

Construction and Validation of a Prognostic Model Based on 11 Lymph Node Metastasis Related Genes for Overall Survival in Endometrial Cancer

Hong Wu (✉ wjq0609@126.com)

Handan Central Hospital <https://orcid.org/0000-0002-8174-2959>

haiqin Feng

Handan Central Hospital

Xiaoli Miao

Handan Central Hospital

jiancai Ma

Handan Central Hospital

Cairu Liu

Handan Central Hospital

Lina Zhang

Handan Central Hospital

Liping Yang

Handan Central Hospital

Primary research

Keywords: lymph node metastasis, risk signature, endometrial cancer, mutation, HMGB3

Posted Date: August 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-734385/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Endometrial cancer (EC) is one of the most common malignant tumor in the female reproductive system. The incidence of lymph node metastasis (LNM) is only about 10% in clinically suspected early-stage EC patients. Discovering prognostic model and effective biomarkers for early diagnosis is important to reduce the mortality rate.

Methods: We downloaded the RNA-sequencing data and clinical information from the TCGA database. Gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were used to analyze the differentially expressed genes (DEGs). A least absolute shrinkage and selection operator (LASSO) regression was conducted to identify the characteristic dimension decrease and distinguish prognosis-related LNM related genes signature. Subsequently, a novel prognosis-related nomogram to predict overall survival (OS). A survival analysis was carried out to explore the individual prognostic significance of the risk model and key gene was validated in vitro.

Results: In total, 89 LRGs were identified. Based on the LASSO Cox regression, 11 genes were selected for the development of a risk evaluation model. The Kaplan–Meier curve indicated that patients in the low-risk group had considerably better OS ($P = 3.583e-08$). The area under the curve (AUC) of this model was 0.718 at 5 years of OS. Then, we developed an OS-associated nomogram that included the risk score and clinicopathological features. The concordance index of the nomogram was 0.769. The survival verification performed in three subgroups from the nomogram demonstrated the validity of the model. The AUC of the nomogram was 0.787 at 5 years OS. Proliferation and metastasis of HMGB3 were explored in EC cell line. Finally, we revealed that the most frequently mutated genes in the low-risk and high-risk groups are PTEN and TP53, respectively.

Conclusions: Our results suggest that LNM plays an important role in the prognosis, and HMGB3 was potential as a biomarker for EC patients. By detecting the mutation of the risk signature, clinicians can accurately treat patients with targeted therapy, thereby improving their survival rate.

Introduction

Endometrial cancer (EC) is one of the most common gynecologic malignancy and also the sixth most common type of cancer in women worldwide. It was estimated that there were 382,069 new cases and 89,929 deaths in 2018 [1]. In China, EC ranks as the second most common cancer of the female genital system, and the five-year overall survival rate is 55.1%[2]. To stratify patients into distinct prognostic groups, EC and other cancers of the female gynecologic system are most commonly staged using guidelines provided by the International Federation of Gynecology and Obstetrics (FIGO) based on findings at either clinical examination and surgical exploration[3]. However, traditional clinical criteria are not enough to predict EC prognosis accurately because the prognosis of patients with same clinicopathological factor varies differently. Therefore, it is imperative to emphasize the molecular

changes that occur during endometrial cancer progression and develop novel predictive biomarkers to accurately estimate patient outcomes.

Machine learning is a form of artificial intelligence that can automatically analyze patterns from sample data, and make corresponding predictions. Due to its accuracy and predictive performance, the machine learning algorithm is used in different fields, including medical diagnostics and prognostic prediction[4]. Recently, in-depth exploration of the public datasets, including TCGA and GEO, shed light on the heterogeneity in prognostic gene signatures of patients with similar clinicopathological feature but distinct molecular features [5]. With advancement in database mining, numerous biomarkers for EC have been identified as promising tools to classify tumors and predict cancer prognosis. For example, Wang et al. used the gene expression data associated with cellular glycolysis from the TCGA database to build a 9-gene prognostic signature, and this signature identified patients with poor prognosis in EC [6]. Zhou et al. integrated tumor mutation burden (TMB) with immune infiltrates to construct a TMB-related signature (GFAP, EDN3, CXCR3, PLXNA4, SST), which had a better appraisal of prognostic and predictive factors and provided guidance of immunotherapy for EC [7].

Lymph node metastasis is a common problem in EC, which seriously affects the prognosis of patients and decreases the 5-year survival. Therefore, it is increasingly urgent to understand mechanisms and find a specific therapeutic method for these patients. There are few reports concentrating on mRNA combination biomarkers for LNM of EC, so differentially expressed mRNA associated with LNM should be the key to the progression of EC.

To systematically investigate the roles of LNM in EC, we analyzed the expression profile of LNM-related genes, as well as normal tissues in The Cancer Genome Atlas (TCGA) database. Furthermore, we established and validated a multiple-LRG-combined expression signature for the prediction EC patient outcomes. Multivariate Cox regression analysis suggested that risk score might be an independent prognostic indicator for the overall survival of EC patients and a prognostic nomogram model was established that could increase the accuracy of OS prediction. Our results also provide important insights into HMGB3 as a promising biomarker for the progression and provide novel perspectives for the therapeutic strategy in EC.

Materials And Methods

Download of gene expression and clinical information

For endometrial cancers, transcriptome RNA sequencing data from FPKM file as well as clinical data were downloaded from The Cancer Genome Atlas (TCGA) database, containing 35 normal samples and 532 tumor samples from patients with endometrial cancer. We then screened the corresponding clinical data and transcriptome data, and excluded the incomplete data. Our study was in accordance with the publication guidelines provided by TCGA.

Identification of DE-LRGs and enrichment analysis

Differentially expressed of transcriptome RNA-sequencing of LRGs (DE-LRGs) in normal and tumor samples were screened through “limma” package of R (version 3.6.1), and the screening standards were based on false discovery rate (FDR) < 0.05 and \log_2 |fold change| (\log_2 FC) > 1. “Clusterprofiler” R package was used to analyze Gene Ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment of DE-LRGs. The results of GO annotation and KEGG pathway analyses were visualized by the “GOplot” package in R platform.

Establishment and validation of the prognostic model based on the DE-LRGs

The least absolute shrinkage and selection operator (LASSO) regression analysis was used to establish the LRGs signature model. This model was used for subsequent evaluation and analysis of risk measures for the patients’ risk values. We categorized these patients into high- and low-risk groups according to the risk values. The risk score for each patient was computed using the following formula:

$$\text{Risk score} = \sum n_i = \sum (\text{Coef}_i * x_i)$$

Coef_i represented the coefficient and x_i represented the expression level of LRGs. Subsequently, the validity of the LRGs signature model was evaluated by analyzing the difference between high- and low-risk groups. In the two subgroups, each patient’s clinicopathological features and gene expression profiles were shown via the “pheatmap” and “survival” R packages. In addition, the Kaplan-Meier curve analysis and receiver operating characteristic (ROC) curves were performed to estimate the sensitivity and specificity of the prognostic signature.

Gene set enrichment analysis (GSEA) analysis

GSEA (<http://www.broadinstitute.org/gsea/index.jsp>) was conducted to investigate the biological pathways of our prognostic gene signature [8]. The collection of hallmark.all.v6.1.symbols.gmt gene sets in Molecular Signatures Database (MSigDB, <http://software.broadinstitute.org/gsea/msigdb/index.jsp>) was taken as the reference gene sets in GSEA software. The standardized *P*-value < 0.05 and FDR < 0.25 was considered to be significantly enriched when the samples were divided into high- and low-risk groups.

Construction and evaluation the of the nomogram

Univariate and multivariate Cox regression analyses were used to determine whether the LRGs signature was independent risk factors for overall survival. Then we established a clinical practical nomogram to predict individual survival probability by the “rms” package of R. To evaluate whether the actual and the predicted survival in the nomogram close to each other, calibration curves for predicting 3-, 5-, and 7-year survival rate were drawn. The 45° line represented the best prediction [9]. The accuracy of the nomogram was then performed with Kaplan–Meier survival analysis and the area under the tdROC curve (AUC).

Harrel's concordance index (C-index) was measured to validate the predictive ability of the nomogram [10].

