A novel four-gene prognostic signature as a risk biomarker in cervical cancer

Jun Wang (wangxy6@sj-hospital.org)
Shengjing Hospital of China Medical University https://orcid.org/0000-0002-8886-7615
Hua Zheng
The Affiliated Benxi Jinshan Hospital of Dalian Medical University
Yatian Han
Benxi Central Hospital of China Medical University
Geng Wang
Benxi Central Hospital of China Medical University
Yanbin Li
Benxi Central Hospital of China Medical University

Research

Keywords: Cervical cancer, GEO, TCGA, risk score, nomogram

Posted Date: August 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-56408/v1

Abstract

Background: Cervical cancer (CC) is a major malignancy affecting women worldwide, with limited treatment options for patients with advanced disease. The aim of this study was to identify novel prognostic biomarkers for CC by a bioinformatics-based analysis using the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA)-CC cohort.

Methods: RNA-Seq data from four GEO datasets (GSE5787, GSE6791, GSE26511, and GSE63514) were used to identify differentially expressed genes (DEGs) between CC and normal cervical tissues. Functional and enrichment analyses of the DEGs were performed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database and the Database for Annotation, Visualization and Integrated Discovery (DAVID). The Oncomine database, Cytoscape software, and Kaplan–Meier survival analysis were used for in-depth screening for hub DEGs. Cox regression was then used to develop prognostic signature, which was in turn used to create a nomogram.

Results: A total of 207 DEGs were identified in the tissue samples, eight of which were prognostically significant in terms of overall survival (OS). Thereafter, a novel four-gene signature consisting of DSG2, MMP1, SPP1, and MCM2 was developed and validated using stepwise Cox analysis. The area under the receiver operating characteristic (ROC) curve (AUC) values of 0.785, 0.609, and 0.686 in the training, verification, and combination groups, respectively. Moreover, the nomogram analysis showed that a combination of this four-gene signature plus lymph node metastasis (LNM) status effectively predicted the 1- and 3-year OS probabilities of CC patients with accuracies of 69.01% and 83.93%, respectively.

Conclusions: We developed a four-gene signature that can accurately predict the prognosis, in terms of OS, of CC patients, and could be a valuable tool for designing treatment strategies.

Introduction

Cervical cancer (CC) is one of the most common malignancies and a major cause of cancer-related death among women globally [1]. Recently, the incidence of CC has gradually increased, particularly among younger women (35–39 years old) [2]. In > 95% of cases, CC is closely related to the presence of persistent high-risk types of human papillomavirus (HPV) [3], Although the HPV vaccine is effective for the prevention of CC, it does not cover all pathogens associated with CC and is not universally available to women, especially those in low- and middle-income countries where there is a high incidence of and mortality due to CC [4,5]. Therefore, informative biomarkers are needed for CC diagnosis and prognosis prediction.

High-throughput sequencing is an effective method that can be used to screen biomarkers for cancers. With advances in microarrays, small changes at the level of transcription in addition to dysregulation of post-transcriptional signaling in CC can be detected [6]. For instance, Yan et al. used cDNA microarray analysis to show that CXCL8 is overexpressed in cervical cancer tissues relative to tissues from cervical intraepithelial lesions [7]. And Emmanouil et al. found that Minichromosome maintenance protein 2

(MCM2) could significantly improve the sensitivity and specificity of the diagnosis of cervical lesions linked to HPV infection [8]. However, few studies have identified prognostic and predictive signatures by combining with multi genes, and thus a comprehensive analysis for the identification of a robust signature for CC is still needed.

To explore potential biomarkers of poor overall survival (OS) among CC patients in greater detail, we used four Gene Expression Omnibus (GEO) datasets (GSE5787, GSE6791, GSE26511 and GSE63514) to improve the accuracy of the results. By screening and conducting validation based on the Database for Annotation, Visualization and Integrated Discovery (DAVID) database, Oncomine database, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database, and Molecular Complex Detection (MCODE) plug-in in Cytoscape, 40 hub differentially expressed genes (DEGs) were identified between CC and normal cervical tissues. Thereafter, mRNA expression data on the hub DEGs in CC patients (who had corresponding clinical data) in The Cancer Genome Atlas (TCGA)-CC cohort were used together with a stepwise Cox regression analysis to develop a robust four-gene prognostic signature. This signature involved Desmoglein 2 (DSG2), Matrix Metallopeptidase 1 (MMP1), Secreted Phosphoprotein 1 (SPP1), and MCM2. Nomogram analysis suggested that this four-gene signature and lymph node metastasis (LNM) status could accurately predict 1- and 3-year overall survival (OS) among CC patients. In summary, the novel four-gene signature not only about CC pathogenesis but also represents a new method for prognostic evaluation of this type of cancer.

Materials And Methods

Study design

We collected and collated messenger RNA (mRNA) expression datasets based on the Affymetrix Human Genome U133 Plus 2.0 Array platform in GEO (https://www.ncbi.nlm.nih.gov/gds/). Probe information for the microarrays was read and normalized using the “affy” package in R software. The batch effects in the microarray experiments were removed using the “sva” package [9]. Principal component analysis (PCA) was used to assess whether the samples in each group (CC tissues [n = 98] and normal cervical tissues [n = 32]) were clustered, prior to using the samples to identify DEGs. Thereafter, both mRNA expression data and corresponding clinical data for patients with CC (n = 304) were obtained from the TCGA database (https://cancergenome.nih.gov) for additional analyses.

Identification of DEGs

The differential expression matrix of the GEO samples included in the analysis was extracted from the total gene expression matrix, with DEGs between the CC and normal cervical tissues being identified using the “limma” package. Genes with |log2(fold change)| > 1.5 and $p < 0.05$ were considered to be potentially relevant DEGs and were subjected to further analysis.

Functional and pathway

To analyze the DEGs in terms of functional and pathway enrichment, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed using the DAVID database (https://david.ncifcrf.gov). The results were visualized using the "GOplot" package, and the terms were sorted by *p*-value.

Protein–Protein interaction (PPI) network construction

The STRING database (https://string-db.org/) is an online search tool that is frequently used to identify regulatory hub genes. Cytoscape (version 3.6.1) allows visualization and analysis of PPI networks based on the STRING database. We identified candidate hub DEGs using the Cytoscape plug-in MCODE with degree cutoff = 2, node density cutoff = 0.1, node score cutoff = 0.2, and k-core = 2.

Expression validation and survival analysis of the individual hub DEGs

To validate the candidate hub DEGs, the mRNA expression of these DEGs was validated using the Oncomine database (https://www.oncomine.org/), employing a threshold *p*-value of 1×10^{-4} and a fold-change of 2 in five Oncomine microarray datasets. Additionally, the TCGA samples (n = 304) were divided into high- and low-expression groups using the median expression level of each individual candidate hub DEGs as the cutoff value, and Kaplan–Meier survival analysis was then performed for the high- and low-expression groups using the "survival" package in R software.

Cox proportional hazards regression model and risk score

The TCGA-CC samples (n = 304) were randomly divided into a training group (n = 152) and a verification group (n = 152). We then carried out a stepwise Cox regression analysis to identify significant hub DEGs, establish the best model based on the Akaike information criterion (AIC), with the verification and combination groups used for validation. Based on the best model, we then calculated the risk scores for predicting poor OS among CC patients using the following formula (which was used to create the four-gene prognostic signature):

$$Risk\ score = \sum_{n=1}^{\infty}(e_{\mathrm{n}} * \beta_{\mathrm{n}})$$

where β is the estimated stepwise Cox regression coefficient of the mRNA and e is the mRNA expression level.

Based on the median risk score, CC patients were categorized into high- and low-score (risk) groups. A receiver operating characteristic (ROC) curve analysis was used to assess the prognostic performance of the four-gene prognostic signature. Kaplan–Meier survival curves and the log-rank test were used to determine associations between the risk score and OS among patients with CC.

