Supporting Information

Corces-Zimmerman et al. 10.1073/pnas.1324297111

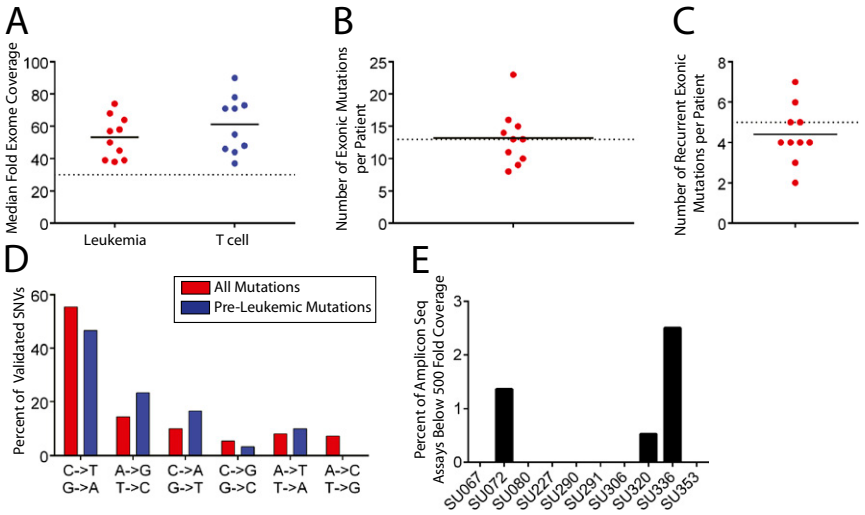

Fig. S1. Sequencing statistics from exome sequencing and targeted amplicon sequencing studies. (*A*) The median fold coverage of each individual exome sequencing analysis is plotted for both the leukemia exome and the T-cell (germ-line control) exome. Each patient was sequenced to >30-fold median coverage (dotted line). The total number of exonic mutations per patient (*B*) and the total number of recurrent exonic mutations per patient (*C*) are shown in comparison with the averages obtained in previous sequencing studies from The Cancer Genome Atlas (TCGA) Research Network (11) (dotted lines). (*D*) The rate of various transitions and transversions is shown for all leukemia-specific mutations and for the subset of mutations classified as preleukemic mutations. Both of the mutation spectrums are dominated by C-to-T transitions. (*E*) Targeted amplicon sequencing was performed on all leukemia-specific mutations from each patient in multiple cellular subpopulations. The total number of assays that were not covered to >500-fold coverage is shown for each patient. One hundred percent of all assays of recurrently mutated genes and more than 99% of all assays were covered to >500-fold.

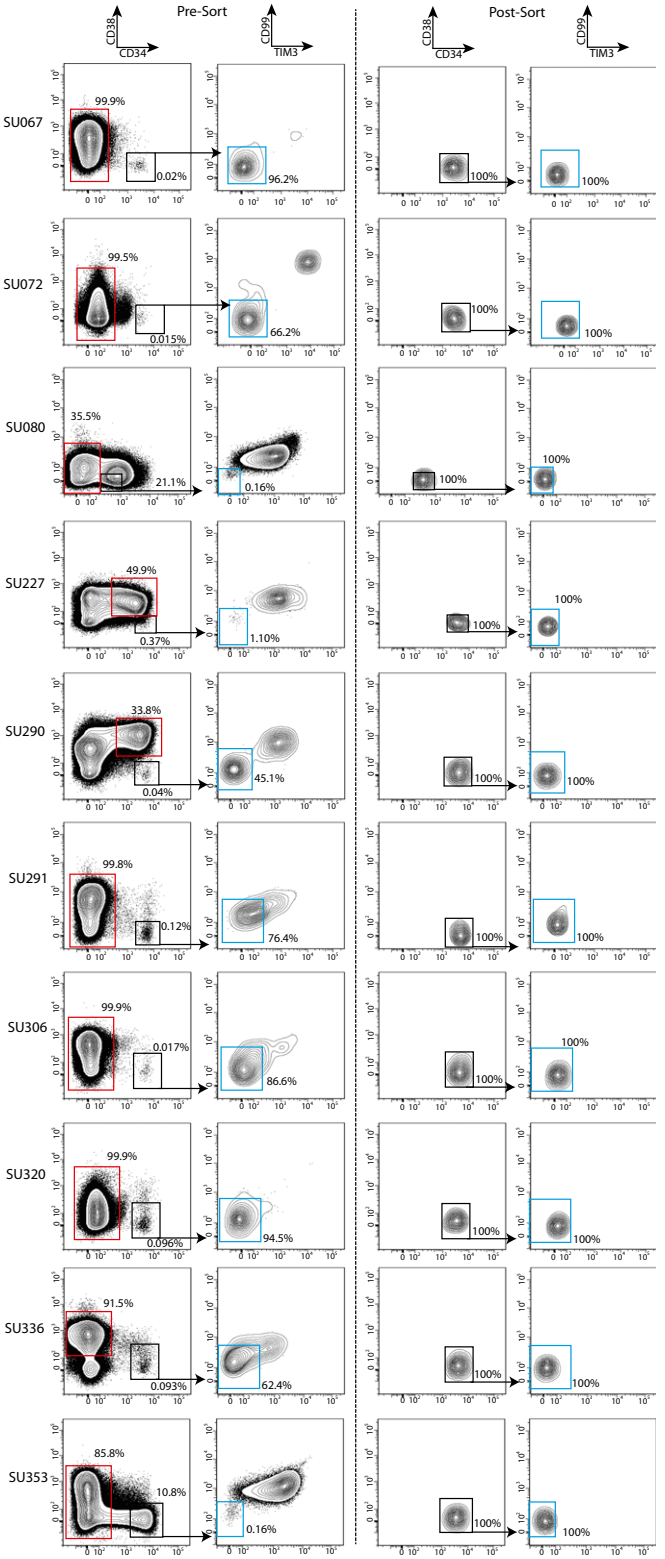

Fig. S2. Hematopoietic stem cell (HSC) sorting schemes for all 10 patients profiled by exome sequencing. CD34+CD38− cells from each patient were profiled for expression of leukemia-specific markers TIM3 and CD99. The CD34+/CD38−/TIM3−/CD99− population (blue box) was sorted to high purity. Leukemia cells (red box) were sorted as TIM3+/CD99+. Presort and postsort analyses of sorted HSCs are shown to indicate sort purity and rule out leukemia cell contamination.

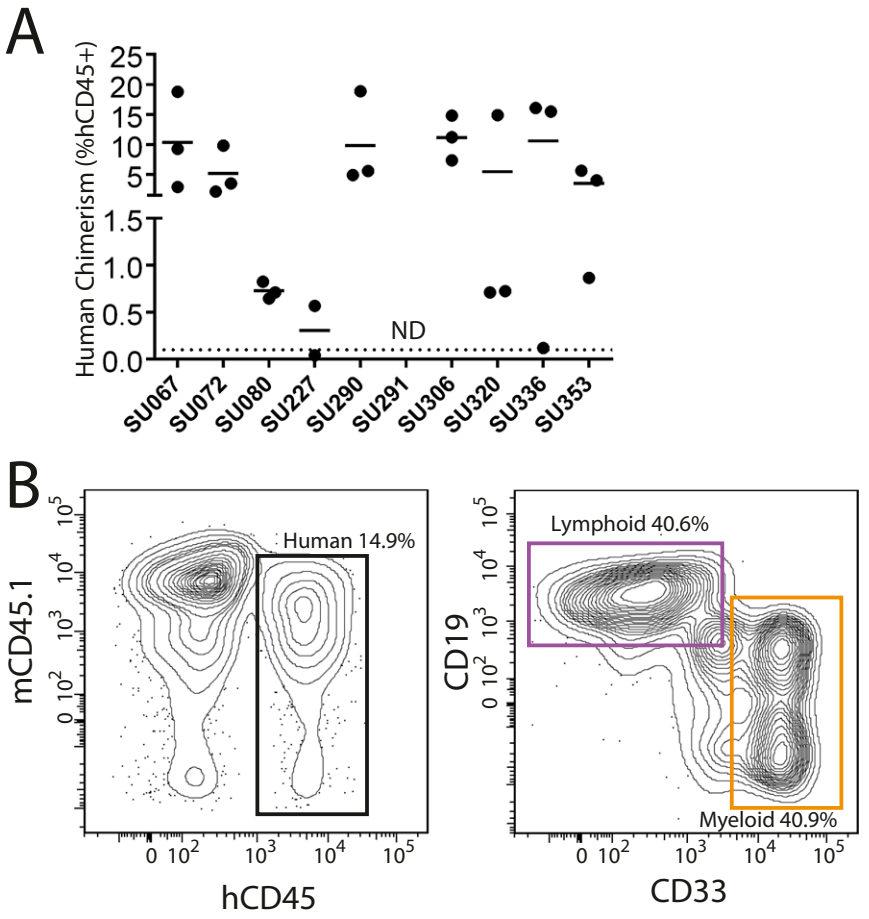

Fig. S3. Xenotransplantation studies of HSCs isolated from leukemia patients. HSCs from leukemia patients were transplanted in neonatal NOD/SCID/IL2R-γ (NSG) mice via facial vein injection. (*A*) Levels of human chimerism, indicated by hCD45+ cells, are shown for each of three mice transplanted with HSCs from each individual case. HSCs from case SU291 were not transplanted due to an insufficient number of cells. The average level of human chimerism is indicated by a horizontal bar, and the lower limit of detectable engraftment (0.1% hCD45+ cells) is indicated by the dotted line. (*B*) An example of the long-term bilineage engraftment observed in all cases is shown. Engrafted cells contain both CD33+ myeloid and CD19+ lymphoid cells indicating engraftment of bona fide stem cells.

