Am J Transl Res 2026;18(9):8501-8521
www.ajtr.org /ISSN:1943-8141/AJTR0171935

Original Article

Machine learning-based prognostic model and immune landscape analysis in multiple myeloma

Wenjing Li, Li Li, Yu Sun, Tianhua Zhao

Department of Hematology, The 980th Hospital of the People's Liberation Army Joint Logistics Support Force, Shijiazhuang 050082, Hebei, China

Received February 7, 2026; Accepted July 29, 2026; Epub September 15, 2026; Published September 30, 2026

Abstract: Objective: Multiple myeloma (MM) is a plasma cell malignancy characterized by complex cytogenetic abnormalities and a supportive bone marrow microenvironment that promotes tumor growth and drug resistance. Despite the lack of curative options, novel targeted therapies show promise in overcoming resistance and improving patient outcome. Therefore, this study aimed to identify and validate prognostic biomarkers and underlying mechanisms in MM using integrated bioinformatics, machine learning, and experimental approaches. Methods: This study used bioinformatics analyses to identify key candidate genes and molecular pathways in MM. Differentially expressed genes (DEGs) were identified from datasets (GSE5900, GSE7116, GSE146649), followed by functional enrichment, protein-protein interaction network, and survival analyses. Results: *GAS2L3* was identified as a key gene in MM, strongly associated with prognosis. A *GAS2L3*-based risk model predicted patient survival. Functional assays, including Cell Counting Kit-8, Transwell, and terminal deoxynucleotidyl transferase dUTP nick end labeling, showed that *GAS2L3* silencing suppressed cell proliferation, migration, and invasion, while promoting apoptosis in RPMI-8226 cells. Immune infiltration analysis revealed significant associations between *GAS2L3* and immune cell populations, including memory B cells and natural killer cells. Single-cell transcriptomic analysis identified the HLA-E-KLRK1 signaling axis as a key regulator of MM progression. Conclusions: This study uncovered novel molecular markers and immune regulatory pathways, providing insight into MM prognosis and possible therapeutic strategies.

Keywords: Bioinformatic analysis, multiple myeloma, prognosis, immune infiltration, machine learning

Introduction

Multiple myeloma (MM) is characterized by clonal plasma cell proliferation in the bone marrow, accounting for approximately 10% of all hematologic malignancies [1]. The uncontrolled growth of malignant plasma cells causes severe complications, including osteolytic bone disease, renal impairment, anemia, and hypercalcemia [2]. Although immunomodulatory drugs, proteasome inhibitors, and monoclonal antibodies have significantly improved outcome, MM remains largely incurable, with most patients eventually relapsing and developing drug resistance [3, 4]. Its high mortality rate and inevitable development of treatment resistance underscore the urgent need for novel prognostic biomarkers and therapeutic targets.

Recently, transcriptomic profiling has emerged as a powerful tool for dissecting MM molecular heterogeneity. Studies using microarray and RNA sequencing have identified differentially expressed genes (DEGs), constructed prognostic gene signatures, and characterized the tumor immune microenvironment [5-7]. For instance, Zhou et al. developed a GAS2 family-based prognostic model for MM patients [8]. Other studies have applied machine learning algorithms to public MM datasets to develop risk stratification models and examine associations between gene expression and clinical outcome [9-11]. Single-cell RNA sequencing (scRNA-seq) studies further clarify cellular heterogeneity and intercellular communication within the MM bone marrow microenvironment [12]. Despite these advances, several limitations persist. Many prognostic models rely on a single dataset or analytical approach, limiting generalizability across independent cohorts. Furthermore, most transcriptomic-based studies emphasize model construction without fun-

https://doi.org/10.62347/EOZA1443

ctional experimental validation, leaving the biologic relevance of identified biomarkers largely unclear [13-16].

To address these limitations, this study adopted an integrated analytical approach combining multi-dataset transcriptomic analysis, parallel machine learning feature selection, immune infiltration characterization, and single-cell transcriptomic profiling. Three independent Gene Expression Omnibus (GEO) datasets were selected based on the following criteria: (I) inclusion of MM and normal control samples for differential expression analysis, (II) use of purified CD138-positive plasma cells to minimize contamination from non-malignant cells, and (III) adoption of widely used Affymetrix microarray platforms to ensure cross-platform compatibility. Reanalyzing these publicly available datasets using advanced bioinformatic methods can provide novel biologic insight beyond the original studies and integrate independent cohorts to increase statistical power and result reproducibility.

In this study, DEGs were identified across the three GEO datasets, followed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses and protein-protein interaction (PPI) network construction. Four complementary machine learning algorithms (LASSO, SVM-RFE, Boruta, and XGBoost) were applied in parallel to identify high-confidence prognostic genes, thereby reducing the risk of model overfitting associated with single-method approaches. A Cox regression-based prognostic risk model was then constructed and validated in an independent cohort. CIBERSORT-based immune infiltration analysis was performed to examine associations between core genes and immune cell subsets. scRNA-seq analysis was used to characterize cellular heterogeneity within the MM bone marrow microenvironment and identify key ligand-receptor signaling axes, including HLA-E-KLRK1. The biologic function of the top candidate gene, *GAS2L3*, was validated through *in vitro* knockdown experiments in RPMI-8226 MM cells, addressing a step often omitted in computational studies. Compared to previous transcriptomic studies in MM, this study aimed to integrate multi-cohort data, multi-algorithm feature selection, single-cell analysis, and functional experimental validation within a unified framework, providing robust prognostic biomarkers and mechanistic insight into MM pathogenesis.

Materials and methods

Microarray data

Three gene expression datasets (GSE5900, GSE7116, GSE146649) comprising MM and normal control samples were obtained from NCBI GEO. GSE5900 (GPL5170) includes 12 MM and 22 normal purified plasma cell samples, GSE7116 (GPL96) includes 21 MM and 5 normal samples, and GSE146649 (GPL5188) includes 12 MM and 10 normal bone marrow plasma cell samples. All datasets contain purified CD138+ plasma cells, ensuring that transcriptomic signals reflect malignant plasma cell biology. The use of different Affymetrix platforms (Exon 1.0 ST and U133 series) enables cross-platform validation. An independent dataset, GSE24080 (565 MM patients with survival data), was reserved for prognostic model construction and validation to prevent overfitting and ensure unbiased performance assessment.

Differentially expressed gene identification

DEGs between patients with MM and normal controls in the GSE5900 and GSE7116 and GSE146649 datasets were identified using the limma package, with thresholds set at adjusted $P < 0.05$ and $|\log_2 FC| > 1$. DEGs were visualized using heatmaps and volcano plots generated using the *circlize package* and *ggplot2*, respectively [16, 17].

Differentially expressed genes related to multiple myeloma and functional enrichment analysis

A Venn diagram was used to identify overlapping DEGs across the three datasets. GO and KEGG pathway enrichment analyses were performed using *GOplot* and *ccgraph* [18], respectively. GO functions were categorized into biological process (BP), molecular function (MF), and cellular component (CC). This functional enrichment analysis was performed to determine whether DEGs are involved in MM-related BPs and pathways, providing a biologic basis for their inclusion in downstream prognostic modeling.

Establishment of protein-protein interaction network and identification of hub genes

The STRING database [18] (http://string-db.org/cgi/input.pl) was used to predict PPIs. Briefly, MM-related DEGs were uploaded to STRING, with a minimum interaction score threshold of 0.4 applied to generate the interaction network. The network was visualized using Cytoscape, and the top 10 MCC-scoring genes were identified as hub genes using the cytoHubba plug-in. Hub genes were prioritized because highly connected network nodes are more likely to play key regulatory roles and serve as prognostic biomarkers.

Development of the multiple myeloma random forest Cox score (CRFCS)

An MM-associated gene set was identified as a candidate feature set through correlation analysis in the TCGA-LIHC cohort. The *randomForestSRC* R package (rfsrc method) was used to construct a random forest model and select relevant features. A Cox regression model was then developed using these features, and regression coefficients were estimated using the survival package (coxph function).

Kaplan-Meier survival curve

In the GSE24080 dataset, the *survival* packagewas used to analyze survival based on high or low expression of the hub gene and to assess its prognostic significance across different clinical subgroups. Patients with MM were stratified into high- and low-expression groups according to median gene expression level.

Machine learning feature selection

To identify high-confidence genes from candidate genes, four complementary machine learning feature selection methods were applied in parallel to ensure robust and reliable results. First, Least Absolute Shrinkage and Selection Operator (LASSO) regression was performed using the *glmnet* package (v4.1-7) [19] with the penalty parameter set to $\alpha = 1$ (L1 regularization). A 10-fold cross-validation (cv.glmnet, nfolds = 10) was used to determine the optimal λ value. This approach shrank coefficients of non-informative genes to zero while maintaining predictive performance, yielding a sparse gene set. Next, Support Vector Machine Recursive Feature Elimination (SVM-RFE) was applied using the e1071 package [20] with a linear SVM classifier (kernel = "linear"). The classifier was trained on tumor and normal sample labels, and feature weights (|w|) were iteratively calculated after each training round. In each iteration, the bottom 10% of genes with the smallest weights were removed, and the process continued until the optimal gene subset achieved maximum accuracy under leave-one-out cross-validation. The final feature set was recorded. The XGBoost model (xgboost package, v1.7.5 [21]) was trained with default parameters (eta = 0.3, max_depth = 6, nrounds = 100). The average gain of each gene across decision trees was calculated, and the top 10% ranked based on gain were selected. The *Boruta* package (v8.0.0) [22] was then used to generate shadow features equal in number to real genes. A random forest (ntree = 1000) was iteratively trained, and Z-scores of real genes were compared to those of shadow features. Genes with Z-scores significantly exceeding the maximum shadow Z-score (labeled "Confirmed") were retained. After running the four algorithms independently, gene lists from each output were extracted and intersected using a four-set Venn diagram created using the *VennDiagram* package [23]. The intersecting genes were defined as final high-confidence prognostic biomarkers, ensuring consistent importance across multiple feature selection methods.

Establishment and evaluation of a prognostic risk model

The prognostic risk model was developed using the GSE24080 dataset, which contains survival information, based on the selected core genes (*CDH16*, *GAS2L3*, and *HSD17B6*). A Cox proportional hazards model was fitted using the survival package in R, and regression coefficients (β) were estimated for each gene. The patient risk score was calculated as:

$$\text{Risk score} = \Sigma (\beta_i \times \text{Exp}_i)$$

Where Exp_i represents normalized gene expression, and β_i is the corresponding Cox regression coefficient. Patients were stratified into high- and low-risk groups based on the median risk score. Kaplan-Meier survival curves were used to compare overall survival (OS) between groups, with significance assessed using the log-rank test. Risk score distribution and survival

status were visualized to assess model discrimination. A two-sided $P < 0.05$ was considered significant.

Immune infiltration analysis

To investigate whether the prognostic risk model reflects differences in the MM immune microenvironment, immune cell proportions in all Bulk dataset samples were estimated using the *CIBERSORT* R package [24], excluding samples with $P > 0.05$. The Wilcoxon test was used to compare the proportions of 22 immune cell subsets between disease and control groups ($P < 0.05$), and differentially abundant immune cells were identified. Pearson correlation analysis (*cor.test*, R *stats* package) was performed across all Bulk samples to evaluate associations between key genes, differentially abundant immune cells, and intercellular correlations. The results were visualized using the *pheatmap* package ($|cor| > 0.3$, $P < 0.05$) [25]. To compare infiltration levels of 28 immune cells across samples in the training set, the Wilcoxon rank-sum test was applied ($P < 0.05$). Differential immune cells were identified, and boxplots were generated using the ggplot2 package (v3.5.2) [26]. Spearman correlation analysis (*psych* package, v2.4.6.26) [27] was performed to assess associations between differentially abundant immune cells across all training samples [28]. A correlation heatmap was generated using thresholds of $|cor| > 0.3$ and $P < 0.05$ to define significance. Additionally, Spearman correlation analysis (*psych* package) was also used to evaluate the correlation between biomarkers and differentially abundant immune cells, and results were visualized using a correlation forest plot.

