OncoTargets and Therapy Dovepress

Open Access Full Text Article

ORIGINAL RESEARCH

Exploring Potential Biomarkers and Molecular Mechanisms of Cutaneous Squamous Cell Carcinoma Based on Bioinformatics

Jiayue Qi[1,2], Qingqing Guo[1,2], Jia Bai[1], Xiaoqiang Liang[1], Wenwei Zhu[1], Chengxin Li[1,2,*], Fang Xie[1,*]

[1]Department of Dermatology, First Medical Center, Chinese PLA General Hospital, Beijing, People's Republic of China; [2]School of Medicine, Nankai University, Tianjin, People's Republic of China

*These authors contributed equally to this work

Correspondence: Chengxin Li; Fang Xie, Department of Dermatology, First Medical Center, Chinese PLA General Hospital, Beijing, 100853, People's Republic of China, Email chengxinderm@163.com; xiefang@301hospital.com.cn

Video abstract

Point your SmartPhone at the code above. If you have a QR code reader the video abstract will appear. Or use: https://youtu.be/BGVYrTDRXDk

OncoTargets and Therapy downloaded from https://www.dovepress.com/
For personal use only.

Purpose: Cutaneous squamous cell carcinoma (cSCC) ranks as the second most common malignancy in clinical practice and poses a significant threat to public health due to its high malignancy. In this study, we aimed to explore potential biomarkers and molecular mechanisms of cSCC.

Methods: Differentially expressed genes (DEGs) from GSE66359 and GSE117247 datasets were identified using R software. We conducted enrichment analyses and screened hub genes through protein-protein interaction (PPI) analysis and weighted gene co-expression network analysis (WGCNA). To assess the diagnostic performance of these genes, we generated ROC curves using both internal and external datasets (GSE45164) and validated the expression levels of these genes in cSCC tissues through immunohistochemistry. Subsequently, we predicted the target miRNAs and lncRNAs for hub genes using online databases and constructed competing endogenous RNA (ceRNA) networks.

Results: In total, we identified 505 upregulated DEGs and 522 downregulated DEGs. Through PPI and WGCNA analyses, we identified four hub genes exhibiting robust diagnostic performance in internal and external datasets (AUC > 0.9) and selected three previously unreported genes for further analysis. Immunohistochemistry demonstrated significantly elevated CCNA2, CCNB2, and UBE2C expression in cSCC tissues compared to normal skin tissues. Finally, we constructed three ceRNA networks, namely NEAT1/H19-hsa-miR-148a-3p-CCNA2 and NEAT1-hsa-miR-140-3p-UBE2C.

Conclusion: In conclusion, we have identified CCNA2, CCNB2, and UBE2C as novel biomarkers for cSCC, and the NEAT1/H19-hsa-miR-148a-3p-CCNA2 and NEAT1-hsa-miR-140-3p-UBE2C ceRNA networks may represent molecular mechanisms under-lying cSCC progression. The findings of this study offer new diagnostic and therapeutic options for cSCC patients.

Keywords: weighted gene co-expression network analysis, protein-protein interaction analysis, competitive endogenous RNA, pathogenesis

Introduction

Cutaneous squamous cell carcinoma (cSCC), originating from epidermal keratinocytes, ranks as the second most prevalent malignant cutaneous neoplasm in clinical practice, constituting approximately 20–50% of all skin malignancies.[1] It exhibits a propensity for metastasis, disseminating to lymph nodes and various organs, including the liver, lungs, and bones, posing a substantial threat to life.[2] In recent years, the incidence of cSCC has gradually increased, owing to its heightened malignancy, thereby representing a significant menace to public health.[3] Despite certain advancements in diagnosis and treatment, there remains a lack of comprehensive understanding regarding the pathogenesis of cSCC, and satisfactory therapeutic efficacy still needs to be discovered, with long-term survival rates of treated patients falling below 20%.[4] Consequently, there is a pressing need for further exploration of novel, effective biomarkers,

Received: 8 May 2024
Accepted: 13 October 2024
Published: 26 October 2024

© 2024 Qi et al. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v3.0) License (http://creativecommons.org/licenses/by-nc/3.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

augmenting our comprehension of the pathogenesis of cSCC, and offering fresh diagnostic and therapeutic alternatives for cSCC patients.

Currently, transcriptomics and chip analysis have found widespread application in various diseases.[5–7] Differentially expressed genes analysis is a quantitative approach for investigating changes in gene expression levels between experimental and control groups in multiple samples.[8] It is extensively employed to identify potential biomarkers and therapeutic targets for cancer. Another crucial analysis in transcriptomics is weighted gene co-expression network analysis (WGCNA), which allows the examination of gene expression patterns, identification of co-expression modules containing highly correlated genes, and modules associated with clinical traits.[9] Consequently, integrating WGCNA and differentially expressed gene analysis results can enhance the precision in identifying potential hub genes in cSCC. Furthermore, competitive endogenous RNA (ceRNA) networks can elucidate novel mechanisms that promote cSCC development within the transcriptional regulatory network.[10] The combination of microarray data and bioinformatics analysis allows for exploring potential hub genes and regulatory networks closely associated with disease progression.

In this study, we retrieved microarray datasets for cSCC from the GEO database. Following data integration and batch effect correction, we identified differentially expressed genes (DEGs) between cSCC and normal skin tissues. Gene ontology (GO) analysis was employed to unveil critical biological functions implicated in the pathogenesis of cSCC. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis elucidated molecular pathways associated with cSCC. Protein-protein interaction (PPI) analysis and WGCNA were utilized to explore hub modules and ultimately identify hub genes linked to cSCC, with subsequent validation through immunohistochemistry. Lastly, target miRNAs and lncRNAs of hub genes were predicted using online databases, and ceRNA networks were constructed. The workflow of our study is shown in Figure 1. This work provides an in-depth exploration of the disease development mechanisms of cSCC at the transcriptomic level.

Materials and Methods

Data Preparation

Gene expression data for many diseases can be publicly accessed in the GEO database (www.ncbi.nlm.nih.gov/geo). The GSE66359 dataset was generated using the Affymetrix Human Genome U133 Plus 2.0 Array microarray platform and comprises eight cSCC samples and five normal skin samples. The GSE117247 dataset was generated using the Affymetrix Human Genome U95 Version 2 Array platform and includes eight cSCC samples and ten normal skin

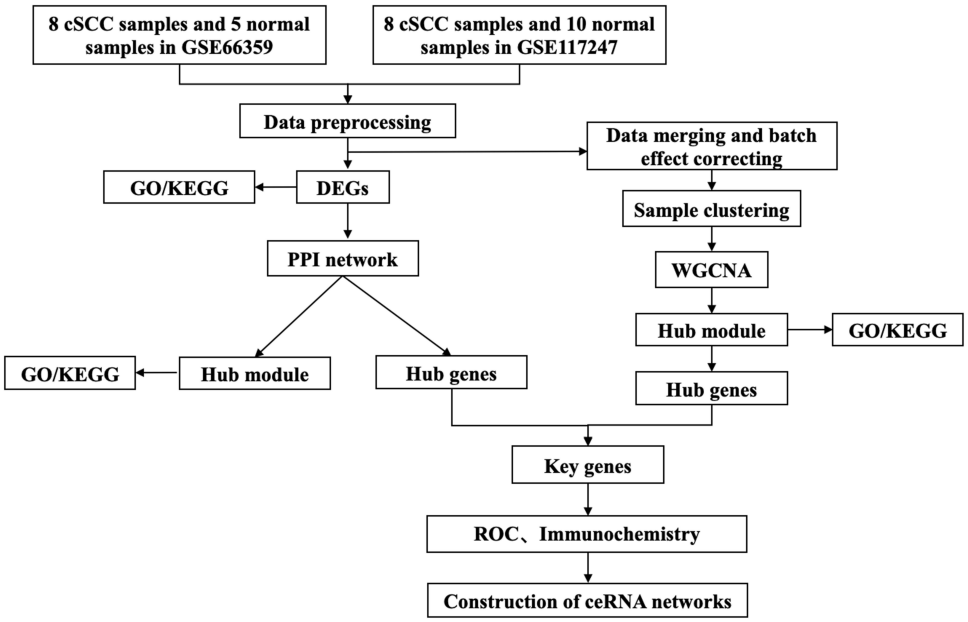

Figure 1 The workflow of our study to explore potential biomarkers and molecular mechanisms of cSCC based on bioinformatics.

samples. The GSE45164 dataset was generated using the Affymetrix Human Genome U133A 2.0 Array platform and consists of ten cSCC samples and three normal skin samples. We merged the samples from the GSE66359 and GSE117247 datasets and corrected batch effects using the Combat package in R software (version 4.1.2).[11] The merged dataset was designated as GSE66359-117247, comprising sixteen cSCC samples and fifteen normal samples. GSE66359, GSE117247, GSE45164, and GSE66359-117247 will be utilized for subsequent analyses.

