

RESEARCH

Open Access



Identification of clinical implications and potential prognostic models of chromatin regulator mutations in multiple myeloma

Lina Zhang , Run Zhang, Jing Wang, Ying Chen, Chun Qiao, Qinglin Shi, Yuanyuan Jin, Xuxing Shen, Jianyong Li and Lijuan Chen 

Abstract

Background: With the rapid development of next-generation sequencing (NGS) technologies, researchers are making efforts to reveal the genomic landscape of multiple myeloma (MM). However, the clinical significance of many mutations remains poorly defined due to the genetic heterogeneity of MM. To systematically explore the clinical implications of gene mutations and build practical prognostic models, we performed DNA sequencing in newly diagnosed MM patients.

Methods: MM cells were purified from bone marrow aspirates using CD138 microbeads and subjected to sequencing with a 387-gene Panel. Nomogram was developed using Cox's proportional hazards model, and candidate variables were screened by stepwise regression. Internal validation was carried out by the bootstrap method.

Results: Between July 2016 and December 2020, a total of 147 patients were included in our study. We found patients with a higher mutational load had a significantly shorter progress-free survival (PFS) (19.0 vs. 32.0 months, $P=0.0098$) and overall survival (OS) (3-year OS rates were 66.1% and 80.0%, $P=0.0290$). Mutations in chromatin regulators (CRs) including *KMT2C* (14.3%), *KMT2D* (14.3%), *EP300* (11.6%) and ARID gene family (31.3%) were highly frequent in newly diagnosed MM patients. Interestingly, proteins encoded by these genes could form a complex called KMT2C/D COMPASS (K_{CD} COMs). Patients with mutations of ARID gene family had a significantly shorter PFS (15.5 vs. 34.0 months, $P=0.0003$) and OS (3-year OS rates were 64.9% and 81.0%, $P=0.0351$) than patients without ARID gene mutations. Incorporating ARID gene mutations into the current staging system could successfully improve their prognostic performance. The PFS and OS nomogram models (including 1q21 copies, ARID gene mutations, extramedullary disease, mutational load and *TP53* mutations) showed good predicting performance in both training and validation sets.

Conclusion: Our findings emphasized the importance of CRs mutations in newly diagnosed MM patients and indicated the mutations affecting K_{CD} COMs might promote the development of MM. High mutational load and harboring mutations in the ARID gene family were novel predictors of adverse prognosis in MM. Prognostic models based on gene mutations were commendably prognostic evaluation methods that could provide a reference for clinical practices.

Keywords: Next-generation sequencing, Chromatin regulator, Epigenetic, Multiple myeloma, Nomogram

*Correspondence: chenljb@126.com

Department of Hematology, The First Affiliated Hospital of Nanjing Medical University, Jiangsu Province Hospital, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing 210029, China



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Multiple myeloma (MM) is a common hematological malignancy characterized by neoplastic proliferation of monoclonal plasma cells [1]. Despite the emergence of numerous new drugs, it is still an incurable disease. Therefore, the biological characteristics of MM require further exploration to find new predictive markers and therapeutic approaches. The advent of next-generation sequencing (NGS) technologies has provided a powerful tool for studies on cancer genetics in the past 15 years. MM showed a moderate level of tumor mutational load (1.6 mutations/Mb) across various cancer types, which suggested an important role of gene mutations in the occurrence and development of MM [2].

Several studies [3–5] have explored the genetic mutation profile of MM using NGS technologies since 2014, and the authors were consistent with the conclusion that the top recurrent mutated genes were *KRAS* (mutation rates, 20–23%) and *NRAS* (19–20%) in MM. Other common mutated genes included *DIS3*, *FAM46C*, *BRAF*, *TP53* and *TRAF3*. These genes with a higher mutation rate were concentrated in five pathways, including the MAPK pathway (*KRAS*, *NRAS* and *BRAF*), plasma cell differentiation pathway (*IRF4* and *PRDM1*), the NF- κ B pathway (*TRAF3*, *CYLD* and *LTB*), cell-cycle control pathway (*RBI* and *CCND1*) and DNA repair pathway (*ATM*, *ATR*, *TP53*) [5, 6]. Previous studies confirmed that gene mutations also had considerable clinical and prognostic significance in MM. Mutations of *IRF4* and *EGR1* were found to be correlated with a favorable prognosis, while mutations of *TP53*, *ATM* and *ATR* were associated with a poor prognosis [4–6]. However, the genetics of MM are complex and the clinical significance of many mutations remains unknown.

In 2020, Ordoñez et al. [7] compared malignant plasma cells versus normal B cells using a multi-epigenomics approach and found extensive activation of regulatory elements in tumor cells, which led to the overexpression of members involved in the NOTCH, NF- κ B, mTOR and p53 signaling pathways and promoted the proliferation of plasma cells. These data demonstrated epigenetic changes played an important role in MM pathogenesis. Epigenetics refers to the regulation of gene expression without changes in DNA sequence, and chromatin regulators (CRs) are important research objects in epigenetics [8]. Based on the different functions, CRs can be grouped into three classes: (1) DNA methylators, including DNMT1, DNMT3A and DNMT3B; (2) histone-modifying enzymes, including KMT2C, KMT2D, p300 and HAT1; (3) chromatin remodelers, including ARID1A, ARID1B, ARID2, CHD5 and ACF1 [9]. Although mutations of CRs are common across multiple cancers, we still have limited knowledge of their roles in MM.

To further investigate the clinical significance of gene mutations in MM, we performed DNA sequencing in 147 patients with newly diagnosed multiple myeloma (NDMM) by NGS technologies and found mutations of CRs have important prognostic significance. We also developed nomogram models based on mutations of CRs to predict the risk of relapse or mortality in MM patients, which can be used conveniently in clinical practice.

Materials and methods

Patients and data collection

The study was approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University. The use of tumor specimens was approved by the Ethics Committee, and informed consent was obtained from all patients. The diagnoses of patients were made based on the International Myeloma Working Group 2014 criteria [10]. Clinical staging was based on the Durie–Salmon (DS) staging system [11], International Staging System (ISS) and Revised International Staging System (R-ISS) [12].

Laboratory examination data such as hemoglobin (Hb), serum creatinine, bone marrow plasma cells proportion, lactate dehydrogenase (LDH) and β_2 microglobulin (β_2 -MG) were collected in our study. The presence of extramedullary disease (EMD) was detected by PETCT, whole-body CT or whole-body MRI. High-risk cytogenetic abnormalities were defined as at least one of the following: del(17p), t(4; 14), or t(14; 16) using cytoplasmic immunoglobulin fluorescence in situ hybridization (cIg-FISH). The numerical aberrations of the chromosomal regions 1q21 were also detected by cIg-FISH. Progress-free survival (PFS) was calculated from the time of diagnosis to disease progression (PD) or death due to any reason. Overall survival (OS) was estimated as the time from diagnosis to death.

Next-generation sequencing

Bone marrow samples were collected before treatment, and tumor cells were purified by positive selection with anti-CD138 magnetic microbeads (Miltenyi Biotec, Germany). Genomic DNA was extracted from tumor cells using the QIAamp® Blood DNA Mini Kit (QIAGEN, Germany). The quantity of DNA was measured using a Qubit 2.0 fluorometer (Life Technologies, USA). Genotyping was performed using a sequencing panel of 387 genes (see Additional file 2 for details) by NGS technologies, and the average sequencing depth was 1000 \times . Genomic DNA (200 ng) was sheared with Enzyme Plus Library Prep Kit (iGenetech, China) and sequencing libraries were constructed using probes and TargetSeq One Kit (iGenetech, China) according to the instructions. Sequencing runs were performed

on NovaSeq 6000 sequencing platform (Illumina, USA) using a NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) (Illumina, USA). The raw sequencing data were available in the SRA database (<https://www.ncbi.nlm.nih.gov/sra>) under the accession number PRJNA826654.

