XHGG, Volume *2*

Supplemental Information

Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders

Susan M. Hiatt, James M.J. Lawlor, Lori H. Handley, Ryne C. Ramaker, Brianne B. Rogers, E. Christopher Partridge, Lori Beth Boston, Melissa Williams, Christopher B. Plott, Jerry Jenkins, David E. Gray, James M. Holt, Kevin M. Bowling, E. Martina Bebin, Jane Grimwood, Jeremy Schmutz, and Gregory M. Cooper

Supplemental Methods

CDKL5 cDNA Amplicon Sequencing

Amplicons were generated with CDKL5_Exon_2_Forward and either CDKL5_Exon_5_Reverse or CDKL5_Exon_6_Reverse. PCRs were performed with OneTaq 2x MM (NEB, M0482) with 10uM primers and 1uL cDNA input. Thermocycler conditions were as follows: 94^{o}C for 30 s, 40 cycles of 98^{o}C for 30 s, 56.4^{o}C for 30 s, 68^{o}C for 30 s, and a final extension of 68^{o}C for 5 min. To achieve full resolution of each amplicon, each PCR reaction was loaded onto a 2% agarose gel and ran at 120mV/hr for 2 hours prior to imaging. The CDKL5_Exon_5_Reverse primer containing reactions resulted in an expected ~240bp amplicon was observed for each sample in addition to a ~275bp amplicon uniquely observed in the proband. Similarly, the CDKL5_Exon_6_Reverse primer containing reactions resulted in an expected ~360bp amplicon for each sample in addition to a ~390bp amplicon uniquely observed in the proband. Each amplicon was gel extracted with a DNA gel recovery kit (Zymogen, #D4007) according to manufacturer's protocol and eluted in 20uL of pre-warmed elution buffer. To ensure sufficient input for Sanger sequencing, a second round of PCR that was identical the first round was performed with 1uL of gel extracted DNA as input from each amplicon. The products of this second round of PCR were again gel extracted and eluted in 20uL of pre-warmed elution buffer. Each elution was submitted to MCLAB (www.mclab.com) for Sanger sequencing.

The forward and reverse primer sequences were as follows:

CDKL5_Exon_2_Forward (NM_003159 cDNA location 222-245):

TGTGGCTTGCATCAAAAGAGGAGT

CDKL5_Exon_5_Reverse (NM_003159 cDNA location 441-462): TCCTGCTTGAGAGTCCGAAGCA

CDKL5_Exon_6_Reverse (NM_003159 cDNA location 561-582):

TCAGGTGGAACTCCATTTGGCA

CDKL5 Genomic DNA PCR

We performed PCR to amplify a product spanning the downstream junction (at GRCh38/hg38 chrX:18510868) using the genomic DNA (gDNA) of the proband and parents as template, in two separate reactions. PCR was performed using Q5 polymerase (NEB #M0494S) with 100 ng gDNA template and 0.5 uM each primer, with 30 seconds initial denaturation at 98°C, 35 cycles of 10 seconds 98°C denaturation, 20 seconds 68°C annealing, 1 minute 72°C extension, and a final 2 minutes 72°C extension. One reaction utilized primer LONG1_FWD and primer JUNC1_REV, resulting in an amplicon of 729 bp, only with proband gDNA. The second reaction utilized primer LONG2_FWD and primer JUNC1_REV, resulting in an amplicon of 1,207 bp, only with proband gDNA. We performed agarose gel electrophoresis and observed no bands in these reactions with parents' gDNA as template, and the correct size bands with the proband gDNA as template. The amplicons were purified with QiaQuick columns (Qiagen #28104) and were Sanger sequenced (MCLab) using the same primers utilized in the PCR; the sequences were aligned to and matched the assembled proband sequence.

We performed a long PCR to amplify from the position of JUNC1_REV across the entirety of the insert sequence and into the upstream gDNA sequence of the proband. PCR was performed using Q5 polymerase (NEB #M0494S) with 100 ng gDNA template and 0.5 uM each primer, with 30 seconds initial denaturation at 98°C, 35 cycles of 10 seconds 98°C denaturation, 8 minutes

68°C annealing/extension, 8 minutes 72°C extension. We increased the time at 68°C primarily for polymerase extension across the A/T-rich sequence in the insert (Dhatterwal, et al., PMID: 29183338). The reaction utilized the primers LONG3_FWD; reverse complement of JUNC1_REV and LONG3_REV, resulting in an amplicon of 7,989 bp only with proband gDNA. We performed agarose gel electrophoresis and observed no bands in these reactions with parents' gDNA as template, and the correct size band with the proband gDNA as template. The amplicon was purified with a QiaQuick column (Qiagen #28104) and was Sanger sequenced (MCLab). For Sanger sequencing, we utilized seven primers: LONG3_FWD and LONG3_REV (the primers used to generate the amplicon), and LONGVAL1-LONGVAL5. This resulted in five non-contiguous regions of the amplicon, comprising in total approximately 57% of its length, to be sequence verified.

To verify PPEF1 sequence in the proband insert, we performed two paired PCR assays. PCR was performed using Q5 polymerase (NEB #M0494S) with 100 ng gDNA template and 0.5 uM each primer, with 30 seconds initial denaturation at 98°C, 35 cycles of 10 seconds 98°C denaturation, 20 seconds 68°C annealing, 1 minute 72°C extension, and a final 2 minutes 72°C extension. One assay utilized the primer JUNC3_REV (matches PPEF1 insert sequence) with the primer PPEF1_FWD2 (matches PPEF1 sequence not present in insert), resulting in an amplicon of 1,122 bp in proband and parental gDNA; this was paired with a PCR using the same JUNC3_REV and the primer JUNC3_FWD (matches CDKL5 sequence outside the insert), resulting in an amplicon of 1,074 bp in the proband gDNA but not in the parents' gDNA. The other assay utilized the primer JUNC4_REV (matches PPEF1 insert sequence) with the primer PPEF1_FWD3 (matches

PPEF1 sequence not present in insert), resulting in an amplicon of 1,004 bp in proband and parental gDNA; this was paired with a PCR using the same JUNC4_REV and the primer JUNC4_FWD (matches CDKL5 sequence outside the insert), resulting in an amplicon of 821 bp in the proband gDNA but not in the parents' gDNA. We performed agarose gel electrophoresis and observed the correctly sized bands (and appropriate absence of bands) in all these reactions.

Primers:

LONG1_FWD: 5'-AACCTGTACATGCCCACACG-3'

JUNC1_REV: 5'-GCCCCGTTGTGTCTGTTTTC-3'

LONG2_FWD: 5'-TATGAGGTGCACGGCATAGG-3'

JUNC1_REV: 5'-GCCCCGTTGTGTCTGTTTTC-3'

LONG3_FWD: 5'-GAAAACAGACACAACGGGGC-3'

LONG3_REV: 5'-acacacCCCTGTCAAGCAAA-3'

LONGVAL1: 5'-TCTCACGTGCAGAGACACAC-3'

LONGVAL2: 5'-GTGTGTCTCTGCACGTGAGA-3'

LONGVAL3: 5'-GAGGCCAGGAGTTTGAGACC-3'

LONGVAL4: 5'-CACCACCGATCCCACAGAAA-3'

LONGVAL5: 5'-TGAGGAATCGCCACACTGAC-3'

JUNC3_REV: 5'-ATCTGTCACGGCTTCCCTTG-3'

PPEF1_FWD2: CTTCCACCCACTCCCCATTC-3'

JUNC3_FWD: GTACATGCCCACACGCAAAG-3'

JUNC4_REV: GCAGGCAATGGAGGTGTAGT-3’

PPEF1_FWD3: 5’-agcagcagccagactcaaat-3’

JUNC4_FWD: 5’-TCATTATGAGGTGCACGGCA-3’

Proband 4 Genomic DNA PCR

We performed multiple PCRs to amplify products spanning the junctions of the chr6 chromothripsis-like region in Proband 4. Using the genomic DNA (gDNA) of the proband and parents as template, we performed PCR using coordinates identified in the paternal contigs of the trio-binned assembly (in Table S6B). PCR was performed using Phusion High-Fidelity Master Mix (ThermoFisher #F531) with 100 ng gDNA template and 0.5 uM each primer, with 20 seconds initial denaturation at 98°C, 30 cycles of 10 seconds 98°C denaturation, 20 seconds 63°C annealing, 1 or 2 minutes (depending on predicted amplicon size) 72°C extension, and a final 5 minutes 72°C extension.

Primers and expected amplicon sizes are listed below. We performed 1% agarose gel electrophoresis, using a 1 kb Plus DNA ladder (Invitrogen 10787-018) and observed no bands in these reactions with parents’ gDNA as template, and the correct size bands with the proband gDNA as template, except where noted. A subset of amplicons were gel eluted with QiaQuick Gel Extraction kit (Qiagen #28704) and a subset were Sanger sequenced (MCLab) using the same primers utilized in the PCR; the sequences were aligned to and matched the assembled proband sequence.

Targeted Junction	Upstream primer sequence	Downstream Primer sequence	Expected Amplicon size in proband	Actual amplicon size in proband
REF/D	REF_FWD 5'-CTGCCAAGCCCCTACATTTG	D_RVS 5'- TGTGCCATTGCTGACAATCAG	790 BP	~790
D/C	D_FWD 5'- AGCATAGGCAAGTACACACATC	C_RVS 5'- CTGGGTGTCAAGCTCCTGAG	657 BP	NO AMP
C/A	C_FWD 5'- TGTGGTCTTTAGAGTTTTGCATCC	A_RVS 5'- TGCTTCTCCTCTTGTCGTCTC	1298 BP	~1298
A/G	A_FWD 5'- GCCAGAGCTCTCTTCCTGAC	G_RVS 5'- TTGCGTACATGGTGGGACTC	1738 BP	~1738
G/H	G_FWD 5'- TACAGCCCCAGAGCCATTTC	H_RVS 5'- ACGCTGTGTCTAGGCAGTTC	1589/2087 BP*	~2087
H/F	H_FWD 5'- GAGCACCAGTGGAGGAGTTC	F_RVS 5'- AGCAAACCAGGTCTGACAGC	1235 BP	~1235
F/E	F_FWD 5'- ACCCACATTTGACCCGATCTAG	E_RVS 5'- GGACTCGGTCTTATCGCTGAG	800 BP	~800
E/B	E_FWD 5'- GCTCAACCTTGCCGCTTATC	B_RVS 5'- TGCAAGGCTTTTTCCTTAAGGAC	753 BP	~753
B/REF	B_FWD 5'- TTTTGCCCACCCCTCTACAC	REF_RVS 5'- CTTACCATGGGCTGCAAAGC	798/1238 BP*	~1238

*For amplicons G/H and B/REF, two breakpoints were identified in CCS data, but the larger amplicon was confirmed by PCR. See also Table S6.

PCR amplification of Alu Insertions

Using the genomic DNA (gDNA) of probands and parents as template, we performed PCR to amplify sequence surrounding called Alu insertions (Figure S2). PCR was performed using Phusion High-Fidelity Master Mix (ThermoFisher #F531) with 100 ng gDNA template and 0.5 uM each primer, with 20 seconds initial denaturation at 98°C, 30 cycles of 10 seconds 98°C denaturation, 20 seconds 63°C annealing, 2 minutes 72°C extension, and a final 5 minutes 72°C extension. Primers are listed below. We performed 1% agarose gel electrophoresis and observed bands in the respective samples as described in the manuscript. The Alu in Proband 6, chr7 was not amplified due to surrounding repetitive sequence.

Proband1_chr2Alu_fwd TTGGGGTTCATGGGGCATAC

Proband1_chr2Alu_rev AGAGATGATTGGCGGTGGTC

Proband4_chr3Alu_fwd AGACAGCTGTTCCTTCATGTTTG

Proband4_chr3Alu_rev ACTCTTGGAAAACTGGAACGG

Proband6_chr6Alu_fwd TGTGAAGCAAACATGGCATGAC

Proband6_chr6Alu_rev AGTCTTGTGCCCCCACTATG

Proband6_chr7Alu_fwd CTGCCCAGAAGGACCATCTG

Proband6_chr7Alu_rev AAACAGCATGGAGGGGATGG

Resolution of tig66 mis-assembly in Proband 4

One of the paternal contigs in the proband's *de novo* assembly (tig66) appeared to be misassembled, likely due to the chr7 sequence inversion. Manual inspection of the proband's *de novo* assembly vs. hg38 assembly alignments (and vice versa) and SNVs in these regions showed the paternal assembly contained both paternal and maternal sequence. This is a documented artifact of the current version of hifiasm (https://github.com/chhylp123/hifiasm/issues/10; https://github.com/chhylp123/hifiasm/issues/21). Curation of SNVs in contigs and CCS reads allowed manual reassembly of this region. Visualization of the non-trio-binned hifiasm assembly also confirmed the proper alignments (Supplemental Figure 6).

Supplemental Note: Clinical Summaries

Proband 1 is a Caucasian female with seizures, facial dysmorphism, and hypotonia. Before enrolling in the CSER study around the age of 9 years, she had the following negative/normal genetic testing performed: Microarray, karyotype, *UBE3A* and *MECP2* single gene testing.

Proband 2 is an African American female with hypotonia, severe intellectual disability, and seizures. She has had a normal brain MRI. Before enrollment in CSER (around age 4 years), she had the following negative/normal genetic testing performed: Microarray, mitochondrial DNA sequencing, and single gene testing for *SCN1A, MECP2, ARX, TSC1, TSC2, CDKL5*, and *SLC2A1*.

Proband 3 is a Caucasian male who has seizures and intellectual disability. Before enrollment in CSER (7-8 years), he had the following negative/normal genetic testing performed: Microarray, Fragile X, and epilepsy panel, and single gene testing for *MECP2* and *PTEN*.

Proband 4 is a Caucasian/African-American female born at 36 weeks to a 24-year old gravida 2 para 1011 female who reported no prenatal complications. She was born via vaginal delivery weighing 4 lbs and 13 ounces, measuring 18 inches in length, and with a head circumference of 13 inches. She was determined to have jaundice and failure to thrive in the newborn period. Her parents reported that the failure to thrive/difficulty gaining weight resolved around 9 months of age. By age 3 yo she had normal bone age studies and a normal head CT. She did not talk until 18-24 months of age. She did not stand independently until 21 months and started walking at 2 years of age, with a reportedly wide gait. She received early intervention services until age 3 yo. As of last available records, proband was in 1st grade after repeating kindergarten, where she has an IEP for math, reading, and speech and also receives speech and behavioral therapy. The proband was enrolled into the CSER study at eight years old. At enrollment, she was noted to have moderate ID, hypotonia, and facial dysmorphism (long face with prominent chin, epicanthal folds, deep-set eyes, fifth finger clinodactyly, and shallow creases). Before enrollment in CSER, the proband had negative/normal clinical testing for the following: Microarray, Fragile X, and *NSD1* single gene testing.

Proband 5 is a Caucasian male with moderate intellectual disability, seizures, speech delay, mild dysarthria, optic nerve hypoplasia, a history of lactic acidosis, and hydrocephalus with VP shunt. Brain MRI showed parenchymal thinning of the left cerebral hemisphere, thinned optic chiasm and thin corpus callosum. Before enrollment in CSER (5-6 years), he had the following negative/normal clinical testing: Microarray, MitoMetPlus Microarray, Dual Genome Panel, MELAS testing, and single gene testing for PDHA1 and GLA.

