Observational Study

Medicine

OPEN

Identification and verification of a 4-gene signature predicting the overall survival of cervical cancer

Lu Yuan, MD[a], Zijun Lu, MD[b], Guoqiang Sun, MD[a,*], Dongmei Cao, MD[a,*]

Abstract

Cervical cancer (CC) is one of the most common gynecological malignancies, ranking fourth in both incidence and mortality in women worldwide. Early screening and treatment are of great significance in reducing the incidence and mortality of CC. Due to the complex molecular mechanisms of tumor progression, the predictive power of traditional clinical information is limited. In this study, an effective molecular model is established to assess prognosis of patients with CC and guide clinical treatment so as to improve their survival rate. Three high quality datasets (GSE138080, GSE52904, GSE67522) of expression profiling were obtained from gene expression omnibus (GEO) database. Another mRNA expression and clinicopathological data of CC were obtained from The Cancer Genome Atlas (TCGA) dataset. The bioinformatic analyses such as univariate analysis, multivariate Cox proportional-hazards model (Cox) analysis and lasso regression analysis were conducted to select survival-related differentially expressed genes (DEGs) and further establish a prognostic gene signature. Moreover, the performance of prognostic gene signature was evaluated based on Kaplan–Meier curve and receiver operating characteristic (ROC) curve. Gene set enrichment analysis (GSEA) and tumor immunity analysis were carried out to elucidate the molecular mechanisms and immune relevance. A 4-gene signature comprising procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (PLOD2), spondin1 (SPON1), secreted phosphoprotein 1 (SPP1), ribonuclease H2 subunit A (RNASEH2A) was established to predict overall survival (OS) of CC. The ROC curve indicated good performance of the 4-gene signature in predicting OS of CC based on the TCGA dataset. The 4-gene signature classified the patients into high-risk and low-risk groups with distinct OS rates of CC. Univariate analysis and multivariate Cox regression analysis revealed that the 4-gene signature was an independent factor affecting the prognosis of patients with CC. Our study developed a 4-gene signature capable of predicting the OS of CC. The findings may be beneficial to individualized clinical treatment and timely follow-up for patients with CC.

Abbreviations: AUC = area under curve, CC = cervical cancer, COX = Cox proportional-hazards model, DEGs = differentially expressed genes, FIGO = international federation of gynecology and obstetrics, GEO = gene expression omnibus, GO = gene ontology, GSEA = gene set enrichment analysis, KEGG = Kyoto encyclopedia of genes and genomes, OS = overall survival, PLOD2 = procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2, RNASEH2A = ribonuclease H2 subunit A, ROC = receiver operating characteristic, SPON1 = spondin1, SPP1 = secreted phosphoprotein 1, TCGA = the cancer genome atlas, TNM = tumor node metastasis.

Keywords: cervical cancer, gene expression, gene signature, overall survival

1. Introduction

Cervical cancer (CC) is one of the most common gynecological malignancies, ranking fourth in both incidence and mortality in women worldwide.[1] Comprehensive treatment such as surgery, radiotherapy and chemotherapy are currently the main treatment modes for CC. Under the guidance on the stage and pathological classification of CC proposed by International Federation of Gynecology and Obstetrics (FIGO), we can initially make treatment and judgement for prognosis of CC.[2] However, cancer is a heterogeneous disease, of which the outcomes can differ greatly even for patients having the same clinical characteristics and the same clinical treatment.[3] It suggests that current classifications and clinicopathologic characteristics

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

The data were downloaded from GEO and TCGA database. There was no requirement for ethics committee approval.

Supplemental Digital Content is available for this article.

[a] From the Department of Obstetrics, Maternal and Child Health Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, No. 745, Wuhan, China, [b] From the Department of Gynecology and Obstetrics, Wuhan University of Science and Technology, No. 2, Wuhan, China.

** Correspondence: Dongmei Cao, Department of Obstetrics, Maternal and Child Health Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, No. 745, Wuluo Road, Hongshan District, Wuhan 430070, China (e-mail: cdm090506@163.com), and Guoqiang Sun, Department of Obstetrics, Maternal and Child Health Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, No. 745, Wuluo Road, Hongshan District, Wuhan 430070, China (e-mail: sunguoqiang@hbfy.com).*

How to cite this article: Yuan L, Lu Z, Sun G, Cao D. Identification and verification of a 4-gene signature predicting the overall survival of cervical cancer. Medicine 2022;101:42(e31299).

Received: 2 May 2022 / Received in final form: 17 September 2022 / Accepted: 20 September 2022

http://dx.doi.org/10.1097/MD.0000000000031299

are inadequate for making an accurate prognostication and risk stratification. Hence, the identification of new biomarkers with higher predictive value is an important way to improve the prognosis of CC.

Recently, methods based on RNA-seq and bioinformatics have been developed to identify the key genes that influence the occurrence, progression, diagnosis and prognosis in cancer.[4] Many database-based analyses have been carried out to predict the prognosis of cancer patients. For example, the Oncotype DX assay, a breast cancer recurrence score based on 21-gene expression, was developed to address the need for optimizing the selection of adjuvant systemic therapy for patients with estrogen receptor-positive, lymph node-negative breast cancer.[5] Major oncology societies and entities, including the American Society of Clinical Oncology, the National Comprehensive Cancer Network, the European Society for Medical Oncology, have included Oncotype DX into breast cancer guidelines.[6–10] Moreover, Coloprint, an 18-gene expression signature designed to predict disease relapse in patients with early-stage colorectal cancer, has been used to predict the development of distant metastasis of patients with stage II colon cancer and facilitate the identification of patients who may be safely managed without chemotherapy.[11–13] In face of a large number of sequencing data, the application of bioinformatics analysis to identify effective prognostic molecular markers has indicative significance for the prognosis and treatment of cancer patients.

In this study, differentially expressed genes (DEGs) between CC tissues and normal cervix were screened by integrating 3 gene expression omnibus (GEO) databases. Subsequently, univariate and Lasso-Cox regression analyses were carried out to identify overall survival (OS)-related DEGs in The Cancer Genome Atlas (TCGA) cohort. According to the gene expression and clinical data, a 4-gene prognostic signature was proposed, and its predictive capacity was tested by survival analysis and receiver operating characteristic (ROC) curve. Multivariate Cox survival analysis was carried out to identify the independent prognostic factors of OS. The relevance between prognostic gene signature and tumor immunity was investigated to identify its application potential in guiding immune therapy.

2. Materials and Methods

2.1. Data collection

Three GEO datasets, GSE138080, GSE52904, GSE67522 were downloaded from GEO website (https://www.ncbi.nlm.nih.gov/geo/) via GEO query package. The RNA-seq data and corresponding clinical information of CC patients were obtained from TCGA databases (https://portal.gdc.cancer.gov/). After removing the patients whose survival time was less than 90 days, 259 patients were finally enrolled and randomly assigned to a training set (n = 129) and a testing set (n = 130) for further analysis.

2.2. Identification of DEGs

The DEGs were calculated using the limma package. Fold change > 2 and $P < .05$ were set as the cutoffs to screen out DEGs. DEGs were displayed via volcano plots and heat maps. The intersecting DEGs among 3 datasets were examined using the VennDiagram package.

2.3. Functional enrichment analysis

To reveal the functions of the intersecting DEGs, cluster profiler package was used for Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. The difference was considered statistically significant when $P < .05$.

2.4. Construction of prognostic model

To investigate the prognostic value of the DEGs, univariate Cox proportional-hazards model (Cox) analysis was performed using survival package. To explore the prognostic model, Lasso and multivariate Cox regression analyses were performed to assess the relationship between prognostic DEGs expression and OS. Risk scores of each patient were acquired based on gene expression amount multiplied a linear regression coefficient obtained from the multivariate Cox regression. Patients were divided into high-risk and low-risk group according to the median of risk scores.

