Figure S1

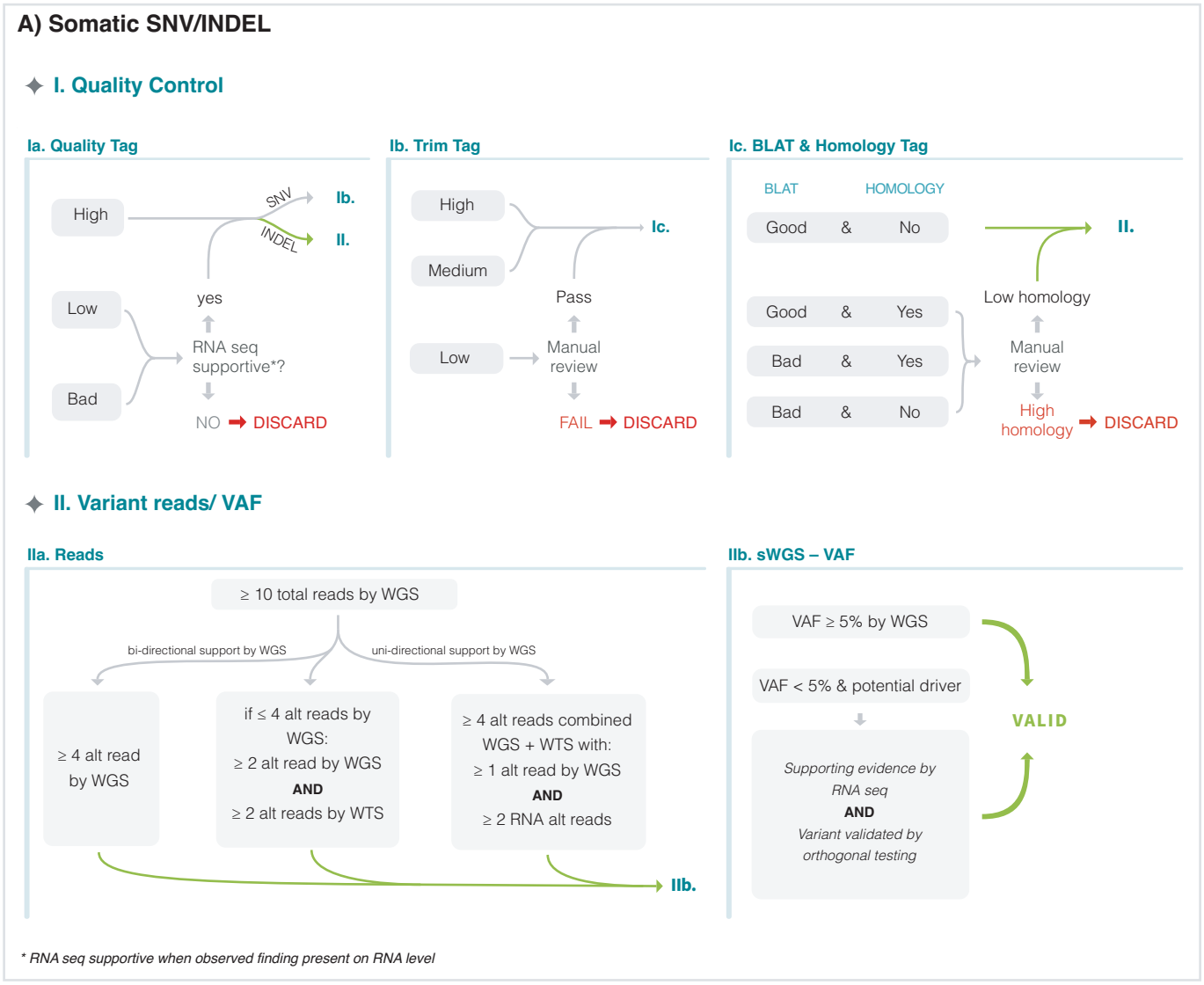


Figure S1. Variant process and auto calling rules. A) SNVs/Indels in tumor testing. This figure details the process and curation rules for Single Nucleotide Variants (SNVs) and small insertion deletions (indels) for Whole Genome Sequencing (WGS). The process starts with rigorous quality control (I) involving three specific tags: (Ia) Quality Tag to ensure base quality, (Ib) BLAT Tag to confirm correct genomic alignment, and (Ic) a Trim Tag to ensure proper trimming of sequencing reads, whereas Ib and Ic only apply for SNVs. After quality control, variant verification is based on mutation reads (II). IIa) A minimum of 10 total reads and at least 4 alternative (alt) reads are required, which may be substituted by RNA seq alt reads. If WGS alt reads show only one-sided support, at least 2 supporting alternative reads by RNA seq are required. IIb) Variants with a Variant Allele Frequency (VAF) of ≥ 5% are considered reliable and classified, and SNVs with a VAF below 5% require verification by an alternative sequencing method (panel/WES) if RNA evidence is present. (IIc).

Figure S1

Abbreviations: CNC, copy number change (absolute); CNV, copy number variant; kb, kilobases; SV, structural variant; WGS, whole genome sequencing

Figure S1. D) Copy number Variants (CNVs). Curation rules for CNVs detected in tumor Whole Genome Sequencing (WGS): Large-scale CNVs, defined as larger than 5,000 kb, require copy number change (CNC) thresholds of ≥ 0.25 to 0.5 depending on their size and whether they are associated with a structural variant (SV). Focal CNVs (smaller than 5,000 kb) require stricter thresholds for CNVs not associated with SVs. Soft-clipped reads provide additional evidence, particularly for CNVs associated with SVs, to ensure accurate curation of complex variants. **WGS germline copy number alterations:** the focus of germline CNV analysis lies in the detection of CNVs, which cover genes associated with cancer/BMF predisposition. First, CNC is evaluated, and if cut-off is met, Germline Reportable (GR) gene list (Supplementary Table S7A) is consulted to determine whether a GR gene lies within the copy number alteration. If so, event may be classified per ACMG recommended rules.

Figure S1

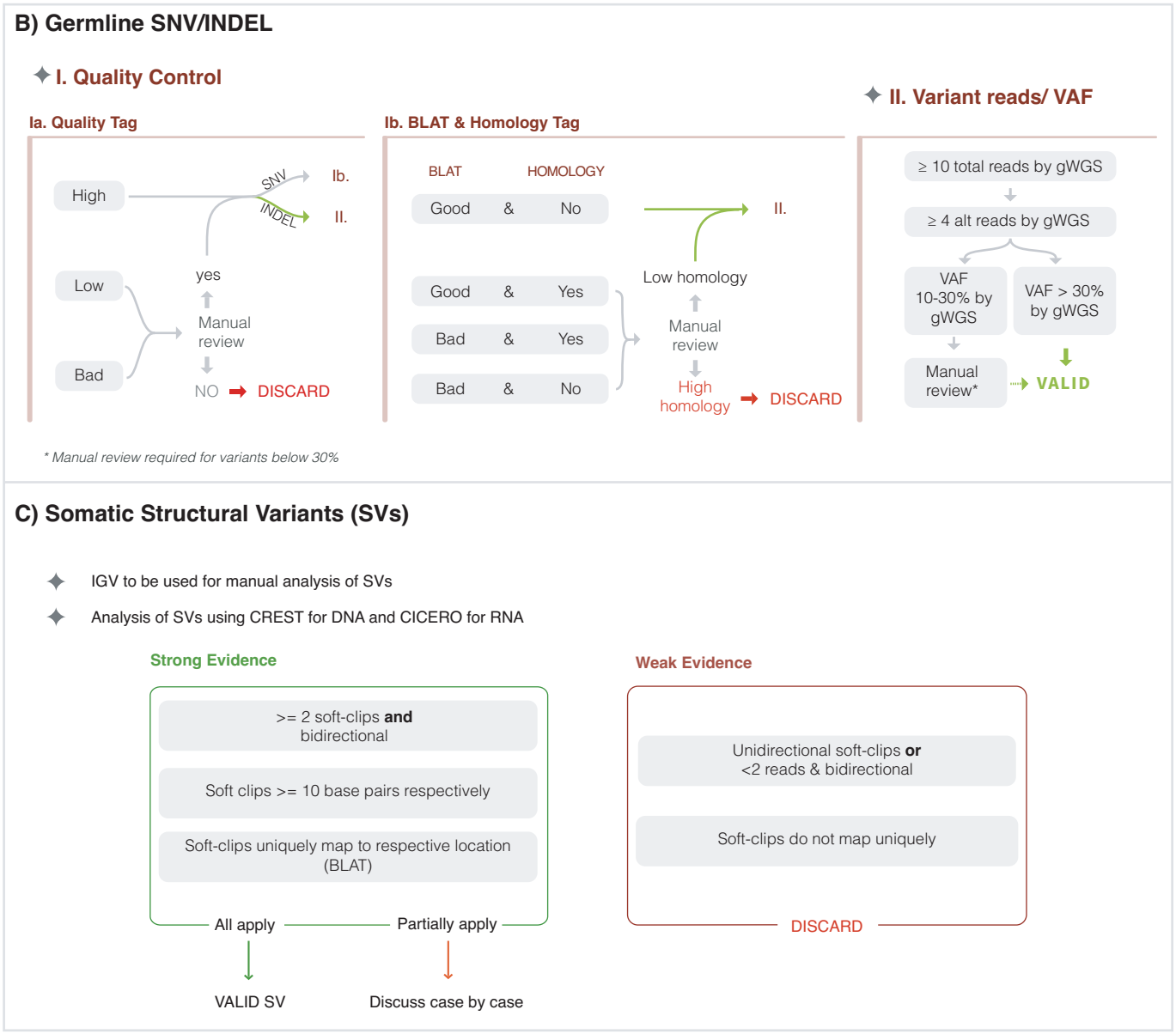


Figure S1. B) Germline SNV/INDEL. For germline variants, the workflow begins with quality control (I), similar to that used for tumor testing but includes an additional 'Homology Tag' to account for sequence homology, reducing false positives from repetitive regions. After passing quality control, germline variants are evaluated based on the total number of alternative (alt) and reference reads. II) A minimum of 10 total reads and more than 4 alternative reads at a VAF of at least 30% are required in WGS data before proceeding to variant classification. Variant with a VAF between 10-30% require manual review. **C) Somatic structural variants.** This figure outlines the curation for Structural Variants (SVs) in tumor Whole Genome Sequencing (WGS). For strong evidence of a valid SV, at least 2 soft-clipped reads must be present on both sides of the variant junction, with each soft-clip being at least 10 base pairs long and uniquely mapping to the genomic location. If only one-sided support or fewer than 2 reads are present, the variant is classified as weak evidence, requiring further manual review. Whole transcriptome sequencing (WTS) may provide additional evidence, with CICERO soft-clips used as support when WGS data is insufficient. Variants that fail to meet any validation criteria are discarded, while borderline cases undergo case-by-case discussion for curation.

Figure S2

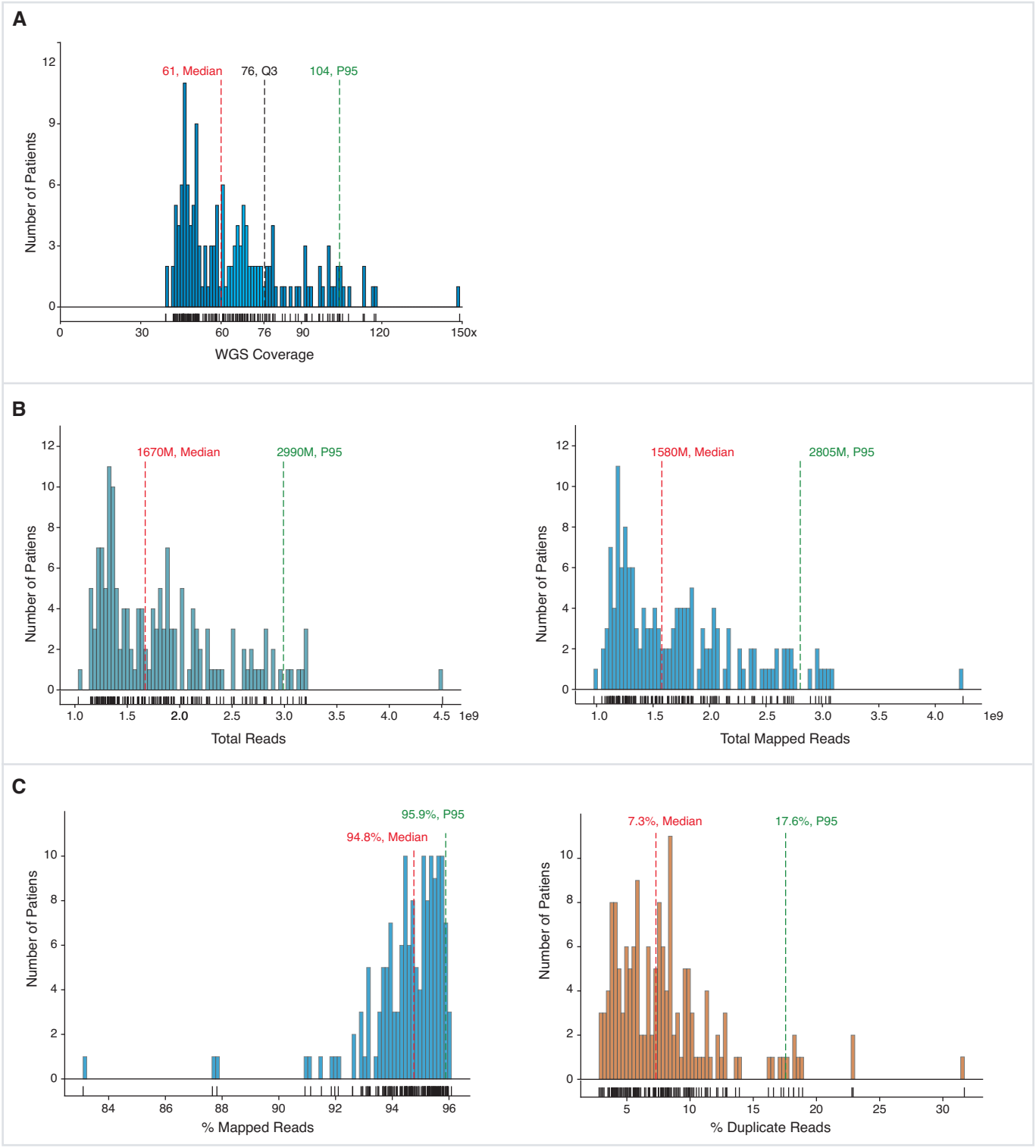


Figure S2. WGS QC metrics. A) This figure illustrates WGS average coverage within our study cohort with a median of 61X and a minimum average coverage of 40X. Q3 is indicated by black dashed line, 95th percentile (P95) is shown as green, dashed lines. **B)** Tallied total number of reads across patients is shown on left, total mapped reads on the right. Median and P95 are indicated by red and green lines, respectively. **C)** Percent of mapped reads are shown on the left, percent duplicate reads on the right. Mapping rate and duplication rate are indicated by median, respectively. P95 are shown as green, dashed lines.

