

Identifying a cervical cancer survival signature based on mRNA expression and genome-wide copy number variations

Nan Li^{1,2,3,*}, Kai Yu^{4,*}, Zhong Lin⁵ and Dingyuan Zeng⁶ 

¹Liuzhou Maternity and Child Healthcare Hospital, Liuzhou 545001, China; ²Affiliated Maternity Hospital and Affiliated Children's Hospital of Guangxi University of Science and Technology, Liuzhou 545001, China; ³Guangxi Health Commission Key Laboratory of Birth Cohort Study in Pregnant Women of Advanced Age, Liuzhou 545001, China; ⁴College of Animal Science and Technology, Guangxi University, Nanning 530004, China; ⁵The Guangxi Zhuang Autonomous Region Reproductive Hospital, Nanning, Guangxi Zhuang Autonomous Region 530021, China; ⁶The Department of Obstetrics and Gynecology, Liuzhou Maternity and Child Health Care Hospital, Liuzhou 545001, China

Corresponding authors: Zhong Lin. Email: uters@126.com; Dingyuan Zeng. Email: zdy13607723199@163.com

*These authors contributed equally to this article.

Impact statement

We explored a new prognostic signature with six genes based on CNV data and mRNA data. The robustness of the six-gene signature in patient risk stratification, that is, categorization of patients into low- and high-risk groups, was confirmed. Moreover, risk score and immune score were closely associated. These six genes could be employed as a novel prognostic signature in predicting cervical cancer prognosis.

Abstract

Cervical cancer mortality is the second highest in gynecological cancers. This study developed a new model based on copy number variation data and mRNA data for overall survival prediction of cervical cancer. Differentially expressed genes from The Cancer Genome Atlas dataset detected by univariate Cox regression analysis were further simplified to six by least absolute shrinkage and selection operator (Lasso) and stepwise Akaike information criterion (stepAIC). The study developed a six-gene signature, which was further verified in independent dataset. Association between immune infiltration and risk score was investigated by immune score. The relation between the signature and functional pathways was examined

by gene set enrichment analysis. Ninety-nine differentially expressed genes were detected, and *C11orf80*, *FOXP3*, *GSN*, *HCCS*, *PGAM5*, and *RIBC2* were identified as key genes to construct a six-gene signature. The prognostic signature showed a significant correlation with overall survival (hazard ratio, HR = 3.45, 95% confidence interval (CI) = 2.08–5.72, $p < 0.00001$). Immune score showed a negative correlation with the risk score calculated by the signature ($p < 0.05$). Four immune-related pathways were closely associated with risk score ($p < 0.0001$). The six-gene prognostic signature was an effective tool to predict overall survival of cervical cancer. In conclusion, the newly identified six genes may be considered as new drug targets for cervical cancer treatment.

Keywords: Copy number variants, cervical cancer, differentially expressed genes, prognostic signature, risk score, survival analysis, immune score

Experimental Biology and Medicine 2022; 247: 207–220. DOI: 10.1177/15353702211053580

Introduction

Primary effective preventions of cervical cancer such as human papillomavirus (HPV) vaccination have been implemented long ago; cervical cancer incidence still ranks the fourth highest among women.¹ In 2020, about 604,127 new cervical cancer cases were diagnosed all over the world.¹ Recommended therapies have been proposed

by the International Federation of Gynecology and Obstetrics (FIGO) based on the FIGO staging system.² Although lymph node metastasis, tumor size, and invasion are key factors in deciding the recurrence risk of cervical cancer, cervical cancer patients evaluated as having a low risk of recurrence still develop cancer recurrence within two years of first diagnosis.³ Therefore, an effective

prediction system is needed to improve current personalized treatment.

A series of prognostic biomarkers have been discovered for survival prediction of cervical cancer patients, for instance, immune signatures,⁴⁻⁶ a 10-gene prognostic model for aggressive cervical cancer,⁷ biomarkers of long non-coding RNAs,⁸⁻¹⁰ biomarkers of microRNAs,^{11,12} biomarkers of histone genes,¹³ and some other prognostic genes or pathways.¹⁴⁻¹⁶ However, to date, a CNVs-based prognostic signature for cervical cancer has not been explored. According to previous studies, CNVs are closely associated with tumorigenesis in cancers such as lung cancer, leukemia, breast cancer.¹⁷⁻²⁰ According to the correlation between differential gene expression and CNVs, Shao *et al.* explored the relationship between CNVs and its downstream effect, particularly on tumor suppressor genes and oncogenes.²¹ Advancement in DNA array technology has gradually uncovered small duplications or deletions related to cancers. Database of the Catalogue of Somatic Mutations in Cancer (COSMIC) collected 1,179,545 CNVs from previous literature (published August 2018) with tumor samples and genomic data.²²

Considering close relationship between CNVs and cancer development, this study integrated CNV data and mRNA data to identify prognostic genes in cervical cancer. The objective of the present work was to develop a novel effective signature for overall survival prediction and assisting clinical treatments of cervical cancer. CNV and mRNA expression data were mined from available database through bioinformatics, and we demonstrated the

relation between immune response and prognostic signature. In addition, the robustness of the prognostic signature developed in this study was compared with the existing ones.

Materials and methods

Data source and preprocessing

Cervical cancer samples with CNV data, mRNA data, clinical features, and patients' survival data were collected from TCGA database (<https://portal.gdc.cancer.gov/>). GSE44001 profiles with mRNA data were downloaded from Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). We excluded samples without survival data or follow-up information. After preprocessing, 291 samples in TCGA dataset were included, and another 300 samples came from GSE44001 profiles (Supplementary Table S1). The workflow of constructing prognostic model of cervical cancer is shown in Figure 1.

Identifying DEGS from TCGA dataset

CNVs were converted to genes by BEDTools,²³ and genes with a CNV of $|\text{segment_mean}| > 0.2$ were retained for further analysis. After conducting *Chi*-square test, differential CNVs from tumor samples and normal samples were detected, and differential genes were outputted ($p < 0.05$). According to false discovery rate (FDR) < 0.01 and $|\log_2(\text{fold change, FC})| > 1$, DEGs were screened from mRNA data using Limma R package.^{24,25} Intersected DEGs between CNV and mRNA data were collected for further analysis.

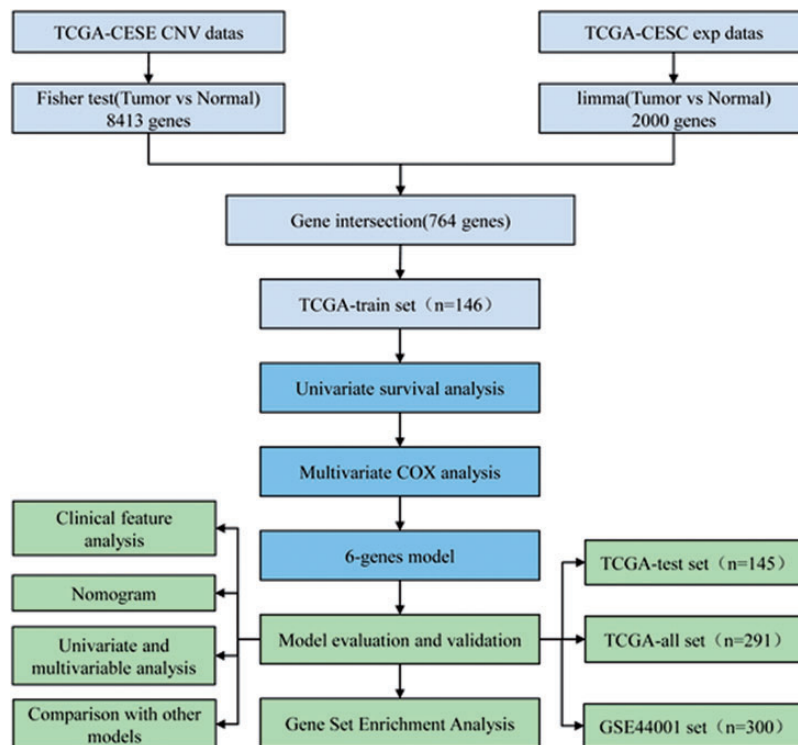


Figure 1. The flow diagram of developing a prognostic model of cervical cancer. (A color version of this figure is available in the online journal.) TCGA: The Cancer Genome Atlas; CNV: copy number variation.

Construction of prognostic model

TCGA dataset was randomly classified into training group and test group with a proportion of 1:1 using sample function in R with random seed 2119. The data showed no significant difference between two groups except for Grade item. GSE44001 served as a validation dataset. Univariate Cox regression analysis was conducted to analyze DEGs in the training dataset. For the purpose of model simplification, Lasso together with five-fold cross-validation and AIC were used to reduce variations and to construct the optimized model.^{26,27} Risk score was defined as coefficient 1*mRNA expression of gene 1 + coefficient 2*mRNA expression of gene 2 + ...coefficient n*mRNA expression of gene n. Then, the risk score was converted to z-score. The samples were classified into two risk groups (low-risk and high-risk) with z-score = 0 as the cut-off. The prognostic model was validated using test dataset and validation dataset. Group overall survival was assessed by plotting Kaplan-Meier survival curve. Survival analysis was conducted with log-rank test. $p < 0.05$ was defined as significant.

