MALAYSIAN JOURNAL OF BIOCHEMISTRY & MOLECULAR BIOLOGY

The Official Publication of The Malaysian Society For Biochemistry & Molecular Biology (MSBMB)
http://mjbmb.org

GENE EXPRESSION MINING AND *in silico* ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES IN HEAD AND NECK CANCER

Hui Shi Saw[a], Bernard Kok Bang Lee[b, c], Hwei-San Loh[a*]

[a]*School of Biosciences, Faculty of Science and Engineering, University of Nottingham Malaysia, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia*
[b]*Genomics and Bioinformatics Research Unit, Cancer Research Malaysia, Sime Darby Medical Centre Subang Jaya, 2nd Floor, Outpatient Centre, 47500 Subang Jaya, Selangor, Malaysia*
[c]*Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*

*Corresponding Author: Sandy.Loh@nottingham.edu.my

History
Received: 3 September 2022
Accepted: 16 November 2023

Keywords:

Biomarker; Differential Gene Expression; GEO; Head and Neck Squamous Cell Carcinoma; Prognosis; TCGA

Abstract

The advancement of high-throughput transcriptome profiling techniques, such as next-generation sequencing and microarray, has led to the development of bioinformatics tools and databases for functional genomics. Integrated bioinformatics analysis has emerged as a promising strategy to address the major cause of morbidity and death globally: cancer. In this study, we aimed to use an integrated bioinformatics pipeline to identify potential molecular biomarkers for diagnosis and prognosis in cancer studies. Specifically, we focused on head and neck squamous cell carcinoma (HNSCC). To achieve this, we performed a meta-analysis on expression datasets from the Gene Expression Omnibus (GEO) using GEO2R to derive differentially expressed genes (DEGs). Subsequently, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Protein-protein interaction networks of the up-regulated and down-regulated genes were constructed using the STRING database, and the top ten hub genes for each group were identified using cytoHubba. The relative mRNA expression of the identified DEGs was validated with GEPIA2, and their correlation with the overall survival of HNSCC patients was assessed using Kaplan-Meier analysis. Combining our findings with published evidence, we observed that the up-regulated genes primarily function in the extracellular matrix and cell cycle regulation. In contrast, the down-regulated genes are involved in muscle contraction. Our results suggest that six down-regulated genes (*MYL1, MYL3, MYH6, MYLPF, ACTA1, TTN*) and five up-regulated genes (*CDC20, CCNB1, MAD2L1, TOP2A, MMP9*) have the potential to serve as diagnosis biomarkers. In contrast, five up-regulated genes (*FN1, CDK1, PLK1, AURKA, CD44*) could be used for prognosis and diagnosis in the clinical analysis of HNSCC. This study demonstrated the effectiveness of an integrated bioinformatics approach in identifying clinically relevant biomarkers for HNSCC, and the pipeline could be applied to other cancer datasets. Further investigation of the identified biomarkers will enrich our understanding of their involvement in the molecular mechanism of carcinogenesis and provide potential therapeutic targets for HNSCC.

INTRODUCTION

Derivation of Differentially Expressed Genes (DEGs)

Bioinformatics analysis is widely recognised for its application in deriving differentially expressed genes (DEGs), which can serve as potential biomarkers to predict cancer development and progression. The emergence of transcriptomic profiling techniques, public repositories for transcriptomic data, and differential gene expression analysis tools have significantly contributed to the development of meta-analysis in cancer research. In this population-based study, we describe and demonstrate the design of a robust analysis pipeline that integrates protein-protein interaction (PPI) networks and leverages literature evidence to gain a better understanding of individual gene functions.

Transcriptome Profiling and Meta-analysis in Integrative Bioinformatics

The advancement of transcriptome profiling has facilitated the exploration of differential gene expression, providing insights into the underlying molecular mechanism of diseases. DEGs are genes that exhibit distinct up-regulation or down-regulation in response to disease or treatment, making them potential biomarkers. Transcriptomic data, which captures the complete set of RNAs transcribed by the DNA of cells or tissues, can reveal differences in gene expression during disease development and specific physiological conditions. High-throughput transcriptomic technologies, such as microarray and next-generation sequencing (NGS), have played crucial roles in generating quality transcriptomic data for analysis. The ultimate goal of generating extensive transcriptomic data is elucidating the interactions among cellular components that govern cancer pathogenesis [1, 2].

The accessibility of transcriptomic data from online databases has contributed to the rapid development of bioinformatics tools. Archived experimental datasets from various diseases, synthesised by the research community, are deposited on platforms such as the Gene Expression Omnibus (GEO) [3] and The Cancer Genome Atlas (TCGA), a publicly accessible repository for cancer-related genomic profiles [4]. The meta-analysis, which combines retrospective results from multiple comparable datasets obtained from clinical samples, offers advantages in overcoming the limitation of small sample sizes, increasing statistical power, and producing clinically relevant results [5, 6]. The choice of software tools for identifying DEGS depends on the upstream processing of transcriptomic data obtained from different platforms and sources. However, downstream analysis of protein-protein interactions and functional prediction could be unified, with shared analysis routes for DEGs derived from both RNA-seq and microarray data. This highlights the flexibility of meta-analysis in identifying DEGs.

Features of a Biological Network

While studying gene regulation on an individual basis is essential, understanding the patterns of interactions between genes is equally essential to obtaining a conceptual view of functionally interacting molecules. Analysing differential gene expression with biological network studies has attracted attention to identifying hypothetical driver genes specific to certain biological pathways in cancer. The identification of targets based on biological networks, such as gene regulations, co-expression of gene targets, and protein-protein interactions, has been demonstrated in cancer studies, including lung cancers and hepatocellular carcinoma [7, 8, 9].

Biomolecular interaction networks are essential in predicting and conducting association studies of disease-causing genes. Network biology aims to establish connections between genes associated with specific functions or phenotypes, such as disease causation, with the positioning of each node not occurring randomly [10, 11, 12]. This approach has revealed increased interactions between protein products and tissue-specific co-expression in genetic disease research. Regarding the function of disease-related genes, it has been found that molecular functions, cellular components, and biological processes identified through Gene Ontology (GO) exhibit predictive capabilities. This theory could be further applied to predict the severity of disease outcomes following targeted treatment or to differentiate histopathologically similar diseases. It is postulated that neighbouring genes of a pathologically related gene within a network may function conjunctively in the same or closely related diseases [13, 14].

Building upon the advantages mentioned earlier, integrating network-based meta-analysis strengthened by incorporating literature evidence can further enhance the robustness of cancer biomarker discovery through bioinformatics analysis. Cancer biomarkers are measurable molecules crucial in detecting aberrant expression, distinguishing normal tissues from disease, predicting treatment outcomes, and assessing the risk of cancer progression or recurrence before histological alterations occur. Collectively, this information can guide clinical decision-making, justifying the application of appropriate therapeutic strategies and improving success rates [15].

Differentially Expressed Genes (DEGs) and Cancer Biomarker

The applications of integrative bioinformatics in identifying core DEGs involved in tumourigenesis, malignancy, or disease progression are considered prominent in cancer research. It has been widely applied to derive clinically

relevant biomarkers, as demonstrated in hepatocellular carcinoma [9], prostate cancer [16], and breast cancer [17]. Head and neck squamous cell carcinoma (HNSCC) ranks Asia's eleventh most common cancer, with approximately 931, 931 new incidences detected and 467, 125 deaths reported worldwide in 2020 [18]. The incidence rate and mortality of SCC in the lip, larynx, nasopharynx, oral cavity, oropharynx, hypopharynx, and salivary glands have not declined despite the advancement in the clinical management of oncology, including surgical removal of the tumour, radiotherapy, and chemotherapy [19, 20]. A subcategory of HNSCC, oral squamous cell carcinoma (OSCC), is a malignant neoplasia arising from the oral cavity's mucosal lining [21]. It could lead to dysfunctions in chewing, swallowing, and speech, significantly impacting the quality of life in patients.

Furthermore, recurrence of OSCC is potent due to the rich lymph vessels in the oral cavity, and thus, a close association with local expansion and lymphatic metastasis [22, 23]. The poor prognosis of OSCC is evident in the lack of significant improvement in the overall five-year survival rate, which has remained the lowest among malignancies (<50%) for the past thirty years [24]. Hence, searching for clinically applicable biomarkers is crucial in preventing late diagnosis, metastasis, and recurrences contributing to the poor prognosis of HNSCC [25].

Aim and Objectives of the Study

This study aims to demonstrate the application of an integrated bioinformatics pipeline for identifying DEGs in cancer studies. Its objective is to identify molecular biomarkers with potential diagnostic and prognostic value in HNSCC: i) to identify potential biomarkers for the diagnosis and prognosis of HNSCC, and ii) to reveal the molecular mechanism underlying HNSCC through integrated DEG analysis. The study describes a general workflow for network-based meta-analysis using HNSCC as the subject of study; however, its application of this approach can also extend to other types of cancer.

MATERIALS AND METHODS

Data Processing and Differential Expression Analysis

The differential expression analysis was initiated using .CEL files obtained from three separate datasets available on the NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). These three datasets, namely GSE13601, GSE30784, and GSE31056, comprise the microarray profiling studies of OSCC (Table 1). For each dataset, the control group consisted of normal oral tissues without preneoplastic oral lesions. Dysplasia and margin samples that do not fit the normal and tumour definitions were excluded from this study. Each dataset was subjected to GEO2R analysis (https://www.ncbi.nlm.nih.gov/geo/geo2r/), which is a web analysis tool that uses GEOquery (version 2.40.0) and limma package (version 3.26.8) to identify DEGs across two experimental groups: cancer and control [3]. The output from GEO2R was downloaded and subjected to joint filtering based on fold change and *p*-value. DEG with a significant threshold of $|\log_2 \text{fold change}| > 1$ and $p\text{-value} < 0.01$ were selected [26]. The analysis was performed using the R statistical programming software (version 4.0.2). The outputs of the differential gene expression analysis from GEO2R were visualised as a volcano plot, considering both statistical significance (*p*-value on the y-axis) and biological relevance (fold change as the x-axis) to identify significantly up-regulated and down-regulated genes.

Table 1. Details of the normal and tumour samples used for the three Gene Expression Omnibus (GEO) datasets

Dataset	Normal (n)	Tumour (n)	Reference
GSE13601	27	31	Estilo *et al*., 2009
GSE30784	45	167	Chen *et al*., 2008
GSE31056	24	23	Reis *et al*., 2011

After identifying the DEGs, additional data cleaning was performed. Microarray probe identity (ID) annotated with multiple gene names (e.g., annotated with "///") was updated using Gene Set Enrichment Analysis (GSEA, version 4.1.0). Probe IDs that still had multiple gene names were cross-referenced with BioGPS (http://biogps.org/#goto=welcome) for secondary verification and replacement with the latest gene name [30, 31, 32]. Metascape (https://metascape.org/gp/index.html#/main/step1) [33] and HUGO Gene Nomenclature Committee (HGNC, https://www.genenames.org/) [34] were then used to update the remaining gene to the latest gene symbols. In cases of genes with multiple copies, the genes showing the highest up-regulation or down-regulation, as indicated by its respective $\log_2$ fold change value, was retained. Lastly, the Venn diagram generated with Venny 2.1 (https://bioinfogp.cnb.csic.es/tools/venny/) was used to identify the DEGs common to all three datasets [35].

The overall workflow for section 2.1 is shown in Figure 1.

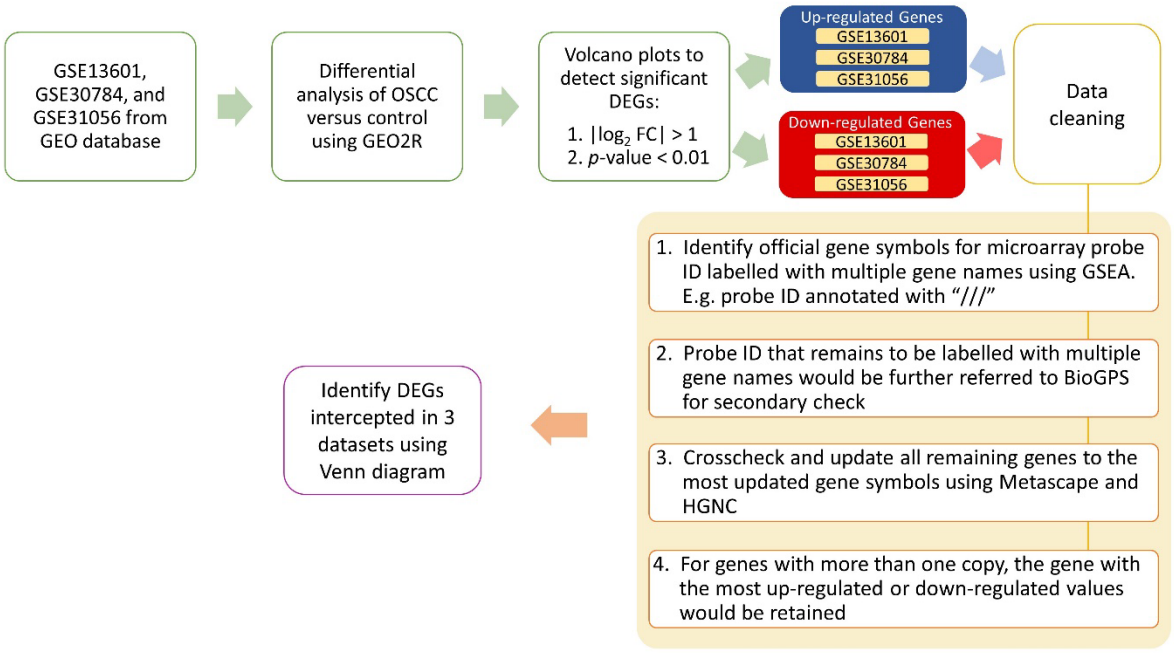


Figure 1. Workflow to shortlist significantly up-regulated and down-regulated genes using various bioinformatics tools.

Functional Enrichment Analysis of Gene Products

The official gene symbols of the common DEGs (*Homo sapiens* species) obtained from Venn diagrams were subjected to functional annotation using the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 (https://david.ncifcrf.gov/) [36, 37]. The parameters used for functional enrichment analysis were: count threshold = 2; modified Fisher exact *p*-value = 0.05. A report consisting of enriched pathways with a minimum of two genes involved and a statistically significant *p*-value (< 0.05) was generated. The top ten gene ontology (GO) terms for biological process, cellular component, molecular function, and the top ten enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were examined [38, 39]. Each enriched GO term or pathway name identified using DAVID is associated with its respective gene count, rich factor, and Fisher exact *p*-value. These values are utilised to understand the relationship between gene functions and cancer development. Bubble charts are created using the ggplot2 package (version 3.3.3) in R for visualisation [40].

