PNAS

Supplementary Information for

Aging and comprehensive molecular profiling in acute myeloid leukemia

Jian-Feng Li[a,1], Wen-Yan Cheng[a,1], Xiang-Jie Lin[b,c,1], Li-Jun Wen[d,e,1], Kai Wang[f,1], Yong-Mei Zhu[a,1], Hong-Ming Zhu[a], Xin-Jie Chen[a], Yu-Liang Zhang[a], Wei Yin[a], Jia-Nan Zhang[a], Xiao Yi[a], Fan Zhang[a], Xiang-Qin Weng[a], Sheng-Yue Wang[a], Lu Jiang[a], Hui-Yi Wu[a], Jia-Qi Ren[a], Xiao-Jing Lin[a], Niu Qiao[a], Yu-Ting Dai[a], Hai Fang[a], Yun Tan[a], Xiao-Jian Sun[a], Gang Lv[a], Xiao-Yu Yan[a], Su-Ning Chen[d,e], Zhu Chen[a,2], Jie Jin[b,c,g,2], De-Pei Wu[d,e,2], Rui-Bao Ren[a,f,2], Sai-Juan Chen[a,2], and Yang Shen[a,2]

[a] 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; [b] Department of Hematology, The First Affiliated Hospital, Zhejiang University College of Medicine, Hangzhou, Zhejiang, 310003, China; [c] Key Laboratory of Hematologic Malignancies, Diagnosis and Treatment, Hangzhou, Zhejiang, 310003, China; [d] National Clinical Research Center for Hematologic Diseases, Jiangsu Institute of Hematology, The First Affiliated Hospital of Soochow University, Suzhou, 215006, China; [e] Institute of Blood and Marrow Transplantation, Collaborative Innovation Center of Hematology, Soochow University, Suzhou, 215006, China; [f] International Center for Aging and Cancer, Department of Hematology of The First Affiliated Hospital, Hainan Medical University, Haikou, 571199, China; [g] Zhejiang University Cancer Center, Hangzhou, Zhejiang, 310003, China;

[1] J.-F.L., W.-Y.C., X.J.L., L.J.W., K.W., and Y.-M.Z. contributed equally to this work.

[2] To whom correspondence may be addressed. Email: zchen@stn.sh.cn, jiej0503@zju.edu.cn, wudepei@suda.edu.cn, rbren@sjtu.edu.cn, sjchen@stn.sh.cn or yang_shen@sjtu.edu.cn.

Correspondence authors: Chen Zhu, Jie Jin, De-Pei Wu, Rui-Bao Ren, Sai-Juan Chen, Yang Shen

This PDF file includes:

Supplementary Information Text
Figures S1 to S10
Tables S1 to S6
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S11

Supplementary Information Text

SI Materials and Methods

Treatment protocols

For non-M3 acute myeloid leukemia (AML), young patients (< 60 years) were given standard intensive "3+7" IA/DA-based regimens as initial induction, which contained idarubicin/daunorubicin (10–12/45–60 mg/m^2, D1–3) and cytarabine (100 mg/m^2, D1–7). When complete remission (CR) was achieved, 4 cycles of high-dose cytarabine (HDAC, 2g/m^2 q12h×6, D1–3) were delivered as consolidation therapy. Elderly patients (≥ 60 years) were evaluated by the treating physician. Fit patients received reduced IA/DA-based induction chemotherapy comprising idarubicin (6 mg/m^2 D1–3) and cytarabine (100 mg/m^2, D1–7), and reduced the consolidation to 2 cycles of HDAC (2 g/m^2 q12h×6, D1–3). While unfit patients were assigned to other less intensive therapies, e.g., venetoclax-containing regimens, hypomethylation agents (HMA)-based regimens at the discretion of the treating physician. Eligible patients received hematopoietic stem cell transplantation (HSCT) as consolidation.
For patients with acute promyelocytic leukemia (APL), the combination of All-trans retinoic acid (ATRA) and Arsenic trioxide (ATO) with or without chemotherapy was administered based on Sanz risk stratification.

Gene fusion calling

Potential gene fusion events were detected by RNA sequencing (RNA-Seq), karyotyping, and/or fluorescence in situ hybridization (FISH). Two methods including fusioncatcher (v1.33) and arriba (v2.4.0) (1) were used to call gene fusions from RNA-Seq data. The majority of reported terms were validated by PCR method. The *FLT3*-ITD and *KMT2A*-PTD events were called based on the arriba method for patients without DNA panel sequencing or PCR validation.

Gene expression quantification

Quantification of transcript read counts was based on the alignment-free method kallisto (v0.46.2) (2) and GENCODE v43 reference transcriptome/gene models using the raw FASTQ files. The fastp (v0.23.2) was used for basic quality control with the parameters "-Q -c -L". The tximport (v.1.28.0) was used to merge transcript counts for quantification of gene expression at the gene level. The gene expression matrix was generated by DESeq2 (v1.28.0) (3) based on the count table files and internal normalization with variance-stabilizing transformation

(VST). Transcripts Per Kilobase Million (TPM) were also generated for partial downstream analysis, mainly CIBERSORTx-based immune cell deconvolution (4). The ComBat function in the R sva package (v3.40.0) was used to adjust batch effect in both the discovery and external validation cohorts. In addition, before calculating age correlation of gene expressions, only protein-coding genes were retained and the 'adjust_matrix' function in the cola (5) R package was used to remove rows with low variance, leaving 18,383 genes. The R package ComplexHeatmap (v 2.16.0) (6) was used to conduct the matrix-based clustering based on the ward.D or ward.D2 methods and '1-cor(t(x)))/2' distance measure.

To identify age-related genes, we first calculated Pearson correlation coefficients using the normalized gene expression matrix and age. We then performed the differential expression gene (DEG) analysis between possible combinations of age groups using the limma package (v3.56.2), while the <40 age groups (<20, 20-29 and 30-39) were combined as single group to improve the power of the test statistics. All DEGs of age groups [adjusted P<0.05 and |log2 (fold change)| > 1] were merged to narrow down the candidate genes.

Gene sets and pathway enrichment analysis

We also retrieved and constructed aging-related gene sets based on public databases, publications and our AML cohort. The first part is the hallmarks of ageing, which includes the Aging Atlas (7), MsigDB (8), and the epigenetic genes from the EpiFactors database (9). The second part is the gene sets from published works, which provides the genes associated with differentiation stages of AML and hematopoietic cells (10, 11). In addition, we also construct in-house age-related gene sets based on the Spearman coefficients of gene-age pairs (greater than or less than 0.15) and their pathway enrichments. Pathway-level clustering of single-sample gene set enrichment score (GSVA tool, v1.48.0) was used to integrate age-related features from clinical, gene fusion, genetic mutations and marker genes. Pathway scores were normalized using the absolute difference between the minimum and maximum values. The STRING website was used to perform pathway analysis of age-related genes and to visualize these age-related gene/proteins.

Variant calling from DNA and RNA sequencing data

The variants calling of targeted exome sequencing and RNA-Seq data were described in our previous published work (12). Briefly, paired-end reads were aligned to the human hg19 reference genome using BWA (v0.7.17) and STAR (v2.7.10) (13) two-pass mode for DNA and RNA-Seq data, respectively. The GATK HaplotypeCaller (v4.1.7.0), GATK UnifiedGenoTyper

(v3.8.0), Lofreq (v2.1.2) (14), Freebayes (v.1.3.2), Vardict (v1.8.3) (15), Varscan (v2.4.5), Strelka (v2.9.10), and Pindel (0.2.5b9) were combined to create variant call format (VCF) files. Then, Multiple calls were then merged based on genotypes and maximum variant allele frequency (VAF). The generated VCF files were annotated and converted to MAF format files by using the VEP (v105) (16) and vcf2maf (v1.6.18). The whole-exome sequencing calling set was obtained from our previously published APL paper (17).

SI Figures

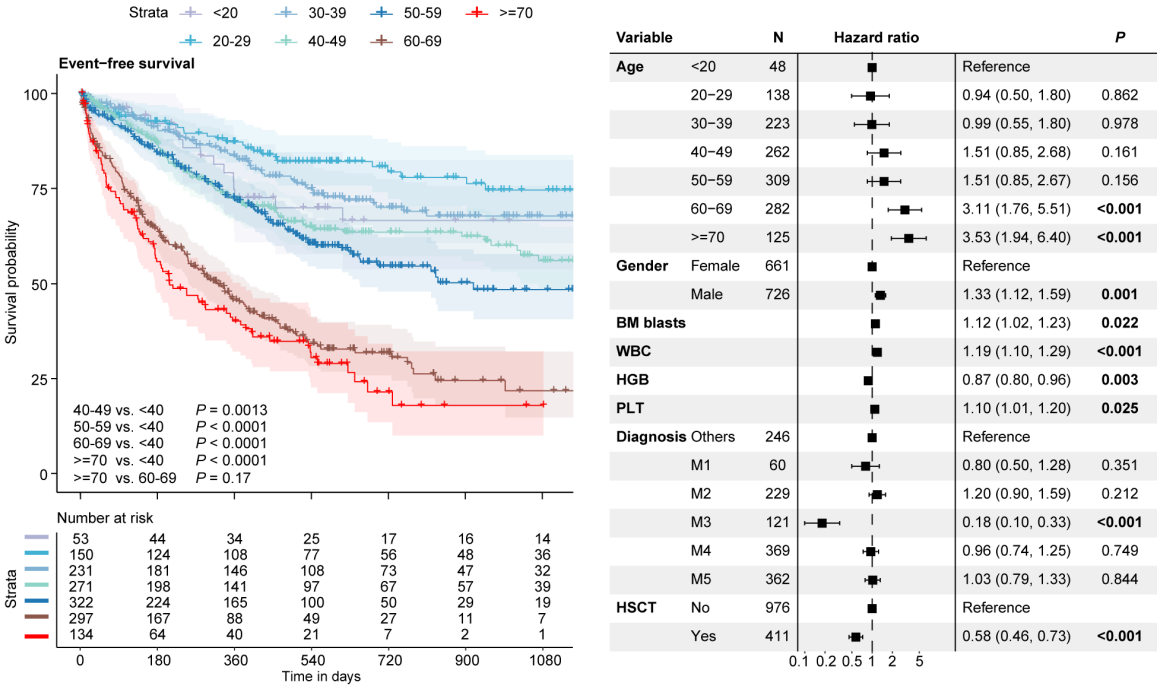

Fig. S1. Event-free survival for age groups and multivariate analysis in AML. The left panel shows the three-year EFS survival curve (Kaplan-Meier) of AML patients. Different age groups are represented by different colors. Statistical significance of survival between the two groups was calculated based on the log-rank test. The right panel shows the forest plot of basic clinical information of AML patients. The age groups 60-69 and >=70 years, male, WBC, HGB and PLT can predict poor prognosis, while M3 diagnosis and HSCT have the protective effect in AML patients. EFS, event-free survival. WBC, white blood count. HGB, hemoglobin. PLT, platelet.

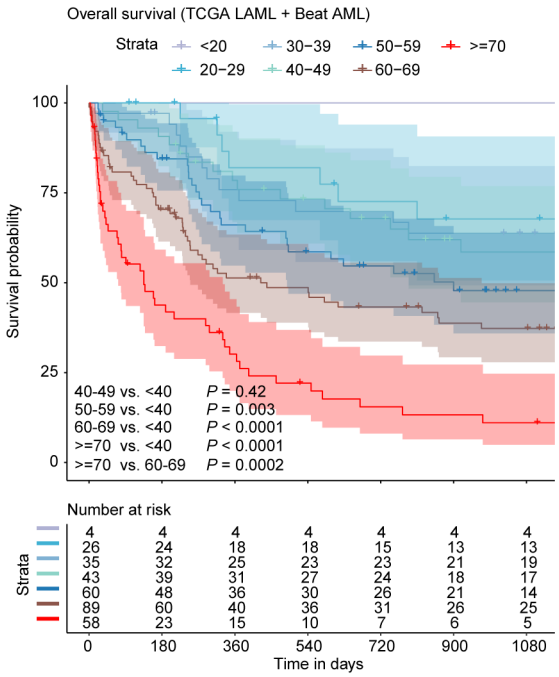

Fig. S2 Overall survival of age groups based on pooled TCGA LAML and Beat AML cohorts. Different age groups are represented by different colors. It is consistent with our cohorts, AML patients older than or equal to 70 years of age have the worst prognosis.