Analysis of mutations

FASTQ files were pre-processed and assessed for quality with fastp. Sequencing reads were aligned to the human reference genome hg19 using BWA v. 0.7.17 software. SAMtools v. 1.11 software was used to convert Sequence Alignment/Map (SAM) files to Binary Alignment Map (BAM) [13], and gatk v.4.1.3.0 was applied to perform indel realignment and base quality recalibration. BAM files were sorted with SortSam, duplicate reads were removed with gatk MarkDuplicates, mapped reads were extracted with SAMtools, and base quality score was recalibrated with gatk BaseRecalibrator and ApplyBQSR [14]. Subsequently, we annotated the variants using Annovar and SnpEff [15]. Single nucleotide polymorphisms (SNPs) were identified using information from ExAC_ALL, ExAC_EAS and 1000g2015aug_all databases, and they were excluded from further analysis.

Derivation and validation of predictive nomogram models

The predictive nomogram was developed using Cox's proportional hazards model. Candidate variables were screened by stepwise regression. For the model fitting, we used the *R* package "survival". Model diagnostics were then performed by package "survminer", including Deviance residues to assess the effect of outlier cases and Schoenfeld residuals to test the proportional risk assumption. The area under the receiver operator curve (ROC) (AUC/C-statistic) and Brier score were calculated to assess the performance of the model. Finally, the model was retested for internal validation using the bootstrap method, with 100 replications.

Statistical methods

Statistical analysis and plotting were performed by *R* software v. 4.1.1 and GraphPad Prism v. 5.0. Numerical variables were analyzed by *t*-test or ANOVA, and non-normal distribution variables were analyzed by Mann–Whitney *U* test or Kruskal–Wallis test. Categorical variables were analyzed using Chi-square test. K–M method and Log-rank test were used to plot and analyze the survival curves. Cox proportional hazards model was used for multifactorial analysis. For all analyses, a *P* value less than 0.05 was considered significant.

Table 1 Clinical characteristics of 147 patients

Clinical characteristics	Proportion (%)
<i>Gender</i>	
Male	60.5 (89/147)
Female	39.5 (58/147)
<i>Subtype</i>	
IgG	48.3 (71/147)
IgA	24.5 (36/147)
IgD	6.1 (9/147)
IgE	0.7 (1/147)
κ light chain	11.6 (17/147)
λ light chain	8.8 (13/147)
<i>DS stage</i>	
I	4.1 (6/147)
II	13.6 (20/147)
III	82.3 (121/147)
<i>ISS stage</i>	
I	15.6 (23/147)
II	32.0 (47/147)
III	52.4 (77/147)
<i>R-ISS stage*</i>	
I	13.0 (18/138)
II	64.5 (89/138)
III	22.6 (31/138)
<i>1q21*</i>	
Gain	25.7 (35/136)
Amplification	24.3 (33/136)
<i>High-risk karyotype*</i>	
del(17p)	12.5 (17/136)
t(4; 14)	17.6 (24/136)
t(14; 16)	0.7 (1/136)
Anemia [†]	64.6 (95/147)
Renal insufficiency [‡]	31.3 (46/147)
Extramedullary disease	20.4 (30/147)
<i>Induction therapy regimen</i>	
PIs-based	50.3 (74/147)
IMiDs-based	9.5 (14/147)
PIs + IMiDs	38.1 (56/147)
Completed ASCT	25.9 (38/147)

ISS International Staging System, *PIs* proteasome inhibitors, *IMiDs* immunomodulatory drugs, *ASCT* autologous hematopoietic stem cell transplantation

*Some patients did not undergo the FISH examinations, so the proportions are calculated by the numbers of patients with FISH results

[†] Anemia refers to hemoglobin < 100 g/L

[‡] Renal insufficiency refers to creatinine > 177 μmol/L or creatinine clearance < 40 ml/min

Results

Clinical characteristics of our study population

From July 2016 to December 2020, a total of 147 NDMM patients were included in the study and the clinical characteristics of patients are summarized in

Table 1. The median age was 63.0 years (range, 38.0–84.0 years), and 60.5% of patients were male. The majority of patients (50.3%) were treated with proteasome inhibitors (PIs)-based regimen as first-line chemotherapy, such as PCD (bortezomib/cyclophosphamide/dexamethasone), PAD (bortezomib/doxorubicin/dexamethasone) and VD (bortezomib/dexamethasone). 38.1% of patients received combination therapy with PIs and immunomodulatory drugs (IMiDs), 9.5% of patients received IMiDs-based regimen, and the remaining three patients were treated symptomatically.

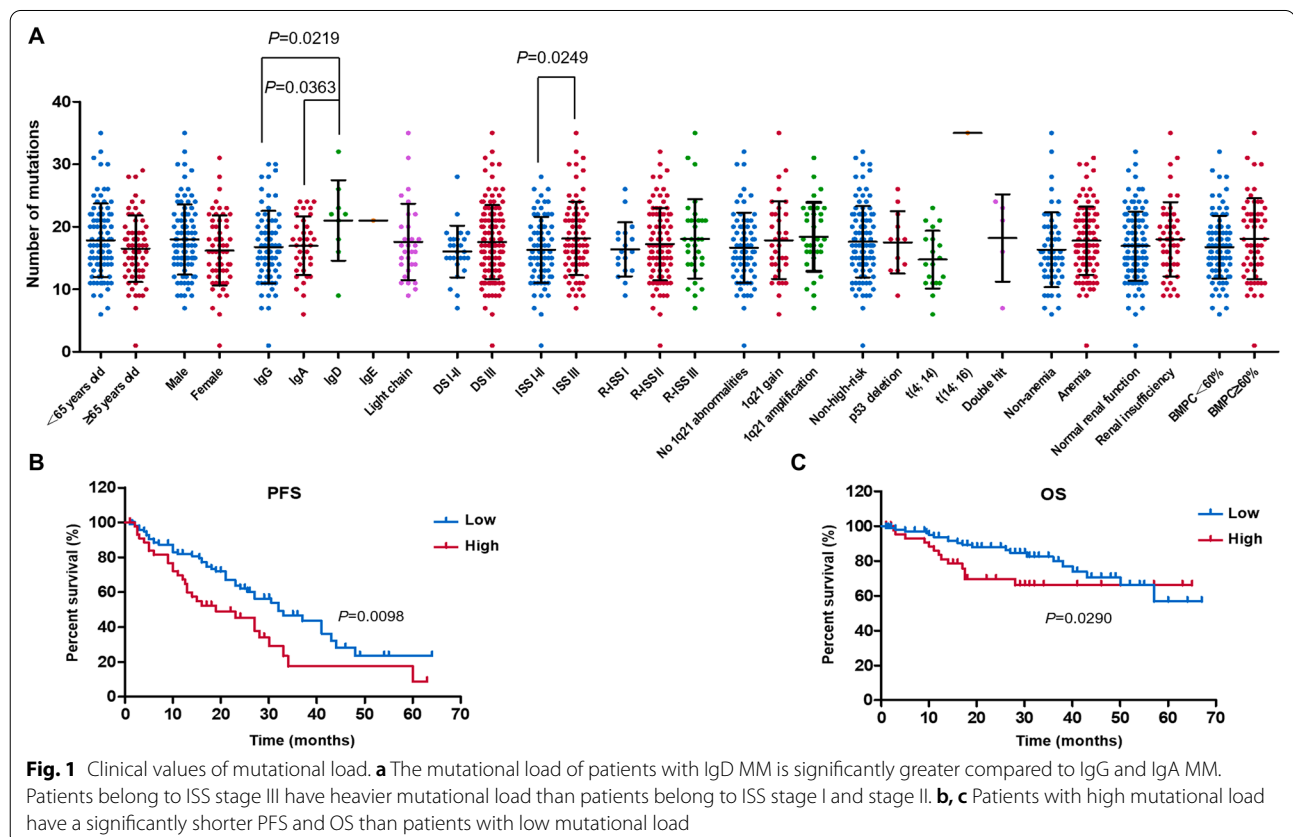
Follow-up time ranged from 2.0 to 67.0 months, with a median time of 26.0 months. During follow-up, 49.7% of patients developed PD and 22.4% of patients died. The 1-year, 2-year and 3-year PFS rates were 77.3%, 56.9% and 37.4%, respectively (the median PFS was 28.0 months). The 1-year, 2-year and 3-year OS rates were 90.7%, 82.3% and 75.4%, respectively (the median OS was not reached).

