

Histone Methylation Modification Patterns and Relevant M-RiskScore in Acute Myeloid Leukemia

Dade Rong

rongdd@mail2.sysu.edu.cn

Sun Yat-Sen University <https://orcid.org/0000-0001-9387-6789>

Xiaomin Chen

Sun Yat-Sen University

Daiyuan Liu

Sun Yat-sen university

Xiangna Ni

Sun Yat-sen University

Zhuomao Mo

Sun Yat-sen University

Xiuzhen Tong

Sun Yat-sen University

Haihe Wang

Sun Yat-sen University

Research

Keywords: acute myeloid leukemia, histone, methylation, classification, chemotherapy

Posted Date: June 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-629632/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Heliyon on September 1st, 2022. See the published version at <https://doi.org/10.1016/j.heliyon.2022.e10610>.

Abstract

Background: Acute myeloid leukemia (AML) is a type of heterogeneous disease with varied prognosis, but current classification methods for AML do not play an ideal role in guiding the therapy. Emerging studies shown alteration of histone methylation is closely related to leukemogenesis. This study aimed at identifying the molecular subtypes associated with histone methylation and establishing a relevant score to predict treatment respond and prognosis in AML.

Methods: Gene expression data and clinical characteristics of patients with AML were obtained from The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database. Molecular subtypes were identified by consensus clustering based on the expression of 24 histone methylation modification regulators (HMMRs). The clinical and biological features of each pattern were investigated by unsupervised clustering, survival, principle component (PCA), somatic mutations, gene set variation (GSVA), tumor mutation burden (TMB) and immune cell infiltration analyses. Different expression analysis and lasso regression analysis was conducted to establish the scoring system that was explored in the role of prognosis by using receivers operating curve (ROC) analysis and univariate/multivariate Cox regression analyses. Moreover, correlation analysis was performed to investigate the value of scoring system in chemotherapeutic prediction. Finally, an independent GSE dataset was used as a reference to validate the established clustering system.

Results: Two distinct histone methylation modification patterns had been identified that exhibit remarkable differences in several clinical and biological characteristics, including HMMRs' expression, AML-M0 distribution, mutations of *NPM1*, survival, tumor mutation burden, somatic mutations, pathways activation and immune cell infiltration. Besides, based on the clustering, we established the scoring system, M-RiskScore. Integrated analysis demonstrated that M-RiskScore-low patients displayed a prominent survival advantage and a better respond to decitabine treatment, while the opposite site was reported to M-RiskScore-high patients but they could benefit more from IA regimen therapy.

Conclusion: Our results demonstrate that detection of HMMRs' expression is potentially useful to AML therapy decisions, and targeting histone methylation would be a more promising strategy for either AML-M0 or *NPM1* mutant patients. M-RiskScore is a hopeful independent poor prognostic biomarker and be able to benefits the treatment decisions in AML.

Background

AML is an age-distributed hematopoietic stem cell differentiation disorder, which is mainly manifested in the inhibition of hematopoietic stem cell differentiation and the accumulation of immature cells at various stages, as well as the reduced production of active hematopoietic elements and cytokines [1, 2]. Meanwhile, AML is also a highly heterogeneous disease because its pathogenetic factors include various chromosomal and molecular abnormalities [3, 4]. Thanks to the progress in basic research, including the development of new chemotherapeutic drugs, the application of allogeneic stem cell transplantation and

the proposal and implementation of the supportive treatment, the outlook of patients with AML has significantly improved [1]. However, more than half of young adult patients and approximately 90% of elderly patients still die due to the disease itself or the toxicity of the therapeutic drugs, and thereby the heterogeneity of this disease remains the main obstacle [1, 2].

Two of the main systems that have been used to classify AML into subtypes are the French-American-British (FAB) classification and the newer World Health Organization (WHO) classification [5, 6]. The FAB classification comprehensively divides AML into subtypes, M0-M7, based on the types and maturity or differentiation of leukemia cells in blood or bone marrow, which is based largely on the morphological features of leukemia cells looked under the microscope after routine staining [5, 6]. However, this comprehensive classification does not play an ideal role in guiding the therapy for the patients with AML because there are almost no differences in the treatment regimens for other subtypes, except AML-M3 subtype [5–7]. The World Health Organization (WHO) classification of AML was established in 2001 with the advances of sequencing technology [8]. After two updated in 2008 and 2016, respectively, the WHO system divides AML into 11 subtypes based on chromosomal and genetic mutations [9, 10]. Compared with the FAB classification, the WHO system displays more advantages in guiding individual treatment because it takes account many of the factors that affect diagnosis and prognosis of AML [5, 6]. Nevertheless, the WHO classification does not cluster AML comprehensively.

Epigenetics refers to the covalent modification of DNA, RNA and histones without changing the DNA sequence to affect the expression of genes [11]. Epigenetic therapy targeting DNA methylation (decitabine and azacytidine) and histone acetylation (Chidamide) benefit the patients with AML, which not only proves the value of targeting epigenetic regulators in the treatment of AML but also urges the need to better understand epigenetic regulation and explore novel critical targets for effective AML treatment [12–16]. Histone methylation modification is one kind of epigenetic modification that modifies chromatin structure by regulating the methylation of lysine or arginine residues in histone tails to achieve gene expression or silencing [17–19]. Similar to other epigenetic modifications, there are three types of regulators involved in histone methylation modification, namely histone methyltransferases (“Writers”), histone demethylases (“Erasers”) and histone methyl modification recognition factors (“Readers”) [17]. In recent years, emerging studies have shown that aberrant histone methylation is closely related to leukemogenesis [20, 21]. Meanwhile, a series of lead compounds targeting histone methylation regulators have completed preclinical and even entered clinical studies [16, 22–25]. However, the current studies of histone methylation modification in AML are usually limited to the changes in one or a few single genes, while the mutation patterns of genes related to leukemia often involve multiple genes that interact in a highly coordinated manner in clinical practice [1, 2]. Therefore, it is necessary to systematically study the expression characteristics of histone methylation regulators in AML, which likely favors the diagnosis and treatment of AML.

In this study, we established a comprehensive classification approach for AML based on the expression level of 24 HMMRs in AML patients. We identified two distinct histone methylation modification patterns that exhibit remarkable differences in several clinical and biological characteristics, including AML-M0

distribution, mutations of *NPM1*, survival, TMB, somatic mutations, pathways activation and immune cell infiltration. Besides, based on the clustering, we established the scoring system, M-RiskScore, which not only acts as an independent poor prognostic predictor but also benefits the treatment decisions.

Results

Landscape of genetic variation of histone methylation modification regulators in AML

The overview of this work was shown in the form of a flowchart (Fig. 1). The clinical profiling of TCGA-LAML cohorts was summarized in Table 1.

Table 1
Clinical characteristics of AML cohort for classification and validation

Clinical Characteristics		Number	Percent (%)
TCGA-LAML (n = 124)			
Survival status	Survival	47	37.9
	Death	77	62.1
Age	≥ 60 years	51	41.1
	< 60 years	73	58.9
Gender	Female	57	46.0
	Male	67	54.0
IDH2-R132 mutation	Negative	112	90.3
	Positive	12	9.7
IDH2-R140 mutation	Negative	114	91.9
	Positive	10	8.1
IDH2-R172 mutation	Negative	122	98.4
	Positive	2	1.6
NPM1 mutation	Negative	93	75.0
	Positive	31	25.0
FLT3 mutation	Negative	88	71.0
	Positive	36	29.0
Activating RAS mutation	Negative	116	93.5
	Positive	8	6.5
FAB subtype	M0	12	9.7
	M1	30	24.2
	M2	27	21.8
	M3	13	10.4
	M4	27	21.8
	M5	12	9.7
	M6	2	1.6

Clinical Characteristics		Number	Percent (%)
	M7	1	0.8
Risk status	Favorable	28	22.6
	Intermediate	71	57.3
	Poor	25	20.1
GEO-GSE71014 (n = 104)		Number	Percent (%)
Survival status	Survival	68	65.4
	Death	36	34.6

A total of 24 HMMRs were finally identified in this study, including 13 writers, 7 erasers and 4 readers. Figure 2A summarized the dynamic reversible process of histone methylation mediated by HMMRs. The differences in gene expression levels between normal and cancer tissues are the critical messages for researchers to assess its attribute, oncogene or tumor suppressor [26, 27]. However, due to the special distribution of hematological malignancy, it is almost impossible for us to obtain the expression messages of HMMRs of the patients with AML before illness. Thence, we chosen to explore the gene expression of 24 regulators in AML patients with different RiskStatus. The result showed that there were 6 regulators with remarkable gene expression differences in different RiskStatus, including 1 writer (CARM1), 2 readers (ATRX and PC) and 3 erasers (KDM2A, KDM4A and KDM5B) (Fig. 2B). Among them, the up-regulation of ATRX, PC, KDM2A and CARM1 implied poor RiskStatus while the up-regulation of EZH2 and KDM5B suggested favorable RiskStatus (Fig. 2B).

Considering the interaction of the HMMRs, we preformed the co-expression analysis. As showed in Fig. 2C, there were two positive co-expression zones (as shown in red boxes) and the bigger one displayed that SETD2, a histone methyltransferase, expressed a positive correlation with ATRX, KDM5A, NSD1 and KMT2A, respectively in gene expression, while the smaller one exhibited 3 writers, including SUV39H1, DOT1L and EHMT2, taking together to form a co-expression loop. Meanwhile, a significant negative correlation in gene expression was discovered between KDM5B and SETD7, KDM5A and SUV39H1, along with ATRX and DOT1L (Fig. 2C). Subsequently, we summarized the incidence of the somatic mutation and copy number variation (CNV) of 24 regulators in the patients with AML. The result showed that there was no obvious genetic alteration in other regulators except for KMT2A that exhibited a significant CNV alteration, including fusion and amplification in copy number (Fig. 2D).

All in all, the above analyses displayed the basic landscape of 24 HMMRs in AML and indicated that cross-talking among the regulators of writers, readers, and erasers may play an important role in the formation of different histone methylation modification patterns in AML.

