



Identification of prognostic and driver gene mutations in acute myeloid leukemia by a bioinformatics analysis

Lifang Li¹, Jingxiong Ruan², Ning Zhang², Jia Dai³, Xin Xu², Xudong Tian², Jian Hu²

¹Emergency Department, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; ²Clinical Laboratory, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; ³Clinical Laboratory, Shaanxi Provincial Cancer Hospital, Xi'an, China

Contributions: (I) Conception and design: L Li, J Hu; (II) Administrative support: X Tian; (III) Provision of study materials or patients: J Ruan; (IV) Collection and assembly of data: N Zhang; (V) Data analysis and interpretation: L Li, J Dai, X Xu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jian Hu, PhD, MD. Clinical Laboratory, First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, China. Email: hobbyhujian@xjtufh.edu.cn.

Background: The representative gene mutation in the prognostic groups of acute myeloid leukemia (AML) patients is not yet known. The purpose of this study is to identify representative mutations that can help physicians better predict patient prognosis and thus develop better treatment plans.

Methods: The Cancer Genome Atlas (TCGA) database was queried for clinical and genetic information, and individuals with AML were classified into 3 groups based on their AML Cancer and Leukemia Group B (CALGB) cytogenetic risk category. Each group's differentially mutated genes (DMGs) were evaluated. Simultaneously, Gene Ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were used to assess the function of DMGs within the 3 distinct groups. We used the driver status and protein impact of DMGs as additional filters to narrow down the list of significant genes. Cox regression analysis was used to examine the survival features of gene mutations in these genes.

Results: A cohort of 197 AML patients were divided into 3 groups according to their prognostic subtype: favorable (n=38), intermediate (n=116) and poor (n=43). There were significant differences in age and tumor metastasis rates among the three groups of patients. Patients in the favorable group had the highest rate of tumor metastasis. Different prognosis groups' DMGs were detected. The DMGs were examined for the driver and harmful mutations. We considered the DMGs that had driver and harmful mutations and that affected the survival outcomes in the prognostic groups as the key gene mutations. The group with a favorable prognosis carried specific gene mutations for *KIT* and *WT1*. The intermediate prognostic group contained mutations in the genes *IDH2*, *NRAS*, *NPM1*, *FLT3*, *RUNX1*, *DNMT1A*, and *MUC16*. In the group with a poor prognosis, the representative genes were *KRAS*, *TP53*, *IDH1*, *IDH2*, and *DNMT3A*, with *TP53* mutations substantially correlated with overall patient survival.

Conclusions: We performed the systemic analysis of the gene mutation in patients with AML and identified representative and driver mutations between the prognostic group. Identification of representative and driver mutations between the prognostic group can help predict the prognosis of patients with AML and guide treatment decisions.

Keywords: Acute myeloid leukemia (AML); prognosis; gene mutation; differentially mutated genes (DMGs); driver mutation

Submitted Apr 07, 2023. Accepted for publication Jun 13, 2023. Published online Jun 21, 2023.

doi: 10.21037/tcr-23-587

View this article at: <https://dx.doi.org/10.21037/tcr-23-587>

Introduction

Acute myeloid leukemia (AML) is a condition involving hematopoietic stem cells and is caused by genetic changes in blood cell progenitors that leads to the overproduction of neoplastic clonal myeloid stem cells (1). AML begins in the bone marrow (the soft interior of some bones that produces new blood cells), and in most cases, it also enters the bloodstream, where it can spread to the lymph nodes, liver, spleen, central nervous system (brain and spinal cord), and testes (2). In the majority of cases, AML develops in leukocytes (excluding lymphocytes) although it can also occur in other types of blood-forming cells (3). Some cases will be detected by regular blood tests, while others will manifest as clinical problems such as infection, bleeding, or widespread disease (4). Bone marrow examination is essential for confirming the diagnosis and acquiring tissue for study to better define the AML subtype and prognostic severity (1,5,6). In addition, more studies are screening for diagnostic markers of AML based on bioinformatics analysis (7). The prognostic biomarkers of AML have been studied extensively. New AML biomarkers have been shown to enhance the molecular understanding of the disease, and to provide significant benefits for screening, diagnosis, prognosis and monitoring of AML, as well as for predicting the treatment response of each patient (2). However, there

are still knowledge gaps and limitations in prior studies (2). Advancing age (≥ 60 years) at the time of AML diagnosis has been associated consistently with worse survival in cooperative group trials and in population-based studies (8).

The AML CALGB cytogenetics risk category was developed by the Cancer and Leukemia Group B (CALGB). It classifies AML into 3 prognostic risk groups: favorable, intermediate, and poor, based on the presence and type of cytogenetic abnormalities and karyotypic complexity (9,10). For the poor group, the probability of complete response is one-fourth and one-twelfth that of the intermediate and favorable prognostic groups, respectively;

the recurrence rate is 3.0 times higher than the intermediate prognostic group and 4.4 times higher than the favorable prognostic group; and the mortality rate is 2.1 times higher than the intermediate prognostic group and 4.3 times higher than the favorable prognostic group (11). Intermediate I is a normal karyotype, whereas intermediate II is an aberrant karyotype (12).

In order to understand the somatic mutations properties of these 3 prognosis groups of AML we investigated the mutated genes, mutation frequency, mutation characteristics, and mutation types in samples of the distinct prognostic subtypes of AML. We completed prognostic analysis, cancer driver gene prediction, mutation identification, and visualization of the mutation sites of the representative genes and subsequently analyzed the targeted chemical drugs for these mutations. This study provides a scientific and theoretical basis for understanding the mechanism of AML development and the specific mutations in individuals with varying prognostic outcomes, as well as the development of their targeted therapies. We present this article in accordance with the STREGA reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-587/rc>).

Methods

Data sets

The Cancer Genome Atlas database (TCGA) was used to retrieve the clinical data and somatic mutation information of patients with AML. For details, we obtained 197 patients with overall survival information (death from any cause). The clinical characteristics including gender, race, age, the FAB (French-American-British) classification, and cytogenetic risk group (established by the European LeukemiaNet, related with the prognosis of patients).

Highlight box

Key findings

- We identified specific differentially mutated genes (DMGs) for prognostic groups of patients with acute myeloid leukemia (AML).
- The group with a favorable prognosis carried specific gene mutations, including for *KIT* and *WT1*.
- The intermediate prognostic group contained mutations in the genes *IDH2*, *NRAS*, *NPM1*, *FLT3*, *RUNX1*, *DNMT1A*, and *MUC16*.
- In the group with a poor prognosis, the representative genes were *KRAS*, *TP53*, *IDH1*, *IDH2*, and *DNMT3A*, with *TP53* mutations substantially correlated with overall patient survival.

What is known and what is new?

- AML is divided into 3 prognostic risk groups: favorable, intermediate, and poor.
- We investigated the somatic mutational properties of these 3 AML prognosis groups and the mutation types in samples of distinct prognostic subtypes.

What is the implication, and what should change now?

- This study provides a theoretical basis for the development of targeted therapies of AML based on gene-specific mutations.

Cytogenetic risk group consisted of favorable (n=38), intermediate (n=116), and poor (n=43).

This study involved 197 samples and detected 9,905 somatic mutations. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Identification of differentially mutated genes

The *clinicalEnrichment* function of the “maftools” package in R software (The R Foundation of Statistical Computing) was used to determine the differentially mutated genes (DMGs) for the favorable, moderate, and poor prognostic subtypes of AML. A P value <0.05 was deemed to indicate a statistically significant difference.

Functional annotation

The R package “clusterProfiler” was used to annotate Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for the acquired genes. GO annotation was performed based on biological process (BP), molecular function (MF), and cellular component (CC) categories. Fisher exact test was used to evaluate the significance level (P value) for identifying GO with substantial enrichment of distinct genes. A P value <0.01 indicated significant enrichment, while other values indicated nonsignificant enrichment. Meanwhile, for the KEGG database, a P value <0.05 was used to indicate signal pathway enrichment of significantly differentially expressed genes.

