Identification of a three-gene-based prognostic model in multiple myeloma using bioinformatics analysis

Ying Pan*, Ye Meng*, Zhimin Zhai and Shudao Xiong

Department of Hematology, The Second Affiliated Hospital of Anhui Medical University, Hefei, China
* These authors contributed equally to this work.

ABSTRACT

Background: Multiple myeloma (MM), the second most hematological malignancy, has high incidence and remains incurable till now. The pathogenesis of MM is poorly understood. This study aimed to identify novel prognostic model for MM on gene expression profiles.

Methods: Gene expression datas of MM (GSE6477, GSE136337) were downloaded from Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) in GSE6477 between case samples and normal control samples were screened by the limma package. Meanwhile, enrichment analysis was conducted, and a protein-protein interaction (PPI) network of these DEGs was established by STRING and cytoscape software. Co-expression modules of genes were built by Weighted Correlation Network Analysis (WGCNA). Key genes were identified both from hub genes and the DEGs. Univariate and multivariate Cox congression were performed to screen independent prognostic genes to construct a predictive model. The predictive power of the model was evaluated by Kaplan–Meier curve and time-dependent receiver operating characteristic (ROC) curves. Finally, univariate and multivariate Cox regression analyse were used to investigate whether the prognostic model could be independent of other clinical parameters.

Results: GSE6477, including 101 case and 15 normal control, were screened as the datasets. A total of 178 DEGs were identified, including 59 up-regulated and 119 down-regulated genes. In WGCNA analysis, module black and module purple were the most relevant modules with cancer traits, and 92 hub genes in these two modules were selected for further analysis. Next, 47 genes were chosen both from the DEGs and hub genes as key genes. Three genes (LYVE1, RNASE1, and RNASE2) were finally screened by univariate and multivariate Cox regression analyses and used to construct a risk model. In addition, the three-gene prognostic model revealed independent and accurate prognostic capacity in relation to other clinical parameters for MM patients.

Conclusion: In summary, we identified and constructed a three-gene-based prognostic model that could be used to predict overall survival of MM patients.

INTRODUCTION

Multiple myeloma (MM) is the second most common hematological malignancy, and caused by abnormal monoclonal plasma cells. It accounts for 10% of all hematological malignancies (*Mateos & San Miguel, 2017*). Many novel therapy strategies have been developed, and the survival of MM patients has improved considerably (*Nooka et al., 2019*; *Ocio et al., 2014*). However, the prognosis remains poor and the disease is still incurable (*Nijhof et al., 2018*). It is of great significance to identify novel cancer-specific prognostic biomarkers and construct prognostic model to find high-risk MM patients.

The prognostic models of MM usually include clinical information and cytogenetic abnormalities (*Qi et al., 2020*). International Staging System (ISS) and the revised ISS (R-ISS) are the most commonly used prognostic models. Recently, some prognostic models based on gene expression profile have been used to predict prognosis in MM, and even better than R-ISS (*Liu et al., 2019*; *Bai et al., 2020*). Besides, new technologies and methods such as single cell RNA-Seq (scRNA-Seq) have been used to predict the prognosis of MM (*Jang et al., 2019*). Weighted gene co-expression network (WGCNA) is also a novel method which uses numerous data in public databases to explore the underlying mechanism of gene networks, and usually used for gene prediction in large-scale datasets in recent years (*Zhang & Horvath, 2005*; *Langfelder & Horvath, 2008*). It first divides gene co-expression networks of complicated biological processes into several modules, then analyze their potential association with clinical traits to look for key modules and hub genes (*Zhang & Horvath, 2005*). It has been widely utilized in many malignant diseases as well as non-malignant diseases (*Li et al., 2020b*; *Xu et al., 2020*). However, it is rarely used in MM.

In the present study, we aimed to use WGCNA method to identify co-expression networks associated with MM and to find potential prognostic biomarkers and construct prognostic model for MM patients on gene expression profiles.

MATERIALS AND METHODS

Data preparation

The raw data of GSE6477 and GSE136337 were downloaded from the gene expression omnibus (GEO) website (http://www.ncbi.nlm.nlh.gov/geo/). There were 101 case and 15 normal control in GSE6477, was based on GPL96Platforms ([HG-U133A] Affymetrix Human Genome U133A Array). There were 401 cases with overall survival (OS) and other clinical information was chosen to identify prognostic genes and construct the prognostic model in GSE136337. All the patients were randomly divided into 2 groups (60% as the training set and 40% as the test set). The Robust multi-array average (RMA) method was used to normalize data.

Analysis of differentially expressed genes (DEGs)

Limma package was used to identify differentially expressed genes (DEGs) (*Ritchie et al., 2015*), and the threshold value was |log2foldchange| >1.0 and adj.P.Val <0.05.

GO term and KEGG pathway enrichment analysis

To better explore the biological significance of DEGs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed (*Yu et al., 2012*). q value <0.05 was used as the cutoff value.

A protein-protein interaction (PPI) network

A PPI network of these DEGs was constructed by using the Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/) (*Szklarczyk et al., 2015*).

Co-expression network analysis

The co-expression network was constructed by WGCNA to explore the correlation of genes (Top50%mad) and search for important interacted gene modules. PickSoftThreshold function of the WGCNA was used to calculate the soft thresholding power β value, and the β value was set as 6. Then, we constructed the co-expression modules using WGCNA algorithms with R software package. Next, Gene significance (GS) and Module membership measures were used to identify genes with high clinical significance as well as high Module membership. Module membership was the correlation between gene expression profile and module eigengene. While, GS was the absolute value of the correlation between gene expression and module traits.

Identification of prognostic key genes and construction of the prognostic model

Key genes were identified both from hub genes and the DEGs (*Xia et al., 2018*). All the patients were randomly divided into two groups (60% as the training set and 40% as the test set) and univariate and multivariate Cox regression analyses were used to investigate the correlation between OS and the expression level of key genes. *P*-value <0.05 was considered of significant difference in the univariate analysis. Key genes with $P < 0.05$ based on the univariate analysis were further included in the multivariate Cox regression analysis. Finally, a prognostic gene signature of MM patients was established based on a linear combination of the regression coefficient derived from the multivariate Cox regression model (β) multiplied with its expression level. The patients were separated into low- and high-risk groups based on the optimal cut-off value (*Camp, Dolled-Filhart & Rimm, 2004*) Kaplan–Meier (K–M) survival curves, time-dependent receiver operating characteristic (ROC) curve analyses and area under the curve (AUC) were used to evaluate the prediction accuracy of our model (*Heagerty & Zheng, 2005*). We also validated the prognostic model in the test set and whole set by K–M curves, ROC curve analyses, and AUC.

