Supplementary data to:

Cell lineage level-targeted sequencing to identify acute myeloid leukemia with myelodysplasia related changes

Kazuaki Yokoyama,[1] Eigo Shimizu,[2*] Nozomi Yokoyama,[3*] Sousuke Nakamura,[5*] Rika Kasajima,[4*] Miho Ogawa,[5] Tomomi Takei,[5] Mika Ito,[5] Asako Kobayashi,[5] Rui Yamaguchi,[2] Seiya Imoto,[4] Satoru Miyano[2] & Arinobu Tojo[1, 5+]

*These authors contributed equally to this work.

[1]Department of Hematology/Oncology, Research Hospital, Institute of Medical Science, University of Tokyo, Tokyo, Japan.

[2] Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan.

[3]Department of Applied Genomics, Research Hospital, Institute of Medical Science, University of Tokyo, Tokyo, Japan.

[4] Division of Health Medical Data Science, Health Intelligence Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan.

[5]Division of Molecular Therapy, Advanced Clinical Research Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan.

[+]Corresponding author: A.T.

Arinobu Tojo

Department of Hematology/Oncology, Research Hospital, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

E-mail: a-tojo@ims.u-tokyo.ac.jp

Phone: 81-3-5449-5633

Fax: 81-3-5540-5429

ONLINE METHODS:

Patients

Our study protocol was approved by the ethical evaluation committee (24-27/26-112-270402) of the Institute of Medical Science at the University of Tokyo and was in accordance with the Declaration of Helsinki. Forty-three patients were recruited for our study, all of whom provided written informed consent.

Cell lineage level-targeted sequencing workflows

A summary of the cell lineage level-targeted sequencing workflow is shown in Supplementary Table 2; the workflow consisted of five steps with their associated methodologies, which are described below.

STEP 1: Flow cytometry sorting of trio fractions

Forty-three fresh or 10% DMSO cryopreserved mononuclear cells from diagnostic samples (bone marrow or peripheral blood) were used. Single-cell suspensions consisting of 1×10^7 cells suspended in 1 mL of ice-cold phosphate-buffered saline

(PBS) supplemented with 2% calf serum and 1 mM EDTA, were stained with an anti-human fluorescence-conjugated specific antibody for CD34 (AlexaFluor647, BD Biosciences, San Jose, CA, USA), CD66b-Pacific Blue, CD3-allophycocyanin (APC)-CY7, CD14-Brilliant Violet 605 (all three from Biolegend, San Diego, CA, USA), or CD45-Violet 500 (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions. After filtering with a cell strainer (BD Biosciences) and addition of CD235a/propidium iodide (PI) (Biolegend), trio fractions were sorted on a FACS Aria (BD Biosciences) using the following gating: granulocytes: CD235a/$PI^{negative}$ $CD45^{dull}$ sideways scatter $(SSC)^{high}$ $CD14^{negative}$ $CD66b^{high}$; blast cells: CD235a/$PI^{negative}$ $CD45^{dull}$ SSC^{low} $CD14^{negative}CD66b^{negative}$ CD34+; and T cells: CD235a/$PI^{negative}$ $CD45^{high}$ SSC^{low} $CD14^{negative}$ CD3+. In cases of CD34-negative blast-enriched cases with AML, we sorted the CD235a/$PI^{negative}$ $CD45^{dull}$ $CD66b^{negative}$ $CD34^{negative}$ blast fraction as a representation of the blast fraction. Fifty-thousand sorted trios and residual unfractionated samples were stored as cell pellets at -20 °C for future use. Representative sorting plots are shown in Supplementary Figure 1. The sorting purity of granulocytes from either fresh or cryopreserved samples were confirmed by cytospin

preparations followed by May-Grünwald-Giemsa staining (Supplementary Figure 2) with observed purities of 99.3 ± 0.3% (n = 3) or 98 ± 0.5% (n = 3), respectively.

STEP 2: Initial mutational screening of flow-sorted trios by targeted deep sequencing

DNA was extracted using a Gentra Puregene Blood kit (Qiagen, Valencia, CA, USA) and quantified using a Qubit DNA HS assay kit (Life Technologies, Carlsbad, CA, USA). Enrichment of 568 amplicons in 54 genes (141 kb) commonly mutated in MDS and AML was performed using 20 ng of genomic DNA from flow-sorted trios (and buccal swab) via the TruSight Myeloid Panel (Illumina, San Diego, CA, USA). Targeted exons of genes included in the myeloid panel are listed in Supplementary Table 3. Targeted deep sequencing of 250-bp paired-end reads was performed on a MiSeq with Reagent kit v3 (Illumina).

STEP 3: Variant calling of targeted deep sequencing data

We used human genome build 19 (hg19) as the reference sequence in our analysis. To minimize the risk of false positives and to focus on the most relevant somatic events, we used two independent analysis pipelines for identifying somatic and germline alterations. Single nucleotide variants (SNVs) were identified using a pipeline consisting of MiSeq Reporter v2.5 (http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqcustomamplicon/truseq-custom-amplicon-15-reference-guide-15027983-02.pdf) and Genomon2 (http://genomon.hgc.jp/exome/en/), whereas structural variants (SVs) such as *FLT3* internal tandem duplications (ITD) were identified using a pipeline composed of Pindel (http://gmt.genome.wustl.edu/packages/pindel/) and Genomon-SV (https://github.com/Genomon-Project/GenomonSV).

For our analysis using the MiSeq Reporter pipeline, paired-end read fastq files were automatically processed using MiSeq Reporter v2.5, according to the manufacturer's provided workflow, which requires both fastq files and a manifest file for input. The manifest file contains information including the sequences of the primer pairs, the expected sequence of the amplicons, and the coordinates relative to the reference

genome (hg19). As a workflow output, a VCF file that contains a list of variants is generated for each sample. Briefly, each read pair is separately processed to individually identify the corresponding primer pair (allowing one mismatch). Next, it is aligned to the expected amplicon sequence (primers excluded) via a banded Smith-Waterman algorithm, allowing gaps a maximum of one third of its length. Variant calling was performed using the Genome Analysis ToolKit (GATK) v1.6 (https://software.broadinstitute.org/gatk/) using generated BAM files.

STEP 4: Annotation and filtering of variants using an in-house pipeline

The Integrative Genomics Viewer (IGV) version 2.3.57 (https://software.broadinstitute.org/software/igv/download) was used to visualize and inspect the read alignments and variant calls. For detection of somatic mutations, we analyzed matched T cells in all cases and, if available, buccal swab (n = 35). Annotations and high-stringency filtering of all variants were completed through an in-house pipeline. The filtering strategies to identify candidate overlapping and non-overlapping mutations are described in Supplementary Table 2. For our analysis, the

databases used to annotate variants included RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/), the 1000 Genomes Project as of August 2015 (http://www.internationalgenome.org/data), dbSNP131 (http://www.ncbi.nlm.nih.gov/projects/SNP/), ToMMo version 1 (https://ijgvd.megabank.tohoku.ac.jp/), the Human Genetic Variation Database (HGVD) as of October 2013 (http://www.hgvd.genome.med.kyoto-u.ac.jp/), ClinVar as of June 2015 (https://www.ncbi.nlm.nih.gov/clinvar/), the Human Gene Mutation Database Professional (HGMD Pro) as of March 2015 (http://www.hgmd.cf.ac.uk/ac/index.php), the cBioPortal for Cancer Genomics as of September 2015 (http://cbioportal.org), the TumorPortal as of September 2015 (http://www.tumorportal.org/), the Catalogue Of Somatic Mutations In Cancer (COSMIC) version 70 (http://cancer.sanger.ac.uk/cosmic), and ICGC Data Portal (https://dcc.icgc.org/). The computational algorithms, Sorting Intolerant From Tolerant (SIFT available at http://sift-dna.org), PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/), and MutationTaster (http://www.mutationtaster.org/) were used to predict whether mutations were damaging.