Analysis of driver genes using mutation location clustering

Mutations in oncogenes often happen in specific domains of proteins (known as mutation hot spots), and these mutations increase the growth or proliferation of cancer cells. We used the OncodriveCLUST method to cluster the mutation sites of gene bases and find cancer genes. The key information we calculated included the number of mutation hot spots, the number of mutations in the hot spots that were clustered, the length of amino acids corresponding to the protein, the fraction of clustered mutations in all mutations of the gene, the P value, and the false discovery rate (FDR) values. The lower the scores were, the higher the driving force was.

Mutation damage assessment with PROVEAN and SIFT software

We used the PROVEAN and SIFT database to find

homologous proteins, and selected protein sequences with high similarity and consistent function for multisequence PSI-BLAST alignment to evaluate the conserved protein sites. The risk was assessed using the PROVEAN/SIFT database score.

Survival analysis

The information on overall survival (OS) was taken from TCGA database. The hazard ratio (HR) was calculated using Kaplan-Meier analysis, and the survival curve was drawn. A P value <0.05 was considered to indicate significant correlation of the gene mutation with patient prognosis. Based on AML patient survival data, the R package “survival” was used to conduct univariate Cox regression analysis on the significant DMGs in the favorable, intermediate, and poor prognostic groups, respectively. We screened for genes significantly correlated with the OS of patients (P<0.01). Least absolute shrinkage and selection operator (LASSO) Cox regression analysis was performed using the R package “glmnet” on genes that were significantly correlated with OS (13). The 10-fold cross-validation selection procedure was used to find the appropriate punishment parameter lambda in order to eliminate genes with strong correlation and reduce model complexity. With R package “survival”, a multivariate Cox regression analysis was performed on the genes identified by LASSO analysis to determine the prognostic genes in the favorable, intermediate, and poor subtypes, respectively.

Targeted drugs for gene mutations

Targeted drugs for gene mutations, including *KRAS*, *TP53*, *KIT* and *DNMT3A* were screened according to the clinical data on cancer mutations provided by the Clinical Interpretation of Variants in Cancer (CIViC) database (<https://civicdb.org/home>).

Results

Clinical characteristics of patients and prognosis

Data from a total of 197 patients with AML were obtained from TCGA database, and the patients were classified into 3 groups according to their CALGB prognostic subtypes: favorable (38 patients), intermediate (116 patients), and poor (43 patients). The majority of patients were Caucasian. The age of the patients and the rate of tumor metastasis in

Table 1 Comparison of clinical characteristics of the patients between the favorable, intermediate, and poor prognostic groups

Characteristics	Prognostic group			P	Test
	Favorable (n=38)	Intermediate (n=116)	Poor (n=43)		
OS, median (IQR), days	623.50 (222.00, 1401.75)	304.00 (90.00, 678.00)	243.00 (31.00, 441.00)	2.90E-03	nonnorm
Gender, n (%)				4.94E-01	
Female	18 (47.4)	55 (47.4)	16 (37.2)		
Male	20 (52.6)	61 (52.6)	27 (62.8)		
Race, n (%)				7.31E-01	fisher_exact
Asian	0 (0.0)	2 (1.7)	0 (0.0)		
Black	1 (2.7)	8 (7.0)	4 (9.3)		
White	36 (97.3)	105 (91.3)	39 (90.7)		
NA	1 (2.6)	1 (0.9)	0 (0.0)		
Age (years), median (IQR)	49.50 (35.25, 58.50)	58.50 (45.00, 67.00)	63.00 (50.50, 69.50)	5.43E-03	nonnorm
Fab_category, n (%)				1.34E-12	fisher_exact
M0	0 (0.0)	9 (7.8)	10 (23.8)		
M1, M2, M4, M5	19 (50.0)	101 (87.8)	30 (71.4)		
M3	18 (47.4)	3 (2.6)	0 (0.0)		
M6, M7	1 (2.6)	2 (1.7)	2 (4.8)		
NA	0 (0.0)	1 (0.9)	1 (2.3)		

OS, overall survival; IQR, interquartile range; NA, not available.

the 3 groups were significantly different (Table 1). However, patients in the favorable group had the highest rate of tumor metastasis, although this group had a better prognosis and the longest survival. Comparison of OS among the 3 groups showed that the favorable group had the best prognosis, the intermediate group had a slightly better prognosis, and the poor group had the worst prognosis, which was consistent with the prognosis of CALGB prognostic classification (Figure 1A).

Landscape of gene somatic mutation of AML

We analyzed the somatic mutations in 197 patients with AML. Overall, missense mutations were the most common in AML, and the remaining other mutation types were very rare. Single-nucleotide mutations were the most common mutation mode, and the most common nucleotide substitution was cytosine ribonucleotide substitution for thymine ribonucleotide (C>T), followed by thymine ribonucleotide substitution for cytosine ribonucleotide (T>C). The median number of somatic mutations in

patients with AML was 10. The most common mutated genes in AML were *DNMT3A* (13%) and *FLT3* (11%), followed by *RUNX1* (9%), *NPM1* (9%), *TTN* (8%), *MUC16* (8%), *TP53* (8%), *IDH2* (8%), *KIT* (6%), and *NRAS* (6%) (Figure 1B, Figure S1).

The poor prognostic group had the highest mutation frequency (median =13.5), the favorable group had the second highest (median =10), and the intermediate group had the lowest (median =8) (Figure 1C). The *TP53* gene had the highest mutation frequency in all subtypes and was followed by the *TTN* gene. Based on the comparative analysis of mutation frequency and number of mutations, we obtained the top 10 genes in each prognostic group (Figure S2). In the favorable prognostic group, *KIT* (25%) and *TTN* (17%) had the highest mutation frequency, followed by *ZFH4* (12%), *RAPGEF2* (12%), *GRIK4* (12%), *MUC16* (8%), *UBR4* (8%), *SPEN* (8%), *AMT* (8%), and *AHNAK* (8%). Missense mutation was predominant in all genes except the *KIT* gene, which was more complex and dominated by in-frame insertions (ins), and the rest were missense mutations, nonsense mutations, and in-frame

