

Identification of novel drug candidates for treating tongue squamous cell carcinoma using computational approaches

Yinghua Li

Southern Medical University

Zihao Chen

Southern Medical University

Lu Chen

Baotou Medical College

Weizhong Li (✉ drweizhong@163.com)

Southern Medical University <https://orcid.org/0000-0001-7188-0904>

Research

Keywords: TSCC, WGCNA, hub proteins, virtual screening, drug candidates

Posted Date: February 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-177053/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Tongue squamous cell carcinoma (TSCC) is one of the most common oral squamous cell carcinoma (OSCC) with a high occurrence and a poor prognosis, yet its molecular mechanisms are largely unknown and novel drug candidates are needed. The purpose of this study was to construct gene co-expression networks to identify hub proteins significantly associated with tumor grades and the overall survival (OS) of TSCC patients and provide potential drug candidates.

Methods

The mRNA-sequencing and clinical data were obtained from The Cancer Genome Atlas (TCGA) dataset. Weighted gene co-expression network analysis (WGCNA) was used for identifying the co-expression module related to the tumor grade of TSCC. The hub proteins were selected by the interaction number with other proteins, and their correlations with prognosis and tumor grades were calculated. Virtual screening of compounds by the hub protein structures was used to identify the drug candidates.

Results

WGCNA identified ten co-expression modules, in which the brown module consisted of 163 genes was most significantly correlated with the tumor grade. Six hub genes/proteins (BUB1, CCNB2, CDC6, CDC20, CDK1, and MCM2) tended to be in the central hub of the network. And higher expression levels of these hub genes were associated with tumor grades and worse overall survival. Three compounds, targeting hub proteins, demonstrated high binding affinities, favorable pharmacologic properties, and low toxicity.

Conclusion

The gene co-expression network-based study could provide additional insight into tumorigenesis and progression of TSCC, and our study might provide novel drug candidates.

1. Introduction

Oral squamous cell carcinoma (OSCC) represents the eighth most common malignancy in the world, with a poor prognosis of < 60% patient survival for 5-year survival rates [1]. In 2018, 354,864 new cases of lip and oral cavity cancer were identified, and 177,384 people died from these types of cancer worldwide [2]. OSCC can occur due to many etiological factors, but a majority of these cancers are associated with lifestyle risk habits including smoking, excessive alcohol consumption, and betel quid chewing [3]. The most common subtype of OSCC is tongue squamous cell carcinoma (TSCC), which is also associated with a poor treatment outcome [4]. TSCC comprises 41% of the OSCC cases and has much more aggressive clinical behaviors and a worse prognosis than other cancers of the oral cavity, which makes TSCC one of the most lethal cancer types [5-6].

The conventional approaches for TSCC treatment involve surgery, radiation therapy, and chemotherapy [7]. However, surgical resection will lead to permanent disfigurement, altered sense of self and debilitating physiological consequences, substantial functional impairment, and morbidity, while chemo- and radio-therapies result in significant toxicities, all affecting patient wellbeing and quality of life. Besides, even with combined treatment involving surgery, radiation, and chemotherapy, the 5-year survival rate is still unsatisfactory [8]. Recently, increased

expression of PD-1/PD-L1 and induced immune suppression status was reported in TSCC tissues [9]. But, implementation of immunotherapy could be challenging since the auto-immune side effects, low response rate, and financial cost [10]. Thus, the identification of novel therapeutic targets and potential drug candidates is urgently needed.

In recent years, microarray technology is widely employed to identify hub genes/proteins and abnormal regulation pathways of cancers [11]. Through microarray technology, biological function and target genes related to tumor development may bring to reveal the underlying mechanisms of cancers and identify potential therapeutic targets [12]. After elucidating and validating fundamental potential therapeutic targets, investigators could further perform molecular docking to screen small molecular drugs that could combine with selected targets [13]. Virtual screening and molecular docking were widely applied on drug discovery and design and medicinal chemistry [14]. The binding sites, binding affinity, and a series of pharmacological properties could also be calculated by virtual screening analysis [15]. Therefore, microarray technology combined with virtual screening analysis could accelerate drug discovery.

In this study, bioinformatics analysis methods were used to analyze gene expression profiles of TSCC cases from The Cancer Genome Atlas database (TCGA). Differentially Expressed Genes (DEGs) and their enriched pathways between tumor and normal control were identified. Subsequently, weighted gene co-expression network analysis (WGCNA) of DEGs was applied and the gene module related to tumor development (tumor grades) was identified. Furthermore, survival analysis was applied to screen out hub genes related to overall survival. Next, we applied molecular docking method to identify three FDA approved drugs, which are novel small molecule inhibitor of hub proteins, as the potential drug for TSCC.

2. Material And Methods

2.1 Data and Pre-processing and Differentially Expressed Genes (DEGs)

The mRNA-Seq and clinical data from patients with TSCC were obtained from The Cancer Genome Atlas (TCGA) public data portal (<https://cancergenome.nih.gov>). The expression data containing 19645 mRNAs from 126 tumor and 13 adjacent normal tissue samples were selected and further analyzed. 'edgeR' [16], the R package designed to calculate DEGs, was used to identify differentially expressed mRNAs between the tumor and normal samples. The DEGs with $\log_2\text{Foldchange}>1$ and $P.\text{value}<0.05$ were considered for subsequent analysis.

2.2 Functional and Pathway Enrichment Analysis of DEGs

The R package of 'clusterProfiler' [17], which is a comprehensive set of functional annotation tools, has been used for systematic and integrative analysis of large gene lists. In this work, the significant gene ontology (GO) biological process terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of DEGs were performed using 'clusterProfiler' with the threshold of $P.\text{value}<0.05$.

2.3 Weighted Gene Co-expression Network Analysis (WGCNA)

Processed by the R package "WGCNA" [18], a scale-free gene co-expression network on DEGs was constructed. The clustering analysis was performed based on DEGs expression matrix of 117 tumor samples since 9 outlier tumor samples were removed. The soft threshold power β was selected in accordance with scale-free topology fitting indices $R^2 (>0.9)$ and mean connectivity for maximizing scale-free topology, maintaining a high average number of

connections, and eliminating small correlations. A hierarchical clustering dendrogram was plotted with identified gene modules after setting minModulesize of 60 and mergeCutHeight of 0.25. Genes in a module were thought to share similar expression patterns. The module eigengene (ME) was the first principal component of a module and it represented the gene expression profiling. Module-trait associations were estimated using MEs and clinical-trait to identify gene modules highly correlated to phenotype.

2.4 Identification hub genes

The Search Tool for the Retrieval of Interacting Genes (STRING) database [19] and Cytoscape software [20] were utilized to construct PPI (Protein-Protein Interaction) networks on genes from the selected module. PPI networks from experiments were selected with the cut-off criterion of confidence score ≥ 0.7 . Hub genes, highly interconnected with other genes in a module, have been shown to be functionally significant [21]. In this study, hub genes were defined by module connectivity, measured by absolute value of degree (>10).

