nature genetics

Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia

Xiao-Jing Yan[1,2,4], Jie Xu[1,4], Zhao-Hui Gu[3,4], Chun-Ming Pan[1,4], Gang Lu[1,4], Yang Shen[1], Jing-Yi Shi[1], Yong-Mei Zhu[1], Lin Tang[1], Xiao-Wei Zhang[1], Wen-Xue Liang[1], Jian-Qing Mi[1], Huai-Dong Song[1], Ke-Qin Li[1], Zhu Chen[1,3] & Sai-Juan Chen[1,3]

Abnormal epigenetic regulation has been implicated in oncogenesis. We report here the identification of somatic mutations by exome sequencing in acute monocytic leukemia, the M5 subtype of acute myeloid leukemia (AML-M5). We discovered mutations in *DNMT3A* (encoding DNA methyltransferase 3A) in 23 of 112 (20.5%) cases. The DNMT3A mutants showed reduced enzymatic activity or aberrant affinity to histone H3 *in vitro*. Notably, there were alterations of DNA methylation patterns and/or gene expression profiles (such as *HOXB* genes) in samples with *DNMT3A* mutations as compared with those without such changes. Leukemias with *DNMT3A* mutations constituted a group of poor prognosis with elderly disease onset and of promonocytic as well as monocytic predominance among AML-M5 individuals. Screening other leukemia subtypes showed Arg882 alterations in 13.6% of acute myelomonocytic leukemia (AML-M4) cases. Our work suggests a contribution of aberrant DNA methyltransferase activity to the pathogenesis of acute monocytic leukemia and provides a useful new biomarker for relevant cases.

Individuals with acute monocytic leukemia, or AML-M5, have a poor prognosis associated with hyperleukocytosis and extramedullary involvement[1–5]. The 3-year disease-free survival rate for individuals with AML-M5 is ~25% (ref. 6). Two morphological entities of AML-M5 are recognized: the bone marrow and peripheral blood may be overwhelmed either by monoblasts (in acute monoblastic leukemia) or by more differentiated promonocytes and monocytes (in acute promonocytic or monocytic leukemia)[1]. Chromosomal translocations involving *MLL* on 11q23 are mostly found in AML-M5 (ref. 7). Mutations in *NPM1*, *FLT3* and *NRAS* have also been reported in this disease[8–10]. However, these genetic changes occur only in a subset of AML-M5 leukemia cases.

The development of massively parallel sequencing technologies makes it feasible to catalog all classes of somatically acquired mutations in a cancer[11–13]. However, a major challenge of cancer genome analysis is to identify 'driver' mutations[12], and several recent genome studies of leukemias and solid tumors have concentrated analysis on coding regions (exomes) to increase the likelihood of identifying driver mutations[14,15]. To gain new insight into leukemogenesis and the molecular basis underlying the clinical heterogeneity of AML-M5, we carried out exome sequencing and subsequent Sanger sequencing analysis in a large series of individuals with this disease.

RESULTS

Discovery of somatic mutations through exome sequencing

We captured and sequenced the exomes from nine paired samples of AML-M5 cases (initial sequencing set; **Supplementary Table 1**) with bone marrow samples obtained at the time of diagnosis and control peripheral blood specimens obtained after complete remission. The captured target in each exome was 24 Mb. The average coverage of each base in the targeted regions was 100-fold, and 95.3% of these bases were covered sufficiently deeply for variant calling (≥10× coverage) (**Supplementary Table 2**). We used an in-house software system to identify somatic mutations by comparing variants identified in bone marrow exome data set against dbSNP and germline variants present in peripheral blood control samples (see Online Methods). We identified 266 potential somatic sequence changes including 220 single-nucleotide variations (SNVs) and 46 small insertions or deletions (indels; **Supplementary Table 3** and **Supplementary Fig. 1**).

We focused our analysis on changes predicted to affect protein-coding sequence, including 59 non-synonymous substitutions and 10 indels affecting integrity of the open reading frame (ORF) (**Supplementary Table 3**). For validation, we amplified the corresponding genomic region directly from original samples using a PCR assay and carried out Sanger sequencing. Of the non-synonymous

[1]State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [2]Department of Hematology, The First Hospital of China Medical University, Shenyang, China. [3]Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China. [4]These authors contributed equally to this work. Correspondence should be addressed to Z.C. (zchen@stn.sh.cn) or S.-J.C. (sjchen@stn.sh.cn).

Received 13 October 2010; accepted 15 February 2011; published online 13 March 2011; doi:10.1038/ng.788

Table 1 Mutations of systematically investigated genes in AML-M5 samples

Annotated gene	Mutation type	Position	Allele change	Amino acid change	Case numbers with mutation	Frequency (%)
ATP2A2	Missense	Chr12: 109,261,491	A>G	p.Lys454Arg[a]	1/112	2/112 (1.8)
	Missense	Chr12: 109,268,521	G>T	p.Val971Leu[b]	1/112	
C10orf2	Missense	Chr10: 102,738,655	A>G	p.Tyr233Cys[a]	1/112	2/112 (1.8)
	Missense	Chr10: 102,739,451	A>G	p.Gln435Arg[b]	1/112	
CCND3	Frameshift	Chr6: 42,011,723	*/+G	p.Arg271*fs**23[a]	2/112	3/112 (2.7)
	Nonsense	Chr6: 42,011,709	G>A	p.Gln276[b]	1/112	
DNMT3A	Missense	Chr2: 25,310,701	A>T	p.Val897Asp[a]	1/112	23/112[c] (20.5)
	Missense	Chr2: 25,310,747	G>A	p.Arg882Cys[a]	7/112	
	Missense	Chr2: 25,310,747	G>T	p.Arg882Ser[b]	1/112	
	Missense	Chr2: 25,310,746	C>T	p.Arg882His[a]	13/112	
	Missense	Chr2: 25,320,953	C>A[d]	p.Gly543Cys[b]	1/112	
	Missense	Chr2: 25,322,435	G>A	p.Arg478Trp[b]	1/112	
GATA2	Missense	Chr3: 129,683,410	C>T	p.Arg362Gln[a]	2/112	4/112 (3.6%)
	In-frame	Chr3: 129,685,454	*/+AGG	p.Cys319SerCys[a]	1/112	
	Missense	Chr3: 129,685,499	G>T	p.Pro304His[b]	1/112	
NSD1	Missense	Chr5: 176,627,199	C>T	p.Pro1726Leu[a]	1/112	2/112 (1.8%)
	Missense	Chr5: 176,654,869	T>A	p.Ser2632Thr[b]	1/112	

[a]Confirmed as somatic mutations. [b]Confirmed as most probable mutations. [c]One individual had both p.Arg478Trp and p.Arg882His alterations. [d]Homozygous mutation.

SNVs, 58 (98.3%) were validated as somatic mutations, whereas among indels, 8 (80.0%) were confirmed. Thus, we identified 66 somatic mutations in 63 genes, with 3 genes bearing changes in two cases (**Supplementary Table 4**). Among the alterations resulting from mutations, we identified the p.Gly12Ser and p.Gly12Asp alterations in NRAS and the p.Asp835Val mutation in FLT3 (**Supplementary Table 4**). We also detected a *MLL-MLLT4* fusion gene, which was confirmed at RNA level. Detection of these known mutations in AML-M5 confirmed our approach and suggested we could identify unknown genetic defects.