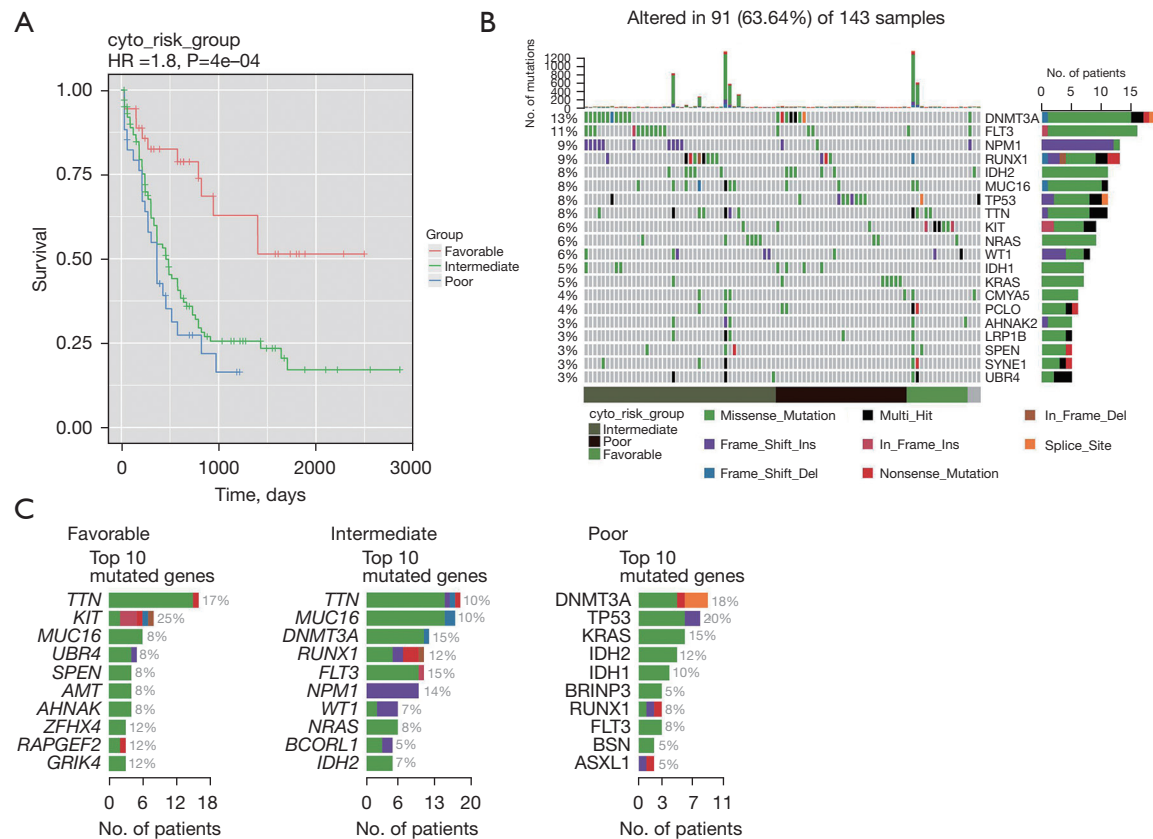


Figure 1 Survival and somatic mutations in 197 patients with AML. (A) Overall survival of the 3 prognostic groups. (B) Heat map showing the genes and mutation types of the somatic mutations in patients with AML. Bottom color block: prognostic groups. Top bar: number of mutations in the sample. Right ID: gene IDs of the 20 most frequent mutations. Left side: frequency of mutations in the corresponding gene. (C) Bar-plot showing the top 10 mutated genes and percentage of mutation occurrence of the 3 prognostic groups. AML, acute myeloid leukemia.

deletions (dels). In the intermediate prognostic group, *FLT3* (15%), *DNMT3A* (15%), and *NPM1* (14%) had the highest mutation frequencies, followed by *RUNX1* (12%), *TTN* (10%), *MUC16* (10%), *NRAS* (8%), *IDH2* (7%), *WT1* (7%), and *BCORL1* (5%). Mutated genes were dominated by missense mutations, except for *NPM1* and *WT1* mutations. All mutations in *NPM1* were frame-shift ins, and *WT1* mutations were also dominated by frame-shift ins. In the poor prognostic group, *TP53* (20%) and *DNMT3A* (18%) had the highest mutation frequencies, followed by *KRAS* (15%), *IDH2* (12%), *IDH1* (10%), *RUNX1* (8%), *FLT3* (8%), *BRINP3* (5%), *BSN* (5%), and *ASXL1* (5%). The mutation types of all genes except the *ASXL1* gene were dominated by missense mutation, and the mutation types of the *ASXL1* gene were mainly frame-shift ins and nonsense mutations.

DMGs in patients with AML with different prognosis

To better investigate the prognosis-related mutation, we compared differences in mutation of patients with AML with different prognoses. We obtained a total of 45 genes with significant differences in the number of mutations between the 3 groups, including 15 genes for each group respectively (Figure 2A, Table S1). In the favorable prognostic group, the *KIT*, *TTN*, *PCLO*, and *WT1* mutations had the most prominent occurrence. In the intermediate prognostic group, the *NPM1*, *FLT3*, *RUNX1*, *NIRAS*, *TTN*, *MUC16*, *DNMT3A*, and *WT1* mutations had the most frequent occurrence. In the poor risk group, mutations of *TP53*, *KRAS*, *IDH1*, *IDH2*, and *DNMT3A* were the most common.

The obtained representative genes of each group

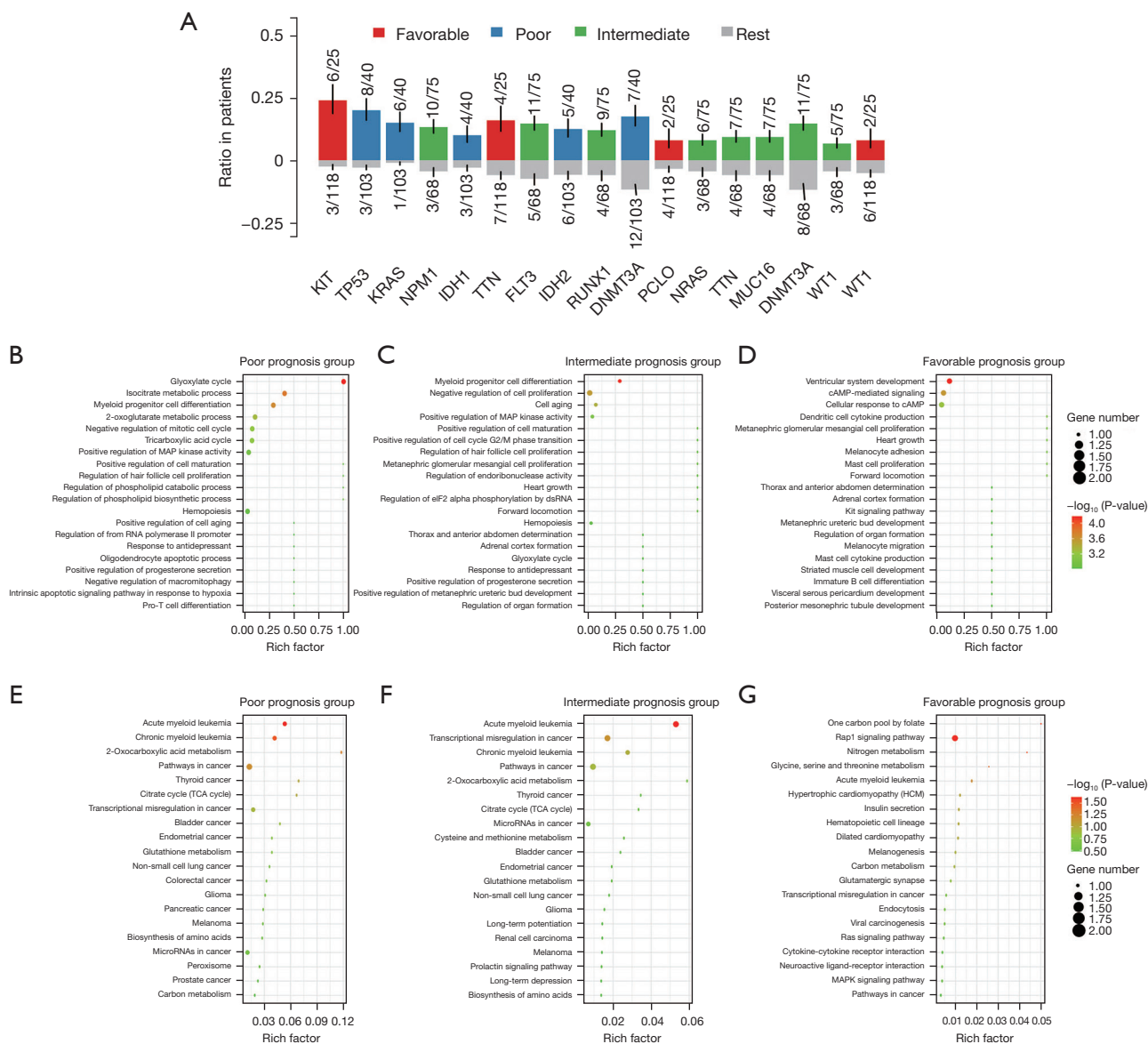


Figure 2 Functional enrichment analysis of genes and genes with significantly different mutations between the 3 prognostic groups. (A) Genes with P value <0.05 for differential mutation significance in a given group are shown, sorted from left to right by significance of the difference in the subgroups. The horizontal axis shows the gene names, the vertical axis shows the proportion of mutations in different subgroups, and the bar chart colors show the subgroups corresponding to the figure notes on the right. (B-G) GO and KEGG pathway enrichment analysis of the differentially mutated genes. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAPK, Mitogen-Activated Protein Kinase.

