Identification and Verification of a Four-Gene Signature Predicting Overall Survival for Cervical Cancer

Lu Yuan
Maternal and Child Health Hospital of Hubei Province: Hubei Province Women and Children Hospital
Guoqing Li
Huazhong University of Science and Technology Tongji Medical College https://orcid.org/0000-0002-6557-1301
Guoqiang Sun
Maternal and Child Health Hospital of Hubei Province: Hubei Province Women and Children Hospital
Dongmei Cao (237544804@qq.com)
Maternal and Child Health Hospital of Hubei Province: Hubei Province Women and Children Hospital

Research

Keywords: gene signature, overall survival, cervical cancer, clinical prognosis

Posted Date: June 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-587544/v1

Abstract

Background. Cervical cancer is one of the most common gynecological malignancies, ranking fourth for both incidence and mortality in women worldwide. Early screening and treatment have significantly reduced the incidence and mortality of cervical cancer. Many studies have shown that cervical cancer still presents a high risk of death from recurrence and metastasis. However, due to the complex molecular mechanisms of tumor progression, the predictive power of traditional clinical information is limited. In order to improve the survival rate of patients with cervical cancer, establishing an effective molecular model to assess prognosis of patients and guide clinical treatment are very important.

Results. four-gene signature comprising PLOD2, SPON1, SPP1, RNASEH2A was established to predict overall survival of cervical cancer. The ROC curve indicated good performance of the four-gene signature at predicting overall survival in the TCGA dataset. The four-gene signature classify patients into high-risk and low-risk groups with distinct overall survival. Univariate and multivariate Cox regression revealed that the four-gene signature was an independent prognostic factor in cervical cancer.

Conclusion. Our study developed a four-gene signature that reliably predict overall survival in cervical cancer. The findings may be beneficial to individualized clinical treatment and timely follow-up.

Background:

Cervical cancer (CC) is one of the most common gynecological malignancies, ranking fourth for both incidence and mortality in women worldwide[1]. Comprehensive treatment such as surgery, radiotherapy and chemotherapy are currently the main treatment model. The current International Federation of Gynecology and Obstetrics (FIGO) stage and pathological classification can initially provide us with the guidance of treatment and judgement for prognosis[2]. However, cancer is a heterogeneous disease, and its outcomes can differ greatly even for patients with the same clinical characteristics and the same clinical treatment[3]. It suggested us that current classifications and clinicopathologic characteristics are restricted or limited for 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 used 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]. Many 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 their breast cancer guidelines[6–10]. Another example is Coloprint, an 18-gene expression signature designed to predict disease relapse in patients with early-stage colorectal cancer, can effectively predict the development of distant metastasis

of patients with stage II colon cancer and facilitates 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 comprehensive and reliable bioinformatics analysis to identify effective prognostic molecular markers has indicative significance for the prognosis and treatment of cancer patients.

Here, differentially expressed genes (DEGs) between CC tissues and normal cervix were screened by integrating three Gene Expression Omnibus (GEO) databases. Subsequently, we used univariate and Lasso-Cox regression analyses to identify overall survival (OS)-related DEGs in The Cancer Genome Atlas (TCGA) cohort. According to the gene expression and clinical data, we propose a 4-gene prognostic signature. Survival analysis and ROC curve were used to test the prediction effectiveness. Multivariate Cox survival analysis was used to identify the independent prognostic factors of OS and the relevance between prognostic signature and tumor immunity was investigated to identify its potential in guidance of immune therapy.

Methods

Data collection

Three GEO datasets, GSE138080, GSE52904, GSE67522 were downloaded from GEO website (https://www.ncbi.nlm.nih.gov/geo/) via GEOquery 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 randomly assigned to a training set (n = 129) and a testing set (n = 130) for further analysis.

Differentially expressed genes identification

The DEGs were calculated using the limma package. Fold change > 2 and $P < 0.05$ were set as the cut-offs to screen for DEGs. Volcano plots and heat maps were performed to display DEGs. The intersecting DEGs among three datasets were examined using the VennDiagram package.

Functional enrichment analysis

To reveal the functions of the intersecting DEGs, clusterProfiler package was used for GO enrichment and KEGG pathway analysis, adjusted $P < 0.05$ was considered statistically significant.

Construction of prognostic model

To investigate the prognostic value of the DEGs, univariate 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 genes expression multiplied a linear combination of regression coefficient obtained from the multivariate Cox regression. Patients were divided into high-risk and low-risk group according to their medium risk score.

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. (TIMER), a useful resource for comprehensive analysis of tumor infiltrating immune cells, was employed to analyze the association between prognostic signature and six 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.

Gene set enrichment analysis

To analysis the hallmark gene sets of the prognostic signature, GSEA was performed between high-risk and low-risk phenotype. Hallmark gene sets were downloaded from Molecular Signatures Database (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp)[33].

Statistical analysis

All statistical analysis of this research was conducted using R software (version 3.6.3). Survival analyses were carried out using survival package and Log-Rank test was utilized to test. 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 undertaken for continuous variables using Mann-Whitney *U* test. $P < 0.05$ was considered statistically significant.

Results

Identification of DEGs

Three datasets of GEO (GSE138080, GSE52904, GSE67522) comprised of 951, 875 and 1048 DEGs were identified between CC tissues and normal cervix (Supplementary Fig. 1). Among them, 183 DEGs were present in all three datasets (Fig. 1a).

Functional Enrichment of the DEGs

GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analyses were applied to discover the functions of the 183 intersected DEGs (Fig. 1b-c). The results of GO analysis revealed that 183 DEGs were associated with DNA replication, cell cycle, chromosomal 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, TNF signaling pathway, bladder cancer, microRNAs in cancer, NF-kappa B signaling pathway.