Figure S3

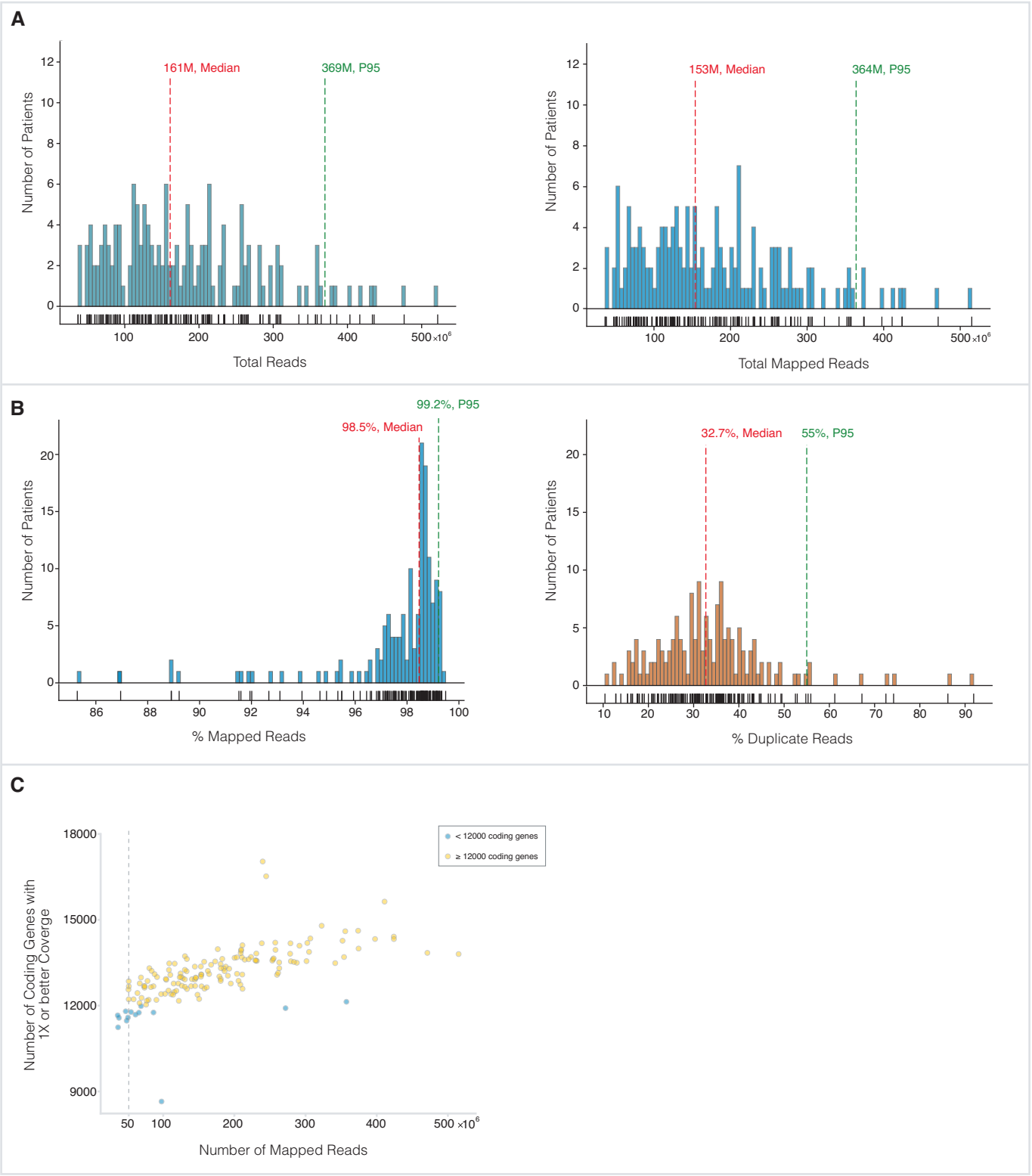


Figure S3. WTS QC metrics. A) Tallied total number of whole transcriptome sequencing (WTS) reads across patients is shown on the left, total mapped reads on the right. Median values are indicated by dashed red lines, 95th percentile (P95) by dashed, green lines. **B)** Percent of mapped reads are shown on the left, percent of duplicate reads on the right. Mapping rate and duplicate rate are indicated by median respectively. The 95th percentiles are shown as green dashed lines. **C)** Number of mapped reads per number of coding genes with 1X or better coverage are depicted for all 154 patients. Yellow dots represent reads with more than 12000 coding genes per mapped read at a minimum of 50 million reads, blue dots represent findings below that threshold. *Gencode v19* was used as a reference gene-list.

Figure S4

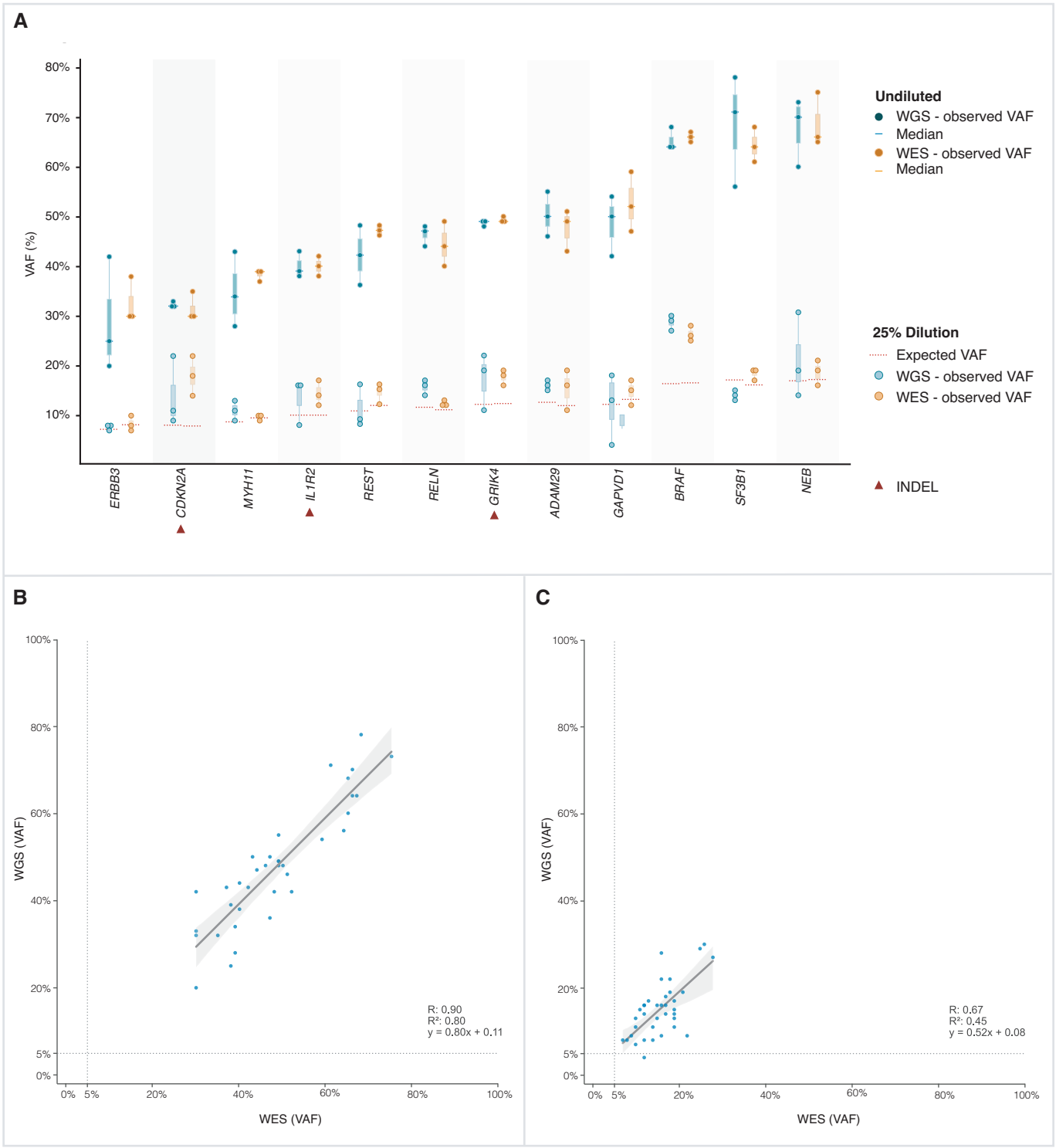


Figure S4. Variant detection by WGS in dilution series using tumor cell line. A) SNV/INDEL detection. Variant allele frequency (VAF, y-axis) is shown for selected variants (x-axis) for whole genome sequencing (blue) and whole exome sequencing (orange) for triplicates of one representative run using cell line COLO 829. Names of genes which contain variants are listed on x-axis. Faded colors indicate results from diluted sample with 25% tumor DNA content (COLO829 diluted with matched normal cell line COLO 829BL), dashed red lines indicate expected VAF after 25% dilution. Genes with INDELs are indicated by triangles. **B)** and **C) Correlation of VAF estimates between WGS and WES. B)** VAF of triplicates from undiluted COLO 829 sample is shown for WGS and WES on x and y axes, respectively. High concordance in VAF estimates is observed (r=0.9). **C)** Results of triplicates for the sample diluted to 25% tumor DNA content are shown for WGS and WES. Correlation of VAF estimates between WGS and WES is reduced but remains moderate (r=0.67).