Experimental validation of HMGB3

Gene Expression Profiling Interactive Analysis (GEPIA) database (<http://gepia.cancer-pku.cn/>) is an interactive website application that recruits the transcriptome data in TCGA and GTEx projects by integrating them in a widely accepted process [11]. We inputted the hub gene into the GEPIA and validated these hub gene.

Endometrial cancer cell line ishikawa was obtained from ATCC (American Type Culture Collection). The cell line were cultured in DMED/F12 medium with 10% FBS according to the provider's instructions and incubated at 37°C with 5% CO₂. HMGB3 siRNA plasmid was designed and synthesized by GenePharma (Shanghai, China). The plasmid was then transfected into the ishikawa cell line. Opti-MEM medium and Lipofectamine RNAiMAX reagent were used for the transfection of siRNA according to the manufacturer's instruction. Then total protein of EC cell line in different groups were extracted and used for western blot as describe in this article [12]. Cell proliferation assays were performed by Cell Counting Kit-8 (Dojindo Molecular Technology, Japan). 10µL CCK8 solution was added into each 96-well plate and incubated for 2 hours. The optical density was measured at 450 nm by microplate spectrophotometer.

By evaluating the effect of HMGB3 on the metastatic and invasive ability of EC cells, we conducted gap closure and transwell invasion assay. Cells were seeded in the 6-well plate and cultured to about 85% density confluence. A straight wound was scratched carefully using 200µl sterile pipette tips. After scratching, cell debris was washed away with PBS. For transwell assay, at a density of 1×10^5 per well, cells were seeded into the upper chamber, and then incubated for 24 hours. Cancer cells were fixed with 4% paraformaldehyde for 30minutes, stained with 0.1% crystal violet for 5 minutes, washed 3 times with PBS, and counted in 6 fields under the microscope. Images were taken at 0h and 24h, and analyzed by Image J software (Rawak Software, Inc. Germany). All of the experiments were repeated for 3 times.

Mutation analysis

The mutation data of EC patients were obtained from the TCGA database mentioned above. To compare the mutational loading between low- and high risk groups, we extracted the mutation data of somatic variants from Mutation Annotation Format (MAF), and analyzed it by MAF tools package [13].

Results

Differentially Expressed LRGs in EC

The process for this study is shown in **Figure 1**. A total of 532 patients were involved in the development and validation of the prognostic signature, including 35 normal tissues. First, we performed the Wilcoxon test with a $\log_2FC > 1$ and $P < 0.05$ to detect the differentially expressed genes (DEGs). Then, 5,155 DEGs

were found between 35 normal samples and 532 tumor samples. Next, we downloaded the list of LRGs from the MSigDB. These IRGs intersected with the DEGs, and 89 differentially expressed IRGs were obtained (**Figure 2A**), including 32 down-regulated and 57 up-regulated genes (**Figures 2B**). The DE-LRGs list, including \log_2FC and the adjusted P -values of each gene was provided in Supplementary Table S1. Afterwards, we performed GO and KEGG pathway analysis for the DE-LRGs and the top 10 GO and KEGG pathway enrichment terms shown in **Figures 2C-D**.

GO analysis revealed that the DE-LRGs were mainly enriched in mitotic nuclear division, nuclear division, chromosome segregation and so on. The KEGG analysis indicated that the genes were mainly involved in positive regulation of cell cycle, progesterone-mediated oocyte maturation, p53 signaling pathway. et al. Thus, through the combined analysis, we revealed 89 DE-LRGs that were significantly associated with EC. Most of the DEGs were associated with the cell cycle and p53 pathway.

Identification of prognostic immune-related DEGs

To further screened out the LRGs with potentially prognostic value for EC patients, LASSO regression model was used to select key prognosis-associated genes. In the LASSO-penalized Cox regression, as $\log \lambda$ (a tuning parameter) changed, the corresponding coefficients of certain genes were reduced to zero, suggesting that their effects on the model could be omitted because they were shrinking parameters (**Figure 3A**). Following cross-validation, 11 genes (C14orf28, HNRNPA3P1, DSP, ACACB, TPX2, HMGB3, ATP8B4, PPP1R14C, CIRBP, CDC6, DTWD1) achieved the minimum partial likelihood deviance and were identified as key prognostic immune-related genes in OS model (**Figure 3B**). As described in the methods. the calculation formula = $(ACACB * 0.1335) - (ATP8B4 * 0.0144) - (C14orf28 * 0.0007) - (CDC6 * 0.0727) - (CIRBP * 0.1112) + (DSP * 0.0196) - (DTWD1 * 0.0318) + HMGB3 * 0.1201 - (HNRNPA3P1 * 0.1167) + (PPP1R14C * 0.0185) + (TPX2 * 0.1833)$. The correlations between these genes were calculated in EC using the Spearman correlation analysis. We found they were significantly relevant. For instance, expression levels of DTWD1 and ABACB, ATP8B4 and CDC6, ATP8B4 and HMGB3 genes were closely correlated with each other (**Figure 3C**). The expression of these 11 genes in tumor and adjacent normal tissues were shown in **Figure 3D**. The results showed that the expression of ACACB, ATP8B4, C14orf28, CIRBP, and DTWD1 were higher in normal tissues, and the content of CDC6, DSP, HMGB3, HNRNPA3P1, PPP1R14C, and TPX2 were higher in tumor tissues. Furthermore, the transcript message of patients stratified by risk score into high- and low- risk subgroups were analyzed by GSEA. In the prognostic model, representative hallmark in high-risk patients were "androgen response", "DNA repair", "estrogen response early", "fatty acid metabolism", "glycolysis". What's more, "MTORC1 signaling", "notch signaling", "PI3K-AKT-MTOR signaling", "TNFA signaling via NFkB", "WNT- β -catenin- signaling" were enriched in low-risk group patients (**Figure 3E**).

Validation and the efficacy of the 11-LRGs prognostic signature

Based on the mean risk score from risk signature, the patients were divided into high-risk and low-risk groups. Then, the expression of the 11 genes in low- and high-risk patients in the TCGA dataset was also

demonstrated in the heatmap (**Figure 4A**). We found significant differences between the high- and low-risk groups associated with LNM, cancer status, peritoneal cytology, recurrence, grade, menopausal status, stage, and stage (all $p < 0.05$). Then risk score of each individual and survival status were ranked and displayed on the dot plot, which showed significant differences in OS between the groups (**Figure 4B-C**). Likewise, the Kaplan-Meier curve analysis demonstrated that the OS of the high-risk group was significantly shorter than that of the low-risk group ($p = 3.583e-08$) (**Figure 4D**). ROC curve analysis revealed that the area under the ROC curve (AUC) of 1-, 3-, 5-year survival of the prognostic LRG model was 0.669, 0.702, and 0.718 (**Figure 4E**). Furthermore, we analyzed the 11 genes for different status of LNM, respectively. We found that expression levels of C14orf28, CIRBP, DTWD1, HMGB3, and TPX2 were significantly different (**Figure S1**). Above all, these results revealed that the LRGs risk model could be served as an effective and accurate prognostic signature in EC.

Nomogram building and validation

For supplying the clinicians a practical formula to estimate EC patients' survival probability, a comprehensive prognostic nomogram based on the patients' risk scores and clinical features was built. First of all, univariate and multivariate Cox regression analyses of survival were performed by the risk signature and related clinicopathological factors (age, menopausal status, stage, histology, grade, peritoneal cytology, LNM, and 11-gene risk model) in the TCGA dataset to determine whether the 11-gene risk score can be used as an independent prognostic factor. Both univariate and multivariate analyses showed that the risk model could be used as a prognostic indicator ($P < 0.001$, **Figure 5A-B**). Afterwards, five independent prognostic parameters, including age, grade, stage, peritoneal cytology, and 11-gene risk model were integrated into the nomogram (**Figure 5C**). The specific score of each factor was shown in **Table 3**. The C-index of the model was 0.769 (C-index=0.712 for clinical prognostic factors alone). The calibration plots showed excellent consistency between the nomogram predictions and actual observations in terms of the 1-, 3-, and 5-year survival rates in the TCGA cohort (**Figure 5D**). Furthermore, we divided the cohort into 3 subgroups (low-score, moderate-score, and high-score groups) evenly according to their risk scores from the nomogram. The survival curve of high-score group had a worse OS than the moderate- and low-score groups (**Figure 5E**). In order to find out whether the risk signature was an effective prognostic indicator, tdROC was plotted. Similar to the performance in the cohort, the AUCs were 0.784, 0.814, and 0.787 for 1-, 3-, and 5-year survival time, respectively (**Figure 5F**). These results found that the 11-gene LRGs model improved the predictive accuracy of OS in EC patients.

