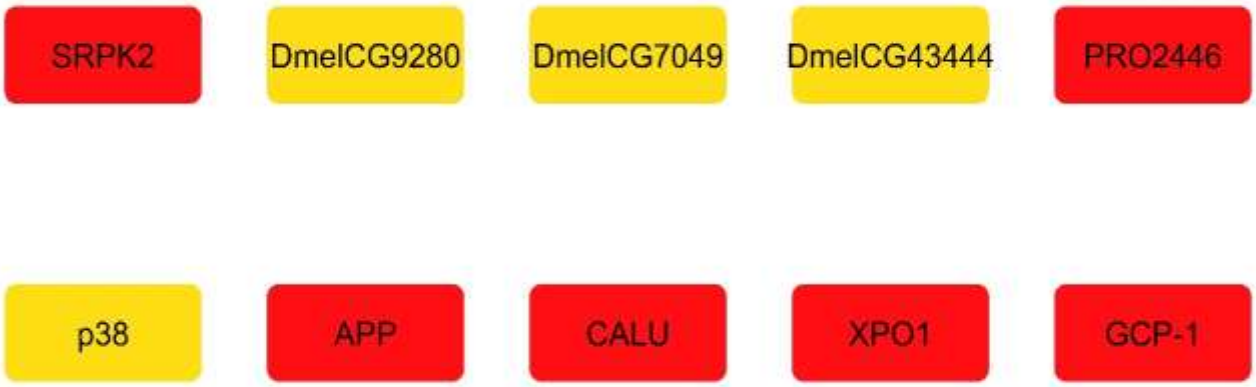

Figure 2F: DMNC

Figure 2G: EcCentricity

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

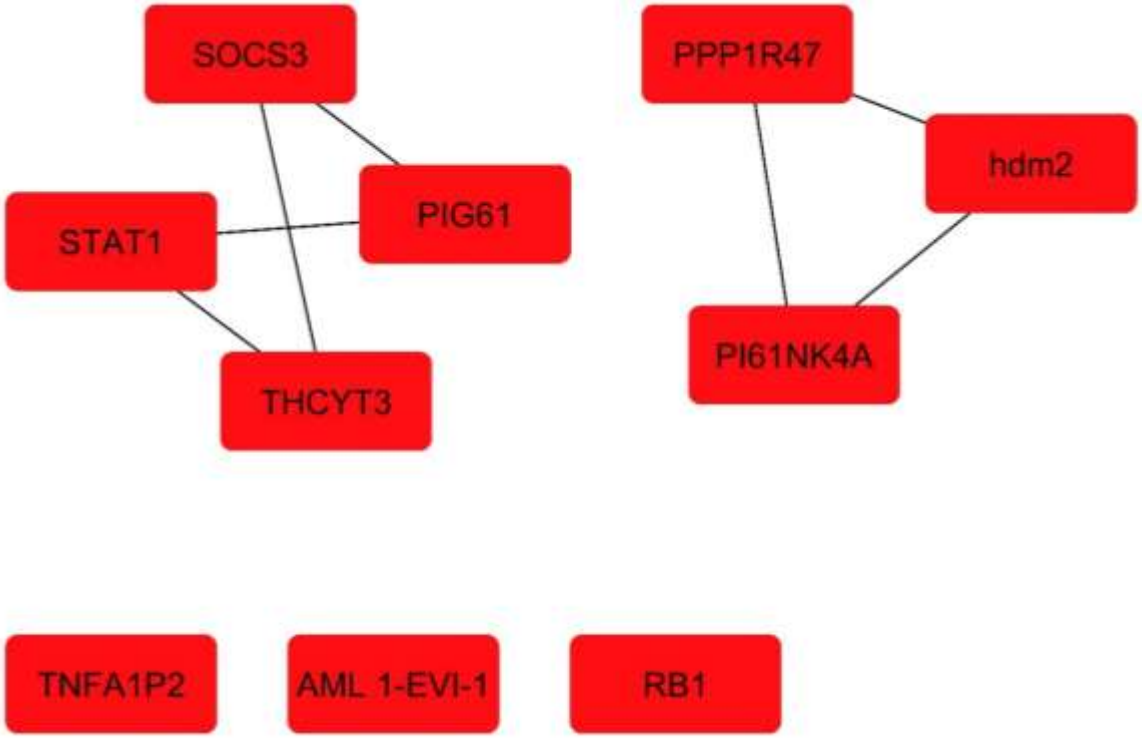

Figure 2H: EPC

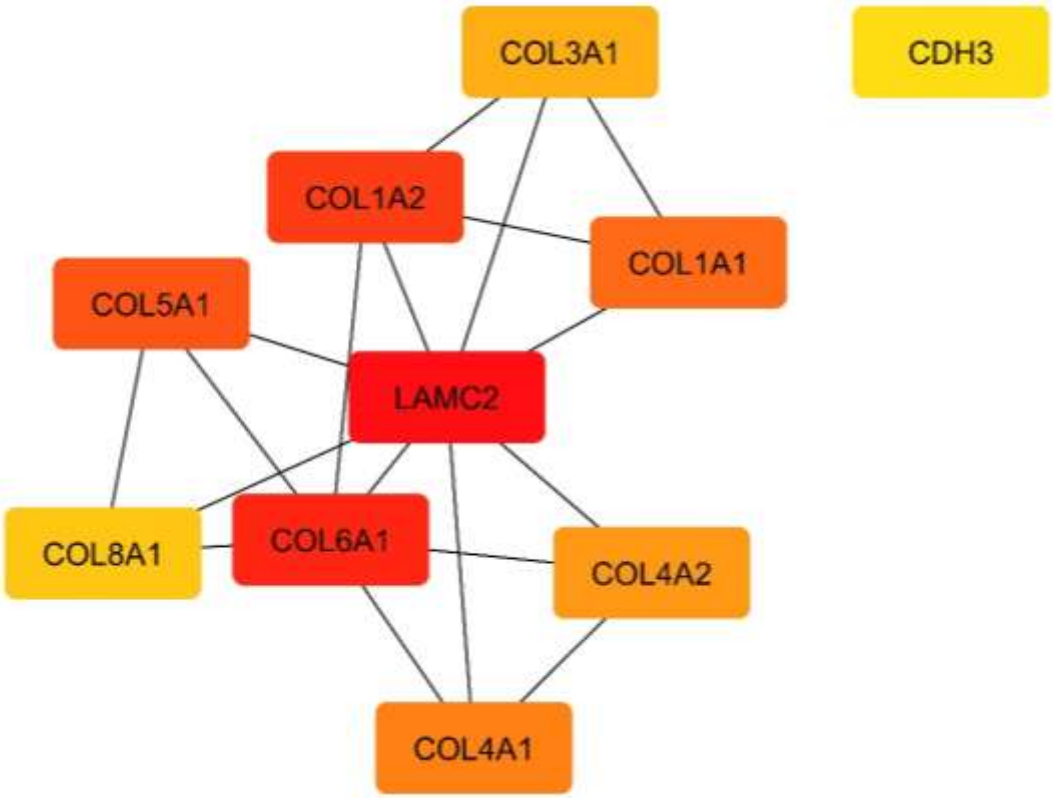

Figure 2I: MCC

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

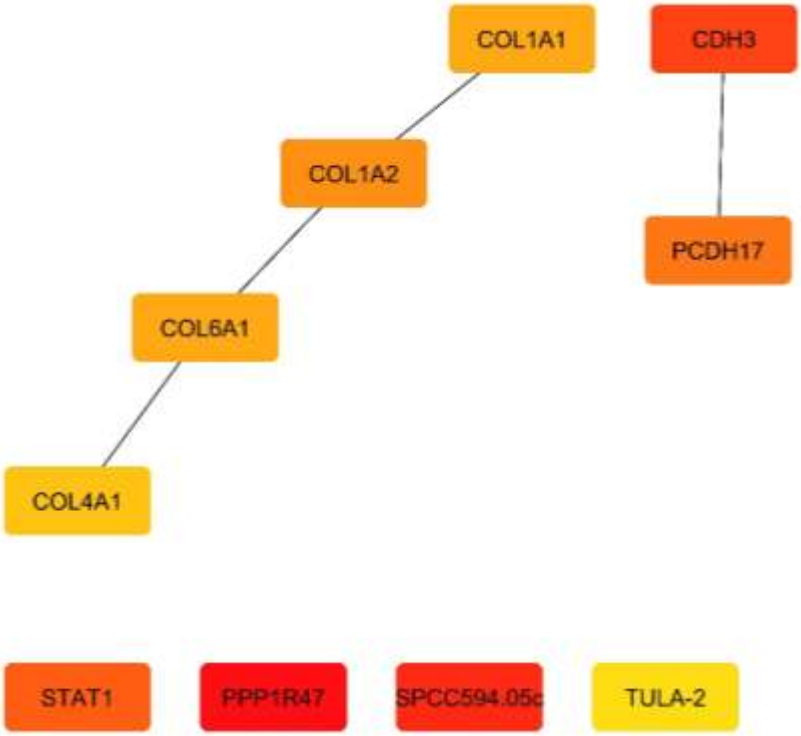

Figure 2J: MNC

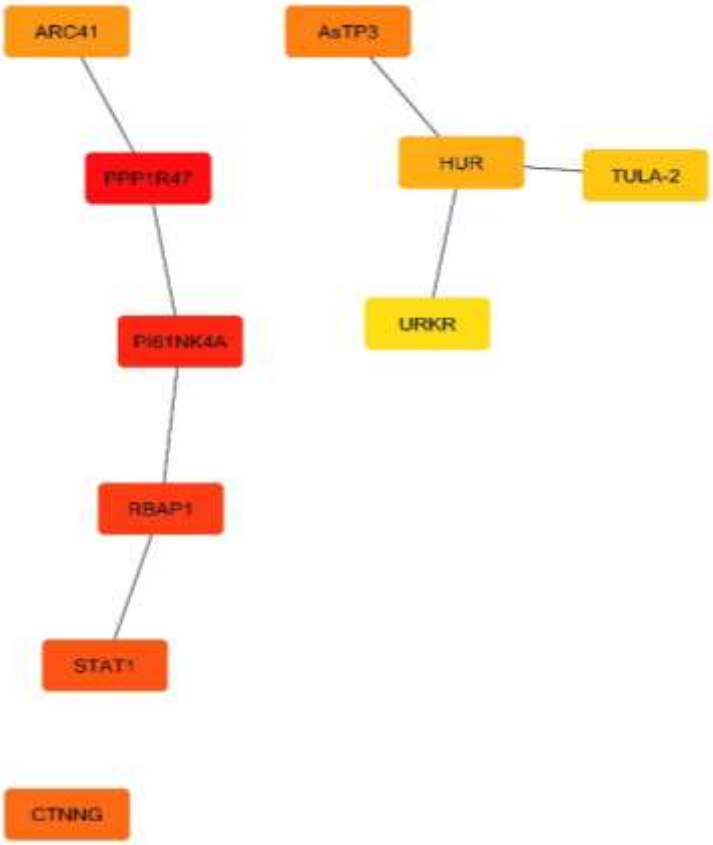

Figure 2K: Radiality

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

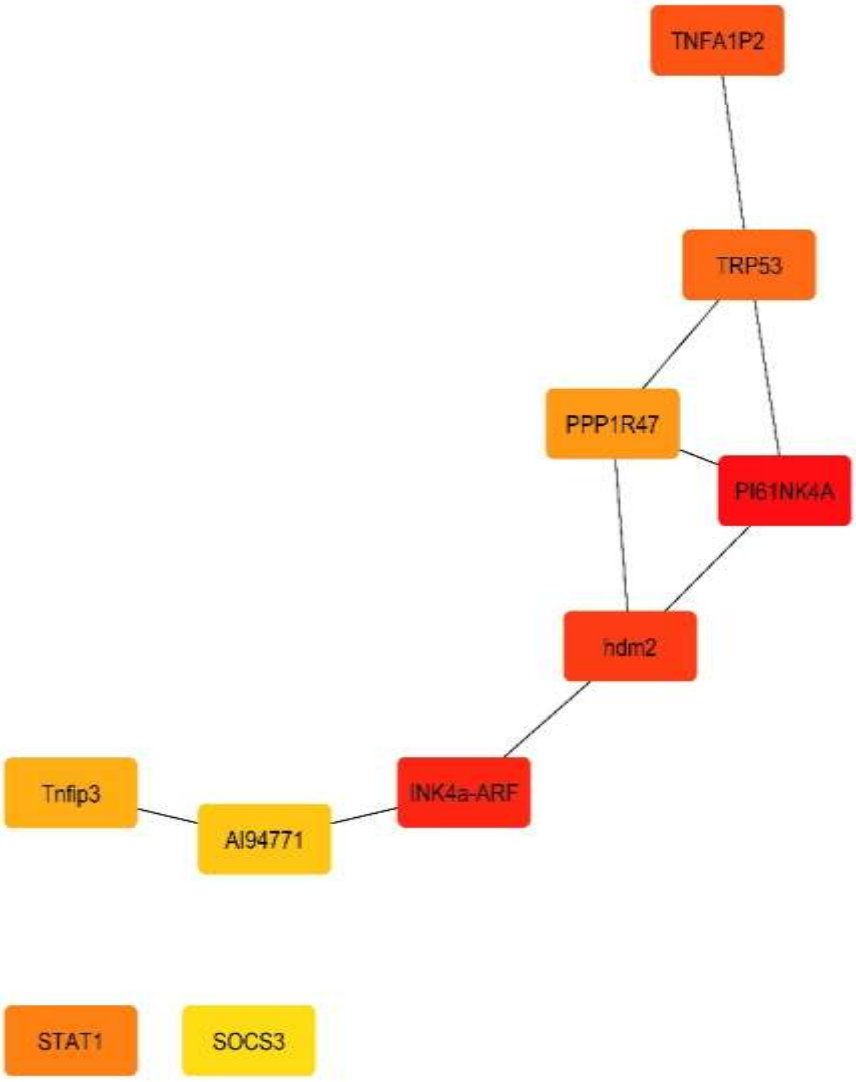

Figure 2L: Stress

Role of STAT1 in OSCC Patient Survival

The survival of the oral cancer patient with respect to the expression of the gene STAT1 was analyzed to determine how effectively this gene coordinate with other cancer gene expression leading to cancer progression. Based on analysis using available data from TCGA, the following Box-Whisker and Kaplan-Meier plot was constructed (Figure 3 and 4) which revealed the role of STAT1 in OSCC progression as crucial. As the expression of STAT1 increased, the progression of OSCC in patient increased with decrease in prognosis of the patient. In contrast, decreased level of STAT1 in OSCC patient lead to decreased level of OSCC progression and finally a higher prognosis. Thus, increased STAT1 correlates with poor prognosis of OSCC.

