A gene signature to predict the survival of patients with multiple myeloma

Dan Cao; Huan Liu*

*Corresponding Author: Huan Liu
Center of Infectious Diseases, West China Hospital of Sichuan University, Chengdu 610041, China.
Email: liuhuanxcw@163.com

Abstract

Background: Multiple Myeloma (MM) is an incurable hematological malignancy. Genomic studies have provided solid foundation for understanding the molecular mechanisms underlying progression of multiple myeloma. This study aimed to identify novel molecular markers associated with MM prognosis.

Methods: The GSE136337 dataset was downloaded from GEO database and taken as the training cohort. The GSE24080 and GSE57317 datasets also obtained from GEO database and were set as the external validation cohorts. In training cohort, Kaplan–Meier analysis and univariate Cox regression analyses were performed to preliminary identify prognostic genes. Multivariate Cox regression analysis was applied to build prognostic signature, which was then verified in the validation cohorts through Kaplan–Meier, Cox, and ROC analyses. Gene Set Enrichment Analysis (GSEA) and tumor immunity analysis were used to elucidate the molecular mechanisms and immune relevance. The half-maximal inhibitory concentration (IC50) of chemotherapeutic drugs in MM patients were estimated by pRRophetic algorithm.

Results: A total of fifteen genes were individually associated with overall survival via Kaplan–Meier analysis and univariate Cox regression analyses. A novel five-gene signature (CTPS1, FABP5, STK26, HECA and TMEM167B) was constructed for MM prognosis prediction. The ROC curve showed that the prognostic signature performed well at predicting overall survival in both the training cohort and the external validation cohorts. The K-M curve revealed low risk group had significantly higher survival rates compared with the high risk group. This signature was further proven to be an independent prognostic factor compared to other clinic-pathological parameters via the multivariate Cox regression analysis. Gene set enrichment analysis revealed the pathways enriched in the high risk group were mainly associated with the cell proliferation and glycolysis. Tumor-infiltrating cells had notably differential levels of expression between high and low risk groups. Moreover, the patients in the high risk group were more sensitive to chemotherapy.
**Introduction**

Multiple Myeloma (MM) is a malignant plasma cell disease that ranks the second most common hematological tumor in the world, which is mainly characterized by abnormal amplification of malignant plasma cells in the bone marrow, resulted in excessive production of monoclonal immunoglobulins in blood and urine impaired renal function, anemia, osteolytic bone lesions, and repeated infections in patients [1-3]. The data shows that the number of new cases in the world is over 150,000 and almost 100,000 patients died from their disease in 2018 [4]. Previously, the overall 5-year survival rate of patients with multiple myeloma is less than 50%. However, with the use of proteasome inhibitors, immunomodulators and monoclonal antibodies in the past decade, the prognosis of patients with MM had improved significantly, but it still remains an incurable disease.

The researches on the molecular mechanism of MM had made important contribution to understanding the development and prognosis of MM. Abnormal expression of genes in MM, their clinical significance, biological functions and potential molecular mechanisms have also been reported. For example, high expression of *UBE2T* predicts poor prognosis and survival in MM by affecting cell division pathway [5]. The high *PTTG1* expression increased cellular proliferation and associated with poor outcomes in MM patients [6]. However, the predictive ability of a single molecular marker is limited, and a prognostic model consisting of multiple indicators is needed clinically for a comprehensive clinical evaluation of MM prognosis.

With the advance of genome-sequencing technologies, accumulating evidences shown that prognostic models combining several gene signatures had great potential in predicting overall survival of MM patients. For example, Liu et al. established nine-gene signature that predicted overall survival of MM patients [7]. Jang et al. founded 44 signature gene set through single cell RNA-Seq that accurately predicted the progression and prognosis of MM patients [8]. Deep mining of genomic data to explore novel robust gene prognostic signatures to help optimize prognosis evaluation system of MM, which is an efficient method to guide patients’ prognostic stratification and individualized treatment.

In this study, we downloaded GSE136337 dataset from Gene Expression Omnibus (GEO) database, after which we used these data to screen out prognostic-related genes and construct prognostic signature based upon prognostic-related genes. The prognostic signature was validated in GSE24080 and GSE57317 datasets. We also evaluated the relationship between prognostic signature and tumor immunity and clinical treatment. Collectively, our results suggested the prognostic signature might help effectively to predict overall survival of MM patients.
Materials and Methods

Data collection

The gene expression data and of related clinical information GSE136337 dataset was retrieved from GEO database. The GSE24080 and GSE57317 datasets were downloaded from the GEO database used as references to validate the established prognostic signature.

