# Supplemental data

# Supplemental Methods

## Cell lines

NB4 cells were kindly provided by M. Lanotte (St. Louis Hospital, Paris, France). HEK-293T cells were obtained from the Cell Bank of the Chinese Academy of Science. NB4 cells were cultured in the RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS) (Gibco, Carlsbad, CA, USA) . HEK-293T cells were cultured in DMEM supplemented with 10% FBS. All cell lines were authenticated by STR sequencing. Mycoplasma contamination was routinely detected by the one-step Quickcolor Mycoplasma detection Kit (Cat. # MD001; Shanghai Yise Medical Technology, Shanghai, China,).

## Antibodies

The anti-H3K27ac (Cat. # 39133; Active Motif, Carlsbad, CA, USA) and anti-MYB (Cat. # ab45150; Abcam, Cambridge, UK) antibodies were used in chromatin immunoprecipitation (ChIP) assays. The anti-MYB (Cat. # sc-8558; Santa Cruz, CA, USA) and anti-β-actin (Cat. # 66009; Proteintech, Wuhan, China) antibodies were used in Western blot experiments.

## Whole genome sequencing

Whole genome sequencing (WGS) was performed on the paired tumor and normal samples from the same patient. High throughput sequencing was performed using the Illumina Hiseq X or NovaSeq 6000 platform. Raw reads were aligned to the reference human genome hg38 (Genome Reference Consortium GRCh38) using the BWA-MEM algorithm[1]. For datasets with multiple data lanes, the aligned files were merged using Picard MergeSamFiles (v25.0), and duplicate reads were marked with Picard MarkDuplicates (v25.0). In addition, the local realignment around indels and the base quality score recalibration functions were performed using GATK (v3.8). The quality of the alignments was assessed by metrics determined by Samtools stats (v1.6)[2]; the coverage and depth of alignments were assessed through each base depth calculated by Samtools depth (v1.6). The mean depth of all samples was 52×. Detailed statistics on alignment and coverage are given in supplemental Table 3.

## ChIP-seq library construction

APL blasts and NB4 cells were cross-linked with 1% formaldehyde (Cat. # F8775; Sigma, St Louis, MO, USA) for 10 min at room temperature, and the fixation reactions were quenched by adding glycine to a final concentration of 125 mM. Cells were washed twice in PBS, then incubated in lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% NP-40, 1 mM phenylmethylsulfonyl fluoride (PMSF, cat. # P7626, Sigma) and 1× protease inhibitor cocktail (Cat. # 11873580001; Roche, Mannheim, Germany) for 10 min. The pellets were re-suspended in the sonication buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 1% Triton X-100, 150 mM NaCl, 1 mM PMSF and 1× protease inhibitor cocktail). Chromatin was sonicated using the Biorupter Pico ultrasonicator device (Diagenode, Liege, Belgium) with 25 cycles of 30-second ON and 30-second OFF. Sheared chromatin was then incubated overnight at 4 °C with the indicated antibody. Protein G magnetic beads (Cat. # 10004D; Thermo Fisher Scientific, Norcross, GA, USA) were added to the immunoprecipitation reaction and incubated for another 2 hours at 4 °C. After finishing incubation, beads were washed for one time with low salt buffer (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 150 mM NaCl), high salt buffer (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 500 mM NaCl), LiCl buffer (250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0) and two washes with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The ChIP DNA was eluted using 100 μL of elution buffer (10 mM Tris, pH 8.0, 1 mM EDTA, pH 8.0, 1 % SDS,) at 55 °C for 10min. Cross-linking was reversed by adding 5 μL of proteinase K (Cat. # P8107; New England Biolabs, Ipswich, MA, USA) and incubating overnight at 65 °C. Input and ChIPed DNA were purified using the DNA purification kit (Cat. # 28106; QIAGEN, Valencia, CA, USA). ChIP-qPCR was conducted in triplicate using the primers listed in supplemental Table 17. ChIP-seq libraries were prepared with the MicroPlex Library Preparation Kit v2 (Cat. # C05010014; Diagenode) according to the manufacturer's instructions. High throughput sequencing was performed on the Illumina Hiseq X or NovaSeq 6000 platform.

## Collection of ChIP-seq data

ChIP-seq data for GFI1, IRF1, RUNX1, PML/RARα, KLF13, MEF2D, NFE2 and

ETV6 were collected from the Gene Expression Omnibus (GEO) database under the accession numbers: GSM935505, GSM2026066, SRP103029, ENCSR608HVP_2, ENCSR647ZXA_1, GSM2527371 and GSM2527376. ChIP-seq data for MYB in MOLT-3, Jurkat and K562 were collected from the GEO database under the following accession numbers: GSM1519643, GSM1519641 and GSM2825506. ChIP-seq for H3K4me1, H3K4me3, H3K27ac and H3K27me3 in K562 were collected from the following accession code: GSM1782706, GSM2534289, GSM2877120 and GSM608166. ChIP-seq for H3K4me1, H3K4me3, H3K27ac and H3K27me3 in Kasumi-1 were collected from the GEO database under the following accession numbers: GSM3165518, GSM1534445, GSM3165517, GSM1534446. H3K27ac ChIP-seq for hematopoietic cell lines were collected from the GEO database under the following access numbers, i.e., GSM2836487 (HL-60), GSM2108046 (THP-1), GSM2136946 (MV4-11), GSM2136938 (MOLM-13), GSM3094374 (EOL-1), GSM3436213 (SET-2), GSM2445788 (SKNO-1), GSM1003462 (DND-41), GSM1246865 (Jurkat), and GSM3425377 (Nalm-6).

**ChIP-seq data processing**

The ChIP-seq raw data obtained from sequencing reactions or the above collections were aligned using Bowtie2[3] version 2.3.5.1 against the human genome hg38. Picard MarkDuplicates version 25.0 was used to remove the duplicate reads. The aligned reads were further normalized to the same library size by MACS2[4] randsample with NUMBER = 1e-7, SEED = 614, and default settings.