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

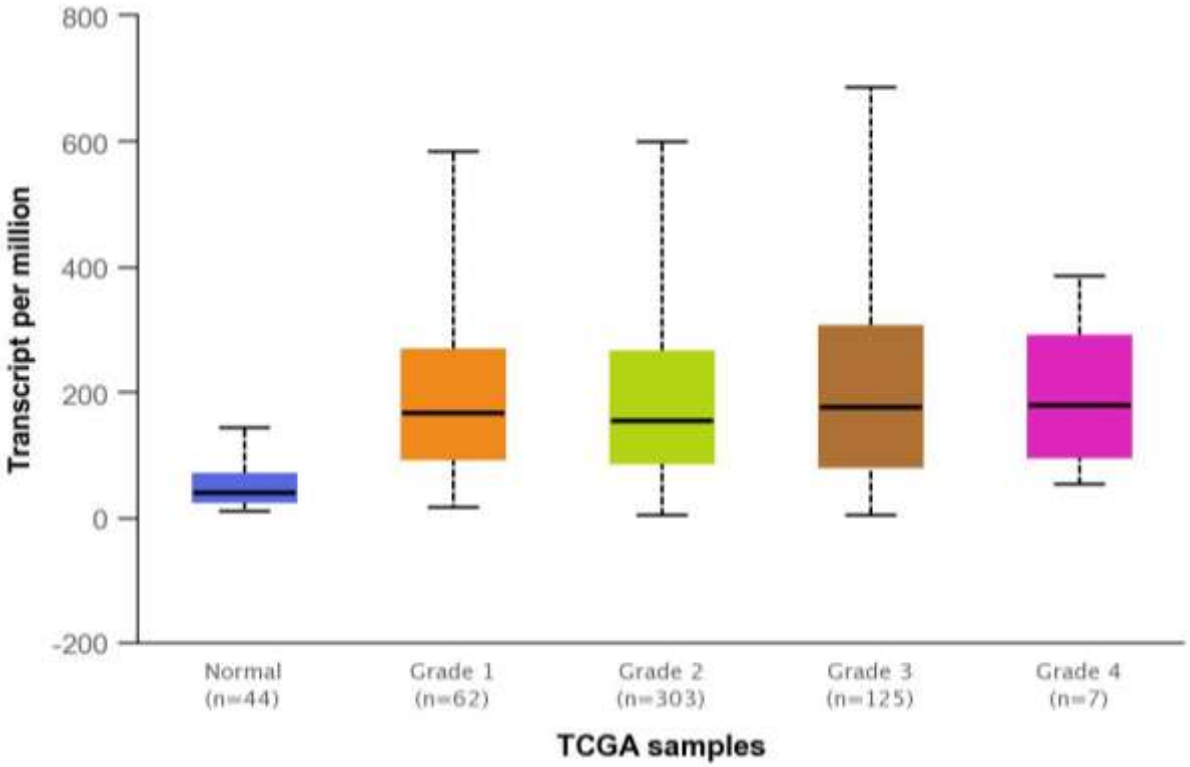

Figure 3A: Expression of STAT1 in HNSC Based on Tumor Grade

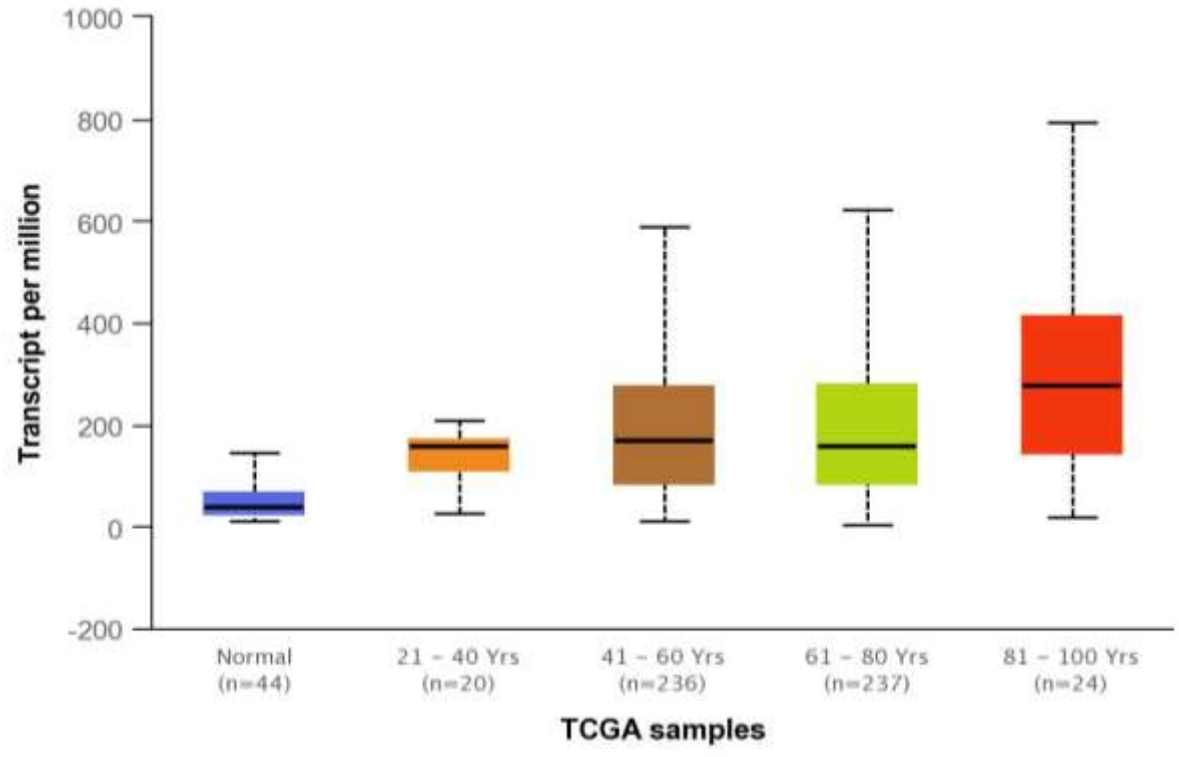

Figure 3B: Expression of STAT1 in HNSC based on patient's age

