Establishment of a Prognostic Stemness Signature for Cervical Squamous Cell Carcinoma

Xiwen Sun
Zhejiang University School of Medicine Second Affiliated Hospital
Fang Wang
Zhejiang University School of Medicine Second Affiliated Hospital
Jiayao Zhao
Zhejiang University School of Medicine Second Affiliated Hospital
Jianwei Zhou (2195045@zju.edu.cn)
The Second Affiliated Hospital of Zhejiang University School of Medicine

Research article

Keywords: cervical carcinoma, TCGA, stemness, prognosis□cervical squamous cell carcinoma

Posted Date: November 19th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-108185/v1

Abstract

Background: The aim of this study was to construct a robust stemness-related gene signature for predicting prognosis of cervical squamous cell carcinoma (CSCC).

Methods: Expression data for the PCBC database-derived pluripotent stem cell (PSC) samples were collected using the one-class logistic regression (OCLR) method to calculate stemness indexes (mRNAsi) of samples derived from the TCGA dataset. Functions of possible mRNAsi-related stemness genes extracted through WGCNA were then examined by enrichment analysis. Most representative stemness genes for prognosis prediction were screened to construct a stemness-related gene signature by shrinkage estimate and univariate analysis. Next, the TCGA dataset and the GSE44001 external dataset were incorporated into that model and classified to evaluate the model efficiency and stability in patient prognosis prediction and classification according to the Riskscore. The associations between the Riskscore and clinical characteristics as well as relevant signaling pathways were also explored. Moreover, the prognosis predicting efficiency of the stemness-related gene signature was compared with those of CSCC prognostic signatures reported in other studies.

Results: According to the findings, mRNAsi showed significant correlation with key oncogene mutation degrees (including DMD, KMT2C, EP300 and MUC4), infiltrating stroma cells and the CIMP classification for CSCC cases. The 8-stemness gene signature in this study achieved high stability and accuracy in prognosis prediction for CSCC cases. In the meanwhile, the model provided diverse therapeutic targets to precisely treat CSCC in clinical practice based on various subtype-specific stemness genes.

Conclusion: Our present study suggested that the 8 stemness gene signature can help to screen out novel stem-related diagnostic indicators, therapeutic targets and prognostic predictors in CSCC.

Background

Cervical squamous cell carcinoma (CSCC) has been identified as a highly aggressive malignant tumor in the reproductive system of women, and it has greatly threatened the health and quality of life of women(1). CSCC has a high morbidity among the developing countries, which takes up 60–90% of the total cases in the world. At present, radical hysterectomy has been recognized to be the preferred treatment among early cervical cancer patients(2–4). Meanwhile, cervical cancer screening is becoming more and more popular, which substantially improves the efficacy and prognosis for patients at the early stage of disease(5, 6).

Postoperative CSCC metastasis and recurrence are still the primary factors causing death in clinic(7, 8). The middle-advanced CSCC patients usually receive chemotherapy and/or radiotherapy, but no satisfactory efficacy is achieved(9). Currently, the International Federation of Gynecology and Obstetrics (FIGO) classification system has been used as the primary criterion to predict the prognosis for cervical cancer patients(10). However, there may be diverse prognostic results for

CSCC patients at close clinical stages. As a result, hierarchical processing for the high-risk subgroups is necessary to optimize postoperative treatment and achieve individualized monitoring, making it urgently needed to discover the new prognostic markers.

Stemness is the ability to differentiate and self-renew from original cells, and this is mainly because that normal stem cells can differentiate to different cell types in the adult organisms(11). Tumor progression is related to the loss of differentiation phenotypes and the gradual acquisition of progenitors- and stem cells-like characteristics. Cancer cell distant metastasis is quite commonly seen in the non-differentiated primary tumors, which results in disease progression and poor prognosis, and this is especially because of resistance to existing treatments(12–14). Some studies indicate that, CSCC cells constituted by the highly heterogeneous cancer cells are not the same. Only a small portion of different tumor cells display the stemness features in the self-replication and maintenance of tumor maintenance. Such cells are the cell bank of CSCC, which have great influence on cancer metastasis and relapse(15, 16).

An increasing number of genomic, epigenomic, transcriptomic and proteomic signatures have been discovered to be related to cancer stemness. Such signatures show correlation with some tumorigenesis-related signaling pathways, thus modulating transcriptional networks and maintaining cancer cell growth and proliferation. The dysregulated transcription of cancer cells frequently leads to the acquisition of stemness characteristics as well as de-differentiation of oncogenes

through altering key signaling pathways that regulate the phenotypes of normal stem cells(17–19). Over the last one decade, The Cancer Genome Atlas (TCGA) and the GEO databases have displayed the landscapes of primary cancers through generating the comprehensive molecular profiles comprised of histopathological and clinical annotations, together with the genomic, epigenomic, transcriptomic and post-translational proteomic characteristics(20, 21). The Progenitor Cell Biology Consortium (PCBC) database(22) covers the expression profiles of the embryonic stem cells (ESCs) and the induced pluripotent stem cells (iPSCs), which can be used for identifying the CSCC stemness genes in later analysis.

In this work, we constructed a model based on the PCBC database-derived stem cell sequencing data for estimating the stemness index (mRNAsi) values for CSCC samples obtained via TCGA and the GEO databases, examining the relationships with patient clinicopathological features and prognosis, and identifying the mRNAsi-related stemness genes extracted through WGCNA. Besides, those most representative stemness genes were extracted for constructing the model to predict the prognosis for CSCC by Lasso and stepwise regression analysis. Notably, our model will assist the clinicians in accurately predicting the prognosis for CSCC patients and selecting individualized treatments.

Methods

Database sources and pre-processing

The overall flow diagram of present study was summarized in Fig. 1. We acquired the level 3 standardized data for the RNA sequences from TCGA via the GDC Data Portal using the TCGAbiolinks functions of R package, namely, GDCdownload, GDCquery, and GDCprepare (http://www.r-project.org). A total of 304 samples were collected in this work for analysis. At the same time, GSE44001 dataset enrolling 97 CSCC samples was obtained from the GEO database(23). Moreover, the RNA sequencing profiles were acquired from the PCBC database through the PCBC Synapse Portal (https://www.synapse.org/pcbc), including 16 ESC and 77 iPSC samples.

We processed the TCGA-derived CSCC RNA-Seq data by the following steps: First of all, we retained primary solid tumor samples with available expression data. Secondly, we converted the Ensembl ID into the gene symbol, and retained protein-encoding genes. For samples having several gene symbols, we used the average expression value for further analysis. For survival analysis, we removed samples with survival time < 0 day.

GEO data were preprocessed according to the following steps: Firstly, we retained samples that had complete expression data, then, converted information on the survival time of samples in the unit of day. Later, we transformed the probe to gene symbol, but eliminated probes related to several genes. The average expression value was used in the situation that having several gene symbols. Table 1 presents the clinical statistics of both TCGA and GSE44001 datasets after preprocessing.

Calculation of OCLR-derived mRNAsi

To calculate mRNAsi based on the mRNA expression data, we constructed a prediction model using the PCBC-derived pluripotent stem cell samples (ESC and iPSC) through the one-class logistic regression (OCLR) method(24). In order to make sure that the TCGA cohort was comparable, gene names in the form of Ensembl IDs were mapped to the Human Genome Organization (HUGO) (25), and genes with no mapping were removed. As for the training matrix obtained, altogether 12998 mRNA expression profiles were collected, which were obtained on the basis of every available PCBC sample. For all samples, the respective average expression values were determined and used in centralization. Finally, we used the R package gelnet (v1.2.1) OCLR function to calculate the gene weight vector based on those processed data.

For all CSCC samples, their respective expression profiles were determined by respective Spearman correlation coefficient and model gene weight vector. Then, all coefficients of Spearman correlation were transformed linearly, while the minimal coefficient of samples was eliminated, and then the value acquired was divided by maximum for calculating mRNAsi of each respective sample, yielding the range of distribution of [0,1].

TCGA data were utilized to compare diverse mRNAsi values among samples with different ages, genders, tumor grades and stages. To investigate the relationship of mRNAsi with mutations related to CSCC in the key oncogenes, the R package TCGAbiolinks(26) (V2.14.0) was utilized to detect the mutations in TCGA-CSCC, while the R package maftools (V2.0.16) (27) was adopted for plotting the mutation

frequency diagram. To explore the associations between mRNAsi and molecular subtypes of CSCC samples, TCGA-CSCC samples were clustered in accordance with gene expression profiles collected by the R package TCGAbiolinks (v2.14.0). Later, R package estimate was used to calculate the immune and stromal scores of each sample, so as to further explore the correlations of immune and stromal scores with the sample miRNAsi.