STEP 5: Mutation validation

Sanger sequencing. To experimentally validate identified mutations, we performed Sanger sequencing with standard procedures using unfractionated residual samples and, if available, paired controls. Briefly, PCR reactions were performed using primers and Platinum Taq DNA Polymerase High Fidelity (Invitrogen, San Diego, CA). DNA (50 ng) was amplified and PCR products were purified using a QIAquick PCR purification kit (Qiagen), followed by direct sequencing with an ABI3130 Genetic Analyzer (Applied Biosystems, CA, USA) using BigDye Terminator v3.1 chemistry (Thermo Fisher Scientific, Waltham, MA, USA). PCR conditions and primer sequences are available upon request.

In those cases in which limited residual samples were available, droplet digital PCR (ddPCR) or targeted deep sequencing using a different probe (QIASEQ) was performed for mutation confirmation as described below.

ddPCR. Gene-specific primers and hydrolysis probes for wild-type (6-carboxy-

2,4,4,5,7,7-hexachlorofluorescein succinimidyl ester (HEX)-labeled) and mutant specific probes (6-carboxyfluorescein (FAM)-labeled) with a nonfluorescent quencher (black-hole quencher 1 (BHQ1)) were designed using Primer Express version 5.0 (Applied Biosystems, Inc., Foster City, CA, USA). Droplets were generated and run on a QX20 ddPCR platform (Bio-Rad, Hercules, CA, USA) according to manufacturer's instructions as previously described elsewhere.

Targeted deep sequencing using a different probe. The enrichment of 5,500 amplicons in 141 genes using the Myeloid Neoplasms Panel (Qiagen) followed by targeted deep sequencing was performed as described in STEP 3.

Supplementary Tables

Supplementary Table 1 (related to Figure 3). Clinicopathological characteristics of the patients in the study.

Patient (UPN)	Age/ Gender	Clinical diagnosis by WHO2008 criteria	IPSS-R (MDS) or MRC (AML) risk	Prior at least 6month -lasing cytopenia before dignosis	Prior documented MDS or MDS/MPN	Cytogenetics	Karyotype abnormality supporting AML-MRC including –7/del(7q)	Dysplasia > 50% in multilinages, supporting AML-MRC	Dysplasia > 10% in lineage(s), suppporting MDS
1	83/M	RARS	Very Low	Yes	No	45,X,-Y[20]	No	Yes	Yes
2	46/M	MDS/MPN (CMML-1)	Very Low	Yes	Yes	46,XX[20]	No	Yes	Yes
3	71/F	MDS/MPN-u	Low	Yes	No	46,XX[20]	No	No	Yes
4	66/M	RCUD	Low	Yes	No	46,XY[20]	No	No	Yes
5	68/F	RCUD	Int	Yes	No	46,XX[20]	No	No	Yes
6	63/M	RCMD	Int	Yes	No	46,XY, del(20)(q11.2q13.3)[20]	No	Yes	Yes
7	57/M	RCMD	Int	Yes	Yes	46,XY[20]	No	Yes	Yes
8	50/M	RAEB-2	High	Yes	Yes	46,XY,-dup(1)(q21q42)[5]	No	Yes	Yes
9	65/F	RAEB-1	High	Yes	Yes	46,XX,der(5)t(1;5)(q21;q31)[11]	No	Yes	Yes
0	72/M	RAEB-2	Very High	Yes	Yes	46,XY,+1,der(1;7)(q10;p10)[20]	Yes	Yes	Yes
11	74/M	RAEB-2	Very High	Yes	No	43,XY,-3,-5,add(7)(q22),-12,-16, -17,-19,+mar1,+mar2,+mar3[8]	Yes	Yes	Yes

(Supplementary Table 1 continues)

(Supplementary Table 1 continued)

Patient (UPN)	Age/ Gender	Clinical diagnosis by WHO2008 criteria	IPSS-R (MDS) or MRC (AML) risk	Prior at least 6month -lasing cytopenia before dignosis	Prior documented MDS or MDS/MPN	Cytogenetics	Karyotype abnormality supporting AML-MRC including –7/del(7q)	Dysplasia > 50% in multilinages, supporting AML-MRC	Dysplasia > 10% in lineage(s), suppporting MDS
12	55/F	RAEB-2	Very High	Yes	Yes	47,XX,+del(1)(p?),+mar1[5]	Yes	Yes	Yes
13	62/M	RAEB-2	Very High	Yes	Yes	46,XY,dup(1)q21;q42)[49	No	Yes	Yes
14	46/M	APL with t(15;17)(q22;q12)	Favorable	No	No	46,XY,t(15;17)(q22;q12)[17]	No	No	No
15	35/F	APL with t(15;17)(q22;q12)	Favorable	No	No	46,XX,t(15;17)(q22;q12)[12]	No	No	No
16	39/M	AML with inv(16)(p13.1q22)	Favorable	No	No	46,XY, inv(16)(p13.1q22)[20]	No	No	No
17	53/F	APL with t(15;17)(q22;q12)	Favorable	No	No	46,XX,t(15;17)(q22;q12)[20]	No	No	No
18	37/M	APL with t(15;17)(q22;q12)	Favorable	No	No	46,XY,t(15;17)(q22;q12)[20]	No	No	No
19	70/F	AML with inv(16)(p13.1q22)	Favorable	No	No	48,XX,+9, inv(16)(p13.1q22), +22[11]	No	No	No
20	49/M	APL with t(15;17)(q22;q12)	Favorable	No	No	46,XY,t(15;17)(q22;q12)[20]	No	No	No
21	49/M	AML with t(8;21)(q22;q22)	Favorable	No	No	46,XY,t(8;21)(q22;q22)[12]	No	No	No
22	53/M	AML with inv(16)(p13.1q22)	Favorable	No	No	46,XY, inv(16)(p13.1q22) [12]	No	No	No
23	81/F	AML with mutated *NPM1*	Intermediate	No	No	46,XX[20]	No	No	No
24	43/M	AML with mutated *NPM1*	Intermediate	No	No	46,XY[20]	No	No	No
25	64/F	AML-NOS	Int	No	No	46,XX[20/20]	No	No	Yes
26	34/F	AML-NOS	Int	No	No	46,XX[20]	No	No	Yes
27	59/M	AML-NOS	Int	Yes	No	45,X,-Y[15]	No	No	Yes
28	78/M	AML-NOS	Int	No	No	46,XY,t(1;14)(q42;q32)[19]	No	No	Yes
29	35/F	AML-MRC	Int	Yes	Yes	46,XX[20]	No	No	Yes
30	75/M	AML-MRC	Int	Yes	Yes	46,XY[20]	No	No	Yes
31	81/M	AML-MRC	Int	Yes	Yes	46,XY[20]	No	Yes	Yes
32	76/M	AML-MRC	Adverse	Yes	Yes	44,XY,del(5)(q?), add(7)(q22), dic(10;19)(q11.2;q11) dic(16;17) (p11.2;p11.2)[13]	Yes	Yes	Yes

(Supplementary Table 1 continues)

(Supplementary Table 1 continued)