Defining mutation frequencies in AML-M5 samples

We next carried out exome sequencing in five additional AML-M5 cases without matched normal samples (expanded sequencing set; **Supplementary Table 1**). We focused on the 63 genes found to have somatic mutations in our initial sequencing set. To exclude rare SNPs for any sequence variations detected, we analyzed 509 control samples from unrelated healthy individuals (control validation set). Then, all the sequence changes detected in the 63 genes were genotyped in bone marrow DNA samples from 98 additional cases of newly diagnosed (94 cases) or relapsed (4 cases) AML-M5 (M5 validation set) by Sequenom analysis. As a result, we found somatic mutations (confirmed by sequencing of paired peripheral blood control samples at complete remission in some cases in our M5 validation set) or very probable somatic mutations (defined as those not identified in 509 samples in the control validation set) in 14 genes, each detected in at least two cases, among a total of 112 AML-M5 cases (**Supplementary Table 5**).

Integration of the above data, together with literature searches for genes whose structure and/or expression are altered in cancer and other human diseases, led to the selection of six genes (*DNMT3A*, *NSD1*, *GATA2*, *CCND3*, *ATP2A2* and *C10orf2*) for sequencing of their whole coding regions in our M5 validation set (**Table 1**). Because only a limited number of paired peripheral blood samples were available, all sequence changes discovered in unpaired leukemia samples were genotyped in the control validation set to rule out the possibility of rare SNPs.

In *DNMT3A*, encoding a member of the DNA methyltransferase 3 family[16], we identified three heterozygous variants in the same codon for Arg882, which led to three distinct substitutions, p.Arg882His, p.Arg882Cys and p.Arg882Ser, detected, in 13, 7 and 1 of 112 samples, respectively (**Fig. 1a**). We also detected a heterozygous p.Val897Asp variant and a homozygous p.Gly543Cys variant in two samples using exome sequencing, and one heterozygous p.Arg478Trp variant in a sample containing p.Arg882His. Among these sequence changes, p.Arg882His, p.Arg882Cys and p.Val897Asp were confirmed as resulting from somatic mutations, whereas p.Arg478Trp, p.Gly543Cys

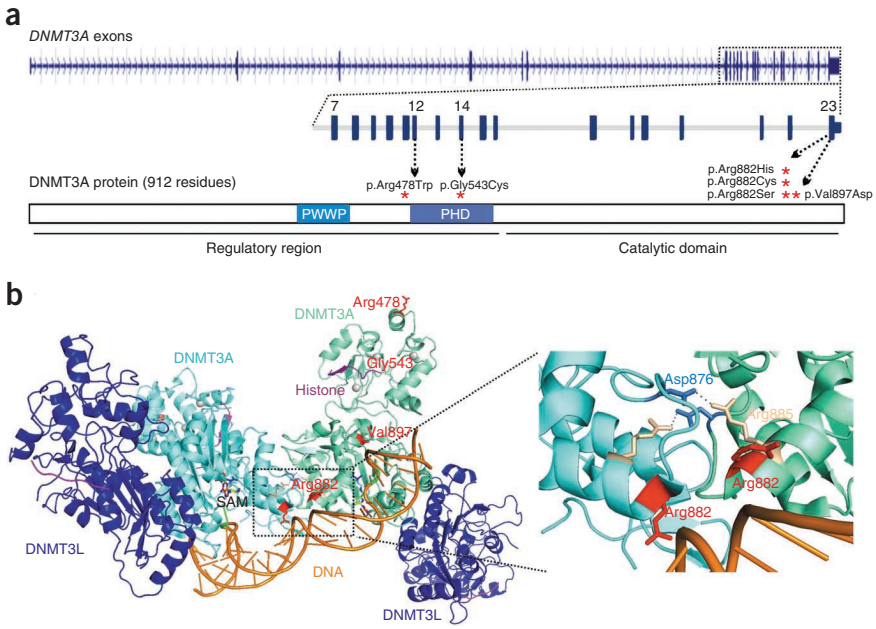

Figure 1 Locations of DNMT3A mutations and structure of DNMT3A protein. (**a**) Genomic organization of *DNMT3A* locus, alternative exons and protein domain structure. Locations of the mutations affecting Arg478 in exon 12, Gly543 in exon 14, Arg882 and Val897 in exon 23 of the *DNMT3A* gene (top) and protein (bottom) are indicated with arrows and red asterisks, respectively. (**b**) Structural prediction of DNMT3A alterations. The structure of DNMT3A dimer is shown in cyan, the structure of two DNMT3L molecules bound to both sides of the DNMT3A dimer (3A-3A) is shown in blue, and the DNA double helix is shown in orange. Purple ribbons, histone H3 N-peptide. Rainbow ribbons, SAM cofactor. Red stick residues, mutations in AML-M5 leukemia. Mutation residues are involved in 3A-3A dimerization or DNA binding (Arg882), SAM cofactor binding (Val897), protein-protein interaction (Arg478), and histone H3 peptide binding (Gly543). Arg882 is near the 3A-3A interface with two pairs of salt bridges formed between Arg885 and Asp876 (counterparts in mouse protein, Arg881 and Asp872, respectively) and very close to the DNA double helix (enlarged).

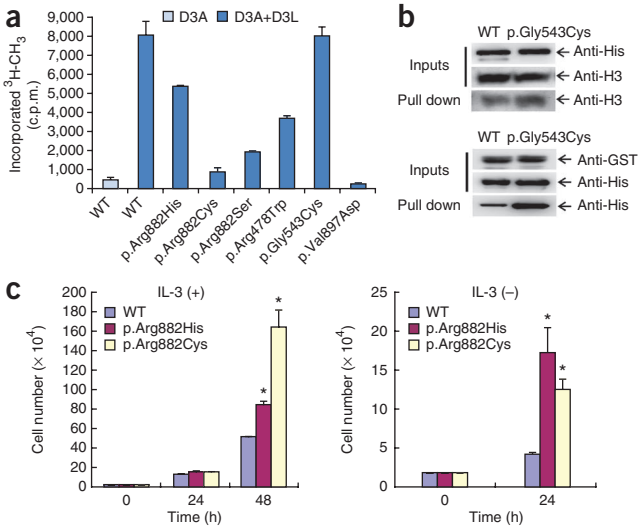

Figure 2 Functional analysis of DNMT3A mutants. (**a**) *In vitro* methyltransferase activity of DNMT3A mutants in AML-M5. DNMT3A, D3A; DNMT3L, D3L. Data for incorporation in counts per minute (c.p.m.) are mean ± standard deviation (s.d.) for three independent experiments. (**b**) Interaction of the p.Gly543Cys mutant with histone H3. Histone H3 in core histone was pulled down by His-labeled DNMT3A (top). His-labeled DNMT3A was pulled down by GST-labeled N-terminal histone H3 (bottom). WT, wild type. (**c**) Cell proliferation assay. 32D cells overexpressing WT, p.Arg882His or p.Arg882Cys DNMT3A grew in 12-well plates with (left) or without (right) IL-3. Cell numbers were counted at indicated time points, beginning at 24 h after transfection. $^*P \leq 0.05$. Results are mean ± s.d. from three independent experiments.

and p.Arg882Ser were not present in the control validation set and are thus very probably the result of somatic mutations (**Fig. 1a** and **Supplementary Fig. 2**). The frequency of *DNMT3A* mutations in our entire AML-M5 series was 23 out of 112 (20.5%) (**Table 1**). In an effort to explore rare SNPs in *DNMT3A*, particularly the non-synonymous ones, we also sequenced the whole coding regions of 31 samples from healthy individuals but detected no such variations.

