



Research article

Integrated multiplex network based approach for hub gene identification in oral cancer

S. Mahapatra ^{a,*}, R. Bhuyan ^b, J. Das ^c, T. Swarnkar ^a^a Department of Computer Application, Siksha O Anusandhan Deemed to be University, Bhubaneswar, India^b Department of Oral Pathology & Microbiology, Siksha O Anusandhan Deemed to be University, Bhubaneswar, India^c Centre for Genomics & Biomedical Informatics, Siksha O Anusandhan Deemed to be University, Bhubaneswar, India

ARTICLE INFO

Keywords:

Hub genes
Multiplex network
Multi-omics data
Oral cancer

ABSTRACT

Background: The incidence of Oral Cancer (OC) is high in Asian countries, which goes undetected at its early stage. The study of genetics, especially genetic networks holds great promise in this endeavor. Hub genes in a genetic network are prominent in regulating the whole network structure of genes. Thus identification of such genes related to specific cancer types can help in reducing the gap in OC prognosis.**Methods:** Traditional study of network biology is unable to decipher the inter-dependencies within and across diverse biological networks. Multiplex network provides a powerful representation of such systems and encodes much richer information than isolated networks. In this work, we focused on the entire multiplex structure of the genetic network integrating the gene expression profile and DNA methylation profile for OC. Further, hub genes were identified by considering their connectivity in the multiplex structure and the respective protein-protein interaction (PPI) network as well.**Results:** 46 hub genes were inferred in our approach with a high prediction accuracy (96%), outstanding Matthews coefficient correlation value (93%) and significant biological implications. Among them, genes PIK3CG, PIK3R5, MYH7, CDC20 and CCL4 were differentially expressed and predominantly enriched in molecular cascades specific to OC.**Conclusions:** The identified hub genes in this work carry ontological signatures specific to cancer, which may further facilitate improved understanding of the tumorigenesis process and the underlying molecular events. Result indicates the effectiveness of our integrated multiplex network approach for hub gene identification. This work puts an innovative research route for multi-omics biological data analysis.

1. Introduction

Cancer has emerged as a deadly disease killing people worldwide. Head and neck squamous cell carcinoma (HNSCC) is the most occurring cancer in south Asia with an annual incidence of more than 58% in India Kulkarni (2013); Joshi et al. (2014). Oral cancer (OC) is the most common subtype of HNSCC. In India, more than two-third of the oral cancer patients present in the advanced stage of cancer Kulkarni (2013). Despite the significant progress in molecular diagnosis, early detection of OC is still not available. Early detection of any cancer type is challenging Schiffman et al. (2015). Hence, precisely predicting the disease will add significant value in treating patients. Historically it is known that cancer occurs due to mutation in genes Hanahan and Weinberg (2011). This boosts the extensive analysis of genetic networks using data mining

algorithms and machine learning tools for the identification of significant genes involved in the disease progression. Furthermore, the study of gene expression levels to detect or discriminate cancer or its subtypes justifies the existence of a strong correlation between gene expression and disease types Marisa et al. (2013); Randhawa and Acharya (2015). However, gene expression data is limited with high noise, fewer number of samples in comparison to a large number of genes. DNA methylation is a heritable epigenetic mark that is essential for normal cell development Jin et al. (2011). Aberrant methylation pattern significantly modulates gene expression levels and results in the development of cancer Wajed et al. (2001). Hence, a combined study of gene expression and DNA methylation pattern can help in identifying potential genetic indicators of cancer and allow for prevention and early detection, which is crucial to cancer treatment.

* Corresponding author.

E-mail address: saswatimahapatra2015@gmail.com (S. Mahapatra).<https://doi.org/10.1016/j.heliyon.2021.e07418>

Received 12 October 2020; Received in revised form 27 January 2021; Accepted 23 June 2021

In network theory, a group of highly connected nodes is known as hub nodes and these sets of nodes are found to be significant in many networks. Hub nodes in a genetic network are expected to play an influential role in terms of regulating the whole network structure. Langfelder et al. addressed the identification of intramodular hubs in multiple genomic networks Langfelder et al. (2013). The selected hub genes were biologically meaningful and can be considered for prognostic applications. The growing availability of diverse omics data enhances the study of biological systems at different levels. In this context, different integrative pipelines of multi-omics data analysis have been suggested. Study reveals that transcriptomic data along with the know-how of the existing gene network provides a new way of exploring the genetic associations for candidate gene selection and functional module identification. This combination gives in-depth biological interpretation and enhanced statistical analysis. A dense subgraph based integrated approach was deployed in Swarnkar et al. (2015), that combined expression analysis of genes with protein-protein interaction (PPI) networks to identify functionally enriched genes. The extracted dense subgraphs spoke for densely connected and highly co-expressed clusters from biological networks. Gadaleta et al. (2017) integrated gene expression data and DNA methylation data for glioblastoma multiforme using Regression2Net and identified potential candidate genes showing significant over representation in different cancer related pathways. Eight most connected hub genes and candidate differentially expressed genes were identified in Liu et al. (2018) by integrating multiple cohort profile data sets of B-cell lymphoma. The identified biomarkers were associated with significant biological communications and could be the therapeutic targets to mark the prognostic difference between the subtypes of B-cell lymphoma. A gene co-expression network based approach was suggested in Randhawa and Acharya (2015) to identify few vital genes with prognostic significance in various stages of Oral Squamous Cell Carcinoma (OSCC). Predictive accuracy of 81% was observed for the selected genes when tested against the developed predictive model to classify the preliminary and advanced stage of OSCC tumor. Lei Chen et al. proposed an integrated computational method based on random walk with restart and shortest path algorithms to identify novel genes associated with oral squamous cell carcinoma which opens an opportunity to the in-depth study of OC Chen et al. (2017). Data from different experiments, web sources, and PPIs were integrated in Kumar et al. (2017) to identify candidate genes of OC. The 39 selected candidate genes from this approach were significantly enriched with different biological processes involved in OC. Li et al. (2018) compared the mean expression levels of genes between diseased and control samples to identify differentially expressed and differentially variable genes for OSCC. In our previous work Mahapatra et al. (2018), we had proposed an integrated pipeline of biological networks for the identification of candidate genes specific to a disease. In the first step of the pipeline, the constructed network of co-expressed genes in combination with a known PPI network resulted in an induced network that preserved the interplay and relationship of genes/proteins. Further, densely connected modules were extracted and hub genes in the modules were identified. The selected dense modules were observed to be biologically significant and statistically significant with high predictive ability.

Nevertheless, the in-depth analysis of biological networks holds a great promise in the identification of such candidate biomarkers and plays a decisive role in deciphering the biological mechanisms behind complex diseases. However, these biological networks consist of different types of links among the interacting units, each link encoding for a type of interaction. As a result, these networks show a higher complexity describing inter-dependencies within and across various networks which may fluctuate. Subsequently, it is not able to record the information characterizing the complex biological system Doncheva et al. (2018). Traditional study of genetic networks assumes that gene-gene associations are exclusive, unvarying and encapsulating all information between them. Simultaneously it ignores the presence of multiple types of gene-gene interactions (multiplexity), which may lead to misleading

of results and generation of false positives De Domenico et al. (2013); Kanawati (2015). In this context, multiplex network provides a way to model and analyze such type of complex real world networks by integrating multiple types of interactions among different biological units. A multiplex framework was proposed to study the effect of social relationships on employee performance, which represented the entire employee social network as a multiplex structure Cai et al. (2018). In comparison to the singleton network model, better reasoning of employee performance was found through the multiplex network model that integrated the various types of social relations. Finn et al. outlined the use of multilayer network in animal behavior analysis Finn et al. (2019). The effectiveness of the multilayer network was analyzed to uncover eco-evolutionary dynamics of animal social behavior. Chierici et al. Chierici et al. (2020) proposed an integrative network fusion pipeline, using Similarity Network Fusion (SNF) in a machine learning framework for biomarker identification in cancer. The exploitation of multilayer network analysis in biological networks is sparse; however, clearly relevant with the context. In a very recent study, a constraint-based PageRank algorithm: iRank was proposed to prioritize cancer genes for hepatocellular carcinoma (HCC) using a multiplex network of omics data Shang and Liu (2020). Gene Regulatory Network (GRN) was utilized as a core-level network in the multiplex system and other omics level networks were used to define interactions across different levels (DNA level, RNA level). The identified cancer genes were observed with improved rank and better classifier performance by empirically integrating more levels of omics data. Wang et al. proposed a multiplex network based data integrative approach and identified 3 subtypes for glioblastoma multiforme (GBM) and breast invasive carcinoma (BIC) Wang et al. (2016). The result obtained was highly consistent with the state-of-the-art techniques (Normalized Mutual Information > 0.8). A multiple network based algorithm: EMDN, was proposed in Ma et al. (2017) that integrates multiple networks of gene differential co-expression and gene differential co-methylation to identify the epigenetic modules. EMDN algorithm proceeds in three steps: (i) prioritizing the seed (ii) identification of module by seed expansion (iii) refinement of candidate modules. The spotted epigenetic modules were highly enriched by known biological pathways and can serve as biomarkers to predict breast cancer subtypes. The multi-network analysis of different types of omics data was highly effective in correlating and integrating multiple data levels and discovering complex disease patterns with multiple facets. To the best of our knowledge, the outlined studies of biomarker identification for OC potentially investigated the driving force of a gene on a single omics level for cancer progression. Gene-gene co-expression was adopted as the most common type of correlation to understand the network of genes. The comprehensive contribution of gene correlation at multiple biological levels was not considered.