Clinical experimental validation-qPCR

First of all, the differential expression of 11 genes was validated based on GEPIA website (**Figure S2**), including HMGB3 (**Figure 6A**).

We knocked down the expression of HMGB3 in ishikawa cell line and conducted the following functional experiments. Then western blot was performed to examine the efficiency of knockdown. The results suggested that the expression of MMP12 in the siRNA group was significantly down-regulated compared with the negative control group (Ctrl) and the blank control group (siCtrl) in ishikawa cells ($p < 0.01$, **Figure**

6B). CCK-8 assay suggested that the proliferation rate of ishikawa was significantly decreased in HMGB3 knockdown group (**Figure 6C**). Transwell and gap closure assay were performed to investigate the effects of HMGB3 on the invasion and metastatic behaviors of EC cells in vitro. The results demonstrated that ishikawa in HMGB3 knockdown group exhibited significant declines in invasion capabilities (**Figure 6D-E**) and migration (**Figure 6F-G**) compared with respective control groups. These results suggested that HMGB3 played a crucial role in the progression of EC cells.

The immune related risk signature and mutation profile

Gene mutations are an important cause of tumorigenesis and development. Hence, we evaluated tumor mutation burden (TMB) of patients in low- and high-risk groups with somatic mutation data. In the risk model, the low-risk group had somatic mutations in the following order: PTEN> ARID1A> PIK3CA> TTN> CTNNB1> PIK3R1> CTCF> MUC16> KMT2D> ZFH3 (**Figure 7A**). Meanwhile, in the high-risk group, somatic mutations were listed in the following order: TP53> PIK3CA> PTEN> TTN> ARID1A> PIK3R1> KMT2D> MUC16> PPP2R1A> CHD4 (**Figure 7B**). The TMB scores of patients stratified by the LRGs prognostic model was therefore investigated. The *t* test demonstrated a significant difference between the low-risk and high-risk groups ($p<0.01$, **Figure S3**). These results indicated that different risk groups in LRGs risk model had different TMB features.

Discussion

The prognosis of EC is great, however, extensive heterogeneity has been reported in a number of studies associated with recurrent disease and high mortality [14]. Recent studies showed that clinicopathological features such as age and metastatic diagnosis are not sufficient to precisely predict the outcome of patients with cancer. Thus, there remains a need to elucidate the molecular mechanisms underlying tumor progression or prognosis. Accumulating evidences show that a comprehensive understanding of EC requires attention not only to clinicopathological features but also to the gene expression. An increasing number of evidences showed that EC patients with lymph node metastasis always exhibited poor responses to standard treatments and thus tend to have poor clinical outcomes [15].

In this study, we performed a comprehensive analysis of lymph node metastasis related genes (LRGs) in EC and linked the data to clinical outcomes and prognosis of patients with EC. First, we systematically studied the LRGs in EC and identified 89 differentially expressed LRGs (DE-LRGs). Based on multivariate Cox coefficients derived from LASSO analysis, we developed a 11-gene signature-based risk score model associated with overall survival in EC patients, namely, C14orf28, HNRNPA3P1, DSP, ACACB, TPX2, HMGB3, ATP8B4, PPP1R14C, CIRBP, CDC6, and DTWD1. Subsequently, both of time-dependent ROC curves and Kaplan–Meier plots revealed that the risk signature outperformed in the prediction of OS for EC patients. By combining the established clinicopathological factors with the signature, we developed a nomogram to predict the survival probability of EC patients. The predictive power was measured by the time-dependent area under the ROC curve (AUC), and the result showed that the integrated nomogram

model had higher predictive power than other models. Lastly, HMGB3 was identified as an oncogene to promote the proliferation and migration of EC cells.

Finally, we analyzed and compared the somatic mutation of low- and high-risk group patients to visualize the mutation annotation format of the two groups. PTEN was the highest mutated genes in low risk group, and TP53 was the highest in high-risk group.

As far as we know, this is the first study concentrating on lymph node metastasis related genes and risk signature construction in EC. There are studies using the LNM related genes to establish the prognostic model in other cancers. One study proved a four-gene signature can be used as a combined biomarker for independent prognosis of colorectal cancer. In this model, high-risk group patients were associated with neuroactive ligand-receptor interaction and estrogen signaling pathway. AUC of the overall survival of the risk model reached to 0.730 [16]. Wang et al. developed a robust mRNA signature as an independent factor to effectively classify LUAD patients with LNM into low- and high-risk groups using TCGA database, and validated by an external cohort from Gene Expression Omnibus (GEO). GSEA showed the hallmarks correlated with the high-risk group were EMT, Hypoxia, and MYC targets [17]. These studies including ours all suggested that LNM related genes were promising for many types of cancer in the prediction of overall survival.

To study the potential molecular mechanism of prognostic effects of the 11-gene signature, GSEA analysis was conducted. The results of GSEA showed that the signature-identified high-risk group was significantly correlated with certain hallmarks of cancer, such as fatty acid metabolism, glycolysis, Notch signaling pathway, and PI3K-AKT-MTOR signaling pathway, indicating the potential molecular mechanisms underlying the lethal tendency of patients in the risk signature. Metabolic disorder had long been reported in EC, and over-expression of fatty acid synthase is associated with cancer progression and upper body fat distribution in EC [18]. Glycolysis of cellular respiration is a complex reaction and is the first step in most carbohydrate catabolism, which was proved to participate in EC. One study used 9 glycolysis-related genes to predict the survival of patients with EC, and the predictive accuracy of overall survival proved to be high for EC patients (AUC of 1-year OS = 0.763) [19]. Evolutionary conserved Notch signaling pathway regulated diverse cellular processes including proliferation, differentiation, and cell invasion. Accumulating evidence links aberrant Notch signaling with EC. For example, FOXA1 promotes cell proliferation by androgen receptor and activates Notch pathway in EC [20]]. Activation of the PI3K/AKT/mTOR pathway reversed FAM83B knockdown-induced autophagy promotion and inhibition of proliferation, migration, and invasion in EC cells. This results indicated that proliferation and metastasis cells inhibited autophagy via activating the PI3K/AKT/mTOR pathway in EC cells [21].

Molecular classification of EC has been shown to be reproducible and associated with clinical outcomes [22, 23]. An integrated genomic-pathologic classification of EC has been proposed by, which defined four major classes of EC as POLE-ultramutated, microsatellite instability–hypermutated (MSI-H), copy-number-low (CNL), and copy-number-high(CNH). It also needs to be noteworthy that some of these genes have been reported in previous studies of cancer [24]. As for the characteristics of these signature genes,

higher expression levels of ACACB, DSP, HMGB3, PPP2R14C, and TPX2 are associated with poor prognosis. On the other hand, higher expression levels of the remaining C14orf28, HNRNPA3P1, ATP8B4, CIRBP, CDC6, and DTWD1 are associated with longer OS.

ACACB was associated with HER2 positive breast cancer with brain metastasis [24]. TPX2 represents a novel prognostic factor for esophageal cancer, and the 5-year overall survival rate of TPX2 high expression group was significantly lower than that of TPX2 low expression group [25]. Several studies have demonstrated that TPX2 also plays an important role in the development of prostate cancer [26], non-small cell lung cancer [27], and breast cancer [28]. Loss of CIRBP expression is correlated with the malignant progression and poor prognosis in nasopharyngeal carcinoma [29]. Among the 11 coding genes, HMGB3 was widely reported to promote malignant progress and predict poor survival in various cancer types [30–32]. In addition, Gu et al. reported that HMGB3 silencing could inhibit the cell proliferation in vitro and suppress tumor growth in vivo levels. The antitumor effects of HMGB3 silencing were mediated by interacting with the HIF1 α [33]. But the expression and function of HMGB3 in EC cells were still ambiguous. Therefore, we further explored the molecular function of HMGB3 in EC cell line Ishikawa. The results revealed that silencing expression of HMGB3 inhibited proliferation, migration and invasion of EC cell lines.