Figure S4

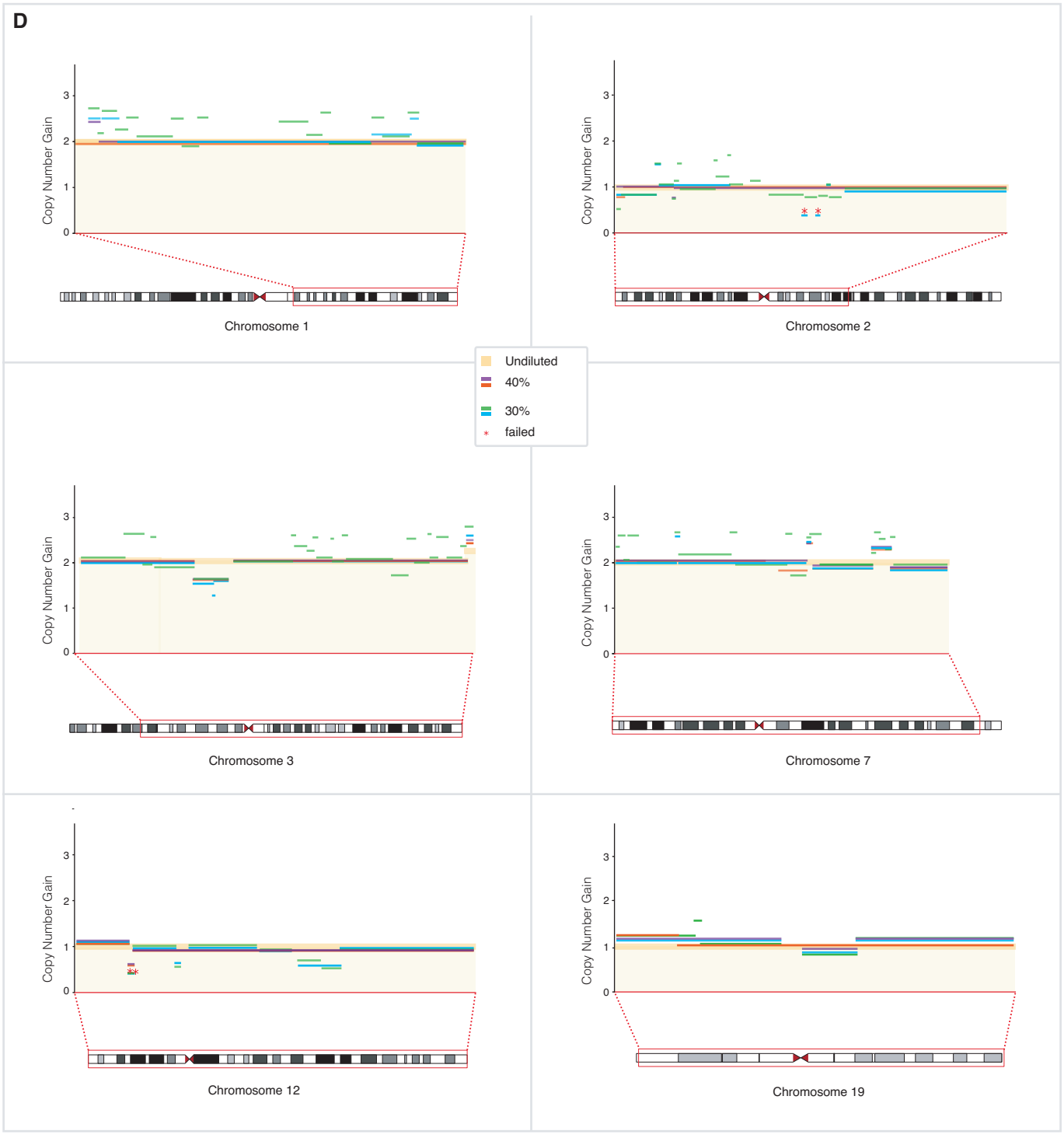


Figure S4. D) Detection of large scale copy number alterations. Representative large scale copy number gains detected by WGS are shown for the original COLO829 cell line, as well as dilutions with 40% and 30% tumor DNA content (COLO829 diluted with COLO 829BL), respectively. Regions of copy number gain are indicated schematically on chromosomes depicted below each graph. The x-axis represents the segment of respective gain, the y-axis indicates absolute copy number gain as detected by WGS. Results from undiluted COLO829 samples are depicted in faded yellow, duplicates of sample diluted to 40% original tumor DNA content is shown in purple and red; duplicates of sample diluted to 30% original tumor DNA content are shown in blue and green. Red asterisk indicates that respective segment failed previously established rules to call copy number change.

Figure S5

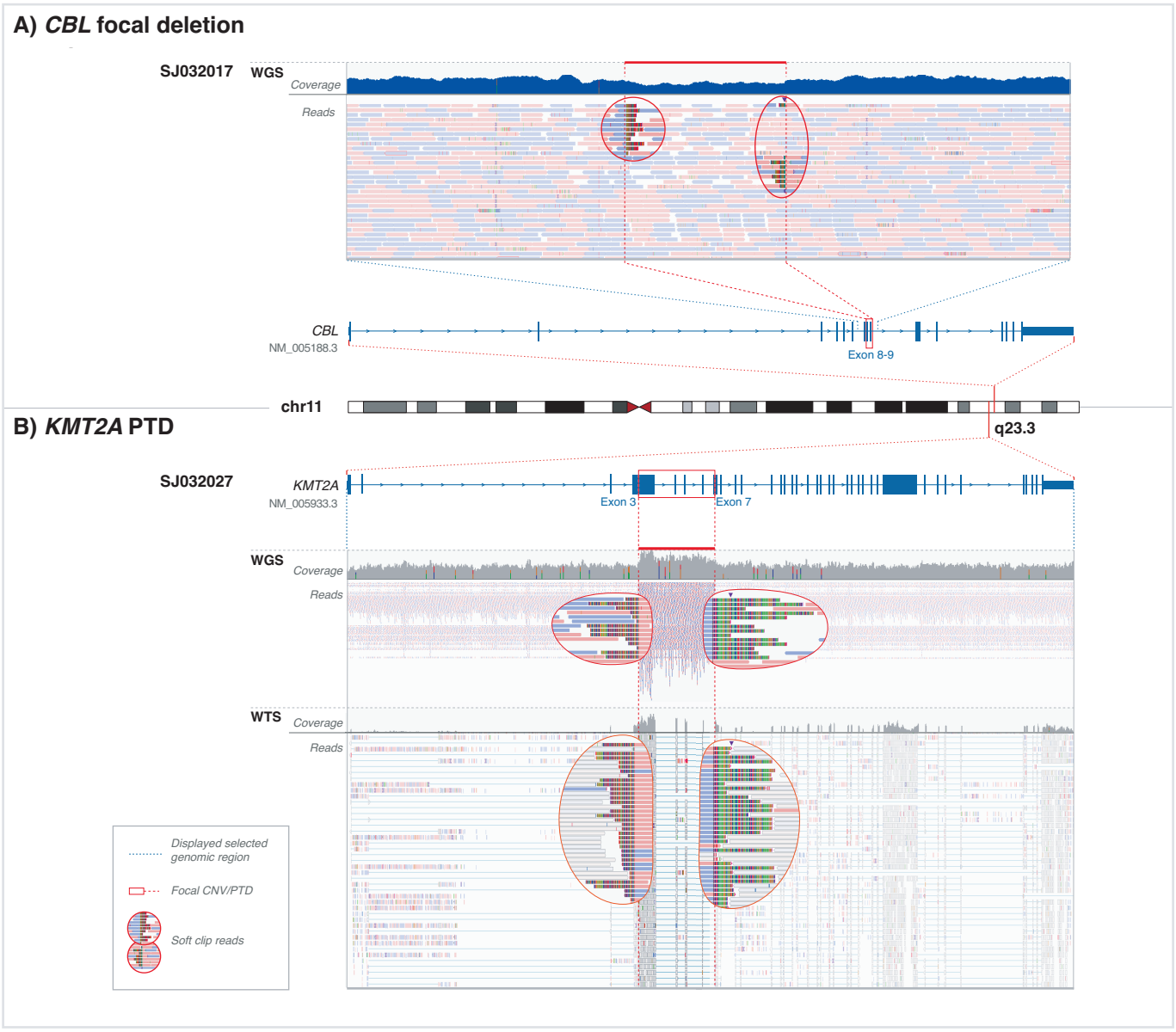


Figure S5. Example cases for focal CNV detection by WGS. A) Intragenic deletion in *CBL* in a representative patient, affecting exons 8 and 9. Soft-clipped bases associated with the focal deletion are circled, a drop in WGS coverage can be observed in the highlighted segment (red line) between soft-clipped bases. **B) *KMT2A*-PDT**. Partial tandem duplication (PTD) affecting exons 3-7 of *KMT2A* is shown in a representative patient. Reads covering the entire *KMT2A* gene are shown, softclipped bases for WGS and WTS are circled in affected area. An increase in WGS coverage is visible on the coverage track (red line).

Figure S6

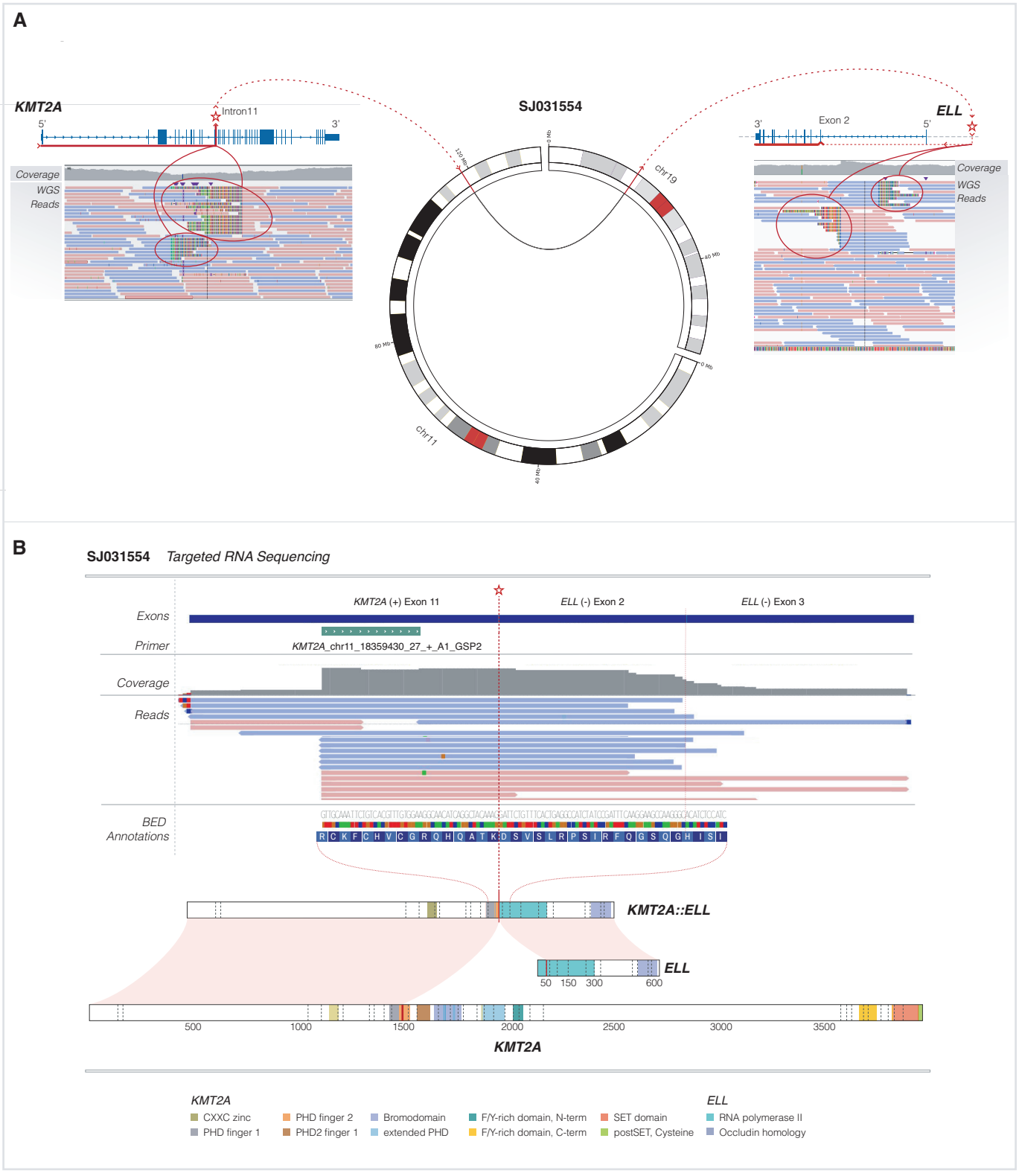


Figure S6. *KMT2A::ELL* fusions in cases SJ031554 and SJ03135. A) SJ031554 – CIRCOS plot shows breakpoints on chromosomes 11 and 19. Soft-clipped reads by WGS indicate *KMT2A* breakpoint in intron 11 and breakpoint on chromosome 19 is found 3.4kb upstream of ELL. **B) SJ031554** – *KMT2A::ELL* fusion verified by targeted RNA sequencing. Fusion transcript and schematic diagram of the predicted chimeric protein are presented.

Figure S6

Figure S6. C) SJ031359 – CIRCOS plot shows breakpoints on chromosomes 11 and 19, indicating a *KMT2A::ELL* fusion. Soft-clipped reads for *KMT2A* are found at the exon-intron junction of exon 9/intron 10. Softclipped reads for *ELL* are found within intron 8. **D) SJ031359** – *KMT2A::ELL* fusion verified by targeted RNA sequencing. Partial sequence of the fusion transcript and schematic diagram of the predicted chimeric protein are presented. Refer to Figure S6B for the information of functional domains of protein.