Single-cell data preprocessing and downstream analysis

To resolve cellular heterogeneity in the MM bone marrow microenvironment and to identify cell-type-specific gene expression patterns relevant to prognostic biomarkers, scRNA-seq data (GSE271107) were processed using R (v4.3.2) and the *Seurat* package (v4.3.0) [29]. Quality control was performed using *PercentageFeatureSet* to calculate mitochondrial content (percent.mt), number of detected genes (*nFeatureRNA*), and UMI counts (*nCountRNA*). Cells were filtered using thresholds of nFeature_RNA > 200, percent.mt ≤ 20%, and genes expressed in at least three cells. Data distributions before and after filtering were visualized using violin plots (*ggplot2*).

Data were normalized using the LogNormalize method with a scaling factor of 10,000. The *FindVariableFeature*s function was used to identify the top 2,000 highly variable genes (HVGs) based on variance-stabilizing transformation, and the LabelPoints function was applied to annotate the top 10 genes with the highest coefficient of variation on feature scatter plots.

Prior to principal component analysis (PCA), the ScaleData function was applied to center and scale the expression matrix of HVGs. The statistical significance of each principal component was evaluated using the JackStraw procedure in combination with permutation testing, with significance defined as $P < 0.05$. In addition, the ElbowPlot method was used to assist in identifying the optimal number of significant principal components. Based on these analyses, the top 20 principal components were selected for further clustering analyses.

Cell clustering was performed using the FindNeighbors and FindClusters functions based on the shared nearest neighbor graph, with the resolution parameter set to 2.0. Dimensionality reduction and data visualization were performed using the RunUMAP algorithm, which generated a two-dimensional uniform manifold approximation and projection (UMAP) plot.

For cell type annotation, cluster-specific marker genes were identified using the FindAllMarkers function with the following parameters: $\log_2$FC > 0.25, min.pct = 0.1, only.pos = TRUE. These annotations were further corrected manually using the *SingleR* package with reference to established cell marker information reported in previous studies [30]. Cell type proportions were then calculated and visualized using the *ggplot2* package.

Cell-to-cell communication analysis

Cell-to-cell communication networks were constructed using the CellChat package (v1.6.1) [29]. The standardized gene expression matrix and cell type labels were used as input data to infer the probabilities of ligand-receptor interactions. The selection criteria were set as $P \leq 0.05$ and $\log_2$ mean (ligand + receptor) ≥ 0.1.

Table 1. shRNA sequences used in the study

Target Gene	Sequence Type	Sequence (5'→3')
GAS2L3	shRNA-1	CCTCTTTGAGTCGGAAGGTTT
	shRNA-2	CGAAAGGTAAGAATACAGTTT
Scramble	sh-NC	CCTAAGGTTAAGTCGCCCTCG

For visualization, an overall intercellular communication network diagram, interaction strength heatmap, and bubble plots representing the top five signaling pathways were generated.

Cell culture

RPMI-8226 cells were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA) [31] and cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS). Cell cultures were maintained at 37°C in a humidified incubator containing 5% CO_2.

Quantitative reverse transcription polymerase chain reaction

TRIzol (Thermo Fisher Scientific) was used to extract total RNA from cells and tissues. The isolated total RNA was reverse-transcribed with PrimeScript™ RT Master Mix (Takara, Kusatsu, Japan). Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed using TB Green™ Premix Ex Taq™ II (Takara) on a QuantStudio 5 real-time PCR system (Applied Biosystems, Warrington, UK) to quantify the expression levels of lncRNAs and mRNAs. The endogenous control GAPDH was used to normalize the expression of lncRNAs or mRNAs. Subcellular fractionation of cytoplasmic and nuclear RNA was performed using the PARIS Kit (Thermo Fisher Scientific, Waltham, MA, USA) following the protocols of the manufacturer. A qRT-PCR assay was subsequently used to detect the expression of *GAS2L3* in the cytoplasm and nucleus. The results were normalized to those of *GAPDH* and U6. The $2^{-\Delta\Delta CT}$ method was used to calculate the relative expression. **Table 1** shows all specific PCR and reverse transcription primer sequences.

Cell proliferation assay

MM cells were seeded into 96-well plates at a density of 3×10^3 cells/well and incubated at 37°C in a CO_2 incubator. Cell viability was assessed at 24-h intervals. At each time point, 10 μL of Cell Counting Kit-8 (CCK-8, Dojindo, Kumamoto, Japan) reagent was added to each well, followed by incubation at 37°C for 3 h. Absorbance was subsequently measured at 450 nm using a microplate reader to determine cell viability.

Transwell assay

Cell migration and invasion assays were performed using 8-μm Transwell cell culture chamber filters (Corning, Cambridge, MA, USA). Prior to the invasion assay, the chamber membranes were precoated with Matrigel. Cells were resuspended in serum-free medium and adjusted to a density of 5×10^4 cells/mL before being seeded into the upper chamber. The lower chambers were filled with culture medium supplemented with 20% FBS. After incubation for 36-48 h, the cells were fixed with 4% paraformaldehyde and stained with 0.1% crystal violet. The cells in the upper chamber were then removed, and the other cells were photographed using an inverted microscope (CKX53, Olympus, Japan). Cell numbers were quantified using ImageJ software [32, 33].

Terminal deoxynucleotidyl transferase-mediated deoxyuridine triphosphate nick-end labeling assays

Terminal deoxynucleotidyl transferase-mediated deoxyuridine triphosphate nick-end labeling (TUNEL) assays were performed using the In Situ Cell Death Detection Kit, AP (Roche Molecular Biochemicals, IN, USA) following the instructions of the manufacturer. For each slide, at least 300 cells were evaluated from three randomly selected microscopic fields. Images were captured using an inverted microscope (model IV/00, Leica Microsystems, Bannockburn, IL, USA) equipped with a 40×/0.17 NA objective lens and a Spot Insight digital camera (model 3.2.0, Diagnostic Instruments, Sterling Heights, MI, USA). Image acquisition and processing were performed using PAX-IT v.5 (PAXcam, Franklin Park, IL, USA) and Adobe Photoshop CS software (Adobe, San Jose, CA, USA), respectively.

Statistical analysis

All statistical analyses were performed using the SPSS software version 19.0 (SPSS Inc., Chicago, IL, USA). Data were presented as mean ± standard deviation (SD). Statistical sig-

nificance was evaluated using either a two-tailed paired Student's *t* test or repeated measures analysis of variance (ANOVA). A *P* value < 0.05 was considered significant.

Results

Differentially expressed gene screening analysis

DEGs between patients with MM and healthy donors were identified using the GEO databases (GSE5900, GSE7116, and GSE146649). Based on the screening results, volcano plots and heatmaps were generated to visualize the top 20 DEGs in each dataset. In the GSE5900 dataset, 1680 downregulated genes and 5580 upregulated genes were identified. This dataset includes 12 samples from patients with MM and 22 samples from healthy donors. Among the significantly upregulated genes in the disease group were NUDT22, KCNS3, KIT, and HVCN1, whereas PF4V1, KRT35, MAGEB18, and MBD3L2 were markedly downregulated in the disease group (**Figure 1A**, **1B**). In the GSE7116 dataset, 568 upregulated genes and 1068 downregulated genes were identified. This dataset included 21 samples from patients with MM and 5 samples from healthy donors. Genes such as CXCL8, EPB42, LTF, and SLC4A3 showed significantly higher expression in the disease group, while genes including CSF3, DSC1, NOG, and HOXB2 were significantly downregulated (**Figure 1C**, **1D**). Additionally, in the GSE146649 dataset, 805 upregulated genes and 620 downregulated genes were identified. This dataset included 12 samples from patients with MM and 10 samples from healthy donors. Genes, including ITGAV, GREM1, and TMEM200A, showed significant upregulation in the disease group, while genes such as EMX2, NFE2, and PITX1 were significantly downregulated (**Figure 1E**, **1F**). These results demonstrated distinct gene expression patterns between healthy donors and patients with MM, and enabled clear visualization of DEGs across the analyzed datasets.

Functional annotation of multiple myeloma-related differentially expressed genes and identification of hub genes

Initially, a Venn intersection analysis was performed across the three datasets mentioned above. The results identified 63 DEGs (**Figure 2A**). Subsequently, GO and KEGG enrichment analyses were conducted on the MM-related DEGs. The results showed that the enriched BP terms were mainly related to olefin compound metabolism, leukocyte activation involved in immune responses, and cell activation in immune responses. The enriched CC terms included centriolar satellite structure, flagellum/caput (spermatogenesis-related structures), and U2-type catalytic pre-spliceosome. The enriched MF terms included catalytic activity acting on DNA, endonuclease activity, and heme binding (**Figure 2B**, **2C**). Additionally, KEGG enrichment analysis revealed multiple signaling pathways associated with MM-related DEGs (**Figure 2D**, **2E**), including keratinization membrane formation and Shigella infection pathways. Furthermore, a PPI network was constructed using the STRING database to illustrate interaction relationships among the encoded proteins (**Figure 2F**). These enrichment and network analyses provided essential biological context for the identified DEGs and served as a prioritization step for selecting candidate genes for subsequent machine learning-based prognostic feature selection.

Independent prognostic significance of hub genes and survival analysis under different expression states

To determine the prognostic significance of DEGs in MM, univariate Cox proportional hazards regression analysis was performed using the GSE24080 dataset [11] to assess the relationship between the gene expression levels and OS. The analysis results are presented as a forest plot (**Figure 3A**). As shown in the figure, elevated expression of *RRN3*, *ZNF232*, *GAS2L3*, *TWNK*, *EXO1*, and *HSD17B6* was significantly associated with poor OS, acting as risk factors (hazard ratio [HR] > 1, P < 0.05). Among them, *RRN3* (HR = 1.43, 95% confidence interval [CI]: 1.097-1.866, P = 0.0083), *ZNF232* (HR = 1.3, 95% CI: 1.02-1.656, P = 0.0341), and *GAS2L3* (HR = 1.28, 95% CI: 1.041-1.575, P = 0.0195) demonstrated particularly strong associations with adverse survival outcome. In contrast, high expression of *RNASE3*, *SPAG17*, *PGRMC2*, and *CDH16* was significantly associated with improved OS, suggesting a protective role (HR < 1, P < 0.05). Among these, CDH16 (HR = 0.744, 95% CI: 0.567-0.975, P = 0.0323) showed the most pronounced protective effect. Finally, based on the Cox proportional hazards regression