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

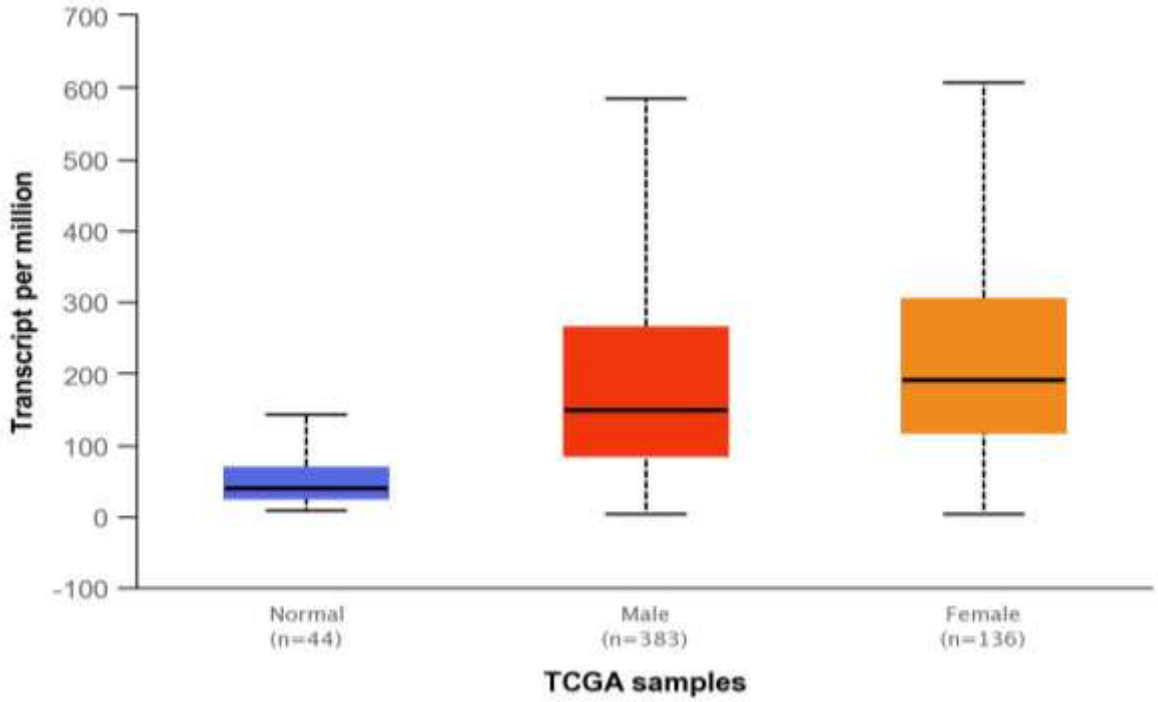

Figure 3C: Expression of STAT1 in HNSC Based on Patient's Gender

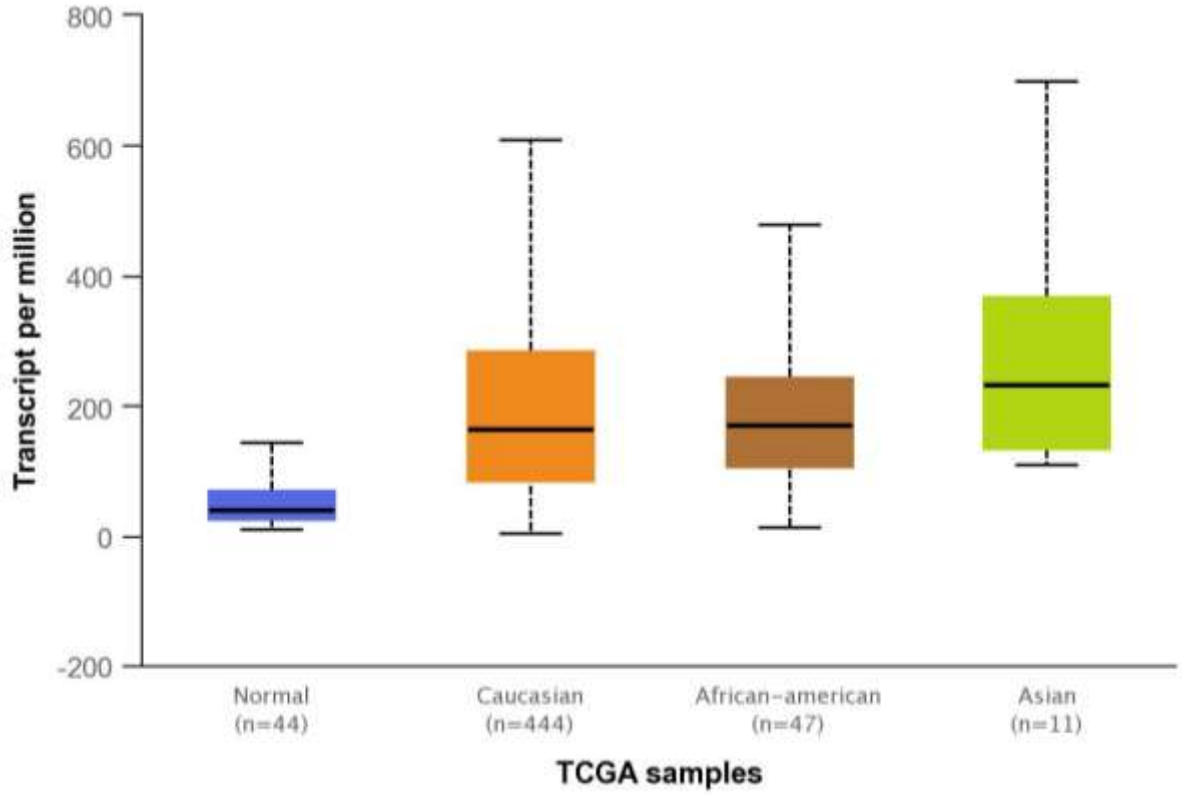

Figure 3D: Expression of STAT1 in HNSC Based on Patient's Race