Construction and validation of the prognostic signature

Combined clinical information and gene expression data of GSE136337 dataset, the prognostic-related genes were screened out with P value lower than 0.0001 via univariate Cox regression analysis and Kaplan-Meier analysis. The multivariate Cox regression analysis was performed to construct a prognostic signature. The regression coefficient was obtained from multivariate Cox regression analysis and the risk score based on prognostic signature = (Coefficient_{Gene1}*expression level of Gene1) + (Coefficient_{Gene2}*expression level of Gene2) +⋯+ (Coefficient_{Genen}*expression level of Genen). Patients were divided into high-risk and low-risk groups according to the median risk score. The Kaplan–Meier Plotter was drawn to compare the survival difference in the high and low risk groups. The time Receiver Operating Characteristic (ROC) curves was applied to evaluate the predictive ability of this prognostic signature. The GSE24080 and GSE57317 datasets were used for external validation. Risk scores of patients in GSE24080 and GSE57317 datasets were calculated using the above established prognostic signature. The significance threshold was set as P value <0.05.

Identification of independent prognostic parameters of MM

To investigate whether the prognostic signature could be an independent prognostic parameter in MM, univariate and multivariate analyses were performed in GSE24080 and GSE136337 datasets on the prognostic signature and clinicopathological parameters such as age, gender, International Staging System (ISS). The significance threshold was set as P value<0.05.

Building and validating a predictive Nomogram

The independent prognostic factors selected by the multivariable Cox regression analysis were applied to construct a nomogram to a predict the probability of 1-year, 2-year, and 3-year overall survival of MM. The calibration curves were drawn to assess the concordance between actual and predicted survival.

Tumor microenvironment estimation

To further explore the relationship between the risk score and the Tumor Microenvironment (TME). The ImmuneScore, StromalScore, and ESTIMATEScore of each MM patients in GSE136337 dataset was calculated by applying the ESTIMATE algorithm in R. The relationship between the risk score and the ImmuneScore, StromalScore, and ESTIMATEScore was analysis through spearman correlation analysis. The significance threshold was set as P value <0.05.
Investigation of Tumor-Infiltrating Immune Cells (TICs)

To explore the relationship between the prognostic signature and immune-cell characteristics, the proportion of tumor-infiltrating immune subsets was analyzed using CIBERSORT algorithm among the patients of GSE136337 dataset. The differences in immune infiltrating cell content between high and low risk groups were analyzed by Wilcoxon test. The relationship between the risk score calculated by prognostic signature and the immune infiltrated cells was explored through spearman correlation analysis. The significance threshold was set as P value <0.05.

Evaluation of the sensitivity of chemotherapeutic drugs

To evaluate role of the prognostic signature in clinical treatment of patients with MM, we calculated the half-maximal inhibitory concentration (IC50) of some chemotherapeutic drugs in patients with MM in GSE136337 dataset. The difference in the IC50 between the high and low risk groups was compared by “pRRophetic” packages of R. The significance threshold was set as P value <0.05.

Results

Data collection

The flowchart of this study is shown in Figure 1. The dataset GSE136337 with 426 MM patients was taken as the training cohort. The datasets GSE24080 and GSE57317 were used as the validation cohort.

Construction of prognostic signature in training cohort

Kaplan-Meier and univariate Cox regression analysis were performed in GSE136337 dataset to screen out the prognostic-related genes. Fifteen genes significantly associated with overall survival were extracted via Kaplan-Meier and univariate Cox regression analysis, shown in Figure 2A. These genes were then subjected to multivariate Cox regression analysis to construct the prognostic signature. The five genes with P value lower than 0.05 were eventually included in this prognostic signature, and regression coefficients were calculated. The coefficient of each gene was plotted in Figure 2B.

The risk score of each patient was calculated based on the expression levels of the five genes and the risk coefficient of each gene from multivariate Cox regression analysis. The distribution of risk scores and the relationship between survival time and the risk scores of the patients in GSE136337 dataset were represented by scatterplots (Figure 3A). The median risk score was the cut-off value that divided patients into high-risk or low-risk groups. K-M survival analysis revealed the low risk group was markedly higher survival probability compared with the high risk group (Figure 3A). The area under the ROC curve (AUC) was used to assess the prognostic accuracy of this signature. The higher the AUC, the better the model performance. The AUCs for 1-, 2-, and 3-year overall survival of the prognostic signature was 0.691, 0.709 and 0.751. The maximum AUC value reached 0.751, which suggested that this prognostic signature had a moderate sensitivity and specificity (Figure 3A).
Prognostic value of the prognostic signature in validation cohorts