2.5 Survival Analysis of Hub genes

We further assessed whether hub genes were correlated with the overall survival of TSCC patients. Patients were classified into high or low expression groups based on the median expression level of the genes. We performing survival analysis using Kapla-Meier curve (K-M curve) method [22]. Log-rank analysis was used to compare two groups, P .value <0.05 is considered to represent statistical significance.

2.6 Preparation for the virtual screening

The material for virtual screening contains the structures of proteins and ligands. Three-dimensional (3D) structure of the protein corresponding to the hub gene is crucial for its interaction with other molecules and biological functions. 3D structures of proteins and their binding sites could be directly obtained from Protein Data Bank (PDB) dataset. In the current study, FDA approved drugs were selected as the ligands. The FDA approved drugs (a total of 1615 compounds) in the ZINC15 database [23] were downloaded and selected as the resource for potential hits in our virtual screening.

2.7 Structure-based virtual screening using LibDock

Virtual screening was carried out using the LibDock module of Discovery Studio 2019. LibDock [24] is a rigid-based docking module and it calculates hotspots for the protein through a grid placed into the binding site and polar probes. The hotspots are further used to align the ligands to form favorable interactions. The Smart Minimiser algorithm and CHARMM force field were performed for ligand minimization, and all ligand poses were ranked based on the ligands score.

2.8 ADME (Adsorption, Distribution, Metabolism, and Excretion) Properties and Toxicity Prediction

As a successful small molecular drug, it should not only be active against a target but also possess appropriate ADME and toxicity properties [25]. Pharmacologic properties of all the 1615 FDA approved drugs were predicted using the ADME module of Discovery Studio 2019, including aqueous solubility level, blood-brain barrier level, CYP2D6 binding, hepatotoxicity, human intestinal absorption level, and plasma protein binding properties.

To confirm the safety of 1615 FDA approved drugs (compounds), different kinds of toxicity indicators including rodent carcinogenicity, developmental toxicity potential properties, Ames mutagenicity were predicted by using a

computational method in the TOPKAT module of Discovery Studio 2019.

3. Results

3.1 Differentially Expressed Genes (DEGs) and Enriched pathways

A total of 126 tumor and 13 normal matched samples were included in DEGs analysis. DEGs analysis was performed using R package 'edgeR' and 3728 genes (1526 upregulated and 2202 downregulated genes) were identified with $P.value < 0.05$ and $\log_2 foldchange > 1$ as the cutoff criterion (**Figure 1A-B**). The locations of the 100 DEGs with the lowest P -value were visualized on the human chromosomes using the 'Rcircos' package (**Figure 1C**). All DEGs were used to identify overrepresented GO terms and KEGG pathways (**Table 1**). GO analysis results showed that the most overrepresented GO terms in biological processes were enriched in *mitotic nuclear division* and *chromosome segregation*. In addition, the most enriched GO terms in molecular function and cellular component were *DNA-dependent ATPase activity* and *chromosomal region*, respectively. KEGG results showed that the most enriched KEGG pathway terms were *Cell cycle*, *DNA replication*, *p53 signaling pathway*, *Human T-cell leukemia virus 1 infection*, and *Viral carcinogenesis*.

3.2 Weighted Gene Co-expression Network Analysis (WGCNA)

After removed the outlier tumor samples, a total of 117 tumor samples were selected to do the WGCNA (**Figure S1A**). According to the scale-free topology fitting indices R^2 and mean connectivity generated from 3728 DEGs expressions of 117 samples, threshold soft power β was set to be 5 (**Figure 2A-B**). All DEGs were submitted to WGCNA to construct the gene co-expression network and assigned to different modules by clustering dendrogram trees. We got 10 modules After setting the parameters and the module size were range from 127 to 825 (**Figure 2C**). The relationships between clinical-trait and gene modules were presented in **Figure 2D**. In the WGCNA results, the brown module showed a significant relationship to tumor grade and contains 263 genes. The genes from the brown module showed positive correlation with the tumor grade (**Figure S1B**).

3.3 Identification of protein-protein interactions (PPIs) and hub genes

In addition, PPIs of genes from the brown module were examined using the STRING database. We found that 40 genes formed a complex functional network, indicating that each of them has at least one functionally similar or interacted gene as the neighbor (**Fig. 3**). Remarkably, six genes (*BUB1*, *CCNB2*, *CDC6*, *CDC20*, *CDK1*, and *MCM2*) tended to be in the central hub of the network generated using the STRING database, thereby demonstrating the importance of these genes. We also evaluated the prognostic significance of six hub genes in patients from TCGA dataset. The K-M survival analysis revealed that the higher expression levels of *BUB1*, *CCNB2*, *CDC6*, *CDC20*, *CDK1*, and *MCM2* were associated with the worse overall survival (**Fig. 4A-F**). The boxplots demonstrated that these hub genes were all highly expressed in tumor samples compared with the normal samples (**Figure S2**). The hub genes were also positively correlated with the tumor grades (**Figure S3**). Overall, these results suggested that the hub genes might play an oncogenic role in TSCC occurrence and development.

3.4 Virtual screening of compounds

The 3D structure of the protein is crucial for its interaction with other molecules and biological functions. In the current study, the 3D structures of hub proteins (*CDC20*, *CDK1*, and *MCM2*) were obtained from 5LCW, 4YC3, and 6XTX of Protein Data Bank dataset (**Fig. 5**). The active sites of *CDC20*, *CDK1*, and *MCM2* were defined with

SER377:HIS380, ALA145:PHE147, and PRO525:GLN531. A total of 1615 FDA approved drugs were taken from the ZINC15 database. After virtual screening by Libdock, 541, 841, and 1591 drugs were found to be eligible to bind stably with CDC20, CDK1, and MCM2, respectively. The top 3 drugs with the highest Libdock score of hub proteins were listed in **Table 2**.

3.5 Pharmacologic properties of compounds

Pharmacologic properties of all 9 selected compounds in **Table 2** were predicted using the ADME module of Discovery Studio 2019, including aqueous solubility level, blood-brain barrier level, CYP2D6 binding, hepatotoxicity, human intestinal absorption level, and plasma protein binding properties (**Table 3**). The aqueous solubility prediction (defined in water at 25°C) indicated that all the compounds were soluble in water. Only 4 compounds were predicted to be inhibitors with CYP2D6, which was an essential enzyme in drug metabolism. As to hepatotoxicity, 3 compounds were found to be toxic. For human intestinal absorption, 6 compounds have good absorption level.

Toxicity results indicated that 3 compounds did not have developmental toxicity potential (**Table 4**). Considering all the results in **Table 3** and **Table 4**, ZINC000100052685, ZINC000008214703 and ZINC000085537014 were identified to be the ideal leading compounds with high Solubility level, non-CYP2D6 inhibitors, less rodent carcinogenicity, together with Ames mutagenicity, and developmental toxicity potential compared with other compounds. Therefore, ZINC000100052685, ZINC000008214703, and ZINC000085537014 were confirmed as safe drug candidates and selected for the subsequent research.