Histone methylation modification patterns mediated by 24 regulators in AML

In order to verify whether Histone methylation regulators affect the formation of histone methylation modification patterns in AML, the R package of ConsensusClusterPlus was used to classify patients with qualitatively different m6A modification patterns based on the expression of 24 HMMRs. It is obvious that the optimal consensus clustering can be obtained when K value of the consensus matrix was equal to 2 (Fig. 3A, S1), which was confirmed by the results of Consensus Cumulative Distribution Function (CDF) Plot (Fig. 3B) and Delta Area Plot (Fig. 3C). Therefore, two distinct histone methylation modification patterns were eventually identified, including 95 cases patterns A and 56 cases in patterns B. We termed these two patterns as Cluster A and Cluster B, respectively. Principal component analysis (PCA) for the transcriptome profiles of these two modification patterns also showed that there was significant distinction existed on them (Fig. 3D). To validate the established histone methylation modification patterns, we repeated the correlated analyses using another independent AML-cohort, GSE71014, of which the clinical profiling was summarized in Table 1. As shown in Figure S2A-C, the best consensus clustering also obtained when K value of the consensus matrix was equal to 2, which was consistent with the previous clustering analysis (Fig. 3A-D, S1). What's more, the survival analysis displayed that the patients of Cluster A exhibited a more prominent survival advantage than Cluster B in both TCGA-LAML cohort and GSE71014 cohort (Fig. 3E-F).

Subsequently, unsupervised clustering of 24 HMMRs was conducted to explore the clinical features of these two histone methylation modification patterns in TCGA-LAML cohort while the two modification patterns, FAB subtypes, mutations of several genes, gender, age, fustat and futime were used as patient annotations. As showed in Fig. 3G, the expression of 24 histone methylation regulators were remarkably upregulated in Cluster A, comparing with Cluster B. It is worth noting that there was a significant difference between Cluster A and Cluster B in the distribution of FAB category, that was, AML-M0 subtype was only distributed in Cluster A but not in Cluster B. Besides, more patients with *NPM1* mutation distributed in Cluster B than Cluster A. However, there were no obvious differences between the patients of Cluster A and Cluster B in *FLT3* mutation, *IDH1* mutation, activating *Ras*, gender, age, fustat and futime.

The above results showed that the patients with AML could be clustered effectively based on the expression of histone methylation modification regulators. What's more, the significant clinical characteristics' differences of the two histone methylation modification patterns also indicated histone methylation modification plays a critical role in AML progression.

Biological characteristics of distinct histone methylation modification patterns in AML

To determine the biological behavior differences between Cluster A and Cluster B, we firstly analyzed their distribution differences of somatic mutation in TCGA-AML cohort using maftools package. It is believed that higher TMB results in more neo-antigens, increasing chances for T cell recognition and clinically correlates with better immune checkpoint inhibitors (ICI) outcomes [28, 29]. The result of somatic mutation analysis obviously shown that Cluster A presented more extensive tumor mutation burden than Cluster B, with the altering frequency 62.5% against 48.15% (Figure S3A-B). Furthermore, the tumor

mutation burden (TMB) quantification analysis also confirmed that Cluster A was markedly correlated with a higher TMB value (Fig. 4A) although there was no significant difference in microsatellite instability between Cluster A and Cluster B (Fig. 4B).

To further investigate the characteristics of Cluster A and Cluster B, we performed GSVA enrichment analysis. As showed in Fig. 4C, Cluster A was remarkably enriched in protein and RNA metabolism pathways, such as RNA degradation, spliceosome pathway, valine leucine and isoleucine biosynthesis pathway and so on while Cluster B presented an enrichment pathway associated with energy metabolism, including oxidative phosphorylation pathway, pantothenate and CoA biosynthesis pathway and lysosome pathway.

It has been well known that tumor microenvironment (TME), which is also composed of tumor cells, stromal cells, immune cells and multiple secreted factors, plays a crucial role in tumor progression [30, 31]. Therefore, we firstly performed the MCP-counter method to investigate the immune cell infiltration of the two histone modification patterns. The result showed that Cluster A presented a higher immune cell infiltration in T cells, B cells, NK cells and endothelial cells while a lower infiltration in monocytes and neutrophils that were the noumenon of tumor cells (Fig. 4D). Then, we conducted the CIBERSORT method to verify the result of MCP-counter. Consistently, the result also showed that the infiltration of CD4+ memory T cells, naive B cells, resting NK cells and plasma cells in Cluster A were all higher than in Cluster B but the infiltration of monocytes was lower (Figure S3C).

Enlighten by the results of immune cell infiltration analyses, we compared the expression of some immune cell markers, chemokines and cytokines between Cluster A and Cluster B so that we could learn more about the immune characteristics of histone modification patterns in AML. As showed in Fig. 4E, a series of immune-activating factors were up-regulated in Cluster A in comparison with Cluster B, including CD244 that is a marker of NK cells, inducible T cell costimulatory ligand (ISOSLG), CD96 that plays a role in the adhesive interaction of activated T and NK cells, TNFRSF18 and CD47. Meanwhile, some immune inhibiting factors were down-regulated in Cluster A, such as HAVCR2, CD86 and LGALS9. Combining the previous results that the expression of 24 histone methylation modification regulators was generally up-regulated in Cluster A (Fig. 3G), we speculated that HMMRs could promote the expression of a series of immune-activating factors in AML.

Overall, we had made a rather detailed description of the biological characteristics of the two histone methylation modification patterns in AML and found that Cluster A displayed more remarkable immune cell infiltration that should be a key factor for its prominent survival advantage.

Generation of histone methylation modification-related score in AML

The expression differences of histone methylation modification regulators between Cluster A and Cluster B would directly lead to the differences in the transcriptome of Cluster A and Cluster B, which resulted in

the clinical and biological characteristics differences in these two histone methylation modification patterns of AML.

To further investigate the changes of the transcriptome of Cluster A and Cluster B, we firstly conducted the empirical Bayesian approach of limma package to determine the differentially expressed genes (DEGs) between them. The volcano plots exhibited that there were lots of DEGs between Cluster A and Cluster B, remaining 61 DEGs even with the \log_2 (fold change) value was equal to ± 3 (Fig. 5A). Then, the clusterProfiler package was used to perform GO enrichment analysis for the DEGs. As shown in Fig. 5B, lots of pathways directly related to immune system were enriched in Cluster A compared with Cluster B, which indicated again that histone methylation modification played a non-negligible role in regulating TME landscapes in AML.

Although we had proved that histone methylation modification patterns were strongly associated with the prognosis of AML, it would be difficult to directly predict individual therapeutic effect and prognosis according to the histone methylation modification patterns due to its complexity. Therefore, we established a scoring system to quantify the histone methylation modification patterns of the patients with AML based on the DEGs discovered in Fig. 5A. We termed as M-RiskScore. Univariate Cox analysis was performed to identify the genes that were eligible for lasso regression analysis and the results showed that a total of 8 genes were selected with their p-value was less than 0.01 (Fig. 5C). The subsequent lasso regression analysis displayed the formula for calculating the M-RishScore was composed of 6 genes finally and they were ADAMTS15, CADM1, CDO1, SYT17, FAM163A and POF1B (Fig. 5D-E).

Taken together, we identified the DEGs between Cluster A and Cluster B, by which we constructed a scoring system, M-RiskScore successfully. We would try to explore the value of M-RiskScore for predicting the therapeutic effect and prognosis of the patients with AML in the subsequent analysis.

The value of M-RiskScore in prognosis prediction for the patients with AML

After we have established the scoring system, we started to explore its potential value in predicting the outcome of the patients with AML.

With the medium value was set as the cutoff value, we divided the patients with AML into high or low M-RiskScore group. The prognosis analysis showed that the low M-RiskScore group exhibited a better survival advantage than the high M-RiskScore group (Fig. 6A). Importantly, the results of correlation analysis between histone methylation modification patterns and M-RiskScore displayed that Cluster A, which had been proved to possess a more prominent survival benefit than Cluster B (Fig. 3E), was with a lower M-RiskScore (Figure S4A), indicating the reliable and trustful of M-RiskScore due to these consistent results. Therefore, the above analyses implied that M-RiskScore was likely to be an independent poor prognostic marker for the patients with AML.

To verify and compare the predictive accuracy of M-RiskScore with those of other well-known prognostic markers in AML, two receiver-operating characteristics (ROC) curves were created. As showed in Fig. 6B-C, M-RiskScore had the highest AUC value both in the cohort of 3 years and 5 years, indicating its remarkable predictive accuracy compared with other prognostic markers of AML. What's more, both the results of univariate and multivariate Cox regression analysis showed that the Hazard ratio of M-RiskScore was significantly higher than those of other prognostic markers of AML (Fig. 6D-E). Logically, it was no doubt that M-RiskScore could serve as an independent poor prognostic marker for the patients with AML.

Finally, to further improve the prognostic value of M-RiskScore in AML, we performed the nomogram associating M-RiskScore with other prognostic markers of AML, including FAB subtype, gender, RiskStatus and age (Fig. 6F). The result of decision curve analysis (DCA) for the nomogram showed that M-RiskScore had a common higher Net Benefit value than other independent prognostic markers in the range of Risk Threshold (Fig. 6G). Moreover, all the calibration plots for the probability of survival of 1 year, 3 years and 5 years ran very close to the diagonal, showing outstanding calibration (Fig. 6H). However, the result of ROC analysis for nomogram exhibited that there was no significant increase in the predictive accuracy compared with independent M-RiskScore in the cohorts of both 3 years and 5 years (Fig. 6B-C, Figure S4B).

In a word, these results proved that M-RiskScore played an important role in predicting the outcome of the patients with AML and could be an independent poor prognostic marker of AML.