Proband 6 is a Caucasian female who is a product of a 39 weeks-by-dates singleton pregnancy. APGAR scores were 8 at one minute, 9 at five minutes, and 9 at ten minutes. She was born via C-section due to failure to progress. The proband was enrolled into the CSER study at 15 years old. She presented with moderate intellectual disability, seizures (focal/partial and complex partial seizures, with mention of intractable epilepsy), speech delay, hypothyroidism, ADHD, and follicular cysts. Epilepsy onset was reported to be in infancy. Before enrollment, she had a

negative clinical microarray, normal brain MRI, normal head CT. This participant had a working diagnosis of “Williams syndrome without genetic confirmation” prior to participation in the research study. At time of study, she was attending high school in special education programming that included life skills, independent self-help and community skills as well as minimal job skills under an extended IEP. Counting and reading were reported to be very challenging. Participant is very social.

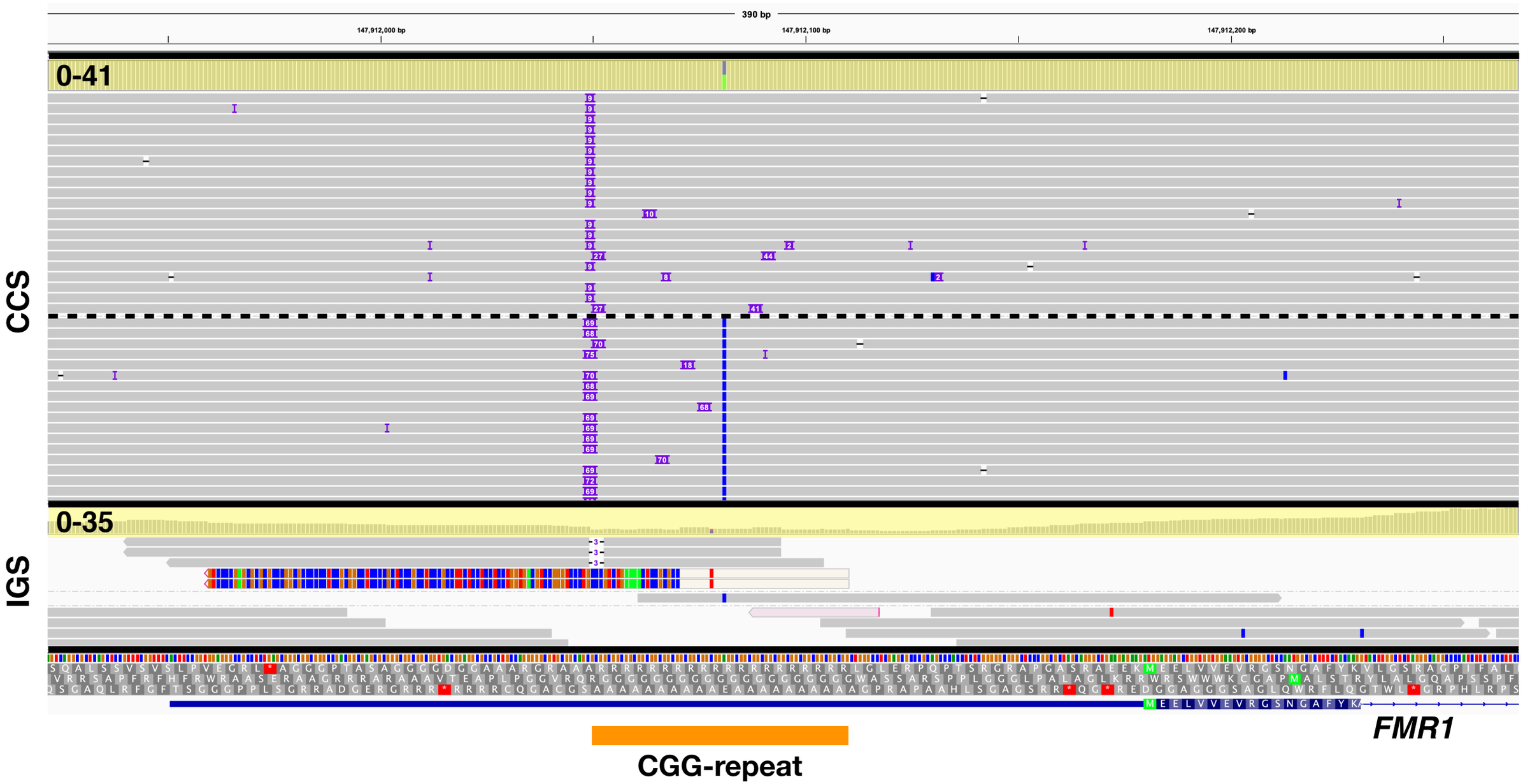
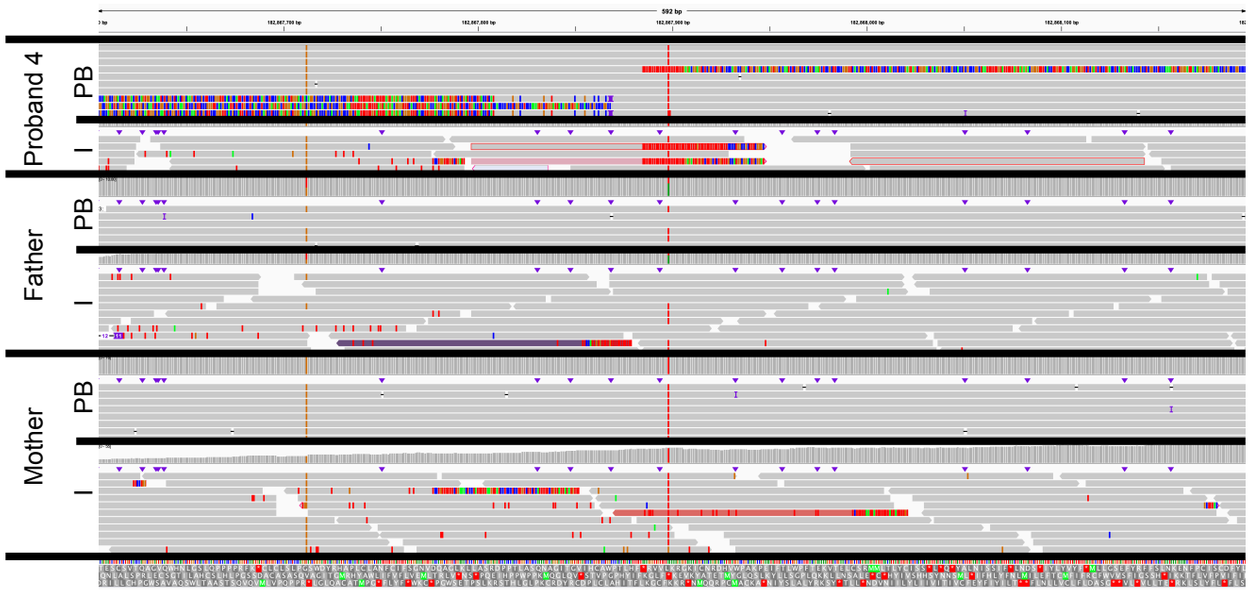


Figure S1. Visualization of a subset of CCS and IGS reads in Proband 6 aligned to the 5’ end of FMR1 which contains a CGG-repeat region (orange bar). CCS reads group into two bins to represent two distinct repeat alleles (69 nt and 9 nt alleles), separated by a dashed line. Note read depth highlighted by the yellow box, with depth scale shown at the left end of the box.

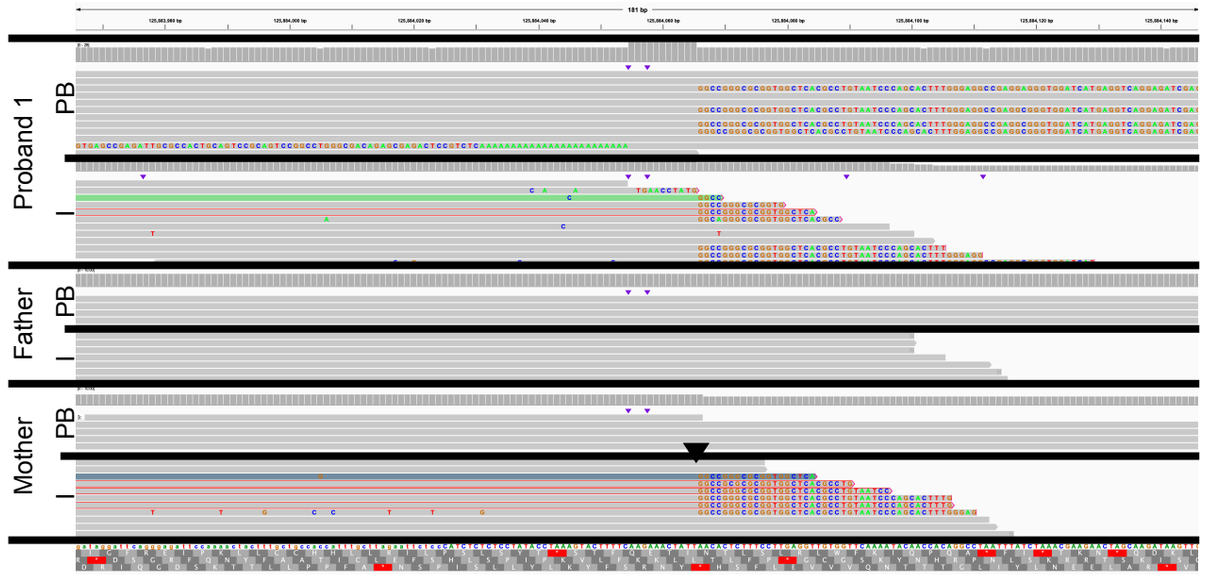
A

chr3:182,867,604-182,868,196



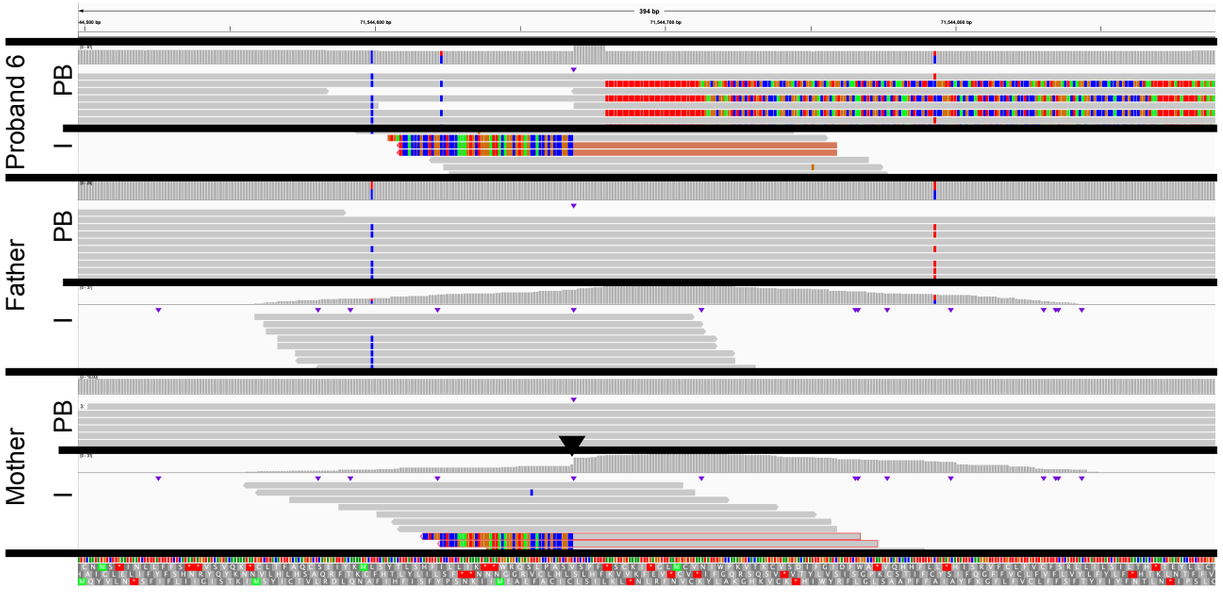
B

chr2:125,883,966-125,884,146



C

chr6:71,544,498-71,544,892



D

chr7:100,290,624-100,291,100

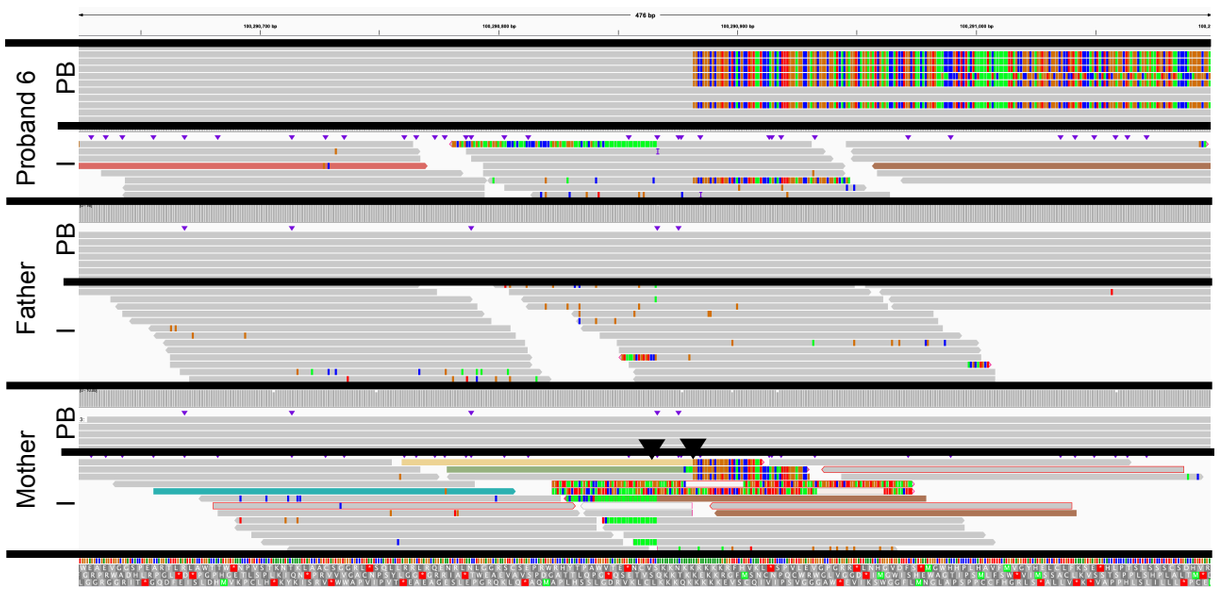


Figure S2. Four *de novo* Alu insertions were called from CCS data in the six probands. In one case, CCS and IGS reads both appear to support *de novo* status (A), while in the three other cases (B, C, D), there are reads supporting the *Alu* insertion in the mom's Illumina reads (black triangle).