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

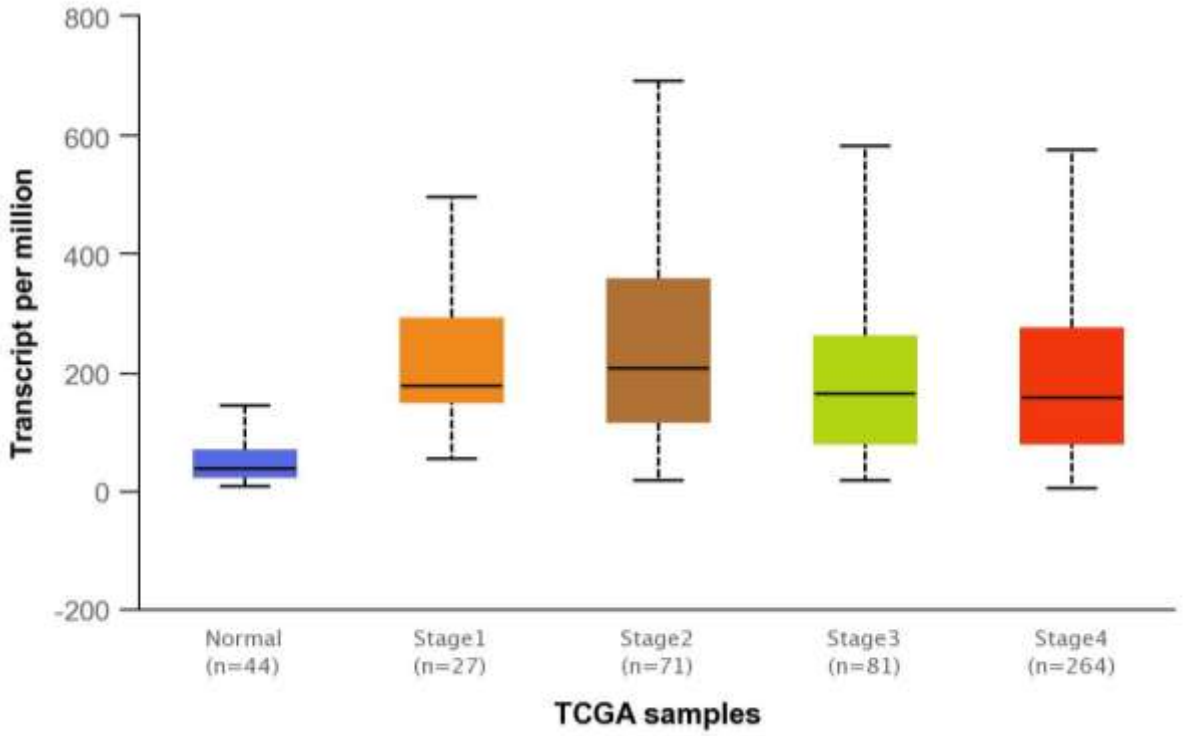

Figure 3E: Expression of STAT1 in HNSC Based on Individual Cancer Stages

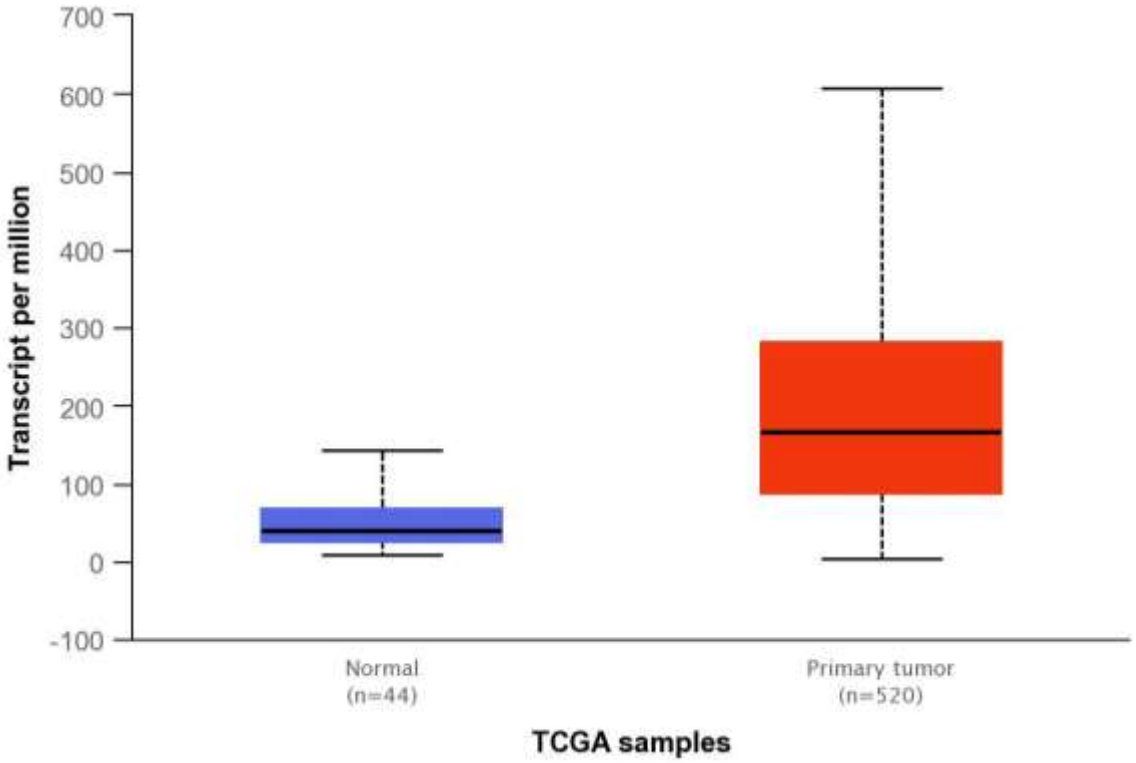

Figure 3F: Expression of STAT1 in HNSC Based on Sample Types