Independence of final signature from conventional clinical feature

Using CC patients (n = 304) in the TCGA-CC cohort with survival status information and detailed clinicopathological information, comprising age, LNM status, Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) stage, and tumor grade, univariate and multivariate Cox regression analyses were conducted to identify whether the four-gene prognostic signature was independent of conventional clinical characteristics.

Analysis of nomogram predicting OS

On the basis of the independent prognostic factors identified in the final multivariate Cox regression analysis, nomograms were used to predict OS (n = 142, 56, and 29 for the 1-, 3-, and 5-year analyses, respectively) among CC patients in the TCGA-CC cohort. The nomograms were visually assessed using calibration plots comparing the predicted and actual survival probabilities among CC patients. The prognostic performance of the nomogram was determined based on the area under the ROC curve (AUC), which can range from 0.5 (no discrimination) to 1 (perfect discrimination).

Statistical analysis

R software (version 3.5.3) with R Studio (version 1.1.463) and the Perl scripting tool (version 5.26.3) were used for data analysis. Differences in clinical features (age, race, tumor grade, FIGO stage, and LNM status) between the training group (n = 152) and verification group (n = 152; formed the 304 TCGA-CC samples) were assessed using the χ^2 test or Fisher's exact test. Two-tailed Student's unpaired t tests were used to compare mean risk scores between pairs of subgroups stratified by clinical features (age, tumor grade, FIGO, and LNM status). The risk scores are presented as the mean ± standard error of the mean (SEM), and a violin plot was used to show the differential distribution of the risk scores in each of the above mentioned subgroups. The optimal cutoff age of the TCGA-CC patients was determined based on survival status using X-tile software (version 3.6.1) [10]. Kaplan–Meier survival analysis was used to assess the association between median risk score and survival (including in the two age subgroups), with the results presented using survival curves and significant differences being determined using the log-rank test. Significance was defined as $p < 0.05$.

Results

Screening for DEGs

The study was constructed as shown in the flow chart in Figure 1. To identify prognostic genes that play a role in CC pathogenesis, we used CC tissues and normal cervical tissues (Table S1). We stabilized the error rate estimates and improved the reproducibility of the gene expression matrix using surrogate variables for removing batch effects (Fig. S1). PCA showed that the two groups of samples were obviously clustered (Fig. 2A). A total of 207 DEGs between the two groups were observed, with 106 being up regulated and 101 being down regulated (Fig. 2B, Table S2).

Functional and pathway enrichment analyses of DEGs

GO and KEGG analyses were performed using the DAVID database for annotation. Regarding GO terms, the DEGs were primarily enriched in extracellular exosome (58 proteins), serine-type endopeptidase activity (16 proteins), and peptide cross-linking (11 proteins) (Fig. 2C). KEGG analysis revealed that the DEGs were enriched in several pathways such as cytokine–cytokine receptor interaction (10 proteins), chemokine signaling (8 proteins), and tumor necrosis factor (TNF) signaling (6 proteins), as well as transcriptional dysregulation in various cancers such as bladder cancer (4 proteins) (Fig. 2D).

Identification of hub DEGs in CC

Although enrichment analyses reveal the biological processes and pathways related to DEGs, they do not provide information about interactions among the DEGs. Thus, we examined the interactions among the proteins using STRING, and visualized the PPI network using Cytoscape software. To construct the PPI network, 112 significantly enriched DEGs were submitted to STRING (Fig. 3A), and the PPI network was subsequently imported into Cytoscape to construct the sub-networks. Using the MCODE plug-in in Cytoscape, we analyzed the top three sub-modules (MCODE scores ≥ 10) of proteins to identify the hub DEGs. There were 40 hub DEGs in these modules, with those in Modules 1 and 2 primarily being up-regulated DEGs, and those in Module 3 primarily being down-regulated DEGs (Fig. 3B-D). The Oncomine co-expression analysis showed that the mRNA expression levels of 22 of the candidate hub DEGs were consistent with our initial analyses (Fig. 4A and Fig. S2).

To examine the hub DEGs in greater detail, TCGA CC samples (n = 304) were used for survival analyses of the individual hub DEGs (with a cutoff value of $p < 0.05$). The expression of eight hub DEGs (CXCL1, CXCL8, DSG2, MMP1, SPP1, MCM2, Lymphoid-specific helicase [HELLS], and Vascular cell adhesion molecule 1 [VCAM1]) was significantly associated with OS among CC patients. Interestingly, MCM2, HELLS, and VCAM1 up-regulation played protective roles (Fig. 4B and Fig. S3).

Cox proportional hazards model and risk score

The TCGA-CC samples (n = 304) were randomly divided into a training group (n = 152) and a verification group (n = 152). Among the 304 patients, 223 (73.4%) and 87 (28.3%) had complete follow-up data of clinical features for at least 1- and 3-years, respectively, but only 40 (13.2%) had detailed follow-up data for ≥ 5 years. There were no significant differences in age, race, tumor grade, FIGO stage, or LNM status between the training and verification groups (Table 1). Next, we assessed the significance of the eight above mentioned hub DEGs in a Cox proportional hazards model and consequently developed a novel four-gene prognostic signature. This signature allowed us to determine the high- and low-risk patients, as follows:

Riskscore = (0.58 * expression value of DSG2) + (0.27 * expression value of MMP1) +

(0.33 * expression value of SPP1) + (-0.48 * expression value of MCM2)

With validation using the verification and combination groups, we built the best fitting Cox proportional hazards model using a combination of four high-power prognostic genes (DSG2, MMP1, SPP1, and MCM2) (Fig. 5A). The ROC curves showed that this four-gene signature achieved an AUC value of 0.785 (95% CI: 0.670–0.879), 0.609 (95% CI: 0.507–0.711), and 0.686 (95% CI: 0.612–0.761) for the training, verification, and combination groups, respectively (Fig. 5B). These outcomes suggest that this four-gene signature demonstrates good performance regarding predicting OS among CC patients (Fig. 5C).

OS prediction and evaluation

To further evaluate whether the four-gene prognostic signature can serve as a prognostic factor, we performed univariate and multivariate Cox regression analyses comparing high- and low-risk CC patients. Covariates besides the risk score included clinical risk factors such as age, tumor grade, FIGO stage, and LNM status (Fig. S4). The univariate Cox regression analysis showed that risk score (hazard ratio [HR]: 3.186; 95% CI: 1.513–6.711; $p = 0.003$) and LNM status (HR: 2.886; 95% CI: 1.435–5.803; $p = 0.003$) were risk factors, while the multivariate Cox regression analysis confirmed that the risk score (HR: 2.743; 95% CI: 1.285–5.856; $p = 0.009$) and LNM status (HR: 2.660; 95% CI: 1.290–5.489; $p = 0.008$) were both independent risk factors (Table 2).

The risk score was then compared between the pairs of subgroups stratified by clinical features to explore whether it was significantly different between the various subgroups. It was only significantly different between LNM-negative and LNM-positive patients, being higher in the latter (1.056±0.053 vs 1.341±0.138, $p = 0.019$) (Fig. 6).

We also constructed a nomogram to predict 1- and 3-year OS for CC patients using the four-gene signature and LNM status (Fig. 7A). The calibration plots showed good agreement between the predicted and actual probabilities regarding 1- and 3-year but not 5-year OS based on the TCGA-CC cohort (Fig. 7B). The resulting AUC values regarding 1- and 3-year OS were 0.746 (95% CI: 0.635–0.857, n = 142) and 0.748 (95% CI: 0.551–0.944, n = 56), respectively, and the prognostic accuracy values were 69.01% and 83.93%, respectively (Fig. 7C).