To further validate the efficacy of the prognostic signature in predicting overall survival in MM patients, the patients in GSE24080 and GSE57317 datasets were classified into high and low risk groups based on the same cut-off value of GSE136337 dataset, respectively. The distribution of risk scores and the relationship between survival time and the risk scores of the patients in GSE24080 and GSE57317 datasets were illustrated by scatterplots (Figure 3B & 3C). GSE24080 dataset included 559 patients with MM. The K-M survival curve revealed low risk group had significantly higher survival rates compared with the high risk group (Figure 3B). The AUCs for 1-, 2-, and 3-year of the prognostic signature was 0.620, 0.693 and 0.733 (Figure 3B), the maximum AUC value reached 0.733. GSE57317 dataset consisted of 55 MM patients. The K-M survival curve showed a higher survival probability of the low-risk group compared with the high risk group (Figure 3C). The AUCs of the prognostic signature were 0.749 in at 1 year, 0.863 at 2 years and 0.9 at 3 year (Figure 3C), the maximum AUC value reached 0.9. In general, the prognostic signature performed well at predicting overall survival in MM patients.

Correlation between risk score and clinico pathological characteristics

Univariate and multivariate Cox regression analysis were carried out to evaluate the independent predictive value of the prognostic signature. As a result, the risk score, ISS stage and age were independent prognostic factor for overall survival (Figure 4A). Consistent with the results from the GSE136337 dataset, the risk score, ISS stage and age also were independent prognostic factor for overall survival in GSE24080 cohort (Figure 4B). Then we established a nomogram including these three prognostic factors, which could predict the mortality of MM patients (Figure 5A). Every patient would gain a total point from each prognostic factor based on the nomogram. The higher total point, the higher mortality the patients were. Furthermore, Calibration curves showed a good consistency between the predictive probability of the nomogram and the actual survival rate (Figure 5B).

Gene Set Enrichment Analysis (GSEA)

To explore the molecular mechanism of the prognosis signature, GSEA compared the high and low risk groups (Figure 6). KEGG pathways enriched in the high-risk group consisted of DNA replication, cell cycle, oxidative phosphorylation, glycolysis, pentose phosphate pathway, citrate cycle TCA cycle, P53 signaling pathway, ubiquitin mediated proteolysis. KEGG pathways enriched in the low risk group included cytokine-cytokine receptor interaction. These results suggested that molecular alteration in the high-risk group was closely related to cell proliferation and metabolic reprogramming.

Correlation between risk score and Tumor Microenvironment (TME)

The ImmuneScore, StromalScore, and ESTIMATEScore of patients from the GSE136337 dataset was calculated by the ESTIMATE algorithm. The higher score estimated in ImmuneScore or StromalScore, the larger amount of the immune or stromal components in TME. ESTIMATEScore is the sum of ImmuneScore and StromalScore, which indicated the comprehensive proportion of the two components in TME. As the results of correlation analysis, ImmuneScore, StromalScore, and ESTIMATEScore had negative correlation
with risk score (Figure 7). These results implied that risk score might be a potential indicator for the status of TME.

**Correlation of risk Score with the Proportion of TICs**

To explore the correlation of risk score with the tumor immune microenvironment, we adopted CIBERSORT algorithm to analyze the proportion of tumor-infiltrating immune subsets and constructed 22 kinds of immune cell profiles in MM patients (Figure 8). Combined the results of difference and correlation analyses, the results showed that a total of four kinds of TICs were associated with risk score (Figure 9). Among them, three kinds of TICs were positively associated with risk score, including CD8+ T cells, Tregs cells and resting mast cells; while plasm cells were negatively correlated with risk score. These results further suggested that the prognostic signature affected the immune status of MM patients.

**Correlation of prognostic signature with clinical treatment**

In order to explore the application of this prognostic signature in clinical treatment, the IC50 of six chemotherapeutic drugs including bortezomib, cisplatin, doxorubicin, etoposide, lenalidomide, vorinostat were compared in high and low risk groups (Figure 10). The result showed that bortezomib, doxorubicin and etoposide had higher IC50 in low risk group. It can be indicated that the patients in the high risk group were more sensitive to the three chemotherapeutic drugs.