were annotated based on GO terms (Figure 2B-2D). GO enrichment results showed that in the poor prognostic subgroup, DMGs were involved in glyoxylate cycle, isocitrate metabolic process, myeloid progenitor cell differentiation and other molecular processes, mitochondria, PR-DUB complex, presynaptic cytoskeletal

matrix, nuclear matrix, and isocitrate dehydrogenase. For the intermediate prognostic group, DMGs were enriched in myeloid progenitor cell differentiation, negative regulation of cell proliferation, cell aging, structural molecular activity conferring elasticity, DNA (cytosine-5-) methyltransferase activity, acting on CpG substrates, and

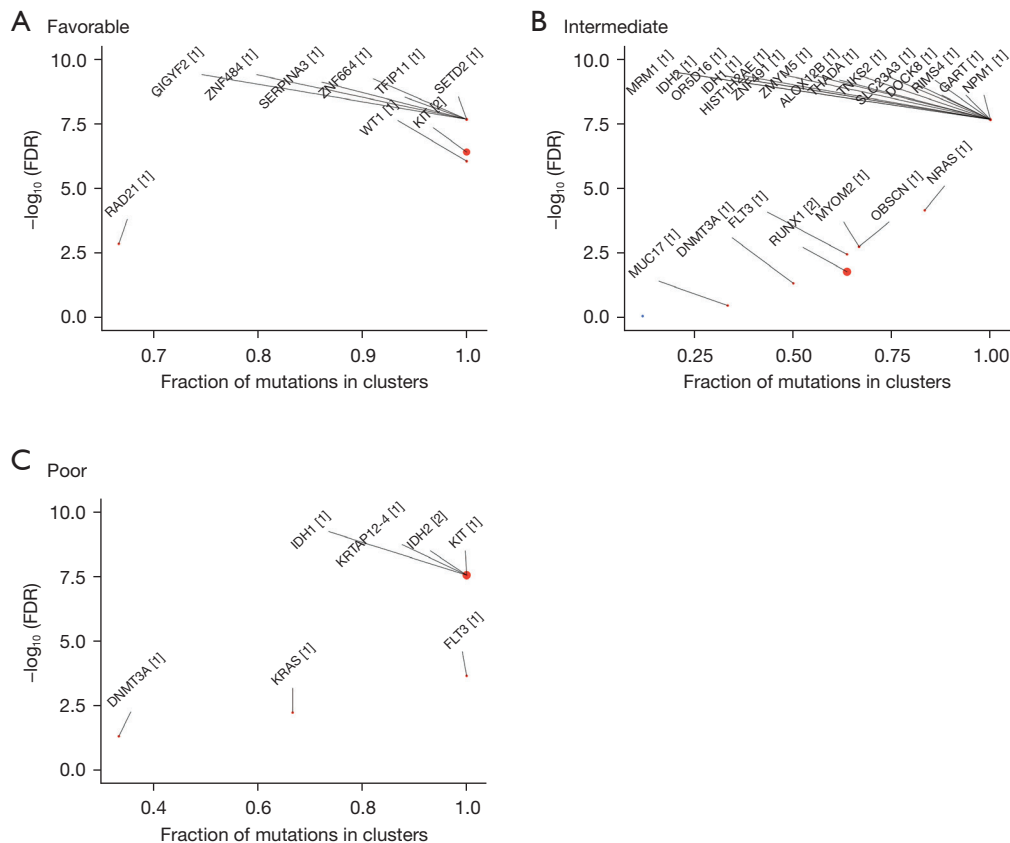


Figure 3 Analysis of driver genes based on mutation location clustering for each prognostic group. The red scatters represent significant ($P < 0.05$) driver genes in the group, the label format is gene name (number of mutation hotspots), the horizontal axis represents the frequency of mutations clustered on hotspots relative to all mutations, and the vertical axis represents $-\log_{10}$ (P value). FDR, false discovery rate.

isocitrate dehydrogenase (NADP⁺) activity. In the favorable prognostic group, DMGs were enriched in ventricular system development, cAMP-mediated signaling, cellular response to cAMP, and presynaptic cytoskeletal matrix assembled at the active zone and synapse.

The obtained representative genes of each group were annotated based on pathway annotation terms derived from the KEGG database (Figure 2E-2G). The results showed that the representative genes in poor prognostic group were enriched in the pathways of AML, chronic myeloid leukemia, and 2-oxocarboxylic acid metabolism. In the intermediate prognostic group, they were related to the pathways of AML, transcriptional misregulation in cancer, and chronic myeloid leukemia. In the favorable prognostic group, there was enrichment for one-carbon pool by folate, Rap1 signaling pathway, nitrogen metabolism, glycine, serine, and threonine metabolism.

Driver mutations based on the mutation clustering algorithm

We used the OncodriveCLUST algorithm to distinguish driver genes in each prognostic group (Figure 3). The algorithm is designed to detect mutated genes clustered in specific regions of the amino acid sequence, which often occur in specific protein regions or active sites. The driver genes in the poor prognostic group were *IDH1*, *IDH2*, *KIT*, *KRTAP12-4*, *FLT3*, *KRAS*, and *DNMT3A*. The driver genes in the intermediate prognostic group were *ALOX12B*, *DOCK8*, *GART*, *HIST1 H2AE*, *IDH1*, *IDH2*, *MRM1*, *NPM1*, *OR5D16*, *RIMS4*, *SLC23A3*, *THADA*, *TNKS2*, *ZMYM5*, *ZNF491*, *NRAS*, *MYOM2*, *OBSCN*, *FLT3*, *RUNX1*, *DNMT3A*, *MUC17*, and *MUC16*. The driver genes in the favorable prognostic group were *GIGYF2*, *SERPINA3*, *SETD2*, *TFIP11*, *ZNF484*, *ZNF664*, *KIT*, *WT1*, and *RAD21*.

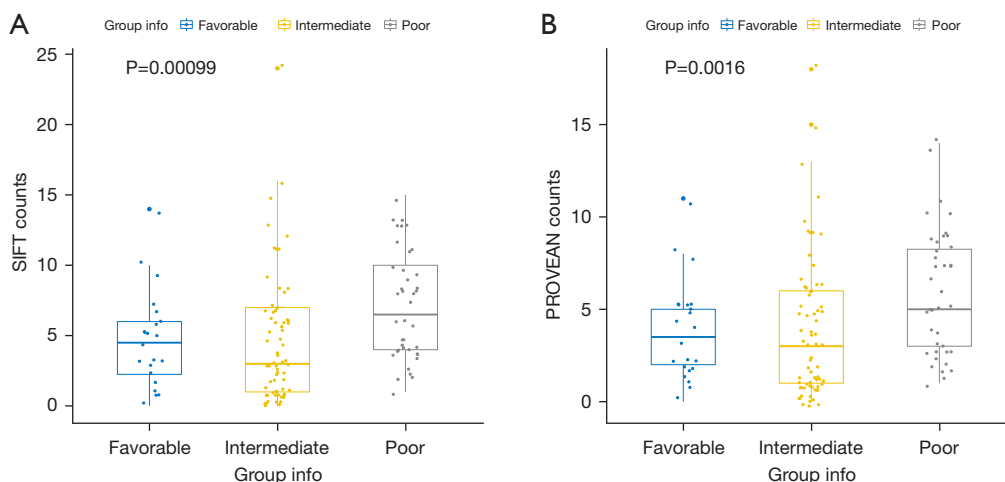


Figure 4 The box plot showing the harmfulness of mutations in patients according to the prognostic group. The left panel (A) is the result predicted by SIFT and the right panel (B) is the result predicted by PROVEAN.

