

Genomic landscape of pediatric B-other acute lymphoblastic leukemia in a consecutive European cohort

Marketa Zaliova,^{1,2,3} Jan Stuchly,^{1,2} Lucie Winkowska,^{1,2} Alena Musilova,^{1,2} Karel Fiser,^{1,2} Martina Slamova,^{1,2} Julia Starkova,^{1,2} Martina Vaskova,^{1,2} Ondrej Hrusak,^{1,2,3} Lucie Sramkova,^{1,2,3} Jan Stary,^{2,3} Jan Zuna^{1,2,3} and Jan Trka^{1,2,3}

¹CLIP - Childhood Leukaemia Investigation Prague; ²Department of Paediatric Haematology and Oncology, Second Faculty of Medicine, Charles University and ³University Hospital Motol, Prague, Czech Republic

©2019 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2018.204974

Received: August 21, 2018.

Accepted: December 20, 2018.

Pre-published: January 10, 2019.

Correspondence: *JAN TRKA/MARKETA ZALIOVA*

jan.trka@fmotol.cuni.cz/

marketa.zaliova@fmotol.cuni.cz

Genomic landscape of pediatric B-other acute lymphoblastic leukemia in a consecutive European cohort

Marketa Zaliova^{1,2,3}, Jan Stuchly^{1,2}, Lucie Winkowska^{1,2}, Alena Musilova^{1,2}, Karel Fiser^{1,2}, Martina Slamova^{1,2}, Julia Starkova^{1,2}, Martina Vaskova^{1,2,3}, Ondrej Hrusak^{1,2,3}, Lucie Sramkova^{1,2,3}, Jan Stary^{2,3}, Jan Zuna^{1,2,3} and Jan Trka^{1,2,3}

¹ CLIP - Childhood Leukaemia Investigation Prague

² Department of Paediatric Haematology and Oncology, Second Faculty of Medicine, Charles University, Prague, Czech Republic

³ University Hospital Motol, Prague, Czech Republic

Supplementary data

Contents

Methods	3
Supplementary Tables.....	5
Supplementary Table 1. Demographic and clinical characteristics of BCP-ALL patients (stratified into subgroups and subtypes) treated according to AIEOP-BFM ALL 2009 trial protocol	5
Supplementary Table 10. Early-treatment-response-based risk assignment of B-other ALL stratified by the presence/absence of PAX5 fusion.....	6
Supplementary Figures.....	7
Supplementary Figure 1. The <i>DUX4</i> gene expression.	7
Supplementary Figure 2. Total number of CNAs and SNVs/indels in B-other ALL stratified by subtype.....	9
Supplementary Figure 3. Schematic presentation of proteins encoded by <i>PAX5</i> and its partner genes identified in the present study.....	10
Supplementary Figure 4. The <i>ZCCHC7-PAX5</i> and <i>PAX5-ZCCHC7</i> fusions resulting from an intrachromosomal inversion within 9p in the patient 2486.	11
Supplementary Figure 5. Intragenic <i>PAX5</i> amplification detected in 6 patients from B-rest group by SNP-array.....	12
Supplementary Figure 6. Aberrant <i>PAX5</i> transcripts in the patients with intragenic <i>PAX5</i> amplification identified by RNA-seq.....	13
Supplementary Figure 7. Results of HCA analyses.	16
Supplementary Figure 8. Expression of <i>P2RY8-CRLF2</i> transcript in patients with <i>P2RY8-CRLF2</i> fusion.....	17

Supplementary Tables 2-9 are provided in a separate Excel file:

Supplementary Table 2: RNA-seq data metrics

Supplementary Table 3: WES data metrics

Supplementary Table 4: Gene sets used for ALL subtype classification

Supplementary Table 5: Genes included in ALL subtype classifying gene sets

Supplementary Table 6: Overview of clinical characteristics, performed analyses and genomic findings

Supplementary Table 7: Copy number aberrations and regions of uniparental disomy (UPD) identified by SNPa

Supplementary Table 8: Mutations detected by WES and RNA-seq

Supplementary Table 9: Fusion transcripts detected by RNA-seq

Methods

Treatment stratification

According to the AIEOP-BFM ALL 2009 treatment protocol, patients with *BCR-ABL1*-negative BCP-ALL were stratified into three branches (standard, medium and high risk). Stratification algorithm included minimal residual disease (MRD) monitoring with at least one MRD target with quantitative range of $\leq 10^{-4}$. Patients MRD-negative on treatment days 33 (TP1) and 78 (TP2) and without any high-risk criteria were stratified to standard-risk arm. Patients with poor-response to prednisone (≥ 1000 blast cells/ μl in peripheral blood on day 8), blast cells $\geq 10\%$ in bone marrow on day 15 as measured by FCM, non-remission on day 33, positive for *KMT2A-AFF1* or *t(4;11)*, positive for hypodiploidy (< 45 chromosomes), with MRD $\geq 10E^{-3}$ on day 78 or with MRD $\geq 10^{-3}$ on day 33 and MRD positive at a level of $< 10E^{-3}$ on day 78 were stratified to high-risk arm. Remaining patients were stratified to medium risk arm.

Sample processing

Mononuclear cells (MNCs) were isolated by Ficoll-Paque (Pharmacia, Germany) gradient centrifugation from bone marrow or peripheral blood samples. In samples with $\leq 70\%$ leukemia involvement, leukemic cells were sorted using a fluorescence-assisted-cell-sorter (FACS). Total DNA and RNA were isolated from MNCs or from sorted cells as part of the routine sample processing procedure. The integrity of RNA was analyzed by chip electrophoresis using the RNA 6000 Nano Kit on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The concentration of the nucleic acids was determined by spectrophotometry using a NanoDrop 2000 or by fluorometry using a Qubit® dsDNA BR Assay Kit on a Qubit® 2.0 Fluorometer (Life Technologies, USA). DNA isolated from bone marrow aspirates collected in remission or from separated T lymphocytes served as individual germline controls.

Single-nucleotide polymorphism array (SNPa)

HumanOmni Express BeadChip

DNA labeling and hybridization were performed according to the Infinium HD assay Ultra protocol from Illumina. The GenomeStudio software v2011.1 (Illumina) was used for genotype calling and quality control. Copy number variations (CNV) and UPD were called using the CNV Partition 2.4.4 algorithm plug-in within the GenomeStudio software. The resulting data (Log R ratio corresponding to copy number and B allele frequency corresponding to SNP genotype) were visually inspected in the Illumina Chromosome Browser.

CytoScan HD arrays

Analysis was performed as a service in the Laboratory for Molecular Biology and Tumor Cytogenetics at the Department of Internal Medicine of Hospital Barmherzige Schwestern (Linz, Austria). The Chromosome Analysis Suite software (Affymetrix) was used for quality control, genotype calling, CNV/UPD identification and data visualization.

The results from both platforms were manually curated. Deletions corresponding to somatic rearrangements of the immunoglobulin and T-cell receptor gene loci, germline CNV/UPD (present in remission samples), and common population variations were excluded.

Analysis of genomic variants

Somatic variants in non-coding regions and synonymous variants in coding regions were excluded. Variants supported by < 3 reads and/or with $< 10\%$ allele frequency were also excluded except for variants in the following panel of genes: *BRAF*, *CRLF2*, *CSF1R*, *FLT3*, *IL2RB*, *IL7R*, *JAK1*, *JAK2*, *JAK3*, *KRAS*, *NF1*, *NRAS*, *PTPN11*, *SH2B3*, *TYK2*. For this panel, all variants that occur recurrently in cancer according to publicly available data sources (COSMIC, <https://cancer.sanger.ac.uk/cosmic>; PeCan, <https://pecan.stjude.cloud/home>) were included when identified by WES and RNA-seq simultaneously.