Data Preprocessing and DEGs Selection

Quality control and preprocessing of the raw files (*.CEL) obtained from the GEO database were conducted using the AffyPLM package for data quality assessment.[12] Relative log expression (RLE) plots were employed to evaluate data trend consistency. Normalization was performed using the Robust Multi-Array Average (RMA) algorithm based on R software (version 4.1.2),[13] and the results were validated using principal components analysis (PCA). Probes were converted to gene symbols using the hgu133plus2.db, hgu95av2.db, and hgu133a2.db packages. In cases where one gene symbol corresponded to multiple probes, the probe with the median expression value was selected. DEGs between samples were conducted using the limma package,[14] with criteria set at $|\log2FC| > 2$ and $p < 0.05$ for gene selection. Volcano plots and heatmaps were generated using the heatmap package to visualize the expression patterns of differentially expressed genes.

Enrichment Analysis

To perform functional-level analysis of DEGs, we conducted enrichment analysis for GO[15] and KEGG[16] pathways using the clusterProfiler package[17] in R software. GO analysis encompasses biological processes, cellular components, and molecular functions. A significance threshold of $p < 0.05$ was considered statistically meaningful.

Construction of PPI Network

We utilized the STRING online database (https://string-db.org)[18] to construct the PPI network, setting a threshold interaction score of >0.4. Subsequently, we employed Cytoscape software (version 3.9.1) for network visualization.[19] We utilized the MCODE plugin to extract densely connected modules from the PPI network, with filtering criteria set as degree cutoff = 2, node score cutoff = 0.2, K-core = 2, and maximum depth = 100. The CytoHubba plugin was employed to identify key genes within the PPI network. We extracted the top 30 genes ranked by five algorithms (Degree, MCC, MNC, Closeness, and Stress). Genes that appeared at the intersection of all five methods and were present in the top-ranked module were considered potential key genes.

Construction of WGCNA and Identification of Hub Genes

WGCNA is a practical algorithm for discovering highly correlated gene modules and identifying phenotypes-associated modules.[20] We employed the WGCNA package to construct a weighted gene co-expression network. The gene expression matrix, GSE66359-117247, obtained after data integration and batch effect correction, was used as input for WGCNA. Initially, we conducted hierarchical clustering of samples to remove outliers and ensure the accuracy of the analysis. To establish a scale-free network, we determined the soft threshold, β, using the pickSoftThreshold function. Subsequently, we constructed the weighted network, created a hierarchical clustering dendrogram, and delineated modules using the blockwiseModules function. To further identify functional modules within the co-expression network, module-trait associations were calculated, and modules with high correlation coefficients were considered cSCC-related modules for subsequent analysis. For each gene within a module, the membership (MM) reflects its correlation with module eigengenes, while the gene significance (GS) represents the correlation of the gene with corresponding clinical traits. Genes with $MM > 0.8$ and $GS > 0.6$ were considered candidate key genes, and those overlapping with key genes from the PPI analysis were designated as the final hub genes.

Validation of Receiver Operating Characteristic (ROC) Curves

GSE66359+117247 were utilized as internal datasets, while GSE45164 was employed as an external dataset to evaluate the diagnostic efficacy of the hub genes through ROC curve analysis. The Area Under the Curve (AUC), representing the area under the ROC curve, serves as a metric for assessing the performance of various predictive models.

Samples Collection

We selected 30 patients diagnosed with cSCC who received treatment at our institution between April 2017 and April 2022. Surgical excision was performed to obtain corresponding tumor tissues and adjacent normal tissues. All patients underwent routine examinations, including complete blood counts, and had no history of other malignancies. Histopathological diagnosis confirmed the presence of cSCC in all cases.

Immunohistochemistry

The cSCC and adjacent normal tissues were obtained through surgical excision, followed by paraffin embedding and sectioning for immunohistochemical staining. The tissue sections were incubated in 0.1% BSA, followed by overnight incubation at 4°C with antibodies against CCNA2 (Proteintech, 66087-1-Ig), CCNB2 (Proteintech, 66391-1-Ig), and UBE2C (Proteintech, 66726-1-Ig). Biotinylated IgG was used as the secondary antibody. Color development was achieved using the DAB staining kit following the secondary antibody reaction, and the sections were observed and captured using an optical microscope. Finally, analysis was conducted using ImageJ software.

Construction of the ceRNA Networks

We employed five online miRNA databases, RNA22, DIANA-micro T, miRWalk, miRDB, and miRcode, to predict the target miRNAs for the hub genes. Subsequently, we selected miRNAs that were identified in at least four of these databases for further analysis. StarBase (version 3.0) (http://starbase.sysu.edu.cn/index.php) was used to predict the interactions of selected miRNAs with lncRNAs.[21] We selected lncRNAs that were present in the majority of miRNA prediction results for further analysis. Cytoscape was utilized to construct interaction networks, namely the ceRNA networks, based on interactions among mRNAs, miRNAs, and lncRNAs.

Data Statistics

Statistical analysis of experimental data was conducted using R software (version 4.1.2) and Prism software (version 9.3.1). Student's *t*-test was used to compare the differences between the two groups. A paired *t*-test was employed to validate significant differences in Immunohistochemistry. *** indicated $P < 0.001$.

Results

Data Quality Control

Regression calculations were performed using the affyPLM package in R language. RLE boxplots were generated to validate the homogeneity between the datasets. RLE plots demonstrated enrichment of gene expression values around zero in GSE66359 and GSE117247, indicating high consistency and their suitability for subsequent analysis (Figure 2).

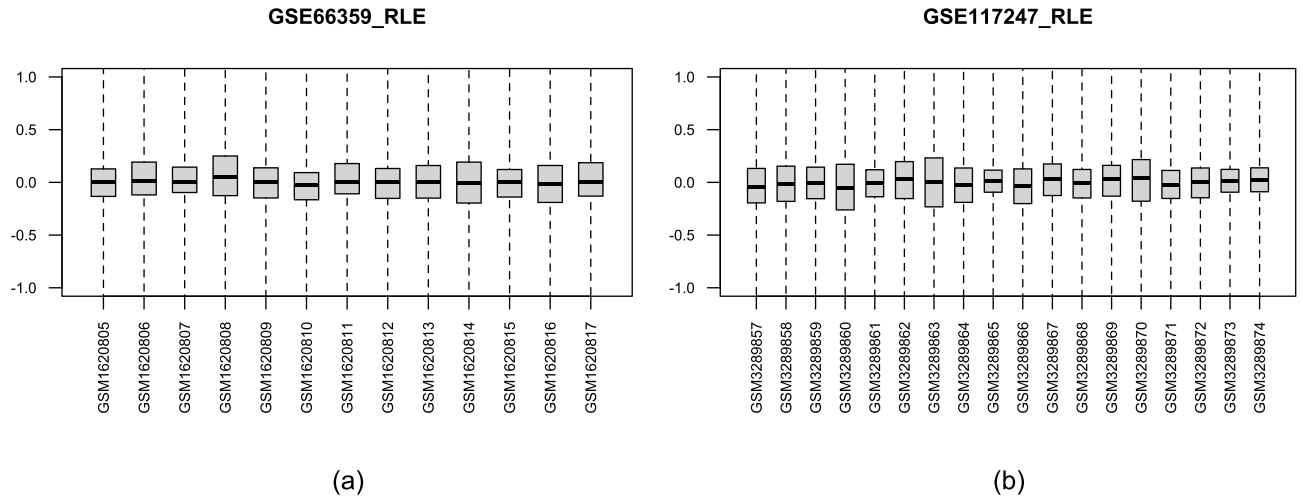

Figure 2 Sample distribution by regression calculation. (**a**) RLE boxplot for GSE66359; (**b**) RLE boxplot for GSE117247.

Normalization of Data

The RMA algorithm was employed for background correction and normalization of the samples. Gene expression density curves (Figure 3a and b) and boxplots (Figure 3c and d) revealed that gene expression values in both datasets ranged between 0 and 15, indicating similar expression trends between the two datasets. Furthermore, PCA illustrated significant differences between the tumor and normal samples in the two datasets (Figure 3e and f).