Calculation of immune score

Immune score was calculated by Estimation of STromal and Immune cells in MAlignant Tumors using Expression data (ESTIMATE) in the R package.²⁸ ESTIMATE could deduce the fraction of stromal and immune cells in cancer samples based on gene expression signature. Additionally, Microenvironment Cell Populations-counter (MCP-counter) was employed to quantify abundance score of various immune-related cells, including B lymphocytes, CD3⁺ T cell, neutrophils, CD8⁺ T cell, endothelial cell, NK cells, fibroblast, monocytic lineage, cytotoxic lymphocyte, and myeloid dendritic cell.²⁹

Gene enrichment analysis

WebGestalt R package was employed to analyze GO (gene ontology) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of DEGs.^{30,31} Relation between KEGG pathways and risk score was analyzed by GSVA R package.^{32,33} Functional pathways showing a correlation coefficient > 0.3 were seen as correlated with the risk score. Statistical differences in different pathways between two risk groups were analyzed by Wilcoxon test. Significant results were defined with a $p < 0.05$.

Results

Screening DEGs based on CNV and mRNA datasets

In CNV dataset, we screened 8413 genes showing differential CNVs through comparing the CNV of cervical cancer samples with the normal samples using *Chi*-square test ($p < 0.05$). In mRNA dataset, Limma in the R package was used to detect DEGs. A total of 2000 DEGs incorporating 961 down-regulated and 1039 up-regulated genes were filtered with the conditions of $FDR < 0.01$ and $|\log_2 FC| > 1$ (Figure 2(a)). Seven hundred and seventy-four DEGs incorporating 359 down-regulated genes and 405 up-regulated

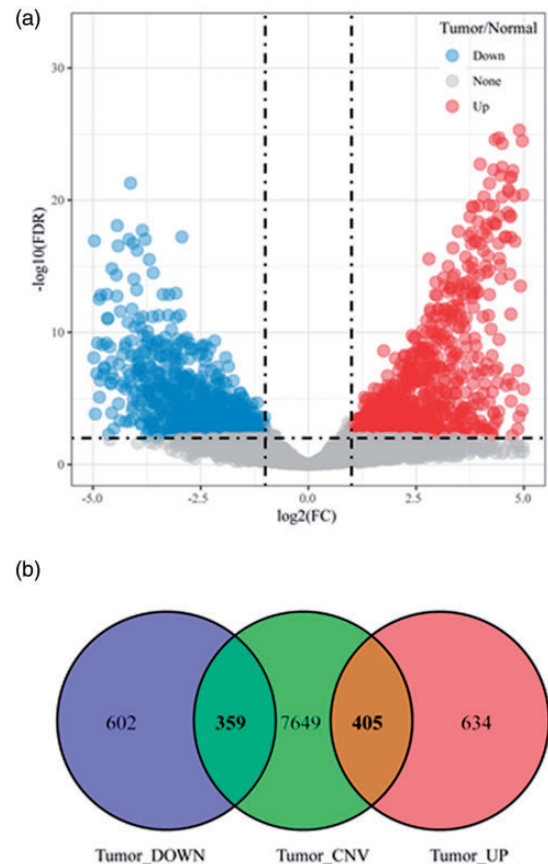


Figure 2. Screened DEGs from mRNA data of TCGA dataset. (a) Volcano plot of DEGs (1039 up-regulated genes and 961 down-regulated genes). Tumor/normal represents the expression ratio of tumor samples and normal samples. Down means down-regulated samples (blue) and up means up-regulated samples (red). None means no significant change of expression (grey). (b) Venn plot within down-regulated genes (violet), up-regulated genes (red), and differential CNVs converted as genes (green). (A color version of this figure is available in the online journal.)

FDR: false discovery rate; FC: fold change; CNV: copy number variants.

genes were identified from genes showing CNVs and mRNA dataset (Figure 2(b)). Furthermore, the enrichment analysis of 764 genes in GO function showed that these genes were related to seven molecular function terms, 148 biological process terms, 48 cellular component terms ($FDR < 0.05$). Among these terms, “condensed chromosome kinetochore,” “DNA-dependent DNA replication,” and “DNA replication origin binding” were the most enriched terms of cellular components, biological processes, and molecular functions, respectively (Figure 3(a) to (c)). In KEGG pathways, seven pathways, including “p53 signaling pathway,” “Oocyte meiosis,” “DNA replication,” “Fanconi anemia pathway,” “Cell cycle,” “One carbon pool by folate,” and “Homologous recombination” were found to be enriched to these genes ($FDR < 0.05$, Figure 3 (d)).

Constructing a prognostic model of cervical cancer

The samples of TCGA dataset were randomly grouped into training dataset (146 samples) and test dataset (145 samples). The results did not show any significant difference between the two groups except for Grade item (G1 to G4,

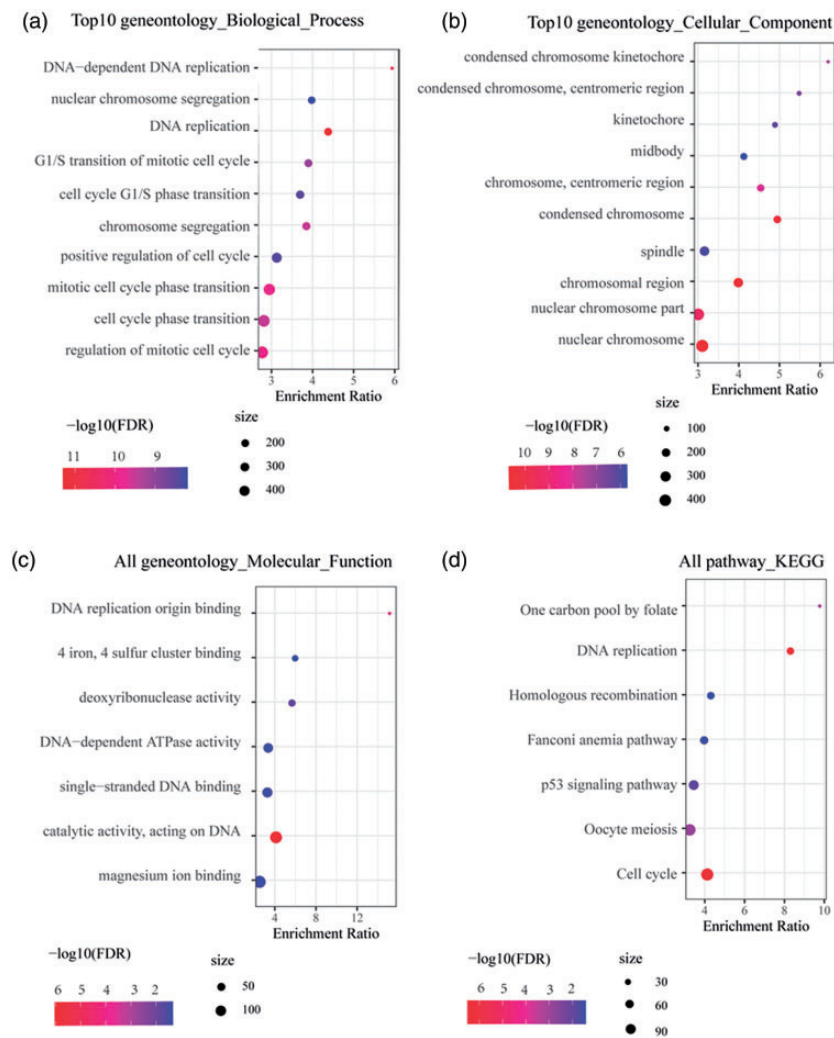


Figure 3. Gene expression enrichment of DEGs in GO function and KEGG pathways. (a) The top 10 biological process terms. (b) The top 10 cellular component terms. (c) Seven molecular function terms. (d) Seven KEGG pathways. Size means counts of genes in different terms. (A color version of this figure is available in the online journal.)
FDR: false discovery rate.

and GX, $p > 0.05$, Supplementary Table S2). The survival coxph function in the R package and univariate Cox regression analysis showed that 764 DEGs and outputted 99 genes were overall survival related ($p < 0.05$). For a further simplification of the gene set with high accuracy, Lasso analysis was used here to reduce gene numbers. Genes showing a coefficient close to zero with the increasing lambda (λ) value were retained to construct a model with the least number of genes or variables (Figure 4(a)). The confidence interval of λ is shown in Figure 4(b). When $\lambda = 0.059$, the model had 12 genes. Then, we used stepAIC R package to further optimize the model. Finally, a total of six genes, namely, *C11orf80*, *FOXP3*, *GSN*, *HCCS*, *PGAM5*, and *RIBC2*, were enrolled in our prognostic model. The formula of risk score was as follows

$$\begin{aligned} \text{Risk Score} = & 1.256 \times C11orf80 - 0.846 \times FOXP3 \\ & - 0.489 \times GSN - 0.553 \times HCCS - 1.081 \\ & * PGAM5 - 0.473 \times RIBC2 \end{aligned}$$