Positions of the variants passing these filtering criteria were inspected in Integrative Genomics Viewer (IGV) and further filtered:

- Variants in non-uniquely mapped reads were excluded except for variants in pseudo-autosomal regions on sex chromosomes
- Variants supported by poor quality reads were excluded

If only RNA-seq data were available, the results from variant calling were analyzed for the presence of single nucleotide and frameshift mutations (i.e., variants which do not occur in healthy European population with >0.1% allele frequency according to the ExAC database) in the following panel of genes: *BRAF*, *BTG1*, *CREBBP*, *CRLF2*, *CSF1R*, *ETV6*, *FLT3*, *IKZF1*, *IL2RB*, *IL7R*, *JAK1*, *JAK2*, *JAK3*, *KMT2D*, *KRAS*, *MYC*, *NF1*, *NRAS*, *PAX5*, *PTPN11*, *RUNX1*, *SETD2*, *SH2B3*, *TBL1XR1*, *TP53*, *TYK2*, *ZEB2*.

Identification of the *IGH-DUX4* and *IGH-CRLF2* fusions

To analyze presence of *DUX4*-involving fusion transcripts that could remain unrevealed using TopHat and deFuse, mapped reads were visualized in IGV, and the following reads supporting the presence of *DUX4* gene rearrangements were manually searched at both D4Z4 repeat regions (4q,10q): reads with unmapped parts (“soft-clipped” sequences) that match the *IGH* gene reference (or chromosome region other than 4q and 10q), reads with mates mapped to the *IGH* gene locus (or to chromosome region other than 4q and 10q). The region of a potential fusion partner (identified in the previous step) was then inspected analogously. A similar approach was used to analyze the presence of *IGH-CRLF2* rearrangement.

Differential gene expression analysis

Differential gene expression analysis was performed using R package Deseq2.

Statistics

Unpaired two-tailed t-test, Mann-Whitney test and two-tailed Fisher exact probability test were used for statistical analyses.

Supplementary Tables

Supplementary Table 1. Demographic and clinical characteristics of BCP-ALL patients (stratified into subgroups and subtypes) treated according to AIEOP-BFM ALL 2009 trial protocol

	BCP-ALL non-B-other (n=300) n (%)	Total B-other (n=110) n (%)	B-rest (n=47) n (%)	DUX4r (n=30) n (%)	BCR-ABL1-like (n=16) n (%)	ZNF384r (n=6) n (%)	ETV6-RUNX1-like (n=5) n (%)	iAMP21 (n=4) n (%)	MEF2Dr (n=2) n (%)	P value							
										BCP-ALL non-B-other versus total B-other	B-rest versus DUX4r	B-rest versus BCR-ABL1-like	DUX4r versus BCR-ABL1-like	B-rest versus non-B-rest B-other	DUX4r versus non-DUX4r B-other	BCR-ABL1-like versus non-BCR-ABL1-like B-other	
Sex																	
male	160 (53%)	71 (65%)	36 (77%)	19 (63%)	5 (31%)	6 (100%)	3 (60%)	2 (50%)	0 (0%)								
female	140 (41%)	39 (35%)	11(23%)	11 (37%)	11 (69%)	0 (0%)	2 (40%)	2 (50%)	2 (100%)	0.04	0.3	0.002	0.06	0.03	1	0.004	
Age at diagnosis (years)																	
<10	270 (90%)	76 (69%)	34 (72%)	21 (70%)	10 (63%)	5 (83%)	4 (80%)	1 (25%)	1 (50%)								
≥10	30 (10%)	34 (31%)	13 (28%)	9 (30%)	6 (38%)	1 (17%)	1 (20%)	3 (75%)	1 (50%)	<.0001	1	0.5	0.7	0.5	1	0.5	
Presenting WBC¹ count/μl																	
<50,000	282 (94%)	94 (85%)	38 (81%)	26 (87%)	15 (94%)	5 (83%)	5 (100%)	4 (100%)	1 (50%)								
≥50,000	18 (6%)	16 (15%)	9 (19%)	4 (13%)	1 (6%)	1 (17%)	0 (0%)	0 (0%)	1 (50%)	0.008	0.5	0.3	0.6	0.3	1	0.5	
Prednisone response²																	
good	288 (96%)	102 (93%)	46 (98%)	24 (80%)	15 (94%)	6 (100%)	5 (100%)	4 (100%)	2 (100%)								
poor	11 (4%)	8 (7%)	1 (2%)	6 (20%)	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0.18	0.01	0.4	0.4	0.13	0.005	1	
NA	1	0	0	0	0	0	0	0	0								
FC MRD d15³																	
<10%	272 (92%)	86 (82%)	40 (89%)	19 (68%)	12 (75%)	5 (83%)	5 (100%)	3 (100%)	2 (100%)								
≥10%	23 (8%)	19 (18%)	5 (11%)	9 (32%)	4 (25%)	1 (17%)	0 (0%)	0 (0%)	0 (0%)	0.005	0.04	0.2	0.7	0.1	0.04	0.5	
NA	5	5	2	2	0	0	0	1	0								
PCR MRD⁴																	
TP1 + TP2 neg.	151 (53%)	28 (26%)	16 (35%)	3 (10%)	3 (20%)	0 (0%)	1 (20%)	3 (75%)	2 (100%)								
other	125 (44%)	56 (53%)	24 (52%)	17 (59%)	6 (40%)	4 (80%)	4 (80%)	1 (25%)	0 (0%)								
TP1 ≥10 ⁻³ + TP2 pos. or TP2 ≥10 ⁻³	11 (4%)	22 (21%)	6 (13%)	9 (31%)	6 (40%)	1 (20%)	0 (0%)	0 (0%)	0 (0%)	<.0001*	0.02	0.09	0.5	0.1	0.04	0.2	
NA	13	4	1 ⁵	1 ⁵	1 ⁵	1 ^A	0	0	0								
Final risk group⁵																	
SR	136 (46%)	25 (23%)	15 (33%)	3 (10%)	3 (20%)	0 (0%)	1 (20%)	1 (25%)	2 (100%)								
MR	119 (40%)	52 (49%)	22 (48%)	13 (45%)	6 (40%)	4 (67%)	4 (80%)	3 (75%)	0 (0%)								
HR	40 (14%)	30 (28%)	9 (20%)	13 (45%)	6 (40%)	2 (33%)	0 (0%)	0 (0%)	0 (0%)	<.0001*	0.02	0.3	0.8	0.09	0.03	0.5	
NA	5	3	1 ⁵	1 ⁵	1 ⁵	0	0	0	0								

¹white blood cell; ²good: <1000 leukemic blood blasts/μl on treatment day 8, poor: ≥1000/μl; ³minimal residual disease at day 15 of treatment measured by flow cytometry; ⁴minimal residual disease measured by PCR, TP: time point, TP1: treatment day 33; TP2: treatment day 78; ⁵standard risk (SR), medium risk (MR), high risk (HR), for definitions see Supplementary Methods; P value – p value of Fisher exact test (or of χ^2 test where indicated by asterisk); ⁵patients died before TP2; ^Ano MRD target; neg. – negative; pos. – positive (any positive value); NA – not applicable/not available. Unknown values are not considered for percentage and statistics.

Supplementary Table 10. Early-treatment-response-based risk assignment of B-other ALL stratified by the presence/absence of *PAX5* fusion

	<i>PAX5</i> fusion- positive* (n=12) n (%)	<i>PAX5</i> fusion- negative (n=98) n (%)	P-value
Final risk group¹			
SR	5 (42%)	20 (21%)	0.03
MR	7 (58%)	45 (47%)	
HR	0	30 (32%)	
NA	0	3	