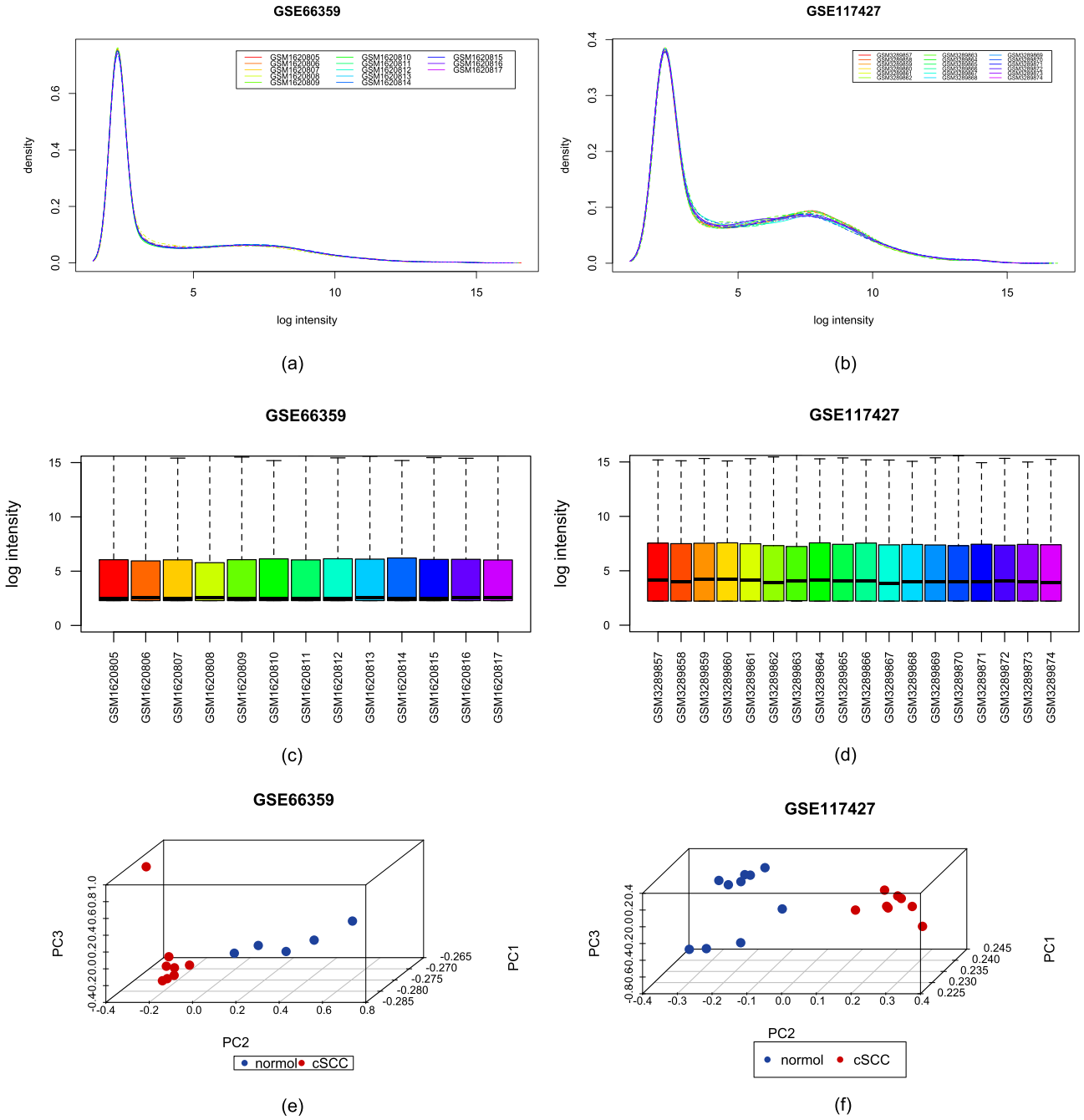

Figure 3 Distribution figures among samples after normalization. (**a**) Gene expression density curve after normalization in GSE66359; (**b**) Gene expression density curve after normalization in GSE117247; (**c**) Gene expression boxplot after normalization in GSE66359; (**d**) Gene expression boxplot after normalization in GSE117247; (**e**) Sample distribution in GSE66359 by PCA analysis. (**f**) Sample distribution in GSE117247 by PCA analysis.

Identification of DEGs

The gene screening was performed by limma package in R language. Based on the cutoff criteria of |log2FC|>2 and p value<0.05, 820 DEGs in the GSE66359 dataset were found, comprising 451 upregulated genes and 369 downregulated genes. 287 DEGs were found in the GSE117247 dataset, including 94 upregulated genes and 193 downregulated genes. Next, heatmaps (Figure 4a and b) and volcano plots (Figure 4c and d) analyses were used to visualize these DEGs. The DEGs found from the two datasets were combined, and 505 upregulated genes and 522 downregulated genes were extracted for the next validation step.

Enrichment Analysis

To further elucidate the potential functions of these DEGs, we conducted an enrichment analysis using the clusterProfiler package. Following GO enrichment analysis (Figure 5a and b), we found that upregulated DEGs were primarily enriched in biological processes related to mitotic sister chromatid segregation and mitotic nuclear division. Regarding cellular components, these DEGs were notably associated with chromosomal regions and centromeres. In molecular functions, microtubule binding and single-stranded DNA binding were identified as relevant to the upregulated DEGs. On the other hand, downregulated DEGs were mainly involved in biological processes such as skin development, epidermis development, and keratinocyte differentiation. They were also associated with cellular components like collagen-containing extracellular matrix and cornified envelope. They

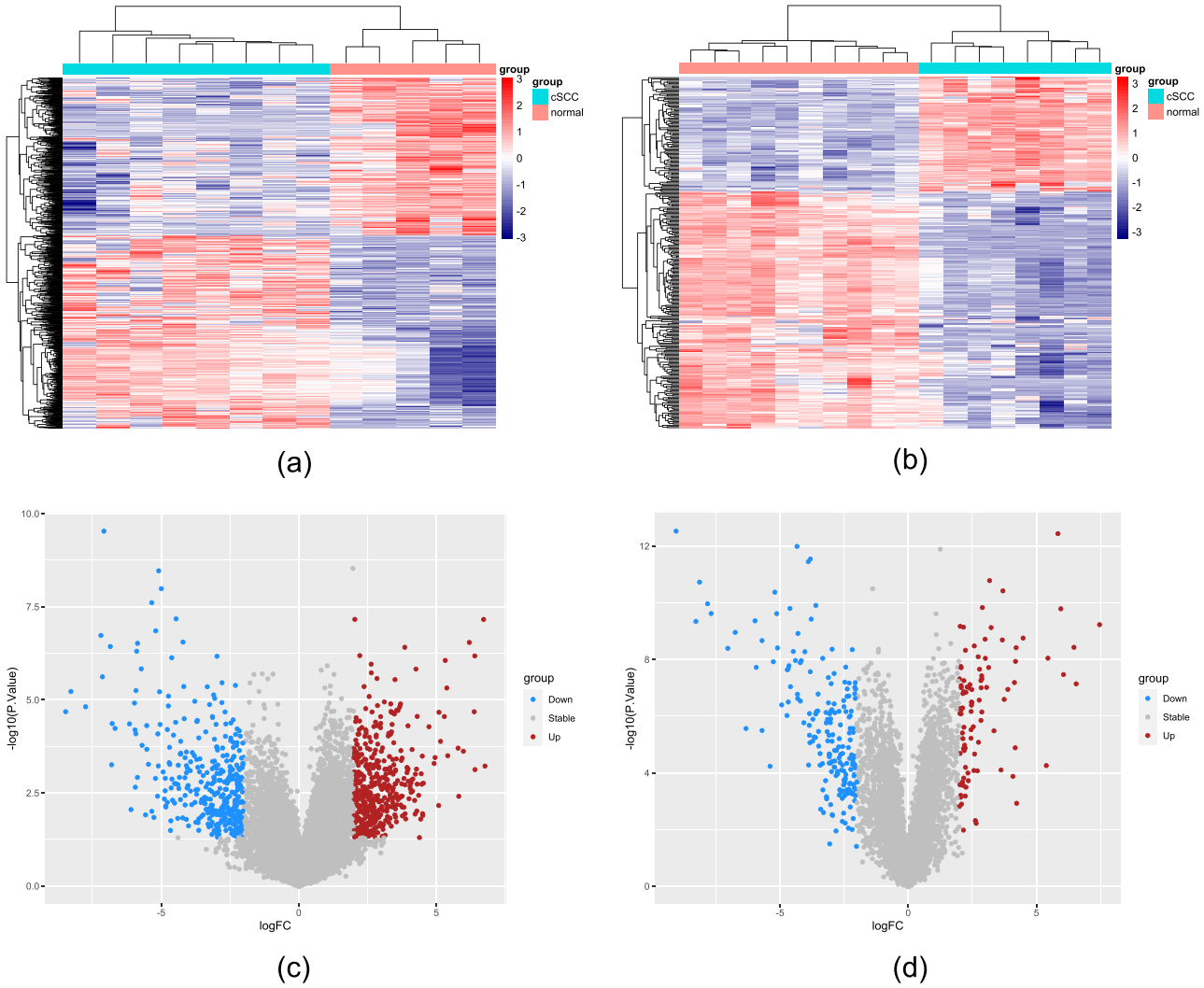

Figure 4 Identification of DEGs by limma package. (**a**) DEGs of GSE66359 in heatmap; (**b**) DEGs of GSE117247 in heatmap; (**c**) DEGs of GSE66359 in volcano plot. (**d**) DEGs of GSE117247 in volcano plot.