Construction of Protein-protein Interaction (PPI) Networks

The interaction of gene products in the up-regulated and down-regulated groups, as identified in section 2.1, was separately searched against the STRING database version 11.0 (https://string-db.org/) [41]. A PPI network was constructed by selecting *Homo sapiens* as the parameter and shortlisting the top ten genes with the highest connectivity. The tabulated result from STRING was processed using the Cytoscape software (version 3.8.1) by transforming it into an organic layout [42]. To optimise the visualisation of the top modules within the PPI network, the Molecular Complex Detection (MCODE) tool within Cytoscape was employed. For both the up-regulated and down-regulated gene groups, the top three modules of network interactions were mapped using specific cut-off values: degree cut-off = 2, node score cut-off = 0.2, maximum depth = 100, and k-score = 2. Furthermore, the cytoHubba plugin in Cytoscape was employed to calculate the node scores and visualise the interactions of the top ten candidate genes ranked by their degree of connectivity [43].

The regulations of the twenty hub genes (ten for each expression group), selected from the cytoHubba analysis, were visualised through the construction of heatmaps using sample data obtained from the three GEO datasets. The Morpheus web application (https://software.broadinstitute.org/morpheus/) was utilised to visualise gene expression patterns across samples with varying experimental conditions. This hierarchical clustering, based on one minus Pearson correlation, was applied to the arranged samples and DEGs for enhanced visualisation.

mRNA Expression of Hub Genes

Following the identification of ten hub genes from each up-regulated and down-regulated group, the relative mRNA expression of each hub gene was validated. Single-gene analysis using the Gene Expression Profiling Interactive Analysis 2 (GEPIA2) tool (http://gepia2.cancer-pku.cn/#analysis) was performed for each shortlisted hub gene [44]. Each gene was matched to the HNSCC databases of The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression Project (GTEx) to generate boxplots. The boxplots were generated using a threshold value of $|\log_2$ fold change$|$ = 1 and p-value = 0.05.

Correlation of Cancer Survival to the Expression of Hub Genes

The correlation between the overall survival of HNSCC patients and the relative expression levels of hub genes, split by median, was identified using the Kaplan-Meier plotter for pan-cancer analysis (https://kmplot.com/analysis/) [45]. Kaplan-Meier mRNA curves derived from RNA-seq datasets obtained from GEO, TCGA, and European Genome-phenome Archive (EGA) were considered statistically significant if they had a p-value < 0.05.

RESULTS

Identification of Up-regulated and Down-regulated Genes from GEO Datasets

Volcano plots were generated using joint filtering of fold change and p-value based on the output generated from the GEO2R analysis (Figure 2). The number of up-regulated and down-regulated genes identified for each dataset can be found by referring to Figure 2, and the summary is provided in Table 2.

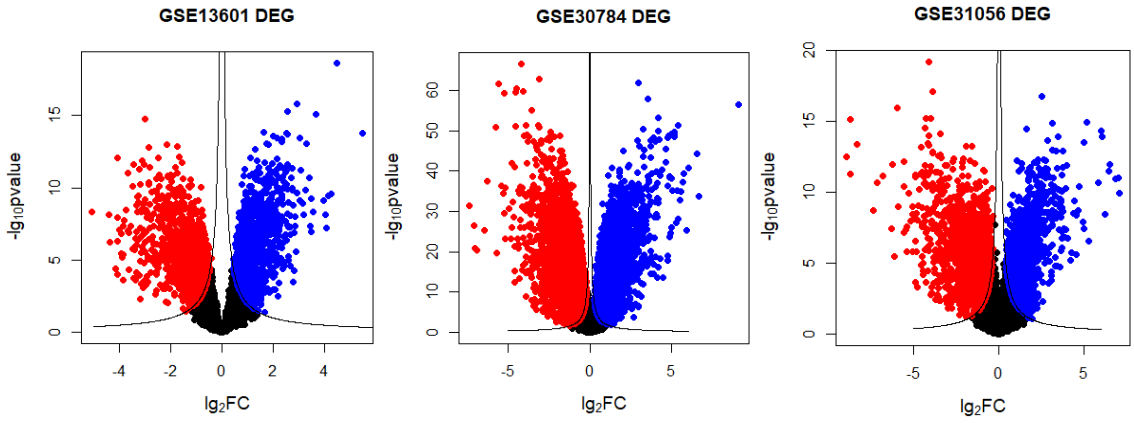


Figure 2. Volcano plots of DEGs generated using the joint filtering of $|\log_2$ fold change$|$ > 1 and $-\log_{10}$ p-value > 2 for three GEO datasets are shown. These scatterplots are useful for visually identifying significantly up-regulated genes (blue dots), down-regulated genes (red dots), and non-significant genes (black dots) across the three datasets. The identification is based on the set threshold of $-\log_{10}$ p-value > 2 and $|\log_2$ fold change$|$ > 1. The volcano plots were generated using the R statistical programming software.

Table 2. Number of up-regulated and down-regulated genes identified from the volcano plots of the respective datasets

Dataset	Up-regulated Genes (n)	Down-regulated Genes (n)
GSE13601	1184	1983
GSE30784	3668	3315
GSE31056	1732	1784

After identifying the up-regulated and down-regulated genes in each dataset, the presence of common DEGs across all three datasets was examined. As depicted in Figure 3, 454 up-regulated genes and 371 down-regulated genes were found to be common among the datasets.

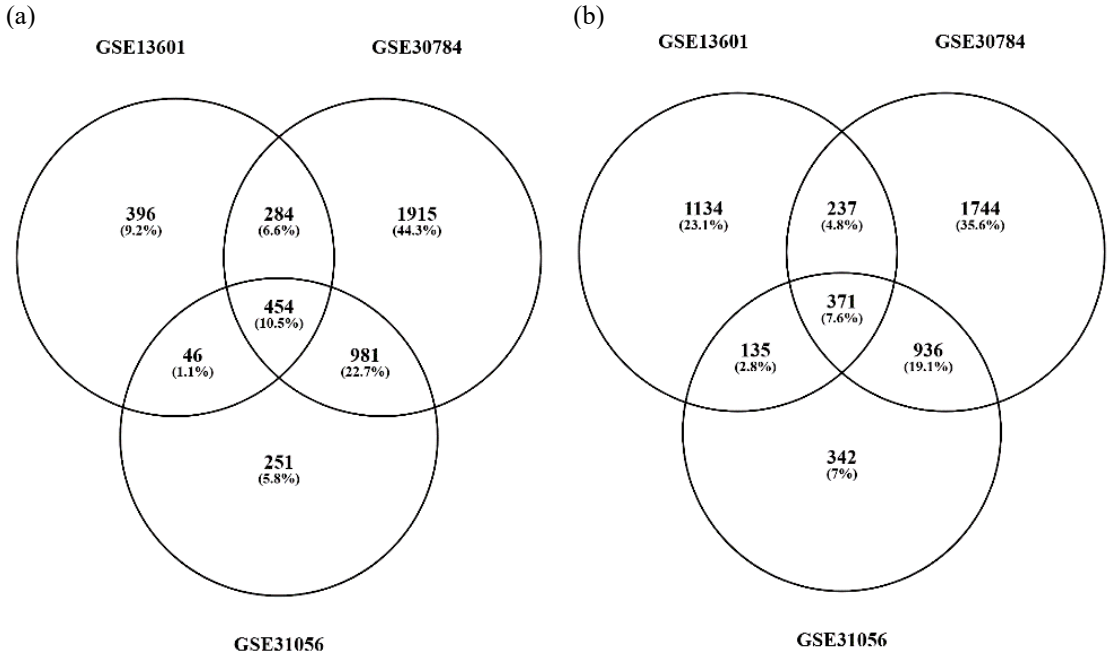


Figure 3. Venn diagrams illustrating the intersected **(a)** up-regulated genes and **(b)** down-regulated genes across the three GEO datasets. The Venn diagrams were created using the Venny web tool. The comprehensive list of these common genes found in all three datasets can be referenced in Supplementary Material 1.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Functional Enrichment Analysis

Figure 4 presents the top ten enriched biological processes, cellular components, molecular functions, and KEGG pathways. The degree of enrichment for each pathway was determined using a *p*-value, where a smaller value indicates a higher level of enrichment in the studied pathway. Detailed information can be found in Supplementary Material 2.

Several enriched GO terms were identified from the analysis of the 454 up-regulated genes in DAVID. In the biological process function group, the top enriched terms were the type I interferon signalling pathway, extracellular matrix (ECM) organisation, collagen catabolic process, and leukocyte migration, all of which are closely associated with the immune defence mechanism in cancer (Figure 4a). The enrichment analysis of cellular component GO terms revealed that up-regulated genes are involved in ECM and focal adhesion, highlighting their relation to the malignant progression of cancer (Figure 4b). Regarding molecular function, up-regulated genes were primarily involved in protein binding, ECM structural constituent, and integrin binding, all of which commonly contribute to tumourigenesis (Figure 4c). Notably, the protein binding GO term (Figure 4c) exhibited the highest gene count and rich factor among all the pathways listed in Figure 4(a-h), indicating that 71.37% of the up-regulated genes are associated with this molecular function. The KEGG enrichment analysis identified pathways such as the cell cycle, ECM-receptor interaction, and focal adhesion, providing further insights into the molecular mechanism underlying HNSCC (Figure 4d).

In contrast to the up-regulated genes, the down-regulated genes showed a stronger correlation with biological processes related to muscle filament sliding and muscle contraction (Figure 4e). Similarly, within the cellular component category, enriched GO terms such as extracellular exosome and muscle myosin complex were identified (Figure 4f). Regarding molecular function, the down-regulated genes were significantly associated with structural constituents of muscle and actin binding (Figure 4g). The enriched KEGG pathways for the down-regulated genes included metabolic pathways and valine, leucine, and isoleucine degradation (Figure 4h). Overall, the results from the DAVID analysis demonstrated a stronger association of up-regulated genes with cancer progression and metastasis compared to the down-regulated genes.

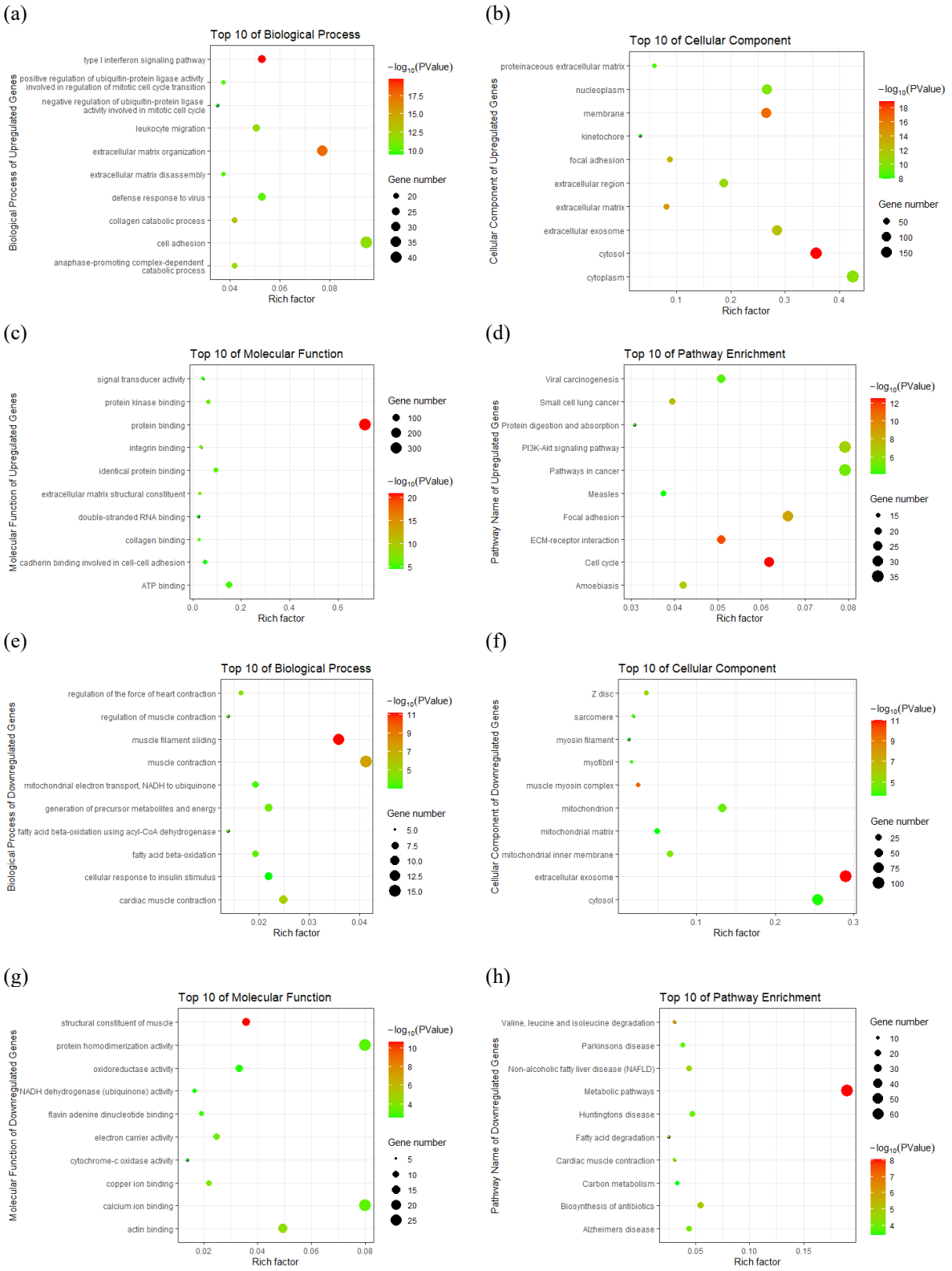
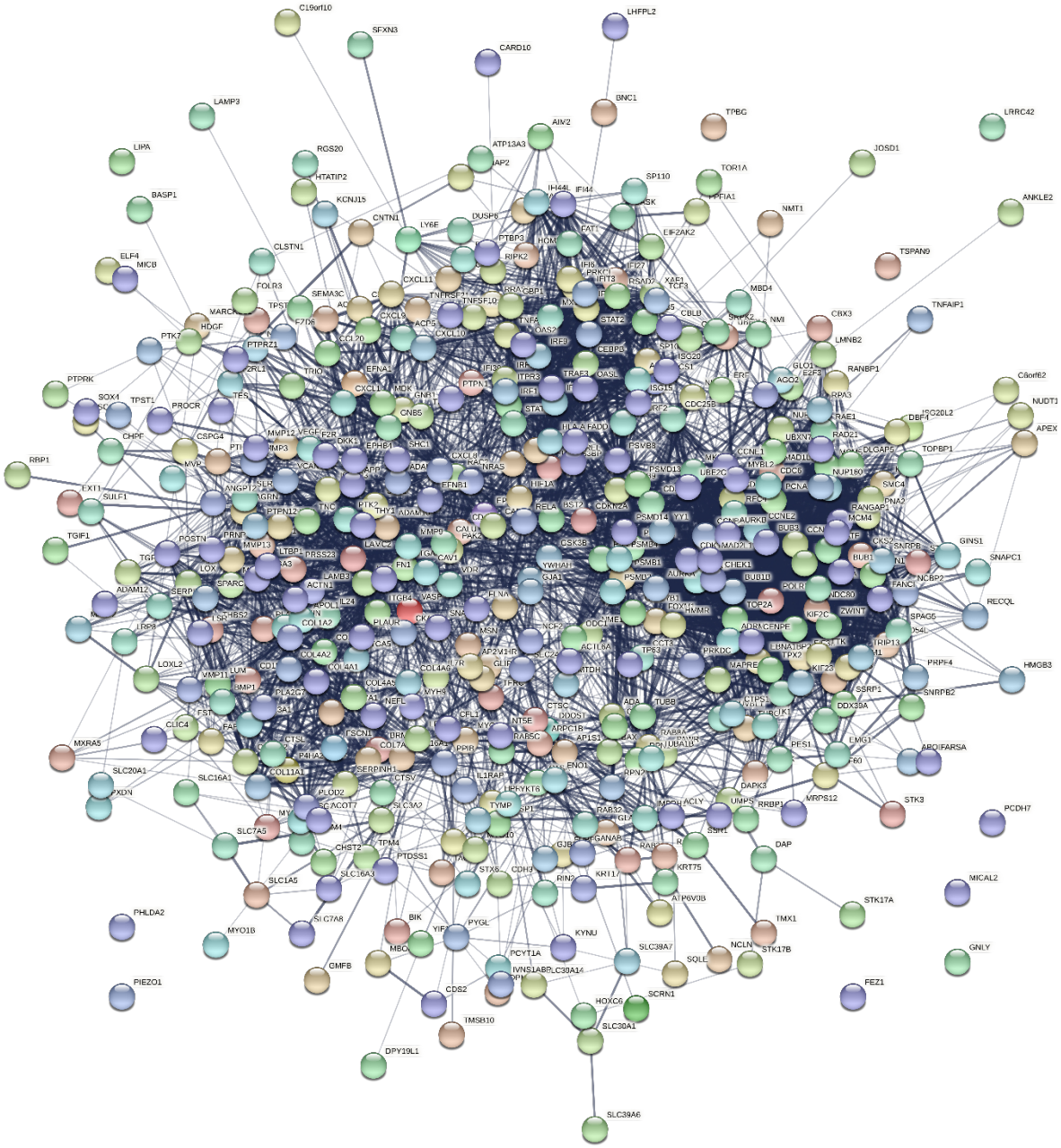


