Gene mutation patterns and their prognostic impact in a cohort of 1,185 patients with acute myeloid leukemia

Running title: GENE MUTATIONS AND PROGNOSIS IN AML

Yang Shen1,3, Yong-Mei Zhu1,3, Xing Fan1,3, Jing-Yi Shi1,3, Qin-Rong Wang1,3, Xiao-Jing Yan1, Zhao-Hui Gu1, Yan-Yan Wang1, Bing Chen1, Chun-Lei Jiang1, Han Yan1, Fei-Fei Chen2, Hai-Min Chen1, Zhu Chen1, Jie Jin2, Sai-Juan Chen1

1. State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, 197 Rui Jin Road II, Shanghai 200025, China.

2. Department of Hematology, Institute of Hematology, the First Affiliated Hospital, Zhejiang University School of Medicine.

3. These authors contributed equally to this work.

Correspondence should be addressed to S-J.C. (sjchen@stn.sh.cn), J.J. (Jie.J@hznc.com) or Z.C. (zchen@stn.sh.cn).

Tel: +86-21-33562658

Fax: +86-21-64743206
Abstract

To evaluate the prognostic value of genetic mutations for acute myeloid leukemia (AML) patients, we examined the gene status for both fusion products such as AML1 (CBFα)-ETO, CBFβ-MYH11, PML-RARα and MLL rearrangement as a result of chromosomal translocations and mutations in genes including FLT3, C-KIT, N-RAS, NPM1, CEBPA, WT1, ASXL1, DNMT3A, MLL, IDH1, IDH2 and TET2 in 1,185 AML patients. Clinical analysis was mainly carried out among 605 cases without recognizable karyotype abnormalities except for 11q23. 452 out of these 605 patients (74.7%) were found to have at least one mutation, and the relationship of gene mutations with clinical outcome was investigated. Notably, we revealed a correlation pattern among NPM1, DNMT3A, FLT3, IDH1, IDH2, CEBPA and TET2 mutations. Multivariate analysis identified DNMT3A and MLL mutations as independent factors predicting inferior overall survival (OS) and event free survival (EFS), while bi-allelic CEBPA mutations or NPM1 mutations without DNMT3A mutations conferred a better OS and EFS in either the whole group or among younger patients aged less than 60 years. The use of molecular markers allows us to subdivide the series of 605 cases into distinct prognostic groups with potential clinical relevance.
**Introduction**

Acute myeloid leukemia (AML) is a group of heterogeneous diseases, with considerable diversities in terms of clinical behavior and prognosis.\(^1\)\(^-\)\(^5\) Clinical and genetic prognostic markers are now crucial in evaluation of AML patients and guiding rational management. Among them, cytogenetic abnormalities are considered as the most important prognostic factors of AML.\(^5\)\(^-\)\(^11\) For example, acute promyelocytic leukemia (APL, AML-M3) with t(15;17) translocation is associated with a favorable prognosis,\(^12\)\(^-\)\(^14\) and core binding factor (CBF) leukemias with t(8;21) translocation in AML-M2 variant (M2v) or inv(16) rearrangement in AML-M4 with eosinophilia (M4eo) have also been reported to have relatively good outcome.\(^5\)\(^,\)\(^7\)\(^,\)\(^15\) Nevertheless, approximately half of the AML patients lack typical prognostic karyotypic changes. To improve clinical outcome, identification of specific and accurate predictors in this group of patients by using molecular approaches is of important value.

It has been proposed that according to their roles in pathogenesis, genetic abnormalities in leukemia can be roughly grouped into two classes: mutations involving signal transduction pathways and giving rise to proliferative advantages to leukemia clones (class I), or those affecting transcription factors or co-factors and causing impaired differentiation (class II).\(^16\) Indeed, numerous gene mutations have been discovered in AML patients without cytogenetic markers and these abnormalities have been considered to belong to either class I, as exemplified by internal tandem duplications (ITDs) or mutations of tyrosine kinase domain (TKD) of both \(C-KIT\)\(^17\)\(^,\)\(^18\) and FMS-like tyrosine kinase 3 (\(FLT3\)) genes\(^11\)\(^,\)\(^19\)\(^,\)\(^20\), and point mutations of
neuroblastoma RAS viral oncogene homolog gene (NRAS)\textsuperscript{17,21-24}, or to class II, such as mutations of nucleophosmin gene (NPM1)\textsuperscript{2,4,22,25}, the CCAAT/enhancer binding protein α gene (CEBPA), Wilm’s tumor (WT1) gene and additional sex combs like 1 (ASXL1).\textsuperscript{26-30} Recently, a new category of gene mutations associated with epigenetic regulation have drawn much attention, including the mutations of isocitrate dehydrogenase1 (IDH1), isocitrate dehydrogenase2 (IDH2) and ten eleven 2 (TET2) mutation which result in a hypermethylation phenotype with impairment of hematopoietic differentiation.\textsuperscript{31-33} The mixed-lineage leukemia (MLL) gene, which can be affected either through chromosomal translocation or via an intragenic partial tandem duplications (PTDs) to form fusion gene, actually encodes a histone methyltransferase.\textsuperscript{22,34-36} Discovery of these genetic events has raised the possibility of a new class of leukemogenic genes.