In the present study, we proposed a multiplex network based integration strategy for the identification of the hub genes for Oral Cancer (OC). The novelty in our approach lies in (1) Construction of three integrated networks for analyzing the comprehensive contribution of gene correlations across transcriptomic and epigenetic levels in OC. (2) Identification of few hub genes fulfilling the defined ensemble norm of multi-degree centrality and intramodular connectivity. Concretely, individual networks of gene co-expression (GCoExNW) and gene co-methylation (GCoMythNW) were constructed out of 120 samples of gene expression and DNA methylation data. Further, along with the multiplex network, an union and intersection networks were constructed by integrating the co-expression and co-methylation relation between genes. Hub genes were identified from the constructed integrated networks employing two different selection criteria defined in the following sections. Extracted hub gene sets from the constructed multiplex network were compared with the hub genes identified from the union network and intersection network. The workflow of our proposed approach is illustrated in Fig. 1. The identified hub genes were evaluated in terms of: i) Statistical Competence ii) Differential Expression Analysis iii) Biological significance analysis.

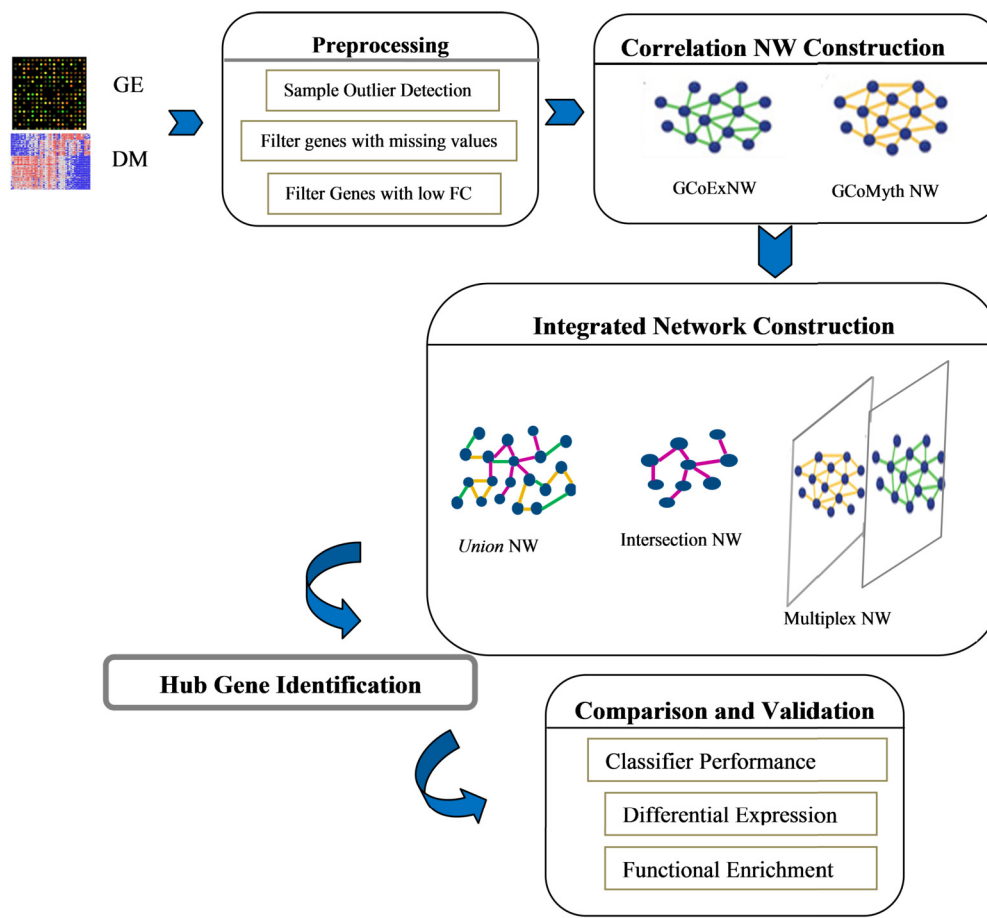


Fig. 1. Systematic illustration of pipeline for Hub Gene selection using integrated multiplex network based approach. Preprocessing step removes noise present in the data; Correlation NW construction step constructs two separate networks of gene co-expression and co-methylation; Integrated network Construction constructs three different integrated networks using the individual networks of gene co-expression and co-methylation; Hubgene identification step filters significant hub genes from three integrated networks; Finally the last step is statistical and biological validation of handpicked hub genes.

2. Materials and methods

2.1. Dataset

Gene expression and DNA methylation data for head and neck squamous cell carcinoma (HNSCC) were considered for our study, comprising of 13 primary sites of head and neck region. The data were downloaded from GDC data portal of TCGA website.¹ Specifically, 120 matched samples for level-3 data set of gene expression and DNA methylation were obtained, consisting of 100 tumor samples and 20 normal samples. The gene expression data were generated using RNA sequencing (RNA-seq) technology in TCGA project. For DNA methylation data, β intensity of the signal of probes was used for analysis. The β values are continuous valued, with a minimum value of 0 (unmethylated) and a maximum value of 1 (completely methylated). Annotation table of the Illumina Human-Methylation27 platform was used to map the probe IDs to corresponding gene symbols. For the gene symbols corresponding to multiple probe IDs, the β values of probes were averaged representing the β value of the gene. In the present study, the NCBI gene dataset² was utilized for biological relevance analysis of the distinguished genes. It includes 8933 cancer related genes and 316 oral cancer genes. STRING interactome database (Version 11.0)³ was used

to further validate the identified genes by mapping them to the respective PPI network of genes/proteins.

2.2. Weighted gene correlation network analysis

Correlation among expression level and methylation intensity of genes was used to decipher their functionality. In this context, Weighted Gene Correlation Network Analysis (WGCNA) Langfelder and Horvath (2008) in R, was used to construct a weighted network of correlation and to discover modules of highly correlated genes. WGCNA is a well established method in the field of biological science for the analysis of gene co-expression networks Padi and Quackenbush (2015); Nangraj et al. (2020). It computes Pearson correlation coefficient to measure the correlation among genes. Construction of scale free network strongly depends on *soft threshold power* β . Co-expression similarity is raised to the power of β to compute adjacency Langfelder et al. (2007). The functions `pickSoftThreshold` and `scaleFreePlot` of the WGCNA package, assists respectively in selecting a proper β value to compute adjacency and for achieving a network with scale free topology. The best value of β is chosen when the curve showing the relation between scale free topology fit index R^2 and β reaches a saturation point.

2.3. Multiplex network construction

A multiplex network can be defined as assemblages of distinct networks. Individual network encodes for a different layer of the multiplex system. Each layer in the multiplex system contains the same set of en-

¹ <https://portal.gdc.cancer.gov/>.

² <http://www.ncbi.nlm.nih.gov/geo/>.

³ <https://string-db.org/>.

tities (nodes) but interconnected by different types of links (edges). The layers are connected (and hence coupled) to each other via interlayer edges Kanawati (2015); Boccaletti et al. (2014). A multiplex network can be utilized for modeling different types of networks such as:

- *Multi relational network*: Multi relational network is a platform for modeling a complex system of networks that encodes different types of association among participating units. The link between nodes in a layer depicts a type of interaction. In this work we have considered co-expression and co-methylation similarity among genes as multiple types of relationships.
- *Dynamic Network*: where edges within a layer represent the network state at a particular timestamp.
- *Attributed Network*: where based on the similarity measure applied to a group of nodes, a similarity graph can be constructed which can be redefined as an additional layer.

MuxViz is an open-source R software package with a graphical user interface, used specially for the visualization and analysis of multiplex networks De Domenico et al. (2015). MuxViz consists of a collection of algorithms for measuring the “node centrality”, “interlayer correlation and reducibility”, “identifying community structure in the network” and “motif analysis” De Domenico et al. (2013). We focused on “node centrality”, specifically *degree centrality* and *eigenvector centrality* of nodes to identify the hub genes in the integrated network.

2.4. Proposed model

The steps of our proposed integrated multiplex network based approach of hub gene identification is explained below. Workflow of the proposed work is demonstrated in Fig. 1.