Identification of Survival-Related DEGs and Establishment of the Four-Gene Prognostic Signature

We randomly divided the TCGA-CESC dataset into training set and testing set. One hundred twenty-nine patients from the training set were included in subsequent survival analyses. By the use of univariate cox regression analysis, 8 DEGs were identified to be associated with OS (Fig. 2a). Patients were divided into high-expressed and low-expressed groups according to the median expression of these survival-related genes. The corresponding survival analysis for each gene is shown as Supplementary Fig. 2 Lasso-penalized Cox regression analysis was applied to further reduce the number of DEGs in the selected panel with best predictive performance based on glmnet package (Fig. 2b-c). After Lasso-penalized Cox regression analysis, a prognostic signature comprising four 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 value of risk scores. Kaplan-Meier survival curves showed that patients in the low-risk group had a longer survival duration compared to the high-risk group (Fig. 3a). The 1-, 3- and 5-year survival rates were evaluated by risk scores and are displayed in Fig. 3b, with AUC values of 0.836, 0.806 and 0.823 respectively. Distribution of the risk scores, survival status and the mRNA expression heat map in the training set are displayed in Fig. 3c-e. These results demonstrated that the four-gene signature had both high sensitivity and specificity and can be used to predict the survival of patients with CC.

Validation of the performance of the four-Gene Signature

The testing set and entire TCGA-CESC dataset were used to validate the robustness of the four-gene signature respectively. Risk scores were calculated with the formula listed above for each patient. According to median value, 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-b). In the testing set, the AUCs for 1-, 3-, and 5-year OS predictions were 0.589, 0.626, and 0.710 (Fig. 4c), while in the entire TCGA-CESC dataset, the AUCs for 1-, 3-, and 5- year OS predictions were 0.680, 0.699, and 0.755, respectively (Fig. 4d). The distribution of the risk scores, the associated survival data and the mRNA expression heat map are respectively displayed in Fig. 4e-j. These two datasets effectively demonstrate that the four-gene signature performs well in predicting OS in patients with CC.

Association between the signature and patients' survival outcomes

The survival curve demonstrated that patients with high-risk were correlated with a trend toward worse OS outcomes in training, testing, and entire sets. Afterwards, patients were divided into different subgroups according to clinical characteristics for survival analysis, we found that the risk signature could predict the OS of subgroup of CC, including patients with 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). We further used univariate and multivariate Cox regression analyses to evaluate prognostic significances of four-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 was independently associated with OS of CC patients (Fig. 5k). These results demonstrated that the model was an independent prognostic factor for CC.

Gene Set Enrichment Analysis (GSAE)

To investigate the underlying molecular mechanism of the prognostic signature, GSEA was performed between high-risk group and low-risk group in 259 patients of the entire set (Fig. 6a, b). In the high-risk group, the enriched hallmark gene sets were mainly focused on various processes associated with tumor progression (including epithelial mesenchymal transition, TNFA signaling via NFκB, hypoxia, apoptosis, inflammatory response). In the low-risk group, four biological processes signatures including E2F targets, oxidative phosphorylation, DNA repair and spermatogenesis were enriched.

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[14], we used ESTIMATE algorithm to analyze the differences in tumor purity, ESTIMATE scores, immune scores and stromal scores between high-risk and low-risk patients. As shown in Fig. 6c-e, we found that ESTIMATE scores and stromal score was 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 (TIMER) database. As shown in Fig. 6f-k, the correlation values of B cells, CD4 + T cells, CD8 + T cells, dendritic, macrophages, neutrophils with risk score were − 0.087 (P = 0.162), − 0.04 (P = 0.518), − 0.041 (P = 0.510), 0.117 (P = 0.061), 0.120 (P = 0.054) and 0.132 (P = 0.034), respectively, suggesting that the infiltration of neutrophil was significantly positive correlated with the prognosis of CC.

Discussion:

The malignancy of CC (such as metastasis, recurrence and drug resistance) is a complex and precise process that requires the abnormal expression of specific genes to empower the cancer cells[15, 16]. Therefore, the underlying molecules that influence 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, we identified a four-gene signature associated with OS in patients with CC by analyzing the expression profiles of 259 CC samples from TCGA. This model was an independent prognostic factor of CC, not related to clinical factors. In addition, we discovered that the risk score was positively related to the infiltration of neutrophil.

Existing studies shown that the four genes are closely related to the development of a variety of cancers. procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 (PLOD2), a key enzyme mediating the formation of stabilized collagen cross-links[17], 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[18–21]. *Xu et al* proved that hypoxia- and TGF-β1-induced PLOD2 expression promotes the migratory, invasive and adhesive capacities of CC cells by participating in TGF-β1 induced EMT and the formation of focal adhesions[22]. Secreted phosphoprotein 1 (SPP1), also called as osteopontin (OPN), plays a role in processes such as immune response, cell adhesion and migration, and tumorigenesis[23]. Several studies have shown that SPP1 is overexpressed in ovarian cancer, gastric, colon, renal, breast, oesophageal and endometrial cancers[23–25]. *Chen et al* demonstrated SPP1 was overexpressed in CC tissues and cell lines and downregulation of SPP1 improved the cisplatin sensitivity of HeLa by inhibiting the PI3K/Akt signaling pathway[26]. Ribonuclease H2 subunit A (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[27]. The expression of RNASEH2A is elevated in gliomas, and it may be involved in the occurrence of human gliomas by regulating cell proliferation and apoptosis[28]. 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[29]. However, its role in CC has not been investigated. Given that cancer progression is a process involving multiple molecules, four genes were statistically integrated, 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 [30, 31]. In current study, the prognostic signature was identified to have significant correlation with neutrophils infiltration. *Wisdom, A. J. et* 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 patients[32]. Therefore, this model may be used to evaluate the individualized immunotherapy of CC, but its specific mechanism needs further study.