The construction of mRNAsi-related gene modules through WGCNA

First of all, transcripts with median absolute deviation (MAD) of > median and over 75% TPM > 1 were selected based on the sample expression profiles, while hierarchical clustering was also conducted for sample clustering analysis. Then, samples with distance > 80000 were taken as the outlier samples for screening. The coefficient of Pearson correlation was used to determine the inter-transcript distance, the R software package Weighted Gene Co-expression Network Analysis (WGCNA) was utilized to establish distance between any two transcripts, whereas 8 was set as the soft threshold to screen co-expression modules. Notably, the co-expression network is suggested to be consistent with scale-free network. For a node with a connectivity k, its logarithm (log(k)) must show negative correlation negatively correlated with the logarithm of specific node occurrence probability (log(P(k)), with a correlation coefficient of > 0.9. Proper β value was adopted for guaranteeing that the network was a scale-free one. Later, the expression matrix was transformed to an adjacent one, which was then additionally transformed to a topological one for gene clustering based on Topological Overlap Matrix (TOM) by using the average-linkage hierarchical clustering approach according to the mixed dynamic shear tree standard. Besides, in every gene network module, the gene number was set as 80 at least. Then,

gene module was determined by the dynamic shear approach, and eigengene values were calculated for all modules successively. Afterwards, clustering analysis were performed on all modules, and adjacent modules were merged into a new one, with reset appropriate height, deepSplit and minModuleSize values. The correlations of the obtained gene modules with mRNAsi were calculated to search for the gene modules with high correlation for further research. Moreover, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analysis was performed by utilizing the WebGestaltR (v0.4.2) R package to examine the signaling pathways affected by the genes involved in these modules.

Survival analysis

Samples containing survival data from TCGA dataset were divided into training and test sets at a 6:4 ratio. For the sake of eliminating the influence of random allocation bias on model stability, each sample was randomly grouped for 100 times with replacement. Univariate analysis was carried out by using the Cox proportional hazards regression model, and survival data was analyzed through the survival coxph function of R package(28). A p-value < 0.05 suggested that the difference was statistically significant.

Construction and validation of the prognostic stemness-related gene model

Lasso regression analysis was conducted using the R package glmnet(29) function. Besides, stepwise regression analysis was carried out by R package MASS function according to the Akaike data criteria. Finally, a risk model was established with genes selected. Thereafter, related gene expression data were selected based on TCGA and GEO datasets, respectively, which were later used in the model to calculate Riskscore values of all samples. Later, Riskscore values were transformed into the zscore, and the Riskscore of 0 was utilized as a threshold to classify samples to high (Risk-H) or low (Risk-L) risk group. In addition, we determined the accuracy, stability and efficiency of the model to classify CSCC prognosis through KM and receiving operating characteristic (ROC) analyses using TCGA test dataset as well as GSE44001 dataset.

Function analysis of Riskscore in CSCC patients

The R package GSVA(30) ssGSEA function was employed to assess the score of KEGG enrichment analysis. At the same time, we determined the relationships between Riskscore values and ssGSEA scores, and conducted clustering analysis for all samples.

Relationships between Riskscore and clinicopathological features in CSCC paitents

We used relevant Riskscore values and clinical characteristics to plot the forest plot and construct the nomogram. Moreover, the relationships between patient survival and Riskscore values as well as clinical characteristics were examined through the univariate and multivariate Cox proportional hazards regression models.

Results

Relationships between mRNAsi and clinicopathological features, molecular classification, immune status and gene mutations in CSCC patients

First of all, mRNAsi of each TCGA-CSCC sample was calculated by OCLR method. Further analysis found that there was no significant difference in mRNAsi across samples with diverse ages, genders, grades and stages (Fig.S1). Then, we screened out 10 most frequent mutations and revealed that samples carrying DMD, KMT2C, EP300 and MUC4 mutations (Fig.S2, $p < 0.05$) showed markedly elevated mRNAsi values compared with samples having no mutation. MUC4 and KMT2C have been recognized as the promoter genes of CSCC, which modulate CSCC occurrence, development and malignant transformation, and their mutations are markedly correlated with tumor metastasis, relapse, and poor patient prognosis(31, 32).

Then, we compared the difference in mRNAsi across various samples. As a result, difference in mRNAsi was not significant across different CSCC subtypes on the basis of gene expression profiles and iCluster subgroups (Fig.S3A, 3C). Subsequently, we compared the difference in mRNAsi among CSCC samples

with different CIMP subtypes. Our results indicated that there was significant difference between groups (Fig.S3B). According to our results, mRNAsi was significantly negatively correlated with the stromal cell infiltration (StromalScore and ESTIMATEScore) in TCGA-CSCC tumor samples, but it showed weak correlation with immune cell infiltration (ImmuneScore) (Fig.S4).

In conclusion, mRNAsi showed tight correlation with gene mutations and infiltration of stroma cells, and it was a creditable molecular marker for CSCC.

Selection of possible stemness-related genes and functional enrichment analysis

According to Fig. 2A, some outlier samples were observed, samples that had > 80000 distance were identified to be outlier samples in screening, and 303 samples were obtained at last (Table S1). Then, WGCNA was used to construct the weight co-expression network for guaranteeing the scale-free network (Fig. 2B). Subsequently, gene modules were determined by dynamic shear approach, which were then conducted clustering analysis. Additionally, modules with close distance were further merged into the new module, having height, deepSplit and minModuleSize set to 0.25, 2 and 80, respectively. Altogether 16 modules were obtained at last (Fig. 2C). It should be noted that, grey module represented gene set that was not clustered to any one of the rest modules. Based on Fig. 2D, the relationships between the 16 module eigenvectors and mRNAsi as well as clinical features were then determined, separately. Clearly, the pink module showed the most significant positive correlation with mRNAsi, while green and cyan modules showed significant negative correlations, among which, the pink module contained 286 genes, the green module covered 493 modules, and the cyan module had 114 gene, yielding a total of 893 genes(Table S1). Notably, these genes were identified as the possible stemness-related genes.

Subsequently, we carried out functional enrichment analysis on genes in the pink (Fig. 3 and Table S2), green and cyan modules (Fig. 4 and Table S3). As a result, these genes were mainly enriched to the pathways that were significantly correlated with cervical cancer genesis and development, including ECM-receptor interaction, the cGMP-PKG signaling pathway, the PI3K-Akt signaling pathway, and Human papillomavirus infection(33–35).

Prognostic risk prediction model establishment on the basis of possible stemness-related genes and accuracy validation

The detailed information of Training (n = 175) and test (n = 116) set samples are displayed in Table S4. Differences in the clinical factors showed no statistical significance between the training and test sets, with the only exception of tumor grade, indicating reasonable sample grouping (Table 2).

Straight after univariate survival analysis, the prognosis-specific stemness-related genes were narrowed through lasso regression to reduce the gene number for the risk model. A stemness-related 8 gene risk model was finally constructed: RiskScore = 0.6457*MRPS24-0.9248*RHOH + 0.5665*TNFRSF11B + 0.2637*PPP1R1 4A + 0.4230*CPQ-1.7806*KCNMA1 + 0.3068*DPYSL4 + 0.7492*AOC3

As shown in Fig. 5, all of these 8 genes can be able to classify samples in the TCGA training set to two groups that had distinct OS ($P < 0.05$).

Then, samples in the training set were divided into Risk-L and -H groups based on the Riskscore. Figure 6A shows survival distributions in Risk-L and Risk-H groups, and the 8 stemness gene expression in samples. According to the figure, samples in training set that had greater Riskscore values had significantly reduced OS compared with those that had lower Riskscore values, which revealed that samples that had higher Riskscore values showed more dismal prognosis. In addition, the high expression levels of PPP1R14A, TNFRSF11B, CPQ, AOC3 and DPYSL4 were associated with dismal prognosis for samples in training set, and they might serve as risk factors. Moreover, the increased KCNMA1 and RHOH expression was associated with the better prognosis for samples in training set and were thus the protective factors. Also, we carried out ROC analysis to classify the prognosis for samples in training set based on Riskscore value. The efficiency of our model in predicting the OS at 1–5 years for training set samples was examined, with the mean AUC of up to 0.87 (Fig. 6B). At last, we drew the KM curve on the basis of OS for samples in Risk-L and Risk-H groups, thus comparing the difference in survival between both groups. As observed from Fig. 6C, Risk-H group of the training set had remarkably shorter OS than that of Risk-L group ($P < 0.0001$).

To validate the reliability of our model, the expression data of the above-mentioned 8 stemness-related genes of TCGA-CSCC samples in test set were collected, and the Riskscore values for all test set samples were measured. Similar to previous analysis results, samples in test set with greater Riskscore values had markedly reduced OS than those with lower Riskscore values, with the mean average AUC of up to 0.8 (Fig.S5). At last, an external data set was used to validate our model accuracy in the prediction of CSCC prognosis and similar findings were obtained (Fig.S6).

In conclusion, our risk model constructed on the basis of sample survival and possible stemness genes using Lasso and stepwise regression analysis was efficient in predicting CSCC prognosis.

Signaling pathways associated with Riskscore

As shown in Fig. 7A, there were 22 related KEGG pathways obtained, all of which had correlation coefficient > 0.3. Next, above-mentioned 22 pathways were chosen for clustering analysis on the basis of sample enrichment scores. 21 out of those 22 pathways showed negative correlation with Riskscore, which were mainly associated with cellular metabolism and immunity (Fig. 7B).