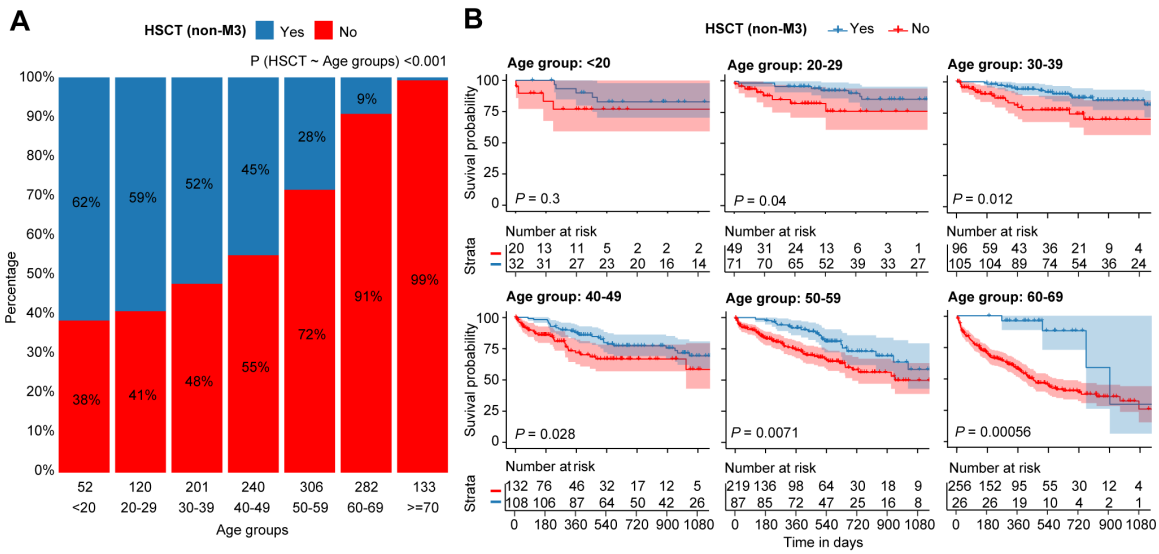

Fig. S3 Hematopoietic stem cell transplantation and the overall survival in age groups of non-M3 AML. (A) Percentage bar chart shows the proportion of patients with HSCT therapy (non-M3/acute promyelocytic leukemia diagnosis). Logistic regression indicates significant decrease (*P*<0.001) in percentage of HSCT with age. (B) Overall survival of age groups with and without HSCT treatment. Note that in patients aged 60-69 years, those who opted for HSCT had much better performance status. HSCT, hematopoietic stem cell transplantation.

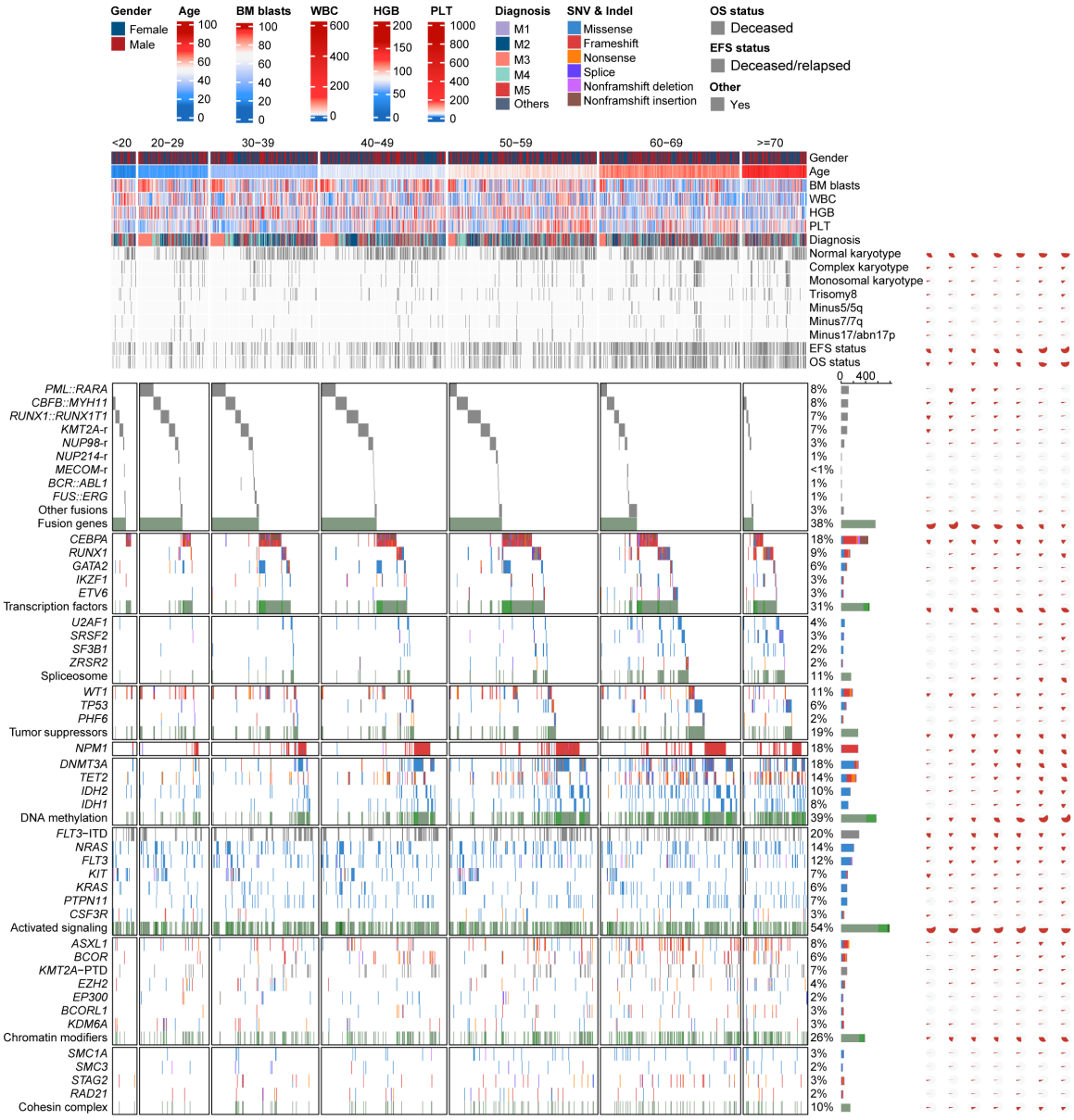

Fig. S4 Genomic landscape of age groups in 1,474 AML patients. Top and bottom heatmap panels show the basic clinical information and molecular events including gene fusions and 35 common mutated gene terms of AML patients. The right bar and pie charts show the overall percentage of positive events in all patients and in each age group. Common gene fusions decrease with age in AML, while most of the gene mutation pathways except activated signaling increase with age.

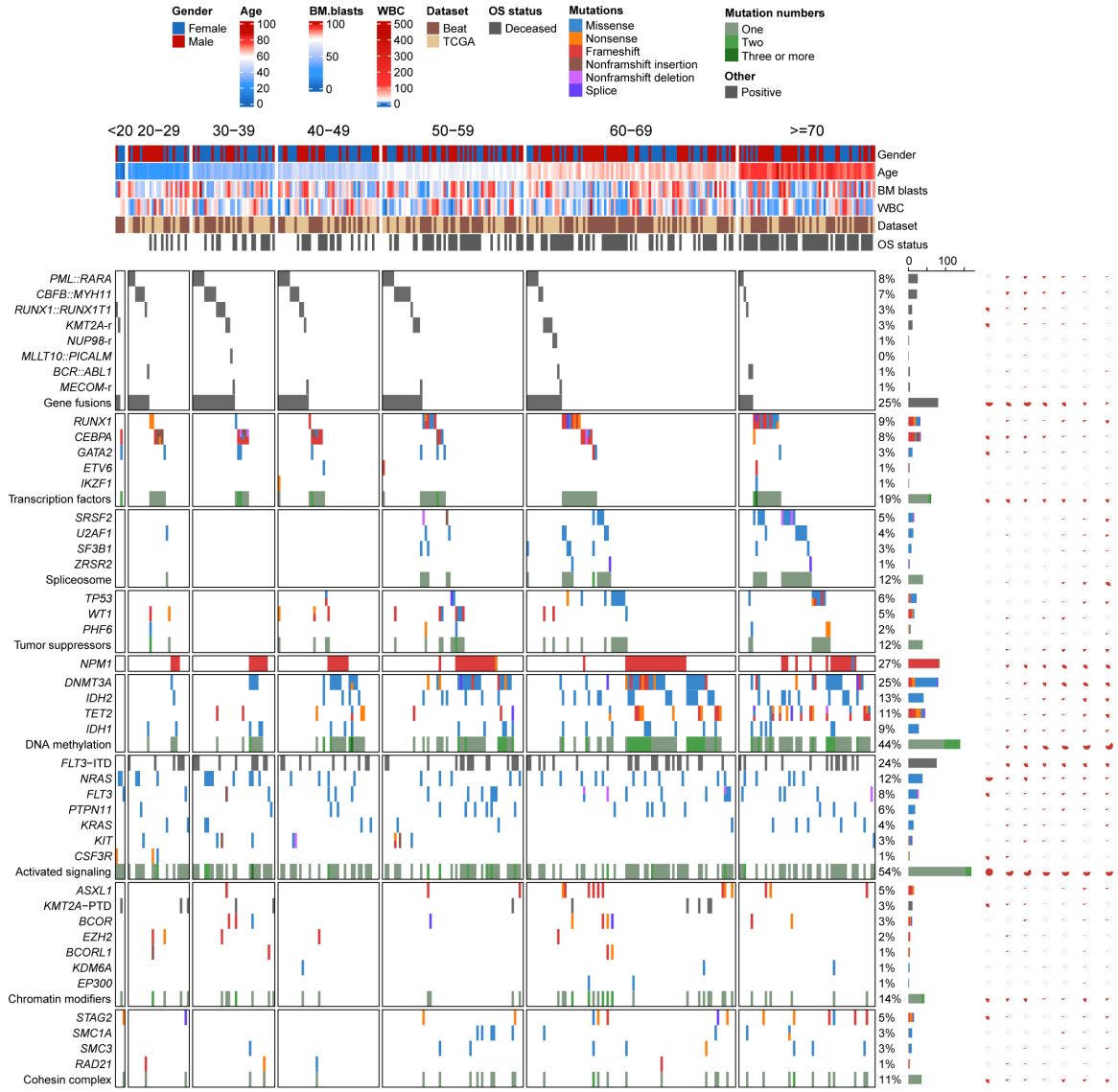

Fig. S5 Genomic landscape of age groups in the pooled TCGA LAML and Beat AML cohorts. It is consistent with our report, similar trend can be found in gene fusions, *NPM1*, spliceosome, tumor suppressors and DNA methylation pathways.

