



The Identification of Stemness-Related Genes in the Risk of Head and Neck Squamous Cell Carcinoma

Guanying Feng^{1,2†}, Feifei Xue^{1†}, Yingzheng He¹, Tianxiao Wang¹ and Hua Yuan^{1,2*}

¹ Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, China, ² Department of Oral and Maxillofacial Surgery, Affiliated Hospital of Stomatology, Nanjing Medical University, Nanjing, China

Objectives: This study aimed to identify genes regulating cancer stemness of head and neck squamous cell carcinoma (HNSCC) and evaluate the ability of these genes to predict clinical outcomes.

Materials and Methods: The stemness index (mRNAsi) was obtained using a one-class logistic regression machine learning algorithm based on sequencing data of HNSCC patients. Stemness-related genes were identified by weighted gene co-expression network analysis and least absolute shrinkage and selection operator analysis (LASSO). The coefficient of LASSO was applied to construct a diagnostic risk score model. The Cancer Genome Atlas database, the Gene Expression Omnibus database, Oncomine database and the Human Protein Atlas database were used to validate the expression of key genes. Interaction network analysis was performed using String database and DisNor database. The Connectivity Map database was used to screen potential compounds. The expressions of stemness-related genes were validated using quantitative real-time polymerase chain reaction (qRT-PCR).

Results: TTK, KIF14, KIF18A and DLGAP5 were identified. Stemness-related genes were upregulated in HNSCC samples. The risk score model had a significant predictive ability. CDK inhibitor was the top hit of potential compounds.

Conclusion: Stemness-related gene expression profiles may be a potential biomarker for HNSCC.

Keywords: head and neck squamous cell carcinoma, cancer stemness, risk, machine learning, compounds

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) was one of the most common cancers worldwide, with ~870,000 new cases and ~440,000 deaths in 2020 (1). It involves cancers of the oral cavity (40%), pharynx (25%), and larynx (15%). The major risk factors for HNSCC include betel nut chewing, alcohol consumption, tobacco consumption, and HPV infection (2).

Cancer stemness (CS) postulates that the growth of a tumor is fueled by limited numbers of dedicated stem cells capable of self-renewal. It represents the degree of oncogenic dedifferentiation (3, 4). Many studies have revealed the molecular mechanisms of CS. Redox and JAK/STAT3

OPEN ACCESS

Edited by:

Parvin Mehdipour,
Tehran University of Medical Sciences,
Iran

Reviewed by:

Fei Zhu,
Soochow University, China
Fengbiao Mao,
University of Michigan, United States

*Correspondence:

Hua Yuan
yuanhua@njmu.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

Received: 31 March 2021

Accepted: 19 May 2021

Published: 11 June 2021

Citation:

Feng G, Xue F, He Y, Wang T and
Yuan H (2021) The Identification
of Stemness-Related Genes in
the Risk of Head and Neck
Squamous Cell Carcinoma.
Front. Oncol. 11:688545.
doi: 10.3389/fonc.2021.688545

signaling pathways were found to influence breast cancer stem cell state (5, 6). The Osteopontin-CD44 signaling pathway was found to enhance cancer stem cell phenotypes in glioma (7). Prostaglandin E2 was found to promote colorectal cancer stem cell expansion (8). Epithelial–mesenchymal transition (EMT) induced mitochondrial fusion through regulation of the miR200c-PGC1 α -MFN1 pathway, directing the asymmetric division of mammary stem cells (9). Previous studies have shown that Oct4, Sox2 and Nanog were associated with oral cancer stem-like status (10, 11). Increasing evidence points to the significance of stemness-related biomarkers in solid tumors, but their role in the risk of HNSCC has not been evaluated. It is essential to identify stemness-related biomarkers and underlying mechanisms associated with CS of HNSCC.

A one-class logistic regression (OCLR) machine learning algorithm published recently was a useful method to quantify cancer stemness index. In this algorithm, a predictive model was constructed and trained in stem cell samples, which can be applied to tumor samples (4). Previous studies have shown the application of the OCLR algorithm in some cancers (12–14). Weighted gene co-expression network analysis (WGCNA) identifies gene modules related to clinical traits through clustering thousands of genes, then uses intramodular connectivity and gene significance based on the correlation of a gene expression profile with a sample trait to identify stemness-related genes for further research. Hence WGCNA identifies significant genes in the biological process more efficiently (15). The least absolute shrinkage and selection operator (LASSO) is a regression analysis method that performs both variable selection and regularization. It enhances the prediction accuracy and interpretability of the resulting statistical model (16, 17).

In this study, a systematic analysis of the relationship between the stemness-related genes and the risk of HNSCC was provided. A flow chart outlining the whole study is shown in **Figure 1A**.

MATERIALS AND METHODS

Study Subjects

Some 44 newly diagnosed head and neck squamous cell carcinoma (HNSCC) patients without any treatment before the operation were enrolled from the Stomatology Hospital of Nanjing Medical University (Nanjing head and neck cancer cohort, NJHNCC). Surgically resected tumors and adjacent normal oral epithelial tissues were obtained from these patients. Clinical information was obtained through medical records while demographic information including age, gender, smoking, alcohol and Betel nut chewing history was collected from each patient using a standard questionnaire. Frozen tumors and matched normal specimens were confirmed by two independent pathologists. Samples were frozen in liquid nitrogen. HNSCC tissues with the malignant cell purity of over 70% were selected for RNA extraction and sequencing. The study was approved by the institutional review board of Nanjing Medical University. Clinical and demographic information is provided in **Table 1**.

Total RNA Extraction and RNA Sequencing

Total RNA was extracted from 44 matched tumor-normal samples using the RNeasy Mini Kit (Qiagen, Hilden, Germany). The quality and quantity of extracted RNA were assessed using the NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA), Qubit 2.0 Fluorometer (Life Technologies, CA, USA) and 1% agarose gel electrophoresis. RNA integrity was assessed using the RNA Nano 6000 Assay Kit (Agilent Technologies, CA, USA).

A total of 3 μ g of high-quality RNA per sample was used for ribosomal RNA removal by the Epicentre Ribo-zero rRNA Removal Kit (Epicentre, USA) and the sequencing library was prepared using the rRNA-depleted RNA by the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, USA)