Discussion

CC is a malignant disease that is the fourth most frequent cancer in the world, with 569,847 new cases and 311,365 deaths in 2018 [11]. When detected early, CC is highly treatable, and these patients have high survival rates and good quality of life. During tumorigenes is as well as during cancer development, mRNA expression levels can exhibit minor changes. During CC progression, multiple mRNAs have been shown to be dysregulated [12,13], although the prognostic values of multi-mRNA signatures based on samples from CC patients has remained unclear. In the present study, we developed a novel four-gene signature and validated it as a biomarker for early diagnosis and predicting 1- and 3-year OS among CC patients. This four-gene signature might constitute an important step forward for treatment decisions and for predicting more accurate and individualized prognoses for CC patients. This four-gene signature also provides a basis for future experimental research.

Several studies have examined the potential of multi-mRNA signatures for clinical research on CC. Huang et al. provided a prediction model of CC recurrence based on the expression patterns of seven genes [14], while Ding et al. developed a prediction model of CC survival based on the expression of three genes [15]. Although they both used a high-throughput molecular identification method, both studies considered only a single dataset. In the present study, we integrated multiple bioinformatics tools and databases to improve the accuracy of the results. The recent change to FIGO staging of CC cases reflects the importance of LNM status, and several reports have demonstrated that positive pathologic LNM is more strongly associated with survival rate than other risk factors such as age, histology, and clinical stage [16,17], though LNM status alone may not predict CC prognosis. In the present study, we demonstrated that the four-gene signature and LNM status both had prognostic value for CC, and we developed a nomogram that integrated the four-gene prognostic signature and LNM status to accurately predict the 1- and 3-year OS rates of patients with CC. According to the four-gene prognostic signature, DSG2, MMP1, and SPP1 are risk factors, whereas MCM2 is a protective factor for patients with CC.

DSG2 is a member of the desmoglein family and the cadherin cell adhesion molecule superfamily. Although its precise role in CC is unclear, it is thought to be involved in the development of several types of cancers [18,19]. It is related to keratinization, developmental biology, and mitogen-activated protein kinase (MAPK) signaling pathways. Our analysis revealed that patients with higher DSG2 expression had poorer prognosis, suggesting that it may play a role in predicting prognosis in terms of poor OS among CC patients.

MMP1, also known as interstitial collagenase, is located on chromosome 11q22.3 and belongs to the matrix metalloproteinase family. It can promote tumor invasion and metastasis through mechanisms involving angiogenesis and immune evasion [20]. Overexpression of MMP1 is strongly associated with unfavorable prognosis in multiple malignancies including breast cancer, esophageal squamous cell carcinoma, and ovarian cancer [21-23]. MMP1 has also previously been proposed as a risk factor in CC [24,25].

SPP1 participates in the regulation of tumor-associated angiogenesis and inflammation [26]. Previous bioinformatics analysis showed that SPP1 is closely related to the incidence and poor prognosis of CC [27], which is consistent with our findings. Moreover, SPP1 down regulation improves the cisplatin sensitivity of HeLa cells by inhibiting the activity of the phosphoinositide 3-kinase (PI3K)/Akt signaling pathway [28].

MCM2 is a component of the DNA replication licensing complex (MCM2-7) that has been found to mainly localize to the nucleus in eukaryotic cells [29]. Overexpression of MCM2 frequently occurs in CC, particularly in cases involving persistent infection with high-risk HPV [8]. MCM2 has been reported to promote tumor proliferation by mediating DNA replication initiation and elongation [30]. In contrast, in our study, MCM2 played a protective role regarding CC progression, although, consistent with previous studies [30], we also observed high expression levels of MCM2 in the CC tissues. Aihemaiti et al. reported that cytoplasmic rather than nuclear accumulation of MCM2 is related to improved survival for patients

with ovarian clear cell carcinoma [31], which may be associated with MCM2-mediated DNA damage-induced apoptosis [32,33]. This pathway may also function in CC, although additional investigation is needed to explore this possibility.

There are some limitations to this study: (i) the sample size was small; (ii) the patients were largely European and American and included few Asian patients, although we are currently collecting tissues from patients treated at the Obstetrics and Gynecology Department of Benxi Central Hospital in China for further analysis; and (iii) additional investigation based on different histological types is needed both to define the detailed mechanisms of the hub DEGs (particularly DSG2 and MCM2) in CC pathogenesis, and to validate the relationship of the four-gene signature with CC prognosis in a larger cohort.

Conclusion

In summary, the mRNA expression levels of four hub DEGs (DSG2, MMP1, SPP1, and MCM2) were significantly associated with OS among CC patients and the novel four-gene signature could have substantial prognostic value, allowing prediction of OS among patients with CC. The efficacy of the four-gene signature for patients with CC is promising and warrants additional investigation.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Publicly available datasets were analyzed in this study, which can be found here:

1. https://www.ncbi.nlm.nih.gov/gds/?term=
2. https://portal.gdc.cancer.gov/

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by grants from the Scientific Fund of the Central Hospital of Benxi (no. 201901).

Authors' contributions

Jun Wang conceived and designed the experiments. Hua Zheng, Yatian Han, Geng Wang, and Yanbin Li analyzed the data. All authors were involved in critically revising the manuscript.

Acknowledgements

We thank the Charlesworth Company for language editing.

Abbreviations

CC: Cervical cancer; GEO: Gene Expression Omnibus datasets; TCGA: The Cancer Genome Atlas database; STRING: Search Tool for the Retrieval of Interacting Genes/Proteins database; DAVID: Database for Annotation, Visualization and Integrated Discovery; PPI: Protein–protein interaction DEGs: Differentially expressed genes; PCA: principal component analysis; FIGO: Fédération Internationale de Gynécologie et d'Obstétrique; OS: Overall survival; ROC: Receiver operating characteristic; AUC: Area under the curve; LNM: lymph node metastasis; CI: confidence interval; DSG2: Desmoglein 2; MMP1: Matrix Metallopeptidase 1; SPP1: Secreted Phosphoprotein 1; MCM2: Minichromosome maintenance protein 2.

References

1. Carla J, Jeffrey S. Cervical Cancer as a Global Concern Contributions of the Dual Epidemics of HPV and HIV. JAMA. 2019;322(16):1558-1560.
2. Pimple S, Mishra S, Shastri S. Global strategies for cervical cancer prevention. Curr Opin Obstet Gynecol. 2016;28(1):4-10.
3. Cardin LT, Prates J, Cunha BR, Tajara EH, Oliani SM. Annexin A1 peptide and endothelial cell-conditioned medium modulate cervical tumorigenesis. FEBS Open Bio. 2019; 9(4):668-681.
4. Beddoe AM. Elimination of cervical cancer: challenges for developing countries. eCancerMedicalScience. 2019;13:975.
5. Zou FW, Tang YF, Liu CY, Ma JA, Hu CH. Concordance Study Between IBM Watson for Oncology and Real Clinical Practice for Cervical Cancer Patients in China: A Retrospective Analysis. Front Genet. 2020;11:200.
6. Barbara P, Daniela DM, Antonio F, Cornelia DG, Guglielmo R, Alessio N. MicroRNAs as markers of progression in cervical cancer: a systematic review. BMC Cancer. 2018;18(1):696.
7. Yan R, Shuai H, Luo X, Wang X, Guan B. The clinical and prognostic value of CXCL8 in cervical carcinoma patients: immunohistochemical analysis. BiosciRep. 2017;37(5):BSR20171021.
8. Emmanouil P, Andrew VK, Livio A, Nicola MZ, Agnieszka M, Brian NR, Matthew F, Surya V, Doreen RM, Ian S, et al. Gene expression profiling informs HPV cervical histopathology but not recurrence/relapse after LEEP in ART-suppressed HIV+HPV+ women. Carcinogenesis. 2019;40(2): 225-233.