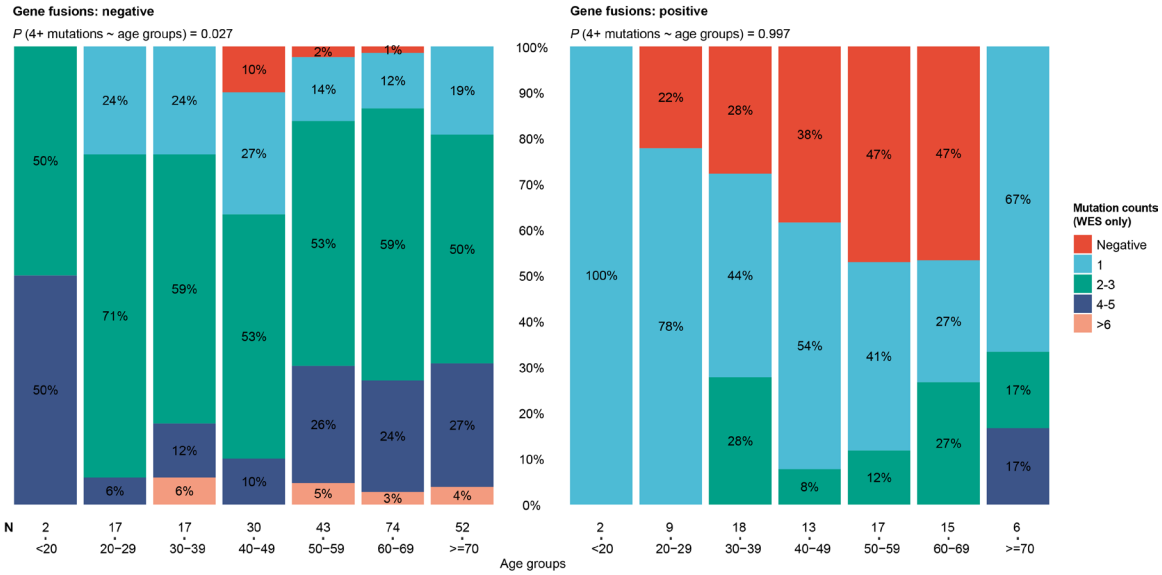

Fig. S6 Comparison of mutation count classes between gene fusion-negative and -positive patients in the pooled TCGA LAML and Beat AML cohorts. The percentage of mutation counts >= 2 is higher in all age groups of gene fusion-negative patients compared to gene fusion-positive patients. The significance of the change in the percentage of patients with four or more mutations with age was inferenced based on logistic regression.

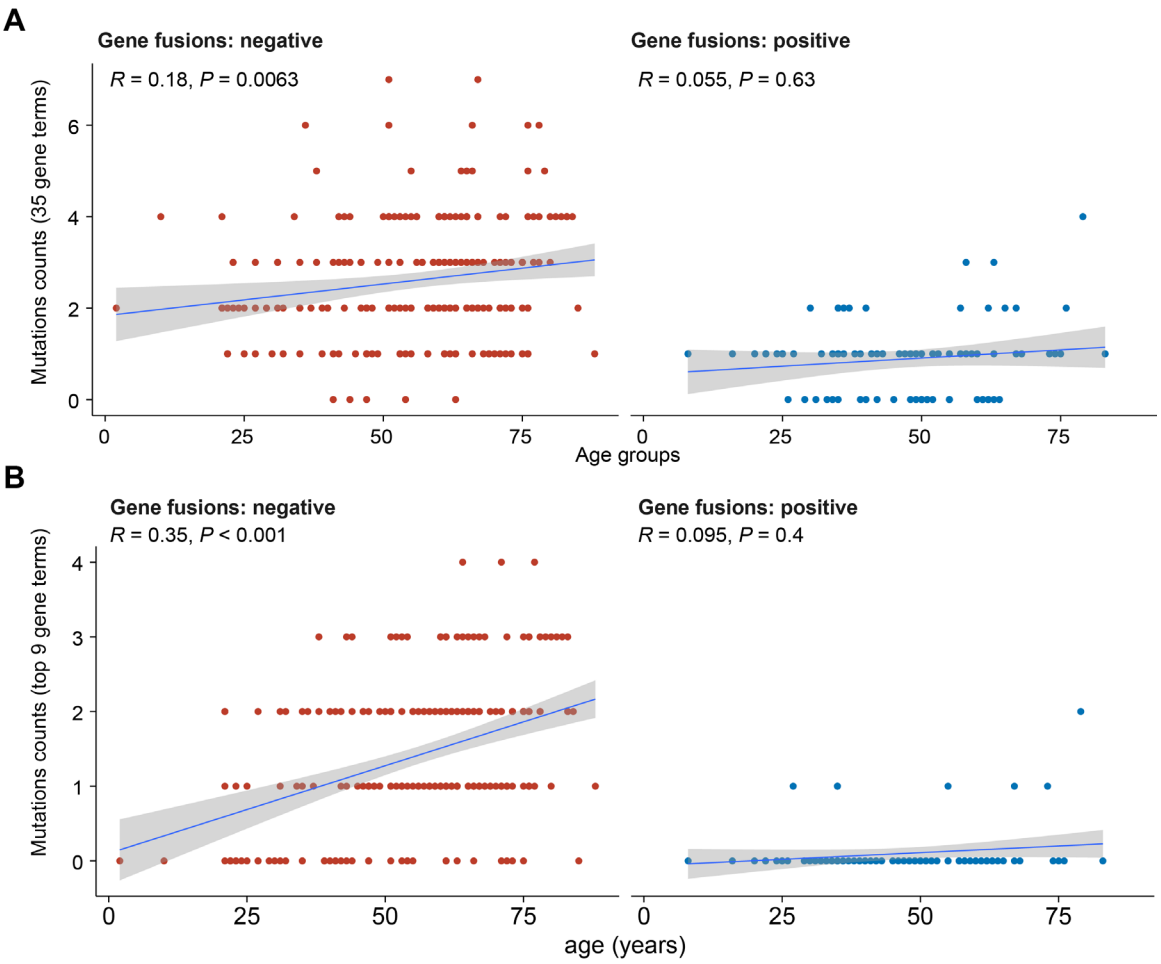

Fig. S7 Age correlation of mutation counts in the pooled TCGA LAML and Beat AML cohorts. (A) Scatter plots of age and mutation counts based on 35 common mutant gene terms. Compared with gene fusion-positive cases (right and blue points), gene fusion-negative (left and red points) patients show more strong correlation between age and mutation counts. (B) Top 9 age-correlated genes enhance the Spearman correlation coefficient of age and mutation counts, especially in gene fusion-negative patients.

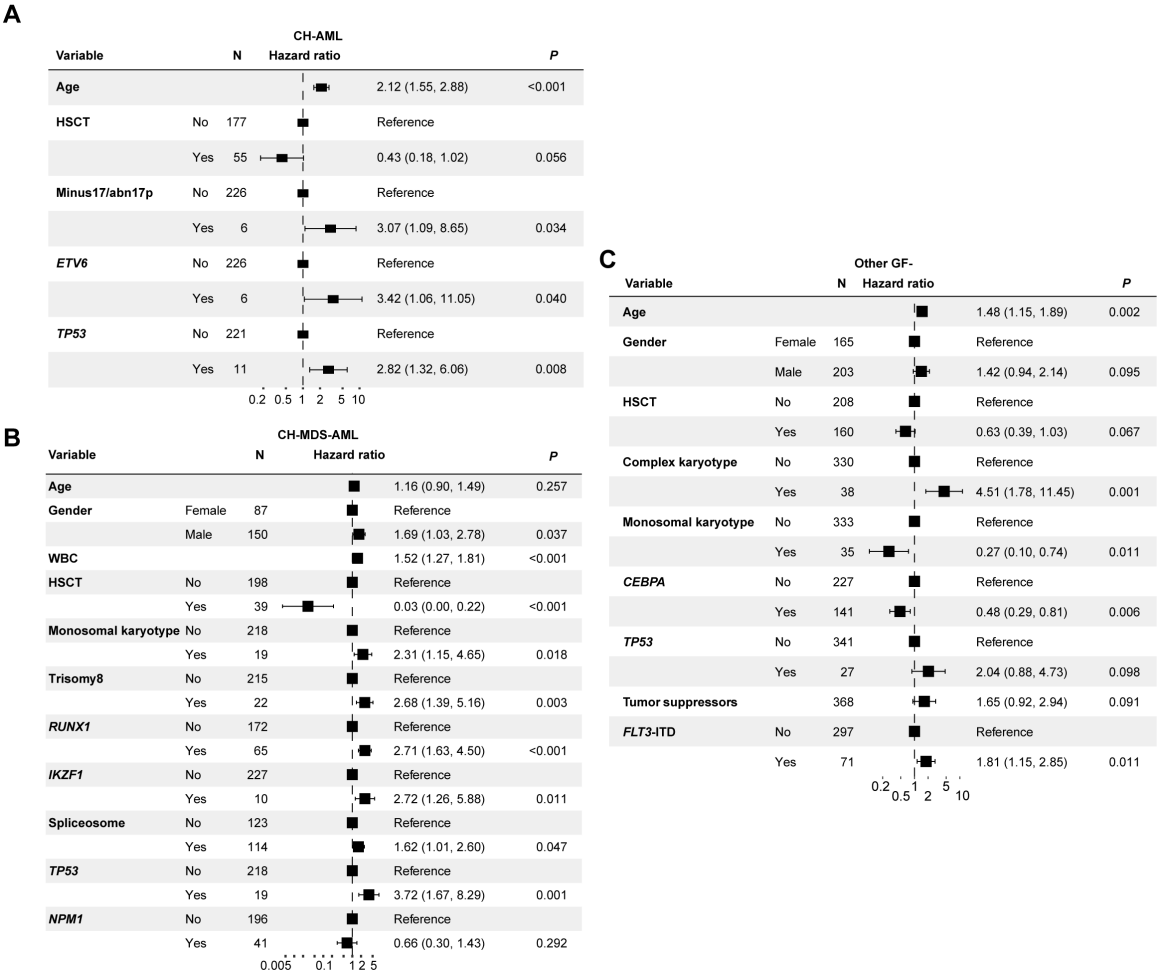

Fig. S8 Multivariant analysis of clinical and molecular events in clonal hematopoiesis groups of AML patients. Left panel are the forest plots of the (A) CH-AML (top) and (B) CH-MDS-AML (bottom) groups. (C) The forest plot of other gene fusion-negative patients is shown on in the right panel. Age, WBC, *ETV6* and *TP53* independently predict poor prognosis in CH-AML, while WBC, monosomal karyotype, trisomy 8, *RUNX1*, *IKZF1*, Spliceosome, and *TP53* predict poor prognosis in CH-MDS-AML. HSCT can significantly improve survival in both CH-AML and CH-MDS-AML groups of AMLs. WBC, white blood cell count. HSCT, hematopoietic stem cell transplantation. CH, clonal hematopoiesis. MDS, myelodysplastic syndromes.