Importantly, the above-mentioned molecular aberrations exert profound effects on individual response to the therapy and treatment outcome of the disease.\textsuperscript{14,16,37} It has been described that CEBPA mutations\textsuperscript{27,28} and NPM1 mutations without FLT3-ITD\textsuperscript{16,20,23,25} are associated with a favorable prognosis; whereas gene mutations related with poor prognosis include C-KIT involvement among CBF AMLs\textsuperscript{17,18,23}, FLT3-ITD without NPM1 mutations,\textsuperscript{16,20,23,25} and MLL-PTD mutations among AML with normal cytogenetics.\textsuperscript{22,34,35} However, the role of leukemic IDH1 and IDH2 mutations as well as WT1, TET2 and ASXL1 mutations in predicting the prognosis of AML are not clearly established.\textsuperscript{31-33,38} In addition, these markers may also provide potential molecular targets for tailored therapies, as recently reported by several
groups demonstrating potential clinical values of sorafenib in the patients with FLT3 ITD and TKD, and azacitidine and decitibine in patients with MLL abnormalities. More recently, genomic sequencing researches in AML and other malignancies have greatly facilitated identification of new oncogenic mutations. Of interest, we and Ley’s group have discovered mutations in DNA methyltransferase 3A gene (DNMT3A) in more than 20% of acute monocytic leukemia by exome sequencing and subsequent Sanger sequencing. The enzyme encoded by DNMT3A gene is responsible for de novo DNA cytosine methylation. Both studies suggested that DNMT3A mutations are associated with hyperleukocytosis at disease presentation, elderly age and poor prognosis. With the accumulation of more new data, decision making of risk-stratified therapy will be possible and should be integrated into the individualized treatment of AML.

We performed this study to systemically investigate the frequencies and the prognostic relevance of previously known genetic events and newly established molecular markers in a large series of adult AML patients. In particular, we intent to stratify the “ambiguous” AML patients who lacked cytogenetic prognostic markers into appropriate prognostic groups by using molecular markers.

Patients, materials, and methods

Patients

Bone marrow (BM) and peripheral blood (PB) samples were collected from 1,185 patients with de novo AML from 1998 to 2010 from the centers of Shanghai Institute
of Hematology (SIH) and Zhejiang Institute of Hematology (ZIH). French-American-British (FAB) criteria were used to define the AML subtypes (M0 through M7, with a few cases not classifiable according to the FAB nomenclature). Patients with leukemia either transformed from myelodysplasia syndrome (MDS) or secondary to other malignancies were excluded from this study.

All the samples were assessed for overview of pattern and distribution of gene mutations, and they were further divided into three groups. Group I containing 605 cases, without prognostic cytogenetic markers except for 11q23, which represented the focus of this study for clinical relevance of gene abnormalities and prognostic analysis. Because 11q23 rearrangements are mostly associated with MLL fusion genes and have been considered to bear similar clinical impact as MLL-PTD mutations\textsuperscript{34-36,44}, which are not recognizable at karyotypic level, patients with these chromosomal changes were included into the series of 605 cases for prognostic analysis. Groups II and III consisted of, respectively, the two most common leukemia subtypes with translocation, namely CBF leukemias with AML1 (CBF\textalpha)-ETO (158 cases) or CBF\beta-MYH11 fusion (18 cases) and APL with PML-RAR\textalpha fusion (387 cases) or the rare variant NPM1-RAR\textalpha (1 case)\textsuperscript{6,14,15}. Although the prognostic value of these two groups have been established, these patients were included in order to overview the frequencies, distribution and correlations of molecular mutations in the whole AML setting. In addition, there was a small group of 16 cases with relatively rare prognostic cytogenetic markers\textsuperscript{45} including t(3;3), t(9;22), -7, del (5q), del (7q) and complex translocation (for detail, see Supplementary Table 1).
This study was approved by the ethnic board of all participating centers. All patients were given informed consent for both treatment and cryopreservation of BM and PB according to the Declaration of Helsinki.

**Treatment protocols**

For APL (AML-M3) patients with t (15; 17), all-trans retinoic acid (ATRA) and arsenic trioxide (ATO) based treatment was given for the induction and consolidation therapy.\(^{14}\)

Other AML patients received standard first line treatment of DA like regimen, which consisted of daunorubicin 45mg/m\(^2\), D1-3; and Ara-C 100-150mg/m\(^2\), D1-7. In the consolidation therapy, young patients were treated with high-dose cytarabine based chemotherapy. Due to small number of the patients (47 cases) received allogenetic stem cell transplantation (allo-SCT), they were not separated for further analysis. The chemotherapy consolidation for elderly patients was mainly decided by the physicians in an individualized manner, as previously described.\(^{46}\)

**Cytogenetic and molecular genetic analysis**

Cytogenetic and molecular studies were performed centrally in SIH and ZIH. The BM samples of *de novo* AML patients were studies mostly by R- and/or G-banding analysis, and were confirmed in most cases with relevant molecular markers. Chromosomal abnormalities were described according to the International System for Human Cytogenetic Nomenclature.\(^{9}\)
Genomic DNA and total RNA were extracted as previously reported. Initially, we screened mutations in a number of genes including *MDR1*, *BCL2*, *P53*, *XPA*, *ATM*, *SULT1C2*, *KIAA1244*, *COL7A1*, *N-RAS*, *NPM1* and *IDH1* in a part of this series of patients. Since sequence variations in the first 8 genes were proved to be single nucleotide polymorphisms (SNPs) when checked in 384 control samples from unrelated healthy individuals (data not shown) and they were not included in current guideline or practice of prognostic predicting, no further analysis were performed on these genes, while the remaining 3 genes showed a certain percentage of mutations. Our efforts were then focused on these 3 genes and other previously known mutations, such as *FLT3*-ITD and -TKD, and those in *C-KIT*, *CEPBA*, *WT1*, *ASXL1*, *DNMT3A*, *MLL*, *IDH2* and *TET2* genes by distinct approaches. Because the mutations of *FLT3*-TKD, *N-RAS*, *NPM1*, *IDH1* and *IDH2* were clearly concentrated, we used a chip-based matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis system (iPLEXTM, Sequenom, San Diego, CA, USA) to perform high throughput genotyping assays for the analysis of mutational status of these genes. For mutations of *FLT3*-ITD (in ITD region), and those in *C-KIT*, *CEPBA*, *WT1*, *ASXL1*, *DNMT3A* and *TET2* genes, samples were analyzed by whole gene sequencing. In addition, a multiplex RT-PCR strategy was used to detect 6 *MLL*-related common fusion genes including *MLL-AF9*, *MLL-AF10*, *MLL-AF6*, *MLL-ELL*, *MLL-ENL* and *MLL-AF17*. Briefly, all samples were screened with 2 parallel multiplex RT-PCR reactions. If there were positive PCR fragments in the samples, split-out PCR was performed to determine the fusion gene type.
mutational status of MLL-PTD and fusion genes such as AML1(CBFα)-ETO, CBFβ-MYH11 and PML-RARα were determined by RT-PCR technique.22