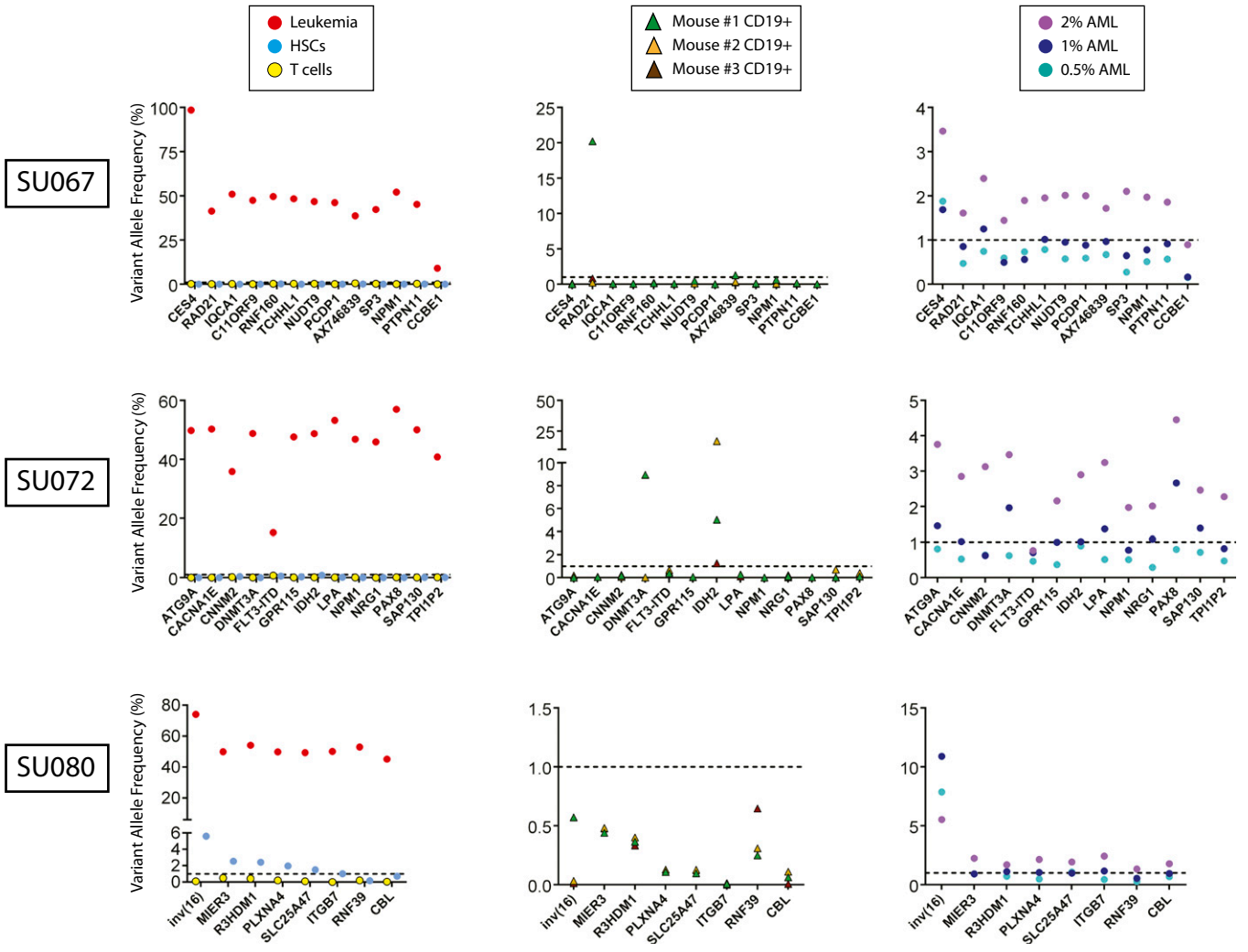

Fig. S4. (Continued)

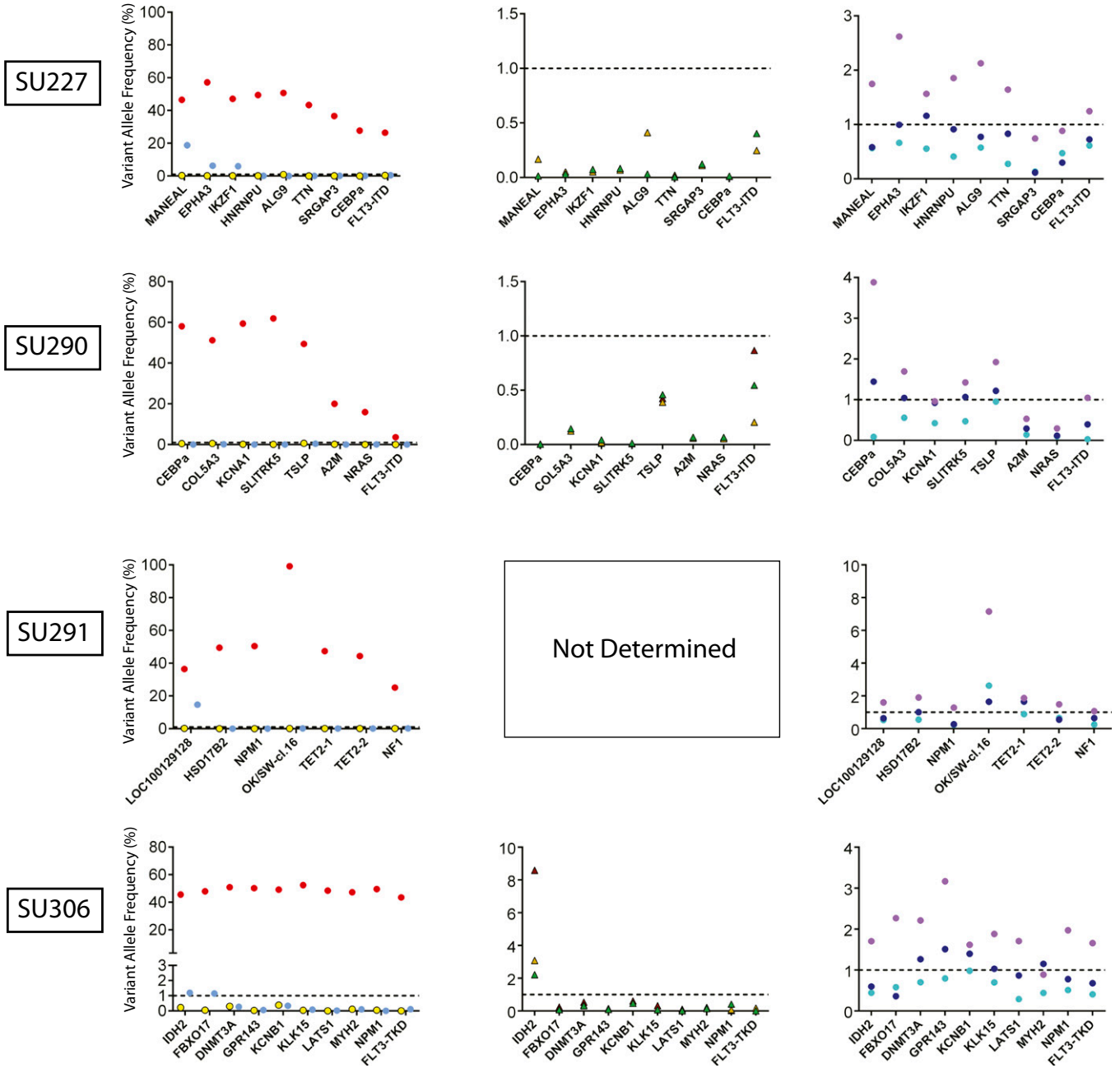

Fig. S4. (Continued)

