List of Supplementary Materials

“Aberrant oligoclonal hematopoiesis in remission AML and relapse from rare cells genomically resembling leukemic blasts"

Supplementary Methods

Exome capture and massively parallel whole exome sequencing

Paired tumor and normal genomic DNA samples were fragmented using a Covaris S2 fragmentation system to a target size of 300bp. The samples were end-repaired, a-tailed, and custom adapters were ligated using the NEBNext DNA Library Prep Kit (NEB) according to the manufacturer's recommended protocols. The barcode base composition was balanced across eight barcodes which were repeated three times across the 24 samples. After ligation, the samples were size selected to 300bp on a 2% agarose gel and 1mm gel slices were retained. Samples were isolated from the gel using the Qiagen QIAquick gel extraction system. Seven µl of each ligation product was enriched using the Phusion master mix kit and custom PCR primers for a total of 14 cycles of PCR amplification. Library QC was performed using the Agilent Bioanalyzer and Q-PCR. Guided by Q-PCR concentrations, 100 ng of each of twelve samples were pooled for a total of four pools. Each pool was captured using the Nextera Rapid Capture Expanded Exome Kit (Illumina) according to the manufacturer's recommended protocols. Pooled samples were then sequenced on the Illumina HiSeq 4000 platform across eight lanes with paired-end reads of 150 bp.

Bioinformatic pipeline analysis of whole exome sequencing data

The exome sequencing data were analyzed by the variant calling pipeline developed by the University of Michigan Bioinformatics Core. For each of the samples, paired-end reads were aligned to the hg19 reference genome using BWA v0.7.8, followed by removal of sequence duplicates using PicardTools v1.79 (http:// picard.sourceforge.net), local realignment around INDELs, and base quality score recalibration using GATK v3.2-2. Read coverage on exome capture target regions was calculated using BEDTools v2.20.1. Normal-tumor paired alignment files were submitted to MuTect v1.1.4, Strelka v1.0.14, and Varscan v2.3.7 (with its false-positive filter) for the detection of somatic and germline SNPs and INDELs. Candidate variant calls across all samples and patients were merged using Jacquard into a single VCF file that included all variant loci whose filter field passed in MuTect or Strelka or VarScan (VarScan calls were limited to somatic variants confirmed in false-positive filter). Variants were annotated using SnpEff v4.0/hg19, dbNSFP v2.4, dbSNP v138, and 1000 Genomes v3. Common variants (at or above 5% overall population allele frequency as reported by 1000 Genomes) were excluded.