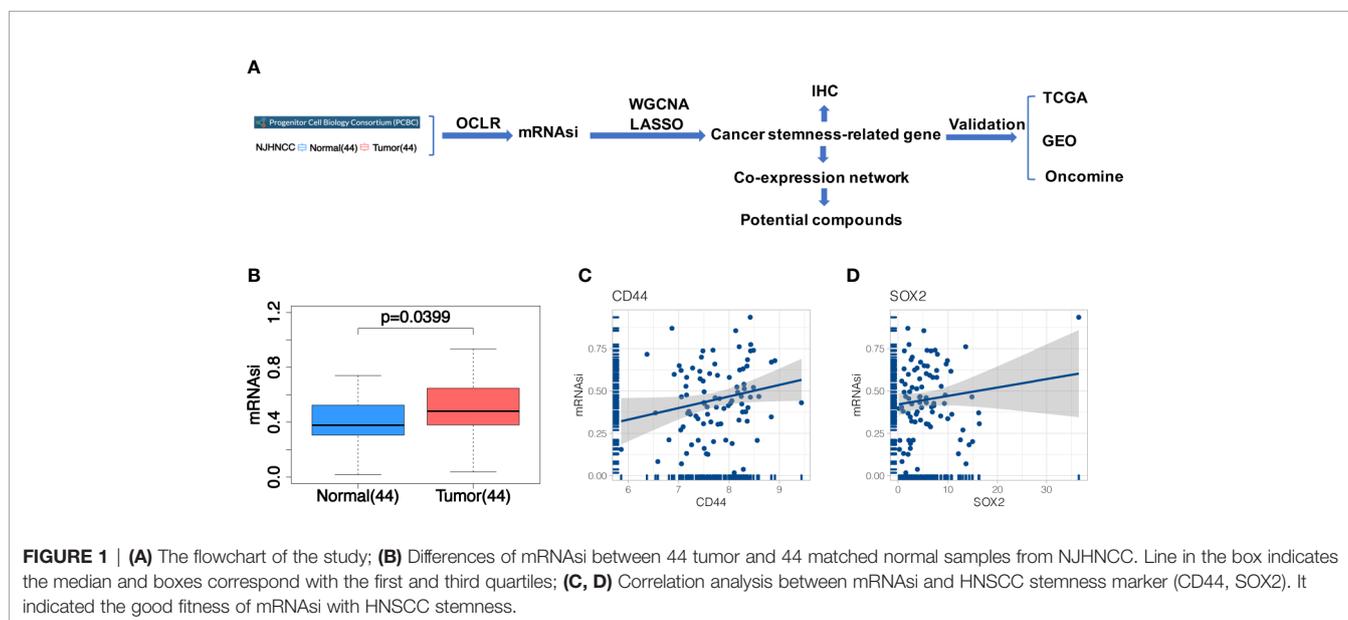


TABLE 1 | Demographic characteristics of 44 HNSCC patients from NJHNCC.

Characteristics	NJHNCC (n = 44)
Age (median; range)	48.5 (27–69)
Gender (frequency; percentage)	
Male	44 (100)
Female	0 (0)
Smoking history (frequency; percentage)	
Ever	40 (91)
Never	4 (9)
Alcohol (frequency; percentage)	
Ever	30 (68)
Never	14 (32)
Betel nut chewing history (frequency; percentage)	
Ever	39 (89)
Never	5 (11)
Clinical stage (frequency; percentage)	
I	0 (0)
II	6 (13.6)
III	21 (47.7)
IV	7 (15.9)
V	8 (18.2)
N/A	2 (4.5)

following manufacturer's recommendations. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, CA, USA), followed by the 150-bp paired-end sequencing on the HiSeq X Ten instrument (Illumina, San Diego, CA, USA) according to the manufacturer's protocols. The result was reported in FPKM.

Data Collection

RNA-seq data from the Progenitor Cell Biology Consortium (PCBC) database (<https://www.synapse.org/#!Synapse:syn2701943>) was obtained (18, 19). RNA-seq data of 500 HNSCC patients was obtained from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov>). The gene expression profiles of 167 HNSCC patients (GSE30784) were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) (20).

Deriving Stemness Index (mRNAsi) Using Machine Learning

Briefly, the original algorithm was obtained from the previously published study (4). We derived stemness signatures using the one-class logistic regression (OCLR) machine learning algorithm trained on gene expression profiles of stem cell samples from the PCBC database. We applied the stemness signatures to RNA-seq data of NJHNCC and obtained stemness index (mRNAsi). The higher mRNAsi, the higher degree of stemness.

The workflow to generate stemness index was as follows: a) download gene expression profiles of stem cell samples from PCBC database (syn2701943); b) map Ensembl IDs to HUGO; c) identify stem cell sample; d) train a one-class logistic regression model; e) perform leave-one-out cross-validation of the one-class logistic regression model; f) export stemness signatures; g) load and preprocess NJHNCC-RNAseq data; h) calculate raw stemness index using Spearman's correlation; i) standardization of raw

stemness index; and j) export stemness index in the interval (0,1). The process was done in R 3.6.1.

The flowchart of deriving stemness index (mRNAsi) was shown in **Supplemental Figure 1**. The R code was shown in **Supplemental File R code**. The stemness signatures of the predictive model were shown in **Supplemental Table 1**.

Wilcoxon test was used to determine the significance of mRNAsi differences between tumor and normal samples. The representative stemness markers of HNSCC were obtained from a published study (21). The correlations between mRNAsi and HNSCC stemness markers were calculated based on the Linear regression (21).

Differential Gene Expression Analysis

The R limma package was used to screen differentially expressed genes between normal and tumor samples. The Wilcoxon signed-rank test was used to identify the differentially expressed genes. The RNA-seq data was transformed by $\log_2(x + 1)$ for normalization. The screening criteria was $|\log_2 \text{ fold change (FC)}| > 1$ and the false discovery rate (FDR) < 0.05 .

Weighted Gene Co-Expression Network Analysis

WGCNA was conducted with differentially expressed genes. Differentially expressed gene profiles were screened to delete outliers. A correlation matrix was constructed with the absolute values in transcriptome data. An adjacency matrix was constructed by the soft threshold. A topological overlap matrix was then constructed. The average linkage hierarchical clustering was conducted on TOM-based dissimilarity. Clusters were detected by the dynamic hybrid cut method.

Module eigengene was regarded as the dominant component in the principal component analysis of gene expression for the module by eigengene network clustering. In general, the correlations between the modules and mRNAsi were evaluated by Pearson's correlation. The correlation between genes and mRNAsi was evaluated by gene significance (GS), which was defined as the Pearson's correlation between gene expression and mRNAsi. The correlation between genes and module was evaluated by module membership (22), which was defined as the Pearson's correlation between gene expression and module eigengenes. The candidate genes were identified with the threshold: correlation coefficients of GS > 0.5 , correlation coefficients of MM > 0.8 .

Least Absolute Shrinkage and Selection Operator Analysis

We constructed a diagnostic risk score model for HNSCC by candidate genes using the LASSO analysis. The diagnostic risk model was constructed with the following formula: risk score = expression level of Gene₁ × β_1 + expression level of Gene₂ × β_2 + ... + expression level of Gene_n × β_n . β is the coefficient calculated by LASSO analysis. The receiver operating characteristic (ROC) curves were used to assess the diagnostic value.

The characteristic of LASSO regression is that when the generalized linear model is established, it includes one-dimensional continuous dependent variable, multi-dimensional continuous dependent variable, non-negative number dependent

variable, binary discrete dependent variable and multiple discrete dependent variable. In addition, whether the dependent variable is continuous or discrete, LASSO regression can solve it. In general, LASSO regression has a wide range of applications. In addition, LASSO regression can also perform variables screen and reduce the complexity of the model, which refers to controlling the complexity of the model through a series of parameters to avoid overfitting. The complexity of LASSO regression is controlled by λ . The larger the λ , the greater the penalty for the linear model with variables. Thus, a model with fewer variables is finally obtained.