We also found sequence variations in *NSD1*, *GATA2*, *CCND3*, *ATP2A2* and *C10orf2*. We discovered two heterozygous mutations, resulting in p.Pro1726Leu and p.Ser2632Thr, in *NSD1*, a gene encoding a histone H3K36 methyltransferase[17]. Notably, we identified three variants in *GATA2*, mutations of which have been reported in a subset of acute myelomonocytic blast crisis of chronic myeloid leukemia[18]. These sequence abnormalities included two missense point mutations and one in-frame insertion of three nucleotides in two zinc finger motifs (ZF1 and ZF2; **Supplementary Fig. 3**). We found two types of deleterious mutations in the cell cycle regulator gene *CCND3* (ref. 19); we identified a one-base (G) insertion causing frameshift in two cases and a nonsense mutation of codon 276 in another case. For *ATP2A2* and *C10orf2*, we found two distinct missense mutations of each gene in two cases.

To clarify molecular abnormalities in AML-M5, we also checked in our series for known gene mutations present at relatively high frequency. In total, we identified *MLL* abnormalities including *MLL* translocations or *MLL* partial tandem duplication (*MLL*-PTD), *NPM1* mutations, *NRAS* mutations and *FLT3* abnormalities including internal tandem duplication (ITD) and Asp835 point mutations, respectively, in 19.6%, 25.9%, 10.7% and 18.8% of the 112 cases (**Supplementary Table 6**). Notably, among 23 cases with various *DNMT3A* mutations and 22 cases with *MLL* variations, only one had both *MLL*-PTD and a *DNMT3A* p.Arg882His mutant and one had both a *MLL-MLLT3* and a *DNMT3A* p.Val897Asp mutant. *DNMT3A* mutations were present with *NPM1* mutations in the same sample in 16 cases, whereas mutations of *NPM1* did not overlap with *MLL* abnormalities (**Supplementary Fig. 4** and **Supplementary Table 7**).

Analysis of biological activities of DNMT3A mutants

Because of its relatively high mutation rate in AML-M5, we focused our attention on the impact of *DNMT3A* mutation. All of the *DNMT3A* mutations we identified were missense mutations and occurred at amino acid residues well conserved through evolution (**Supplementary Fig. 5**). Moreover, the mutant allele was transcriptionally active in all 23 samples that we examined by RT-PCR, which was designed to eliminate genomic DNA contamination (**Supplementary Fig. 2**). Twenty-two samples also showed mRNA expression of the normal allele, whereas in one case with homozygous mutation only the p.Gly543Cys mutant allele was expressed.

We further addressed the possible structure-function relationship of the mutants as the crystal structure of DNMT3A was available[20,21]. DNMT3A protein has three conserved domains (**Fig. 1a**): two in the N-terminal regulatory region (a PWWP domain, which targets the enzyme to nucleic acids, and a cysteine-rich PHD zinc-finger domain, which interacts with unmodified histone H3) and one highly conserved catalytic domain in the C-terminal region[16,22]. By sequence alignment of DNMT3A and DNMT3B we found that they showed high identity, and Arg882 in DNMT3A

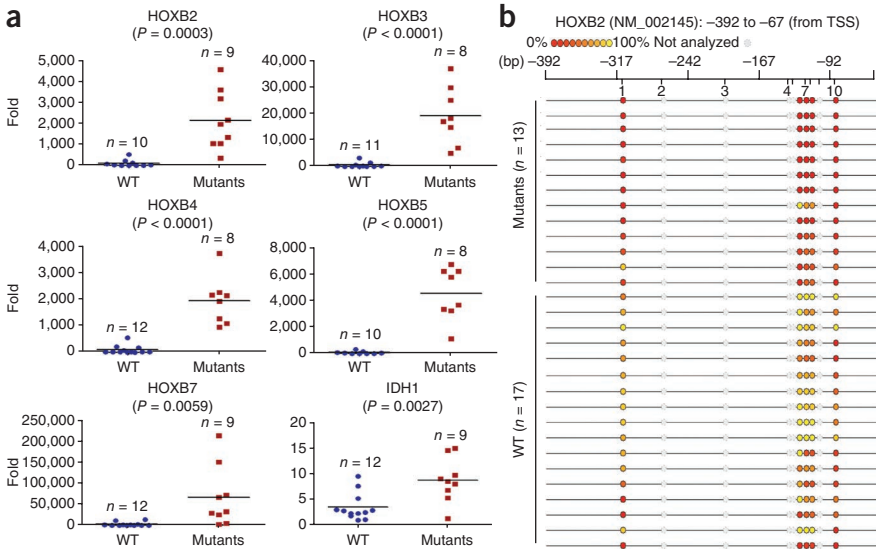

Figure 3 Analysis of gene expression and DNA methylation in individuals with AML-M5. (**a**) Gene expression. Real-time PCR of mRNA levels of *HOXB2–HOXB5*, *HOXB7* and *IDH1* in individuals without (WT) and with *DNMT3A* mutations (mutants). Results are the average of three independent experiments. (**b**) CpG islands adjacent to *HOXB2* were hypomethylated in samples from individuals with *DNMT3A* mutations. Each horizontal line represents methylation status of one individual sample detected by Sequenom EpiTYPER DNA methylation analysis.

Table 2 Clinical features of AML-M5 cases with *DNMT3A* and *MLL* mutations

Variable	WT	Mutations[a]	
		DNMT3A	*MLL*
Mean age at diagnosis (years)	44.8 ± 18.9	54.9 ± 14.5	37.0 ± 17.4
	$P = 0.022$		
		$P = 0.001$	
Gender			
Male (%)	43 (62.3)	13 (56.5)	14 (63.6)
Female (%)	26 (37.7)	10 (43.5)	8 (36.4)
	$P = 0.631$		
		$P = 0.763$	
WBC count			
≥30 × 10^9/L: no. (%)	22 (32.4)	16 (69.6)	13 (59.1)
<30 × 10^9/L: no. (%)	46 (67.6)	7 (30.4)	9 (40.9)
	$P = 0.003$		
		$P = 0.542$	
Monoblast percentage in nonerythroid lineage			
≥80%: no. (%)	29 (42.6)	5 (21.7)	13 (59.1)
<80%: no. (%)	39 (57.4)	18 (78.3)	9 (40.9)
	$P = 0.086$		
		$P = 0.016$	

WT, wild type (individuals without *DNMT3A* or *MLL* mutations). WBC, white blood cells. *MLL* mutations include *MLL* translocation and *MLL*-PTD. A one-way ANOVA test was used to compare the age at the diagnosis in different groups; other variables were analyzed by Fisher's exact test. The *P* value was obtained by comparing wild-type and *DNMT3A* mutations groups or *DNMT3A* and *MLL* abnormalities groups.
[a]Two individuals had both *DNMT3A* and *MLL* mutations and were grouped into two cohorts with mutations.

probably corresponds to Arg823 in DNMT3B (**Supplementary Fig. 5**), which is mutated in immunodeficiency, centromeric instability, facial anomalies (ICF) syndrome[23]. We tried to assess the effect of various mutations on the structure and function of DNMT3A by generating a computational model of the complex of DNMT3A, histone H3 and DNA. According to structural analysis, Arg882 of DNMT3A might participate in the homodimerization of DNMT3A or its interaction with DNA (**Fig. 1b**), whereas Arg478 is located on an α-helix of the PHD domain surface which probably takes part in the protein-protein interaction, and Gly543 in the PHD domain is very close to the histone H3K4 binding surface (**Fig. 1b**). Val897 in the catalytic region is highly conserved among different species (**Supplementary Fig. 5**), and should have an important role in keeping the enzyme in the right conformation for interacting with *S*-adenosylmethionine (SAM) cofactor (**Fig. 1b**).

