GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies

Xinan Wang,[1] Hideki Muramatsu,[1] Yusuke Okuno,[1] Hirotoshi Sakaguchi,[1] Kenichi Yoshida,[2] Nozomu Kawashima,[1] Yinyan Xu,[1] Yuichi Shiraishi,[3] Kenichi Chiba,[3] Hiroko Tanaka,[3] Shoji Saito,[4] Yozo Nakazawa,[4] Taro Masunari,[5] Tadashi Hirose,[6] Shaimaa Elmahdi,[1] Atsushi Narita,[1] Sayoko Doisaki,[1] Olfat Ismael,[1] Hideki Makishima,[7] Asahito Hama,[1] Satoru Miyano,[3,8] Yoshiyuki Takahashi,[1] Seishi Ogawa,[2] and Seiji Kojima[1]

[1]Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan; [2]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Japan; [3]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Japan; [4]Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan; [5]Department of Hematology, Chugoku Central Hospital, Fukuyama, Japan; [6]Department of Hematology, Kawasaki Medical School, Okayama, Japan; [7]Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA; and [8]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Japan.

Correspondence: kojimas@med.nagoya-u.ac.jp
doi:10.3324/haematol.2015.127092

Supplementary Information

Supplementary Methods pp. 2–4

5 **Supplementary Figures** pp. 5–7

Suppl Figure 1. Somatic/germline origin of GATA2 mutations

Suppl Figure 2. Haplotype analysis for *GATA2* mutations

Suppl Figure 3. Splice cassette analysis

10

Supplementary Tables pp. 8–12

Suppl Table 1. Summary of mutational analysis in *GATA2*

Suppl Table 2. Target genes for next-generation sequencing

15 Suppl Table 3. Detected mutations

Suppl Table 4. PCR primers for *GATA2*

References p. 13

Supplementary Methods

Patients

Three families with familial MDS which were consulted to our department for mutation analysis of several MDS-related genes including *GATA2* were enrolled in this study. A total of 75 Japanese pediatric patients (<18 years old) with sporadic AML, 75 patients with AA, 96 patients with JMML were also included in this study. Informed consent was obtained from patients or their guardians. This study was conducted in accordance with the Helsinki Declaration and was approved by the ethics committee of the Nagoya University Graduate School of Medicine.

DNA extraction

Genomic DNA was extracted from peripheral blood or bone marrow mononuclear cells using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). In patient 3, Genomic DNA was extracted from formalin-fixed, paraffin-embedded tissue using QIAamp DNA FFPE Tissue Kit (Qiagen).

Cell separation

Cells were stained with phycoerythrin-conjugated anti-CD3 (Cat no. A07747, Beckman Coulter, Miami, FL) and allophycocyanin-conjugated anti-CD34 antibodies (Cat no. 555824, BD Biosciences, San Jose, CA). $CD34^+$ and $CD3^+$ cells were sorted using a FACSAria II flow cytometer (BD biosciences). The purity of all sorted cells was >90%.

Sanger sequencing

All coding exons plus intron 5 of *GATA2* were amplified by polymerase chain reaction (PCR) with PrimeSTAR GXL DNA polymerase (Takara Bio, Otsu, Japan) using a Mastercycler pro

S thermal cycler (Eppendorf, Hamburg, Germany) according to manufacturer's instructions. Primer sequences are listed in **Supplementary Table 4**. Capillary sequencing was performed using an ABI/PRISM 3130xl Genetic Analyzer (Life Technologies, Carlsbad, USA). RefSeq NM_001145661.1 was used as a reference sequence for *GATA2*.

5

TA cloning

PCR of the target genomic DNA region was performed with PrimeSTAR GXL DNA polymerase. Products were run on an agarose gel, excised, purified with MinElute Gel Extraction Kit (Qiagen), A-tailed with Ex Taq DNA polymerase (Takara bio), and cloned into

10 the PCR2.1-TOPO vector using TOPO TA cloning kit (Life Technologies). Plasmid was cloned into DH5α competent cells and purified with QIAprep Spin Miniprep Kits (Qiagen).

Splice cassette analysis

The effect of splice site mutations was assessed using a splice cassette vector

15 (pSCEGFP2, kindly provided by Ohoshi Murayama, PhD, Laboratory of Molecular Biology, School of Life and Environmental Science, Azabu University), in which EGFP cDNA was divided into two parts (EGFP-N and EGFP-C) by an inserted intron (**Supplementary Figure 2**). We cloned exon 5 and its boundaries (258bp) with or without a mutation into the MCS. Expression of EGFP was measured by transfecting 293T cells with the constructed

20 vectors. Aberrant splicing was also confirmed by the cDNA sequencing of transfected cells.