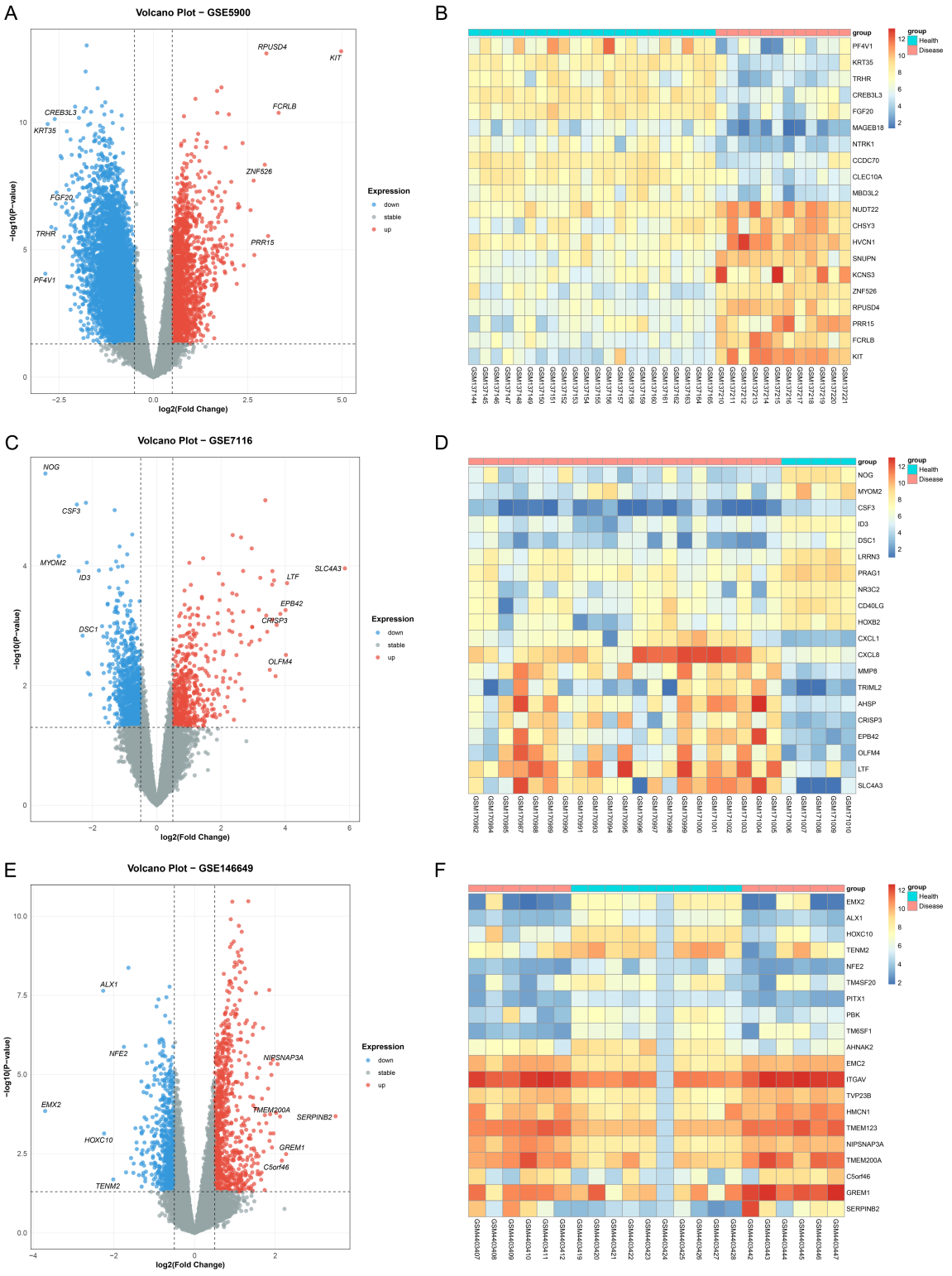


Figure 1. DEGs screening for differentially expressed genes. A, B. Volcano plot and heatmap of DEGs in the GSE5900 dataset. C, D. Volcano plot and heatmap of DEGs in the GSE7116 dataset. E, F. Volcano plot and heatmap of DEGs in the GSE146649 dataset.

results, 10 hub genes were ultimately selected. Based on the median expression level of each gene, the hub genes were stratified into high-expression and low-expression groups, fol-

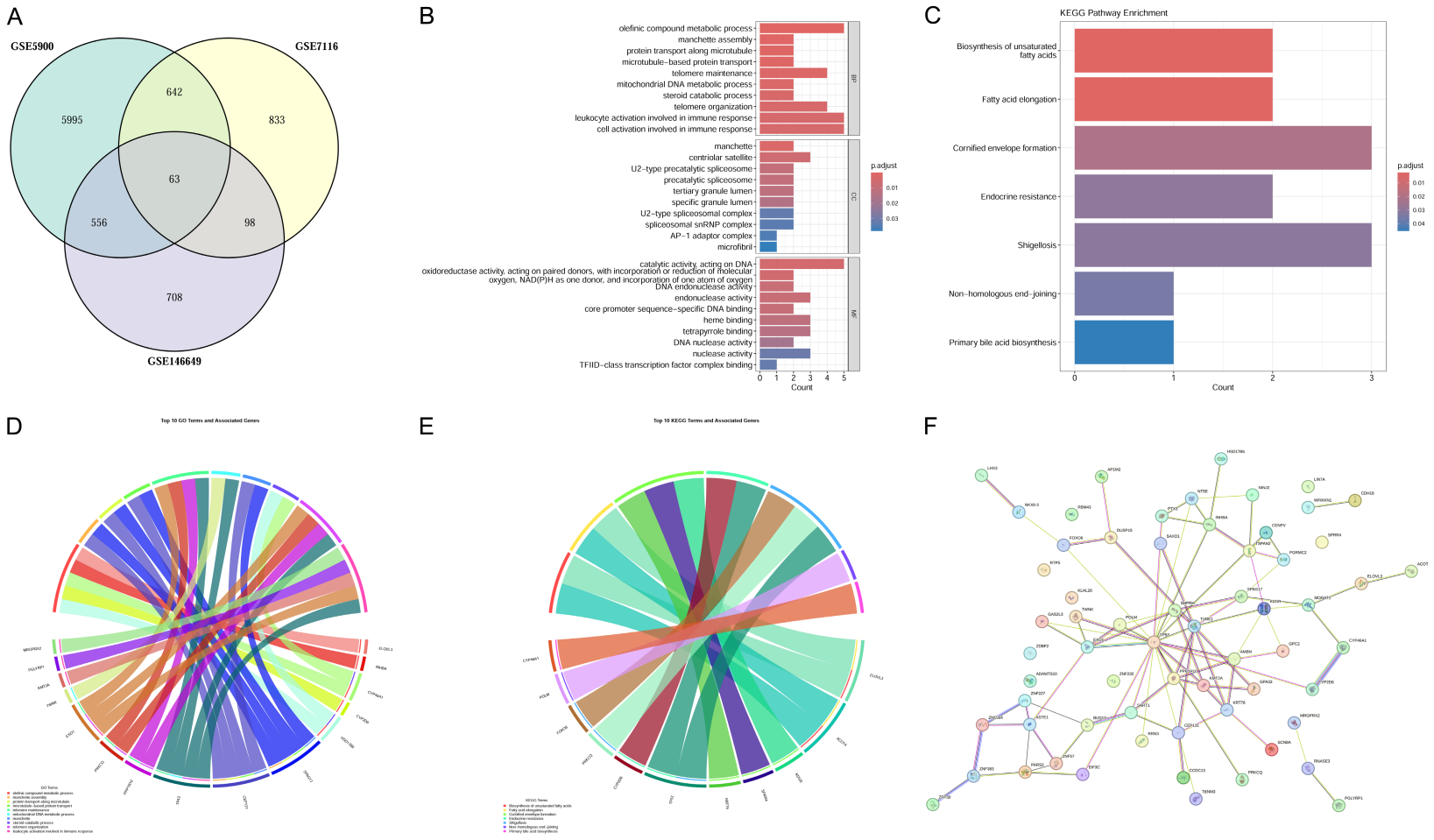


Figure 2. Functional annotation of mm-related DEGs and identification of hub genes. A. Venn plot of DEGs from the GSE5900, GSE7116, and GSE146649 datasets. B. GO analysis bar plot. C. KEGG analysis bar plot. D. Chord plot of GO. E. Chord plot of KEGG. F. PPI protein interaction network diagram.

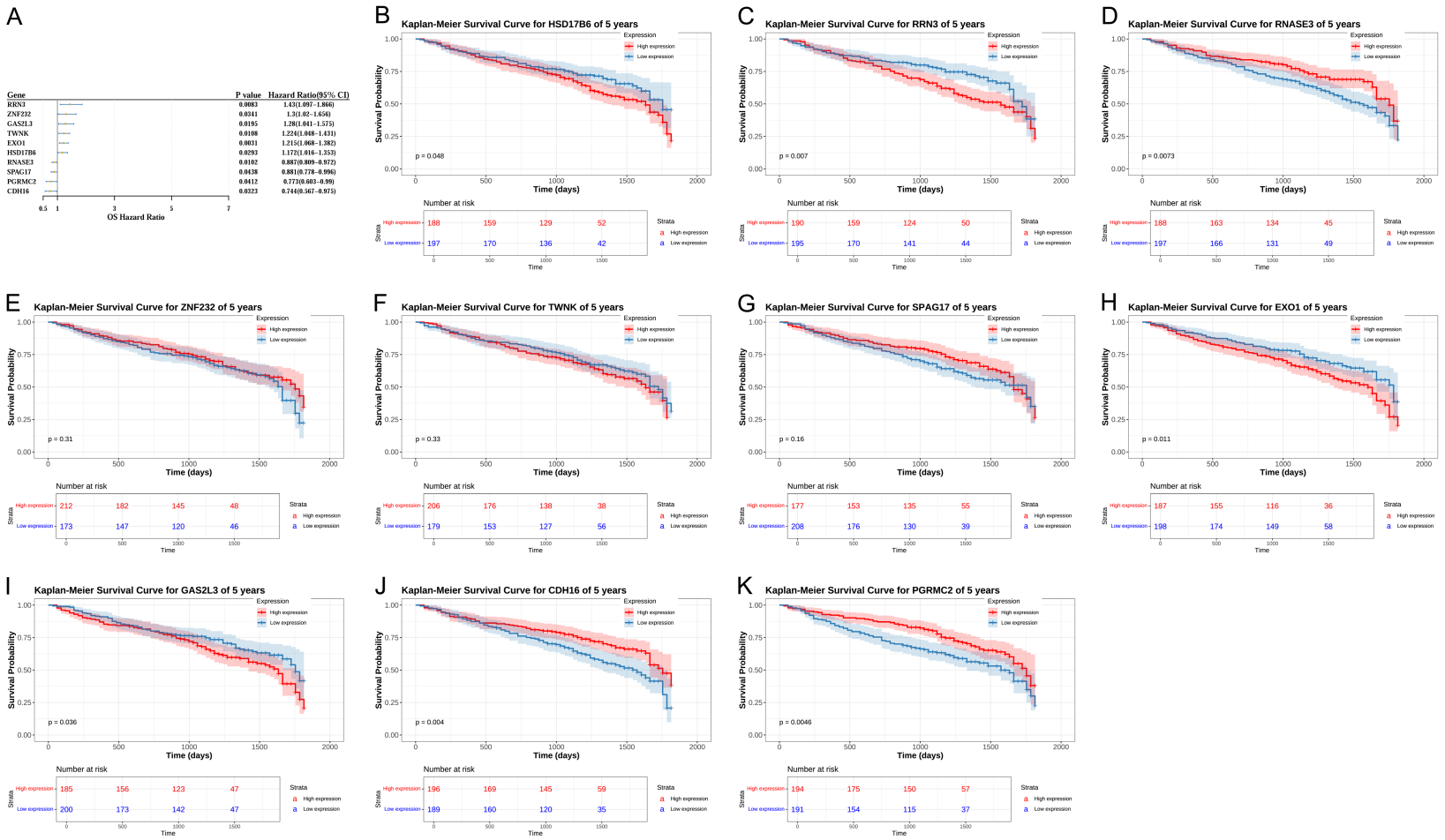


Figure 3. Independent prognostic significance of hub genes and survival analysis under different expression states. A. Forest plot of Cox proportional hazards regression analysis. B-K. 5-year Kaplan-Meier curves for the 10 hub genes, including HSD17B6, RRN3, and RNASE3.

lowed by a 5-year survival analysis. The results showed that only *CDH16*, *EXO1*, *GAS2L3*, *HSD17B6*, *PGRMC2*, *RNASE3*, and *RRN3* exhibited significant differences in prognostic outcome between the high-expression and low-expression groups ($P < 0.05$, **Figure 3B-K**).

Machine learning identification of optimal genes and construction of the risk model

The combined results of four machine learning algorithms were used to identify the core genes in this study. Using LASSO regression to simplify the MM module genes, seven potential prognostic core biomarkers for MM were found (**Figure 4A**, **4B**). The SVM-RFE algorithm identified five genes as important biomarkers (**Figure 4C**), while the Boruta algorithm screened out five core biomarkers (**Figure 4D**). In addition, the XGBoost algorithm identified six core biomarkers (**Figure 4E**). Furthermore, a Venn intersection analysis was performed to identify the most central biomarkers related to the prognosis of MM. The results showed that the four machine learning algorithms collectively identified three core genes as the central genes for this study: *CDH16*, *GAS2L3*, and *HSD17B6* (**Figure 4F**). Based on the Cox proportional hazards regression model and the results obtained from the four machine learning algorithms, a prognostic risk model was constructed. Patients were stratified into high-risk and low-risk groups based on the median risk score. The risk score distribution clearly showed the distinction between the two groups. Survival analysis results showed that patients in the high-risk group exhibited significantly higher mortality rates and shorter survival durations (**Figure 4G**). Additionally, Kaplan-Meier survival curve analysis revealed significant differences in survival outcome between the high- and low-expression groups ($P < 0.0001$, **Figure 4H**).

