

Construction and validation of a robust epigenetic gene-set based signature in ovarian cancer

Wenbin Shen (✉ swb1991@alumni.sjtu.edu.cn)

Fudan University Shanghai Cancer Center <https://orcid.org/0000-0002-1201-6443>

Wei Jiang

Fudan University Shanghai Cancer Center

Shuang Ye

Fudan University Shanghai Cancer Center

Min Sun

Fudan University Shanghai Cancer Center

Huijuan Yang

Fudan University Shanghai Cancer Center

Boer Shan

Fudan University Shanghai Cancer Center

Primary research

Keywords: Ovarian cancer, TCGA, GEO, epigenetic factors, signature

Posted Date: December 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-123005/v1>

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

[Read Full License](#)

Abstract

Background

Epigenetic factors play a critical role in tumor development and progression. The aim of this study was to construct and validate a robust epigenetic gene-set based signature for predicting prognosis of ovarian cancer (OC).

Methods

Public microarray data of OC patients from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases were identified and patients from TCGA dataset were randomized 3:1 into discovery and internal validation series. GSE14764 and GSE26712 from GEO database were combined as the external validation set. LASSO Cox regression model was performed in the discovery set to filter the most useful prognostic epigenetic factors.

Results

Based on LASSO Cox regression model, we built a 26 epigenetic factors based prognostic signature. In the discovery set, patients in high risk group showed significantly poorer overall survival than that patients in low risk group (HR: 2.11, 95% CI: 1.65–2.72, $P < 0.001$). The results were further validated in the internal validation set (HR: 1.69, 95% CI: 1.07–2.63, $P = 0.020$) and external validation set (HR 1.95, 95% CI 1.41–2.69; $p < 0.001$). Survival ROC at 5 year suggested that the epigenetic signature (AUC = 0.700) had better prognostic accuracy than any other clinicopathological factors in the entire cohort. In addition, survival decision curve analysis unveiled a considerable value of clinical utility of the epigenetic signature.

Conclusions

We successfully developed a robust epigenetic signature that can accurately predict prognosis in OC.

Introduction

Ovarian cancer (OC) is one of the leading cause of cancer associated death in women worldwide[1, 2]. For a lack of specific symptoms in the early stage, OC patients showed a dismal 25% five-year survival rate[3]. Consequently, more reliable prognostic biomarkers or predictive models are warranted to be identified to detect the subset of patients with higher mortality patients who deserve more aggressive therapeutic intervention.

Recent studies in many malignant tumors, including OC, suggested that multi-gene classifier or gene signature can make a good prediction of tumor prognosis and survival[4–10]. However, limited robust signature has been developed to predict mortality of OC. To date, due to the public access of gene expression data and biological process specific database, more efforts should be made to construct a prognostic gene signature with unique biological background in OC.

Epigenetic alterations referring to the heritable alterations in gene expression without variations in the DNA sequence, have been confirmed as critical regulatory factors during tumor development and progression[11–13]. Overall, the epigenetic modifiers can be categorized into the group of writers, readers, and erasers. The writers are responsible for the addition of chemical tags on DNA and histones. The readers are in charge of recognizing the specific modifications and the erasers play the role of wiping out these chemical modifications. A great deal of epigenetic alterations mediated by specific epigenetic factors have been proved as prognostic biomarker in OC[14–18]. However, no previous studies analyzed these epigenetic factors comprehensively to assess the prognostic value of the entire gene set of epigenetic factors. We hypothesized that the epigenetic factors gene set based prognostic signature might be of concrete predictive value in the detecting high risk patients with poor overall survival (OS).

In this study, a total of 827 OC patients with microarray data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases were identified to develop and validate a robust prognostic epigenetic signature. Based on appropriate sample-splitting and Cox regression analysis, we successfully constructed a prognostic epigenetic signature with high predictive ability and clinical utility.

Methods

Data source

Due to the samples with available RNA-Sequencing data were relatively small ($n = 374$), we collected Affymetrix Human Genome U133A Array microarray data instead ($n = 519$) from TCGA database (<https://genome-cancer.ucsc.edu>). Another two microarray datasets (GSE14764 and GSE26712) produced by the same Affymetrix Human Genome U133A platform were downloaded from GEO database to validate the result from TCGA. All probes were mapped based on their own EntrezGeneID[19]. When multiple probes were mapped to the same EntrezGeneID, the mean value was used to represent its average expression level. When different microarray datasets were combined to make analysis, the ComBat method was used to remove the potential batch effects.

Epigenetic factors corresponding to genes

A total of 720 genes corresponding to epigenetic factors were downloaded from EpiFactors database (<http://epifactors.autosome.ru/>).

Identification of prognostic epigenetic genes

level of *RB1*) + (-0.099 × expression level of *RSF1*) + (0.043 × expression level of *SETD8*) + (-0.152 × expression level of *SMARCAL1*) + (0.116 × expression level of *SMARCD1*) + (-0.042 × expression level of *TRIM24*) + (-0.298 × expression level of *TRIM27*) + (-0.099 × expression level of *WDR77*).

Table 1
Basic information of patients from TCGA and GEO datasets.

	TCGA		GEO	
	discovery set (N = 422)	internal validation (N = 140)	GSE14764 (N = 80)	GSE26712 N= (185)
FIGO stage				
I	11	4	8	late stage
II	20	7	1	late stage
III	323	109	69	late stage
IV	65	19	2	late stage
Debulking status				
Optimal	81	34	50	90
Suboptimal	296	90	26	95
Grade				
Low grade	57	18	26	0
Moderate/High grade	356	118	54	185

Prognostic value of epigenetic signature in discovery internal validation and external validation series

Based on the median risk score (value = 0.650), 211 patients were categorized into low risk group and 211 into high risk group. The distribution of death status and corresponding risk scores was illustrated in Fig. 2A (left panel) and it suggested that patients with lower risk scores generally had better OS than those patients with higher risk scores. Time-dependent ROC analyses at 5-year revealed a high prognostic accuracy of the epigenetic signature (Fig. 2A, middle panel). The survival rates for patients in low-risk group were 96.7% at 1 year, 74.3% at 3-year, and 47.0% at 5-year, compared with 87.8%, 49.4%, and 19.9% in patients in high-risk group, respectively (HR: 2.11, 95% CI: 1.65–2.72, $P < 0.001$, Fig. 2A, right panel).

According to the same cutoff value determined in discovery set, same analyses were further performed in the internal validation set. The 1 year, 3-year and 5-year OS were 91.4%, 70.8%, and 45.2% for the low-risk

group, and 87.9%, 62.9%, and 24.0% for the high-risk group (HR: 1.69, 95% CI: 1.07–2.63, P = 0.020, Fig. 2B).