3.6 Visualization of docking results from Libdock

There are 2 hydrogen bond interactions observed in the complex of ZINC000100052685 with CDC20 (**Figure 6A**) and 14 van der Waals interactions offered by **Figure 7A**. Four hydrogen bond interactions were observed in the complex of ZINC000008214703 with CDK1 (**Figure 6B**), as well as 14 van der Waals interactions offered by **Figure 7B**. Five hydrogen bond interactions were observed in the complex of ZINC000085537014 with MCM2 (**Figure 6C**), as well as 14 van der Waals interactions offered by **Figure 7C**.

4. Discussion

Although surgical treatment, radiotherapy, and chemotherapy are constantly developing, the mortality rate of TSCC in recent decades is still high [8]. A more in-depth exploration of the molecular mechanisms will help us understand more about TSCC and find better molecular targets for prevention, detection, and treatment. Bioinformatics develop rapidly in the past 10 years, which is a reliable method for exploring gene expression in disease and finding potential molecular targets [26-27]. This study used computational methods such as bioinformatics analysis and virtual screening to identify the potential targets and drug candidates for TSCC.

The characteristics are different in the development and metastasis in different locations of HNSCC, which may lead to inconsistency in the results. To avoid interference, we selected TSCC patients as specific subjects in the present work. Moreover, WGCNA calculates the relationship between genes not based on simple genes counting but their expression changes, which means results are less fluctuated [18]. Genes with similar expression changes, carrying common biological processes, are grouped into the same module. Then relationship coefficients between gene modules and clinical phenotypes can be calculated. Currently, WGCNA has been widely applied to finding the hub genes associated with clinical features in cancers [21].

In this work, we identified 6 potential gene targets in TSCC: *BUB1*, *CCNB2*, *CDC6*, *CDC20*, *CDK1*, and *MCM2*. *BUB1* encodes a serine/threonine-protein kinase that plays a central role in mitosis [28], and *BUB1* overexpression induces tumor formation [29]. In previous studies, *CCNB2* was observed to be the hub gene in smoking of head and neck cancer [30]. Cell division cycle 6 (*CDC6*) is an essential regulator of DNA replication in eukaryotic cells, deregulation of *CDC6* expression in human cells poses a serious risk of carcinogenesis [31]. Downregulation of *CDC6* effectively inhibited the proliferation of TSCC cells [32]. High *CDC20* expression is associated with poor prognosis in oral squamous cell carcinoma [33]. And the previous study showed that development of specific *CDC20* inhibitors could be a novel strategy for the treatment of human cancers [34]. The expression of *CDK1* was significantly correlated with the histological grade of OSCC and the *CDK1* protein was over-expressed in recurrent tumors or in those with lymph node metastasis [35]. Moreover, a significant reduction in the 5-year accumulative survival rate in *CDK1* positive cases compared with *CDK1* negative cases [35]. *CDK1* inhibitor could selectively activate the intrinsic apoptosis pathway and induce apoptosis in cancer cells rather than in normal cells [36]. Quantitative real-time PCR analysis showed that *MCM2* mRNA expression is significantly higher in tongue SCC than in epithelial dysplasia [37]. And the knockdown of *MCM2* has therapeutic applications in enhancing the efficacy of chemotherapy in cancer patients [38].

We have identified three potential inhibitors of hub proteins using high-throughput virtual screening. Three compounds ZINC000100052685, ZINC000008214703, and ZINC000008214703 bind to the common residues of the active site cavity of the *CDC20*, *CDK1*, and *MCM2*, respectively. The AMDET properties of three compounds show favorable pharmacological properties and less toxicity. ZINC000100052685, named as Iloprost, is a synthetic analogue of prostacyclin that affects platelet aggregation and is used to treat pulmonary arterial hypertension (PAH) [39]. Iloprost inhibits the invasion of ovarian cancer cells by downregulating matrix metalloproteinase-2 (MMP-2) [40]. ZINC000008214703 (Unoprostone), a prostaglandin analogue, is believed to reduce elevated intraocular pressure (IOP) and is used to treat Glaucoma [41]. ZINC000085537014 (Cobicistat), marketed under the name Tybost (formerly GS-9350), is indicated for treating the infection with human immunodeficiency virus (HIV) [42]. Cobicistat has been used to enhance the efficacy of anticancer agents [43]. In conclusion, compounds ZINC000100052685, ZINC000008214703, and ZINC000008214703 were anticipated to be promising drug candidates for inhibition of TSCC.

5. Conclusion

In summary, by comprehensively analyzing gene expression profiles of TSCC and adjacent tissues, this study identified the potential therapeutic targets and drug candidates of TSCC. We found that six hub genes (*BUB1*, *CCNB2*, *CDC6*, *CDC20*, *CDK1*, and *MCM2*) played crucial roles in TSCC occurrence and development. Our research explained the pathogenesis of TSCC from the perspective of bioinformatics, and provided novel drug candidates for TSCC. However, further experimental studies are still required to prove our findings and determine the potential clinical value of these drug candidates.

Abbreviations

Tongue squamous cell carcinoma (TSCC); oral squamous cell carcinoma (OSCC); TCGA, The Cancer Genome Atlas; Differentially expressed genes (DEGs); WGCNA, Weighted gene co-expression network analysis; PPI, protein-protein interactions; K-M, Kaplan-Meier; MCC, Matthews correlation coefficient; ORR, objective response rate; KEGG, Kyoto Encyclopedia of genes and Genomes; GO, Gene Ontology; OS, Overall survival; MEs, module eigengenes; The Search Tool for the Retrieval of Interacting Genes (STRING); Protein Data Bank (PDB); Adsorption, Distribution, Metabolism

and Excretion (ADME); Cell division cycle 6 (CDC6); developmental toxicity potential (DTP); human immunodeficiency virus (HIV).

Declarations

Ethical Approval and Consent to participate

All the expression data and clinical information were retrieved from publicly available datasets which were free to download and analyze without limitations. Investigators of each study obtained the approval from their local ethics committee and informed patient consent.

Consent for publication

Not applicable.

Availability of supporting data

The authors declare that the data and code that support the findings of this study are available upon request from drweizhong@163.com.

Competing interests

The authors state that they have no conflicts of interest

Funding

The project was supported by the Guangdong Natural Science Foundation of China (2015A030313309).

Authors' contributions

Yinghua Li and Zihao Chen performed the study and wrote the manuscript; Lu Chen contributed to data preparation; Weizhong Li performed technical modification and conceived the study. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

¹Department of Oral and Maxillofacial Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, People's Republic of China. ²Department of Urology, Nanfang Hospital, Southern Medical University, Guangzhou, People's Republic of China. ³The First Affiliated Hospital, Baotou Medical College, Baotou, People's Republic of China.

References

1. Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin* 2011; 61: 212-236.