**Annotations of CREs in APL patients**

The H3K27ac ChIP-seq data from APL patients was used to define the CREs. The BED files of H3K27ac-positive peaks from patients, called using MACS2 as described above, were used to perform sample saturation analysis. Saturation analysis was based on the permutation process, randomly selecting the number of desired samples in 1,000 iterations to combine and calculate the number of peaks. The permutation process was executed as a loop from 1 to the number of samples. We could evaluate the saturation of H3K27ac-positive regions based on a non-linear regression model fitted on the results of saturation analysis. Then, the H3K27ac-positive peaks of all patients were merged by Bedtools[5] v2.27.1 with the default setting. Considering that CREs are

usually flanked by H3K27ac positive regions[6,7], we further obtained the APL-associated CREs by merging the H3K27ac-positive peaks through extending by ± 500 bp (the average length of transcription factor (TF)-bound regions). The HOMER[8] peak annotate tool (annotatePeaks.pl) version 4.9.1 was used to annotate the relative genomic distribution of CREs, and the nearest neighbor gene of a given CRE was assigned as the CRE-associated gene.

To compare mutated CREs in APL versus those in other cancer types, we used WGS and H3K27ac ChIP-seq data to establish the landscapes of mutated CREs in other types of hematopoietic malignancies and solid cancers. Hematopoietic malignancies included chronic lymphocytic leukemia (CLLE) and malignant lymphoma (MALY). Solid cancers included bone cancer (BOCA), breast cancer (BRCA), liver cancer (LIRI), pancreatic cancer (PACA), pediatric brain cancer (PBCA), and prostate adenocarcinoma (PRAD). The variant call format files of WGS data for the above cancer types were downloaded from the Pan-Cancer Analysis of Whole Genomes project (PCAWG)[9]. The H3K27ac ChIP-seq data of these cancer types were downloaded from the Gene Expression Omnibus (GEO) database. Briefly, we constructed somatic mutation profiles for each cancer type by combining variant call data of multiple samples. We then established the CRE profiles for each cancer type using H3K27ac ChIP-seq data for the corresponding cell line. Then, we identified the mutated CREs for each cancer type by integrating the mutation profiles data and the CRE profiles (Detailed statistics in supplemental Table 7). We performed Gene Ontology (GO) enrichment analysis to identify the enriched pathways within mutated CRE-regualted genes in APL and other cancer types.

**Identification of master TFs based on CRC analysis**

To construct the CRC model in APL, we first used the Rank Ordering of Super Enhancers (ROSE) algorithm[10] to define the supper enhancers (SEs) of 16 APL samples based on the above described H3K27ac ChIP-seq data. The stitching distance was fixed at 12.5 kb to facilitate comparisons between samples. For other parameters, the default settings were used. Genes annotated by the ROSE2 ENHANCER_TO_TOP_GENE.txt file were used for defining the target genes of SEs for subsequent analyses. Then, we applied the CRC mapper algorithm[11] to construct the CRC model in APL based on the

super-enhancer profiles of these 16 APL patients. A total of 25 transcription factors were identified using the 16 CRC models. The transcription factors included in more than 25% of patients were considered as master transcription factors of APL. In the CRC analysis, the refGene.txt of hg38 downloaded from UCSC was used to annotate the genome files, and the motifs of transcription factors scanned by FIMO were obtained from the JASPAR database[12]. The previously reported RARE-half site bound by PML/RARα[13] was included in the scanning process.

**Enrichment of binding regions for master transcription factors within mutated CREs**

We collected the genomic binding regions of the master TFs in the hematopoietic cell lines. The genomic binding regions of MYB, IRF1, PU.1 and RARα were obtained by analyzing their ChIP-seq data in NB4. The genomic binding regions of other TFs were obtained from the Cistrome database[14]. We used bedtools intersect v2.27.1 to calculate the overlap of the TF binding regions with mutated or unmutated CREs. The Fisher's exact test was used to evaluate the significance.

**Motif analysis**

To investigate whether mutated CREs were directly targeted by the identified transcription factors, we used the HOMER[8] motif discovery tool (findMotifsGenome.pl) version 4.9.1 to preform the motif enrichment analysis in the mutated CREs versus non-mutated CREs. We downloaded the position weight matrices of these transcription factors and their paralogs from the JASPAR database. The RARE-half motif for PML/RARα was included in the motif enrichment analysis. Then, we calculated the enrichment of the above motifs within the mutated CREs (as the target regions) versus the non-mutated CREs (as the background regions). The number of the background regions was set to twice the target regions by randomly selecting from non-mutated CREs. This process was iterated 200 times to get the final result.

To define the MYB motif in APL, we used the findMotifsGenome.pl program to perform the motif enrichment analysis based on MYB-specific ChIP-seq data in NB4 cells. The top 1000 intensity peaks were selected as the target regions. The most significant motif was defined as the specific binding motif of MYB in APL.

## Mutation enrichment analysis whitin the motifs and flanking regions

To assess the significance of the mutation enrichment within the motifs and flanking regions, we referred to the previously published method design[7] for mutation enrichment analysis and realized it through self-developed code. First, bedtools v2.27.1 was used to extract the overlap between the ChIP-seq peaks of the identified transcription factors in hematopoietic cell lines and the profile of APL CREs as previously established. We obtained the potential binding regions of each TF in APL as the regions of interest for subsequent analysis. Second, we scanned the motif positions in the regions of interest as the binding sites of the corresponding TFs. MOODS[15] v1.9.4 was used to match the position frequency matrices of motifs from the database JASPAR against DNA sequences in the regions of interest. We set the parameter *P* value to 0.0001. Third, we calculated the mutation frequency at the motif positions of the transcription factor combined with the APL somatic mutation profile as established above and further determined the mutation load by expanding the motif positions according to the specified base number, including 10bp, 20bp, 30bp, 40bp, 50bp, 100bp, 200bp, 300bp, 400bp, 500bp. Fourth, the permutation test was performed to calculate the mutation frequency within randomly selected positions from the APL H3K27ac-positive regions. The randomly selected positions were consistent in number and length with the binding sites of the corresponding TFs in the regions of interest. This process was performed in 5000 iterations for each transcription factor. Finally, we transferred the mutation frequency within the motifs and around regions to the *Z* score according to the permutation test to evaluate the significance of the mutation enrichment.

## RNA extraction, reverse transcription and real time PCR

Total RNA was extracted using the RNeasy mini kit (Cat. # 74106; QIAGEN) according to the manufacturer's instructions. RNA was reverse transcribed into cDNA using the PrimeScript™ RT reagent Kit with gDNA Eraser (Perfect Real Time) (Cat. # RR047B; Takara, Osaka, Japan). RT-qPCR was conducted using the SYBR Green Premix pro Taq HS qPCR Kit (Rox Plus) (Cat. # AG11719; Accurate Biology, Changsha, China) on the ABI ViiA 7 Real-Time PCR System. The relative expression level of each gene was calculated as $2^{-\triangle\triangle Ct}$. Primers for real time-qPCR are listed in supplemental Table 18.