The 8-stemness gene signature was an independent factor for predicting CSCC prognosis

Nomogram has been identified as an approach to effectively and intuitively present risk model results, and it is used conveniently to predict the prognosis for patients. To be specific, the length of straight line in nomogram represents the influences of different parameters together with their significance on prognosis. In this study, a nomogram was constructed using Riskscore, age, tumor grade and T stage. As displayed in Fig. 8, Riskscore most significantly affected the prediction of survival rate. Furthermore, both Riskscore and patient clinical characteristics were utilized for plotting the forest plot. According to Fig. 9, Riskscore had the HR of around 2.43 ($P < 0.001$), which indicated that the stemness-related gene signature might serve as a vital risk factor in predicting CSCC prognosis. Afterward, univariate and multivariate COX regression analyses were conducted to assess the independence of 8-stemness gene signature on survival prediction. As a result, only 8-stemness gene signature showed significant correlation with the prognosis for CSCC patients, both upon univariate (HR = 2.3410, p = 8.3E-12) and multivariate (HR = 2.6570, p = 1.9E-12) analyses (Table 3). Thus, our 8-stemness gene signature might be used to be an independent prognostic factor for CSCC.

Comparison between our 8-stemness gene signature risk model and other reported prognostic models

Three CSCC prognostic models (namely, the 8-, 5- and 3-gene signature) (36–38) reported in literature were selected, which were then compared with the 8-stemness gene signature constructed in our study. The corresponding Riskscore values of each prognostic model were calculated based on gene expression profiles for samples derived from TCGA-CSCC dataset, which were further converted to zscore values. Then, samples were divided into high- (Riskscore > 0) and low-risk (Riskscore > 0) groups, and OS difference were compared. Figure 10 displays the ROC and overall survival KM curves for those three models. Obviously, the 8-stemness gene signature constructed in this study had higher ROC value than other three models, with statistically significant difference in OS between low and high risk subgroups.

Discussion

In current study, expression data for the PCBC database-derived pluripotent stem cell samples were obtained using the OCLR method to calculate stemness indexes (mRNAsi) of samples derived from the TCGA dataset. Meanwhile, corresponding clinicopathological features, gene mutation status and survival data were retrieved. Further analysis results revealed that mRNAsi having significant correlation with key oncogene mutation status, stroma cells infiltrating and CIMP classifications among CSCC cases. Additionally, functions of the 893 potential mRNAsi-related stemness genes within the gene modules were analyzed by enrichment analysis and WGCNA. Besides, we conducted shrinkage estimate and univariate analysis to screen the most representative stemness genes for prognosis to construct the 8-stemness gene signature for CSCC. Then, samples obtained based on the TCGA-CSCC dataset and GSE44001 external dataset were incorporated into the model and classified according to the Riskscore value to evaluate the model efficiency and stability in the prediction of patient prognosis. Subsequently, we examined the associations between Riskscore and relevant clinicopathological characteristics as well

as signal transduction pathways. At last, our 8-stemness gene signature was compared with three CSCC prognostic signatures reported in literature.

As a result, one (KCNMA1) in the eight genes was suggested previously to be involved in the progression of CSCC, pathogenesis and malignant transformation, indicating that our bioinformatic mining results were highly reliable and accurate. (39). However, the relationships of the remaining 7 genes with CSCC are not validated in clinical or basic study. Typically, RHOH participates in the migration of prostate cancer cells(40). PPP1R14A has been demonstrated to evaluate the prognostic risk of gastric cancer(41). DPYSL4, which is associated with mitochondrial supercomplexes, can regulate energy metabolism in tumor cells and be involved in cancer invasion and progression(42). Dysregulation of TNFRSF11B is shown to be involved in breast cancer genesis and development and correlates with poor prognosis in patients(43). Nonetheless the associations of genes with CSCC are not examined yet which should be further investigated.

Nowadays, with the rapid development of high-throughput genome-wide sequencing techniques, the bioinformatic big data for the large-scale CSCC samples can be easily obtained to explore the hidden information. Despite the superiority of our stemness-related gene signature, there are certain limitations in the present work like other bioinformatic studies. First of all, the experimental verification is lacking, and further experimental validation should be performed on the potential mechanisms underlying the signature genes. In addition, we have limited TCGA-CSCC samples and relevant clinical data. Therefore, a larger sample size is needed for identifying the key prognostic stemness gene functions for predicting CSCC progression.

Conclusion

To sum up, this study established a novel stemness-related gene signature which was validated as a new and significant prognostic marker for CSCC. Notably, our stemness-related gene signature connected the molecular features of CSCC stem cells and clinical outcomes, which will help to find novel predictors for therapeutic effect and prognosis. Besides, the methods of our study had general applicability and provided some references value for other cancer types.

Abbreviations

CSCC: cervical squamous cell carcinoma; PSC: pluripotent stem cell; OCLR: one-class logistic regression; FIGO: International Federation of Gynecology and Obstetrics; TCGA: The Cancer Genome Atlas;PCBC: Progenitor Cell Biology Consortium; ESCs: embryonic stem cells; iPSCs: the induced pluripotent stem cells; HUGO: Human Genome Organization; MAD: median absolute deviation; WGCNA: Weighted Gene Co-expression Network Analysis; TOM: Topological Overlap Matrix; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology; ROC: receiving operating characteristic.

Declarations

Acknowledgements

Not applicable.

Funding

We gratefully acknowledge the financial support from the Public Welfare Technology Research Project of Zhejiang Province, China. (No. LGF19H040021).

Availability of data and materials

Not applicable.

Author Contribution

JW Z : Project development, administration and supervision

XW S: Methodology development, data analysis, manuscript writing

F W: Data collection and figure organization

JY Z: Manuscript review and editing

All authors read and approved the final manuscript.

Ethics approval and consent to participate

The datasets used and analysed during the current study are available from the corresponding author on reasonable reguest.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no Competing interests.

References

1. Marth C, Landoni F, Mahner S, et al.: Cervical cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology 29: iv262,
2. Chen SB, Yang XH, Weng HR, Liu DT, Li H and Chen YP: Clinicopathological features and surgical treatment of cervical oesophageal Scientific reports 7: 3272, 2017.
3. Uppal S, Rebecca Liu J, Kevin Reynolds R, Rice LW and Spencer RJ: Trends and comparative effectiveness of inpatient radical hysterectomy for cervical cancer in the United States (2012-2015). Gynecologic oncology 152: 133-138,
4. Gil-Moreno A and Magrina JF: Minimally Invasive or Abdominal Radical Hysterectomy for Cervical The New England journal of medicine 380: 794, 2019.
5. Altobelli E, Rapacchietta L, Profeta VF and Fagnano R: HPV-vaccination and cancer cervical screening in 53 WHO European Countries: An update on prevention programs according to income level. Cancer medicine 8: 2524-2534,
6. Feldman S: Screening Options for Preventing Cervical JAMA internal medicine 179: 879-880, 2019.
7. Nanthamongkolkul K and Hanprasertpong J: Predictive Factors of Pelvic Lymph Node Metastasis in Early-Stage Cervical Oncology research and treatment 41: 194-198, 2018.
8. Alvarado-Cabrero I, Roma AA, Park KJ, Rutgers JKL and Silva EG: Factors Predicting Pelvic Lymph Node Metastasis, Relapse, and Disease Outcome in Pattern C Endocervical Adenocarcinomas. International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists 36: 476-485,
9. Bosque MAS, Cervantes-Bonilla MA and Palacios-Saucedo GDC: Clinical and dosimetric factors associated with the development of hematologic toxicity in locally advanced cervical cancer treated with chemotherapy and 3D conformal radiotherapy. Reports of practical oncology and radiotherapy : journal of Greatpoland Cancer Center in Poznan and Polish Society of Radiation Oncology 23: 392-397, 2018.
10. Matsuo K, Machida H, Mandelbaum RS, Konishi I and Mikami M: Validation of the 2018 FIGO cervical cancer staging system. Gynecologic oncology 152: 87-93,
11. Daily K, Ho Sui SJ, Schriml LM, et al.: Molecular, phenotypic, and sample-associated data to describe pluripotent stem cell lines and derivatives. Scientific data 4: 170030,
12. Flanagan DJ, Hodder MC and Sansom OJ: Microenvironmental cues in cancer stemness. Nature cell biology 20: 1102-1104,
13. Saygin C, Matei D, Majeti R, Reizes O and Lathia JD: Targeting Cancer Stemness in the Clinic: From Hype to Hope. Cell stem cell 24: 25-40,
14. Prasad S, Ramachandran S, Gupta N, Kaushik I and Srivastava SK: Cancer cells stemness: A doorstep to targeted therapy. Biochimica et biophysica acta. Molecular basis of disease 1866: 165424, 2020.