To confirm that the epigenetic signature had similar prognostic value in external cohort, we combined the samples in GSE14764 and GSE26712 and a total of 265 OC patients were further used to validate the 26-mRNA epigenetic signature. Based on the established cutoff point, 178 (67.2%) patients were classified as low risk, and 87 (32.8%) as high risk. 5-year disease-free survival was 27.3% for the high-risk group and 53.2% for the low-risk group (HR 1.95, 95% CI 1.41–2.69; p < 0.001; Fig. 2C).

Prognostic performance and clinical use of the epigenetic signature in predicting OS

After multivariate analysis adjusted by clinicopathological variables, the epigenetic classifier remained a powerful and independent factor in the discovery, internal validation and external validation sets (Table 2). Stratified analysis suggested that the epigenetic classifier was still a clinically and statistically significant prognostic model in stage I-III, stage IV, low grade, high grade and patients with different debulking status (Fig. 3).

Table 2
Multivariate Cox regression analysis of epigenetic signature in predicting OS.

	HR	P*
Discovery set	14.32 (5.69–36.07)	< 0.001
Internal validation set	5.91 (1.37–25.52)	0.017
External validation set	25.26 (7.51–84.94)	< 0.001
*Adjusting for FIGO stage, grade debulking status		

To verify that the epigenetic signature bears stronger predictive ability than clinicopathological features in predicting OS, time-dependent ROC was performed in the entire cohort and the result showed that the epigenetic signature (AUC = 0.700) had significantly higher prognostic accuracy than debulking status (AUC = 0.635), FIGO stage (AUC = 0.505) and grade (AUC = 0.536) at 5 years (Fig. 4A). The decision curve analysis at 5 years was further used to analyze clinical usefulness of epigenetic signature. It suggested that using the epigenetic signature to predict OS brought more benefit than either the treat-all-patients scheme or the treat-none scheme, while the other clinicopathological features added strikingly less benefit (Fig. 4B).

Discussion

The past decades have witnessed the distinct oncogenic outcome in OC patients, which spans less than 5 months to over 10 years. However, no individualized treatment was used in those cases with

significantly poor long-term survival for a lack of predictive models helping to identify them accurately. Consequently, it is urgently needed to construct a prognostic tool which could accurately identify those patients with poor survival. In this study, a novel prognostic classifier based on 26 epigenetic genes was developed to improve the prediction of OS for OC patients after surgical resection. By applying the epigenetic signature to the TCGA discovery set patients, a clear separation was observed in the survival curves between low and high-risk patients. And it was internally validated in the validation series of TCGA patients and the external cohorts of GSE14764 and GSE26712, indicating the good reproducibility of this signature in OC. Stratified by FIGO stage, debulking status, and tumor grade, the epigenetic signature remains a good prognostic model, implying that the signature can be used to refining the current staging and risk evaluation system. In addition, the time-dependent ROC and DCA curve at 5 year suggested that this epigenetic signature has the most significant accuracy and clinical utility in predicting mortality after initial resection of OC. Therefore, our study identified an epigenetic signature that could help identify patients with high risk of mortality and guide individualized treatment of patients with OC, which is credible to be applied to clinic.

The epigenetic modifiers have been highlighted for critical functions in cancer initiation and progression. Accumulating evidence indicated that agents targeting epigenetic alterations have promising outcomes in retarding OC progression. However, no previous study has assessed a subset of epigenetic genes that collectively form an epigenetic signature and serve as an independent prognostic factor for OC. To date, numerous studies have developed gene expression signatures to stratify OC survival based on different cohorts. However, none of them has been incorporated into clinical practice might due to several limitations. First, the sample size was usually small that lacked sufficient validation to prove model stability. Second, most of them generate the prognostic score based on the whole transcriptome profiling, which easily led to overfitting and ignored other casual genes. In this present study, we focused on the epigenetic gene set that is more generic than the normal signature.

Most of genes included in the signature have been experimentally demonstrated to be linked with OC. BRD9 which was previously reported to regulate chromatin remodeling and transcription was found overexpressed in ovarian cancer and depleting BRD9 sensitizes cancer cells to olaparib and cisplatin[21]. PHF20 is a transcription factor, which was originally identified in glioma patients and its phosphorylation plays an important role in tumorigenesis via regulating of p53 mediated signaling in OC[22]. PADI4 is an enzyme that converts both histone arginine and mono-methyl arginine residues to citrulline. It was proved to be able to regulate the proliferation, apoptosis, invasion and migration of ovarian cancer cells[23]. EHMT1 and EHMT2 have roles in epigenetic silencing of gene expression and disrupting EHMT1/2 will sensitizes OC cells to PARPi. DDB2 is a kind of DNA damage-binding protein[24]. Previous studies suggested that DDB2 can repress OC cell dedifferentiation by suppressing ALDH1A1 expression[25]. CARM1, an arginine methyltransferase was proved as an informative prognostic biomarker for predicting resistance to chemotherapy for ovarian cancer[26]. CHD2, chromodomain helicase DNA binding protein could bind to microRNA-144 and partially inhibited its activity, thereby promoting the proliferative and migratory abilities of OC cells[27]. APOBEC3 DNA cytosine deaminase family members normally defend against viruses and transposons and its expression was found to be correlated with T-Cell Infiltration and

Improved Clinical Outcomes in OC based on public genetic data[28]. As for the rest epigenetic genes integrated in our signature, further clinical and basic research should be conducted to explore their value in OC.

Though this is the first study investigating the prognostic role of epigenetic genes and successfully developing a epigenetic signature in OC, some limitations are Inevitable in our study. Firstly, our study was based on the data from public-available datasets without testing prospectively in an independent cohort. Furthermore, the information of several other important clinicopathological features, like histology, was not available in these datasets. Finally, mechanisms of the identified epigenetic genes on the progression of OC are still needed to be further investigated.

Conclusion

In conclusion, we successfully construct a robust epigenetic signature that can reliably predict the emergence of death postoperatively in OC. A prospective validation should be conducted to confirm the prognostic and predictive role of the developed epigenetic signature.

Declarations

Acknowledgement:

We thank the TCGA and GEO databases for providing these valuable data.

Data and materials availability:

Source data and reagents are available from the corresponding author upon reasonable request.

Authors' contribution:

B.S. designed this study. W.S. and W.J. were responsible for the interpretation of the results. M.S. and H.Y. processed the data processing and performed the statistical analysis. W.S. wrote the manuscript. B.S. revised and improved the manuscript. All authors approved the submission.

Competing interests:

The authors declared that they have no competing interest.