Identification of independent prognostic parameters of MM

To investigate independent prognostic parameters and to identify whether the prognostic model was the independent prognostic parameter, univariate and multivariate Cox regression analyses were performed using the Cox regression model method with forwarding stepwise procedure. Other clinical parameters including age, gender, number
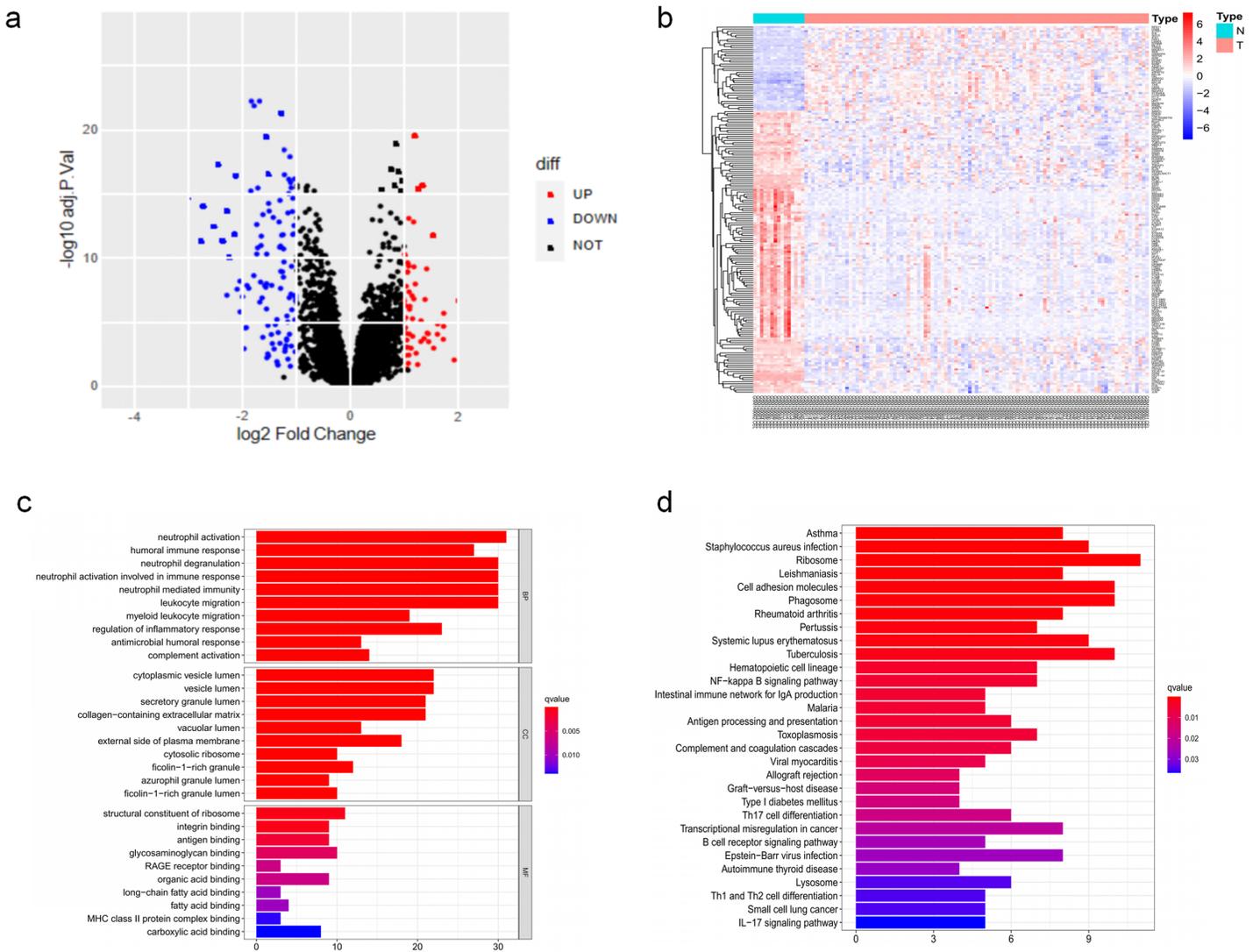
Figure 1 The DEGs identified in GSE6477. (A) Volcano map of DEGs between MM case group and normal control group. (B) Heatmap of the DEGs. (C) The top 20 GO terms in the enrichment analysis of DEGs. (D) The top 30 KEGG pathways in the enrichment analysis of DEGs.

of transplantation (NOT), tumor staging (ISS and R-ISS), cytogenetic abnormalities, lactate dehydrogenase (LDH) level, albumin level, β2-microglobulin level, and percentage of abnormal plasma cell in bone marrow. *P*-value <0.05 was considered of significant difference. Also, we compared the AUC of our model with other conventional models such as age, gender, ISS, and R-ISS models.

RESULTS

Identification of DEGs in MM and the enrichment of these DEGs

The DEGs of GSE6477 were screened by the limma package. A total of 178 DEGs were identified, including 59 upregulated genes and 119 downregulated genes in MM case samples compared to normal control samples (Fig. 1A). The DEGs were shown on a

Figure 2 Analysis of the PPI network. A total of 178 DEGs were filtered into the DEGs PPI network complex. Full-size 🖼 DOI: 10.7717/peerj.11320/fig-2

heatmap (Fig. 1B). GO analysis showed that these genes were significantly enriched in neutrophil activation, humoral immune response, and structural constituent of ribosome (Fig. 1C). In KEGG analysis, the DEGs were mainly enriched in Asthma, *Staphylococcus aureus* infection, and Ribosome (Fig. 1D). Next, A PPI network, as demonstrated in Fig. 2, was constructed using the online STRING website to understand the interaction among the related DEGs in MM.

Co-expression analysis and hub genes identification

WGCNA was used to construct co-expression network and to find the modules of highly correlated genes. Firstly, we calculated the soft thresholding power β, and the power of β was set at 6 to ensure a scale-free network in our research (Figs. 3A–3D). Next, we identified 15 modules for further analysis (Fig. 4A). The heatmap and meta-modules were used to visualize the gene network (Figs. 4B and 4C). Among these 15 modules, module black and module purple were the most relevant modules with cancer traits (Fig. 4D). Subsequently, intramodular analysis of the genes in the 15 modules was followed. And genes in the black and purple modules were found to have high correlations with MM (Figs. 4E and 4F). Finally, genes in these two modules were selected as hub genes according to the cut-off criteria: module membership values >0.80 and GS values >0.2. Totally, there were 92 hub genes from black and purple modules that were chosen for next analysis.

Identification of survival related key genes and construction of three-gene prognostic signature