## RNA sequencing and data processing

RNA-seq libraries were constructed according to the manufacture's instruction using the TruSeq RNA Sample Preparation Kit v2 (Cat. # RS-122-2001 or RS-122-2002; Illumina, Hayward, CA, USA). The purified library was quantified using Qubit 4 Fluorometer (Invitrogen, Carlsbad, CA, USA), and size distribution was analyzed by Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA). High throughput sequencing was performed on the Illumina HiSeq 2500 or NovaSeq 6000 platform. The software Hisat2[16] was used to align the reads from RNAseq to the reference human genome hg38 (Genome Reference Consortium GRCh38). The gene counts of each sample were generated using the HTseq-count[17] and were normalized to transcripts per million (TPM).

## Identification of the functional regions harboring somatic non-coding mutations within mutated CREs

To identify the functional regions harboring somatic non-coding mutations, we performed a comprehensive analysis based on the recurrence occurring within a small region, the transcriptional influences, and the disease relevance. First, we identified the recurrently mutated loci by clustering all somatic non-coding mutations within 50 bp. Second, we assigned the target gene to each recurrently mutated locus within CREs according to the principle of proximity, and then used Phenolyzer[18] to assess the priority of those target genes related to leukemia. Third, we evaluated the transcriptional effects of the recurrently mutated loci within CREs by comparing the expression levels of the respective target gene in mutated and non-mutated samples. Then, each recurrently mutated locus within CREs was placed in a 3D feature space taking into account mutation recurrence, leukemia relevance, and gene expression. Finally, we calculated the Euclidean distances based on the three features described above and obtained the functional regions harboring somatic non-coding mutations within mutated CREs.

## PCR and Sanger sequencing

Non-coding *WT1* somatic or germline variants were validated by PCR amplification and Sanger sequencing in tumor and paired normal samples. A somatic mutation refers to an alteration that occurs in the tumor sample but not in the paired normal sample. A

germline variant refers to an alteration that exists in both the tumor and the paired normal sample. The PCR primers were designed using the Primer3 software (supplemental Table 19). The PCR products were sequenced on the ABI 3730XL DNA sequencer (Applied Biosystems).

**Luciferase reporter assay**

The *WT1* promoter and enhancer region with strong H3K27ac signals were amplified using genomic DNA from APL patients with or without non-coding *WT1* variants. The primers used were described in supplemental Table 20. The amplified regions were cloned into the pGL3-basic vector. The luciferase constructs, in combination with the pRL-SV40 renilla plasmid and MYB overexpressing plasmids (pENTER-MYB (Cat. # CH806231; WZ Biosciences Inc., China) and empty vector) or knockdown siRNAs (si-MYB and si-NC), were delivered into HEK-293T or NB4 cells. After transfection for 24 hours, both firefly luciferase activity and renilla luciferase activity were detected with the GloMax 20/20 Luminometer (Promega) using the Dual-Luciferase Reporter Assay System (Cat. # E1910; Promega, Madison, WI, USA) following the manufacturer's instructions. The firefly luciferase activity was normalized with the renilla luciferase activity to control the transfection efficiency.

**DNA pulldown assay**

The 5'-biotinylated double-stranded DNA probes (supplemental Table 21) were incubated with the whole cell lysate of HEK-293T cells overexpressing MYB overnight at 4 °C. The magnetic streptavidin beads were then added into the complexes of reactions, followed by rotation at 4 °C for another 4 hours. Beads were washed three times with cold lysis buffer (20 mM Tris-HCl pH 8.0; 2 mM EDTA; 1% Triton X100; 150 mM NaCl) and resuspended in SDS loading buffer for western blot analysis with anti-MYB antibody.

**CRISPR/Cas9-mediated editing and knockout**

The Cas9-expressing NB4 cells were generated by lentiviral transduction of LentiCRISPR v2 GFP (Addgene plasmid #82416), followed by flow cytometry sorting of GFP-positive cells. The sorted cells were then sub-cloned and selected for the best CRISPR/Cas9 efficiency clone using a lentiviral reporter pKLV-sgGFP.

The sgRNAs oligonucleotides containing the BbSI restriction sites (supplemental Table 22) were constructed into the lentiviral vector pKLV-U6gRNA(BbsI)-PGKpuro2ABFP (Addgene plasmid #50946). Supernatants containing lentivirus were generated by co-transfecting 16 μg of the pKLV construct, 9 μg of the psPAX2 package plasmid and 6 μg of the pMD2G envelop plasmid into HEK-293T cells using the HilyMax reagent (Cat. # H357-10; Dojindo, Tokyo, Japan) and were harvested at 48 or 72 hours after transfection. Cas9-expressing NB4 cells were infected using the lentivirus with polybrene at 8 μg/mL (Cat. # H9268; Sigma) by centrifuging directly at 1200 × g for 90 min at 37 °C. For cells transduced with sgRNAs targeting the MYB motif or control sgRNAs targeting regions surrounding the MYB motif within the intron 3, single cells were sorted into 96 cell plates. Once single cells had grown into colonies, genomic DNA was extracted and analyzed for mutations by Sanger sequencing.

