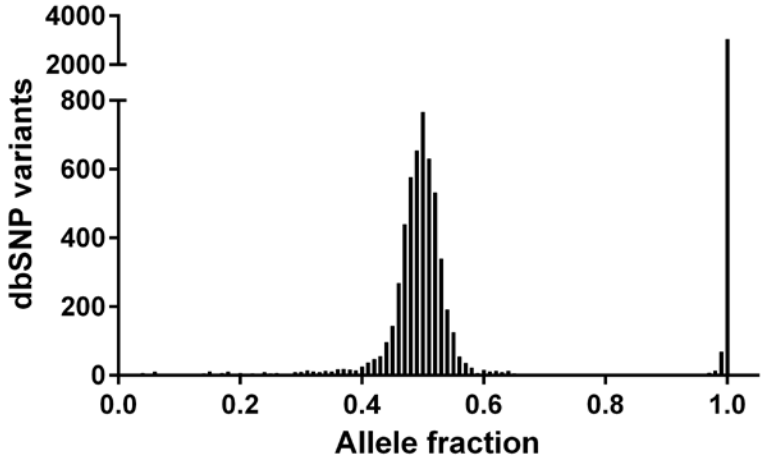
Supplementary Methods

Patient identification. The diagnosis and severity of AA were made according to established criteria. (1) All patients lacked dysplasia in the myeloid and megakaryocytic lineages on bone marrow examination, as well as clonal cytogenetic abnormalities by metaphase karyotyping and/or fluorescence *in situ* hybridization (FISH). No cases had T-cell receptor or immunoglobulin rearrangement, or clonal abnormalities identified by flow cytometry, with the exception of those cases with paroxysmal nocturnal hemoglobinuria (PNH) clones based on absence of CD55/CD59 expression.

Next-generation sequencing. Targeted exon capture followed by next-generation sequencing was performed as previously described. (2) Briefly, a set of unique RNA baits were selected using the Agilent SureSelect E-array program that capture the coding sequence of 219 genes (Supplementary Table 1) previously identified to be recurrently altered in hematologic malignancies. (3-17) Biotinylated RNA baits were synthesized by Agilent for the SureSelect Target Enrichment system. DNA was extracted from each sample (QIAamp DNA Blood Mini Kit, Qiagen) and quantified by Quant-iT PicoGreen®dsDNA Assay Kit (Invitrogen). Barcoded genomic DNA libraries were constructed from 100 ng of dsDNA each, pooled in batches of 24 and subjected to solution-phase hybrid capture utilizing the biotinylated RNA baits by the DFCI Center for Cancer Genome Discovery, as previously described. (2) Paired-end sequencing was performed on a single lane of an Ilumina HiSeq 2000 instrument and the sequence reads were aligned to human reference genome hg19.

Mutant allele validation. Single nucleotide variants (SNVs) and small insertions/deletions (InDels) were identified using the Genome Analysis Toolkit (Broad Institute) and individually visualized with the Integrative Genome Viewer (IGV, Broad Institute). The allele fractions of all dbSNP variants called from the 39 samples are shown in the Figure. As expected, nearly all (96%)



known germline variants from the dbSNP database were called between 40-60% or >95% allele fraction.

To confirm that variants within this range were likely to represent germline SNPs, we selected 15 variant calls present between 40-60% allele fraction in genes previously described to be somatically mutated in MDS for which we had germline DNA (buccal swab) or peripheral blood to sort CD3-positive and CD3-negative cells (see below). All 15 variants in that range were either known germline variants in dbSNP or were confirmed to be present in the germline. Therefore, we focused our validation on variants not within the dbSNP database and present in <40% of reads.

Sanger sequencing was performed to validate all variants and to test paired germline specimens. PCR primers (Supplementary Table 2) were designed with Primer3 (version 0.4.0, http://frodo.wi.mit.edu/primer3/). In addition, for all variants called in <10% of reads, mass spectrometry-based validation was performed using the MASSarray platform (Sequenom) with iPlex chemistry by the DFCI Center for Cancer Genome Discovery. Primers for mass-spectrometry-based validation were designed using Typer 4 software (Sequenom), as previously described.(2) Sanger sequencing was performed at the DF/HCC DNA Sequencing Facility. The mixing experiment with wild-type and mutant *BCR* was performed by TOPO cloning (Invitrogen) of PCR product followed by sequencing of individual bacterial recombinants to identify clones harboring wild-type or mutant *BCR*. These were mixed at 50:50, PCR amplified and subjected to Sanger sequencing.