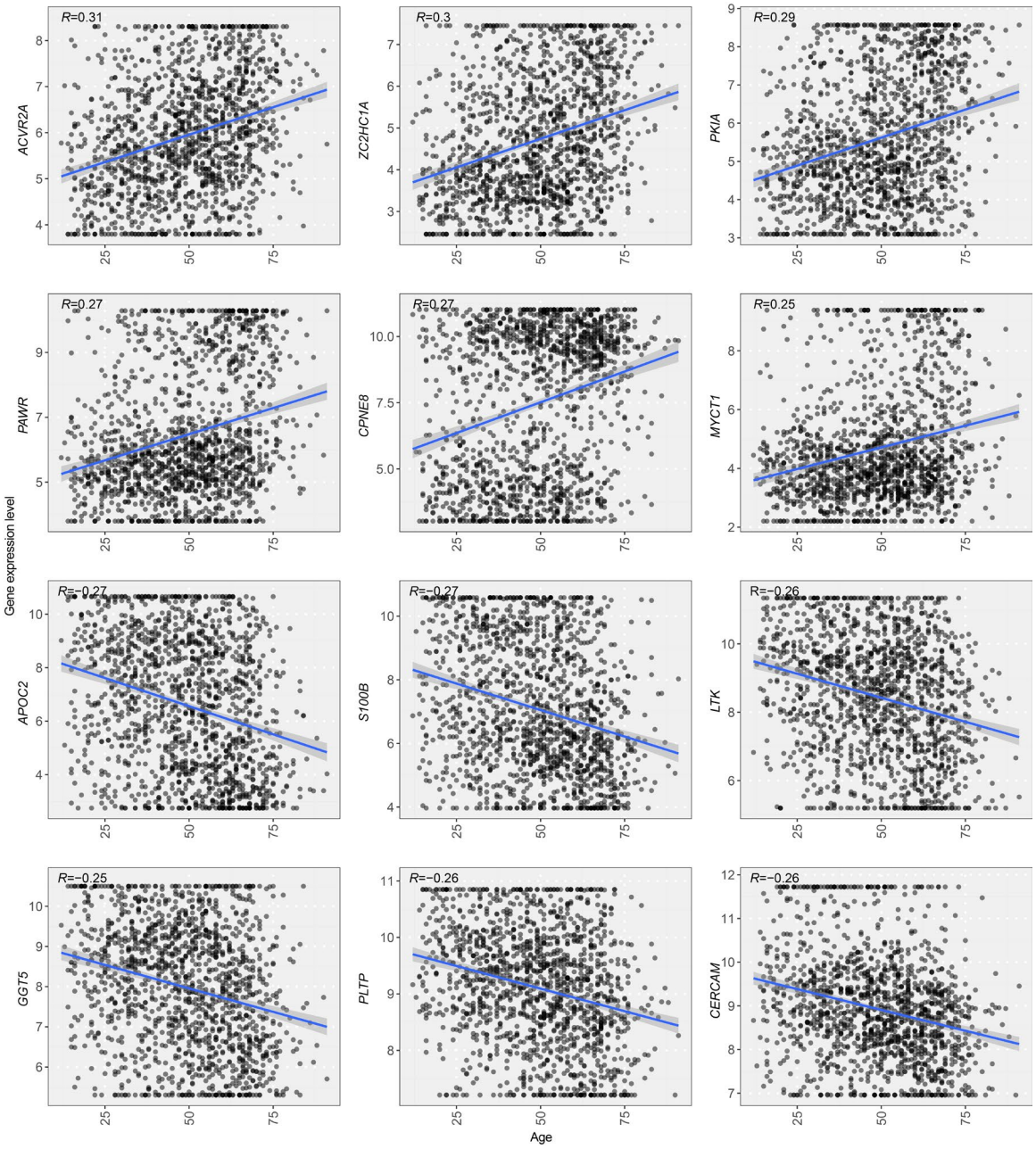

Fig. S9 Representative genes positively or negatively correlated with age. Pearson correlation coefficient was labeled at top-left in each plot.

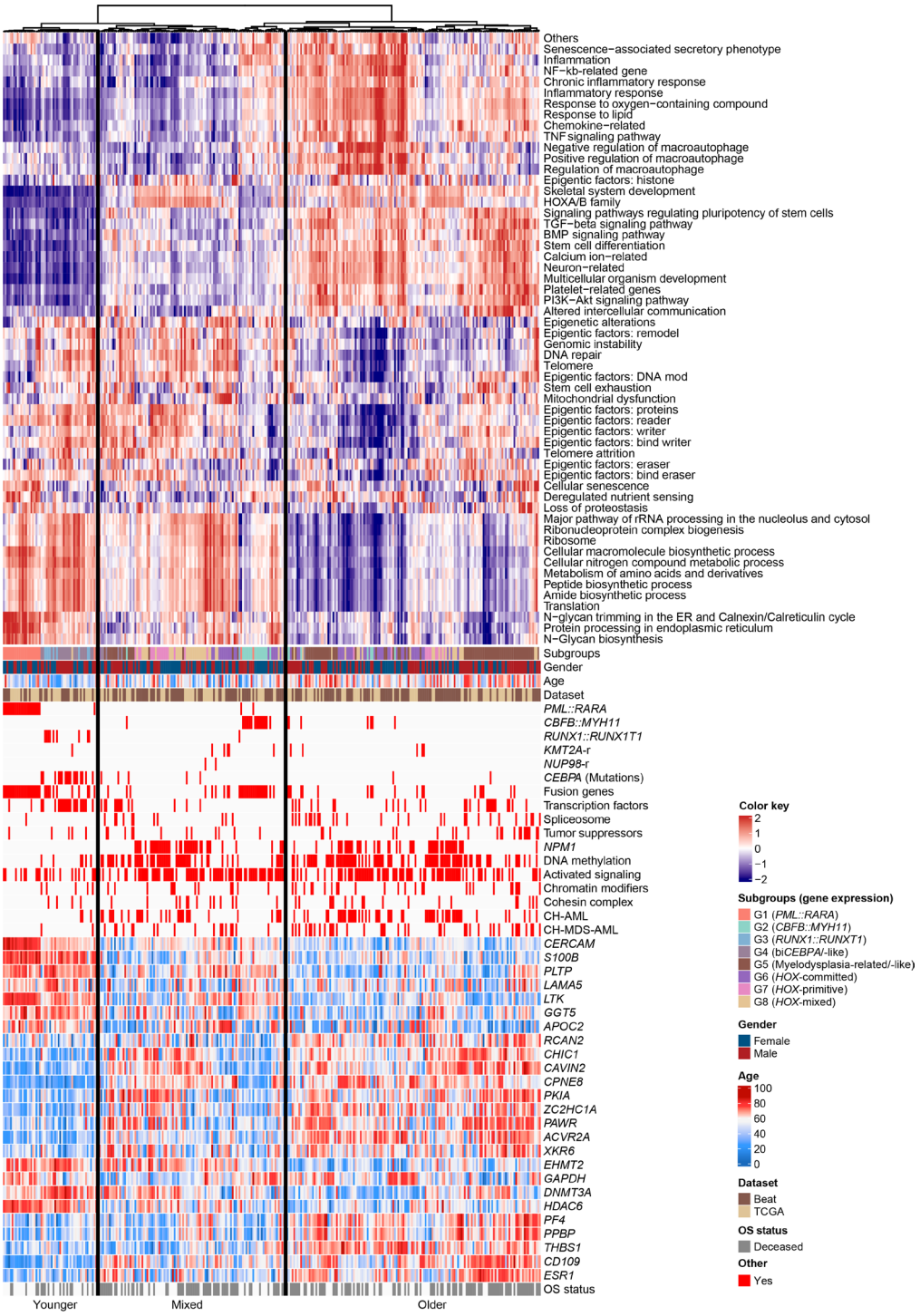

Fig. S10 Reproduction of pathways clustering and age-correlated genes in the pooled TCGA LAML and Beat AML cohorts. It is consistent with our finding that at least three types of ageing pathways can be found in the external cohorts. Inflammatory-, platelet- and other pathways highly correlated with age and prognosis in AML patients.

SI Tables

Table S1. Comparison of clinical and molecular features of different age groups in 1,474 patients with AML.

Variable	**Overall**, N = 1,474[1]	**<20**, N = 53[1]	**20-29**, N = 151[1]	**30-39**, N = 232[1]	**40-49**, N = 272[1]	**50-59**, N = 322[1]	**60-69**, N = 304[1]	**>=70**, N = 140[1]	*P*[2]
Age	50 (37 − 62)	16 (15 − 18)	25 (22 − 27)	34 (32 − 37)	45 (42 − 47)	55 (52 − 57)	65 (62 − 67)	74 (71 − 76)	<0.001
Gender									0.85
Female	695 (47)	24 (45)	67 (44)	113 (49)	134 (49)	158 (49)	136 (45)	63 (45)	
Male	779 (53)	29 (55)	84 (56)	119 (51)	138 (51)	164 (51)	168 (55)	77 (55)	
BM blasts	67 (46 − 84)	73 (54 − 81)	71 (48 − 87)	72 (52 − 85)	72 (53 − 86)	66 (46 − 81)	62 (38 − 80)	58 (40 − 79)	<0.001
NA	8	2	1	1	1	2	0	1	
WBC	13 (3 − 43)	32 (11 − 91)	13 (4 − 45)	20 (5 − 53)	15 (4 − 44)	12 (3 − 44)	8 (2 − 30)	7 (2 − 28)	<0.001
NA	54	2	6	7	6	10	14	9	
HGB	83 (67 − 103)	78 (70 − 103)	88 (70 − 107)	84 (67 − 107)	86 (67 − 106)	86 (70 − 106)	80 (64 − 97)	74 (60 − 93)	<0.001
NA	63	3	11	7	7	11	15	9	
PLT	41 (22 − 82)	32 (19 − 55)	34 (21 − 61)	33 (20 − 74)	42 (21 − 81)	50 (25 − 96)	45 (24 − 86)	47 (25 − 81)	<0.001
NA	59	2	10	7	6	11	14	9	
Diagnosis									

(Continued on next page…)

Variable	**Overall**, N = 1,474[1]	**<20**, N = 53[1]	**20-29**, N = 151[1]	**30-39**, N = 232[1]	**40-49**, N = 272[1]	**50-59**, N = 322[1]	**60-69**, N = 304[1]	**>=70**, N = 140[1]	*P*[2]
M1	64 (4.3)	2 (3.8)	10 (6.6)	10 (4.3)	18 (6.6)	13 (4.0)	7 (2.3)	4 (2.9)	
M2	245 (17)	9 (17)	25 (17)	47 (20)	47 (17)	58 (18)	38 (13)	21 (15)	
M3	123 (8.3)	1 (1.9)	30 (20)	30 (13)	31 (11)	16 (5.0)	14 (4.6)	1 (0.7)	
M4	398 (27)	15 (28)	36 (24)	59 (25)	73 (27)	98 (30)	80 (26)	37 (26)	
M5	388 (26)	16 (30)	28 (19)	47 (20)	66 (24)	93 (29)	90 (30)	48 (34)	
Others	256 (17)	10 (19)	22 (15)	39 (17)	37 (14)	44 (14)	75 (25)	29 (21)	
Normal karyotype	641 (46)	17 (33)	49 (34)	87 (40)	118 (46)	156 (51)	146 (50)	68 (51)	0.002
NA	74	2	7	16	14	14	14	7	
Complex karyotype	103 (7.4)	5 (9.8)	11 (7.6)	15 (6.9)	9 (3.5)	17 (5.5)	29 (10)	17 (13)	
NA	74	2	7	16	14	14	14	7	
Monosomal karyotype	95 (6.8)	3 (5.9)	9 (6.3)	12 (5.6)	8 (3.1)	17 (5.5)	30 (10)	16 (12)	
NA	74	2	7	16	14	14	14	7	
Trisomy8	84 (6.0)	3 (5.9)	5 (3.5)	14 (6.5)	12 (4.7)	13 (4.2)	21 (7.2)	16 (12)	
NA	74	2	7	16	14	14	14	7	
Minus5/5q	31 (2.2)	2 (3.9)	3 (2.1)	1 (0.5)	3 (1.2)	4 (1.3)	12 (4.1)	6 (4.5)	
NA	74	2	7	16	14	14	14	7	
Minus7/7q	46 (3.3)	3 (5.9)	5 (3.5)	3 (1.4)	6 (2.3)	9 (2.9)	14 (4.8)	6 (4.5)	
NA	74	2	7	16	14	14	14	7	

(Continued on next page…)