2.5. Correlation analysis between DEGs signature and immune cells infiltration

To explore the relationship between prognostic signature and immune cells infiltration, The Stromal Score, Immune Score, ESTIMATE Score, and Tumor Purity were analyzed by ESTIMATE algorithm. Tumor Immune Estimation Resource, a useful resource for comprehensive analysis of tumor infiltrating immune cells, was employed to analyze the association between prognostic signature and 6 types of immune cells (B cells, $CD4^+$ T cells, $CD8^+$ T cells, dendritic cells, macrophages). The correlation between risk scores and immune infiltration was calculated by Pearson correlation.

2.6. Gene set enrichment analysis

To analyze the hallmark gene sets of the prognostic signature, GSEA was performed between high-risk phenotype and low-risk phenotype. Hallmark gene sets were downloaded from Molecular Signatures Database (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp).[14]

2.7. Statistical analysis

All statistical analyses in this study were conducted using R software (version 3.6.3). Survival analyses were carried out using survival package, with Log-Rank test conducted for verification. ROC curves were plotted using survival ROC package. Univariate and multivariate Cox proportional hazards regression analyses were performed to valuate prognostic value of clinical features and the risk scores. Group comparisons were conducted for continuous variables using Mann–Whitney U test. $P < .05$ indicated statistical significance of the difference.

3. Results

3.1. Identification of DEGs

Three datasets of GEO (GSE138080, GSE52904, GSE67522) comprising 951, 875 and 1048 DEGs were identified between CC tissues and normal cervix (Fig. S1, http://links.lww.com/MD/H724). Among them, 183 DEGs are presented in all 3 datasets (Fig. 1A).

3.2. Functional enrichment of the DEGs

GO and KEGG pathway enrichment analyses were conducted to discover the functions of the 183 intersected DEGs (Fig. 1B and C). The results of GO analysis revealed that 183 DEGs were associated with DNA replication, cell cycle, chromosomal

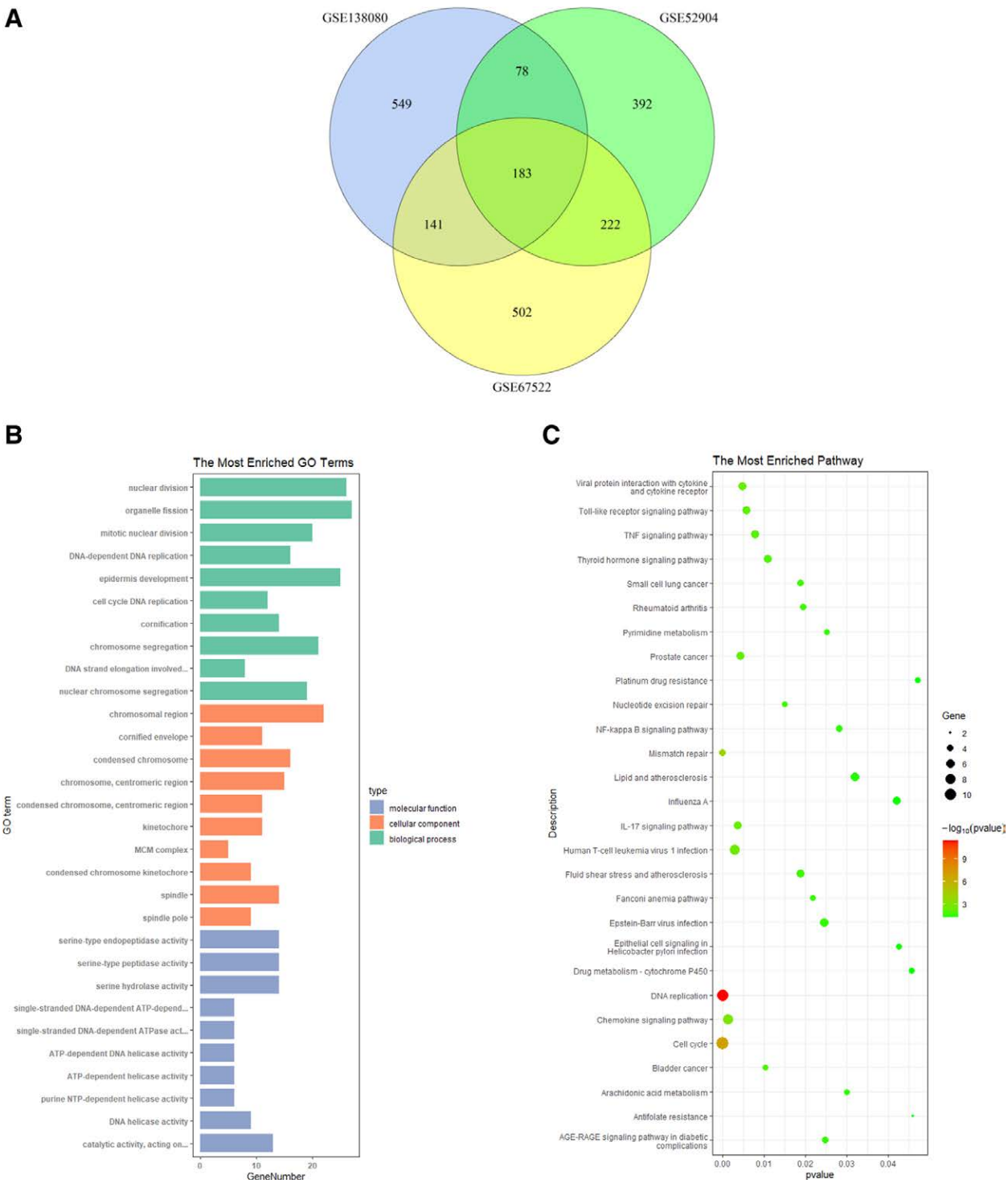

Figure 1 Functional enrichment analysis of the DEGs. (A) The Venn diagram demonstrating intersecting DEGs in 3 GEO datasets. (B) GO Functional enrichment analysis of the 183 intersecting DEGs. The *x*-axis represents the number of genes in each term, and the y-axis represents enriched GO terms. (C) KEGG Functional enrichment analysis of the 183 intersecting DEGs. The *x*-axis represents the *P* value for each pathway in the enrichment analysis, and the *y*-axis represents enriched KEGG pathways. DEGs = differentially expressed genes, GEO = gene expression omnibus, GO = gene ontology, KEGG = Kyoto encyclopedia of genes and genomes.

region, helicase activity from the categories of biological process, cellular component and molecular function, respectively. KEGG pathway analysis revealed that the DEGs participated in DNA replication, cell cycle, mismatch repair, prostate cancer, tumor necrosis factor signaling pathway, bladder cancer, microRNAs in cancer, nuclear factor-kappa B signaling pathway.