Target gene sequencing

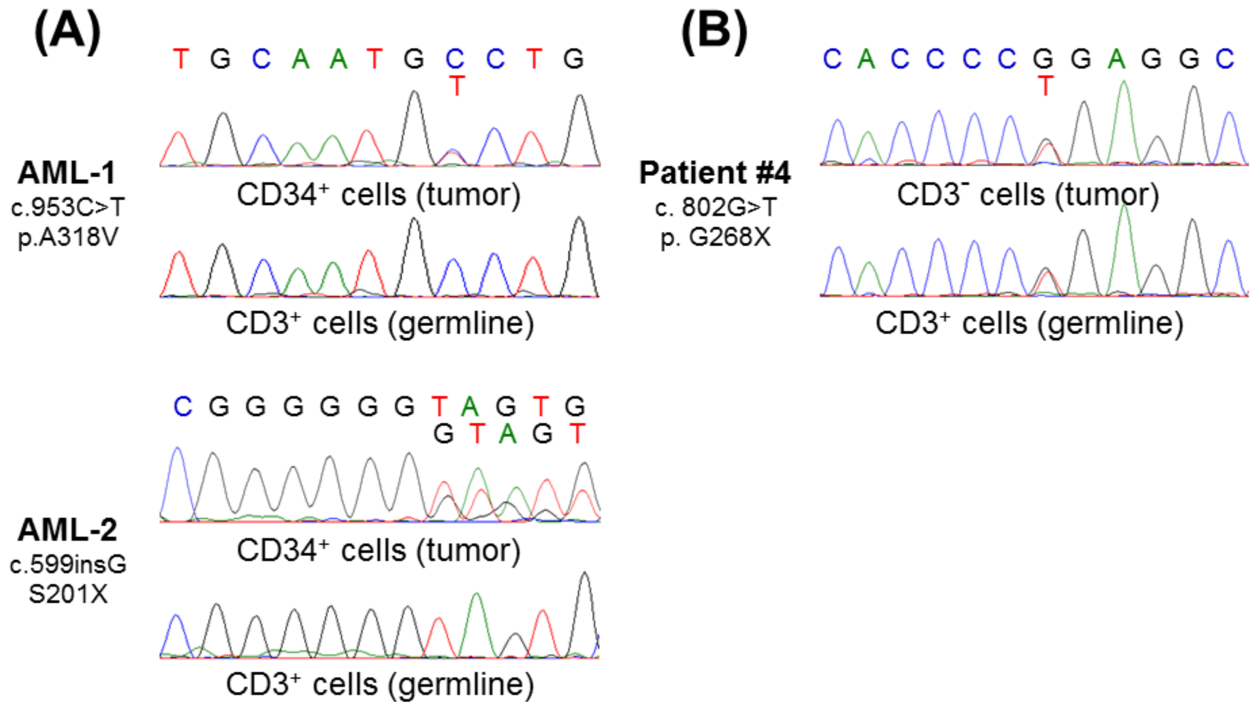
The target sequencing of selected genes was performed as previously described[1]. Briefly, target enrichment was performed using SureSelect (Agilent Technologies, Santa Clara, CA)

25 custom enrichment kits according to manufacturer's instructions. Massively-parallel sequencing was performed using HiSeq 2500 or MiSeq (Illumina, San Diego, CA) to

cover >97% of the target region 100 times. Sequence reads were aligned to a GRCh37 reference genome using Burrows-Wheeler Aligner 0.5.8[2]. Variants were detected using Samtools[3] and an in-house constructed pipeline. Variants with variant allele frequency (VAF) > 0.02 (2%) supported by ≥8 variant reads were considered positive. If the same variant was observed with VAF > 0.01 in 16 unrelated control samples from healthy volunteers, the variant was considered a false positive. Remaining variants were annotated with ANNOVAR[4]. Synonymous variants and common variants [defined as those with >1% frequency in ESP6500 (http://evs.gs.washington.edu/EVS/) database as of 12/2014] were excluded. To classify the somatic/germline origin of each variant, we used the following criteria: variants with VAF 0.4–0.6 (suggesting germline SNPs) were considered germline variants unless previously reported as somatic mutations in the COSMIC (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) database; or variants with VAF 0.02–0.39 (suggesting somatic mutations) or 0.61–0.99 (suggesting somatic mutations with loss of heterozygosity) were considered somatic variants unless they were registered in our in-house SNP database. This classification criterion was verified in patients for whom germline control samples were available (data not shown). For the determination of driver mutations, we performed an extensive literature and database search. We considered a mutation to be a driver if the mutation was reported to be a driver in the literature or the mutation was an inactivating (nonsense, splice site, or insertion/deletion) mutation and the gene was a known tumor-suppressor gene in hematological malignancies.