The relationship between mutation profile and the signature was also performed to explore the possible mechanisms of the signature's prognostic value. Mutation profiles of the 11-IRG signature high-risk group and low-risk group were different. The most frequently mutated genes in the low-risk and high-risk groups are PTEN and TP53, respectively. Phosphatase and Tensin homolog (PTEN) is a tumor suppressor gene. Loss of its function is the most frequent genetic alteration in endometrioid endometrial cancers (70–80%) and high grade tumors (90%). However, TP53 mutations are considered a surrogate biomarker of the serous-like 'copy number high' molecular subtype of endometrial carcinoma (EC) [34]. Understanding the complex interaction of mutated genes in endometrial cancer will help to better select patients that are likely to respond to some of the new and costly targeted therapies.

In this study, we constructed a valid LNM related risk signature that can comprehensively predict survival of EC patients. Furthermore, the proliferation and metastasis of HMGB3 was also validated in vitro, and HMGB3 is potential to become a new biomarker for EC patients. To the best of our knowledge, there are no studies exploring the function of HMGB3 in EC, and this is the first LNM related predictive model for EC patients using RNA-sequencing technology. Additionally, our signature can predict patient survival and disease progression. Despite these promising results, there are several limitations in this study. First, the signature was constructed from public data and retrospective studies. Second, in this TCGA EC cohort, the proportion of Chinese patients was small. It is unclear whether this signature will function effectively for Asian patients. Future studies should incorporate a larger number of Chinese EC samples. Finally, external validation is also needed in other centers or datasets.

Conclusion

In conclusion, we identified differently expressed LNM-related genes that may involve in the prognosis in EC patients. These genes have significant values in predicting the patients' OS and HMGB3 may be a therapeutic target for EC. Further studies are necessary to verify these results in our study.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

Conceptualization: Hong Wu, Liping Yang.

Data curation: Haiqin Feng

Formal analysis: Haiqin Feng, Xiaoli Miao

Investigation: Jiancai Ma, Cairu Liu

Methodology: Hong Wu, Lina Zhang

Resources: Liping Yang, Xiaoli Miao

Validation: Haiqin Feng

Writing – original draft: Hong Wu

Writing – review & editing: Liping Yang, Hong Wu

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Additional data not presented in the manuscript can be obtained by contacting the authors.

Ethics approval and consent to participant

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Statement

The study conformed to the provisions of the Declaration of Helsinki (as revised in 2013), and we confirm that all figures and tables are original.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018, *68*, 7-30.
2. Chen, W.; Zheng, R.; Baade, P.D.; Zhang, S.; Zeng, H.; Bray, F.; Jemal, A.; Yu, X.Q.; He, J. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016, *66*, 115-132.
3. Pecorelli, S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet.* 2009, *105*, 103-104.
4. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* 2015, *13*, 8-17.
5. Pfaff, E.; Aichmuller, C.; Sill, M.; Stichel, D.; Snuderl, M.; Karajannis, M.A.; Schuhmann, M.U.; Schittenhelm, J.; Hasselblatt, M.; Thomas, C.; et al. Molecular subgrouping of primary pineal parenchymal tumors reveals distinct subtypes correlated with clinical parameters and genetic alterations. *Acta Neuropathol.* 2020, *139*, 243-257.
6. Wang, Z.H.; Zhang, Y.Z.; Wang, Y.S.; Ma, X.X. Identification of novel cell glycolysis related gene signature predicting survival in patients with endometrial cancer. *Cancer Cell Int.* 2019, *19*, 296.
7. Zhou, H.; Chen, L.; Lei, Y.; Li, T.; Li, H.; Cheng, X. Integrated analysis of tumor mutation burden and immune infiltrates in endometrial cancer. *Curr Probl Cancer.* 2021, *45*, 100660.
8. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005, *102*, 15545-15550.
9. Pencina, M.J.; D'Agostino, R.B. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med.* 2004, *23*, 2109-2123.
10. Royston, P.; Altman, D.G. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol.* 2013, *13*, 33.
11. Tang, Z.; Li, C.; Kang, B.; Gao, G.; Li, C.; Zhang, Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017, *45*, W98-W102.
12. Li, X.; Yang, X.; Fan, Y.; Cheng, Y.; Dong, Y.; Zhou, J.; Wang, Z.; Li, X.; Wang, J. A ten-gene methylation signature as a novel biomarker for improving prediction of prognosis and indicating gene targets in endometrial cancer. *Genomics.* 2021, *113*, 2032-2044.
13. Mayakonda, A.; Lin, D.C.; Assenov, Y.; Plass, C.; Koeffler, H.P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018, *28*, 1747-1756.
14. Bendifallah, S.; Ouldamer, L.; Lavoue, V.; Canlorbe, G.; Raimond, E.; Coutant, C.; Graesslin, O.; Touboul, C.; Collinet, P.; Darai, E.; et al. Patterns of recurrence and outcomes in surgically treated women with endometrial cancer according to ESMO-ESGO-ESTRO Consensus Conference risk groups: Results from the FRANCOGYN study Group. *Gynecol Oncol.* 2017, *144*, 107-112.
15. Woolston, C.M.; Zhang, L.; Storr, S.J.; Al-Attar, A.; Shehata, M.; Ellis, I.O.; Chan, S.Y.; Martin, S.G. The prognostic and predictive power of redox protein expression for anthracycline-based chemotherapy

- response in locally advanced breast cancer. *Mod Pathol.* 2012,25, 1106-1116.
16. Li, X.; Zhang, Q.; Zhao, L.; Jiang, L.; Qi, A.; Wei, Q.; Song, X.; Wang, L.; Zhang, L.; Zhao, Y.; et al. A Combined four-mRNA Signature Associated with Lymphatic Metastasis for Prognosis of Colorectal Cancer. *J Cancer.* 2020,11, 2139-2149.
 17. Wang, Y.; Zhang, Q.; Gao, Z.; Xin, S.; Zhao, Y.; Zhang, K.; Shi, R.; Bao, X. A novel 4-gene signature for overall survival prediction in lung adenocarcinoma patients with lymph node metastasis. *Cancer Cell Int.* 2019,19, 100.
 18. Tsuji, T.; Yoshinaga, M.; Togami, S.; Douchi, T.; Nagata, Y. Fatty acid synthase expression and clinicopathological findings in endometrial cancer. *Acta Obstet Gynecol Scand.* 2004,83, 586-590.
 19. Liu, J.; Li, S.; Feng, G.; Meng, H.; Nie, S.; Sun, R.; Yang, J.; Cheng, W. Nine glycolysis-related gene signature predicting the survival of patients with endometrial adenocarcinoma. *Cancer Cell Int.* 2020,20, 183.
 20. Qiu, M.; Bao, W.; Wang, J.; Yang, T.; He, X.; Liao, Y.; Wan, X. FOXA1 promotes tumor cell proliferation through AR involving the Notch pathway in endometrial cancer. *BMC Cancer.* 2014,14, 78.
 21. Lin, Q.; Chen, H.; Zhang, M.; Xiong, H.; Jiang, Q. Knocking down FAM83B inhibits endometrial cancer cell proliferation and metastasis by silencing the PI3K/AKT/mTOR pathway. *Biomed Pharmacother.* 2019,115, 108939.
 22. Salinas, E.A.; Miller, M.D.; Newton, A.M.; Sharma, D.; McDonald, M.E.; Keeney, M.E.; Smith, B.J.; Bender, D.P.; Goodheart, M.J.; Thiel, K.W.; et al. A Prediction Model for Preoperative Risk Assessment in Endometrial Cancer Utilizing Clinical and Molecular Variables. *Int J Mol Sci.* 2019,20.
 23. Liu, J.; Wan, Y.; Li, S.; Qiu, H.; Jiang, Y.; Ma, X.; Zhou, S.; Cheng, W. Identification of aberrantly methylated differentially expressed genes and associated pathways in endometrial cancer using integrated bioinformatic analysis. *Cancer Med.* 2020,9, 3522-3536.
 24. Cancer Genome Atlas Research, N.; Kandoth, C.; Schultz, N.; Cherniack, A.D.; Akbani, R.; Liu, Y.; Shen, H.; Robertson, A.G.; Pashtan, I.; Shen, R.; et al. Integrated genomic characterization of endometrial carcinoma. *Nature.* 2013,497, 67-73.
 25. Lu, X.; Gao, C.; Liu, C.; Zhuang, J.; Su, P.; Li, H.; Wang, X.; Sun, C. Identification of the key pathways and genes involved in HER2-positive breast cancer with brain metastasis. *Pathol Res Pract.* 2019,215, 152475.
 26. Sui, C.; Song, Z.; Yu, H.; Wang, H. Prognostic significance of TPX2 and NIBP in esophageal cancer. *Oncol Lett.* 2019,18, 4221-4229.
 27. Zhou, F.; Wang, M.; Aibaidula, M.; Zhang, Z.; Aihemaiti, A.; Aili, R.; Chen, H.; Dong, S.; Wei, W.; Maimaitiaili, A. TPX2 Promotes Metastasis and Serves as a Marker of Poor Prognosis in Non-Small Cell Lung Cancer. *Med Sci Monit.* 2020,26, e925147.
 28. Tan, G.Z.; Li, M.; Tan, X.; Shi, M.L.; Mou, K. MiR-491 suppresses migration and invasion via directly targeting TPX2 in breast cancer. *Eur Rev Med Pharmacol Sci.* 2019,23, 9996-10004.
 29. Lin, T.Y.; Chen, Y.; Jia, J.S.; Zhou, C.; Lian, M.; Wen, Y.T.; Li, X.Y.; Chen, H.W.; Lin, X.L.; Zhang, X.L.; et al. Loss of Cirbp expression is correlated with the malignant progression and poor prognosis in