Patients with heavier mutational load had a poor prognosis

A total of 343 mutated genes were detected, and missense mutation was the most common mutation type. The median total number of mutations in each patient

was 17.0 (range 1.0–35.0). In order to fully understand the clinical value of mutational load, we performed an integrated analysis of clinical features and the number of mutations (Fig. 1a). The number of mutations in patients with IgD MM was significantly greater compared to IgG ($P=0.0219$) and IgA ($P=0.0363$) MM. Patients with ISS stage I and stage II were combined because of the low number of patients in ISS stage I group. The mutational load of patients with ISS stage III was significantly heavier than those of patients with ISS stage I and stage II ($P=0.0249$).

With 20 as the threshold, patients were divided into a high mutational load group (number of mutations ≥ 20 , $n=47$) and a low mutational load group (number of mutations <20 , $n=100$). Patients in the low mutational load group had a significantly longer PFS than patients in the high mutational load group (median PFS, 32.0 vs. 19.0 months, $P=0.0098$, Fig. 1b). In a similar manner, OS of the low mutational load group was significantly longer than that of the high mutational load group (median OS was not reached, 3-year OS was 80.0% and 66.1%, $P=0.0290$, Fig. 1c). These results indicated that a high mutational load was associated with poor prognosis in MM patients.



Profile of gene mutations

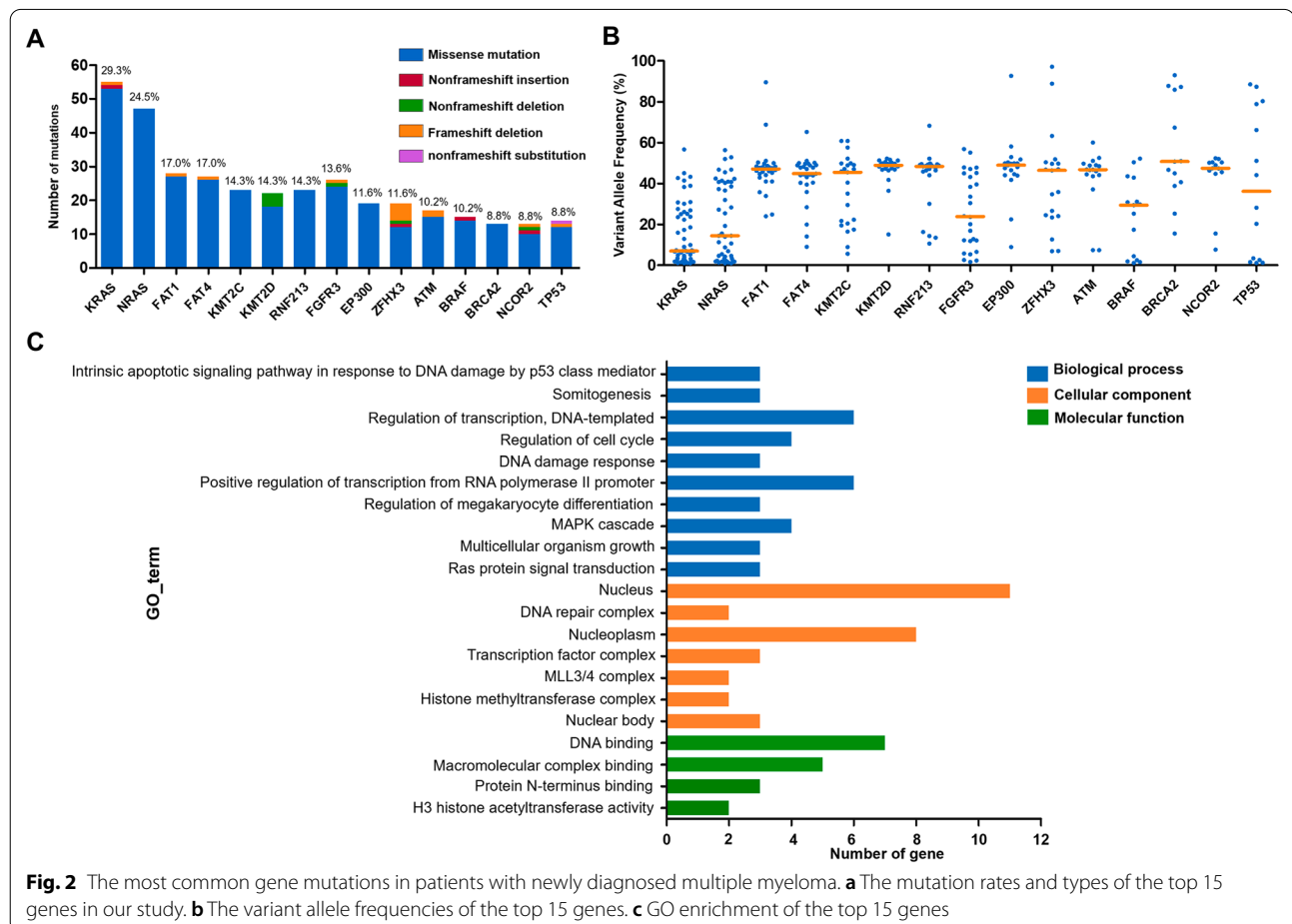
In this study, the top 15 genes with the highest mutation frequencies were *KRAS* (29.3%), *NRAS* (24.5%), *FAT1* (17.0%), *FAT4* (17.0%), *KMT2C* (14.3%), *KMT2D* (14.3%), *RNF213* (14.3%), *FGFR3* (13.6%), *EP300* (11.6%), *ZFH3* (11.6%), *ATM* (10.2%), *BRAF* (10.2%), *BRCA2* (8.8%), *NCOR2* (8.8%) and *TP53* (8.8%) (Fig. 2a). The variant allele frequencies (VAFs) are presented in Fig. 2b, and the median VAFs for each gene were *KRAS* (7.0%, range 1.0–56.7%), *NRAS* (14.4%, range 1.0–56.4%), *FAT1* (47.1%, range 24.0–89.6%), *FAT4* (44.8%, range 9.0–65.3%), *KMT2C* (45.5%, range 5.6–60.9%), *KMT2D* (48.9%, range 15.1–52.2%), *RNF213* (48.4%, range 10.6–68.3%), *FGFR3* (23.8%, range 1.7–56.9%), *EP300* (49.0%, range 8.9–92.7%), *ZFH3* (46.4%, range 6.9–97.2%), *ATM* (46.8%, range 7.4–60.1%), *BRAF* (29.4%, range 1.2–52.2%), *BRCA2* (50.8%, range 15.5–93.0%), *NCOR2* (47.4%, range 7.6–52.4%) and *TP53* (36.2%, range 1.2–88.5%).

