

# An interactive nomogram based on clinical and molecular signatures to predict prognosis in multiple myeloma patients

Linxin Liu<sup>1,\*</sup>, Jian Qu<sup>2,\*</sup>, Yuxin Dai<sup>3</sup>, Tingting Qi<sup>2</sup>, Xinqi Teng<sup>2</sup>, Guohua Li<sup>2</sup>, Qiang Qu<sup>4,5</sup>

<sup>1</sup>Department of Hematology, Xiangya Hospital, Central South University, Changsha, China

<sup>2</sup>Department of Pharmacy, The Second Xiangya Hospital, Central South University, Institute of Clinical Pharmacy, Central South University, Changsha, China

<sup>3</sup>Department of Biochemistry and Molecular Biology, School of Life Sciences, Central South University, Changsha, China

<sup>4</sup>Department of Pharmacy, Xiangya Hospital, Central South University, Changsha, China

<sup>5</sup>Institute for Rational and Safe Medication Practices, National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

\*Equal contribution

**Correspondence to:** Qiang Qu; email: [quqiang@csu.edu.cn](mailto:quqiang@csu.edu.cn)

**Keywords:** multiple myeloma, WGCNA, LASSO, overall survival, prognosis

**Received:** January 30, 2021

**Accepted:** June 23, 2021

**Published:** July 14, 2021

**Copyright:** © 2021 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Although novel drugs and treatments have been developed and improved, multiple myeloma (MM) is still recurrent and difficult to cure. In the present study, the magenta module containing 400 hub genes was determined from the training dataset of GSE24080 through weighted gene co-expression network analysis (WGCNA). Then, using the least absolute shrinkage and selection operator (Lasso) analysis, a fifteen-gene signature was firstly selected and the predictive performance for overall survival (OS) was favorable, which was identified by Receiver Operating Characteristic (ROC) curves. The risk score model was constructed based on survival-associated fifteen genes from the Lasso model, which classified MM patients into high-risk and low-risk groups. Areas under the curve (AUC) of ROC curve and log-rank test showed that the high-risk group was correlated to the dismal survival outcome of MM patients, which was also identified in testing dataset of GSE9782. The calibration plot, the AUC value of the ROC curve and Concordance-index showed that the interactive nomogram with risk score could favorably predict the probability of multi-year OS of MM patients. Therefore, it may help clinicians make a precise therapeutic decision based on the easy-to-use tool of the nomogram.

## INTRODUCTION

Multiple myeloma (MM) is an aggressive neoplastic disease characterized by a collection of several diverse cytogenetically distinct plasma cell malignancies with a high degree of heterogeneity, accounting for about 10% of all hematological malignancies [1, 2]. Trisomies and Immunoglobulin heavy chain gene (IgH) translocations are considered primary cytogenetic abnormalities, and other cytogenetic

changes, such as gain(1q), del(1p), del(17p), del(13), RAS mutations, and translocations involving MYC, termed secondary cytogenetic abnormalities, also reflect high risk towards MM [3]. Overall survival of MM has significantly improved in the last 15 years, which has moved to the forefront of clinical interest because of the significant advances in medical treatment [4]. MM malignant cells and stromal cells secrete cytokines and growth factors, which explain the biological and clinical manifestations of

the disease, including hypercalcemia, renal failure, anemia, bone destruction, and infection [5].

Immunomodulatory drugs such as lenalidomide and pomalidomide, proteasome inhibitors such as bortezomib, autologous stem cell transplantation, and monoclonal antibodies have significantly improved the survival rate of MM patients [6, 7], but recurrence and resistance are still a major problem [8]. The staging and classification of MM have always been the key to MM individualized precision therapy. In clinical practice, even if the International Staging System (ISS) or Durian Salmon Staging (DSS) are similar, the prognosis of MM patients is still different. The differences between these patients may result from a lot of reasons, of which difference in molecular variation may be one of the most important reasons [9]. Thus, a better understanding of the molecular pathogenesis of MM could help us to identify new prognostic and therapeutic targets. With the extensive development, validation, and clinical applications of molecular techniques such as fluorescence *in situ* hybridization and next-generation sequencing, several prognostic and predictive biomarkers have been used to predict progression-free survival, overall survival (OS), and treatment response.

Weighted gene co-expression network analysis (WGCNA), a system biology algorithm, is used to find the correlation patterns among genes across microarray samples, identify modules of high related genes, and relate modules to certain clinical phenotypes [10, 11]. WGCNA is widely used to facilitate the screening or identification of candidate biomarkers or therapeutic targets that are critically associated with clinical traits [12]. The combined analysis of a panel of biomarkers, rather than an individual signature, is the most promising approach that is powerful enough to change clinical management [13–15]. The least absolute shrinkage and selection operator (Lasso), one of the machine learning methods, which uses L1 penalty for penalizing the squared error loss function of the coefficients, is an advanced variable selection algorithm for multi-collinear data or high-dimensional data [16, 17].

In the present study, we combined WGCNA with Lasso regression to simplify the complexity of the network of genes and improve the predictive accuracy of genes to the OS of MM patients. We also constructed a risk score model and an interactive nomogram to predict the prognosis of MM, which may help the clinicians in the treatment of MM in the future. The workflow chart is summarized in Supplementary Figure 1.

## RESULTS

### Construction of WGCNA

The dendrogram of 554 samples was shown in Supplementary Figure 2, and four outlier samples (GSM592558, GSM592552, GSM592499, and GSM592597) were excluded using the `flashclust` function of R package “WGCNA”. Heat map of the sample-associated clinical trait was also displayed in the different levels of red color (Supplementary Figure 2). To construct the WGCNA network, we determined the soft threshold power ( $\beta=16$ ) to define the adjacency matrix based on the criterion of approximate scale-free topology and mean connectivity (Figure 1A). To figure out the interactions among these thirteen co-expressed modules, the connectivity of eigengenes was analyzed. To construct the co-expressed network and identify the key module, a hierarchical clustering tree for the module was produced using the best-fit  $\beta$ -value ( $\beta=16$ ) (Figure 1B). Module blue and turquoise below the cut-off line at 0.25 were merged, resulting in twelve merged modules at last (Figure 1C).

### Identification of most significant modules and eigengenes

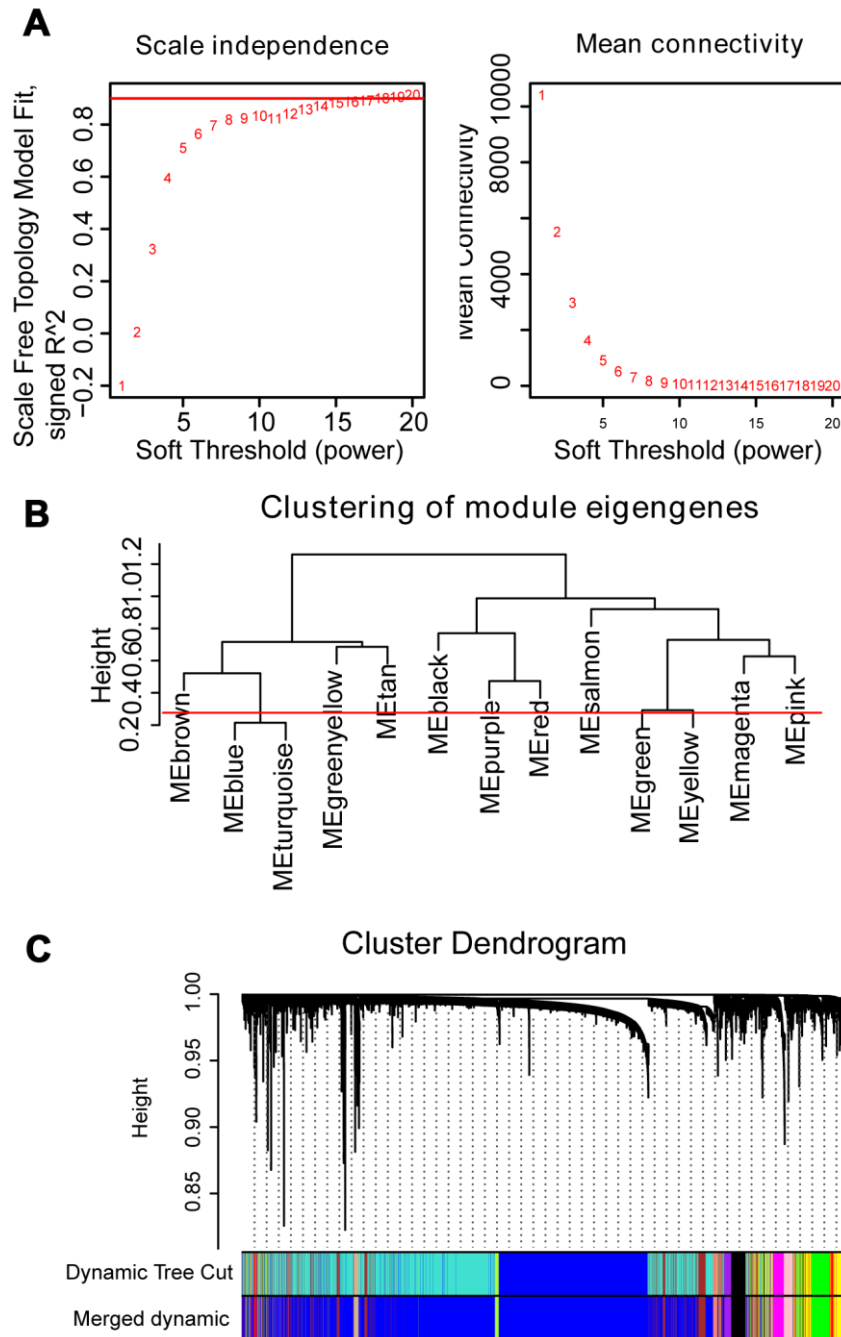
The correlation between module eigengenes and clinical traits of MM was showed in the heat map of Figure 2A. Module magenta was significantly related to event-free survival (EFS) ( $p=4e-08$ ) and OS ( $p=6e-09$ ). In addition, the module magenta had the highest GS relating to EFS ( $p$  value= $2.5e-130$ ) and OS ( $p$  value= $9e-251$ ) in boxplot (Supplementary Figure 3). Therefore, the magenta module containing 400 eigengenes was used for subsequent analysis. As a result, the relationships between module membership and gene significance of 400 hub eigengenes from the magenta module for OS and EFS were significant, implying that hub genes in the scatter plot also tend to be highly correlated with OS and EFS (Figure 2B, 2C).

### Lasso penalized Cox regression

After the Lasso was performed using 400 hub eigengenes using the training dataset, the coefficients were shown in Figure 3A. Genes with non-zero coefficients were considered to have strong prognostic potential in the Lasso penalized regression model. Two tuning parameters ( $\text{Log } \lambda_{\text{min}} = -2.86$  and  $\text{Log } \lambda_{\text{lse}} = -1.96$ ) were obtained using 10-fold cross-validation via minimum criteria (Figure 3B). Therefore, the fifteen-gene group and one-gene group were obtained based on  $\text{Log } \lambda_{\text{min}}$  and  $\text{Log } \lambda_{\text{lse}}$ . Wilcoxon test showed that the fifteen-gene group had better survival prediction potential than the one-gene group in MM patients

( $p < 0.05$ ) (Figure 3C). The ROC curves showed that the AUCs of the fifteen-gene group and the one-gene group were 0.756 and 0.69, respectively, indicating that the fifteen-gene group had a more predictive ability of prognosis (Figure 3D).