- nasopharyngeal carcinoma. *Cancer Manag Res.* 2019,11, 6959-6969.
30. Fang, J.; Ge, X.; Xu, W.; Xie, J.; Qin, Z.; Shi, L.; Yin, W.; Bian, M.; Wang, H. Bioinformatics analysis of the prognosis and biological significance of HMGB1, HMGB2, and HMGB3 in gastric cancer. *J Cell Physiol.* 2020,235, 3438-3446.
31. Li, Z.; Zhang, Y.; Sui, S.; Hua, Y.; Zhao, A.; Tian, X.; Wang, R.; Guo, W.; Yu, W.; Zou, K.; et al. Targeting HMGB3/hTERT axis for radioresistance in cervical cancer. *J Exp Clin Cancer Res.* 2020,39, 243.
32. Xie, X.; Pan, J.; Han, X.; Chen, W. Downregulation of microRNA-532-5p promotes the proliferation and invasion of bladder cancer cells through promotion of HMGB3/Wnt/beta-catenin signaling. *Chem Biol Interact.* 2019,300, 73-81.
33. Gu, J.; Xu, T.; Huang, Q.H.; Zhang, C.M.; Chen, H.Y. HMGB3 silence inhibits breast cancer cell proliferation and tumor growth by interacting with hypoxia-inducible factor 1alpha. *Cancer Manag Res.* 2019,11, 5075-5089.
34. Singh, N.; Piskorz, A.M.; Bosse, T.; Jimenez-Linan, M.; Rous, B.; Brenton, J.D.; Gilks, C.B.; Kobel, M. p53 immunohistochemistry is an accurate surrogate for TP53 mutational analysis in endometrial carcinoma biopsies. *J Pathol.* 2020,250, 336-345.

Tables

Table 1: 11 lymph node metastasis associated genes and corresponding coefficient value

LNM associated genes	Coefficient
ACACB	0.13351004
ATP8B4	-0.0144339
C14orf28	-0.0007171
CDC6	-0.0727459
CIRBP	-0.1112859
DSP	0.01966864
DTWD1	-0.0318811
HMGB3	0.12010494
HNRNPA3P1	-0.1167104
PPP1R14C	0.01855752
TPX2	0.18337651
Risk score	Low: <3.75
	High: ≥3.75

Table 2: Corresponding risk score for each variable and total score

Variables	Category	Score
Age	<60	0
	≥60	35
Grade	G1	0
	G2	75
	G3	100
Stage	Stage I	0
	Stage II	20
	Stage III	45
	Stage IV	55
Peritoneal cytology	Negative	0
	Positive	32.5
Risk signature	Low	0
	High	72.5
Total score	Low risk	0-110
	Moderate risk	112.5-207.5
	High risk	≥212.5

Supplementary

Supplementary Table S1 is not available with this version

Figures

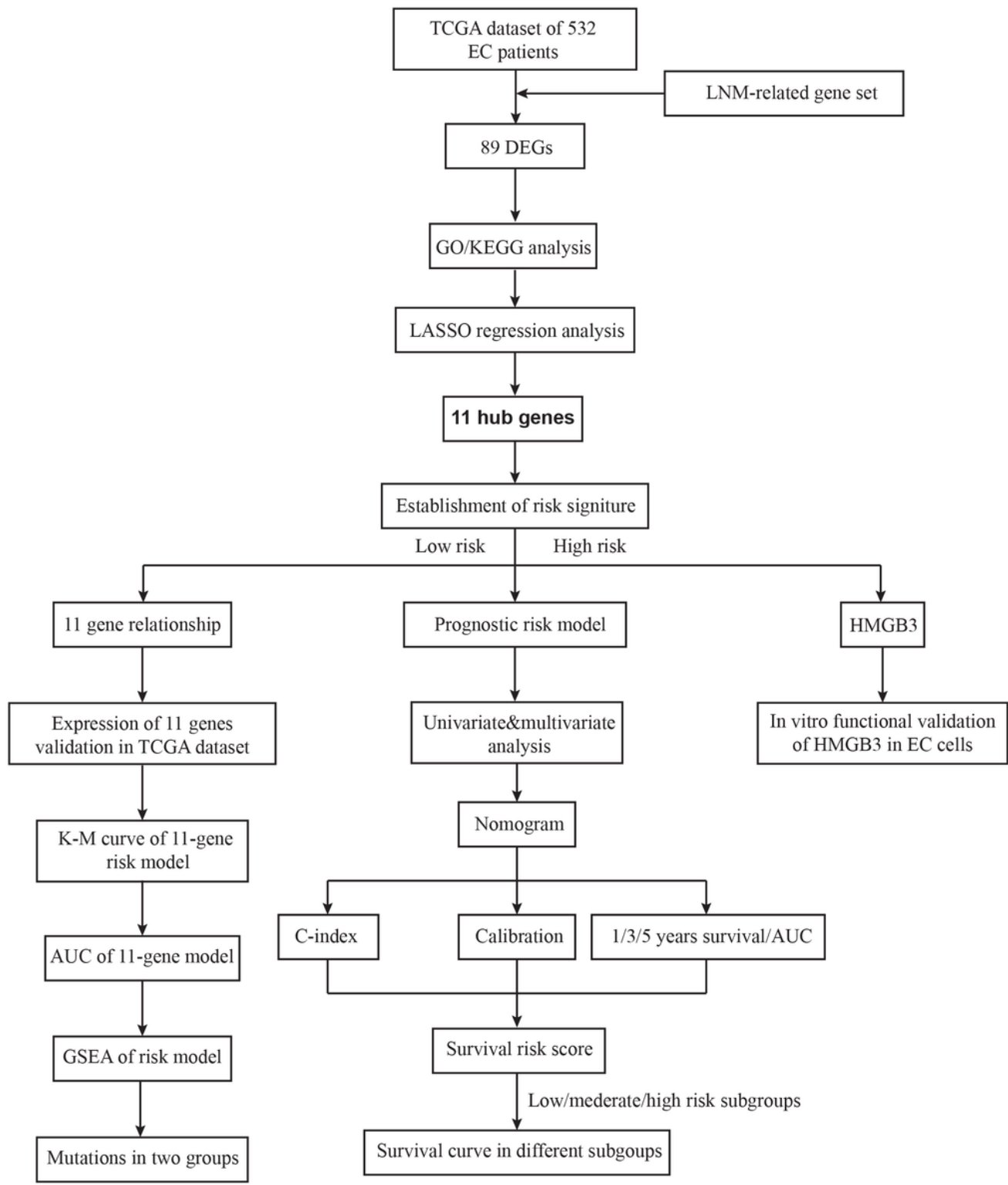


Figure 1

Flowchart for identifying the LNM-related prognostic signature.