Figure S7

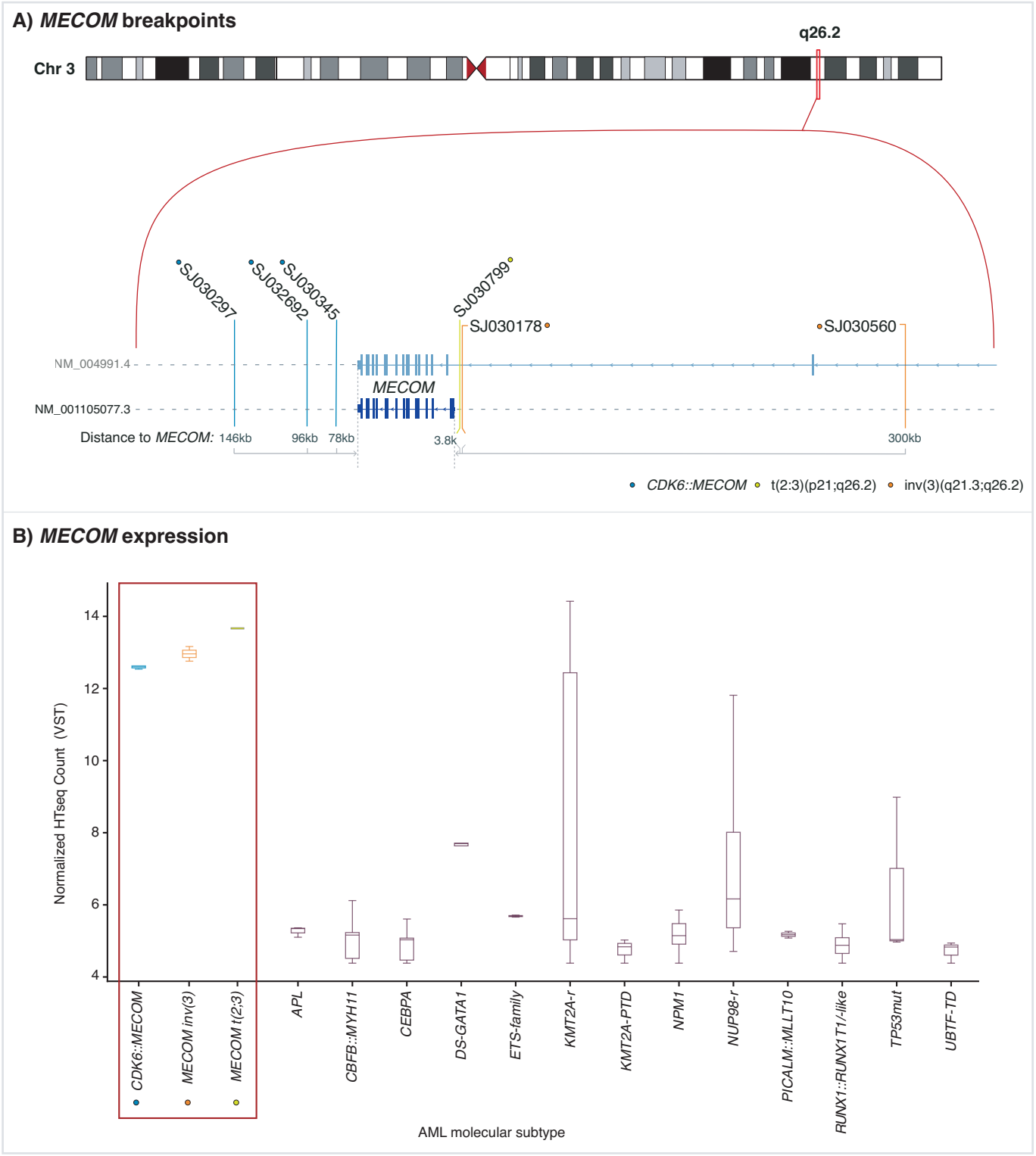


Figure S7. Cases with enhancer hijacking structural alterations. A) Breakpoints in proximity to *MECOM*. Affected region in chromosome 7 is highlighted and *MECOM* transcripts are shown schematically, with lines indicating breakpoints and colors representing subtypes of *MECOM* rearrangements for respective patients. Distance to *MECOM* 5' and 3' ends are shown below. **B) Normalized expression of *MECOM*.** Boxplots framed in red indicate high expression of *MECOM* in patients with *MECOM*-rearranged AML, in contrast to patients defined by other subtypes of AML in this study cohort. Note: subtypes having 2 or more patients in the study

Figure S7

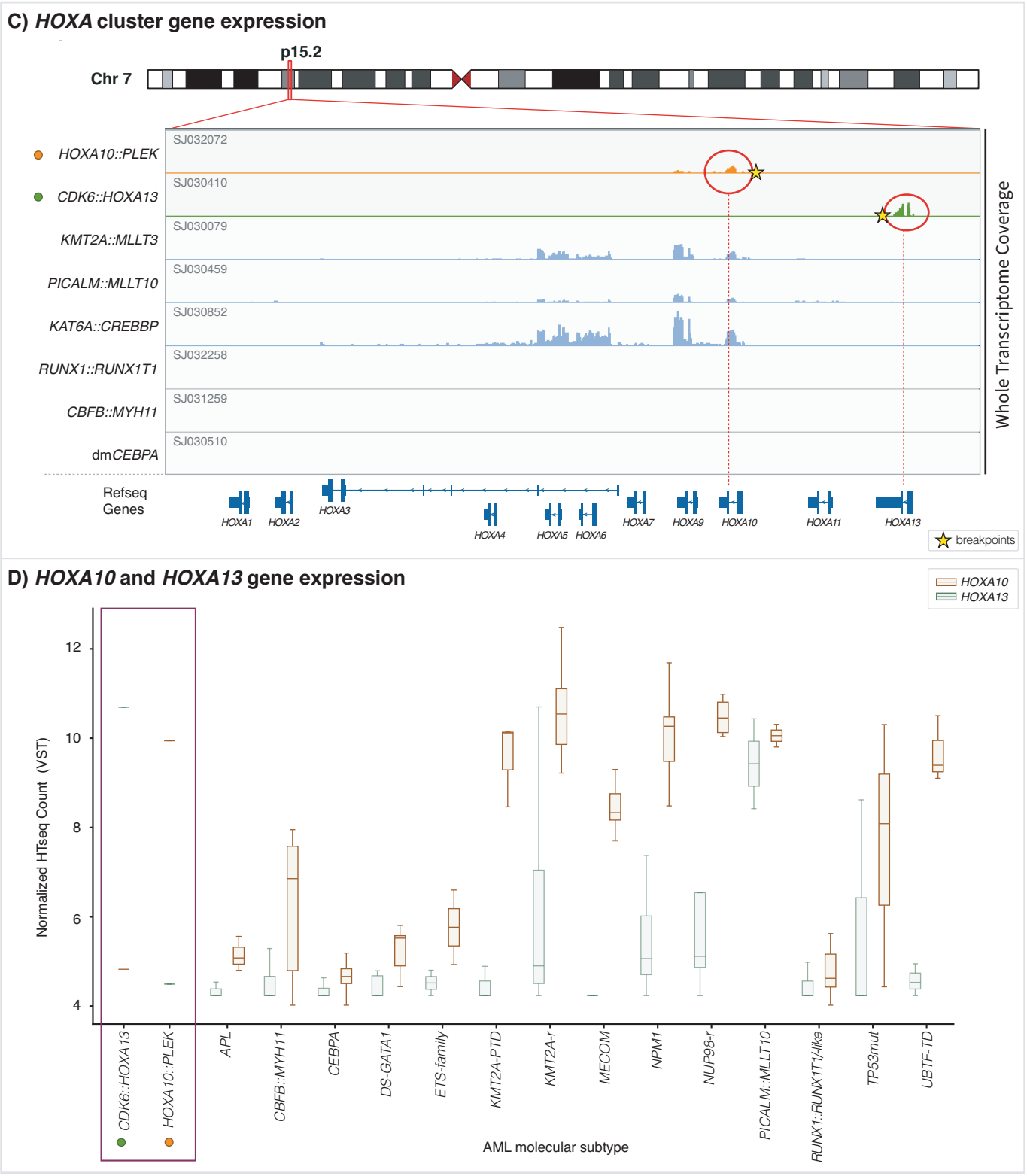


Figure S7. C) WTS coverage (expression) of *HOXA* gene cluster is shown for two patients with *HOXA*-r, in contrast to selected patients with other AML-driver genetic alterations: three patients from molecular subtypes of which high *HOX* expression is expected (*KMT2A*-r, *PICALM::MLLT10*, *KAT6A::CREBBP*), as well as three patients from subtypes of which no increased expression of *HOXA* cluster genes is expected (*RUNX1::RUNX1T1*, *CBFB::MYH11*, *CEBPA-bZIP*). Stars indicate breakpoints. **D) Normalized expression of *HOXA10* and *HOXA13* genes.** *HOX* rearranged AML cases are shown on the left side of the plot. Expression is shown in contrast to patients defined by other subtypes of AML in this study cohort, framed box-plots indicate expression of *HOXA10* (orange) and *HOXA13* (blue).

Figure S7

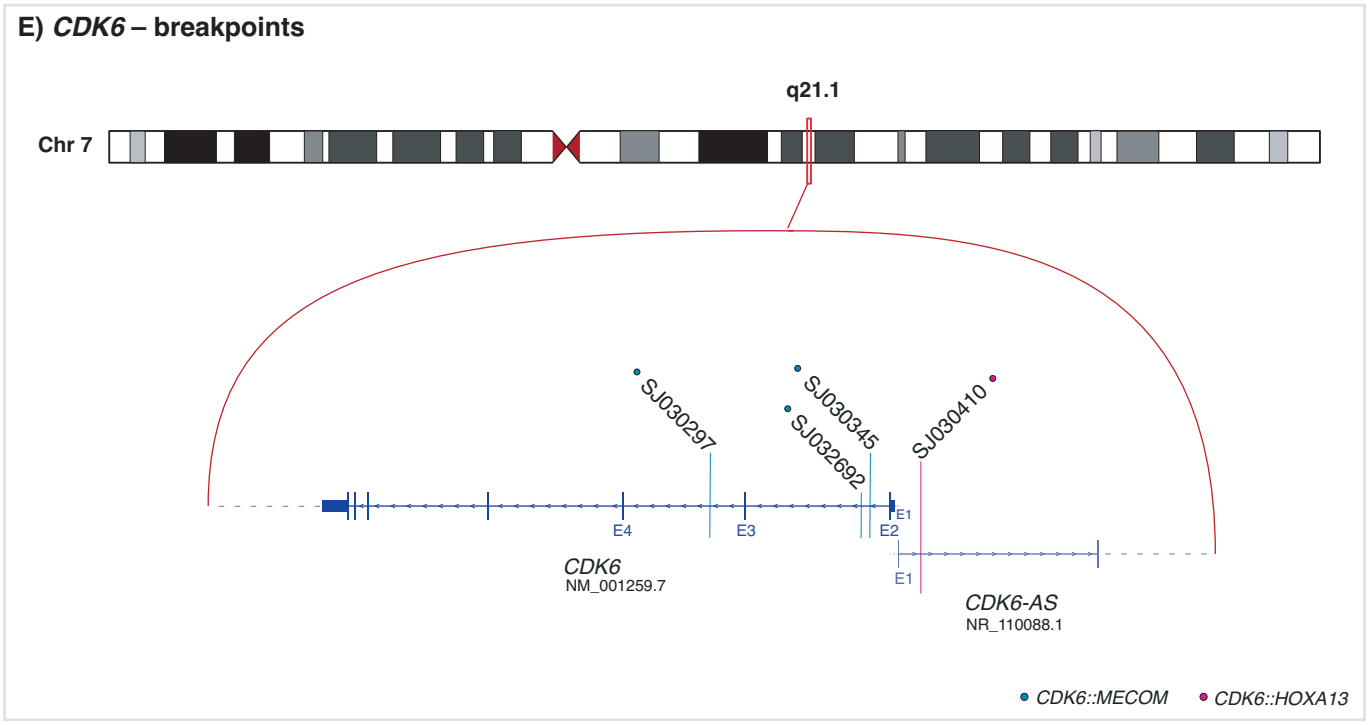


Figure 7. E) Breakpoints in genomic DNA affecting *CDK6*. Affected region within chromosome 7 is marked, *CDK6* and *CDK6-AS* genes are shown schematically. Each vertical line indicates the breakpoint for a given case, colors refer to the type of underlying gene fusion.

Figure S8

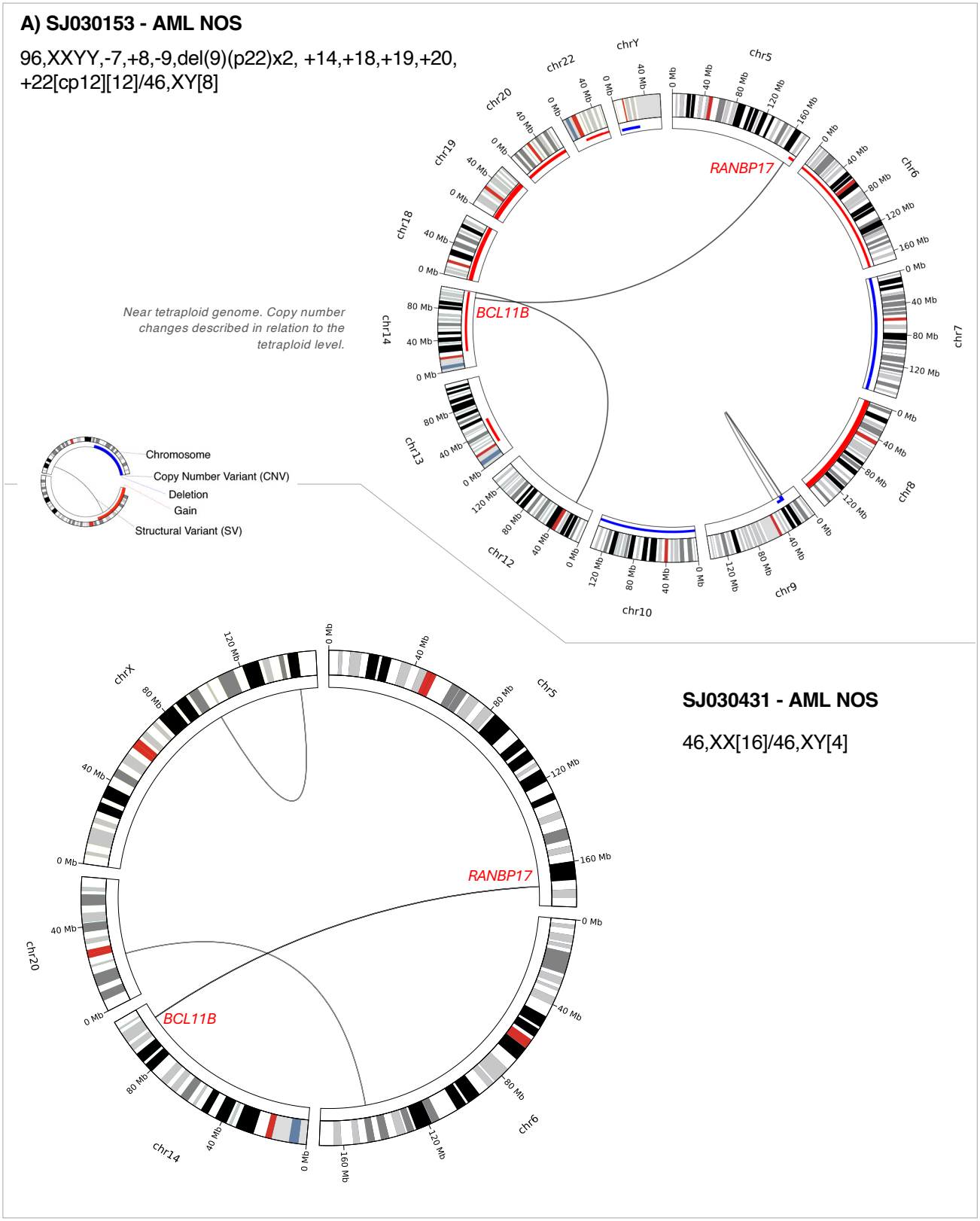


Figure 8. Two cases with t(5;14)(q35;q32.2) by WGS. A) A translocation t(5;14)(q35;q32) identified by WGS only in two patients is shown in CIRCOS plots, suggesting enhancer hijacking fusions involving *BCL11B or TLX3 or NKX2-5*.

