ARTICLE

Targeted long-read sequencing identifies missing disease-causing variation

Danny E. Miller,[1,2,*] Arvis Sulovari,[1,21] Tianyun Wang,[1,21] Hailey Loucks,[2] Kendra Hoekzema,[1] Katherine M. Munson,[1] Alexandra P. Lewis,[1] Edith P. Almanza Fuerte,[2,22] Catherine R. Paschal,[3,4] Tom Walsh,[1,5] Jenny Thies,[2] James T. Bennett,[2,3,6,7] Ian Glass,[2] Katrina M. Dipple,[2,7,8] Karynne Patterson,[1] Emily S. Bonkowski,[2] Zoe Nelson,[2] Audrey Squire,[2] Megan Sikes,[2] Erika Beckman,[2] Robin L. Bennett,[5] Dawn Earl,[2] Winston Lee,[9,10] Rando Allikmets,[10,11] Seth J. Perlman,[12] Penny Chow,[13] Anne V. Hing,[13] Tara L. Wenger,[2] Margaret P. Adam,[2] Angela Sun,[2,8] Christina Lam,[2,7,14] Irene Chang,[2] Xue Zou,[15] Stephanie L. Austin,[16] Erin Huggins,[16] Alexias Safi,[16] Apoorva K. Iyengar,[17,18] Timothy E. Reddy,[17] William H. Majoros,[17] Andrew S. Allen,[17] Gregory E. Crawford,[16] Priya S. Kishnani,[16] University of Washington Center for Mendelian Genomics, Mary-Claire King,[1,5] Tim Cherry,[6] Jessica X. Chong,[2,7] Michael J. Bamshad,[1,2,7] Deborah A. Nickerson,[1,7] Heather C. Mefford,[2,7,22] Dan Doherty,[2,7,19] and Evan E. Eichler[1,7,20,*]

Summary

Despite widespread clinical genetic testing, many individuals with suspected genetic conditions lack a precise diagnosis, limiting their opportunity to take advantage of state-of-the-art treatments. In some cases, testing reveals difficult-to-evaluate structural differences, candidate variants that do not fully explain the phenotype, single pathogenic variants in recessive disorders, or no variants in genes of interest. Thus, there is a need for better tools to identify a precise genetic diagnosis in individuals when conventional testing approaches have been exhausted. We performed targeted long-read sequencing (T-LRS) using adaptive sampling on the Oxford Nanopore platform on 40 individuals, 10 of whom lacked a complete molecular diagnosis. We computationally targeted up to 151 Mbp of sequence per individual and searched for pathogenic substitutions, structural variants, and methylation differences using a single data source. We detected all genomic aberrations—including single-nucleotide variants, copy number changes, repeat expansions, and methylation differences—identified by prior clinical testing. In 8/8 individuals with complex structural rearrangements, T-LRS enabled more precise resolution of the mutation, leading to changes in clinical management in one case. In ten individuals with suspected Mendelian conditions lacking a precise genetic diagnosis, T-LRS identified pathogenic or likely pathogenic variants in six and variants of uncertain significance in two others. T-LRS accurately identifies pathogenic structural variants, resolves complex rearrangements, and identifies Mendelian variants not detected by other technologies. T-LRS represents an efficient and cost-effective strategy to evaluate high-priority genes and regions or complex clinical testing results.

Introduction

Routine use of genetic testing in clinical and research settings has improved diagnostic rates and uncovered the genetic basis for many rare genetic conditions, yet approximately half of individuals with a suspected Mendelian condition remain undiagnosed.[1–4] Broadly, undiagnosed individuals who have undergone testing by DNA sequencing fall into two main categories: (1 those with a DNA sequence variant or structural difference that does not fully fit their phenotype (i.e., variant of unknown significance) and (2) those in whom routine clinical evaluation—including exome sequencing—failed to reveal any candidate variants or identified only a single variant for a recessive condition that fits the phenotype. Thus, new tools and technologies that provide a comprehensive and accurate survey of genetic variation have the potential to improve diagnostic rates.

[1]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; [2]Department of Pediatrics, Division of Genetic Medicine, University of Washington and Seattle Children's Hospital, Seattle, WA 98105, USA; [3]Department of Laboratories, Seattle Children's Hospital, Seattle, WA 98105, USA; [4]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA 98195, USA; [5]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA; [6]Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA 98101, USA; [7]Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA; [8]Center for Clinical and Translational Research, Seattle Children's Research Institute, Seattle, WA 98101, USA; [9]Department of Genetics and Development, Columbia University, New York, NY 10032, USA; [10]Department of Ophthalmology, Columbia University, New York, NY 10032, USA; [11]Department of Pathology and Cell Biology, Columbia University, New York, NY 10032, USA; [12]Department of Neurology, Seattle Children's Hospital, University of Washington, Seattle, WA 98105, USA; [13]Department of Pediatrics, Division of Craniofacial Medicine, University of Washington, Seattle, WA 98195, USA; [14]Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA 98101, USA; [15]Program in Computational Biology & Bioinformatics, Duke University, Durham, NC 27710, USA; [16]Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC 27708, USA; [17]Department of Biostatistics and Bioinformatics, Duke University; Durham, NC 27708, USA; [18]University Program in Genetics and Genomics, Duke University; Durham, NC 27708, USA; [19]Department of Pediatrics, Division of Developmental Medicine, University of Washington and Seattle Children's Hospital, Seattle, WA 98105, USA; [20]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA
[21]These authors contributed equally
[22]Present address: Center for Pediatric Neurological Disease Research, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
*Correspondence: danny.miller@seattlechildrens.org (D.E.M.), eee@gs.washington.edu (E.E.E.)
https://doi.org/10.1016/j.ajhg.2021.06.006.

Clinical testing methods such as chromosomal microarray (CMA) and exome sequencing do not provide a complete view of human genetic variation. Structural variants (SVs) such as repeat expansions, insertions, deletions, or rearrangements may account for many of the pathogenic variants that go undetected,[5] but they are challenging to identify using existing short-read sequencing technology. Long-read sequencing (LRS) technology, which sequences native DNA molecules, can generate reads from 1,000 to over 1 million base pairs in length while also providing information on DNA methylation.[6] The improved performance of LRS for SV detection has been demonstrated.[5,7–9] However, generating sufficient LRS data for genome-wide analysis remains prohibitively expensive, which makes studies comparing short-read sequencing to long-read sequencing challenging and slows clinical implementation.

Current methods allow LRS of targeted genomic regions using targeted long-read sequencing (T-LRS) either by PCR enrichment or Cas9-mediated isolation of targets.[10–12] However, these methods typically remove critical information such as methylation status, take time to design and optimize, and are restricted to a relatively modest number of genomic targets. To overcome these limitations, we implemented a computational method to select and sequence native DNA using Oxford Nanopore Technologies (ONT). This method, known as adaptive sampling, accepts or rejects DNA molecules for sequencing based on set target sequences and can be modified in real time.[13,14]

We assessed the specificity and sensitivity of T-LRS using adaptive sampling to detect known pathogenic SVs, such as copy number variants (CNVs), repeat expansions, and translocations by sequencing 30 individuals in whom such variants were identified in the course of clinical testing and identified the known variant in all cases (Table S1). These individuals acted as control subjects and allowed us to evaluate whether T-LRS could better characterize previously identified structural changes. In 8/8 persons with complex structural rearrangements, T-LRS enabled more precise resolution of the mutation, which led, in one case, to a change in clinical management. In addition, we sequenced ten persons with a known or suspected autosomal-recessive or X-linked Mendelian condition in whom either only one (n = 8) or no (n = 2) pathogenic variants were found by standard clinical testing. We identified pathogenic or likely pathogenic variants in six and variants of uncertain clinical significance in two of these ten. Our results demonstrate the potential added value of T-LRS as a clinical test to efficiently and cost-effectively evaluate individuals with complex SVs or to identify causal variants in high-priority candidate genes.

Material and methods

Study design

Individuals were identified based on previous clinical or research testing results, which included chromosomal microarray, karyotype, clinical exome sequencing, or research WGS. Individuals with complex copy number changes were defined as those with two or more CNVs or one CNV and at least one translocation. Persons with "missing" variants were defined as those in whom clinical testing had identified one pathogenic variant in a gene associated with an autosomal-recessive disorder or no variants in a gene associated with an X-linked disorder.

DNA isolation and library preparation

DNA for sequencing was isolated from blood, saliva, or fibroblasts using standard methods (Table S1). Extracted DNA was quantified and sheared to a target fragment size of 8–12 kbp using a Covaris g-TUBE. Approximately 1.5 μg of sheared DNA was used to make sequencing libraries using the ONT Ligation Sequencing Kit (SQK-LSK109) following the manufacturer's instructions, except that for each library the short fragment buffer was used during cleanup, and all elutions were done for 10 min at 37°C. All 15 μL of each library was loaded onto a release 9.4.1 flow cell for sequencing on an ONT GridION running MinKNOW control software v18.04.1.

Sequencing and selection of target regions

Target regions were enriched using ReadFish v.0.0.4.[13] In this mode, the software analyzes the signal after a DNA molecule enters a pore to determine whether that molecule lies within a specified genomic region of interest. If it does, the pore continues to sequence the molecule; if not, the DNA molecule is ejected from the pore. In cases with complex CNVs, we targeted large genomic regions on either side of the known aberration. For cases in which a single gene was suspected, at least 100 kbp of DNA surrounding the gene was targeted for sequencing (Table S2). In all cases, standard regions were targeted on multiple chromosomes to serve as internal copy number and coverage controls. ReadFish was run with guppy 3.4.5 and configured to use the dna_r9.4.1_450bp_fast model with min_chunks = 0 and max_chunks = 12. The sequencing_MIN106_DNA file was modified to set break_reads_after_seconds = 0.4. For each experiment, at least 100 kbp and up to several Mbp on either side of the gene or region of interest were targeted (Table S2). Sequencing experiments were run for up to 72 h and, in some cases, a second DNA library was loaded onto the same flow cell after washing at approximately 24 h into a sequencing experiment in order to increase output (Table S1).

Sequence analyses

FASTQ files were generated using guppy 4.0.11 and aligned to GRCh38 using both minimap2 (v.2.17)[15] and NGMLR (v.0.2.7)[16] with default parameters. Variants were called using Longshot (v.0.4.1),[17] Clair (v.4.0.0),[18] and medaka (v.1.2.3). VCF files that combined all variant calls were annotated with variant effect predictor annotations[19] and CADD v.1.6 scores.[20] Novel intronic variants or those with allele frequencies < 2% were annotated using SpliceAI (v.1.3.1).[21] Variants for analysis were filtered based on allele frequency < 2%, CADD score > 15, and SpliceAI prediction > 0.1. If no causative variant was identified with these parameters, the filters were removed, and all variants were manually inspected in the specific gene of interest. Variants were phased using both Longshot and medaka. Copy number changes and breakpoint transitions were identified using circular binary segmentation.[22] SVs were identified using both Sniffles (v.202006)[16] and SVIM

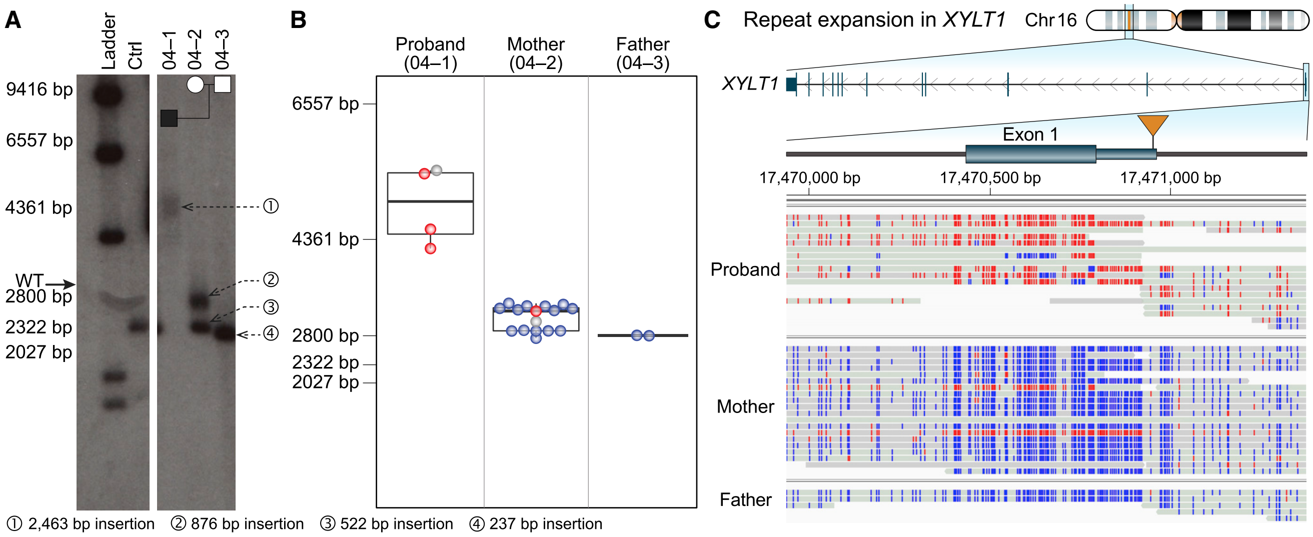


Figure 1. Targeted long-read sequencing simultaneously detects repeat expansion and methylation status
Expansion and methylation of a GGC repeat in the 5′ UTR of *XYLT1* is a common cause of Baratela-Scott syndrome.
(A) Southern blot of family 04 reported by LaCroix and colleagues[26] demonstrates that the proband (04-01) carries an expansion (1) of a region defined by two KpnI restriction enzyme sites containing a GGC repeat, the mother (04-02) carries one premutation (2) and one wild-type allele (3), and the father (04-03) carries two wild-type alleles (4). Both panels are from the same Southern blot on day 6 of exposure.
(B) T-LRS of the trio revealed that the length of fragments from single reads spanning both KpnI cut sites used in (A) was consistent with the results from the Southern blot. Colored dots in (B) correspond to methylated (red) and non-methylated (blue) reads shown in (C); gray represents reads where methylation status was not determined.
(C) Expansion of the GGC repeat in the proband results in methylation of the 5′ UTR and exon 1. Two reads in the mother are methylated (red), one of which spans the region between the KpnI cut sites and whose length is consistent with a premutation allele as shown in (B). The second methylated read terminates within the repeat and the length cannot be assayed.

(v.1.0.1)[23] on both minimap2 and NGMLR alignments. Only those SVs supported by four or more reads within the regions targeted for sequencing were analyzed. For cases in which CpG methylation was assayed, methylation changes were identified in select samples using Nanopolish (v.0.8.4),[24] and BAM files were subsequently converted for visual analysis using Nanopore methylation utilities (commit ece6507).[25]

The complex rearrangements in individuals S014, S020, and S036 were identified by searching the variant files generated by Sniffles and SVIM for SVs that occurred near the deletion breakpoints identified by microarray. We then filtered each file for inversion or translocation events with at least three supporting reads. These events were manually evaluated to ensure that the reconstructed path resulted in a structurally normal chromosome that contained one centromere and two telomeres. Subway plots in Figures 2 and S29 were manually drawn.

PacBio CLR sequencing of family 04

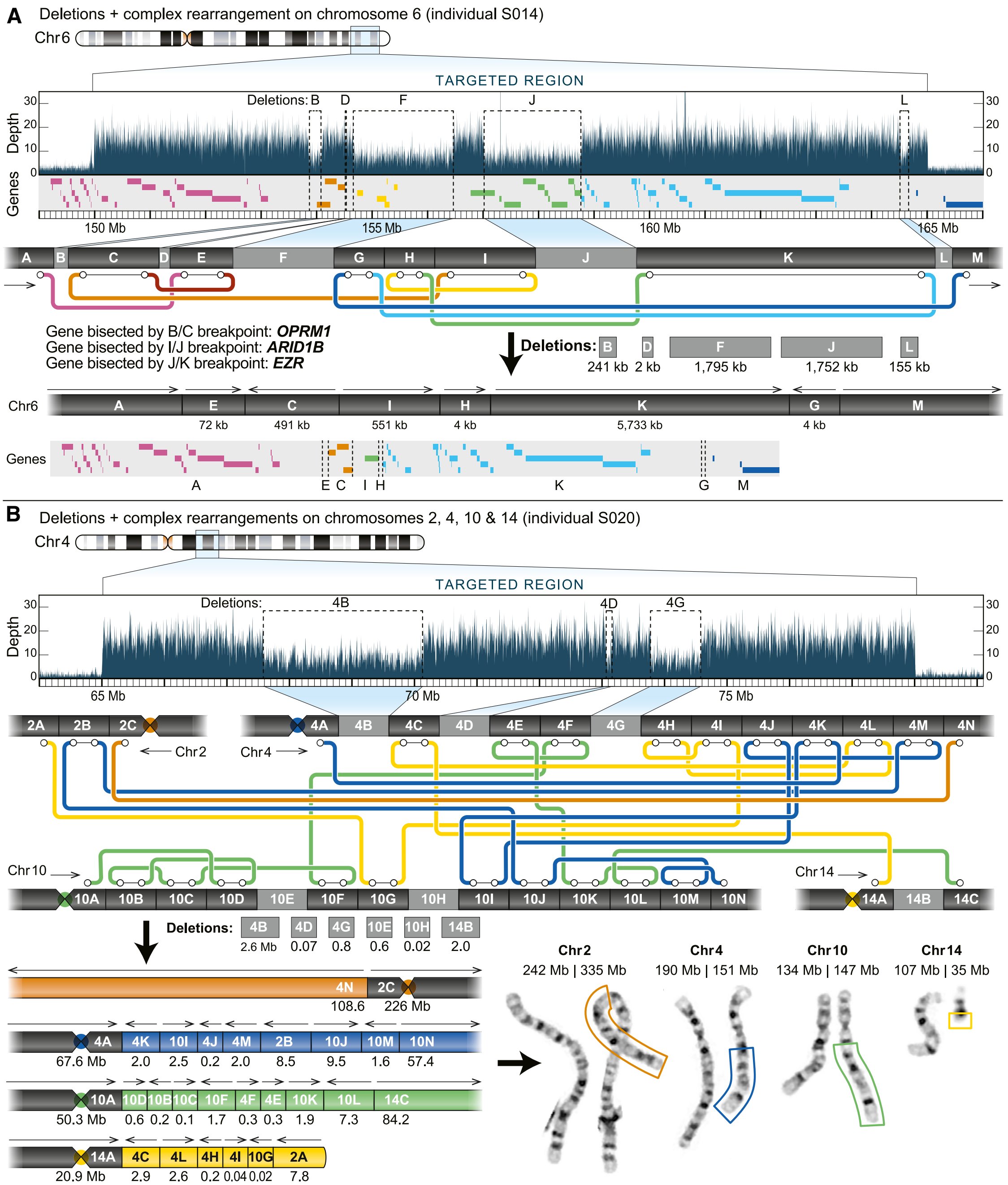
PacBio CLR libraries were generated according to manufacturer's instructions and as described in Chaisson et al.[7] with some modifications. Briefly, high-molecular-weight DNA was sheared using Megaruptor (Diagenode) using the 50 kbp setting. After adaptor ligation with the SMRTbell Express Template Prep Kit, samples were size-selected on a BluePippin instrument using a high-pass cutoff of 35 kbp or 40 kbp, resulting in average library sizes (measured with FEMTO Pulse) of 61 and 72 kbp, respectively. Each library was loaded on three SMRT Cell 1Ms on the Sequel platform using v3 chemistry with 10 h movie times. Final data yield was 32 Gbp Reads of Insert (ROI) (10× coverage) for 38-2 and 38 Gbp ROI (12× coverage) for 38-4, with mean subread lengths of 23 kbp and N50 subread read lengths of 40 kbp.

HiFi sequencing of individual S020 and analysis for additional rearrangement breakpoints

A PacBio HiFi library was generated as in Wenger et al.[27] with the following modifications: high-molecular-weight DNA was sheared using g-TUBE (Covaris) to a mode size of 26 kbp. After adaptor ligation with the SMRTbell Express Template Prep Kit 2.0 and removal of imperfect SMRTbells with the Enzyme Clean Up Kit, the library was size-fractionated on a SageELF platform (Sage Science) using the 1–18 kbp protocol and the fraction's size was measured on a FEMTO Pulse instrument (Agilent) and quantified with the Qubit dsDNA HS (High Sensitivity) Assay Kit (ThermoFisher). A fraction with a roughly 22 kbp average size was sequenced on one SMRT Cell 8M on a Sequel II instrument (PacBio) using v.2.0 bind and sequencing chemistry, with 4 h pre-extension and 30 h movie time. CCS analysis was performed through SMRT Link v9.0 with default settings (3 full passes, estimated quality 0.99) except the maximum read-length cutoff was extended to 100 kbp. Final data yield was 12.2 Gbp of sequence (~4× coverage) with an average length of 21.6 kbp and median estimated quality (Phred scaled) of Q28. Reads were aligned to GRCh38 and SVs were detected as described in Audano et al.[28] We searched for genome-wide translocations or rearrangements missed by T-LRS by filtering out BND variants overlapping a segmental duplication, near a reference gap, or near a contig end. Variants that passed this filter were visually evaluated with IGV and none identified were missed by T-LRS.

Calculation of average read length within and outside of targeted regions

Average read length both genome-wide and within target regions (Table S2) was calculated using a custom script. Briefly,



A
Deletions + complex rearrangement on chromosome 6 (individual S014)
Chr6
TARGETED REGION
Depth
30
20
10
0
Deletions: B
D
F
J
L
Genes
150 Mb
155 Mb
160 Mb
165 Mb
A B C D E F G H I J K L M
Gene bisected by B/C breakpoint: OPRM1
Gene bisected by I/J breakpoint: ARID1B
Gene bisected by J/K breakpoint: EZR
Deletions: B D F J L
241 kb 2 kb 1,795 kb 1,752 kb 155 kb
Chr6
A E C I H K G M
72 kb 491 kb 551 kb 4 kb 5,733 kb 4 kb
Genes
A E C I H K G M
B
Deletions + complex rearrangements on chromosomes 2, 4, 10 & 14 (individual S020)
Chr4
TARGETED REGION
Depth
30
20
10
0
Deletions:
4B
4D
4G
65 Mb
70 Mb
75 Mb
2A 2B 2C
4A 4B 4C 4D 4E 4F 4G 4H 4I 4J 4K 4L 4M 4N
Chr2
Chr4
Chr10
Chr14
10A 10B 10C 10D 10E 10F 10G 10H 10I 10J 10K 10L 10M 10N
14A 14B 14C
Deletions: 4B 4D 4G 10E 10H 14B
2.6 Mb 0.07 0.8 0.6 0.02 2.0
4N 2C
108.6 226 Mb
4A 4K 10I 4J 4M 2B 10J 10M 10N
67.6 Mb 2.0 2.5 0.2 2.0 8.5 9.5 1.6 57.4
10A 10D 10B 10C 10F 4F 4E 10K 10L 14C
50.3 Mb 0.6 0.2 0.1 1.7 0.3 0.3 1.9 7.3 84.2
14A 4C 4L 4H 4I 10G 2A
20.9 Mb 2.9 2.6 0.2 0.04 0.02 7.8
Chr2
242 Mb | 335 Mb
Chr4
190 Mb | 151 Mb
Chr10
134 Mb | 147 Mb
Chr14
107 Mb | 35 Mb

Figure 2. Targeted long-read sequencing identifies additional structural differences not observed by standard clinical testing
(A) T-LRS of individual S014 revealed two additional deletions and one rearrangement (inversion) not reported by CMA. Reanalysis of the CMA data confirmed deletion L. The "subway" plot shows how each region is connected and allows for reconstruction of the new DNA sequence and gene order in the individual.
(B) Clinical CMA of individual S020 identified three deletions on chromosomes 4 and 14 and the subsequent karyotype revealed a complex translocation involving chromosomes 2, 4, 10, and 14. T-LRS identified 11 translocations, 13 rearrangements, and 6 deletions directly affecting 12 genes. Reconstruction of each derivative chromosome estimates the size of each event, as represented by the boxes surrounding part of the derivative chromosomes on the karyotype and is consistent with expected sizes based on karyotype.

the average length of all reads in all FASTQ files from a sample was used to calculate the genome-wide average read length. To calculate the average length of reads within target regions, SAMtools was used to isolate reads that mapped to the target region. Read IDs were then extracted and the length of the read in the FASTQ file was calculated. Each read ID was counted once. Because two flow cells with two different target regions were run for samples S020 and S036, the genome-wide read length was calculated using reads separated by experiment.

Depth-of-coverage calculations within target regions and genome-wide

Coverage of target regions and genome-wide coverage was calculated using SAMtools depth with the -a and -Q 0 flag, which calculates coverage using reads with quality scores of 0 or above.[29]

Refinement of copy number variant breakpoints using binary segmentation

We used sequence depth information from the ONT reads to refine the CNV breakpoints. Specifically, we processed the read-depth information through a binary segmentation, implemented in the R package changepoint.[22] The function cpt.meanvar() was used, which considers both mean and variance of sequencing depth to identify the transition points in the data (i.e., point of sudden increase or decrease in depth). The Bayesian Information Criterion was used to identify the best fit for the optimal regions of distinct depth profiles. This approach helped us refine coordinates for deletions and duplications. All analyses were done using R.3.6.1, and the scripts used for breakpoint refinement are publicly available on GitHub. Results are in Figure S1.

Generation of coverage plots

Data for coverage plots were generated using SAMtools depth with the -a and -Q 0 flags; a custom script then calculated the average coverage in 1 kbp nonoverlapping windows. Plots were generated using average coverage in karyoploteR.[30]

Southern blot of family 04

Southern blot was performed using standard methods as previously described.[26] DNA was digested with KpnI restriction enzyme (New England Biolabs), followed by electrophoresis (0.8% agarose), overnight capillary transfer of the separated DNA fragments via charged nylon membrane (GE Amersham), and crosslinking by exposure to ultraviolet light. The probe (chr16:17,563,659–17,564,191, GRCh37/hg19) was prepared by PCR amplification, cloned into a plasmid, labeled with p32-alpha-dCTP (MegaPrime), and hybridized to the membrane at 6°C overnight. The membrane was washed two times for 15 min each time in 2× SSC, 0.1% SDS and once with 0.2× SSC, 1% SDS at 6°C. Probes were exposed to film for 6 days at −80°C before development.

Estimating the size of reads spanning the KpnI cut sites in family 04

The number of base pairs between KpnI cut sites in family 04 (Figure 1; Tables S4 and S6) was estimated by first determining the genomic position of both KpnI sites by computationally digesting 5 kbp of reference genome using restriction analyzer. This resulted in a 2,589 bp fragment that aligned to chr16:17,468,735–17,471,324 using BLAT (GRCh38 coordinates). A custom script was then used to extract reads from the minimap2 assembly that spanned a 500 or 50 bp interval around the repeat expansion site (Table S4) and to count the total number of nucleotides within that interval by parsing the CIGAR string. All reads spanned the complete interval between the two KpnI sites. The length of the read in the targeted interval was then reduced by the additional target space (either 500 or 50 bp) and 2,589 bases were added to this value, which represented the difference between the length of the interval within the KpnI cut sites and the 1 bp interval that was targeted for counting.

Calculation of repeat lengths

To estimate the size of repeats, we analyzed regions within *FMR1* (MIM: 300805), *ATXN3* (MIM: 607047), and *ATXN8OS* (MIM: 603680) identified as tandem repeats by Tandem Repeats Finder[31] using sensitive parameter settings to maximize tandem repeat discovery despite potential sequence errors in the ONT reads: trf dna_sequence.fa 2 7 7 80 10 20 50 -h -d. For *FXN* we defined the repeat window as the region of the reference genome containing the GAA repeat, and for *XYLT1* (MIM: 608124) we used the position given in LaCroix et al.[26] All targets can be found in Table S4. We used a reference-guided approach to estimating the size of the repeat length. Prior to analysis we re-aligned reads to GRCh38 (without alternative contigs) using minimap2 with the -r 50000, -end-bonus 10000, and–no-end-flt options to optimize the number of reads that spanned the repeats. This reduced the number of reads split by the aligner (Figures S16–S21). A custom script was then used to identify reads that spanned the target region plus a variable number of repeats that depended on the quality of the alignment (given in Table S4). For each read, the CIGAR string was then parsed to determine the length of the sequence that spanned the interval and the length of the additional sequence analyzed was subtracted from the length to get the estimated repeat size. The supplemental alignment of read b0a508ce-069d-43ac-865e-7b7cd900eb70 in sample 04-02 was manually removed, leaving 16 reads remaining for that sample. Repeats were then grouped by their length and the average was calculated (Tables S5 and S6). Reads spanning the interval chrX:32,554,300–32,555,300 were isolated and used to estimate the number of AGAA repeats using Tandem Repeats Finder within *DMD* (MIM: 300377). Nine reads were used to estimate the number of AGAA repeats within the interval; three reads were excluded because the repeat in those reads contained a mix of AGAA and TGTT repeats (Table S17).

Validation of variants

To validate that the splice variant in S004 indeed affected splicing, we assayed for a 50 bp insertion between *NPHP4* (MIM: 607215) exons 5 and 6 with PCR of cDNA from fibroblasts using two primer pairs. The first pair flanks the exon junction (forward: 5′-CTCCTGCACCCGCTTCTC-3′ and reverse: 5′-GGATTCTCCATGA GCTGGAA-3′); the second pair uses the same reverse primer, but the forward primer (5′-CAGCACTCACTGCTCTCGTG-3′) falls within the expected 50 bp insertion of intron 5–6. RNA was extracted using the Aurum Total RNA Kit (Bio-Rad) with a spin-mediated protocol. cDNA was synthesized using iScript cDNA Synthesis Kit (Bio-Rad). PCR was performed on the cDNA using 15 μL 2x Failsafe PCR Buffer J (2X) (Lucigen), 5.6 μL of water, 2.5 μL of 10 μM forward and reverse primer mix, 2 μL of 50 ng/μL template, and 0.4 μL Platinum Taq (5 U/μL) per reaction. A touchdown cycling protocol was used: the first 10 cycles had a variable annealing temperature from 65°C to 56°C, and the next 25 cycles had an annealing temperature of 55°C, for a total of 35 cycles. Extension time was 30 s per cycle. Bands were then excised and extracted using Monarch DNA gel extraction Kit (New England Biolabs). For the first primer pair, two bands were seen and were extracted for separate sequencing. Due to low yield after gel extraction, the extracted bands were run again using the same PCR protocol and primers, then un-purified PCR products or column-purified products (Monarch PCR & DNA Cleanup Kit, New England Biolabs) were submitted to Genewiz with the reverse primer for Sanger sequencing (Figure S33D).

For validation of segregation in S004, PCR was performed using the above protocol with the same temperature and extension time on DNA samples from the proband and parents. Proband DNA was extracted using Gentra Puregene Cell kit (QIAGEN) from fibroblasts, and for parents the prepIT-L2P reagent (DNA Genotek Inc.) was used to extract from saliva. Primers used were forward (5′-TTGAGAACCACTGCTCCAGA-3′) and reverse (5′-ACGAAAC ATCTGCCAAAACC-3′). Unpurified PCR products or column-purified PCR products were submitted to Genewiz for Sanger sequencing using the forward primer, and the splice variant was confirmed to be maternally inherited.

The ~1,900 bp deletion breakpoint in sample S013 was validated by PCR. Briefly, 12.5 μL of Roche FastStart PCR Master mix was combined with 10.5 μL of water, 1 μL of genomic DNA, and 1 μL each of forward (5′-CCCCTTAGAGCAGAAAGGGAC-3′) and reverse (5′-TCATTACCTGACACCCGCAC-3′) primers. PCR was run at an annealing temperature of 55°C for a total of 35 cycles with an extension time of 2 min.

Sanger sequencing was performed to validate the *WDR19* (MIM: 608151) intronic variant. Parental DNA was extracted from saliva using the prepIT-L2P reagent (DNA Genotek Inc.). The same PCR protocol described above for S004 was used, with the forward (5′-CTCCTCCCCATCACCTTTC-3′) and reverse (5′-ACATCCTTGCTT CCTGACCA-3′) primers. The forward primer was used for Sanger sequencing (Genewiz).

Phasing of individual S025 by linkage disequilibrium

Using physical phasing information from the ONT reads that span the 1,450 bp insertion, we determined that a nearby SNV (rs2184339, T>C) had its alternative allele, C, on the same haplotype as reads with the insertion. Using the 1000 Genomes Project SNV genotypes, we calculated linkage disequilibrium between rs2184339 and the missense mutation rs61750120 (G>A) using the R^2 and D′ statistics.[32] Among 2,504 total unrelated individuals representing 26 world populations,[33] we observed no haplotypes containing both the C and A alleles ($D' = 0$, and $R^2 = 0.0002$) (Figure S39). These observations suggest that the insertion allele within intron 1 of *ABCA4* (MIM: 601691) and the missense allele in exon 22 reside on different haplotypes.

Study approval

This study was approved by the institutional review board at the University of Washington under protocols 7064 (University of Washington Repository for Mendelian Disorders), 4125, and 28853. All participants or their legal guardians provided written informed consent. The procedures followed in this study were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and proper informed consent was obtained from all individuals or their guardians.

Results

T-LRS using the adaptive sampling approach allows for rapid selection and real-time discovery of pathogenic variants from candidate genomic regions.[13] We applied this method by direct sequencing of DNA from blood, saliva, or cell lines from 40 affected individuals (and 4 unaffected parents) clinically diagnosed with a variety of genetic conditions (Tables S1 and S2). Among the 40 affected individuals, 14 had a known (i.e., detected by prior genetic testing) pathogenic or suspected pathogenic SV such as a CNV, mobile element insertion, or translocation; 6 had known pathogenic repeat expansions; 8 had known complex rearrangements; 2 had variants identified by clinical testing that could not be phased; and 10 had a clinical diagnosis of an X-linked or recessive Mendelian condition but either no known pathogenic variants (n = 2) or only one known variant out of the expected two (n = 8). All previously known pathogenic SVs were identified in 30 individuals (28 affected individuals and two parents) by T-LRS, including 14 individuals with single or simple CNVs, 8 individuals with multiple CNVs or translocations, 6 individuals with repeat expansions, and 2 parents carrying repeat expansions classified as premutation alleles (Table S1).

Detecting known pathogenic SVs

Fourteen individuals were previously found to have a single pathogenic or suspected pathogenic CNV, translocation, or transposable element insertion detected by CMA, karyotype, short-read sequencing, or long-read sequencing (Table S3). This set includes, for example, frequently observed recurrent deletions or duplications associated with autism and developmental delay (chromosomes 15q11, 16p11, 22q11, and 1q21). We generated 10–62× coverage of the target regions (1–40 Mbp) using a single flow cell for each individual. This sequencing-based approach identified SVs in the expected regions for all 14 persons (Table 1, Figures S1–S15, Table S3). In 5/14 affected individuals, T-LRS provided additional information, including further refining the breakpoint region (n = 4; BK144-03, BK364-03, S046, S060), clarifying the orientation of a duplication (BK364-03, Figure S5), and identifying a previously unknown unbalanced translocation (BK506-03, Figure S10). Evaluation of the underlying genic sequence on the normal homologous chromosomes overlapping the deleted segments found no pathogenic or likely pathogenic variants, consistent with a dominant effect of these SVs.

We also independently identified, in individual S063, an SVA insertion in *BRCA1* (MIM: 113705), originally identified by CRISPR-Cas9 targeting of ~200 kbp regions at tumor suppressor genes followed by long-read sequencing on a PacBio platform.[34] Using our method, we identified the pathogenic SVA insertion after a standard Nanopore ligation library preparation (approximately 2 h) followed by 48 h of sequencing and 1 h of analysis (Figure S15).

Our method also allowed us to identify the precise breakpoints of a translocation that was suspected but not confirmed to be pathogenic. Individual S060 had a clinical diagnosis of campomelic dysplasia (MIM: 114290) and was known to carry a translocation between chromosomes 12 and 17 that was suspected to affect *SOX9* (MIM: 608160). Unfortunately, this could not be confirmed using additional clinical testing such as CMA and single-gene testing. T-LRS of the regions near the known translocation breakpoints at 12q13.3 and 17q25 (35 Mbp total, Table S2) allowed us to identify a translocation breakpoint on

Table 1. Structural variants identified in this study

	Deletions	Duplications	Translocations	Inversions or rearrangements	Total
Events identified by clinical testing	22	15	8	1	46
Events identified by clinical testing and missed by T-LRS	0	0	0	0	0
Events newly identified by T-LRS and not reported on clinical testing	6	0	13	22	41
Total	28	15	21	23	87

Among 14 affected individuals with simple SVs and 8 affected individuals with complex SVs, targeted long-read sequencing detected all 46 structural variants previously identified by clinical testing as well as an additional 41 events not identified by clinical testing.

chromosome 17 located 164 kbp from *SOX9* (Figure S14). While mutations within *SOX9* are the most common cause for campomelic dysplasia, SVs such as translocations and inversions that fall within 1 Mbp of *SOX9* have been associated with it as well, suggesting that the translocation in this individual is the likely pathogenic variant.[35]

Detecting triplet repeat expansions and methylation status

Next, we focused on six persons carrying known repeat expansions associated with spinocerebellar ataxia (MIM: 608768), Friedreich's ataxia (MIM: 229300), fragile X (MIM: 300624), and Baratela-Scott syndromes (MIM: 615777). Repeat expansions in the latter two are, in particular, difficult to detect and resolve using standard sequencing because of the length and high GC content of the repeats. Detecting hyper-expansion and methylation typically require time-consuming Southern blotting with methylation-sensitive enzymes to diagnose.[26,36] We generated a minimum of 8× coverage for all six samples carrying pathogenic expansions in *FMR1*, *FXN* (MIM: 606829), *ATXN3*, *ATXN8OS*, or *XYLT1*. In each sample, we detected pathogenic repeat-sized alleles, and at least one read spanned the complete expansion, providing a more precise estimate on the allele size. We were also able to determine the exact sequence of the expanded allele (Figures S16–S21, Tables S4–S6). In some instances, especially with DNA from cell lines, the length of the expansion was more variable than anticipated. For example, a cell line heterozygous for an expansion within *FXN* was reported to have predominant alleles at 750 and 1,030 repeat units while our sequencing-based estimate identified predominant repeats of 333 and 1,049 repeat units. This finding is consistent with previous work showing repeat length instability in cell lines or somatic mosaicism of expanded alleles.[36]

Expansion of a GGC repeat in the 5′ untranslated region (UTR) of *XYLT1* was recently shown to be a common cause of Baratela-Scott syndrome mediated by methylation and transcriptional silencing (Figure 1A).[26] T-LRS of two affected families from that study (family 04 and 06) allowed us to simultaneously assay repeat length, sequence content, and methylation using a single test (Figures 1B and 1C). Comparing read length and methylation in each individual revealed that some reads for the premutation haplotype in the proband's mother (individual 04-02) were methylated, suggesting that some, but not all, of her cells have silenced the expansion. Thus, T-LRS of native DNA molecules provides additional information not available when repeat length and methylation are assayed separately. Interestingly, methylation analysis in the *FMR1*-expanded CGG repeat obtained from a cell line revealed that the disease locus was no longer methylated despite containing an expansion of nearly 400 repeats (Figure S17). This finding is consistent with a recent observation that methylation status of fragile X full-mutation alleles between 200 to 400 is not stably maintained and, if observed in primary material from an individual, may predict a less severe phenotype.[37]

Phasing of clinically identified variants

We tested whether T-LRS could be used to phase previously identified variants that had exhausted clinical testing options. In an individual (S071) with global developmental delay, clinical trio exome sequencing identified two variants of uncertain significance (VUSs) in *METTL5* (MIM: 618628). These variants were approximately 3.3 kbp apart and could not be phased given one variant was paternally inherited while the other was *de novo*; thus, it was unclear whether both alleles were affected. Using T-LRS we recovered reads spanning both variant positions, allowing us to determine that the variants were in fact in *trans* (Figure S22).

A second individual (S086) with epilepsy was found to have two pathogenic variants in KIAA1109 (MIM: 611565): a maternally inherited deletion and a second *de novo* mosaic missense variant. The variants were approximately 52 kbp apart and clinical exome sequencing suggested that the variant allele fraction of the mosaic variant was 16%. We targeted a 2.1 Mbp region around the gene and recovered approximately 29× coverage of the target region. Medaka was used to phase the variants into two different haplotypes, suggesting that the variants are indeed in *trans* (Figure S23).

Characterization of complex structural rearrangements

To assess the added diagnostic value of T-LRS, we selected eight individuals in whom routine clinical testing using

CMA or karyotype revealed complex structural changes classified as pathogenic, such as multiple noncontiguous CNVs or rearrangements affecting multiple chromosomes. We hypothesized that T-LRS would identify additional rearrangements or CNVs that would be clinically informative. Samples were sequenced by targeting 15–151 Mbp of genomic sequence generating 7–39× coverage of each target region. We identified and refined deletion and duplication breakpoints using a binary segmentation algorithm to delineate transitions in read-depth (Figure S1). Our analysis identified all previously reported events, further refining the rearrangements in four of eight individuals: uncovering a common duplication (S021), refining the breakpoints of a focal amplification (S022), identifying a duplication as tandem (S035), and clarifying the orientation of a terminal deletion and duplication event (S083) (Table 1; Figures S24–S31; Tables S7–S9).

In the four other individuals, we detected additional CNVs, rearrangements, and translocations of potential clinical relevance. For example, in individual S014, a CMA identified three noncontiguous deletions of chromosome 6 spanning a 5 Mbp interval. T-LRS of 15 Mbp surrounding the known deletions revealed two additional deletions and an additional rearrangement not associated with a deletion (Figure 2A; Tables S10 and S11). Thus, the analysis resolved the structure of the region and identified new candidate genes for further consideration, such as *IPCEF1* (MIM: 617476) and *CNKSR3* (MIM: 617476). In individual S082, CMA identified a likely pathogenic deletion on chromosome 10 as well as multiple deletions and duplications on chromosome 17 that included a pathogenic duplication of *RAI1* (MIM: 607642) and *PMP22* (MIM: 601097). We identified the known CNVs using T-LRS and were able to determine that two CNVs on chromosome 17 were associated with inversions, which revealed the complex structure of the chromosome (Figure S30).

We identified more extensive chromosomal differences in two individuals. Clinical testing of individual S020 identified multiple deletions and translocations involving four different chromosomes. To evaluate these differences further, we targeted 74 Mbp of sequence around the known CNVs and obtained approximately 27× coverage of four target regions using one ONT flow cell (Table S2). However, because analysis of these regions indicated rearrangements involving regions outside the targeted area, a second flow cell was run, targeting an additional 77 Mbp of sequence flanking the previously targeted region. In total, we analyzed 151 Mbp of genomic space and identified the precise position of 11 translocations, 13 intrachromosomal rearrangements, and 6 deletions that directly impacted 12 genes, 11 of which were not reported by clinical testing (Figure S25). All 30 of these structural breakpoints were subsequently validated by low-coverage PacBio HiFi whole-genome sequencing (WGS) (Figure 2B; Table S12). Reconstruction of all translocation and rearrangement breakpoints resulted in derivative chromosomes of lengths expected based on karyotype (Figures 2C and 2D). Among the 12 genes disrupted by an SV, two may be associated with autosomal-dominant arrhythmogenic right ventricular dysplasia (*CTNNA3* [MIM: 607667]) and thoracic aortic aneurysms (*PRKG1* [MIM: 176894]) (Table S13). As a result, this individual was referred to cardiology for additional evaluation, which did not reveal any abnormalities, and for anticipatory monitoring for dysrhythmias. Similar to individual S020, clinical testing identified multiple SVs in individual S036. T-LRS identified two additional deletions, five rearrangements, and six translocations not previously detected. In total, these events bisected seven genes, only two of which were reported on prior clinical testing (Figure S29, Tables S14 and S15).

Identifying missing variants in recessive and X-linked Mendelian conditions

We performed T-LRS on ten individuals in whom clinical testing or follow-up research studies revealed only a single variant in a gene associated with a recessive condition (n = 8) or no variants in genes associated with an X-linked condition (n = 2) (Table 2). Each of these individuals had a strongly suspected clinical diagnosis but the molecular diagnosis was missing or incomplete. Using ACMG criteria,[38] T-LRS revealed a pathogenic or likely pathogenic variant in six of ten persons with suspected recessive or X-linked conditions, and a VUS in two of ten; no second candidate variant was found in two others (S004 and S018) (Figure 3; Table 2; Figures S32–S41; Tables S16–S18). The newly discovered variants included deletions, mobile element insertions, inversions, repeat expansions, and intronic variants predicted to affect splicing. In 50% of cases, we generated the data using a single ONT flow cell (Table S1).

Sequencing of two individuals with suspected recessive disorders, S003 (nephronophthisis, *NPHP4*) and S056 (cranioectodermal dysplasia, *WDR19*), revealed that both carried rare intronic variants predicted by SpliceAI[21] to affect splicing located on the opposite haplotype from the known pathogenic variant (Figure 3A). In a fibroblast cell line from S003, we confirmed aberrant splicing by PCR and Sanger sequencing (Figure S27). In S003, we also identified heterozygous intronic GA-rich tandem repeat expansions with both haplotypes fully spanned by at least one long read. Because both expansions are within the range previously observed in control subjects,[39] we were able to exclude them as candidate second hits, which would have been challenging to conclude using short reads alone (Figure S33).

Using T-LRS we identified two deletions missed by previous testing. In an individual with Hermansky-Pudlak syndrome (MIM: 203300) (S013) and a known paternally inherited stop-gain variant, T-LRS revealed a novel 1,900 bp deletion on the maternal haplotype not identified by clinical CMA or exome sequencing. The deletion spanned all of exon 3, resulting in a frameshift and was subsequently clinically validated with an exon-level array (Figure 3B, Figure S37). An individual with glycogen storage disease III (MIM: 232400) (S047) was found by clinical

Table 2. Missing disease-causing variants

Individual (gene)	Inheritance	Prior genetic testing	Known variant identified by T-LRS	Missing variant identified by T-LRS	Category of variant	ACMG criteria met	Confirmation
S002 (*ALMS1*)	AR	SNP, ES, ELA	p.Ser745*	*Alu* insertion in exon 20	P	PVS1, PM3, PP4	clinically confirmed
S003 (*NPHP4*)	AR	SNP, ES, ELA	p.Gln45*	NM_015102.4:c.517+50C>G; splice site variant	P	PS3, PM2, PM3, PP3, PP4	confirmed to affect splicing by RT-qPCR
S004 (*VARS2*)	AR	SNP, ES, ELA	p.Ala420Thr	none identified	–	–	–
S008 (*HPRT1*)	X-linked	karyotype, TS of *HPRT1*	N/A	17 Mbp paracentric inversion bisecting *HPRT1*	P	PVS1	clinically confirmed
S009 (*DMD*)	X-linked	SNP, ELA, TS of *DMD*	N/A	AGAA expansion in intron 16	VUS	PM2	observed in mother and absent in unaffected brother of proband
S013 (*HPS1*)	AR	SNP, ES	p.Arg439*	~1,900 bp deletion that includes exon 3 (first coding exon)	LP	PVS1, PM3	clinically confirmed
S018 (*PAH*)	AR	PKU panel	c.1066−11G>A	none identified	–	–	–
S025 (*ABCA4*)	AR	SNP, ES, research WGS	p.Arg1108Cys	~1,500 bp transposable element insertion in intron 1	VUS	PM3, PP3, PP4	confirmed by reanalysis of short-read WGS
S047 (*AGL*)	AR	TS of *AGL*, research WGS	p.Val426*	1,525 bp deletion including part of exon 3	P	PVS1	confirmed by reanalysis of short-read WGS data
S056 (*WDR19*)	AR	ciliopathy panel, ELA	p.Arg1178Gln	NM_025132.3:c.1250-197C>T; splice site variant	LP	PM2, PM3, PP3, PP4	variant confirmed by PCR

In eight of ten individuals with suspected genetic diseases, T-LRS identified six pathogenic or likely pathogenic disease-causing variants and two variants of uncertain clinical significance not identified by clinical or research testing. Prior testing of individual S009 included a muscle biopsy and immunohistochemistry, which found minimal dystrophin present.

AR, autosomal recessive; SNP, single-nucleotide polymorphism array; ELA, exon-level array; ES, exome sequencing; TS, targeted sequencing; PKU, phenylketonuria; P, pathogenic; LP, likely pathogenic; VUS, variant of uncertain significance; RT-qPCR, reverse transcription quantitative PCR; WGS, whole-genome sequencing.

testing to have a single-nucleotide deletion in *AGL* (MIM: 610860) leading to a frameshift, with no second variant identified after research-based WGS. T-LRS revealed a 1,525 bp deletion that removed part of exon 3 resulting in a frameshift and permitted phasing of both variants onto different haplotypes (Figure S40). Review of the short-read WGS data confirmed the presence of a deletion (Figure S40).

We were also able to identify other types of SVs using T-LRS. In an individual with Alström syndrome (MIM: 203800) and a known paternally inherited stop-gain variant (S002), we identified a novel *Alu* repeat mobile element insertion in exon 20 not identified by clinical exome sequencing, which was confirmed by a clinical laboratory as a pathogenic second hit (Figure 3C, Figure S32). S008 was an individual with biochemically confirmed Lesch-Nyhan syndrome (MIM: 300322) in whom T-LRS identified a 187 bp deletion within intron 3 of *HPRT1* (MIM: 308000), where evaluation of the flanking reads suggested a 17 Mbp paracentric inversion that was clinically confirmed using FISH (Figure 3D, Figure S35). Research-based WGS and targeted sequencing of *ABCA4* and locus in S025, an individual with Stargardt disease (MIM: 248200), failed to identify a 1,500 bp composite retrotransposable element insertion consisting of *Alu*J (SINE) and partial L2a, L2c, L2d2, and L1HS (LINEs) mapping within the first intron of *ABCA4*. We identified the event using both SV callers applied in this study and found that it mapped to a different haplotype than the known pathogenic variant. We categorized this as a VUS; however, consistent with previous work on similar insertions, *in silico* analysis with SpliceAI strongly suggests the insertion results in aberrant splicing of the first exon of *ABCA4* (Figure 3E, Figure S25, Table S18).

Finally, we used T-LRS to evaluate *DMD* in a family with multiple individuals affected by X-linked Duchenne

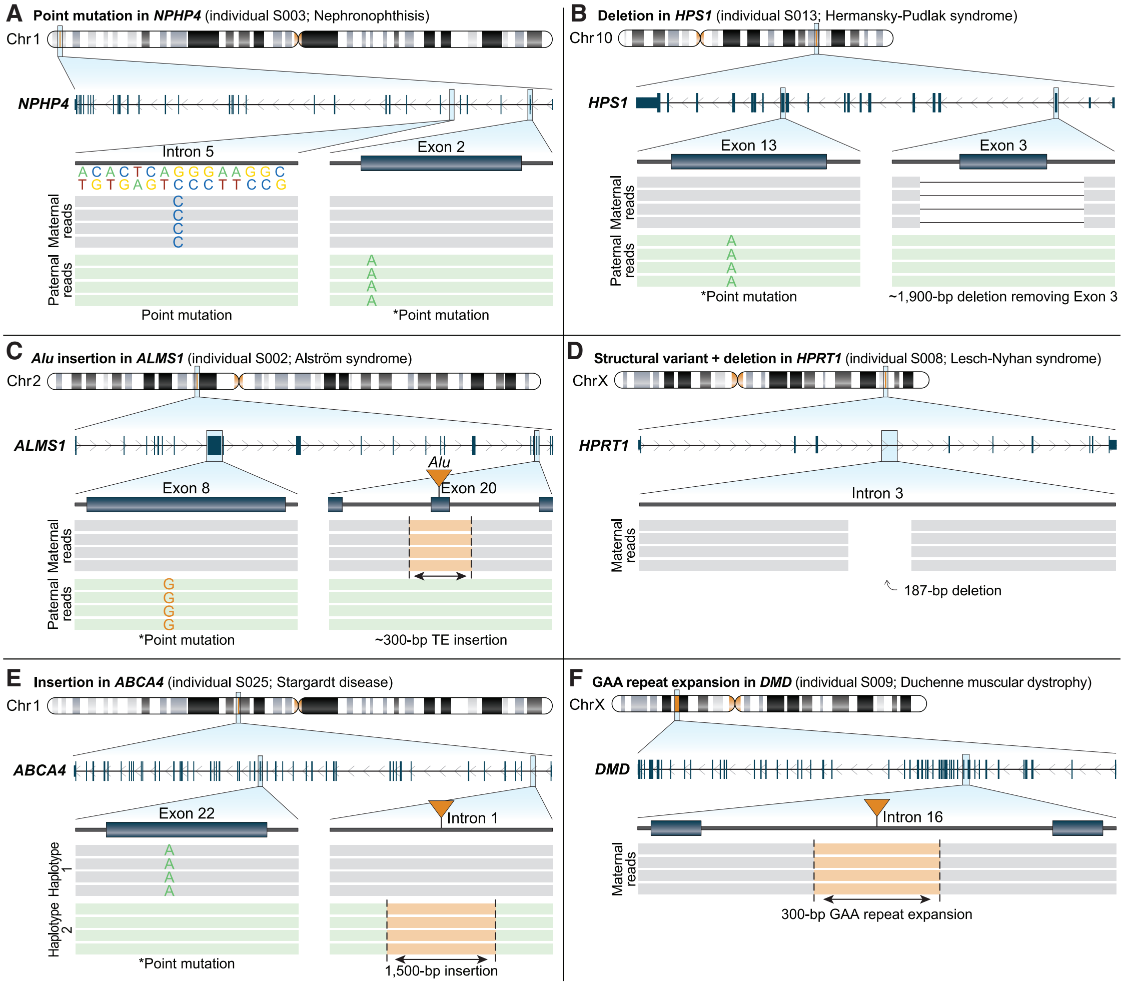

A Point mutation in NPHP4 (individual S003; Nephronophthisis)
Chr1
NPHP4
Intron 5
Exon 2
ACACTCAGGGAAGGC
TGTGAGTCCCTTCCG
C
A
Maternal reads
Paternal reads
Point mutation
*Point mutation
B Deletion in HPS1 (individual S013; Hermansky-Pudlak syndrome)
Chr10
HPS1
Exon 13
Exon 3
A
Maternal reads
Paternal reads
*Point mutation
~1,900-bp deletion removing Exon 3
C Alu insertion in ALMS1 (individual S002; Alström syndrome)
Chr2
ALMS1
Alu
Exon 8
Exon 20
G
Maternal reads
Paternal reads
*Point mutation
~300-bp TE insertion
D Structural variant + deletion in HPRT1 (individual S008; Lesch-Nyhan syndrome)
ChrX
HPRT1
Intron 3
Maternal reads
187-bp deletion
E Insertion in ABCA4 (individual S025; Stargardt disease)
Chr1
ABCA4
Exon 22
Intron 1
A
Haplotype 1
Haplotype 2
*Point mutation
1,500-bp insertion
F GAA repeat expansion in DMD (individual S009; Duchenne muscular dystrophy)
ChrX
DMD
Intron 16
Maternal reads
300-bp GAA repeat expansion

Figure 3. Targeted long-read sequencing identifies variants not detected by clinical testing
Pathogenic, likely pathogenic, or variants of uncertain significance (VUSs) identified by T-LRS along with variants identified by prior clinical testing (denoted by an asterisk).
(A) T-LRS detected a candidate intronic splice acceptor variant as well as the known paternally inherited stop-gain. Long-read phasing demonstrates that these variants are in *trans*.
(B) A 1,900 bp deletion within *HPS1* removes exon 3; phasing revealed that this variant and the previously known paternally inherited stop-gain occur on different haplotypes. Clinical testing with an exon-level array confirmed the deletion.
(C) T-LRS reveals a previously known paternally inherited stop-gain as well as a novel *Alu* insertion in exon 20 of *ALMS1*. Subsequent clinical testing confirmed the *Alu* was pathogenic and maternally inherited.
(D) A 187 bp deletion and 17 Mbp inversion disrupts *HPRT1*. Clinical testing confirmed the presence of an inversion.
(E) Insertion of a 1,500 bp composite retrotransposable element is predicted to create multiple splice acceptor and donor sites and represents a candidate second hit. Linkage disequilibrium phasing suggests the variants are on different haplotypes.
(F) Expansion of an AGAA repeat within *DMD* represents a VUS in an individual with Duchenne muscular dystrophy and a family history lacking a genetic diagnosis.

muscular dystrophy (MIM: 310200) lacking a precise genetic diagnosis. T-LRS of *DMD* in the proband (S009) revealed no candidate single-nucleotide variants (SNVs) but did reveal an intronic 117 AGAA repeat expansion (Figure 3F). The proband's mother was heterozygous for this expansion but it was not found in his unaffected older brother (Figure S36). To determine the frequency of this expansion in a population sample, we analyzed nearly 9,000 short-read genomes[40] using ExpansionHunter,[41] identifying 72 individuals with 117 AGAA repeats or longer for an estimated population allele frequency of 0.4%. Remarkably, 71 (98.6%) of the individuals with large alleles are female—an observation inconsistent with Hardy-Weinberg equilibrium (OR = 52, p = 3e−16, Fisher's exact test). Based on this information, we categorize this expansion as a high-priority VUS for future research investigation.

Discussion

Here, we show that T-LRS using adaptive sampling on the ONT platform can be used for phasing and detection of clinically relevant variants, such as SNVs, CNVs, repeat expansions, and methylation differences. Because target regions are computationally defined for sequencing, this technique is flexible and can be used to interrogate any part of the genome without the need to design specific experimental assays. Drawbacks do exist, such as the need to shear DNA prior to sequencing to increase coverage and the limited ability to assay for mosaic variants compared to exome sequencing. In addition, the analysis of complex structural changes is challenging to fully automate, which may limit its adoption by clinical laboratories, although methods are being developed to both systemically call SVs from long-read phased genome assemblies and merge them to better define their precise breakpoints.[28,42] Regardless, T-LRS removes a substantial barrier to widespread clinical use of long-read technology by reducing per-sample costs of sequencing selected genes or regions to a price point comparable to short-read WGS. When reagents are purchased at scale, the per-sample materials cost of T-LRS is approximately $650 USD when a single ONT flow cell and library is used. Current materials costs for short-read WGS can vary significantly from institution to institution but, on average, are likely around $1,000 USD. The immediate potential clinical uses of T-LRS include screening of candidate genes in which existing technologies have failed to provide a precise genetic diagnosis, refinement of isolated or complex structural breakpoints, phasing of known variants, and evaluation of repeat structure.

Clinical evaluation of SVs typically ends after identification of a single pathogenic CNV or a complex series of both CNVs and rearrangements. Here, we demonstrate that among 22 individuals with known simple or complex SVs, clinical testing identified only 53% (46/87) of the SVs found by T-LRS (Table 1). Additional SVs were recovered in 27% of persons (6/22) and in two persons this information revealed 16 additional genes directly disrupted by an SV. In one individual, the discovery of additional affected genes associated with dysrhythmia and aortic dilation resulted in further clinical evaluation and establishment of a surveillance plan. Detailed understanding of these events also provides key information for understanding the mechanisms behind their formation.[43]

Our understanding of the normal SV spectrum is only beginning to emerge from population-based LRS of individuals without a known condition.[7,28,44] As a result, the pathogenicity of many variants remain uncertain. For example, in case S009 with X-linked Duchenne muscular dystrophy, the intronic AGAA repeat expansion is not only rare in population samples but also found almost exclusively in females. Whether this expansion perturbs the function of *DMD*, perhaps by blocking transcript elongation,[45] acting as a novel transcription factor binding site,[46] or inducing cellular death through a process such as RAN translation,[47] remains to be determined. However, its low prevalence in males makes it a compelling candidate for further evaluation, and if determined to be pathogenic, a potential target for therapeutic intervention.[48] We anticipate that more widespread application of T-LRS will lead to discovery of many more SVs of unknown significance. Assessment of pathogenicity of these variants will benefit from greater public sharing of SVs (e.g., establishment of a database, development of robust mechanisms for matching, etc.), as has been the case for SNVs and indels discovered by short-read exomes and genomes.[49,50] The availability of haplotype-resolved genomes[42] and improvements in reference genomes, such as those made possible by complete telomere-to-telomere assemblies of human chromosomes[51] as well as the characterization of thousands of human genomes as part of initiatives such as All of Us,[52] will also help with characterization of potentially pathogenic SVs identified by clinical and research testing.

In our cohort of individuals with a clinical diagnosis of a recessive or X-linked condition, in whom a single variant or no candidate variants were identified by prior clinical or research testing, T-LRS revealed a pathogenic variant, likely pathogenic variant, or VUS in 80% of affected individuals. Among the eight affected individuals in whom a second hit was identified, two had undergone research WGS that did not identify the causative variants because of filtering of data that reduced the sensitivity of the analysis. Identifying SVs in short-read sequencing data is an active area of research and challenges are well known.[53] While short-read WGS technology may have revealed the candidate second variant in 7/8 affected individuals, our results suggest that T-LRS may be a better next step after clinical genetic testing when a candidate locus of interest is known and has increased sensitivity to detect SVs over short-read WGS in these cases. While large-scale, prospective studies of varied populations will be required to fully assess the advantages of T-LRS over conventional testing strategies, we anticipate that T-LRS may be used to increase the diagnostic rate for Mendelian conditions. Indeed, given that short-read WGS results in only a small increase in the diagnostic rate of unsolved conditions, T-LRS could be a more sensitive and cost-effective approach to screening candidate genes or regions for disease-causing variants in high-priority regions.[54] Additional studies will be needed to understand the sensitivity of T-LRS compared to either short- or long-read WGS in syndromic cases with negative clinical testing that are known to be associated with multiple genes. Individual evaluation of cases with nondiagnostic T-LRS will determine the next best evaluation, which could include either short- or long-read WGS or RNA studies.[3,55]

We predict eventual implementation of whole-genome LRS (WG-LRS) will have a major impact on clinical genetic testing, because as a single test WG-LRS has the potential to replace nearly every other genetic test currently offered, excepting perhaps analysis by karyotype.[56] For example, in a person suspected to have a Mendelian condition, WG-LRS data could first be used to evaluate sequence variants within a specific gene or genes. If no explanatory variant was found, the same dataset could reflexively be analyzed to interrogate sequence variants in all exons and high-priority noncoding regulatory regions, as well as search genome-wide for SVs and mutated repetitive elements. This testing strategy would replace the often-used stratified approach to testing (i.e., single gene testing, CMA, followed by exome sequencing). Moreover, these steps are computationally applied to the same LRS data, so such a stepwise analysis could be completed in hours or days compared to weeks to months for conventional stratified testing strategies. Clinical adoption of T-LRS or WG-LRS is likely to increase the diagnostic rate, reduce the cost, and shorten the time to diagnosis for families with rare genetic conditions.

Data and code availability

Data generated in this project will be available at dbGaP accession number phs000693. Analysis scripts used in this study are available on GitHub at https://github.com/danrdanny/targetedLongReadSequencing. Please see Table S19 or contact corresponding author D.E.M. for information regarding data not available under this accession number.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.06.006.

Acknowldgements

We are grateful to the individuals and their families who participated in this study. We thank Angela Miller for figure preparation, Tonia Brown for assistance in editing this manuscript, and Sunday Stray and Lemlen Ghile from the laboratory of Mary-Claire King for isolation of DNA from whole blood for the SAGE BK samples. This work was supported by a grant to D.E.M. and E.E.E. from the Brotman Baty Institute for Precision Medicine. This work was also supported, in part, by grants from the US National Institutes of Health (NIH R01 MH101221 to E.E.E.) and the Simons Foundation (SFARI 608045 to E.E.E.). J.T.B. is supported by NIH R01 HL130996-05 and BWF CAMS #1014700. R.A. and T.C. are supported by NIH R01 EY29315. Sequencing and analysis were supported in part by the University of Washington Center for Mendelian Genomics (UW-CMG) and funded by NHGRI and NHLBI grants UM1 HG006493 and U24 HG008956. H.L. and D.D. are supported by NIH R01HD100730 and U54HD083091. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. M.-C.K. is an American Cancer Society Research Professor. E.E.E. is an investigator of the Howard Hughes Medical Institute.

Declaration of interests

P.S.K. reports receiving consulting fees from Sanofi Genzyme, Amicus Therapeutics, Maze Therapeutics, JCR Pharmaceutical, and Asklepios Biopharmaceutical, Inc; research and grant support from Sanofi Genzyme, Valerion Therapeutics, and Amicus Therapeutics; has equity in Asklepios Biopharmaceutical, Inc., and Maze Therapeutics; and is a member of the Pompe and Gaucher Disease Registry Advisory Board for Sanofi Genzyme, Amicus Therapeutics, and Baebies.

Received: April 6, 2021
Accepted: June 7, 2021
Published: July 2, 2021

Web resources

dbGaP, https://www.ncbi.nlm.nih.gov/gap
Medaka, https://github.com/nanoporetech/medaka
OMIM, https://www.omim.org/
Restriction analyzer, https://www.molbiotools.com/restrictionanalyzer.html
Scripts used in this study, available on GitHub, https://github.com/danrdanny/targetedLongReadSequencing

References

1. Lowther, C., Valkanas, E., Giordano, J.L., Wang, H.Z., Currall, B.B., O'Keefe, K., Collins, R.L., Zhao, X., Austin-Tse, C.A., Evangelista, E., et al. (2020). Systematic evaluation of genome sequencing as a first-tier diagnostic test for prenatal and pediatric disorders. bioRxiv. https://doi.org/10.1101/531210.
2. Boycott, K.M., Rath, A., Chong, J.X., Hartley, T., Alkuraya, F.S., Baynam, G., Brookes, A.J., Brudno, M., Carracedo, A., den Dunnen, J.T., et al. (2017). International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am. J. Hum. Genet. *100*, 695–705.
3. Frésard, L., and Montgomery, S.B. (2018). Diagnosing rare diseases after the exome. Cold Spring Harb. Mol. Case Stud. *4*, a003392.
4. Ewans, L.J., Schofield, D., Shrestha, R., Zhu, Y., Gayevskiy, V., Ying, K., Walsh, C., Lee, E., Kirk, E.P., Colley, A., et al. (2018). Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. Genet. Med. *20*, 1564–1574.
5. Eichler, E.E. (2019). Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. N. Engl. J. Med. *381*, 64–74.
6. Logsdon, G.A., Vollger, M.R., and Eichler, E.E. (2020). Long-read human genome sequencing and its applications. Nat. Rev. Genet. *21*, 597–614.
7. Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat. Commun. *10*, 1784.
8. Hiatt, S.M., Lawlor, J.M.J., Handley, L.H., Ramaker, R.C., Rogers, B.B., Partridge, E.C., Boston, L.B., Williams, M., Plott, C.B., Jenkins, J., et al. (2021). Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders. HGG Adv *2*, 100023.
9. Mitsuhashi, S., and Matsumoto, N. (2020). Long-read sequencing for rare human genetic diseases. J. Hum. Genet. *65*, 11–19.
10. Gilpatrick, T., Lee, I., Graham, J.E., Raimondeau, E., Bowen, R., Heron, A., Downs, B., Sukumar, S., Sedlazeck, F.J., and Timp, W. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat. Biotechnol. *38*, 433–438.
11. Karamitros, T., and Magiorkinis, G. (2015). A novel method for the multiplexed target enrichment of MinION next

generation sequencing libraries using PCR-generated baits. Nucleic Acids Res. *43*, e152, e152.

12. Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. *12*, 1261–1276.
13. Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B.J., and Loose, M. (2021). Readfish enables targeted nanopore sequencing of gigabase-sized genomes. Nat. Biotechnol. *39*, 442–450.
14. Loose, M., Malla, S., and Stout, M. (2016). Real-time selective sequencing using nanopore technology. Nat. Methods *13*, 751–754.
15. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.
16. Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods *15*, 461–468.
17. Edge, P., and Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. Nat. Commun. *10*, 4660.
18. Luo, R., Wong, C.-L., Wong, Y.-S., Tang, C.-I., Liu, C.-M., Leung, C.-M., and Lam, T.-W. (2020). Exploring the limit of using a deep neural network on pileup data for germline variant calling. Nat Mach Intell *2*, 220–227.
19. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.
20. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47* (D1), D886–D894.
21. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. Cell *176*, 535–548.e24.
22. Killick, R., and Eckley, I.A. (2014). changepoint : An R Package for Changepoint Analysis. J. Stat. Softw. *58*. https://doi.org/10.18637/jss.v058.i03.
23. Heller, D., and Vingron, M. (2019). SVIM: structural variant identification using mapped long reads. Bioinformatics *35*, 2907–2915.
24. Loman, N.J., Quick, J., and Simpson, J.T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. Methods *12*, 733–735.
25. Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Gershman, A., Sadowski, N., Sedlazeck, F.J., Hansen, K.D., Simpson, J.T., and Timp, W. (2020). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. Nat. Methods *17*, 1191–1199.
26. LaCroix, A.J., Stabley, D., Sahraoui, R., Adam, M.P., Mehaffey, M., Kernan, K., Myers, C.T., Fagerstrom, C., Anadiotis, G., Akkari, Y.M., et al.; University of Washington Center for Mendelian Genomics (2019). GGC Repeat Expansion and Exon 1 Methylation of XYLT1 Is a Common Pathogenic Variant in Baratela-Scott Syndrome. Am. J. Hum. Genet. *104*, 35–44.
27. Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. *37*, 1155–1162.
28. Audano, P.A., Sulovari, A., Graves-Lindsay, T.A., Cantsilieris, S., Sorensen, M., Welch, A.E., Dougherty, M.L., Nelson, B.J., Shah, A., Dutcher, S.K., et al. (2019). Characterizing the Major Structural Variant Alleles of the Human Genome. Cell *176*, 663–675.e19.
29. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
30. Gel, B., and Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics *33*, 3088–3090.
31. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.
32. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics *31*, 3555–3557.
33. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
34. Walsh, T., Casadei, S., Munson, K.M., Eng, M., Mandell, J.B., Gulsuner, S., and King, M.-C. (2020). CRISPR-Cas9/long-read sequencing approach to identify cryptic mutations in *BRCA1* and other tumour suppressor genes. J. Med. Genet. https://doi.org/10.1136/jmedgenet-2020-107320.
35. Pfeifer, D., Kist, R., Dewar, K., Devon, K., Lander, E.S., Birren, B., Korniszewski, L., Back, E., and Scherer, G. (1999). Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. Am. J. Hum. Genet. *65*, 111–124.
36. Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J.A., Fenwick, R.G., Jr., Warren, S.T., et al. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell *67*, 1047–1058.
37. Zhou, Y., Kumari, D., Sciascia, N., and Usdin, K. (2016). CGG-repeat dynamics and *FMR1* gene silencing in fragile X syndrome stem cells and stem cell-derived neurons. Mol. Autism *7*, 42.
38. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al.; ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.
39. Sulovari, A., Li, R., Audano, P.A., Porubsky, D., Vollger, M.R., Logsdon, G.A., Warren, W.C., Pollen, A.A., Chaisson, M.J.P., Eichler, E.E.; and Human Genome Structural Variation Consortium (2019). Human-specific tandem repeat expansion and differential gene expression during primate evolution. Proc. Natl. Acad. Sci. USA *116*, 23243–23253.
40. Trost, B., Engchuan, W., Nguyen, C.M., Thiruvahindrapuram, B., Dolzhenko, E., Backstrom, I., Mirceta, M., Mojarad, B.A.,

Yin, Y., Dov, A., et al. (2020). Genome-wide detection of tandem DNA repeats that are expanded in autism. Nature *586*, 80–86.

41. Dolzhenko, E., Deshpande, V., Schlesinger, F., Krusche, P., Petrovski, R., Chen, S., Emig-Agius, D., Gross, A., Narzisi, G., Bowman, B., et al. (2019). ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics *35*, 4754–4756.
42. Ebert, P., Audano, P.A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M.J., Sulovari, A., Ebler, J., Zhou, W., Serra Mari, R., et al. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science *372*, eabf7117.
43. Beck, C.R., Carvalho, C.M.B., Akdemir, Z.C., Sedlazeck, F.J., Song, X., Meng, Q., Hu, J., Doddapaneni, H., Chong, Z., Chen, E.S., et al. (2019). Megabase Length Hypermutation Accompanies Human Structural Variation at 17p11.2. Cell *176*, 1310–1324.e10.
44. Beyter, D., Ingimundardottir, H., Oddsson, A., Eggertsson, H.P., Bjornsson, E., Jonsson, H., Atlason, B.A., Kristmundsdottir, S., Mehringer, S., Hardarson, M.T., et al. (2020). Long read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. bioRxiv. https://doi.org/10.1101/848366.
45. Punga, T., and Bühler, M. (2010). Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. EMBO Mol. Med. *2*, 120–129.
46. Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.-L., Ruan, Y., Wei, C.-L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752–1762.
47. Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.C., et al. (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. USA *108*, 260–265.
48. Levin, A.A. (2019). Treating Disease at the RNA Level with Oligonucleotides. N. Engl. J. Med. *380*, 57–70.
49. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
50. Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., et al. (2015). The Matchmaker Exchange: a platform for rare disease gene discovery. Hum. Mutat. *36*, 915–921.
51. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., et al. (2020). Telomere-to-telomere assembly of a complete human X chromosome. Nature *585*, 79–84.
52. Denny, J.C., Rutter, J.L., Goldstein, D.B., Philippakis, A., Smoller, J.W., Jenkins, G., Dishman, E.; and All of Us Research Program Investigators (2019). The "All of Us" Research Program. N. Engl. J. Med. *381*, 668–676.
53. Mahmoud, M., Gobet, N., Cruz-Dávalos, D.I., Mounier, N., Dessimoz, C., and Sedlazeck, F.J. (2019). Structural variant calling: the long and the short of it. Genome Biol. *20*, 246.
54. Palmer, E.E., Sachdev, R., Macintosh, R., Melo, U.S., Mundlos, S., Righetti, S., Kandula, T., Minoche, A.E., Puttick, C., Gayevskiy, V., et al. (2021). Diagnostic Yield of Whole Genome Sequencing After Nondiagnostic Exome Sequencing or Gene Panel in Developmental and Epileptic Encephalopathies. Neurology *96*, e1770–e1782.
55. Murdock, D.R., Dai, H., Burrage, L.C., Rosenfeld, J.A., Ketkar, S., Müller, M.F., Yépez, V.A., Gagneur, J., Liu, P., Chen, S., et al. (2020). Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. J. Clin. Invest *131*, e141500, 33001864.
56. Hochstenbach, R., Liehr, T., and Hastings, R.J. (2021). Chromosomes in the genomic age. Preserving cytogenomic competence of diagnostic genome laboratories. Eur. J. Hum. Genet. *29*, 541–552.

The American Journal of Human Genetics, Volume *108*

Supplemental information

Targeted long-read sequencing identifies missing disease-causing variation

Danny E. Miller, Arvis Sulovari, Tianyun Wang, Hailey Loucks, Kendra Hoekzema, Katherine M. Munson, Alexandra P. Lewis, Edith P. Almanza Fuerte, Catherine R. Paschal, Tom Walsh, Jenny Thies, James T. Bennett, Ian Glass, Katrina M. Dipple, Karynne Patterson, Emily S. Bonkowski, Zoe Nelson, Audrey Squire, Megan Sikes, Erika Beckman, Robin L. Bennett, Dawn Earl, Winston Lee, Rando Allikmets, Seth J. Perlman, Penny Chow, Anne V. Hing, Tara L. Wenger, Margaret P. Adam, Angela Sun, Christina Lam, Irene Chang, Xue Zou, Stephanie L. Austin, Erin Huggins, Alexias Safi, Apoorva K. Iyengar, Timothy E. Reddy, William H. Majoros, Andrew S. Allen, Gregory E. Crawford, Priya S. Kishnani, University of Washington Center for Mendelian Genomics, Mary-Claire King, Tim Cherry, Jessica X. Chong, Michael J. Bamshad, Deborah A. Nickerson, Heather C. Mefford, Dan Doherty, and Evan E. Eichler

Supplemental Data

Supplemental Text

Clinical summary and prior testing for simple structural variant cases

All nine cases beginning with "BK" are from the Study of Autism Genetics Exploration (SAGE) collection, five cases of which have been previously described[5] and four (BK397-101, BK430-103, BK482-101, BK487-101) are samples collected after the publication.

Individual S016 has a known duplication in *CTNND2* confirmed by clinical mate-pair sequencing to be tandem.[1]

Individual S023 presented with developmental delay, epilepsy, and hypothyroidism. A SNP array identified a mosaic region of 18p11.32 and 18q21.31q23 with copy number between 1 and 2. Evaluation of chromosomes revealed that this individual was mosaic for a ring chromosome 18 in ~40% of cells.

Individual S046 presented with developmental delay and was found to have an unbalanced translocation between chromosomes 4 and 15. A single-nucleotide polymorphism (SNP) array identified a copy state of 1 at 4q35.1q35.2 and a copy state of 3 at 15q26.1q26.3.

Individual S060 was diagnosed with campomelic dysplasia, vertebral anomalies, neurogenic bowel and bladder, and bilateral moderate conductive hearing loss. Karyotype revealed 46,XY,t(12;17)(q13.3;q25). Because of high suspicion for campomelic dysplasia associated with a *SOX9* variant sequencing of *SOX9* was undertaken which was negative. SNP microarray was used in an attempt to localize the translocation breakpoints if they were associated with a deletion or duplication but was unrevealing.

Individual S063 was from a family with severe bilateral breast and ovarian cancer in which clinical evaluation failed to identify a pathogenic variant in either *BRCA1* or *BRCA2*. Long-read sequencing on a PacBio platform after CRISPR-Cas9 excision of *BRCA1* and *BRCA2* revealed a SINE-VNTR-Alu retrotransposon insertion that created a pseudoexon in *BRCA1*.

Clinical summary and prior testing for cases that underwent phasing

Individual S071 was referred to genetics for global developmental delay. Trio exome sequencing revealed two variants of uncertain clinical significance, one a *de novo* variant, in *METTL5* that could not be phased.

Individual S086 carried a diagnosis of epilepsy. Panel testing on an exome backbone revealed an inherited pathogenic variant in *KIAA1109* and a *de novo* mosaic variant. These variants could not be phased by the clinical laboratory.

Clinical summary and prior testing for repeat expansion cases

Individual S011 was an individual with clinically confirmed expansions in *ATXN3* (74 and 28 repeats) and *ATXN8OS* (80 and 25 repeats). Individual 04-01 (proband), 04-02 (mother), 04-03 (father), 06-01 (proband), 06-02 (mother), and 06-03 (father) were previously described.[3]

Clinical summary and prior testing for complex structural variant cases

Individual S014 presented prenatally with agenesis of the corpus callosum and ventriculomegaly. At birth the child was noted to have mild dysmorphic features and pelviectasis without hydroureter. A SNP array identified a complex pathogenic heterozygous deletion within 6q25.2 to 6q25.3, which included *ARID1B* and perhaps explained most of his clinical findings.

Individual S020 presented with hypotonia, developmental delay, epilepsy, and dysmorphic features. A SNP array revealed deletions of 4q13.2q13.3, 4q13.3, and 14q11.2. Karyotyping identified rearrangements among chromosomes 2, 14, 10, and 4 that involved the deleted regions and also 2p23, 2p25, 10q21.2, 10q21.1, and 10q22.3.

Individual S021 presented with developmental delay and was found by microarray and karyotype to be mosaic for complex changes to chromosome 8 that included loss of 8p23.2–pter (copy number 1), mosaic loss of a segment within 8p23.2–8p23.1 (copy number 1–2), and mosaic gain of 8q22.1–qter (copy number 2–3). Karyotype identified one cell line with a derivative chromosome 8 that consists of a terminal duplication of 8q, from 8q22.1 to qter, and a terminal deletion of 8p, from 8p23.2 to pter. The duplicated region of 8q is present on distal 8p, such that 8q22.1 to qter is present at both ends of the derivative chromosome 8. The other cell line has a terminal deletion of 8p, from 8p23.1 to pter.

Individual S022 presented with hemihypertrophy of unclear etiology. A standard Beckwith-Wiedemann workup including *CDKN1C* sequencing and deletion/duplication analysis was unremarkable. A SNP array identified a focal amplification of 4q with the copy state reported as more than 4 with an adjacent 27.5 Mbp region of homozygosity on 4q. A SNP array also identified a duplication of 15q11.2. Metaphase and interphase FISH confirmed that the focal amplification of 4q was more than 4.

Individual S035 presented with developmental delay and was found on array to have duplications of both 8q24.3 and 16p13.11.

Individual S036 underwent genetic testing for expressive language delay, microcephaly, and mild dysmorphic features. A SNP array revealed four noncontiguous deletions of chromosome 10 at 10p12.2p12.1, 10p11.21, and 10q21.1 (two deletions in this interval). Karyotype revealed a translocation between chromosomes 6 and 18 at 6q22.2 and 18p11.2 as well as a pericentric inversion of chromosome 10 at 10p11.2q11.2.

Individual S082 was referred for genetic testing because of gross motor delay and multiple congenital anomalies. SNP array revealed two pathogenic CNVs, a 10q25.2 deletion involving *RBM20*, and a complex chromosome 17 duplication and deletion involving *RAI1* and *PMP22*.

Individual S083 was prenatally diagnosed with multiple congenital anomalies, including imperforate anus and vertebral anomalies. SNP array after birth revealed a 4p16.3 deletion and 4p16.3p15.2 duplication. The deletion overlaps the Wolf-Hirschhorn syndrome critical region.

Clinical summary and prior testing for missing variant cases

Individual S002 presented with early-onset obesity, type 2 diabetes, cone-rod dystrophy, and sensorineural hearing loss. Clinical testing by SNP array was unremarkable and exome sequencing identified a single paternally inherited stop-gain variant in *ALMS1*, the gene associated with Alström syndrome, a recessive disorder that fit the phenotype well. Subsequent exon-level array revealed no deletions or duplications in the gene.

Individual S003 was an individual with renal failure, retinal degeneration, and essential tremor. A SNP array was unremarkable, trio exome revealed a single variant in *NPHP4*, and deletion/duplication analysis of NPHP4 was unremarkable. Clinical RNA testing was sent, which was interpreted as indeterminate.

Individual S004 presented with agenesis of the corpus callosum, microcephaly, poor growth, lactic acidosis, and global developmental delay. The SNP array was unremarkable, trio exome sequencing revealed a single pathogenic variant in *VARS2*, and deletion/duplication analysis of the gene was unremarkable.

Individual S008 was an individual with biochemically confirmed Lesch-Nyhan syndrome. Clinical testing included karyotype and sequencing of exons on an exome backbone, both of which were unremarkable.

Individual S009 presented with concern for Duchenne muscular dystrophy due to a family history of the disease. The proband's maternal uncle passed away at age 29 from the disease without a molecular diagnosis. Molecular testing in the proband included SNP array, targeted exon sequencing, and deletion/duplication analysis, all of which were unremarkable. Analysis of muscle biopsy by immunohistochemistry revealed staining indicative of a dystrophinopathy, but dystrophin 1 antibody staining (rod domain) was more than classically seen in Duchenne type dystrophy.

Individual S013 presented with oculocutaneous albinism and platelet dysfunction, suggesting Hermansky-Pudlak syndrome, but SNP array and exome sequencing only identified a single paternally inherited stop in *HPS1*, one of several genes associated with this recessive disorder.

Individual S018 presented with elevated phenylalanine consistent with phenylketonuria. Panel testing revealed a single pathogenic variant in *PAH*.

Individual S025 presented in their early twenties with disease characteristics in the macula of both eyes consistent with recessive Stargardt disease. No other systemic issues or family history was reported. Research sequencing revealed a single inherited pathogenic variant and research whole-genome short-read sequencing failed to identify a second hit. After identification of a 1,500 bp insertion in the first intron of *ABCA4* with LRS visual reanalysis of the short-read data confirmed an 11 bp target site duplication at the same position.

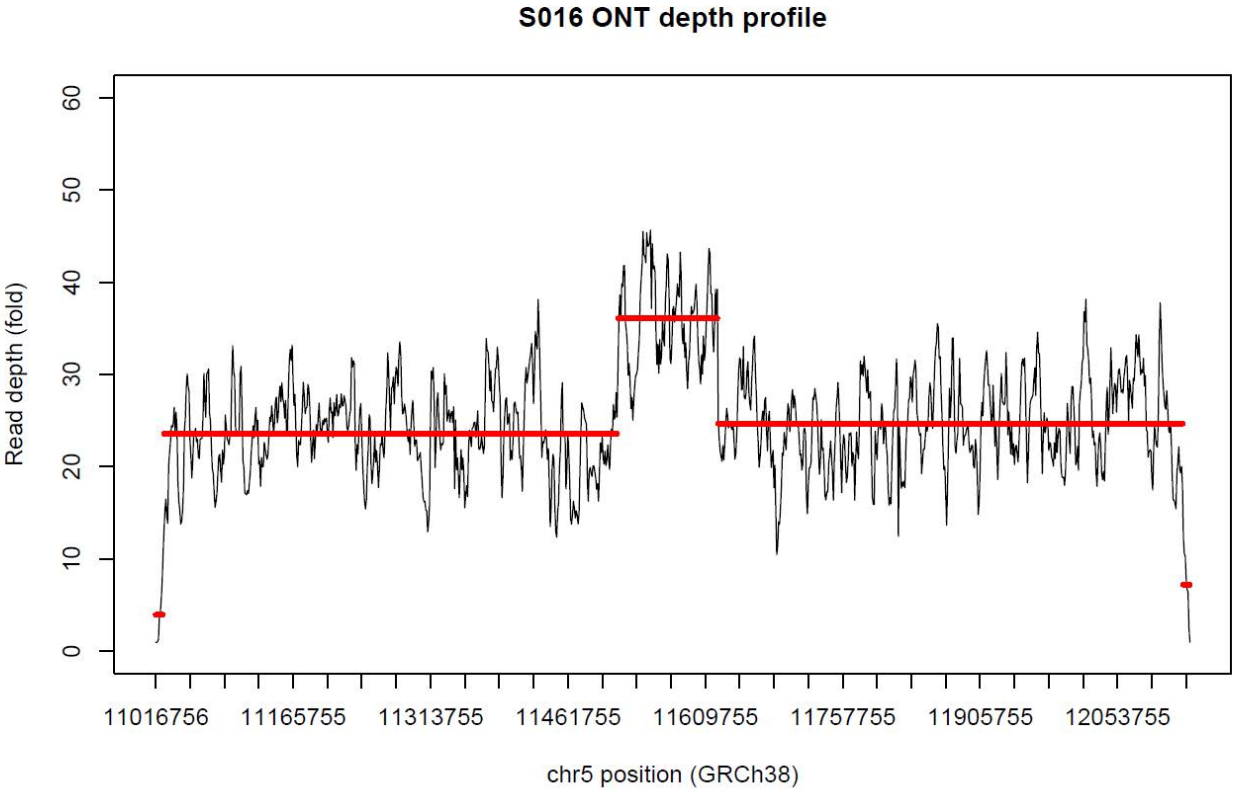
Individual S047 presented at 18 months old with failure to thrive, marked hepatomegaly, and fasting hypoglycemia. A liver biopsy was performed revealing increasing glycogen content, decreased glucose-1-phosphate to glucose ratio, and no measurable debranching enzyme. A diagnosis of Glycogen Storage Disease Type IIIa was suspected. Sequencing of the *AGL* gene revealed a single heterozygous frameshift variant in exon 10 resulting in a premature stop codon (c.1276delG, p.Ile437X) on one allele. A second variant was not identified.

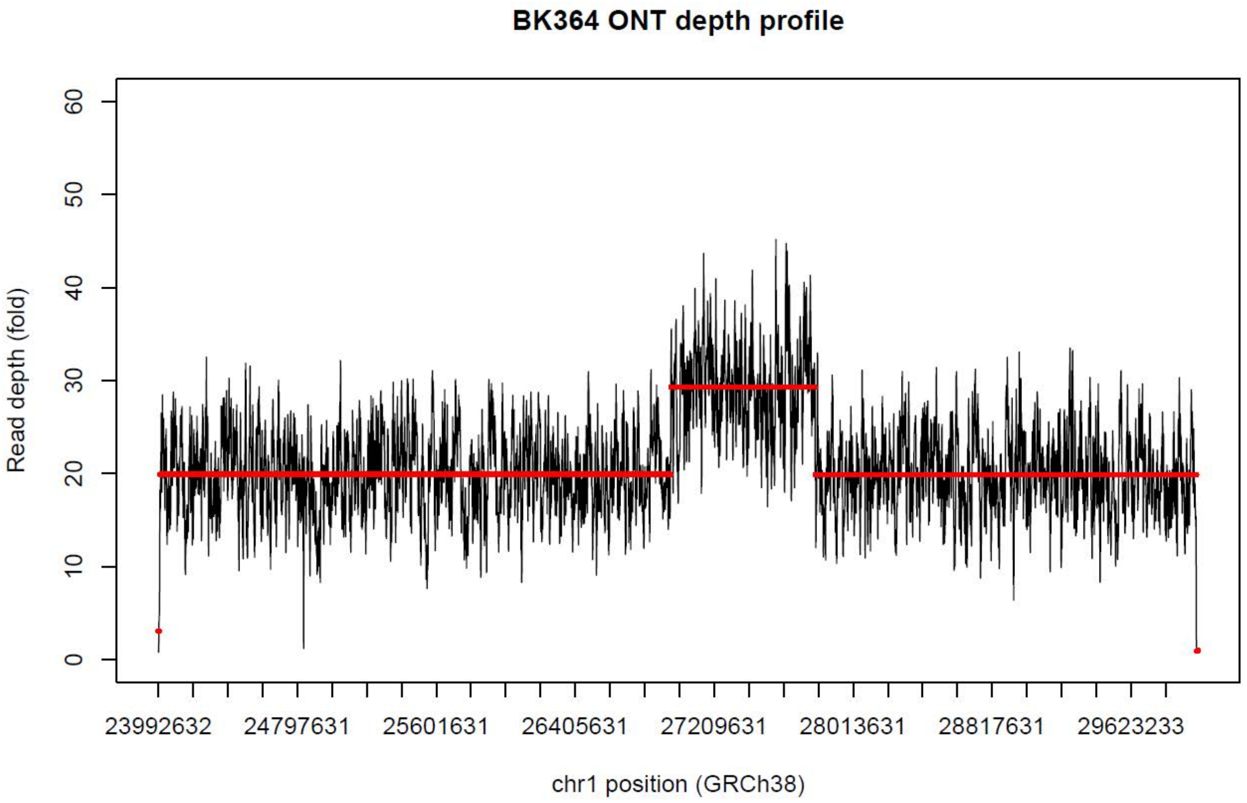
Individual S056 presented with end-stage renal disease related to nephronophthisis, rhizomelic/metaphyseal skeletal dysplasia, retinal dystrophy, developmental delays, hyperparathyroidism, and hepatic fibrosis. A ciliopathy gene panel revealed a single heterozygous pathogenic variant in *WDR19* and deletion/duplication analysis of the gene was unremarkable.