Immune infiltration analysis of the prognostic risk model

Following the establishment of the prognostic risk model, immune infiltration analysis was subsequently performed based on this model. The heatmap results showed significant differences in the proportions of most immune cell subpopulations between the high-risk and low-risk groups (**Figure 5A**). Furthermore, boxplot analysis based on the Wilcoxon rank-sum test demonstrated that the infiltration levels of several immune cell subpopulations, including activated $CD4^+$ T cells, memory B cells, natural killer (NK) cells, Th2 T cells, effector memory $CD8^+$ T cells, activated $CD8^+$ T cells, and central memory $CD8^+$ T cells, differed significantly between the high-risk and low-risk groups. These results showed that these immune cell subpopulations may play an important immunoregulatory role in the development and progression of MM (**Figure 5B**). Correlation heatmap analysis of immune-infiltration cell subpopulations showed strong correlations among immune cell subpopulations that differed significantly between the high-risk and low-risk groups. Specifically, activated $CD4^+$ T cells showed significant positive correlations with Th2 immune subpopulations, effector memory $CD4^+$ T cells, and memory B cells. Additionally, NK cells were significantly positively correlated with effector memory $CD8^+$ T cells. These results showed that alterations in the immune microenvironment were closely related to disease progression in MM (**Figure 5C**).

Furthermore, immune correlation analysis was performed on the three differential genes identified through machine learning. The results are presented as a heatmap (**Figure 5D**). Among these genes, the core gene *CDH16* showed significant positive correlations with NK cells and central memory $CD8^+$ T cells, with correlation coefficients ranging from 0.2 to 0.4. In addition, CDH16 and HSD17B6 also showed a significant negative correlation with activated CD4 T cells and Memory B cell (**Figure 5E**, **5F**).

Single-cell transcriptomics revealed the cellular heterogeneity of multiple myeloma

To elucidate the disease mechanisms and alterations in the immune microenvironment of MM, single-cell transcriptomic analysis was performed using the GSE271107 dataset. This dataset comprised human bone marrow samples from patients with MM, including four MM samples and five healthy control samples, and was used to characterize the overall cellular landscape of MM. The UMAP was then applied to classify the core cells into 22 distinct clusters (**Figure 6A**). Cell-type annotation was performed using the SingleR package, the CellMarker database, and relevant literature, resulting in the identification of 10 major cell types, including B cells, myeloid progenitor cells, megakaryocyte-erythroid progenitor cells, mono-

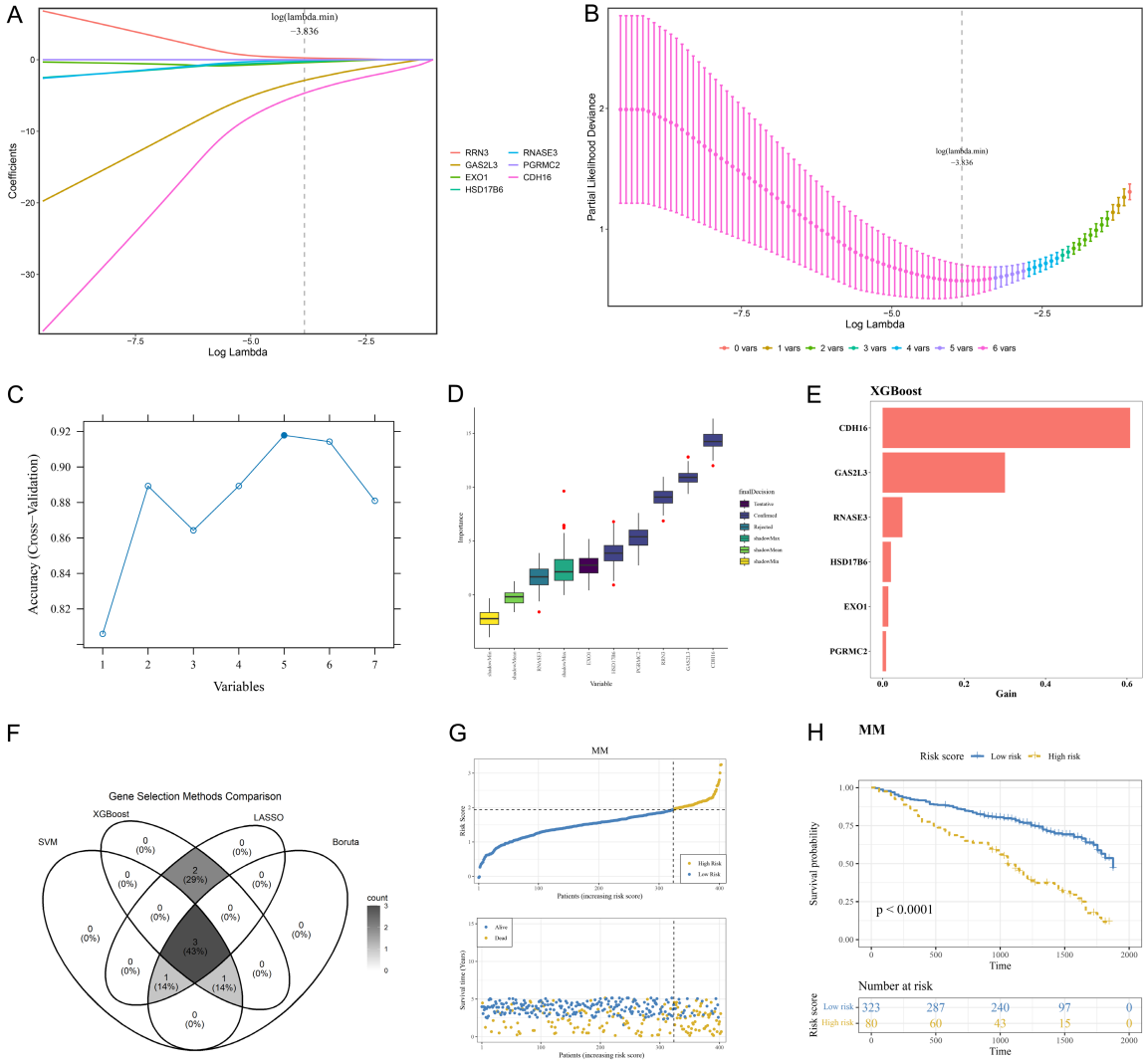


Figure 4. Machine learning identification of optimal genes and construction of the risk model. A. Cross-validation curve of the Lasso model. B. Lasso coefficient path diagram. C. SVM-RFE algorithm screening of prognostic genes. D. Boruta algorithm screening of prognostic differential genes. E. XGBoost algorithm screening of prognostic differential genes. F. Venn intersection analysis of the four machine learning algorithms. G. Construction of the risk model based on machine learning screening results. H. Kaplan-Meier curve verification of high-risk and low-risk groups in the risk model.

cytes, NK cells, platelets, pre-B lymphocytes (CD34-negative), pre-B lymphocytes (CD34-positive), T cells, and tissue stem cells. Among these, B cells were represented by clusters 2, 8, 10, 12, and 17, the myeloid progenitor cells by cluster 11, megakaryocyte-erythroid progenitor cells by cluster 13, monocytes by clusters 1, 4, 7, 14, 15, and 19, platelets by cluster 21, NK cells by cluster 6, CD34-negative pre-B lymphocytes by cluster 3, CD34-positive pre-B lymphocytes by cluster 18, T cells by clusters 0, 5, 9, 16, and 20, and tissue stem cells by cluster 22.

Intercellular communication analysis identified key signaling axes driving immune regulation and tumor progression

In both normal and MM samples, the strength and extent of intercellular communications exhibited distinct patterns. The intensity and quantity of cell communication in the disease group are both lower than those in the normal group (**Figure 6B**). In the normal group, monocytes displayed larger node sizes and extensive connections with nearly all immune cell populations, accompanied by significant self-loop