The testing dataset and whole samples of GSE24080 were used to validate the constructed fifteen-gene and one-gene prognosis models respectively. Wilcoxon test showed that the fifteen-gene prognostic model had better survival prediction potential than the one-gene

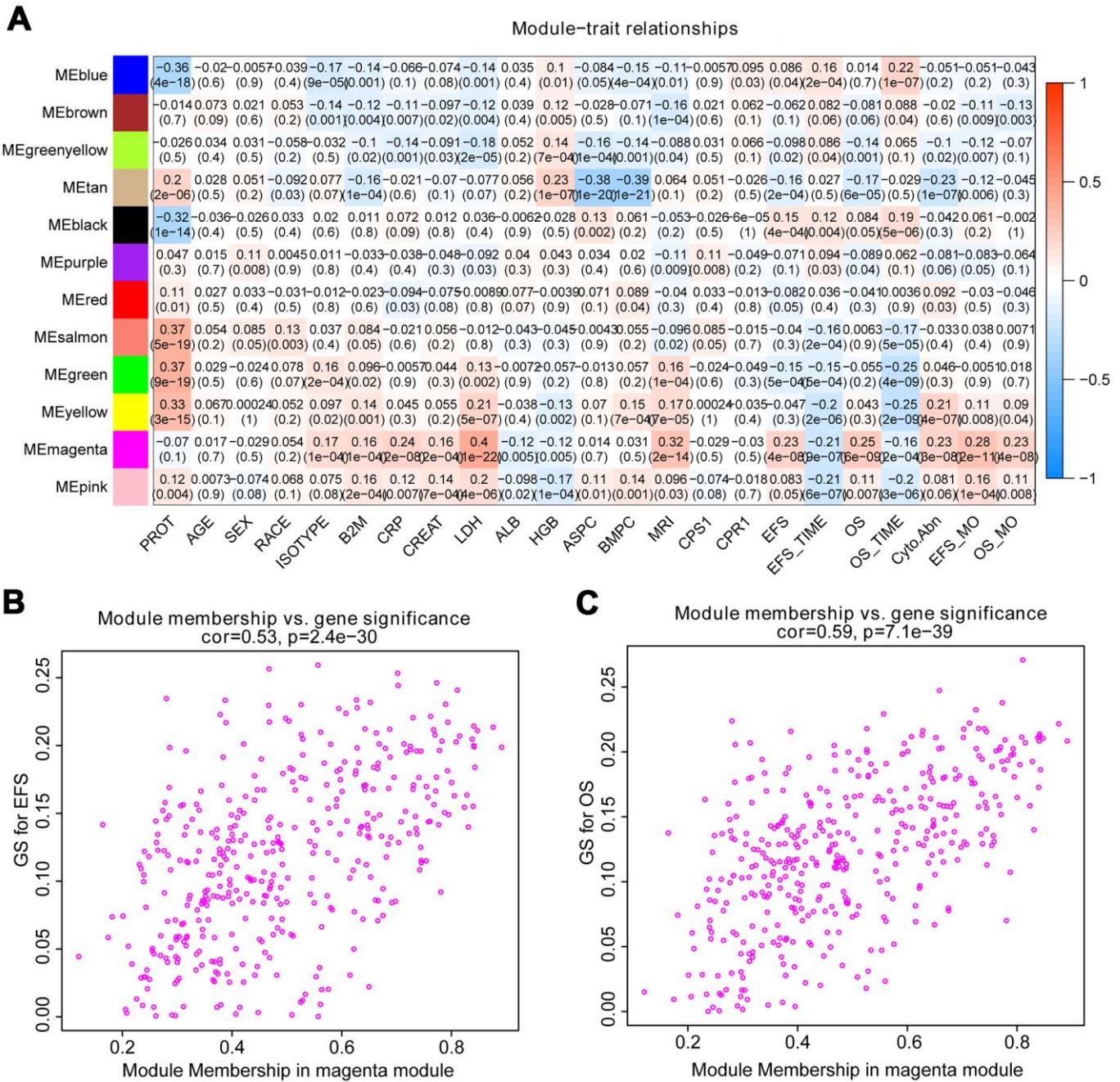


**Figure 1. WGCNA constructs a gene network related to the status of MM.** (A) Analysis of the scale-free fit index and mean connectivity for various soft-thresholding powers. The red line indicates the appropriate scale-free topology fit index at 0.9. The best  $\beta$  value was estimated at 0.9. (B) Dendrogram of consensus module eigengenes obtained by WGCNA on the consensus correlation. The red line at 0.25 indicates the merge threshold; groups of eigengenes below the threshold represent modules whose expression profiles were merged owing to their similarity. (C) Merged modules were identified by the Dynamic Tree Cutting method of WGCNA. Each module is assigned a color as an identifier. According to the correlation between the modules, twelve modules are generated after the merge.

group in both datasets ( $p < 0.05$ ) (Figure 4A, 4B). The AUCs of the fifteen-gene group and the one-gene group were 0.65 and 0.62 in the testing dataset, respectively, and were 0.71 and 0.66 in the whole sample dataset of GSE24080 (Figure 4C, 4D). As the fifteen-gene model has more accurate prognostic power than the one-gene model, the following studies focused on the fifteen-gene prognostic model.

### Univariate and multivariate Cox regression for fifteen hub genes

To obtain a more accurate and sensitive risk score model, univariate and multivariate Cox regression analyses for fifteen genes were performed. The univariate Cox regression showed that 14 genes could be independent biomarkers for the prediction of OS



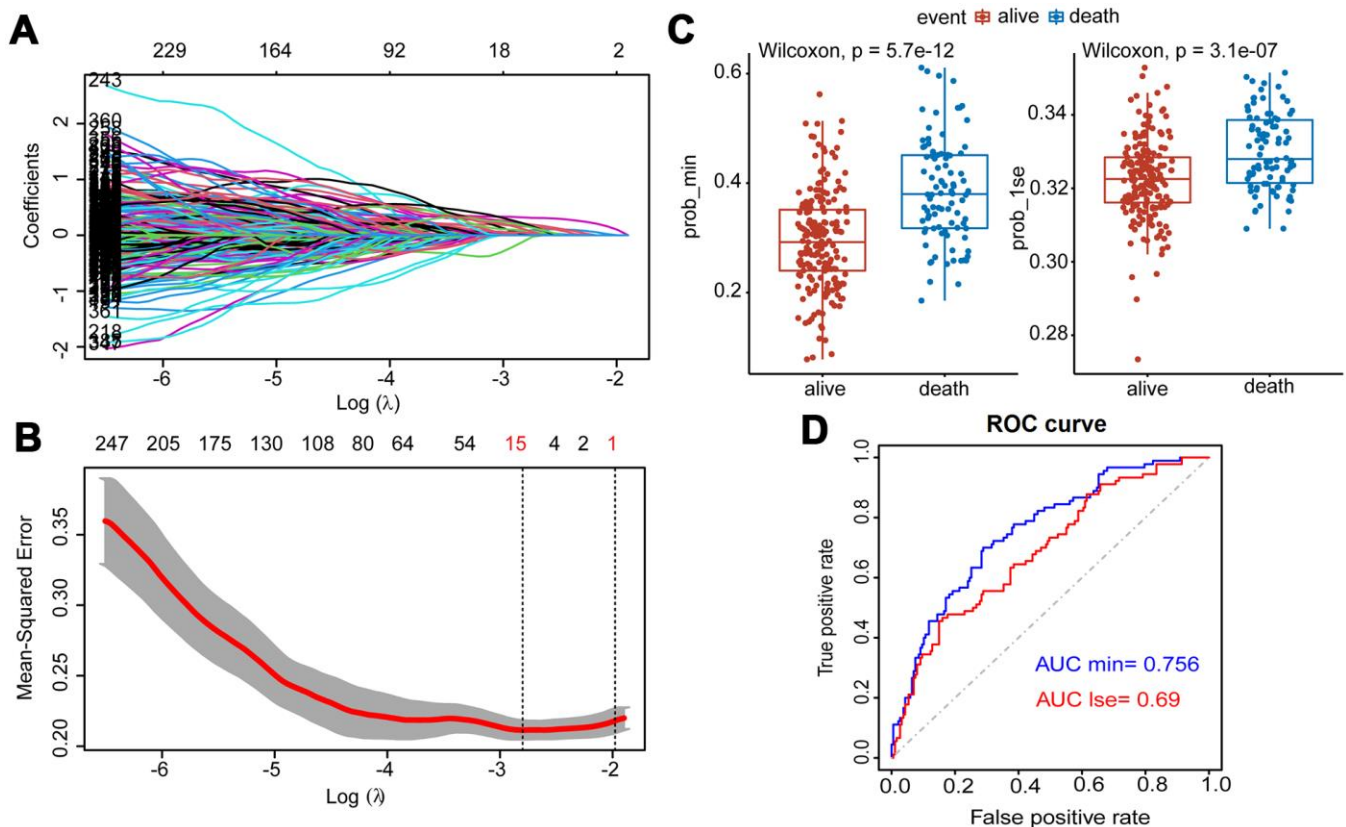
**Figure 2. Identify modules related to the clinical features of MM.** (A) Heat map of the correlation between module eigengenes and clinical traits of MM. Red means high adjacency and blue means low adjacency. The correlation coefficient and p-value were listed in the heat map. (B, C) Scatter plot of module eigengenes related to EFS and OS in the magenta module.

( $p < 0.05$ ) (Table 1). Three genes, *AURKA*, *FAM72A*, and *NUF2*, were probably more closely correlated to OS after the multivariate Cox regression performing in the forest plot ( $p < 0.1$ ) (Figure 5A and Table 1). Kaplan-Meier survival analysis also showed that three genes contributed to independent survival prediction (Log-rank  $p < 0.05$ ) (Figure 5B).

### Establishment of a fifteen-gene risk score model

A fifteen-gene risk score model was built by multivariate Cox hazard regression analysis. The Concordance-index value was 0.69 and the global log-rank p-value was  $7.764 \times 10^{-11}$  ( $p < 0.05$ ) in the forest plot (Figure 5A), indicating that the risk score model containing fifteen hub genes could be used to predict OS. MM patients with a risk score  $\geq 0$  were classified into the high-risk group, whereas MM patients with a risk score  $< 0$  were divided into the low-risk group. Compared to the low-risk group, the distribution plot of survival status showed that the samples of death

events were enriched in the high-risk group (Figure 6A, 6B). All genes were overexpressed in the high-risk group in the heat map (Figure 6C). Consistently, the Kaplan-Meier curve showed that the high-risk group had lower survival possibility, compared to the low-risk group (Log-rank  $p < 0.0001$ ) (Figure 6D). Time-dependent ROC curves further verified and depicted the excellent prediction performance of the fifteen-gene risk score model with the AUC of 1-year was 0.692, 3-year AUC = 0.712, 5-year AUC = 0.676, 7-year AUC = 0.757 (Figure 6E). The mean of time-dependent AUC was over 0.7 during the different time, which imply that the risk score model of fifteen genes may contribute to accurate survival prediction (Figure 6F). Six hub genes were selected from the testing dataset of GSE9782 and used to construct a risk score model in Supplementary Table 1. The similar results showed that this risk score model was also beneficial for the prediction of survival based on log-rank test  $p < 0.0001$  and AUC value was 0.638.