Key genes were identified both from hub genes and the DEGs. Totally, 47 key genes were identified (Fig. 5). To evaluate the prognostic values of these key genes in MM patients,
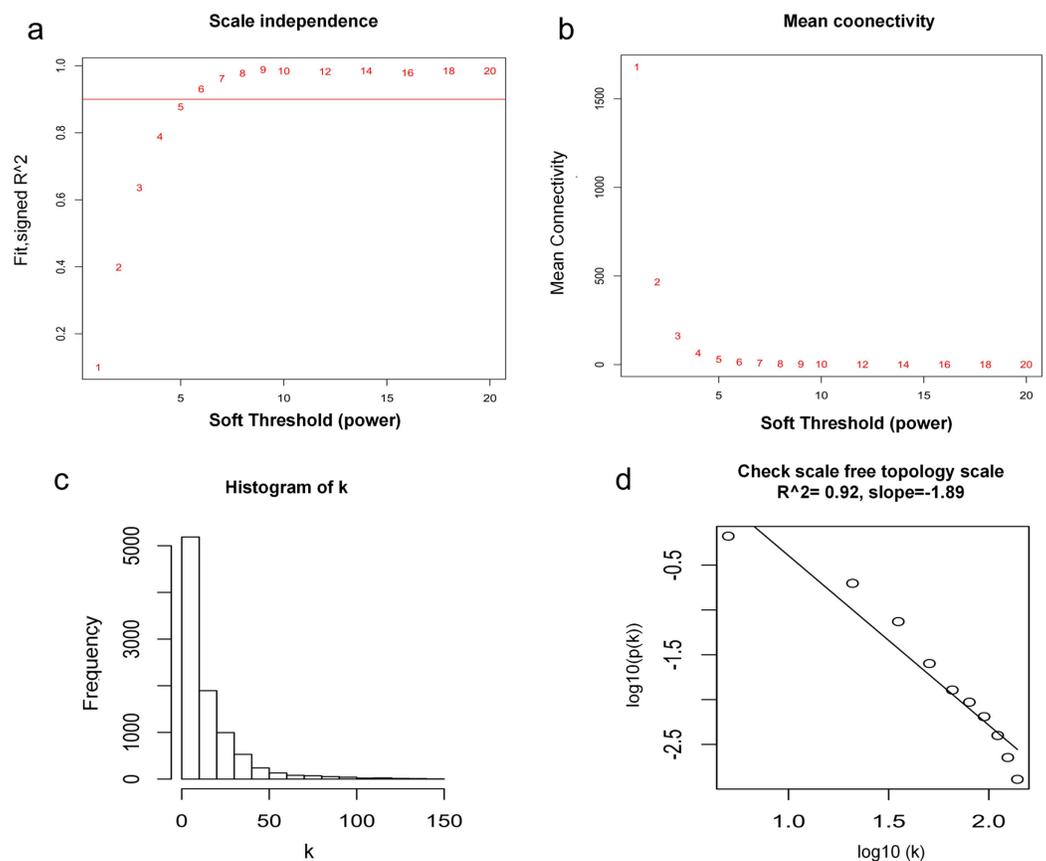
Figure 3 Determination of soft thresholding power. (A) Analysis of the scale-free fitting indices for various soft thresholding powers. (B) Mean connectivity analysis of various soft thresholding powers. (C) Histogram of the connection distribution when β = 6. (D) Check scale-free topology when β = 6.

Full-size 🖻 DOI: 10.7717/peerj.11320/fig-3

we next performed a univariate Cox regression, and identified 8 genes significantly related to OS in MM patients. Then, a stepwise multivariate Cox regression analysis was performed, and three genes (LYVE1, RNASE1, and RNASE2) were eventually selected to construct a prognostic model (Fig. 6). Risk score = (−0.59714 * expression level of LYVE1) + (−1.35971* expression level of RNASE1) + (0.28240 * expression level of RNASE2). Next, we calculated the optimal cut-off value for the three-gene expression risk score with X-Tile software, and divided patients in training set ($n = 241$) into high risk group and low risk group. The K–M curve showed that high risk group had poorer prognosis than low risk group ($P < 0.0001$) (Fig. 7). The AUCs of time-dependent ROC curves were calculated to assess the prognostic capacity of the three-gene signature. The AUCs of risk scores in training set at 8-,10-, and 12-year survival times were 0.717, 0. 720 and 0.800 (Fig. 8A).

To verify the predictive value of the three-gene signature, we used the internal test set ($n = 160$) and whole set ($n = 401$) to assess the results from the training set. Consistent of the findings of the training set, the K–M survival curves of test set and whole set revealed that worse OS existed in the high risk groups (Fig. 7). The AUCs of risk scores in test set
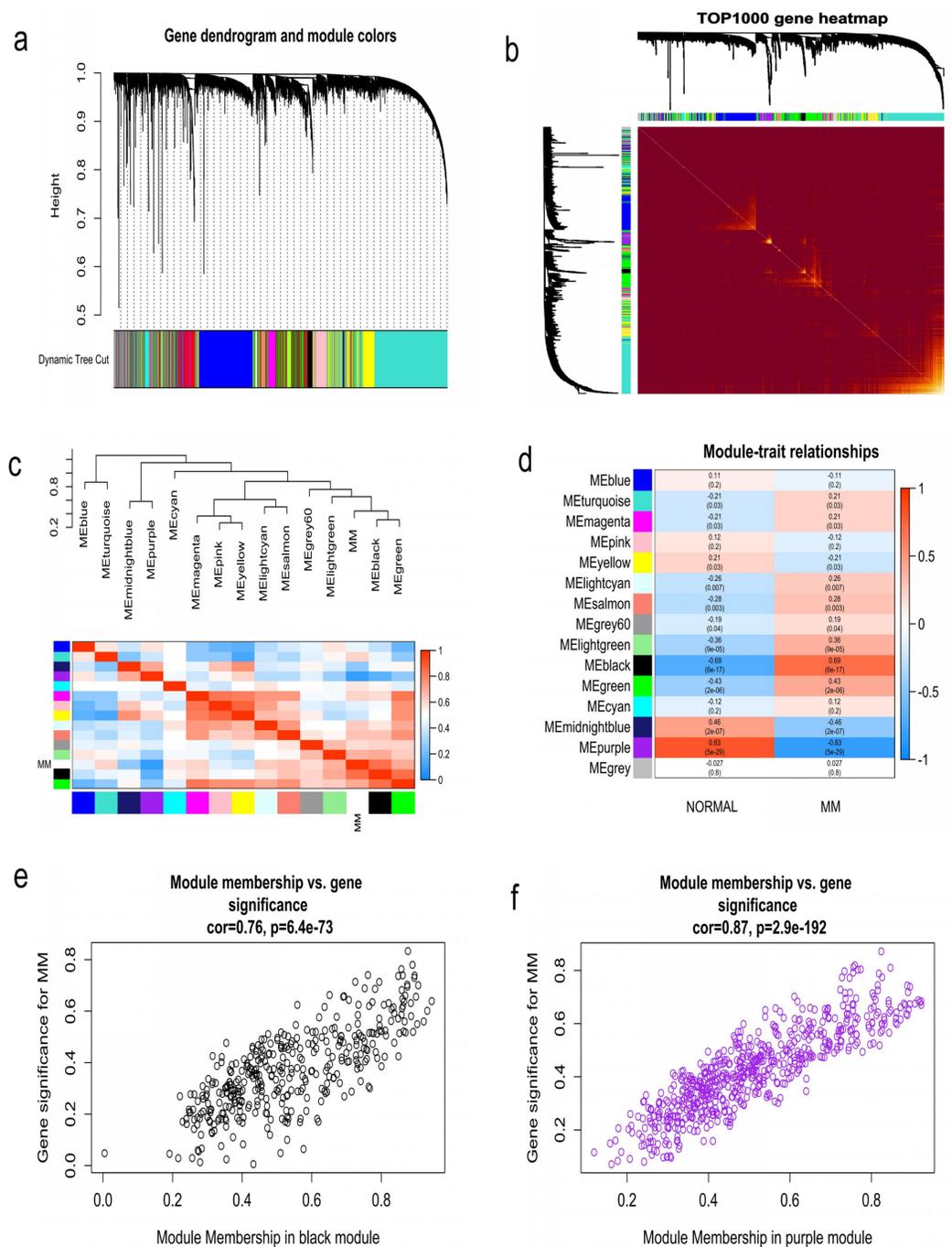
Figure 4 Identification of modules associated with MM clinical traits. (A) Clustering dendrograms of genes based on co-expression network analysis. In total, 15 co-expression modules were constructed and shown in different colors. (B) A heatmap of all the genes. The light color indicated a low overlap, and the darker red color indicated higher overlap. (C) The eigengene dendrogram and eigengene adjacency plot. (D) Module-trait association. The purple and black gene modules were the most relevant modules with cancer traits. (E) A scatter plot of gene significance for MM versus the module membership in the black module. (F) A scatter plot of gene significance for MM versus the module membership in the purple module. Full-size ⬚ DOI: 10.7717/peerj.11320/fig-4
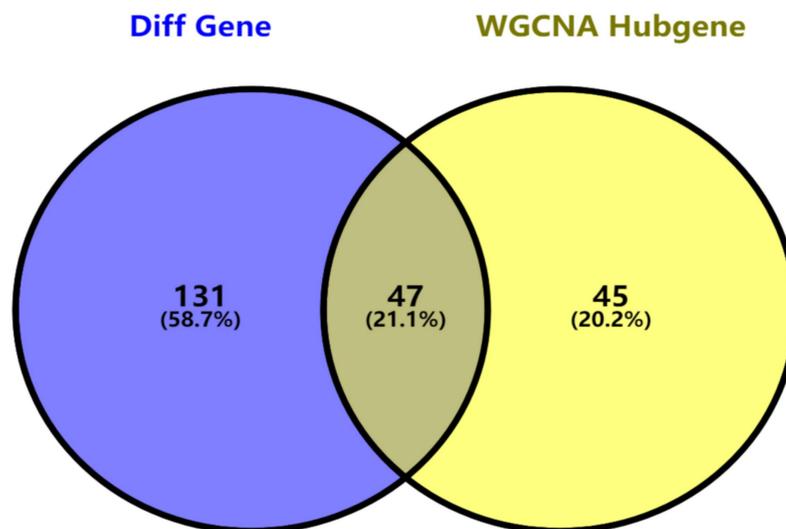
Figure 5 Key genes identified from both the DEGs and the hub genes. A Venn diagram of the DEGs and hub genes revealed 47 key genes.　Full-size ⊡ DOI: 10.7717/peerj.11320/fig-5
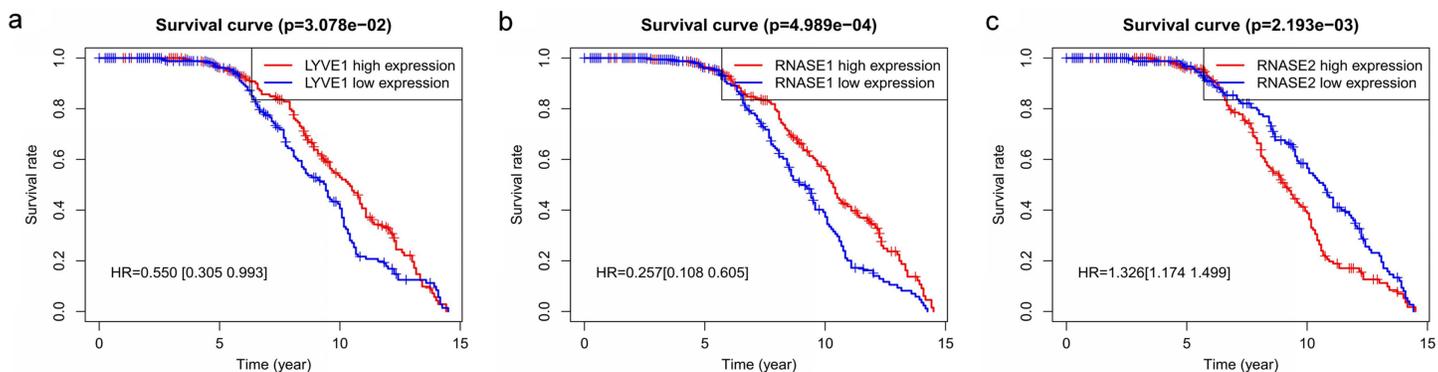