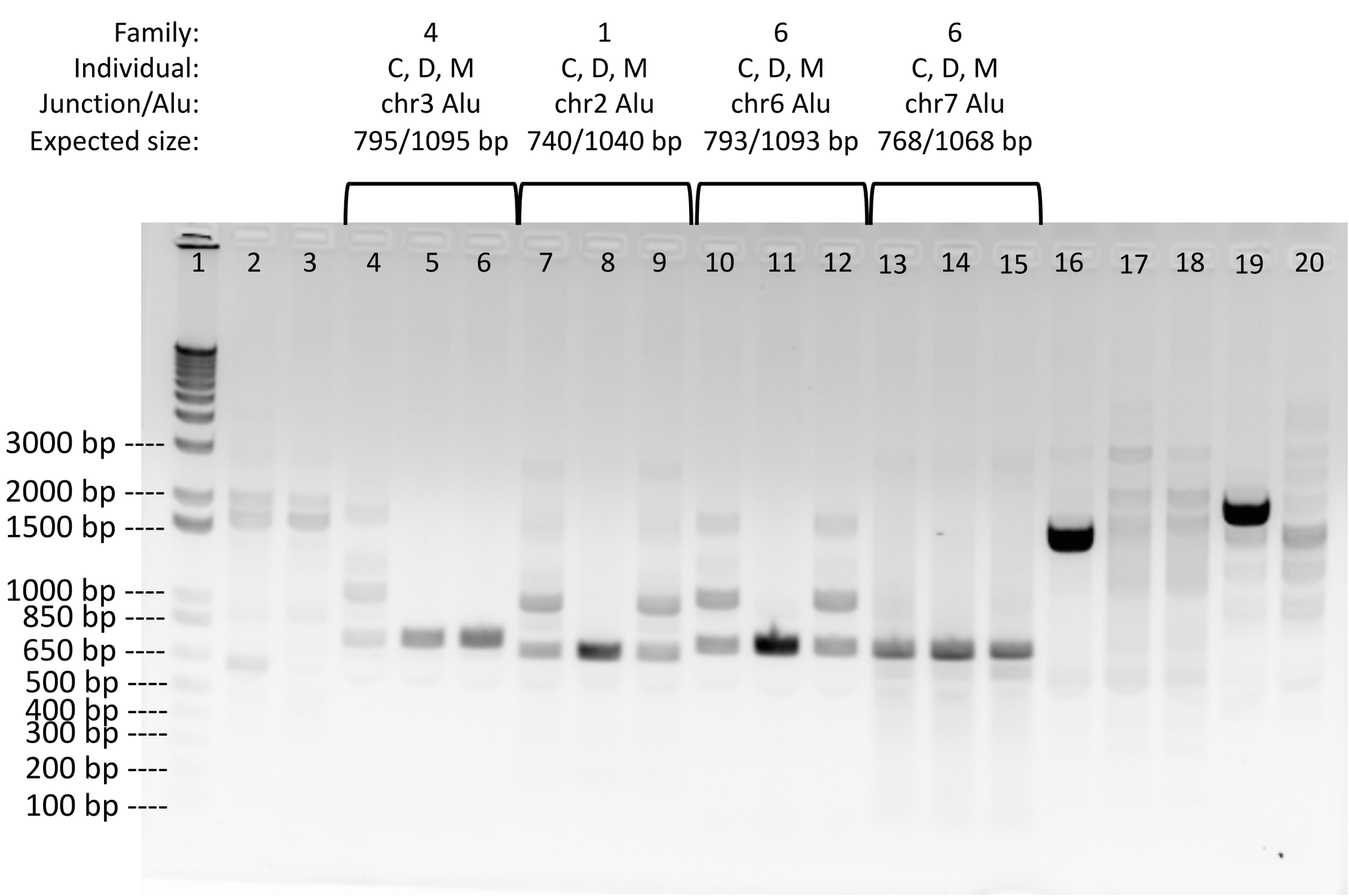


Figure S3. PCR Amplification confirms three of four Alu insertions in various probands. A 1% agarose gel was run with 1 ug Invitrogen 1 Kb Plus DNA Ladder in Lane 1. Lanes 4-15 are sets of three PCR samples; each set of 3 represents child (C), dad (D), and mom (M) as PCR template. While the chr3 Alu in Family 4 appears to be *de novo,* maternal inheritance was confirmed for the chr2 Alu in Family 1, and the chr6 Alu in Family 6. PCR for the chr7 Alu insertion in Family 6 resulted in nonspecific bands only, likely due to surrounding repetitive sequence. Note that this gel image is duplicated as Figure S9 and Lanes 2,3, and 16-20 are described there.

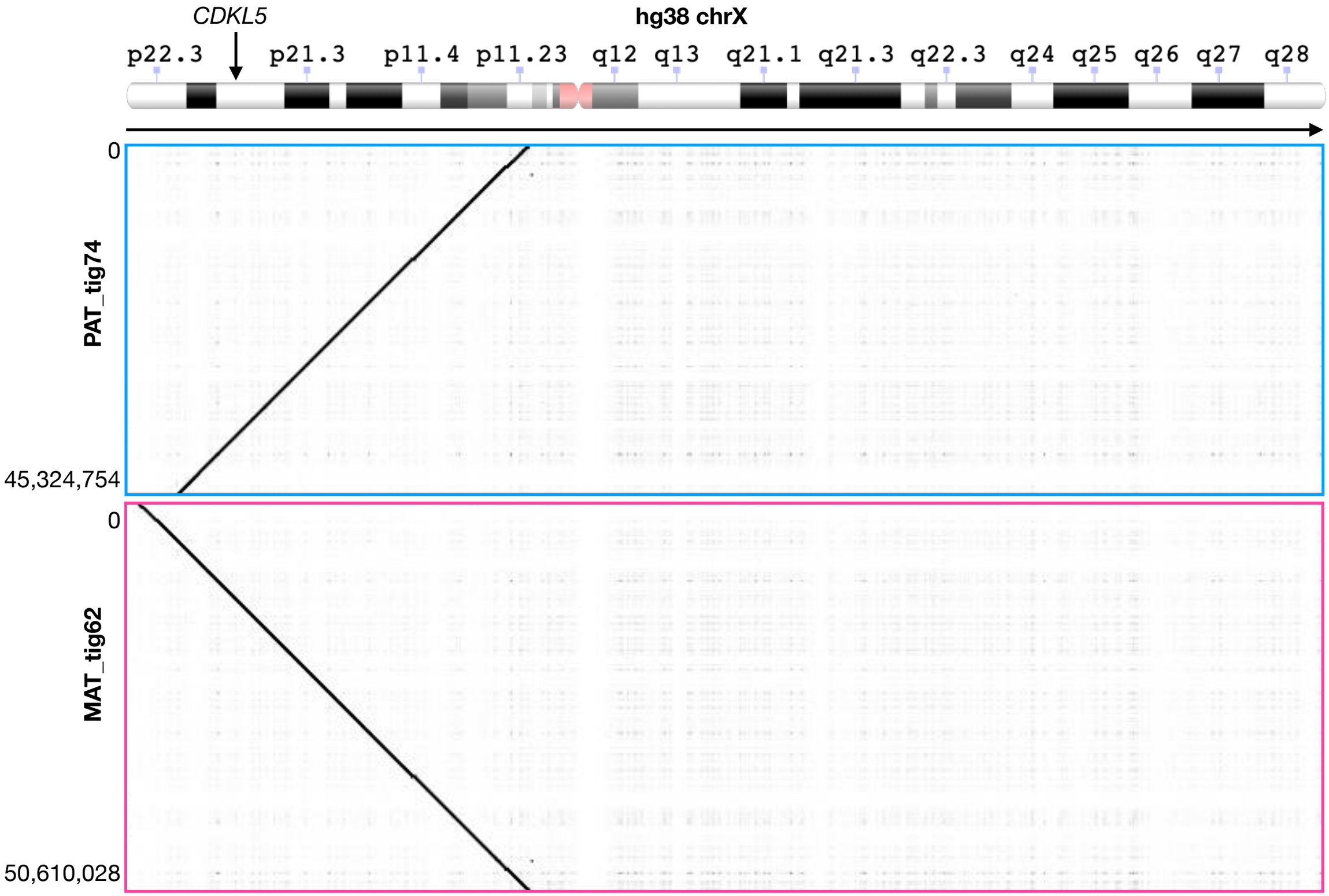


Figure S4. *De novo* trio-based assembly for proband 6 resulted in two large paternal- and maternal-specific contigs (PAT_tig74 and MAT_tig62) covering the majority of the p arm of chromosome X, including *CDKL5*. Ideogram is from the NCBI Genome Decoration Page (https://www.ncbi.nlm.nih.gov/genome/tools/gdp).

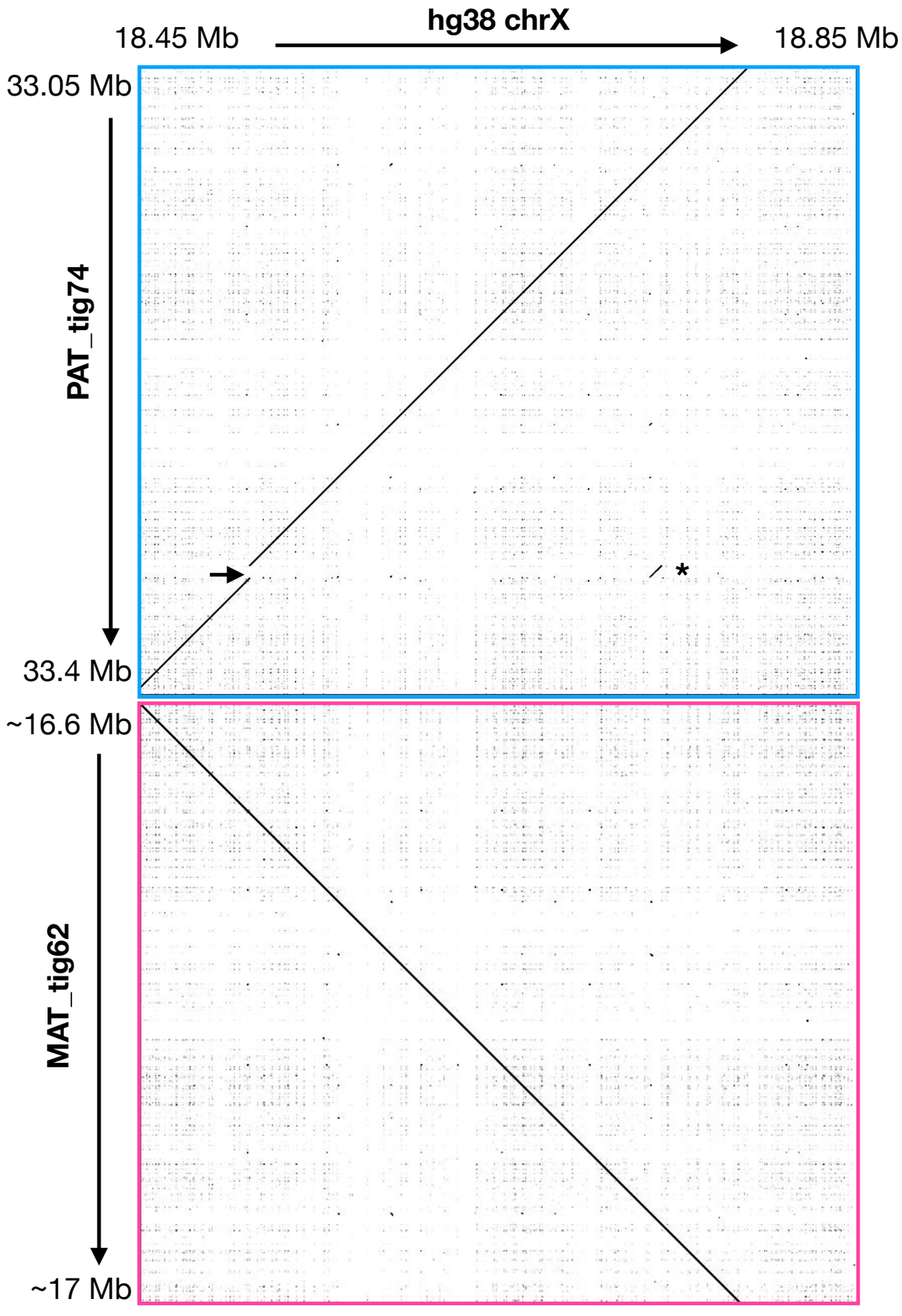


Figure S5. Alignment of Proband 6's paternal and maternal contigs surrounding *CDKL5* to reference chromosome X. The 6993 bp insertion in the proband's primary contig is noted with an arrow, and alignment to a downstream LINE in an intron of *PPEF1* is shown with an asterisk.

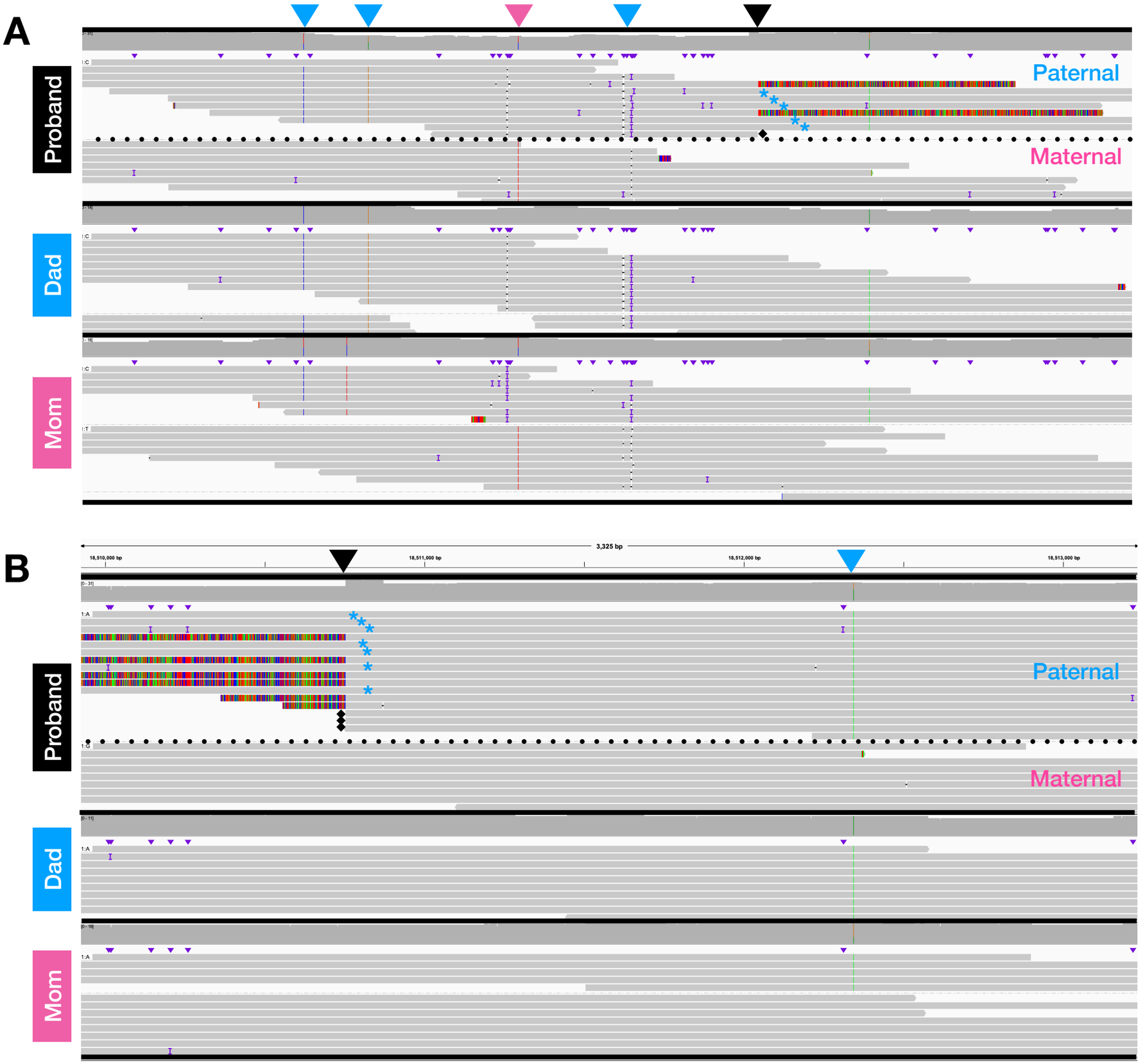


Figure S6. Phasing of single nucleotide variants both upstream (A) and downstream (B) of the insertion (black triangle) in *CDKL5* in proband 6 indicate that the insertion is on the paternal allele. Inherited variants on the paternal allele (blue triangles) and maternal allele (pink triangle) are labeled. A dotted line separates the proband's haplotypes, which are labeled as paternal or maternal. Note the existence of paternal alleles lacking the insertion, suggesting mosaicism (blue asterisks). Black diamonds indicate reads that are hard-clipped and support the insertion. All reads supporting the insertion are shown, but only a subset of other reads are shown here.

