

Identification and validation of a novel prognosis prediction model in adrenocortical carcinoma by integrative bioinformatics analysis and statistics and machine learning

Xin Yan

Wuhan University Zhongnan Hospital

Zi-Xin Guo

Wuhan University Zhongnan Hospital

Dong-Hu Yu

Wuhan University Zhongnan Hospital

Chen Chen

Wuhan University Zhongnan Hospital

Xiao-Ping Liu

Wuhan University Zhongnan Hospital

Jie Yao

Wuhan University Zhongnan Hospital

Tong-Zu Liu

Wuhan University Zhongnan Hospital

Sheng Li (✉ lisheng-znyy@whu.edu.cn)

Wuhan University Zhongnan Hospital <https://orcid.org/0000-0003-4070-7345>

Primary research

Keywords: adrenocortical carcinoma, bioinformatics, machine learning, hub genes, prognosis prediction model

Posted Date: November 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-101881/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Adrenocortical carcinoma (ACC) is a rare malignancy with poor prognosis. Thus, we aimed to establish a gene signature to predict the prognosis for ACC.

Methods: Firstly, “WGCNA” package was used to construct a co-expression network and screen key module. Secondly, survival associated genes were identified by performing survival analysis. Thirdly, regression models were constructed by using the Ridge, ELASTIC-NET, and LASSO methods. Time-dependent ROC analysis, Cox regression analysis, GSEA, DCA and nomogram were performed to validate the model. Fourthly, mutations and CNVs of genes in the model were explored. Finally, LDA, KNN, SVM, PPI network and meta-analysis were used screened and validated meaningful prognostic biomarkers.

Results: Two key modules were selected and 93 survival associated genes were identified. Furthermore, 11 models were constructed and two models were further selected, which were validated in each dataset (training set, internal validation set, GSE19750, and GSE76021). Model 2 was further identified as the best model (training set: survival analysis: $p < 0.0001$; AUC: 0.92 at 1 year, 0.91 at 3 years and 0.95 at 5 years). In genes in the best model, MKI67 was altered most (12%). Six hub genes were further analyzed by constructing a PPI network and validated by meta-analysis.

Conclusion: In summary, we constructed and validated a prognostic multi-gene model and six powerful prognostic biomarkers, which might be useful instruments for predicting the prognosis of ACC patients.

Background

Adrenal cortical carcinoma (ACC) is a relatively rare malignant tumor[1]. 40% of the cases had distant metastasis at the time of diagnosis[2]. Worst of all, most patients with ACC have to face poor prognosis. The morbidity age of ACC showed bimodal distribution, good hair in kids and 40 to 50-year-old people[3]. As for the prognosis of ACC patients, the five-year survival rate for ACC patients was as low as 35% according to a recent research[2]. At present, many studies identified molecular biomarkers associated with ACCs to solve this problem[4]. It means that predicting survival and prognosis of patients with the help of biomarkers has great application prospects. Thus, we aimed to construct a multi-gene signature and some prognostic biomarkers, which perhaps concentrate on predicting the prognosis of patients with ACC.

In the present study, by using GSE76019, a co-expression network was constructed by using R package “WGCNA” (weighted correlation network analysis). Key modules related to survival state were selected. Then we screened out hub genes in co-expression network and further identified survival associated genes among them by using GEPIA (Gene Expression Profiling Interactive Analysis)[5]. Considering that Ridge, ELASTIC-NET, and LASSO (least absolute shrinkage and selection operator) methods were widely used data mining methods especially for regression model construction. Based on survival associated genes identified by WGCNA, all the three methods were included for constructing multiple-gene signatures by using the data from TCGA database. After finishing all the steps, totally 11 models were built and two candidate best models were screened out among them by calculating AUC (area under curve) using internal validation set. Both of the two models could significantly predict the OS (Overall survival) of patients with ACC, validated by using three independent datasets (internal validation set, GSE19750 and GSE76021). Moreover, we constructed two nomograms based on the two multiple-gene signatures, which were verified to predict the probability of the overall survival of patients with ACC well. The calibration curve indicated that nomogram of model 2 showed better prognostic value compared with this of model 1, which was regarded as the best prognostic model. Linear discriminant analysis (LDA), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) methods were used to verify the prognostic values of genes in the best model. We further explore the mutations and copy number variations (CNVs) of genes in the best model and demonstrated that CNVs of these genes were significantly associated with relative mRNA expression levels and survival of patients with ACC. Moreover, six genes were further screened out by protein-protein interaction (PPI) network construction, which were regarded as hub genes in this study. Furthermore, meta-analysis was performed for validation of these hub genes. The potential functions of the best model and hub genes were explored by performing Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and gene set enrichment analysis (GSEA). In conclusion, we constructed a multi-gene signature which might be an independent survival and prognostic biomarker for ACCs. And six hub genes were also screened out, which might be powerful prognostic biomarkers of patients with ACC.

Materials And Methods

Data collection

For WGCNA analysis, dataset GSE76019[6] was downloaded from Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) to construct co-expression network. GSE76019, performed on Affymetrix HT HG-U133 + PM Array Plate, included 34 ACCs.

To search the best regression model by using the Ridge ($\alpha = 0$), ELASTIC-NET (with α varying from 0.1 to 0.9), and LASSO ($\alpha = 1$) methods, we downloaded the mRNA-seq data and clinical information of ACCs from The Cancer Genome Atlas (TCGA) database. After weeding out incomplete samples, 79 ACCs were included in this study. We randomly split them into two groups in the ratio of 2:1. The training set including 52 samples was selected to construct regression models, meanwhile the rest 27 samples were used to preliminarily screen the best model and further validate our results (as an internal validation set). We also downloaded two independent datasets for external validation. GSE19750[7] that performed on

Affymetrix Human Genome U133 Plus 2.0 Array included 44 ACCs and 4 normal adrenal glands. Dataset GSE76021[8], performed on Affymetrix Human Genome U133A Array, included 29 ACC samples.

To explore mutations and copy number variations (CNVs) of genes in the best model, we also downloaded the CNV data of ACC samples from TCGA database.

Data preprocessing

A workflow of the study was shown in Fig. 1. For WGCNA analysis, 34 adrenocortical carcinoma samples from GSE76019 were included for WGCNA analysis. Probesets were filtered by their variances across all samples, only probes with variances ranked in top 2,500 were selected for subsequent analyses. Microarray quality was assessed by sample clustering according to the distance between different samples in Pearson's correlation matrices and average linkage.

Moreover, we used "DEseq2"[9] in R for normalization and \log_2 transformation for TCGA-ACC data which displayed as count number. The expression values of the genes we identified in WGCNA were extracted to construct the regression models by using the Ridge, ELASTIC-NET, and LASSO methods. As for GSE19750 and GSE76021, the normalization and \log_2 transformation were performed by using method "rma"[10] based on R package "affy".

Co-expression network constructing and key module screening

Firstly, gsg (goodSamplesGenes) and sample network methods were used to check the expression data profile of the top 2,500 genes from GSE76019. Samples with $Z.Ku < -2.5$ were considered as array outliers, which were not included for further analysis. Based on "WGCNA" in R software, we constructed a co-expression network. In order to classify genes into gene modules, we used three independent branch cutting methods (manual (interactive) branch cutting approach, automatic single block analysis and 2 block analysis) in WGCNA. Parameters for branch splitting were set as follows: minClusterSize = 30, and deepSplit = 2. Moreover, a cu line for merging high associated models was set through calculating the dissimilarity of module eigengenes (MEs). In order to identify key modules correlated to survival state (the trait interested us most), we firstly calculated the Gene Significance (GS) with the aim of quantifying the correlation between genes and trait. Then we further defined the module significance (MS), which was the average GS of all the genes in a module. Based on above analyses, the most positive correlation module and the most negative correlation module were regarded as key modules.

Functional enrichment analysis

To explore the potential functions of genes in key modules, Gene Ontology (GO) enrichment analysis[11] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis[12] were performed by using "clusterProfiler"[13] in R software. For the GO part, we only identified the biological process (BP). Gene sets were significantly enriched only when $P < 0.05$.

Survival associated gene identification

In the present study, genes in the key modules which reached the standard of $|cor.geneModuleMembership| > 0.8$ and $|cor.geneTraitSignificance| > 0.2$ were regarded as hub genes in co-expression network. In order to find out genes associated with survival, we further performed overall survival (OS) analysis and diseases-free survival (DFS) analysis for these hub genes by using GEPIA[5]. Genes were considered to be survival associated genes when they showed significant P value in both OS and DFS.

Regression model construction and the best model screening

Based on the training set, the expression values of survival associated genes were extracted to construct models by using the Ridge, ELASTIC-NET, and LASSO methods. All of the three methods were performed by using R package "glmnet"[14]. A total of 11 models would be constructed. We further got a gene signature from each model which contained the most useful prognostic biomarkers the model thought. To identify the best model among the 11 models, we calculated the risk score (RS) of every samples in the internal validation set based on these models. In the internal validation set, we used the risk score of each model to represent the model and further worked out the AUC to distinguish alive ACCs from dead ACCs by performing receiver operating characteristic curve (ROC) analysis by using R package "plotROC"[15]. The model with the highest AUC was regarded as the best model, which was included for further analysis.

Survival analysis and time-dependent ROC analysis

We further worked out the risk score of each samples in all datasets based on the best model. ACCs in training set, internal validation set, entire set, GSE19750 and GSE76021 were divided into high- and low- risk groups according to their own median risk score. Then we used R package "survival"[16] to performed survival analysis. In order to investigate the performance of the signature, R package "timeROC"[17] was used to perform time-dependent (1-year, 3-year, and 5-year) ROC analysis. The AUC was also performed to predictive accuracy.

Cox proportional hazards regression analysis

To verify the prognostic value of the gene signature (best model), we included the risk score of the best model and other important clinical features (gender, age, tumor stage, T stage, N stage, M stage, and Weiss score) for univariable Cox analysis of overall survival (OS) by using TCGA-ACC data.

Factors were included for multivariate Cox analysis when P value < 0.05 . To do this, we could make sure if this gene signature was irrelevant to other clinical features for predicting OS of ACCs. "We used R package "forestplot"[18] in R package was applied to visualization.

Gene set enrichment analysis (GSEA)

In order to understand the lurking functions of the best multi-gene signature, we evaluated the median risk score based on GSE76019 (the dataset which was used for WGCNA analysis). After that, totally 34 ACCs were divided into two groups accurately (high-risk group and low-risk group). "c2.cp.kegg.v6.2.symbols.gmt" was set as the reference gene sets. GSEA[19] was conducted between the two groups. In this study, KEGG signaling pathways reached the standards (nominal $P < 0.05$, $|ES| > 0.6$, gene size ≥ 100 and FDR $< 25\%$) were significantly enriched.

Decision curve analysis (DCA) and nomogram construction

Beyond ROC, we also used R package "rmda" to perform DCA to make sure if the signature was of great value for clinical application by using TCGA-ACC data[20]. Moreover, based on TCGA-ACC data, "rms" in R was used to construct the nomogram[21]. Calibrate curve was plotted to test the nomogram. The calibrate curve was also used for visualization and the 45° line in the curve represents the best prediction. Moreover, we calculated the consistency index (C-index) between actual probability and predicted probability to evaluate the prediction effectiveness of the nomogram[22]. Cross-validation could deal with over-fitting of the model. Thus, we performed cross-validation before nomogram construction.

Mutations and CNVs of genes in the best model

To explore mutations and CNVs of genes in the best model, we included all the ACC samples with CNV data from TCGA database for this analysis. Based on CBio Cancer Genomics Portal (<http://www.cbiportal.org/>), we first explored the genetic alterations of these genes. Furthermore, we identified the association between genetic alterations and the clinical features of patients with ACC. We used chi-square or Mann-Whitney U test to analysis the statistical significance of the result. Combined with relative mRNA expression levels of genes in the best model, we also explored the relationship between CNVs and mRNA expression levels of these genes. One-way analysis of variance (ANOVA) or Kruslal-Wallis test were used to test the result. Moreover, the association between CNVs of genes in the best model and survival of patients with ACC was also identified by performing survival analysis based on R package "survival".

Linear discriminant analysis (LDA), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM)

To verify the prognostic value of genes in the best model, we regarded genes as variables and relative mRNA expression values as variable values. Then we performed LDA, KNN, and SVM analyses. We used package "MASS"[23] in R software for LDA. For KNN algorithm, we chose the best K parameter by adopting cross validation method based on R package "caret"[24]. Then we performed KNN algorithm by using R packages "class"[23] and "kknn"(Hechenbichler and Schliep, 2004). For SVM analysis, four kinds of algorithms were used in the present study based on "e1071"(Meyer, 2015) in R software (linear-SVM, polynomial-SVM, radial basis function (RBF) SVM, and sigmoid-kernel SVM). R package "kernlab"(Karatzoglou, 2004) was used to assist with SVM feature selection. We used the data from TCGA as training set to first construct classifier in each analysis, and further verify the value of the classifier by calculate the classification rate based on TCGA-ACC data, GSE19750, GSE76019, and GSE76021. In this study, we regarded genes in the best model as meaningful prognostic biomarkers when the accuracy of classification ≥ 0.50 .

Protein-protein interaction (PPI) network construction to screen hub genes in the best model

To find out some relative genes in the best model, we first constructed a PPI network of genes formed this model by using the Search Tool for the Retrieval of Interacting Genes (STRING)[25]. Then we imported the PPI network into Cytoscape software (<https://cytoscape.org/>) and further calculated the degree of connectivity of each gene. In this study, genes with the highest degree were regarded as hub genes. We also analyzed and visualized the correlation among genes in the best model by performing Spearman analysis using R package "corrplot"(Wei, 2017) and "PerformanceAnalytics"[26].

Results

Constructing a co-expression network to identify key modules

We firstly included 33 ACC samples to perform WGCNA analysis after weeding out the outlier samples (Figure S1). According to the result, we regarded the beta (β) = 5 (scale free $R^2 = 0.89$) as soft-thresholding for further adjacencies calculation (Figure S2). Eventually, totally 8 modules were identified after classifying genes into gene modules and merging modules as shown in Figure S3. Genes not including in any other significant modules were included in the grey module, which were removed from subsequent analysis. Among the 8 modules, the blue module was the most associated module with survival status positively ($P = 5e-05$, $R^2 = 0.65$), meanwhile the red module was the most associated module negatively ($P = 4e-06$, $R^2 = -0.71$) (Fig. 2A). As shown in Fig. 2B, the MS of the two modules were also higher than MS of any other module. The relationships between MM and GS in the blue module ($cor = 0.64$) and red module ($cor = 0.74$) were also significant suggested by Fig. 2C-D. Thus, we considered blue and red modules to be key modules in this study. As shown in Figure S4A-B, we also created a network heatmap and a classical MDS plot.

Genes in key modules associated biological pathways

According to GO biological processes analysis, genes in red module were enriched in 4 BPs including leukocyte migration, positive regulation of cell migration, regulation of G-protein coupled receptor protein signaling pathway, and chemokine-mediated signaling pathway (Fig. 3A). As for genes in blue module, totally they were enriched in 197 BPs (Table S1). The top 10 were organelle fission, nuclear division, chromosome segregation, nuclear chromosome segregation, mitotic nuclear division, sister chromatid segregation, mitotic sister chromatid segregation, microtubule cytoskeleton organization involved in mitosis, regulation of chromosome segregation, and mitotic spindle organization (Fig. 3B). When talking about KEGG pathway analysis, genes in blue module were enriched in only 3 KEGG pathways including cell cycle, oocyte meiosis, and p53 signaling pathway (Fig. 3C).

Survival associated gene identification

With the cut-off criterion ($|cor.geneModuleMembership| > 0.8$ and $|cor.geneTraitSignificance| > 0.2$), 107 genes (12 genes in red module and 95 genes in blue module) were identified in co-expression network in total. Then we performed OS and DFS analysis for these 107 genes by using GEPIA. According to the results, high expression levels of 95 genes were obviously associated with poor OS for ACCs. Meanwhile, 100 genes showed significant *P* value in DFS analysis. 93 genes were common among them, which were regarded as survival associated genes (Table S2).

Establishment of two multi-gene signatures for predicting OS

We identified 93 biomarkers significantly related to survival of ACCs by using WGCNA analysis. With the aim of setting up a multiple-gene signature to predict the prognosis of patients with ACC, we calculated the relative regression coefficient of each genes by using the Ridge, ELASTIC-NET, and LASSO methods. The most powerful prognostic markers of the 93 biomarkers were screened out in each model as shown in Fig. 4. After that, 11 prognostic models were constructed and the relative regression coefficients of the most powerful prognostic markers were shown in Table S3. Combine the relative expression levels of the mRNA in the classifier and corresponding LASSO coefficients, we first calculated the risk score of each sample in the internal validation set. Furthermore, we worked out the AUC based on the risk score (Table 2). Two models contained model 1 (AUC = 0.676; $\alpha = 0.1$) and model 2 (AUC = 0.676; $\alpha = 0.6$) were finally identified, which were thought as candidate best models.

Table 1
Clinical characteristics of patients with ACC in each study.

Characteristics	Training dataset	Internal validation dataset	Entire TCGA dataset	GSE76019	GSE19750	GSE76021
Number of patients	52	27	79	34	44	29
Age(median, IQR)	48.5(31, 58)	52(36, 62)	50(35, 60)	0	52.5(41,58.5)	0
≥ 65	7	4	11	0	6	0
< 65	45	23	68	0	34	0
NA	0	0	1	34	4	29
Gender						
Male	19	12	31	0	17	0
Female	33	15	48	0	26	0
NA	0	0	0	34	1	29
Stage						
I	5	4	9	10	7	12
II	23	14	37	7	9	5
III	9	7	16	10	4	10
IV	14	1	15	7	8	2
NA	1	1	2	0	16	0
Grade						
1	0	0	0	0	7	0
2	0	0	0	0	9	0
3	0	0	0	0	4	0
4	0	0	0	0	8	0
NA	52	27	79	34	16	29
ACC: adrenocortical carcinoma; IQR: interquartile range; NA: not available.						

Table 2
AUC of risk scores calculated by the Ridge ($\alpha = 0$), ELASTIC-NET (with varying from 0.1 to 0.9), and least absolute, shrinkage, and selection operator (LASSO; $\alpha = 1$) methods by using internal validation set.

Models	Alive vs Dead	
	AUC	95% CI
Ridge ($\alpha = 0$)	0.670	0.445–0.896
ELASTIC-NET ($\alpha = 0.1$)	0.676	0.447–0.905
ELASTIC-NET ($\alpha = 0.2$)	0.670	0.440–0.901
ELASTIC-NET ($\alpha = 0.3$)	0.636	0.404–0.869
ELASTIC-NET ($\alpha = 0.4$)	0.574	0.347–0.801
ELASTIC-NET ($\alpha = 0.5$)	0.574	0.349–0.798
ELASTIC-NET ($\alpha = 0.6$)	0.676	0.457–0.895
ELASTIC-NET ($\alpha = 0.7$)	0.563	0.333–0.792
ELASTIC-NET ($\alpha = 0.8$)	0.551	0.323–0.780
ELASTIC-NET ($\alpha = 0.9$)	0.551	0.326–0.776
LASSO ($\alpha = 1$)	0.534	0.308–0.760

In model 1 and model 2, ACCs in training set were further split into two groups (low-risk, and high-risk) according to median-risk score (5.0406 in model 1 and 16.5329 in model 2). In both of the two models, high-risk group was associated with poor OS of patients with ACC (Model 1: $P < 0.0001$, Fig. 5A; model 2: $P < 0.0001$, Fig. 5D). We also performed time-dependent ROC analysis. The results suggested that the AUC values of model 1 in the training set were 0.95 at 1 year, 0.98 at 3 years and 0.96 at 5 years (Fig. 5A). Meanwhile the AUC values of model 2 in the training set were 0.92 at 1 year, 0.91 at 3 years and 0.95 at 5 years (Fig. 5D).

Validation of model 1 and model 2

In the present study, we used 3 validation sets (inter validation set, GSE19750, and GSE76021) to validate the results we got from training set. With the same method we mentioned in training set, we calculated the risk score for each sample in these three validation sets based on model 1 and model 2. Samples in internal validation set, GSE19750, and GSE76021 were split into two groups relying on the median-risk score in each set. In model 1, according to the survival analysis, the high-risk group had lower survival rate compared with the low-risk group in all the validation sets (internal validation set: $P = 3E-04$; GSE19750: $P = 0.011$; GSE76021: $P = 0.012$) as the training set suggested (Fig. 5B, Fig. 6A-B). As for the results of time-dependent ROC analysis, the prognostic accuracy of model 1 in the internal validation set was 0.73 at 1 year, 0.82 at 3 years and 0.77 at 5 years (Fig. 5B); the prognostic accuracy of model 1 in GSE19750 was 0.61 at 1 year, 0.88 at 3 years and 0.91 at 5 years (Fig. 5A); the prognostic accuracy of model 1 in GSE76021 was 0.84 at 1 year, 0.78 at 3 years and 0.76 at 5 years (Fig. 6B). As for model 2, the high-risk group was associated with poorer survival compared with low-risk group (internal validation set: $P = 0.018$; GSE19750: $P = 0.0035$; GSE76021: $P = 0.012$; Fig. 5E, Fig. 6C-D). Moreover, the prognostic accuracy of model 2 in the internal validation set was 0.78 at 1 year, 0.74 at 3 years and 0.76 at 5 years (Fig. 5E); the prognostic accuracy of model 2 in GSE19750 was 0.60 at 1 year, 0.89 at 3 years and 0.89 at 5 years (Fig. 5C); the prognostic accuracy of model 2 in GSE76021 was 0.90 at 1 year, 0.78 at 3 years and 0.77 at 5 years (Fig. 6D). Both of the two models showed great prognostic values of patients with ACC.

Prognostic value of model 1 and model 2

Because we had identified two best models before, we respectively included the risk score of model 1 (or model 2) and important factors we mentioned before for the cox regression analysis. Risk score of model 1 (Hazard Ratio = 4.822, 95%CI of ratio: 2.636–8.821, $P < 0.001$), risk score of model 2 (Hazard Ratio = 4.465, 95%CI of ratio: 2.449–8.143, $P < 0.001$), tumor stage (Hazard Ratio = 6.097, 95%CI of ratio: 2.433–15.282, $P < 0.001$), and M stage (Hazard Ratio = 5.806, 95%CI of ratio: 2.471–13.640, $P < 0.001$) were influence features of OS as suggested by univariate Cox analysis (Fig. 7A-C). In model 1, the results of multivariate Cox analysis suggested that even being adjusted by other features, risk scores of model 1 (Hazard Ratio = 4.663, 95%CI of ratio: 2.238–9.713, $P < 0.001$) was still relevant to OS among patients with ACC as shown in Fig. 7B. In model 2, the results of multivariate Cox analysis suggested that even being adjusted by other features, risk scores of model 2 (Hazard Ratio = 5.603, 95%CI of ratio: 2.410–13.024, $P < 0.001$) and tumor stage (Hazard Ratio = 3.803, 95%CI of ratio: 1.223–11.822, $P = 0.021$) were still relevant to OS among patients with ACC (Fig. 7D).

Identification of model 1 and model 2 associated biological pathways

With the cut-off criteria we set before, only one risk score-related KEGG signaling pathways were enriched in model 1 including cell cycle (nominal $P < 0.001$, $|ES| = 0.738$, gene size = 123 and FDR = 0.903%) as shown in Table 3. In model 2, two risk score-related KEGG signaling pathways were significantly enriched including cell cycle (nominal $P < 0.001$, $|ES| = 0.722$, gene size = 123 and FDR = 0.884%) and oocyte meiosis (nominal $P < 0.001$, $|ES| = 0.604$, gene size = 111 and FDR = 0.913%; Table 3). Interestingly, these two pathways were consistent with the results of KEGG analysis of genes in blue module.

Table 3
Gene set enrichment analysis of model 1 and model 2 by using GSE76019.

Category	Term	Count	ES	PValue	FDR
Model 1	Cell cycle	123	0.738	< 0.001	0.009025
Model 2	Cell cycle	123	0.722	< 0.001	0.008841
	Oocyte meiosis	111	0.604	< 0.001	0.009127

KEGG: Kyoto Encyclopedia of Genes and Genomes; ES: Enrichment score; FDR: False Discovery Rate

Clinical utility of model 1 and model 2 and nomogram building

Both of the two candidate models showed great prognostic values of patients with ACC according to previous analyses. In order to distinguish the best prognostic model, we further constructed two nomograms separately according to risk scores of model 1 (Fig. 8A) and model 2 (Fig. 9A) and other significant features in multivariate Cox analysis. Both of the two models showed good potential for clinical application. The model 1 related signature showed high potential except Threshold Probability (Pt) was approximately between 0.75 and 0.82 (Fig. 8E). Signature of model 2 also showed good value for clinical application when $0.00 < Pt < 0.72$ or $0.90 < Pt < 1$ (Fig. 9E). But as for the result of calibration curves (Fig. 8B-D; Fig. 9B-

D), the nomogram of model 2 showed better value compared with ideal model and model 1. Especially for nomogram's 3-year and 5-year OS, the nomogram showed great performance almost as the ideal model did (Fig. 9C-D). Thus, we regarded model 2 ($\alpha = 0.6$) as the best prognostic model in the present study. Model 2 contained 14 most powerful prognostic markers (ASPM (abnormal spindle microtubule assembly), AURKA (aurora kinase A), CCNB2 (cyclin B2), CDC20 (cell division cycle 20), CENPA (centromere protein A), EXO1 (exonuclease 1), FBXO5 (F-box protein 5), HJURP (Holliday junction recognition protein), KIF2C (kinesin family member 2C), MKI67 (marker of proliferation Ki-67), NUF2 (NUF2 component of NDC80 kinetochore complex), PARPBP (PARP1 binding protein), TACC3 (transforming acidic coiled-coil containing protein 3), and TROAP (trophinin associated protein)), which were selected for further analysis.