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394-424.
3. Chai A, Lim KP, Cheong SC. Translational genomics and recent advances in oral squamous cell carcinoma. *Semin Cancer Biol* 2020; 61: 71-83.
4. Listl S, Jansen L, Stenzinger A, Freier K, Emrich K, Holleczeck B *et al.* Survival of patients with oral cavity cancer in Germany. *Plos One* 2013; 8: e53415.
5. Patel RS, Clark JR, Dirven R, Wyten R, Gao K, O'Brien CJ. Prognostic factors in the surgical treatment of patients with oral carcinoma. *Anz J Surg* 2009; 79: 19-22.
6. Bello IO, Soini Y, Salo T. Prognostic evaluation of oral tongue cancer: means, markers and perspectives (II). *Oral Oncol* 2010; 46: 636-643.
7. Ketabat F, Pundir M, Mohabatpour F, Lobanova L, Koutsopoulos S, Hadjiiski L *et al.* Controlled Drug Delivery Systems for Oral Cancer Treatment-Current Status and Future Perspectives. *Pharmaceutics* 2019; 11.
8. Zumsteg ZS, Cook-Wiens G, Yoshida E, Shiao SL, Lee NY, Mita A *et al.* Incidence of Oropharyngeal Cancer Among Elderly Patients in the United States. *Jama Oncol* 2016; 2: 1617-1623.
9. Yu GT, Bu LL, Huang CF, Zhang WF, Chen WJ, Gutkind JS *et al.* PD-1 blockade attenuates immunosuppressive myeloid cells due to inhibition of CD47/SIRPalpha axis in HPV negative head and neck squamous cell carcinoma. *Oncotarget* 2015; 6: 42067-42080.
10. Lubek JE. Head and Neck Cancer Research and Support Foundations. *Oral Maxillofac Surg Clin North Am* 2018; 30: 459-469.
11. Shi W, Hu D, Lin S, Zhuo R. Five-mRNA Signature for the Prognosis of Breast Cancer Based on the ceRNA Network. *Biomed Res Int* 2020; 2020: 9081852.
12. Liu G, Chen Z, Danilova IG, Bolkov MA, Tuzankina IA, Liu G. Identification of miR-200c and miR141-Mediated lncRNA-mRNA Crosstalks in Muscle-Invasive Bladder Cancer Subtypes. *Front Genet* 2018; 9: 422.
13. Chen XR, Wang XT, Hao MQ, Zhou YH, Cui WQ, Xing XX *et al.* Homology Modeling and Virtual Screening to Discover Potent Inhibitors Targeting the Imidazole Glycerophosphate Dehydratase Protein in *Staphylococcus xylosum*. *Front Chem* 2017; 5: 98.
14. Kontoyianni M. Docking and Virtual Screening in Drug Discovery. *Methods Mol Biol* 2017; 1647: 255-266.
15. Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS, Olson AJ. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin Drug Discov* 2010; 5: 597-607.
16. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; 26: 139-140.
17. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 2012; 16: 284-287.
18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* 2008; 9: 559.
19. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017; 45: D362-D368.
20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498-2504.

21. Chen Z, Liu G, Hossain A, Danilova IG, Bolkov MA, Liu G *et al.* A co-expression network for differentially expressed genes in bladder cancer and a risk score model for predicting survival. *Hereditas* 2019; 156: 24.
22. Adachi K, Tamada K. [Genome-Based Biomarkers of Immune Checkpoint Blockade Therapy]. *Gan To Kagaku Ryoho* 2016; 43: 143-150.
23. Sterling T, Irwin JJ. ZINC 15–Ligand Discovery for Everyone. *J Chem Inf Model* 2015; 55: 2324-2337.
24. Zhou X, Yu S, Su J, Sun L. Computational Study on New Natural Compound Inhibitors of Pyruvate Dehydrogenase Kinases. *Int J Mol Sci* 2016; 17: 340.
25. Alqahtani S. In silico ADME-Tox modeling: progress and prospects. *Expert Opin Drug Metab Toxicol* 2017; 13: 1147-1158.
26. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; 5: R80.
27. Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB. Bioinformatics challenges for personalized medicine. *Bioinformatics* 2011; 27: 1741-1748.
28. Dun B, Sharma A, Xu H, Liu H, Bai S, Zeng L *et al.* Transcriptomic changes induced by mycophenolic acid in gastric cancer cells. *Am J Transl Res* 2013; 6: 28-42.
29. Ricke RM, Jeganathan KB, van Deursen JM. Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *J Cell Biol* 2011; 193: 1049-1064.
30. Shaikh I, Ansari A, Ayachit G, Gandhi M, Sharma P, Bhairappanavar S *et al.* Differential gene expression analysis of HNSCC tumors deciphered tobacco dependent and independent molecular signatures. *Oncotarget* 2019; 10: 6168-6183.
31. Borlado LR, Mendez J. CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis* 2008; 29: 237-243.
32. Feng CJ, Lu XW, Luo DY, Li HJ, Guo JB. Knockdown of Cdc6 inhibits proliferation of tongue squamous cell carcinoma Tca8113 cells. *Technol Cancer Res Treat* 2013; 12: 173-181.
33. Moura IM, Delgado ML, Silva PM, Lopes CA, Do AJ, Monteiro LS *et al.* High CDC20 expression is associated with poor prognosis in oral squamous cell carcinoma. *J Oral Pathol Med* 2014; 43: 225-231.
34. Wang L, Zhang J, Wan L, Zhou X, Wang Z, Wei W. Targeting Cdc20 as a novel cancer therapeutic strategy. *Pharmacol Ther* 2015; 151: 141-151.
35. Chen X, Zhang FH, Chen QE, Wang YY, Wang YL, He JC *et al.* The clinical significance of CDK1 expression in oral squamous cell carcinoma. *Med Oral Patol Oral Cir Bucal* 2015; 20: e7-e12.
36. Zhang S, Bao Y, Ju X, Li K, Shang H, Ha L *et al.* BA-j as a novel CDK1 inhibitor selectively induces apoptosis in cancer cells by regulating ROS. *Sci Rep* 2015; 5: 13626.
37. Li JN, Feng CJ, Lu YJ, Li HJ, Tu Z, Liao GQ *et al.* mRNA expression of the DNA replication-initiation proteins in epithelial dysplasia and squamous cell carcinoma of the tongue. *Bmc Cancer* 2008; 8: 395.
38. Deng M, Sun J, Xie S, Zhen H, Wang Y, Zhong A *et al.* Inhibition of MCM2 enhances the sensitivity of ovarian cancer cell to carboplatin. *Mol Med Rep* 2019; 20: 2258-2266.
39. Hill NS, Preston IR, Roberts KE. Inhaled Therapies for Pulmonary Hypertension. *Respir Care* 2015; 60: 794-802, 802-805.
40. Ahn JH, Lee KT, Choi YS, Choi JH. Iloprost, a prostacyclin analog, inhibits the invasion of ovarian cancer cells by downregulating matrix metalloproteinase-2 (MMP-2) through the IP-dependent pathway. *Prostaglandins Other Lipid Mediat* 2018; 134: 47-56.

41. Fung DS, Whitson JT. An evidence-based review of unoprostone isopropyl ophthalmic solution 0.15% for glaucoma: place in therapy. *Clin Ophthalmol* 2014; 8: 543-554.
42. von Hentig N. Clinical use of cobicistat as a pharmacoenhancer of human immunodeficiency virus therapy. *HIV AIDS (Auckl)* 2016; 8: 1-16.
43. Hohmann N, Bozorgmehr F, Christopoulos P, Mikus G, Blank A, Burhenne J *et al.* Pharmacoenhancement of Low Crizotinib Plasma Concentrations in Patients with Anaplastic Lymphoma Kinase-Positive Non-Small Cell Lung Cancer using the CYP3A Inhibitor Cobicistat. *Clin Transl Sci* 2020.