Analysis of Gene Expression

The expressions of stemness-related genes between normal and tumor samples in RNA-seq data of NJHNCC were evaluated. The validation of stemness-related genes in the TCGA and GEO database was performed. The validation of stemness-related genes in the Oncomine database (<http://www.oncomine.org>) was performed with threshold: $P = 0.05$, fold change = 2 and gene rank = 10%. The expression of the proteins encoded by the stemness-related genes was analyzed using clinical specimens from the Human Protein Atlas database (<https://www.proteinatlas.org>) (22).

Network Analysis of Co-Expressed Genes

Co-expressed genes were obtained from the WGCNA gene module with the threshold: GS >0.5 and MM >0.8. The relationships of co-expressed genes were evaluated in RNA-seq data of NJHNCC using Pearson's correlation. The protein-protein interactions of co-expressed genes were evaluated in the STRING database (<https://string-db.org/>). The causal interactions of co-expressed genes were evaluated in the DisNor database (<https://disnor.uniroma2.it/>). The neighbor genes indicated possible upstream or downstream genes.

Functional Annotation of Co-Expressed Genes

Co-expressed genes obtained from the WGCNA gene module were used to perform biological function analysis. Gene Ontology (GO) (18) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses were conducted. Biological process (BP), cellular components (CC) and molecular function (MF) were included in GO enrichment analysis. The threshold was $P < 0.05$.

Screening of Potential Compounds

The connectivity map (CMap) database (<https://clue.io/cmap>) was a comprehensive resource to explore the underlying associations among genes, chemicals, and biological conditions. We utilized the CMap to screen potential compounds that targeting HNSCC stemness genes. We evaluated co-expressed genes and obtained relative compounds.

Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) Assay

The primary HNSCC tissues and adjacent normal oral epithelial tissues were obtained from 35 HNSCC patients who received

radical surgery in the Stomatology Hospital of Nanjing Medical University from January 2019 to June 2020. Frozen tumors and matched normal specimens were confirmed by two independent pathologists. Patients who received chemotherapy, radiotherapy, or any other medical intervention were excluded. Tissue specimens were stored in the -80°C freezer. The study was approved by the institutional review board of Nanjing Medical University.

We extracted the total RNA using TRIzol reagent (Invitrogen, USA) and reagent Kit (TAKARA, Japan) according to the manufacturer's protocol. The cDNA was reversed by PrimeScriptTM RT reagent Kit (TAKARA, Japan). The qRT-PCR was performed on a 7300HT system (ABI, USA) using SYBR Premix Ex Taq II kit (TAKARA, Japan). GAPDH was the reference gene. The relative expression level was calculated by the $2^{-\Delta\Delta\text{Ct}}$ method. The sense and antisense primers used are listed below:

GAPDH	5'-GCAAGTTCAACGGCACAG-3', 5'-GCCAGTAGACTCCACGACAT-3'
TTK	5'-GTGGAGCAGTACCACTAGAAATG-3', 5'-CCCAAGTGAACCGGAAAATGA-3'
KIF14	5'-CCTCACCCACAGTAGCCGA-3', 5'-AAGTGCCAACTACTACAGGA-3'
KIF18A	5'-TGCTGGGAAGACCCACACTAT-3', 5'-GCTGGTGTAAGTAAGTCCATGA-3'
DLGAP5	5'-AAGTGGGTCGTTATAGACCTGA-3', 5'-TGCTCGAACATCACTCTCGTTAT-3'

RESULTS

Obtaining and Evaluation of Stemness Index (mRNAsi)

After constructing the predictive model in PCBC stem cell samples and applying it in NJHNCC, we obtained the stemness index (**Supplemental Table 2**). The difference of mRNAsi between 44 tumor and 44 matched normal samples from NJHNCC was evaluated (**Figure 1B**). The mRNAsi of 44 matched normal samples were significantly lower than 44 tumor samples (Paired t-test, $P = 0.0399$). Representative stemness markers (CD44, SOX2) were selected. The good fitness between mRNAsi and stemness markers was observed (**Figures 1C, D**).

Identification of Candidate Stemness-Related Genes by WGCNA

Some 2,925 differential expressed genes were obtained by differential gene expression analysis of RNA-seq data from NJHNCC (Wilcoxon test, FDR <0.05), including 1,294 upregulated genes and 1,631 downregulated genes (**Figure 2A** and **Supplemental Table 3**).

To identify biologically significant gene modules related to stemness, WGCNA was performed. Unsupervised hierarchical clustering showed that gene modules were obtained (**Figure 2B**)

with soft threshold $\beta = 10$ and scale-free $R^2 = 0.9$ (**Supplemental Figure 2**). Focusing on cancer stemness, the correlations between modules and mRNAsi were evaluated (**Figure 2C**). The black module turned out to be the most significant module positively correlated with mRNAsi ($\text{cor} = 0.43$, $P = 0.01$). The tan module exhibited a negative correlation with mRNAsi ($\text{cor} = -0.69$, $P < 0.01$). Thus, the link between gene module and cancer stemness was established.

With threshold that $\text{GS} > 0.5$ and $\text{MM} > 0.8$, 8 candidate stemness-related genes in the black module were then identified (**Figure 2D**): cyclin dependent kinase inhibitor 3 (CDKN3), TTK protein kinase (TTK), kinesin family member 14 (KIF14), kinesin family member 18A (KIF18A), DLG associated protein 5 (DLGAP5), kinesin family member (KIF23), cytoskeleton associated protein 2 like (CKAP2L), BUB1 mitotic checkpoint serine/threonine kinase (BUB1). Details of the eight candidate stemness-related genes were shown in **Supplemental Table 4**.

Identification of Stemness-Related Genes by LASSO and Construction of a Predictive Model for Diagnosis

Some four genes from candidate stemness-related genes were identified using LASSO in RNA-seq data from NJHNSC: TTK, KIF14, KIF18A, and DLGAP5. The coefficients of stemness genes and the minimize λ method screening out four genes were shown (**Figures 2E, F**). The diagnostic risk scores for HNSCC were

calculated using the following formula: Risk score = $-1.081 \times \text{TTK}_{\text{exp}} + 1.364 \times \text{KIF14}_{\text{exp}} + 1.651 \times \text{KIF18A}_{\text{exp}} + 0.965 \times \text{DLGAP5}_{\text{exp}}$. The coefficients were obtained from LASSO.