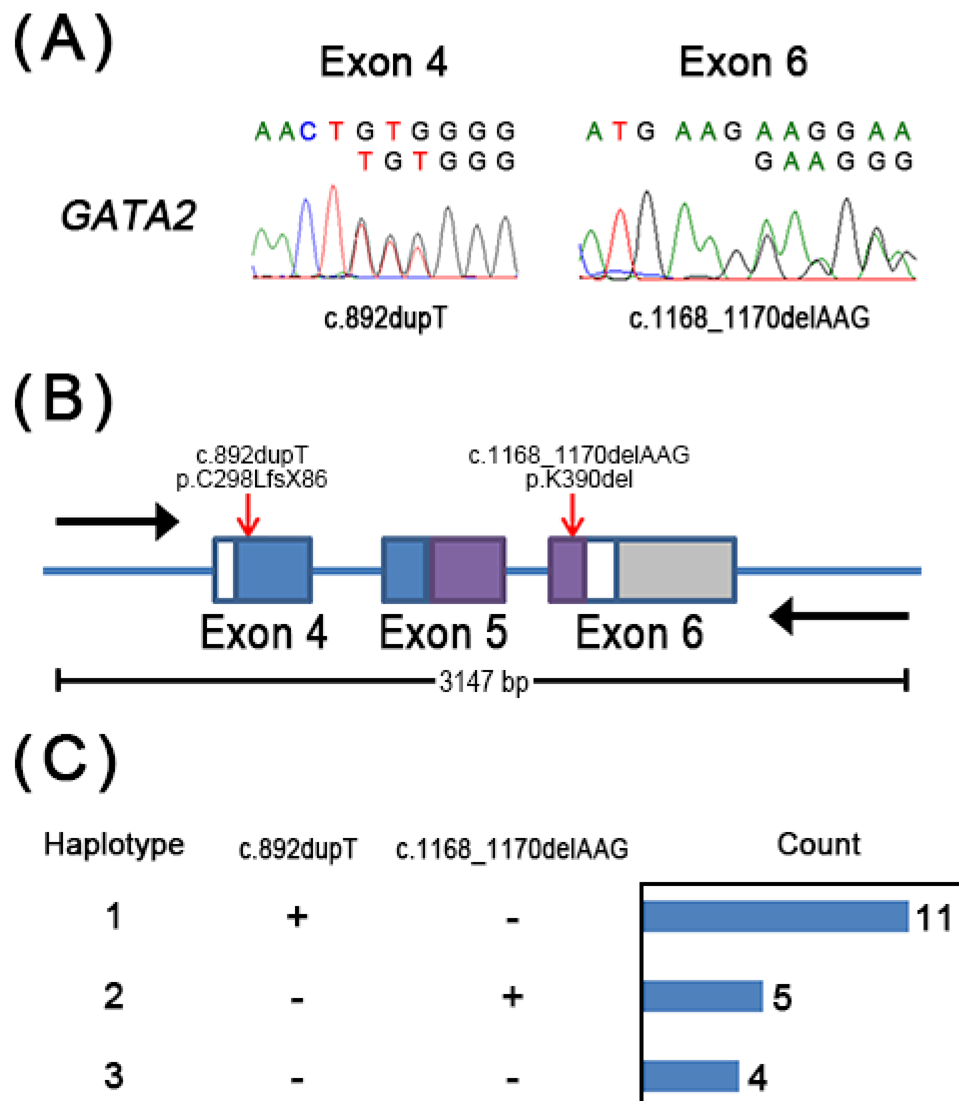
Supplementary Figures

Supplementary Figure 1. Somatic/germline origin of *GATA2* mutations



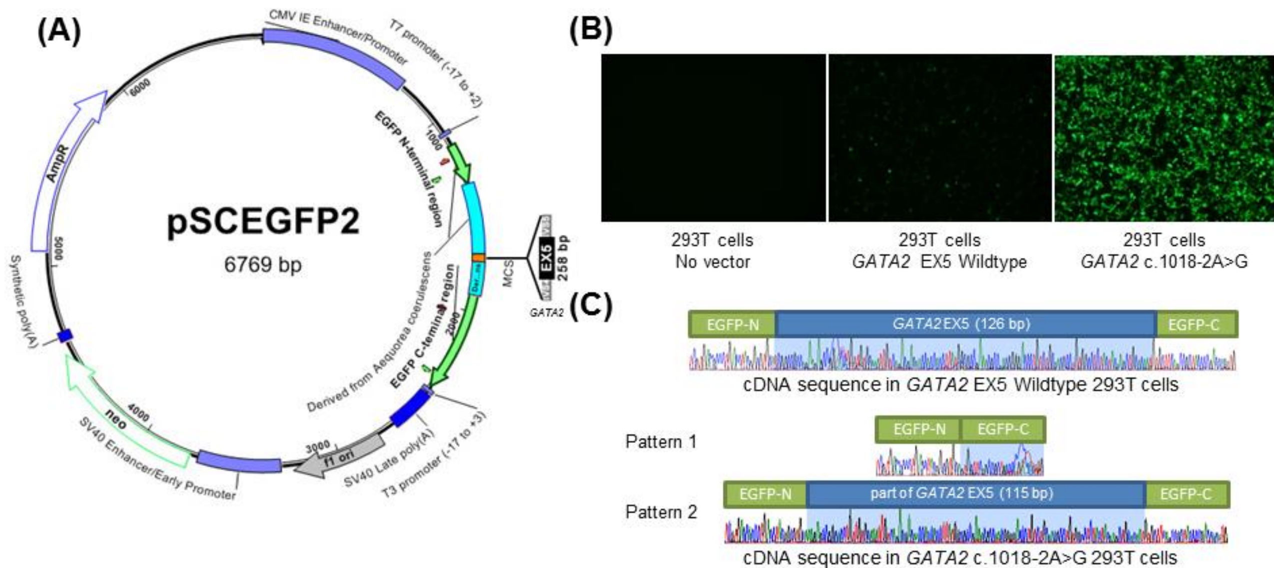
5 Peripheral blood mononuclear cells were sorted into $CD34^+$ or $CD3^-$ (containing tumor) and $CD3^+$ (germline) cells. DNA from each cell subset was extracted and was subjected to Sanger sequencing for each mutation.

Supplementary Figure 2. Haplotype analysis for *GATA2* mutations



(A) The Sanger sequencing of *GATA2* in Patient 2. Two *GATA2* mutations were identified: a c.892dupT mutation in exon 4 and a c.1168_1170delAAG mutation. (B) Schema of TA cloning. Arrows indicate primers. A 3147-bp region including exons 4–6 was amplified using PCR and cloned into a TA cloning vector. (C) Results of TA cloning. Haplotype 1 contains only the c.892dupT mutation, whereas haplotype 2 contains only the c.1168_1170delAAG mutation. Haplotype 3 (without mutation) was also observed.

Supplementary Figure 3. Splice cassette analysis



(A) A map of SCEGFP2 vector. EGFP cDNA was split into two parts (EGFP-N and EGFP-C) by an inserted intron sequence. A multi-cloning site was placed within the intron to allow the insertion of sequences of interest. A minigene containing exon 5 and part of IVS 4 and IVS 5 of *GATA2* (258 bp) was inserted into the vector. If the Exon 5 is skipped, functional EGFP is expressed; otherwise exon insertion inhibits the expression of functional EGFP. (B) The measurement of GFP expression. 293T cells were transfected with vehicle, *GATA2* Exon 5 wild type sequence-containing vector, or *GATA2* c.1018-2A>G mutation-containing vectors. (C) cDNA sequencing. In cells transfected with wild type *GATA2* minigene, full-length exon 5 of *GATA2* was transcribed. In cells transfected with c.1018-2A>G-containing minigene, two cDNA clones were isolated. In pattern 1, *GATA2* exon 5 was skipped (leading to the expression of functional EGFP). In pattern 2, the aberrantly spliced exon 5 transcript (115bp) is inserted between EGFP-N and EGFP-C.