Tables

Table 1. Functional and pathway enrichment analysis of DEGs in TSCC.

ID	Term	Count	P-value
Has:04510	Cell cycle	22	<0.01
Has:04512	DNA replication	7	<0.01
Has:05200	p53 signaling pathway	8	<0.01
Has:05205	Human T-cell leukemia virus 1 infection	9	<0.01
Has:04151	Viral carcinogenesis	8	<0.01
GO_BP:0030198	nuclear division	59	<0.01
GO_BP:0030574	mitotic nuclear division	51	<0.01
GO_BP:0007155	organelle fission	60	<0.01
GO_BP:0022617	chromosome segregation	52	<0.01
GO_BP:0030335	sister chromatid segregation	40	<0.01
GO_CC:0070062	chromosomal region	51	<0.01
GO_CC:0031012	chromosome, centromeric region	41	<0.01
GO_CC:0005578	kinetochore	35	<0.01
GO_CC:0016020	condensed chromosome	39	<0.01
GO_CC:0005788	condensed chromosome, centromeric region	32	<0.01
GO_MF:0005178	single-stranded DNA-dependent ATPase activity	7	<0.01
GO_MF:0005201	DNA-dependent ATPase activity	12	<0.01
GO_MF:0005515	microtubule binding	14	<0.01
GO_MF:0004222	catalytic activity, acting on DNA	12	<0.01
GO_MF:0005102	ATPase activity, coupled	14	<0.01

Notes: Top five terms were selected according to P-value.

Abbreviations: DEGs, different expressed genes; TSCC, tongue squamous cell carcinoma; GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function.

Table 2. The Sequence and structures used in obtaining structures of hub proteins

Protein	ZincID	Libdock score	Smile
CDC20	ZINC000003799072	116.1	<chem>OCc1cc([C@@H](O)CNCCCCCOCCCCc2ccccc2)ccc1O</chem>
CDC20	ZINC000100052685	105.9	<chem>CC#CC[C@H](C)[C@@H](O)/C=C/[C@@H]1[C@@H](O)C[C@@H]2C/C(=C/CCCC(=O)OCC(=O)c3ccccc3)C[C@H]21</chem>
CDC20	ZINC000003785268	105.6	<chem>OCc1cc([C@H](O)CNCCCCCOCCCCc2ccccc2)ccc1O</chem>
CDK1	ZINC000008214703	126.7	<chem>CCCCCCC(=O)CC[C@H]1[C@H](O)C[C@H](O)[C@@H]1C/C=C\CCCC(=O)O</chem>
CDK1	ZINC000021297660	113.4	<chem>COc1c(C)c2c(c(O)c1C/C=C\C)CCC(=O)OCCN1CCOCC1)C(=O)OC2</chem>
CDK1	ZINC000000601250	111.8	<chem>Clc1ccc(CS[C@H](Cn2ccnc2)c2ccc(Cl)cc2Cl)cc1</chem>
MCM2	ZINC000028232755	171.2	<chem>CCCC(=O)OCC(=O)[C@]1(O)Cc2c(O)c3c(c(O)c2[C@@H](O[C@H]2C[C@H](NC(=O)C(F)(F)F)[C@@H](O)[C@@H](C)O2)C1)C(=O)c1c(OC)cccc1C3=O</chem>
MCM2	ZINC000085537014	166.8	<chem>CC(C)c1nc(CN(C)C(=O)N[C@@H](CCN2CCOCC2)C(=O)N[C@H](CC[C@H](Cc2ccccc2)NC(=O)OCc2cnsc2)Cc2ccccc2)cs1</chem>
MCM2	ZINC000028639340	161.7	<chem>CC[C@@H]([C@H](C)O)n1ncn(-c2ccc(N3CCN(c4ccc(OC[C@H]5CO[C@@](Cn6cnnc6)(c6ccc(F)cc6F)C5)cc4)CC3)cc2)c1=O</chem>

Table 3. Adsorption, distribution, metabolism, and excretion properties of compounds.

Compounds	Solubility level	BBB level	CYP2D6	Hepatotoxicity	Absorption level	PPB level
ZINC000003799072	3	2	1	0	0	1
ZINC000100052685	2	4	0	0	1	1
ZINC000003785268	3	2	1	0	0	1
ZINC000008214703	3	4	0	0	0	1
ZINC000021297660	3	3	0	0	0	1
ZINC000000601250	1	0	1	1	1	1
ZINC000028232755	1	4	0	0	3	0
ZINC000085537014	3	4	1	1	2	0
ZINC000028639340	3	4	0	1	2	1

BBB, blood-brain barrier; CYP2D6, cytochrome P-450 2D6; PPB, plasma protein binding.

Aqueous-solubility level: 0, extremely low; 1, very low, but possible; 2, low; 3, good.

BBB level: 0, very high penetrant; 1, high; 2, medium; 3, low; 4, undefined.

CYP2D6 level: 0, noninhibitor; 1, inhibitor.

Hepatotoxicity: 0, nontoxic; 1, toxic.

Human-intestinal absorption level: 0, good; 1, moderate; 2, poor; 3, very poor.

PPB: 0, absorbent weak; 1, absorbent strong.

Table 4. Toxicities of compounds.

Compounds	Mouse NTP (Female)	Mouse NTP (Male)	Rat NTP (Female)	Rat NTP (Male)	Ames	DTP
ZINC000003799072	0	0	0	0	0	1
ZINC000100052685	0	0	0	0	0	0
ZINC000003785268	0	0	0	0	0	1
ZINC000008214703	0	0	0	0	0	0
ZINC000021297660	0	1	0	1	0	1
ZINC000000601250	0	1	0	0	0	1
ZINC000028232755	0	0	0	0	0	1
ZINC000085537014	0	0	0	0	0	0
ZINC000028639340	0	0	0	0	0	1

NTP, U.S. National Toxicology Program; DTP, developmental toxicity potential.

NTP:0 (noncarcinogen); 1 (carcinogen).

Ames: 0 (nonmutagen); 1 (mutagen).