Funding:

This article was funded by Shanghai Anticancer Association EYAS PROJECT: SACA-CY1B08

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2018.***CA Cancer J Clin* 2018, **68**:7-30.
2. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020.***CA Cancer J Clin* 2020, **70**:7-30.
3. Siegel RL, Miller KD, Jemal A: **Cancer Statistics, 2017.***CA Cancer J Clin* 2017, **67**:7-30.
4. Hinchcliff E, Paquette C, Roszik J, Kelting S, Stoler MH, Mok SC, Yeung TL, Zhang Q, Yates M, Peng W, et al: **Lymphocyte-specific kinase expression is a prognostic indicator in ovarian cancer and correlates with a prominent B cell transcriptional signature.***Cancer Immunol Immunother* 2019, **68**:1515-1526.
5. Shen S, Wang G, Zhang R, Zhao Y, Yu H, Wei Y, Chen F: **Development and validation of an immune gene-set based Prognostic signature in ovarian cancer.***EBioMedicine* 2019, **40**:318-326.
6. Muinao T, Deka Boruah HP, Pal M: **Diagnostic and Prognostic Biomarkers in ovarian cancer and the potential roles of cancer stem cells - An updated review.***Exp Cell Res* 2018, **362**:1-10.
7. Sun H, Cao D, Ma X, Yang J, Peng P, Yu M, Zhou H, Zhang Y, Li L, Huo X, Shen K: **Identification of a Prognostic Signature Associated With DNA Repair Genes in Ovarian Cancer.***Front Genet* 2019, **10**:839.
8. An Y, Bi F, You Y, Liu X, Yang Q: **Development of a Novel Autophagy-related Prognostic Signature for Serous Ovarian Cancer.***J Cancer* 2018, **9**:4058-4071.
9. Zhao Y, Yang SM, Jin YL, Xiong GW, Wang P, Snijders AM, Mao JH, Zhang XW, Hang B: **A Robust Gene Expression Prognostic Signature for Overall Survival in High-Grade Serous Ovarian Cancer.***J Oncol* 2019, **2019**:3614207.
10. Wang R, Ye XH, Zhao XL, Liu JL, Zhang CY: **Development of a five-gene signature as a novel prognostic marker in ovarian cancer.***Neoplasma* 2019, **66**:343-349.
11. Costa-Pinheiro P, Montezuma D, Henrique R, Jerónimo C: **Diagnostic and prognostic epigenetic biomarkers in cancer.***Epigenomics* 2015, **7**:1003-1015.
12. Okugawa Y, Grady WM, Goel A: **Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers.***Gastroenterology* 2015, **149**:1204-1225.e1212.
13. Takeshima H, Ushijima T: **Accumulation of genetic and epigenetic alterations in normal cells and cancer risk.***NPJ Precis Oncol* 2019, **3**:7.
14. Weberpals JI, Koti M, Squire JA: **Targeting genetic and epigenetic alterations in the treatment of serous ovarian cancer.***Cancer Genet* 2011, **204**:525-535.
15. Zeng M, Kwiatkowski NP, Zhang T, Nabet B, Xu M, Liang Y, Quan C, Wang J, Hao M, Palakurthi S, et al: **Targeting MYC dependency in ovarian cancer through inhibition of CDK7 and CDK12/13.***Elife* 2018, **7**.

16. Sharma A, Albahrani M, Zhang W, Kufel CN, James SR, Odunsi K, Klinkebiel D, Karpf AR: **Epigenetic activation of POTE genes in ovarian cancer.***Epigenetics* 2019, **14**:185-197.
17. Jones BA, Varambally S, Arend RC: **Histone Methyltransferase EZH2: A Therapeutic Target for Ovarian Cancer.***Mol Cancer Ther* 2018, **17**:591-602.
18. Salgado-Albarrán M, González-Barrios R, Guerra-Calderas L, Alcaraz N, Estefanía Sánchez-Correa T, Castro-Hernández C, Sánchez-Pérez Y, Aréchaga-Ocampo E, García-Carrancá A, Cantú de León D, et al: **The epigenetic factor BORIS (CTCF) controls the androgen receptor regulatory network in ovarian cancer.***Oncogenesis* 2019, **8**:41.
19. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.***Biostatistics* 2003, **4**:249-264.
20. Tibshirani R: **The lasso method for variable selection in the Cox model.***Stat Med* 1997, **16**:385-395.
21. Zhou Q, Huang J, Zhang C, Zhao F, Kim W, Tu X, Zhang Y, Newsheem S, Zhu Q, Deng M, et al: **The bromodomain containing protein BRD-9 orchestrates RAD51-RAD54 complex formation and regulates homologous recombination-mediated repair.***Nat Commun* 2020, **11**:2639.
22. Li Y, Park J, Piao L, Kong G, Kim Y, Park KA, Zhang T, Hong J, Hur GM, Seok JH, et al: **PKB-mediated PHF20 phosphorylation on Ser291 is required for p53 function in DNA damage.***Cell Signal* 2013, **25**:74-84.
23. Cui YY, Yan L, Zhou J, Zhao S, Zheng YB, Sun BH, Lv HT, Rong FN, Chang XT: **The role of peptidylarginine deiminase 4 in ovarian cancer cell tumorigenesis and invasion.***Tumour Biol* 2016, **37**:5375-5383.
24. Watson ZL, Yamamoto TM, McMellen A, Kim H, Hughes CJ, Wheeler LJ, Post MD, Behbakht K, Bitler BG: **Histone methyltransferases EHMT1 and EHMT2 (GLP/G9A) maintain PARP inhibitor resistance in high-grade serous ovarian carcinoma.***Clin Epigenetics* 2019, **11**:165.
25. Cui T, Srivastava AK, Han C, Wu D, Wani N, Liu L, Gao Z, Qu M, Zou N, Zhang X, et al: **DDB2 represses ovarian cancer cell dedifferentiation by suppressing ALDH1A1.***Cell Death Dis* 2018, **9**:561.
26. Nakayama N, Sakashita G, Nariai Y, Kato H, Sinmyozu K, Nakayama JI, Kyo S, Urano T, Nakayama K: **Cancer-related transcription regulator protein NAC1 forms a protein complex with CARM1 for ovarian cancer progression.***Oncotarget* 2018, **9**:28408-28420.
27. Sheng M, Wei N, Yang HY, Yan M, Zhao QX, Jing LJ: **CircRNA UBAP2 promotes the progression of ovarian cancer by sponging microRNA-144.***Eur Rev Med Pharmacol Sci* 2019, **23**:7283-7294.
28. Leonard B, Starrett GJ, Maurer MJ, Oberg AL, Van Bockstal M, Van Dorpe J, De Wever O, Helleman J, Sieuwerts AM, Berns EM, et al: **APOBEC3G Expression Correlates with T-Cell Infiltration and Improved Clinical Outcomes in High-grade Serous Ovarian Carcinoma.***Clin Cancer Res* 2016, **22**:4746-4755.