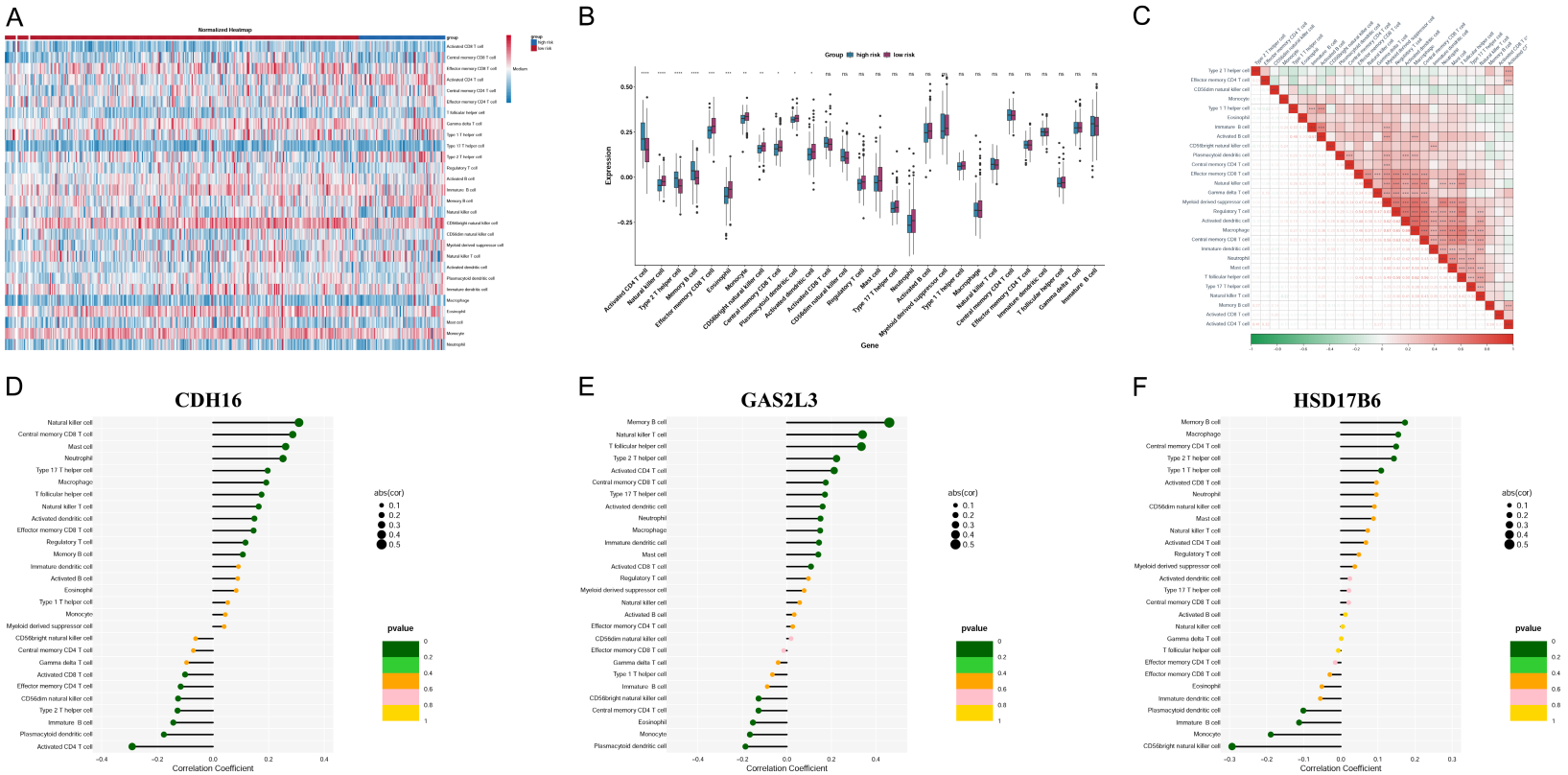
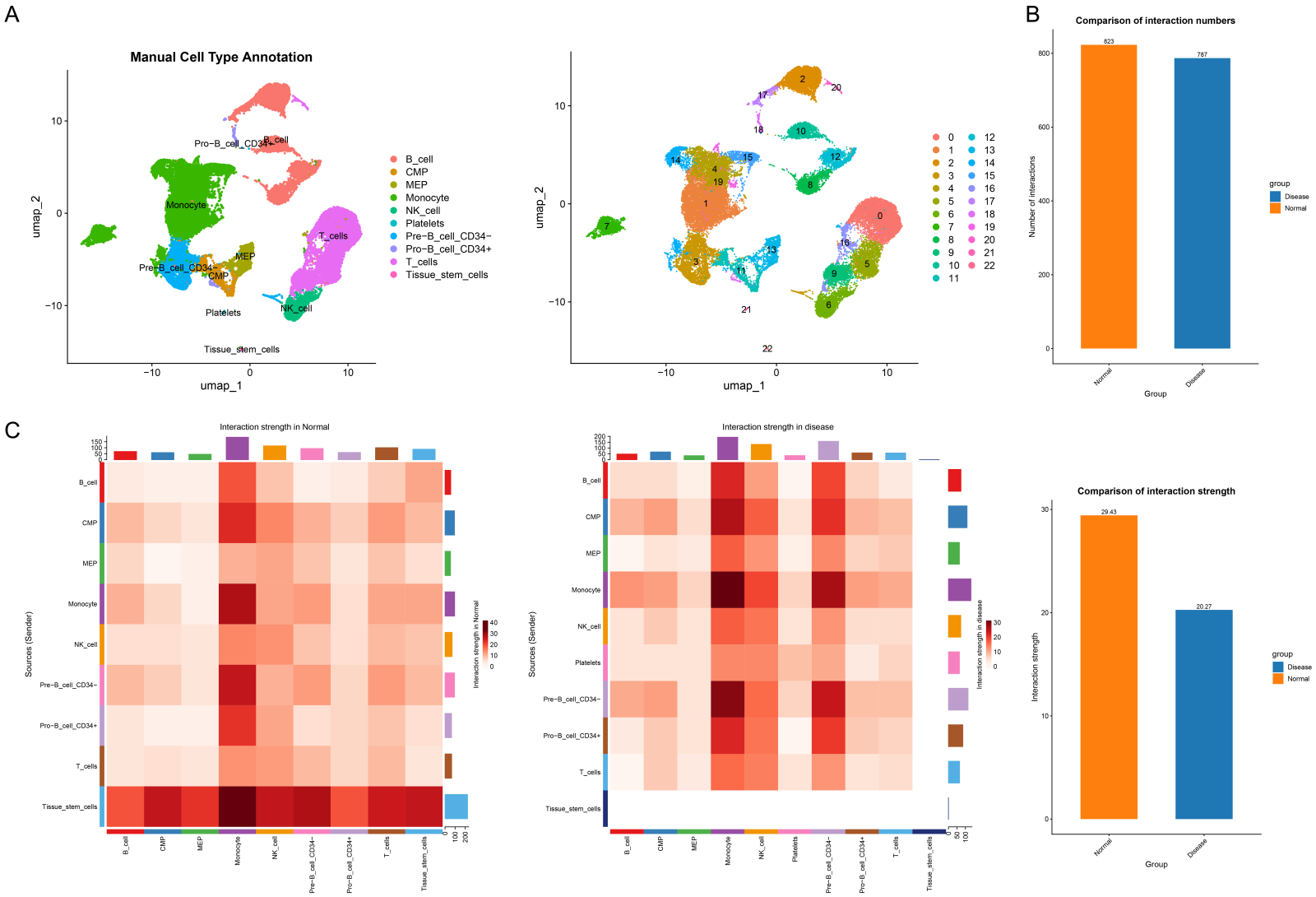


Figure 5. Immune infiltration analysis of the prognostic risk model. A. Heatmap of immune infiltration analysis. B. Boxplot from Wilcoxon rank-sum test showing the differential expression of immune cell subpopulations. C. Heatmap of correlation analysis between immune cell subpopulations. D-F. Lollipop of correlation analysis between the 3 core genes selected by machine learning and immune cell subpopulations.


A
Manual Cell Type Annotation
B_cell
CMP
MEP
Monocyte
NK_cell
Platelets
Pre−B_cell_CD34−
Pro−B_cell_CD34+
T_cells
Tissue_stem_cells
umap_1
umap_2
B
Comparison of interaction numbers
Number of interactions
Group
Normal
Disease
823
787
C
Interaction strength in Normal
Interaction strength in disease
Sources (Sender)
Comparison of interaction strength
Interaction strength
29.43
20.27

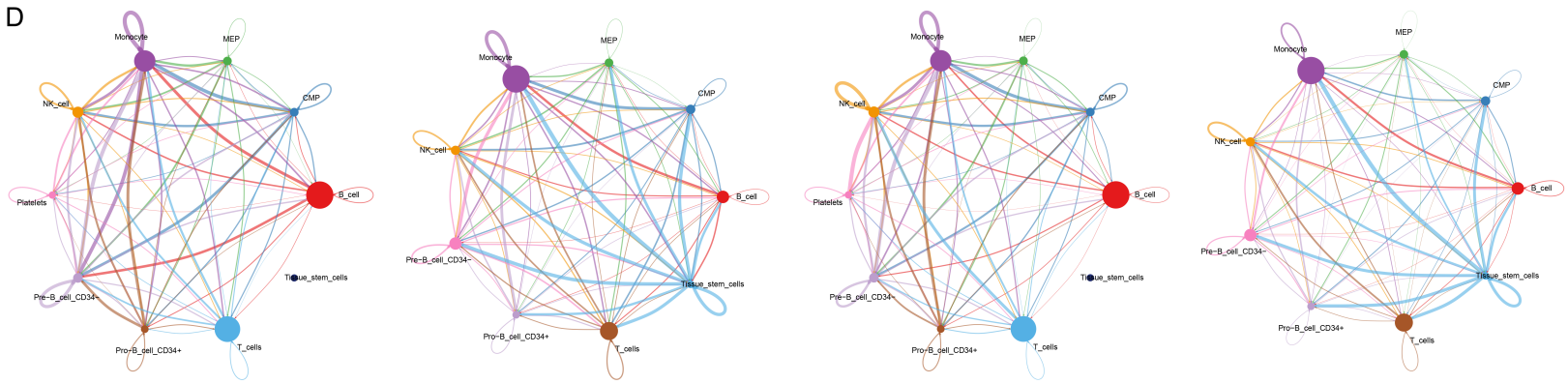


Figure 6. Single-cell transcriptomics reveals cellular heterogeneity in multiple myeloma. A. UMAP of scRNA before and after anno. B. Bar plot showing interaction counts and strengths in normal and disease groups. C. Heatmap plot showing interaction counts and strengths in normal and disease group. D. Circle plot showing interaction counts and strengths in normal and disease groups.

interactions. These findings suggest that monocytes play an organizing role and may maintain functional state through autocrine signaling. Tissue stem cells also emerged as an important signaling hub, showing widespread interactions with immune and progenitor cell groups. Substantial autocrine signaling was also observed in tissue stem cells. In the disease group, monocytes, B cells, and T cells exhibited larger node sizes and frequently served as ligands in intercellular communication. This suggests that these cell types function as major organizational hubs. Among them, monocytes and CD34-negative pre-B lymphocytes exhibited prominent self-loop interactions, indicating the potential involvement of autocrine signaling in maintaining their functional states. The results are presented as network and heatmap visualizations (**Figure 6C**, **6D**). To explore the key roles of regulatory signaling axes across different cell types in MM, an intercellular communication network was constructed. This diagram illustrates the communication probabilities of multiple ligand-receptor pairs among different cell types under normal and disease conditions. Point colors represent the communication probability between cell types, with a gradient ranging from blue (low communication probability) to red (high communication probability). These color variations reflect the communication strength of each ligand-receptor pair across different cell types. In both the normal and disease groups, the MIF-(CD74+CXCR4) and MIF-(CD74+CD44) signaling axes played crucial roles in most intercellular communications. However, in the normal group, the APP-CD74 and CXCL12-CXCR4 signaling axis exhibited higher communication probabilities in interactions between tissue stem cells and B cells. In contrast, the HLA-E-KLRK1 signaling axis was particularly prominent in the disease group, especially in platelet-NK cell interactions, T cell-NK cell interactions, and NK cell autocrine signaling, all of which exhibited higher communication probabilities (Figure S1).

Effect of GAS2L3 silencing on multiple myeloma cell proliferation, migration, invasion, and apoptosis

Based on previous machine learning analyses, risk model construction, and immune infiltration profiling, the *GAS2L3* gene was selected. In the Cox regression model, it exhibited an HR of 1.28 P=0.0195. This indicates that *GAS2L3* is a significant risk-associated gene in MM development and several immune cell subpopulations, such as Tfh cells, memory B cells, and NK T cells. However, whether *GAS2L3* modulation influences biologic behaviors related to MM remains unclear. Therefore, the role of *GAS2L3* in the biological behavior of the RPMI-8226 MMMM cell line was investigated. We first verified the knockdown efficiency using qPCR (**Figure 7A**). CCK-8 assay results showed that *GAS2L3* silencing significantly reduced the viability of RPMI-8226 cells (P < 0.0001, **Figure 7B**). Transwell assays further revealed that GAS2L3 silencing markedly suppressed the migratory and invasive capacities of RPMI-8226 cells (P < 0.001, **Figure 7C**, **7E**). Additionally, the TUNEL assay revealed that *GAS2L3* silencing significantly promoted apoptosis in RPMI-8226 cells (P < 0.001, **Figure 7D**, **7F**). These findings suggest that *GAS2L3* silencing suppresses the malignant phenotype of MM cells.

The effect of overexpressing GAS2L3 on proliferation, migration, invasion, and apoptosis in multiple myeloma

To further validate the functional role of *GAS2L3* in multiple myeloma, we established stable *GAS2L3*-overexpressing cell lines (OE-*GAS2L3*-1 and OE-*GAS2L3*-2), with an empty vector control group (OE-NC) as reference, to systematically evaluate the effects of *GAS2L3* overexpression on the malignant biologic behaviors of RPMI-8226 cells, thereby corroborating the findings from our silencing experiments. qPCR results confirmed that *GAS2L3* mRNA levels were significantly elevated in both overexpressing cell lines, verifying the successful construction of the overexpression model (P < 0.001, **Figure 8A**). CCK-8 assay results demonstrated that, compared to the OE-NC group, the proliferative viability of OE-*GAS2L3* cells was significantly enhanced (P < 0.001, **Figure 8B**). Furthermore, Transwell and TUNEL assay results revealed that overexpression of *GAS2L3* significantly promoted the migration and invasion abilities of RPMI-8226 cells relative to the OE-NC group (P < 0.05, **Figure 8C-E**), while the apoptosis rate was markedly reduced (P < 0.001, **Figure 8F**, **8G**). Taken together, the gain-of-function data from overexpression experiments corroborate the phenotypes observed in our prior knockdown studies, confirming that