A summary of mutations and CNVs of genes in the best model

Among the 92 ACC patients with sequencing data, only ten independent samples existed mutations of genes in the best model (Table 4). As for genetical alteration of the 14 genes, MKI67 altered most (12%) and the main type was mRNA High (Fig. 10). Combined with the relative mRNA expression values of these genes, these genes seemed to highly express when there existed alteration of them (Fig. 10). Among the 90 ACC samples with CNV data, CNVs of the fourteen genes were constantly existed (Fig. 11A, Table 5). Among the fourteen genes, PARPBP was the most associated gene, which had the highest frequency of CNV events (77.78%, 70/90) (Fig. 11A, Table 5). Moreover, gain of copy number was the most common CNV event (62.32%, 344/552, Fig. 11B).

Table 4
Mutations of genes in model 2 in ACC patients from TCGA database.

ACC Sample ID	ASPM	AURKA	CDC20	EXO1	FBXO5	HJURP	KIF2C	MKI67	TACC3
TCGA-PK-A5HB	I1717*								
TCGA-OU-A5PI		Q2415H							
TCGA-OR-A5K4	R3198C		V361I						
TCGA-OR-A5JA		V27M		C695F				D2208Y	D709Y
TCGA-OR-A5JC				L127fs*6					
TCGA-OR-A5L3					P459L				
TCGA-OR-A5K9						T575I			
TCGA-OR-A5JP							R2786Q		
TCGA-OR-A5LP							T284N		
TCGA-OR-A5JG							P3145H		
Note: ACC: adrenocortical carcinoma.									

Table 5
Different CNV patterns occur in ACC samples from TCGA database.

	Diploid	Deep deletion	Shallow deletion	Copy number gain	Amplification	CNV sum	Percentage
deletion							
ASPM	63	0	13	13	1	27	30.00%
AURKA	37	0	3	49	1	53	58.89%
CCNB2	56	0	22	11	1	34	37.78%
CDC20	55	0	32	2	1	35	38.89%
CENPA	58	0	18	14	0	32	35.56%
EXO1	59	0	15	16	0	31	34.44%
FBXO5	54	0	19	16	1	36	40.00%
HJURP	59	2	16	13	0	31	34.44%
KIF2C	54	0	32	3	1	36	40.00%
MKI67	52	0	10	28	0	38	42.22%
NUF2	63	0	11	14	2	27	30.00%
PARPBP	20	1	6	63	0	70	77.78%
TACC3	41	0	10	37	2	49	54.44%
TROAP	21	0	1	65	3	69	76.67%

Relationship between alterations of genes in the best model and the clinical features

Based on the data from TCGA database, six clinical features (gender, age, pathologic stage, T stage, N stage, and M stage) were collected for this analysis. As the result suggested (Table 6), alterations of genes were significantly associated with pathologic stage (Chi-square = 10.644, $P= 0.014$), T stage (Chi-square = 11.008, $P= 0.012$), and M stage (Chi-square = 10.687, $P= 0.001$).

Table 6
Clinical pathological parameters of ACC patients with or without mutation/CNV of genes in model 2.

		With mutation and/or CNV	Without mutation and CNV	Total number	Chi-square	P-value
Age	≥ 65	6	8	14	0.200	0.655
	< 65	24	50	74		
Gender	Female	19	40	59	0.284	0.594
	Male	11	18	29		
Pathologic stage	I	3	6	9	10.644	0.014
	II	10	33	43		
	III	5	13	18		
	IV	12	6	18		
T stage	T1	3	6	9	11.008	0.012
	T2	11	37	48		
	T3	3	8	11		
	T4	13	7	20		
N stage	N0	24	54	78	2.195	0.138
	N1	6	4	10		
	Nx	0	0	0		
M stage	M0	18	52	70	10.687	0.001
	M1	12	6	18		
	Mx	0	0	0		

Note: With mutation or CNV: Cases have mutant or CNV or mutant + CNV, confirmed through TCGA database. Without mutation and CNV: Cases with neither mutant nor CNV, confirmed through TCGA database.

The effects of alterations in the fourteen genes on the relative mRNA expression levels

According to the result, copy number gains of ASPM ($P < 0.0001$), CCNB2 ($P = 0.0437$), CENPA ($P < 0.0001$), EXO1 ($P = 0.0292$), HJURP ($P = 0.007$), MKI67 ($P = 0.0054$), NUF2 ($P = 0.0004$), PARPBP ($P = 0.0037$), TACC3 ($P < 0.0001$), and TROAP ($P = 0.0024$) were significantly associated with higher mRNA expression comparing with these in copy number shallow deletions (or deep deletions, Fig. 12). As for AURKA ($P = 0.0768$) and FBXO5 ($P = 0.1043$), gains of copy number led to the trend of higher mRNA expression. There was no copy number gain in CDC20, but samples with shallow deletion were significantly lower expressed compared with the diploid ones ($P = 0.0451$). As for KIF2C, only one sample existed copy number gain, which perhaps meant that the result might not be convincing.

Correlation between CNVs of the fourteen genes and survival of patients with ACC

In this step, survival analysis was performed to explore the prognostic value of CNVs in the fourteen genes. According to the result, ACC patients with alterations in the fourteen genes had poor OS ($P = 7.286E-4$, Fig. 13A) and DFS ($P = 0.0131$, Fig. 13B). Unfortunately, there was no association between patients with CNV or without CNV and OS ($P = 0.58$, Fig. 13C). Then we separately explored the association between OS in each gene and CNVs. The result demonstrated that shallow deletions in ASPM ($P = 0.0170$, Fig. 13D), CENPA ($P = 0.0071$, Fig. 13H), EXO1 ($P = 0.0069$, Fig. 13I), HJURP ($P = 0.0200$, Fig. 13K), and NUF2 ($P = 0.0450$, Fig. 13N) led to better OS of patients with ACC compared those affected by copy number gains. While patients with shallow deletions in AURKA ($P = 0.0150$, Fig. 13E), MKI67 ($P = 0.0039$, Fig. 13M), PARPBP ($P < 0.0001$, Fig. 13O), and TACC3 ($P = 0.0016$, Fig. 13P) had poorer OS. As for MKI67, patients affected by CNVs (shallow deletion or gain) had poorer OS comparing with these affected by diploid ($P = 0.0039$, Fig. 13M).

Validation of the fourteen genes in the best model

According to the result, all the classifiers constructed and verified by TCGA-ACC data showed good performance because the accuracy of classification was more than 0.70 (Table 7). Especially the polynomial-SVM-based classifier, its accuracy was 0.9870. And we validated all the classifiers by using GSE19750, GSE76019, and GSE76021. All the classifiers showed good performance in GSE76019 and GSE76021 (accuracy of classification ≥ 0.50 , Table 7). But in GSE19750, KNN-based classifier and all the SVM-based classifiers did not play well as we expected (Table 7).

Perhaps because only 22 ACC samples from GSE19750 were included for validation. All in all, all the classifiers constructed based on the fourteen genes showed great performance to distinguish dead ACC samples from alive ACC samples, which meant all the fourteen genes were significantly associated with prognosis of patients with ACC.

Table 7

The accuracy of classification of LDA-based classifier, KNN-based classifier, linear-SVM-based classifier, polynomial-SVM-based classifier, RBF-SVM-based classifier, and sigmoid-kernel-SVM based classifier.

		LDA	KNN	Kappa	Linear-SVM	Kappa	Polynomial-SVM	Kappa	RBF-SVM	Kappa	Sigmoid-kernel SVM	Kappa
TCGA-ACC	Training set	0.8701	0.8961	0.7463	0.8831	0.7111	0.9870	0.9694	0.9351	0.8469	0.8182	0.5664
GSE19750	Testing set 1	0.6364	0.2727	-0.0353	0.4091	0.0403	0.1818	0	0.2727	0.0435	0.4545	0.0704
GSE76019	Testing set 2	0.7647	0.7059	0.2056	0.7059	0.3561	0.6471	0	0.7353	0.3319	0.7647	0.5211
GSE76021	Testing set 3	0.6207	0.7241	0.7241	0.7931	0.5272	0.6897	0.2162	0.5862	-0.0235	0.7931	0.5272

Note: LDA: Linear discriminant analysis; KNN: K-Nearest Neighbor; RBF: radial basis function; SVM: Support Vector Machine.

Hub genes in the best model

All the fourteen genes from the best model were included in this PPI network (Figure S5A). As shown in Figure S5B, the degrees of six genes (ASPM, AURKA, CCNB2, CDC20, KIF2C, and NUF2) were higher than any other of the remaining nine genes (Degree = 13). Thus, the five genes were considered as hub genes in the best model. Expression of ASPM was significantly related to CCNB2, CDC20, FBXO5, HJURP, MKI67, NUF2, PARPBP, TACC3, and TROAP (Figure S5C). Among them, MKI67 was the most related gene (Figure S5D). Expression of AURKA was significantly associated with FBXO5, HJURP, MKI67, PARPBP, TACC3, and TROAP (Figure S5C). Moreover, PARPBP was associated with AURKA best (Figure S5D). Expression of CCNB2 was relevant to ASPM, CDC20, CENPA, FBXO5, HJURP, PARPBP (Figure S5C) and PARPBP was the most associated gene with CCNB2 (Figure S5D). Expression of CDC20 was associated with CCNB2, ASPM, FBXO5, HJURP, KIF2C, MKI67, PARPBP, TROAP (Figure S5C) and the most related gene was HJURP (Figure S5D). The expression of NUF2 was significantly related to HJURP, FBXO5, CENPA, CDC20, AURKA, and ASPM (Figure S5C). Among them, ASPM was the most associated gene (Figure S5D).

Discussion

ACC is a relatively rare malignancy in urinary system associated with a poor prognosis[27]. Five-year survival rate for patients with ACC was 35% as reported[2]. Similarly, it is not optimistic that 40% of the cases have distant metastasis when diagnosis[2], which means that they have missed the best opportunity of treatment. Thus, there is a great need to develop novel molecular biomarkers for diagnosis and prognosis of ACC patients.

Over the past several years, the discoveries of novel and informative genes by using bioinformatics methods have provided valuable information in diagnosis of malignancies[4]. In present, WGCNA has been widely used to screen out novel and effective molecular biomarkers in bioinformatics field[28, 29]. In our previous study, we have made great efforts to use WGCNA as the main method to identify prognosis biomarkers for bladder cancer[30]. Thus, we firstly identified genes associated with survival of ACC in the present study. And 93 biomarkers significantly relevant to survival of ACC were further identified combining with overall survival analysis and disease-free survival analysis. Interestingly, some studies thought that a single biomarker might not predict the prognosis of patients with ACC well[31]. Nowadays, more and more studies have paid their attention on developing a multi-gene signature for predicting prognosis in malignancies instead of just a single biomarker[32, 33]. Given that the Ridge, ELASTIC-NET, and LASSO methods have been widely used to construct a Cox proportional hazard regression model[34], we used all the methods to establish 11 multiple-gene signatures for predicting prognosis of ACC patients based on the 93 survival associated genes. Two candidate models were further screened out by calculating the AUC by using internal validation set. The two signatures were further verified by using three independent datasets (internal validation set, GSE19750, GSE76021). In both model 1 and model 2, we firstly calculated the risk score of each patient according to the relative expression levels of the mRNA in the classifier and corresponding LASSO coefficients. Then we divided patients into high- and low-risk groups in each dataset based on the risk score for further validation and analysis. In the training set, the two signatures could distinguish between the high- and low-risk patients according to the results of survival analysis and ROC analysis. In the internal validation set, GSE19750, and GSE76021, the results were entirely consistent with what we found in the training set.

To test whether the prognostic value of the two signatures were independent of other clinical features (gender, age, tumor stage, T stage, N stage, M stage, and weiss score), we included these factors for the Cox regression analysis. The result indicated that tumor stage, risk score of model 1, and risk score of model 2 were significant prognostic factors for ACC patients. Moreover, risk score of model 1 and risk score of model 2 were independent of tumor stage. Tumor stage including TNM staging system reflects the internal characteristics of tumors, which is of great value for evaluation of the degree of tumor deterioration and prediction of prognosis according to previous findings[35]. In our study, model 1 and model 2

could not only discriminate between low- and high- risk groups like tumor stage did, but also showed more effective prognostic value compared with tumor stage, which made our results reliable.

To better understand the biological role of the two signatures, we performed GSEA analysis. Two KEGG pathways associated with model 1 were identified including cell cycle and oocyte meiosis meanwhile one KEGG pathway associated with model 2 was identified including cell cycle. Cell cycle was the basic process of cell proliferation. According to a study by Kaistha., et al[36], the deregulation of cell cycle was obviously associated with tumor progression. As for oocyte meiosis, maturation and development of oocyte is a complex biologic process in mammalian, which is the key event in the reproductive process[37]. Any errors in this process can lead to the oocyte developmental abnormalities or infertility[37]. Meiosis is specialized cell division with the reduction in genetic content, which guaranteed the stability of chromosome numbers of gamogenetic organisms[37].

To screen out the best model among them, we further construct two nomograms based on risk score of model 1 and risk score of model 2 for prediction of the possibility of ACC patients' OS, the calibration curve suggested that the nomogram of model 2 showed better performance especially for nomogram's 3-year and 5-year OS compared with the nomogram of model 1. Thus, model 2, contained 14 genes, was regarded as the best prognostic model.

We also evaluated the effect of mutations and CNVs of genes in the best model. In the present study, ten independent samples existed mutations of these genes. MKI67 altered most and the main type was mRNA High. Further analysis demonstrated that high expression levels of genes in the best model was associated with gene alteration. Moreover, PARPBP had the highest frequency of CNV events and gain of copy number was the most common CNV event. In detail, alterations of genes were related to pathologic stage, T stage, and M stage. Copy number gains of ASPM, CCNB2, CENPA, EXO1, HJURP, MKI67, NUF2, PARPBP, TACC3, and TROAP led to high mRNA expression comparing with these in loss of copy number (shallow deletions).

Alterations of these genes also had strong effect on the survival of ACC patients. Alterations in the fourteen genes were significantly associated with poor OS and DFS. Further analysis indicated that shallow deletions in ASPM, CENPA, EXO1, HJURP, and NUF2 caused better OS of ACC patients, meanwhile shallow deletions in AURKA, MKI67, PARPBP, and TACC3 led to poorer OS. Three methods (LDA, KNN, and SVM) made these results convincing.

A PPI network was constructed based on the fourteen genes from the best model, six hub genes were further identified according to the degree of connectivity. The results of meta-analysis indicated that all the six hub genes were associated with survival status and time of patients with ACC.

There still existed some limitations in our study. Firstly, though our study was a relatively preliminary exploration of establishing a multi-gene signature by combing bioinformatics methods and statistics and machine learning, there might be lack of novelty in methods. Therefore, we will explore more novel bioinformatics and machine learning methods to integrate prognostic gene signatures. Secondly, the best prognostic model was developed and validated by using data derived from TCGA and GEO databases. Although the signature could distinguish high- and low- risk groups well, it was not clear if it could show good performance as we expected in clinical trials. Thus, we will further apply and test this signature in clinical judgement for prognosis of ACCs. Thirdly, according to the results of DCA, the best signature showed high potential when $0.00 < Pt < 0.72$ or $0.90 < Pt < 1$. But when Pt equaled other values, this signature did not play well as we expected. Perhaps due to the small size of these datasets. Therefore, we will further validate the multi-gene signature in larger and richer datasets.

To sum up, we performed WGCNA analysis to construct a co-expression network for ACC by using GSE76019. 93 hub genes in the co-expression network were identified as survival associated genes. Moreover, we constructed an ELASTIC-NET model including fourteen genes, which were validated to show good performance by using three independent datasets. This fourteen-gene signature probably became a promising tool for prediction of survival and prognosis for ACCs, which was independent of some important factors such as tumor stage. Moreover, six hub genes were identified among the fourteen genes from the best model, which were the most powerful biomarkers of prognosis of patients with ACC. However, the best model and six hub genes needs to be verified by using more novel bioinformatics methods, clinical trials and larger datasets.

Abbreviations

ACC: Adrenocortical carcinoma; AUC: area under curve; ANOVA: One-way analysis of variance; C-index: consistency index; CNV: copy number variation; DFS: diseases-free survival; GEO: Gene Expression Omnibus; GEPIA: Gene expression profiling interactive analysis; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: gene set enrichment analysis; KNN: K-Nearest Neighbor; LASSO: least absolute shrinkage and selection operator; LDA: Linear discriminant analysis; OS: Overall survival; PPI: protein-protein interaction; ROC: receiver operating characteristic curve; RS: risk score; STRING: the Search Tool for the Retrieval of Interacting Genes; SVM: support vector machine; TCGA: The Cancer Genome Atlas; WGCNA: weighted correlation network analysis.

Declarations

Acknowledgments

Not applicable.

Author contributions

S. L., and T. L. conceived and designed the study, X. Y., Z. G., D. Y., and X. L. performed the analysis procedures, X. Y., Z. G., C. C., and X. L. analyzed the results, S. L., J. Y., and T. L. contributed analysis tools, X. Y. contributed to the writing of the manuscript. All authors reviewed the manuscript.

Funding

This work was supported by the Zhongnan Hospital of Wuhan University Science, Technology and Innovation Seed Fund (znpj2017050) and the 351 Talent Project of Wuhan University (Luojia Young Scholars: SL); and National Natural Science Foundation of China (81802541).

Availability of data and material

The data (GSE76019, GSE76021) that support the findings of this study are openly available in Gene Expression Omnibus (GEO) database at <http://doi.org/10.1158/1078-0432.CCR-15-2738>, reference number [000391472400031PM 27307598EREF].

The data (GSE19750) that support the findings of this study are openly available in Gene Expression Omnibus (GEO) database at <http://doi.org/10.1016/j.surg.2013.06.058>, reference number [000327571200067PM 24238056EREF].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest.

References

1. Golden SH, Robinson KA, Ian S, Blair A, Ladenson PW: **Clinical review: Prevalence and incidence of endocrine and metabolic disorders in the United States: a comprehensive review.** *J Clin Endocrinol Metab* 2009, **94**(6):1853-1878.
2. Guillaume A, Eric L, Martin F, Anne J, Windy L, Hanin O, Stéphanie R, Karine P, Fernande RC: **Integrated genomic characterization of adrenocortical carcinoma.** *NAT GENET* 2014, **46**(6):607-612.
3. B A, M F: **Clinical review: Adrenocortical carcinoma: clinical update.** *Journal of Clinical Endocrinology and Metabolism* 2006, **91**(6):2027.
4. He Z, Sun M, Ke Y, Lin R, Xiao Y, Zhou S, Zhao H, Wang Y, Zhou F, Zhou Y: **Identifying biomarkers of papillary renal cell carcinoma associated with pathological stage by weighted gene co-expression network analysis.** *Oncotarget* 2017, **8**(17):27904-27914.
5. 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**(Web Server issue).
6. Pinto EM, Rodriguez-Galindo C, Choi JK, Pounds S, Liu Z, Neale G, Finkelstein D, Hicks JM, Pappo AS, Figueiredo BC *et al*: **Prognostic Significance of Major Histocompatibility Complex Class II Expression in Pediatric Adrenocortical Tumors: A St. Jude and Children's Oncology Group Study.** *CLIN CANCER RES* 2016, **22**(24):6247-6255.
7. Demeure MJ, Coan KE, Grant CS, Komorowski RA, Stephan E, Sinari S, Mount D, Bussey KJ: **PTTG1 overexpression in adrenocortical cancer is associated with poor survival and represents a potential therapeutic target.** *SURGERY* 2013, **154**(6):1405-1416, 1416.
8. Pinto EM, Rodriguez-Galindo C, Choi JK, Pounds S, Liu Z, Neale G, Finkelstein D, Hicks JM, Pappo AS, Figueiredo BC *et al*: **Prognostic Significance of Major Histocompatibility Complex Class II Expression in Pediatric Adrenocortical Tumors: A St. Jude and Children's Oncology Group Study.** *CLIN CANCER RES* 2016, **22**(24):6247-6255.
9. Anders S: **Differential gene expression analysis based on the negative binomial distribution.** *Journal of Marine Technology & Environment* 2009, **2**(2).
10. Gautier L, Cope LB, Bolstad BM, Irizarry RA: **affy - analysis of Affymetrix GeneChip data at the probe level.** *BIOINFORMATICS* 2004, **20**(3):307-315.
11. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *NAT GENET* 2000, **25**(1):25-29.
12. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopaedia of Genes and Genomes.** *NUCLEIC ACIDS RES* 2000, **volume 28**(1):27-30.
13. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *OMICS* 2012, **16**(5):284-287.
14. Friedman J: **fit a GLM with lasso or elasticnet regularization.**

15. Sachs MC: **plotROC**: Generate Useful ROC Curve Charts for Print and Interactive Use. 2015.
16. Therneau TM: **survival**: Survival Analysis. *TECHNOMETRICS* 2015, **46**(1):111-112.
17. Blanche P: **timeROC**: Time-Dependent ROC Curve and AUC for Censored Survival Data. 2015.
18. Aut MG, Aut TL: **forestplot**: Advanced Forest Plot Using 'grid' Graphics. 2016.
19. A S, P T, VK M, S M, BL E, MA G, A P, SL P, TR G, ES L *et al*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. 2005.
20. Vickers A, Elkin E: **Decision curve analysis: a novel method for evaluating prediction models**. *Medical Decision Making An International Journal of the Society for Medical Decision Making* 2006, **26**(6):565.
21. Yizhou W, Jun L, Yong X, Renyan G, Kui W, Zhenlin Y, Xuying W, Guanghua L, Dong W, Lehua S: **Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy**. *Journal of Clinical Oncology Official Journal of the American Society of Clinical Oncology* 2013, **31**(9):1188-1195.
22. Michael J P, Ralph B D: **Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation**. *STAT MED* 2010, **23**(13):2109-2123.
23. Venables WN, Ripley BD: **Modern Applied Statistics with S**. *Statistics & Computing* 2002, **52**(1):704-705.
24. Kuhn M: **Caret: Classification and regression training**. *Astrophysics Source Code Library* 2015, **129**(1):291-295.
25. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*: **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets**. *NUCLEIC ACIDS RES* 2019, **47**(D1):D607-D613.
26. Cre BGP, Aut, Aut PC, Ctb KB, Ctb RB, Ctb JU, Ctb EZ: **PerformanceAnalytics: Econometric tools for performance and risk analysis**. 2014.
27. McCabe MJ, Pinese M, Chan CL, Sheriff N, Thompson TJ, Grady J, Wong M, Gauthier MA, Puttick C, Gayevskiy V *et al*: **Genomic stratification and liquid biopsy in a rare adrenocortical carcinoma (ACC) case, with dual lung metastases**. *Cold Spring Harb Mol Case Stud* 2019, **5**(2).
28. Gu HY, Yang M, Guo J, Zhang C, Lin LL, Liu Y, Wei RX: **Identification of the Biomarkers and Pathological Process of Osteoarthritis: Weighted Gene Co-expression Network Analysis**. *FRONT PHYSIOL* 2019, **10**:275.
29. Li Q, Chen W, Song M, Chen W, Yang Z, Yang A: **Weighted gene co-expression network analysis and prognostic analysis identifies hub genes and the molecular mechanism related to head and neck squamous cell carcinoma**. *CANCER BIOL THER* 2019:1-10.
30. Yan X, Guo ZX, Liu XP, Feng YJ, Zhao YJ, Liu TZ, Li S: **Four novel biomarkers for bladder cancer identified by weighted gene coexpression network analysis**. *J CELL PHYSIOL* 2019.
31. Chen L, Luo Y, Wang G, Qian K, Qian G, Wu CL, Dan HC, Wang X, Xiao Y: **Prognostic value of a gene signature in clear cell renal cell carcinoma**. *J CELL PHYSIOL* 2019, **234**(7):10324-10335.
32. Qiao GJ, Chen L, Wu JC, Li ZR: **Identification of an eight-gene signature for survival prediction for patients with hepatocellular carcinoma based on integrated bioinformatics analysis**. *PEERJ* 2019, **7**:e6548.
33. Wu J, Jin S, Gu W, Wan F, Zhang H, Shi G, Qu Y, Ye D: **Construction and Validation of a 9-Gene Signature for Predicting Prognosis in Stage III Clear Cell Renal Cell Carcinoma**. *FRONT ONCOL* 2019, **9**:152.
34. Lu HX, Du J, Wen DL, Sun JH, Chen MJ, Zhang AQ, Jiang JX: **Development and validation of a novel predictive score for sepsis risk among trauma patients**. *WORLD J EMERG SURG* 2019, **14**:11.
35. Veeratterapillay R, Simren R, El-Sherif A, Johnson MI, Soomro N, Heer R: **Accuracy of the revised 2010 TNM classification in predicting the prognosis of patients treated for renal cell cancer in the north east of England**. *J CLIN PATHOL* 2012, **65**(4):367-371.
36. Kaistha BP, Lorenz H, Schmidt H, Sipos B, Pawlak M, Gierke B, Kreider R, Lankat-Buttgereit B, Sauer M, Fiedler L: **PLAC8 Localizes to the Inner Plasma Membrane of Pancreatic Cancer Cells and Regulates Cell Growth and Disease Progression through Critical Cell-Cycle Regulatory Pathways**. *CANCER RES* 2015, **76**(1):96-107.
37. Jaffe LA, Egbert JR: **Regulation of Mammalian Oocyte Meiosis by Intercellular Communication Within the Ovarian Follicle**. *ANNU REV PHYSIOL* 2017, **79**:237-260.