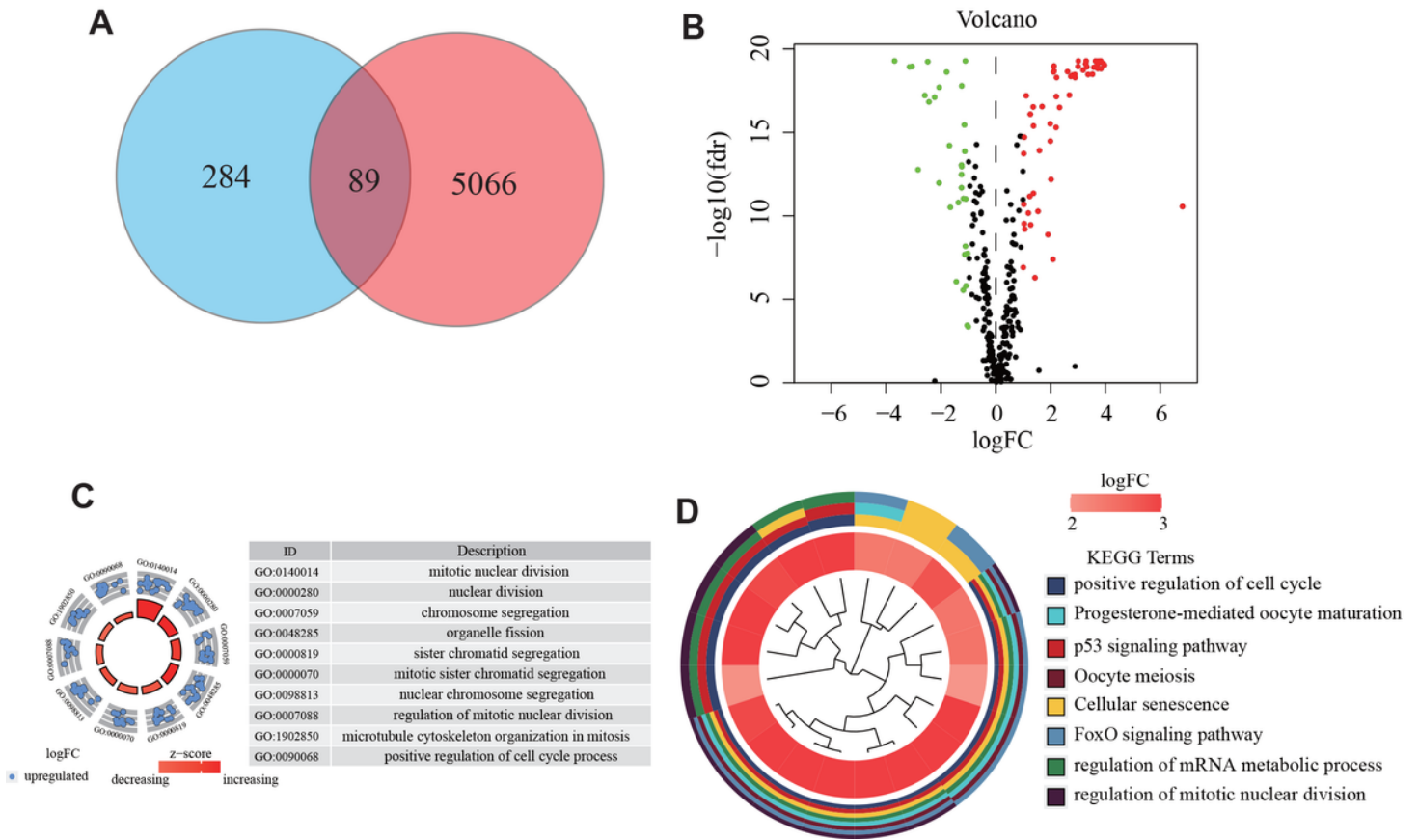


Figure 2

Identification and functional analysis of LRGs in EC. A. The intersect LNM-related genes were associated with DEGs between normal and EC tissues. B. Volcano plot was drawn to show the differentially expressed LNM related genes. C. Enrichment analysis reveals the top 10 GO terms and D. KEGG pathways.

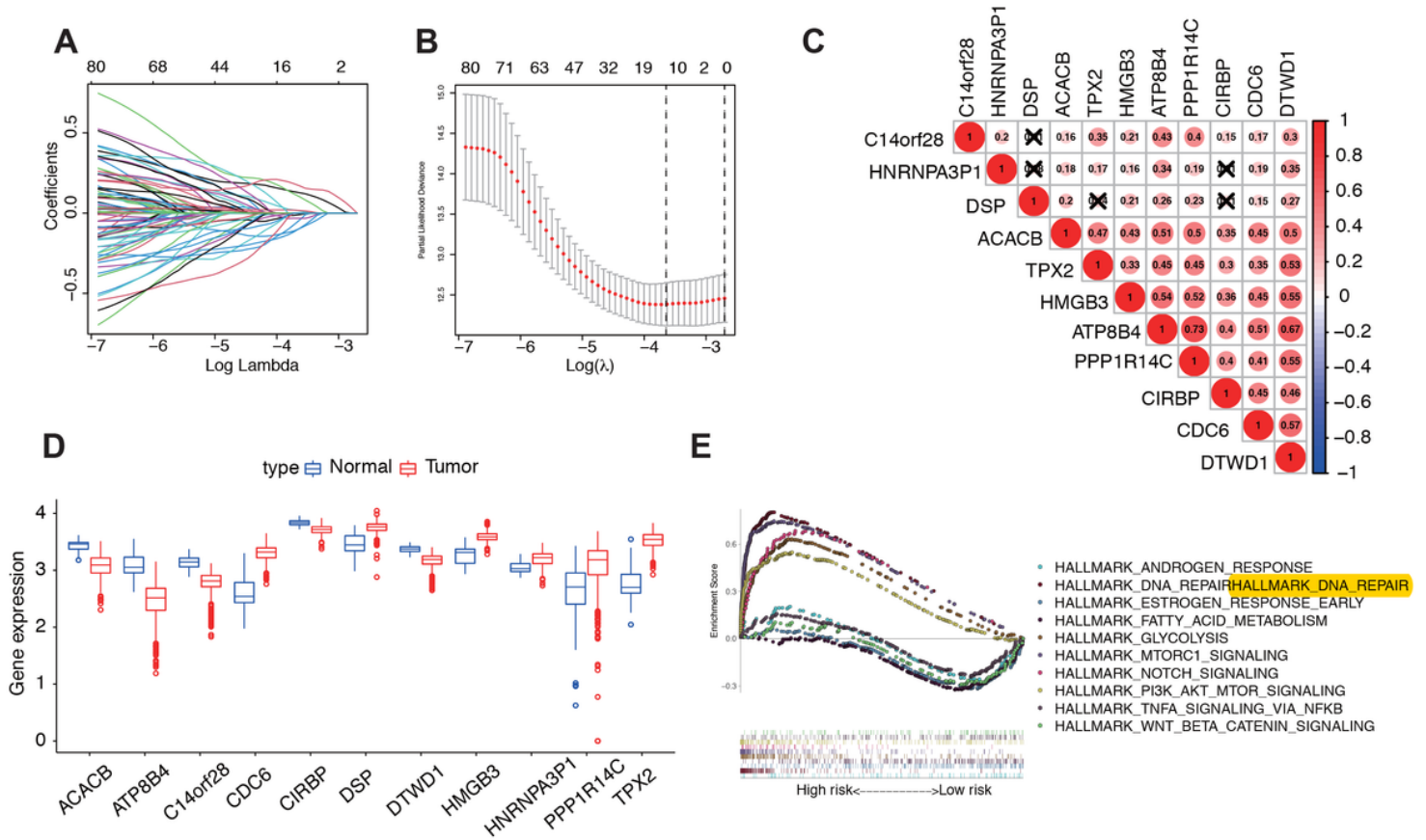


Figure 3

Construction of risk signature with LRGs. A. LASSO coefficients. B. Plots of the cross-validation error rates. The dashes signify the value of the minimal error and greater λ value. C. Spearman correlation analysis of the 11 LNM-related genes. D. The expression of 11 LRGs between TCGA endometrial cancer (EC) and normal tissues. E. GSEA showed significant enrichment of the tumor-related signaling pathways.