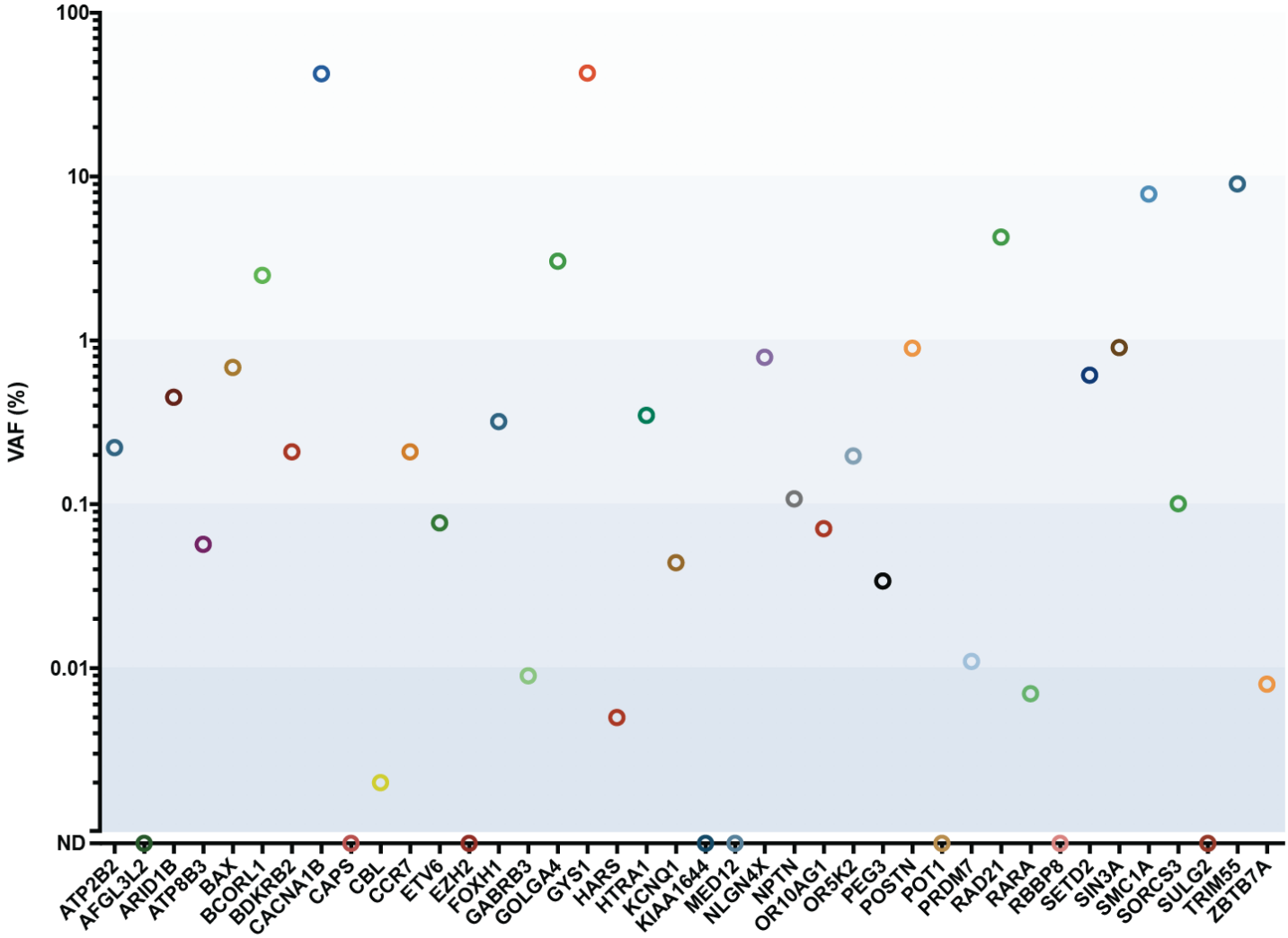

Figure S1: VAF in complete remission samples of genes mutated in only a single case. VAF = variant allele frequency, ND = not detected. The mean of duplicate measurements is depicted. VAF% is on a $\log_{10}$-scale.