Patient (UPN)	Age/ Gender	Clinical diagnosis by WHO2008 criteria	IPSS-R (MDS) or MRC (AML) risk	Prior at least 6month -lasing cytopenia before dignosis	Prior documented MDS or MDS/MPN	Cytogenetics	Karyotype abnormality supporting AML-MRC including –7/del(7q)	Dysplasia > 50% in multilinages, supporting AML-MRC	Dysplasia > 10% in lineage(s), suppporting MDS
33	76/M	AML-MRC	Adverse	No	Yes	46,XY, del(7)(p?)[20]	Yes	No	Yes
34	46/M	AML-MRC	Adverse	Yes	Yes	46,XY,dup(3)(q21.q27)[20/20]	Yes	Yes	Yes
35	41/F	AML-MRC	Adverse	Yes	Yes	45,XX,-2,inv(q21q26.2)del(5)(q?), -7,add(8)(q13),add(17)(q11.2), +mar1[3]	Yes	Yes	Yes
36	58/M	AML-MRC	Adverse	Yes	Yes	45,XY,-7[11]	Yes	Yes	Yes
37	64/F	AML-MRC	Adverse	Yes	Yes	47,XX,+21,del(3)(q?)[19]	Yes	Yes	Yes
38	73/M	AML-MRC	Adverse	No	Yes	46,X,-Y,-7,del(9)(q?), add(17)(p11.2),+mar1,+,mar2[5]	Yes	No	Yes
39	65/M	AML-MRC	Adverse	Yes	Yes	43,XY,t(2;11)(p21;q23), del(5)(q?), -1,i(7)(q10), dic(13;22)(p11.2;p11.2), der(15)t(1;15)(q21;q22)ins(15;?) (q22;?),add(16)(p11.2),-18 [13]	Yes	Yes	Yes
40	68/F	AML-MRC	Adverse	Yes	Yes	46,XX[20]	Yes	Yes	Yes
41	72/F	AML-MRC	Adverse	No	No	45,XY,-7[7]	Yes	Yes	Yes
42	36/F	AML-MRC	Adverse	No	No	46,XX,add(16)(p11.2)[20]	Yes	Yes	Yes
43	65/M	AML-MRC	Adverse	No	No	44,XY,del(5)(q?),-7,-11,add(11)(q13),add(12)(p11.2), -14,add(15)(p11.2), -16,add(19)(q13.1),+2mar[1]	Yes	Yes	Yes
44	67/F	AML-MRC	Adverse	No	No	45,XX,del(5)(q?),add(14)(p11.2), -21,-22,+mar1[12]	Yes	No	Yes

Abbreviations: UPN, unique patient number; RCUD, refractory cytopenia with single lineage dysplasia; RCMD, refractory cytopenia with multilineage dysplasia; RARS, refractory anemia with ring sideroblasts; RAEB-1, refractory anemia with excess of

blasts-1; RAEB-2, refractory anemia with excess of blasts-2; MDS/MPN-u, myelodysplastic/myeloproliferative neoplasms, unclassifiable; CMML-1, chronic myelomonocytic leukemia-1; AML, acute myeloid leukemia; AML-NOS, AML-not otherwise specified; AML-MRC, AML with myelodysplasia-related changes; IPSS-R, revised international prognostic scoring system; MRC, prognostic scoring system by Medical Research Council Acute Myeloid Leukemia 10; Int, intermediate risk.

Supplementary Table 2. List of genes and exons included in the myeloid panel.

Function	Gene	Targeted exon	Function	Gene	Targeted exon
Splicesome	*SF3B1*	13–16	Signal Transduction	*CSF3R*	14–17
	SRSF2	1		*MPL*	10
	U2AF1	2 + 6		*NOTCH1*	26–28 + 34
	ZRSR2	full		*BRAF*	15
Chromatin Modifier	*ATRX*	8–10 & 17–31		*CBL*	8 + 9
	ASXL1	12		*CBLB*	9 + 10
	BCOR	full		*CBLC*	9 + 10
	BCORL1	full		*GNAS*	8 + 9
	EZH2	full		*HRAS*	2 + 3
	KDM6A	full		*KRAS*	2 + 3
	MLL	5–8		*NRAS*	2 + 3
Cohesin Complex	*RAD21*	full		*MYD88*	3–5
	STAG2	full		*PTPN11*	3 + 13
	SMC3	2 , 11 , 16 + 17	Tyrosine Kinase	*FLT3*	14 + 15 + 20
	SMC1A	10, 13 , 19 ,23 ,25 + 28		*KIT*	2, 8–11, 13 + 17
Transcription Factor	*CEBPA*	full		*PDGFRA*	12, 14, 18
	CUX1	full		*ABL1*	4–6
	ETV6	full		*JAK2*	12 + 14
	GATA1	2		*JAK3*	13
	GATA2	2–6	DNA Methylation	*DNMT3A*	full
	IKZF1	full		*IDH1*	4
	RUNX1	full		*IDH2*	4
Tumor Suppressor	*PTEN*	5+7		*TET2*	3–11
	TP53	2-11	Cell Cycle	*CDKN2A*	full
	WT1	7+9		*FBXW7*	9 + 10 + 11
Cellular Process	*CALR*	9			
	NPM1	12			
	PHF6	full			
	SETBP1	4			

Supplementary Table 3. Overview of cell lineage level-targeted sequencing workflow.

STEP 1. Flowcytometry sorting by FACS AriaII
- Blast/Granulocyte/T cell
STEP 2. Targeted deep sequencing
- 20ng DNA from tumor (blast and granulocyte) and control (T cell and/or buccal swab)
- Library prepared by Trusight Myeloid Panel (Illumina)
- Sequencing by reagents V3 (600 cycles) chemistry with 250 bp paired-end reads on MiSeq platform (Illumina)
STEP 3. Single nucleotide variants (SNVs) calling and structural variants (SVs) calling by two independent pipelines:
- SNVs calling
- Sequence-alignment (hg19) and SNVs calling by MiSeq Reporter pipeline
- Sequence-alignment (hg19) by Burrows-Wheeler Aligner mem (BWA) and SNVs calling by Genomon2 pipeline
- SVs calling
- Sequence-alignment (hg19) by BWA and SVs calling by Genomon-SV pipeline
- Sequence-alignment (hg19) by BWA and SVs calling by Pindel pipeline
STEP 4. Annotation and filtering variants by in-house pipeline
- 1. Annotation of all variants in STEP3 performed by in-house pipelines
- 2. Discard variants following high stringency filtering:
- Phred quality score less than 40 and/or Illumina QC filter: not "passing" results
- Indel repeat greater than 8
- Variant allele frequency (VAF) < 2%
- Less than 50x total read depth and/or less than 10 alternate support reads
- Strand biased support reads
- In sequencing error prone region by in-house database
- Regions of segmental duplications
- Synonyomous, intronic (excluding splicing region), 3'-untranslated region (UTR), and 5'-UTR
- Observed in controls
- Tumor/Control (T cell and/or buccal swab) with p value ≥ 0.05 by Fisher's exact test
- A minor allele frequency ≥ 1% in single nucleotide polymorphism (SNP) database (excluding clinically associated SNP)
- Database: 1000 Genomes Project/NHLBI-ESP 5400 exomes/ToMMo/HGVD/dbSNP131nonflagged
- Functionally benign
- SIFT score > 0.050 or PolyPhen2 score < 0.957 (excluding clinically associated SNVs/SVs registered in database)
- Database: ClinVar or HGMD or COSMIC or cBioPortal or Tumor Portal or ICGC
- 3. Pick up the overlapping SNVs (MiSeq Reporter and Genomon2) and SVs (Genomon-SV and Pindel).
- 4. Pick up only reliable variants by IGV inspection.
- 5. Rescue variants discarded in STEP 4-2 or 4-3, which were highly reliable and highly deleterious by manual inspection.
(e.g., germline SNVs and SVs with reliable support reads, clinical association of which were reported in ClinVar or HGMD database)
- 6. Pick up "overlapping" and "non-overlapping driver" variants following VAF criteria:
- Tier 1 "overlapping driver"
- VAF ≥ 20% in blast fraction with the same variant observed in granulocyte fraction (VAF ≥ 2%)
- Tier 2 "overlapping driver" (in case Tier1 overlapping drivers not available)
-10% < VAF < 20% in blast fraction with the same variant observed in granulocyte fraction (VAF ≥ 2%)
- Tier 1 "non-overlapping driver"
- VAF ≥ 20% in blast fraction with the same variant not observed in granulocyte fraction
- Tier 2 "non-overlapping driver" (in case Tier1 non-overlapping drivers not available)
- 10% < VAF < 20% in blast fraction with the same variant not observed in granulocyte fraction
STEP 5. Validation of Tier 1 and Tier 2 variants in STEP4-5 using residual unfractionated samples in STEP1 either by:
- 1. Sanger sequencing
- 2. Droplet digital PCR (ddPCR) or another sequencing modality if Sanger sequencing was not available

Supplementary Table 4. Validated mutations identified in the patients in the study.

