Front. Med. 2022, 16(4): 627–636
https://doi.org/10.1007/s11684-020-0815-4

RESEARCH ARTICLE

Distinct gene expression pattern of *RUNX1* mutations coordinated by target repression and promoter hypermethylation in acute myeloid leukemia

Jingming Li[1,*], Wen Jin[1,2,*], Yun Tan[1], Beichen Wang[1], Xiaoling Wang[1], Ming Zhao[1], Kankan Wang (✉)[1,2]

[1]*Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China;* [2]*CNRS-LIA Hematology and Cancer, Sino-French Research Center for Life Sciences and Genomics, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China*

Abstract **Runt-related transcription factor 1 (RUNX1) is an essential regulator of normal hematopoiesis. Its dysfunction, caused by either fusions or mutations, is frequently reported in acute myeloid leukemia (AML). However, *RUNX1* mutations have been largely under-explored compared with *RUNX1* fusions mainly due to their elusive genetic characteristics. Here, based on 1741 patients with AML, we report a unique expression pattern associated with *RUNX1* mutations in AML. This expression pattern was coordinated by target repression and promoter hypermethylation. We first reanalyzed a joint AML cohort that consisted of three public cohorts and found that *RUNX1* mutations were mainly distributed in the Runt domain and almost mutually exclusive with *NPM1* mutations. Then, based on RNA-seq data from The Cancer Genome Atlas AML cohort, we developed a 300-gene signature that significantly distinguished the patients with *RUNX1* mutations from those with other AML subtypes. Furthermore, we explored the mechanisms underlying this signature from the transcriptional and epigenetic levels. Using chromatin immunoprecipitation sequencing data, we found that RUNX1 target genes tended to be repressed in patients with *RUNX1* mutations. Through the integration of DNA methylation array data, we illustrated that hypermethylation on the promoter regions of RUNX1-regulated genes also contributed to dysregulation in *RUNX1*-mutated AML. This study revealed the distinct gene expression pattern of *RUNX1* mutations and the underlying mechanisms in AML development.**

Keywords *RUNX1*; gene mutation; acute myeloid leukemia; transcriptional repression; DNA methylation

Introduction

Hematopoietic transcription factors (TFs) are essential regulators of hematopoiesis, and their genetic alterations, including mutations and chromosomal translocations, cause the differentiation blockage observed in many types of leukemia. These TFs exert their regulatory activities via downstream targets controlled in coordination by epigenetic regulations, including DNA methylation [1]. Given their essential roles in controlling hematopoiesis and their genetic abnormalities in the development of leukemia, elucidating the molecular classification of hematopoietic TF-associated leukemia with the integration of the ensemble of genome-wide targets and DNA methylation status is crucial.

Runt-related TF 1 (*RUNX1*, also named *AML1*, *CBFA2*, and *PEBP2aB*) encodes a TF that plays pivotal roles in normal hematopoiesis, including definitive hematopoietic stem cell formation and megakaryocyte maturation and granulocytic differentiation [2,3]. It can form a heterodimeric complex with the core-binding factor β subunit (CBFB) to exert its transactivation function on target genes [4]. *RUNX1* abnormalities include fusions caused by chromosomal translocations and acquired mutations, among which point mutations in *RUNX1* are frequently detected in hematologic malignancies, such as myelodysplastic syndromes, particularly acute myeloid leukemia

Received February 14, 2020; accepted July 8, 2020

Correspondence: Kankan Wang, kankanwang@shsmu.edu.cn

*Equal contribution.

(AML). The intensive investigations of the genome-wide binding maps of RUNX1-associated fusion proteins, such as RUNX1–RUNX1T1 generated by t(8;21), ETV6–RUNX1 by t(12;21), and CBFB–MYH11 by inv(16), have considerably improved our understanding of the molecular signatures and mechanisms of RUNX1 fusion-mediated leukemogenesis. In contrast to favorable prognosis-associated RUNX1 fusions, *RUNX1* mutations usually portend poor outcomes and chemotherapy resistance in patients with AML [5,6], suggesting that the molecular mechanisms underlying *RUNX1* mutations are distinct from those underlying fusion forms. Genetic and clinical analyses show that *RUNX1* mutations rarely coexist with *NPM1* and *CEBPA* mutations [7]. Given their independent prognosis, *RUNX1* mutations have been adopted as a new but provisional entity in the latest (2016) revision of the World Health Organization (WHO) AML classification to stratify patients for risk-adapted treatments [8]. However, the molecular signatures associated with *RUNX1* mutations are limited, and few common signatures have been identified to depict the mechanisms of *RUNX1* mutations in AML.

In this study, we identified a 300-gene signature for *RUNX1*-mutated AML that was highly associated with genes directly repressed by RUNX1 and genes with the hypermethylated status. We first investigated the spectrum of *RUNX1* mutations in a large combined AML cohort. Next, we performed differential gene expression analysis and developed a gene signature that could distinguish *RUNX1*-mutated AML from wild-type cases. Moreover, we integrated chromatin immunoprecipitation sequencing (ChIP-seq) data to evaluate the effects of *RUNX1* mutations on RUNX1 target genes and analyzed the association of DNA hypermethylation with *RUNX1* mutations. We collectively demonstrated that *RUNX1*-mutated AML possessed a distinct gene expression pattern coordinated by target repression and promoter hypermethylation. Our study provides insights into the pathogenic mechanisms caused by *RUNX1* mutations and suggests promising targets for prognostic and therapeutic approaches.

Materials and methods

Data resources

The data of patients with AML with mutation information and clinical features were collected from three public cohorts containing 1741 qualified AML cases: TCGA (LAML project, 144 cases), EGA (European Genome-phenome Archive, EGAS00001000275, 1328 cases), and GEO (Gene Expression Omnibus, GSE23312, 269 cases). For the AML cohort from TCGA, RNA-seq, gene mutation, clinical, and DNA methylation data were downloaded from the GDC Data Portal website. For the AML cohort from EGA, gene mutation and clinical data were downloaded from the supplementary information of Papaemmanuil's work [9] and website of Gerstung's work [10]. The cohort from EGA contained *de novo* AML cases and secondary or therapy-related AML cases given that the original source (n = 1540) of the EGA cohort contained 8.3% (n = 128) secondary or therapy-related AML cases [9].

In the analysis of *RUNX1* mutation distribution, we used all the 1741 cases from the three cohorts (TCGA, GSE23312, and EGAS00001000275). In the analysis of gene expression, we used 413 *de novo* AML cases with gene expression data from the TCGA AML cohort and GEO cohort. In the analysis of DNA methylation, we used 113 *de novo* AML cases with gene expression and DNA methylation data from the TCGA AML cohort. In the analysis of prognostic impact, we used 135 *de novo* non-APL cases with gene expression data and detailed survival information from TCGA, in which the treatment protocols had a low impact on prognosis (Tables S2–S4).