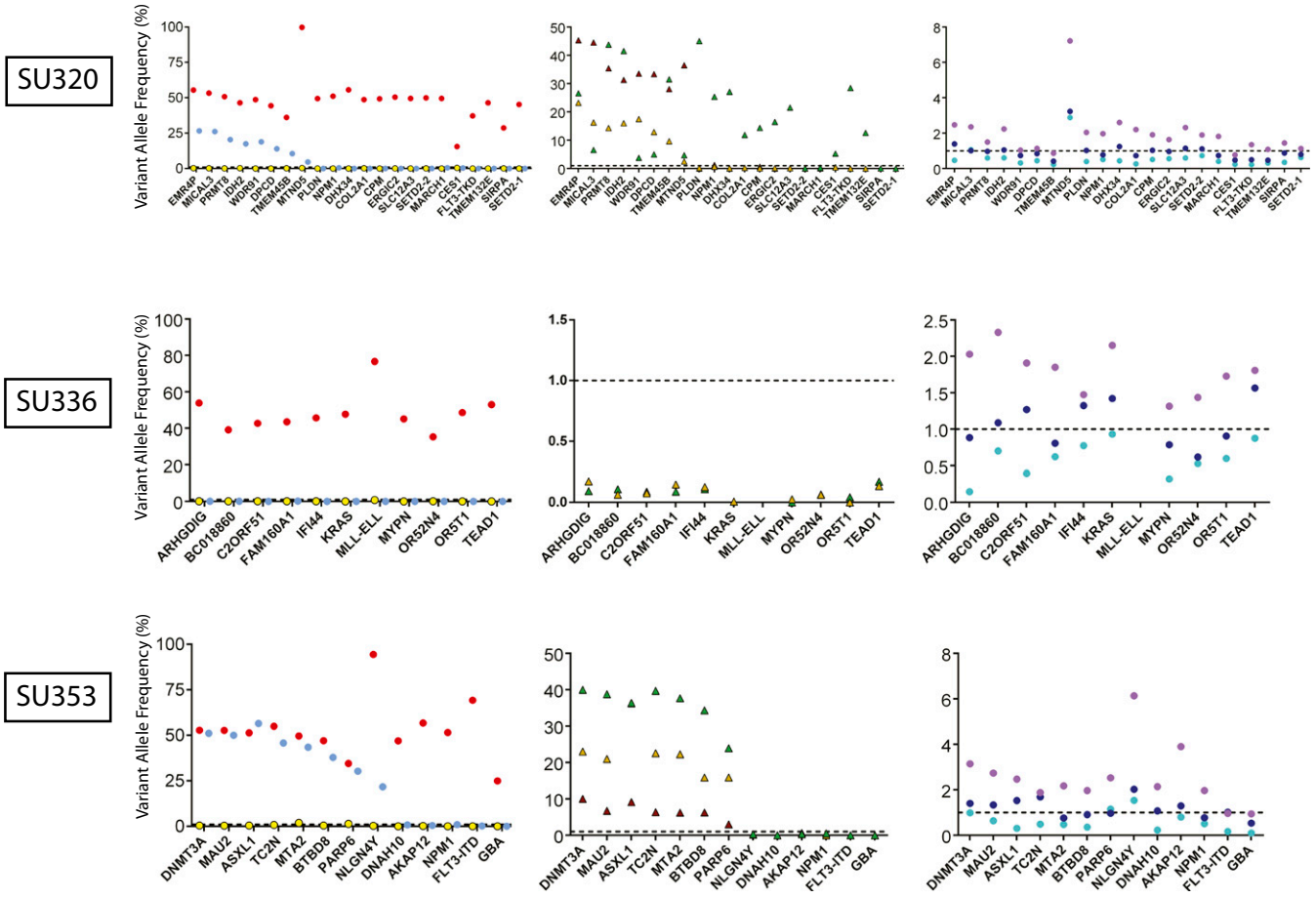

Fig. S4. Targeted amplicon sequencing of all leukemia-specific mutations in FACS-purified cells. FACS-purified leukemia cells, T cells, and HSCs were profiled by targeted amplicon sequencing for all leukemia-specific mutations (*Left*). Similarly, FACS-purified hCD45+CD19+ lymphoid and hCD45+CD33+ myeloid cells from mice engrafted with patient HSCs were profiled by targeted amplicon sequencing for all leukemia-specific mutations (*Center*). These assays were used to determine which mutations could be classified as preleukemic. Each assay was also subjected to quality control measurements, which involved the sequencing of admixtures of leukemia genomic DNA (gDNA) and nonleukemic control gDNA (*Right*). These assays were used to determine the accuracy of our sequencing methodologies in determining the variant allele frequency of low-abundance mutations.

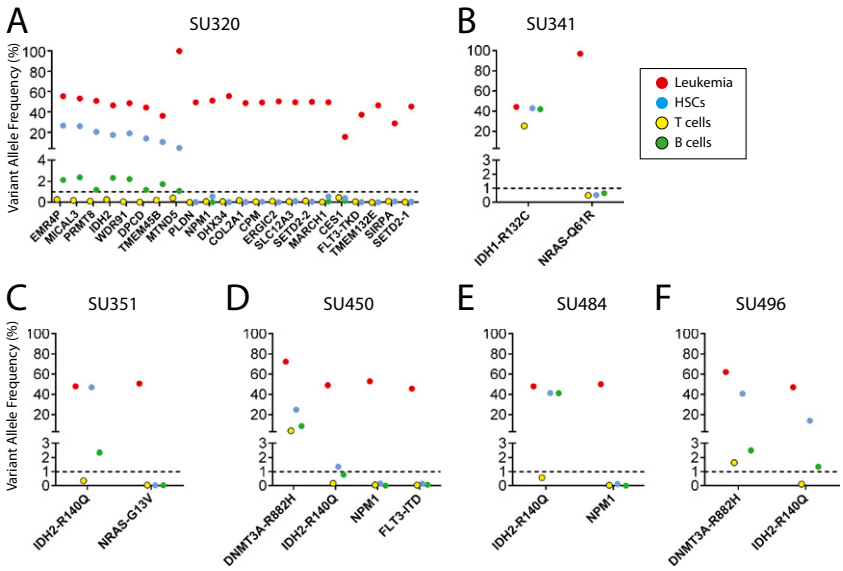

Fig. S5. Preleukemic mutations are detectable in B and T lymphocytes from acute myeloid leukemia (AML) patients. FACS-purified CD19+/CD20+ B cells and CD3+ T cells were profiled by targeted amplicon sequencing for all leukemia-specific mutations. (*A*–*F*) Cases in which preleukemic mutations could be detected in these lymphocytes are shown. In almost all cases, a late mutation was also profiled to demonstrate that detectable preleukemic mutations were not the consequence of leukemia cell contamination.