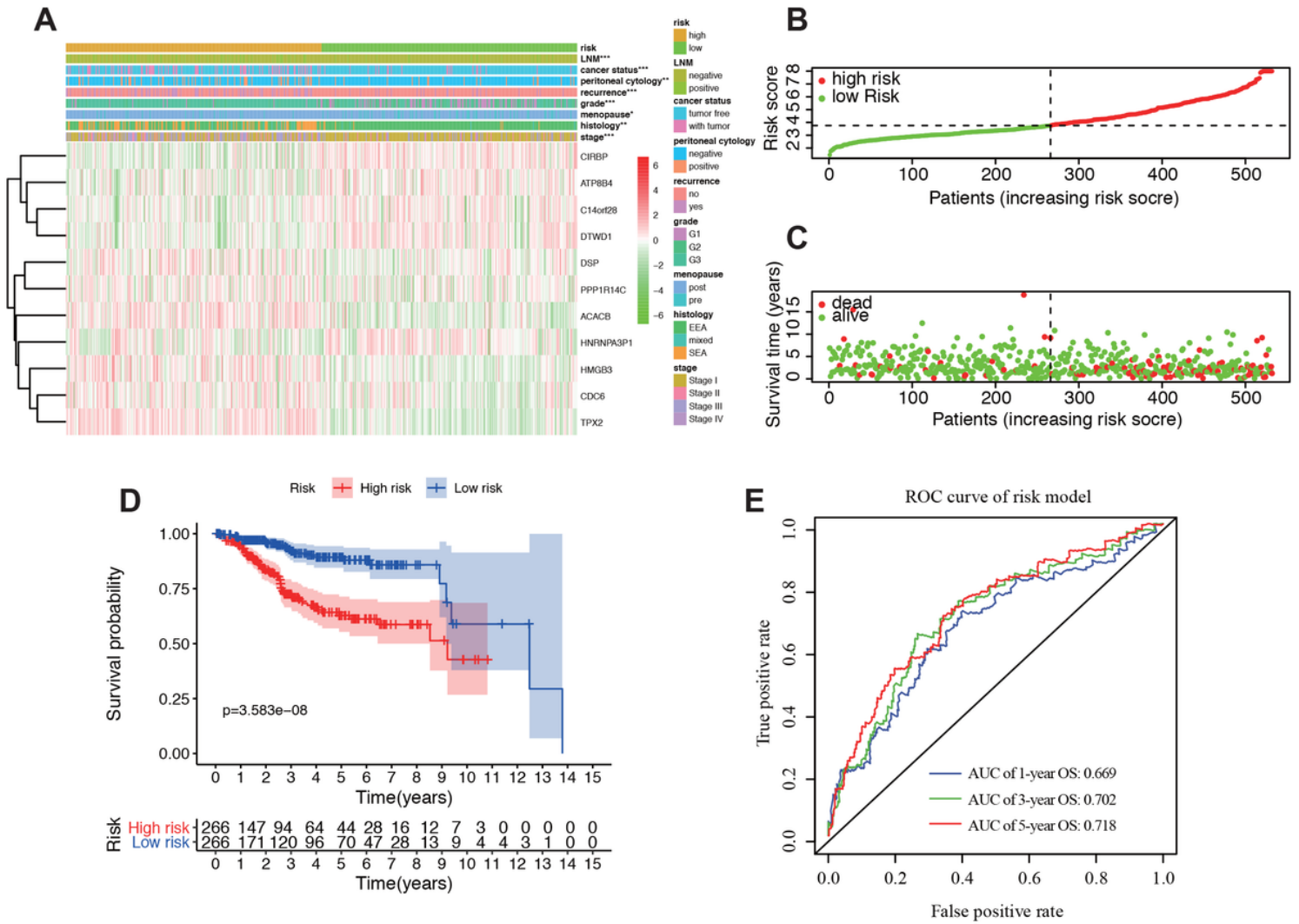


Figure 4

Prognostic analysis of the risk model in the TCGA patients. A. The heatmap shows the expression of the 11 genes in high-risk and low-risk group of EC. The distribution of clinicopathological characteristics was compared between the high-risk and low-risk groups. B-C. The distributions of the five-gene signature and survival status of the patients in the risk signature. D. Kaplan-Meier survival analysis with patients in low and high-risk groups. E. 1-, 3-, 5-year survival time dependent receiver operating characteristic (ROC) curve. $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. ROC, receiver operating characteristic; AUC, area under the ROC curve. E. the 1-, 3-, and 5-year AUC of ROC curves.

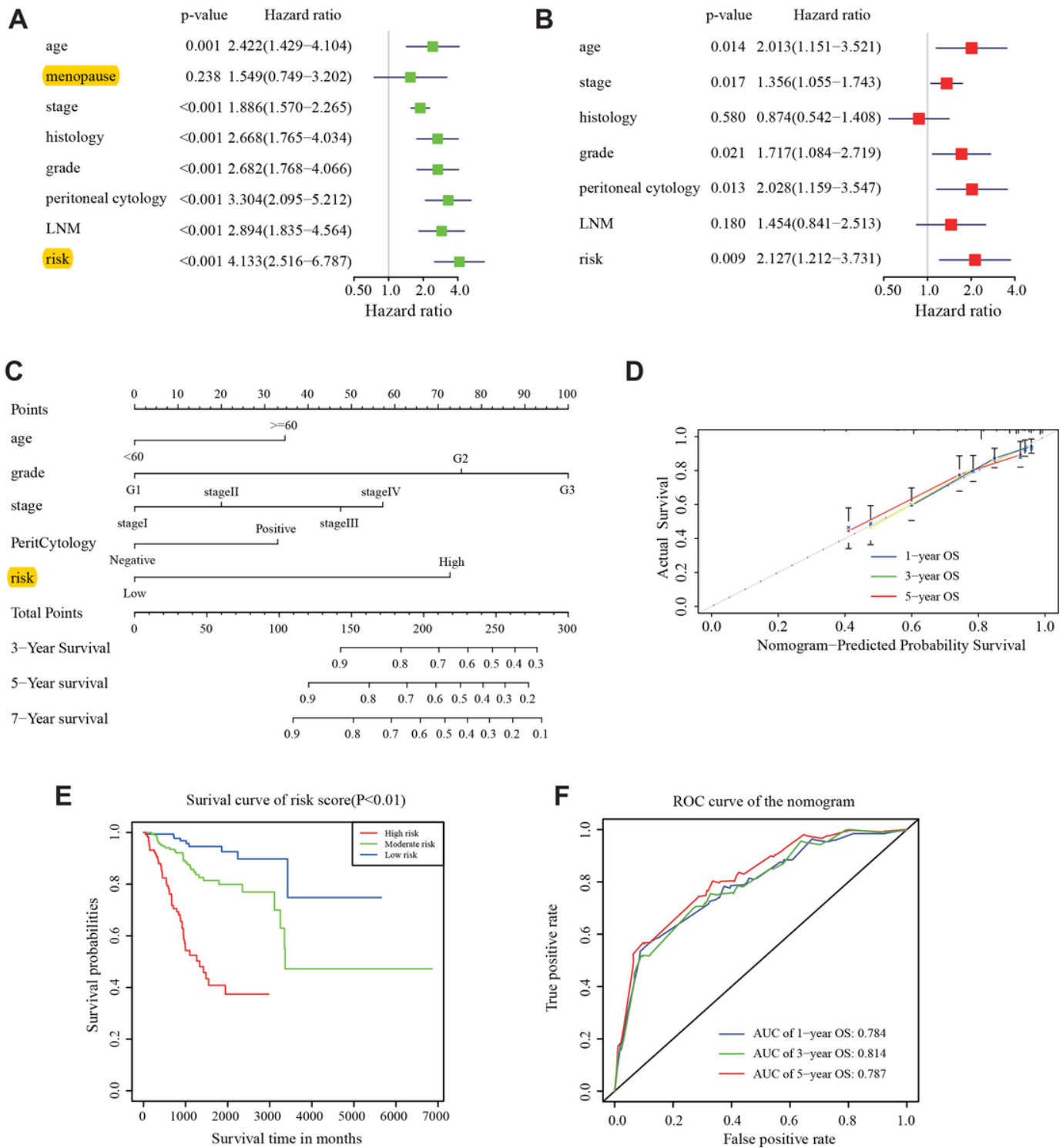


Figure 5

Construction and validation of nomogram. A. Univariate and multivariate Cox analyses of cervical cancer. B. Nomogram used to predict prognosis in patients with endometrial cancer at 1, 3, and 5 years. C. Calibration curves for the nomogram at 1-, 3-, and 5-year overall survival. D. Survival curve of patients in low-, moderate-, and high-score according to the total score of the nomogram. E. ROC curve of 1-, 3-, 5-year survival depend on the nomogram.