To explore the functional consequences of structural alteration, we purified wild-type and mutant DNMT3A proteins including the PHD and catalytic domains from *Escherichia coli*. Because the enzymatic activity of DNMT3A-DNMT3L complex is very high and the mutant residues are not on the binding surface with DNMT3L, we used the co-purified DNMT3A-DNMT3L complex to carry out DNA methyltransferase activity assay *in vitro*. The interaction between DNMT3A and DNMT3L was not hampered by the mutations of DNMT3A (**Supplementary Fig. 6**). The enzymatic assays showed that, compared with wild-type DNMT3A, all Arg882 mutants consistently showed a marked reduction in their DNA methylation activities (**Fig. 2a**). In agreement with the structural prediction, p.Arg478Trp and p.Val897Asp mutants also had substantially reduced enzyme activities *in vitro* (**Fig. 2a**). However, the p.Gly543Cys mutant had no influence on DNA methylation activity, probably because the mutation resided in the functional domain interacting with histone but outside the catalytic region of DNMT3A. Indeed, the p.Gly543Cys mutant showed increased ability to interact with histone H3 *in vitro* (**Fig. 2b**). These biochemical results suggest that *DNMT3A* mutations observed in our sequencing study could confer abnormal function of DNA methylation or histone binding *in vivo*. To further investigate the potential role of *DNMT3A* mutations in leukemogenesis, we carried out experiments in 32D cells, an interleukin-3 (IL-3)-dependent mouse myeloid cell line. Overexpression of p.Arg882His and p.Arg882Cys mutants promoted the proliferation of 32D cells even without IL-3 as compared with wild-type DNMT3A (**Fig. 2c**).

Aberrant DNA methyltransferase activity could change DNA methylation and alter gene expression. Thus, we determined the gene expression profile and DNA methylation patterns in six AML-M5 samples with *DNMT3A* Arg882 mutations and four AML-M5 samples with wild-type *DNMT3A*. We observed differences of expression level in 889 of 20,723 (4.3%) annotated genes by using an Affymetrix microarray (**Supplementary Fig. 7a**). Notably, we found that several *HOX* family genes and the *IDH1* (isocitrate dehydrogenase 1), which are associated with acute myeloid leukemia pathogenesis[24,25], were upregulated in individuals with the *DNMT3A* mutant (**Supplementary Fig. 7b**). In validation analysis using RT-PCR among 21 cases, we found that the expression of

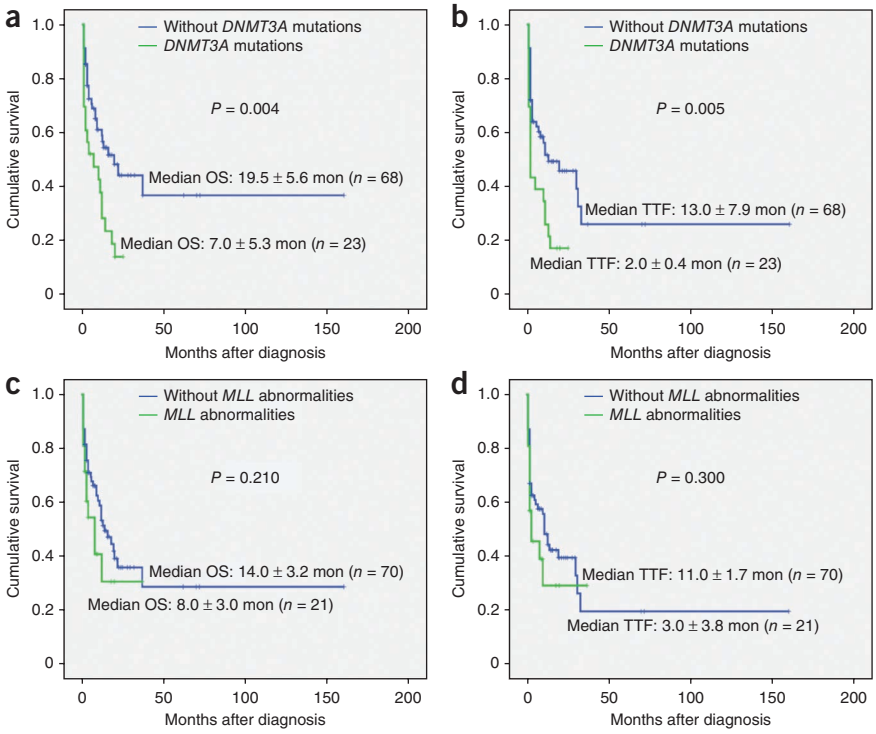

Figure 4 Kaplan-Meier analysis of the survival of individuals with AML-M5. (**a**,**b**) Overall survival (OS) and TTF of individuals with or without *DNMT3A* mutations. Individuals with *DNMT3A* mutations had a poorer overall survival or TTF than those without the mutations. (**c**,**d**) Overall survival and TTF of individuals with or without *MLL* mutations. *MLL* mutations did not influence overall survival or TTF of individuals with AML-M5. Mon, months.

these genes significantly increased in cases with *DNMT3A* mutations as compared with those with wild-type *DNMT3A* (**Fig. 3a**). Moreover, by using methyl-DNA immunoprecipitation (MeDIP)-chip analysis, we found that a total of 3,878 genomic regions had significantly different methylation patterns between the two groups; this might affect gene expression of various pathways (**Supplementary Tables 8** and **9** and **Supplementary Fig. 8**). We validated the presence of hypomethylated CpG islands in the *HOXB* cluster including a region adjacent to *HOXB2* by Sequenom assay (**Fig. 3b** and **Supplementary Fig. 9**).

Table 3 *DNMT3A* mutations in acute leukemia

Subtype of acute leukemia samples	No. samples analyzed	No. samples with *DNMT3A* mutations					
		R882S	R882C	R882H	Others[a]	Total (%)	95% CI
AML-M1	6	0	0	0	0	0 (0.0)	0.0–4.8
AML-M2	77	0	0	0	0	0 (0.0)	0.0–3.8
AML-M3	94	0	0	0	0	0 (0.0)	0.0–3.1
AML-M4	66	0	4	5	0	9 (13.6)	6.8–23.2
AML-M5	112	1	7	13	3	23 (20.5)[b]	13.7–28.6
ALL	19	0	0	0	0	0 (0.0)	0.0–13.9

AML, acute myeloid leukemia. ALL, acute lymphocytic leukemia.
[a]Others include the p.Arg478Trp, p.Gly543Cys and p.Val897Asp alterations.[b]One individual had both p.Arg478Trp and p.Arg882His alterations.