Figure 4. (a-d) The top ten enriched GO and KEGG pathways of the up-regulated and (e-h) down-regulated genes. The enriched pathways are categorised into biological process, cellular component, and molecular function for GO analysis, while KEGG enriched pathways are shown in (d) and (h). The plots were generated using the ggplot2 package in R statistical software.

Hub Genes Identification from PPI networks

The interactions between the translated proteins of 454 up-regulated and 371 down-regulated common genes within their respective expression groups were investigated using STRING. The PPI networks depicted in Figure 5 are highly interactive, with a higher density of nodes observed in the clustered up-regulated genes compared to the down-regulated genes.

(a)



(b)

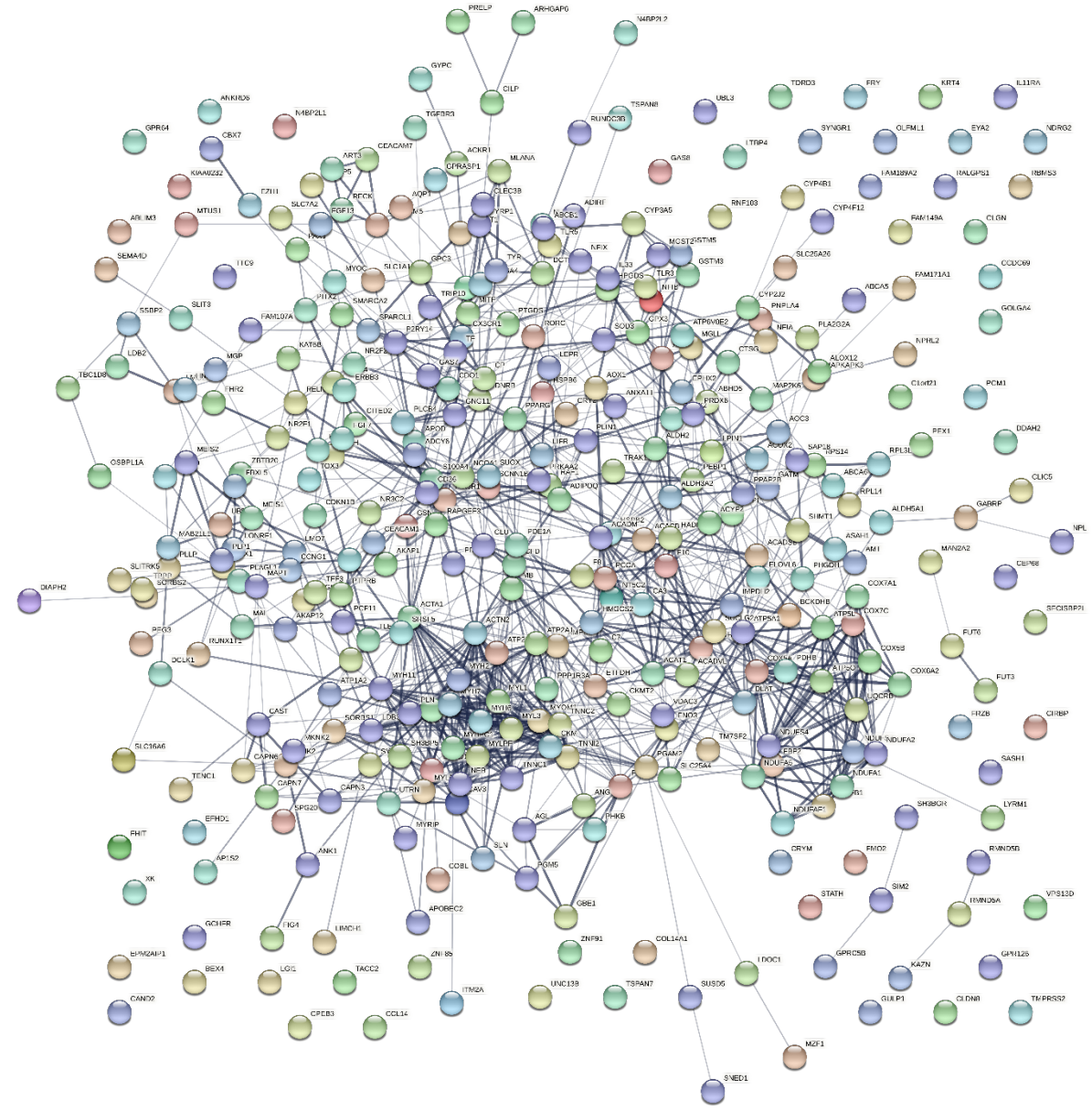


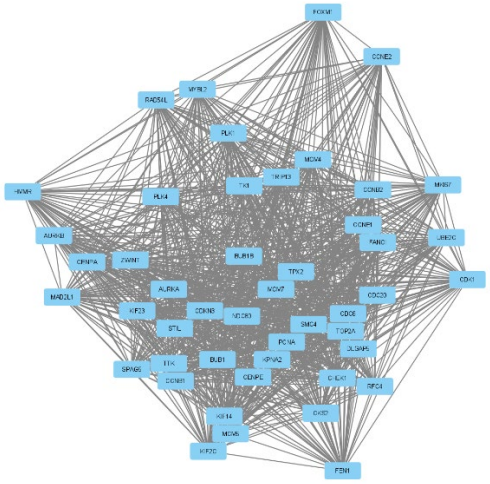
Figure 5. Protein-protein interaction (PPI) networks of **(a)** 454 up-regulated genes and **(b)** 371 down-regulated genes generated from the STRING database. The thickness of the lines indicates edge confidence, with a stronger connection between nodes represented by thicker connectors.

To focus on the PPI more specifically, MCODE was utilised to identify the top three highly interactive modules for the up-regulated and down-regulated gene groups, as depicted in Figure 5. Each PPI was assigned a score, indicating the connectivity level among the interacting proteins (refer to Supplementary Material 3). The top module of up-regulated genes exhibited a score of 44.39, comprising 47 nodes and 1021 edges (Figure 6a), followed by a score of 22.089, with 46 nodes and 497 edges (Figure 6b). Module three consisted of 27 nodes and 153 edges, with a score of 11.769 (Figure 6c). In comparison, the top module of down-regulated genes displayed a lower score of 17.5, encompassing 21 nodes and 175 edges (Figure 6e). The second-ranked module achieved a score of 13.571, with 15 nodes and 95 edges (Figure 6f), while module 3 comprised 11 nodes and 26 edges, attaining a score of 5.2 (Figure 6g).

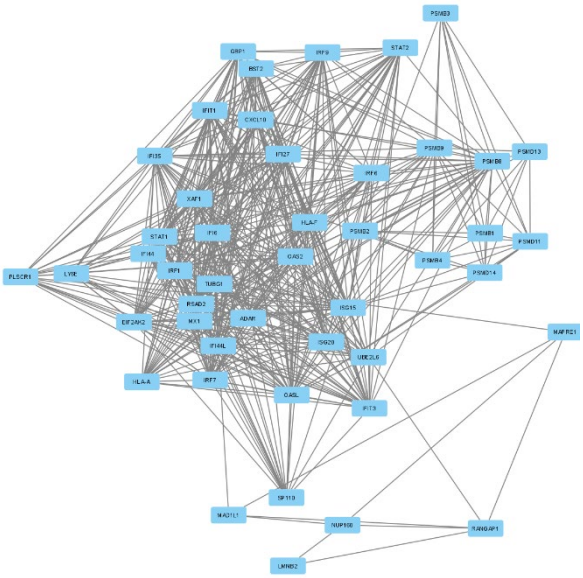
The subsequent selection of the top ten candidate genes from each up-regulated and down-regulated group was based on their degree of connectivity within their respective PPI networks, using cytoHubba. As illustrated in Figure 6d, the

most prominent up-regulated gene, represented by the highest intensity of colour ranging from yellow to red, corresponds to fibronectin-1 (*FN1)*, with a top score of 111. Among the ten up-regulated and ten down-regulated genes, the highest score observed for a down-regulated gene is 36 (Titin, *TTN*), which is considerably lower than the score of 75 achieved by the tenth-ranked up-regulated gene, matrix metallopeptidase 9 (*MMP9)*. This implies that the ten up-regulated genes are more likely to function as hub genes, producing proteins that exhibit more intensive interactions than the down-regulated genes.

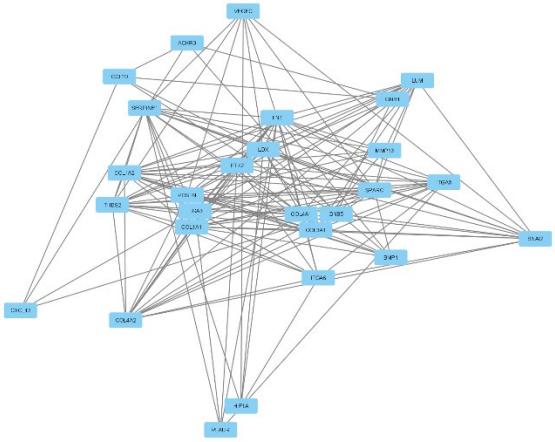
(a)



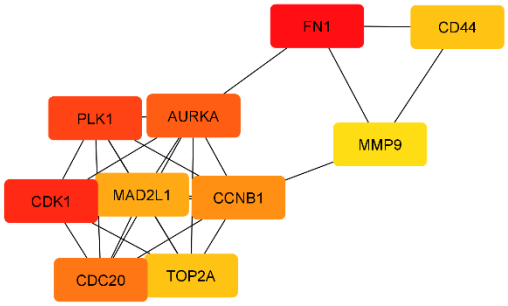
(b)



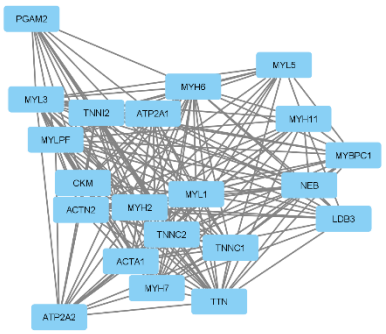
(c)



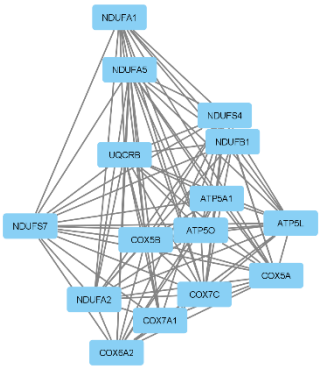
(d)



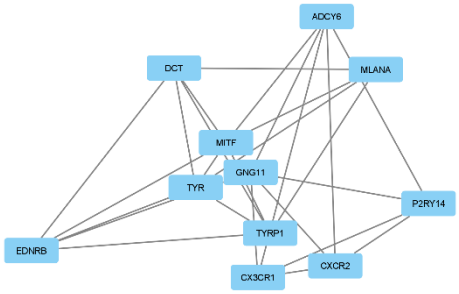
(e)



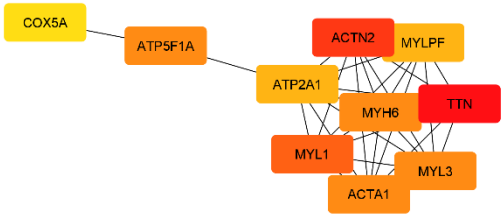
(f)



(g)



(h)



(i)

