**Supplementary Figure 1. Cell type compositions in each sample.**

(A) UMAP projection showing the cell-type composition of all samples respectively. Each cluster is indicated with a different color in Figure 1b. HSC, hematopoietic stem cell; LC, leukemic cell; GMP, granulocyte-monocyte progenitor; Promono, promonocyte; Mono, monocyte; cDC, conventional dendritic cell; pDC, plasmacytoid dendritic cell; EPC, erythrocyte precursor cell; Pro-E, proerythroblast; baso-E, basophilic erythroblast; Poly-E, polychromatophilic erythroblast; ortho-E, orthochromatic erythroblast; LMPP, lymphoid-primed multipotential progenitor.

(B) The proportion of cell-types in each sample. The number above the graph indicates the percentage of the proportion. The Proportion percentage, which was less than 1%, were not indicated. GMP, cluster 3-4; MPS, cluster 5-13; Erythrocyte, cluster 14-20; B, cluster 22-24; T / NK, cluster 25-32. Dx, diagnosis; CR, complete remission; Rel, relapse; CON, healthy donor.
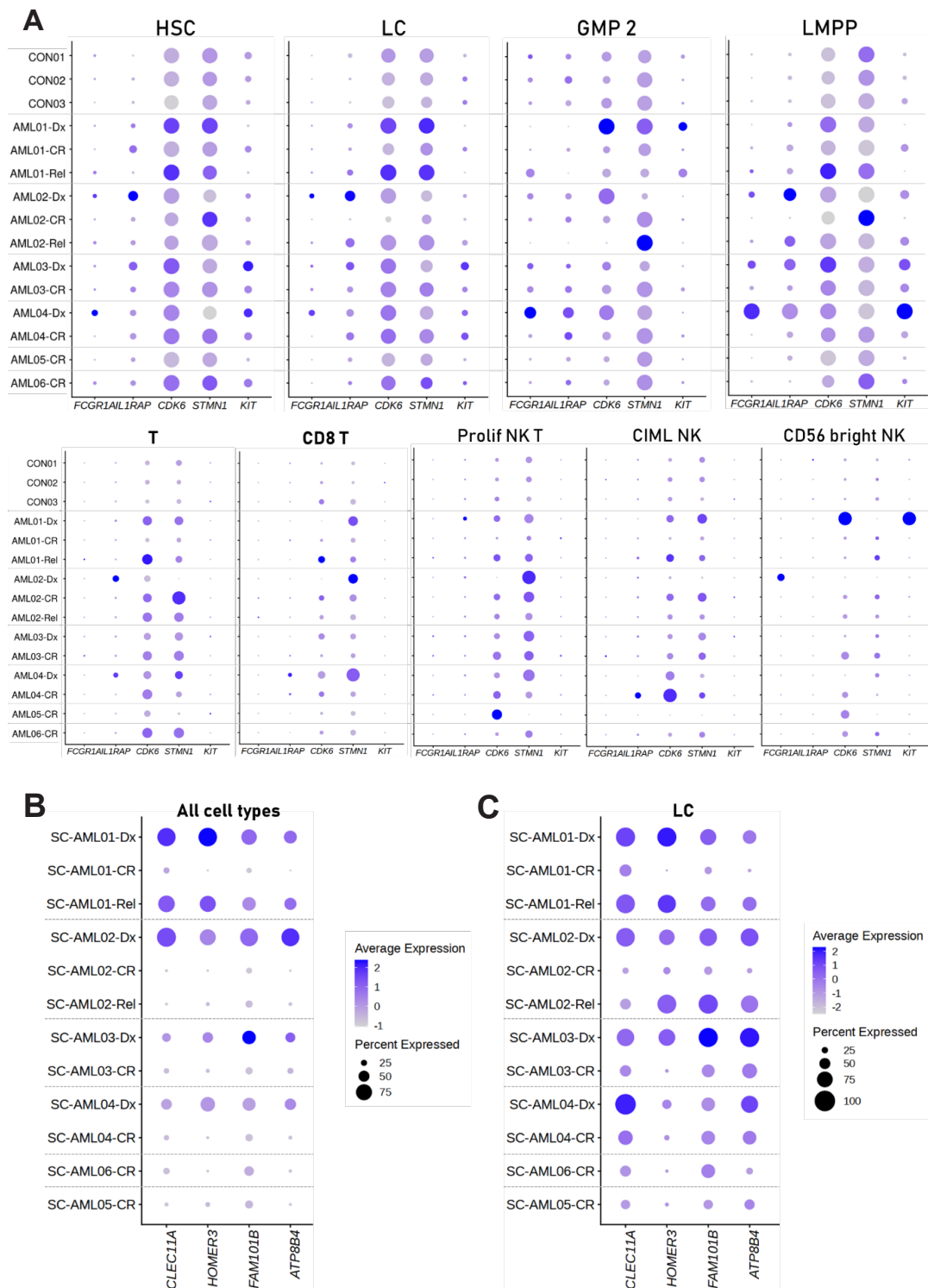
Supplementary Figure 2. Candidate genes to distinguish between diagnosis, relapse and remission.