Quantitative real-time RT-PCR

After Turbo DNase (Ambion) treatment, 1 μg total RNA was used for cDNA synthesis using M-MLV First Strand Kit (Invitrogen). Real-time PCR was performed in ABI PRISM7900HT using SYBR Premix Ex Taq™ (TaKaRa). Fold change was calculated based on $2^{-\Delta Ct} \times 1000$ method after normalization to the transcript level of housekeeping gene GAPDH. Primer sequences used in the real-time RT-PCR are listed in supplementary Table 2.

Microarray expression profiling and methylation analysis

Affymetrix Human Genome U133 Plus 2.0 Array GeneChip microarrays were used to assess the total RNA samples (Affymetrix). DNA samples were extracted for the HG18 Methylation 2.1M Deluxe Promoter Array (NimbleGen) to identify the methylated DNA regions. The procedure and statistical analysis were performed as previously reported.40 All microarray data are available in the NCBI Gene Expression Omnibus (GEO) under accession number GSE27244.

Statistical analyses

For clinical analysis, complete remission (CR) was defined according to the criteria of International Working Group.48 Fisher’s exact P test was used to compare
the difference of CR rates. One way ANOVA test was used to compare the age, WBC count, BM blasts at diagnosis in different groups. Relationship between different gene mutations was analyzed by Kendall’s tau-b correlation coefficients. OS was measured from the date of disease diagnosis to death (failure) or alive at last follow-up (censored).\(^{48}\) Event free survival (EFS) was defined as from disease diagnosis to treatment failure such as relapse, refractory disease, death, or alive in CR at last follow-up (censored). Kaplan-Meier analysis was used to calculate the distribution of OS and EFS.\(^{48}\) Log rank comparison was performed to compare the difference of survivals. Binary logistic regression and COX model was used for the multivariate analysis of associations between mutational status and the achievement of CR and OS and EFS, respectively.\(^{49}\) A limited backward selection procedure was used to exclude redundant variates.\(^{49}\) All above statistical procedures were performed with the SPSS statistical software package, version 16.0.

**Results**

**Patient characteristics**

The characteristics of the 1,185 de novo AML patients are summarized in Table 1. Particular attentions for clinical analysis were given to 605 cases who lacked cytogenetic prognostic markers other than 11q23 abnormalities (group I, 51.1%). The relatively high proportion of patients with group III (APL, AML-M3) could result from a preferred choice for SIH by patients owing to the successful ATRA/ATO based therapy. And, our data showed a relatively low incidence of M4eo given a careful
examination of \textit{CBF-MYH11} was performed in all morphologically M4 patients, which might reflect the difference in genetic backgrounds between Chinese and Caucasian populations, concurrent to a previous report on a large Chinese AML series.\textsuperscript{50}

**Frequencies and distribution of gene mutations**

Among the 605 Group I patients, \textit{FLT3} mutations were found in 61 (10.8%), \textit{C-KIT} in 30 (5.4%), \textit{N-RAS} in 34 (5.9%); \textit{NPM1} in 122 (20.9%), \textit{CEBPA} in 123 (22.0%), \textit{WT1} in 20 (3.7%), \textit{ASXL1} in 27 (5.2%), \textit{DNMT3A} in 73 (12.3%), \textit{MLL} in 83 (14.0%), \textit{IDH1} in 52 (9.3%), \textit{IDH2} in 53 (9.8%), and \textit{TET2} in 65 (12.7%). Four hundred and fifty two (74.7%) cases were found to have at least one mutation. In Group II with \textit{CBF} leukemias, the most frequent mutations in addition to those caused by chromosomal translocations were \textit{C-KIT} (25.6%) and \textit{N-RAS} (9.7%), which differed from the situation in Group III (APL, AML-M3) where additional mutations than \textit{PML-RAR\alpha} were mainly those of \textit{FLT3} (13.4%) and \textit{N-RAS} (5.4%) (Supplementary Figure 1).

As we recently reported\textsuperscript{40}, mutant \textit{DNMT3A} gene was mostly associated to a myelomonocytic or monocytic morphology in FAB classification (P<0.001) with a frequency of 10.0% (22/219) and 18.2% (38/209) in M4 and M5, respectively. Of note, we identified 13 potential new types of \textit{DNMT3A} mutations which are denoted in Figure 1A. All these sequence variations were checked in a series of 384 control samples from unrelated healthy individuals and none of them was observed. We found three M1, one M2 with \textit{AML1 (CBF\alpha)-ETO} fusion, one M3 with \textit{PML-RAR\alpha} fusion,
two M6 and two M2 without CBF fusion patients carried previously reported DNMT3A mutations\textsuperscript{37, 40} as well. Of note, the two patients with M6 presented monocytic features in non-erythroid lineages, while each of the M2 CBF and M3 patients experienced very soon central nervous system (CNS) involvement in spite of a standard intrathecal prophylaxis.

