RESEARCH ARTICLE

Exploring gene regulatory interaction networks and predicting therapeutic molecules for hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma

Abanti Bhattacharjya[1], Md Manowarul Islam[1], Md Ashraf Uddin[2], Md Alamin Talukder[3] (iD),
AKM Azad[4], Sunil Aryal[2], Bikash Kumar Paul[5,6], Wahia Tasnim[5] (iD),
Muhammad Ali Abdulllah Almoyad[7] and Mohammad Ali Moni[8,9,10] (iD)

1 Department of Computer Science and Engineering, Jagannath University, Dhaka, Bangladesh
2 School of Information Technology, Deakin University, Geelong, Australia
3 Department of Computer Science and Engineering, International University of Business Agriculture and Technology, Dhaka, Bangladesh
4 Department of Mathematics and Statistics, Faculty of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia
5 Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh
6 Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh
7 Department of Basic Medical Sciences, College of Applied Medical Sciences, King Khalid University, Abha, Saudi Arabia
8 Artificial Intelligence & Data Science, Faculty of Health and Behavioural Sciences, The University of Queensland, Brisbane, Australia
9 AI & Digital Health Technology, Artificial Intelligence and Cyber Futures Institute, Charles Sturt University, Bathurst, Australia
10 Rural Health Research Institute, Charles Sturt University, Orange, Australia

Hypopharyngeal cancer is a disease that is associated with EGFR-mutated lung adenocarcinoma. Here we utilized a bioinformatics approach to identify genetic commonalities between these two diseases. To this end, we examined microarray datasets from GEO (Gene Expression Omnibus) to identify differentially expressed genes, common genes, and hub genes between the selected two diseases. Our analyses identified potential therapeutic molecules for the selected diseases based on 10 hub genes with the highest interactions according to the degree topology method and the maximum clique centrality (MCC). These therapeutic molecules may have the potential for simultaneous treatment of these diseases.

Abbreviations

CSV, comma-separated values; DE, differential expression; DEGs, differentially expressed genes; FDR, false discovery rate; GEO, Gene Expression Omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MCC, maximum clique centrality; miRNAs, microRNAs; PPI, protein–protein interactions; SVG, scalable vector graphic; TF, transcription factor.

Bioinformatics, which combines the capabilities of computer science with biology, has expanded significantly in recent years [1]. Several bioinformatics toolkits are leveraged to achieve the desired result for the experiment. Bioinformatics can research the molecular causes of sickness, describe the disease's situation from the gene's nook, and reduce the amount of time and money spent on the process by utilizing computer abilities to narrow the scope of the investigation and improve the quality of the results [1]. Two hundred different cell types and 100 different cancers have been found among the 100 trillion cells in the human body [2].

The prognosis of tumors that originate from other head & neck sites is better than that of hypopharyngeal cancer, a less frequent type of head & neck cancer [3]. With only 15–30% of patients living for more than 5 years, hypopharyngeal carcinoma, which makes up around 5% of all head & neck malignancies, has a horrible prognosis [4,5]. Two common risk factors for hypopharyngeal carcinoma are alcohol consumption and smoking [6]. According to the American Cancer Society, the human papillomavirus also causes hypopharyngeal carcinoma.

The epidermal growth factor receptor gene is the most commonly mutated gene in lung cancer (*EGFR*) [7]. Lung squamous cell carcinoma (SCC), which has an estimated frequency of 3% to 18%, is comparatively uncommon compared to lung adenocarcinoma (1–10) [7]. Lung cancer, which comprises both small and non-small cells, is the leading cause of cancer-related death globally [8–11]. The world's highest incidence and fatality rates are associated with the most prevalent kinds of cancer [12]. Risk factors for lung cancer include smoking, passive smoking, age, gender, family history, chronic lung disease, chest radiotherapy, diet, obesity, physical activity, alcohol consumption, employment, education, and income [13]. The human papillomavirus might potentially increase the risk of developing lung cancer [14].

Head and neck cancers, as well as lung cancers, pose significant challenges to global health [15]. Head and neck cancer is among the most frequent malignancies to migrate to the lungs [16–18]. Following bone and soft-tissue sarcomas [19], recognized head–neck cancer is the third most frequent reason for pulmonary mastectomy. Ferlay *et al.* [20] reported in 2012 that there were 686,000 new instances of head and neck cancer, 1,825,000 new cases of lung cancer, and a combined mortality rate of 5% and 19%, respectively. Hypopharyngeal cancer is a type of head and neck cancer and *EGFR*-mutated lung adenocarcinoma is also one form of lung cancer. Thus, we can claim that patients

with EGFR-mutated lung adenocarcinoma may have the potential to develop hypopharyngeal cancer, according to the preceding statistic. Also, hypopharyngeal cancer may potentially spread to lung adenocarcinoma with EGFR mutation, and lung adenocarcinoma with EGFR mutation may potentially metastasize to hypopharyngeal cancer [21]. This article's researchers analyzed the general population's lung metastases in newly diagnosed hypopharyngeal cancer. According to the Canadian Cancer Society, lung cancer may develop if hypopharyngeal cancer progresses [22]. Therefore, this concludes that they are related genetically because they share genes. This set of shared genes is restrained by regulatory interaction network pathways.

In this research, we aimed to look into common DEGs (differentially expressed genes), hub genes, various gene regulatory networks, and therapeutic molecule for hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma using bioinformatics technology.

We used two datasets for EGFR-mutated lung adenocarcinoma and hypopharyngeal cancer. Each of these datasets have eight samples. The DEGs shared by these two datasets were extracted using the R programming language (Vienna, Austria). These widely distributed DEGs help to identify GO terms, pathways, PPI networks, and TF-miRNA. Based on the hub genes of patients with hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma who have these two diseases concurrently, certain therapeutic compounds are envisaged. Hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma can be associated with each other directly or indirectly. These diseases have some common interrelated genes. Gene regulatory interaction networks are developed by using different types of bioinformatics tools. The PPI network is visualized, and common drugs are developed for the selected two associated diseases. A PPI network describes the connections between proteins in a biological system in the context of biological study. The process of visualizing this network usually entails producing a map or graphical depiction that shows the connections and interactions between various proteins. By illustrating these relationships graphically, scientists can better understand the intricate biological mechanisms involving proteins, and possibly pinpoint important hubs or nodes in the network that are essential to cellular operations or disease processes. Visualizing the PPI network is, all things considered, a step towards better understanding the complexities of molecular interactions through the analysis of biological data. Developing common drugs indicates the creation of common medications for the selected diseases. In addition, designing one common drug for two associated

diseases decreases the amount of drug one should absorb for the diseases separately.

There are some earlier works based on different diseases using the various bioinformatics approaches as follows [23]: Molecular biomarker identification to suggest therapeutic targets for the creation of medicines to treat esophageal cancer. The authors collected these GSE93756, GSE94012, GSE104958, and GSE143822 datasets from the National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus. Using the R Language Limma Package, DEGs were collected by applying the adjusted value < 0.05. After that, DEGs' GO and pathway enrichment analysis was done. And PPI and clustering analysis were also done based on the DEGs. Taz *et al.* [24] discussed idiopathic pulmonary fibrosis (IPF) people who have SARS-CoV-2 infections who are genetically more likely to develop IPF. The GEO NCBI database was used to collect the GSE147507 and GSE35145 datasets. DEGs for GSE35145 were retrieved using the GEO2R tool that is included by default with the dataset in GEO, while DEGs for GSE147507 were gathered using the R programming language. Adjusted P-value < 0.05 and log2-fold change (absolute) > 1.0 were used as the cutoff criteria. Common genes for these two diseases were used for gene set enrichment analysis, PPIs network construction, hub gene searching and module examination, TF-miRNA identification, and candidate drug suggestion [25]. Identification of the SARS-CoV-2 biomarkers and pathways that complicate the condition in patients with pulmonary arterial hypertension. The R programming language's *limma* and DESeq2 packages were used to gather DEGs of GSE147507 for SARS-CoV-2 infection in human lung epithelial cells. DEGs of GSE117261 for PAH lung were found through the GEO2R tool of the GEO NCBI database. Adjusted P-value < 0.05 and log2-fold change (absolute) > 1.0 were used as the cutoff criteria. For the objective of controlling the false discovery rate, the Benjamini–Hochberg approach was used on both datasets. Common DEGs from these two datasets were used for further analysis.

Chen *et al.* [26] employed bioinformatics analysis to screen for and identify potential target genes in head and neck cancer. The authors used the dataset GSE58911 from GEO. An interactive web tool called GEO2R, which is by default available on GEO, was used to identify the DEGs. The cutoff criteria were an adjusted P < 0.05 and a log fold-change (FC) ≥ 1 or ≤ -1. KEGG and GO enrichment analysis was performed using the DEGs. Additionally, a PPI network was constructed, and hub genes were identified using the DEGs [27]. Bioinformatics analysis was used

to identify relevant HNSCC (head and neck squamous cell carcinoma) genes from public databases. In this research work, DEGs were deemed significant if their logFC ≥ 1 or ≤ -1 and an adjusted P < 0.05. After retrieving the DEGs, several analyses were conducted, including GO, KEGG, PPI, DEG survival analyses, verification of key genes via Oncomine, specimens, and real-time PCR. Jin and Yang [28], using integrated bioinformatics techniques, used the identification and analysis of genes linked to head and neck squamous cell carcinoma. GSE13601, GSE31056, and GSE30784 datasets from GEO were downloaded and analyzed using the R language. To identify the DEGs from three different datasets, P < 0.01 and $|\log(FC)| > 1$ were chosen as the cutoff. Common genes of the three different datasets were also identified by using the Venn Diagram package in R language. Further analysis (analysis of KEGG pathways and gene ontologies, top modules and hub genes in a PPI network identification, validation of hub gene relative mRNA expression levels, examining the hub genes' protein levels in the human protein atlas database, hub gene survival analysis using the TCGA database, RNA extraction and real-time quantitative PCR, and analysis of statistics) was done based on the common DEGs [29]. By using integrated bioinformatics analysis, hub genes associated with the development of head and neck squamous cell carcinoma were found. From TCGA and GEO, the gene expression profiles for HNSCC were retrieved. Using WGCNA, key coexpression modules were identified, and DEGs were defined as genes with the cutoff criteria of $|\log FC| \geq 1.0$ and adj. P < 0.05. Functional analysis of interest genes, PPI construction, and hub gene screening, validation of hub gene expression patterns and prognostic values, and validation of survival-related hub gene protein expressions by the HPA database were also conducted in this study [30]. Using TCGA and GEO datasets, a study was conducted to identify potential biomarkers and analyze survival data for head and neck squamous cell carcinoma. Using R, the GSE6631 dataset for head and neck squamous cell carcinoma was analyzed. Here, adj. P < 0.05 was applied to differential gene screening in order to control the number of false-positive results. The heat and volcano maps were also constructed for the corresponding DEGs. Enrichment analysis, PPI analysis, hub genes survival analysis, key genes verifications, analysis of immunohistochemical, and finding potential small molecules were also identified in this study.

Tu *et al.* [31] used bioinformatics to investigate lung adenocarcinoma prognostic biomarkers. There are various datasets such as GSE31210, GSE32665,

GSE32863, GSE43458, and GSE72094 datasets for lung adenocarcinoma that were used in that study. $|\log_2 FC| \geq 1.5$ and $P < 0.05$ were the cutoff criteria to retrieve the DEGs from these datasets. Enrichment analysis, finding and verifying the prognostic gene signature, interactive analysis of gene expression, analysis of the prognostic model's independence, and nomogram construction were all conducted in an article [32]. The use of bioinformatics to identify important biomarkers in patients with lung adenocarcinoma. The GSE10072 dataset from the GEO database was used in that article. The adjusted $P < 0.05$ and $|\log_2 FC| \geq 1$ were the cutoff criteria to retrieve DEGs. All of these steps, such as the analysis of KEGG pathways and gene ontologies, the top five upregulated and top five downregulated comparison, The top five downregulated and top five upregulated stages of overall survival (OS), analysis of the PPI network and modules were done here [33]. Microarray data analysis using bioinformatics to find potential lung adenocarcinoma biomarkers. The GEO database was used to download the datasets (GSE118370, GSE32863, GSE85841, and GSE43458) for lung adenocarcinoma. DEGs were defined with $|\log_2 FoldChange| \geq 1$ and FDR < 0.05. Analysis of GO and KEGG enrichment, analyzing modules and building a ppi network, and analyses of hub genes were examined here [34]. Using bioinformatics analysis, elevated mRNA levels of the genes *AURKA*, *CDC20*, and *TPX2* are linked to a poor prognosis for lung adenocarcinoma caused by smoking. GSE31210, GSE32863, GSE40791, GSE43458, and GSE75037 datasets from the GEO database were analyzed. Using the cutoff criteria of $P < 0.05$ and absolute fold change > 1.5, the DEGs were retrieved. The functional enrichment analysis was done for 58 DEGs. After that, the validation of data and statistical analysis steps were performed in this research work.

Microarray data exploration is among the most well-known techniques used for extensive investigations of gene expression, and high-throughput technologies are becoming more and more important in the field of biomedical research [35]. Researchers in genetics can analyze gene expression simultaneously with the help of microarray studies [36]. This research attempts to discover the relationship between the selected two diseases. GSE212398 for hypopharyngeal cancer and GSE198672 for EGFR-mutated lung adenocarcinoma datasets were used for the investigation. The NCBI's GEO database served as the source for both dataset selections. Shared DEGs are collected from those two datasets. Figure 1 presents the proposed methodology's flow diagram.

This work aimed to identify targeted therapeutic molecules for these two diseases. Targeted therapy has very often a remarkable effect against cancer. Drug compounds can serve multiple purposes in cancer treatment, including reducing the size of tumors before surgery, removing any remaining cancer cells after surgery, or, as a last resort, when other treatments are ineffective or cancer recurs.

Our contributions are summarized as follows:

- To propose a bioinformatics framework for integratively analyzing expression profiles of lung adenocarcinoma and hypopharyngeal cancer samples to find commonly found biomarkers.
- To conduct detailed downstream analyses based on commonly found biomarkers.
- Finally, propose some therapeutic agents for those biomarkers via drug-target analyses.

Materials and methods

In this section we present the methodology of our experiments. We introduced a process of designing gene superintendent interaction networks, including PPI networks, interactions between TFs and genes, Network regulating gene-miRNA interactions, network of the gene–diseases for hypopharyngeal cancer EGFR-mutated lung adenocarcinoma, and suggested common drug compounds for these two associated diseases.

The steps in the proposed methodology are described below.