Because the prognosis of CC patients varies widely and there are no effective biomarkers, we are often cannot predict therapeutic effects of patients. With the development of high-throughput sequencing technology, molecular research on the occurrence and development of CC has been rapidly developed. We constructed a gene model that can be used to evaluate the prognosis of patients with CC to guide postoperative treatment. In addition, our model could also be used to predict neutrophils infiltration in the progression of CC. However, the prediction model needs further validation in multicenter, large-scale clinical trials and prospective studies.

Conclusion:

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

Abbreviations

CC, Cervical cancer

FIGO, Federation of Gynecology and Obstetrics

DEGs, differentially expressed genes

GEO, Gene Expression Omnibus

OS, overall survival

TCGA, The Cancer Genome Atlas

GO, Gene Ontology

KEGG, Kyoto Encyclopedia of Genes and Genomes

GSEA, Gene Set Enrichment Analysis

TIMER, Tumor Immune Estimation Resource

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the TCGA (https://portal.gdc.cancer.gov/), and GEO website (https://www.ncbi.nlm.nih.gov/geo/).

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no funding for this work.

Authors' contributions

DC conceived the experiments, LY and GL downloaded and analyzed the data, GS prepared the figures and tables, LY prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
2. Bhatla N, Aoki D, Sharma DN, Sankaranarayanan R. 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. Nature Reviews Clinical Oncology. 2018;15:81–94.
4. Kamps R, Brandao RD, Bosch BJ, Paulussen AD, Xanthoulea S, Blok MJ, Romano A. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. Int J Mol Sci. 2017;18.
5. Siow ZR, De Boer RH, Lindeman GJ, Mann G. B. Spotlight on the utility of the Oncotype DX((R)) breast cancer assay. Int J Womens Health. 2018;10:89–100.
6. Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, Hammond EH, Kuderer NM, Liu MC, Mennel RG, Van Poznak C, Bast RC, Hayes DF. and American Society of Clinical O. 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, Buys SS, Farmer M, Friedman S, Garber JE, Kauff ND, Khan S, Klein C, Kohlmann W, Kurian A, Litton JK, Madlensky L, Merajver SD, Offit K, Pal T, Reiser G, Shannon KM, Swisher E, Vinayak S, Voian NC, Weitzel JN, Wick MJ, Wiesner GL, Dwyer M, Darlow S 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, Penault-Llorca F, Poortmans P, Rutgers E, Zackrisson S, Cardoso F, Committee EG. 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, Thomssen C. St. 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, Pandor A, Harnan S, Evans P, Wyld L. 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(R)) for colon cancer prognosis. Nat Rev Clin Oncol. 2011;8:131–3.
12. Kopetz S, Tabernero J, Rosenberg R, Jiang ZQ, Moreno V, Bachleitner-Hofmann T, Lanza G, Stork-Sloots L, Maru D, Simon I, Capella G, Salazar R. 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, Roepman P, Snel M, Glas AM, Schuster T, Keller G, Zeestraten E, Goossens I, Janssen KP. Friess H. and Rosenberg R. Independent validation of a prognostic genomic signature (ColoPrint) for patients with stage II colon cancer. Ann Surg. 2013;257:1053–8.
14. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S, Liu JS, Liu XS. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. 2016;17:174.
15. Bahrami A, Hasanzadeh M, Shahidsales S, Farazestanian M, Hassanian SM, Moetamani Ahmadi M, Maftouh M, Gharib M, Yousefi Z, Kadkhodayan S, Ferns GA, Avan A. Genetic susceptibility in cervical cancer: From bench to bedside. J Cell Physiol. 2018;233:1929–39.
16. Burki TK. Novel mutations in cervical cancer. The Lancet Oncology. 2017;18.
17. Yamauchi M, Sricholpech M. Lysine post-translational modifications of collagen. Essays Biochem. 2012;52:113–33.
18. Eisinger-Mathason TS, Zhang M, Qiu Q, Skuli N, Nakazawa MS, Karakasheva T, Mucaj V, Shay JE, Stangenberg L, Sadri N, Pure E, Yoon SS, Kirsch DG, Simon M. C. Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. Cancer Discov. 2013;3:1190–205.
19. Gilkes DM, Bajpai S, Wong CC, Chaturvedi P, Hubbi ME, Wirtz D, Semenza G. L. Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. Mol Cancer Res. 2013;11:456–66.
20. Du H, Chen Y, Hou X, Huang Y, Wei X, Yu X, Feng S, Wu Y, Zhan M, Shi X, Lin S, Lu L, Yuan S. and Sun L. PLOD2 regulated by transcription factor FOXA1 promotes metastasis in NSCLC. Cell Death Dis. 2017;8:e3143.
21. Chen Y, Terajima M, Yang Y, Sun L, Ahn YH, Pankova D, Puperi DS, Watanabe T, Kim MP, Blackmon SH, Rodriguez J, Liu H, Behrens C, Wistuba II, Minelli R, Scott KL, Sanchez-Adams J, Guilak F, Pati D, Thilaganathan N, Burns AR, Creighton CJ, Martinez ED, Zal T, Grande-Allen KJ, Yamauchi M. and Kurie JM. Lysyl hydroxylase 2 induces a collagen cross-link switch in tumor stroma. J Clin Invest. 2015;125:1147–62.