DTP: 0 (nontoxic); 1 (toxic)

Figures

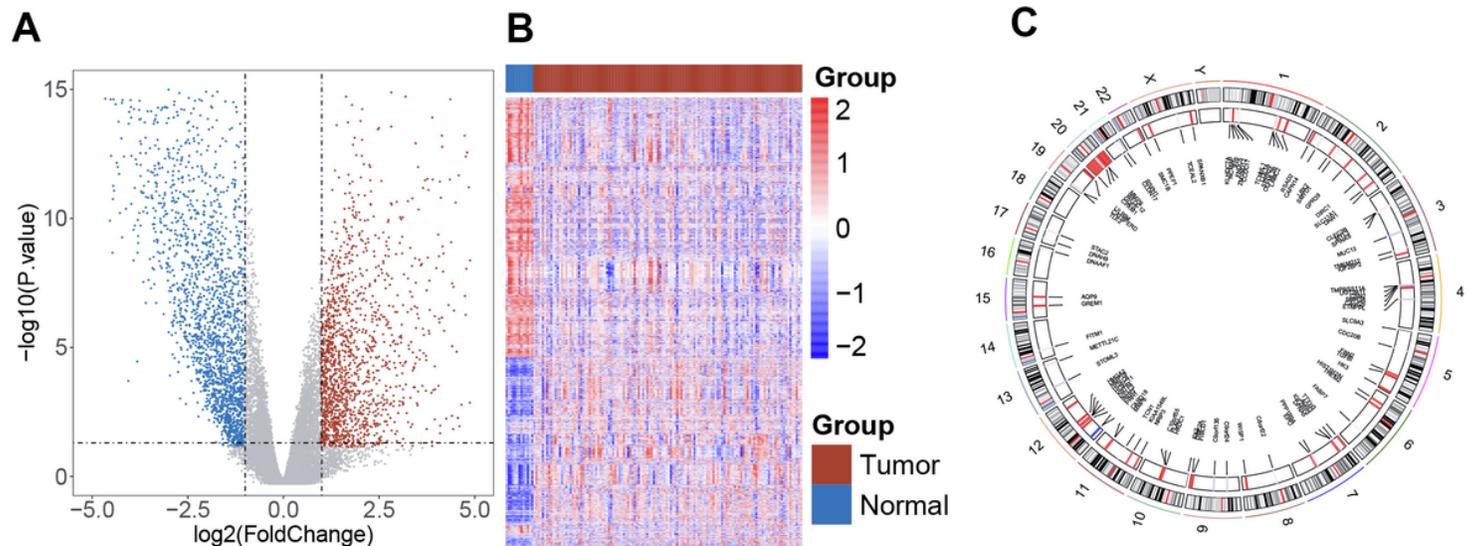


Figure 1

DEG analysis of TSCC in TCGA database. (a) Volcano Plot visualizing the DEGs. The vertical lines demarcate the $\log_2\text{Foldchange}$ values while the horizontal line marks a $-\log_{10}\text{P-value}$ of 0.05. Red represents the upregulated genes, while green represents the downregulated genes. (b) Heatmap of the DEGs. The right longitudinal axis showed the clustering information of samples. The samples were mainly divided into two major clusters and these two clusters were the tumor tissue and adjacent normal tissue. (c) Circos plot showing the distribution in chromosome and expression change of DEGs (the DEG with the highest P-value). Different color of the short line represents the upregulation (red) or downregulation (blue) of DEGs. Abbreviation: DEG, different expressed genes; TSCC, tongue squamous cell carcinoma; TCGA, The Cancer Genome Atlas.

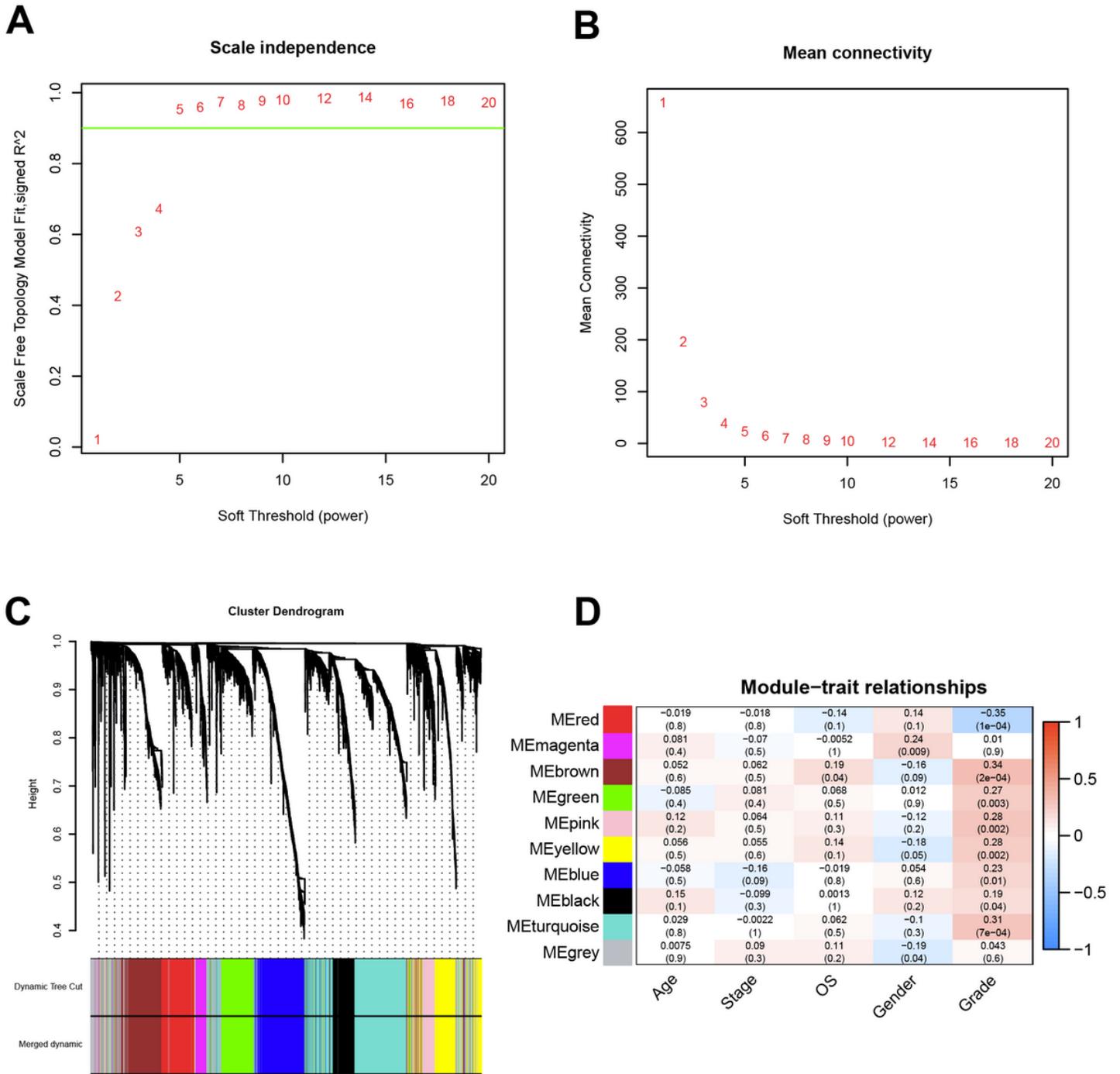


Figure 2

(a) Analysis of the scale-free fit index for various soft-thresholding powers (β). (b) Analysis of the mean connectivity for various soft-thresholding powers. In all, 5 was the most fit power value. (c) The cluster dendrogram of TSCC patients. Each branch in the figure represents one gene, and every color below represents one co-expression module. (d) PCC matrix between gene module and clinical characteristics. The PCC values range from -1 to 1 depending on the strength of the relationship. A positive value indicates that the genes within a particular co-expression module increase as the clinical trait increases, whereas the opposite is true if the PCC is negative. Abbreviation: WGCNA, weighted gene co-expression network analysis; TSCC, tongue squamous cell carcinoma; PCC, Pearson Correlation Coefficient.