Dataset selection

The NCBI (National Center for Biotechnology Information) is an online platform from which we can collect many forms of biological data in a variety of formats; these data are also accessible in a variety of computer-readable formats. Datasets used in this research were gathered from the NCBI platform's GEO (Gene Expression Omnibus) database. The GEO database for high-throughput gene expression analysis can be accessed through the National Center for Biotechnology Information platform [37]. RELA is dependent on CD271 expression and stem-like features in hypopharyngeal cancer, according to the dataset (GSE212398). The dataset (GSE198672) contains EGFR-mutated lung adenocarcinomas that develop from preexisting tumor cells and persist in a specialized stromal milieu as drug-tolerant persisters after erlotinib treatment. The RNA sequence from GSE198672 and GSE212398 was extracted using the GPL10558 (Illumina [San Diego, CA, USA] HumanHT-12 V4.0 expression bead chip) and GPL20844 (Agilent-072363 [Santa Clara, CA, USA] SurePrint G3 Human GE v3 8x60K Microarray 039494 [Feature Number Version]) platforms,
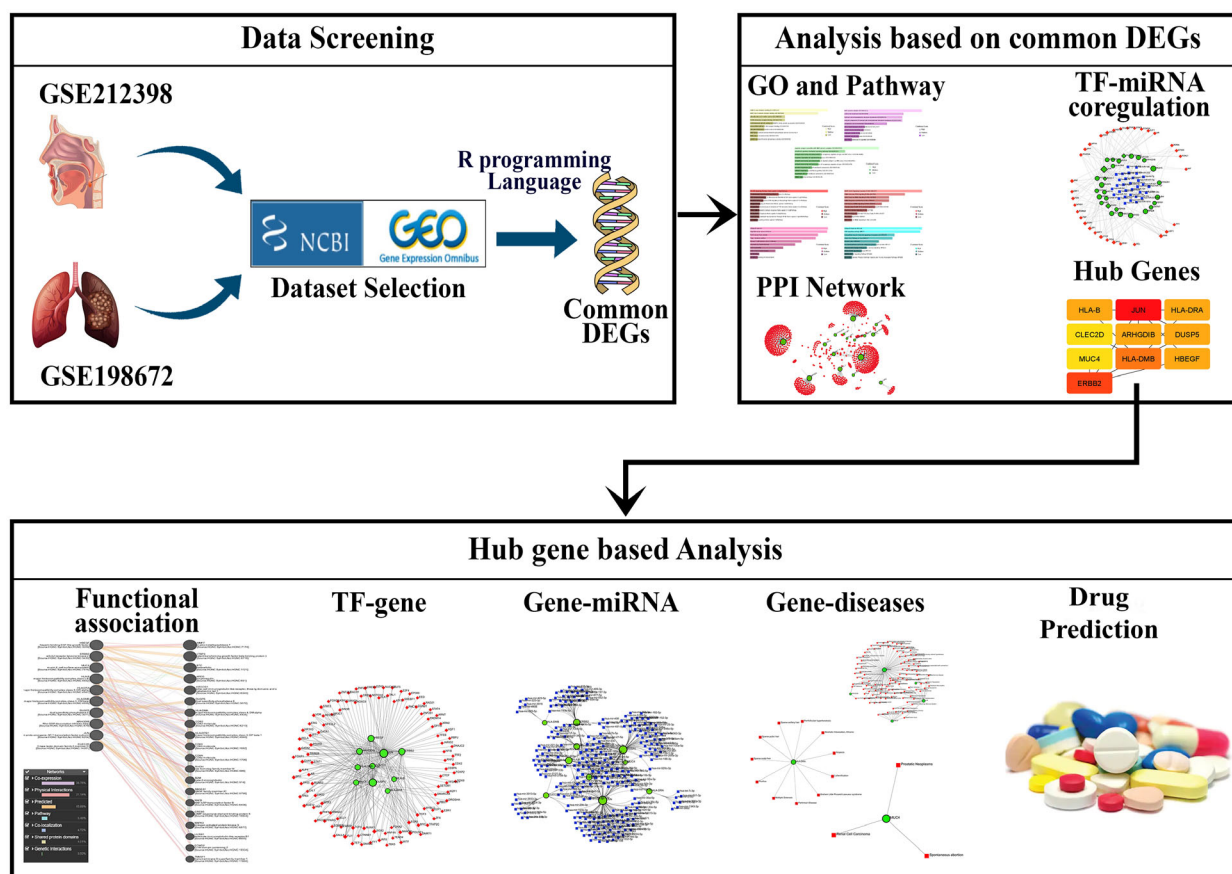
Fig. 1. Diagram representing the proposed methodology of the current research. For hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma, two datasets are used. Each dataset has eight samples. Using the R programming language, the DEGs (differentially expressed genes) from those two datasets are retrieved. The VENNY tool is used to find out the common genes between these two diseases. With the aid of these widespread DEGs, GO terms, pathways, PPI networks, TF-miRNA, and hub genes are identified. Functional association, TF-gene, gene-miRNA, gene–disease, and some therapeutic compounds are anticipated based on the hub genes of individuals with hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma who have these two diseases concurrently.

respectively. The GSE212398 dataset is a subseries of the GSE212399 dataset. For our investigation, we chose the GSE212398 dataset because this dataset contains eight samples, including four samples for control and four samples for KO. The GSE198672 dataset also has eight samples.

Differential expression analysis

Finding DEGs from these microarray datasets is the initial step for this particular research. The *GEOquery* [38] R package was used for retrieving gene expression datasets for both diseases from the NCBI GEO [39] database. Next, the *limma* [40] R package with empirical Bayes statistics was used for differential expression (DE) analysis. The DE output was formatted as a comma-separated values (CSV) file containing information, including Gene Symbol, logFC, *P*.value, and adjusted *P*.Val for the corresponding disease dataset and collected. Both datasets FDRs were controlled

using the Benjamini–Hochberg [41] method. An adjusted *P* < 0.05 and log2-fold change (absolute) > 1 are the cutoff criteria for obtaining DEGs for both datasets. Using the Venny tool [42], the shared DEGs between these two diseases were visualized.

Enrichment analysis of shared DEGs

Enrichment of gene sets is the study of gene sets with connected chromosomal locations, molecular activities, and biological functions [43]. Gene Ontology (GO), which is divided into the three categories of biological process, molecular function, and cellular component, is used for gene product annotation [44]. Understanding the molecular activity, cellular function, and the location in a cell where the genes perform their functions serves as the main foundation for choosing GO keywords [24]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, which

has a significant advantage over gene annotation, is frequently used to study metabolic pathways [45]. For extensive route analysis, databases from Reactome [46], BioCarta [47], and WikiPathways [48] were used in addition to the KEGG pathway.

TF-miRNA coregulatory network

To determine which transcription factors (TFs) bind with shared DEGs in the regulatory regions, target gene relationships between TFs and TFs were examined [49]. MiRNAs that attempt to bind on a gene transcript to negatively affect protein expression have been identified using miRNAs target gene interactions [50]. The RegNetwork repository [51] provided interactions for TF-miRNA coregulatory interactions, which make it easier to identify regulatory TFs and miRNAs that regulate DEGs of interest during the transcriptional and post-transcriptional phases [24]. Utilizing the NetworkAnalyst platform, we constructed the TF-miRNA coregulatory network [52]. Researchers can browse complex datasets with the help of NetworkAnalyst to find biological traits and functionalities that can be used to generate useful biological hypotheses [53]. The minimum network option was selected among the different available formats to construct the TF-miRNA coregulatory network.

PPI network

PPI activity is thought to be the main area of interest in cellular biology research and is necessary for system biology [54]. With the aid of cutting-edge research on PPI networks, the number of complex biological processes is identified [55,56]. Proteins operate inside of cells through interactions with other proteins, and information produced by a PPI network aids in our understanding of how proteins function [57]. Given to the STRING [58] database (https://string-db.org/), shared DEGs of hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma are used to create a PPI network and discover the genes that are directly associated among the common genes. Some basic settings are set on the STRING to get the desired result, such as setting the network type as a full STRING network, selecting the meaning of network edges by evidence, also selecting active interaction sources by text mining, experiments, databases, coexpression, neighborhood, gene fusion, and co-occurrence. Interactive SVG (network is a scalable vector graphic [SVG]; interactive) is selected for network display mode in advanced settings. The information provided by STRING is based on expected and experimental interactions, and the interactions generated by the web tool are characterized by 3D structures, supplementary data, and evidence scores [59]. After constructing this PPI network from STRING, this STRING PPI network was further reconstructed in CYTOSCAPE to identify only the

interconnected genes and remove the disconnected genes among those 32 shared genes. With the help of the web-based NETWORKANALYST [52] software (https://www.networkanalyst.ca/), identified directly interconnected genes (from CYTOSCAPE PPIs network) were entered into InnateDB [60] to design additional PPIs for interconnected genes. Here, auto layout was selected in the layout option to build this network.

Retrieving hub genes

Hub nodes are referred to be the highly connected nodes in a large-scale PPI network [61]. The cytoHubba plugin for the CYTOSCAPE program is used to locate hub nodes. The user-friendly cytoHubba interface makes it the most popular hub identification plugin for CYTOSCAPE, and it comes with 11 topological analysis methods [62]. Among the 11 topological methods on CytoHubba, the degree method and the maximum clique centrality (MCC) were chosen to identify the hub genes. In the degree topology method, the degree is counted according to the number of interactions among the genes. Higher interacted genes from the given input genes are easily identified. The gene has the highest number of degree scores ranking as the top among all genes. The most important candidate genes among the shared DEGs, which may be crucial in physiological regulatory functions, were found using the maximal clique centrality (MCC), which demonstrated better accuracy in predicting critical proteins in the PPI network.

The MCC technique was found to be the most efficient way to locate hub nodes in a PPI network [63]. Also, the authors of the article [64] mentioned that the most efficient technique for identifying hub nodes was thought to be the MCC algorithm. So, these two methods (the degree topology method, one of the most popular topological methods [65] and the MCC, the most efficient method among the available 11 methods [62]) that were chosen to identify hub genes out of the 11 available.

Functional association network

A bioinformatics program called GeneMANIA displays functional association information, genetic relationships, pathways, and coexpression for a given set of input data [66]. Gene sets' functions can be predicted with the aid of GeneMANia [67]. The percentage of coexpression, physical interactions, predicted, pathway, colocalization, shared protein domains, and genetic interactions for the given input genes are easily identified through the functional association network. Physical interactions between two or more proteins can produce binary interactions and complex proteins [68]. Genes are associated in gene coexpression networks, which are transcription factor-transcription factor association networks that are typically presented as undirected graphs [69]. Unlike most coexpression networks, which are undirected

graphs, this network showed a close relationship between two genes [70]. In studies on protein–protein interactions, two genes are connected if they are found to interact. These ligand-based protein networks, which foresee the ability of nearby proteins to bind connected substances indirectly, may be used to enhance genetically orientated gene networks, which foretell the significance of a procedure or a disease [71]. Using a data analysis technique called gene coexpression analysis, it is possible to find groupings of genes that have comparable expression patterns under various conditions [72]. The link between the genes' functions is referred to as genetic interaction [73]. The top 10 hub genes were used to demonstrate a functional network from Gene-Mania [74].

TF-gene interactions

By analyzing the TF-gene interaction using the discovered 10 hub genes, one may determine the impact of TF on functional pathways and gene expression levels [75]. Users can do a meta-analysis and analyze gene expression for numerous species with the use of NETWORKANALYST [52]. The control of gene transcription as well as the establishment of cellular identity and activity are assumed to depend on transcriptional factors (TFs), the TF (transferrin) gene-producing proteins [76]. The TF-gene interaction investigates how TF affects functional pathways and levels of gene expression [77]. Finding the important TF-gene interactions is crucial for comprehending the roles of pleiotropic global regulators [78]. Through direct or indirect interactions with other TFs, specific TFs help regulate the expression of a variety of target genes [76]. To control life activities, several transcription factors interact [79]. The 10 hub genes are utilized to evaluate the impact of TF on the functional pathways and expression levels of the genes through TF-gene interaction analysis. To find TF-gene interactions with well-known genes, researchers use the NETWORKANALYST platform [52]. NETWORKANALYST includes activities that are typical of network topologies and can be used to analyze biological modules [80]. The NETWORKANALYST platform's ChEA [81] database inspired the network built for the TF-gene interaction network [24].

Gene-miRNA interactions

By analyzing the Gene-miRNA interaction using the discovered 10 hub genes, one may determine the impact of TF on functional pathways and gene expression levels by base-pairing with their target mRNAs, microRNAs, a class of brief, noncoding RNA molecules with a length of 21–25 nucleotides, regulate the expression of genes, primarily by silencing or downregulating the target genes [82]. Natural single-stranded tiny RNA molecules known as microRNAs control the expression of genes by attaching to certain mRNAs and either starting the translation of the target

mRNA or starting the destruction of the target mRNA [83]. Small noncoding RNAs known as microRNAs (miRNAs) were discovered to promote mRNA degradation or prevent post-transcriptional translation [83]. Evidence is mounting that miRNAs have a role in carcinogenesis and cancer metastasis [84]. More and more varieties and uses for small noncoding RNAs are being discovered. This implies that there may be regulatory mechanisms that are far more complex than those now employed in the analysis and creation of gene regulatory networks. Finding new therapeutic targets can benefit from the analysis of inter-pathway regulatory factors. Since noncoding miRNAs are important for activating pathways, their activity is crucial in this regulatory environment. In the control of transcriptome processes, microRNAs are crucial [85]. For many biological processes in both plants and animals, post-transcriptional mediators of gene expression such as microRNAs are crucial [86]. To fully comprehend the miRNA's biological function, it is crucial to pinpoint the genes that it regulates [73]. MiRNAs can be retrieved by using the TarBase database. TarBase is a comprehensive repository of animal microRNA targets supported by experimental data. The database is also functionally connected to several other helpful resources, including GO and the UCSC Genome Browser. TarBase provides a rich dataset from which to evaluate characteristics of miRNA targeting that will be helpful for the upcoming generation of target prediction tools. TarBase reveals substantially more empirically supported targets than even recent evaluations indicate [87]. The network of gene-miRNA interactions is created using the web-based tool TarBase under NetworkAnalyst for those 10 hub genes (*JUN, ERBB2, HLA-DMB, HBEGF, HLA-B, HLA-DRA, DUSP5, ARHGDIB, MUC4, CLEC2D*).

Gene–disease interactions

Gene–disease interactions network focuses mostly on the most recent understanding of human genetic illnesses, including complex, Mendelian, and ecological diseases [88]. Gene–disease interactions network helps to identify those diseases that can occur due to the input genes. This network helps us to identify the risk factors that should be cured by therapeutic molecules. Gene–disease interactions network focuses primarily on the most recent knowledge of complex and ecological diseases, as well as other human genetic ailments [89]. DisGeNET is a sizable database of gene–disease interactions that combines information from several sources and covers a range of biological traits linked to diseases [88]. The hub genes were linked to related diseases and their chronic states by the network analysis of gene–disease correlations. DisGeNET [88] is a large database of gene–disease interactions that incorporates links from several sources and covers a variety of biological features associated with illnesses [49]. The investigation of

gene–disease correlations using NetworkAnalyst identified associated diseases and their chronic conditions with the hub genes (*JUN, ERBB2, HLA-DMB, HBEGF, HLA-B, HLA-DRA, DUSP5, ARHGDIB, MUC4, CLEC2D*) [53].