[GSE76019, GSE76021] Pinto EM, Rodriguez-Galindo C, Choi JK, Pounds S *et al*; 2016; Prognostic Significance of Major Histocompatibility Complex Class II Expression in Pediatric Adrenocortical Tumors: A St. Jude and Children's Oncology Group Study; Gene Expression Omnibus (GEO) database; DOI: 10.1158/1078-0432.CCR-15-2738.

[GSE19750] Demeure MJ, Coan KE, Grant CS, Komorowski RA *et al*; 2013; PTTG1 overexpression in adrenocortical cancer is associated with poor survival and represents a potential therapeutic target; Gene Expression Omnibus (GEO) database; DOI: 10.1016/j.surg.2013.06.058.

Supplemental Material

Table S1: GO biological processes of genes in blue module.

Table S2: Survival analysis of hub genes in co-expression network.

Table S3: The corresponding regression coefficients of each gene in models constructed by the Ridge, ELASTIC-NET, and LASSO methods.

Figures

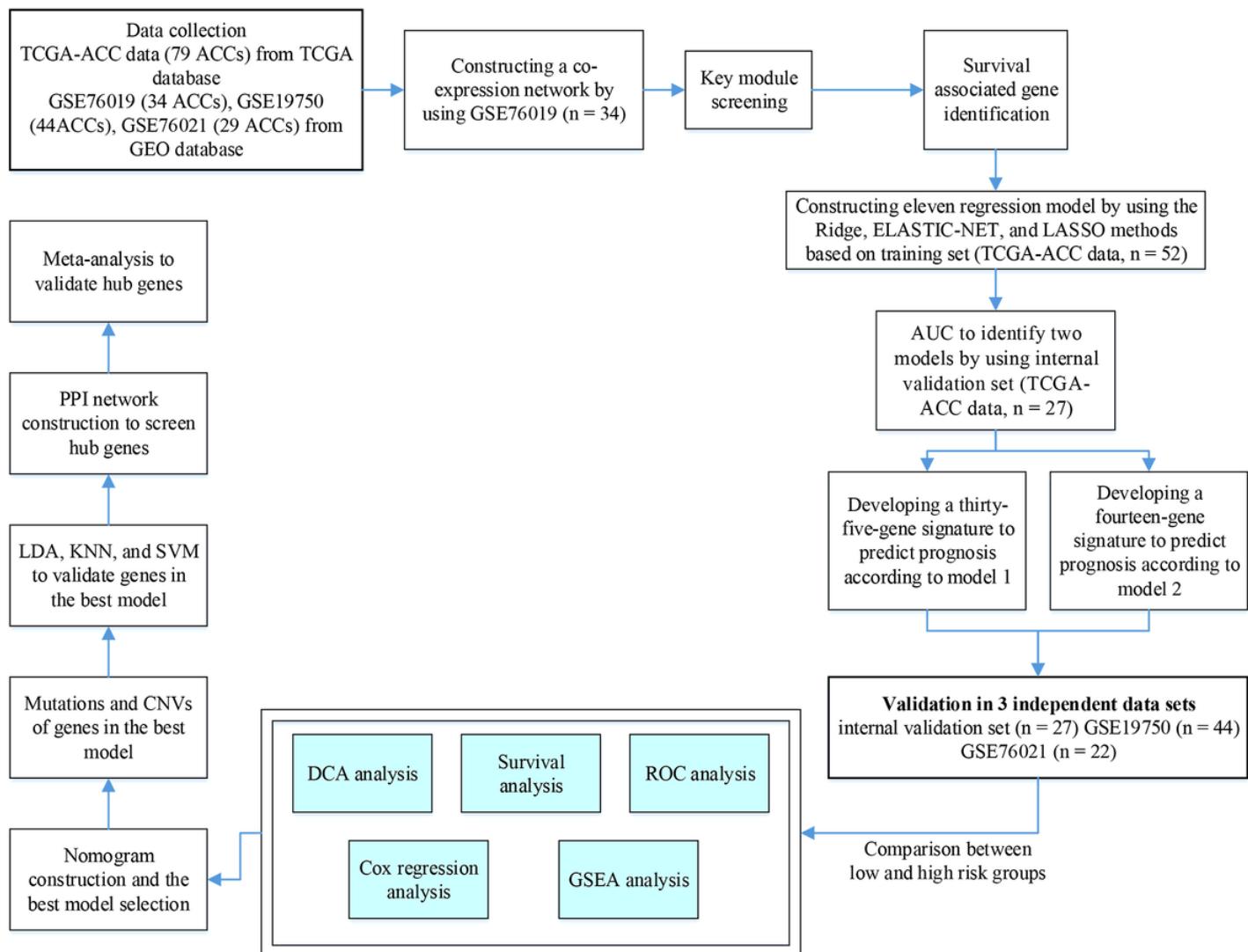


Figure 1

Flow chart indicating the process used to select target genes included in the analysis.

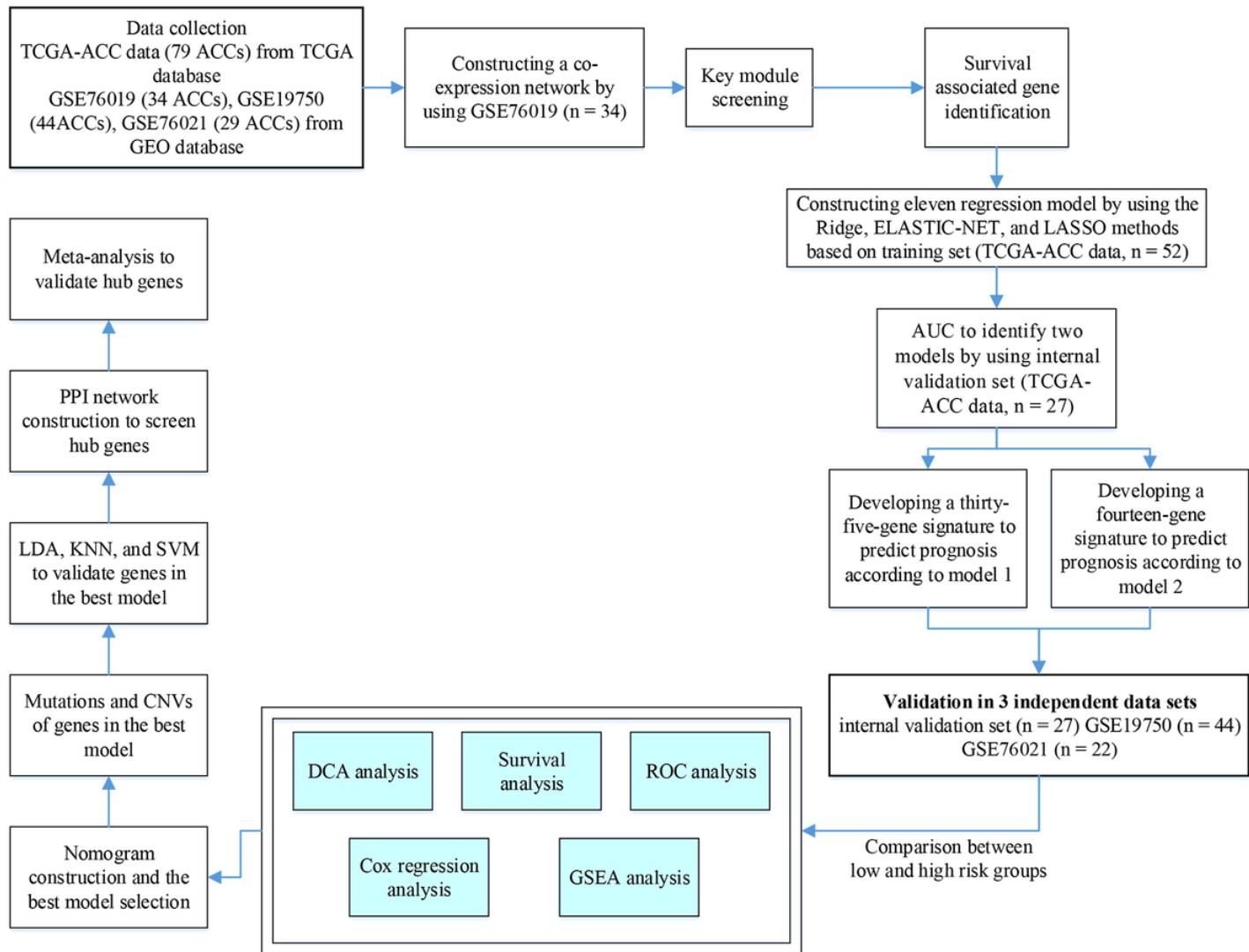


Figure 1

Flow chart indicating the process used to select target genes included in the analysis.

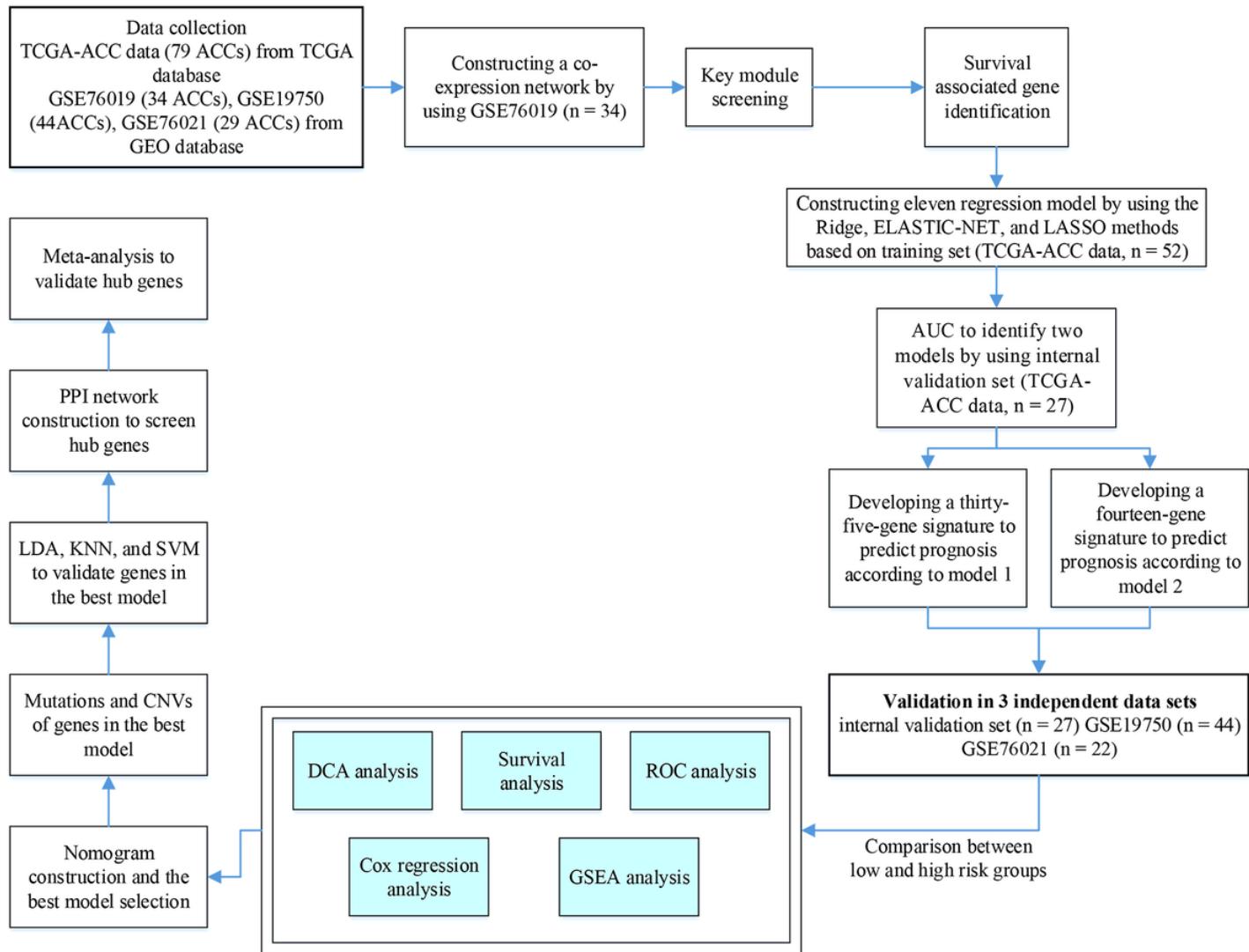


Figure 1

Flow chart indicating the process used to select target genes included in the analysis.

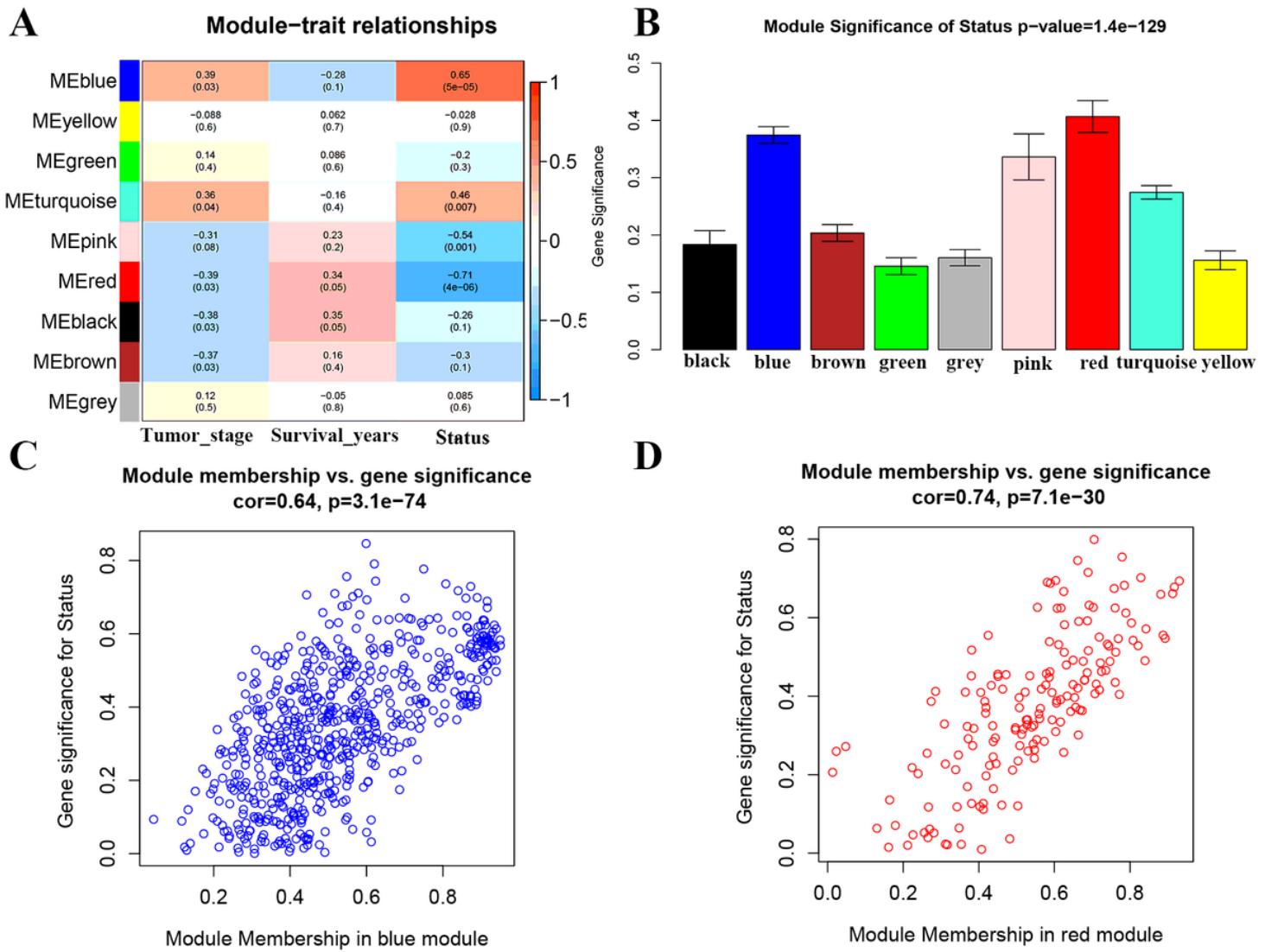
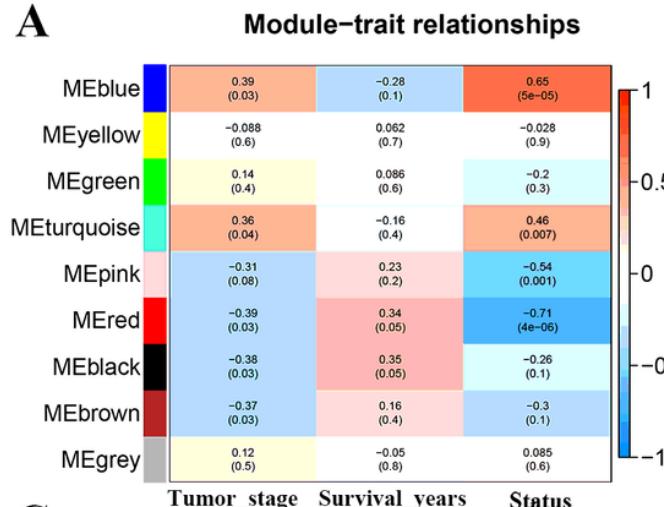
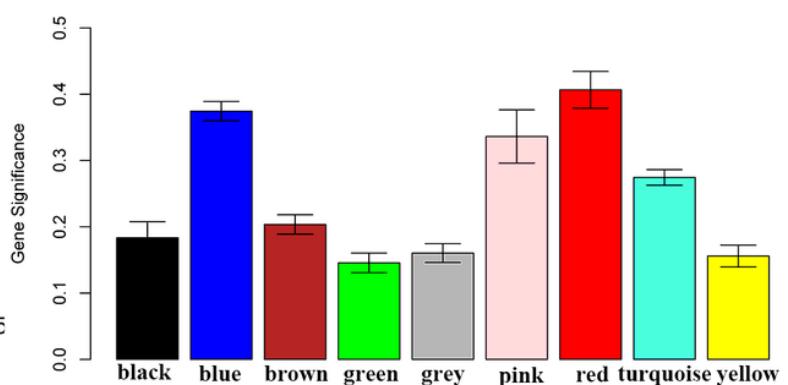
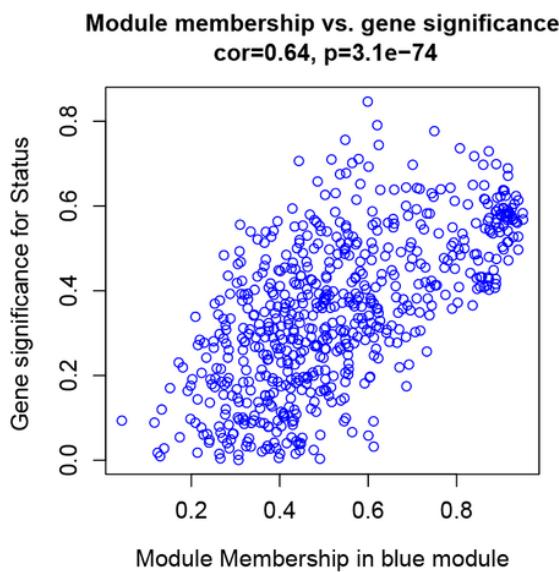
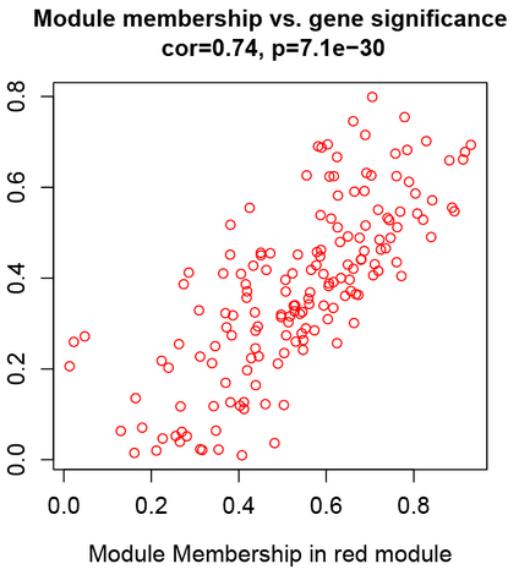


Figure 2

Identification of modules associated with the clinical traits of ACC. (A) Heatmap of the correlation between module eigengenes and clinical traits of ACC. (B) Distribution of average gene significance and errors in the modules associated with survival status of ACC. (C) Scatter plot of module eigengenes related to survival status in the blue module. (D) Scatter plot of module eigengenes related to survival status in the red module.

A**B**

Module Significance of Status p-value=1.4e-129

**C****D****Figure 2**

Identification of modules associated with the clinical traits of ACC. (A) Heatmap of the correlation between module eigengenes and clinical traits of ACC. (B) Distribution of average gene significance and errors in the modules associated with survival status of ACC. (C) Scatter plot of module eigengenes related to survival status in the blue module. (D) Scatter plot of module eigengenes related to survival status in the red module.

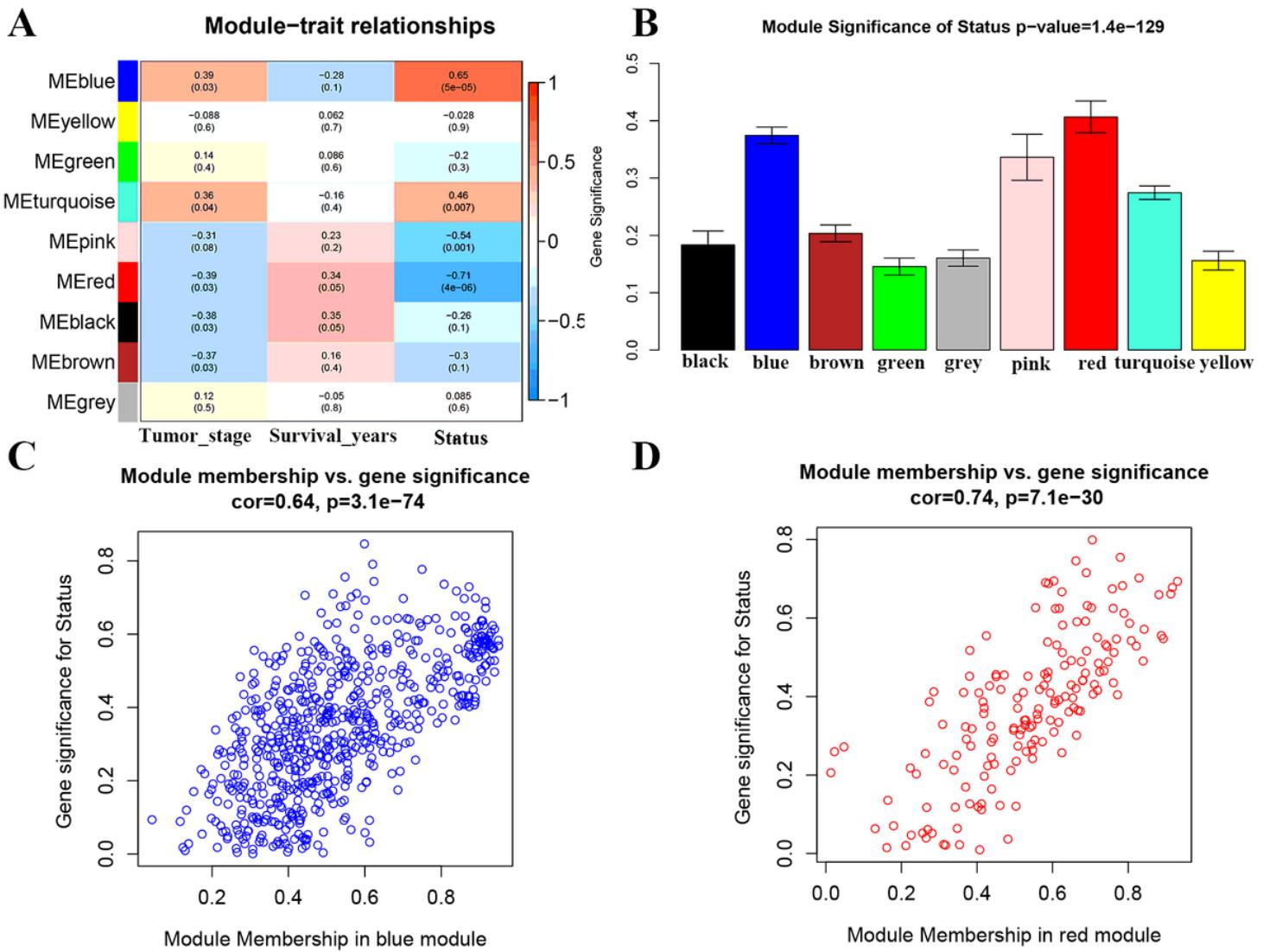


Figure 2

Identification of modules associated with the clinical traits of ACC. (A) Heatmap of the correlation between module eigengenes and clinical traits of ACC. (B) Distribution of average gene significance and errors in the modules associated with survival status of ACC. (C) Scatter plot of module eigengenes related to survival status in the blue module. (D) Scatter plot of module eigengenes related to survival status in the red module.