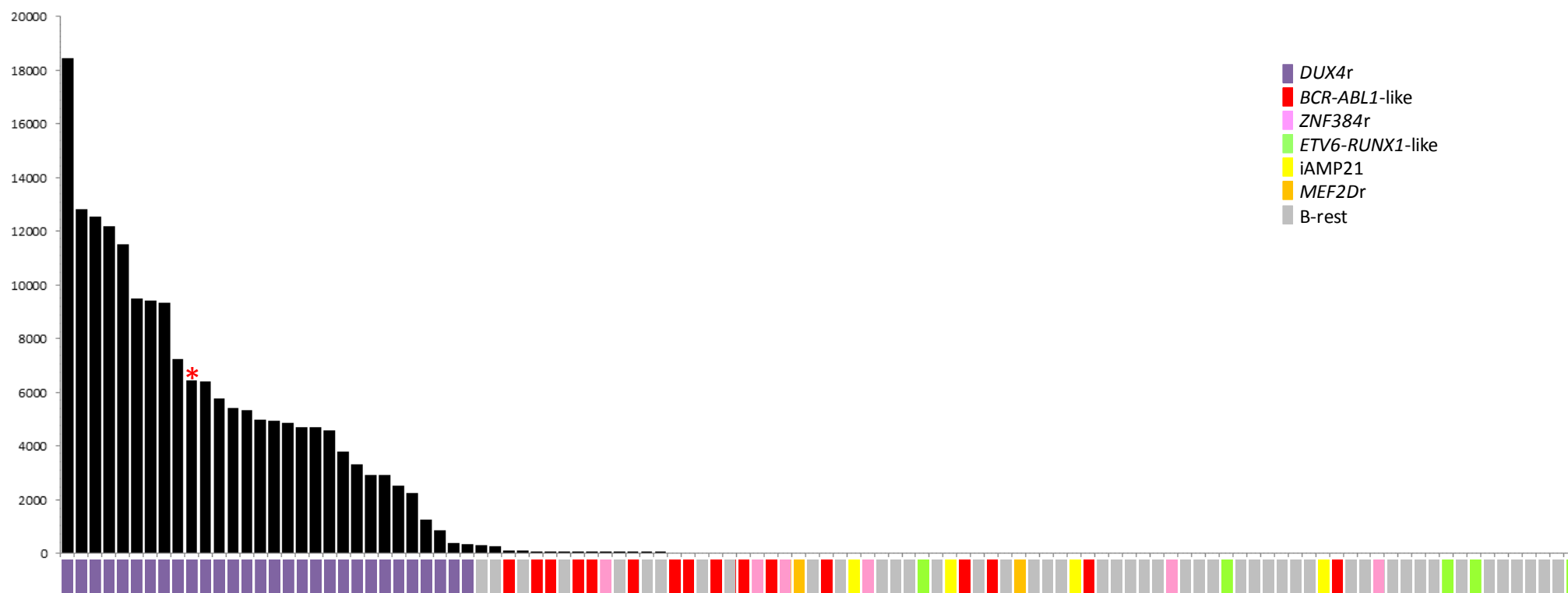
¹standard risk (SR), medium risk (MR), high risk (HR), for definitions see Supplementary Methods; *does not include *ZCCHC7-PAX5* fusion; P value – p value of Fisher exact test; NA – not applicable/not available; unknown values are not considered for percentage and statistics.

Supplementary Figures

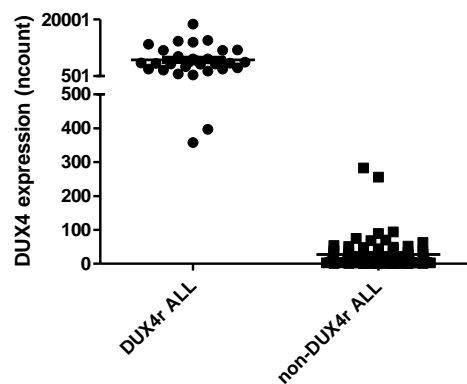
Supplementary Figure 1. The *DUX4* gene expression.

A total of 110 B-other samples were ranked according to the level of the *DUX4* gene expression analyzed by RNA-seq (A) and the *DUX4* gene expression in *DUX4r* ALL was compared to non-*DUX4r* ALL (B). Although *DUX4r* ALL expressed *DUX4* at significantly higher levels compared to non-*DUX4r*, and all individual *DUX4r* ALL samples expressed *DUX4* at higher levels compared to non-*DUX4r*, the difference between *DUX4r* ALL samples with the lowest expression levels and non-*DUX4r* ALL samples with the highest expression levels was low. With a single exception (the patient 2584 indicated by a red asterisk), reads supporting the presence of *IGH-DUX4* fusion were detected by RNA-seq in all patients with *DUX4r* ALL, but not in any patient with non-*DUX4r* ALL. (A,B) y-axis shows normalized *DUX4* counts. SEM – standard error of the mean.

A



B

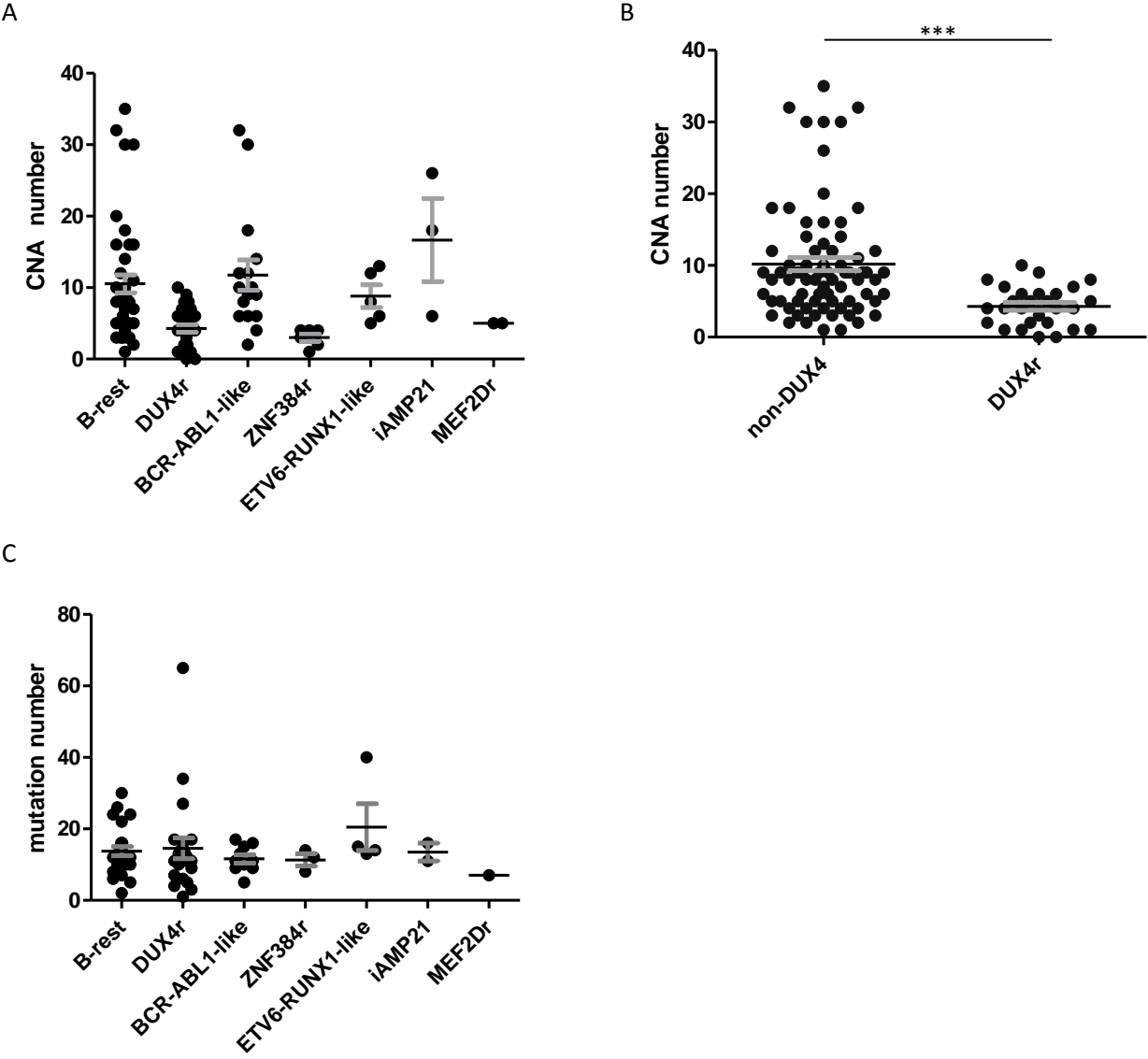


Mean \pm SEM *DUX4r* ALL 6056 \pm 779 (n=30)
 Mean \pm SEM non-*DUX4r* ALL 27 \pm 5 (n=80)