Variable	**Overall**, N = 1,474[1]	**<20**, N = 53[1]	**20-29**, N = 151[1]	**30-39**, N = 232[1]	**40-49**, N = 272[1]	**50-59**, N = 322[1]	**60-69**, N = 304[1]	**>=70**, N = 140[1]	*P*[2]
Minus17/abn17p	37 (2.6)	2 (3.9)	7 (4.9)	5 (2.3)	2 (0.8)	9 (2.9)	10 (3.4)	2 (1.5)	
NA	74	2	7	16	14	14	14	7	
PML::RARA	123 (8.3)	1 (1.9)	30 (20)	30 (13)	31 (11)	16 (5.0)	14 (4.6)	1 (0.7)	
CBFB::MYH11	112 (7.6)	5 (9.4)	16 (11)	21 (9.1)	24 (8.8)	24 (7.5)	16 (5.3)	6 (4.3)	
RUNX1::RUNX1T1	106 (7.2)	9 (17)	18 (12)	12 (5.2)	25 (9.2)	28 (8.7)	10 (3.3)	4 (2.9)	
***KMT2A*-r**	96 (6.5)	8 (15)	13 (8.6)	17 (7.3)	22 (8.1)	20 (6.2)	13 (4.3)	3 (2.1)	
***NUP98*-r**	50 (3.4)	2 (3.8)	7 (4.6)	9 (3.9)	10 (3.7)	13 (4.0)	6 (2.0)	3 (2.1)	0.63
***NUP214*-r**	8 (0.5)	0 (0)	2 (1.3)	1 (0.4)	1 (0.4)	4 (1.2)	0 (0)	0 (0)	0.29
***MECOM*-r**	6 (0.4)	1 (1.9)	0 (0)	1 (0.4)	1 (0.4)	1 (0.3)	2 (0.7)	0 (0)	0.60
BCR::ABL1	8 (0.5)	0 (0)	2 (1.3)	3 (1.3)	1 (0.4)	1 (0.3)	1 (0.3)	0 (0)	0.51
FUS::ERG	12 (0.8)	2 (3.8)	1 (0.7)	4 (1.7)	1 (0.4)	1 (0.3)	2 (0.7)	1 (0.7)	0.14
Other fusions	41 (2.8)	1 (1.9)	4 (2.6)	5 (2.2)	4 (1.5)	6 (1.9)	17 (5.6)	4 (2.9)	
Fusion genes	559 (38)	29 (55)	93 (62)	102 (44)	120 (44)	114 (35)	79 (26)	22 (16)	<0.001
CEBPA	267 (18)	11 (21)	19 (13)	52 (22)	47 (17)	67 (21)	49 (16)	22 (16)	0.16
RUNX1	135 (9.2)	0 (0)	7 (4.6)	13 (5.6)	18 (6.6)	29 (9.0)	38 (13)	30 (21)	
GATA2	91 (6.2)	2 (3.8)	7 (4.6)	25 (11)	19 (7.0)	25 (7.8)	8 (2.6)	5 (3.6)	
IKZF1	40 (2.7)	1 (1.9)	0 (0)	6 (2.6)	8 (2.9)	7 (2.2)	10 (3.3)	8 (5.7)	
ETV6	39 (2.6)	0 (0)	1 (0.7)	3 (1.3)	3 (1.1)	7 (2.2)	21 (6.9)	4 (2.9)	

(Continued on next page…)

Variable	**Overall**, N = 1,474[1]	**<20**, N = 53[1]	**20-29**, N = 151[1]	**30-39**, N = 232[1]	**40-49**, N = 272[1]	**50-59**, N = 322[1]	**60-69**, N = 304[1]	**>=70**, N = 140[1]	***P***[2]
Transcription factors	462 (31)	13 (25)	30 (20)	79 (34)	77 (28)	107 (33)	102 (34)	54 (39)	0.009
U2AF1	59 (4.0)	1 (1.9)	2 (1.3)	8 (3.4)	10 (3.7)	6 (1.9)	22 (7.2)	10 (7.1)	
SRSF2	45 (3.1)	0 (0)	1 (0.7)	1 (0.4)	3 (1.1)	5 (1.6)	16 (5.3)	19 (14)	
SF3B1	35 (2.4)	0 (0)	1 (0.7)	3 (1.3)	4 (1.5)	9 (2.8)	10 (3.3)	8 (5.7)	
ZRSR2	25 (1.7)	0 (0)	0 (0)	3 (1.3)	3 (1.1)	5 (1.6)	11 (3.6)	3 (2.1)	0.12
Spliceosome	161 (11)	1 (1.9)	4 (2.6)	14 (6.0)	19 (7.0)	25 (7.8)	59 (19)	39 (28)	<0.001
WT1	165 (11)	7 (13)	21 (14)	32 (14)	22 (8.1)	41 (13)	33 (11)	9 (6.4)	0.14
TP53	82 (5.6)	0 (0)	3 (2.0)	5 (2.2)	10 (3.7)	14 (4.3)	30 (9.9)	20 (14)	
PHF6	36 (2.4)	0 (0)	3 (2.0)	5 (2.2)	7 (2.6)	6 (1.9)	11 (3.6)	4 (2.9)	0.79
Tumor suppressors	276 (19)	7 (13)	27 (18)	41 (18)	38 (14)	58 (18)	72 (24)	33 (24)	0.051
NPM1	271 (18)	2 (3.8)	12 (7.9)	32 (14)	44 (16)	71 (22)	72 (24)	38 (27)	<0.001
DNMT3A	271 (18)	2 (3.8)	7 (4.6)	23 (9.9)	42 (15)	69 (21)	88 (29)	40 (29)	<0.001
TET2	207 (14)	2 (3.8)	8 (5.3)	14 (6.0)	33 (12)	43 (13)	65 (21)	42 (30)	<0.001
IDH2	151 (10)	2 (3.8)	9 (6.0)	8 (3.4)	15 (5.5)	39 (12)	57 (19)	21 (15)	<0.001
IDH1	114 (7.7)	0 (0)	4 (2.6)	9 (3.9)	19 (7.0)	27 (8.4)	32 (11)	23 (16)	
DNA methylation	573 (39)	6 (11)	25 (17)	47 (20)	86 (32)	136 (42)	183 (60)	90 (64)	<0.001
***FLT3*-ITD**	292 (20)	12 (23)	22 (15)	52 (22)	60 (22)	71 (22)	57 (19)	18 (13)	0.12

(Continued on next page…)

NRAS	206 (14)	5 (9.4)	21 (14)	33 (14)	38 (14)	47 (15)	38 (13)	24 (17)	0.84

Variable	**Overall**, N = 1,474[1]	**<20**, N = 53[1]	**20-29**, N = 151[1]	**30-39**, N = 232[1]	**40-49**, N = 272[1]	**50-59**, N = 322[1]	**60-69**, N = 304[1]	**>=70**, N = 140[1]	***P***[2]
FLT3	174 (12)	4 (7.5)	20 (13)	29 (13)	30 (11)	41 (13)	33 (11)	17 (12)	0.92
KIT	96 (6.5)	9 (17)	13 (8.6)	16 (6.9)	21 (7.7)	22 (6.8)	13 (4.3)	2 (1.4)	
KRAS	94 (6.4)	3 (5.7)	6 (4.0)	12 (5.2)	18 (6.6)	16 (5.0)	30 (9.9)	9 (6.4)	
PTPN11	98 (6.6)	0 (0)	6 (4.0)	18 (7.8)	20 (7.4)	17 (5.3)	22 (7.2)	15 (11)	
CSF3R	47 (3.2)	4 (7.5)	7 (4.6)	6 (2.6)	4 (1.5)	11 (3.4)	9 (3.0)	6 (4.3)	
Activated signaling	790 (54)	30 (57)	75 (50)	126 (54)	151 (56)	184 (57)	155 (51)	69 (49)	0.53
ASXL1	125 (8.5)	2 (3.8)	7 (4.6)	12 (5.2)	15 (5.5)	24 (7.5)	45 (15)	20 (14)	
BCOR	91 (6.2)	1 (1.9)	2 (1.3)	8 (3.4)	20 (7.4)	18 (5.6)	29 (9.5)	13 (9.3)	
***KMT2A*-PTD**	96 (6.5)	1 (1.9)	6 (4.0)	12 (5.2)	18 (6.6)	24 (7.5)	25 (8.2)	10 (7.1)	
EZH2	56 (3.8)	1 (1.9)	7 (4.6)	12 (5.2)	2 (0.7)	10 (3.1)	13 (4.3)	11 (7.9)	
EP300	32 (2.2)	0 (0)	3 (2.0)	6 (2.6)	7 (2.6)	8 (2.5)	5 (1.6)	3 (2.1)	0.96
BCORL1	38 (2.6)	0 (0)	1 (0.7)	1 (0.4)	6 (2.2)	12 (3.7)	12 (3.9)	6 (4.3)	
KDM6A	44 (3.0)	2 (3.8)	8 (5.3)	6 (2.6)	6 (2.2)	10 (3.1)	7 (2.3)	5 (3.6)	0.61
Chromatin Modifiers	386 (26)	5 (9.4)	31 (21)	50 (22)	64 (24)	84 (26)	103 (34)	49 (35)	<0.001
SMC1A	45 (3.1)	0 (0)	1 (0.7)	5 (2.2)	6 (2.2)	12 (3.7)	15 (4.9)	6 (4.3)	
SMC3	27 (1.8)	0 (0)	3 (2.0)	4 (1.7)	4 (1.5)	7 (2.2)	4 (1.3)	5 (3.6)	0.72
STAG2	50 (3.4)	2 (3.8)	0 (0)	7 (3.0)	4 (1.5)	11 (3.4)	20 (6.6)	6 (4.3)	
RAD21	32 (2.2)	0 (0)	2 (1.3)	5 (2.2)	5 (1.8)	13 (4.0)	4 (1.3)	3 (2.1)	0.36

(Continued on next page…)

Variable	**Overall**, N = 1,474[1]	**<20**, N = 53[1]	**20-29**, N = 151[1]	**30-39**, N = 232[1]	**40-49**, N = 272[1]	**50-59**, N = 322[1]	**60-69**, N = 304[1]	**>=70**, N = 140[1]	***P***[2]
Cohesin complex	148 (10)	2 (3.8)	6 (4.0)	20 (8.6)	18 (6.6)	41 (13)	43 (14)	18 (13)	0.001

[1]n (%); Median (IQR)

[2]Pearson's Chi-squared test; Kruskal-Wallis rank sum test; Fisher's exact test

Table S2. Comparison of clinical and molecular features in different age groups of the pooled TCGA LAML and Beat AML cohorts.

Variable	**Overall**, N = 315[1]	**<20**, N = 4[1]	**20-29**, N = 26[1]	**30-39**, N = 35[1]	**40-49**, N = 43[1]	**50-59**, N = 60[1]	**60-69**, N = 89[1]	**>=70**, N = 58[1]	***P***[2]
Age	58 (43, 66)	9 (7, 12)	24 (22, 26)	35 (33, 37)	45 (42, 47)	54 (51, 57)	63 (61, 66)	75 (73, 78)	<0.001
Gender									
Female	151 (48%)	3 (75%)	10 (38%)	22 (63%)	25 (58%)	28 (47%)	38 (43%)	25 (43%)	
Male	164 (52%)	1 (25%)	16 (62%)	13 (37%)	18 (42%)	32 (53%)	51 (57%)	33 (57%)	
BM blasts	75 (50, 86)	80 (68, 83)	72 (56, 90)	75 (52, 86)	75 (48, 86)	70 (48, 86)	76 (60, 85)	76 (40, 90)	>0.9
NA	7	0	1	2	1	1	2	0	
WBC	23 (5, 57)	32 (23, 44)	30 (11, 69)	19 (5, 45)	29 (11, 58)	23 (4, 47)	24 (5, 59)	14 (5, 59)	0.7
NA	15	0	1	1	2	3	3	5	
OS status	172 (55%)	0 (0%)	7 (27%)	15 (43%)	17 (40%)	33 (55%)	53 (60%)	47 (81%)	
PML::RARA	25 (7.9%)	0 (0%)	3 (12%)	5 (14%)	5 (12%)	5 (8.3%)	5 (5.6%)	2 (3.4%)	0.4
CBFB::MYH11	23 (7.3%)	0 (0%)	4 (15%)	5 (14%)	4 (9.3%)	7 (12%)	2 (2.2%)	1 (1.7%)	0.018
RUNX1::RUNX1T1	10 (3.2%)	1 (25%)	1 (3.8%)	4 (11%)	2 (4.7%)	1 (1.7%)	0 (0%)	1 (1.7%)	0.005
***KMT2A*-r**	11 (3.5%)	1 (25%)	0 (0%)	2 (5.7%)	1 (2.3%)	3 (5.0%)	4 (4.5%)	0 (0%)	0.14
***NUP98*-r**	2 (0.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2.2%)	0 (0%)	0.8
MLLT10::PICALM	1 (0.3%)	0 (0%)	0 (0%)	1 (2.9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0.2
BCR::ABL1	4 (1.3%)	0 (0%)	1 (3.8%)	0 (0%)	0 (0%)	0 (0%)	1 (1.1%)	2 (3.4%)	0.4
***MECOM*-r**	4 (1.3%)	0 (0%)	0 (0%)	1 (2.9%)	1 (2.3%)	1 (1.7%)	1 (1.1%)	0 (0%)	0.8