3.3. Identification of survival-related DEGs and establishment of the 4-gene prognostic signature

The TCGA-CESC dataset was randomly divided into training set and testing set, respectively. One hundred twenty-nine patients from the training set were included in subsequent survival analyses. Through univariate cox regression analysis, 8

DEGs associated with OS were identified (Fig. 2A). Patients were divided into high-expressed and low-expressed groups according to the median of these survival-related genes. The corresponding survival analysis for each gene is shown in Figure S2, http://links.lww.com/MD/H725. Lasso-penalized Cox regression analysis was carried out to further reduce the number of DEGs in the selected panel with best predictive performance using 10-fold cross validation based on glmnet

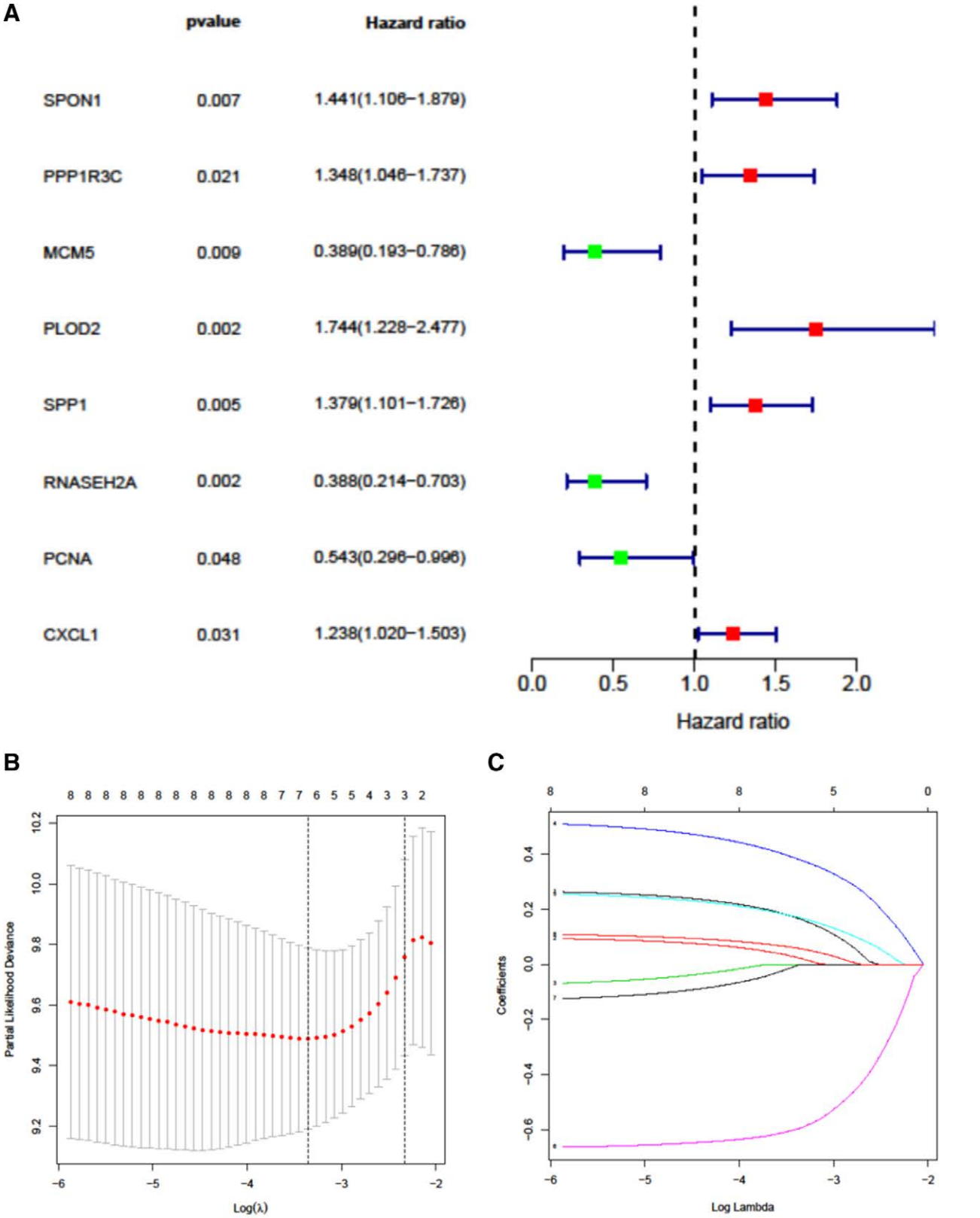

Figure 2 Evaluation of DEGs with prognostic value. (A) The forest plot shows the hazard ratio (HR) values for 183 DEG calculated by univariate COX analysis. Only the results for $P < .05$ are shown. (B and C) Lasso analysis of the prognostic DEGs in cervical cancer. COX = Cox proportional-hazards model, DEGs = differentially expressed genes.

package (Fig. 2B and C). After Lasso-penalized Cox regression analysis, a prognostic signature comprising 4 genes, including procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (PLOD2), spondin1 (SPON1), secreted phosphoprotein 1 (SPP1), ribonuclease H2 subunit A (RNASEH2A), was developed by multivariate Cox analysis. The risk score was calculated as follows:

[(0.62218) × Expression value of PLOD2] + [(0.24936) × Expression value of SPON1] + [(0.27333) × Expression value of SPP1] + [(−0.88808) × Expression value of RNASEH2A].

Patients were divided into high-risk and low-risk groups according to median of risk scores. Kaplan–Meier survival curves showed that the patients in the low-risk group had a longer survival duration that those in the high-risk group (Fig. 3A). The 1-, 3-, and 5-year survival rates were evaluated by risk scores, with area under curve values of 0.836, 0.806, and 0.823 respectively, as shown in Figure 3B. Distribution of the risk scores, survival status and the mRNA expression heat map in the training set are shown in Figure 3C–E. These results demonstrated that the 4-gene signature had solid performance in both high sensitivity and specificity, which can well predict the survival of patients with CC.