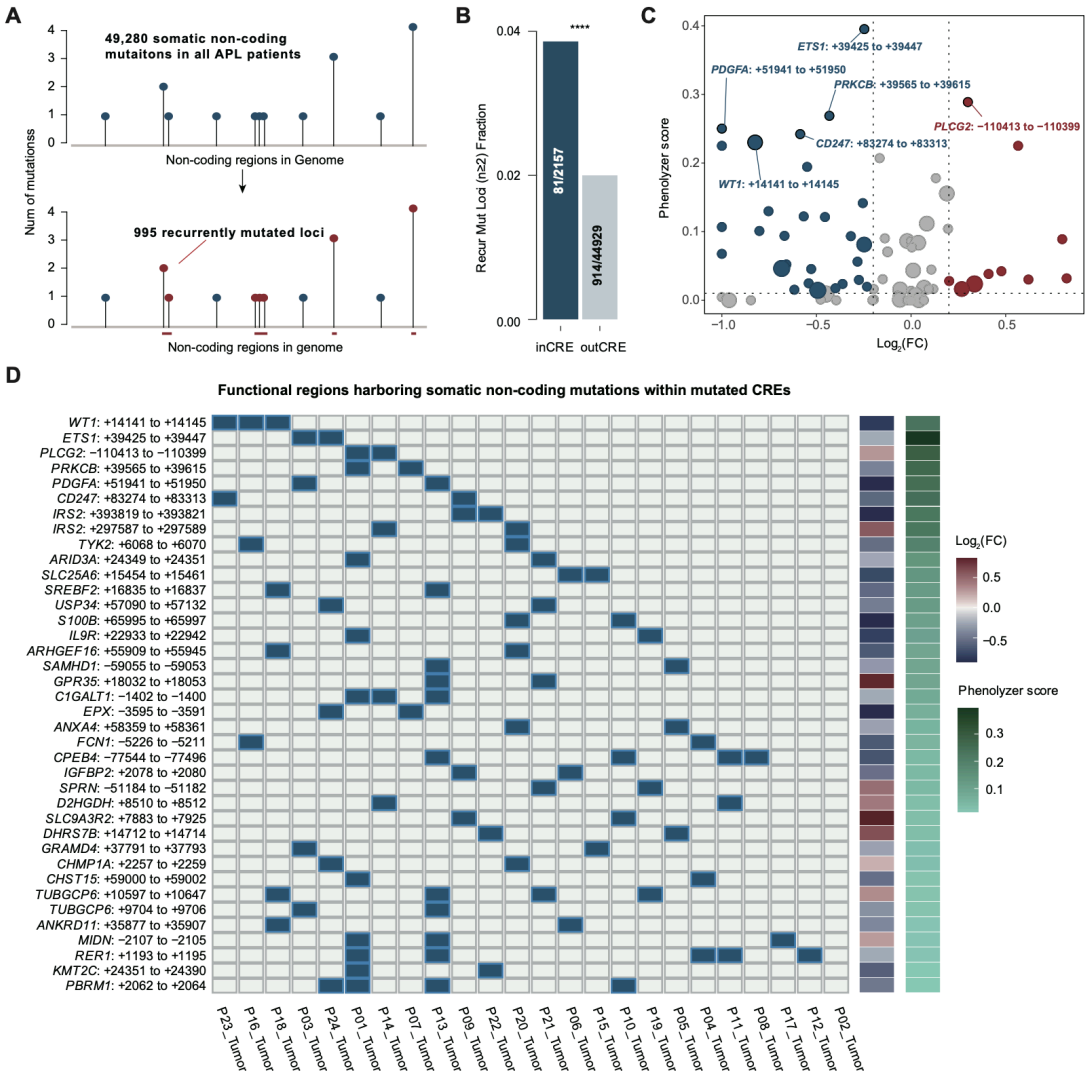
**Chromatin conformation capture (3C) experiment**

The 3C-qPCR experiments were performed as previously described[19] with some modifications. Briefly, cells were cross-linked with 1% formaldehyde and were quenched with 125 mM glycine. Cells were then lysed and resuspended in 0.3% SDS in 1 × NEBuffer 2.1 (10 μL of 10 × NEBuffer 2, 3 μL of 10% SDS, 87 μL of $H_2O$) with shaking at 37 °C for 30 min. After adding Triton X-100 to a final concentration of 1% (v/v), the genomic DNA was digested overnight with XbaI (Cat. # R0145; New England Biolabs). DNA ligation was performed with the T4 DNA ligase (Cat. # M0202; New England Biolabs) at RT for 6 hours. The crosslinks were then reversed by overnight incubation at 65 °C with proteinase K. The digestion and ligation efficiencies were assessed using gel electrophoresis before proceeding to DNA purification by the phenol-chloroform extraction.

The purified 3C DNA was used to perform 3C-qPCR nearby the XbaI recognition sites among the regulatory regions of the *WT1* locus. The positive control sample was generated by amplifying, digesting and ligating fragments containing each of these seven XbaI sites with primers nearby the restriction sites, which was sequentially diluted to correct the PCR efficiency. The fragment within two adjacent XbaI sites in the *WT1* locus served as the loading control. The interaction frequency was calculated as the ratio relative to the region A in the vector. All the primers were designed using

Primer3 (supplemental Table 23). qPCR was performed using SsoFast EvaGreen Supermix (Cat. #172-5211; Bio-Rad, Hercules, CA, USA).