Figure 6 Relationship of the three prognostic genes with OS. Kaplan–Meier curve showing that the three genes (LYVE1, RNASE1, and RNASE2) expression were associated with OS of MM patients (A–C).　Full-size ⊡ DOI: 10.7717/peerj.11320/fig-6

at 8-,10-, and 12-year survival times were 0.709, 0.768 and 0.766 (Fig. 8B), and the whole set were 0.714, 0.741 and 0.783, respectively (Fig. 8C). We also compared the prognostic capacity of our model with other conventional models, including age, gender, ISS, and R-ISS stage in whole set. The AUCs for 8-, 10-, and 12-year OS predictions for age were 0.500, 0.518, and 0.478; and for gender were 0.516, 0.506, and 0.450 (Figs. 8D–8F). The AUCs of ISS stage at 8-, 10-, and 12-year survival times were 0.503, 0.547, and 0.633; and of R-ISS stage were 0.613, 0.616, and 0.634 (Figs. 8D–8F). Our model had the highest AUCs, indicating a superior prognostic value. Together, these results demonstrated that our model was capable of predicting OS of MM patients.

The prognostic model is independent for MM patients

The univariate and multivariate Cox regression analyses were utilized to investigate independent prognostic parameters and identify whether the three-gene prognostic model
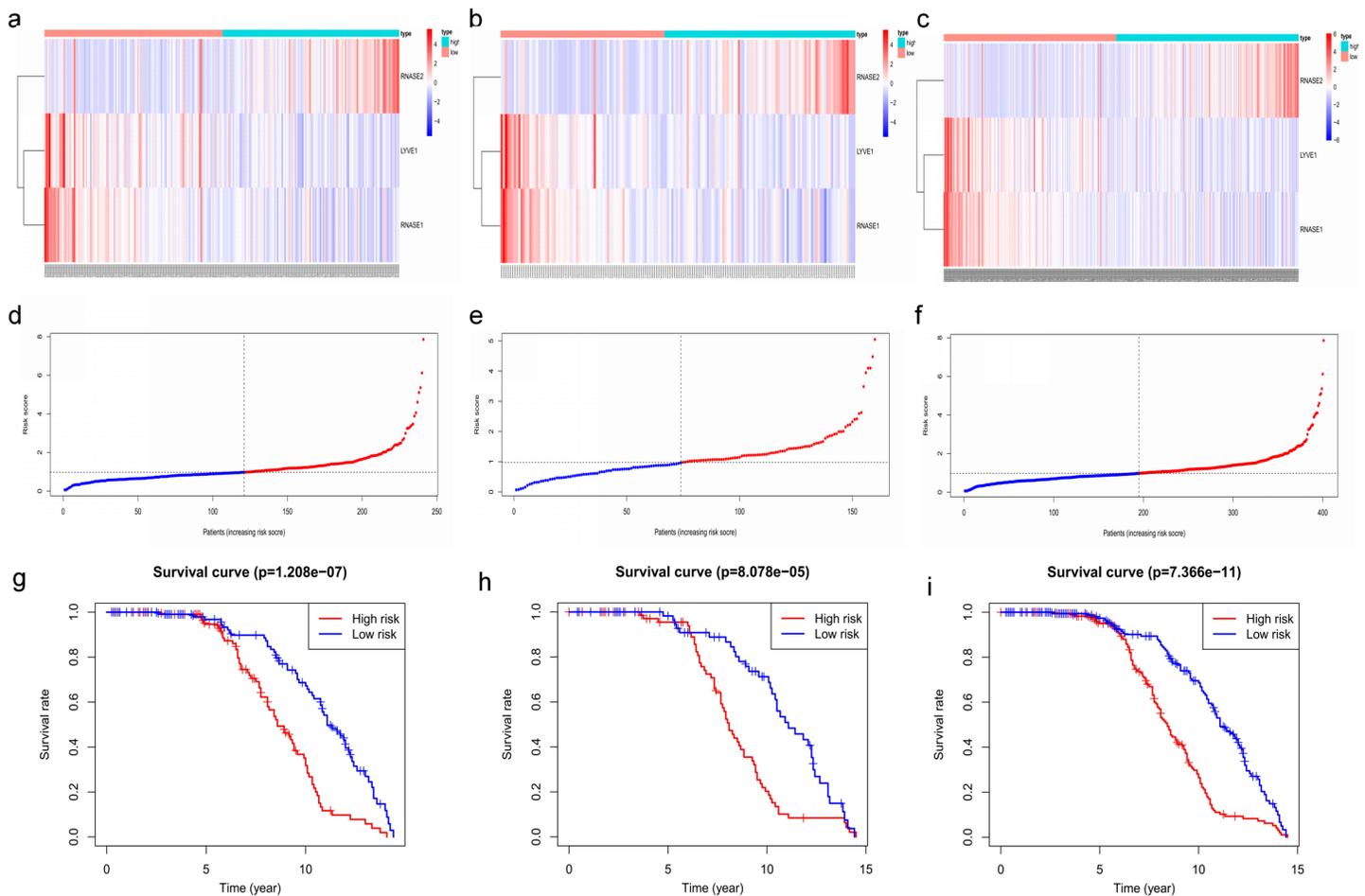
Figure 7 Construction and validation of prognostic risk model for MM patients. (A–C) Heatmaps of the three prognostic genes in training set, test set, and whole set, respectively. (D–F) Risk score distributions in training set, test set, and whole set, respectively. (G–I) Kaplan–Meier curve showing that OS is significantly shorter for patients in high-risk group than those in low-risk group in training set, test set, and whole set, respectively.   Full-size 🖾 DOI: 10.7717/peerj.11320/fig-7