Data integration and differential gene expression analysis

For the TCGA AML data, the RNA-seq files of 151 cases (Level 3, HTSeq counts) were downloaded. The whole-exon sequencing files of 149 cases (Level 3, somatic mutations) were downloaded. Mutation calls were retained if detected by at least two of the following software packages: MuTect, MuSE, SomaticSniper, and VarScan. Our analyses were based on the gene expression and mutation data produced using the updated reference genome (GRCh38/hg38). For the cohort from GEO (GSE23312), the microarray data were filtered out to remove batch effects and failed samples (with $> 10\%$ failed probes) before hierarchical clustering.

Differentially expressed genes (DEGs) were identified by using RNA-seq read count data and the R package DESeq2 (version 1.22.1) [11]. We identified DEGs with a false discovery rate of less than 0.05 and absolute LFC ($\log_2$ fold change) of more than 0.5849 (FC > 1.5 or $< 2/3$).

Sample clustering

Hierarchical clustering based on the gene expression data of 413 patients (144 from TCGA and 269 from GSE23312) was conducted by using the pheatmap package (version 1.0.12). The variance-stabilizing transformation provided by DESeq2 was used to normalize RNA-seq count data for calculating sample distances. The complete linkage method and Euclidean distance were used to measure the distances. Principal components analysis (PCA) was conducted with the plotPCA function in DESeq2.

ChIP-seq analysis

The target genes of RUNX1 were retrieved from the published ChIP-seq data of three AML cell lines, U937 (GSM1632640 and GSM1632641), HL60 (GSM2871154 and GSM2871145), and THP-1 (GSM2108052), and reanalyzed as follows. Raw sequencing data were downloaded from the NCBI Sequence Read Archive. Short read data were aligned to the UCSC human reference genome hg38 by using Bowtie2 (version 2.3.4.3) [12]. PCR duplicates were removed by using SAMtools (version 1.9) [13]. RUNX1 binding peaks were called by utilizing MACS2 (version 2.1.2) [14] at a cut-off of q = 0.05 (data with input) or P = 10e–6 (data without input). High-confidence binding peaks were defined as overlapping regions in at least two cell lines. UCSC RefGene data were applied to extract gene promoter regions (–2 kb/+0 kb to transcription start sites (TSS) for genes shorter than 500 bp and –2 kb/+0.5 kb to TSS otherwise). BEDTools [15], which assigned binding peaks to RUNX1 target genes, were used to overlap high-confidence peak regions with gene promoter regions. The Cistrome Data Browser and the UCSC genome browser were used for visualization [16,17].

DNA methylation analysis

We reanalyzed the Illumina Infinium 450k DNA Methylation Array data of 113 patients from the TCGA AML cohort from scratch to update methylation profiles on the basis of the reference genome hg38. The raw IDAT files (Level 1) for each sample were downloaded from the online supplementary material of TCGA's work [18]. The probe annotation data (released on September 9, 2018) of hg38 were from Zhou *et al.*'s paper [19]. The R package minfi (version 1.28.3) [20] was used to adjust the background and normalize the red/green dye bias across samples to produce the beta values of each probe. Poorly performing probes with detection $P < 0.05$ and probes on the sex chromosomes were filtered out first. On the basis of the probe annotation data, the probes were further removed if they had SNPs close to the 3′ end, multiple mappable locations, or partial overlaps with repeated elements in the bisulfite-converted genome. We finally obtained 407 071 high-quality probes on the array. On the basis of the beta values of the probes, we performed moderated *t*-test with the limma package (version 3.38.3) [21] to evaluate probe-wise differences. The probe coordinates in the genome were then used to generate CpG clusters and detect differentially methylated regions (DMRs) with the bumphunter package (version 1.24.5) [22]. We set the parameter maxGap to 500 bp and the parameter B to 4000 and defined family-wise error rate < 0.1 to be statistically significant. The gene promoter regions were extracted as mentioned in the ChIP-seq analysis.

Bioinformatics analyses

Gene Ontology (GO) enrichment analysis was performed by using the clusterProfiler package [23]. We conducted gene set enrichment analysis (GSEA) in accordance with the user guide and ranked all the genes by gene expression fold change (*RUNX1*-mutated vs. *RUNX1* wild-type). Survival analysis was conducted by using the packages survival and survminer. Prognostic impacts were evaluated with the Kaplan–Meier method and the log-rank test. Statistical analyses were performed with the environment R version 3.5.1.

Results

RUNX1 mutations were mainly in the Runt domain and exclusive with *NPM1* mutations

To investigate the characteristics of *RUNX1* mutations in AML, we carried out a meta-analysis that included 1741 patients with AML collected from three AML cohorts (detailed in the section of "Materials and methods"). First, we found that *RUNX1* mutations were mutually exclusive with frequently occurring fusions in core-binding factor AML (CBF-AML), including RUNX1–RUNX1T1 generated by t(8;21) and CBFB–MYH11 generated by inv(16) (Fig. 1A). Given that RUNX1–RUNXT1 and CBFB–MYH11, which are both driver fusions that dominantly inhibit CBF functions, are considered as independent factors promoting the development of AML [24], the data indicated that *RUNX1* mutations might independently promote leukemogenesis. Second, the detailed analysis of the distribution of *RUNX1* mutations showed that 125 patients with *RUNX1* mutations carried 135 *RUNX1* mutations, including 57 frameshifts (42.2%) and 18 nonsense variants (13.3%). Most of the mutations (n = 102, 75.6%) were located in the highly conserved Runt domain (Fig. 1B) that binds to the Runt-binding DNA element [25]. Additionally, we calculated the frequency of mutated loci and found that the top recurrent mutations included p.R166Q, p.D198G, p.R201*, p.R204*, and p.R204Q (5 times), and p.R162k (4 times). Interestingly, *RUNX1* mutations were far less frequently observed in the Runx1 inhibition domain than in the Runt domain (Fig. 1B). Third, we compared the concurrent mutations with mutant *RUNX1* vs. wild-type RUNX1.We used samples with cytogenetically normal AML (CN-AML) to exclude the interference of chromosomal abnormalities. As shown in Fig. 1C, the following genes were more frequently mutated among patients with AML and *RUNX1* mutations than those with the wild-type counterpart: *SRSF2* ($P < 0.001$, the Fisher's exact test), *ASXL1* ($P < 0.001$), *BCOR* (P = 0.001), and *SF3B1* (P = 0.038). By contrast, significantly fewer *NPM1* ($P < 0.001$), *CEBPA* (P =