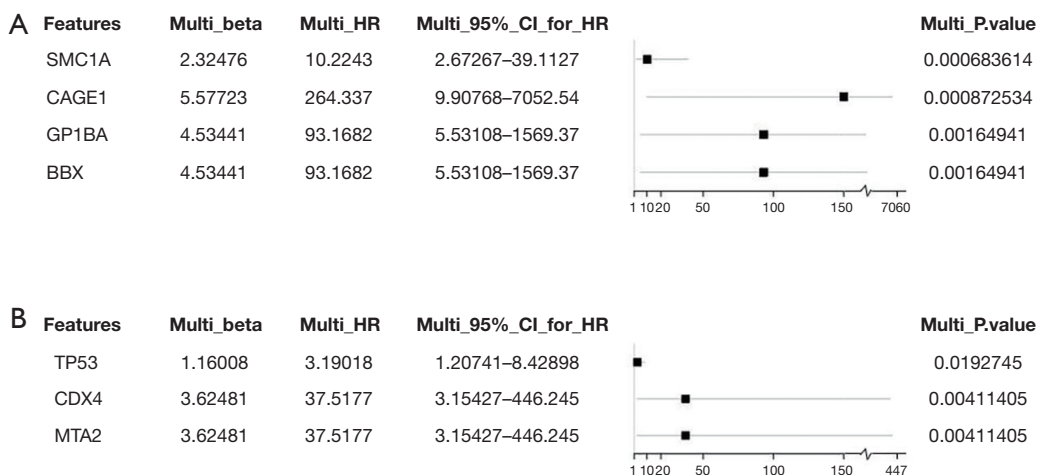


Figure 5 The best Cox multivariate regression model obtained with the LASSO algorithm. (A) Multivariate Cox regression of the gene-based hazard model for the intermediate prognostic group. (B) Multivariate Cox regression of gene-based hazard model for the poor prognostic group. The horizontal line shows the confidence interval, and the dots represent the hazard ratio.

Mutational deleteriousness analysis based on the PROVEAN and SIFT algorithms

PROVEAN/SIFT algorithms can assess the deleteriousness of mutations based on protein conservativeness and tolerance for amino acid substitutions. A total of 7,581 genes with deleterious mutations were obtained by using PROVEAN and SIFT deleterious analysis in AML. Among the different prognostic groups, the most number of deleterious genes were found in the poor prognostic group (Figure 4).

Correlation with survival for DMGs in AML

Univariate Cox regression analysis was performed for the association of deleterious mutations in DMGS significantly associated with patient survival. There was a total of 31 genes associated with survival in the poor prognostic group, 34 genes associated with survival in the intermediate prognostic group, and 14 genes associated with survival in the favorable prognostic group. The best Cox multiple regression model was selected with the LASSO algorithm. No eligible model was found for the favorable subtype. The

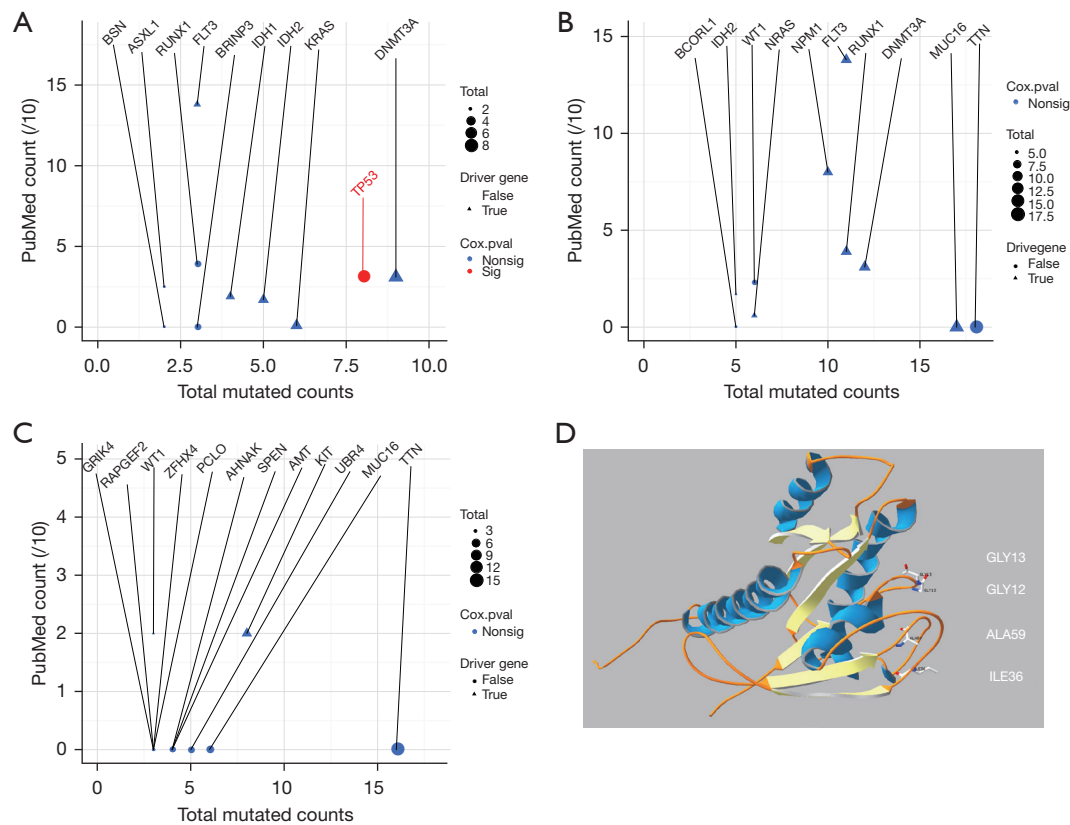


Figure 6 Scatterplot showing the representative mutated genes in each group, summarizing their mutation frequency, association with prognosis, cancer driver genes, and results from previous studies. Horizontal axis: total number of gene mutations. Vertical axis: number of previous publications. Scatter color: P value of significance of hypothesis test obtained from Cox regression analysis; the red point indicates that the mutation had a significant association with prognosis. Point shape: the triangle means indicates it is the cancer driver gene, while the circle indicates it is not. Gene name label color: red indicates the gene in the best multiple Cox regression model, and yellow indicates the significantly differentially mutated genes ($P < 0.05$). (A) Representative mutated genes for the favorable prognostic group. (B) Representative mutated genes for the intermediate prognostic group. (C) Representative mutated genes for the poor prognostic group. (D) Location of the mutant site on the 3D structure of the KRAS protein.

gene mutation-based model for the intermediate prognostic subtype consisted of *SMC1A*, *CAGE1*, *GP1BA*, and *BBX*. The model for the poor prognostic group consisted of *TP53*, *CDX4*, and *MTA2* (Figure 5).

Screening for representative mutations for the prognostic groups

To select representative mutated genes in each group, we integrated the results of mutation frequency, survival analysis, and driver gene analysis (Figure 6A-6C). Among the representative genes in the favorable prognostic group, the DMGs were *KIT*, *WT1*, *PCLO*, and *TTN*, of which the driver genes were *WT1* and *KIT*. Among the representative

genes in the intermediate prognostic group, the DMGs were *WT1*, *NRAS*, *NPM1*, *FLT3*, *RUNX1*, *DNMT3A*, *MUC16*, and *TTN*, of which the driver genes were *NRAS*, *NPM1*, *FLT3*, *RUNX1*, *DNMT3A* and *MUC16*. Among the genes represented in the poor prognostic group, the DMGs were *KRAS* and *TP53*, among which *TP53* was significantly associated with survival and *KRAS* was a cancer driver gene with significant hypermutation (15% of samples). The specific mutations of *KRAS* were p.G12D, p.G12V, p.G13D, p.I36M, and p.A59E. The mutation sites were concentrated on irregularly coiled structures, which are important regions for the functional implementation and conformation formation of protein molecules (Figure 6D). Those representative genes were enriched