9. Leek JT, Johnson WE, Parker HS, Jaffe AJ, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics (Oxford, England). 2012;28(6):882-883.
10. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res. 2014;10(21):7252-7259.
11. Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre T, Jemal J. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
12. Sun J, Li HL, Ma H, Yang S, Li RY. SMYD2 promotes cervical cancer growth by stimulating cell proliferation. Cell Biosci. 2019;9:75.
13. Zhong Y, Yang J, Xu WW, Wang Y, Zheng CC, Li B, He QY. KCTD12 promotes tumorigenesis by facilitating CDC25B/CDK1/Aurora A-dependent G2/M transition. Oncogene. 2017;36(44):6177-6189.
14. Huang L, Zheng M, Zhou QM, Zhang MY, Yu YH, Yun JP, Wang HY. Identification of a 7-gene signature that predicts relapse and survival for early stage patients with cervical carcinoma. Med Oncol. 2012;29(4):2911-2918.
15. Ding T, Ma H, Feng J. A three-gene novel predictor for improving the prognosis of cervical cancer. Oncol Lett. 2019;18(5):4907-4915.
16. Yan DD, Tang Q, Chen JH, Tu YQ, Lv XJ. Prognostic value of the 2018 FIGO staging system for cervical cancer patients with surgical risk factors. Cancer Manage Res. 2019;11:5473-5480.
17. Jeong SY, Park H, Kim MS, Kang JH, Paik ES, Lee YY, Kim TJ, Lee JW, Kim BG, Duk SB, et al. Pretreatment Lymph Node Metastasis as a Prognostic Significance in Cervical Cancer: Comparison between Disease Status. Cancer Res Treat. 2020;52(2): 516-523.
18. Katharina H, Julian Z, Lena S, Steffen O, Volker S. Loss of desmoglein 2 promotes tumorigenic behavior in pancreatic cancer cells. Mol Carcinog. 2017;56(8):1884-1895.
19. Runsen J, Wang XF, Zang RH, Liu CM, Zheng SF, Li HC, Sun N, He J. Desmoglein-2 modulates tumor progression and osimertinib drug resistance through the EGFR/Src/PAK1 pathway in lung adenocarcinoma. Cancer Lett. 2020;483:46-58.
20. Winer A, Adams A, Mignatti P. Matrix Metalloproteinase Inhibitors in Cancer Therapy: Turning Past Failures Into Future Successes. Mol Cancer Ther. 2018;17(6):1147-1155.
21. Vizoso FJ, Gonzalez LO, Corte MD, Rodríguez JC, Vázquez J, Lamelas ML, Junquera S, Merino AM, García-Muñiz JL. Study of matrix metalloproteinases and their inhibitors in breast cancer. Br J Cancer. 2007;96(6):903-911.
22. Liu M, Hu Y, Zhang MF, Luo KJ, Xie XY, Wen J, Fu JH, Yang H. MMP1 promotes tumor growth and metastasis in esophageal squamous cell carcinoma. Cancer Lett. 2016;377(1):97-104.
23. Akira Y, Yusuke Y, Yusuke Y, Mitsuya I, Shun-Ichi I, Tomoyasu K, Tohru K, Fumitaka T, Hiroaki K, Fumitaka K, et al. Malignant extracellular vesicles carrying MMP1 mRNA facilitate peritoneal dissemination in ovarian cancer. Nat Commun. 2017;8:14470.

24. Solovyeva NI, Timoshenko OS, Gureeva TA, Kugaevskaya EV. Matrix metalloproteinases and their endogenous regulators in squamous cervical carcinoma. Biomed Khim. 2015;61(6):694-704.
25. Zhang ZF, Wang LL, Du J, Li YB, Yang HL, Li XC, Li H, Hu HY. Lipid Raft Localization of Epidermal Growth Factor Receptor Alters Matrix metalloproteinase-1 Expression in SiHa Cells via the MAPK/ERK Signaling Pathway. Oncol Lett. 2016;12(6):4991-4998.
26. Neil EH, Qian JC, Laura KS, Meghan BW, Jeffrey PG, Ninnie MA, Jesse AE, Judith EW, Alexander DB. Transgenic mammary epithelial osteopontin (spp1) expression induces proliferation and alveologenesis. Genes Cancer. 2013;4(5-6):201-212.
27. Zhao MH, Huang WB, Zou SW, Shen Q, Zhu XQ. A Five-Genes-Based Prognostic Signature for Cervical Cancer Overall Survival Prediction. Int J Genomics. 2020;20:ID834763.
28. Chen X, Xiong SH, Ye L, Yang HC, Mei SS, Wu JH, Chen SS, Mi RR. SPP1 inhibition improves the cisplatin chemo-sensitivity of cervical cancer cell line. Cancer Chemother Pharmacol. 2019;83(4):603-613.
29. Parker MW, Botchan MR, Berger JM. Mechanisms and regulation of DNA replication initiation in eukaryotes. Crit Rev Biochem Mol Biol. 2017;52(2):107-144.
30. Liu D, Zhang XX, Xi BX, Wan DY, Li L, Zhou J, Wang W, Ma D, Wang H, Gao QL. Sine oculis homeobox homolog 1 promotes DNA replication and cell proliferation in cervical cancer. Int J Oncol. 2014;45(3):1232-1240.
31. Gulinisha A, Morito K, Daichi N, Akiko Y, Tatsunori M, Iichiroh O, Yuko K, Jin XH, Anna T, Naoyuki M, et al. Subcellular localization of MCM2 correlates with the prognosis of ovarian clear cell carcinoma. Oncotarget. 2018;9(46):28213-28225.
32. Maki H, Morito K, Kouhei Y, Kazuko Y, Shirou A, Masanobu K. A novel role for acinus and MCM2 as host-specific signaling enhancers of DNA-damage-induced apoptosis in association with viral protein gp70. Leuk Res. 2009;33(8):1100-1107.
33. Suzuki, Kurata, Shinya, Miyazawa, Murayama, Hidaka, Yamamoto, Kitagawa. Overexpression of MCM2 in myelodysplastic syndromes: association with bone marrow cell apoptosis and peripheral cytopenia. Exp Mol Pathol. 2012;92(1):160-166.

Tables

Table 1. Clinical features of CC patients in the training and verification groups.

Feature	Training group (n = 152)	Verification group (n = 152)	*p*-value
Age			
≤64	129	136	0.304
>64	23	16	
Grade			
G1–G2	79	74	0.834
G3	58	61	
Unknown	15	17	
FIGO stage			
I–II	112	119	0.607
III–IV	36	29	
Unknown	4	4	
LNM status			
Positive	60	73	0.278
Negative	34	26	
Unknown	58	53	
Race			
White	118	121	0.700
Asian	9	11	
Other	4	5	
Not reported	21	15	

***CC**: cervical cancer; **FIGO**: Fédération Internationale de Gynécologie et d'Obstétrique; **LNM**: lymph node metastasis*

Table 2. Univariate/multivariate Cox regression analyses.