Figure S7. Partial sequence from the proband's paternal *de novo* assembly ('PAT_tig74' in Figure S4) that surrounds the L1-mediated insertion in *CDKL5*. The 119-nt TSD is in yellow/orange, and exon 3 sequence within the TSD is underlined. Novel L1-inserted sequence is in green, and PPEF1 inserted sequence is in blue. Sanger confirmed sequence is in bold.

5'-
ACAGAATCTTCTGAGTTAATTTTCCTTTATTCTCAAAGTTCTTGAGACCAAATCTTGTCTCCATAAATGTGTAACCCTACATTTAAACGTATCGCCAGTGAAATTTTCTTTAAAAAGAAACATTTAAAAATACTTGTTGGAAGAAATGACTTGTCTATACTCTTGATTTA
TAATAAATGTATAATAAAGTATGTTCCATTTGTATGTGGCTCATTATTGCATTATCACTATAGAATCTTGTCACTTTAGTAGTCATTATTATAACACTGAGAAGCAATGTCAGTATAGCAGAGCTTTGTAGTTTGTATGCGTGCCCTTGATTGTTTACTTCTTTTTATTATAGG
AGCCTATGGAGTTGTACTTAAATGCAGACACAAGGCAAGTACATTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTATTTTTTTTTTTTTTTATTATACTCTAAGTTTTAGGGTACATGTGC
ACATTGTGCAGGTTAGTTACATATGTATACATGTGCCATGCTGGTGCGCTGCACCCACTAATGTGTCATCTAGCATTAGGTATATCTCCCAATGCTATCCCTCCCCCCTCCCCCGACCCCACCACAGTCCCCAGAGTGTGATATTCCCCTTCCTGTGTCCATGTGATCTCATTG
TTCAATTCCCACCTATGAGTGAGAATATGCGGTGTTTGGTTTTTTGTTCTTGCGATAGTTTACTGAGAATGATGGTTTCCGATTTCATCCATGTCCCTACAAAGGATATGAACTCATCATTTTTTATGGCTGCATAGTATTCCATGGTGTATATGTGCCACATTTTCTTAATCC
AGTCTATCATTGTTGGACATTTGGGTTGGTTCCAAGTCTTTGCTATTGTGAATAGTGCCGCAATAAACATACGTGTGCATGTGTCTTTATAGCAGCATGATTTATACTCATTTGGGTATATACCCAGTAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCCTG
AGGAATCGCCACACTGACTTCCACAATGGTTGAACTAGTTTACAGTCCCACCAACAGTGTAAA**AGTGTTCCTATTTCTCCGCATCCTCTCCAGCACCTGTTGTTTCCTGACTTTTTAATGATTGCCATTCTAACTGGTGTGAGATGATATCTCATAGTGGTTTTGATTTGC**
ATTTCTCTGATGGCCAGTGATGATGAGCATTTCTTCATGTGTTTTTTGGCTGCATAAATGTCTTCTTTTGAGAAGTGTCTGTTCATGTCCTTCGCCCACTTTTTGATGGGGTTGTTTGTTTTTTTCTTGTAAATTTGTTTGAGTTCATTGTAGATTCTGGATATTAGCCCT
TTGTCAGATGAGTAGGTTGCAAAAATTTTCTCCCATGTTGTAGGTTGCCTGTTCACTCTGATGGTAGTTTCTTTTGCTGTGCAGAAGCTCTTTAGTTTAATTAGATCCCATTTGTCAATTTTGTCTTTTGTTGCCATTGCTTTTGGTGTTTTGGACATGAAGTCCTTGCC
CACGCCTATGTCCTGAATGGTAATGCCTAGGTTTTCTTCTAGGGTTTTTATGGTTTTAGGTTTAACGTTTAAATCTTTAATCCATCTTGAATTGATTTTTGTATAAGGTGTAAGGAAGGGATCCAGTTTCAGCTTTCTACATATGGCTAGCCAGTTTTCCCAGCACCATT
TATTAAATAGGGAATCCTTTCCCCATTGCTTGTTTTTCTCAGGTTTGTCAAAGATCAGATAGTTGTAGATAAGCGGCATTATTTCTGAGGGCTCTGTTCTGTTCCATTGATCTATATCTCTGTTTTGGTACCAGTACCATGCTGTTTTGGTTACTGTAGCCTTGTAGTATA
GTTTGAAGTCAGGTAGTGTGATGCCTCCAGCTTTGTTCTTTTGGCTTAGGATTGACTTGGCAATGCGGGCTCTTTTTGGTTCCATATGAACTTTAAAGTAGTTTTTTCCAATTCTGTGAAGAAAGTCATTGGTAGCTTGATGGGGATGGCATTGAATCTGTAAATTA
CCTTGGGCAGTATGGCCATTTTCACGATATTGATTCTTCCTACCCATGAGCATGGAATGTTCTTCCATTTGTTTGTCTCCTCTTTTATTTCCTTGAGCAGTGGTTTGTAGTTCTCCTTGAAGAGGTCCTTCACATCCCTTGTAAGTTGGATTCCTAGGTATTTTATTCTCTTTG
AAGCAATTGTGAATGGGAGTTCACCCATGATTTGGCTCTCTGTTTGTCTGTTGTTGGTGTATAAGAATGCTTGTGATTTTTGTACATTGATTTTGTATCCTGA**GACTTTGCTGAAGTTGCTTATCAGCTTAAGGAGATTTTGGGCTGAGACGATGGGGTTTTCTAGATA**
AACAATCATGTCGTCTGCAAACAGGGACAATTTGACTTCCTCTTTTCCTAATTGAATACCCTTTATTTCCTTCTCCTGCCTGATTGCCCTGGCCAGAACTTCCAACACTATGTTGAATAGGAGCGGTGAGAGAGGGCATCCCTGTCTTGTGCCGGTTTTCAAAGGGA
ATGCTTCCAGTTTTTGCCCATTCAGTATGATATTGGCTGTGGGTTTGTCATAGATAGCTCTTATTATTTTGAAATACGTCCCATCAATACCTAATTTATTGAGAGTTTTTAGCATGAAGGGTTGTTGAATTTTGTCAAAGGCTTTTTCTGCATCTATTGAGATAATCATGT
GGTTTTTGTCTTTGGCTCTGTTTATATGCTGGATTACATTTATTGATTTGCGTATATTGAACCAGCCTTGCATCCCAGGGATGAAGCCCACTTGATCATGGTGGATAAGCTTTTTGATGTGCTGCTGGATTCGGTTTGCCAGTATTTTATTGAGGATTTTTGCATCAATG
TTCATCAAGGATATTGGTCTAAAATTCTCTTTTTTGGTTGTGTCTCTGCCTGGCTTTGGTATCAGAATGATGCTGGCCTCATAAAATGAGTTAGGGAGGATTCCCTCTTTTTCTATTGATTGGAATAGTTTCAGAAGGAATGGTACCAGTTCCTCTTTGTACCTCTGGT
AGAATTCGGCTGTGAATCCATCTGGTCCTGGACTCTTTTTGGTTGGTAAACTATTGATTATTGCCACAATTTCAGAGCCTGTTATTGGTCGATTCAGAGATTCAACTTCTTCCTGGTTTAGTCTTGGGAGAGTGTATGTGTTGAGGAATGTATCCATTTCTTCTAGATT
TTCTAGTTTATTTGCGTAGAGGTGTTTGTAGTATTCTCTGATGGTAGTTTGTATTTCTGTGGGATCGGTGGTGATATCCCCTTTATCATTTTTTATTGTGTCTATTTGATTCTTCTCTCTTTTTTTTCTTTATTAGTCTTGCTAGCGGTCTATCAATTTTGTTGATCCTTTCAAAAA
ACCAGCTCCTGGATTCATTGATTTTTTGAAGGGTTTTTTGTGTCTCTATTTCCTTCAGTTCTGCTCTGATTTTAGTTATTTCTTGCCTTCTGCTAGCTTTTGAATGTGTTTGCTCTTGCTTTTCTAGTTCTTTTAATTGTGATGTTAGGGTGTCAATTTTGGATCTTTCCTGCTT
TCTCTTGTAGGCATTTAGTGCTATAAATTTCCCTCTACACACTGCTTTGAATGCGTCCCAGAGATTCTGGTATGTGGTGTCTTTGTTCTCGTTGGTTTCAAAGAACATCTTTATTTCTGCCTTCATTTCGTTATGTACCCAGTAGTCATTCAGGAGCAGGTTGTTCAGTTTCC
ATGTAGTTGAGCGGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGATAGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGTGGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAA
AATGTATATTCTGTTGATTTGGGGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCTGGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAAAGTCTCCCATTATTAATGTGTGGGAGTCTA
AGTCTCTTTGTAGGTCACTGAGGACTTGCTTTATGAATCTGGGTGCTCCTGTATTGGGTGCATAAATATTTAGGATAGTTAGCTCCTCTTGTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTTTATCAGAGA
CTAGGATTGCAACCCCTGCCTTTTTTTGTTTTCCATTGGCTTGGTAGATCTTCCTCCATCCTTTTATTTTGAGCCTATGTGTGTCTCTGCACGTGAGATGGGTTTCCTGAATACAGCACACTGATGGGTCTTGAC**TCTTTATCCAACTTGCCAGTCTGTGTCTTTTAATTGC**
AGAATTTAGTCCATTTATATTTAAAGTTAATATTGTTATGTGTGAATTTGATCCTGTCATTATGATGTTAGCTGGTGATTTTGCTCATTAGTTGATGCAGTTTCTTCCTAGTCTCGATGGTCTTTACATTTTGGCATGATTTTGCAGCGGCTGGTACCGGTTGTTCCTTTCC
ATGTTTAGCGCTTCCTTCAGGAGCTCTTTTAGGGCAGGCCTGGTGGTGACAAAATCTCTCAGCATTTGCTTGTCTTTAAAGGATTTTGTTTCTCCTTCACTTACGAAGCTTAGTTTGGCTGGATATGAAATTCTGGGTTGAAAATTCTTTTCTTTAAGAATGTTGAAT
ATTGGCTCCCACCCTCTTCTGGCTTGTAGAGTTTCTGCCGAGAGATCCGCTGTTAGTCTGATAGGCTTCCCTTTGTGGGTAACCCGACCTTTCTCTCTGGCTGCCCTTAACATTTTTTCCTTCATTTCCAACTTTGGTGAATCTGACAATTATGTGTCTTGGAGTTGTTC
TTCTCGAGGAGTATCTTTGTGGCATTCTCTGTATTTCCTGAATCTGAATGTTGGCCTGCCTTGCTAGATTGGGGAAGTTCTCCTGGATAATATCCTGAAGAGTGTTTTCCAACTTGGTTCCATTCTCCCAGTCACTTTCAGGTACACCAATCAGATGCAGATTTAGTCT
TTTCACATAGTCCCATATTTCTTGGAGGCTTTGTTCATTTCTTATTACTCTTTTTTCTCTAAACTTCTCTTCTCGCTTCATTTCATTCATTTGATCTTCAATCACTGATACCCTTTCTTCCACTTGATCAAATCAGCTACTGAAGCTTGTGCATGCGTCATGTAATTCTCGTGC
CGTGGTTTTCAGCTCCATCAGGTCATTTAAGGTCTTCTCTACGCTGTTTATTCTAGTTAGCCATTCGTCTAATCTTTTTTCAAGGTTTTTAACTTCTTTGCGATGGGTTCGAACATCCTCCTTTAGCTCGGAGAAGTTTGTTATTACCGATCATCCGAGGCCTACGGCTGTCA
ACTCATCAAAGTCCTTCTCTGTCCAGCTTTGTTCCATTGCTGGCGAGGAGCTGCGTTCCTTTGGAGGAGAAGAGGCACTCTGAGTTTTAGAATTTTCAGCTTTTCTGCTCTGGTTTCTCCGCATCTTTGTGGTTTTATCTGCCTTTGGTCTTTGATGATGGTGACCTACA
AATGGGGTTTTGGTGTGGATGTCCTTTCTGTTTGTTAGTTTTCCTTCTAACAGTCAGGACCCTCAGCTGCAGGTCTGTTGGAATTTGCTGGAGGTCCACTCCAGACCCTGTTTGCCTGAGTATCACCAGCAGAGGCTTCAGAACAGCAAATATTGCAGAACGGCAAAT
GTTGCTGCCTGATCCTTTCTCTGG**AAACTTCGTCTCAGAGGGGTACCTGGCTGTATGAGGTGTCATTCGGCCCCTCCTGGGAGTTGTCTCCCAGTTAGGCTACTCGGGGGTCAGGGACCCACTTGAGGAGGCAGTCTGTCCGTTGTCAGATCTCATACTCTGTGCT**
GGGAGAACCACTACTCTCTTCAAAGCTGTCAGACAGGGATGTTTAAGTCTGCAGAAGTTTCTGCTGCCTTTTGTTCAGCTATGCCCTGCCCCCAGAGGTGGCGTCTATAGAGGCAGGCAGGCCTCCTTGAGCTGCAATGGGCTCCACCCAGTTCAAGCTTCCCGG
CTGCTTTGTTTACCTACTCAAGCCTCAGCAATGGTGGATGCCCCTCCCCCACCCTTGCTGCTGCCTTGCAGTTCAATCTTGGACTGCTGTGCTAGCAGTGAGCGAGGCTCCGTGGGCGTGGGACCCTCCAAGCCAGGTGTGGGATATAATCTCCTAGTGTGCCATTT
GCTAAGACCATTGGAAAAGTGCAGTATTAGAGTGGGAGTGTCCTGATTTTCCAGGTACCATCTGTCACGGCTTCCCTTGGCTAGGAAAGGGAATTCCCCGACCCCTTGCCCTTCCTGGGTGAGGCGATGCCCAGCCCTGCTTCTGCTCACACTCCGTGGGCTGCA
CCCACTGTCCAATAAGCCCGAGTGAGATGAACCCGGTGCCTCAGTTGGAAATACAGAAATCACCTGTCTTCTGCGTCGCTGACGCTGGGAGCTGTAGACTGGAGCTGTTCCTATTCGGCCATCTTGGAACCAAGATCAAGGGAACTTCATTTTTCATCCACAAAAC
TTTTTTGCTGAGAGATCAAAATATTTTGAGTATTTCATGCAGCAGATCTAGGGGAAAATGGCAAATTAAACACAGTTGGCCCCGTTGTGTCTGTTTTCTGTCCTCTCTCTCTCTCTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTCTGTTTCCCAGGCTGCAGTGCAGTGG
CACGGTCTCGGCTCACTGCAACCTCTGCCTCCCGGATTCAGGTGATTCTCCTGCCTCAGCCTCCTGAGTGTCTGGGATTACAGGTGCTTGCTACCACACCCAGCTAATTTTTTGTATTTTTAGTAGAGACAGGGTTTGAGAATGTTGACCAGGTTGGTCTTGAACCCCT
GACCTCGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCACCCGGCCTCTCTTCTCTTTGATCTGTCTGTCTGTCTACCTATCATCTAGTCATTCTGAAGCAGAGGCAGGCAATGGAGGTGTAGTTCAAAAATGATTTAAAATTATA
ATTTTTTACAACCATGTATTTCTCCTGCAAATTAAGATCTGCCTGTTATTATCTCCAAGACAGATGCCAGACTTAAACGCTCATTTTGCTTTAATTTGGGCCATGTCAGTATAGCAGAGCTTTGTAGTTTGTATGCGTGCCCTTGATTGTTTACTTCTTTTTATTATAGGAGCC
TATGGAGTTGTACTTAAATGCAGACACAAGGCAAGTACATTATTTTTAAAAAGAAATATCTGTATATGTTTAACTGTTTTGAAACTAATGTAGTGGTCTTTGCGTGTGGGCATGTACAGGTTGAGCATTCCTTATCTGAAATGCTTGAGACCAGAAATGTTTTGGATTTTT
TTGGATTTTGGAATATTTCCATATATGTAATGAGATACCTTGGGGATGGGACCCAAGGCTAGACACAAAATTCATTTATGTTTCATATATACTTAATACATATAGCTTGAAGGTAATTTTATGCAATATTTTAAGTTTTTGCCTAAAACAAAGTTTAAATACTTATATGTGGAAT
TTTCCACTTGTGGCATCATATCAGTACTCAAAGTTTTGGATTTTGGAGCATTTTGGATTTTTGAATTAGGGATGCTCAACCTGTAATTAAAACTCTGGTTATGAGGATGTTGTTTAATGAAGTGGAATTAATATGTGAATGCTGCCAGAAAGAGACTACTTTTAGATAAGC
TTGATCATGGTTGGGTAGTCCTTAGTTCTTAGCAATCACGTGCTATTTTTTTTTTTCAACTGACCTATGCCGTGCACCTCATAATGACATTTAGATCAACAGTGGACAGCATGTATGATGGTGGTCCCGTAAGATTATGATATGGTATTTTTTGTTTAGATAAAAAACATTTT
CAGATACACAAATACTTATCATTGACTATGGTACTCAGTAAGGTAACATGCTATACAGGTTTGTAGCCTTGAAACAATAGGCTATACCATGTAGCCTAGATGTGTAGTAGGCTATACCATCTAGGTTTGTGTAAGTACACTCTATGATGTTTGCACAGTGATAAAAATCACCT
AATGACATATTTTTCAGAATGTATTCTTGTCATTAAGTGATGCATGACTGTATAAGTGAAAAGCTGCCTAATGGGAAAAGAGAAAAGGTGGTGGGACAGGAATAAGAAGGCAGGGCAGTGACAGATTTAATTTATGTTCAAGTACTTATGAGCAAAGAGTTTTTTTAA
AAATTATTTTTCATGTAGACCCATTAACTTAGATTGCTTGATGCATTTGAATCCAGATTTTGAAGGGTAATAAGGTAGGTGTTTGGATATGAGCTAGAACTTGGGATAATAGTGTAGATTATGATTACTGGCTATAGTTGATTACCCTTGAACAGCTGTGAAATGAACCTTA
CTAGAAATTATTTTACTATGGTAATAATCACAAATTATTCAGAGAAGTTATTTGAAGTTTCCATAGCTGTTATTCAGTTGTAAGCATGTTTTACTGGGAGAGTAACTTTCAGATGGCAAGTGTTGAAACTGTAACAAACATTAGTTTAAATATTTTTCTATTATGGCATACCTA
GTCTGAGTATTTGGCTTCATTTGCATATAATTTGTTAAGACCAAGGTTTTAAGTTTTTTCGTGATGTTTCAGAATGATATAATTTATGTCTGTGGTGATTTTAGCAATTTAACCCAAAACATTTAAACTCTTGAATCTGTCATCTATTGAAAATAATGTATGTGTGTGTATGTAT
GTAAATATGCACACACATAAATAAAGATGATACTCTTATATTTGCTTTTTCTTTCCTATTGTACAAAATACTACCTGCCTTTTGAATTTAAGCCTTCGTGACACTT