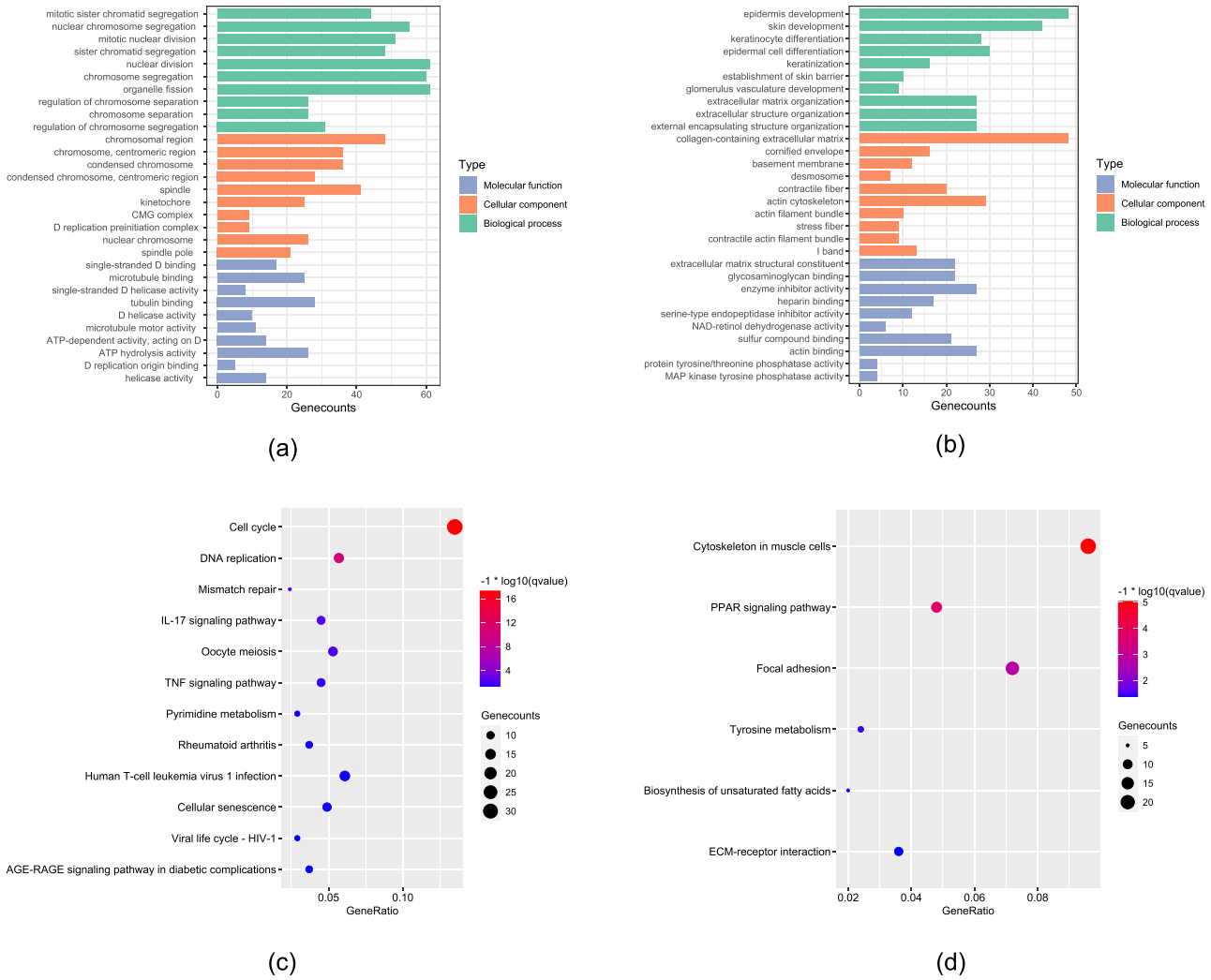

Figure 5 Functional enrichment analysis of DEGs. (**a**) GO enrichment analysis of the upregulated DEGs; (**b**) GO enrichment analysis of the downregulated DEGs; (**c**) KEGG analysis of the upregulated DEGs; (**d**) KEGG analysis of the downregulated DEGs.

were linked to extracellular matrix structural constituent and glycosaminoglycan binding in molecular functions. KEGG analysis revealed that upregulated DEGs were significantly enriched in signaling pathways such as the cell cycle, DNA replication, mismatch repair, and the IL-17 signaling pathway. In contrast, downregulated DEGs were primarily enriched in pathways, including focal adhesion and PPAR signaling (Figure 5c and d).

PPI Construction and MCODE Module Analysis

Protein-protein interactions constitute a significant source of functional complexity within cells, and PPI analysis can provide deeper insights into disease progression. Using the STRING online database, we constructed an interaction network of proteins encoded by DEGs. This network consisted of 585 nodes and 4789 edges and was visualized using Cytoscape. We employed five algorithms (Degree, MCC, MNC, Closeness, Stress) from the cytoHubba plugin for gene selection and listed the top 30 genes for each method (Table 1). Genes simultaneously selected by all five algorithms were considered key genes (AURKA, BIRC5, CCNA2, CCNB2, CDK1, UBE2C). Subsequently, we conducted module analysis using the Molecular Complex Detection (MCODE) plugin and identified 25 clusters. We selected the top-ranked module (score = 62.551) as the key module, noteworthy for containing the six key genes mentioned earlier. This module played a crucial role in the constructed PPI network, consisting of 20 nodes and 2158 edges (Figure 6a). We then performed GO and KEGG analyses on the genes within

Table 1 Key Genes Identified Using the cytoHubba Plugin

Degree	MCC	MNC	Closeness	Stress
CDK1	**AURKA**	**CDK1**	**CDK1**	CPEB2
CCNA2	DLGAP5	**CCNA2**	**CCNA2**	PKP1
CCNB1	NCAPG	CCNB1	**UBE2C**	**CDK1**
CCNB2	CCNB1	KIF11	CCNB1	IL6
KIF11	KIF11	**CCNB2**	**AURKA**	CEBPA
TOP2A	BUB1B	TOP2A	**CCNB2**	MYC
BUB1	NUSAP1	BUB1	AURKB	LPL
NCAPG	**CCNB2**	NCAPG	KIF11	**UBE2C**
BUB1B	**CCNA2**	BUB1B	CDC20	EPHB2
AURKB	KIF2C	AURKB	RRM2	MMP2
AURKA	AURKB	MAD2L1	NCAPG	DSG1
MAD2L1	TPX2	**AURKA**	TOP2A	CDSN
RRM2	TOP2A	RRM2	**BIRC5**	FADS2
CDC20	**CDK1**	CDC20	MAD2L1	**AURKA**
UBE2C	BUB1	**UBE2C**	BUB1	FOS
CDC6	**UBE2C**	CDC6	BUB1B	DSC1
ASPM	ASPM	ASPM	CDC6	PRKCA
DLGAP5	KIF20A	DLGAP5	DLGAP5	RASA1
NDC80	TTK	NDC80	PLK1	**BIRC5**
KIF20A	CENPA	KIF20A	ASPM	PCNA
TTK	PBK	TTK	CDCA8	PLK1
CDCA8	**BIRC5**	CDCA8	KIF20A	**CCNB2**
CDC45	CDC45	CDC45	CDC45	PPL
TPX2	OIP5	TPX2	NDC80	TK1
BIRC5	CEP55	**BIRC5**	NUSAP1	CDKN1A
KIF2C	CDCA8	KIF2C	TTK	CDC6
NUSAP1	CENPF	NUSAP1	TPX2	PDGFRB
PBK	RRM2	PBK	KIF2C	LMNB1
PLK1	NDC80	KIF23	PBK	DDX58
KIF23	ZWINT	PLK1	MKI67	**CCNA2**

Notes: The bold values represent the genes that appear in all five ranking methods used here.
Abbreviations: MCC, maximal clique centrality; MNC, maximum neighborhood component.

the key module (Figure 6b and c). In terms of biological processes, the key module genes were mainly enriched in nuclear division and organelle fission. Regarding cellular components, these genes were primarily associated with chromosomal and centromeric regions. In molecular functions, these genes were mainly related to microtubule binding. KEGG analysis indicated that the genes within the key module were enriched in pathways such as the cell cycle and oocyte meiosis.

WGCNA and Hub Module Identification

During the sample clustering process, two samples (GSM1620814 and GSM1620816) were identified as outliers and subsequently excluded (Figure 7a). Additionally, we determined that $\beta = 10$ and $R2 = 0.85$ were the optimal soft-thresholding parameters to ensure the scale-free nature of the topology model (Figure 7b). In total, 26 modules were obtained, and we computed the correlation of each module with cSCC (Figure 7c). Among these modules, the blue module exhibited the strongest correlation with cSCC (Figure 7d). Furthermore, module-trait analysis confirmed a highly positive correlation between the blue module and cSCC ($r=0.58$, $p<0.001$) (Figure 7e). It suggests that genes within the blue module may play a critical role in the development of cSCC. Molecular functional analysis (Figure 7f) revealed that genes within the blue module were significantly enriched in pathways associated with single-stranded DNA binding,