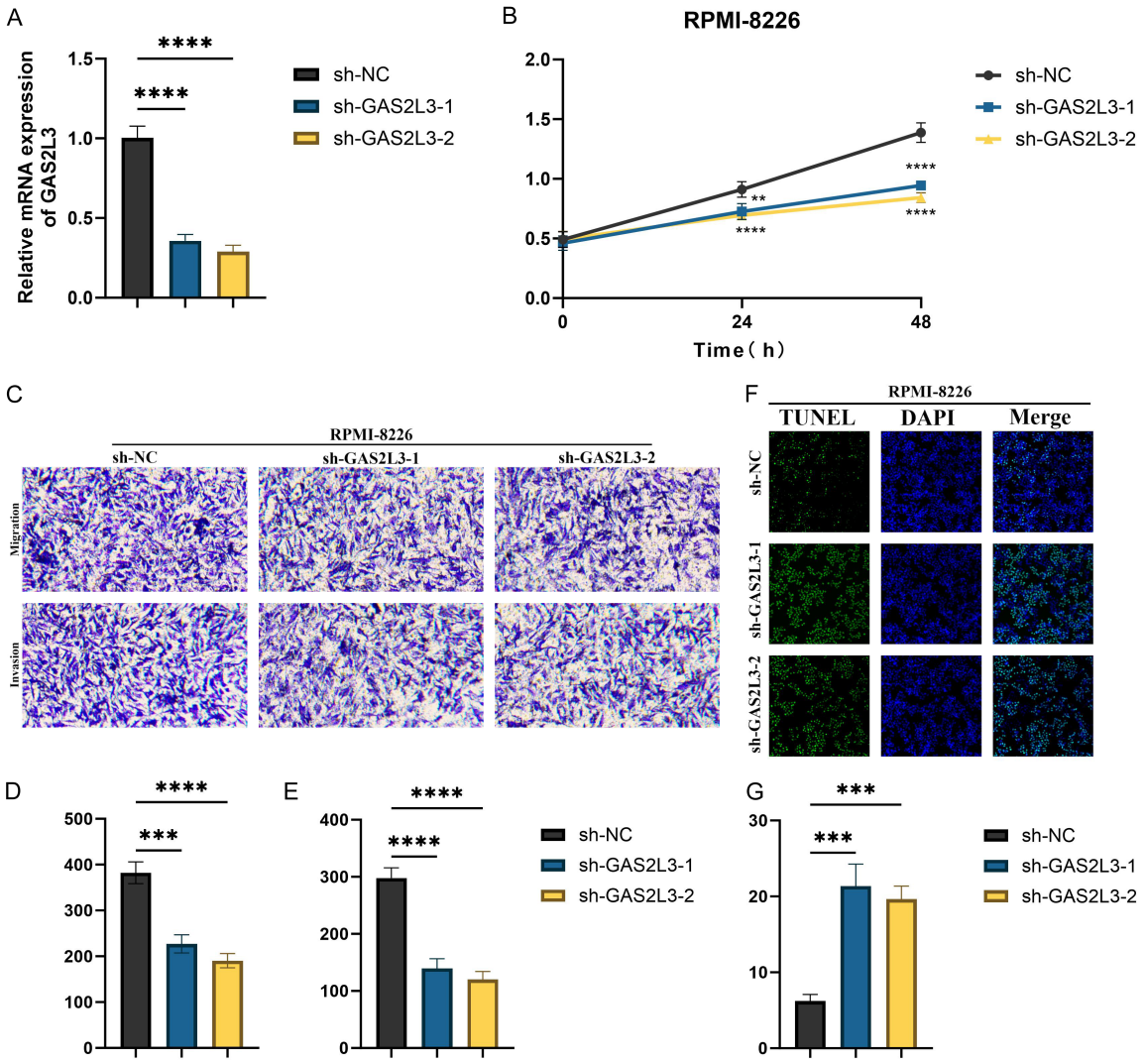


Figure 7. Effect of silencing *GAS2L3* on proliferation, migration, invasion, and apoptosis in multiple myeloma. A. qRT-PCR experiment verifying the silencing efficiency of *GAS2L3*. B. CCK-8 assay testing the effect of *GAS2L3* silencing on the viability of RPMI-8226 cells. C. Transwell assay testing the effect of *GAS2L3* silencing on the migration and invasion of RPMI-8226 cells. D, E. Quantitative statistical charts for Transwell migration and invasion. F. TUNEL assay testing the effect of *GAS2L3* silencing on apoptosis in RPMI-8226 cells. G. Quantitative statistical chart of TUNEL assay showing the apoptosis rate of RPMI-8226 cells after *GAS2L3* silencing (**P < 0.01, ***P < 0.001, ****P < 0.0001) (n = 3).

GAS2L3 functions as an oncogene in multiple myeloma by promoting proliferation, migration, and invasion while simultaneously suppressing apoptosis.

Discussion

This study employed a multi-step analytical method, with each component designed to address specific questions toward the development of a robust machine learning-based prognostic model. The initial differential expression analysis across three independent datasets revealed consistently dysregulated genes in MM. Subsequent functional enrichment and PPI network analyses provided biologic context, confirming that the identified DEGs are involved in pathways relevant to MM pathogenesis and enabling prioritization of genes with plausible mechanistic roles. The integration of four complementary machine learning algorithms (LASSO, SVM-RFE, Boruta, and XGBoost) was a deliberate strategy to mitigate limitations inherent to any single feature selection method. Only

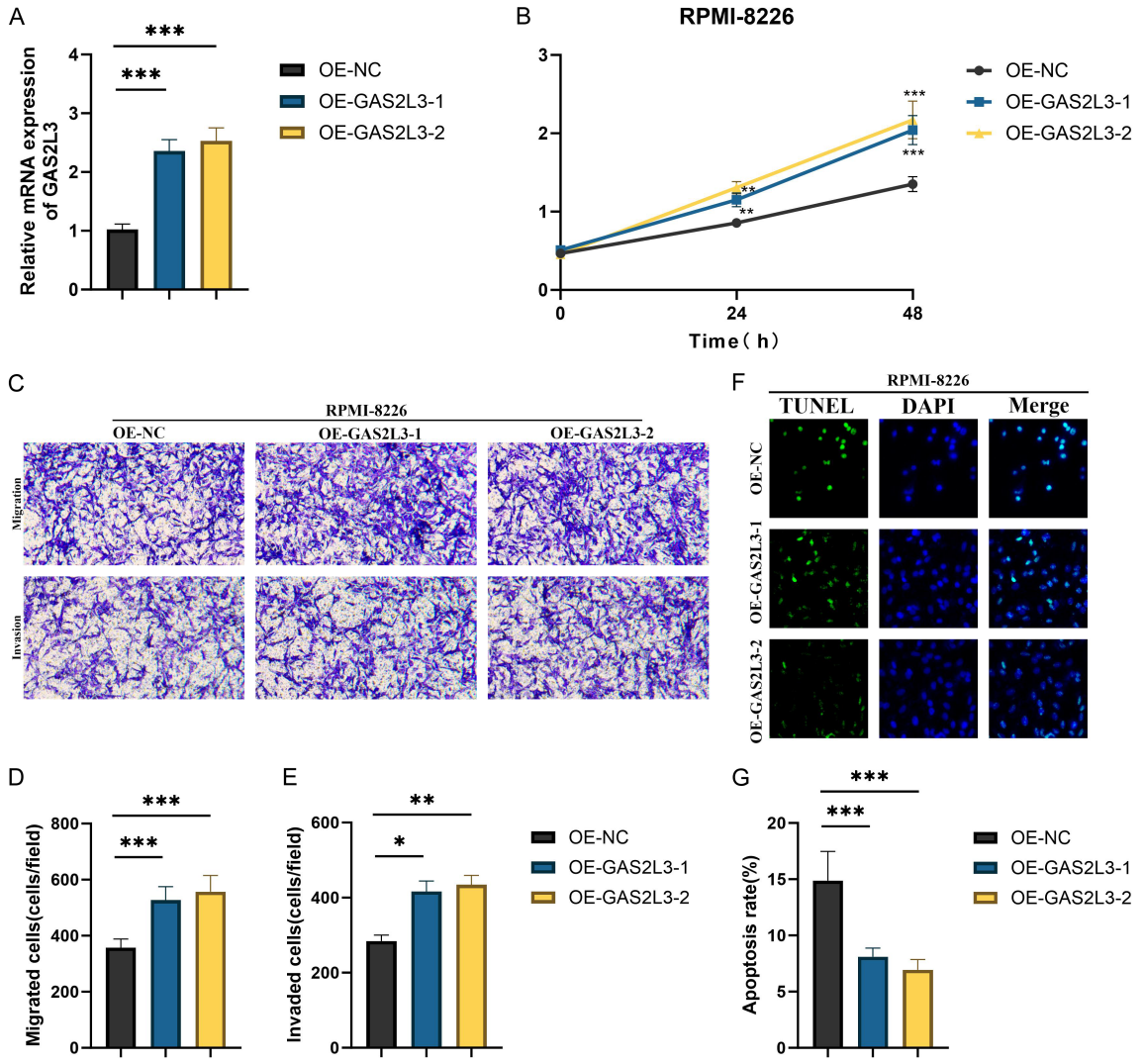


Figure 8. Effect of overexpressing *GAS2L3* on proliferation, migration, invasion, and apoptosis in multiple myeloma. A. qRT-PCR experiment verifying the overexpressing efficiency of *GAS2L3*. B. CCK-8 assay testing the effect of *GAS2L3* overexpression on the viability of RPMI-8226 cells. C. Transwell assay testing the effect of *GAS2L3* overexpression on the migration and invasion of RPMI-8226 cells. D, E. Quantitative statistical charts for Transwell migration and invasion. F. TUNEL assay testing the effect of *GAS2L3* overexpression on apoptosis in RPMI-8226 cells. G. Quantitative statistical chart of TUNEL assay showing the apoptosis rate of RPMI-8226 cells after *GAS2L3* silencing (**P < 0.01, ***P < 0.001, ****P < 0.0001) (n).

genes consistently identified across all four algorithms were retained as core biomarkers, thereby reducing overfitting risk. Immune infiltration analysis was performed to determine whether the prognostic model captured immunologically relevant differences between risk groups, thereby linking the gene signature to the tumor microenvironment. Single-cell transcriptomic analysis further resolved the cellular origin of the prognostic signals and identified intercellular communication axes potentially driving immune evasion. Finally, *in vitro* functional experiments provided biological validation of the top candidate gene, *GAS2L3*, confirming the functional relevance of the computational predictions in MM cells. Each analytical step, therefore, served a distinct and crucial role within the overall study design. Additional panels (e.g., individual GO/KEGG enrichment dot plots, correlation heatmaps for immune cell subpopulations, and detailed intercellular communication maps) are shown as supplementary materials to maintain clarity and focus on the main figures.

MM is a malignant tumor that remains largely incurable in clinical practice, with its incidence and prevalence increasing annually [34]. This poses new challenges for disease management. A defining feature of MM is the accumulation of malignant plasma cells in the bone marrow [35]. These cells originate from B cells through an antigen-driven maturation process involving germinal center and memory B cell stages [36]. Therefore, changes in B cells and memory B cells are inevitably implicated in disease development. The single-cell transcriptomic analysis in this study supports this perspective. In this study, gene expression datasets (GSE5900, GSE7116, and GSE146649) were selected for analysis. These datasets comprise mainly purified CD138-positive plasma cells, making them suitable for exploring the intrinsic transcriptional characteristics of MM. Numerous studies show that targeting CD138-positive plasma cells plays a crucial role in MM therapy [37]. For example, activation of the Dll1-Notch signaling pathway can promote the proliferation of CD138-positive cells and accelerate MM progression [38]. In addition, targeted nanomedicines and monoclonal antibodies, such as VIS832, show strong therapeutic potential [39].