Figure S8

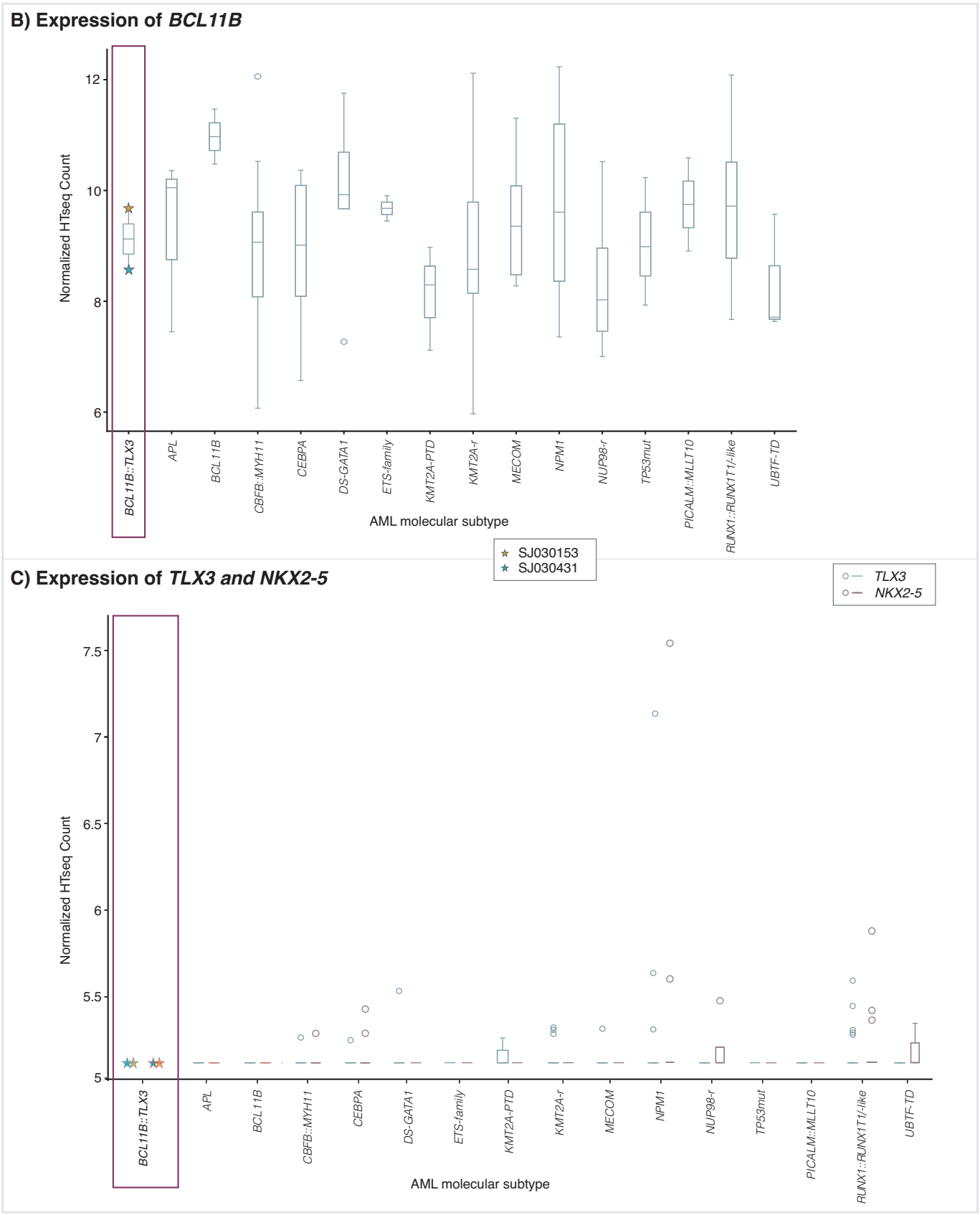


Figure S8. B) and **C) Normalized expression of *BCL11B, TLX3* and *NKX2-5.*** Cases characterized by other AML subtypes used for comparison were sourced from this study cohort, except for the *BCL11B*-AML subtype, which were obtained from St. Jude Cloud (https://www.stjude.cloud/) and included to serve as positive control.

Figure S8

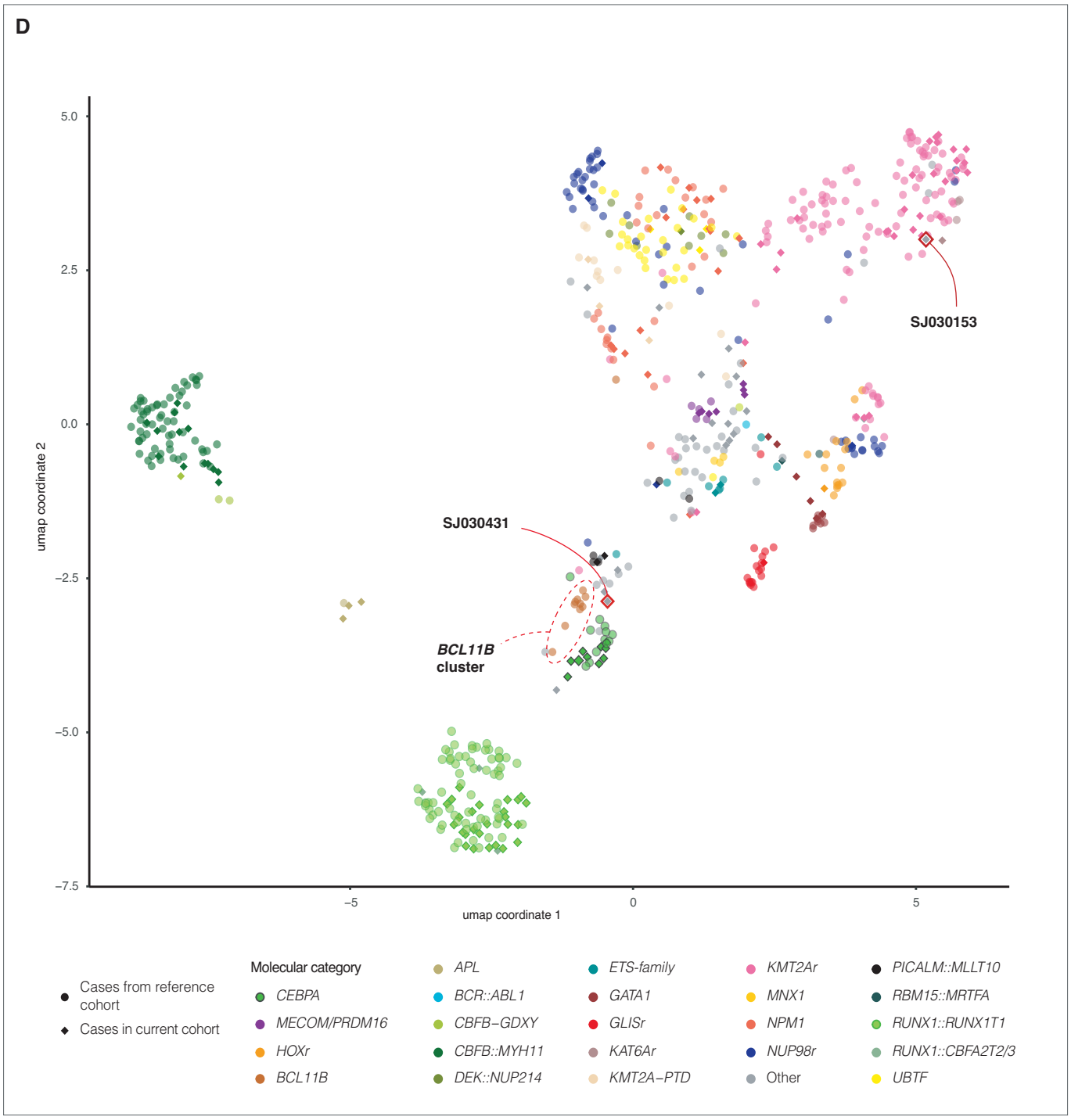


Figure S8. D) Global gene expression profiling by Uniform Manifold Approximation and Projection for Dimentional Reduction (UMAP). Global gene expression profiling is shown in the context of an extended cohort of 420 AML patients including 154 AML patients in this study cohort (indicated as diamonds) and 266 cases of a reference cohort generated from data available through St. Jude Cloud (link). Selected patients are highlighted.

Figure S9

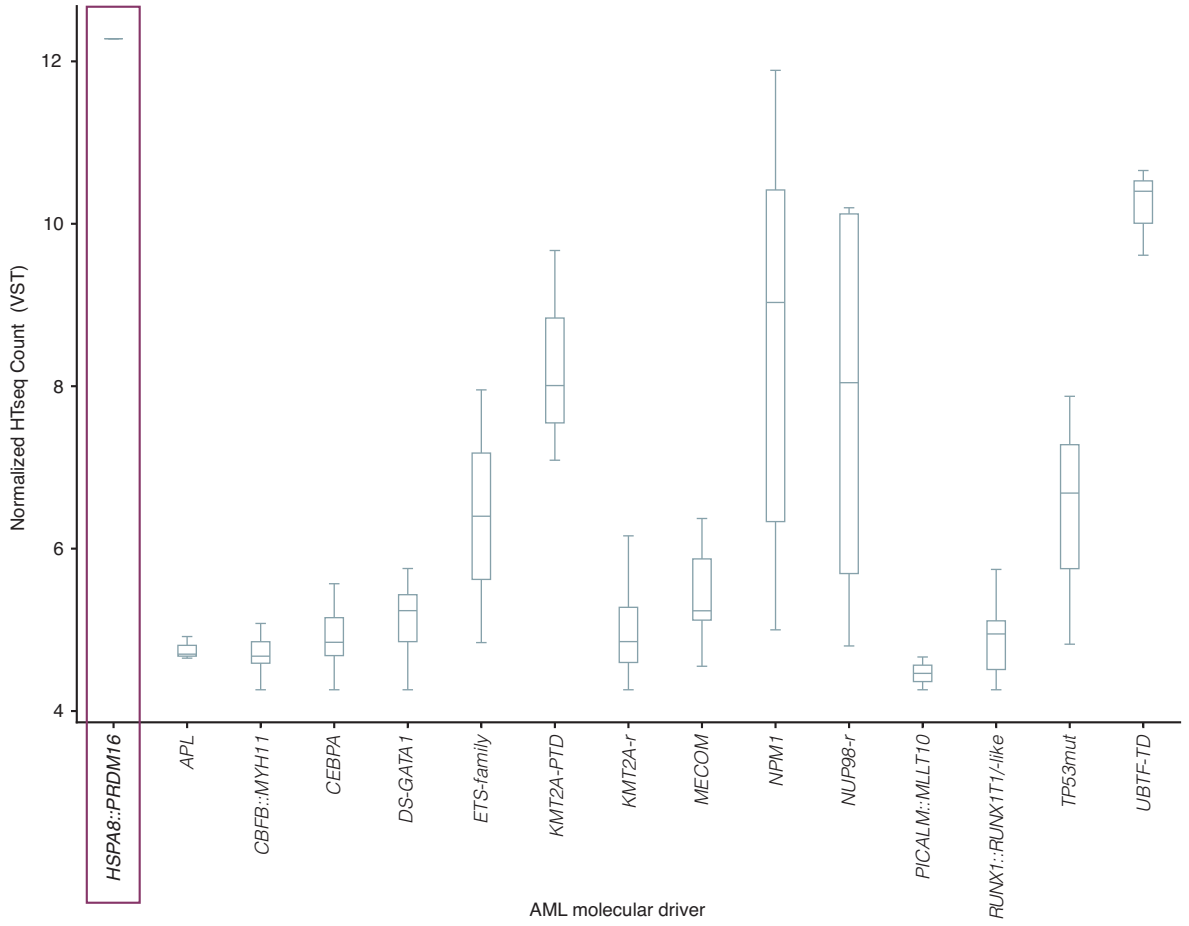


Figure S9. Case SJ031206 with *HSPA8::PRDM16* fusion. A) Normalized Expression of *PRDM16.*