Therapeutic molecule suggestion for selected diseases

Therapeutic molecule suggestion is the pointer step of this research. DSigDB is used for drug suggestion. Users can access the DSigDB database via the Enrichr platform (https://amp.pharm.mssm.edu/Enrichr/) [24]. Enrichr is mostly used as an enrichment analysis tool, which offers substantial graphical data on the combined functions of the input genes [90]. There are 19,531 genes, 22,527 gene sets, and 17,389 unique chemicals in DSigDB [91]. A new gene set resource called Drug Signatures Database (DSigDB) connects medicines and compounds with their target genes. To forecast drugs, DSigDB largely employs gene expression-based datasets, and each group of genes is seen as being targeted when taking a molecule into account [91].

Results

Differential expression analysis identifies common DEGs between hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma

We found 605 identical DEGs for hypopharyngeal cancer (GSE212398) and 1062 identical DEGs for EGFR-mutated lung adenocarcinoma (GSE198672) by using the R programming language. Among those identical DEGS, 32 common genes were identified between hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma through the Venny tool. The Venn diagram of shared DEGs between the two diseases is shown in Fig. 2.

Enrichment of functional pathways and gene ontology terms

The analysis of gene set enrichment was performed using the online tool Enrichr [24]. Many databases, including the GO Consortium [92], Reactome [46], KEGG [93], WikiPathways [94], and BioCarta [47] were used by Taz *et al.* [25] to find GO keywords and cell-informing pathways. The GO database was used to find the biological process, molecular function, and cellular components. Analysis of biological process, molecular function, and cellular component data revealed notable involvement in peptide antigen assembly with the MHC protein complex, ErbB-3 class receptor binding, and MHC protein complex in shared DEGs, respectively. MAPK family signaling cascades,
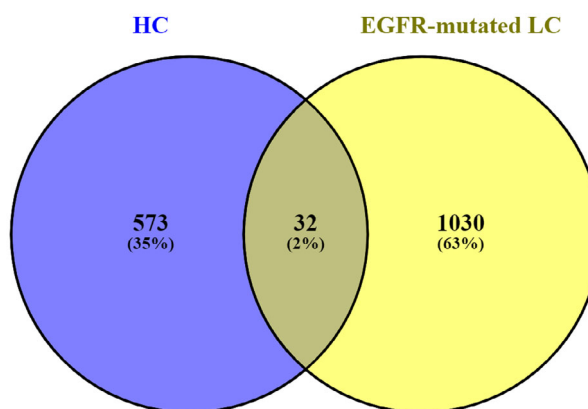

Fig. 2. Venn diagram of shared DEGs. Thirty-two common genes were found between HC and EGFR-mutated LC. Common DEGs were 2% among 1667 DEGs.

allograft rejection, allograft rejection, and the D4-GDI signaling pathway were highly enriched among all identified when Reactome, KEGG, WikiPathways, and BioCarta databases were used, respectively. Table 1 shows the top five biological terms, cellular terms, and molecular terms and Table 2 shows the top five pathways from Reactome, KEGG, WikiPathways, and BioCarta with correspondent *P*-value and genes. Top 10 GO terms concomitant to biological process, molecular function, and cellular component pinpointing entrenched on the combined score is in Fig. 3, and also based on a combined score, the top 10 pathways from Reactome, KEGG, WikiPathways, and BioCarta are mentioned in Fig. 4. To get the combination score, multiply the z score, which represents the deviation from the predicted rank, by the log of the *P*-value from Fisher's exact test. The 'combined scores' for Figs. 3 and 4 are automatically calculated in the Enrichr platform. In the biological process, peptide antigen assembly with MHC protein complex (GO:0002501) indicates 'Peptide attachment to an MHC protein complex's antigen-binding groove [95].' The interferon-gamma-mediated signaling pathway (GO:0060333) means the cascade of molecular signals that begins when interferon-gamma binds to its receptor on a target cell's surface and ends with the control of a cell's transcription, among other downstream cellular processes. The only type II interferon so far discovered is interferon-gamma. The antigen processing and presentation of exogenous peptide antigen via MHC class II (GO:0019886) is the process by which an MHC class II protein complex collaborates with an antigen-presenting cell to express a peptide antigen of external origin on the cell surface. Typically, but not always, a complete protein is used to digest the peptide

Table 1. Biological terms, cellular terms, and molecular terms with correspondent *P*-values and genes.

Type	Term	*P*-value	Genes
GO Biological Process	Peptide antigen assembly with MHC protein complex (GO:0002501)	3.71E-05	*HLA-DMB;HLA-DRA*
	Interferon-gamma-mediated signaling pathway (GO:0060333)	1.74E-04	*HLA-DRB4;HLA-B;HLA-DRA*
	Antigen processing and presentation of exogenous peptide antigen via MHC class II (GO:0019886)	5.10E-04	*HLA-DMB; HLA-DRB4; HLA-DRA*
	Negative regulation of reproductive process (GO:2000242)	5.11E-04	*ARHGDIB;NKX3-1*
	Antigen processing and presentation of peptide antigen via MHC class II (GO:0002495)	5.41E-04	*HLA-DMB; HLA-DRB4; HLA-DRA*
GO Cellular Component	MHC protein complex (GO:0042611)	4.16E-06	*HLA-DMB;HLA-B;HLA-DRA*
	Endosome membrane (GO:0010008)	1.12E-05	*HLA-DMB;HLA-DRB4;ERBB2;HLA-B; HLA-DRA;RHOD*
	Luminal side of endoplasmic reticulum membrane (GO:0098553)	1.19E-05	*HLA-DRB4;HLA-B;HLA-DRA*
	Integral component of luminal side of endoplasmic reticulum membrane (GO:0071556)	1.19E-05	*HLA-DRB4;HLA-B;HLA-DRA*
	Cytoplasmic vesicle membrane (GO:0030659)	2.70E-05	*HLA-DRB4;ERBB2;HLA-B;HLA-DRA; RHOD;HBEGF*
GO Molecular Function	ErbB-3 class receptor binding (GO:0043125)	0.007975	*ERBB2*
	MHC class II protein complex binding (GO:0023026)	3.32E-04	*HLA-DMB;HLA-DRA*
	Phosphatidic acid transfer activity (GO:1990050)	0.011148	*PITPNC1*
	CCR6 chemokine receptor binding (GO:0031731)	0.011148	*DEFB1*
	Oxidoreductase activity, acting on NAD(P)H, heme protein as acceptor (GO:0016653)	0.012731	*CYB5R2*

Table 2. Five pathways from Reactome, KEGG, WikiPathways, and BioCarta and with correspondent *P*-values and genes.

Database	Pathway	*P*-value	Genes
Reactome	MAPK family signaling cascades R-HSA-5683057	1.39E-04	*DUSP5;JUN;ERBB2;PEA15;HBEGF*
	ERBB2 activates PTK6 Signaling R-HSA-8847993	1.91E-04	*ERBB2;HBEGF*
	GRB2 events in ERBB2 signaling R-HSA-1963640	2.57E-04	*ERBB2;HBEGF*
	ERBB2 regulates cell motility R-HSA-6785631	2.57E-04	*ERBB2;HBEGF*
	PI3K events in ERBB2 Signaling R-HSA-1963642	2.93E-04	*ERBB2;HBEGF*
KEGG	Allograft rejection	3.83E-07	*HLA-DMB;HLA-DRB4;HLA-B;HLA-DRA*
	Staphylococcus aureus infection	3.96E-07	*HLA-DMB;HLA-DRB4;KRT16; HLA-DRA;DEFB1*
	Graft-versus-host disease	5.79E-07	*HLA-DMB;HLA-DRB4;HLA-B;HLA-DRA*
	Type I diabetes mellitus	6.37E-07	*HLA-DRB4;HLA-B;HLA-DRA*
	Human T-cell leukemia virus 1 infection	1.15E-06	*JUN;EGR2;HLA-DMB;HLA-DRB4; HLA-B;HLA-DRA*
WikiPathway	Allograft rejection WP2328	3.85E-04	*HLA-DMB;HLA-B;HLA-DRA*
	ErbB signaling pathway WP673	4.11E-04	*JUN;ERBB2;HBEGF*
	Extracellular vesicle-mediated signaling in recipient cells WP2870	0.001049058	*TSPAN8;ERBB2*
	Ebola virus pathway on host WP4217	0.001134032	*HLA-DMB;HLA-B;HLA-DRA*
	Bladder cancer WP2828	0.001862402	*ERBB2;HBEGF*
BioCarta	D4-GDI signaling pathway *Homo sapiens* h d4gdiPathway	3.71E-05	*JUN;ARHGDIB*
	T-cell receptor signaling pathway *Homo sapiens* h tcrPathway	0.003493127	*JUN;HLA-DRA*
	TSP-1 induced apoptosis in microvascular endothelial cell *Homo sapiens* h tsp1Pathway	0.011147938	*JUN*
	Pertussis toxin-insensitive CCR5 signaling in macrophage *Homo sapiens* h Ccr5Pathway	0.014310901	*JUN*
	antigen processing and presentation *Homo sapiens* h mhcPathway	0.01903698	*HLA-DRA*

antigen. The negative regulation of reproductive processes (GO:2000242) indicates any procedure that slows down, prevents, or lessens the number of times, how often, or how much the reproductive process occurs. The antigen processing and presentation of peptide antigen via MHC class II (GO:0002495) is the process by which an MHC class II protein complex collaborates with an antigen-presenting cell to express
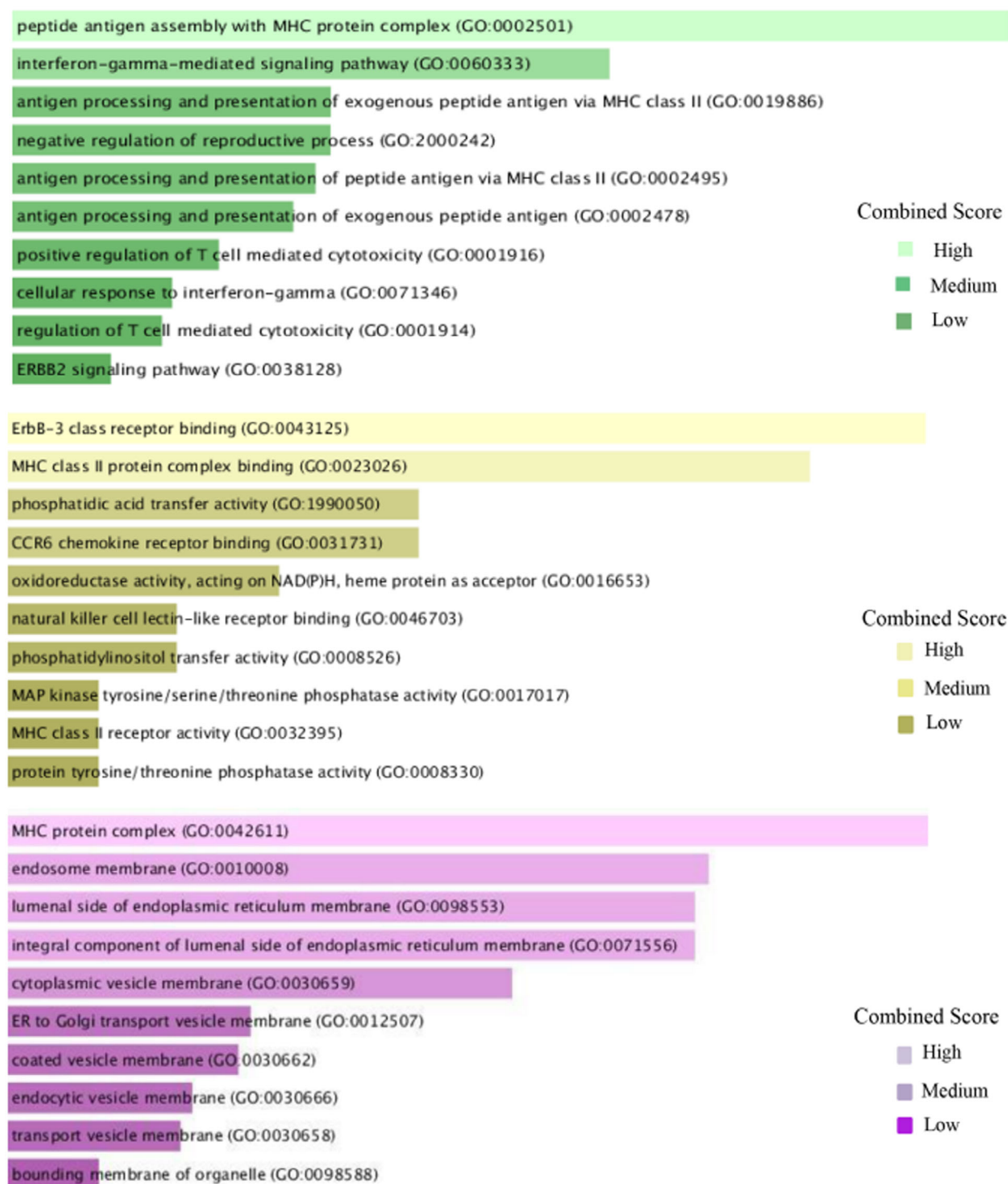
peptide antigen assembly with MHC protein complex (GO:0002501)

interferon–gamma–mediated signaling pathway (GO:0060333)

antigen processing and presentation of exogenous peptide antigen via MHC class II (GO:0019886)

negative regulation of reproductive process (GO:2000242)

antigen processing and presentation of peptide antigen via MHC class II (GO:0002495)

antigen processing and presentation of exogenous peptide antigen (GO:0002478)

positive regulation of T cell mediated cytotoxicity (GO:0001916)

cellular response to interferon–gamma (GO:0071346)

regulation of T cell mediated cytotoxicity (GO:0001914)

ERBB2 signaling pathway (GO:0038128)

Combined Score
High
Medium
Low

ErbB-3 class receptor binding (GO:0043125)

MHC class II protein complex binding (GO:0023026)

phosphatidic acid transfer activity (GO:1990050)

CCR6 chemokine receptor binding (GO:0031731)

oxidoreductase activity, acting on NAD(P)H, heme protein as acceptor (GO:0016653)

natural killer cell lectin–like receptor binding (GO:0046703)

phosphatidylinositol transfer activity (GO:0008526)

MAP kinase tyrosine/serine/threonine phosphatase activity (GO:0017017)

MHC class II receptor activity (GO:0032395)

protein tyrosine/threonine phosphatase activity (GO:0008330)