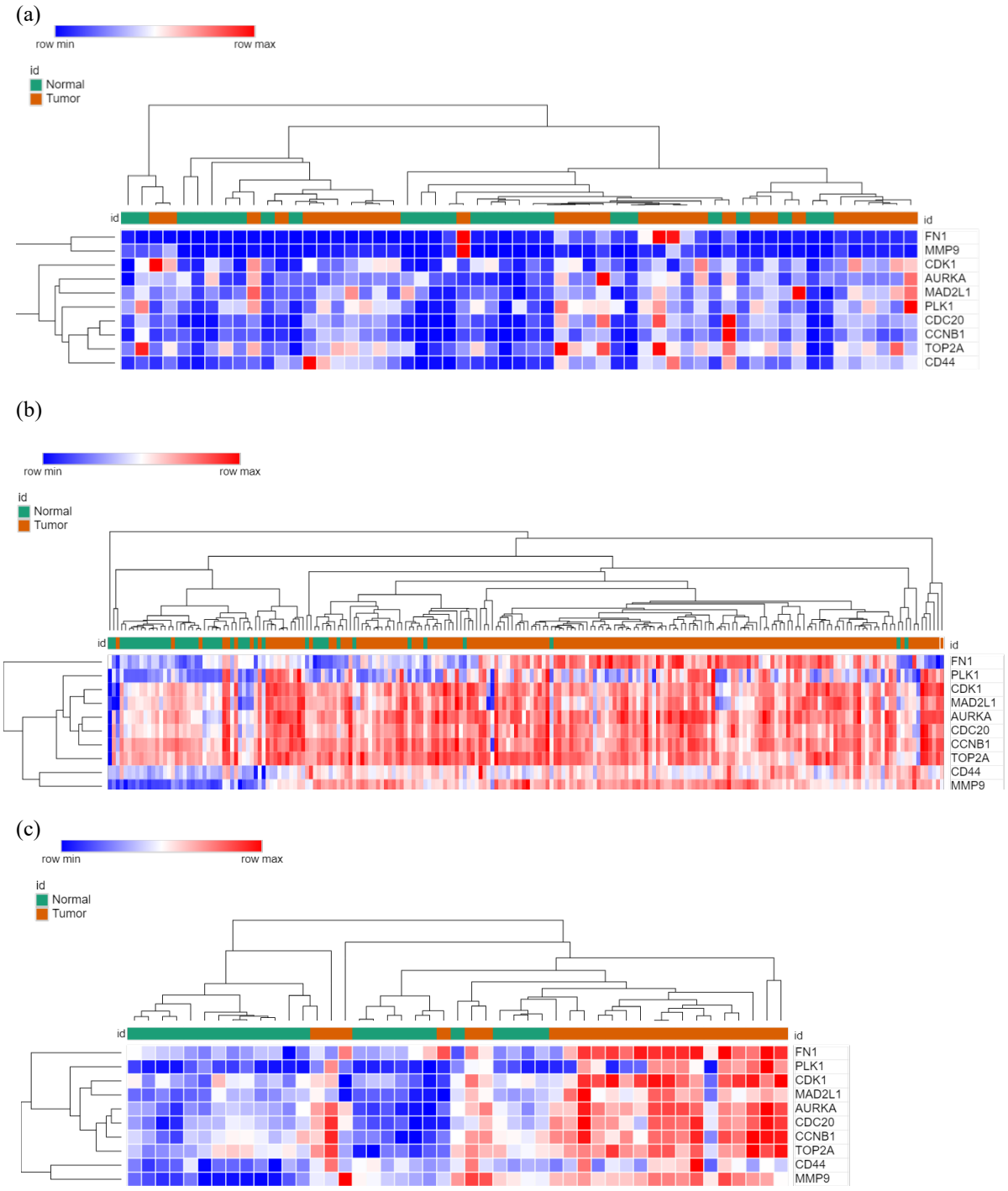
Rank	Name	Score
1	FN1	111
2	CDK1	92
3	PLK1	90
4	AURKA	85
5	CDC20	82
6	CCNB1	81
7	MAD2L1	78
8	CD44	76
8	TOP2A	76
10	MMP9	75

(j)

Rank	Name	Score
1	TTN	36
2	ACTN2	30
3	MYL1	28
4	ATP5F1A	27
4	ACTA1	27
4	MYL3	27
4	MYH6	27
8	ATP2A1	26
8	MYLPF	26
10	COX5A	25

Figure 6. Module analysis of PPI networks was conducted using Cytoscape, focusing on **(a-d)** up-regulated genes and **(e-h)** down-regulated genes. **(a-c; e-g)** The MCODE application was employed to identify the top three modules with the highest connectivity in each group. **(d, h)** The nodes' scores calculated from cytoHubba were used to assign colours and annotations to the top ten candidate DEGs. The intensity of colour ranges from yellow to red, indicating an increasing degree of connectivity, with dark red representing the gene with the highest score and light yellow representing the gene with the least connection among the ten hub genes in the PPI network. (i-j) The top ten (i) up-regulated and (j) down-regulated genes, ranked by their degree of connectivity, are displayed.

To examine the mRNA expression trends of the candidate genes, heatmaps were generated for both normal and OSCC samples (refer to Supplementary Material 4). Figure 7(a-f) illustrates the distinct patterns of expression observed for the up-regulated and down-regulated genes. In Figure 7(a-c), the ten up-regulated candidate genes consistently exhibited increased mRNA expression (indicated by red cells) in tumour samples across the three datasets. Similarly, Figure 7(d-f) displays decreased transcript expression (indicated by blue cells) of the down-regulated genes in tumour samples. These findings align with the classification of the experimental data into up-regulated and down-regulated groups, thereby validating the successful identification and categorisation of DEGs using the R script.



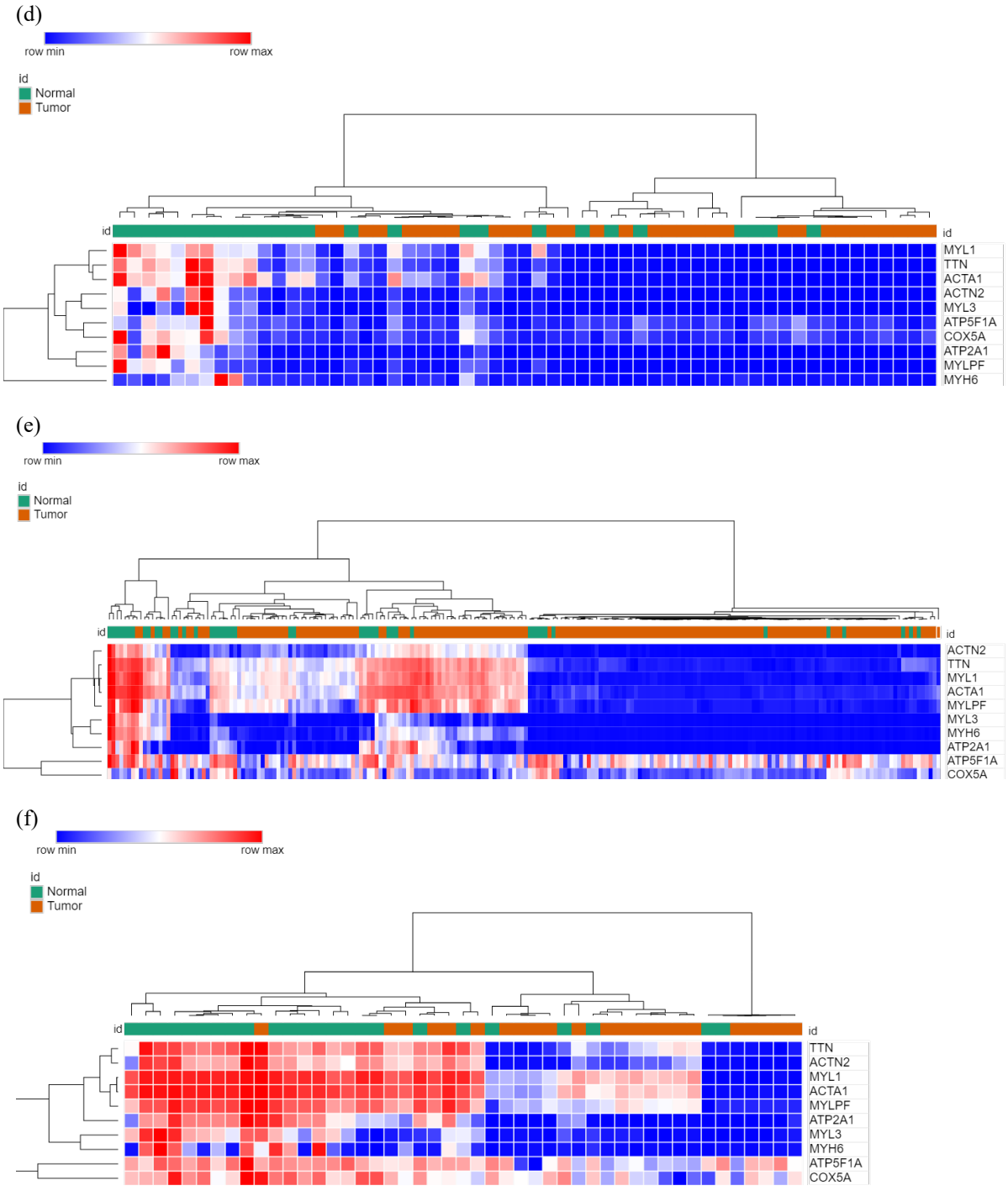


Figure 7. Heatmaps depicting the expression levels of the top ten up-regulated hub genes **(a-c)** and the ten down-regulated hub genes **(d-f)** were created using samples from the GSE13601, GSE30784, and GSE31056 datasets. In the heatmaps, red cells indicate samples where the corresponding genes were up-regulated, while blue cells indicate samples where the genes were down-regulated. The data used for generating the heatmaps were obtained from the GEO database, and the plots were created using the Morpheus online tool.

Transcriptional Expression of Hub Genes

The boxplots generated using expression data obtained from TCGA and GTEx databases corroborate the results obtained from the differential analysis. Genes classified as up-regulated exhibit higher expression levels in the tumour samples (Figure 8a), while down-regulated genes show a lower median in the tumour samples (Figure 8b). Except for *CD44, ATP5F1A, MYH6,* and *COX5A*, all the identified hub genes display significant up-regulation or down-regulation in HNSCC. This suggests that the differential expression of these hub genes in HNSCC samples may contribute to the progression and development of cancer. Furthermore, these findings indicate that the experimental pipeline employed in this study can identify genes with regulatory patterns consistent with those observed in TCGA and GTEx databases despite the databases being independent. Therefore, the DEGs obtained from GEO datasets with small sample sizes can reflect the gene regulation patterns observed in larger clinical databases.

Correlation of Cancer Survival to the Expression of Hub Genes

The relationship between the expression levels of the identified hub genes and the probability of survival over time in HNSCC patients was examined to determine their prognostic significance. Figure 9a demonstrates that higher expressions of *FN1, CDK1, PLK1, AURKA,* and *CD44* (up-regulated hub genes) are associated with statistically significant lower overall survival in HNSCC (p-value < 0.05). Similarly, Figure 9b shows that lower expression of *TTN, ACTN2, MYL1, ACTA1, MYL3,* and *MYLPF* (down-regulated hub genes) significantly correlates with lower overall survival in HNSCC. However, the expression levels of the remaining hub genes did not exhibit a statistically significant impact on the survival duration of the sampled HNSCC patients.

DISCUSSION

***In silico* Derivation of Hub Genes**

This study aimed to identify DEGs that could serve as highly predictive biomarkers for the prevention and therapeutic interventions of HNSCC. Given the global prevalence of HNSCC, identifying potential biomarkers is crucial for early prevention and developing therapeutic interventions. This study represents a preliminary step in addressing the lack of biomarkers in clinical HNSCC patients.

The study commenced with a meta-analysis of microarray gene expression to derive significantly up-regulated (n=454) and down-regulated (n=371) genes across all three GEO datasets (Figure 3). The functional properties of the up-regulated and down-regulated gene cohorts were then observed through GO and KEGG pathway enrichment analysis (Figure 4). Subsequently, PPI networks were constructed using STRING (Figure 5), and hub genes, consisting of ten up-regulated and ten down-regulated genes, were shortlisted from cytoHubba (Figure 6). These hub genes are hypothesised to be major drivers of HNSCC due to their strong correlation with other DEGs.

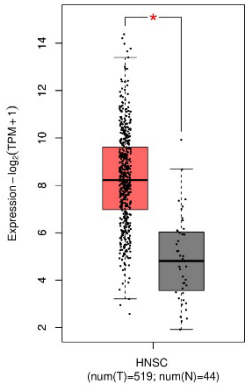
Furthermore, heatmaps generated using data from the GEO database revealed increased transcript levels in patient samples for the up-regulated genes and decreased transcript levels for the down-regulated genes across all three datasets (Figure 7). The correct classification of DEGs into the up-regulated or down-regulated groups was cross-verified by analysing their relative gene expressions with GEPIA2 (Figure 8). The findings demonstrate that the designed bioinformatics pipeline can identify gene candidates with distinguishable expression patterns in normal and cancer tissues.

To merit DEGs for future research, supporting the biological relevance of the shortlisted genes generated from bioinformatics analysis with reported findings is essential. Understanding the etiological role of each hub gene in HNSCC, such as its involvement in carcinogenesis, metastasis, and invasion at a molecular level, is crucial in inferring their potential as HNSCC biomarkers for prognosis, diagnosis, or both.

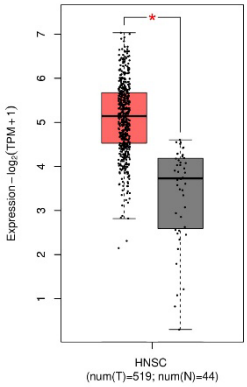
Further literature studies on the ten up-regulated hub genes *(FN1, CDK1, PLK1, AURKA, CDC20, CCNB1, MAD2L1, CD44, TOP2A, MMP9*) are consistent with the results, showing a positive correlation between their up-regulation and HNSCC. Based on survival analysis, half of the identified up-regulated genes, including *FN1, CDK1, PLK1, AURKA,* and *CD44* (refer to Figure 9a), are significantly associated with poorer survival in response to amplified expression. This further supports their potential role in HNSCC. However, regardless of their prognosis value, the study also identified the effect of gene expression for all genes in relation to HNSCC.