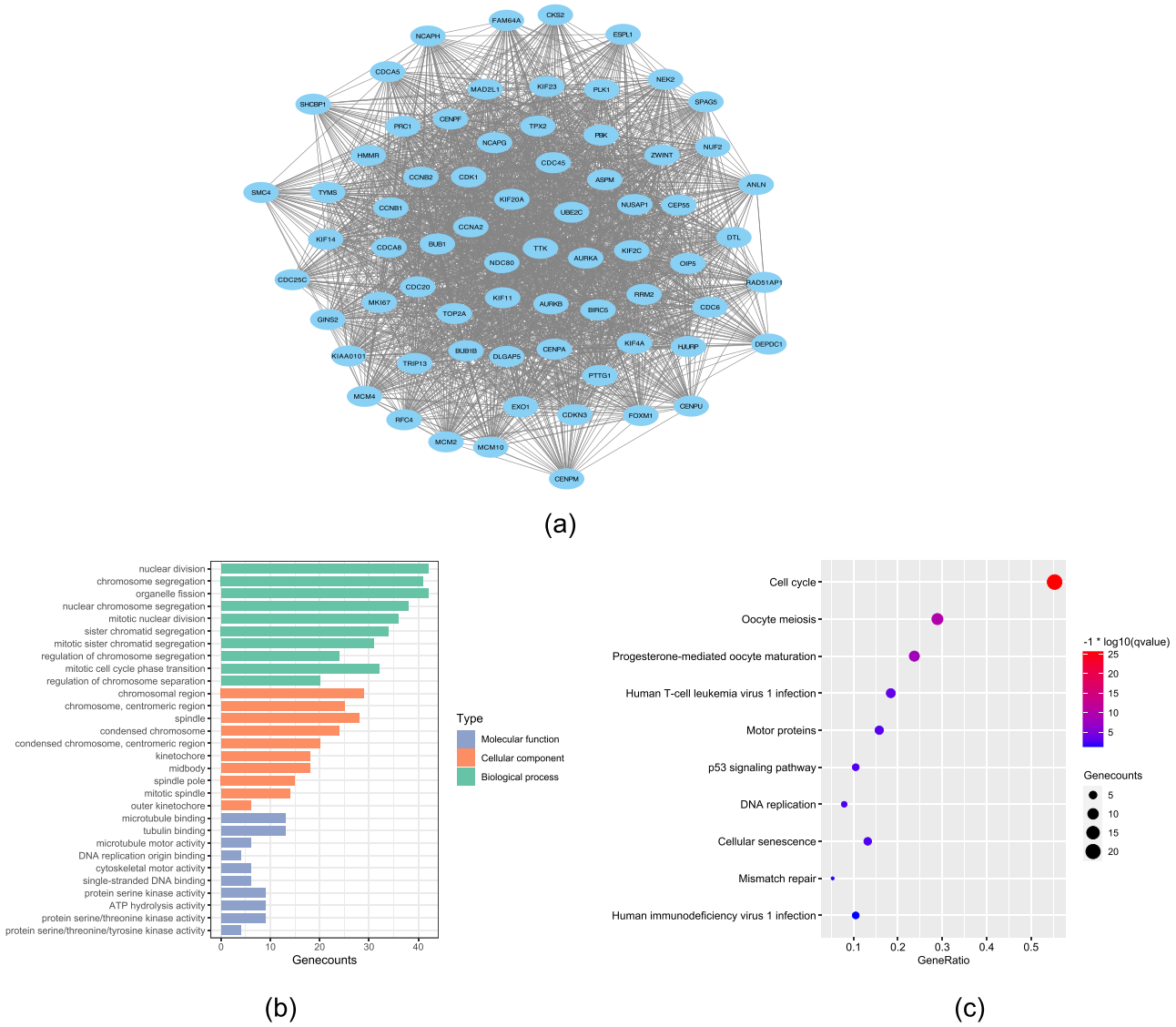

Figure 6 PPI network construction and module analyses. (**a**) The top-ranked module (score = 62.551) contains 20 nodes and 2158 edges; (**b**) GO enrichment analysis of genes in the top-ranked module; (**c**) KEGG analysis of genes in the top-ranked module.

catalytic activity on DNA, single-stranded DNA helicase activity, and microtubule binding. Cellular component analysis indicated a strong correlation of genes within the blue module with chromatin regions and condensed chromatin components. Furthermore, these genes were found to be actively involved in key biological processes, including DNA replication. KEGG analysis underscored the enrichment of genes in the blue module within pathways related to DNA replication, the cell cycle, and the proteasome pathway (Figure 7g). We examined the relationship between MM and GS to select key genes, revealing a positive correlation (r=0.37, p=5.6E-48) (Figure 7h). Genes with GS greater than 0.6 and MM greater than 0.8 were selected as candidate key genes. An intersection was taken between these candidates and those identified through PPI analysis, ultimately determining key genes, including CDK1, CCNA2, CCNB2, and UBE2C.

ROC Validation

To evaluate the diagnostic performance of the key genes in distinguishing cSCC from normal skin tissue, GSE66359 +117247 were used as internal validation sets (Figure 8a), while GSE45164 served as an external validation set to generate ROC curves for these four key genes (Figure 8b). These four genes exhibit high diagnostic value, with the AUCs exceeding 90%. Therefore, CDK1, CCNA2, CCNB2, and UBE2C are hub genes in cSCC.

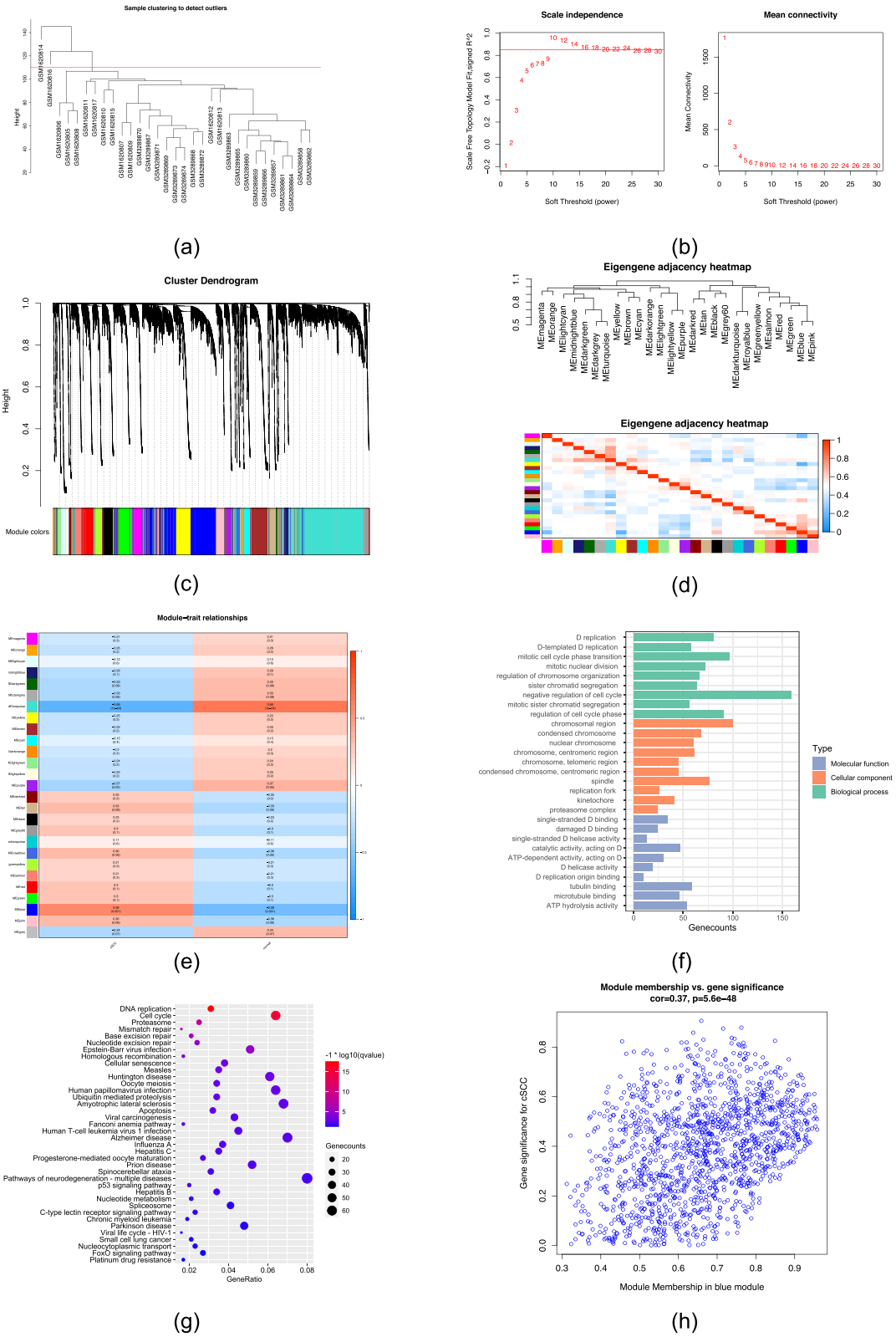

Figure 7 Co-expression network analysis based on WGCNA. (**a**) Sample clustering to detect outliers; (**b**) Graphs of scale-free topology model; (**c**) Clustering dendrogram for identifying cSCC-specific modules; (**d**) K Heatmap of the eigengene network indicates correlations between different modules; (**e**) Heatmap of associations among module eigengenes in normal and cSCC samples; (**f**) GO enrichment analysis of genes in the blue module; (**g**) KEGG analysis of genes in the blue module; (**h**) Scatter plots highlighting the association between GS and MM based on genes in the blue module.