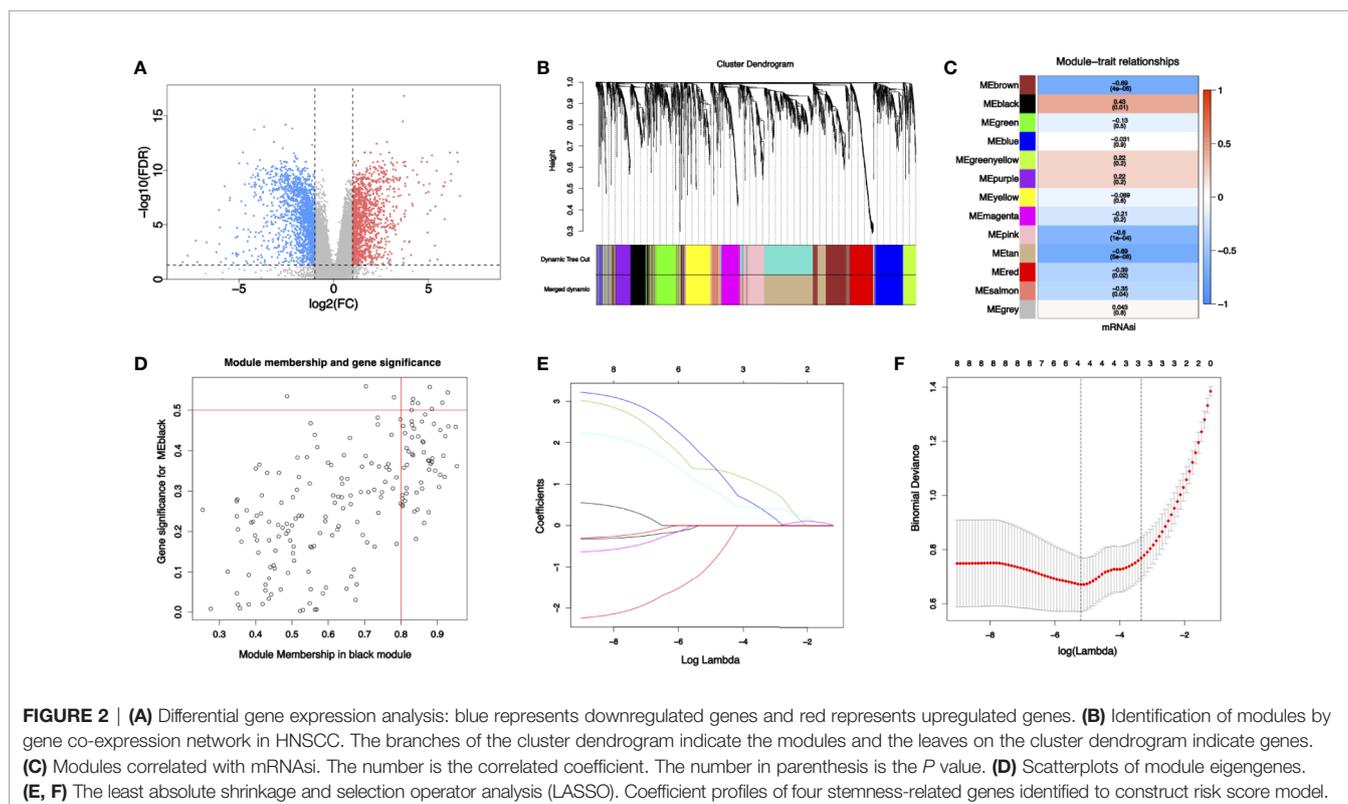
Evaluation of Stemness-Related Genes' Expression and Risk Score Model

The expression of stemness-related genes was evaluated in gene expression profiles from NJHNSC (**Figure 3A**), GEO database (**Figure 3B**), TCGA database (**Figure 3C**). Four stemness-related genes were significantly high-expressed in tumor samples ($P < 0.01$). In the Oncomine database, the stemness-related genes were upregulated not only in HNSCC but also in other cancers (**Figure 3D**). All of them were in the top 1% of gene rank and each of them had evidence from multiple cancers.

Moreover, we evaluated the risk score model in NJHNSC (AUC = 0.954), TCGA database (AUC = 0.949) and GEO database (AUC = 0.899) (**Figure 3E**).

Detected Expression of Stemness-Related Genes in HNSCC Tissues

We analyzed the expression of the proteins encoded by the stemness-related genes using clinical specimens from the HPA database. TTK (**Figure 4A**) was moderately positive and KIF14 (**Figure 4B**) was strongly positive in HNSCC metastatic lymph nodes. KIF18A (**Figure 4C**) was moderately positive and DLGAP5 (**Figure 4D**) was strongly positive in HNSCC tissue relative to their expression levels in normal tissue.



Network Analysis of Co-Expressed Genes

We explored the correlation between eight co-expressed genes by Pearson’s correlation in RNA-seq data of NJHNSCC. The result showed that the correlations between the expression of co-expressed genes were significant ($P < 0.05$). Each of the stemness genes was highly correlated with other genes (Figure 5A). Also, we explored the relationship of co-expressed genes in the STRING database and it showed that they had significant correlations with other genes (Figure 5B and Supplemental Figure 3).

To investigate the causal interactions of co-expressed genes, we inferred neighbors of them in the DisNor database. The result revealed that PLK1, CENPE and TP53 might be downstream genes. DNA damage might lead to the upregulation of TTK. BUB1 might upregulate mitotic activity (Figure 5E).

Biological Function Analysis of Co-Expressed Genes

To explore the biological function of co-expressed genes, GO and KEGG analyses were performed. The results showed that co-expressed

genes were enriched in biological processes including chromosome segregation, nuclear division and organelle fission in GO analysis and cell cycle pathway in KEGG analysis (Figures 5C, D).