UPN	Chr	Start	End	Ref	Alt	Gene	Class	Amino acid change	Databse	SIFT	Poly-phen2	Mutation Taster	Driver type	Confirmed by
1	chr2	198266834	198266834	T	C	*SF3B1*	nonsynonymous SNV	p.K700E	COSM84677	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
1	chr4	106164863	106164863	T	G	*TET2*	nonsynonymous SNV	p.L1244R	COSMIC+-2	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
2	chr20	31022441	31022441	-	G	*ASXL1*	frameshift insertion	p.G646fs	COSM1411076	NA NA	NA NA	NA NA	Overlapping	Sanger
2	chr4	55599321	55599321	A	T	*KIT*	nonsynonymous SNV	p.D816V	COSM1314	0.00 D	0.96 D	1.00 A	Overlapping	Sanger
3	chr1	115258747	115258747	C	T	*NRAS*	nonsynonymous SNV	p.G12D	COSM564	0.00 D	0.40 B	1.00 D	Overlapping	Sanger
3	chr21	36252878	36252878	T	C	*RUNX1*	nonsynonymous SNV	p.R162G	COSM24718	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
4	chr2	25523008	25523008	-	G	*DNMT3A*	frameshift insertion	p.P59fs	COSMIC+-2	NA NA	NA NA	NA NA	Non-overlapping	Target seq
4	chr20	31022438	31022444	CGGAGGG	-	*ASXL1*	frameshift deletion	p.I641fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
5	chr9	21971161	21971161	T	C	*CDKN2A*	nonsynonymous SNV	p.H66R	COSM14253	0.26 T	0.01 B	1.00 D	Overlapping	Sanger
6	chr21	44524456	44524456	G	A	*U2AF1*	nonsynonymous SNV	p.S34F	COSM166866	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
7	chr11	119149376	119149376	C	T	*CBL*	stopgain	p.R462X	COSM34079	0.72 T	NA NA	1.00 A	Overlapping	Sanger
7	chr2	25467099	25467099	G	-	*DNMT3A*	stopgain	p.Y592X	COSM231573	NA NA	NA NA	NA NA	Overlapping	Sanger
7	chr2	25467101	25467101	A	T	*DNMT3A*	nonsynonymous SNV	p.Y592N	COSMIC+-2	0.05 D	0.85 P	1.00 D	Overlapping	Sanger
7	chr2	198266834	198266834	T	C	*SF3B1*	nonsynonymous SNV	p.K700E	COSM84677	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
7	chr7	148512035	148512035	C	-	*EZH2*	frameshift deletion	p.C548fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
8	chr20	31022441	31022441	-	G	*ASXL1*	frameshift insertion	p.G646fs	COSM1411076	NA NA	NA NA	NA NA	Overlapping	Target seq
9	chr2	198267484	198267484	G	A	*SF3B1*	nonsynonymous SNV	p.R625C	COSM110696	0.00 D	0.99 D	1.00 D	Overlapping	Sanger
9	chr4	106196914	106196914	G	-	*TET2*	frameshift deletion	p.M1749fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Target seq
10	chr12	12006443	12006449	AACTCT	-	*ETV6*	frameshift deletion	p.G137fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger

(Supplementary Table 4 continues)

(Supplementary Table 4 continued)

UPN	Chr	Start	End	Ref	Alt	Gene	Class	Amino acid change	Databse	SIFT	Poly-phen2	Mutation Taster	Driver type	Confirmed by
10	chr7	101882818	101882818	C	G	*CUX1*	nonsynonymous SNV	p.L1281V	NO	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
11	chr17	7578190	7578190	T	C	*TP53*	nonsynonymous SNV	p.Y220C	COSM99718	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
12	chr17	7579882	7579882	C	G	*TP53*	nonsynonymous SNV	p.E11Q	COSM327260	0.03 D	0.97 D	1.00 N	Overlapping	Sanger/ddPCR
12	chr18	42531907	42531907	G	A	*SETBP1*	nonsynonymous SNV	p.D868N	COSM1318400	0.00 D	1.00 D	1.00 D	Non-overlapping	Target seq
12	chr21	44514777	44514777	T	G	*U2AF1*	nonsynonymous SNV	p.Q157P	COSM1318797	0.00 D	0.74 P	1.00 D	Overlapping	ddPCR/Target seq
13	chr21	44524456	44524456	G	A	*U2AF1*	nonsynonymous SNV	p.S34F	COSM166866	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
14	chr4	106193989	106193992	ACAG	-	*TET2*	frameshift deletion	p.N1484fs	COSMIC+-2	NA NA	NA NA	NA NA	Non-overlapping	Target seq
15	chr11	32417908	32417911	CCGA	-	*WT1*	frameshift deletion	p.S313fs	COSMIC+-2	NA NA	NA NA	NA NA	Non-overlapping	Target seq
15	chr11	32417917	32417923	CAAGAGT	-	*WT1*	frameshift deletion	p.T309fs	COSMIC+-2	NA NA	NA NA	NA NA	Non-overlapping	Target seq
16	chr1	115258747	115258747	C	G	*NRAS*	nonsynonymous SNV	p.G12A	COSM565	0.00 D	0.59 P	1.00 D	Non-overlapping	Target seq
16	chr4	106197287	106197287	G	A	*TET2*	nonsynonymous SNV	p.E1874K	COSM5945072	0.00 D	1.00 D	1.00 D	Non-overlapping	Target seq
17	chr13	28608251	28608251	-	TGAGATCA TATTCATAT TCTC	*FLT3*	nonframeshift insertion	p.K602delins REYEYDLK	COSM1317921	NA NA	NA NA	NA NA	Non-overlapping	Sanger
18	chr13	28592642	28592642	C	A	*FLT3*	nonsynonymous SNV	p.D835Y	COSM783	0.00 D	1.00 D	1.00 D	Non-overlapping	Sanger
21	chr4	55599320	55599320	G	C	*KIT*	nonsynonymous SNV	p.D816H	COSM1311	0.01 D	0.27 B	1.00 D	Non-overlapping	Sanger
22	chr13	28592641	28592641	T	G	*FLT3*	nonsynonymous SNV	p.D835A	COSM27650	0.00 D	1.00 D	1.00 D	Non-overlapping	Sanger
23	chr4	106182956	106182956	-	A	TET2	frameshift insertion	p.L1332fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
23	chr4	106196537	106196537	C	T	TET2	stopgain	p.Q1624X	COSM42020	0.22 T	NA NA	1.00 D	Overlapping	Sanger
23	chr5	170837542	170837542	-	TCTG	NPM1	frameshift insertion	p.W288fs	COSM17559	NA NA	NA NA	NA NA	Non-overlapping	Sanger
24	chr2	25457243	25457243	G	A	DNMT3A	nonsynonymous SN\	p.R882C	COSM53042	0.00 D	1.00 D	1.00 D	Overlapping	Sanger

(Supplementary Table 4 continues)

(Supplementary Table 4 continued)