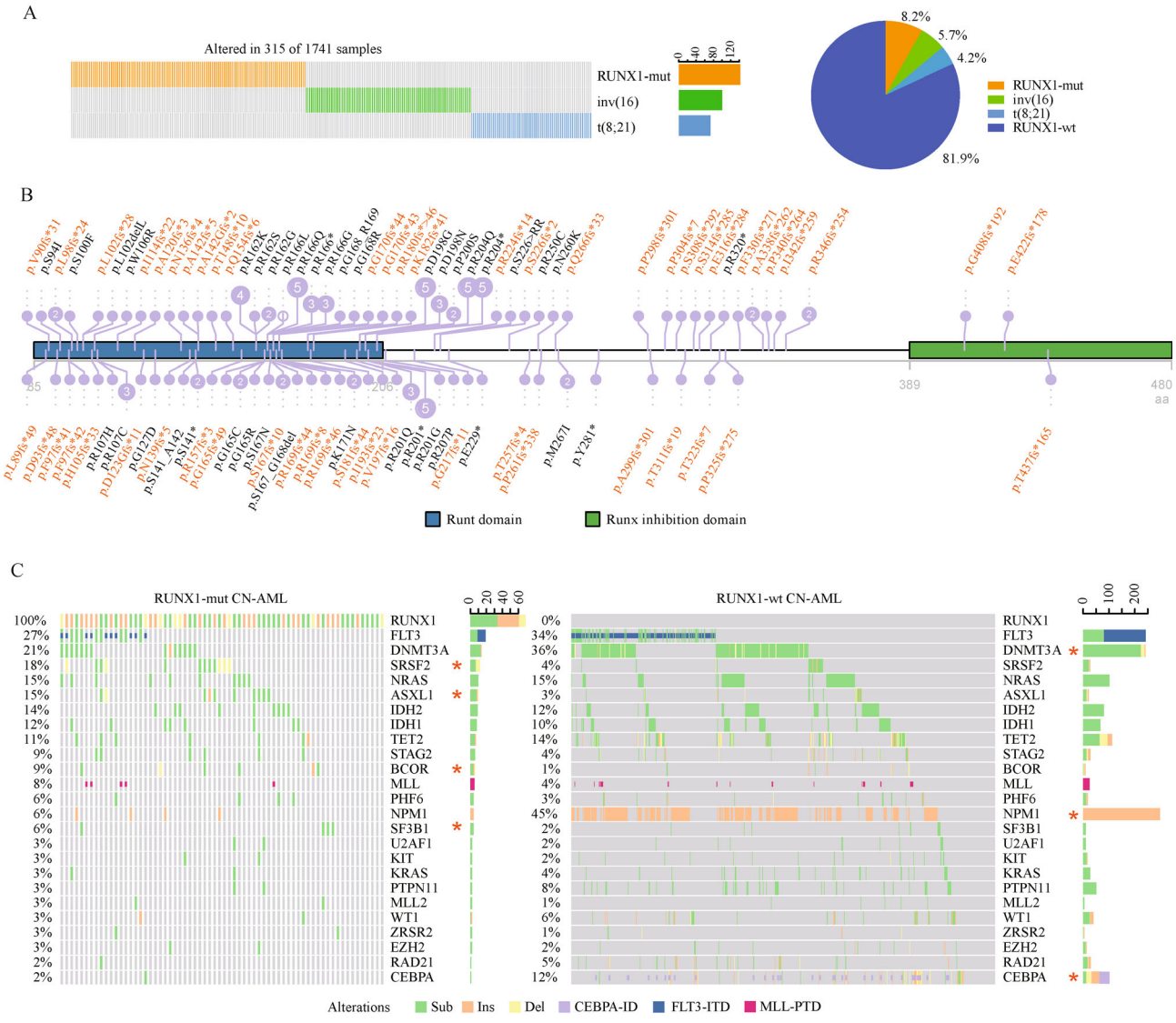

Fig. 1 *RUNX1* mutations are mainly located in the Runt domain and exclusive with *NPM1* mutations. (A) RUNX1 mutations were exclusive with t(8;21) and inv(16) AML in 1741 patients with AML collected from three published cohorts (GEO, GSE23312; EGA, EGAS00001000275; and TCGA AML). mut, mutated; wt, wild-type. (B) Distribution of mutations in the different domains of the RUNX1 protein (NP_001745.2). Frameshift mutations are highlighted in orange. (C) Distribution of mutations in *RUNX1* and 25 additional cancer genes in patients with CN-AML. Concurrent mutations are shown for patients with *RUNX1* mutations ($n = 66$, left panel) or wild-type *RUNX1* ($n = 666$, right panel). The starred genes are those with higher mutation frequencies compared to the other group. Sub, substitution; Ins, insertion; Del, deletion; ITD, internal tandem duplication; PTD, partial tandem duplication.

0.006), and *DNMT3A* ($P = 0.02$) mutations were observed in *RUNX1*-mutated cases than in other cases.

Patients with AML and *RUNX1* mutations show a distinct gene signature

Considering that *RUNX1* mutations showed marked distribution characteristics in AML cohorts, we hypothesized that *RUNX1*-mutated AML might exhibit common gene expression features. Given that *NPM1* mutations were frequently observed (49%) in patients with wild-type *RUNX1* but rarely observed in patients with *RUNX1* mutations (Fig. 1C), DEGs might result from either *NPM1* mutations or *RUNX1* mutations. Therefore, we removed *NPM1*-mutated samples to compare patients with *RUNX1* mutations and wild-type *RUNX1*. DEG analysis (FDR < 0.01, FC > 1.5 or $< 2/3$) showed that 476 genes were significantly upregulated in *RUNX1*-mutated AML, whereas 531 were significantly downregulated (Fig. 2A). On the basis of these DEGs, we further developed a gene signature for *RUNX1* mutations to characterize the most noticeable differences between patients with *RUNX1*