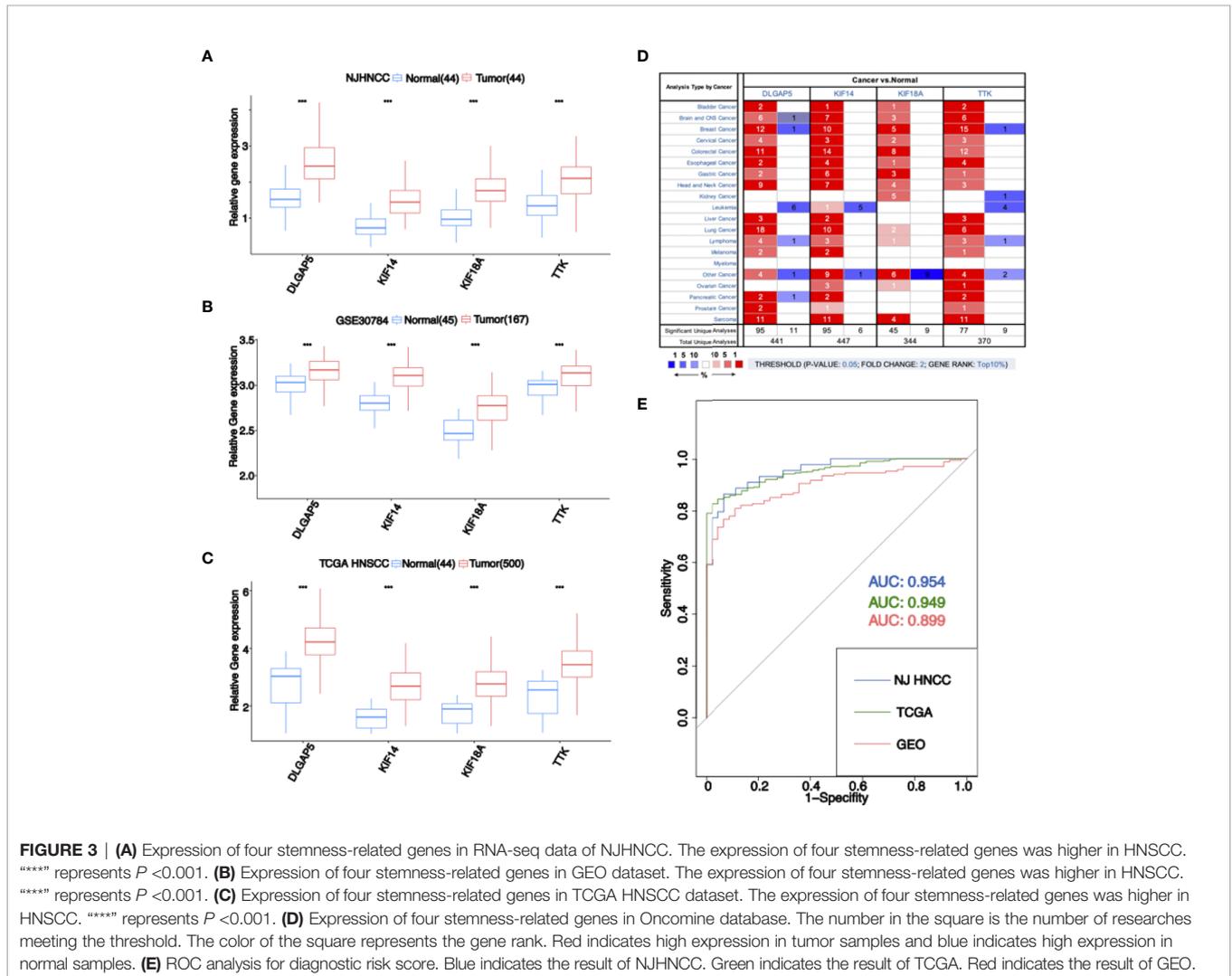
In addition, genes in the module that negatively related to stemness were also analyzed and the result showed that they were related to extracellular matrix organization (Supplemental Figures 4A, B).

Screening of Potential Compounds for Stemness

We screened potential compounds targeting the stemness of HNSCC in the CMap database. The top compounds were exhibited, along with mechanisms of action (Figure 6A). The top mechanism of action was the CDK inhibitor (Figure 6B).

The Expression of Stemness-Related Genes in Tissues

The expressions of stemness-related genes in HNSCC and adjacent normal oral epithelial tissues were evaluated by qRT-



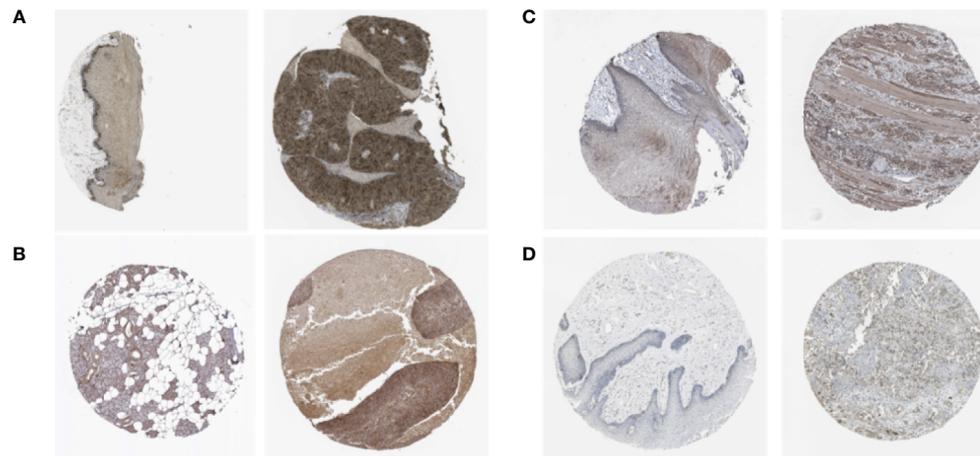


FIGURE 4 | The expression profiles of the proteins encoded by TTK (A), KIF14 (B), KIF18A (C) and DLGAP5 (D) in normal and HNSCC tissues using clinical specimens from the Human Protein Profiles. The left is normal specimens and the right is specimens from HNSCC patients.

PCR (Supplemental Figure 5). GAPDH was the reference gene. The expressions of TTK, KIF14, KIF18A and DLGAP5 were upregulated in HNSCC tissues compared to normal tissues ($P < 0.001$), which were consistent with our findings.

DISCUSSION

A recent study has reported that cancer stem cell plays an important role in cancer growth, progression and therapy resistance (23). It indicates that the role of cancer stemness in HNSCC is worth investigating. In this study, we used the machine learning algorithm to learn the gene signature of stem cells and applied it in RNA-seq data of NJHNCC. Then, we identified stemness-related genes of HNSCC and constructed a diagnostic risk score model.

From normal tissue to precancerous lesions and cancer, one essential signature is the process of oncogenic dedifferentiation. The outcome that the stemness of tumor samples was higher than normal samples corroborated this. Some certain mutations contribute to the oncogenesis of dedifferentiated cells rather than mature cells, indicating that cell dedifferentiation status could be one of the oncogenic factors (23). In stemness index analysis, it was associated with stemness marker CD44 and SOX2. The results showed that the stemness index had the ability to distinguish the degree of cancer stemness efficiently.

Next, we obtained candidate stemness-related genes using WGCNA and stemness index. Then we constructed the diagnostic risk score model for HNSCC using LASSO. The predictive model was externally validated and it indicated a robust predictive ability.

With regard to the stemness-related genes, TTK was reported to regulate colon cancer progression *via* PI3K/AKT pathway (24). Kinesins are a family that plays many roles in intracellular transport or cell division. KIF14 and KIF18A were involved in the progression from prometaphase to metaphase stage of

mitosis and cytokinesis (25). DLGAP5 was involved in the progress of centrosome-independent mitotic spindle assembly, which is essential for the survival and proliferation of SMARCA4/BRG1 mutant cells. This mechanism was reported in non-small cell lung cancer (26). Among the stemness-related genes, TTK was the cancer driver gene with a significant driver level according to the OncoVar database (27). Though the majority of somatic mutations are harmless, somatic alterations in specific cancer driver genes confer a selective growth advantage for the cells and can lead to cancer (28). It indicates the possibility of investigating cancer stemness in terms of somatic alterations and related cancer driver genes.