was the independent prognostic parameter in 401 MM patients with complete clinical information from GSE136337. Univariate Cox regression analysis showed that NOT, tumor stage (ISS and R-ISS), t(11,14), albumin level, β2-microglobulin level, and the prognostic model had prognostic values (Fig. 9A). After the multivariate Cox regression analysis, NOT, R-ISS, t(11,14), LDH level and the prognostic model were proved to be independent prognostic factors for OS of MM patients (Fig. 9B).

Correlation between the prognostic model with clinical characters

Correlation between risk score and various clinical parameters were shown in Fig. 10. Our results indicated that risk score was significantly correlated with ISS, and R-ISS stage; patients in phase III of ISS or R-ISS had the highest risk score when compared with patients in phase I and II (Fig. 10). However, no differences were found between risk score and age or gender or NTO (Table 1). Besides, the relationships between the three genes in the prognostic model with clinical characters were also analyzed. We found that as LYVE1
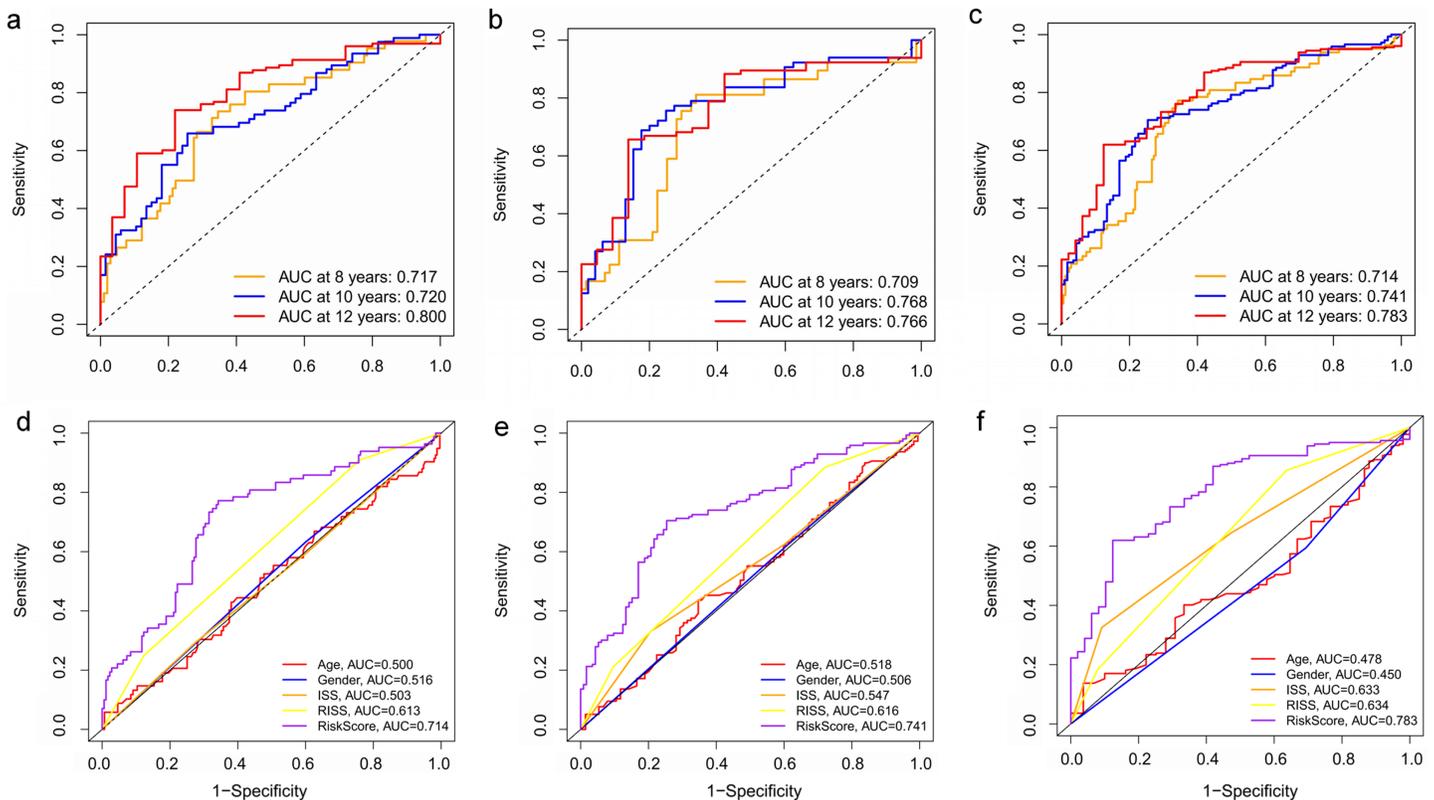
Figure 8 Time-dependent ROC curves for the prognostic model in the GEO MM cohort. (A–C) Time-dependent ROC curve analysis showing the accuracy and reliability of the prognostic model in training set, test set, and whole set, respectively. (D–F) The ROC curves for 8-, 10-, and 12-year overall survival predictions for the risk model in compare with age, gender, ISS, and R-ISS models in whole set.

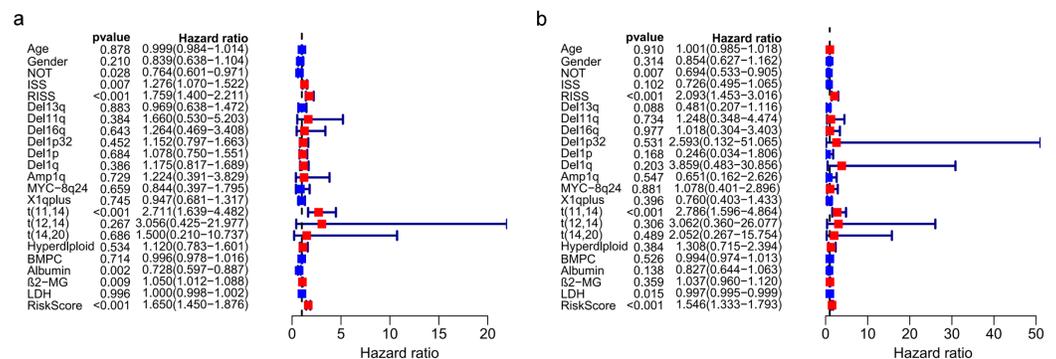Full-size 🖼 DOI: 10.7717/peerj.11320/fig-8

Figure 9 Prognostic risk model could predict prognosis independently. (A). Univariate association of the prognostic model and clinical characteristics with OS. (B). Multivariate association of the prognostic model and clinical characteristics with OS.