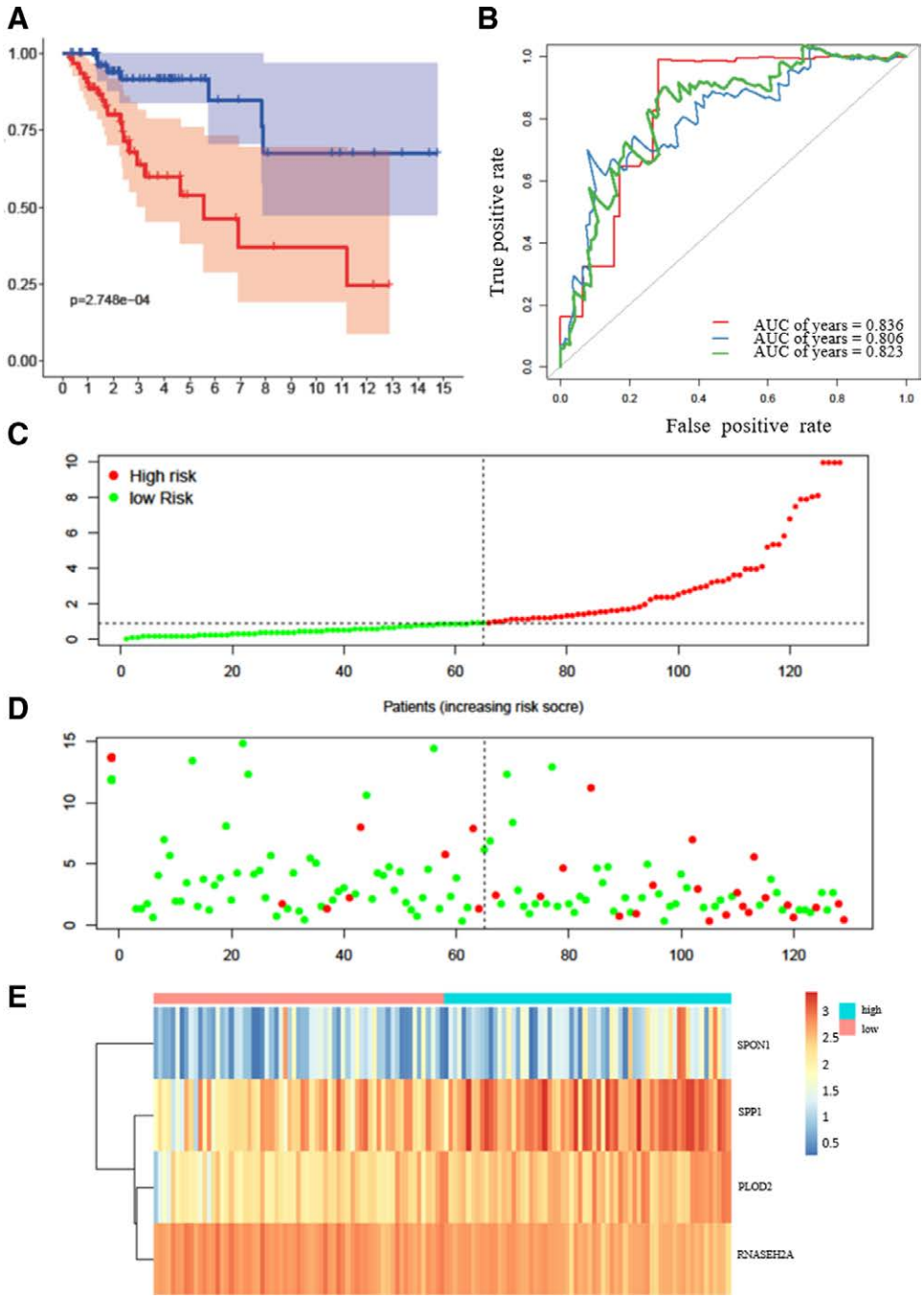

Figure 3 Development of risk score based on the 4-gene signature of patients with CC in the training set. (A) Kaplan–Meier survival curves of the CC samples based on the 4-gene signature in the training set. The *x*-axis represents survival time, and the *y*-axis represents survival probability. (B) AUC of time-dependent ROC curves verified the prognostic performance of the risk scores. The *x*-axis represents the false positive rate, and the y-axis represents the true positive rate. (C–E) Distribution of the risk score and the associated survival data and mRNA expression heat map in Training Set. AUC = area under curve, CC = cervical cancer, ROC = receiver operating characteristic.

3.4. Validation of the performance of the 4-gene signature

The robustness of the 4-gene signature was validated based on the testing set and entire TCGA-CESC dataset, respectively. Risk scores of each patient were calculated according to the formula listed above. According to median of risk scores, the patients were divided into high-risk and low-risk groups. The outcome of the low-risk group was significantly better than that of the high-risk group (Fig. 4A and B). In the testing set, the area under curves for 1-, 3-, and 5-year OS predictions were 0.589, 0.626, and 0.710 (Fig. 4C), respective, while those were 0.680, 0.699, and 0.755

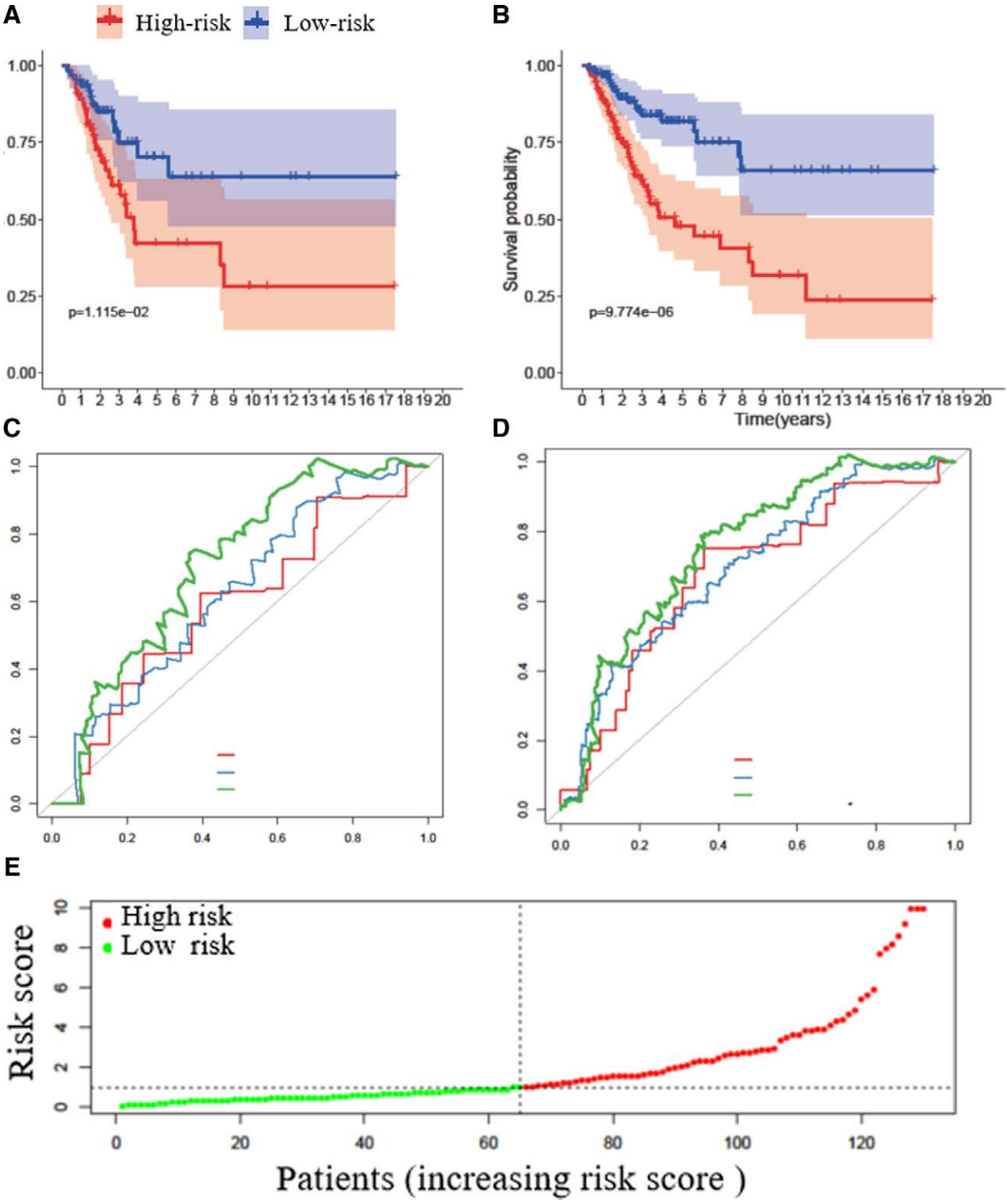

Figure 4 Validation of the 4-gene signature in patients with CC in the testing set and entire TCGA datasets. (A and B) Kaplan–Meier survival curves of the CC samples based on the 4-gene signature in the testing set and the entire TCGA datasets. (C and D) AUC of time-dependent ROC curves verified the prognostic performance of the risk scores. The *x*-axis represents the false positive rate, and the y-axis represents the true positive rate. (E, G, and I) Distribution of the risk scores and the associated survival data and mRNA expression heat map in the testing set. (F, H, and J) Distribution of the risk score and the associated survival data and mRNA expression heat map in the entire TCGA dataset. AUC = area under curve, CC = cervical cancer, ROC = receiver operating characteristic, TCGA = the cancer genome atlas.