UPN	Chr	Start	End	Ref	Alt	Gene	Class	Amino acid change	Databse	SIFT	Poly-phen2	Mutation Taster	Driver type	Confirmed by
24	chr5	170837542	170837542	-	TCTG	NPM1	frameshift insertion	p.W288fs	COSM17559	NA NA	NA NA	NA NA	Non-overlapping	Sanger
24	chr15	90631934	90631934	C	T	IDH2	nonsynonymous SN\	p.R140Q	COSM41590	0.00 D	1.00 D	1.00 D	Non-overlapping	Sanger
25	chr13	28608246	28608246	-	CCATTTGA GATCATATT CATATTCTC T	*FLT3*	nonframeshift insertion	p.E604delinsREY EYDLKWE	COSMIC+-2	NA NA	NA NA	NA NA	Non-overlapping	Sanger
25	chr4	106156729	106156729	C	T	*TET2*	stopgain	p.R544X	COSM41850	1.00 T	NA NA	1.00 D	Overlapping	Target seq
25	chr4	106157755	106157755	C	T	*TET2*	stopgain	p.Q886X	COSM100053	0.07 T	NA NA	1.00 D	Overlapping	Sanger
25	chr5	170837543	170837543	-	TCTG	*NPM1*	frameshift insertion	p.W288fs	COSM158604	NA NA	NA NA	NA NA	Non-overlapping	ddPCR
26	chr11	32417913	32417913	-	GTACAAGA	*WT1*	frameshift insertion	p.R312fs	COSM27300	NA NA	NA NA	NA NA	Overlapping	Target seq
26	chr12	112926852	112926852	C	A	*PTPN11*	nonsynonymous SNV	p.P491H	In HGMD	0.35 T	0.84 P	1.00 D	Overlapping	Sanger
26	chr2	25457242	25457242	C	T	*DNMT3A*	nonsynonymous SNV	p.R882H	COSM52944	0.03 D	0.04 B	1.00 D	Overlapping	ddPCR
26	chr21	36164567	36164567	-	T	*RUNX1*	frameshift insertion	p.S436fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Target seq
27	chr15	90631934	90631934	C	T	*IDH2*	nonsynonymous SNV	p.R140Q	COSM41590	0.00 D	1.00 D	1.00 D	Overlapping	Sanger/ddPCR
27	chr3	105438902	105438902	T	C	*CBLB*	nonsynonymous SNV	p.N466D	COSM149439	0.00 D	0.04 B	1.00 D	Overlapping	Sanger
27	chr5	170837543	170837543	-	TCTG	*NPM1*	frameshift insertion	p.W288fs	COSM158604	NA NA	NA NA	NA NA	Non-overlapping	ddPCR
28	chr1	36933434	36933434	G	A	*CSF3R*	nonsynonymous SNV	p.T618I	COSM1737962	0.06 T	1.00 D	0.68 D	Overlapping	Sanger
28	chr19	33793138	33793138	G	-	*CEBPA*	frameshift deletion	p.I62fs	COSM18534	NA NA	NA NA	NA NA	Overlapping	Target seq
28	chr4	106190855	106190855	G	A	*TET2*	nonsynonymous SNV	p.C1378Y	COSM211731	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
28	chr4	106190874	106190874	C	A	*TET2*	nonsynonymous SNV	p.D1384E	COSMIC+-2	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
29	chr3	128200134	128200145	CCTTCTTC ATGG	-	*GATA2*	nonframeshift deletion	p.T387_E391del	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
30	chr17	7579882	7579882	C	G	*TP53*	nonsynonymous SNV	p.E11Q	COSM327260	0.03 D	0.97 D	1.00 N	Overlapping	Sanger/Target seq

(Supplementary Table 4 continues)

(Supplementary Table 4 continued)

UPN	Chr	Start	End	Ref	Alt	Gene	Class	Amino acid change	Databse	SIFT	Poly-phen2	Mutation Taster	Driver type	Confirmed by
30	chr17	74732959	74732959	G	C	*SRSF2*	nonsynonymous SNV	p.P95R	COSM211661	0.26 T	0.99 D	1.00 D	Overlapping	Target seq
30	chr21	36231792	36231792	C	T	*RUNX1*	nonsynonymous SNV	p.D198N	COSM24721	0.00 D	1.00 D	1.00 D	Overlapping	Sanger/Target seq
30	chr4	106162564	106162564	A	T	*TET2*	nonsynonymous SNV	p.I1160F	COSMIC+-2	0.00 D	0.99 D	1.00 D	Overlapping	Sanger/Target seq
30	chr4	106196937	106196937	-	GGCCC	*TET2*	frameshift insertion	p.H1757fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Target seq
30	chr9	21971161	21971161	T	C	*CDKN2A*	nonsynonymous SNV	p.H66R	COSM14253	0.26 T	0.01 B	1.00 D	Overlapping	Sanger/Target seq
30	chrX	39933575	39933575	G	A	*BCOR*	stopgain	p.R342X	COSM518558	1.00 T	NA NA	1.00 A	Overlapping	Sanger/Target seq
31	chr19	33793183	33793183	-	G	*CEBPA*	frameshift insertion	p.P45fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Target seq
31	chr20	31022442	31022443	GG	-	*ASXL1*	frameshift deletion	p.G646fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
31	chr3	128202759	128202759	G	A	*GATA2*	nonsynonymous SNV	p.L321F	COSM249853	0.01 D	1.00 D	1.00 D	Overlapping	Target seq
31	chr4	55599320	55599320	G	T	*KIT*	nonsynonymous SNV	p.D816Y	Mutation Id:	0.00 D	0.94 D	1.00 D	Overlapping	Sanger
31	chr4	106157998	106157998	C	-	*TET2*	frameshift deletion	p.Q967fs	COSM1411076	NA NA	NA NA	NA NA	Overlapping	Sanger
31	chr4	106197366	106197366	T	G	*TET2*	nonsynonymous SNV	p.V1900G	COSMIC+-2	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
32	chr17	7578413	7578413	C	T	*TP53*	nonsynonymous SNV	p.V80M	COSM98966	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
32	chr19	33793253	33793253	G	-	*CEBPA*	frameshift deletion	p.P23fs	COSM18544	NA NA	NA NA	NA NA	Non-overlapping	Sanger/ddPCR
32	chr4	106157832	106157832	-	GC	*TET2*	frameshift insertion	:p.A911fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
33	chr1	115258747	115258747	C	T	*NRAS*	nonsynonymous SNV	p.G12D	COSM564	0.00 D	0.40 B	1.00 D	Overlapping	Sanger
33	chr12	12022592	12022592	-	C	*ETV6*	frameshift insertion	p.Y233fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
33	chr17	74732959	74732959	G	A	*SRSF2*	nonsynonymous SNV	p.P95L	COSM146288	0.10 T	0.46 P	1.00 D	Overlapping	Sanger
33	chr18	42531907	42531907	G	A	*SETBP1*	nonsynonymous SNV	p.D868N	COSM1318400	0.00 D	1.00 D	1.00 D	Overlapping	Sanger

(Supplementary Table 4 continues)

(Supplementary Table 4 continued)