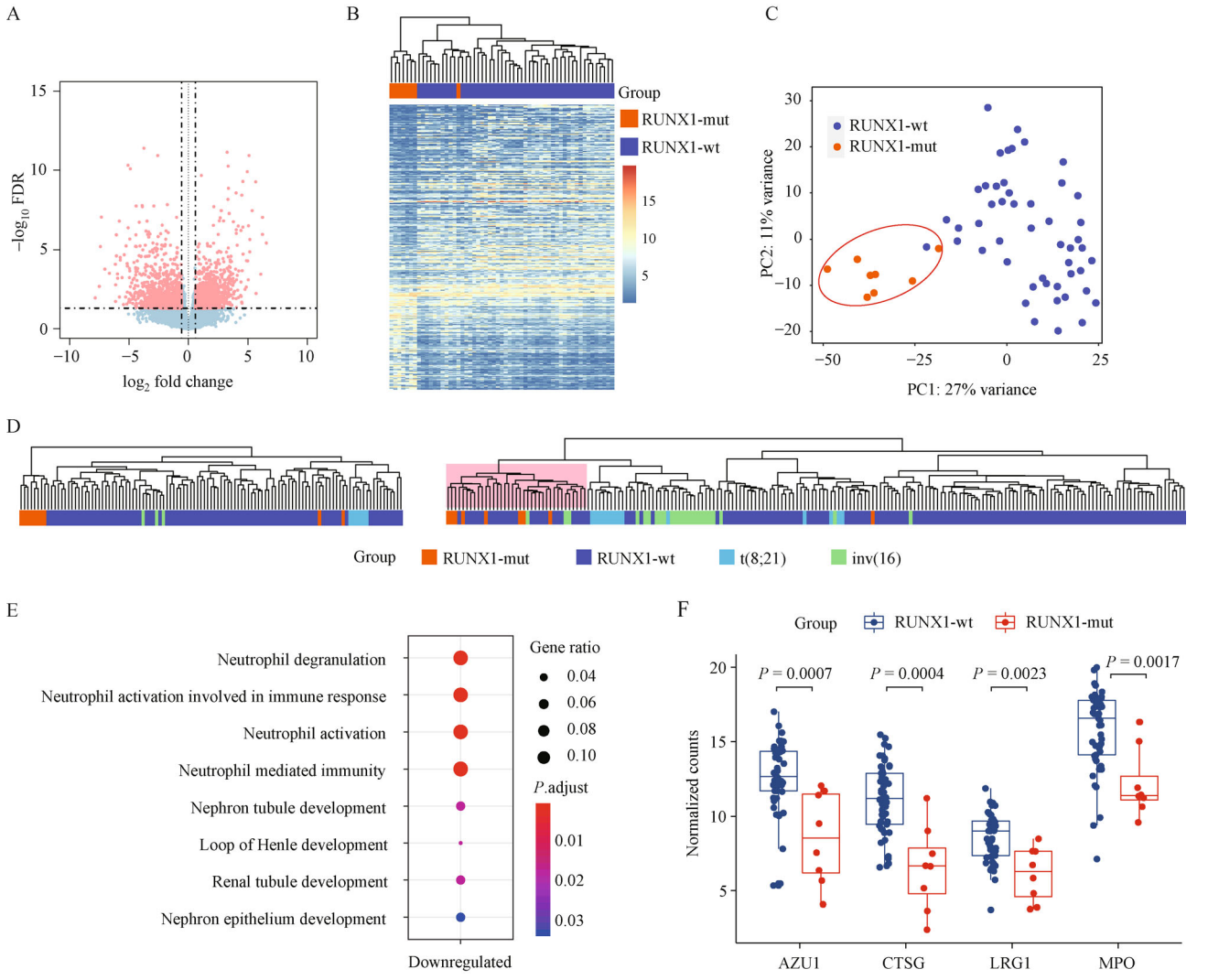

Fig. 2 Patients with AML and *RUNX1* mutations have a distinct gene signature. (A) Volcano plot representing the DEGs between the patients with *RUNX1* mutations and wild-type *RUNX1*. (B) *RUNX1*-mutated samples were distinguished in the TCGA *NPM1* wild-type CN-AML cohort. Heat map representing the expression (normalized counts) of the 300-gene signature. mut, mutated; wt, wild-type. (C) PCA results based on the 300-gene signature demonstrated the clear separation of *RUNX1*-mutated samples from wild-type *RUNX1* samples. (D) Hierarchical clustering based on the 300-gene signature separated *RUNX1*-mutated samples and t(8;21) samples from others in the TCGA AML cohort (left panel, $n = 144$) and the GEO cohort GSE23312 (right panel, $n = 269$). The *RUNX1*-mutation-associated cluster is highlighted in pink. (E) GO analysis revealed that neutrophil-associated pathways were downregulated in *RUNX1*-mutated AML. (F) Four genes in the neutrophil-associated pathways were repressed in *RUNX1*-mutated samples of the TCGA AML cohort. *P* values were calculated by using the Wilcoxon signed-rank test.

mutations and wild-type *RUNX1*. The clustering performance of several gene signature settings (100, 200, 250, 300, and 400 of the most variable DEGs) was compared (Fig. S1). The 300-gene signature was chosen for the minimum number of genes that yielded the best classification in hierarchical clustering: *RUNX1*-mutated samples clustered as a distinct class, except only one *RUNX1*-mutated sample (Fig. 2B; Table S1). We performed PCA to test the capability of the 300-gene signature to identify *RUNX1*-mutated samples and found that the *RUNX1*-mutated samples separated from the other samples in the CN-AML cohort (Fig. 2C). Furthermore, we used this gene signature for hierarchical clustering in the whole TCGA AML cohort and the GEO AML cohort (GSE23312). As shown in Fig. 2D, the majority of *RUNX1*-mutated samples clustered as one group, confirming the performance of this signature. Interestingly, the t(8:21) AML samples also formed a separate cluster (Fig. 2D), demonstrating the extended utility of this signature. Thus, the gene signature consisting of the top 300 DEGs represented the distinct gene expression pattern of *RUNX1*-mutated AML.

We performed GO enrichment analysis to explore the

biological function associated with this signature. The results obtained for 199 downregulated genes showed the enrichment of GO terms related to myeloid functions, including neutrophil degranulation, neutrophil activation, and neutrophil-mediated immunity (Fig. 2E). Specially, the expression of four vital genes (*AZU1*, *CTSG*, *LRG1*, and *MPO*) in these neutrophil-associated GO-terms were significantly repressed in *RUNX1*-mutated cases (Fig. 2F), suggesting that *RUNX1* mutations might dysregulate neutrophil functions in myeloid cells to contribute to the development of AML. These findings were also supported by a recent work showing that *RUNX1* mutations could lead to the blockage of granulocytic differentiation in *in vitro* models and primary AML samples [26]. For upregulated genes, no pathways were significantly enriched.