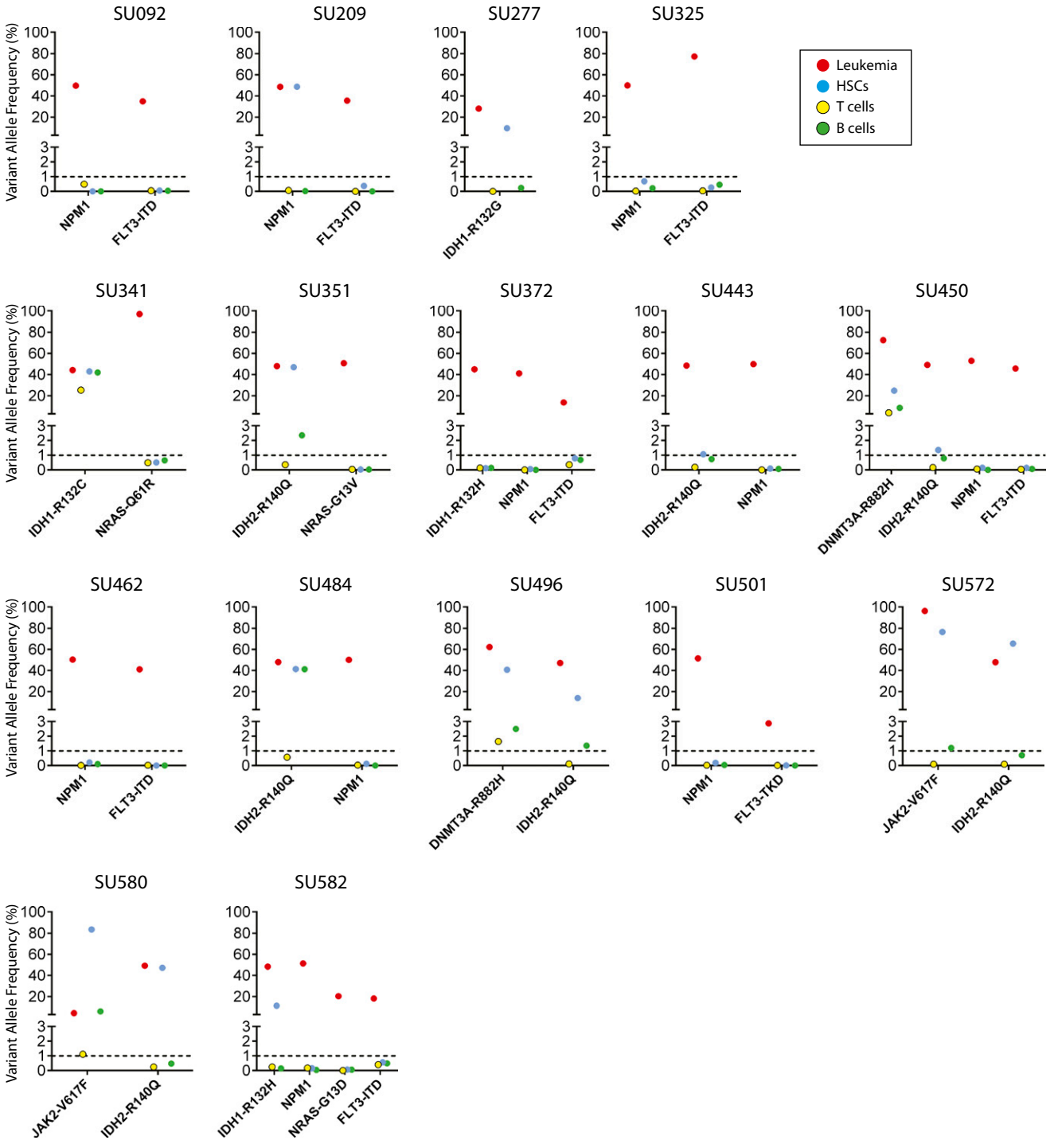

Fig. S6. Targeted amplicon sequencing of patients with known mutations in *IDH1*/*IDH2*, *DNMT3A*, *NPM1*, *FLT3*, or *KRAS*/*NRAS*. Targeted amplicon sequencing of FACS-purified leukemia cells, HSCs, T cells, and B cells from patients with known recurrent mutations in *IDH1*/*IDH2*, *DNMT3A*, *NPM1*, *FLT3*, or *KRAS*/*NRAS* was performed in 16 additional patients. Variant allele frequencies for each mutation profiled in each patient are shown.

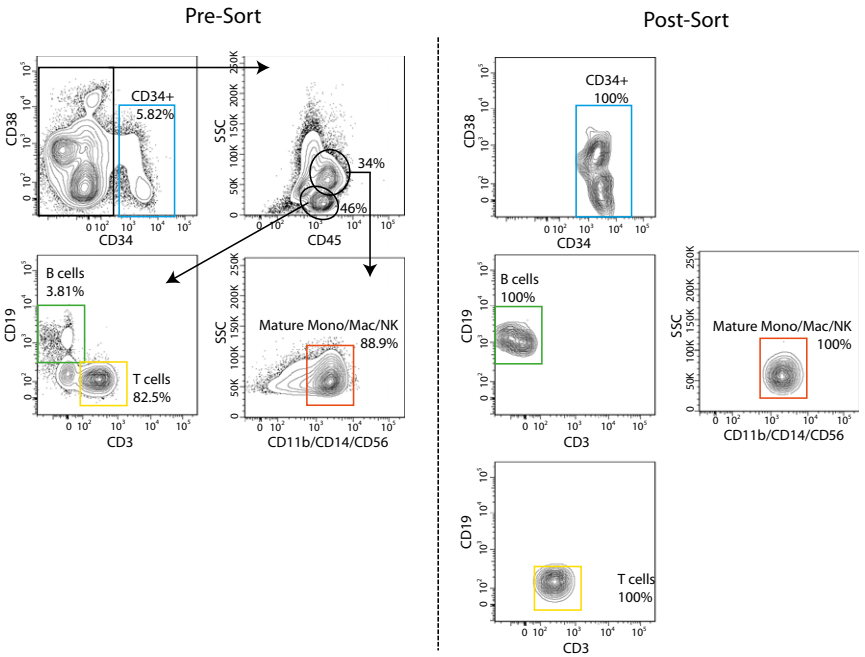

Fig. S7. Sorting scheme for remission bone marrow samples. Bone marrow mononuclear cells from patient SU353 were FACS purified for CD34+ cells, CD19+ B cells, CD3+ T cells, and CD11b+/CD14+/CD56+ monocytes, macrophages, and natural killer cells. Sample postsort analyses are presented for patient SU353 to show purity of sorted cells.

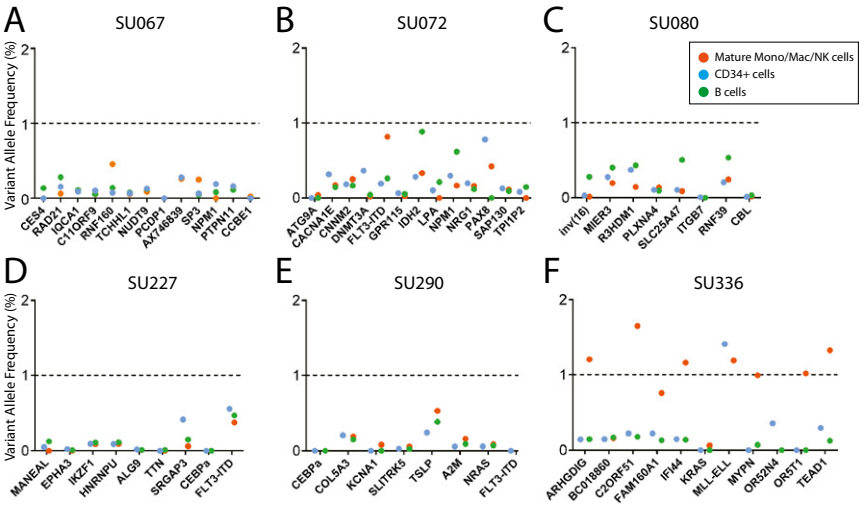

Fig. S8. Targeted amplicon sequencing of leukemia-specific mutations in remission cells. Bone marrow mononuclear cells from each patient were FACS purified. (*A–F*) Targeted amplicon sequencing of all leukemia-specific mutations was performed on gDNA from remission CD34+ cells, B cells, and mature monocytes, macrophages, and natural killer cells. Results from six patients are indicated.

Table S1. Recurrent and preleukemic mutations identified by exome and targeted amplicon sequencing of 10 AML patients

Patient	Age	Sex	Recurrent mutations	Preleukemic recurrent mutations	Evidence for preleukemia
SU067	52	F	NPM1, RAD21, PTPN11, CES4	RAD21	CD19+ lymphoid cells from transplant
SU072	57	M	NPM1, FLT3-ITD, IDH2, DNMT3A, CACNA1E, SAP130, LPA	IDH2, DNMT3A	CD19+ lymphoid cells from transplant
SU080	37	M	inv(16), CBL, MIER3	inv(16), MIER3	Sorted HSCs
SU227	29	M	FLT3-ITD, CEBPa, IKZF1, EPHA3	IKZF1, EPHA3	Sorted HSCs
SU290	64	M	FLT3-ITD, CEBPa, NRAS	—	No preleukemia identified
SU291	35	F	NPM1, TET2, NF1, trisomy 8	—	Sorted HSCs
SU306	32	F	NPM1, FLT3-D835, IDH2, DNMT3A	IDH2	Sorted HSCs and CD19+ lymphoid cells from transplant
SU320	68	M	NPM1, FLT3-ITD, IDH2, COL2A1, DHX34	IDH2, COL2A1, DHX34	Sorted HSCs and CD19+ lymphoid cells from transplant
SU336	29	F	MLL-ELL, KRAS-G13D	—	No preleukemia identified
SU353	65	M	NPM1, FLT3-ITD, DNMT3A, ASXL1, AKAP12, PARP6	DNMT3A, ASKL1, PARP6	Sorted HSCs and CD19+ lymphoid cells from transplant
SU008	64	M	FLT3-ITD	—	Sorted HSCs
SU014	59	M	IDH1, SMC1A, ZBTB33, C10ORF76, NPM1, FLT3-ITD	NPM1, SMC1A, ZBTB33	Sorted HSCs
SU030	52	F	SLC12A1, NPM1, FLT3-ITD	—	Sorted HSCs
SU043	38	F	NPM1, FLT3-ITD	—	No preleukemia identified
SU048	79	F	TET2, SMC1A, NPM1, FLT3-ITD	TET2, SMC1A	Sorted HSCs and CD19+ lymphoid cells from transplant
SU070	72	F	TET2, PHF6, CTCF, KALRN, DOCK9, PLA2G4D, FLT3-ITD	TET2, PHF6, CTCF, KALRN, DOCK9, PLA2G4D	Sorted HSCs and CD19+ lymphoid cells from transplant

Patient demographic and recurrent mutations identified in each case are indicated. In addition, recurrent mutations that were found to be preleukemic are indicated alongside the sequencing information used to inform that classification.