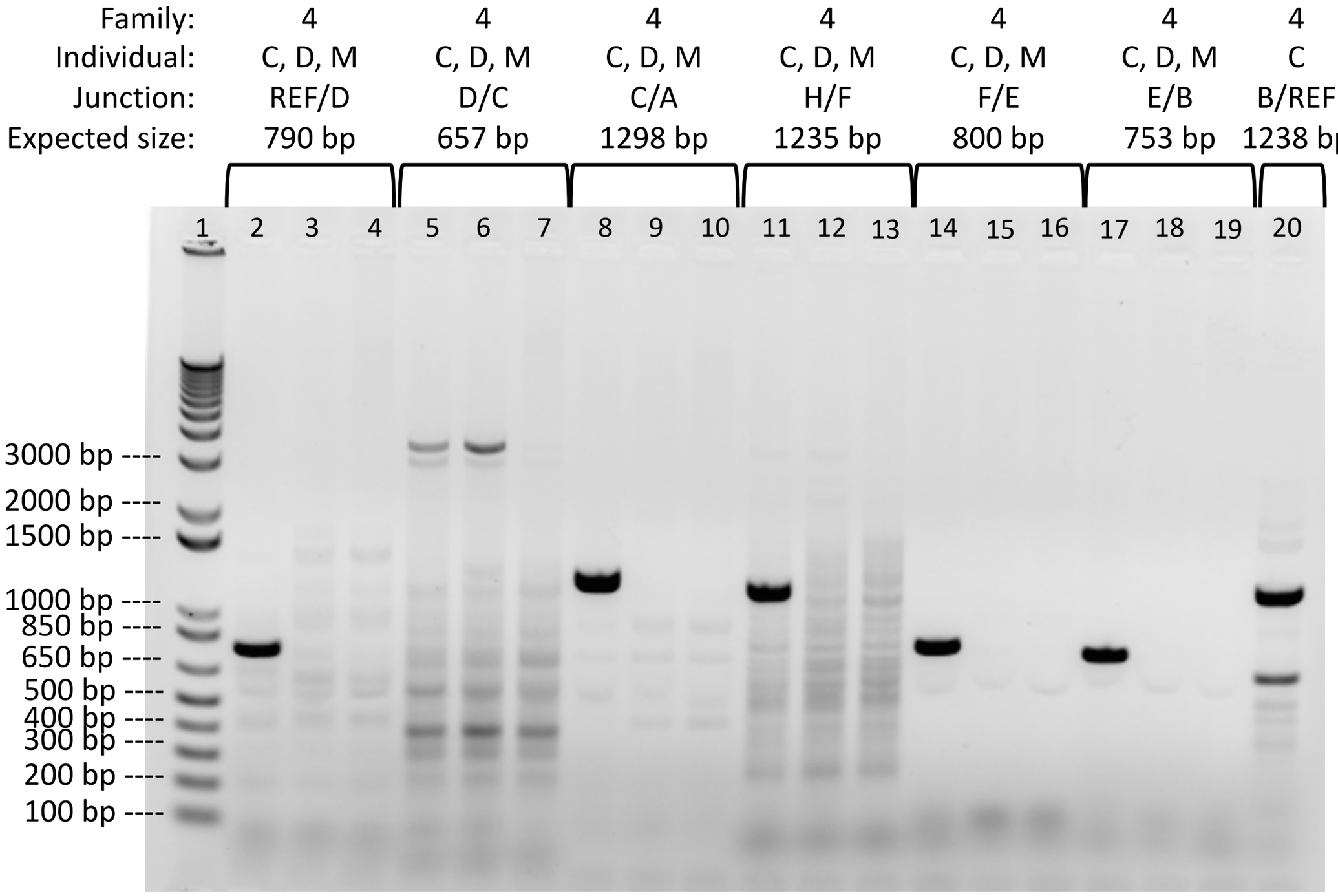


Figure S8. PCR Amplification confirms the majority of the breakpoints in the chromothripsis-like region on chr6 in Proband 4. A 1% agarose gel was run with 1 ug Invitrogen 1 Kb Plus DNA Ladder in Lane 1. Lanes 2-19 are sets of three PCR samples; each set of 3 represents child (C), dad (D), and mom (M) as PCR template. Each set of 3 PCRs spans across the junction listed, and should result in a band of the expected size listed if the rearrangement is present (i.e., only in C). Lane 20 is C only; D and M are present in Figure S9. Note that PCR for junction D/C resulted in nonspecific bands only.

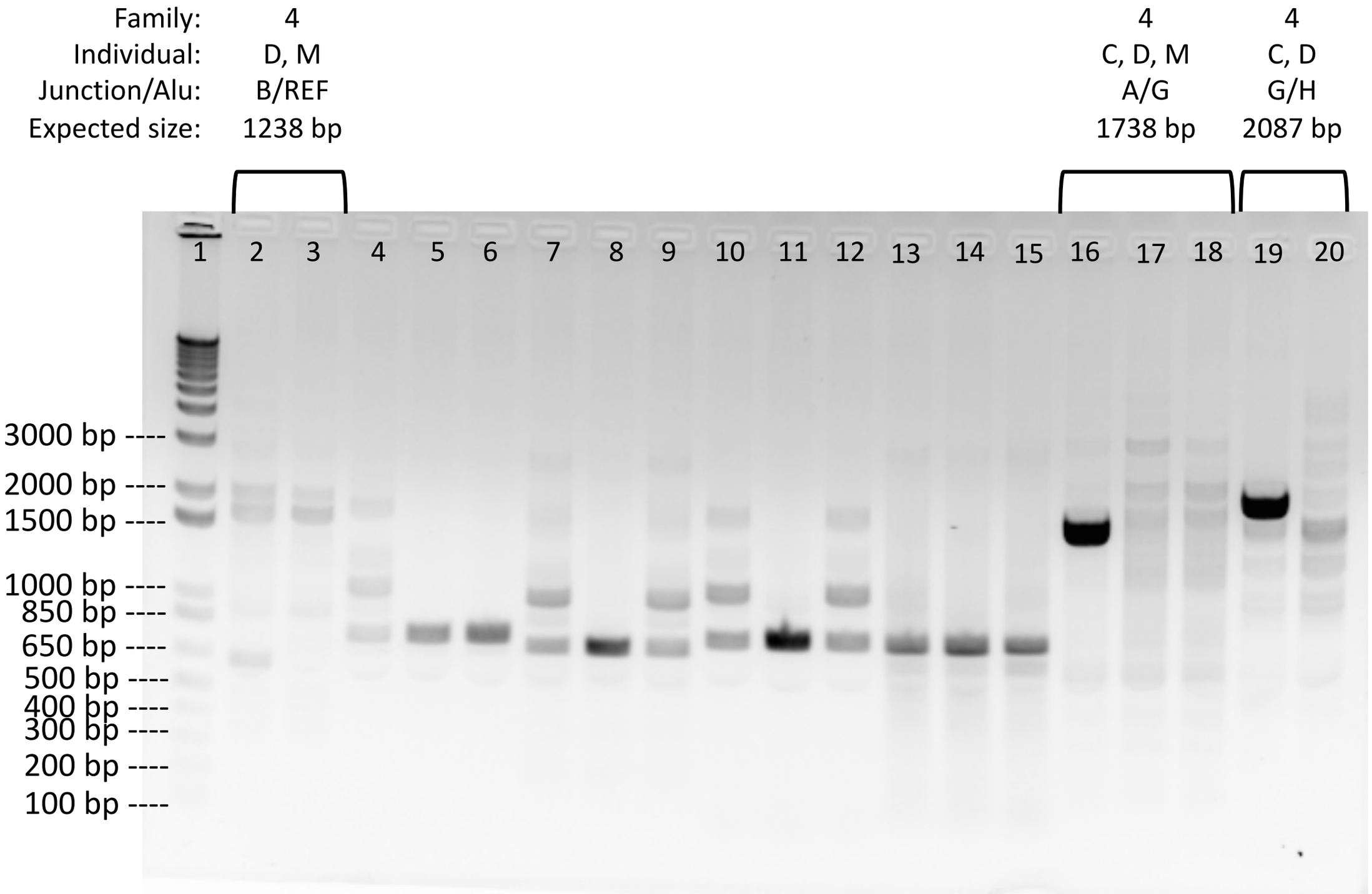
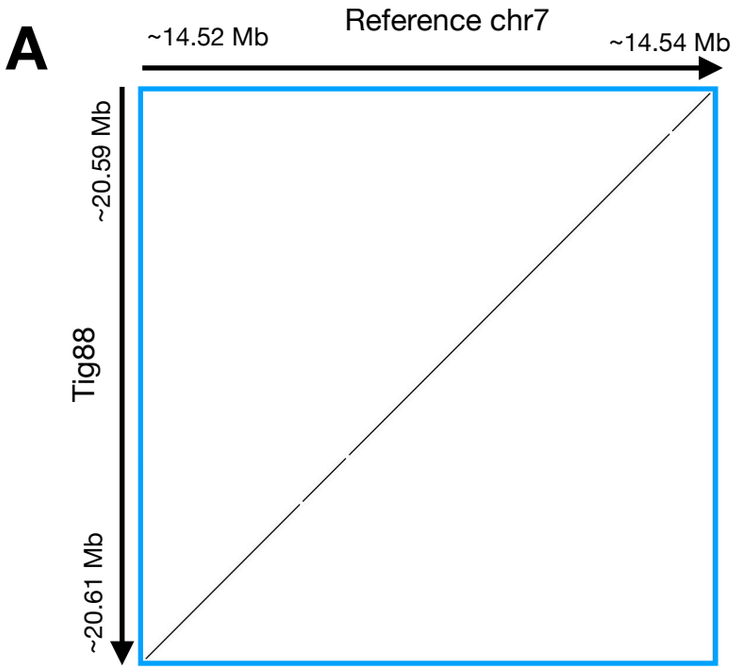
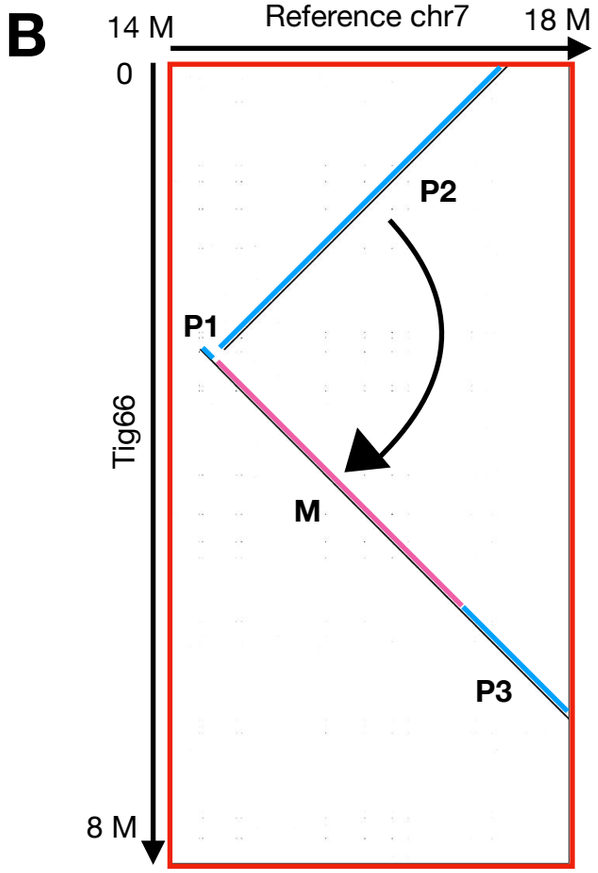
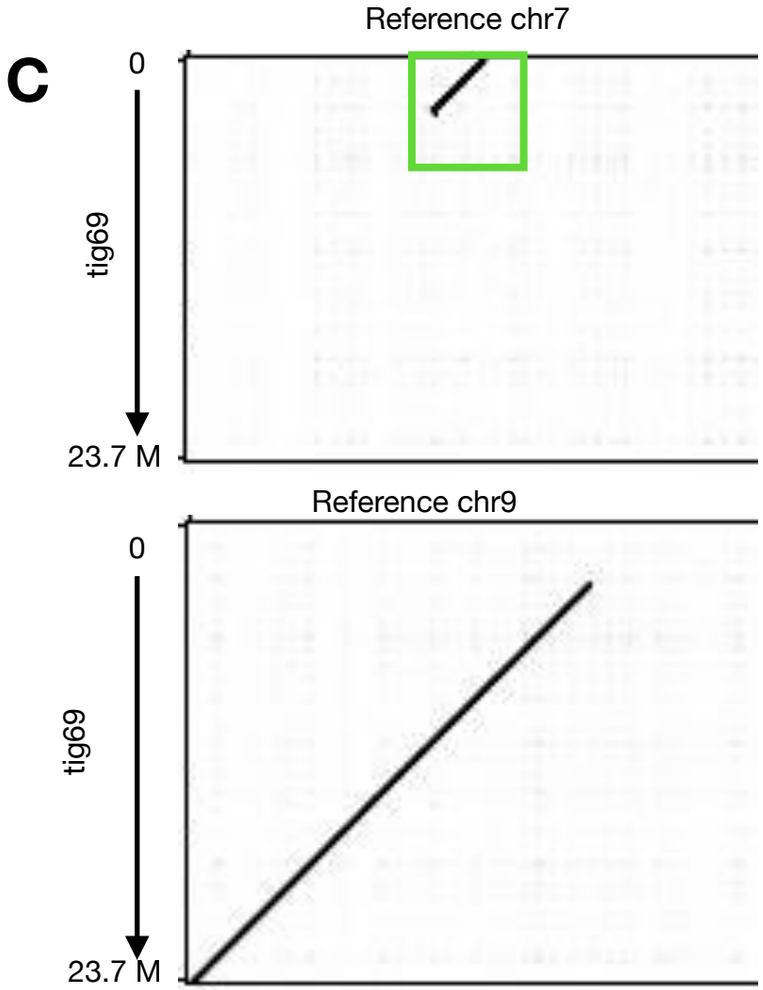