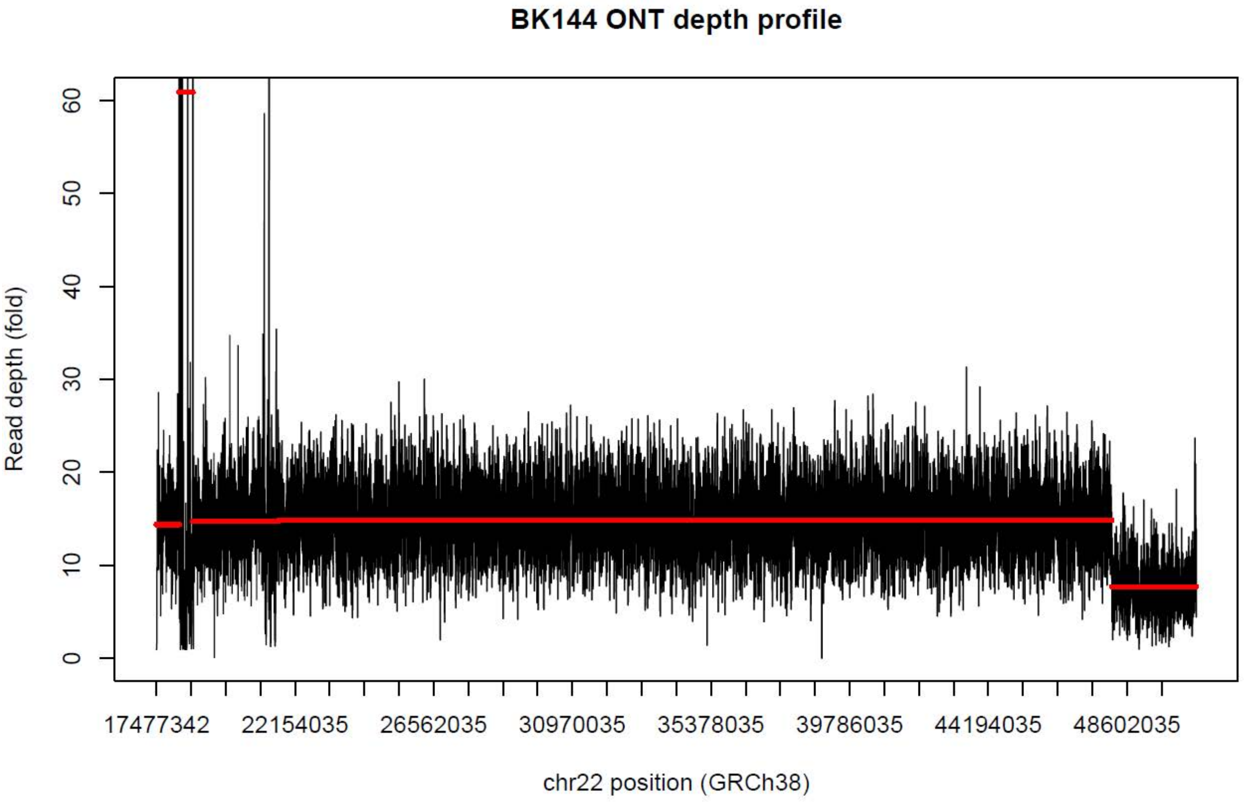
Supplementary Figures

Figure S1. Binary segmentation figures.

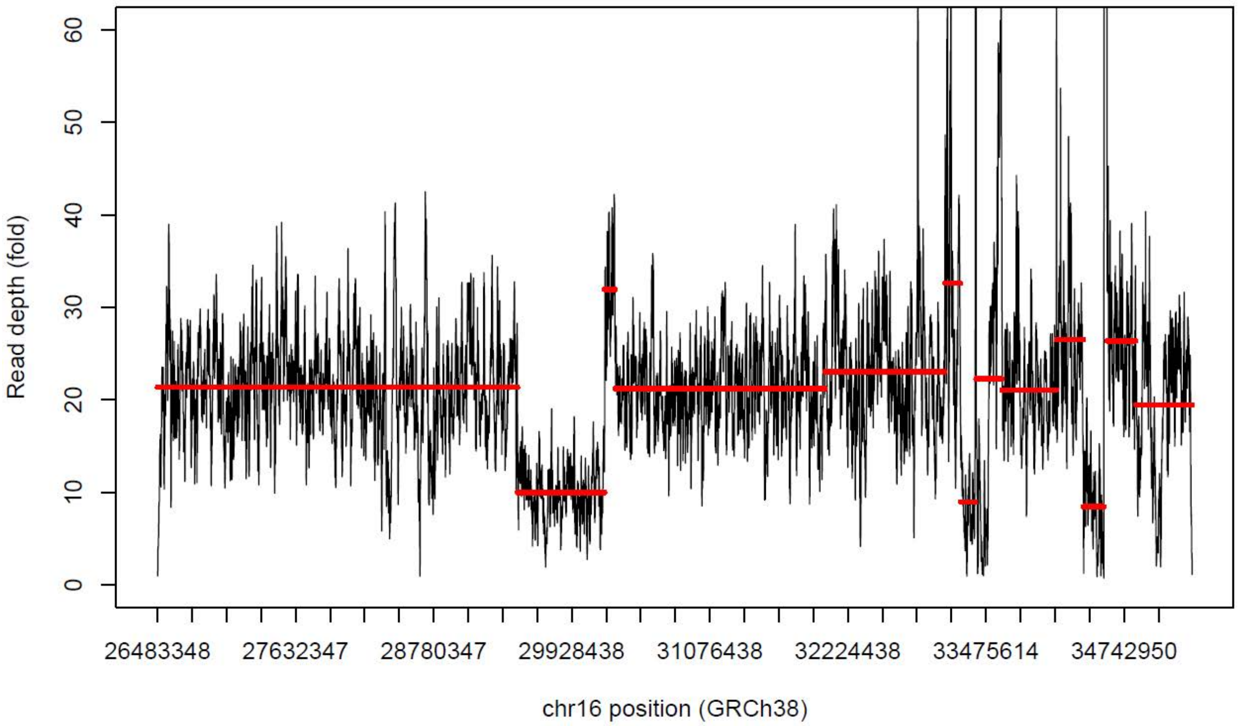
To complement existing SV callers and independently guide the refinement of CNV breakpoints, we applied the binary segmentation method to the sequencing depth profiles of each CNV region. The coordinates of each region were further refined through visual validation. Due to the sensitivity of the binary segmentation method to duplicated genomic sequences, 6 of the 29 regions could not have their breakpoints refined by this method; these were refined using a combination of existing SV callers and visual refinement of sequencing depth profile. The script used to generate these images as well as the coverage for each target region is available on GitHub.



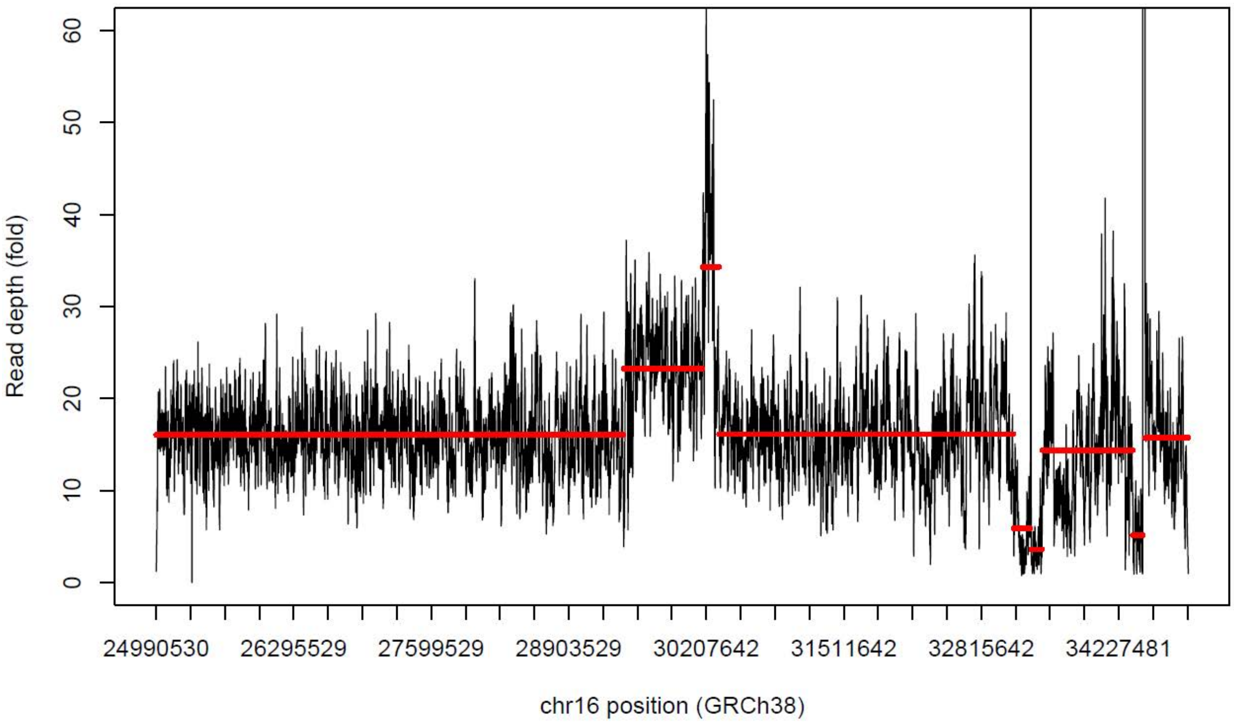

BK364 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
23992632
24797631
25601631
26405631
27209631
28013631
28817631
29623233
chr1 position (GRCh38)

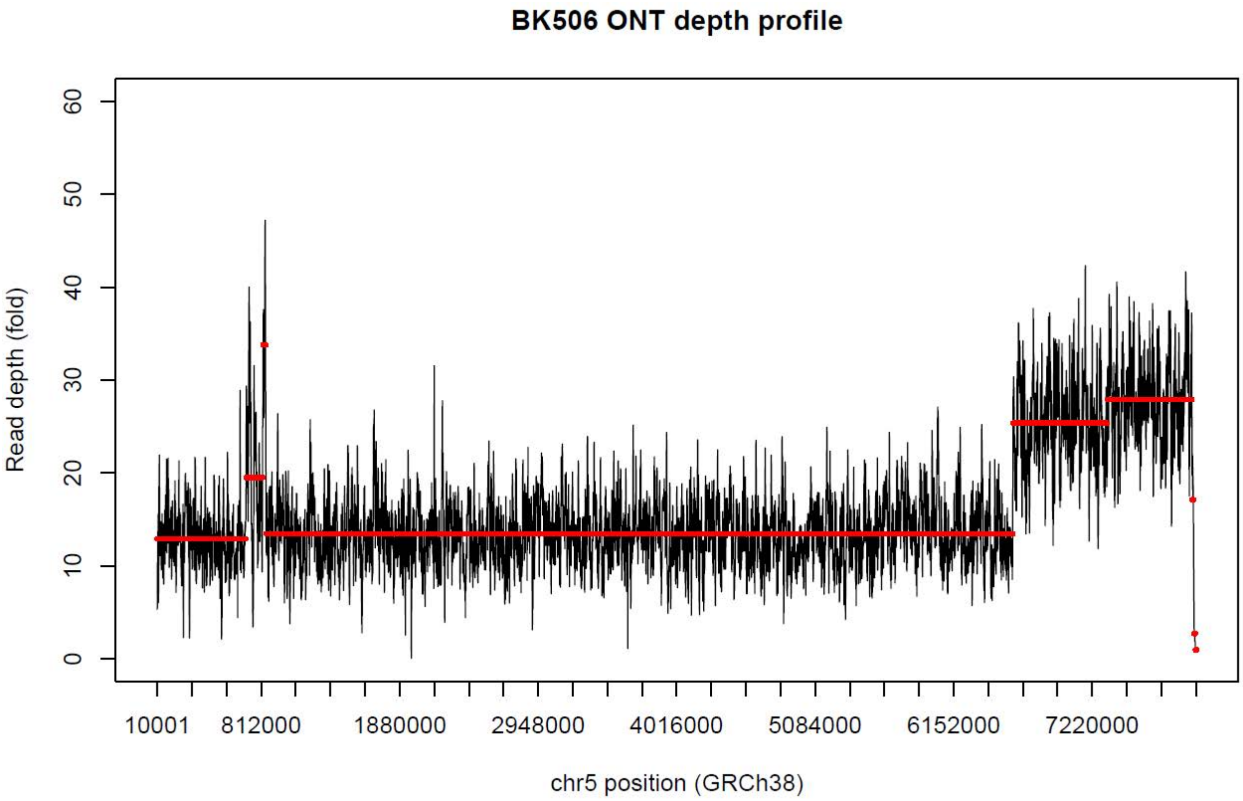

BK144 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
17477342
22154035
26562035
30970035
35378035
39786035
44194035
48602035
chr22 position (GRCh38)

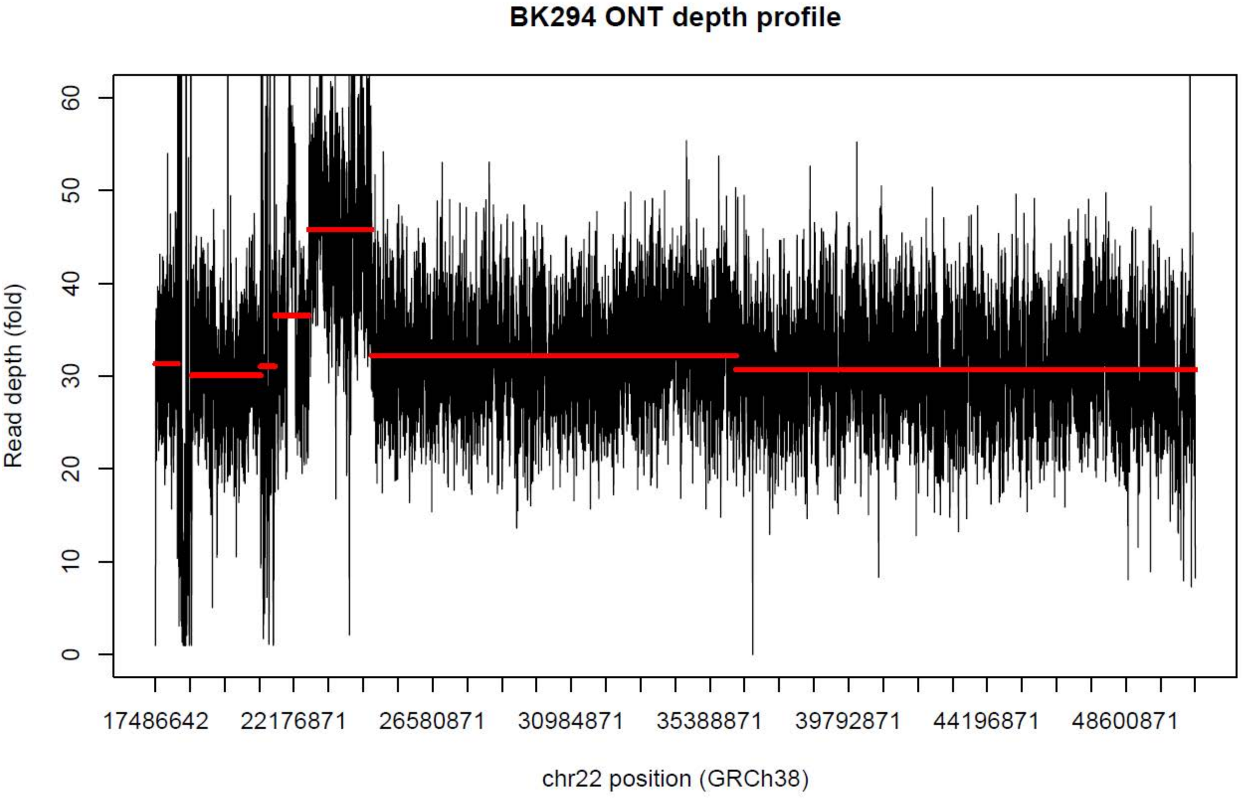
BK397 ONT depth profile

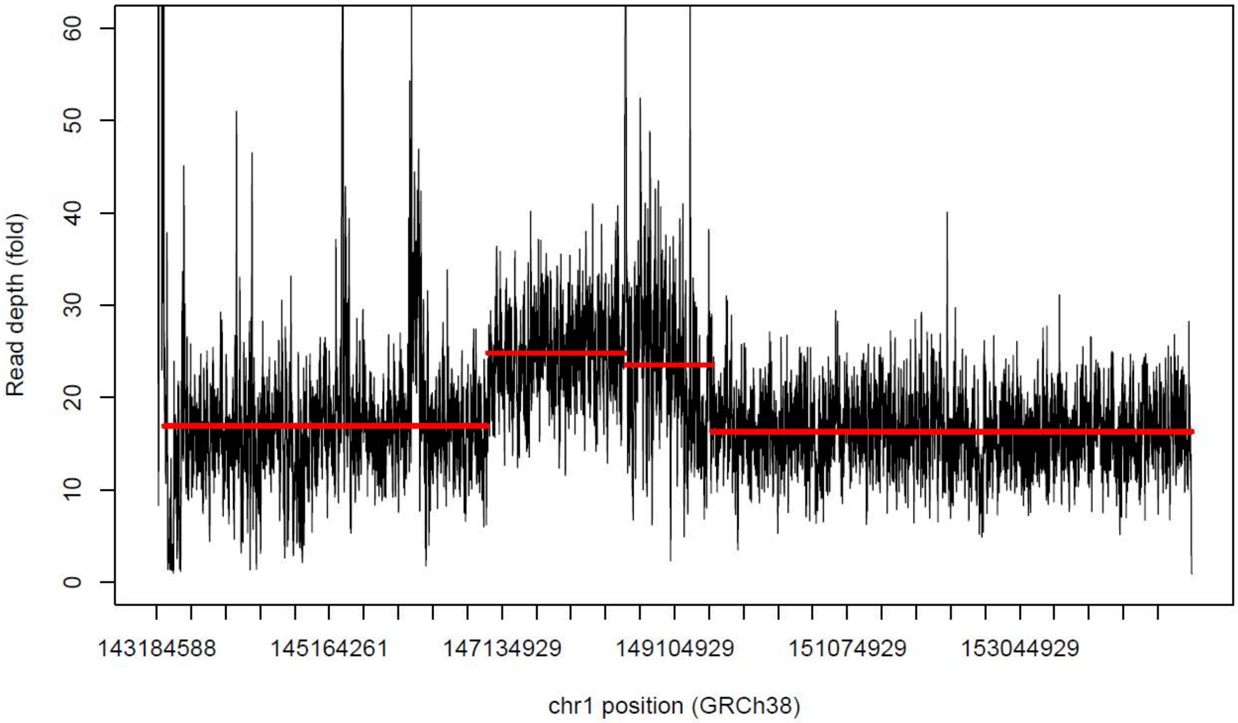


BK430 ONT depth profile

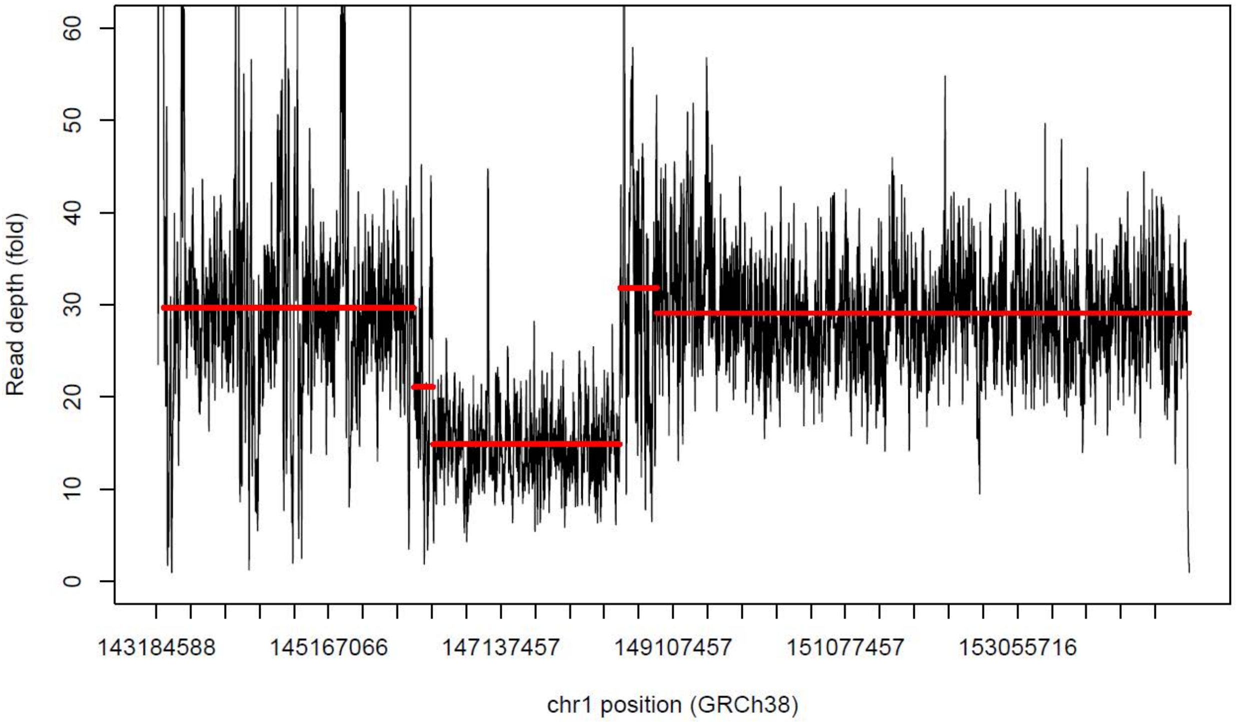



BK506 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
10001
812000
1880000
2948000
4016000
5084000
6152000
7220000
chr5 position (GRCh38)

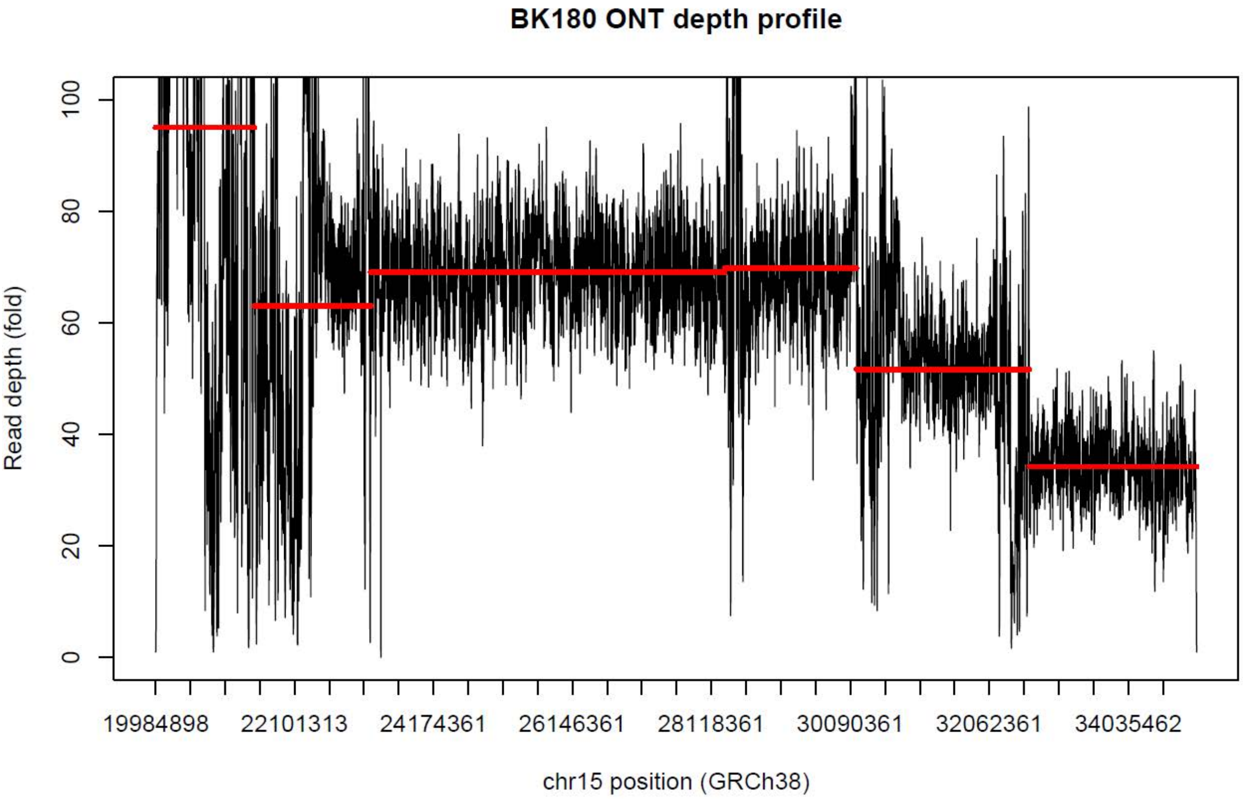

BK294 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
17486642
22176871
26580871
30984871
35388871
39792871
44196871
48600871
chr22 position (GRCh38)

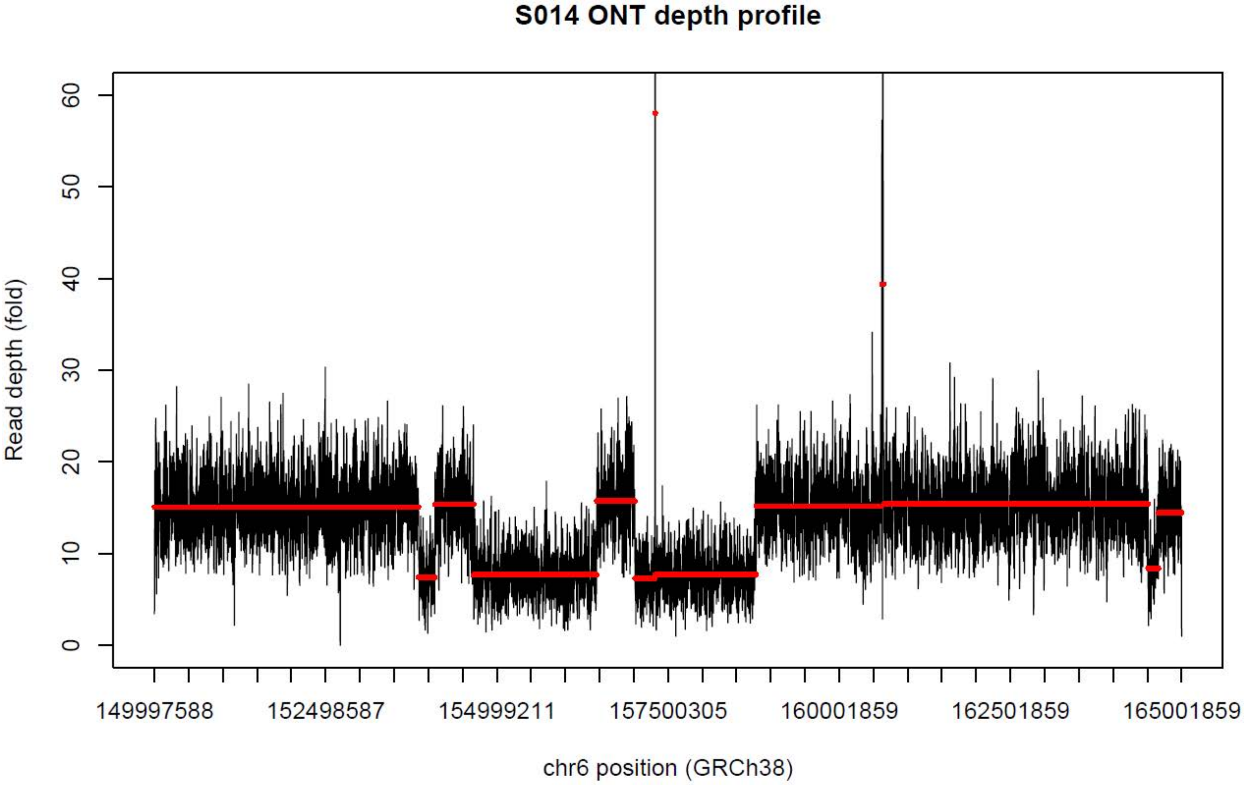


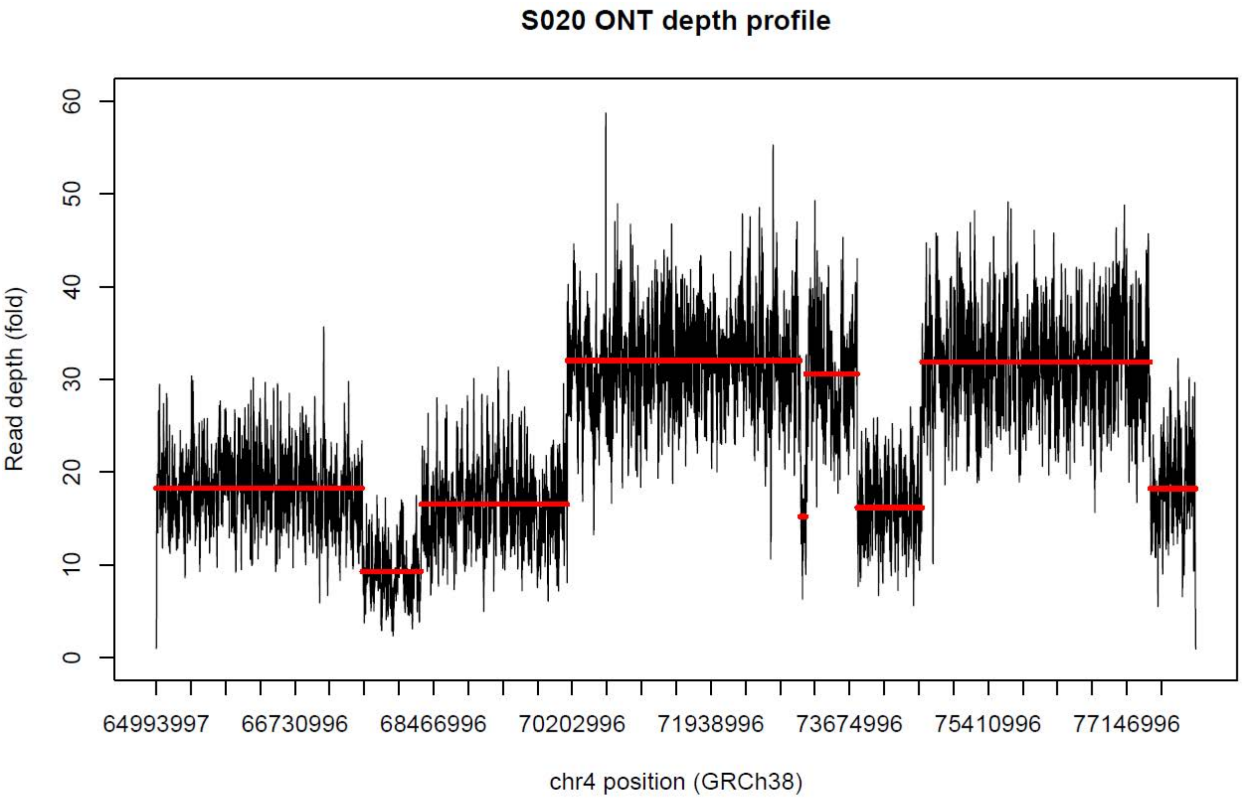
BK482 ONT depth profile

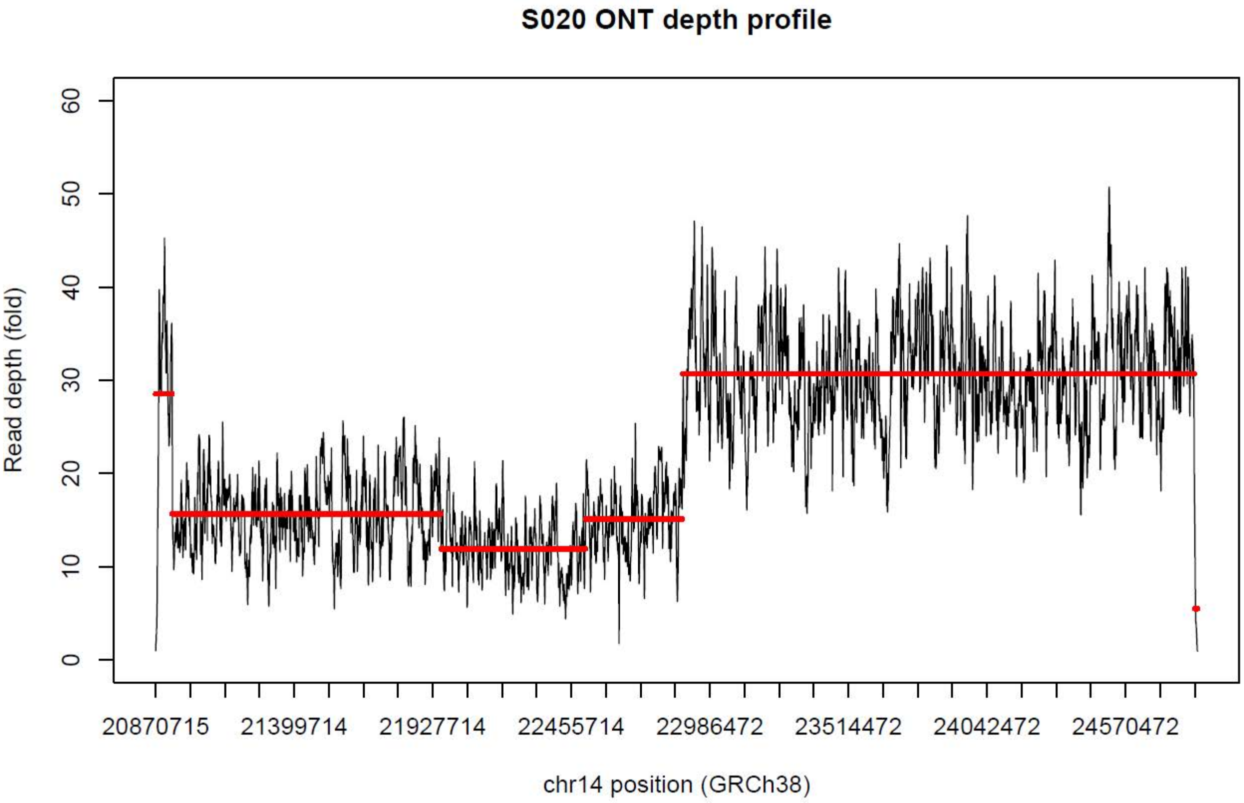


BK487 ONT depth profile

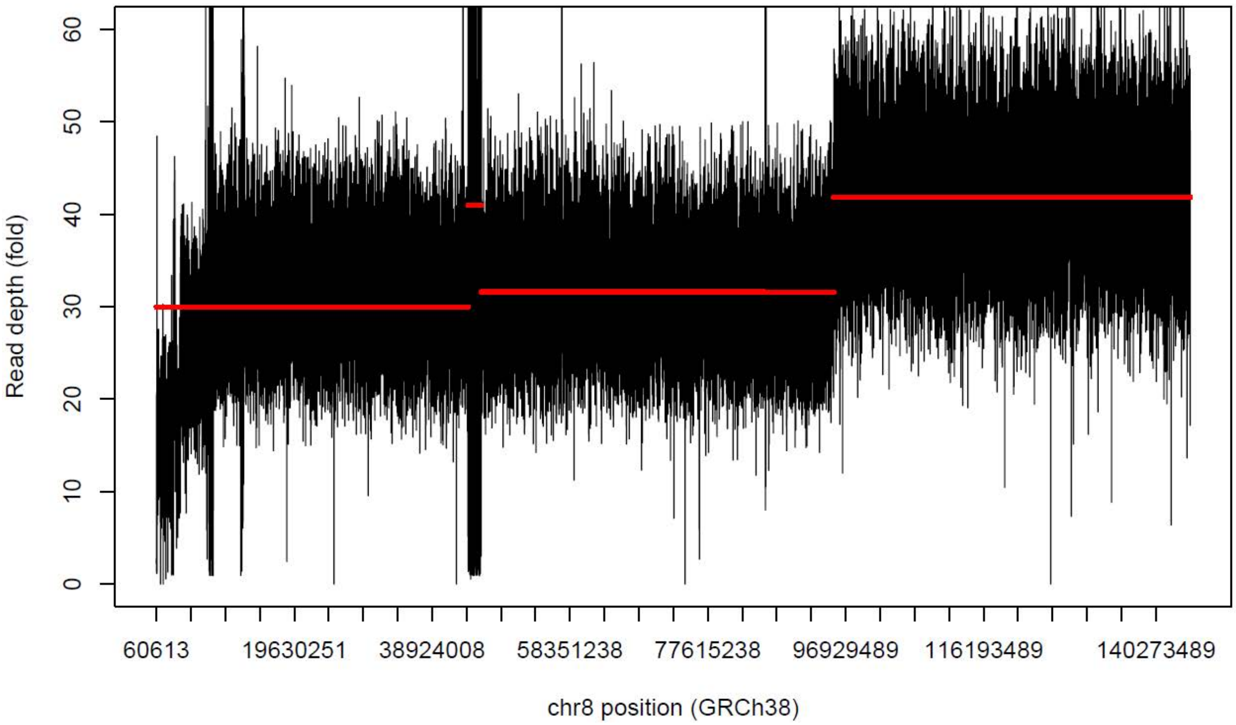

BK180 ONT depth profile
Read depth (fold)
0
20
40
60
80
100
19984898
22101313
24174361
26146361
28118361
30090361
32062361
34035462
chr15 position (GRCh38)


S014 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
149997588
152498587
154999211
157500305
160001859
162501859
165001859
chr6 position (GRCh38)

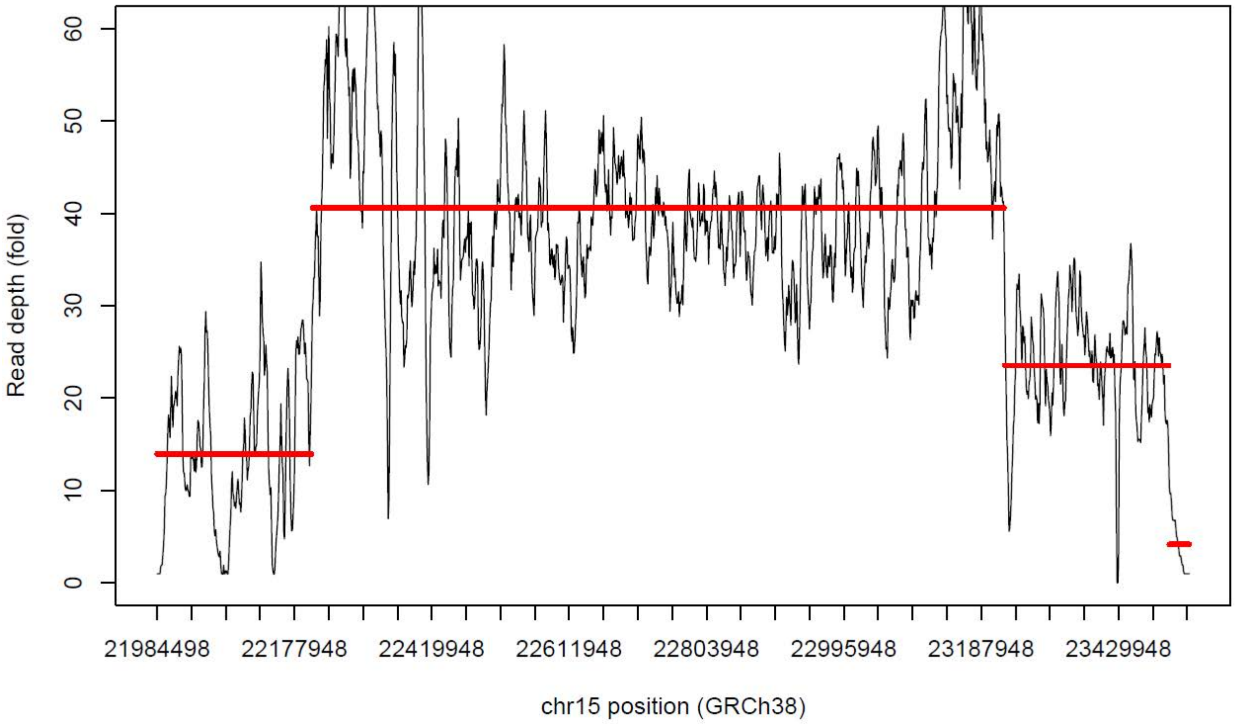

S020 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
64993997
66730996
68466996
70202996
71938996
73674996
75410996
77146996
chr4 position (GRCh38)


S020 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
20870715
21399714
21927714
22455714
22986472
23514472
24042472
24570472
chr14 position (GRCh38)

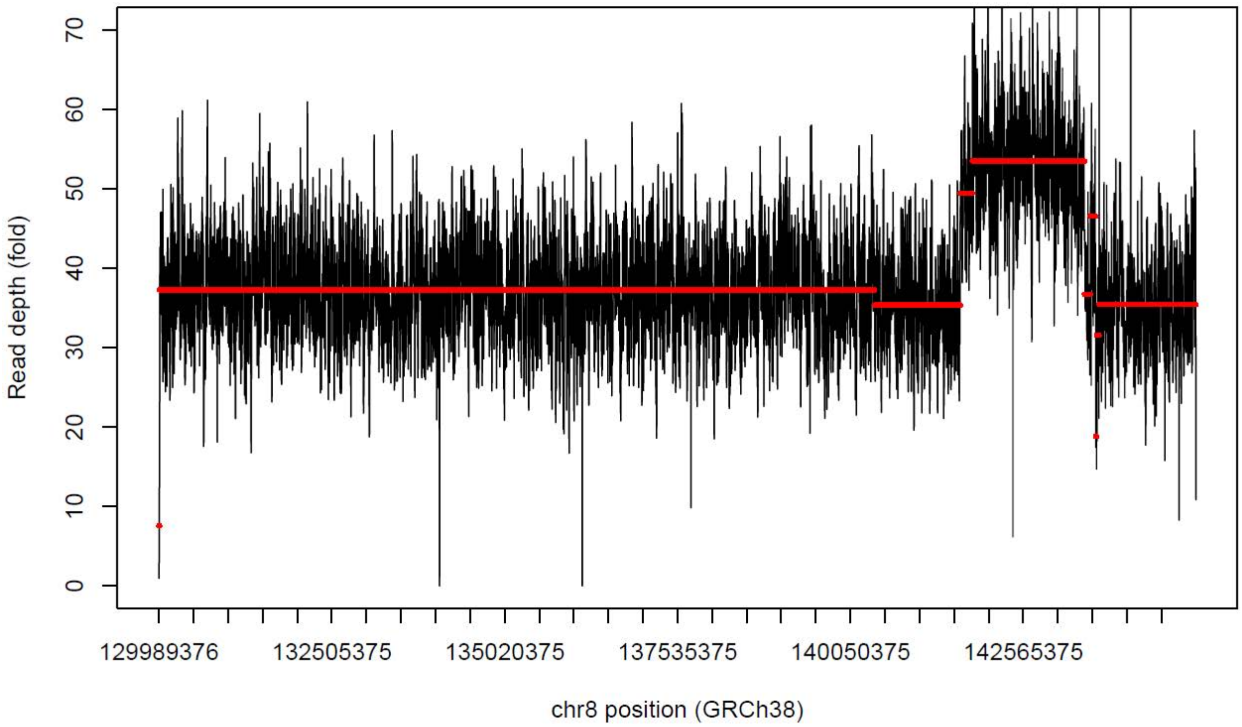
S021 ONT depth profile



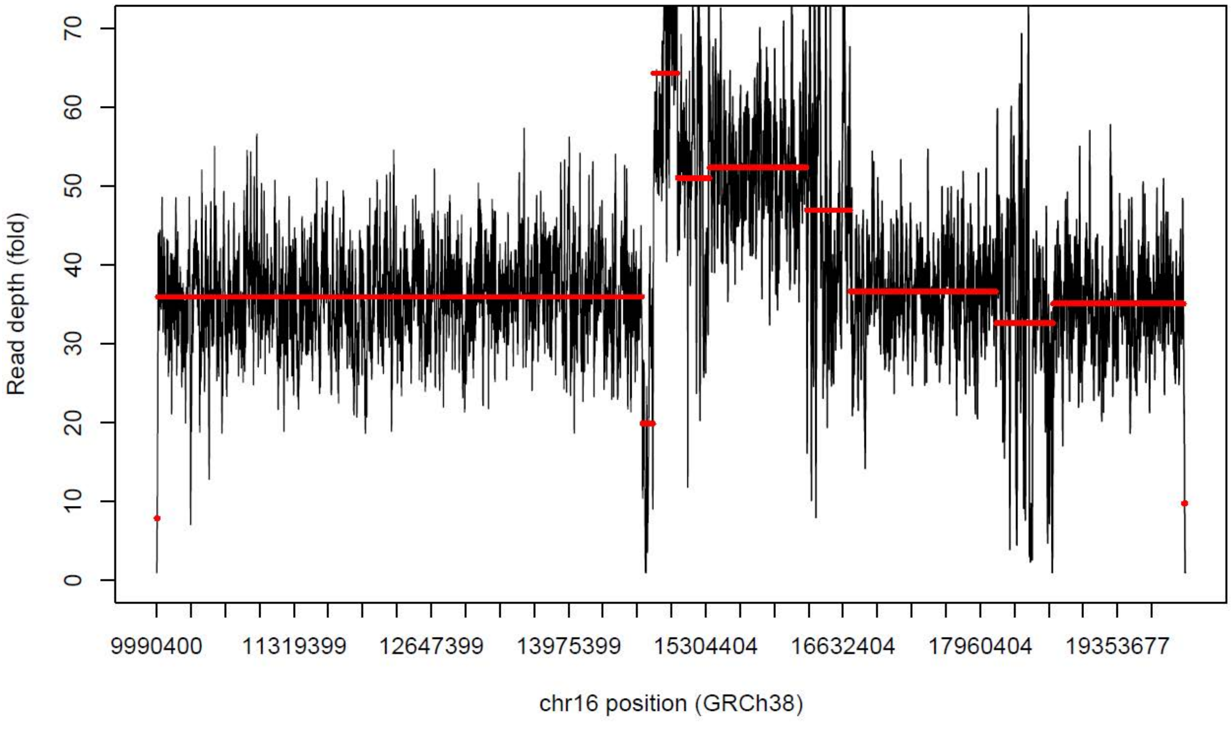
S022 ONT depth profile



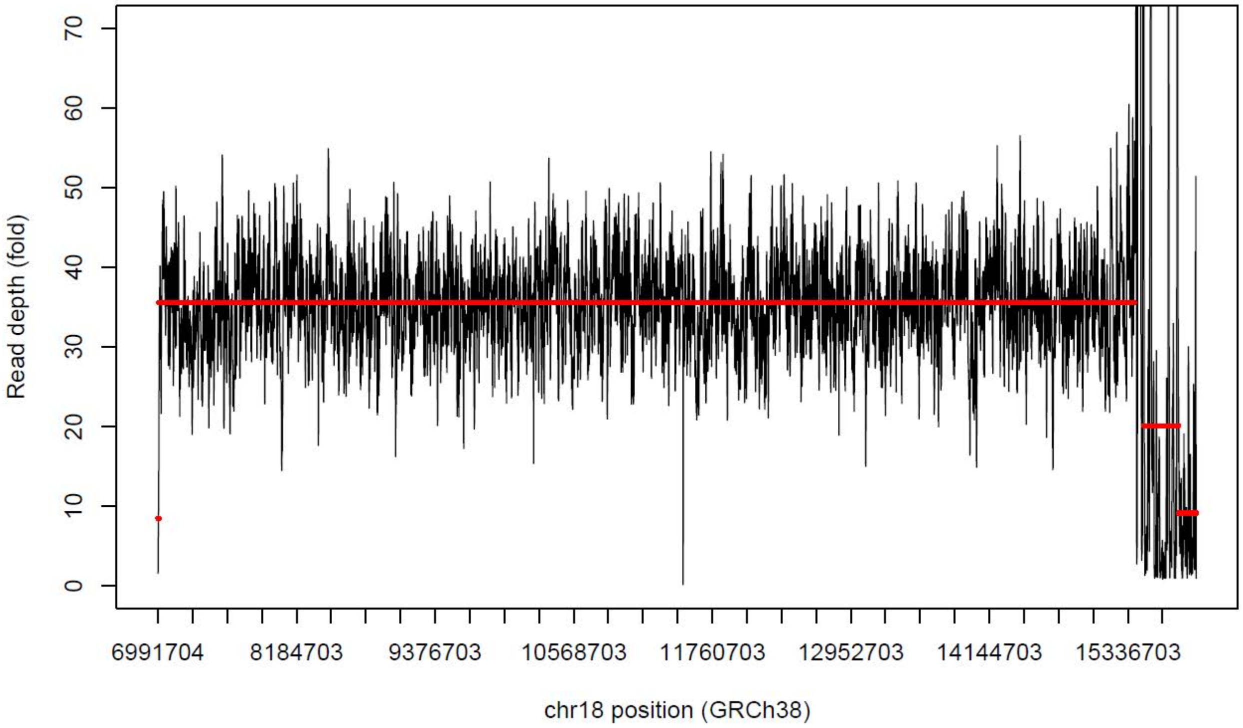
S035 ONT depth profile



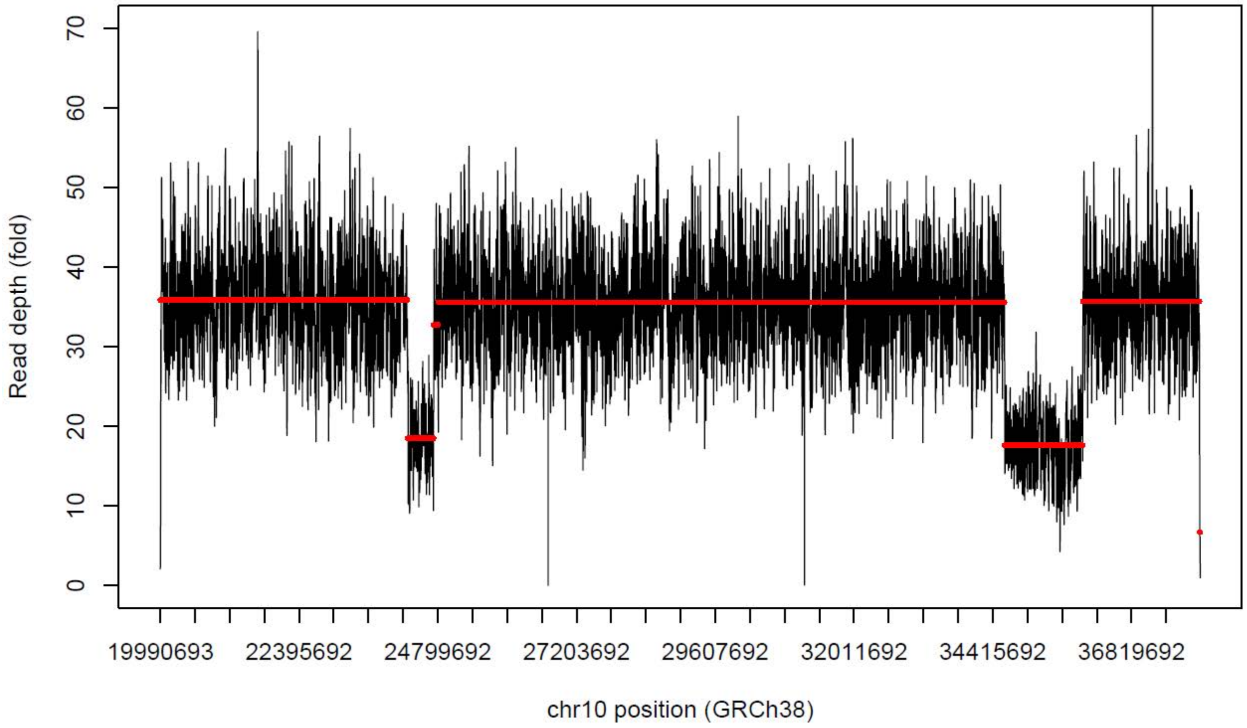
S035 ONT depth profile

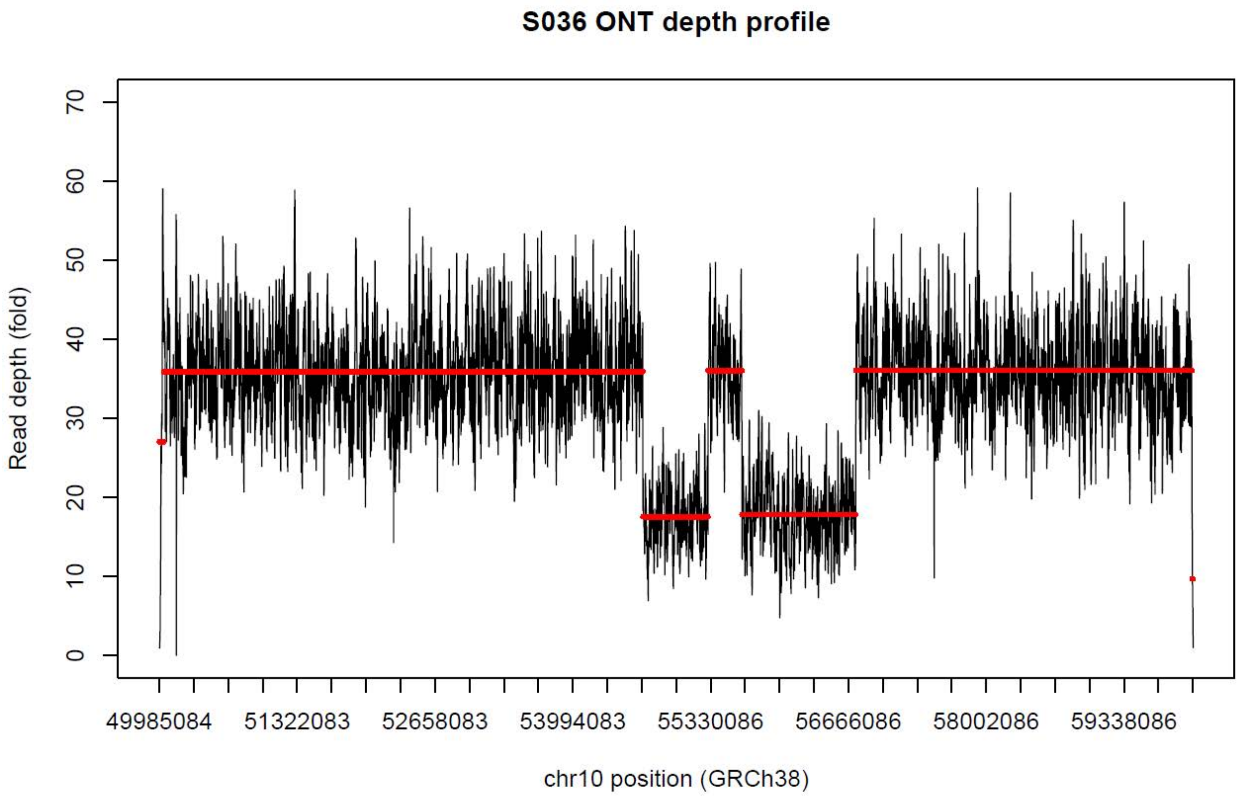


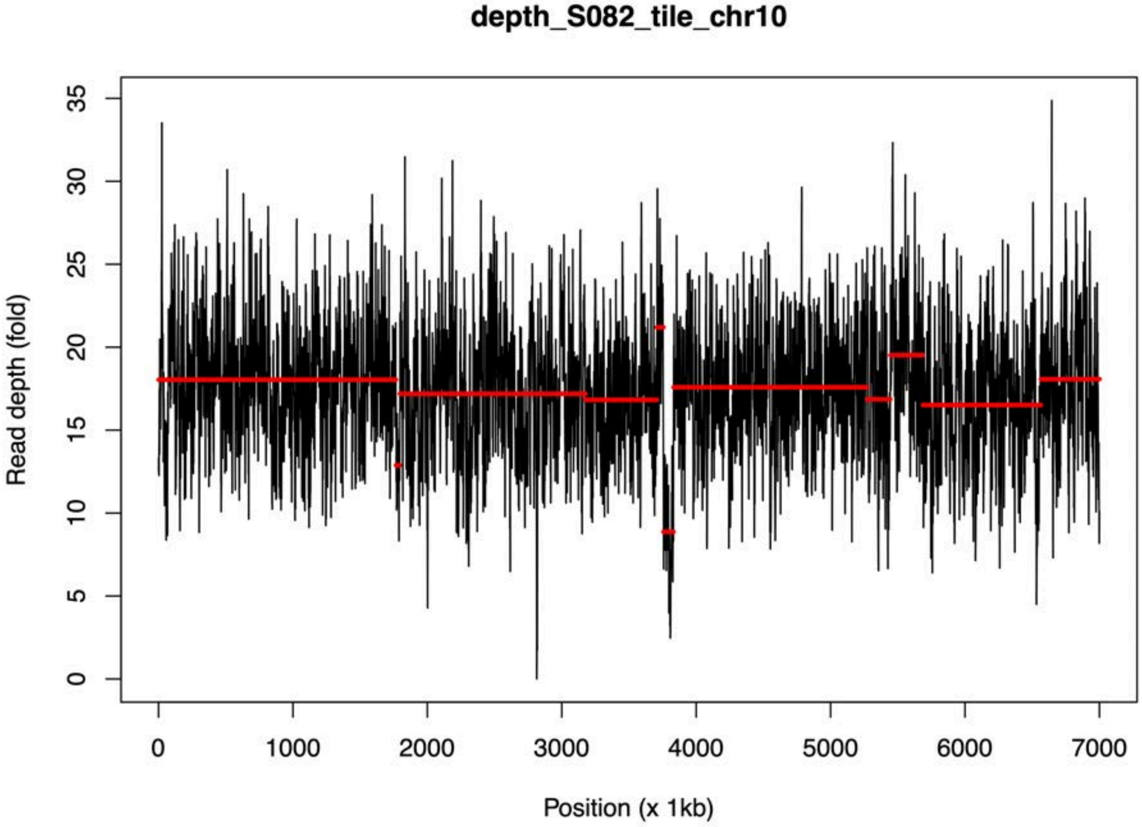
S036 ONT depth profile

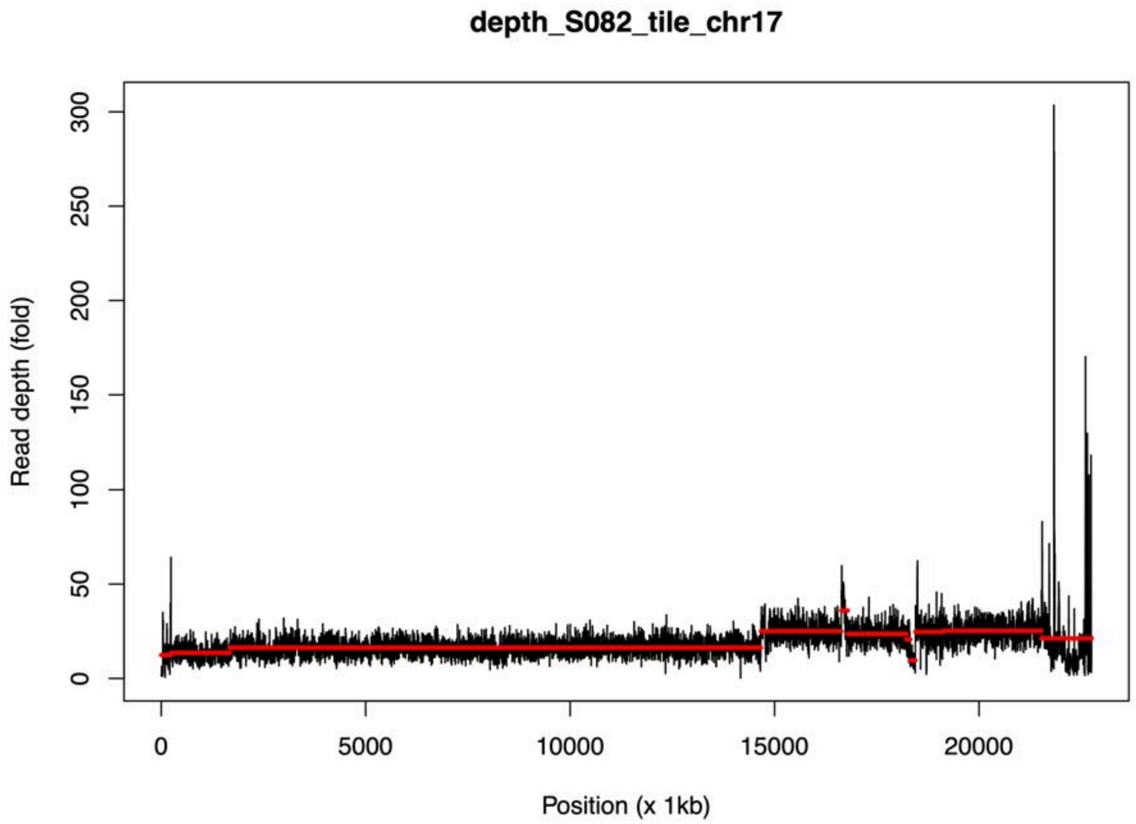


S036 ONT depth profile




S036 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
70
49985084
51322083
52658083
53994083
55330086
56666086
58002086
59338086
chr10 position (GRCh38)


depth_S082_tile_chr10
Read depth (fold)
0
5
10
15
20
25
30
35
0
1000
2000
3000
4000
5000
6000
7000
Position (x 1kb)


depth_S082_tile_chr17
Read depth (fold)
0
50
100
150
200
250
300
0
5000
10000
15000
20000
Position (x 1kb)

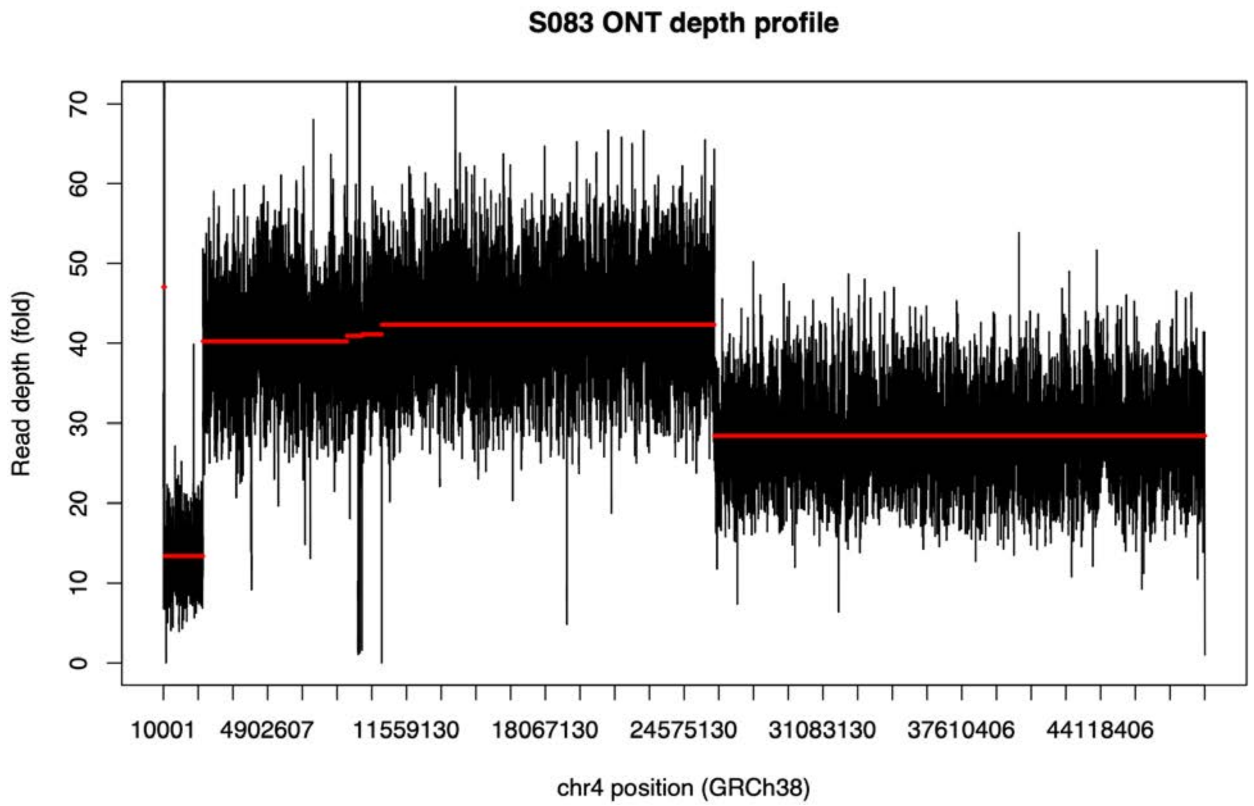
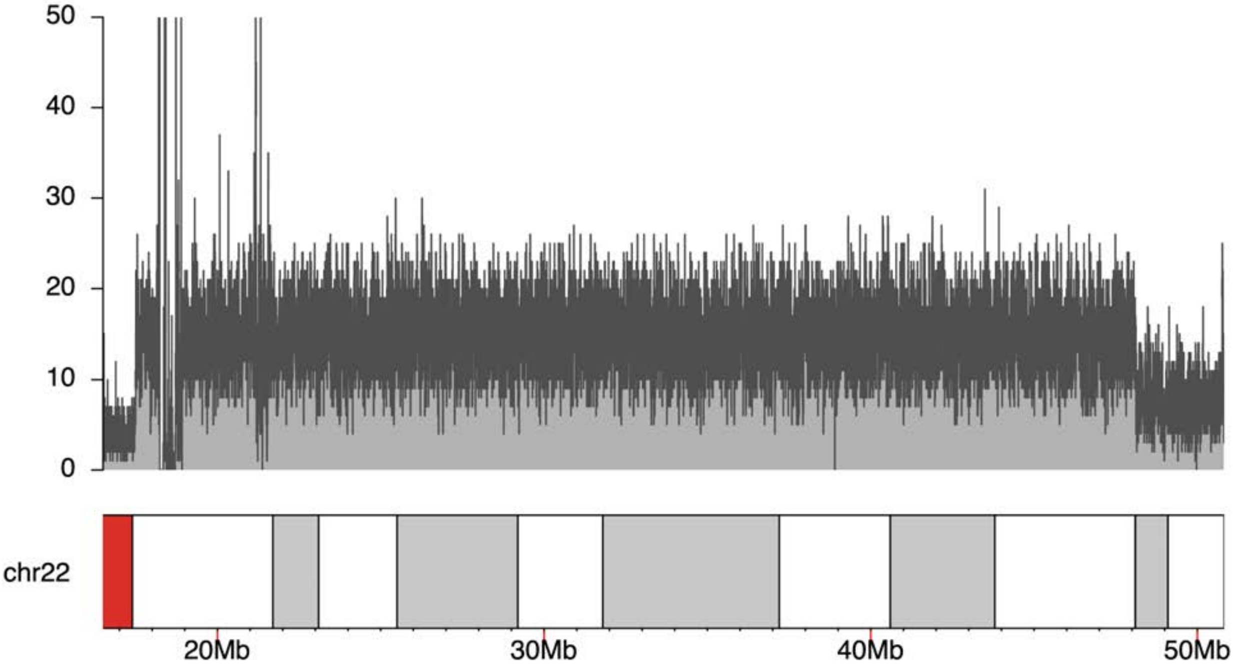
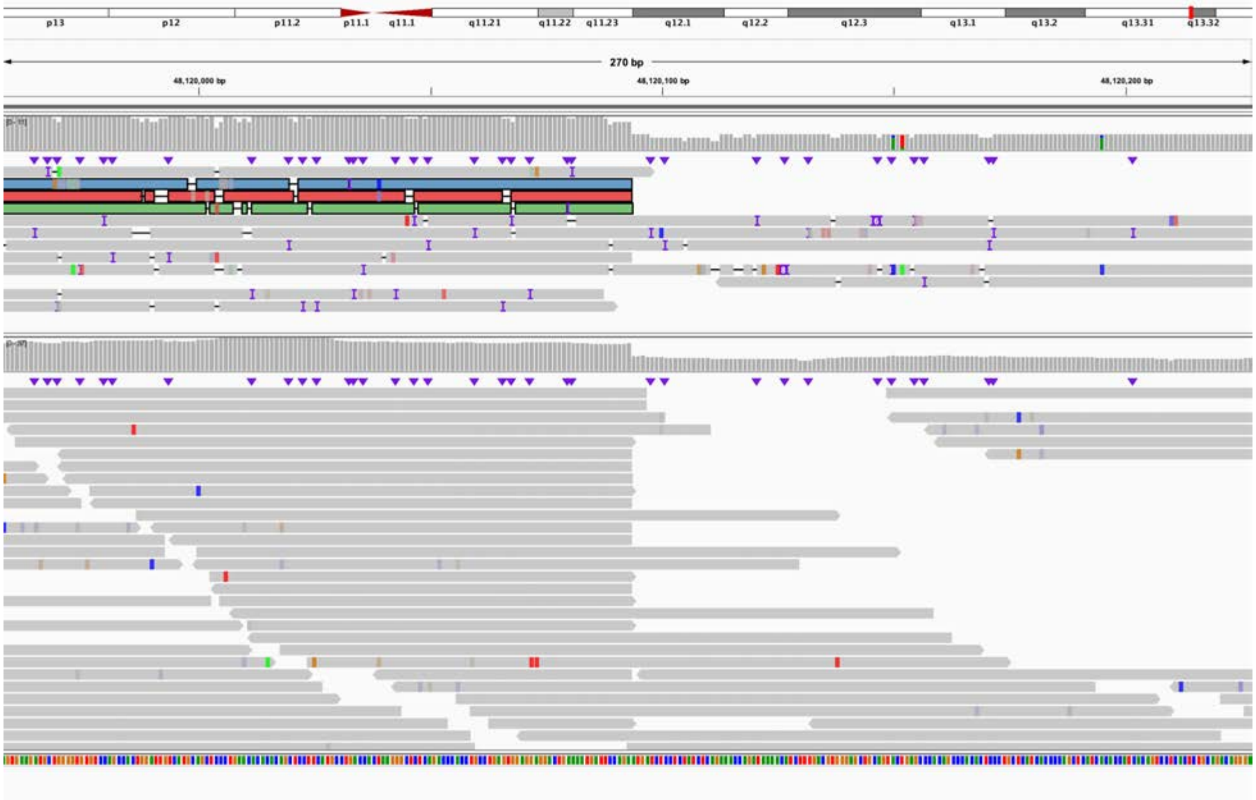

S083 ONT depth profile
Read depth (fold)
0
10
20
30
40
50
60
70
10001
4902607
11559130
18067130
24575130
31083130
37610406
44118406
chr4 position (GRCh38)

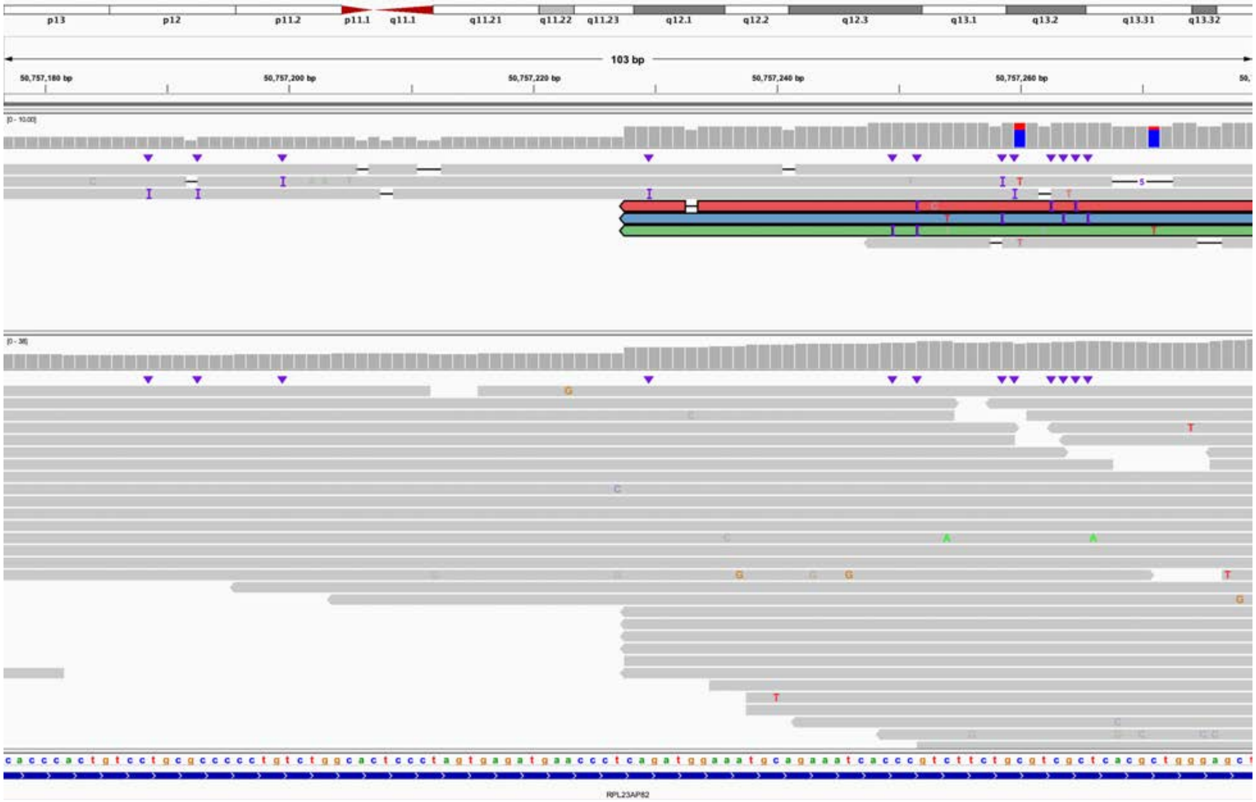
Figure S2. BK144-03, known 22q13.3 deletion.



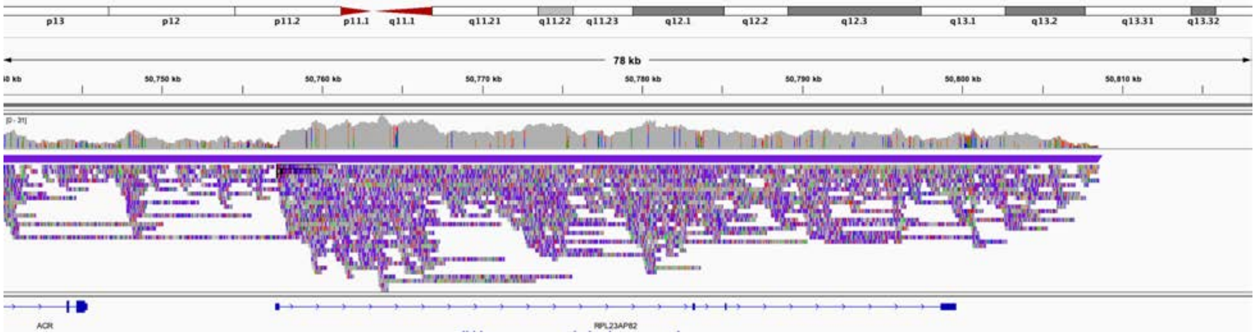
A. Coverage of the target region.



B. Centromere proximal end of deletion, long-read BAM file is top track, short-read BAM is bottom track. Long reads that span the deletion breakpoint are highlighted with color; IGV view is chr22:48,119,958-48,120,228.

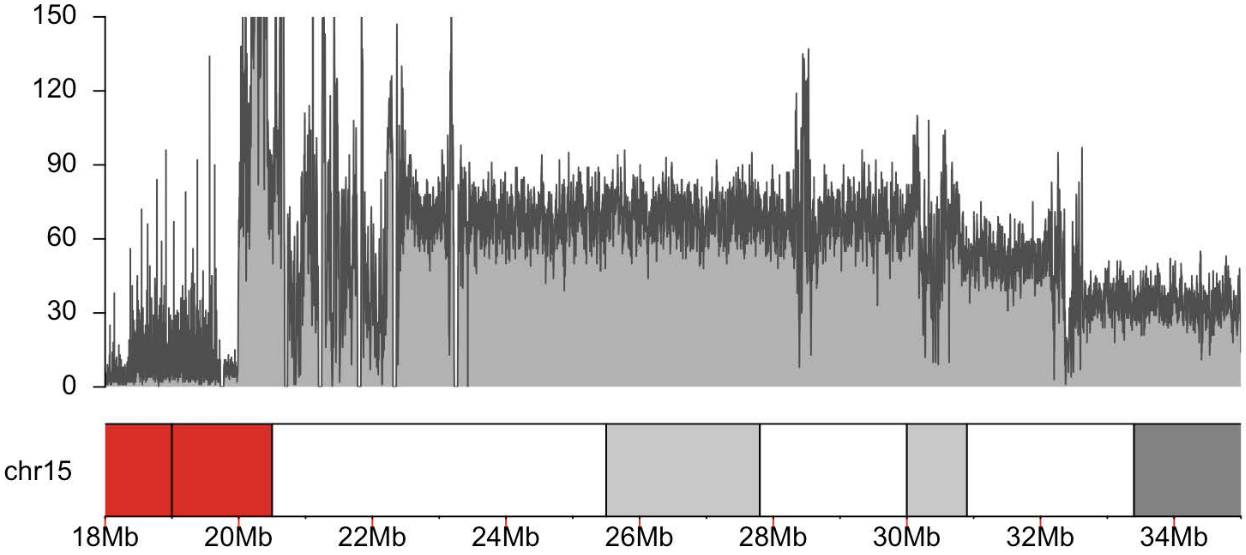


C. Telomere proximal end of deletion, long-read BAM file is the top track while short-read file is the bottom track. Long reads that span the deletion are colored as in (B); IGV view is chr22:50,757,177-50,757,279.



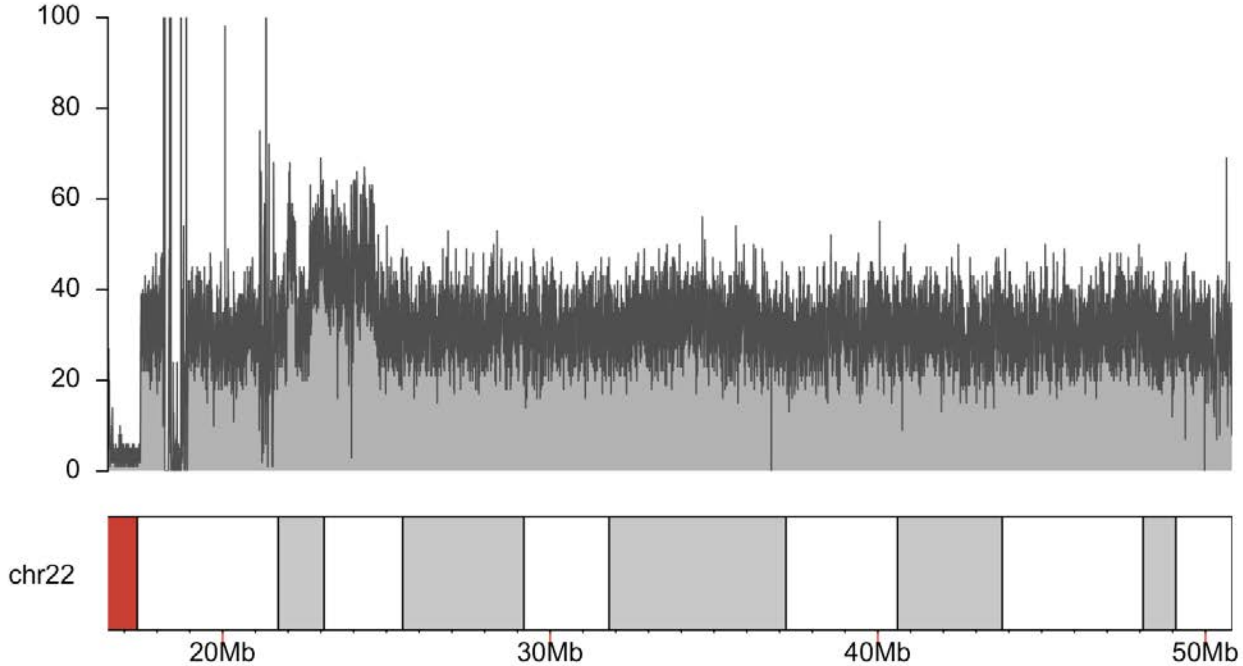
D. IGV view of only long reads at the end of chromosome 22; view is chr22:50,739,987-50,818,468.

Figure S3. BK180-03, known 15q11-q13 duplication.



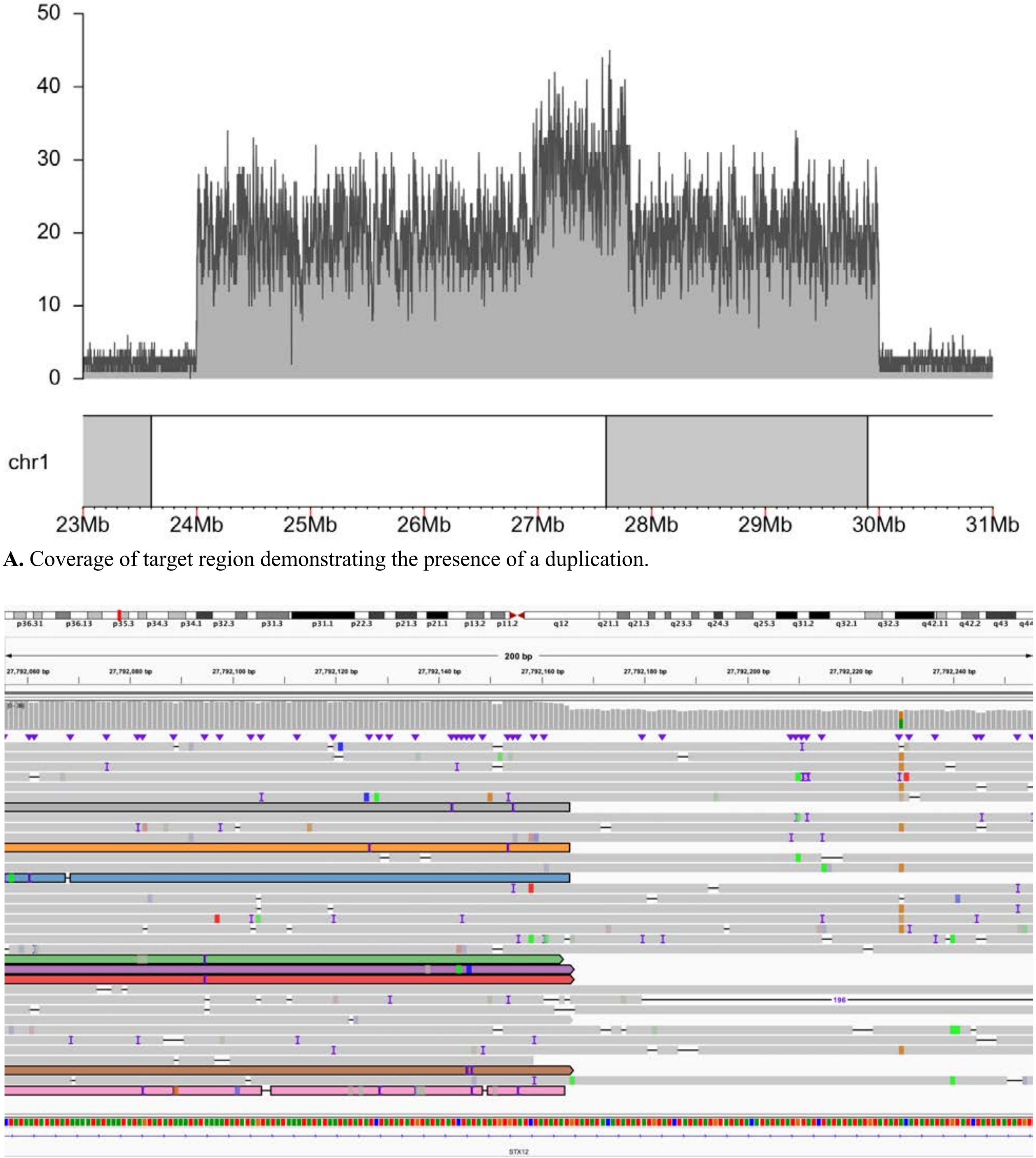
Coverage of target region in 15q. No definitive duplication breakpoint was found using targeted long reads.

Figure S4. BK294-03, known 22q11.2 duplication.



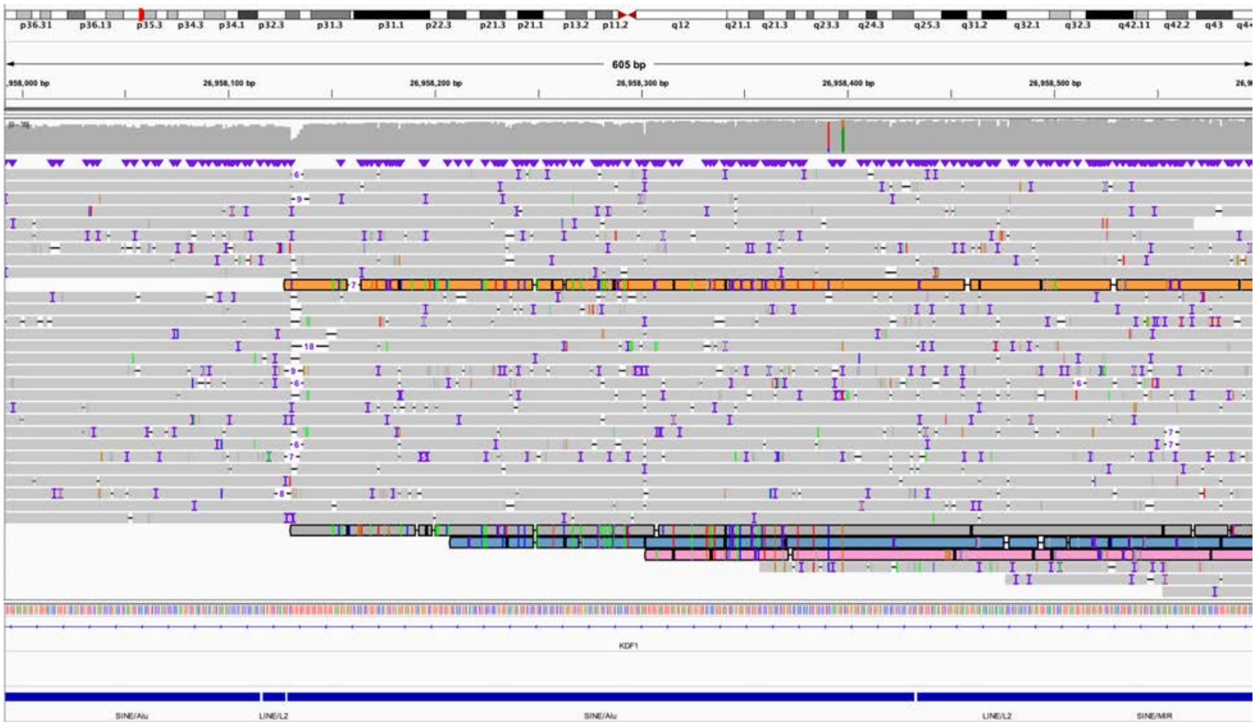
Coverage of chromosome 22 target region. No definitive duplication breakpoint was found using targeted long reads.

Figure S5. BK364-03, known 1p36.11 duplication.

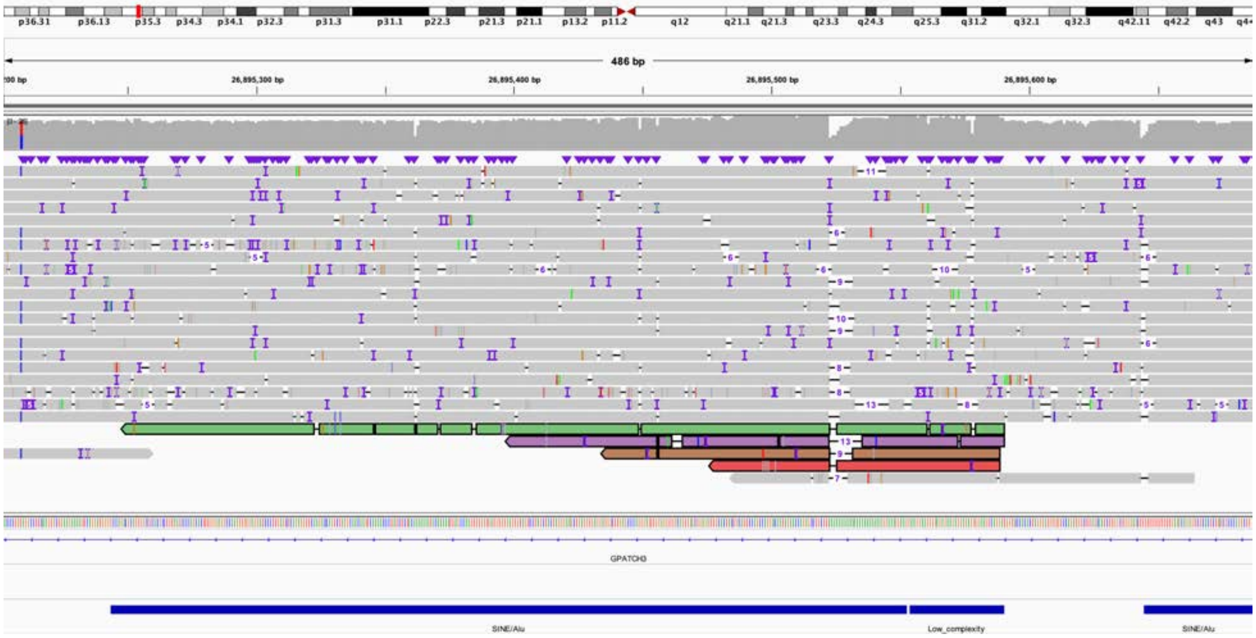


A. Coverage of target region demonstrating the presence of a duplication.

B. IGV view of 3' end, or centromere-proximal end of the duplication, reads that define the deletion breakpoint are represented by colors. IGV view is of chr1:27,792,056-27,792,256.

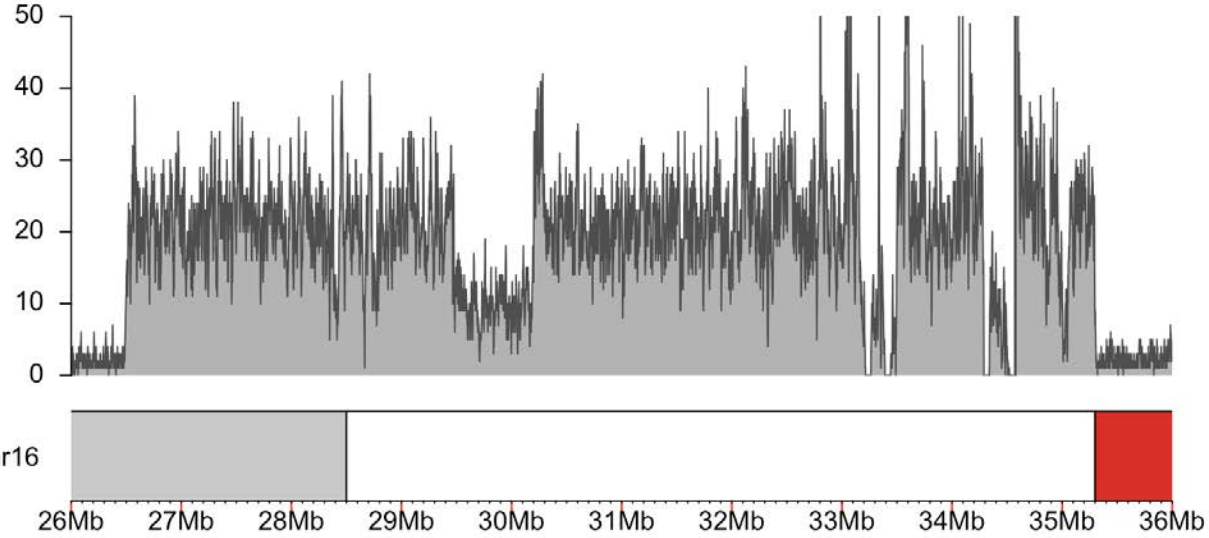


C. Reads from B are split and align to the 5' end, or telomere-proximal end of the duplication. The region contains several repetitive elements, as shown in the bottom track. IGV view is of chr1:26,956,878-26,959,178.