(A) Dot plots showing expression of genes, reported to be highly expressed in AML, in HSC, LC, GMP2, LMPP, T, CD8 T, prolif NK T, CIML NK, and CD56 bright NK. (B-C) Dot plots showing expression of new candidate genes to distinguish complete remission from diagnosis and relapse in total cell types (B) or LC (C). Dot size represents the percentage of cells expressing the genes within the samples. Blue and white represent high and low expression, respectively.
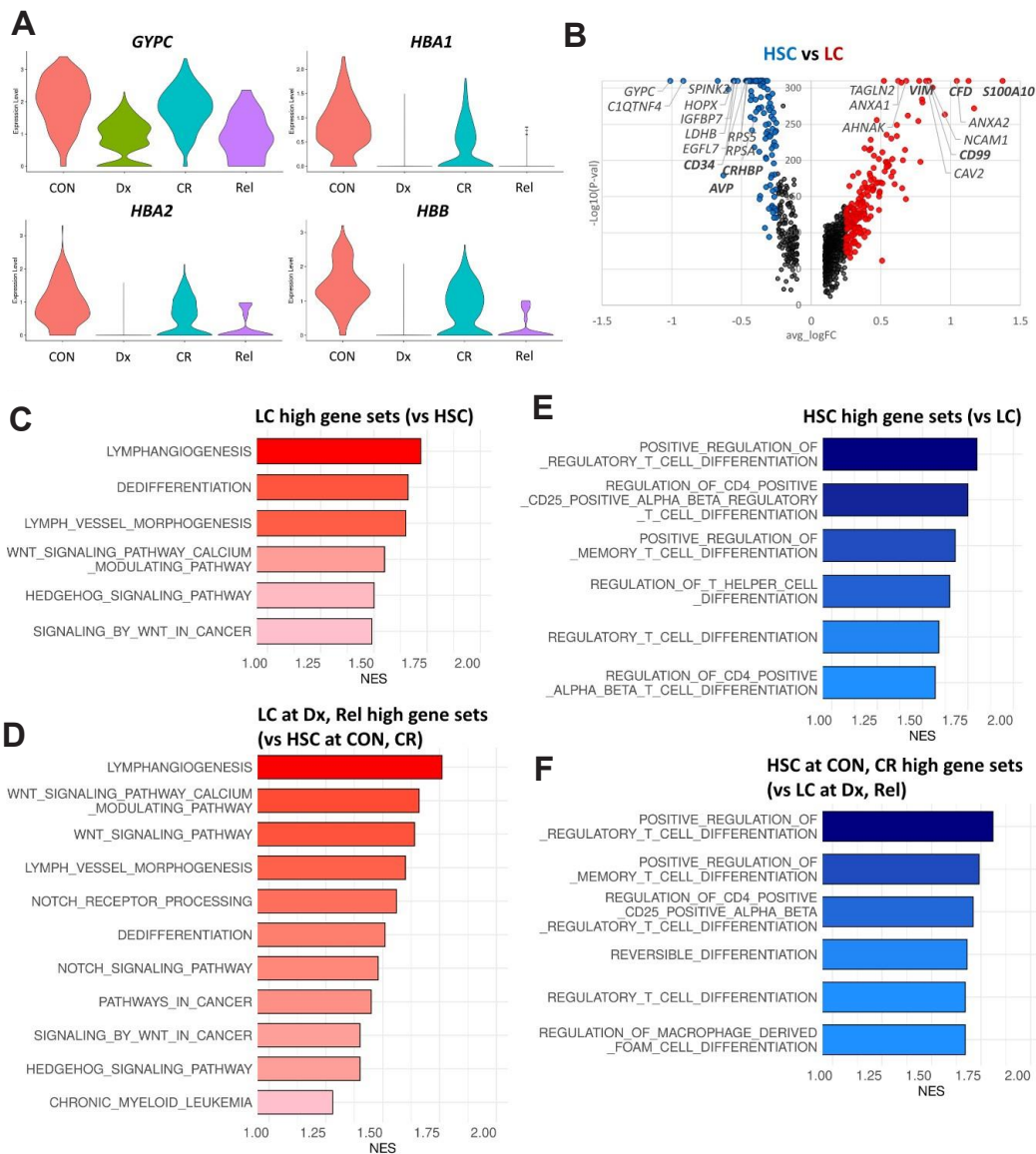
Supplementary Figure 3. Transcriptional landscape of LC and heterogeneity of HSC, related to Figure 2.

(A) Violin plot showing the expression of genes involved in red blood cell development in HSC over disease stages.

(B) Differentially expressed genes of LC vs HSC. X-axis represents the log2 fold change, and y-axis represents -log10 (*P*-value). Red dot indicates the upregulated genes in LC. Blue dot indicates the upregulated genes in HSC. Genes with a log2 fold change of less than 0.25 were indicated by black dots. Known markers for each cell type are shown in bold.

(C) Significantly enriched gene sets associated with cancer or self-renewal in LC compared to HSC. Gene sets were filtered by p-value < 0.05 and FDR < 0.25 and ordered by NES. NES, normalized enrichment score; *P*, *P*-value; FDR, false discovery rate.

(D) Significantly enriched gene sets associated with cancer or self-renewal in LC at Dx, Rel compared to HSC at CON, CR.

(E) Significantly enriched gene sets associated with differentiation in HSC compared to LC.

(F) Significantly enriched gene sets associated with differentiation in HSC at CON, CR compared to LC at Dx, Rel.
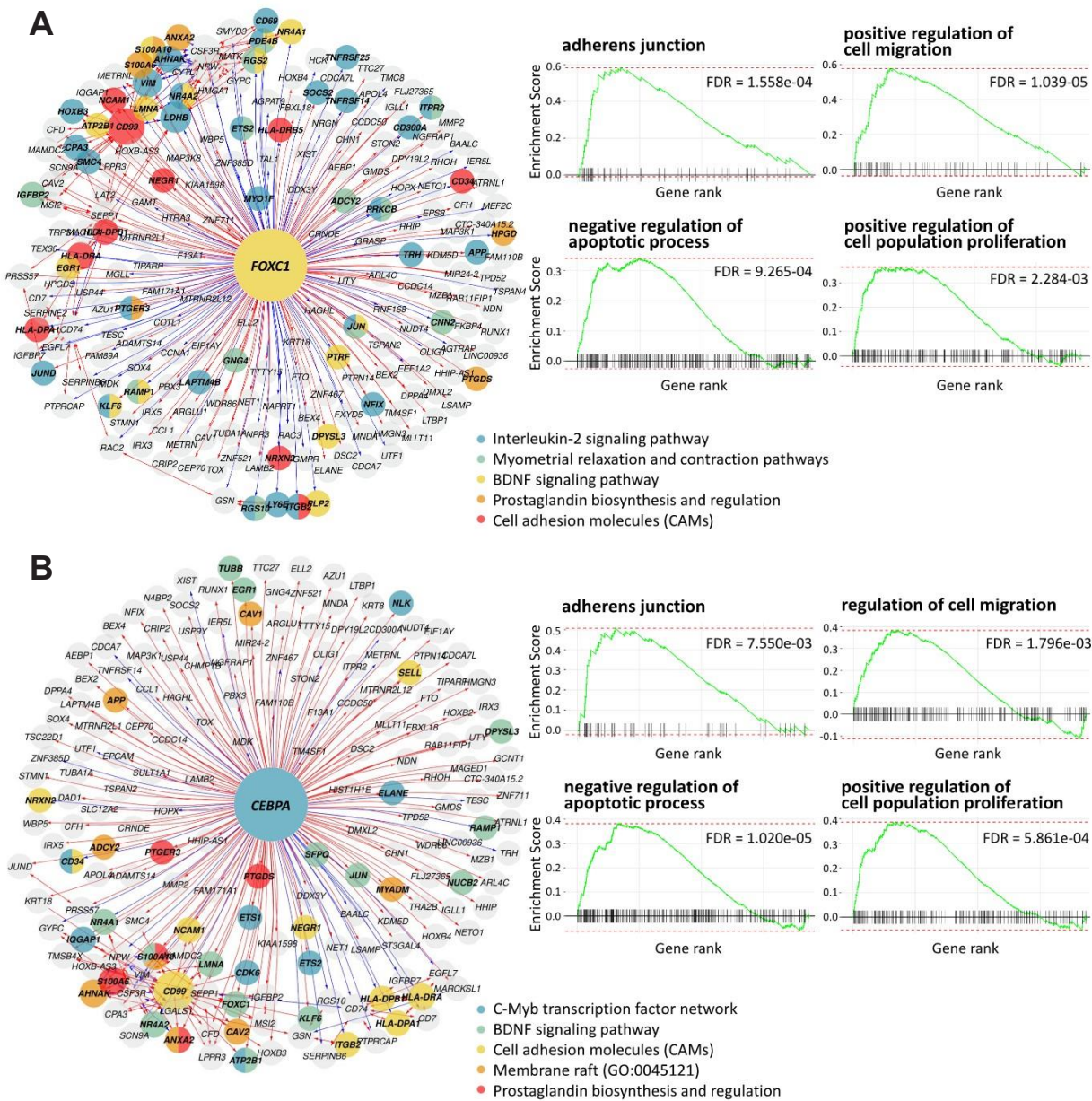
Supplementary Figure 4. Virtual KO analysis of *FOXC1* and *CEBPA* on LC, related to Figure 2.