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	*p*-value	HR	95% CI	*p*-value
Age ≤ 64 / >64	2.567	0.888–7.419	0.082	1.543	0.514–4.631	0.439
Grade G1–2 / G3	1.022	0.502–2.077	0.953	0.979	0.479–2.002	0.953
FIGO stage I–II / III–IV	0.948	0.288–3.124	0.931	0.764	0.219–2.669	0.673
LNM status Positive / negative	2.886	1.435–5.803	**0.003**	2.660	1.290–5.489	**0.008**
Risk score High risk / low risk	3.186	1.513–6.711	**0.003**	2.743	1.285–5.856	**0.009**

***CI**: confidence interval; **FIGO**: Fédération Internationale de Gynécologie et d'Obstétrique; **HR**: hazard ratio; **LNM**: lymph node metastasis*

Figures

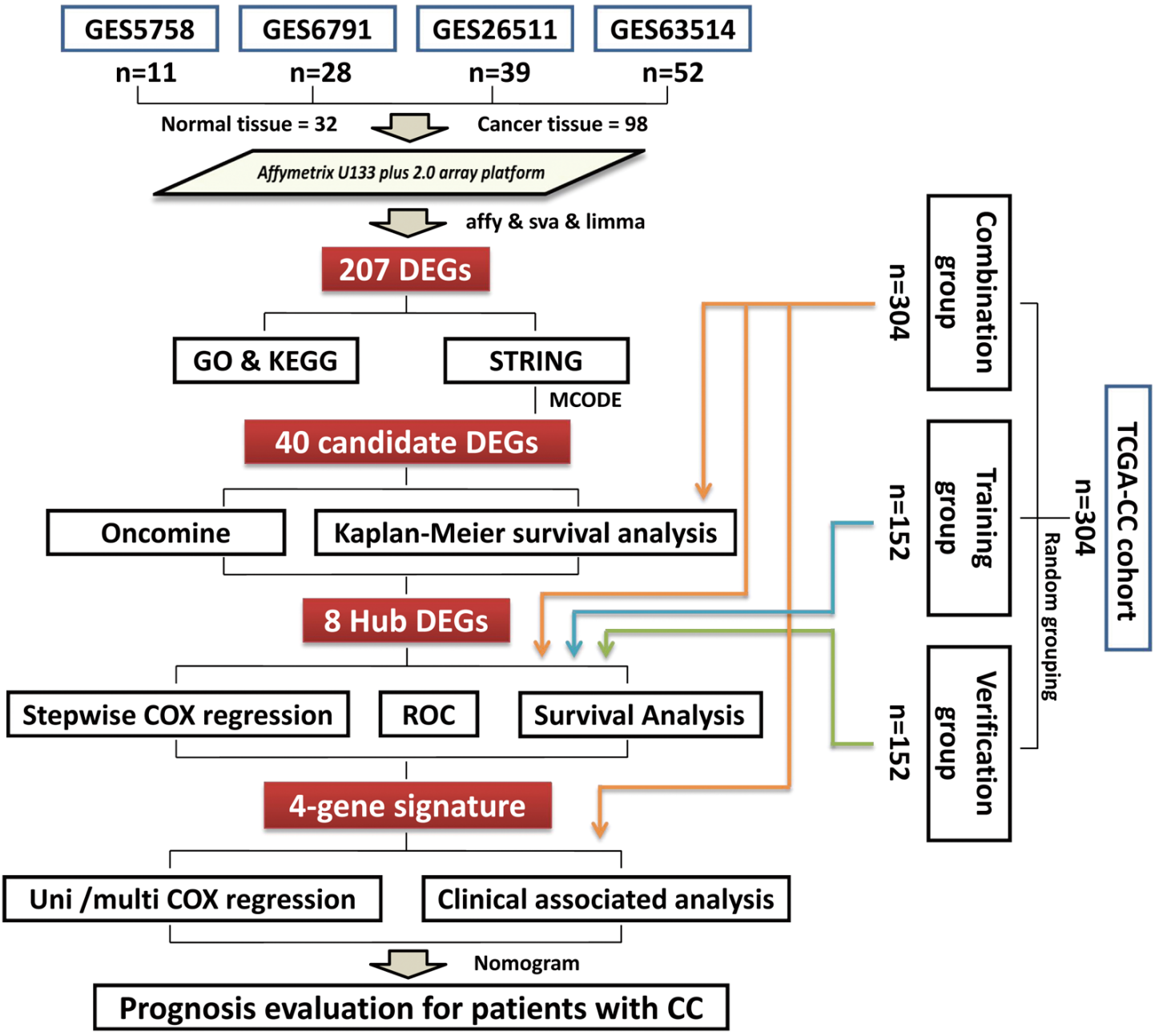

Figure 1

Flow chart of our study on a multi-mRNA prognostic signature for CC. CC: cervical cancer; DEG: differentially expressed gene; ROC: receiver operating characteristic.

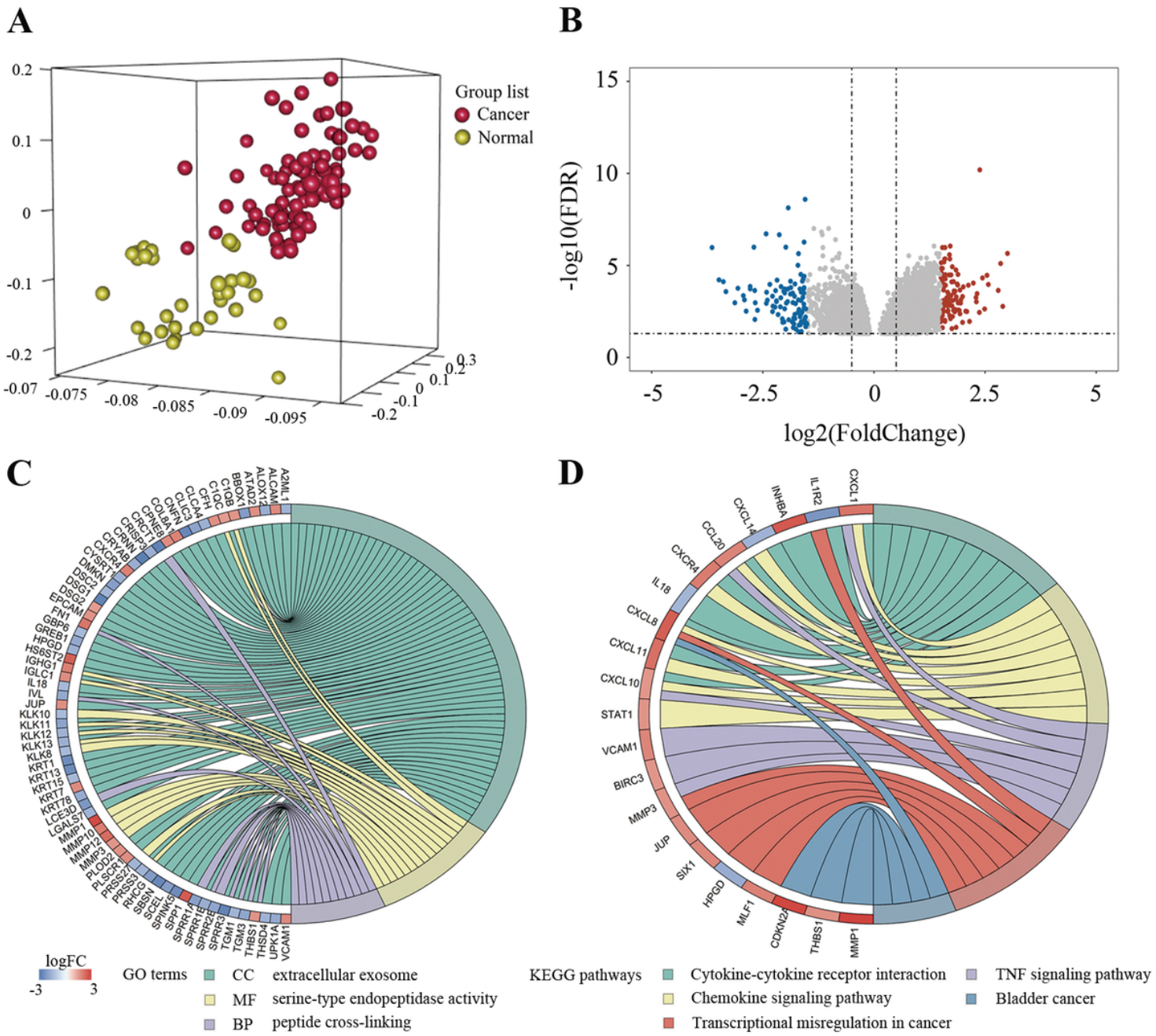

Figure 2

Screening for DEGs in samples from patients with CC. (A) PCA plot showing significant differential clustering between CC (red spheres) and normal cervical (yellow spheres) tissue samples. (B) Volcano plots depicting changes in mRNA expression between normal cervical tissue (n = 32) and CC (n = 98) groups. There were 106 and 101 mRNAs with |log2(fold change)| > 1.5 that were significantly ($p < 0.05$) up-regulated (red) and down-regulated (blue) in the CC group compared to the normal cervical tissue group. (C-D) GO and KEGG enrichment analyses of DEGs in CC samples. CC: cervical cancer; DEG: differentially expressed gene; PCA: principal component analysis; FC: fold change; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.