2.4.1. Preprocessing

The sample matched data of Gene expression and DNA methylation were inputted separately to the preprocessing step of our proposed approach. Firstly, genes with more than 30% of missing values across the samples were filtered and further, missing values in the data were replaced with the mean of the sample Mahapatra et al. (2018). It brought about 18283 genes and 120 samples in gene expression data and 20523 genes and 120 samples in DNA methylation data. Hierarchical clustering was performed on both the data sets to remove outliers in the samples, producing 116 gene expression samples and 115 DNA methylation samples. Out of these 114 samples were common in both data set consisting of 100 tumor and 14 normal samples, which was further considered in the next step of network construction. Additionally, genes were filtered showing fold change (FC) less than 0.5 in their expression levels in normal and tumor samples, producing 10376 genes Dembele and Kastner (2014); Zhao et al. (2018).

2.4.2. Correlation network construction

Preprocessed data of gene expression and DNA methylation with 10376 genes and 20523 genes respectively were separately inputted to WGCNA functions for the construction of gene co-expression network (GCoExNW) and gene co-methylation network (GCoMythNW). WGCNA defines a correlation similarity which is raised to power β to compute adjacency similarity, so that degree distribution fits a small world network. The power value of 6 was taken where the relation between R^2 and β reaches saturation. Additionally, to obtain the similarity between genes at network topology level, a Topological Overlap Matrix (TOM) was computed from the adjacency similarity matrix. The resultant network meets the scale free independence criterion. Keeping a threshold of 0.1, the edges in the GCoEx network and GCoMyth network were scrutinized to facilitate the formation of integrated networks. BlockWiseModules function performs automatic network construction followed by identification of clusters of interconnected genes termed as modules of co-expressed genes in a block-wise manner. If the number of

genes in the data set exceeds maxBlockSize, genes will be pre-clustered into blocks whose size should not exceed maxBlockSize and the function constructs network in a block wise manner. We have set maxBlockSize to be 30. WGCNA utilizes unsupervised hierarchical clustering for module detection. Later modules in blocks are merged based on mergeCutHeight threshold. mergeCutHeight is the dendrogram cut height for module merging. Observing the cluster dendrogram and experimenting with different values for mergeCutHeight, the cut height value .30 resulted in a good set of modules of correlated genes. The intramodular connectivity of the genes within the module was computed for all genes in both the networks of GCoExNW and GCoMythNW. Intramodular connectivity is a measure of connection strength or co-expression level of a given gene with respect to all the genes of a particular module.

2.4.3. Integrated network construction

Three different integrated networks were constructed from the correlation similarity networks i.e. GCoExNW and GCoMythNW. The integrated networks formed were:

1. *Union Network*: Considering the union of all the edges in the generated GCoExNW and GCoMythNW, an integrated Union Network was formed consisting of 6189 nodes and 304656 edges. The edges in the created Union Network convey either co-methylation, co-expression or co-methylation along with the co-expression relationships among genes. Alteration of genes can be at single or multiple biological levels.
 2. *Intersection Network*: An integrated Intersection Network was created by taking the common edges between the two singleton networks i.e. GCoExNW and GCoMythNW. This intersection network consisted of 259 nodes and 357 edges. Here, genes that alter at multiple biological levels were focused. i.e. genes in this Intersection Network not only co-expressed but co-methylated also. Genes in this Intersection Network co-alter in the epigenetic as well as transcriptomic levels.
 3. *Multiplex Network*: Multiplex network provides a platform for modeling of the complex system of networks where nodes in the network are connected through the variety of links. It is represented as a network of networks, each encoding different types of association among participating units/nodes Kivela et al. (2014). Each network encodes for different layers in a multiplex system with coupling links between layers. More specifically, in the present study, we exploited the multiplex networks to model the association between functionally similar genes at multiple biological levels. Different types of functional dependencies exist among the essential molecular components of the cell across diverse biological levels Elling and Deng (2009). Because of this interdependency, the disruption caused due to aberrant methylation of a single gene can be quickly propagated to the PPI network. It leads to abnormal functioning of tissue or cell which culminate in disease. In this work, the co-expression and the co-methylation relationship between genes was utilized to construct the multiplex system. MuxViz software De Domenico et al. (2015) was used for the construction and visualization of the multiplex network.
- The multiplex network was constructed as an edge colored multi-graph consisting of two layers by taking into consideration the co-expression and co-methylation similarity among genes, where each layer contains the same sets of nodes. The two layers of the multiplex network were namely GCoEx (representing gene co-expression network) and GCoMyth (representing gene co-methylation network). 1068 number of genes were found to be common in the constructed GCoExNW and GCoMythNW. This set of 1068 genes and their involved edges in the respective GCoExNW and GCoMythNW were considered for the construction of a multiplex network. An aggregate network was also generated, where data from different interactions or layers is added or somehow packed up into a multiplex structure. However, the aggregated network may lead to loss

Table 1. Hub genes marked in the three integrated networks by deploying selection parameters C1 and C2.

Criteria 1 (C1)				Criteria 2 (C2)			
Hub	NoG	CR	OC	Hub	NoG	CR	OC
C1-HubUn	1710	706(41%)	25(2%)	C2-HubUn	721	279(38%)	10(2%)
C1-HubIn	78	33(42%)	0(0%)	C2-HubIn	40	14(35%)	0(0%)
C1-HubMN	527	218(42%)	8(2%)	C2-HubMN	233	97(42%)	8(4%)

C1-HubUn, C1-HubIn, C1-HubMN are three hub gene sets identified from Union Network, Intersection Network and Multiplex Network using criteria 1; C2-HubUn, C2-HubIn, C2-HubMN are three hub gene sets identified from Union Network, Intersection Network and Multiplex Network using criteria 2; NoG Number of genes, CR Number of Cancer related genes, OC Number of Oral cancer genes.

processes and pathways Levine and Steffen (2001). The functionality of a gene can not be studied completely as an isolated component. Because genes always interact with each other to be expressed. A group of genes can be marked as functionally analogous by mapping them to a known network. Study suggests, direct protein-protein interactions are the strongest indicator of functional association among genes and this information can be exploited for the identification of candidate genes Vinayagam et al. (2014). In this study, a set of 28 universally accepted OC genes that play a crucial role in Oral Cancer progression were considered as seed genes (Supplementary Table S2). The marked hub genes along with the seed genes were mapped to the PPI network to unveil the functionality of hub genes and strengthen the study of disease pathogenesis. STRING database Szklarczyk et al. (2015) and Cytoscape Shannon et al. (2003) were used to construct and visualize the connectivity of the seed genes and hub genes in the PPI network. Cytoscape is an open-source software used for the analysis and visualization of biological networks. stringApp of Cytoscape plug-in was employed for the analysis.

- (b) Gene Ontology (GO) functional enrichment analysis of putative hub genes: DAVID Sherman et al. (2009) was adopted to uncover and further analyze the biological meaning behind the identified optimal hub genes. DAVID is a tool for functionally classifying genes using a set of functional annotation tools and analyzing the biological role of genes. It is widely used to perform GO and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analyses of genes.

3. Results

The Hub Gene Identification step of our proposed work gave rise to six different sets of hub genes selected from the three integrated networks: Union network, Integration network and Multiplex network. The different sets of hub genes, marked by deploying criteria 1 and criteria 2 of our approach are given in Table 1.

Before moving further we did a priori validation of six hub gene sets by investigating their correspondence with the disease. The spotted hub gene sets were examined for the presence of cancer related (CR) genes and oral cancer (OC) genes as described in the section 2.4.5 (C) "Biological validation - Correspondence between the hub genes and the disease of interest". The percentage of CR and OC genes found in the hub gene sets is summarized in Table 1. All the designated hub gene sets were observed with a good proportion of co-expressed cancer candidate genes. This witness the fact that disease related genes have high interacting partners compared to other genes Wu et al. (2008). The Criteria 2 of our approach was narrower than that of Criteria 1. In Criteria 2, genes were mined based on an ensemble of gene ranking and intramodular connectivity, which results in a reduced fraction of CR genes in the respective hub gene sets. However, the hub gene

sets namely C1-HubMN and C2-HubMN identified from the constructed multiplex network were showing a highest fraction of CR genes which is consistent both for Criteria 1 and Criteria 2. Although small, a fraction of OC genes were also spotted in the hub gene sets extracted from the integrated Union network and Multiplex network, whereas no OC genes were spotted in the intersection network. A set of 74 genes, named as C1-CmnHub found to be common in three hub gene sets extracted using criteria 1 and 38 genes were found common in the six hub gene sets extracted from the integrated networks. Out of them genes ACY2, AKAP6, ASPM, F8, FCGR3A, FOXO4, HSDL2, LILRB1, ME3, MUC21, NDRG2, PIK3CG, TNNC1, SIGLEC9 were examined as related to cancer. Additionally, eight genes: CCL18, CCL4, TERT, RECK, LEPR, CDC20, LTA, HLADQA which are known to be oral cancer were spotted both in the hubs C1-HubMN and C2-HubMN. An intersection chart showing the common genes spotted in different hub gene sets is illustrated in Figs. 3a and 3b. These genes were highly connected as well as showing a significant correlation between expression and methylation levels in the multiplex network which may have a regulatory effect on related genes. The identified hub gene sets were further validated in terms of their prediction accuracy in the succeeding section.