With regard to the class I gene mutations, FLT3-ITD and/or -TKD mutations exhibited an extremely lower incidence in CBF leukemias (2/176, 1.1%) than in other subtypes of AML (P<0.001), whereas a high frequency being present in addition to PML-RAR\textalpha fusion in M3 subtype. On the other hand, C-KIT mutations were most commonly seen in CBF leukemias (45/176, 25.6%, P<0.001). In contrast, N-RAS mutations were distributed evenly in different AML subtypes. Interestingly, among 259 patients bearing class I mutations, only 8 (3.1%) had overlapping of these markers while all others carried mutation of only one gene.

Concerning the class II mutations involving NPM1 and CEBPA, as well as mutants of genes related to epigenetic regulation such as DNMT3A, MLL, IDH1, IDH2 and TET2, they tended to occur in AML patients without chromosome translocations (All P<0.001). However, ASXL1 (P=0.116) and WT1 (P=0.296) mutations seemed to be distributed equally in cytogenetically normal or abnormal AML groups. Interestingly, in contrast to DNMT3A (P=0.002) and MLL (P=0.004) mutations which were closely related to a M4 or M5 phenotype, CEBPA mutations favored a non-M4 and non-M5 phenotype (P<0.001). Mutation of another epigenetic regulator, TET2, also showed this tendency (P=0.056). The phenomenon of overlapping mutations in the same case
was quite frequent. Some mutations were even statistically correlated. \textit{DNMT3A} (43/129, \textit{P}<0.001), \textit{FLT3} (23/129, \textit{P}<0.001), \textit{IDH1} (17/129, \textit{P}=0.046), \textit{IDH2} (26/129, \textit{P}<0.001) and \textit{TET2} (16/129, \textit{P}=0.030) were significantly associated with \textit{NPM1} mutations. The association between \textit{DNMT3A} and \textit{FLT3} as well as \textit{IDH1} and \textit{IDH2} was also observed (\textit{P}=0.003, 0.010, and 0.001 respectively). Among Group I patients, frequent associations between the mutations of \textit{NPM1} and \textit{FLT3} (\textit{P}<0.001), \textit{DNMT3A} (\textit{P}<0.001), \textit{IDH1} (\textit{P}=0.056) and \textit{IDH2} (\textit{P}<0.001), but not \textit{TET2} (\textit{P}=0.710), were observed. While mutations of \textit{CEBPA} and \textit{TET2} were highly correlated in all the patients (\textit{P}<0.001), only marginal correlation was identified in Group I patients (\textit{P}=0.053). Notably, \textit{MLL} abnormalities seldom co-existed with other mutations.

Figure 2 represents the gene mutational status, including their distribution and frequencies among distinct AML subtypes.

**Correlation between gene expression levels and DNA methylation status associated with \textit{DNMT3A} mutations**

We previously reported that aberrant DNA methyltransferase activity due to \textit{DNMT3A} mutations could change DNA methylation and alter gene expression.\textsuperscript{40} Interestingly, through in-depth exploration of the possible consequences of the \textit{DNMT3A} mutations using microarray data of expression profile and whole genomic DNA methylation, we identified a correlation between gene expression levels and the DNA methylation status in some genes. 28 up-regulated genes with DNA hypomethylation and 47 down-regulated genes accompanied with DNA
hypermethylation were revealed (Figure 1B). Among them, some genes were associated with hematopoiesis and epigenetics regulation, including CCDC56, DCXR, TNFSF13 and SLC25A11 (up-regulated) and SCL25A37 and EIF4G3 (down-regulated). A correlation was revealed between the gene expression levels of CCDC56, DCXR and TNFSF13 and the mutation of DNMT3A, whereas no up-regulation of these 3 genes was observed in patients with MLL rearrangement (Figure 1C).

It has been reported that increased expression of multiple homeobox (HOX) genes such as HOXA7, HOXA9, HOXA10 and MEIS1 due to MLL abnormalities play an important role in leukemogenesis. Recently, we demonstrated an up-regulation of HOXB family genes and IDH1 gene among patients with DNMT3A mutations in our previous work. We also observed from this study that though MLL and DNMT3A mutations seldom overlapped, they shared a common feature of poor prognosis in distinct patient populations. Therefore, we tried to address possible association of expression levels of different HOX genes with MLL or DNMT3A abnormalities. Interestingly, members of the HOXB family were found over-expressed only in a group with DNMT3A mutations, while up-regulation of MEIS1 gene could be observed only in patients with MLL abnormalities. Over-expression of HOXA7, HOXA9 and HOXA10 genes was observed in both DNMT3A and MLL abnormalities groups, contrarily to the situation of HOXA5 and HOXA13 of which the expression was not affected by DNMT3A and MLL abnormalities (Figure 1D). Although we previously found hypomethylation in genomic sequences of some HOXB family members, which might contribute to over-expression status of these genes, this time no obvious
changes were found in *HOXA* family genes and the exact molecular mechanisms for their up-regulation need further investigation.