Clinical analysis. Clinical and demographic data were collected by chart review. Factors were compared in univariate analysis by two-sided Fisher's exact test for categorical variables and two-sided t-test for continuous variables, except where indicated. Analyses were conducted using SAS version 9.2, Cary, NC.

CD3 cell sorting. Samples were sorted using anti-human CD3 magnetic beads (Miltenyi, #130-050-101) and an MS MACS column (Miltenyi). CD3-positive and CD3-negative fraction purity was verified by flow cytometry using anti-human CD3 (clone SK7, Becton Dickinson) on a FACS Aria through the DFCI Division of Hematologic Neoplasia Flow Cytometry Core Facility.

Immunohistochemistry. Immunohistochemistry was performed on 5-µm formalin-fixed, paraffin-embedded tissue sections using anti-beta 2 microglobulin antibody (Dako, #A007202-2) at 1:1800 dilution, as previously described. (18)

References

1. Guinan EC. Diagnosis and management of aplastic anemia. *Hematology Am Soc Hematol Educ Program* 2011; 2011: 76-81.

2. Wagle N, Berger MF, Davis MJ, Bluemensteil B, Defelice M, Pochanard P*, et al.* High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer discovery* 2012; 2(1): 82-93.

3. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R*, et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature genetics* 2010 Feb; 42(2): 181-185.

4. Cheung KJ, Johnson NA, Affleck JG, Severson T, Steidl C, Ben-Neriah S*, et al.* Acquired TNFRSF14 mutations in follicular lymphoma are associated with worse prognosis. *Cancer Res* 2010 Nov 15; 70(22): 9166-9174.

5. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD*, et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 2011 Aug 18; 476(7360): 298-303.

6. Kridel R, Meissner B, Rogic S, Boyle M, Telenius A, Woolcock B*, et al.* Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. *Blood* 2012 Jan 18.

7. Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Kuppers R*, et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 2001 Jul 19; 412(6844): 341-346.

8. Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V*, et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* 2011 Mar 10; 471(7337): 189-195.

9. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A*, et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nature genetics* 2011 Sep; 43(9): 830-837.

10. Cheung KJ, Shah SP, Steidl C, Johnson N, Relander T, Telenius A*, et al.* Genome-wide profiling of follicular lymphoma by array comparative genomic hybridization reveals prognostically significant DNA copy number imbalances. *Blood* 2009 Jan 1; 113(1): 137-148.

11. Rui L, Emre NC, Kruhlak MJ, Chung HJ, Steidl C, Slack G*, et al.* Cooperative epigenetic modulation by cancer amplicon genes. *Cancer cell* 2010 Dec 14; 18(6): 590-605.

12. Steidl C, Shah SP, Woolcock BW, Rui L, Kawahara M, Farinha P*, et al.* MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. *Nature* 2011 Mar 17; 471(7338): 377-381.

13. Lenz G, Davis RE, Ngo VN, Lam L, George TC, Wright GW*, et al.* Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* 2008 Mar 21; 319(5870): 1676-1679.

14. Kato M, Sanada M, Kato I, Sato Y, Takita J, Takeuchi K*, et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* 2009 Jun 4; 459(7247): 712-716.

15. Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB*, et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* 2010 Jan 7; 463(7277): 88-92.

16. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J*, et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* 2012 Mar 22; 366(12): 1079-1089.

17. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G*, et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med* 2011 Jun 30; 364(26): 2496-2506.

18. Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty VV*, et al.* Combined genetic inactivation of beta2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. *Cancer Cell* 2011 Dec 13; 20(6): 728-740.

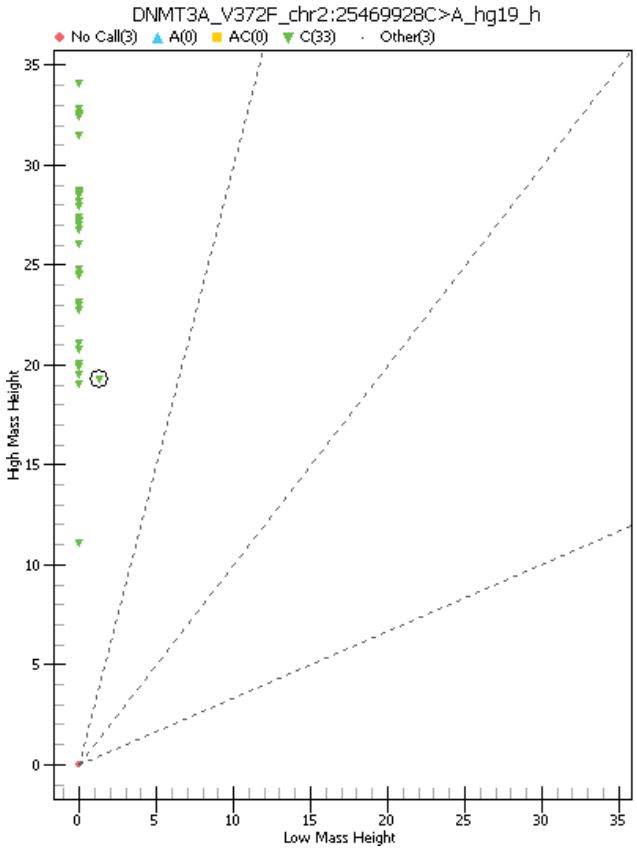
Supplementary Table 1. 219 genes included in targeted exon capture selected based on previous reports that they are recurrently altered in hematologic malignancies.