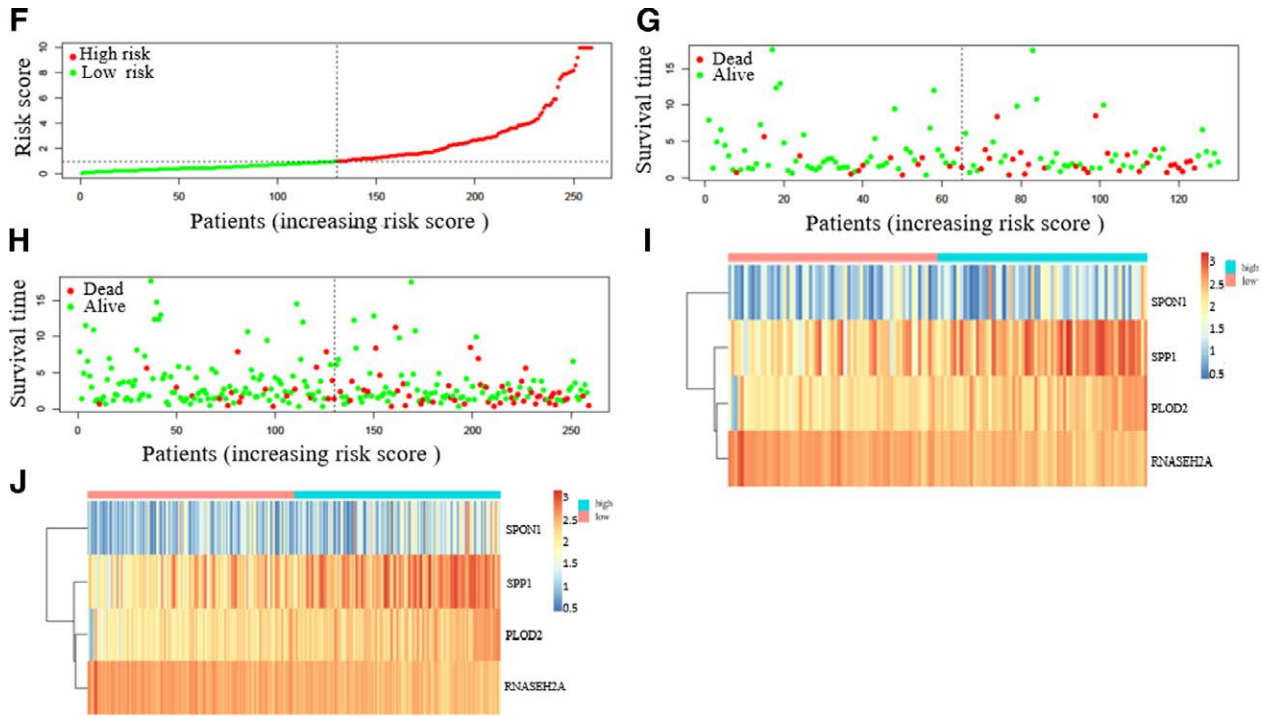

Figure 4 Continued

in the entire TCGA-CESC dataset, respectively (Fig. 4D). The distributions of the risk scores, the associated survival data and the mRNA expression heat map are displayed in Figure 4E–J, respectively. These 2 datasets demonstrate that the 4-gene signature performs well in predicting OS for patients with CC.

3.5. Association between the 4-gene signature and patients' survival outcomes

The survival curve demonstrated that patients with high-risk were associated with a downward trend of OS outcomes in training, testing, and entire sets. Afterwards, patients were divided into different subgroups according to clinical characteristics for survival analysis, and it was found that the 4-gene signature could predict the OS of subgroup of CC, including patients in G1–G2 grade, G3 grade, M0 stage, N1 stage, T1 stage, T2–T3 stage, FIGO stage I–II, FIGO stage III–IV (Fig. 5A–G). However, there were no correlation between risk scores and OS in N0 stage, N1 stage (Fig. 5H–I). Further, univariate and multivariate Cox regression analyses were carried out to evaluate prognostic significances of the 4-gene signature and various clinicopathologic characteristics. Univariate Cox regression analysis indicated that FIGO stage and risk scores were correlated with OS of CC patients (Fig. 5J). Subsequent multivariate Cox regression analysis showed that risk scores were independently associated with OS of CC patients (Fig. 5K). These results demonstrate that the proposed signature model is an independent factor affecting the prognosis of patients with CC.

3.6. Gene set enrichment analysis

To investigate the underlying molecular mechanism of the prognostic signature, GSEA was performed between high-risk group and low-risk group among totally 259 patients in the entire set (Fig. 6A and B). In the high-risk group, the enriched hallmark gene sets were mainly concentrated on various processes associated with tumor progression, including epithelial mesenchymal transition, tumor necrosis factor signaling via nuclear factor kappa B, hypoxia, apoptosis, and inflammatory response). In the low-risk group, 4 biological processes signatures including E2F targets, oxidative phosphorylation, DNA repair and spermatogenesis were enriched.

3.7. Correlation between DEGs prognostic signature for CC and the infiltration of immune cells

Considering the interaction between tumor and host immune system influencing patient prognosis,[15] ESTIMATE algorithm was adopted to analyze the differences in tumor purity, ESTIMATE scores, immune scores and stromal scores between high-risk and low-risk patients. As shown in Figure 6C–E, ESTIMATE scores and stromal score were higher in high-risk group. Subsequently, we analyzed the correlation between the prognostic signature and the infiltration of immune cell subtypes in CC using the data from Tumor Immune Estimation Resource database. As shown in Figure 6F–K, the correlation values of B cells, CD4 + T cells, CD8 + T cells, dendritic, macrophages, neutrophils with risk score were –0.087 (P = .162), –0.04 (P = .518), –0.041 (P = .510), 0.117 (P = .061), 0.120 (P = .054) and 0.132 (P = .034), respectively, suggesting that the infiltration of neutrophil was significantly positive correlated with the prognosis of CC.

4. Discussion

The malignancy of CC (such as metastasis, recurrence, and drug resistance) is a complex and precise process based on abnormal expression of specific genes.[16,17] Therefore, the underlying molecules influencing the prognosis of patients with CC may be altered before detectable clinicopathologic abnormalities occur. It is of significance to screen prognostic molecular markers, which is critical to the individualized prevention, treatment and timely follow-up of CC patients. In our study, a 4-gene signature associated with OS in CC patients was identified by analyzing the expression profiles of 259 CC samples from TCGA. This model was an independent prognostic factor of CC. In addition,