T test: P value < 0.0001

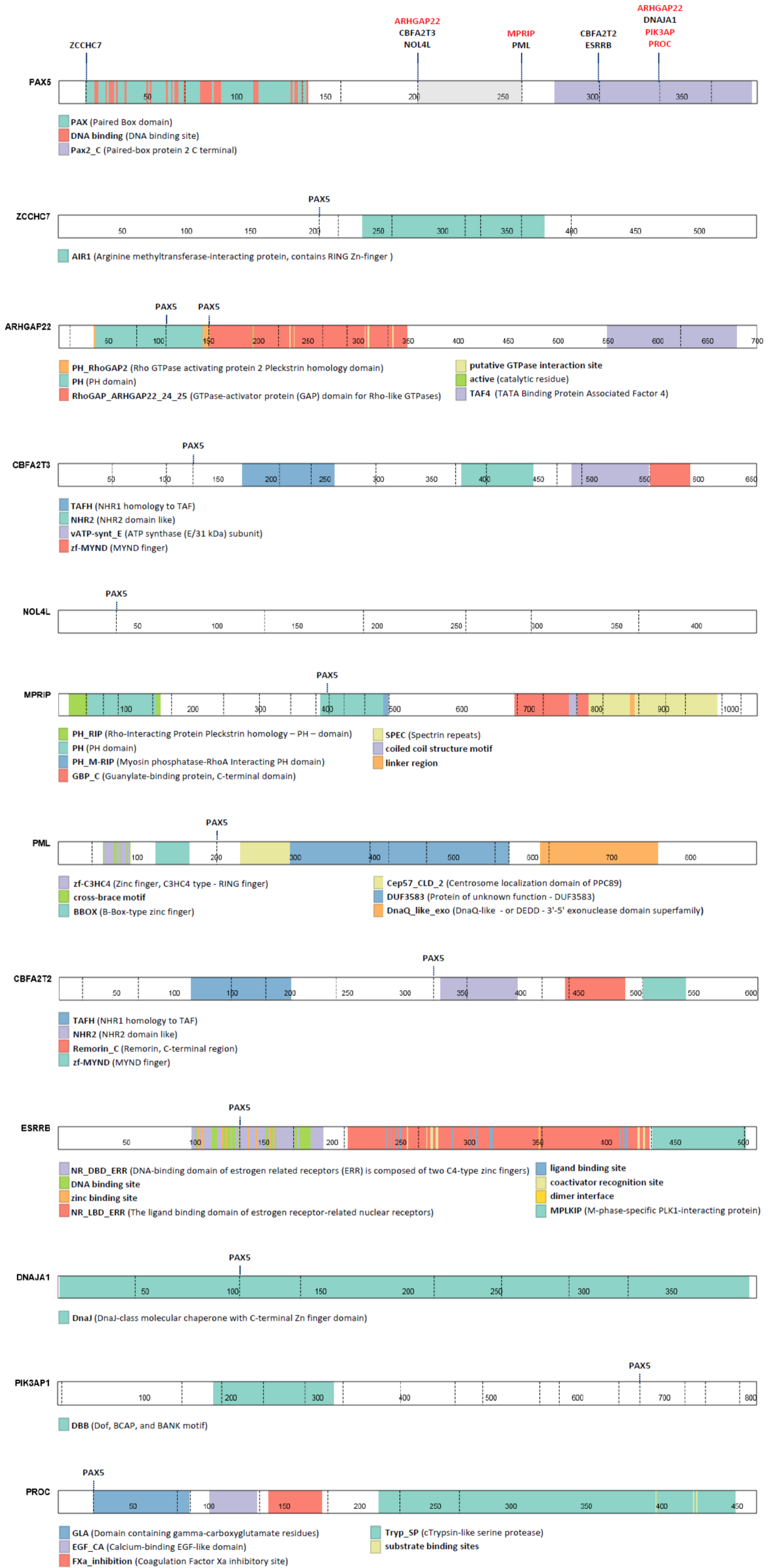
Supplementary Figure 2. Total number of CNAs and SNVs/indels in B-other ALL stratified by subtype.

The total number of copy number aberrations (CNAs; y-axis) in individual patients classified to different ALL subtypes (A), and in *DUX4r* ALL compared to non-*DUX4r* ALL (B). The total number of SNVs/indels (analyzed by WES) in individual patients classified to different ALL subtypes (C). (A-C) Means with their standard errors are shown. *** Mann-Whitney U test $p < 0.0001$.



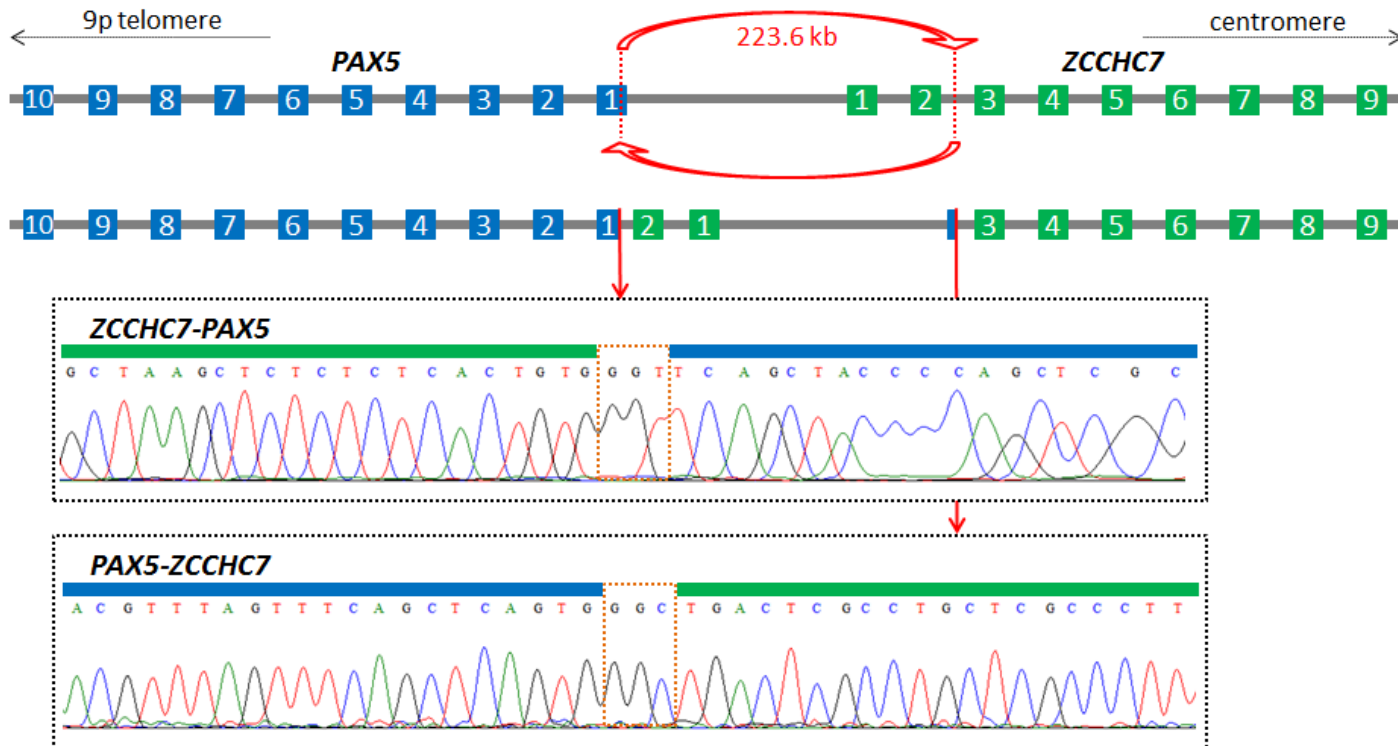
Supplementary Figure 3. Schematic presentation of proteins encoded by *PAX5* and its partner genes identified in the present study.

Schemes of *PAX5* and its fusion partner proteins were created by ProteinPaint (Jinghui Zhang, Ph.D. and Xin Zhou, Ph.D., Nature Gen Dec 29, 2015. doi:10.1038/ng.3466; https://pecan.stjude.org/proteinpaint/). Dashed lines separate regions encoded by individual exons, numbers depict protein lengths (in number of amino-acids). Known protein domains and fusion points identified in present study are shown. Novel fusion partners of *PAX5* are highlighted in red.