Full-size 🖼 DOI: 10.7717/peerj.11320/fig-9

expression, which was higher in males than in females, decreased, ISS stage in MM patients increased significantly (Fig. 10A, Table 1). Additionally, as LYVE1 and RNASE1 expression decreased, R-ISS stage in MM patients increased (Figs. 10B and 10C). These findings showed that LYVE1 and RNASE1 exerted protective effects against disease progression.
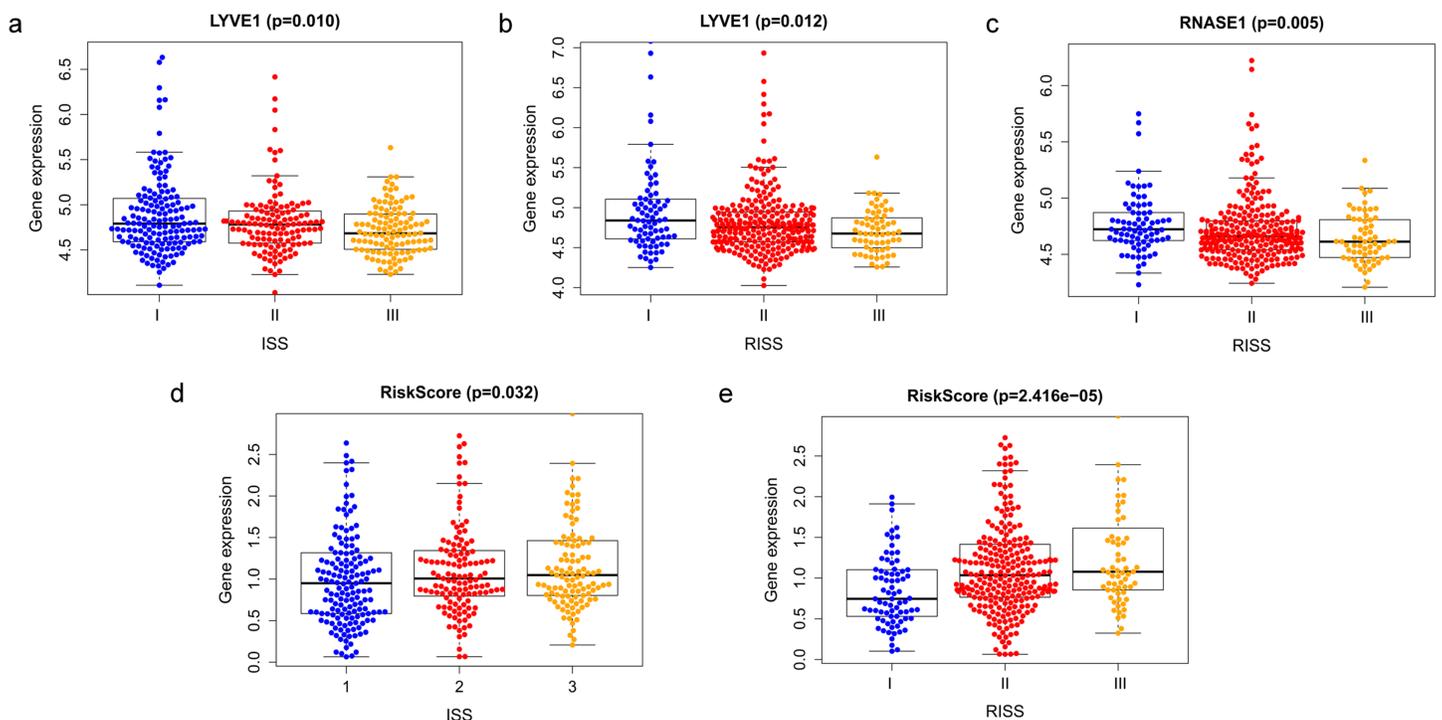
Figure 10 Relationship of prognostic genes and risk model with clinical variables for MM. (A and B) LYVE1 expression was correlated with ISS and R-ISS stage in MM patients. (C) RNASE1 expression was correlated with R-ISS stage. (D and E) Risk score was significantly correlated with ISS and R-ISS stage.

Full-size 🖼 DOI: 10.7717/peerj.11320/fig-10

Table 1 Correlation analysis between three genes and clinical characters for MM.

Variables	Age (≤60, >60) t (p)	Gender (female, male) t (p)	NTO (≤1, >1) t (p)	ISS (I, II, III) t (p)	RISS (I, II, III) t (p)
LYVE1	236.601	−2.477*	3.895	9.293*	8.888*
RNASE1	208.63	−1.529	5.652	5.13	10.782**
RNASE2	239.084	−2.524*	7.929	0.167	3.094
Risk score	238.985	−0.208	5.607	6.887*	21.262***

Notes:
  * $p < 0.05$.
  ** $p < 0.01$.
  *** $p < 0.001$.
  t: t value from Student's t test; p: p-value from Student's test.

DISCUSSION

WGCNA is an important and helpful method which describes the pattern of gene association between different samples, and usually used for gene prediction in large-scale datasets (*Zhang & Horvath, 2005*; Langfelder, 2008). The main advantage of this method is that it can cluster genes into co-expression modules, as well as build a bridge between gene expression changes and sample characteristics. Thus, WGCNA has been widely used in various malignant and non-malignant diseases (*Chen et al., 2020*; *Chen et al., 2019*; *Chen et al., 2018*; *Li et al., 2020a*; *Zheng et al., 2020*). In cancer research, more and more researchers have applied WGCNA for developing new prognostic biomarkers

(*Xu et al., 2020*; *Chen et al., 2020*; *Chen et al., 2019*; *Chen et al., 2018*). In this study, we firstly used an integrated analysis from both WGCNA and DEGs to screen potential biomarkers related to prognosis of MM in GEO databases.

Totally, 178 DEGs were identified, including 59 up-regulated genes and 119 down-regulated genes in GSE6477. Next, we applied WGCNA to construct a co-expression network for evaluating the relationships between genes and modules. In WGCNA, genes were clustered into 15 modules, and the blackmodule and purple module were identified to be the most relevant modules with cancer traits. According to the cut-off criteria: module membership values >0.80 and GS values >0.2, a total of 92 hub genes from these two modules were selected. Key genes were excavated from both the hub genes and the DEGs. Totally, 47 key genes were identified and selected for survival analysis. Using univariate and multivariate Cox regression, we screened three independent prognostic genes and constructed a predictive model. The prognostic genes were LYVE1, RNASE1, and RNASE2. The patients were divided into high- and low-risk groups based on the risk score. Training set, test set and whole set were utilized to validate the prognostic capacity of the our model. Besides, the three-gene prognostic model revealed independence and accuracy in relation to other clinical parameters for MM patients. Our model was the first prognostic model for survival times of MM via integrative bioinformatics analysis.