(A-B) The egocentric plot (left) shows the interactions between *FOXC1* (A) or *CEBPA* (B) and virtual KO perturbed genes (FDR <0.05). GSEA plots (right) show enriched functions associated with virtual KO perturbed genes upon the deletion of *FOXC1* (A) or *CEBPA* (B).
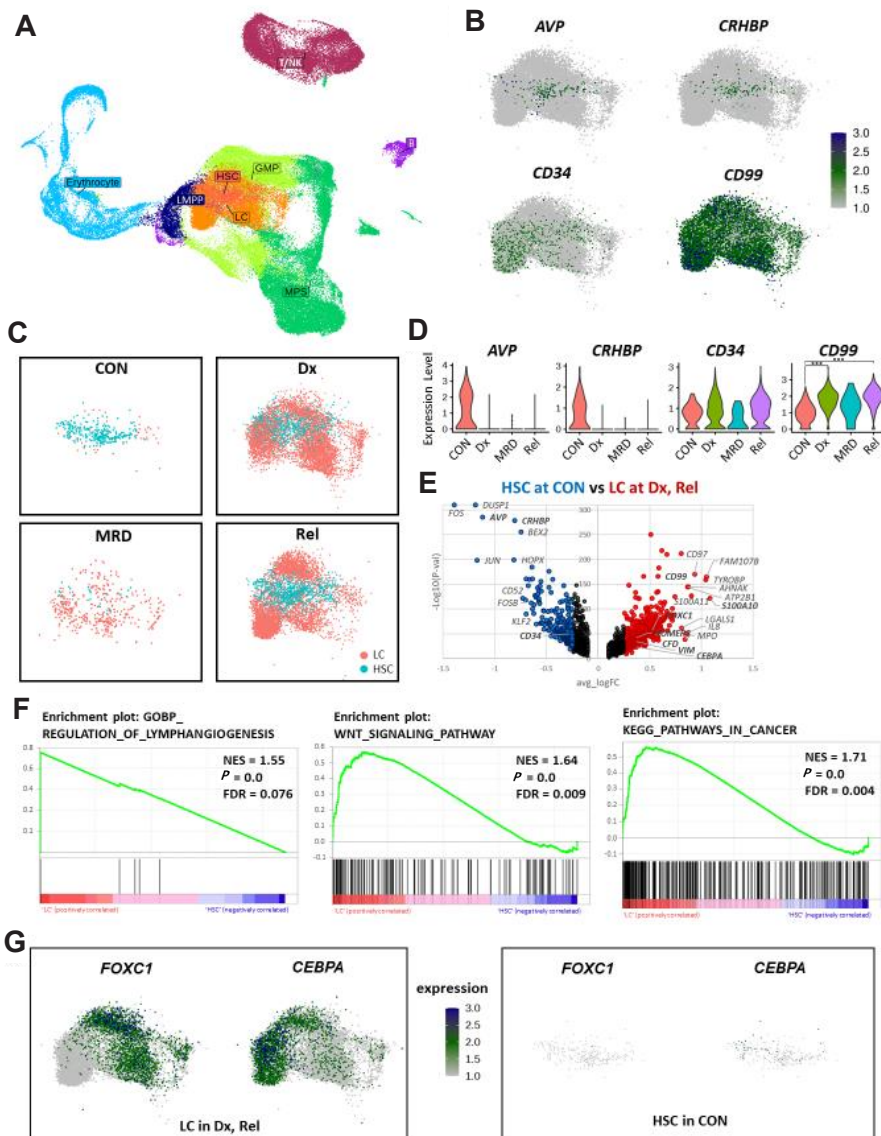
Supplementary Figure 5. Analysis of public AML scRNAseq data for validation of the results in Figure 2.

(A) UMAP visualization of public AML scRNAseq data (Dx 4, MRD 4, Rel 4) and our CON. Colors indicate cell types. Dx, diagnosis; MRD, minimal residual disease; Rel, relapse; CON, healthy donor; HSC, hematopoietic stem cell; LC, leukemic cell; GMP, granulocyte-monocyte progenitor; MPS, mononuclear phagocyte system; LMPP, lymphoid-primed multipotential progenitor.

(B) Expression of known HSC and LC marker genes in HSC and LC. *AVP, CRBHP,* and *CD34* for HSC, *CD99* for LC.

(C) UMAP visualization showing the distribution of HSCs and LCs by disease stage.

(D) Violin plot showing the expression of known HSC (*AVP*, *CRHBP*, *CD34*) and LC (*CD99*) marker genes in HSCs by disease stage. ***P-value < 0.001 by Wilcoxon rank-sum test using FindMarkers function.

(E) DEGs in LCs at Dx and Rel vs. HSCs at CON. Red dots represent upregulated genes in LCs at Dx and Rel. Blue dots represent upregulated genes in HSCs at CON and CR. Black dots represent genes with a log2 fold change < 0.25. Known markers of each cell- type and genes of interest are shown in bold.

(F) GSEA plots showing enrichment of cancer or stem cell associated gene sets in LCs at Dx and Rel compared to HSCs at CON. NES, normalized enrichment score; *P*, *P*-value; FDR, false discovery rate.

(G) Expression of *FOXC1* and *CEBPA* in LCs at Dx and Rel or HSCs at CON.
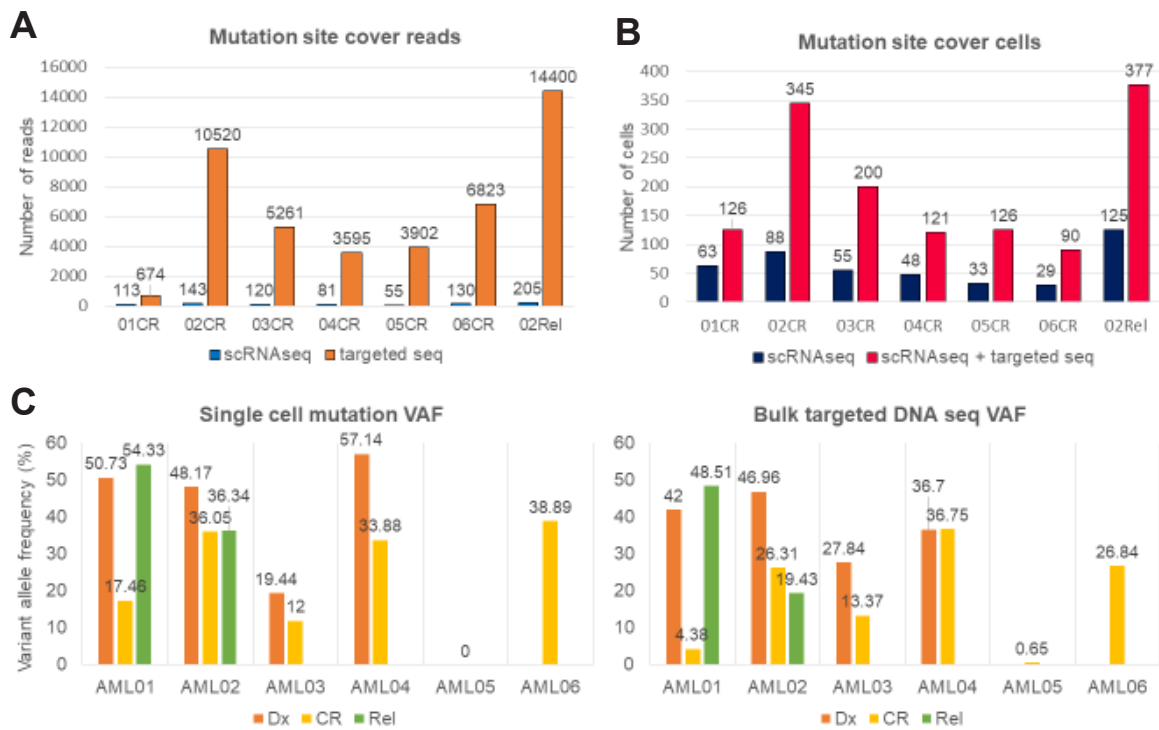
Supplementary Figure 6. Integration of targeted seq with the existing scRNA-seq to increase the number of *DNMT3A* mutation site cover reads and cells.