(a)

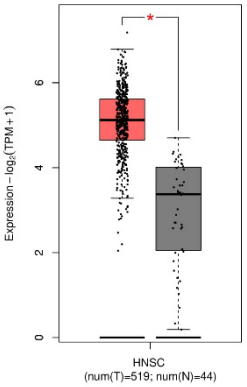
FN1



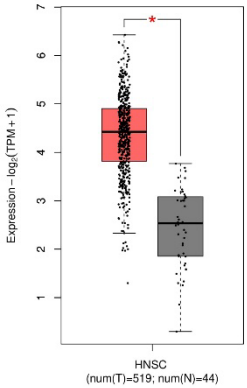
CDK1



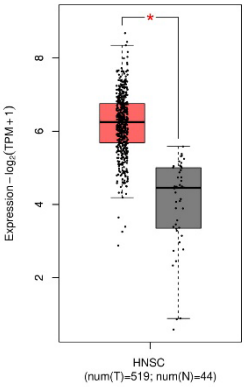
PLK1



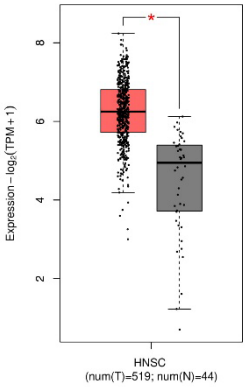
AURKA



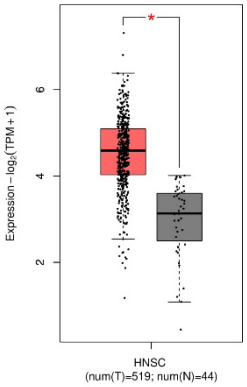
CDC20



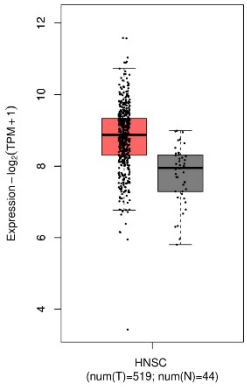
CCNB1



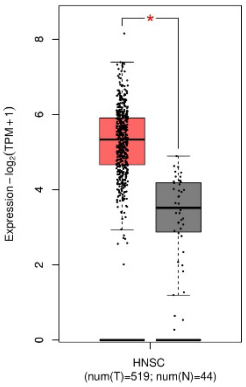
MAD2L1



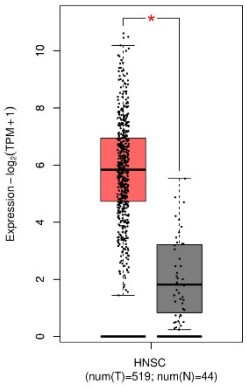
CD44



TOP2A



MMP9



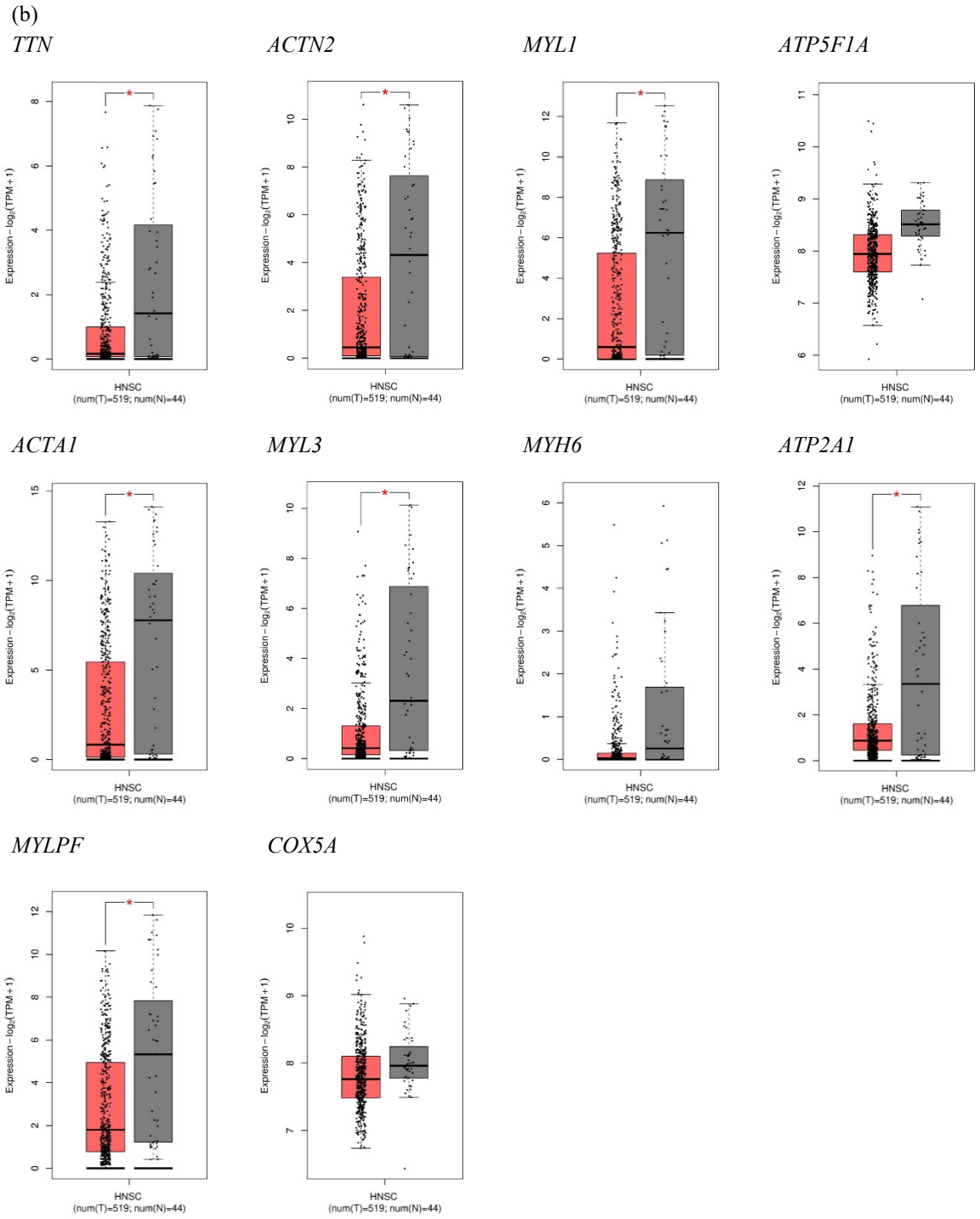
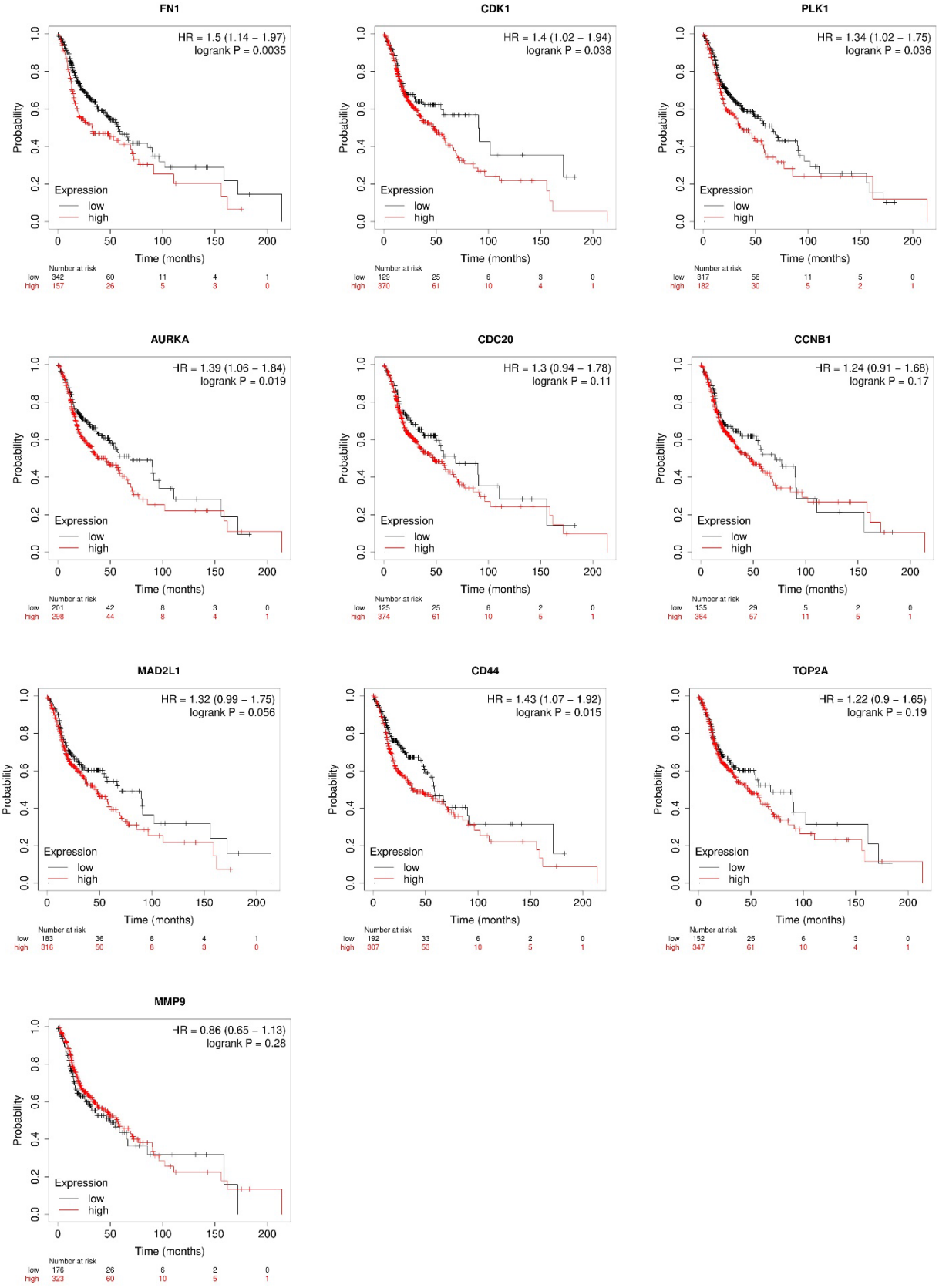


Figure 8. Expression levels of genes belonging to the (a) up-regulated and (b) down-regulated groups. Pink boxes represent patient samples diagnosed with HNSCC, while grey boxes indicate normal tissue samples. ATP5A1, the previous symbol for ATP5F1A, was used to search against the GEPIA2 web tool for analysis. Each of the identified DEGs was searched in GEPIA2, and the boxplots were generated with thresholds set at |log2 fold change| = 1 and p-value = 0.05. Significantly different expression levels with p-value < 0.05 are denoted by an asterisk (*).

(a)


FN1
HR = 1.5 (1.14 – 1.97)
logrank P = 0.0035
CDK1
HR = 1.4 (1.02 – 1.94)
logrank P = 0.038
PLK1
HR = 1.34 (1.02 – 1.75)
logrank P = 0.036
AURKA
HR = 1.39 (1.06 – 1.84)
logrank P = 0.019
CDC20
HR = 1.3 (0.94 – 1.78)
logrank P = 0.11
CCNB1
HR = 1.24 (0.91 – 1.68)
logrank P = 0.17
MAD2L1
HR = 1.32 (0.99 – 1.75)
logrank P = 0.056
CD44
HR = 1.43 (1.07 – 1.92)
logrank P = 0.015
TOP2A
HR = 1.22 (0.9 – 1.65)
logrank P = 0.19
MMP9
HR = 0.86 (0.65 – 1.13)
logrank P = 0.28
Probability
Expression
low
high
Time (months)
Number at risk

(b)

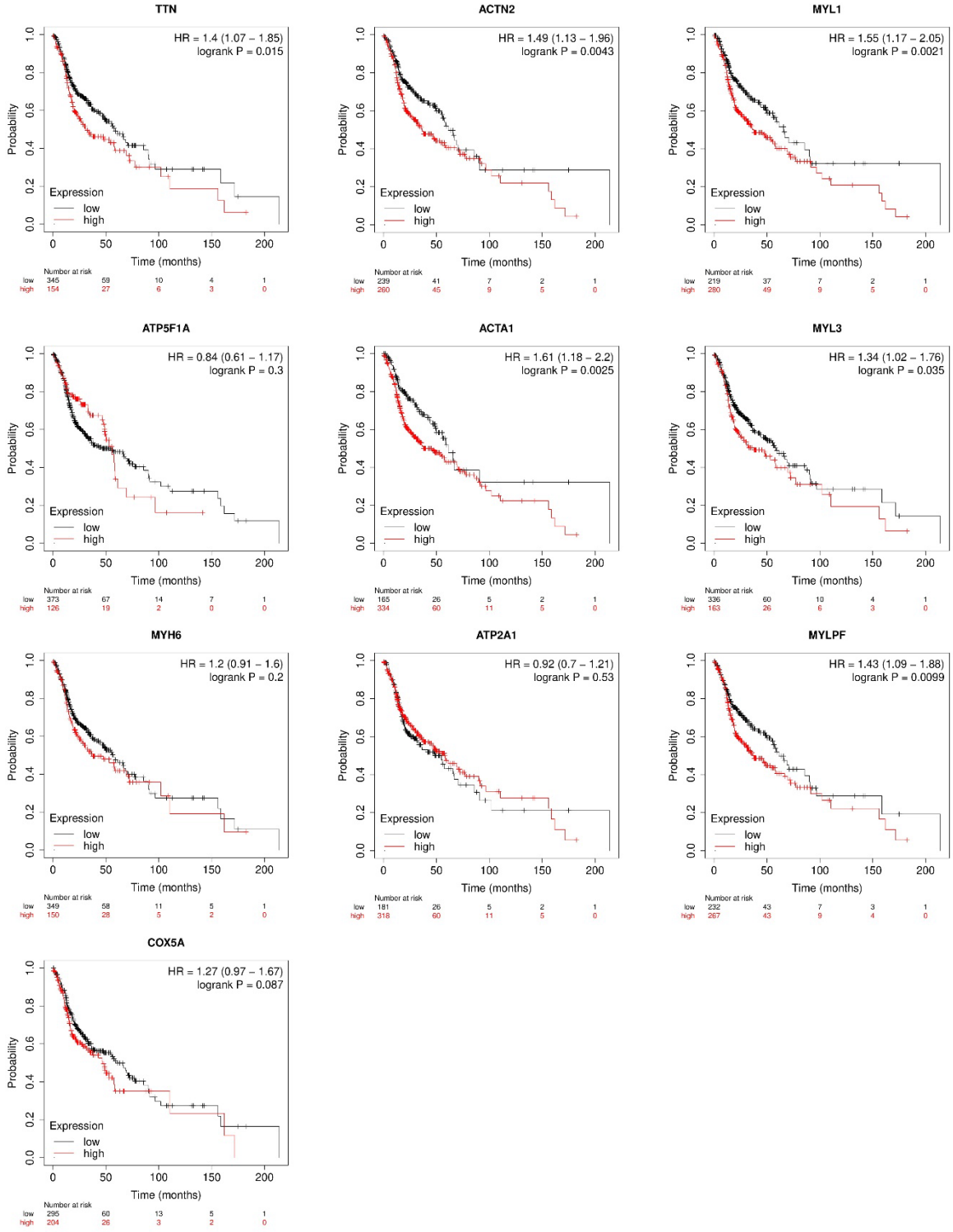


Figure 9. The prognostic significance of the identified up-regulated and down-regulated genes in HNSCC was assessed using Kaplan-Meier plots depicting the overall survival of HNSCC patients. Figure 9a displays the Kaplan-Meier plot for the up-regulated genes, while Figure 9b shows the plot for the down-regulated genes. Each of the identified DEGs was analysed using the Kaplan-Meier Plotter web application, utilizing the built-in RNA-seq data of HNSCC. Curves with a *p*-value < 0.05 were deemed statistically significant, indicating that changes in gene expression could influence the survival of HNSCC patients.