Supplementary Tables

Supplementary Table 1. Summary of mutational analysis in *GATA2*

Disease	*GATA2* wildtype (patients)	*GATA2* mutated (patients)	Total (patients)
JMML	96	0	96
AA	75	0	75
AML	73	2	75
Familial MDS	0	6	6

5 MDS, myelodysplastic syndromes; JMML, juvenile myelomonocytic leukemia; AA, aplastic anemia; AML, acute myeloid leukemia

Supplementary Table 2. Target genes for next-generation sequencing

AEBP2	*FLT3*	*PRPF8*
ASXL1	*GATA1*	*PTPN11*
ATM	*GATA2*	*RAD21*
ATRX	*GNAS*	*RB1*
B2M	*GPRC5A*	*RIT1*
BCOR	*IDH1*	*RUNX1*
BCORL1	*IDH2*	*SETBP1*
BRAF	*IKZF1*	*SF3B1*
BRCA2	*IRF1*	*SH2B3*
BRCC3	*JAK1*	*SMC1A*
CBL	*JAK2*	*SMC3*
CDH23	*JAK3*	*SRP72*
CEBPA	*JARID2*	*SRSF2*
CREBBP	*KANSL1*	*STAG2*
CSMD1	*KDM6A*	*STAT3*
CTCF	*KIT*	*SUZ12*
CUX1	*KRAS*	*TERF1*
DAXX	*LAMB4*	*TERF2*
DCAF7	*LUC7L2*	*TET2*
DDX41	*MAP3K4*	*TP53*
DIDO1	*MPL*	*U2AF1*
DIS3	*NCOR2*	*U2AF2*
DNMT3A	*NF1*	*UMODL1*
EED	*NPM1*	*WT1*
ETNK1	*NRAS*	*ZRSR2*
ETV6	*PHF6*	*ZSWIM4*
EZH2	*PIGA*	
FBXW7	*PRF1*	

Supplementary Table 3. Detected mutations

Patient	Gene	Variant effect	Nucleic acid change	Amino acid change	Somatic/ germline	VAF	Driver
#1	*GATA2*	indel	892dupT	C298LfsX86	germline	0.53	yes
#1	*LAMB4*	missense	911C>T	P304L	germline	0.51	no
#1	*NCOR2*	missense	173T>A	I58N	germline	0.52	no
#1	*NCOR2*	missense	4396C>T	L1466F	germline	0.46	no
#2	*BRCA2*	missense	4427A>G	D1476G	germline	0.53	no
#2	*GATA2*	indel	892dupT	C298LfsX86	germline	0.51	yes
#2	*GATA2*	indel	1168_1170delAAG	K390del	somatic	0.39	yes
#2	*NCOR2*	missense	4396C>T	L1466F	germline	0.47	no
#2	*NCOR2*	missense	173T>A	I58N	germline	0.46	no
#2	*RUNX1*	indel	709_710delCA	Q237EfsX335	somatic	0.37	yes
#3	*AEBP2*	missense	719G>A	R240H	somatic	0.02	no
#3	*ASXL1*	indel	2676dupT	N893X	somatic	0.24	yes
#3	*ATM*	missense	4673C>T	T1558M	somatic	0.05	no
#3	*ATRX*	nonsense	634C>T	R212X	somatic	0.06	yes
#3	*B2M*	missense	95G>A	R32H	somatic	0.03	no
#3	*BRCA2*	nonsense	8773C>T	Q2925X	somatic	0.02	yes
#3	*BRCC3*	missense	766G>A	G256R	somatic	0.05	no
#3	*CSMD1*	missense	6334G>A	E2112K	somatic	0.04	no
#3	*CTCF*	missense	58G>A	E20K	somatic	0.03	no
#3	*CUX1*	missense	4414G>A	G1472S	somatic	0.02	no
#3	*GATA2*	indel	892dupT	C298LfsX86	germline	0.50	yes
#3	*GPRC5A*	indel	1016dupG	A340GfsX12	somatic	0.33	yes
#3	*IDH2*	missense	418C>T	R140W	somatic	0.03	yes
#3	*LAMB4*	missense	911C>T	P304L	germline	0.38	no
#3	*NRAS*	missense	187G>A	E63K	somatic	0.06	yes
#3	*SF3B1*	missense	587G>A	R196Q	somatic	0.03	no
#3	*SMC3*	missense	1010T>G	I337R	somatic	0.04	no
#3	*STAG2*	nonsense	3034C>T	R1012X	somatic	0.16	yes
#3	*SUZ12*	missense	1523G>A	R508H	somatic	0.05	no
#3	*TET2*	missense	217C>T	R73C	somatic	0.04	no
#3	*TET2*	missense	427G>A	D143N	somatic	0.04	no
#3	*WT1*	missense	1057C>T	R353C	somatic	0.02	no
#3	*ZSWIM4*	missense	1141G>A	D381N	somatic	0.05	no
#3	*ZSWIM4*	missense	733G>A	G245S	somatic	0.02	no
#4	*ASXL1*	missense	2888C>T	P963L	germline	0.45	no