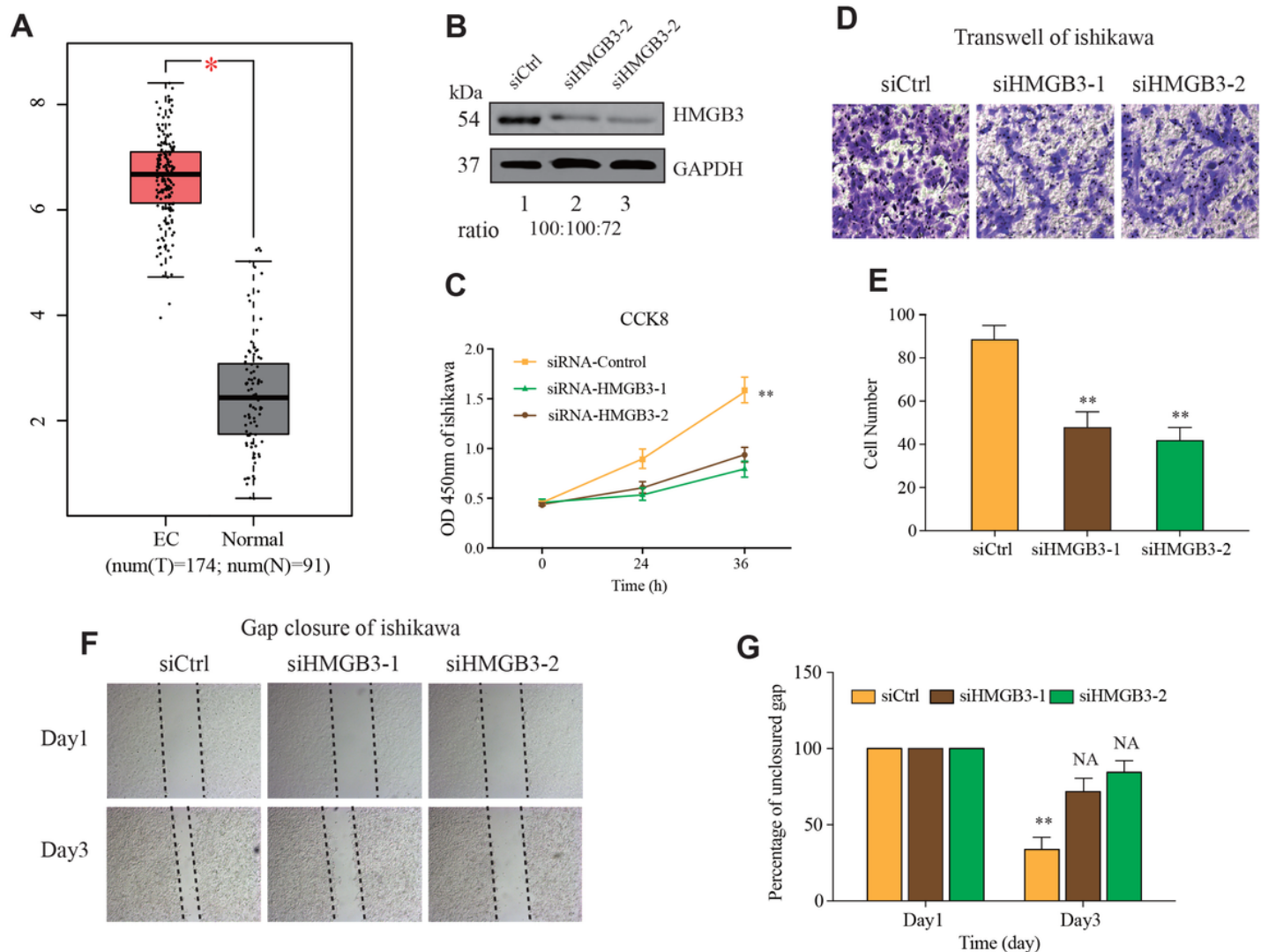


Figure 6

In vitro functional validation of the HMGB3. A. The relative expression of HMGB3 in GEPIA. B. Knockdown efficiency of HMGB3 by two small interferon RNA transfection. C. Proliferative effect of HMGB3 on ishikawa evaluated by Cell Counting Kit-8 tests. D. Effects of HMGB3 on the invasion of ishikawa cells evaluated by Transwell assays. E. Statistical analysis of the Transwell invasion. (F) Effects of MMP12 on the migration of ishikawa cells evaluated by gap closure assays. (G) Statistical analysis of the gap closure. * $p < 0.05$, ** $p < 0.01$.

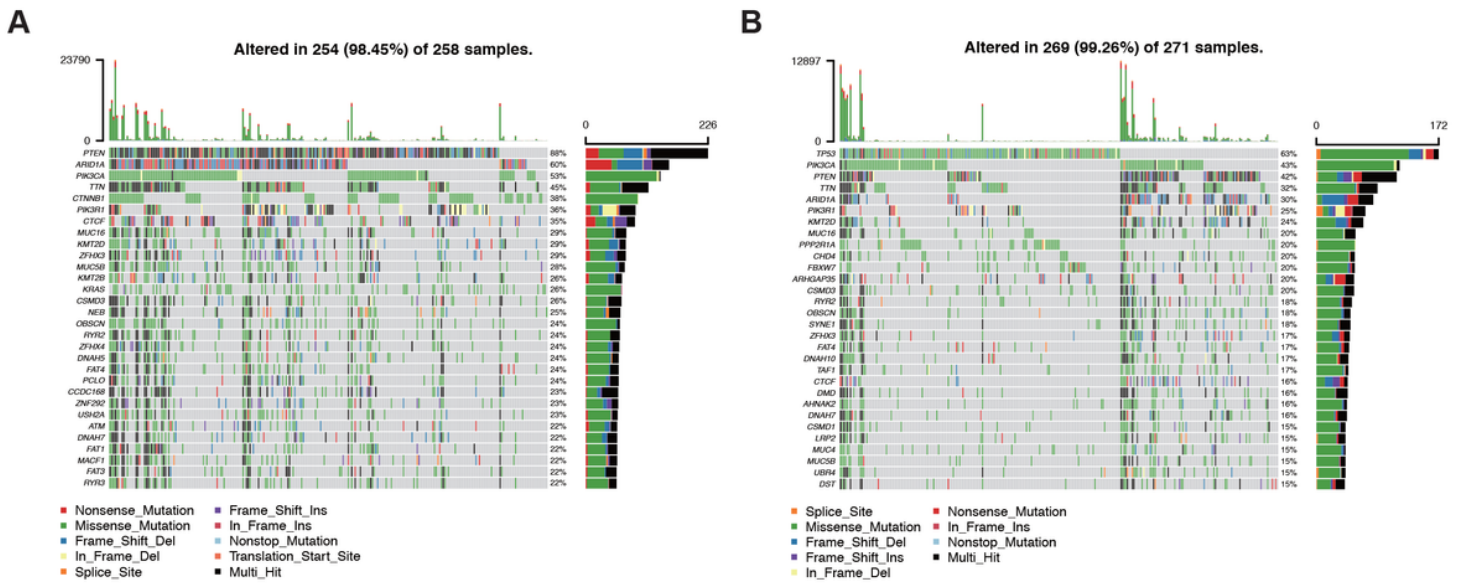


Figure 7

Landscape of mutation genes in low and high risk groups. Waterfall plot showing mutation profiles of each gene in each endometrial cancer sample. A. Low risk group. B. High risk group.

Supplementary Files

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

- [FigureS1.pdf](#)
- [FigureS2.pdf](#)
- [FigureS3.pdf](#)
- [originaldata.docx](#)