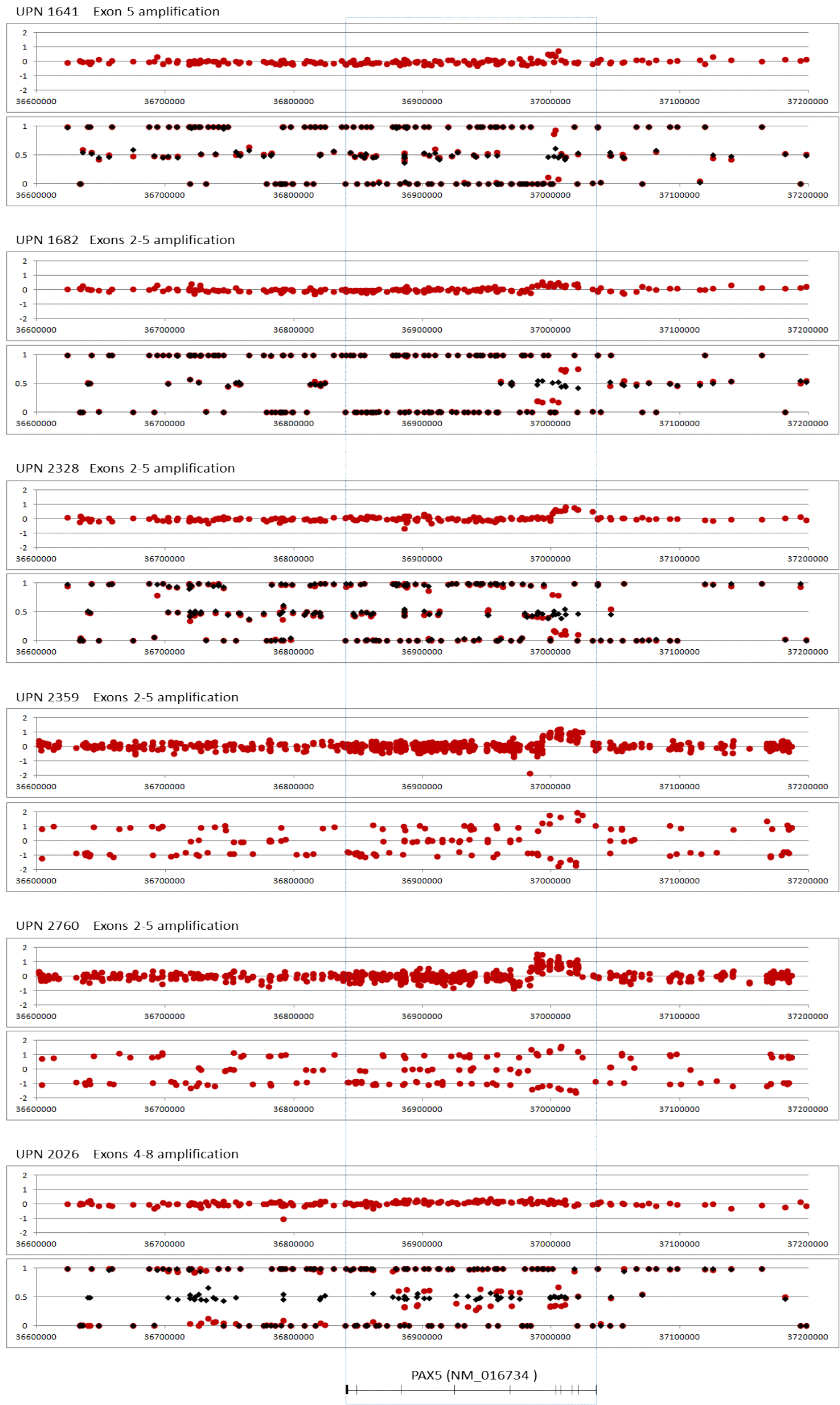
Supplementary Figure 4. The *ZCCHC7-PAX5* and *PAX5-ZCCHC7* fusions resulting from an intrachromosomal inversion within 9p in the patient 2486.

The genomic fusion *ZCCHC7-PAX5* utilizing a telomeric breakpoint in the *PAX5* exon 1 and a centromeric breakpoint in the *ZCCHC7* intron 2 was identified by WES. Both *ZCCHC7-PAX5* and the reciprocal fusion were confirmed by PCR. The upper part of the figure shows schematic structure and localization of partner genes (which are located next to each other on 9p) and the result of inversion; numbered blue/green boxes represent exons, red vertical lines indicate breakpoints. Sequences of junction regions obtained by Sanger sequencing of PCR products are shown in the lower part of the figure. The orange boxes (dashed lines) demarcate 3 and 3 non-templated nucleotides that were inserted at the junction points. A single nucleotide was lost at the breakpoint in *PAX5*, while all nucleotides surrounding *ZCCHC7* breakpoint were preserved.



Supplementary Figure 5. Intragenic *PAX5* amplification detected in 6 patients from B-rest group by SNP-array.

Figure shows the log R ratio (upper panels) and B-allele frequency (lower panels) in the *PAX5* gene region. Data from diagnostic samples are shown in red, B-allele frequency from remission sample (germline genotype) is shown in black. In the patient 2026, the atypical *PAX5* amplification neighbors with pter UPD (whose part can be seen here).

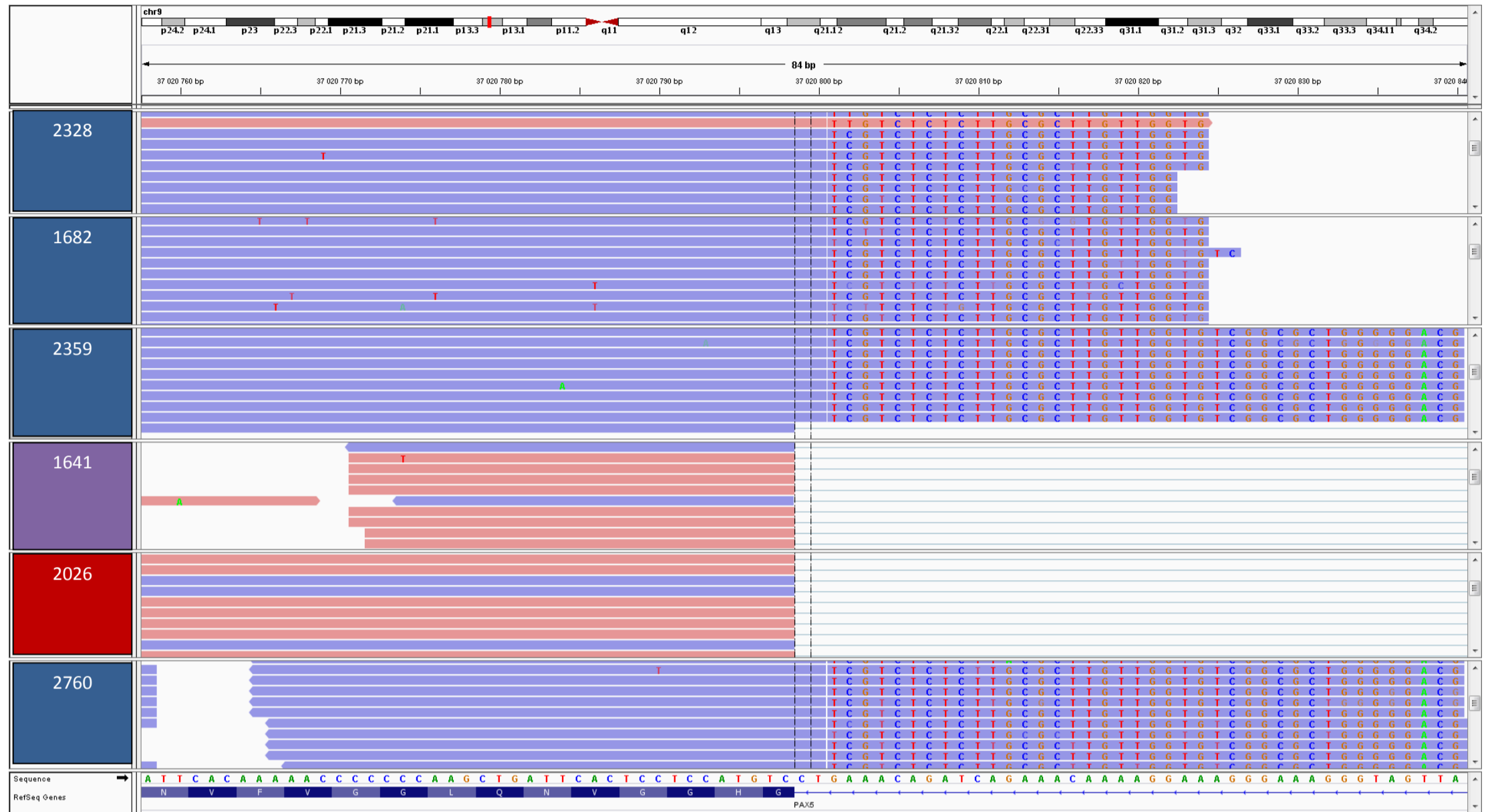


Supplementary Figure 6. Aberrant PAX5 transcripts in the patients with intragenic PAX5 amplification identified by RNA-seq.