RUNX1 target genes were generally repressed in *RUNX1*-mutated AML

Given that RUNX1 is a crucial TF in hematopoiesis, *RUNX1* mutations in AML may impact its target genes. We analyzed high-confidence RUNX1 binding peaks from the ChIP-seq data of three AML cell lines without *RUNX1* aberrations to identify RUNX1 target genes (Fig. 3A). High-confidence binding peaks were defined as those occurring in at least two of the cell lines. Genes with high-confidence RUNX1 binding peaks in their promoters were considered as RUNX1 target genes. Accordingly, a total of 1371 RUNX1 target genes were identified.

Next, we performed GSEA to investigate the general effects of *RUNX1* mutations on these RUNX1 target genes. RUNX1 targets were significantly enriched in the genes

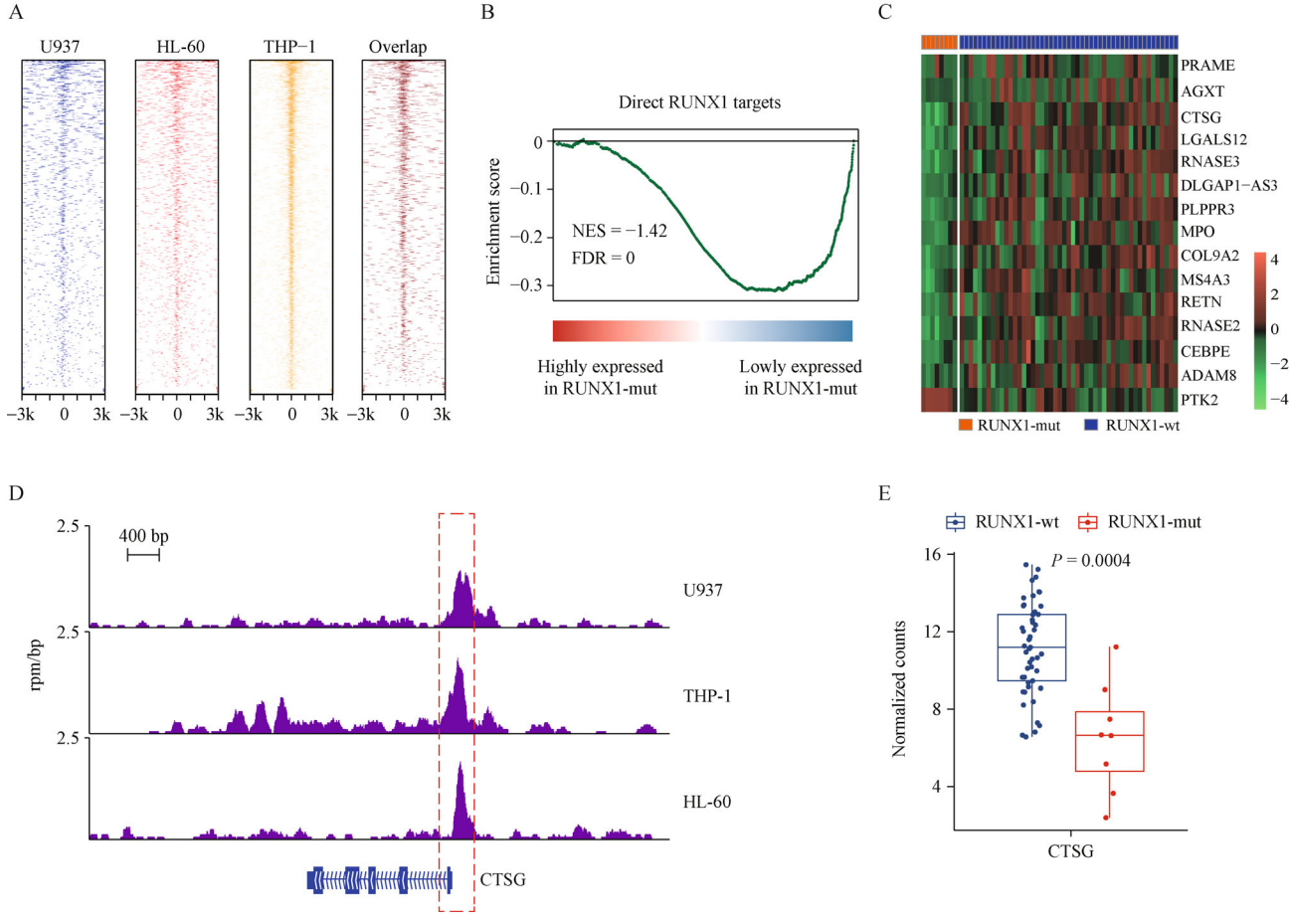

Fig. 3 RUNX1 target genes are generally repressed in *RUNX1*-mutated AML. (A) Heat maps illustrating RUNX1 binding signals in the regions near the TSSs in AML cell lines. Overlaps represent high-confidence binding peaks. (B) GSEA plot representing the repression of RUNX1 target genes. A total of 1216 RUNX1 target genes expressed in the TCGA cohort (mean RNA-seq counts > 2) were used as the gene set. FDR, false discovery rate; NES, normalized enrichment score. (C) Heat map illustrating the expression of 15 RUNX1 target genes included in the gene signature in CN-AML from the TCGA cohort. (D) Direct binding of RUNX1 on the *CTSG* promoter in AML cell lines. The *CTSG* transcript shown in this plot is NM_001911.2. (E) *CTSG* was significantly downregulated in patients with *RUNX1* mutations (Wilcoxon signed-rank test).

with the negative $\log_2$ fold change in *RUNX1*-mutated AML (Fig. 3B), indicating that RUNX1 target genes were generally repressed in AML with *RUNX1* mutations. The aforementioned 300-gene signature that represented the distinct expression pattern in *RUNX1*-mutated AML was chosen for in-depth investigation to determine the effects of mutated *RUNX1* on the genes with the most differential expression. The intersection between the RUNX1 targets and the signature yielded 15 genes that were deregulated RUNX1 targets in *RUNX1*-mutated AML (Fig. 3C). Among these 15 RUNX1 targets, 14 were downregulated and one was upregulated, suggesting that the direct effect of *RUNX1* mutations on transcription was predominantly repressive (Fig. 3C). For example, *CTSG*, which is associated with neutrophil-associated host defense and immune response, was a target gene of RUNX1 in all three AML cell lines (Fig. 3D) and was significantly downregulated in patients with *RUNX1* mutations relative to in patients with wild-type *RUNX1* (Fig. 3E). Our previously published study has demonstrated that the suppression of *CTSG* in t(8;21) AML is caused by RUNX1–RUNX1T1 direct targeting and plays a critical role in the pathogenesis of AML [27]. The results of the current work showed that the role of mutant RUNX1 in transcriptional repression might be similar to that of RUNX1–RUNX1T1 and might contribute to the progression of AML by inhibiting the expression of *CTSG*.

