

# Supplementary data

## SUPPLEMENTARY MATERIALS AND METHODS

### FLT3-ITD analysis

Presence of FLT3-ITD was verified by PCR on genomic DNA as previously described.<sup>1</sup> PCR was performed using primers 5'-GCAATTTAGGTATGAAAGCCAGC-3' (forward) and 5'-CTTTCAGCATTGACGGCAACC-3' (reverse) and analyzed on 2.5% agarose gel.

### Immunoblotting

Protein lysates were prepared by lysing cells in 2x SDS sample buffer, resolved on 7% SDS-PAGE gel and transferred to Immobilon-P PVDF membrane (Millipore). Anti-ARID1B antibody (1:250 dilution) (Santa Cruz; sc-32762) was used to determine ARID1B protein expression. The membrane was stripped before probing with anti- $\beta$  actin antibody (1:5000 dilution) (Sigma; A1978).

### Growth assay, colony assay and differentiation of NB4 cells

To assess growth in liquid medium, cells were plated in 96-well plate and viability was assessed using MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide] reagent. Fold change was calculated by normalization of absorbance values (570 nm) at each time point to absorbance readings on the day of plating. For clonogenic assay, cells were cultured in methylcellulose-based semisolid media (H4230; StemCell Technologies). Cells were plated in duplicate in 24-well plate and colonies were enumerated after 9 days. For differentiation assay, NB4 cells at a density of 0.5 million/ml, were incubated with different concentrations of ATRA. Cells were harvested on indicated days and washed before staining with anti-human CD11b antibody. Cells

were acquired on BD LSR II (BD Biosciences) and Sytox Blue Dead Cell Stain (Invitrogen) was used to exclude non-viable cells.

### **Gene Ontology and GSEA analysis**

Gene Ontology (GO) was performed on differentially expressed genes using goseq Bioconductor package (version 1.20.0), which accounts for bias due to gene length.<sup>2</sup> Resulting p-values were adjusted for False Discovery Rate (FDR). GO terms were considered as enriched based on  $FDR < 0.05$ . For Gene Set Enrichment Analysis (GSEA), we used all measured genes and ranked them according to the fold change. This ranked file was then uploaded into GSEAPreranked (version 2.2) module on GenePattern public server (<http://www.broadinstitute.org/cancer/software/genepattern/>) for enrichment analysis.<sup>3</sup>

### **Allele-specific RT-PCR analysis for RARA and PML mutations**

RT-PCR primers specific for wild-type PML, wildtype RARA and PML-RARA fusion alleles were used to amplify the transcript sequences harboring mutations of PML and RARA genes. For analysis of RARA gene, nested PCR were performed in some cases with insufficient quantity of DNA. PCR were analyzed on agarose gel, DNA was extracted and analyzed by Sanger sequencing. Primer sequences are provided below.

### **RNA sequencing and expression analysis**

RNA was extracted from ARID1B knockdown and control NB4 cells using Qiagen RNeasy kit. cDNA libraries were prepared from poly-A selected RNA using TruSeq sample preparation kit (Illumina). Libraries were sequenced on HiSeq 2000 and 100 bp paired-end reads were aligned to human reference genome (NCBI build 37/hg19) using STAR version 2.4.1d splice-aware aligner with default parameters.<sup>4</sup> For gene

expression analysis, FPKM expression units were calculated using Cufflinks program.<sup>5</sup> To determine differential gene expression, uniquely aligned reads overlapping transcript exons and junctions were assigned to Ensembl transcripts (version 75 gene model) using featureCounts (version 1.4.6-p3).<sup>6</sup> These read counts were then used to classify genes as differentially expressed (False Discovery Rate; FDR < 0.05) using DESeq2 (1.8.1) or DEseq (version 1.20.0) Bioconductor package.<sup>7,8</sup> All other statistical tests were done in R (version 3.2.1).

The RNA-Seq data have been deposited in the NCBI Gene Expression Omnibus under accession code GSExxxxx.

### **Code availability**

Code used for read alignment and variant calling is hosted at: [https://github.com/PoisonAlien/somatic\\_pipeline](https://github.com/PoisonAlien/somatic_pipeline)

Code used for generating oncoplots at: <https://github.com/PoisonAlien/maftools>

## List of primers

### RT-PCR primers for allele-specific mutation analysis

#### PML mutations

h PML exon 2 For	GACTTCTGGTGCTTTGAGTGCG
h RARA exon 3 Rev	TGGGCACTATCTCTTCAGAACTGC
h PML exon 4 Rev	GGGCTGGCTTTCTTGATACAG

#### RARA mutations

h PML exon 3 For	GATGGCTTCGACGAGTTCAAGG
h RARA exon 9 Rev	CTGTCCGCTCAGAGTGTCCAG
h RARA exon 2 For	CTCTGACCACTCTCCAGCACC
h RARA exon 5/6 For	CAGCTGGGCAAATACACTACGA
h RARA exon 7 Rev	GTTGGCGAAGGCAAAGACCA
h RARA exon 4 For	CAGAAGTGCTTTGAAGTGGGCA

### qRT-PCR primers for ARID1B and ARID1A expression analysis

h ARID1B qPCR F1	CCAGTCAACTGGCAGCAATTC
h ARID1B qPCR R1	CACTCACATCTGAGAATGGGTCCG
h ARID1B qPCR F2	CTCAGTAGAAGCATCACAACCCAGTC
h ARID1B qPCR R2	CAAAGACCTGATTTAAGGGACATC
h ARID1A qPCR F1	AAGCCACCAACTCCAGCATC
h ARID1A qPCR R1	CGCTTCTGGAATGTGGAGTCAC
h ARID1A qPCR F2	GCATTAGATACCATCAACATCCTGC
h ARID1A qPCR R2	AAAGATCTCAATCAGGCATCGTC

## qRT-PCR primers for ARID1B target genes

h GATA1 F	GCAGCCTATTCCTCTCCCAAG
h GATA1 R	ACAGTGGAGTGGCTGTTGCTC
h GATA2 F	CAGCAAGGCTCGTTCCTGTTC
h GATA2 R	CGGTTCTGCCCATTCATCTTG
h NFE2 F	AGCTAGGAGAGATGGAAGTACTTG
h NFE2 R	CTCAAATGATGGCTCACTTGGAG
h IKZF2 F	CTGCTGTCATAGAGAAGCTCACG
h IKZF2 R	TGGGTAGCTGAATCGCATGAG
h PRG2 F	GTCAAGTCTGGATTGGAGGCAG
h PRG2 R	CAGCCCAGTACGCAAAGTTC
h BPI F	AAGCACTCAAATATTGGCCCCTTC
h BPI R	TCTGTAGTTTCTCGTTAACCCCTGG
h PRTN3 F	CATAACATTTGCACTTTCGTCCCT
h PRTN3 R	CCCAGATCACGAAGGAGTCTATTC
h IRF5 F	TGAGCTCATCCTGTTCCAAAAGG
h IRF5 R	GAGCTTCTTCTCTCGGGGTTTG
h WNT7B F	GCCTCATGAACCTGCATAACAATG
h WNT7B R	CGCGTTGTA CTTCCTTCAG
h HOXC4 F	AAGCAACCCATAGTCTACCCATG
h HOXC4 R	TTCTCTAATTCCAGGACTTGCTGC
h GAPDH F	GGAAGGTGAAGGTCGGAGT
h GAPDH R	CCATGGGTGGAATCATATTGGA

## REFERENCES

1. Frohling S, Schlenk RF, Breitruck J, Benner A, Kreitmeier S, Tobis K, *et al.* Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. *Blood* 2002; **100**: 4372-4380.

2. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 2010; **11**: R14.
3. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; **102**: 15545-15550.
4. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; **29**: 15-21.
5. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012; **7**: 562-578.
6. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; **30**: 923-930.
7. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010; **11**: R106.
8. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550.

## SUPPLEMENTARY FIGURE LEGENDS

### Supplementary Figure S1.

Schematic shows the approach used to identify somatic variants in APL samples without matched germline controls.

### Supplementary Figure S2.

Box plots show proportion of bases covered at 10x and 30x sequencing depth for samples analyzed by exome sequencing in discovery cohort (Dx; primary APL; CR: complete remission; Rel: Relapse APL).

### Supplementary Figure S3.

Venn diagrams depict genes which harbor non-silent somatic mutations at diagnosis and relapse of eight paired APL samples analyzed by exome sequencing.

### Supplementary Figure S4.

**Coverage metrics of APL frequency set.** Box plots depict coverage for targeted sequencing of newly diagnosed (Dx), complete remission (CR) and relapse (Rel) samples. Proportion of bases covered by either at least 10 or 30 reads are shown.

### Supplementary Figure S5.

Bars represent proportion of different types of base substitutions observed at primary and relapsed APL. Disease samples with matched germline controls analyzed either with whole exome or targeted exome sequencing were used for this analysis.

### Supplementary Figure S6.

Schematic representation of somatic mutations identified in known leukemia genes in our APL cohort. Location and type of mutations validated by Sanger sequencing in *FLT3*, *WT1*, *NRAS* and *KRAS* genes at both primary and relapse APL are shown.

Circles are color coded for different types of mutations. Ig: Immunoglobulin, ZnF: Zinc finger

#### **Supplementary Figure S7.**

**Comparison of mutational frequencies of selected leukemia genes between APL and other AML subtypes.** Frequencies of FLT3-ITD and somatic mutations in *FLT3*, *NRAS*, *KRAS*, *EZH2* and *ETV6* genes validated in primary APL (n=165) in this study were compared to the frequencies in non-M3 AML (n=176) from TCGA dataset.

#### **Supplementary Figure S8.**

**RT-PCR analysis to determine the allele harboring *RARA* mutations.** Wild-type *RARA* and *PML-RARA* fusion alleles were amplified from cDNA of 3 relapse cases (4 mutations), and the presence of mutation was examined by Sanger sequencing of both amplicons. Red arrows indicate the presence of mutated nucleotide, while the black arrows show the corresponding germline nucleotide.

#### **Supplementary Figure S9.**

**Allele specific analysis for *PML* mutations.** RT-PCR analysis was performed for wild-type *PML* and *PML-RARA* alleles in 3 relapse cases, and the allele harboring the mutation was determined using Sanger sequencing. Somatic mutations are indicated by red arrows.

#### **Supplementary Figure S10.**

***ARID1A* expression in *ARID1B* knockdown cells.** qRT-PCR analysis for *ARID1A* transcript levels in NB4 cells stably expressing shRNAs (either sh2 or sh3) targeting *ARID1B*. Levels of *GAPDH* transcript were used to normalize for RNA. Data are represented as mean  $\pm$  SEM.

#### **Supplementary Figure S11.**

**Stable knockdown of *ARID1B* does not alter the growth of NB4 cells in liquid culture.** Growth kinetics of control and *ARID1B* knockdown (either sh2 or sh3) cells in liquid culture. Cells were plated in 96-well plate in six replicates, and viability was



assessed using MTT assay every 24 hours (n=8). Absorbance (570 nM) values at each time point were normalized to the respective readings on the day of plating.

### **Supplementary Figure S12.**

**Flow cytometric analysis of CD11b expression in ARID1B knockdown cells.** NB4 cells transduced with either control shRNA or ARID1B sh2 or sh3 were analyzed by flow cytometry for surface expression of CD11b marker. Representative overlay of histograms shows expression of CD11b in untreated control and ARID1B knockdown cells.

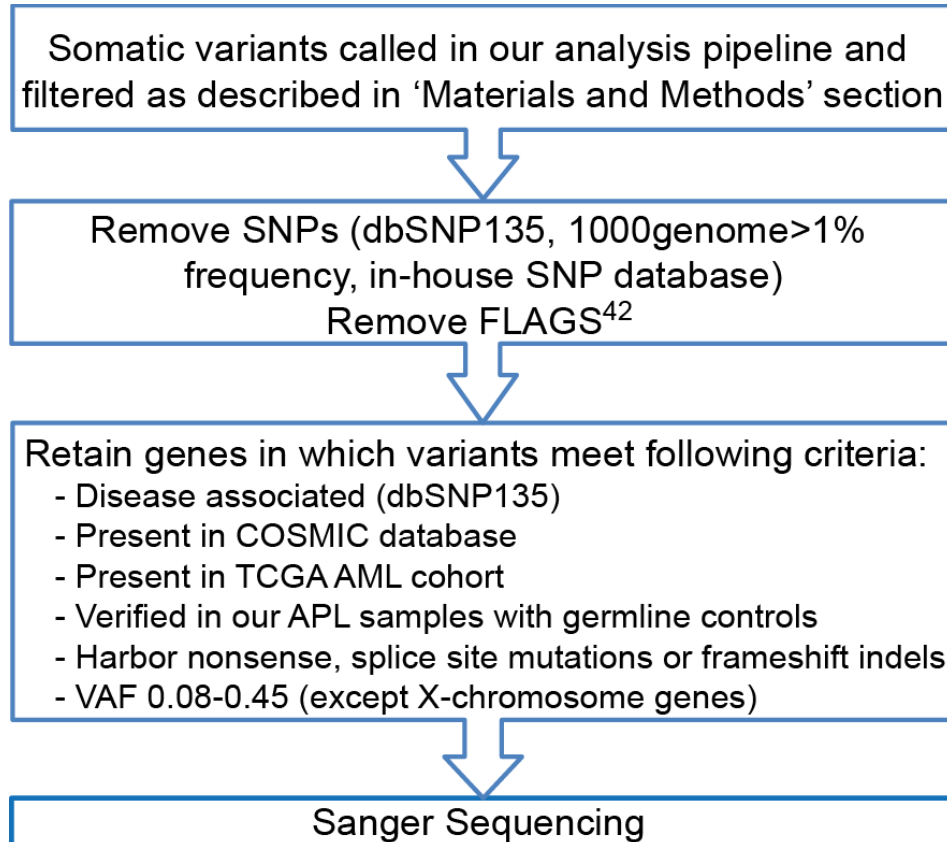
### **Supplementary Figure S13.**

**Silencing of ARID1B predominantly results in upregulation of target genes.** Distribution of genes upregulated or downregulated (FDR<0.05) in ARID1B knockdown (sh2 and sh3) vs control NB4 cells in RNA-Sequencing is depicted.