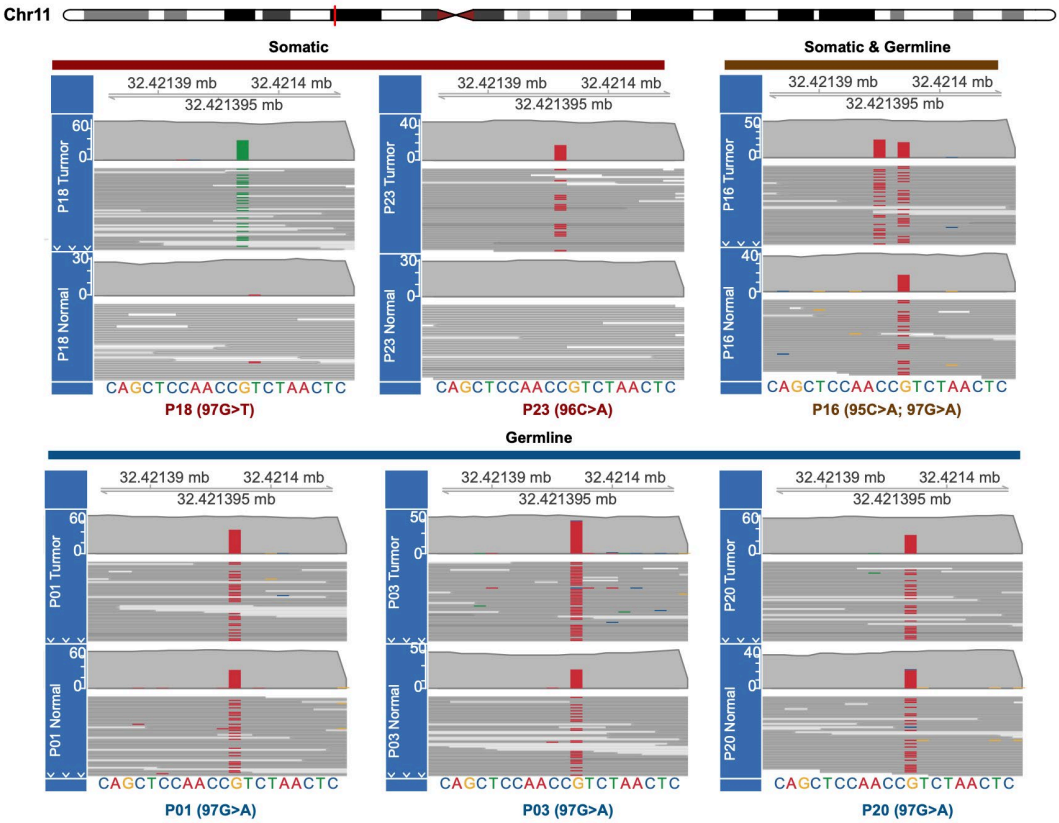
# Supplemental Figures




**Supplemental Figure 1. Identification of the functional regions harboring somatic non-coding mutations within mutated CREs.**

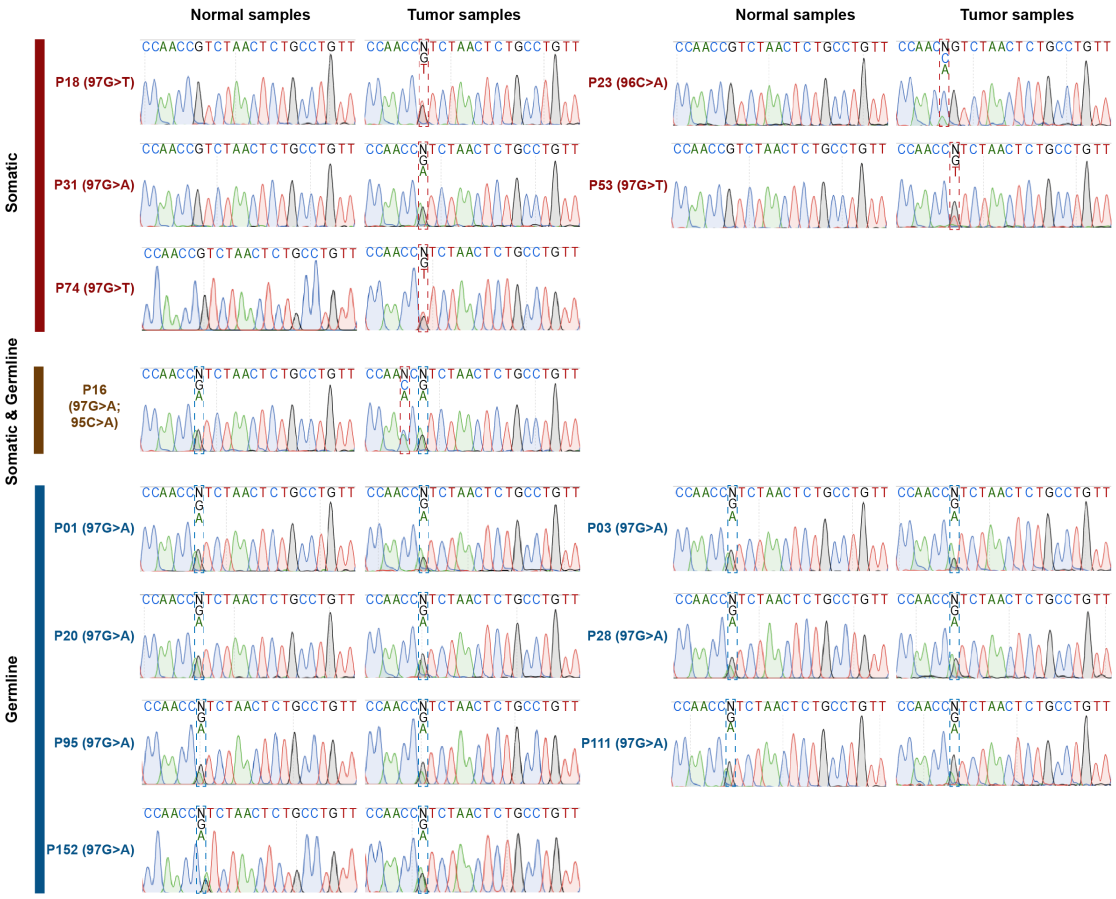
**(A)** Diagram of the clustering of somatic non-coding mutations within 50 bp to identify the recurrently mutated loci. A total of 995 recurrently mutated loci were found from 49,280 somatic non-coding mutations. **(B)** Distribution of the recurrently mutated loci according to whether they are located within CREs. The bar plot shows the distribution of the recurrently mutated loci within CREs. We identified a total of 81 recurrently mutated loci, which were significantly enriched within CREs relative to other non-coding regions. The statistical significance was calculated by the Fisher's exact test. ****$P < 0.0001$. **(C)** Presentation of each recurrently mutated locus within CREs in a

3D feature space, considering the mutation recurrence, leukemia relevance, and gene expression. The scatter plot shows the fold change and Phenolyzer score of 81 recurrently mutated loci within CREs. The size of the point represents the number of the mutation recurrences. The blue and red points represent the candidate functional mutated loci that inhibit and promote the expression of the related target gene, respectively. **(D)** Displaying identified functional regions harboring somatic non-coding mutations within mutated CREs. The abscissa represents each patient, the ordinate represents the functional regions harboring somatic non-coding mutations for the indicated gene, the blue represents the sample with the specified mutation, and the gray represents the sample that does not contain the specified mutation. Also included is the log2 fold change of associated genes by comparing the expression levels of respective target genes in mutated and non-mutated samples. Different shades of green squares represent the Phenolyzer score.