(A) The graph shows a comparison of the number of reads covering the mutation site in scRNA-seq and Targeted seq.

(B) The graph shows that the number of mutation site cover cells is improved by integrating scRNA-seq and targeted seq compared to scRNA-seq alone.

(C) Comparison of *DNMT3A* VAF in the single cell RNA (left) and bulk targeted DNA (right) sequencing data.
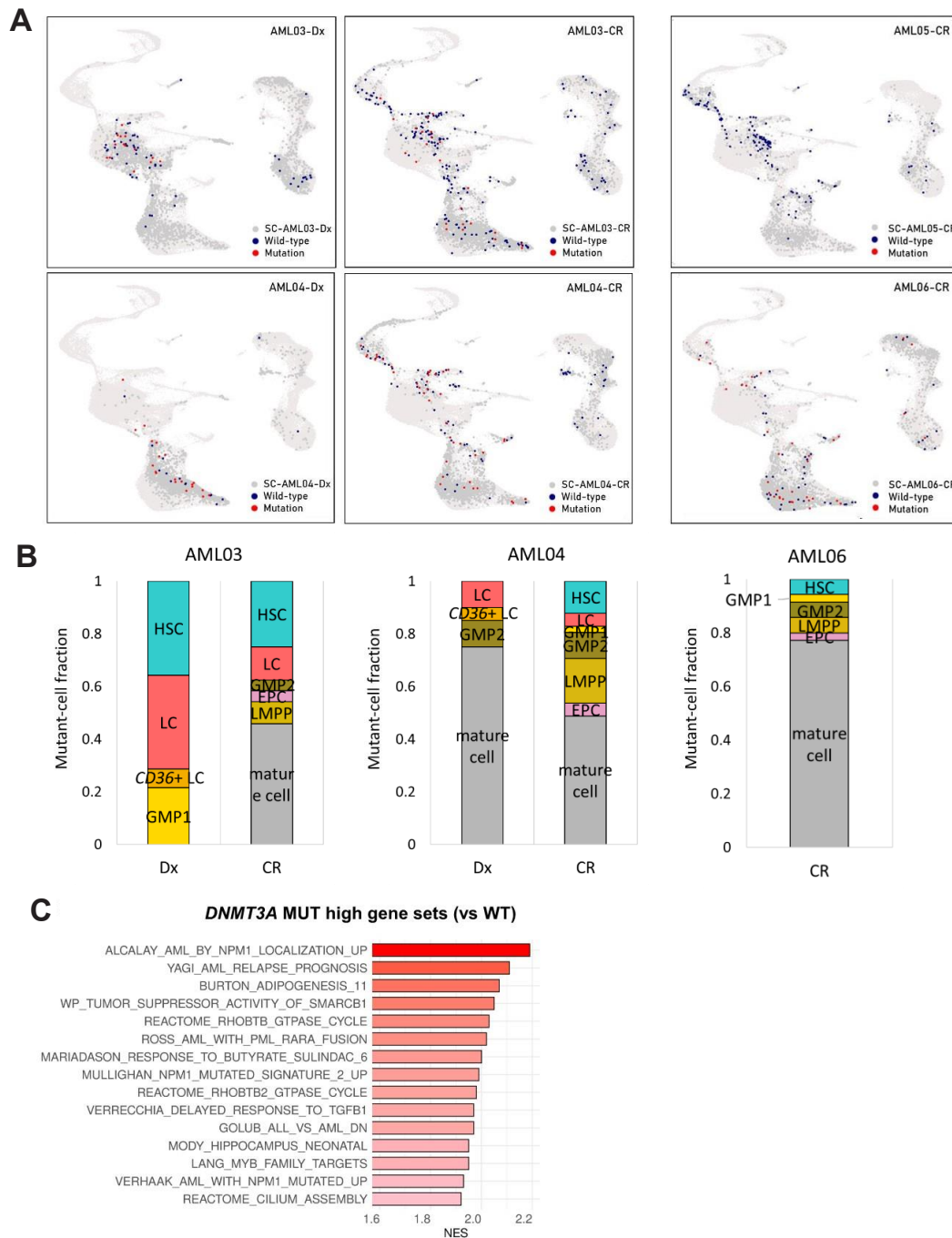
Supplementary Figure 7. Identification of cell type and signatures of *DNMT3A* mutant cells, related to Figure 3.

(A) Identification of *DNMT3A* mutant and WT cells in AML 03 – 06. red, *DNMT3A* mutant cell; navy, *DNMT3A* WT cell; dark gray, no coverage of *DNMT3A*.

(B) Fraction of *DNMT3A* mutant cells for each cluster in different time point samples of AML03, AML04 and AML06. There were no mutant cells detected at CR of AML05.

(C) Top 15 unregulated gene sets in *DNMT3A* mutant cells compared to WT cells in AML samples. Gene sets were filtered by p-value < 0.05 and FDR < 0.25 and ordered by NES. NES, normalized enrichment score; P, p-value; FDR, false discovery rate.