Table 2 Potential targeted chemical drugs for representative mutated genes

Subtype	Gene	Variant	Disease	Drugs
Poor	<i>TP53</i>	DELETERIOUS MUTATION, MUTATION, TRUNCATING MUTATION, DNA BINDING DOMAIN MUTATION, WILD TYPE, OVEREXPRESSION, ALTERATION, CONSERVED DOMAIN MUT, G245S, M237I, R158H, R249S, R280K, R280T, R158L	Head and neck squamous cell carcinoma, esophagus squamous cell carcinoma, precursor B lymphoblastic lymphoma/leukemia, myelodysplastic syndrome, myeloma, oral squamous cell carcinoma, gastric adenocarcinoma, breast cancer, colorectal cancer, esophageal carcinoma, acute myeloid leukemia, adrenocortical carcinoma, non-small cell lung carcinoma, sarcoma, chronic lymphocytic leukemia, ovarian cancer, stomach cancer, stomach carcinoma, cancer, leukemia	Doxorubicin, "Cetuximab, Capecitabine, Oxaliplatin", "Docetaxel, Selumetinib (AZD6244)", Docetaxel, Adjuvant Chemotherapy, Pazopanib, Alemtuzumab, "Cisplatin, Carboplatin", Tamoxifen, "Cisplatin, Etoposide, Mitomycin", EAP Protocol, Nutlin-3a, AMGMDS3, RG7112
	<i>KRAS</i>	G12V, G12D, G12A	Colorectal cancer, lung cancer, lung adenocarcinoma, non-small cell lung carcinoma, pancreatic carcinoma, pancreatic cancer, tumor of exocrine pancreas, hairy cell leukemia, cancer, pancreatic ductal carcinoma, multiple myeloma, melanoma, ovarian cancer	Cetuximab, "Gefitinib, Erlotinib", "BEZ235 (NVP-BEZ235, Dactolisib), ARRY-142886", PD0332991, MK-2206, "Docetaxel, Selumetinib (AZD6244)", Vemurafenib, Crizotinib, Adoptive T-cell Transfer, "Selumetinib (AZD6244), BEZ235", Panitumumab, Gefitinib, Melphalan, Regorafenib, "Cetuximab, Panitumumab", "Panitumumab, Cetuximab"
	<i>DNMT3A</i>	R882, MUTATION	Acute myeloid leukemia, T-cell acute lymphoblastic leukemia, myelodysplastic syndrome, cancer	Daunorubicin, Idarubicin, Decitabine, "Atezolizumab, Pembrolizumab, Nivolumab"
Favorable	<i>KIT</i>	D816V	Acute myeloid leukemia, systemic macrocytosis	Midostaurin

in mitogen-activated protein kinase (MAPK) signaling pathway, PI3K-Akt signaling pathway, Sphingolipid signaling pathway, and apoptosis.

Precision medicine, which refers to the targeted provision of individually tailored prevention and treatment strategies for patients, has emerged as popular topic in the field of medicine. In the context of cancer, this involves cancer-specific mutations and the corresponding targeted therapies. We obtained potential targeted drugs for these mutations, including *KRAS*, *TP53*, and *DNMT3A* mutations in the poor subgroup and *KIT* mutations in the favorable group according to the clinical data on cancer mutations provided by the CIViC database (<https://civicdb.org/home>) (Table 2).

Discussion

The accumulation of abnormal blast cells in the bone marrow causes AML. These cells interfere with normal

blood cell production, which results in the most common underlying cause of AML-related death, bone marrow failure. AML blasts can spread to organs, especially the central nervous system and lungs. The cytogenetics and mutational status of the *NPM1*, *FLT3*, and *CEBPA* genes are associated with the development of resistance to standard therapy (12).

The European Leukemia Net (ELN) system defines 4 prognostic groups for AML. The favorable group includes patients with mutant *NPM1* without *FLT3* ITDs or *CEBPA* mutations. The intermediate 1 group (int-1, closest to favorable) includes patients with a normal karyotype who are *FLT3* ITD negative, and the intermediate 2 group includes patients with *FLT3* ITD and a normal karyotype (14).

FLT3 ITD usually affects the prognosis of AML by increasing the relapse rate rather than decreasing the complete response rate. Besides the presence of an *FLT3* ITD, the proportion of alleles with an ITD was also considered (15). The classification genes *FLT3* and *NPM1*

were also detected in our study of the significant DMGs between prognostic groups, although the mutation of *CEBPA* was not statistically different.

Most of the knowledge about the impact of NPM and FLT mutations comes from younger (under 60 years old) patients with *de novo* AML and normal karyotype (16). The frequency of NPM1 mutation decreases with age (17). However, not only do older adults constitute the majority of patients with AML, they are also more prone to demonstrate resistance. Older patients account for a higher percentage of the adverse (about 31%), intermediate (approximately 49%), and less favorable (approximately 20%) prognosis groups (14). Consequently, it is crucial to search for further particular gene mutations within these prognostic groups. This study focused on the examination of mutant genes correlated with AML prognosis categories. We found unique mutations in the prognostic categories of patients with AML.

We assessed the genes with driver and harmful mutations among these variants. In addition to mutations in genes that are prevalent in leukemia and used for classification, such as *FLT3* and *NPM1*, we also observed somatic mutations in other genes with frequent mutations, including *KRAS*, *NRAS*, and *TP53*. These genes also appear to occur often in solid tumors, but hematologic malignancies have been investigated less systematically. Mutations in *TP53* and *KRAS/NRAS* were reported to be related to resistance to current standard chemotherapy, including for azacitidine + venetoclax, decitabine + venetoclax, and low-dose cytarabine + venetoclax (18). Mutations in the *DNMT3A* and *TP53* genes are more prevalent in adult patients with AML than in younger patients, while they are nearly entirely absent in pediatric cases (19). Meanwhile, *NRAS*, *KRAS*, and *WT1* mutations are prevalent in pediatric AML (19). The adverse prognostic impact of *KRAS* mutations was confirmed in adult mixed-lineage leukemia rearranged (MLL-r) AML, although there was no significant difference in prognosis between the *KRAS* wild type and mutated type in patients with non-MLL-r AML (20). Although *TP53* mutations have been identified in only 5–10% of AML cases, they are strongly correlated with therapy-related AML and serve as a crucial prognostic indicator (21).

There is an association between the immunophenotype and prognosis in AML. Several studies have identified the relationship between antigens CD7, CD9, CD11b, CD13, CD14, CD15, CD33, CD34 and CD56 and AML

prognosis. Some studies have reported conflicting results regarding the relationship between AML prognosis. One study found that CD56 expression in AML patients was associated with poor prognosis (22). Another study found that CD34 expression in AML patients was associated with a favorable prognosis (23). However, another study found that CD34 expression in AML patients was associated with a poor prognosis (24).

We systematically investigated the specific mutations and driver mutations in 3 prognostic groups of patients; we believe these mutations can be used to focus the targeted treatment of patients with AML according to the particular subtypes. *IDH1* and *IDH2* were found to be the specific mutations and driver mutant genes for the poor prognosis subtypes, indicating that targeted therapy for *IDH1* and *IDH2* can be developed for patients with poor prognostic subtypes. Mutations in the *IDH* genes in critical metabolic enzymes result in the generation of oncometabolite 2-hydroxyglutarate, which promotes leukemogenesis by interfering with normal myeloid differentiation (25). Enasidenib and ivosidenib are potent and selective inhibitors of mutant *IDH2* and *IDH1*, respectively, serve as differentiation agents, and have demonstrated therapeutic activity in relapsed/refractory (R/R) AML (25,26). *DNMT3A* mutations are early events in the genesis of cancer and appear to be associated with poor prognosis in patients with AML, making this gene a potential target for novel therapeutics (27). They screened targeted chemical drugs including daunorubicin, idarubicin, decitabine, atezolizumab, pembrolizumab, and nivolumab for their possible inhibitory effects on AML by retrieving information from drug databases, which may aid in the precise treatment of AML with *DNMT3A* mutations.

Conclusions

We systematically identified the specific and driver mutations that served as prognostic indicators in 3 prognostic groups of patients.