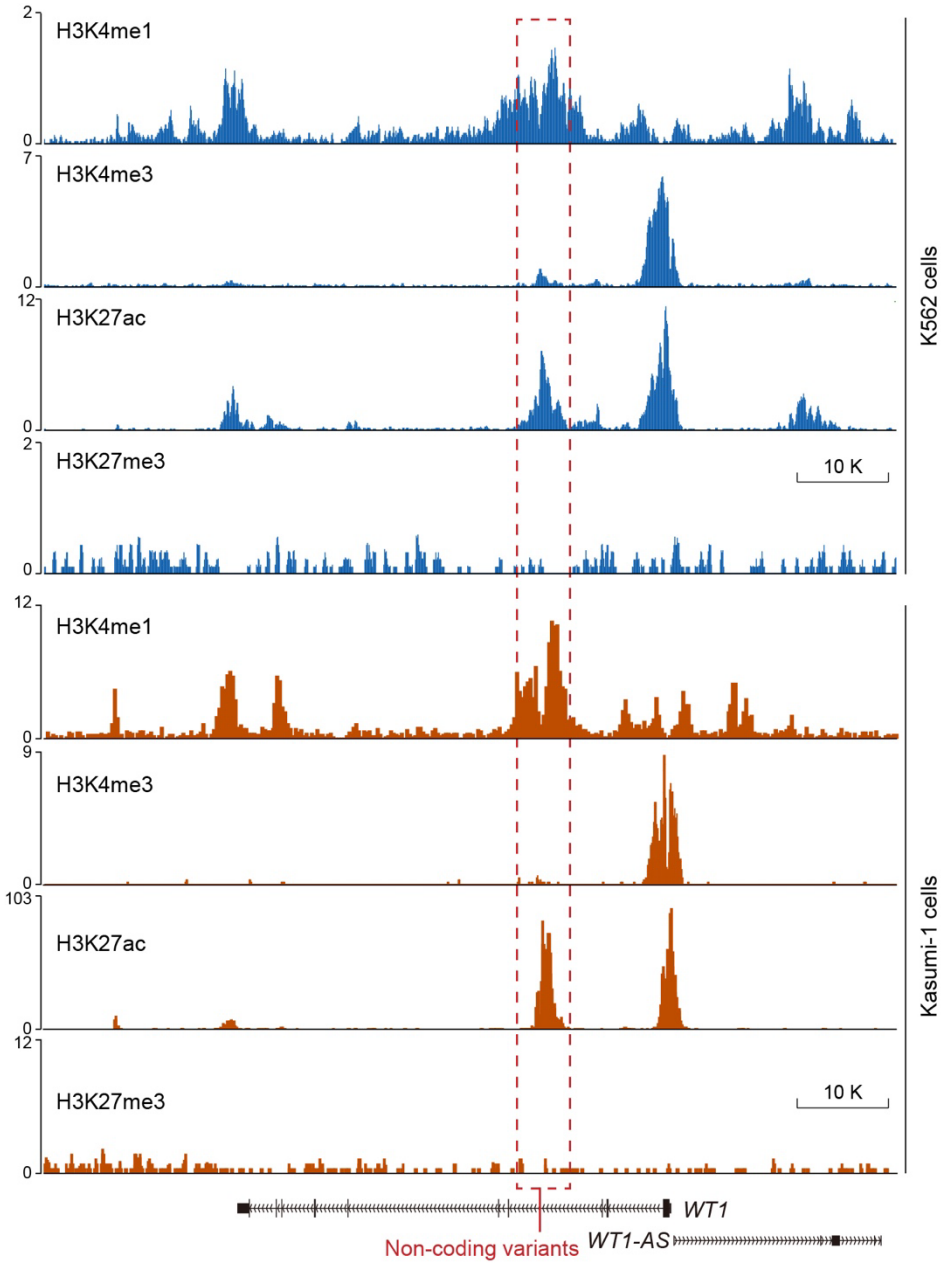



**Supplemental Figure 2. Visualization of the alignment pileup near the non-coding *WT1* somatic mutations and/or germline variants in the tumor sample and the paired normal sample from the same patient.**

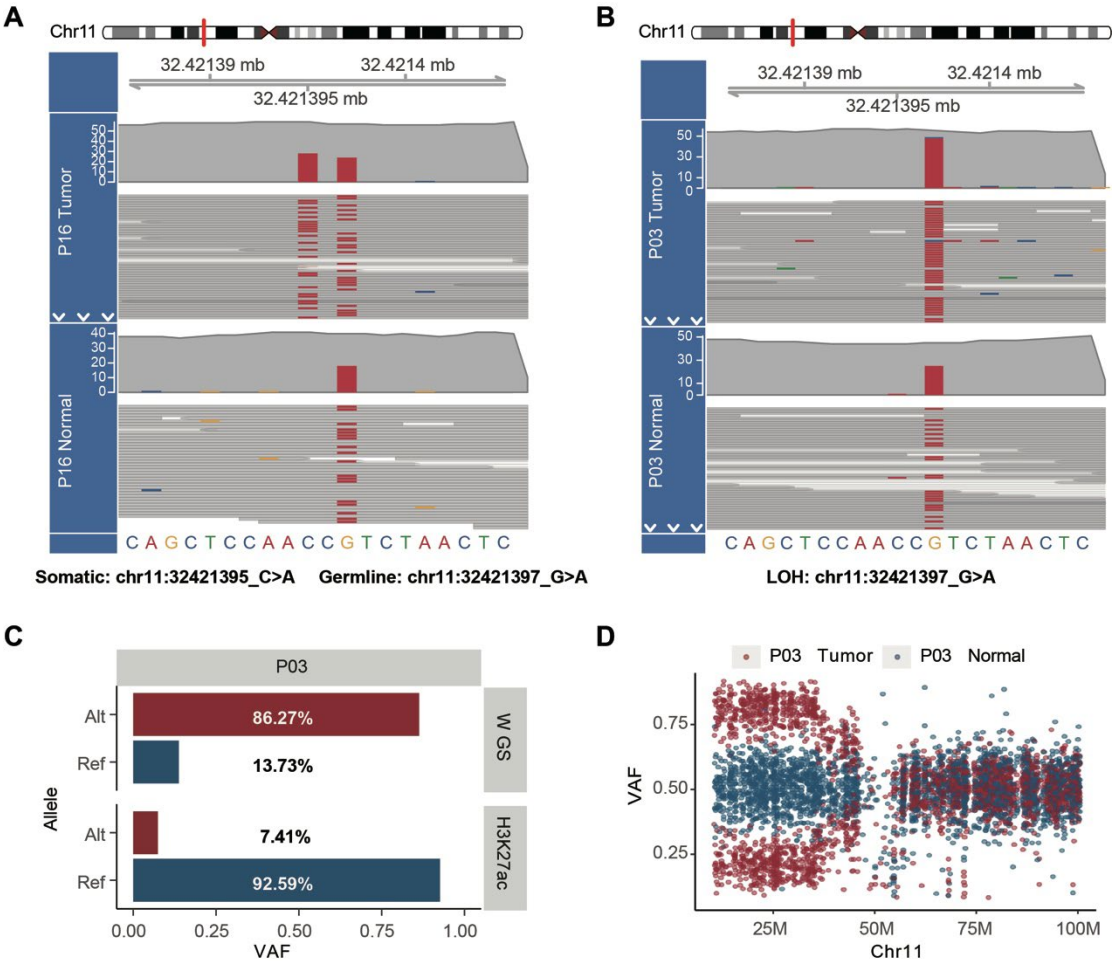
P18 and P23 contained somatic *WT1* non-coding mutations. P16 harbored a somatic *WT1* non-coding mutation and a germline *WT1* non-coding variant on two alleles, respectively. P01, P03, and P20 contained germline *WT1* non-coding variants. A somatic mutation refers to an alteration that occurs in the tumor sample but not in the paired normal sample. A germline variant refers to an alteration that exists in both the tumor and the paired normal sample.