Combined Score
High
Medium
Low

MHC protein complex (GO:0042611)

endosome membrane (GO:0010008)

lumenal side of endoplasmic reticulum membrane (GO:0098553)

integral component of lumenal side of endoplasmic reticulum membrane (GO:0071556)

cytoplasmic vesicle membrane (GO:0030659)

ER to Golgi transport vesicle membrane (GO:0012507)

coated vesicle membrane (GO:0030662)

endocytic vesicle membrane (GO:0030666)

transport vesicle membrane (GO:0030658)

bounding membrane of organelle (GO:0098588)

Combined Score
High
Medium
Low

Fig. 3. Top 10 GO terms concomitant to biological process, molecular function, and cellular component pinpointing entrenched on the combined score (the log of the *P*-value from Fisher's exact test and multiplying that by the z-score of the deviation from the expected rank).

a peptide antigen on the cell surface. Usually, but not always, the protein in its whole serves as the source of the peptide antigen. In cellular components, the MHC protein complex (GO:0042611) is an MHC class II beta chain or an invariant beta2-microglobulin chain, along with or without a bound peptide, lipid, or
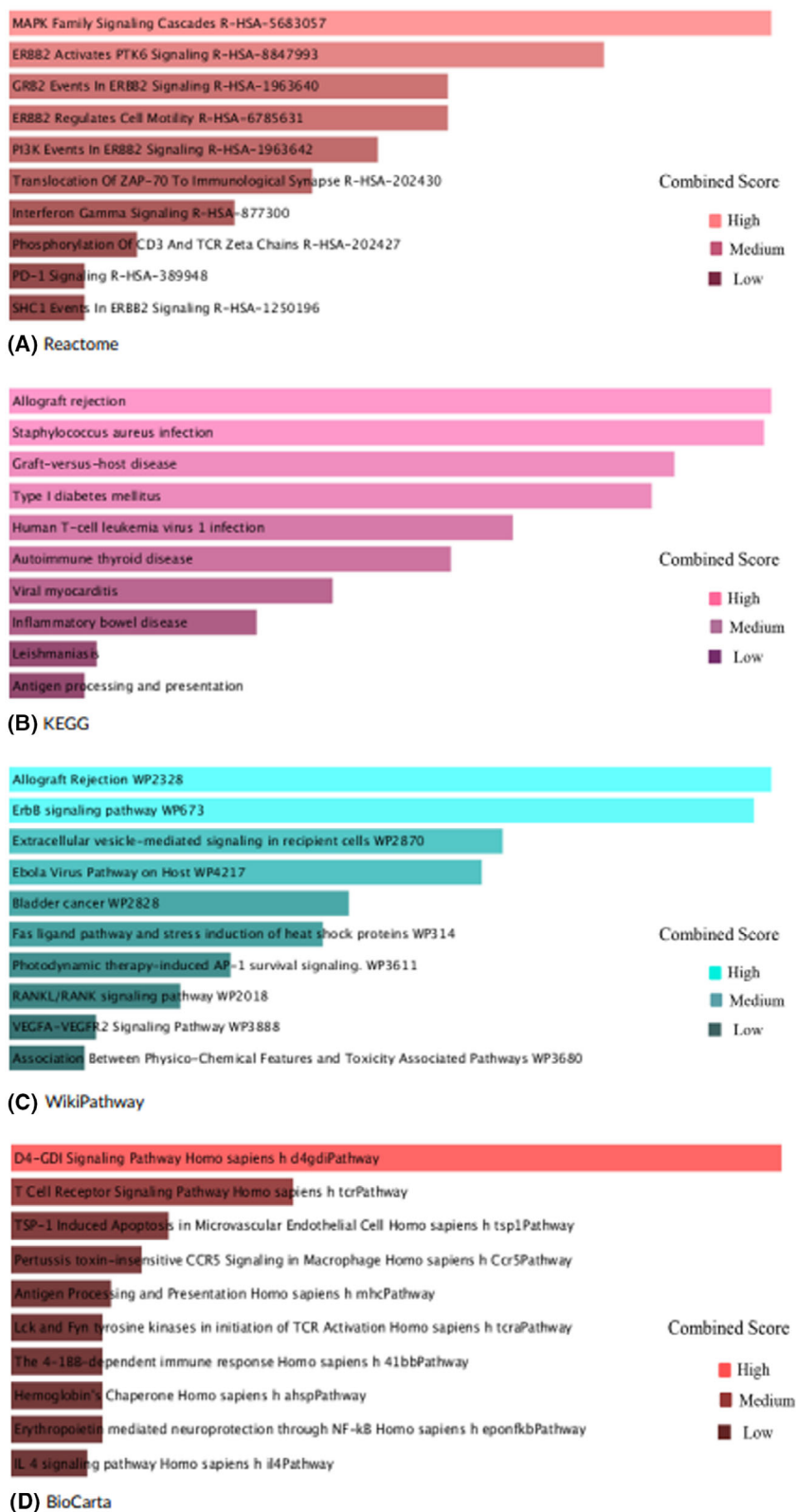
Fig. 4. Top 10 pathways from (A) Reactome, (B) KEGG, (C) WikiPathways, and (D) BioCarta pinpointing entrenched on the combined score (the log of the *P*-value from Fisher's exact test and multiplying that by the z-score of the deviation from the expected rank).

polysaccharide antigen, makes up a transmembrane protein complex. The endosome membrane (GO:0010008) indicates a lipid bilayer that envelops an endosome. The luminal side of the endoplasmic reticulum membrane (GO:0098553) indicates the leaflet-shaped side of the plasma membrane that is facing the lumen. The integral component of the luminal side of the endoplasmic reticulum membrane (GO:0071556) indicates a portion of the endoplasmic reticulum membrane made up of gene products that can only pass through the membrane's luminal side. The cytoplasmic vesicle membrane (GO:0030659) is a cytoplasmic vesicle's protective lipid bilayer. In molecular function, ErbB-3 class receptor binding (GO:0043125) indicates ErbB-3/HER3 protein-tyrosine kinase receptor binding. The MHC class II protein complex binding (GO:0023026) is the main histocompatibility complex of class II. The phosphatidic acid transfer activity (GO:1990050) means phosphatidic acid is taken out of a membrane or a monolayer lipid particle, transported through the aqueous phase while being sheltered in a hydrophobic pocket, and then brought to a membrane or lipid particle that will accept it. Phosphatidic acid is a type of glycophospholipid that typically has a phosphate group attached to carbon-3, an unsaturated fatty acid attached to carbon-2, and a saturated fatty acid attached to carbon-1. The CCR6 chemokine receptor binding (GO:0031731) is chemokine CCR6 receptor binding. The oxidoreductase activity, acting on NAD(P)H, heme protein as acceptor (GO:0016653) indicates an oxidation–reduction (redox) reaction that uses NADH or NADPH as a hydrogen or electron donor to reduce a heme protein is catalyzed.

TF-miRNA coregulatory network construction

To comprehend how TF and miRNA regulate with shared DEGs, a TF-miRNA coregulatory network was developed. Common DEGs (*GRAMD3, SYNGR1, CMTM7, HLA-DMB, HLA-DRB4, CLEC2D, CCNB1IP1, HLA-DRA, DEFB1, ERBB2, MUC4, LOC145837, RPS24, RHOD, HBEGF, ARHGDIB, PITPNC1, PEA15, KRT16, GJB3, JUN, FHL2, CYB5R2, HLA-B, EGR2, HERC6, DUSP5, HBA2, CHST15, NKX3-1, LBH, TSPAN8*) were utilized to create the network of TF-miRNA coregulators. Thirty-two common genes were given as the 'Gene Input List' in NETWORKANALYST. Then 'H.sapiens (human)' and 'Official Gene Symbol' were chosen for 'Specify organism' and 'Set ID type' attributes correspondingly. After uploading this information, TF-miRNA coregulatory interactions were selected from the gene regulatory interaction. The literature-curated regulatory interaction information was collected from the RegNetwork repository. After selecting the minimum network from the network tools option, the TF-miRNA coregulatory network was constructed. Background as white and layout as circular bi/tripartite were chosen to better visualize the network. Also, opacity, thickness, color, label, and size were customized from edge and node options. Red color for TF, green-black highlighted for seeds, and blue for miRNA were chosen from The Global Node Styles. The network shown in Fig. 5 comes with 93 nodes, 223 edges, and 28 seeds.

PPI network

The network of protein–protein interactions was built using the STRING. Interconnected genes and disconnected genes are easily differentiated from this in Fig. 6. This network was further evoked to cytoscape for better visualization. This PPI network, shown in Fig. 7, contains only 17 connected genes. Another PPI network was contrived by IMEx Interactome of NetworkAnalyst using the corresponding connected genes to understand the infection state by those corresponding genes. This network is shown in Fig. 8, contains 972 nodes, 1110 edges, and 16 seeds. These 16 seeds are those 17 interconnected genes except 'CLEC2D.' These 16 seeds have a higher degree of interaction with the protein. As '*CLEC2D*' had no significant interactions in the network, so, the IMEx interactome automatically removed 'CLEC2D' as the network seed.

Pinpointing hub genes

For this research, the top 10 hub genes were taken, because these top 10 hub genes are considered to be the most responsible genes among all. If these 10 hub genes are cured by a therapeutic molecule, then all other affected genes may have the possibility to recover as well, as these hub genes have interaction with other genes. Ten hub genes had been pinpointed from the reconstructed PPI network, as shown in Fig. 7 of Cytoscape using the degree topology and the MCC method. Tables 3 and 4 show the top 10 hub genes according to the degree topology method and the MCC method, respectively. The JUN has the highest interaction among the retrieved 10 hub genes. The same hub genes were retrieved using two different methods, the degree topology method and the MCC method. *JUN* and CLEC2D have the highest and lowest scores, respectively, in both methods. Seeing that from the TF-gene interactions network and the gene-miRNA network, the hub genes with score 1 have contributed to those networks. miRNAs are discovered to promote mRNA degradation or prevent post-
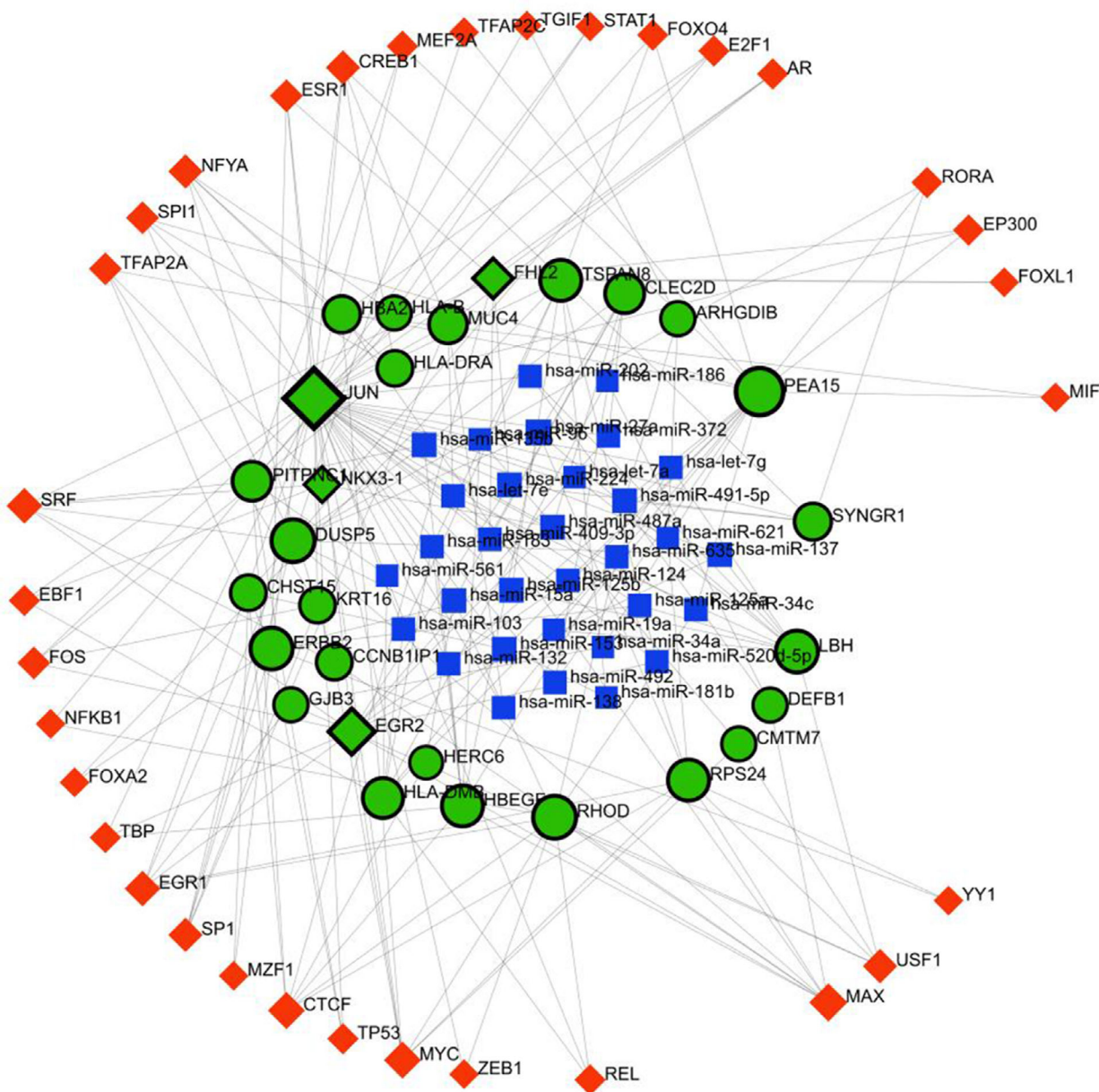
Fig. 5. Visualization of TF-miRNA coregulatory network through NetworkAnalyst. Green-black highlighted nodes indicate seeds, red diamond-shaped nodes for TF, and blue box-shaped nodes for miRNA.

transcriptional translation, understanding the functions of pleiotropic global regulators requires identifying the significant TF-gene interactions. After that, two networks were drawn through the CYTOSCAPE. The grid layout was chosen to extract Figs. 9 and 10.