ABCA7	BTLA	DKK2	HUWE1	MPDZ	POSTN	SUZ12
ABL1	CARD11	DLEU1	IDH1	MUC16	POU2F2	SYK
ACTB	CCDC80	DLEU2	IDH2	MUM1	PRBM1	SYNE1
ALMS1	CCND1	DMD	IKZF1	MYC	PRDM1	TAF1
ALPK2	CCND3	DNMT3A	IKZF2	MYCN	PRKDC	TAL1
APC	CD200	DST	IKZF3	MYD88	PTEN	TBL1XR1
ARID1A	CD36	EBF1	IL7R	NF1	PTPN11	TCF4
ARID1B	CD58	ECT2L	IRAK1	NFKB2	RAPGEF1	TET2
ARID2	CD70	EED	IRF4	NFKBIA	RB1	TMEM30A
ARID3A	CD79A	EP300	IRF8	NOTCH1	REL	TNFAIP3
ASXL1	CD79B	EPHA7	JAK1	NOTCH2	RELN	TNFRSF14
ATM	CDH23	ETS1	JAK2	NOTCH3	RFTN1	TNFSF9
B2M	CDKN2A	ETV6	JAK3	NPM1	RFX7	TP53
BCAS3	CDKN2B	EZH2	KDM4C	NR3C1	RFXAP	TRAF3
BCL10	CDKN3	FAM40B	KDM6A	NRAS	RNF213	TTC27
BCL11B	CELSR2	FAS	KIF20B	NUP214	RUNX1	TYK2
BCL2	CHD1	FBXO11	KIT	OR6K3	SAMD9	U2AF1
BCL6	CHD2	FBXW7	KLHL6	P2RY8	SENP6	UBE2A
BCL7A	CHD8	FLT3	KRAS	PARD3	SETD2	UGGT1
BCL9	CIITA	FOXO1	LEF1	PASD1	SF3B1	ULK4
BCORL1	CMYA5	FYB	LMO2	PASK	SGK1	UNC13B
BCR	CNOT1	G6PD	LRRK1	PAX5	SH2B3	UNC5C
BLNK	COL4A2	GATA1	MALT1	PCLO	SLITRK6	UNC5D
BRAF	COL6A3	GATA2	MAP3K14	PDGFC	SMARCA2	VPS13A
BRD2	CREBBP	GATA3	MAP4K1	PDGFRA	SMARCA4	WT1
BRD4	CRLF2	GNA13	MAPK1	PDGFRB	SMARCB1	XPO1
BRD7	CTCF	GNAS	MEF2B	PDS5B	SOCS1	ZFHX3
BTG1	CTSS	GNB1	MEF2C	PHF6	SRSF2	ZMYM3
BTG2	CUL9	HCK	MIR155	PHLPP1	SRSF8	ZNF608
BTG3	CXCR4	HDAC7	MKI67	PIK3R1	STAT3	ZNF708
BTK	DDX3X	HEATR1	MLL	PIM1	STAT6	ZRSR2
			MLL2	PLCG2		