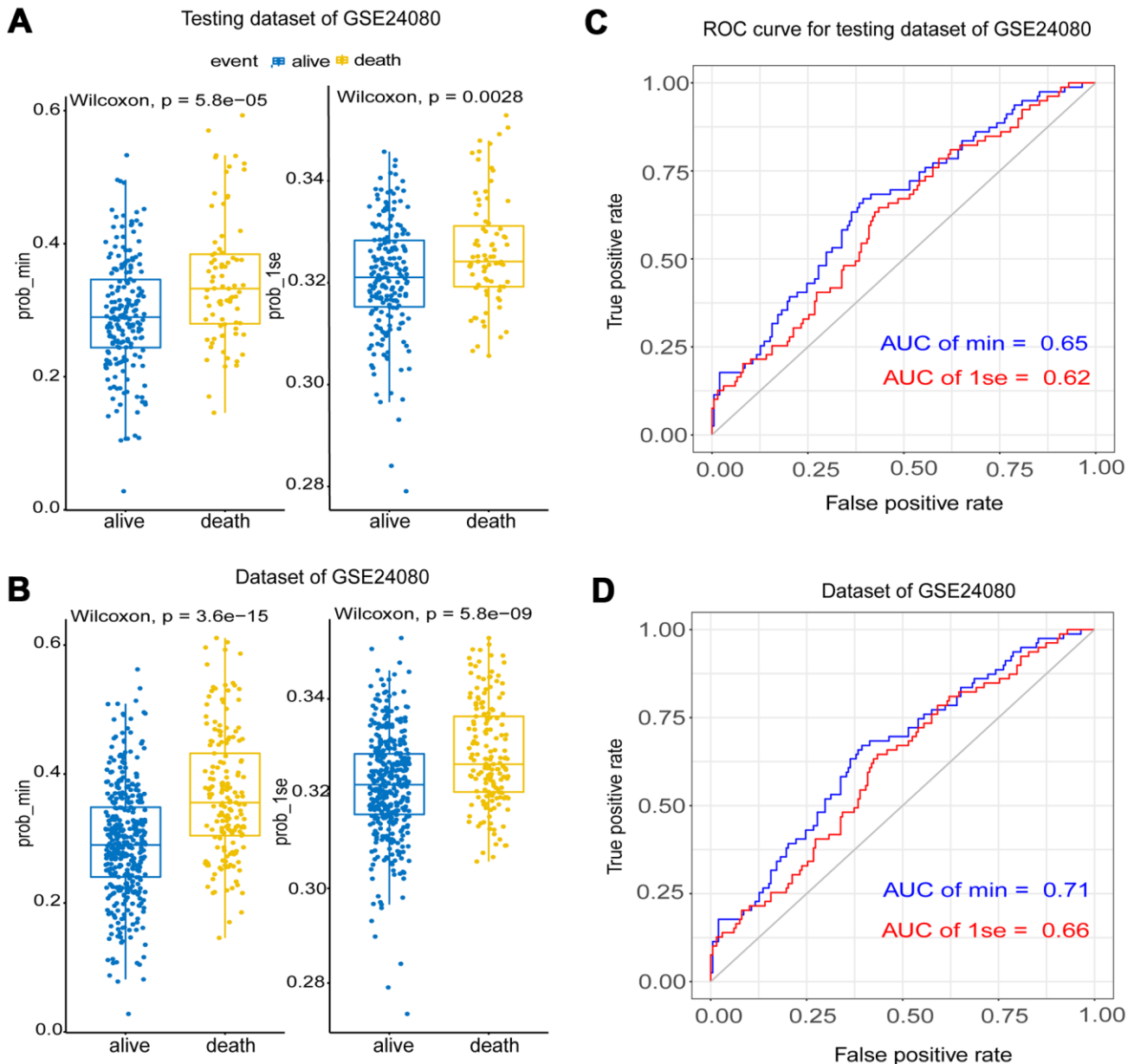


**Figure 3. Construction of Lasso Cox regression model using training dataset of MM.** (A) A coefficient profile plot was produced against the  $\log(\lambda)$  sequence. (B) Tuning parameter ( $\lambda$ ) selection in the Lasso model used 10-fold cross-validation via minimum criteria. Dotted vertical lines were drawn at the optimal values by using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria). (C) Survival probabilities (Prob-min and Prob-1se) were predicted by two Lasso models based on two ideal parameters  $\log \lambda_{1se}$  and  $\log \lambda_{min}$ . The Wilcoxon test was used to compare the different survival outcomes. (D) ROC curves analysis and the values of AUC were used to compare two Lasso models based on  $\log \lambda_{1se}$  and  $\log \lambda_{min}$ .

## Establishment and validation of the nomogram

Clinical traits and the fifteen-gene risk score were also analyzed by univariate and multivariate Cox regression to identify the survival-related variable. The risk score and clinical signatures including AGE, CREAT, LDH, ALB, Cyto\_Abn were found to be significantly related to OS ( $p < 0.05$ ) (Supplementary Table 2). Since five clinical traits and risk scores were vital variables for the prediction of OS of MM, we rebuilt the multivariate

Cox regression model and visualized it by nomogram. Total points were summated by adding each point of AGE, CREAT, LDH, ALB, Cyto\_Abn, and the fifteen-gene risk score (Figure 7A). The calibration curves of 1 to 7-year survival demonstrated favorable prediction performance of nomogram (Figure 7B). To illustrate the significant influence of the fifteen-gene risk score in the nomogram, the ROC curve further verified that AUC for the model containing the risk score was 0.785, higher than 0.730 without risk score (Figure 7C, 7D).



**Figure 4. The validation of the Lasso model using testing and the whole dataset of GSE24080.** (A, B) Survival probabilities (Probmin and Prob1se) were predicted by two Lasso models based on two ideal parameters  $\text{Log } \lambda_{1se}$  and  $\text{Log } \lambda_{min}$  using testing and the whole dataset of GSE24080. The Wilcoxon test was used to compare the different survival outcomes. (C, D) ROC curves analysis and the values of AUC were used to compare two Lasso models based on  $\text{Log } \lambda_{1se}$  and  $\text{Log } \lambda_{min}$  using testing and the whole dataset of GSE24080.

**Table 1. Univariate and multivariate Cox regression analysis for fifteen genes.**

Variable	Univariate Cox analysis			P	Multivariate Cox analysis			P
	Beta	HR	95% CI HR		Beta	HR	95% CI HR	
ABHD3	3.5	34	2.7-440	0.006*	1.205	3.336	0.157-70.765	0.440
ACN9	3.5	32	6.3-160	0.000**	1.461	4.310	0.754-24.640	0.101
ANLN	1.9	6.6	2.1-21	0.001**	-0.097	0.908	0.262-3.145	0.879
AURKA	5	150	31-710	0.000**	2.442	11.43	1.412-93.551	0.023*
CTNNA1	0.79	2.2	0.34-14	0.410	-0.931	0.394	0.059-2.628	0.336
FAM210A	6.5	660	24-18000	0.000**	-0.292	0.747	0.017-32.382	0.880
FAM72A	4.1	62	19-200	0.000**	2.301	9.983	1.282-77.727	0.028*
HAUS3	4.9	140	5.6-3400	0.003*	1.835	6.264	0.204-191.915	0.293
KNSTRN	7.3	150	54-39000	0.000**	2.991	19.99	0.433-915.781	0.126
LINC00998	5.3	200	16-2600	0.000**	1.084	2.958	0.159-54.955	0.467
LOC81691	1.3	3.7	1.9-7.5	0.000**	0.387	1.472	0.666-3.253	0.339
NUF2	2.3	10	4.1-26	0.000**	-1.105	0.331	0.099-1.108	0.073#
SMCHD1	5.1	160	10-2400	0.000**	2.310	10.08	0.421-240.952	0.154
TIMM21	6.6	710	41-12000	0.000**	1.788	5.980	0.154-232.174	0.338
ZNF92	6	380	27-5600	0.000**	1.026	2.791	0.148-52.590	0.493

#, p<0.1, \*, p<0.05, \*\*, p<0.001.

Meanwhile, the C-index of the model with the risk score was 0.754, higher than 0.722 of the C-index without the risk score (Table 2). The patient sample of GSM592833 was randomly selected and used to illustrate the nomogram (Figure 8A, 8B). It should be noted that the patient died on the following day of 2016. The death probability of this sample was 0.661 predicted by nomogram with a risk score, but only 0.445 predicted by the nomogram without risk score. These results indicate that the nomogram with risk score has better predictive performance thus is more reliable. According to the equation of nomogram, after inputting the values of clinical traits and risk score, the prediction of OS was estimated via nomogram tool in Microsoft Excel (Supplementary Table 3).

## DISCUSSION

Although several new drugs have been introduced into clinical practice, MM is still incurable with dismal clinical outcomes in MM patients [18, 19]. Molecular risk stratification basing on hub gene expression of MM has opened an avenue for clinicians to conduct personalized medicine [20, 21]. Therefore, it is critical to developing novel molecular biomarkers that are closely related to the clinicopathological characteristics and clinical survival outcomes of MM patients.

To reveal the relationship between clinical information and gene expression matrix, many bioinformatics methods are developed. A prognosis-associated long noncoding RNA-mRNA network for multiple myeloma was constructed by WGCNA based on the microarray

of MM (GSE24080) [22]. In the present study, we firstly comprehensively investigate the influence of genes on the prognosis of MM patients via WGCNA. The key module magenta was screened out according to the p-value and Pearson's coefficients in the module-trait heat map. The bar plots of gene significance also showed that the module magenta was highly correlated to EFS and OS (p<0.05). Dot plot of gene significance and module membership revealed that 400 hub genes within the magenta module were significant.

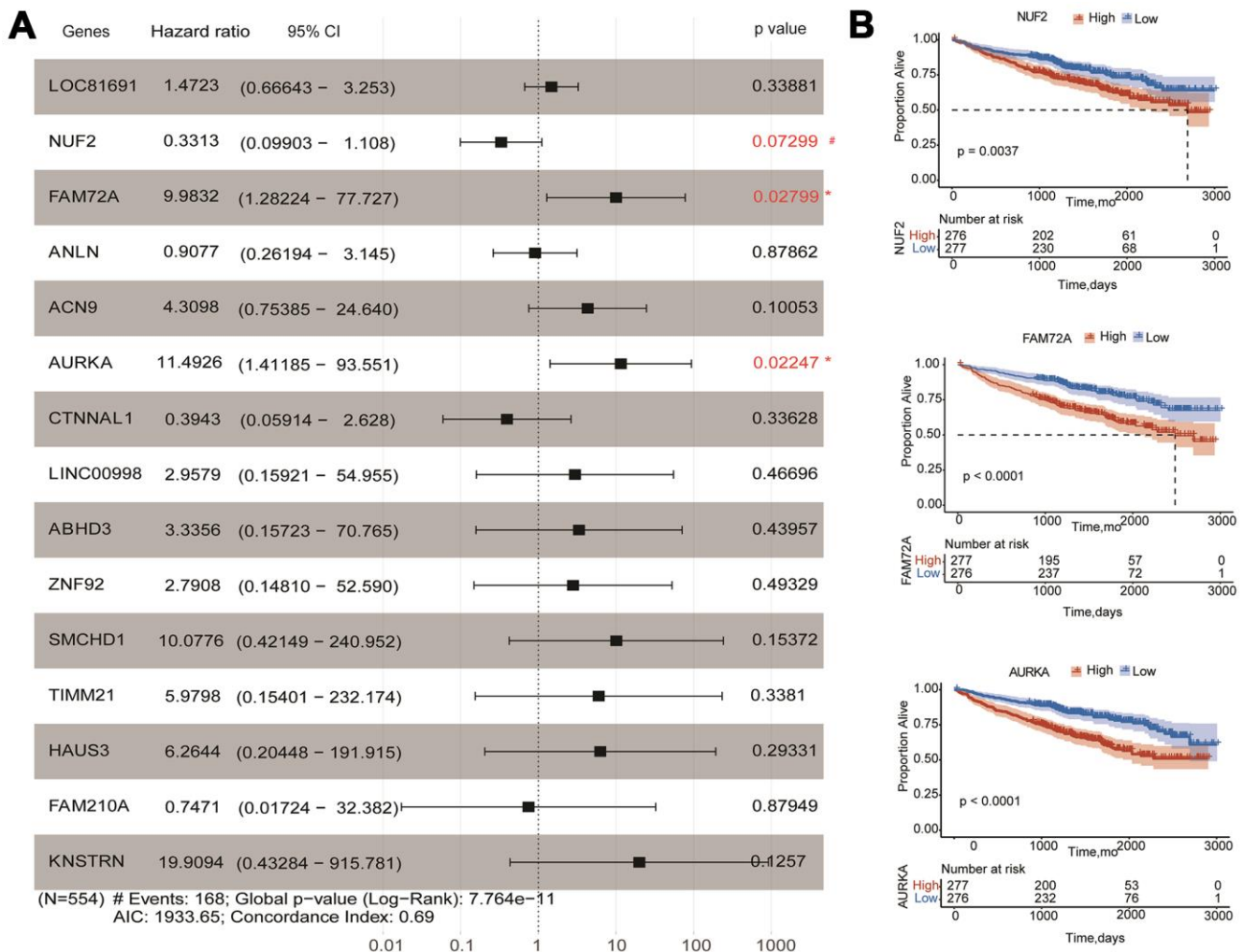
To further screen hub genes involved in the prognosis of MM, Lasso was introduced to improve the prediction accuracy by forcing the sum of the absolute value of the regression coefficients to be less than a fixed value. We incorporated the 400 hub genes into the Lasso model, fifteen hub genes with non-zero coefficients in the model were identified based on log  $\lambda_{\min}$ . The ROC curves indicated that the fifteen-gene model showed better performance of prediction ability compared to the one-gene model based on log  $\lambda_{\text{lse}}$ . Multivariate Cox regression analyzed the predicted attribution of individual genes within the fifteen-gene model and three genes (*AURKA*, *NUF2*, and *FAM72A*) were found to be significantly correlated to survival outcome (p<0.1) the same as independent indicators using the K-M curve (Log-rank p<0.01). Considering the interaction and influence among hub genes, we constructed the risk score model to weigh the efficacy of fifteen genes on OS. The low-risk score group had a better survival outcome compared to the high-risk score group. The classifier based on risk score was well performed to distinguish the status of OS by time ROC curve. The

risk score model containing six hub genes was constructed using the testing dataset of GSE9782, which also showed favorable prediction ability.