22. Xu F, Zhang J, Hu G, Liu L, Liang W. 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.
23. Zhao H, Chen Q, Alam A, Cui J, Suen KC, Soo AP, Eguchi S, Gu J, Ma D. The role of osteopontin in the progression of solid organ tumour. Cell Death Dis. 2018;9:356.
24. Zeng B, Zhou M, Wu H, Xiong Z. SPP1 promotes ovarian cancer progression via Integrin beta1/FAK/AKT signaling pathway. Onco Targets Ther. 2018;11:1333–43.
25. Liu J, Xu K, Chase M, Ji Y, Logan JK, Buchsbaum RJ. Tiam1-regulated osteopontin in senescent fibroblasts contributes to the migration and invasion of associated epithelial cells. J Cell Sci. 2012;125:376–86.
26. Chen X, Xiong D, Ye L, Yang H, Mei S, Wu J, Chen S. and Mi R. SPP1 inhibition improves the cisplatin chemo-sensitivity of cervical cancer cell lines. Cancer Chemother Pharmacol. 2019;83:603–13.
27. Shen J, Lin S, Liu L, Wang C. Ribonuclease H2 Subunit A impacts invasiveness and chemoresistance resulting in poor survivability of breast cancer in ER dependent manner. American journal of translational research. 2020;12:2281–94.
28. Dai B, Zhang P, Zhang Y, Pan C, Meng G, Xiao X, Wu Z, Jia W, Zhang J, Zhang L. RNaseH2A is involved in human gliomagenesis through the regulation of cell proliferation and apoptosis. Oncol Rep. 2016;36:173–80.
29. Chang H, Dong T, Ma X, Zhang T, Chen Z, Yang Z, Zhang Y. Spondin 1 promotes metastatic progression through Fak and Src dependent pathway in human osteosarcoma. Biochem Biophys Res Commun. 2015;464:45–50.
30. Shah W, Yan X, Jing L, Zhou Y, Chen H, Wang Y. 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.
31. Wang J, Li Z, Gao A, Wen Q, Sun Y. The prognostic landscape of tumor-infiltrating immune cells in cervical cancer. Biomed Pharmacother. 2019;120:109444.
32. Wisdom AJ, Hong CS, Lin AJ, Xiang Y, Cooper DE, Zhang J, Xu ES, Kuo HC, Mowery YM, Carpenter DJ, Kadakia KT, Himes JE, Luo L, Ma Y, Williams N, Cardona DM, Haldar M, Diao Y, Markovina S, Schwarz JK, Kirsch DG. Neutrophils promote tumor resistance to radiation therapy. Proc Natl Acad Sci U S A. 2019;116:18584–9.
33. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25.