15. Hou T, Zhang W, Tong C, et al.: Putative stem cell markers in cervical squamous cell carcinoma are correlated with poor clinical outcome. BMC cancer 15: 785,
16. Zhang Y, An J, Liu M, et al.: Efficient Isolation, Culture, Purification and Stem Cell Expression Profiles of Primary Tumor Cells Derived from Uterine Cervical Squamous Cell Carcinoma. American journal of reproductive immunology: e13251,
17. Zhao M, Chen Z, Zheng Y, et : Identification of cancer stem cell-related biomarkers in lung adenocarcinoma by stemness index and weighted correlation network analysis. Journal of cancer research and clinical oncology2020.
18. Qin B, Zou S, Li K, et : CSN6-TRIM21 axis instigates cancer stemness during tumorigenesis. British journal of cancer2020.
19. Bai KH, He SY, Shu LL, et al.: Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness Cancer medicine2020.
20. Feng G, Beilei Z, Caizhi C and Wen Z: Analysis of CASP12 diagnostic and prognostic values in cervical cancer based on TCGA database. Bioscience reports
21. Wang XB, Cui NH, Liu XN, et al.: Identification of DAPK1 Promoter Hypermethylation as a Biomarker for Intra-Epithelial Lesion and Cervical Cancer: A Meta-Analysis of Published Studies, TCGA, and GEO Datasets. Frontiers in genetics 9: 258,
22. Salomonis N, Dexheimer PJ, Omberg L, et al.: Integrated Genomic Analysis of Diverse Induced Pluripotent Stem Cells from the Progenitor Cell Biology Consortium. Stem cell reports 7: 110-125, 2016.
23. Lee YY, Kim TJ, Kim JY, et : Genetic profiling to predict recurrence of early cervical cancer. Gynecologic oncology 131: 650-654, 2013.
24. Sokolov A, Paull EO and Stuart JM: One-Class Detection Of Cell States In Tumor Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing 21: 405-416, 2016.
25. Liu ET: The Human Genome Organisation (HUGO). The HUGO journal 3: 3-4,
26. Colaprico A, Silva TC, Olsen C, et al.: TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic acids research 44: e71,
27. Mayakonda A, Lin DC, Assenov Y, Plass C and Koeffler HP: Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome research 28: 1747-1756, 2018.
28. Lv Y, Duanmu J, Fu X, Li T and Jiang Q: Identifying a new microRNA signature as a prognostic biomarker in colon PloS one 15: e0228575, 2020.
29. Engebretsen S and Bohlin J: Statistical predictions with Clinical epigenetics 11: 123, 2019.
30. Hanzelmann S, Castelo R and Guinney J: GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics 14: 7,

31. Xu D, Liu S, Zhang L and Song L: MiR-211 inhibits invasion and epithelial-to-mesenchymal transition (EMT) of cervical cancer cells via targeting MUC4. Biochemical and biophysical research communications 485: 556-562,
32. Scholl S, Popovic M, de la Rochefordiere A, et al.: Clinical and genetic landscape of treatment naive cervical cancer: Alterations in PIK3CA and in epigenetic modulators associated with sub-optimal outcome. EBioMedicine 43: 253-260,
33. Gong L, Lei Y, Tan X, et al.: Propranolol selectively inhibits cervical cancer cell growth by suppressing the cGMP/PKG Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 111: 1243-1248, 2019.
34. Zhang W, Zhou Q, Wei Y, et al.: The exosome-mediated PI3k/Akt/mTOR signaling pathway in cervical International journal of clinical and experimental pathology 12: 2474-2484, 2019.
35. Hemmat N and Bannazadeh Baghi H: Association of human papillomavirus infection and inflammation in cervical Pathogens and disease 772019.
36. Ding TT, Ma H and Feng JH: A three-gene novel predictor for improving the prognosis of cervical Oncology letters 18: 4907-4915, 2019.
37. Xie F, Dong D, Du N, et : An 8gene signature predicts the prognosis of cervical cancer following radiotherapy. Molecular medicine reports 20: 2990-3002, 2019.
38. Liu J, Nie S, Gao M, et al.: Identification of EPHX2 and RMI2 as two novel key genes in cervical squamous cell carcinoma by an integrated bioinformatic analysis. Journal of cellular physiology 234: 21260-21273, 2019.
39. Ramirez A, Vera E, Gamboa-Dominguez A, Lambert P, Gariglio P and Camacho J: Calcium-activated potassium channels as potential early markers of human cervical Oncology letters 15: 7249-7254, 2018.
40. Tajadura-Ortega V, Garg R, Allen R, et al.: An RNAi screen of Rho signalling networks identifies RhoH as a regulator of Rac1 in prostate cancer cell migration. BMC biology 16: 29,
41. Hou JY, Wang YG, Ma SJ, Yang BY and Li QP: Identification of a prognostic 5-Gene expression signature for gastric Journal of cancer research and clinical oncology 143: 619-629, 2017.
42. Nagano H, Hashimoto N, Nakayama A, et al.: p53-inducible DPYSL4 associates with mitochondrial supercomplexes and regulates energy metabolism in adipocytes and cancer cells. Proceedings of the National Academy of Sciences of the United States of America 115: 8370-8375,
43. Luo P, Lu G, Fan LL, et al.: Dysregulation of TMPRSS3 and TNFRSF11B correlates with tumorigenesis and poor prognosis in patients with breast Oncology reports 37: 2057-2062, 2017.

Tables

Due to technical limitations, table 1, table 2 and table 3 are only available as a download in the Supplemental Files section.

Figures

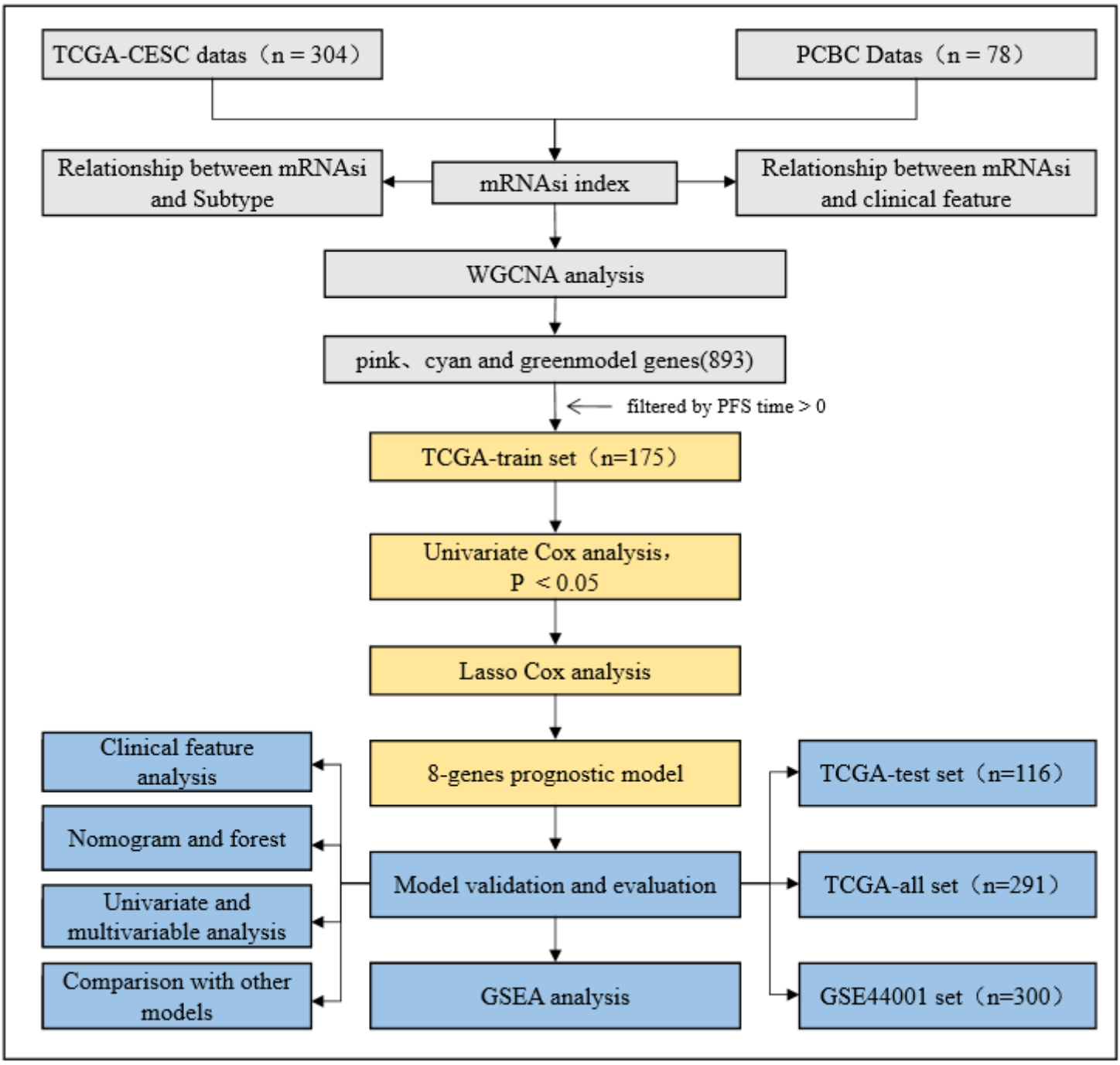

Figure 1

Flow diagram of methods for establishing of prognostic stemness-related gene signature in CSCC