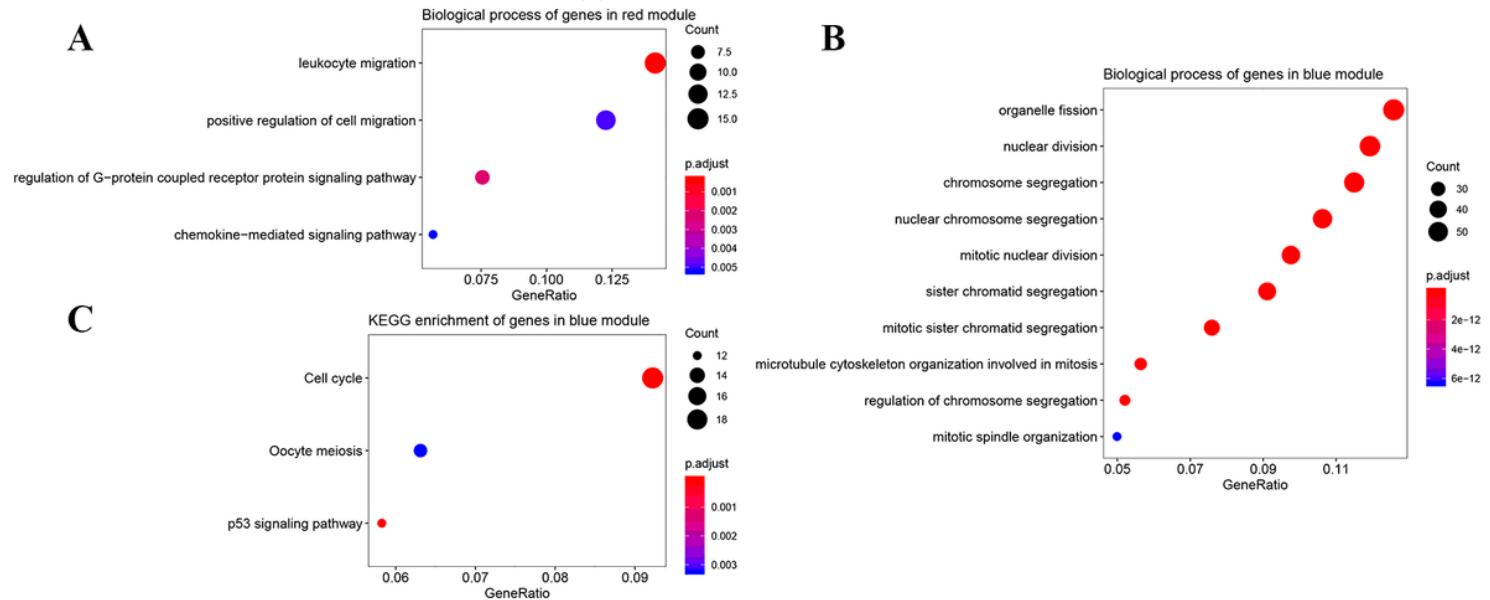


Figure 3

Bioinformatics analysis of genes in the hub modules and ten-time cross-validation for tuning parameter selection. (A) Biological process of genes in the red module. (B) Biological process of genes in the blue module. (C) KEGG pathway enrichment of genes in the blue module.

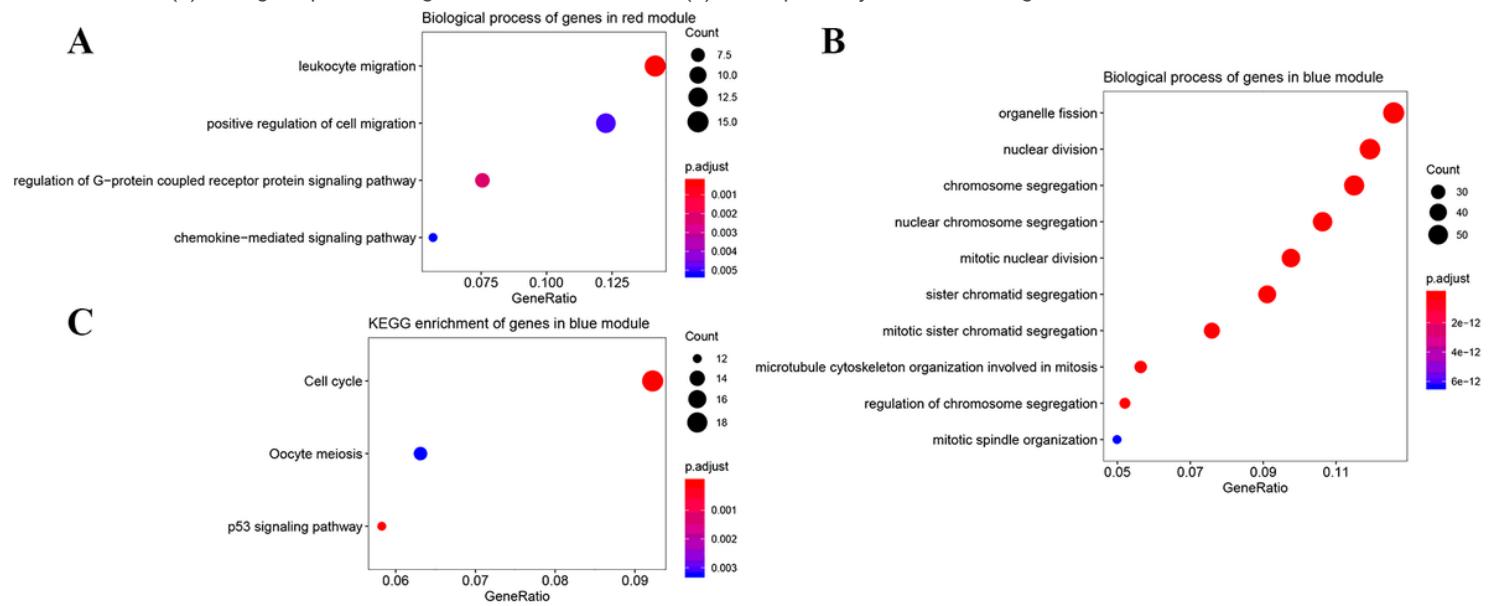


Figure 3

Bioinformatics analysis of genes in the hub modules and ten-time cross-validation for tuning parameter selection. (A) Biological process of genes in the red module. (B) Biological process of genes in the blue module. (C) KEGG pathway enrichment of genes in the blue module.

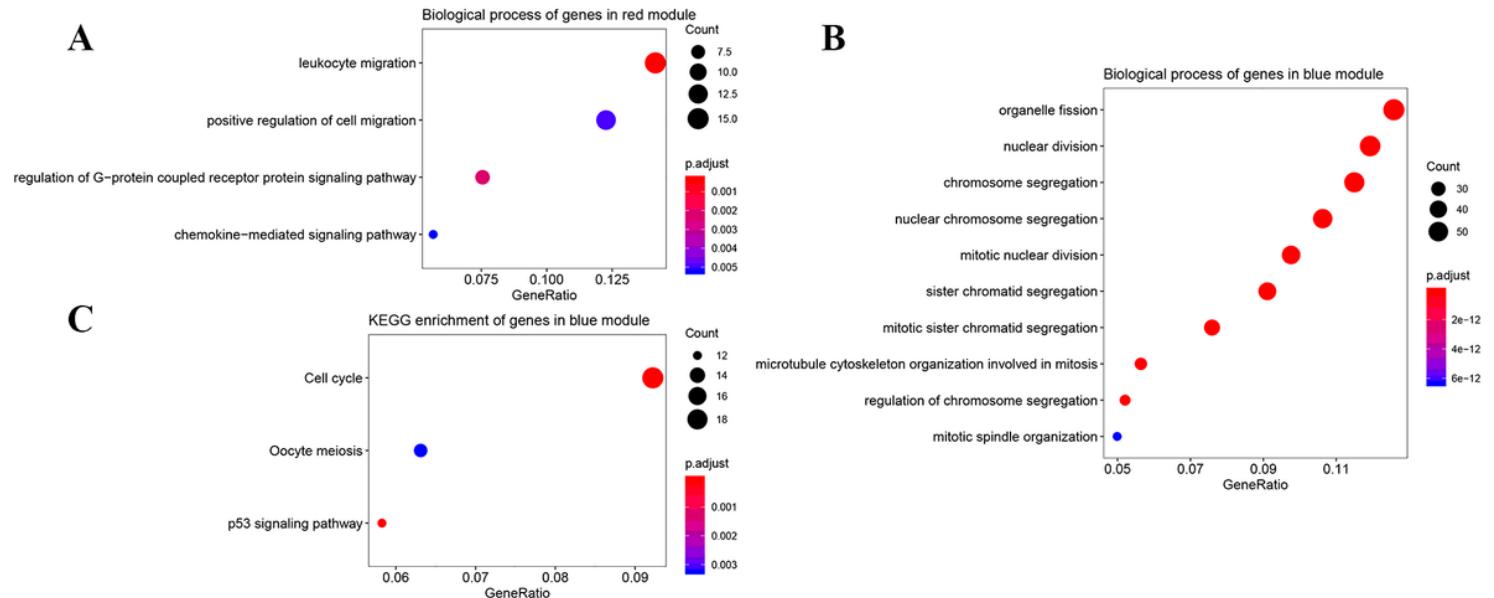


Figure 3

Bioinformatics analysis of genes in the hub modules and ten-time cross-validation for tuning parameter selection. (A) Biological process of genes in the red module. (B) Biological process of genes in the blue module. (C) KEGG pathway enrichment of genes in the blue module.

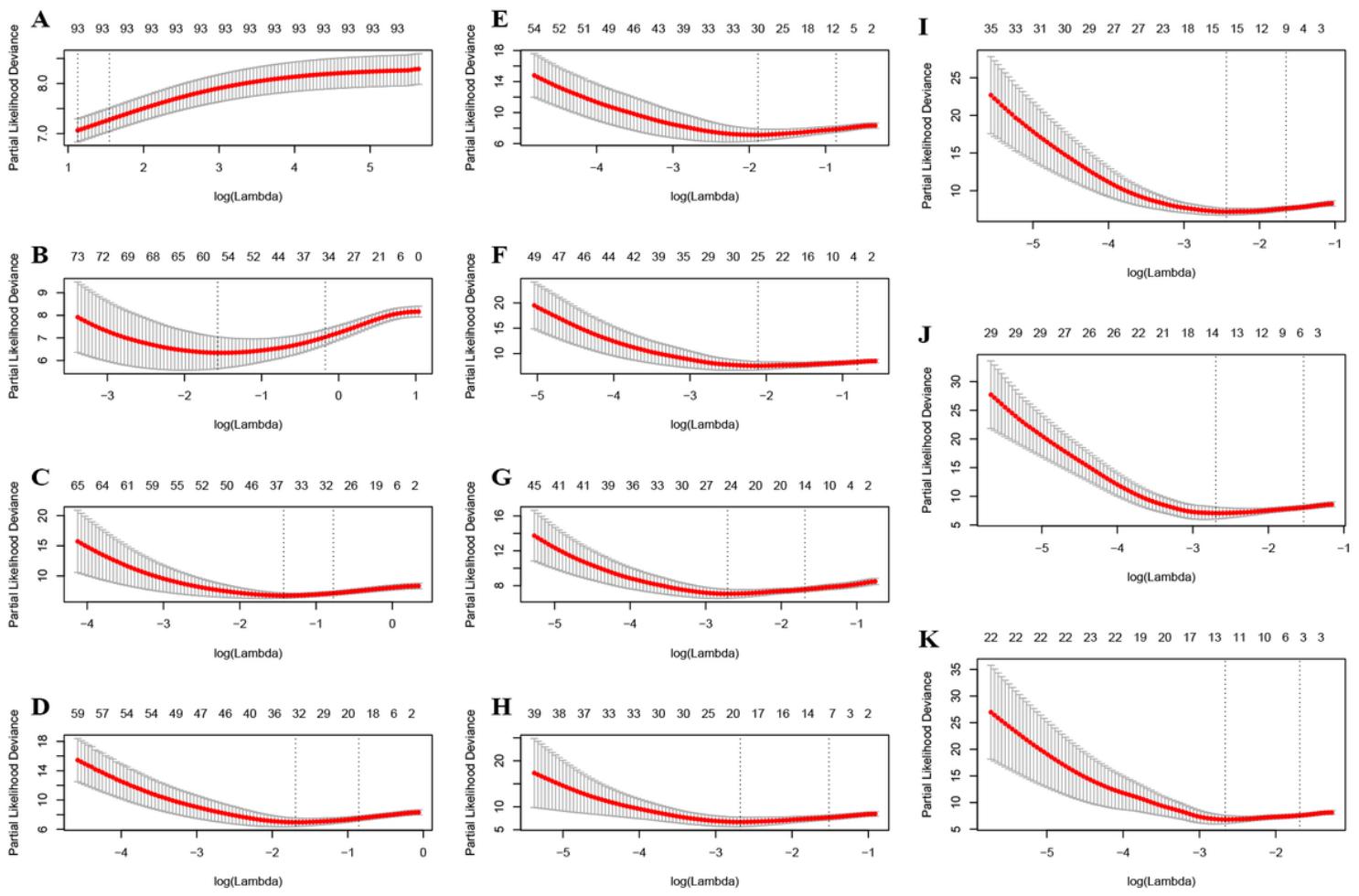


Figure 4

Ten-time cross-validation for tuning parameter selection in the Ridge (A), ELASTIC-NET (with α varying from 0.1 to 0.9 (E - J), and lasso model (K). The solid vertical lines are the partial likelihood deviance SE. The dotted vertical lines are drawn at the optimal values by minimum criteria and 1-SE criteria.

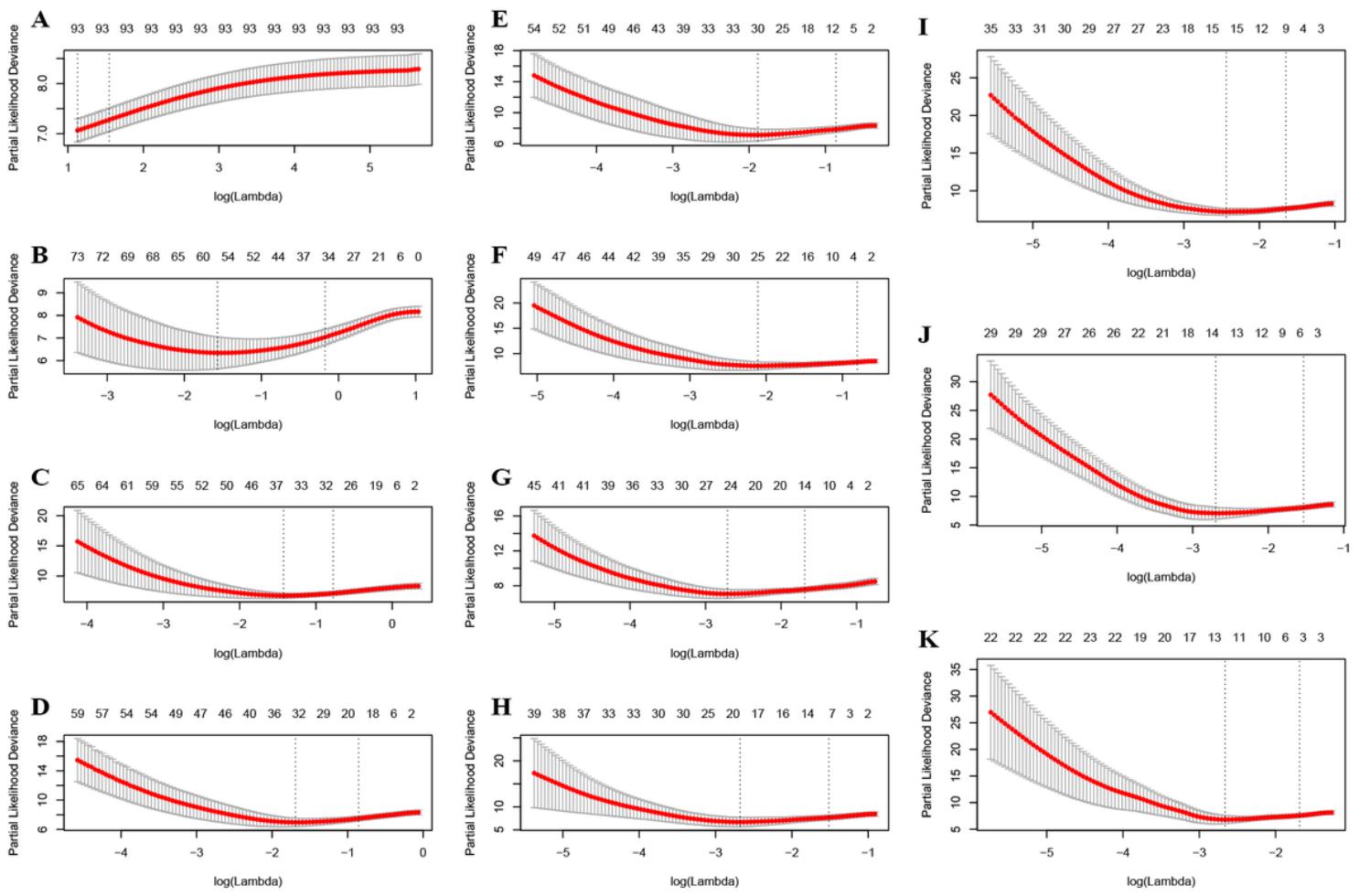


Figure 4

Ten-time cross-validation for tuning parameter selection in the Ridge (A), ELASTIC-NET (with α varying from 0.1 to 0.9 (E - J), and lasso model (K). The solid vertical lines are the partial likelihood deviance SE. The dotted vertical lines are drawn at the optimal values by minimum criteria and 1-SE criteria.

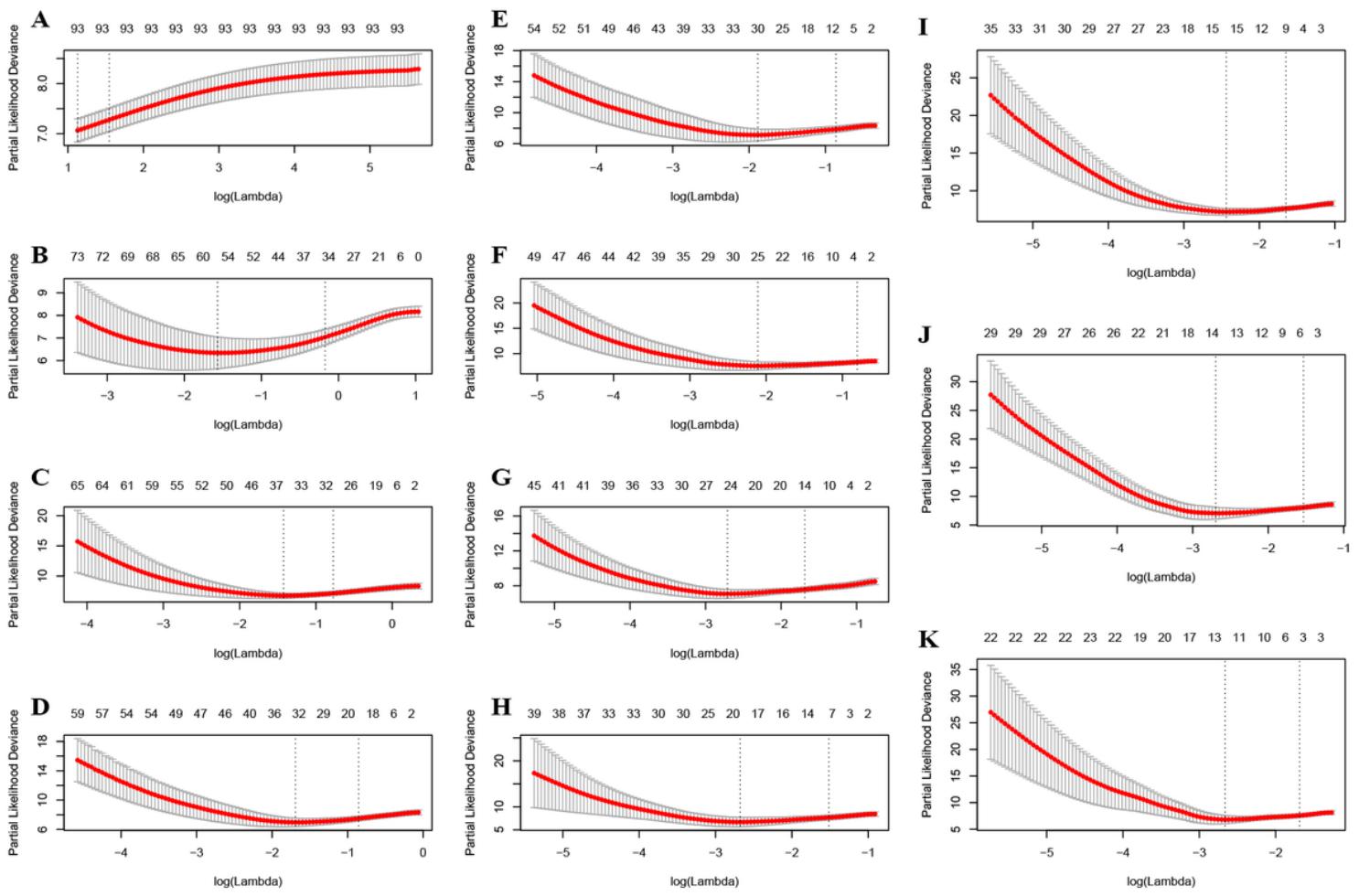


Figure 4

Ten-time cross-validation for tuning parameter selection in the Ridge (A), ELASTIC-NET (with α varying from 0.1 to 0.9 (E - J), and lasso model (K). The solid vertical lines are the partial likelihood deviance SE. The dotted vertical lines are drawn at the optimal values by minimum criteria and 1-SE criteria.

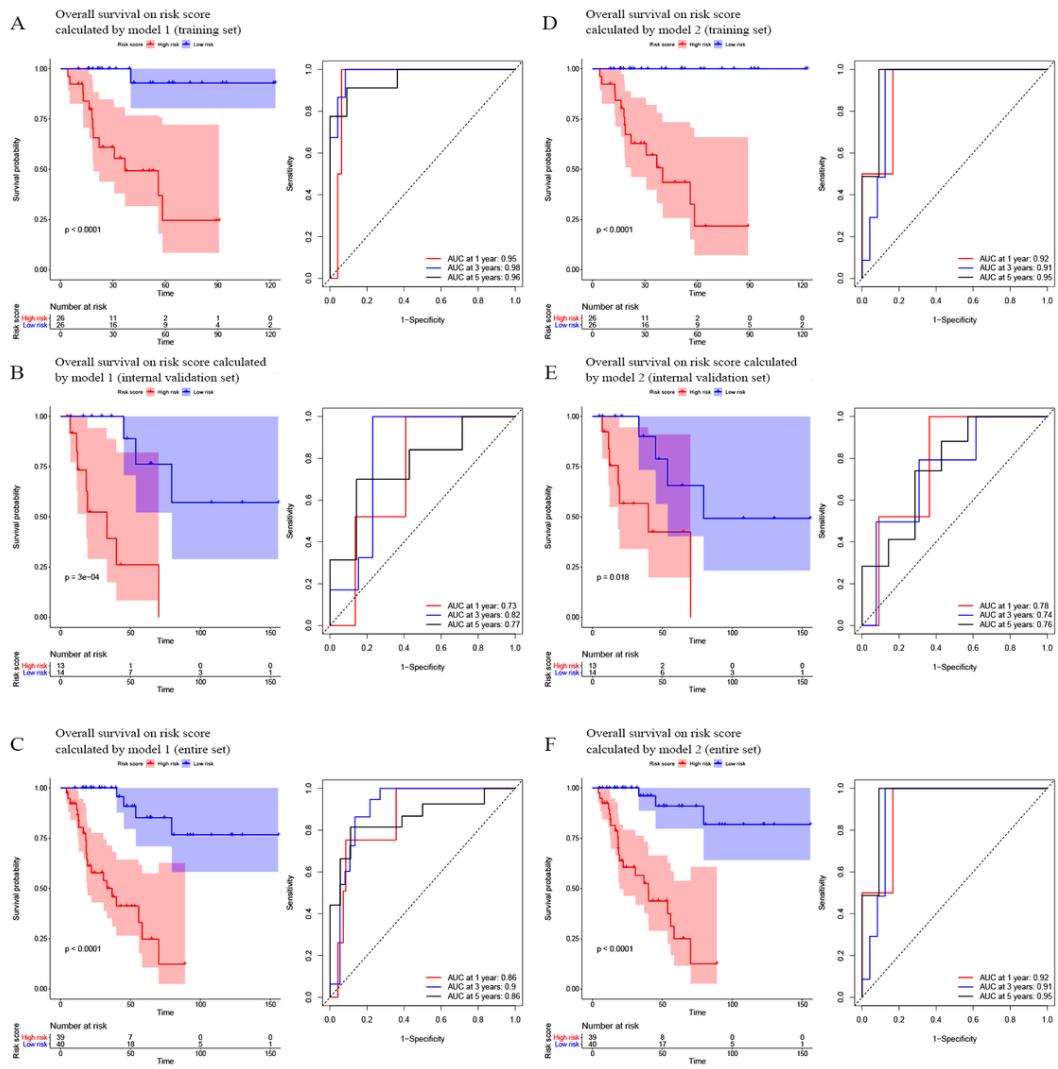


Figure 5

Survival analysis of the association between risk score calculated by model 1 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in (A) training set, (B) internal validation set, (C) entire set. Survival analysis of the association between risk score calculated by model 2 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in (D) training set, (E) internal validation set, (F) entire set.