Figures

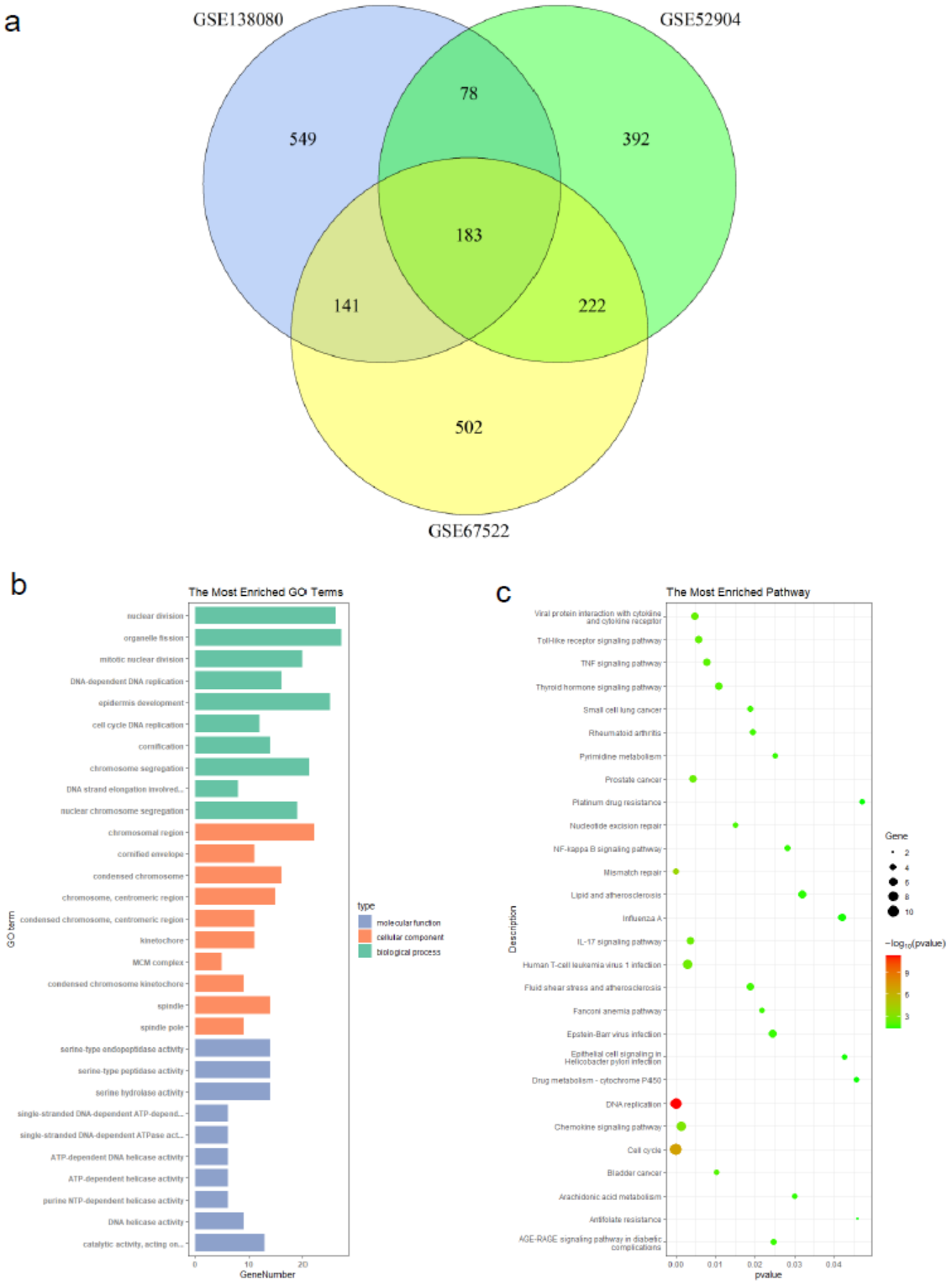

Figure 1

Functional enrichment analysis of the DEGs. a. A total of 183 DEGs were identified after integrated analysis of three GEO datasets. b. GO Functional enrichment analysis of the DEGs. c. KEGG Functional enrichment analysis of the DEGs.

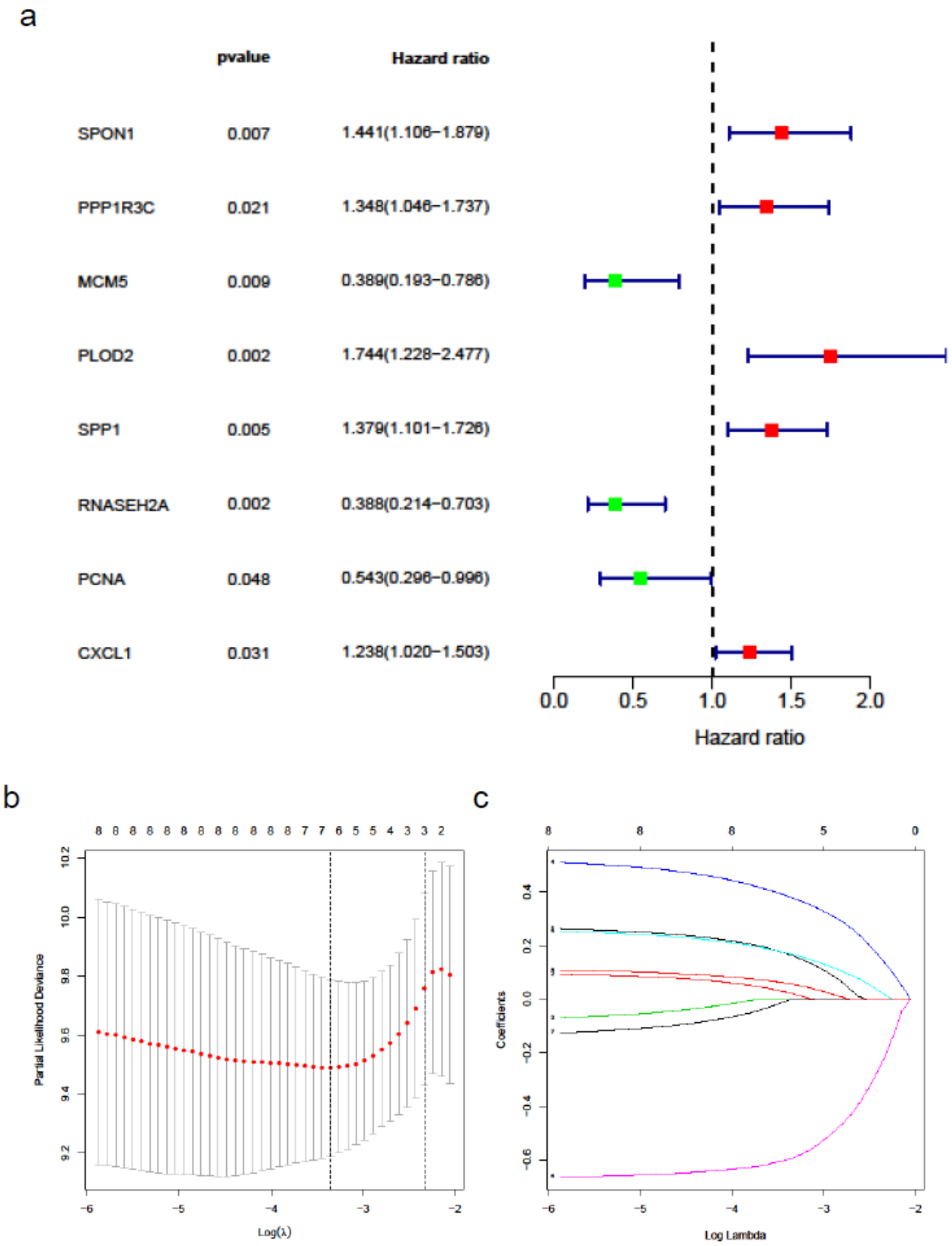

Figure 2

Evaluation of DEGs With Prognostic Value. a. DEGs associated with overall survival using univariate Cox regression model. b, c. Lasso analysis of the prognostic DEGs in cervical cancer.