With regard to the co-expression network of stemness-related genes, PLK1 promotes EMT and metastasis in gastric carcinoma (29). PLK1 has been identified as a potential target to enhance the therapeutic sensitivity of paclitaxel-resistant prostate cancer (30). CENPE is the largest kinesin protein. The role of CENPE in microtubule kinetochore capture contributes to chromosome congression and alignment (25). Cell division cycle protein was reported to impair ionizing radiation-induced cell apoptosis and promote EMT in nasopharyngeal carcinoma. The expression of CDC6 was upregulated steadily after acute ionizing radiation exposure (31). It indicates that ionizing radiation could lead to oncogenic dedifferentiation, but the mechanism remains unclear.

The signaling pathway that stemness-related genes and co-expression network are most involved in is PI3K/AKT pathway. PI3K/AKT pathway has been reported to regulate EMT (32). Also, the oncogenesis of hepatocellular carcinoma could be activated *via* PI3K/AKT pathway (33). With regard to cancer stemness, research on PI3K/AKT pathway is still lack. Signaling pathways including Hedgehog, Notch, and Wnt pathways have been reported to maintain cancer stemness status (34). Among stemness-related genes and their co-expression network, TTK and PLK1 have been reported to involve in these pathways. In B-cell acute lymphoblastic leukemia, activation of Notch signaling leads to cell-cycle arrest and

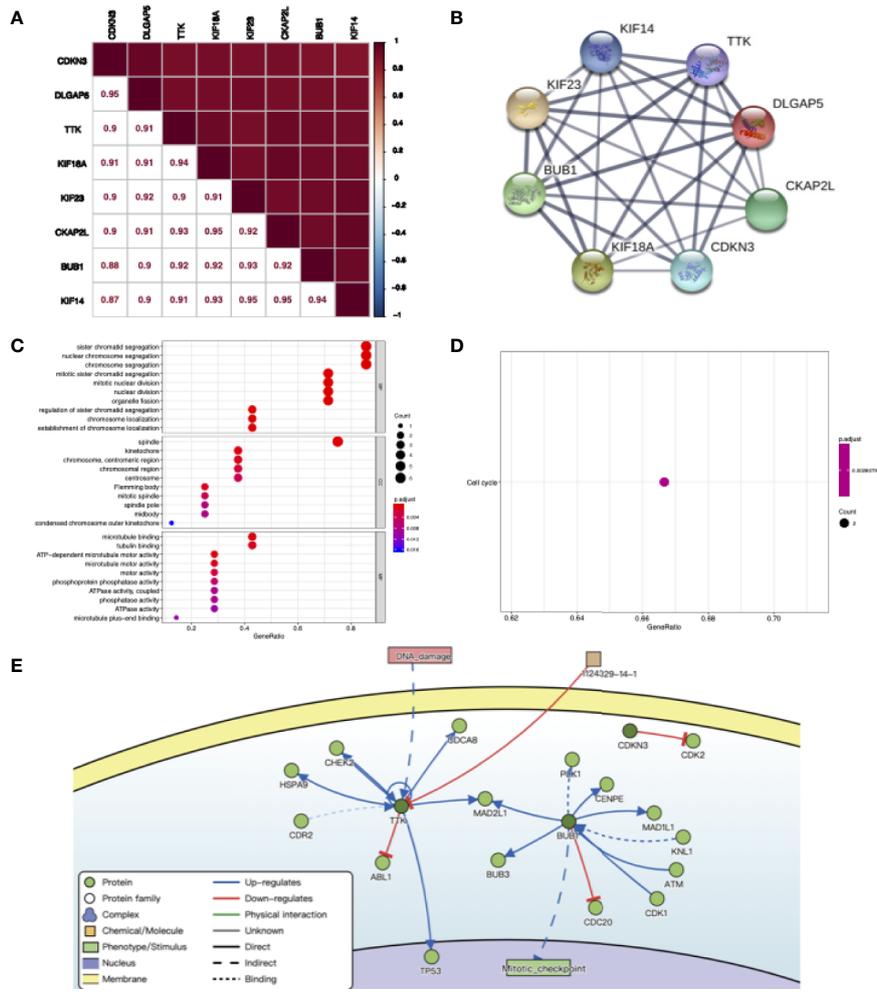


FIGURE 5 | (A) The correlation between co-expression genes of stemness-related genes calculating by RNA-seq data of NJHNSC. Pearson correlation coefficient is shown. The depth of square color was determined by the correlation coefficient. (B) The relationship of co-expression genes in the STRING database. The thickness of the solid line represents the strength of the correlation. (C) Results for GO enrichment analysis. The color of the dot was determined by *P* value. (D) Results for KEGG enrichment analysis. The color of the dot was determined by *P* value. (E) The causal interactions of co-expression genes in DisNor database. The dark green circles represent co-expression genes. The light green circles represent the neighbor of co-expression genes.

suppresses PLK1 (35). PLK1 inhibition or depletion enhances the level of cytosolic and nuclear β -catenin in human prostate cancer cells (36). Cell lines harboring activating mutations in the CTNNB1 gene, encoding the Wnt pathway signaling regulator β -catenin, were on average up to five times more sensitive to TTK inhibitors than cell lines wild-type for CTNNB1 (37). EMT has also been involved in the generation of cancer stem cells (38). High expression of the EMT-TFs ZEB1 in cancer cells activated the expression of stemness factors SOX2 (39). The ZEB1/miR-200/BMI1 pathway was involved in pancreatic cancer stem cells, which explained how EMT enabled stemness (40). Mesenchymal and stemness traits are known to characterize cancer stemness cell, which is responsible for tumor metastasis and resistance to conventional therapy (41, 42). In stemness-related genes and co-expressed genes we identified, PLK1 and CDC6 have been reported to involve in

the progress of EMT (29, 31). The role of other stemness-related genes in EMT is worth investigating.

The potential compounds targeting the HNSCC stemness were screened. Top potential drugs and mechanisms of action were found. The CDK inhibitor was the top hit, following by the PLK inhibitor. To date, High-dose cisplatin, given concurrently with radiotherapy as part of a definitive chemoradiotherapy regimen, is the standard of care, with established survival benefits for patients (43). The compounds targeting CDK and PLK inhibitors may bring novel insights into HNSCC chemotherapy.

In summary, the comprehensive transcriptome profiling of HNSCC using the OCLR algorithm was conducted and stemness-related genes were identified. We also constructed a risk score model for HNSCC. The co-expression network of stemness-related genes was analyzed. CDK and PLK inhibitors