Clinical relevance of *DNMT3A* versus *MLL* abnormalities

Because *DNMT3A* and *MLL* are two epigenetic regulation genes commonly mutated in AML-M5 and mutations of these two genes rarely overlapped in cases, we divided our AML-M5 cases into three groups: cases with *DNMT3A* mutations, those with *MLL* abnormalities and those without mutations of the two genes. We found that both the *DNMT3A* mutation group and the *MLL* rearrangement or PTD group were associated with hyperleukocytosis (**Table 2**). However, these two groups differed significantly in terms of bone marrow morphology and mean age of disease onset: 59.1% of cases with *MLL* abnormalities had ≥80% monoblasts in bone marrow, whereas only 21.7% of cases with *DNMT3A* mutations showed a high percentage of monoblasts, and the bone marrow of most *DNMT3A* mutation cases was infiltrated with promonocytes and monocytes (**Table 2**). Moreover, the mean age of disease onset was 54.9 years in the group with *DNMT3A* mutations, but was 37.0 years in the group with *MLL* abnormalities (**Table 2**).

Notably, we found that individuals with *DNMT3A* mutations had a much worse prognosis as compared with individuals without mutations of this gene, as reflected by the overall survival curve (**Fig. 4a**) and the time to treatment failure (TTF) (**Fig. 4b**). The group with *MLL* abnormalities had a slightly but not significantly inferior prognosis (**Fig. 4c**,**d**). In contrast, we observed no differences in overall survival between cases with or without *NPM1*, *FLT3* or *NRAS* mutations (**Supplementary Fig. 10**). With regard to age at disease onset, *DNMT3A* mutations were associated with a poor prognosis in relatively young cases, though their prognostic value in elderly cases deserves further analysis in a larger number of individuals (**Supplementary Fig. 11**). Furthermore, multivariate analysis showed that *DNMT3A* mutations and white blood cell count were the two independent prognostic variables associated with overall survival and TTF (**Supplementary Table 10**). Binary logistic regression also indicated that *DNMT3A* mutations were associated with lower complete remission rate (**Supplementary Table 10**). These data suggested that *DNMT3A* mutations represent a poor prognostic factor and are probably involved in the pathogenesis of acute promonocytic or monocytic leukemia in elderly individuals, whereas *MLL* abnormalities might be associated with acute monoblastic leukemia in relatively young individuals.

DNMT3A mutations in acute myelomonocytic leukemia

To determine whether the mutational status of *DNMT3A* shows lineage specificity or selectivity, we also tested for *DNMT3A* mutations in other subtypes of leukemia (**Table 3**). No mutations were detected in acute myeloid leukemia subtypes M1 through M3 or in acute lymphocytic leukemia. Notably, for *DNMT3A*, we identified Arg882 mutations in 9 of 66 cases (13.6%) of AML-M4 (acute myelomonocytic leukemia). Analysis of bone marrow cytological features showed that the bone marrow of these nine cases was characterized by a major monocytic cell infiltration, though a smaller contingent of myeloblasts was distinguishable, whereas among 31 cases without Arg882 mutations eligible for bone marrow cytological analysis, 14 cases had a major component of myeloblasts. This finding indicates that *DNMT3A* mutations are relatively specific to acute leukemia with monocytic features and represent driver mutations with an important role in the pathogenesis of acute myeloid leukemia with monocytic lineage involvement.

DISCUSSION

Acute myeloid leukemia subtypes are clonal hematopoietic diseases caused by somatic alterations of genomic information[26]. Genome analysis on these distinct subtypes allows not only a better understanding of leukemogenesis but also the identification of new biomarkers and/or drug targets, as recently reported in individuals with AML-M1[11,27].

Here we identified recurrent mutations of *DNMT3A* in about one-fifth of AML-M5 cases. Among the six different mutations in our case series, those of Arg882 were the most frequent, in agreement with a recent report describing the *DNMT3A* mutations in a large series of acute myeloid leukemia cases[28]. In fact, the p.Arg882His substitution has also been reported recently in an array-based sequencing study of $CD34^+$ blasts of acute myeloid leukemia and myeloproliferative disorder[29]. No non-synonymous *DNMT3A* SNPs were found in 192 healthy individuals[30], similar to the observation we made in this study. Indeed, although we identified three *DNMT3A* variations in cases in our study without paired control samples, they were not identified in 509 normal samples, and sequencing coding regions of 31 samples from healthy persons showed no SNPs causing amino acid substitutions, suggesting that the three variations probably represent somatic mutations.

DNA methylation is a crucial epigenetic modification of the genome that is involved in regulating many cellular processes, including gene expression regulation and chromatin structural remodeling[16]. A link between DNA methylation and cancer has been shown[31]. DNMT3A and DNMT3B are paralogous enzymes responsible for *de novo* DNA cytosine methylation during development[32]. Mutations in *DNMT3B* cause ICF syndrome[23,33]. Notably, the alterations of Arg882 in DNMT3A are analogous to the p.Arg823Gly mutant in DNMT3B, which was discovered in ICF syndrome and has reduced methylation activity[23]. In this study, we found that the enzymatic activity of the Arg882 mutant was significantly reduced, possibly because of interrupted homodimerization, as Arg882 is adjacent to Val873, Asp872 and Arg881, which are located at the homodimer surface[20], or because of inappropriate interaction of DNMT3A with DNA according to the structural model. We also found a reduced enzyme activity in p.Arg478Trp and p.Val897Asp

mutants, probably because of distinct but related biochemical mechanisms. Although the homozygous mutant p.Gly543Cys had similar enzyme activity to wild-type DNMT3A, its increased affinity for histone H3 could lead to sequestration of the enzyme and subsequently disturbed *de novo* DNA methylation in relevant chromatin domains. Notably, these potential functional alterations of DNMT3A are consistent with the three-dimensional structure of the protein, which suggests a putative negative impact on the structure-function integrity of the proteins. In 22 of 23 cases, both normal and aberrant alleles of *DNMT3A* were expressed, suggesting possible dominant-negative effects of the mutants against the wild-type proteins in these heterozygous cases. In addition, the expression of *DNMT3A* is low in AML-M4 and AML-M5 compared with other subtypes of acute myeloid leukemia[34]. Thus, even in individuals without mutation of *DNMT3A*, decreased expression of this gene due to an unknown abnormality in transcriptional regulation could contribute to disease mechanism. Notably, we found that *DNMT3A* mutations enabled 32D hemopoietic cells to acquire growth and survival advantage even without growth factor, although the leukemogenic potential of these mutations remains to be proven at an organism level.