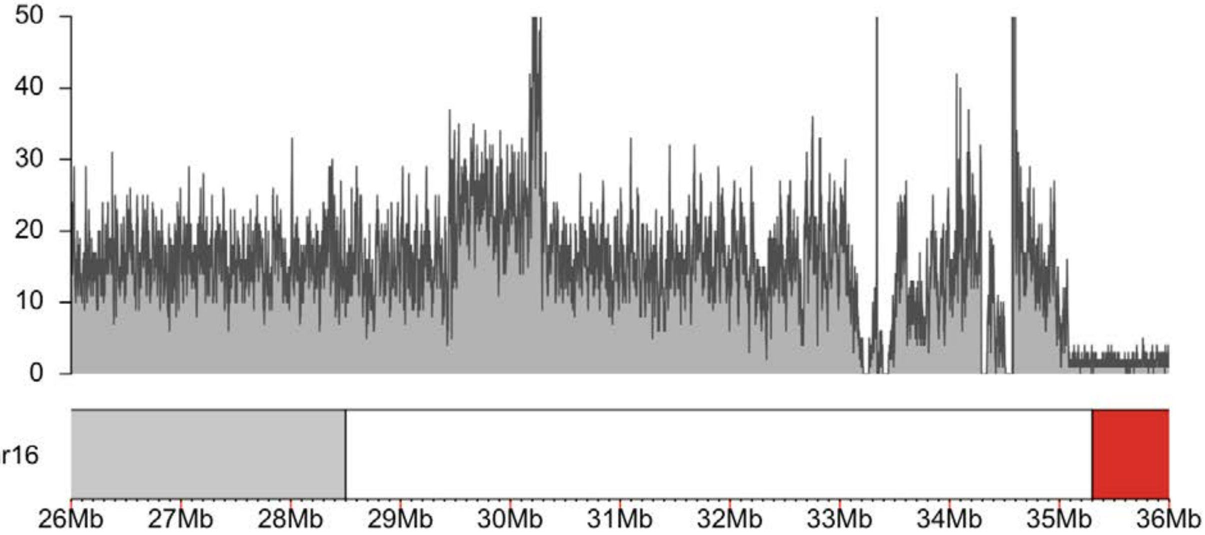
D. Fragments of four reads seen in B align incorrectly to a region outside of the duplicated region that includes a SINE/Alu and low-complexity region (bottom track); these fragments are notably shorter than those in C. IGV view is of chr1:26,895,202-26,895,689.

Figure S6. BK397-101, known 16p11.2 deletion.



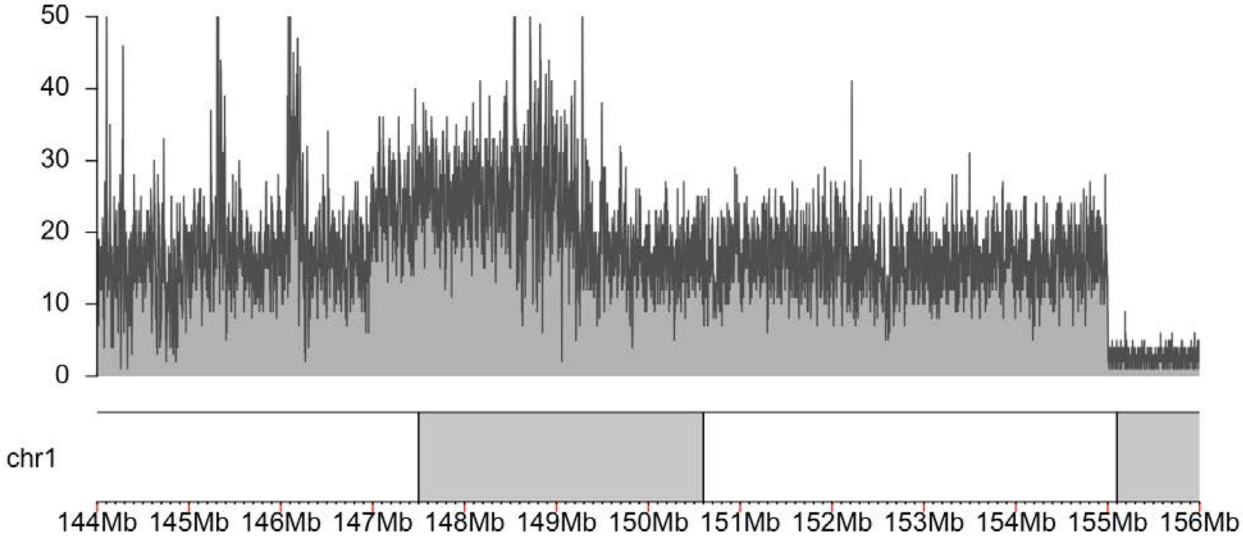
Coverage of chromosome 16 target region. No definitive deletion breakpoint was found using targeted long reads.

Figure S7. BK430-103, known 16p11.2 duplication.



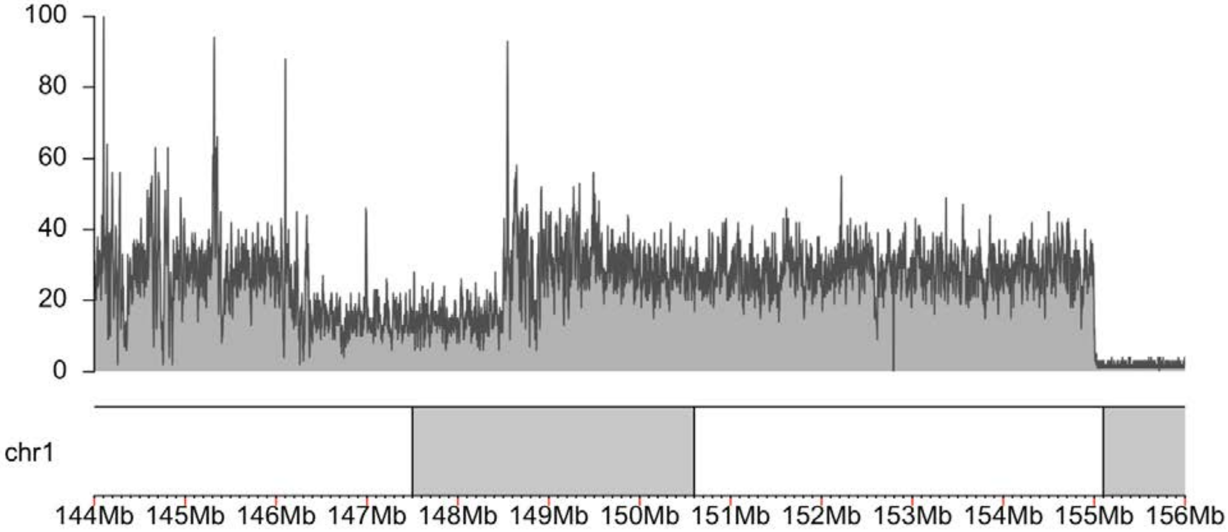
Coverage of chromosome 16 target region. No definitive duplication breakpoint was found using targeted long reads.

Figure S8. BK482-101, known 1q21.1 duplication.



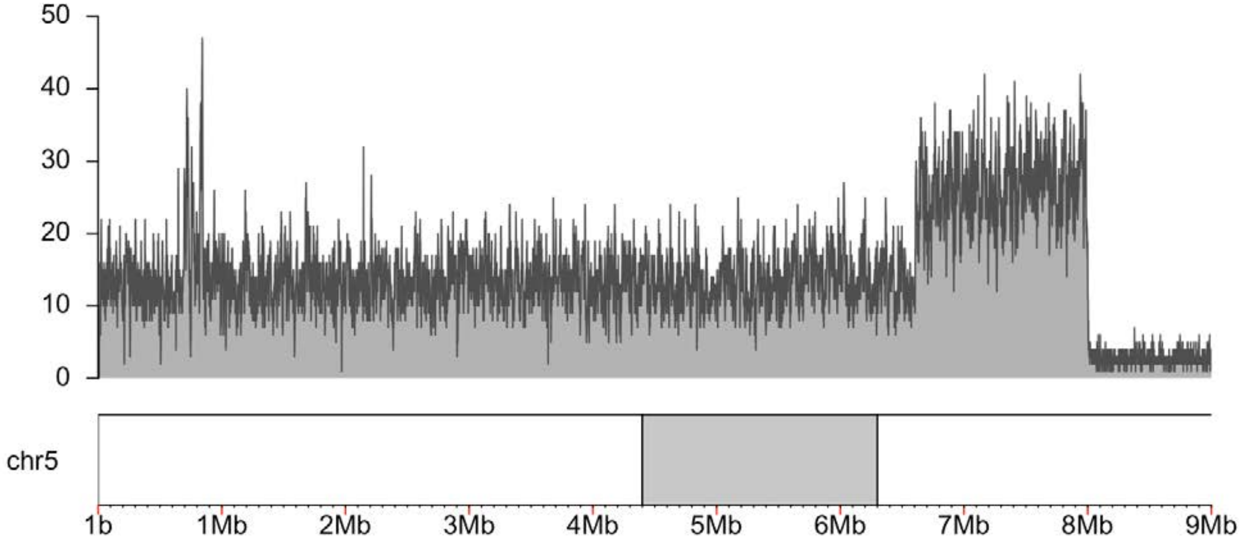
Coverage of chromosome 1 target region. No definitive duplication breakpoint was found using targeted long reads.

Figure S9. BK487-101, known 1q21 deletion.

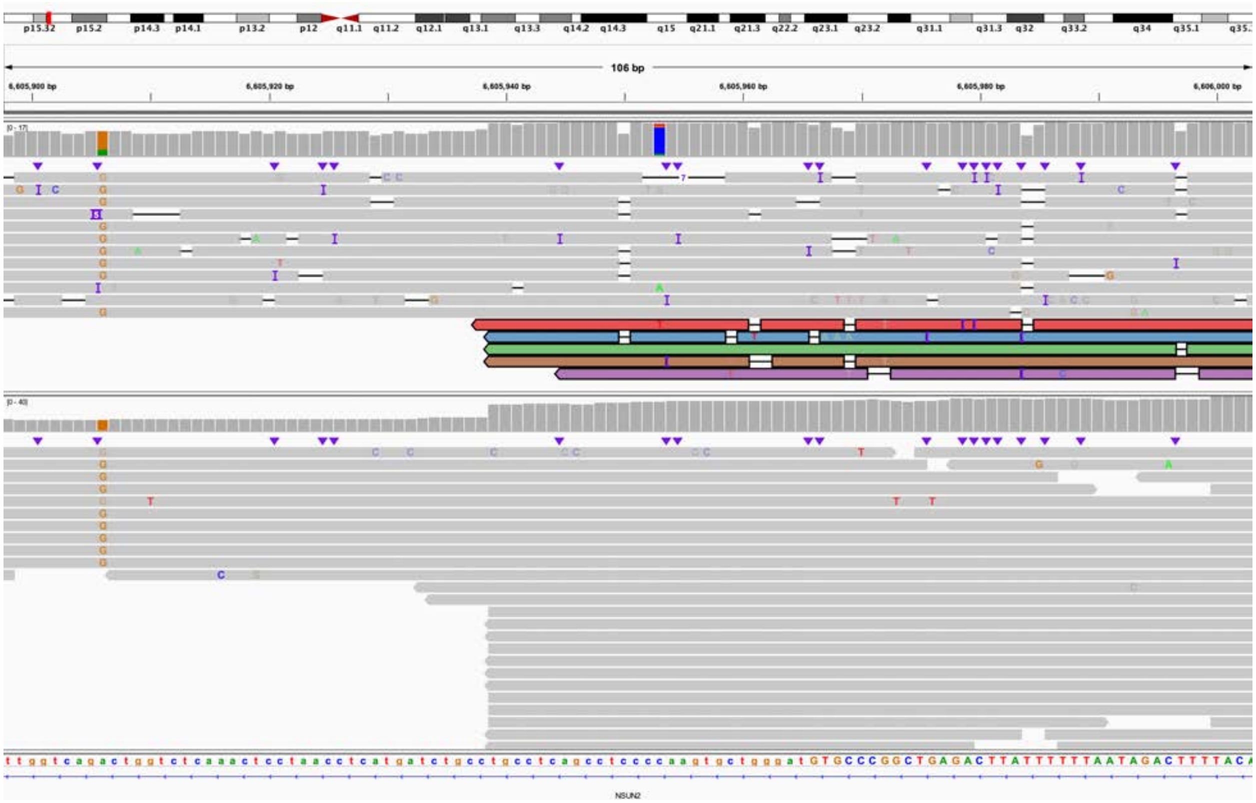


Coverage of chromosome 1 target region. No definitive deletion breakpoint was found using targeted long reads.

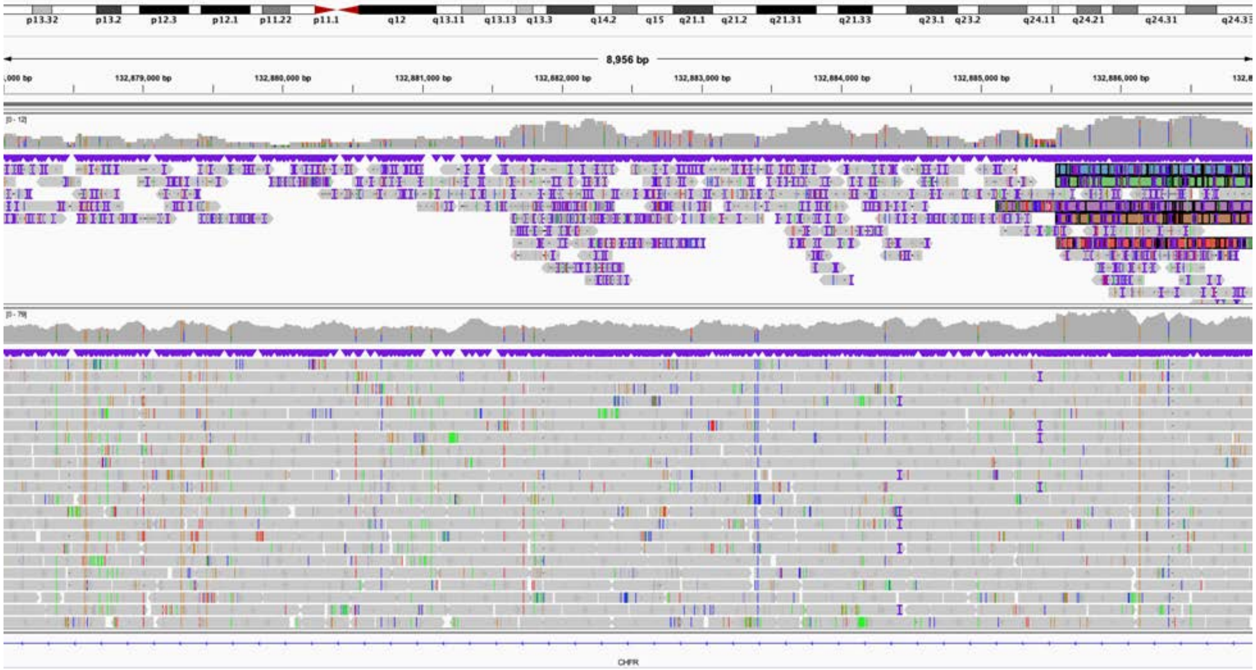
Figure S10. BK506-03, known 5p15.33 deletion.



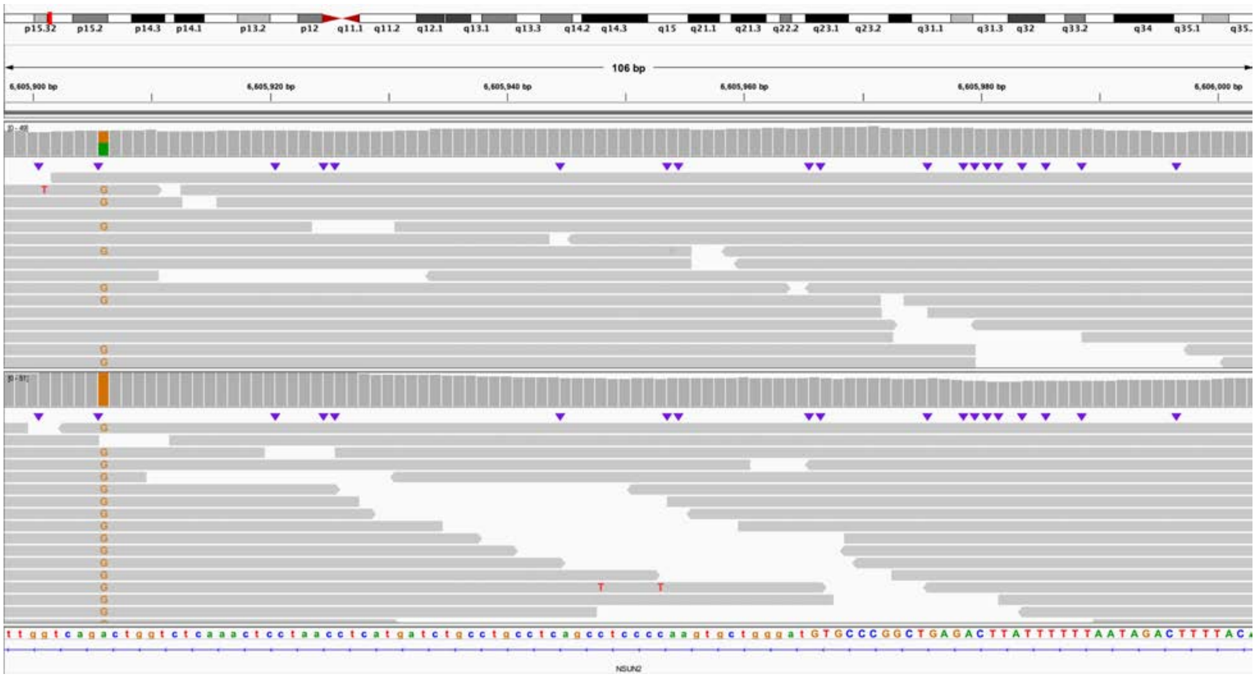
A. Coverage of chromosome 5 target region.



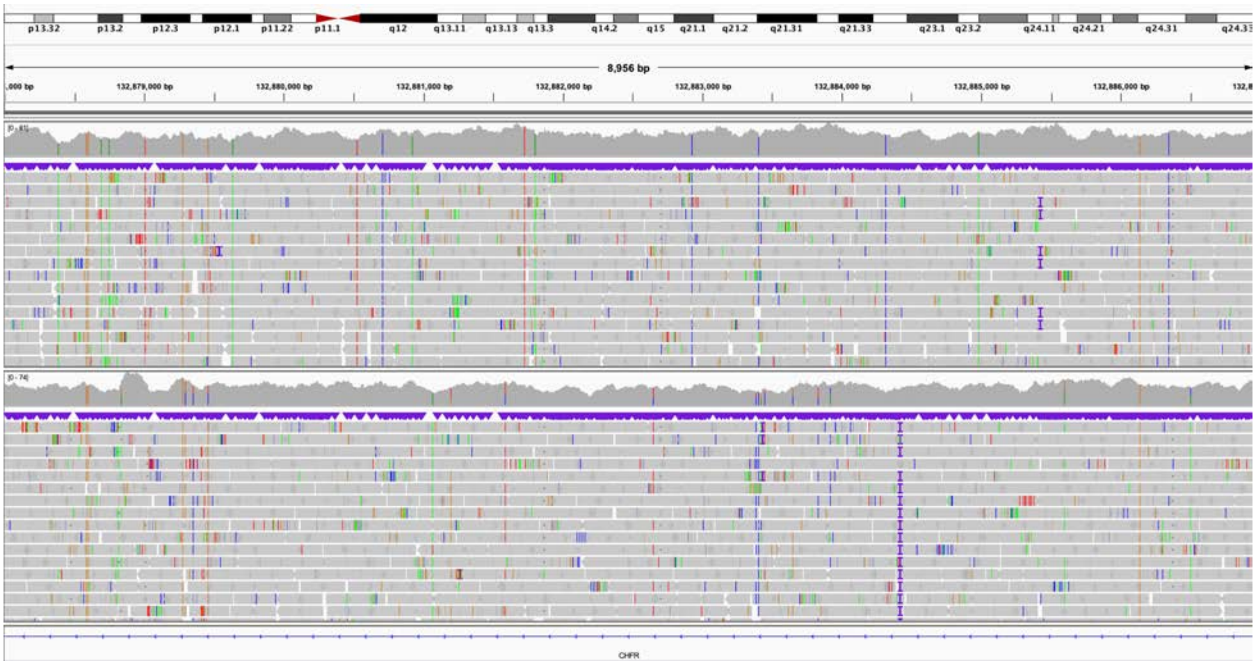
B. Breakpoint of the chromosome 5 deletion in IGV: long-read data is top track; short-read data is bottom track. Colors in top track correspond to colors of reads in C. View is of chr5:6,605,898-6,606,003.



C. View of both long-read (top track) and short-read (bottom track) data from BK506-03 showing the centromere-proximal sequence is not deleted, but the telomere-proximal sequence is present in three copies. Long reads in the top track can be matched with those from B by their color. IGV view is chr12:132,877,990-132,886,995.

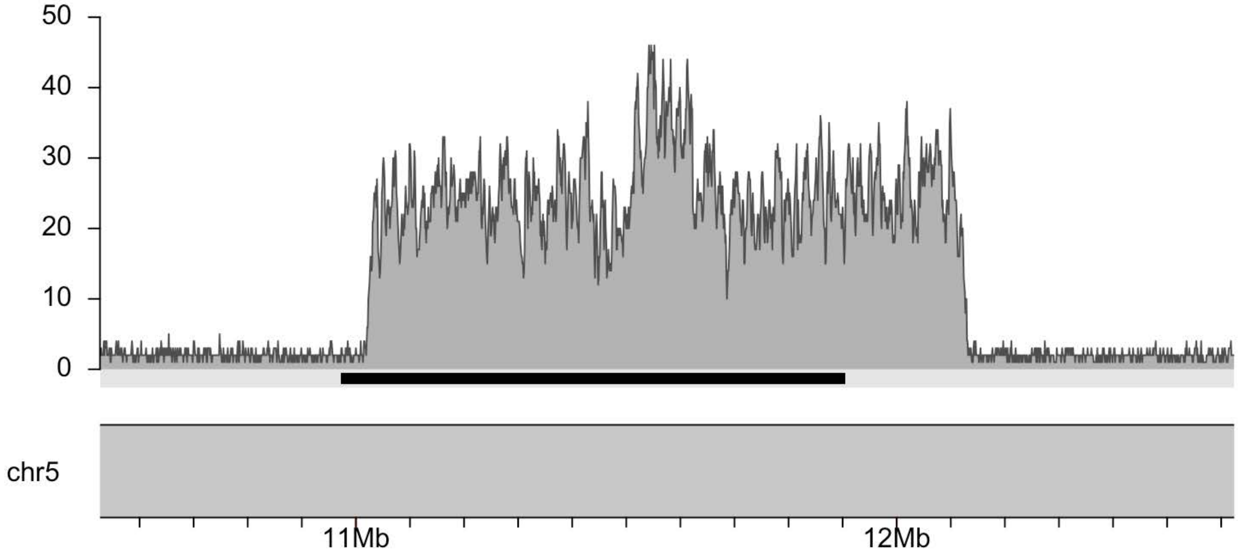

p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q11.2 q12.1 q13.1 q13.3 q14.2 q14.3 q15 q21.1 q21.3 q22.2 q23.1 q23.2 q31.1 q31.3 q32 q33.2 q34 q35.1 q35.
106 bp
6,605,900 bp
6,605,920 bp
6,605,940 bp
6,605,960 bp
6,605,980 bp
6,606,000 bp
[0 - 48]
[0 - 51]
NSUN2

D. View of short-read sequencing data at the chromosome 5 breakpoint from the parents of BK506-03 (BK506-01 and BK506-02). IGV view is chr5:6,605,898-6,606,003.

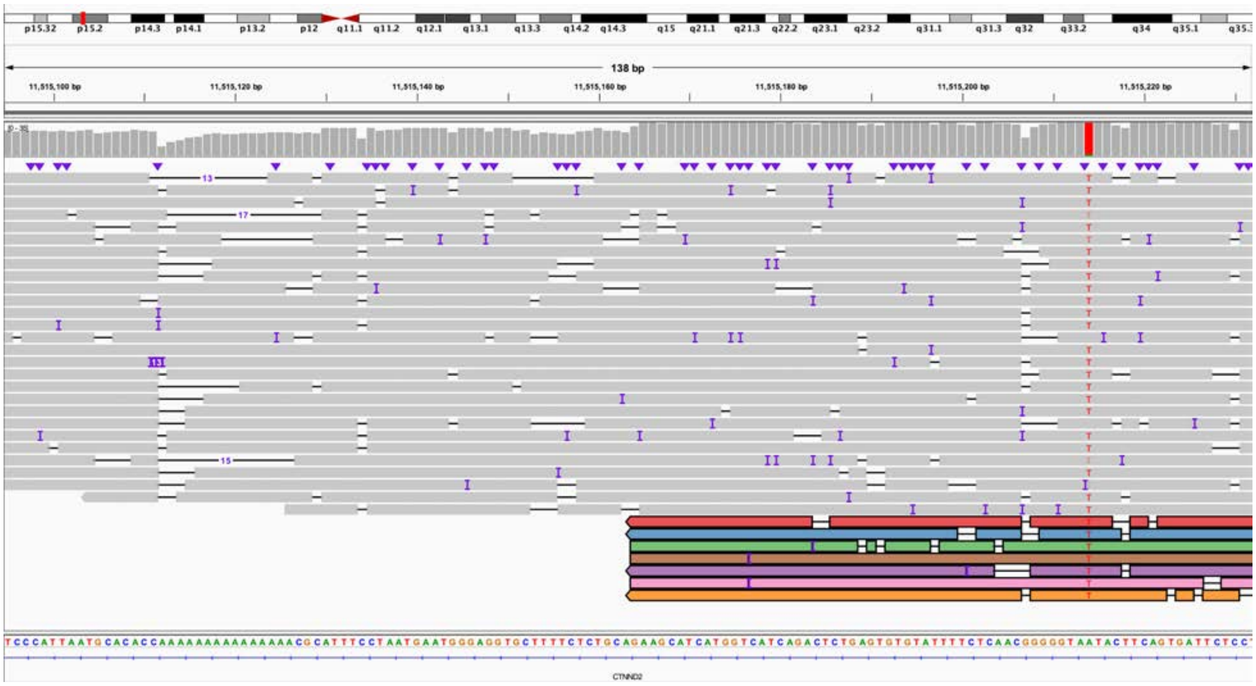


E. View of short-read sequencing data at the chromosome 12 breakpoint from the parents of BK506-03. IGV view is chr12:132,877,990-132,886,995.

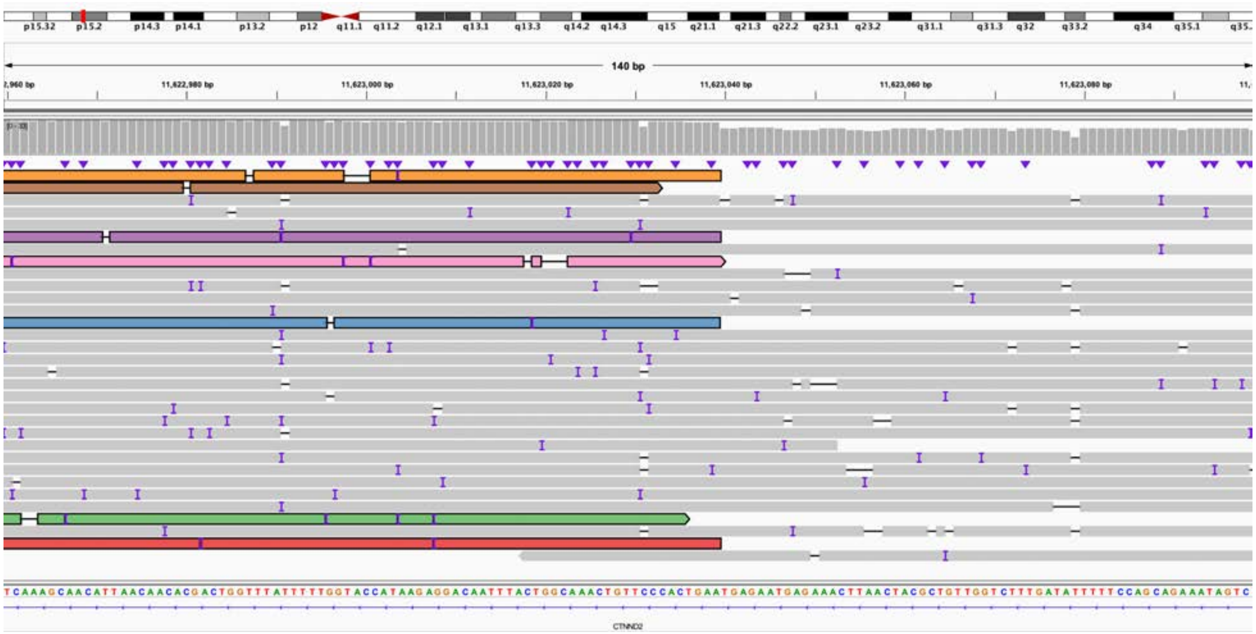
Figure S11. S016, known tandem duplication within *CTNND2*.



A. Coverage of duplication within *CTNND2*, gene body of *CTNND2* is represented by the black bar below the coverage graph.

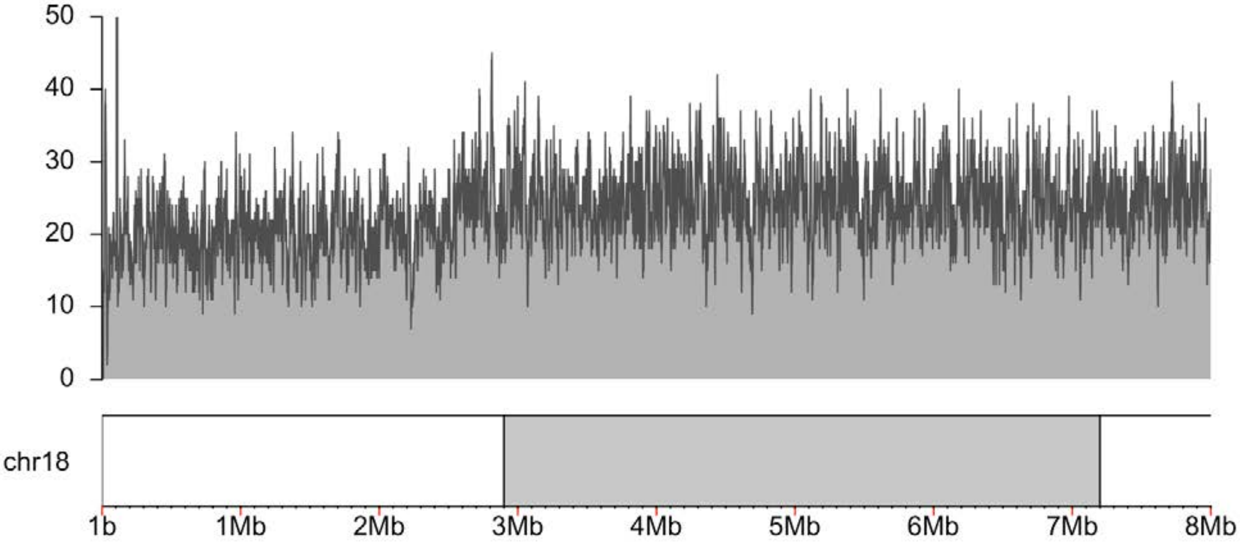


B. Reads in the telomere-proximal end of the duplication, colors correspond to reads in C. The orientation of reads in B and C confirm that it is a tandem duplication as previously shown.[1] IGV view of chr5:11,515,095-11,515,232.

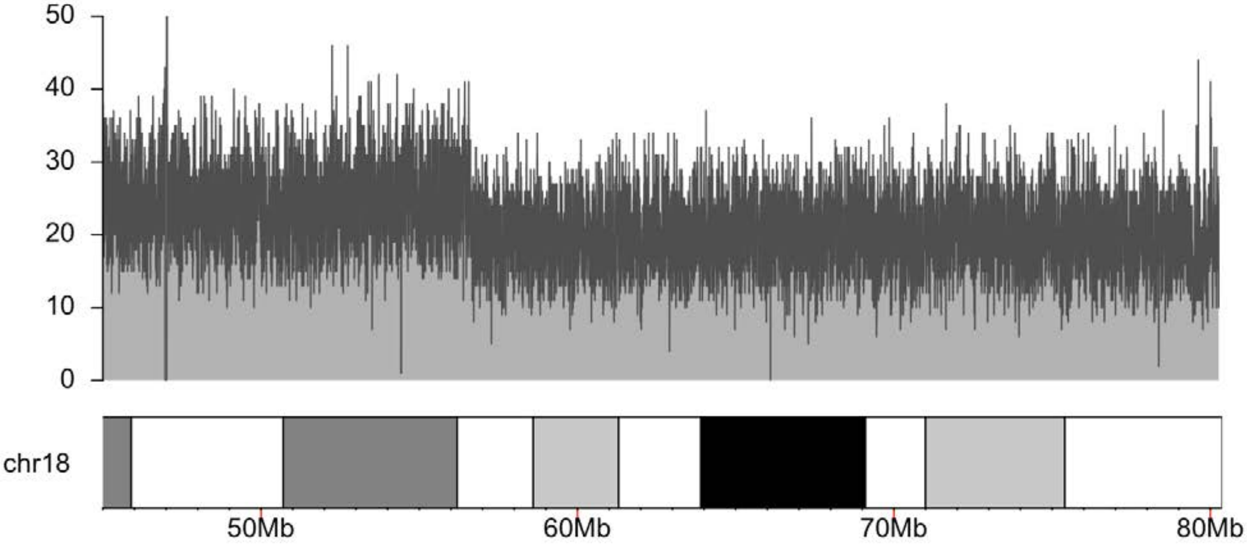


C. Reads from the centromere-proximal end of the duplication, colors correspond to reads in B. IGV view is of chr5:11,622,960-11,623,099.

Figure S12. S023, mosaic ring 18 present in 40% of cells.

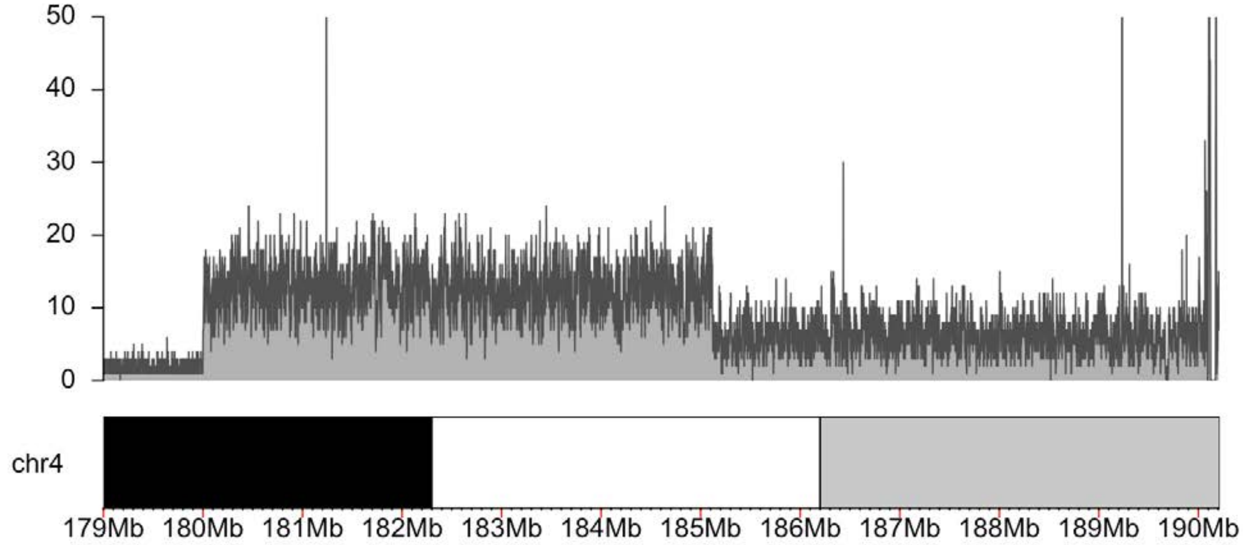


A. Coverage of chromosome 18p target region shows mosaic copy loss.

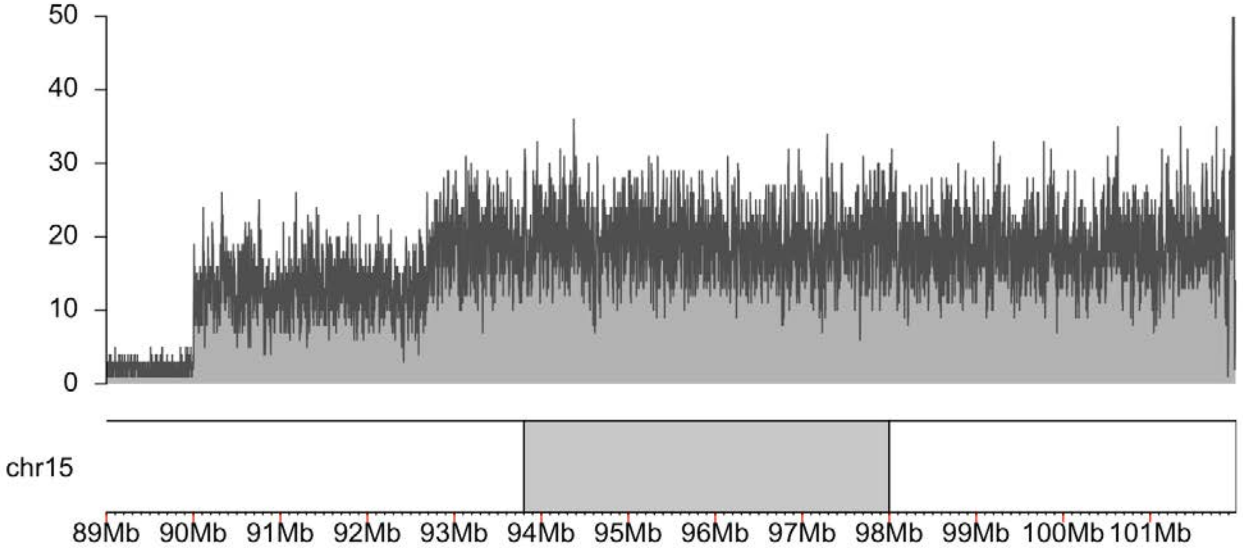


B. Coverage of chromosome 18q target region shows mosaic copy loss.

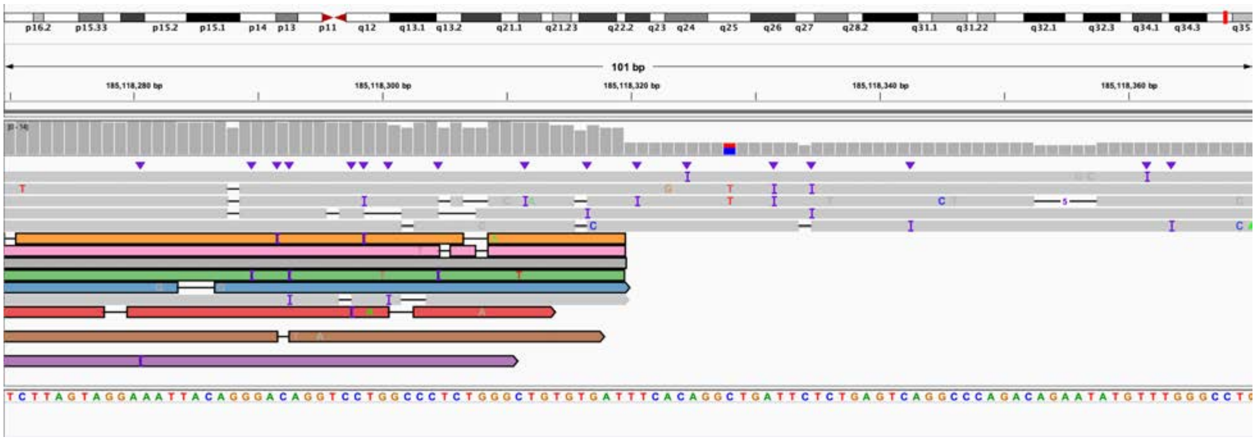
Figure S13. S046, known unbalanced translocation between chromosomes 4 and 15.



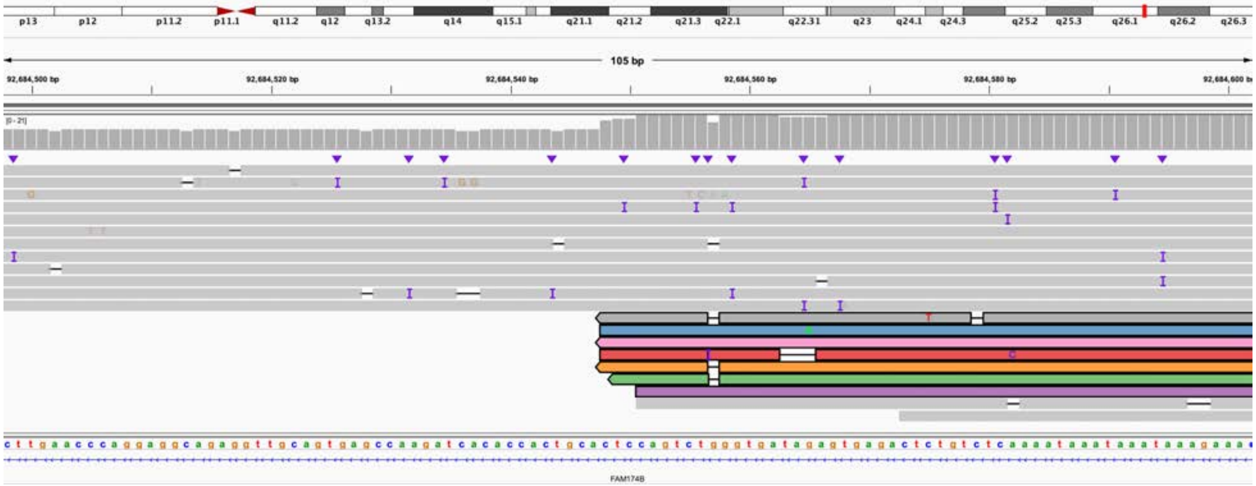
A. Coverage of chromosome 4 target region showing a distal deletion.



B. Coverage of chromosome 15 target region showing a distal gain.

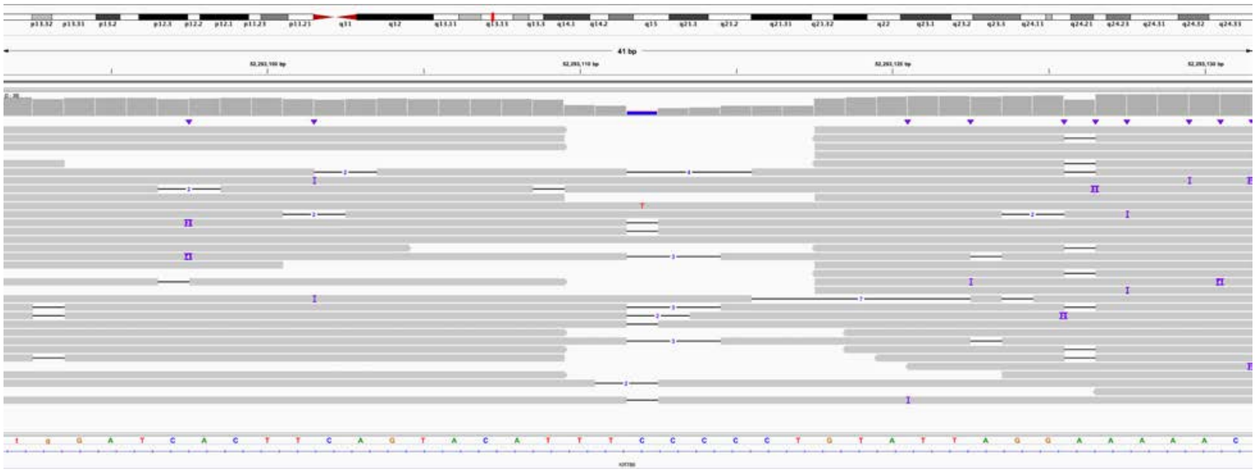

p16.2 p15.33 p15.2 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
101 bp
185,118,280 bp
185,118,300 bp
185,118,320 bp
185,118,340 bp
185,118,360 bp
[0 - 14]
TCTTAGTAGGAAATTACAGGGACAGGTCCTGGCCCTCTGGGCTGTGTGATTTCACAGGCTGATTCTCTGAGTCAGGCCCAGACAGAATATGTTTGGGCCTC

C. IGV view of reads spanning the translocation breakpoint. Reads are colored to correspond to reads in D. IGV view is of chr4:185,118,270-185,118,370.

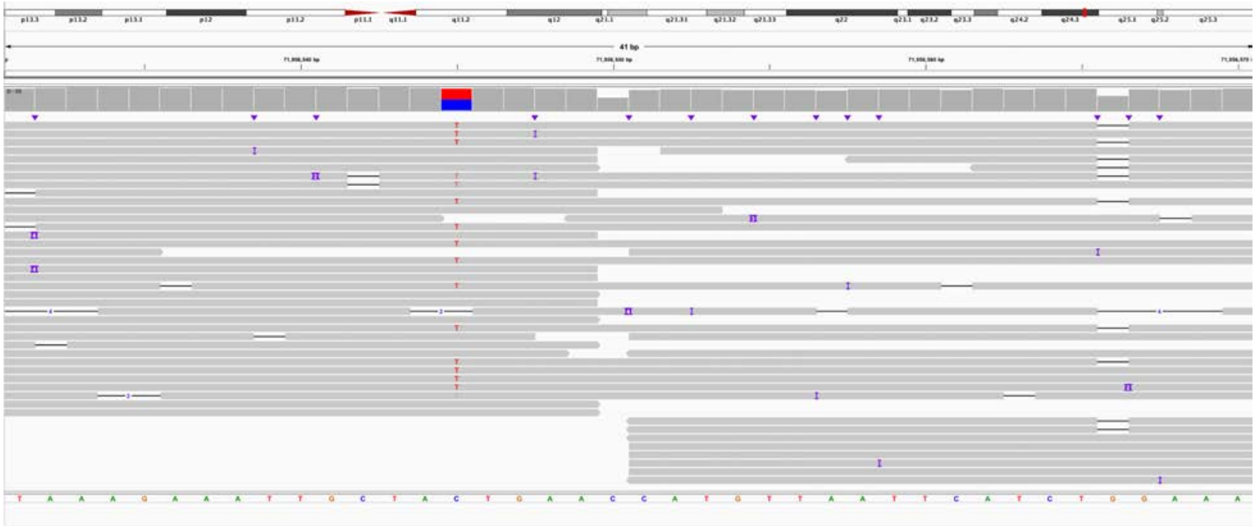

p13 p12 p11.2 p11.1 q11.2 q12 q13.2 q14 q15.1 q21.1 q21.2 q21.3 q22.1 q22.31 q23 q24.1 q24.3 q25.2 q25.3 q26.1 q26.2 q26.3
105 bp
92,684,500 bp
92,684,520 bp
92,684,540 bp
92,684,560 bp
92,684,580 bp
92,684,600 bp
[0 - 21]
cttgaacccaggaggcagaggttgcagtgagccaagatcacaccactgcactccagtctgggtgatagagtgagactctgtctcaaaataaataaataaagaaa
FAM174B

D. View of reads spanning the translocation breakpoint on chromosome 15. IGV view is of chr15:92,684,498-92,684,602.

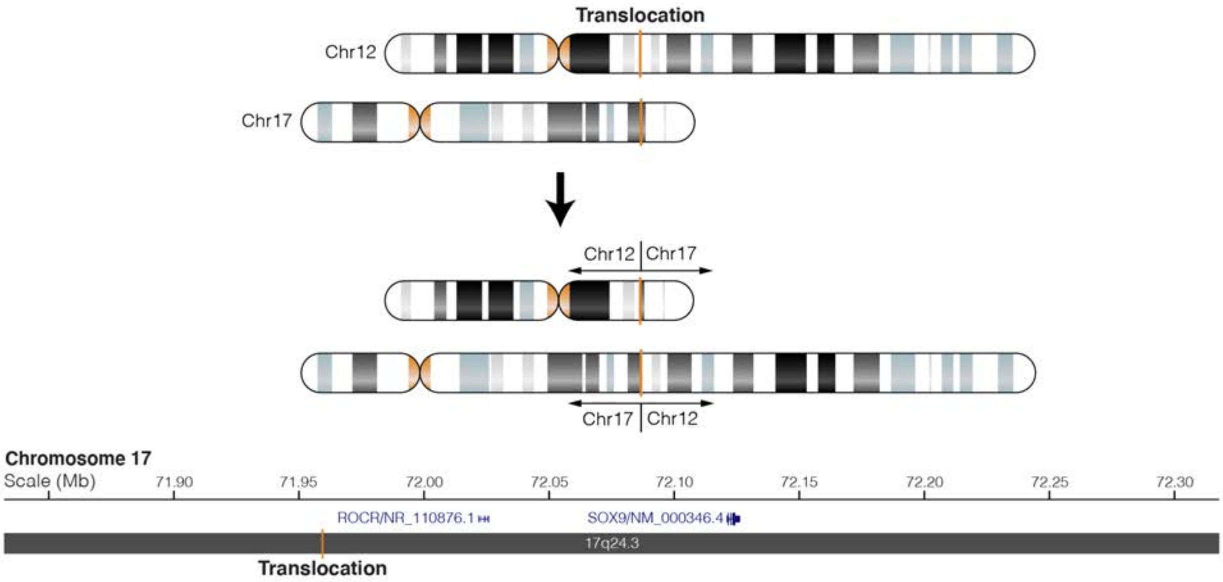
Figure S14. S060, known translocation between chromosomes 12 and 17 thought to affect *SOX9*.



A. Screenshot of chromosome 12 showing translocation breakpoints. IGV view is of chr12:52,293,092-52,293,131.

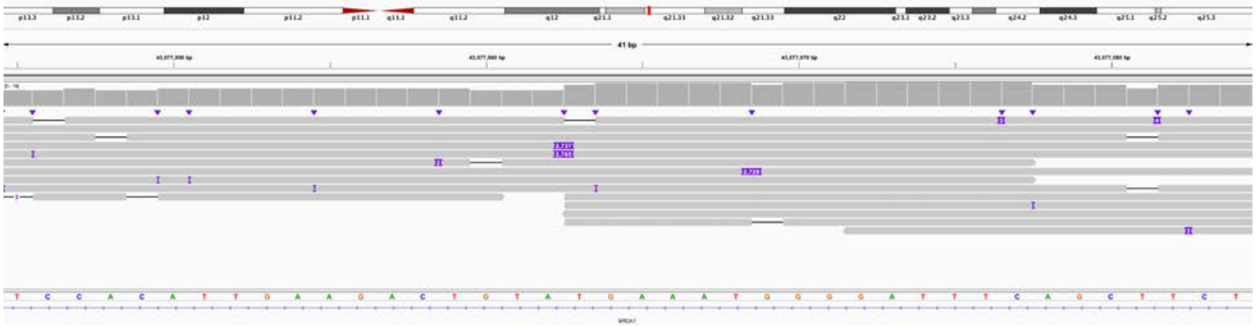


B. Screenshot of chromosome 17 showing translocation breakpoints. IGV view is of chr17:71,956,531-71,956,570.



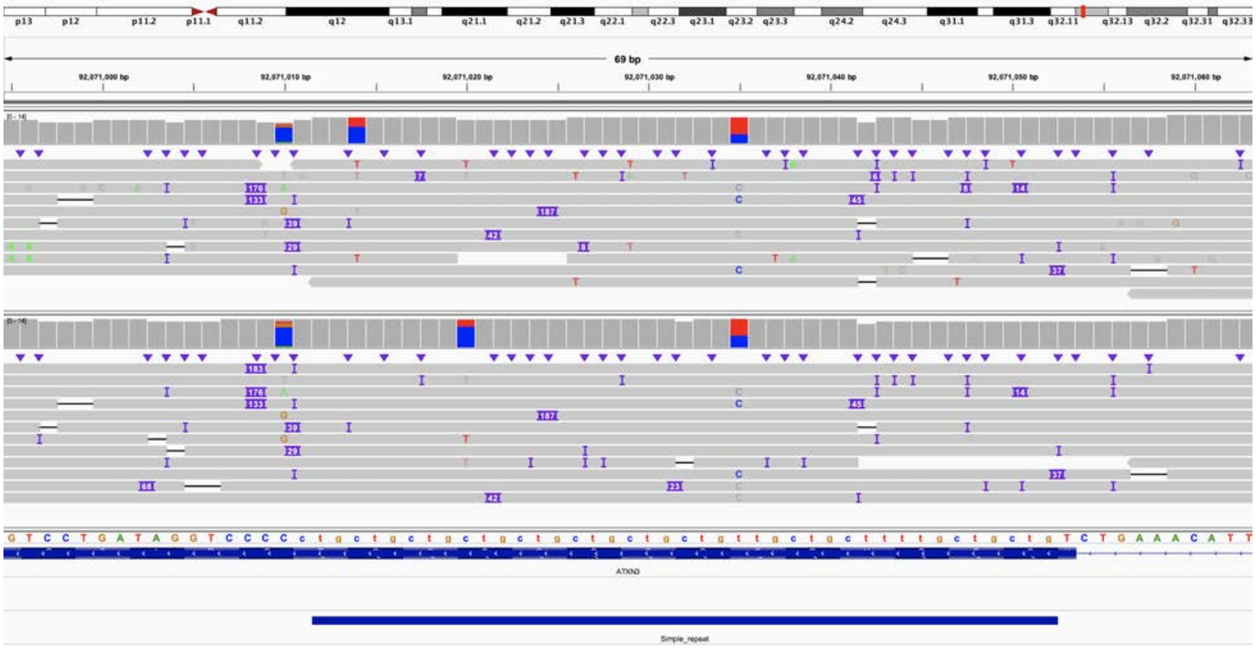
C. Cartoon of chromosome 12 and 17 translocations and breakpoints.

Figure S15. S063, known SVA insertion in *BRCA1*.

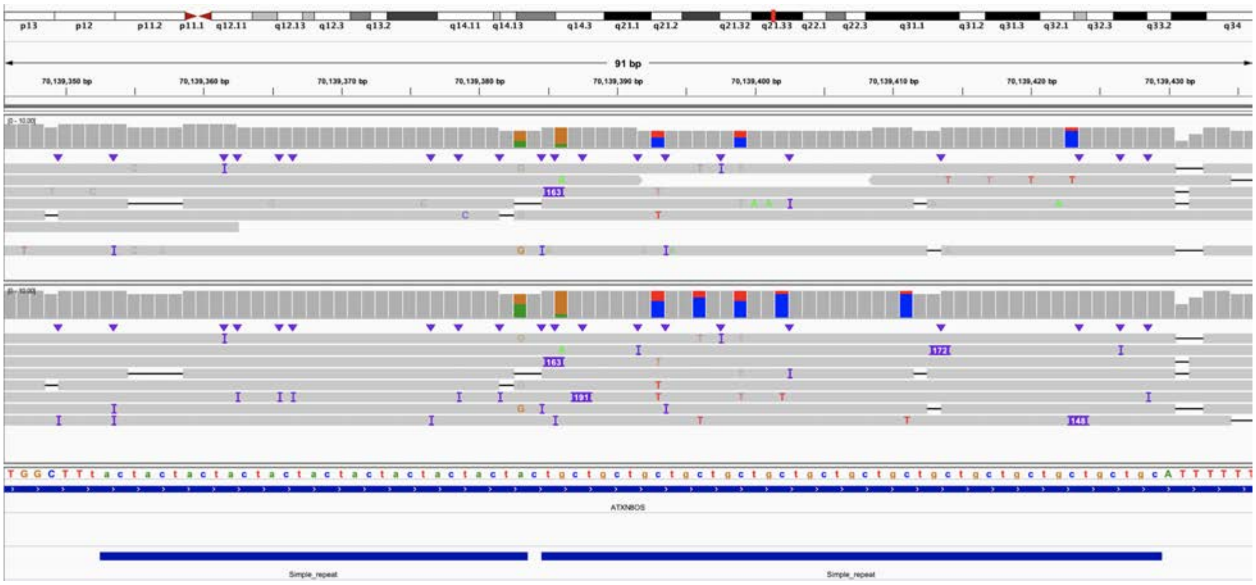


IGV view of known SVA insertion into *BRCA1*. IGV view is of chr17:43,077,045-43,077,084. This insertion was previously reported by Walsh and colleagues.[2]

Figure S16. S011 (*ATXN3* and *ATXN8OS*), evaluation of repeat length in an individual.

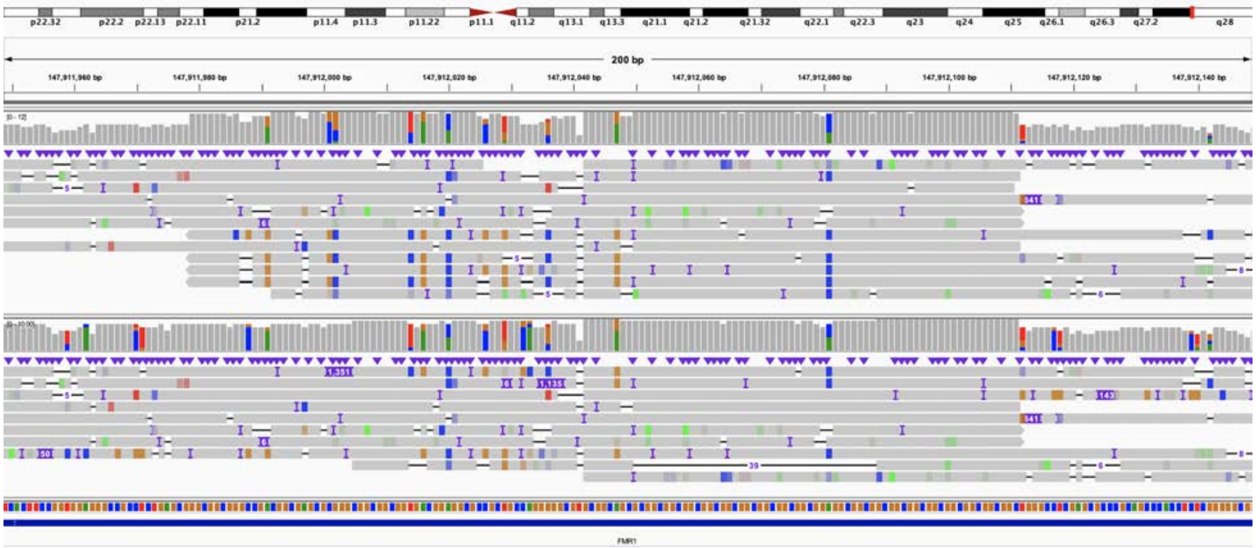


A. T-LRS of an individual known to carry a heterozygous repeat expansion in *ATXN3*. Three reads carry the expansion and span the region. IGV view of chr14:92,070,974-92,071,079. For both A and B reads aligned to the reference using minimap2 with default parameters is the top panel and reads aligned with modified parameters as outlined in the methods is the bottom panel.

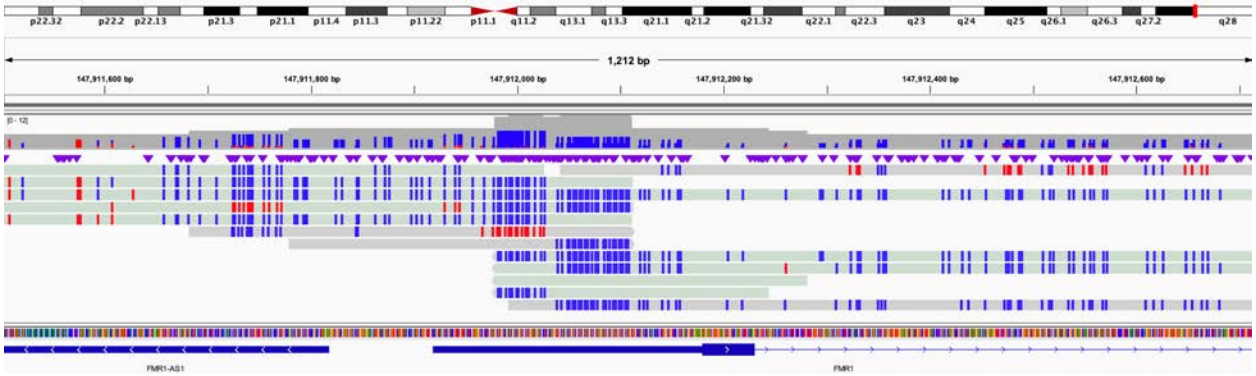


B. This individual was also known to carry a heterozygous expansion in *ATXNOS8*. In the minimap2 alignment with default parameters, a single read carries the full expansion and spans the region (top panel), but with modified parameters four reads can be found spanning the interval (bottom panel). IGV window is chr13:70,139,346-70,139,436.

Figure S17. S039 (*FMR1*), evaluation of repeat length and methylation status.

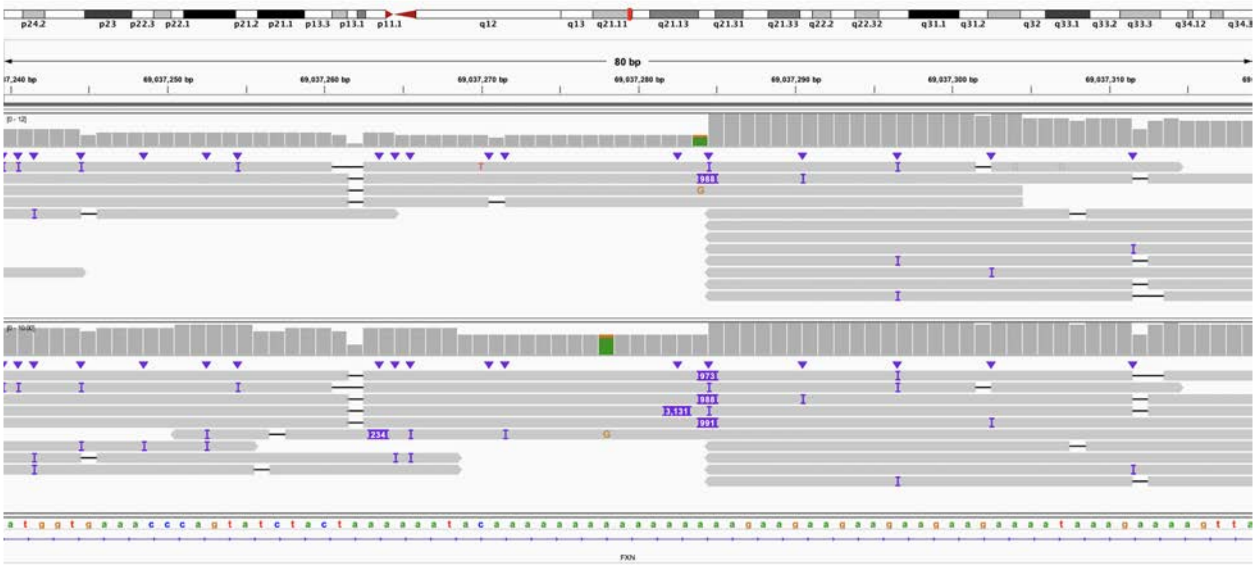


A. Screenshot of CGG repeat expansion in 5' UTR of *FMR1* from an individual known to carry an expansion. DNA was obtained from Coriell (Table S1). Top panel represents minimap2 run with default parameters; bottom panel is minimap2 run with parameters meant to reduce split reads. In the bottom panel, two reads with insert sizes of approximately 1,350 bp can be seen that are not present in the top panel. IGV view is chrX:147,911,949-147,912,149.



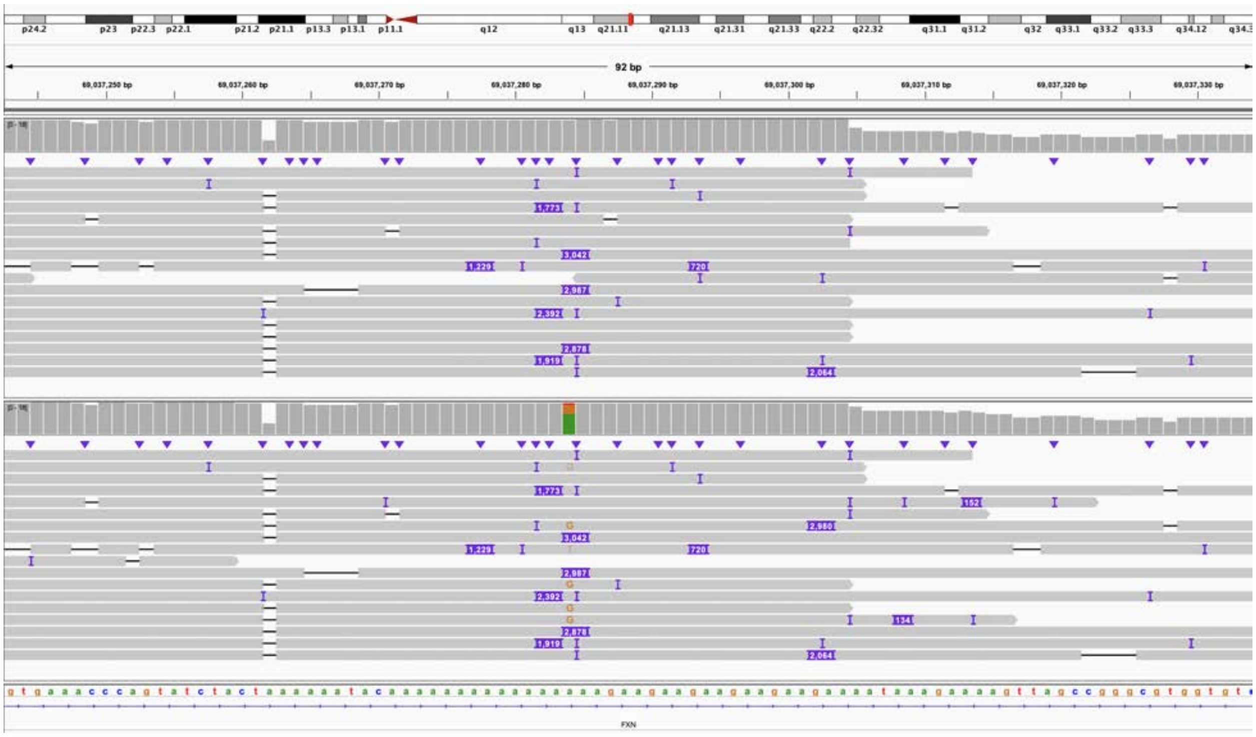
B. Methylation analysis of 5' UTR of *FMR1* reveals that the region around the CGG expansion is no longer methylated. The region shown is larger than in panel A. IGV view is chrX:147,911,502-147,912,719.

Figure S18. S040 (*FXN*), known repeat expansion.


p24.2 p23 p22.3 p22.1 p21.2 p21.1 p13.3 p13.1 p11.1 q12 q13 q21.11 q21.13 q21.31 q21.33 q22.2 q22.32 q31.1 q31.2 q32 q33.1 q33.2 q33.3 q34.12 q34.3
80 bp
7,240 bp
69,037,250 bp
69,037,260 bp
69,037,270 bp
69,037,280 bp
69,037,290 bp
69,037,300 bp
69,037,310 bp
[0 - 12]
988
[0 - 10.00]
973
988
3,131
991
234
a t g g t g a a a c c c a g t a t c t a c t a a a a a a t a c a a a a a a a a a a a a a a a a g a a g a a g a a g a a g a a g a a a a t a a a g a a a a g t t a
FXN

IGV view of a sample from an individual with a heterozygous expansion in *FXN*. Sample is from a cell line obtained from Coriell (Table S1). Top panel is reads aligned with minimap2 using default parameters; bottom panel is parameters that reduce the number of split reads. IGV view is from chr9:69,037,240-69,037,319.

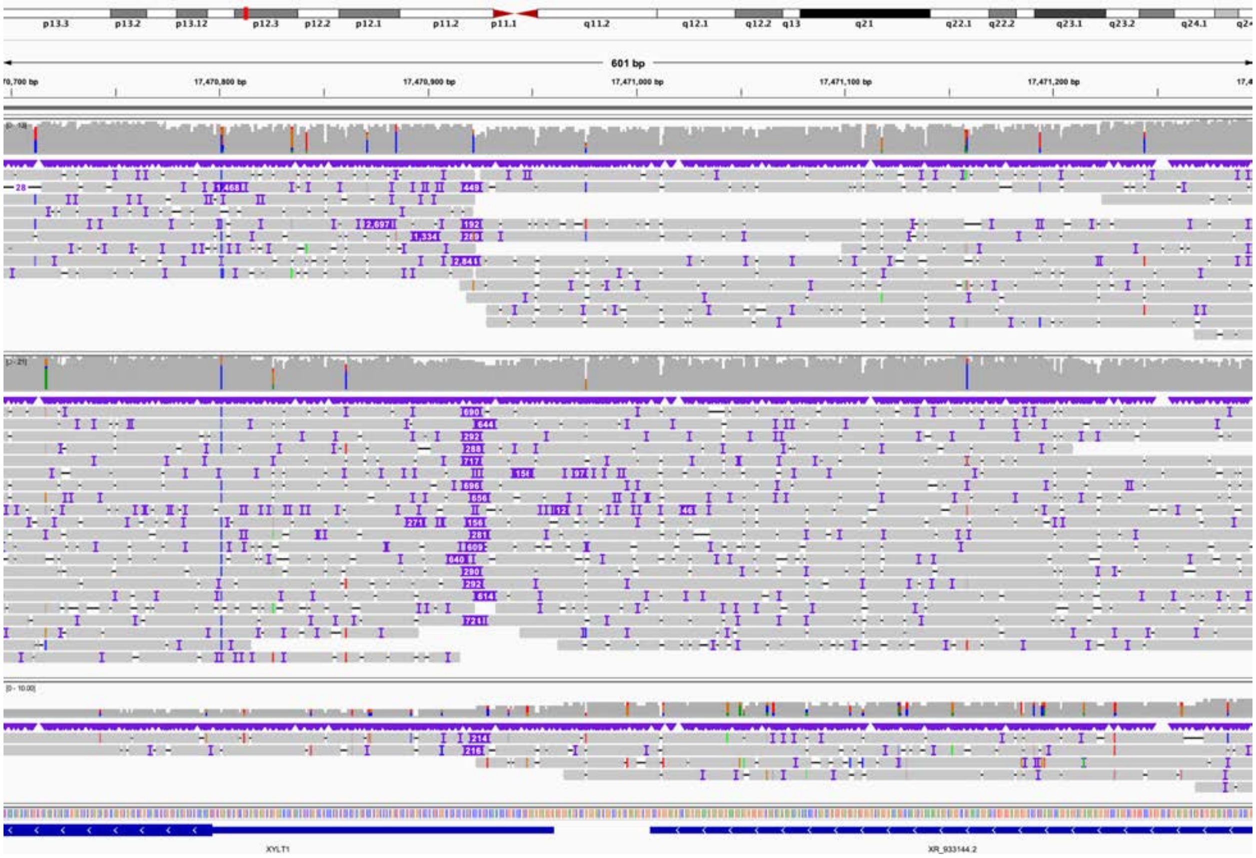
Figure S19. S041 (*FXN*), known repeat expansion.



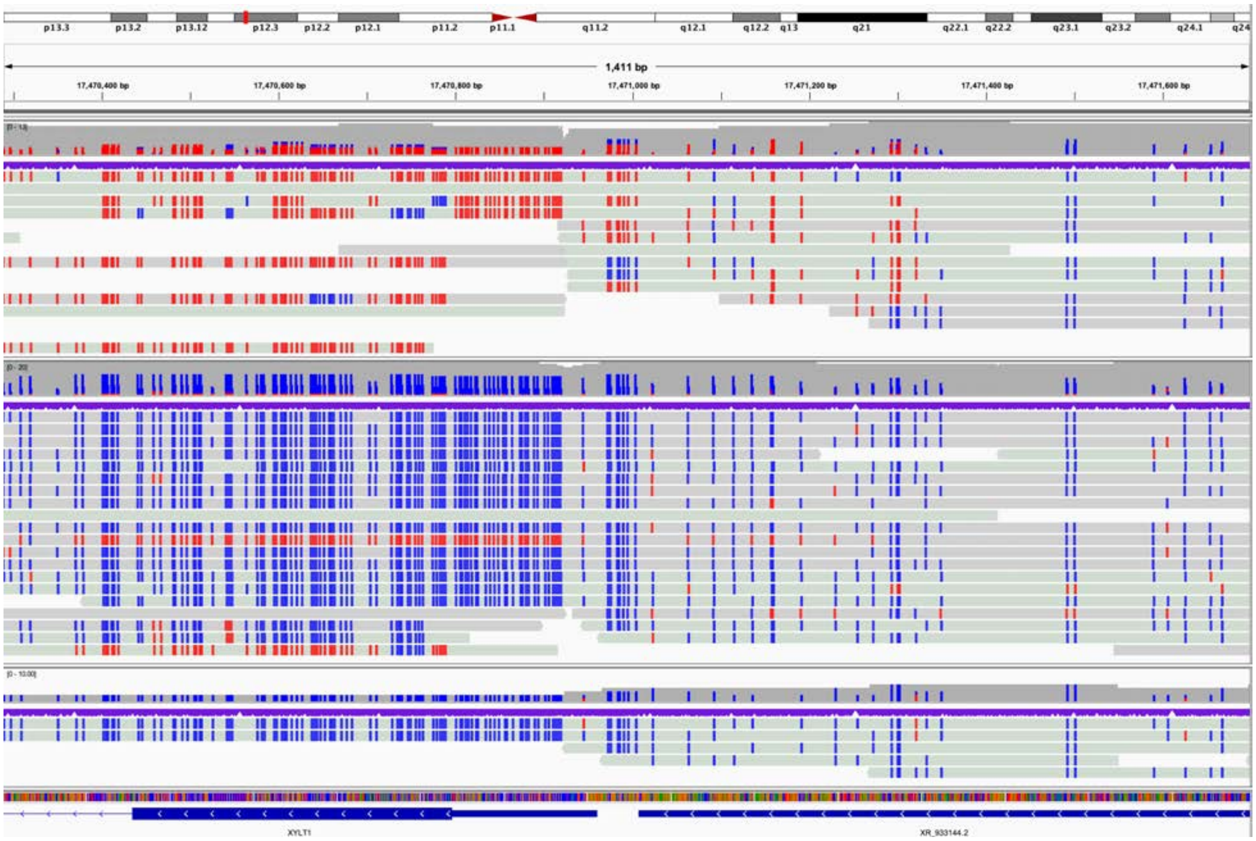
IGV view of a case known to be heterozygous for two different repeat expansions in *FXN*; DNA was obtained from Coriell (Table S1). Reads aligned with minimap2 and default parameters are in the top panel; reads aligned with minimap2 and parameters meant to decrease the number of reads broken by the aligner are in the bottom panel. IGV view is from chr9:69,037,243-69,037,334.

Figure S20. 04-01, 04-02, and 04-03 (*XYLT1*), evaluation of repeat length in family 04 from LaCroix *et al*. 2019.

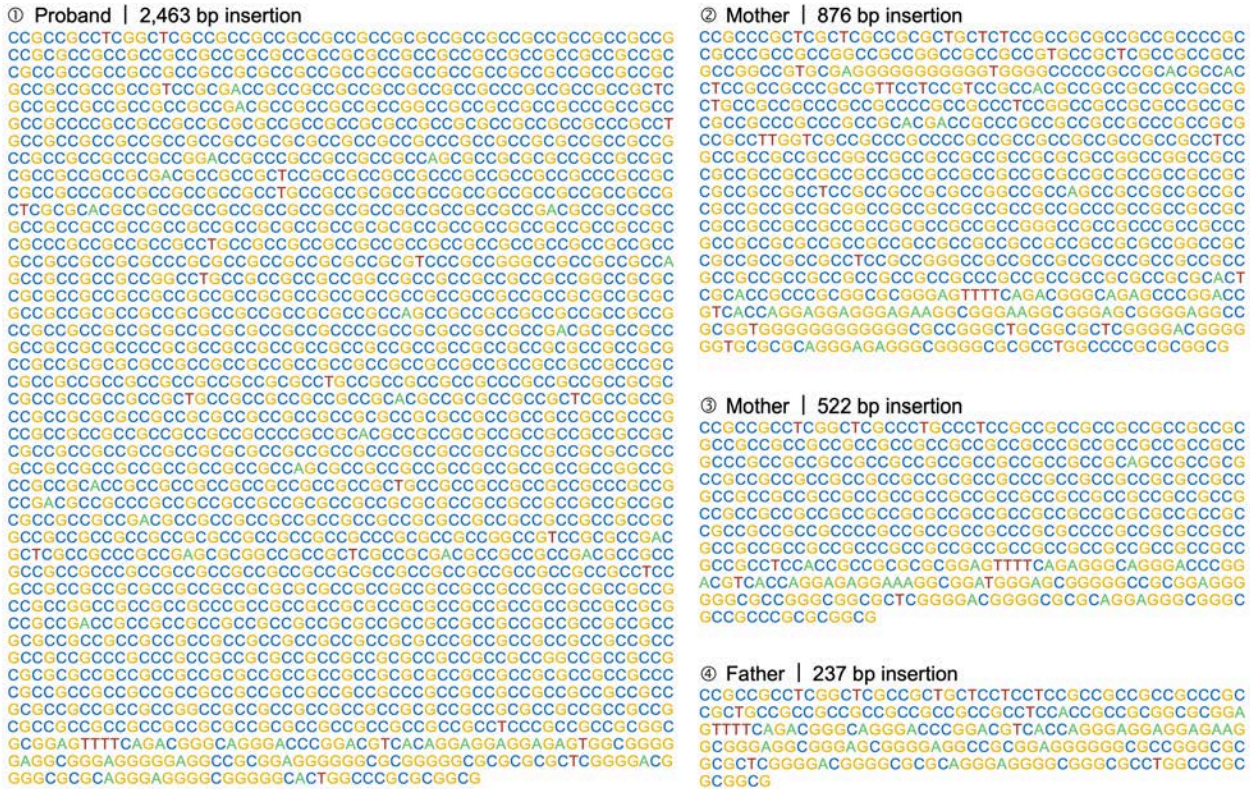
3



A. Screenshot of expansion in 5' UTR of *XYLT1* in family 04 from LaCroix *et al*., 2019.[32] The proband (top) inherited a permutation allele from his mother and a *de novo* deletion from his father; the mother (middle) has a wild-type allele and permutation allele, while the father has two wild-type alleles (bottom).

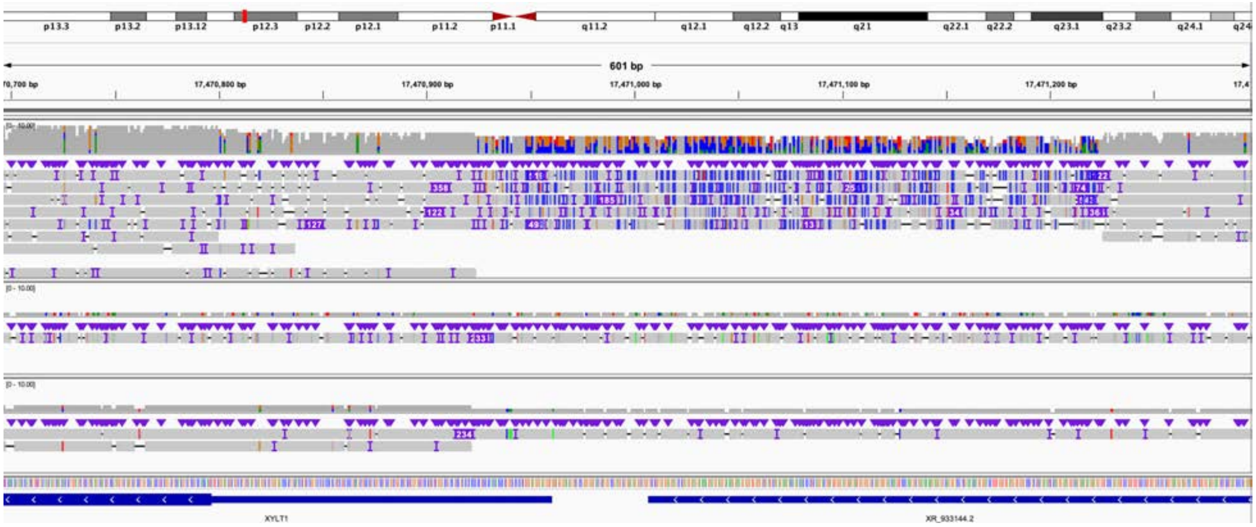


B. Methylation was called and assigned to each read and shows that all reads in the proband are methylated (red), while most in the mother are unmethylated (blue), and both in the father are unmethylated. IGV view from chr16:17,470,696-17,471,299. B. IGV view from chr16:17,470,288-17,471,705.

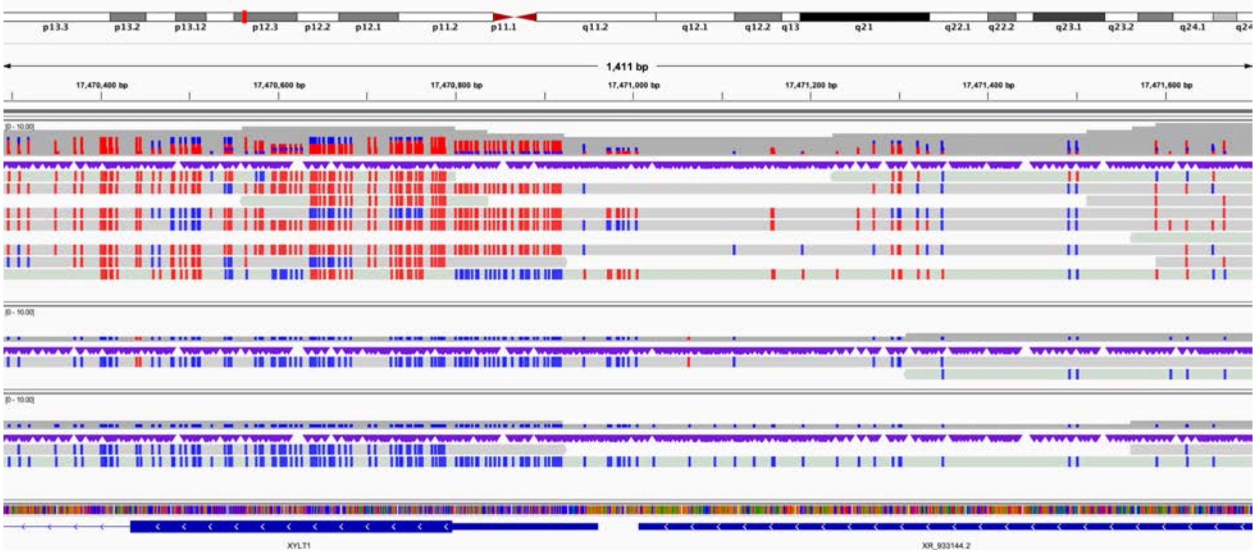

① Proband | 2,463 bp insertion
② Mother | 876 bp insertion
③ Mother | 522 bp insertion
④ Father | 237 bp insertion

C. Sequence of the GCC repeat haplotypes as determined by PacBio CLR sequencing is consistent with the Southern blots and the lengths identified using ONT.

Figure S21. 06-01, 06-02, and 06-03 (*XYLT1*), evaluation of repeat length in family 06 from LaCroix *et al*. 2019.

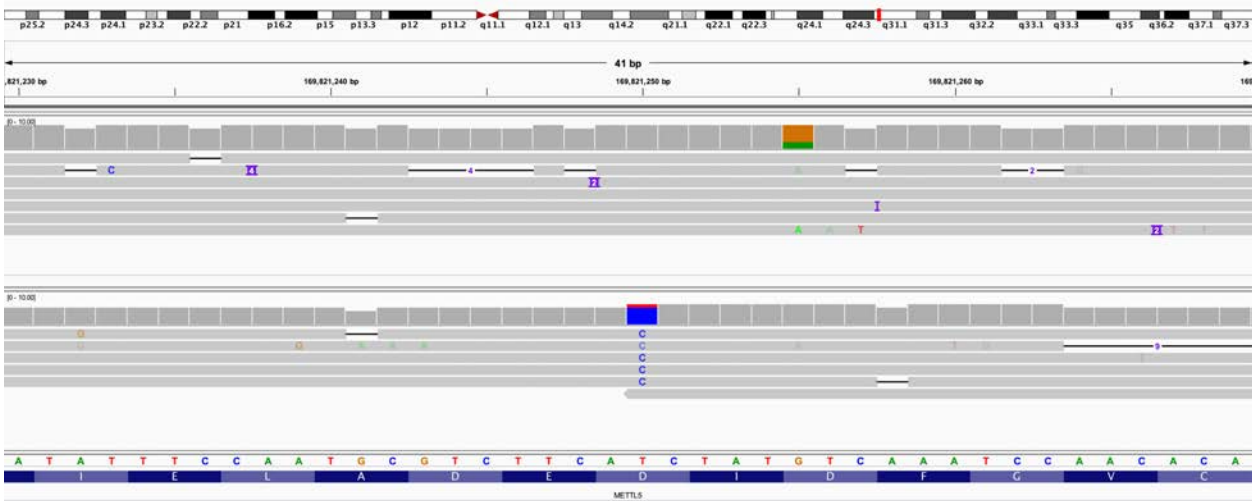


A. IGV view of proband (top), mother (middle), and father (bottom) from family 06 from LaCroix *et al*., 2019.[33] Only a single read was recovered spanning the GGC repeat in both the mother and father. IGV view is from chr16:17,470,696-17,471,299.

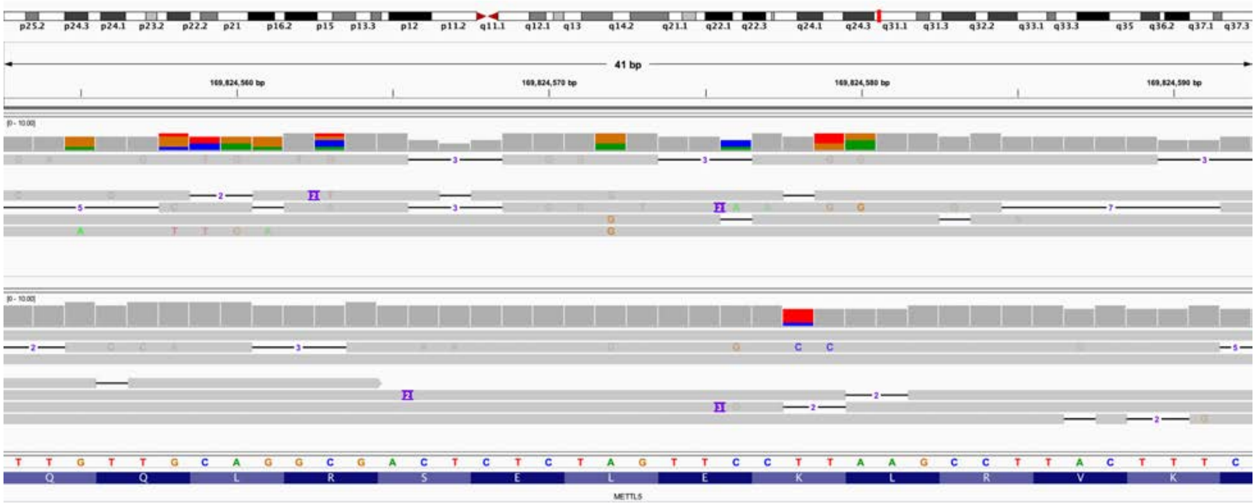


B. IGV view of reads converted to show methylation status; order is the same as in A but the view is slightly larger to show methylation of the first exon of *XYLT1* in the proband. IGV view from chr16:17,470,288-17,471,705.

Figure S22. S071, Phasing of inherited and *de novo* variants in METTL5.

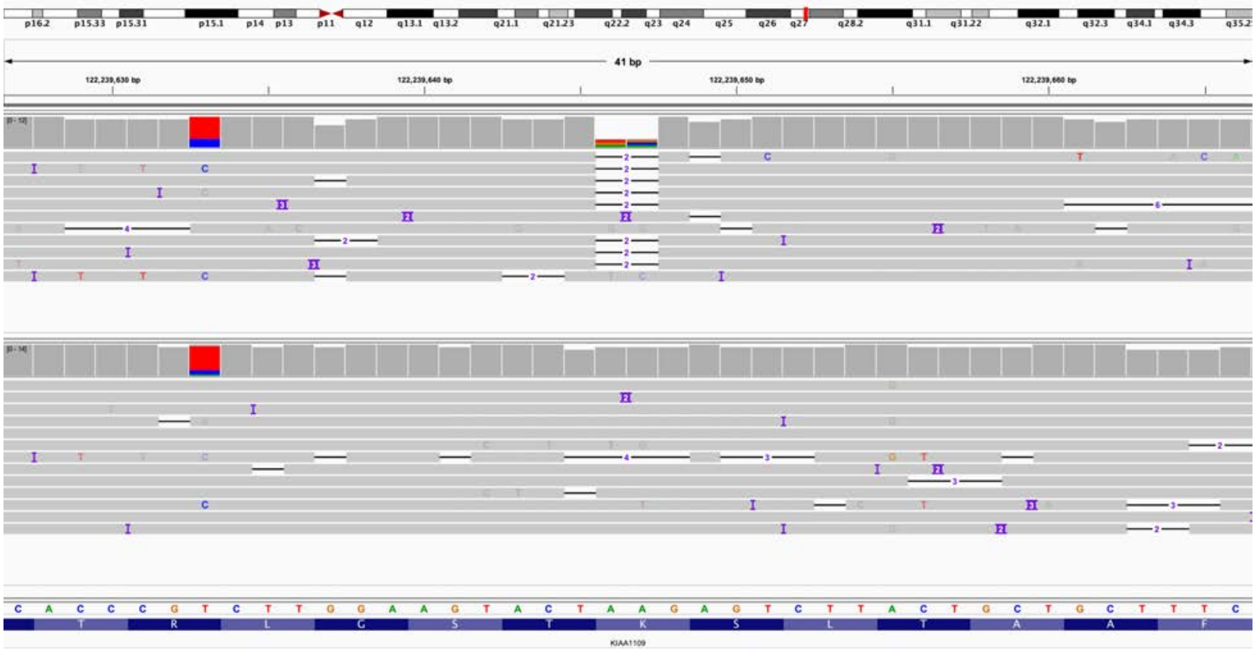


A. View of paternally inherited c.248 A>G (p.D83G), IGV view is chr2:169,821,230-169,821,269. Haplotype 1 is the top track, haplotype 2 is the bottom.