Acknowledgments

Funding: This work received fund support from the Clinical Research Award of the First Affiliated Hospital of Xi'an Jiaotong University (No. XJTU1AF-CRF-2018-013) and the Natural Science Foundation of Shaanxi Province, Youth

Project (No. 2022JQ-866).

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-587/rc>

Peer Review File: Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-587/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-587/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Pelcovits A, Niroula R. Acute Myeloid Leukemia: A Review. *R I Med J* (2013) 2020;103:38-40.
- Prada-Arismendy J, Arroyave JC, Röthlisberger S. Molecular biomarkers in acute myeloid leukemia. *Blood Rev* 2017;31:63-76.
- Chopra M, Bohlander SK. The cell of origin and the leukemia stem cell in acute myeloid leukemia. *Genes Chromosomes Cancer* 2019;58:850-8.
- Saleh K, Khalifeh-Saleh N, Kourie HR. Acute myeloid leukemia transformed to a targetable disease. *Future Oncol* 2020;16:961-72.
- Reville PK, Kadia T. Individualizing Treatment for Newly Diagnosed Acute Myeloid Leukemia. *Curr Treat Options Oncol* 2020;21:34.
- Kayser S, Levis MJ. Clinical implications of molecular markers in acute myeloid leukemia. *Eur J Haematol* 2019;102:20-35.
- Chen W, Liu D, Wang G, et al. Screening diagnostic markers for acute myeloid leukemia based on bioinformatics analysis. *Transl Cancer Res* 2022;11:1722-9.
- Kantarjian H, Kadia T, DiNardo C, et al. Acute myeloid leukemia: current progress and future directions. *Blood Cancer J* 2021;11:41.
- Byrd JC, Mrózek K, Dodge RK, et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood* 2002;100:4325-36.
- Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017;129:424-47.
- Lagunas-Rangel FA, Chávez-Valencia V, Gómez-Guijosa MÁ, et al. Acute Myeloid Leukemia-Genetic Alterations and Their Clinical Prognosis. *Int J Hematol Oncol Stem Cell Res* 2017;11:328-39.
- Döhner H, Estey EH, Amadori S, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 2010;115:453-74.
- Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33:1-22.
- Estey E. Acute myeloid leukemia: 2016 Update on risk-stratification and management. *Am J Hematol* 2016;91:824-46.
- Linch DC, Hills RK, Burnett AK, et al. Impact of FLT3(ITD) mutant allele level on relapse risk in intermediate-risk acute myeloid leukemia. *Blood* 2014;124:273-6.
- Prébet T, Boissel N, Reutenauer S, et al. Acute myeloid leukemia with translocation (8;21) or inversion (16) in elderly patients treated with conventional chemotherapy: a collaborative study of the French CBF-AML intergroup. *J Clin Oncol* 2009;27:4747-53.
- Lazenby M, Gilkes AF, Marrin C, et al. The prognostic relevance of *flt3* and *npm1* mutations on older patients treated intensively or non-intensively: a study of 1312

- patients in the UK NCRI AML16 trial. *Leukemia* 2014;28:1953-9.
18. Stahl M, Menghrajani K, Derkach A, et al. Clinical and molecular predictors of response and survival following venetoclax therapy in relapsed/refractory AML. *Blood Adv* 2021;5:1552-64.
 19. Bolouri H, Farrar JE, Triche T Jr, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* 2018;24:103-12.
 20. Matsuo H, Yoshida K, Nakatani K, et al. Fusion partner-specific mutation profiles and KRAS mutations as adverse prognostic factors in MLL-rearranged AML. *Blood Adv* 2020;4:4623-31.
 21. Hunter AM, Sallman DA. Current status and new treatment approaches in TP53 mutated AML. *Best Pract Res Clin Haematol* 2019;32:134-44.
 22. Galera PK, Jiang C, Braylan R. Immunophenotyping of Acute Myeloid Leukemia. *Methods Mol Biol* 2019;2032:281-96.
 23. Ouyang G, Xu Z, Jiang D, et al. Clinically useful flow cytometry approach to identify immunophenotype in acute leukemia. *J Int Med Res* 2019;47:1483-92.
 24. Rasheed HM, Donia HM, Nadwan EA, et al. Identifying Leukemia-associated Immunophenotypes in Acute Myeloid Leukemia Patients Using Multiparameter Flow Cytometry. *Oman Med J* 2021;36:e323.
 25. Issa GC, DiNardo CD. Acute myeloid leukemia with IDH1 and IDH2 mutations: 2021 treatment algorithm. *Blood Cancer J* 2021;11:107.
 26. Martelli MP, Martino G, Cardinali V, et al. Enasidenib and ivosidenib in AML. *Minerva Med* 2020;111:411-26.
 27. Brunetti L, Gundry MC, Goodell MA. DNMT3A in Leukemia. *Cold Spring Harb Perspect Med* 2017;7:a030320.
- (English Language Editor: J. Gray)

Cite this article as: Li L, Ruan J, Zhang N, Dai J, Xu X, Tian X, Hu J. Identification of prognostic and driver gene mutations in acute myeloid leukemia by a bioinformatics analysis. *Transl Cancer Res* 2023;12(6):1552-1564. doi: 10.21037/tcr-23-587

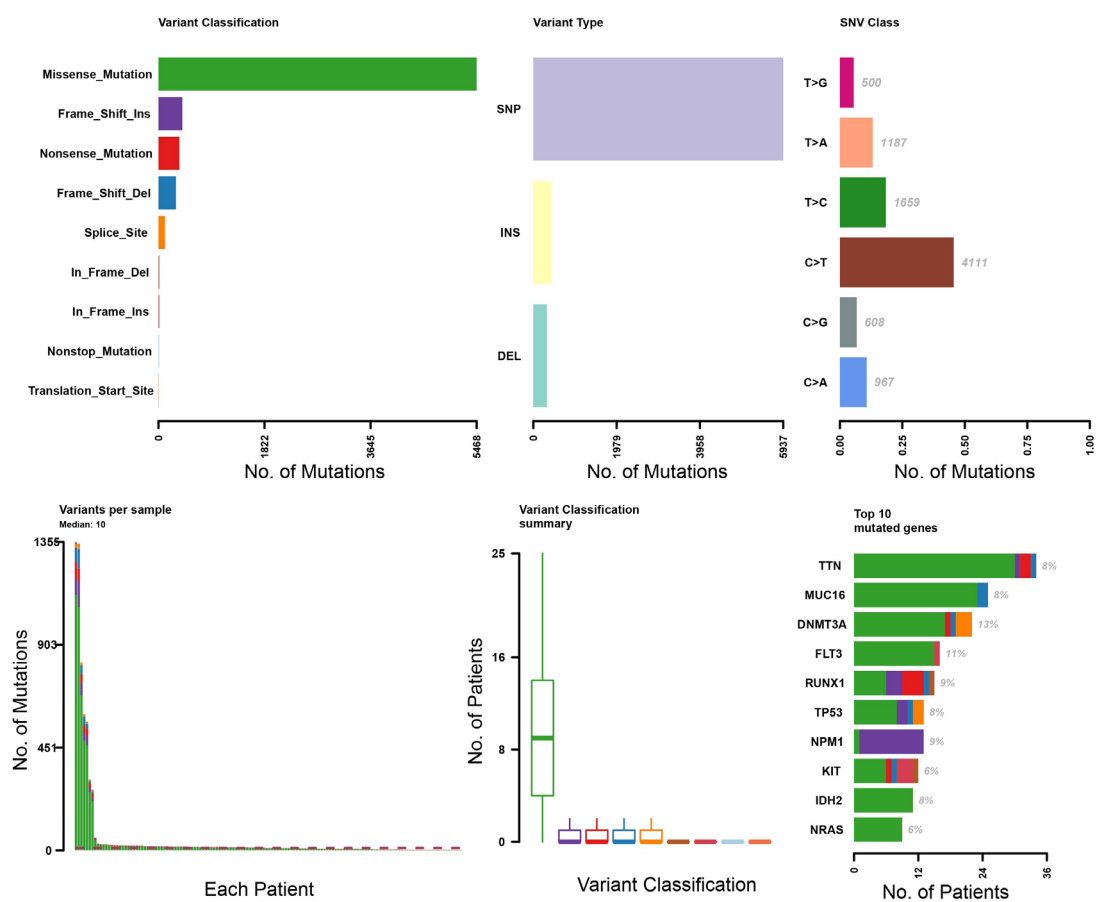
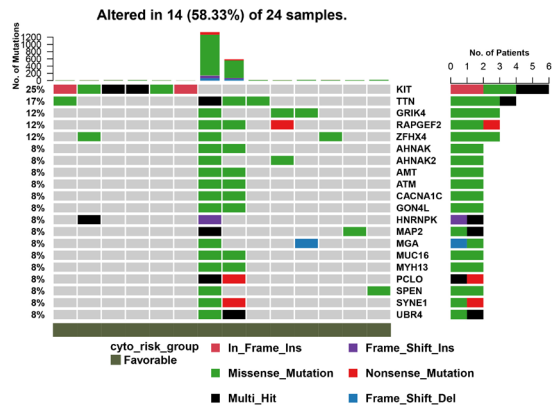
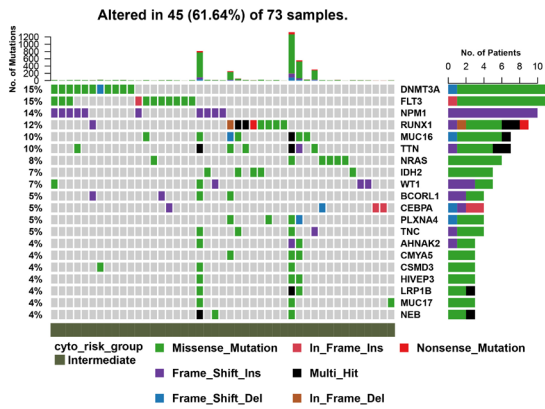