Figures

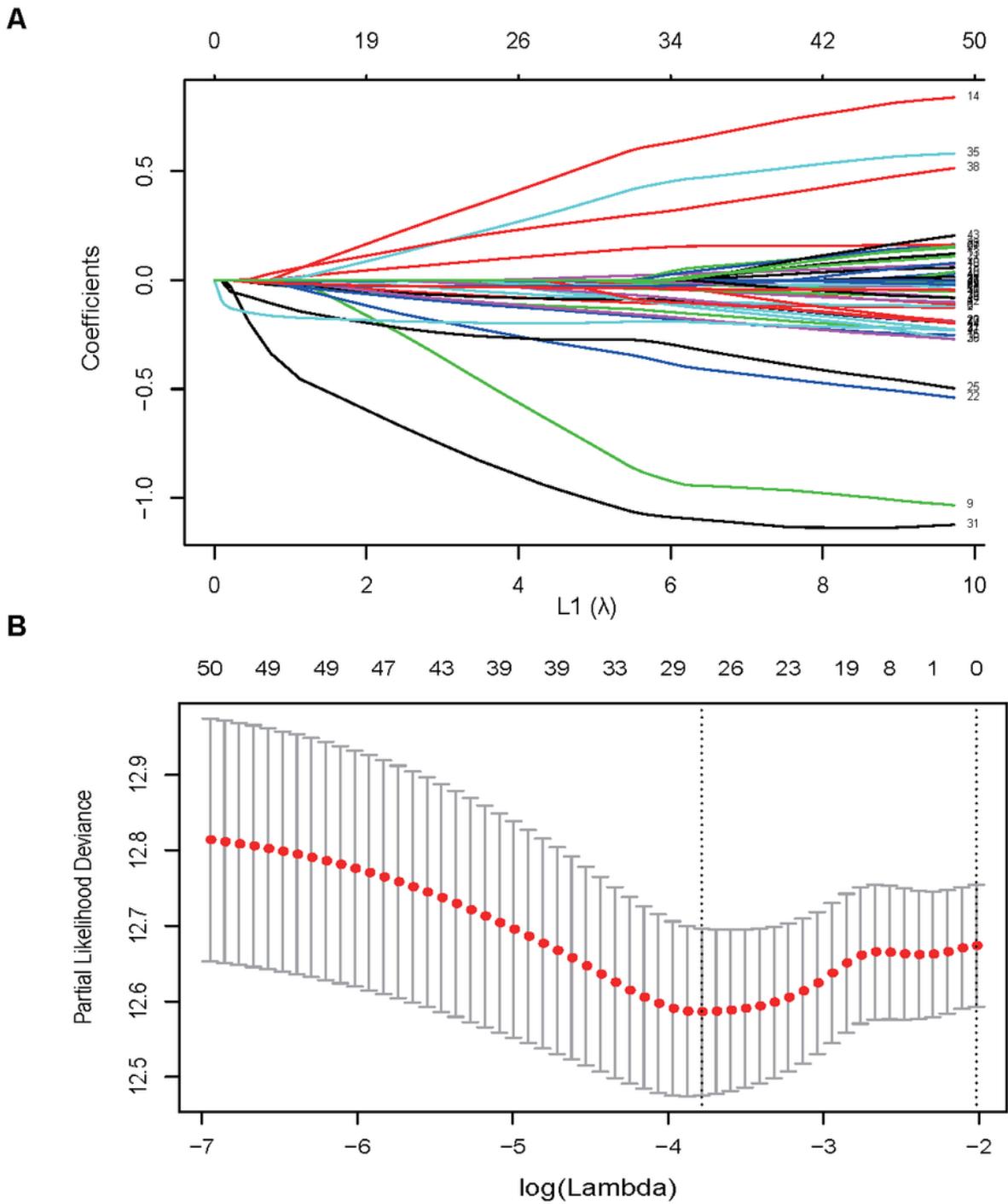


Figure 1

Figure 1

(A) LASSO coefficient profiles of the 51 OS associated epigenetic genes. (B) Tuning parameter (λ) selection in the LASSO model used 10-fold cross-validation via minimum criteria.

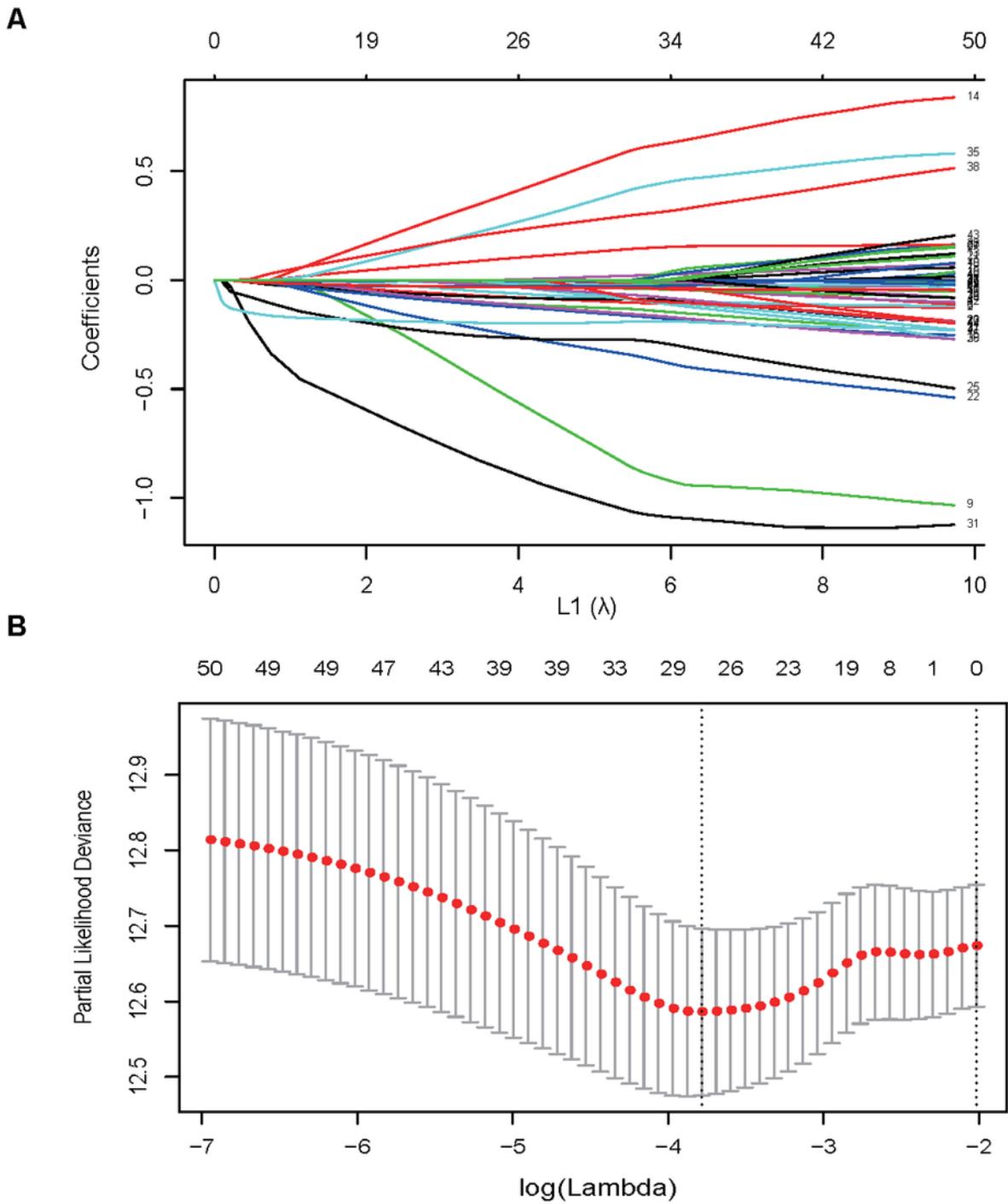
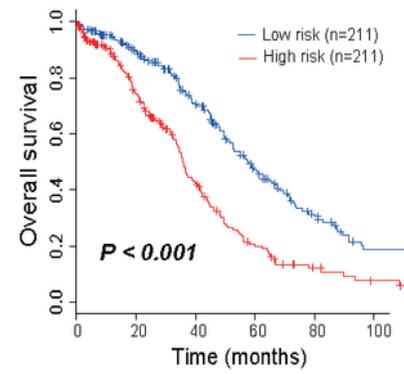
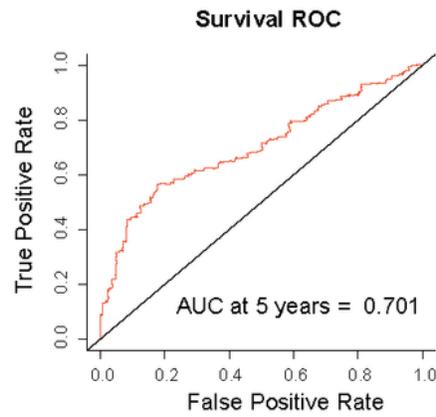
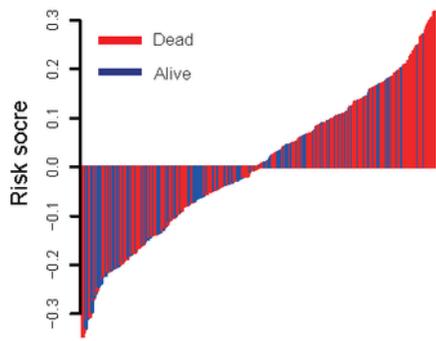