3.1. Classifier performance

Hub genes in a gene network are high degree nodes having more interacting partners than non-hub nodes. These genes direct the whole network structure and play a very important role in multiple regulatory systems He and Zhang (2006). Three different classifiers KNN (k-nearest neighbors, with $k=3$, based on the different repeated experimental trials), Random Forest (RF) and Support Vector Machine (SVM) had been applied for evaluating the identified sets of hub genes to determine how good or bad they were in predicting the disease state of a test data. To avoid one time occasionality, the class outcome of the hub genes as predictor entity were examined using 10 fold cross validation. All the sets of hub genes showed a low variance in MCC for all three classifiers, which is summarized in Table 2. Each set of hub genes had achieved a standard measure of MCC value of ≥ 0.5 . However, the hub genes selected from the integrated multiplex networks were showing the highest value of MCC. All the hub gene sets achieved a comparable prediction accuracy between 92% and 96% and showed good performing scores of sensitivity, specificity, precision and F-measure with respect to various classes for different classifiers. However, among all, hubs C1-HubMN and C2-HubMN score highest in terms of sensitivity and specificity. In the medical context, sensitivity and specificity are the key measures for evaluating a diagnostic model. It is noteworthy to mention here that the genes marked as the hub in the multiplex network were co-related across transcriptomic as well as epigenetic levels and highly connected in the constructed multiplex system. Results revealed that the integrated multiplex network approach effectively identifies hub genes with improved prediction accuracy.

The objective of this work was to identify few driver genes which are crucial for understanding the genetic etiology behind a complex disease and potential candidate for identifying drug targets for OC. As all the hub gene sets showed competitive performance as a predictor

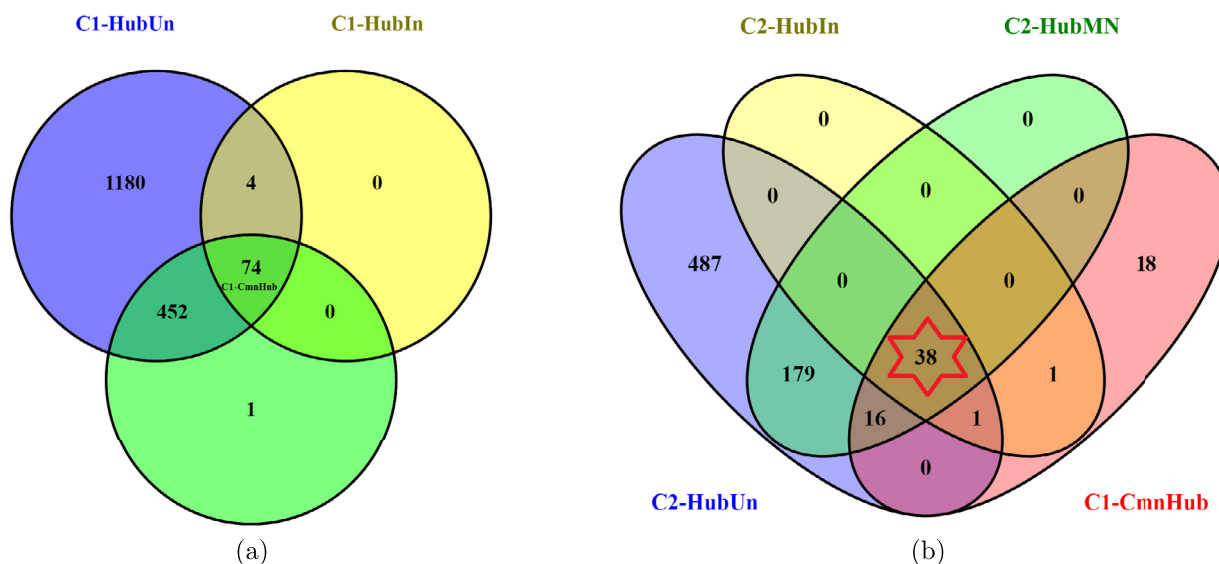


Fig. 3. Venn diagram showing different hub gene sets and their intersection. (a) Intersection of hub gene sets C1-HubUn, C1-HubIn, C1-HubMN extracted using criteria 1 shows 74 common genes: C1-CmnHub. (b) Intersection of hub gene sets C2-HubUn, C2-HubIn, C2-HubMN extracted using criteria 1 and C1-CmnHub shows 38 number of genes.

Table 2. Classifier performance of Hub genes marked in the integrated networks.

Hub	NoG	Cl	3NN					RF					SVM				
			Sn	Sp	Pr	FM	MCC	Sn	Sp	Pr	FM	MCC	Sn	Sp	Pr	FM	MCC
C1-HubUn	1710	N	0.90	0.95	0.95	0.92	0.89	0.89	0.94	0.94	0.91	0.90	0.92	0.98	0.95	0.94	0.91
		T	0.95	0.90	0.93	0.94	0.89	0.94	0.89	0.92	0.93	0.90	0.98	0.92	0.97	0.98	0.91
C1-HubIn	78	N	0.82	0.99	0.97	0.89	0.86	0.77	0.89	0.96	0.87	0.84	0.82	0.99	0.97	0.89	0.86
		T	0.99	0.82	0.93	0.96	0.86	0.89	0.77	0.92	0.96	0.84	0.99	0.82	0.93	0.96	0.86
C1-HubMN	527	N	0.85	0.99	0.97	0.90	0.96	0.87	1.00	1.00	0.93	0.93	0.92	1.00	1.00	0.96	0.95
		T	0.99	0.85	0.94	0.97	0.96	1.00	0.88	0.95	0.98	0.93	1.00	0.92	0.97	0.99	0.95
C2-HubUn	721	N	0.92	0.97	0.92	0.92	0.89	0.87	1.00	1.00	0.93	0.91	0.82	1.00	1.00	0.90	0.88
		T	0.97	0.92	0.97	0.97	0.89	1.00	0.87	0.95	0.98	0.91	1.00	0.82	0.94	0.97	0.88
C2-HubIn	40	N	0.77	0.84	0.95	0.87	0.84	0.77	0.99	0.97	0.86	0.82	0.69	1.00	1.00	0.82	0.79
		T	0.85	0.77	0.92	0.96	0.84	0.99	0.77	0.92	0.95	0.82	1.00	0.69	0.89	0.94	0.79
C2-HubMN	233	N	0.82	0.99	0.97	0.89	0.98	0.80	1.00	1.00	0.89	0.93	0.95	0.99	0.97	0.96	0.95
		T	0.99	0.82	0.93	0.96	0.98	1.00	0.80	0.93	0.96	0.93	0.99	0.95	0.98	0.99	0.95
Cmn-Hub	46	N	0.96	0.98	0.98	0.91	0.98	0.90	0.95	0.97	0.95	0.92	0.99	1.00	0.96	0.96	
		T	0.99	0.95	0.91	0.98	0.98	0.85	0.90	0.99	0.99	0.95	0.99	0.92	0.97	0.99	0.96

The hub gene sets marked in bold showing significant predictive accuracy measures.

C1-HubUn, C1-HubIn, C1-HubMN are three hub gene sets identified from Union Network, Intersection Network and Multiplex Network using criteria 1; C2-HubUn, C2-HubIn, C2-HubMN are three hub gene sets identified from Union Network, Intersection Network and Multiplex Network using criteria 2; Cmn-Hub is the hub gene set consisting of 46 common hub genes from all six integrated networks; NoG Number of genes, Cl class label, 3NN 3 nearest neighbors, RF random forest, SVM support vector machine, Sn sensitivity, Sp specificity, Pr precision, FM F-measure, MCC Matthews correlation coefficient; N normal and T tumor Oral cancer sample.

variable, the proposal of identifying hub genes boiled down to forming another hub gene set: Cmn-Hub, consisting of 46 genes. These 46 hub genes comprised of 38 common genes (found in all six hub gene sets) and 8 OC genes (spotted in both the hubs C1-HubMN and C2-HubMN). To our surprise, these 46 hub genes were outperforming with average sensitivity value of 93% and specificity value of 94% when applied to the classifiers. The observed prediction accuracy of hub gene set Cmn-Hub was 96% with an average MCC value of 0.96. ROC curve showing the comparative performance of Cmn-Hub, C1-HubMN, C2-HubMN is illustrated in Fig. 4. The common hub gene set with a smaller size (46 genes) was observed with significantly better performance as a classifier in comparison to other identified hubs. Additionally, the eigenvector centrality of hub genes in the multiplex network was observed to be greater than 0.5, indicating the strong connectivity of these hub genes with other hub genes in the network. This suggests that these genes are highly influential in the biological network and contribute most as a predictor variable. The selected hub genes were further validated in our next step in terms of their connectivity and functional association with the known set of genes for oral cancer. The list of hub genes is

given as Supplementary Table S1. Further, the differential expression analysis and biological importance of these hand picked 46 number of genes is discussed in succeeding sections.