Based on the Cox proportional hazards regression model and machine learning analyses, three core genes most relevant to MM were identified: *CDH16*, *GAS2L3*, and *HSD17B6*. Among these, *GAS2L3* exhibited an HR of 1.28 in the Cox model, representing the highest risk ratio among prognostic genes, and the strongest association with poor prognosis in patients with MM. *GAS2L3* belongs to the GAS2 family, whose members are involved primarily in cross-linking actin and microtubule filaments during interphase and in growth-arrested cells. As a newly identified cytoskeleton-related protein, *GAS2L3* differs from other GAS2 family members because it is specifically expressed during mitosis. It localizes to the mitotic spindle during metaphase, to the midbody during cytokinesis, and to the constriction site during cell abscission [20, 21]. Studies show that *GAS2L3* loss in cardiomyocytes leads to heart failure and death in mice [40] and that its silencing inhibits tumor progression in hepatocellular carcinoma [41]. However, the role of *GAS2L3* in multiple myeloma has not been reported. Other studies show that the p53 target gene Necdin regulates the response of hematopoietic stem cells to genotoxic stress through cell cycle-dependent and cell cycle-independent mechanisms in a *GAS2L3*-dependent manner. In this regulatory context, Necdin functions as a molecular switch in adult hematopoiesis and, as a p53 target gene, promotes hematopoietic stem cell quiescence under steady state conditions in a p53-like manner [42]. In this process, *GAS2L3* also plays an essential role [4]. MM does not originate from hematopoietic stem cells, but its development is closely related to the bone marrow microenvironment where these cells reside, and can profoundly influence their fate [43]. During disease progression, rapid tumor cell proliferation leads to occupation of the hematopoietic stem cell niche and suppression of normal hematopoiesis. This mechanism explains why patients with MM frequently develop anemia and immune dysfunction [44]. During this process, tumor cells secrete large amounts of pro-inflammatory factors that can inhibit the normal differentiation of hematopoietic stem cells. This alteration indirectly leads to reduced function or decreased numbers of NK cells. Therefore, immune infiltration analysis revealed a strong correlation between *GAS2L3* expression and memory B and NK cells. These findings provide evidence for the important regulatory role of *GAS2L3* in the MM immune microenvironment. Furthermore, this study experimentally confirmed that *GAS2L3* silencing inhibits the malignant behavior of MM cells.

In addition, this study showed that the HLA-E-KLRK1 signaling axis exhibited high communication probabilities in platelet-NK cell interactions, T cell-NK cell interactions, and NK cell autocrine signaling. These findings indicate that the HLA-E-KLRK1 signaling axis plays an important role in immune evasion in MM. Studies show that NK cells and macrophages may contribute to chemotherapy resistance in acute myeloid leukemia, and that the HLA-E-KLRK1 signaling axis is among the major ligand-receptor pairs mediating communication between these two cell types [45]. In head and neck squamous cell carcinoma, PD-1 blockade therapy was associated with significant changes in regulatory T cells and several ligand-receptor pairs, including HLA-E-KLRK1 signaling axis. Similarly, clinical studies of *Cutibacterium acnes* infection report significant alterations in HLA-E-KLRK1-mediated cell communication

among keratinocytes, muscle cells, epithelial cells, endothelial cells, and monocytes [46]. Therefore, these findings suggest that the HLA-E-KLRK1 signaling axis may represent a promising therapeutic target for overcoming chemotherapy resistance and advancing immunotherapeutic strategy in MM.

In summary, this study showed that *GAS2L3* has value for prognostic assessment in MM and may serve as a new candidate target for therapeutic intervention. Furthermore, investigation of the HLA-E-KLRK1 signaling axis provides new insights into the immune evasion mechanisms underlying MM. In addition, immune infiltration analysis suggested roles of NK and memory B cells in chemotherapy resistance, offering further insight into the contribution of immune cells to MM treatment.

Acknowledgements

This study was supported by Medical Science Research Project of Hebei (No. 20231308).

Disclosure of conflict of interest

None.

Address correspondence to: Wenjing Li, Department of Hematology, The 980th Hospital of the People's Liberation Army Joint Logistics Support Force, No. 398 Zhongshan West Road, Shijiazhuang 050082, Hebei, China. E-mail: m2vendy@163.com

References

[1] Cowan AJ, Green DJ, Kwok M, Lee S, Coffey DG, Holmberg LA, Tuazon S, Gopal AK and Libby EN. Diagnosis and management of multiple myeloma: a review. JAMA 2022; 327: 464-477.

[2] Rajkumar SV. Multiple myeloma: 2022 update on diagnosis, risk stratification, and management. Am J Hematol 2022; 97: 1086-1107.

[3] Kumar SK, Rajkumar V, Kyle RA, van Duin M, Sonneveld P, Mateos MV, Gay F and Anderson KC. Multiple myeloma. Nat Rev Dis Primers 2017; 3: 17046.

[4] Zhou Y, Zhang L, Song S, Xu L, Yan Y, Wu H, Tong X and Yan H. Elevated GAS2L3 expression correlates with poor prognosis in patients with glioma: a study based on bioinformatics and immunohistochemical analysis. Front Genet 2021; 12: 649270.

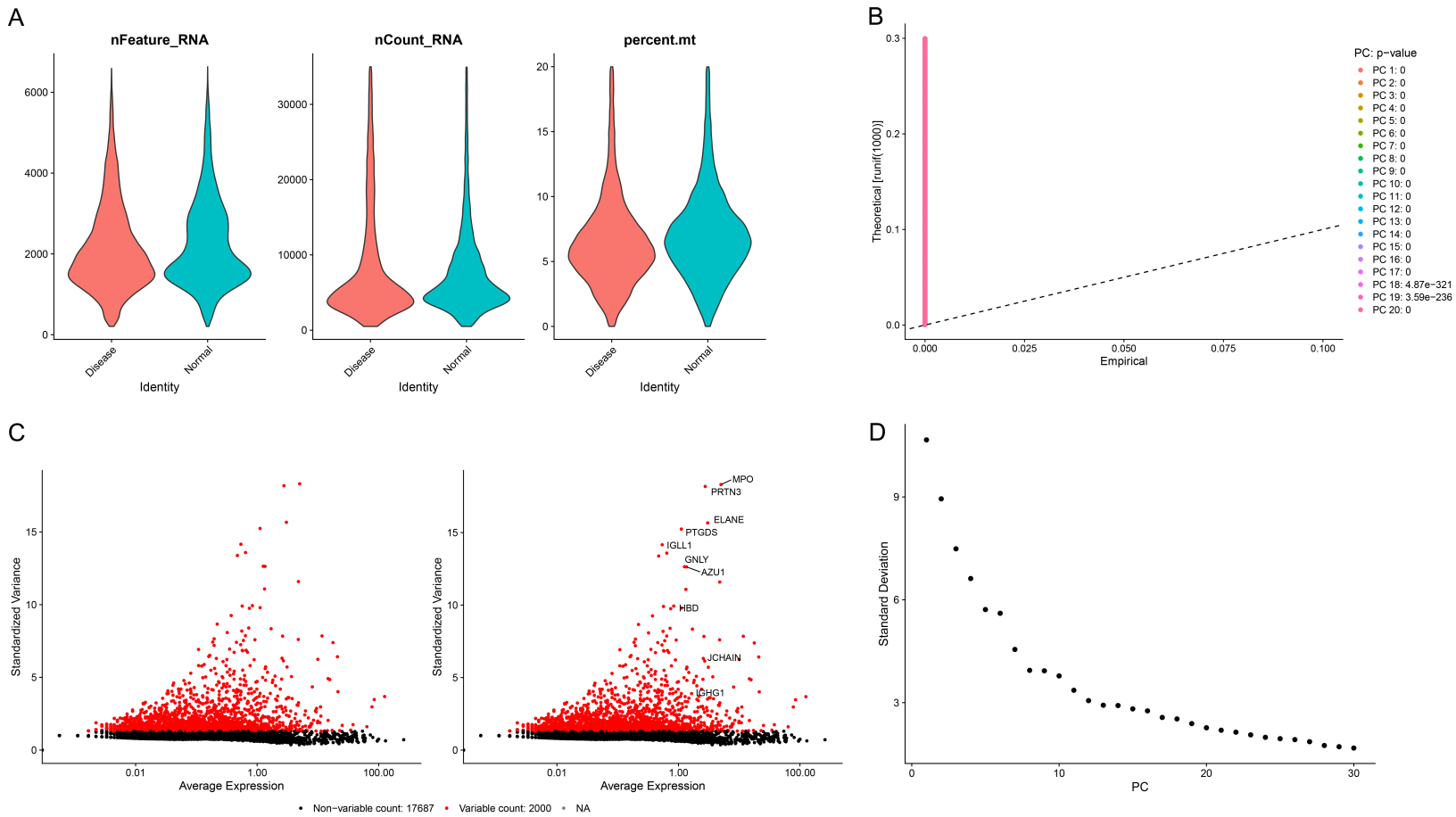
[5] Zhan F, Hardin J, Kordsmeier B, Bumm K, Zheng M, Tian E, Sanderson R, Yang Y, Wilson C, Zangari M, Anaissie E, Morris C, Muwalla F, van Rhee F, Fassas A, Crowley J, Tricot G, Barlogie B and Shaughnessy J Jr. Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. Blood 2002; 99: 1745-1757.

[6] Shi L, Campbell G, Jones WD, Campagne F, Wen Z, Walker SJ, Su Z, Chu TM, Goodsaid FM, Pusztai L, Shaughnessy JD Jr, Oberthuer A, Thomas RS, Paules RS, Fielden M, Barlogie B, Chen W, Du P, Fischer M, Furlanello C, Gallas BD, Ge X, Megherbi DB, Symmans WF, Wang MD, Zhang J, Bitter H, Brors B, Bushel PR, Bylesjo M, Chen M, Cheng J, Cheng J, Chou J, Davison TS, Delorenzi M, Deng Y, Devanarayan V, Dix DJ, Dopazo J, Dorff KC, Elloumi F, Fan J, Fan S, Fan X, Fang H, Gonzaludo N, Hess KR, Hong H, Huan J, Irizarry RA, Judson R, Juraeva D, Lababidi S, Lambert CG, Li L, Li Y, Li Z, Lin SM, Liu G, Lobenhofer EK, Luo J, Luo W, McCall MN, Nikolsky Y, Pennello GA, Perkins RG, Philip R, Popovici V, Price ND, Qian F, Scherer A, Shi T, Shi W, Sung J, Thierry-Mieg D, Thierry-Mieg J, Thodima V, Trygg J, Vishnuvajjala L, Wang SJ, Wu J, Wu Y, Xie Q, Yousef WA, Zhang L, Zhang X, Zhong S, Zhou Y, Zhu S, Arasappan D, Bao W, Lucas AB, Berthold F, Brennan RJ, Buness A, Catalano JG, Chang C, Chen R, Cheng Y, Cui J, Czika W, Demichelis F, Deng X, Dosymbekov D, Eils R, Feng Y, Fostel J, Fulmer-Smentek S, Fuscoe JC, Gatto L, Ge W, Goldstein DR, Guo L, Halbert DN, Han J, Harris SC, Hatzis C, Herman D, Huang J, Jensen RV, Jiang R, Johnson CD, Jurman G, Kahlert Y, Khuder SA, Kohl M, Li J, Li L, Li M, Li QZ, Li S, Li Z, Liu J, Liu Y, Liu Z, Meng L, Madera M, Martinez-Murillo F, Medina I, Meehan J, Miclaus K, Moffitt RA, Montaner D, Mukherjee P, Mulligan GJ, Neville P, Nikolskaya T, Ning B, Page GP, Parker J, Parry RM, Peng X, Peterson RL, Phan JH, Quanz B, Ren Y, Riccadonna S, Roter AH, Samuelson FW, Schumacher MM, Shambaugh JD, Shi Q, Shippy R, Si S, Smalter A, Sotiriou C, Soukup M, Staedtler F, Steiner G, Stokes TH, Sun Q, Tan PY, Tang R, Tezak Z, Thorn B, Tsyganova M, Turpaz Y, Vega SC, Visintainer R, von Frese J, Wang C, Wang E, Wang J, Wang W, Westermann F, Willey JC, Woods M, Wu S, Xiao N, Xu J, Xu L, Yang L, Zeng X, Zhang J, Zhang L, Zhang M, Zhao C, Puri RK, Scherf U, Tong W and Wolfinger RD; MAQC Consortium. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. Nat Biotechnol 2010; 28: 827-838.

[7] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.