Figure S1 The profile of somatic mutations of patients with AML. Each row from top to bottom shows the statistics of mutation types, the type and number of mutation bases, and the box plot of mutation number and mutation species in each sample. AML, acute myeloid leukemia.

Favorable



Intermediate



Poor

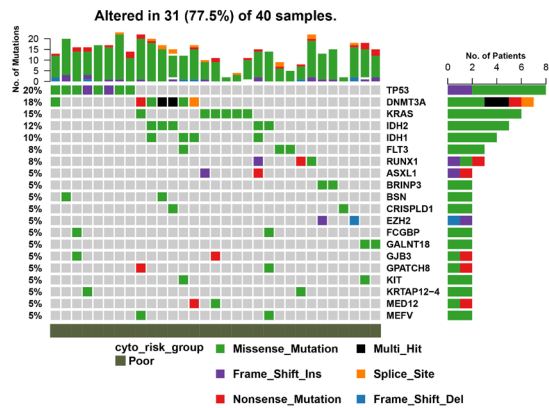


Figure S2 Heat maps showing the genes and mutation types of somatic mutations that occurred in the favorable (A), intermediate (B), and poor (C) prognostic groups.

Table S1 Significantly different gene mutations between the 3 prognostic groups

Hugo symbol	Group 1	Group 2	n_mutated_group1	n_mutated_group2	P value	OR low	OR high	FDR
KIT	Favorable	Rest	6 of 25	3 of 118	0.000951641	0	0.351055818	0.029712055
TP53	Poor	Rest	8 of 40	3 of 103	0.001751797	0	0.449673544	0.029712055
KRAS	Poor	Rest	6 of 40	1 of 103	0.001980804	0	0.38770562	0.029712055
NPM1	Intermediate	Rest	10 of 75	3 of 68	0.057048427	0	1.038299839	0.6417948
IDH1	Poor	Rest	4 of 40	3 of 103	0.095849297	0	1.310572294	0.761718878
TTN	Favorable	Rest	4 of 25	7 of 118	0.101562517	0	1.324849622	0.761718878
FLT3	Intermediate	Rest	11 of 75	5 of 68	0.131026151	0	1.315537773	0.819553313
IDH2	Poor	Rest	5 of 40	6 of 103	0.159169344	0	1.544620924	0.819553313
RUNX1	Intermediate	Rest	9 of 75	4 of 68	0.163910663	0	1.464221526	0.819553313
DNMT3A	Poor	Rest	7 of 40	12 of 103	0.252433328	0	1.703181146	1
PCLO	Favorable	Rest	2 of 25	4 of 118	0.281843854	0	3.19290701	1
NRAS	Intermediate	Rest	6 of 75	3 of 68	0.298063041	0	2.113526306	1
TTN	Intermediate	Rest	7 of 75	4 of 68	0.325340695	0	2.079845999	1
MUC16	Intermediate	Rest	7 of 75	4 of 68	0.325340695	0	2.079845999	1
DNMT3A	Intermediate	Rest	11 of 75	8 of 68	0.397359577	0	1.959977219	1
WT1	Intermediate	Rest	5 of 75	3 of 68	0.415276387	0	2.762824043	1
WT1	Favorable	Rest	2 of 25	6 of 118	0.423153167	0	4.479665159	1
PCLO	Poor	Rest	2 of 40	4 of 103	0.53820923	0	5.945930578	1
MUC16	Favorable	Rest	2 of 25	9 of 118	0.607037732	0	6.495630532	1
CMYA5	Favorable	Rest	1 of 25	5 of 118	0.691438614	0	25.74294736	1
CMYA5	Intermediate	Rest	3 of 75	3 of 68	0.704919526	0	6.323236075	1
RUNX1	Poor	Rest	3 of 40	10 of 103	0.762374247	0	5.906795442	1
KIT	Poor	Rest	2 of 40	7 of 103	0.774464068	0	9.610431269	1
NRAS	Poor	Rest	2 of 40	7 of 103	0.774464068	0	9.610431269	1
IDH2	Intermediate	Rest	5 of 75	6 of 68	0.787196266	0	4.733441448	1
IDH1	Intermediate	Rest	3 of 75	4 of 68	0.817681358	0	7.859942089	1
NRAS	Favorable	Rest	1 of 25	8 of 118	0.832242472	0	39.67256167	1
CMYA5	Poor	Rest	1 of 40	5 of 103	0.866108002	0	47.56655708	1
MUC16	Poor	Rest	2 of 40	9 of 103	0.86761872	0	12.18457926	1
FLT3	Poor	Rest	3 of 40	13 of 103	0.881844882	0	7.691615127	1
TP53	Favorable	Rest	1 of 25	10 of 118	0.889190796	0	49.39612747	1
PCLO	Intermediate	Rest	2 of 75	4 of 68	0.916541047	0	17.42343265	1
NPM1	Poor	Rest	2 of 40	11 of 103	0.924912763	0	14.85867244	1
RUNX1	Favorable	Rest	1 of 25	12 of 118	0.927319108	0	59.69609609	1
WT1	Poor	Rest	1 of 40	7 of 103	0.933082745	0	64.84321244	1
FLT3	Favorable	Rest	1 of 25	15 of 118	0.96191648	0	75.67068368	1
KRAS	Intermediate	Rest	1 of 75	6 of 68	0.99536161	0	164.9300569	1
KIT	Intermediate	Rest	1 of 75	8 of 68	0.999075354	0	219.5410318	1
TP53	Intermediate	Rest	1 of 75	10 of 68	0.999822459	0	278.0257502	1
DNMT3A	Favorable	Rest	0 of 25	19 of 118	1	0	Inf	1
NPM1	Favorable	Rest	0 of 25	13 of 118	1	0	Inf	1
TTN	Poor	Rest	0 of 40	11 of 103	1	0	Inf	1
IDH2	Favorable	Rest	0 of 25	11 of 118	1	0	Inf	1
IDH1	Favorable	Rest	0 of 25	7 of 118	1	0	Inf	1
KRAS	Favorable	Rest	0 of 25	7 of 118	1	0	Inf	1