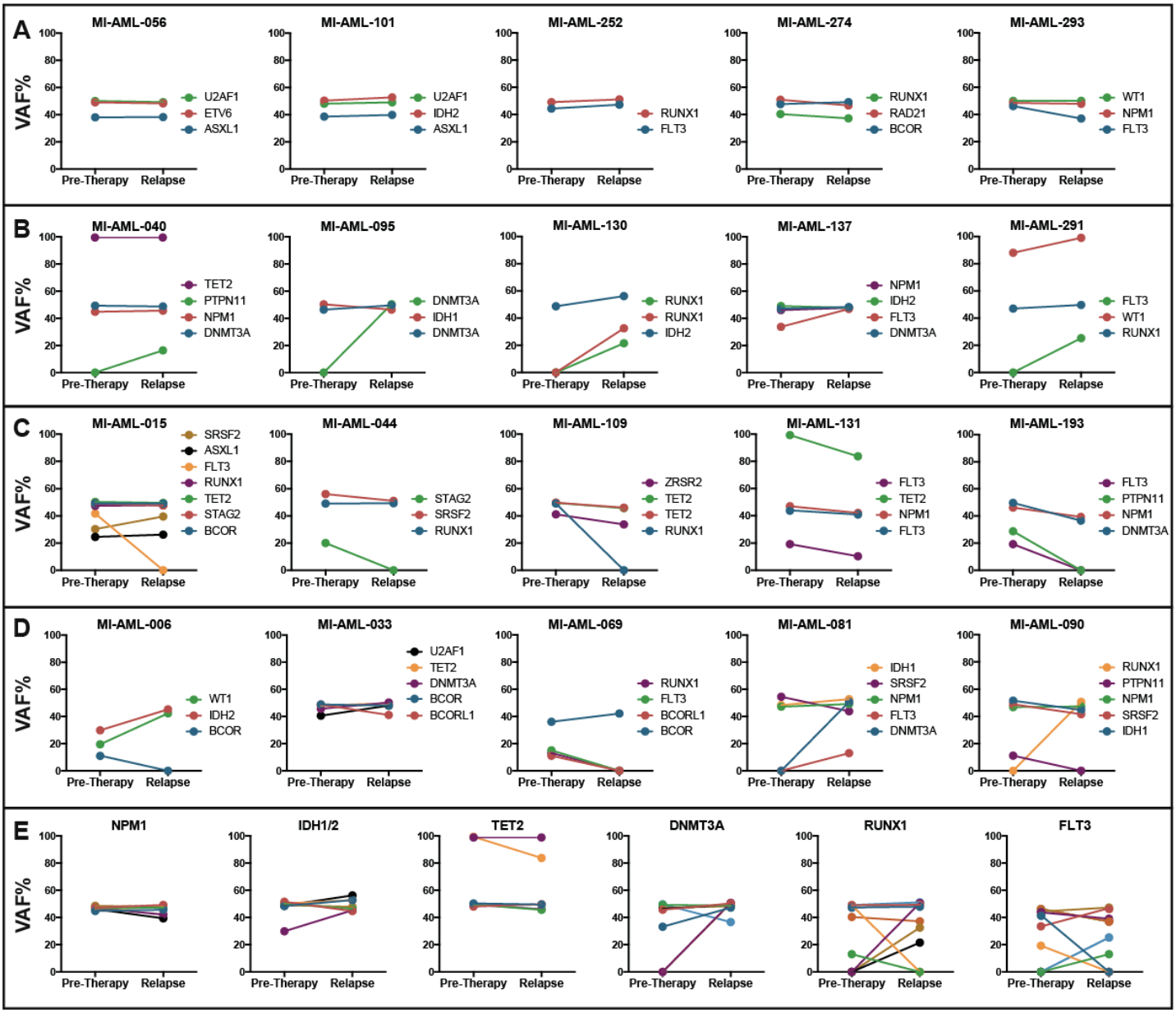

Figure S2: VAF of gene mutations in paired pre-therapy and relapse samples. Samples demonstrated patterns of stable VAF (**A**), acquisition or enrichment of subclones (**B**), loss or decrease of dominant pre-therapy clone (**C**), or combined loss and gain of clones (**D**). Comparison of allele frequency between paired pre-therapy and relapse specimens for specific genes (**E**). The mean of duplicate measurements is depicted. VAF = variant allele frequency.