Figure S9. PCR Amplification confirms the majority of the breakpoints in the chromothripsis-like region on chr6 in Proband 4. A 1% agarose gel was run with 1 ug Invitrogen 1 Kb Plus DNA Ladder in Lane 1. Sets of PCR samples represent child (C), dad (D), and mom (M) as PCR template as noted. Note that Lanes 2 and 3 are D and M only; C is present in Figure S8. Note that this gel image is duplicated as Figure S3 and Lanes 4-15 are described there.







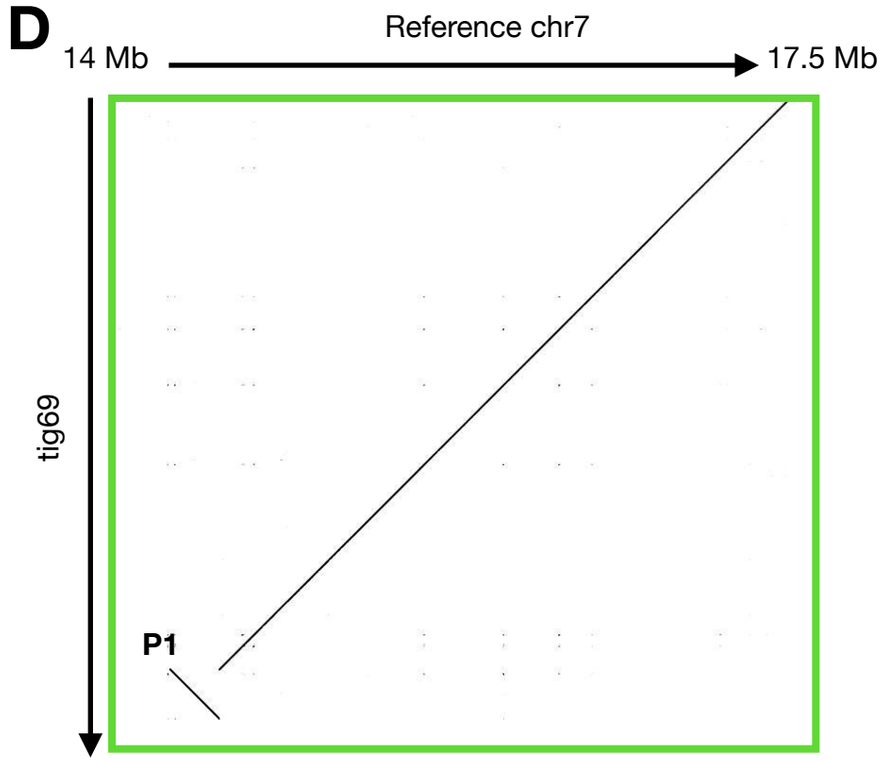
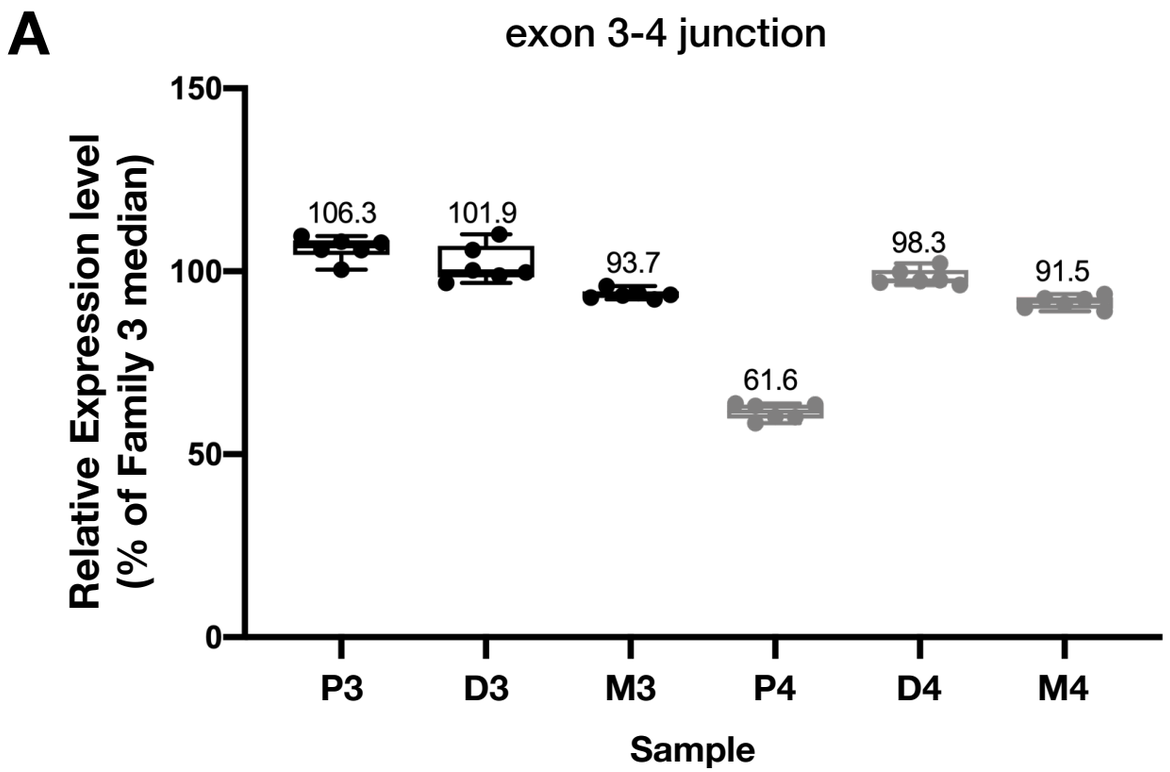
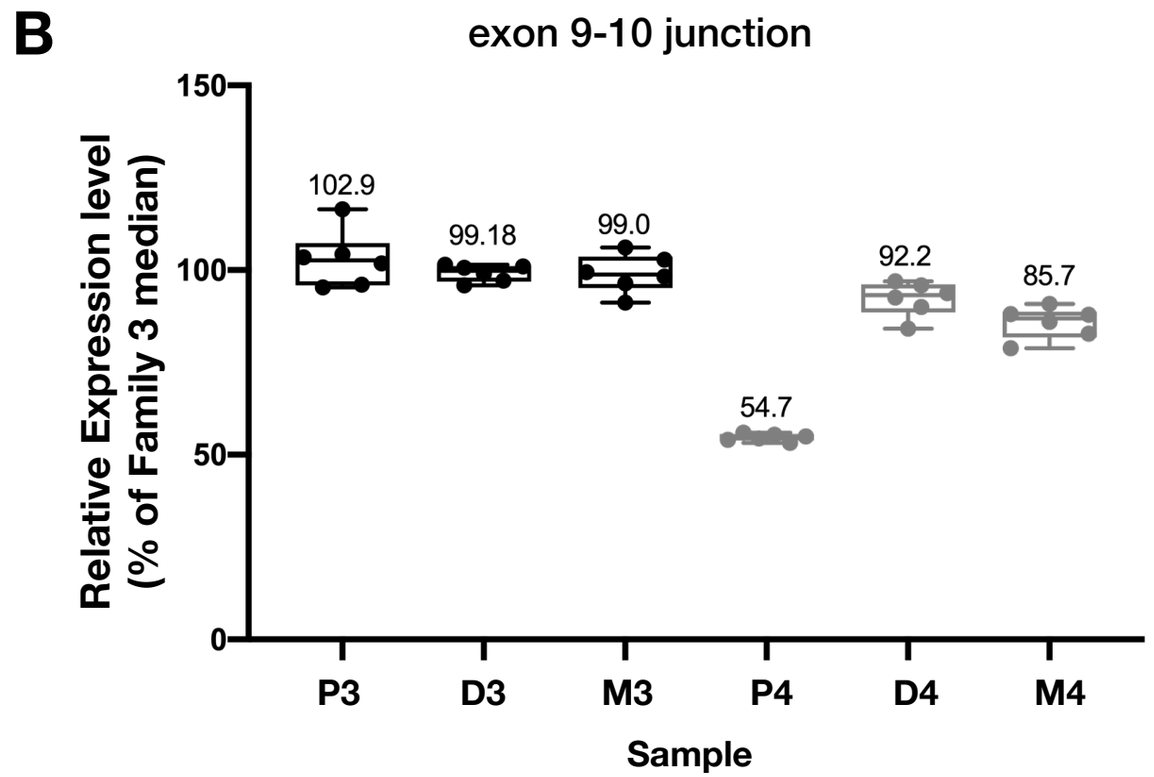
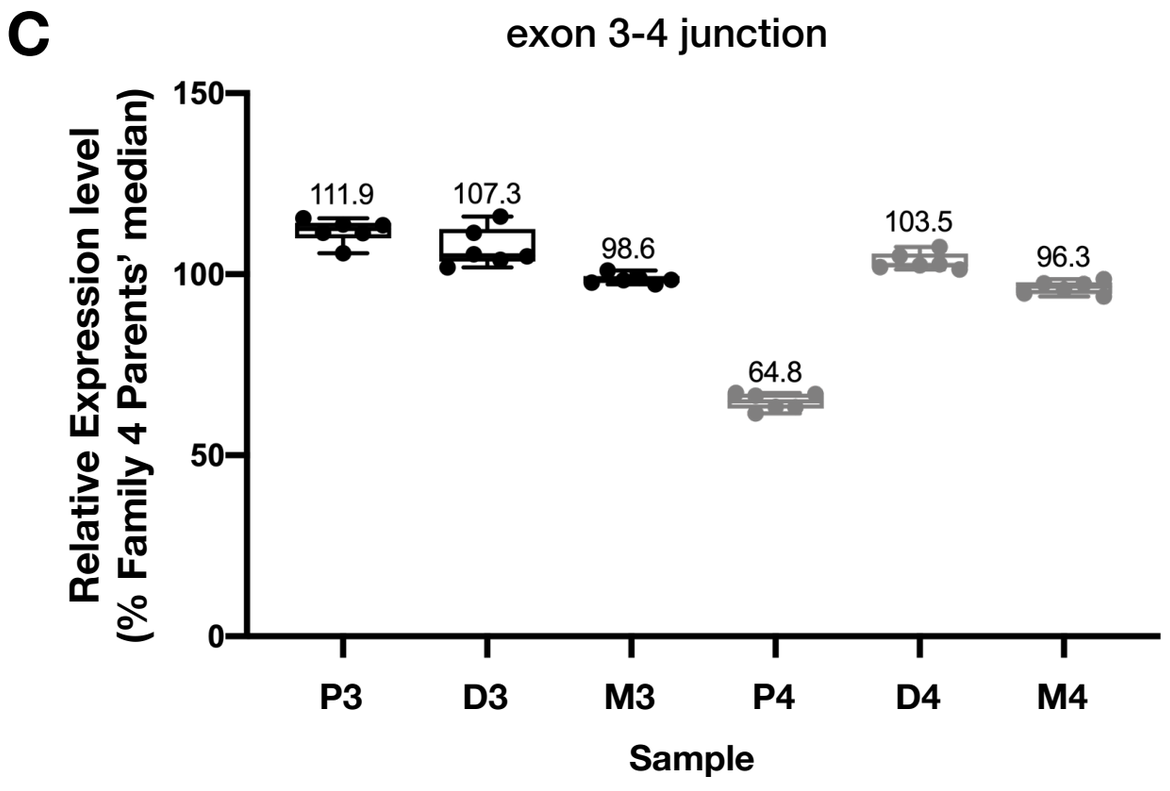


Figure S10. Manual curation of an assembly artifact was required for PAT_tig66. **A.** Zoomed in view of blue box from Main Figure 4C. This region shows the alignment at the 3' end of tig88 to chromosome 7, and the sequence is inverted compared to the neighboring, upstream chr9 sequence. **B.** Zoomed in view of red box from Figure 4C. Manual inspection of this paternally-binned contig revealed both maternal and paternal sequence. P1, P2 and P3 represent paternal sequences, M represents maternal sequence. If the maternal sequence is removed, and P2 is aligned contiguously with P3, P1 represents an inversion with respect to these sequences. **C**. A contig (tig69) from same region from the non-trio-binned hifiasm assembly does not contain this error. **D** Zoomed in region from C. Tig69 aligned to chromosome 7 sequence shows the inversion (P1).