A GMP 1

B

GATA2   PRTN3

TOP2A

C

SC-AML01-CR
DNMT3A R882 mutant cell

SC-AML03-CR
DNMT3A R882 mutant cell

D LMPP

E

VPREB1   FLT3

MPO

F

SC-AML02-CR
DNMT3A R882 mutant cell

SC-AML04-CR
DNMT3A R882 mutant cell

SC-AML06-CR
DNMT3A R882 mutant cell
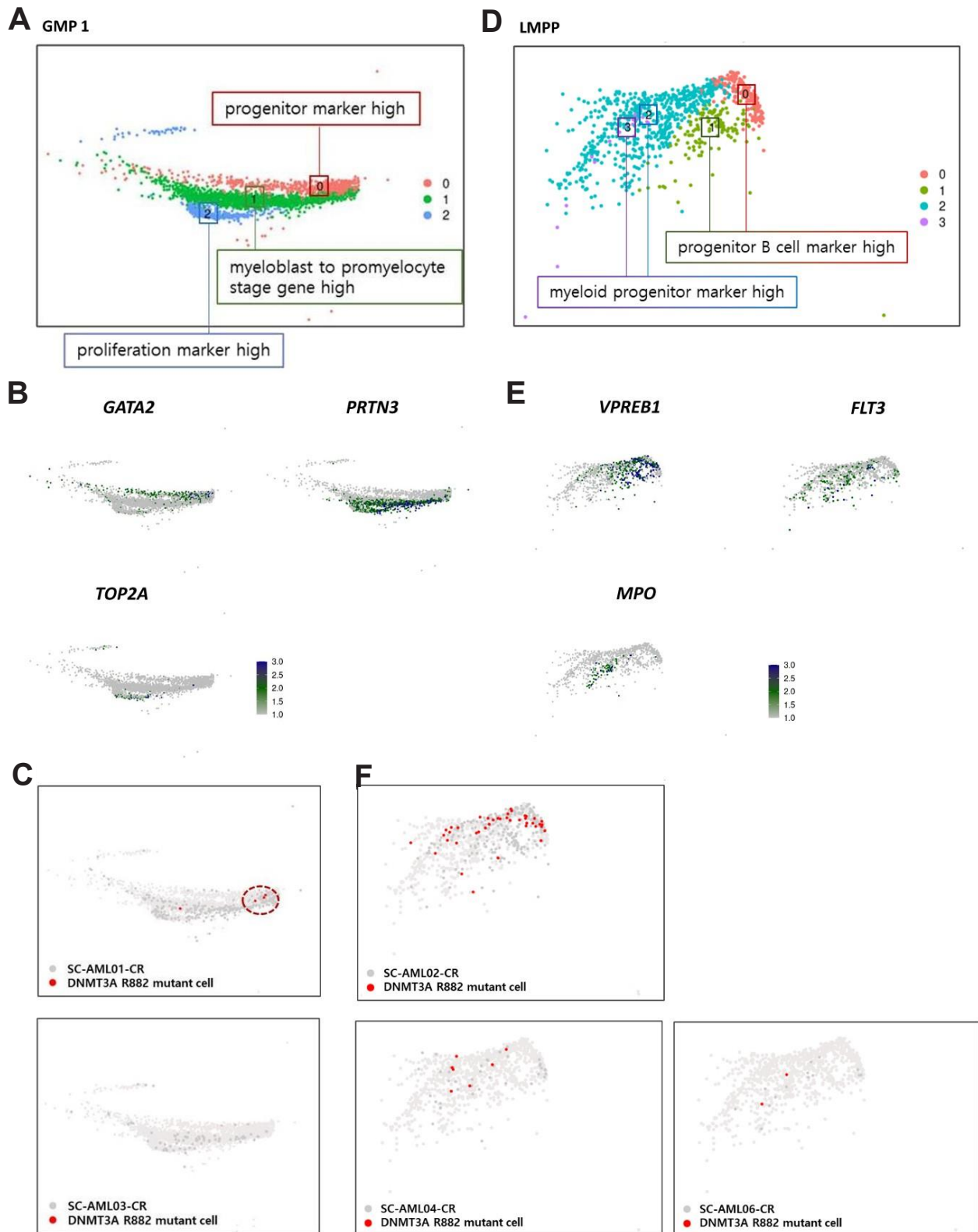
Supplementary Figure 8. Subclustering of *DNMT3A* mutant cell rich cell-types at CR of AML01 and AML02.

(A) UMAP projection showing subclusters of GMP1, a cell type *DNMT3A* mutant cell rich cell-type at CR in AML01.

(B) Expression levels of representative markers for subclusters of GMP1 are plotted onto UMAP at the middle. *GATA2* for progenitor marker high GMP1, *PRTN3* for myeloblast to promyelocyte stage gene high GMP1, *TOP2A* for proliferation marker high GMP1.

(C) UMAP projection shows the comparison of mutant cells in GMP1 (top) at CR between relapsed (AML01) and non-relapsed patients (AML03). Red dots indicate that at least one *DNMT3A* R882 mutant read was detected. Dark gray, no coverage of *DNMT3A*.

(D) UMAP projection showing subclusters of LMPP, a cell type *DNMT3A* mutant cell rich cell-type at CR in AML02.

(E) Expression levels of representative markers for subclusters of LMPP are plotted onto UMAP at the middle. *VPREB1* for progenitor B cell marker high LMPP, *FLT3* and *MPO* for myeloid progenitor marker high LMPP.

(F) UMAP projection shows the comparison of mutant cells in LMPP at CR between relapsed (AML02) and non-relapsed patients (AML04 and AML06). Red dots indicate that at least one *DNMT3A* R882 mutant read was detected. Dark gray, no coverage of *DNMT3A*.

A

B

HSC at CR vs LC at Dx, Rel

C

LC at Dx, Relhigh gene sets (vs HSC at CR)

D

HSC at CR high gene sets (vs LC at Dx, Rel)

E

F

G
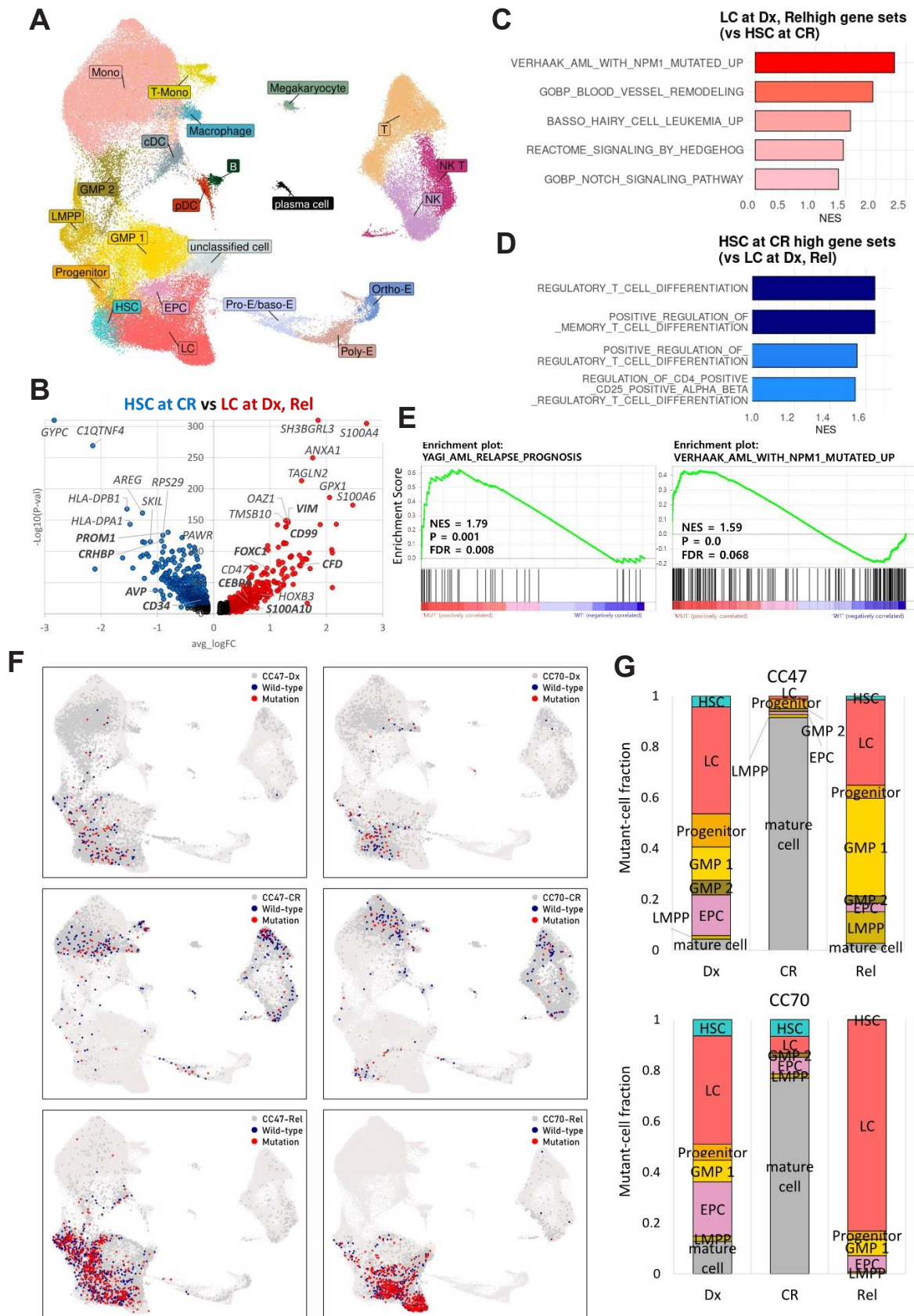
Supplementary Figure 9. scRNAseq analysis of additional AML patients to examine if our results are reproducible or heterogeneous.