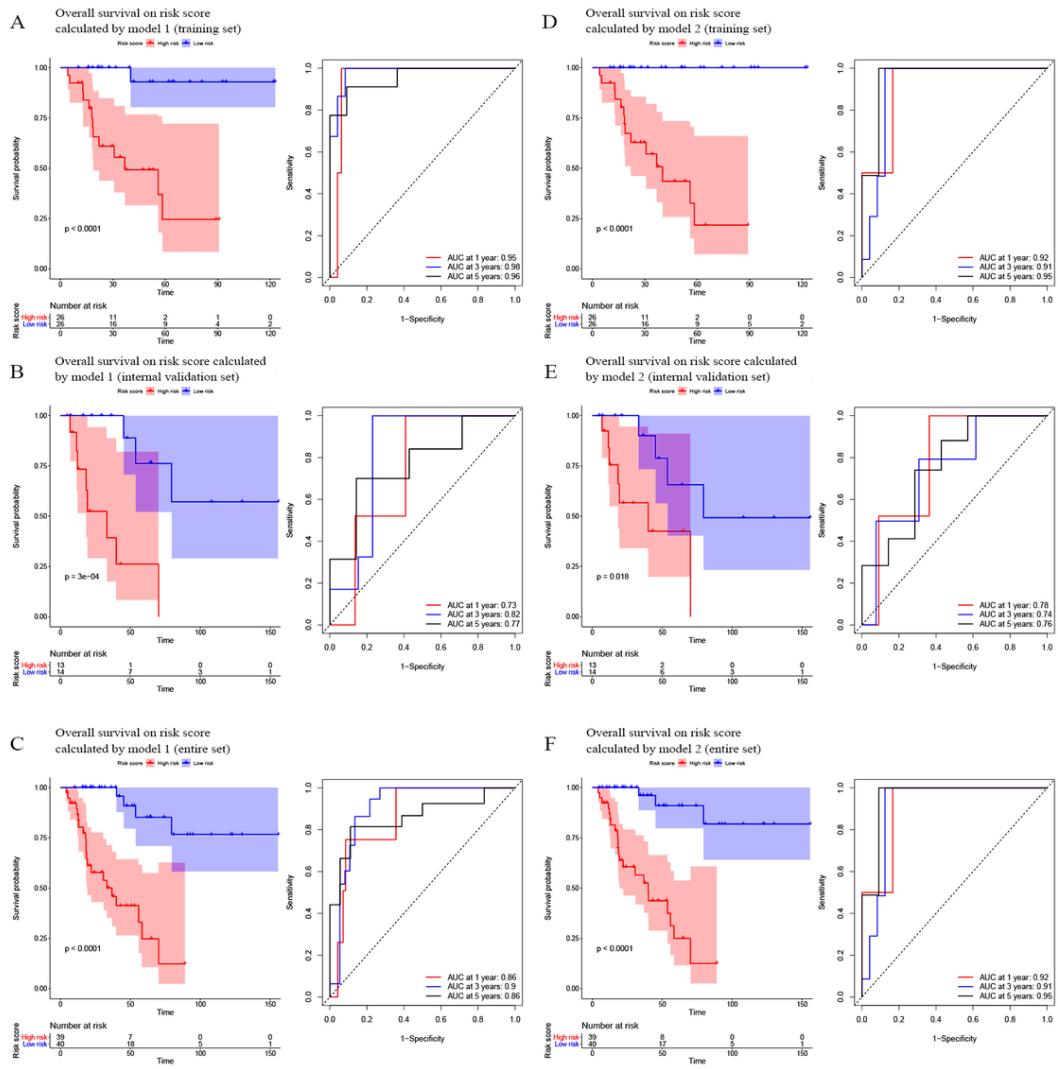


Figure 5

Survival analysis of the association between risk score calculated by model 1 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in (A) training set, (B) internal validation set, (C) entire set. Survival analysis of the association between risk score calculated by model 2 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in (D) training set, (E) internal validation set, (F) entire set.

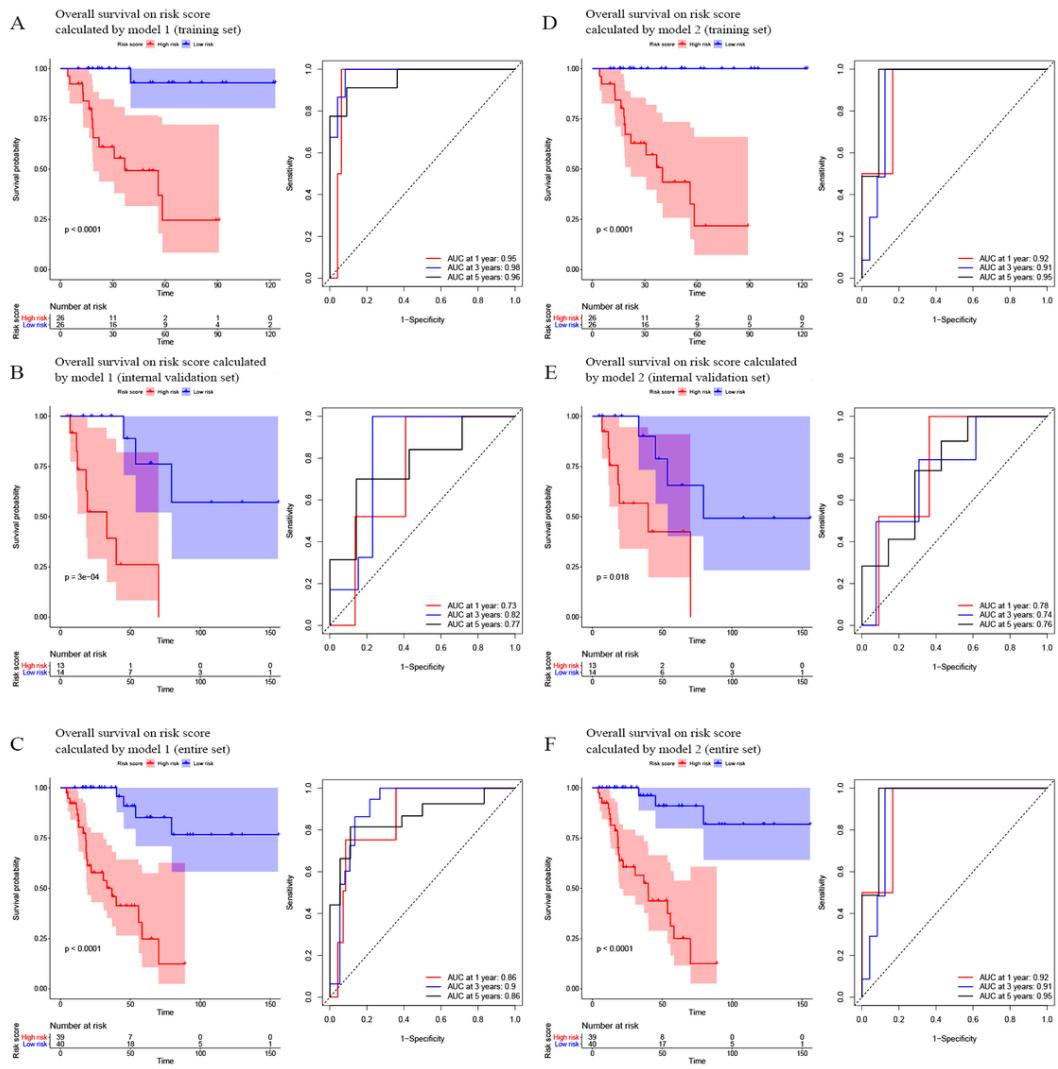
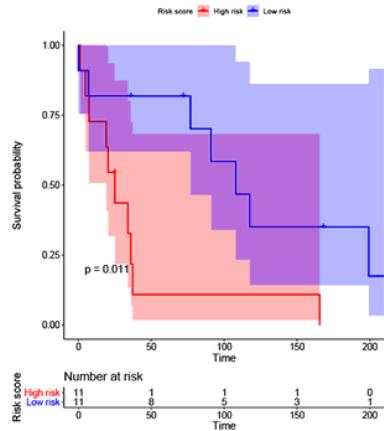


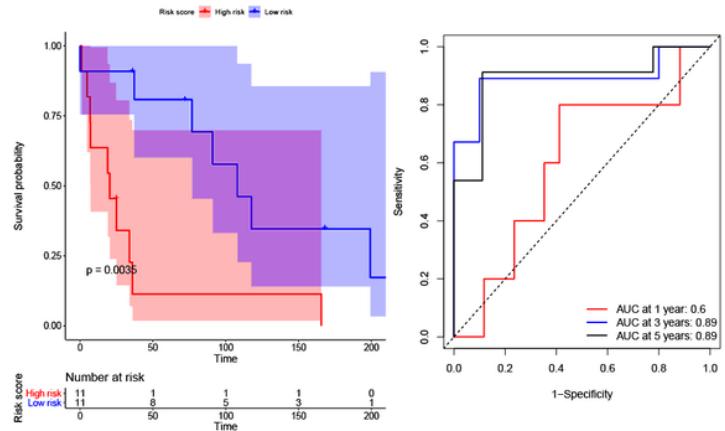
Figure 5

Survival analysis of the association between risk score calculated by model 1 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in (A) training set, (B) internal validation set, (C) entire set. Survival analysis of the association between risk score calculated by model 2 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in (D) training set, (E) internal validation set, (F) entire set.

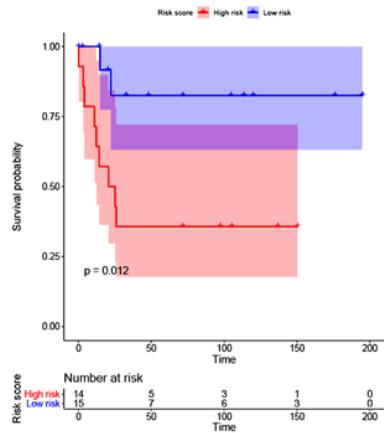
A Overall survival on risk score calculated by model 1 (GSE19750)



C Overall survival on risk score calculated by model 1 (GSE19750)



B Events free survival on risk score calculated by model 1 (GSE76021)



D Events free survival on risk score calculated by model 2 (GSE76021)

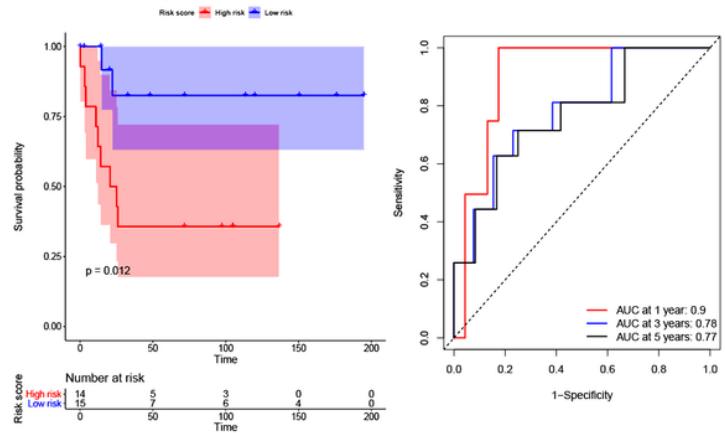
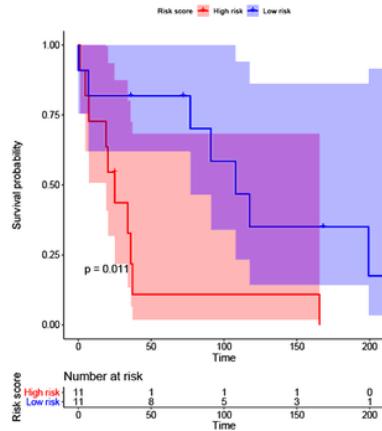


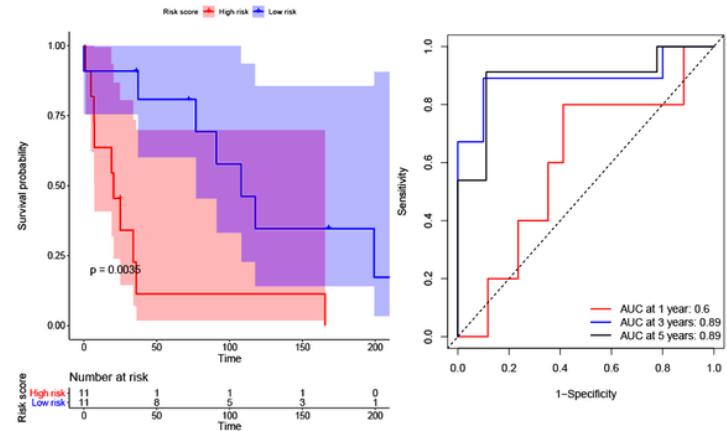
Figure 6

(A) Survival analysis of the association between risk score calculated by model 1 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE19750; (B) Survival analysis of the association between risk score by model 1 and events-free survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE76021; (C) Survival analysis of the association between risk score calculated by model 2 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE19750; (D) Survival analysis of the association between risk score by model 2 and events-free survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE76021.

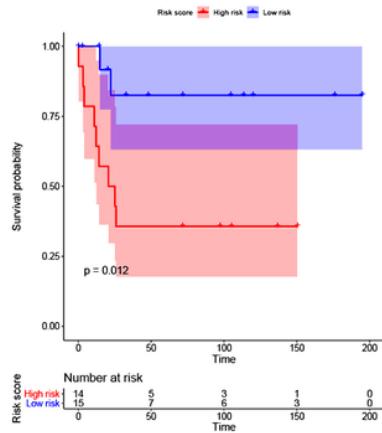
A Overall survival on risk score calculated by model 1 (GSE19750)



C Overall survival on risk score calculated by model 1 (GSE19750)



B Events free survival on risk score calculated by model 1 (GSE76021)



D Events free survival on risk score calculated by model 2 (GSE76021)

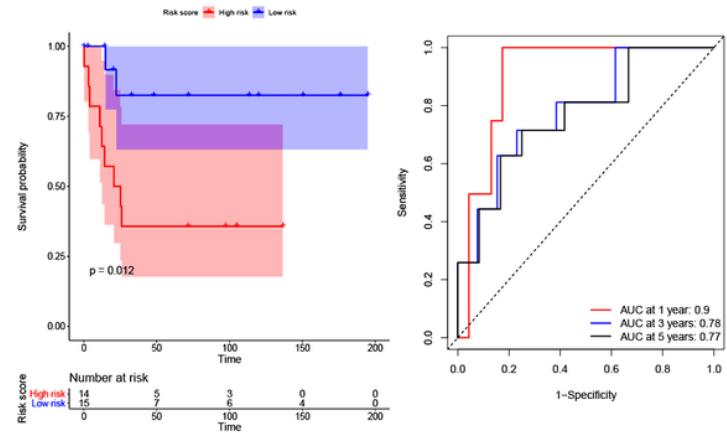
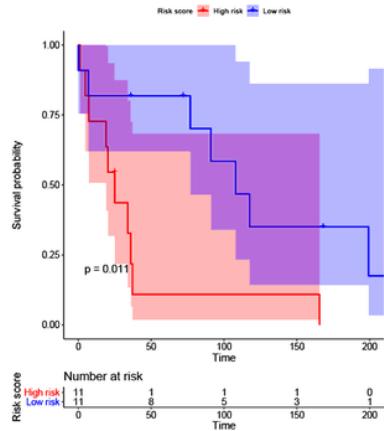


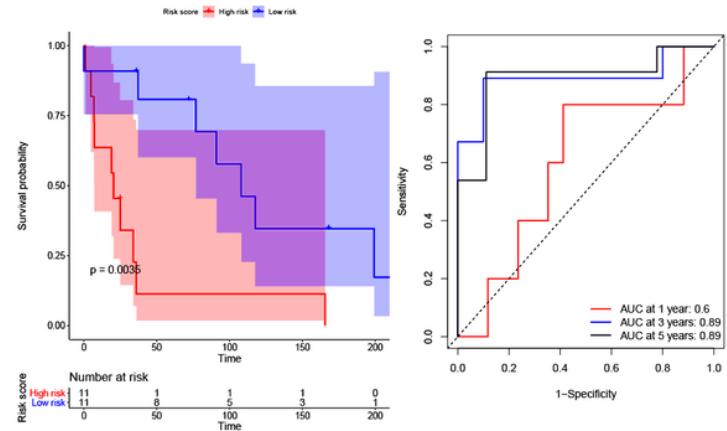
Figure 6

(A) Survival analysis of the association between risk score calculated by model 1 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE19750; (B) Survival analysis of the association between risk score by model 1 and events-free survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE76021; (C) Survival analysis of the association between risk score calculated by model 2 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE19750; (D) Survival analysis of the association between risk score by model 2 and events-free survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE76021.

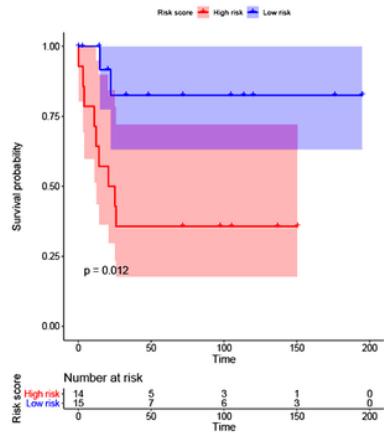
A Overall survival on risk score calculated by model 1 (GSE19750)



C Overall survival on risk score calculated by model 1 (GSE19750)



B Events free survival on risk score calculated by model 1 (GSE76021)



D Events free survival on risk score calculated by model 2 (GSE76021)

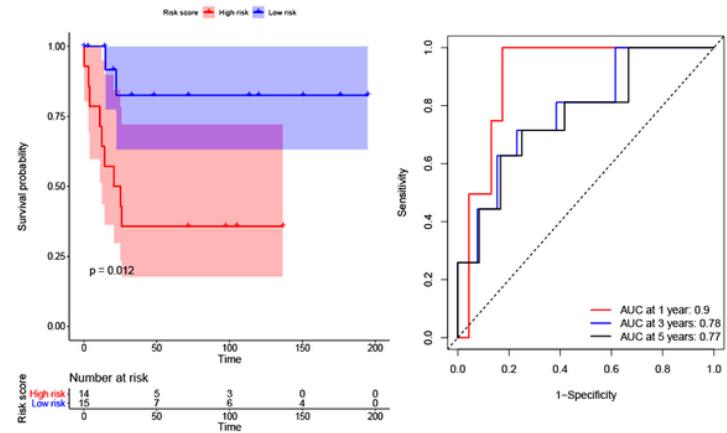


Figure 6

(A) Survival analysis of the association between risk score calculated by model 1 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE19750; (B) Survival analysis of the association between risk score by model 1 and events-free survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE76021; (C) Survival analysis of the association between risk score calculated by model 2 and overall survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE19750; (D) Survival analysis of the association between risk score by model 2 and events-free survival time in ACC, and time dependent ROC analyses at 1,3, and 5 years in GSE76021.

A

Variable	P-value		HR	LCI	UCI
Risk_score (a = 0.1)	<0.001	(blue square)	4.822	2.636	8.821
Gender (female/male)	0.98	(blue square)	1.01	0.453	2.252
Age (<65 vs =65)	0.372	(blue square)	1.631	0.557	4.776
Tumor_stage	<0.001	(blue square)	6.097	2.433	15.282
T_stage (Ta-T1 vs T2-T4)	0.134	(blue square)	4.632	0.622	34.465
N_stage (N0 vs N1-N3)	0.268	(blue square)	1.839	0.626	5.409
M_stage (M0 vs M1)	<0.001	(blue square)	5.806	2.471	13.64
Weiss_score (=6 vs >6)	0.124	(blue square)	1.906	0.838	4.333

0 5 10 15 20 25 30 35

B

Variable	P-value		HR	LCI	UCI
Risk_score (a = 0.1)	<0.001	(blue square)	4.663	2.238	9.713
Tumor_stage	0.234	(blue square)	2.065	0.626	6.808
M_stage (M0 vs M1)	0.566	(blue square)	1.351	0.483	3.774

0 1 2 3 4 5 6 7 8 9 10

C

Variable	P-value		HR	LCI	UCI
Risk_score (a = 0.6)	<0.001	(blue square)	4.465	2.449	8.143
Gender (female/male)	0.98	(blue square)	1.01	0.453	2.252
Age (<65 vs =65)	0.372	(blue square)	1.631	0.557	4.776
Tumor_stage	<0.001	(blue square)	6.097	2.433	15.282
T_stage (Ta-T1 vs T2-T4)	0.134	(blue square)	4.632	0.622	34.465
N_stage (N0 vs N1-N3)	0.268	(blue square)	1.839	0.626	5.409
M_stage (M0 vs M1)	<0.001	(blue square)	5.806	2.471	13.64
Weiss_score (=6 vs >6)	0.124	(blue square)	1.906	0.838	4.333

0 5 10 15 20 25 30 35

D

Variable	P-value		HR	LCI	UCI
Risk_score (a = 0.6)	<0.001	(blue square)	5.603	2.41	13.024
Tumor_stage	0.021	(blue square)	3.803	1.223	11.822
M_stage (M0 vs M1)	0.867	(blue square)	0.909	0.299	2.767

0 1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 7

(A) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 1, gender, age, tumor stage, TNM stage and Weiss score by using TCGA-ACC data. (B) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 1, tumor stage and M stage by using TCGA-ACC data. (C) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 2, gender, age, tumor stage, TNM stage and Weiss score by using TCGA-ACC data. (D) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 2, tumor stage and M stage by using TCGA-ACC data.HR: hazard ratio; OS: overall survival.

A

Variable	P-value		HR	LCI	UCI
Risk_score (a = 0.1)	<0.001	(blue square)	4.822	2.636	8.821
Gender (female/male)	0.98	(blue square)	1.01	0.453	2.252
Age (<65 vs =65)	0.372	(blue square)	1.631	0.557	4.776
Tumor_stage	<0.001	(blue square)	6.097	2.433	15.282
T_stage (Ta-T1 vs T2-T4)	0.134	(blue square)	4.632	0.622	34.465
N_stage (N0 vs N1-N3)	0.268	(blue square)	1.839	0.626	5.409
M_stage (M0 vs M1)	<0.001	(blue square)	5.806	2.471	13.64
Weiss_score (=6 vs >6)	0.124	(blue square)	1.906	0.838	4.333

0 5 10 15 20 25 30 35

B

Variable	P-value		HR	LCI	UCI
Risk_score (a = 0.1)	<0.001	(blue square)	4.663	2.238	9.713
Tumor_stage	0.234	(blue square)	2.065	0.626	6.808
M_stage (M0 vs M1)	0.566	(blue square)	1.351	0.483	3.774

0 1 2 3 4 5 6 7 8 9 10

C

Variable	P-value		HR	LCI	UCI
Risk_score (a = 0.6)	<0.001	(blue square)	4.465	2.449	8.143
Gender (female/male)	0.98	(blue square)	1.01	0.453	2.252
Age (<65 vs =65)	0.372	(blue square)	1.631	0.557	4.776
Tumor_stage	<0.001	(blue square)	6.097	2.433	15.282
T_stage (Ta-T1 vs T2-T4)	0.134	(blue square)	4.632	0.622	34.465
N_stage (N0 vs N1-N3)	0.268	(blue square)	1.839	0.626	5.409
M_stage (M0 vs M1)	<0.001	(blue square)	5.806	2.471	13.64
Weiss_score (=6 vs >6)	0.124	(blue square)	1.906	0.838	4.333

0 5 10 15 20 25 30 35

D

Variable	P-value		HR	LCI	UCI
Risk_score (a = 0.6)	<0.001	(blue square)	5.603	2.41	13.024
Tumor_stage	0.021	(blue square)	3.803	1.223	11.822
M_stage (M0 vs M1)	0.867	(blue square)	0.909	0.299	2.767

0 1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 7

(A) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 1, gender, age, tumor stage, TNM stage and Weiss score by using TCGA-ACC data. (B) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 1, tumor stage and M stage by using TCGA-ACC data. (C) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 2, gender, age, tumor stage, TNM stage and Weiss score by using TCGA-ACC data. (D) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 2, tumor stage and M stage by using TCGA-ACC data.HR: hazard ratio; OS: overall survival.