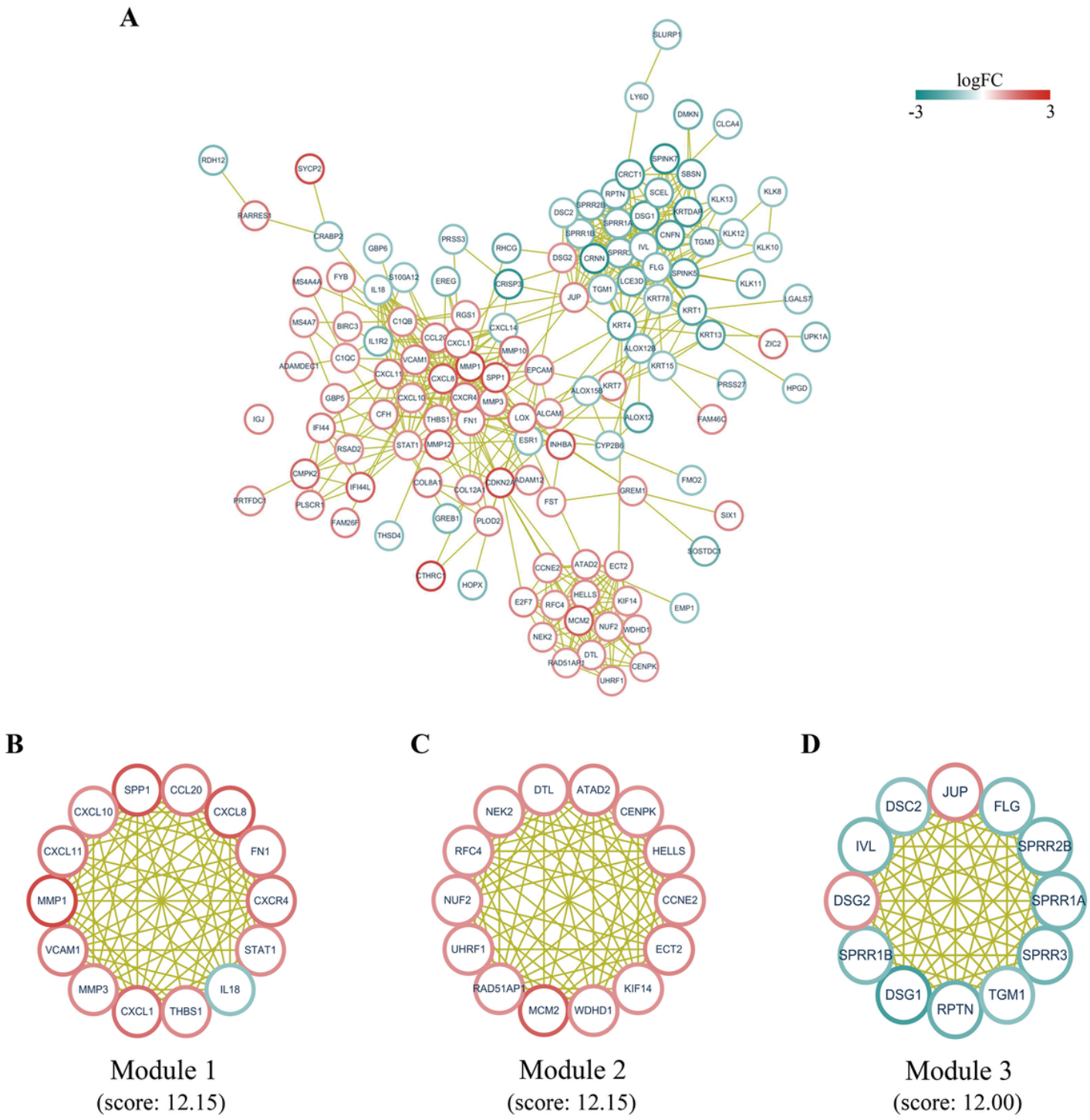

Figure 3

PPI network of DEGs in CC. (A) PPI network demonstrating the interactions between up regulated (red) and down regulated (green) proteins. (B-D) The three most significant sub-modules (MCODE scores ≥ 10) contained 14 (B), 14 (C) and 12 (D) candidate DEGs, respectively. Node color is positively related to fold-change in expression; yellow link indicates specific or meaningful association among nodes. CC: cervical cancer; DEG: differentially expressed gene; PPI: protein–protein interaction.