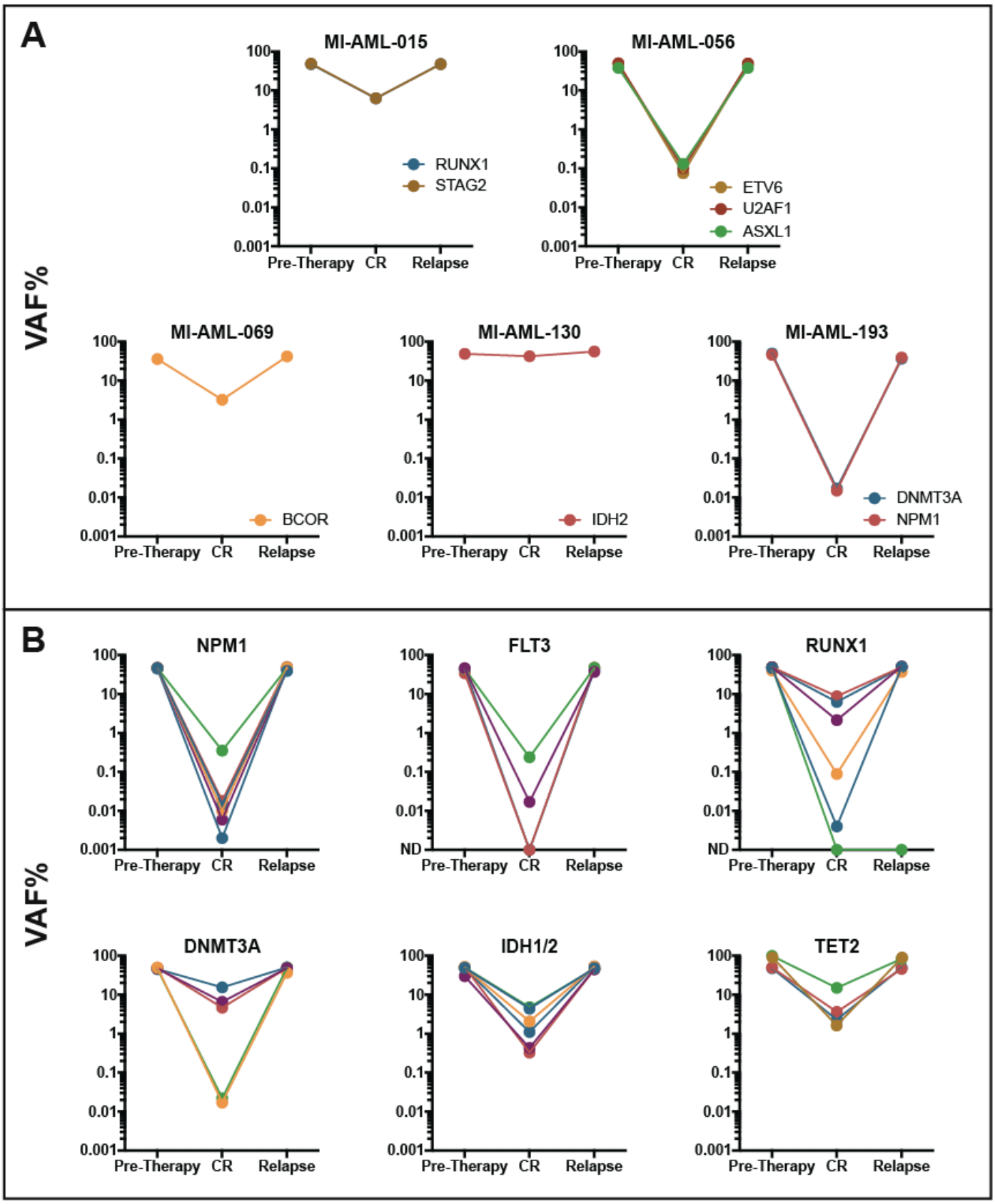

Figure S3: VAF of gene mutations in additional matched pre-therapy, complete remission, and relapse specimens. (A) Additional cases with concordant or single gene mutation measurements. **(B)** Comparison of allele frequency between matched pre-therapy, complete remission, and relapse specimens for specific genes. The mean of duplicate measurements is depicted. VAF% is on a log10-scale. VAF = variant allele frequency, CR = complete remission, ND = not detected. VAF of genes present on the X chromosome were corrected for male patients by a factor of two to accurately depict clonal composition of pre-therapy and relapse specimens.