The six genes were distributed in chromosome 9 (*GSN*), chromosome 11 (*C11orf80*), chromosome 12 (*PGAM5*), chromosome 22 (*RIBC2*), and chromosome X (*HCCS* and *FOXP3*) (Supplementary Figure S1). Compared with normal samples, cervical cancer samples had higher expression of the six genes ($p < 0.05$, Supplementary Figure S2). The curve of Kaplan-Meier survival demonstrated that the training set samples were divided into high- and low-risk groups by *FOXP3*, *GSN*, *HCCS* and *RIBC2* ($p < 0.05$), but the samples were not sensitive to the other two genes *C11orf80* and *PGAM5* (Supplementary Figure S3). Sixty-eight out of 146 samples in the training set were grouped into high-risk and the remaining 78 samples were grouped into low-risk group through calculating sample risk score (Figure 5). The mRNA expression of *C11orf80* was higher in high-risk group, whereas the remaining five genes showed a higher expression in low-risk group (z-score of mRNA expression = 0 was selected as the cut-off for defining relatively low and high gene expression, Figure 5(a)). From Kaplan-Meier survival curve, it was shown that low-risk patients tended to live longer

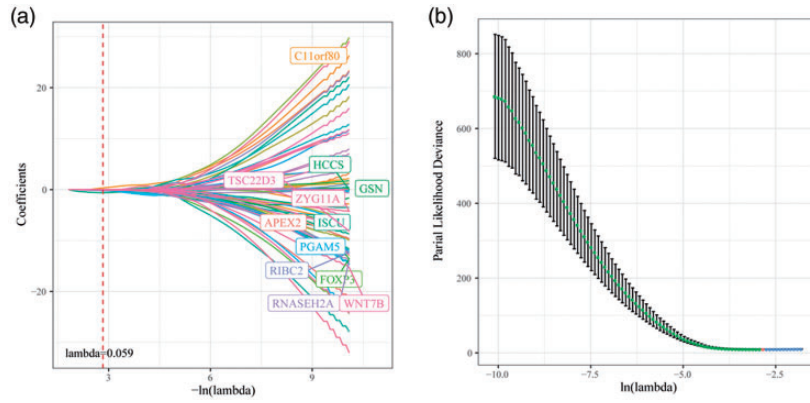


Figure 4. Optimization of prognostic model by Lasso. (a) Coefficient variation following with λ variation of 99 differentially expressed genes in Lasso model. One curve represents one gene. Dashed line means the site when $\lambda = 0.059$. (b) Partial likelihood deviance of λ (green and blue line). Confidential interval was shown in black line. Red dot means the site when $\lambda = 0.059$. (A color version of this figure is available in the online journal.)

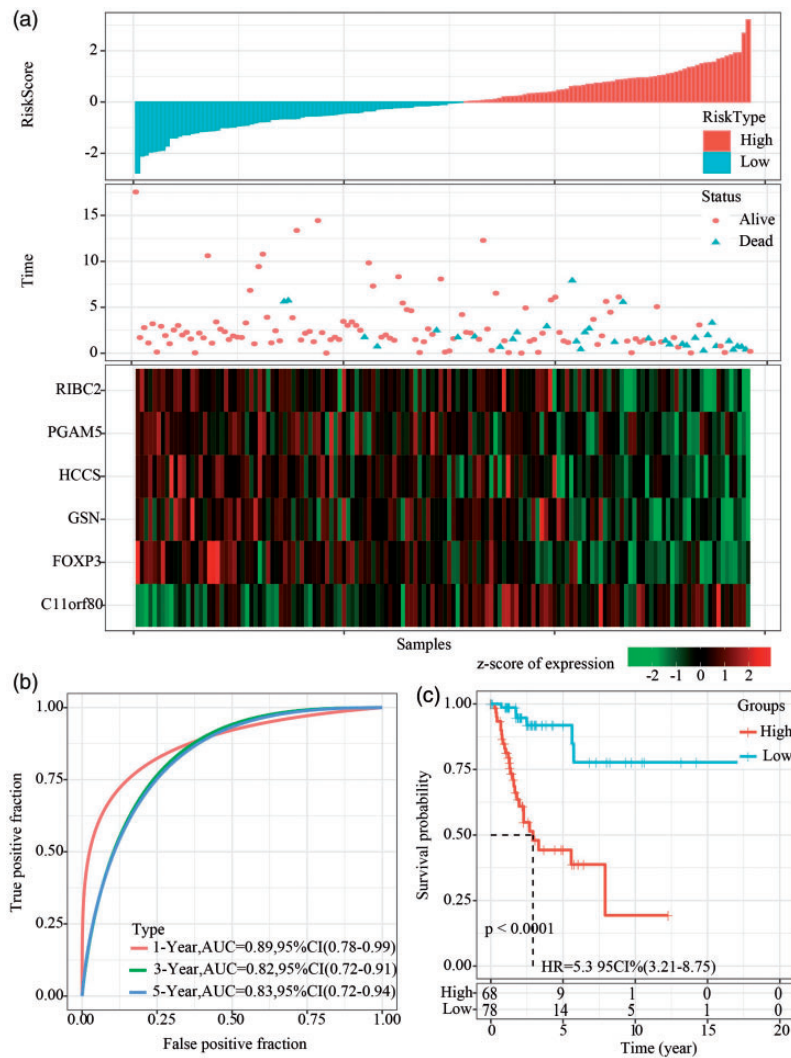


Figure 5. Performance of six-gene prognostic signature in the training dataset. (a) Classification of low-risk group (blue) and high-risk group (red) by risk score which was converted as z-score. The distribution of dead and alive samples in two risk types was shown under risk types. Association between risk types and expression of six genes. (b) ROC curves of one-year, three-year, and five-year survival with an AUC of 0.89, 0.82, and 0.83, respectively. (c) Kaplan-Meier survival curve of low-risk and high-risk groups ($p < 0.0001$, HR = 5.3, 95%CI = 3.21-8.75). (A color version of this figure is available in the online journal.) CI: confidential interval; HR: hazard ratio.

than high-risk patients (HR = 5.3, 95%CI = 3.21–8.75, $p < 0.0001$). Moreover, the robust effectiveness of the prognosis model in overall survival prediction of cervical cancer patients was confirmed by the area under curve (AUC), as AUC for five years, three years, and one year was 0.83, 0.82, and 0.89, respectively (Figure 5(b) and (c)).

Validating the six-gene prognostic signature

Prediction ability of the six-gene prognostic signature was evaluated using the test dataset from TCGA and validation dataset from GSE44001. In the test dataset, 145 samples were grouped into low-risk group (73 samples) and high-risk group (72 samples) ($p < 0.001$, Supplementary Figure S4). The accuracy of the signature in evaluating five-year, three-year, one-year overall survival was shown by an AUC of 0.84, 0.71, and 0.78, respectively (Supplementary Figure S4(b)). Similarly, the signature was robust in classifying a total of 291 samples from TCGA into two risk groups ($p < 0.0001$, Supplementary Figure S5). In the validation

dataset, 144 out of 300 samples were predicted as having a low risk, while the rest 156 samples were evaluated as having a high risk ($p < 0.05$, Figure 6). One-year, three-year, and five-year AUC of overall survival was 0.73, 0.64, and 0.63, respectively (Figure 6(b)).

Correlation analysis between risk score and clinical features

Based on TCGA dataset, we also analyzed correlation between risk score and clinical features. Distribution analysis of different clinical features in the two risk groups showed that the mortality was noticeably higher in high-risk group ($p < 0.05$, Figure 7(a)). Significantly different distribution of N stage, stage I to IV, and grade was detected between two risk groups ($p < 0.05$, Figure 7(c), (e), (g)), while there was no difference in T stage, M stage or age (Figure 7(b), (d), (f)). Survival analysis also showed a higher overall survival of low-risk group with different clinical features (Figure 8). Although no significant