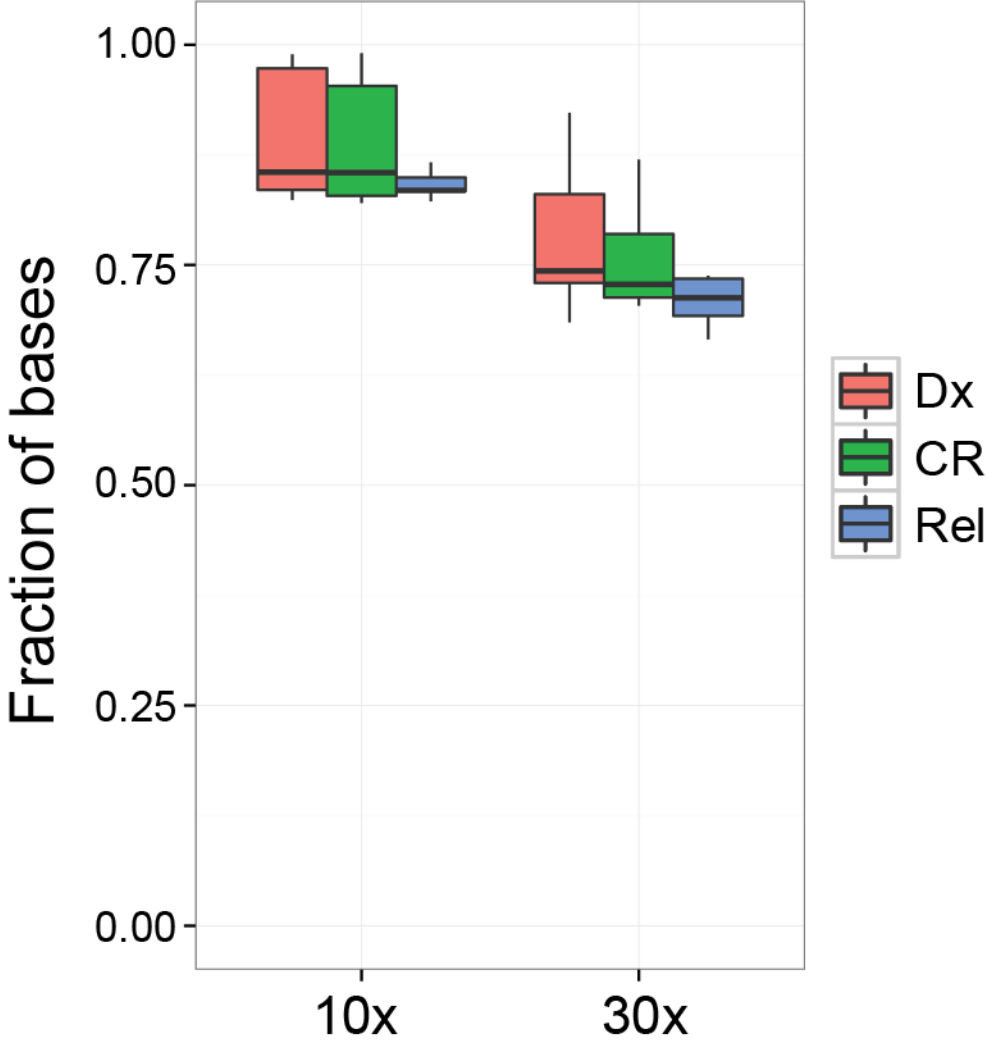
### **Supplementary Figure S14.**

**GSEA analysis of genes upregulated in ARID1B knockdown NB4 cells.** Genes upregulated in ARID1B knockdown NB4 cells were analyzed by GSEA and enriched genesets (FDR<0.25) were identified. Enrichment plots for genesets representing downstream targets of GATA1 and GATA2 are shown.