We also carried out experiments to explore the possible epigenetic consequences of *DNMT3A* mutations. Indeed, gene expression profiles and DNA methylation patterns showed some differences between patients with and without *DNMT3A* mutations. Notably, we found that members of *HOXB* genes were extensively upregulated at mRNA level in individuals with mutations in *DNMT3A*, and we detected hypomethylation of certain CpG islands in the *HOXB* locus. HOX family proteins have an important role in regulation of normal hematopoiesis[35]. Substantial evidence also indicates that abnormal HOX protein expression is functionally important in the pathogenesis of acute myeloid leukemia and other cancers[24,35]. In leukemia, dysregulated *HOX* gene expression can occur because of chromosomal translocation involving upstream regulators such as *MLL*[24]. Taking these results into consideration, we assume that *HOX* family members might be target genes of both *DNMT3A* and *MLL*, and dysregulated *HOX* genes because abnormalities of these two driver genes might contribute to the pathogenesis of AML-M5. DNMT3A can facilitate gene transcription by methylating nonpromoter regions including intergenic regions and gene bodies[36]. Therefore, the potential mechanism of aberrant DNA methylation due to DNMT3A alterations needs further investigation because of the complex function of this enzyme. Furthermore, we found increased expression of *IDH1* in individuals with *DNMT3A* mutations, consistent with a gain of function of this enzyme caused by *IDH1* mutations in malignances including acute myeloid leukemia[25,37].

The fact that abnormalities of *DNMT3A* are common and are restricted to acute myeloid leukemia with major monocytic lineage involvement in both M5 and a subset of M4 suggests two concepts. One is that the *DNMT3A* mutations may be pathogenic rather than random events. The other is that these leukemia types may more likely develop as a result of aberrant epigenetic regulation. We also found mutations in another epigenetic regulatory gene, *NSD1*, a member of the histone methyltransferase family in AML-M5, although with a relatively low rate. These *NSD1* mutations were different from those causing Sotos syndrome, a rare genetic disorder. Notably, the *NSD1-NUP98* fusion gene derived from t(5;11)(q35;p15.5) translocation has been found in *de novo* childhood acute myeloid leukemia[38]. This indicates that abnormal NSD1 is probably involved in the pathogenesis of AML-M5. The finding of mutations of the hematopoietic transcription factor *GATA2* is also notable because its p.Leu359Val mutation in ZF2 and deletion from Ala341 to Gly346 in ZF1 have been reported by our group to be involved in acute myelomonocytic blastic crisis of chronic myeloid leukemia and are molecularly involved in the aberrant signaling of PU.1, a transcription factor essential for the differentiation of monocytic lineage[18].

With the discovery of new somatic mutations, the pattern of genomic abnormalities in AML-M5 is notable: on one hand, *DNMT3A* mutations should represent altered regulation of DNA methylation; on the other hand, *MLL* abnormalities, and probably also *NSD1* mutations, could give rise to aberrant status of histone methylation with emergence of newly acquired transcriptional activities. Recently, mutations of histone methyltransferase genes, such as *EZH2* in some hematological malignancies and *MLL2* in Kabuki syndrome (an inherited disorder), have been reported[15,39,40]. These studies indicate that mutations in the epigenetic regulatory pathway have an important role in the pathogenesis of human disorders, including malignant diseases. The identification of genes whose expressions are specifically modulated by *DNMT3A* inactivation and the mechanism by which mutations in *DNMT3A* enable the pathogenesis of monocytic leukemia may be the next crucial steps in this line of research.

Correlation between *NPM1* and/or *FLT3* mutations with prognosis has been controversial[10]. In our AML-M5 series, mutations of *NPM1*, *FLT3* and *NRAS* were not statistically confirmed to be correlated with disease prognosis. Notably, by using *DNMT3A* mutations and *MLL* abnormalities as biomarkers, we identified case groups with different disease prognoses. These two types of variations probably have different characteristics in that individuals bearing *DNMT3A* mutations differed from those with *MLL* abnormalities in terms of median ages at disease onset and of differentiation stages of leukemic components in bone marrow. The *DNMT3A* mutant group was also associated with very poor outcome as compared to individuals without *DNMT3A* mutations. Although these findings need to be further studied in a larger sample size, the information obtained in this study based on analysis of clinical relevance could help improve the clinical management of AML-M5. To this end, the newly identified and previously established major genomic abnormalities together can be found in about three of five (60.7%) individuals with AML-M5, whereas genetic defects in the remaining cases remain to be identified.

URLs. EpiDesigner, http://www.epidesigner.com/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. NCBI Sequence Read Archive: exome sequencing data, SRP005624. NCBI Gene Expression Omnibus: SNP and expression microarray data, GSE27244.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was supported in part by the China National High Tech Program for Biotechnology (863:2006AA02A405), Chinese National Key Basic Research Project (973: 2010CB529203), the National Natural Science Foundation of China (30772744, 30821063, 81000213) and the Shanghai Municipal Commission for Science and Technology (07DZ05908,08DZ2200100). We thank Jiangsu Institute of Hematology of the First Affiliated Hospital of Soochow University, Department of Hematology of the First Affiliated Hospital of Zhejiang University, Department of Hematology of the Tongji Hospital of Tongji University and Department of Hematology of Luwan division of Ruijin Hospital for providing AML-M5 samples. We are grateful to W.-L. Zhao, F. Yang, Y.-Y. Wang and X. Fan for assistance in clinical analysis of the AML-M5 samples, and to B. Chen and S.-M. Xiong for carrying out the hematological morphological analysis.