Association of *RUNX1* mutations with DNA hypermethylation

RUNX1 mutations might also influence epigenetic modifications, such as DNA methylation. Therefore, the DNA methylation microarray data of the TCGA AML cohort were reanalyzed. We improved the analysis accuracy by utilizing the updated hg38 human reference genome and a newly developed probe-masking method for Infinium HumanMethylation450K BeadChip probes [19]. Accordingly, we identified 210 DMRs, among which 183 were hypermethylated, and 27 were hypomethylated in *RUNX1*-mutated AML. This result indicated that *RUNX1* mutations might be associated with DNA hypermethylation.

We assigned these DMRs to gene promoter regions and detected 51 differentially methylated genes (DMGs). Next, we intersected DEGs and DMGs to investigate whether the methylation status change caused dysregulated gene expression. The presence of 10 DEGs among 51 DMGs might partially suggest the contribution of the altered DNA methylation levels to gene expression changes caused by *RUNX1* mutations. The majority of these genes (8/10) were in the aforementioned gene signature of *RUNX1* mutations: one was hypomethylated and upregulated, whereas seven were hypermethylated and downregulated due to *RUNX1* mutations (Fig. 4A). For example, *MS4A3* is involved in the regulation of innate immune pathways and the G1/S cell cycle in hematopoiesis [28]. This result suggested that *RUNX1* mutations might lead to promoter hypermethylation and *MS4A3* expression repression (Fig. 4B). These events might affect cell cycle regulation and contribute to the differentiation blockage of AML.

Next, the TCGA AML cohort was used to evaluate the prognostic influence of the expression of hypermethylated and downregulated genes (see the section of "Materials and methods"). As shown in Fig. 4C, the low expression of *MS4A3* was associated with increased hazard ratio and a poor prognosis, indicating that *MS4A3* downregulation might be crippling in *RUNX1*-mutated AML. Furthermore, the low expression of two additional hypermethylated and downregulated genes, namely, *CD96* and *LTK*, was also correlated with the poor prognosis of patients with AML (Fig. S2). Additionally, the clinical outcome was less influenced by treatment choices than other factors (Table S2, Table S3, and Table S4). In summary, the results suggested that DNA hypermethylation was associated with *RUNX1* mutations and contributed to gene repression caused by *RUNX1* mutations.

Discussion

Somatic gene mutations constitute key events in AML pathogenesis, and *RUNX1*-mutated AML has been added as a provisional category in accordance with the updated (2016) WHO classification. In this study, we investigated the genetic characteristics of *RUNX1* mutations in AML and showed that patients with *RUNX1*-mutated AML had a distinct gene expression pattern. We found that *RUNX1* mutations generally repressed RUNX1 target genes and were associated with DNA hypermethylation. Our investigation helps understand the pathobiology of *RUNX1*-mutated AML.

Using a large number of samples helps us obtain an improved understanding of the characteristics of mutation distribution. First, we identified several previously unreported recurrent mutations in *RUNX1*, such as p.R166Q, p.D198G, p.R201fs, and p.R204fs. In addition to recurrent mutations, we found some interesting results regarding co-occurrent or mutually exclusive mutations with *RUNX1* mutations. Cases with CN-AML were chosen to analyze the co-occurrent or mutually exclusive mutations with *RUNX1* mutations because *RUNX1* mutations mainly occur in patients with AML and normal karyotypes [6]. In addition to confirming the previous finding on the mutual exclusivity of *RUNX1* mutations with *NPM1* mutations [7], we obtained several novel findings: (1) mutations previously reported as concurrent, such as *IDH1*, *IDH2*, and *DNMT3A* [29], did not show a significantly higher frequency in the *RUNX1*-mut cohort than in the *RUNX1*-wt cohort and (2) *DNMT3A* mutations even had a statistically significantly lower frequency in the