Figure 1

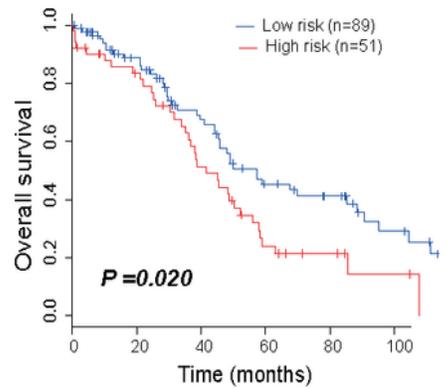
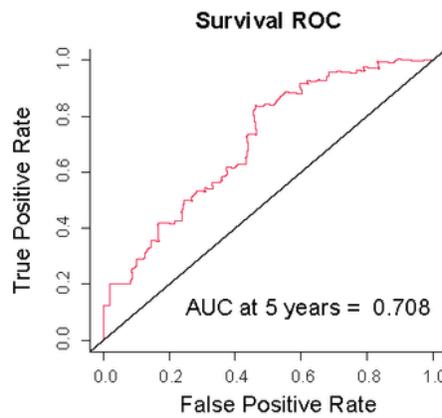
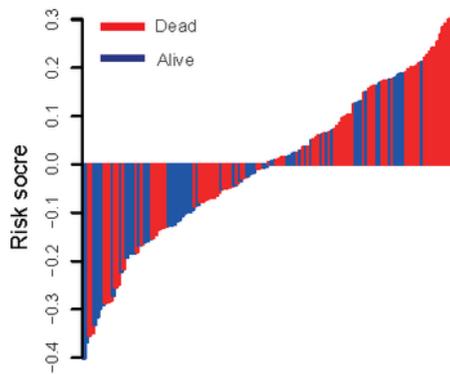
Figure 1

(A) LASSO coefficient profiles of the 51 OS associated epigenetic genes. (B) Tuning parameter (λ) selection in the LASSO model used 10-fold cross-validation via minimum criteria.

A. Training set



B. Internal validation set



C. External validation set

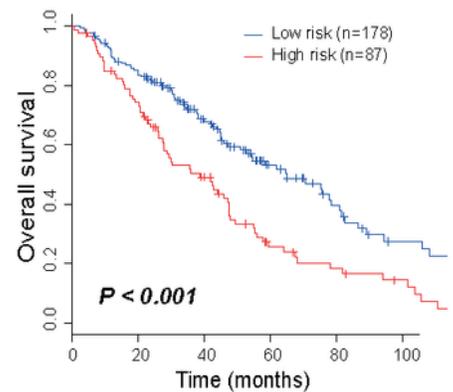
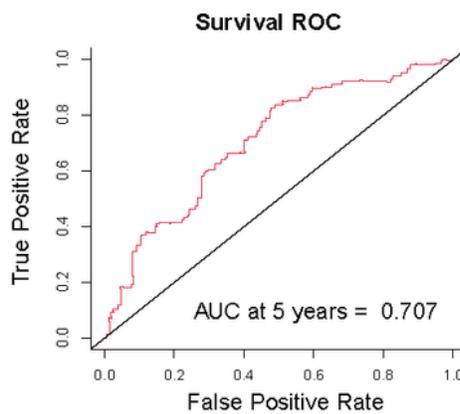
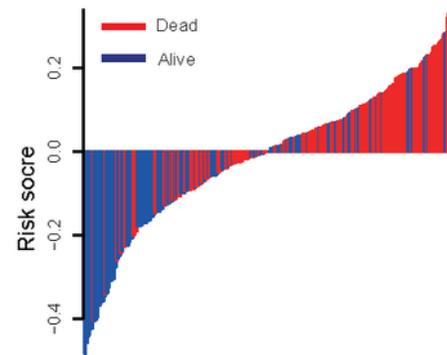
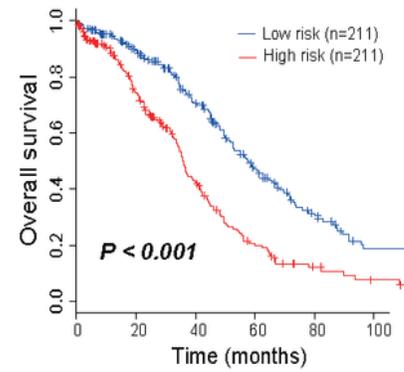
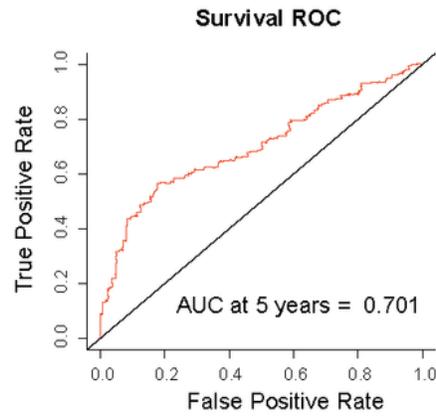
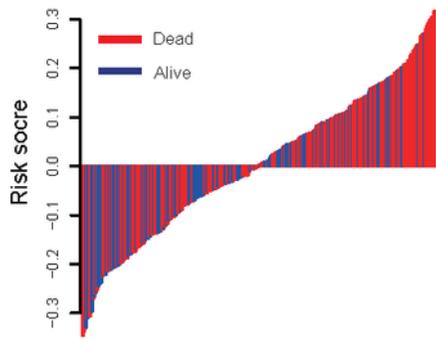