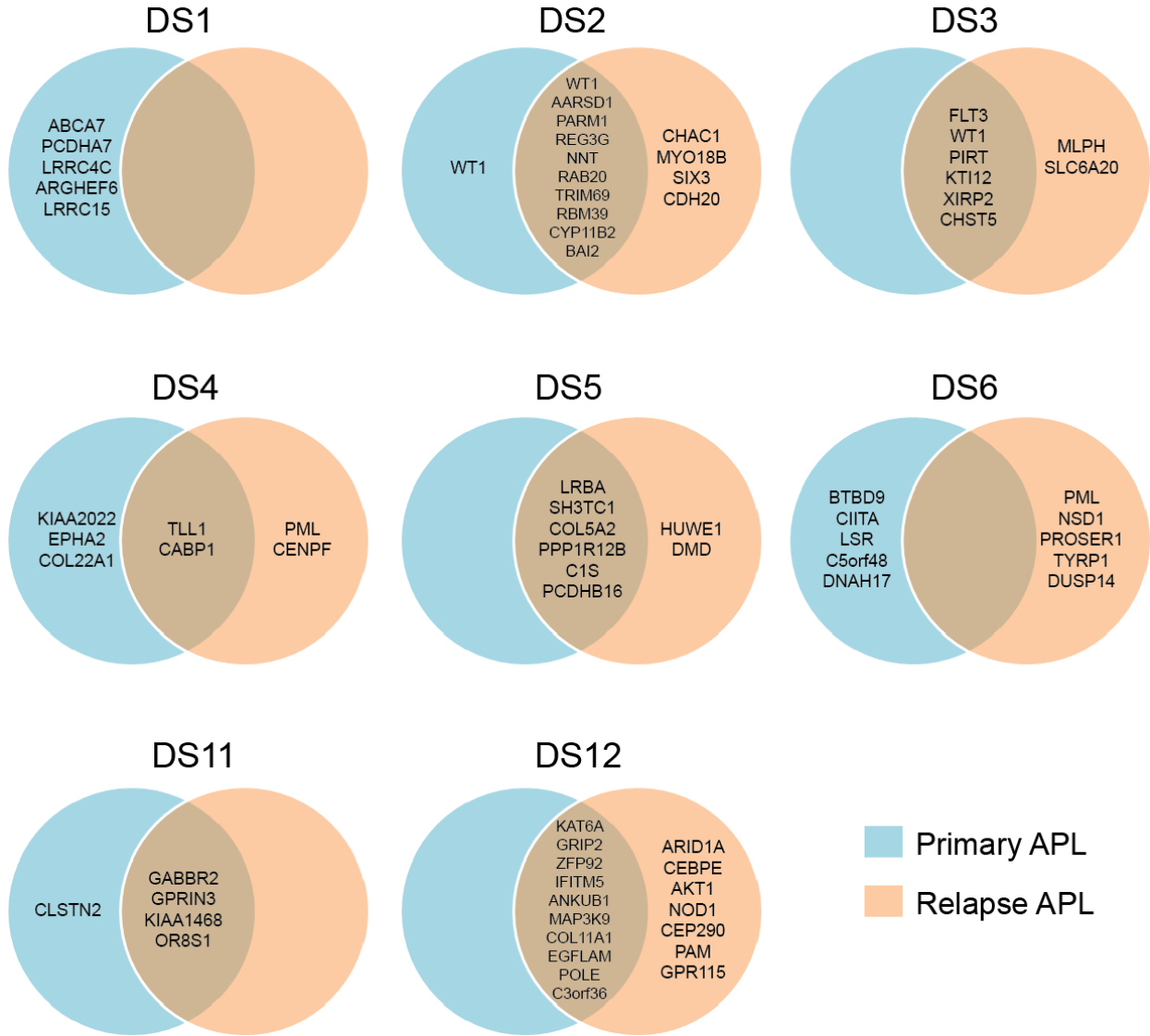
## Supplementary Figure S1



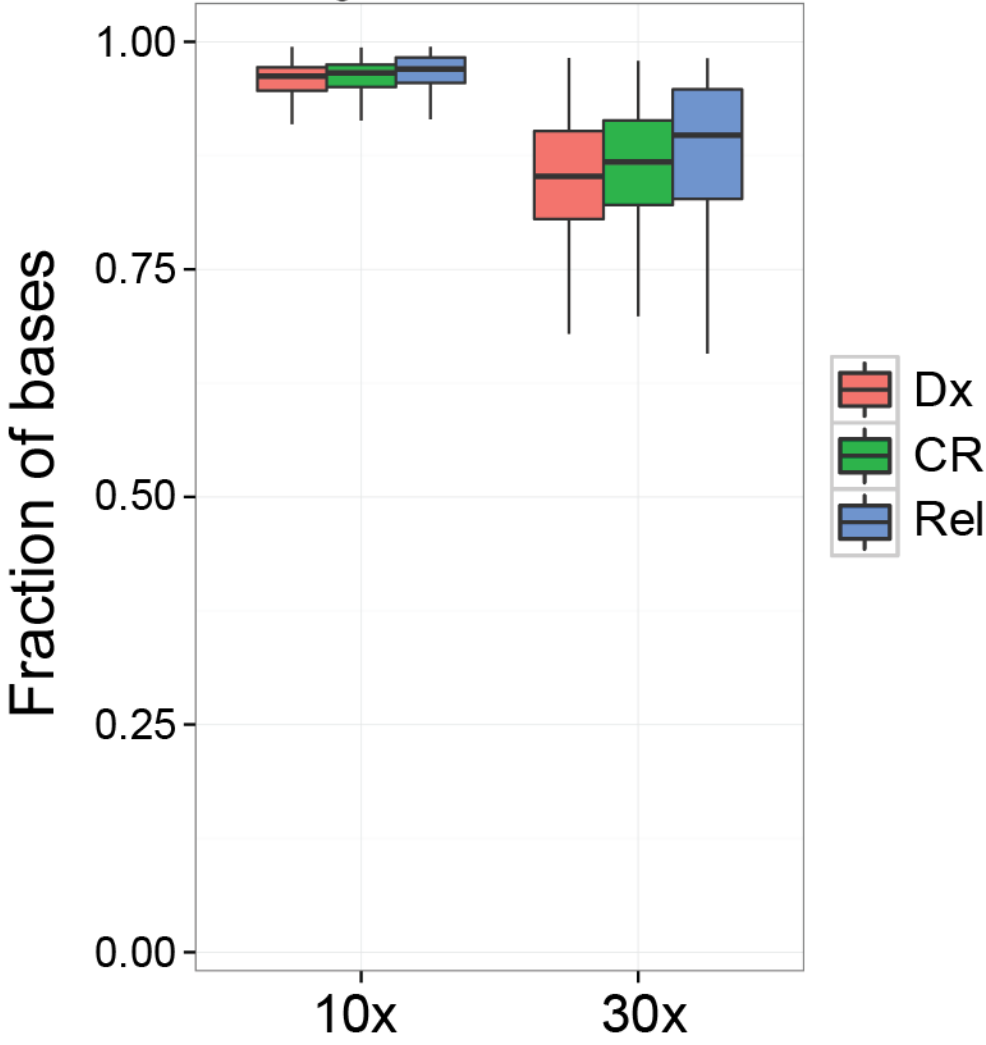
# Supplementary Figure S2



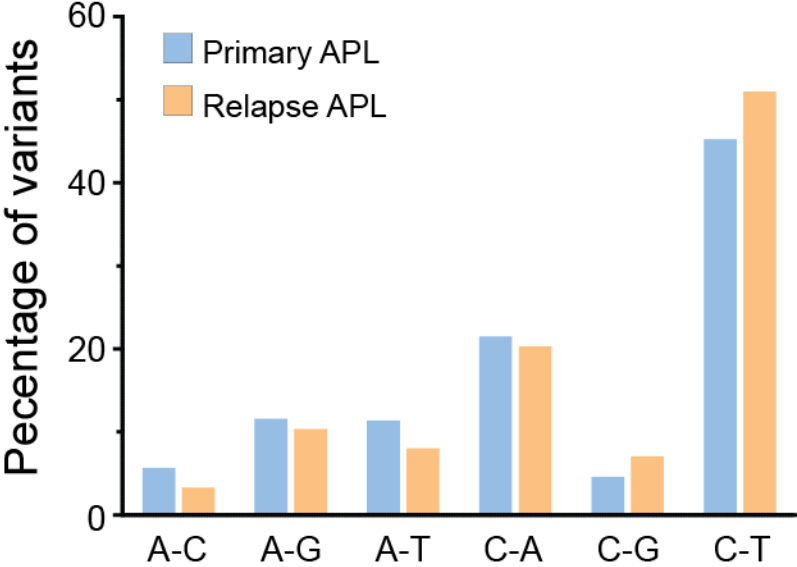
# Supplementary Figure S3



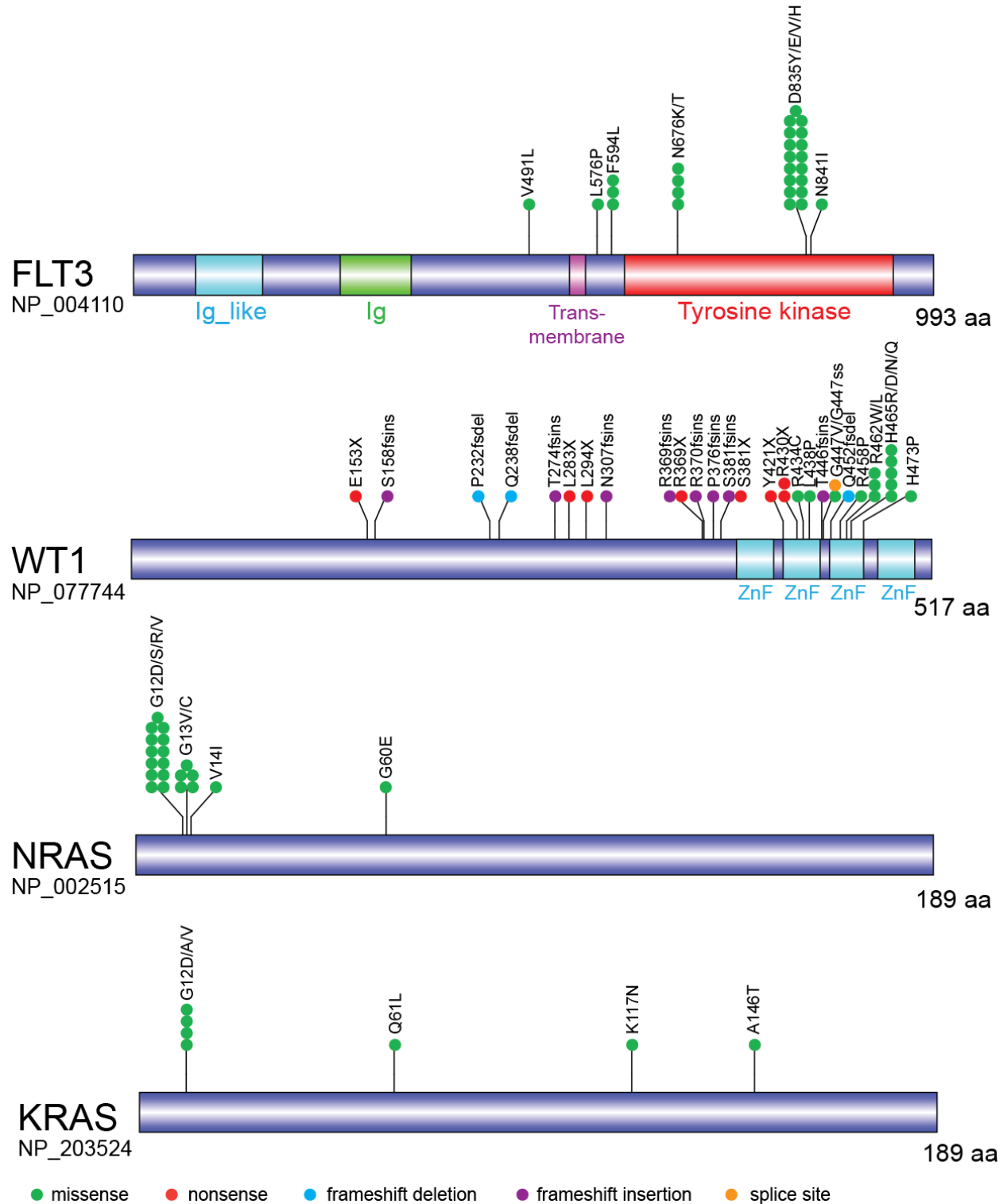
# Supplementary Figure S4



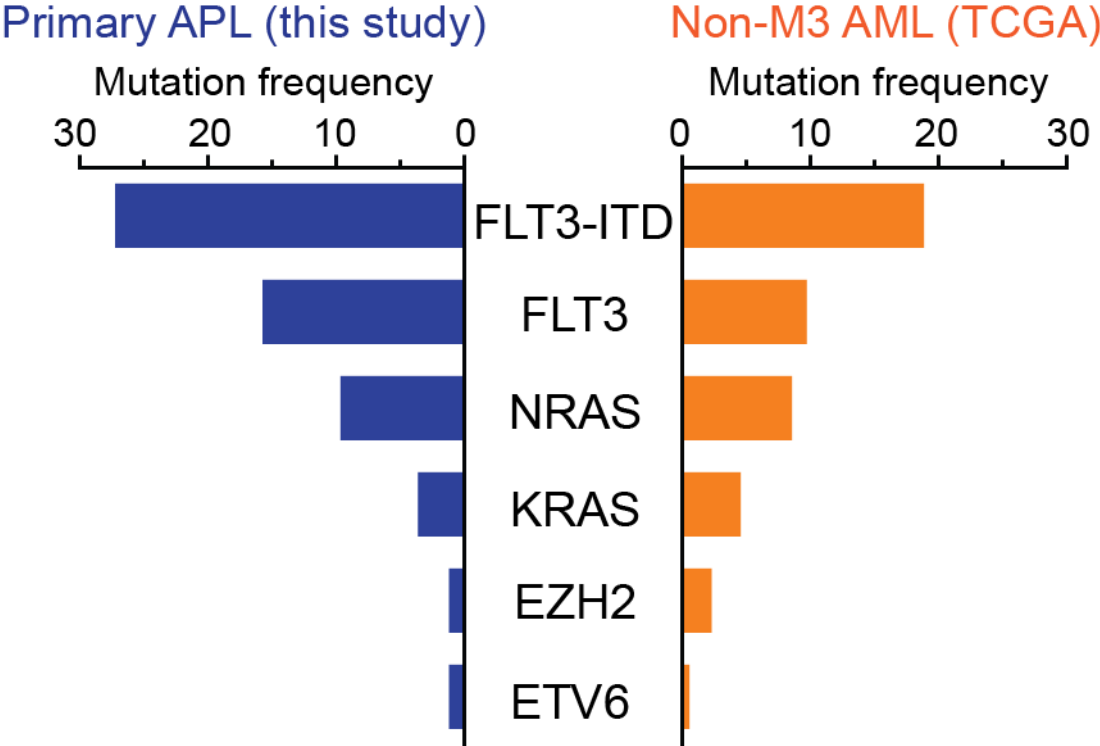