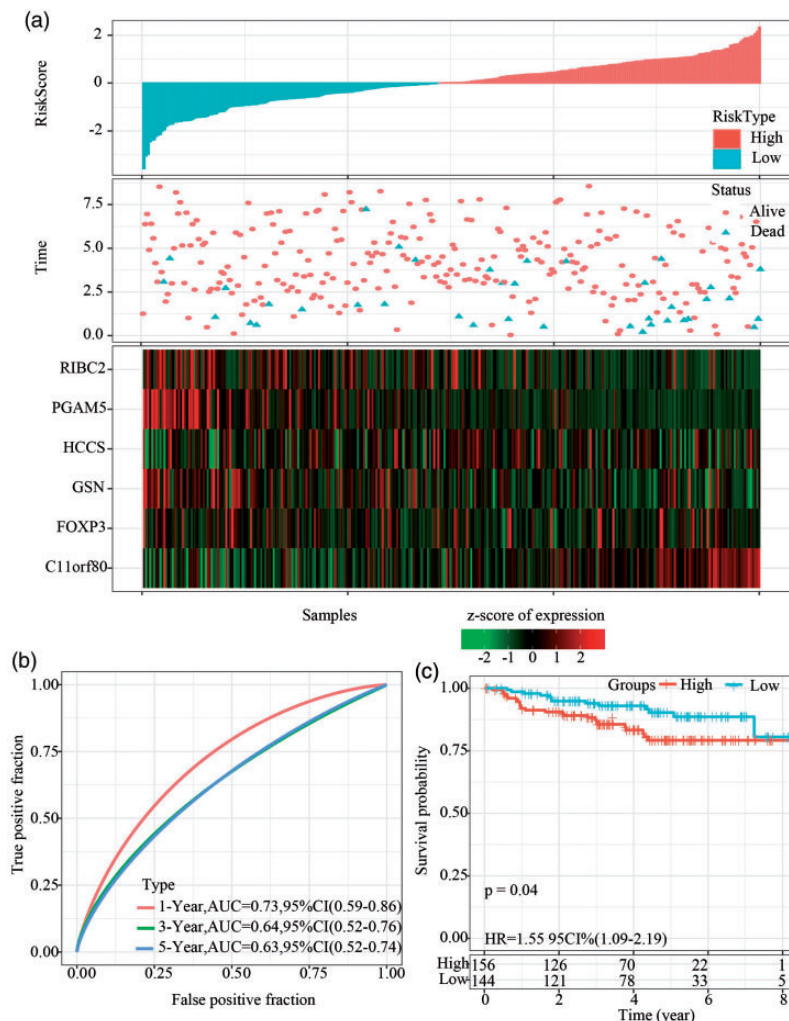


Figure 6. Performance of the six-gene prognostic signature in the validation dataset (GSE44001). (a) Classification of low-risk group (blue) and high-risk group (red) by risk score which was converted as z-score. The distribution of dead and alive samples in two risk types was shown under risk types. Association between risk types and expression of six genes. (b) ROC curves of one-year, three-year, and five-year survival with an AUC of 0.73, 0.64, and 0.63, respectively. (c) Kaplan-Meier survival curve of low-risk and high-risk groups ($p = 0.04$, HR = 1.55, 95%CI = 1.09–2.19). (A color version of this figure is available in the online journal.) CI: confidential interval; HR: hazard ratio.

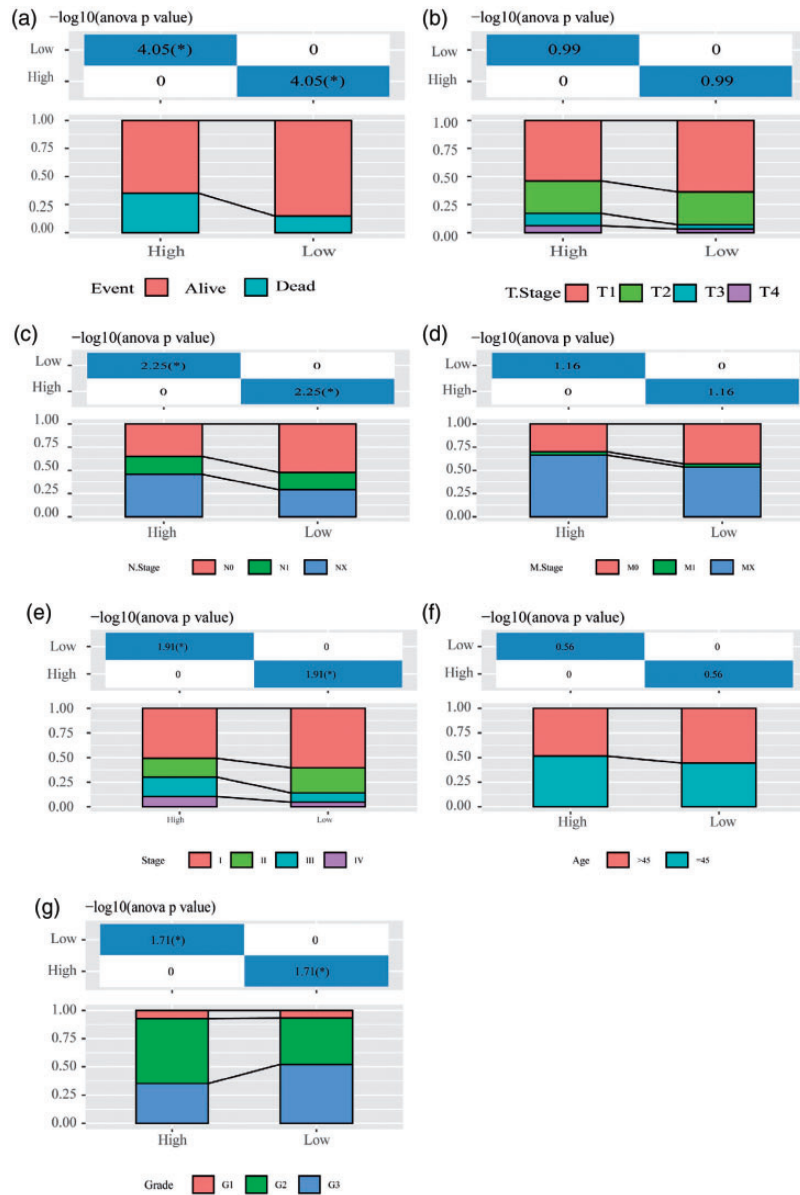


Figure 7. Comparison of different clinical features including survival status (a), T stage (b), N stage (c), M stage (d), stage I to IV (e), age (f), and grade 1 to 3 (g) in low-risk and high-risk groups. Group differences were analyzed by ANOVA. * $p < 0.05$. (A color version of this figure is available in the online journal.)

difference in N0 stage or M1 stage was detected, the risk score was closely associated with patients' overall survival in most clinical features, including stage I to IV, M0 stage, N1 stage, T stage, and grade 1 to 4, age ($p < 0.05$, Figure 8). Univariate Cox regression analysis revealed that risk was closely related to overall survival (95%CI = 2.08–5.72, HR = 3.45, $p < 0.00001$), and that M stage, N stage, T stage were positively related to overall survival (HR > 1, $p < 0.05$) (Figure 9(a)). From the data of multivariable Cox regression analysis, overall survival was significantly related to risk type (HR = 2.52, 95%CI = 0.89–7.11, $p < 0.00001$) (Figure 9(b)).

A nomogram for predicting overall survival

We combined T stage and risk score to develop a nomogram based on TCGA dataset. From Figure 10(a), total points of

risk score and T stage could be calculated, and mortality in five years, three years, and one year is shown. The predicted overall survival was corrected by the actual overall survival (Figure 10(b)). We conducted decision curve analysis (DCA) to evaluate overall survival prediction performance of the nomogram, T stage and risk score, and observed a high prediction accuracy of the nomogram (Figure 10(c)).

Correlation analysis between immune score and risk score

Immune score could act as a prognostic factor; therefore, its relationship with between risk score was assessed. Immune score of all the TCGA samples was calculated with the ESTIMATE in the R package. Figure 11 shows the difference of immune score in low- and high-risk groups. Low-risk

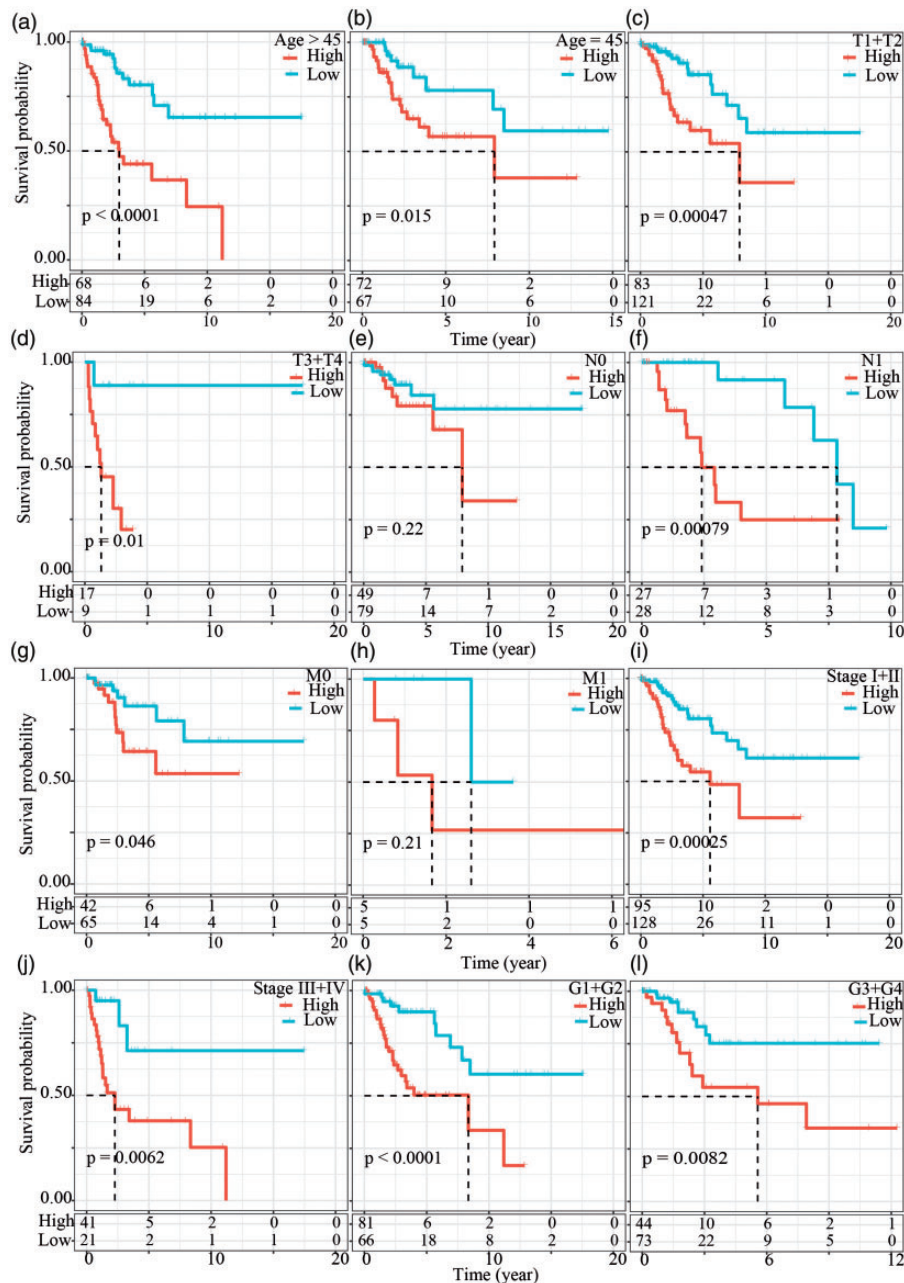


Figure 8. Kaplan-Meier survival curve of different clinical features including age (a–b), T stage (c–d), N stage (e–f), M stage (g–h), stage I to IV (i–j), grade 1 to 4 (k–l) in low-risk and high-risk groups. Log-rank test was performed. $p < 0.05$ was considered to be significant. (A color version of this figure is available in the online journal.)