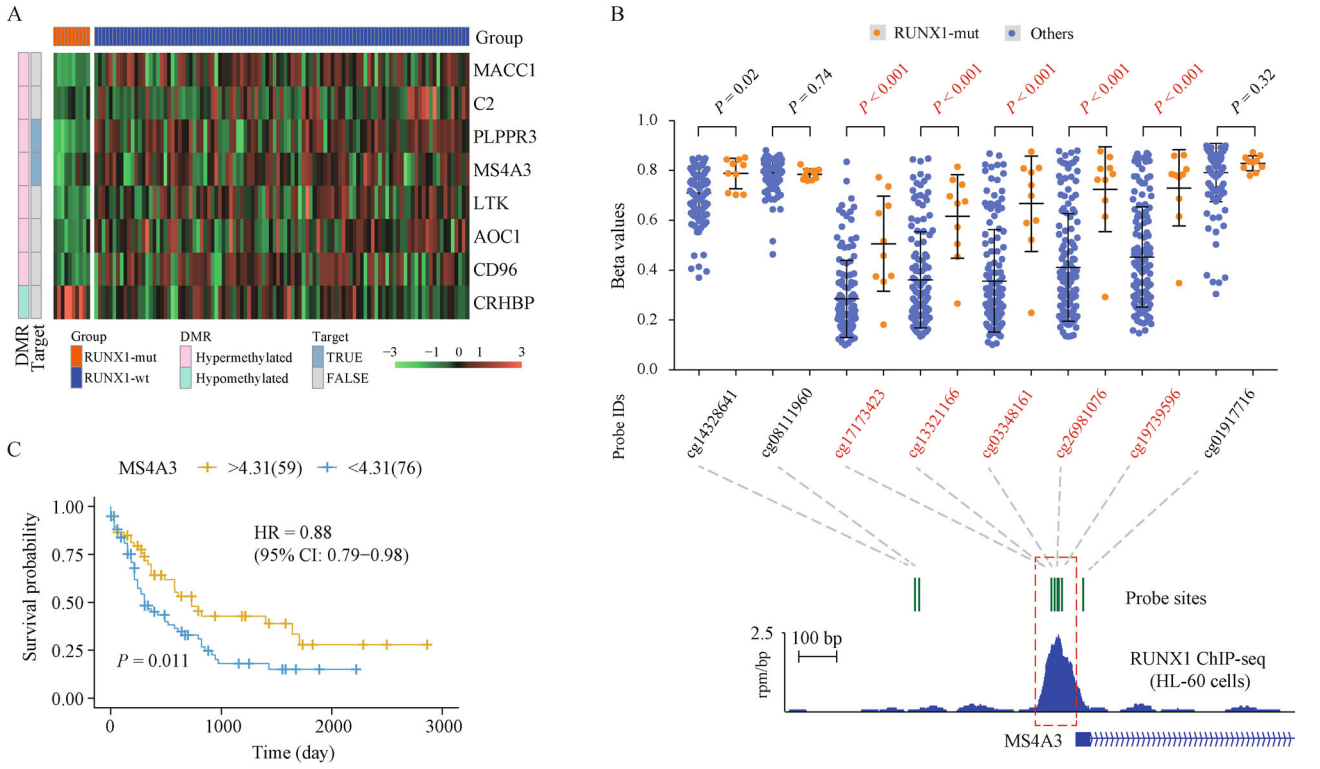

Fig. 4 *RUNX1* mutations are associated with DNA hypermethylation. (A) Heat map representing the gene expression of eight DMGs included in the gene signature of *RUNX1* mutations. A total of 113 patients from TCGA were included in the analysis. DMR represents the DNA methylation status of the gene promoters. Target represents RUNX1 binding status on the promoters. mut, mutated; wt, wild-type. (B) Schematic of the hypermethylated promoter region of *MS4A3*. The scatter plot represents the levels of DNA methylation beta values in the TCGA AML cohort. Beta values for specific CpG residues are shown as a proportion ranging from 0 (unmethylated, 0%) to 1 (fully methylated, 100%). (C) Kaplan–Meier plot of the overall survival of patients with AML from TCGA grouped by the expression levels of *MS4A3*. The hazard ratio was calculated by utilizing univariate Cox regression. Gene expression levels were evaluated by using FPKM values converted from read counts by DESeq2. The cut-off value ($\log_2$FPKM = 4.31) was selected via the method of maximum selected rank statistics provided in the survminer package.

RUNX1-mut cohort than in other cohorts. These new findings were mainly attributed to the effective analysis based on CN-AML because the exclusion of AML cases with abnormal karyotypes removed biases and avoided underestimating the frequency of mutations, such as *DNMT3A* and *IDH1/2*, that are highly prevalent in CN-AML [30,31].

We developed a gene expression signature that represented the most distinguishable expression features of patients with *RUNX1* mutations and could separate these patients from patients with wild-type *RUNX1* in sample clustering. The performance of this signature had improved compared with that of a previously reported signature [6] because the effects of *NPM1* mutations were excluded. High-frequency *NPM1* mutations were almost exclusively enriched in *RUNX1* wild-type AMLs. The removal of *NPM1*-mutated cases prior to gene signature development could clearly identify the differences caused by *RUNX1* rather than the combinatory effects of *RUNX1* and *NPM1* mutations. Moreover, compared with the noncommercial microarray data used in previous studies [6,32], we used the updated TCGA RNA-seq data of hg38 (release on September 27, 2018), which covered more of the genome and had high levels of reproducibility. In addition, GO enrichment analysis based on our gene signature showed that pathways related to neutrophil functions were downregulated in *RUNX1*-mutated AML; this result was consistent with the finding of a recently published work showing that *RUNX1* mutations lead to a blockage in granulocytic differentiation in human $CD34^+$ progenitor cells [26].

Our results offer insights into the mechanisms underlying gene expression alterations in cases with *RUNX1* mutations. First, we considered that *RUNX1* mutations might lead to the dysfunction of RUNX1 in transcriptional regulation. When we took all the expression differences into consideration using GSEA, RUNX1 target genes tended to be downregulated in *RUNX1*-mutated AML.

When we used relatively loose cut-offs to conduct differential gene expression analysis (FDR < 0.05, FC > 1.5 or < 2/3), approximately 7.3% (n = 100) of RUNX1 targets were found to be differentially expressed with statistical significance and generally downregulated (n = 68). This result was suggestive of noticeable direct gene repression caused by *RUNX1* mutations. Second, we investigated the effects of *RUNX1* mutations on DNA methylation. We found that 10 out of 1007 DEGs were differentially methylated, indicating that DNA methylation partially accounted for the altered gene expression levels. Three DEGs, namely, *MS4A3*, *PLPPR3*, and *C20orf197*, were found to be RUNX1 targets that were hypermethylated and downregulated in *RUNX1*-mutated AML. In normal hematopoietic cells, RUNX1 regulates DNA demethylation by recruiting the DNA demethylation enzymes TET2, TET3, TDG, and GADD45 [33]. We may reasonably hypothesize that in *RUNX1*-mutated AML, *RUNX1* mutations led to a mechanism that blocked the recruitment of demethylation modifiers. This effect increased the methylation of RUNX1 target genes and reduced their gene expression. However, only bioinformatics analyses were conducted in our study, and further wet experiments on regulatory targets and methylation status are required for in-depth investigation.

Our work not only indicates that *RUNX1*-mutated AML is an entity with a distinct gene expression pattern but also provides insights into the pathogenic mechanisms caused by *RUNX1* mutations. Furthermore, the repression of *RUNX1* and its targets has been reported to inhibit growth and induce the apoptosis of AML cells expressing mutant *RUNX1* [34], suggesting that expression mimickers that deplete *RUNX1* and its target gene expression levels may become novel therapeutic agents. Therefore, our findings on the gene signatures that are especially dysregulated in *RUNX1*-mutated AML may help identify promising targets with prognostic and therapeutic utilities.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Nos. 81890994, 81770153, 81530003, and 81911530240) and the National Key Research and Development Program of China (No. 2019YFA0905900).

Compliance with ethics guidelines

Jingming Li, Wen Jin, Yun Tan, Beichen Wang, Xiaoling Wang, Ming Zhao, and Kankan Wang declare that they have no conflict of interest. This article does not contain any studies with human or animal subjects.

Electronic Supplementary Material Supplementary material is available in the online version of this article at https://doi.org/10.1007/s11684-020-0815-4 and is accessible for authorized users.