**Molecular markers and clinical aspects**

As far as the possible association with clinic features, we found that mutations of *NPM1*, *CEBP A*, and nearly all the epigenetics regulatory genes (*DNMT3A*, *IDH1*, *IDH2* and *TET2*) except for *MLL* rearrangements were related with an elderly age at diagnosis; *FLT3*, *NPM1*, *CEBP A*, *DNMT3A*, *MLL*, and *IDH2* mutations were related with high WBC count at presentation; and *NPM1*, *FLT3*, *DNMT3A*, *MLL* mutations were associated to a higher percentage of blasts in BM (Table 2).

**Response to induction therapy**

In univariate analysis of the 605 cases in Group I, *DNMT3A* mutations, *MLL* abnormalities and *N-RAS* mutations were associated with a statistically significant lower CR rate (45.8%, 48.2%, and 41.2%, P=0.014, 0.022 and 0.030 respectively) in contrast to *CEBP A* mutations (67.5%) which conferred a higher CR rate (P=0.030). However, only bi-allelic (75.3%), but not mono-allelic (56.0%), *CEBP A* mutations were associated with favorable response to the treatment (P=0.003 and 1.000, respectively). Because *NPM1* mutations were frequently associated with abnormalities of *DNMT3A*, *FLT3*, *IDH1* and *IDH2*, we made further analysis using the combination of *NPM1* mutations with each of the above-mentioned mutations to investigate their potential prognostic impact. Obviously, *DNMT3A* mutations could separate the
patients with *NPM1* mutations into two distinct prognostic groups: a subpopulation of *NPM1* mutations without *DNMT3A* mutations (*NPM1m+/DNMT3Am-*) was related to a significantly higher CR rate (72.3%, *P*=0.017); however, mutations of *IDH1* (*P*=0.265) and *IDH2* (*P*=0.218), and *FLT3*-ITD (*P*=0.164) or TKD (*P*=0.318) did not add more prognostic value among patients with *NPM1* mutation (supplementary Table 3). And in M4 and M5 patients, similar results were achieved (*P*=0.056 for *DNMT3A* mutations, and *P*=0.742, 0.736, 0.661 and 0.737 for *IDH1*, *IDH2*, *FLT3*-ITD and TKD mutations respectively).

Of note, when 89 cases without gene mutations were grouped, they seemed to have an intermediate prognosis in terms of CR (64.0%).

A complete list of covariates that entered multivariate model was indicated in table 3. Multivariate analysis of Group I patients indicated that *NPM1m+/DNMT3Am-* and bi-allelic *CEBP A* mutations (*CEBP Am+*) were independent factors associated with favorable CR rate, and *DNMT3A* mutations (*DNMT3Am+*) were associated with a lower CR rate. Two other independent clinical factors, WBC count and age, were also unfavorable for CR rate. Of note, in age adjusted population (n=481) who were younger than 60 years and treated intensively, *MLL* abnormalities, *TET2* mutations and BM blasts proportion predicted unfavorable CR rate independently (Table 4).

**Survival analysis**

Among 605 Group I cases, the median OS and EFS were 15.0±1.9 and 8.0±1.2 months, respectively. In univariate analysis, *DNMT3A* mutations and *MLL*
abnormalities suggested a poor prognosis (both $P<0.001$ for OS; and $P=0.001$ and $P<0.001$ for EFS, respectively). Although $N$-RAS mutation cases showed an inferior EFS ($P=0.016$), OS was only marginally affected ($P=0.084$). Favorable OS and EFS were achieved in the $CEBPAm+$ patients ($P=0.002$ and 0.005 for OS and EFS, respectively). Further analysis showed only bi-allelic $CEBPAm+$ status was associated with better treatment outcome (both $P<0.001$ for OS and EFS, respectively). There was no statistical significance of OS and EFS in $FLT3$, $C-KIT$, $WT1$, $ASXL1$, $IDH1$, $IDH2$ and $TET2$ mutations ($P=0.169$ and 0.371, 0.317 and 0.165, 0.815 and 0.587, 0.765 and 0.717, 0.257 and 0.731, 0.339 and 0.770, 0.148 and 0.074, respectively). $NPM1$ mutation did not predict OS ($P=0.409$) and EFS ($P=0.274$). However, in patients with $NPM1$ mutations, it was the mutation of $DNMT3A$ ($P<0.001$ and $P=0.002$ for OS and EFS, respectively), but not of $FLT3$ (ITD: $P=0.231$ and 0.156, for OS and EFS, respectively; TKD: $P=0.314$ and 0.204, for OS and EFS, respectively), $IDH1$ ($P=0.457$ and 0.552 for OS and EFS, respectively) or $IDH2$ ($P=0.863$ and 0.843 for OS and EFS, respectively) that helped to discriminate two different prognostic groups. Figure 3 shows the Kaplan-Meier curves for OS and EFS according to genotypes with statistical significance in univariate analysis. And in M4 and M5 patients, similarly, only $DNMT3A$ mutations ($P=0.001$ and 0.006 for OS and EFS, respectively) could sub-divide the patients with $NPM1$ mutations, and the existence of the mutation of $IDH1$ ($P=0.585$ and 0.941 for OS and EFS, respectively), $IDH2$ ($P=0.275$ and 0.118 for OS and EFS, respectively), $FLT3$-ITD ($P=0.807$ and 0.507 for OS and EFS, respectively), and TKD ($P=0.988$ and 0.621, respectively) could not help to further
stratify the NPM1 m+ patients.