3.2. Functional association analysis

3.2.1. Association of the selected hub genes with the seed genes

To further validate the filtered gene set, we looked for the interconnection and functional association of our studied gene sets with the 28 seed genes (Supplementary Table S2) taken based on its association in the progression of OC or Head and Neck Squamous Carcinoma (HN-SCC). Identified 46 number of hub genes along with the 28 seed oral cancer genes were mapped to PPI network using stringApp of Cytoscape which maps the genes to STRING database of interacting proteins Szklarczyk et al. (2015). STRING includes both physical interactions from experimental data and functional associations from curated pathways, automatic text mining, and prediction methods. A minimum interaction score greater than 0.5 (optimum confidence) was used and only query proteins were displayed. The co-expressed hub genes identified in the

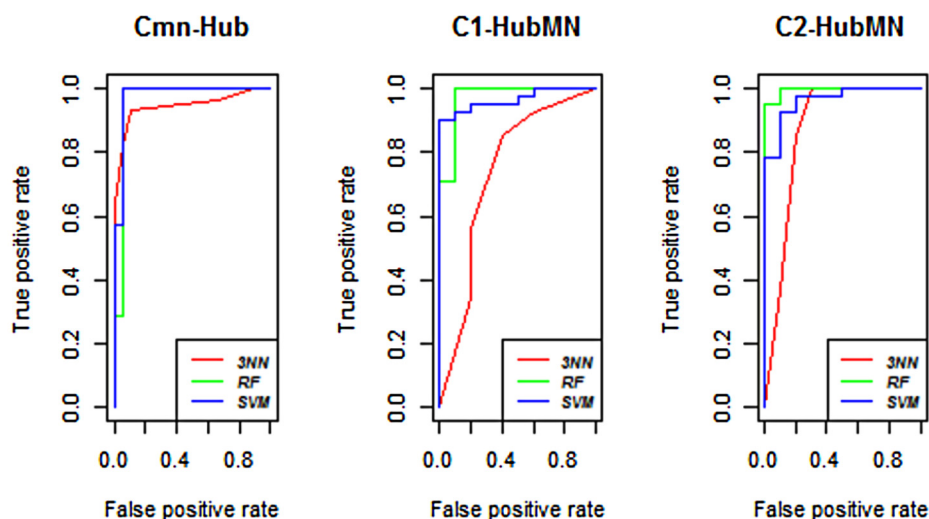


Fig. 4. Comparative performance of the identified Hub gene sets (Cmn-Hub, C1-HubMN, C2-HubMN) for three different classifiers: *k*-nearest neighbor for *k* = 3 (3NN), random forest (RF) and support vector machine (SVM).

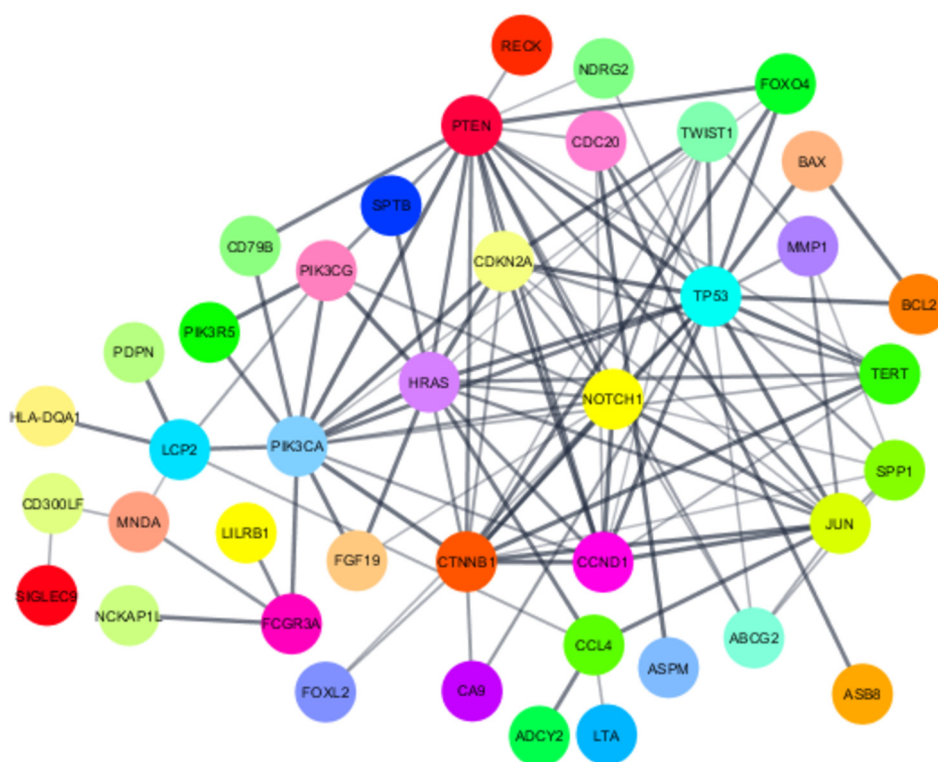


Fig. 5. Association of 46 significant hub genes and 28 seed genes in PPI network.

integrated multiplex network were also relatively closer in the PPI network, which is shown in Fig. 5. Genes viz. FOXO4, NDRG2 which are known to get involved in cancer were marked as strongly connected with the widely known cancer gene TP53. PIK3CG, CCL4 were strongly connected with HRAS, PIK3CA, NOTCH1, JUN which are known specific to OC. CDC20, RECK, NDRG2 were connected with seed gene PTEN. Hub gene TERT was highly co-related with PTEN, TP53, HRAS, PIK3CA which are benchmark genes for OC. This shows, genes that are co-altered at multiple levels of expression, also are more closer in the known biological network. These genes can be considered significant for further study of prognostic influence in oral cancer. To more accurately select the hub genes, a biological significance analysis of these hub genes had been done in our next step.

3.2.2. Identification of differentially expressed hub genes (DEHs)

DE analysis was carried out to identify the hub genes showing significant expression changes between tumor and normal samples. A filtering threshold of $\text{padj}(\text{FDR}) < 0.05$ and $\text{Log2Fold-change} > 2$ was set to identify the DEHs. Out of 46 identified hub genes, 28 genes were found to be significantly differentially expressed, which is shown in Fig. 6a. The list of DEHs is given in Supplementary Table S3. Fig. 6b demonstrates the 16 up-regulated DEHs and 12 down-regulated DEHs. The most significant up-regulated genes were CDC20, CCL4, LCP2, ASPM, FCGR3A, whereas genes NDRG2, MUC21, PERM1, TNNC1, HSPB7 were observed to be significantly down-regulated. Among them, genes CCL4, CDC20 are Oral cancer genes and ASPM, FCGR3A, TNNC1 are related to cancer.

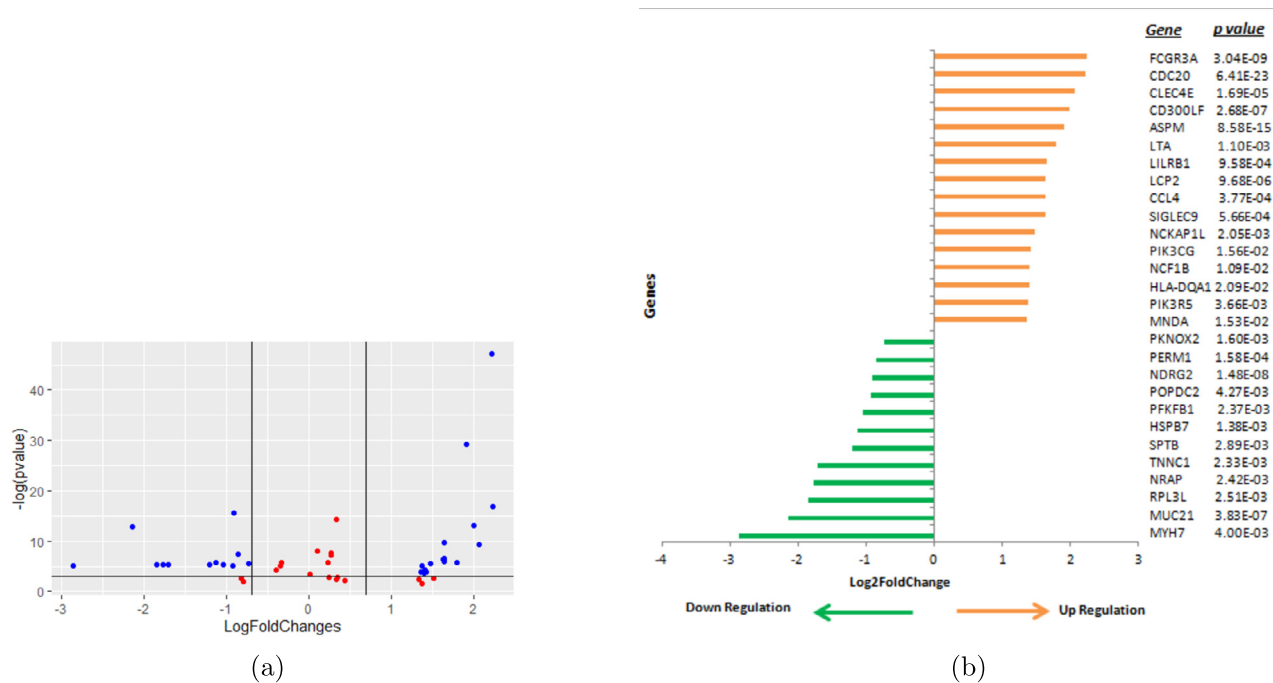


Fig. 6. Differential expression analysis of hub genes. (a) A Volcano plot: Genes represented by blue color dots are showing significantly altered expression (FDR < 0.05 and LogFC > 2) (b) 16 up-regulated and 12 down-regulated DEHs with their p-values.