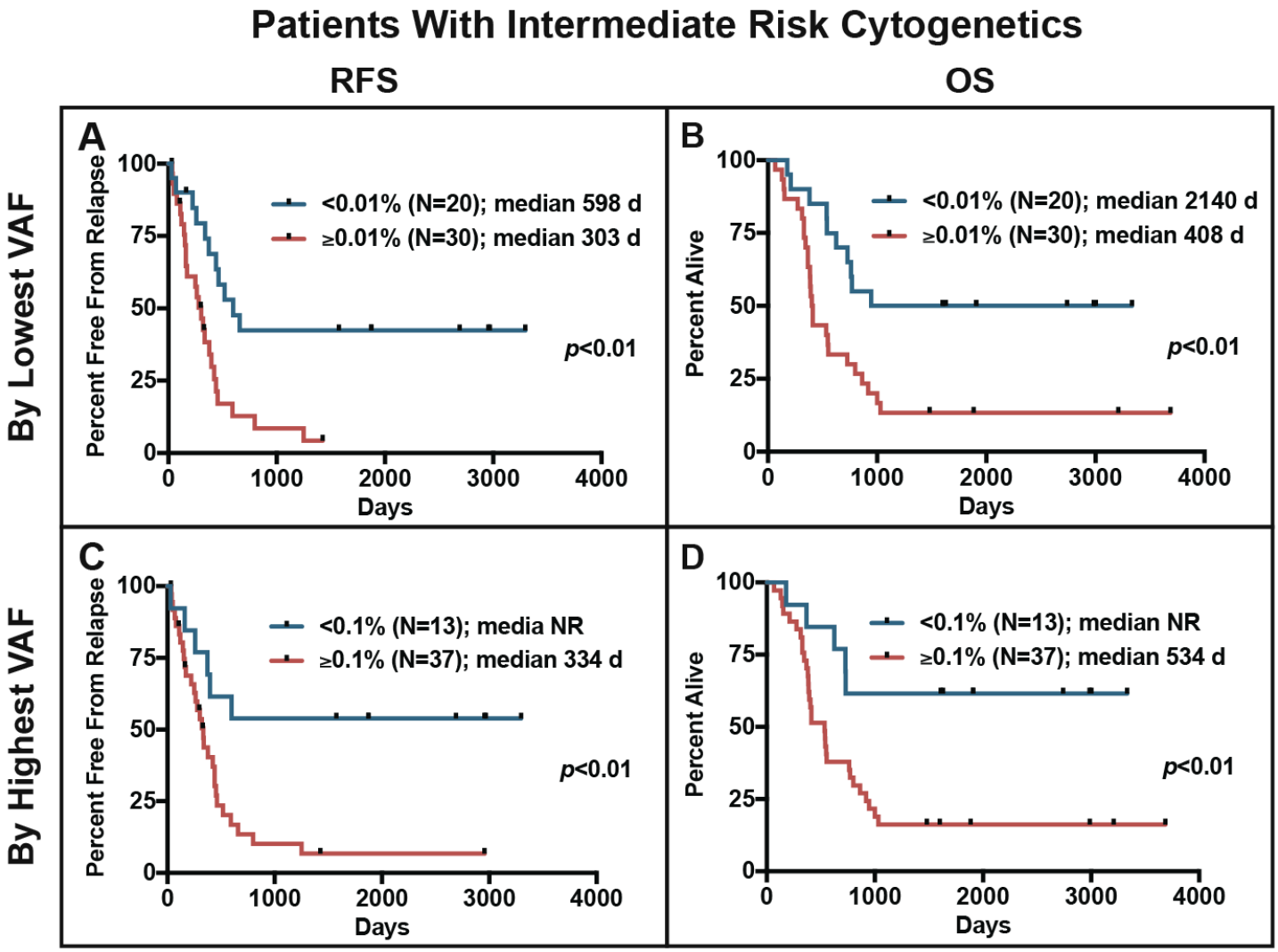

Figure S4: RFS and OS of patients with intermediate risk cytogenetics by VAF of gene mutations in Initial CR. RFS (**A**) and OS (**B**) using lowest measured VAF per case categorized by a dichotomous cutoff of 0.01%. RFS (**C**) and OS (**D**) using highest measured VAF (excluding *DNMT3A*) per case categorized by a dichotomous cutoff of 0.1%. Significance was determined using the log-rank test (Mantel-Cox). VAF = variant allele frequency, RFS = relapse-free survival, OS = overall survival, d = days, N = number of cases.