In multivariate analysis, DNMT3A m+ and MLL rearrangements (MLLm+) were independent factors predicting poor prognosis; bi-allelic CEBPAm+ and NPM1 m+/DNMT3A m- conferred a better OS and EFS (table 3). Age and WBC count were also independent factors related to prognosis. Notably, among younger patients who received standard induction and consolidation, the results were similar to the entire group, and the mutational status of the above-mentioned 4 genes still bore prognostic significance while WBC count no longer predicted OS and EFS.

Using molecular markers which were proved significantly related to prognosis in multivariate analysis, we could stratify AML patients without cytogenetic markers into three prognostic groups: Favorable group with bi-allelic CEBPAm+ or NPM1 m+/DNMT3A m- status; Poor group with DNMT3A m+ or MLLm+ cases; Intermediate group which contained all other remaining cases (Figure 4). Of note, the prognosis of the 89 patients without detectable gene mutations also corresponded to an intermediate status in terms of both OS and EFS (supplementary Figure 3). Hence, according to the gene mutational status, the 605 Group I patients could be clearly classified into distinct prognostic subgroups.

Discussion

It has long been appreciated that cytogenetic factors are independent predictors for the prognosis of AML patients. However, there are still more than half of AML patients in whom no cytogenetic markers can be found. Genetic mutations that escape
cytogenetic detection have increasingly been discovered and these mutations may serve as potential markers to extend the prognostic parameters in AML. Enormous efforts have been made to clarify the correlation between molecular changes and the clinical outcome of AML patients, allowing further dissection of AML into molecular subtypes with distinctive prognosis and therapy responses.\textsuperscript{16,37,45} Nevertheless, the clinical value of some genetic mutations remained controversial and the frequency and prognostic impact of some newly discovered mutations have not yet been well documented. A systematic investigation of genetic mutations in large series of patients should be essential in determining their clinical relevance.

In this study, we attempted to clarify the value of a cluster of molecular markers other than cytogenetic factors in stratification of AML patients into different prognostic groups. One important finding was that the recently reported \textit{DNMT3A} mutations indeed had a higher incidence in M4 and M5 subtypes (10.0\% and 18.2\%, respectively, \(P<0.001\)) in a much larger patient cohort in this study.\textsuperscript{40} The fact that \textit{DNMT3A} mutations were also identified in 2 cases of M6 with erythromonocytic leukemia provides further evidence that \textit{DNMT3A} mutations are restricted to the monocytic lineage involvement in AML. We also detected, for the first time, \textit{DNMT3A} mutations in one M2 case carrying \textit{AML1(CBF}\(\alpha\))-\textit{ETO} and one M3 individual with \textit{PML-RAR}\(\alpha\). These two cases quickly developed CNS leukemia even under intrathecal prophylaxis, reminiscent of the characteristic extramedullary involvement in monocytic leukemia. In addition, in support of the report by Ley\textsuperscript{37}, and extending the work of our own\textsuperscript{40}, patients with \textit{DNMT3A} mutations were highly related to a very
poor prognosis in AML.

Combining the mutations discovered by Ley’s study and the potential sequence variations identified in this series, an interesting situation emerged: the sequence changes occurring in the PHD domain of DNMT3A protein were of micro-deletion or nonsense mutation in nature resulting in loss-of-function of the protein, except for a homozygous G543C mutation with abnormal interaction to histone H3 as we already reported. In contrast, the sequence changes found in the catalytic domain consisted mostly of missense alterations, leading to reduction of enzymatic activity according to biochemistry assay. DNA methylation is a crucial epigenetic modification of the genome that is involved in many cellular processes, including gene expression regulation and chromatin structural remodeling. In this work, we further analyzed the aberrant DNA methylation status due to DNMT3A mutations in relationship to gene expression patterns in AML patients. These deregulated genes directly or indirectly due to abnormal DNA methyltransferase activity might further contribute to the pathogenesis of leukemia. Notably, we found that over-expression of HOXB genes could be detected only in DNMT3A mutated group whereas MEIS1 only in MLL abnormal patients, though up-regulation of HOXA7, HOXA9 and HOXA10 genes was found in both DNMT3A mutated and MLL abnormal cases. The leukemogenesis of DNMT3A mutations and MLL aberration might share some common pathways, albeit their specific characteristics. Some other important genes associated with hematopoiesis and epigenetics regulation, such as CCDC56, DCXR and TNFSF13 were also proved to be affected by DNMT3A mutation, which deserves mechanistic
study with regard to leukemogenesis. Considering a poor prognosis of patients with DNMT3A mutations, together with the biological data\textsuperscript{40}, we propose DNMT3A mutation as a “driver” one which should play an essential role in pathogenesis of leukemia involving monocytic lineage.

We found that leukemia samples from 452 out of 605 cases (74.7\%) in Group I contained at least one of the mutations. It has been generally accepted that two classes of gene mutations cooperate in AML pathogenesis. Class I mutations such as those of C-KIT, FLT3 and N-RAS are associated with activated signal transduction and provide a proliferative and survival advantage to the hematopoietic progenitors.\textsuperscript{16} However, they often show subtype-restricted distribution in AML, as evidenced by a rather specific C-KIT mutation as the second hit in the pathogenesis of CBF leukemias\textsuperscript{18} and the unique high incidence of FLT3 mutations in APL (AML-M3)\textsuperscript{14}, though N-RAS mutations displayed an even distribution across all major subtypes of AML. The class II gene mutations affecting transcription regulation and causing impaired differentiation were often overlapping with other molecular defects and co-existence patterns of some mutations were recognizable. Nevertheless, MLL mutations seemed to be mutually exclusive events, and seldom overlapping, with other mutations. Interestingly, in our preliminary morphology analysis, CEBPA, MLL and DNMT3A seemed to be correlated to distinct morphologic phenotypes within the AML-M4 subtype: patients with DNMT3A mutations and MLL variants tended to have a major monocytic lineage involvement in BM, contrarily to those with CEBPA mutations who had more blasts of granulocytic lineage. This might be explained by respective roles of
CEBPA in myeloid differentiation and of MLL as well as DNMT3A in the control of differentiation/growth of monocytic cells. Taken together, our results suggest a necessity of cooperation between distinct genetic events in leukemogenesis, although further investigation of the underlying mechanism is warranted.