Functional association network

The 10 hub genes were used in the GeneMania Functional Network. Figure 11 helped us to predict how

certain gene sets will behave. Utilizing a massive collection of functional association data, GeneMANIA discovers additional genes that are connected to a set of input genes. Protein and genetic relationships, pathways, coexpression, colocalization, and protein domain similarity are all examples of association data. Gene-MANIA can be used to discover new components of a pathway or complex, discover extra genes that may have escaped existing screens, or discover novel genes that have a particular function, such as protein
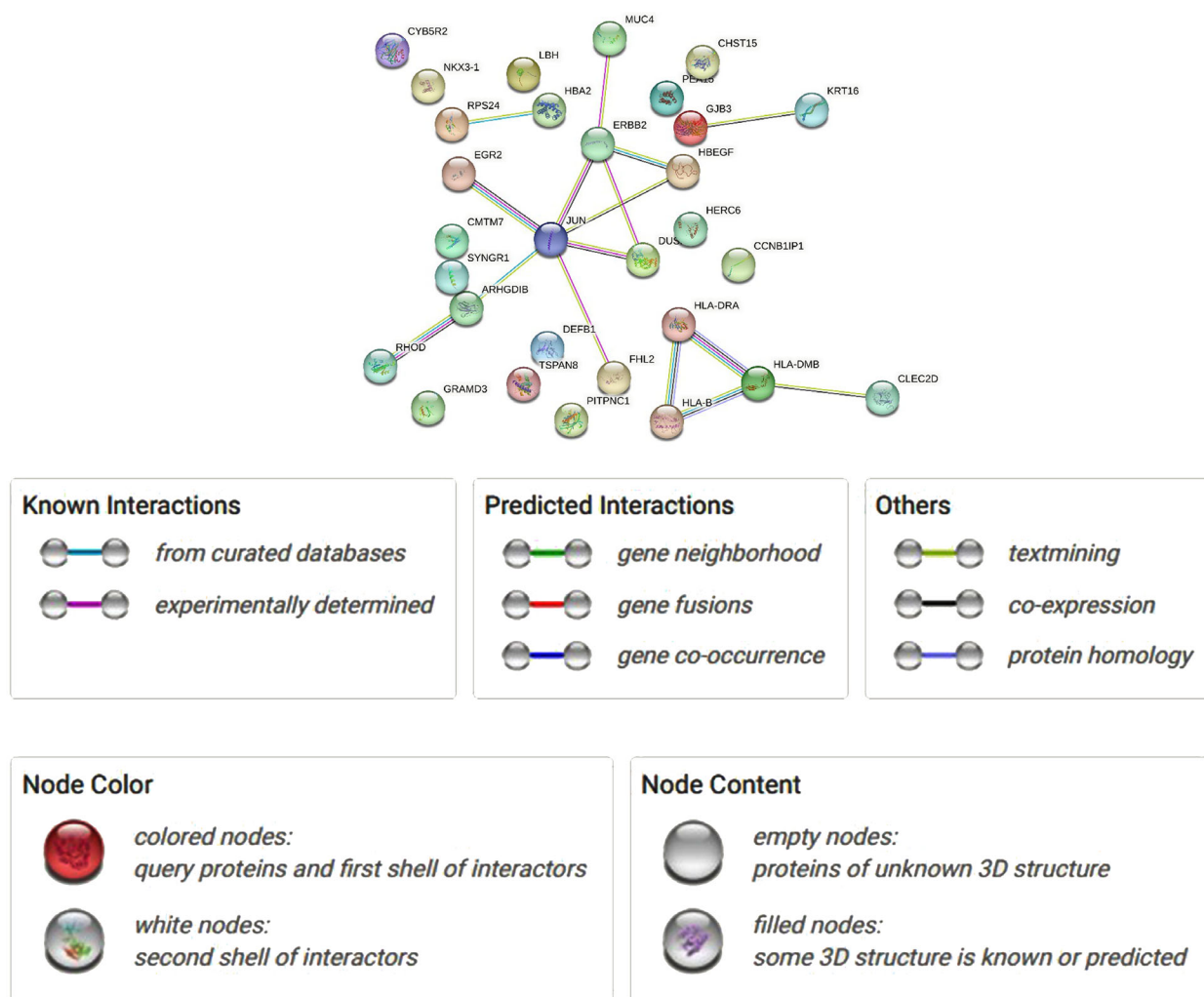
Fig. 6. Protein–protein interaction network through string.

kinases. If the input gene list has five or more genes, GeneMANIA uses the 'assigned based on query gene' technique to assign weights to enhance connectivity between all of the given input genes. To maximize the interaction between genes on a given list and minimize the interaction with genes not on a given list, the weights are automatically selected using linear regression. As our input gene list had more than five genes (10 hub genes), this default method was done here. This network has 30 nodes and displays functional keywords such as shared protein domains, coexpression, physical interactions, predicted, pathways, and genetic interactions [96] and also shows the percentage of the functional keywords for our interested gene set. In all, 36.76% coexpression, 31.14% physical interactions, 15.88% predicted, 6.48% pathway, 4.72% colocalization, 4.51% shared protein domains and 0.50%

genetic interactions were found in the functional association network. A higher level of coexpression (36.76%) between the transcripts associated with the two selected diseases and a consistent 31.14% physical interaction. This implies a more pronounced connection at the genetic and molecular levels between the diseases. The acknowledgment of these strong associations is then used to support the idea that it is relatively straightforward to customize or modify a generic medication to effectively treat both distinct illnesses. The thesis here is that a single medicine can target a shared biological basis, shared by the two diseases because of the significant coexpression and physical interaction between the genes and proteins linked to them. It might be easier to create a drug that treats both illnesses at once because of their similar foundation.
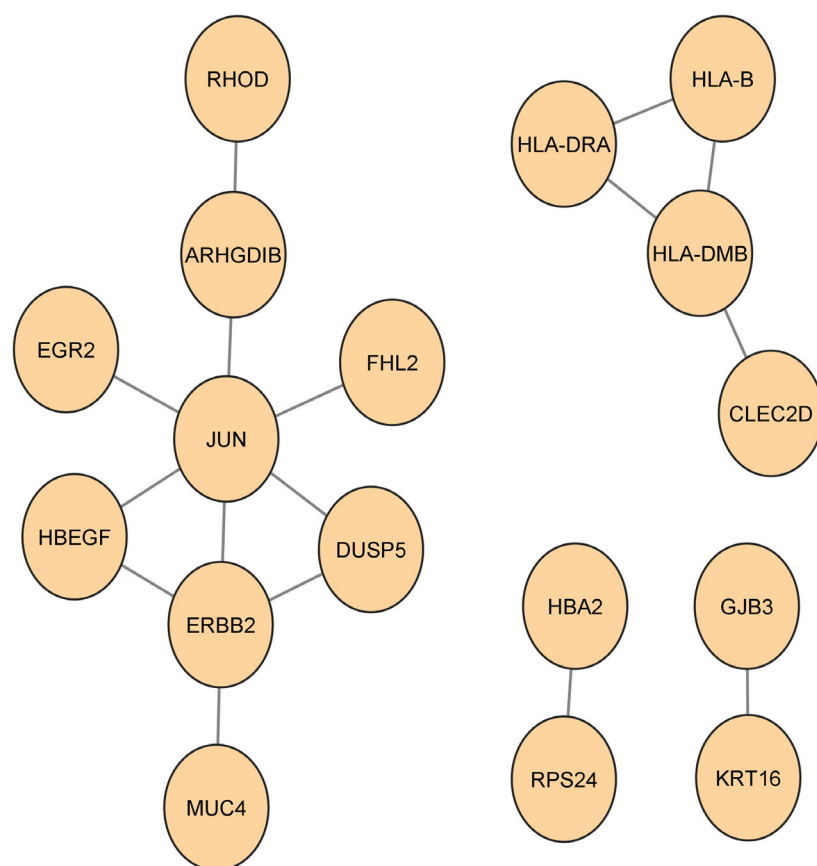
Fig. 7. PPIs network through CYTOSCAPE using the directly interconnected genes.

Association of the TF-gene

The TF-gene association network was fabricated to understand the transcription factor produced by the interested gene set. It was contrived using NetworkAnalyst. Hub genes (JUN, ERBB2, HLA-DMB, HBEGF, HLA-B, HLA-DRA, DUSP5, ARHGDIB, MUC4, CLEC2D) were used to build a TF-gene interaction network. Ten hub genes were given as the 'Gene Input List' in NetworkAnalyst. Then 'H.sapiens (human)' and 'Official Gene Symbol' were chosen for 'Specify organism' and 'Set ID type' attributes correspondingly. By uploading that information, TF-gene interactions were selected from the gene regulatory interaction. From the available three options (ENCODE, JASPAR, and ChEA) ChEA was selected to draw this network. The ChEA database is a transcription factor that targets a database inferred from integrating literature curated Chip-X data. After proceeding further, the network is constructed via the ChEA database. Background as white and layout as circular bi/tripartite were chosen to better visualize the network. Also, opacity, thickness, color, label, and size were customized from edge and node options. Red color for TF-gene and green with the

black highlight for seeds were selected from the global node styles for better visualization of this network. There are 119 nodes, 226 edges, and 10 seeds in the TF-gene network, as shown in Fig. 12. In our constructed TF-gene interaction, JUN is regulated by 60 TFs, the 47 TFs that control DUSP5, 24 TFs for *HBEGF*, and 24 TFs for *ERBB2*, and *HLA-DMB, HLA-B, HLA-DRA, ARHGDIB, MUC4, CLEC2D* are regulated by 13 TFs, 7 TFs, 8 TFs, 19 TFs, 12 TFs, and 13 TFs respectively.

Gene-miRNA interactions

Using NetworkAnalyst, the same hub genes (JUN, ERBB2, HLA-DMB, HBEGF, HLA-B, HLA-DRA, DUSP5, ARHGDIB, MUC4, CLEC2D) were used as input to display the gene-miRNA interaction network. The network as shown in Fig. 13 comes with 271 nodes, 401 edges, and 10 seeds. The gene-miRNA interaction network shows miRNAs originated from those 10 hub genes. JUN is regulated by 118 miRNAs. The 96 miRNAs that control DUSP5, 43 miRNAs, and 47 miRNAs, respectively, control *ERBB2* and *HLA-B*.
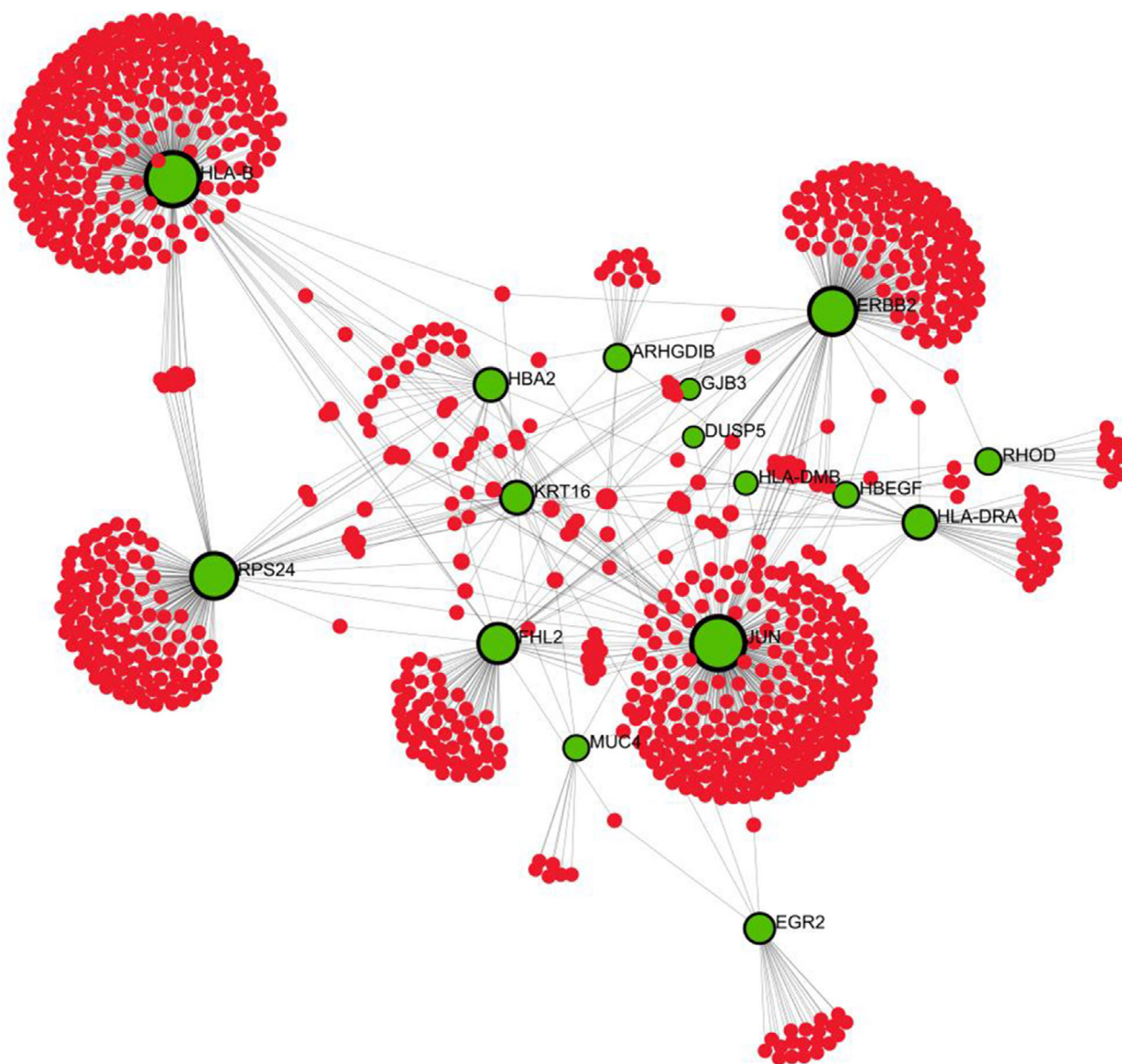
Fig. 8. PPIs network obtained through IMeX intercome of InnateDB database contains 972 nodes, 1110 edges, and 16 seeds.

Interactions of gene–disease

DisGeNET database under NetworkAnalyst was used to demonstrate the gene–disease association network. Hub genes (JUN, ERBB2, HLA-DMB, HBEGF, HLA-B, HLA-DRA, DUSP5, ARHGDIB, MUC4, CLEC2D) were also used to construct a gene–disease association network. DisGeNET database under NetworkAnalyst was used here to build the gene–disease correlations network. Ten hub genes were given as 'Gene Input List' in NETWORKANALYST. Then 'H.sapiens (human)' and 'Official Gene Symbol' were chosen for 'Specify organism' and 'Set ID type' attributes
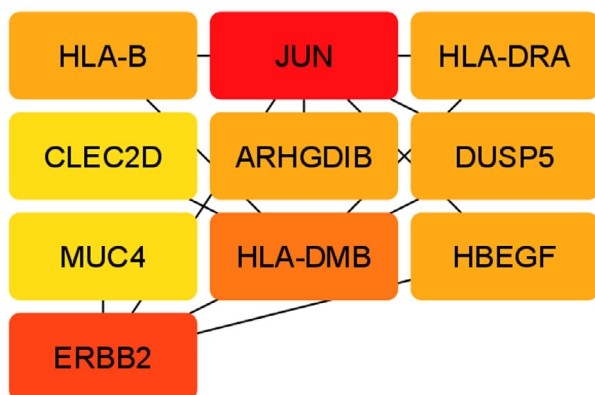
correspondingly. By uploading that information, gene–disease associations were selected from diseases, drugs & chemicals. A gene–disease association database has information about gene–disease associations that have been curated by the literature gathered from the Dis-GeNET database that only applies to human data. After proceeding further, the network is constructed via the DisGeNET database. The network contains three subnetworks. Each subnetwork contains the genes with their associated disease. This network is divided into three subnetworks because each subnetwork has no common associated diseases and contains

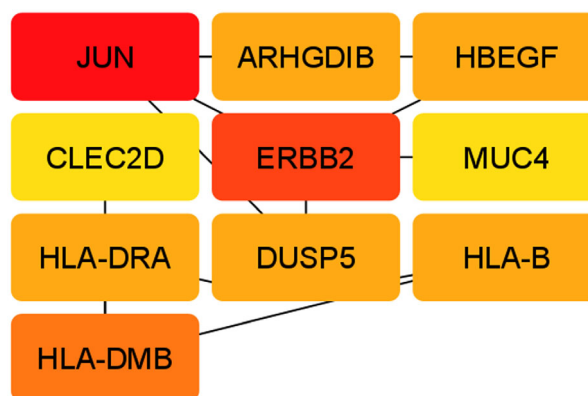**Table 3.** The 10 hub genes, ordered by degree of importance.