Table S2. Patient clinical information

Sample	Age	Sex	Primary/ secondary	De novo/ relapsed	Cytogenetics	FLT3-ITD	WHO classification	FAB	Induction treatment	Consolidation treatment	Relapse?	Transplant?	Current disease status
SU067	52	F	Primary	De novo	Normal	Negative	AML–not otherwise specified	M5	3+4	ME × 2, HiDAC × 2	Yes, 1 y 5 mo	Yes, myeloablative MUD in CR2	Alive, CR2, 2 y 10 mo
SU072	58	M	Primary	De novo	Normal	Positive	AML–not otherwise specified	M1	3+7	HiDAC × 4	N/A	No	Alive, CR1, 2 y 8 mo
SU080	37	M	Primary	De novo	inv(16)	Negative	AML with inv(16)(p13.1q22) or *t*(16;16)(p13.1;q22)	M5	3+7	HiDAC × 3	Unknown	Unknown	Unknown, lost to f/u 4 mo
SU227	29	M	Primary	De novo	Normal	Positive	AML with myelodysplasia-related changes	N/A	G-CLAC	G-CLAC × 3	No	Yes, Haplo in CR1	Deceased, 7 mo, TRM from haplo
SU290	64	M	Primary	De novo	Normal	Weak	AML–not otherwise specified	M5	G-CLAC	G-CLAC × 1, Int Ara-C × 1	No	Yes, NMA MUD in CR1	Alive, 1 y 10 mo
SU291	35	F	Primary	De novo	+8	Negative	AML–not otherwise specified	M5	3+7	HiDAC × 4	Yes, 1 y 6 mo	Yes, myeloablative MRD with residual disease	Alive, 1 y 10 mo
SU306	32	F	Primary	De novo	No analyzable metaphases	Negative	AML–not otherwise specified	M5a	G-CLAC	G-CLAC × 3	Unknown	Unknown	Unknown, lost to f/u 7 mo
SU320	68	M	Primary	De novo	No analyzable metaphases	Negative	AML–not otherwise specified	N/A	Decitabine, 3+7	Ida + Int Ara-C × 2	Yes, 5 mo	Yes, NMA MUD in CR1	Deceased, 7 mo
SU336	29	F	Primary	De novo	*t*(11;19)(q23;p13.1)	Negative	AML with 11q23 (MLL)	M5	3+4	HiDAC × 1	No	Yes	Alive, CR1, 1 y 5 mo
SU353	65	M	Primary	De novo	Normal	Positive	AML–not otherwise specified	M4	3+7	None	Yes, 2 mo	No	Deceased, 2 mo
SU092	53	F	Primary	De novo	Normal	Positive	AML–not otherwise specified	N/A				Yes, NMA MRD in CR1	Alive, CR1, 2 y 6 mo
SU209	65	F	Primary	De novo	Normal	Positive	AML–not otherwise specified	M2		Did not achieve CR	Primary refractory	No	Deceased, 1 y
SU277	69	F	Primary	De novo	Normal	Negative	AML with myelodysplasia-related changes	M5			Yes, 11 mo	No	Deceased, 1 y 4 mo
SU325	59	F	Primary	De novo	Unknown	Positive	AML–not otherwise specified	M0/M1					Unknown, lost to f/u 6 mo
SU341	28	M	Primary	De novo	Normal	Negative	AML–not otherwise specified	M1			No	Yes, myeloablative MUD in CR1	Alive, CR1, 1 y 5 mo
SU351	74	M	Primary	Relapsed	der(10)*t*(X;10), +8	Negative	AML with myelodysplasia-related changes	M5					Unknown, lost to f/u 1 mo
SU372	53	F	Primary	De novo	Normal	Positive	AML–not otherwise specified	M5a			Yes, 7 mo	No	Deceased, 10 mo
SU443	68	M	Primary	De novo	Normal	Positive	AML–not otherwise specified	M1					Unknown, lost to f/u 4 mo
SU450	41	M	Primary	Primary refractory	Unknown	Positive	AML–not otherwise specified	M1				Yes, myeloablative 9/10 mismatch	Hospice

Table S2. Cont.

Sample	Age	Sex	Primary/ secondary	De novo/ relapsed	Cytogenetics	FLT3-ITD	WHO classification	FAB	Induction treatment	Consolidation treatment	Relapse?	Transplant?	Current disease status
SU462	85	F	Primary	De novo	Unknown	Positive	AML–not otherwise specified	N/A				No	Deceased, 1 mo
SU484	70	F	Primary	De novo	Normal	Negative	AML–not otherwise specified	M5		None	Yes, 7 mo	No	Alive, 7 mo
SU496	23	M	Primary	Relapsed	+6, +8, +11	Negative	AML–not otherwise specified	M5					Deceased
SU501	80	M	Primary	De novo	Normal	Negative	AML–not otherwise specified	M5				No	Hospice
SU572	72	M	Secondary	From PV	Unknown	Unknown	Unknown	N/A					Unknown
SU580	67	M	Secondary	From PV/MF	Complex	Negative	AML–not otherwise specified	N/A					Alive
SU582	71	M	Primary	De novo	Normal	Positive	AML–not otherwise specified	N/A					Deceased
SU008	63	M	Primary	De novo	Normal	Positive	AML–not otherwise specified	M1					Alive, 7 y 6 mo
SU014	58	M	Primary	De novo	Normal	Positive	AML–not otherwise specified	N/A					Unknown, lost to f/u 1 mo
SU030	52	F	Primary	De novo	Normal	Positive	AML with myelodysplasia-related changed	M4			Yes		Unknown, lost to f/u 9 mo
SU043	38	F	Primary	De novo	Normal	Positive	AML–not otherwise specified	M4			No	Yes, myeloablative MUD in CR1	Deceased, 2 y 11 mo
SU048	75	F	Primary	De novo	Unknown	Positive	AML–not otherwise specified	N/A			No	No	Deceased, 1 mo
SU070	72	F	Primary	De novo	Normal	Positive	AML with myelodysplasia-related changed	M4			No	No	Deceased, 1 mo

Patient clinical information is provided for each case investigated. Ten patients were profiled by exome sequencing in this study. Six patients were profiled by exome sequencing in our previous study (1). These 16 patients were analyzed and presented in Fig. 2*A*. Sixteen additional patients were profiled using targeted amplicon sequencing for the presence of recurrent mutations in *IDH1*/*IDH2*, *DNMT3A*, *NPM1*, *FLT3*, and *KRAS*/*NRAS* in HSCs. These 16 patients were analyzed and presented in Fig. 2*B*. Additionally, any of the 16 patients profiled by exome sequencing were also presented in Fig. 2*B* if the patient had a leukemic mutation in one of the profiled genes. CR, complete remission; FAB, French American British; f/u, follow-up; G-CLAC, clofarabine with high dose cytarabine and granulocyte colony stimulating factor priming; Haplo, haploidentical transplant; HiDAC, high dose cytarabine; Ida, idarubicin; ME, mitocantrone and etoposide; MF, myelofibrosis; MLL, mixed lineage leukemia; MRD, matched related donor; MUD, matched unrelated donor; NMA, non-myelo ablative; PV, polycythemia vera; TRM, treatment related mortality; WHO, World Health Organization.