(A) UMAP visualization of scRNAseq data from BM-MNCs of additional six AML patients. Colors indicate cell types. HSC, hematopoietic stem cell; LC, leukemic cell; GMP, granulocyte-monocyte progenitor; Mono, monocyte; cDC, conventional dendritic cell; pDC, plasmacytoid dendritic cell; EPC, erythrocyte precursor cell; Pro-E, proerythroblast; baso-E, basophilic erythroblast; Poly-E, polychromatophilic erythroblast; ortho-E, orthochromatic erythroblast; LMPP, lymphoid-primed multipotential progenitor.

(B) DEGs between *DNMT3A*-mutant and WT cells in HSC of additional AML samples. Red dots indicate upregulated genes in *DNMT3A*-mutant cells. Blue dots indicate downregulated genes. Black dots indicate genes with a log2 fold change < 0.25. Genes of interest are marked in bold.

(C) Significantly enriched gene sets associated with cancer or self-renewal in LC compared to HSC of additional AML patients. Gene sets were filtered by p-value < 0.05 and FDR < 0.25 and ordered by NES. NES, normalized enrichment score; *P, P*-value; FDR, false discovery rate.

(D) Significantly enriched gene sets associated with differentiation in HSC compared to LC of additional patients.

(E) GSEA plots of enriched gene sets in *DNMT3A*-mutant cells compared to WT cells of additional patients.

(F) Plot showing *DNMT3A*-mutant and WT cells projected onto the UMAP cluster from CC47 and CC70 according to disease stages. Red, *DNMT3A*-mutant cell; navy, *DNMT3A*-WT cell; dark gray, no *DNMT3A* coverage.

 (G) Fraction of *DNMT3A*-mutant cells in each cluster at different disease stages in CC47 and CC70. Mutant cell fraction: number of mutant cells in each cell-type divided by the total number of cells that cover the mutation site. Mature cell types excluding stem/progenitor cell types were merged.
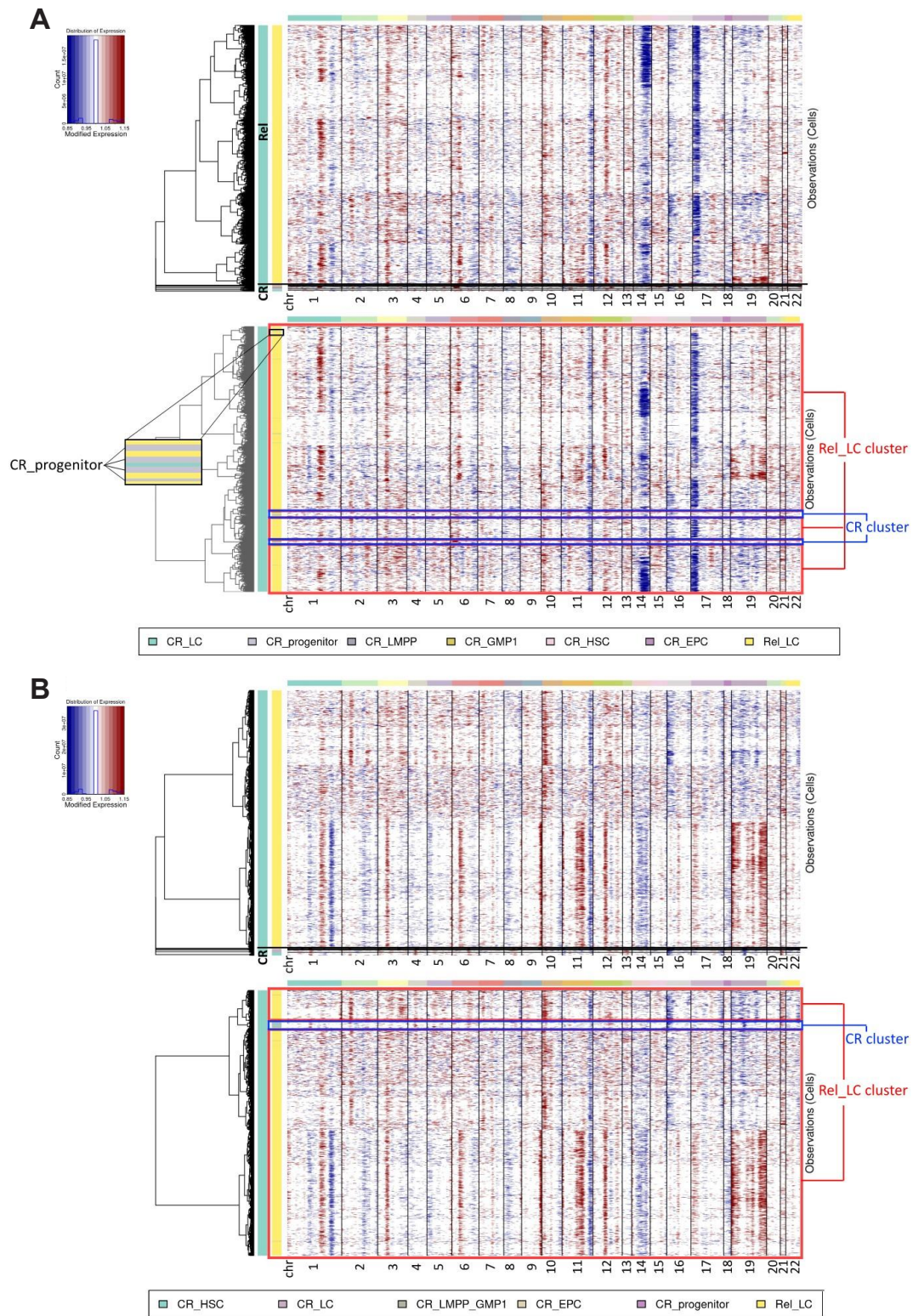
Figure S10. CNV analysis to identify clone involved in relapse in additional patients with relapsed AML.