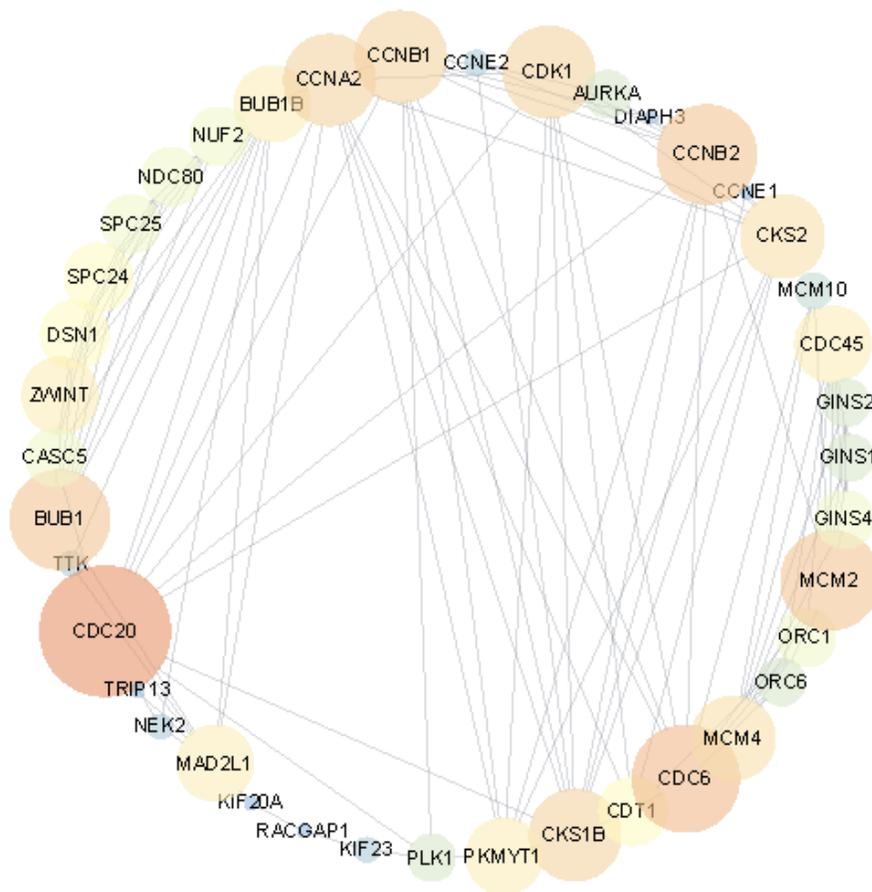


Figure 3

Protein–protein interaction network of genes in the brown module. The color intensity and the size of nodes were positively correlated with the degree score.

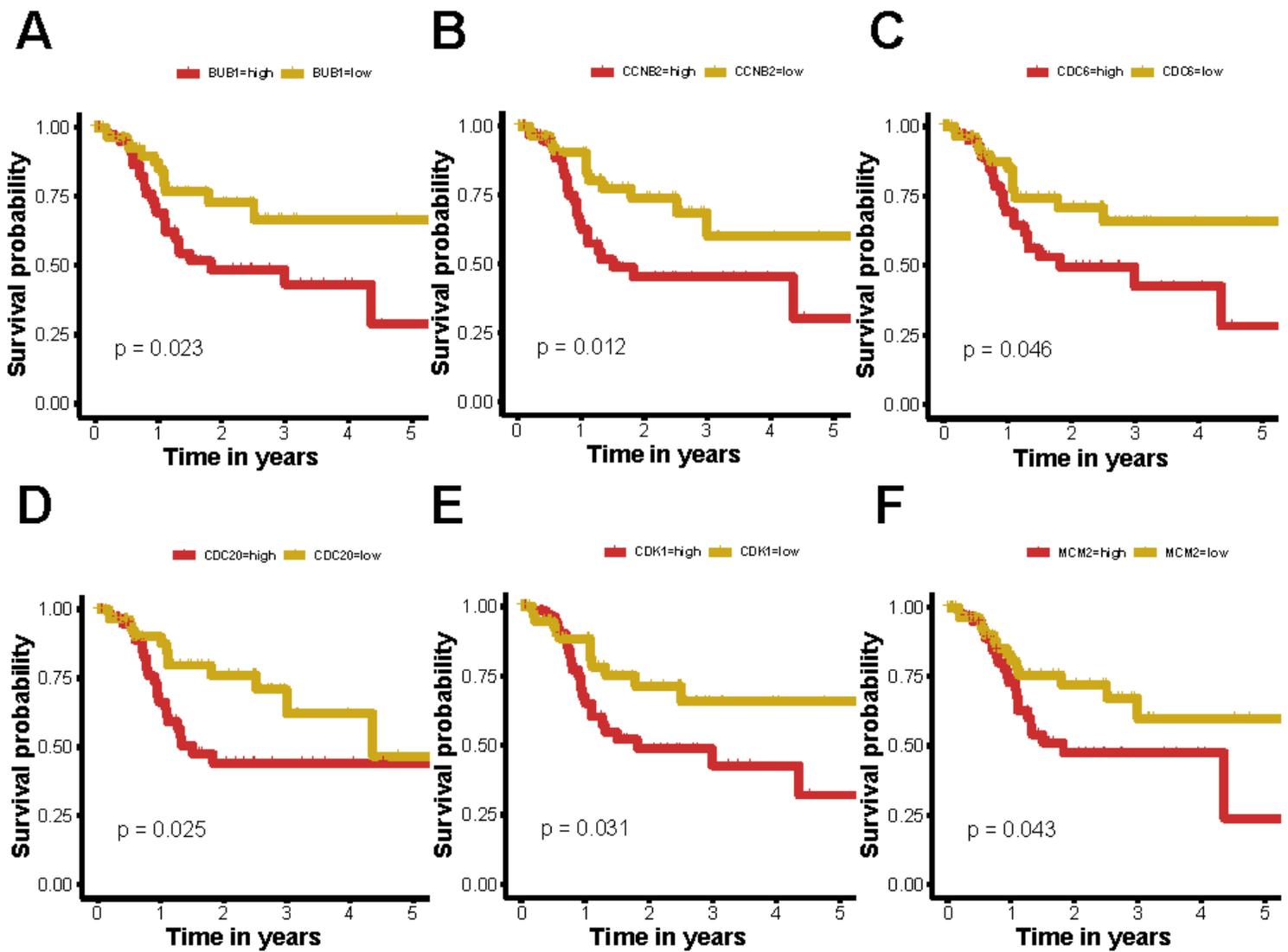


Figure 4

Survival analysis of the association between the expression of hub genes and overall survival time in TSCC. a-f: The Kaplan-Meier survival curve revealed that high BUB1, CCNB2, CDC6, CDC20, CDK1, and MCM2 expression conferred the worse overall survival in patients with TSCC ($P < 0.05$). Abbreviation: TSCC, tongue squamous cell carcinoma.