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

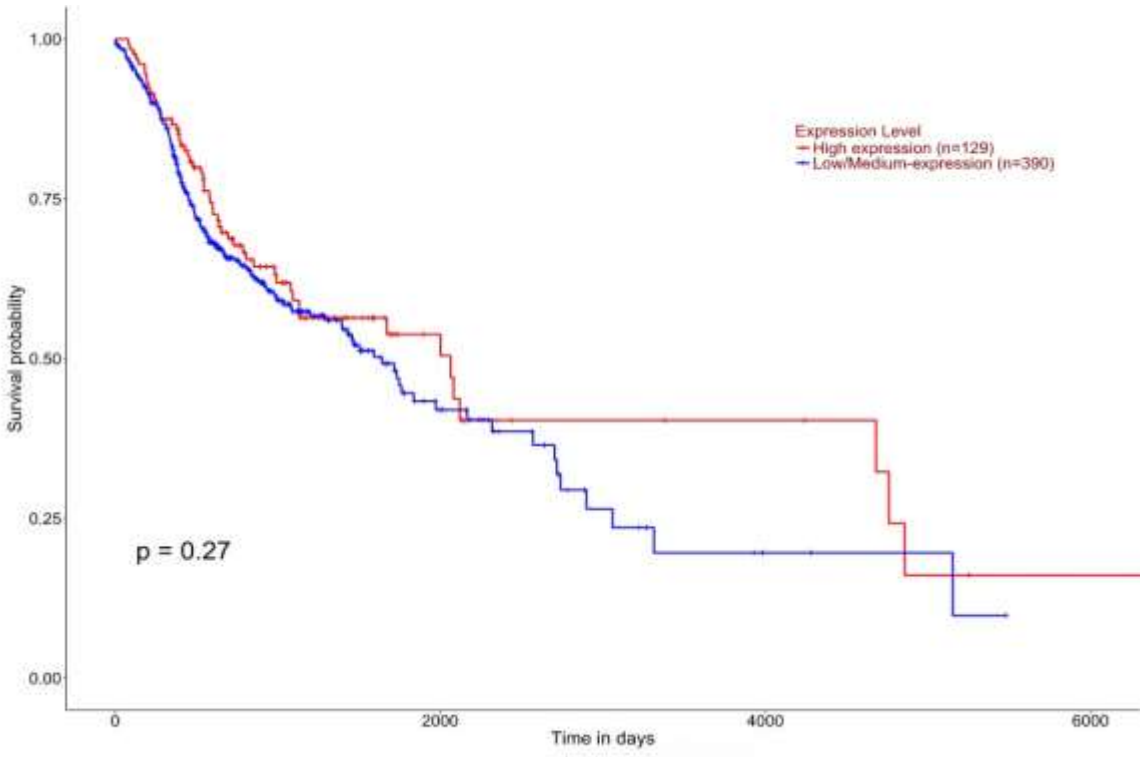

Figure 4A: Effect of STAT1 Expression Level on HNSC Patient Survival

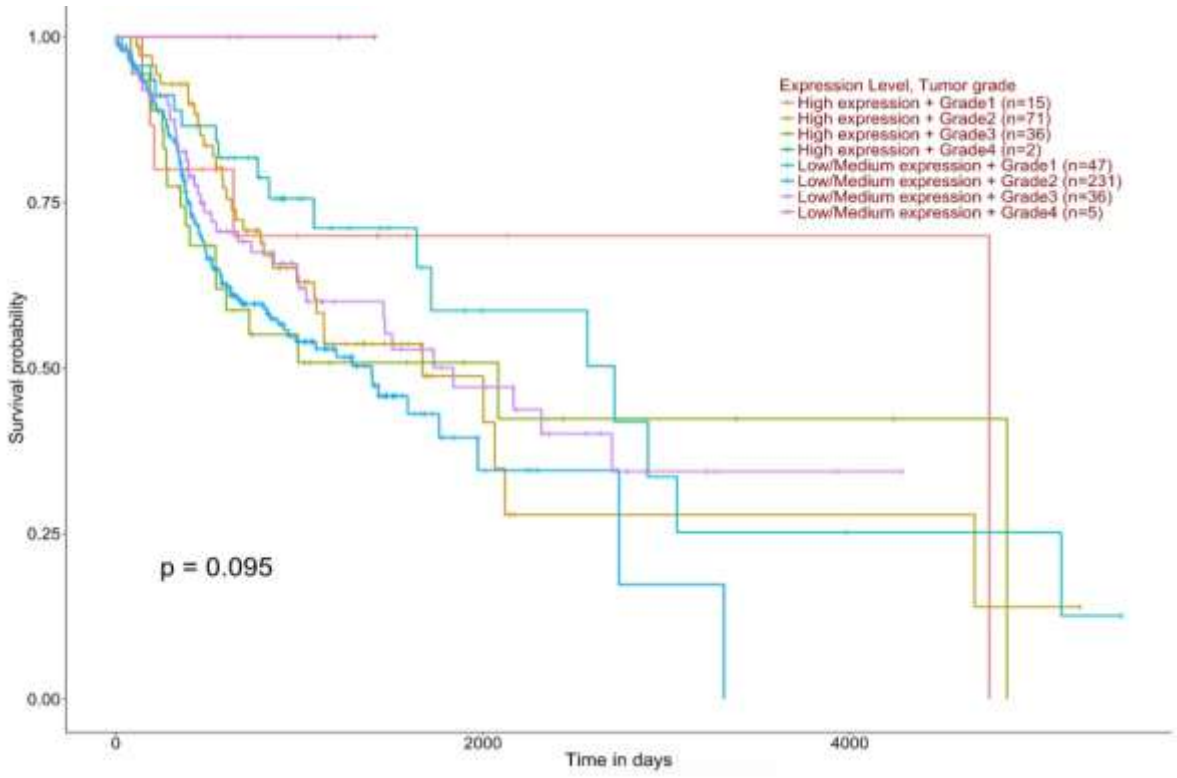

Figure 4B: Effect of STAT1 Expression Level and Tumor Grade on HNSC Patient Survival