**Supplemental Figure 3.** The Sanger sequencing chromatograms of the third intron of *WT1* in tumor and matched normal samples from APL patients with non-coding *WT1* variants. Somatic and germline variants were identified by comparing the DNA sequences of tumor to normal samples.

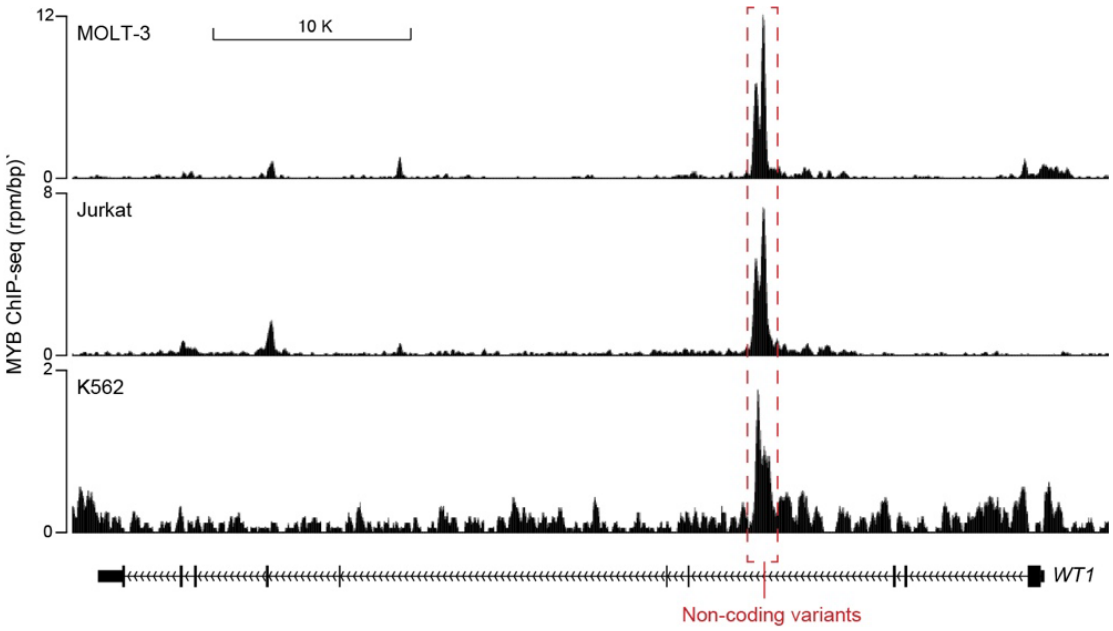



**Supplemental Figure 4.** ChIP-seq tracks of H3K4me1, H3K4me3, H3K27ac and H3K27me3 occupancy at the *WT1* locus in K562 and Kasumi-1 cells without non-coding *WT1* variants. The recurrently mutated site on the third intron of *WT1* is marked with a red line.