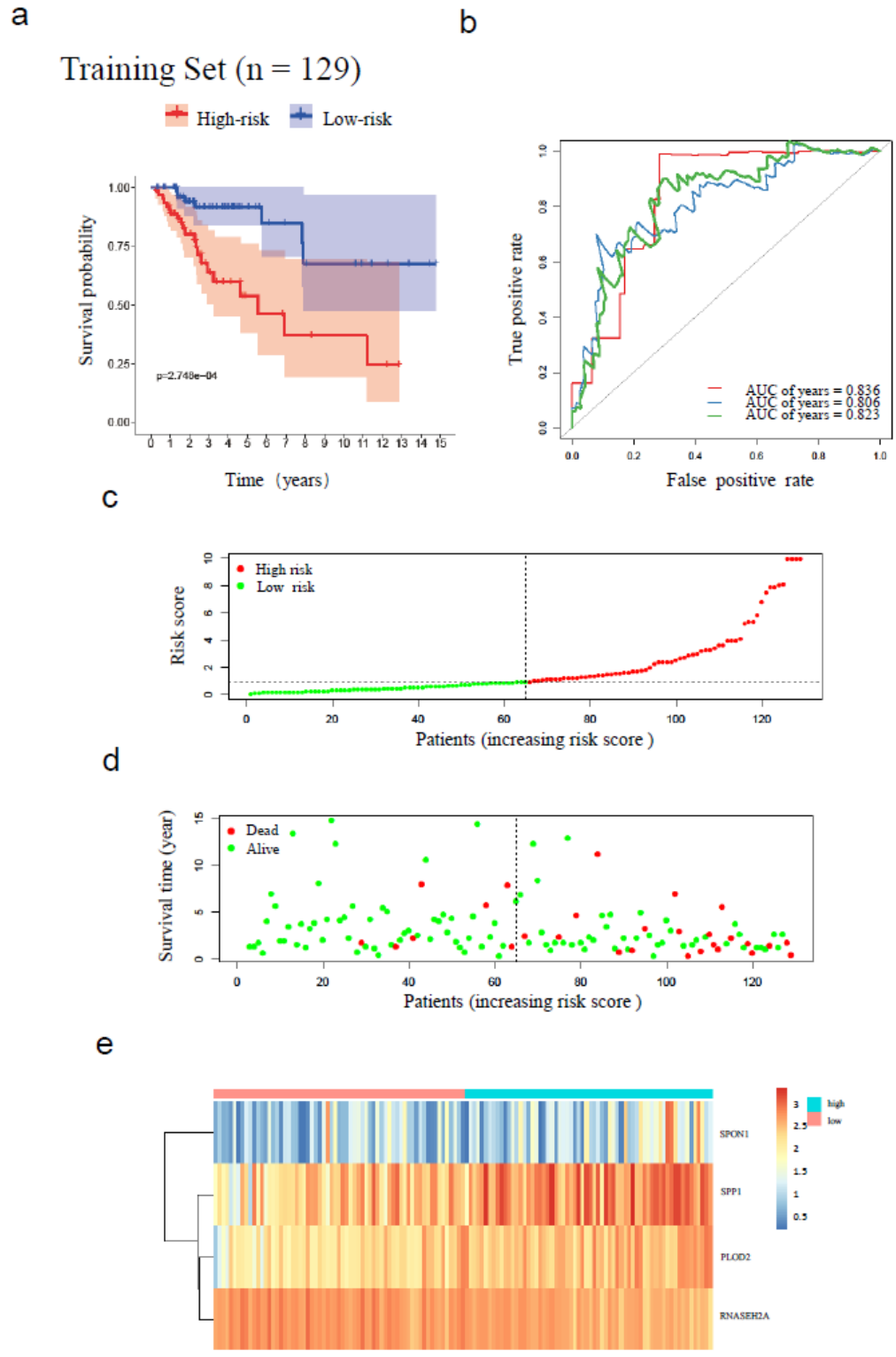

Figure 3

Development of risk score based on the four-gene signature of patients with CC. a. Kaplan-Meier survival curves of the CC samples based on the four-gene signature in training set. b. AUC of time-dependent ROC curves verified the prognostic performance of the risk scores. c d, e. Distribution of the risk score and the associated survival data and mRNA expression heat map in Training Set.