It is worth pointing out that abnormalities of epigenetic regulation seems to play an essential role in the pathogenesis of AML, as evidenced by the fact that DNMT3A is a DNA methyltransferase while MLL is a histone methyltransferase. In addition, IDH1, IDH2 and TET2 mutations were found to result in a DNA hypermethylation\textsuperscript{31-33,51}, and a group of histone methyltransferase genes, such as MLL2, UTX and SETD2, have recently been found to be mutated in hematologic malignancies. Importantly, mutations of EVI1, of which the protein product is assumed to interact with DNMT3A and DNMT3B, was recently reported to associate with poor prognosis in AML\textsuperscript{1}, while mutations of EZH2, encoding a histone methyltransferase which could interact with DNMT3A, were described in lymphomas and MDS.\textsuperscript{52,53} Taking into consideration of these data, we would like to make a proposal that gene mutations involved in epigenetic regulation may be considered as a third class, apart from classes I and II mutations, in that they not only belong to a distinct regulatory network but also might share common features of aggressive disease, poor prognosis and older age onset (with the exception of MLL abnormalities).

In the present study, logistic-regression analyses showed that DNMT3A mutations represented independent unfavorable prognostic factor for remission induction with conventional daunorubicin and cytarabine–based chemotherapy; in contrast, bi-allelic
CEBPA or NPM1 mutations without DNMT3A mutations were associated with a favorable response. Importantly, age and WBC count were also proved to be independent factors of adverse outcome. However, in relatively young patient population, only MLL and TET2 abnormalities were independently related with an unfavorable induction results. In Cox regression, DNMT3A mutations and MLL rearrangements were associated with an inferior OS and EFS, whereas bi-allelic CEBPA mutations or NPM1 mutations without DNMT3A changes independently predicted favorable OS and EFS. Again, these four genes were proved to be the only independent prognostic factors in relatively young patients in predicting OS and EFS. NPM1 mutations with wild type FLT3-ITD was shown to be an important favorable genotype in AML patients without cytogenetic changes in previous reports.20-22 Our series failed to show the difference between NPM1 m+/FLT3-ITD m+ and NPM1 m+/FLT3-ITD m- groups, possibly due to the lower frequency of FLT3-ITD mutations in our series in cytogenetically normal patients (23/605, 3.8%), which might hamper further stratification of NPM1 m+ patients and decreased the sensitivity of statistics.

Finally, we have established a new stratification system to sub-classify the prognosis of AML without cytogenetic prognostic factors according to the mutations patterns of the four genes. Therefore, evaluation of molecular markers in AML, especially through detection of DNMT3A, MLL, NPM1 and CEBPA mutations, should be recommended. Further studies will be focused on using different treatment strategies according to AML genotypes. For example, enhanced therapy such as high dose anthracyclins and DNA methylation regulatory agents in elderly patients with
DNMT3A mutations might be tried to improve their clinical outcome.
Acknowledgements

This work was supported in part by the National High Tech Program for Biotechnology (863:2006AA02A405), Chinese National Key Basic Research Project (973: 2010CB529200), Mega-projects of Science Research for the 11th Five-Year Plan (2008ZX09312-026), the National Natural Science Foundation of China (30772744, 30830119, 30821063), the Shanghai Municipal Commission for Science and Technology (10411965600), the Shanghai Rising Star Program (11QA1404300), Samuel Waxman Cancer Research Foundation Co-PI Program. We are grateful to Wei-Li Zhao for assistance in clinical analysis of the AML samples, and to Shu-Min Xiong for performing the hematological morphological analysis.

Author contributions

S.-J.C, J.J. and Z.C. were the principal investigators who conceived the study. S.-J.C., J.J., Z.C. and Y.S. coordinated and oversaw the study. Y.S., Y.-M.Z., X.F., Q.-R.W. and J.-Y.S. performed most of the experiments. Z.-H.G and X.-J.Y. were responsible for bioinformatics investigation. Q.-R.W. and J.-Y.S. participated in the validation experiments. C.-L.J and H.Y. contributed in sample treating and PCR amplification. Y.S., X.F., F.-F.C. Y.-Y.W, B.C. and H.M.C. gathered detailed clinical information for the study and helped to perform clinical analysis. Y.-M.Z. and J.-Y.S. participated in the PCR assay and Sequenom analysis. Z.C., S.-J.C. and Y.S wrote the manuscript. All the authors declared no relevant financial conflict of interest was involved in this work.
References

Figure Legends

Figure 1. **DNMT3A** mutations in acute myeloid leukemia (AML).