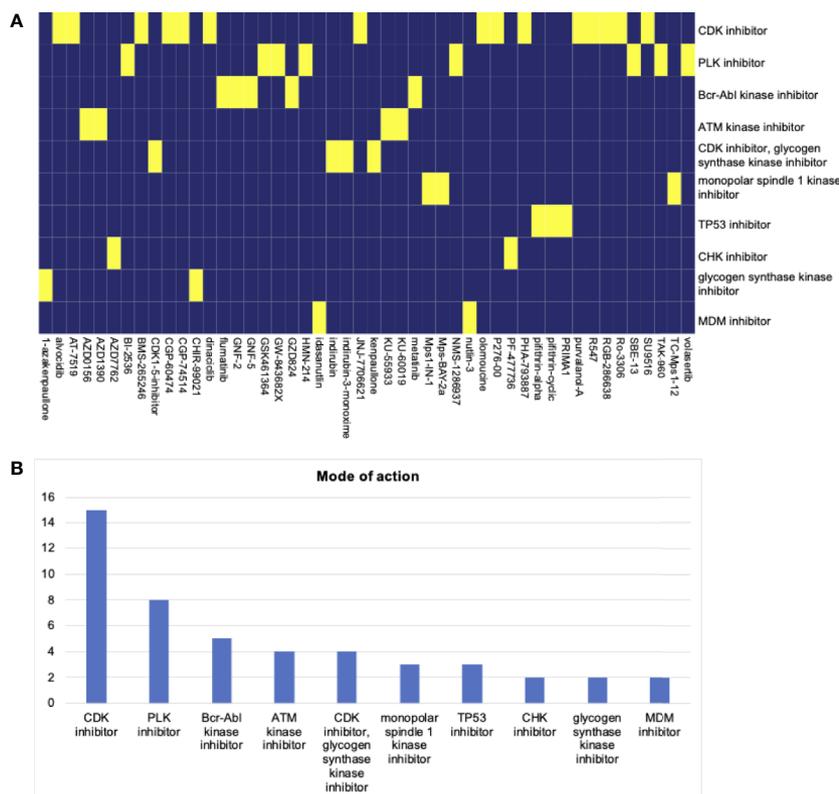


FIGURE 6 | (A) The heatmap of compounds from the CMap that share mechanism of action based on co-expression genes. **(B)** The number of genes enriched in each mechanism of action. The top hit is the CDK inhibitor.

play an important role in HNSCC chemotherapy. The application of machine learning in HNSCC transcriptome research brings new perspectives into tumor research. Considering gaps of cell based exploration remained in this study, more experiments need to be conducted to evaluate the effect of stemness-related genes, which will strengthen the elucidation of the diversity and heterogeneity in cancer research. We foresee conducting more researches to further investigate cancer stemness.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Figshare Database and accession number NJHNCC-RNaseq-stemsig [14540610] (https://figshare.com/articles/dataset/NJHNCC-RNaseq-stemsig_txt/14540610).

ETHICS STATEMENT

The study was approved by the Institutional Review Board of Nanjing Medical University. The patients/participants

provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (81672678), Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD, 2018-87), the Project of Invigorating Health Care through Science, Technology and Education (Jiangsu Provincial Medical Youth Talent QNRC2016852) and sponsored by Qing Lan Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.688545/full#supplementary-material>