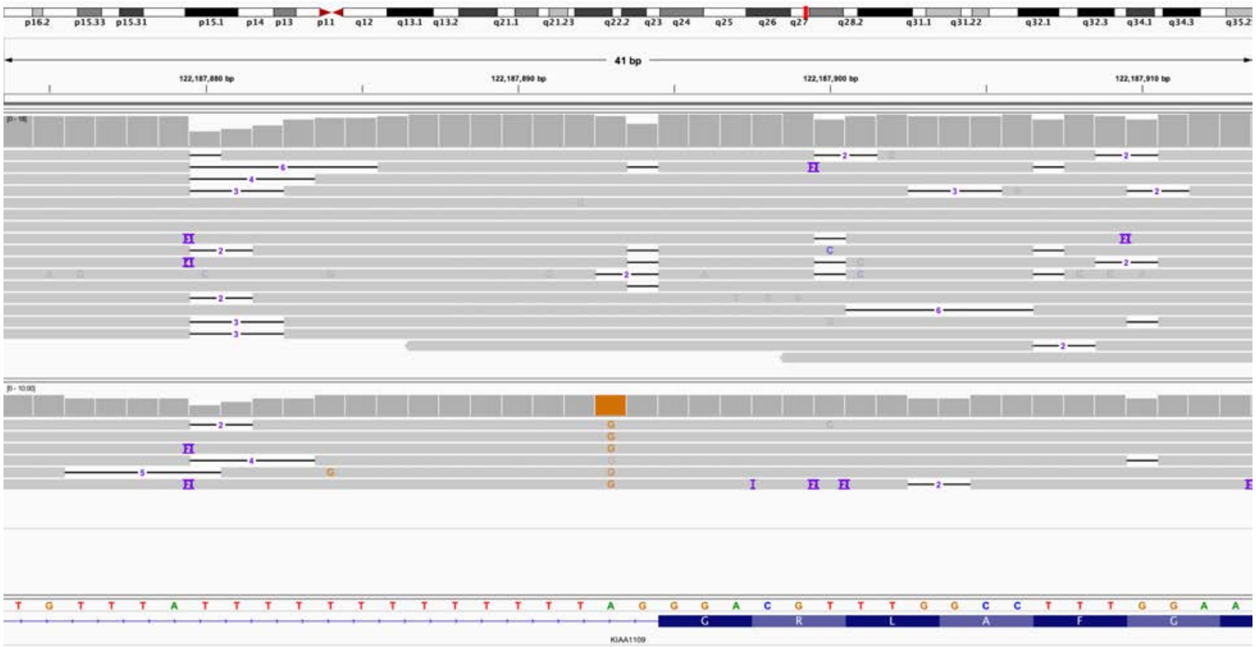


B. View of *de novo* c.26T>C(p.L9P), IGV view is chr2:169,824,553-169,824,592. Haplotype 1 is the top track, haplotype 2 is the bottom.

Figure S23. S086, Phasing of known inherited and *de novo* mosaic variant in KIAA1109.

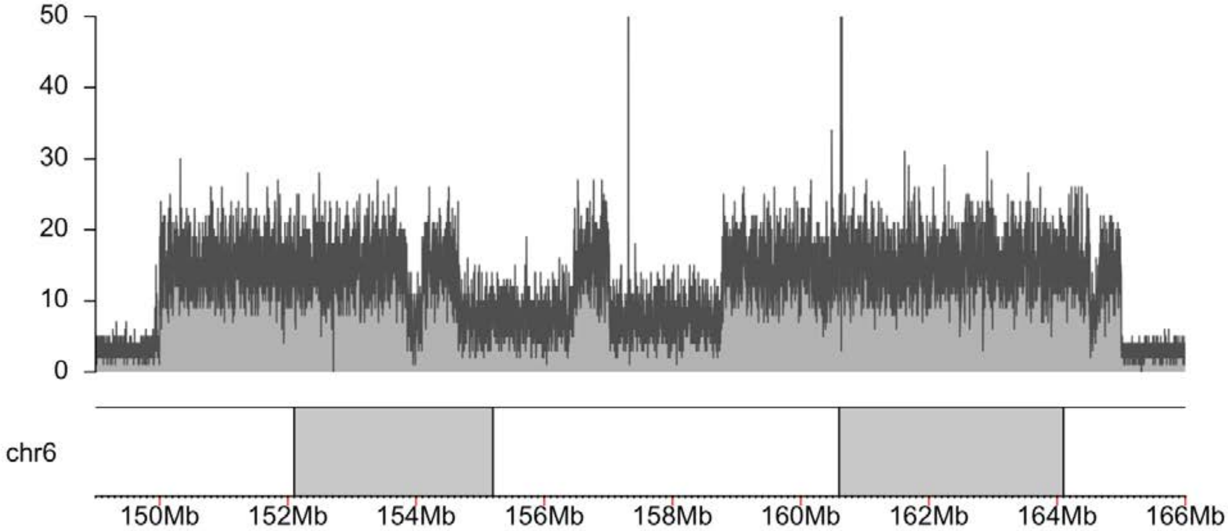

p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35.2
41 bp
122,239,630 bp
122,239,640 bp
122,239,650 bp
122,239,660 bp
C A C C C G T C T T G G A A G T A C T A A G A G T C T T A C T G C T G C T T T C
T R L G S T K S L T A A F
KIAA1109

A. View of maternally inherited 2bp deletion, IGV view is of chr4:122,239,627-122,239,666. Haplotype 1 is top track, haplotype 2 is bottom.

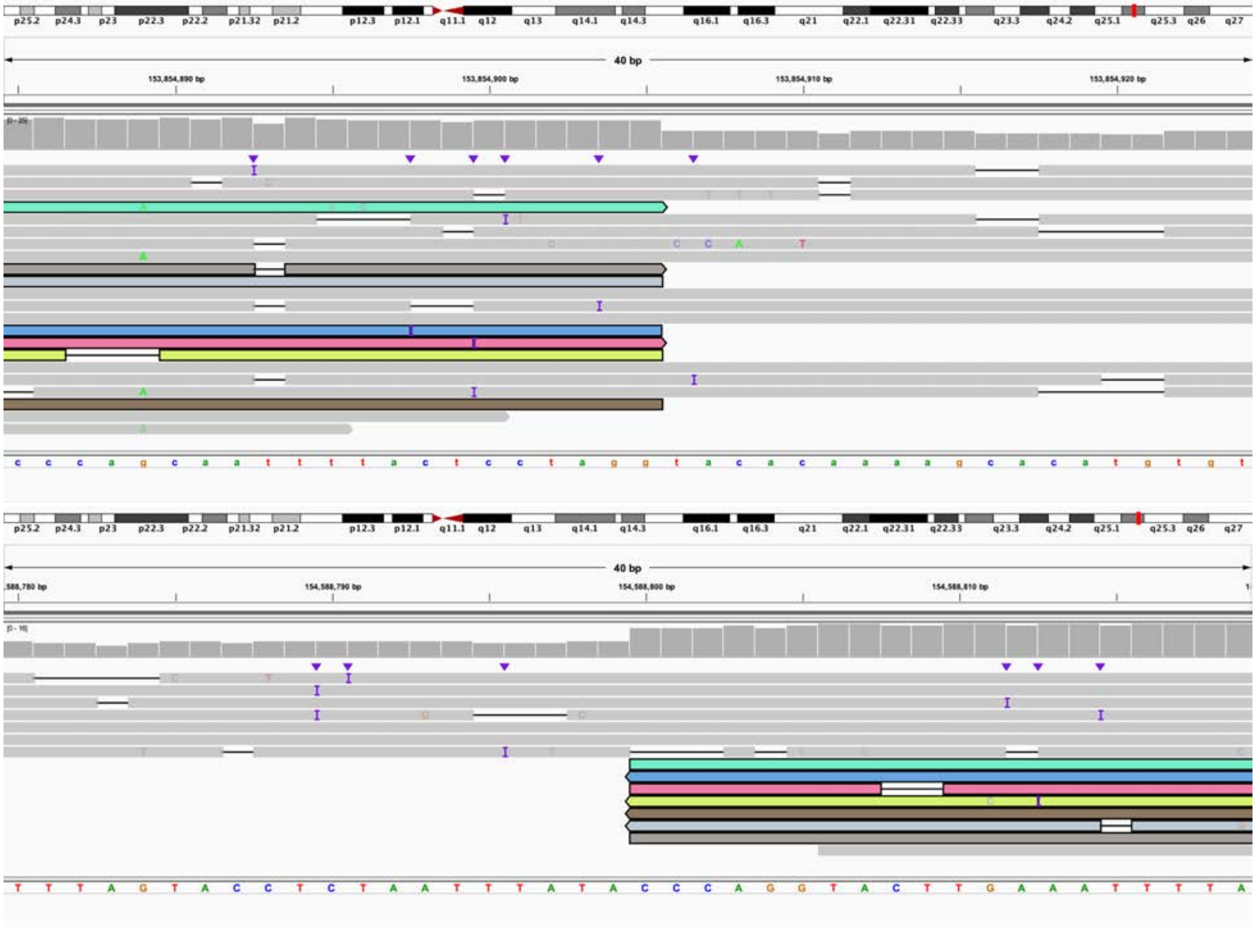

p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35.2
41 bp
122,187,880 bp
122,187,890 bp
122,187,900 bp
122,187,910 bp
T G T T T A T T T T T T T T T T T T T T A G G G A C G T T T G G C C T T T G G A A
G R L A F G
KIAA1109

B. View of *de novo* A>G with reported allele fraction of 16%. IGV view is of chr4:122,187,874-122,187,913. Haplotype 1 is top track, haplotype 2 is bottom.

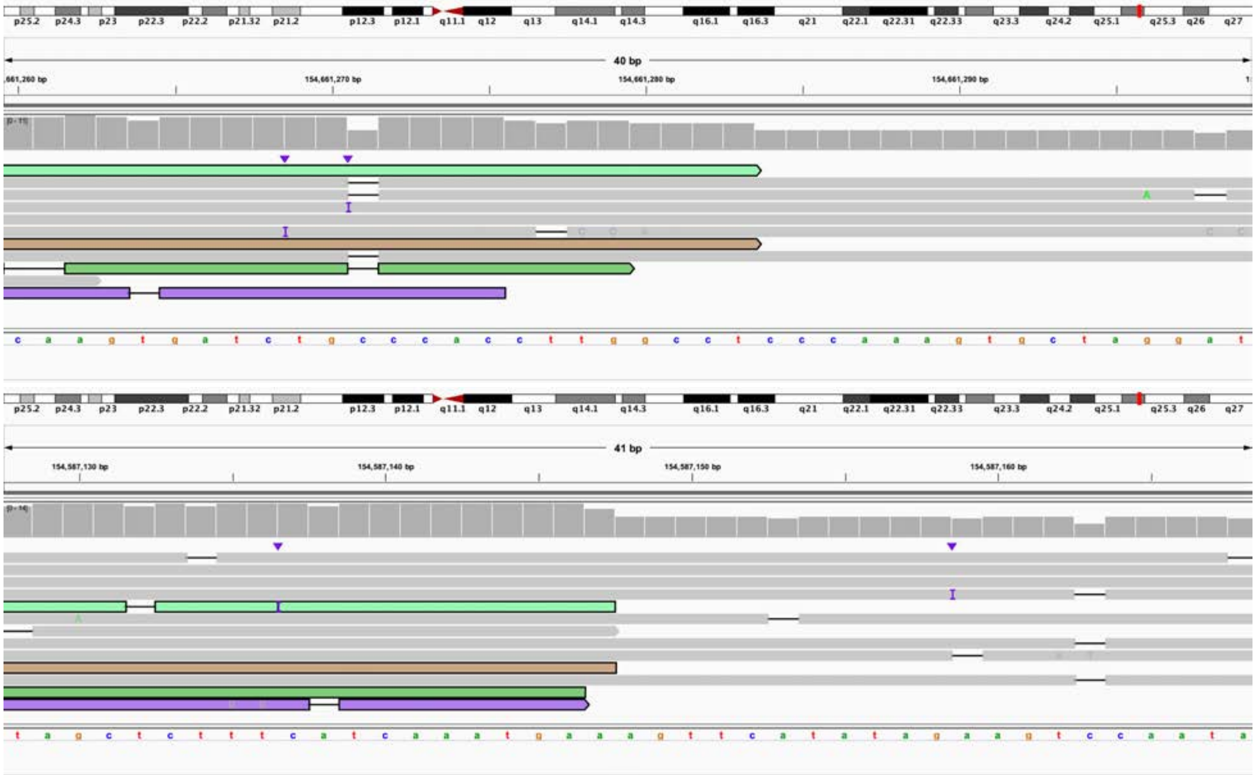
Figure S24. S014, three noncontiguous deletions of chromosome 6 identified by CMA.



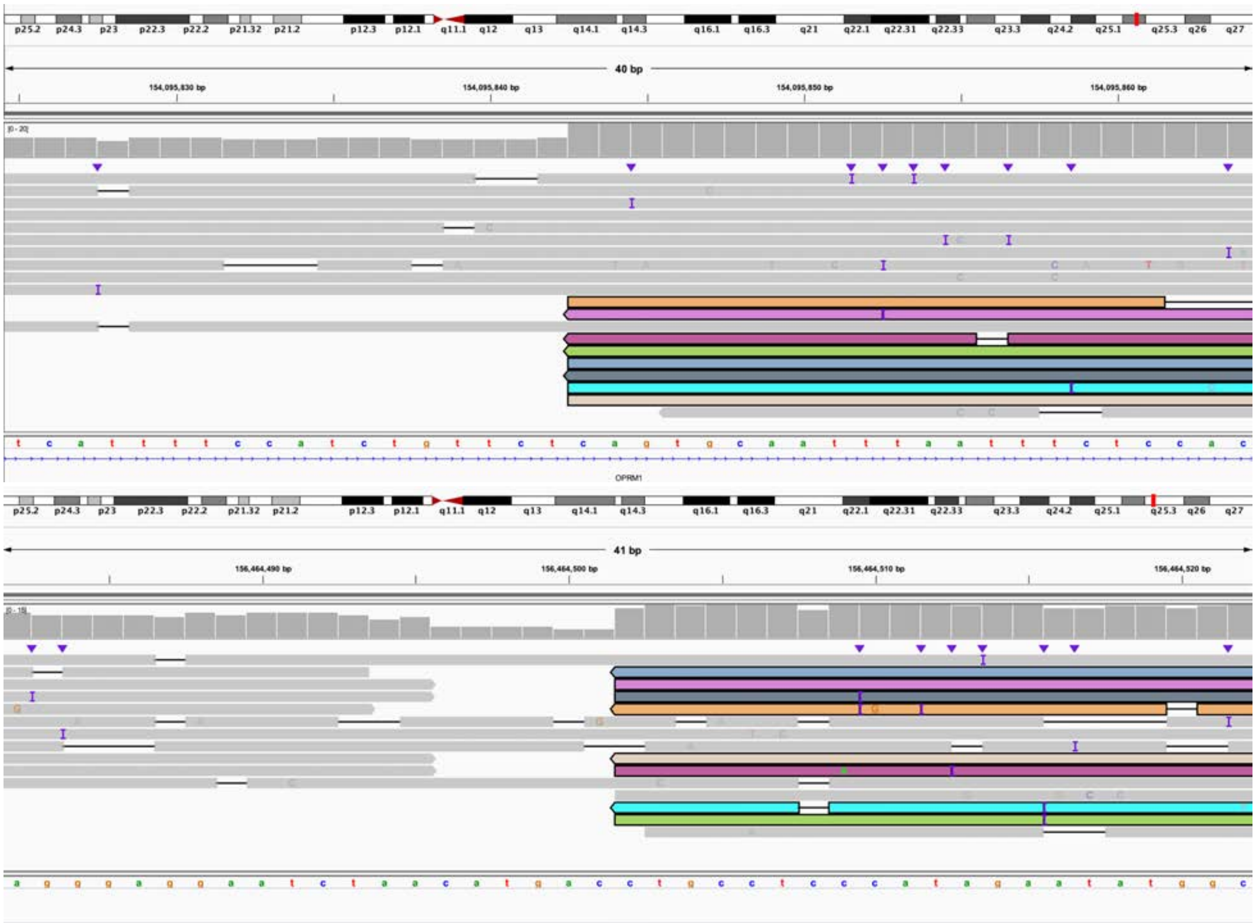
A. Coverage of chromosome 6 target region.


p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
153,854,890 bp
153,854,900 bp
153,854,910 bp
153,854,920 bp
c c c a g c a a t t t t a c t c c t a g g t a c a c a a a a g c a c a t g t g t
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
588,780 bp
154,588,790 bp
154,588,800 bp
154,588,810 bp
T T T A G T A C C T C T A A T T T A T A C C C A G G T A C T T G A A A T T T T A

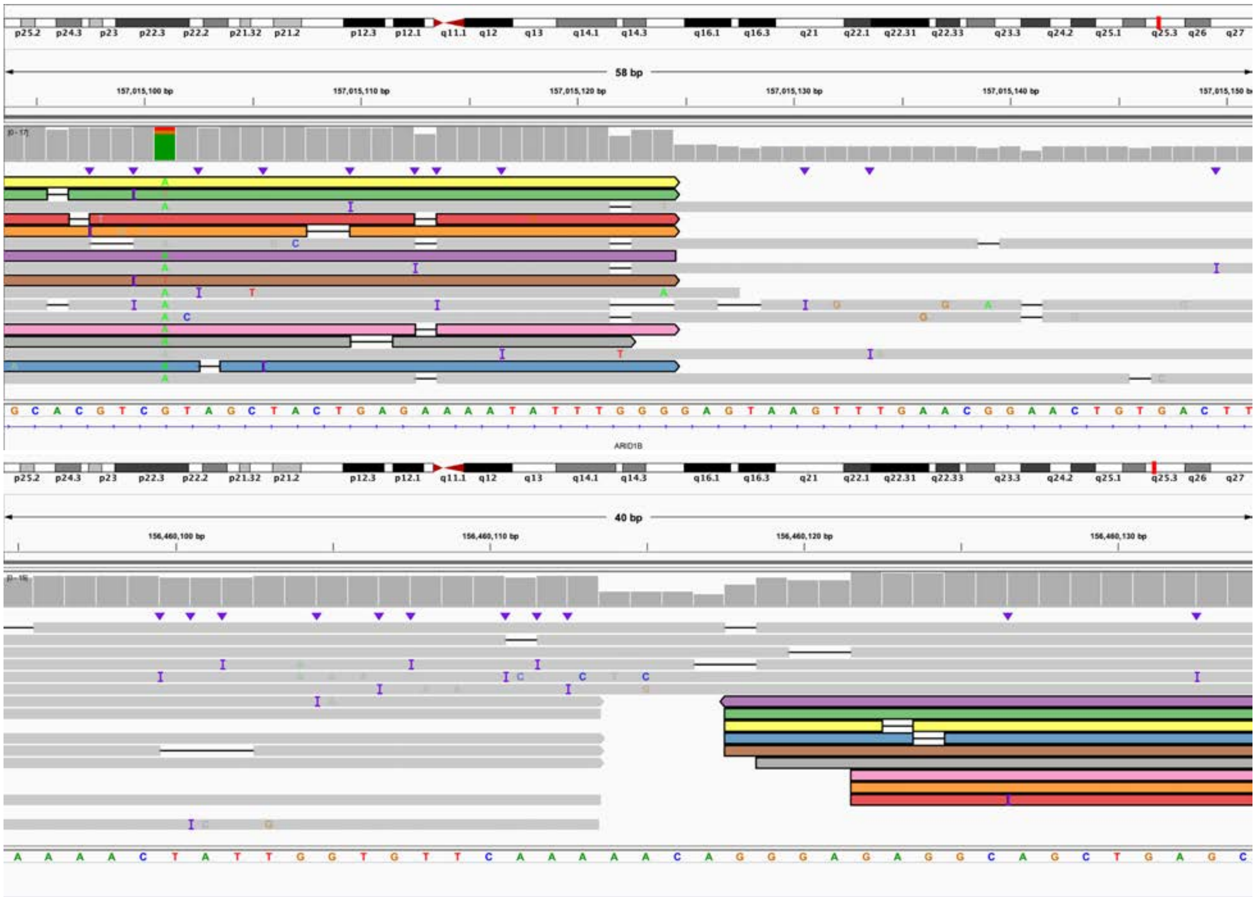
B: 3' end of fragment A (IGV: chr6:153,854,885-153,854,924) connects to 5' end of fragment E (IGV: chr6:154,588,780-154,588,819).


p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
,661,260 bp
154,661,270 bp
154,661,280 bp
154,661,290 bp
[0 - 11]
c a a g t g a t c t g c c c a c c t t g g c c t c c c a a a g t g c t a g g a t
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
41 bp
154,587,130 bp
154,587,140 bp
154,587,150 bp
154,587,160 bp
[0 - 14]
t a g c t c t t t c a t c a a a t g a a a g t t c a t a t a g a a g t c c a a t a

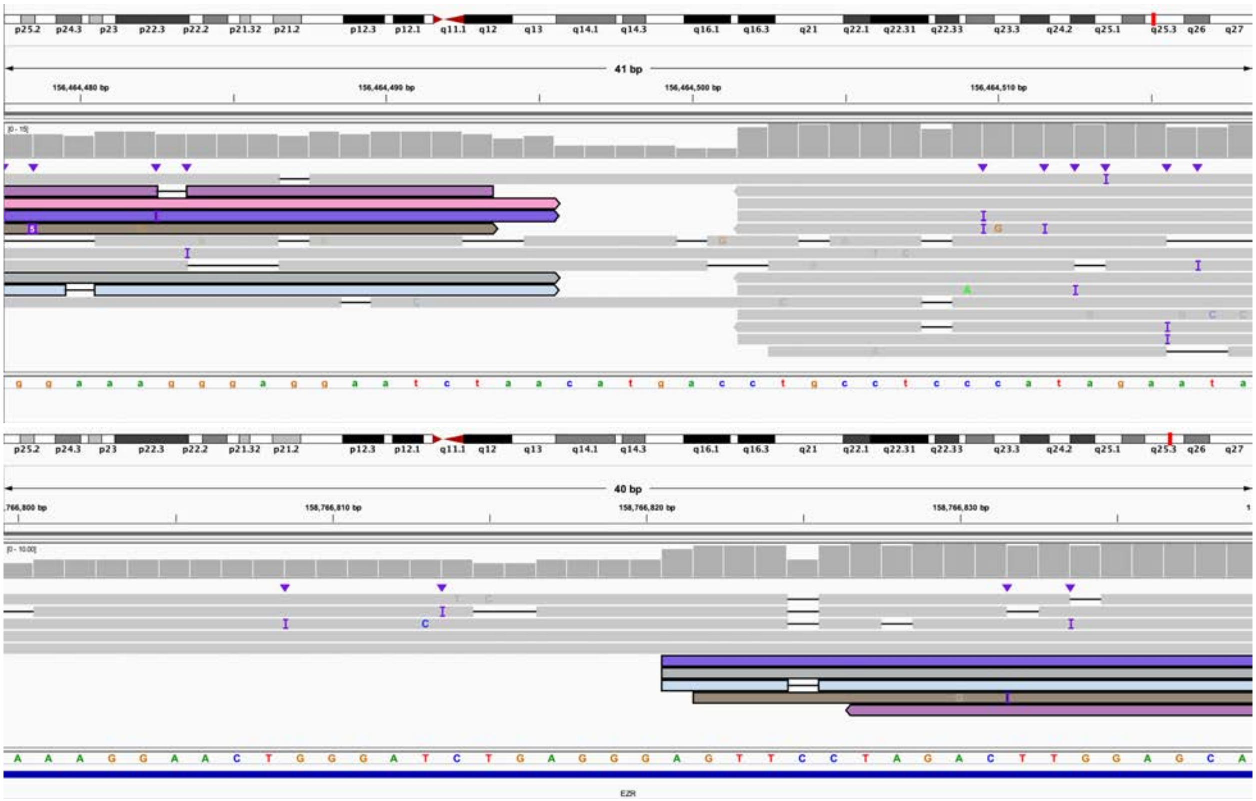
C: 3' end of fragment E (IGV: chr6:154,661,260-154,661,299) connects to 3’ end of fragment C (IGV: chr6:154,587,128-154,587,168).


p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
154,095,830 bp
154,095,840 bp
154,095,850 bp
154,095,860 bp
[0 - 20]
t c a t t t t c c a t c t g t t c t c a g t g c a a t t t a a t t t c t c c a c
OPRM1
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
41 bp
156,464,490 bp
156,464,500 bp
156,464,510 bp
156,464,520 bp
[0 - 18]
a g g g a g g a a t c t a a c a t g a c c t g c c t c c c a t a g a a t a t g g c

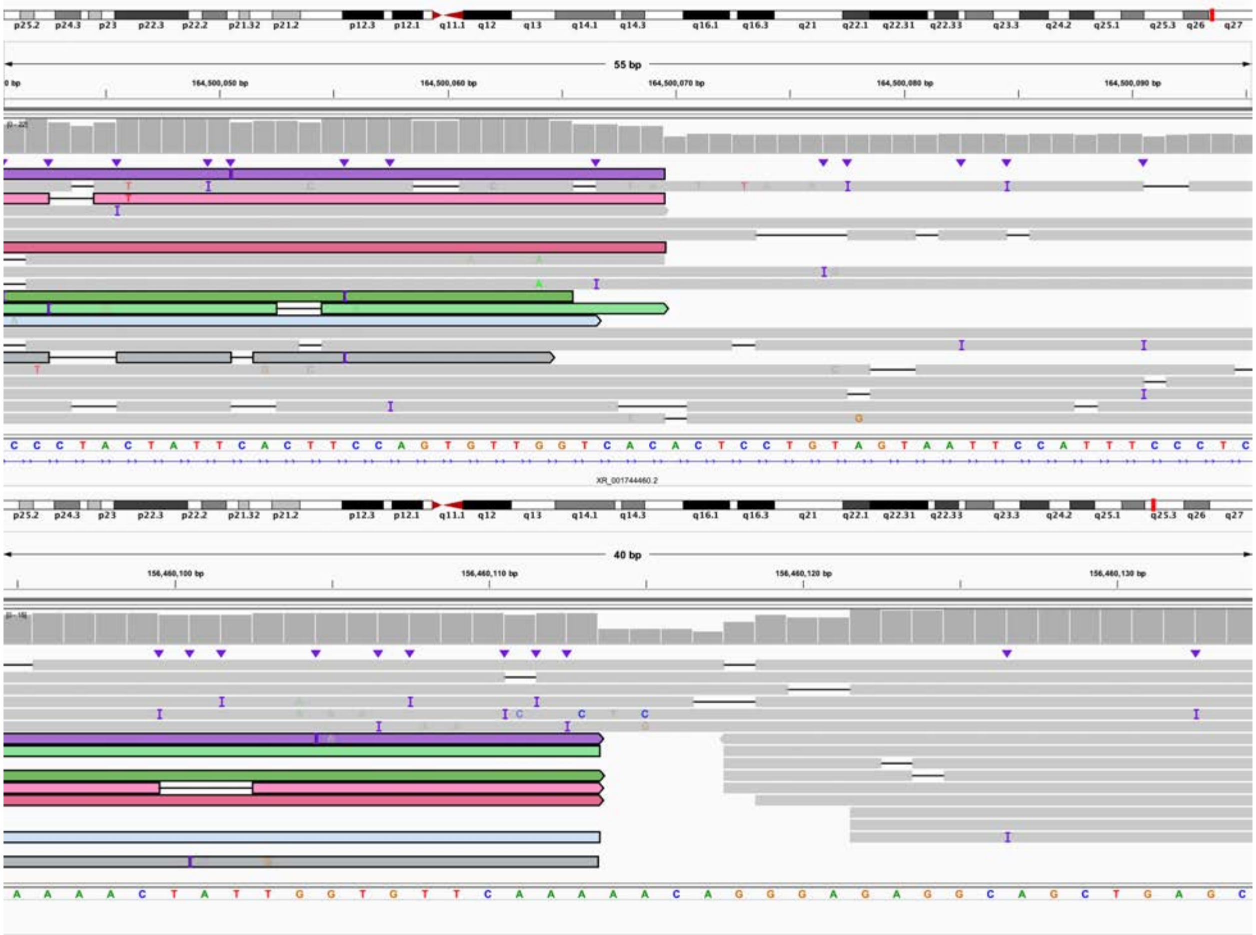
D: 5' end of fragment C bisects *OPRM1* (IGV: chr6:154,095,825-154,095,864) and connects to 5' end of fragment I (IGV: chr6:156,464,482-156,464,522).


p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
58 bp
157,015,100 bp
157,015,110 bp
157,015,120 bp
157,015,130 bp
157,015,140 bp
157,015,150 bp
[0 - 17]
G C A C G T C G T A G C T A C T G A G A A A A T A T T T G G G G A G T A A G T T T G A A C G G A A C T G T G A C T T
ARID1B
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
156,460,100 bp
156,460,110 bp
156,460,120 bp
156,460,130 bp
[0 - 16]
A A A A C T A T T G G T G T T C A A A A A C A G G G A G A G G C A G C T G A G C

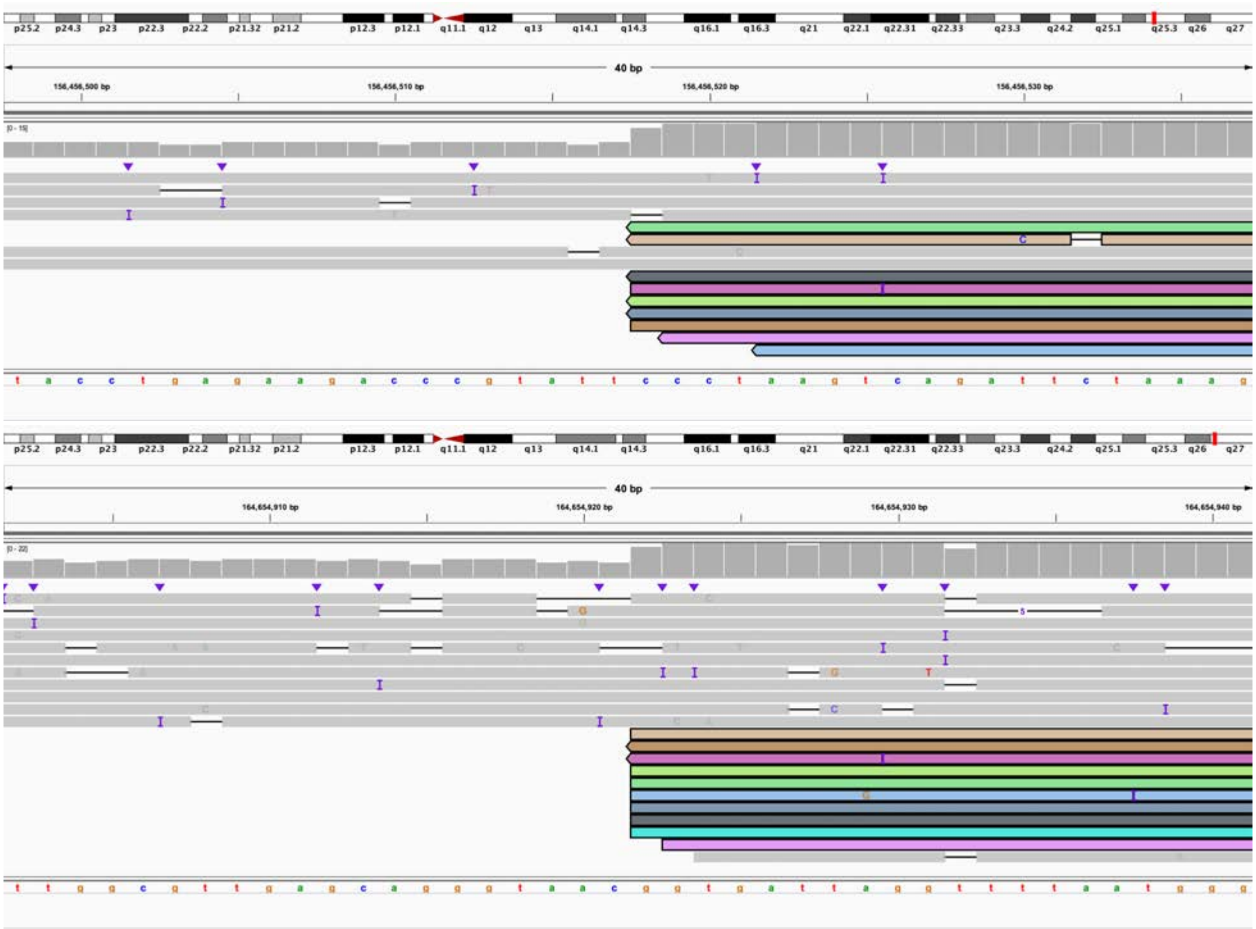
E: 3' end of fragment I bisects *ARID1B* (IGV: chr6:157,015,094-157,015,151) and connects to 5' end of fragment H (IGV: chr6:156,460,095-156,460,134).


p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
41 bp
156,464,480 bp
156,464,490 bp
156,464,500 bp
156,464,510 bp
[0 - 15]
g g a a a g g g a g g a a t c t a a c a t g a c c t g c c t c c c a t a g a a t a
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
,766,800 bp
158,766,810 bp
158,766,820 bp
158,766,830 bp
[0 - 10.00]
A A A G G A A C T G G G A T C T G A G G G A G T T C C T A G A C T T G G A G C A
EZR

F: 3' end of fragment H (IGV: chr6:156,464,478-156,464,518) connects to 5' end of fragment K and bisects *EZR* (IGV: chr6:158,766,800-158,766,839).


p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
55 bp
164,500,050 bp
164,500,060 bp
164,500,070 bp
164,500,080 bp
164,500,090 bp
C C C T A C T A T T C A C T T C C A G T G T T G G T C A C A C T C C T G T A G T A A T T C C A T T T C C C T C
XR_001744460.2
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
156,460,100 bp
156,460,110 bp
156,460,120 bp
156,460,130 bp
A A A A C T A T T G G T G T T C A A A A A C A G G G A G A G G C A G C T G A G C

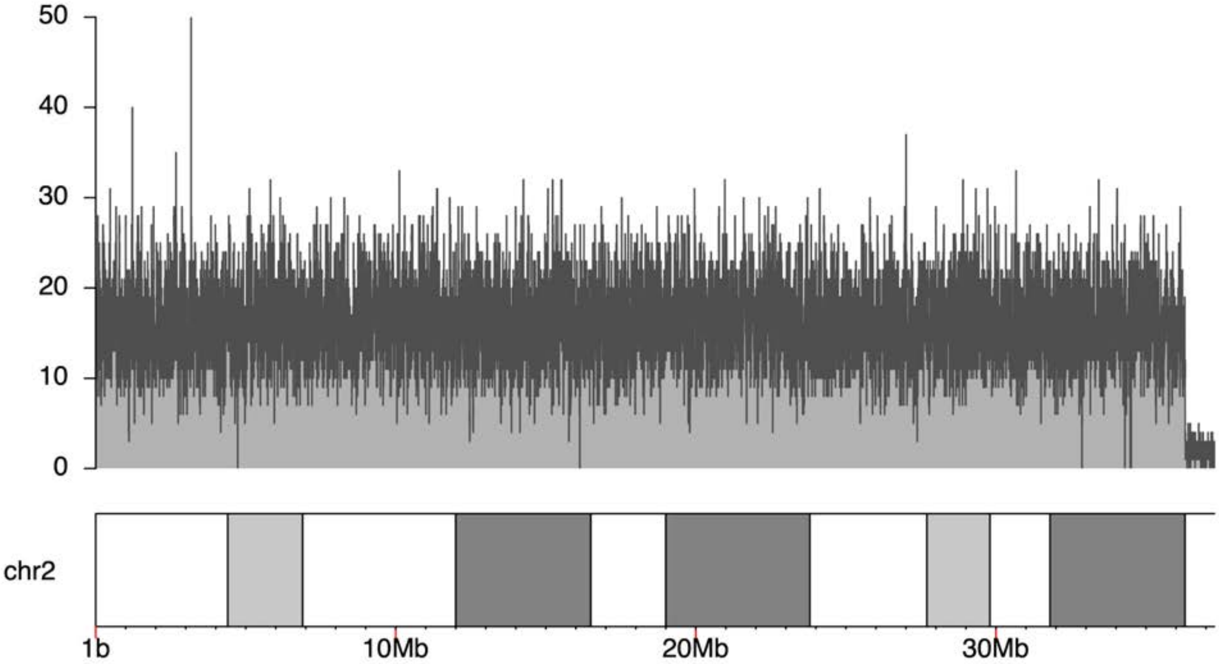
G: 3' end of fragment K bisects *XR_001744460.2* (IGV: chr6:164,500,041-164,500,095) and connects to 3' end of fragment G (IGV: chr6:156,460,095-156,460,134).


p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
156,456,500 bp
156,456,510 bp
156,456,520 bp
156,456,530 bp
[0 - 15]
t a c c t g a g a a g a c c c g t a t t c c c t a a g t c a g a t t c t a a a g
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
164,654,910 bp
164,654,920 bp
164,654,930 bp
164,654,940 bp
[0 - 22]
t t g g c g t t g a g c a g g g t a a c g g t g a t t a g g t t t t a a t g g g

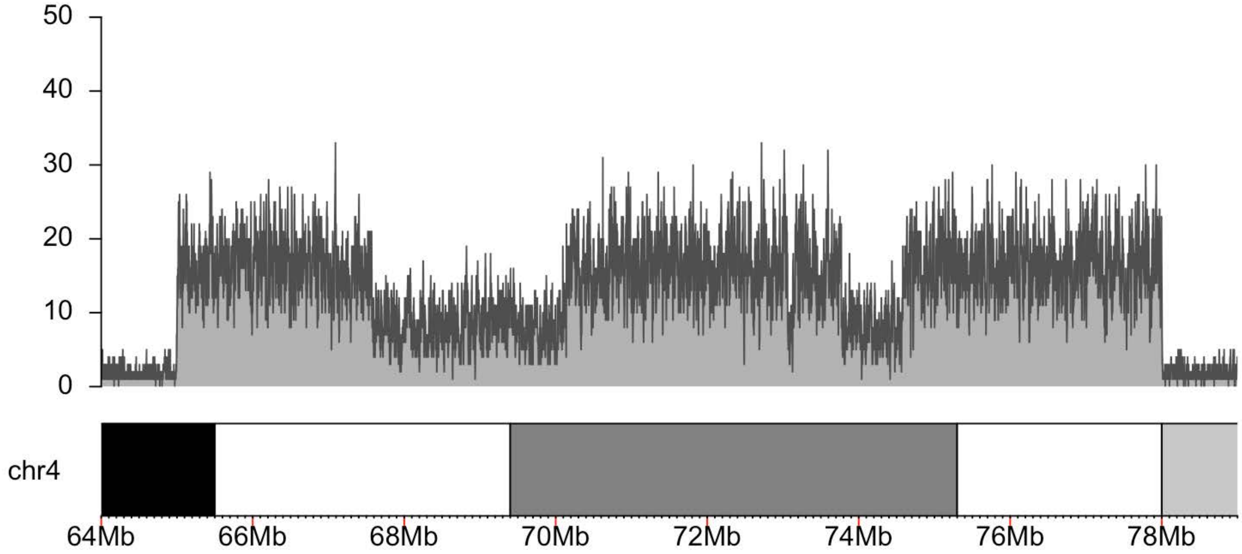
H: 5' end of fragment G (IGV: chr6:156,456,498-156,456,537) connects to 5' end of fragment M (IGV: chr6:164,654,902-164,654,941).

Figure S25. S020, individual with three deletions identified on array and multiple rearrangements on karyotype.

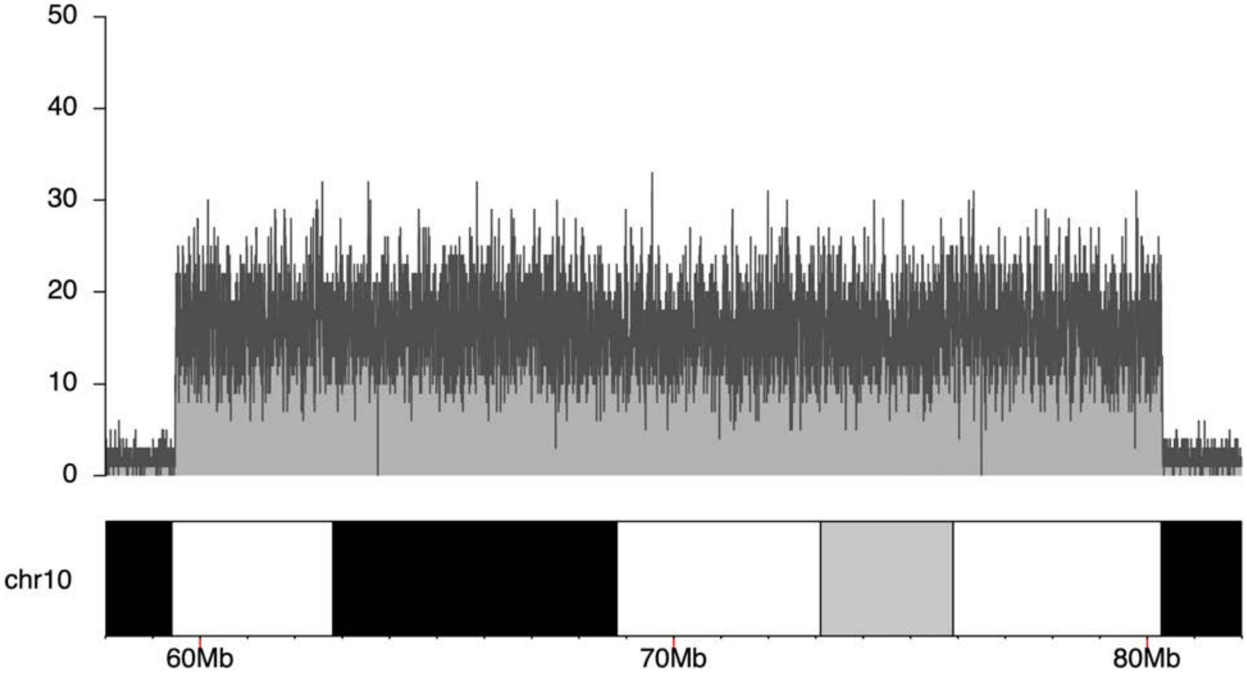
In panels J-AG Nanopore data is the top track and PacBio HiFi data is the bottom track.



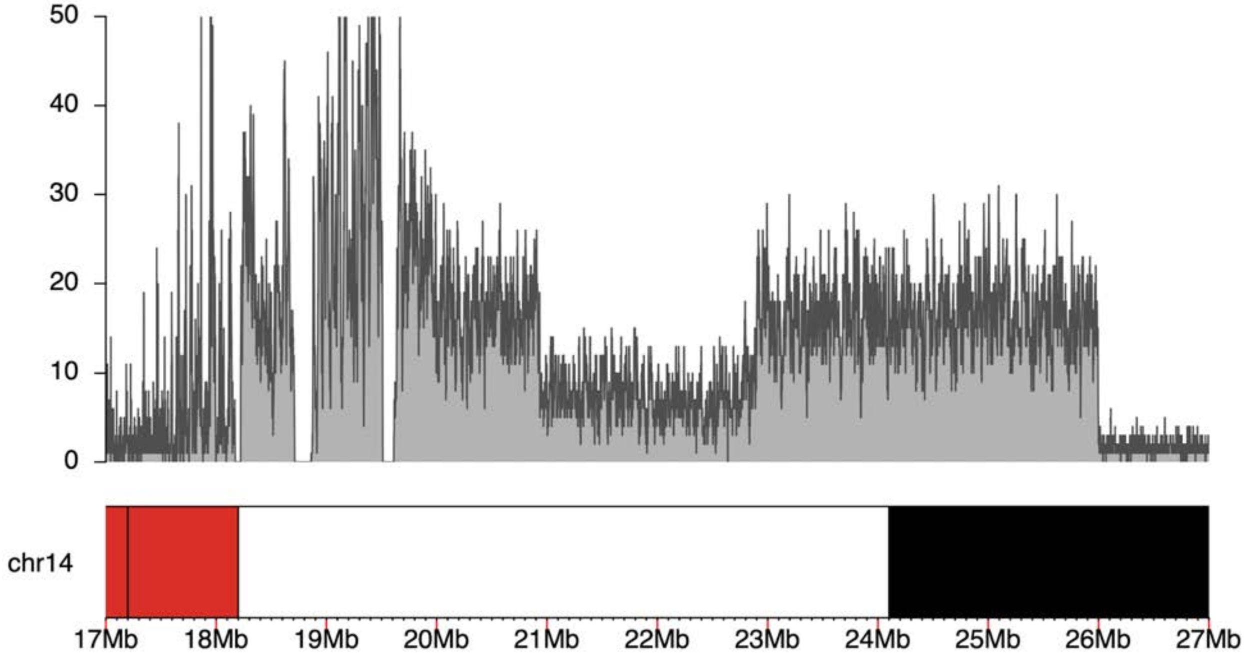
A: Coverage of chromosome 2 target region from readfish experiment 1.



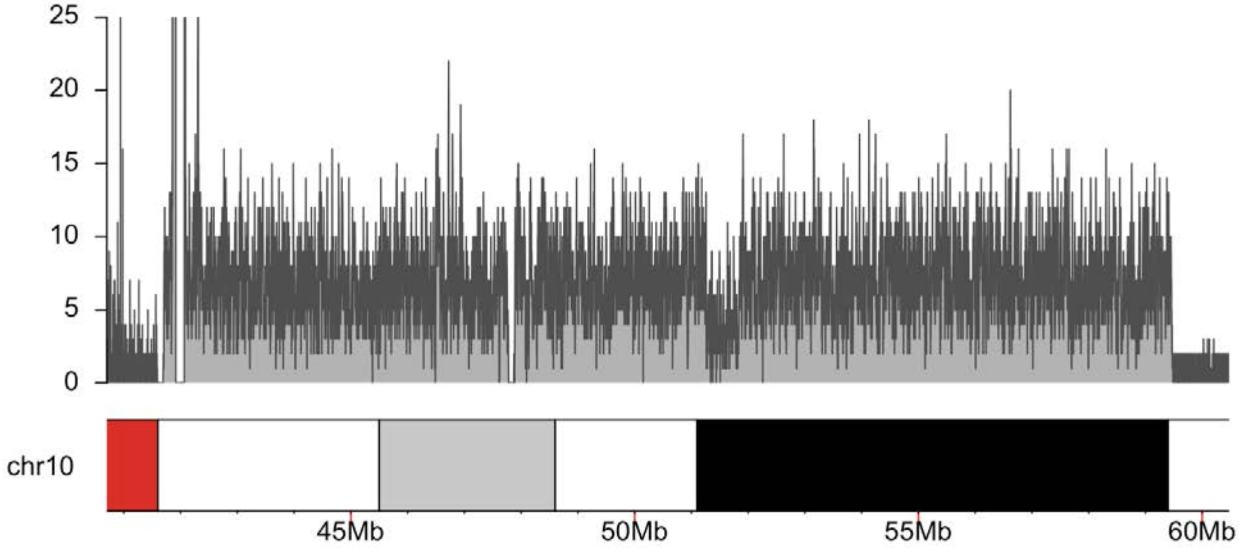
B: Coverage of chromosome 4 target region from readfish experiment 1.



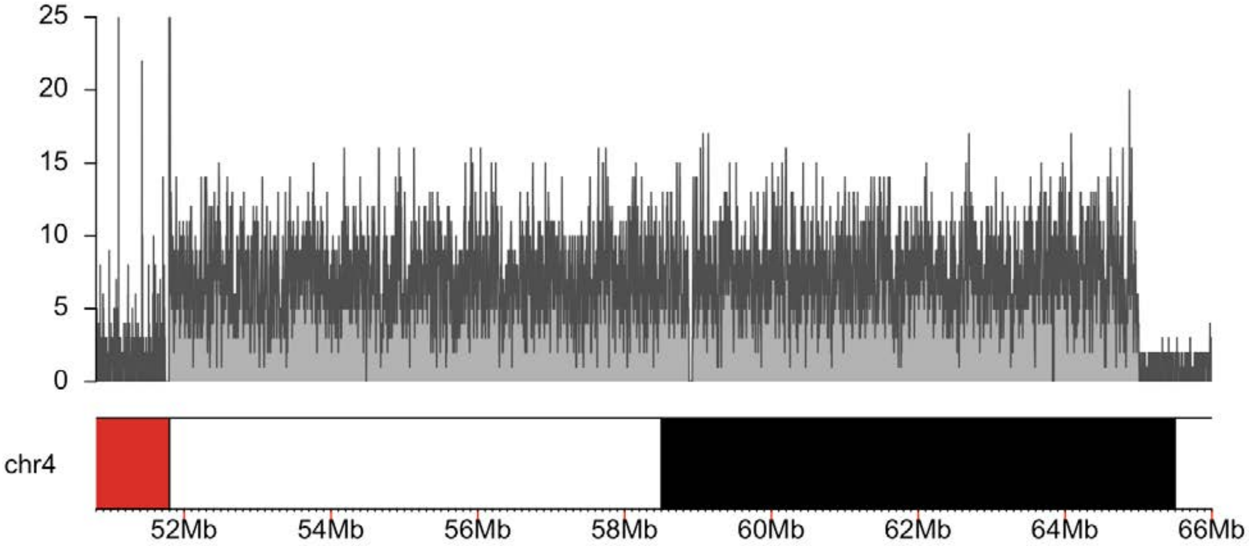
C: Coverage of chromosome 10 target region from readfish experiment 1.



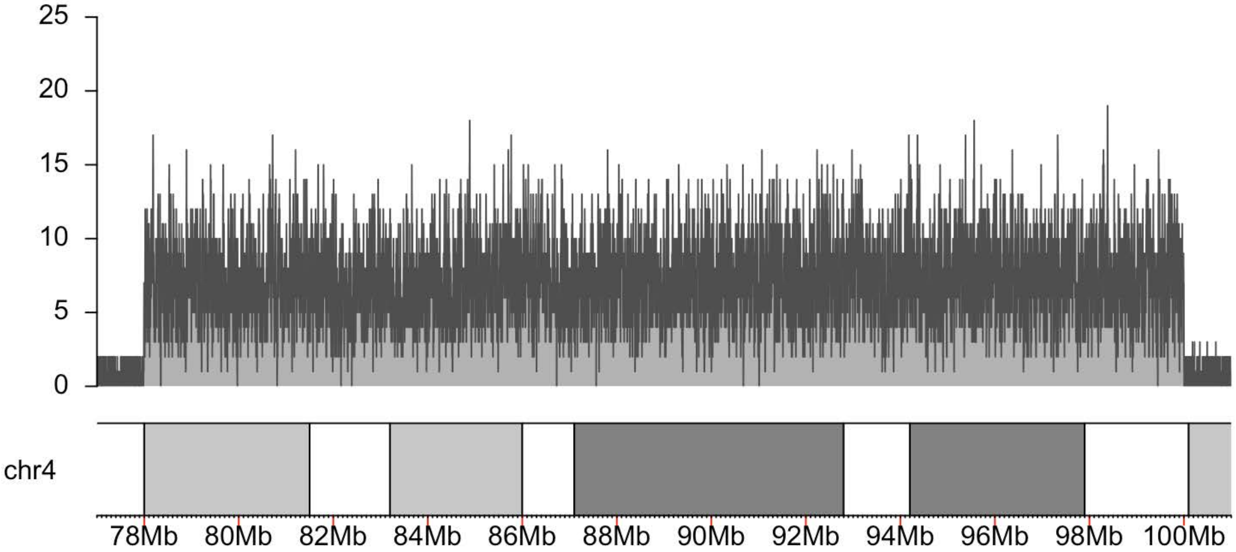
D: Coverage of chromosome 14 target region from readfish experiment 1.



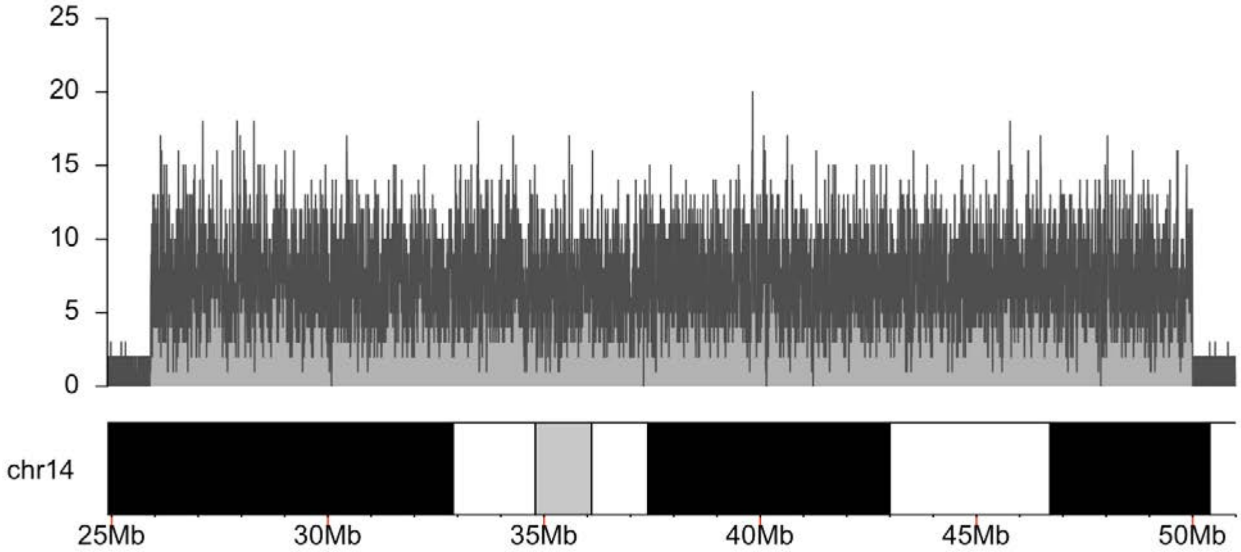
E: Coverage of chromosome 10 target region from readfish experiment 2.



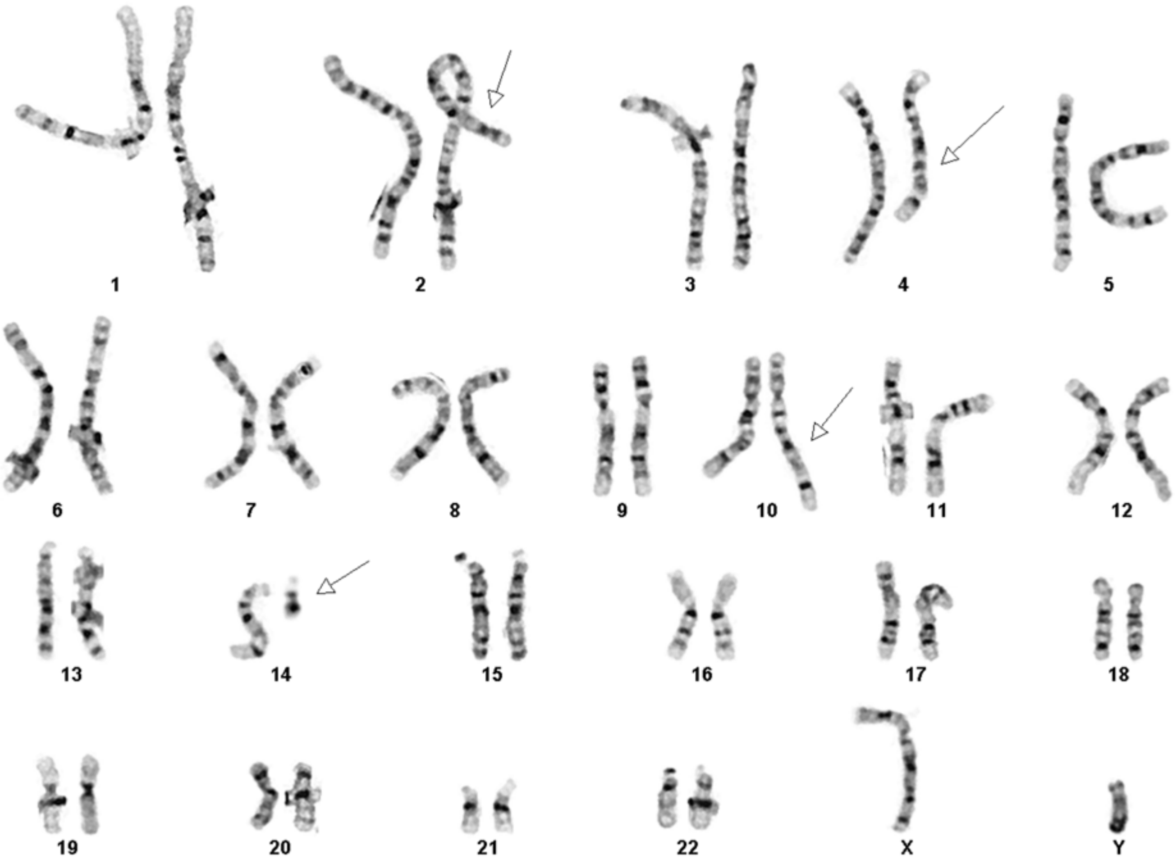
F: Coverage of first chromosome 4 target region from readfish experiment 2.



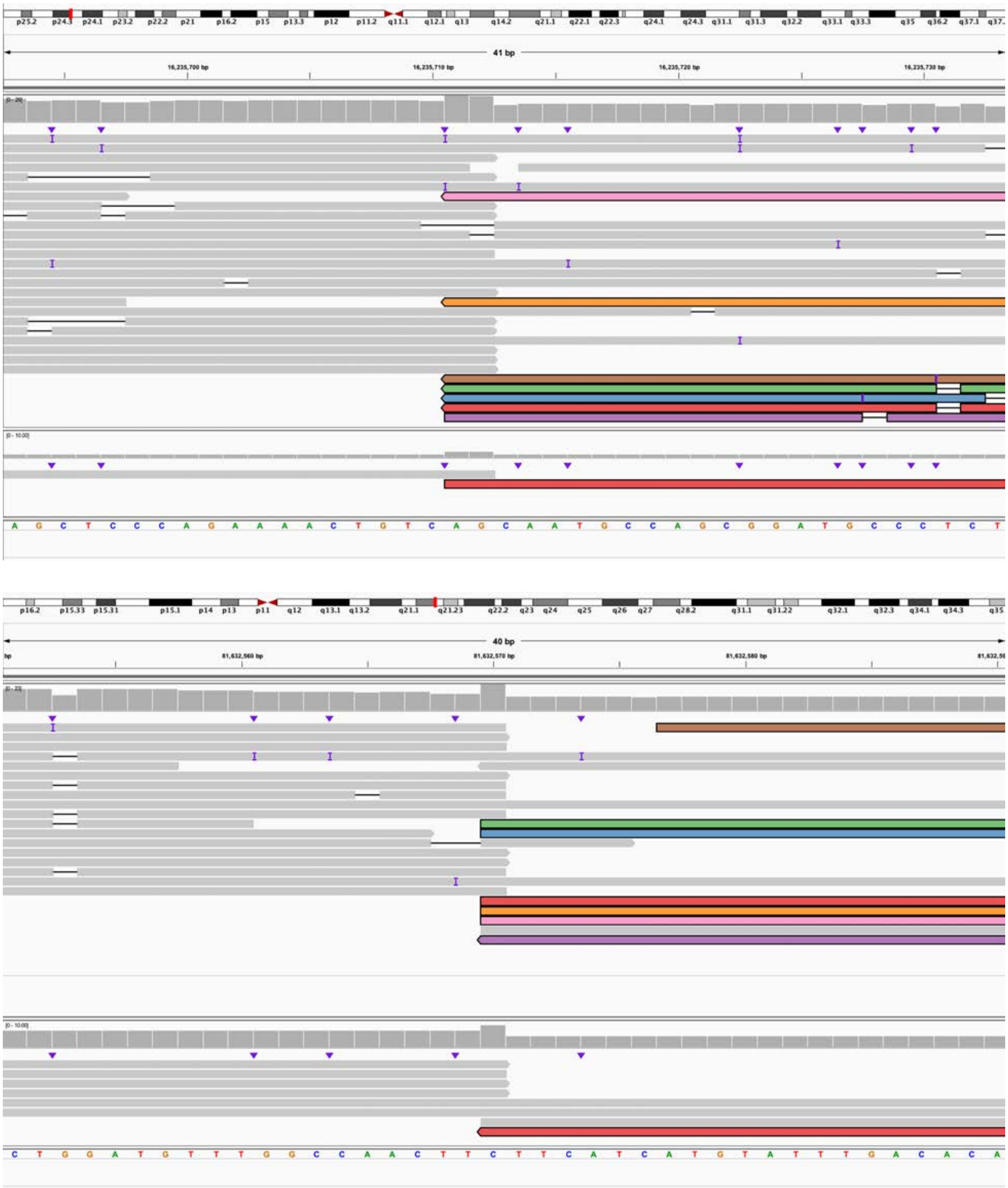
G: Coverage of second chromosome 4 target region from readfish experiment 2.



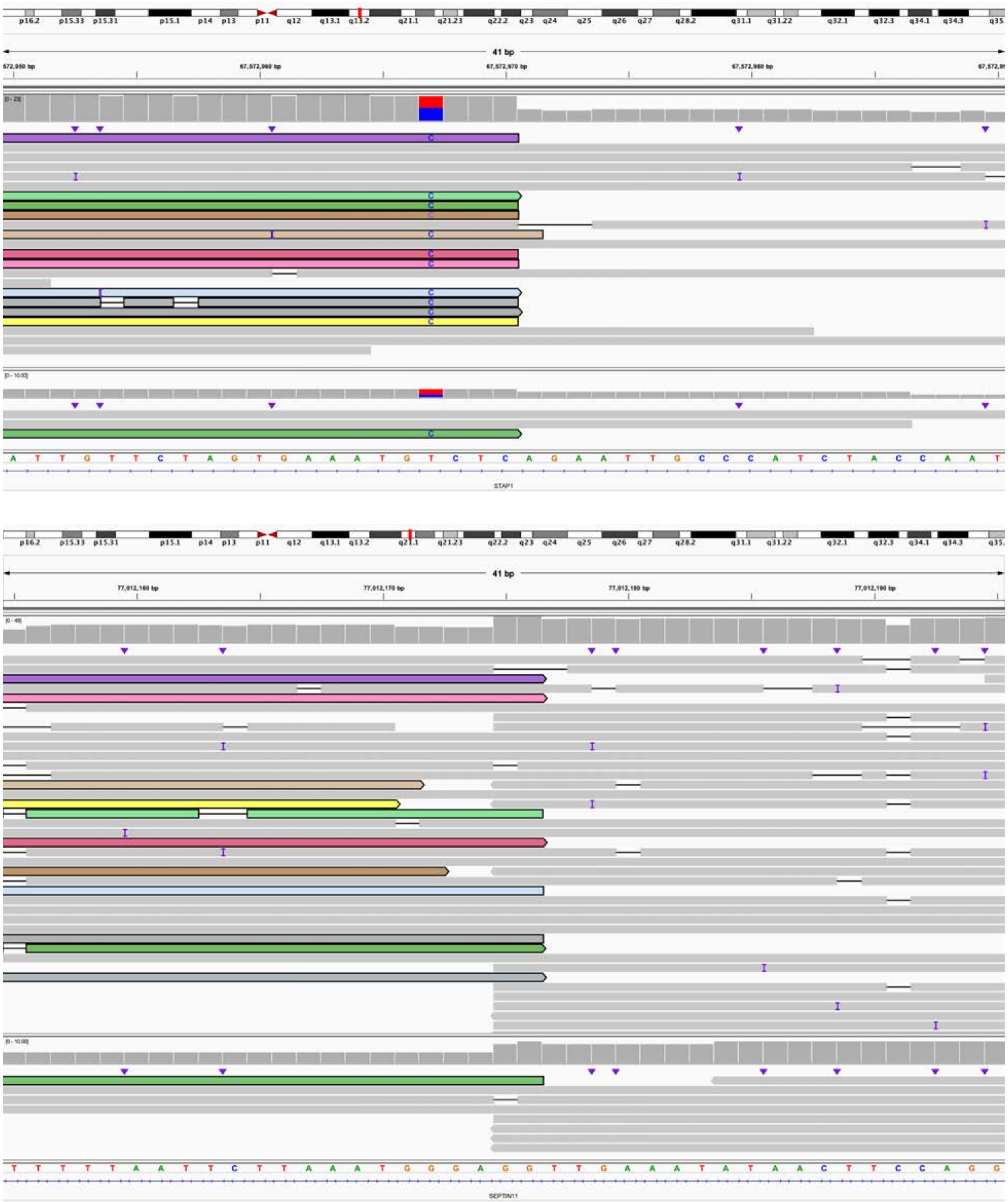
H: Coverage of chromosome 14 target region from readfish experiment 2.



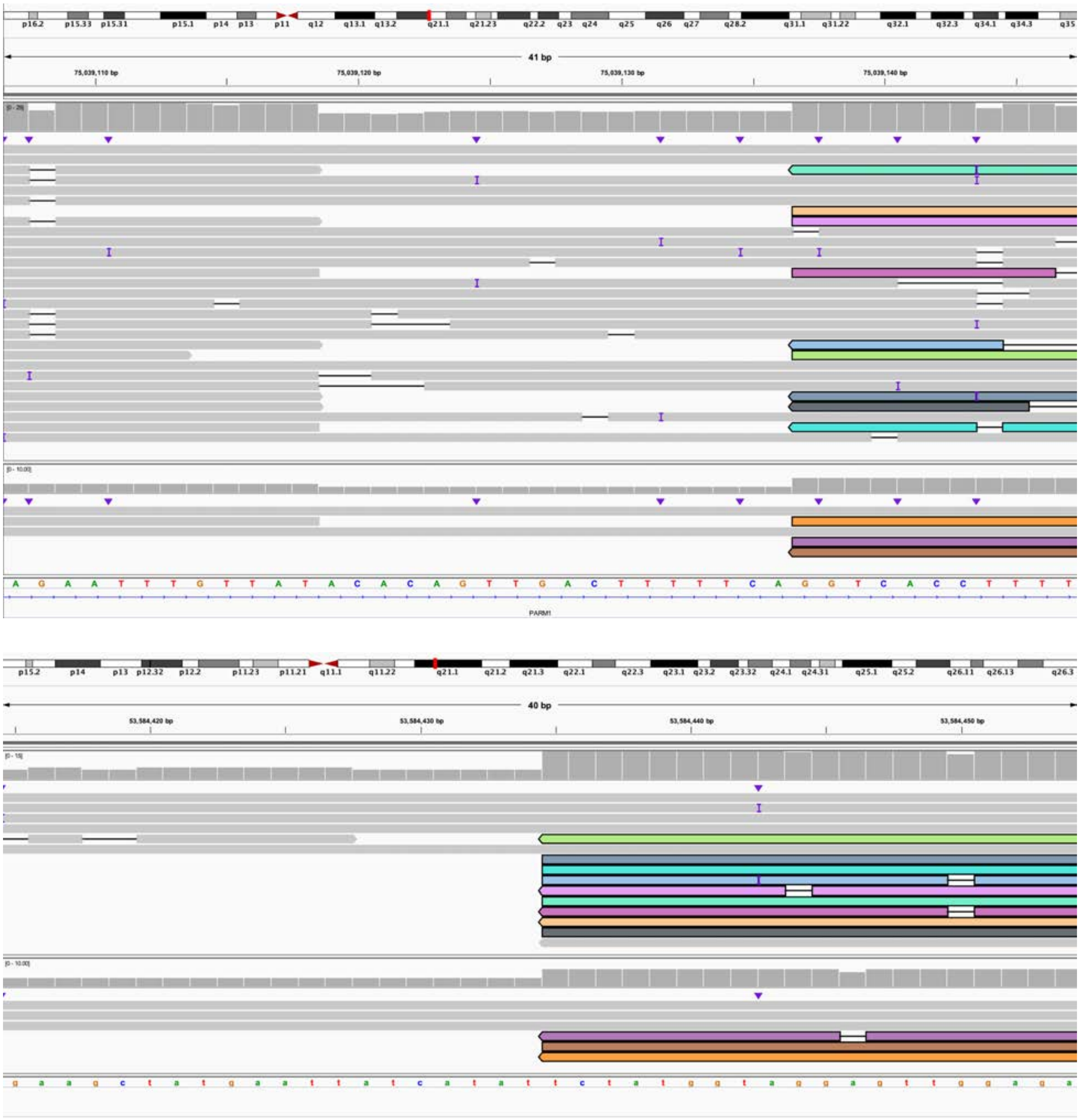
I: Karyotype of the case, arrows point to abnormal chromosomes.


p25.2 p24.3 p24.1 p23.2 p22.2 p21 p16.2 p15 p13.3 p12 p11.2 q11.1 q12.1 q13 q14.2 q21.1 q22.1 q22.3 q24.1 q24.3 q31.1 q31.3 q32.2 q33.1 q33.3 q35 q36.2 q37.1 q37.
41 bp
16,235,700 bp
16,235,710 bp
16,235,720 bp
16,235,730 bp
[0 - 29]
[0 - 10.00]
A G C T C C C A G A A A A C T G T C A G C A A T G C C A G C G G A T G C C C T C T
p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
40 bp
81,632,560 bp
81,632,570 bp
81,632,580 bp
81,632,59
[0 - 23]
[0 - 10.00]
C T G G A T G T T T G G C C A A C T T C T T C A T C A T G T A T T T G A C A C A

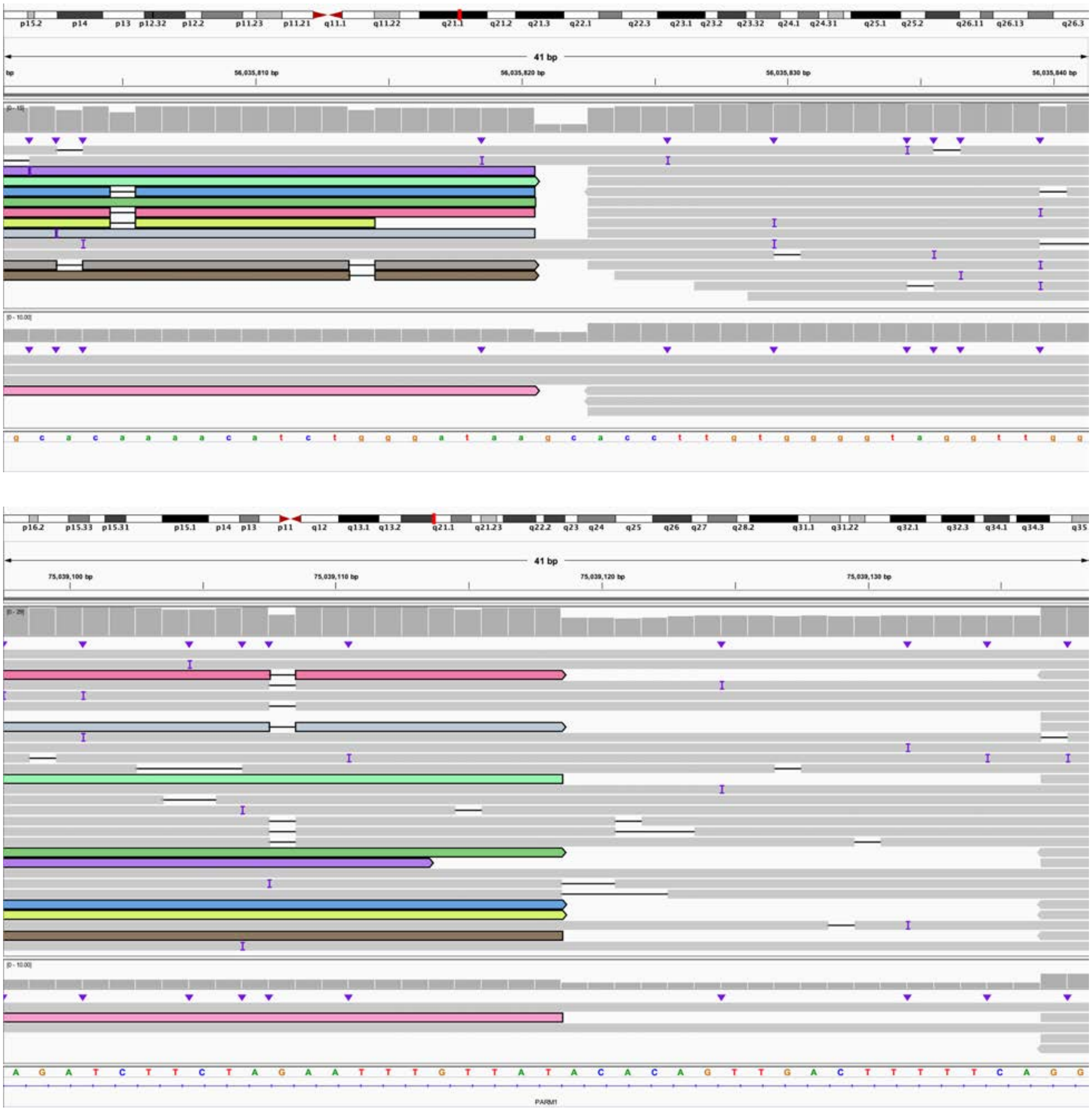
J. The 5' end of 2:C (IGV: chr2:16,235,693-16,235,733) is connected to the 5' end of 4:N (IGV: chr4:81,632,551-81,632,590). This is the entirety of derivative chromosome 2.


p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
572,950 bp
67,572,960 bp
67,572,970 bp
67,572,980 bp
[0 - 23]
[0 - 10.00]
A T T G T T C T A G T G A A A T G T C T C A G A A T T G C C C A T C T A C C A A T
STAP1
p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
77,012,160 bp
77,012,170 bp
77,012,180 bp
77,012,190 bp
[0 - 40]
[0 - 10.00]
T T T T T A A T T C T T A A A T G G G A G G T T G A A A T A T A A C T T C C A G G
SEPTIN11

K. The 3' end of 4:A bisects *STAP1* (IGV: chr4:67,572,950-67,572,990) and is connected to the 3' end of 4:K and bisects *SEPTIN11* (IGV: chr4:77,012,155-77,012,195). This is the beginning of derivative chromosome 4.

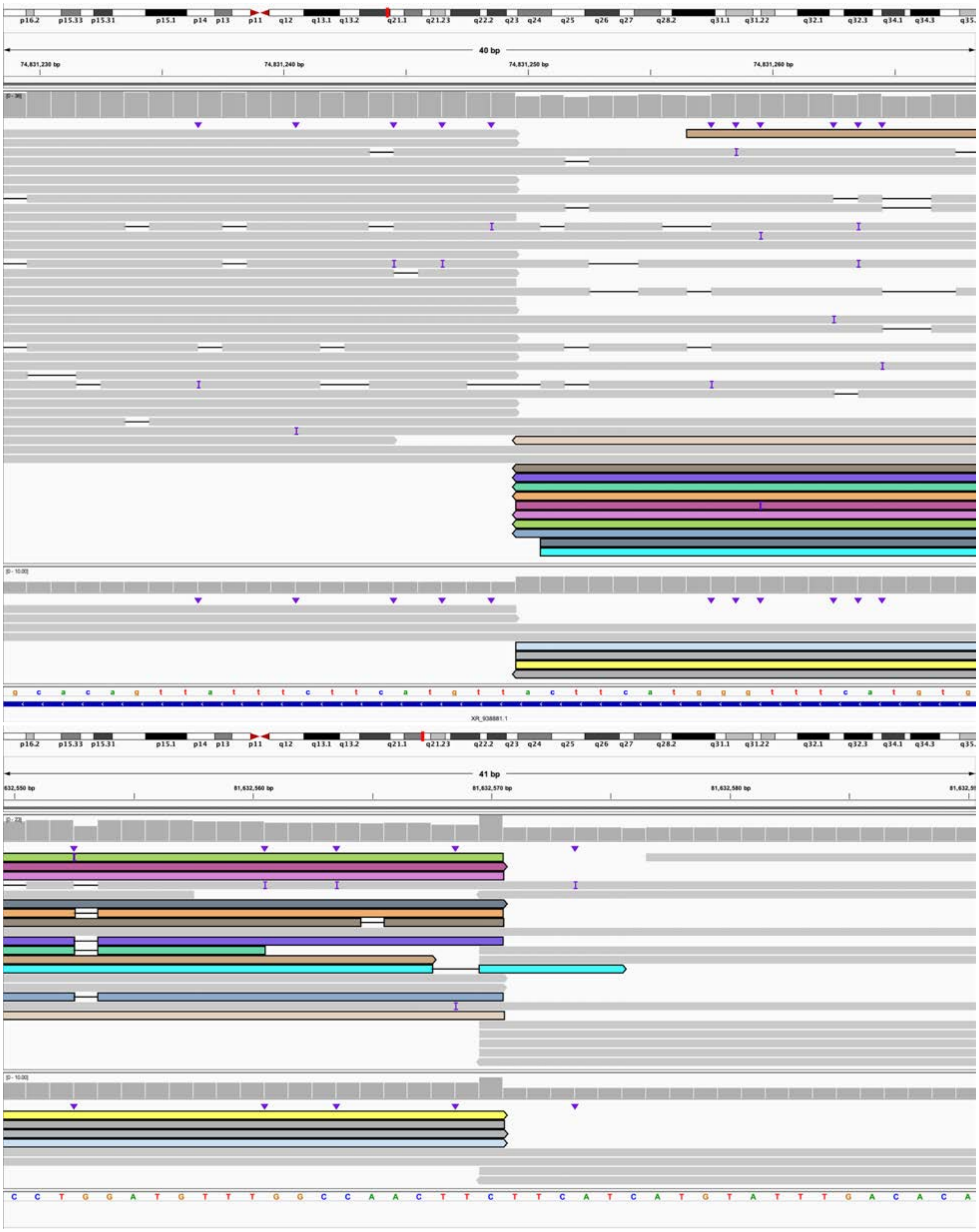

41 bp
75,039,110 bp
75,039,120 bp
75,039,130 bp
75,039,140 bp
A G A A T T T G T T A T A C A C A G T T G A C T T T T T C A G G T C A C C T T T T
PARM1
40 bp
53,584,420 bp
53,584,430 bp
53,584,440 bp
53,584,450 bp
g a a g c t a t g a a t t a t c a t a t t c t a t g g t a g g a g t t g g a g a

L. The 5' end of 4:K bisects *PARM1* (IGV: chr4:75,039,107-75,039,147) and is connected to the 5' end of 10:I (IGV: chr10:53,584,415-53,584,454). This is within derivative chromosome 4.

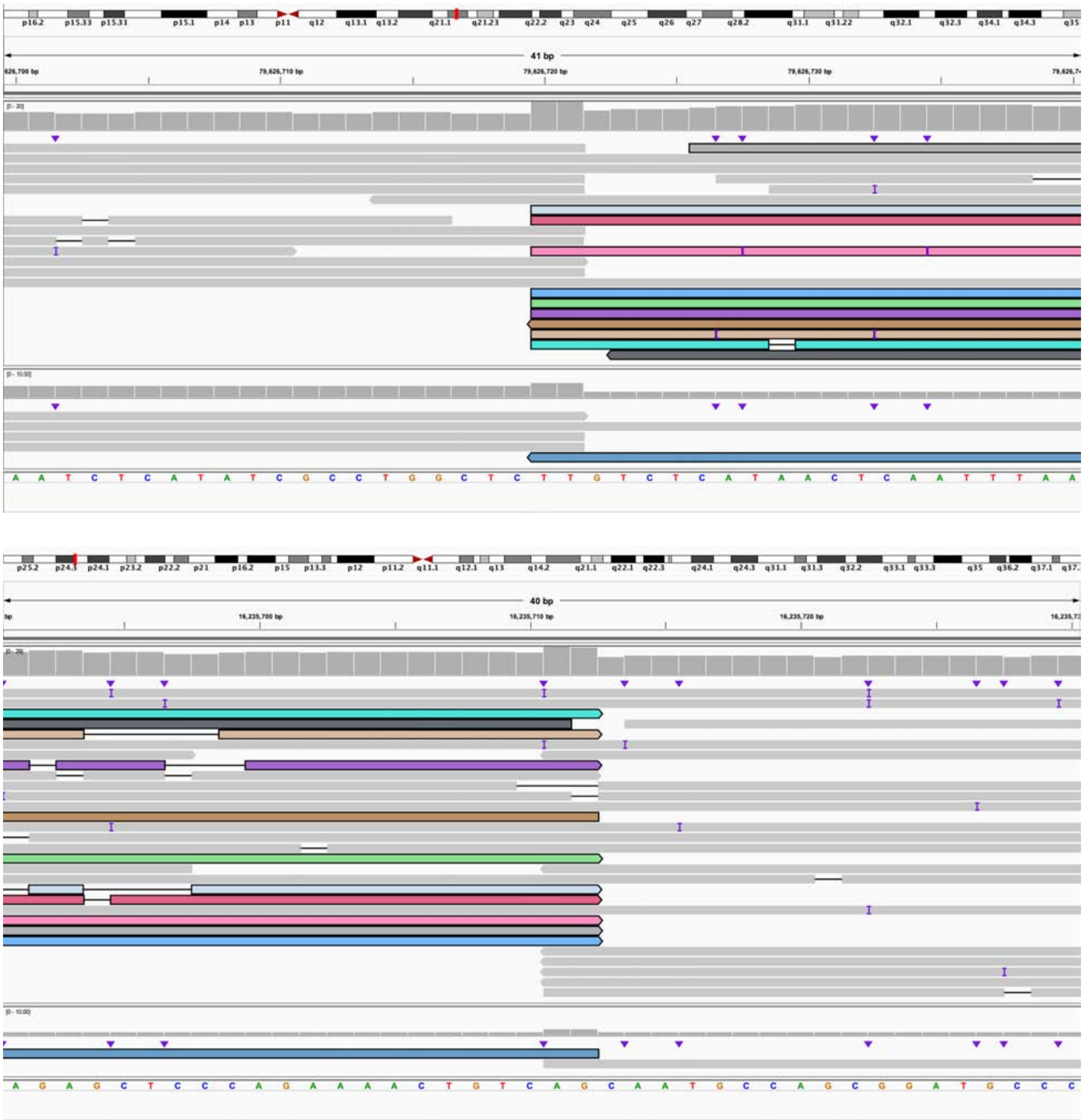


p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
56,035,810 bp
56,035,820 bp
56,035,830 bp
56,035,840 bp
g c a c a a a a c a t c t g g g a t a a g c a c c t t g t g g g g t a g g t t g g
p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
75,039,100 bp
75,039,110 bp
75,039,120 bp
75,039,130 bp
A G A T C T T C T A G A A T T T G T T A T A C A C A G T T G A C T T T T T C A G G
PARM1

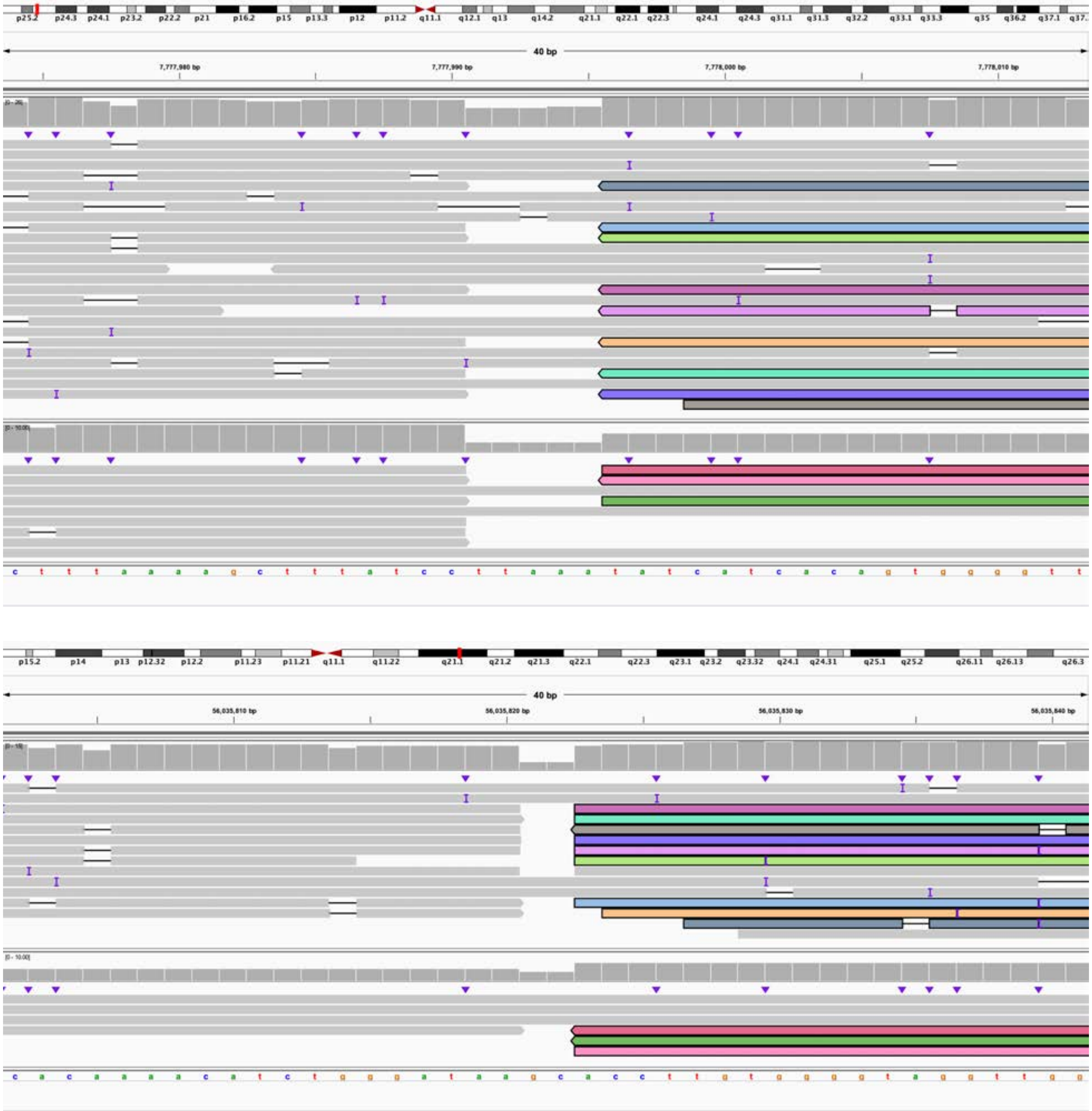
M. The 3' end of 10:I (IGV: chr10:56,035,801-56,035,841) is connected to the 3' end of 4:J and bisects *PARM1* (IGV: chr4:75,039,098-75,039,138). This is within derivative chromosome 4.


p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35.
40 bp
74,831,230 bp
74,831,240 bp
74,831,250 bp
74,831,260 bp
g c a c a g t t a t t t c t t c a t g t t a c t t c a t g g g t t t c a t g t g
XR_938881.1
p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35.
41 bp
632,550 bp
81,632,560 bp
81,632,570 bp
81,632,580 bp
C C T G G A T G T T T G G C C A A C T T C T T C A T C A T G T A T T T G A C A C A

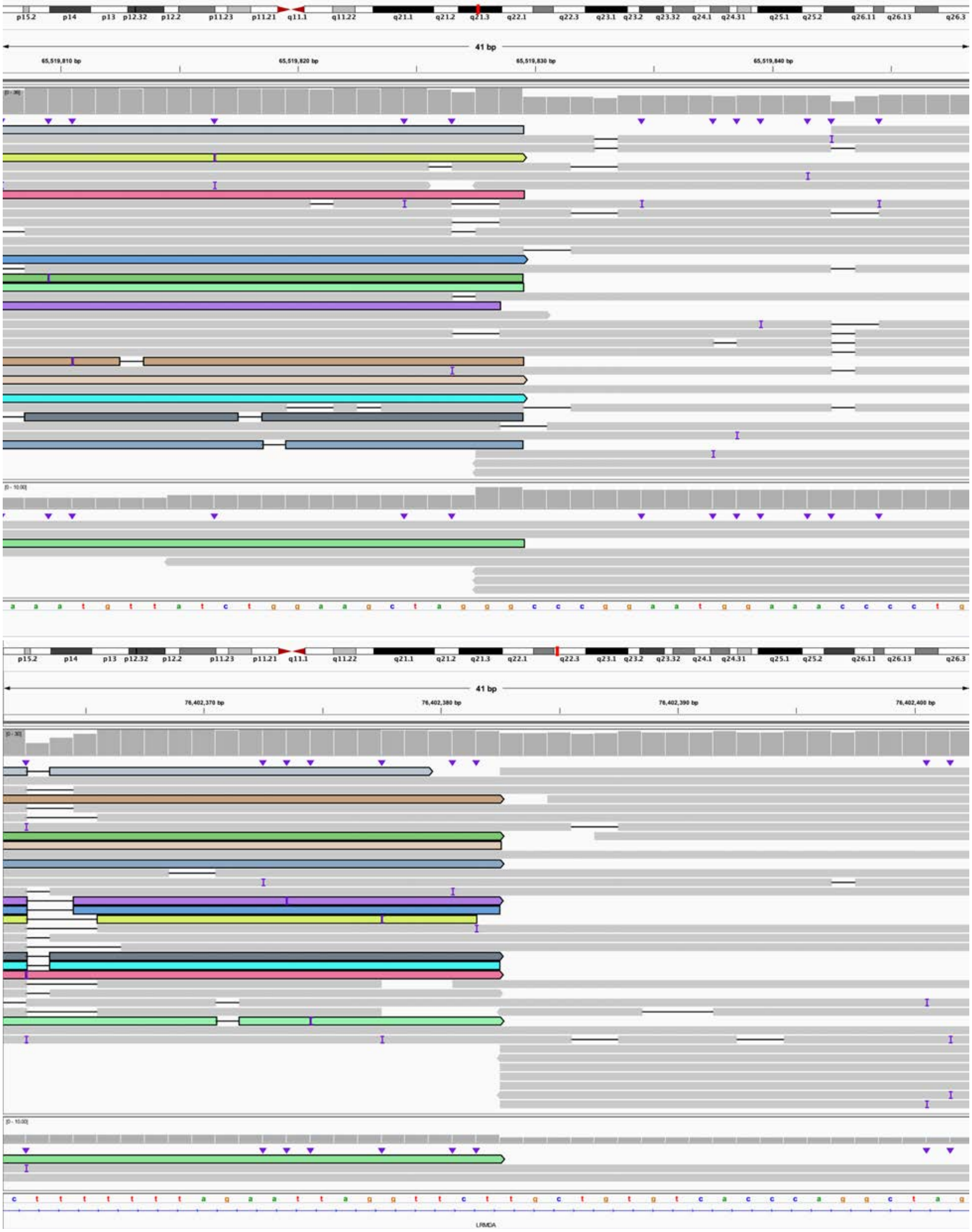
N. The 5' end of 4:J bisects *XR_938881.1* (IGV: chr4:74,831,229-74,831,268) and is connected to the 3' end of 4:M (IGV: chr4:81,632,550-81,632,590). This is within derivative chromosome 4.


p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
626,700 bp
79,626,710 bp
79,626,720 bp
79,626,730 bp
79,626,7
[0 - 20]
[0 - 10.00]
A A T C T C A T A T C G C C T G G C T C T T G T C T C A T A A C T C A A T T T A A
p25.2 p24.3 p24.1 p23.2 p22.2 p21 p16.2 p15 p13.3 p12 p11.2 q11.1 q12.1 q13 q14.2 q21.1 q22.1 q22.3 q24.1 q24.3 q31.1 q31.3 q32.2 q33.1 q33.3 q35 q36.2 q37.1 q37.
40 bp
bp
16,235,700 bp
16,235,710 bp
16,235,720 bp
16,235,73
[0 - 26]
[0 - 10.00]
A G A G C T C C C A G A A A A C T G T C A G C A A T G C C A G C G G A T G C C C

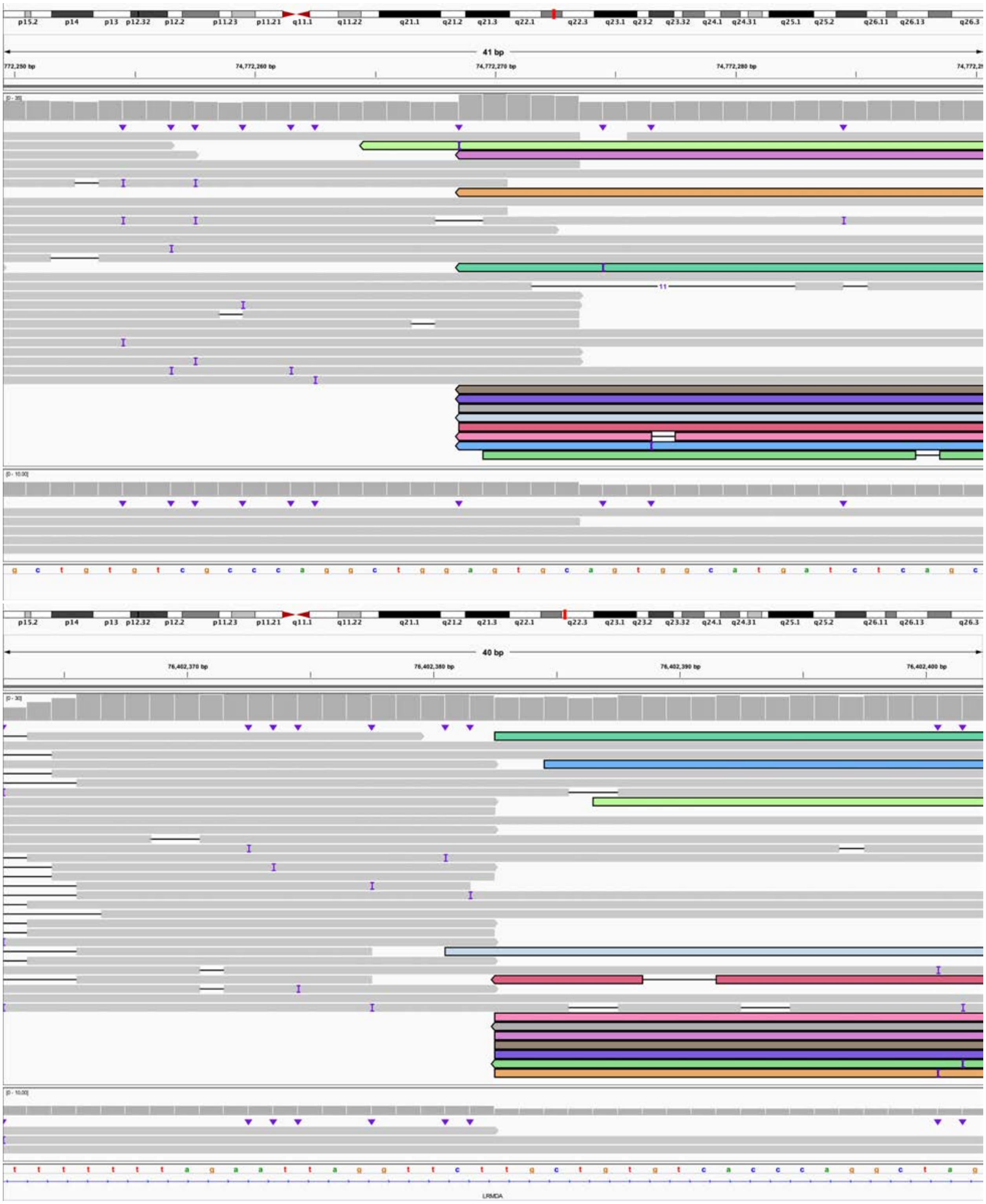
O. The 5' end of 4:M (IGV: chr4:79,626,700-79,626,740) is connected to the 3' end of 2:B (IGV: chr2:16,235,691-16,235,730). This is within derivative chromosome 4.