Figure 2

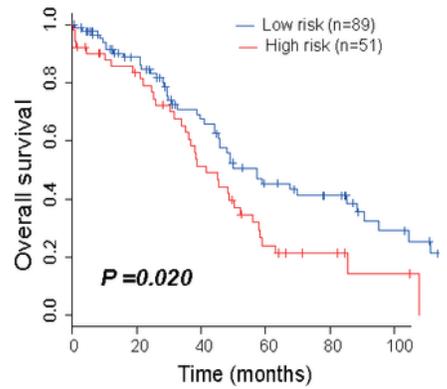
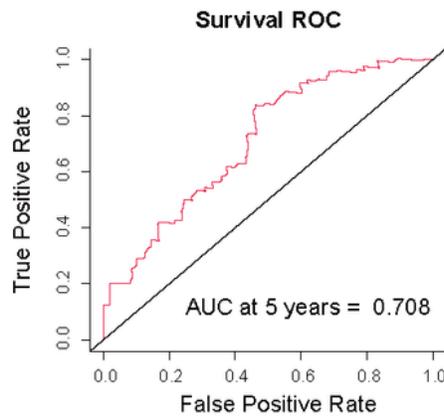
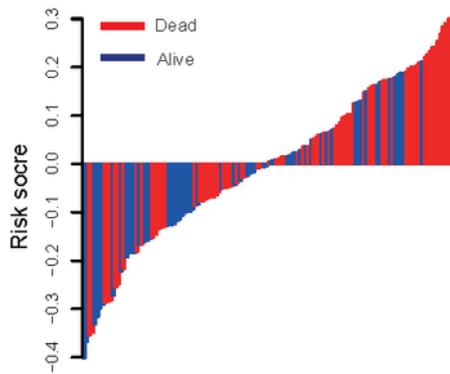
Figure 2

Distribution of risk score, time dependent ROC curves at 5 years and Kaplan-Meier survival curves between patients in low and high risk group in training set (A), internal validation set (B) and external cohort (C).

A. Training set



B. Internal validation set



C. External validation set

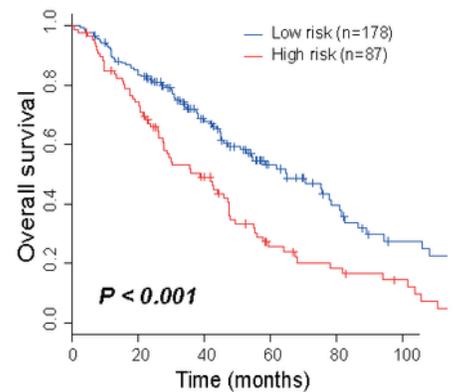
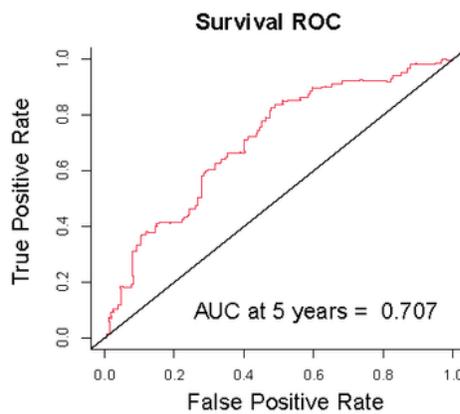
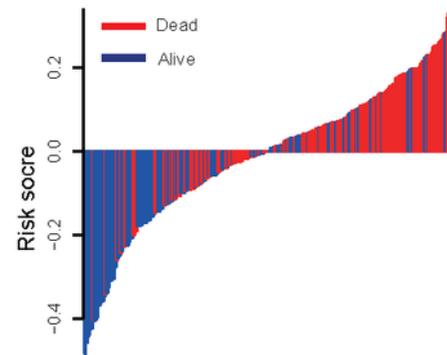
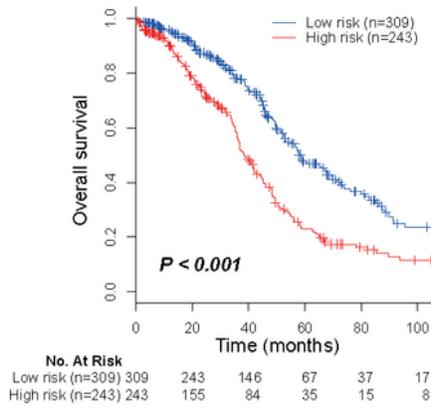
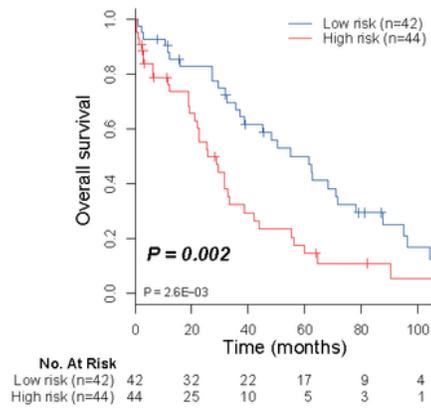
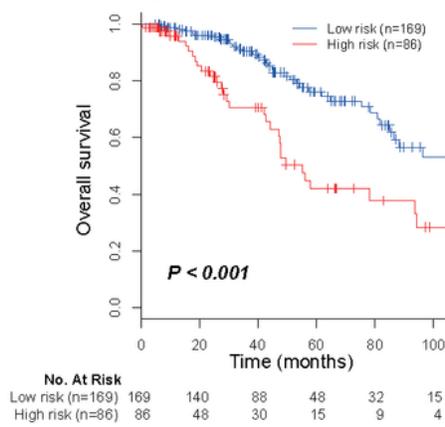
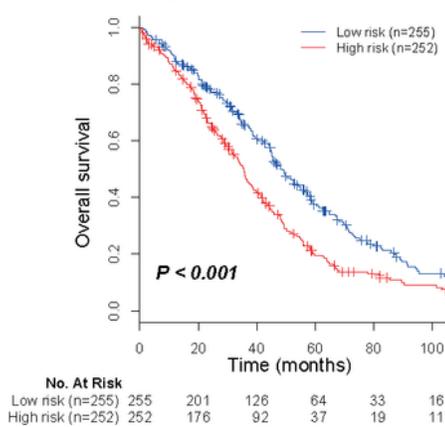
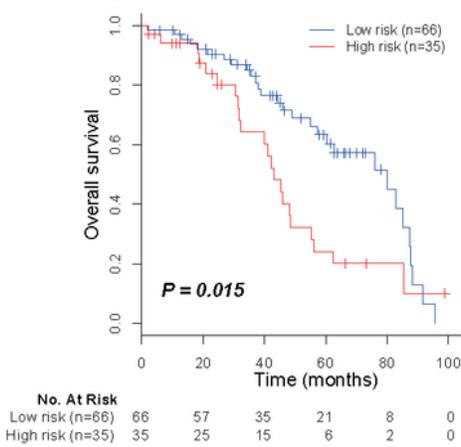
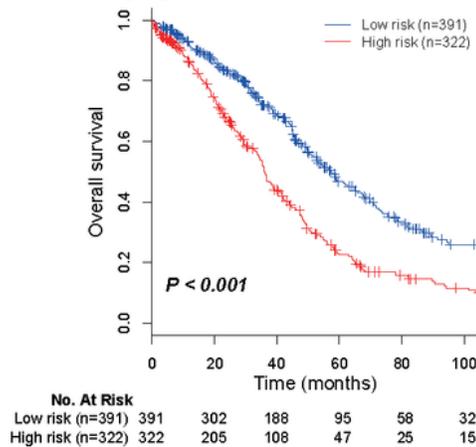