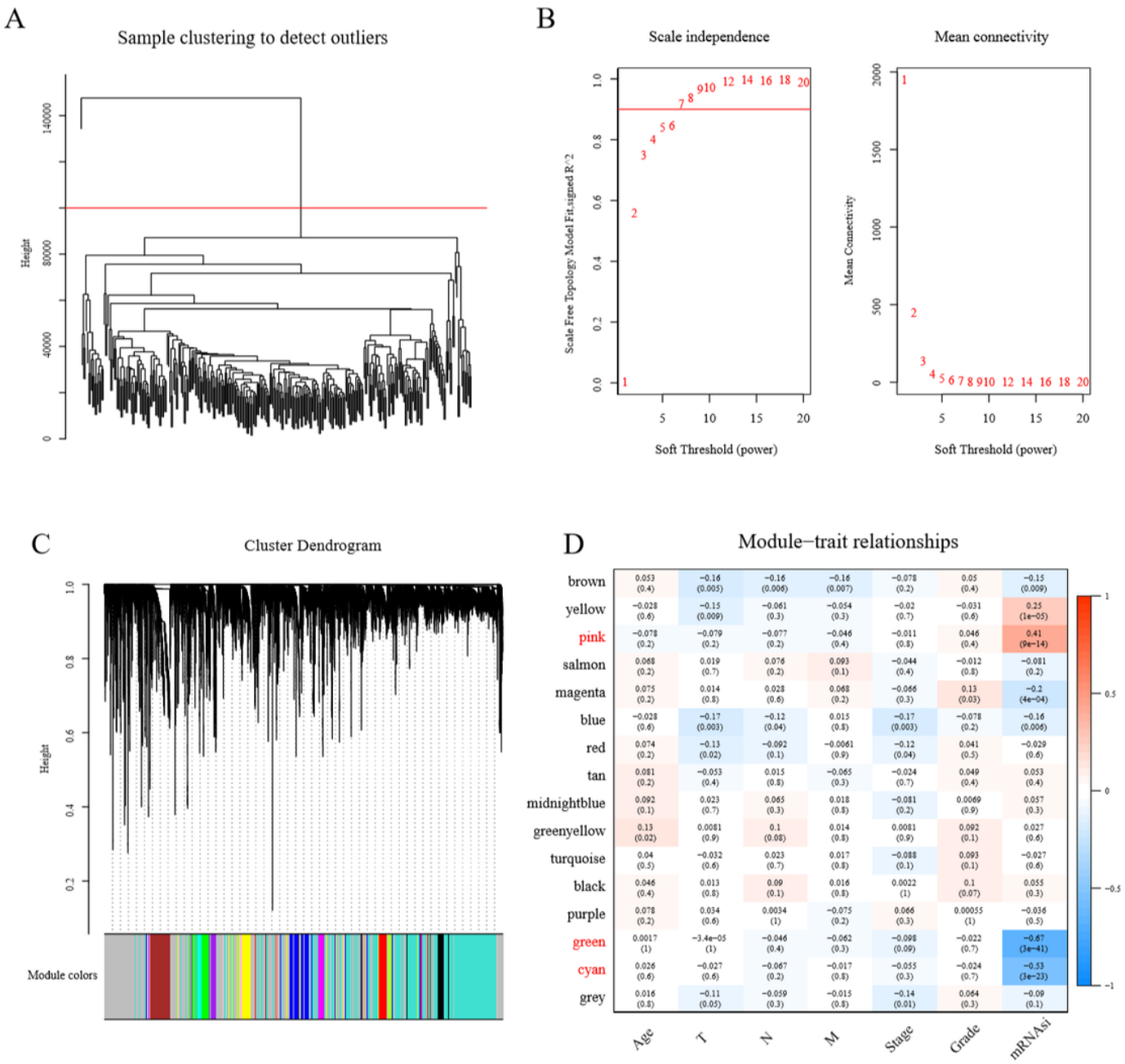

Figure 2

Stemness-related gene modules mined through WGCNA (A) Sample clustering analysis. (B) Analysis of network topology for various soft-thresholding powers. (C) Gene dendrogram and module colors. (D) Correlation between each module and clinicopathological features.

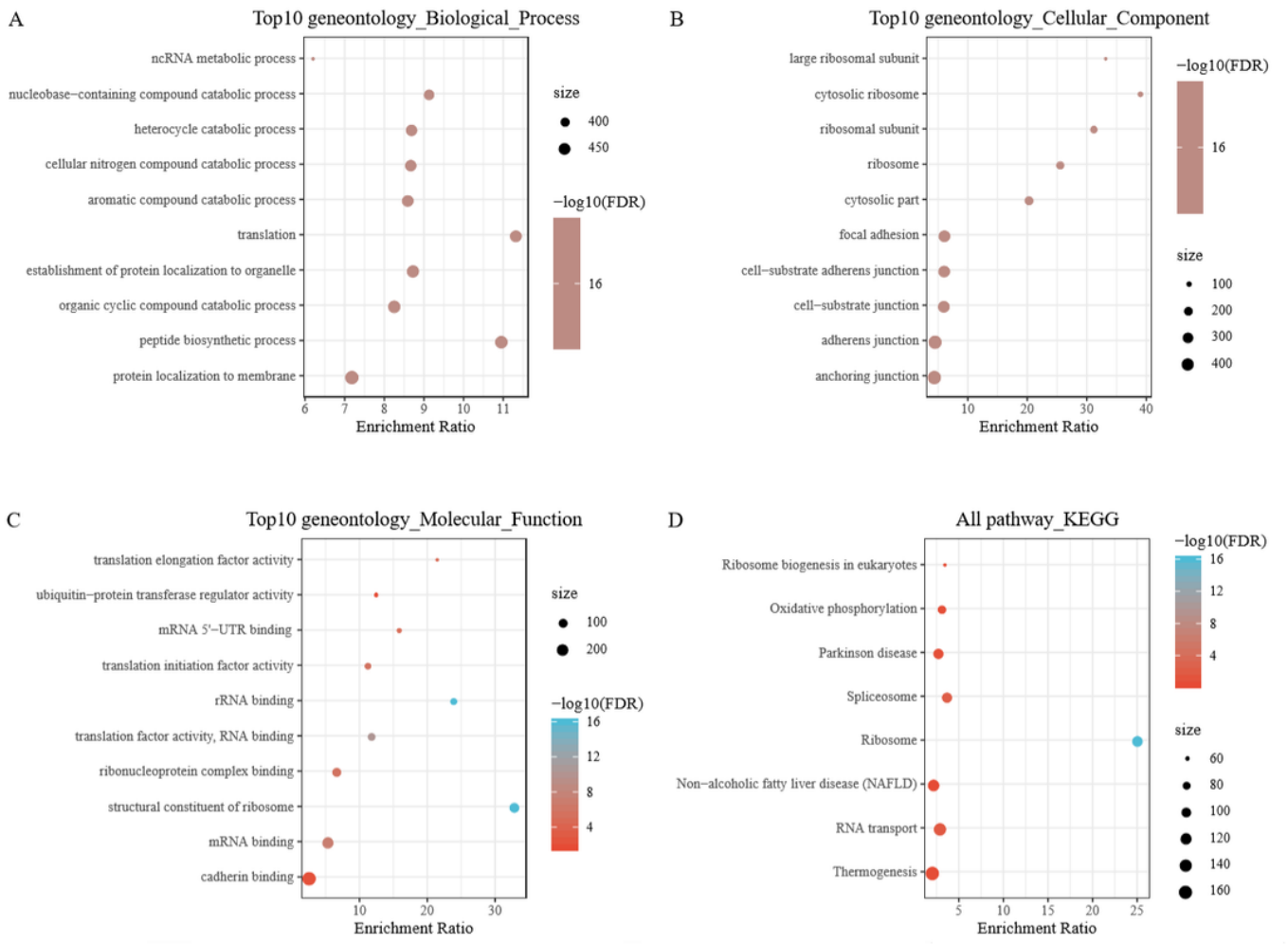

Figure 3

The KEGG pathway and GO enrichment analysis of the genes in pink module (A) Top10 BP terms enriched by the genes in pink module. (B) Top10 CC terms enriched by the genes in pink module. (C) Top10 MF terms enriched by the genes in pink module. (D) Top10 GO KEGG pathways enriched by the genes in pink module. BP: biological peocess; CC: cellular component; MF: molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes

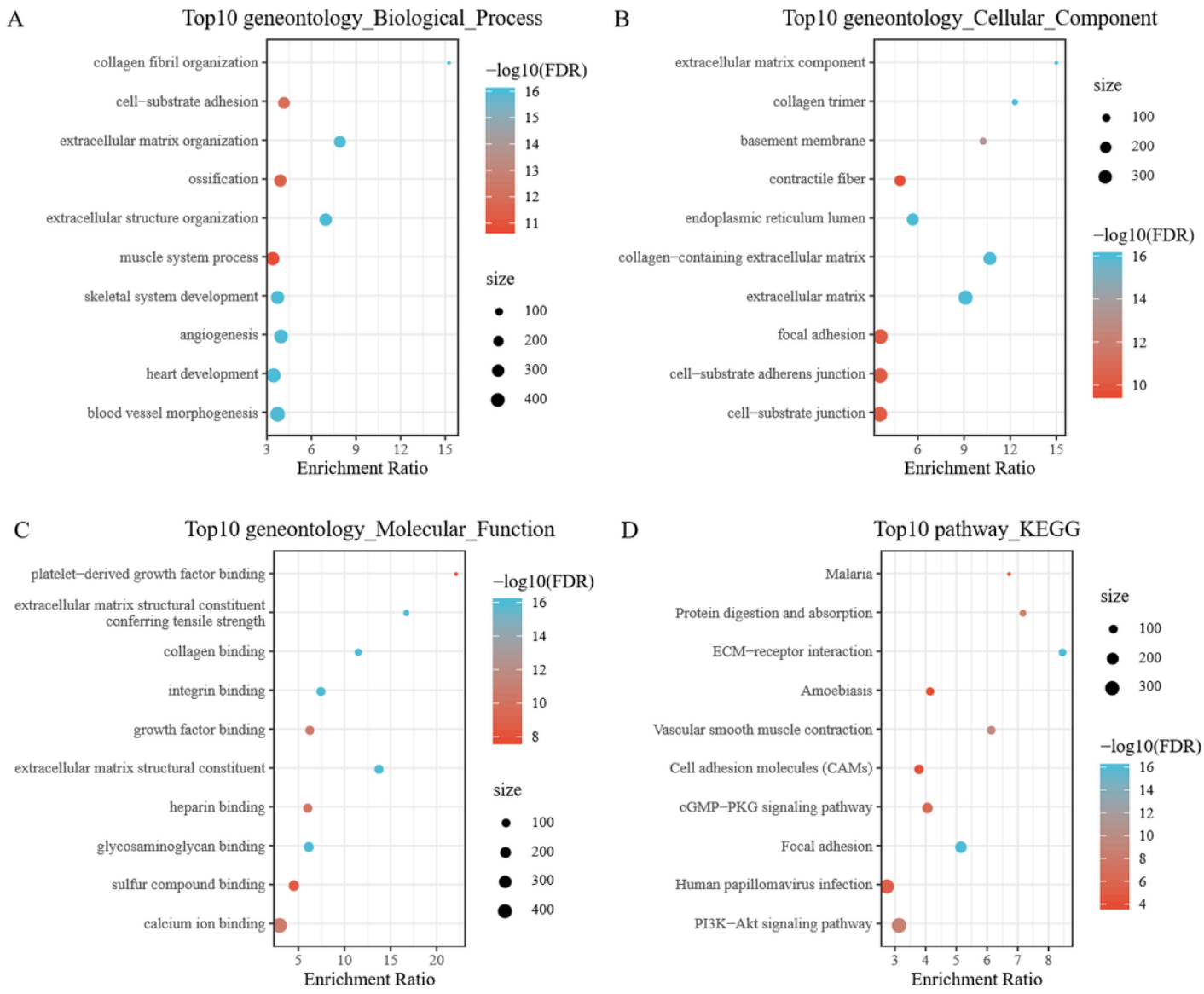

Figure 4

The KEGG pathway and GO enrichment analysis of the genes in green and cyan modules (A) Top10 BP terms enriched by the genes in green and cyan modules. (B) Top10 CC terms enriched by the genes in green and cyan modules. (C) Top10 MF terms enriched by the genes in green and cyan modules. (D) Top10 GO KEGG pathways enriched by the genes in green and cyan modules. BP: biological peocess; CC: cellular component; MF: molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes

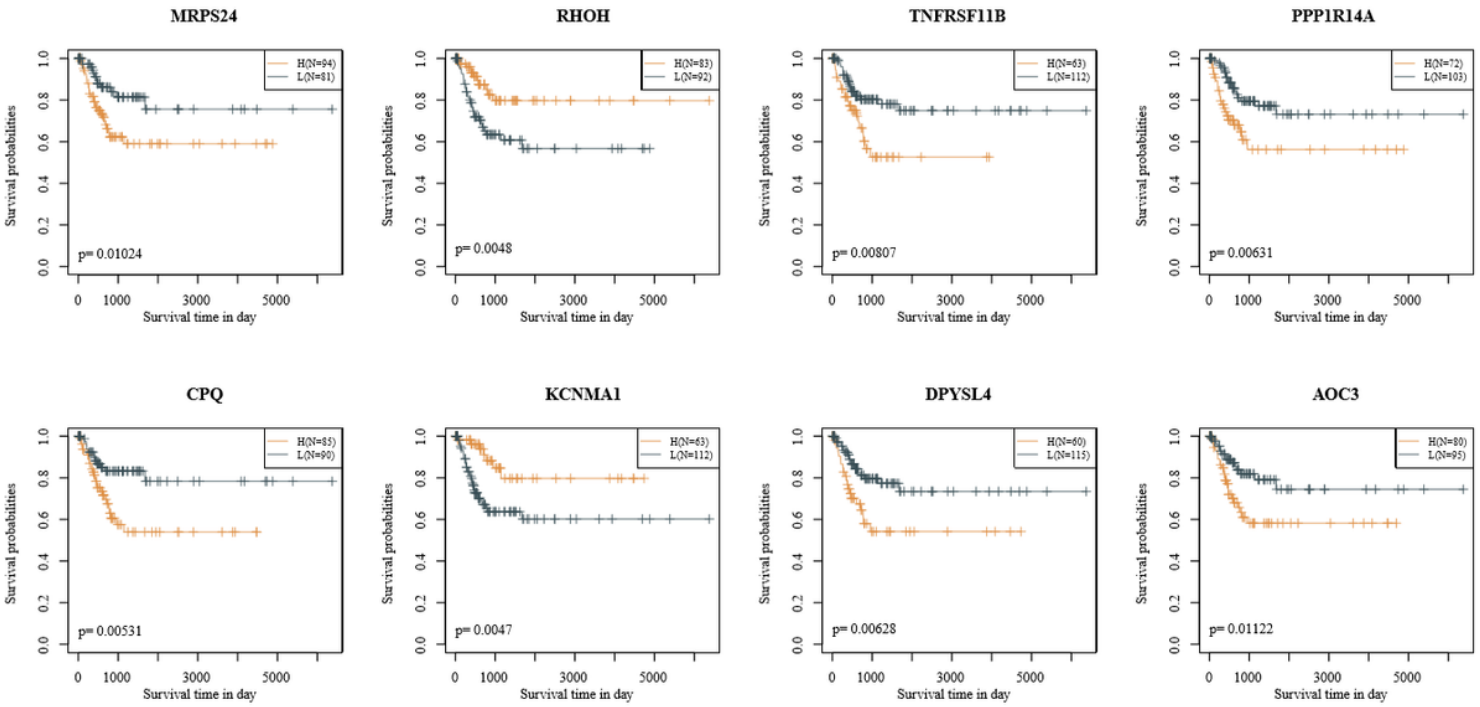

Figure 5

The KM curve on the basis of expression level of each 8 gene in the risk model

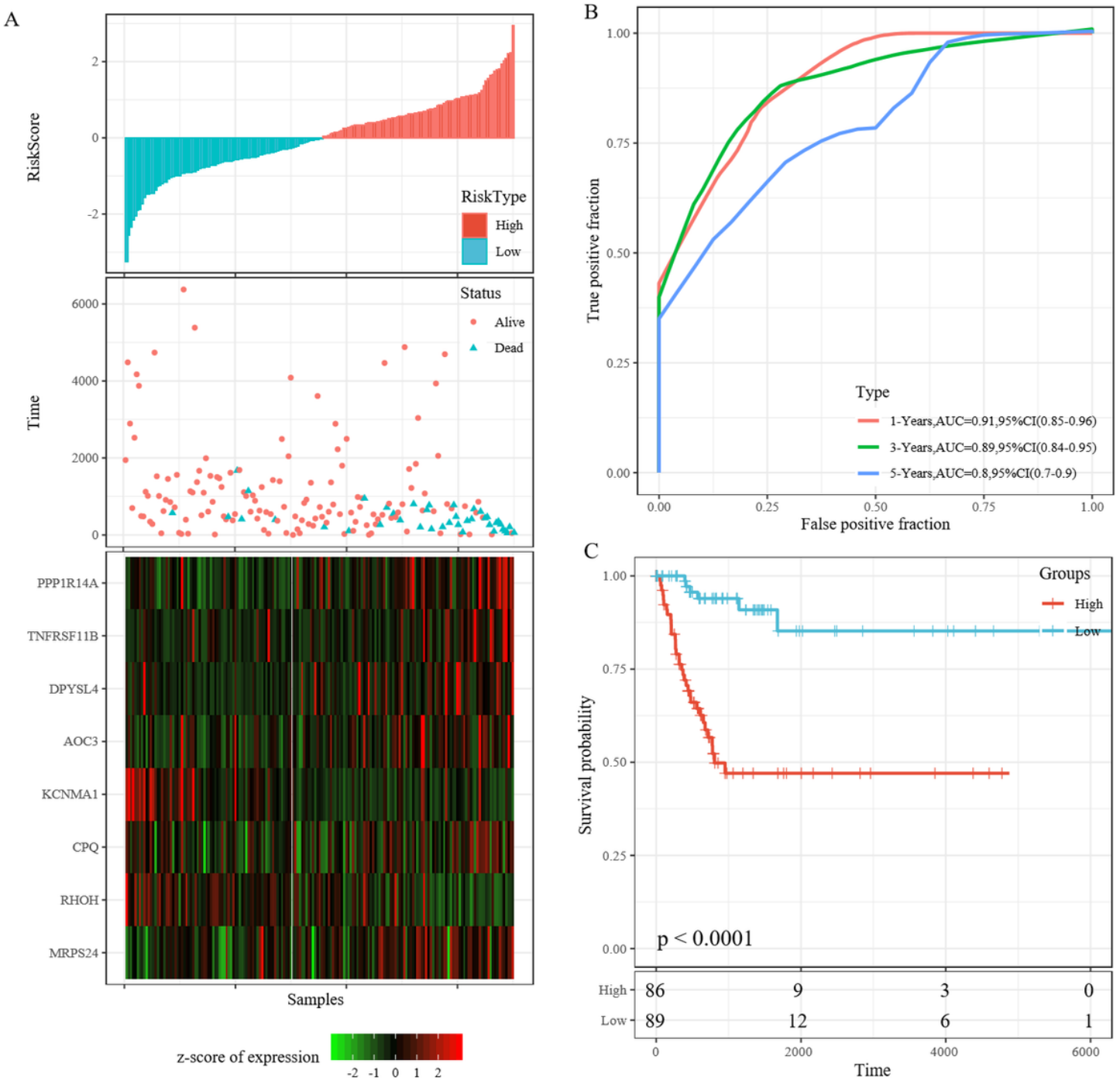

Figure 6

The ROC curves of risk model and KM curve on the basis of OS in the training set (A) Survival distributions in Risk-L and Risk-H groups, and the 8 stemness gene expression in samples. (B) The ROC curves and corresponding AUC in predicting the OS at 1-5 years of the risk model. (C) The KM curve on the basis of OS for samples in Risk-L and Risk-H groups.

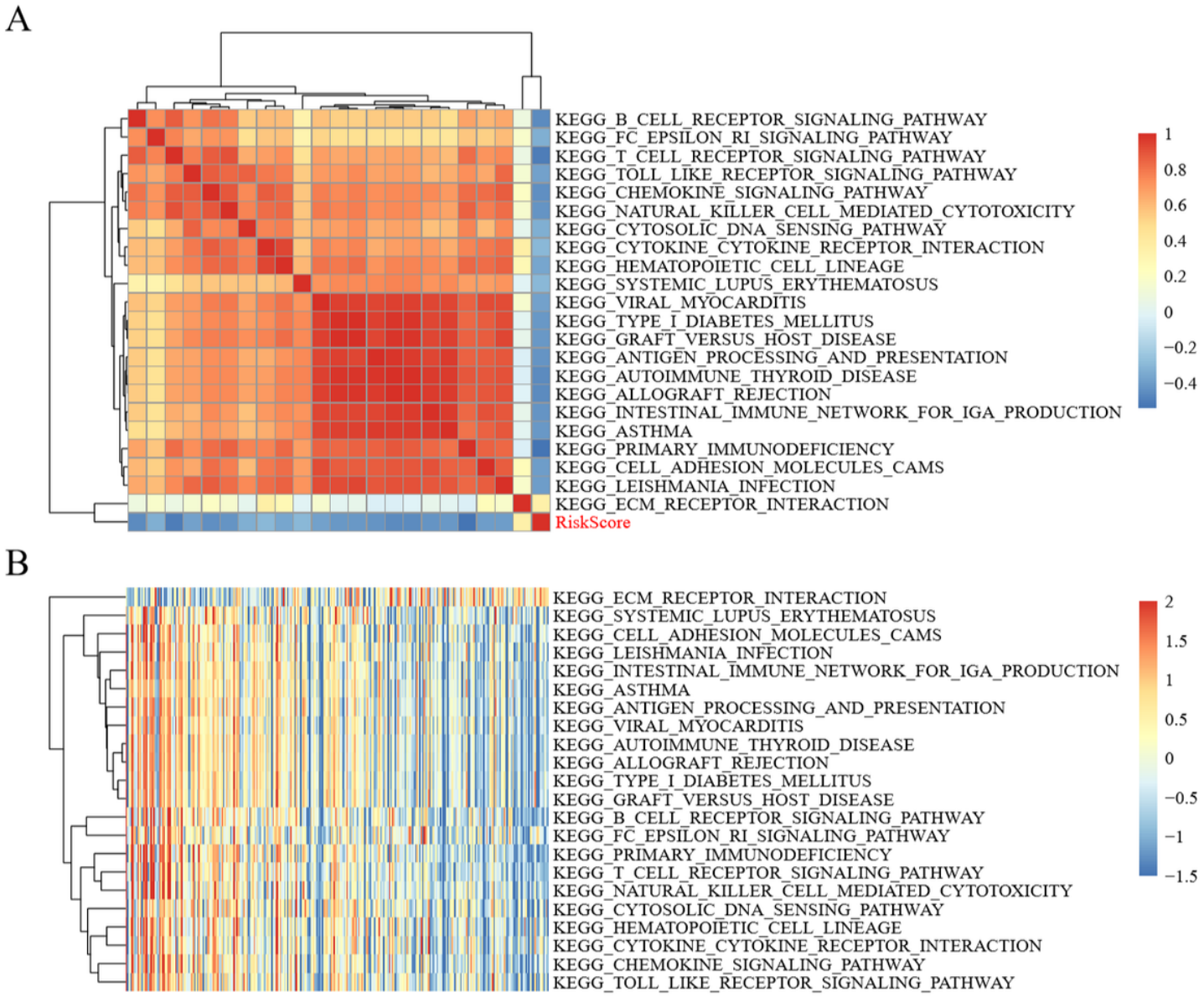

Figure 7

Signaling pathways associated with Riskscore (A) 22 related KEGG pathways with correlation coefficient >0.3. (B) Clustering analysis results of 22 KEGG pathways on the basis of sample enrichment scores

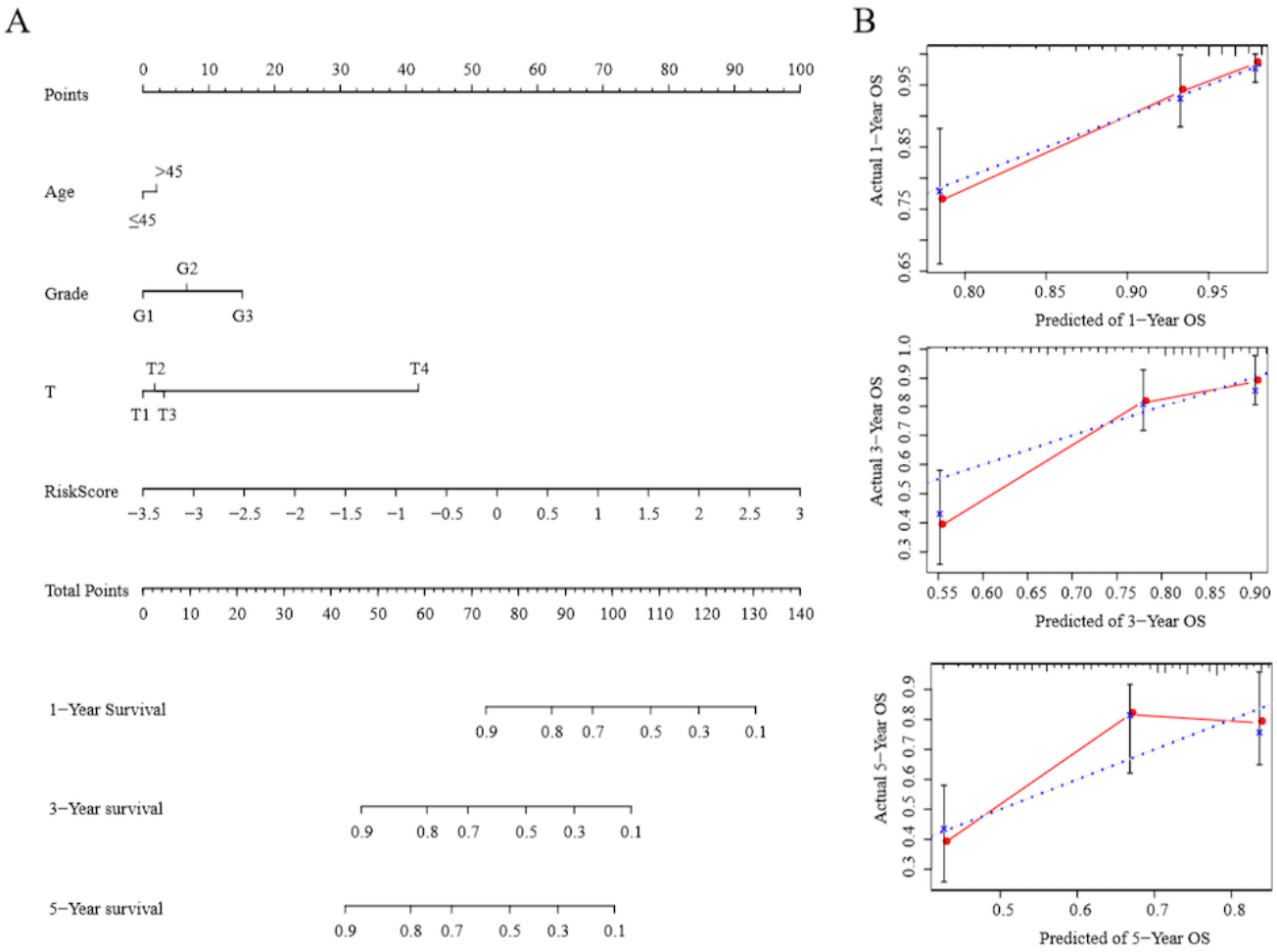

Figure 8

The stemness-related gene signature had more impacts than other clinicopathological features on survival prediction for CSCC (A) Nomogram constructed with Riskscore, age, tumor grade and T stage. (B) Adjusted 1-, 3- and 5 years overall survival based on nomogram data.