AUTHOR CONTRIBUTIONS

S.-J.C. and Z.C. were the principal investigators who conceived the study. S.-J.C., Z.C. and X.-J.Y. coordinated and oversaw the study. X.-J.Y. and J.X. carried out most of the experiments. Z.-H.G. and G.L. were responsible for bioinformatics investigation. C.-M.P. and H.-D.S. carried out the exome sequencing and participated in the validation experiments. J.-Q.M. and L.T. participated in the preparation of biological samples. Y.S. helped gather detailed clinical information for the study and helped to carry out clinical analysis. Y.-M.Z. and J.-Y.S. participated in the PCR assay and Sequenom analysis. X.-W.Z. and W.-X.L. helped to carry out the biochemical experiments. K.-Q.L. carried out the structural analysis and guided the biochemical experiments. Z.C., S.-J.C. and X.-J.Y. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Bennett, J.M. *et al.* Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br. J. Haematol.* **33**, 451–458 (1976).
2. Fenaux, P. *et al.* Acute monocytic leukaemia in adults: treatment and prognosis in 99 cases. *Br. J. Haematol.* **75**, 41–48 (1990).
3. Bennett, J.M. *et al.* Long-term survival in acute myeloid leukemia: the Eastern Cooperative Oncology Group experience. *Cancer* **80**, 2205–2209 (1997).
4. Porcu, P. *et al.* Hyperleukocytic leukemias and leukostasis: a review of pathophysiology, clinical presentation and management. *Leuk. Lymphoma* **39**, 1–18 (2000).
5. Peterson, L., Dehner, L.P. & Brunning, R.D. Extramedullary masses as presenting features of acute monoblastic leukemia. *Am. J. Clin. Pathol.* **75**, 140–148 (1981).
6. Tallman, M.S. *et al.* Acute monocytic leukemia (French-American-British classification M5) does not have a worse prognosis than other subtypes of acute myeloid leukemia: a report from the Eastern Cooperative Oncology Group. *J. Clin. Oncol.* **22**, 1276–1286 (2004).
7. Schoch, C. *et al.* AML with 11q23/MLL abnormalities as defined by the WHO classification: incidence, partner chromosomes, FAB subtype, age distribution, and prognostic impact in an unselected series of 1897 cytogenetically analyzed AML cases. *Blood* **102**, 2395–2402 (2003).
8. Koh, Y. *et al.* Different clinical importance of FLT3 internal tandem duplications in AML according to FAB classification: possible existence of distinct leukemogenesis involving monocyte differentiation pathway. *Ann. Hematol.* **88**, 1089–1097 (2009).
9. Boissel, N. *et al.* Prevalence, clinical profile, and prognosis of NPM mutations in AML with normal karyotype. *Blood* **106**, 3618–3620 (2005).
10. Schlenk, R.F. *et al.* Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. *N. Engl. J. Med.* **358**, 1909–1918 (2008).
11. Ley, T.J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
12. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. *Nature* **458**, 719–724 (2009).
13. Pleasance, E.D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
14. Dalgliesh, G.L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360–363 (2010).
15. Morin, R.D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* **42**, 181–185 (2010).
16. Bestor, T.H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9**, 2395–2402 (2000).
17. Lucio-Eterovic, A. K. *et al.* Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc. Natl. Acad. Sci. USA* **107**, 16952–16957 (2010).
18. Zhang, S.J. *et al.* Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. *Proc. Natl. Acad. Sci. USA* **105**, 2076–2081 (2008).
19. Weng, A.P. & Aster, J.C. No T without D3: a critical role for cyclin D3 in normal and malignant precursor T cells. *Cancer Cell* **4**, 417–418 (2003).
20. Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. *Nature* **449**, 248–251 (2007).
21. Ooi, S.K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
22. Xie, S. *et al.* Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene* **236**, 87–95 (1999).
23. Xie, Z.H. *et al.* Mutations in DNA methyltransferase DNMT3B in ICF syndrome affect its regulation by DNMT3L. *Hum. Mol. Genet.* **15**, 1375–1385 (2006).
24. Shah, N. & Sukumar, S. The Hox genes and their roles in oncogenesis. *Nat. Rev. Cancer* **10**, 361–371 (2010).
25. Gross, S. *et al.* Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J. Exp. Med.* **207**, 339–344 (2010).
26. Gilliland, D.G. & Tallman, M.S. Focus on acute leukemias. *Cancer Cell* **1**, 417–420 (2002).
27. Mardis, E.R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
28. Ley, T.J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
29. Yamashita, Y. *et al.* Array-based genomic resequencing of human leukemia. *Oncogene* **29**, 3723–3731 (2010).
30. El-Maarri, O. *et al.* A systematic search for DNA methyltransferase polymorphisms reveals a rare DNMT3L variant associated with subtelomeric hypomethylation. *Hum. Mol. Genet.* **18**, 1755–1768 (2009).
31. Jones, P.A. DNA methylation and cancer. *Oncogene* **21**, 5358–5360 (2002).
32. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**, 247–257 (1999).
33. Xu, G.L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191 (1999).
34. Bullinger, L. *et al.* Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N. Engl. J. Med.* **350**, 1605–1616 (2004).
35. Eklund, E.A. The role of HOX genes in malignant myeloid disease. *Curr. Opin. Hematol.* **14**, 85–89 (2007).
36. Wu, H. *et al.* Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444–448 (2010).
37. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
38. Panarello, C., Rosanda, C. & Morerio, C. Cryptic translocation t(5;11)(q35;p15.5) with involvement of the NSD1 and NUP98 genes without 5q deletion in childhood acute myeloid leukemia. *Genes Chromosom. Cancer* **35**, 277–281 (2002).
39. Ng, S.B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793 (2010).
40. Nikoloski, G. *et al.* Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat. Genet.* **42**, 665–667 (2010).

ONLINE METHODS

DNA sample preparation. Genomic DNA samples for exome sequencing were freshly obtained from mononuclear cells isolated by Ficoll gradient centrifugation from bone marrow samples at primary diagnosis and matching peripheral blood during complete remission after standard chemotherapy. DNA or RNA samples for validation were obtained from our frozen leukemia bank.

Targeted sequence capture. Genomic DNA was captured on a NimbleGen 2.1M human exome array according to the manufacturer's protocols (Roche/NimbleGen). We aimed to capture most of the exome part from the DNA sample with NimbleGen the 2.1M chip, which contains 24 Mb CCDS (~85% of US National Center for Biotechnology Information CCDS Database) region cross ~17,000 genes in 34 Mb targeted nucleotides. DNA was sheared by sonication and adaptors were ligated to the resulting fragments. The adaptor-ligated templates were fractionated by agarose gel electrophoresis and fragments of the desired size were excised. Extracted DNA was hybridized to the capture array at 42.0 °C using the manufacturer's buffer. The array was washed twice at 47.5 °C and three more times at room temperature (20–25 °C) using the manufacturer's buffers. Bound genomic DNA was eluted using 125 mM NaOH for 10 min at room temperature. The resulting fragments were amplified by ligation-mediated PCR, purified and subjected to DNA sequencing on the Illumina platform.

Massively parallel sequencing. The workflow of Cluster generation using the Illumina cluster station was as follows: template hybridization, isothermal amplification, linearization, blocking, denaturation and sequencing primer hybridization. Then, deep sequencing was carried out for the captured libraries with the Illumina Genome Analyzer IIx platform (GAIIx), and 75–120 bp paired- and single-end reads were output according to the manufacturer's protocols because the software and hardware were upgraded in the process. Image analysis and base calling were carried out by Illumina RTA versions 1.5 and 1.6 with default parameters.

Alignment, SNV or indel calling and quality control. The software BWA[41] was used to align both single- and paired-end reads to the reference human genome (hg18, downloaded from http://genome.ucsc.edu/) with default parameters. After the alignment, only those uniquely mapped single reads or confidently mapped (Phred Quality ≥10) paired end reads were kept. Variations including SNVs and indels were called with the Samtools software package[42] and filtered with recommended threshold (SNV quality ≥20, indel quality ≥50 and ≥3 reads covered) for cases. To ensure the filter power and minimize the false discovery rate, loose criteria were applied to filter control variations (SNV quality ≥10, indel quality ≥10 and ≥3 reads covered). After the variation calling, SNVs with ≥20% mutation rate, that is, ≥20% of the reads mapped to the SNV point support the variation, were kept as qualified point mutations (15% was set for indels because of their low sensitivity).

Genome-wide SNP genotyping and sensitivity evaluation. All the nine bone marrow samples in initial sequencing sets were genotyped with Illumina human 610 or 660W SNP array and analyzed using GenomeStudio V2010.2 with GT module 1.6.3 (Illumina). With the SNP information, we evaluated the heterozygous SNVs located in the targeted region after the first round of quality control and a high sensitivity (99.2% on average) was observed. With the guidance of the pattern concluded from these positive heterozygous SNVs, which were the dominant form of somatic mutants[27,43], we tested different sets of parameters and finally exerted a more strict but still sensitive (97.0%) criterion to filter those false positive SNVs.