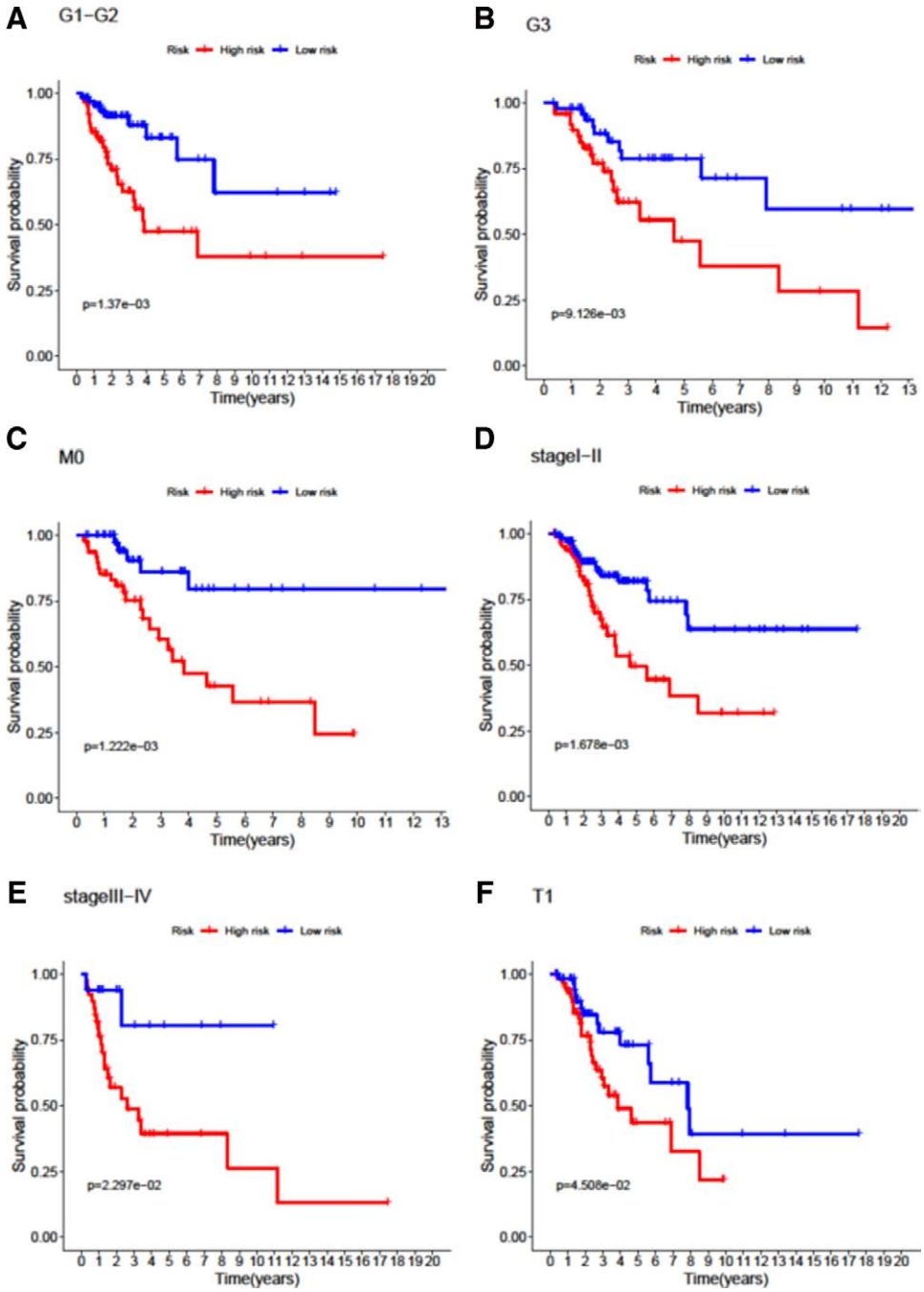

Figure 5. (A–K) Kaplan–Meier survival analysis for the patients divided by each clinical feature. (A–I.) Kaplan–Meier plots of OS between low-risk and high-risk groups based on subgroups according to TNM staging, histological grade and FIGO stage. (J and K) The univariate (J) and multivariate (K) Cox regression analysis of risk score, age, TNM stage, grade, and disease type. COX = Cox proportional-hazards model, FIGO = international federation of gynecology and obstetrics, OS = overall survival, TNM = tumor node metastasis.

we discovered that the risk score was positively related to the infiltration of neutrophil.

Existing studies have shown that the 4 genes are closely related to the development of a variety of cancers. PLOD2, a key enzyme mediating the formation of stabilized collagen cross-links,[18] was confirmed to mediate hypoxia-induced cancer metastasis via collagen modification and ECM remodeling in a variety of cancers, such as sarcoma, breast cancer and lung cancer.[19–22] Xu et al proved that hypoxia- and transforming growth factor β1-induced PLOD2 expression promoted the migratory, invasive and adhesive capacities of CC cells by participating in TGF-β1-induced epithelial mesenchymal transition and the formation of focal adhesions.[23] SPP1, also known as osteopontin, plays a role in processes such as immune response, cell adhesion and migration, and tumorigenesis.[24] Several studies have shown that SPP1 is overexpressed in ovarian cancer, gastric, colon, renal, breast, esophageal and endometrial cancers.[24–26] Chen et al demonstrated SPP1 was overexpressed in CC tissues and cell lines and the downregulation of SPP1 improved the cisplatin sensitivity of HeLa by inhibiting the phosphatidylinositol-3-kinases/protein kinase B signaling pathway.[27] RNASEH2A, a member of the RNase HII family, participates in DNA replication by mediating removal of lagging-strand Okazaki fragment RNA primers and impacts invasiveness and chemoresistance, resulting in poor survivability of breast cancer in ER dependent manner.[28] RNASEH2A may be associated with the occurrence

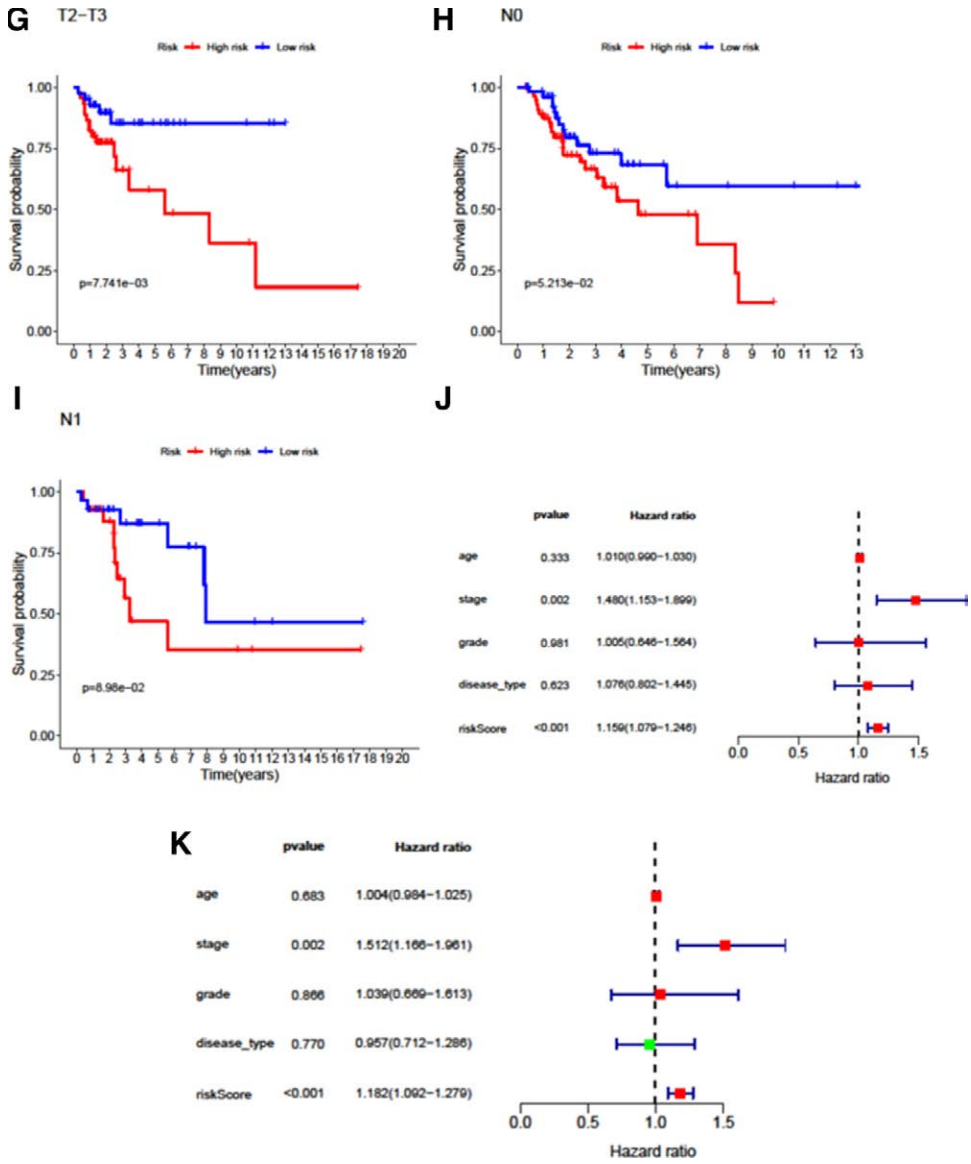

Figure 5. Continued

of human gliomas by regulating cell proliferation and apoptosis.[29] However, there are few reports on the role of RNASEH2A in CC. SPON1, an important member of the thrombospondin family, has been reported to promote metastasis in human osteosarcoma.[30] However, its role in CC has not been investigated. Given that cancer progression is a process involving multiple molecules, a 4-gene signature was developed in this study, which may assess prognostic risk of patients more accurately.

Previous study has reported that immune infiltration is vital in response to treatment and prognosis of CC.[31,32] In this study, the prognostic signature was identified to have significant correlation with neutrophils infiltration. Wisdom, A. J. et al found that high levels of neoplastic infiltrating neutrophils were associated with poorer OS in CC and inhibition of neutrophils may be one of the mechanisms to improve the prognosis of CC patients.[33] Therefore, this model may be used to evaluate the individualized immunotherapy of CC, though further study of its specific mechanism is needed.