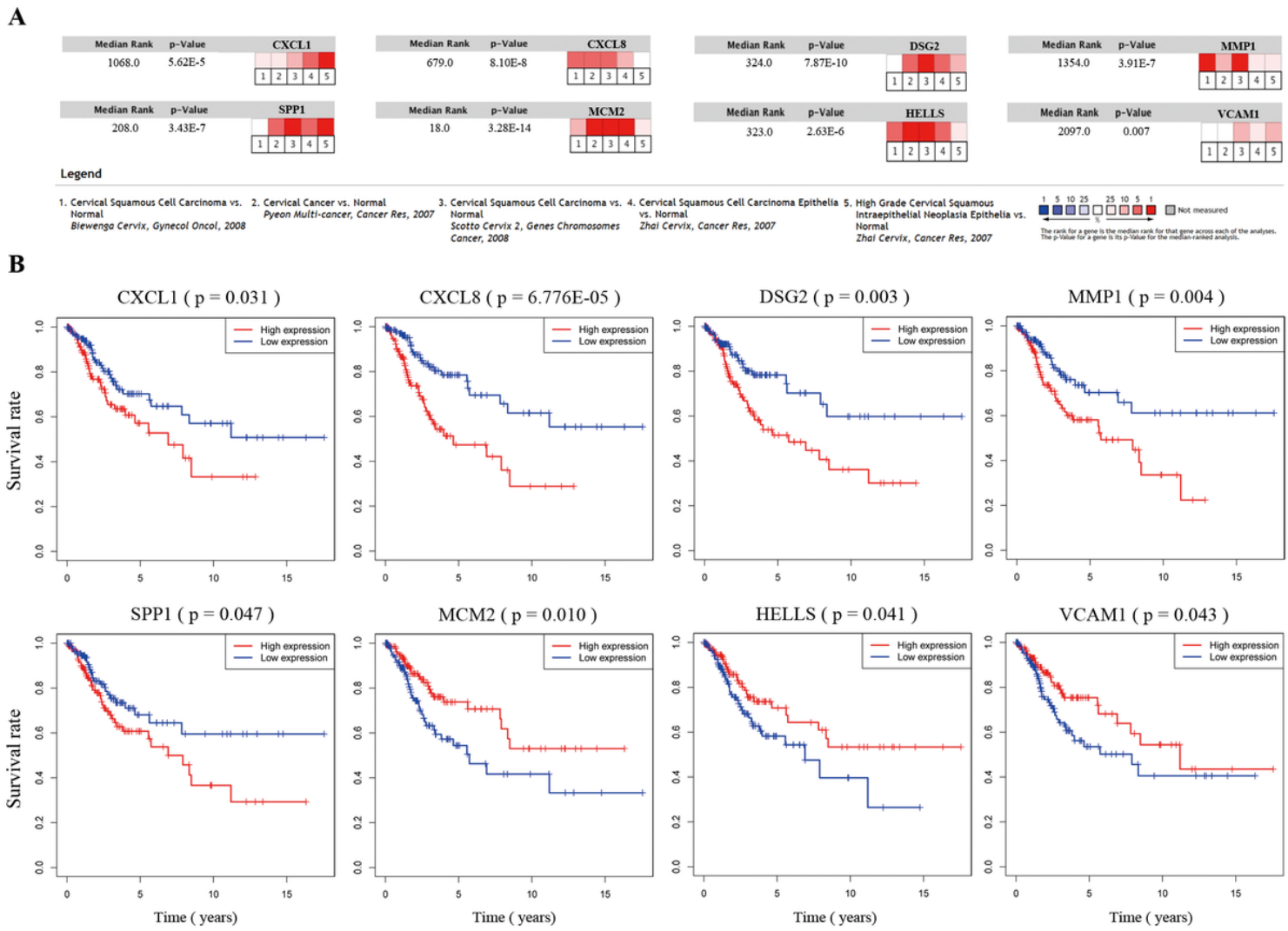

Figure 4

In-depth screening for hub DEGs in CC. (A) mRNA expression levels of candidate hub DEGs evaluated using the Oncomine microarray database. (B) Kaplan–Meier survival analysis and the log-rank test. mRNA levels of CXCL1 (p = 0.031), CXCL8 (p < 0.001), DSG2 (p = 0.003), MMP1 (p = 0.004), SPP1 (p = 0.047), MCM2 (p = 0.010), HELLS (p = 0.041), and VCAM1 (p = 0.043) were significantly associated with OS among CC patients. CC: cervical cancer; DEG: differentially expressed gene; OS: overall survival.

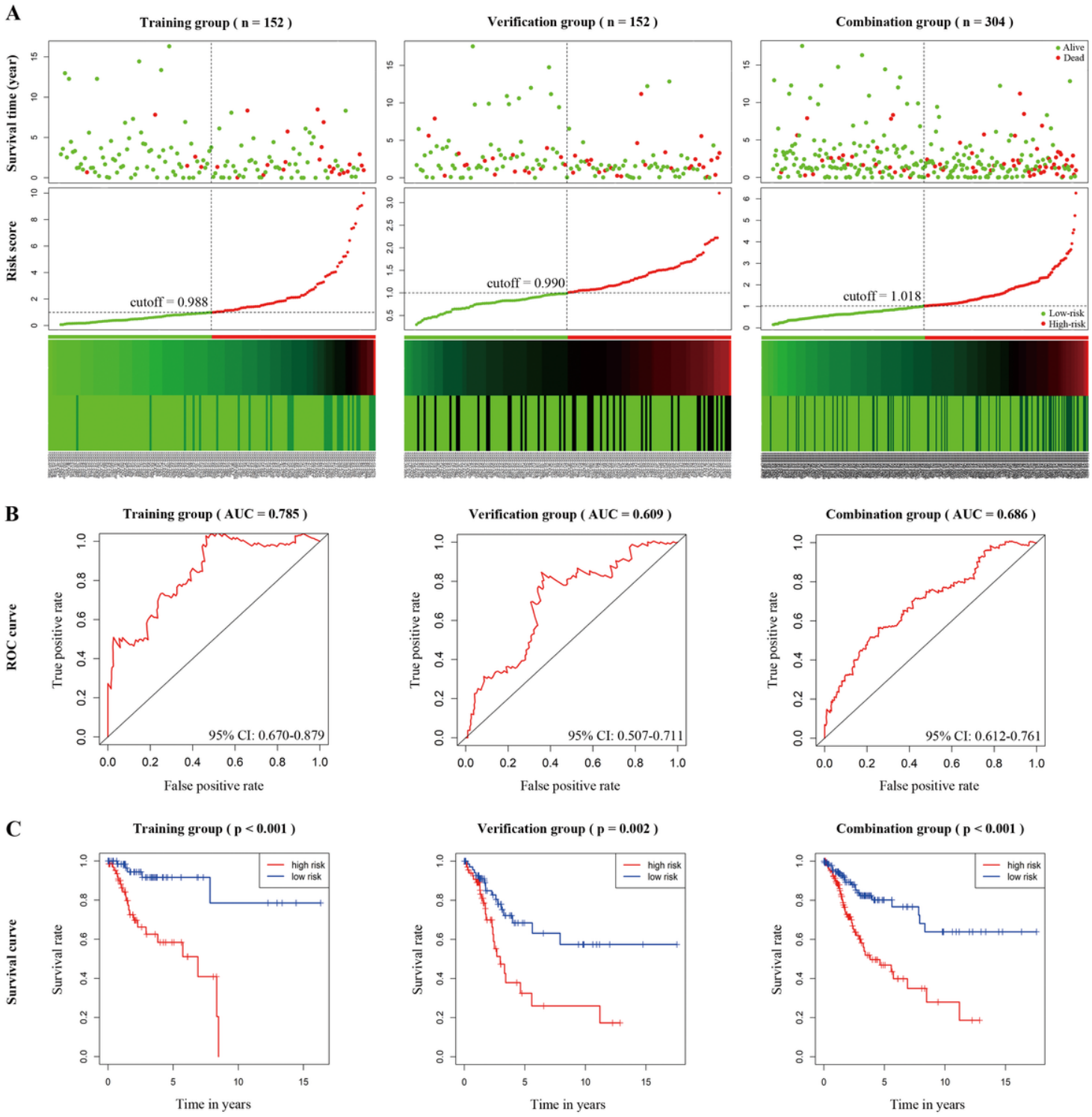

Figure 5

Construction of a four-gene prognostic signature for CC. (A) Survival status of patients with CC and four-gene signature risk score distribution for the training, verification, and combination groups. The black-dotted line represents the median cutoff. (B) AUC of ROC curve evaluating the ability of the four-gene signature to predict prognosis in terms of OS. (C) Kaplan–Meier curves of the four-gene signature indicating that high-risk patients had shorter survival. AUC: area under the curve; CC: cervical cancer; CI: confidence interval; OS: overall survival; ROC: receiver operating characteristic.