**Discussion**

As stated above, with the introduce of proteasome inhibitors, immunomodulators and monoclonal antibodies into clinical practice, the clinical outcome of patients with MM had improved, but the disease remained incurable [9,10]. Therefore, accurate prediction of prognosis could judge whether patients can benefit from clinical treatment. Traditional clinicopathological parameters have been used to reflect and predict disease progression. International staging system is currently the most common tool for classification and stratification of MM patient, which has been applied to predict the prognosis of MM patients [11]. However, considering the heterogeneity and complexity of this disease, it is of vital importance to developed molecular biomarkers or signature as a beneficial supplement to traditional clinicopathological parameters to further improve the accuracy of prognosis prediction. Molecular prognostic biomarkers vary from tumor progress and can dynamically reflect the prognosis of patients. In addition, they may also play crucial roles in the progression of multiple myeloma and become new therapeutic targets.

In the current study, we screened out fifteen genes associated with overall survival through Kaplan-Meier and univariate Cox regression analysis. A novel prognostic signature predicting overall survival of multiple myeloma was established via multivariate Cox regression analysis. Among these genes, **HECA** and **TMEM167B** were identified as protective genes whereas **CTPS1**, **FABP5** and **STK26** were related to poor survival. This prognostic signature also was an independent prognostic factor of multiple myeloma. Patients in the low risk group was markedly higher survival probability compared with the high risk group. Prognostic efficacy of this gene signature was validated in the GSE136337 dataset and the external datasets GSE24080 and GSE57317. A nomogram integrated with risk score and traditional clinicopathological factors was esta-
published, which accurately predicted overall survival of multiple myeloma patients. GSEA indicated that the two risk groups possessed significantly different metabolic features. The CIBERSORT algorithm and ESTIMATE algorithm were adopted for the first time in multiple myeloma to evaluate the relationship between prognostic signature and immune cell infiltration and tumor microenvironment. Moreover, for the first time that we used genes profiles to predict the sensitivity of patients to chemotherapeutic agents by pRRophetic algorithm.

There were five genes included in our prognostic signature. \textit{FABP5} can enhance the transcriptional activity of nuclear receptor peroxisome proliferator-activated receptor \( \beta/\delta \) and involve in cell proliferation and angiogenesis and migration [12-14]. \textit{FABP5} has been reported to implicate in the progress of cancer and its expression increased in abundance in multiple tumors such as hepatocellular carcinoma, clear cell renal cell carcinoma, breast cancer and prostate cancer [15-17]. In multiple myeloma, \textit{FABP5} was linked to disease progression and poor clinical outcome [18,19]. The result was consistent with our result which suggested \textit{FABP5} is risk factor in multiple myeloma. The roles of \textit{STK26}, \textit{CTPS1} and \textit{HECA} in multiple myeloma have not been reported. \textit{STK26}, also named as MST4, is a new member of the germinal center kinase III family and plays an important role in regulating cell growth, transformation, and apoptosis [20,21]. Studies had showed that \textit{STK26} functioned as a cancer-promoting gene in multiple cancers such as hepatocellular carcinoma and glioblastoma [22,23]. \textit{CTPS1} is responsible for DNA synthesis in lymphocytes [24]. Some studies have reported \textit{CTPS1} acted as a specific role in sustaining the proliferation of activated lymphocytes and might be a therapeutic target in lymphoma via inhibiting pathological T cell proliferation [24,25]. \textit{HECA} belongs to a new class of cell differentiation regulators, which regulates the proliferation and differentiation of cells [26]. It had reported that \textit{HECA} functioned as a tumor suppressor in oral squamous-cell carcinoma. In breast cancer, the high mRNA expression level of \textit{HECA} was correlated with good prognoses [27]. The biological function of \textit{TMEM167B} in cancer development have not yet been elucidated and further research is needed.

Gene set enrichment analysis revealed the difference of enrichment pathway between high and low risk groups. Pathways enriched in the high-risk group were mainly associated with the cell proliferation and glycolysis, while it was related to cytokine-cytokine receptor interaction in the low risk group. The results suggested there were different metabolic features between high and low risk groups. Our prognostic signature could be acted as a cost-effective complementary tool to predict the prognosis of MM patients from a metabolic microenvironment perspective.