patients showed higher Immune score, Stromal score, and ESTIMATE score than high-risk ones ($p < 0.001$, Figure 11). Furthermore, we evaluated immune infiltration of tumor through MCP-counter. Consistent with the ESTIMATE results, the immune score calculated by MCP-counter showed a higher enrichment of most immune cells (CD3⁺ T cells, endothelial cells, B lymphocytes, CD8⁺ T cells, NK cells, myeloid dendritic cells, cytotoxic lymphocytes, fibroblasts, and neutrophils, monocytic lineage) in low-risk group ($p < 0.05$, Figure 12). These two methods, for analyzing the distribution of immune cells, revealed a lower immune infiltration in high-risk group than low-risk

group, suggesting that a worse survival was associated with a lower level of immune infiltration.

Identifying functional pathways related to risk score

To explore the relation between functional pathways and risk score, we conducted single sample gene set enrichment analysis (ssGSEA) analysis to calculate ssGSEA score of TCGA dataset samples. An Association study between risk score and ssGSEA score detected 16 pathways in negative correlation with risk score (Figure 13(a)); here, four immune-related pathways, namely, natural killer cell mediated pathway, B cell receptor signaling pathway, T cell

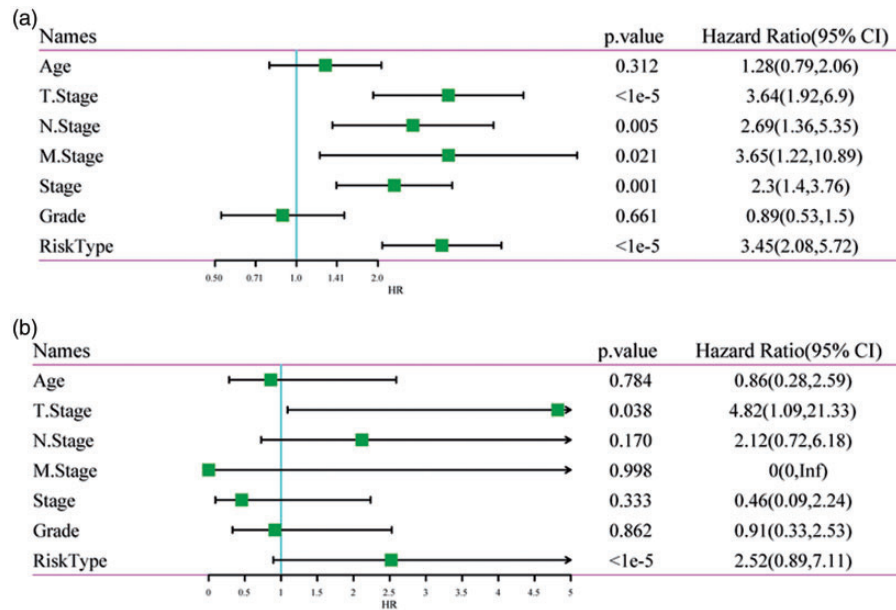


Figure 9. Relation between risk score and clinical features. (a) Univariate Cox regression analysis of clinical features and risk score. (b) Multivariate Cox regression analysis of risk score and clinical features. Green diamond means the media of HR. (A color version of this figure is available in the online journal.) CI: confidential interval; HR: hazard ratio.

receptor signaling pathway, and chemokine signaling pathway, were negatively related to risk score ($p < 0.0001$, Figure 13(b) to (e)).

Comparison with other prognostic signatures

We also compared the current gene signature with previously developed ones. We selected the gene signature for cervical cancer developed in recent two years. Considering that the number of prognostic genes can significantly affect AUC, we included four signatures with similar gene numbers compared with our signature. Therefore, we selected four prognostic signatures of cervical cancer, specifically, a five-gene signature (*GALNTL6*, *ARSE*, *DPAGT1*, *GANAB*, and *FURIN*) of Ju *et al.*,³⁴ a six-gene signature (*ANGPTL4*, *GOT1*, *HSPA5*, *IER3*, *PFKM*, and *PFKFB4*) by Cai *et al.*,³⁵ a five-gene signature (*EPHX2*, *CHAF1B*, *KIAA1524*, *CDC45*, and *RMI2*) by Liu *et al.*,³⁶ an eight-gene signature (*CCDC136*, *ABCG2*, *CYP26A1*, *TNNI3*, *SYT13*, *FOXC2*, *CXCL5*, and *TMEM233*) by Xie *et al.*³⁷ To ensure the comparability, the same methodology was applied to determine the risk score of 291 samples in TCGA dataset using the four signatures. The results showed that all the samples were clearly categorized into low- and high-risk groups by the four signatures ($p < 0.05$, Figure 14). The receiver operating characteristic (ROC) demonstrated that the five-gene signature developed by Liu *et al.* had the highest AUC of 0.76 in predicting one-year survival (95%CI = 0.65–0.88, Figure 14(e)), and that the five-gene signature by Cai *et al.* had the highest AUC of 0.72 in predicting three-year survival (95%CI = 0.64–0.81, Figure 14(c)); moreover, the five-gene signature of Ju *et al.* showed the highest AUC of 0.75 in predicting five-year survival status (95%CI = 0.65–0.84, Figure 14(a)). As the five-gene signature of Ju *et al.* had a higher AUC of five-year survival than our signature, we further calculated C-index of these two signatures. Our

signature was found to have a relatively higher C-index value in all three datasets (training dataset, test dataset, and validation dataset) than Ju *et al.*'s signature, suggesting that our six-gene signature was more effective in predicting overall survival of cervical cancer patients (Supplementary Figure S6). Compared with an AUC of 0.71, 0.72, 0.79 in predicting five-year, three-year, and one-year survival of our six-gene signature, respectively, the overall effectiveness of these four signatures for predicting survival was all less preponderant than the prognostic signature in this study (Supplementary Figure S5).

Discussion

Studies have previously developed prognostic signatures of cervical cancer based on long non-coding RNAs, microRNAs, mRNAs, immune genes, histone genes; however, signatures developed based on CNVs have not been explored. The study established a prognostic model based on CNV data along with mRNA data of cervical cancer samples from GEO and TCGA databases for the characterization and prediction of overall survival of cervical cancer. This was the first study that combined CNV data and gene expression data to identify differential genes. We detected 99 DEGs from TCGA dataset and further simplified the gene set using Lasso and AIC analysis. Most genes in the six-gene signature (*C11orf80*, *FOXP3*, *GSN*, *HCCS*, *PGAM5*, and *RIBC2*) could characterize patients' overall survival, and the prediction system showed a robust performance in the independent dataset (GSE44001). Further analysis between overall survival and the prognostic signature revealed that the risk score was relevant to immune response, with low-risk group showing a high immune score (Figures 11 and 12), indicating that one or several prognostic genes were involved in immune-related process.