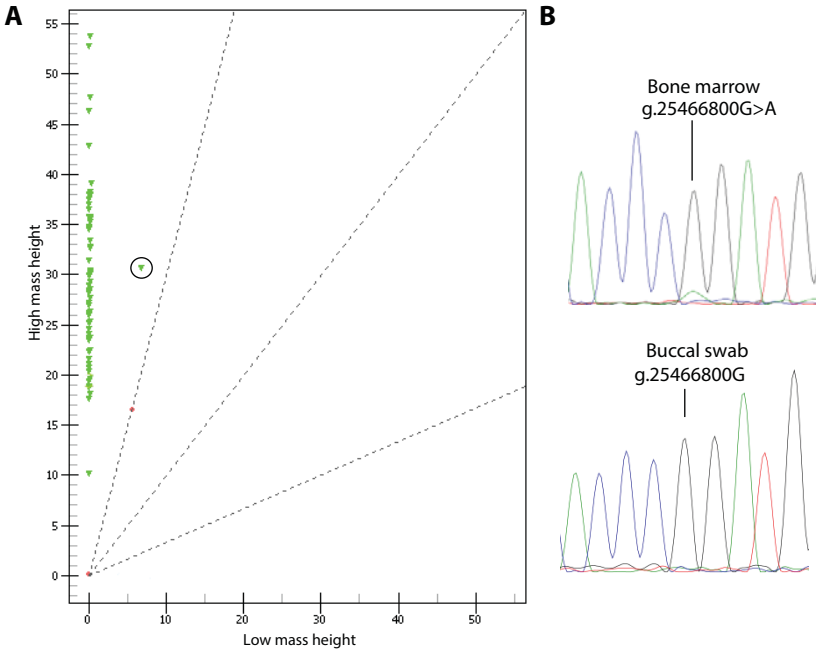
Supplementary Table 2. Primer sequences used to PCR amplify and validate mutant alleles. The forward primer is listed first.

Mutation	**Primers (forward/reverse)**
CELSR2 p.I1556T	TGCCGCCACCATCTTTG / GCTGTCACACACGTTCTTCTTG
SMARCA4 p.E670K	GGTGAGGGCTGTAAGAAGATCAC / AGGAAGCTGGCACAGAACAAC
ASXL1 p.S770*	CAACTACTGCCGCCTTATCCTC / GTGGGTCCTTGCTCCTCATC
PHF6 p.R129*	TCCAGCTAGTGAATTGGGTGAAG / ACAGAGCAAGACTTCGTTTCAGG
DNMT3A p.R635W	TCATTTCGTTTTGCCAGAGTTG / GACCAGGAATTTGTGAGTGCTG
BCR p.R1080C	TACCTTCGTCACCGTCCTCAC / CATTGCCCACACCTCCATC
IKZF3 p.R458C	CATAGGAGCAATCCCTGAGACC / TGAAGAACGCCAGAATCACATC
B2M p.M1I	AAGGAAACTGAAAACGGGAAAG / AAGGAGAACTTGGAGAAGGGAAG
ASXL1 p.D969Efs*4	GTTGGAGAGGAGTGGGAGAAAG / CAGTGGGGCAGATTGGTTC

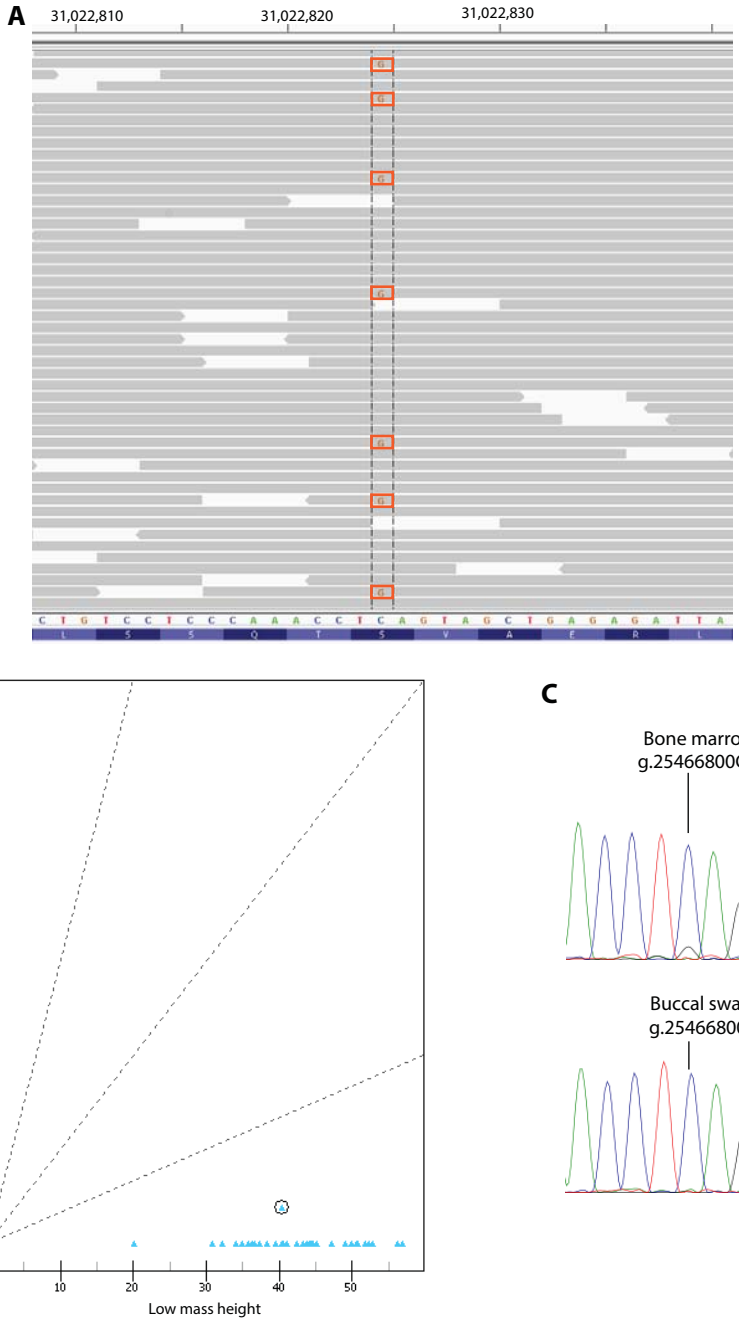
Supplementary Figure 1. Mass spectrometry-based genotyping demonstrates a subclone from a specimen (circled) that harbors *DNMT3A* g.25469280C>A.