To better understand the biological functions of the top 15 genes, we performed gene ontology (GO) enrichment analysis using DAVID (<https://david.ncifc>

[rf.gov/](https://david.ncifc.gov/)) (Fig. 2c). In biological process (BP) category, the mutated genes were significantly enriched in the terms “regulation of transcription, DNA-templated” and “positive regulation of transcription from RNA polymerase II promoter”. Correspondingly, these genes were markedly enriched in “nucleus” of cellular component (CC) category and “DNA binding” of molecular function (MF) category. The enrichment results indicated that genes involved in regulation of gene transcription (*KMT2D*, *ZFH3*, *KMT2C*, *EP300*, *BRCA2* and *TP53*) might play an important role in MM pathogenesis. Interestingly, among the enriched genes, *KMT2D*, *KMT2C* and *EP300* belong to CRs and are implicated in the formation of KMT2C/D COMPASS complex (K_{CD}COMs) [16].

Correlation between chromosome abnormalities and mutated genes

The most common cytogenetic abnormalities in MM include copy number gains of 1q21, del(17p) as well as IgH translocations t(4; 14), t(11; 14), t(14; 16) and t(14; 20). However, only t(4; 14) was included in the further

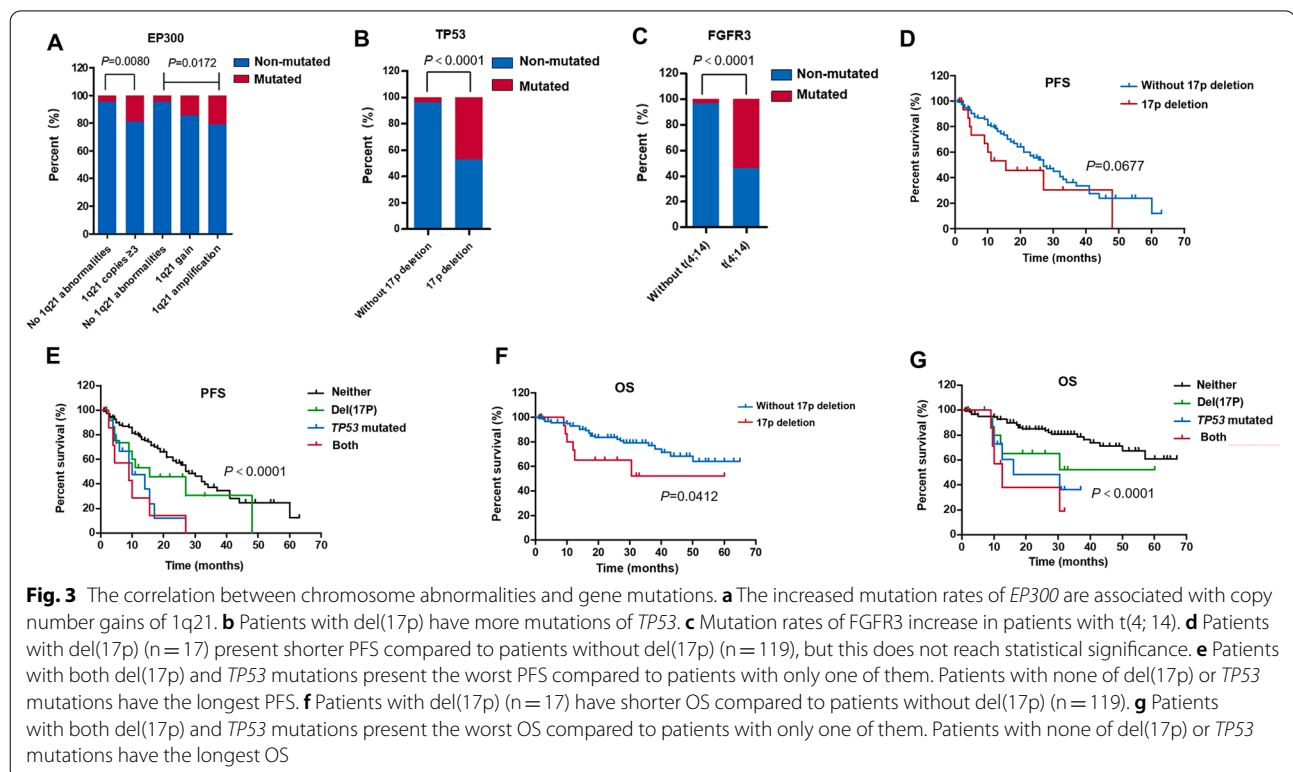


analysis for there were very few patients with t(11;14), t(14;16) and t(14;20).

The increased mutation rate of *EP300* (19.1% vs. 4.4%, $P=0.0080$) was associated with copy number gains of 1q21. Three copies of 1q21 were defined as 1q21 gain, and four or more copies were defined as 1q21 amplification. With the increase of 1q21 copies, the mutation rates of *EP300* also gradually increased (the mutation rates were 4.4%, 14.3% and 21.2%, respectively; $P=0.0172$, Fig. 3a). Moreover, there was a clear correlation between gene mutations and chromosome abnormalities on which the gene was located. Patients with del(17p) had noticeably more mutations of *TP53* (47.1% vs. 4.2%, $P<0.0001$) (Fig. 3b) and mutation rates of *FGFR3* obviously increased in patients with t(4; 14) (54.2% vs. 3.6%, $P<0.0001$) (Fig. 3c).

The prognostic value of chromosome abnormalities and gene mutations was evaluated using Kaplan–Meier survival analysis. In patients with del(17p), there was a tendency toward a reduction in PFS, but this did not reach statistical significance (median PFS, 15.5 vs. 27.0 months, $P=0.0677$) (Fig. 3d). We re-grouped patients into four groups according to del(17p) and *TP53* mutations: neither group contained patients without del(17p) and *TP53* mutations ($n=114$),

del(17p) group contained patients with del(17p) positive ($n=17$), *TP53* mutated group contained patients harboring *TP53* mutations ($n=13$), and both group contained patients harboring del(17p) and *TP53* mutations concomitantly ($n=8$). The median PFS for these four groups were 28.0, 15.5, 10.0 and 9.0 months. A significant difference in PFS among the four groups was observed ($P<0.0001$) (Fig. 3e), which suggested that patients could be better stratified by the combined use of del(17p) and *TP53* mutations. Analysis for OS was performed next in the same manner. Patients with del(17p) had shorter OS compared to patients without del(17p) (median OS was not reached, 3-year OS was 52.1% and 76.9%, $P=0.0412$) (Fig. 3f). The 3-year OS for neither group, del(17p) group, *TP53* mutated group and both group were 77.1%, 52.1%, 36.4% and 19.0%, respectively ($P<0.0001$) (Fig. 3g). Similar to the PFS, the combined use of del(17p) and *TP53* mutations allowed a better prediction of OS. Moreover, our results showed the effect of *TP53* mutations on patients' prognosis was more pronounced than del(17p). However, further studies were needed to draw the final conclusion. Combined analysis of 1q21 gain and *EP300* mutations, t(4; 14) and *FGFR3* mutations did not show better discriminating abilities for predicting the outcomes of patients.