Since the prognosis of CC patients varies widely and there are no effective biomarkers, the therapeutic effects of patients cannot be predicted accurately. With the development of high-throughput sequencing technology, molecular researches on the occurrence and development of CC have been carried out increasingly. In this study, a gene model was constructed to evaluate the prognosis of CC patients and guide postoperative treatment. The results showed that the proposed model can well predict neutrophils infiltration in the progression of CC. However, the prediction model needs to be further validated based on multicenter, large-scale clinical trials, and further prospective studies are required.

5. Conclusion

In conclusion, by analyzing RNA sequence-based gene expression signatures and clinical data in TCGA and GEO patients, a 4-gene signature prognostic stratification system was developed, which can reliably predict OS in CC and may facilitate individualized treatment and timely follow-up for patients with CC.

6. Limitation

The main limitation of this study is the lack of molecular validation. In the next study, we will detect the expression levels of these 4 genes in CC cells (SiHa, HeLa, Caski) by Western Blot and PCR, detect the expression levels of these 4 genes in CC tissues and paracancerous tissues by immunohistochemical experiments.

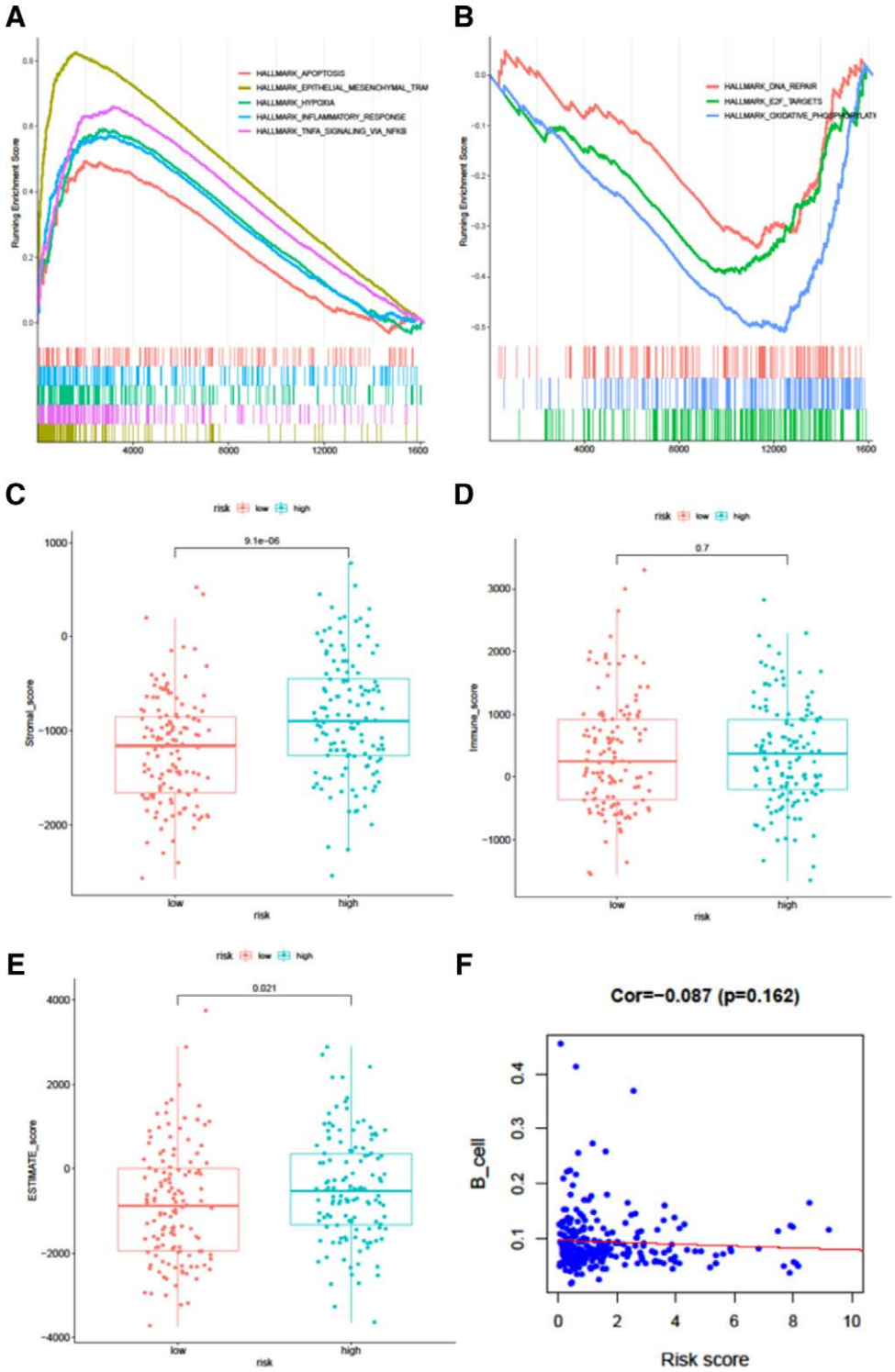

Figure 6. (A–K) Correlation of the risk score with infiltrative immune cells. (A) GSEA analysis demonstrating the enriched hallmark in the high-risk group. (B) GSEA analysis demonstrating the enriched hallmark in the low-risk group. (C–E) Boxplots show the immune scores, stromal scores and ESTIMATE scores between high and low-risk groups. (F–K) Correlation between the 4-gene prognostic signature for cervical cancer and the infiltration of immune cell subtypes (B cells, CD4 + T cells, CD8 + T cells, neutrophil, macrophage, and dendritic cells). The *x*-axis represents risk scores, and the *y*-axis represents the infiltration scores of each type of immune cell. GSEA = gene set enrichment analysis.

Author contributions

Conceptualization: Lu Yuan, Zijun Lu.
Data curation: Lu Yuan, Zijun Lu.
Methodology: Zijun Lu.
Supervision: Guoqiang Sun.
Writing – original draft: Lu Yuan.
Writing – review & editing: Dongmei Cao.

Acknowledgments

The authors are grateful to the support from the GEO and TCGA dataset.