Univariate and multivariate Cox regression analyses were also used to identify favorable clinical risk characteristics, at last, AGE, LDH, CREAT, Cyo\_Abn, and ALB were selected due to the high correlation to prognosis ( $p < 0.05$ ). Compared to the previously published nomogram, it is worth mentioning that we considered gene risk score as a continuous variable which was more accurate to predict prognosis. This research not only considered one aspect of gene expression, but the clinical characters and risk gene combination, eventually integrated into the nomogram, intuitively reflect effects of each part on the prognosis

of patients with MM, and ultimately to the scores, for 1 to 7 years survival rate of prediction. Thus, an interactive nomogram was established, taking into account a variety of survival-related five clinical traits and the fifteen-gene risk score. ROC curve analysis and C-index showed that the nomogram with risk score had higher reliability than the model without a risk score. Meanwhile, the nomogram model without the fifteen-gene risk score was also validated in the same sample of GSM592833, but with less accuracy.

Compared to the R-ISS system, LDH and cyto\_abnormality were also considered as indicators in the nomogram. The continuous values of LDH were transferred to points of the nomogram, which was more accurate for the prediction of the survival of each



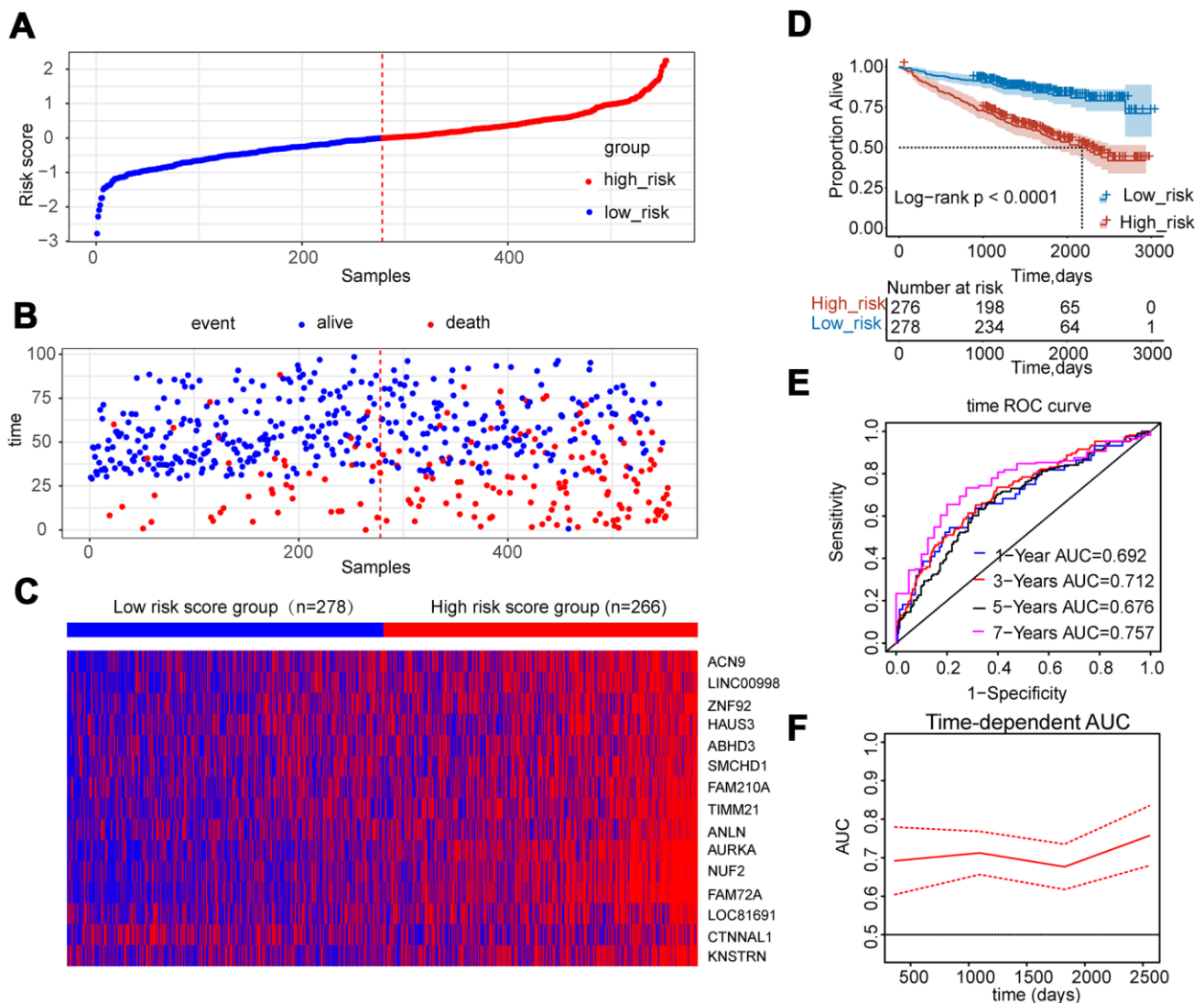
**Figure 5. Effects of hub genes on survival outcome of MM patients of GSE24080. (A)** Forest plot of multivariate Cox regression analysis for fifteen genes. FAM72A, AURKA, and NUF2 were significantly associated with survival in multiple myeloma patients ( $p < 0.1$ ). **(B)** Kaplan-Meier analysis of FAM72A, AURKA and NUF2 to predict patient survival. MM patients with the high and low expression levels of three genes were divided into two groups according to the cut-off value of medium expression levels. p-value of the Log-rank test less than 0.05 was considered as a statistical difference.



patient. Another prognostic model containing the ISS system, the expressions of two miRNA (let-7b and miR-18a) in serum and cytogenetics, could improve the identification of patients with newly diagnosed MM with poor outcomes [23]. The AUC was near 0.73 lower than the AUC value of 0.785 in our nomogram model. Other models using Gaussian process regression and random forests model also illustrated that the AUC value was near 0.70 lower than the AUC value of 0.785 in our nomogram model [24].

The molecular signature of these hub genes has been demonstrated to be associated with the proliferation or

invasion of several human cancers. *AURKA* encodes a serine/threonine kinase located in the centrosomes and plays a vital role in the distribution of chromosomes to two daughter cells in mitosis by participating in the replication, separation, and maturation of centrosomes [25, 26]. Studies have shown that overexpression of *AURKA* leads to chromosome instability and promotes malignant transformation of cells [27]. *AURKA* phosphorylates p53 serine 315 residues, promoting MDM2-mediated degradation, while *AURKA* silencing reverses this process. Besides, reduced *AURKA* levels led to greater stability, while increased *AURKA* expression undermined the response to cisplatin-

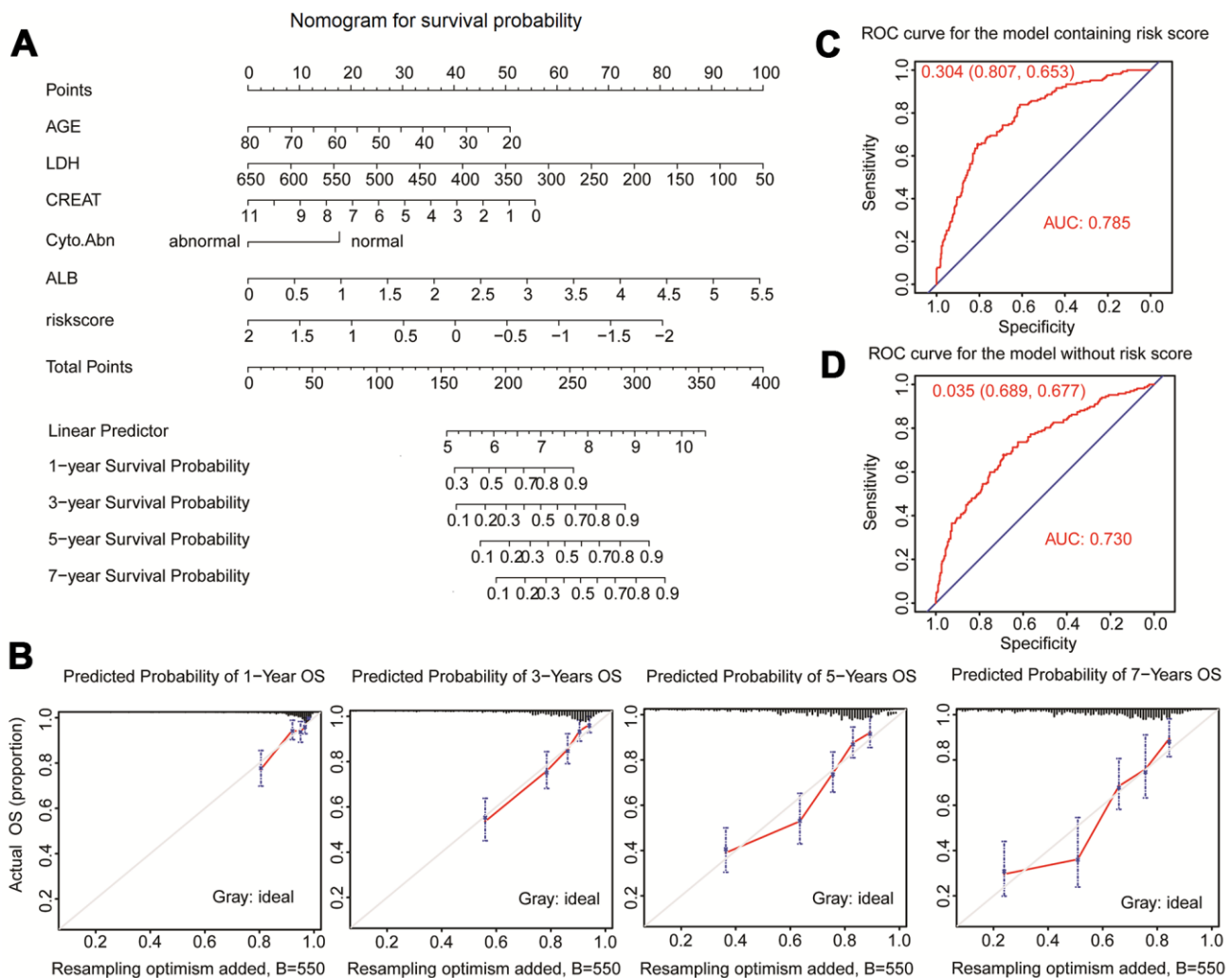


**Figure 6. Construction of risk score model using the fifteen-gene expression profile in MM patients of GSE24080.** (A) Fifteen-gene risk score distribution of MM patients based on risk score levels. The best cut-off value was used to divide the patients into two groups. (B) The survival status of all patients was distributed and classified by risk score group. (C) The expression profiles of the fifteen genes in high-risk and low-risk groups. (D) Kaplan-Meier analysis of the fifteen-gene risk score model. (E) Time-dependent ROC curves for the fifteen-gene risk score model to predict overall survival of MM patients. AUCs for 1 to 7-year survival were shown in the figure. (F) Time-dependent AUCs for the risk assessment model.

induced apoptosis [28, 29]. AURKA protected ovarian cancer cells from chemotherapeutic drug-induced apoptosis by activating the Akt pathway in a p53-dependent manner [30]. There is increasing evidence that AURKA is expressed to varying degrees in newly diagnosed and recurrent MM patients and MM cell lines [30, 31]. AURKA has been reported to be associated with myeloma resistance and early disease recurrence [32]. AURKA may disrupt the DNA damage repair response by regulating DNA repair proteins such as CHK1/2 [33]. Inhibition of AURKA expression in MM cells induced apoptosis and death [31, 34, 35]. Some small molecule inhibitors against AURKA are currently being studied in clinical trials in MM or other cancer patients [36, 37]. In the present study, we also showed

that AURKA could be an independent indicator for the prediction of the survival outcomes of MM patients.