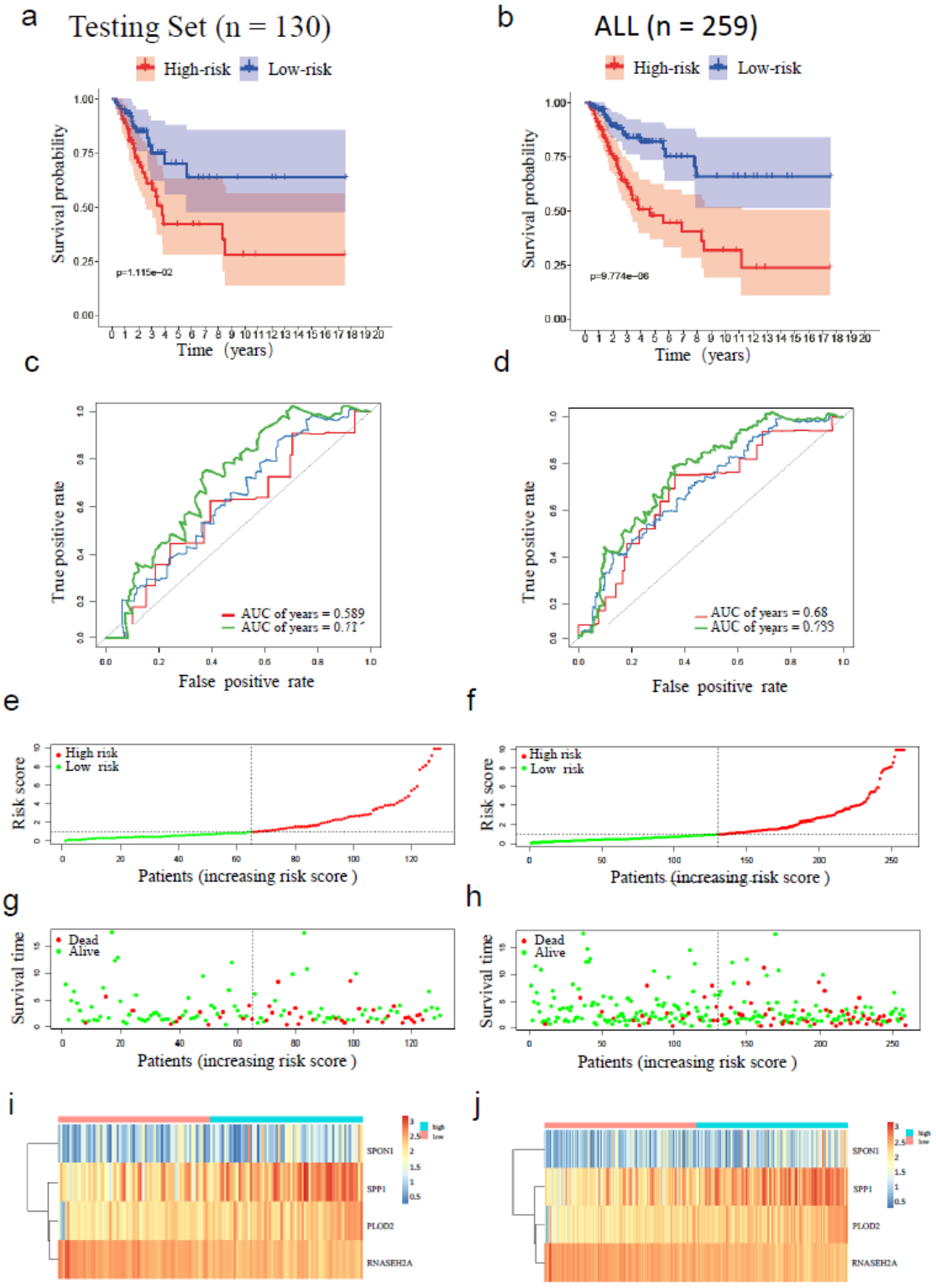

Figure 4

validation of the four-gene signature in patients with CC in testing set and entire TCGA datasets. a, b. Kaplan-Meier survival curves of the CC samples based on the four-gene signature in Testing Set and entire set. c, d. AUC of time-dependent ROC curves verified the prognostic performance of the risk scores. e, g, i. Distribution of the risk scores and the associated survival data and mRNA expression heat map in

testing set. f, h, j. Distribution of the risk score and the associated survival data and mRNA expression heat map in entire TCGA dataset.