M-RiskScore in the role of chemotherapy of AML

Chemotherapeutic drugs will continue playing an important role in AML treatment by the virtue of their universal and highly effective cytotoxicity [2]. Therefore, we specifically tested the ability of M-RiskScore to predict the efficacy of induction chemotherapy in patients with AML. Three GEO-AML cohorts with the chemotherapeutic information were download and processed. They were GSE84344, having the respond information for decitabine treatment of the patients with AML, GSE103424 that contained the clinical characteristics of the patients with AML following the treatment of IA regimen (idarubicin + cytarabine) and GSE10087, which processed the clinical information of the patients with AML treated with different standard induction chemotherapy, including IA regimen, CIA regimen (clofarabine + idarubicin + cytarabine), FAI regimen (fludarabine + cytarabine + idarubicin), decitabine, BID FA regimen (twice daily fludarabine + cytarabine) and CECA regimen (cyclophosphamide + etoposide + carboplatin + cytarabine). The basic clinical characteristics of these three GEO-AML cohorts were summarized in Table 2. We found that patients with low M-RiskScore showed more obvious therapeutic advantages both in the complex-chemotherapy-treatment cohort and the decitabine-treatment cohort compared to those with high M-RiskScore (Fig. 7A-B). In contrast, the analysis of the IA regimen cohort exhibited the AML-patients with high M-RiskScore benefited more from the related treatment relatively (Fig. 7C). The somatic mutation analysis for AML-TCGA cohort displayed that the top three most frequently mutated genes in AML-patients with low M-RiskScore were *DNMT3A*, *KIT* and *WT1*, while they were *NPM1*, *DNMT3A* and *RUNX1* in the high M-RiskScore group (Fig. 7D-E). Given DNMT3A, a DNA methyltransferase encoded by

DNMT3A, is one of the well-known target of decitabine [32], we speculated that a higher mutation frequency of *DNMT3A* could be the cause of its poorer chemotherapeutic response to the treatment with decitabine in the high M-RiskScore group (Fig. 7B, D). What's more, given these mutation characteristics, we proposed a hypothesis that the mutated *KIT* and/or mutated *WT1* led to the oncogenesis in AML-patients with low M-RiskScore. Meanwhile, the leukemogenesis in the patients with high M-RiskScore should be attributed to the mutations of *NPM1* and/or *RUNX1*.

Table 2
Clinical characteristics of AML cohort for correlation analyses.

Patient series	GSE103424	GSE103424	GSE103424
No. of patients	52	45	41
Age			
≥ 60 years	NA	45	25
< 60 years	NA	0	16
Gender			
Female	24	17	14
Male	28	28	27
Risk status			
Favorable	5	0	4
Intermediate	30	23	15
Poor	10	14	22
NA	7	8	0
Response to chemotherapy			
Yes	17	18	11
No	35	27	30

All in all, the above results indicated that M-RiskScore was associated with the chemotherapeutic response of the AML-patients with different chemotherapy treatments and could serve as an evaluated index for chemotherapeutic drugs' usage.

Discussion

Individualized treatment is the ultimate goal of cancer treatment, which can effectively alleviate the disease, minimize the risk of complications or side effects and is economical, basing on a comprehensive consideration of various physiological or clinical characteristics of the patients, including gender, age,

genetic characteristics, and treatment histories [33–35]. The achievement of individualized cancer therapy requires both extensive pathologic subtyping of the tumor for its heterogeneity and diagnosis of genome alterations of the patients with cancer [35]. Advances in sequencing techniques provide an unprecedented view of the complex genetic and nongenetic heterogeneity within individual tumors, which not only remove the restriction on understanding the alterations of gene expression in the patients but also provides an opportunity for tumor subtype classification [33–35].

Herein, we established a repeatable clustering and scoring system for AML basing on 24 HMMRs' expression of the patients with AML. The patients with a lower expression of 24 histone methylation regulators were mainly divided into Cluster A, exhibiting a better survival advantage, a higher TMB value and a more significant immune cell infiltration, compared with those who possessed a higher expression of related genes in Cluster B. M-RiskScore, a scoring system for quantifying the two histone methylation modification patterns, was constructed based on the expression of 6 DEGs with different weights between Cluster A and Cluster B. A close connection between the histone methylation modification patterns and M-RiskScore was reflected in the obvious overlap of the patients in Cluster A and the patients with low M-RiskScore and vice versa, which was confirmed by a remarkable survival advantage of the patients with lower M-RiskScore. It is of note that M-RiskScore was proved to be a dependent poor prognostic marker of AML for its excellent prognostic accuracy. Moreover, we also found that E-RiskScore could serve as a diagnostic index for the choice of chemotherapeutic drugs from therapy.

Although the FAB classification is the first and the most comprehensive classification method for AML, morphologically [5]. However, the AML-M3 also referred to acute promyelocytic leukemia (APL), which has been captured successfully with the combination treatment of two targeted therapies: retinoic acid (RA) and arsenic [7], the other seven subtypes still share similar treatment proposals to date. Interestingly, our study shown that almost all the patients who exhibited AML-M0 subtype were divided into Cluster A. AML-M0 means acute leukemias with minimal signs of myeloid differentiation [5]. Based on this, we deduced that activity of histone methylation regulators is closely related to AML-M0. Even if no study has reported on the relevance of the pathogenesis of AML-M0 and histone methylation modification, the recent advances in dissecting the molecular regulation and targeting histone methylome in AML together with the success in developing lead compounds specific to key histone methylation-modifying enzymes [16] indicated epigenetic therapies by targeting histone methylome will shed light on the treatment of the patients with AML-M0.

Mutations of *NPM1* are well known to be associated with leukemogenesis and have already represented a distinct entity in the WHO classification of AML. In our study, we found that most of the patients in Cluster B displayed mutations of *NPM1* while *NPM1* was one of the top three frequently mutated genes in the high M-RiskScore subtype, which indicated mutant *NPM1* should be associated with an inactive gene expression of HMMRs in AML. To date, although few studies have reported on the relationship between histone methylation and mutations of *NPM1*, the existing research results almost all can be served as the experimental evidence for our conjecture. For example, it has been reported that KDM1A antagonist SP2509 induced more apoptosis in AML cells expressing mutant *NPM1* than mixed-lineage leukemia

fusion oncoproteins [36]. Another study also pointed out an inhibitor for histone methyltransferase DOT1L exhibited a potent cytotoxicity to the *NPM1*-mutated AML cells [37]. More researches are needed to ascertain the causality between mutation of *NPM1* and the inactive expression of histone methylation regulators, but it is no doubt that histone methylation regulators are promising targets for *NPM1*-mutated AML therapy.

In our analyzed setting, we speculated that the mutations of *RUNX1* (*RUN family transcription factor 1*), which also is a distinct entity in the WHO classification of AML [38] and one of the top three genes with the most frequent mutation rate in the high M-RiskScore subtype, led to leukemogenesis in the patients with high M-RiskScore individually or combination with mutant *NPM1*, and meanwhile, mutated *KIT* (*KIT proto-oncogene, receptor tyrosine kinase*) and/or *WT1* (*WT1 transcription factor*) could be the cause of oncogenesis in the AML-patients with low M-RiskScore. The fact that both mutations of *NPM1* and *RUNX1* have already been developed as distinct entities in the WHO classification of AML is sufficient to prove the first half of our hypothesis. Regardless of neither mutated *KIT* nor mutated *WT1* has become the entity for classification of AML, multiple previous elegant studies have shown that they are critically associated with leukemogenesis [39–45]. The purpose of identifying the driver mutations of AML is to discover novel entities for classification and finally find out corresponding therapy. We have already proposed our insights on the treatment of *NPM1*-mutated AML in the above analysis. As for *RUNX1* and *WT1*, because both of them are transcription factors, attempts to directly disrupt their function are likely to meet limited success for their commonly “undruggable” protein structures and unbearable side effects. However, it is possible to conduct drug development according to the products they regulate, especially enzymes. In contrast to *RUNX1* and *WT1*, *KIT*, which is a protein tyrosine kinase receptor, is very likely to be a direct chemotherapeutic target for the AML-patients with mutant *KIT*. Imatinib, a famous inhibitor of BCR-ABL tyrosine kinase inhibitors (TKI) and the first-line chemotherapeutic drug for most patients with chronic myelogenous leukemia (CML) [46, 47], also exhibit an effective inhibition on *KIT* kinase activity [48, 49], which implies that imatinib could be applied for the chemotherapy in *KIT*-mutated AML.

It is of note that our results showed that the patients with low M-RiskScore possessed a better response to the chemotherapy in an AML-cohort that containing 41 samples from 38 patients who had received different chemotherapeutic regimens, which implied that a lower M-RiskScore was likely to associate with a better chemotherapeutic respond. Furthermore, the continued analyses revealed that the patients with a lower M-RiskScore were likely to benefit from the treatment of decitabine, but not IA therapy regimen. Because the cohorts used for the above analyses do not contain abundant patient samples, there are no significant statistical differences in the analyses mentioned in this section. However, these results still indicated that M-RiskScore may be able to serve as a diagnostic marker for patients with AML.

Conclusions

In conclusion, our study demonstrated that histone methylation modification plays a curial role in the diagnosis and prognosis of AML from the perspective of bioinformatics. The defined clustering and scoring system may contribute to the individual treatment of the patients with AML.

Materials And Methods

1. Data collection

Public gene expression, mutation and clinical annotation data were obtained in The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) and Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). LAML cohort from TCGA database was used as the training dataset. GSE110087, GSE84334, GSE103424 and GSE71014 cohorts were acquired from GEO database. Among them, the first three cohorts were used for analyses of chemotherapy in AML while the last one was for validation.

24 HMMRs, including 13 writers (KMT2A, KMT2D, KMT5A, SETD2, NSD1, SMYD3, NSD2, DOT1L, EZH2, SETD7, CARM1, SUV39H1, EHMT2), 7 erasers (KDM1A, KDM2A, KDM4A, KDM5A, KDM5B, KDM6A, KDM6B) and 4 readers (ATRX, EED, PC, RAG2), were identified from the related studies for further analyses [17].