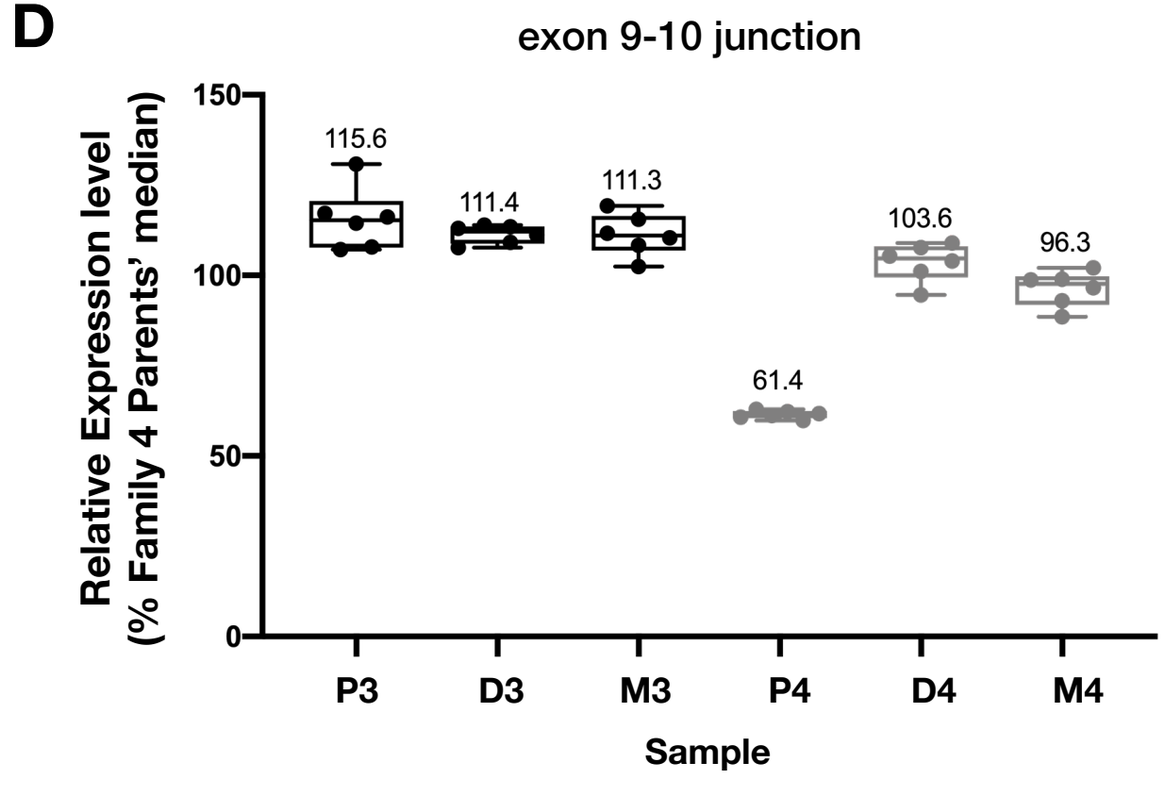


Figure S11. *MLLT3* shows decreased expression in Proband 4 (P4). qRT-PCR using TaqMan probes targeting the *MLLT3* exon 3-4 (**A, C**) and exon 9-10 (**B, C**) splice junctions, normalized to either the median of Family 3 values (**A, B**), or the median of Family 4 Parent values (**C, D**). Samples include Proband (P), Dad (D), and Mom (M) from two different trios in this study, 3 and 4.

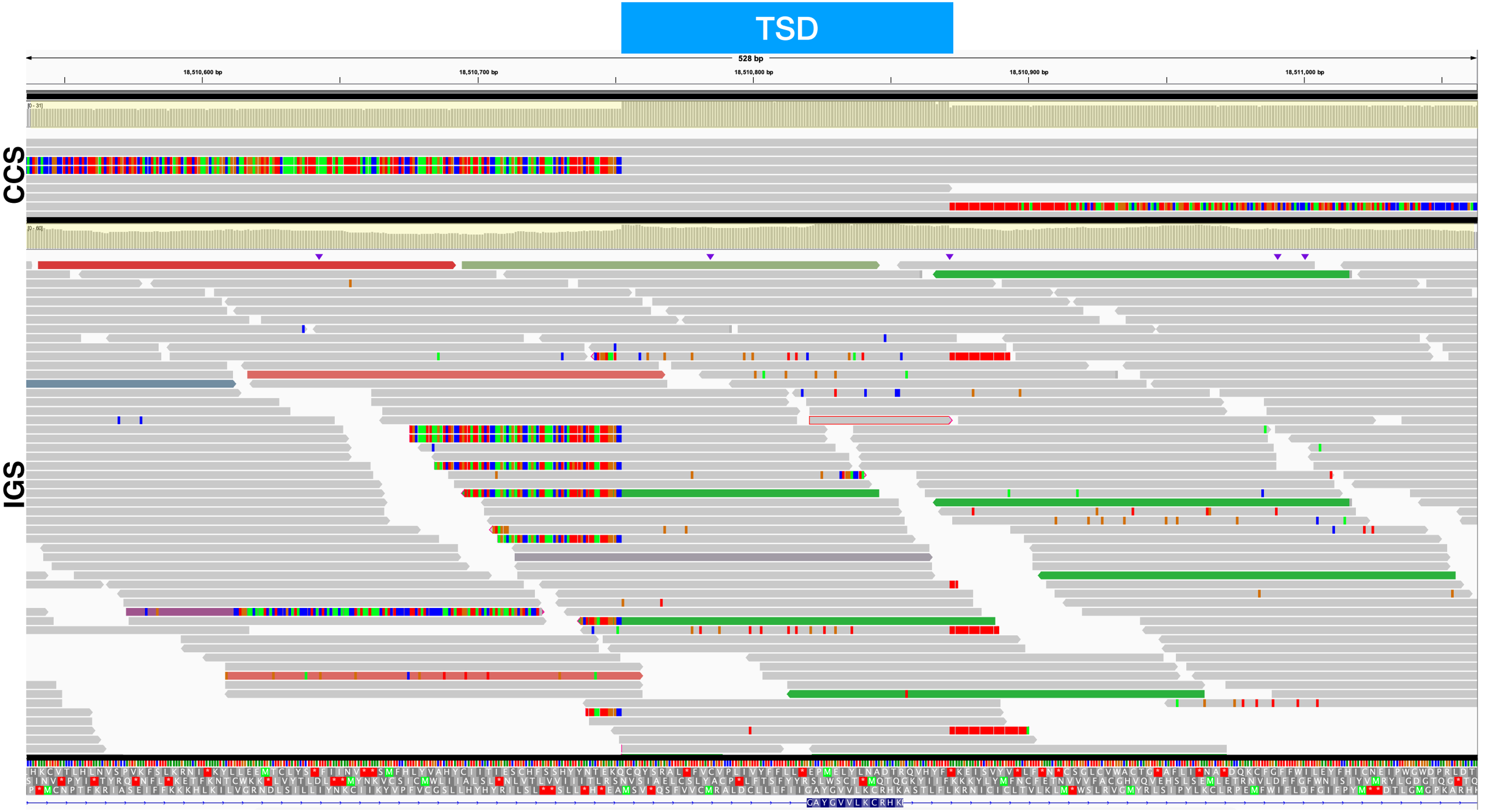


Figure S12. Visualization of a subset of CCS and IGS reads in Proband 6 at the L1-mediated insertion site in *CDKL5*. CCS reads consist of those representing reference allele (gray), or reads with one unaligned end (multicolor), similar to those in Figure 1 of the main text. Note that CCS reads show a sharp increase in read depth (highlighted by the yellow box), spanning the target site duplication (TSD, blue box). IGS reads also include reference alleles (gray), reads with unaligned ends representing the insertion (multicolor), and reads whose mapped pairs indicate a longer than expected insert length (green). There is a slight increase in read depth in the TSD, but the edges are not sharply demarcated.

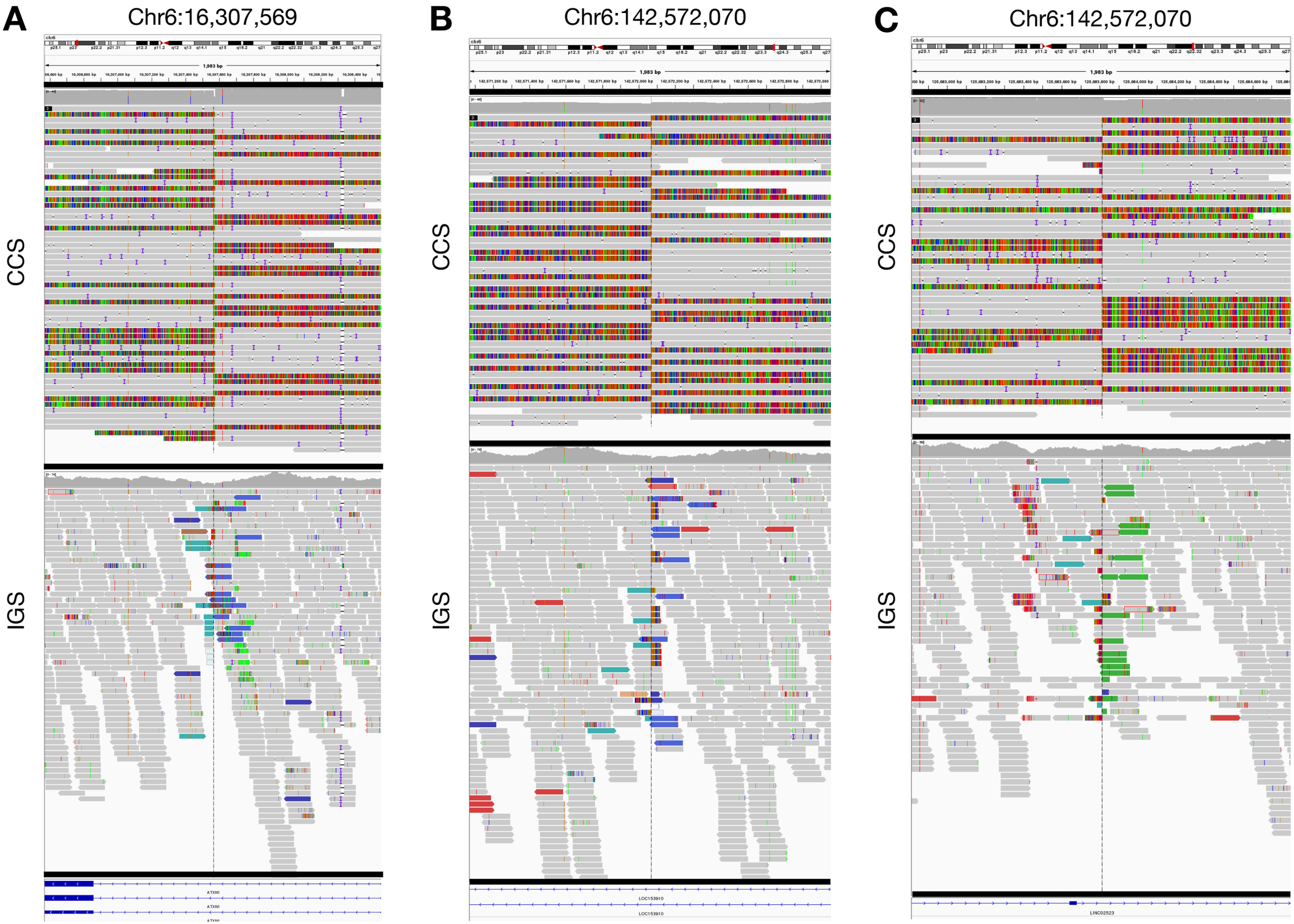


Figure S13. Visualization of a subset of CCS and IGS reads in Proband 4 on chromosome 6, including the two sites of the pericentric inversion (A,B), and a representative breakpoint in the complex chromothripsis-like region (C). In both CCS and IGS, gray bars represents alignment of the read to reference and multicolored segments represent unaligned sequence. The shorter IGS reads also include reads whose mapped pairs indicate a longer than expected insert length (blue, turquoise or green reads), suggesting a structural variant.

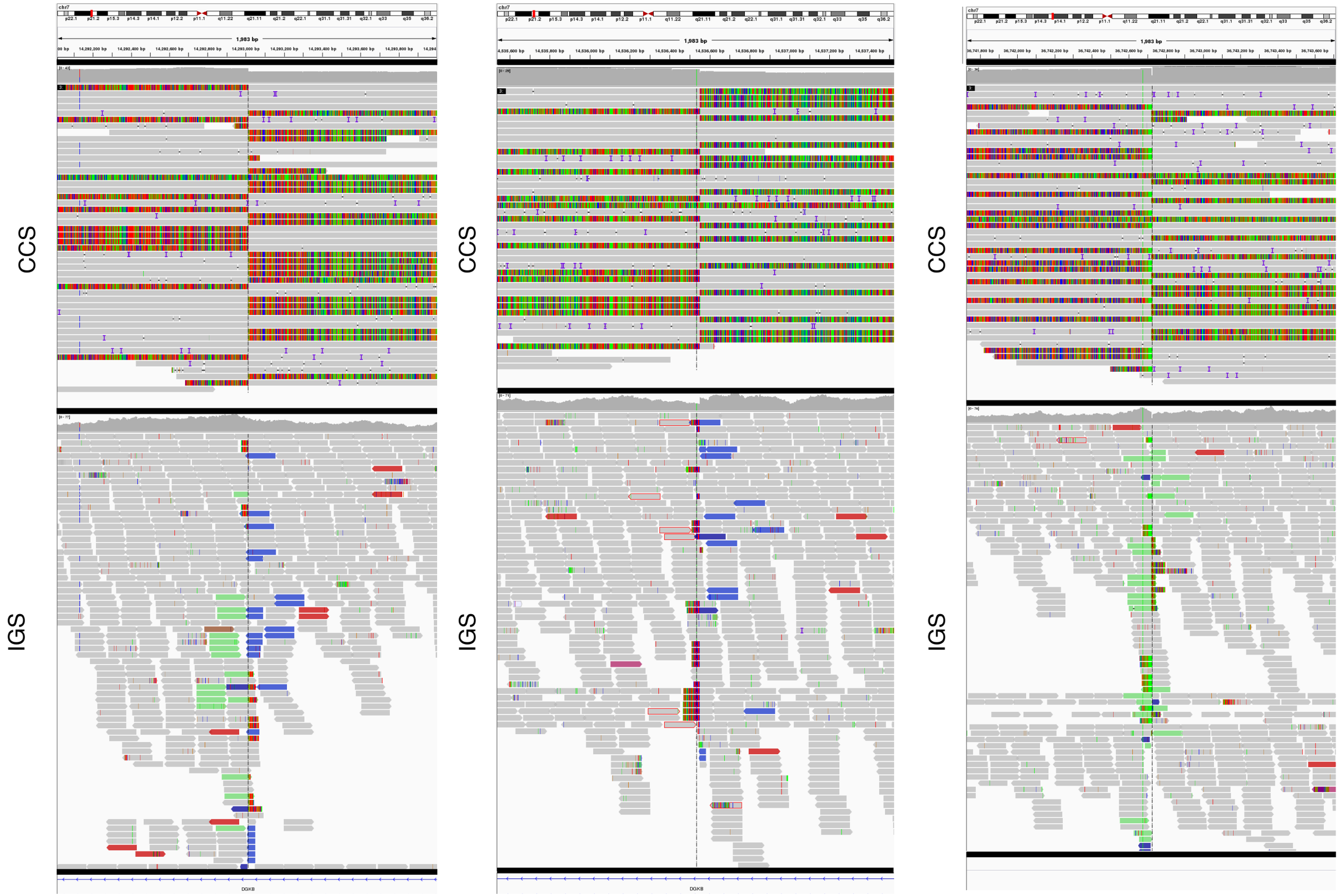


Figure S14. Visualization of a subset of CCS and IGS reads in Proband 4 at the three breakpoints on chromosome 7. In both CCS and IGS, gray bars represents alignment of the read to reference and multicolored segments represent unaligned sequence. The shorter IGS reads also include reads whose mapped pairs indicate a longer than expected insert length or a mate pair on a different chromosome (blue or green reads), suggesting a structural variant.