Figure 2

Figure 2

Distribution of risk score, time dependent ROC curves at 5 years and Kaplan-Meier survival curves between patients in low and high risk group in training set (A), internal validation set (B) and external cohort (C).

A. Stage I-III**B. Stage IV****C. Debulking status: Optimal****D. Debulking status: Suboptimal****E. Low grade****F. High grade****Figure 3****Figure 3**

Kaplan-Meier survival analysis for the entire dataset (N=827) based on the epigenetic signature stratified by FIGO stage (A), debulking status (B), and tumor grade (C).

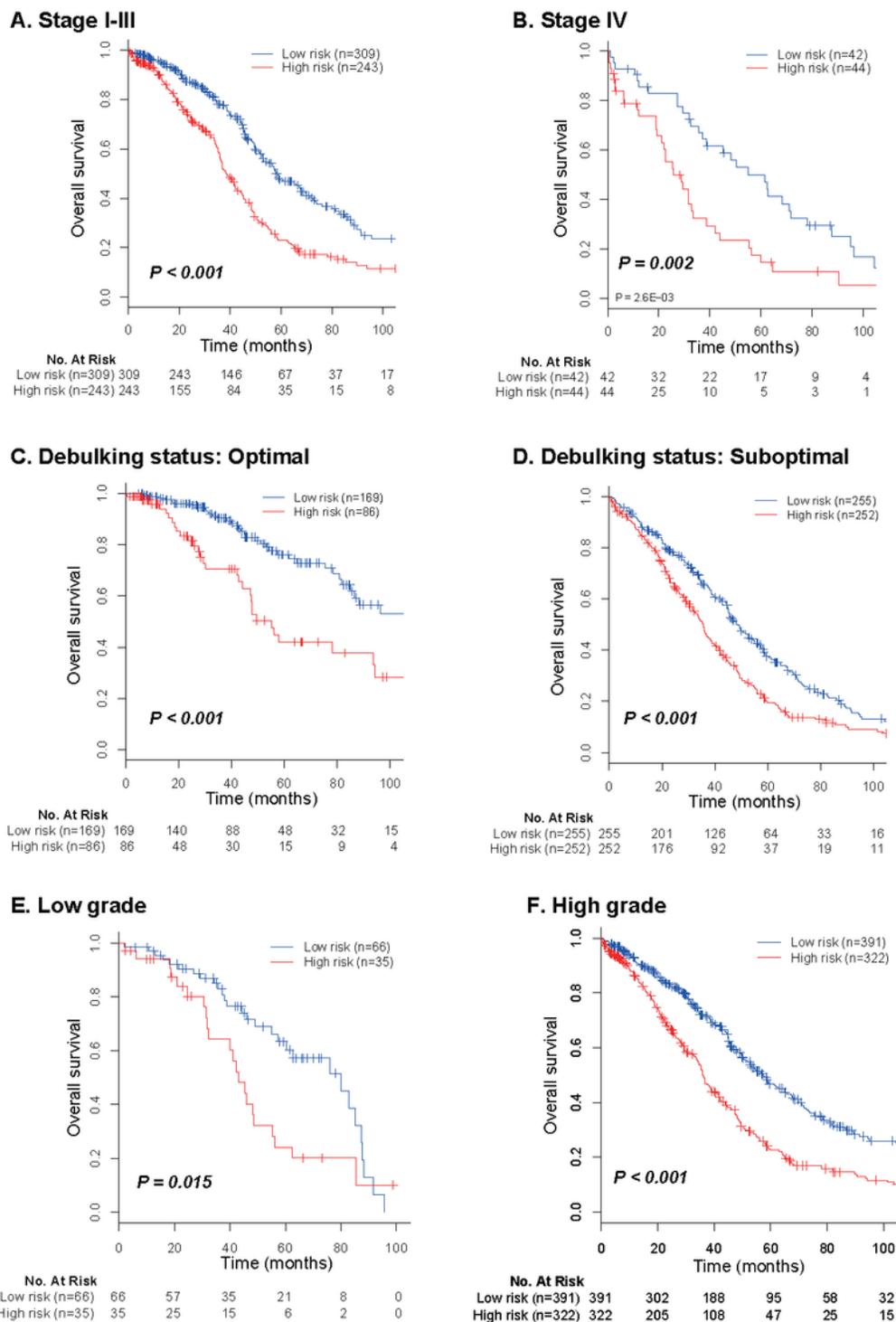


Figure 3

Figure 3

Kaplan-Meier survival analysis for the entire dataset (N=827) based on the epigenetic signature stratified by FIGO stage (A), debulking status (B), and tumor grade (C).