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

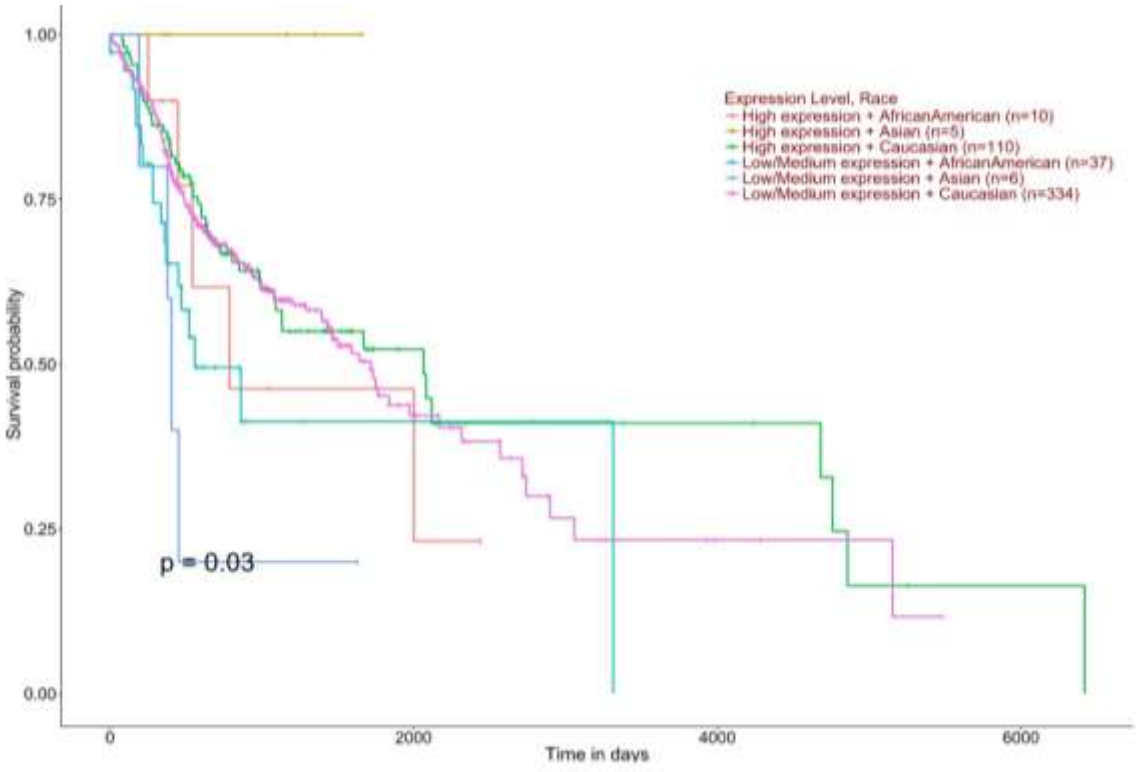

Figure 4C: Effect of STAT1 Expression Level and Race on HNSC Patient Survival

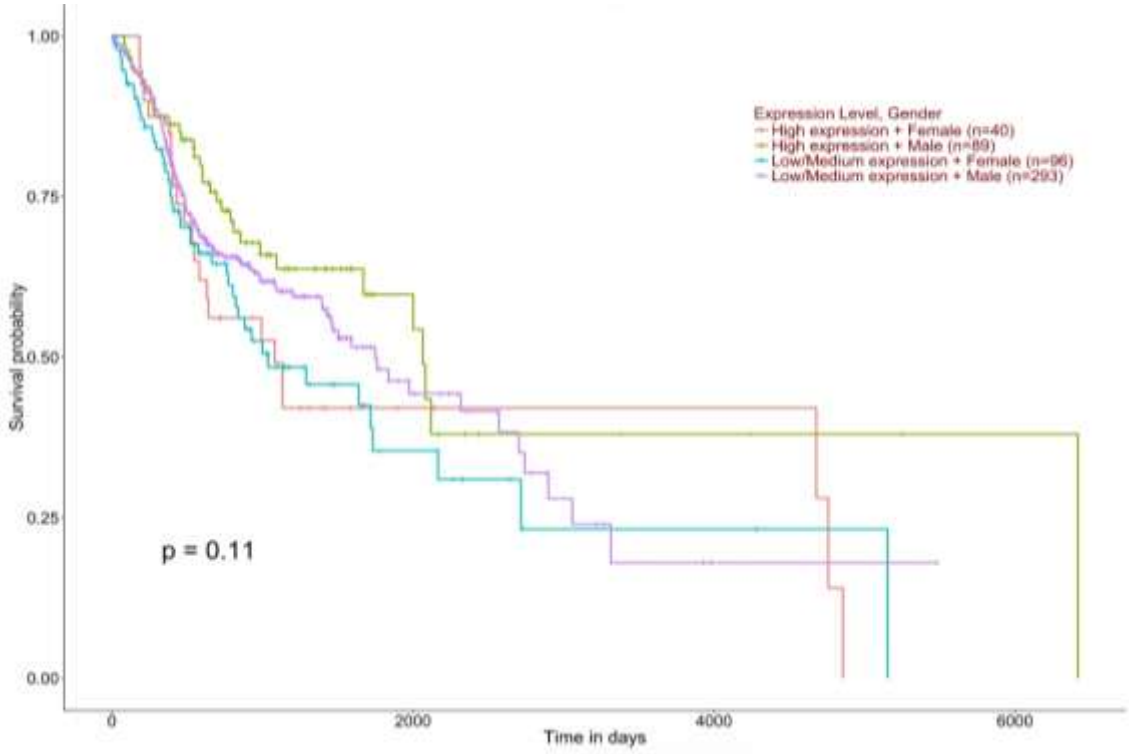

Figure 4D: Effect of STAT1 Expression Level and Gender on HNSC Patient Survival

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

Cross-Validating the Identified Gene Expression

The role of STAT1 expression in other cancer was also cross validated using an integrated TCGA analysis. A PanCancer analysis was carried out finally to determine the proficiency of STAT1. As shown in the figure 5, STAT1 expression was seen to be significant in different cancers. Hence it is clear that in most of the cancers, the level of STATA1 was found to be increased than normal.