(Continued on next page…)

Variable	**Overall**, N = 315[1]	**<20**, N = 4[1]	**20-29**, N = 26[1]	**30-39**, N = 35[1]	**40-49**, N = 43[1]	**50-59**, N = 60[1]	**60-69**, N = 89[1]	**>=70**, N = 58[1]	***P***[2]
Gene fusions	80 (25%)	2 (50%)	9 (35%)	18 (51%)	13 (30%)	17 (28%)	15 (17%)	6 (10%)	
RUNX1	29 (9.2%)	0 (0%)	2 (7.7%)	1 (2.9%)	1 (2.3%)	6 (10%)	8 (9.0%)	11 (19%)	0.11
CEBPA	24 (7.6%)	1 (25%)	4 (15%)	5 (14%)	5 (12%)	3 (5.0%)	5 (5.6%)	1 (1.7%)	0.045
GATA2	11 (3.5%)	1 (25%)	1 (3.8%)	2 (5.7%)	1 (2.3%)	3 (5.0%)	2 (2.2%)	1 (1.7%)	0.3
ETV6	3 (1.0%)	0 (0%)	0 (0%)	0 (0%)	1 (2.3%)	1 (1.7%)	0 (0%)	1 (1.7%)	0.6
IKZF1	2 (0.6%)	0 (0%)	0 (0%)	0 (0%)	1 (2.3%)	0 (0%)	0 (0%)	1 (1.7%)	0.4
Transcription factors	61 (19%)	1 (25%)	7 (27%)	6 (17%)	8 (19%)	12 (20%)	15 (17%)	12 (21%)	>0.9
SRSF2	16 (5.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (3.3%)	4 (4.5%)	10 (17%)	0.004
U2AF1	13 (4.1%)	0 (0%)	1 (3.8%)	0 (0%)	0 (0%)	2 (3.3%)	4 (4.5%)	6 (10%)	0.2
SF3B1	8 (2.5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (3.3%)	4 (4.5%)	2 (3.4%)	0.7
ZRSR2	3 (1.0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2.2%)	1 (1.7%)	0.8
Spliceosome	39 (12%)	0 (0%)	1 (3.8%)	0 (0%)	0 (0%)	6 (10%)	13 (15%)	19 (33%)	
TP53	19 (6.0%)	0 (0%)	0 (0%)	0 (0%)	1 (2.3%)	2 (3.3%)	9 (10%)	7 (12%)	0.077
WT1	15 (4.8%)	0 (0%)	2 (7.7%)	0 (0%)	3 (7.0%)	7 (12%)	3 (3.4%)	0 (0%)	0.038
PHF6	6 (1.9%)	0 (0%)	1 (3.8%)	0 (0%)	0 (0%)	2 (3.3%)	0 (0%)	3 (5.2%)	0.2
Tumor suppressors	38 (12%)	0 (0%)	2 (7.7%)	0 (0%)	4 (9.3%)	10 (17%)	12 (13%)	10 (17%)	0.12
NPM1	84 (27%)	0 (0%)	4 (15%)	8 (23%)	9 (21%)	19 (32%)	27 (30%)	17 (29%)	0.5
DNMT3A	79 (25%)	0 (0%)	1 (3.8%)	4 (11%)	10 (23%)	19 (32%)	29 (33%)	16 (28%)	

(Continued on next page…)

Variable	**Overall**, N = 315[1]	**<20**, N = 4[1]	**20-29**, N = 26[1]	**30-39**, N = 35[1]	**40-49**, N = 43[1]	**50-59**, N = 60[1]	**60-69**, N = 89[1]	**>=70**, N = 58[1]	***P***[2]
IDH2	41 (13%)	0 (0%)	1 (3.8%)	0 (0%)	4 (9.3%)	6 (10%)	19 (21%)	11 (19%)	
TET2	34 (11%)	0 (0%)	0 (0%)	2 (5.7%)	4 (9.3%)	3 (5.0%)	11 (12%)	14 (24%)	
IDH1	28 (8.9%)	0 (0%)	2 (7.7%)	3 (8.6%)	3 (7.0%)	5 (8.3%)	9 (10%)	6 (10%)	>0.9
DNA methylation	139 (44%)	0 (0%)	4 (15%)	8 (23%)	16 (37%)	26 (43%)	47 (53%)	38 (66%)	
***FLT3*-ITD**	76 (24%)	0 (0%)	6 (23%)	10 (29%)	9 (21%)	15 (25%)	23 (26%)	13 (22%)	>0.9
NRAS	38 (12%)	2 (50%)	3 (12%)	8 (23%)	4 (9.3%)	5 (8.3%)	11 (12%)	5 (8.6%)	0.13
FLT3	26 (8.3%)	1 (25%)	1 (3.8%)	2 (5.7%)	3 (7.0%)	6 (10%)	8 (9.0%)	5 (8.6%)	0.8
PTPN11	18 (5.7%)	0 (0%)	1 (3.8%)	1 (2.9%)	2 (4.7%)	6 (10%)	5 (5.6%)	3 (5.2%)	0.9
KRAS	14 (4.4%)	0 (0%)	2 (7.7%)	2 (5.7%)	2 (4.7%)	2 (3.3%)	1 (1.1%)	5 (8.6%)	0.3
KIT	10 (3.2%)	0 (0%)	1 (3.8%)	3 (8.6%)	2 (4.7%)	3 (5.0%)	0 (0%)	1 (1.7%)	0.10
CSF3R	3 (1.0%)	1 (25%)	2 (7.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	<0.001
Activated signaling	169 (54%)	4 (100%)	15 (58%)	22 (63%)	21 (49%)	31 (52%)	45 (51%)	31 (53%)	0.5
ASXL1	15 (4.8%)	0 (0%)	0 (0%)	1 (2.9%)	0 (0%)	2 (3.3%)	9 (10%)	3 (5.2%)	0.2
***KMT2A*-PTD**	11 (3.5%)	1 (25%)	2 (7.7%)	2 (5.7%)	0 (0%)	1 (1.7%)	5 (5.6%)	0 (0%)	0.034
BCOR	10 (3.2%)	0 (0%)	0 (0%)	3 (8.6%)	0 (0%)	1 (1.7%)	4 (4.5%)	2 (3.4%)	0.4
EZH2	5 (1.6%)	0 (0%)	2 (7.7%)	1 (2.9%)	1 (2.3%)	0 (0%)	1 (1.1%)	0 (0%)	0.14
BCORL1	4 (1.3%)	0 (0%)	1 (3.8%)	1 (2.9%)	0 (0%)	0 (0%)	2 (2.2%)	0 (0%)	0.4
KDM6A	3 (1.0%)	0 (0%)	0 (0%)	0 (0%)	1 (2.3%)	0 (0%)	1 (1.1%)	1 (1.7%)	0.8

(Continued on next page…)

Variable	**Overall**, N = 315[1]	**<20**, N = 4[1]	**20-29**, N = 26[1]	**30-39**, N = 35[1]	**40-49**, N = 43[1]	**50-59**, N = 60[1]	**60-69**, N = 89[1]	**>=70**, N = 58[1]	***P***[2]
EP300	2 (0.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2.2%)	0 (0%)	0.8
Chromatin modifiers	43 (14%)	1 (25%)	4 (15%)	7 (20%)	2 (4.7%)	4 (6.7%)	19 (21%)	6 (10%)	
STAG2	15 (4.8%)	1 (25%)	1 (3.8%)	0 (0%)	0 (0%)	2 (3.3%)	5 (5.6%)	6 (10%)	0.064
SMC1A	9 (2.9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (8.3%)	2 (2.2%)	2 (3.4%)	0.2
SMC3	9 (2.9%)	0 (0%)	0 (0%)	1 (2.9%)	0 (0%)	2 (3.3%)	3 (3.4%)	3 (5.2%)	0.8
RAD21	4 (1.3%)	0 (0%)	1 (3.8%)	1 (2.9%)	1 (2.3%)	0 (0%)	1 (1.1%)	0 (0%)	0.3
Cohesin complex	35 (11%)	1 (25%)	2 (7.7%)	2 (5.7%)	1 (2.3%)	9 (15%)	10 (11%)	10 (17%)	0.14

[1]n (%); Median (IQR)

[2]Kruskal-Wallis rank sum test; Fisher's exact test

Table S3. Comparison of clinical and molecular features in clonal hematopoiesis groups of 915 gene fusion-negative patients with AML.

Variable	**Overall**, N = 915[1]	**CH-AML**, N = 257[1]	**CH-MDS-AML**, N = 266[1]	**Other GF-**, N = 392[1]	***P***[2]
Age	54 (41 − 65)	56 (46 − 66)	61 (49 − 68)	47 (34 − 58)	<0.001
Gender					<0.001
Female	415 (45)	141 (55)	101 (38)	173 (44)	
Male	500 (55)	116 (45)	165 (62)	219 (56)	
BM blasts	65 (44 − 81)	68 (48 − 81)	55 (35 − 74)	68 (48 − 84)	<0.001
NA	4	1	0	3	
WBC	13 (3 − 46)	20 (5 − 56)	6 (2 − 34)	15 (4 − 43)	<0.001
NA	29	4	14	11	
HGB	84 (67 − 102)	85 (68 − 106)	78 (65 − 93)	87 (70 − 105)	<0.001
NA	35	5	14	16	
PLT	47 (23 − 93)	63 (29 − 108)	49 (25 − 93)	40 (20 − 85)	<0.001
NA	32	4	14	14	
Diagnosis					<0.001
M1	58 (6.3)	9 (3.5)	6 (2.3)	43 (11)	
M2	153 (17)	37 (14)	35 (13)	81 (21)	
M4	248 (27)	80 (31)	57 (21)	111 (28)	
M5	266 (29)	92 (36)	101 (38)	73 (19)	
Others	190 (21)	39 (15)	67 (25)	84 (21)	
Normal karyotype	548 (63)	169 (72)	154 (60)	225 (60)	0.010
NA	52	21	11	20	
Complex karyotype	80 (9.3)	18 (7.6)	22 (8.6)	40 (11)	0.40
NA	52	21	11	20	
Monosomal karyotype	76 (8.8)	18 (7.6)	21 (8.2)	37 (9.9)	0.57
NA	52	21	11	20	
Trisomy8	55 (6.4)	13 (5.5)	25 (9.8)	17 (4.6)	0.025
NA	52	21	11	20	

(Continued on next page…)