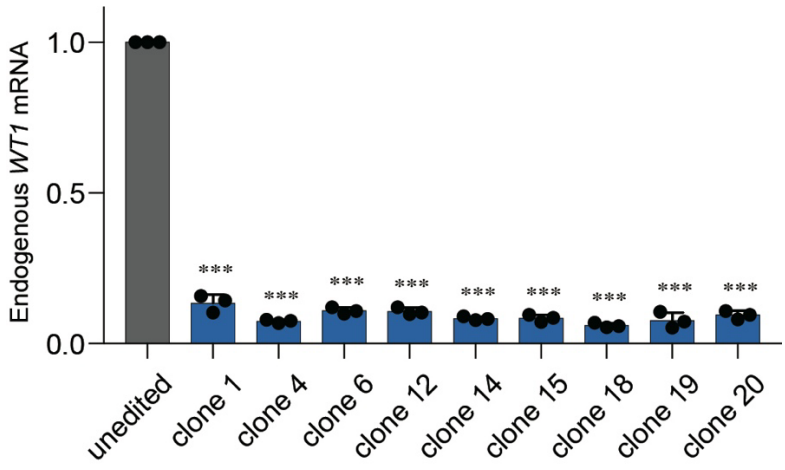



**Supplemental Figure 5. Two APL patients homozygous for the non-coding *WT1* variants.**

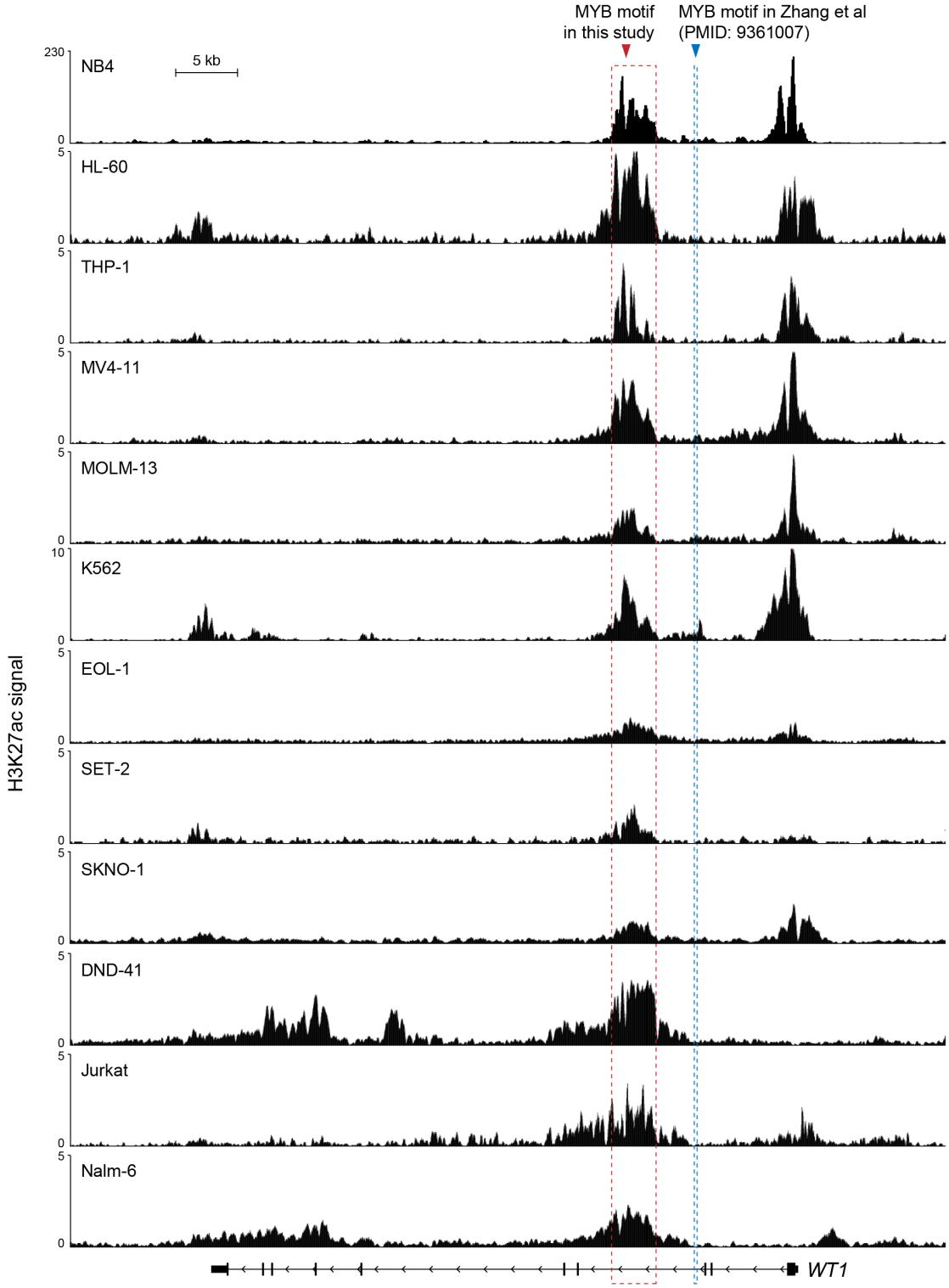
**(A-B)** Visualizing the alignment of WGS reads for non-coding *WT1* variants. **(A)** The alignment track shows that P16 harbors a somatic mutation and a germline variant on two alleles. **(B)** The alignment track shows that P03 contains the same germline variant in a homozygous state due to copy-neutral LOH. **(C)** The variant allele frequency (VAF) of the non-coding *WT1* variants in P03 based on the WGS data confirmed the homozygous status. Still, the variant allele frequency shows extremely low levels in H3K17ac ChIP-seq data, indicating the inhibitory effect of the non-coding *WT1* variant on H3K27ac binding. **(D)** Tumor-specific loss of heterozygosity (LOH) for chromosome 11p in P03 where the non-coding *WT1* variant is located. LOH was assessed by examining variant allele frequencies (VAFs) for variants found to be heterozygous in the normal sample. The red and blue points represent the VAFs of single nucleotide polymorphisms (SNPs) in the tumor and the paired normal samples.




**Supplemental Figure 6.** ChIP-seq tracks showing MYB binding to the third intron of the *WT1* gene in hematopoietic cell lines without non-coding *WT1* variants, i.e., MOLT-3, Jurkat and K562. The recurrently mutated site on the third intron of *WT1* is marked with a red line.




**Supplemental Figure 7.** The *WT1* mRNA levels in control and the MYB motif-mutated clones. Data are represented as the fold change relative to the expression of the control cells. ***$P < 0.001$.




**Supplemental Figure 8. ChIP signals for H3K27ac at the WT1 locus in hematopoietic cells.** Our identified enhancer and a previously reported MYB motif-containing region[20] are shown by the red and blue dashed-line squares, respectively. The MYB motif is indicated by the red arrow.

## Supplemental References


1. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595.
2. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
3. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.
4. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
5. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842.
6. Kron KJ, Murison A, Zhou S, et al. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nat Genet*. 2017;49(9):1336-1345.
7. Mazrooei P, Kron KJ, Zhu Y, et al. Cistrome Partitioning Reveals Convergence of Somatic Mutations and Risk Variants on Master Transcription Regulators in Primary Prostate Tumors. *Cancer Cell*. 2019;36(6):674-689 e676.
8. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-589.
9. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82-93.
10. Loven J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013;153(2):320-334.
11. Saint-Andre V, Federation AJ, Lin CY, et al. Models of human core transcriptional regulatory circuitries. *Genome Res*. 2016;26(3):385-396.
12. Fornes O, Castro-Mondragon JA, Khan A, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2020;48(D1):D87-D92.
13. Wang K, Wang P, Shi J, et al. PML/RARalpha targets promoter regions containing PU.1 consensus and RARE half sites in acute promyelocytic leukemia. *Cancer Cell*. 2010;17(2):186-197.

14. Liu T, Ortiz JA, Taing L, et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol*. 2011;12(8):R83.
15. Korhonen JH, Palin K, Taipale J, Ukkonen E. Fast motif matching revisited: high-order PWMs, SNPs and indels. *Bioinformatics*. 2017;33(4):514-521.
16. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907-915.
17. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-169.
18. Yang H, Robinson PN, Wang K. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. *Nat Methods*. 2015;12(9):841-843.
19. Tan Y, Wang X, Song H, et al. A PML/RARalpha direct target atlas redefines transcriptional deregulation in acute promyelocytic leukemia. *Blood*. 2021;137(11):1503-1516.
20. Zhang X, Xing G, Fraizer GC, Saunders GF. Transactivation of an intronic hematopoietic-specific enhancer of the human Wilms' tumor 1 gene by GATA-1 and c-Myb. *J Biol Chem*. 1997;272(46):29272-29280.