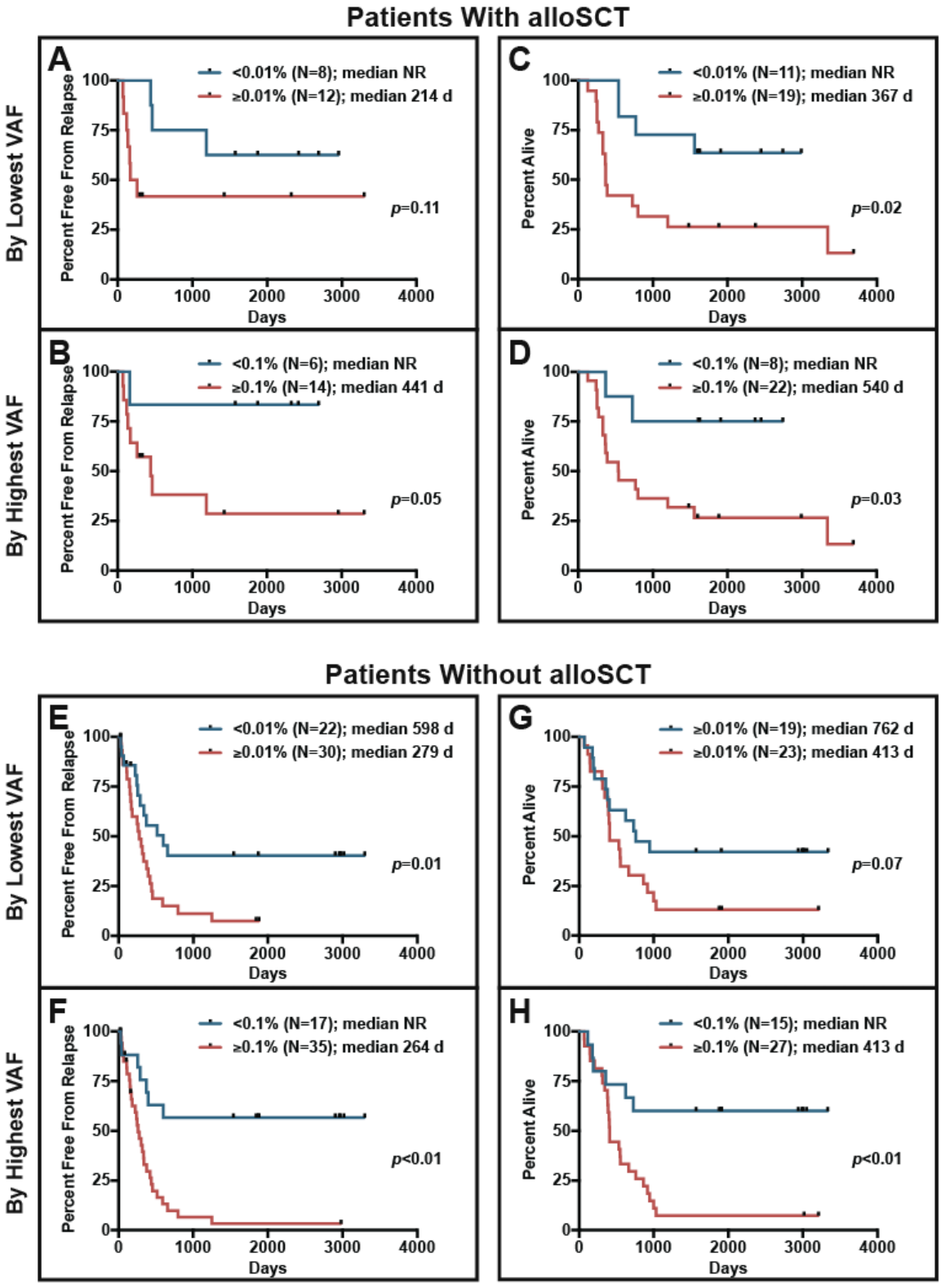

Figure S5. RFS and OS of patients with or without allogeneic stem cell transplantation (alloSCT) by VAF of gene mutations in Initial CR. RFS (**A, E**) and OS (**C, G**) using lowest measured VAF per case categorized by a dichotomous cutoff of 0.01%. RFS (**B, F**) and OS (**D, H**) using highest measured VAF (excluding *DNMT3A*) per case categorized by a dichotomous cutoff of 0.1%. Patients were included in the alloSCT group (**A-D**) for RFS analysis if they received alloSCT in Initial CR and for OS analysis if they received an alloSCT at any time after Initial CR. Patients not meeting these criteria for RFS or for OS analyses were included in the no alloSCT RFS or OS groups, respectively (**E-H**). Significance was determined using the log-rank test (Mantel-Cox). VAF = variant allele frequency, RFS = relapse-free survival, OS = overall survival, d = days, N = number of cases.