2. Landscape for histone methylation modification regulators in AML

Correlation analysis of the gene expression among the 24 histone methylation modification regulators in AML was explored by the “corrplot” package. The Wilcoxon rank-sum test was used to investigate the expression difference of histone methylome regulators in AML patients with different RiskStatus. Meanwhile, the information of somatic mutation and copy number variations (CNV) of included genes were generated from cBioPortal website (<https://www.cbioportal.org/>).

3. Consensus clustering for 24 histone methylation modification regulators

We conducted the Consensus clustering analysis to identify distinct histone methylation modification patterns based on the expression of 24 histone methylome regulators so that we could classify the patients with AML for further analysis. The number of clusters and their stability was determined by the consensus clustering algorithm and 1000 times repetitions were conducted for guaranteeing the stability of classification [50].

4. Features of distinct histone methylation modification patterns

A series of analyses were performed to validate the histone methylation modification patterns after finishing the consensus clustering. Principal component analysis (PCA), a technique for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss [51], was conducted to verify the quality of consensus clustering. “Survival” package was applied to explore the time-dependent prognostic value of the clusters.

Tumor mutational burden (TMB), a new promising biomarker emerged recently, is classically defined as the number of non-synonymous exonic mutations per megabase (Mb) [28]. The total number of mutations counted was divided by the exome size (38 Mb was utilized as the exome size), by which we calculated the TMB of each case. TMB correlation analysis was executed to explore the association between TMB and the distinct clusters.

To investigate the difference in biological process activity between the clusters, we performed gene set variation analysis (GSVA) enrichment analysis using “GSVA” R packages, which is an unsupervised and non-parametric method for estimating the variation in biological process and pathway activity in the samples of an expression dataset [52]. The gene sets of “h.all.v7.1.symbols” were downloaded from the MSigDB database (<https://www.gsea-msigdb.org/gsea/index.jsp>) for GSVA analysis. The result was considered to be statistically significant while its p-value was less than 0.05. Moreover, we explored the somatic gene mutations in the different clusters by the “maftools” package.

5. Estimation of immune cell infiltration

We explored the different immune cell infiltration between the distinct clusters using MCP-counter and CIBERSORT methods. MCP-counter is a methodology based on transcriptomic markers that assess the proportion of immune and stromal cell populations in the tumor microenvironment (TME) from transcriptomic data. There are 10 cell populations estimated by MCP-counter, including T cells, CD8 + T cells, cytotoxic lymphocytes, B lineage, NK cells, Myeloid dendritic cells, Neutrophils, Endothelial cells and Fibroblasts [53]. CIBERSORT, a deconvolution algorithm reported, is used to quantify the 22 infiltrated immune cells according to normalized gene expression profiles. The 22 immune cells included memory B cells, naive B cells, plasma cells, resting/activated DCs, resting/activated NK cells, resting/activated mast cells, eosinophils, neutrophils, monocytes, M0–M2 macrophages, and 7 T-cell types (CD8 + T cells, regulatory T cells (Tregs), resting/activated memory CD4 + T cells, follicular helper T cells, naive CD4 + T cells and $\gamma\delta$ T cells) [54].

What’s more, 20 immune checkpoint genes (*CD244*, *PDCD1*, *PD-L 1*, *CTLA4*, *CD80*, *CD86*, *CD28*, *TIGIT*, *PVR*, *CD96*, *SIRPA*, *CD47*, *LGALS9*, *HAVCR2*, *ICOS*, *ICOSLG*, *TNFSF18*, *TNFRSF18*, *KLRD1*, *KLRC1*) were retrieved from previous study [55], and correlation analysis was conducted to determine the difference expression of immune checkpoint genes between the two clusters.

6. Construction of histone methylation modification-related score (M-RiskScore) in AML

We firstly conducted the empirical Bayesian approach of limma package to determine the differentially expressed genes (DEGs) between the histone methylation modification patterns. Then, we performed univariate Cox regression analysis to identify the prognostic genes with a p-value less than 0.01 for further analysis.

The lasso regression analysis was subsequently applied to construct the histone methylation modification-related score (RiskScore) by “glmnet” and “survival” package. In this analysis, a lasso

penalty was used to account for shrinkage and variable selection. The optimal value of the lambda penalty parameter was defined by performing 10 cross-validations. The calculation formula for the acetylation-related score was as follows:

$$\begin{aligned} \text{RiskScore} = & (\text{coefficient mRNA1} \times \text{expression of mRNA1}) \\ & + (\text{coefficient mRNA2} \times \text{expression of mRNA2}) + \dots \\ & + (\text{coefficient mRNA}_n \times \text{expression of mRNA}_n) \end{aligned}$$

According to the median of the RiskScore, the patients with AML were divided into two groups (high-risk or low-risk group). We then performed the survival analysis based on this grouping strategy. A receiver operating characteristic (ROC) curve, which is a plot of the sensitivity versus 1 – specificity of a diagnostic test [56], was constructed to examine the prognostic accuracy for purpose of verifying the RiskScore. Finally, to validate whether the RiskScore could be an independent prognostic marker in AML, we carried out the univariate and multivariate Cox regression analyses.

7. Nomogram construction and validation

In order to improve the clinical application of RiskScore, we constructed the nomogram based on Cox regression model. There were 5 components of the nomogram, including gender, FAB subtype, RiskStatus, age and RiskScore. Decision curve analysis was performed to compare the net benefits of different prognostic models (FAB subtype, RiskStatus, RiskScore and nomogram). The concordance index, calibration plot, and ROC curve were used to verify the nomogram.

Nomograms are a pictorial representation of a complex mathematical formula [57]. Medical nomograms use biological and clinical variables, such as tumor grade and patient age, to determine a statistical prognostic model that generates a probability of a clinical event, such as cancer recurrence or death, for a particular individual [58].

Model performance was evaluated through calibration and discrimination [59]. Bias corrected calibration for 3 years and 5 years overall survival rate was performed by 1000 bootstrap resamples to evaluate the consistency between the observed and estimated survival probability by “rms” package. Discrimination was evaluated by Harrell’s concordance index (C-index) and ROC curve. A higher C-index value demonstrated better model-fitting performance. Area under the ROC curve (AUC) value is an effective way to summarize the overall diagnostic accuracy of the test, taking values from 0 to 1, where a value of 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test, indicating a higher AUC value revealed superior model discriminative ability [56].

Decision curve analysis (DCA) was further performed to measure and compare the clinical utilities of the different prognostic models. DCA is a method for evaluating the benefit of a diagnosis test across a range of patient preferences for accepting the risk of undertreatment and overtreatment to facilitate decisions about test selection and use [60].

8. Biological characteristics of RiskScore

GSVA analysis, TMB analysis, Somatic gene mutations analysis and immune cell infiltration analysis were performed to exhibit the biological characteristics of the high-RiskScore group and low-RiskScore group.

9. Validation of Molecular Classification

Here, we employed the RECA-EU cohort from the ICGC database for score validation. Survival analysis, ROC curve, univariate and multivariate Cox regression analyses were also performed to estimate the clinical value of the acetylation-related score

Abbreviations

AML	Acute myeloid leukemia
HMMRs	Histone methylation modification regulators
ROC	receiver operating characteristic
AUC	Area under the ROC curve
TMB	Tumor mutational burden
PCA	Principal component analysis
ICI	immune checkpoint inhibitors
CNV	copy number variations
GSVA	gene set variation analysis
TME	tumor microenvironment
DEGs	differentially expressed genes
WHO	World Health Organization
TKI	tyrosine-kinase inhibitors
CML	chronic myelogenous leukemia

Declarations

Acknowledgements

We are grateful to Wen-Chien Chou (Department of Laboratory Medicine, National Taiwan University Hospital, Taiwan), Luciano Castiello (Pasteur Italy Institute, Italy), Chieh-Lin Jerry Teng (Division of Hematology/Medical Oncology, Taichung Veterans General Hospital, Taiwan) and Stephan Rainer Bohl (Internal Medicine III, University Hospital of Ulm, Germany) for shared the genetic and clinical information of the patients with AML.

Authors' contributions

Conception and design of studies: DR, ZM, XC, XN. Drafting article: DR, XC, HW, XT. Review and discussion: ZM, HW, XT.

Funding

This work was supported by Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515010294) to TX, and the National Natural Science Foundation of China (No. 81472730) to WH.

Availability of data and materials

All data and materials are available via DR.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no financial conflicts of interest.

References

1. Ferrara F, Schiffer CA. Acute myeloid leukaemia in adults. *The Lancet*. 2013;381:484–95.
2. Short NJ, Rytting ME, Cortes JE. Acute myeloid leukaemia. *Lancet*. 2018;392:593–606.
3. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med*. 2016;374:2209–21.
4. Prada-Arismendy J, Arroyave JC, Rothlisberger S. Molecular biomarkers in acute myeloid leukemia. *Blood Rev*. 2017;31:63–76.
5. Vardiman J. The classification of MDS: from FAB to WHO and beyond. *Leuk Res*. 2012;36:1453–8.
6. Hwang SM. Classification of acute myeloid leukemia. *Blood Res*. 2020;55:1-s4.
7. de Thé H, Pandolfi PP, Chen Z. Acute Promyelocytic Leukemia: A Paradigm for Oncoprotein-Targeted Cure. *Cancer Cell*. 2017;32:552–60.