NUF2, also named CDCA1 as a part of a protein complex associated with the centromeres, is essential for normal centromere microtubule attachment [38] and chromosome instability in tumor cells [39]. Abnormal expression of NUF2 leads to mitotic dysregulation [40, 41]. The dysexpression of NUF2 is closely related to the development of tumors and it can be used as a biomarker for poor prognosis [42]. Small interfering RNA inhibited NUF2 expression in pancreatic cancer, glioma, and liver cancer and reduced the growth of tumors [43, 44]. Besides, the sub-G1 proportion in the cell cycle was significantly increased [44, 45]. The



**Figure 7. Nomogram predicting 1-year, 3-year, 5-year, and 7-year overall survival of multiple myeloma patients. (A)** The nomogram consists of the fifteen-gene risk score and five clinical risk indicators. Add the points from these 6 variables together and find the location of the Total Points. The Total Points projected on the bottom scales indicate the probability of 1-year, 3-year, 5-year, and 7-year overall survival. **(B)** The calibration curve for predicting 1-year, 3-year, 5-year, and 7-year overall survival. **(C)** ROC curves for the nomogram model with risk score to predict patient survival. **(D)** ROC curves for the nomogram model without risk score to predict patient survival.

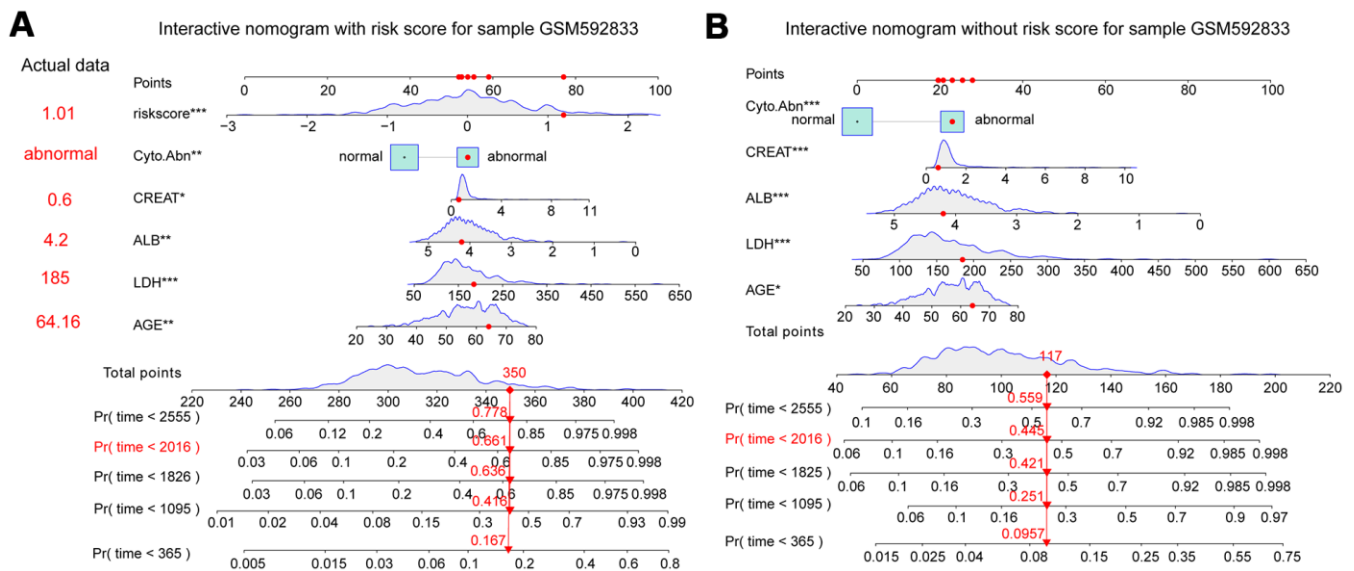
**Table 2. Nomogram model parameters.**

Variable	Nomogram with risk score			p	Nomogram without risk score			p
	Beta	OR	95%CI		Beta	OR	95%CI	
AGE	0.024	1.024	1.007-1.042	0.005*	0.021	1.021	1.004-1.039	0.013543*
LDH	0.004	1.004	1.002-1.006	0.000**	0.005	1.005	1.004-1.007	6.08e-09**
ALB	-0.334	0.716	0.572-0.897	0.004*	-0.450	0.637	0.512-0.793	5.32e-05**
CREAT	0.101	1.106	1.013-1.208	0.025*	0.146	1.157	1.062-1.261	0.000895**
Cyto-Abn	-0.511	0.600	0.436-0.826	0.001*	-0.699	0.497	0.365-0.677	9.39e-06**
risk score	0.647	1.910	1.501-2.430	0.000000145**	NA	NA	NA	NA
C-index	0.754±0.019				0.722±0.02			
Likelihood ratio test	122.6, p<0.001				94.85, p<2e-16			
Wald test	138.6, p<0.001				111.5, p<2e-16			
Score (Log-rank test)	150, p<0.001				119.3, p<2e-16			

\*, p<0.05, \*\*, p<0.001.

prognostic impact of NUF2 correlates well with the dismal prognosis of MM [46]. After the multivariate Cox regression analysis was performed, it appears that lower expression of NUF2 was associated with worse outcomes in the forest plot, but the K-M curve showed the improved outcome for low expressions, which was different may due to the differences in statistical methods and the influence by gene-gene interaction. It implied that our interactive nomogram with risk scores could more favorably predict the probability of multi-year OS of MM patients than single gene expression.

FAM72A (family with sequence similarity 72 member A), known as p17 or LMP1P, is composed of 149 amino acids as a kind of neuron protein [47]. Under physiological conditions, FAM72A is expressed at low levels; but the overexpression can cause neuronal cell death. This protein has been shown to have high clinical relevance for survival/death outcomes in cancer patients, as it may be associated with tumorigenic effects in non-neuronal tissue including breast, colorectal, and lung cancer [48–50]. Preliminary data show that FAM72A acts downstream of PKC as a uracil



**Figure 8. Interactive nomogram predicts the survival probability of overall survival for the GSM592833 sample.** The patient died on the following day of 2016. \*\* (p<0.01), \*\*\* (p<0.001) indicates a significant correlation. (A) Nomogram containing five clinical traits and fifteen-gene risk score predicts 365-day, 1095-day, 1825-day, 2016-day, and 2555-day overall survival probability. (B) A nomogram containing five clinical traits without a fifteen-gene risk score predicts overall survival probability.

DNA glycosylase-2 (UNG2) binding protein and mediates tumorigenic action [47]. Cell cycle analysis showed that FAM72A drove cells into the G0/G1 phase, which may explain the physiological low expression level of FAM72A in proliferating cells and its role in neurons [49]. Tumor cells may attempt to compensate for the inhibition of p53 and the interference of base excision repair (BER) mechanism, so it may be characterized by the increased expression level of FAM72A protein [48]. This was consistent with previous cell cycle analysis by Wang et al. which showed that FAM72A shortened G1/S transition in nasopharyngeal carcinoma [51].

Other hub genes were also reported in various types of cancers as independent indicators for survival prediction. Rare homozygotes in the *ANLN* (rs12535394) SNP pair are prognostic of favorable breast cancer survival [52]. High levels of ANLN contributed to the poor prognosis of anthracycline-based chemotherapy in breast cancer patients [53, 54]. A three protein-coding genes prognostic model including *SMCHD1* predicts overall survival in bladder cancer patients [55]. Univariate and multivariate survival analysis showed that high expression of *HAUS3* was an independent prognostic factor for the dismal outcome of overall survival of HCC patients [55]. High expression of HAUS3 is also associated with poor prognosis and HCC progression [56]. KNSTRN is significantly associated with dismal survival status in endometrial cancer [57].

In summary, this study used a variety of analytical methods to establish a scoring system as a prognostic indicator for MM patients. The prognostic model was an independent classifier for MM. Moreover, several hub genes in the model could be utilized as an effective drug target for MM treatment.

## MATERIALS AND METHODS

### Data collection and preprocessing

The microarray expression data of GSE24080 as the training dataset and GSE9782 as the testing dataset of MM patients were obtained from the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo>). According to the annotation information, the probes are converted into gene symbols on the array platform, and the gene expression is calculated by R package "limma". All samples with complete follow-up clinical information are listed in Supplementary Figure 4.

### Construction of WGCNA

The details of the WGCNA algorithm were conducted by the R package "WGCNA" and were described in

our previous publications [58–60]. Firstly, we used the sample dendrogram clustering method to remove the obvious outlier samples with too many missing entries for outlier detection and showed in a clinical trait heat map. Secondly, based on the criteria of approximate scale-free topology and mean connectivity, the soft-threshold power was calculated by the `pickSoftThreshold` function of WGCNA. Thirdly, we calculated adjacencies using best-fit soft-threshold power, transformed the adjacency into a topological overlap matrix (TOM), and calculated the corresponding dissimilarity TOM (dissTOM). Finally, after hierarchical clustering analysis based on the dissTOM, modules were generated by the dynamic tree cut method for module merging (the cut-off line at 0.25).

### Construction of module-trait relationships and identification of hub modules associated with clinical traits

Pearson's correlation coefficients and p values between modules and MM clinical traits were calculated and visualized in a module-trait heat map. The module membership represents the correlation between the clinical traits and MM's clinical status. The average gene significance (GS) in each module was displayed in a bar graph, which also reflected the relationship between the module and MM clinical status. The module with the highest module significance and module membership was considered as hub clinical module. To further identify hub eigengenes within key modules, a scatter plot of GS related to module membership for clinical traits was visualized. The p values and Pearson's coefficients were also calculated.

### Lasso penalized regression analysis

The expression matrix of eigengenes within the key module was divided into the training and testing datasets of GSE24080 by the R package of "caret". Lasso penalized regression was used to select the most powerful predictive eigengenes from the training dataset. By performing L1 norm regularization through the R package "glmnet", different coefficients corresponding to different  $\lambda$  values were obtained. The "cv.glmnet" function was used to select the optimal  $\lambda$  value by ten-fold cross-validation. Gene screening using  $\text{Log } \lambda_{\text{min}}$  or  $\text{Log } \lambda_{1\text{se}}$  was conducted by "coef" function. Comparison analysis of survival probability of the two ideal models was conducted by the Wilcoxon test and ROC curves through the R package of "pROC" using the training dataset, testing dataset, and the whole expression matrix of GSE24080.

## Univariate and multivariate Cox regression analysis

To further identify the influence of hub genes and clinical traits on the survival outcome of MM, univariate and multivariate Cox regression analyses were conducted using R packages “survival” and “survminer”. A p-value less than 0.05 was considered statistical significance. Kaplan-Meier (K-M) curve and Log-rank test were also conducted to compare survival outcomes between high-expression and low-expression groups by R packages “survival” and “survminer”. Log-rank  $p < 0.05$  was considered statistical significance.

## Establishment of risk score model based on fifteen-gene signature

The risk score for MM patients was calculated and evaluated by Concordance-index (C-index) based on the Cox proportional-hazards model [21, 58]. Then all patients were divided into a high-risk group and a low-risk group according to the cut-off value of the risk score. The risk score distribution, survival status, and expression pattern of the fifteen hub genes were displayed in the dot plot and heat map. Kaplan-Meier analysis was used to compare survival outcomes between the high-risk group and low-risk group. Log-rank  $p < 0.05$  was considered as statistical significance. Time-dependent ROC analysis was performed and the time-dependent area under the curve (AUC) was calculated by the R package “timeROC”. The testing dataset of GSE9782 was used to testify the performance of the risk score model following the above process.