UPN	Chr	Start	End	Ref	Alt	Gene	Class	Amino acid change	Databse	SIFT	Poly-phen2	Mutation Taster	Driver type	Confirmed by
33	chr20	31022403	31022425	CACCACTG CCATAGAG AGGCGGC	-	*ASXL1*	frameshift deletion	p.E630fs	COSM3720455	NA NA	NA NA	NA NA	Overlapping	Sanger
33	chr7	148523705	148523705	G	A	*EZH2*	stopgain	p.Q250X	COSMIC+-2	0.63 T	NA NA	1.00 A	Overlapping	Sanger
34	chr1	115258747	115258747	C	T	*NRAS*	nonsynonymous SNV	p.G12D	COSM564	0.00 D	0.40 B	1.00 D	Overlapping	Sanger/ddPCR
34	chr11	119148990	119148990	T	C	*CBL*	nonsynonymous SNV	p.C404R	COSM34059	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
34	chr18	42531907	42531907	G	A	*SETBP1*	nonsynonymous SNV	p.D868N	COSM1318400	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
34	chr21	36164630	36164630	C	G	*RUNX1*	nonsynonymous SNV	p.Q415H	COSMIC+-2	0.06 T	1.00 D	1.00 D	Overlapping	Sanger
34	chr21	44514777	44514777	T	C	*U2AF1*	nonsynonymous SNV	p.Q157R	COSM1724986	0.00 D	0.99 D	1.00 D	Overlapping	Sanger
34	chr21	44524456	44524456	G	A	*U2AF1*	nonsynonymous SNV	p.S34F	COSM166866	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
35	chr2	198266834	198266834	T	C	*SF3B1*	nonsynonymous SNV	p.K700E	COSM84677	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
36	chr2	25463175	25463177	AGA	-	*DNMT3A*	nonframeshift deletion	p.F772_L773del	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
36	chr2	209113113	209113113	G	A	*IDH1*	nonsynonymous SNV	p.R132C	COSM28747	0.00 D	0.13 B	1.00 D	Overlapping	ddPCR
36	chr20	31022441	31022441	-	G	*ASXL1*	frameshift insertion	p.G646fs	COSM1411076	NA NA	NA NA	NA NA	Overlapping	Sanger
36	chr21	36164697	36164697	C	-	*RUNX1*	frameshift deletion	p.G393fs	COSMIC+-2	NA NA	NA NA	NA NA	Non-overlapping	Sanger
37	chr2	198266834	198266834	T	C	*SF3B1*	nonsynonymous SNV	p.K700E	COSM84677	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
37	chr21	36231792	36231792	C	T	*RUNX1*	nonsynonymous SNV	p.D198N	COSM24721	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
38	chr17	7577120	7577120	C	T	*TP53*	nonsynonymous SNV	p.R273H	COSM99729	0.01 D	0.99 D	1.00 A	Overlapping	Sanger
38	chr21	36252866	36252866	G	A	*RUNX1*	stopgain	p.R166X	COSM24769	1.00 T	NA NA	1.00 A	Non-overlapping	Sanger
38	chr21	44514777	44514777	T	C	*U2AF1*	nonsynonymous SNV	p.Q157R	COSM1724986	0.00 D	0.99 D	1.00 D	Overlapping	Sanger
38	chr9	5073770	5073770	G	T	*JAK2*	nonsynonymous SNV	p.V617F	COSM12600	0.00 D	0.93 D	1.00 D	Overlapping	Sanger

(Supplementary Table 4 continues)

(Supplementary Table 4 continued)

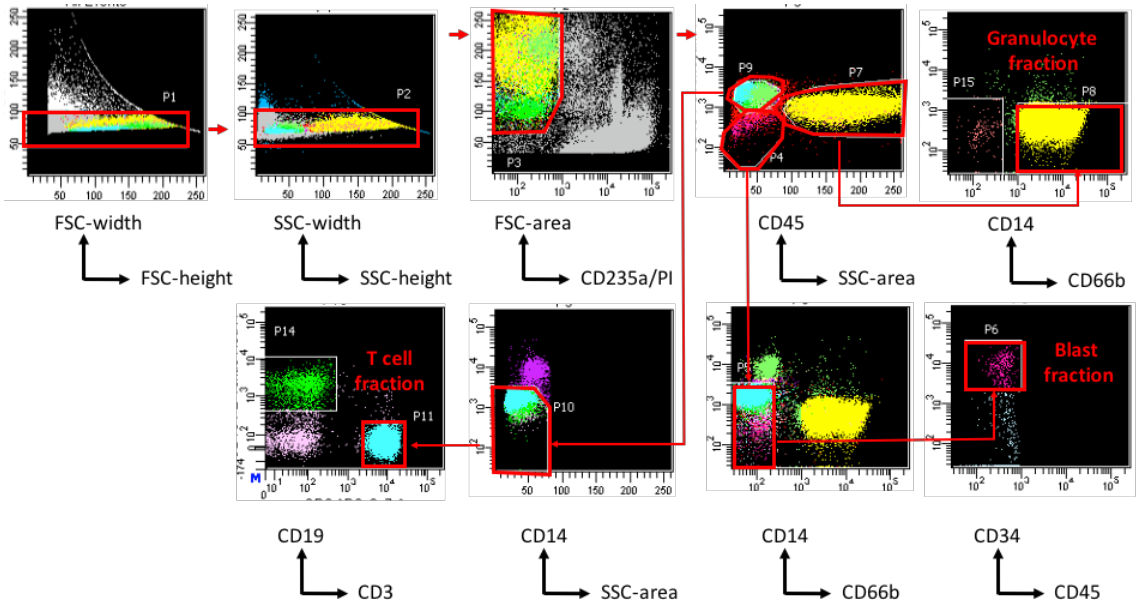
UPN	Chr	Start	End	Ref	Alt	Gene	Class	Amino acid change	Databse	SIFT	Poly-phen2	Mutation Taster	Driver type	Confirmed by
39	chr17	7578458	7578458	G	C	*TP53*	nonsynonymous SNV	p.R65G	COSM318153	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
39	chr17	7579882	7579882	C	G	*TP53*	nonsynonymous SNV	p.E11Q	COSM327260	0.03 D	0.97 D	1.00 N	Overlapping	Sanger
40	chr1	115258747	115258747	C	T	*NRAS*	nonsynonymous SNV	p.G12D	COSM564	0.00 D	0.40 B	1.00 D	Overlapping	Sanger/ddPCR
40	chr21	44514777	44514777	T	C	*U2AF1*	nonsynonymous SNV	p.Q157R	COSM1724986	0.00 D	0.99 D	1.00 D	Overlapping	Sanger
40	chr21	44524456	44524456	G	A	*U2AF1*	nonsynonymous SNV	p.S34F	COSM166866	0.00 D	1.00 D	1.00 D	Overlapping	Sanger/ddPCR
41	chr4	106157946	106157946	-	G	*TET2*	frameshift insertion	p.H949fs	COSMIC+-2	NA NA	NA NA	NA NA	Overlapping	Sanger
41	chr4	106180796	106180796	G	A	*TET2*	nonsynonymous SNV	p.G1275E	COSM110791	0.00 D	1.00 D	1.00 D	Overlapping	Sanger
42	chr1	115258747	115258747	C	T	*NRAS*	nonsynonymous SNV	p.G12D	COSM564	0.00 D	0.40 B	1.00 D	Overlapping	ddPCR/Target seq
42	chr11	32417914	32417914	-	T	*WT1*	frameshift insertion	p.R312fs	COSM49053	NA NA	NA NA	NA NA	non-Overlapping	Target seq
42	13	28608487	28608487	-	AGACAAACACCAATTGTTGCATAGAATGAGATGTTGTCTTGGATGAAAGGGAAGGGGCCTGCAACAAAAGAGAGGACTCCAATA	*FLT3*	nonframeshift insertion	p.L552delinsLIGVLSFVAGPFPFIQDNISFYATIGVCL	COSMIC+-2	NA NA	NA NA	NA NA	non-overlapping	Target seq
43	chr17	7577545	7577545	T	C	*TP53*	nonsynonymous SNV	p.M246V	COSM251429	0.00 D	0.99 D	1.00 D	Overlapping	Sanger
43	chr17	7578406	7578406	C	T	*TP53*	nonsynonymous SNV	p.R175H	COSM3355994	0.00 D	1.00 D	1.00 A	Overlapping	Sanger
44	chr17	7574018	7574018	G	A	*TP53*	nonsynonymous SNV	p.R337C	COSM117591	0.00 D	0.34 B	1.00 D	Overlapping	Sanger
44	chr17	7577504	7577504	G	-	*TP53*	frameshift deletion	p.S260fs	COSM46112	NA NA	NA NA	NA NA	Overlapping	Sanger

Abbreviations: UPN, unique patient number; Chr, chromosome; Start, variant base start position in hg19 coordinates; End, variant base end position in hg19 coordinates; Ref, reference base represented in hg19 coordinates; Alt, variant base in hg19 coordinates; Gene, gene symbol defined by the HUGO Gene Nomenclature Committee; Class, classification based on amino acid change pattern; Amino acid change, predicted

amino acid change for the variation; Database, evidence of variant by database registration (e.g., COSMIC ID); SIFT, Score (left column) and Deleterious (D)/Tolerated (T) status (right column) assigned by SIFT; Polyphen2, Score (left column) and Deleterious (D)/Probably deleterious (P)/Benign (B) status (right column) assigned by Polyphen2; Mutation Taster, Score (left column) and Deleterious (D)/Tolerated(T) status (right column) assigned by Mutation Taster; Driver type, overlapping status of mutation; Confirmed by: modality used for mutation confirmation; SNV, single nucleotide variation; COSMIC+-2; Variation identified not at the specific site, but within 2 bp, in COSMIC; Sanger, sanger sequencing; ddPCR, droplet digital PCR; WES, whole exome sequencing; RNA seq, RNA sequencing; Target seq, targeted sequencing using different probe; NA, not available.