(A, B) InferCNV heatmap of CR and Rel in CC47 (A) or CC70 (B). In the heatmaps on the top, cells are grouped by cell type. Cell types are labeled in different colors. In the heatmaps on the bottom, cells were ordered according to similar CNV patterns.

Supplementary Table 1. Clinical Information of the patients/controls and sampling time.

Patients	Age/sex	FAB classification	Disease status	Days from Dx	%Blast	Sample extraction Source	Further Treatment	Description of chemotherapy
AML01	67/male	M1	Dx	0	91	BM	Intensive induction chemotherapy	Idarubicin (IDA) 12 mg/m^2/day (D1-3) Cytarabine (Ara-C) 100 mg/m^2/day (D1-7)
			CR	31	2.5	BM	Consolidation chemotherapy for 3 cycles	Ara-C 1.0 g/m^2/12hours (D1,3,5)
			Relapse	378	80	BM	Intensive reinduction chemotherapy	Fludarabine 30 mg/m^2/day (D1-5) Ara-C 2.0 g/m^2/day (D1-5)
			Death	749				
AML02	61/female	M0	Dx	0	97	BM	Intensive induction chemotherapy for 2 cycles	Fludarabine 25 mg/m^2/day (D1-4) Ara-C 1.0 g/m^2/day (D1-4) IDA 5 mg/m^2/day (D1-3)
			CR	112	0.8	BM	Consolidation chemotherapy for 3 cycles	Ara-C 2.0 g/m^2/12hours (D1,3,5)
			Relapse	469	36	BM	Intensive reinduction chemotherapy	Idarubicin (IDA) 12 mg/m^2/day (D1-3) Cytarabine (Ara-C) 100 mg/m^2/day (D1-7)
			Death	567				
AML03	63/female	M4	Dx	0	66	BM	Intensive induction chemotherapy	IDA 12 mg/m^2/day (D1-3) Cytarabine (Ara-C) 100 mg/m^2/day (D1-7)
			CR	33	2.1	BM	Consolidation chemotherapy for 3 cycles	Ara-C 1.0 g/m^2/12hours (D1,3,5)
			Alive	3,371				
AML04	56/male	M4	Dx	0	22	BM	Intensive induction chemotherapy	IDA 12 mg/m^2/day (D1-3) Cytarabine (Ara-C) 100 mg/m^2/day (D1-7)
			CR	67	1		Consolidation chemotherapy for 3 cycles	Ara-C 3.0 g/m^2/12hours (D1,3,5)
			After 3rd consolidation	207	0.5	BM		
			Alive	3,164				
AML05	65/male	M2	Dx	0	41	BM	Intensive induction chemotherapy for 2 cycles	Fludarabine 25 mg/m^2/day (D1-4) Ara-C 1.0 g/m^2/day (D1-4) Idarubicin (IDA) 5 mg/m^2/day (D1-3)
			CR	89	1.9	BM	Consolidation chemotherapy for 3 cycles	Ara-C 2.0 g/m^2/12hours (D1,3,5)
			Relapse	364	11		Re-induction therapy	Idarubicin (IDA) 12 mg/m^2/day (D1-3) Cytarabine (Ara-C) 100 mg/m^2/day (D1-7)
			Death	679				
AML06	33/male	M4	Dx	0	57		Intensive induction chemotherapy	Idarubicin (IDA) 12 mg/m^2/day (D1-3) Cytarabine (Ara-C) 100 mg/m^2/day (D1-7)
			CR	33	4	BM	Consolidation chemotherapy for 3 cycles	Ara-C 3.0 g/m^2/12hours (D1,3,5)
			Alive	4,903				
Control1	21/female					BM		
Control2	44/male					BM		
Control3	32/male					BM		

ID	Age/Sex	FAB	Status			Sample	Treatment	Regimen
CC20	54/male	M1	Dx			BM	Intensive induction chemotherapy, and re-induction therapy	Daunorubicin 12 mg/m^2/day (D1-3) Ara-C 100 mg/m^2/day (D1-7), and Fludarabine 30 mg/m^2/day (D1-5) Ara-C 2.0 g/m^2/day (D1-5)
			CR			BM	Consolidation chemotherapy, and haploidentical HCT	Fludarabine 30 mg/m^2/day (D1-5) Ara-C 2.0 g/m^2/day (D1-5)
CC47	60/female	M0	Dx			BM	Intensive induction chemotherapy	Daunorubicin 12 mg/m^2/day (D1-3) Ara-C 100 mg/m^2/day (D1-7)
			CR			BM	Consolidation chemotherapy for 2cycles, and unrelated HCT	Ara-C 2.0 g/m^2/12hours (D1,3,5)
			Relapse				Reinduction chemotherapy	Fludarabine 30 mg/m^2/day (D1-5) Ara-C 2.0 g/m^2/day (D1-5)
CC64	54/female	M4	Dx			BM	Intensive induction chemotherapy	Daunorubicin 12 mg/m^2/day (D1-3) Ara-C 100 mg/m^2/day (D1-7)
			CR			BM	Consolidation chemotherapy for 1 cycle, and sibling HCT	Ara-C 2.0 g/m^2/12hours (D1,3,5)
			Alive	912				
CC70	68/female	M4	Dx	0	37	BM	Intensive induction chemotherapy	Daunorubicin 12 mg/m^2/day (D1-3) Ara-C 100 mg/m^2/day (D1-7)
			CR	25	1	BM	Consolidation chemotherapy for 3 cycles	Ara-C 1.0 g/m^2/12hours (D1,3,5) Fludarabine 30 mg/m^2/day (D1-5) Ara-C 2.0 g/m^2/day (D1-5), and decitabine 20 mg/m^2/day (D1-5) venetoclax 200mg/day (D1-28)
			Relapse	202	87		Reinduction chemotherapy	
			Death	294				
CC76	49/male	M5	Dx	0	62	BM	Intensive induction chemotherapy	Daunorubicin 12 mg/m^2/day (D1-3) Ara-C 100 mg/m^2/day (D1-7)
			CR	49	2	BM	Consolidation chemotherapy for 3 cycles	Ara-C 3.0 g/m^2/12hours (D1,3,5)
			Alive	782				
CC80	59/female	M2	Dx	0	95	BM	Intensive induction chemotherapy	Daunorubicin 12 mg/m^2/day (D1-3) Ara-C 100 mg/m^2/day (D1-7)
			CR	37	1	BM	Consolidation chemotherapy for 3 cycles	Ara-C 3.0 g/m^2/12hours (D1,3,5)

Abbreviations: FAB, French-American-British; Dx, diagnosis; BM, bone marrow; CR, complete remission; HCT, hematopoietic cell transplantation.

Supplementary Table 2. Primer sequences for targeted sequencing of *DNMT3A*.