Understanding the Roles of DEGs in Relation to Cancer

The association of FN1 with adverse effects on HNSCC is evident due to its role as an ECM protein. *FN1* encodes soluble plasma and insoluble cellular forms of FN1, with the insoluble cellular FN1 being a glycoprotein involved in cellular adhesion [20]. The expression of *FN1* can lead to epithelial-mesenchymal transition (EMT) of tissue, characterised by the loss of E-cadherin and activation of Vimentin [46]. Consequently, epithelial cells lose their phenotypic and high differentiating characteristics while acquiring features of re-differentiation, such as stem cell characteristics, apoptosis resistance, migratory abilities, and invasiveness, which facilitate tumour invasion and metastasis [47]. In a study on HNSCC biomarkers, FN1 was found to be degraded by the p62-dependent autophagy-lysosome pathway. In another study, FN1 was implicated in regulating apoptosis through the NF-κB/P65 pathway in nasopharyngeal carcinoma (NPC) [46, 48]. Increased mRNA expression of *FN1* has also been reported in many cancers, including renal [49], thyroid [50], and colorectal cancer [51].

The role of FN1 as a positive mediator in cancer is also demonstrated by its ability to induce metalloproteinases, promoting invasion and metastasis. Among the members of the metalloproteinase family, MMP9 is identified as one of the up-regulated genes in our study, consistent with its detection in the immunohistochemical analysis of HNSCC [52]. MMP9 is a gelatinase, specifically a type IV collagenase, that degrades collagen Type IV (ColIV), a principal component of the ECM [53]. During tumour invasion, metastasis, and angiogenesis, secretion of MMPs by stromal cells can dynamically manipulate the surrounding microenvironment to facilitate tumour spread [54]. An increase in MMP9 is detected in tumour epithelium and stroma upon tumour progression, leading to the gradual fragmentation of the basement membrane, thereby allowing the invasion of lamina propria in oral tongue squamous cell carcinoma (OTSCC) [55]. This suggests a close association between malignancy progression and the expression of MMP9, where the extensive dissolution of ColIV can lead to morphological changes in the basement membrane and surrounding tissue, and tumour angiogenesis. Therefore, elevated expression of MMPs in stromal cells is related to lymphatic metastasis, resulting in poorer survival rates for cancer patients [55]. Given the detection of MMP9 in the serum of HNSCC patients, it is suggested that MMP9 could serve as a marker for the diagnosis and early detection of HNSCC [56].

Matrix metalloproteinase can also serve as ligands for CD44, a cell surface adhesion receptor. The binding of CD44 to various ECM ligands, including hyaluronan, osteopontin (OPN), and collagens, modulates the proliferation, invasion, and migration processes of tumour cells through signalling pathways [57, 58]. CD44 is a multifunctional transmembrane glycoprotein, and its various alternatively spliced variants are correlated with tumour subtypes and often used as markers of cancer stem cells (CSCs) that promote tumour progression. Strong overexpression of *CD44* detected in high-grade HNSCC is correlated with the enrichment of pro-angiogenic factors, stimulating early angiogenesis and promoting HNSCC progression by facilitating the formation of new blood vessels to nourish tumour growth [59, 60, 61, 62]. In addition to its resemblance to the disease recurrence properties of CSCs, the elevation of CD44 levels in the serum of patients makes it an ideal marker of metastasis, potentially for HNSCC and other cancer types [57].

Furthermore, our study also detected up-regulation of genes essential in cell cycle regulation. Cyclin-dependent kinase 1 (CDK1) activation by cyclin A2 and cyclin B during the G2 phase is required to enter the mitotic event. The formation of the CDK1 complex with B-type cyclins upon the onset of mitosis emphasises the importance of cyclin B1 (CCNB1) in controlling the G2-M transition of the cell cycle [63]. Up-regulation of *CCNB1* expression, which plays a role in tumour proliferation, has been reported in squamous cell carcinoma (SCC) of the lung and tongue and is also clinically relevant to histological grade in OSCC [64, 65].

During nuclear translocation, polo-like kinase 1 (PLK1) phosphorylates CCNB1, crucial for activating cyclin B-CDK1 complexes. In addition to suppressing pro-apoptotic function through the phosphorylation of p53, PLK1 can compromise cell cycle checkpoints and induce genetic instability, leading to tumourigenesis [66, 67]. Its importance in CDK1 reactivation in response to DNA damage recovery, and its role in activating chromosome segregation during the metaphase-anaphase transition, highlights its potential for the exploitation of cancer therapy [68].

PLK1 also interacts with another identified gene, aurora kinase A (AURKA), during centrosome maturation in the G2 phase [69]. AURKA is a serine/threonine kinase that can activate PLK1 and CDK1 through phosphorylation, thereby guarding the mitotic entry and G2 arrest upon DNA damage [70, 71]. Overexpression of AURKA has been observed to increase the frequency of tetraploidy and centrosome amplification, along with defective p53-dependent DNA damage checkpoints, which could lead to the formation of mammary cancer [72, 73]. This arises from the simultaneous inactivation of both the spindle assembly checkpoint (during mitosis) and the DNA damage checkpoint (during the G2 phase) [74, 75]. Consistent with our bioinformatics analysis, elevated levels of AURKA in cancer have also been reported in studies of HNSCC [76] and hepatocellular carcinoma [77].

Similar to the findings above, other genes related to the mitotic spindle play a significant role in tumourigenesis. Overexpression of cell division cycle 20 (CDC20), a spindle checkpoint protein, can cause premature anaphase, resulting in aneuploidy in tumour cells [78]. Increased *CDC20* expression has been correlated with aneuploidy in OSCC, leading to reduced oral cancer-specific survival and suggesting its potential as an independent therapeutic target

[79, 80]. On the other hand, overexpression of mitotic arrest deficient 2 like 1 (MAD2L1), which functions in the mitotic spindle assembly checkpoint, negatively affects chromosomal stability by hyper-stabilising kinetochore microtubules [81]. The CDC20 and MAD2L1 cellular levels balance is crucial in regulating the anaphase-promoting complex/cyclosome (APC/C). MAD2L1 is required to counteract the CDC20-dependent activation of APC/C, halting anaphase initiation when the activated spindle assembly complex (SAC) detects unattached kinetochores. High expression of *CDC20* and *MAD2L1* is associated with aneuploidy in malignant tumours and worse prognosis in oral cancer, non-small cell lung cancer, gastric cancer, pancreatic ductal adenocarcinoma, and hepatocellular carcinoma [80, 82, 83, 84, 85].

DNA topoisomerase II alpha (TOP2A) is a nuclear enzyme that controls transcription and DNA replication and is involved in modifying DNA topologic states [86]. TOP2A plays a role in chromosome condensation, re-ligation of DNA strands, and is a marker for proliferating cells [87, 88]. Its up-regulation has been implicated in several tumours types, including NPC [87, 89, 90]. Importantly, TOP2A is an established molecular target for anthracyclin-based chemotherapy and is clinically relevant in the context of anticancer drugs, such as topoisomerase inhibitors like etoposide [91, 92, 93].

In this study, GO and the KEGG pathway analyses revealed a closer relationship of up-regulated genes to cancer than down-regulated genes (Figure 4). Figure 4(a-d) shows a higher enrichment of EMT and immune system-related pathways among the up-regulated genes, which play a crucial role in the function of cancer cells. According to Wachi et al. (2005), a study on DEGs in lung squamous cell carcinoma (LUSC), the human PPI networks of up-regulated genes tend to exhibit higher network connectivity and centrality in cancerous tissues. These common topological characteristics shared among DEGs in cancerous tissue suggest that up-regulated genes are essential for cancer cell proliferation.

In contrast to the up-regulated genes, the ten down-regulated hub genes (*TTN, ACTN2, MYL1, ATP5F1A, ACTA1, MYL3, MYH6, ATP2A1, MYLPF, COX5A*) selected in our analysis have not been extensively characterised in the context of cancer, and the available findings directly associating these genes with HNSCC are limited. However, the relevance of genes encoding for myosin (*MYL1, MYL3, MYLPF, MYH6)* to HNSCC can be established. Consistent with our findings of myosin light chain 1 (*MYL1*) as a down-regulated gene, *MYL1* down-regulation has also been observed in the bioinformatics analysis of buccal cancer, which was cross-validated with qPCR analysis of cancer tissue [95]. In another study on OSCC, *MYL1* and *MYL3* were down-regulated, and MYL3 down-regulation was also identified in OTSCC [96, 97]. Although the molecular pathogenesis of myosin light chains (MLC) is yet to be elucidated, He et al. (2004) suggested that the decreased expression of MLC reflects muscle destruction in tongue tumour tissue during cancer invasion and migration. The study of myosin light chain, phosphorylatable, fast skeletal muscle (*MYLPF*), which also showed down-regulation in the same study of HNSCC, further suggests its possible role in remodelling muscle function in HNSCC tissues [98]. The identification of myosin heavy chain 6 (*MYH6*) as a hub gene in HNSCC samples indicates its unique function in disease development, independent of tobacco smoking [99]. This tumour microenvironment-related gene predicts tumour progression and patient survival, thus considered a novel putative oncogene in HNSCC [100, 101].

Consistent with our findings, actin alpha 1, skeletal muscle (*ACTA1*) is also identified as a down-regulated gene that plays a role in cell integrity, structure, and mobility, highlighting its relevance to aggressiveness and clinical malignancy in oral epithelial dysplasia and HNSCC. *ACTA1* down-regulation has also been observed in colorectal cancer [98, 102]. The down-regulation of *ACTA1,* along with the co-identification of *MYL1* in the same study of buccal cancer and *MYL3* in OSCC, suggests the reduced activity of actin and myosin in the remodelling of tumour tissue, particularly in HNSCC [95, 96]. Frequent mutation of *TTN*, another identified gene encoding a striated muscle protein, has been reported in lung squamous cell carcinoma, lung, colon, and breast cancer [103]. *TTN* has also been identified as a hub gene, along with actinin alpha 2 *(ACTN2),* in HNSCC [99]. Genes involved in muscle contraction (such as actin, myosin, and TTN) could be constituents of stromal myofibroblasts, which are associated with invasion and metastasis in OSCC [104, 105, 106, 107].

There is a less established correlation between other identified genes with HNSCC. *ACTN2* has been reported to be overexpressed in hepatocellular carcinoma, whereas the hypermethylation of *ATP2A1* in oropharyngeal cancer indicates reduced disease-free survival [108, 109]. Low expression of *ATP5F1A* may promote tumour development with microsatellite instability (MSI), while high expression of ATP5F1A is associated with glioblastoma and earlier-onset prostate cancer [110, 111]. A limited study on another identified gene, cytochrome C oxidase subunit 5A *(COX5A)*, suggests its involvement in the migration and invasion of non-small cell lung carcinoma (NSCLC), highlighting its relevance in the tumorigenesis of LUSC [112, 113].

Although the down-regulated genes identified in our bioinformatics analysis are consistent with the down-regulation of mRNA levels in TCGA HNSCC samples (n = 519) validated with GEPIA2 (Figure 8b), the results from the Kaplan-Meier plot suggest better overall survival for higher expression of the down-regulated genes identified in this study (Figure 9b). Using *MYLPF* as an example, its prognostic value is significantly increased in hypopharyngeal, laryngeal, and HNSCC when mRNA expression is higher [114]. This finding contradicts the results of the current study, where lower expression of *MYLPF* in TCGA HNSCC showed better overall survival

with a *p*-value = 0.0099. Similar contradicting trends, associating better overall survival with higher expression of *ACTA1* and *MYL1,* have also been reported in HNSCC [98]. Such conflicts might arise from differences in the sociodemographic characteristics of patients, suggesting the possibility of stratifying HNSCC patients based on these unique molecular biomarkers [20]. However, further research is needed to accurately predict the prognostic value of the down-regulated genes to specific clinical parameters, as the tumorigenesis of these genes is yet to be fully understood.

Limitations and Future Prospects

In this study, RNA-seq data of HNSCC profiles were used in bioinformatics tools such as GEPIA2 and Kaplan-Meier Plotter, and the reported literature primarily relied on data sourced from the TCGA database. It is important to note that the genetic landscape of non-Caucasian cohorts in the TCGA database is limited, with only 3.86% of the head and neck cancer archives belonging to Asians [115, 116]. Therefore, there is a need to increase the sample size and include a broader spectrum of the world's population to improve data representation and enhance the credibility of the results.

Another limitation of this study is related to the technical constraints of constructing the PPI network. The construction of the PPI network relies on experimental data and literature-based information catalogued in protein databases. This dependence on established data could introduce bias in identifying hub genes, as proteins that have been extensively studied may appear to have a stronger correlation to biological relevance, potentially misleading the identification process [117, 118].

While *in silico* analysis is beneficial for identifying potential hub genes with crucial roles in HNSCC, it is important to complement this study with an experimental-based approach. After the discovery phase, the mRNA expression of DEGs in HNSCC samples can be validated using real-time polymerase chain reaction (RT-PCR). Additionally, the protein expression and localisation of DEGs in HNSCC tissues can be analysed using immunohistochemical staining, and these results can be compared with the findings obtained through bioinformatics analysis [16]. Validated DEGs have the potential to serve as biomarkers for predicting cancer recurrence, disease progression, and individual responses to therapeutic interventions, such as tumour resistance to radiotherapy [15, 20]. Furthermore, research can be expanded to understand the relevance between the dysregulation of DEGs in HNSCC and the modulation of the tumour microenvironment by integrating CRISPR technology, which can be useful in designing DNA vaccines for immunotherapy [119, 120].