## Nomogram

Five indicators including AGE, lactate dehydrogenase (LDH), creatinine (CREAT), cytogenetic abnormalities (Cyto\_Abn), and the fifteen-gene risk score were enrolled to establish the nomogram by the R package “rms”. The interactive nomogram was calculated by the function of R package “glm” and showed using “regplot”. Each indicator had its point, and the total points were calculated by adding all the points together. The total points of these indicators and survival probability were calculated by the R package “nomogramEx”. C-index was measured to evaluate the discrimination performance of nomogram with or without risk score by R package “Hmisc”. The calibration curve and ROC curve were also plotted to evaluate the prediction efficacy of the nomogram with or without risk score. The nomogram was validated using a randomly selected sample of GSM592833 to predict the 1-year, 3-year, 5-year, and 7-year OS. A predictive tool of a nomogram based on Microsoft Excel was developed for convenient clinical use. After the input of clinical information and the risk score of

MM patients, the survival probability is output (Supplementary Table 4).

## Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

## Abbreviations

MM: multiple myeloma; WGCNA: weighted gene co-expression network analysis; Lasso: least absolute shrinkage and selection operator; ROC: receiver operating characteristic; K-M: Kaplan-Meier; AUC: Areas under the curve; ISS: International Staging System; DSS: Durian Salmon Staging; TOM: topological overlap matrix; dissTOM: dissimilarity TOM; GS: gene significance; C-index: Concordance-index; LDH: lactate dehydrogenase; CREAT: creatinine; Cyto\_Abn: cytogenetic abnormalities; EFS: event-free survival; OS: overall survival; UNG2: uracil DNA glycosylase-2.

## AUTHOR CONTRIBUTIONS

QQ designed the study. LL, JQ, and YD analyzed the results and wrote the manuscript. TQ, XT, and GL revised the text of the article. All authors have read and agreed to the published version of the manuscript.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This work was supported by the National Natural Science Foundation of China (No. 82073944).

## REFERENCES

1. Bergsagel PL. Prognostic factors in multiple myeloma: it's in the genes. *Clin Cancer Res.* 2003; 9:533–34. PMID:[12576415](https://pubmed.ncbi.nlm.nih.gov/12576415/)
2. Bladé J, Rosiñol L, Cibeira MT. Prognostic factors for multiple myeloma in the era of novel agents. *Ann Oncol.* 2008 (Suppl 7); 19:vii117–20. <https://doi.org/10.1093/annonc/mdn437> PMID:[18790932](https://pubmed.ncbi.nlm.nih.gov/18790932/)
3. Rajkumar SV. Multiple myeloma: Every year a new standard? *Hematol Oncol.* 2019 (Suppl 1); 37:62–65. <https://doi.org/10.1002/hon.2586> PMID:[31187526](https://pubmed.ncbi.nlm.nih.gov/31187526/)

4. Ludwig H, Novis Durie S, Meckl A, Hinke A, Durie B. Multiple Myeloma Incidence and Mortality Around the Globe; Interrelations Between Health Access and Quality, Economic Resources, and Patient Empowerment. *Oncologist*. 2020; 25:e1406–13. <https://doi.org/10.1634/theoncologist.2020-0141> PMID:32335971
5. Bataille R, Harousseau JL. Multiple myeloma. *N Engl J Med*. 1997; 336:1657–64. <https://doi.org/10.1056/NEJM199706053362307> PMID:9171069
6. João C, Bergantim R, Neves M, Chacim S, Afonso C, Barradas J, Bernardo M, Coelho H, Esteves G, Fraga C, Geraldés C, Gonçalves C, Jorge A, et al. Multiple myeloma in elderly patients—a Portuguese multicentric real-life study. *Ann Hematol*. 2019; 98:1689–701. <https://doi.org/10.1007/s00277-019-03640-y> PMID:30963200
7. Song X, Wilson KL, Kagan J, Panjabi S. Cost of peripheral neuropathy in patients receiving treatment for multiple myeloma: a US administrative claims analysis. *Ther Adv Hematol*. 2019; 10:2040620719839025. <https://doi.org/10.1177/2040620719839025> PMID:30967926
8. Mitsiades CS, Davies FE, Laubach JP, Joshua D, San Miguel J, Anderson KC, Richardson PG. Future directions of next-generation novel therapies, combination approaches, and the development of personalized medicine in myeloma. *J Clin Oncol*. 2011; 29:1916–23. <https://doi.org/10.1200/JCO.2010.34.0760> PMID:21482978
9. Sonneveld P, Avet-Loiseau H, Lonial S, Usmani S, Siegel D, Anderson KC, Chng WJ, Moreau P, Attal M, Kyle RA, Caers J, Hillengass J, San Miguel J, et al. Treatment of multiple myeloma with high-risk cytogenetics: a consensus of the International Myeloma Working Group. *Blood*. 2016; 127:2955–62. <https://doi.org/10.1182/blood-2016-01-631200> PMID:27002115
10. Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics*. 2007; 8:22. <https://doi.org/10.1186/1471-2105-8-22> PMID:17250769
11. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005; 4:Article17. <https://doi.org/10.2202/1544-6115.1128> PMID:16646834
12. Liu Z, Zhang X, Lei H, Lam N, Carter S, Yockey O, Xu M, Mendoza A, Hernandez ER, Wei JS, Khan J, Yohe ME, Shern JF, Thiele CJ. CASZ1 induces skeletal muscle and rhabdomyosarcoma differentiation through a feed-forward loop with MYOD and MYOG. *Nat Commun*. 2020; 11:911. <https://doi.org/10.1038/s41467-020-14684-4> PMID:32060262
13. Birkhahn M, Mitra AP, Cote RJ. Molecular markers for bladder cancer: the road to a multimarker approach. *Expert Rev Anticancer Ther*. 2007; 7:1717–27. <https://doi.org/10.1586/14737140.7.12.1717> PMID:18062746
14. Croner RS, Förtsch T, Brückl WM, Rödel F, Rödel C, Papadopoulos T, Brabletz T, Kirchner T, Sachs M, Behrens J, Klein-Hitpass L, Stürzl M, Hohenberger W, Lausen B. Molecular signature for lymphatic metastasis in colorectal carcinomas. *Ann Surg*. 2008; 247:803–10. <https://doi.org/10.1097/SLA.0b013e31816bcd49> PMID:18438117
15. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, Ma ZL, Liu ZY. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol*. 2016; 34:2157–64. <https://doi.org/10.1200/JCO.2015.65.9128> PMID:27138577
16. Chaturvedi N, de Menezes RX, Goeman JJ. Fused Lasso algorithm for Cox' proportional hazards and binomial logit models with application to copy number profiles. *Biom J*. 2014; 56:477–92. <https://doi.org/10.1002/bimj.201200241> PMID:24496763
17. Li Y, Chen X, Wang Y, Hu J, Shen Z, Ding X. Application of group LASSO regression based Bayesian networks in risk factors exploration and disease prediction for acute kidney injury in hospitalized patients with hematologic malignancies. *BMC Nephrol*. 2020; 21:162. <https://doi.org/10.1186/s12882-020-01786-w> PMID:32370757
18. Gozzetti A, Papini G, Defina M, Bocchia M. Extramedullary myeloma relapses. *Ann Hematol*. 2012; 91:1511–12. <https://doi.org/10.1007/s00277-012-1432-3> PMID:22362123
19. Chan HS, Chen CI, Reece DE. Current Review on High-Risk Multiple Myeloma. *Curr Hematol Malig Rep*. 2017; 12:96–108. <https://doi.org/10.1007/s11899-017-0368-z> PMID:28317082

20. Zhang Y, Chen XL, Chen WM, Zhou HB. Prognostic Nomogram for the Overall Survival of Patients with Newly Diagnosed Multiple Myeloma. *Biomed Res Int.* 2019; 2019:5652935. <https://doi.org/10.1155/2019/5652935> PMID:31080823
21. Qi T, Qu J, Tu C, Lu Q, Li G, Wang J, Qu Q. Super-Enhancer Associated Five-Gene Risk Score Model Predicts Overall Survival in Multiple Myeloma Patients. *Front Cell Dev Biol.* 2020; 8:596777. <https://doi.org/10.3389/fcell.2020.596777> PMID:33344452
22. Zhu FX, Wang XT, Ye ZZ, Gan ZP, Lai YR. Construction of a prognosis-associated long noncoding RNA-mRNA network for multiple myeloma based on microarray and bioinformatics analysis. *Mol Med Rep.* 2020; 21:999–1010. <https://doi.org/10.3892/mmr.2020.10930> PMID:32016443
23. Manier S, Liu CJ, Avet-Loiseau H, Park J, Shi J, Campigotto F, Salem KZ, Huynh D, Glavey SV, Rivotto B, Sacco A, Roccaro AM, Bouyssou J, et al. Prognostic role of circulating exosomal miRNAs in multiple myeloma. *Blood.* 2017; 129:2429–36. <https://doi.org/10.1182/blood-2016-09-742296> PMID:28213378
24. Sun C, Li H, Mills RE, Guan Y. Prognostic model for multiple myeloma progression integrating gene expression and clinical features. *Gigascience.* 2019; 8:giz153. <https://doi.org/10.1093/gigascience/giz153> PMID:31886876
25. Dar AA, Goff LW, Majid S, Berlin J, El-Rifai W. Aurora kinase inhibitors--rising stars in cancer therapeutics? *Mol Cancer Ther.* 2010; 9:268–78. <https://doi.org/10.1158/1535-7163.MCT-09-0765> PMID:20124450
26. Vader G, Lens SM. The Aurora kinase family in cell division and cancer. *Biochim Biophys Acta.* 2008; 1786:60–72. <https://doi.org/10.1016/j.bbcan.2008.07.003> PMID:18662747
27. Nikonova AS, Astsaturov I, Serebriiskii IG, Dunbrack RL Jr, Golemis EA. Aurora A kinase (AURKA) in normal and pathological cell division. *Cell Mol Life Sci.* 2013; 70:661–87. <https://doi.org/10.1007/s00018-012-1073-7> PMID:22864622
28. Katayama H, Sasai K, Kawai H, Yuan ZM, Bondaruk J, Suzuki F, Fujii S, Arlinghaus RB, Czerniak BA, Sen S. Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. *Nat Genet.* 2004; 36:55–62. <https://doi.org/10.1038/ng1279> PMID:14702041
29. Dar AA, Zaika A, Piauzelo MB, Correa P, Koyama T, Belkhir A, Washington K, Castells A, Pera M, El-Rifai W. Frequent overexpression of Aurora Kinase A in upper gastrointestinal adenocarcinomas correlates with potent antiapoptotic functions. *Cancer.* 2008; 112:1688–98. <https://doi.org/10.1002/cncr.23371> PMID:18311783
30. Yang H, He L, Kruk P, Nicosia SV, Cheng JQ. Aurora-A induces cell survival and chemoresistance by activation of Akt through a p53-dependent manner in ovarian cancer cells. *Int J Cancer.* 2006; 119:2304–12. <https://doi.org/10.1002/ijc.22154> PMID:16894566
31. Görgün G, Calabrese E, Hideshima T, Ecsedy J, Perrone G, Mani M, Ikeda H, Bianchi G, Hu Y, Cirstea D, Santo L, Tai YT, Nahar S, et al. A novel Aurora-A kinase inhibitor MLN8237 induces cytotoxicity and cell-cycle arrest in multiple myeloma. *Blood.* 2010; 115:5202–13. <https://doi.org/10.1182/blood-2009-12-259523> PMID:20382844
32. Zhou W, Yang Y, Xia J, Wang H, Salama ME, Xiong W, Xu H, Shetty S, Chen T, Zeng Z, Shi L, Zangari M, Miles R, et al. NEK2 induces drug resistance mainly through activation of efflux drug pumps and is associated with poor prognosis in myeloma and other cancers. *Cancer Cell.* 2013; 23:48–62. <https://doi.org/10.1016/j.ccr.2012.12.001> PMID:23328480
33. Qin Y, Zhang S, Deng S, An G, Qin X, Li F, Xu Y, Hao M, Yang Y, Zhou W, Chang H, Qiu L. Epigenetic silencing of miR-137 induces drug resistance and chromosomal instability by targeting AURKA in multiple myeloma. *Leukemia.* 2017; 31:1123–35. <https://doi.org/10.1038/leu.2016.325> PMID:27857131
34. Shi Y, Reiman T, Li W, Maxwell CA, Sen S, Pilarski L, Daniels TR, Penichet ML, Feldman R, Lichtenstein A. Targeting aurora kinases as therapy in multiple myeloma. *Blood.* 2007; 109:3915–21. <https://doi.org/10.1182/blood-2006-07-037671> PMID:17213289
35. Hose D, Rème T, Meissner T, Moreaux J, Seckinger A, Lewis J, Benes V, Benner A, Hundemer M, Hielscher T, Shaughnessy JD Jr, Barlogie B, Neben K, et al. Inhibition of aurora kinases for tailored risk-adapted treatment of multiple myeloma. *Blood.* 2009; 113:4331–40. <https://doi.org/10.1182/blood-2008-09-178350> PMID:19171872
36. Michaelis M, Selt F, Rothweiler F, Wiese M, Cinatl J Jr. ABCG2 impairs the activity of the aurora kinase inhibitor tozasertib but not of alisertib. *BMC Res Notes.* 2015; 8:484.