Name	Score
JUN	6
ERBB2	4
HLA-DMB	3
HBEGF	2
HLA-B	2
HLA-DRA	2
DUSP5	2
ARHGDIB	2
MUC4	1
CLEC2D	1

Table 4. The 10 hub genes, ordered by MCC (maximal clique centrality) method.

Name	Score
JUN	7
ERBB2	5
HLA-DMB	3
HBEGF	2
HLA-B	2
HLA-DRA	2
DUSP5	2
ARHGDIB	2
MUC4	1
CLEC2D	1

Fig. 9. The top 10 hub genes network according to the degree topology method through CYTOSCAPE.

Fig. 10. The top 10 hub genes network accordance with the maximal clique centrality (MCC) method through CYTOSCAPE.

Common drug suggestion

Those 10 hub genes (JUN, ERBB2, HLA-DMB, HBEGF, HLA-B, HLA-DRA, DUSP5, ARHGDIB, MUC4, CLEC2D) by the degree topology method and the MCC were used to recommend common drugs for the selected two diseases. These 10 genes are the most responsible genes of all. From the gene–disease association network, it is clear that the 10 hub genes are also responsible for some new diseases. That means some new diseases will occur due to those affected genes. Thus, these most affected genes must be cured by therapeutic molecules. Ten hub genes were used to recommend common drugs for the selected two diseases. These 10 hub genes are considered to be the most responsible among all genes. Thus, these 10 hub genes must be cured by any molecules. When any molecule is used to cure these hub genes, it also affects other connected genes with those hub genes. Ten hub genes were given as input in the Enrichr platform. From the available options, the DiseasesDrugs option was chosen. Our suggested drugs were retrieved from the DSigDB database under the DiseaseDrugs option in the Enrichr platform. Table 5 shows some predicated drug compounds for hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma patients who have these two diseases simultaneously. Here, eight well-known therapeutic molecules were suggested from the DSigDB database. We suggest these eight common drugs based on the number of hub genes cured.

Discussion

A dangerous constituent of hypopharyngeal cancer is lung adenocarcinoma with EGFR mutation, which are reported to have some associations with each other. Therefore, we hypothesized that this association is due
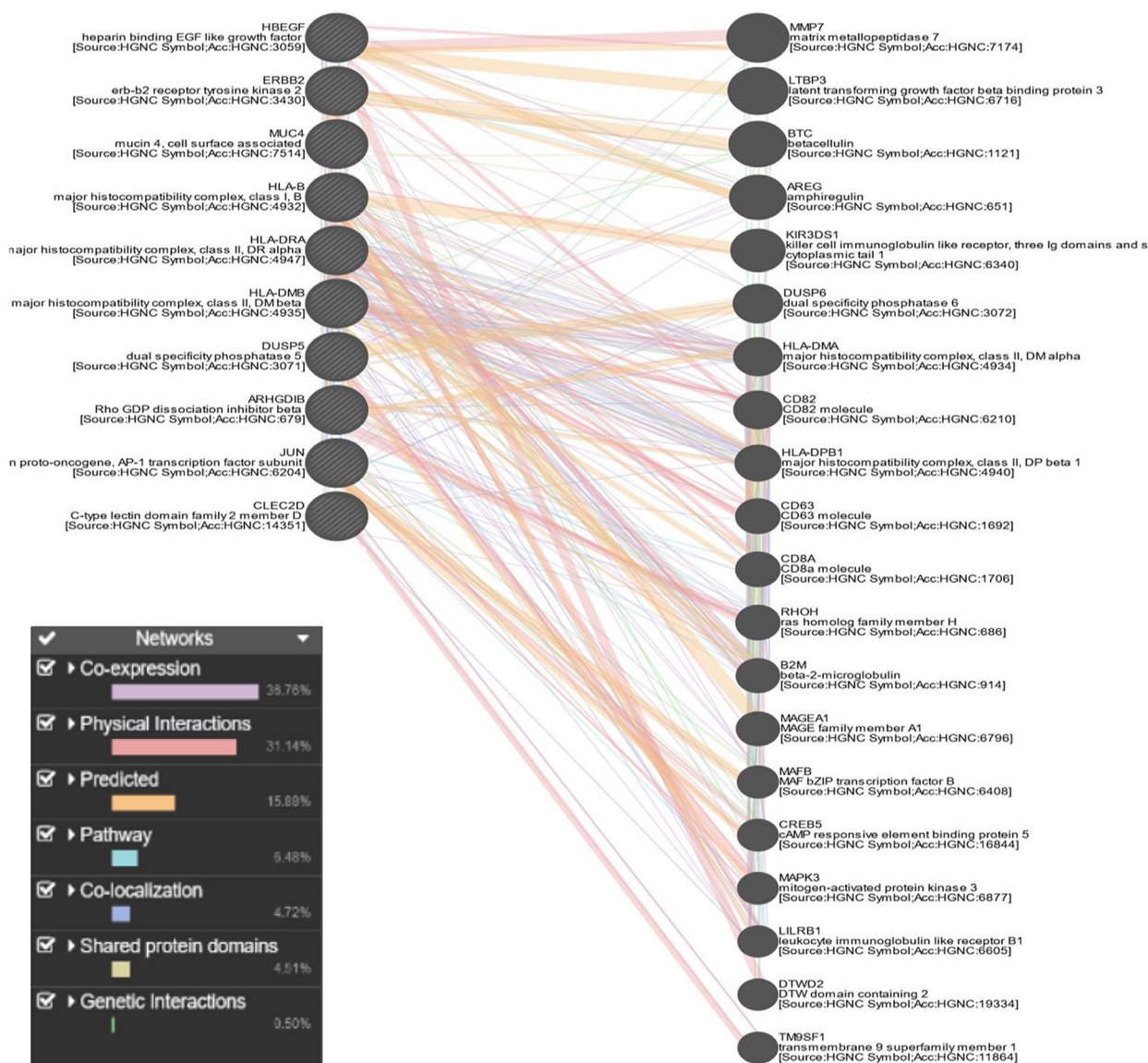
its corresponding diseases individually. SubNetwork1 contains 117 nodes, 118 edges, and five seeds, and Subnetwork2 contains 12 nodes, 11 edges, one seed, and Subnetwork3 contains four nodes, three edges, and one seed. The network as shown in Fig. 14 has 133 nodes, 132 edges, and seven seeds in total.

Fig. 11. Functional association network through GENEMANIA. 36.76% coexpression, 31.14% physical interactions, 15.88% predicted, 6.48% pathway, 4.72% colocalization, 4.51% shared protein domains and 0.50% genetic interactions are found here.

to some common genes/proteins as drivers. To do the investigation, microarray datasets were collected accordingly, followed by differential expression analysis with the Benjamini–Hochberg FDR approach. Common DEGs from the two datasets were then gathered once the differentially expressed genes had been found from the dataset. The threshold for FDR was chosen as <0.05, as it is quite common in statistical hypothesis testing the relevant literature. While we use a stringent FDR < 0.05 cutoff to retrieve DEGs, a substantially large number of common genes were retrieved. If we apply stringent FDR (i.e., <0.10), the necessary (false-positive) genes have been also used to deduct. If

important genes are mistakenly excluded from the analysis, it can lead to inaccurate identification of differentially expressed genes between conditions. These deducted genes can also affect the entire downstream analysis pipeline with false-positive results. This may result in erroneous drug prediction, which was very crucial for our aim, as we wanted to find out the targeted molecules for two different diseases. So, we have to find out the most effective responsible genes rather than retrieving a large number of common genes. Thirty-two genes (GRAMD3, SYNGR1, CMTM7, HLA-DMB, HLA-DRB4, CLEC2D, CCNB1IP1, HLA-DRA, DEFB1, ERBB2, MUC4, LOC145837, RPS24, RHOD,
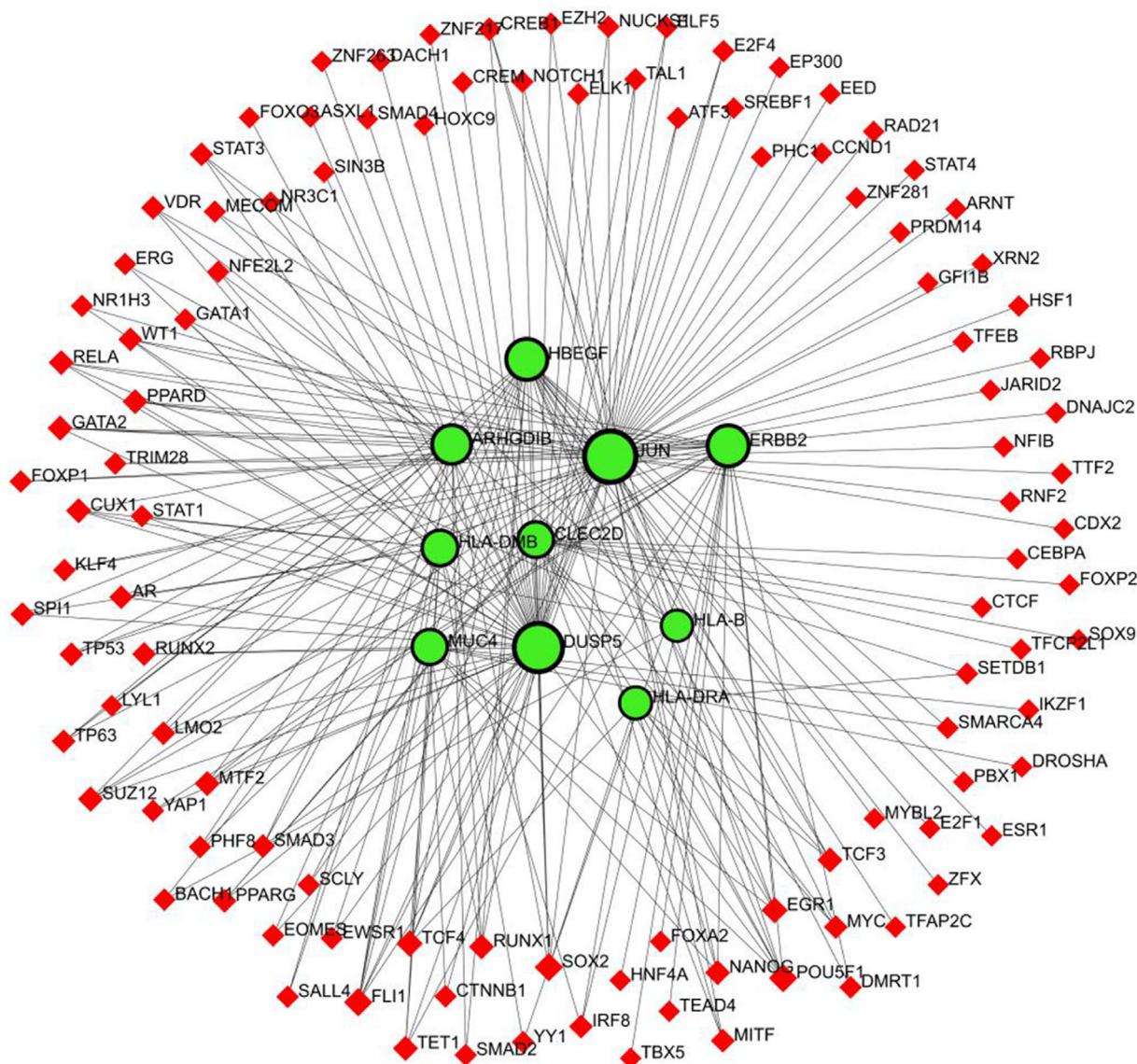
Fig. 12. Visualization of TF-gene association network through NETWORKANALYST. Red diamond-shaped nodes indicate TF-gene and green-black highlighted circle-shaped nodes for seeds.

HBEGF, ARHGDIB, PITPNC1, PEA15, KRT16, GJB3, JUN, FHL2, CYB5R2, HLA-B, EGR2, HERC6, DUSP5, HBA2, CHST15, NKX3-1, LBH, TSPAN8) are common in hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma. Further analysis was done by using these common DEGs.

To discover GO terms and pathways, an analysis of gene set enrichment was performed. The PPI network is the most notable part for detecting interconnected genes, disconnected genes, and hub genes identification. HLA-B, JUN, ERBB2, RPS24, FHL2, HLA-DRA, KRT16, HBA2, EGR2, ARHGDIB, HLA-

DMB, CLEC2D, MUC4, RHOD, HBEGF, GJB3, DUSP5 were pinpointed as directly interconnected genes and JUN, ERBB2, HLA-DMB, HBEGF, HLA-B, HLA-DRA, DUSP5, ARHGDIB, MUC4, CLEC2D were spotted as hub genes according to the degree method and the MCC.