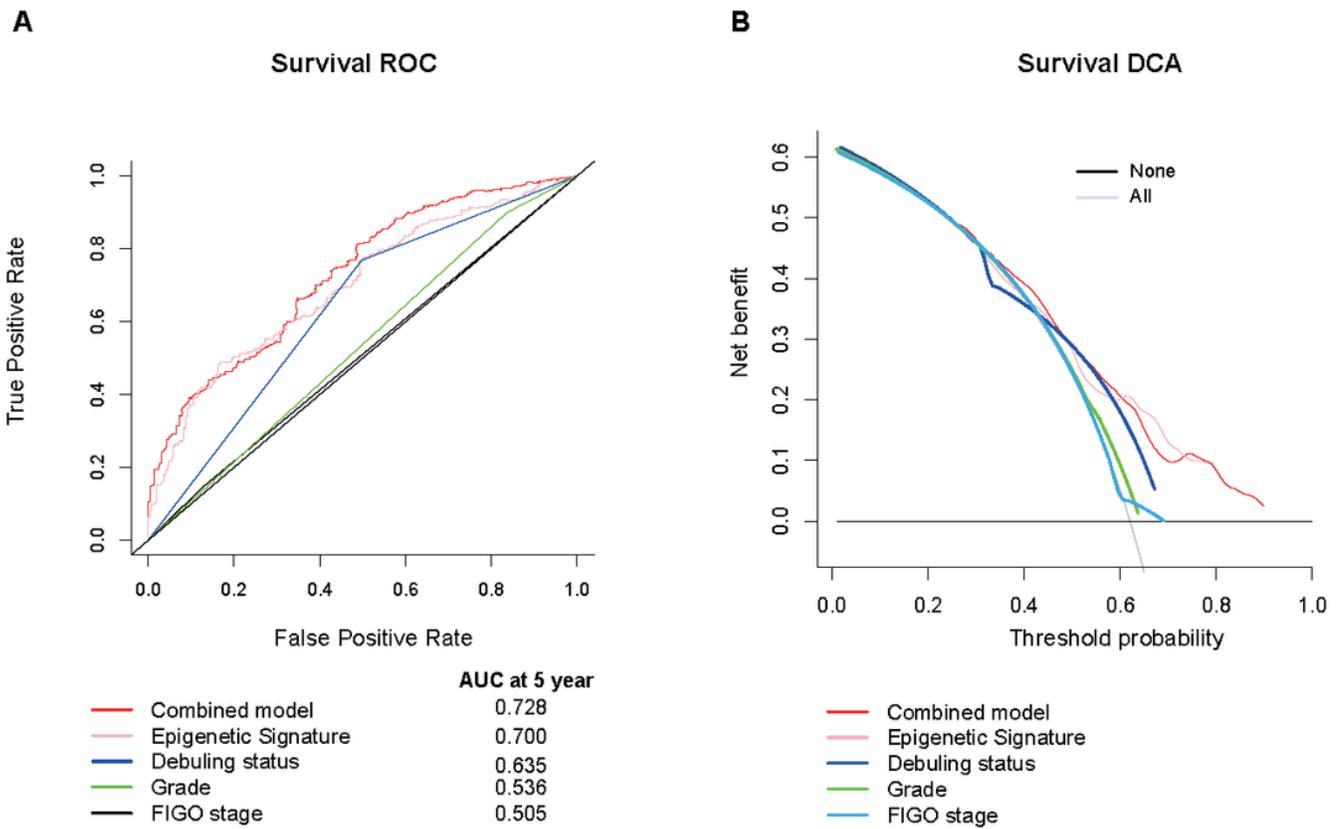


Figure 4

Figure 4

(A) Time-dependent ROC curves at 5 year of the epigenetic signature, debulking status, FIGO stage and tumor grade in the entire cohort; Decision curve analysis at 5 year for the epigenetic signature, debulking status, FIGO stage and tumor grade and the combined model (B). The y-axis measures the net benefit.

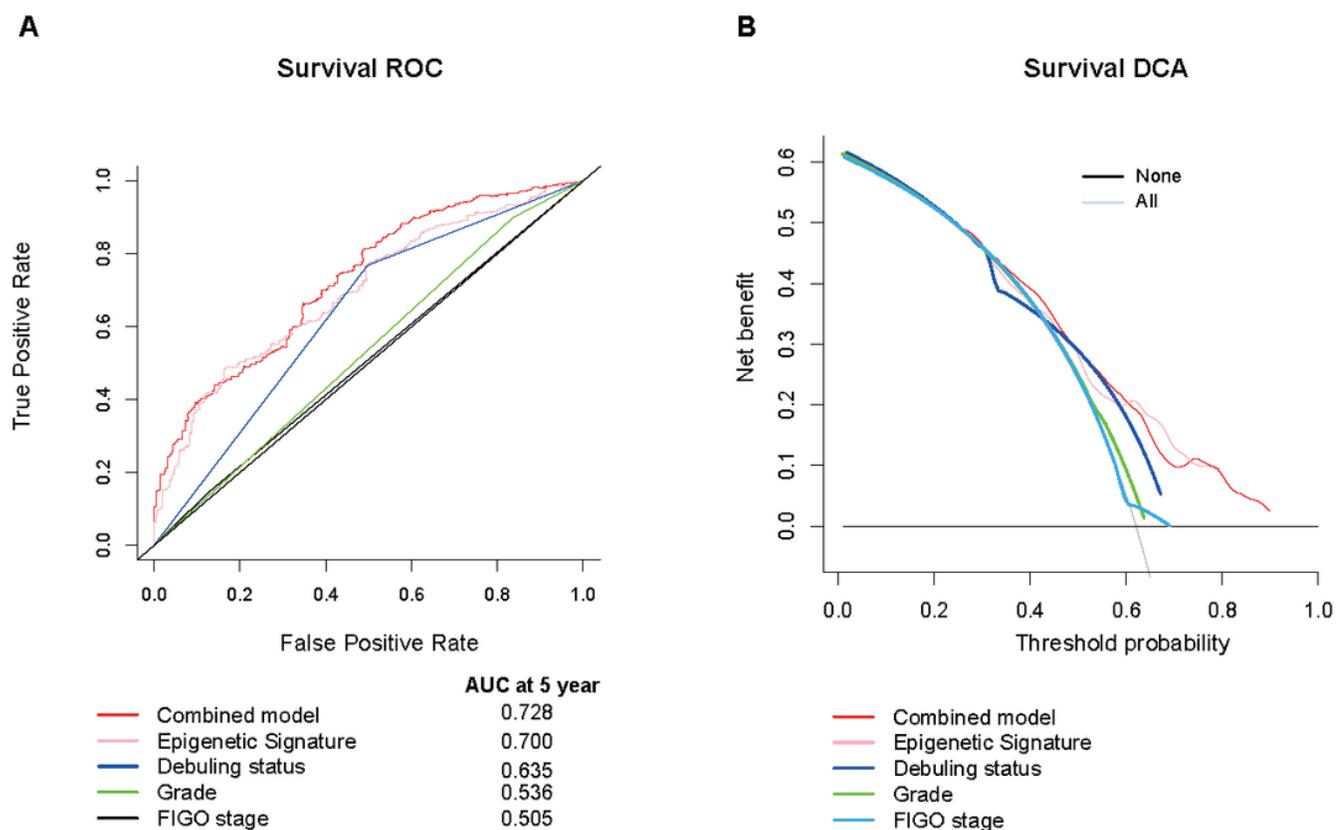


Figure 4

Figure 4

(A) Time-dependent ROC curves at 5 year of the epigenetic signature, debulking status, FIGO stage and tumor grade in the entire cohort; Decision curve analysis at 5 year for the epigenetic signature, debulking status, FIGO stage and tumor grade and the combined model (B). The y-axis measures the net benefit.

Supplementary Files

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

- [TableS1.xlsx](#)
- [TableS1.xlsx](#)