Primer name	Primer Sequence (5' to 3')	Amplicon size
Read1 (Forward)	CTACACGACGCTCTTCCGATCT	
DNMT3A_1_01cr	GCGCAGAATGCTGGGTATTTGGTTTCCCAGTCC	1553 bp
DNMT3A_1_02cr	ATCTTACCGAGGGGGTATTTGGTTTCCCAGTCC	1553 bp
DNMT3A_1_03cr	TATGGTGTACCGGGGTATTTGGTTTCCCAGTCC	1553 bp
DNMT3A_1_04cr	CGAACCTGTAGAGGGTATTTGGTTTCCCAGTCC	1553 bp
DNMT3A_1_05cr	ATAGGCCACTGTGGGTATTTGGTTTCCCAGTCC	1553 bp
DNMT3A_1_06cr	TCTCAGTGAAGCGGGTATTTGGTTTCCCAGTCC	1553 bp
DNMT3A_1_02rel	GAGACTATCGTCGGGTATTTGGTTTCCCAGTCC	1553 bp
DNMT3A_2_01cr	GCGCAGAATGCTACTGACGTCTCCAACATGAGC	1527 bp
DNMT3A_2_02cr	ATCTTACCGAGGACTGACGTCTCCAACATGAGC	1527 bp
DNMT3A_2_03cr	TATGGTGTACCGACTGACGTCTCCAACATGAGC	1527 bp
DNMT3A_2_04cr	CGAACCTGTAGAACTGACGTCTCCAACATGAGC	1527 bp
DNMT3A_2_05cr	ATAGGCCACTGTACTGACGTCTCCAACATGAGC	1527 bp
DNMT3A_2_06cr	TCTCAGTGAAGCACTGACGTCTCCAACATGAGC	1527 bp
DNMT3A_2_02rel	GAGACTATCGTCACTGACGTCTCCAACATGAGC	1527 bp

Supplementary Table 4. Mutation information at Dx and CR of patients.

Patient	Chr	Start	End	Ref	Alt	Func.refGene	Gene.refGene	Dx_var_freq	CR_var_freq
AML01	2	25457242	25457242	C	T	exonic	DNMT3A	42	4.38
	4	106197378	106197378	A	G	exonic	TET2	46.07	2.51
	5	170837543	170837543	-	TCTG	exonic	NPM1	33.68	0
	7	151875097	151875097	T	-	splicing	KMT2C	34.21	0
	21	44514780	44514780	C	T	exonic	U2AF1	4.2	0
AML02	2	25457242	25457242	C	T	exonic	DNMT3A	46.96	21.84
	1	115258744	115258744	C	T	exonic	NRAS	43.85	0
	2	209113113	209113113	G	A	exonic	IDH1	43.2	0.17
	11	32413556	32413556	T	G	exonic	WT1	46.4	0.66
	X	39932019	39932019	C	-	exonic	BCOR	48.15	0.21
	X	70340966	70340966	G	A	exonic	MED12	45.87	0
AML03	2	25457243	25457243	G	T	exonic	DNMT3A	27.84	1
	4	106158269	106158269	A	-	exonic	TET2	17.72	0
	4	106190770	106190770	G	T	exonic	TET2	32.7	0.99
	5	170837543	170837543	-	TCTG	exonic	NPM1	23.81	0
	6	75857422	75857422	A	G	exonic	COL12A1	33.02	0.99
AML04	2	25457242	25457242	C	T	exonic	DNMT3A	36.7	36.75
	1	115258748	115258748	C	A	exonic	NRAS	32.24	0.12
	4	106158444	106158444	-	A	exonic	TET2	25.87	0
	4	106193931	106193931	C	T	exonic	TET2	28.75	0
	5	170837543	170837543	-	TCTG	exonic	NPM1	35.46	0
AML05	2	25457242	25457242	C	T	exonic	DNMT3A	11.16	0.65
	5	82817300	82817300	G	A	exonic	VCAN	8.67	0.17
	15	90631934	90631934	C	T	exonic	IDH2	7.34	1.2
	X	39921391	39921391	C	T	splicing	BCOR	22.89	0
	X	129149392	129149392	-	ACGG	exonic	BCORL1	7.77	0
AML06	2	25457242	25457242	C	T	exonic	DNMT3A	27.21	26.84
	5	170837543	170837543	-	TCTG	exonic	NPM1	14.29	0
	15	33842404	33842404	C	T	exonic	RYR3	22.3	0.53
	15	90631934	90631934	C	T	exonic	IDH2	18.34	0

Patient	Chr	Nucleotide	Amino acid	Gene.refGene	Dx_var_freq	CR_var_freq
CC20	2	c.2644C>T	p. R882H	DNMT3A	46.52	10.66
	7	c.4559C>T	p.A1520V	MGA	48.86	49.32
	15	c.515G>A	p.R172K	IDH2	43.37	0
	21	c.578A>G	p.H193R	U2AF1	48.89	50.82
	X	c.1288C>T	p.Q430*	BCOR	93.89	0
CC47	2	c.2645G>A	p. R882C	DNMT3A	47.94	26.85
	5	c.860_863dupTCTG	p.W288fs	NPM1		
	13	c.1716_1793dupTGAAAGCCAGCTACAGATGGTACAGGTGACCGGCTCCTCAGATAATGAGTACTTCTACGTTGATTTCAGAGAATATGA	p.Y597_E598insDESQLQMVQVTGSSDNEYFYVDFREY	FLT3	20.96	0
	11	c.2530A>C	p.S844R	CBL	54.53	52.51
	11	c.846G>C	p.K282N	KMT2A	47.26	48.98
	4	c.4082G>A	p.G1361D	TET2	46.24	0
CC64	2	c.2645G>T	p. R882S	DNMT3A	47.03	0
	13	c.2505T>G	p.D835E	FLT3	42.71	0
	5	c.860_863dupTCTG	p.W288fs	NPM1	45.49	0
	9	c.302C>T	p.P101L	CDKN2A	49.85	49.33
	9	c.3401A>G	p.Q1134R	NOTCH1	46.32	49.54
	8	c.1590_1591delGA	p.K531fs	RAD21	45.11	0
CC70	2	c.2644C>T	p. R882H	DNMT3A	46.58	26.35
	13	c.1773_1805dupCGTTGATTTCAGAGAATATGAATATGATCTCAA	p.L601_K602insNVDFREYEYDL	FLT3	13.02	0
	2	c.395G>A	p.R132H	IDH1	10.96	0
	15	c.418C>T	p.R140W	IDH2	16.56	0
	5	c.860_863dupTCTG	p.W288fs	NPM1	50.00	0
	1	c.35G>C	p.G12A	NRAS	5.22	0
	7	c.82T>G	p.L28V	MGA	53.82	0
CC76	2	c.2644C>T	p. R882H	DNMT3A	45.66	7.59
	13	c.2503G>T	p.D835Y	FLT3	2.22%	0
	5	c.860_863dupTCTG	p.W288fs	NPM1	41.35%	0
	1	c.34G>A	p.G12S	NRAS	4.25%	0
	4	c.1421C>T	p.P474L	TET2	50.68	48.79
CC80	2	c.2645G>A	p. R882C	DNMT3A	0	1.36
	5	c.860_863dupTCTG	p.W288fs	NPM1	48.97	0
	4	c.556G>T	p.E186*	TET2	47.2	0
	7	c.1795G>A	p.E599K	CUX1	46.38	0

Chr, Chromosome number; Start, Start position; End, End position; Ref, Reference base(s); Alt, Alternate non-reference alleles called on at least one of the samples; Func.refGene, Regions (e.g., exonic, intronic, non-coding RNA)) that one variant hit; Gene.refGene, Gene name associated with one variant; Dx_var_freq, variant allele frequency in diagnosis; CR_var_freq, variant allele frequency in complete remission