Variable	**Overall**, N = 915[1]	**CH-AML**, N = 257[1]	**CH-MDS-AML**, N = 266[1]	**Other GF-**, N = 392[1]	***P***[2]
Minus5/5q	24 (2.8)	4 (1.7)	9 (3.5)	11 (3.0)	0.45
NA	52	21	11	20	
Minus7/7q	32 (3.7)	6 (2.5)	12 (4.7)	14 (3.8)	0.45
NA	52	21	11	20	
Minus17/abn17p	30 (3.5)	6 (2.5)	5 (2.0)	19 (5.1)	0.070
NA	52	21	11	20	
CEBPA	252 (28)	49 (19)	50 (19)	153 (39)	<0.001
RUNX1	114 (12)	16 (6.2)	72 (27)	26 (6.6)	<0.001
GATA2	77 (8.4)	12 (4.7)	13 (4.9)	52 (13)	<0.001
IKZF1	36 (3.9)	5 (1.9)	11 (4.1)	20 (5.1)	0.13
ETV6	27 (3.0)	6 (2.3)	16 (6.0)	5 (1.3)	0.002
Transcription factors	401 (44)	78 (30)	128 (48)	195 (50)	<0.001
U2AF1	46 (5.0)	0 (0)	46 (17)	0 (0)	<0.001
SRSF2	41 (4.5)	0 (0)	41 (15)	0 (0)	<0.001
SF3B1	30 (3.3)	0 (0)	30 (11)	0 (0)	<0.001
ZRSR2	17 (1.9)	0 (0)	17 (6.4)	0 (0)	<0.001
Spliceosome	131 (14)	0 (0)	131 (49)	0 (0)	<0.001
WT1	101 (11)	22 (8.6)	20 (7.5)	59 (15)	0.003
TP53	64 (7.0)	13 (5.1)	22 (8.3)	29 (7.4)	0.33
PHF6	30 (3.3)	4 (1.6)	16 (6.0)	10 (2.6)	0.009
Tumor suppressors	189 (21)	38 (15)	56 (21)	95 (24)	0.014
NPM1	266 (29)	141 (55)	45 (17)	80 (20)	<0.001
DNMT3A	248 (27)	189 (74)	59 (22)	0 (0)	<0.001
TET2	171 (19)	106 (41)	65 (24)	0 (0)	<0.001
IDH2	137 (15)	41 (16)	49 (18)	47 (12)	0.067
IDH1	99 (11)	26 (10)	37 (14)	36 (9.2)	0.15
DNA methylation	496 (54)	257 (100)	158 (59)	81 (21)	<0.001
***FLT3*-ITD**	212 (23)	95 (37)	38 (14)	79 (20)	<0.001

(Continued on next page…)

Variable	**Overall**, N = 915[1]	**CH-AML**, N = 257[1]	**CH-MDS-AML**, N = 266[1]	**Other GF-**, N = 392[1]	***P***[2]
NRAS	113 (12)	20 (7.8)	47 (18)	46 (12)	0.002
FLT3	87 (9.5)	38 (15)	24 (9.0)	25 (6.4)	0.002
KIT	18 (2.0)	5 (1.9)	0 (0)	13 (3.3)	0.011
KRAS	41 (4.5)	13 (5.1)	19 (7.1)	9 (2.3)	0.011
PTPN11	74 (8.1)	29 (11)	21 (7.9)	24 (6.1)	0.061
CSF3R	32 (3.5)	6 (2.3)	15 (5.6)	11 (2.8)	0.074
Activated signaling	472 (52)	168 (65)	133 (50)	171 (44)	<0.001
ASXL1	90 (9.8)	0 (0)	90 (34)	0 (0)	<0.001
BCOR	75 (8.2)	0 (0)	75 (28)	0 (0)	<0.001
***KMT2A*-PTD**	78 (8.5)	30 (12)	22 (8.3)	26 (6.6)	0.079
EZH2	41 (4.5)	0 (0)	41 (15)	0 (0)	<0.001
EP300	24 (2.6)	6 (2.3)	6 (2.3)	12 (3.1)	0.77
BCORL1	28 (3.1)	3 (1.2)	20 (7.5)	5 (1.3)	<0.001
KDM6A	16 (1.7)	5 (1.9)	6 (2.3)	5 (1.3)	0.62
Chromatin modifiers	282 (31)	44 (17)	192 (72)	46 (12)	<0.001
SMC1A	24 (2.6)	12 (4.7)	3 (1.1)	9 (2.3)	0.035
SMC3	18 (2.0)	6 (2.3)	4 (1.5)	8 (2.0)	0.78
STAG2	40 (4.4)	0 (0)	40 (15)	0 (0)	<0.001
RAD21	21 (2.3)	3 (1.2)	6 (2.3)	12 (3.1)	0.29
Cohesin complex	101 (11)	21 (8.2)	52 (20)	28 (7.1)	<0.001
Age groups					<0.001
<20	24 (2.6)	4 (1.6)	1 (0.4)	19 (4.8)	
20-29	58 (6.3)	6 (2.3)	11 (4.1)	41 (10)	
30-39	130 (14)	23 (8.9)	28 (11)	79 (20)	
40-49	152 (17)	49 (19)	27 (10)	76 (19)	
50-59	208 (23)	66 (26)	53 (20)	89 (23)	
60-69	225 (25)	69 (27)	95 (36)	61 (16)	
>=70	118 (13)	40 (16)	51 (19)	27 (6.9)	

[1]n (%); Median (IQR)

Variable	**Overall**, N = 915[1]	**CH-AML**, N = 257[1]	**CH-MDS-AML**, N = 266[1]	**Other GF-**, N = 392[1]	***P***[2]

[2]Pearson's Chi-squared test; Kruskal-Wallis rank sum test; Fisher's exact test

Table S4. Comparison of clinical and molecular features in clonal hematopoiesis groups of the pooled gene fusion-negative patients from TCGA LAML and Beat AML cohorts.

Variable	**Overall**, N = 235[1]	**CH-AML**, N = 82[1]	**CH-MDS-AML**, N = 59[1]	**Other GF-**, N = 94[1]	***P***[2]
Age	61 (47 − 67)	61 (51 − 66)	65 (56 − 76)	54 (39 − 63)	<0.001
BM blasts	75 (47 − 86)	76 (48 − 87)	68 (35 − 80)	75 (56 − 88)	0.091
NA	6	1	1	4	
WBC	26 (5 − 61)	46 (8 − 75)	15 (4 − 56)	21 (5 − 43)	0.011
NA	13	5	7	1	
RUNX1	29 (12)	1 (1.2)	18 (31)	10 (11)	<0.001
CEBPA	24 (10)	5 (6.1)	3 (5.1)	16 (17)	0.019
GATA2	10 (4.3)	2 (2.4)	2 (3.4)	6 (6.4)	0.47
ETV6	2 (0.9)	1 (1.2)	1 (1.7)	0 (0)	0.52
IKZF1	1 (0.4)	0 (0)	1 (1.7)	0 (0)	0.25
Transcription factors	58 (25)	8 (9.8)	22 (37)	28 (30)	<0.001
SRSF2	16 (6.8)	0 (0)	16 (27)	0 (0)	<0.001
U2AF1	13 (5.5)	0 (0)	13 (22)	0 (0)	<0.001
SF3B1	6 (2.6)	0 (0)	6 (10)	0 (0)	<0.001
ZRSR2	3 (1.3)	0 (0)	3 (5.1)	0 (0)	0.015
Spliceosome	37 (16)	0 (0)	37 (63)	0 (0)	<0.001
TP53	18 (7.7)	2 (2.4)	3 (5.1)	13 (14)	0.017
WT1	11 (4.7)	5 (6.1)	0 (0)	6 (6.4)	0.12
PHF6	5 (2.1)	1 (1.2)	2 (3.4)	2 (2.1)	0.74
Tumor suppressors	32 (14)	8 (9.8)	5 (8.5)	19 (20)	0.054
NPM1	84 (36)	44 (54)	8 (14)	32 (34)	<0.001
DNMT3A	77 (33)	69 (84)	8 (14)	0 (0)	<0.001
IDH2	41 (17)	13 (16)	12 (20)	16 (17)	0.78
TET2	32 (14)	23 (28)	9 (15)	0 (0)	<0.001
IDH1	26 (11)	14 (17)	6 (10)	6 (6.4)	0.076
DNA methylation	134 (57)	82 (100)	30 (51)	22 (23)	<0.001

(Continued on next page…)

Variable	**Overall**, N = 235[1]	**CH-AML**, N = 82[1]	**CH-MDS-AML**, N = 59[1]	**Other GF-**, N = 94[1]	***P***[2]
***FLT3*-ITD**	59 (25)	26 (32)	9 (15)	24 (26)	0.084
NRAS	25 (11)	7 (8.5)	10 (17)	8 (8.5)	0.19
FLT3	24 (10)	7 (8.5)	7 (12)	10 (11)	0.80
PTPN11	17 (7.2)	8 (9.8)	3 (5.1)	6 (6.4)	0.57
KRAS	11 (4.7)	5 (6.1)	4 (6.8)	2 (2.1)	0.28
KIT	2 (0.9)	0 (0)	1 (1.7)	1 (1.1)	0.72
CSF3R	2 (0.9)	0 (0)	1 (1.7)	1 (1.1)	0.72
Activated signaling	127 (54)	50 (61)	29 (49)	48 (51)	0.29
ASXL1	14 (6.0)	0 (0)	14 (24)	0 (0)	<0.001
***KMT2A*-PTD**	11 (4.7)	3 (3.7)	2 (3.4)	6 (6.4)	0.73
BCOR	8 (3.4)	0 (0)	8 (14)	0 (0)	<0.001
EZH2	3 (1.3)	0 (0)	3 (5.1)	0 (0)	0.015
BCORL1	4 (1.7)	0 (0)	3 (5.1)	1 (1.1)	0.089
KDM6A	2 (0.9)	1 (1.2)	1 (1.7)	0 (0)	0.52
EP300	2 (0.9)	1 (1.2)	1 (1.7)	0 (0)	0.52
Chromatin modifiers	37 (16)	5 (6.1)	25 (42)	7 (7.4)	<0.001
STAG2	15 (6.4)	0 (0)	15 (25)	0 (0)	<0.001
SMC1A	7 (3.0)	3 (3.7)	1 (1.7)	3 (3.2)	0.89
SMC3	9 (3.8)	7 (8.5)	2 (3.4)	0 (0)	0.006
RAD21	3 (1.3)	2 (2.4)	0 (0)	1 (1.1)	0.62
Cohesin complex	32 (14)	12 (15)	16 (27)	4 (4.3)	<0.001
Age groups					
<20	2 (0.9)	0 (0)	1 (1.7)	1 (1.1)	
20-29	17 (7.2)	1 (1.2)	4 (6.8)	12 (13)	
30-39	17 (7.2)	4 (4.9)	2 (3.4)	11 (12)	
40-49	30 (13)	13 (16)	1 (1.7)	16 (17)	
50-59	43 (18)	19 (23)	8 (14)	16 (17)	
60-69	74 (31)	30 (37)	19 (32)	25 (27)	
>=70	52 (22)	15 (18)	24 (41)	13 (14)	

Variable	**Overall**, N = 235[1]	**CH-AML**, N = 82[1]	**CH-MDS-AML**, N = 59[1]	**Other GF-**, N = 94[1]	***P***[2]

[1]Median (IQR); n (%)

[2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test; Fisher's exact test

Table S5. Comparison of clinical and molecular features in gender groups of 1,474 patients with AML.