40 bp
7,777,980 bp
7,777,990 bp
7,778,000 bp
7,778,010 bp
c t t t a a a a g c t t t a t c c t t a a a t a t c a t c a c a g t g g g g t t
40 bp
56,035,810 bp
56,035,820 bp
56,035,830 bp
56,035,840 bp
c a c a a a a c a t c t g g g a t a a g c a c c t t g t g g g g t a g g t t g g

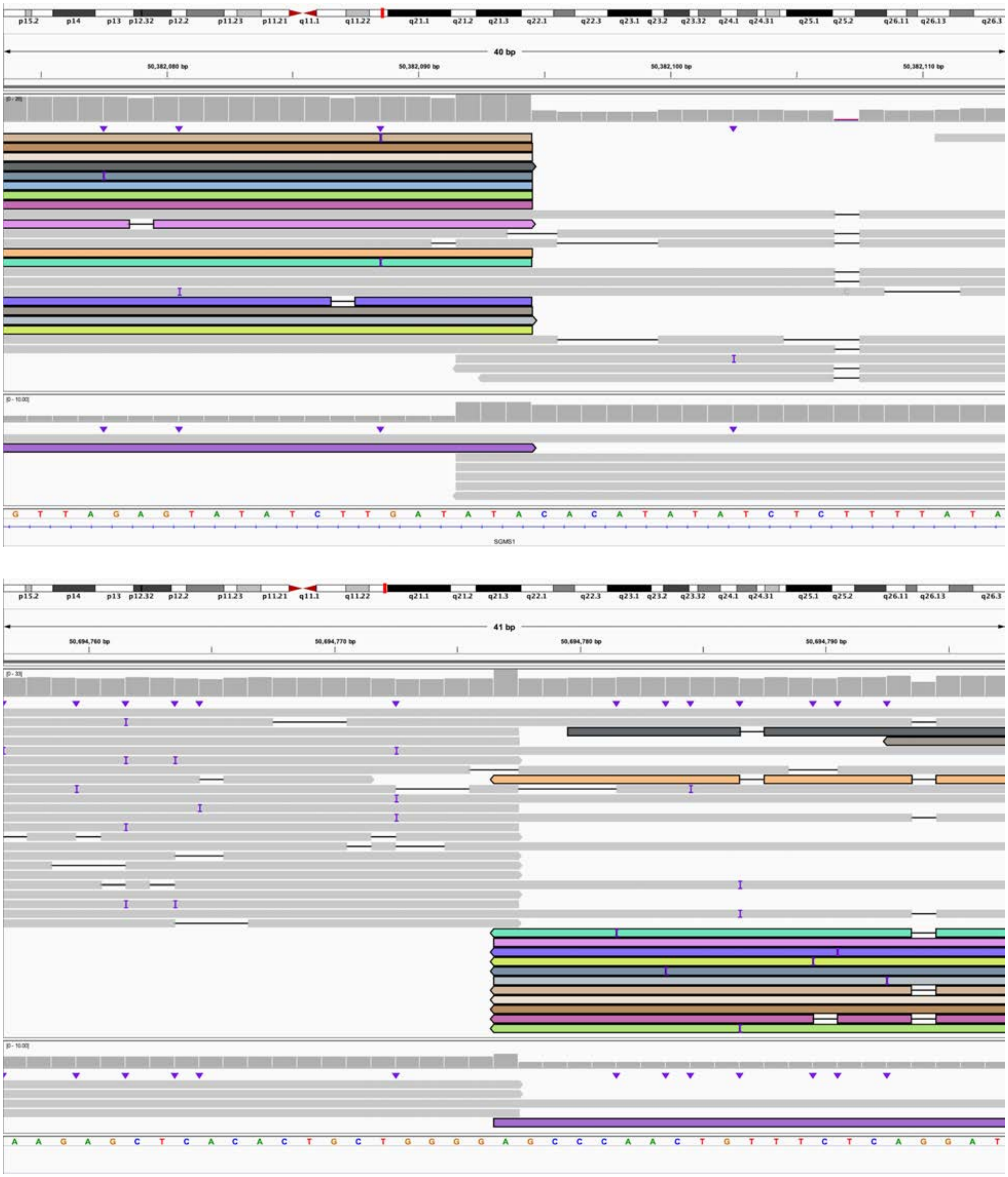
P. The 5' end of 2:B (IGV: chr2:7,777,974-7,778,013) is connected to the 5' end of 10:J (IGV: chr10:56,035,802-56,035,841). This is within derivative chromosome 4.


p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
65,519,810 bp
65,519,820 bp
65,519,830 bp
65,519,840 bp
[0 - 36]
[0 - 10.00]
a a a t g t t a t c t g g a a g c t a g g g c c c g g a a t g g a a a c c c c t g
p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
76,402,370 bp
76,402,380 bp
76,402,390 bp
76,402,400 bp
[0 - 30]
[0 - 10.00]
c t t t t t t t a g a a t t a g g t t c t t g c t g t g t c a c c c a g g c t a g
LRMDA

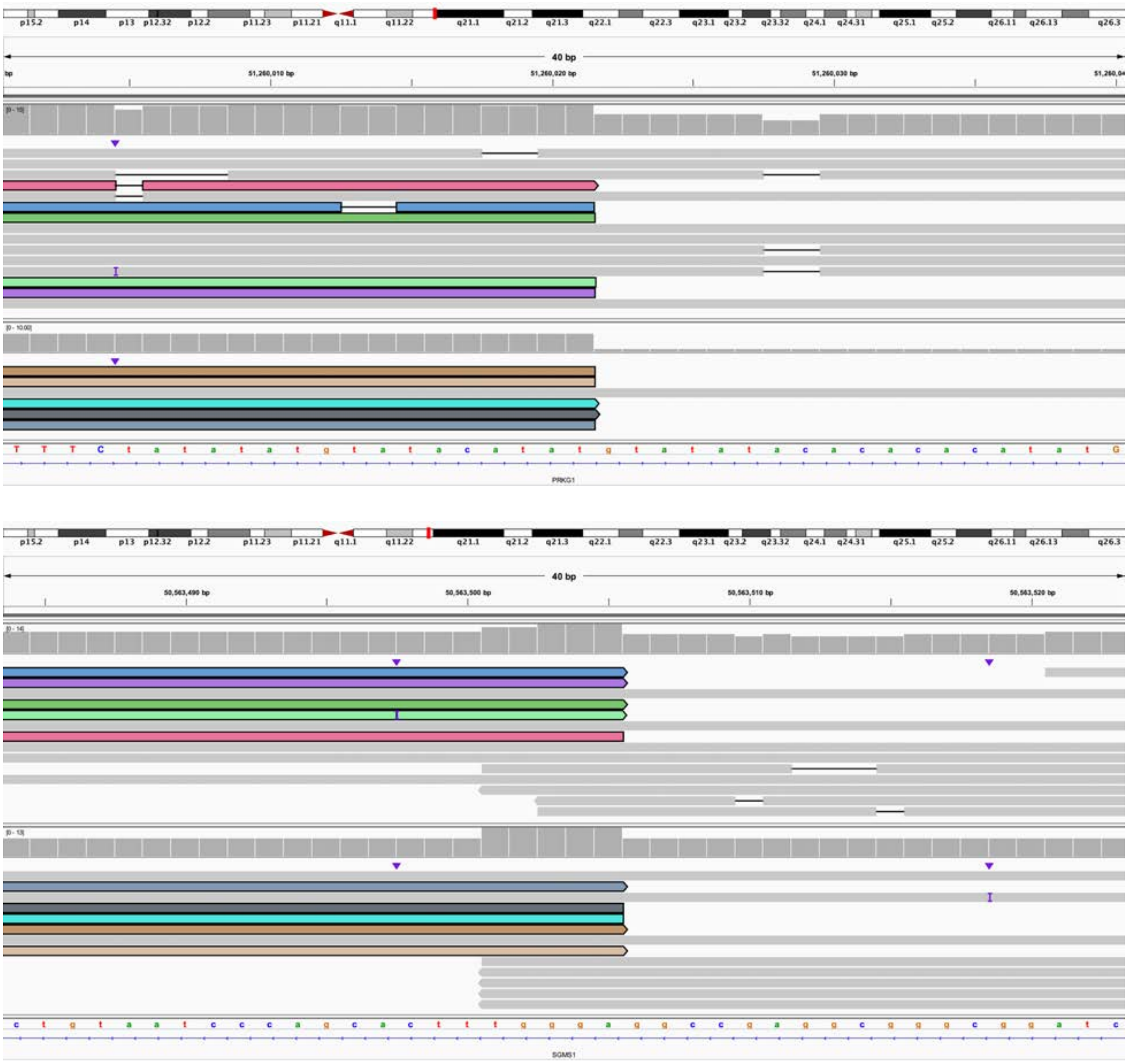
Q. The 3' end of 10:J (IGV: chr10:65,519,808-65,519,848) is connected to the 3' end of 10:M and bisects *LRMDA* (IGV: chr10:76,402,362-76,402,402). This is within derivative chromosome 4.


p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
772,250 bp
74,772,260 bp
74,772,270 bp
74,772,280 bp
74,772,2
[0 - 30]
11
[0 - 10.00]
g c t g t g t c g c c c a g g c t g g a g t g c a g t g g c a t g a t c t c a g c
p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
40 bp
76,402,370 bp
76,402,380 bp
76,402,390 bp
76,402,400 bp
[0 - 30]
[0 - 10.00]
t t t t t t t a g a a t t a g g t t c t t g c t g t g t c a c c c a g g c t a g
LRMDA

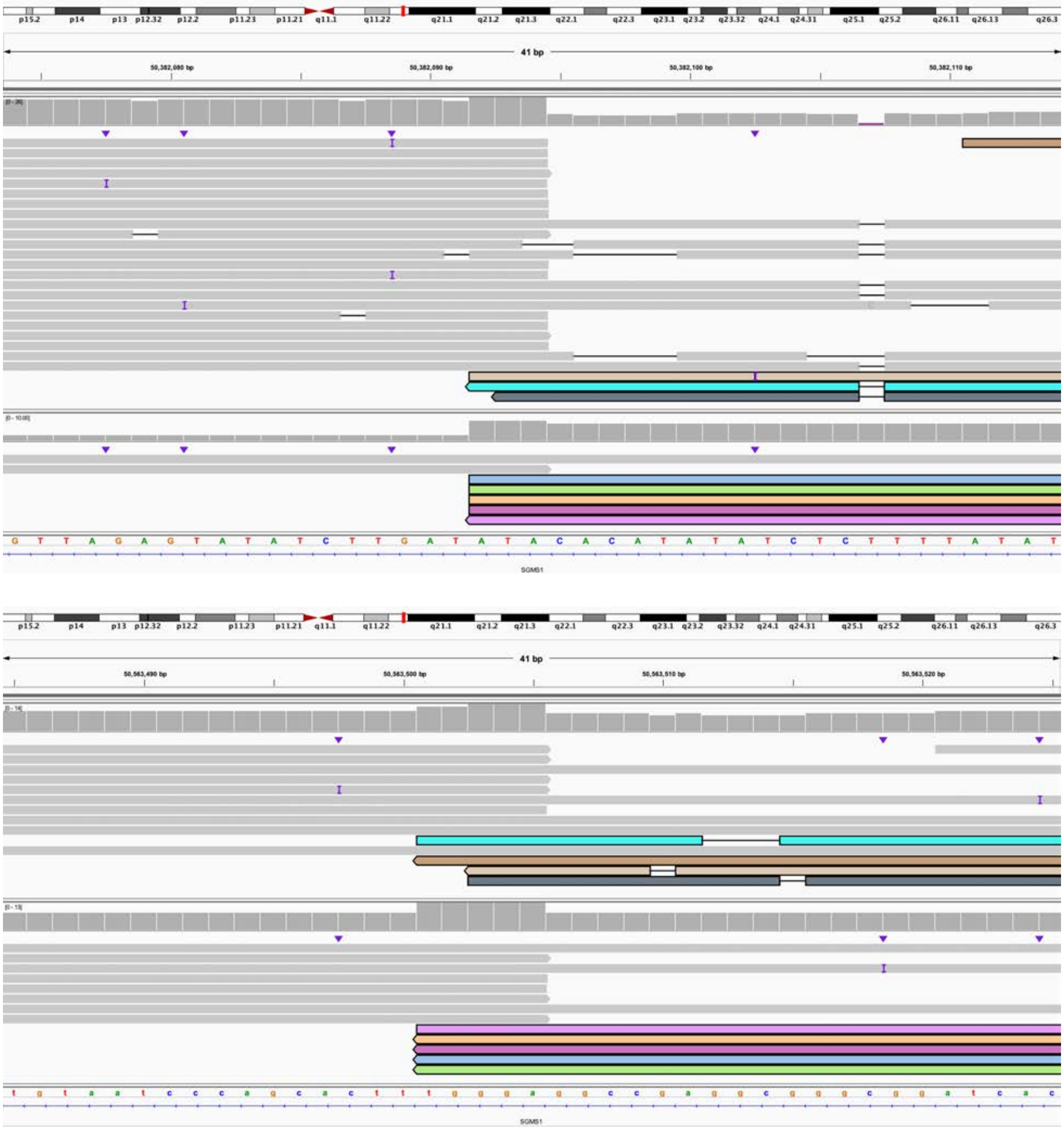
R. The 5' end of 10:M (IGV: chr10:74,772,250-74,772,290) is connected to the 5' end of 10:N and bisects *LRMDA* (IGV: chr10:76,402,363-76,402,402). This is the end of derivative chromosome 4.


p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
40 bp
50,382,080 bp
50,382,090 bp
50,382,100 bp
50,382,110 bp
[0 - 20]
[0 - 10.00]
G T T A G A G T A T A T C T T G A T A T A C A C A T A T A T C T C T T T T T A T A
SGMS1
p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
50,694,760 bp
50,694,770 bp
50,694,780 bp
50,694,790 bp
[0 - 33]
[0 - 10.00]
A A G A G C T C A C A C T G C T G G G G A G C C C A A C T G T T T C T C A G G A T

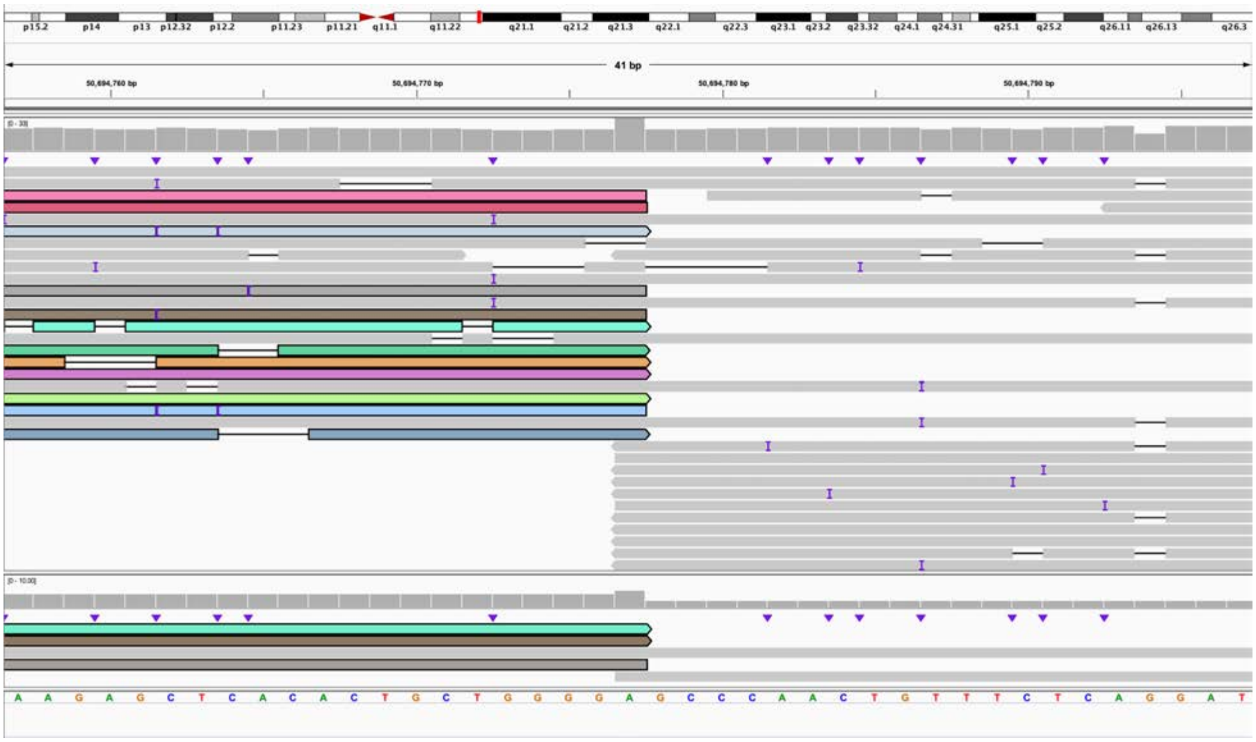
S. The 3' end of 10:A bisects *SGMS1* (IGV: chr10:50,382,074-50,382,113) and is connected to the 5' end of 10:D (IGV: chr10:50,694,757-50,694,797). This is the beginning of derivative chromosome 10.

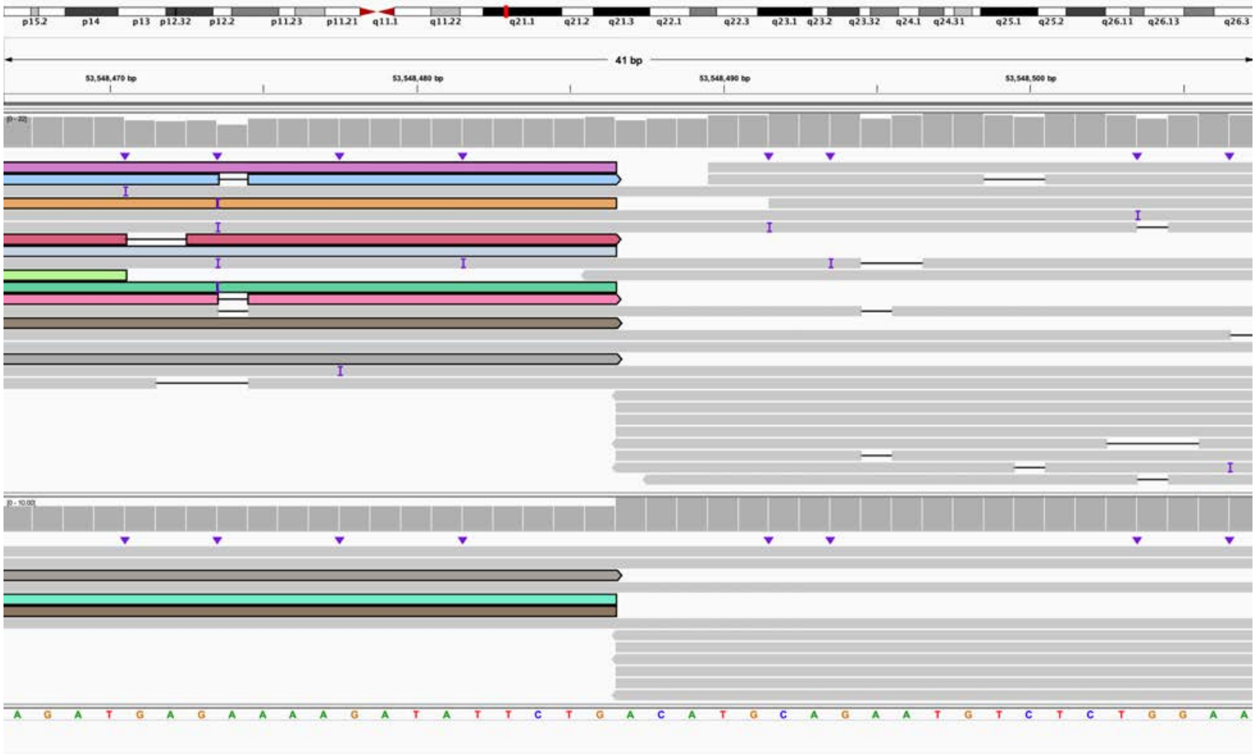

p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
40 bp
51,260,010 bp
51,260,020 bp
51,260,030 bp
T T T C t a t a t a t g t a t a c a t a t g t a t a t a c a c a c a c a t a t G
PRKG1
40 bp
50,563,490 bp
50,563,500 bp
50,563,510 bp
50,563,520 bp
c t g t a a t c c c a g c a c t t t g g g a g g c c g a g g c g g g c g g a t c
SGMS1

T. The 3' end of 10:D bisects *PRKG1* (IGV: chr10:51,260,001-51,260,040) and is connected to the 3' end of 10:B which bisects *SGMS1* (IGV: chr10:50,563,484-50,563,523). This is within derivative chromosome 10.

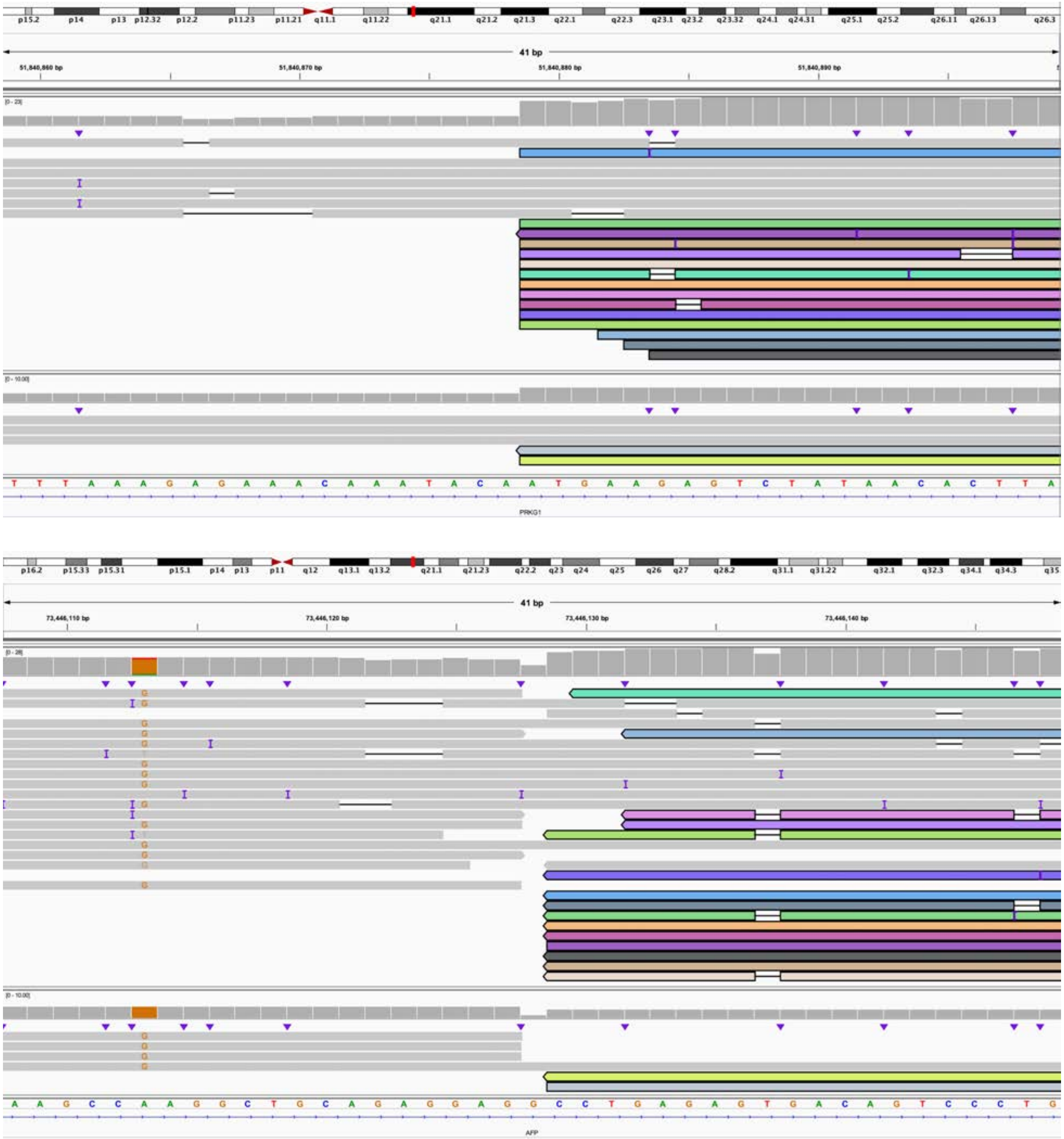

p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
50,382,080 bp
50,382,090 bp
50,382,100 bp
50,382,110 bp
[0 - 26]
[0 - 10.00]
G T T A G A G T A T A T C T T G A T A T A C A C A T A T A T C T C T T T T A T A T
SGMS1
p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
50,563,490 bp
50,563,500 bp
50,563,510 bp
50,563,520 bp
[0 - 14]
[0 - 13]
t g t a a t c c c a g c a c t t t g g g a g g c c g a g g c g g g c g g a t c a c
SGMS1

U. The 5' end of 10:B bisects *SGMS1* (IGV: chr10:50,382,074-50,382,114) and is connected to the 5' end of 10:C and bisects *SGMS1* (IGV: chr10:50,563,485-50,563,525). This is within derivative chromosome 10.


p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
50,694,760 bp
50,694,770 bp
50,694,780 bp
50,694,790 bp
[0 - 33]
[0 - 10.00]
A A G A G C T C A C A C T G C T G G G G A G C C C A A C T G T T T C T C A G G A T

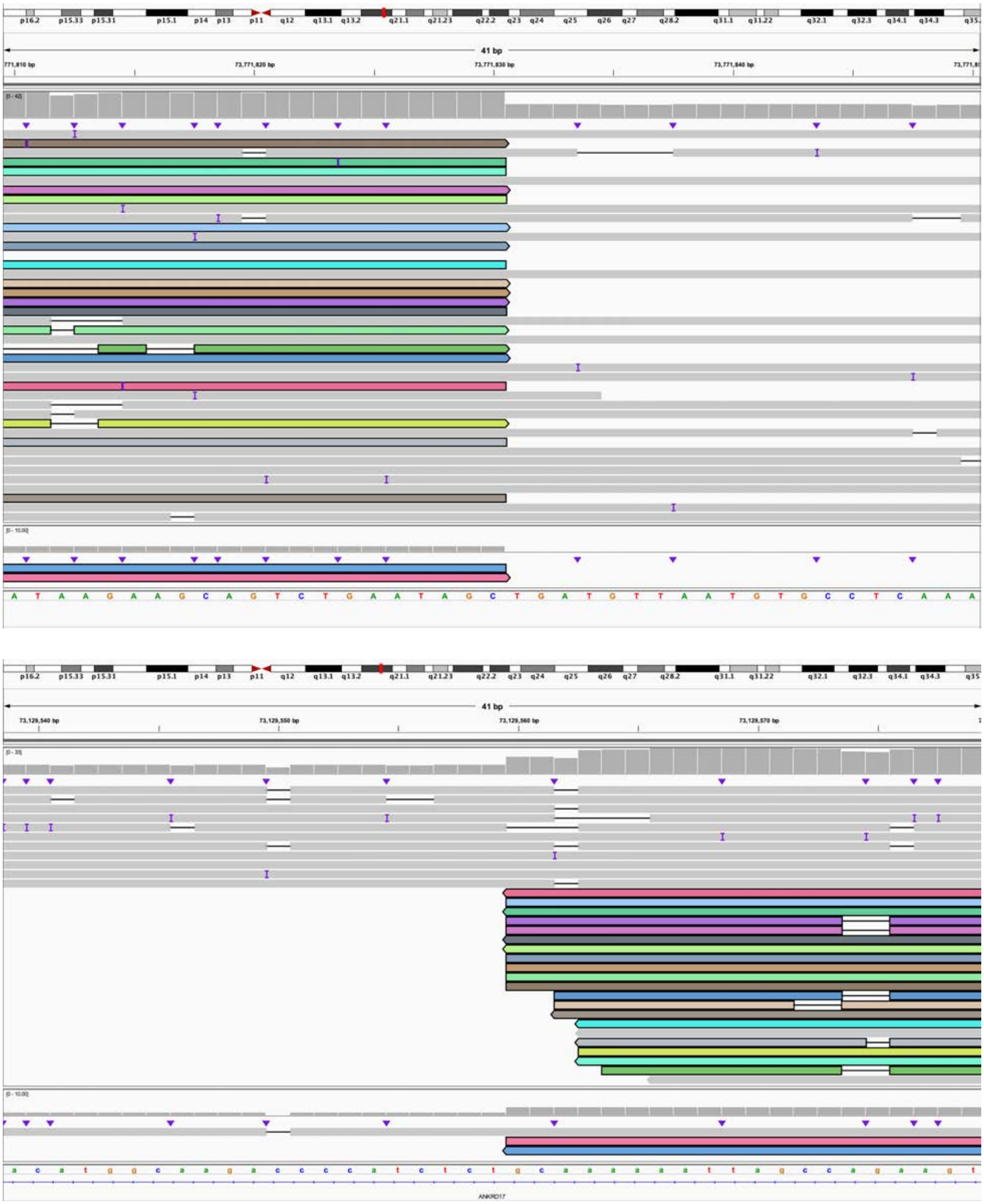

p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
53,548,470 bp
53,548,480 bp
53,548,490 bp
53,548,500 bp
[0 - 22]
[0 - 10.00]
A G A T G A G A A A A G A T A T T C T G A C A T G C A G A A T G T C T C T G G A A

V. The 3' end of 10:C (IGV: chr10:50,694,757-50,694,797) is connected to the 3' end of 10:F (IGV: chr10:53,548,467-53,548,507). This is within derivative chromosome 10.

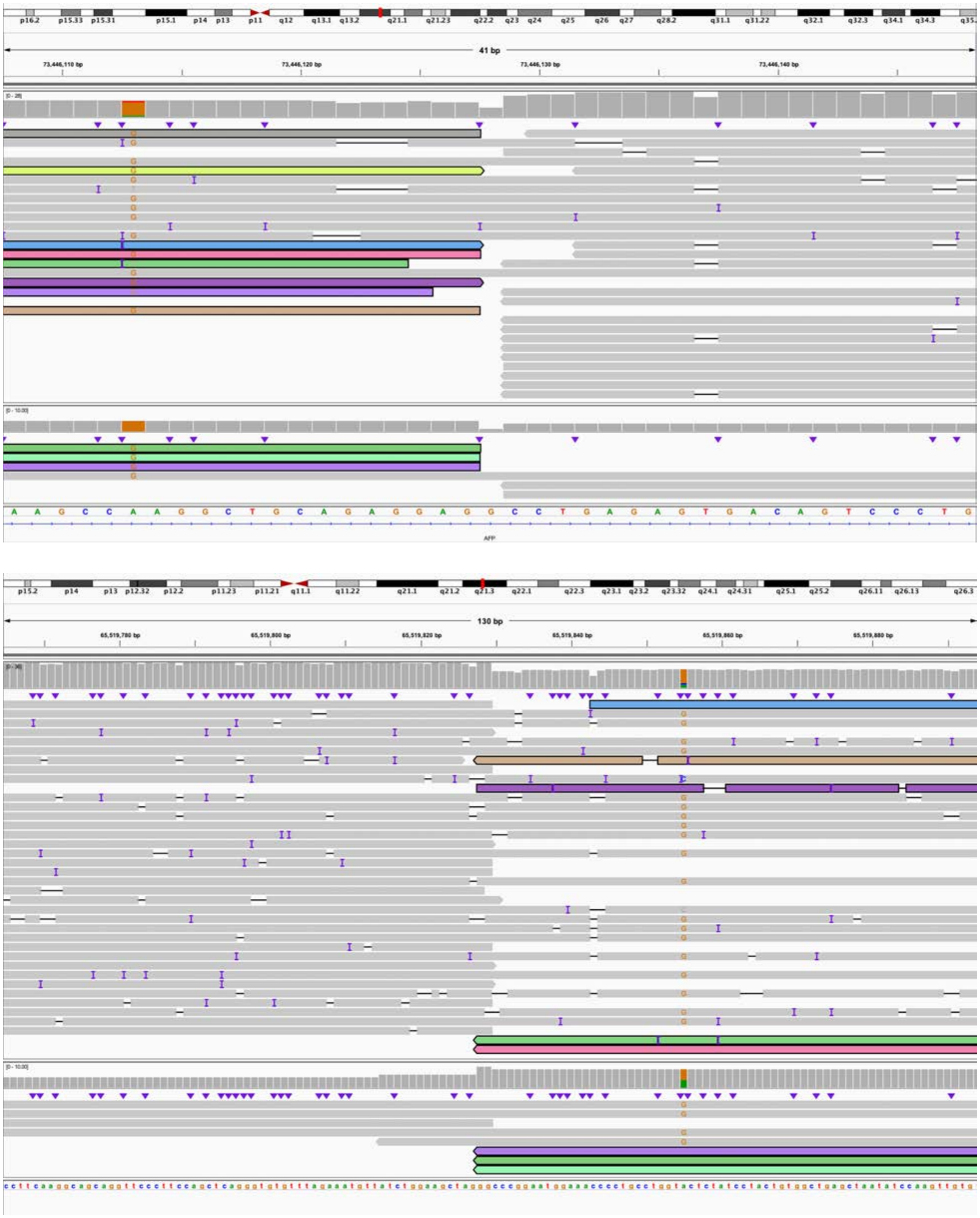


p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
51,840,860 bp
51,840,870 bp
51,840,880 bp
51,840,890 bp
[0 - 23]
[0 - 10.00]
T T T A A A G A G A A A C A A A T A C A A T G A A G A G T C T A T A A C T T A
PRKG1
p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
73,446,110 bp
73,446,120 bp
73,446,130 bp
73,446,140 bp
[0 - 28]
[0 - 10.00]
A A G C C A A G G C T G C A G G A G G C C T G A G A G T G A C A G T C C C T G
AFP

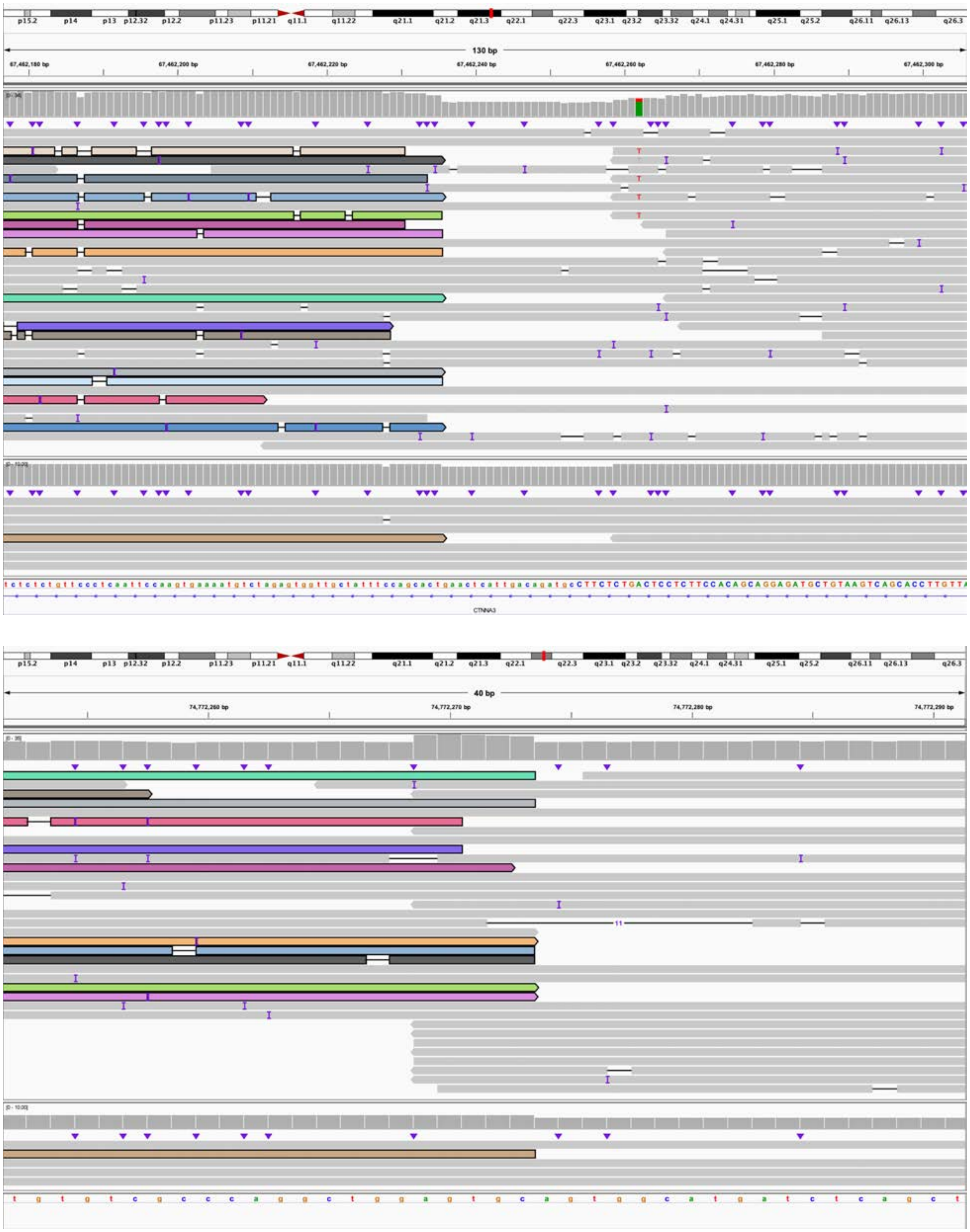
W. The 5' end of 10:F bisects *PRKG1* (IGV: chr10:51,840,859-51,840,899) and is connected to the 5' end of 4:F and bisects *AFP* (IGV: chr4:73,446,108-73,446,148). This is within derivative chromosome 10.


p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
771,810 bp
73,771,820 bp
73,771,830 bp
73,771,840 bp
[0 - 42]
[0 - 10.00]
A T A A G A A G C A G T C T G A A T A G C T G A T G T T A A T G T G C C T C A A A
p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
73,129,540 bp
73,129,550 bp
73,129,560 bp
73,129,570 bp
[0 - 33]
[0 - 10.00]
a c a t g g c a a g a c c c c a t c t c t g c a a a a a a t t a g c c a g a a g t
ANKRD17

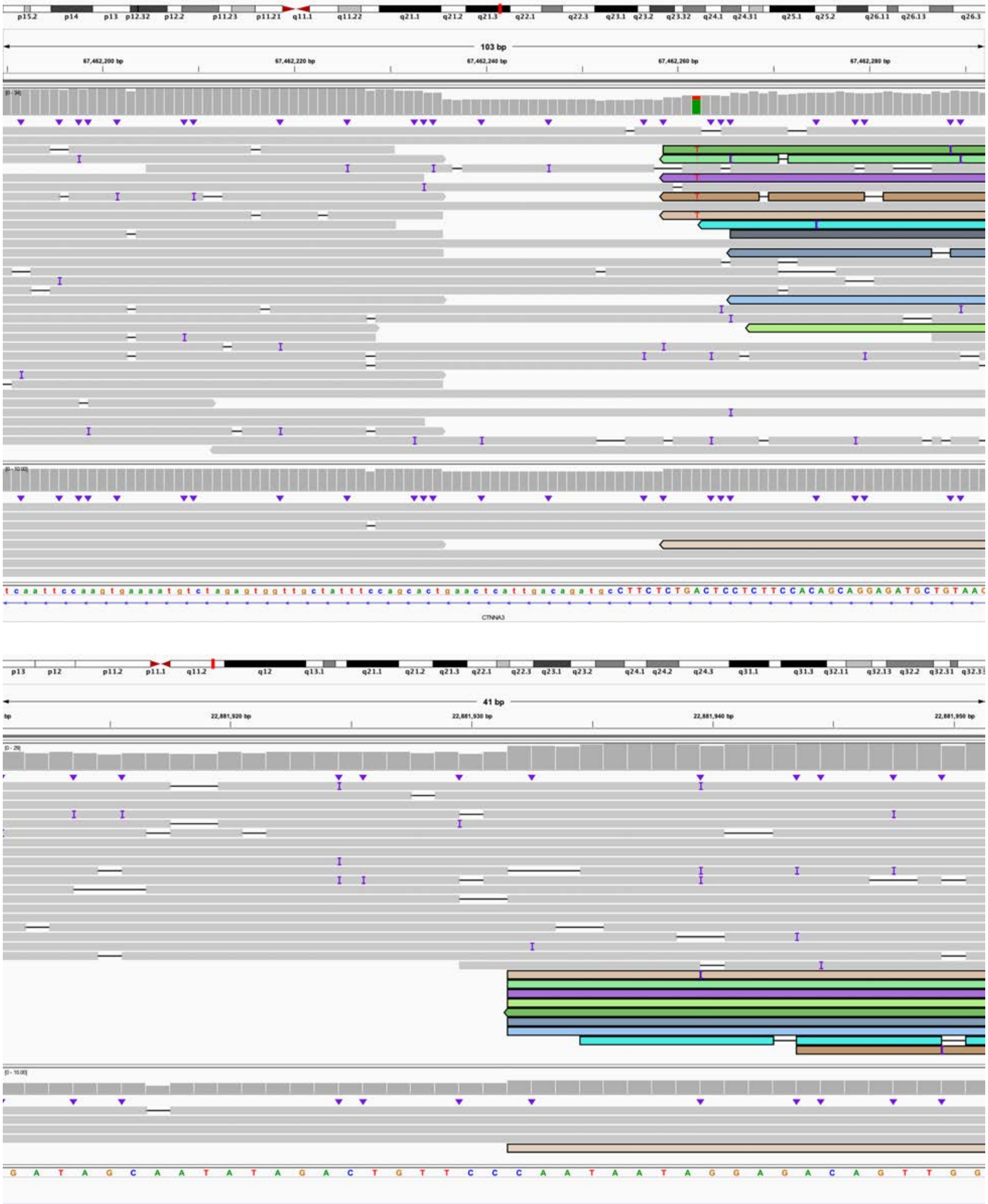
X. The 3' end of 4:F (IGV: chr4:73,771,810-73,771,850) is connected to the 5' end of 4:E which bisects *ANKRD17* (IGV: chr4:73,129,539-73,129,579). This is within derivative chromosome 10.



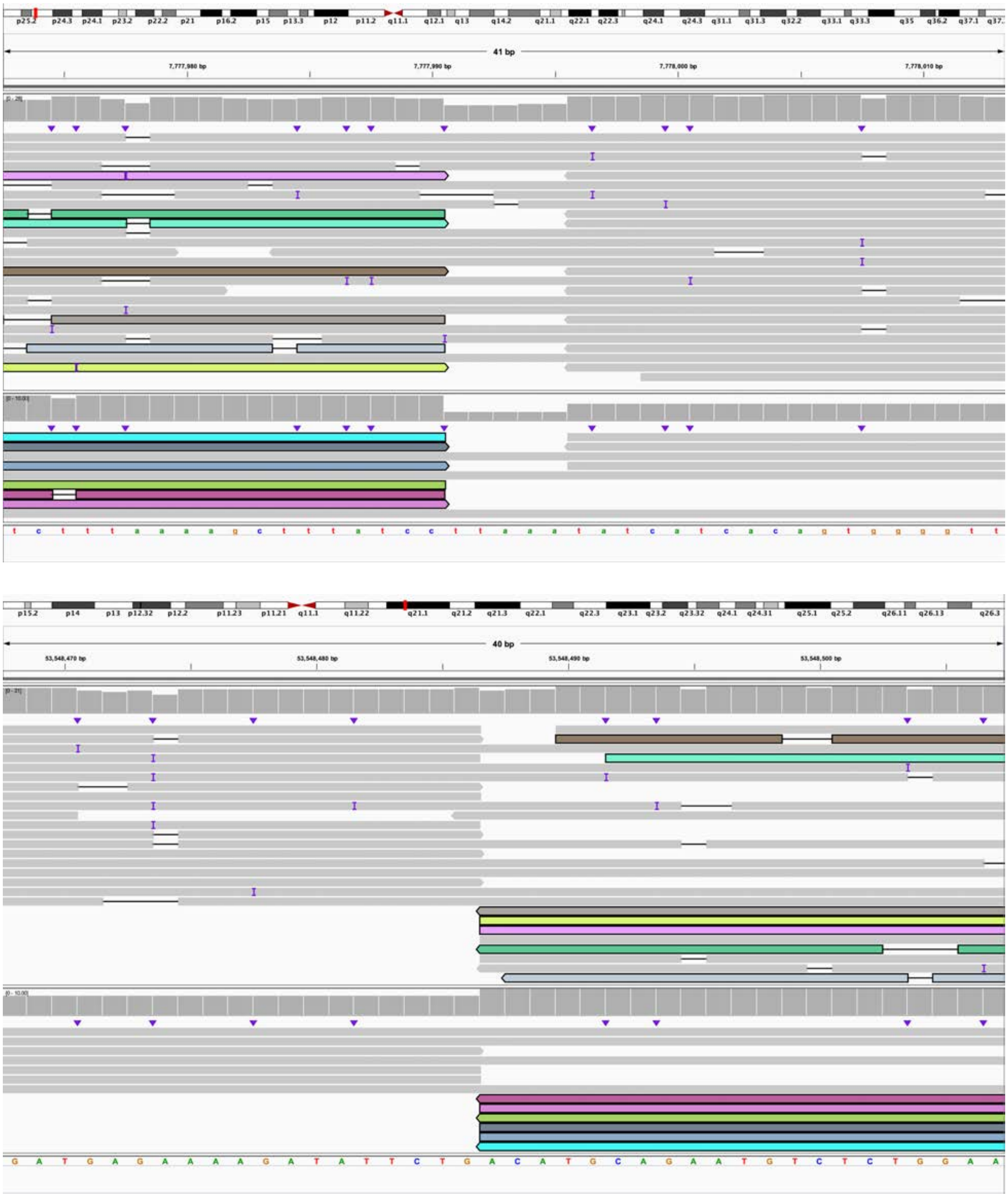
Y. The 3' end of 4:E bisects *AFP* (IGV: chr4:73,446,108-73,446,148) and is connected to the 5' end of 10:K (IGV: chr10:65,519,765-65,519,894). This is within derivative chromosome 10.


p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
130 bp
67,462,180 bp
67,462,200 bp
67,462,220 bp
67,462,240 bp
67,462,260 bp
67,462,280 bp
67,462,300 bp
tctctctgttccctcaattccaagtgaaaatgtctagagtggttgctatttccagcactgaactcattgacagatgcCTTCTCTGACTCCTCTTCCACAGCAGGAGATGCTGTAAGTCAGCACCTTGTTA
CTNNA3
p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
40 bp
74,772,260 bp
74,772,270 bp
74,772,280 bp
74,772,290 bp
11
t g t g t c g c c c a g g c t g g a g t g c a g t g g c a t g a t c t c a g c t

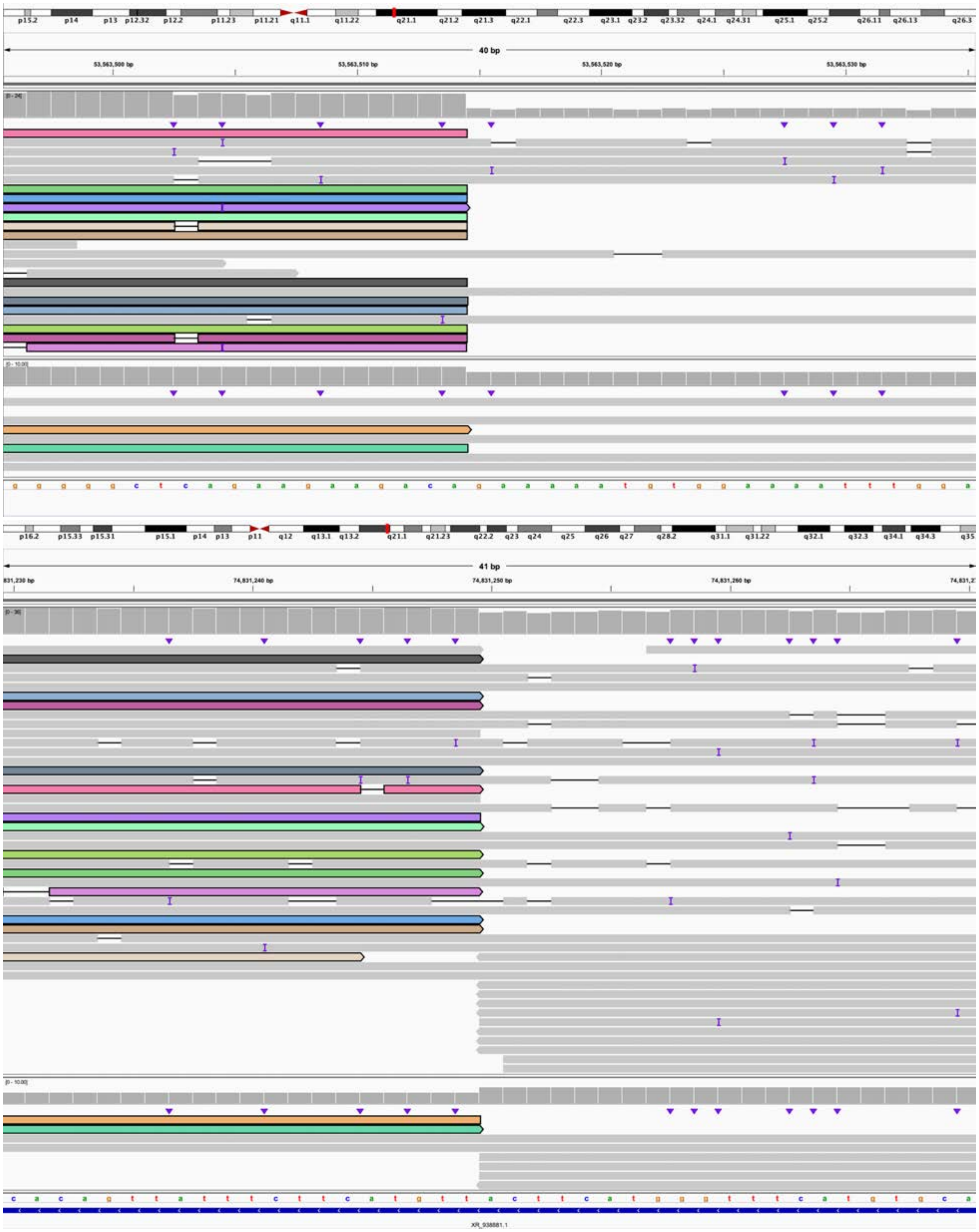
Z. The 3' end of 10:K bisects *CTNNA3* (IGV: chr10:67,462,177-67,462,306) and is connected to the 3' end of 10:L (IGV: chr10:74,772,252-74,772,291). This is within derivative chromosome 10.


103 bp
67,462,200 bp
67,462,220 bp
67,462,240 bp
67,462,260 bp
67,462,280 bp
tcaattccaagtgaaaatgtctagagtggttgctatttccagcactgaactcattgacagatgcCTTCTCTGACTCCTCTTCCACAGCAGGAGATGCTGTAAC
CTNNA3
41 bp
22,881,920 bp
22,881,930 bp
22,881,940 bp
22,881,950 bp
G A T A G C A A T A T A G A C T G T T C C C A A T A A T A G G A G A C A G T T G G

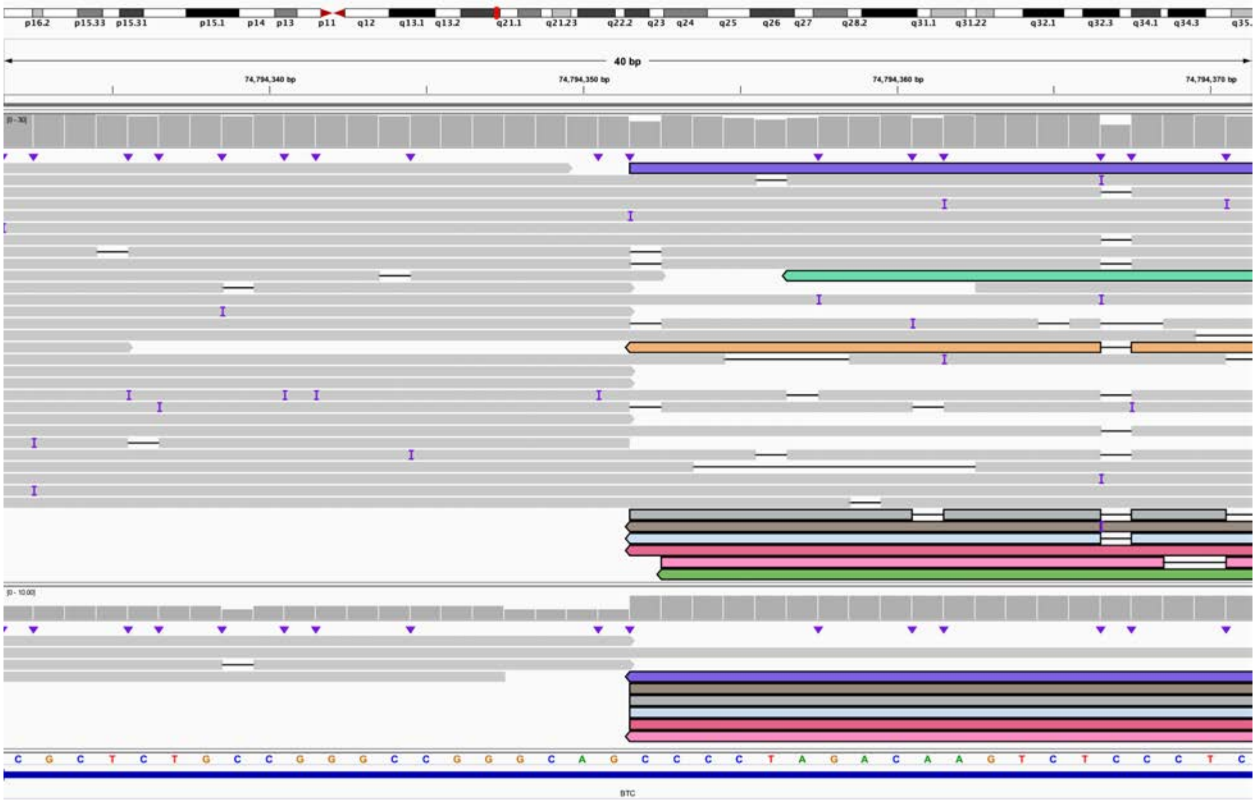
AA. The 5' end of 10:L bisects *CTNNA3* (IGV: chr10:67,462,190-67,462,292) and is connected to the 5' end of 14:C (IGV: chr14:22,881,911-22,881,951). This is the end of derivative chromosome 10.

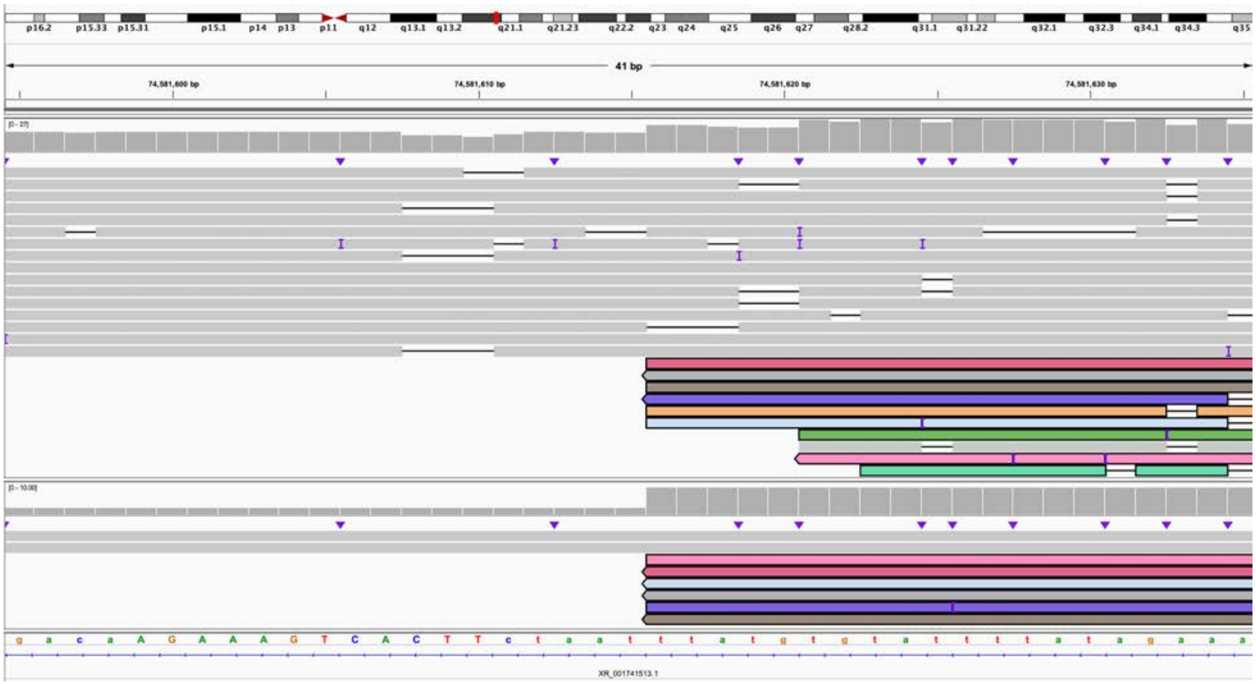

41 bp
7,777,980 bp
7,777,990 bp
7,778,000 bp
7,778,010 bp
t c t t t a a a a g c t t t a t c c t t a a a t a t c a t c a c a g t g g g t t
40 bp
53,548,470 bp
53,548,480 bp
53,548,490 bp
53,548,500 bp
G A T G A G A A A A G A T A T T C T G A C A T G C A G A A T G T C T C T G G A A

AB. The 3' end of 2:A (IGV chr2:7,777,973-7,778,013) is connected to the 5' end of 10:G (IGV chr10:53,548,468-53,548,507). This is the beginning of derivative chromosome 14.

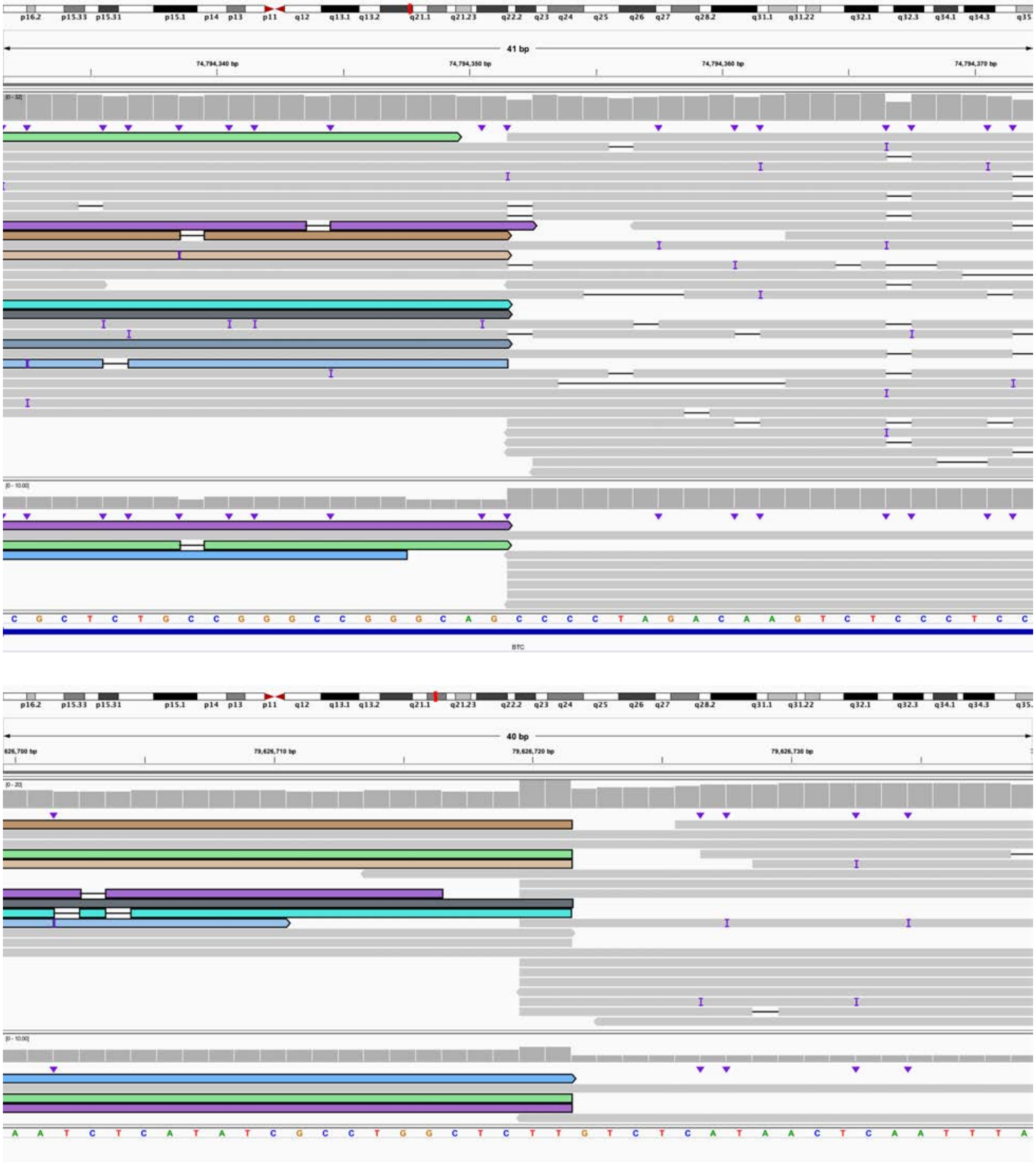


AC. The 3' end of 10:G (IGV: chr10:53,563,496-53,563,535) is connected to the 3' end of 4:I and bisects *XR_938881.1* (IGV: chr4:74,831,230-74,831,270). This is within derivative chromosome 14.

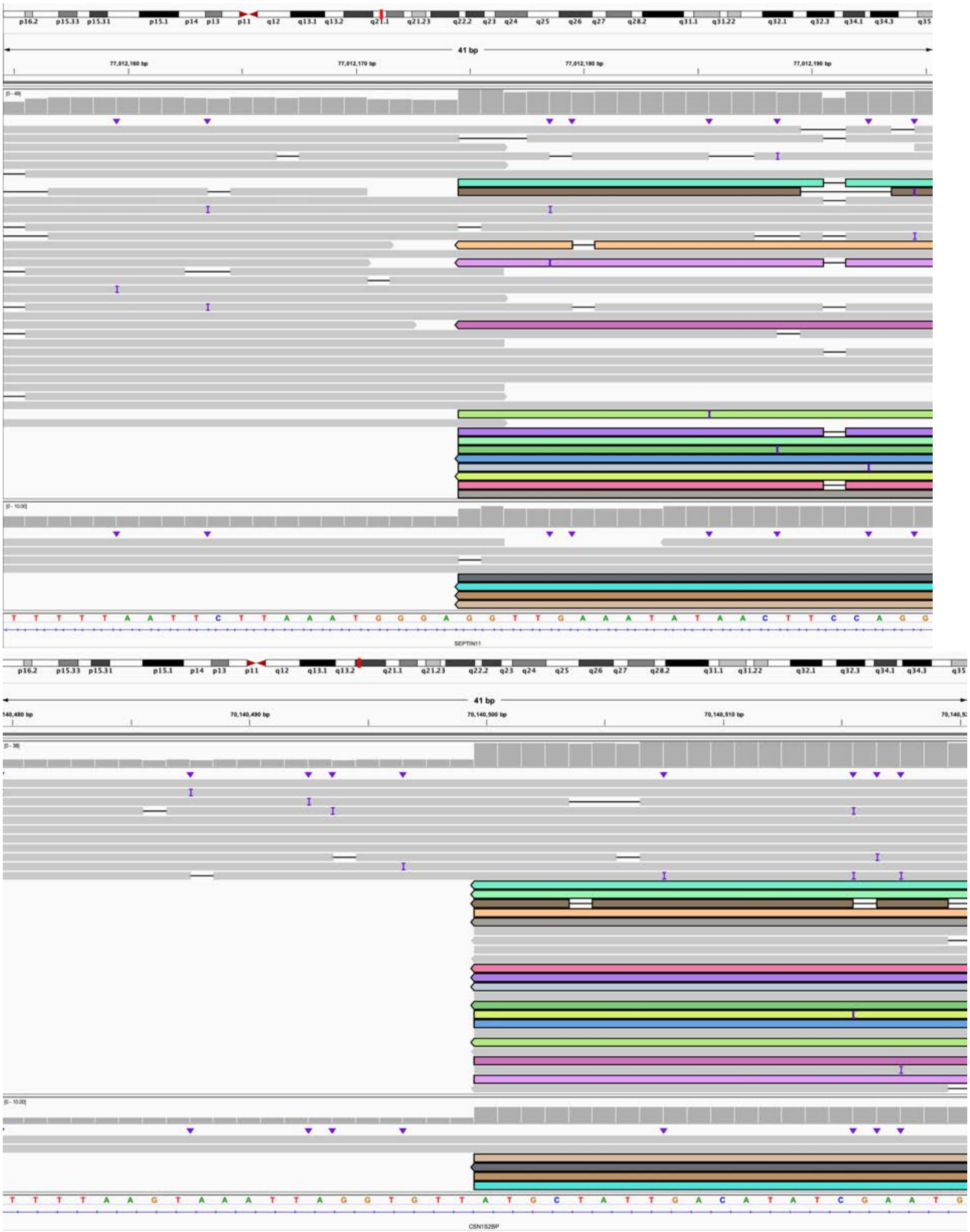

p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
40 bp
74,794,340 bp
74,794,350 bp
74,794,360 bp
74,794,370 bp
C G C T C T G C C G G G C C G G G C A G C C C C T A G A C A A G T C T C C C T C
BTC


p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
74,581,600 bp
74,581,610 bp
74,581,620 bp
74,581,630 bp
g a c a A G A A A G T C A C T T c t a a t t t a t g t g t a t t t t a t a g a a a
XR_001741513.1

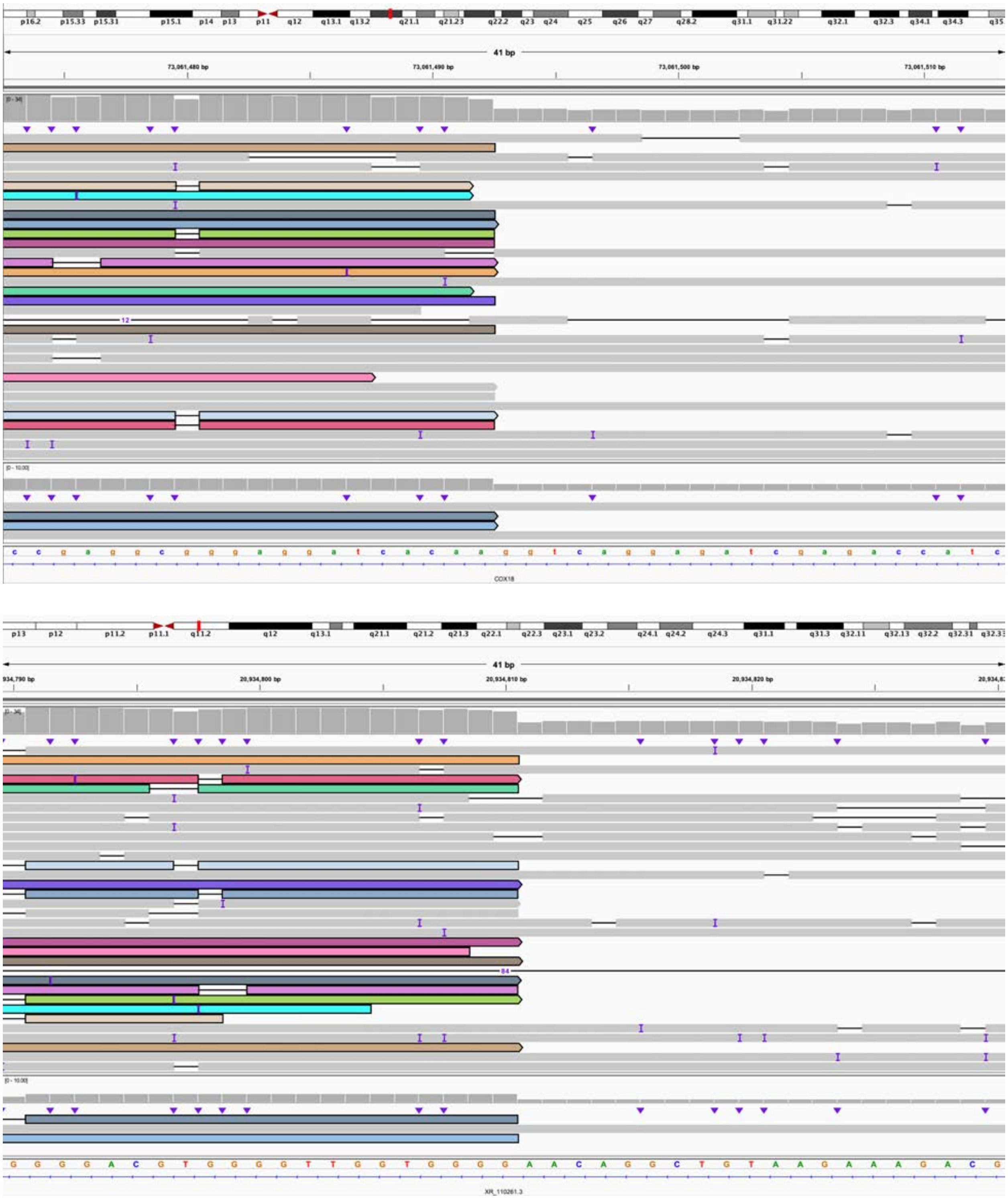
AD. The 5' end of 4:I bisects *BTC* (IGV: chr4:74,794,332-74,794,371) and is connected to the 5' end of 4:H and bisects *XR_001741513.1* (IGV: chr4:74,581,595-74,581,635). This is within derivative chromosome 14.


p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
74,794,340 bp
74,794,350 bp
74,794,360 bp
74,794,370 bp
[0 - 32]
[0 - 10.00]
C G C T C T G C C G G G C C G G G C A G C C C C T A G A C A A G T C T C C C T C C
BTC
40 bp
626,700 bp
79,626,710 bp
79,626,720 bp
79,626,730 bp
[0 - 20]
[0 - 10.00]
A A T C T C A T A T C G C C T G G C T C T T G T C T C A T A A C T C A A T T T A

AE. The 3' end of 4:H bisects *BTC* (IGV: chr4:74,794,332-74,794,372) and is connected to the 3' end of 4:L (IGV: chr4:79,626,700-79,626,739). This is within derivative chromosome 14.

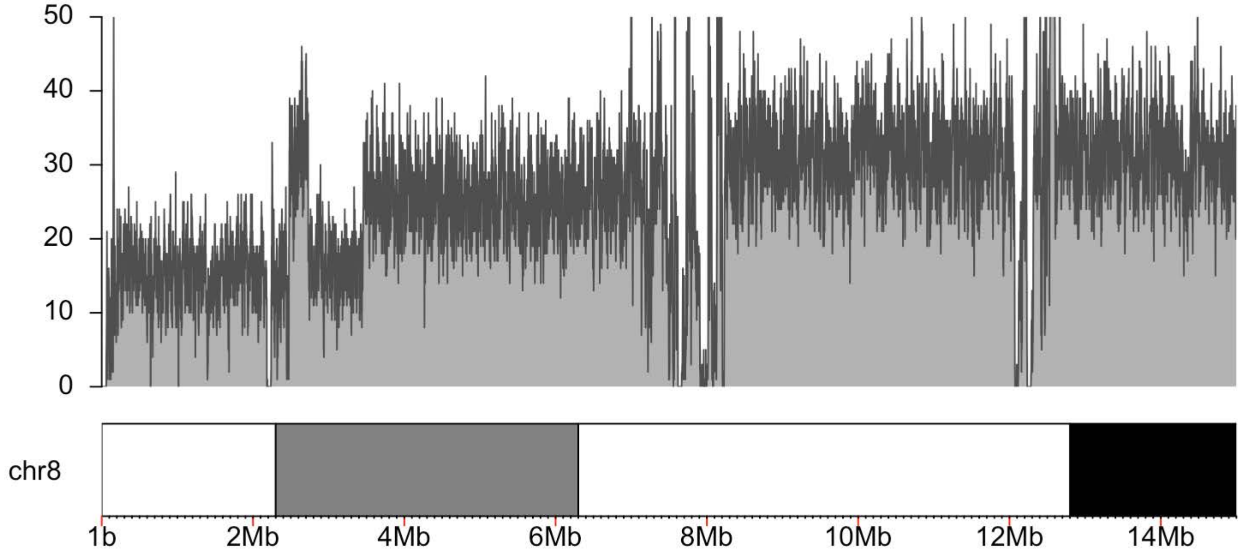

p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
77,012,160 bp
77,012,170 bp
77,012,180 bp
77,012,190 bp
[0 - 49]
[0 - 10.00]
T T T T T A A T T C T T A A A T G G G A G G T T G A A A T A T A A C T T C C A G G
SEPTIN11
p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
140,480 bp
70,140,490 bp
70,140,500 bp
70,140,510 bp
70,140,5
[0 - 38]
[0 - 10.00]
T T T T T A A G T A A A T T A G G T G T T A T G C T A T T G A C A T A T C G A A T G
CSN1S2BP

AF. The 5' end of 4:L bisects *SEPTIN1* (IGV: chr4:77,012,155-77,012,195) and connects to the 5' end of 4:C and bisects *CSN1S2BP* (IGV: chr4:70,140,480-70,140,520). This is within derivative chromosome 14.

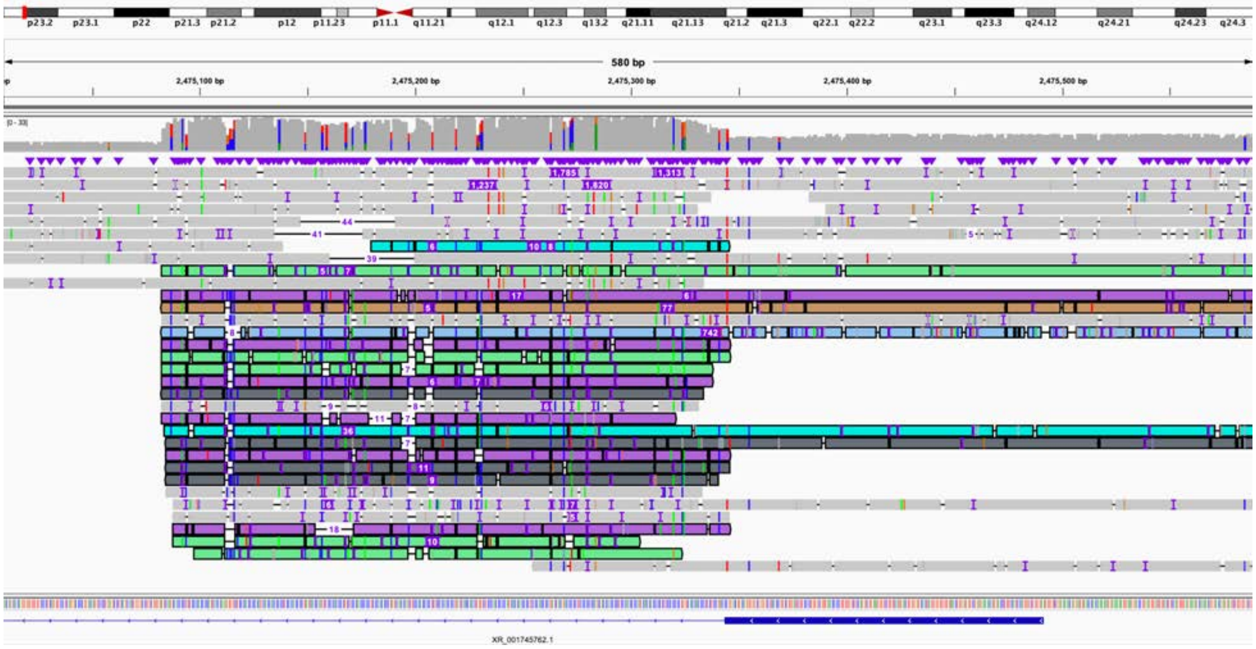

p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
41 bp
73,061,480 bp
73,061,490 bp
73,061,500 bp
73,061,510 bp
[0 - 34]
12
[0 - 10.00]
c c g a g g c g g g a g g a t c a c a a g g t c a g g a g a t c g a g a c c a t c
COX18
p13 p12 p11.2 p11.1 q11.2 q12 q13.1 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q24.1 q24.2 q24.3 q31.1 q31.3 q32.11 q32.13 q32.2 q32.31 q32.33
41 bp
934,790 bp
20,934,800 bp
20,934,810 bp
20,934,820 bp
[0 - 34]
84
[0 - 10.00]
G G G G A C G T G G G G T T G G T G G G G A A C A G G C T G T A A G A A A G A C G
XR_110261.3

AG. The 3' end of 4:C bisects *COX18* (IGV: chr4:73,061,473-73,061,513) and is connected to the 3' end of 14:A and bisects *XR_110261.3* (IGV: chr14:20,934,790-20,934,830). This is the end of derivative chromosome 14.

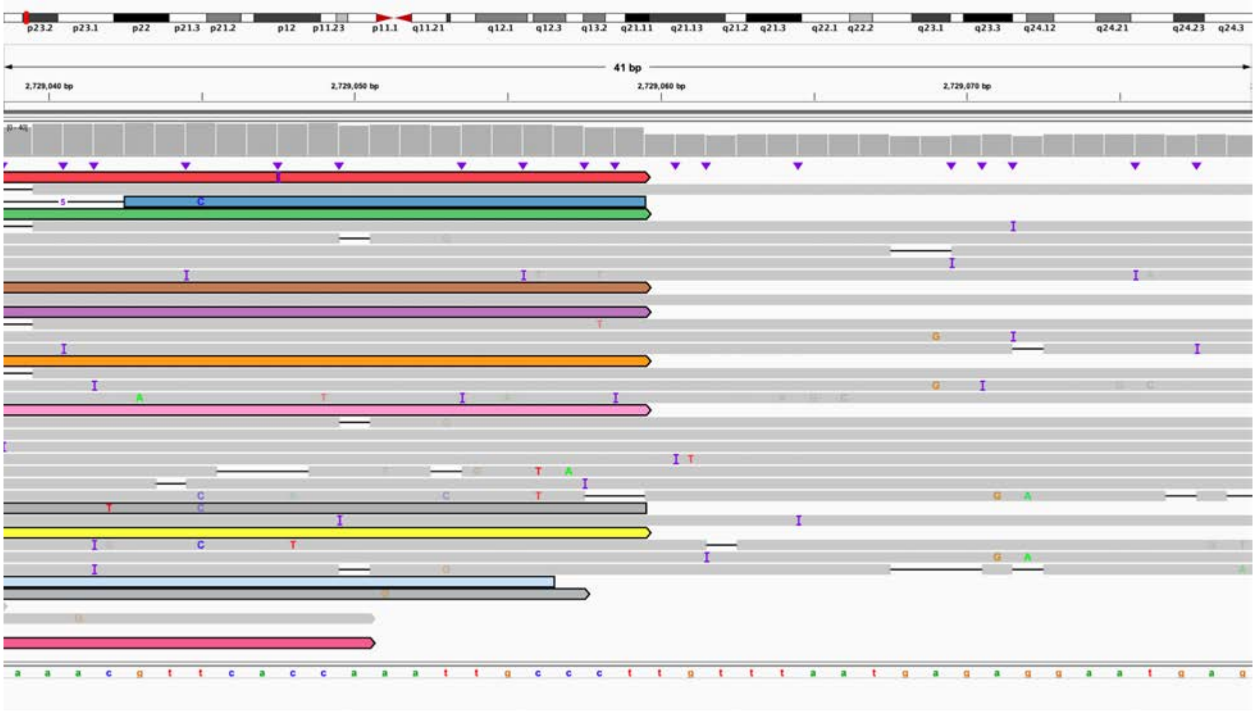
Figure S26. S021, mosaic loss of 8p and mosaic gain of 8q.



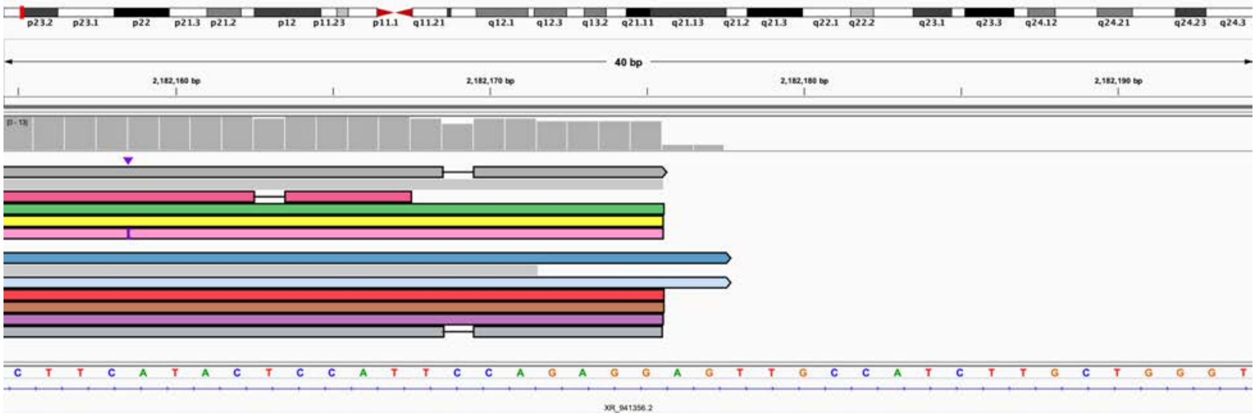
A. Coverage of chr8:1-15,000,000. This demonstrates the presence of a deletion of one chromosome from the telomere to 3,458,035. Also apparent is the mosaic region between 3,469,752 and 7,186,524, then the normal copy state after the defensin locus. The duplication around 2 Mbp is frequently observed in SNP array studies and is often not reported on the array as it is considered benign (see panels B–D for details).


p23.2 p23.1 p22 p21.3 p21.2 p12 p11.23 p11.1 q11.21 q12.1 q12.3 q13.2 q21.11 q21.13 q21.2 q21.3 q22.1 q22.2 q23.1 q23.3 q24.12 q24.21 q24.23 q24.3
580 bp
2,475,100 bp
2,475,200 bp
2,475,300 bp
2,475,400 bp
2,475,500 bp
[0 - 33]
XR_001745762.1

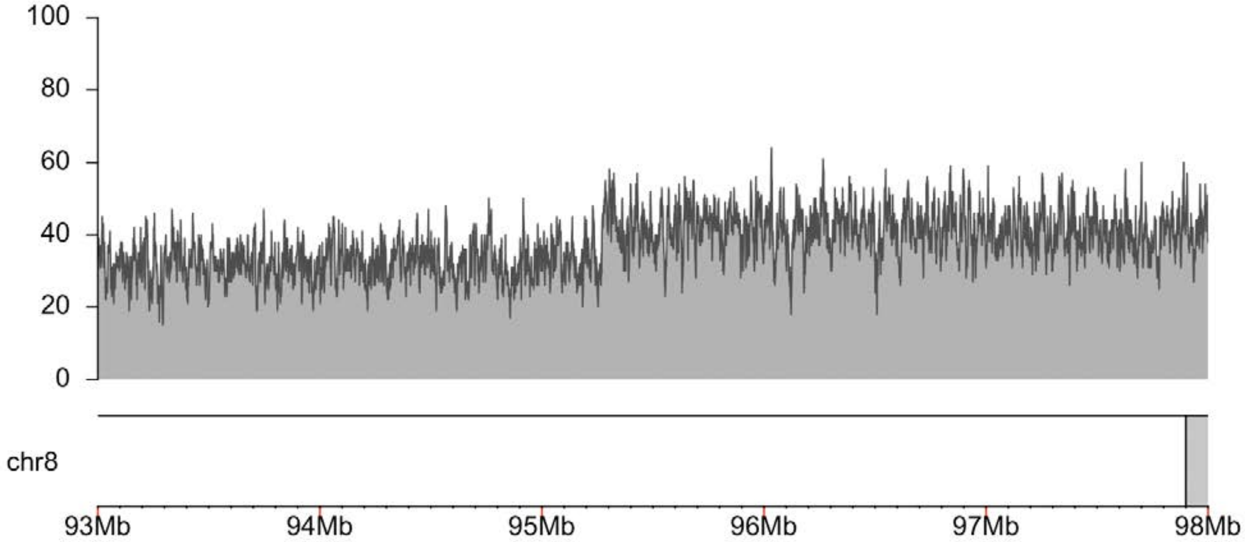
B. The 5' end of the duplication observed in the coverage plot in A begins at approximately chr8:2,475,083, within a VNTR. Read color shows that reads were split and re-aligned to the repetitive region. IGV view is of chr8:2,475,009-2,475,591.


p23.2 p23.1 p22 p21.3 p21.2 p12 p11.23 p11.1 q11.21 q12.1 q12.3 q13.2 q21.11 q21.13 q21.2 q21.3 q22.1 q22.2 q23.1 q23.3 q24.12 q24.21 q24.23 q24.3
41 bp
2,729,040 bp
2,729,050 bp
2,729,060 bp
2,729,070 bp
[0 - 43]
a a a c g t t c a c c a a a t t g c c c t t g t t t a a t g a g a g g a a t g a g

C. The 3' end of the duplication observed in A ends at chr8:2,729,059, meaning the duplication is approximately 254 kbp. IGV view is of chr8:2,729,039-2,729,079. Reads are linked to those shown in panel D and color is consistent between the two.

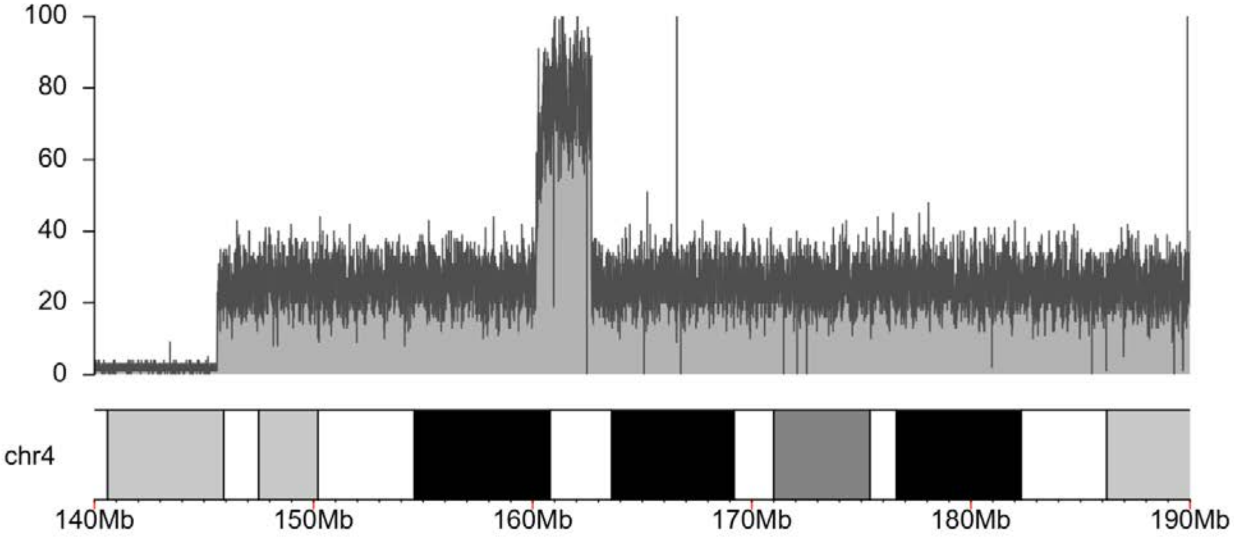

p23.2 p23.1 p22 p21.3 p21.2 p12 p11.23 p11.1 q11.21 q12.1 q12.3 q13.2 q21.11 q21.13 q21.2 q21.3 q22.1 q22.2 q23.1 q23.3 q24.12 q24.21 q24.23 q24.3
40 bp
2,182,160 bp
2,182,170 bp
2,182,180 bp
2,182,190 bp
[0 - 13]
C T T C A T A C T C C A T T C C A G A G G A G T T G C C A T C T T G C T G G G T
XR_941356.2

D. Reads from the 3' end of the duplication shown in C link to the beginning of a deletion on chr8 at chr8:2,182,175. The orientation of the reads suggests that the duplication is inverted. IGV view is of chr8:2,182,155-2,182,194.