Supplementary Figures and Figure Legends

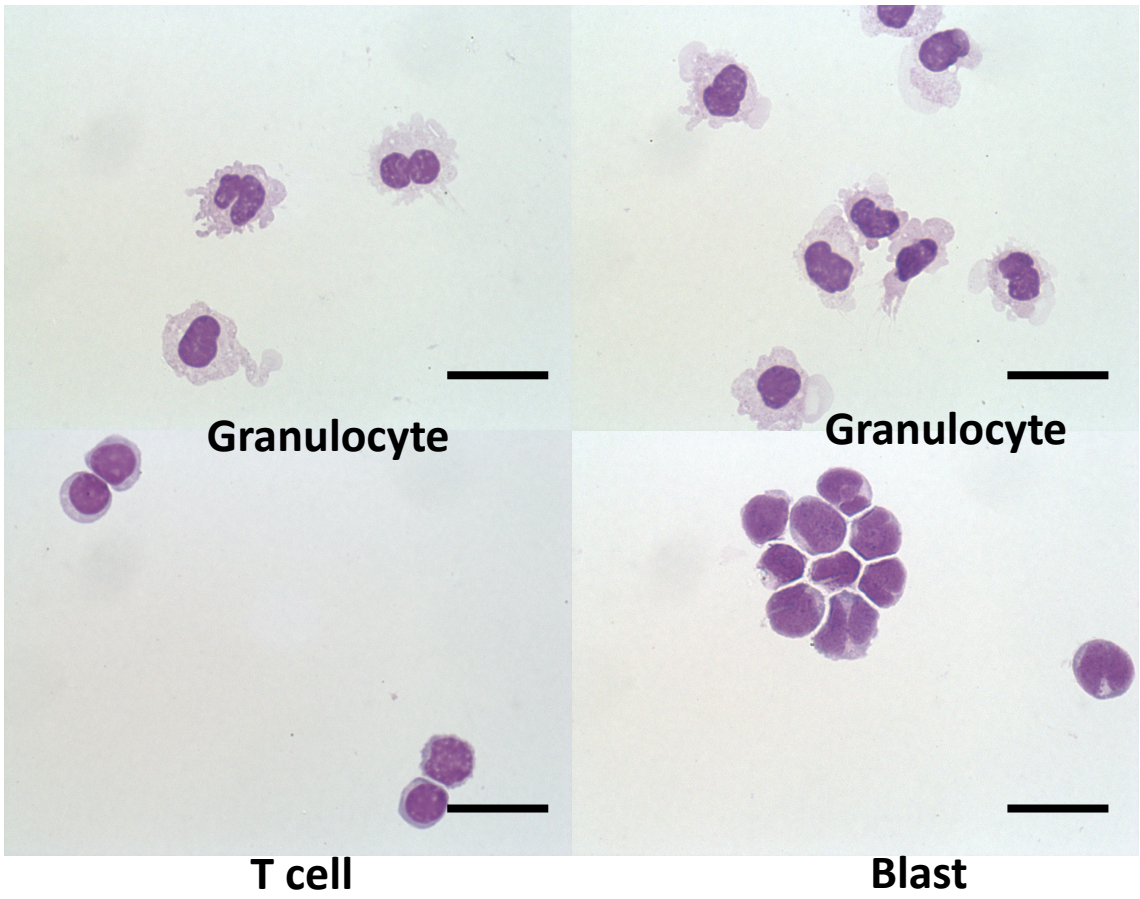
Figure S1.



Supplementary Figure 1. Representative sorting plot of granulocyte, blast, and T cells in a patient with myelodysplastic syndrome (unique patient number (UPN10)).

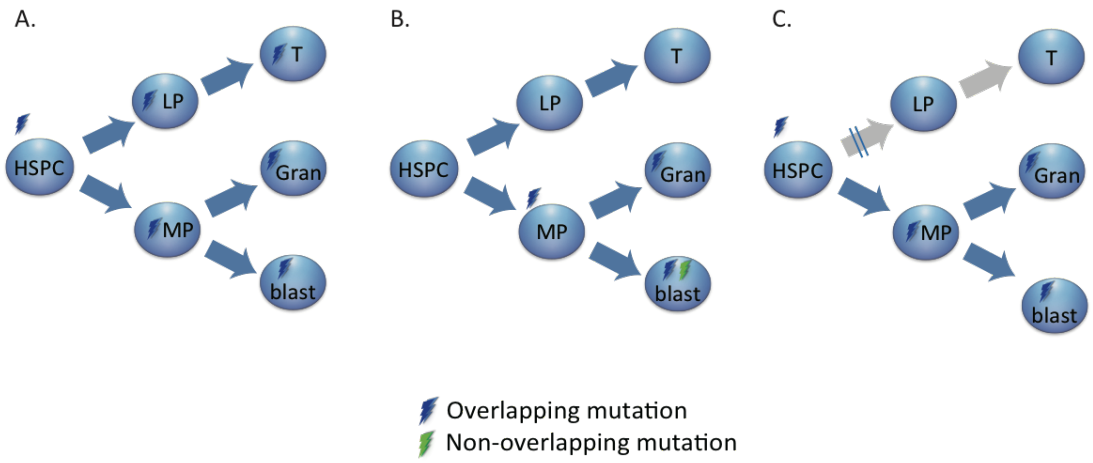
Abbreviations: FSC, forward scatter; SSC, side scatter.

Figure S2.



Supplementary Figure 2. Representative images of sorted trios in a patient with AML-MRC (UPN37) showing granulocytes, blasts, and T cells by cytospin followed by May-Grünwald-Giemsa staining. Scale bar indicate 20 μm.

Figure S3.

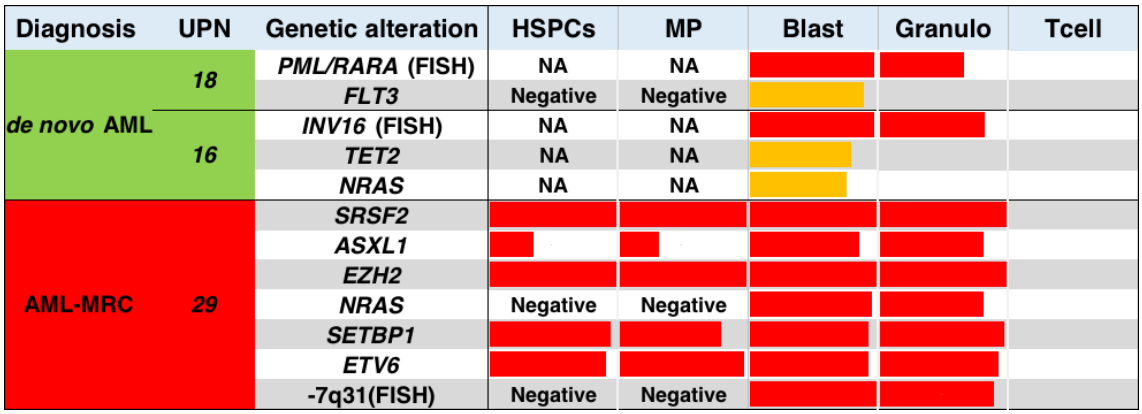


Supplementary Figure 3. Model depicting clonal origin of overlapping mutation.