3.2.3. Functional association analysis of the selected hub genes

To further unveil the biological implications behind the selected genes, Gene Ontology (GO) functional and pathway enrichment analysis of hand picked hub genes along with the considered seed genes were done using DAVID Sherman et al. (2009). A count ≥ 2 and EASE > 0.1 were set as cut off parameters. EASE is a GO enrichment analysis score based on the Fisher Exact Test Hosack et al. (2003). A threshold of $p < 0.05$ was used to indicate significant functionality and pathway categories. GO functional enrichment analysis was carried out based on three functional groups: (1) biological processes, (2) cellular components and (3) Molecular functions.

Enriched list of the biological processes involving the hub genes is summarized in Supplementary Table S4. Result revealed that the identified hub genes were significantly enriched in different biological processes including immune response, cell proliferation, apoptosis, cell cycle arrest, signal transduction, Ras protein signal transduction along with the benchmark OC genes NOTCH1, JUN and HRAS. JUN is responsible for OC metastasis whereas deviated expression of NOTCH1 is responsible for cell proliferation. Proper functioning of the immune system requires a balanced interaction between immune and non-immune cells. Chemokines regulate leukocyte migration and trafficking to their destined tissues Chakraborty et al. (2018). Encoded proteins from hub genes CCL4, CCL18 has chemokinetic and inflammatory functions. Aberrant expression of these genes may disrupt the immune system and lead to resistant tumor formation. The presence of cancer related gene PIK3CG in many biological processes regulating the immune system gives an indication to the researchers for further analysis to uncover the genetic etiology behind the disease. The rate of growth in normal and cancer tissues is determined by a balance between cell proliferation and apoptosis. As cancer cell continues to grow uncontrollably, biological activities are also disturbed in cell cycle regulation. Somatic mutation in TP53 plays a significant role in regulating cell cycle, cellular differentiation, DNA repair, apoptosis Nanda et al. (2011); Polyak et al. (1994). Gene CDC20 which was observed to be significantly up-regulated is significantly enriched in the biological process related to cell proliferation. Hub gene CDC20 was observed to be significantly up-regulated in OC. This implies that deregulated expression of CDC20 can mediate a series of downstream genes to drive cellular

proliferation and activate cell death apoptosis. Genes CASQ2, MYH6, MYH7 which are not known to be cancer related, were found to be highly enriched in GO terms related to muscle contraction. Muscle contraction has been reported as a key biological process in OC, which witnesses the presence of myofibroblasts in tumor stroma of patients with lymph node involvement Shaikh et al. (2019). A series of genes including LILRB1, FOXO4 (cancer related) were found to be predominantly enriched in cell cycle arrest, immune response, regulation of the apoptotic process. Apoptosis is one of the key processes in oral cancer progression. Additionally, gene PIK3CG, TERT, CCL4, CCL18 were found to be highly enriched with biological processes such as neutrophil chemotaxis and angiogenesis. Angiogenesis is one of the crucial factors in cancer progression Nishida et al. (2006). The presence of neutrophils at the primary tumor site activates a broad spectrum of functions leading to cancer. It generates the secretion of proangiogenic factors as well as the proteolytic activation of proangiogenic factors which causes angiogenesis Granot and Jablonska (2015). Fig. 7 shows the most significant biological processes observed in the identified hub genes.

The enumerated list of cellular components (CC) and molecular functions (MF) involved in the putative list of hub genes are summarized in Supplementary Tables S5 and S6 respectively. Our list of hub genes was found to be mostly enriched in the cytosol (CC), phosphatidylinositol 3-kinase complex (CC), plasma membrane (CC), protein complex (CC), muscle myosin complex (CC) along with the verified list of seed genes such as TP53, NOTCH1, JUN, PIK3CA, HRAS. Moreover, we also noted protein binding (MF), enzyme binding (MF), transcription factor binding (MF), kinase activity (MF), phosphatidylinositol-4,5-bisphosphate 3-kinase activity (MF) in the enriched list of molecular functions entailing genes PIK3CG, CDC20, CCL4, ADCY2 and TERT. Results suggest that there is a direct or indirect association between the functional activity and localization of the proteins encoded by the scrutinized hub genes and OC.

DAVID also produce the KEGG pathway enrichment analysis for the 46 identified hub genes with respect to the widely accepted seed genes. The highlighted pathways were p53 signaling pathway, Chronic myeloid leukemia, small cell lung cancer, Pathways in cancer, ErbB signaling pathway, PI3K-Akt signaling pathway and focal adhesion. In general, these pathways are responsible for cell cycle regulation, signal

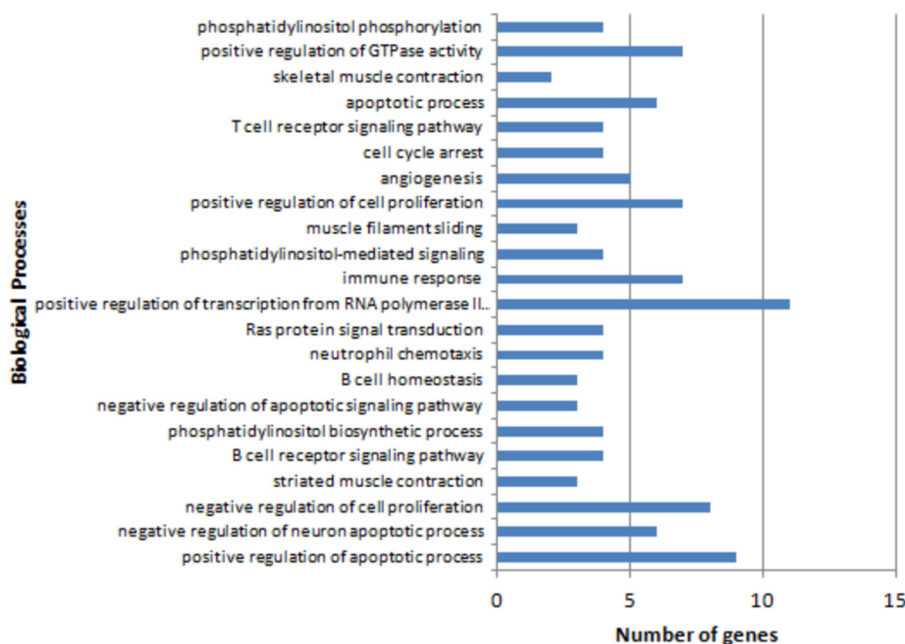


Fig. 7. Most significant biological processes (P -value<0.05) observed in the identified hub genes.

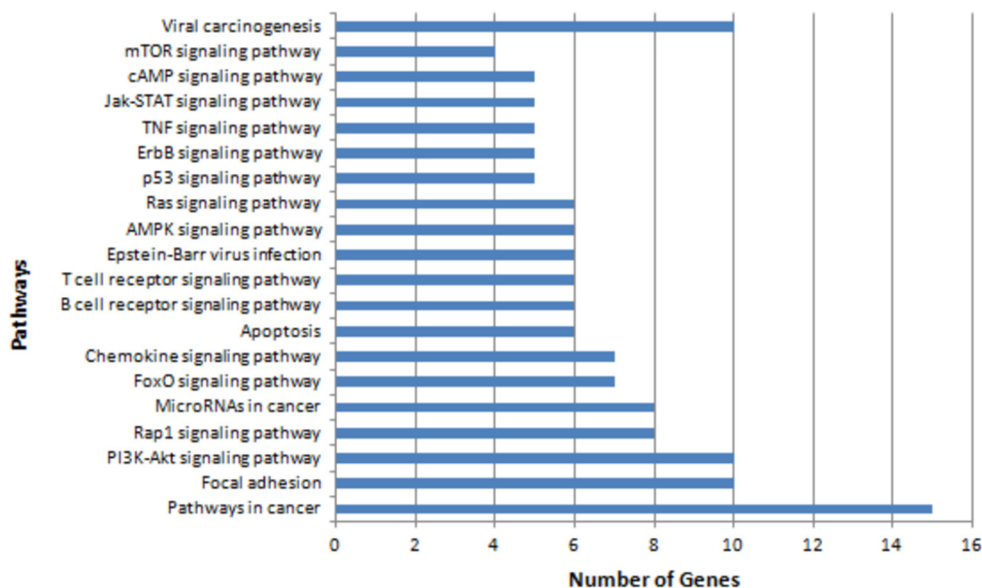


Fig. 8. Highly enriched list of pathways (P -value<0.05) observed in the identified hub genes.