Figure S9

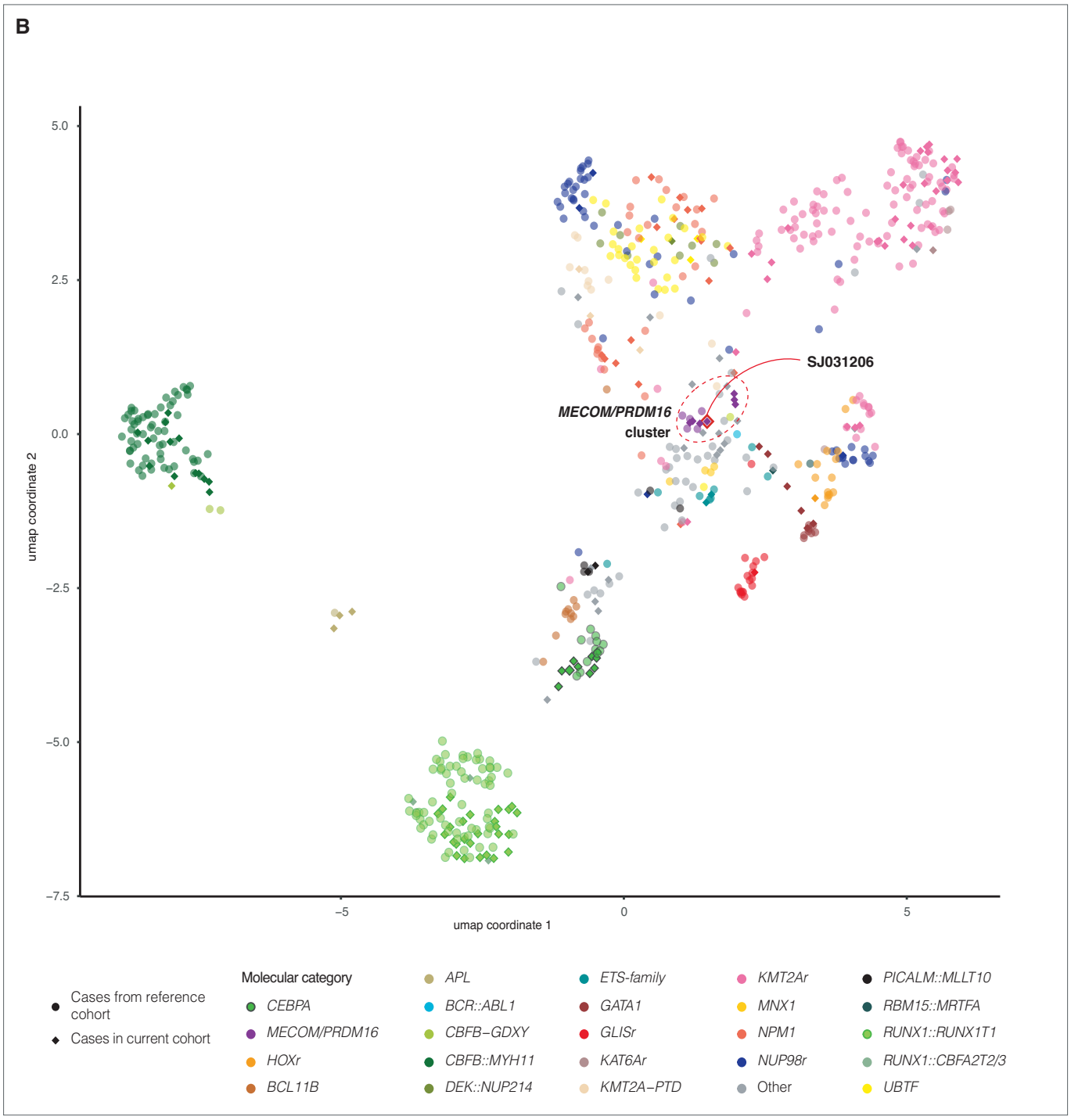


Figure S9. B) Global expression profiling by UMAP. Global expression profiling is shown for an extended cohort of 420 AML patients (see supplementary figure S8D legend for details). The patient with *HSPA8::PRDM16* is highlighted and lies within *MECOM/PRDM16* cluster.

Figure S10

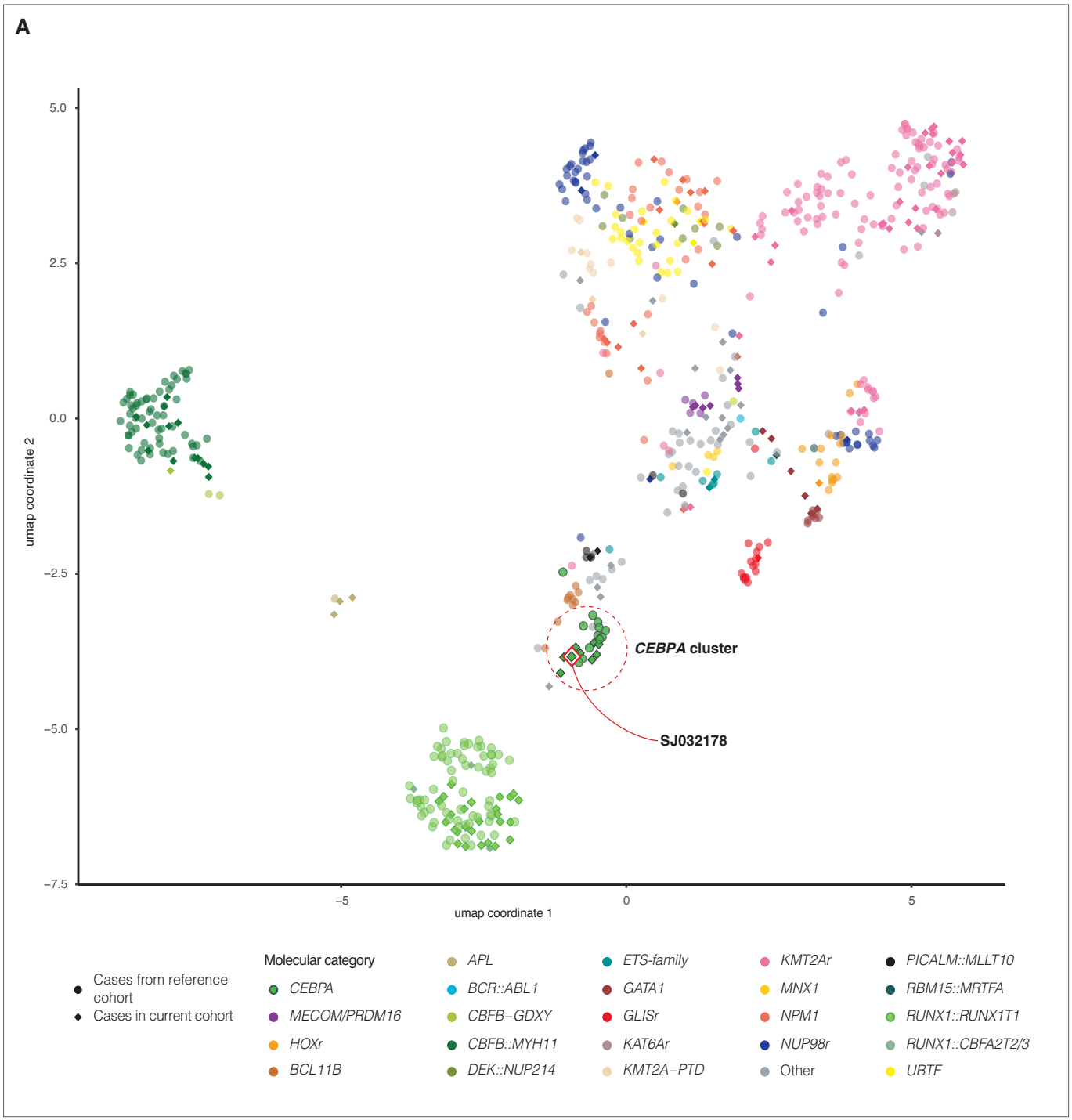


Figure S10. Global gene expression profiling generated from WTS. A) Global gene expression profiling using UMAP shows that a case (SJ032178) with a heterozygous variant in the 5' region of *CEBPA* clusters within *CEBPA bZIP* group.

Figure S10

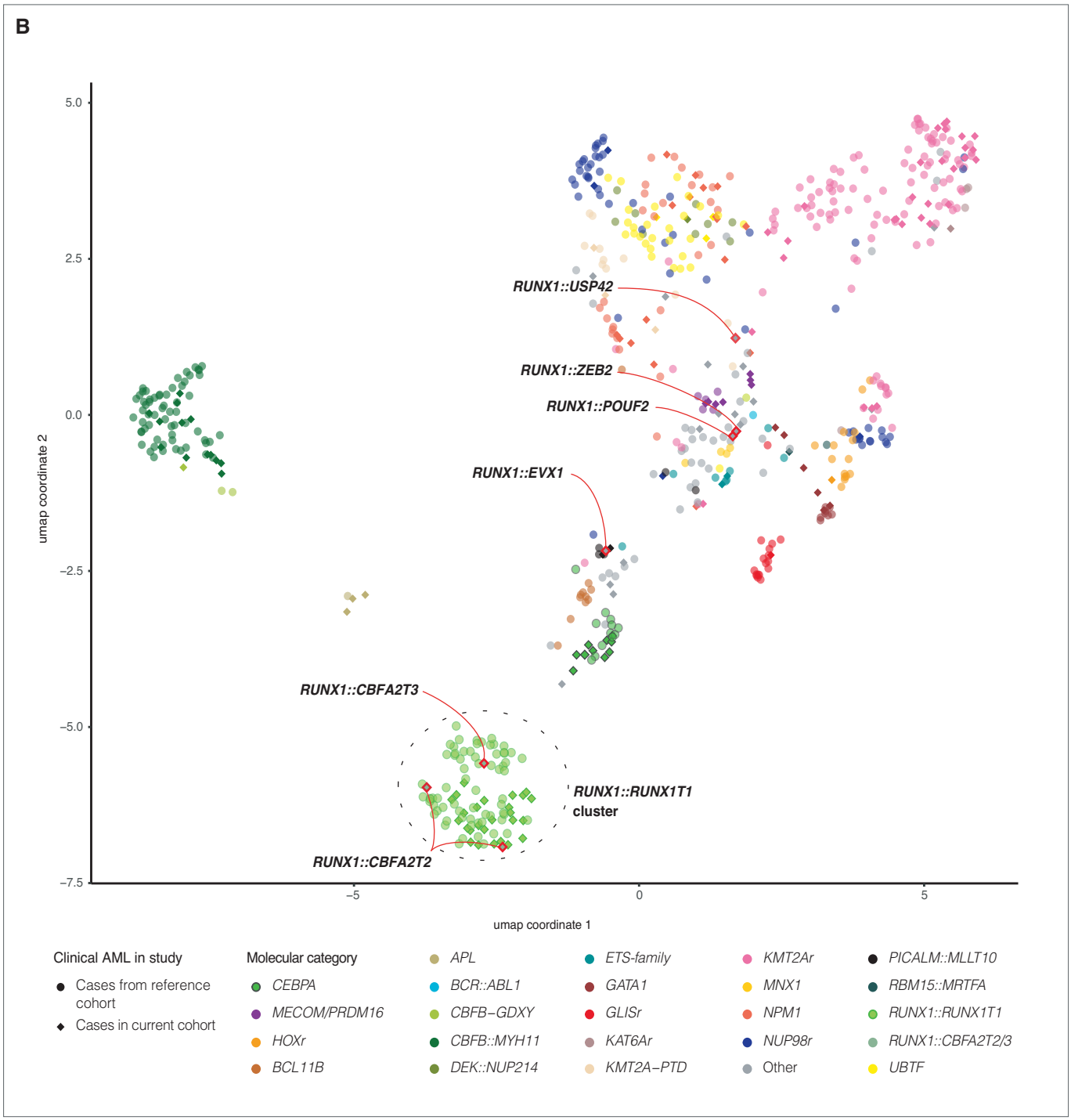


Figure S10. B) Patients with *RUNX1*-r are highlighted. Global gene expression profiling clustered three *RUNX1::CBFA2T3/2* cases with the *RUNX1::RUNX1T1* group; in contrast, four cases with *RUNX1* fused to other genes did not cluster with the *RUNX1::RUNX1T1* group, nor form a distinct cluster.