8. Jaffe ES, Organization WH, Cancer IAfRo. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. IARC Press; 2001.
9. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellstrom-Lindberg E, Tefferi A, Bloomfield CD. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 2009;114:937–51.
10. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127:2391–405.
11. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004;429:457–63.
12. Cashen AF, Schiller GJ, O'Donnell MR, DiPersio JF. Multicenter, phase II study of decitabine for the first-line treatment of older patients with acute myeloid leukemia. *J Clin Oncol*. 2010;28:556–61.
13. Ball B, Zeidan A, Gore SD, Prebet T. Hypomethylating agent combination strategies in myelodysplastic syndromes: hopes and shortcomings. *Leuk Lymphoma*. 2017;58:1022–36.
14. Stahl M, DeVeaux M, Montesinos P, Itzykson R, Ritchie EK, Sekeres MA, Barnard JD, Podoltsev NA, Brunner AM, Komrokji RS, et al. Hypomethylating agents in relapsed and refractory AML: outcomes and their predictors in a large international patient cohort. *Blood Adv*. 2018;2:923–32.
15. Wang H, Liu YC, Zhu CY, Yan F, Wang MZ, Chen XS, Wang XK, Pang BX, Li YH, Liu DH, et al. Chidamide increases the sensitivity of refractory or relapsed acute myeloid leukemia cells to anthracyclines via regulation of the HDAC3 -AKT-P21-CDK2 signaling pathway. *J Exp Clin Cancer Res*. 2020;39:278.
16. Tsai CT, So CW. Epigenetic therapies by targeting aberrant histone methylome in AML: molecular mechanisms, current preclinical and clinical development. *Oncogene*. 2017;36:1753–9.
17. Audia JE, Campbell RM. Histone Modifications and Cancer. *Cold Spring Harb Perspect Biol*. 2016;8:a019521.
18. Cheung N, So CW. Transcriptional and epigenetic networks in haematological malignancy. *FEBS Lett*. 2011;585:2100–11.
19. Hammond CM, Stromme CB, Huang H, Patel DJ, Groth A. Histone chaperone networks shaping chromatin function. *Nat Rev Mol Cell Biol*. 2017;18:141–58.
20. Schenk T, Chen WC, Gollner S, Howell L, Jin L, Hebestreit K, Klein HU, Popescu AC, Burnett A, Mills K, et al. Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. *Nat Med*. 2012;18:605–11.
21. Salvatori B, Iosue I, Djodji Damas N, Mangiavacchi A, Chiaretti S, Messina M, Padula F, Guarini A, Bozzoni I, Fazi F, Fatica A. Critical Role of c-Myc in Acute Myeloid Leukemia Involving Direct Regulation of miR-26a and Histone Methyltransferase EZH2. *Genes Cancer*. 2011;2:585–92.
22. Maes T, Mascaro C, Tirapu I, Estiarte A, Ciceri F, Lunardi S, Guibourt N, Perdones A, Lufino MMP, Somerville TCP, et al. ORY-1001, a Potent and Selective Covalent KDM1A Inhibitor, for the Treatment

- of Acute Leukemia. *Cancer Cell*. 2018;33:495–511 e412.
23. Kruger RG, Mohammad H, Smitheman K, Cusan M, Liu Y, Pappalardi M, Federowicz K, Van Aller G, Kaspavec J, Tian X, et al. Inhibition Of LSD1 As a Therapeutic Strategy For The Treatment Of Acute Myeloid Leukemia. *Blood*. 2013;122:3964–4.
24. Swords RT, Perez A, Rodriguez A, Watts JM, Vargas F, Elias R, Rienhoff HY Jr, Zelent A. The Small Molecule Img-98, a Potent and Selective Inhibitor of the Lysine Demethylase Lsd-1, Effectively Augments the Pro-Differentiation Effects of ATRA in a Pre-Clinical Model of AML. *Blood*. 2015;126:460–0.
25. Wouters BJ, Delwel R. Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. *Blood*. 2016;127:42–52.
26. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med*. 2004;10:789–99.
27. Kontomanolis EN, Koutras A, Syllaios A, Schizas D, Mastoraki A, Garmpis N, Diakosavvas M, Angelou K, Tsatsaris G, Pagkalos A, et al. Role of Oncogenes and Tumor-suppressor Genes in Carcinogenesis: A Review. *Anticancer Res*. 2020;40:6009–15.
28. Fumet JD, Truntzer C, Yarchoan M, Ghiringhelli F. Tumour mutational burden as a biomarker for immunotherapy: Current data and emerging concepts. *Eur J Cancer*. 2020;131:40–50.
29. Jardim DL, Goodman A, de Melo Gagliato D, Kurzrock R. The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. *Cancer Cell*. 2021;39:154–73.
30. Yuan Y. **Spatial Heterogeneity in the Tumor Microenvironment**. *Cold Spring Harb Perspect Med* 2016, 6.
31. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res*. 2019;79:4557–66.
32. Watts J, Nimer S: **Recent advances in the understanding and treatment of acute myeloid leukemia**. *F1000Res* 2018, 7.
33. Mendelsohn J. Personalizing oncology: perspectives and prospects. *J Clin Oncol*. 2013;31:1904–11.
34. Buettner R, Wolf J, Thomas RK. Lessons learned from lung cancer genomics: the emerging concept of individualized diagnostics and treatment. *J Clin Oncol*. 2013;31:1858–65.
35. Senft D, Leiserson MDM, Ruppin E, Ronai ZA. Precision Oncology: The Road Ahead. *Trends Mol Med*. 2017;23:874–98.
36. Fiskus W, Sharma S, Shah B, Portier BP, Devaraj SG, Liu K, Iyer SP, Bearss D, Bhalla KN. Highly effective combination of LSD1 (KDM1A) antagonist and pan-histone deacetylase inhibitor against human AML cells. *Leukemia*. 2014;28:2155–64.
37. Zhang W, Zhao C, Zhao J, Zhu Y, Weng X, Chen Q, Sun H, Mi JQ, Li J, Zhu J, et al. Inactivation of PBX3 and HOXA9 by down-regulating H3K79 methylation represses NPM1-mutated leukemic cell survival. *Theranostics*. 2018;8:4359–71.
38. Bullinger L, Dohner K, Dohner H. Genomics of Acute Myeloid Leukemia Diagnosis and Pathways. *J Clin Oncol*. 2017;35:934–46.

39. Rocquain J, Carbuccia N, Trouplin V, Raynaud S, Murati A, Nezri M, Tadriz Z, Olschwang S, Vey N, Birnbaum D, et al. Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC Cancer*. 2010;10:401.
40. Niktoreh N, Walter C, Zimmermann M, von Neuhoff C, von Neuhoff N, Rasche M, Waack K, Creutzig U, Hanenberg H, Reinhardt D. Mutated WT1, FLT3-ITD, and NUP98-NSD1 Fusion in Various Combinations Define a Poor Prognostic Group in Pediatric Acute Myeloid Leukemia. *J Oncol*. 2019;2019:1609128.
41. Wang Y, Xiao M, Chen X, Chen L, Xu Y, Lv L, Wang P, Yang H, Ma S, Lin H, et al. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. *Mol Cell*. 2015;57:662–73.
42. Pronier E, Bowman RL, Ahn J, Glass J, Kandoth C, Merlinsky TR, Whitfield JT, Durham BH, Gruet A, Hanasoge Somasundara AV, et al. Genetic and epigenetic evolution as a contributor to WT1-mutant leukemogenesis. *Blood*. 2018;132:1265–78.
43. Ayatollahi H, Shajiei A, Sadeghian MH, Sheikhi M, Yazdandoust E, Ghazanfarpour M, Shams SF, Shakeri S. Prognostic Importance of C-KIT Mutations in Core Binding Factor Acute Myeloid Leukemia: A Systematic Review. *Hematol Oncol Stem Cell Ther*. 2017;10:1–7.
44. Jawhar M, Döhner K, Kreil S, Schwaab J, Shoumariyeh K, Meggendorfer M, Span LLF, Fuhrmann S, Naumann N, Horny HP, et al. KIT D816 mutated/CBF-negative acute myeloid leukemia: a poor-risk subtype associated with systemic mastocytosis. *Leukemia*. 2019;33:1124–34.
45. Coombs CC, Tallman MS, Levine RL. Molecular therapy for acute myeloid leukaemia. *Nat Rev Clin Oncol*. 2016;13:305–18.
46. Braun TP, Eide CA, Druker BJ. Response and Resistance to BCR-ABL1-Targeted Therapies. *Cancer Cell*. 2020;37:530–42.
47. Moslehi JJ, Deininger M. Tyrosine Kinase Inhibitor-Associated Cardiovascular Toxicity in Chronic Myeloid Leukemia. *J Clin Oncol*. 2015;33:4210–8.
48. Heinrich MC, Griffith DJ, Druker BJ, Wait CL, Ott KA, Zigler AJ. Inhibition of c-kit receptor tyrosine kinase activity by STI 571, a selective tyrosine kinase inhibitor. *Blood*. 2000;96:925–32.
49. Guida T, Anaganti S, Provitera L, Gedrich R, Sullivan E, Wilhelm SM, Santoro M, Carlomagno F. Sorafenib inhibits imatinib-resistant KIT and platelet-derived growth factor receptor beta gatekeeper mutants. *Clin Cancer Res*. 2007;13:3363–9.
50. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26:1572–3.
51. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374:20150202.
52. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.

53. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautes-Fridman C, Fridman WH, de Reynies A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17:218.
54. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12:453–7.
55. Burugu S, Dancsok AR, Nielsen TO. Emerging targets in cancer immunotherapy. *Semin Cancer Biol.* 2018;52:39–52.
56. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5:1315–6.
57. Grimes DA. The nomogram epidemic: resurgence of a medical relic. *Ann Intern Med.* 2008;149:273–5.
58. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *The Lancet Oncology.* 2015;16:e173–80.
59. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *Jama.* 2017;318:1377–84.
60. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *Jama.* 2015;313:409–10.