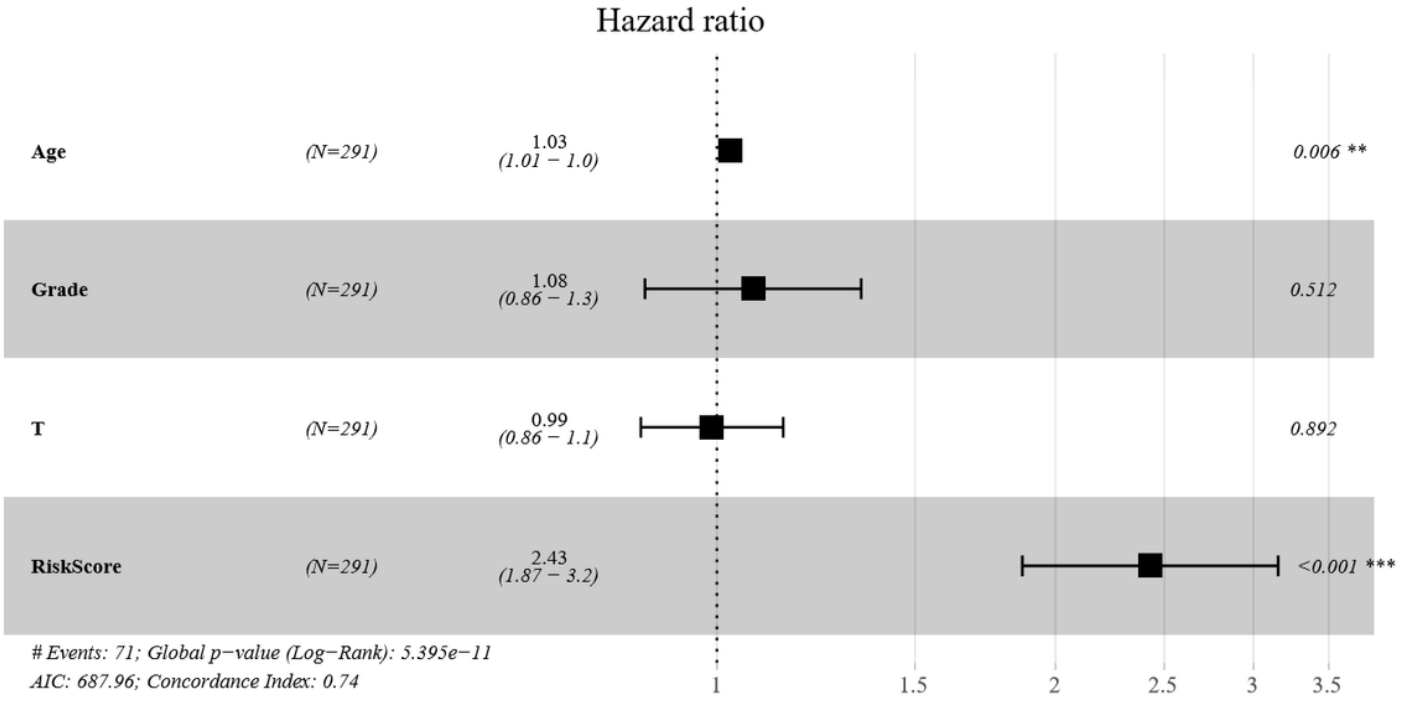

Figure 9

Forest plot constructed with Riskscore, age, tumor grade and T stage

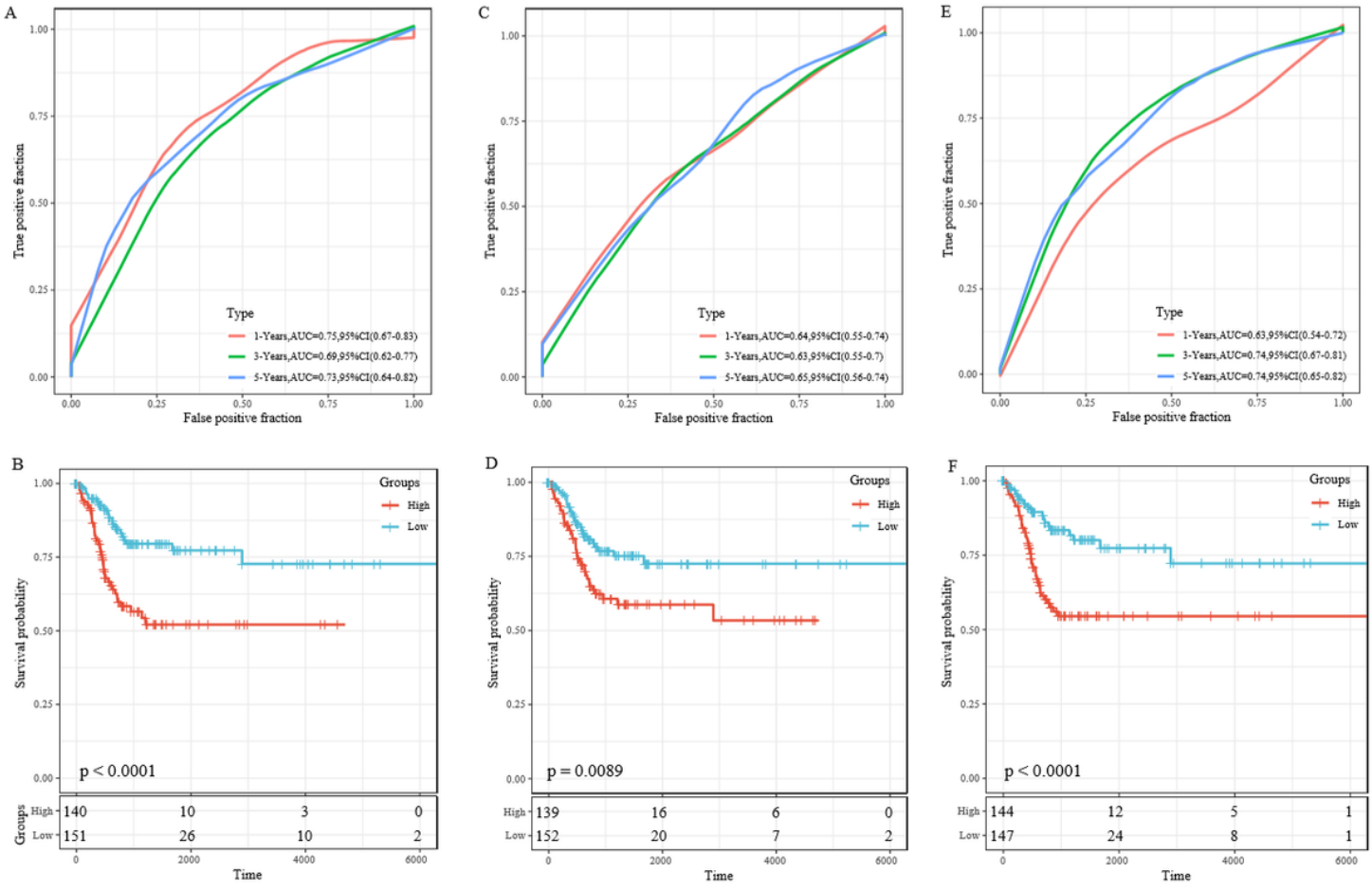

Figure 10

The ROC and overall survival KM curves for those three models The ROC and prognosis KM curves of 8-gene signature (A, B, PMID: 31432147), 5-gene signature (C, D, PMID: 31041817) and 3-gene signature (E, F, PMID: 31612001) in predicting the Risk-H and Risk-L groups on the TCGA dataset.

Supplementary Files

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

- Fig.S1.pdf
- Fig.S2.pdf
- Fig.S3.pdf
- Fig.S4.pdf
- Fig.S5.pdf
- Fig.S6.pdf
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- BMCBioinformaticsTable1.pdf
- BMCBioinformaticsTable2.pdf
- BMCBioinformaticsTable3.pdf