[8] Gu Z, Gu L, Eils R, Schlesner M and Brors B. Circlize implements and enhances circular visualization in R. Bioinformatics 2014; 30: 2811-2812.
[9] Wickham H. Ggplot2: elegant graphics for data analysis. New York: Springer-Verlag, 2016.
[10] Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16: 284-287.
[11] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-D613.
[12] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT and Lin CY. CytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014; 8 Suppl 4: S11.
[13] Shaughnessy JD Jr, Zhan F, Burington BE, Huang Y, Colla S, Hanamura I, Stewart JP, Kordsmeier B, Randolph C, Williams DR, Xiao Y, Xu H, Epstein J, Anaissie E, Krishna SG, Cottler-Fox M, Hollmig K, Mohiuddin A, Pineda-Roman M, Tricot G, van Rhee F, Sawyer J, Alsayed Y, Walker R, Zangari M, Crowley J and Barlogie B. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. Blood 2007; 109: 2276-2284.
[14] Ishwaran H, Kogalur UB, Blackstone EH and Lauer MS. Random survival forests. Ann Appl Stat 2008; 2: 841-860, 820.
[15] Terry M Therneau PMG. Modeling survival data: extending the cox model. New York: Springer, 2000.
[16] Friedman J, Hastie T and Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33: 1-22.
[17] Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C and Lin C. e1071: misc functions of the department of statistics, probability theory group. R package version 1.7-16; 2024
[18] Chen T and Guestrin C. XGBoost: a scalable tree boosting system. Proc ACM SIGKDD Int Conf Knowl Discov Data Min 2016: 785-794
[19] Kursa MB and Rudnicki WR. Feature selection with the boruta package. J Stat Softw 2010; 36: 1-13.
[20] Chen H and Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 2011; 12: 35.
[21] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M and Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015; 12: 453-457.
[22] Kolde R. Pheatmap: pretty heatmaps. R package version 1.0.12; 2019.
[23] Revelle W. Psych: procedures for psychological, psychometric, and personality research. R package version 2.4.6; 2024.
[24] Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P and Satija R. Integrated analysis of multimodal single-cell data. Cell 2021; 184: 3573-3587, e3529.
[25] Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ and Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019; 20: 163-172.
[26] Zhang D ZM and Wu J. Single-cell RNA sequencing reveals the immune landscape of multiple myeloma. Acta Pharmacologica Sinica 2023; 44: 1350-1365.
[27] Sorrentino G DSH and Ryi SY. Single-cell transcriptomic characterization of the retinal immune response to Zika infection. Cell Reports 2020; 30: 618-630.
[28] Zheng J, Zhang T, Guo W, Zhou C, Cui X, Gao L, Cai C and Xu Y. Integrative analysis of multi-omics identified the prognostic biomarkers in acute myelogenous leukemia. Front Oncol 2020; 10: 591937.
[29] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV and Nie Q. Inference and analysis of cell-cell communication using CellChat. Nat Commun 2021; 12: 1088.
[30] Iannozzi NT, Giuliani N and Storti P. Deciphering the bone marrow microenvironment's role in multiple myeloma immunotherapy resistance. Front Immunol 2025; 16: 1613265.
[31] Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, Bai J, Zhao T, Li X and Xiao Y. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res 2019; 47: D721-D728.
[32] Schneider CA, Rasband WS and Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods 2012; 9: 671-675.
[33] McInnes L, Healy J and Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv 2018; 1802.03426.
[34] Mafra A, Laversanne M, Marcos-Gragera R, Chaves HVS, Mcshane C, Bray F and Znaor A. The global multiple myeloma incidence and

mortality burden in 2022 and predictions for 2045. J Natl Cancer Inst 2025; 117: 907-914.
[35] Lin CH, Tariq MJ, Ullah F, Sannareddy A, Khalid F, Abbas H, Bader A, Samaras C, Valent J, Khouri J, Anwer F, Raza S and Dima D. Current novel targeted therapeutic strategies in multiple myeloma. Int J Mol Sci 2024; 25: 6192.
[36] Ochi T, Konishi T and Takenaka K. Bispecific antibodies for multiple myeloma: past, present and future. Int J Hematol 2024; 120: 23-33.
[37] Riccardi F, Tangredi C, Dal Bo M and Toffoli G. Targeted therapy for multiple myeloma: an overview on CD138-based strategies. Front Oncol 2024; 14: 1370854.
[38] Xu D, Hu J, Xu S, De Bruyne E, Menu E, Van Camp B, Vanderkerken K and Van Valckenborgh E. Dll1/Notch activation accelerates multiple myeloma disease development by promoting CD138+ MM-cell proliferation. Leukemia 2012; 26: 1402-1405.
[39] Yu T, Chaganty B, Lin L, Xing L, Ramakrishnan B, Wen K, Hsieh PA, Wollacott A, Viswanathan K, Adari H, Cho SF, Li Y, Chen H, Yang W, Xu Y, An G, Qiu L, Munshi N, Babcock G, Shriver Z, Myette JR, Anderson KC and Tai YT. VIS832, a novel CD138-targeting monoclonal antibody, potently induces killing of human multiple myeloma and further synergizes with IMiDs or bortezomib in vitro and in vivo. Blood Cancer J 2020; 10: 110.
[40] Stopp S, Gründl M, Fackler M, Malkmus J, Leone M, Naumann R, Frantz S, Wolf E, von Eyss B, Engel FB and Gaubatz S. Deletion of *Gas2l3* in mice leads to specific defects in cardiomyocyte cytokinesis during development. Proc Natl Acad Sci U S A 2017; 114: 8029-8034.
[41] Zhao Y, Zheng M and Li P. *GAS2L3* silencing inhibits hepatocellular carcinoma progression. J Adv Res 2025.
[42] Asai T, Liu Y, Di Giandomenico S, Bae N, Ndiaye-Lobry D, Deblasio A, Menendez S, Antipin Y, Reva B, Wevrick R and Nimer SD. Necdin, a p53 target gene, regulates the quiescence and response to genotoxic stress of hematopoietic stem/progenitor cells. Blood 2012; 120: 1601-1612.
[43] Lu Q, Yang D, Li H, Niu T and Tong A. Multiple myeloma: signaling pathways and targeted therapy. Mol Biomed 2024; 5: 25.
[44] Gao J, Yan X, Fan D and Li Y. Single-cell data revealed the function of natural killer cells and macrophage cells in chemotherapy tolerance in acute myeloid leukemia. PeerJ 2024; 12: e18521.
[45] Miraki Feriz A, Bahraini F, Khosrojerdi A, Azarkar S, Sajjadi SM, HosseiniGol E, Honardoost MA, Saghafi S, Silvestris N, Leone P, Safarpour H and Racanelli V. Deciphering the immune landscape of head and neck squamous cell carcinoma: a single-cell transcriptomic analysis of regulatory T cell responses to PD-1 blockade therapy. PLoS One 2023; 18: e0295863.
[46] Church EC, Bishop E, Fiore-Gartland A, Yu KKQ, Chang M, Jones RM, Brache JK, Ballweber Fleming L, Phan JM, Makatsa MS, Heptinstall J, Chiong K, Dintwe O, Naidoo A, Voillet V, Mayer-Blackwell K, Nwanne G, Andersen-Nissen E, Vary JC Jr, Tomaras GD, McElrath MJ, Sherman DR, Murphy SC, Kublin JG and Seshadri C. Probing dermal immunity to mycobacteria through a controlled human infection model. Immunohorizons 2024; 8: 695-711.


A
nFeature_RNA
nCount_RNA
percent.mt
Disease
Normal
Identity
B
Theoretical [runif(1000)]
Empirical
PC: p-value
PC 1: 0
PC 2: 0
PC 3: 0
PC 4: 0
PC 5: 0
PC 6: 0
PC 7: 0
PC 8: 0
PC 9: 0
PC 10: 0
PC 11: 0
PC 12: 0
PC 13: 0
PC 14: 0
PC 15: 0
PC 16: 0
PC 17: 0
PC 18: 4.87e-321
PC 19: 3.59e-236
PC 20: 0
C
Standardized Variance
Average Expression
MPO
PRTN3
ELANE
PTGDS
IGLL1
GNLY
AZU1
HBD
JCHAIN
IGHG1
Non-variable count: 17687
Variable count: 2000
NA
D
Standard Deviation
PC

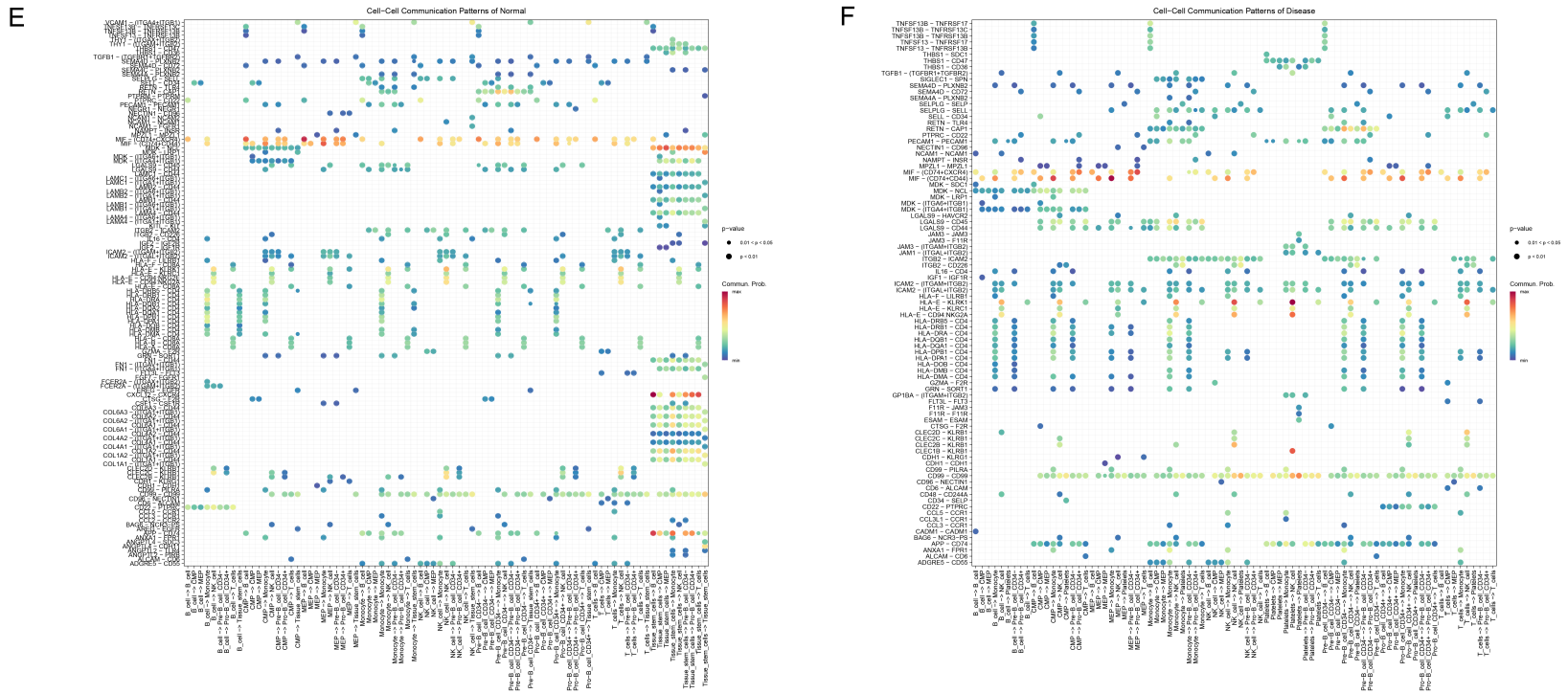


Figure S1. Intercellular communication analysis of ligand-receptor signaling pathways in normal and multiple myeloma samples. A, B. Bubble plots showing the communication probabilities and interaction strengths of ligand-receptor pairs among different cell populations in the normal and disease groups. C, D. Bubble plots illustrating the major signaling pathways mediating intercellular communication in the normal and disease groups, respectively. E, F. Detailed visualization of representative signaling axes, including MIF-(CD74+CXCR4), MIF-(CD74+CD44), APP-CD74, CXCL12-CXCR4, and HLA-E-KLRK1. The color gradient from blue to red represents increasing communication probability, while bubble size indicates interaction strength. Enhanced HLA-E-KLRK1 signaling was observed in the disease group, particularly in platelet-NK cell interactions, T cell-NK cell interactions, and NK cell autocrine signaling.