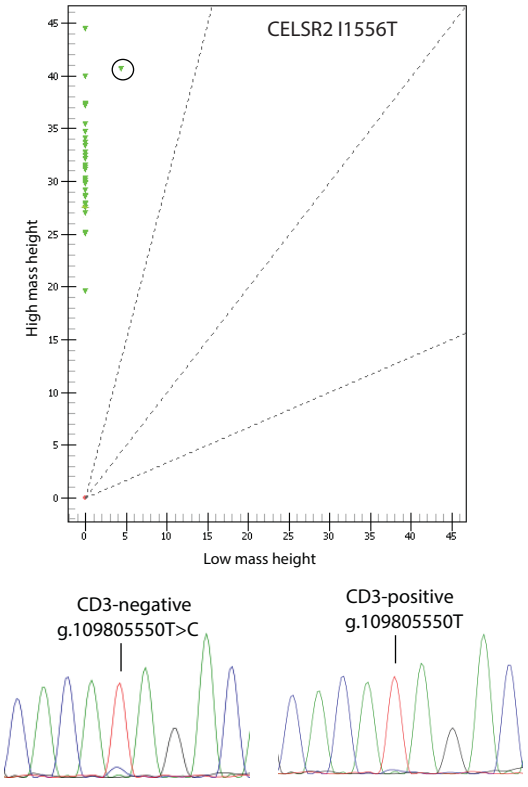
Supplementary Figure 2. **A.** Mass spectrometry-based validation of the *DNMT3A* g.25466800G>A calls. Each triangle represents an individual specimen examined for the mutation. The abundance of the wild-type allele is plotted on the y-axis and the mutant allele on the x-axis. All specimens contained only the wild-type allele except AA13, which is circled. **B.** Sanger sequencing traces showing subtle evidence of the mutation from whole bone marrow and the absence of the mutation in buccal DNA.



Supplementary Figure 3. **A.** Snapshot of sequencing reads from patient AA9 that overlap *ASXL1* g.31022824 showing multiple mutant calls. **B.** Mass spectrometry-based validation of the mutant calls. Each triangle represents an individual specimen examined for the *ASXL1* g.31022824C>G mutation. The abundance of the wild-type allele is plotted on the y-axis and the mutant allele on the x-axis. All specimens contained only the wild-type allele except AA9, which is circled. **C.** Sanger sequencing traces showing subtle evidence of the mutation from whole bone marrow and the absence of the mutation in buccal DNA.



Supplementary Figure 4. Mass spectrometry-based genotyping demonstrates a subclone of specimen AA6 (circled) that harbors *CELSR2* g.109805550T>C mutation. Sanger sequencing of CD3-negative and CD3-positive sorted populations shows a notable mutation in the CD3-negative fraction but little if any mutation present in the CD3-positive population.



Supplementary Figure 5. Next-generation sequencing can identify expansion of an ASXL1 mutation. **A.** Reads spanning *ASXL1* g.31023451 from a specimen obtained 9 months after diagnosis demonstrate a recurrent deletion (indicated by black horizontal bars within individual reads). **B.** The *ASXL1* g.31023451delCATT deletion is detectable by Sanger sequencing in unsorted, CD3-negative, and CD3-positive populations. **C.** Sequencing of a specimen obtained 9 months previously, at the time of diagnosis, demonstrates a single read containing the *ASXL1* g.31023451delCATT deletion from among a subset of the 624 reads covering that position.