(A) Three conserved domains in **DNMT3A** are shown: the PWWP domain that targets the enzyme to nucleic acids, the cysteine-rich PHD zinc-finger domain which interacts with unmodified histone H3 and the highly conserved catalytic domain in the C-terminal region. The mutations in AML previously reported by us are marked in black, and the newly detected mutations in this work are in red. The most common missense mutations are predicted to affect amino acid R882. A total of 37 AML had the R882H mutation, 24 with R882C, 1 with R882S and 1 with R882P in our 1,178 samples. (B) Correlation analysis of gene expression and DNA methylation. The CpG content of the promoter sequences of the genes presented is depicted as color code in a separate column (left lane), including low CpG content (LCP), intermediate CpG content (ICP), high CpG content (HCP). Hypomethylation or hypermethylation in the middle lane indicated the CpG methylated level of genes in **DNMT3A** mutated samples compared to samples without **DNMT3A** mutations. Cluster of differently expressed genes were shown in the right. Raw microarray data of gene expression and DNA methylation were available. (C) Quantitative PCR analysis of genes associated with hematopoiesis and epigenetics regulation which were up- or down-regulated and were accompanied with DNA methylation changes in microarray analysis in patients with **DNMT3A** mutations (DNMT3A), **MLL** abnormalities (MLL) or without these two types of aberrations (WT). (D) Quantitative PCR analysis of genes in distinct **HOX** families in patients with **DNMT3A** mutations (DNMT3A), **MLL** abnormalities (MLL) or without these two types of aberrations (WT).


Figure 3. Kaplan-Meier curves for overall survival (OS) and event free survival (EFS) according to genotypes with statistical significance in univariate analysis. (A, B) The median OS and EFS of patients with or without **CEBPA** mutations (**CEBPA**m+ or m-) were 21.0±5.8 months (mo) and 12.0±1.5 mo (P=0.002), 11.0±5.1 mo and 5.0±0.5 mo (P=0.004), respectively. (C, D) The median OS and EFS of patients with bi-allelic or mono-allelic **CEBPA** mutations were not reached
(NR) and 10.0±1.6 months (P<0.001), and NR and 3.0±0.7 months (P=0.001), when compared to wild type CEBPA patients. (E, F) The median OS and EFS of patients with or without DNMT3A mutations (DNMT3Am+ or m-) were 7.0±2.1 mo and 18.0±2.3 mo (P<0.001), 3.0±0.3 mo and 8.0±1.2 mo (P=0.001), respectively. (G, H) The median OS and EFS of patients with or without MLL abnormalities (MLLm+ or m-) were 8.0±2.2 mo and 17.0±2.4 mo (P<0.001), 3.0±0.2 mo and 8.0±1.4 mo (P<0.001), respectively. (I, J) The median OS and EFS of patients with or without N-RAS mutations (N-RASm+ or m-) were 10.0±4.2 mo and 17.0±2.1 mo (P=0.084), 3.0±0.3 mo and 8.0±1.3 mo (P=0.006), respectively. (K, L) The median OS and EFS of patients with NPM1 mutation but no DNMT3A mutation (NPM1m+/DNMT3Am-) were NR and 34.0 mo, while NPM1 mutation cases with DNMT3A mutations (NPM1m+/DNMT3Am+) had inferior OS (7.0±3.4 mo, P<0.001) and EFS (3.0±0.6 mo, P=0.002).

**Figure 4.** Kaplan-Meier curves for OS and EFS according to genotypes with statistical significance in multivariate analysis. All AML patients without cytogenetic prognostic markers could be divided into three prognostic groups using four marker combinations: Low risk, bi-allelic CEBPAm+ and/or NPM1 m+/DNMT3Am-; High risk, DNMT3Am+ and/or MLLm+; Intermediate, all remaining cases. Very few patients were repeatedly calculated in each group due to the concurrence of different mutations.
Figures

Figure 1
Figure 2
Figure 3

A

B

C

D

E

F

Survival (%)

Survival (%)

Survival (%)

Survival (%)

Survival (%)

Event-free survival (%)

Event-free survival (%)

Event-free survival (%)

Event-free survival (%)

Event-free survival (%)

Survival (%)

Survival (%)

Survival (%)

Survival (%)

Survival (%)

Event-free survival (%)

Event-free survival (%)

Event-free survival (%)

Event-free survival (%)

Event-free survival (%)

CEBPA+ (n=123)

CEBPA+ (n=123)

CEBPA+ (n=73)

CEBPA+ (n=73)

DNMT3A+ (n=513)

DNMT3A+ (n=513)

DNMT3A- (n=72)

DNMT3A- (n=72)

CEBPA- (n=435)

CEBPA- (n=435)

CEBPA- (n=435)

CEBPA- (n=435)

Bi-allelic CEBPA+ (n=73)

Bi-allelic CEBPA+ (n=73)

Mono-allelic CEBPA+ (n=50)

Mono-allelic CEBPA+ (n=50)

P=0.002

P=0.004

P<0.001

P=0.001

P<0.001

P=0.001

DNMT3A- (n=513)

DNMT3A- (n=513)

DNMT3A+ (n=72)

DNMT3A+ (n=72)

P=0.001

P=0.001
Figure 4

A

Low risk: median OS not reached (n=148)

Intermediate risk: median OS 17.0±2.9 mo (n=311)

High risk: median OS 7.0±1.8 mo (n=151)  \( P<0.001 \)

B

Low risk: median EFS 34.0 mo (n=148)

Intermediate risk: median EFS 8.0±1.6 mo (n=311)

High risk: median EFS 3.0±0.1 mo (n=151)  \( P<0.001 \)
Gene mutation patterns and their prognostic impact in a cohort of 1,185 patients with acute myeloid leukemia

Yang Shen, Yong-Mei Zhu, Xing Fan, Jing-Yi Shi, Qin-Rong Wang, Xiao-Jing Yan, Zhao-Hui Gu, Yan-Yan Wang, Bing Chen, Chun-Lei Jiang, Han Yan, Fei-Fei Chen, Hai-Min Chen, Zhu Chen, Jie Jin and Sai-Juan Chen