Tumor microenvironment is composed of a series of cells and molecules, which provides necessary conditions for tumor growth and development. Immune cells and stromal cells are two main types of non-tumor components in tumor microenvironment, and have been proved to be valuable for cancer prognosis evaluation. The ImmuneScore or StromalScore represented the amount of the immune or stromal cells in tumor microenvironment, respectively. ESTIMATE Score is the sum of the amount of the immune or stromal cells. Our prognostic signature not only showed the differences in metabolic features, but also revealed corresponding changes in the tumor microenvironments. The risk score calculated by prognostic signature was negative associated with ImmuneScore, StromalScore, and ESTIMATEScore. Some studies have
reported that ImmuneScore and StromalScore were an important indicative to judge prognosis. Franck et al. observed that patients with higher ImmuneScore tended to have a lower recurrence rate and a better prognosis in colon cancer [28]. Wang et al. reported that high ImmuneScore predicted a prolonged overall survival, disease-free survival and distant metastasis-free survival in nasopharyngeal carcinoma patients [29]. Pu et al. founded that high stromal score was verified to be significantly associated with better overall survival in pancreatic ductal adenocarcinoma [30]. These findings were in line with that of our study. The lower ImmuneScore and StromalScore, the higher risk score and the poorer prognosis. The survival difference between high and low risk patients might be due to changes of immune components in the tumor microenvironment.

Infiltrated immune cells constitute important parts of tumor immune microenvironment and have been widely studied in recent years. In our study, CD8+ T cell, T cells regulatory and resting mast cells infiltration was significantly upregulated in high-risk group, whereas plasm cells infiltration was significantly downregulated. CD8+ T cell specifically recognize tumor-derived antigens presented by Major Histocompatibility Complex (MHC)-I and play an antitumor role. Many studies have concluded that CD8+ T cell has a favorable effect on patient survival in most types of cancers such as breast and ovarian cancers [31]. However, the contrary results were observed in clear cell renal cell carcinoma and prostate cancer, which may be due to the reason that the increase densities of CD8+T cells might be associated with more advanced tumors [31]. Our results showed that the higher level of CD8+T cells infiltration in high-risk group. Therefore, further studies are needed on whether CD8+T cells are related to the progression of multiple myeloma. The regulatory T cells are enriched within primary and metastatic tumors and play an important role in suppressing normal immune responses and promoting tumoral immune escape. A study has reported abnormal regulatory T cells activity in myeloma patients could contribute to immune dysfunction in MM, which might be the reason that the higher infiltration of regulatory T cells associated with poor prognosis [32]. Consistent with our results, higher regulatory T cells infiltration, the lower survival rate in multiple myeloma patients. Mast cells had differential prognostic impact in cancers. In prostate cancer and breast cancer, higher level of mast cells infiltration had positive prognostic value, while higher densities of they were related to reduce overall survival in bladder cancer and gastric cancer [31]. The increased mast cells infiltration has been reported to associated with lymph node invasion by tumor cells, possibly suggesting a role for them in tumor dissemination. The immune cells affected the prognosis of patients via a variety of ways. Therefore, it deserves further explore to uncover the specific mechanism of tumor immune microenvironment on prognosis.

We further explored the application of this prognostic signature in clinical treatment. For the first time, we used available molecular data of multiple myeloma to evaluate clinical drug response. Our result showed that bortezomib, doxorubicin and etoposide had higher IC50 in low risk group. Therefore, this prognosis signature could act as specific biomarkers to evaluate the efficacy of the bortezomib, doxorubicin and etoposide in the clinical treatment. This study provides a new avenue for researchers to develop drugs with high efficacy and contributes to realizing individualized medicine.

There are some limitations in our study. First, although the prognostic signature constructed in our
study using public data from GSE136337 dataset and test in GSE24080 and GSE57317 datasets have reliable robustness, this signature needed to be further verified in our own clinical data. Second, we explored the tumor immune microenvironment and molecular mechanisms in patients at different risk groups and predicted their effects of clinical treatment, the study still lacked experimental verification, while we provided a direction for studying the prognosis of multiple myeloma.

**Conclusions**

To make a long story short, our study identified and validated a novel signature including five genes (*CTPS1, FABP5, STK26, HECA* and *TMEM167B*), which could act as a prognostic predictor of MM patients. In addition, this signature proved to be independently related to overall survival in both the derivation and validation cohorts, providing insight into the prediction of MM prognosis. Furthermore, this signature could predict the sensitivity of MM patients to chemotherapeutic drugs (bortezomib, doxorubicin and etoposide) and help doctors make more rational treatment decisions.

**Declarations**

**Funding:** The authors declare that there are no sources of funding to be acknowledged.

**Author contributions:** Dan Cao: Review of the literature and draft this article. Huan Liu: Critical revision and final editing. All authors have read and agreed to the published version of the manuscript.

**Competing interests:** The authors declare that there are no conflicts of interest in this work.

**Data availability:** The raw and normalized microarray data reported in this article have been deposited in the Gene Expression Omnibus database (accession number GSE136337, GSE24080 and GSE57317)

**Consent for publication:** Not applicable.

**References**