The RNA-seq did not allow us to clarify the structure of complete PAX5 transcript(s). However, reads with aberrant exon-exon junctions that suggest a presence of aberrant PAX5 transcripts containing tandemly arrayed “extra” exons from amplified regions were detected by RNA seq in all 6 patients with intragenic PAX5 amplification. (A-E) Snapshots from the Integrative Genome Viewer (IGV) showing reads mapped to selected PAX5 exons. Non-mapped (softclipped) read parts correspond to last/first exon of amplified region. While inclusion of amplified exon set does not theoretically disrupt reading frame in the patients with PAX5^{AMP} (amplification of exons 2-5 or exon 5), the reading frame is disrupted at the exons 8-4 junction in the patient with the amplification of exons 4-8.

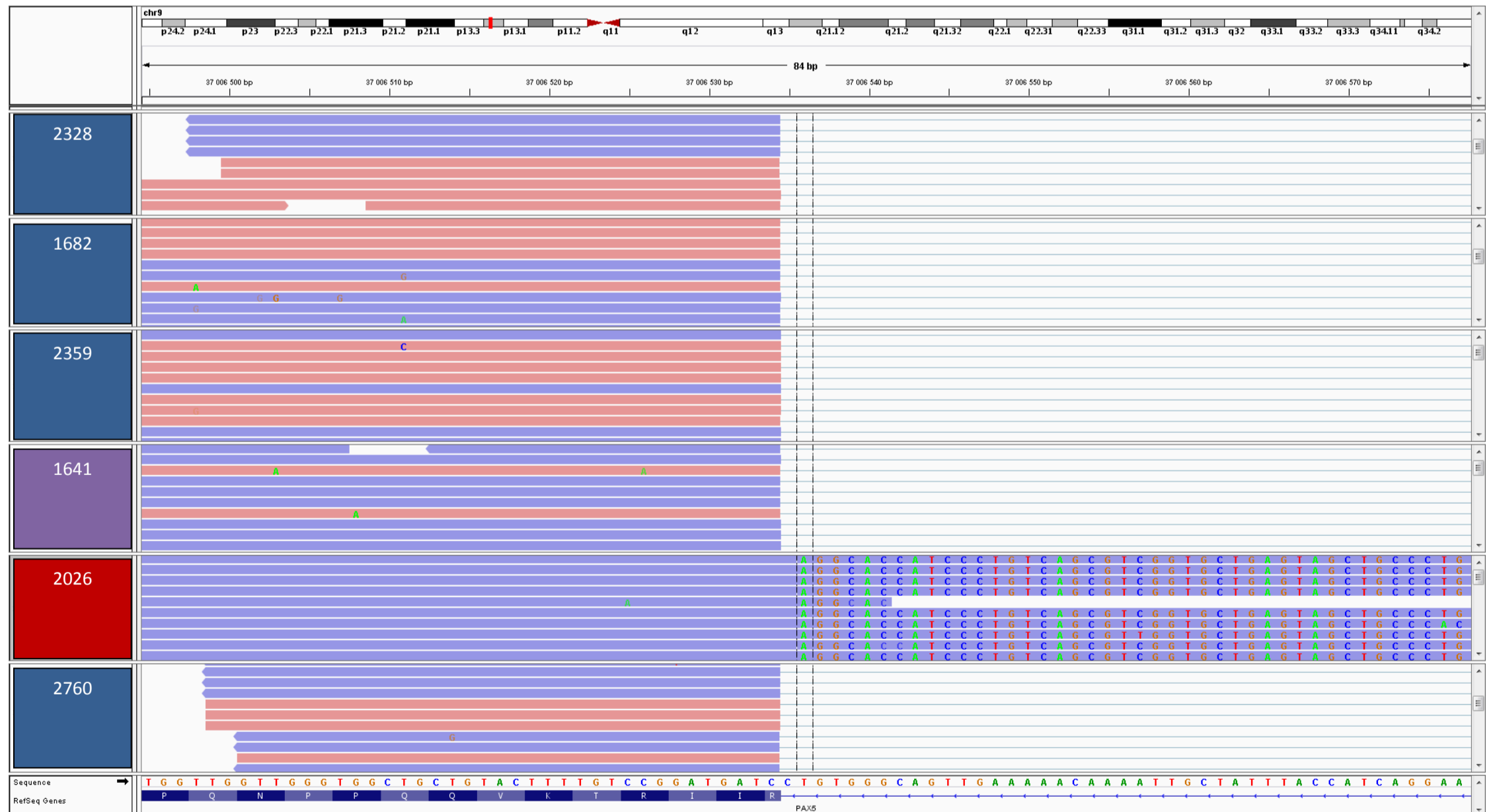
A. IGV snapshot from intron 1 / exon 2 border

Reads spanning exon 5 / exon 2 junction are present in the samples 2328, 1682, 2359, 2760 with PAX5^{AMP} involving exons 2-5. Reads are partially mapped to exon 2, parts of the reads corresponding to exon 5 are softclipped.



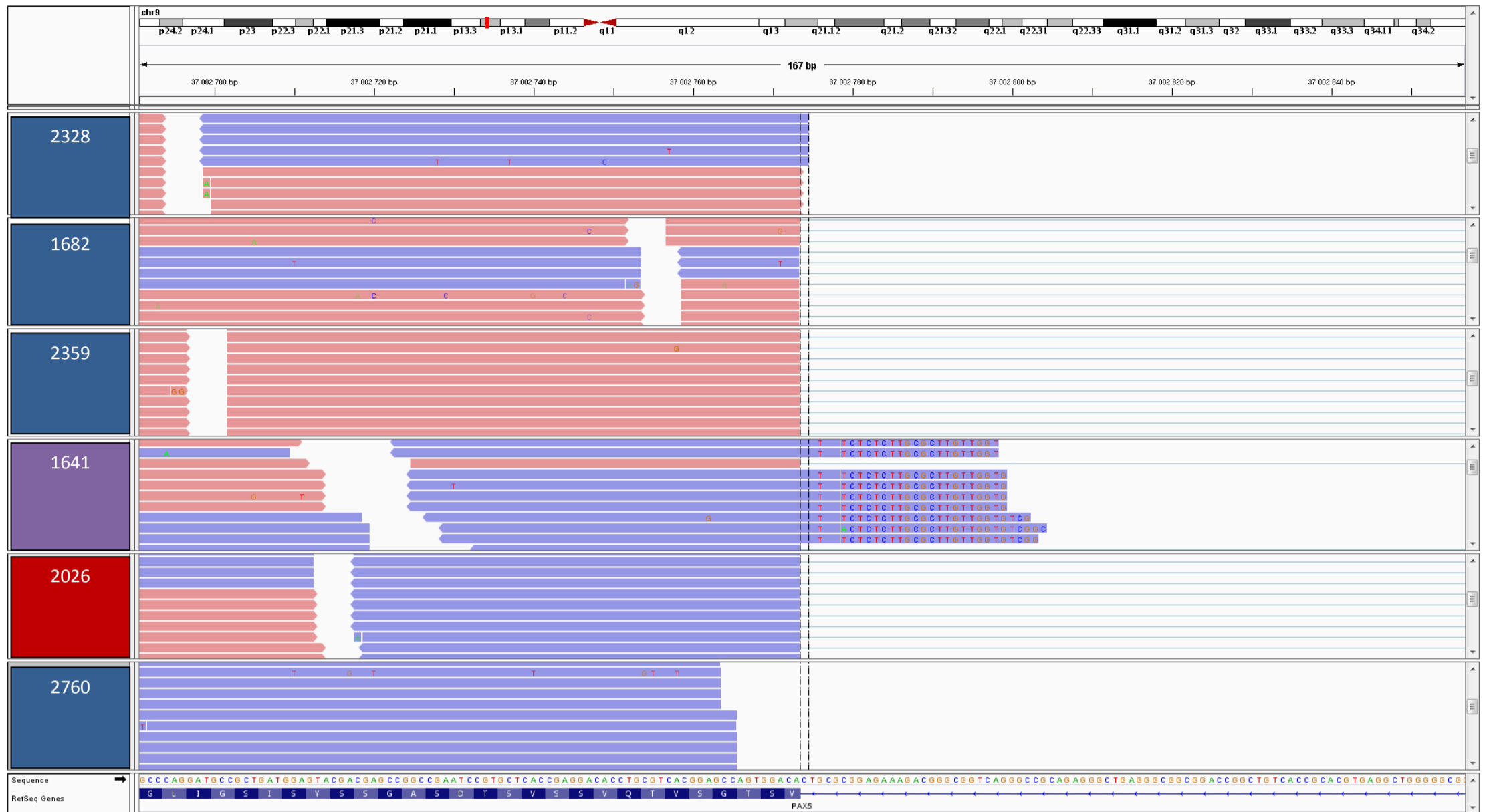
B. IGV snapshot from intron 3 / exon 4 border