Figure S11

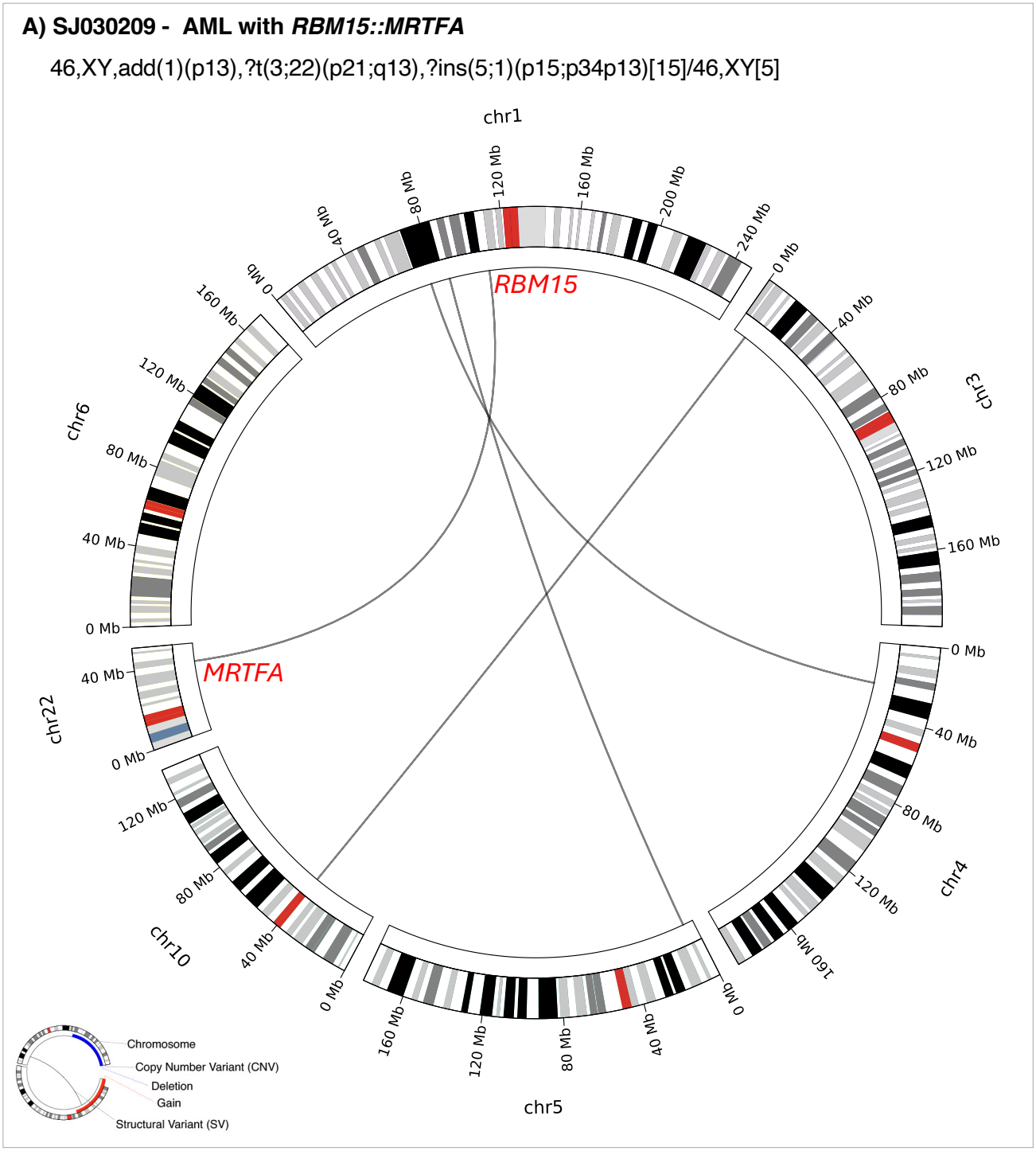
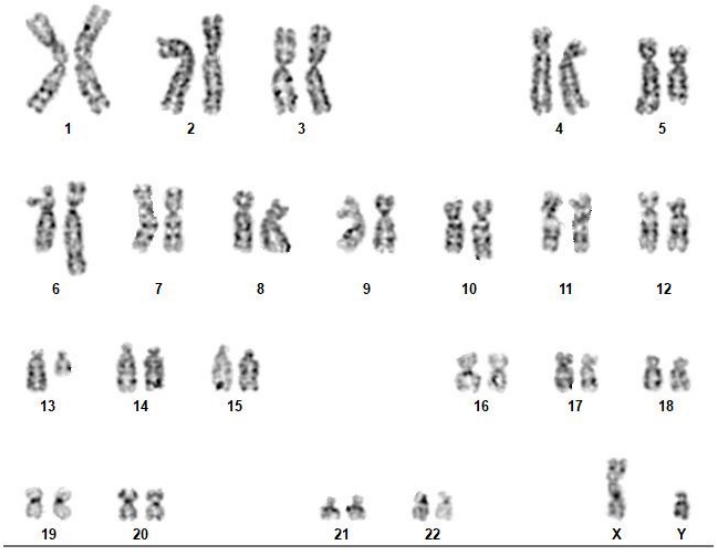


Figure S11. Example cases demonstrating that the integrated WGS and WTS approach resolves complex genomes and identifies drivers in cases with complex karyotype. A) SJ030209. WGS revealed multiple balanced chromosome translocations across the tumor genome and elucidated the complex karyotype reported by conventional cytogenetics. In the complex genome, the integrated WGS and WTS (iWGS-WTS) analysis identified an AML driver, *RBM15::MRTFA* fusion.

Figure S11

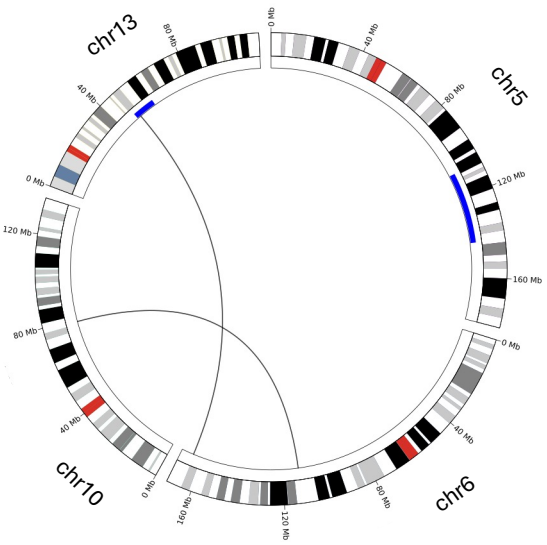
B) SJ032210 – AML with *NPM1*mut

Conventional karyotyping (cKT)



46,XY,del(5)(q22q33),t(6;13)(q25;q12),?**inv(10)(q22q26)**[13]/ 46,XY[7]

WGS



Spectral Karyotyping

Figure 11. B) SJ032210. A possible inv(10)(q22q26) was observed by G-banded karyotyping. Whole genome sequencing refined this as a t(6;10)(q21;q22.3), SKY (spectral karyotyping) verified the translocation between 6q and 10q. The iWGS-WTS analysis further suggested that the chromosome translocation t(6;10)(q21;q22.3) does not result in a fusion oncogene or disrupt a tumor suppressor gene (not shown). Instead, an AML-defining *NPM1* mutation was identified (Supplementary Table S2).

Figure S12

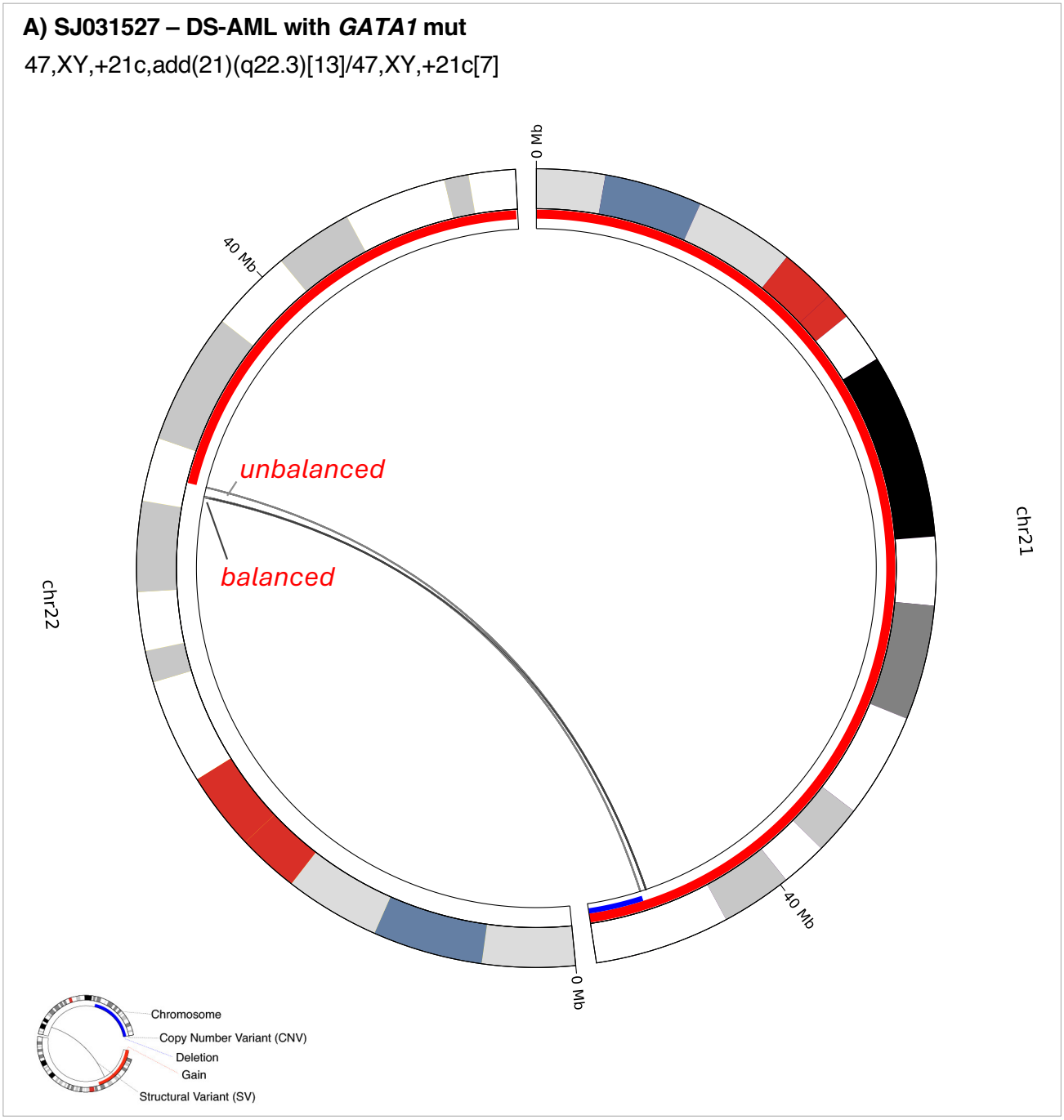


Figure S12. **Abnormal chromosomes observed by conventional karyotyping (cKT) further clarified by WGS**. CIRCOS plots of 5 patients, SJ031527, SJ031601, SJ031719, SJ032364 and SJ032526. Copy number variants and structural variants detected by WGS are shown for affected chromosomes. Outer ring represents chromosomes, bars within represent cytobands (centromere in red). Inner ring shows copy number alterations, loss in blue and gain in red. Lines represent inter- or intra- chromosomal rearrangements. **A) SJ031527**, a case with constitutional trisomy 21 and mutation in *GATA1* (Supplementary Table S2). The add(21)(q22.3) observed by karyotype analysis was further defined by WGS as a der(21)t(21;22)(q22.3;q12.2).

Figure S12

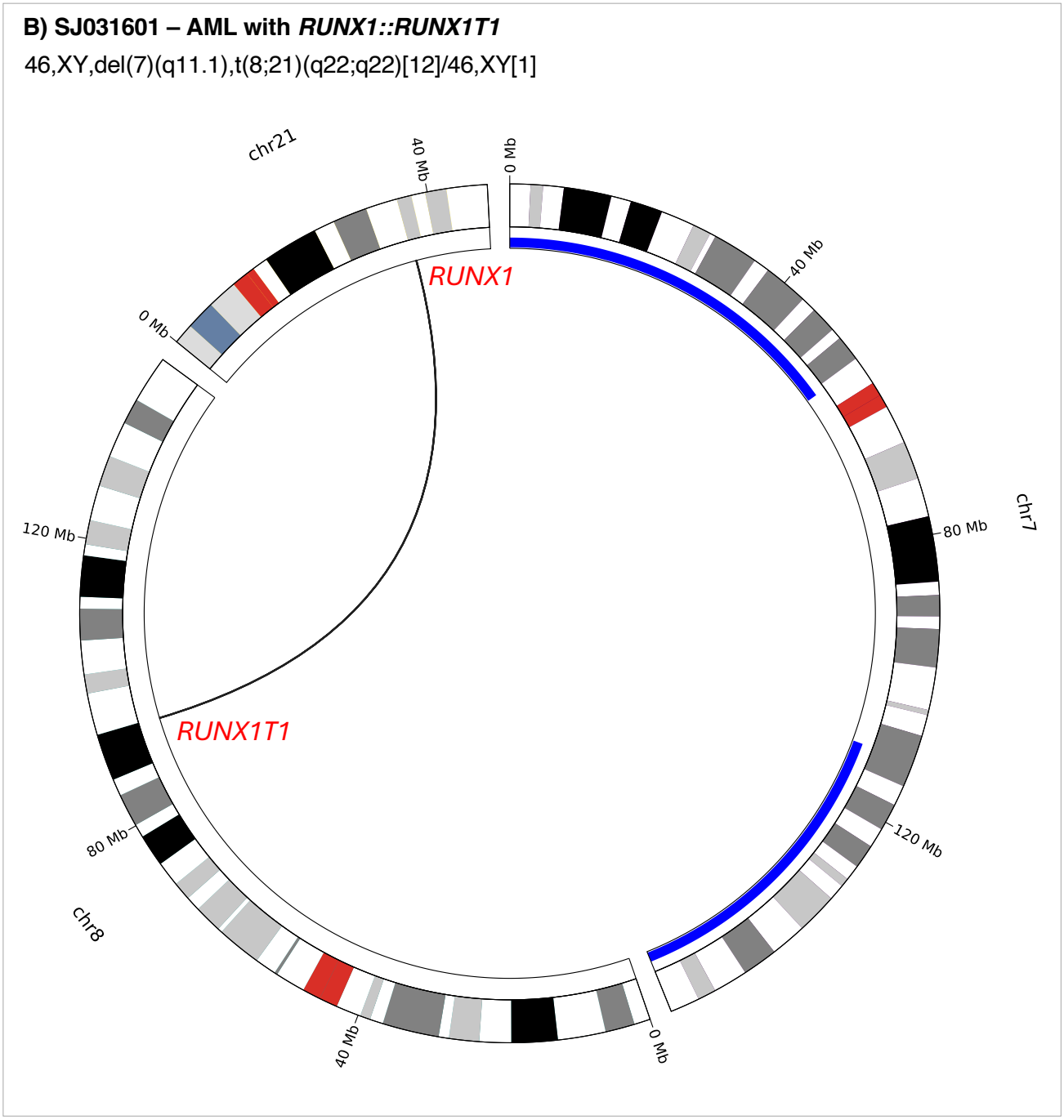


Figure S12. B) SJ031601, AML with *RUNX1::RUNX1T1*. A del(7)(q11.1) observed by karyotype analysis was further defined by WGS as der(7)del(7)(p11.2)del(7)(q31.1).