Mutations in ARID gene family were potential indicators of a poor prognosis

$K_{CD}COMs$ complex might play an important role in the development of MM as previously shown. Unfortunately, survival analysis found no difference in PFS or OS between patients with and without mutations of *KMT2C*, *KMT2D* or *EP300*. Therefore, we focused on another component in this complex, called ARID family (Fig. 4a). In our study population, 46 patients (31.3%) had one or more mutations in ARID gene family and missense mutation was the most common mutation type (Fig. 4b). The median VAFs for each gene were *ARID1A* (48.3%, range 9.5–52.3%), *ARID1B* (30.7%, range 19.3–51.5%), *ARID2* (46.3%, range 10.2–48.6%), *ARID3A* (49.8%, range 20.9–67.1%), *ARID3C* (45.5%, range 41.7–49.2%), *ARID4A* (45.2%, range 2.7–53.4%), *ARID4B* (37.4%, range 27.7–47.2%), *ARID5A* (50.7%, range 5.8–50.9%), *ARID5B* (49.9%, range 48.3–51.4%), *JARID1A* (45.7%, range 21.3–50.4%), *JARID1C* (49.6%, range 5.0–94.1%) and *JARID2* (54.1%, range 45.6–63.4%) (Fig. 4c).

To better understand the effects of ARID gene family on the prognosis of MM, we compared PFS and OS between patients with and without mutation of ARID gene family. We divided patients into two groups: the ARID mutated group ($n=46$) referred to patients who carried mutations of ARID family and the non-ARID-mutated group

($n=101$) referred to patients without mutations of ARID family. PFS was significantly shorter in the ARID mutated group compared to the non-ARID-mutated group (median PFS, 15.5 vs. 34.0 months, $P=0.0003$) (Fig. 4d). Similarly, the ARID mutated group had obviously shorter OS than the non-ARID-mutated group (median OS was not reached, 3-year OS was 64.9% and 81.0%, $P=0.0351$) (Fig. 4e).

To further evaluate the prognostic significance of ARID family mutations, we combined them with ISS and R-ISS staging system. Patients who belonged to ISS/R-ISS stage I and did not have mutations of ARID family were still divided into stage I; patients who belonged to ISS/R-ISS stage III and have mutations of ARID family were divided into stage III; all other situations were divided into stage II (Table 2).

According to the ISS, 23 patients were stage I, 47 patients were stage II, and 77 patients were stage III. The PFS could not be precisely distinguished (median PFS were 41.0, 24.0 and 27.0 months, respectively; $P=0.1820$), especially between stage II and stage III (Fig. 5a). In the ISS + ARID staging system, 19 patients had stage I, 113 patients had stage II and 15 patients had stage III. A significant reduction of PFS was observed as the stage progressed (median PFS were 41.0, 28.0 and 15.0 months, respectively; $P=0.0101$) supporting a

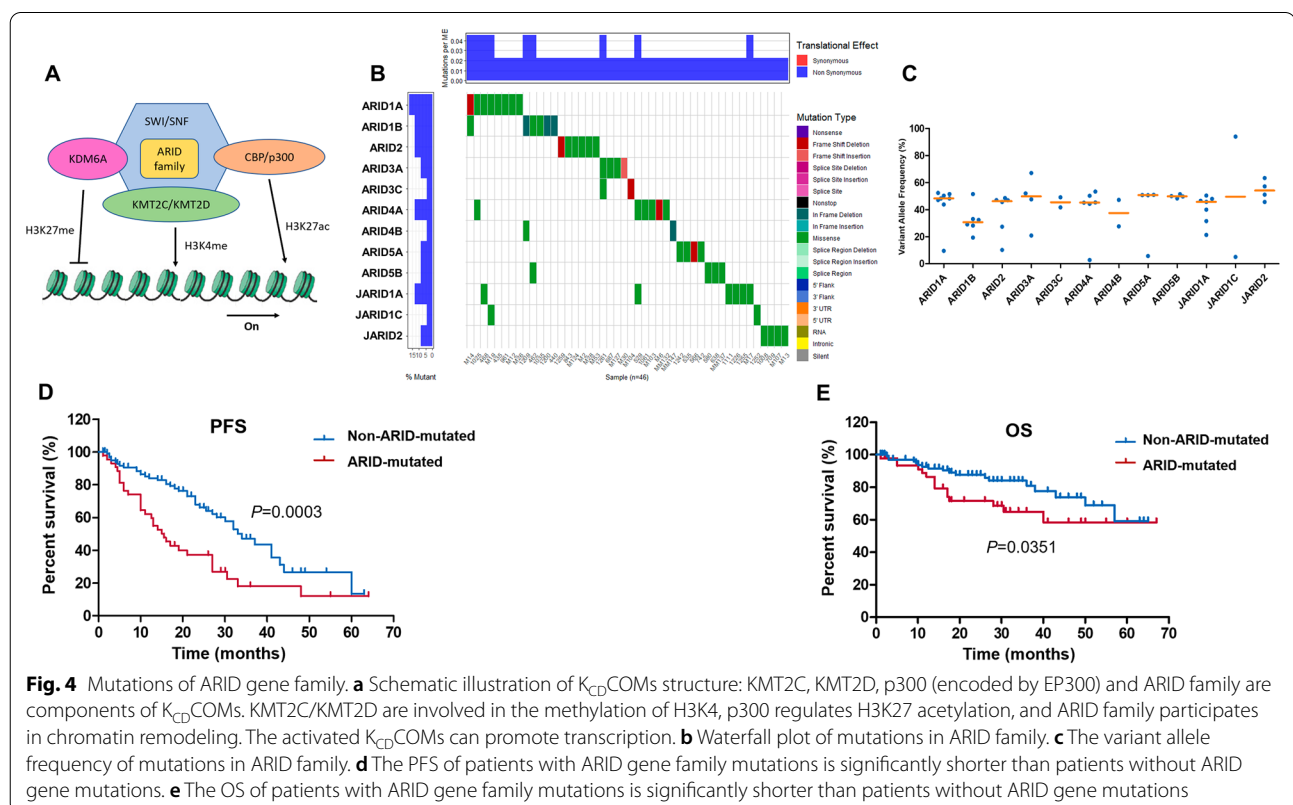
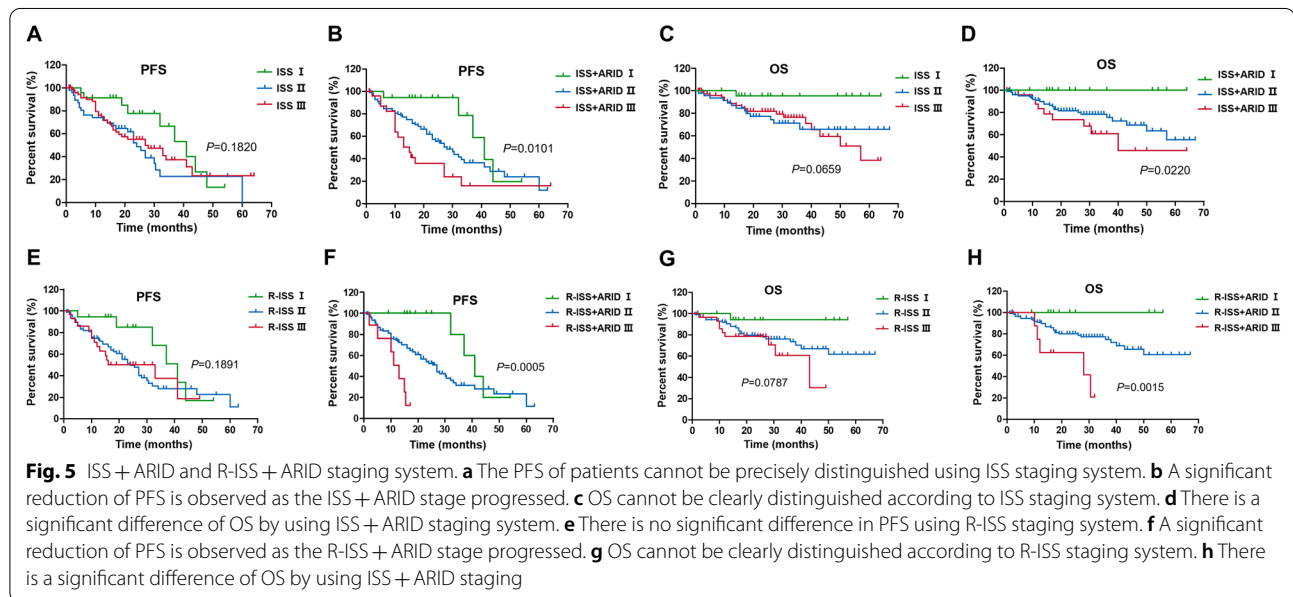