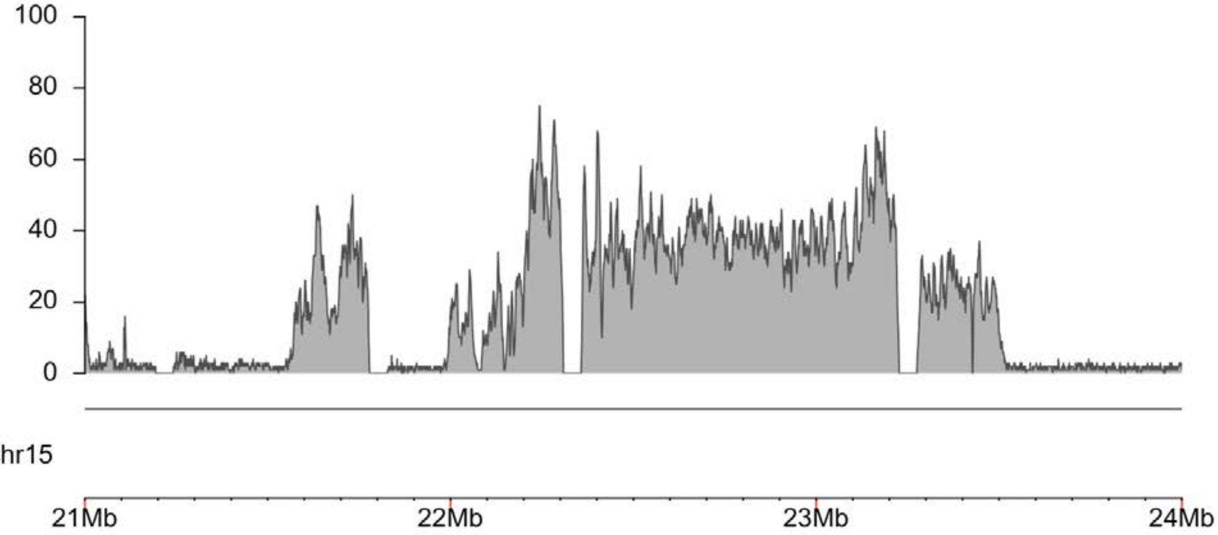


E: Coverage of chr8:93,000,000-98,000,000 demonstrating the increase of copy state from 2 to 3 observed on array at 95,270,423. No definitive breakpoint could be identified in the long-read data.

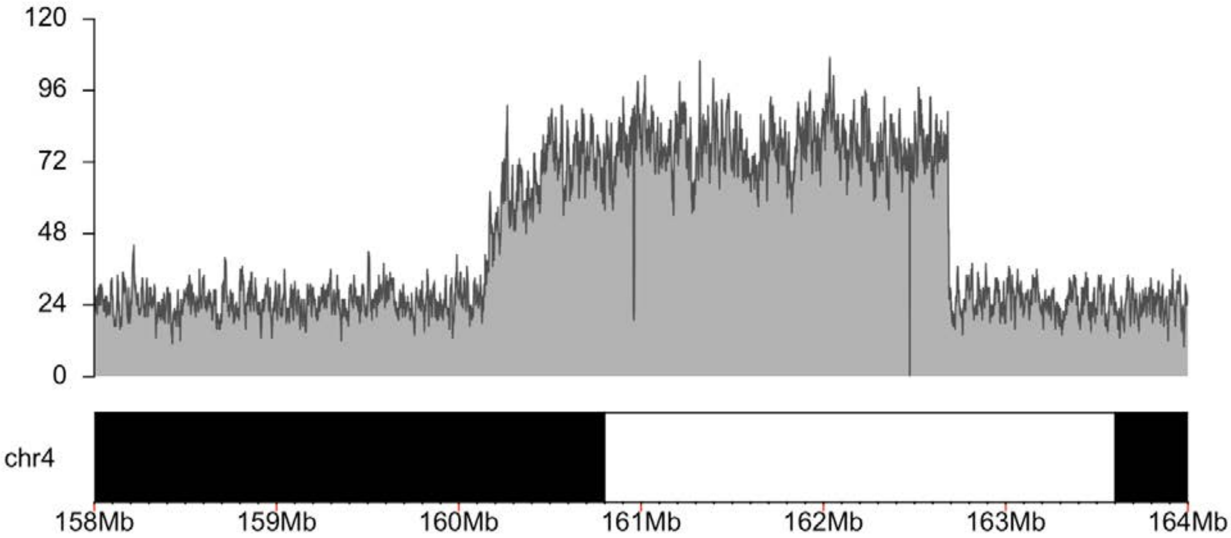
Figure S27. S022, focal amplification of 4q with adjacent region of homozygosity, duplication of 15q11.2.



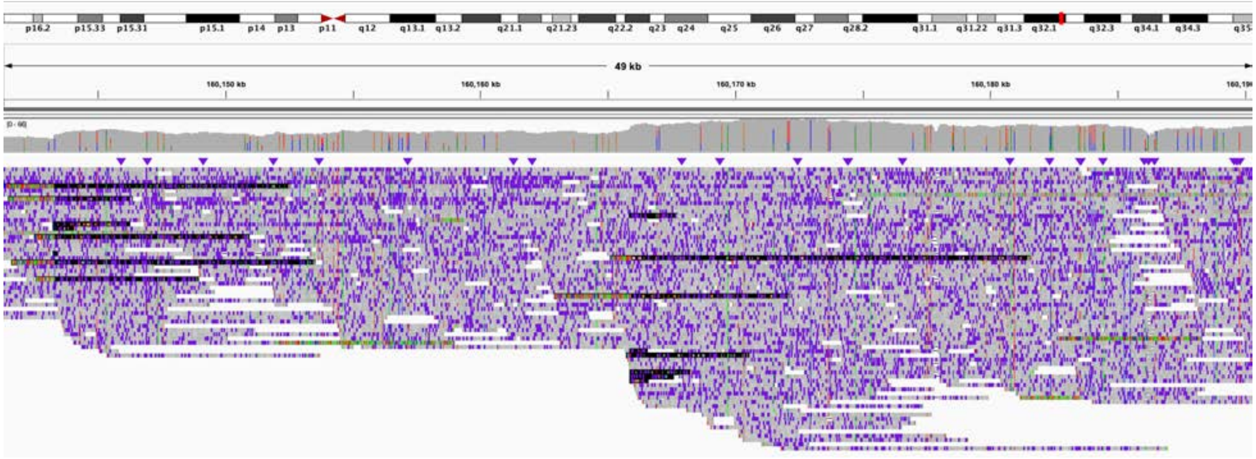
A. Coverage of the 4q target region shows the presence of an amplification.



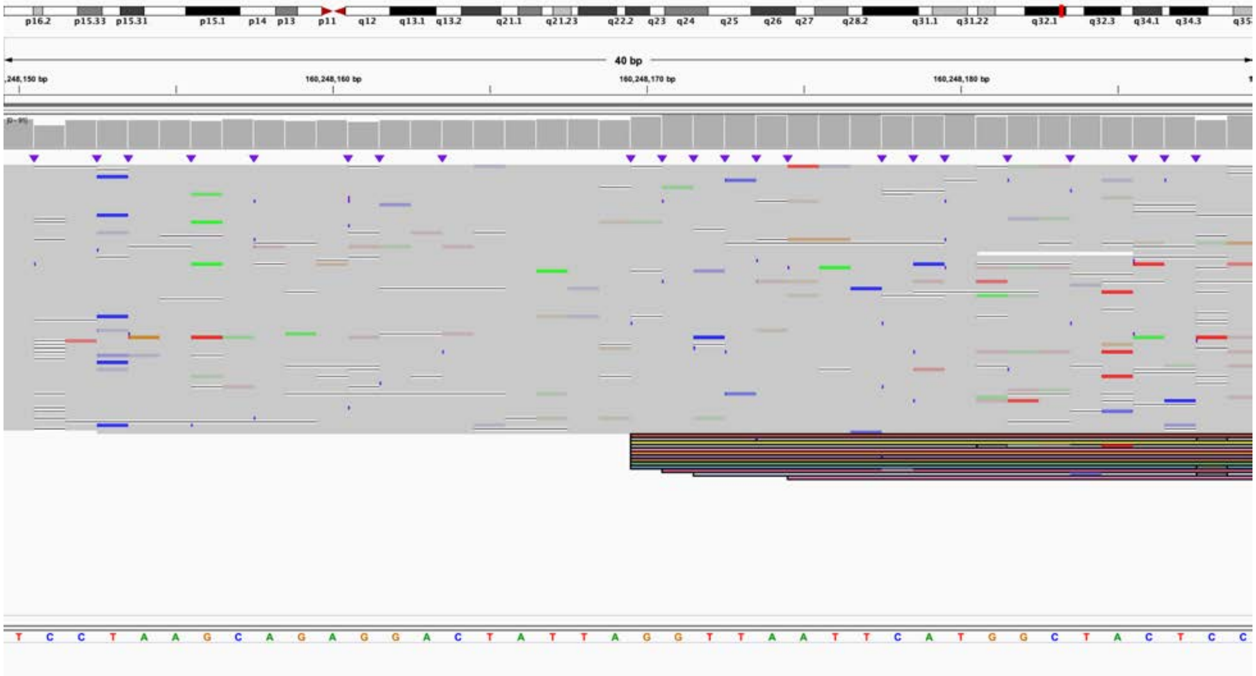
B. Coverage of 15q11.2 target region, regions with no coverage represent regions of the reference genome denoted as ‘N’.



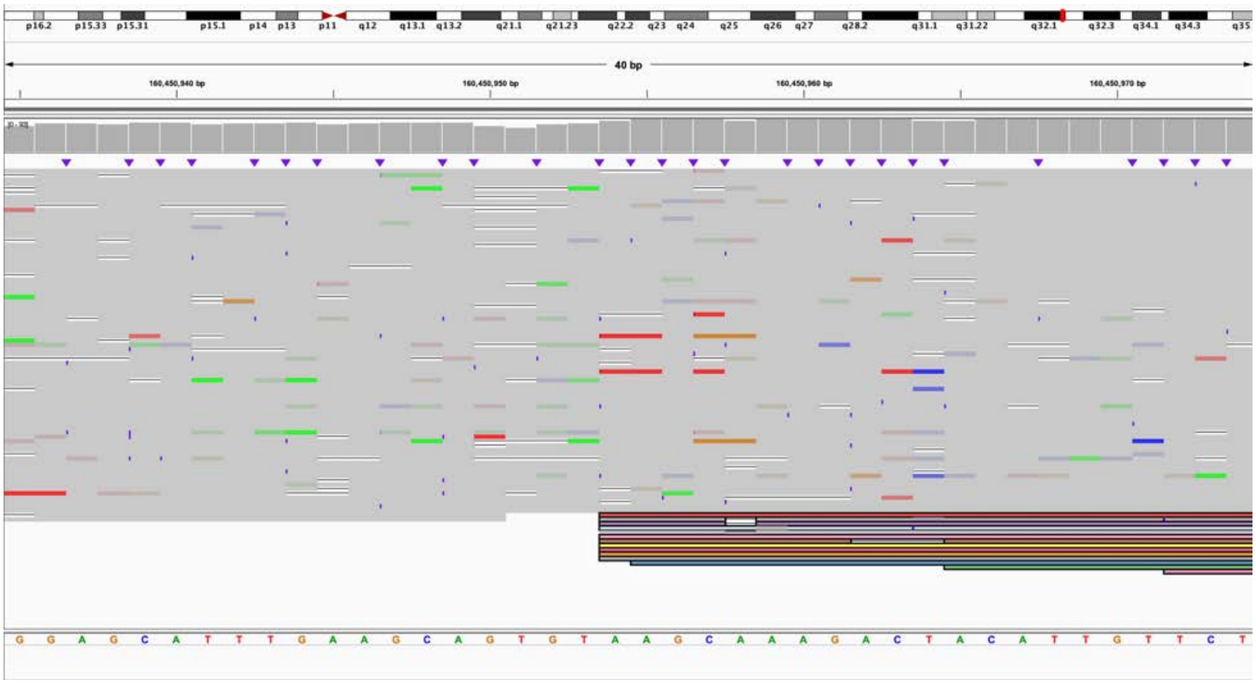
C. Coverage of the focal amplification of 4q.



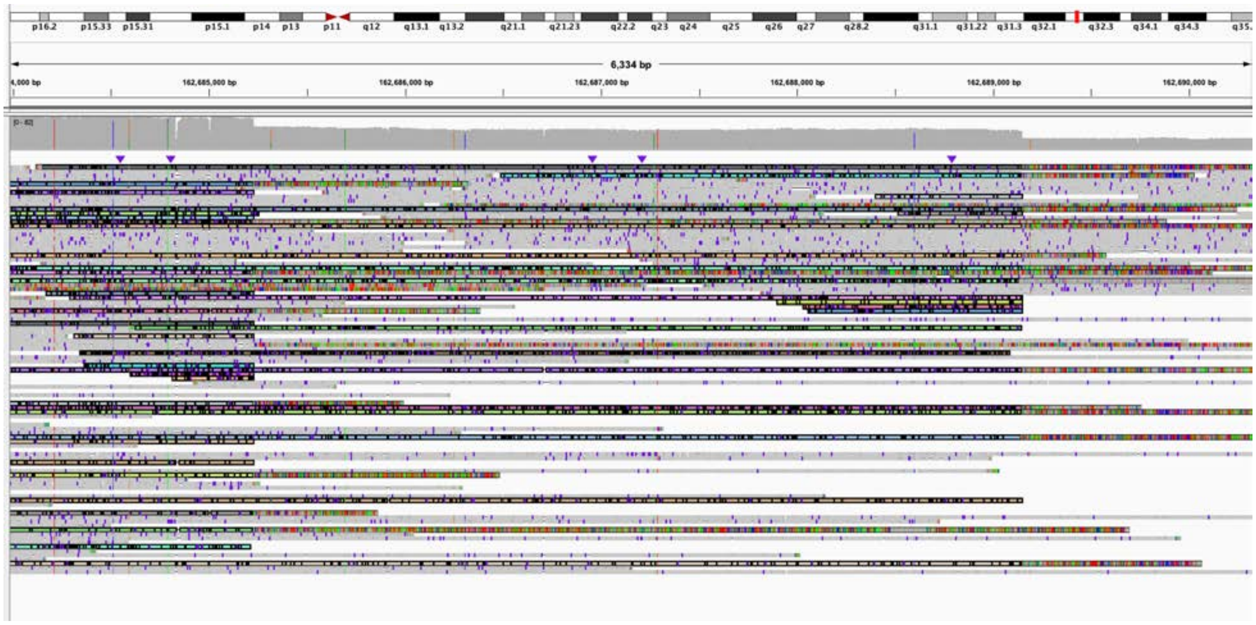
D. View of the beginning of the focal amplification of 4q, this is the centromere-proximal side and includes the second increase in coverage. Within this region, copy number estimate increases from 2x to 4x in a stepwise fashion (Table S9). Reads that span both breakpoints are highlighted. IGV view is of chr4:160,141,260-160,190,550.


p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
40 bp
,248,150 bp
160,248,160 bp
160,248,170 bp
160,248,180 bp
T C C T A A G C A G A G G A C T A T T A G G T T A A T T C A T G G C T A C T C C

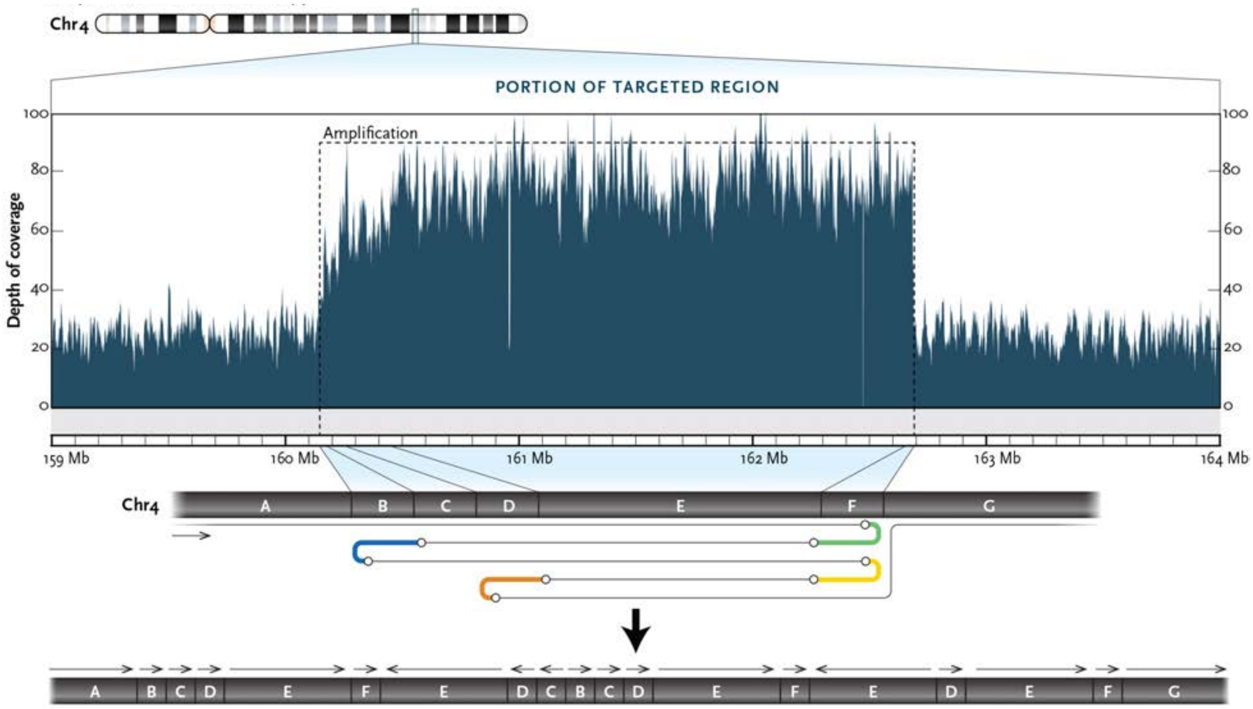
E. View of the third centromere-proximal breakpoint and amplification. Copy number estimate in this region increases from 4x to 5x (Table S9). IGV view is of chr4:160,248,150-160,248,189.


p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q21.1 q21.23 q22.2 q23 q24 q25 q26 q27 q28.2 q31.1 q31.22 q32.1 q32.3 q34.1 q34.3 q35
40 bp
160,450,940 bp
160,450,950 bp
160,450,960 bp
160,450,970 bp
G G A G C A T T T G A A G C A G T G T A A G C A A A G A C T A C A T T G T T C T

F. View of the fourth centromere-proximal breakpoint and amplification. Copy number estimate increases from 5x to 6x in this region (Table S9). IGV view is of chr4:160,450,935-160,450,974.

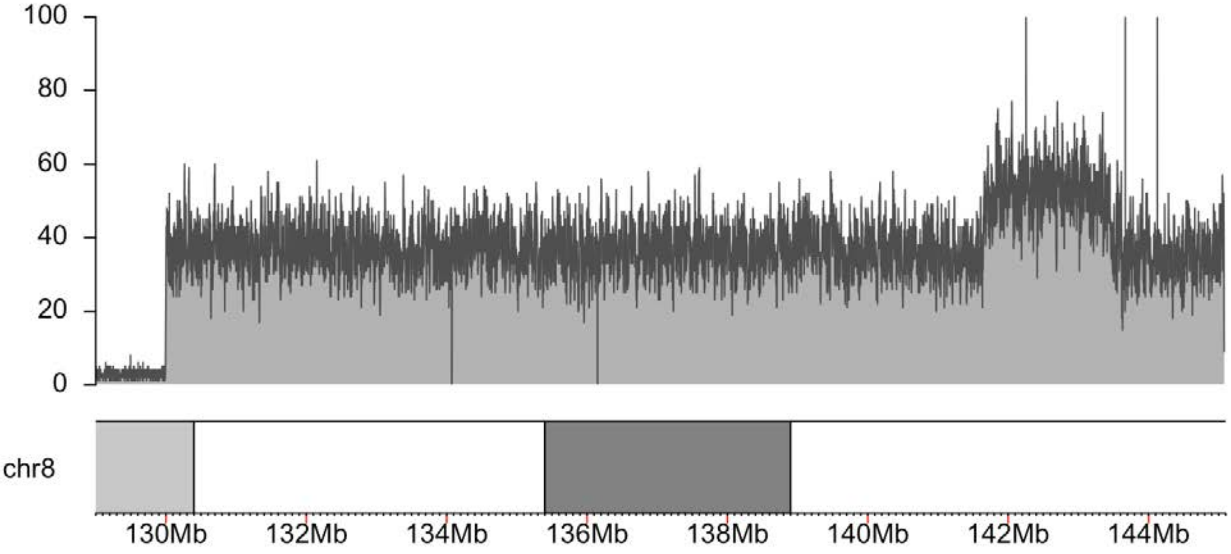


G. View of the telomere-proximal region of the focal amplification of 4q. Reads that span both breakpoints are highlighted. Copy number estimate decreases from 6x to 4x, then from 4x to 2x over a 3,920 bp interval (Table S9). IGV view is of chr4:162,683,982-162,690,349.

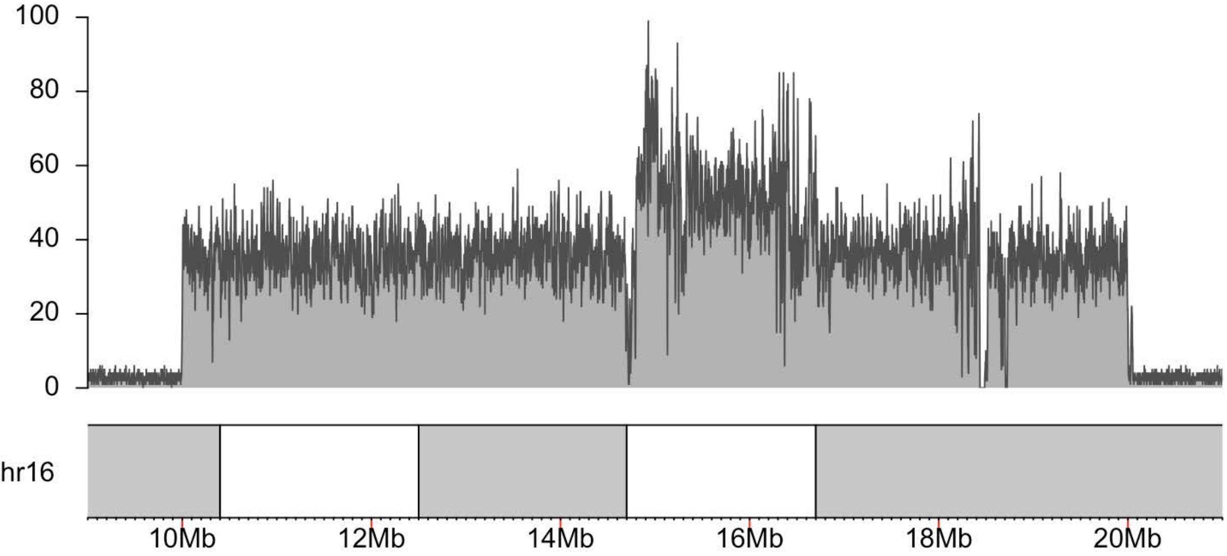


H. Estimate of the possible structure of the 4q amplification. We were not able to determine the exact order in which segments were duplicated. A possible arrangement of the region is shown based on this estimate.

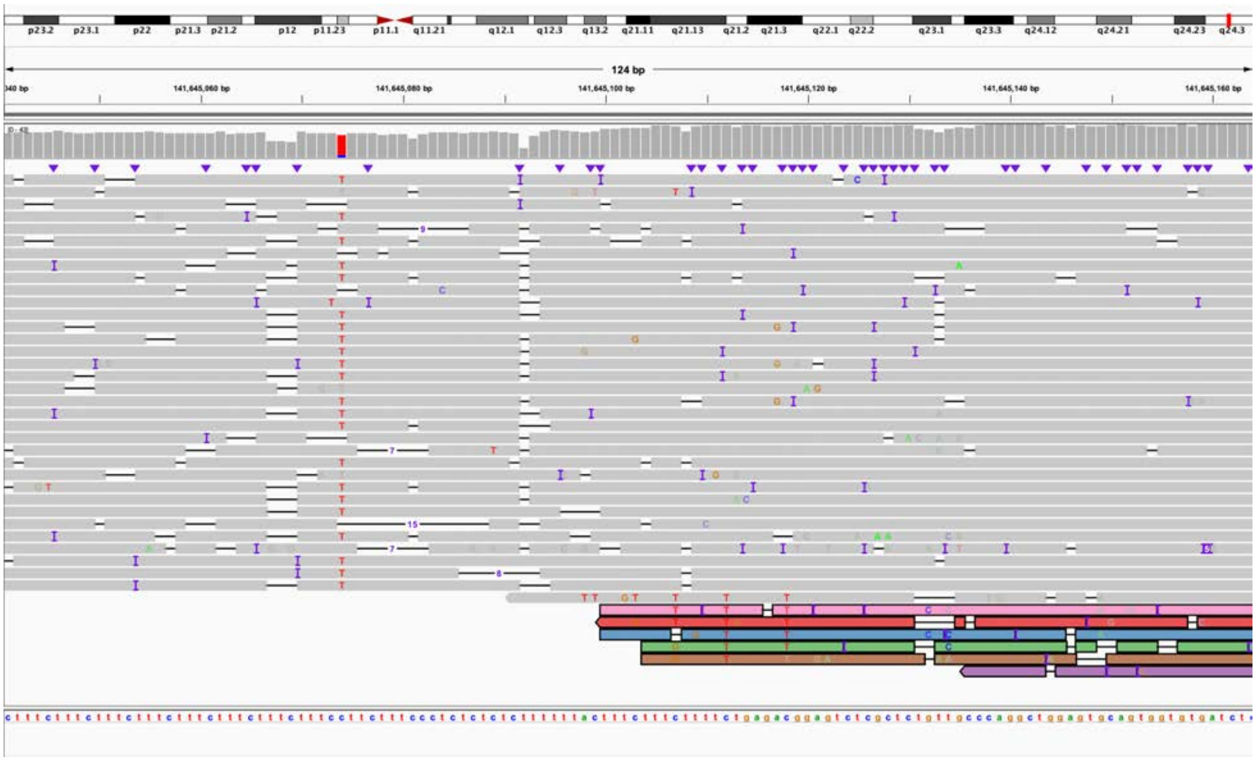
Figure S28. S035, duplications of 8q24 and 16p13.11 identified by clinical testing.



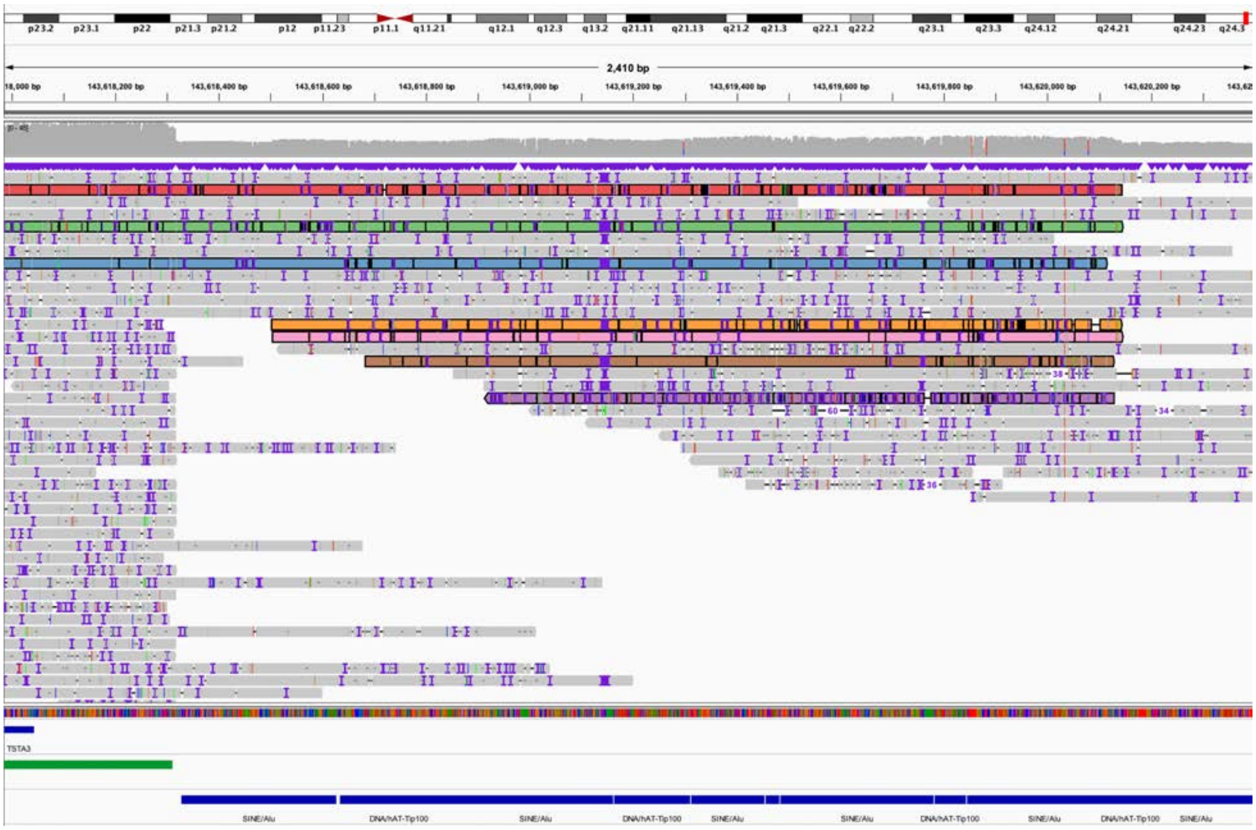
A. Coverage of chromosome 8 target region.



B. Coverage of chromosome 16 target region.

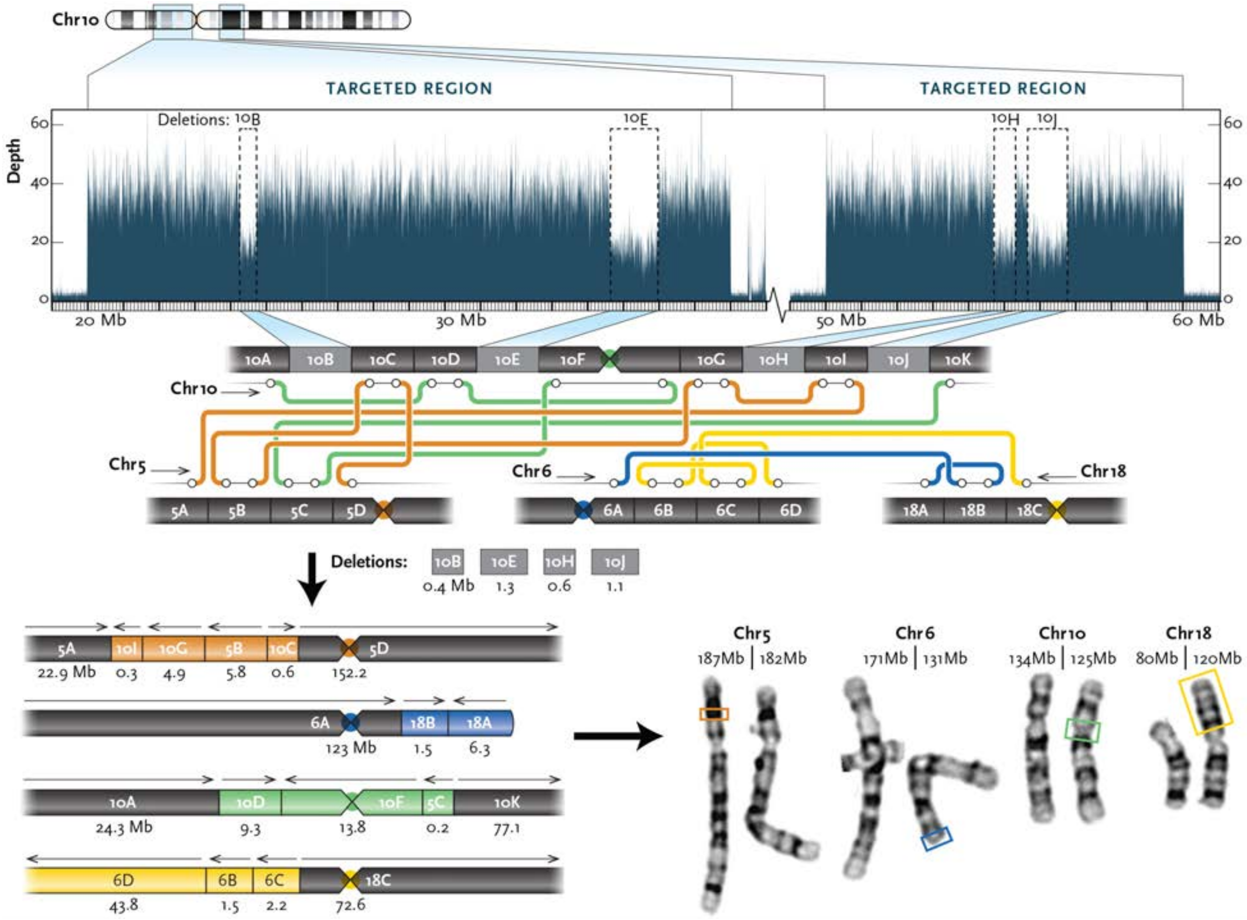


C. Centromere-proximal end of the chromosome 8 duplication, reads spanning the duplication are indicated by color. IGV coordinates are chr8:141,645,041-141,645,164.

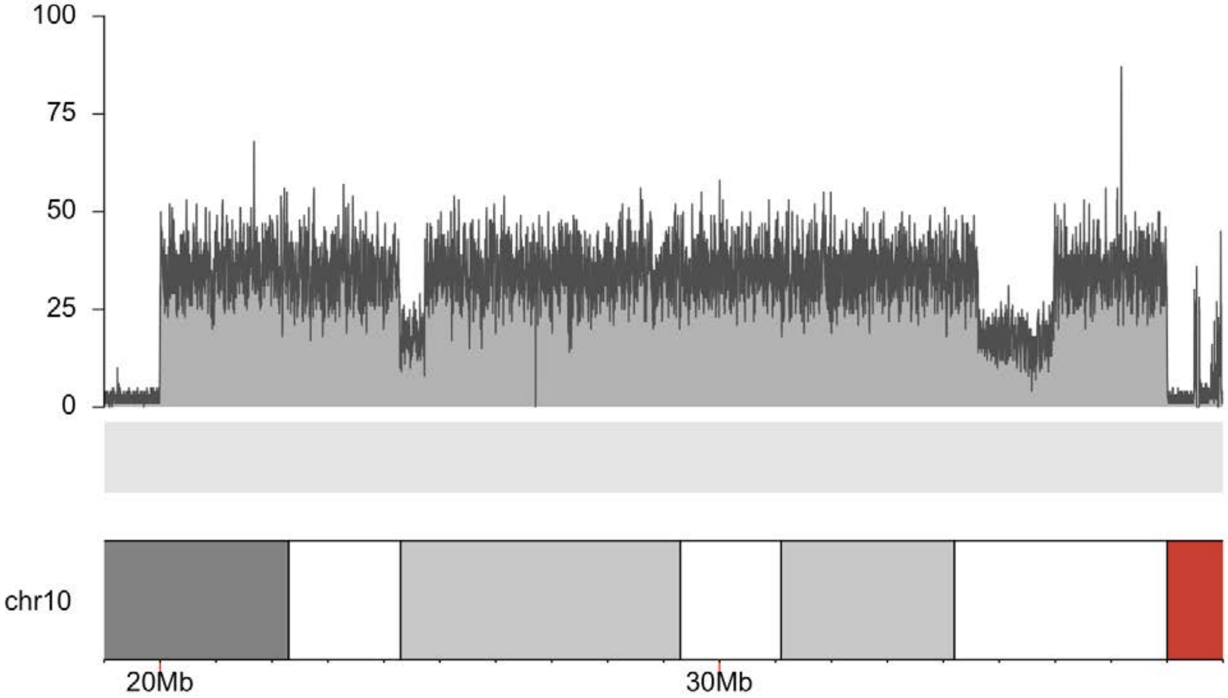


D. Telomere-proximal end of the chromosome 8 duplication shows that the duplication occurred within a TE-dense region. Reads on the 5' end of the image span an approximately 8 kbp deletion. Reads colored as in C. IGV coordinates are chr8:143,617,985-143,620,407.

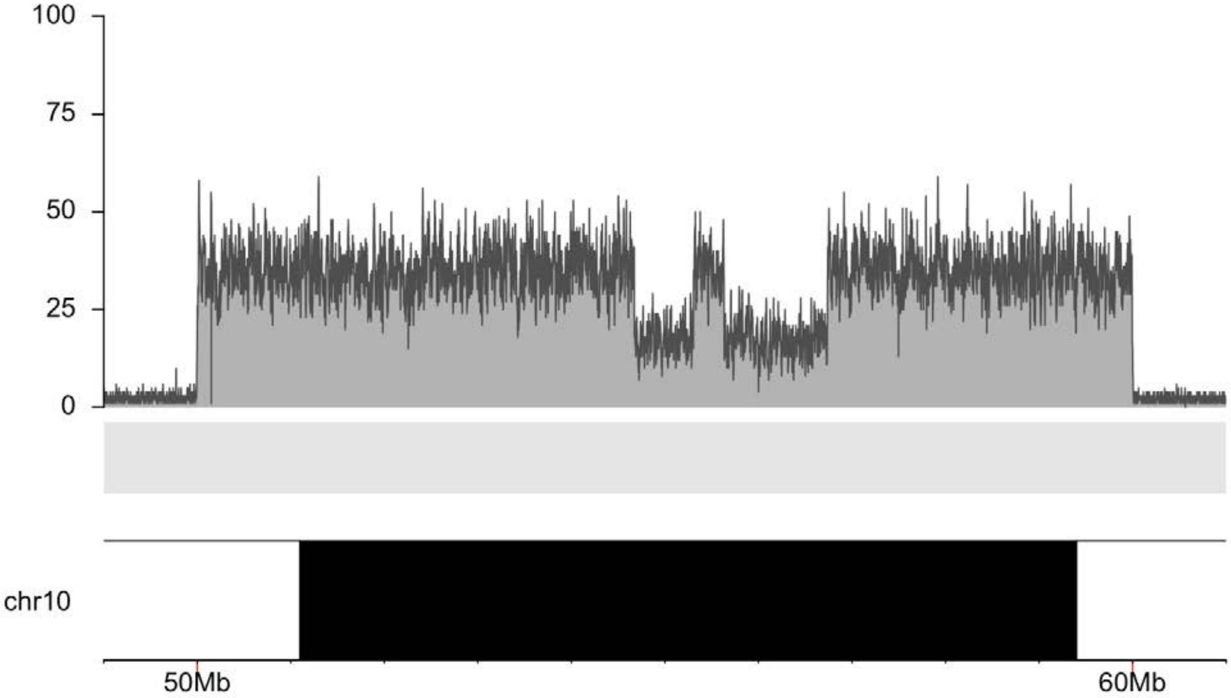
Figure S29. S036, individual with multiple rearrangements and translocations of chromosomes 5, 6, 10, and 18.



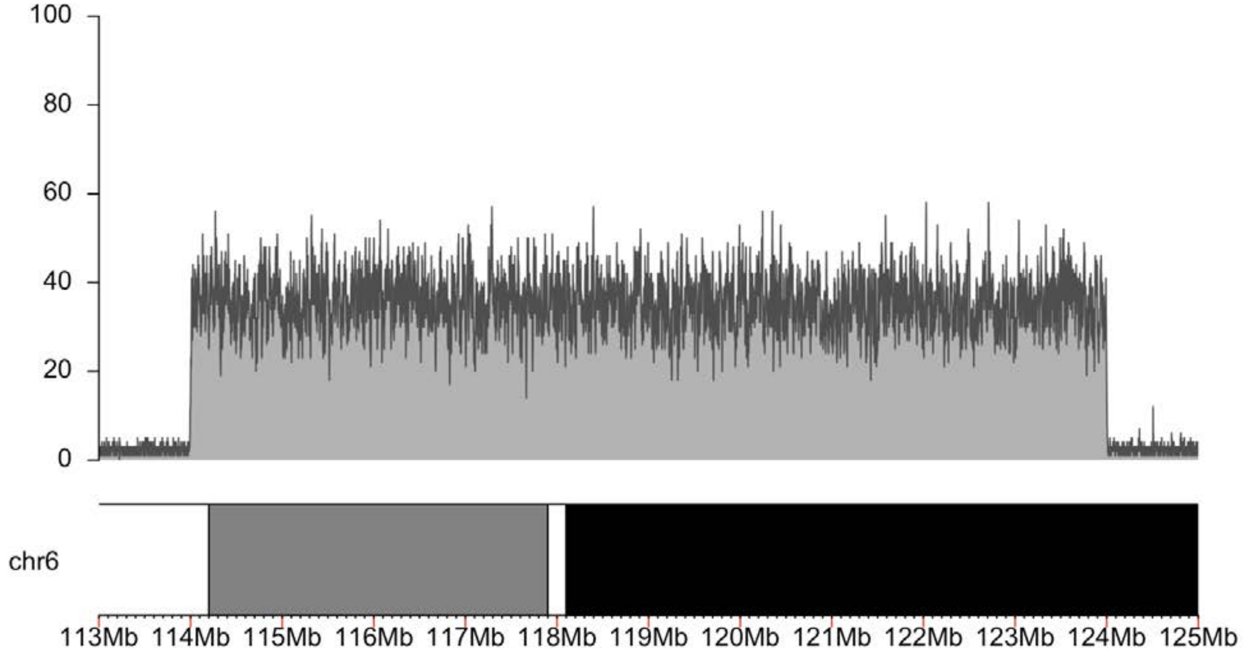
A. In a child with multiple deletions of chromosome 10, karyotyping revealed translocations between chromosomes 6 and 18 along with a pericentric inversion of chromosome 10. LRS identified additional translocations involving chromosomes 10 and 5 as well as several additional intrachromosomal rearrangements. Rearrangements can be resolved by following the subway plot for each derivative chromosome. Estimates of the reconstructed chromosome sizes are shown along with where they are estimated to correspond to on the karyotype (colored boxes).



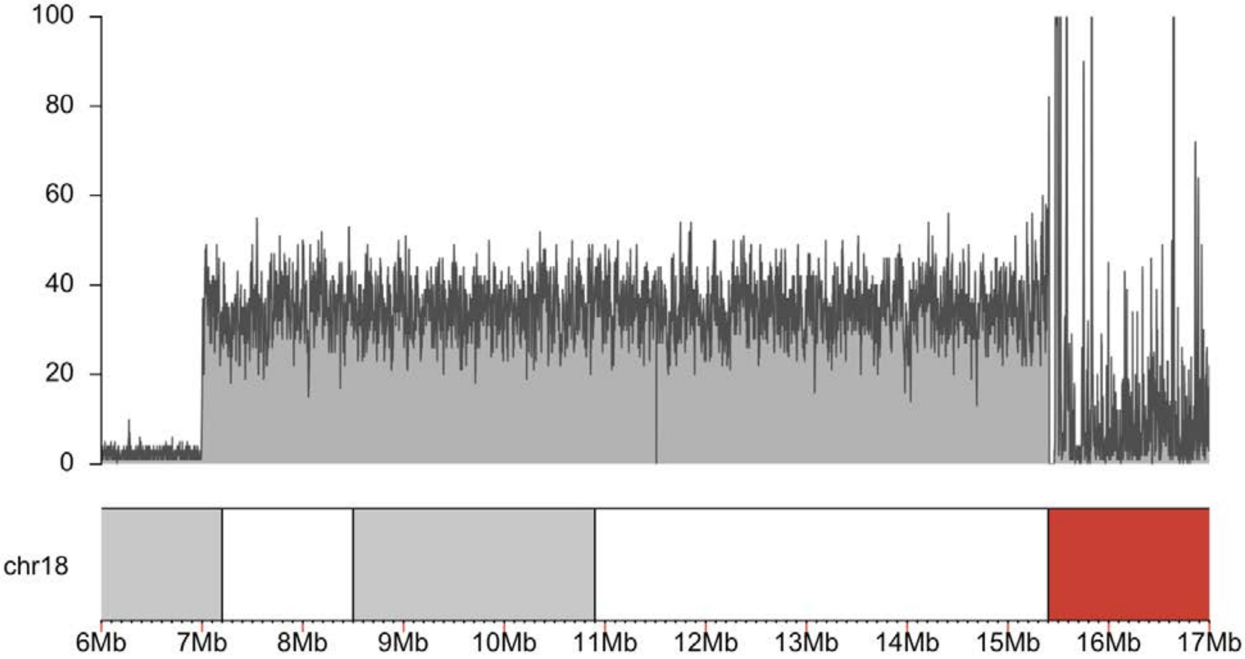
B. Coverage of chromosome 10p target region from the first readfish run.



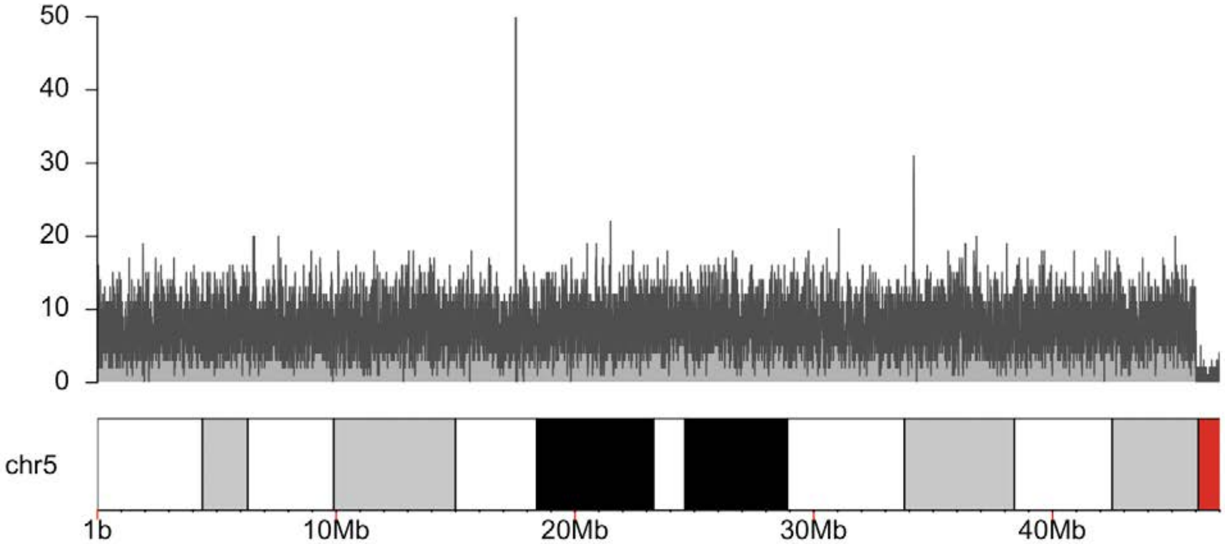
C. Coverage of chromosome 10q target region from the first readfish run.



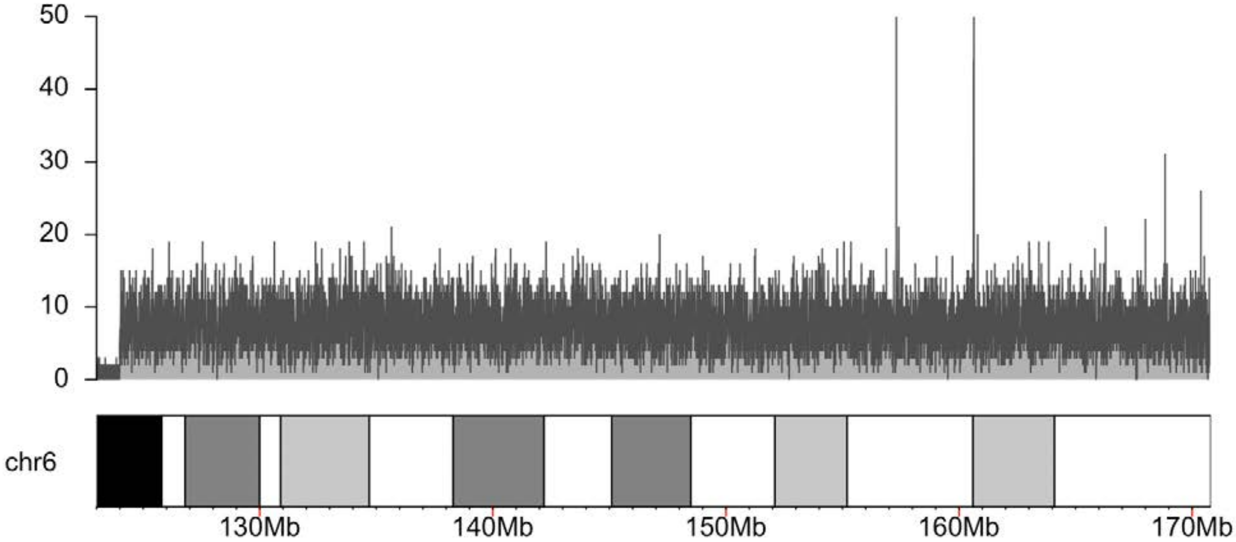
D. Coverage of chromosome 6 target region from the first readfish run.



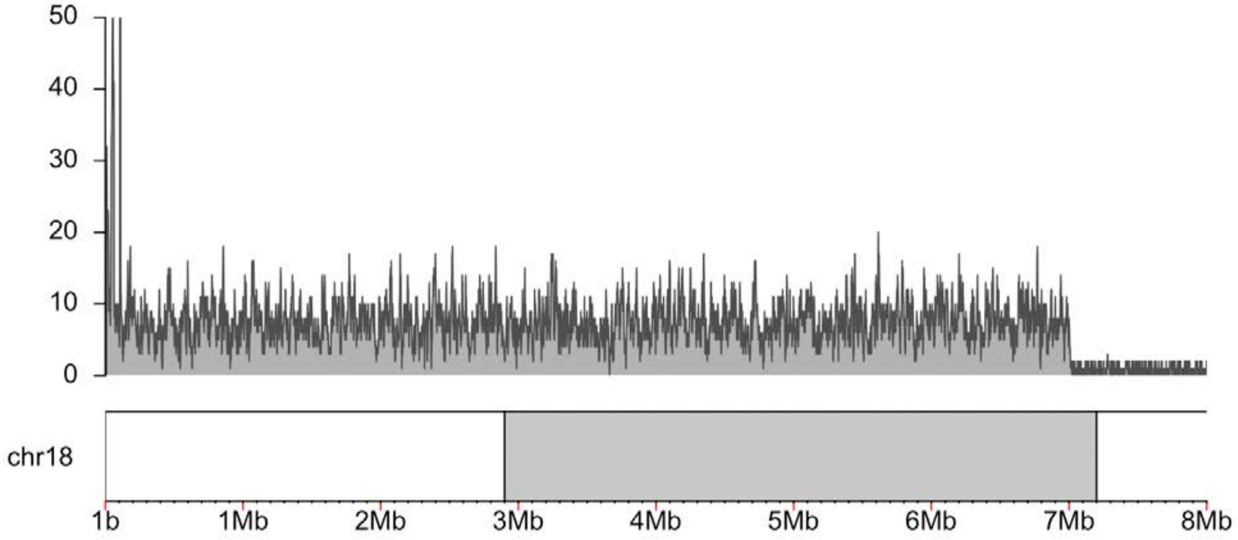
E. Coverage of chromosome 18 target region from the first readfish run.



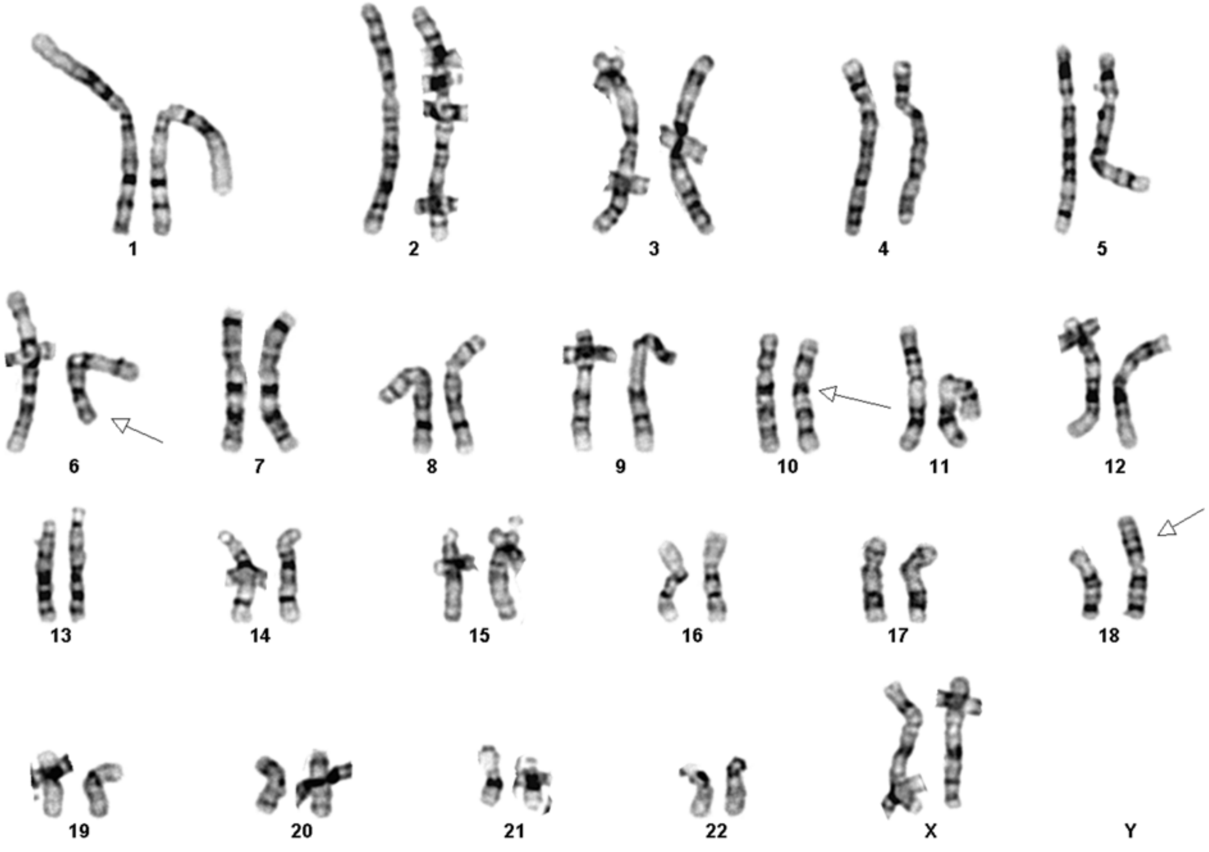
F. Coverage of chromosome 5 target region from the second readfish run.



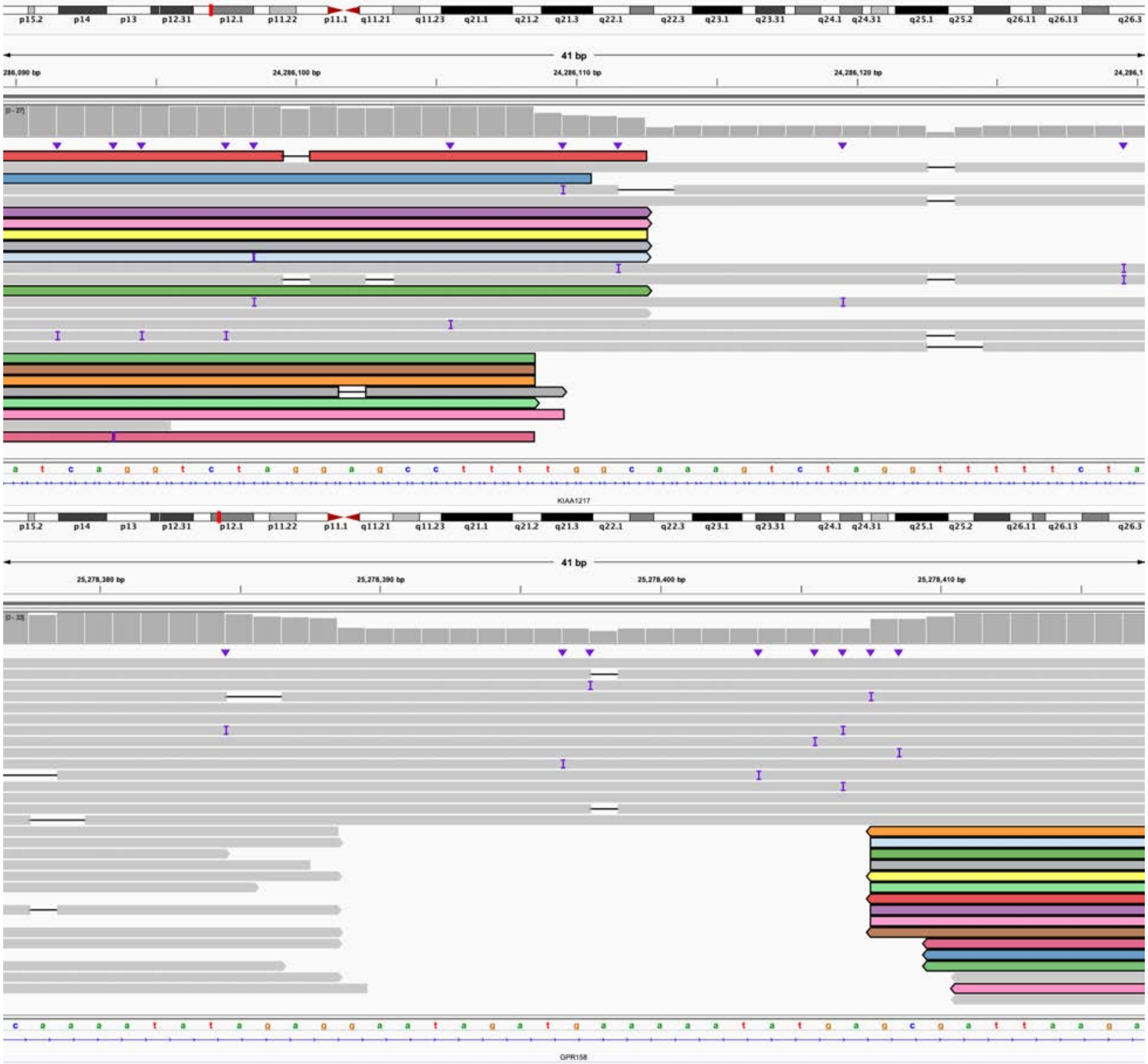
G. Coverage of chromosome 6 target region from the second read fish run.



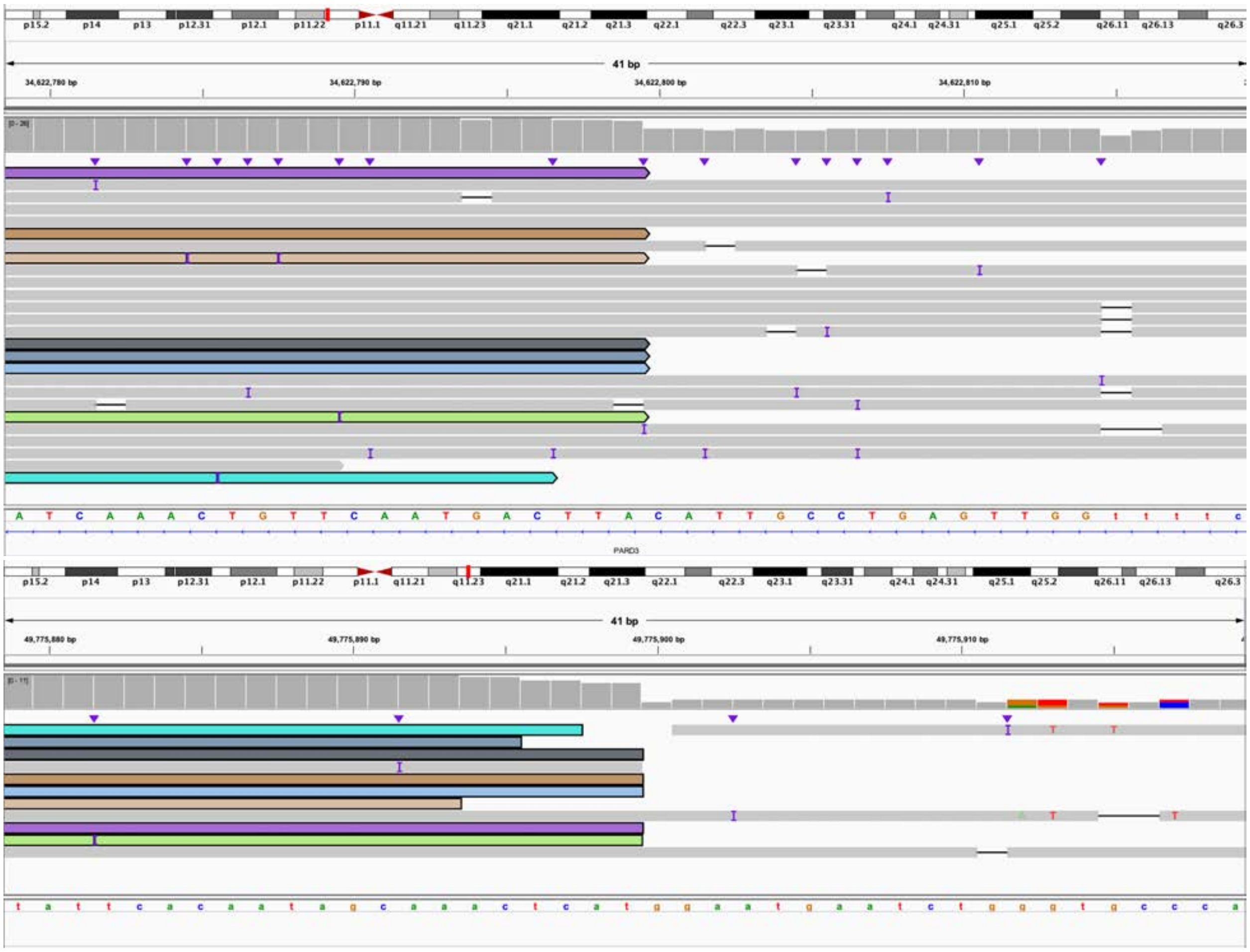
H. Coverage of chromosome 18 target region from the second readfish run.



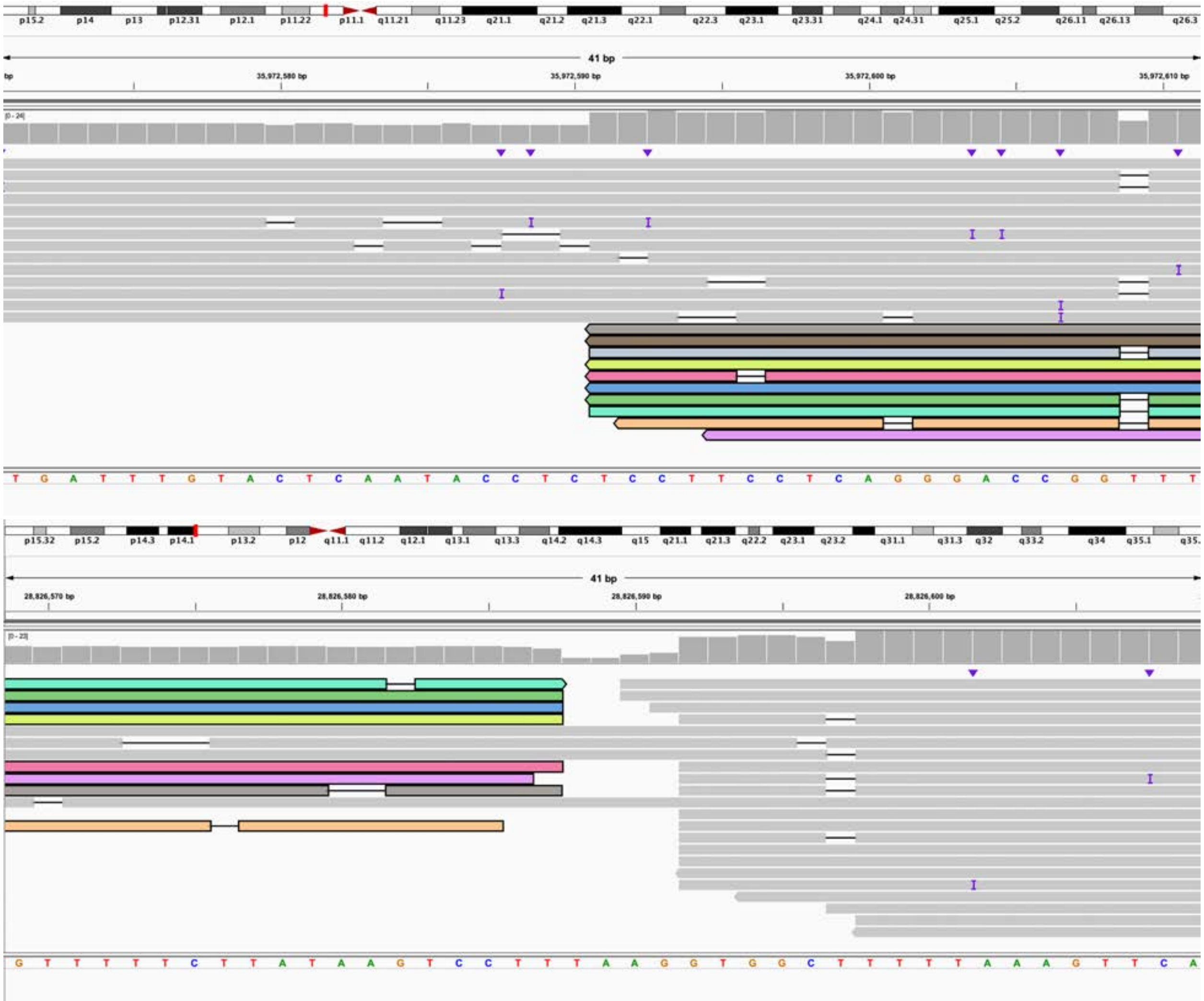
I. Karyotype of the individual, arrows denote derivative chromosomes.


p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
286,090 bp
24,286,100 bp
24,286,110 bp
24,286,120 bp
24,286,1
[0 - 27]
a t c a g g t c t a g g a g c c t t t t g g c a a a g t c t a g g t t t t t c t a
KIAA1217
p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
25,278,380 bp
25,278,390 bp
25,278,400 bp
25,278,410 bp
[0 - 33]
c a a a a t a t a g a g g a a t a g a t g a a a a a t a t g a g c g a t t a a g a
GPR158

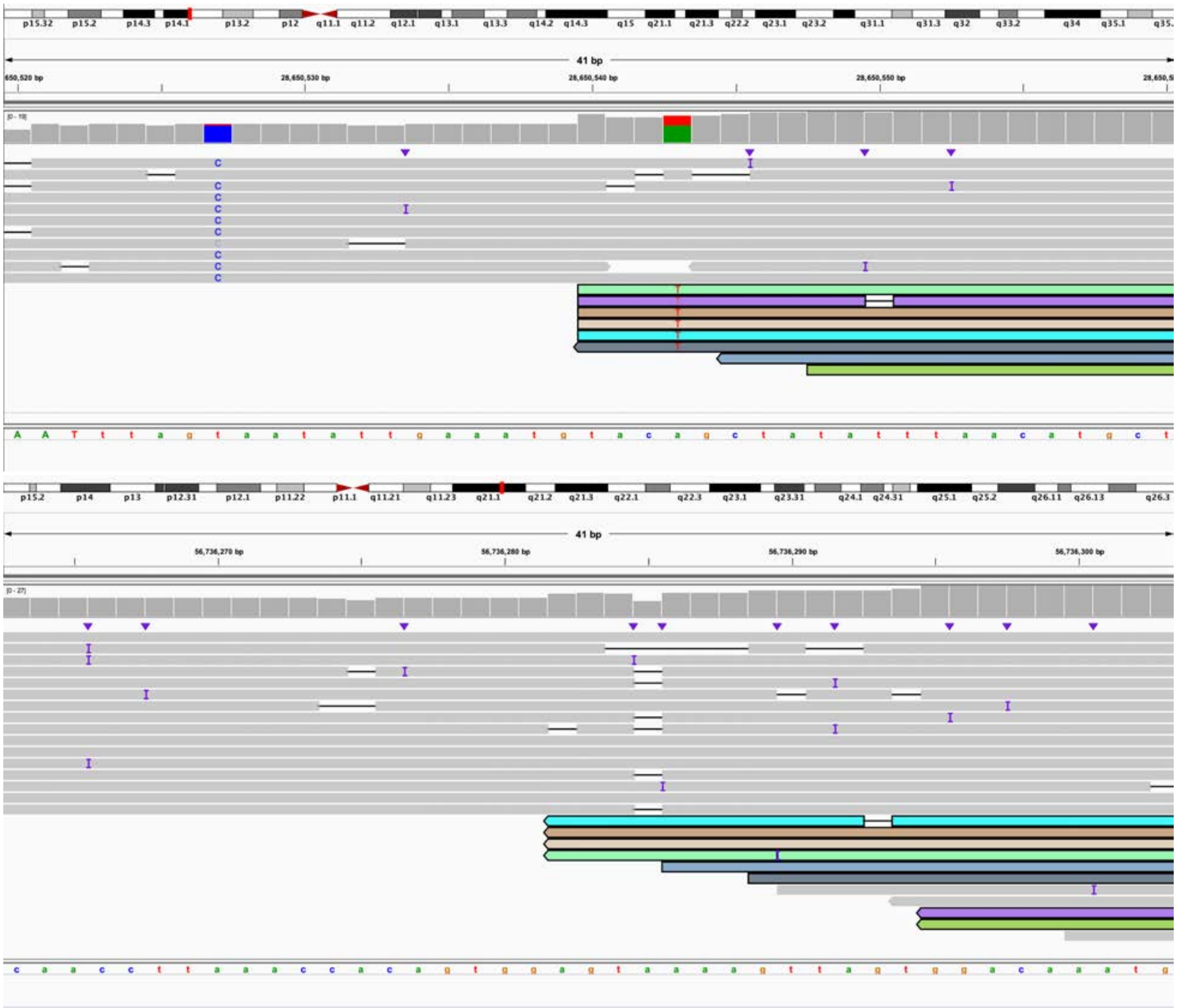
J: 3' end of 10:A bisects *KIAA1217* (IGV: chr10:24,286,090-24,286,130) and is linked to the 5' end of 10:D which bisects *GPR158* (IGV: chr10:25,278,377-25,278,417). This is the beginning of derivative chromosome 10.


p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
34,622,780 bp
34,622,790 bp
34,622,800 bp
34,622,810 bp
A T C A A A C T G T T C A A T G A C T T A C A T T G C C T G A G T T G G t t t t c
PARD3
p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
49,775,880 bp
49,775,890 bp
49,775,900 bp
49,775,910 bp
t a t t c a c a a t a g c a a a a c t c a t g g a a t g a a t c t g g g t g c c c a

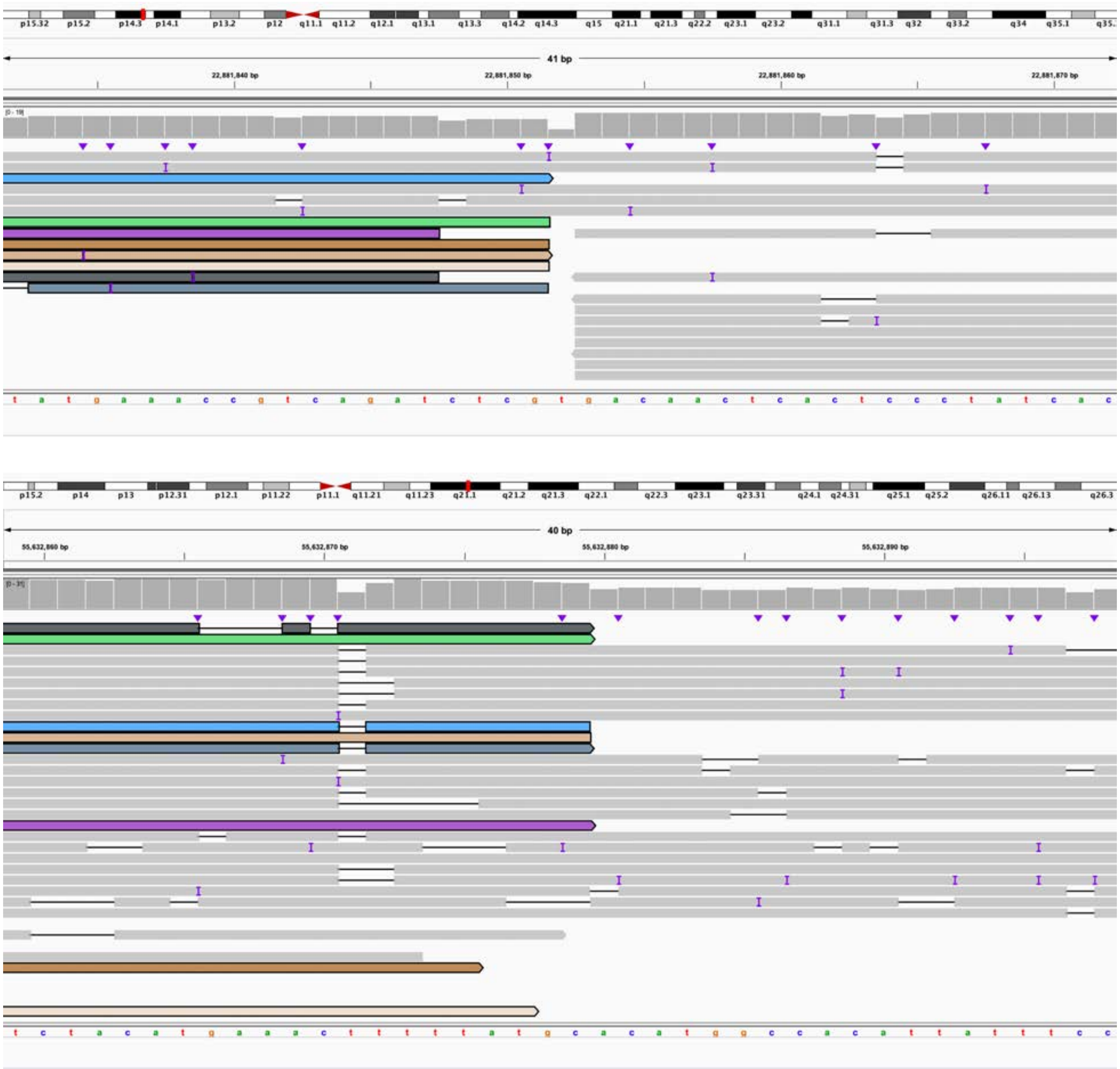
K: 3' end of 10:D bisects *PARD3* (IGV: chr10:34,622,779-34,622,819) and is linked to the 3' end of 10:F (IGV: chr10:49,775,879-49,775,919). This is within derivative chromosome 10.


p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
35,972,580 bp
35,972,590 bp
35,972,600 bp
35,972,610 bp
[0 - 24]
T G A T T T G T A C T C A A T A C C T C T C C T T C C T C A G G G A C C G G T T T
p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q11.2 q12.1 q13.1 q13.3 q14.2 q14.3 q15 q21.1 q21.3 q22.2 q23.1 q23.2 q31.1 q31.3 q32 q33.2 q34 q35.1 q35.
41 bp
28,826,570 bp
28,826,580 bp
28,826,590 bp
28,826,600 bp
[0 - 23]
G T T T T T C T T A T A A G T C C T T T A A G G T G G C T T T T T A A A G T T C A

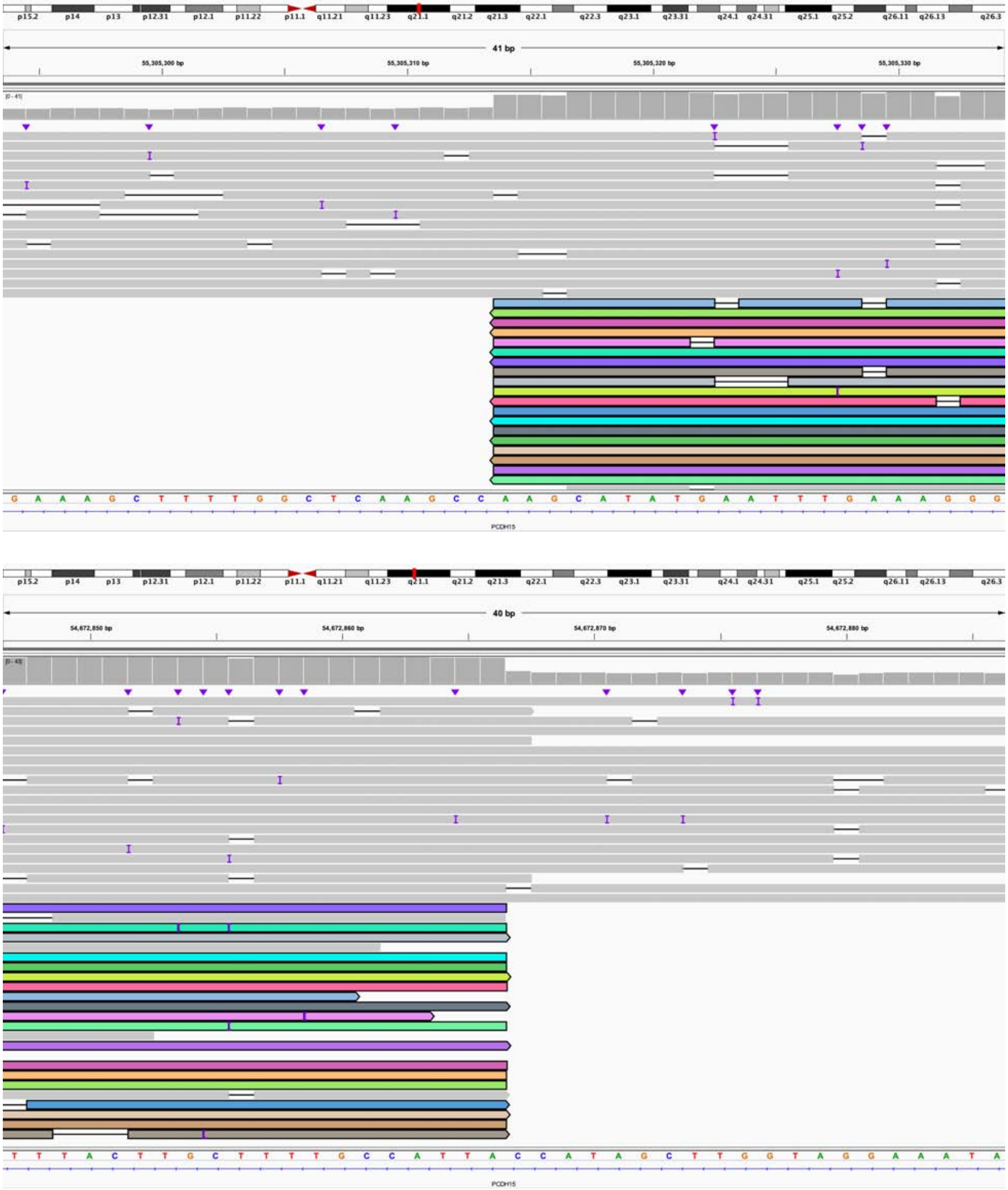
L: 5' end of 10:F (IGV: chr10:35,972,571-35,972,611) is linked to the 3' end of 5:C (IGV: chr5:28,826,569-28,826,609). This is within derivative chromosome 10.


41 bp
650,520 bp
28,650,530 bp
28,650,540 bp
28,650,550 bp
A A T t t a g t a a t a t t g a a a t g t a c a g c t a t a t t t a a c a t g c t
41 bp
56,736,270 bp
56,736,280 bp
56,736,290 bp
56,736,300 bp
c a a c c t t a a a c c a c a g t g g a g t a a a a g t t a g t g g a c a a a t g

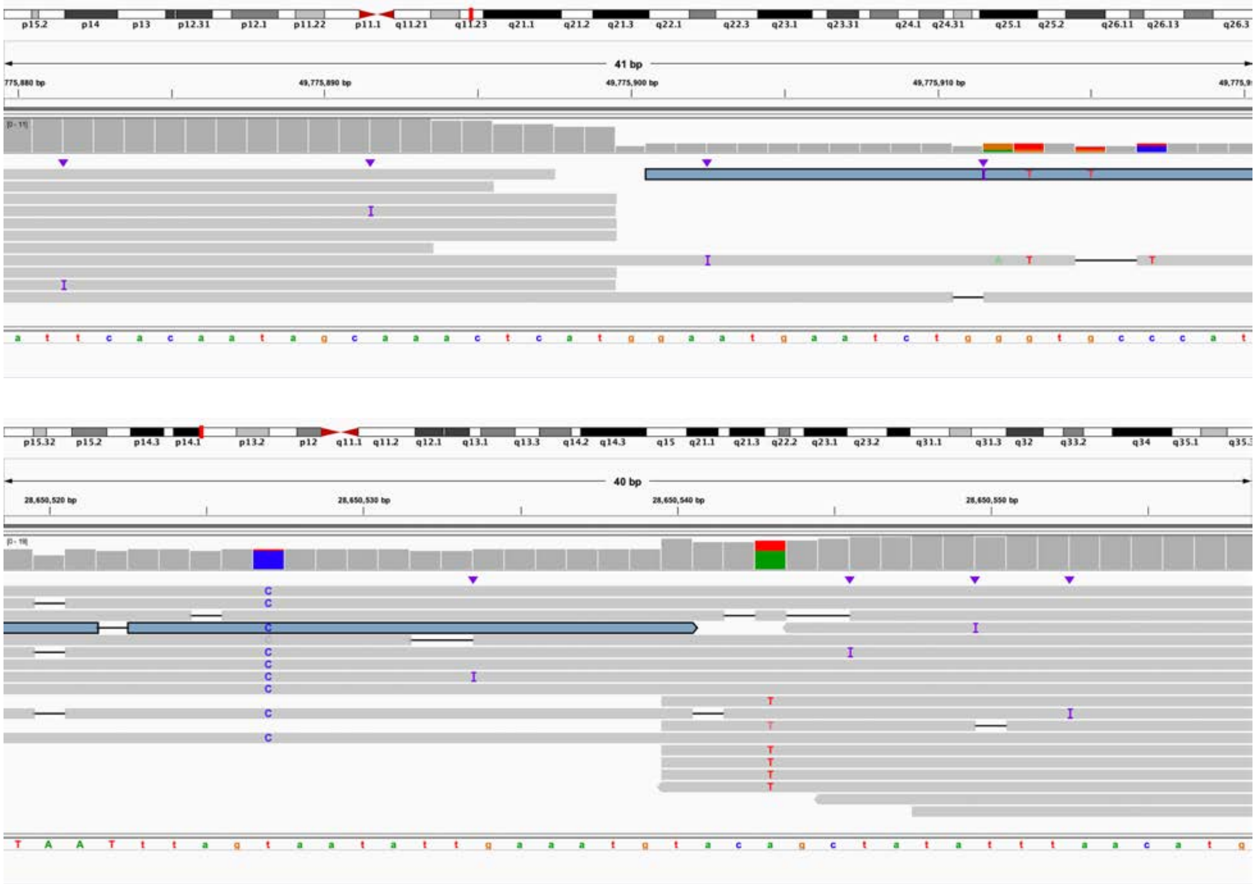
M: 5' end of 5:C (IGV: chr5:28,650,520-28,650,560) is linked to the 5' end of 10:K (IGV: chr10:56,736,263-56,736,303). This is the end of derivative chromosome 10.


p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q11.2 q12.1 q13.1 q13.3 q14.2 q14.3 q15 q21.1 q21.3 q22.2 q23.1 q23.2 q31.1 q31.3 q32 q33.2 q34 q35.1
41 bp
22,881,840 bp
22,881,850 bp
22,881,860 bp
22,881,870 bp
[0 - 19]
t a t g a a a c c g t c a g a t c t c g t g a c a a c t c a c t c c c t a t c a c
p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
40 bp
55,632,860 bp
55,632,870 bp
55,632,880 bp
55,632,890 bp
[0 - 31]
t c t a c a t g a a a c t t t t t a t g c a c a t g g c c a c a t t a t t t c c

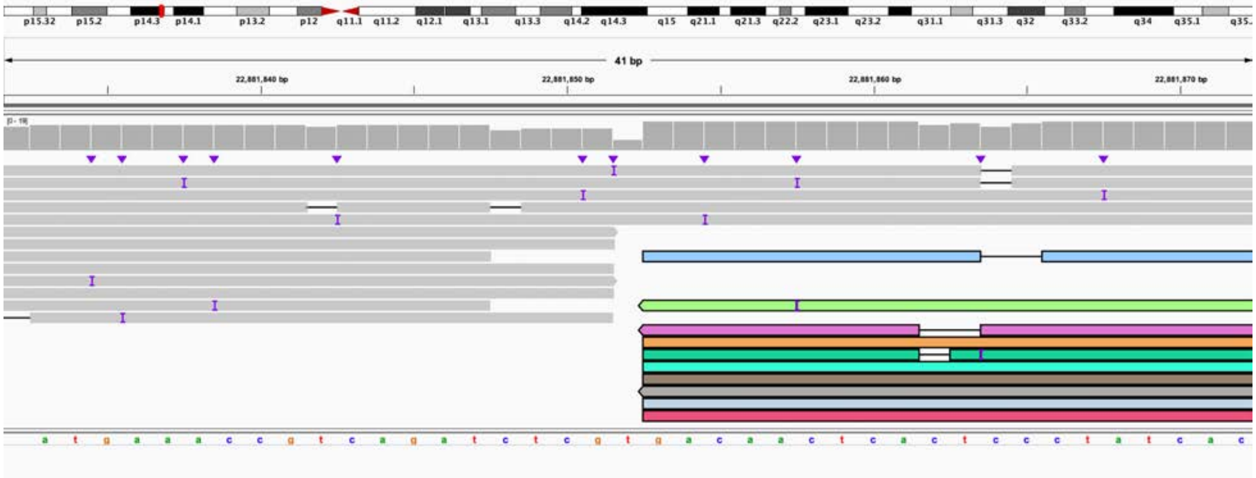
N: 3' end of 5:A (IGV: chr5:22,881,832-22,881,872) linked to 3' end of 10:I (IGV: chr10:55,632,859-55,632,898). This is the beginning of derivative chromosome 5.

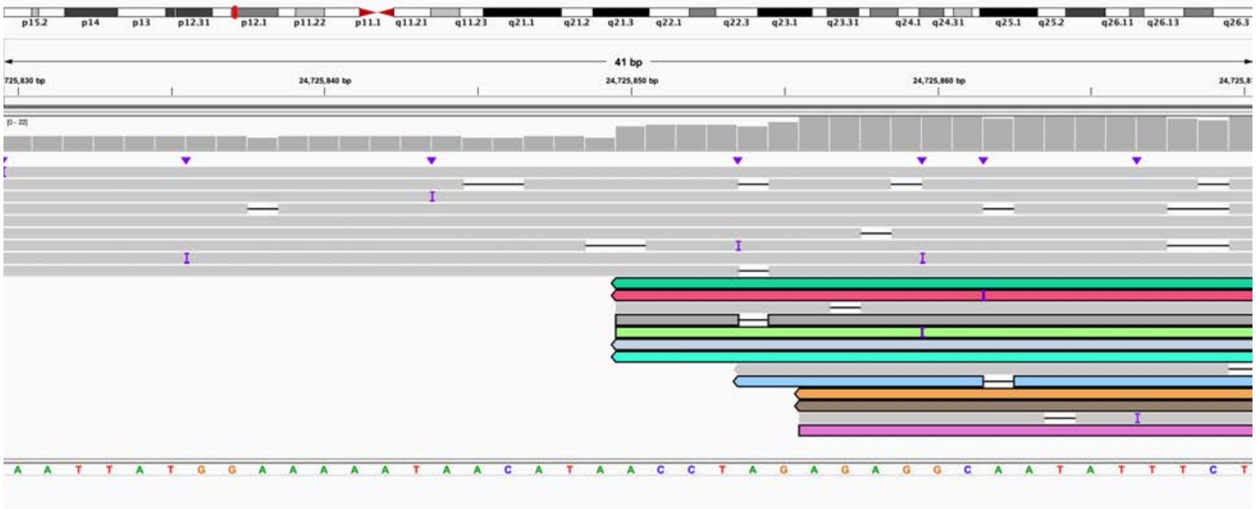


O: 5' end of 10:I bisects *PCDH15* (IGV: chr10:55,305,294-55,305,334) linked to 3' end of 10:G which also bisects *PCDH15* (IGV: chr10:54,672,847-54,672,886). This is within derivative chromosome 5.