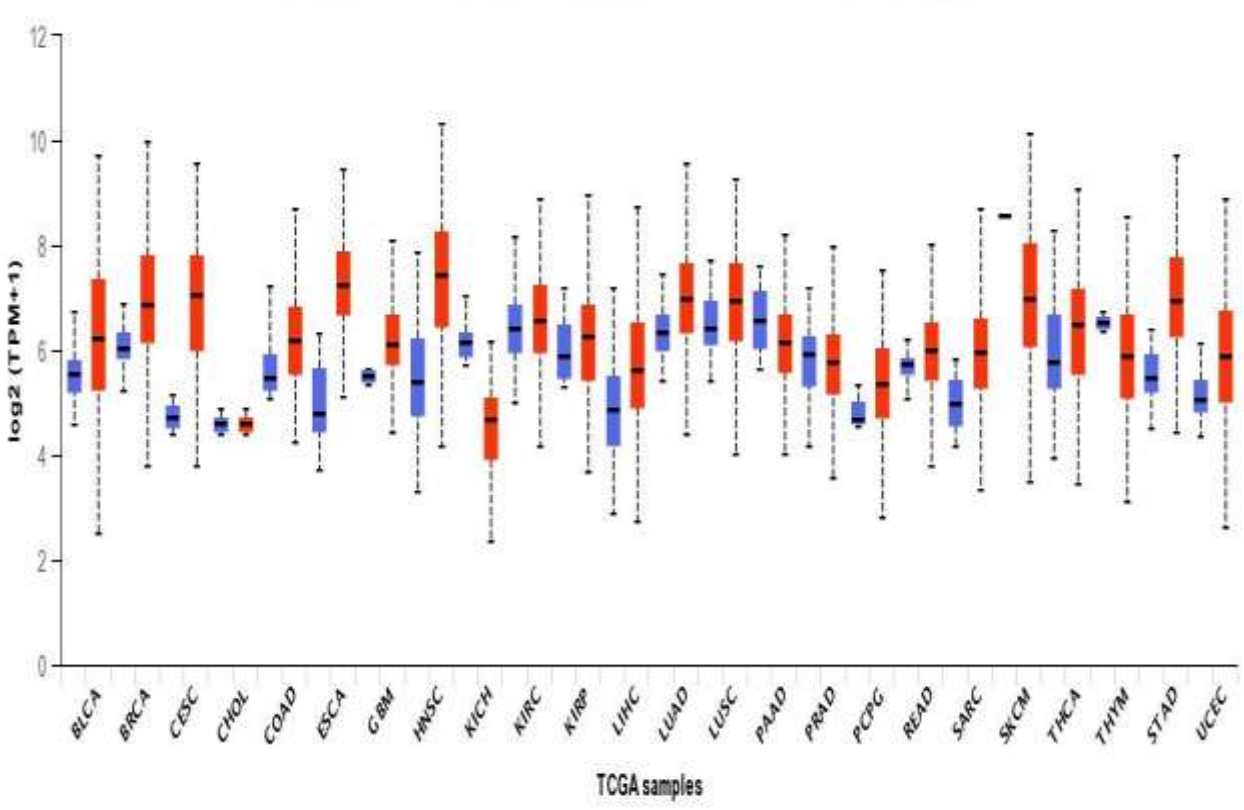

Figure 5: Expression of STAT1 Across TCGA Cancers (with Tumor and Normal Samples)

STAT1 in Tumor

Among the seven different STAT family members, the role of STAT1 in cancer is crucial. It plays a vital role in interferon signaling and thus maintaining several cellular functions. This is aided by the stimulation mechanisms working in hand with cytokines, hormones and growth factors. Several experiments proved the role of STAT1 in promoting cancer by working as oncogene (18). The signal transduction for STAT1 happens from the cytoplasmic domain found in transmembrane receptor. From here, the signal passes into the nucleus and thus the gene expression is regulated. This in turn modulates several cellular processes like cell death, differentiation and proliferation (19).

IV. CONCLUSION

The gene expression data analysis using microarray expression profiles from twelve samples of normal and OSCC cancer patients were carried out. Protein interaction network analysis and followed by topological interaction studies revealed Signal Transducer and Activator of Transcription 1 (STAT1) as one of the major master regulators

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

of oral squamous cell carcinoma. It was found that increased expression of STAT1 leads to increase in tumor size and stage and consequently, poor prognosis of OSCC. Along with using STAT1 as a biomarker for OSCC, drug discovery and development aiming to reduce the expression of STAT1 in OSCC patients will have a better effect in increasing the cancer prognosis.

REFERENCES

[1] Johnson NW, Jayasekara P, Amarasinghe AA. Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and etiology. *Periodontol* 2000. 2011;57:19–37.

[2] Ahmed SP, Jayan L, Dineshlumar T, Raman S. Oral Squamous Cell Carcinoma under microscopic vision: A review of histological variants and its prognostic indicaators. *SRM J Res Dent Sci* 2019;10:90-7.

[3] Pires FR, Ramos AB, Oliveira JB, Tavares AS, Luz PS, Santos TC. Oral squamous cell carcinoma: clinicopathological features from 346 cases from a single oral pathology service during an 8-year period. *Journal of applied oral science: revista FOB*. 2013;21(5):460–7. 10.1590/1679-775720130317

[4] Ohno S. Major sex-determining genes. Springer-Verlag; *Berlin, Germany:* 1979

[5] Oghumu S, Knobloch TJ, Terrazas C, Varikuti S et al. Deletion of macrophage migration inhibitory factor inhibits murine oral carcinogenesis: Potential role for chronic pro-inflammatory immune mediators. *Int J Cancer* 2016 Sep 15;139(6):1379-90.

[6] Edgar Ron, Michael Domrachev and Alex E. Lash. Gene Expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acid Research;* 2002; 30(1):207-10.

[7] Barrett Tanya, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013; 41:D991-5.

[8] Gentleman Robert C, Vincent J Carey, Douglas M Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;5(10):R80.

[9] Benjamini Yoav and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*. Series B (Methodological) 1995; 57(1):289-300.

[10] G. Dennis, Jr, Brad T Sherman, Douglas A Hosack, Jun Yang, Wei Gao, H Clifford Lane, and Richard A Lempicki. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003; 4(9): R60.

[11] M. Kanehisa, Yoko Sato, Masayuki Kawashima, Miho Furumichi and Mao Tanabe. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44: D1, 2016, D457–D462.

[12] E. Eden, Roy Navon, Israel Steinfeld, Doron Lipson and Zohar Yakhini. G Orilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48 (2009).

[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.

[14] Villaveces Jose M, Rafael C. Jimenez, and Bianca H. Habermann. PsicquicGraph, a BioJS component to visualize molecular interactions from PSICQUIC servers. Version 1F1000Res. 2014; 3-44.

[15] Chin Chia-Hao, Shu-Hwa Chen, Hsin-Hung Wu, Chin-Wen Ho, Ming-Tat Ko and Chung-Yen Lin. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014; 8(4):1-7.

[16] Darshan S. Chandrashekar, Bhuwan Bashel, Sai Akshaya Hodigere Balasubramanya, Chad J. Creighton, Israel Ponce-Rodriguez and Balabhadrapatruni V.S.K.Chakravarthi. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 2017; 19(8): 649–658.

[17] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015; 347(6220):1260419.

[18] Zhang Y, Liu Z. STAT1 in cancer: friend or foe? Discov Med. 2017, 24 (130): 19-29.

[19] Meissl K, Maschler SM and Strobl MB. The good and the bad faces of STAT1 in solid tumours. *Cytokine*. 2017, 89: 12-20.

DOI: 10.37200/IJPR/V24I5/PR201706
Received: 07 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020