- <https://doi.org/10.1186/s13104-015-1405-4>  
PMID:26415506
37. Long ZJ, Wang LX, Zheng FM, Chen JJ, Luo Y, Tu XX, Lin DJ, Lu G, Liu Q. A novel compound against oncogenic Aurora kinase A overcomes imatinib resistance in chronic myeloid leukemia cells. *Int J Oncol.* 2015; 46:2488–96.  
<https://doi.org/10.3892/ijo.2015.2960>  
PMID:25872528
38. DeLuca JG, Dong Y, Hergert P, Strauss J, Hickey JM, Salmon ED, McEwen BF. Hec1 and nuf2 are core components of the kinetochore outer plate essential for organizing microtubule attachment sites. *Mol Biol Cell.* 2005; 16:519–31.  
<https://doi.org/10.1091/mbc.e04-09-0852>  
PMID:15548592
39. Sansregret L, Vanhaesebroeck B, Swanton C. Determinants and clinical implications of chromosomal instability in cancer. *Nat Rev Clin Oncol.* 2018; 15: 139–50.  
<https://doi.org/10.1038/nrclinonc.2017.198>  
PMID:29297505
40. Obara W, Sato F, Takeda K, Kato R, Kato Y, Kanehira M, Takata R, Mimata H, Sugai T, Nakamura Y, Fujioka T. Phase I clinical trial of cell division associated 1 (CDCA1) peptide vaccination for castration resistant prostate cancer. *Cancer Sci.* 2017; 108:1452–57.  
<https://doi.org/10.1111/cas.13278> PMID:28498618
41. Sugimasa H, Taniue K, Kurimoto A, Takeda Y, Kawasaki Y, Akiyama T. Heterogeneous nuclear ribonucleoprotein K upregulates the kinetochore complex component NUF2 and promotes the tumorigenicity of colon cancer cells. *Biochem Biophys Res Commun.* 2015; 459:29–35.  
<https://doi.org/10.1016/j.bbrc.2015.02.043>  
PMID:25701787
42. Xu W, Wang Y, Wang Y, Lv S, Xu X, Dong X. Screening of differentially expressed genes and identification of NUF2 as a prognostic marker in breast cancer. *Int J Mol Med.* 2019; 44:390–404.  
<https://doi.org/10.3892/ijmm.2019.4239>  
PMID:31198978
43. Huang SK, Qian JX, Yuan BQ, Lin YY, Ye ZX, Huang SS. SiRNA-mediated knockdown against NUF2 suppresses tumor growth and induces cell apoptosis in human glioma cells. *Cell Mol Biol (Noisy-le-grand).* 2014; 60:30–36.  
PMID:25481014
44. Hu P, Chen X, Sun J, Bie P, Zhang LD. siRNA-mediated knockdown against NUF2 suppresses pancreatic cancer proliferation *in vitro* and *in vivo*. *Biosci Rep.* 2015; 35:e00170.
- <https://doi.org/10.1042/BSR20140124>  
PMID:25370920
45. Liu Q, Dai SJ, Li H, Dong L, Peng YP. Silencing of NUF2 inhibits tumor growth and induces apoptosis in human hepatocellular carcinomas. *Asian Pac J Cancer Prev.* 2014; 15:8623–29.  
<https://doi.org/10.7314/apjcp.2014.15.20.8623>  
PMID:25374179
46. Hose D, Rème T, Hielscher T, Moreaux J, Messner T, Seckinger A, Benner A, Shaughnessy JD Jr, Barlogie B, Zhou Y, Hillengass J, Bertsch U, Neben K, et al. Proliferation is a central independent prognostic factor and target for personalized and risk-adapted treatment in multiple myeloma. *Haematologica.* 2011; 96:87–95.  
<https://doi.org/10.3324/haematol.2010.030296>  
PMID:20884712
47. Nehar S, Mishra M, Heese K. Identification and characterisation of the novel amyloid-beta peptide-induced protein p17. *FEBS Lett.* 2009; 583:3247–53.  
<https://doi.org/10.1016/j.febslet.2009.09.018>  
PMID:19755123
48. Pramanik S, Kutzner A, Heese K. Lead discovery and in silico 3D structure modeling of tumorigenic FAM72A (p17). *Tumour Biol.* 2015; 36:239–49.  
<https://doi.org/10.1007/s13277-014-2620-7>  
PMID:25234718
49. Heese K. The protein p17 signaling pathways in cancer. *Tumour Biol.* 2013; 34:4081–87.  
<https://doi.org/10.1007/s13277-013-0999-1>  
PMID:23900679
50. Guo C, Zhang X, Fink SP, Platzer P, Wilson K, Willson JK, Wang Z, Markowitz SD. Ugene, a newly identified protein that is commonly overexpressed in cancer and binds uracil DNA glycosylase. *Cancer Res.* 2008; 68:6118–26.  
<https://doi.org/10.1158/0008-5472.CAN-08-1259>  
PMID:18676834
51. Wang LT, Lin CS, Chai CY, Liu KY, Chen JY, Hsu SH. Functional interaction of Ugene and EBV infection mediates tumorigenic effects. *Oncogene.* 2011; 30:2921–32.  
<https://doi.org/10.1038/onc.2011.16> PMID:21317926
52. Dai X, Chen X, Hakizimana O, Mei Y. Genetic interactions between ANLN and KDR are prognostic for breast cancer survival. *Oncol Rep.* 2019; 42:2255–66.  
<https://doi.org/10.3892/or.2019.7332> PMID:31578580
53. Wang Z, Chen J, Zhong MZ, Huang J, Hu YP, Feng DY, Zhou ZJ, Luo X, Liu ZQ, Jiang WZ, Zhou WB. Overexpression of ANLN contributed to poor prognosis of anthracycline-based chemotherapy in breast cancer patients. *Cancer Chemother Pharmacol.* 2017; 79: 535–43.



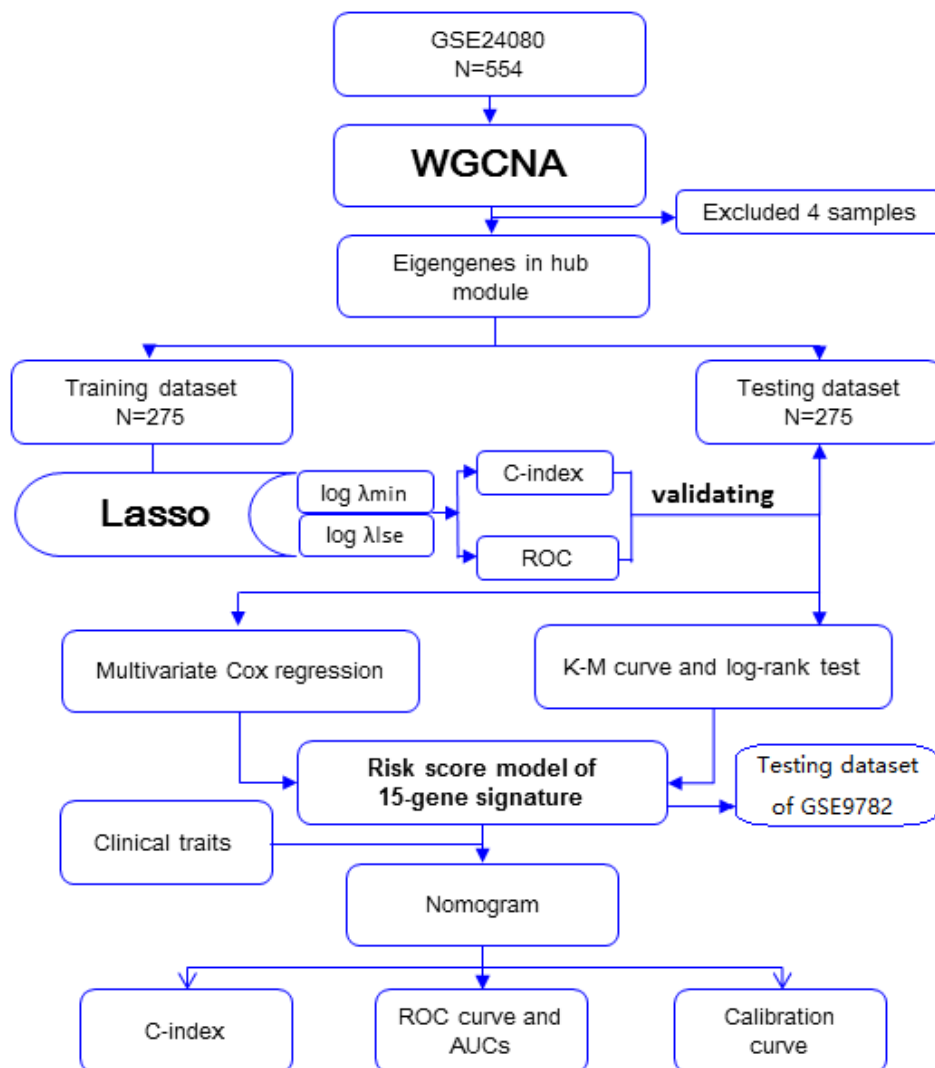
<https://doi.org/10.1007/s00280-017-3248-2>

PMID:[28243684](https://pubmed.ncbi.nlm.nih.gov/28243684/)