1. Jan M, et al. (2012) Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med* 4(149):149ra118.

Table S3. Patient leukemia-specific mutations

AML case	Chromosome	Position	Gene name	Accession no.	Mutation type	Ref allele	Var allele	Annotation	T-cell variant ratio	T-cell VAF	Leukemia variant ratio	Leuk. VAF
SU067	chr3	195,700,761	AX746839	AX746839	SNP	A	G	UTR	655/119,886	0.55	44,039/113,681	38.74
SU067	chr11	61,311,105	C11orf9	NM_013279	SNP	G	A	UTR	44/30,276	0.15	40,864/86,126	47.45
SU067	chr18	55,273,163	**CCBE1**	NM_133459	SNP	C	A	G185V	5/9,865	0.05	1,008/11,103	9.08
SU067	chr16	54,356,252	**CES4**	NR_003276	SNP	G	A	UTR	4/2,342	0.17	2,611/2,649	98.57
SU067	chr2	237,039,038	IQCA1	NM_024726	SNP	C	T	V259M	108/46,026	0.23	32,092/62,942	50.99
SU067	chr5	170,770,149	**NPM1**	NM_199185	INDEL	T	+TCTG	ins(+TCTG)	67/48,592	0.14	26,698/51,209	52.14
SU067	chr4	88,598,209	NUDT9	NM_024047	SNP	C	T	UTR	212/76,608	0.28	41,807/89,316	46.81
SU067	chr2	120,075,379	PCDP1	NM_001029996	SNP	C	G	UTR	25/19,844	0.13	3,416/7,391	46.22
SU067	chr12	111,372,581	**PTPN11**	NM_002834	SNP	G	A	A72T	42/13,422	0.31	7,312/16,160	45.25
SU067	chr8	117,929,042	**RAD21**	NM_006265	SNP	G	A	Q592*	52/48,538	0.11	22,528/54,420	41.40
SU067	chr21	29,279,099	RNF160	NM_015565	SNP	C	T	V167I	3/1,118	0.27	707/1,425	49.61
SU067	chr2	174,528,942	SP3	NM_003111	SNP	G	A	Q182*	96/30,452	0.32	19,199/45,400	42.29
SU067	chr1	150,325,022	**TCHHL1**	NM_001008536	SNP	C	T	G587D	61/18,691	0.33	14,116/29,194	48.35
SU072	chr2	219,795,669	ATG9A	NM_024085	SNP	C	A	G597V	0/10,535	0.00	5,311/10,670	49.78
SU072	chr1	180,034,849	**CACNA1E**	NM_000721	SNP	G	A	UTR	2/9,111	0.02	3,876/7,712	50.26
SU072	chr10	104,826,926	CNNM2	NM_017649	SNP	A	G	*876W	2/1,220	0.16	7,769/21,651	35.88
SU072	chr2	25,320,998	**DNMT3A**	NM_175629	INDEL	GTACT	GTGT	Q338fs(+G, −AC)	75/103,254	0.07	55,444/113,732	48.75
SU072	chr13	27,506,223	**FLT3**	NM_004119	INDEL	—	ins(+81bp)	ITD	7/948	0.74	93/612	15.20
SU072	chr6	47,789,910	**GPR115**	NM_153838	SNP	G	A	V324I	21/40,479	0.05	16,703/35,083	47.61
SU072	chr15	88,432,938	**IDH2**	NM_002168	SNP	C	T	R140Q	12/9,576	0.13	4,783/9,818	48.72
SU072	chr6	160,947,547	**LPA**	NM_005577	SNP	G	A	P913S	72/81,050	0.09	31,344/58,920	53.20
SU072	chr5	170,770,152	**NPM1**	NM_199185	INDEL	G	+GTCA	ins(+GTCA)	5/7,686	0.07	1,755/3,746	46.85
SU072	chr8	32,624,969	NRG1	NM_013959	SNP	C	T	A64V	5/15,171	0.03	7,783/16,953	45.91
SU072	chr2	128,460,977	**SAP130**	NM_001145928	SNP	C	T	V630M	104/107,365	0.10	59,525/118,961	50.04
SU072	chr7	128,483,345	TPI1P2	NR_002187	SNP	T	A	UTR	60/49,814	0.12	22,920/56,120	40.84
SU080	chr11	118,654,216	**CBL**	NM_005188	Indel	AGGTAC	A	Q409fs (-GGTAC)	10/24,035	0.04	8,088/17,901	45.18
SU080	chr16	65,683,825==15,722,691	**inv(16)**	—	Translocation	—	—	Translocation	30/33,254	0.09	20,680/27,881	74.17
SU080	chr12	51,876,091	ITGB7	NM_000889	SNP	C	A	V326L	4/23,622	0.02	3,759/7,493	50.17
SU080	chr5	56,253,580	**MIER3**	NM_152622	SNP	A	G	UTR	24/4,814	0.50	1,904/3,811	49.96
SU080	chr7	131,503,712	**PLXNA4**	NM_020911	SNP	C	T	V1393M	806/399,887	0.20	200,704/ 402,685	49.84
SU080	chr2	136,060,541	**R3HDM1**	NM_015361	SNP	A	G	UTR	139/32,248	0.43	16,371/30,263	54.10
SU080	chr6	30,146,845	RNF39	NM_170769	SNP	A	G	UTR	67/33,772	0.20	15,036/28,371	53.00
SU080	chr14	99,863,324	SLC25A47	NM_207117	SNP	C	T	S64F	206/197,162	0.10	106,419/ 215,530	49.38
SU227	chr11	111,185,674	ALG9	NM_001077692	SNP	A	T	L368M	4/18,455	0.02	48,323/95,348	50.68
SU227	chr19	38,484,844	**CEBPa (allele 1)**	NM_004364	INDEL	—	nt317_318dupTT	F106fs(+TT)	ND	ND	ND	ND
SU227	chr19	38,484,190	**CEBPa (allele 2)**	NM_004364	INDEL	—	nt971_972ins36	L324fs(+36)	0/3,335	0.00	4,835/17,432	27.74
SU227	chr3	89,472,890	**EPHA3**	NM_005233	SNP	C	A	P317T	1/6,886	0.01	14,427/25,232	57.18
SU227	chr13	27,506,251	**FLT3**	NM_004119	INDEL	—	ins(+21bp)	ITD	7/1,915	0.37	2,119/8,020	26.42
SU227	chr1	243,088,102	HNRNPU	NM_004501	SNP	G	A	P424L	9/11,823	0.08	46,757/94,501	49.48
SU227	chr7	50,417,782	**IKZF1**	NM_006060	SNP	G	A	G158S	14/17,198	0.08	62,853/133,294	47.15

Table S3. Cont.