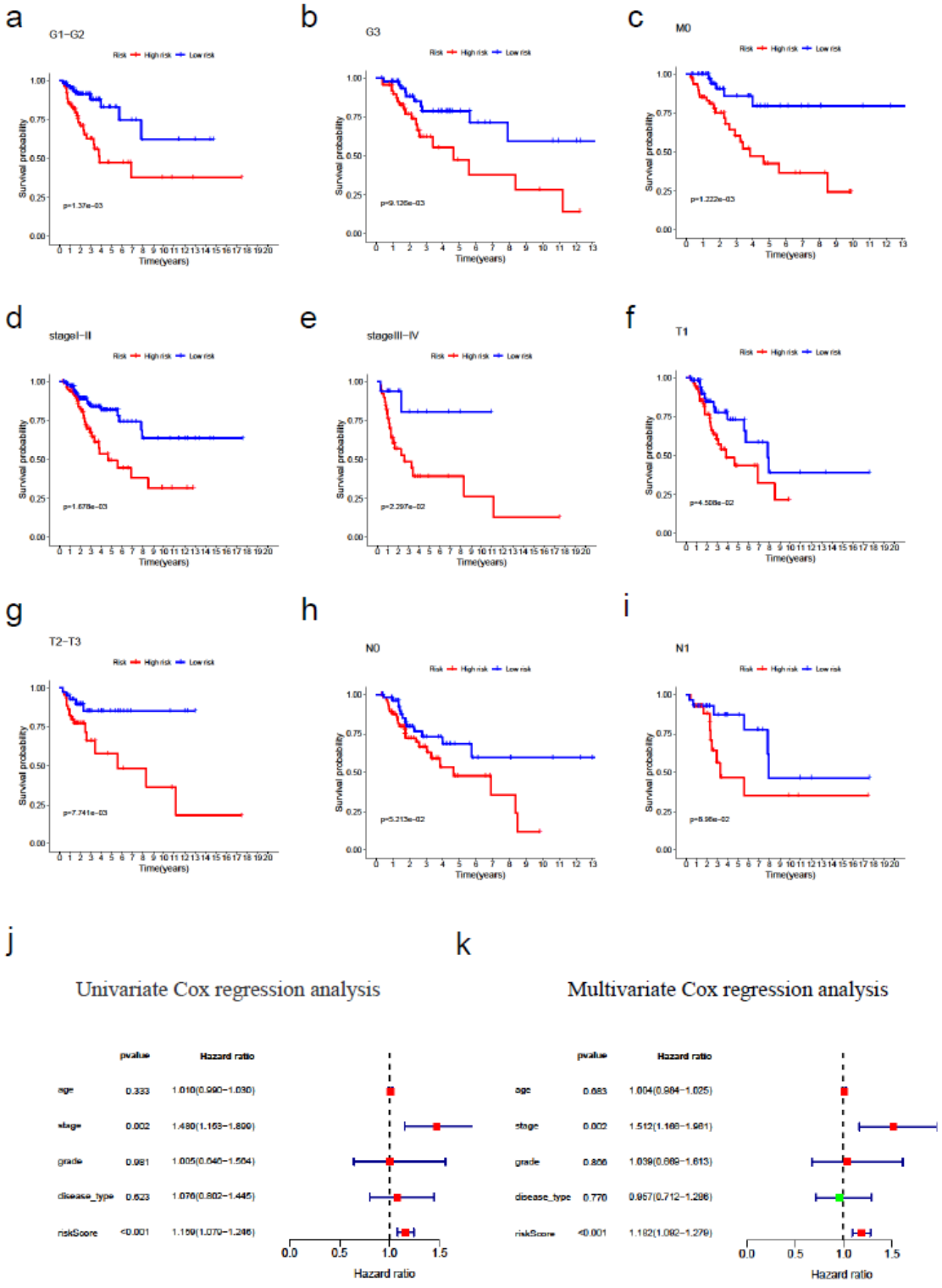

Figure 5

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, k. The univariate (j) and multivariate (k) Cox regression analysis of risk score, age, TNM stage, grade, disease type.

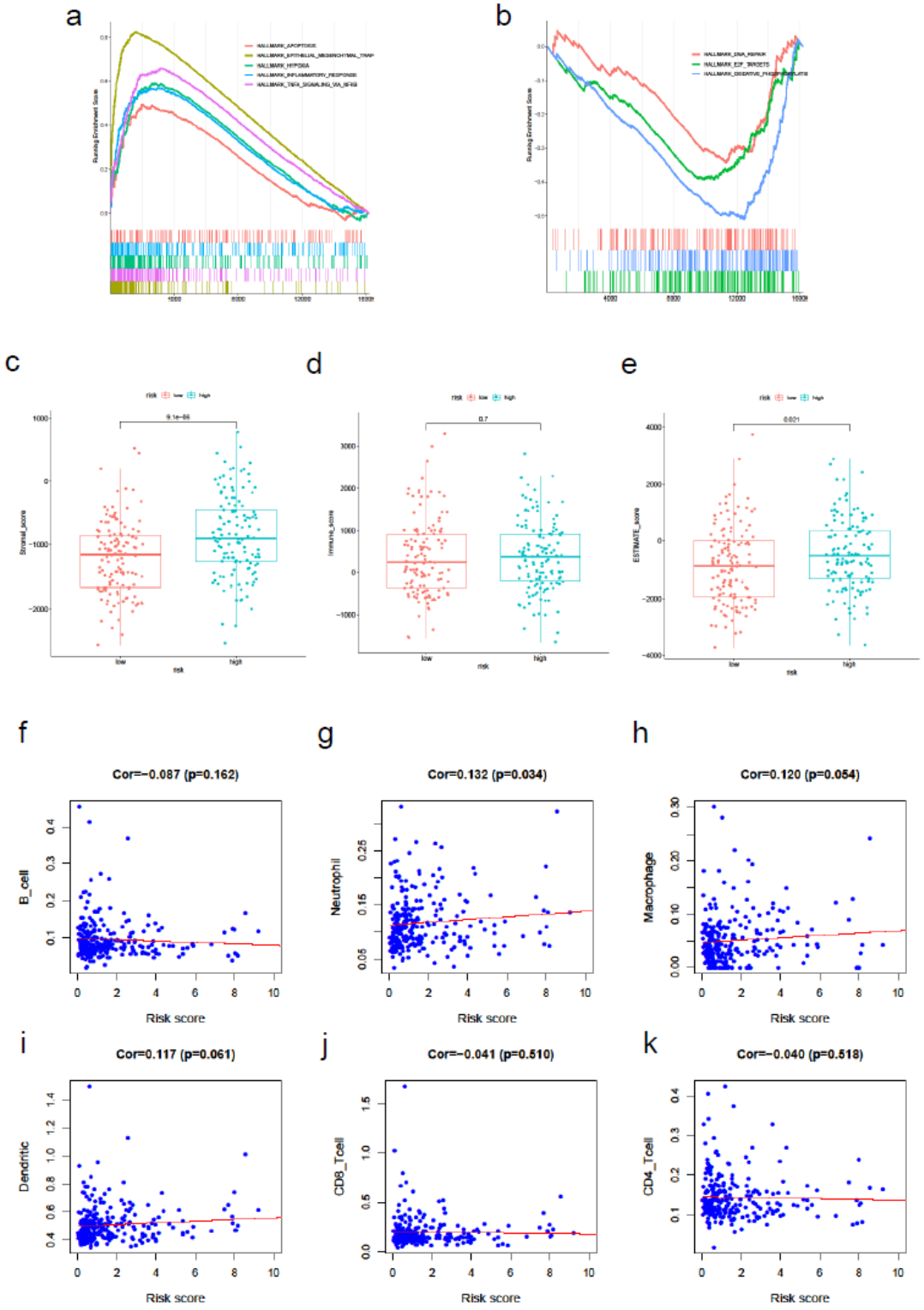

Figure 6

Correlation of the risk score with infiltrative immune cells. a, b. GSEA analysis between high and low risk groups. c, d, e. Boxplots show the immune scores, stromal scores and ESTIMATE scores between high

and low risk groups. f-k. Correlation between the four-gene prognostic signature for cervical cancer and the infiltration of immune cell subtypes.

Supplementary Files

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

- FigureS1.pdf
- FigureS2.pdf