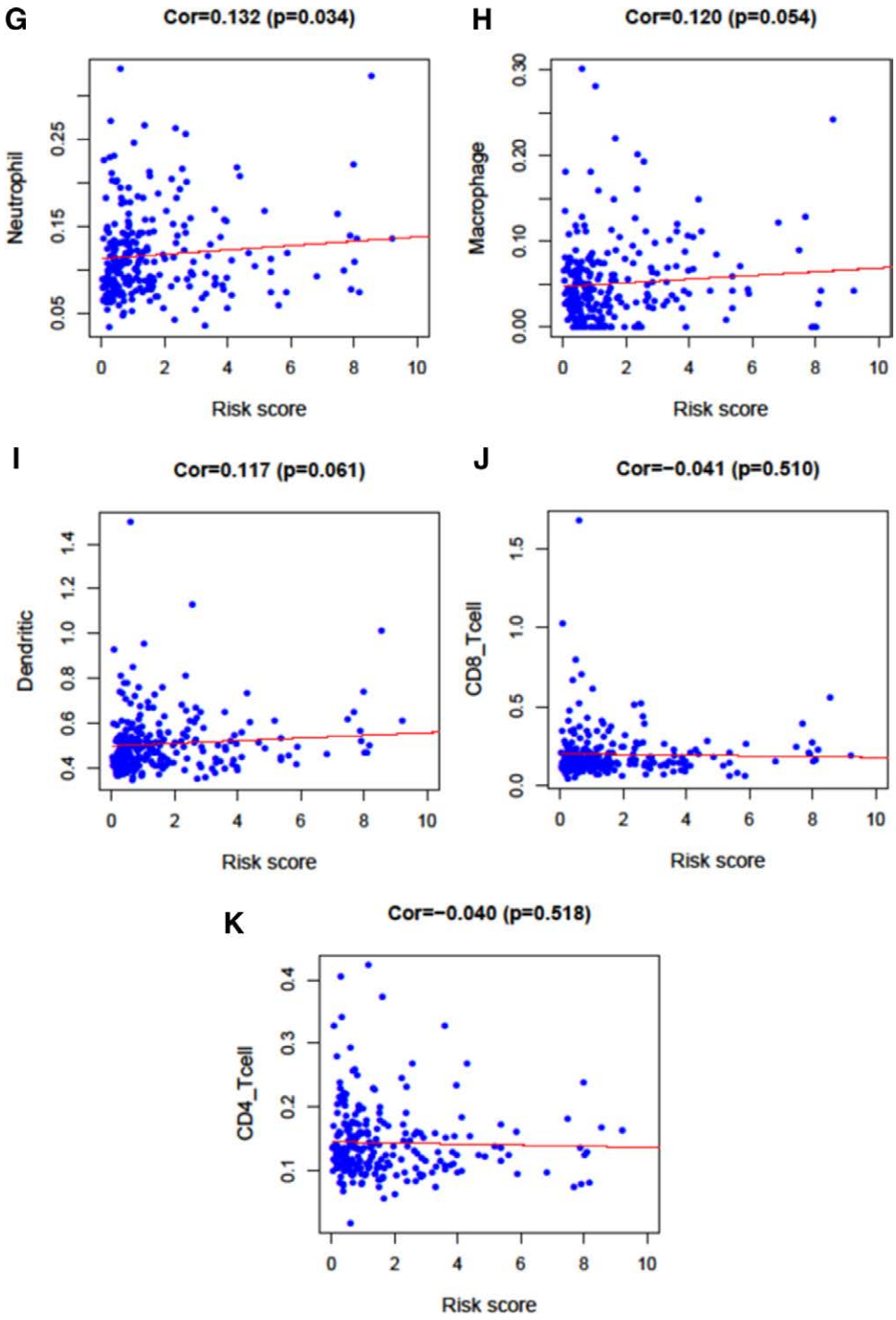

Figure 6. Continued

References

[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
[2] Bhatla N, Aoki D, Sharma DN, et al. Cancer of the cervix uteri. Int J Gynaecol Obstet. 2018;143(Suppl 2):22–36.
[3] Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15:81–94.
[4] Kamps R, Brandao RD, Bosch BJ, et al. Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification. Int J Mol Sci. 2017;18:308.
[5] Siow ZR, De Boer RH, Lindeman GJ, et al. Spotlight on the utility of the Oncotype DX® breast cancer assay. Int J Womens Health. 2018;10:89–100.
[6] Harris LN, Ismaila N, Mcshane LM, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol. 2016;34:1134–50.
[7] Daly MB, Pilarski R, Berry M, et al. NCCN guidelines insights: genetic/familial high-risk assessment: breast and ovarian, version 2.2017. J Natl Compr Canc Netw. 2017;15:9–20.
[8] Senkus E, Kyriakides S, Ohno S, et al. Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2015;26(Suppl 5):v8–30.

[9] Gnant M, Harbeck N, St TC. Gallen/Vienna 2017: a brief summary of the consensus discussion about escalation and de-escalation of primary breast cancer treatment. Breast Care (Basel). 2017;12:102–7.
[10] Ward S, Scope A, Rafia R, et al. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2013;17:1–302.
[11] Tan IB, Tan P. Genetics: an 18-gene signature (ColoPrint®) for colon cancer prognosis. Nat Rev Clin Oncol. 2011;8:131–3.
[12] Kopetz S, Tabernero J, Rosenberg R, et al. Genomic classifier ColoPrint predicts recurrence in stage II colorectal cancer patients more accurately than clinical factors. Oncologist. 2015;20:127–33.
[13] Maak M, Simon I, Nitsche U, et al. Independent validation of a prognostic genomic signature (ColoPrint) for patients with stage II colon cancer. Ann Surg. 2013;257:1053–8.
[14] Liberzon A, Birger C, Thorvaldsdottir H, et al. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25.
[15] Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. 2016;17:174.
[16] Bahrami A, Hasanzadeh M, Shahidsales S, et al. Genetic susceptibility in cervical cancer: from bench to bedside. J Cell Physiol. 2018;233:1929–39.
[17] Burki TK. Novel mutations in cervical cancer. Lancet Oncol. 2017;18:e137.
[18] Yamauchi M, Sricholpech M. Lysine post-translational modifications of collagen. Essays Biochem. 2012;52:113–33.
[19] Eisinger-Mathason TS, Zhang M, Qiu Q, et al. Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. Cancer Discov. 2013;3:1190–205.
[20] Gilkes DM, Bajpai S, Wong CC, et al. Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. Mol Cancer Res. 2013;11:456–66.
[21] Du H, Chen Y, Hou X, et al. PLOD2 regulated by transcription factor FOXA1 promotes metastasis in NSCLC. Cell Death Dis. 2017;8:e3143–e3143.
[22] Chen Y, Terajima M, Yang Y, et al. Lysyl hydroxylase 2 induces a collagen cross-link switch in tumor stroma. J Clin Invest. 2015;125:1147–62.
[23] Xu F, Zhang J, Hu G, et al. Hypoxia and TGF-beta1 induced PLOD2 expression improve the migration and invasion of cervical cancer cells by promoting epithelial-to-mesenchymal transition (EMT) and focal adhesion formation. Cancer Cell Int. 2017;17:54.
[24] Zhao H, Chen Q, Alam A, et al. The role of osteopontin in the progression of solid organ tumour. Cell Death Dis. 2018;9:356.
[25] Zeng B, Zhou M, Wu H, et al. SPP1 promotes ovarian cancer progression via Integrin beta1/FAK/AKT signaling pathway. Onco Targets Ther. 2018;11:1333–43.
[26] Liu J, Xu K, Chase M, et al. Tiam1-regulated osteopontin in senescent fibroblasts contributes to the migration and invasion of associated epithelial cells. J Cell Sci. 2012;125(Pt 2):376–86.
[27] Chen X, Xiong D, Ye L, et al. SPP1 inhibition improves the cisplatin chemo-sensitivity of cervical cancer cell lines. Cancer Chemother Pharmacol. 2019;83:603–13.
[28] Shen J, Lin S, Liu L, et al. Ribonuclease H2 subunit A impacts invasiveness and chemoresistance resulting in poor survivability of breast cancer in ER dependent manner. Am J Transl Res. 2020;12:2281–94.
[29] Dai B, Zhang P, Zhang Y, et al. RNaseH2A is involved in human gliomagenesis through the regulation of cell proliferation and apoptosis. Oncol Rep. 2016;36:173–80.
[30] Chang H, Dong T, Ma X, et al. Spondin 1 promotes metastatic progression through Fak and Src dependent pathway in human osteosarcoma. Biochem Biophys Res Commun. 2015;464:45–50.
[31] Shah W, Yan X, Jing L, et al. A reversed CD4/CD8 ratio of tumor-infiltrating lymphocytes and a high percentage of CD4(+)FOXP3(+) regulatory T cells are significantly associated with clinical outcome in squamous cell carcinoma of the cervix. Cell Mol Immunol. 2011;8:59–66.
[32] Wang J, Li Z, Gao A, et al. The prognostic landscape of tumor-infiltrating immune cells in cervical cancer. Biomed Pharmacother. 2019;120:109444.
[33] Wisdom AJ, Hong CS, Lin AJ, et al. Neutrophils promote tumor resistance to radiation therapy. Proc Natl Acad Sci USA. 2019;116:18584–9.