transduction and many cancer mediated process. Pathways in cancer, p53 signaling pathway, small cell lung cancer are directly attributed to OC. p53 protein regulates cell cycle and functions and acts as a tumor suppressor. Disrupted regulation of p53 signaling elements deregulates numerous tumor suppressing process including DNA repair, cell cycle arrest and apoptosis Stegh (2012). p53 gene is one of the most frequently mutated genes in oral cancer Rowley et al. (1998). Isoforms of oncoprotein BCR-ABL activates tyrosine kinase and cause chronic myeloid leukemia. Deregulation of these isoforms causes altered cellular adhesion, activation of mitogenic signaling pathways, inhibition of apoptosis process Melo and Deininger (2004). Hub genes PIK3CG and PIK3R5 were found as markers in chronic myeloid leukemia and play an important role in cell growth, proliferation, differentiation, motility, survival and intracellular trafficking. Mutation of proto-oncogene PIK3CA is a key factor in cell proliferation and survival through activation of the PI3K/Akt signaling pathway. PIK3CA along with PTEN

plays a pivotal role in the PI3K-Akt signaling pathway and hence accepted as a biomarker for oral carcinogenesis Chang et al. (2013). Hub genes PIK3CG and PIK3R5 along with the seed gene PIK3CA were observed in many regulating pathways in OC. Overexpression of PIK3CG in tumor cells suppresses cell proliferation. Loss of function of this protein contributes to tumorigenesis process Semba et al. (2002). Focal adhesion kinase is a cytoplasmic tyrosine kinase identified as a key mediator of signaling by integrins, a major family of cell surface receptors and other receptors in both normal and tumor cells Zhao and Guan (2009). It is known to be an important mediator of cell growth, cell proliferation, cell survival and cell migration, all of which are often dysfunctional in cancer cells. The pathway enrichment analysis unveils that the hub genes marked by our approach are related to the development of cancer. The highly significant pathways observed in this analysis are demonstrated in Fig. 8. A detailed list of enriched pathways is given in Supplementary Table S7.

4. Discussion

Oral cancer ranked among the top three cancers and counts about thirty percent of prevalent cases in India. Early diagnosis and prevention of OC are prioritized at the global health forums Kulkarni (2013). Genomic studies show that hub genes in a genetic network plays an influential role in regulating the whole network structure. Thus, identification of such genes related to specific cancer types can help in reducing the gap in OC prognosis.

In this study, a biological network based integrative approach was proposed for the identification of hub genes related to OC. Data from both genomic and epigenomic levels were employed for the study. Gene co-expression network and gene co-methylation network were constructed using WGCNA. In the next step three integrated networks: Union Network, Integration Network and Multiplex network were generated from the constructed correlation networks. Multiplex network was constructed where isolated networks of gene co-expression and co-methylation encode for different layers in the multiplex system with coupling edges in between. Hence interplay among the interacting units across layers is preserved. Degree centrality (multi-degree centrality) and eigenvector centrality of genes were computed to find the connectivity of genes in the multiplex network. Based on the connectivity of genes in the integrated networks and the correlation networks, criteria 1 and criteria 2 were fixed for the identification of hub genes from the integrated network which results in six sets of hub genes. Among all, 38 hub genes were found to be common in all the hub gene sets.

To examine the performance of these genes as a predictor variable, the classifier accuracy of these gene sets was determined on the classifiers KNN, RF and SVM. The classifier performance indicates that all the hub gene sets attained good values for MCC. However, hub gene sets constructed from Multiplex network (C1-HubMN and C2-HubMN) achieved significantly high scores for MCC, Sensitivity and Specificity. This corroborates the ability of the Multiplex network in exploiting different types of interactions among biological entities. We further identified 46 genes as our putative hub genes, comprising of 38 common genes (found common in all six hub gene sets) and 8 OC genes (spotted in both the hubs C1-HubMN and C2-HubMN). To our surprise, the classifier performance of Cmn-Hub with 46 genes was significantly high with a prediction accuracy of 96% in comparison to other hub gene sets whose size varies between 40 - 1710. We also investigated the interconnection and functional association of our studied hub genes (46 number) with the 28 seed genes which are known to get associated with OC, by mapping them to the PPI network using stringApp plugin of Cytoscape. Multiplex centrality analysis considers the participation of a gene at multiple levels of the biological system. These genes are significantly influential both in transcriptomic and epigenetic level with eigenvector centrality >0.5. The differential expression analysis identified 28 hub genes showing a significant difference in expression levels between two experimental groups. Among them, genes CDC20, CCL4, LCP2, ASPM, FCGR3A were observed to be most significantly up-regulated and genes NDGR2, MUC21, PERM1, TNNC1, HSPB7 were significantly down-regulated. The DAVID biological enrichment analysis conveyed that the marked set of genes including PIK3CG, PIK3R5, CCL4, CDC20, TERT, LTA, MYH7 significantly takes part in various biological processes including immune response, cell proliferation, apoptosis, cell cycle arrest, signal transduction, muscle contraction, neutrophil chemotaxis, angiogenesis and pathways such as p53 signaling pathway, Chronic myeloid leukemia, Pathways in cancer, ErbB signaling pathway, PI3K-Akt signaling pathway along with the seed genes. More importantly, this approach focuses on identifying the hub genes that co-express at multiple levels including both genetic and epigenetic. The proposed multi-layered data integration approach provides the researchers with new insight for detailed exploration of tumorigenesis and progression in OC by systematically enhancing the study of gene action.

5. Conclusions

Diagnosis of OC at its early stage results in better therapeutic implications. The study reveals that genes encoded for hub proteins are enriched for disease genes and preferentially targeted by pathogens. Thus identification of such hub genes may give a further clue to the researchers for a thorough understanding of the disease pathogenesis.

This work aimed at identifying a set of hub genes crucial for OC using an integrated framework of multi-omics data. Three integrated networks: Union Network, Integration Network and Multiplex Network were constructed using expression level as well as methylation intensity of genes. A set of 46 genes were identified as hub genes which are validated in terms of their statistical competence and biological importance. We spotted five genes PIK3CG, PIK3R5, MYH7, CDC20 and CCL4, which were observed to be predominantly enriched in various biological processes and pathways crucial for OC. This work suggests an alternative research direction for biological network analysis. The identified genes in our approach can be a potential prognostic biomarker in the process of tumorigenesis and underlying molecular events in OC. Our findings enlighten the researchers for a more extensive investigation of OC.

Declarations

Author contribution statement

S. Mahapatra: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

R. Bhuyan, J. Das: Analyzed and interpreted the data.

T. Swarnkar: Conceived and designed the experiments; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data associated with this study can be found at The Cancer Genome Atlas (TCGA).

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2021.e07418>.

References

- Barabasi, A.-L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5 (2), 101.
- Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C.I., Gómez-Gardenes, J., Romance, M., Zanin, M., 2014. The structure and dynamics of multilayer networks. *Phys. Rep.* 544 (1), 1–122.
- Cai, M., Wang, W., Cui, Y., Stanley, H.E., 2018. Multiplex network analysis of employee performance and employee social relationships. *Physica A Stat. Mech. Appl.* 490, 1–12.
- Chakraborty, P., Karmakar, T., Arora, N., Mukherjee, G., 2018. Immune and genomic signatures in oral (head and neck) cancer. *Heliyon* 4 (10), e00880.
- Chang, K.-Y., Tsai, S.-Y., Chen, S.-H., Tsou, H.-H., Yen, C.-J., Liu, K.-J., et al., 2013. Dissecting the egfr-pi3k-akt pathway in oral cancer highlights the role of the egfr variant iii and its clinical relevance. *J. Biomed. Sci.* 20 (1), 43.
- Chen, L., Yang, J., Xing, Z., Yuan, F., Shu, Y., Zhang, Y., Cai, Y.-D., 2017. An integrated method for the identification of novel genes related to oral cancer. *PLoS ONE* 12 (4), e0175185.