Targeted exon resequencing. Genomic DNA from samples in the mutation discovery sets (the initial and expanded sequencing sets) were subjected to targeted exon, amplified by PCR and resequenced by means of Sanger sequencing. Damaging prediction of mutations was carried out by PolyPhen 2 as described[44]. To determine the mutations of six highlighted genes (*DNMT3A*, *NSD1*, *GATA2*, *CCND3*, *ATP2A2* and *C10orf2*) in AML-M5, we carried out whole-exon sequencing in the M5 validation set. Sequenom assay was carried out to determine the frequency of mutations in AML-M5 and other subtypes of acute leukemia. Identification of *MLL* abnormalities and *FLT3*-ITD were subjected to targeted exon amplification and Sanger sequencing.

Computational modeling of DNMT3A wild-type and mutants. The available protein crystal structures of DNMT3L, DNMT3A, HhaI and DNMT3A-DNMT3L complex with histone H3 peptide or DNA (PDB 2PVC, 2QRV, 3A1B and 1MHT, respectively) were identified as the template[20,21,45]. A three-dimensional model of the histone H3 peptide, DNMT3L, DNMT3A and DNA complex was generated using CCP4 Molecular Graphics Superimposition and molecular docking program AutoDock 4.0 (ref. 46). The structural images were drawn using PyMOL software (DeLano Scientific).

Protein purification and DNA methylation assay *in vitro*. DNMT3A mutant constructs were generated using site-directed mutagenesis with His-tagged DNMT3A plasmid constructed from commercial human DNMT3A gene template purchased from ATCC (catalog no. 10436367). Wild-type DNMT3A and each of the six mutant constructs were co-purified with or without wild-type GST-tagged DNMT3L. These proteins were used to carry out the DNA methyltransferase activity assay *in vitro*.

For the detection of *in vitro* methylation activity, a biotinylated 1,179-bp PCR fragment amplified from the EBNA1 region of p220.2 was used as substrate. The methylation reaction contained DNA fragments (100 ng) and DNMT3A (100 ng) with or without DNMT3L (100 ng) in 20 mM HEPES, pH 7.5, 30 mM NaCl, 1 mM EDTA, 0.2 mM DTT, 50 mg/ml BSA and 1.25 mM *S*-[methyl-^{3}H]AdoMet (80 Ci/mmol) (Amersham Biosciences) in a total volume of 20 µl. Incorporation of methyl-^{3}H into the substrate DNA was determined by liquid scintillation.

Pull-down assay. GST-histone H3 protein and wild-type or p.Gly543Cys mutant His-DNMT3A were purified from *E. coli* BL21 (DE3) carrying an expression construct. For GST pull-down assay, 200 µg of purified GST fusion protein was incubated with GST-Sepharose 4B beads (Amersham Biosciences) in 1× PBS buffer at 4 °C for 30 min, and then the beads were washed five times with 1× PBS buffer. Purified DNMT3A products (100 µg) were then added to the slurry and incubated at 4 °C for another 30 min. The Sepharose beads were washed five times with 1× PBS buffer, and bound proteins were resolved on an SDS-PAGE gel followed by protein blotting.

For the Ni-NTA pull-down assay, 200 µg of purified His-fusion DNMT3A proteins (wild-type and p.Gly543Cys mutant) were incubated with Ni-NTA beads (Amersham Biosciences) in binding buffer (500 mM NaCl, 20 mM Tris, pH 8.5, 2 mM DTT) at 4 °C for 30 min, and then the beads were washed five times with binding buffer. Core histones (50 µg; Roche) were then added to the slurry and incubated at 4 °C for another 30 min. The Sepharose beads were washed five times with binding buffer, and bound proteins were resolved on an SDS-PAGE gel followed by western blotting.

Cell proliferation analysis. 32D cells were transfected using Cell Line Nucleofector Kit V according to the manufacturer's instructions. At 24 h after transfection, the cells with overexpression of GFP-tagged wild-type DNMT3A and mutant DNMT3A were analyzed by flow cytometry assay. The average transfection efficiency was ~50%. Transfected cells (2×10^4) were then incubated with or without IL-3 in a 12-well plate for cell proliferation analysis.

Microarray expression profiling. Total RNA from bone marrow samples of AML-M5 cases were prepared and, after quality control, subjected to Affymetrix Human Genome U133 Plus 2.0 Array GeneChip microarrays according to the manufacturer's instructions (Affymetrix). A series of software (Expression Console software, Affymetrix; Partek GS 6.5) was used for data analysis. Probes with fold change of >2 and significant difference ($P \leq 0.05$) between the DNMT3A mutation group and the wild-type DNMT3A group were retained for further analysis.

Methylation analysis. DNA samples were extracted for the HG18 Methylation 2.1M Deluxe Promoter Array (NimbleGen) to identify the methylated DNA regions. Methyl-DNA immunoprecipitation (MeDIP) was carried out using Biomag magnetic beads coupled to mouse monoclonal antibody to 5-methylcytidine. The immunoprecipitated DNA was eluted and purified by phenol-chloroform extraction and ethanol precipitation. The total input and immunoprecipitated DNA were labeled with Cy3- and Cy5-labeled random nonamers, respectively, and hybridized to the assay chips. Scanning was carried out with

doi:10.1038/ng.788

the Axon GenePix 4000B microarray scanner. The raw data were extracted as pair files by NimbleScan software and further processed through median centering, quantile normalization and linear smoothing by Bioconductor packages Ringo, limma and MEDME. After normalization, a normalized $\log_2$-ratio data were created for each sample. From the normalized $\log_2$-ratio data, a sliding-window peak-finding algorithm provided by NimbleScan v2.5 (Roche-NimbleGen) was applied to find the enriched peaks with specified parameters (sliding window width, 750 bp; minimal probes per peak, 2; *P* value (after $-\log_{10}$ transformation) minimum cutoff, 2; maximum spacing between nearby probes within peak, 500 bp). Then, the identified peaks were mapped to genomic features including transcripts and CpG islands. Genes with significant changes were revealed with DEP analysis using M′ method. From this analysis, 2,246 peaks (1,946 unique genes) were detected as hypermethylated in *DNMT3A* mutation sample as compared to samples with wild-type *DNMT3A*, whereas 1,602 peaks (1,291 unique genes) were hypomethylated.

Bisulfite conversion followed by MALDI-TOF mass spectrometry. Bisulfite conversion of 1 μg genomic DNA using EZ DNA methylation-Gold Kit (Zymo) were carried out according to manufacturer's instructions. Hot-start PCR was used to amplify the region of interest from the bisulfite-converted genomic DNA. Bisulfite conversion followed by Sequenom MassARRAY MALDI-TOF mass spectrometry–based quantitative DNA methylation analysis was carried out in triplicate by standard protocol (Sequenom EpiTYPER DNA methylation analysis). Primers were designed using Sequenom's EpiDesigner tool (see URLs). Primer sequences (without linker sequences) used in Sequenom analysis are listed in **Supplementary Table 11**. A T7-promoter tag was incorporated into each reverse PCR Page 3 amplification primer, and a decamer linker sequence was added to each forward primer to balance the primer length (**Supplementary Table 11**).

41. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
42. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
43. Mullighan, C.G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
44. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
45. O'Gara, M., Klimasauskas, S., Roberts, R.J. & Cheng, X. Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaL methyltransferase-DNA-AdoHcy complexes. *J. Mol. Biol.* **261**, 634–645 (1996).
46. Garrett, M. *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662 (1998).