#4	*DAXX*	missense	22G>A	G8S	germline	0.49	no
#4	*FLT3*	missense	1130C>A	P377Q	somatic	0.46	no
#4	*GATA2*	nonsense	802G>T	G268X	germline	0.55	yes
#4	*NF1*	missense	1048G>A	V350M	somatic	0.44	no
#4	*NRAS*	missense	35G>A	G12D	somatic	0.44	yes
#4	*TP53*	missense	46C>A	Q16K	somatic	0.34	yes
#4	*WT1*	indel	1090_1093dupTCGG	p.A365VfsX20	somatic	0.98	yes
#5	*ASXL1*	indel	1935dupG	G646WfsX12	somatic	0.14	yes
#5	*GATA2*	splice	1018-2A>G	(exon 5)	germline	0.50	yes
#5	*LUC7L2*	missense	863G>C	R288P	germline	0.58	no
#5	*MAP3K4*	missense	2996A>G	N999S	germline	0.48	no
#5	*SETBP1*	missense	2602G>A	D868N	somatic	0.17	yes
#5	*STAT3*	indel	1846_1848delGAA	E616del	somatic	0.05	no
#5	*ZRSR2*	indel	1279_1290dupAGGGACCGCAGC	R317_S320dup	germline	1.00	no
#6	*GATA2*	splice	1018-2A>G	(exon 5)	germline	0.55	yes
#6	*KDM6A*	missense	C3212G	S1071C	germline	0.50	no
#6	*ZRSR2*	indel	1279_1290dupAGGGACCGCAGC	R317_S320dup	germline	0.47	no
AML-1	*CEBPA*	indel	247delC	Q83SfsX77	somatic	0.45	yes
AML-1	*CEBPA*	indel	910_915dupGCCAAG	A303_K304dup	somatic	0.20	yes
AML-1	*GATA2*	missense	953C>T	A318V	somatic	0.47	yes
AML-1	*NCOR2*	missense	5861C>G	S1954C	germline	0.48	no
AML-1	*NRAS*	missense	38G>A	G13D	somatic	0.25	yes
AML-1	*ZSWIM4*	missense	801G>T	Q267H	germline	0.55	no
AML-2	*ATM*	missense	8246A>T	K2749I	germline	0.53	no
AML-2	*GATA2*	indel	599insG	S201X	somatic	0.42	yes
AML-2	*KIT*	missense	2446G>T	D816Y	somatic	0.29	yes
AML-2	*SMC1A*	indel	392delinsTTCT	R131delinsLL	somatic	0.79	yes

VAF, variant allele frequency

Supplementary Table 4. PCR primers for *GATA2*

Target	Forward primer	Reverse primer	Anneal temp.	Product size (bp)
***GATA2* Exon 1**	5'-CCCGCAAAGTGATGTCGAA-3'	5'-ACCAAATACTCCCTCTGGTTAAAG-3'	60	531
***GATA2* Exon 2**	5'-ACCTCGTGGTGGGACTTTG-3'	5'-GATCCTACATCCGGGAAGC-3'	62	400
***GATA2* Exon 3a**	5'-GTCCCTAGCTCTGCCTACCC-3'	5'-CTCCTCGGGCTGCACTAC-3'	64	412
***GATA2* Exon 3b**	5'-ACCTTTTCGGCTTCCCAC-3'	5'-CTCTCCCAAGTCACAGCTCC-3'	64	406
***GATA2* Exon 4**	5'-GACTCCCTCCCGAGAACTTG-3'	5'-TGTAATTAACCGCCAGCTCC-3'	64	229
***GATA2* Exon 5**	5'-GTGGAGCGAGGGTCAGG-3'	5'-CACAAAGCGCAGAGGTCC-3'	64	223
***GATA2* Exon 6**	5'-AGGAATGTTGCTGGAGGAAG-3'	5'-GCTGGCAGGAGTGGTGTC-3'	64	415
***GATA2* Intron 5**	5'-ATGGAGTCACCTATACTGTGTATTT-3'	5'-TTTGCAGAGTGGAGGGTATTAG-3'	60	217

GATA2 Exon 4 forward and *GATA2* Exon 6 reverse primers were used for TA cloning.

References

1. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241-247.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-1760.
3. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-2079.
4. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010;38(16):e164.