- Chierici, M., Bussola, N., Marcolini, A., Francescato, M., Zandonà, A., Trastulla, L., Furlanello, C., 2020. Integrative network fusion: a multi-omics approach in molecular profiling. *Front. Oncol.* 10, 1065.
- Das, S., Meher, P.K., Rai, A., Bhar, L.M., Mandal, B.N., 2017. Statistical approaches for gene selection, hub gene identification and module interaction in gene co-expression network analysis: an application to aluminum stress in soybean (glycine max L). *PLoS ONE* 12 (1), e0169605.
- De Domenico, M., Porter, M.A., Arenas, A., 2015. Muxviz: a tool for multilayer analysis and visualization of networks. *J. Complex Netw.* 3 (2), 159–176.
- De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivela, M., Moreno, Y., Porter, M.A., Arenas, A., 2013. Mathematical formulation of multilayer networks. *Phys. Rev. X* 3 (4), 041022.
- Dembélé, D., Kastner, P., 2014. Fold change rank ordering statistics: a new method for detecting differentially expressed genes. *BMC Bioinform.* 15 (1), 1–15.
- Doncheva, N.T., Morris, J.H., Gorodkin, J., Jensen, L.J., 2018. Cytoscape stringapp: network analysis and visualization of proteomics data. *J. Proteome Res.* 18 (2), 623–632.
- Elling, A.A., Deng, X.W., 2009. Next-generation sequencing reveals complex relationships between the epigenome and transcriptome in maize. *Plant Signal. Behav.* 4 (8), 760–762.
- Finn, K.R., Silk, M.J., Porter, M.A., Pinter-Wollman, N., 2019. The use of multilayer network analysis in animal behaviour. *Anim. Behav.* 149, 7–22.
- Gadaleta, F., Bessonov, K., Van Steen, K., 2017. Integration of gene expression and methylation to unravel biological networks in glioblastoma patients. *Genet. Epidemiol.* 41 (2), 136–144.
- Granot, Z., Jablonska, J., 2015. Distinct functions of neutrophil in cancer and its regulation. *Mediat. Inflamm.* 2015.
- Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell* 144 (5), 646–674.
- He, X., Zhang, J., 2006. Why do hubs tend to be essential in protein networks? *PLoS Genet.* 2 (6), e88.
- Hosack, D.A., Dennis, G., Sherman, B.T., Lane, H.C., Lempicki, R.A., 2003. Identifying biological themes within lists of genes with ease. *Genome Biol.* 4 (10), 1–8.
- Jin, B., Li, Y., Robertson, K.D., 2011. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* 2 (6), 607–617.
- Joshi, P., Dutta, S., Chaturvedi, P., Nair, S., 2014. Head and neck cancers in developing countries. *Rambam Maim. Med. J.* 5 (2).
- Kanawati, R., 2015. Multiplex network mining: a brief survey. *IEEE Intel. Inf. Bull.* 16 (1), 24–27.
- Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A., 2014. Multilayer networks. *J. Complex Netw.* 2 (3), 203–271.
- Kulkarni, M.R., 2013. Head and neck cancer burden in India. *Int. J. Head Neck Surg.* 4 (1), 29–35.
- Kumar, R., Samal, S.K., Routray, S., Dash, R., Dixit, A., 2017. Identification of oral cancer related candidate genes by integrating protein-protein interactions, gene ontology, pathway analysis and immunohistochemistry. *Sci. Rep.* 7 (1), 2472.
- Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* 9 (1), 559.
- Langfelder, P., Mischel, P.S., Horvath, S., 2013. When is hub gene selection better than standard meta-analysis? *PLoS ONE* 8 (4), e61505.
- Langfelder, P., Zhang, B., Horvath, S., 2007. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinformatics* 24 (5), 719–720.
- Levine, A.E., Steffen, D.L., 2001. OrCGDB: a database of genes involved in oral cancer. *Nucleic Acids Res.* 29 (1), 300–302.
- Li, J., Yang, C., Ji-Jiang, Y., Chen, S., Wang, Q., Liang, S., et al., 2018. Introduce a new approach to detect genes associated to oral squamous cell carcinoma, *bioRxiv*, 377788.
- Liu, Z., Meng, J., Li, X., Zhu, F., Liu, T., Wu, G., Zhang, L., 2018. Identification of hub genes and key pathways associated with two subtypes of diffuse large B-cell lymphoma based on gene expression profiling via integrated bioinformatics. *BioMed Res. Int.* 2018.
- Ma, X., Liu, Z., Zhang, Z., Huang, X., Tang, W., 2017. Multiple network algorithm for epigenetic modules via the integration of genome-wide DNA methylation and gene expression data. *BMC Bioinform.* 18 (1), 1–13.
- Mahapatra, S., Mandal, B., Swarnkar, T., 2018. Biological networks integration based on dense module identification for gene prioritization from microarray data. *Gene Rep.* 12, 276–288.
- Marisa, L., de Reyniès, A., Duval, A., Selves, J., Gaub, M.P., Vescovo, L., et al., 2013. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 10 (5), e1001453.
- Melo, J.V., Deininger, M.W., 2004. Biology of chronic myelogenous leukemia—signaling pathways of initiation and transformation. *Hematol./Oncol. Clin.* 18 (3), 545568.
- Nanda, M.S., Shah, Z.A., Sameer, A.S., Syeed, N., Murtaza, I., Siddiqi, M.A., Ali, A., 2011. Tp53-molecular soldier's mutations in bladder cancer in the Kashmiri population. *Asian Pac. J. Cancer Prev.* 12 (1), 67–72.
- Nangraj, A.S., Selvaraj, G., Kaliampurthi, S., Kaushik, A.C., Cho, W.C., Wei, D.Q., 2020. Integrated PPI-and WGCNA-retrieval of hub gene signatures shared between Barrett's esophagus and esophageal adenocarcinoma. *Front. Pharmacol.* 11, 881.
- Negre, C.F., Morzan, U.N., Hendrickson, H.P., Pal, R., Lisi, G.P., Loria, J.P., Batista, V.S., 2018. Eigenvector centrality for characterization of protein allosteric pathways. *Proc. Natl. Acad. Sci.* 115 (52), E12201–E12208.
- Nishida, N., Yano, H., Nishida, T., Kamura, T., Kojiro, M., 2006. Angiogenesis in cancer. *Vasc. Health Risk Manag.* 2 (3), 213.
- Padi, M., Quackenbush, J., 2015. Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC Syst. Biol.* 9 (1), 80.
- Pan, W., 2002. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* 18 (4), 546–554.
- Polyak, K., Lee, M.-H., Erdjument-Bromage, H., Koff, A., Roberts, J.M., Tempst, P., Massagué, J., 1994. Cloning of p27kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals. *Cell* 78 (1), 59–66.
- Randhawa, V., Acharya, V., 2015. Integrated network analysis and logistic regression modeling identify stage-specific genes in oral squamous cell carcinoma. *BMC Med. Genom.* 8 (1), 1–20.
- Rowley, H., Sherrington, P., Helliwell, T.R., Kinsella, A., Jones, A.S., 1998. p53 expression and p53 gene mutation in oral cancer and dysplasia. *Otolaryngol. Head Neck Surg.* 118 (1), 115–123.
- Schiffman, J.D., Fisher, P.G., Gibbs, P., 2015. Early detection of cancer: past, present, and future. *Am. Soc. Clin. Oncol. Educ. Book* 35 (1), 57–65.
- Semba, S., Itoh, N., Ito, M., Youssef, E.M., Harada, M., Moriya, T., Yamakawa, M., 2002. Down-regulation of PIK3CG, a catalytic subunit of phosphatidylinositol 3-oh kinase, by CpG hypermethylation in human colorectal carcinoma. *Clin. Cancer Res.* 8 (12), 3824–3831.
- Shaikh, I., Ansari, A., Ayachit, G., Gandhi, M., Sharma, P., Bhairappanavar, S., Das, J., 2019. Differential gene expression analysis of HNSCC tumors deciphered tobacco dependent and independent molecular signatures. *Oncotarget* 10 (58), 6168.
- Shang, H., Liu, Z.-P., 2020. Network-based prioritization of cancer genes by integrative ranks from multi-omics data. *Comput. Biol. Med.* 119, 103692.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504.
- Sherman, B.T., Lempicki, R.A., et al., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4 (1), 44–57.
- Stegh, A.H., 2012. Targeting the p53 signaling pathway in cancer therapy—the promises, challenges and perils. *Expert Opin. Ther. Targets* 16 (1), 67–83.
- Swarnkar, T., Simoes, S.N., Anura, A., Brentani, H., Chatterjee, J., Hashimoto, R.F., Mitra, P., 2015. Identifying dense subgraphs in protein-protein interaction network for gene selection from microarray data. *Netw. Model. Anal. Health Inform. Bioinform.* 4 (1), 33.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., 2015. String v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. Crossref, PubMed.
- Vinayagam, A., Zirin, J., Roesel, C., Hu, Y., Yilmazel, B., Samsonova, A.A., Perrimon, N., 2014. Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nat. Methods* 11 (1), 94.
- Wajed, S.A., Laird, P.W., DeMeester, T.R., 2001. DNA methylation: an alternative pathway to cancer. *Ann. Surg.* 234 (1), 10.
- Wang, H., Zheng, H., Wang, J., Wang, C., Wu, F.-X., 2016. Integrating omics data with a multiplex network-based approach for the identification of cancer subtypes. *IEEE Trans. Nanobiosci.* 15 (4), 335–342.
- Wu, X., Jiang, R., Zhang, M.Q., Li, S., 2008. Network-based global inference of human disease genes. *Mol. Syst. Biol.* 4 (1).
- Zhao, B., Erwin, A., Xue, B., 2018. How many differentially expressed genes: a perspective from the comparison of genotypic and phenotypic distances. *Genomics* 110 (1), 67–73.
- Zhao, J., Guan, J.-L., 2009. Signal transduction by focal adhesion kinase in cancer. *Cancer Metastasis Rev.* 28 (1–2), 35–49.