CONCLUSION

The network-based meta-analysis pipeline applied to HNSCC datasets has effectively revealed a strong association between the identified hub genes and the underlying molecular mechanisms of cancer. This study's findings underscore the potential significance of eleven genes in the context of HNSCC. Among these, the six down-regulated genes (*MYL1, MYL3, MYH6, MYLPF, ACTA1, TTN*) and five up-regulated genes (*CDC20, CCNB1, MAD2L1, TOP2A, MMP9*) exhibit promise as biomarkers for HNSCC diagnosis. Furthermore, an additional set of five up-regulated genes (*FN1, CDK1, PLK1, AURKA, CD44*) could serve as valuable biomarkers for prognosis and diagnosis in clinical HNSCC analyses. These differentially expressed genes warrant in-depth investigation to characterise their precise molecular functions and role in the pathogenesis of HNSCC. The integrated bioinformatics approach, utilizing transcriptomic data, has demonstrated its robustness as a methodology for exploring potential biomarkers applicable across various cancer types. This work enhances our understanding of HNSCC and opens avenues for broader applications in cancer research.

ACKNOWLEDGEMENTS

The authors would like to thank the academics from the University of Nottingham Malaysia School of Biosciences who had suggested improvements on this research study during the final year project symposium.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests regarding the publication of this manuscript. Data curation and manuscript writing were performed by H.S.S. The study was conceived and supervised by B.K.B.L. and H.-S.L. The initial manuscript was reviewed and edited by B.K.B.L. and H.-S.L. The revised version was critically reviewed by B.K.B.L.

REFERENCES

1. Barabási, A.-L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**(2), 101-113.
2. Cieślik, M. and Chinnaiyan, A.M. (2018) Cancer transcriptome profiling at the juncture of clinical translation. *Nat. Rev. Genet.* **19**(2), 93-109.
3. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S. and Soboleva, A. (2012) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**(D1), D991-D995.

4. Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. (Pozn)*. **19**(1A), A68-A77.

5. Jha, P.K., Vijay, A., Sahu, A. and Ashraf, M.Z. (2016) Comprehensive gene expression meta-analysis and integrated bioinformatic approaches reveal shared signatures between thrombosis and myeloproliferative disorders. *Sci. Rep.* **6**(1), 37099.

6. Pierre, M., DeHertogh, B., Gaigneaux, A., DeMeulder, B., Berger, F., Bareke, E., Michiels, C. and Depiereux, E. (2010) Meta-analysis of archived DNA microarrays identifies genes regulated by hypoxia and involved in a metastatic phenotype in cancer cells. *BMC Cancer* **10**(1), 176.

7. Maharjan, M., Tanvir, R.B., Chowdhury, K., Duan, W. and Mondal, A.M. (2020) Computational identification of biomarker genes for lung cancer considering treatment and non-treatment studies. *BMC Bioinform.* **21**(9), 218.

8. Ma, Q., Xu, Y., Liao, H., Cai, Y., Xu, L., Xiao, D., Liu, C., Pu, W., Zhong, X. and Guo, X. (2019) Identification and validation of key genes associated with non-small-cell lung cancer. *J. Cell. Physiol.* **234**(12), 22742-22752.

9. Shen, S., Kong, J., Qiu, Y., Yang, X., Wang, W. and Yan, L. (2019) Identification of core genes and outcomes in hepatocellular carcinoma by bioinformatics analysis. *J. Cell. Biochem.* **120**(6), 10069-10081.

10. Ideker, T. and Sharan, R. (2008) Protein networks in disease. *Genome Res.* **18**(4), 644-652.

11. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G.M., Blake, J.A., Bult, C., Dolan, M., Drabkin, H., Eppig, J.T., Hill, D.P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J.M., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R.S., Sethuraman, A., Theesfeld, C.L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S.Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E.M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T. and White, R. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**(D1), D258-D261.

12. Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., Vidal, M. and Barabási, A.-L. (2007) The human disease network. *PNAS*. **104**(21), 8685-8690.

13. Shim, J.E. and Lee, I. (2015) Network-assisted approaches for human disease research. *Anim. Cells Syst.* **19**(4), 231-235.

14. Oti, M., Snel, B., Huynen, M.A. and Brunner, H.G. (2006) Predicting disease genes using protein-protein interactions. *J. Med. Genet*. **43**(8), 691-698.

15. Goossens, N., Nakagawa, S., Sun, X. and Hoshida, Y. (2015) Cancer biomarker discovery and validation. *Transl. Cancer Res.* **4**(3), 256-269.

16. He, Z., Tang, F., Lu, Z., Huang, Y., Lei, H., Li, Z. and Zeng, G. (2018) Analysis of differentially expressed genes, clinical value and biological pathways in prostate cancer. *Am. J. Transl. Res.* **10**(5), 1444-1456.

17. Deng, J.-L., Xu, Y.-H. and Wang, G. (2019) Identification of potential crucial genes and key pathways in breast cancer using bioinformatic analysis. *Front. Genet.* **10**, 695.

18. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J. Clin.* **71**(3), 209-249.

19. 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. Genomic Med.* **7**(8), e857.

20. Qadir, F., Lalli, A., Dar, H.H., Hwang, S., Aldehlawi, H., Ma, H., Dai, H., Waseem, A. and Teh, M.-T. (2019) Clinical correlation of opposing molecular signatures in head and neck squamous cell carcinoma. *BMC Cancer* **19**(1), 830.

21. Rivera, C. (2015) Essentials of oral cancer. *Int. J. Clin. Exp. Pathol.* **8**(9), 11884-11894.

22. Sasahira, T., Bosserhoff, A.K. and Kirita, T. (2018) The importance of melanoma inhibitory activity gene family in the tumor progression of oral cancer. *Pathol. Int.* **68**(5), 278-286.

23. Sasahira, T. and Kirita, T. (2018) Hallmarks of cancer-related newly prognostic factors of oral squamous cell carcinoma. *Int. J. Mol. Sci.* **19**(8), 2413.

24. Kim, J.W., Park, Y., Roh, J.-L., Cho, K.-J., Choi, S.-H., Nam, S.Y. and Kim, S.Y. (2016) Prognostic value of glucosylceramide synthase and P-glycoprotein expression in oral cavity cancer. *Int. J. Clin. Oncol.* **21**(5), 883-889.

25. Duprez, F., Berwouts, D., De Neve, W., Bonte, K., Boterberg, T., Deron, P., Huvenne, W., Rottey, S. and Mareel, M. (2017) Distant metastases in head and neck cancer. *Head Neck* **39**(9), 1733-1743.

26. Xiao, Y., Hsiao, T.H., Suresh, U., Chen, H.I., Wu, X., Wolf, S.E. and Chen, Y. (2014) A novel significance score for gene selection and ranking. *Bioinformatics* **30**(6), 801-807.

27. Estilo, C.L., O-charoenrat, P., Talbot, S., Socci, N.D., Carlson, D.L., Ghossein, R., Williams, T., Yonekawa, Y., Ramanathan, Y., Boyle, J.O., Kraus, D.H., Patel, S., Shaha, A.R., Wong, R.J., Huryn, J.M., Shah, J.P. and Singh, B. (2009) Oral tongue cancer gene expression profiling: Identification of novel potential prognosticators by oligonucleotide microarray analysis. *BMC Cancer* **9**, 11.

28. Chen, C., Méndez, E., Houck, J., Fan, W., Lohavanichbutr, P., Doody, D., Yueh, B., Futran, N.D., Upton, M., Farwell, D.G., Schwartz, S.M. and Zhao, L.P. (2008) Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **17**(8), 2152-2162.

29. Reis, P.P., Waldron, L., Perez-Ordonez, B., Pintilie, M., Galloni, N.N., Xuan, Y., Cervigne, N.K., Warner, G.C., Makitie, A.A., Simpson, C., Goldstein, D., Brown, D., Gilbert, R., Gullane, P., Irish, J., Jurisica, I. and Kamel-Reid, S. (2011) A gene signature in histologically normal surgical margins is predictive of oral carcinoma recurrence. *BMC Cancer* **11**, 437.

30. Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamayo, P., Spiegelman, B., Lander, E.S., Hirschhorn, J.N., Altshuler, D. and Groop, L.C. (2003) PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**(3), 267-273.

31. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS.* **102**(43), 15545-15550.

32. Wu, C., Jin, X., Tsueng, G., Afrasiabi, C. and Su, A.I. (2016) BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res.* **44**(D1), D313-D316.

33. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C. and Chanda, S.K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**(1), 1523.

34. Tweedie, S., Braschi, B., Gray, K., Jones, T.E.M., Seal, Ruth L., Yates, B. and Bruford, E.A. (2020) Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* **49**(D1), D939-D946.

35. Oliveros, J.C. (2015). Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html

36. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**(1), 44-57.

37. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**(1), 1-13.

38. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2016) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**(D1), D353-D361.

39. The Gene Ontology Consortium. (2014) Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**(D1), D1049-D1056.

40. Wickham, H. (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York.

41. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J. and Mering, C.V. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**(D1), D607-D613.

42. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498-2504.

43. 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**(Suppl 4), S11.

44. Tang, Z., Kang, B., Li, C., Chen, T. and Zhang, Z. (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**(W1), W556-W560.

45. Nagy, Á., Munkácsy, G. and Győrffy, B. (2021) Pancancer survival analysis of cancer hallmark genes. *Sci. Rep.* **11**(1), 6047.

46. Liu, X., Meng, L., Li, X., Li, D., Liu, Q., Chen, Y., Li, X., Bu, W. and Sun, H. (2020) Regulation of FN1 degradation by the p62/SQSTM1-dependent autophagy–lysosome pathway in HNSCC. *Int. J. Oral Sci.* **12**(1), 34.

47. Brabletz, T., Kalluri, R., Nieto, M.A. and Weinberg, R.A. (2018) EMT in cancer. *Nat. Rev. Cancer* **18**(2), 128-134.

48. Wang, J., Deng, L., Huang, J., Cai, R., Zhu, X., Liu, F., Wang, Q., Zhang, J. and Zheng, Y. (2017) High expression of Fibronectin 1 suppresses apoptosis through the NF-κB pathway and is associated with migration in nasopharyngeal carcinoma. *Am. J. Transl. Res.* **9**(10), 4502-4511.

49. Waalkes, S., Atschekzei, F., Kramer, M.W., Hennenlotter, J., Vetter, G., Becker, J.U., Stenzl, A., Merseburger, A.S., Schrader, A.J., Kuczyk, M.A. and Serth, J. (2010) Fibronectin 1 mRNA expression correlates with advanced disease in renal cancer. *BMC Cancer* **10**(1), 503.

50. Sponziello, M., Rosignolo, F., Celano, M., Maggisano, V., Pecce, V., De Rose, R.F., Lombardo, G.E., Durante, C., Filetti, S., Damante, G., Russo, D. and Bulotta, S. (2016) Fibronectin-1 expression is increased in aggressive thyroid cancer and favors the migration and invasion of cancer cells. *Mol. Cell Endocrinol.* **431**, 123-132.

51. Yi, W., Xiao, E., Ding, R., Luo, P. and Yang, Y. (2016) High expression of fibronectin is associated with poor prognosis, cell proliferation and malignancy via the NF-κB/p53-apoptosis signaling pathway in colorectal cancer. *Oncol. Rep.* **36**(6), 3145-3153.

52. Patel, B.P., Shah, S.V., Shukla, S.N., Shah, P.M. and Patel, P.S. (2007) Clinical significance of MMP-2 and MMP-9 in patients with oral cancer. *Head Neck* **29**(6), 564-572.

53. Stamenkovic, I. (2000) Matrix metalloproteinases in tumor invasion and metastasis. *Semin. Cancer Biol.* **10**(6), 415-433.

54. Kessenbrock, K., Plaks, V. and Werb, Z. (2010) Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141**(1), 52-67.

55. Fan, H.-X., Li, H.-X., Chen, D., Gao, Z.-X. and Zheng, J.-H. (2012) Changes in the expression of MMP2, MMP9, and ColIV in stromal cells in oral squamous tongue cell carcinoma: relationships and prognostic implications. *J. Exp. Clin. Cancer Res.* **31**(1), 90.

56. Stanciu, A.E., Zamfir-Chiru-Anton, A., Stanciu, M.M., Popescu, C.R. and Gheorghe, D.C. (2016) Serum level of Matrix Metalloproteinase-9 in patients with head and neck squamous cell carcinoma. *Clin. Lab.* **62**(8), 1569-1574.

57. Senbanjo, L.T. and Chellaiah, M.A. (2017) CD44: a multifunctional cell surface adhesion receptor is a regulator of progression and metastasis of cancer cells. *Front. Cell Dev. Biol.* **5**, 18.

58. Chen, C., Zhao, S., Karnad, A. and Freeman, J.W. (2018) The biology and role of CD44 in cancer progression: therapeutic implications. *J. Hematol. Oncol.* **11**(1), 64.

59. Ludwig, N., Szczepanski, M.J., Gluszko, A., Szafarowski, T., Azambuja, J.H., Dolg, L., Gellrich, N.C., Kampmann, A., Whiteside, T.L. and Zimmerer, R.M. (2019) CD44(+) tumor cells promote early angiogenesis in head and neck squamous cell carcinoma. *Cancer Lett.* **467**, 85-95.

60. Ludwig, N., Yerneni, S.S., Razzo, B.M. and Whiteside, T.L. (2018) Exosomes from HNSCC promote angiogenesis through reprogramming of endothelial cells. *Mol. Cancer Res*. **16**(11), 1798-1808.

61. Krump, M. and Ehrmann, J. (2013) Differences in CD44s expression in HNSCC tumours of different areas within the oral cavity. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc. Czech Repub.* **157**(4), 280-283.

62. Shakib, P.A., Ensani, F., Abdirad, A., Valizadeh, B., Seyedmajidi, M. and Sum, S. (2015) CD44 and CD74: The promising candidates for molecular targeted therapy in oral squamous cell carcinoma. *Dent. Res. J.* **12**(2), 181-186.

63. Gavet, O. and Pines, J. (2010) Progressive activation of CyclinB1-Cdk1 coordinates entry to mitosis. *Dev. Cell* **18**(4), 533-543.

64. Hassan, K.A., El-Naggar, A.K., Soria, J.-C., Liu, D., Hong, W.K. and Mao, L. (2001) Clinical significance of Cyclin B1 protein expression in squamous cell carcinoma of the tongue. *Clin. Cancer Res.* **7**(8), 2458-2462.

65. Watanabe, S., Watanabe, R., Oton-Leite, A.F., Alencar Rde, C., Oliveira, J.C., Leles, C.R., Batista, A.C. and Mendonça, E.F. (2010) Analysis of cell proliferation and pattern of invasion in oral squamous cell carcinoma. *J. Oral Sci.* **52**(3), 417-424.

66. Strebhardt, K. and Ullrich, A. (2006) Targeting polo-like kinase 1 for cancer therapy. *Nat. Rev. Cancer* **6**(4), 321-330.

67. Kanaji, S., Saito, H., Tsujitani, S., Matsumoto, S., Tatebe, S., Kondo, A., Ozaki, M., Ito, H. and Ikeguchi, M. (2006) Expression of polo-like kinase 1 (PLK1) protein predicts the survival of patients with gastric carcinoma. *Oncology* **70**(2), 126-133.