The presence of overlapping mutation is indicated by blue thunder allows; non-overlapping mutation, green thunder allows. The dashed arrow with double strikethrough indicates that further lineage differentiation is inhibited. (A) If a mutation in the blast fraction is shared by both the granulocyte and T cell fraction, then presumable clonal origin is hematopoietic stem/multipotent progenitor cells (HSPC). If a mutation in the blast fraction is shared only by granulocyte, then presumable clonal

origin is either (B) myeloid progenitor (MP) or, alternatively, (C) HSPC, in which differentiation to lymphoid lineage might be inhibited. **Abbreviations**: Tc, T cell; Gran, granulocyte; LP, lymphoid progenitor cell.

Figure S4.



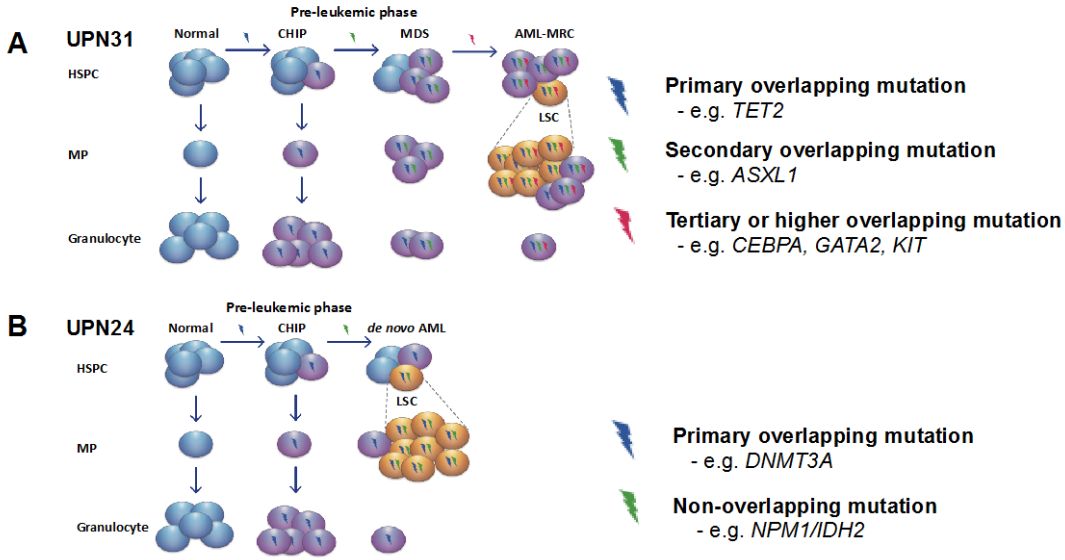
Supplementary Figure 4. Result of mutational screening of hematopoietic stem/multipotent progenitor fractions, along with trio fractions (granulocyte, blast, and T cell fractions) in representative cases of *de novo* AML and AML-MRC.

Targeted deep sequencing combined with fluorescence in situ hybridization (FISH) analysis were performed. Overlapping and non-overlapping genetic alterations are shown as red and yellow rectangles, respectively. The variant allele frequency of each mutation or the frequency of cytogenetic alteration is indicated by the horizontal width of the rectangle, with a maximum value set at 50% or 100%, respectively.

Abbreviations: UPN, unique patient number; NA, not applicable; Negative, genetic alteration not detected either by FISH or CL; HSPCs, hematopoietic stem/multipotent

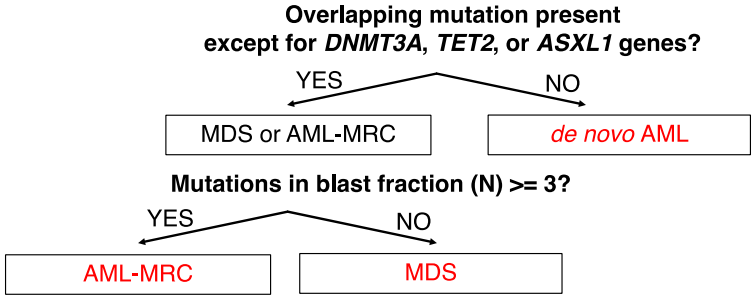
progenitor cells-enriched fraction (CD34+CD38-); MP, myeloid progenitor enriched fraction (CD34+CD38+); Granulo, granulocyte.

Figure S5.



Supplementary Figure 5. Proposed model depicting (B) MDS and AML-MRC clonal ontogeny in contrast to (A) *de novo* AML. (A) Shown is an inferred model of UPN24 depicting pre-leukemic phase (CHIP) initiated by a single overlapping mutation, which rapidly evolved into *de novo* AML by acquisition of non-overlapping mutations. (B) Shown is an inferred model of UPN31 depicting pre-leukemic phase including CHIP and MDS which gradually evolved into AML-MRC by stepwise acquisition of overlapping mutations. The thunder allows indicates mutations in each fractions. Abbreviations: HSPC, hematopoietic stem/multipotent progenitor cells; MP, myeloid progenitors.

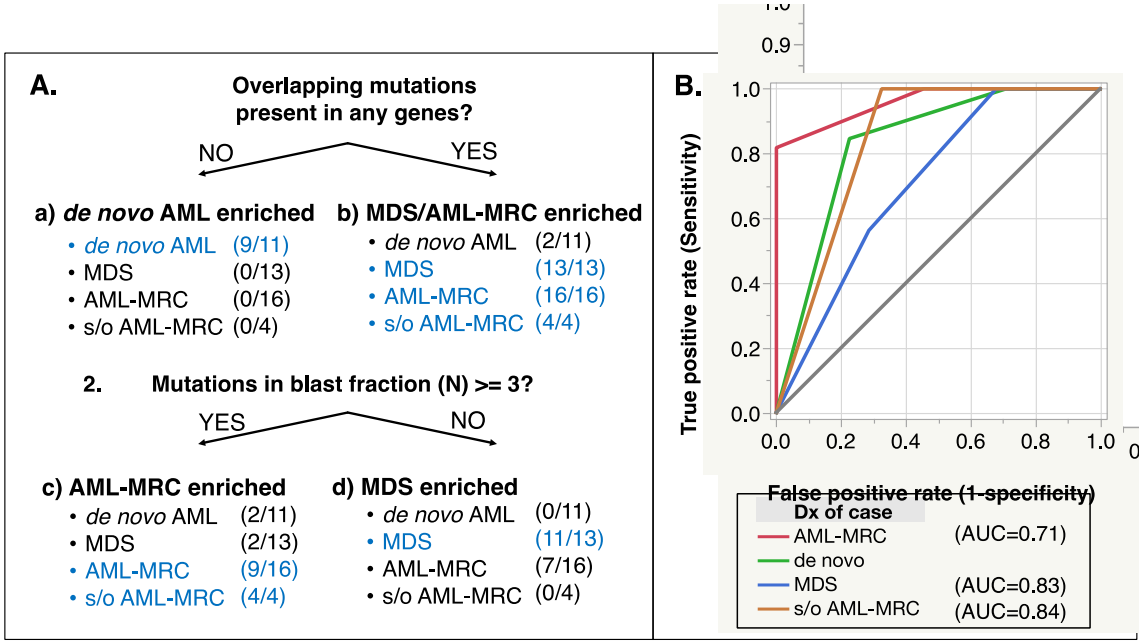
Figure S6.



Supplementary Figure 6.

Shown is a hierarchical classification model for the three disease subsets including 1) *de novo* AML, 2) AML-MRC, and 3) MDS, based on recursive partitioning analysis.

Figure S7.



Supplementary Figure 7. Disease classification model according to two mutational indices identified by cell lineage level-targeted sequencing. (A) Shown is the result of a hierarchical classification model for three disease subsets when the mutational index, the presence of overlapping mutation except for *DNMT3A*, *TET2*, and *ASXL1*, could be removed in relative to Figure 7. In each subset generated by the partitioning analysis, the enriched diseases are highlighted by blue characters. (B) Note that C-statistics (AUC) of AML-MRC and *de novo* AML is slightly lower compared to the result obtained in Figure 7B.