Figures

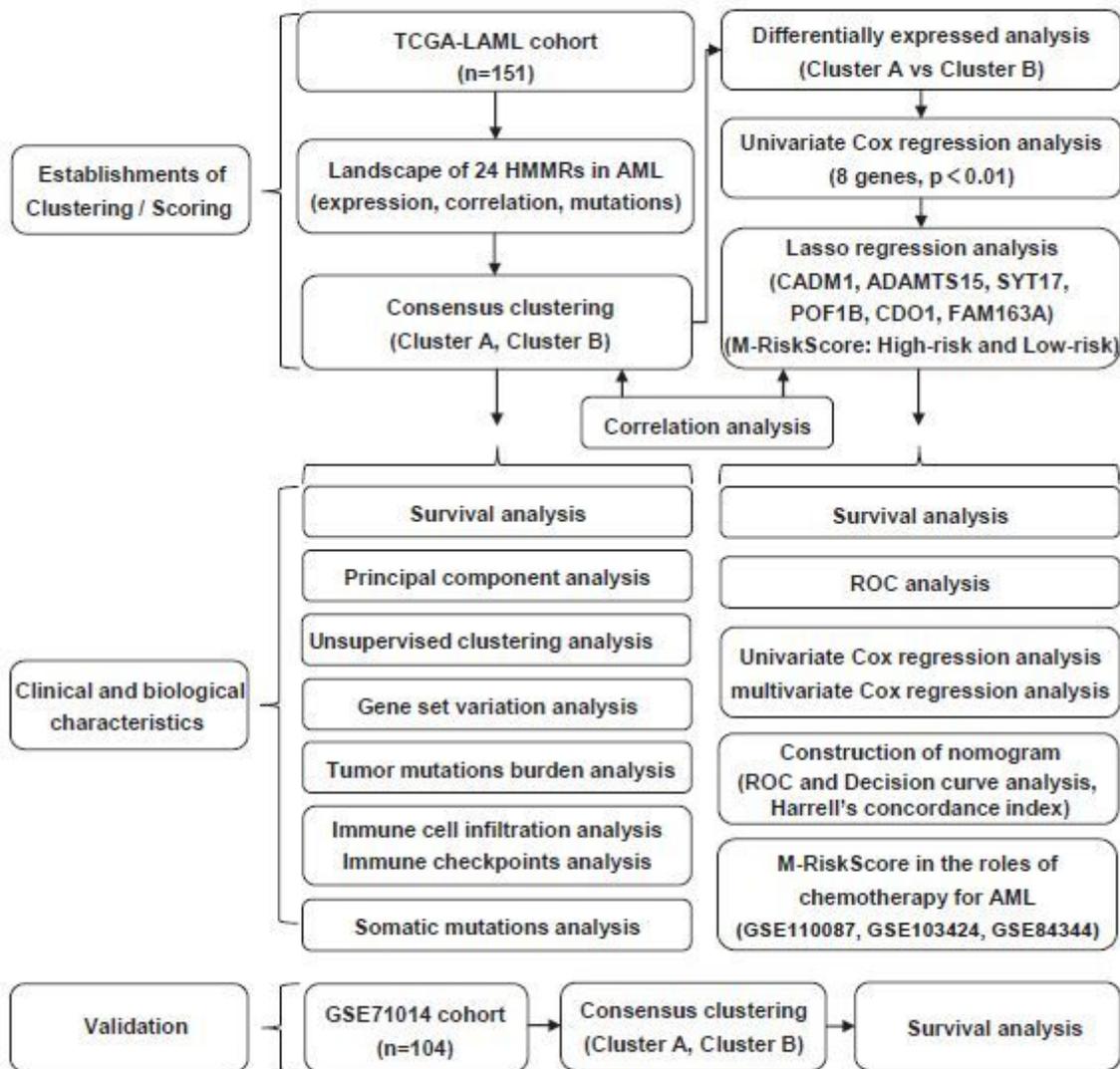


Figure 1

Overview of this work.

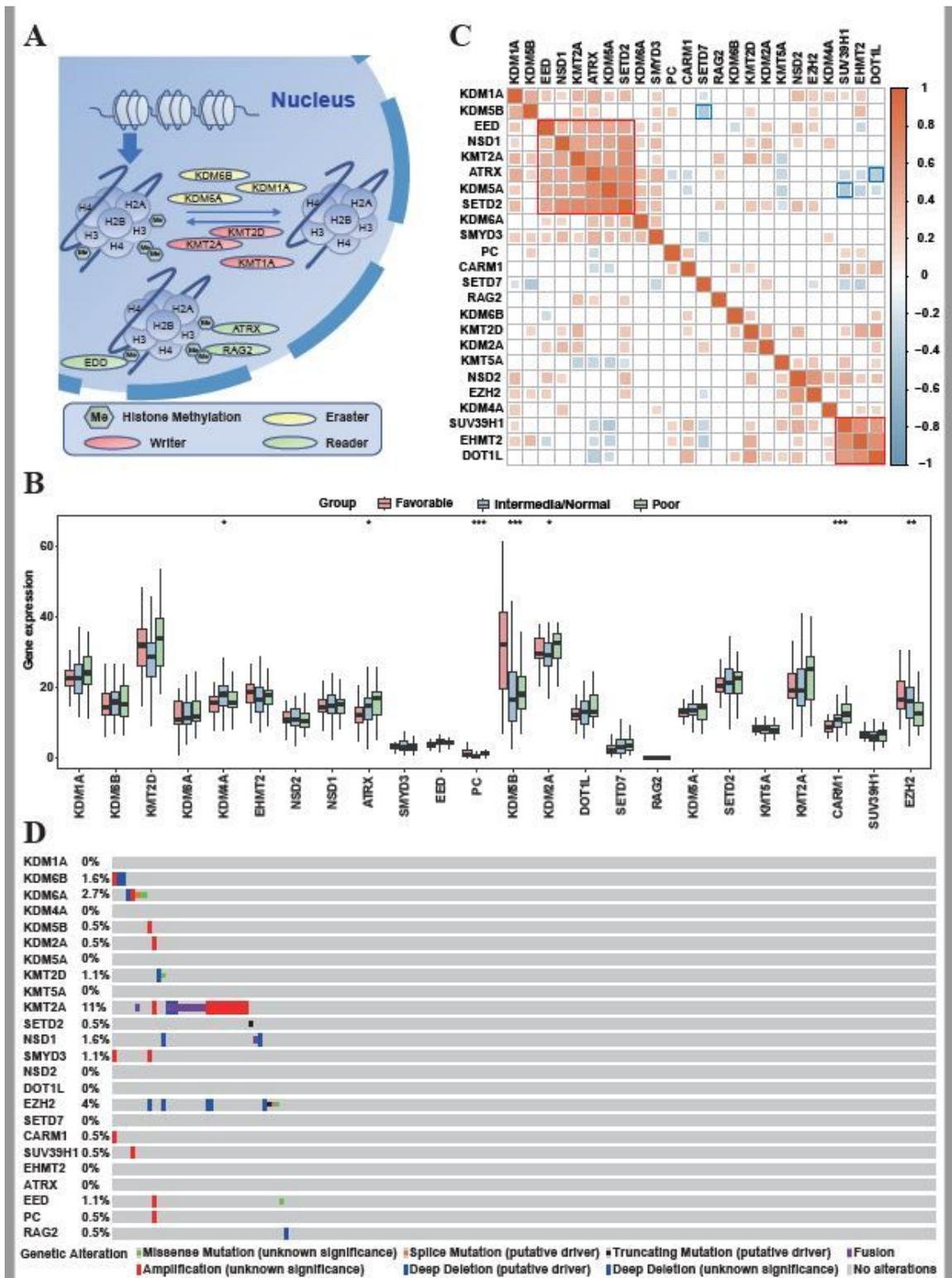


Figure 2

Landscape of genetic variation of 24 HMMRs in AML. A. Graphical summary of dynamic process of histone methylation modification. B. The expression of 24 HMMRs among AML with different Riskstatus. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. C. Co-expression analysis of 24 HMMRs in AML. Positive co-expression, red; Negative co-expression, blue. D. Overview of somatic mutations and chromosomal variations of 24 HMMRs in AML.