68. Otto, T. and Sicinski, P. (2017) Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* **17**(2), 93-115.

69. De Luca, M., Lavia, P. and Guarguaglini, G. (2006) A functional interplay between Aurora-A, Plk1 and TPX2 at spindle poles: Plk1 controls centrosomal localization of Aurora-A and TPX2 spindle association. *Cell Cycle* **5**(3), 296-303.

70. Seki, A., Coppinger, J.A., Jang, C.Y., Yates, J.R. and Fang, G. (2008) Bora and the kinase Aurora a cooperatively activate the kinase Plk1 and control mitotic entry. *Science* **320**(5883), 1655-1658.

71. Macůrek, L., Lindqvist, A., Lim, D., Lampson, M.A., Klompmaker, R., Freire, R., Clouin, C., Taylor, S.S., Yaffe, M.B. and Medema, R.H. (2008) Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. *Nature* **455**(7209), 119-123.

72. Meraldi, P., Honda, R. and Nigg, E.A. (2002) Aurora-A overexpression reveals tetraploidization as a major route to centrosome amplification in p53-/- cells. *EMBO J.* **21**(4), 483-492.

73. Wang, X., Zhou, Y.X., Qiao, W., Tominaga, Y., Ouchi, M., Ouchi, T. and Deng, C.X. (2006) Overexpression of aurora kinase A in mouse mammary epithelium induces genetic instability preceding mammary tumor formation. *Oncogene* **25**(54), 7148-7158.

74. Marumoto, T., Hirota, T., Morisaki, T., Kunitoku, N., Zhang, D., Ichikawa, Y., Sasayama, T., Kuninaka, S., Mimori, T., Tamaki, N., Kimura, M., Okano, Y. and Saya, H. (2002) Roles of aurora-A kinase in mitotic entry and G2 checkpoint in mammalian cells. *Genes Cells* **7**(11), 1173-1182.

75. Anand, S., Penrhyn-Lowe, S. and Venkitaraman, A.R. (2003) AURORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol. *Cancer Cell* **3**(1), 51-62.

76. Wu, J., Yang, L., Shan, Y., Cai, C., Wang, S. and Zhang, H. (2016) AURKA promotes cell migration and invasion of head and neck squamous cell carcinoma through regulation of the AURKA/Akt/FAK signaling pathway. *Oncol. Lett.* **11**(3), 1889-1894.

77. Jeng, Y.-M., Peng, S.-Y., Lin, C.-Y. and Hsu, H.-C. (2004) Overexpression and amplification of Aurora-A in hepatocellular carcinoma. *Clin. Cancer Res.* **10**(6), 2065-2071.

78. Choi, J.-W., Kim, Y., Lee, J.-H. and Kim, Y.-S. (2013) High expression of spindle assembly checkpoint proteins CDC20 and MAD2 is associated with poor prognosis in urothelial bladder cancer. *Virchows Arch.* **463**(5), 681-687.

79. Wang, L., Zhang, J., Wan, L., Zhou, X., Wang, Z. and Wei, W. (2015) Targeting Cdc20 as a novel cancer therapeutic strategy. *Pharmacol. Ther.* **151**, 141-151.

80. Mondal, G., Sengupta, S., Panda, C.K., Gollin, S.M., Saunders, W.S. and Roychoudhury, S. (2007) Overexpression of Cdc20 leads to impairment of the spindle assembly checkpoint and aneuploidization in oral cancer. *Carcinogenesis* **28**(1), 81-92.

81. Kabeche, L. and Compton, D.A. (2012) Checkpoint-independent stabilization of kinetochore-microtubule attachments by Mad2 in human cells. *Curr. Biol.* **22**(7), 638-644.

82. Chang, D.Z., Ma, Y., Ji, B., Liu, Y., Hwu, P., Abbruzzese, J.L., Logsdon, C. and Wang, H. (2012) Increased CDC20 expression is associated with pancreatic ductal adenocarcinoma differentiation and progression. *J. Hematol. Oncol.* **5**, 15.

83. Kato, T., Daigo, Y., Aragaki, M., Ishikawa, K., Sato, M. and Kaji, M. (2012) Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients. *J. Surg. Oncol.* **106**(4), 423-430.

84. Wang, L., Yin, F., Du, Y., Du, W., Chen, B., Zhang, Y., Wu, K., Ding, J., Liu, J. and Fan, D. (2009) MAD2 as a key component of mitotic checkpoint: a probable prognostic factor for gastric cancer. *Am. J. Clin. Pathol.* **131**(6), 793-801.

85. Zhang, S.H., Xu, A.M., Chen, X.F., Li, D.H., Sun, M.P. and Wang, Y.J. (2008) Clinicopathologic significance of mitotic arrest defective protein 2 overexpression in hepatocellular carcinoma. *Hum. Pathol.* **39**(12), 1827-1834.

86. Nuncia-Cantarero, M., Martinez-Canales, S., Andrés-Pretel, F., Santpere, G., Ocaña, A. and Galan-Moya, E.M. (2018) Functional transcriptomic annotation and protein-protein interaction network analysis identify NEK2, BIRC5, and TOP2A as potential targets in obese patients with luminal A breast cancer. *Breast Cancer Res. Treat.* **168**(3), 613-623.

87. de Resende, M.F., Vieira, S., Chinen, L.T.D., Chiappelli, F., da Fonseca, F.P., Guimarães, G.C., Soares, F.A., Neves, I., Pagotty, S., Pellionisz, P.A., Barkhordarian, A., Brant, X. and Rocha, R.M. (2013) Prognostication of prostate cancer based on TOP2A protein and gene assessment: TOP2A in prostate cancer. *J. Transl. Med.* **11**(1), 36.

88. Zhang, R., Xu, J., Zhao, J. and Bai, J.H. (2018) Proliferation and invasion of colon cancer cells are suppressed by knockdown of TOP2A. *J. Cell. Biochem.* **119**(9), 7256-7263.

89. Zheng, H., Li, X., Chen, C., Chen, J., Sun, J., Sun, S., Jin, L., Li, J., Sun, S. and Wu, X. (2016) Quantum dot-based immunofluorescent imaging and quantitative detection of TOP2A and prognostic value in triple-negative breast cancer. *Int. J. Nanomedicine* **11**, 5519-5529.

90. Lan, J., Huang, H.Y., Lee, S.W., Chen, T.J., Tai, H.C., Hsu, H.P., Chang K.Y. and Li, C.F. (2014) TOP2A overexpression as a poor prognostic factor in patients with nasopharyngeal carcinoma. *Tumor Biol.* **35**(1), 179-187.

91. Kaplan, E. and Gündüz, U. (2012) Expression analysis of TOP2A, MSH2 and MLH1 genes in MCF7 cells at different levels of etoposide resistance. *Biomed. Pharmacother.* **66**(1), 29-35.

92. Pritchard, K.I., Messersmith, H., Elavathil, L., Trudeau, M., O'Malley, F. and Dhesy-Thind, B. (2008) HER-2 and Topoisomerase II as predictors of response to chemotherapy. *J. Clin. Oncol.* **26**(5), 736-744.

93. Champoux, J.J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70**, 369-413.

94. Wachi, S., Yoneda, K. and Wu, R. (2005) Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* **21**(23), 4205-4208.

95. Sajnani, M.R., Patel, A.K., Bhatt, V.D., Tripathi, A.K., Ahir, V.B., Shankar, V., Shah, S., Shah, T.M., Koringa, P.G., Jakhesara, S.J. and Joshi, C.G. (2012) Identification of novel transcripts deregulated in buccal cancer by RNA-seq. *Gene* **507**(2), 152-158.

96. Thiel, U.J.E., Feltens, R., Adryan, B., Gieringer, R., Brochhausen, C., Schuon, R., Fillies, T., Grus, F., Mann, W.J. and Brieger, J. (2011) Analysis of differentially expressed proteins in oral squamous cell carcinoma by MALDI-TOF MS. *J. Oral Pathol. Med.* **40**(5), 369-379.

97. He, Q.-Y., Chen, J., Kung, H.-F., Yuen, A.P.-W. and Chiu, J.-F. (2004) Identification of tumor-associated proteins in oral tongue squamous cell carcinoma by proteomics. *Proteomics* **4**(1), 271-278.

98. Yang, K., Zhang, S., Zhang, D., Tao, Q., Zhang, T., Liu, G., Liu, X. and Zhao, T. (2019) Identification of SERPINE1, PLAU and ACTA1 as biomarkers of head and neck squamous cell carcinoma based on integrated bioinformatics analysis. *Int. J. Clin. Oncol.* **24**(9), 1030-1041.

99. Shaikh, I., Ansari, A., Ayachit, G., Gandhi, M., Sharma, P., Bhairappanavar, S., Joshi, C.G. and Das, J. (2019) Differential gene expression analysis of HNSCC tumors deciphered tobacco dependent and independent molecular signatures. *Oncotarget* **10**(58), 6168-6183.

100. Meng, D., Liu, T., Ma, F. and Wang, M. (2021) Screening the key genes of prognostic value in the microenvironment for head and neck squamous cell carcinoma. *Medicine* **100**(4), e24184-e24184.

101. Onken, M.D., Winkler, A.E., Kanchi, K.L., Chalivendra, V., Law, J.H., Rickert, C.G., Kallogjeri, D., Judd, N.P., Dunn, G.P., Piccirillo, J.F., Lewis, JS. Jr., Mardis, E.R. and Uppaluri, R. (2014) A surprising cross-species conservation in the genomic landscape of mouse and human oral cancer identifies a transcriptional signature predicting metastatic disease. *Clin. Cancer Res.* **20**(11), 2873-2884.

102. Multhaupt, H.A.B., Leitinger, B., Gullberg, D. and Couchman, J.R. (2016) Extracellular matrix component signaling in cancer. *Adv. Drug Deliv. Rev.* **97**, 28-40.

103. Kim, N., Hong, Y., Kwon, D. and Yoon, S. (2013) Somatic mutaome profile in human cancer tissues. *Genomics Inform.* **11**(4), 239-244.

104. Mazzoccoli, G., Castellana, S., Carella, M., Palumbo, O., Tiberio, C., Fusilli, C., Capocefalo, D., Biagini, T., Mazza, T. and Lo Muzio, L. (2017) A primary tumor gene expression signature identifies a crucial role played by tumor stroma myofibroblasts in lymph node involvement in oral squamous cell carcinoma. *Oncotarget* **8**(62), 104913-104927.

105. Marsh, D., Suchak, K., Moutasim, K.A., Vallath, S., Hopper, C., Jerjes, W., Upile, T., Kalavrezos, N., Violette, S.M., Weinreb, P.H., Chester, K.A., Chana, J.S., Marshall, J.F., Hart, I.R., Hackshaw, A.K., Piper, K. and Thomas, G.J. (2011) Stromal features are predictive of disease mortality in oral cancer patients. *J. Pathol.* **223**(4), 470-481.

106. Sobral, L.M., Bufalino, A., Lopes, M.A., Graner, E., Salo, T. and Coletta, R.D. (2011) Myofibroblasts in the stroma of oral cancer promote tumorigenesis via secretion of activin A. *Oral Oncol.* **47**(9), 840-846.

107. Luksic, I., Suton, P., Manojlovic, S., Virag, M., Petrovecki, M. and Macan, D. (2015) Significance of myofibroblast appearance in squamous cell carcinoma of the oral cavity on the occurrence of occult regional metastases, distant metastases, and survival. *Int. J. Oral Maxillofac. Surg.* **44**(9), 1075-1080.

108. Lo, L.H., Lam, C.Y., To, J.C., Chiu, C.H. and Keng, V.W. (2021) Sleeping Beauty insertional mutagenesis screen identifies the pro-metastatic roles of CNPY2 and ACTN2 in hepatocellular carcinoma tumor progression. *Biochem. Biophys. Res. Commun.* **541**, 70-77.

109. Misawa, K., Imai, A., Matsui, H., Kanai, A., Misawa, Y., Mochizuki, D., Mima, M., Yamada, S., Kurokawa, T., Nakagawa, T. and Mineta, H. (2020) Identification of novel methylation markers in HPV-associated oropharyngeal cancer: genome-wide discovery, tissue verification and validation testing in ctDNA. *Oncogene* **39**(24), 4741-4755.

110. Xu, G. and Li, J.Y. (2016) ATP5A1 and ATP5B are highly expressed in glioblastoma tumor cells and endothelial cells of microvascular proliferation. *J. Neurooncol.* **126**(3), 405-413.

111. Feichtinger, R.G., Schäfer, G., Seifarth, C., Mayr, J.A., Kofler, B. and Klocker, H. (2018) Reduced levels of ATP Synthase Subunit ATP5F1A correlate with earlier-onset prostate cancer. *Oxid. Med. Cell. Longev.* **2018**, 1347174.

112. Chen, W.-L., Kuo, K.-T., Chou, T.-Y., Chen, C.-L., Wang, C.-H., Wei, Y.-H. and Wang, L.-S. (2012) The role of cytochrome c oxidase subunit Va in non-small cell lung carcinoma cells: association with migration, invasion and prediction of distant metastasis. *BMC Cancer* **12**(1), 273.

113. Fu, W.-N., Shang, C., Huang, D.-F., Xu, Z.-M., Sun, X.-H. and Sun, K.-L. (2006) Average-12.9 chromosome imbalances coupling with 15 differential expression genes possibly involved in the carcinogenesis, progression and metastasis of supraglottic laryngeal squamous cell cancer. *Chin. J. Med. Genet.* **23**(1), 7-11.

114. Liu, J., Zhu, W., Li, Z., Cai, G., Wang, J., Tang, Q., Maroun, C.A. and Zhu, G. (2020) Proteomic analysis of hypopharyngeal and laryngeal squamous cell carcinoma sheds light on differences in survival. *Sci. Rep.* **10**(1), 19459.

115. Spratt, D.E., Chan, T., Waldron, L., Speers, C., Feng, F.Y., Ogunwobi, O.O. and Osborne, J.R. (2016) Racial/ethnic disparities in genomic sequencing. *JAMA Oncol.* **2**(8), 1070-1074.

116. GDC Data Portal. (2021). Harmonized cancer datasets: genomic data commons data portal. https://portal.gdc.cancer.gov/

117. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H. and Wanker, E.E. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**(6), 957-968.

118. Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P. and Vidal, M. (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**(7062), 1173-1178.

119. Nagarsheth, N., Wicha, M.S. and Zou, W. (2017) Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat. Rev. Immunol.* **17**(9), 559-572.

120. Hanash, S. and Schliekelman, M. (2014) Proteomic profiling of the tumor microenvironment: recent insights and the search for biomarkers. *Genome Med.* **6**(2), 12.