Table 2 ARID gene mutation staging system

Stage	ISS + ARID	R-ISS + ARID
I	ISS stage I without ARID mutations*	R-ISS stage I without ARID mutations
II	Not stage I or stage III	Not stage I or stage III
III	ISS stage III with ARID mutations	R-ISS stage III with ARID mutations

ISS international staging system, R-ISS revised international staging system

*ARID mutations refer to gene mutations in ARID gene family



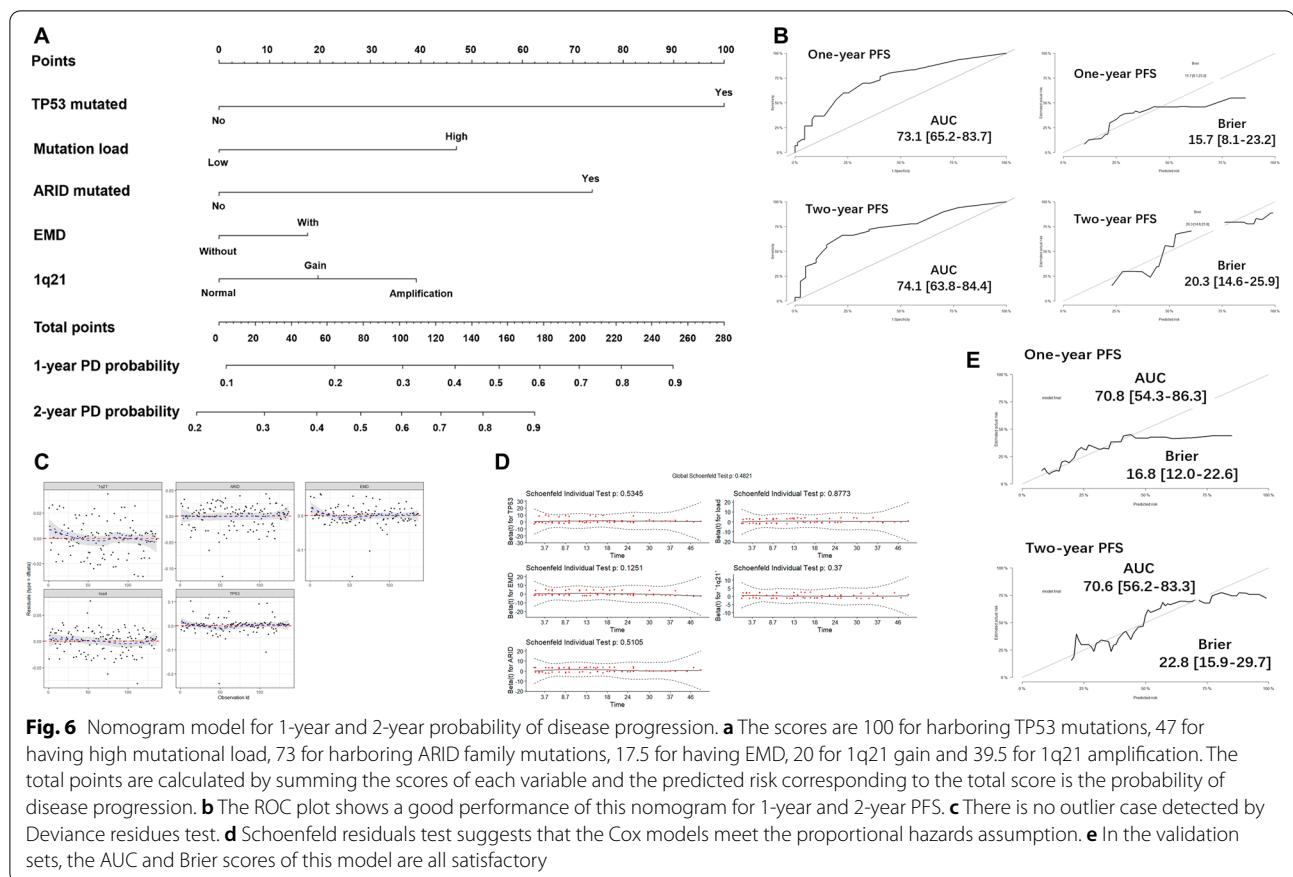
higher prediction power of ISS + ARID staging system (Fig. 5b). This was also the case for OS. Patients could not be clearly distinguished according to ISS staging (median OS was not reached, 3-year OS was 95.4%, 65.8% and 76.5%, $P=0.0659$) (Fig. 5c) while there was a significant difference of OS by using ISS + ARID staging (3-year OS was 100%, 75.6% and 61.0%, $P=0.0220$) (Fig. 5d).

According to the R-ISS, 18 patients were stage I, 89 patients were stage II, and 31 patients were stage III. There was no significant difference in PFS among these three groups (median PFS was 41.0, 24.0 and 33.0 months, respectively; $P=0.1891$). When using ISS + ARID staging system, 15 patients had stage I, 114 patients had stage II and 9 patients had stage III. ISS + ARID staging displayed better performance in distinguishing PFS as the stage progressed (median PFS was 41.0, 27.0 and 13.0 months, respectively; $P=0.0005$) (Fig. 5e). Similar findings were obtained for OS. Patients could not be clearly distinguished according to R-ISS staging (median OS was not reached, 3-year OS was 94.1%, 73.6% and 60.6%, $P=0.0787$) (Fig. 5g), while there was a significant difference of OS by using ISS + ARID staging (3-year OS was 100%, 74.8% and 20.8%, $P=0.0015$) (Fig. 5h).

Construction and validation of the nomogram based on ARID mutation

After removing the missing values, we got 136 patients with complete clinical data. Candidate variables were added into Cox proportional hazards regression models of PFS and OS, including age, gender, EMD, subtype, DS staging, ISS staging, R-ISS staging, LDH level, high-risk cytogenetic abnormalities, anemia, renal insufficiency, 1q21 copy numbers, mutation load, ARID family mutations and *TP53* mutations. The final variables were selected based on coefficients and P -values of stepwise regression as well as clinical values. Regression coefficients and HRs for the final model variables are presented in Additional file 1: Tables S4 and S5.