All of the genes in the three-gene signature have not been previously reported to be associated with MM. LYVE1 is a hyaluronic acid receptor. Its main function is to transport hyaluronidase in extracellular matrix to lymphatic fluid. Also, it involves in the generation of lymphatic vessels and the activation of lymphocytes (*Arimoto et al., 2018*). High expression level of LYVE1 in tumor tissue can induce tumor cell proliferation, lymphangiogenesis and lymphatic metastasis, and was reported to be closely related to poor prognosis in many cancers, including lung cancer, renal cancer, and breast cancer (*Li et al., 2018*; *Schraml et al., 2019*; *Hunter et al., 2019*). However, LYVE1 might also play anti-tumor role in some cancers. In papillary thyroid carcinoma, LYVE1 was proved to be decreased and associated with poor prognosis (*Wu et al., 2019*). *Latil et al. (2003)* reported that LYVE1 was downregulated in prostate cancer and was related to the relapse. In hepatocellular carcinoma and ovarian cancer, its expression was also decreased and played the tumor-suppressive effect (*Llovet et al., 2006*; *Gao et al., 2017*). However, its expression and role in MM has not been reported. In this study, we found that the downregulation of LYVE1 in MM was associated with worse prognosis. In addition, the expression level of LYVE1 was closely related to clinical stage (ISS and R-ISS).

RNASE1, belongs to the ribonuclease (RNase) A gene superfamily, acts to regulate the vascular homeostasis of extracellular RNA (*Bedenbender et al., 2019*). In gastric cancer, RNASE1 acted as a tumor suppressor gene, as its expression was demonstrated to decrease progressively from the normal, primary cancer and metastatic cells (*Wang et al., 2006*). While in prostate cancer, overexpression of RNASE1 was associated with the poor survival (*Gao et al., 2020*).

RNASE2, another member of the RNase A gene family, mainly involves in immune function and plays an important role in toll-like receptor activation (*Ostendorf et al., 2020*). Studies have identified that combination analysis of RNASE2 with other genes has

independent prognostic value in kidney renal clear cell carcinoma (*Xiang et al., 2020*) and renal cell carcinoma (*Qin et al., 2021*). It has also been found to be upregulated in acute lymphoblastic leukemia (*Niini et al., 2002*). The upregulation of RNASE2 in gastric cancer was associated with poorer survival (*Wang et al., 2020*). However, its role in MM is unknown. In our study, overexpressed RNASE2 was found to be associated with worse prognosis in MM patients. The tumor suppressor effect of RNASE2 in MM and its molecular mechanisms need further study.

There are several advantages and limitations in this study. The GEO databases have enough sample size and provide a relatively believable basis for bioinformatics analysis. Besides, in this study, we used an integrated analysis from both WGCNA and DEGs to screen potential biomarkers, which may also make our results more credible. Although our results were validated in other GEO databases, it would be more reliable to validate the significance of prognostic model in real-world clinical MM cohorts. Moreover, biological functions and underlying molecular mechanisms of these key genes in MM are still needed to be explored in the future research.

CONCLUSIONS

In summary, we used comprehensive study including WGCNA, DEGs screening, univariate and multivariate Cox regression to construct a three-gene-based prognostic model, which exhibits an effective prognostic value for distinguishing MM patients with poor OS.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Ying Pan performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.

- Ye Meng performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Zhimin Zhai conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Shudao Xiong conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data of GSE6477 and GSE136337 were downloaded from the gene expression omnibus (GEO) website. Series matrix were downloaded in GSE136337 for validation.

REFERENCES

Arimoto S, Hasegawa T, Takeda D, Saito I, Amano R, Akashi M, Komori T. 2018. Lymphangiogenesis and lymphno demetastasis in oral squamous cell carcinoma. *Anticancer Research* **38(11)**:6157–6162 DOI 10.21873/anticanres.12968.

Bai H, Xu P, Wang L, Chen B. 2020. A centromere-associated gene score for rapid determination of risk in multiple myeloma. *American Journal of Translational Research* **12(6)**:2425–2438.

Bedenbender K, Scheller N, Fischer S, Leiting S, Preissner KT, Schmeck BT, Vollmeister E. 2019. Inflammation-mediated deacetylation of the ribonuclease 1 promoter via histone deacetylase 2 in endothelial cells. *The FASEB Journal* **33(8)**:9017–9029 DOI 10.1096/fj.201900451R.

Camp RL, Dolled-Filhart M, Rimm DL. 2004. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clinical Cancer Research* **10(21)**:7252–7259 DOI 10.1158/1078-0432.CCR-04-0713.

Chen J, Cai Y, Xu R, Pan J, Zhou J, Mei J. 2020. Identification of four hub genes as promising biomarkers to evaluate the prognosis of ovarian cancer in silico. *Cancer Cell International* **20(1)**:270 DOI 10.1186/s12935-020-01361-1.

Chen X, Hu L, Wang Y, Sun W, Yang C. 2019. Single cell gene co-expression network reveals FECH/CROT signature as a prognostic marker. *Cells* **8(7)**:698 DOI 10.3390/cells8070698.

Chen X, Yang C, Xie S, Cheung E. 2018. Long non-coding RNA GAS5 and ZFAS1 are prognostic markers involved in translation targeted by miR-940 in prostate cancer. *Oncotarget* **9(1)**:1048–1062 DOI 10.18632/oncotarget.23254.

Gao Y, Liu X, Li T, Wei L, Yang A, Lu Y, Zhang J, Li L, Wang S, Yin F. 2017. Cross-validation of genes potentially associated with overall survival and drug resistance in ovarian cancer. *Oncology Reports* **37(5)**:3084–3092 DOI 10.3892/or.2017.5534.

Gao L, Meng J, Zhang Y, Gu J, Han Z, Wang X, Gao S. 2020. Development and validation of a six-RNA binding proteins prognostic signature and candidate drugs for prostate cancer. *Genomics* **112(6)**:4980–4992 DOI 10.1016/j.ygeno.2020.08.034.

Heagerty PJ, Zheng Y. 2005. Survival model predictive accuracy and ROC curves. *Biometrics* **61(1)**:92–105 DOI 10.1111/j.0006-341X.2005.030814.x.

Hunter S, Nault B, Ugwuagbo KC, Maiti S, Majumder M. 2019. Mir526b and Mir655 promote tumour associated angiogenesis and lymphangiogenesis in breast cancer. *Cancers* **11(7)**:938 DOI 10.3390/cancers11070938.

Jang JS, Li Y, Mitra AK, Bi L, Abyzov A, van Wijnen AJ, Baughn LB, Van Ness B, Rajkumar V, Kumar S, Jen J. 2019. Molecular signatures of multiple myeloma progression through single cell RNA-Seq. *Blood Cancer Journal 3;* **9(1)**:2 DOI 10.1038/s41408-018-0160-x.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **29(9)**:559 DOI 10.1186/1471-2105-9-559.

Latil A, Bieche I, Chene L, Laurendeau I, Berthon P, Cussenot O, Vidaud M. 2003. Gene expression profiling in clinically localized prostate cancer: a four-gene expression model predicts clinical behavior. *Clinical Cancer Research* **9**:5477–5485.

Li P, Cong Z, Qiang Y, Xiong L, Tang L, Zhang Y, Wu H, Yi J, Jing H, Li D, Shen Y. 2018. Clinical significance of CCBE1 expression in lung cancer. *Molecular Medicine Reports* **17**:2107–2112 DOI 10.3892/mmr.2017.8187.