Variable	**Female**, N = 695[1]	**Male**, N = 779[1]	*P*[2]
Age	50 (37 − 61)	51 (36 − 63)	0.85
BM blasts	67 (45 − 83)	68 (47 − 84)	0.32
NA	3	5	
WBC	11 (3 − 39)	14 (4 − 46)	0.037
NA	25	29	
HGB	80 (66 − 97)	86 (67 − 110)	<0.001
NA	27	36	
PLT	46 (25 − 83)	37 (20 − 79)	<0.001
NA	27	32	
Diagnosis			0.085
M1	19 (2.7)	45 (5.8)	
M2	124 (18)	121 (16)	
M3	55 (7.9)	68 (8.7)	
M4	193 (28)	205 (26)	
M5	185 (27)	203 (26)	
Others	119 (17)	137 (18)	
Normal karyotype	306 (47)	335 (45)	0.54
NA	39	35	
Complex karyotype	41 (6.3)	62 (8.3)	0.14
NA	39	35	
Monosomal karyotype	39 (5.9)	56 (7.5)	0.24
NA	39	35	
Trisomy8	41 (6.3)	43 (5.8)	0.71
NA	39	35	
Minus5/5q	15 (2.3)	16 (2.2)	0.86
NA	39	35	
Minus7/7q	18 (2.7)	28 (3.8)	0.29
NA	39	35	

(Continued on next page…)

Variable	**Female**, N = 695[1]	**Male**, N = 779[1]	***P***[2]
Minus17/abn17p	8 (1.2)	29 (3.9)	0.002
NA	39	35	
PML::RARA	55 (7.9)	68 (8.7)	0.57
CBFB::MYH11	48 (6.9)	64 (8.2)	0.34
RUNX1::RUNX1T1	58 (8.3)	48 (6.2)	0.11
***KMT2A*-r**	58 (8.3)	38 (4.9)	0.007
***NUP98*-r**	29 (4.2)	21 (2.7)	0.12
***NUP214*-r**	5 (0.7)	3 (0.4)	0.49
***MECOM*-r**	3 (0.4)	3 (0.4)	>0.99
BCR::ABL1	6 (0.9)	2 (0.3)	0.16
FUS::ERG	5 (0.7)	7 (0.9)	0.70
Other fusions	15 (2.2)	26 (3.3)	0.17
Fusion genes	280 (40)	279 (36)	0.077
CEBPA	103 (15)	164 (21)	0.002
RUNX1	52 (7.5)	83 (11)	0.035
GATA2	36 (5.2)	55 (7.1)	0.13
IKZF1	22 (3.2)	18 (2.3)	0.31
ETV6	21 (3.0)	18 (2.3)	0.40
Transcription factors	197 (28)	265 (34)	0.019
U2AF1	15 (2.2)	44 (5.6)	<0.001
SRSF2	14 (2.0)	31 (4.0)	0.029
SF3B1	15 (2.2)	20 (2.6)	0.61
ZRSR2	5 (0.7)	20 (2.6)	0.006
Spliceosome	49 (7.1)	112 (14)	<0.001
WT1	87 (13)	78 (10)	0.13
TP53	35 (5.0)	47 (6.0)	0.40
PHF6	8 (1.2)	28 (3.6)	0.002
Tumor suppressors	127 (18)	149 (19)	0.67
NPM1	159 (23)	112 (14)	<0.001
DNMT3A	149 (21)	122 (16)	0.004

(Continued on next page…)

Variable	**Female**, N = 695[1]	**Male**, N = 779[1]	***P***[2]
TET2	89 (13)	118 (15)	0.20
IDH2	68 (9.8)	83 (11)	0.58
IDH1	59 (8.5)	55 (7.1)	0.31
DNA methylation	280 (40)	293 (38)	0.29
***FLT3*-ITD**	155 (22)	137 (18)	0.023
NRAS	93 (13)	113 (15)	0.53
FLT3	83 (12)	91 (12)	0.88
KIT	47 (6.8)	49 (6.3)	0.71
KRAS	49 (7.1)	45 (5.8)	0.32
PTPN11	52 (7.5)	46 (5.9)	0.23
CSF3R	19 (2.7)	28 (3.6)	0.35
Activated signaling	394 (57)	396 (51)	0.024
ASXL1	47 (6.8)	78 (10)	0.025
BCOR	48 (6.9)	43 (5.5)	0.27
***KMT2A*-PTD**	49 (7.1)	47 (6.0)	0.43
EZH2	16 (2.3)	40 (5.1)	0.005
EP300	7 (1.0)	25 (3.2)	0.004
BCORL1	18 (2.6)	20 (2.6)	0.98
KDM6A	28 (4.0)	16 (2.1)	0.026
Chromatin modifiers	176 (25)	210 (27)	0.48
SMC1A	18 (2.6)	27 (3.5)	0.33
SMC3	15 (2.2)	12 (1.5)	0.38
STAG2	25 (3.6)	25 (3.2)	0.68
RAD21	15 (2.2)	17 (2.2)	0.97
Cohesin complex	68 (9.8)	80 (10)	0.76
Age groups			0.85
<20	24 (3.5)	29 (3.7)	
20-29	67 (9.6)	84 (11)	
30-39	113 (16)	119 (15)	
40-49	134 (19)	138 (18)	

(Continued on next page…)

Variable	**Female**, N = 695[1]	**Male**, N = 779[1]	***P***[2]
50-59	158 (23)	164 (21)	
60-69	136 (20)	168 (22)	
>=70	63 (9.1)	77 (9.9)	
CH groups			<0.001
GF+	280 (40)	279 (36)	
CH-AML	141 (20)	116 (15)	
CH-MDS-AML	101 (15)	165 (21)	
Other GF-	173 (25)	219 (28)	

[1]Median (IQR); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Table S6. Comparison of clinical and molecular features in gender groups of the pooled TCGA LAML and Beat AML cohorts.

Variable	**Female**, N = 151[1]	**Male**, N = 164[1]	*P*[2]
Age	54 (41 − 65)	60 (47 − 67)	0.085
BM blasts	74 (51 − 86)	75 (50 − 88)	0.91
NA	3	4	
WBC	28 (8 − 56)	19 (5 − 59)	0.15
NA	6	9	
PML::RARA	14 (9.3)	11 (6.7)	0.40
CBFB::MYH11	12 (7.9)	11 (6.7)	0.67
RUNX1::RUNX1T1	5 (3.3)	5 (3.0)	>0.99
***KMT2A*-r**	3 (2.0)	8 (4.9)	0.16
***NUP98*-r**	1 (0.7)	1 (0.6)	>0.99
MLLT10::PICALM	0 (0)	1 (0.6)	>0.99
BCR::ABL1	1 (0.7)	3 (1.8)	0.62
***MECOM*-r**	2 (1.3)	2 (1.2)	>0.99
Fusion genes	38 (25)	42 (26)	0.93
RUNX1	9 (6.0)	20 (12)	0.056
CEBPA	8 (5.3)	16 (9.8)	0.14
GATA2	6 (4.0)	5 (3.0)	0.66
ETV6	1 (0.7)	2 (1.2)	>0.99
IKZF1	2 (1.3)	0 (0)	0.23
Transcription factors	21 (14)	40 (24)	0.019
SRSF2	1 (0.7)	15 (9.1)	<0.001
U2AF1	4 (2.6)	9 (5.5)	0.21
SF3B1	4 (2.6)	4 (2.4)	>0.99
ZRSR2	0 (0)	3 (1.8)	0.25
Spliceosome	9 (6.0)	30 (18)	<0.001
TP53	5 (3.3)	14 (8.5)	0.052
WT1	8 (5.3)	7 (4.3)	0.67
PHF6	3 (2.0)	3 (1.8)	>0.99

(Continued on next page…)

Variable	**Female**, N = 151[1]	**Male**, N = 164[1]	***P***[2]
Tumor suppressors	16 (11)	22 (13)	0.44
NPM1	52 (34)	32 (20)	0.003
DNMT3A	53 (35)	26 (16)	<0.001
IDH2	21 (14)	20 (12)	0.65
TET2	15 (9.9)	19 (12)	0.64
IDH1	14 (9.3)	14 (8.5)	0.82
DNA methylation	77 (51)	62 (38)	0.019
***FLT3*-ITD**	44 (29)	32 (20)	0.046
NRAS	21 (14)	17 (10)	0.34
FLT3	15 (9.9)	11 (6.7)	0.30
PTPN11	9 (6.0)	9 (5.5)	0.86
KRAS	8 (5.3)	6 (3.7)	0.48
KIT	6 (4.0)	4 (2.4)	0.53
CSF3R	0 (0)	3 (1.8)	0.25
Activated signaling	94 (62)	75 (46)	0.003
ASXL1	3 (2.0)	12 (7.3)	0.026
KMT2A-PTD	5 (3.3)	6 (3.7)	0.87
BCOR	4 (2.6)	6 (3.7)	0.75
EZH2	1 (0.7)	4 (2.4)	0.37
BCORL1	1 (0.7)	3 (1.8)	0.62
KDM6A	2 (1.3)	1 (0.6)	0.61
EP300	1 (0.7)	1 (0.6)	>0.99
Chromatin modifiers	16 (11)	27 (16)	0.13
STAG2	5 (3.3)	10 (6.1)	0.25
SMC1A	4 (2.6)	5 (3.0)	>0.99
SMC3	3 (2.0)	6 (3.7)	0.50
RAD21	3 (2.0)	1 (0.6)	0.35
Cohesin complex	15 (9.9)	20 (12)	0.52
Age groups			
<20	3 (2.0)	1 (0.6)	

(Continued on next page…)

Variable	**Female**, N = 151[1]	**Male**, N = 164[1]	***P***[2]
20-29	10 (6.6)	16 (9.8)	
30-39	22 (15)	13 (7.9)	
40-49	25 (17)	18 (11)	
50-59	28 (19)	32 (20)	
60-69	38 (25)	51 (31)	
>=70	25 (17)	33 (20)	
CH types			<0.001
GF+	38 (25)	42 (26)	
CH-AML	54 (36)	28 (17)	
CH-MDS-AML	18 (12)	41 (25)	
Other GF-	41 (27)	53 (32)	

[1]Median (IQR); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

SI Datasets

Dataset S1. Clinical information of 1,474 primary AML.

Dataset S2. Mutation list of 35 common gene terms in AML.

Dataset S3. Overall age correlation of gene expression markers and the pathways analysis of AML patients.

Dataset S4. Differentially expressed genes analysis of age groups.

Dataset S5. Gene sets enrichment analysis including aging hallmarks, epigenetic factors and age-correlated pathways.

Dataset S6. Clinical and molecular information of merged Beat AML and TCGA LAML cohorts.

Dataset S7. Potential genes related to the occurrence of gene fusions and genetic mutations.

Dataset S9. Gene expression correlations with age and its pathways analysis in gender groups.

Dataset S10. Differentially expressed genes of gene fusion negative patients with AML.

Dataset S11. Immune cells deconvolution and the enrichment scores of AML patients.

SI References

1. S. Uhrig *et al.*, Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome research* **31**, 448-460 (2021).
2. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).
3. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
4. A. M. Newman *et al.*, Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782 (2019).
5. Z. Gu, M. Schlesner, D. Hubschmann, cola: an R/Bioconductor package for consensus partitioning through a general framework. *Nucleic Acids Res* **49**, e15 (2021).
6. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
7. Anonymous, Aging Atlas: a multi-omics database for aging biology. *Nucleic Acids Res* **49**, D825-d830 (2021).
8. A. Liberzon *et al.*, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
9. D. Marakulina *et al.*, EpiFactors 2022: expansion and enhancement of a curated database of human epigenetic factors and complexes. *Nucleic Acids Res* **51**, D564-d570 (2023).
10. P. van Galen *et al.*, Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265-1281 e1224 (2019).
11. Y. Zhang *et al.*, Single-cell transcriptomics reveals multiple chemoresistant properties in leukemic stem and progenitor cells in pediatric AML. *Genome Biol* **24**, 199 (2023).
12. W. Y. Cheng *et al.*, Transcriptome-based molecular subtypes and differentiation hierarchies

improve the classification framework of acute myeloid leukemia. *Proc Natl Acad Sci U S A* **119**, e2211429119 (2022).
13. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
14. A. Wilm *et al.*, LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* **40**, 11189-11201 (2012).
15. Z. Lai *et al.*, VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* **44**, e108 (2016).
16. W. McLaren *et al.*, The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
17. X. Lin *et al.*, Integration of Genomic and Transcriptomic Markers Improves the Prognosis Prediction of Acute Promyelocytic Leukemia. *Clinical cancer research : an official journal of the American Association for Cancer Research* **27**, 3683-3694 (2021).