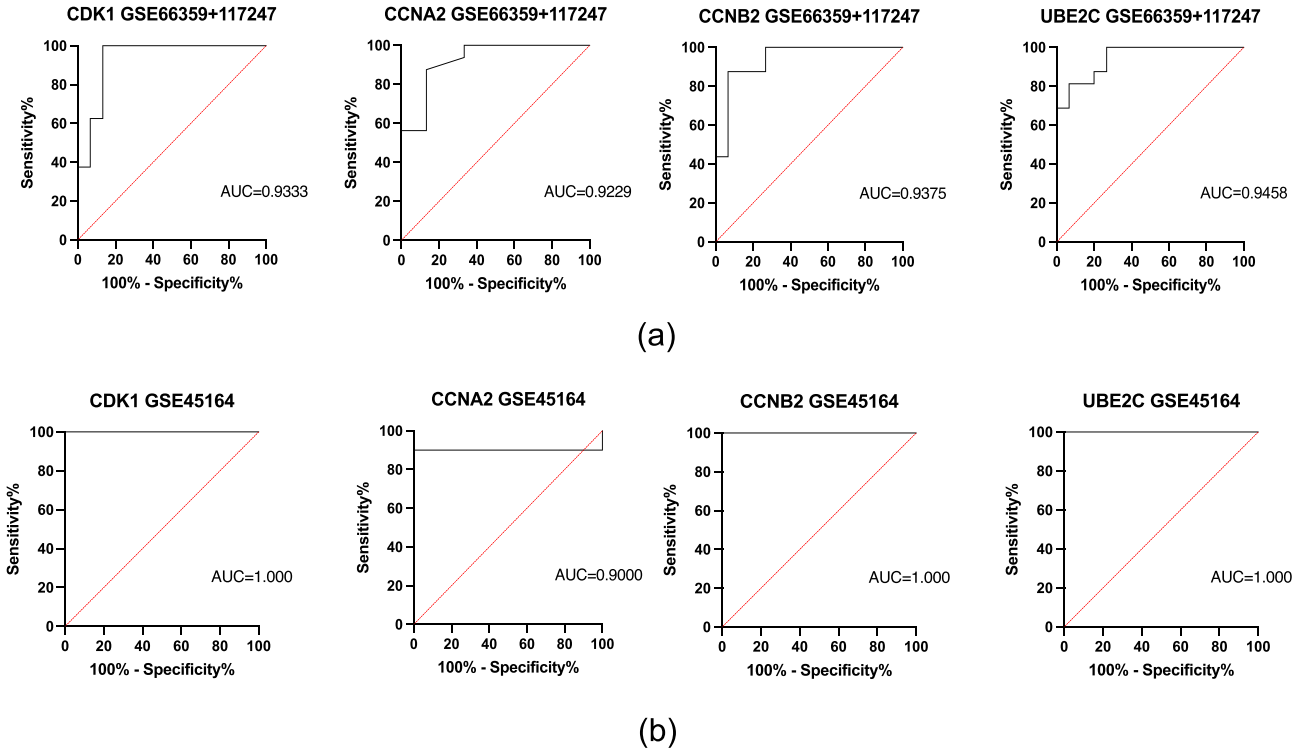

Figure 8 Validation of the diagnostic performance of the hub genes using internal and external datasets. (**a**) ROC curves of the hub genes in the GSE66359+117247 (internal); (**b**) ROC curves of the hub genes in the GSE45164 (external).

Immunohistochemistry

Literature retrieval indicates that CDK1 has been previously reported as an oncogenic gene in cSCC. However, the roles of CCNA2, CCNB2, and UBE2C in cSCC have yet to be reported. Therefore, we selected these three genes for further experimental validation. We performed immunohistochemistry staining on 30 cases of cSCC and adjacent normal tissues. The results revealed high expression of CCNA2, CCNB2, and UBE2C in cSCC tissues, while their expression was significantly reduced in adjacent normal tissues (Figure 9a). Paired *t*-test by prism demonstrated a significant difference (Figure 9b).

Construction of the ceRNA Network

MiRNAs induce gene silencing by binding to their target genes, while their upstream molecules, lncRNAs, can upregulate target gene expression by binding to miRNA response elements. This interaction is known as the ceRNA network. Our study employed five online miRNA databases (miRDB, miDIP, RNA22, TargetScan, and RNAInter) to predict the target miRNAs of hub genes. Based on these predictions, we constructed mRNA-miRNA regulatory networks using Cytoscape (Figure 10a). Subsequently, we employed the online database Starbase to predict lncRNAs interacting with the selected miRNAs. We selected lncRNAs present in most miRNA prediction results as our predicted lncRNAs. Finally, we obtained four target lncRNAs for CCNA2-targeting miRNAs, four target lncRNAs for CCNB2-targeting miRNAs, and three target lncRNAs for UBE2C-targeting miRNAs. Based on these predictions, we constructed three ceRNA networks (Figure 10b). Through literature searches, we selected two downregulated miRNAs and two upregulated lncRNAs in cSCC for further analysis. Specifically, hsa-miR-140-3p and hsa-miR-148a-3p were found to be downregulated in cSCC, while NEAT1 and H19 were upregulated in cSCC. Therefore, we propose that the NEAT1/H19-hsa-miR-148a-3p-CCNA2 and NEAT1-hsa-miR-140-3p-UBE2C axes may represent crucial pathways regulating the progression of cSCC (Figure 10c).

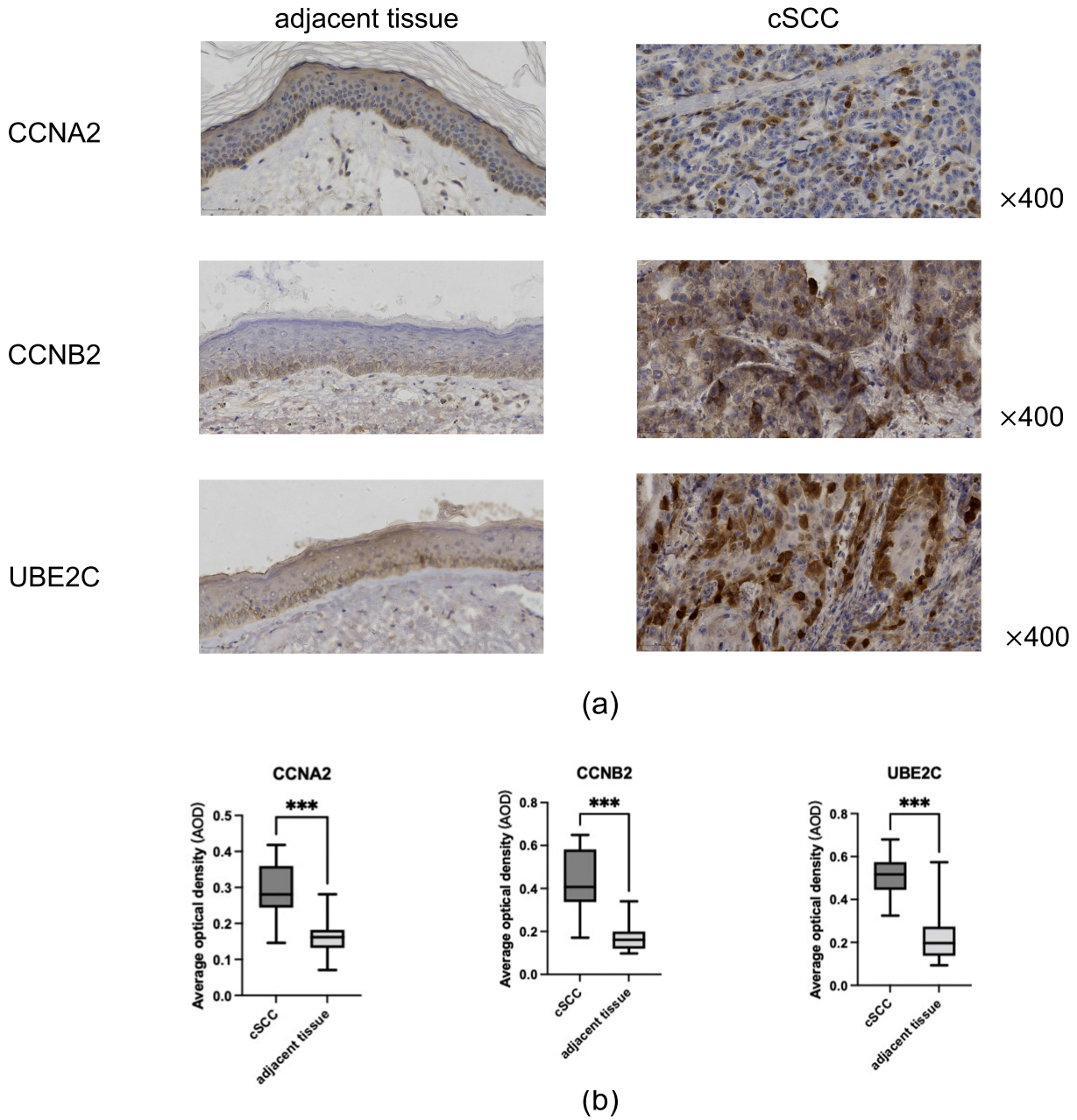

Figure 9 Expression patterns of CCNA2, CCNB2, UBE2C in cSCC and normal tissues. (**a**) Protein levels of CCNA2, CCNB2, UBE2C in cSCC and normal tissues; (**b**) Expression of CCNA2, CCNB2, UBE2C in 30 pairs of cSCC and adjacent normal tissues, *** $P < 0.001$.