MicroRNAs are gene-silencing factors. Common genes were used for retrieving their corresponding miRNAs and TFs. Thirty-two miRNAs and 33 TFs were produced for the targeted gene set. The network has 93 nodes, 223 edges, and 28 seeds. The miRNA has-miR-27a has five, which is the highest interaction
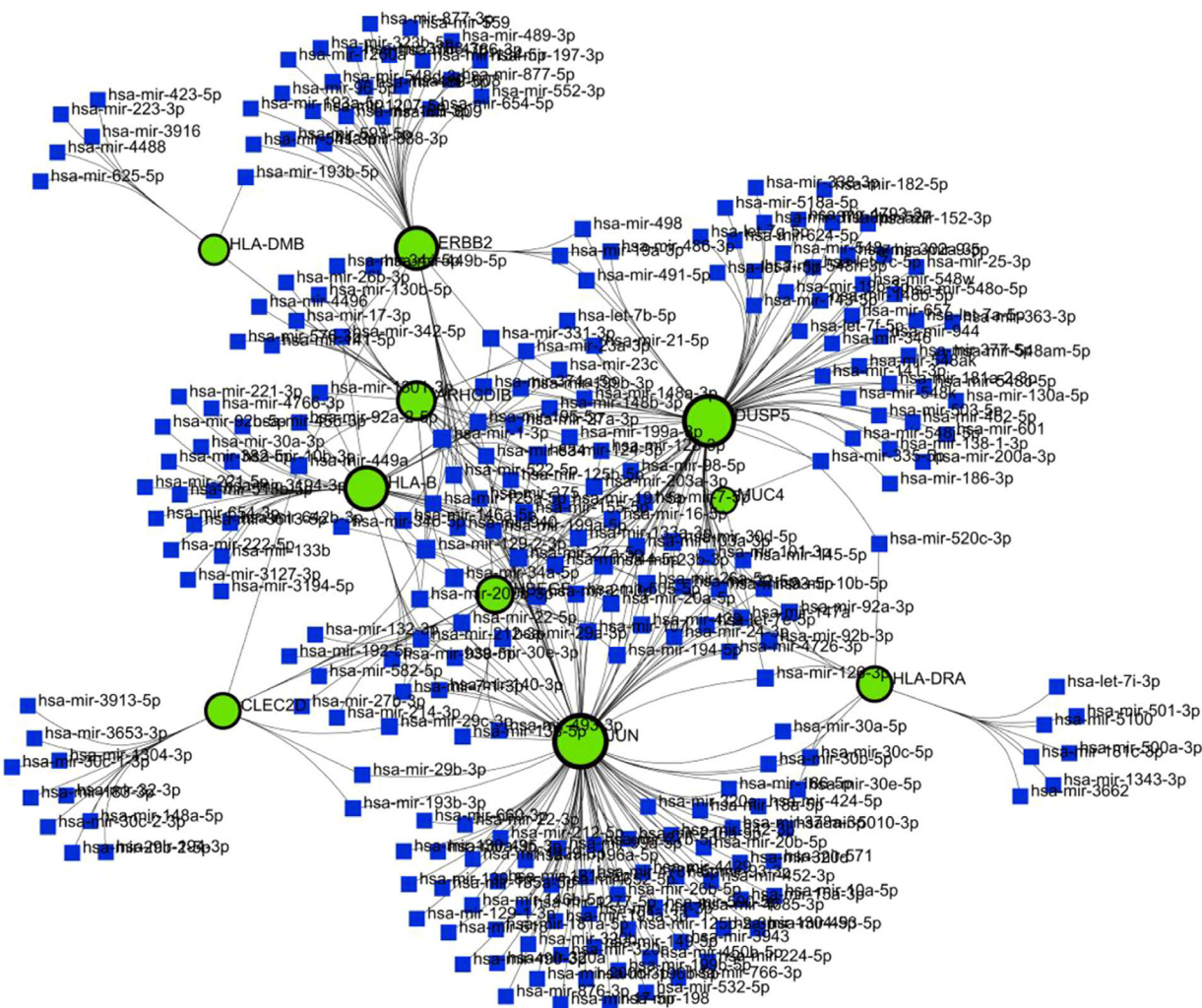
Fig. 13. Visualization of gene-miRNA network through NETWORKANALYST. Green-black highlighted nodes indicate 17 seeds and blue box-shaped nodes for miRNA, edges connect the genes and miRNAs.

among all the miRNAs and the MAX TF gene has eight, which is also the highest interaction among all TFs in TF-miRNA coregulatory network.

A TF-gene network was contrived by 10 hub genes of hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma. TFs are the promoter genes that are responsible for transcription. The control of gene expression is carried out by particular genes that interact with TF genes, which act as reactors for this control. JUN has an elevated interconnection among all the networks in our study. Another gene-miRNA network was analyzed for those hub genes that were identified earlier in the PPI network. miRNAs that can regulate gene expression by slowing down mRNA synthesis [97]. To know the miRNAs produced for those 10 hub genes, a gene-miRNA network was fabricated.

In our investigation, 261 miRNAs are produced for those 10 hub genes. The gene–disease association network was constructed for the same hub genes. To know the risk genes among them, which can be a cause for other associated diseases. If these risk genes can't be cured by any molecules, then these associated diseases can occur in the near future.

Note, in this study we collected the protein–protein interaction information from the STRING database. In this database, PPI information is defined based on various types of gene/protein association evidence, e.g., known interactions (curated databases, experimentally verified interactions), predicted interactions (gene neighborhood, gene fusion, gene co-occurrences), and other types of associations (text-mining, coexpression, protein homology). Therefore, we argue that the
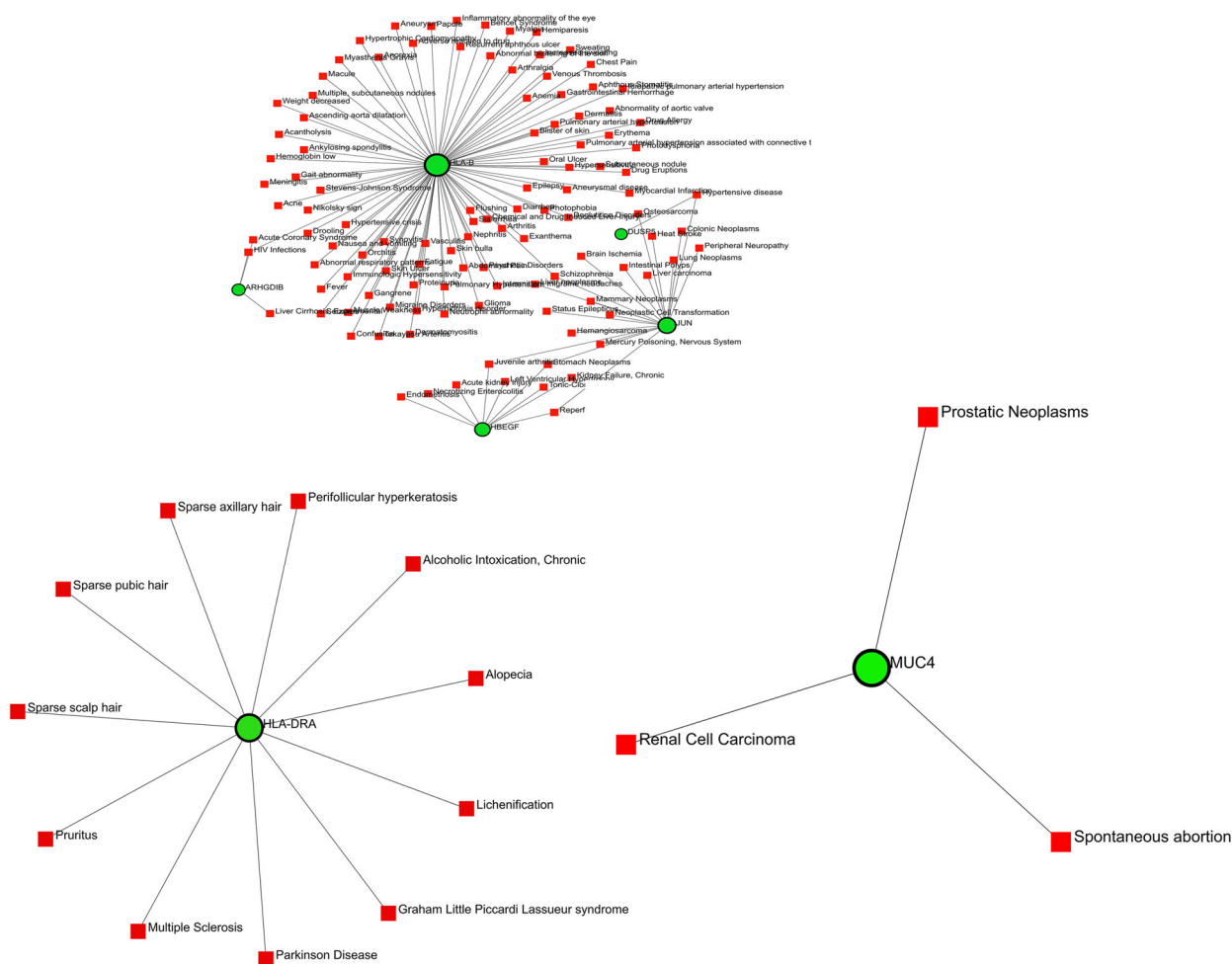
Fig. 14. Gene–disease network is divided into three subnetworks. Subnetwork1 represents genes (HLA-B, DUSP5, ARHGDIB, JUN, HBEGF) and their corresponding associated diseases, Subnetwork2 acts for HLA-DRA genes with its associated diseases and Subnetwork3 focuses on the MUC4 gene with its correspondent genes. Here, green-black highlighted nodes for seeds and red box-shaped nodes for associated diseases.

Table 5. EGFR-mutated lung adenocarcinoma patients and those with hypopharyngeal cancer: predicated drugs.

Drug's name	Adjusted *P*-value	Genes
Retinoic acid CTD 00006918	0.007990864	DUSP5;JUN;ERBB2; ARHGDIB;HLA-B;HLA-DRA;MUC4;HBEGF
Arsenenous acid CTD 00000922	0.010633891	JUN;ERBB2;ARHGDIB; HLA-B;HBEGF
TERT-BUTYL HYDROPEROXIDE CTD 00007349	0.012283126	*DUSP5;JUN;ERBB2; HLA-B;HBEGF*
Carbamazepine CTD 00005574	0.007027547	*JUN;ERBB2;HLA-B;HBEGF*
Etoposide CTD 00005948	0.007027547	*JUN;ERBB2;ARHGDIB; HBEGF*
Tonzonium bromide PC3 UP	0.007027547	*DUSP5;JUN;HBEGF*
NVP-TAE684 CTD 00004657	0.007027547	ERBB2;HBEGF
Prostaglandin J2 CTD 00001744	0.007027547	JUN;ARHGDIB

PPI information used in our study is substantially comprehensive, i.e., covering not only protein–protein association but also functional association (indirectly), and TF-gene interaction. However, for gene-miRNA and gene–disease interaction, the STRING database is not evidence of such types of evidence. However, for our future work we aim to augment those types of gene/protein associations (gene-miRNA and

gene–disease) along with STRING PPI, to form an integrated network, upon which we will conduct further downstream analysis.

Finally, drug molecules based on the 10 hub genes were recommended. These 10 genes are considered the most affected and responsible genes among all. If we cure these genes with any therapeutic molecules, other genes that are connected to these 10 genes will be also cured by the same therapeutic molecules. If these 10 genes are not cured by molecules, other associated diseases can occur due to these genes. From our suggested eight drugs, the retinoic acid CTD 00006918 can affect eight hub genes among the 10 hub genes. Patients having both diseases (hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma) concurrently may have a higher possibility of cure by using our suggested drug compounds. Our suggested drugs may have the potential for the treatment of these two diseases, but this requires experimental validation and further testing. If some related illnesses are found with hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma, then future research in this area aims to create a single generic drug to treat some related illnesses, offering a fresh perspective.

Conclusion

This study mentioned that the selected two diseases may have the possibility to metastasize to one another. Analyzing any disease means analyzing the disease genes. The constructed PPI networks displayed all the directly associated genes, general genes, and a channel that ensures the route to a general remedy map. The Cyto-hubba module was used to identify 10 hub genes using the degree topology approach and the MCC. Only those genes that are interconnected with each other and have a higher interaction among all were taken for this research purpose. If we can recover the directly affected, higher interconnected genes of a disease, we can get rid of those selected diseases (hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma). The next step is to employ GeneMania to develop a new network for the 32 shared genes to learn more about their physical interactions, shared protein domains, shared pathways, and genetic interactions. TF-gene, gene-miRNA, and gene–disease association networks were designed by using the same 10 hub genes. After analyzing those networks, some well-known therapeutic molecules were suggested for hypopharyngeal cancer and EGFR-mutated lung adenocarcinoma by using the 10 hub genes as input. A common drug for selected two associated diseases aims to reduce the amount of drug one should use and also reduce cost. Future research in this area aims to create a single generic drug to treat several related illnesses, offering a fresh perspective.

Acknowledgements

Conflict of interest

The authors declare no conflicts of interest.

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/2211-5463.13807.

Data accessibility

The selected datasets are sourced from free and open-access sources such as the GEO database https://www.ncbi.nlm.nih.gov/geo/.

Author contributions

AB: data curation, methodology, software, writing-original draft preparation; MMI: conceptualization, supervision, methodology, formal analysis, visualization, writing-original draft preparation; MAU: visualization, investigation, formal analysis; MAT: investigation, validation, visualization, writing – reviewing and editing; AKMA: visualization, validation, writing – reviewing and editing; SA, BKP, WT, and MAAA: visualization, validation; MAM: analysis, validation, visualization, writing – reviewing and editing.

References

1 Phoebe Chen Y-P and Chen F (2008) Identifying targets for drug discovery using bioinformatics. *Expert Opin Ther Targets* **12**, 383–389.

2 Hanif MA (2018) Head and neck cancer-Bangladesh perspective. *Bangladesh J Otorhin* **24**, 1–2.

3 Sanders O and Pathak S (2022) Hypopharyngeal cancer.

4 Hoffmann M, Görögh T, Gottschlich S, Lohrey C, Rittgen W, Ambrosch P, Schwarz E and Kahn T (2005)

Human Papillomaviruses in head and neck cancer: 8 year-survival-analysis of 73 patients. *Cancer Lett* **218**, 199–206.

5   Rodrigo JP, Hermsen MA, Fresno MF, Brakenhoff RH, García-Velasco F, Snijders PJ, Heideman DA and García-Pedrero JM (2015) Prevalence of human papillomavirus in laryngeal and hypopharyngeal squamous cell carcinomas in northern Spain. *Cancer Epidemiol* **39**, 37–41.

6   Lindquist D, Romanitan M, Hammarstedt L, Näsman A, Dahlstrand H, Lindholm J, Onelöv L, Ramqvist T, Ye W, Munck-Wikland E *et al.* (2007) Human papillomavirus is a favourable prognostic factor in tonsillar cancer and its oncogenic role is supported by the expression of E6 and E7. *Mol Oncol* **1**, 350–355.

7   Jin R, Peng L, Shou J, Wang J, Jin Y, Liang F, Zhao J, Wu M, Li Q, Zhang B *et al.* (2021) EGFR-mutated squamous cell lung cancer and its association with outcomes. *Front Oncol* **11**, 2262.

8   Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394–424.

9   Talukder MA, Layek MA, Kazi M, Uddin MA and Aryal S (2023) Empowering covid-19 detection: optimizing performance through fine-tuned efficientnet deep learning architecture. *Comput Biol Med* **168**, 107789.

10 Talukder MA, Islam MM, Uddin MA, Akhter A, Hasan KF and Moni MA (2022) Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. *Expert Syst Appl* **205**, 117695.