AML case	Chromosome	Position	Gene name	Accession no.	Mutation type	Ref allele	Var allele	Annotation	T-cell variant ratio	T-cell VAF	Leukemia variant ratio	Leuk. VAF
SU227	chr1	38,033,929	MANEAL	NM_152496	SNP	G	A	UTR	1/1,194	0.08	809/1,739	46.52
SU227	chr3	9,002,005	SRGAP3	NM_001033117	SNP	C	T	UTR	17/12,674	0.13	25,483/69,656	36.58
SU227	chr2	179,181,585	**TTN**	NM_133378	SNP	A	T	D14898E	0/3,517	0.00	2,679/6,179	43.36
SU290	chr12	9,135,105	A2M	NM_000014	SNP	G	A	L810F	64/38,891	0.16	8,461/42,210	20.05
SU290	chr19	38,485,093	**CEBPa**	NM_004364	INDEL	—	nt68delC	P23fs(-C)	ND	ND	ND	ND
SU290	chr18	38,484,251	**CEBPa**	NM_004364	INDEL	—	nt910_912dupAAG	K304(+AAG)	245/47,622	0.51	499/859	58.09
SU290	chr19	9,951,071	**COL5A3**	NM_015719	SNP	G	A	P912L	214/39,565	0.54	18,656/36,403	51.25
SU290	chr13	27,506,225	**FLT3**	NM_004119	INDEL	—	ins(+90bp)	ITD	0/683	0.00	36/991	3.63
SU290	chr12	4,896,293	KCNA1	NM_000217	SNP	T	G	UTR	6/3,478	0.17	2,239/3,766	59.45
SU290	chr1	115,060,270	**NRAS**	NM_002524	SNP	G	A	G12D	98/161,069	0.06	18,535/116,054	15.97
SU290	chr13	87,128,795	SLITRK5	NM_015567	SNP	G	T	UTR	144/129,589	0.11	52,934/85,463	61.94
SU290	chr5	110,436,504	TSLP	NM_033035	SNP	A	G	N64S	166/26,469	0.63	11,834/23,929	49.45
SU291	chr16	80,689,212	HSD17B2	NM_002153	SNP	G	C	K278N	2/78,955	0.00	35,275/71,327	49.46
SU291	chr6	73,975,626	LOC100129128	NT_086342	SNP	G	A	UTR	58/26,720	0.22	10,156/27,886	36.42
SU291	chr17	26,587,134	**NF1**	NM_001042492	SNP	C	T	Q1315*	7/6,821	0.10	1,983/7,894	25.12
SU291	chr5	170,770,149	**NPM1**	NM_199185	Indel	T	+TCTG	ins(+TCTG)	3/41,705	0.01	18,488/36,643	50.45
SU291	chrM	8,335	OK/SW-cl.16	AB064665	SNP	G	A	UTR	1/1,113	0.09	4,557/4,592	99.24
SU291	chr4	106,400,238	**TET2**	NM_001127208	SNP	T	A	C1273S	10/8,131	0.12	4,508/9,521	47.35
SU291	chr4	106,400,290	**TET2**	NM_001127208	SNP	C	T	S1290L	1/1,023	0.10	448/1,010	44.36
SU291	chr8	—	**Trisomy 8**	—	Trisomy	—	—	—				
SU306	chr2	25,323,964	**DNMT3A**	NM_153759	INDEL	A	G	del(V149) (-ACT)	491/164,722	0.30	84,734/166,534	50.88
SU306	chr19	44,125,068	FBXO17	NM_024907	SNP	G	A	UTR	48/105,001	0.05	39,191/81,787	47.92
SU306	chr13	27,490,642	**FLT3**	NM_004119	SNP	C	A	D835Y	1/24,275	0.00	12,906/29,701	43.45
SU306	chrX	9,688,809	GPR143	NM_000273	SNP	G	A	S103L	1/2,935	0.03	2,235/4,454	50.18
SU306	chr15	88,432,938	**IDH2**	NM_002168	SNP	C	T	R140Q	20/8,943	0.22	3,992/8,757	45.59
SU306	chr20	47,424,753	KCNB1	NM_004975	SNP	A	G	S251P	223/57,689	0.39	24,960/50,757	49.18
SU306	chr19	56,020,752	KLK15	NM_138563	SNP	C	T	UTR	9/22,399	0.04	7,387/14,087	52.44
SU306	chr6	150,024,636	LATS1	NM_004690	SNP	T	G	Q1105H	0/10,532	0.00	5,651/11,671	48.42
SU306	chr17	10,368,561	MYH2	NM_017534	SNP	C	T	A1708T	23/20,705	0.11	7,797/16,506	47.24
SU306	chr5	170,770,149	**NPM1**	NM_199185	INDEL	T	+TCTG	ins(+TCTG)	23/45,414	0.05	8,691/17,544	49.54
SU320	chr4	164,726,448	MARCH1	NM_017923	SNP	G	A	S92F	162/479,669	0.03	250,698/ 506,204	49.53
SU320	chr16	54,420,325	CES1	NM_001266	SNP	C	T	V38I	948/208,353	0.45	25,831/166,130	15.55
SU320	chr12	46,654,291	**COL2A1**	NM_033150	SNP	T	A	I1320F	28/15,160	0.18	10,666/21,882	48.74
SU320	chr12	67,550,270	CPM	NM_198320	SNP	C	T	G203D	85/175,702	0.05	98,742/200,371	49.28
SU320	chr19	52,548,106	**DHX34**	NM_014681	SNP	A	C	UTR	2/2,403	0.08	2,288/4,113	55.63
SU320	chr10	103,344,463	DPCD	NM_015448	SNP	C	A	D41E	1/7,118	0.01	4,644/10,462	44.39
SU320	chr19	6,904,054	EMR4P	NR_024075	SNP	T	C	UTR	7/2,630	0.27	1,959/3,533	55.45
SU320	chr12	29,415,848	ERGIC2	NM_016570	SNP	T	G	UTR	1/913	0.11	526/1,043	50.43
SU320	chr13	27,490,635	**FLT3**	NM_004119	SNP	C	G	del(I836)	2/3,073	0.07	979/2,629	37.24
SU320	chr15	88,432,938	**IDH2**	NM_002168	SNP	C	T	R140Q	23/9,075	0.25	5,364/11,553	46.43
SU320	chr22	16,653,634	MICAL3	NM_015241	SNP	T	C	N1958S	33/17,505	0.19	11,119/20,863	53.30
SU320	chrM	13,107	**MTND5**	AF339086	SNP	T	C	—	17/4,189	0.41	3,434/3,442	99.77

Table S3. Cont.

AML case	Chromosome	Position	Gene name	Accession no.	Mutation type	Ref allele	Var allele	Annotation	T-cell variant ratio	T-cell VAF	Leukemia variant ratio	Leuk. VAF
SU320	chr5	170,770,149	**NPM1**	NM_199185	INDEL	T	+TCTG	ins(+TCTG)	25/33,935	0.07	13,854/27,088	51.14
SU320	chr15	43,687,593	PLDN	NM_012388	SNP	T	G	UTR	6/61,759	0.01	40,175/81,253	49.44
SU320	chr12	3,573,358	PRMT8	NM_019854	SNP	C	T	UTR	442/407,976	0.11	209,944/413,041	50.83
SU320	chr3	47,073,404	SETD2 (allele 1)	NM_014159	SNP	G	A	Q2292*	6/42,605	0.01	25,820/56,980	45.31
SU320	chr3	47,133,125	SETD2 (allele 2)	NM_014159	SNP	C	A	M1526I	3/2,144	0.14	4,285/8,578	49.95
SU320	chr20	1,867,292	SIRPA	NM_080792	SNP	G	A	UTR	75/71,993	0.10	22,403/77,941	28.74
SU320	chr16	55,483,538	**SLC12A3**	NM_000339	SNP	G	A	S804N	112/136,738	0.08	48,359/97,575	49.56
SU320	chr17	29,983,847	TMEM132E	NM_207313	SNP	C	G	S408R	3/52,355	0.01	12,746/27,439	46.45
SU320	chr11	129,229,907	TMEM45B	NM_138788	SNP	C	T	A124V	7/3,292	0.21	1,114/3,083	36.13
SU320	chr7	134,539,724	WDR91	NM_014149	SNP	C	A	A243S	10/16,205	0.06	10,834/22,237	48.72
SU336	chr16	272,706	ARHGDIG	NM_001176	SNP	C	T	A190V	7/5,819	0.12	3,272/6,072	53.89
SU336	chrM	6,244	BC018860	BC018860	SNP	G	A	UTR	5/3,490	0.14	5,779/14,756	39.16
SU336	chr2	88,606,297	C2orf51	NM_152670	SNP	C	T	P7L	13/10,036	0.13	28,213/66,096	42.68
SU336	chr4	152,790,420	FAM160A1	NM_001109977	SNP	G	A	V593M	10/10,109	0.10	5,711/13,108	43.57
SU336	chr21	32,293,247	HUNK	NM_014586	SNP	G	A	R675H	3/3,573	0.08	3,348/9,773	34.26
SU336	chr1	78,893,771	**IFI44**	NM_006417	SNP	G	A	R276H	1/1,037	0.10	558/1,220	45.74
SU336	chr12	25,289,548	**KRAS**	NM_004985	SNP	C	T	G13D	2/4,258	0.05	668/1,401	47.68
SU336	chr11–chr19	ND	**MLL-ELL**	—	Translocation				28/3,651	0.77	1,583/2,064	76.70
SU336	chr10	69,640,213	MYPN	NM_032578	SNP	C	T	L1320F	1/1,853	0.05	1,677/3,717	45.12
SU336	chr11	5,733,339	OR52N4	NM_001005175	SNP	C	T	R265C	1/1,026	0.10	3,986/11,297	35.28
SU336	chr11	55,800,090	OR5T1	NM_001004745	SNP	C	A	R134S	0/3,498	0.00	487/1,002	48.60
SU336	chr11	47,333,509	SPI1	NM_001080547	SNP	G	A	R221C	1/455	0.22	44/1,053	4.18
SU336	chr11	12,921,545	TEAD1	NM_021961	SNP	A	G	UTR	0/1,076	0.00	649/1,226	52.94
SU353	chr6	151,716,572	**AKAP12**	NM_005100	SNP	A	G	UTR	11/4,781	0.23	2,596/4,573	56.77
SU353	chr20	30,484,791	**ASXL1**	NM_015338	SNP	C	T	Q377*	44/9,855	0.45	4,723/9,197	51.35
SU353	chr1	92,318,715	BTBD8	NM_183242	SNP	A	T	UTR	134/30,520	0.44	5,637/11,973	47.08
SU353	chr19	4,723,269	C19orf30	NR_027148	SNP	T	C	UTR	7,500/142,318	5.27	66,870/128,230	52.15
SU353	chr15	91,328,582	CHD2	NM_001271	SNP	A	G	M1029V	3/863	0.35	1,382/2,702	51.15
SU353	chr12	122,824,678	**DNAH10**	NM_207437	SNP	G	A	E118K	0/6,003	0.00	529/1,127	46.94
SU353	chr2	25,320,953	**DNMT3A**	NM_022552	SNP	C	A	G543C	256/51,923	0.49	33,086/62,673	52.79
SU353	chr13	27,506,227	**FLT3**	NM_004119	INDEL	—	ins(+57bp)	ITD	0/1,065	0.00	748/1,080	69.26
SU353	chr1	153,473,890	GBA	NM_000157	SNP	C	G	G289R	3/113,963	0.00	20,259/81,220	24.94
SU353	chr19	19,307,544	**MAU2**	NM_015329	SNP	A	T	Q98H	171/44,844	0.38	19,320/36,660	52.70
SU353	chr11	62,120,701	**MTA2**	NM_004739	SNP	T	A	D289V	142/7,516	1.89	1,967/3,966	49.60
SU353	chrY	15,451,498	NLGN4Y	NR_028319	SNP	C	T	UTR	26/7,879	0.33	10,071/10,662	94.46
SU353	chr5	170,770,148	**NPM1**	NM_199185	INDEL	T	+TCTG	ins(+TCTG)	7/71,489	0.01	30,000/58,240	51.51
SU353	chr15	70,344,496	**PARP6**	NM_020214	SNP	G	C	S103C	264/18,980	1.39	1,616/4,673	34.58
SU353	chr14	91,338,428	TC2N	NM_001128596	SNP	T	C	Y131C	753/99,191	0.76	49,937/90,845	54.97