Li Z, Cui Y, Feng J, Guo Y. 2020a. Identifying the pattern of immune related cells and genes in the peripheral blood of ischemic stroke. *Journal of Translational Medicine* **18(1)**:296 DOI 10.1186/s12967-020-02463-0.

Li W, Wang L, Wu Y, Yuan Z, Zhou J. 2020b. Weighted gene co-expression network analysis to identify key modules and hub genes associated with atrial fibrillation. *International Journal of Molecular Medicine* **45(2)**:401–416 DOI 10.3892/ijmm.2019.4416.

Liu Y, Yang N, Peng X, Liu G, Zhong H, Liu L. 2019. One-lincRNA and five-mRNA based signature for prognosis of multiple myeloma patients undergoing proteasome inhibitors therapy. *Biomedicine & Pharmacotherapy* **18(1)**:109254 DOI 10.1016/j.biopha.2019.109254.

Llovet JM, Chen Y, Wurmbach E, Roayaie S, Fiel MI, Schwartz M, Thung SN, Khitrov G, Zhang W, Villanueva A, Battiston C, Mazzaferro V, Bruix J, Waxman S, Friedman SL. 2006. A molecular signature to discriminate dysplastic nodules from early hepatocellular carcinoma in HCV cirrhosis. *Gastroenterology* **131(6)**:1758–1767 DOI 10.1053/j.gastro.2006.09.014.

Mateos MV, San Miguel JF. 2017. Management of multiple myeloma in the newly diagnosed patient. *Hematology* **2017(1)**:498–507 DOI 10.1182/asheducation-2017.1.498.

Niini T, Vettenranta K, Hollmén J, Larramendy ML, Aalto Y, Wikman H, Nagy B, Seppänen JK, Ferrer Salvador A, Mannila H, Saarinen-Pihkala UM, Knuutila S. 2002. Expression of myeloid-specific genes in childhood acute lymphoblastic leukemia: a cDNA array study. *Leukemia* **16(11)**:2213–2221 DOI 10.1038/sj.leu.2402685.

Nijhof IS, van de Donk NWCJ, Zweegman S, Lokhorst HM. 2018. Current and new therapeutic strategies for relapsed and refractory multiple myeloma: an update. *Drugs* **78(1)**:19–37 DOI 10.1007/s40265-017-0841-y.

Nooka AK, Kaufman JL, Hofmeister CC, Joseph NS, Heffner TL, Gupta VA, Sullivan HC, Neish AS, Dhodapkar MV, Lonial S. 2019. Daratumumab in multiple myeloma. *Cancer* **125(14)**:2364–2382 DOI 10.1002/cncr.32065.

Ocio EM, Richardson PG, Rajkumar SV, Palumbo A, Mateos MV, Orlowski R, Kumar S, Usmani S, Roodman D, Niesvizky R, Einsele H, Anderson KC, Dimopoulos MA, Avet-Loiseau H, Mellqvist UH, Turesson I, Merlini G, Schots R, McCarthy P, Bergsagel L, Chim CS, Lahuerta JJ, Shah J, Reiman A, Mikhael J, Zweegman S, Lonial S, Comenzo R, Chng WJ, Moreau P, Sonneveld P, Ludwig H, Durie BG, Miguel JF. 2014. New drugs and novel mechanisms of action in multiple myeloma in 2013: a report from the International Myeloma Working Group (IMWG). *Leukemia* **28(3)**:525–542 DOI 10.1038/leu.2013.350.

Ostendorf T, Zillinger T, Andryka K, Schlee-Guimaraes TM, Schmitz S, Marx S, Bayrak Kübra, Linke R, Salgert S, Wegner J, Grasser T, Bauersachs S, Soltesz L, Hübner MP, Nastaly M, Coch C, Kettwig M, Roehl I, Henneke M, Hoerauf A, Barchet W, Gärtner J, Schlee M, Hartmann G, Bartok E. 2020. Immune sensing of synthetic, bacterial, and protozoan RNA by toll-like receptor 8 requires coordinated processing by RNase T2 and RNase2. *Immunity* **52(4)**:591–605 DOI 10.1016/j.immuni.2020.03.009.

Qi T, Qu J, Tu C, Lu Q, Li G, Wang J, Qu Q. 2020. Super-enhancer associated five-gene risk score model predicts overall survival in multiple myeloma patients. *Frontiers in Cell and Developmental Biology* **8**:596777 DOI 10.3389/fcell.2020.596777.

Qin X, Liu Z, Yan K, Fang Z, Fan Y. 2021. Integral analysis of the RNA binding protein-associated prognostic model for renal cell carcinoma. *International Journal of Medical Sciences* **18(4)**:953–963 DOI 10.7150/ijms.50704.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43(7)**:e47 DOI 10.1093/nar/gkv007.

Schraml P, Athelogou M, Hermanns T, Huss R, Moch H. 2019. Specific immune cell and lymphatic vessel signatures identified by image analysis in renal cancer. *Modern Pathology* **32(7)**:1042–1052 DOI 10.1038/s41379-019-0214-z.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43(D1)**:D447–D452 DOI 10.1093/nar/gku1003.

Wang WS, Zhu JR, Mu ZY, Liu BY, Wang B, Chen DF, Wen ZL. 2020. Expression and clinical significance of immunogenomics in gastric adenocarcinoma. *Clinical Journal of Gastroenterology and Hepatology* **29**:507–514.

Wang L, Zhu JS, Song MQ, Chen GQ, Chen JL. 2006. Comparison of gene expression profiles between primary tumor and metastatic lesions in gastric cancer patients using laser microdissection and cDNA microarray. *World Journal of Gastroenterology* **12(43)**:6949–6954 DOI 10.3748/wjg.v12.i43.6949.

Wu M, Yuan H, Li X, Liao Q, Liu Z. 2019. Identification of a five-gene signature and establishment of a prognostic nomogram to predict progression-free interval of papillary thyroid carcinoma. *Frontiers in Endocrinology* **10**:790 DOI 10.3389/fendo.2019.00790.

Xia L, Su X, Shen J, Meng Q, Yan J, Zhang C, Chen Y, Wang H, Xu M. 2018. ANLN functions as a key candidate gene in cervical cancer as determined by integrated bioinformatic analysis. *Cancer Management and Research* **10**:663–670 DOI 10.2147/CMAR.S162813.

Xiang Y, Zhou S, Hao J, Zhong C, Ma Q, Sun Z, Wei C. 2020. Development and validation of a prognostic model for kidney renal clear cell carcinoma based on RNA binding protein expression. *Aging (Albany NY)* **12(24)**:25356–25372.

Xu X, Xu L, Huang H, Li J, Dong S, Jin L, Ma Z, Li L. 2020. Identification of hub genes as biomarkers correlated with the proliferation and prognosis in lung cancer: a weighted gene co-expression network analysis. *Biomed Research International* **2020(32)**:3416807–3416811 DOI 10.1155/2020/3416807.

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* **16(5)**:284–287 DOI 10.1089/omi.2011.0118.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* **4(1)**:17 DOI 10.2202/1544-6115.1128.

Zheng JN, Li Y, Yan YM, Shi H, Zou TT, Shao WQ, Wang Q. 2020. Identification and validation of key genes associated with systemic sclerosis-related pulmonary hypertension. *Frontiers in Genetics* **11**:816 DOI 10.3389/fgene.2020.00816.