11 Khatun M, Islam MM, Rifat HR, Shahid MSB, Talukder MA and Uddin MA (2023) Hybridized convolutional neural networks and long short-term memory for improved Alzheimer's disease diagnosis from MRI scans. In *2023 26th International Conference on Computer and Information Technology (ICCIT)*, 1–6. IEEE.

12 Wild C, Weiderpass E and Stewart B (2020) *World Cancer Report: Cancer Research for Cancer Prevention*. International Agency for Research on Cancer, World Health Organization, Geneva, Switzerland.

13 Mukti RF, Samadder PD, Emran AA, Ahmed F, Imran IB, Malaker A and Yeasmin S (2014) Score based risk assessment of lung cancer and its evaluation for Bangladeshi people. *Asian Pac J Cancer Prev* **15**, 7021–7027.

14 Klein F, Kotb WFA and Petersen I (2009) Incidence of human Papilloma virus in lung cancer. *Lung Cancer* **65**, 13–18.

15 van Hooren MR, Leunis N, Brandsma DS, Dingemans A-MC, Kremer B and Kross KW (2016) Differentiating

head and neck carcinoma from lung carcinoma with an electronic nose: a proof of concept study. *Eur Arch Otorhinolaryngol* **273**, 3897–3903.

16 Kaifi JT, Gusani NJ, Deshaies I, Kimchi ET, Reed MF, Mahraj RP and Staveley-O'Carroll KF (2010) Indications and approach to surgical resection of lung metastases. *J Surg Oncol* **102**, 187–195.

17 Pfannschmidt J, Egerer G, Bischof M, Thomas M and Dienemann H (2012) Surgical intervention for pulmonary metastases. *Dtsch Arztebl Int* **109**, 652.

18 Kondo H, Okumura T, Ohde Y and Nakagawa K (2005) Surgical treatment for metastatic malignancies. Pulmonary metastasis: indications and outcomes. *Int J Clin Oncol* **10**, 81–85.

19 Saleh W, AlShammari A, Sarraj J, AlAshgar O, Ahmed MH and AlKattan K (2018) Surgical treatment of pulmonary metastasis: report from a tertiary care center. *Asian Cardiovasc Thorac Ann* **26**, 296–301.

20 Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D and Bray F (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**, E359–E386.

21 Shen N, Li T, Zhou L and Zhou X (2021) Lung metastases in newly diagnosed hypopharyngeal cancer: a population-based study. *Eur Arch Otorhinolaryngol* **278**, 4469–4476.

22 Pillay J, Rahman S, Klarenbach S, Reynolds DL, Tessier LA, Thériault G, Persaud N, Finley C, Leighl N, McInnes MD *et al.* (2024) Screening for lung cancer with computed tomography: protocol for systematic reviews for the canadian task force on preventive health care. *Syst Rev* **13**, 88.

23 Islam MR, Alam MK, Paul BK, Koundal D, Zaguia A and Ahmed K (2022) Identification of molecular biomarkers and key pathways for esophageal carcinoma (EsC): a bioinformatics approach. *Biomed Res Int* **2022**, 5908402.

24 Taz TA, Ahmed K, Paul BK, Kawsar M, Aktar N, Mahmud SH and Moni MA (2021b) Network-based identification genetic effect of SARS-CoV-2 infections to idiopathic pulmonary fibrosis (IPF) patients. *Brief Bioinform* **22**, 1254–1266.

25 Taz TA, Ahmed K, Paul BK, Al-Zahrani FA, Mahmud SH and Moni MA (2021a) Identification of biomarkers and pathways for the SARS-CoV-2 infections that make complexities in pulmonary arterial hypertension patients. *Brief Bioinform* **22**, 1451–1465.

26 Chen F, Zheng A, Li F, Wen S, Chen S and Tao Z (2019) Screening and identification of potential target genes in head and neck cancer using bioinformatics analysis. *Oncol Lett* **18**, 2955–2966.

27 Ye Y, Wang J, Liang F, Song P, Yan X, Wu S, Huang X and Han P (2021) Identification of key genes for HNSCC from public databases using bioinformatics analysis. *Cancer Cell Int* **21**, 549.

28 Jin Y and Yang Y (2019) Identification and analysis of genes associated with head and neck squamous cell carcinoma by integrated bioinformatics methods. *Mol Genet Genom Med* **7**, e857.

29 Li CY, Cai J-H, Tsai JJ and Wang CC (2020) Identification of hub genes associated with development of head and neck squamous cell carcinoma by integrated bioinformatics analysis. *Front Oncol* **10**, 681.

30 Shen Y, Liu J, Zhang L, Dong S, Zhang J, Liu Y, Zhou H and Dong W (2019) Identification of potential biomarkers and survival analysis for head and neck squamous cell carcinoma using bioinformatics strategy: a study based on TCGA and GEO datasets. *Biomed Res Int* **2019**, 7376034.

31 Tu Z, He X, Zeng L, Meng D, Zhuang R, Zhao J and Dai W (2021) Exploration of prognostic biomarkers for lung adenocarcinoma through bioinformatics analysis. *Front Genet* **12**, 647521.

32 Ni KW and Sun GZ (2019) The identification of key biomarkers in patients with lung adenocarcinoma based on bioinformatics. *Math Biosci Eng* **16**, 7671–7687.

33 Guo T, Ma H and Zhou Y (2019) Bioinformatics analysis of microarray data to identify the candidate biomarkers of lung adenocarcinoma. *PeerJ* **7**, e7313.

34 Zhang M-Y, Liu X-X, Li H, Li R, Liu X and Qu Y-Q (2018) Elevated mRNA levels of AURKA, CDC20 and TPX2 are associated with poor prognosis of smoking related lung adenocarcinoma using bioinformatics analysis. *Int J Med Sci* **15**, 1676–1685.

35 Sturn A, Quackenbush J and Trajanoski Z (2002) Genesis: cluster analysis of microarray data. *Bioinformatics* **18**, 207–208.

36 Lee M-LT, Kuo FC, Whitmore G and Sklar J (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* **97**, 9834–9839.

37 Edgar R, Domrachev M and Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–210.

38 Davis S and Meltzer PS (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and bioconductor. *Bioinformatics* **23**, 1846–1847.

39 Clough E and Barrett T (2016) The Gene Expression Omnibus database. *Methods Mol Biol* **1418**, 93–110.

40 Smyth GK (2005) LIMMA: linear models for microarray data. 397–420.

41 Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol* **57**, 289–300.

42 Oliveros JC (2007) Venny. an interactive tool for comparing lists with Venn diagrams. http://bioinfogp. cnb.csic.es/tools/venny/index.html

43 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al.* (2005) Gene Set Enrichment Analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550.

44 Doms A and Schroeder M (2005) GoPubMed: exploring pubmed with the gene ontology. *Nucleic Acids Res* **33**, W783–W786.

45 Kanehisa M and Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30.

46 Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B *et al.* (2018) The reactome pathway knowledgebase. *Nucleic Acids Res* **46**, D649–D655.

47 Nishimura D (2001) Biocarta. *Biotech Softw Intern Rep Comput Softw J Scient* **2**, 117–120.

48 Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Mélius J, Cirillo E, Coort SL, Digles D *et al.* (2018) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* **46**, D661–D667.

49 Al-Mustanjid M, Mahmud SH, Royel MRI, Rahman MH, Islam T, Rahman MR and Moni MA (2020) Detection of molecular signatures and pathways shared in inflammatory bowel disease and colorectal cancer: a bioinformatics and systems biology approach. *Genomics* **112**, 3416–3426.

50 Hsu S-D, Lin F-M, Wu W-Y, Liang C, Huang W-C, Chan W-L, Tsai W-T, Chen G-Z, Lee C-J, Chiu C-M *et al.* (2011) miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res* **39**, D163–D169.

51 Liu Z-P, Wu C, Miao H and Wu H (2015) RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database (Oxford)* **2015**, bav095.

52 Zhou G, Soufan O, Ewald J, Hancock RE, Basu N and Xia J (2019) NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res* **47**, W234–W241.

53 Xia J, Gill EE and Hancock RE (2015) NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* **10**, 823–844.

54 Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M *et al.* (2007) Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol Syst Biol* **3**, 89.

55 Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes H-W *et al.* (2005) The MIPS mammalian protein–protein interaction database. *Bioinformatics* **21**, 832–834.

56 Chowdhury UN, Islam MB, Ahmad S and Moni MA (2020) Network-based identification of genetic factors in ageing, lifestyle and type 2 diabetes that influence to the progression of Alzheimer's disease. *Inform Med Unlocked* **19**, 100309.

57 Ben-Hur A and Noble WS (2005) Kernel methods for predicting protein–protein interactions. *Bioinformatics* **21**, i38–i46.

58 Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P and Snel B (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* **31**, 258–261.

59 Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P *et al.* (2010) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* **39**, D561–D568.

60 Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS and Lynn DJ (2013) InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. *Nucleic Acids Res* **41**, D1228–D1233.

61 Hsing M, Byler KG and Cherkasov A (2008) The use of Gene Ontology terms for predicting highly-connected 'hub' nodes in protein-protein interaction networks. *BMC Syst Biol* **2**, 80.

62 Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T and Lin C-Y (2014) cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* **8**, S11.

63 Shi L, Wen Z, Li H and Song Y (2021) Identification of hub genes associated with tuberculous pleurisy by integrated bioinformatics analysis. *Front Genet* **12**, 730491.

64 Wang S, Song Z, Tan B, Zhang J, Zhang J and Liu S (2021) Identification and validation of hub genes associated with hepatocellular carcinoma via integrated bioinformatics analysis. *Front Oncol* **11**, 614531.

65 Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A and Raval A (2009) Identifying hubs in protein interaction networks. *PLoS ONE* **4**, e5344.

66 Habib N (2017) Drug design and analysis for bipolar disorder and associated diseases: A bioinformatics approach. *Netw Biol* **7**, 41.

67 Zuberi K, Franz M, Rodriguez H, Montojo J, Lopes CT, Bader GD and Morris Q (2013) Genemania prediction server 2013 update. *Nucleic Acids Res* **41**, W115–W122.

68 Frishman D and Valencia A (2009) *Modern Genome Annotation. The BioSapiens Network*. Springer, New York, NY.

69 Tieri P, Farina L, Petti M, Astolfi L, Paci P and Castiglione F (2019) Network inference and reconstruction in bioinformatics. In *Encyclopedia of Bioinformatics and Computational Biology*

(Ranganathan S, Gribskov M, Nakai K and Schönbach C, eds), pp. 805–813. Elsevier, Amsterdam.

70 Ruan J, Dean AK and Zhang W (2010) A general coexpression network-based approach to gene expression analysis: comparison and applications. *BMC Syst Biol* **4**, 8.

71 Hasan MR, Paul BK, Ahmed K and Bhuyian T (2020) Design protein–protein interaction network and protein-drug interaction network for common cancer diseases: a bioinformatics approach. *Inform Med Unlocked* **18**, 100311.

72 Roy S, Bhattacharyya DK and Kalita JK (2014) Reconstruction of gene coexpression network from microarray data using local expression patterns. *BMC Bioinformatics* **15**, S10.

73 Basar MA, Hosen MF, Paul BK, Hasan MR, Shamim S and Bhuyian T (2023b) Identification of drug and protein–protein interaction network among stress and depression: a bioinformatics approach. *Inform Med Unlocked* **37**, 101174.

74 Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD and Morris Q (2018) GeneMANIA update 2018. *Nucleic Acids Res* **46**, W60–W64.

75 Ye Z, Wang F, Yan F, Wang L, Li B, Liu T, Hu F, Jiang M, Li W and Fu Z (2019) Bioinformatic identification of candidate biomarkers and related transcription factors in nasopharyngeal carcinoma. *World J Surg Oncol* **17**, 60.

76 Ye Y, Gao L and Zhang S (2017) Integrative analysis of transcription factor combinatorial interactions using a Bayesian tensor factorization approach. *Front Genet* **8**, 140.

77 Basar MA, Hasan MR, Paul BK, Shadhin KA and Mollah MS (2023a) A system biology and bioinformatics approach to determine the molecular signature, core ontologies, functional pathways, drug compounds in between stress and type 2 diabetes. In *International Work-Conference on Bioinformatics and Biomedical Engineering* pp. 320–331. Springer,

78 Tran LM, Hyduke DR and Liao JC (2010) Trimming of mammalian transcriptional networks using network component analysis. *BMC Bioinformatics* **11**, 511.

79 Kazemian M, Pham H, Wolfe SA, Brodsky MH and Sinha S (2013) Widespread evidence of cooperative DNA binding by transcription factors in Drosophila development. *Nucleic Acids Res* **41**, 8237–8252.

80 Xia J, Benner MJ and Hancock RE (2014) NetworkAnalyst-integrative approaches for protein–protein interaction network analysis and visual exploration. *Nucleic Acids Res* **42**, W167–W174.

81 Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR and Ma'ayan A (2010) ChEA: transcription factor regulation inferred from integrating genome-wide ChiP-X experiments. *Bioinformatics* **26**, 2438–2444.

82 León LE and Calligaris SD (2017) Visualization and analysis of miRNA–targets interactions networks. *Methods Mol Biol* **1509**, 209–220.

83 Kuhn DE, Martin MM, Feldman DS, Terry AV Jr, Nuovo GJ and Elton TS (2008) Experimental validation of miRNA targets. *Methods* **44**, 47–54.

84 Lynam-Lennon N, Maher SG and Reynolds JV (2009) The roles of microRNA in cancer and apoptosis. *Biol Rev* **84**, 55–71.

85 Shah A, Meese E and Blin N (2010) Profiling of regulatory microRNA transcriptomes in various biological processes: a review. *J Appl Genet* **51**, 501–507.

86 Carthew RW (2006) Gene regulation by microRNAs. *Curr Opin Genet Dev* **16**, 203–208.

87 Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, Papadimitriou D, Kavakiotis I, Maniou S, Skoufos G *et al.* (2018) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res* **46**, D239–D245.

88 Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, Garca-Garca J, Sanz F and Furlong LI (2016) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833–D839.

89 Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F and Furlong LI (2011) Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. *PLoS ONE* **6**, e20284.

90 Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR and Ma'ayan A (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128.

91 Yoo M, Shin J, Kim J, Ryall KA, Lee K, Lee S, Jeon M, Kang J and Tan AC (2015) DSigDB: drug signatures database for gene set analysis. *Bioinformatics* **31**, 3069–3071.

92 Gene Ontology Consortium (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258–D261.

93 Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T *et al.* (2007) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* **36**, D480–D484.

94 Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, Miller A, Digles D, Lopes EN, Ehrhart F *et al.* (2021) WikiPathways: connecting communities. *Nucleic Acids Res* **49**, D613–D621.

95 Batalia MA and Collins EJ (1997) Peptide binding by class I and class II MHC molecules. *Pept Sci* **43**, 281–302.

96 Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214–W220.

97 Zhang H-M, Kuang S, Xiong X, Gao T, Liu C and Guo A-Y (2015) Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Brief Bioinform* **16**, 45–58.