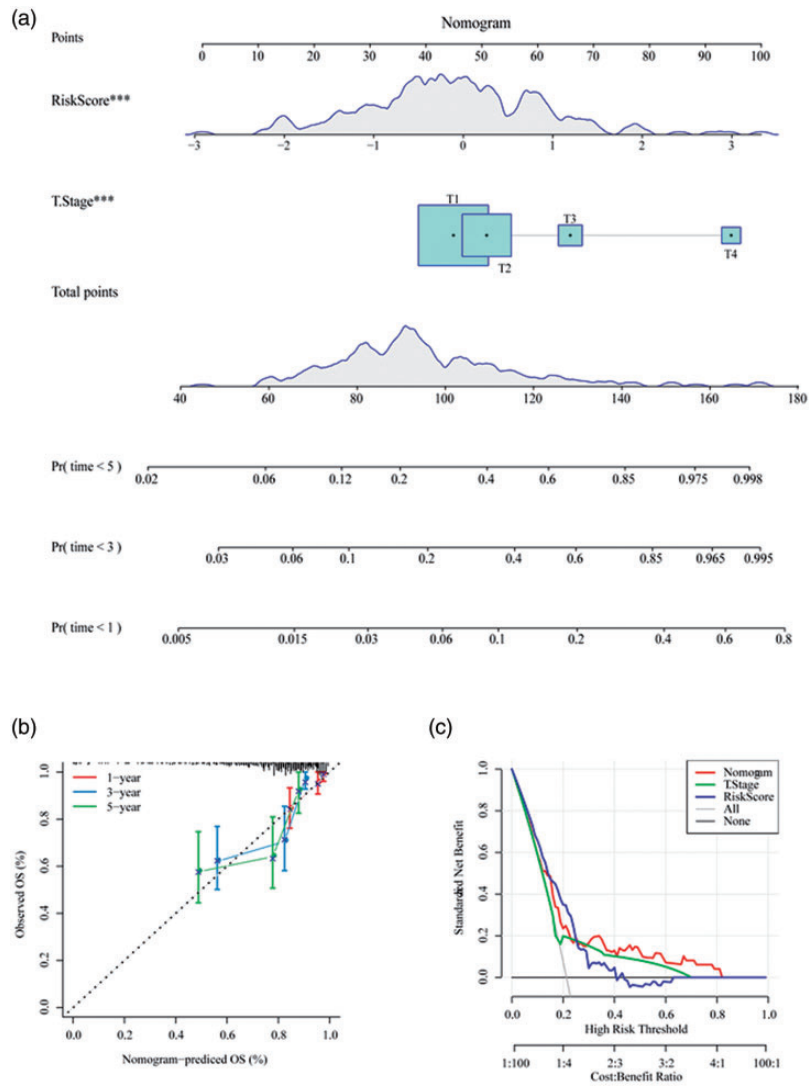


Figure 10. Application of risk score in predicting overall survival. (a) A nomogram to predict one-year, three-year, and five-year death rate. Risk score and T stage were included to calculate total points. (b) The relation between predicted OS by nomogram and observed OS of one year, three years, and five years. (c) DCA plot of nomogram (red), T stage (green), and risk score (blue). (A color version of this figure is available in the online journal.) OS: overall survival.

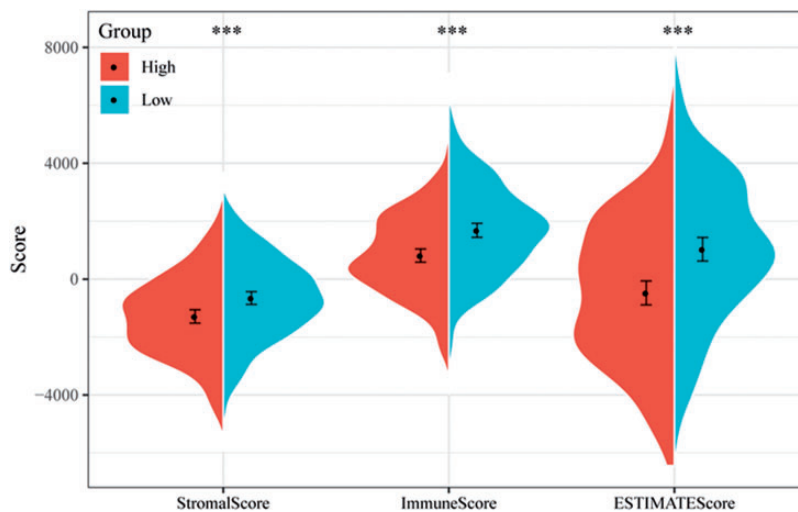


Figure 11. Comparison of immune score between low-risk group (red) and high-risk group (blue) in the TCGA dataset. Student's *t* test was employed to investigate the difference between two groups. ****p* < 0.001. (A color version of this figure is available in the online journal.)

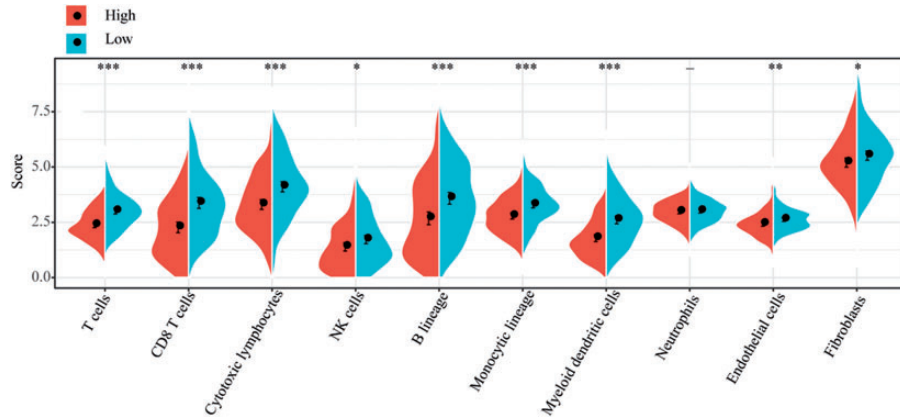


Figure 12. Analyzing the immune score of different immune-related cells by MCP-counter in $CD8^+$ T cells, cytotoxic lymphocytes, $CD3^+$ T cells (T cells), NK cells, myeloid dendritic cells, monocytic lineage, B lymphocytes (B lineage), endothelial cells, fibroblasts, and neutrophils. The difference between two groups was examined by Student's *t* test. $-p > 0.05$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. (A color version of this figure is available in the online journal.)

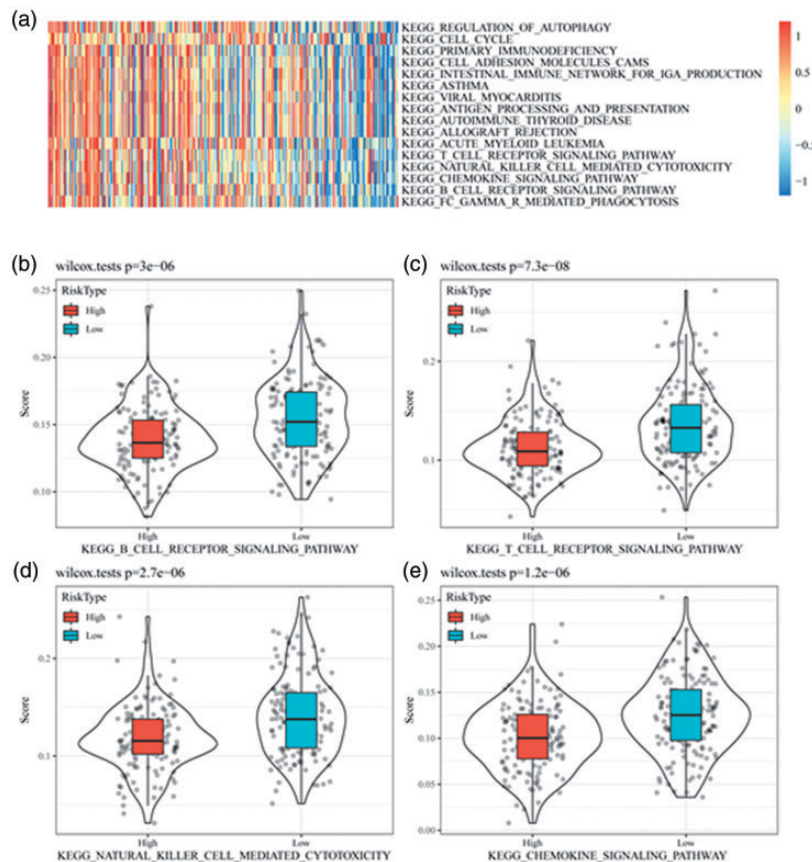


Figure 13. Correlation between risk score and KEGG pathways evaluated by ssGSEA. (a) Sixteen functional pathways negatively related to risk score. (b) Comparison of high- and low-risk groups in four immune-related pathways (natural killer cell-mediated pathway, chemokine signaling pathway, and B cell receptor signaling pathway, and T cell receptor signaling pathway). Wilcoxon test was performed. (A color version of this figure is available in the online journal.)

Of these six prognostic genes, *FOXP3*, *GSN*, and *PGAM5*, were closely involved in the relation between immune score and risk score. As a transcriptional regulator, *FOXP3* functions essentially in activating or suppressing the function of regulatory T-cells through interacting with cytokines and other some transcription factors, NF-kappaB and nuclear factor of activated T cells (NFAT), for instance.³⁸ It has been reported that *FOXP3* is strongly related to many cancers, including colorectal cancer, non-small lung cancer, papillary

and follicular thyroid cancers, breast cancer, prostate cancer, melanoma, gastric cancer, pancreatic cancer, and cervical cancer.^{39,40} In this study, consistent with previous researches, higher expression of *FOXP3* has been detected in tumor samples in comparison with normal ones.^{41,42}

Gelsolin (*GSN*), which participates in immune response, is also involved in macrophage recruitment, phagocytosis, and could exogenously enhance the function of T-lymphocytes.^{43,44} Lower expression of *GSN* in tumor samples was