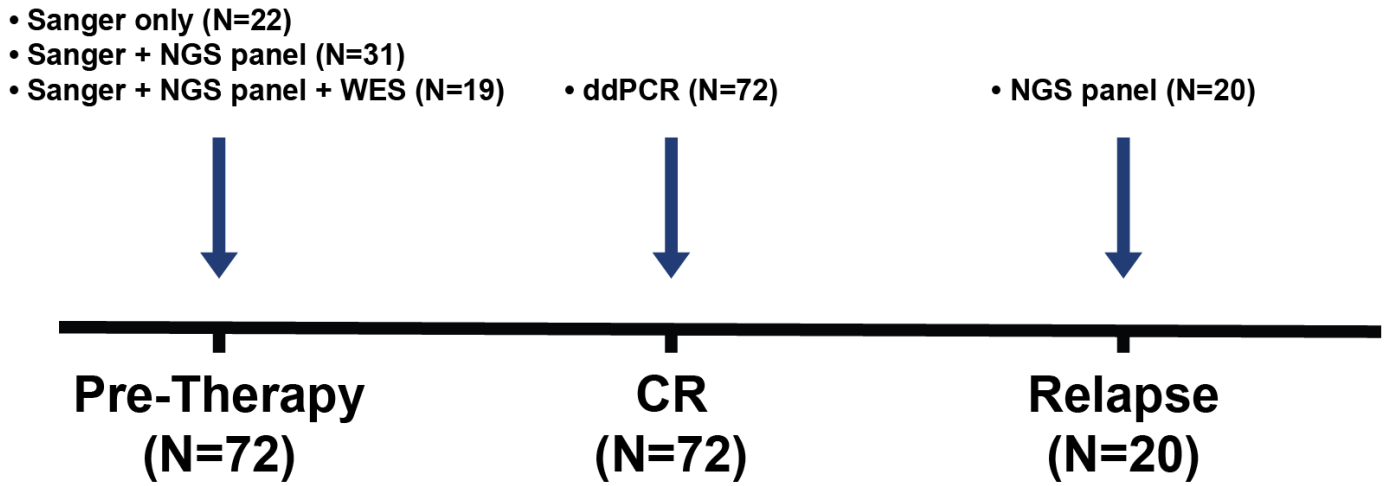

Figure S6. Sequencing methods used to detect gene mutations for pre-therapy, complete remission, and relapse samples. All pre-therapy samples (N=72) were initially screened for somatic mutations using Sanger resequencing of 13 recurrently mutated genes in AML. Pre-therapy samples either with ≤1 mutations after Sanger sequencing or with paired relapse (N=50) underwent targeted NGS sequencing with the 49 gene myeloid panel, which captured most recurrently mutated genes in AML. The remaining pre-therapy cases with ≤1 mutations despite Sanger and the myeloid panel (N=19) underwent WES, which identified ≥2 somatic mutations for each case. Mutations identified in pre-therapy samples were tracked into CR samples using ddPCR for all cases and all mutations assayed. Twenty cases also had paired relapse samples; these relapse samples were assayed with the 49 gene myeloid NGS panel to determine persistence, gain, or loss of mutations in these recurrently mutated genes at relapse. NGS = next generation sequencing, WES = whole exome sequencing, ddPCR = droplet digital PCR, CR = complete remission.

Supplementary Table Legends

Table S1: Detailed Patient Characteristics. AML-RGA = AML with recurrent genetic abnormalities; AML-MRC = AML with myelodysplasia-related changes; AML-TR = therapy-related AML; AML-NOS = AML, not otherwise specified, alloSCT = allogeneic stem cell transplantation, "3+7" = standard induction chemotherapy with 3 days of an anthracycline and 7 days of continuous infusion cytarabine.

Table S2: VAF of Individual Mutations by Case Detected by ddPCR in Complete Remission. Genomic coordinates based on human reference hg18/NCBI36 or hg19/GRCh37 as noted. Remission VAF based on average of duplicate ddPCR measurements for each mutation. ITD = Internal tandem duplication, VAF = variant allele frequency, ND = not detected.

Table S3: VAF of Mutated Genes in Matched Pre-Therapy, CR, and Relapse Samples. Pre-therapy and relapse samples measured with NGS. CR sample measured with ddPCR; average of replicate ddPCR reactions shown. VAF = variant allele frequency, ND = not detected, CR = complete remission.

Table S4: Wildtype Controls and Level of Blank. Median and mean mutant allele signal in ddPCR wildtype only controls. Total number of separate measurements of probes for specific genes and for all probes is shown along with percent of measurements in which no mutant allele signal was detected (% zero). Level-of-blank (LOB) was determined using the formula LOB = $(mean_{wt})+(1.645*SD_{wt})$ where $mean_{wt}$ is the mean mutant signal in wildtype only controls and SD_{wt} is the standard deviation of $mean_{wt}$.