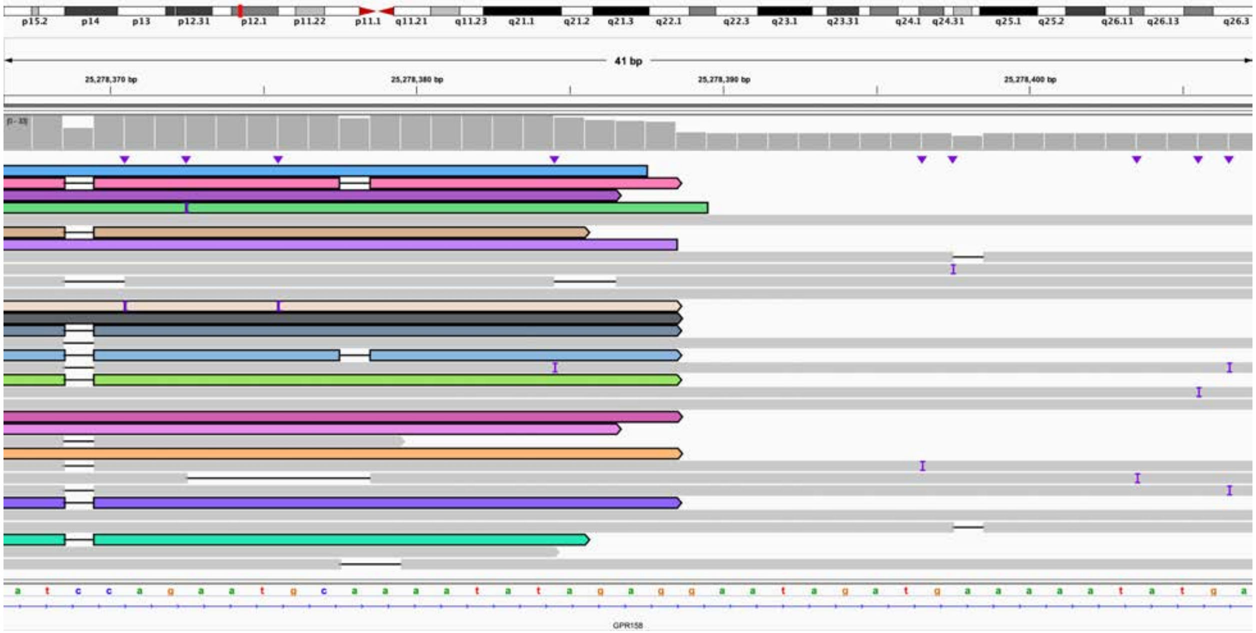

p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
775,880 bp
49,775,890 bp
49,775,900 bp
49,775,910 bp
49,775,9
[0 - 11]
a t t c a c a a t a g c a a a c t c a t g g a a t g a a t c t g g g t g c c c a t
p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q11.2 q12.1 q13.1 q13.3 q14.2 q14.3 q15 q21.1 q21.3 q22.2 q23.1 q23.2 q31.1 q31.3 q32 q33.2 q34 q35.1 q35.3
40 bp
28,650,520 bp
28,650,530 bp
28,650,540 bp
28,650,550 bp
[0 - 19]
T A A T t t a g t a a t a t t g a a a t g t a c a g c t a t a t t t a a c a t g

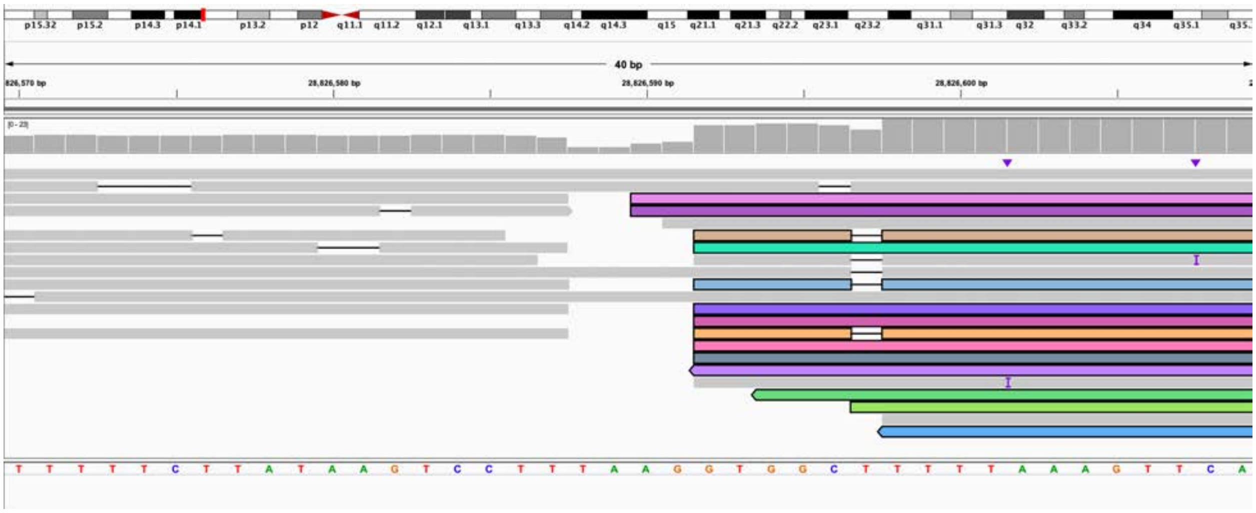
P: 5' end of 10:G (IGV: chr10:49,775,880-49,775,920) linked to 3' end of 5:B (IGV: chr5:28,650,519-28,650,558). This is within derivative chromosome 5.


p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q11.2 q12.1 q13.1 q13.3 q14.2 q14.3 q15 q21.1 q21.3 q22.2 q23.1 q23.2 q31.1 q31.3 q32 q33.2 q34 q35.1 q35.
41 bp
22,881,840 bp
22,881,850 bp
22,881,860 bp
22,881,870 bp
[0 - 19]
a t g a a a c c g t c a g a t c t c g t g a c a a c t c a c t c c c t a t c a c

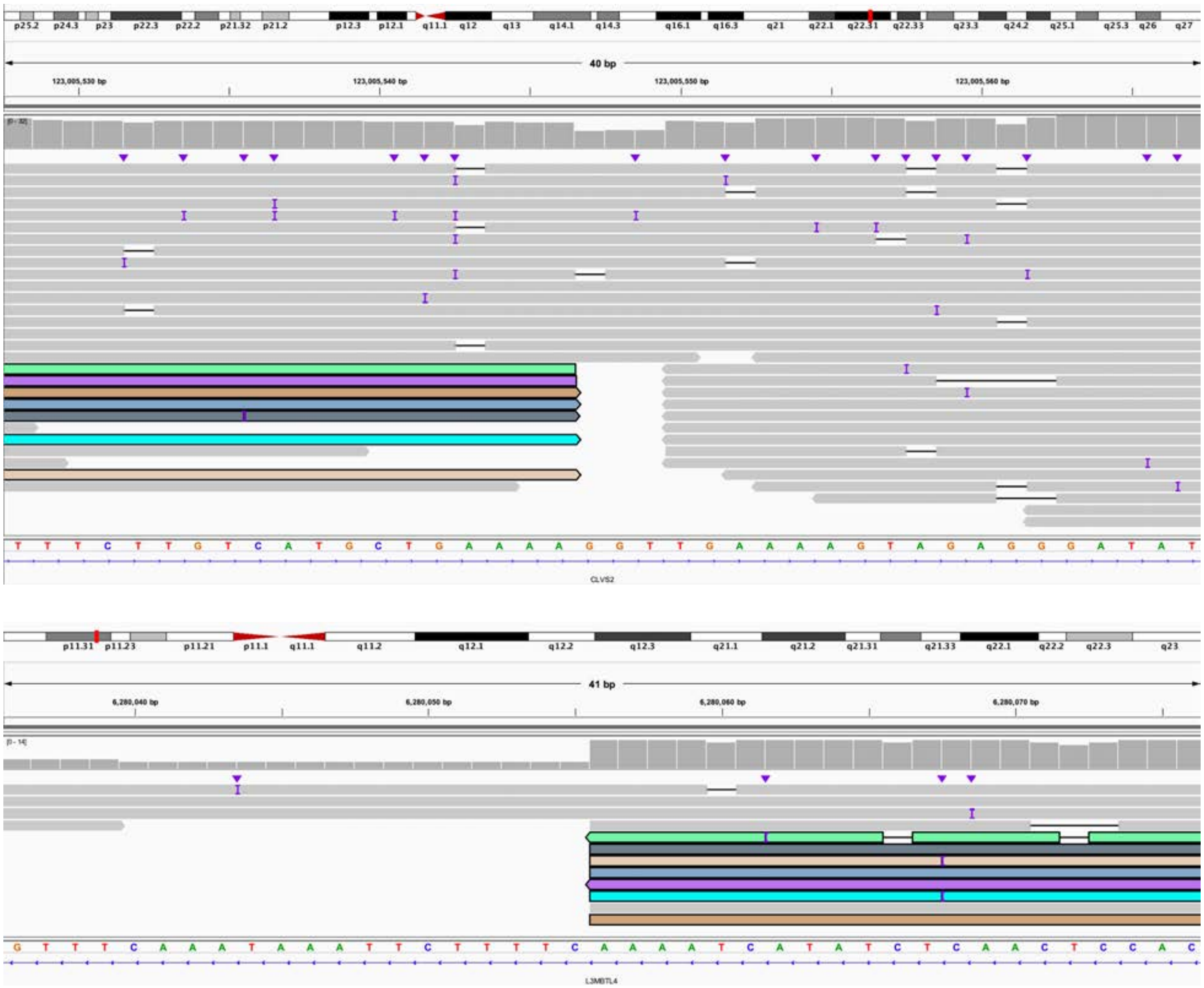

p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
725,830 bp
24,725,840 bp
24,725,850 bp
24,725,860 bp
24,725,8
[0 - 22]
A A T T A T G G A A A A A T A A C A T A A C C T A G A G A G G C A A T A T T T C T

Q: 5' end of 5:B (IGV: chr5:22,881,832-22,881,872) linked to 5' end of 10:C (IGV: chr10:24,725,830-24,725,870). This is within derivative chromosome 5.

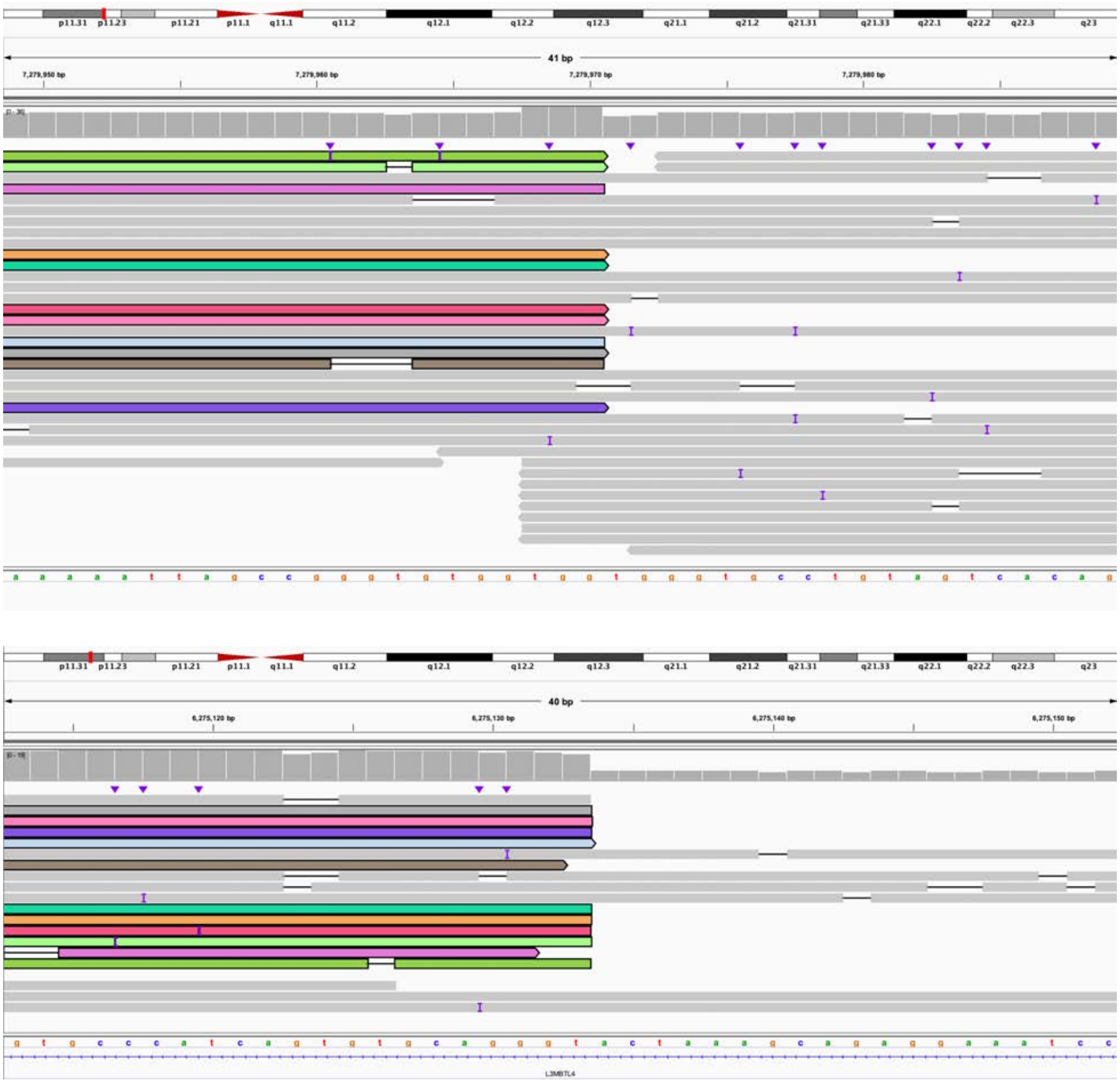

p15.2 p14 p13 p12.31 p12.1 p11.22 p11.1 q11.21 q11.23 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.31 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
25,278,370 bp
25,278,380 bp
25,278,390 bp
25,278,400 bp
[0 - 23]
a t c c a g a a t g c a a a a t a t a g a g g a a t a g a t g a a a a a t a t g a
GPR158


p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q11.2 q12.1 q13.1 q13.3 q14.2 q14.3 q15 q21.1 q21.3 q22.2 q23.1 q23.2 q31.1 q31.3 q32 q33.2 q34 q35.1 q35.
40 bp
826,570 bp
28,826,580 bp
28,826,590 bp
28,826,600 bp
[0 - 23]
T T T T T C T T A T A A G T C C T T T A A G G T G G C T T T T T A A A G T T C A

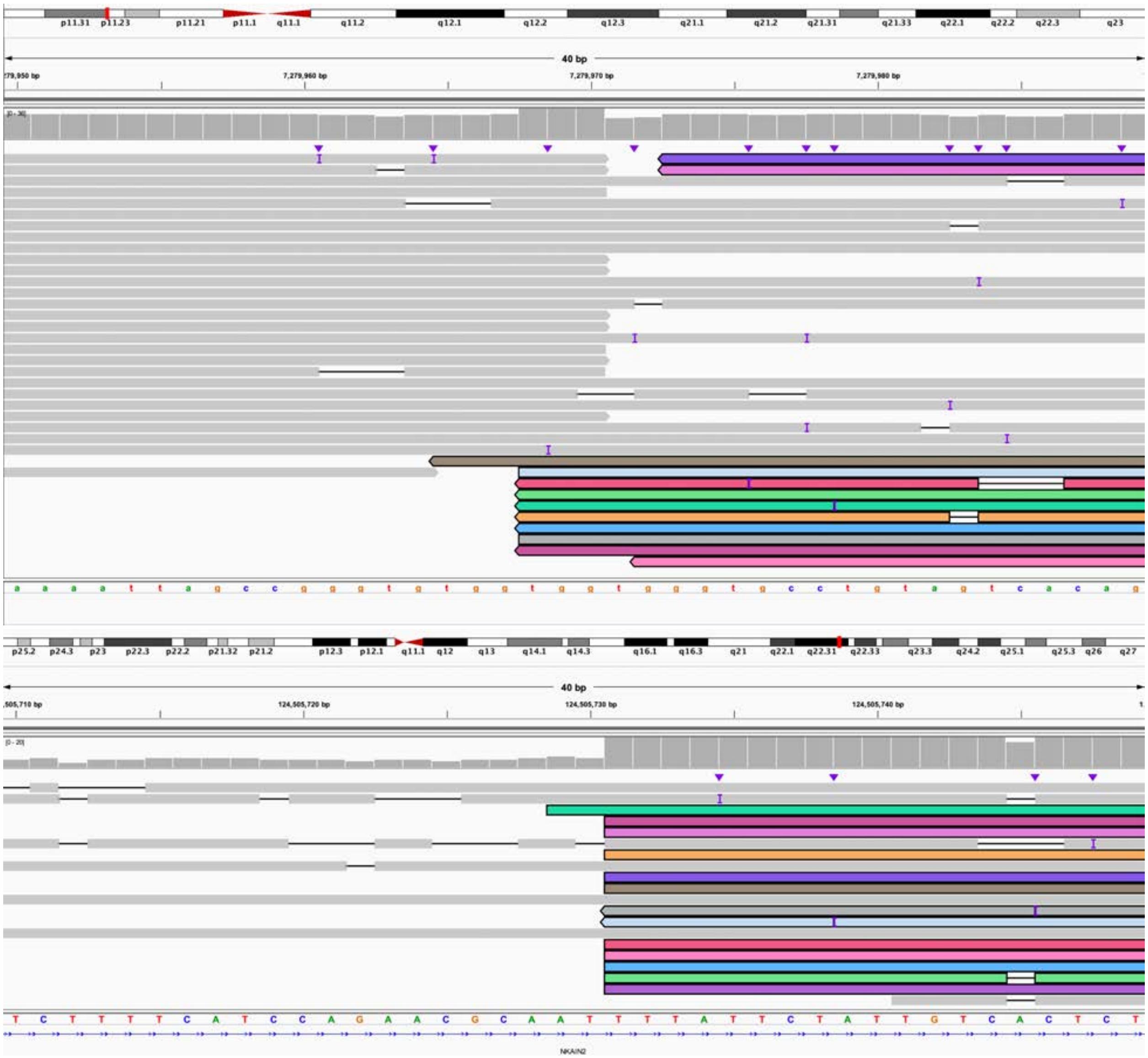
R: 3' end of 10:C is within *GPR158* (IGV: chr10:25,278,367-25,278,407) and is linked to 5' end of 5:D (IGV: chr5:28,826,570-28,826,609). This is the end of derivative chromosome 5.


p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
123,005,530 bp
123,005,540 bp
123,005,550 bp
123,005,560 bp
[0 - 32]
T T T C T T G T C A T G C T G A A A A G G T T G A A A A G T A G A G G G A T A T
CLVS2
p11.31 p11.23 p11.21 p11.1 q11.1 q11.2 q12.1 q12.2 q12.3 q21.1 q21.2 q21.31 q21.33 q22.1 q22.2 q22.3 q23
41 bp
6,280,040 bp
6,280,050 bp
6,280,060 bp
6,280,070 bp
[0 - 14]
G T T T C A A A T A A A T T C T T T T C A A A A T C A T A T C T C A A C T C C A C
L3MBTL4

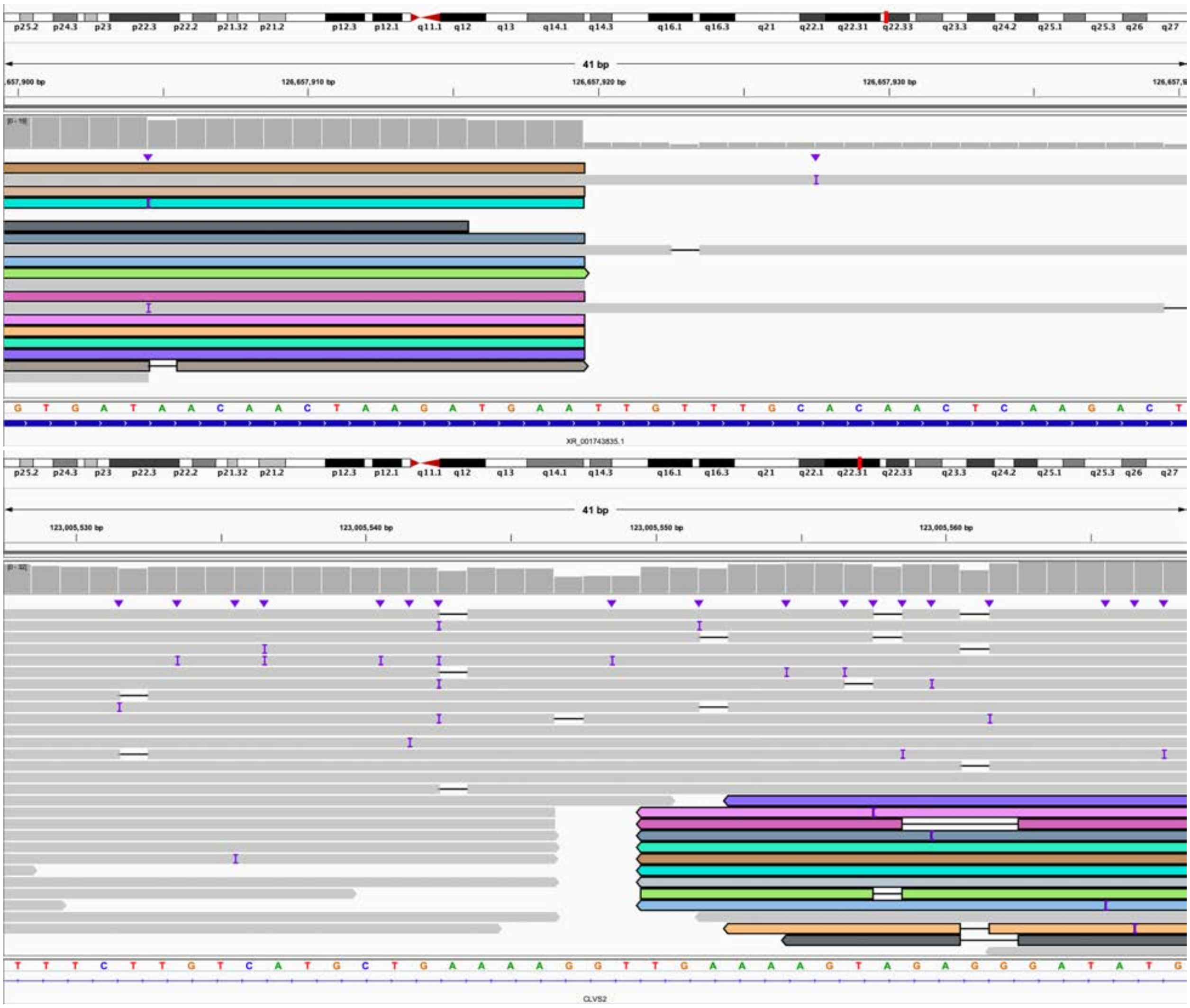
S: 3' end of 6:A bisects *CLVS2* (IGV: chr6:123,005,528-123,005,567) linked to 5' end of 18:C which bisects *L3MBTL4* (IGV: chr18:6,280,036-6,280,076). This is the beginning of derivative chromosome 6.


p11.31 p11.23 p11.21 p11.1 q11.1 q11.2 q12.1 q12.2 q12.3 q21.1 q21.2 q21.31 q21.33 q22.1 q22.2 q22.3 q23
41 bp
7,279,950 bp
7,279,960 bp
7,279,970 bp
7,279,980 bp
a a a a a t t a g c c g g g t g t g g t g g t g g g t g c c t g t a g t c a c a g
p11.31 p11.23 p11.21 p11.1 q11.1 q11.2 q12.1 q12.2 q12.3 q21.1 q21.2 q21.31 q21.33 q22.1 q22.2 q22.3 q23
40 bp
6,275,120 bp
6,275,130 bp
6,275,140 bp
6,275,150 bp
g t g c c c a t c a g t g t g c a g g g t a c t a a a g c a g a g g a a a t c c
L3MBTL4

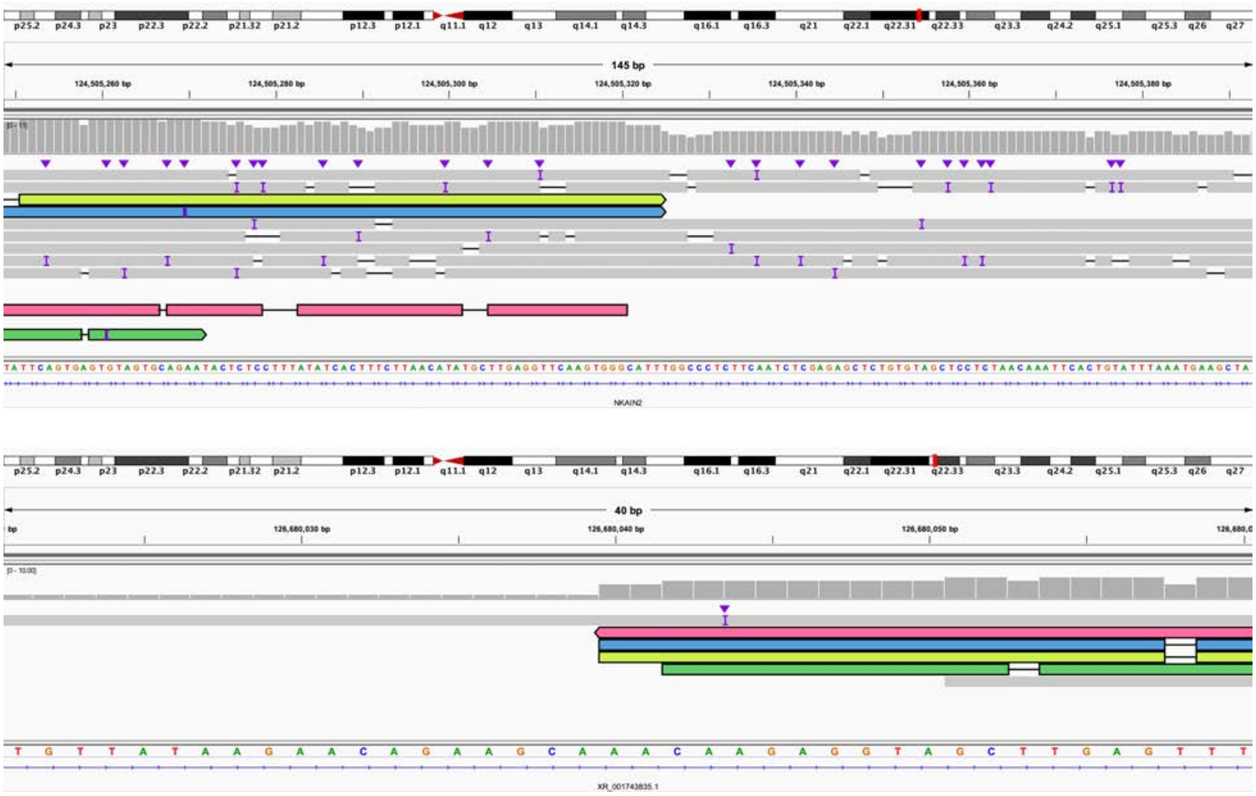
T: 3' end of 18:C (IGV: chr18:7,279,949-7,279,989) linked to 3' end of 18:A which bisects *L3MBTL4* (IGV: chr18:6,275,113-6,275,152). This is the end of derivative chromosome 6.


p11.31 p11.23 p11.21 p11.1 q11.1 q11.2 q12.1 q12.2 q12.3 q21.1 q21.2 q21.31 q21.33 q22.1 q22.2 q22.3 q23
40 bp
279,950 bp
7,279,960 bp
7,279,970 bp
7,279,980 bp
[0 - 30]
a a a a t t a g c c g g g t g t g g t g g t g g g t g c c t g t a g t c a c a g
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
,505,710 bp
124,505,720 bp
124,505,730 bp
124,505,740 bp
[0 - 20]
T C T T T T C A T C C A G A A C G C A A T T T T A T T C T A T T G T C A C T C T
NKAIN2

U: 5' end of 18:D (IGV: chr18:7,279,950-7,279,989) linked to 5' end of 6:C which bisects *NKAIN2* (IGV: chr6:124,505,710-124,505,749). This is the beginning of derivative chromosome 18.

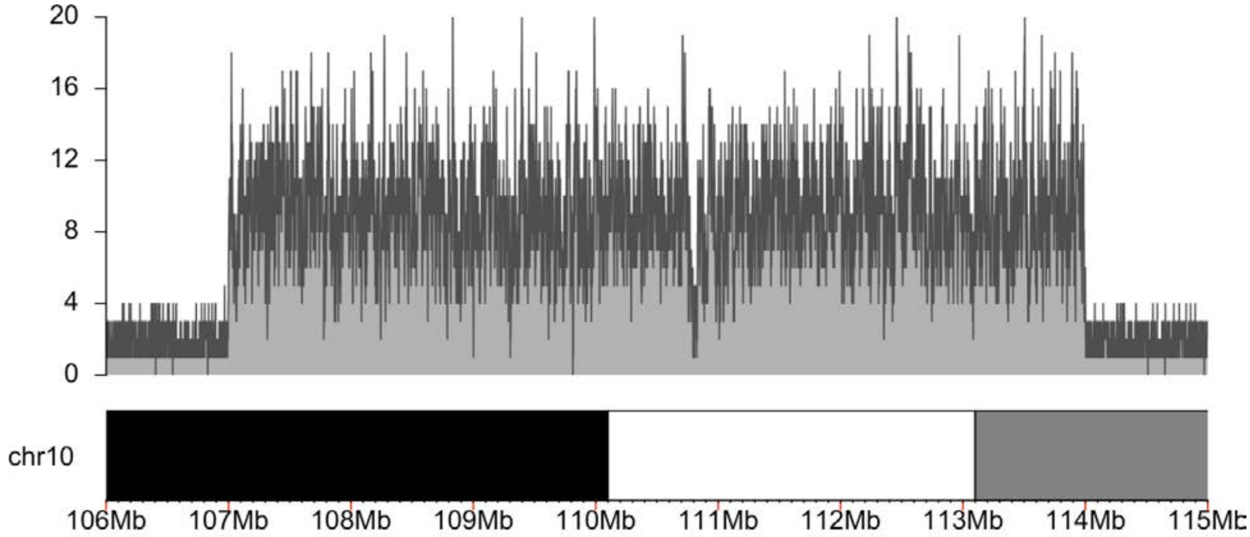

p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
41 bp
657,900 bp
126,657,910 bp
126,657,920 bp
126,657,930 bp
126,657,9
G T G A T A A C A A C T A A G A T G A A T T G T T T G C A C A A C T C A A G A C T
XR_001743835.1
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
41 bp
123,005,530 bp
123,005,540 bp
123,005,550 bp
123,005,560 bp
T T T C T T G T C A T G C T G A A A A G G T T G A A A A G T A G A G G G A T A T G
CLVS2

V: 3' end of 6:C bisects *XR_001743835.1* (IGV: chr6:126,657,900-126,657,940) linked to 5' end of 6:B which bisects *CLVS2* (IGV: chr6:123,005,528-123,005,568). This is within derivative chromosome 18.

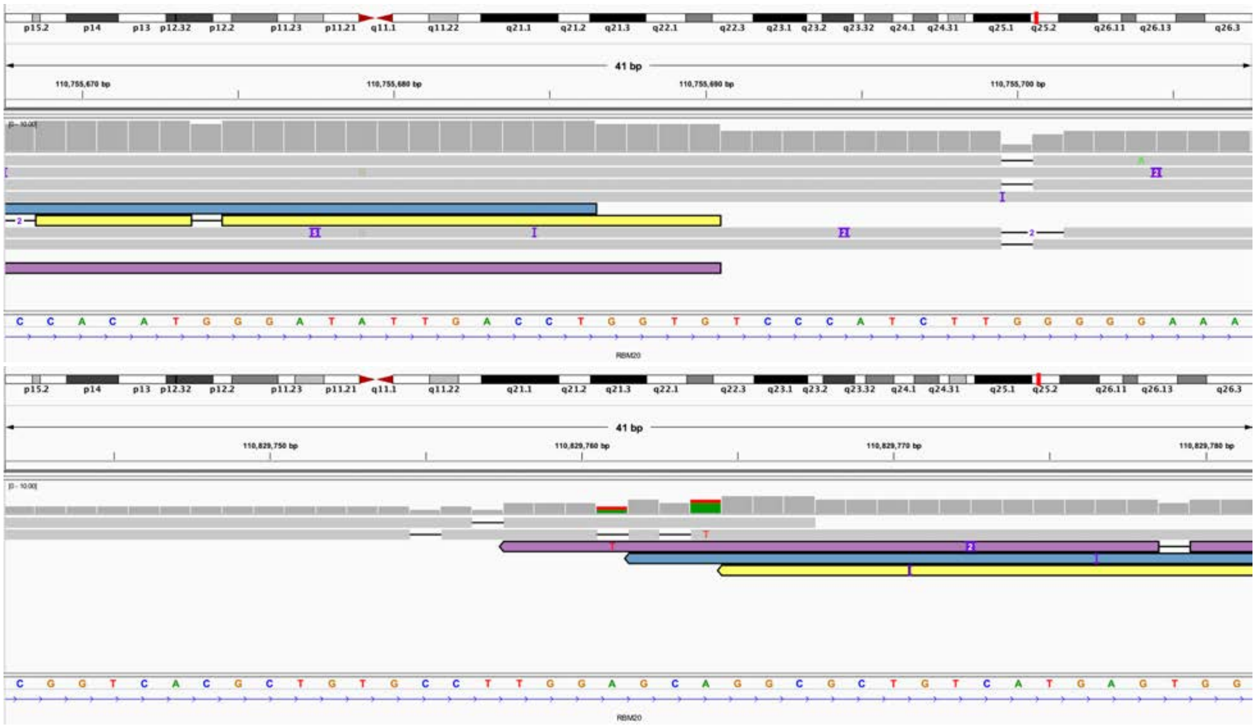

p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
145 bp
124,505,260 bp
124,505,280 bp
124,505,300 bp
124,505,320 bp
124,505,340 bp
124,505,360 bp
124,505,380 bp
TATTCAGTGAGTGTAGTGCAGAATACTCTCCTTTATATCACTTTCTTAACATATGCTTGAGGTTCAAGTGGGCATTTGGCCCTCTTCAATCTCGAGAGCTCTGTGTAGCTCCTCTAACAAATTCACTGTATTTAAATGAAGCTA
NKAIN2
p25.2 p24.3 p23 p22.3 p22.2 p21.32 p21.2 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.3 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
126,680,030 bp
126,680,040 bp
126,680,050 bp
[0 - 10.00]
T G T T A T A A G A A C A G A A G C A A A C A A G A G G T A G C T T G A G T T T
XR_001743835.1

W: 3' end of 6:B bisects *NKAIN2* (IGV: chr6:124,505,249-124,505,393) linked to 5' end of 6:E which bisects *XR_001743835.1* (IGV: chr6:126,680,021-126,680,060). This is the end of derivative chromosome 18.

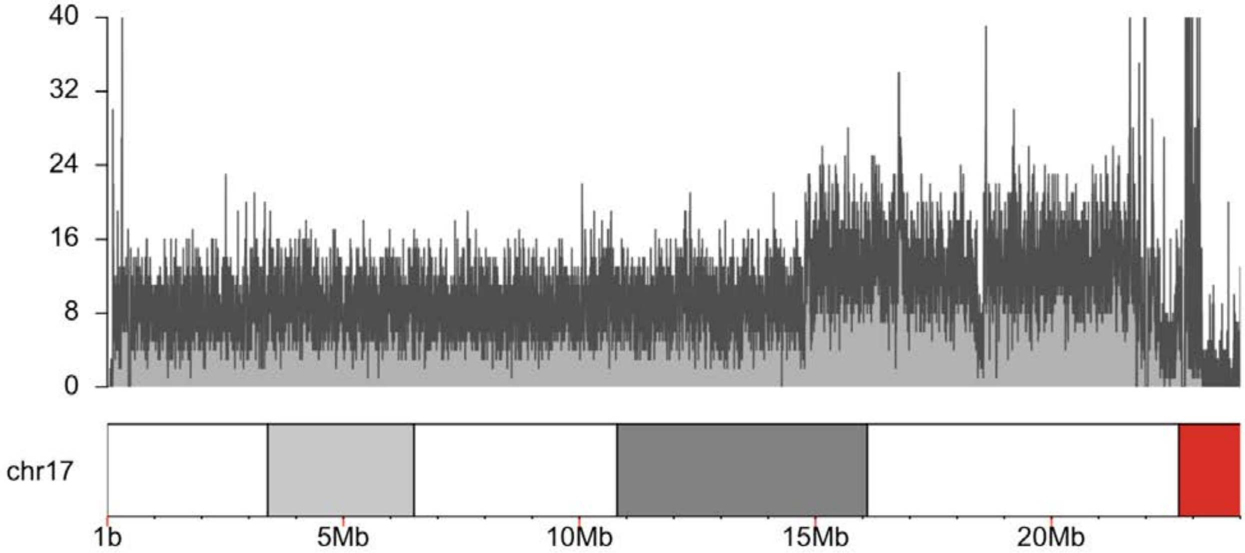
Figure S30. S082, individual with deletion of *RBM20* and duplication involving *RAI1* and *PMP22*.



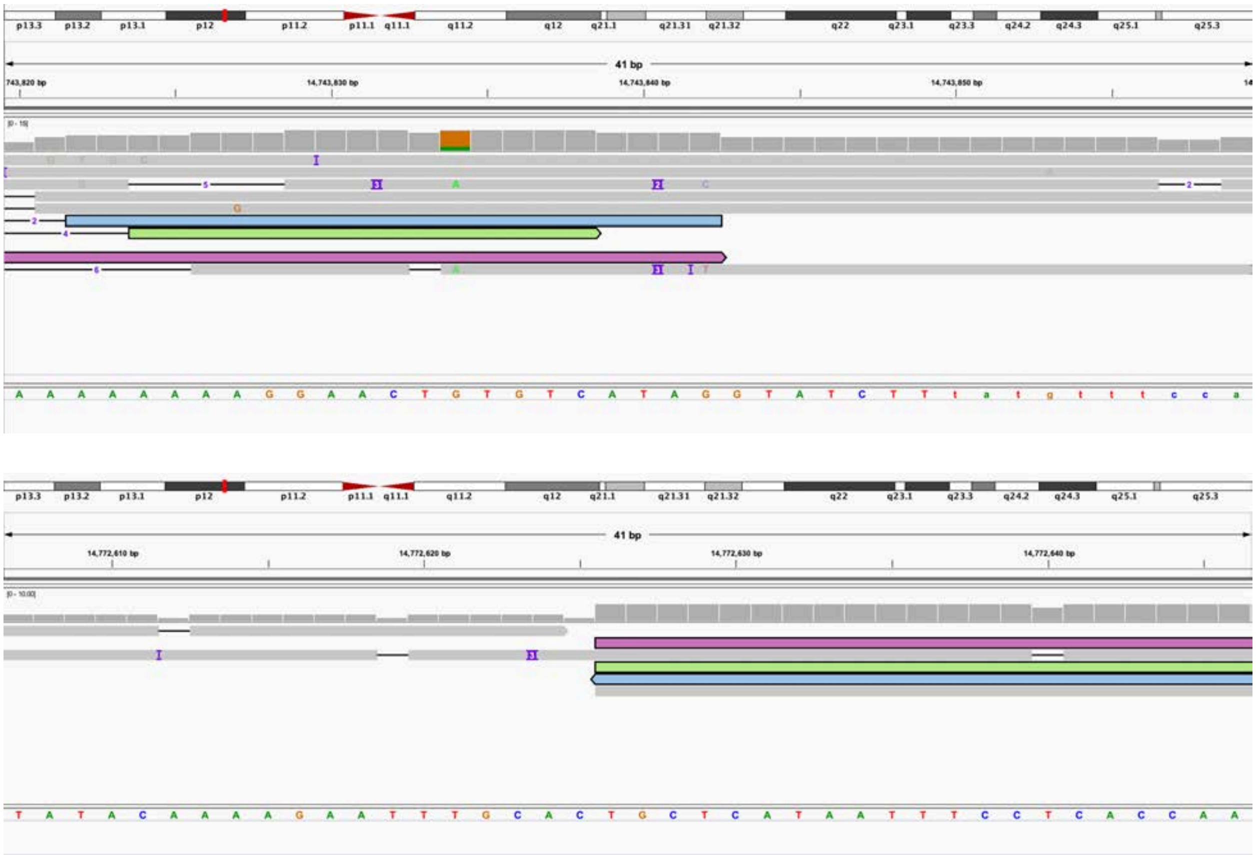
A. Coverage of chromosome 10 target region.


p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
110,755,670 bp
110,755,680 bp
110,755,690 bp
110,755,700 bp
[0 - 10.00]
C C A C A T G G G A T A T T G A C C T G G T G T C C C A T C T T G G G G G A A A
RBM20
p15.2 p14 p13 p12.32 p12.2 p11.23 p11.21 q11.1 q11.22 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.32 q24.1 q24.31 q25.1 q25.2 q26.11 q26.13 q26.3
41 bp
110,829,750 bp
110,829,760 bp
110,829,770 bp
110,829,780 bp
[0 - 10.00]
C G G T C A C G C T G T G C C T T G G A G C A G G C G C T G T C A T G A G T G G
RBM20

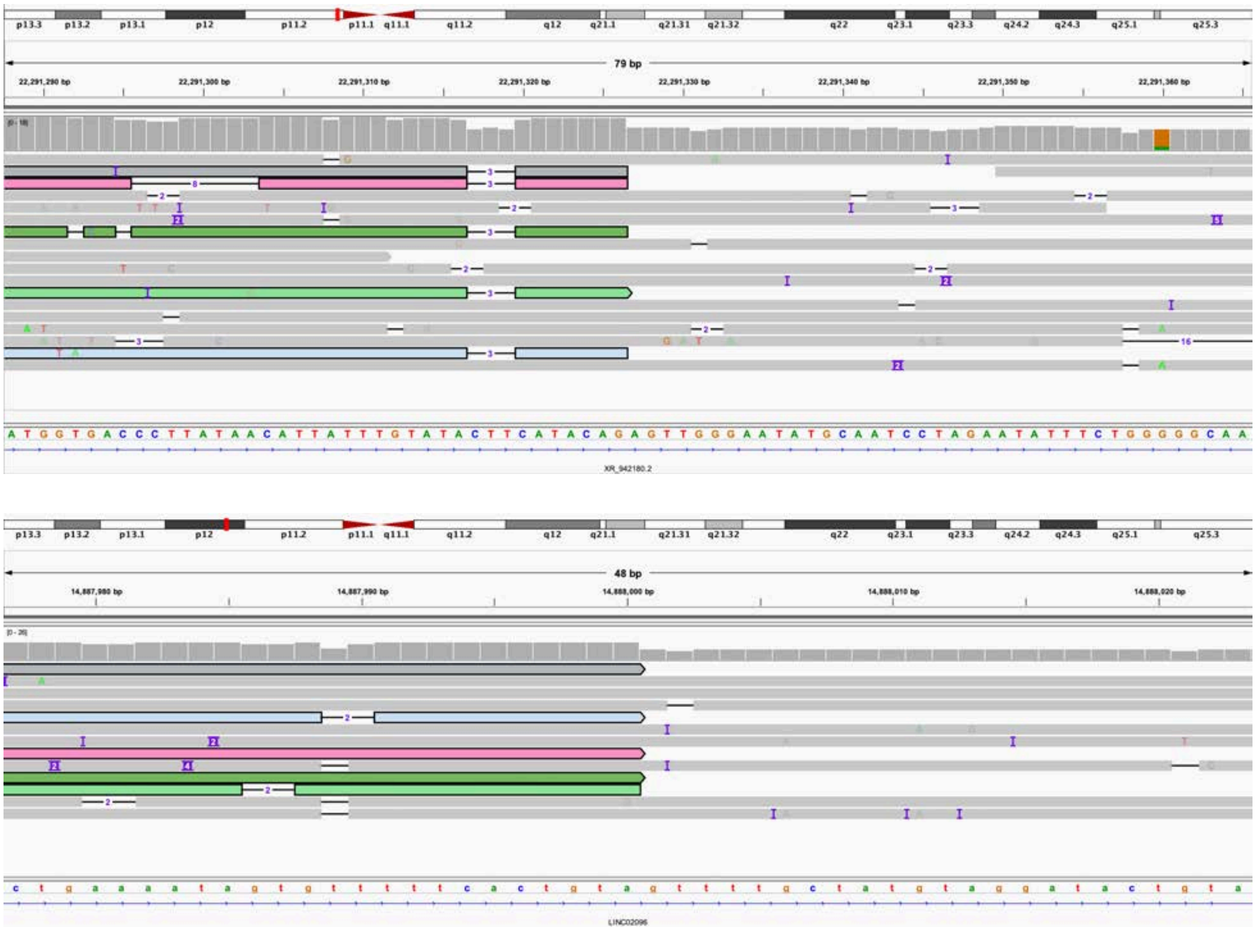
B: 5' end of deletion within RBM20 (IGV: chr10:110,755,668-110,755,707) is linked to 3' end of deletion (IGV: chr10:110,829,742-110,829,781).



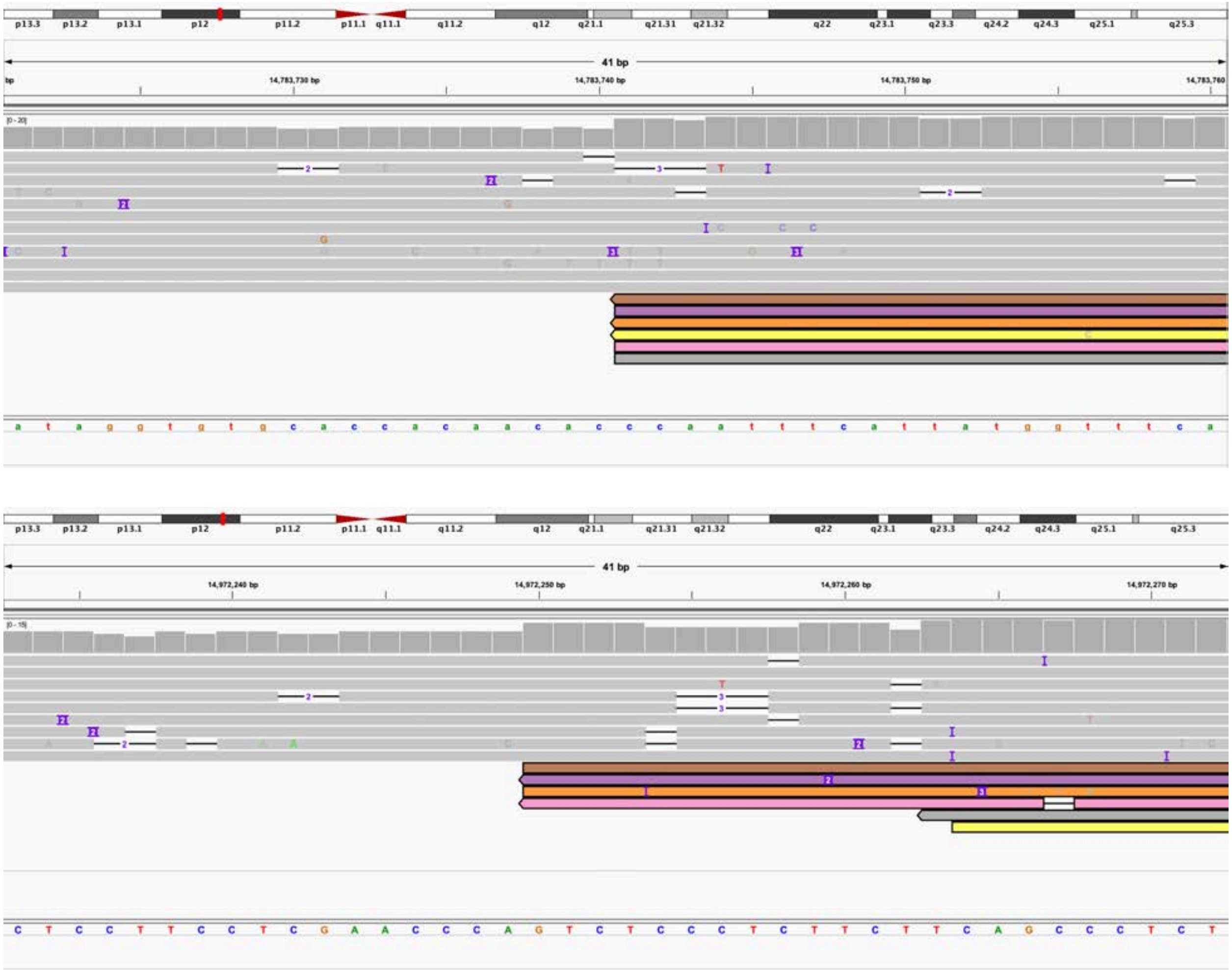
C: Coverage of chromosome 17 target region.


p13.3 p13.2 p13.1 p12 p11.2 p11.1 q11.1 q11.2 q12 q21.1 q21.31 q21.32 q22 q23.1 q23.3 q24.2 q24.3 q25.1 q25.3
41 bp
743,820 bp
14,743,830 bp
14,743,840 bp
14,743,850 bp
A A A A A A A A G G A A C T G T G T C A T A G G T A T C T T t a t g t t t c c a
p13.3 p13.2 p13.1 p12 p11.2 p11.1 q11.1 q11.2 q12 q21.1 q21.31 q21.32 q22 q23.1 q23.3 q24.2 q24.3 q25.1 q25.3
41 bp
14,772,610 bp
14,772,620 bp
14,772,630 bp
14,772,640 bp
T A T A C A A A A G A A T T T G C A C T G C T C A T A A T T T C C T C A C C A A

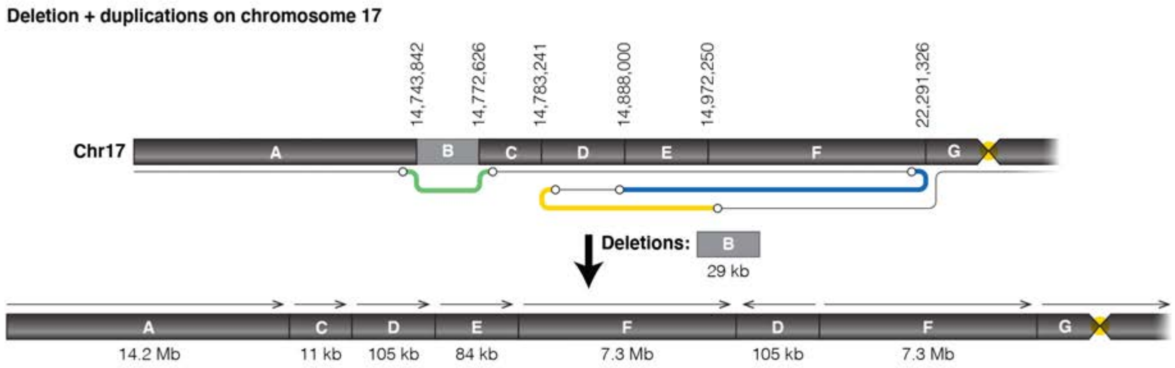
D: 5’ end of A (IGV: chr17:14,743,820-14,743,859) is linked to 3’ end of C (IGV: chr17:14,772,607-14,772,646) and results in a 29kb deletion.


p13.3 p13.2 p13.1 p12 p11.2 p11.1 q11.1 q11.2 q12 q21.1 q21.31 q21.32 q22 q23.1 q23.3 q24.2 q24.3 q25.1 q25.3
79 bp
22,291,290 bp
22,291,300 bp
22,291,310 bp
22,291,320 bp
22,291,330 bp
22,291,340 bp
22,291,350 bp
22,291,360 bp
[0 - 18]
A T G G T G A C C C T T A T A A C A T T A T T T G T A T A C T T C A T A C A G A G T T G G G A A T A T G C A A T C C T A G A A T A T T T C T G G G G G C A A
XR_942180.2
p13.3 p13.2 p13.1 p12 p11.2 p11.1 q11.1 q11.2 q12 q21.1 q21.31 q21.32 q22 q23.1 q23.3 q24.2 q24.3 q25.1 q25.3
48 bp
14,887,980 bp
14,887,990 bp
14,888,000 bp
14,888,010 bp
14,888,020 bp
[0 - 26]
c t g a a a a t a g t g t t t t t c a c t g t a g t t t t g c t a t g t a g g a t a c t g t a
LINC02096

E: 5’ end of F (IGV: chr17:22,291,288-22,291,365) is connected to the 5’ end of D (IGV: chr17:14,887,977-14,888,023).

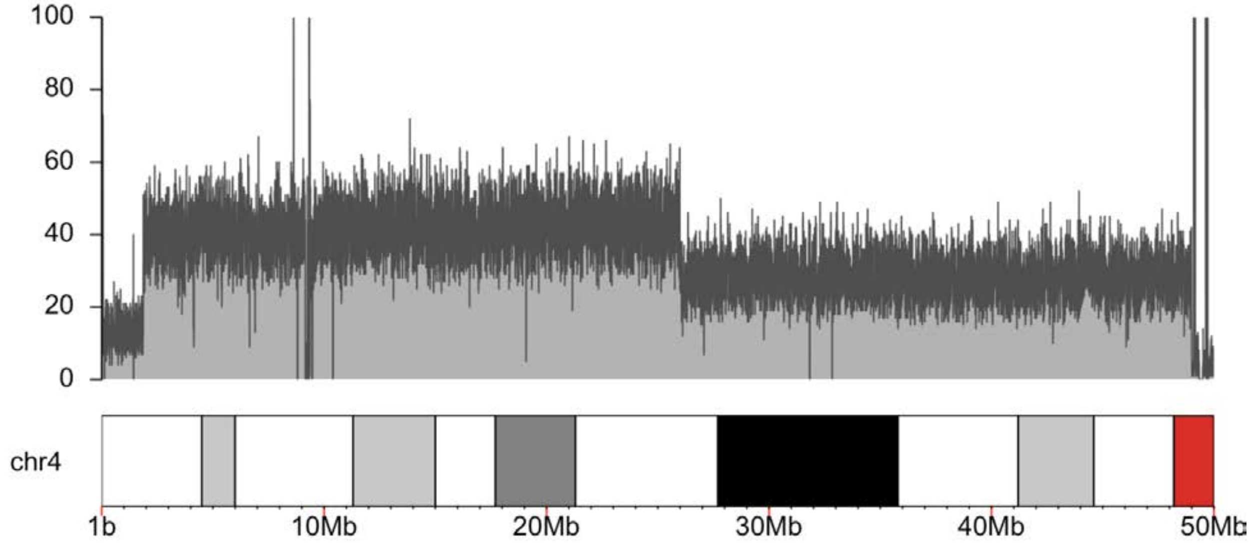


F: 3’ end of D (IGV: chr17:14,783,721-14,783,760) is connected to the 3’ end of F (IGV: chr17:14,972,233-14,972,272).

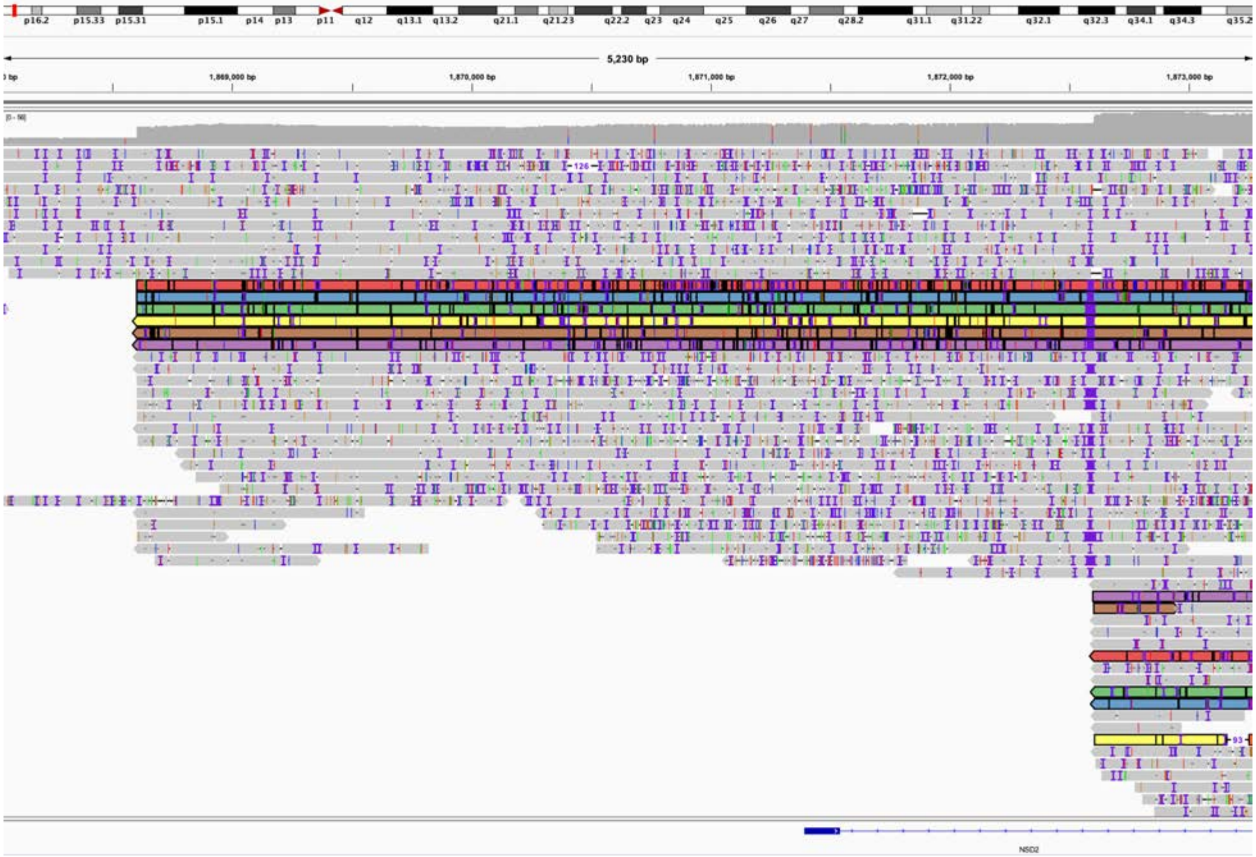


G: Cartoon of complex rearrangement on chromosome 17.

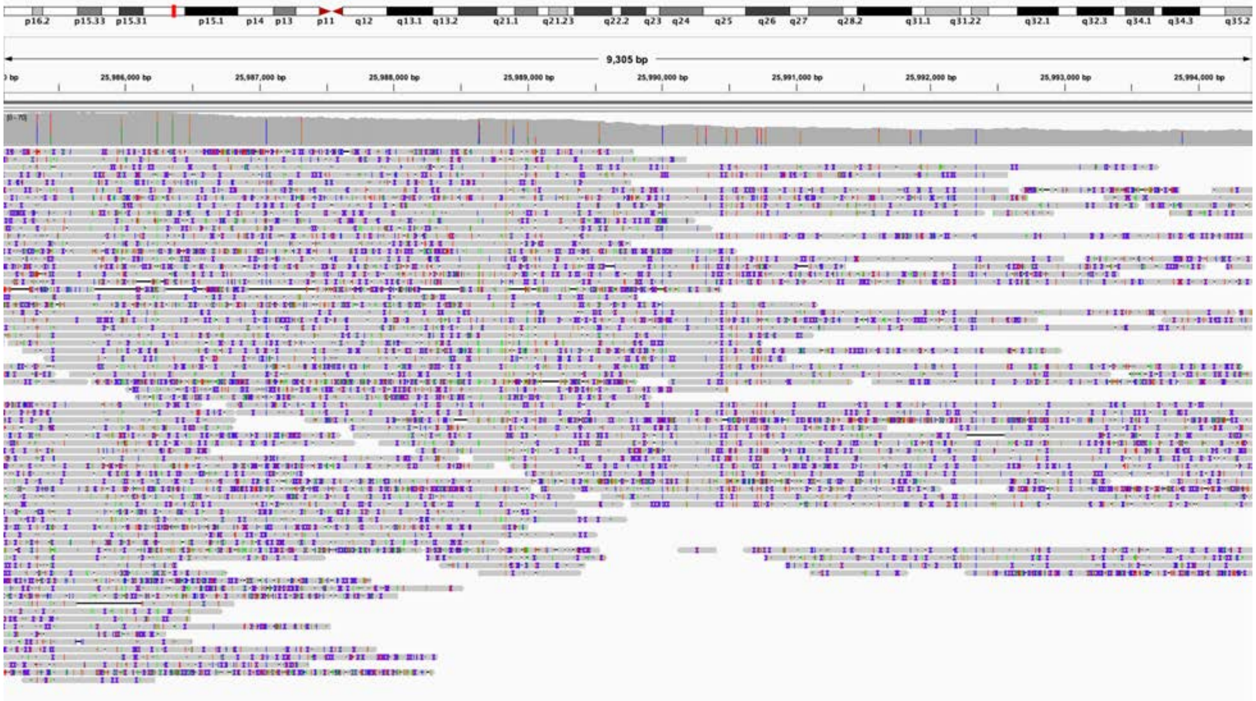
Figure S31. S083, individual with terminal deletion and proximal duplication of chromosome 4.



A. Coverage of target region showing terminal deletion followed by proximal duplication.

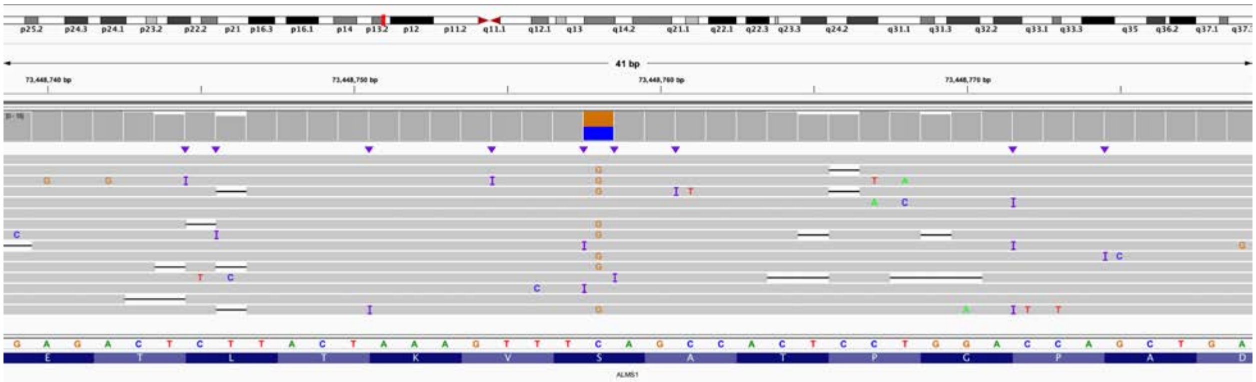


B. IGV view of deletion and duplication breakpoints. A small number of reads are highlighted to show the relationship between the duplicated region and the deletion. IGV view is of chr4:1,867,951-1,873,179.

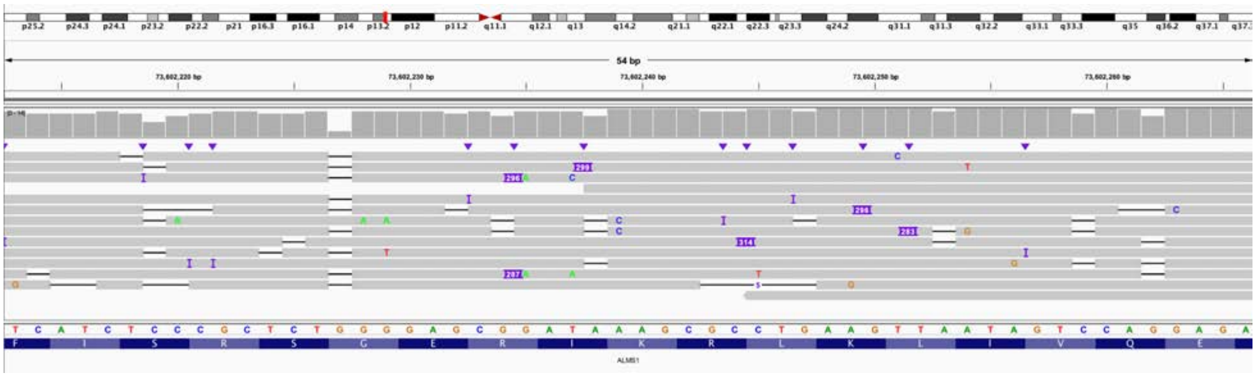


C. IGV view of the end of the duplication breakpoint. There is no clear termination of the duplication, suggesting a terminal event. IGV view is of chr4:25,985,086-25,994,389.

Figure S32. S002 (*ALMS1*), IGV views of known inherited stop variant and *Alu* insertion.

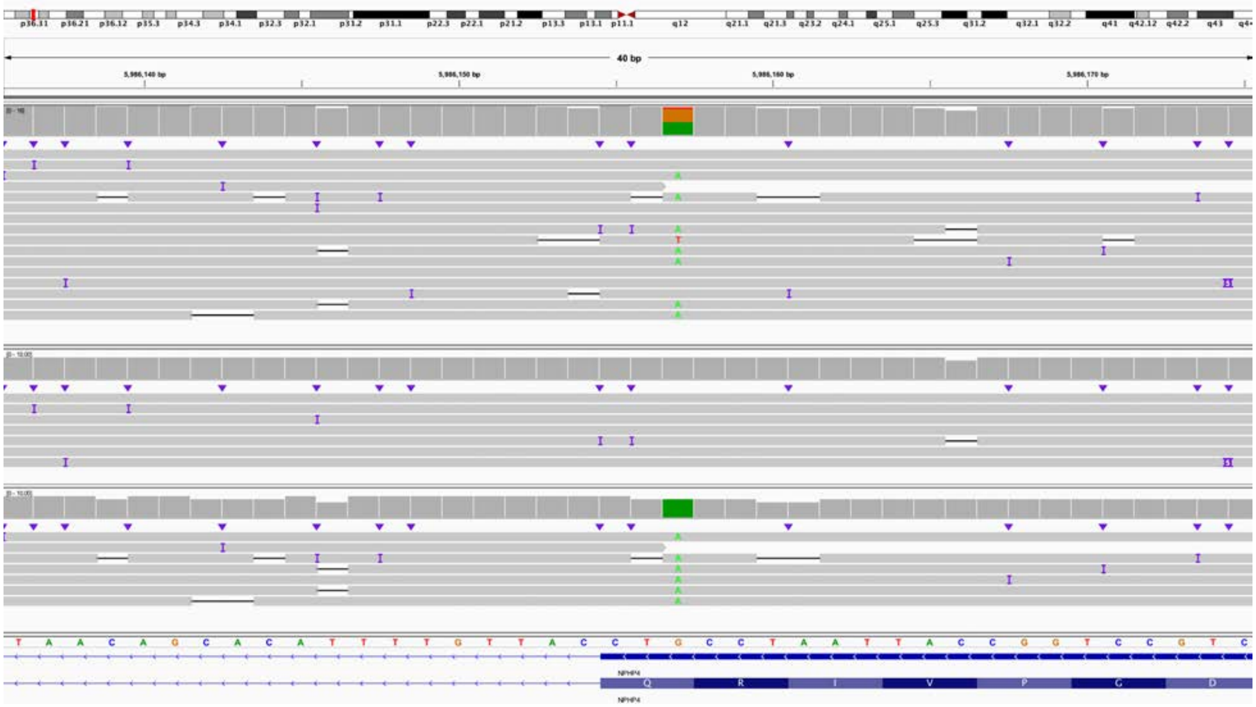

p25.2 p24.3 p24.1 p23.2 p22.2 p21 p16.3 p16.1 p14 p13.2 p12 p11.2 q11.1 q12.1 q13 q14.2 q21.1 q22.1 q22.3 q23.3 q24.2 q31.1 q31.3 q32.2 q33.1 q33.3 q35 q36.2 q37.1 q37.3
41 bp
73,448,740 bp
73,448,750 bp
73,448,760 bp
73,448,770 bp
G A G A C T C T T A C T A A A G T T T C A G C C A C T C C T G G A C C A G C T G A
E T L T K V S A T P G P A D
ALMS1

A. IGV view of all reads showing the known paternally inherited C>G that results in a stop codon. IGV coordinates for the screenshot are chr2:73,448,739-73,448,779.

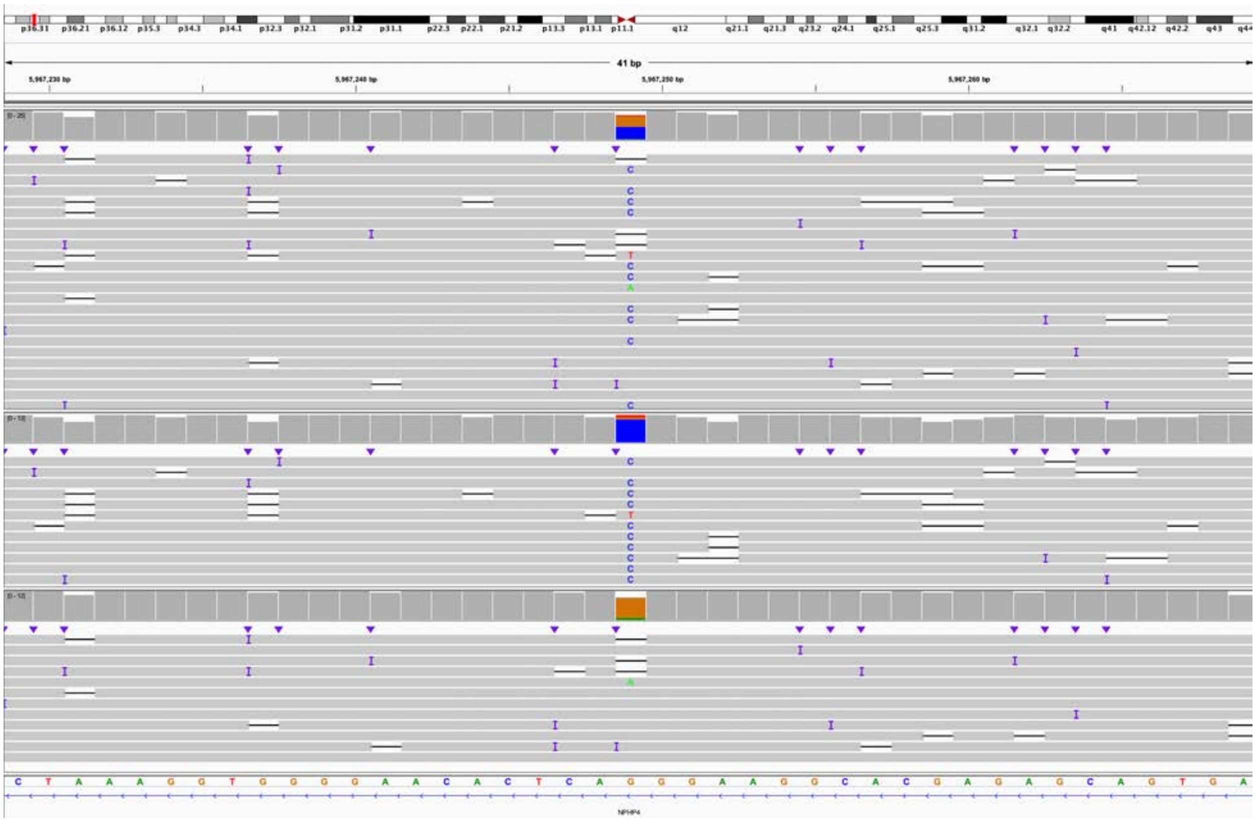

p25.2 p24.3 p24.1 p23.2 p22.2 p21 p16.3 p16.1 p14 p13.2 p12 p11.2 q11.1 q12.1 q13 q14.2 q21.1 q22.1 q22.3 q23.3 q24.2 q31.1 q31.3 q32.2 q33.1 q33.3 q35 q36.2 q37.1 q37.3
54 bp
73,602,220 bp
73,602,230 bp
73,602,240 bp
73,602,250 bp
73,602,260 bp
299
296
296
283
314
287
T C A T C T C C C G C T C T G G G G A G C G G A T A A A G C G C C T G A A G T T A A T A G T C C A G G A G A
F I S R S G E R I K R L K L I V Q E
ALMS1

B. Screenshot of ~300 bp insertion in exon 20 that represents an *Alu* insertion, all reads are shown. IGV view is chr2:73,602,213-73,602,266.

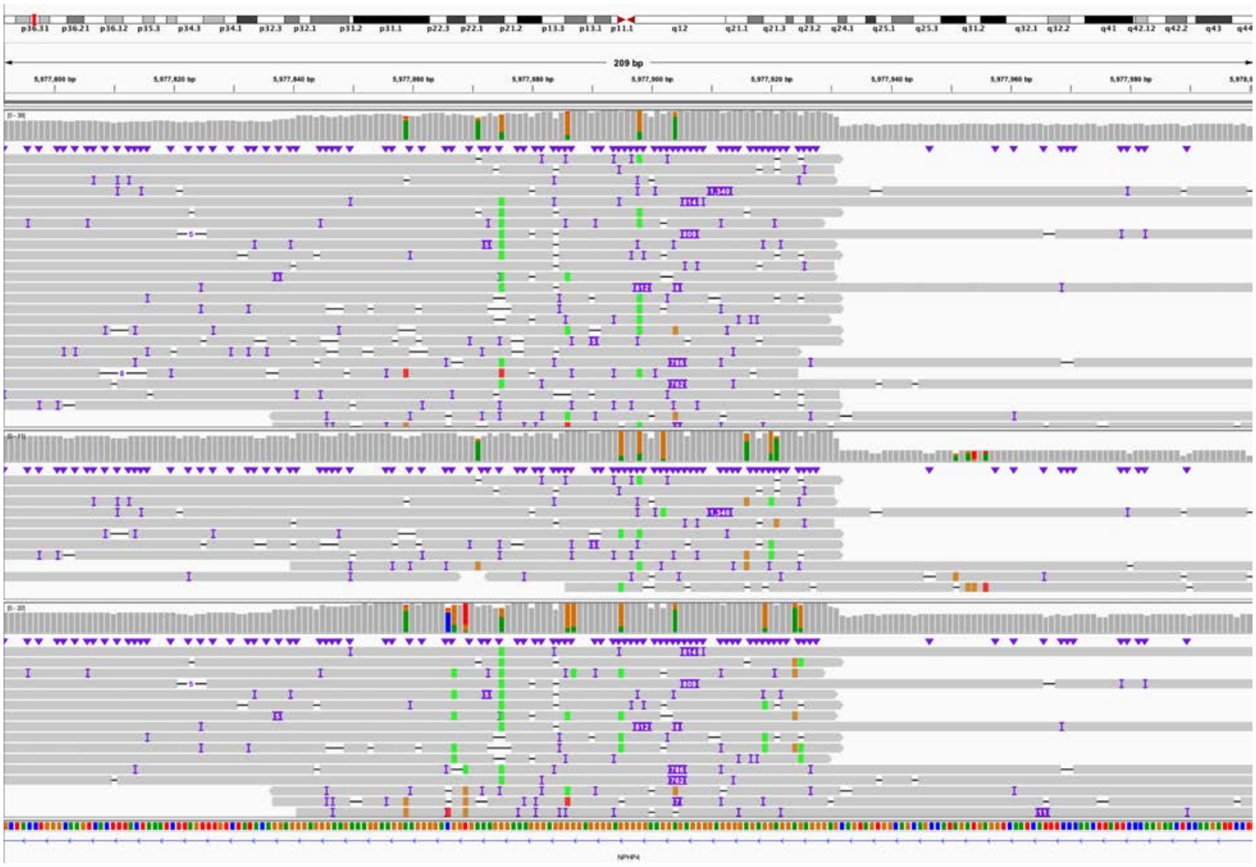
Figure S33. S003 (*NPHP4*), IGV views of inherited stop, splice variant, and data showing variant does affect splicing.


p36.31 p36.21 p36.12 p35.3 p34.3 p34.1 p32.3 p32.1 p31.2 p31.1 p22.3 p22.1 p21.2 p13.3 p13.1 p11.1 q12 q21.1 q21.3 q23.2 q24.1 q25.1 q25.3 q31.2 q32.1 q32.2 q41 q42.12 q42.2 q43
40 bp
5,986,140 bp
5,986,150 bp
5,986,160 bp
5,986,170 bp
T A A C A G C A C A T T T T G T T A C C T G C C T A A T T A C C G G T C C G T C
NPHP4
Q R I V P G D
NPHP4

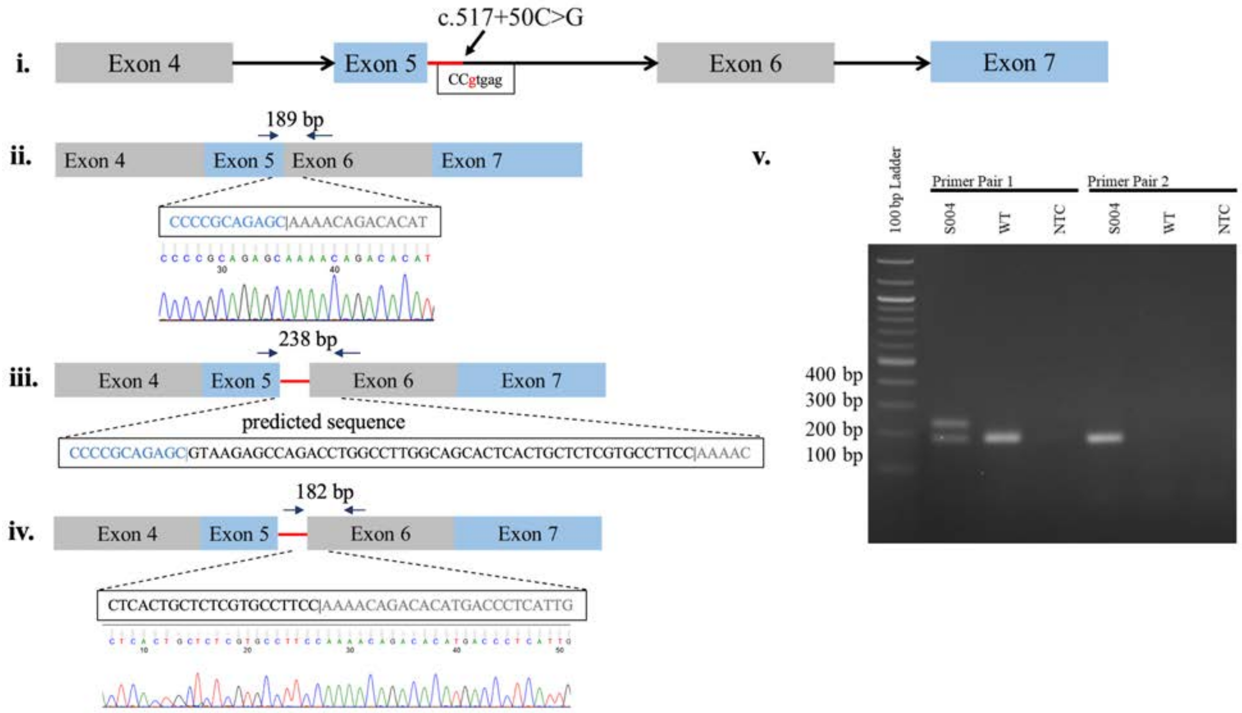
A. IGV view of all reads (top track) and reads phased into two haplotypes (middle and bottom tracks). DNA sequence and gene body are located below. The known paternally inherited G>A is located in the bottom haplotype. IGV screenshot is from chr1:5,986,136-5,986,175.


p36.31 p36.21 p36.12 p35.3 p34.3 p34.1 p32.3 p32.1 p31.2 p31.1 p22.3 p22.1 p21.2 p13.3 p13.1 p11.1 q12 q21.1 q21.3 q23.2 q24.1 q25.1 q25.3 q31.2 q32.1 q32.2 q41 q42.12 q42.2 q43 q44
41 bp
5,967,230 bp
5,967,240 bp
5,967,250 bp
5,967,260 bp
[0 - 25]
C T A A A G G T G G G G A A C A C T C A G G G A A G G C A C G A G A G C A G T G A
NPHP4

B. IGV view with reads as in A showing the G>C predicted to create a novel splice donor site on a different haplotype than the known paternally inherited G>A in A. IGV screenshot is from chr1:5,967,229-5,967,269.

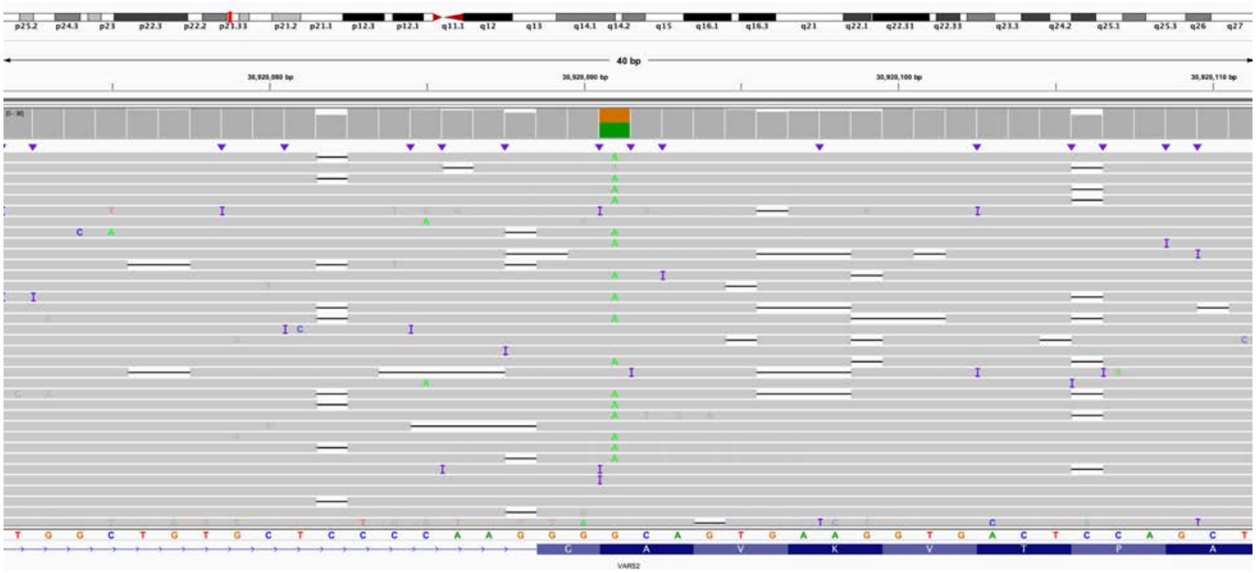


C. Analysis of SVs using both SVIM and Sniffles identified two insertions within an AG-rich repeat in an intron of *NPHP4* at approximately chr1:5,977,890. Phasing of reads separated all reads into a haplotype containing the ~800 bp insert and another containing the ~1,340 bp insert, similar to sizes observed in both nonhuman primates and human samples [4]. IGV screenshot is from chr1:5,977,792-5,978,001.

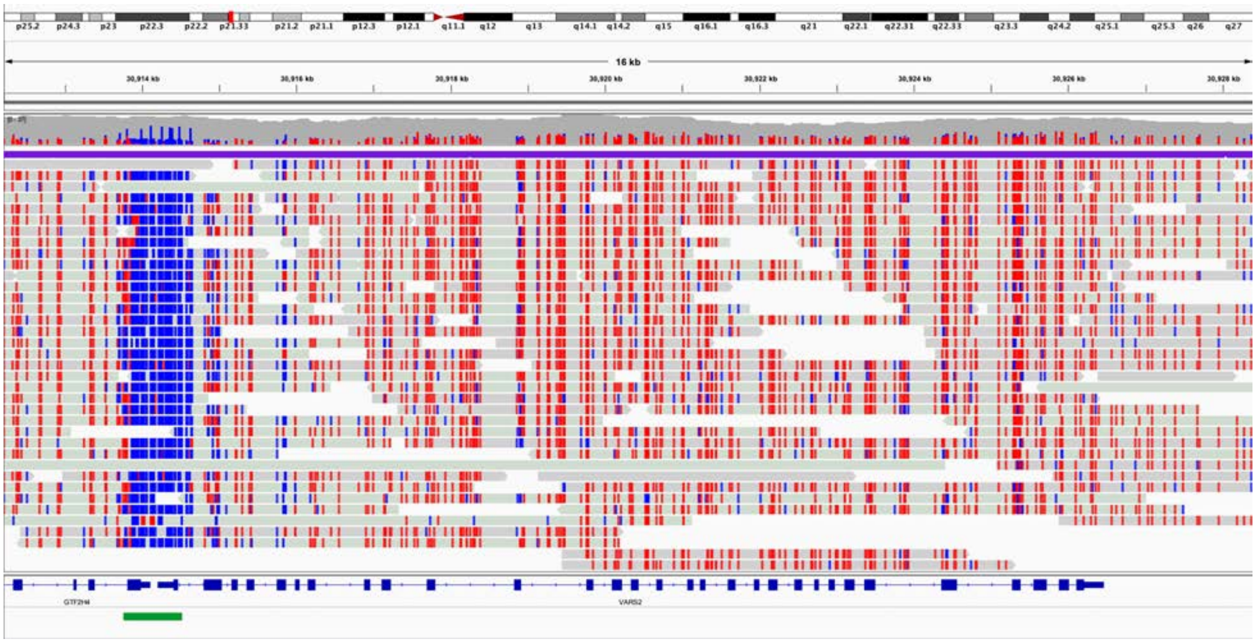


D. Splicing defect caused by deep intronic *NPHP4* variant in individual S004. **i.** *NPHP4* exon structure around the c.517+50C>G variant predicted to generate a new splice donor site (gtgag). **ii.** Normal splice isoform indicating primer pair 1 spanning the Exon 5-6 junction and corresponding Sanger sequencing of the PCR product. **iii.** Predicted aberrant transcript with inclusion of 49 intronic base pairs indicating predicted PCR product size for primer pair 1. **iv.** Predicted aberrant transcript indicating predicted PCR product size for primer pair 2, which should only amplify the aberrant transcript and corresponding Sanger sequencing of the PCR product. **v.** PCR products from S004 and unaffected fibroblast cDNA. Bands match the sizes predicted in B-D. NPHP4 reference sequence: NM_015102.4. bp=base pairs; NTC=No Template Control PCR; WT=wild-type, unaffected fibroblast cDNA.

Figure S34. S004 (*VARS2*), IGV view of known inherited variant.

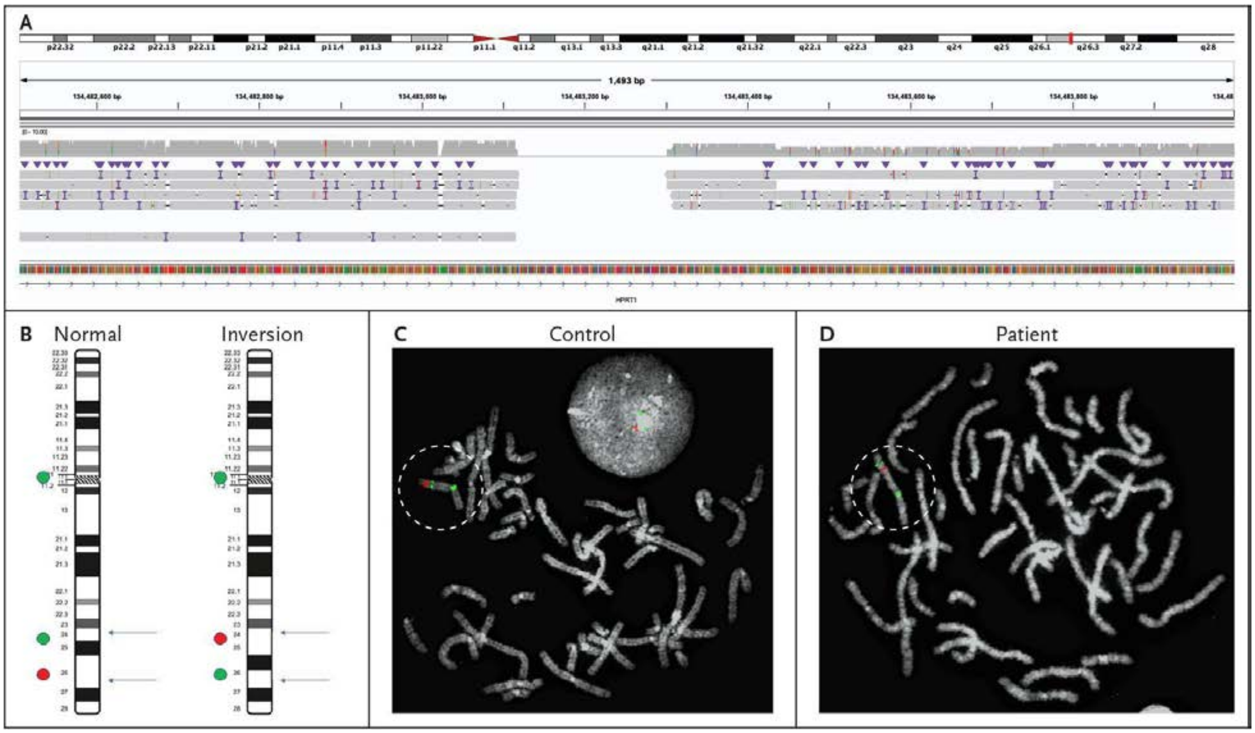

p25.2 p24.3 p23 p22.3 p22.2 p21.33 p21.2 p21.1 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.2 q15 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
40 bp
30,920,080 bp
30,920,090 bp
30,920,100 bp
30,920,110 bp
T G G C T G T G C T C C C C A A G G G G C A G T G A A G G T G A C T C C A G C T
G A V K V T P A
VARS2

A. IGV view of known pathogenic paternally inherited G>A. No second variant was found in this case. IGV view is of chr6:30,920,072-30,920,111.


p25.2 p24.3 p23 p22.3 p22.2 p21.33 p21.2 p21.1 p12.3 p12.1 q11.1 q12 q13 q14.1 q14.2 q15 q16.1 q16.3 q21 q22.1 q22.31 q22.33 q23.3 q24.2 q25.1 q25.3 q26 q27
16 kb
30,914 kb
30,916 kb
30,918 kb
30,920 kb
30,922 kb
30,924 kb
30,926 kb
30,928 kb
GTF2H4
VARS2

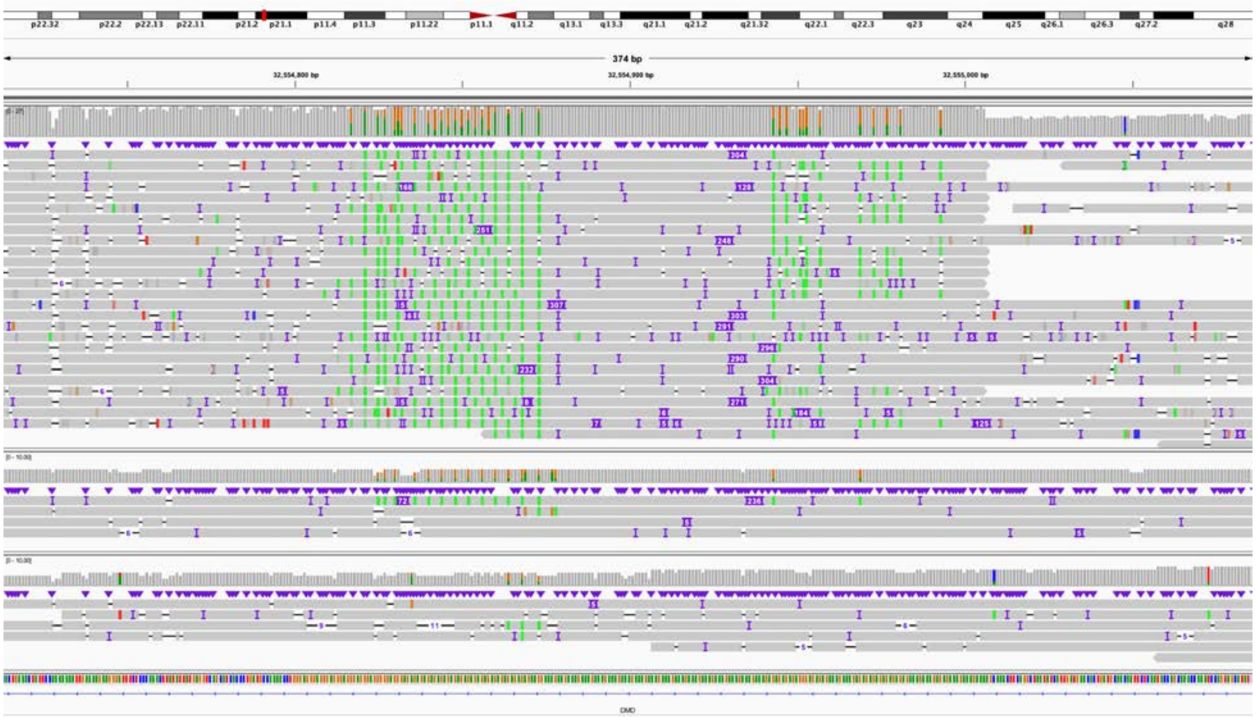
B. Analysis of reads reveals no change in methylation at the 5' UTR of *VARS2*. In this view, blue represents a hypomethylated region, suggesting an open 5' region and promoter. The green box under the promoter indicates the position of the CpG island. IGV view is chr6:30,912,208-30,928,459.

Figure S35. S008 (*HPRT1*), view of inversion and FISH results.