We established a nomogram model for PFS including *TP53* mutations, mutation load, ARID family mutations, EMD and 1q21 copy numbers (Fig. 6a). The total points were calculated by summing the scores of each variable, and the predicted risk corresponding to the total score was the probability of PD. The ROC plot showed a good performance of this model in predicting 1-year and 2-year PFS for the AUC were 0.731 and 0.741. The Brier score obtained from the model was 0.157 and 0.203

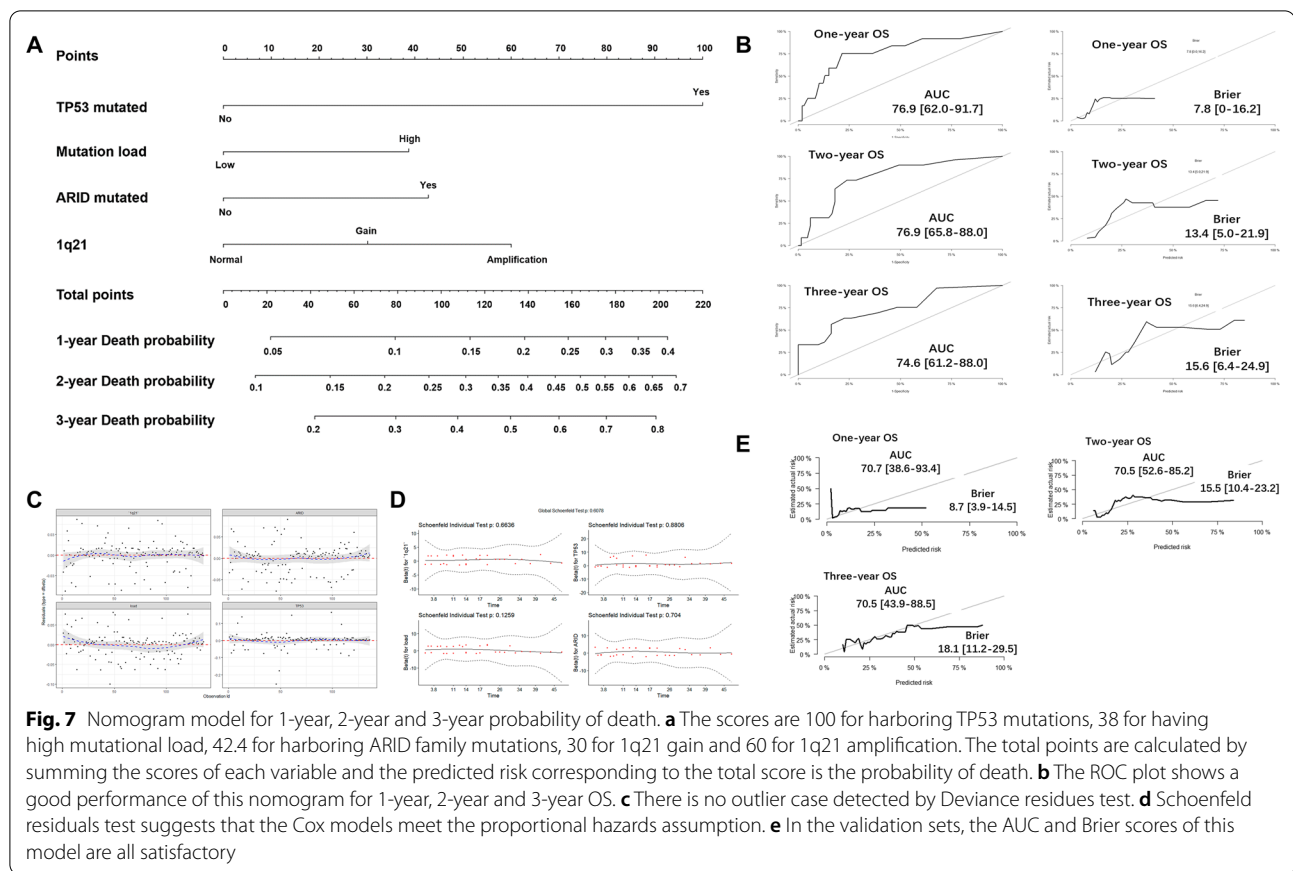


which confirmed the calibration was acceptable (Fig. 6b). There was no outlier case detected by Deviance residues test (Fig. 6c), and Schoenfeld residuals test suggested that the Cox models met the proportional hazards assumption (Fig. 6d). With internal validation, the AUC (0.708 for 1-year PFS and 0.706 for 2-year PFS) and Brier scores (0.168 for 1-year PFS and 0.228 for 2-year PFS) of this model were all satisfactory (Fig. 6e).

Likewise, we developed a nomogram model for OS by using TP53 mutations, mutation load, ARID family mutations and 1q21 copy numbers (Fig. 7a). The predicted risk corresponding to the total score was the probability of death. The model performed well in predicting 1-year, 2-year and 3-year OS for the AUC that were 0.769, 0.769 and 0.746. The Brier score obtained from the model was 0.078, 0.134 and 0.156 which confirmed the calibration was acceptable (Fig. 7b). No outliers were identified by Deviance residues test (Fig. 7c) and Schoenfeld residuals test (Fig. 7d). With internal validation, the AUC (0.707 for 1-year OS, 0.705 for 2-year OS and 0.705 for 3-year OS) and Brier scores (0.087 for 1-year OS, 0.155 for 2-year OS and 0.181 for 3-year OS) confirmed this model was eligible for predicting OS (Fig. 7e).

Discussion

MM is a highly heterogeneous cancer whose development is driven by numerous factors, including cytogenetic abnormalities, changes in the bone marrow microenvironment and aberrant immune regulation [17]. A growing number of studies have revealed that epigenetic alterations, including dysfunction of CRs, play a critical role in the occurrence and progression of MM. Tessoulin et al. [18] found human myeloma cell lines displayed high mutation rates of epigenetic modifiers. The mutation rate of *TET2* (a DNA methylation regulator) was 15%, and the mutation rate of *SETD2* (a chromatin remodeler) was 6%. Ohguchi et al. [19] identified a KDM3A-KLF2-IRF4 axis that maintained myeloma cell survival and dysfunction of KDM3A (a chromatin remodeler) was toxic to MM cells in vitro and in vivo. However, the role of CRs in MM remains poorly studied and relevant clinical studies are particularly lacking. The present study was the first to systematically analyzed the clinical significance of CRs mutations in MM and provided a direction for subsequent mechanistic studies.



KRAS was the most commonly mutated gene in our study followed by *NRAS*, which was consistent with existing reports [3–5]. Among the top 15 genes with the highest mutation rate, *KMT2C*, *KMT2D* and *EP300* are CRs. *KMT2C/KMT2D* are involved in the methylation of H3K4 and p300 (encoded by *EP300*) regulates H3K27 acetylation [9]. It was interesting that *KMT2C*, *KMT2D* and p300 could form a complex called K_{CD} COMs to activate transcription under normal conditions [16], indicating that the dysfunction of K_{CD} COMs might be involved in the pathogenesis of MM. Previous studies demonstrated that mutations affecting components of the K_{CD} COMs were associated with the development of other tumors, including breast cancer, lung cancers and B cell lymphomas [20–23]. The possible mechanism was considered that dysfunction of K_{CD} COMs could promote oncogene expression (e.g., *BCL2*) and decrease tumor suppressor gene expression (e.g., *TP53* and *SOCS3*) [20, 22, 23]. However, the role of K_{CD} COMs in MM is not clear and subsequent functional experiments are needed to be further conducted.

We identified, for the first time, harboring mutations of ARID gene family as a predictor of poor prognosis in MM. It could be incorporated into ISS or R-ISS staging

system for the optimal stratification of patients with MM. Subsequently, we constructed nomogram models to predict PFS and OS based on ARID family mutations and the model performed well with good discrimination and calibration. There are seven subfamilies and 15 members in ARID gene family (see Additional file 1: Table S3 for details). All members contain a DNA-binding domain and have the ability to regulate transcription [24]. Much research effort has been focused on the dysfunction of ARID gene family in a variety of tumors. Loss of *ARID1A* expression was related to poor outcomes in ovarian clear cell carcinoma [25]. Mutations of *ARID5B* might be a potential cofactor in patients with *ETV6*-linked leukemia predisposition [26]. Frequent deletions of *JARID2* promoted the transformation of chronic myeloid malignancies to leukemia [27]. However, the role of this gene family in MM is still unknown, which should be explored in further study.