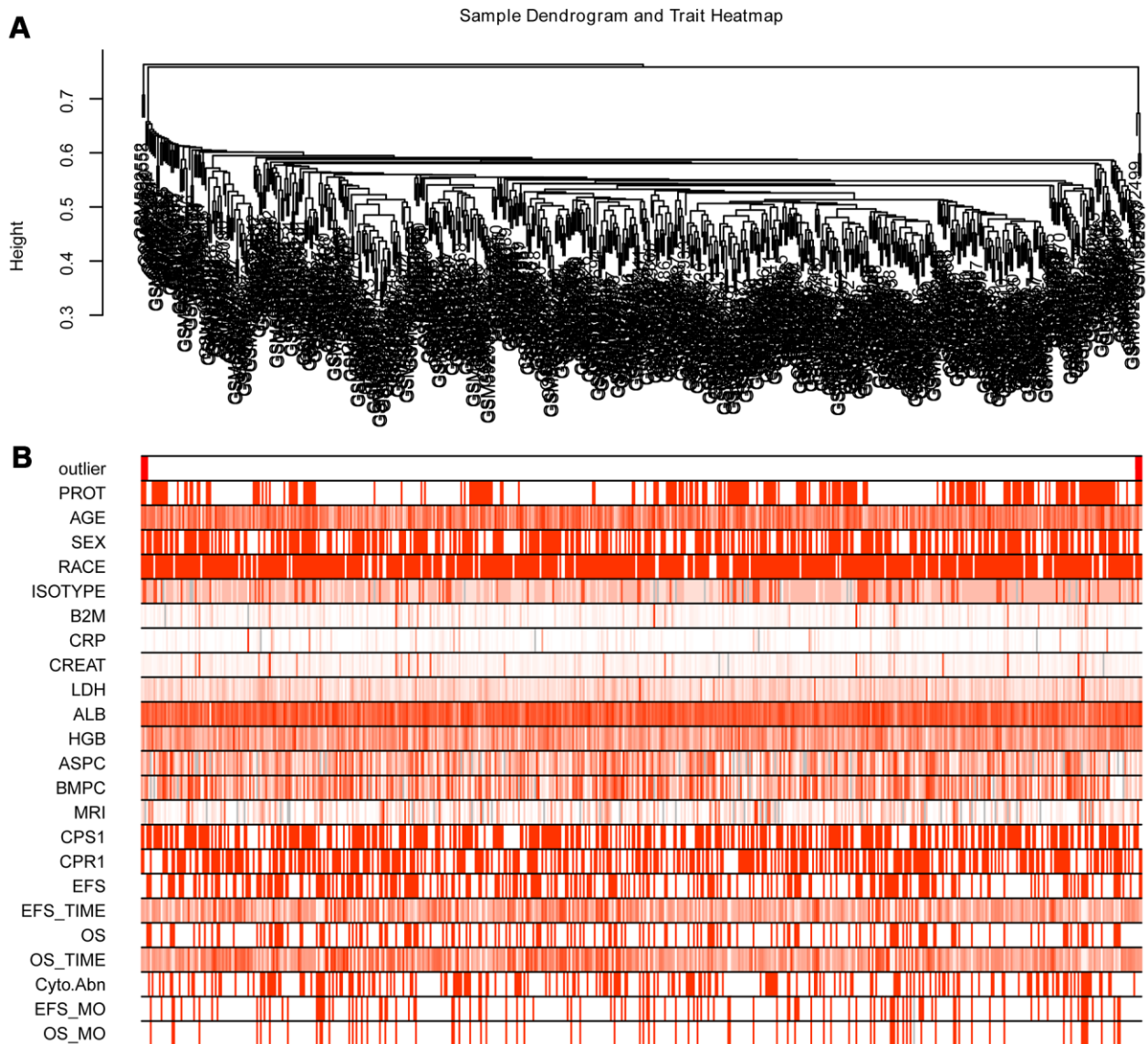
54. Long X, Zhou W, Wang Y, Liu S. Prognostic significance of ANLN in lung adenocarcinoma. *Oncol Lett.* 2018; 16:1835–40.  
<https://doi.org/10.3892/ol.2018.8858> PMID:[30008873](https://pubmed.ncbi.nlm.nih.gov/30008873/)
55. Ning XH, Qi YY, Wang FX, Li SC, Jia ZK, Yang JJ. A Three Protein-Coding Gene Prognostic Model Predicts Overall Survival in Bladder Cancer Patients. *Biomed Res Int.* 2020; 2020:7272960.  
<https://doi.org/10.1155/2020/7272960>  
PMID:[33150179](https://pubmed.ncbi.nlm.nih.gov/33150179/)
56. Zhang X, Zhuang R, Ye Q, Zhuo J, Chen K, Lu D, Wei X, Xie H, Xu X, Zheng S. High Expression of Human AugminComplex Subunit 3 Indicates Poor Prognosis and Associates with Tumor Progression in Hepatocellular Carcinoma. *J Cancer.* 2019; 10:1434–43.  
<https://doi.org/10.7150/jca.28317> PMID:[31031853](https://pubmed.ncbi.nlm.nih.gov/31031853/)
57. Huo X, Sun H, Liu Q, Ma X, Peng P, Yu M, Zhang Y, Cao D, Shen K. Clinical and Expression Significance of AKT1 by Co-expression Network Analysis in Endometrial Cancer. *Front Oncol.* 2019; 9:1147.  
<https://doi.org/10.3389/fonc.2019.01147>  
PMID:[31781484](https://pubmed.ncbi.nlm.nih.gov/31781484/)
58. Ouyang Z, Li G, Zhu H, Wang J, Qi T, Qu Q, Tu C, Qu J, Lu Q. Construction of a Five-Super-Enhancer-Associated-Genes Prognostic Model for Osteosarcoma Patients. *Front Cell Dev Biol.* 2020; 8:598660.  
<https://doi.org/10.3389/fcell.2020.598660>  
PMID:[33195283](https://pubmed.ncbi.nlm.nih.gov/33195283/)
59. Ma H, Qu J, Luo J, Qi T, Tan H, Jiang Z, Zhang H, Qu Q. Super-Enhancer-Associated Hub Genes In Chronic Myeloid Leukemia Identified Using Weighted Gene Co-Expression Network Analysis. *Cancer Manag Res.* 2019; 11:10705–18.  
<https://doi.org/10.2147/CMAR.S214614>  
PMID:[31920381](https://pubmed.ncbi.nlm.nih.gov/31920381/)
60. Dai Y, Lv Q, Qi T, Qu J, Ni H, Liao Y, Liu P, Qu Q. Identification of hub methylated-CpG sites and associated genes in oral squamous cell carcinoma. *Cancer Med.* 2020; 9:3174–87.  
<https://doi.org/10.1002/cam4.2969>  
PMID:[32155325](https://pubmed.ncbi.nlm.nih.gov/32155325/)

SUPPLEMENTARY MATERIALS

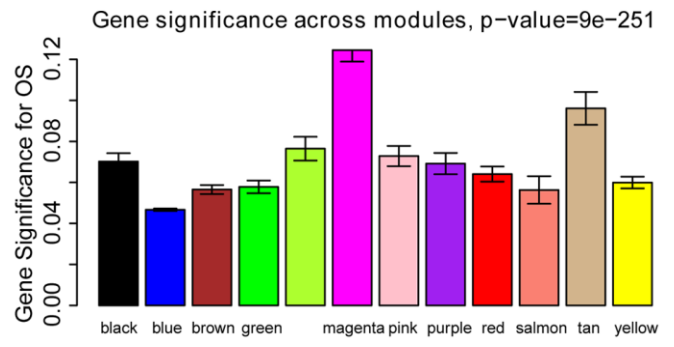
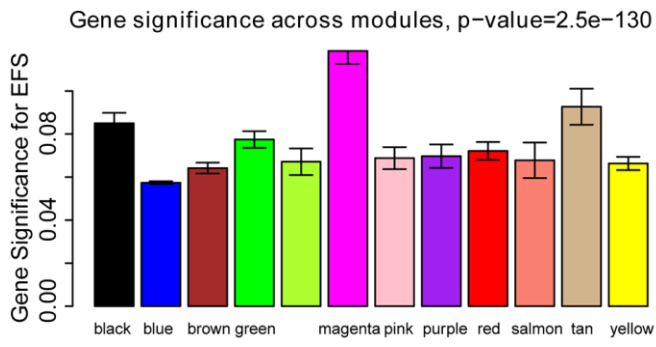
Supplementary Figures



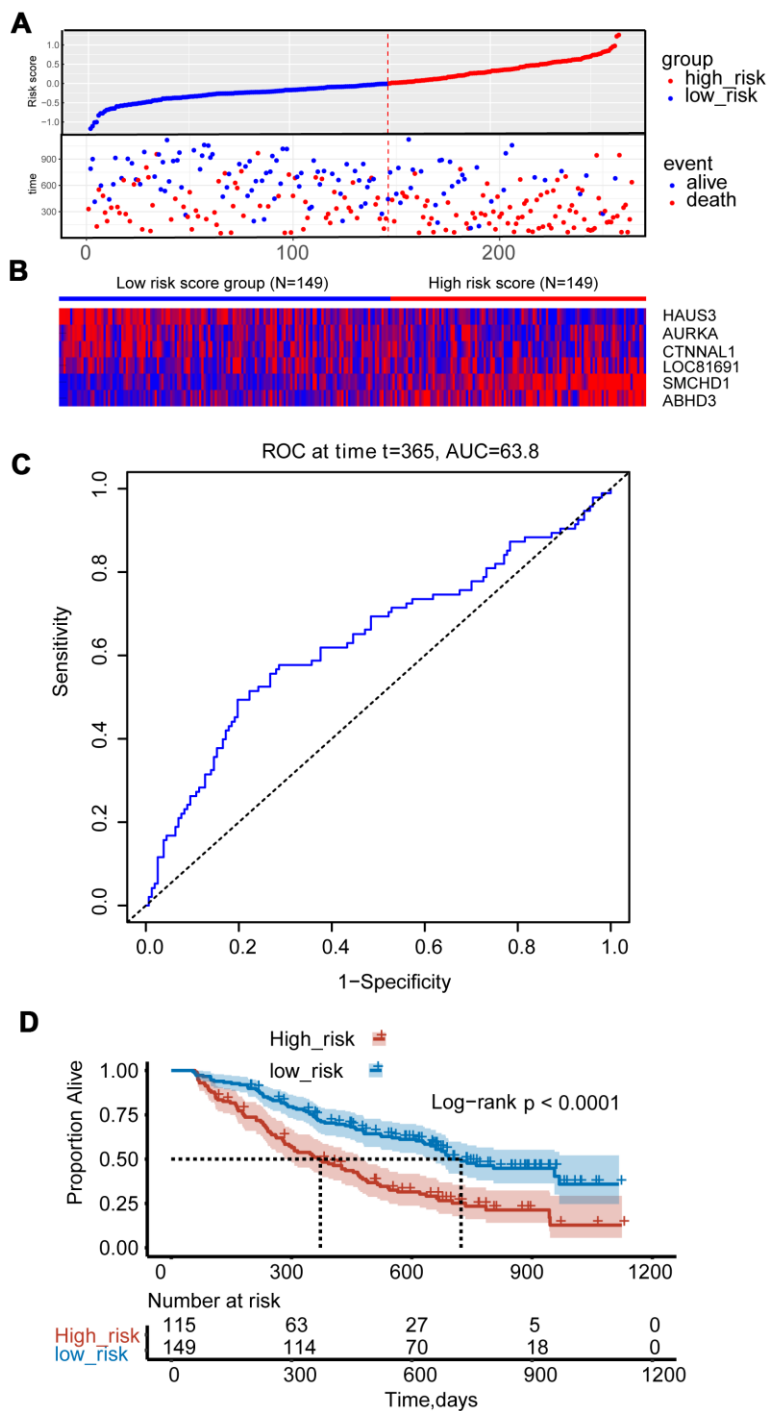
Supplementary Figure 1. The workflow chart of this study.



**Supplementary Figure 2. WGCNA construction.** (A) Clustering dendrogram of 554 samples of MM. Four outlier samples (GSM592558, GSM592552, GSM592499 and GSM592597) were excluded using the flashclust function of R package “WGCNA”. (B) Heat map showed the clinical datasets of MM patients.



**Supplementary Figure 3. Distribution of average gene significance and errors in the modules associated with EFS and OS of MM.**



**Supplementary Figure 4. Risk score model for GSE9782.** (A) Construction of risk score model using the fifteen-gene expression profile in MM patients of GSE9782. (A) The selected six-gene risk score distribution of MM patients based on risk score levels. The best cut-off value was used to divide the patients into two groups. The survival status of all patients was distributed and classified by risk score group. (B) The expression profiles of six genes in high-risk and low-risk groups. (C) ROC curve for the fifteen-gene risk score model to predict overall survival of MM patients. AUC was 0.638. (D) Kaplan-Meier analysis of the six-gene risk score model.

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 2, 4.

**Supplementary Table 1. Univariate and multivariate Cox regression analysis for clinical characteristics and risk score.**

**Supplementary Table 2. An interactive tool of nomogram predicts overall survival of MM patients based on excel.**

**Supplementary Table 3. Summary of clinical characteristics of GSE24080.**

Items		Input data	
AGE	1 year - 80 years	<b>64.16</b>	<b>years</b>
CREAT	Creatinine	<b>0.6</b>	<b>mg/dl</b>
LDH	Lactate dehydrogenase	<b>185</b>	<b>U/l</b>
ALB	Albumin	<b>4.2</b>	<b>g/dl</b>
Cyto_Abn	An indicator of the detection of cytogenetic abnormalities	<b>1</b>	<b>1=abnormal; 0=no detected or absent</b>
Riskscore	Risk score based on genes	<b>1.01</b>	<b>1=high risk; 0=low risk</b>
<b>Results:</b>	1-year Survival Prob		<b>0.858645467</b>
	3-years Survival Prob		<b>0.61493303</b>
	5-years Survival Prob		<b>0.408277594</b>
	7-years Survival Prob		<b>0.271462399</b>

Nomogram predicts overall survival of MM patients.

**Supplementary Table 4. Code for interactive nomogram.**