A

Variable	P-value	HR	LCI	UCI
Risk_score (a = 0.1)	<0.001	4.822	2.636	8.821
Gender (female/male)	0.98	1.01	0.453	2.252
Age (<65 vs =65)	0.372	1.631	0.557	4.776
Tumor_stage	<0.001	6.097	2.433	15.282
T_stage (Ta-T1 vs T2-T4)	0.134	4.632	0.622	34.465
N_stage (N0 vs N1-N3)	0.268	1.839	0.626	5.409
M_stage (M0 vs M1)	<0.001	5.806	2.471	13.64
Weiss_score (=6 vs >6)	0.124	1.906	0.838	4.333

0 5 10 15 20 25 30 35

B

Variable	P-value	HR	LCI	UCI
Risk_score (a = 0.1)	<0.001	4.663	2.238	9.713
Tumor_stage	0.234	2.065	0.626	6.808
M_stage (M0 vs M1)	0.566	1.351	0.483	3.774

0 1 2 3 4 5 6 7 8 9 10

C

Variable	P-value	HR	LCI	UCI
Risk_score (a = 0.6)	<0.001	4.465	2.449	8.143
Gender (female/male)	0.98	1.01	0.453	2.252
Age (<65 vs =65)	0.372	1.631	0.557	4.776
Tumor_stage	<0.001	6.097	2.433	15.282
T_stage (Ta-T1 vs T2-T4)	0.134	4.632	0.622	34.465
N_stage (N0 vs N1-N3)	0.268	1.839	0.626	5.409
M_stage (M0 vs M1)	<0.001	5.806	2.471	13.64
Weiss_score (=6 vs >6)	0.124	1.906	0.838	4.333

0 5 10 15 20 25 30 35

D

Variable	P-value	HR	LCI	UCI
Risk_score (a = 0.6)	<0.001	5.603	2.41	13.024
Tumor_stage	0.021	3.803	1.223	11.822
M_stage (M0 vs M1)	0.867	0.909	0.299	2.767

0 1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 7

(A) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 1, gender, age, tumor stage, TNM stage and Weiss score by using TCGA-ACC data. (B) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 1, tumor stage and M stage by using TCGA-ACC data. (C) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 2, gender, age, tumor stage, TNM stage and Weiss score by using TCGA-ACC data. (D) Forest plot summary of analyses of OS univariate analysis of Risk score calculated by model 2, tumor stage and M stage by using TCGA-ACC data. HR: hazard ratio; OS: overall survival.

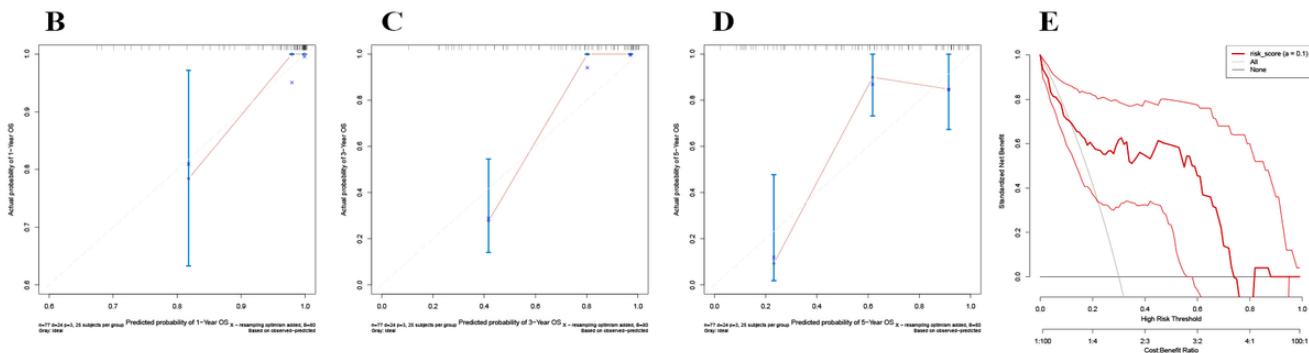
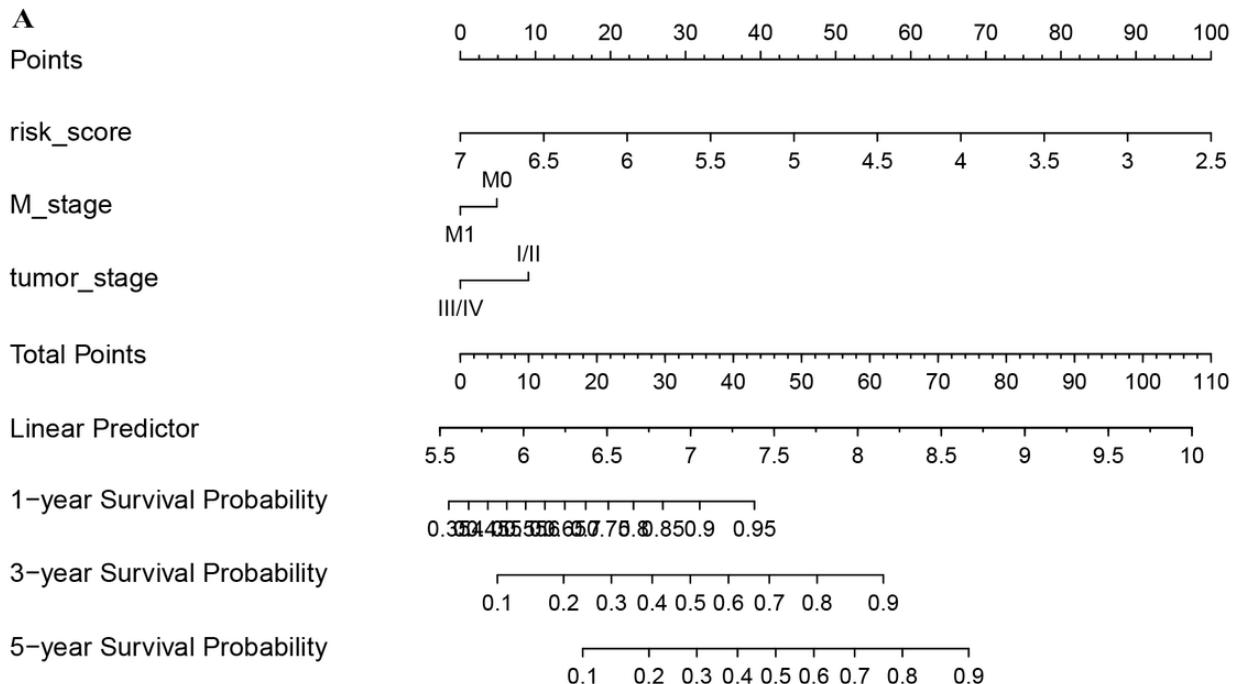


Figure 8

(A) The nomogram constructed with model 1 for predicting proportion of patients with 1-, 3- or 5-year OS. The calibration plots for predicting 1- (B), 3- (C) or 5- (D) year OS. (E) DCA for assessment of the clinical utility of the model 1, the x-axis represents the percentage of threshold probability, and the y-axis represents the net benefit. OS: overall survival; DCA: decision curve analysis.

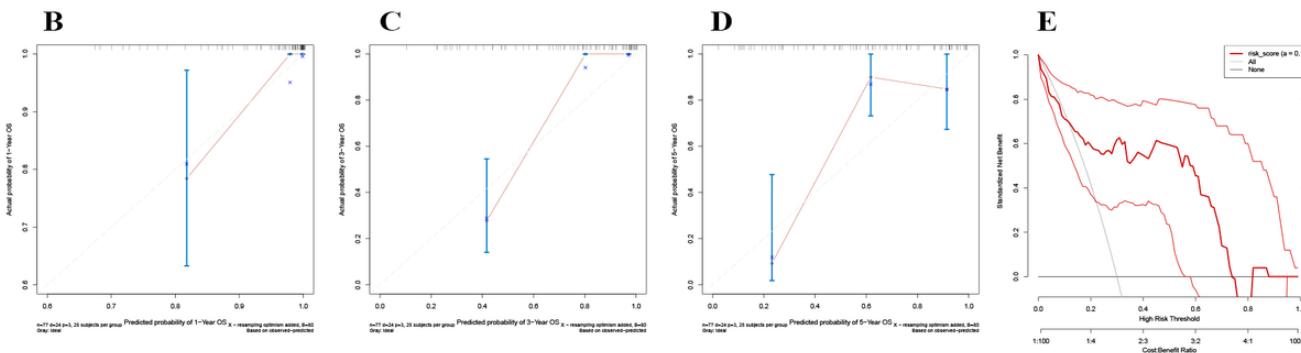
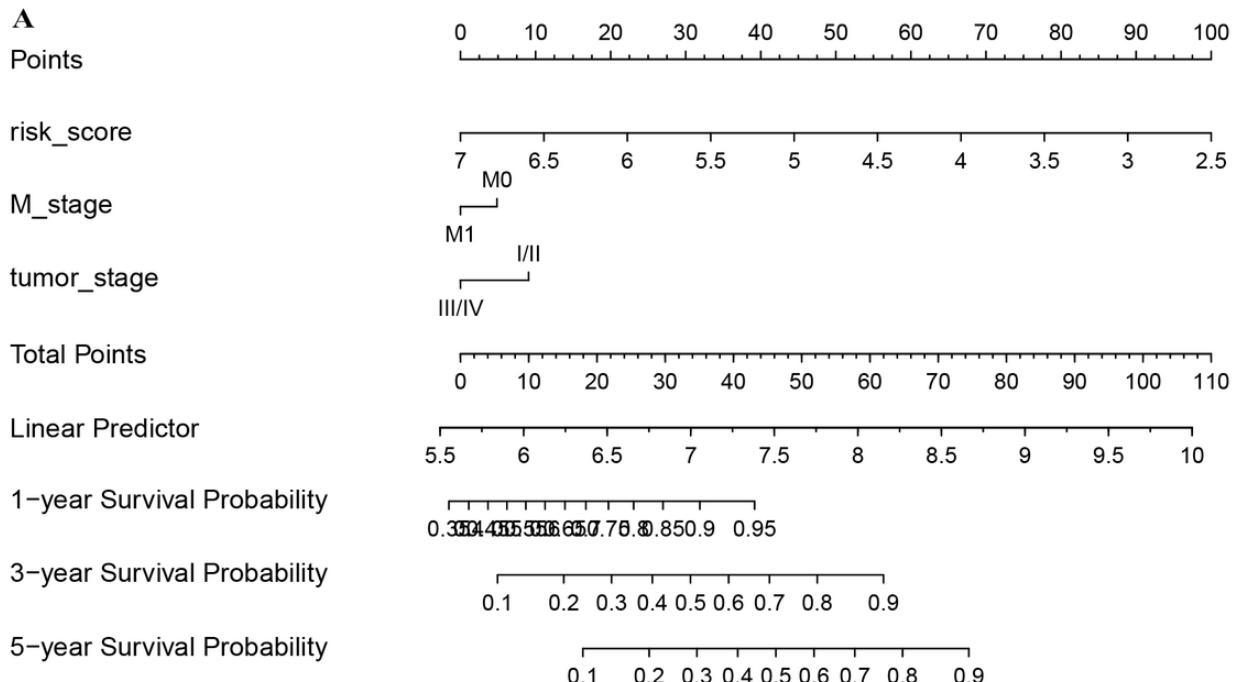


Figure 8

(A) The nomogram constructed with model 1 for predicting proportion of patients with 1-, 3- or 5-year OS. The calibration plots for predicting 1- (B), 3- (C) or 5- (D) year OS. (E) DCA for assessment of the clinical utility of the model 1, the x-axis represents the percentage of threshold probability, and the y-axis represents the net benefit. OS: overall survival; DCA: decision curve analysis.

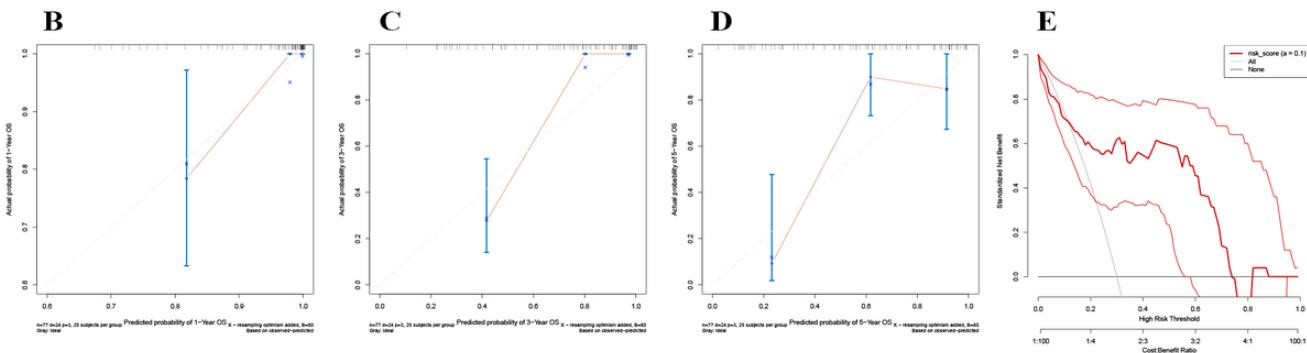
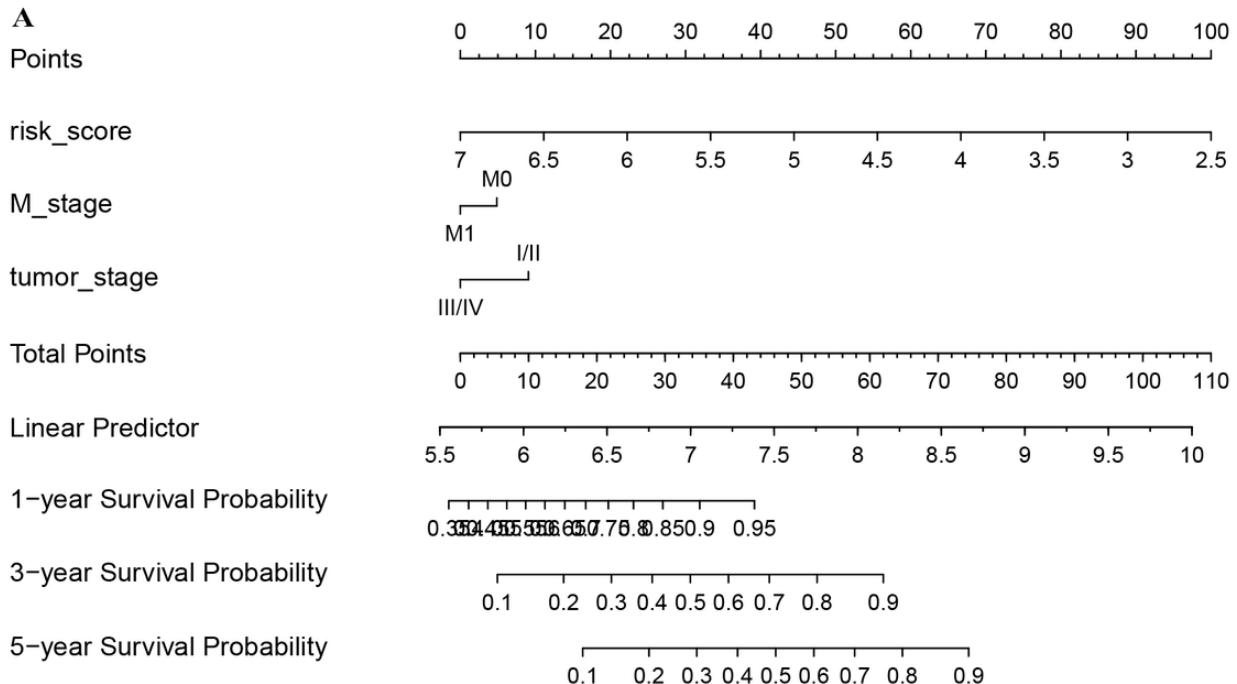


Figure 8

(A) The nomogram constructed with model 1 for predicting proportion of patients with 1-, 3- or 5-year OS. The calibration plots for predicting 1- (B), 3- (C) or 5- (D) year OS. (E) DCA for assessment of the clinical utility of the model 1, the x-axis represents the percentage of threshold probability, and the y-axis represents the net benefit. OS: overall survival; DCA: decision curve analysis.

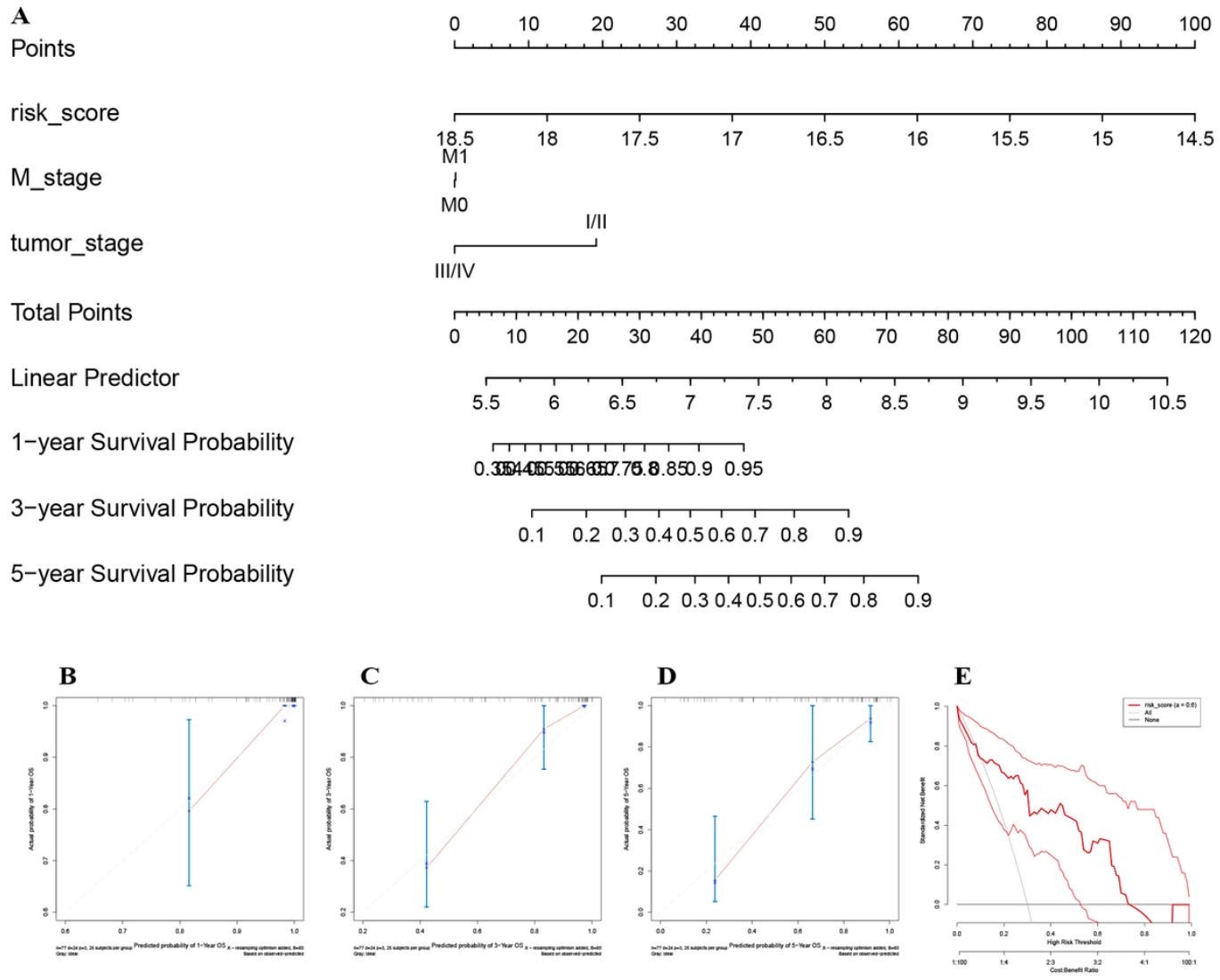


Figure 9

(A) The nomogram constructed with model 2 for predicting proportion of patients with 1-, 3- or 5-year OS. The calibration plots for predicting 1- (B), 3- (C) or 5- (D) year OS. (E) DCA for assessment of the clinical utility of model 2, the x-axis represents the percentage of threshold probability, and the y-axis represents the net benefit. OS: overall survival; DCA: decision curve analysis.

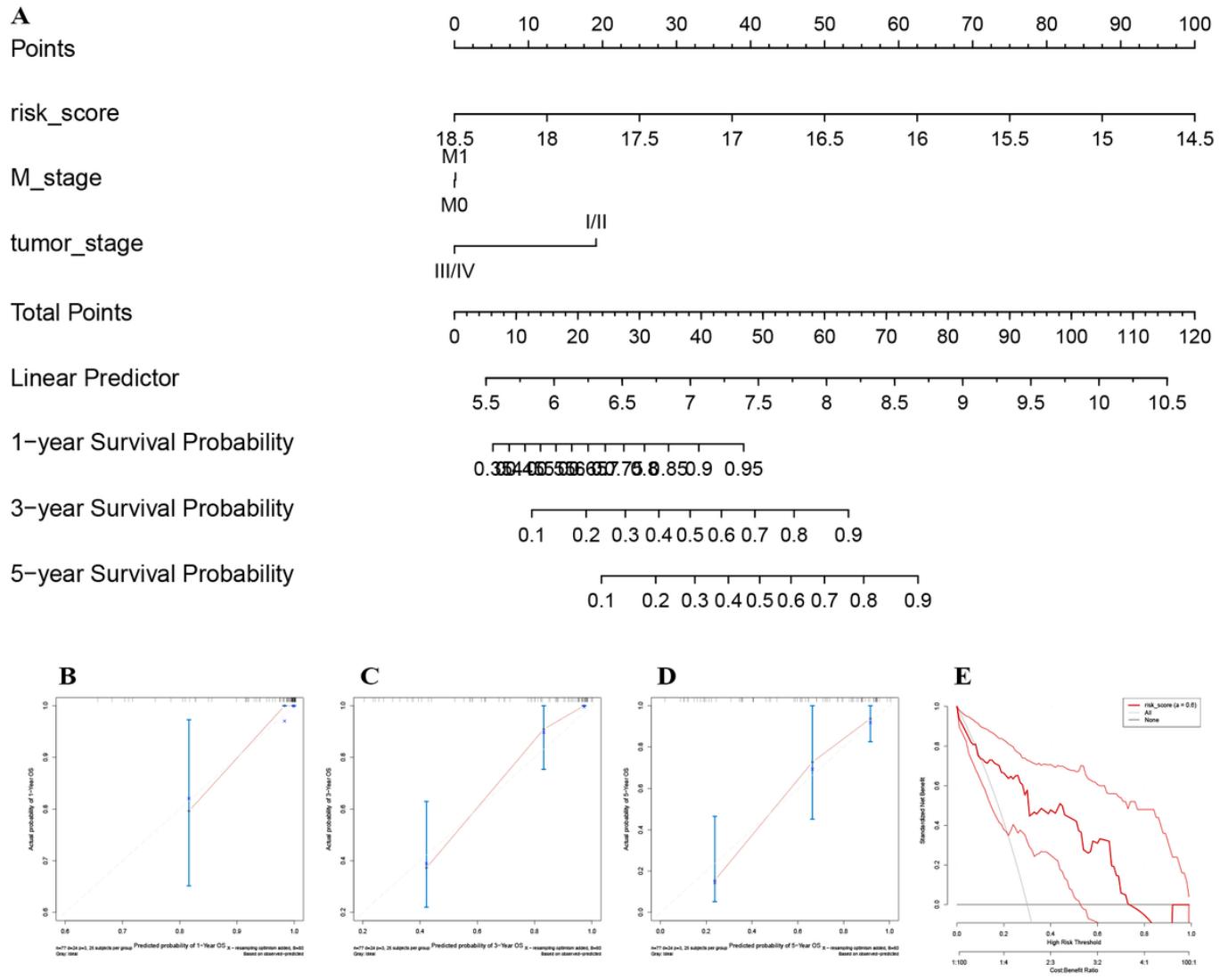


Figure 9

(A) The nomogram constructed with model 2 for predicting proportion of patients with 1-, 3- or 5-year OS. The calibration plots for predicting 1- (B), 3- (C) or 5- (D) year OS. (E) DCA for assessment of the clinical utility of model 2, the x-axis represents the percentage of threshold probability, and the y-axis represents the net benefit. OS: overall survival; DCA: decision curve analysis.