Discussion

CSCC is a prevalent malignant skin tumor, and its incidence has steadily increased in recent years, garnering considerable attention. To date, there remains a scarcity of effective therapeutic strategies for cSCC. It is primarily attributable to the absence of precise molecular targets, resulting in a generally unfavorable prognosis. With the rapid development of bioinformatics, the field of life sciences has entered the era of big data characterized by multi-omics data, including transcriptomics. Compared to traditional experiments, bioinformatics research on transcriptomics helps to rapidly and accurately extract information. It enhances experimental efficiency, reduces costs, and plays a significant role in studying human gene function, human diseases, and precision medicine. Therefore, investigating the pathogenesis of cSCC at the genetic level holds promise for early diagnosis and the development of effective interventions to prevent cSCC progression. Consequently, we have explored potential biomarkers and molecular mechanisms associated with cSCC.

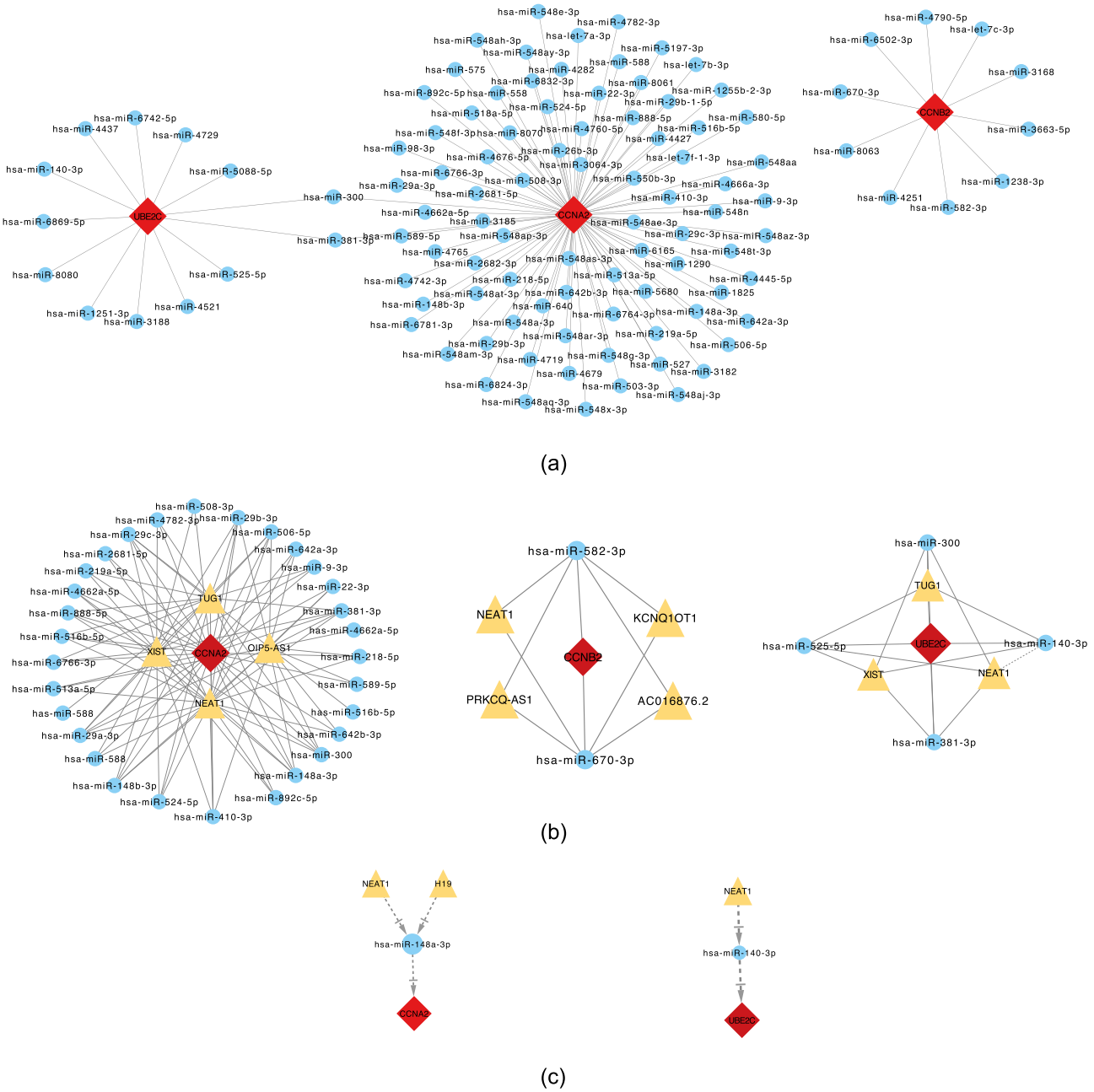

Figure 10 CeRNA networks. (**a**) The mRNA-miRNA co-expressed networks constructed by Cytoscape; (**b**) The ceRNA networks constructed by Cytoscape; (**c**) NEAT1/H19-hsa-miR-148a-3p-CCNA2 and NEAT1-hsa-miR-140-3p-UBE2C.

Initially, we screened DEGs in cSCC, identifying 505 upregulated genes and 522 downregulated genes. Subsequently, we conducted enrichment analyses on these genes. Upregulated DEGs were primarily associated with biological processes such as chromosome segregation and nuclear division, while downregulated DEGs were mainly involved in processes related to skin development, epidermal development, and keratinocyte differentiation, which are fundamental processes in cell proliferation. Our research results indicated that the upregulated DEGs between normal samples and cSCC were mainly associated with cell cycle, DNA replication, mismatch repair, and IL-17 signaling pathways. In fact, a previous study suggested that the development and progression of cSCC might be related to the recruitment and release of IL-17 by Th17 cells.[22]

To further explore the key genes involved in the pathogenesis of cSCC, we constructed a PPI network of DEGs. Utilizing five algorithms, namely Degree, MCC, MNC, Closeness, and Stress, we identified six genes highly correlated with cSCC: AURKA, BIRC5, CCNA2, CCNB2, CDK1, and UBE2C. We found that the expression patterns of these genes were upregulated in cSCC tissues compared to normal tissues, and these genes were also present in the top-scoring key module.

Subsequently, we employed WGCNA to identify key modules associated with cSCC. Within the WGCNA analysis, the blue module emerged as the most relevant module correlated with cSCC. Molecular function analysis revealed that genes within the blue module were enriched in pathways related to single-stranded DNA binding, catalytic activity on DNA, single-stranded DNA helicase activity, and microtubule binding. It has been reported that disruptions in microtubule dynamics can regulate cell proliferation in several human tumors,[23] including squamous cell carcinoma.[24] Cellular component analysis showed that the genes in the blue module were primarily associated with chromatin regions and condensed chromatin, indicating their essential roles in mitosis. KEGG analysis demonstrated that genes in the blue module were enriched in pathways related to DNA replication, the cell cycle, and the proteasome pathway. Notably, the GO/KEGG enrichment analysis results for the blue module closely resembled the most important module in the PPI network.

We identified four hub genes, CDK1, CCNA2, CCNB2, and UBE2C, as the intersection between the genes selected through WGCNA and PPI analyses. ROC analyses conducted on internal and external validation datasets indicated that these 4 hub genes hold significant diagnostic value, potentially serving as biomarkers for accurate diagnosis and treatment of cSCC. Literature research revealed that CDK1 was previously reported as an oncogene in cSCC.[25] However, the roles of CCNA2, CCNB2, and UBE2C in cSCC had not been previously documented. These genes have been confirmed to regulate the pathogenesis of other tumors, including squamous cell carcinoma,[26–30] suggesting a close association between the aberrant overexpression of hub genes and tumorigenesis. To confirm this, we conducted experimental validation in vitro after bioinformatics analysis, and our study revealed a significant upregulation of CCNA2, CCNB2, and UBE2C expression in cSCC patients.