# Supplementary Figure S5



# Supplementary Figure S6

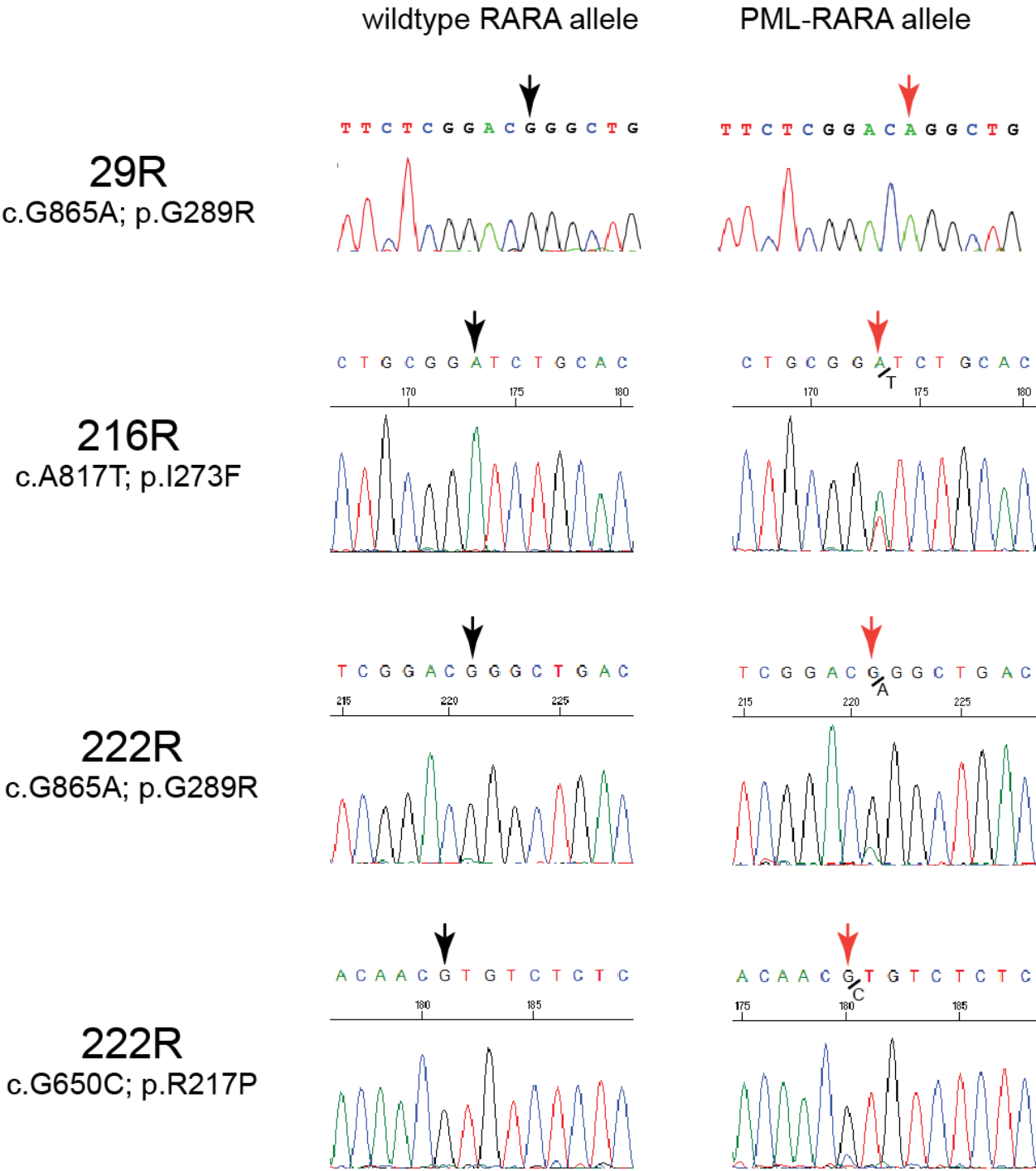


# Supplementary Figure S7

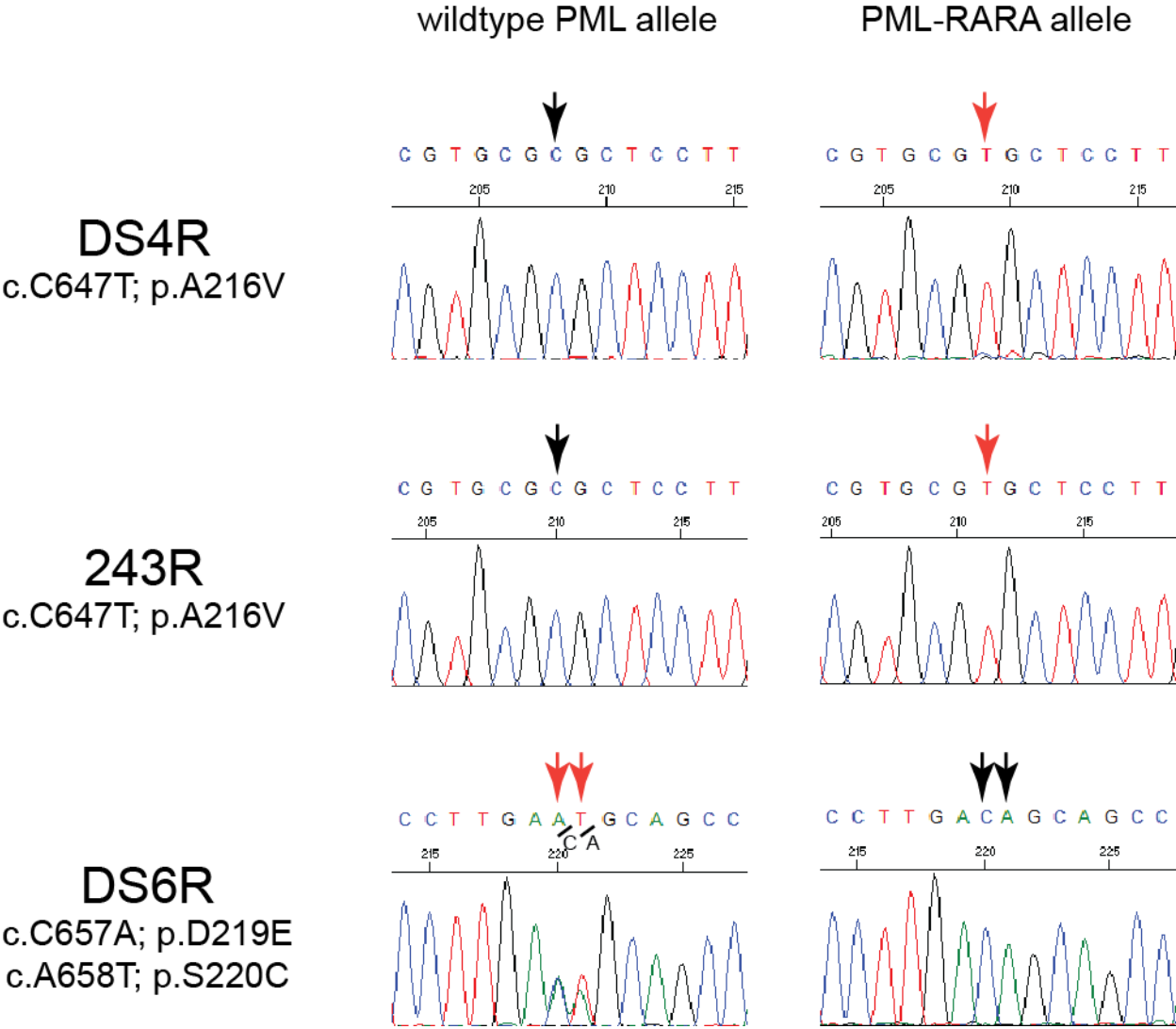




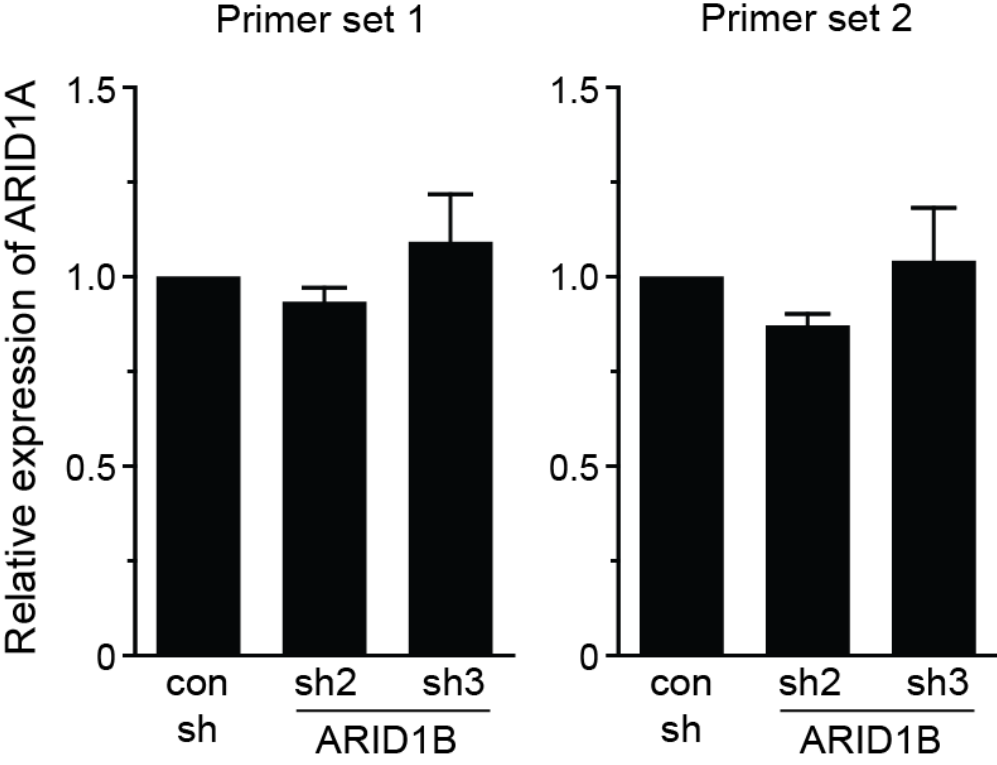
# Supplementary Figure S8



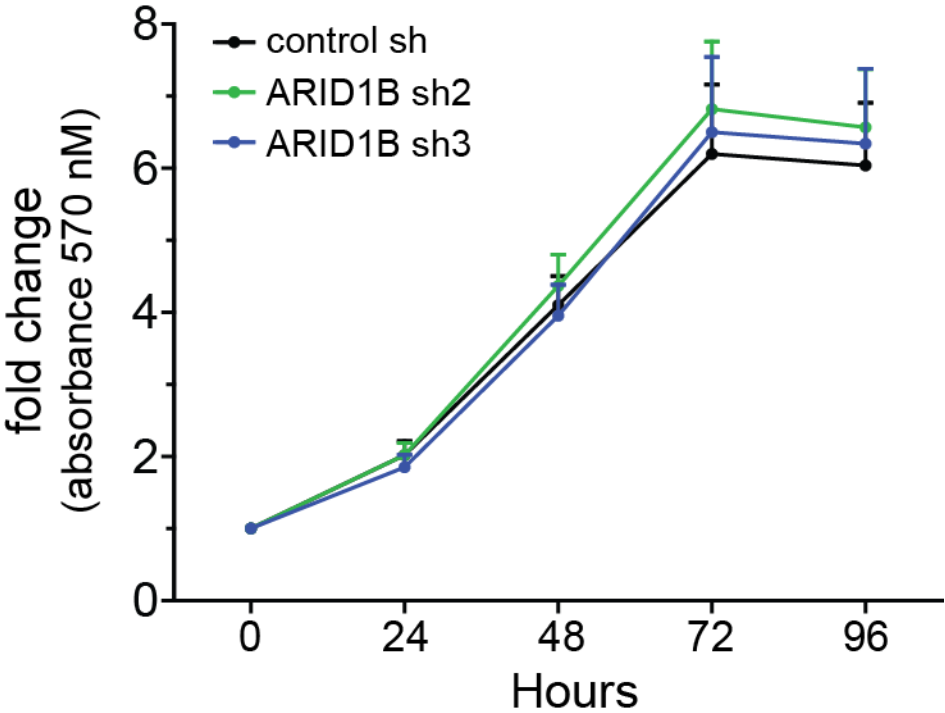
# Supplementary Figure S9



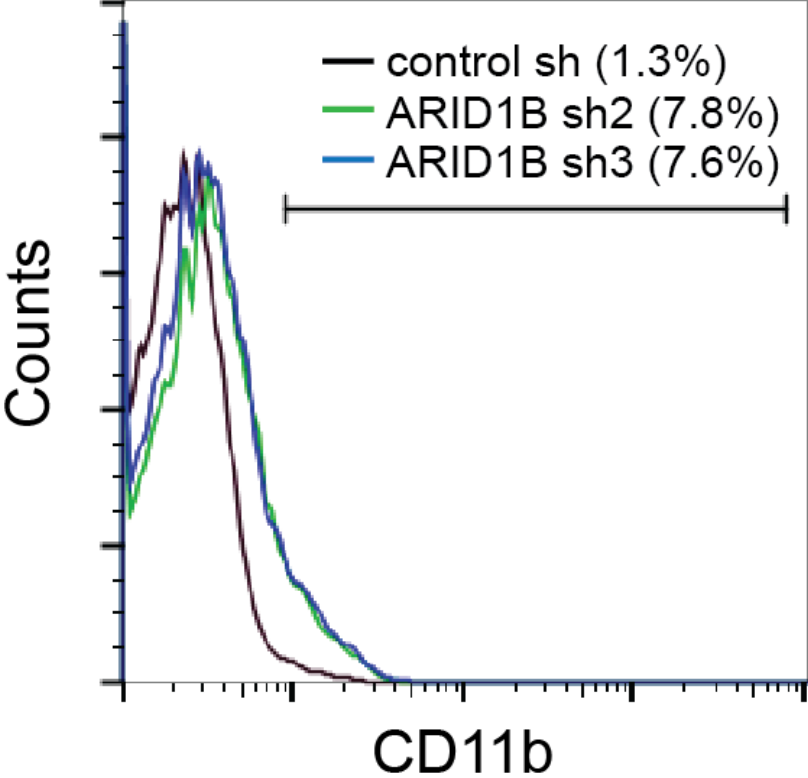