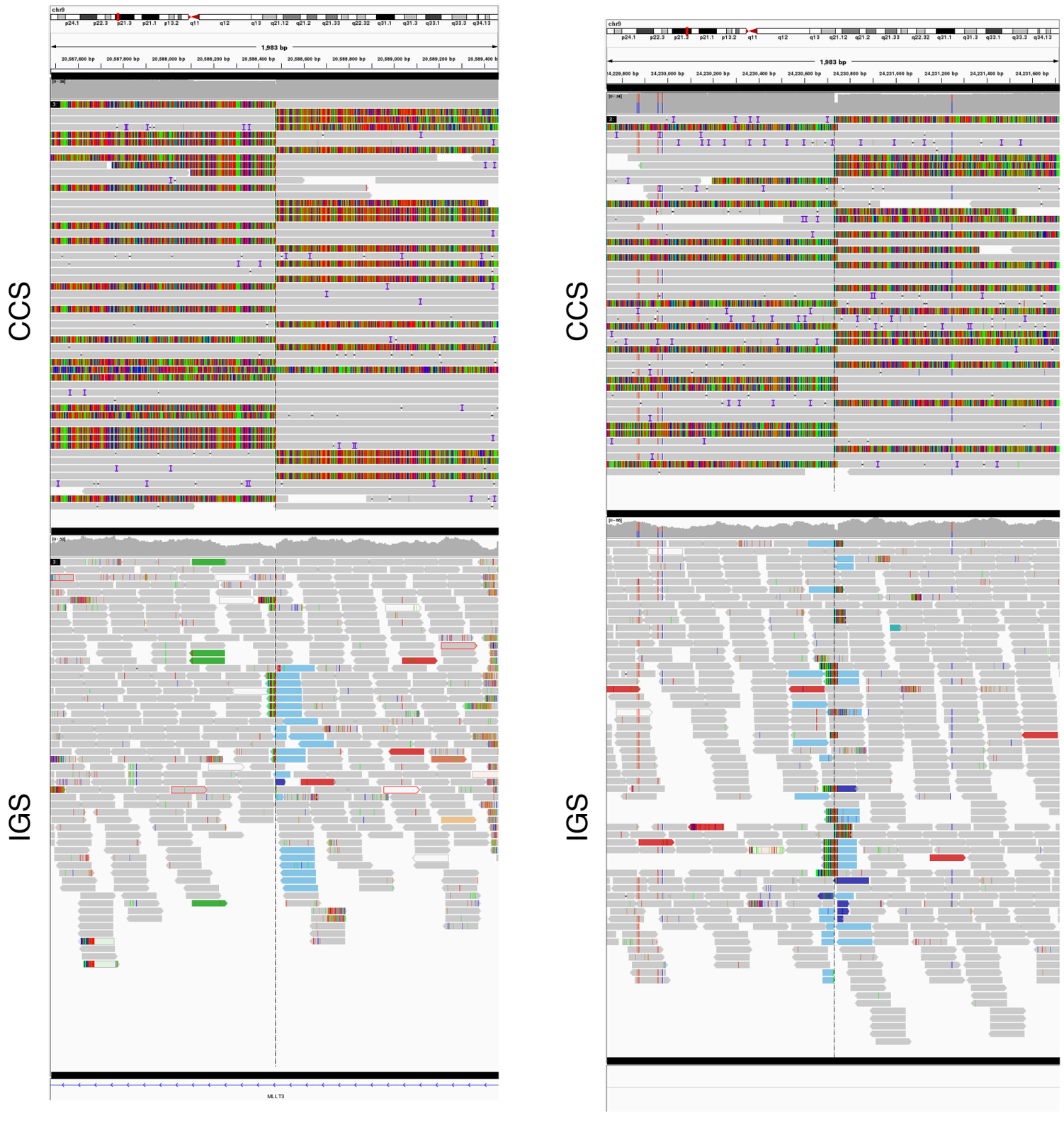


Figure S15. Visualization of a subset of CCS and IGS reads in Proband 4 at the two breakpoints on chromosome 9. In both CCS and IGS, gray bars represents alignment of the read to reference and multicolored segments represent unaligned sequence. The shorter IGS reads also include reads whose mapped pairs indicate a longer than expected insert length or a mate pair on a different chromosome (blue or green reads), suggesting a structural variant.

Proband

Proband Allele Ratios

Father

Mother

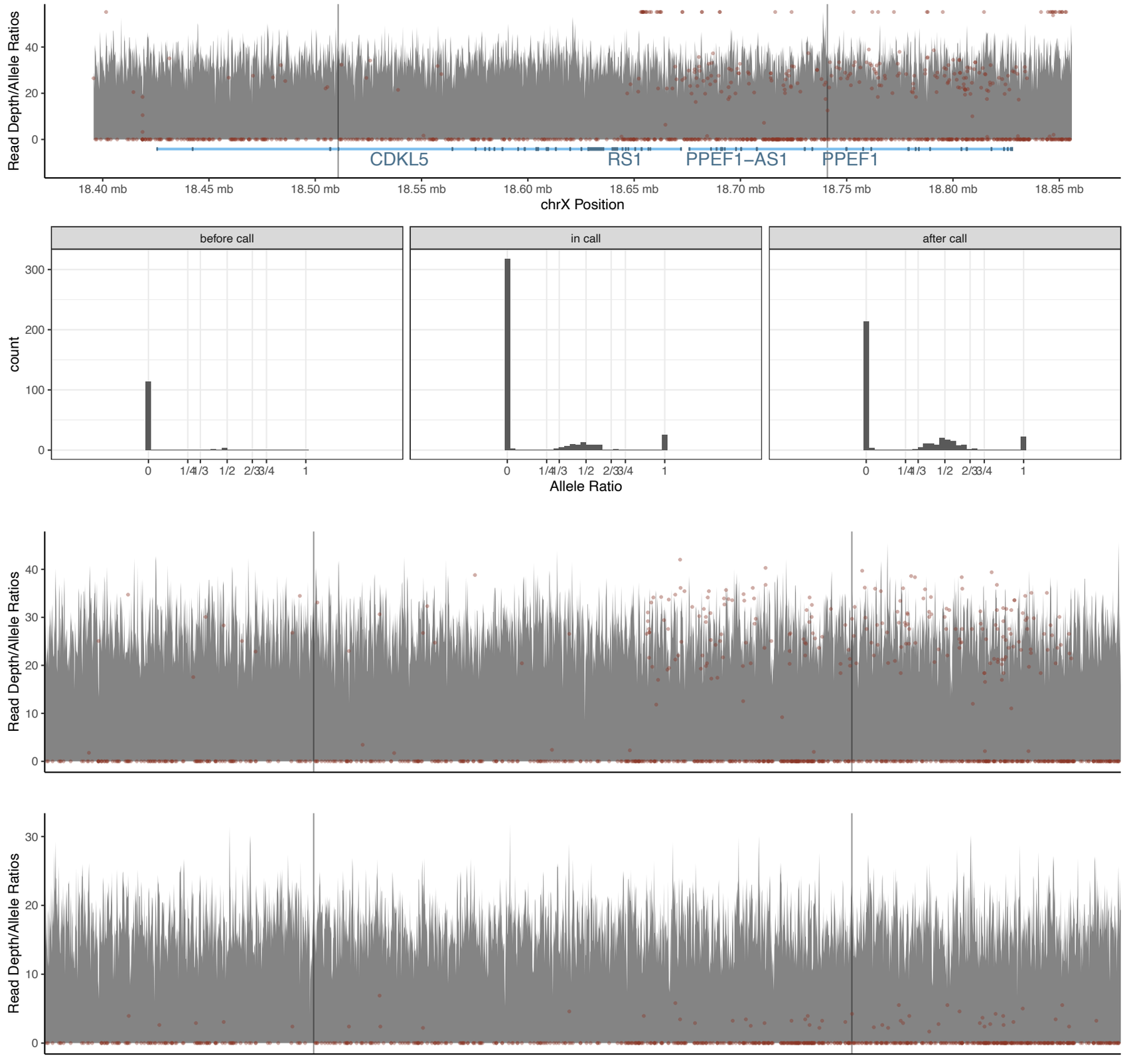
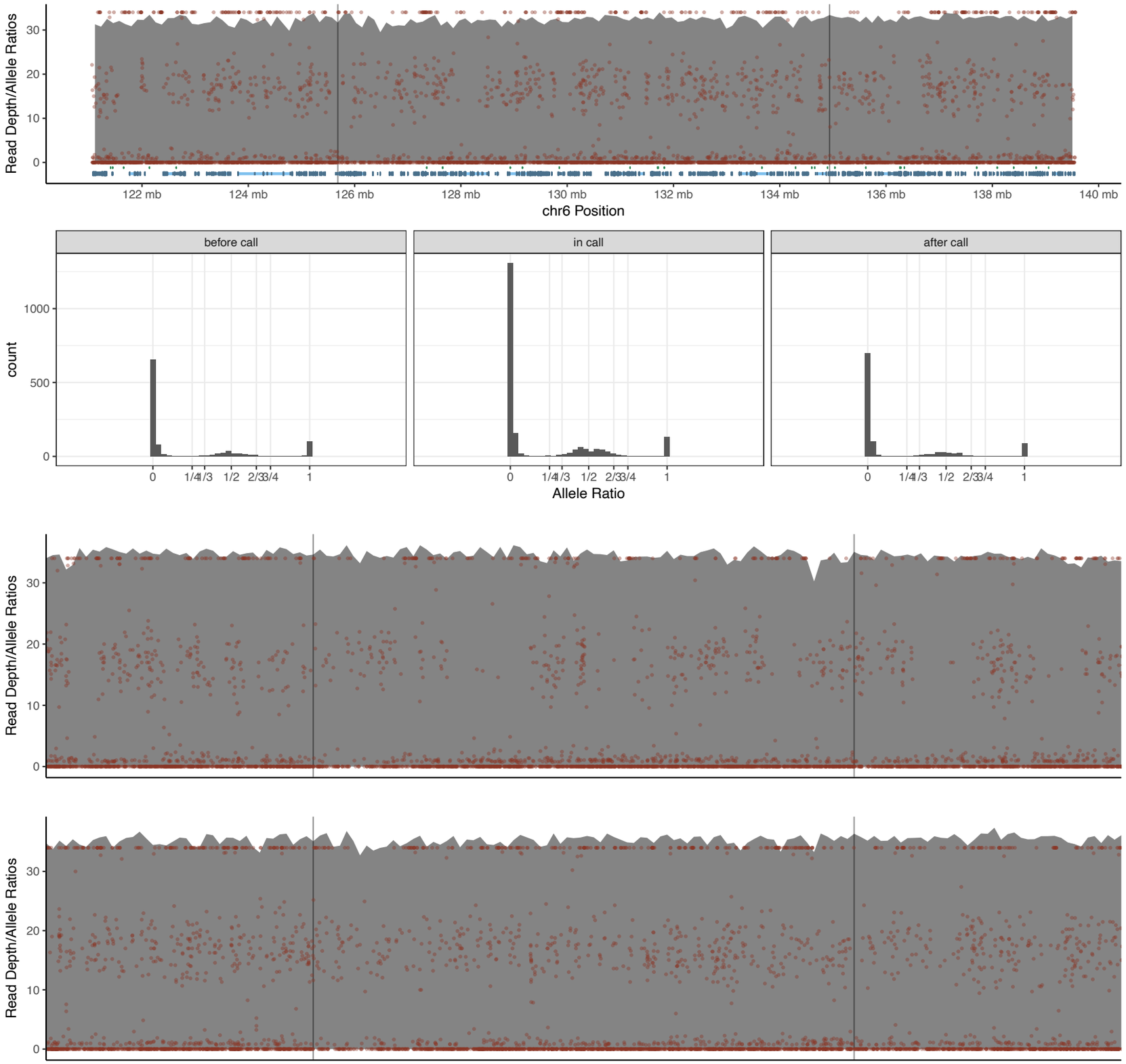


Figure S16. Plots of read depth (gray peaks) and allele ratios (red dots, and plot in second panel from top) in IGS data surrounding the 230 kb duplication event called by Delly and Manta. Called breakpoints are indicated by black vertical lines. Neither depth nor allele ratios support a duplication in the proband. Note that similar patterns are observed for proband, mother and father across the region.

CDKL5 exon 3

Figure S17. Several *de novo* structural variants have been reported in *CDKL5*, each with a breakpoint in intron 3 (gray box), near the breakpoint identified here. B is a zoomed in view of A (black box).

Proband



Father

Mother

Figure S18. The complex rearrangements at 6q22.31-6q23.3 appear to be copy-neutral. Plots of read depth (gray peaks) and allele ratios (red dots and plots in second panel from top) in IGS data show similar patterns for proband, mother and father across the region.

Table S3. Range of insertion sizes for *FMR1* 5' UTR variation compared to GRCh38, based on visualization of reads in IGV.

Family ID	Individual	Gender	Allele 1, nt insertion	Allele 2, nt insertion
1	C	F	75	63-66
	D	M	NA	63
	M	F	72	66
2	C	F	33	27
	D	M	NA	27
	M	F	33	6
3	C	M	30	NA
	D	M	NA	66
	M	F	27-30	27-30
4	C	F	39-42	33
	D	M	NA	33
	M	F	36-39	36-39
5	C	M	30	NA
	D	M	33	NA
	M	F	105	30
6	C	F	69	9

	D	M	NA	9
	M	F	69	45

C, child; D, dad; M, mom; NA, not applicable for hemizygous males.

Table S8. Pathogenic or likely pathogenic structural variants that have been reported near intron 3 of *CDKL5.*

Publication	Proband in Pub	detected by	GRCh38 Coordinates	Flanking sequence	Description	Inheritance
This Study	6	CCS	chrX:18510871	AluSx/Tigger3a/4b	LINE/*PPEF1* insertion and exon 3 dup	*de novo*
Erez, et al.[a]	1	array	chrX:18369553-18526954	AluSq; AluSp	157409 bp del, removes exon 1-3	unknown
Erez, et al. [a]	2	array	chrX:18432728-18570428	AluJb; unique	137701 bp del; removed exons 1-4	*de novo*
Erez, et al. [a]	3	array	chrX:18439733-18514587	AluSx; AluSq	74869 bp del; removes exons 1-3	*de novo*
Bartnik et al.[b]	1	array	chrX:18406582-18521160	L1MB5; unique	114579 bp del; removes exons 1-3	*de novo*
Bartnik et al. [b]	2	array	chrX:18564194-18564780	unknown	exon 4 del (confirmed by RT-PCR)	*de novo*

Cordova-Fletes, et al.[c]	1	karyotype; FISH; RT-PCR; array painting	chrX:18519860-18532094	unknown	t(X;2)(p22.1;p25.3)	unknown
Sanchis-Juan, et al.[d]	4	WGS; ONT long reads	chrX:17774889-18055885	Alu	280 kb dup	*de novo*
Sanchis-Juan, et al.[d]	4	WGS; ONT long reads	chrX:17774889-18514192	Alu	458 kb inversion	*de novo*
Sanchis-Juan, et al.[d]	4	WGS; ONT long reads	chrX:18230835-18514192	Alu	283 kb dup	*de novo*

[a] PMID:19471977, [b] PMID:21293276, [c] PMID:19807736, [d] PMID:30526634