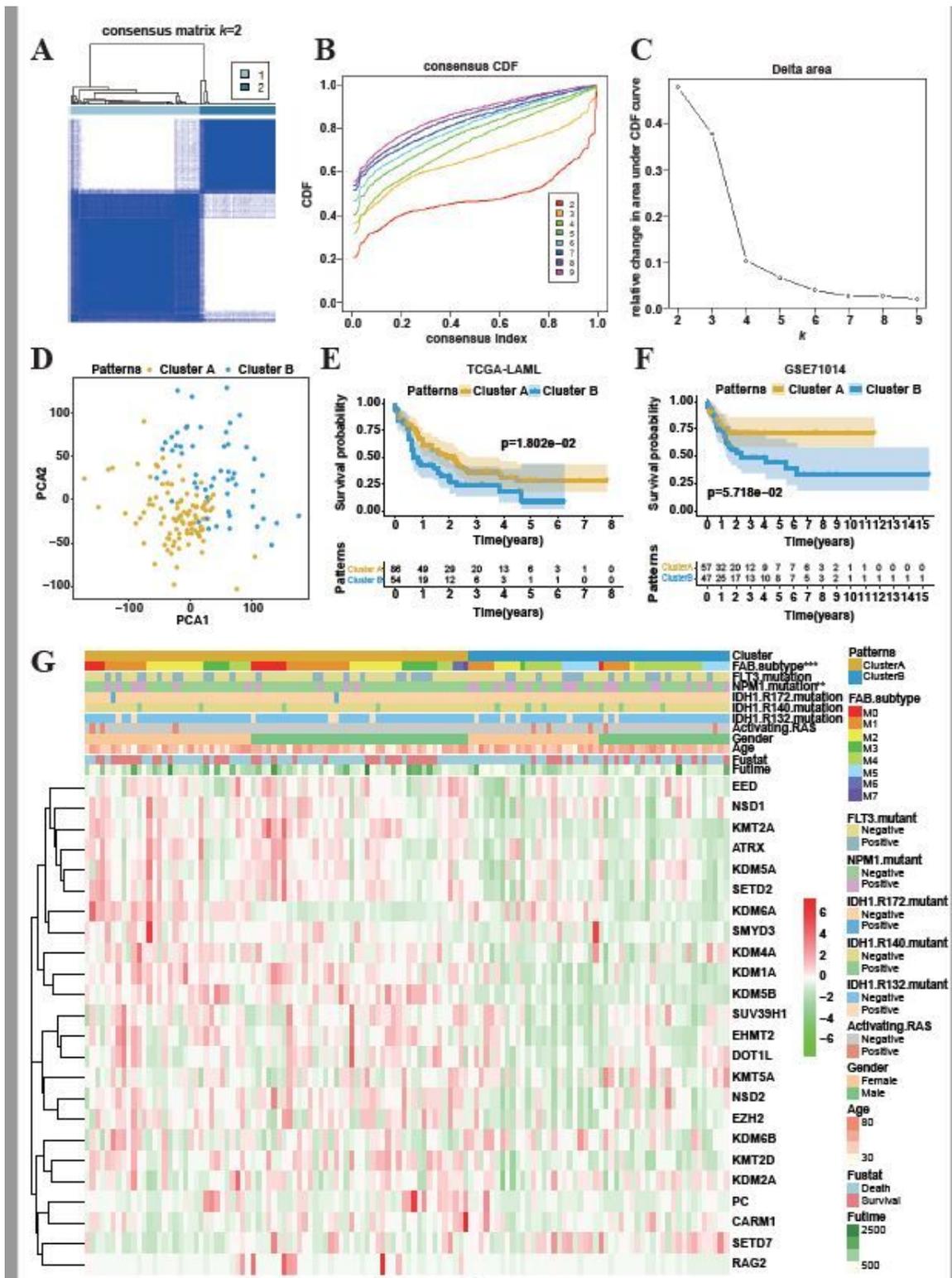


Figure 3

The establishment of histone methylation modification patterns in AML A. Consensus matrix of TCGA-LAML cohort for k value is equal to 2. B. Cumulative distribution function (CDF) plot of the consensus matrices for k = 2-9. C. Delta area plot of CDF plot. D. Principal component analysis for the transcriptome profiles of the two histone methylation modification patterns. E. Survival analysis for the two histone methylation modification patterns of TCGA-LAML cohort. F. Survival analysis for the two histone

methylation modification patterns of GSE71014 cohort. G. Unsupervised clustering of 24 HMMRs in the TCGA-LAML cohort. The histone methylation modification patterns, FAB subtype, activating RAS, gender, age, fustant, fustime and mutations of genes (FLT3, NPM1, IDH1-R172, IDH1-R140, IDH1-R132) were used as annotation. ** p<0.01; *** p<0.001.

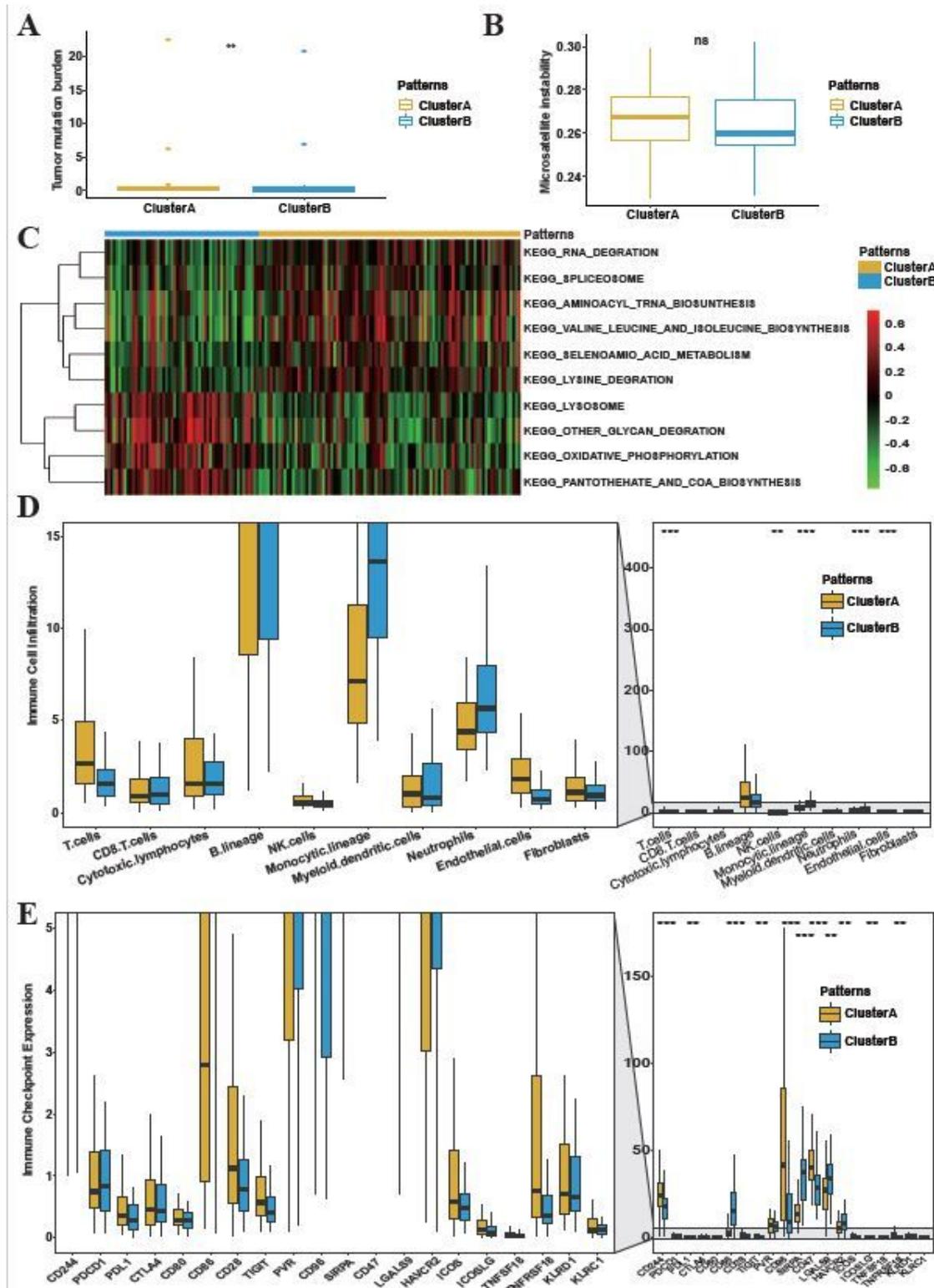


Figure 4

Biological characteristics of the two histone methylation modification patterns A. Tumor mutation burden analysis for two histone methylation modification patterns. ** $p < 0.01$. B. Microsatellite instability analysis for two histone methylation modification patterns. ns: no statistically significant. C. GSVA enrichment analysis showing the activation states of biological pathways in distinct histone methylation modification patterns. The heatmap was used to visualize these biological processes. Activated pathways were represented by red and inhibited pathways were represented by blue. D. The abundance of each TME infiltrating cell in two histone methylation modification patterns. ** $p < 0.01$; *** $p < 0.001$. E. The different expression of several immune checkpoint proteins between the two histone methylation modification patterns. ** $p < 0.01$; *** $p < 0.001$.

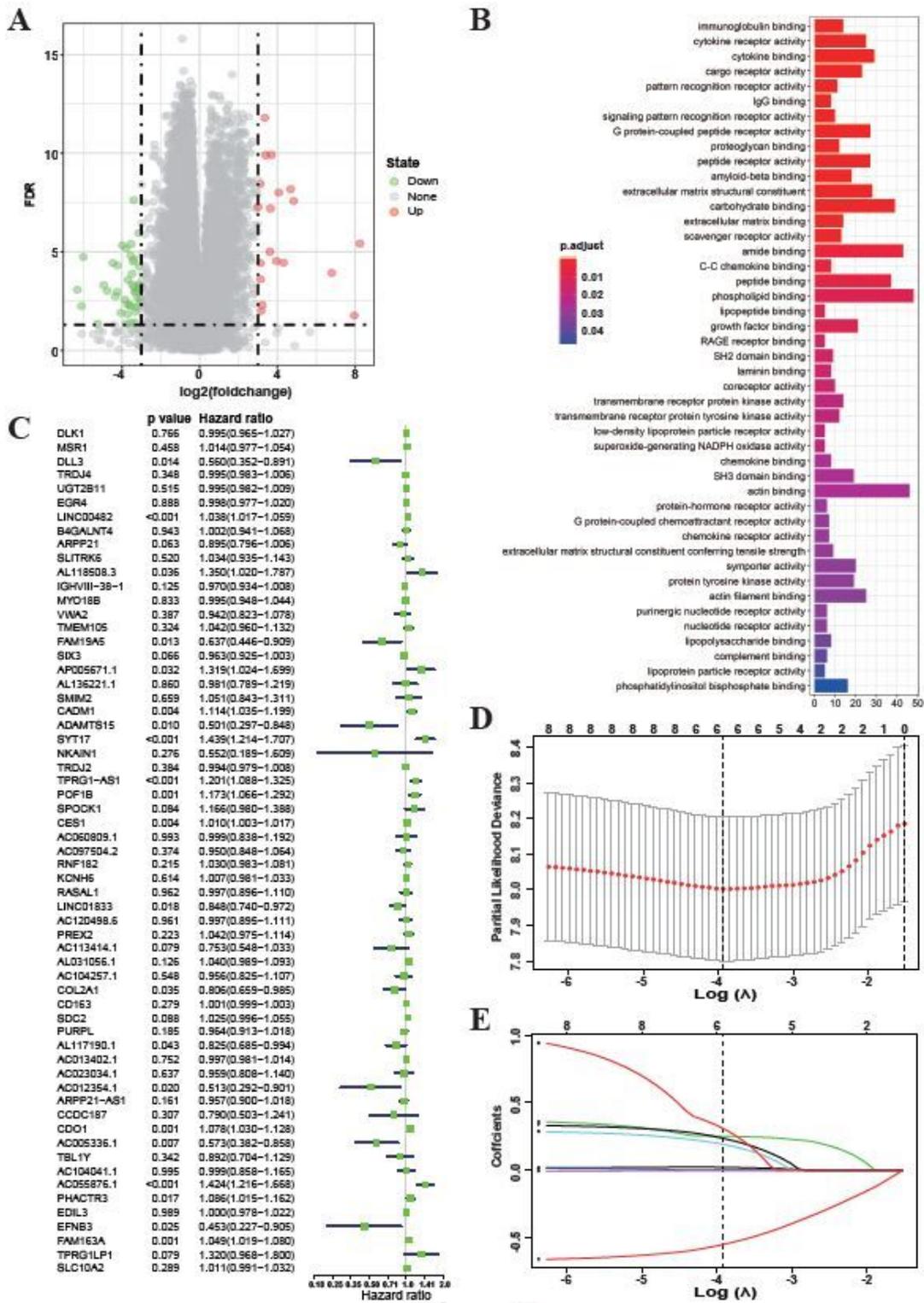


Figure 5

The establishment of M-RiskScore A. A volcano plot showing the different expression genes (DEGs) between the two histone methylation modification patterns. The value of log2 fold change is equal to ± 3 . B. Functional annotation for 61 DEGs using GO enrichment analysis. The color depth of the barplots represented the activation of the pathways. The length of the barplots represented the number of the genes enriched. C. The prognostic analyses for 61 DEGs in the TCGA-LAML cohort using a univariate Cox

regression model. The genes highlighted were selected for lasso regression analysis. D and E. The results of lasso regression analysis. The determination of the value of λ (D) and the determination of the coefficients' values of the selected genes from C (E).