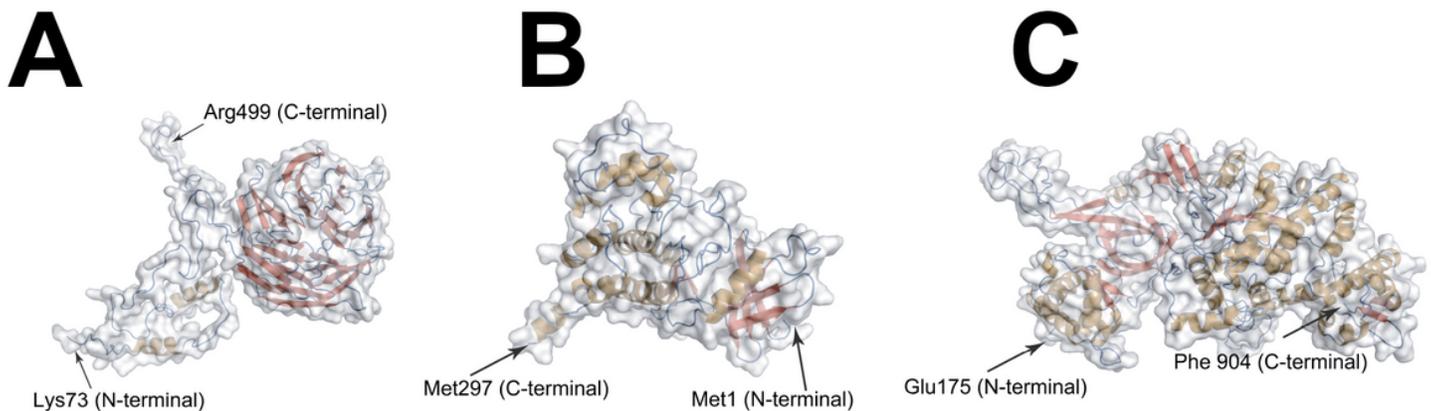


Figure 5

3D structures of hub proteins. (a) CDC20; (b) CDK1; (c) MCM2. In the structure of each protein, secondary structures were represented by colors (β -sheets: red; coil: blue; α helix: brown).

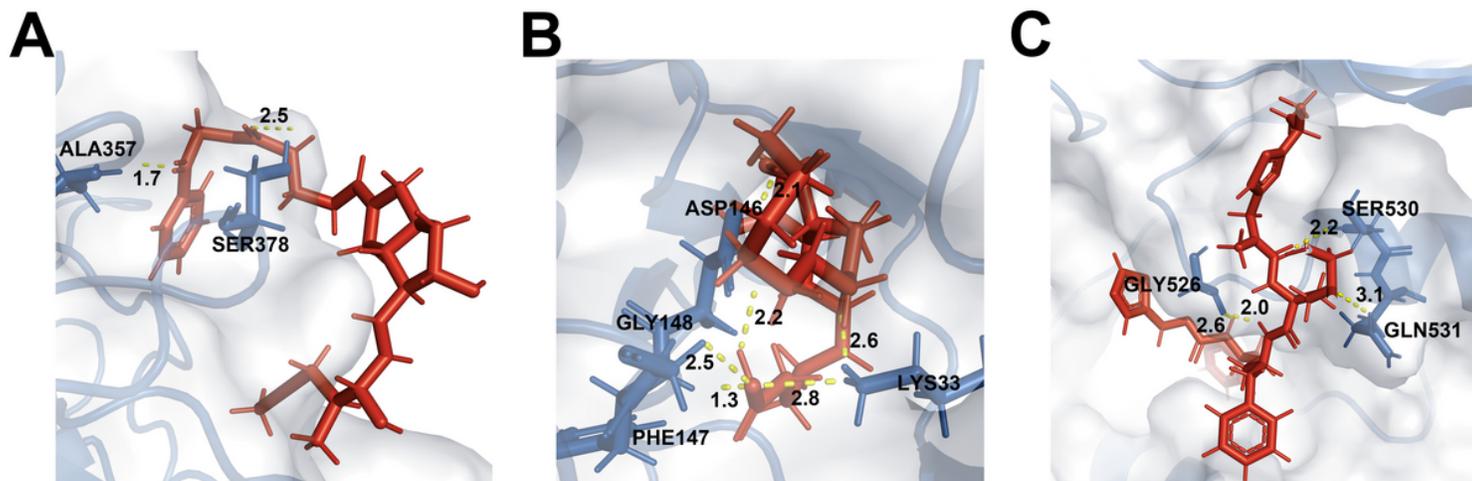


Figure 6

The docking structure of hub proteins and the compounds. (a) ZINC000100052685 with CDC20, (b) ZINC000008214703 with CDK1, (c) ZINC000085537014 with MCM2. The proteins, compounds, and hydrogen bonds were colored by blue, red, and yellow color, respectively.

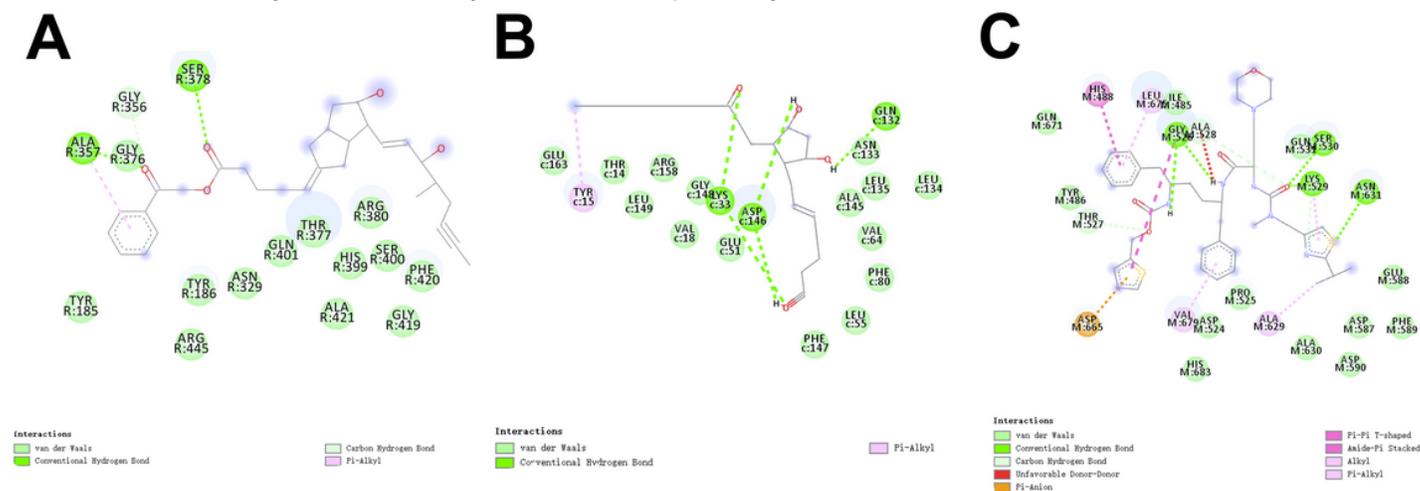


Figure 7

Schematic of intermolecular interaction of the predicted binding modes of (a) ZINC000100052685 with CDC20, (b) ZINC000008214703 with CDK1, (c) ZINC000085537014 with MCM2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.pdf](#)
- [FigS2.pdf](#)

- FigS3.pdf