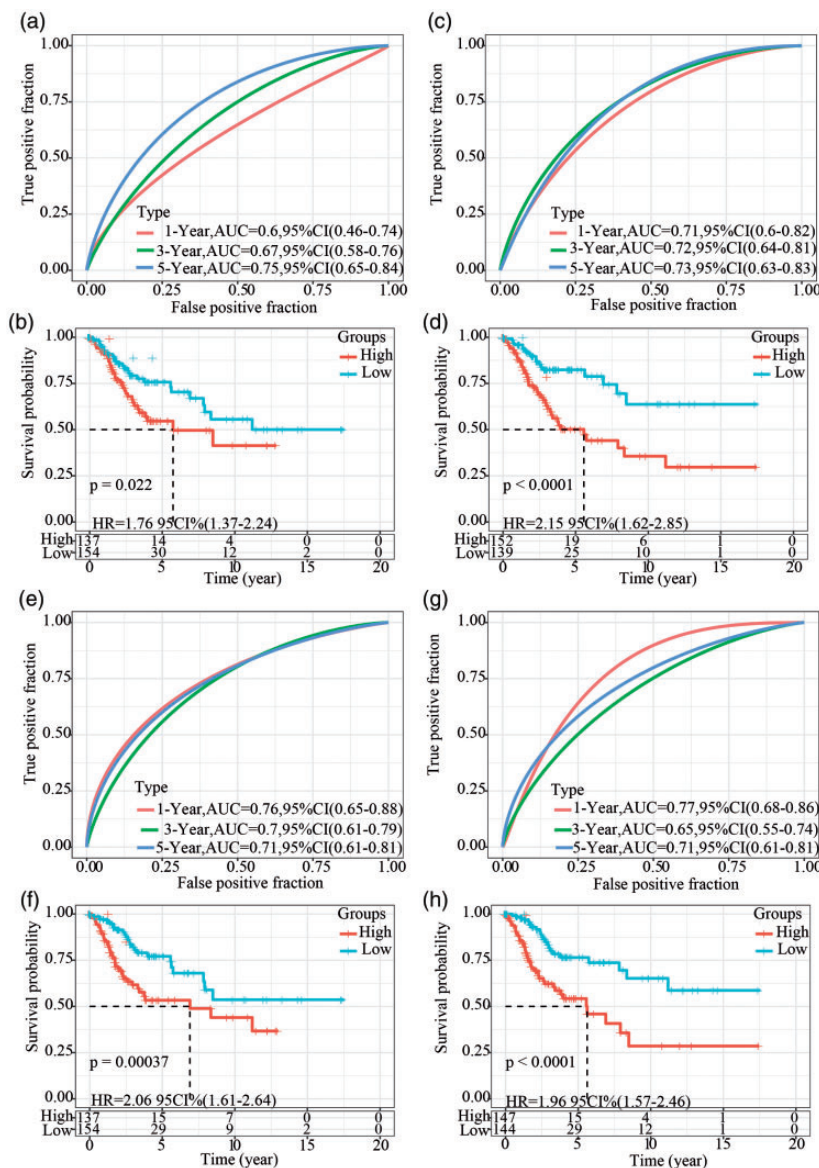


Figure 14. The performance of other prognostic signatures of cervical cancer from literature. (a–b) ROC and Kaplan-Meier survival curve of five-gene signature from Ju *et al.* HR = 1.76, 95%CI = 1.37–2.24, $p = 0.022$. (c–d) ROC and Kaplan-Meier survival curve of six-gene signature from Cai *et al.* HR = 2.15, 95%CI = 1.62–2.85, $p < 0.0001$. (e–f) ROC and Kaplan-Meier survival curve of five-gene signature from Liu *et al.* HR = 2.06, 95%CI = 1.61–2.64, $p = 0.00037$. (g–h) ROC and Kaplan-Meier survival curve of eight-gene signature from Xie *et al.* HR = 1.96, 95%CI = 1.57–2.46, $p < 0.0001$. (A color version of this figure is available in the online journal.) HR: hazard ratio; CI: confidence interval.

detected; interestingly, GSN also had lower expression in high-risk group, indicating a lower immune response in high-risk group and higher tendency of developing tumor progression. Furthermore, it was also found that GSN has tumor-suppressive effect on bladder cancer and breast cancer.^{45,46} Consistently, Abedini *et al.* discovered that over-expressed GSN was significantly related to aggressive gynecological cancers.⁴⁷ It has been demonstrated that NKT cell activation was mediated by the mitochondrial phosphatase phosphoglycerate mutase 5 (PGAM5), and RIPK3-PGAM5-DRP1 signaling axis could mediate the crosslink between host immunity and mitochondrial function in inflammatory diseases.⁴⁸

Interestingly, *RIBC2* has been previously determined as a prognostic gene of cervical cancer through bioinformatics

analysis.^{49,50} Higher expressed *RIBC2* were found to be positively correlated with a more favorable overall survival, which was consistent with our results (Supplementary Figure S3, Figures 8 and 9). However, its function and biological process in tumorigenesis was still unclear. It was indicated that *RIBC2* interacts with *YBX1*, which is associated with cancer proliferation in numerous tissues⁵¹ and also with *TREX1*, aberrant mutations of which are involved in immune-related diseases.⁵² The involvement of *C11orf80* and *HCCS* in cervical cancer or in other cancers has not been reported; thus, their mechanism and function in oncogenesis should be further investigated.

Compared with other prognostic signatures, our six-gene signature based on CNV and mRNA data showed higher AUC and C-index, indicating that it was more

effective in predicting overall survival. Although the methodology used in the present study seemed to be similar to Ju *et al.*, we focused more on CNVs, while Ju *et al.*'s study was based on post-translational modifications (PTMs). CNVs can affect gene expression. In clinical practice, CNVs are more stable and easy to be detected than the variation of gene expression. Detecting the expression of differential genes formed a part of our analysis, but it should be clear that the six-gene signature in our study (C11orf80, FOXP3, GSN, HCCS, PGAM5, and RIBC2) was not overlapped with the five-gene signature developed by Ju *et al.* (GALNTL6, ARSE, DPAGT1, GANAB, and FURIN), as two studies used different strategies when identifying prognostic genes. In addition, our study used three datasets (training dataset, test dataset, and validation dataset) to construct and validate the prognostic model, showing a higher reliability than Ju *et al.*'s study in which only training dataset and test dataset were used. Although screening DEGs from both CNVs and mRNA data can eliminate the effect of false DEGs, using this methodology may omit some certain DEGs at the same time. In addition, as our results were obtained only based on bioinformatics analysis, future validation with more experimental data and clinical patients is also required.

In conclusion, the study developed a novel prognostic signature with six genes according to CNV data and mRNA data. The robustness of the six-gene signature to stratify patients into high- and low-risk groups was confirmed. Moreover, the current findings revealed a close correlation between risk score and immune score, providing new insights into development of new drug targets for treating cervical cancer. In clinical practice, these six genes could be a novel prognostic signature to predict prognosis of cervical cancer patients.

AUTHORS' CONTRIBUTIONS

All authors participated in the design, interpretation of the studies, analysis of the data, and review of the article. ZL and DYZ designed the research; NL performed the statistical analysis; KY analyzed and interpreted the data; ZL drafted the article; DYZ revised the article for important intellectual content.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

This work was supported by Key Research and Development Program of Guangxi (No. Guike AB18126056); Key Research and Development Program of Liuzhou (No. 2018BJ10301); Scientific Research and Technology Development Project of Liuzhou (No. 2018DB20501); and Guangxi Self-Financing Research Program of Guangxi Region Health and Family Planning Commission (No. Z20190019, No. Z20200192, No. Z20210518).

ORCID ID

Dingyuan Zeng  <https://orcid.org/0000-0002-4271-5632>

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**:209–49
- Benedet JL, Bender H, Jones H 3rd, Ngan HY, Pecorelli S. FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO committee on gynecologic oncology. *Int J Gynaecol Obstet* 2000;**70**:209–62
- Peiretti M, Zapardiel I, Zanagnolo V, Landoni F, Morrow CP, Maggioni A. Management of recurrent cervical cancer: a review of the literature. *Surg Oncol* 2012;**21**:e59–66
- Yang S, Wu Y, Deng Y, Zhou L, Yang P, Zheng Y, Zhang D, Zhai Z, Li N, Hao Q, Song D, Kang H, Dai Z. Identification of a prognostic immune signature for cervical cancer to predict survival and response to immune checkpoint inhibitors. *Oncoimmunology* 2019;**8**:e1659094
- Zheng J, Cao B, Zhang X, Niu Z, Tong J. Immune-related four-lncRNA signature for patients with cervical cancer. *BioMed Res Int* 2020;**2020**:3641231
- Chen Y, Lin H, Pi YN, Chen XX, Zhou H, Tian Y, Zhao WD, Xia BR. Development and validation of a prognostic signature based on immune genes in cervical cancer. *Front Oncol* 2021;**11**:616530
- Halle MK, Sodal M, Forsse D, Engerud H, Woie K, Lura NG, Wagner-Larsen KS, Trovik J, Bertelsen BI, Haldorsen IS, Ojesina AI, Krakstad C. A 10-gene prognostic signature points to LIMCH1 and HLA-DQB1 as important players in aggressive cervical cancer disease. *Br J Cancer* 2021;**124**:1690–9
- Mao Y, Dong L, Zheng Y, Dong J, Li X. Prediction of recurrence in cervical cancer using a nine-lncRNA signature. *Front Genet* 2019;**10**:31001325284
- Chen Y, Wang CX, Sun XX, Wang C, Liu TF, Wang DJ. Long non-coding RNA CCHE1 overexpression predicts a poor prognosis for cervical cancer. *Eur Rev Med Pharmacol Sci* 2017;**21**:479–83
- Xia L, Wang H, Cai S, Su X, Shen J, Meng Q, Chen Y, Li L, Yan J, Zhang C, Xu M. Integrated analysis of a competing endogenous RNA network revealing a prognostic signature for cervical cancer. *Front Oncol* 2018;**8**:30237984368
- Liang B, Li Y, Wang T. A three miRNAs signature predicts survival in cervical cancer using bioinformatics analysis. *Sci Rep* 2017;**7**:5624
- How C, Pintilie M, Bruce JP, Hui AB, Clarke BA, Wong P, Yin S, Yan R, Waggott D, Boutros PC, Fyles A, Hedley DW, Hill RP, Milosevic M, Liu FF. Developing a prognostic micro-RNA signature for human cervical carcinoma. *PLoS One* 2015;**10**:e0123946
- Li X, Tian R, Gao H, Yang Y, Williams BRG, Gantier MP, McMillan NAJ, Xu D, Hu Y, Gao Y. Identification of a histone family gene signature for predicting the prognosis of cervical cancer patients. *Sci Rep* 2017;**7**:16495
- Li X, Tian R, Gao H, Yan F, Ying L, Yang Y, Yang P, Gao Y. Identification of significant gene signatures and prognostic biomarkers for patients with cervical cancer by integrated bioinformatic methods. *Technol Cancer Res Treat* 2018;**17**:1533033818767455
- Han GH, Chay DB, Nam S, Cho H, Chung JY, Kim JH. The combination of transient receptor potential vanilloid type 1 (TRPV1) and phosphatase and tension homolog (PTEN) is an effective prognostic biomarker in cervical cancer. *Int J Gynecol Pathol* 2021;**40**:214–23
- Bahrami A, Hasanzadeh M, ShahidSales S, Yousefi Z, Kadkhodayan S, Farazestanian M, Joudi Mashhad M, Gharib M, Mahdi Hassanian S, Avaz A. Clinical significance and prognosis value of Wnt signaling pathway in cervical cancer. *J Cell Biochem* 2017;**118**:3028–33