We hypothesize that CCNA2, CCNB2, and UBE2C may serve as potential biomarkers for the diagnosis and treatment of cSCC. LncRNAs and miRNAs frequently interact in gene expression regulation. To identify regulatory factors associated with hub genes, we constructed lncRNA-miRNA-mRNA regulatory networks, elucidating the pathogenesis of cSCC at the transcriptome level. We conducted a literature search based on the ceRNA hypothesis and selected downregulated miRNAs in cSCC for further analysis. Among the targeted miRNAs for CCNA2, CCNB2, and UBE2C, hsa-miR-140-3p and hsa-miR-148a-3p have been confirmed to be downregulated in cSCC.[31,32] Additionally, it has been reported that two lncRNAs, NEAT1 and H19, are upregulated in cSCC.[33,34] NEAT1 can regulate cSCC cell proliferation,[33] while H19 plays a significant role in cSCC proliferation and invasion by promoting the epithelial-mesenchymal transition process.[34] Therefore, we propose that the NEAT1/H19-hsa-miR-148a-3p-CCNA2 and NEAT1-hsa-miR-140-3p-UBE2C axes may constitute a potential ceRNA regulatory pathway, modulating the progression of cSCC. However, our study has some limitations. The investigation into the pathogenesis was primarily conducted through computational simulations, and further in vivo and in vitro experiments will be necessary to validate these findings in the future.

Conclusion

In this study, we have identified CCNA2, CCNB2, and UBE2C as potential biomarkers for cSCC for the first time. Additionally, we have proposed novel molecular regulatory mechanisms, NEAT1/H19-hsa-miR-148a-3p-CCNA2 and NEAT1-hsa-miR-140-3p-UBE2C, which may offer prospects for targeted therapy with anti-tumor drugs.

Data Sharing Statement

The datasets analyzed in this study were acquired from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Ethical Statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Chinese PLA General Hospital (S2023-685-01). Informed consent was obtained from all subjects or their legal guardians for the participation in this study.

Acknowledgments

This research was funded by The National Science Foundation of China, grant number 81903227.

This paper has been uploaded to ResearchSquare as a preprint: https://www.researchsquare.com/article/rs-3511039/v1. The data is available from the corresponding author upon request.

Author Contributions

All authors made substantial contributions to the conception and design of the study, collection of data, or analyses or interpretation of data; took part in the writing of the manuscript; agreed to submit to the current journal; gave final approval of the version to be published; and agreed to be accountable for all aspects of work.

Disclosure

The author(s) report no conflicts of interest in this work.

References

1. Que SKT, Zwald FO, Schmults CD. Cutaneous squamous cell carcinoma: incidence, risk factors, diagnosis, and staging. *J Am Acad Dermatol*. 2018;78(2):237–247. doi:10.1016/j.jaad.2017.08.059
2. Roberto C, Natalia G, Rogelio G, et al. Cutaneous squamous cell carcinoma: from biology to therapy. *Int J Mol Sci*. 2020;21(8):2956. doi:10.3390/ijms21082956
3. Waldman A, Schmults C. Cutaneous squamous cell carcinoma. *Hematol Oncol Clin North Am*. 2019;33(1):1–12. doi:10.1016/j.hoc.2018.08.001
4. Hillen U, Leiter U, Haase S, et al. Advanced cutaneous squamous cell carcinoma: a retrospective analysis of patient profiles and treatment patterns-Results of a non-interventional study of the DeCOG. *Eur J Cancer*. 2018;96:34–43. PMID: 29665511. doi:10.1016/j.ejca.2018.01.075
5. Li CY, Cai JH, Tsai JJP, et al. Identification of hub genes associated with development of head and neck squamous cell carcinoma by integrated bioinformatics analysis. *Front Oncol*. 2020;10:681. doi:10.3389/fonc.2020.00681
6. Tian L, Chen T, Lu J, et al. Integrated protein-protein interaction and weighted gene co-expression network analysis uncover three key genes in hepatoblastoma. *Front Cell Dev Biol*. 2021;9:631982. doi:10.3389/fcell.2021.631982
7. Cao S, Wang X, Liu X, et al. Integrative analysis of angiogenesis-related long non-coding RNA and Identification of a Six-DEARlncRNA signature associated with prognosis and therapeutic response in esophageal squamous cell carcinoma. *Cancers*. 2022;14(17):4195. doi:10.3390/cancers14174195
8. Segundo-Val IS, Sanz-Lozano CS. Introduction to the GeneExpression Analysis. *Methods Mol Biol*. 2016;1434:29–43.
9. Zhao W, Langfelder P, Fuller T, et al. Weighted gene coexpression network analysis: state of the art. *J Biopharm Stat*. 2010;20(2):281–300. doi:10.1080/10543400903572753
10. Yamamura S, Imai-Sumida M, Tanaka Y, et al. Interaction and cross-talk between non-coding RNAs. *Cell Mol Life Sci*. 2018;75(3):467–484. doi:10.1007/s00018-017-2626-6
11. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–127. doi:10.1093/biostatistics/kxj037
12. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;20(3):307–315. doi:10.1093/bioinformatics/btg405
13. Wu J, Irizarry R, MacDonald J, et al. Gcrma: background adjustment using sequence information. *R Package Version*. 2012;2012:2200.
14. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007
15. Gene Ontology C. Gene ontology consortium: going forward. *Nucleic Acids Res*. 2015;43(Database issue):D1049–56.
16. Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457–62. doi:10.1093/nar/gkv1070
17. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287. doi:10.1089/omi.2011.0118
18. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638–D646. doi:10.1093/nar/gkac1000
19. Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol*. 2019;20(1):185. doi:10.1186/s13059-019-1758-4
20. Langfelder PS, Horvath S. Horvath WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559. doi:10.1186/1471-2105-9-559
21. Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92–7. doi:10.1093/nar/gkt1248
22. Sun L, Ko J, Vidimos A, et al. A distinctive lineage-negative cell population produces IL-17A in cutaneous squamous cell carcinoma. *J Interferon Cytokine Res*. 2020;40(8):418–424. doi:10.1089/jir.2020.0039
23. Zhang D, Zhang B, Zhou LX, et al. Deacetylisovaltratum disrupts microtubule dynamics and causes G2/M-phase arrest in human gastric cancer cells in vitro. *Acta Pharmacol Sin*. 2016;37(12):1597–1605. doi:10.1038/aps.2016.91
24. Gao L, Xue B, Xiang B, et al. Arsenic trioxide disturbs the LIS1/NDEL1/dynein microtubule dynamic complex by disrupting the CLIP170 zinc finger in head and neck cancer. *Toxicol Appl Pharmacol*. 2020;403:115158. doi:10.1016/j.taap.2020.115158

25. Qin S, Yang Y, Zhang HB, et al. Identification of CDK1 as a candidate marker in cutaneous squamous cell carcinoma by integrated bioinformatics analysis. *Transl Cancer Res*. 2021;10(1):469–478. doi:10.21037/tcr-20-2945
26. Jin Z, Zhao X, Cui L, et al. UBE2C promotes the progression of head and neck squamous cell carcinoma. *Biochem Biophys Res Commun*. 2020;523 (2):389–397. doi:10.1016/j.bbrc.2019.12.064
27. Li R, Pang XF, Huang ZG, et al. Overexpression of UBE2C in esophageal squamous cell carcinoma tissues and molecular analysis. *BMC Cancer*. 2021;21(1):996. doi:10.1186/s12885-021-08634-6
28. Li Y, Lin H, Chen L, et al. Novel therapies for tongue squamous cell carcinoma patients with high-grade tumors. *Lif*. 2021;11(8):813.
29. Li H, Weng Y, Wang S, et al. CDCA7 facilitates tumor progression by directly regulating CCNA2 expression in esophageal squamous cell carcinoma. *Front Oncol*. 2021;11:734655. doi:10.3389/fonc.2021.734655
30. Ma Q. MiR-219-5p suppresses cell proliferation and cell cycle progression in esophageal squamous cell carcinoma by targeting CCNA2. *Cell Mol Biol Lett*. 2019;24:4. doi:10.1186/s11658-018-0129-6
31. Sand M, Skrygan M, Georgas D, et al. Microarray analysis of microRNA expression in cutaneous squamous cell carcinoma. *J Dermatol Sci*. 2012;68(3):119–126. doi:10.1016/j.jdermsci.2012.09.004
32. Zhang Z, Deng X. circ_0001821 contributes to the development of cutaneous squamous cell carcinoma by regulating MicroRNA-148a-3p/EGFR axis and activating phosphatidylinositol 3-Kinase/Akt Pathway. *Mol Cell Biol*. 2022;42(3):e0008921. doi:10.1128/mcb.00089-21
33. Jiang S, Liu H, Zhang J, et al. MMP1 regulated by NEAT1/miR-361-5p axis facilitates the proliferation and migration of cutaneous squamous cell carcinoma via the activation of Wnt pathway. *Cancer Biol Ther*. 2021;22(5–6):381–391. doi:10.1080/15384047.2021.1941583
34. Zhang W, Zhou K, Zhang X, et al. Roles of the H19/microRNA-675 axis in the proliferation and epithelial-mesenchymal transition of human cutaneous squamous cell carcinoma cells. *Oncol Rep*. 2021;45(4):39. doi:10.3892/or.2021.7990

OncoTargets and Therapy Dovepress

Publish your work in this journal

OncoTargets and Therapy is an international, peer-reviewed, open access journal focusing on the pathological basis of all cancers, potential targets for therapy and treatment protocols employed to improve the management of cancer patients. The journal also focuses on the impact of management programs and new therapeutic agents and protocols on patient perspectives such as quality of life, adherence and satisfaction. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/oncotargets-and-therapy-journal