Reads spanning exon 8 / exon 4 junction are present in the patient 2026 with amplification of PAX5 exons 4-8. Reads are partially mapped to exon 4, parts of the reads corresponding to exon 8 are softclipped



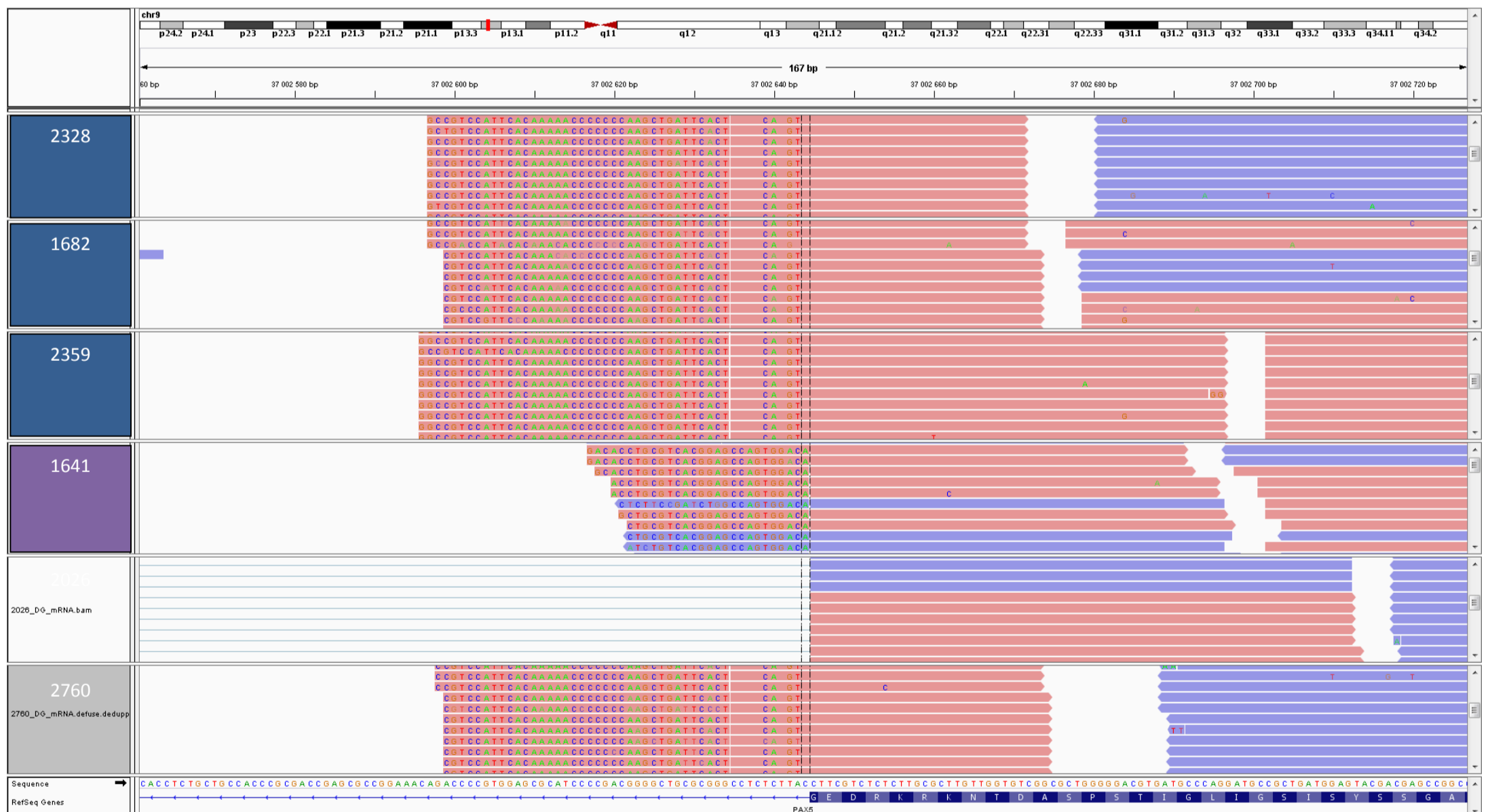
C. IGV snapshot from intron 4 / exon 5 border

Reads spanning exon 5 / exon 5 junction are present in the patient 1641 with *PAX5*^{AMP} involving exon 5. Reads are partially mapped to exon 5, parts of the reads corresponding to extra exon 5 are softclipped.



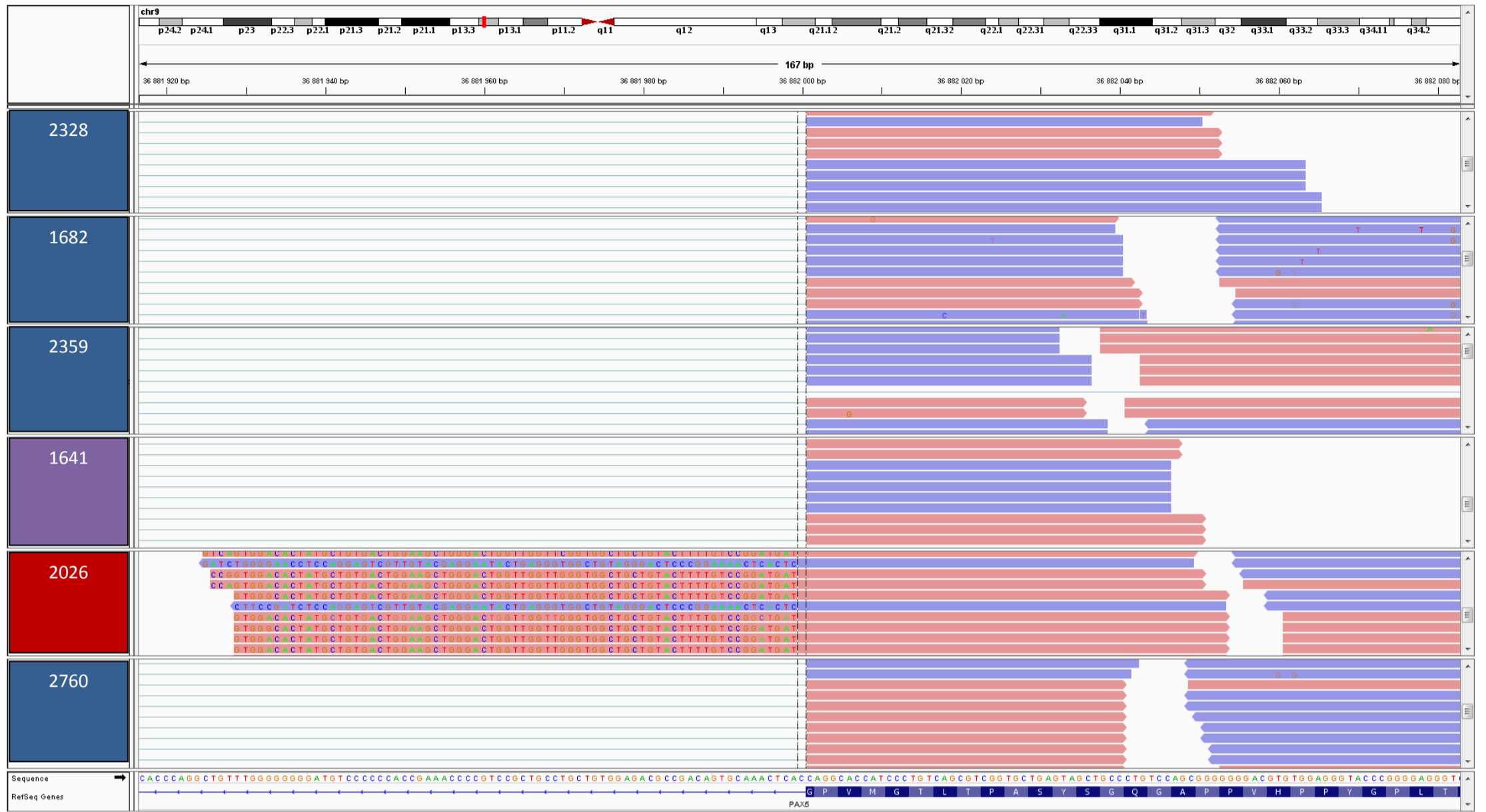
D. IGV snapshot from exon 5 / intron 5 border