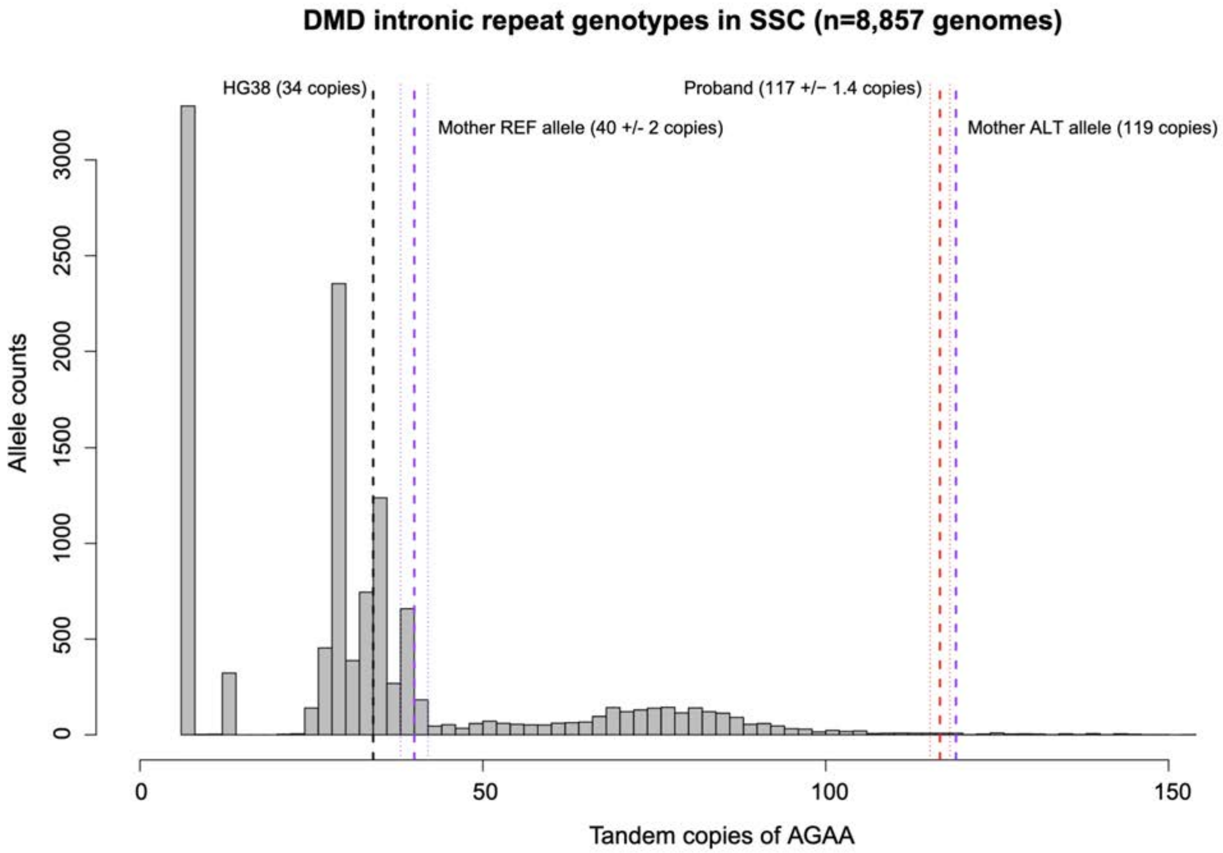


A. T-LRS suggested an approximately 17 Mbp inversion bisected *HPRT1*. Reads on the left partially mapped to approximately chrX:117,359,013 and reads on the right partially mapped to chrX:117,488,245. IGV view is chrX:134,482,505-134,484,004. **B.** FISH probes were designed to confirm the presence of an inversion. **C.** Image from control case, dashed circle highlights X chromosome. **D.** Image from the individual confirmed the presence of a 17 Mbp inversion reported as inversion (X)(q24q26.3), dashed circle highlights the X chromosome.

Figure S36. S009 (*DMD*), IGV view of AGAA expansion and frequency in SSC samples.

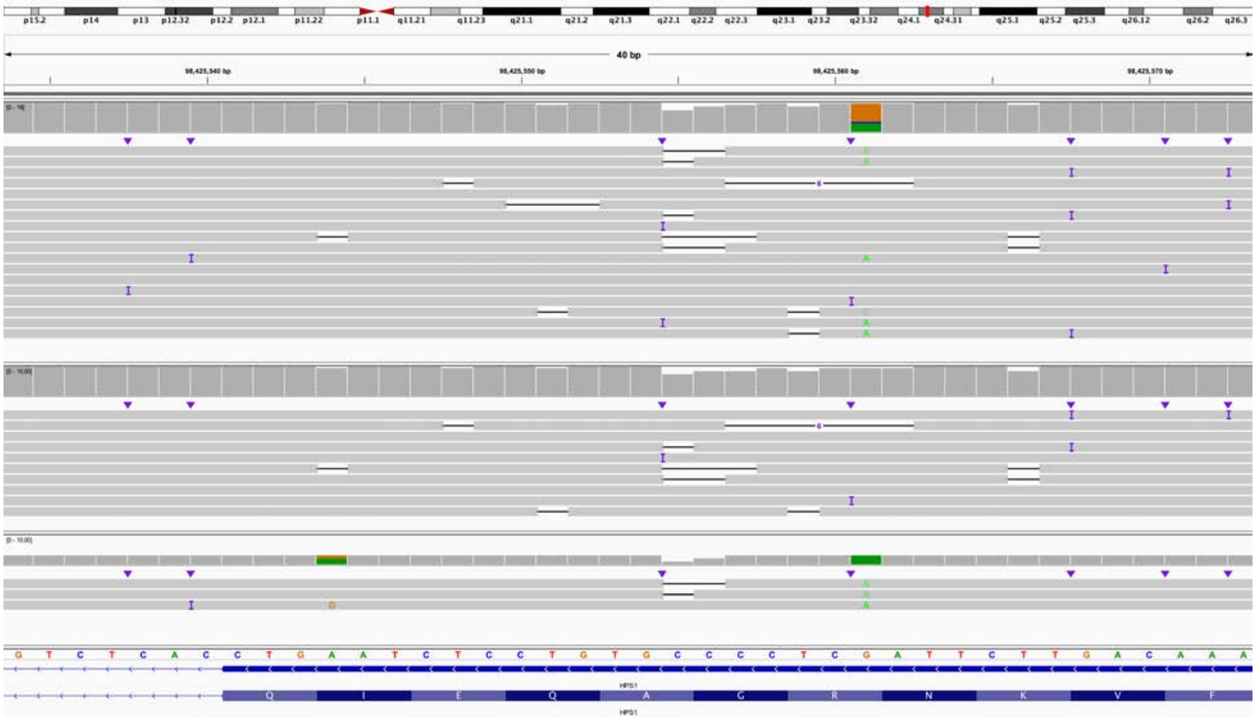


A. Expansion of an AGAA repeat represents a variant of uncertain significance in a child with a clinical diagnosis of Duchenne muscular dystrophy, but no molecular diagnosis. This view shows the insertion in intron 16 of the gene that is homozygous in the proband (top panel) and heterozygous in the bottom panel (the insert is divided into a 72 bp insert and a 236 bp insert). Bottom panel is from the proband’s unaffected full brother who does not have an insertion at this position. IGV view is of chrX:32,554,713-32,555,087.

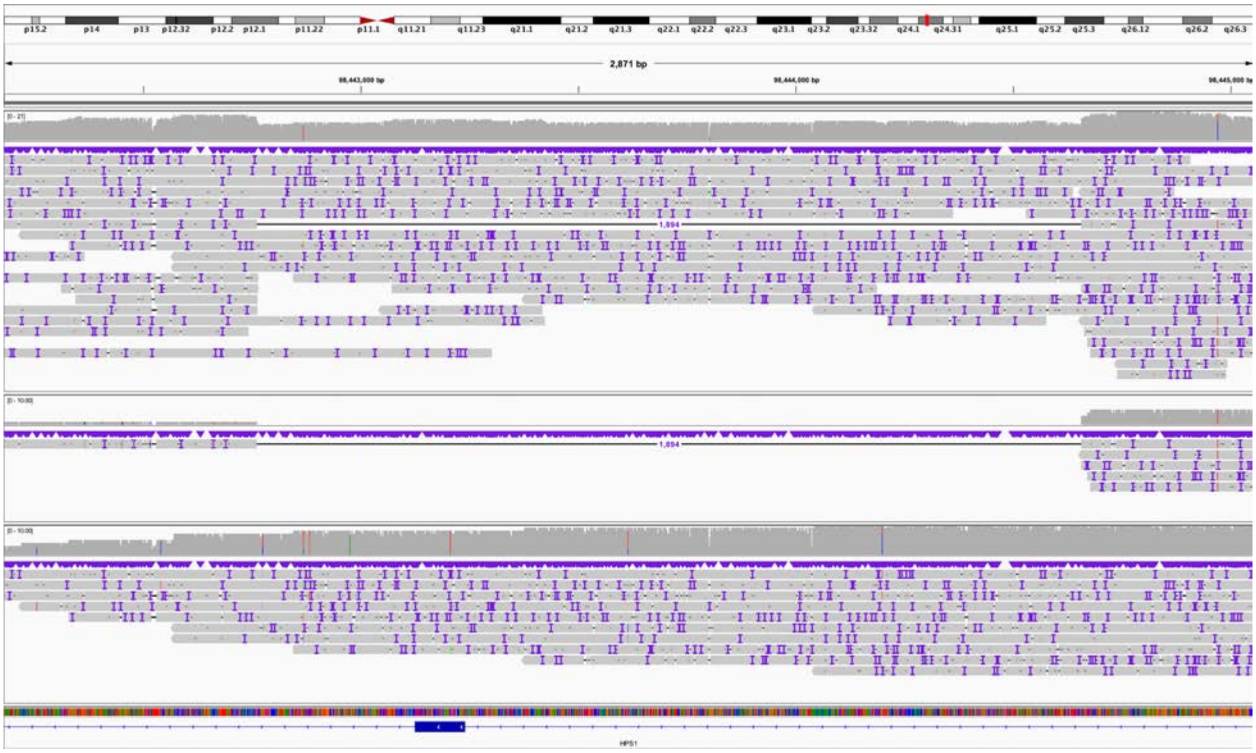


B. Histogram showing the number of individuals predicted to have an expansion of the AGAA repeat at the same position as the proband within the SSC collection. Position of the repeat expansion in the proband is shown in red, the length of the two haplotypes from the proband's mother is shown in red.

Figure S37. S013 (*HPS1*), IGV view of inherited variant and deletion identified by LRS.

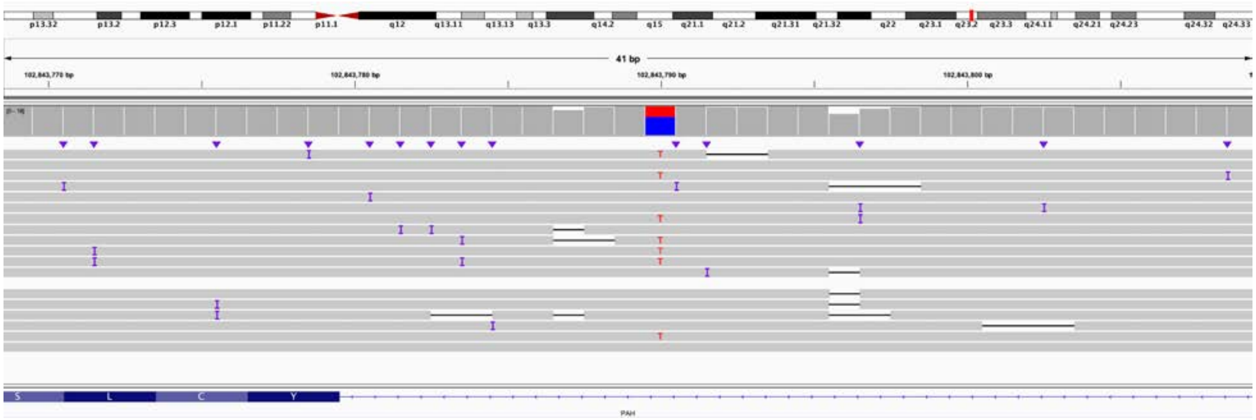


A. Screenshot of known paternally inherited pathogenic G>A variant. IGV view is of chr10:98,425,534-98,425,573. All reads are shown in the top track, haplotype 1 is the middle, and haplotype 2 is the bottom. Haplotype 2 is assumed to be the paternal track as the known paternally inherited G>A is in that track.

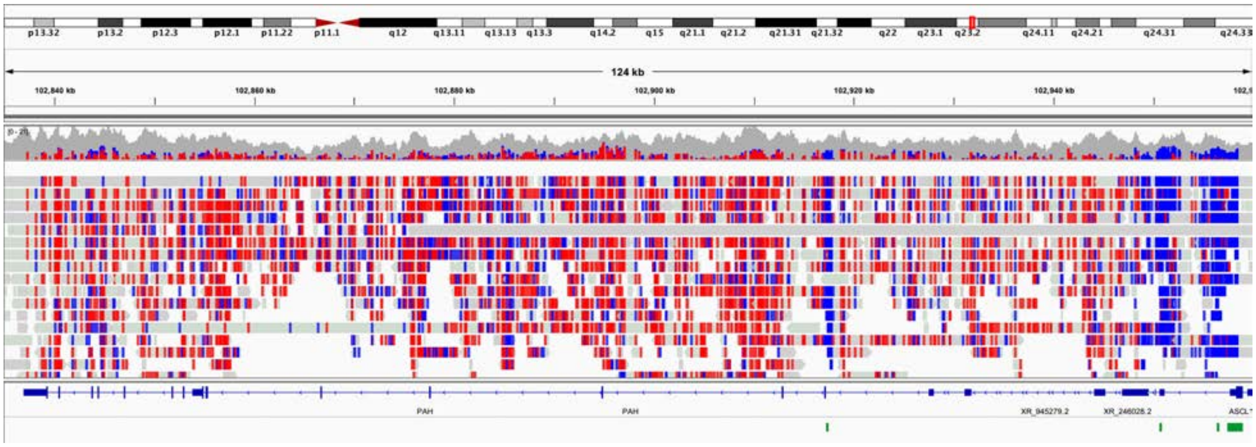


B. SV calling identified an approximately 1,900 bp deletion that included all of exon 3 on a different haplotype than the known G>A. Tracks are the same as in A. Screenshot from chr10:98,442,300-98,445,184.

Figure S38. S018 (*PAH*), known inherited splice variant identified, no second variant found.

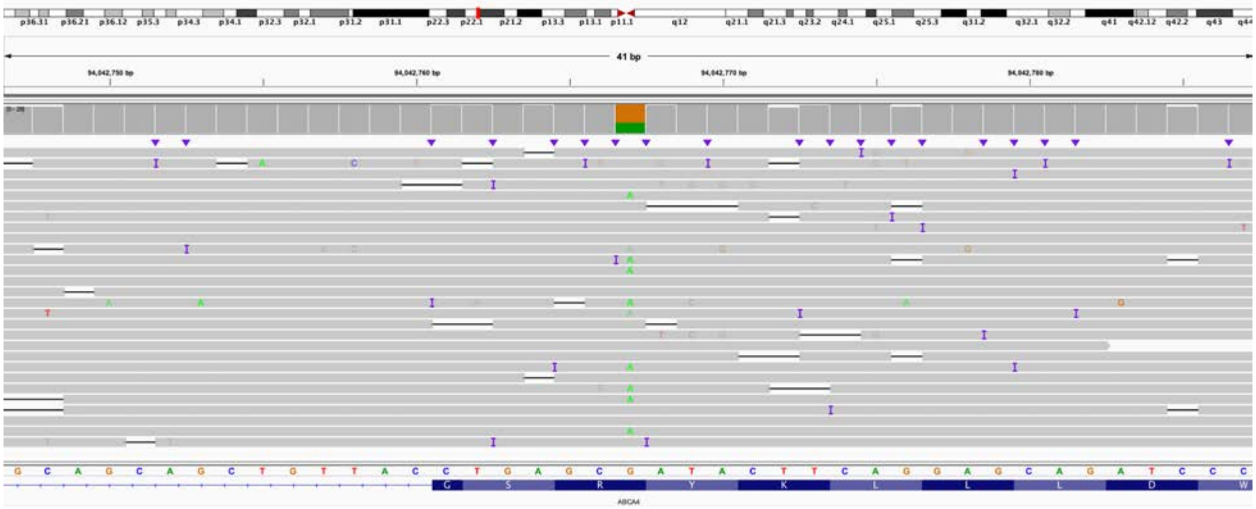


A. IGV view of previously known inherited pathogenic C>T splice variant. No second variant was found in this case. IGV view is of chr12:102,843,769-102,843,809.

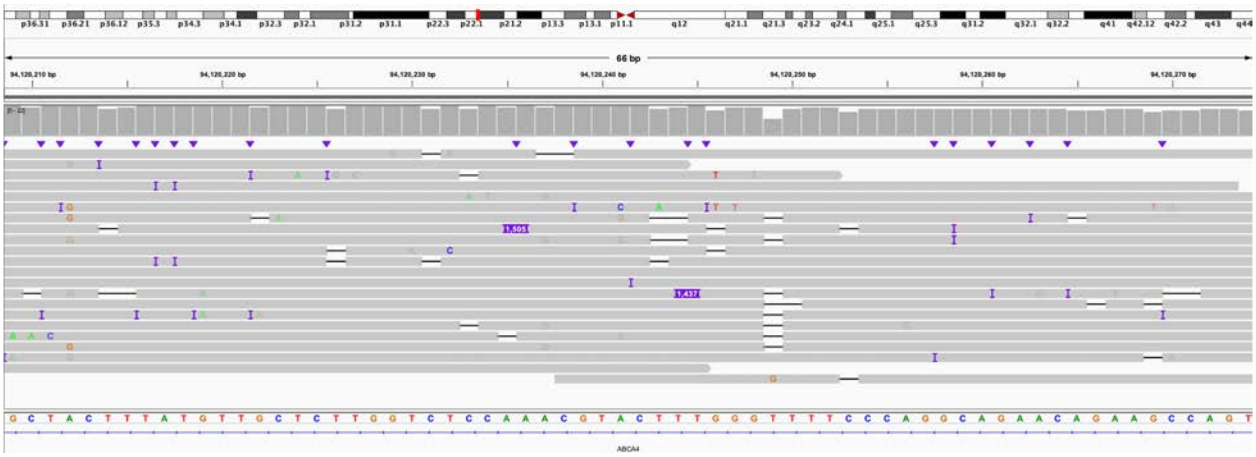


B. Bisulfite view of the entire *PAH* gene showing no hypermethylation near the 5' end of the gene (blue). Green blocks at the bottom of the image represent CpG islands. In this view, blue represents a CpG that is not methylated while red represents a methylated CpG.

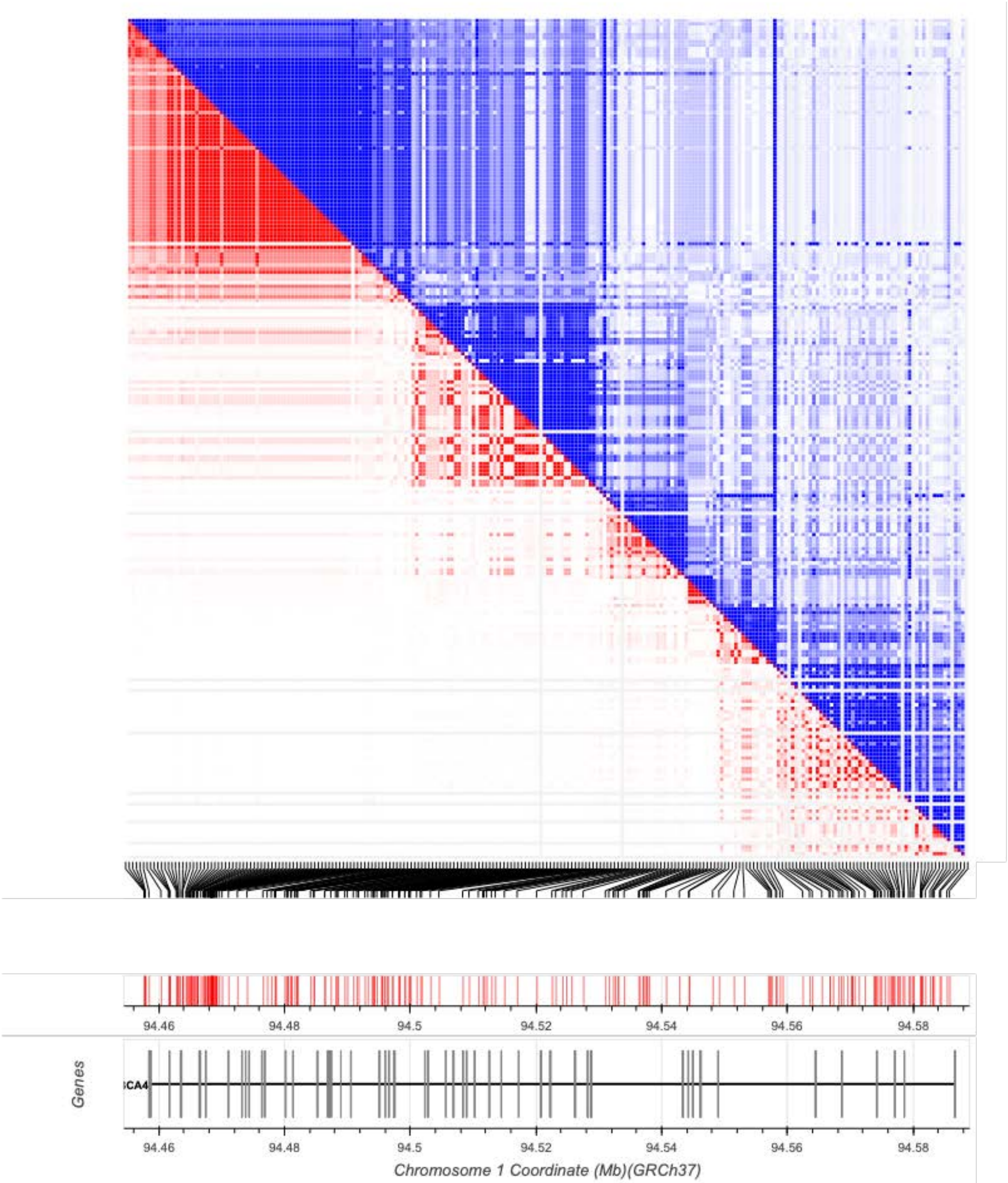
Figure S39. S025 (*ABCA4*), the previously known variant and a 1,500 bp insertion can be phased into different haplotypes.


41 bp
94,042,750 bp
94,042,760 bp
94,042,770 bp
94,042,780 bp
G C A G C A G C T G T T A C C T G A G C G A T A C T T C A G G A G C A G A T C C C
G S R Y K L L L D W
ABCA4

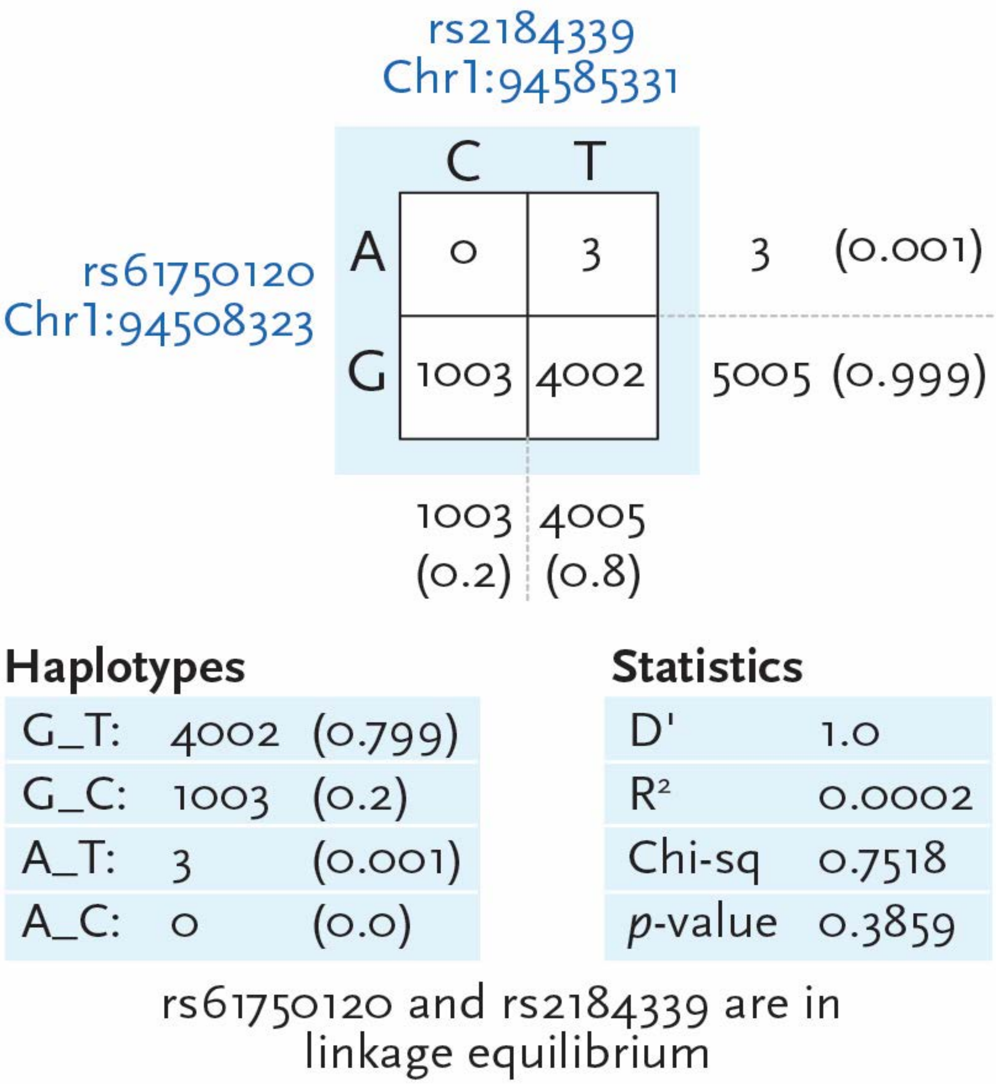
A. IGV screenshot of long reads showing the previously identified inherited pathogenic variant. IGV view is of chr1:94,042,747-94,042,787.


66 bp
94,120,210 bp
94,120,220 bp
94,120,230 bp
94,120,240 bp
94,120,250 bp
94,120,260 bp
94,120,270 bp
1,505
1,437
G C T A C T T T A T G T T G C T C T T G G T C T C C A A A C G T A C T T T G G G T T T T C C C A G G C A G A A C A G A A G C C A G T
ABCA4

B. IGV screenshot of long reads showing two reads that include the 1,500 bp insertion in intron 1 of *ABCA4*. Several reads on either side of the region include the insertion and are soft clipped. IGV view is of chr1:94,120,209-94,120,274.



C. The linkage disequilibrium matrix represents R^2 and D' values in the lower and upper diagonal of the heatmap, respectively. Higher intensity colors represent higher linkage disequilibrium. Using orthogonal Illumina WGS data from S025, we identified all SNVs overlapping *ABCA4* using GATK, followed by calculating all pairwise R^2 and D' values using the 1000 Genomes Project Phase III genotypes.



rs2184339
Chr1:94585331

rs61750120
Chr1:94508323

	C	T	
A	0	3	3 (0.001)
G	1003	4002	5005 (0.999)
	1003 (0.2)	4005 (0.8)	

Haplotypes

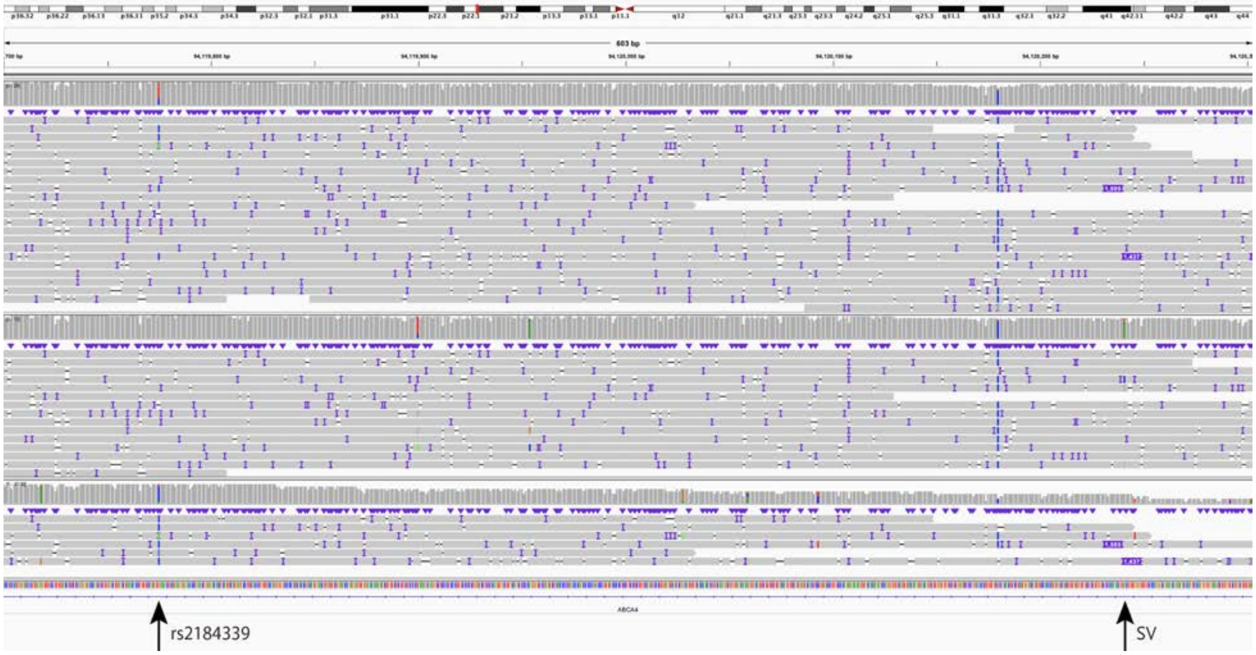
G_T:	4002	(0.799)
G_C:	1003	(0.2)
A_T:	3	(0.001)
A_C:	0	(0.0)

Statistics

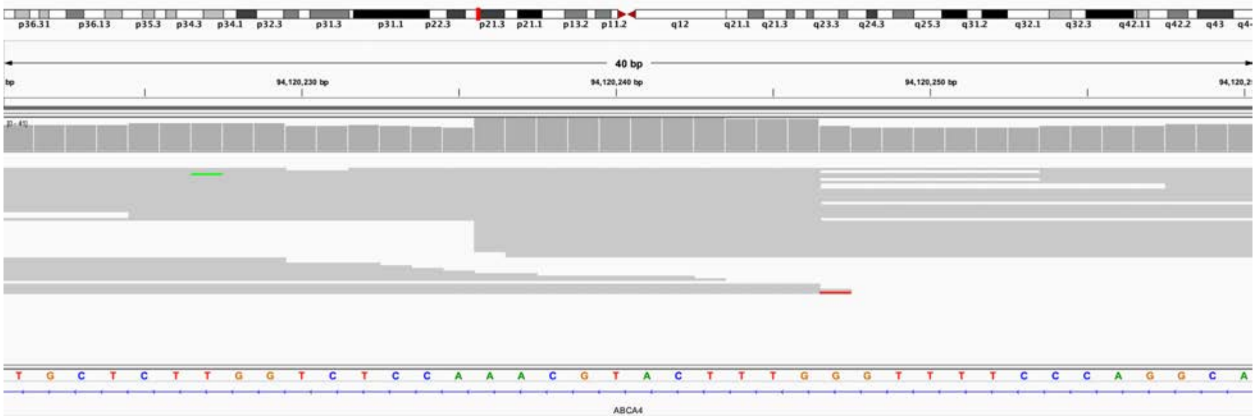
D'	1.0
R^2	0.0002
Chi-sq	0.7518
p-value	0.3859

rs61750120 and rs2184339 are in
linkage equilibrium

D. Phasing detail. The missense mutation on exon 22 corresponds to rs61750120 (G>A), while rs2184339 (T>C) has its alternative allele on the same allele as the 1.5 kbp insertion. Using the 1000 Genomes Project variation data (Machiela and Chanock 2015), we confirmed that the A allele of rs61750120 and the C allele of rs2184339 were never observed on the same haplotype (D'=1 and R^2=0.0001).

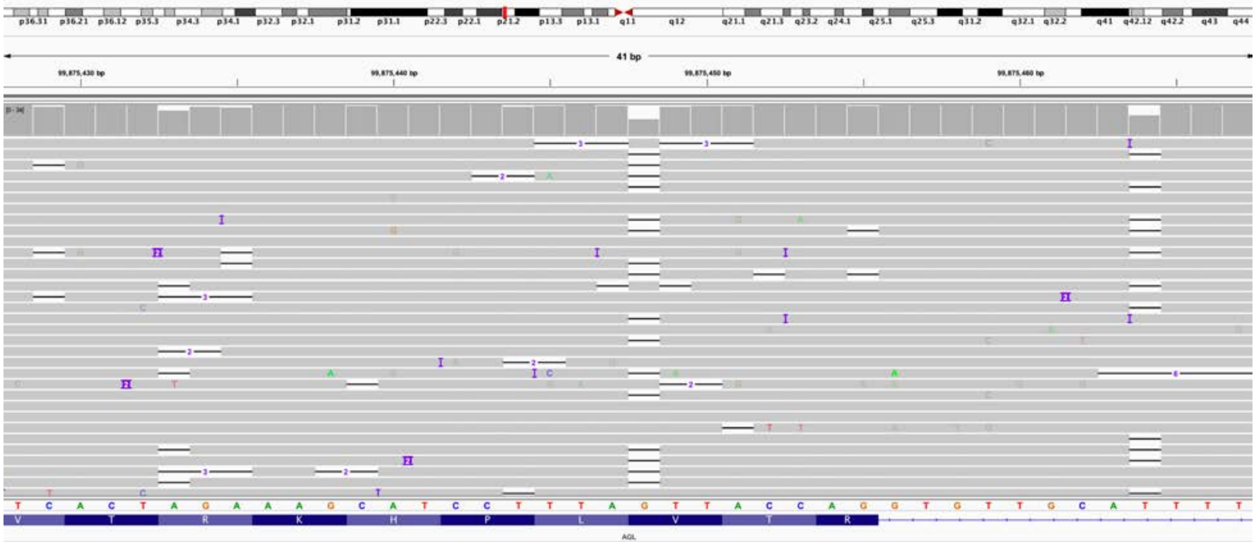


E. Phasing of long reads shows that rs2184339 (T>C) is present on reads with the 1,500 bp insertion. Most reads terminating at the insertion site are soft clipped. IGV view is of chr1:94,119,700-94,120,304.

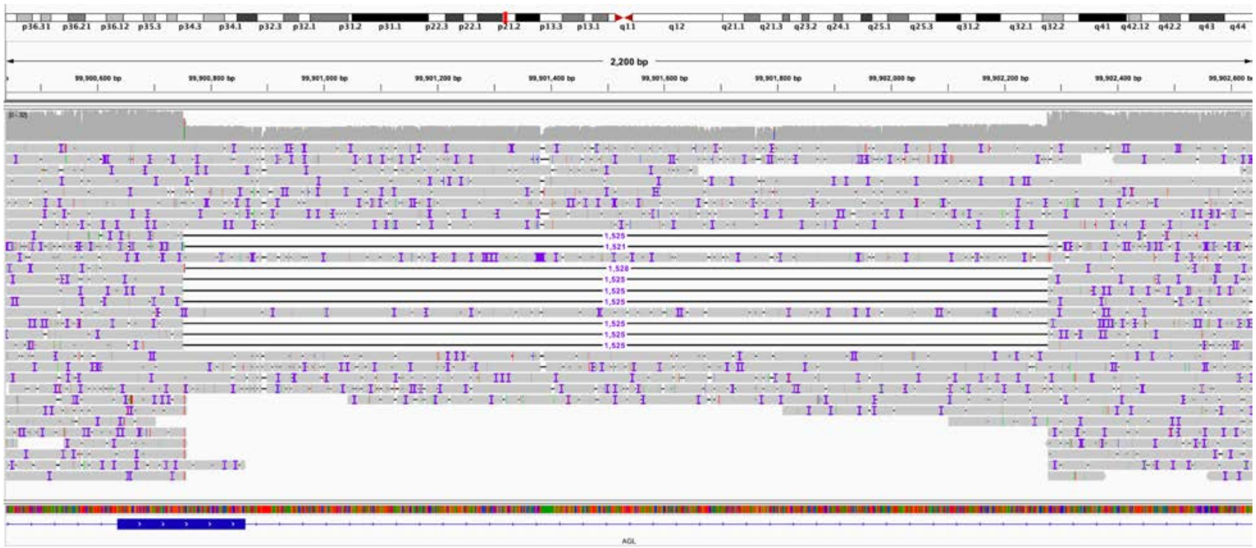


F. Analysis of short-read sequencing data reveals a 9 bp target site duplication at the position of the suspected insertion. IGV view is of chr1:94,120,221-94,120,260.

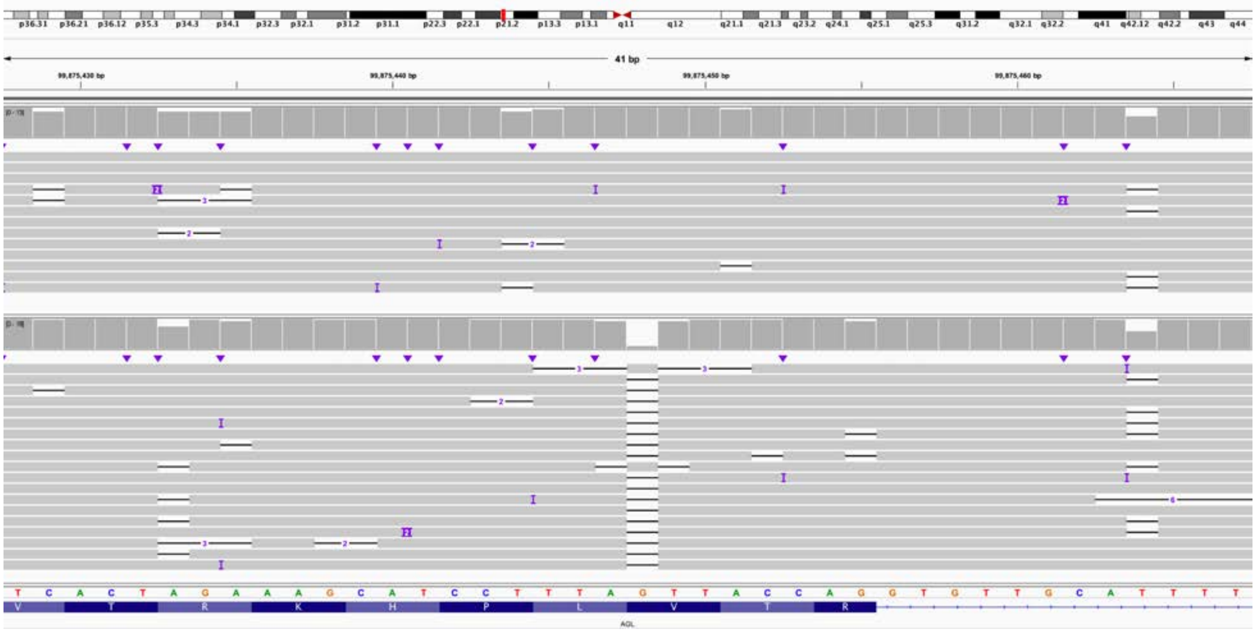
Figure S40. S047 (*AGL*), known single-nucleotide deletion, second hit is 1.5 kbp deletion.



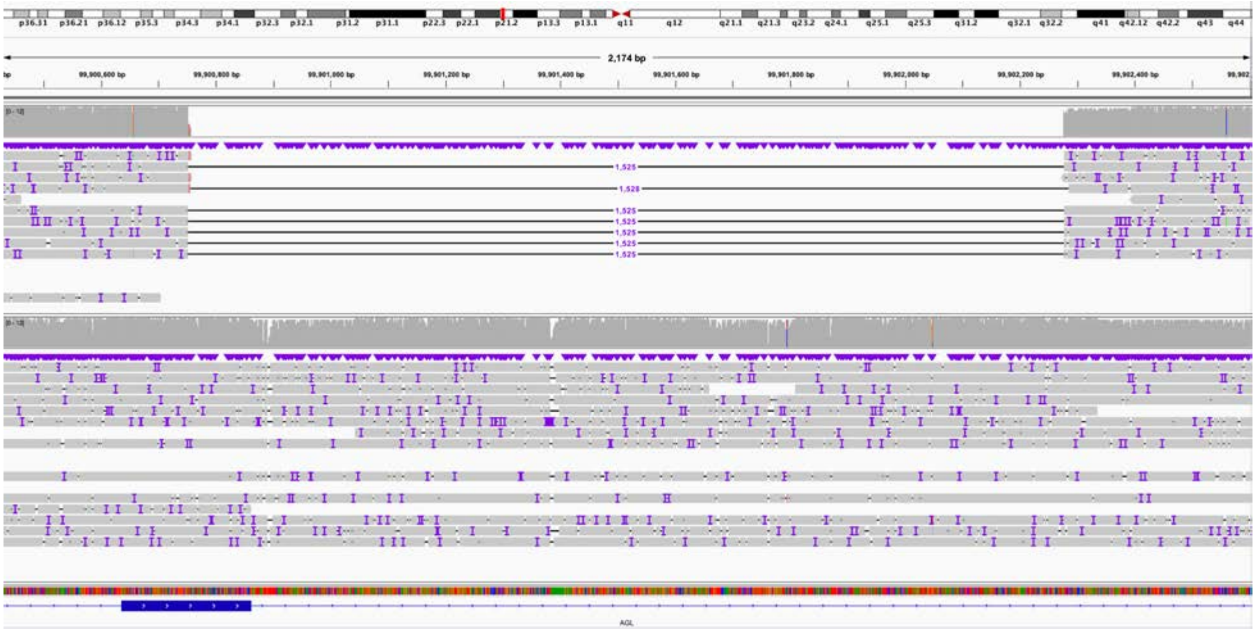
A. IGV screenshot of previously identified single-nucleotide deletion (view is chr1:99,875,428-99,875,467).



B. IGV screenshot of 1,525 bp deletion (coordinates are chr1:99,900,752-99,902,276) identified by Sniffles and SVIM (view is of chr1:99,900,438-99,902,636).

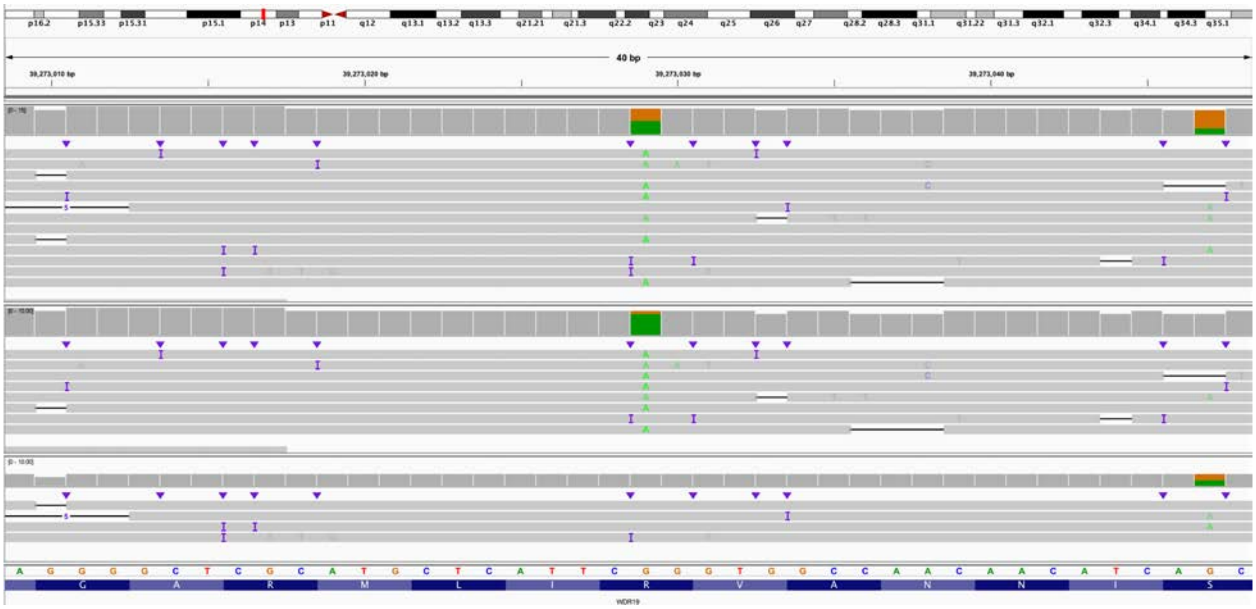

41 bp
99,875,430 bp
99,875,440 bp
99,875,450 bp
99,875,460 bp
T C A C T A G A A A G C A T C C T T T A G T T A C C A G G T G T T G C A T T T T
V T R K H P L V T R
AGL

C. Longshot phased the reads into two haplotypes (HP1 and HP2) and suggests that the single-nucleotide deletion is on a different haplotype than the 1,525 bp deletion. IGV view is chr1:99,875,428-99,875,467.



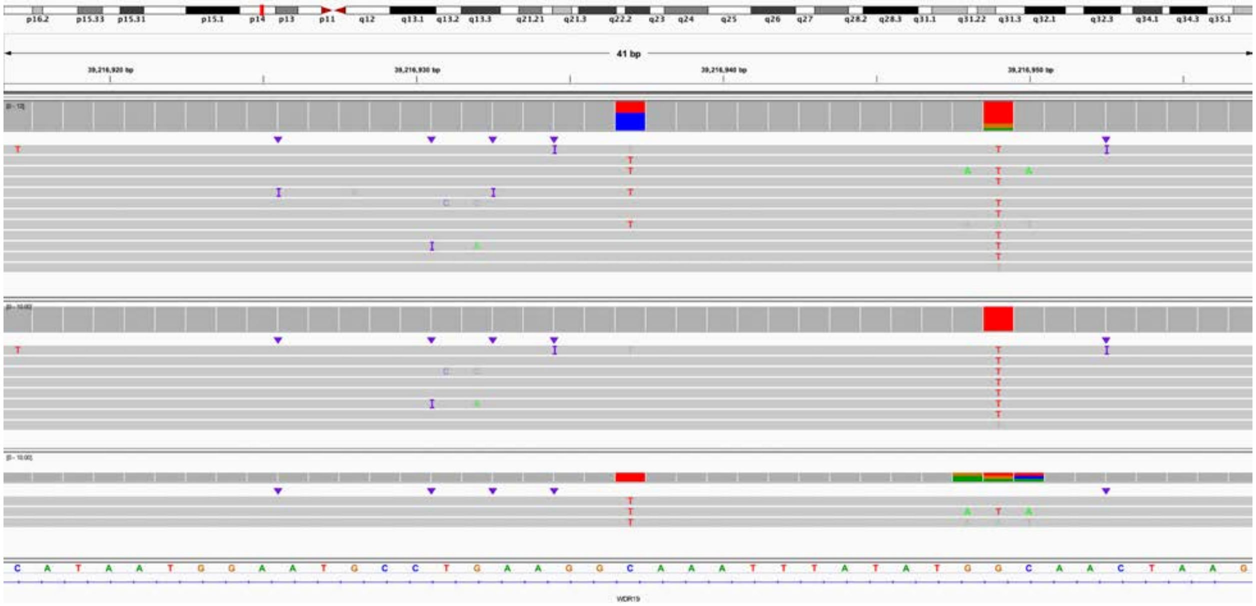
D. Longshot phased the reads into two haplotypes (HP1 and HP2) and suggests that the single-nucleotide deletion is on a different haplotype than the 1,525 bp deletion. IGV view is chr1:99,900,428-99,902,600.

Figure S41. S056 (*WDR19*), IGV views of inherited and splice variants identified by LRS.



p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q13.3 q21.21 q21.3 q22.2 q23 q24 q25 q26 q27 q28.2 q28.3 q31.1 q31.22 q31.3 q32.1 q32.3 q34.1 q34.3 q35.1
40 bp
39,273,010 bp
39,273,020 bp
39,273,030 bp
39,273,040 bp
A G G G G C T C G C A T G C T C A T T C G G G T G G C C A A C A A C A T C A G C
G A R M L I R V A N N I S
WDR19

A. IGV view of known pathogenic inherited G>A. Top panel is all reads, middle panel is haplotype 1, and bottom panel is haplotype 2. IGV view is of chr4:39,273,009-39,273,048.



p16.2 p15.33 p15.31 p15.1 p14 p13 p11 q12 q13.1 q13.2 q13.3 q21.21 q21.3 q22.2 q23 q24 q25 q26 q27 q28.2 q28.3 q31.1 q31.22 q31.3 q32.1 q32.3 q34.1 q34.3 q35.1
41 bp
39,216,920 bp
39,216,930 bp
39,216,940 bp
39,216,950 bp
C A T A A T G G A A T G C C T G A A G G C A A A T T T A T A T G G C A A C T A A G
WDR19

B. An intronic C>A variant in haplotype 2 is predicted to increase the likelihood this position acts as both a splice acceptor and donor. IGV view is of chr4:39,216,917-39,216,957.

Supplementary Tables

Table S1: Sample summary, DNA source, flow cells and libraries used per sample.

Type	Individual	DNA Source	Flow cells used	Libraries used	Summary of sample
Missing Variant (10)	S002	Blood	1	1	Single variant in *ALMS1*
	S003	Fibroblast	2	2	Single variant in *NPHP4*
	S004	Blood	3	3	Single variant in *VARS2*
	S008	Blood	2	3	No variant in *HPRT1*
	S009	Blood	3	4	No variant in *DMD*
	S013	Blood	1	2	Single variant in *HPS1*
	S018	Blood	1	1	Single variant in *PAH*
	S025	Blood	1	1	Single variant in *ABCA4*
	S047	Blood	1	1	Single variant in *AGL*
	S056	Blood	1	1	Single variant in *WDR19*
Phasing (2)	S071	Blood	1	1	Two variants in *METTL5*
	S086	Blood	2	2	Two variants in *KIAA1109*, one mosaic
Repeat Expansion (10)	S011	Saliva	1	1	Expansions of both *ATXN3* and *ATXN8OS*
	S039	Cell line	1	1	Expansion of *FMR1*
	S040	Cell line	1	1	Expansion of *FXN*
	S041	Cell line	1	1	Expansion of *FXN*
	04-01	Blood	1	1	Expansion and methylation of *XYLT1*
	06-01	Saliva	2	2	Expansion and methylation of *XYLT1*
	04-02	Fibroblasts	1	1	Mother of 04-01
	04-03	Saliva	1	1	Father of 04-01
	06-02	Saliva	1	1	Mother of 06-01
	06-03	Saliva	1	1	Father of 06-01
SV Case - Complex (8)	S014	Blood	1	1	Three noncontiguous deletions of chr6
	S020	Blood	2	3	Multiple deletions of chr4 and chr14, multiple translocations
	S021	Blood	2	2	Multiple mosaic deletions of chr8
	S022	Blood	1	1	Amplification of 4q with adjacent region of homozygosity
	S035	Blood	1	1	Two duplications
	S036	Blood	2	2	Four deletions on chr10, multiple translocations
	S082	Blood	1	1	Deletion of chr10, duplication on chr17
	S083	Blood	1	1	Deletion and duplication of chr4
SV Case - Simple (14)	BK144-03	Blood	1	1	22q13.3 deletion
	BK180-03	Blood	1	1	15q11-q13 duplication
	BK294-03	Blood	1	1	22q11.2 duplication
	BK364-03	Blood	1	1	1p36.11 duplication
	BK397-101	Blood	1	1	16p11.2 deletion
	BK430-103	Blood	1	1	16p11.2 duplication
	BK482-101	Blood	1	1	1q21.1 duplication
	BK487-101	Blood	1	1	1q21 deletion
	BK506-03	Blood	1	1	5p15.33 deletion
	S016	Blood	1	1	Tandem duplication in *CTNND2*
	S023	Blood	1	1	Mosaic ring 18
	S046	Blood	1	1	Unbalanced translocation between chr4 and chr15
	S060	Blood	1	1	Translocation between chr12 and chr17 in an individual with campomelic dysplasia
	S063	Fibroblast	1	1	Known SVA insertion in *BRCA1*

Table S2: Per sample sequencing targets, coverage, and average read length.

Individual	Target chr	Target start	Target end	Size of target region (bp)	Target, gene, or target	Average coverage of whole genome (x)	Average coverage of target region (x)	Average length of all reads (bp)	Average length of reads in target region (bp)
Simple SV cases									
BK144-03	22	17,500,000	50,818,000	33,318,000	22q13.3 del	3.5	14.7	539	2,133
BK180-03	15	20,000,000	35,000,000	15,000,000	15q11-q13 dup	3.6	61.5	524	4,289
BK294-03	22	17,500,000	50,818,000	33,318,000	22q11.2 dup	3.5	33.7	524	4,617
BK364-03	1	24,000,000	30,000,000	6,000,000	1p36.11 dup	2.0	21.2	427	3,692
BK397-101	16	26,500,000	35,300,000	8,800,000	16p11.2 del	0.7	12.5	865	6,230
BK430-103	16	25,000,000	35,000,000	10,000,000	16p11.2 dup	1.5	17.0	612	4,057
BK482-101	1	140,000,000	155,000,000	15,000,000	1q21.1 dup	2.2	22.8	417	2,258
BK487-101	1	140,000,000	155,000,000	15,000,000	1q21 del	1.6	36.9	515	6,883
BK506-03	5	1	8,000,000	7,999,999	5p15.33 del	2.7	15.8	443	3,814
S016	5	11,026,788	12,124,078	1,097,290	*CTNND2* dup	1.8	25.3	440	4,626
S023	18	1	5,000,000	4,999,999	8p	2.2	23.1	561	5,106
	18	45,000,000	80,373,285	35,373,285	8q	2.2	21.4	561	5,947
S046	4	180,000,000	190,214,555	10,214,555	translocation	2.0	9.8	436	2,757
	15	90,000,000	101,991,189	11,991,189	translocation	2.0	17.7	436	3,068
S060	12	45,000,000	65,000,000	20,000,000	translocation	2.0	27.2	579	6,574
	17	65,000,000	80,000,000	15,000,000	translocation	2.0	26.3	579	6,801
S063	17	42,000,000	44,000,000	2,000,000	SVA insertion	2.1	18.8	724	5,925
Repeat expansion cases									
S011	14	92,041,011	92,101,011	60,000	*ATXN3*	2.7	8.3	481	1,099
	13	70,109,384	70,169,384	60,000	*ATXN8OS*	2.7	9.2	481	1,197
S039	X	147,862,037	147,962,037	100,000	*FMR1*	2.1	10.0	450	3,772
S040	9	69,007,285	69,067,285	60,000	*FXN*	1.6	21.2	483	5,649
S041	9	69,007,285	69,067,285	60,000	*FXN*	0.9	15.5	484	8,134
04-01	16	16,500,000	18,000,000	1,500,000	*XYLT1*	2.5	18.2	474	5,170
06-01	16	16,500,000	18,000,000	1,500,000	*XYLT1*	2.3	8.0	415	2,121
04-02	16	16,500,000	18,000,000	1,500,000	*XYLT1*	0.9	11.5	511	5,815
04-03	16	16,500,000	18,000,000	1,500,000	*XYLT1*	2.3	4.7	894	2,308
06-02	16	16,500,000	18,000,000	1,500,000	*XYLT1*	0.4	3.0	645	3,276
06-03	16	16,500,000	18,000,000	1,500,000	*XYLT1*	1.1	2.2	470	2,036
Cases with complex copy number changes									
S014	6	150,000,000	165,000,000	15,000,000	-	2.7	13.3	420	2,105
S020	2	1	36,300,000	36,299,999	1st run	1.9	16.3	872	6,639
	4	65,000,000	78,000,000	13,000,000	1st run	1.9	14.4	872	6,556
	10	59,474,150	80,320,091	20,845,941	1st run	1.9	16.3	872	6,569
	14	20,881,587	24,829,792	3,948,205	1st run	1.9	12.1	872	6.549
	10	41700000	59,475,000	17,775,000	2nd run	0.6	9.2	785	7,158
	4	51800000	65,000,000	13,200,000	2nd run	0.6	7.3	785	8,057
	4	78000000	100,000,000	22,000,000	2nd run	0.6	7.2	785	8,072
	14	25900000	50,000,000	24,100,000	2nd run	0.6	7.4	785	8,239
S021	8	1	145,138,636	145,138,635	-	4.9	34.7	642	3,603
S022	4	162611564	190000000	27,388,436	-	2.0	31.2	634	8,158
	15	22,000,000	23,500,000	1,500,000	-	2.0	31.8	634	6,003
S035	8	130,000,000	145,138,636	15,138,636	-	2.7	38.7	538	5,712
	16	10,000,000	20,000,000	10,000,000	-	2.7	38.5	538	5,671
S036	6	114,000,000	124,000,000	10,000,000	1st run	2.6	35.7	602	6,443
	18	7,000,000	16,000,000	9,000,000	1st run	2.6	34.9	602	6,010
	10	20,000,000	38,000,000	18,000,000	1st run	2.6	33.4	602	6,487
	10	50,000,000	60,000,000	10,000,000	1st run	2.6	32.4	602	6,446
	5	1	46,000,000	45,999,999	2nd run	0.7	7.8	828	8,413
	6	124,000,000	170,805,979	46,805,979	2nd run	0.7	7.6	828	8,511
	18	1	7,000,000	6,999,999	2nd run	0.7	7.8	828	7,535
S082	10	107,000,000	114,000,000	7,000,000	-	1.7	9.3	663	3,003
	17	1	23,000,000	23,000,000	-	1.7	10.3	663	2,916
S083	4	1	49,000,000	49,000,000	-	2.2	34.3	628	6,407
Phasing cases									
S071	2	169,780,000	169,850,000	70,000	*METTL5*	0.7	9.6	487	5,937
S086	4	121,200,000	123,300,000	2,100,000	*KIAA1109*	2.8	29.4	480	4,723

Missing variant cases									
S002	2	73,200,000	73,800,000	600,000	*ALMS1*	1.4	17.0	608	6,780
S003	1	5,762,810	6,092,425	329,615	*NPHP4*	2.9	21.6	792	6,232
S004	6	30814208	31026459	212,251	*VARS2*	3.5	41.9	640	6,443
S008	X	134,360,165	134,600,668	240,503	*HPRT1*	2.5	7.2	1,151	7,160
S009	X	31,019,219	33,439,460	2,420,241	*DMD*	6.9	29.5	501	3,344
S013	10	98,316,193	98,546,963	230,770	*HPS1*	3.7	18.7	751	3,484
S018	12	102,736,889	103,058,441	321,552	*PAH*	2.1	11.4	459	2,102
S025	1	93,000,000	95,000,000	2,000,000	*ABCA4*	3.2	23.1	574	3,392
S047	1	98,000,000	102,000,000	4,000,000	*AGL*	1.6	23.5	491	5,753
S056	4	35,000,000	43,000,000	8,000,000	*WDR19*	1.4	22.1	537	8,282

Table S3: Overview of individuals with a known single structural variant.

Individual	Previously known result	Confirmation of known result	Additional information gained with T-LRS	Previously reported event	Events seen by LRS
BK144-03	22q13.3 deletion	Identified the known deletion.	Identified exact position of deletion breakpoints.	1 deletion	1 deletion
BK180-03	15q11-q13 duplication	Identified the known duplication.	None	1 duplication	1 duplication
BK294-03	22q11.2 duplication	Identified the known duplication.	None	1 duplication	1 duplication
BK364-03	1p36.11 duplication	Identified the known duplication.	Identified the exact position of the duplication breakpoints and found that the duplication is tandem.	1 duplication	1 duplication
BK397-101	16p11.2 deletion	Identified the known deletion.	None	1 deletion	1 deletion
BK430-103	16p11.2 duplication	Identified the known duplication.	None	1 duplication	1 duplication
BK482-101	1q21.1 duplication	Identified the known duplication.	None	1 duplication	1 duplication
BK487-101	1q21 deletion	Identified the known deletion.	None	1 deletion	1 deletion
BK506-03	5p15.33 deletion	Identified the known deletion.	Appears to be an unbalanced translocation between chr5 and chr12. Short-read data from the individual agrees and the translocation is not found in parental short-read data.	1 deletion	Translocation between chr5 and chr12
S016	Tandem duplication in *CTNND2*	Identified the known duplication and confirmed it is tandem.	None	1 duplication	1 duplication
S023	Mosaic ring 18 in 40% of cells.	See decreased coverage in regions reported on the array.	None.	2 mosaic deletions	2 mosaic deletions
S046	Unbalanced translocation between chr4 and chr15.	Identified the translocation via coverage.	Identified exact position of translocation breakpoints.	Translocation	Translocation
S060	Translocation between chr12 and chr17	Identified the translocation	Exact position of translocation breakpoints	Translocation	Translocation
S063	SVA insertion in *BRCA1*	Identified SVA insertion	None, this served as a control	SVA insertion	SVA insertion

Table S4: Overview of individuals with repeat expansions, including expected and observed repeat expansion sizes.

Individual	Gene(s)	Repeat	Previously reported repeat size	Repeat position	BP added to window for repeat assay	Number of reads spanning window	Average repeat lengths (% difference from expected if known)
S011	*ATXN3*	CAG	Clinically reported as 74 and 28 repeats	chr14:92071012-92071052	100	10	25 (11%), 77 (4%)
S011	*ATXN8OS*	CAG	Clinically reported as 80 and 25 repeats	chr13:70139385-70139429	25	8	15 (40%), 73 (9%)
S039	*FMR1*	CGG	Coriell sample 06897, reported as 477 repeats.	chrX:147911980-147912111	200	3	386 (19%)
S040	*FXN*	GAA	Coriell sample 16789, reported as 750 and 1000 repeats.	chr9:69,037,285-69,037,302	50	4	333 (56%), 1049 (5%)
S041	*FXN*	GAA	Coriell sample 15850, reported as 650 and 1030 repeats.	chr9:69,037,285-69,037,302	50	9	647 (1%), 958 (7%)
04-01	*XYLT1*	GGC	Previously reported expansion.	chr16:17470921-17470922	500	4	758
04-02	*XYLT1*	GGC	Mother of 04-01, reported as one wild-type and one premutation allele.	chr16:17470921-17470922	50	16	97, 221
04-03	*XYLT1*	GGC	Father of 04-01, reported as two wild-type alleles.	chr16:17470921-17470922	50	2	71
06-01	*XYLT1*	GGC	Previously reported expansion.	chr16:17470921-17470922	500	5	224
06-02	*XYLT1*	GGC	Mother of 06-01, reported as one wild-type and one premutation allele	chr16:17470921-17470922	50	1	81
06-03	*XYLT1*	GGC	Father of 06-01, reported as two wild-type alleles	chr16:17470921-17470922	50	1	78

Table S5: Per-read details of individuals with *ATXN3*, *ATXN8OS*, *FMR1*, and *FXN* repeat expansions.

Read ID	Repeat Length (bp)	Repeats (Estimate)	Repeat Group	Average Number of Repeats
S011 - ATXN3				
c4b0552d-118e-4d6b-9411-09b0861262fe	55	18	1	25
616cf7f8-a13f-429a-b0c5-825f1a15da3d	72	24	1	
e133bc59-4c37-4c5c-8ce9-b206180058eb	77	26	1	
eadb0392-728a-4437-8e5e-c2814638894f	79	26	1	
2f107739-f4a6-40f0-b52b-252132f992f3	85	28	1	
137c7831-3cb4-4f7c-99b6-196a969702dc	218	73	2	77
72ed9321-3c6b-4cab-b508-83d6efc1efc0	230	77	2	
b6a6b164-c767-4514-b723-4a6643a0f3f5	231	77	2	
e6c40588-04c8-4a17-b740-fc1990995267	232	77	2	
0fa6df78-d10e-42f8-9e1e-6db5a907d858	250	83	2	
S011 - ATXN8OS				
9712bb2f-b1fb-40db-9e05-6f708b6443e4	41	14	1	15
3e0a7cc1-ee7f-47ff-97f4-9516cbebb21e	43	14	1	
d5eea8a6-f1d5-459f-b865-bff3c35a26b3	44	15	1	
c03fb6ab-a4a1-4518-b946-ad650f578e25	46	15	1	
70fec058-32ab-4270-aaa6-1984af7b1402	197	66	2	73
4f49b32d-139c-4beb-b831-9c4a5b7a19be	206	69	2	
a7304699-9d25-4843-b587-cb84de848b84	226	75	2	
489bde02-eae6-4270-a7db-1e488b61ccfd	245	82	2	
S039 - FMR1				
5fac5744-2c5c-4211-88c8-5b9ee3a20752	759	253	1	386
b215ab66-d863-46d3-878e-69d3d855a302	1448	483	1	
ef287abe-e2a4-4b0e-8121-1de041a53f77	1268	423	1	
S040 - FXN				
44970178-6c34-48e8-afd9-f529d6bda0e2	985	328	1	333
1e915a81-c955-4c88-8946-b3e73d371b64	1004	335	1	
afddceb3-0802-404d-a61c-24ecdba887db	1007	336	1	
a66018cb-a390-4cb6-89f4-b9c0073e68f8	3147	1049	2	1049
S041 - FXN				
a93466ef-353b-403d-a2ec-d5439867b826	1788	596	1	647
b1ac947d-e98d-41f4-b5ca-d464e3668388	1939	646	1	
8164bc3d-6cae-4928-a975-e695c48bc757	1963	654	1	
0bfda482-97a2-42c8-ad67-798ebf2bd5b8	2079	693	1	
bf8f85f8-f878-4f38-aefc-b7aa0009b8ae	2415	805	2	958
89100931-0b2b-4938-a882-f7c81652858a	2895	965	2	
20ad64ba-a1b7-4371-93f8-e3c425c2c008	2996	999	2	
d13da846-f4f3-45ba-806a-0f77f50bb71a	3000	1000	2	
c0a0f7ad-c30a-4e40-ab51-fe00b1ab699a	3058	1019	2	

Table S6: Per-read details of *XYLT1* repeat expansions.

Read ID	Repeat Length (bp)	Insert between *Kpn*I cut sites (bp)	Repeats	Repeat Group	Average Number of Repeats
04-01 - XYLT1					
71ed6a24-3280-465a-8114-338c0eb6308d	1846	4435	615	1	758
c522d590-4165-47ac-9d4e-fb98f391bc13	2861	5450	954	1	
beb4b5c7-293c-4f69-b275-d393c6a3ae34	1591	4180	530	1	
a09207e2-a9cc-4cb6-9e12-f9fb56214399	2800	5389	933	1	
04-02 - XYLT1					
b2bebdb7-8e03-4b86-833f-445c1c2cd1d0	167	2756	56	1	97
7c629eb4-81f6-4d27-988b-b827b81419a2	282	2871	94	1	
775ca9d2-16b7-43f2-8855-6ef25cfd26a0	283	2872	94	1	
0cce5876-2f35-44f8-835f-867f08e4e571	288	2877	96	1	
879df89f-3293-4bec-93ec-073c985c3a18	288	2877	96	1	
83d4020e-16f1-4b94-95d8-1bf48fca6de6	295	2884	98	1	
970f3e26-1807-42d3-b15e-4c85a25b6ab4	433	3022	144	1	
0edbaf14-3d7e-46a2-a03d-268c784560f3	603	3192	201	2	221
b09a91bf-aee4-4fc0-a7bf-51d7cb1ad126	615	3204	205	2	
40767c26-5d0c-467d-bdcf-c1a6346d4229	641	3230	214	2	
f85efe4f-2d32-4a6d-a217-165c6dbdc661	645	3234	215	2	
b0a508ce-069d-43ac-865e-7b7cd900eb70	653	3242	218	2	
5c6ac009-694d-4f86-9511-8708f952a81c	688	3277	229	2	
4e315aca-60a8-4e61-8cd6-c08f129d701a	693	3282	231	2	
227e44ce-2051-4169-9b25-17cfdb50a4c5	716	3305	239	2	
011e133a-ea64-43fd-a4f4-0c2cfc8fe56e	725	3314	242	2	
04-03 - XYLT1					
bba6d56d-5293-4c5b-bd03-327dc12f7ece	218	2807	73	1	71
370138b9-090a-4fb0-80bc-01723888a41b	209	2798	70	1	
06-01 - XYLT1					
c66d7ab6-2ec4-417d-9445-d1d68ef3cf8e	668	3257	223	1	224
7db4e75c-c2f7-4029-9c6e-a8e06ff269a6	720	3309	240	1	
210471f7-0e15-436e-862f-85557ad3d4ff	734	3323	245	1	
8f3c0400-f5dc-47e2-bc69-611a477181bf	622	3211	207	1	
5e39f080-90c9-4f3d-ba8b-bed01bef0caa	616	3205	205	1	
06-02 - XYLT1					
236a7c09-2c88-4ee8-b2b3-fda20bc625c1	243	2832	81	1	81
06-03 - XYLT1					
9e7e1ed4-ebc2-4b30-a8ea-4c561d87e793	235	2824	78	1	78

Table S7: Summary of individuals with complex SVs, including previously known events and new events identified by T-LRS.

Individual	Previously known result	Confirmation of known result	Additional information gained with T-LRS	Total previously known events	Total new events
S014	Three noncontiguous deletions of chromosome 6.	Identified the three deletions seen on array	Identified two additional deletions and one rearrangement, and found no pathogenic or likely pathogenic variants in deleted regions.	3	3
S020	Two deletions of 4q and one of 14q identified on the array. Karyotype revealed translocation between chromosomes 2, 4, 10, and 14.	Identified the three deletions seen on array and several translocations between chromosomes 2, 4, 10, and 14.	Identified one additional deletion on chromosome 4 and two on chromosome 10. Found the exact position of all translocation breakpoints, revealing two that bisected genes with AD phenotypes. Identified additional rearrangement breakpoints within chromosome 10 and chromosome 14 not involved in a deletion or translocation.	7	22
S021	Terminal 3.2 Mbp loss of 8p23.3 to p23.2 with a copy state of 1. Adjacent interstitial 3.7 Mbp mosaic loss of 8p23.2 to p23.1 with a copy state between 1 and 2. Terminal 50 Mbp mosaic gain of 8q22.1 to q24.3 with a copy state between 2 and 3.	Confirmed mosaic loss and gain seen on array.	Find that a common 400 kbp duplication on 8p not reported on the array appears to be inverted and attached to the chromosome 8 not carrying a deletion.	3	1
S022	Focal amplification of 4q (copy state 4 or greater) with adjacent region of homozygosity and 15q11.2 duplication.	See the amplification of 4q32, homozygosity of 4q, and duplication in 15q11.2.	Determined the structure of the tandem amplification on chr 4 and confirmed 5 copies of the amplification (6 with the wild-type chromosome). Do not see breakpoints of the 15q11.2 duplication, but do see the duplication. No methylation differences or pathogenic variants in the homozygous region.	2	0
S035	Duplications of 8q24 and 16p13.11	Both duplications observed.	Found that 8q duplication is tandem and identified exact breakpoints. Could not identify the exact position of the 16p13.11 duplication that is flanked by segmental duplications.	2	0
S036	Four deletions on chr 10, karyotype found translocation between chr 18 and 6 and pericentric inversion of chr10.	Identified four deletions and found the t(6;18).	Found additional translocations between chromosome 10 and chromosome 5 as well as additional rearrangements on chromosomes 6, 10, and 18.	7	13
S082	Single pathogenic deletion on chr 10. One complex likely pathogenic CNV on chr 17 with two deletions and three duplications.	Identified the deletion on chr 10. Identified the two deletions and three duplications on chr 17.	Determined that the duplications are rearranged and inverted.	6	2
S083	Deletion of distal arm of 4p, duplication of 4p proximal to the deletion.	Identified the deletion and duplication.	Determined the exact position of the deletion and duplication breakpoints.	2	0

Table S8: Known and new events observed for individuals with known complex SVs.

	Known				Observed				New Events			
Individual	Deletion	Duplication	Translocation	Rearrangement	Deletion	Duplication	Translocation	Rearrangement	Deletion	Duplication	Translocation	Rearrangement
S014	3	0	0	0	5	0	0	1	2	0	0	1
S020	3	0	4	0	5	0	11	13	2	0	7	13
S021	2	1	0	0	2	1	0	1	0	0	0	1
S022	0	2	0	0	0	2	0	0	0	0	0	0
S035	0	2	0	0	0	2	0	0	0	0	0	0
S036	4	0	2	1	6	0	8	6	2	0	6	5
S082	3	3	0	0	3	3	0	0	0	0	0	2
S083	1	1	0	0	1	1	0	0	0	0	0	0

Table S9: Details of focal amplification of 4q in individual S022.

Region	Size (bp)	Copy number estimate	Coverage	Increase or decrease in coverage from prior
chr4:160,000,000-160,143,267	143,267	2	31.3	
chr4:160,143,268-160,165,900	22,632	3	43.9	12.6
chr4:160,165,901-160,248,169	82,268	4	62.7	18.8
chr4:160,248,170-160,450,953	202,783	5	76.0	13.3
chr4:160,450,954-162,685,228	2,234,274	6	93.8	17.8
chr4:162,685,229-162,689,149	3,920	4	57.6	-36.2
chr4:162,689,150-162,700,000	110,850	2	30.6	-27.0

Table S10: Rearrangement numbers and sizes for individual S014.

Chr	Segment	Start	End	Size (bp)	Source/Type
6	A	1	153,854,905	153,854,904	-
6	B	153,854,906	154,095,842	240,936	Deletion
6	C	154,095,843	154,587,147	491,304	Inversion
6	D	154,587,148	154,588,799	1,651	Deletion
6	E	154,588,800	154,661,283	72,483	-
6	F	154,661,284	156,456,517	1,795,233	Deletion
6	G	156,456,518	156,460,113	3,595	Inversion
6	H	156,460,118	156,464,495	4,377	-
6	I	156,464,502	157,015,124	550,622	-
6	J	157,015,125	158,766,820	1,751,695	Deletion
6	K	158,766,821	164,500,069	5,733,248	-
6	L	164,500,070	164,654,921	154,851	Deletion
6	M	164,654,922	170,805,979	6,151,057	-

Table S11: Genes impacted by S014 breakpoints.

Chr	Breakpoint	Gene affected by breakpoint	Previously known based on clinical testing?	OMIM phenotype
6	B/C	*OPRM1*	Y	none
6	I/J	*ARID1B*	Y	Autosomal dominant: Coffin-Siris syndrome I
6	K/L	*EZR*	Y	none

Table S12: Rearrangement numbers and sizes for individual S020.

Chr	Segment	Start	End	Size (bp)	Source/Type	Validated with HiFi?
2	A	1	7,777,990	7,777,989	Derivative 14	Yes
2	B	7,777,996	16,235,712	8,457,716	Derivative 4	Yes
2	C	16,235,711	242,193,529	225,957,818	Derivative 2	Yes
4	A	1	67,572,970	67,572,969	Derivative 4	Yes
4	B	67,572,971	70,140,499	2,567,528	Deletion	Yes
4	C	70,140,500	73,061,492	2,920,992	Derivative 14	Yes
4	D	73,061,493	73,129,559	68,066	Deletion	Yes
4	E	73,129,560	73,446,127	316,567	Derivative 10	Yes
4	F	73,446,129	73,771,830	325,701	Derivative 10	Yes
4	G	73,771,831	74,581,615	809,784	Deletion	Yes
4	H	74,581,616	74,794,351	212,735	Derivative 14	Yes
4	I	74,794,352	74,831,249	36,897	Derivative 14	Yes
4	J	74,831,250	75,039,118	207,868	Derivative 4	Yes
4	K	75,039,137	77,012,176	1,973,039	Derivative 4	Yes
4	L	77,012,175	79,626,721	2,614,546	Derivative 14	Yes
4	M	79,626,720	81,632,570	2,005,850	Derivative 4	Yes
4	N	81,632,570	190,214,555	108,581,985	Derivative 2	Yes
10	A	1	50,382,094	50,382,093	Derivative 10	Yes
10	B	50,382,092	50,563,505	181,413	Derivative 10	Yes
10	C	50,563,501	50,649,777	86,276	Derivative 10	Yes
10	D	50,694,777	51,260,021	565,244	Derivative 10	Yes
10	E	51,260,022	51,840,878	580,856	Deletion	Yes
10	F	51,840,879	53,548,486	1,707,607	Derivative 10	Yes
10	G	53,548,487	53,563,514	15,027	Derivative 14	Yes
10	H	53,563,515	53,584,434	20,919	Deletion	Yes
10	I	53,584,435	56,035,820	2,451,385	Derivative 4	Yes
10	J	56,035,823	65,519,829	9,484,006	Derivative 4	Yes
10	K	65,519,828	67,462,235	1,942,407	Derivative 10	Yes
10	L	67,462,259	74,772,273	7,310,014	Derivative 10	Yes
10	M	74,772,269	76,402,382	1,630,113	Derivative 4	Yes
10	N	76,402,383	133,797,422	57,395,039	Derivative 4	Yes
14	A	1	20,934,810	20,934,809	Derivative 14	Yes
14	B	20,934,811	22,881,931	1,947,120	Deletion	Yes
14	C	22,881,932	107,043,718	84,161,786	Derivative 10	Yes
			Total Size	334,539,803	Derivative 2	
			Total Size	151,177,985	Derivative 4	
			Total Size	146,979,108	Derivative 10	
			Total Size	34,512,995	Derivative 14	

Table S13: Genes impacted by S020 breakpoints.

Chr	Breakpoint	Gene affected by breakpoint	Previously known based on clinical testing?	OMIM phenotype
4	A/B	*STAP1*	Y	none
4	B/C	*CSN1S2BP*	N	none
4	C/D	*COX18*	N	none
4	D/E	*ANKRD17*	N	none
4	E/F	*AFP*	N	Autosomal dominant: hereditary persistence of alpha-fetoprotein; autosomal recessive: alpha-fetoprotein deficiency
4	H/I	*BTC*	N	none
4	J/K	*PARM1*	N	none
4	K/L	*SEPTIN11*	N	none
10	A/B, B/C	*SGMS1*	N	none
10	D/E, E/F	*PRKG1*	N	Autosomal dominant: aortic aneurysm, familial thoracic
10	K/L	*CTNNA3*	N	Autosomal dominant: arrhythmogenic right ventricular dysplasia, familial
10	M/N	*LRMDA*	N	Autosomal recessive: Albinism, oculocutaneous, type VII

Table S14: Rearrangement numbers and sizes for individual S036.

Chr	Segment	Start	End	Size (bp)	Source/Type
5	A	1	22,881,851	22,881,850	Derivative 5
5	B	22,881,853	28,650,540	5,768,687	Derivative 5
5	C	28,650,540	28,826,587	176,047	Derivative 10
5	D	28,826,592	181,538,259	152,711,667	Derivative 5
6	A	1	123,005,546	123,005,545	Derivative 6
6	B	123,005,550	124,505,324	1,499,774	Derivative 18
6	C	124,505,731	126,657,919	2,152,188	Derivative 18
6	D	126,657,920	126,680,039	22,119	Deletion
6	E	126,680,040	170,805,979	44,125,939	Derivative 18
10	A	1	24,286,112	24,286,111	Derivative 10
10	B	24,286,113	24,725,849	439,736	Deletion
10	C	24,725,850	25,278,388	552,538	Derivative 5
10	D	25,278,408	34,622,799	9,344,391	Derivative 10
10	E	34,622,800	35,972,590	1,349,790	Deletion
10	F	35,972,591	49,775,899	13,803,308	Derivative 10
10	G	49,775,901	54,672,866	4,896,965	Derivative 5
10	H	54,672,867	55,305,313	632,446	Deletion
10	I	55,305,314	55,632,879	327,565	Derivative 5
10	J	55,632,880	56,736,281	1,103,401	Deletion
10	K	56,736,282	133,797,422	77,061,140	Derivative 10
18	A	1	6,275,133	6,275,132	Derivative 6
18	B	6,275,134	6,280,055	4,921	Deletion
18	C	6,280,056	7,279,970	999,914	Derivative 6
18	D	7,279,968	80,373,285	73,093,317	Derivative 18
			Total Length	187,139,272	Derivative 5
			Total Length	130,280,591	Derivative 6
			Total Length	124,670,997	Derivative 10
			Total Length	120,871,218	Derivative 18

Table S15: Genes impacted by S036 breakpoints.

Chr	Breakpoint	Gene affected by breakpoint	Previously known based on clinical testing?	OMIM phenotype
6	A/B	*CLVS2*	N	none
6	B/C	*NKAIN2*	N	none
10	A/B	*KIAA1217*	N	none
10	C/D	*GPR158*	N	none
10	D/E	*PARD3*	Y	none
10	F/G, G/H, H/I	*PCDH15*	Y	Autosomal dominant and autosomal recessive: Usher syndrome, type 1D/F digenic
18	A/B, C/D	*L3MBTL4*	N	none

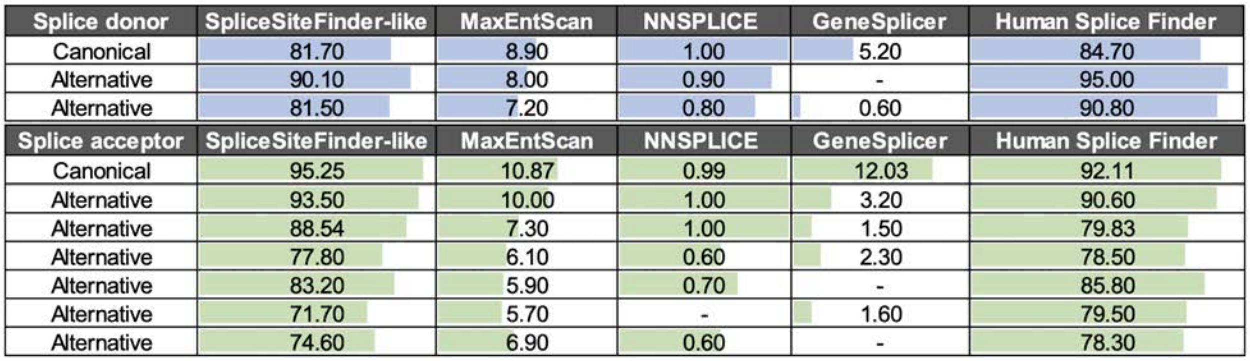
Table S16: Summary of individuals with missing variants.

Individual	Clinical workup	Confirmation of inherited variant	Variant found by T-LRS	Confirmation or supporting findings
S002	CMA normal, exome with paternally inherited stop in *ALMS1*, consistent with suspected diagnosis of Alström syndrome. Deletion/duplication analysis of *ALMS1* negative.	Confirmed the known paternally inherited variant.	*Alu* insertion in exon 20	Clinically confirmed
S003	CMA normal, exome with paternally inherited stop in *NPHP4*, consistent with suspected diagnosis of NPH. Deletion/duplication analysis of *NPHP4* negative.	Confirmed the known paternally inherited variant.	Intronic splice variant	Confirmed by qPCR
S004	CMA normal, exome identified single paternally inherited pathogenic variant in *VARS2*, deletion/duplication analysis of *VARS2* negative.	Confirmed the known paternally inherited pathogenic p.A420T.	No second hit found.	n/a
S008	Biochemical diagnosis of Lesch-Nyhan based on enzyme analysis. CMA normal, sequencing and deletion/duplication analysis of *HPRT1* negative.	No pathogenic SNVs or copy number changes observed, consistent w/clinical testing.	Identified an inversion within *HPRT1* and confirmed with PCR.	Clinically confirmed
S009	Suspected diagnosis of Duchenne muscular dystrophy. Sequencing and deletion/duplication analysis of *DMD* negative. Muscle biopsy consistent with diagnosis. The proband's maternal uncle died of muscular dystrophy.	No pathogenic SNVs or copy number changes observed, consistent w/clinical testing.	Identified a candidate AGAA repeat expansion in intron 16.	Mom is heterozygous, unaffected brother has wild-type allele
S013	CMA normal, exome with paternally inherited stop in *HPS1*, consistent with suspected diagnosis of Hermansky-Pudlak syndrome. No deletion/duplication analysis done.	Confirmed the known paternally inherited variant.	Identified 1,900 bp deletion that removes exon 3.	Clinically confirmed with exon-level array
S018	Elevated phenylalanine, exome with single pathogenic variant in *PAH*.	Confirmed the known inherited variant.	No second hit found.	No second hit found with short-read sequencing
S025	Diagnosed with Stargardt disease by clinical retinal exam. Exome with single pathogenic variant in *ABCA4*. No second hit seen on research WGS.	Confirmed the known pathogenic variant.	~1,500 bp transposable element insertion in the first intron.	Insertion is present in Illumina short-read data.
S047	Diagnosed with glycogen storage disease III. Exome revealed single-nucleotide deletion resulting in frameshift. Research SR-WGS was negative.	Confirmed the known pathogenic variant.	1,525 bp deletion that removed part of exon 3 and results in frameshift.	Deletion present in Illumina SR-WGS data.
S056	CMA normal, exome with single pathogenic variant in *WDR19*, consistent with presumed diagnosis of Sensenbrenner syndrome. Deletion/duplication analysis negative.	Confirmed the known inherited variant.	Intronic splice variant	None

Table S17: Reads used to calculate length of AGAA motif in individual S009.

Read ID/Contig	Start Pos	End Pos	Period Size	Copy Number	Percent Matches	Percent Indels	Score	Motif
268a7f22-1c5e-40a6-b50a-764cc962ecd4:20612-21234	84	562	4	118.5	94	2	753	AAGA
33430113-f292-4c8a-abaa-ef24685b0275:9126-9747	90	540	4	117	79	15	393	AAGA
ab709b56-560e-442b-b415-6538973c978d:4402-4955	66	485	4	107.5	86	10	505	AAGA
157b7b75-7a60-4e4e-8717-6bfa1fc8424f:3823-4453	81	569	4	121.8	92	5	720	AAGA
4683a48e-1bd4-430f-a361-712f000d8ff1:14088-14711	84	562	4	119.8	93	4	727	AAGA
622462ec-1093-4532-91dc-a8366bd9f712:1867-2467	83	537	4	116.8	91	6	620	AAAG
aee7ddf8-5473-4c67-9e45-65ddb6b3512b:1667-2278	78	556	4	120.2	91	5	695	AGAA
4885dba3-b149-4870-b903-eb240a6f34e0:1488-2110	85	562	4	119.5	96	1	762	AAGA
ffc51d09-c8a1-4308-a843-9540a3c812ea:1304-1906	83	546	4	113.8	78	13	297	AAGA
Average				117.2				
Excluded read IDs containing TGTT motif								
1fd6ae08-37dc-47e1-8804-28117b821352								
8f3ea358-76e1-41a7-88a7-ef0e287b4c77								
37a6233f-91bb-4a5f-909d-e3951cc855be								

Table S18: Predicted strength of the canonical splice donor site at the Exon 1–Intron 1 boundary of *ABCA4* (NM_000350) and alternative sites introduced by the ~1,500 bp insertion in individual S025.



Splice donor	SpliceSiteFinder-like	MaxEntScan	NNSPLICE	GeneSplicer	Human Splice Finder
Canonical	81.70	8.90	1.00	5.20	84.70
Alternative	90.10	8.00	0.90	-	95.00
Alternative	81.50	7.20	0.80	0.60	90.80
Splice acceptor	**SpliceSiteFinder-like**	**MaxEntScan**	**NNSPLICE**	**GeneSplicer**	**Human Splice Finder**
Canonical	95.25	10.87	0.99	12.03	92.11
Alternative	93.50	10.00	1.00	3.20	90.60
Alternative	88.54	7.30	1.00	1.50	79.83
Alternative	77.80	6.10	0.60	2.30	78.50
Alternative	83.20	5.90	0.70	-	85.80
Alternative	71.70	5.70	-	1.60	79.50
Alternative	74.60	6.90	0.60	-	78.30

While the SV is absent from all accessible population genetic databases (gnomAD, BRAVO), assessment by multiple *in silico* prediction tools suggest a strong likelihood of pathogenicity. Specifically, the deep residual neural network SpliceAI predicts a splice-altering consequence of the pre-mRNA through the introduction of a *de novo* donor site by the insertion event (Δ score = 0.33, high recall). The ~1,500 bp insertion sequence itself contains two strong alternative splice donors and six alternative splice acceptors. The relative strength of these alternative donors and acceptors indicates a high probability of competition with canonical sites. One alternative donor site in particular exhibits a stronger splice signal than the canonical donor site at the Exon 1–Intron 1 boundary is the most probable site of alternative splicing leading to the inclusion of the 5' portion of the intron in the final mRNA transcript. Length of color bars (blue = donor, green = acceptor) are proportioned to the respective scales of each algorithm: SpliceSiteFinder-Like (0-100), MaxEntScan (0-16), NNSPLICE (0-1), Gene Splicer (0-15), Human Splice Finder v.3.1 (0-100).

Table S19: Accession numbers or contact information for original sequencing data.

Individual	Data Accession or ID	Contact	Email
S002	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S003	phs000693	Danny E. Miller, MD, PhD	danny.miller @seattlechildrens.org
S004	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S008	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S009	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S011	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S013	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S014	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S016	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S018	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S020	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S021	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S022	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S023	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S025	TBD	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S035	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S036	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S039	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S040	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S041	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S046	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S047	TBD	Priya Kishnani, MD	priya.kishnani@duke.edu
S056	phs000693	Danny E. Miller, MD, PhD	danny.miller @seattlechildrens.org
S060	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S063	http://dx.doi.org/10.1136/jmedgenet-2020-107320		
S071	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S082	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S083	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
S086	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
04-01	TBD	Heather Mefford, MD, PhD	heather.mefford@stjude.org
04-02	TBD	Heather Mefford, MD, PhD	heather.mefford@stjude.org
04-03	TBD	Heather Mefford, MD, PhD	heather.mefford@stjude.org
06-01	TBD	Heather Mefford, MD, PhD	heather.mefford@stjude.org
06-02	TBD	Heather Mefford, MD, PhD	heather.mefford@stjude.org
06-03	TBD	Heather Mefford, MD, PhD	heather.mefford@stjude.org
BK144-03	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
BK180-03	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
BK294-03	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
BK364-03	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
BK397-101	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
BK430-103	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
BK482-101	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
BK487-101	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org
BK506-03	phs000693	Danny E. Miller, MD, PhD	danny.miller@seattlechildrens.org

Supplemental References

1. Miller, D.E., Squire, A., and Bennett, J.T. (2019). A child with autism, behavioral issues, and dysmorphic features found to have a tandem duplication within CTNND2 by mate-pair sequencing. Am J Med Genet A.

2. Walsh, T., Casadei, S., Munson, K.M., Eng, M., Mandell, J.B., Gulsuner, S., and King, M.-C. (2020). CRISPR–Cas9/long-read sequencing approach to identify cryptic mutations in BRCA1 and other tumour suppressor genes. J Med Genet jmedgenet-2020-107320.

3. LaCroix, A.J., Stabley, D., Sahraoui, R., Adam, M.P., Mehaffey, M., Kernan, K., Myers, C.T., Fagerstrom, C., Anadiotis, G., Akkari, Y.M., et al. (2019). GGC Repeat Expansion and Exon 1 Methylation of XYLT1 Is a Common Pathogenic Variant in Baratela-Scott Syndrome. Am J Hum Genetics *104*, 35–44.

4. Sulovari, A., Li, R., Audano, P.A., Porubsky, D., Vollger, M.R., Logsdon, G.A., Consortium, H.G.S.V., Warren, W.C., Pollen, A.A., Chaisson, M.J.P., et al. (2019). Human-specific tandem repeat expansion and differential gene expression during primate evolution. Proc National Acad Sci *116*, 201912175.

5. Guo, H., Duyzend, M.H., Coe, B.P., Baker, C., Hoekzema, K., Gerdts, J., Turner, T.N., Zody, M.C., Beighley, J.S., Murali, S.C., et al. (2019). Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. Genet Med *21*, 1611–1620.