Exome sequencing of 10 patients identified leukemia-specific mutations. Genomic position, gene accession number, and genetic alteration are annotated. Variant allele frequencies (VAF) obtained by targeted amplicon sequencing of these leukemia-specific mutations in T-cell and leukemia gDNA are provided. Genes highlighted in bold are recurrently mutated according to the TCGA Research Network's AML study (1). INDEL, insertion or deletion; Leuk, leukemia; Ref, reference; Var, variant.

1. Cancer Genome Atlas Research Network (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368(22):2059–2074.

Table S4. Analysis of preleukemic and recurrent mutations

Mutation no.	Patient	Gene	Mutation	Preleukemic?	Evidence for preleukemia	TCGA category
1	SU067	RAD21	Q592*	Yes	CD19+ lymphoid cells from transplant	Cohesin complex
2	SU067	PTPN11	A72T	No	—	Activated signaling
3	SU067	CES4	UTR	No	—	Unknown
4	SU067	NPM1	ins(+TCTG)	No	—	NPM1
5	SU067	TCHHL1	G587D	No	—	Unknown
6	SU067	CCBE1	G185V	No	—	Unknown
7	SU072	CACNA1E	UTR	No	—	Unknown
8	SU072	DNMT3A	Q338fs	Yes	CD19+ lymphoid cells from transplant	DNA methylation
9	SU072	SAP130	V630M	Yes	CD33+ myeloid cells from transplant	Chromatin modifiers
10	SU072	LPA	P913S	No	—	Unknown
11	SU072	IDH2	R140Q	Yes	CD19+ lymphoid cells from transplant	DNA methylation
12	SU072	FLT3	ITD	No	—	Activated signaling
13	SU072	NPM1	ins(+GTCA)	No	—	NPM1
14	SU072	GPR115	V324I	No	—	Unknown
15	SU072	ODZ3	UTR	Yes	CD33+ myeloid cells from transplant	Unknown
16	SU080	inv(16)	inversion	Yes	Sorted HSCs	Transcription factor Fusions
17	SU080	MIER3	UTR	Yes	Sorted HSCs	Chromatin modifiers
18	SU080	CBL	Q409fs	No	—	Activated signaling
19	SU080	PLXNA4	V1393M	Yes	Sorted HSCs	Unknown
20	SU080	R3HDM1	UTR	Yes	Sorted HSCs	Unknown
21	SU227	TTN	D14898E	No	—	Unknown
22	SU227	IKZF1	G158S	Yes	Sorted HSCs	Myeloid transcription factors
23	SU227	EPHA3	P317T	Yes	Sorted HSCs	Unknown
24	SU227	CEBPa	F106fs(+TT)	No	—	Myeloid transcription factors
25	SU227	CEBPa	L324ins(+12bp)	No	—	Myeloid transcription factors
26	SU227	FLT3	ITD	No	—	Activated signaling
27	SU290	NRAS	G12D	No	—	Activated signaling
28	SU290	CEBPa	P23fs(-C)	No	—	Myeloid transcription factors
29	SU290	CEBPa	K304ins(+AAG)	No	—	Myeloid transcription factors
30	SU290	FLT3	ITD	No	—	Activated signaling
31	SU290	COL5A3	P912L	No	—	Unknown
32	SU306	IDH2	R140Q	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	DNA methylation
33	SU306	NPM1	ins(+TCTG)	No	—	NPM1
34	SU306	FLT3	D835Y	No	—	Activated signaling
35	SU306	DNMT3A	del(V149)	No	—	DNA methylation
36	SU320	COL2A1	I1320F	Yes	CD19+ lymphoid cells from transplant	Unknown
37	SU320	IDH2	R140Q	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	DNA methylation
38	SU320	FLT3	del(I836)	Yes	CD19+ lymphoid cells from transplant	Activated signaling
39	SU320	DHX34	UTR	Yes	CD19+ lymphoid cells from transplant	Unknown
40	SU320	NPM1	ins(+TCTG)	Yes	CD19+ lymphoid cells from transplant	NPM1
41	SU320	MTND5	—	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	Unknown
42	SU320	SLC12A3	S804N	Yes	CD19+ lymphoid cells from transplant	Unknown
43	SU336	KRAS	G13D	No	—	Activated signaling
44	SU336	MLL-ELL	*t*(11;19)	No	—	Chromatin modifiers
45	SU336	IFI44	R276H	No	—	Unknown
46	SU353	DNMT3A	G543C	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	DNA methylation
47	SU353	ASXL1	Q377stop	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	Chromatin modifiers
48	SU353	AKAP12	UTR	No	—	Activated signaling
49	SU353	PARP6	S103C	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	Unknown
50	SU353	NPM1	+TCTG	No	—	NPM1
51	SU353	MAU2	Q98H	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	Cohesin complex

Table S4. Cont.

Mutation no.	Patient	Gene	Mutation	Preleukemic?	Evidence for preleukemia	TCGA category
52	SU353	MTA2	D289V	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	Chromatin modifiers
53	SU353	DNAH10	E118K	Yes	Single cell-derived colony genotyping	Unknown
54	SU008	FLT3	ITD	No	—	Activated signaling
55	SU014	IDH1	R132H	No	—	DNA methylation
56	SU014	NPM1	ins(+TCTG)	Yes	Sorted HSCs	NPM1
57	SU014	SMC1A	R96H	Yes	Sorted HSCs	Cohesin complex
58	SU014	KAISO	C496Y	Yes	Sorted HSCs	Chromatin modifiers
59	SU014	C10orf76	Q380E	No	—	Unknown
60	SU014	FLT3	ITD	No	—	Activated signaling
61	SU030	NPM1	ins(+TCTG)	No	—	NPM1
62	SU030	SLC12A1	S514R	No	—	Unknown
63	SU030	FLT3	ITD	No	—	Activated signaling
64	SU048	TET2	E1357*	Yes	Sorted HSCs	DNA methylation
65	SU048	TET2	D1384V	Yes	Sorted HSCs	DNA methylation
66	SU048	SMC1A	R711G	Yes	Sorted HSCs	Cohesin complex
67	SU048	NPM1	ins(+TCTG)	No	—	NPM1
68	SU048	FLT3	ITD	No	—	Activated signaling
69	SU070	PHF6	G226R	Yes	Sorted HSCs	Myeloid transcription Factors
70	SU070	DOCK9	A1475V	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	Unknown
71	SU070	CTCF	R11Q	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	Chromatin modifiers
72	SU070	TET2	Y1649s	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	DNA methylation
73	SU070	TET2	T1884A	Yes	Sorted HSCs and CD19+ lymphoid cells from transplant	DNA methylation
74	SU070	FLT3	ITD	No	—	Activated signaling

Data from all patients where exome sequencing, targeted amplicon sequencing, and HSC transplant data were available were used to investigate patterns of mutation acquisition. Recurrent mutations were classified as either preleukemic or not preleukemic based on the presence of these mutations in sorted HSCs or in cells that resulted from xenotransplantation of HSCs. Each mutation was classified into one of the mutation categories proposed by the TCGA Research Network (1) or labeled as "unknown" when no function was easily attributable to a given mutation in the context of human AML.

1. Cancer Genome Atlas Research Network (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368(22):2059–2074.