# Supplementary Figure S10



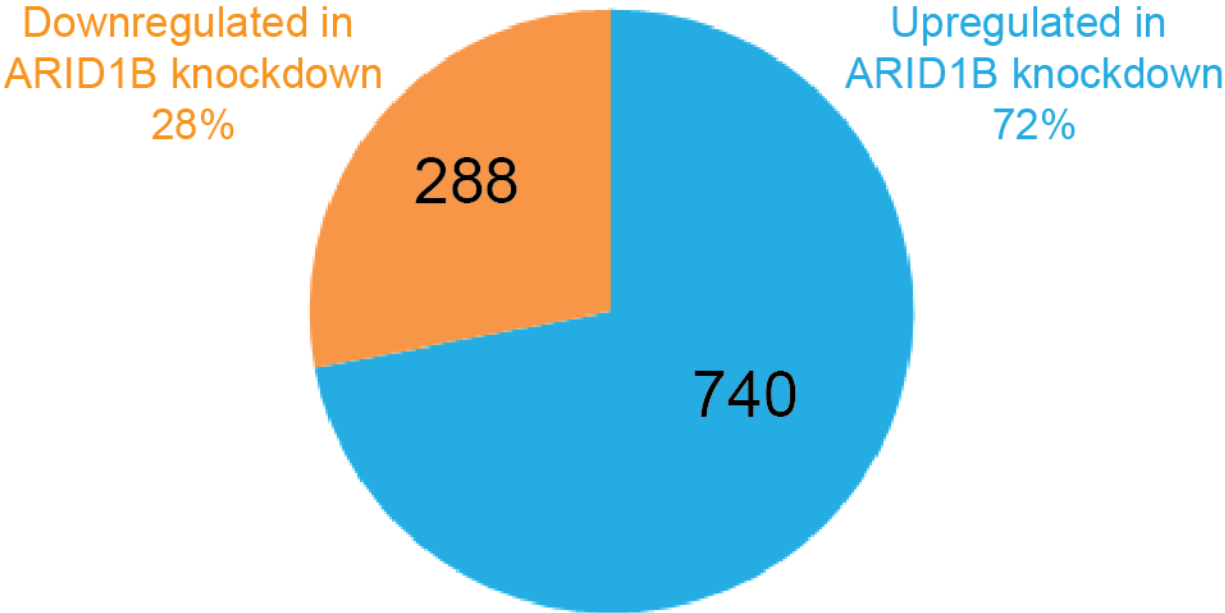
# Supplementary Figure S11



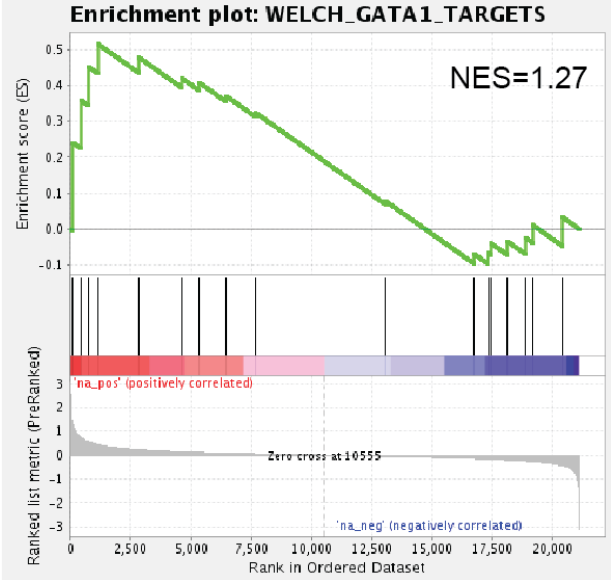
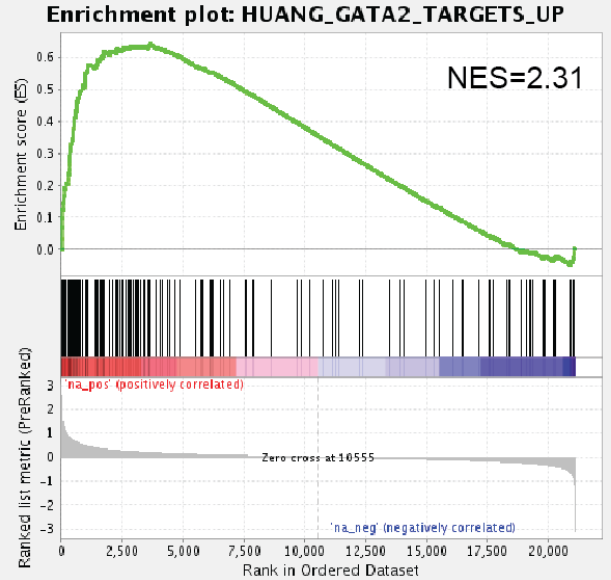
# Supplementary Figure S12



# Supplementary Figure S13



# Supplementary Figure S14



## **SUPPLEMENTARY TABLES**

(Separate excel sheets)

**Supplementary Table S1:** Clinical information of APL samples used for whole exome sequencing (Discovery cohort)

**Supplementary Table S2:** Non-silent somatic mutations validated in APL discovery cohort (exome sequencing)

**Supplementary Table S3:** Genes analyzed by targeted capture approach in APL frequency cohort

**Supplementary Table S4:** Details of APL samples used for targeted sequencing (frequency cohort)

**Supplementary Table S5:** Genes excluded from mutational analysis due to large size

**Supplementary Table S6:** Validated somatic mutations in targeted sequencing of APL (frequency cohort)