17. Shlien A, Malkin D. Copy number variations and cancer. *Genome Med* 2009;**1**:62
18. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukheim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, Barletta JA, Borecki IB, Broderick S, Chang AC, Chiang DY, Chirieac LR, Cho J, Fujii Y, Gazdar AF, Giordano T, Greulich H, Hanna M, Johnson BE, Kris MG, Lash A, Lin L, Lindeman N, Mardis ER, McPherson JD, Minna JD, Morgan MB, Nadel M, Orringer MB, Osborne JR, Ozenberger B, Ramos AH, Robinson J, Roth JA, Rusch V, Sasaki H, Shepherd F, Sougnez C, Spitz MR, Tsao MS, Twomey D, Verhaak RG, Weinstock GM, Wheeler DA, Winckler W, Yoshizawa A, Yu S, Zakowski MF, Zhang Q, Beer DG, Wistuba II, Watson MA, Garraway LA, Ladanyi M, Travis WD, Pao W, Rubin MA, Gabriel SB, Gibbs RA, Varmus HE, Wilson RK, Lander ES, Meyerson M. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007;**450**:893–8
19. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X, Pui CH, Relling MV, Evans WE, Shurtleff SA, Downing JR. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007;**446**:758–64
20. Cybulski C, Wokończyk D, Huzarski T, Byrski T, Gronwald J, Górski B, Debniak T, Masojć B, Jakubowska A, van de Wetering T, Narod SA, Lubiński J. A deletion in CHEK2 of 5,395 bp predisposes to breast cancer in Poland. *Breast Cancer Res Treat* 2007;**102**:119–22
21. Shao X, Lv N, Liao J, Long J, Xue R, Ai N, Xu D, Fan X. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med Genet* 2019;**20**:175
22. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P, Harsha B, Hathaway C, Jupe SC, Kok CY, Noble K, Ponting L, Ramshaw CC, Rye CE, Speedy HE, Stefancsik R, Thompson SL, Wang S, Ward S, Campbell PJ, Forbes SA. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 2019;**47**:D94S1–d47
23. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* 2010;**26**:841–2
24. Smyth GK. *Limma: linear models for microarray data*. *Bioinformatics and computational biology solutions using R and Bioconductor*. Berlin: Springer, 2005, pp.397–420
25. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47
26. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc: Series B (Methodological)* 1996;**58**:267–88
27. Sakamoto Y, Ishiguro M, Kitagawa G. *Akaike information criterion statistics*. Vol. 81. Dordrecht, The Netherlands: D Reidel, 1986, p.26853
28. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Commun* 2013;**4**:2612
29. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Erratum to: estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* 2016;**17**:249
30. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 2005;**33**:W741–W48
31. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* 2019;**47**:W199–W205
32. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 2005;**102**:15545–50
33. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform* 2013;**14**:1–15
34. Ju M, Qi A, Bi J, Zhao L, Jiang L, Zhang Q, Wei Q, Guan Q, Li X, Wang L. A five-mRNA signature associated with post-translational modifications can better predict recurrence and survival in cervical cancer. *J Cell Mol Med* 2020;**24**:6283–97
35. Cai L, Hu C, Yu S, Liu L, Yu X, Chen J, Liu X, Lin F, Zhang C, Li W. Identification and validation of a six-gene signature associated with glycolysis to predict the prognosis of patients with cervical cancer. *BMC Cancer* 2020;**20**:1–15
36. Liu J, Nie S, Gao M, Jiang Y, Wan Y, Ma X, Zhou S, Cheng W. Identification of EPHX2 and RMI2 as two novel key genes in cervical squamous cell carcinoma by an integrated bioinformatic analysis. *J Cell Physiol* 2019;**234**:21260–73
37. Xie F, Dong D, Du N, Guo L, Ni W, Yuan H, Zhang N, Jie J, Liu G, Tai G. An 8-gene signature predicts the prognosis of cervical cancer following radiotherapy. *Mol Med Rep* 2019;**20**:2990–3002
38. Bettelli E, Dastrange M, Oukka M. Foxp3 interacts with nuclear factor of activated T cells and NF-kappa B to repress cytokine gene expression and effector functions of T helper cells. *Proc Natl Acad Sci USA* 2005;**102**:5138–43
39. Szyllberg L, Karbownik D, Marszaek A. The role of FOXP3 in human cancers. *Anticancer Res* 2016;**36**:3789–94
40. Jia H, Qi H, Gong Z, Yang S, Ren J, Liu Y, Li MY, Chen GG. The expression of FOXP3 and its role in human cancers. *Biochimica Biophysica Acta Rev Cancer* 2019;**1871**:170–8
41. Zeng C, Yao Y, Jie W, Zhang M, Hu X, Zhao Y, Wang S, Yin J, Song Y. Up-regulation of Foxp3 participates in progression of cervical cancer. *Cancer Immunol, Immunotherapy: CII* 2013;**62**:481–7
42. Zhang H, Zhang S. The expression of Foxp3 and TLR4 in cervical cancer: association with immune escape and clinical pathology. *Arch Gynecol Obstet* 2017;**295**:705–12
43. Irving AT, Wang D, Vasilevski O, Latchoumanin O, Kozer N, Clayton AH, Szczepny A, Morimoto H, Xu D, Williams BR, Sadler AJ. Regulation of actin dynamics by protein kinase R control of gelsolin enforces basal innate immune defense. *Immunity* 2012;**36**:795–806
44. Feldt J, Schicht M, Garreis F, Welss J, Schneider UW, Paulsen F. Structure, regulation and related diseases of the actin-binding protein gelsolin. *Expert Rev Mol Med* 2019;**20**:e7
45. Tanaka M, Müllauer L, Ogiso Y, Fujita H, Moriya S, Furuuchi K, Harabayashi T, Shinohara N, Koyanagi T, Kuzumaki N. Gelsolin: a candidate for suppressor of human bladder cancer. *Cancer Res* 1995;**55**:3228–32
46. Baig RM, Mahjabeen I, Sabir M, Masood N, Ali K, Malik FA, Kayani MA. Mutational spectrum of gelsolin and its down regulation is associated with breast cancer. *Disease Markers* 2013;**34**:71–80
47. Abedini MR, Wang PW, Huang YF, Cao M, Chou CY, Shieh DB, Tsang BK. Cell fate regulation by gelsolin in human gynecologic cancers. *Proc Natl Acad Sci USA* 2014;**111**:14442–7
48. Kang YJ, Bang BR, Han KH, Hong L, Shim EJ, Ma J, Lerner RA, Otsuka M. Regulation of NKT cell-mediated immune responses to tumours and liver inflammation by mitochondrial PGAM5-Drp1 signalling. *Nature Commun* 2015;**6**:8371
49. Meng H, Liu J, Qiu J, Nie S, Jiang Y, Wan Y, Cheng W. Identification of key genes in association with progression and prognosis in cervical squamous cell carcinoma. *DNA Cell Biol* 2020;**39**:848–63
50. Wang H, Li SW, Li W, Cai HB. Elastic net-based identification of a multigene combination predicting the survival of patients with cervical cancer. *Med Sci Monitor*: 2019;**25**:10105–13
51. Tsofack SP, Garand C, Sereduk C, Chow D, Aziz M, Guay D, Yin HH, Lebel M. NONO and RALY proteins are required for YB-1 oxaliplatin induced resistance in colon adenocarcinoma cell lines. *Mol Cancer* 2011;**10**:145
52. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Bertozzi GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamasas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005;**437**:1173–8