Currently, there is consensus that chromosomal abnormalities are present before gene mutations during the pathogenesis of myeloma [6]. Therefore, marked correlations were observed between chromosomal abnormalities and gene mutations in MM [5, 28]. We found *FGFR3* mutations were associated with t(4; 14)

and TP53 mutations were associated with del(17p), which were in line with other studies [28]. However, the association of gain(1q21) with mutations of *EP300* has not been reported in MM before. Gain/amplification of 1q21 copies is considered a secondary genomic event during MM progression. The incidence of gain(1q21) increased from monoclonal gammopathy of undetermined significance to relapsed MM [29]. So, we speculate that mutations of *EP300* might promote the tumor progression of MM. On the other hand, del(17p) is a well-established marker of poor prognosis in MM, but our results demonstrated the combination of del(17p) and *TP53* mutations could better predict outcomes of MM patients. Patients who had both del(17p) and *TP53* mutations presented with a very poor prognosis and should be paid more attention to by clinicians.

However, there were still some limitations in the present study. It was regretful that we did not perform external validation due to the absence of suitable data. The effectiveness of this model will be externally validated using multi-center data in the future. Moreover, we used the data of normal person in databases but not the same patient's non-tumor tissue to serve as controls, so we were unable to completely eliminate germline variants. By using strict bioinformatic analysis we tried to minimize the influence of germline variants.

In conclusion, our findings emphasized the importance of CRs mutations in NDMM patients and the mutations affecting K_{CD} COMs might promote the development of MM. High mutational load and harboring mutations in the ARID gene family were novel predictors of adverse prognosis in MM. Prognostic models based on 1q21 copies, ARID gene mutations, extramedullary disease, mutational load and TP53 mutations were commendably prognostic evaluation methods for OS and PFS and provided a reference for clinical evaluation.

Abbreviations

AUC: Area under the receiver operator curve; BAM: Binary Alignment Map; BP: Biological process; CC: Cellular component; CR: Chromatin regulator; DS: Durie–Salmon; EMD: Extramedullary disease; GO: Gene ontology; Hb: Hemoglobin; ISS: International Staging System; LDH: Lactate dehydrogenase; MF: Molecular function; MM: Multiple myeloma; NDMM: Newly diagnosed multiple myeloma; NGS: Next-generation sequencing; OS: Overall survival; PD: Disease progression; PFS: Progress-free survival; R-ISS: Revised International Staging System; ROC: Receiver operator curve; SAM: Sequence Alignment/Map; SNP: Single nucleotide polymorphism.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-022-01314-7>.

Additional file 1. Supplementary Material.

Additional file 2. 387 genes selected for sequencing in this study.

Acknowledgements

We thank the patients and their families for their participation.

Author contributions

LC, JL and LZ designed the study. LZ, RZ, JW, YC, CQ, QS, YJ and XS analyzed and interpreted data. LZ and LC wrote and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No. 82070223).

Availability of data and materials

The data used or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval and consent to participate

The study was approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University. The use of tumor specimens was approved by the Ethics Committee, and informed consent was obtained from all patients.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 2 November 2021 Accepted: 16 July 2022

Published online: 23 July 2022

References

- Röllig C, Knop S, Bornhäuser M. Multiple myeloma. *Lancet*. 2015;385(9983):2197–208.
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214–8.
- Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun*. 2014;5:2997.
- Lohr JG, Stojanov P, Carter SL, Cruz-Gordillo P, Lawrence MS, Auclair D, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*. 2014;25(1):91–101.
- Walker BA, Boyle EM, Wardell CP, Murison A, Begum DB, Dahir NM, et al. Mutational spectrum, copy number changes, and outcome: results of a sequencing study of patients with newly diagnosed myeloma. *J Clin Oncol*. 2015;33(33):3911–20.
- Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol*. 2017;14(2):100–13.
- Ordoñez R, Kulis M, Russiñol N, Chapaprieta V, Carrasco-Leon A, García-Torre B, et al. Chromatin activation as a unifying principle underlying pathogenic mechanisms in multiple myeloma. *Genome Res*. 2020;30(9):1217–27.
- Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell*. 2012;150(1):12–27.
- Zhang P, Torres K, Liu X, Liu C-G, Pollock RE. An overview of chromatin-regulating proteins in cells. *Curr Protein Pept Sci*. 2016;17(5):401–10.
- Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos M-V, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*. 2014;15(12):e538–48.
- Durie BG, Salmon SE. A clinical staging system for multiple myeloma. Correlation of measured myeloma cell mass with presenting clinical features, response to treatment, and survival. *Cancer*. 1975;36(3):842–54.
- Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, et al. Revised international staging system for multiple myeloma:

- a report from international myeloma working group. *J Clin Oncol*. 2015;33(26):2863–9.
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
 14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernyt-sky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
 15. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
 16. Lavery WJ, Barski A, Wiley S, Schorry EK, Lindsley AW. KMT2C/D COMPASS complex-associated diseases [KCOM-ADs]: an emerging class of congenital regulopathies. *Clin Epigenetics*. 2020;12(1):10.
 17. Kumar SK, Rajkumar V, Kyle RA, van Duin M, Sonneveld P, Mateos MV, et al. Multiple myeloma. *Nat Rev Dis Primers*. 2017;3:17046.
 18. Tessoulin B, Moreau-Aubry A, Descamps G, Gomez-Bougie P, Maiga S, Gaignard A, et al. Whole-exon sequencing of human myeloma cell lines shows mutations related to myeloma patients at relapse with major hits in the DNA regulation and repair pathways. *J Hematol Oncol*. 2018;11(1):137.
 19. Ohguchi H, Hideshima T, Bhasin MK, Gorgun GT, Santo L, Cea M, et al. The KDM3A-KLF2-IRF4 axis maintains myeloma cell survival. *Nat Commun*. 2016;7:10258.
 20. Fagan RJ, Dingwall AK. COMPASS Ascending: Emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. *Cancer Lett*. 2019;458:56–65.
 21. Gala K, Li Q, Sinha A, Razavi P, Dorso M, Sanchez-Vega F, et al. KMT2C mediates the estrogen dependence of breast cancer through regulation of ER α enhancer function. *Oncogene*. 2018;37(34):4692–710.
 22. Ortega-Molina A, Boss IW, Canela A, Pan H, Jiang Y, Zhao C, et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med*. 2015;21(10):1199–208.
 23. Zhang J, Dominguez-Sola D, Hussein S, Lee J-E, Holmes AB, Bansal M, et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med*. 2015;21(10):1190–8.
 24. Lin C, Song W, Bi X, Zhao J, Huang Z, Li Z, et al. Recent advances in the ARID family: focusing on roles in human cancer. *Onco Targets Ther*. 2014;7:315–24.
 25. Katagiri A, Nakayama K, Rahman MT, Rahman M, Katagiri H, Nakayama N, et al. Loss of ARID1A expression is related to shorter progression-free survival and chemoresistance in ovarian clear cell carcinoma. *Mod Pathol*. 2012;25(2):282–8.
 26. Karastaneva A, Nebral K, Schlagenhaut A, Baschin M, Palankar R, Juch H, et al. Novel phenotypes observed in patients with-linked leukaemia/familial thrombocytopenia syndrome and a biallelic risk allele as leukaemogenic cofactor. *J Med Genet*. 2020;57(6):427–33.
 27. Puda A, Milosevic JD, Berg T, Klampfl T, Harutyunyan AS, Gisslinger B, et al. Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. *Am J Hematol*. 2012;87(3):245–50.
 28. Walker BA, Mavrommatis K, Wardell CP, Ashby TC, Bauer M, Davies FE, et al. Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood*. 2018;132(6):587–97.
 29. Hanamura I. Gain/amplification of chromosome Arm 1q21 in multiple myeloma. *Cancers*. 2021;13(2):256.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