REFERENCES

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
- Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* (2017) 3(4):524–48. doi: 10.1001/jamaoncol.2016.5688
- Clevers H. The Cancer Stem Cell: Premises, Promises and Challenges. *Nat Med* (2011) 17(3):313–9. doi: 10.1038/nm.2304
- Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine Learning Identifies Stemness Features Associated With Oncogenic Dedifferentiation. *Cell* (2018) 173(2):338–54. doi: 10.1016/j.cell.2018.03.034
- Luo M, Shang L, Brooks MD, Jiagge E, Zhu Y, Buschhaus JM, et al. Targeting Breast Cancer Stem Cell State Equilibrium Through Modulation of Redox Signaling. *Cell Metab* (2018) 28(1):69–86. doi: 10.1016/j.cmet.2018.06.006
- Wang T, Fahrman JF, Lee H, Li Y-J, Tripathi SC, Yue C, et al. JAK/STAT3-Regulated Fatty Acid β -Oxidation Is Critical for Breast Cancer Stem Cell Self-Renewal and Chemoresistance. *Cell Metab* (2018) 27(1):136–50. doi: 10.1016/j.cmet.2017.11.001
- Pietras A, Katz AM, Ekström EJ, Wee B, Halliday JJ, Pitter KL, et al. Osteopontin-CD44 Signaling in the Glioma Perivascular Niche Enhances Cancer Stem Cell Phenotypes and Promotes Aggressive Tumor Growth. *Cell Stem Cell* (2014) 14(3):357–69. doi: 10.1016/j.stem.2014.01.005
- Wang D, Fu L, Sun H, Guo L, DuBois RN. Prostaglandin E2 Promotes Colorectal Cancer Stem Cell Expansion and Metastasis in Mice. *Gastroenterology* (2015) 149(7):1884–95. doi: 10.1053/j.gastro.2015.07.064
- Wu M-J, Chen Y-S, Kim MR, Chang C-C, Gampala S, Zhang Y, et al. Epithelial-Mesenchymal Transition Directs Stem Cell Polarity Via Regulation of Mitofusin. *Cell Metab* (2019) 29(4):993–1002.e6. doi: 10.1016/j.cmet.2018.11.004
- Chien C-S, Wang M-L, Chu P-Y, Chang Y-L, Liu W-H, Yu C-C, et al. Lin28B/Let-7 Regulates Expression of Oct4 and Sox2 and Reprograms Oral Squamous Cell Carcinoma Cells to a Stem-Like State. *Cancer Res* (2015) 75(12):2553–65. doi: 10.1158/0008-5472.CAN-14-2215
- Chiou S-H, Yu C-C, Huang C-Y, Lin S-C, Liu C-J, Tsai T-H, et al. Positive Correlations of Oct-4 and Nanog in Oral Cancer Stem-Like Cells and High-Grade Oral Squamous Cell Carcinoma. *Clin Cancer Res* (2008) 14(13):4085–95. doi: 10.1158/1078-0432.CCR-07-4404
- Pan S, Zhan Y, Chen X, Wu B, Liu B. Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices. *Front Oncol* (2019) 9:613. doi: 10.3389/fonc.2019.00613
- Zhang Y, Tseng JT-C, Lien I, Li F, Wu W, Li H. mRNAsi Index: Machine Learning in Mining Lung Adenocarcinoma Stem Cell Biomarkers. *Genes* (2020) 11(3):257. doi: 10.3390/genes11030257
- Lian H, Han YP, Zhang YC, Zhao Y, Yan S, Li QF, et al. Integrative Analysis of Gene Expression and DNA Methylation Through One-Class Logistic Regression Machine Learning Identifies Stemness Features in Medulloblastoma. *Mol Oncol* (2019) 13(10):2227–45. doi: 10.1002/1878-0261.12557
- Langfelder P, Horvath S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* (2008) 9:559. doi: 10.1186/1471-2105-9-559
- Santosa F, Symes WW. Linear Inversion of Band-Limited Reflection Seismograms. *SIAM J Sci Stat Comput* (1986) 7(4):1307–30. doi: 10.1137/0907087
- Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Series B (Methodol)* (1996) 58(1):267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
- Salomonis N, Dexheimer PJ, Omberg L, Schroll R, Bush S, Huo J, et al. Integrated Genomic Analysis of Diverse Induced Pluripotent Stem Cells From the Progenitor Cell Biology Consortium. *Stem Cell Rep* (2016) 7(1):110–25. doi: 10.1016/j.stemcr.2016.05.006
- Daily K, Sui SJH, Schriml LM, Dexheimer PJ, Salomonis N, Schroll R, et al. Molecular, Phenotypic, and Sample-Associated Data to Describe Pluripotent Stem Cell Lines and Derivatives. *Sci Data* (2017) 4(1):1–10. doi: 10.1038/sdata.2017.30
- Chen C, Méndez E, Houck J, Fan W, Lohavanichbutr P, Doody D, et al. Gene Expression Profiling Identifies Genes Predictive of Oral Squamous Cell Carcinoma. *Cancer Epidemiol Biomarkers Prev* (2008) 17(8):2152–62. doi: 10.1158/1055-9965.EPI-07-2893
- Yu SS, Cirillo N. The Molecular Markers of Cancer Stem Cells in Head and Neck Tumors. *J Cell Physiol* (2020) 235(1):65–73. doi: 10.1002/jcp.28963
- Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhorji G, et al. A Pathology Atlas of the Human Cancer Transcriptome. *Science* (2017) 357(6352):eaan2507. doi: 10.1126/science.aan2507
- Lytle NK, Barber AG, Reya T. Stem Cell Fate in Cancer Growth, Progression and Therapy Resistance. *Nat Rev Cancer* (2018) 18(11):669–80. doi: 10.1038/s41568-018-0056-x
- Zhang L, Jiang B, Zhu N, Tao M, Jun Y, Chen X, et al. Mitotic Checkpoint Kinase Mps1/TTK Predicts Prognosis of Colon Cancer Patients and Regulates Tumor Proliferation and Differentiation Via PKC α /ERK1/2 and PI3K/Akt Pathway. *Med Oncol* (2019) 37(1):5. doi: 10.1007/s12032-019-1320-y
- Rath O, Kozielski F. Kinesins and Cancer. *Nat Rev Cancer* (2012) 12(8):527–39. doi: 10.1038/nrc3310
- Tagal V, Wei S, Zhang W, Brekken RA, Posner BA, Peyton M, et al. SMARCA4-Inactivating Mutations Increase Sensitivity to Aurora Kinase A Inhibitor VX-680 in Non-Small Cell Lung Cancers. *Nat Commun* (2017) 8:14098. doi: 10.1038/ncomms14098
- Wang T, Ruan S, Zhao X, Shi X, Teng H, Zhong J, et al. OncoVar: An Integrated Database and Analysis Platform for Oncogenic Driver Variants in Cancers. *Nucleic Acids Res* (2021) 49(D1):D1289–D301. doi: 10.1093/nar/gkaa1033
- Martincorena I, Campbell PJ. Somatic Mutation in Cancer and Normal Cells. *Science* (2015) 349(6255):1483–9. doi: 10.1126/science.aab4082
- Cai XP, Chen LD, Song HB, Zhang CX, Yuan ZW, Xiang ZX. PLK1 Promotes Epithelial-Mesenchymal Transition and Metastasis of Gastric Carcinoma Cells. *Am J Transl Res* (2016) 8(10):4172–83.
- Shin S-B, Woo S-U, Yim H. Cotargeting Plk1 and Androgen Receptor Enhances the Therapeutic Sensitivity of Paclitaxel-Resistant Prostate Cancer. *Ther Adv Med Oncol* (2019) 11:175883591984637. doi: 10.1177/1758835919846375
- Yu X, Liu Y, Yin L, Peng Y, Peng Y, Gao Y, et al. Radiation-Promoted CDC6 Protein Stability Contributes to Radioresistance by Regulating Senescence and Epithelial to Mesenchymal Transition. *Oncogene* (2019) 38(4):549–63. doi: 10.1038/s41388-018-0460-4
- Chen C, Song G, Xiang J, Zhang H, Zhao S, Zhan Y. AURKA Promotes Cancer Metastasis by Regulating Epithelial-Mesenchymal Transition and Cancer Stem Cell Properties in Hepatocellular Carcinoma. *Biochem Biophys Res Commun* (2017) 486(2):514–20. doi: 10.1016/j.bbrc.2017.03.075
- Yang Y-F, Zhang M-F, Tian Q-H, Fu J, Yang X, Zhang CZ, et al. SPAG5 Interacts With CEP55 and Exerts Oncogenic Activities Via PI3K/AKT Pathway in Hepatocellular Carcinoma. *Mol Cancer* (2018) 17(1):117. doi: 10.1186/s12943-018-0872-3
- Saygin C, Matei D, Majeti R, Reizes O, Lathia JD. Targeting Cancer Stemness in the Clinic: From Hype to Hope. *Cell Stem Cell* (2019) 24(1):25–40. doi: 10.1016/j.stem.2018.11.017
- Kannan S, Aitken MJL, Herbrich SM, Golfman LS, Hall MG, Mak DH, et al. Antileukemia Effects of Notch-Mediated Inhibition of Oncogenic PLK1 in B-Cell Acute Lymphoblastic Leukemia. *Mol Cancer Ther* (2019) 18(9):1615–27. doi: 10.1158/1535-7163.MCT-18-0706
- Li J, Karki A, Hodges KB, Ahmad N, Zoubeidi A, Strebhardt K, et al. Cotargeting Polo-Like Kinase 1 and the Wnt/ β -Catenin Signaling Pathway in Castration-Resistant Prostate Cancer. *Mol Cell Biol* (2015) 35(24):4185–98. doi: 10.1128/MCB.00825-15
- Zaman GJR, de Roos JADM, Libouban MAA, Prinsen MBW, de Man J, Buijsman RC, et al. TTK Inhibitors as a Targeted Therapy for (-catenin) Mutant Cancers. *Mol Cancer Ther* (2017) 16(11):2609–17. doi: 10.1158/1535-7163.MCT-17-0342

38. Scheel C, Weinberg RA. Cancer Stem Cells and Epithelial–Mesenchymal Transition: Concepts and Molecular Links. *Semin Cancer Biol* (2012) 22 (5):396–403. doi: 10.1016/j.semcancer.2012.04.001
39. Wellner U, Brabletz T, Keck T. ZEB1 in Pancreatic Cancer. *Cancers* (2010) 2 (3):1617–28. doi: 10.3390/cancers2031617
40. Wilson MM, Weinberg RA, Lees JA, Guen VJ. Emerging Mechanisms by Which EMT Programs Control Stemness. *Trends Cancer* (2020) 6(9):775–80. doi: 10.1016/j.trecan.2020.03.011
41. Visvader JE, Lindeman GJ. Cancer Stem Cells: Current Status and Evolving Complexities. *Cell Stem Cell* (2012) 10(6):717–28. doi: 10.1016/j.stem.2012.05.007
42. Pradella D, Naro C, Sette C, Ghigna C. EMT and Stemness: Flexible Processes Tuned by Alternative Splicing in Development and Cancer Progression. *Mol Cancer* (2017) 16(1):8. doi: 10.1186/s12943-016-0579-2
43. Chow LQM. Head and Neck Cancer. *N Engl J Med* (2020) 382(1):60–72. doi: 10.1056/NEJMra1715715

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Feng, Xue, He, Wang and Yuan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.