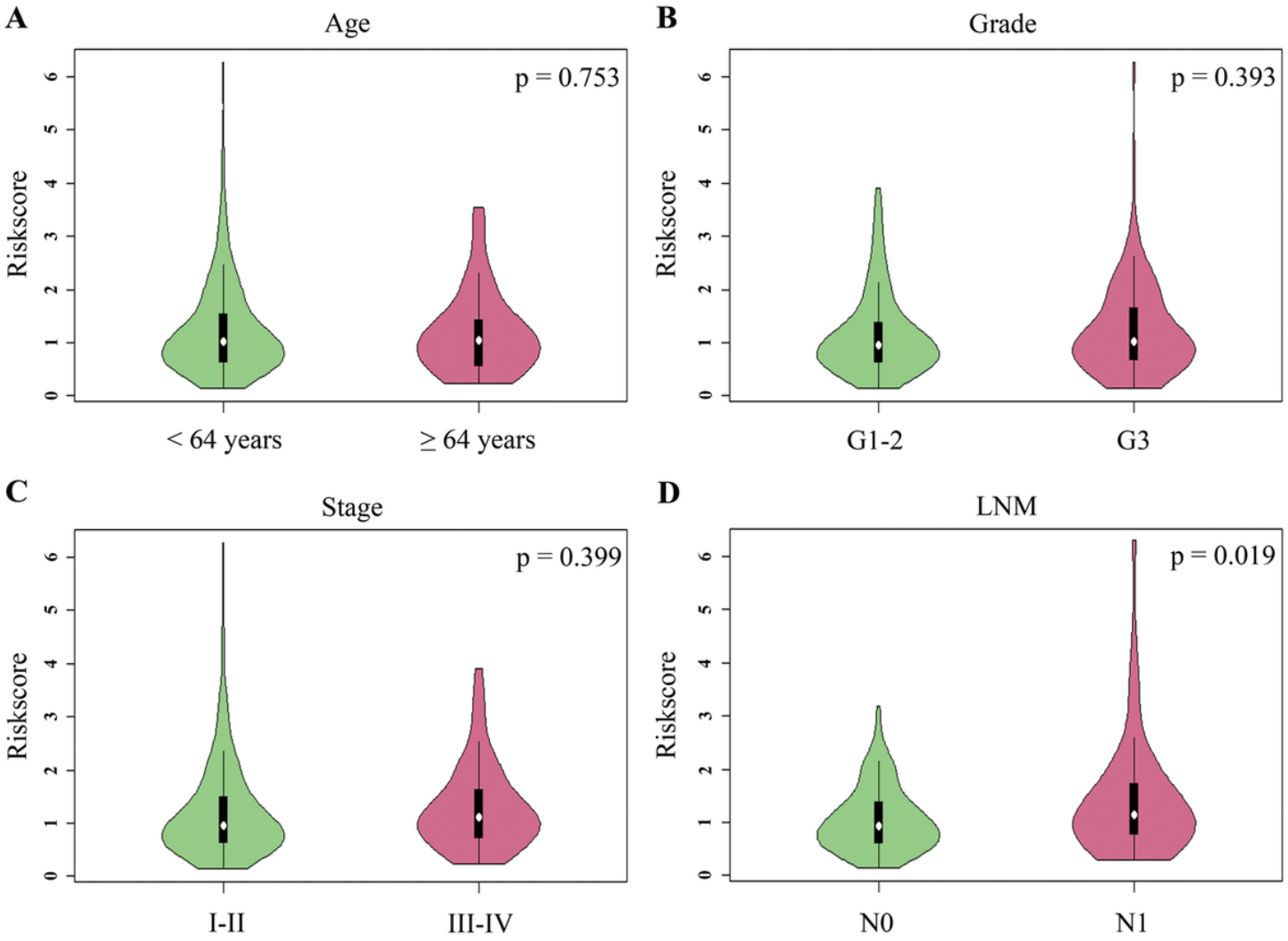

Figure 6

Differences in the risk score among CC patients stratified by clinical features. Violin plots showing differences in the mean risk score in CC patients stratified by (A) age (mean ± SEM: 1.225±0.054 vs 1.183±0.111), (B) tumor grade (1.155±0.062 vs 1.240±0.081), (C) FIGO stage (1.194±0.057 vs 1.295±0.098), and (D) LNM status (1.056±0.053 vs 1.341±0.138) (< 64 years old [n = 256], ≥ 64 years old [n = 48]; grade G1–G2 [n = 153], G3 [n = 119]; stage I–II [n = 189], III–IV [n = 108]; LNM N0 [n = 133], LNM [N1 n= 60]). P < 0.05 represents significance between corresponding subgroups, according to unpaired Student's t test. CC: cervical cancer; FIGO: Fédération Internationale de Gynécologie et d'Obstétrique; LNM: lymph node metastasis (N0 = negative, N1 = positive).

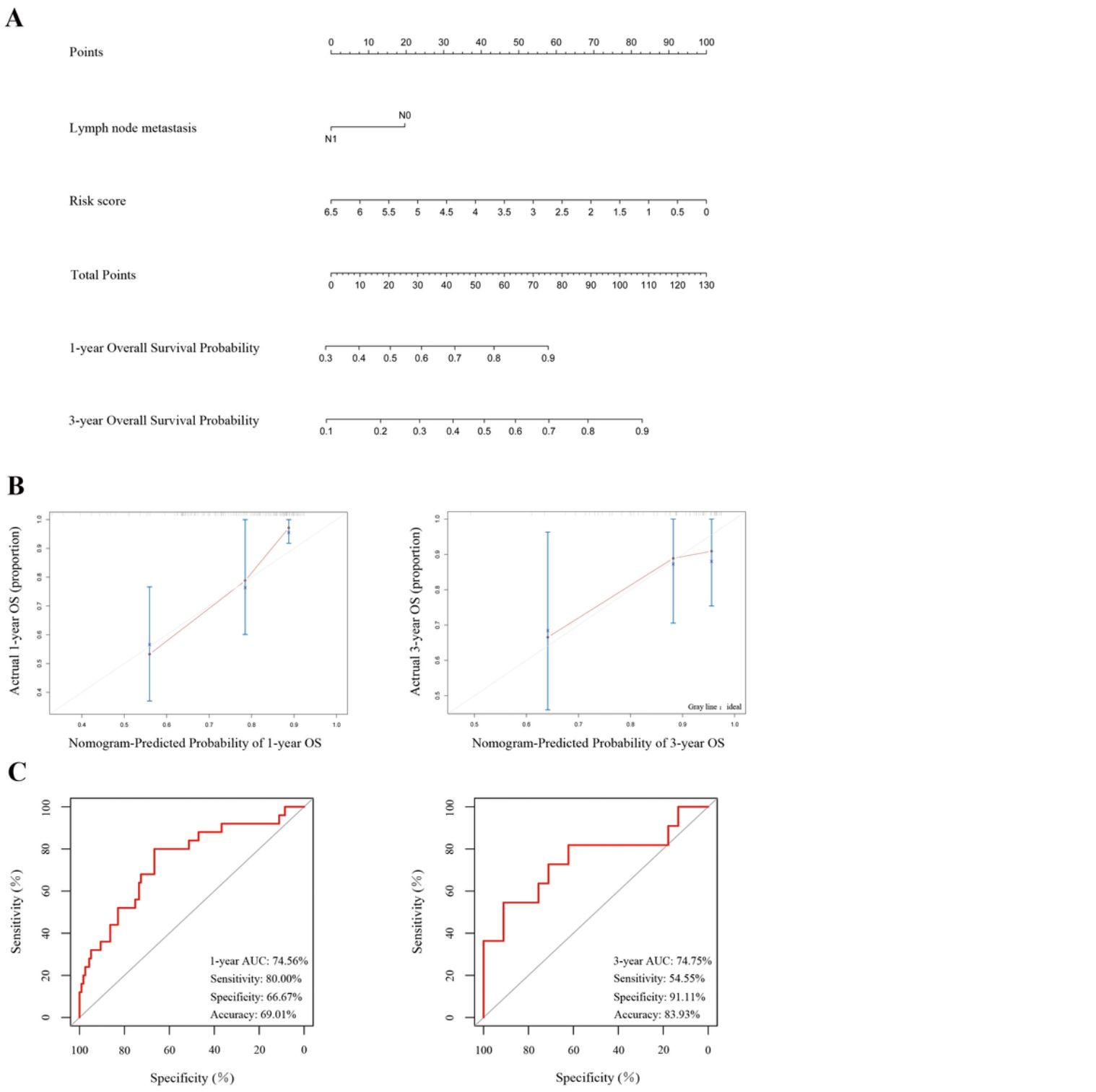

Figure 7

Nomogram analysis predicting OS among patients with CC. (A) Nomogram of two independent risk factors (four-gene signature and LNM status); the scores for each variable were added to obtained the total score for predicting the 1- and 3-year OS of patients with CC. (B) Calibration plots for the nomogram predicting 1-year OS (n = 142) and 3-year OS (n = 56). Reference line represents a perfect match between the predicted and actual survival probabilities. (C) AUC of the ROC curve verifying the prognostic

accuracy of the nomogram for predicting 1- and 3-year OS. AUC: area under the curve; CC: cervical cancer; LNM: lymph node metastasis; OS: overall survival; ROC: receiver operating characteristic.

Supplementary Files

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

- FigureS1.tif
- FigureS2.tif
- FigureS3.tif
- FigureS4.tif
- TableS1.docx
- TableS2.docx