Figure S12

C) SJ031719 – AML with *TBL1XR1::RARB*

46,XX,t(3;3)(p24;q26.2),-16,+mar[13]/46,idem,del(11)(q21q23)[3]

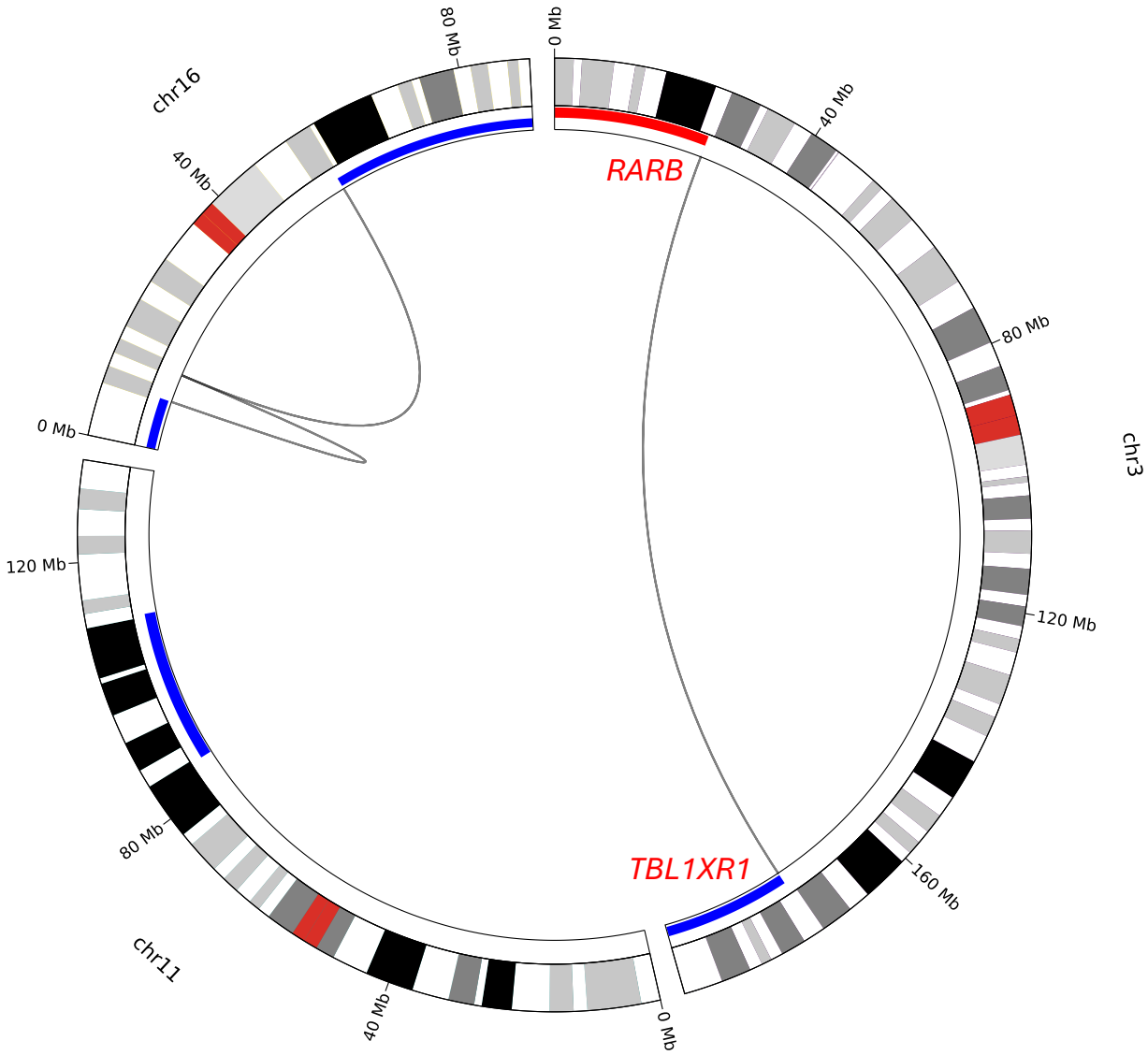


Figure S12. C) SJ031719, AML with *TBL1XR1::RARB*. WGS showed segmental CNVs on 3p and 3q that are associated with the fusion. In addition, deletions in both p and q arms of chr 16, were detected, suggesting the marker chromosome in cKT was derived from chr 16.

Figure S12

D) SJ032364 – AML with *RUNX1::ZEB2*

48,XX,t(2;11)(q21;q22),+6,+7[11]/48,idem,-7,add(12)(p13),+22[4]/46,XX[5]

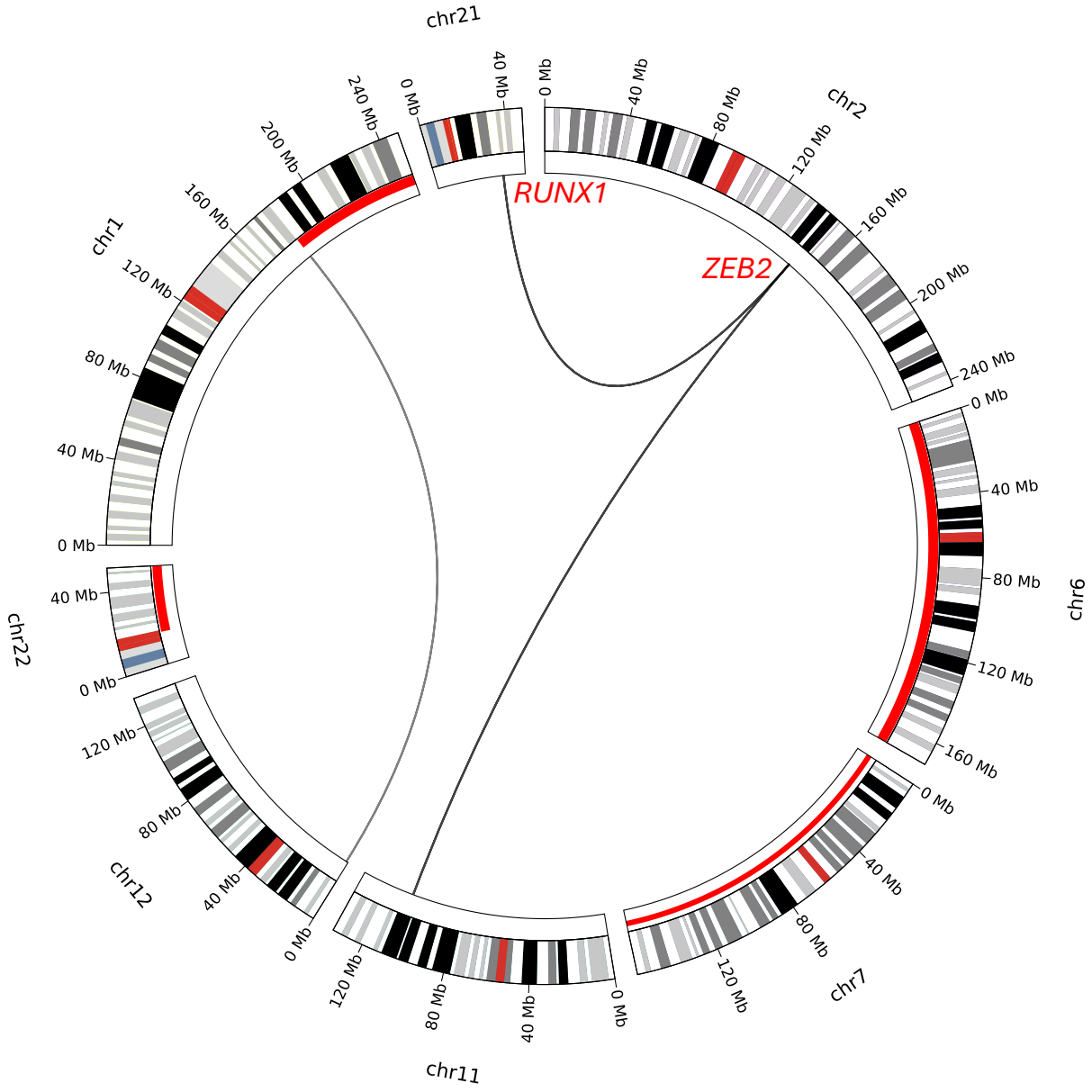


Figure S12. D) SJ032364, AML with *RUNX1::ZEB2*. An add(12)(p13) observed by cKT is further defined by WGS as der(12)t(1;12)(p13.3;q25.3).

Figure S12

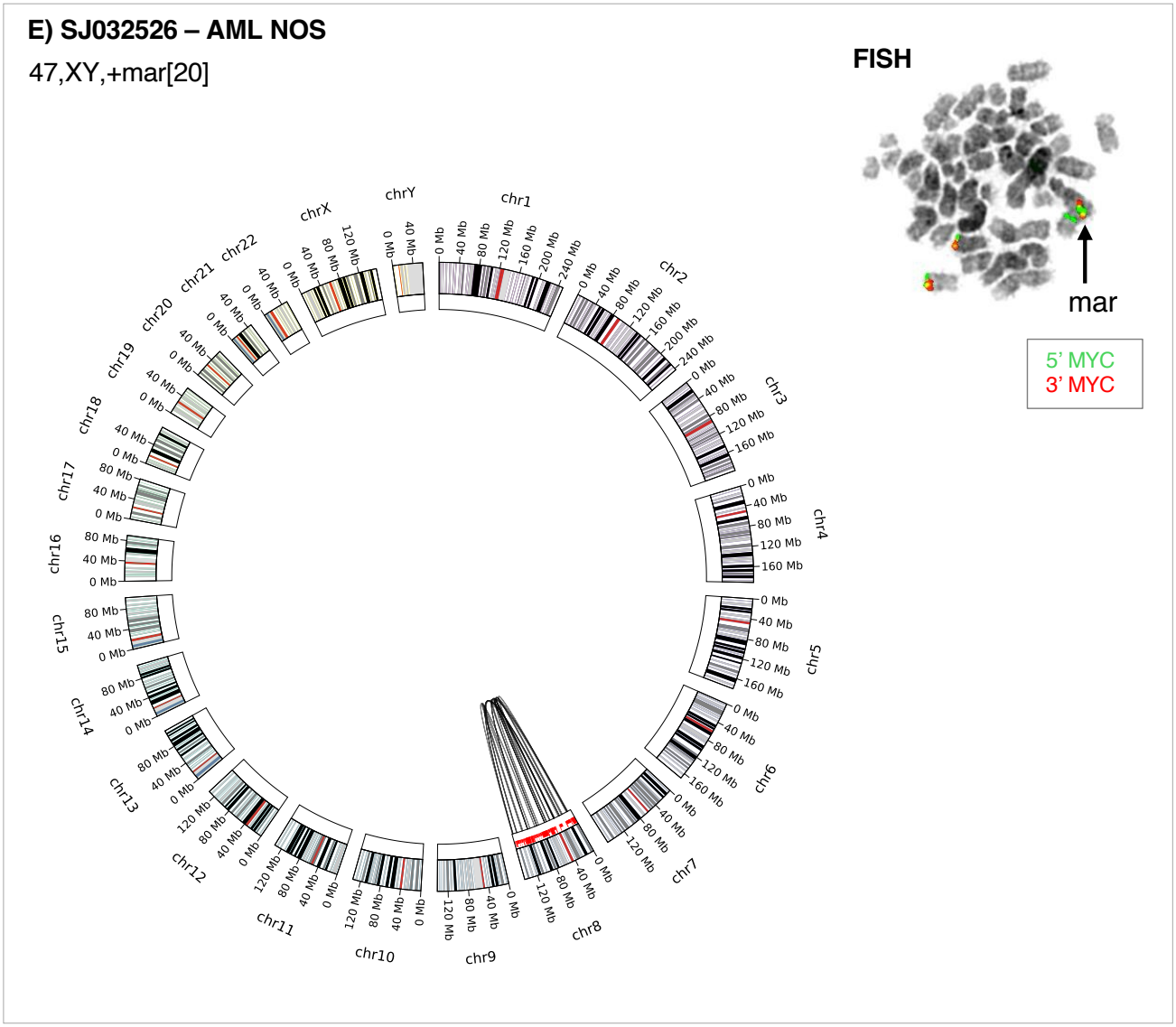


Figure S12. E) SJ032526, AML NOS. WGS revealed copy gains of chromosome 8 and chromothripsis-like changes across chromosome 8, suggesting that the marker chromosome observed by cKT was derived from chromosome 8. This was further supported by FISH analysis using probes targeting *MYC* (8q24) .