References

1. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet 2016; 17(9): 551–565
2. Guo H, Ma O, Speck NA, Friedman AD. Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis. Blood 2012; 119(19): 4408–4418
3. Tober J, Yzaguirre AD, Piwarzyk E, Speck NA. Distinct temporal requirements for Runx1 in hematopoietic progenitors and stem cells. Development 2013; 140(18): 3765–3776
4. Sood R, Kamikubo Y, Liu P. Role of RUNX1 in hematological malignancies. Blood 2017; 129(15): 2070–2082
5. Tang JL, Hou HA, Chen CY, Liu CY, Chou WC, Tseng MH, Huang CF, Lee FY, Liu MC, Yao M, Huang SY, Ko BS, Hsu SC, Wu SJ, Tsay W, Chen YC, Lin LI, Tien HF. AML1/RUNX1 mutations in 470 adult patients with *de novo* acute myeloid leukemia: prognostic implication and interaction with other gene alterations. Blood 2009; 114(26): 5352–5361
6. Gaidzik VI, Bullinger L, Schlenk RF, Zimmermann AS, Röck J, Paschka P, Corbacioglu A, Krauter J, Schlegelberger B, Ganser A, Späth D, Kündgen A, Schmidt-Wolf IG, Götze K, Nachbaur D, Pfreundschuh M, Horst HA, Döhner H, Döhner K. RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group. J Clin Oncol 2011; 29(10): 1364–1372
7. Mendler JH, Maharry K, Radmacher MD, Mrózek K, Becker H, Metzeler KH, Schwind S, Whitman SP, Khalife J, Kohlschmidt J, Nicolet D, Powell BL, Carter TH, Wetzler M, Moore JO, Kolitz JE, Baer MR, Carroll AJ, Larson RA, Caligiuri MA, Marcucci G, Bloomfield CD. RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and microRNA expression signatures. J Clin Oncol 2012; 30(25): 3109–3118
8. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016; 127(20): 2391–2405
9. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, Gundem G, Van Loo P, Martincorena I, Ganly P, Mudie L, McLaren S, O'Meara S, Raine K, Jones DR, Teague JW, Butler AP, Greaves MF, Ganser A, Döhner K, Schlenk RF, Döhner H, Campbell PJ. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016; 374(23): 2209–2221
10. Gerstung M, Papaemmanuil E, Martincorena I, Bullinger L, Gaidzik VI, Paschka P, Heuser M, Thol F, Bolli N, Ganly P, Ganser A, McDermott U, Döhner K, Schlenk RF, Döhner H, Campbell PJ. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet 2017; 49(3): 332–340
11. Love MI, Anders S, Kim V, Huber W. RNA-Seq workflow: gene-level exploratory analysis and differential expression. F1000 Res 2015; 4: 1070
12. Langmead B, Wilks C, Antonescu V, Charles R. Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics 2019; 35(3): 421–432
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth

G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25(16): 2078–2079

14. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008; 9(9): R137
15. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26(6): 841–842
16. Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, Zhu M, Wu J, Shi X, Taing L, Liu T, Brown M, Meyer CA, Liu XS. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res 2017; 45(D1): D658–D662
17. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, Gibson D, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res 2019; 47(D1): D853–D858
18. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. N Engl J Med 2013; 368(22): 2059–2074
19. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res 2017; 45(4): e22
20. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014; 30(10): 1363–1369
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43(7): e47
22. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, Clark SJ, Molloy PL. *De novo* identification of differentially methylated regions in the human genome. Epigenetics Chromatin 2015; 8(1): 6
23. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16 (5): 284–287
24. Duployez N, Marceau-Renaut A, Boissel N, Petit A, Bucci M, Geffroy S, Lapillonne H, Renneville A, Ragu C, Figeac M, Celli-Lebras K, Lacombe C, Micol JB, Abdel-Wahab O, Cornillet P, Ifrah N, Dombret H, Leverger G, Jourdan E, Preudhomme C. Comprehensive mutational profiling of core binding factor acute myeloid leukemia. Blood 2016; 127(20): 2451–2459
25. Fukunaga J, Nomura Y, Tanaka Y, Amano R, Tanaka T, Nakamura Y, Kawai G, Sakamoto T, Kozu T. The Runt domain of AML1 (RUNX1) binds a sequence-conserved RNA motif that mimics a DNA element. RNA 2013; 19(7): 927–936
26. Gerritsen M, Yi G, Tijchon E, Kuster J, Schuringa JJ, Martens JHA, Vellenga E. RUNX1 mutations enhance self-renewal and block granulocytic differentiation in human *in vitro* models and primary AMLs. Blood Adv 2019; 3(3): 320–332
27. Jin W, Wu K, Li YZ, Yang WT, Zou B, Zhang F, Zhang J, Wang KK. AML1-ETO targets and suppresses cathepsin G, a serine protease, which is able to degrade AML1-ETO in t(8;21) acute myeloid leukemia. Oncogene 2013; 32(15): 1978–1987
28. Kutok JL, Yang X, Folkerth R, Adra CN. Characterization of the expression of HTm4 (MS4A3), a cell cycle regulator, in human peripheral blood cells and normal and malignant tissues. J Cell Mol Med 2011; 15(1): 86–93
29. Khan M, Cortes J, Kadia T, Naqvi K, Brandt M, Pierce S, Patel KP, Borthakur G, Ravandi F, Konopleva M, Kornblau S, Kantarjian H, Bhalla K, DiNardo CD. Clinical outcomes and co-occurring mutations in patients with RUNX1-mutated acute myeloid leukemia. Int J Mol Sci 2017; 18(8): E1618
30. O'Brien EC, Brewin J, Chevassut T. DNMT3A: the DioNysian MonsTer of acute myeloid leukaemia. Ther Adv Hematol 2014; 5 (6): 187–196
31. Ok CY, Loghavi S, Sui D, Wei P, Kanagal-Shamanna R, Yin CC, Zuo Z, Routbort MJ, Tang G, Tang Z, Jorgensen JL, Luthra R, Ravandi F, Kantarjian HM, DiNardo CD, Medeiros LJ, Wang SA, Patel KP. Persistent *IDH1/2* mutations in remission can predict relapse in patients with acute myeloid leukemia. Haematologica 2019; 104(2): 305–311
32. Greif PA, Konstandin NP, Metzeler KH, Herold T, Pasalic Z, Ksienzyk B, Dufour A, Schneider F, Schneider S, Kakadia PM, Braess J, Sauerland MC, Berdel WE, Büchner T, Woermann BJ, Hiddemann W, Spiekermann K, Bohlander SK. RUNX1 mutations in cytogenetically normal acute myeloid leukemia are associated with a poor prognosis and up-regulation of lymphoid genes. Haematologica 2012; 97(12): 1909–1915
33. Suzuki T, Shimizu Y, Furuhata E, Maeda S, Kishima M, Nishimura H, Enomoto S, Hayashizaki Y, Suzuki H. RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. Blood Adv 2017; 1(20): 1699–1711
34. Mill CP, Fiskus W, DiNardo CD, Qian Y, Raina K, Rajapakshe K, Perera D, Coarfa C, Kadia TM, Khoury JD, Saenz DT, Saenz DN, Illendula A, Takahashi K, Kornblau SM, Green MR, Futreal AP, Bushweller JH, Crews CM, Bhalla KN. RUNX1-targeted therapy for AML expressing somatic or germline mutation in RUNX1. Blood 2019; 134(1): 59–73