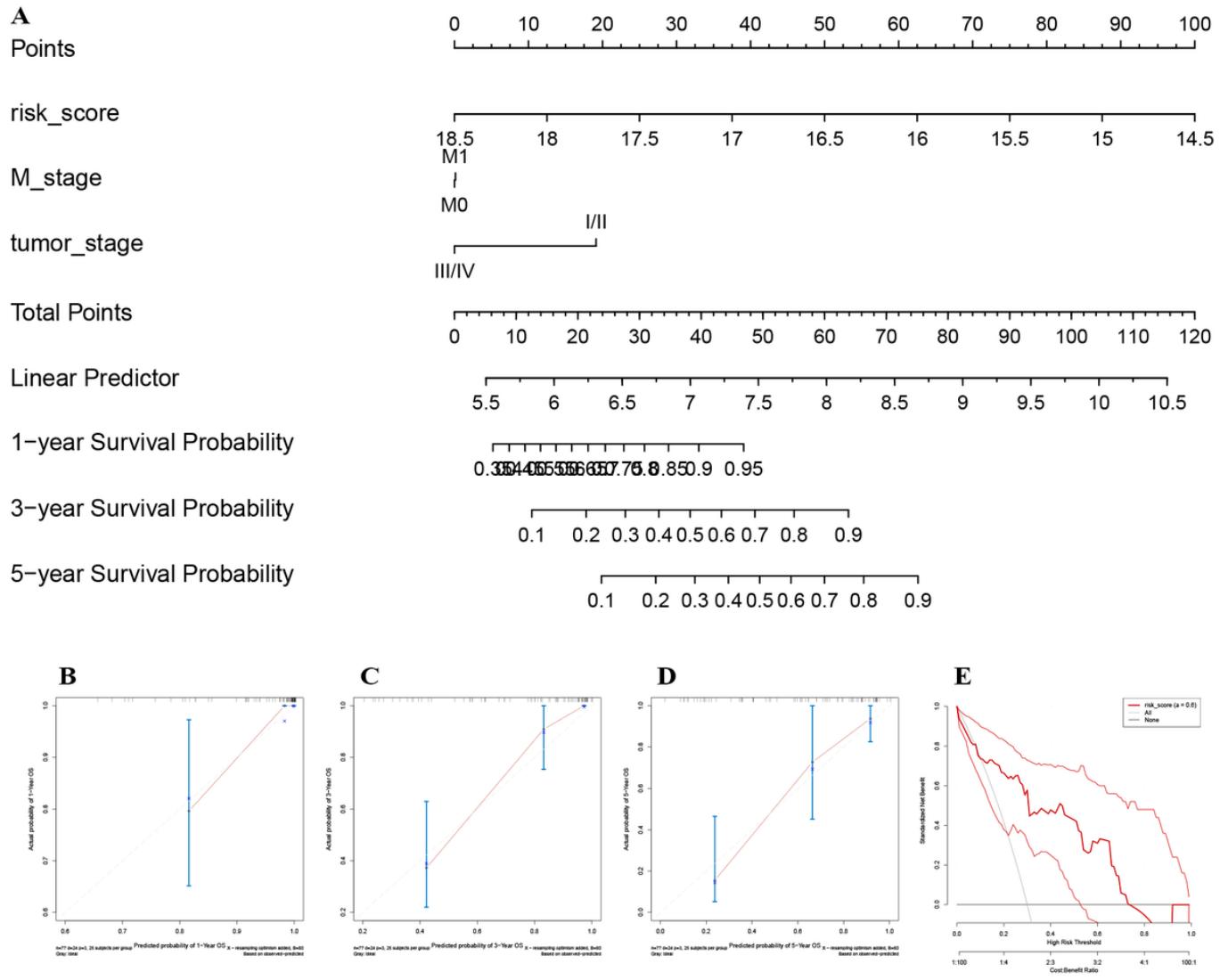


Figure 9

(A) The nomogram constructed with model 2 for predicting proportion of patients with 1-, 3- or 5-year OS. The calibration plots for predicting 1- (B), 3- (C) or 5- (D) year OS. (E) DCA for assessment of the clinical utility of model 2, the x-axis represents the percentage of threshold probability, and the y-axis represents the net benefit. OS: overall survival; DCA: decision curve analysis.

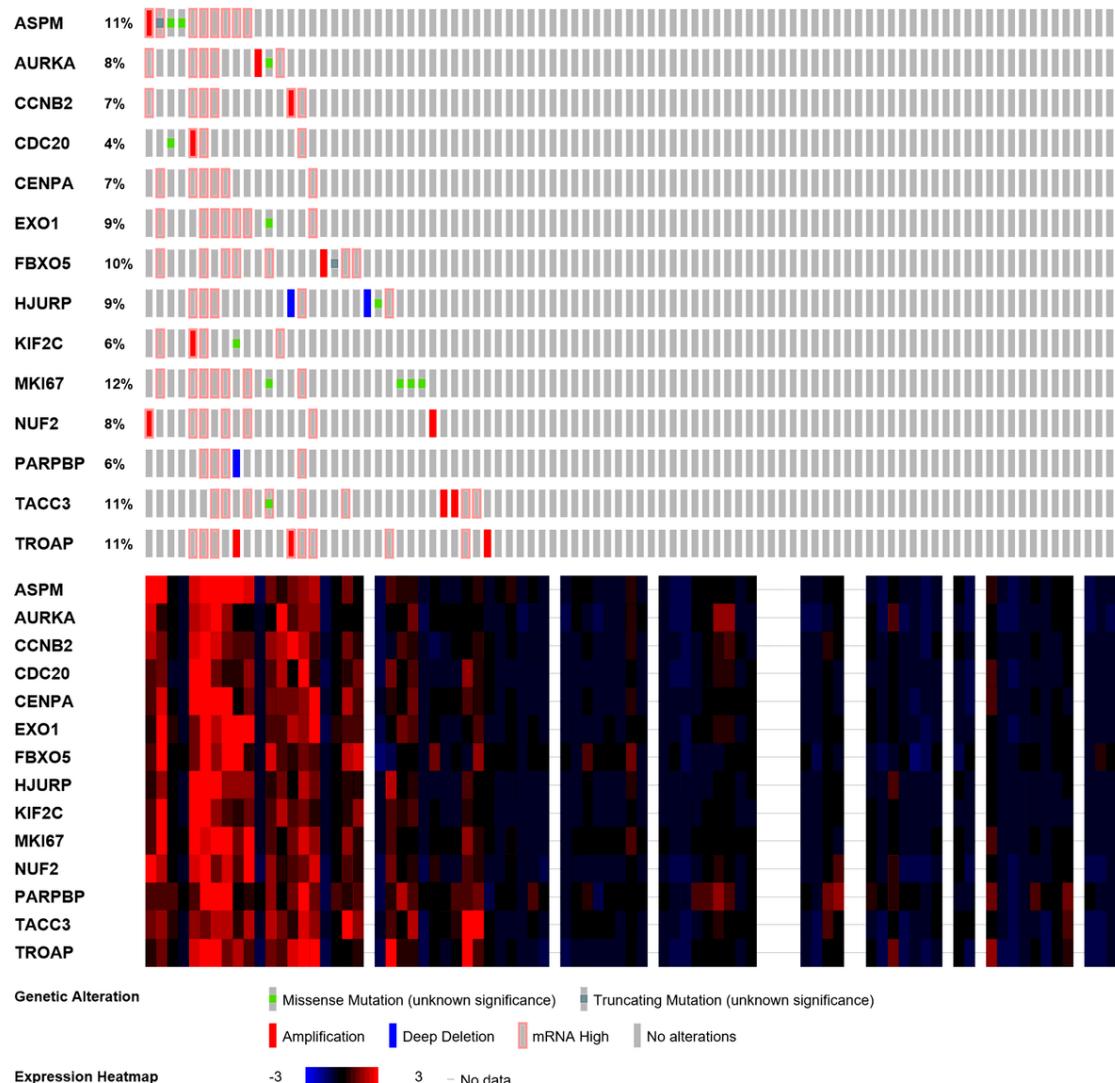


Figure 10

Genetic alterations associated with genes in model 2 and expression heatmap of genes in model 2 based on the data from TCGA.

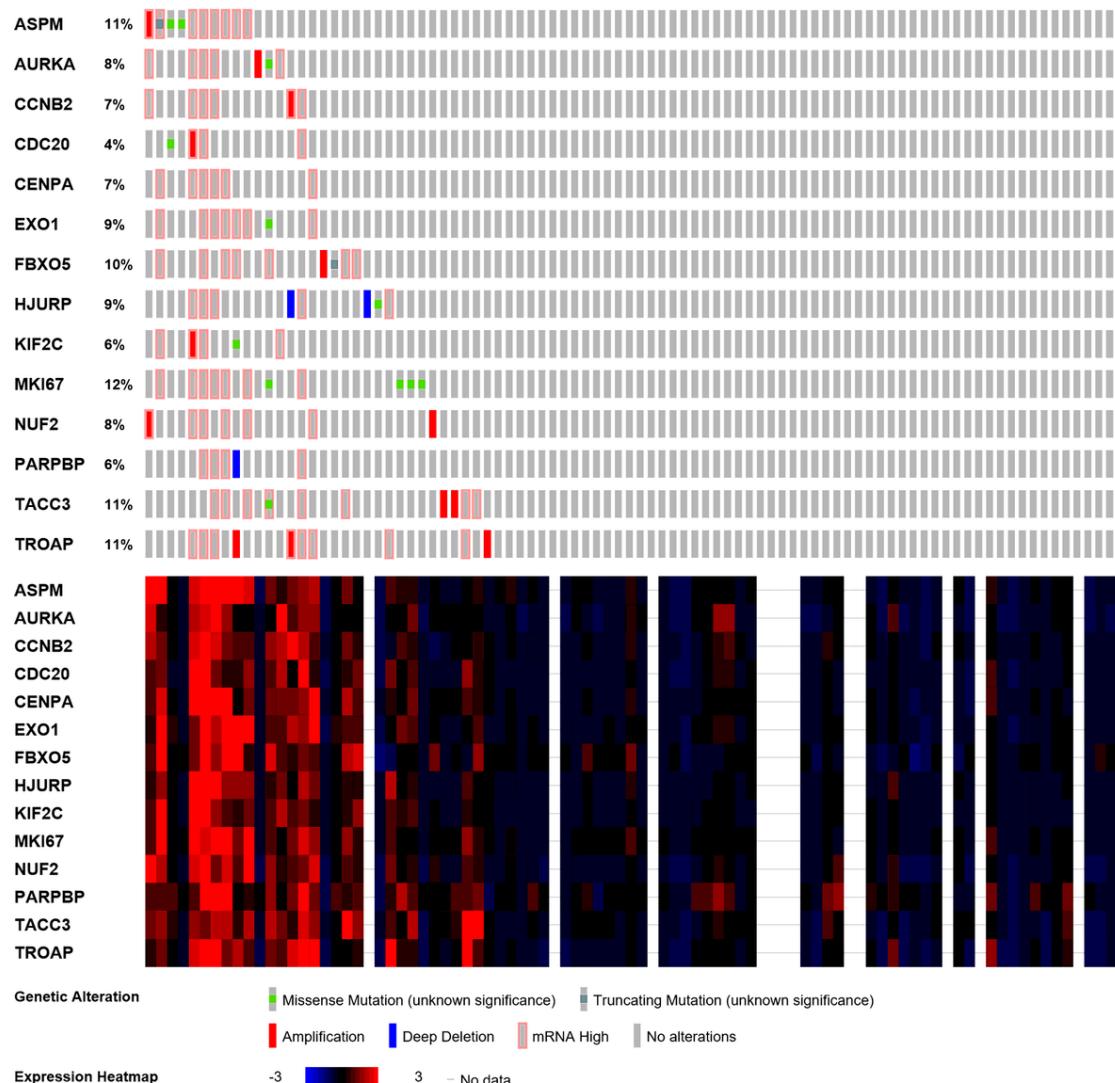


Figure 10

Genetic alterations associated with genes in model 2 and expression heatmap of genes in model 2 based on the data from TCGA.

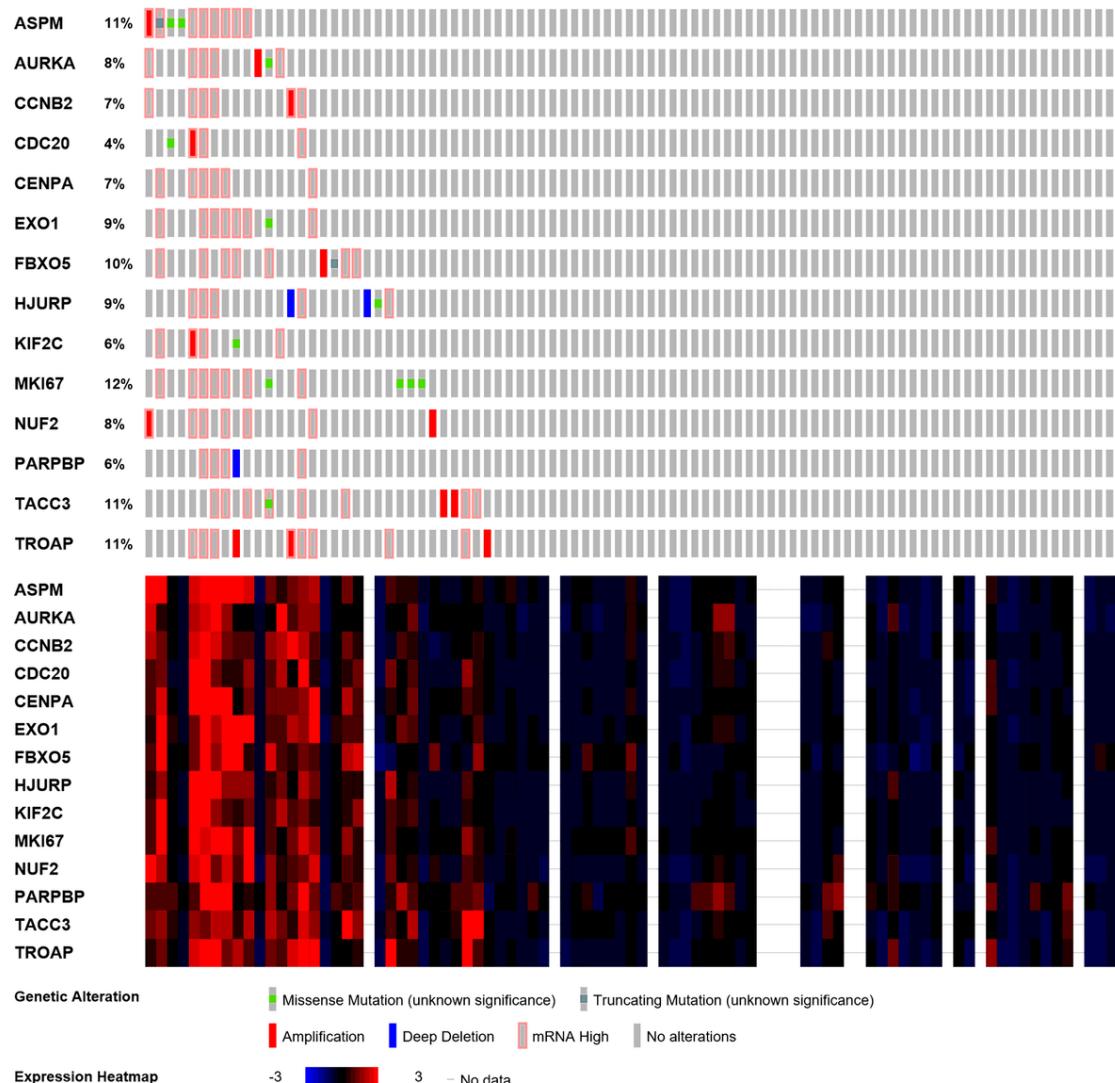
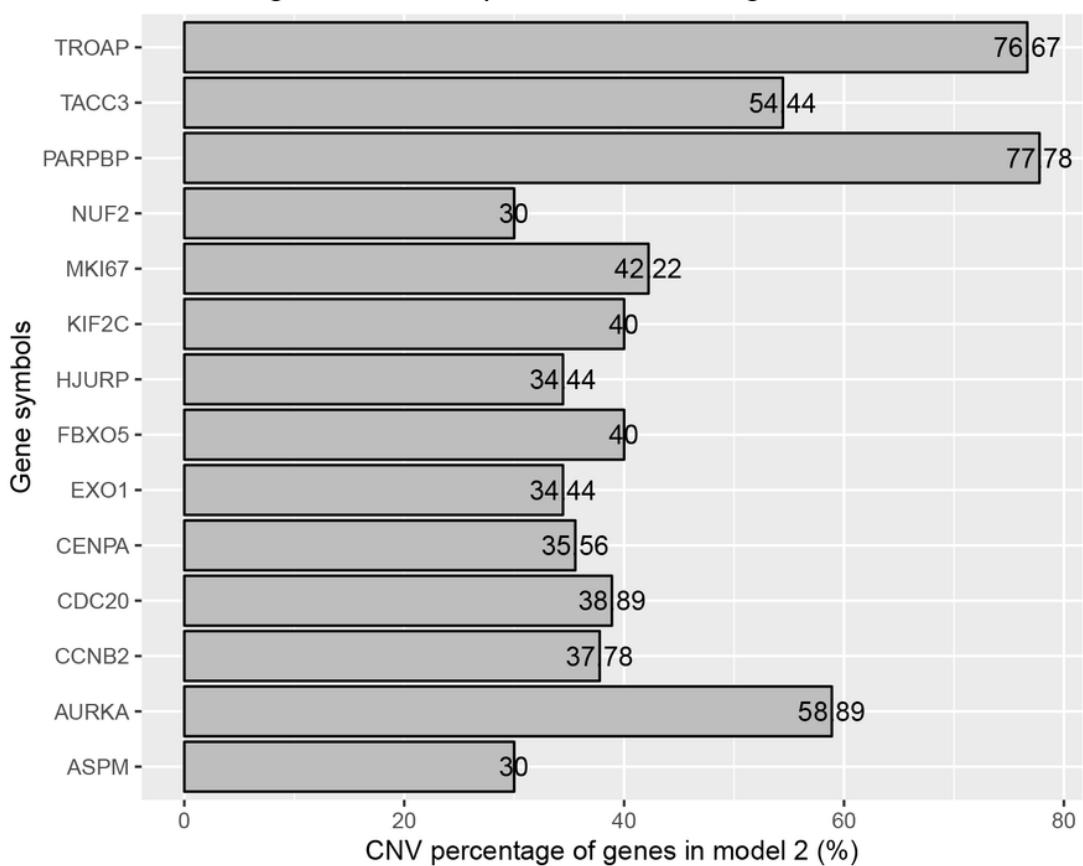
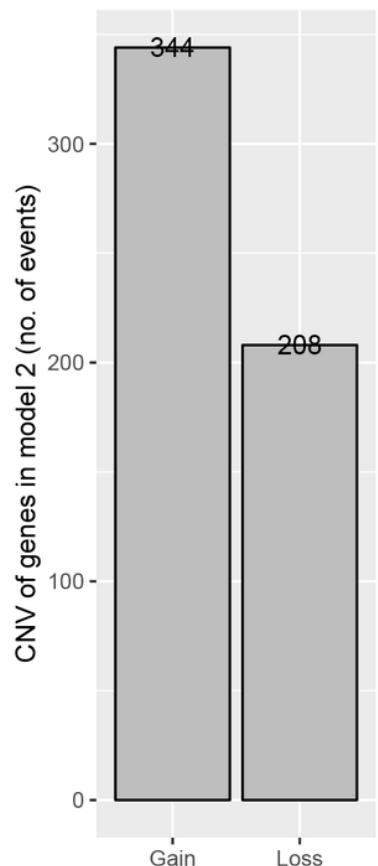


Figure 10

Genetic alterations associated with genes in model 2 and expression heatmap of genes in model 2 based on the data from TCGA.

A

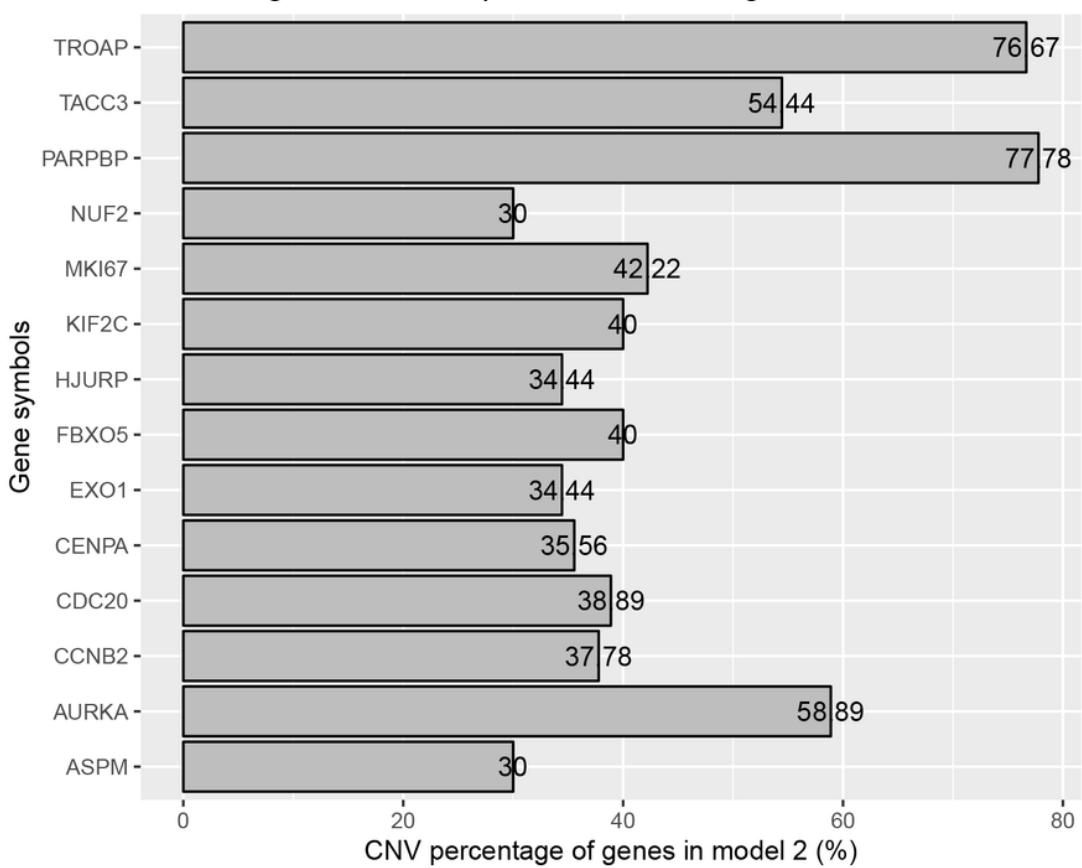
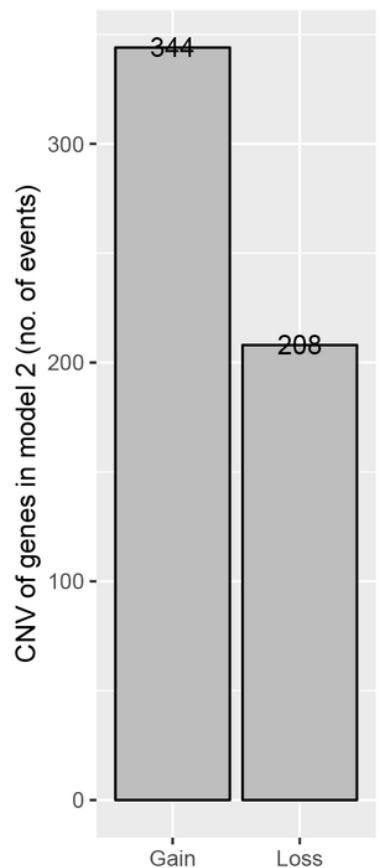
Percentage of ACC samples with CNVs of genes in model 2

**B****Figure 11**

CNVs of genes in model 2 in ACC. (A) Percentage of ACC samples with CNVs of genes in model 2 based on TCGA ACC data. (B) Events of copy number gain or loss of genes in model 2 in ACC samples.