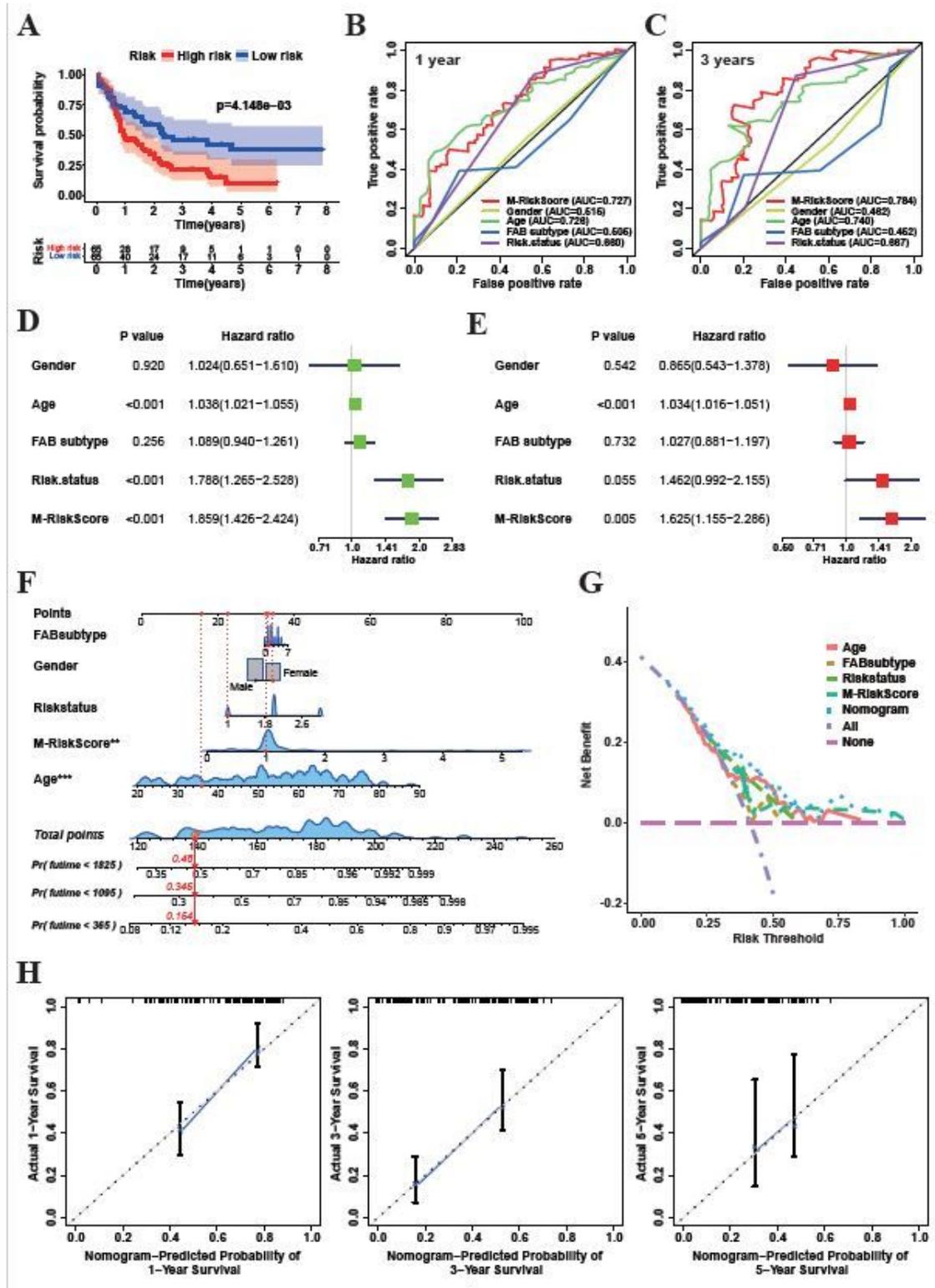


Figure 6

The value of M-RiskScore in prognostic prediction of AML. A. Survival analysis for low and high M-RiskScore patient groups in the TCGA-LAML cohort. B-C. ROC analyses for prognostic prediction accuracy

of 3-years survival rate (B) and 5-years survival rate (C) of M-RiskScore, Gender, Age, FAB subtype and Riskstatus in TCGA-LAML cohort. D-E. Univariate Cox regression model (D) and multivariate Cox regression model (E) for prognostic analyses of M-RiskScore, Gender, Age, FAB subtype and Riskstatus in TCGA-LAML cohort. F. Construction of nomogram making up of M-RiskScore, Gender, Age, FAB subtype and Riskstatus. G. The decision curve analysis for nomogram and its components. H. The calibration plot of nomogram for the probability of survive of 1 year, 3 years and 5 years.

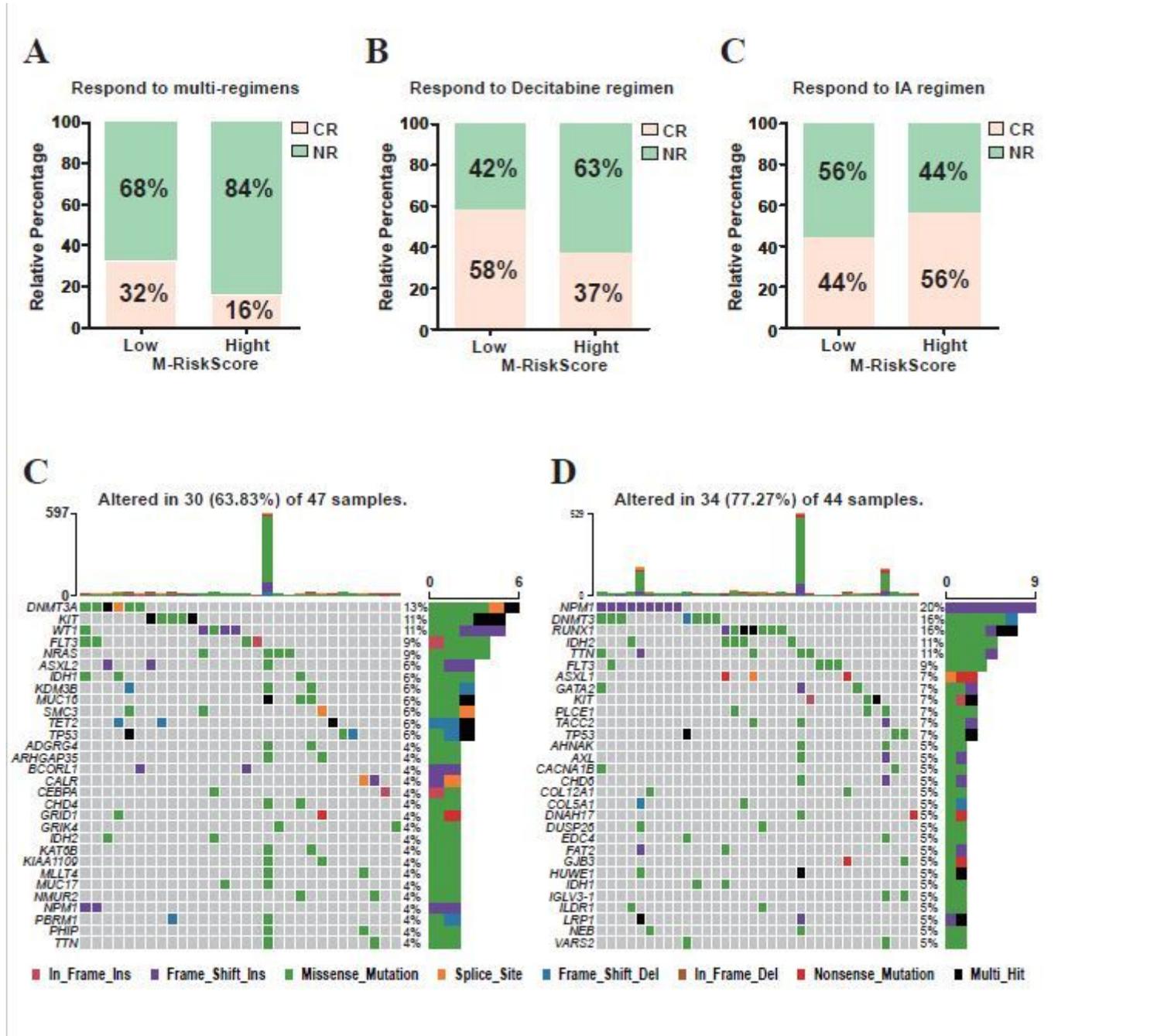


Figure 7

M-RiskScore in the role of chemotherapy of AML A. Correlation analysis of M-RiskScore with complex-chemotherapeutic respond in GSE110087-AML cohort. B. Correlation analysis of M-RiskScore with decitabine treatment respond in GSE84334-AML cohort. C. Correlation analysis of M-RiskScore with IA

regimen treatment respond in GSE103424-AML cohort. D. Somatic mutation analysis of low (left) and high (right) M-RiskScore patients' groups with AML in TCGA-LAML cohort.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.pdf](#)