Reads spanning exon 5 / exon 2 junction are present in the samples 2328, 1682, 2359, 2760 with *PAX5*^{AMP} involving exons 2-5, while reads spanning exon 5 / exon 5 junction are present in the patient 1641 with *PAX5*^{AMP} involving exon 5. Reads are partially mapped to exon 5, parts of the reads corresponding to exon 2 or extra exon 5 are softclipped.



E. IGV snapshot from exon 8 / intron 8 border

Reads spanning exon 8 / exon 4 junction are present in the patient 2026 with amplification of *PAX5* exons 4-8. Reads are partially mapped to exon 8, parts of the reads corresponding to exon 4 are softclipped.

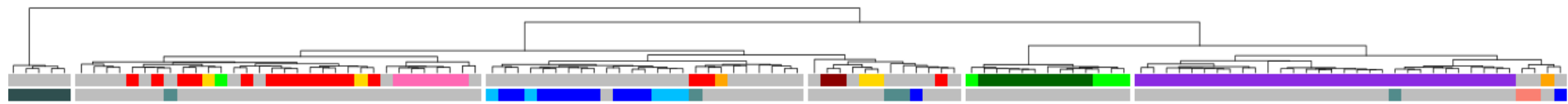


Supplementary Figure 7. Results of HCA analyses.

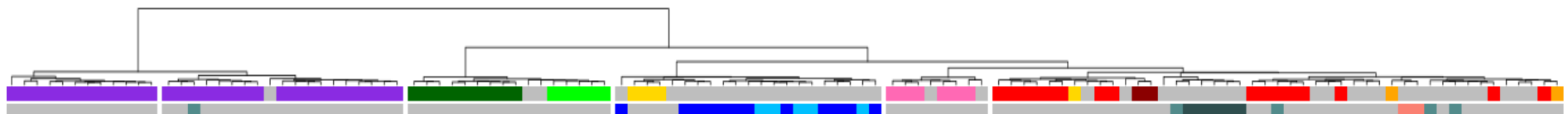
A total of 110 B-other ALLs were clustered hierarchically based on the expression of genes belonging to defined gene sets (A-E) and resulting dendrograms are shown. (A-C) gene sets generated by differential gene expression analysis comparing patients with B-other ALL harboring *PAX5* P80R (A) or *PAX5* fusion (B) or *PAX5*^{AMP} (C) to remaining patients; (D,E) – gene sets from published studies; gene sets are further described in Supplementary Tables 4-5. Genetic annotations are split in two lanes, 1st lane shows classification into previously established subtypes, 2nd lane shows selected recurrent genetic aberrations (mutually exclusive with each other and with the exception of *PAX5* mutations other than P80R also with established subtypes).

* does not include *ZCCHC7-PAX5* fusion.

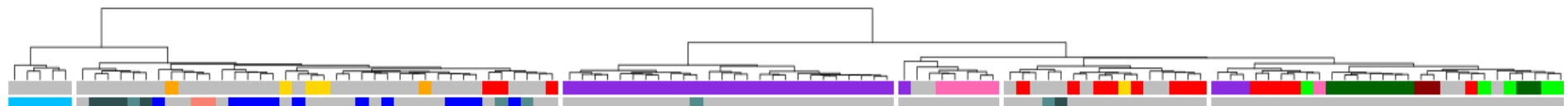
A HCA with *PAX5*-P80R ALL specific gene set



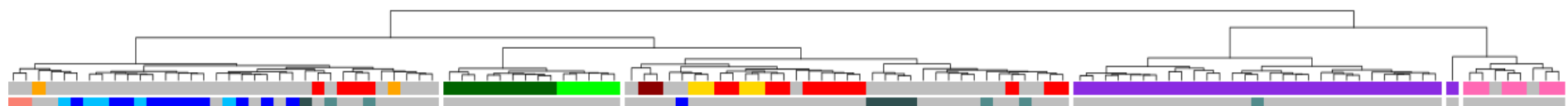
B HCA with *PAX5*-fusion-positive ALL specific gene set



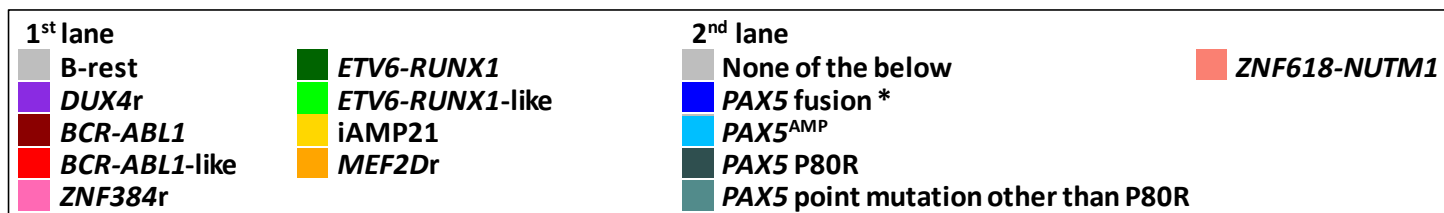
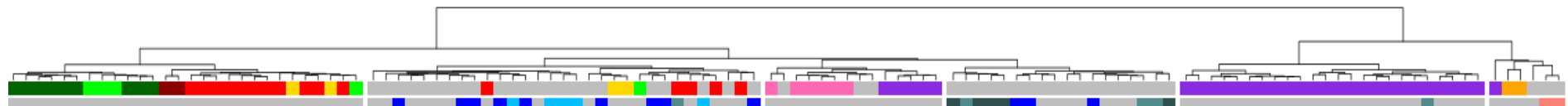
C HCA with *PAX5*^{AMP}-positive ALL specific gene set



D HCA with *ZNF384r* ALL specific gene set



E HCA with *MEF2Dr* ALL specific gene set



Supplementary Figure 8. Expression of *P2RY8-CRLF2* transcript in patients with *P2RY8-CRLF2* fusion.

11/12 patients with *P2RY8-CRLF2* detectable by RNA-seq and indirectly by SNP-array (blue dots), and 10/10 patients with *P2RY8-CRLF2* detectable by RT-PCR but with negative results of RNA-seq and SNP-array (*P2RY8-CRLF2*-low, red dots) were analyzed by qPCR. Y-axis shows relative expression of *P2RY8-CRLF2* (expression of *P2RY8-CRLF2* was normalized to the *GUS* gene expression; for the graphical presentation shown here all expression levels were subsequently normalized to the highest one).