A

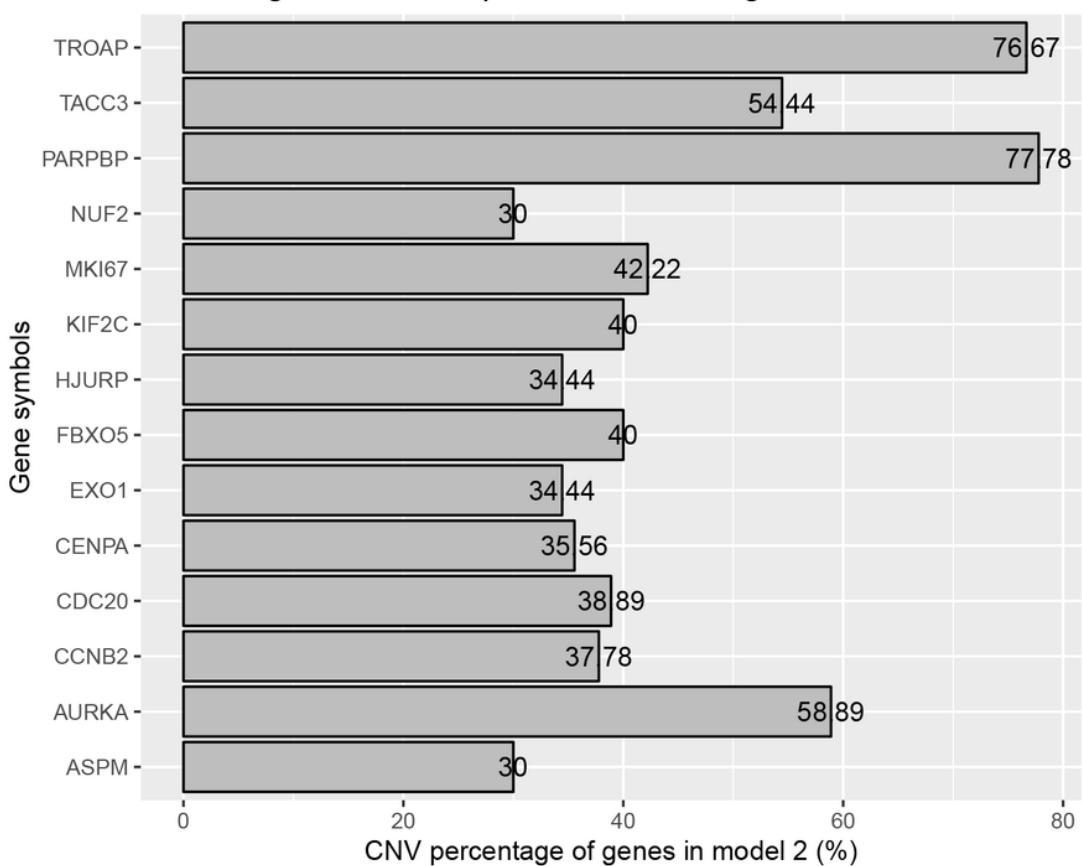
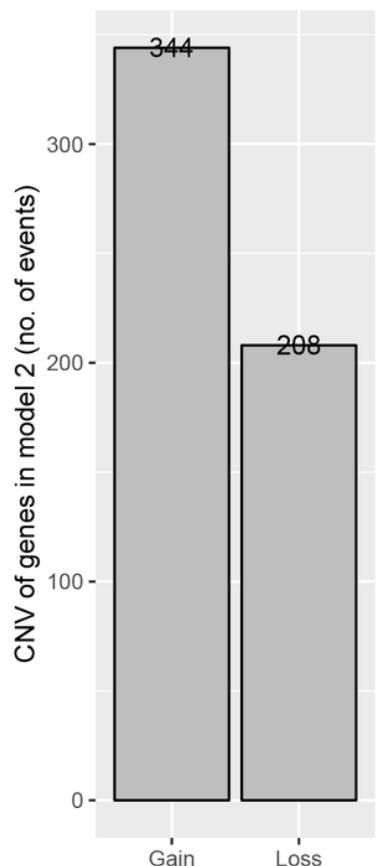
Percentage of ACC samples with CNVs of genes in model 2

**B****Figure 11**

CNVs of genes in model 2 in ACC. (A) Percentage of ACC samples with CNVs of genes in model 2 based on TCGA ACC data. (B) Events of copy number gain or loss of genes in model 2 in ACC samples.

A

Percentage of ACC samples with CNVs of genes in model 2

**B****Figure 11**

CNVs of genes in model 2 in ACC. (A) Percentage of ACC samples with CNVs of genes in model 2 based on TCGA ACC data. (B) Events of copy number gain or loss of genes in model 2 in ACC samples.

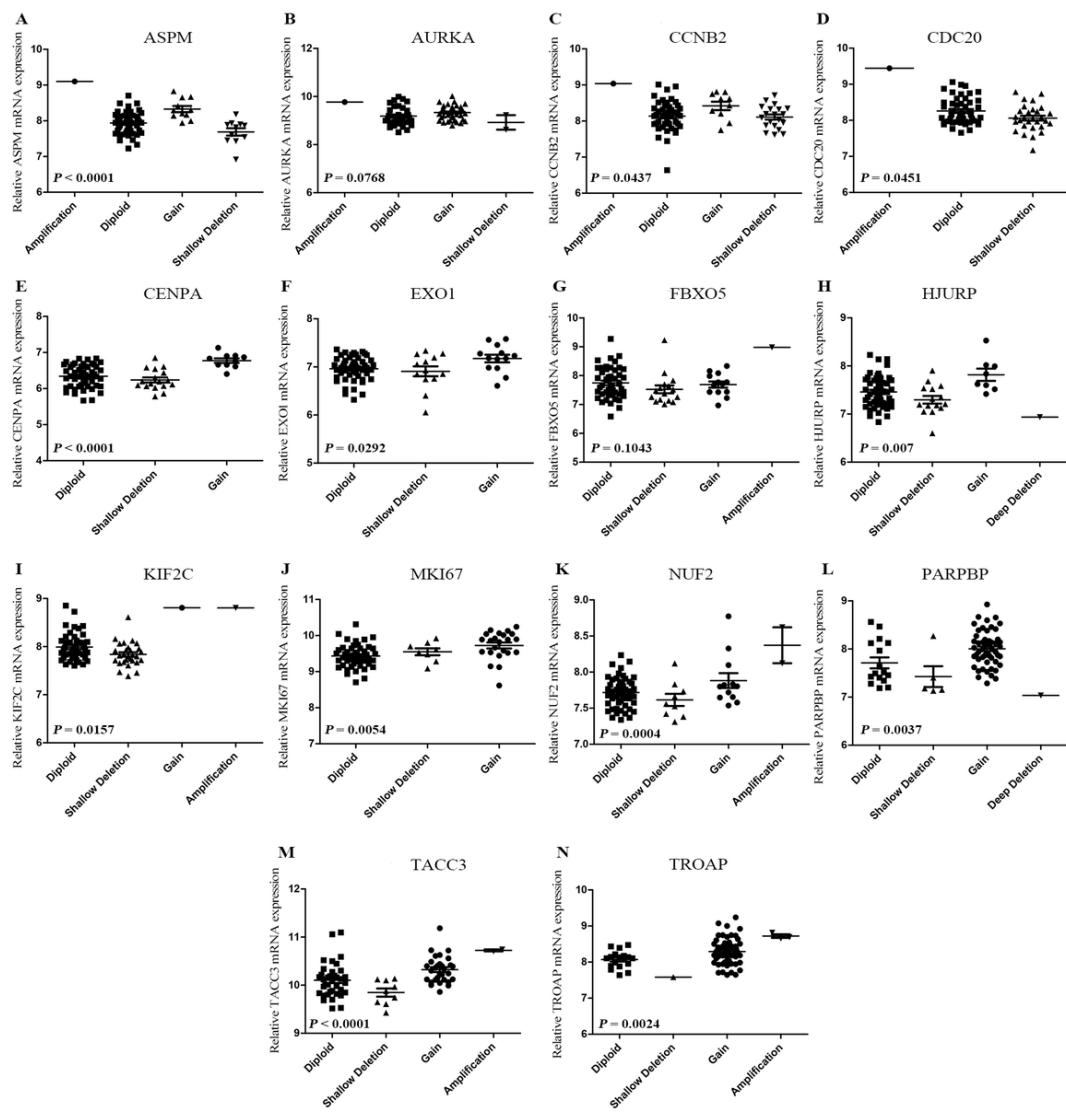


Figure 12

Correlation between different CNV patterns and mRNA expression levels of genes in model 2 respectively.

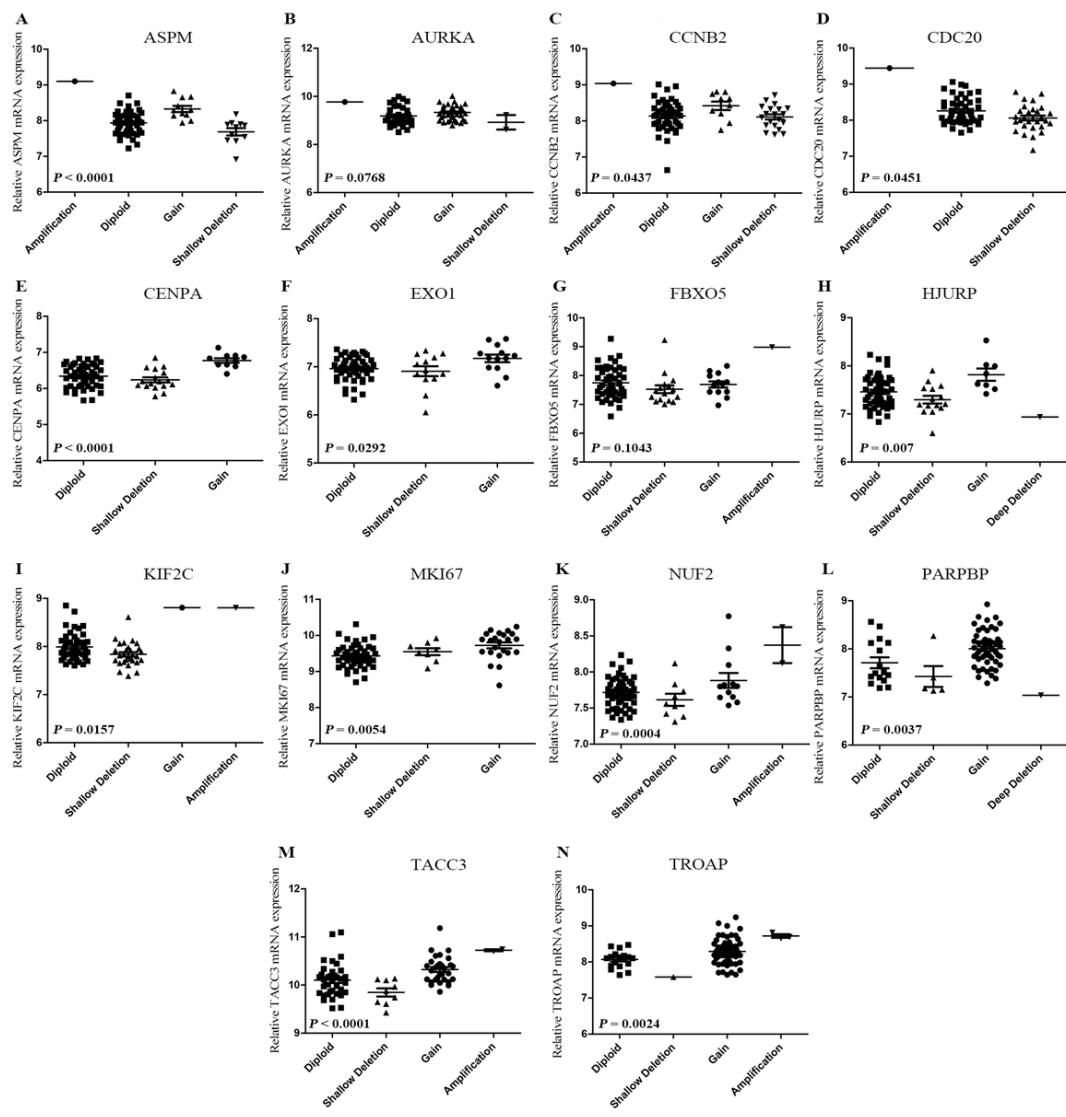


Figure 12

Correlation between different CNV patterns and mRNA expression levels of genes in model 2 respectively.

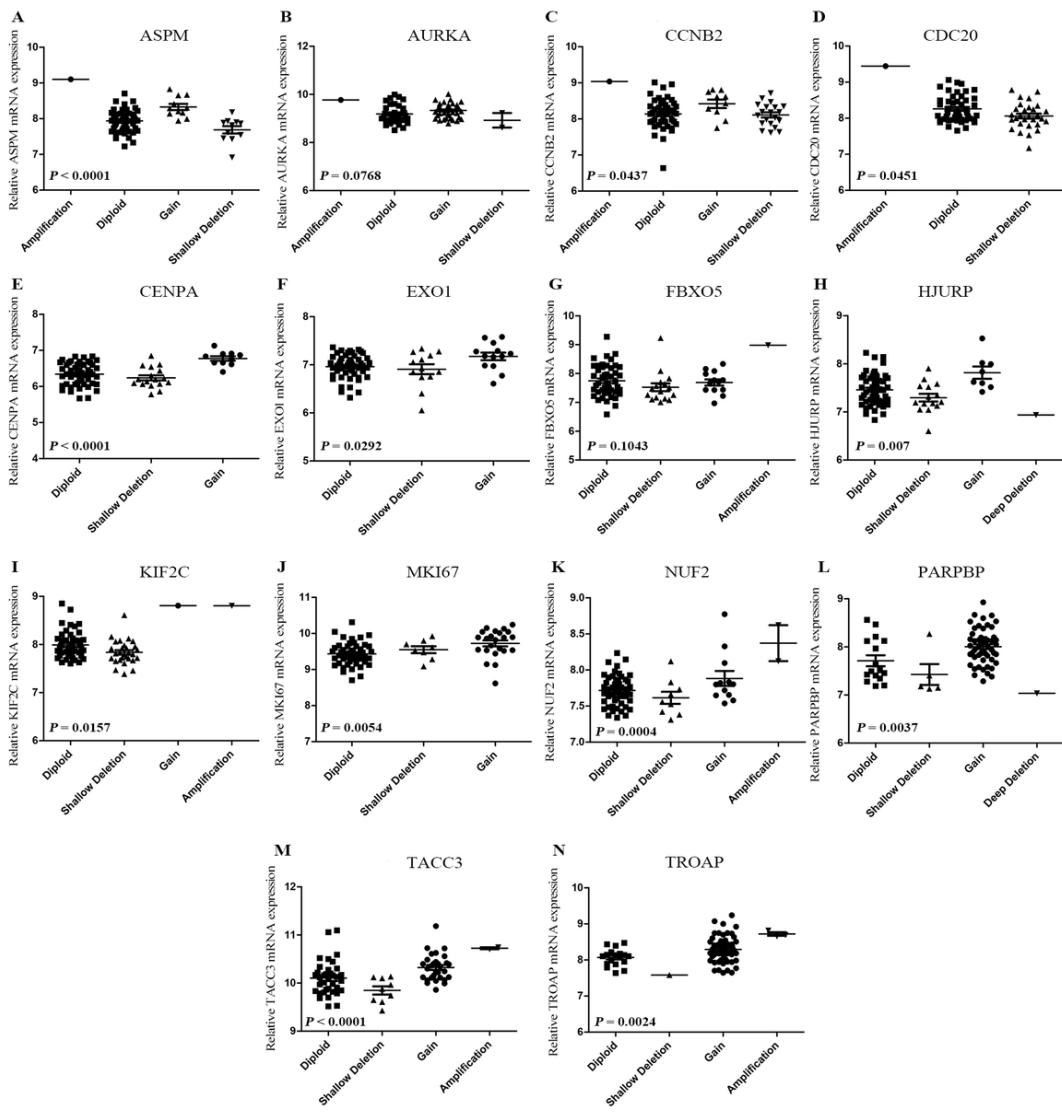


Figure 12

Correlation between different CNV patterns and mRNA expression levels of genes in model 2 respectively.

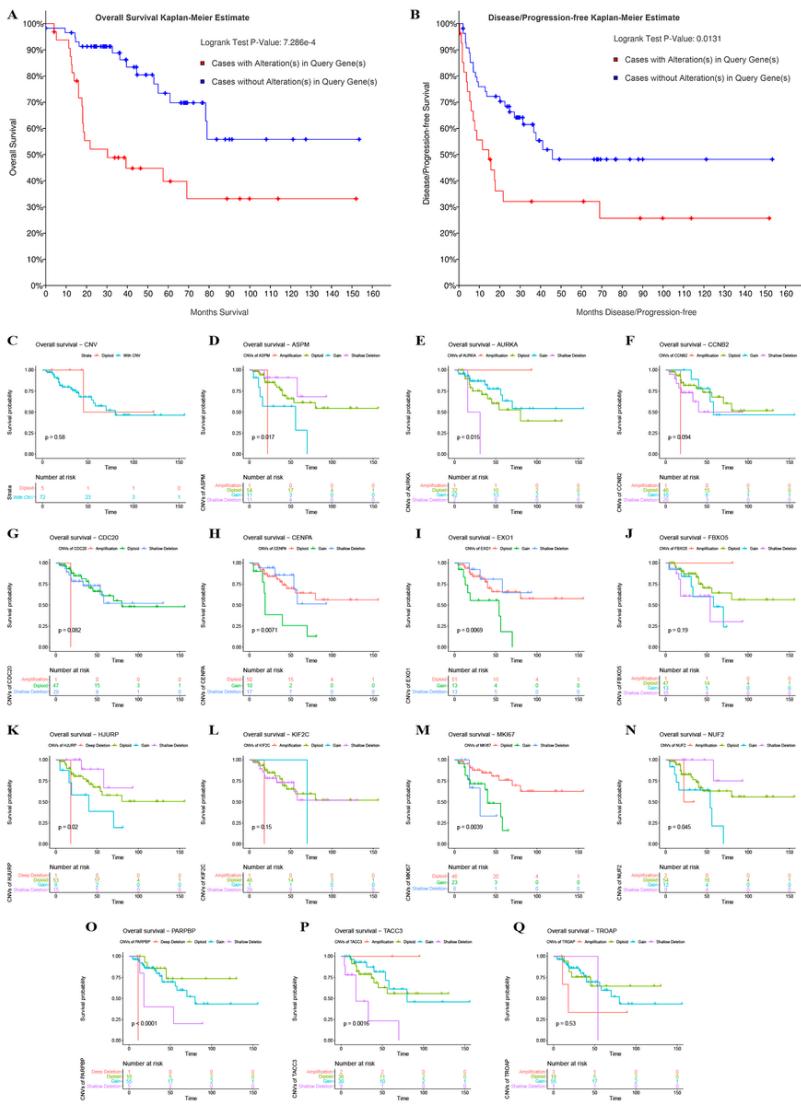


Figure 13

Survival analysis of ACC patients with CNVs of genes in model 2 based on TCGA ACC data. Overall survival (A) and Disease-free survival (B) of patients with alterations in genes in model 2 or patients without alterations in genes in model 2. (C) Overall survival of patients with any CNVs of genes in model 2 or patients with diploid. Overall survival of patients with different CNV types of ASPM (D), AURKA (E), CCNB2 (F), CDC20 (G), CENPA (H), EXO1 (I), FBXO5 (J), HJURP (K), KIF2C (L), MKI67 (M), NUF2 (N), PARPBP (O), TACC3 (P), and TROAP (Q).

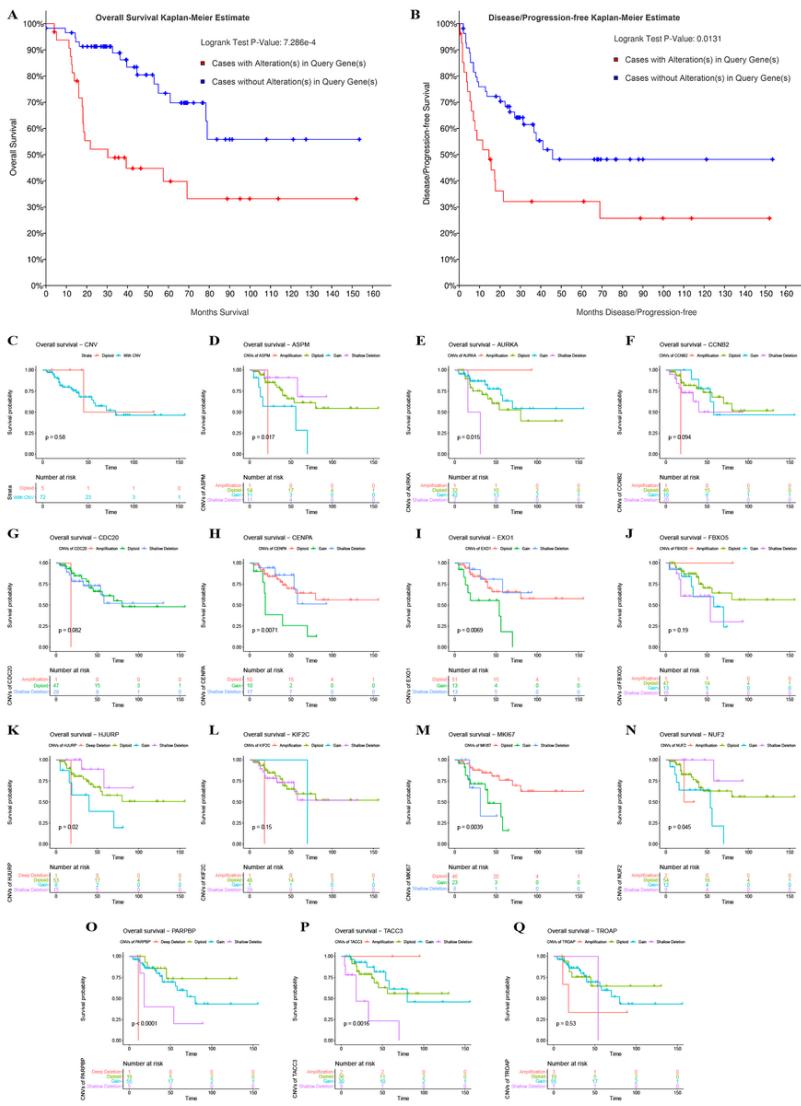


Figure 13

Survival analysis of ACC patients with CNVs of genes in model 2 based on TCGA ACC data. Overall survival (A) and Disease-free survival (B) of patients with alterations in genes in model 2 or patients without alterations in genes in model 2. (C) Overall survival of patients with any CNVs of genes in model 2 or patients with diploid. Overall survival of patients with different CNV types of ASPM (D), AURKA (E), CCNB2 (F), CDC20 (G), CENPA (H), EXO1 (I), FBXO5 (J), HJURP (K), KIF2C (L), MKI67 (M), NUF2 (N), PARPBP (O), TACC3 (P), and TROAP (Q).

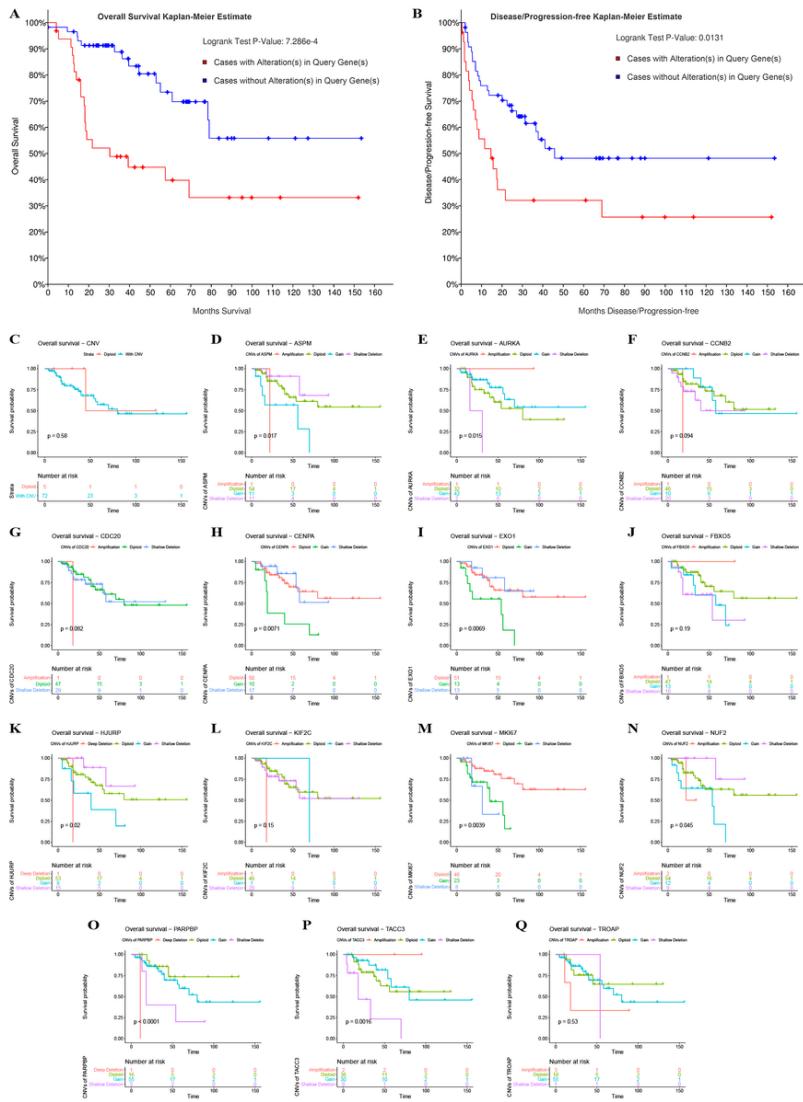


Figure 13

Survival analysis of ACC patients with CNVs of genes in model 2 based on TCGA ACC data. Overall survival (A) and Disease-free survival (B) of patients with alterations in genes in model 2 or patients without alterations in genes in model 2. (C) Overall survival of patients with any CNVs of genes in model 2 or patients with diploid. Overall survival of patients with different CNV types of ASPM (D), AURKA (E), CCNB2 (F), CDC20 (G), CENPA (H), EXO1 (I), FBXO5 (J), HJURP (K), KIF2C (L), MKI67 (M), NUF2 (N), PARPBP (O), TACC3 (P), and TROAP (Q).

Supplementary Files

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

- Supplementarymaterials.doc
 - Supplementarymaterials.doc
 - Supplementarymaterials.doc