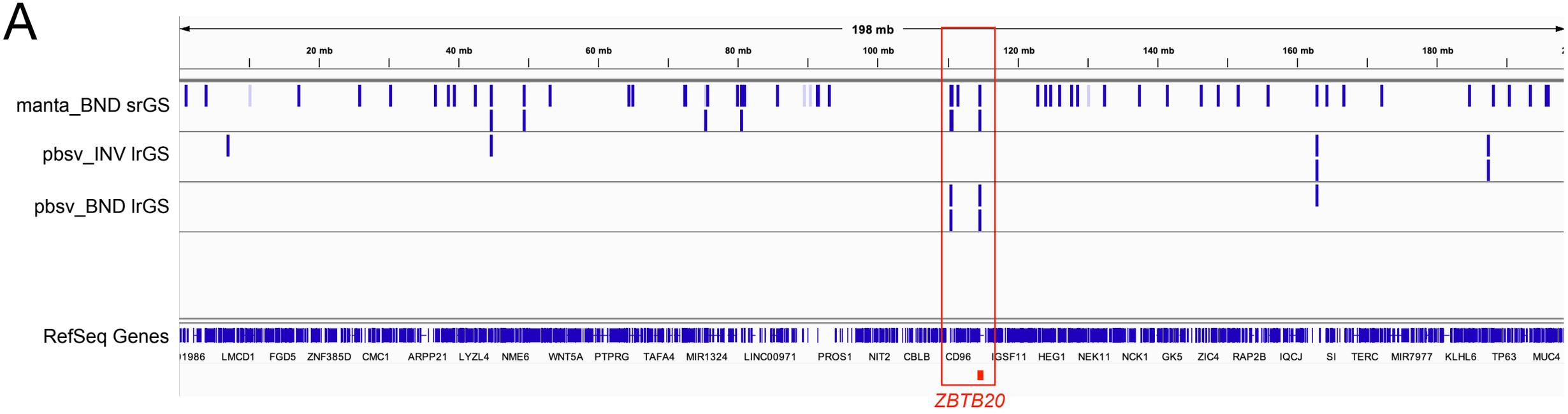
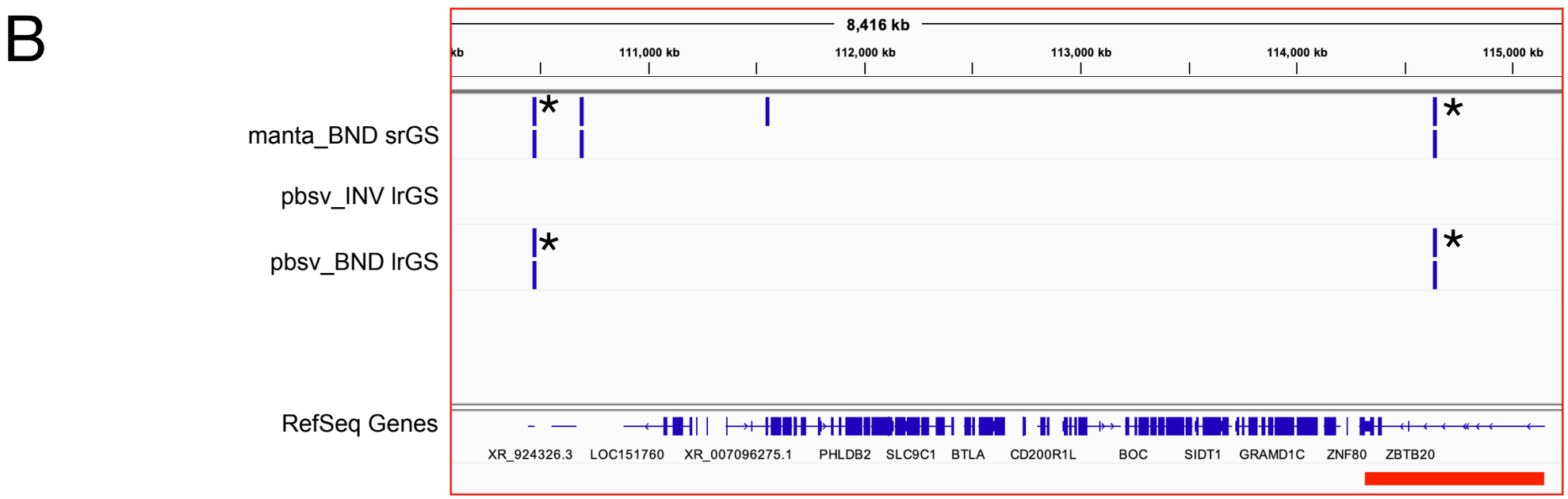
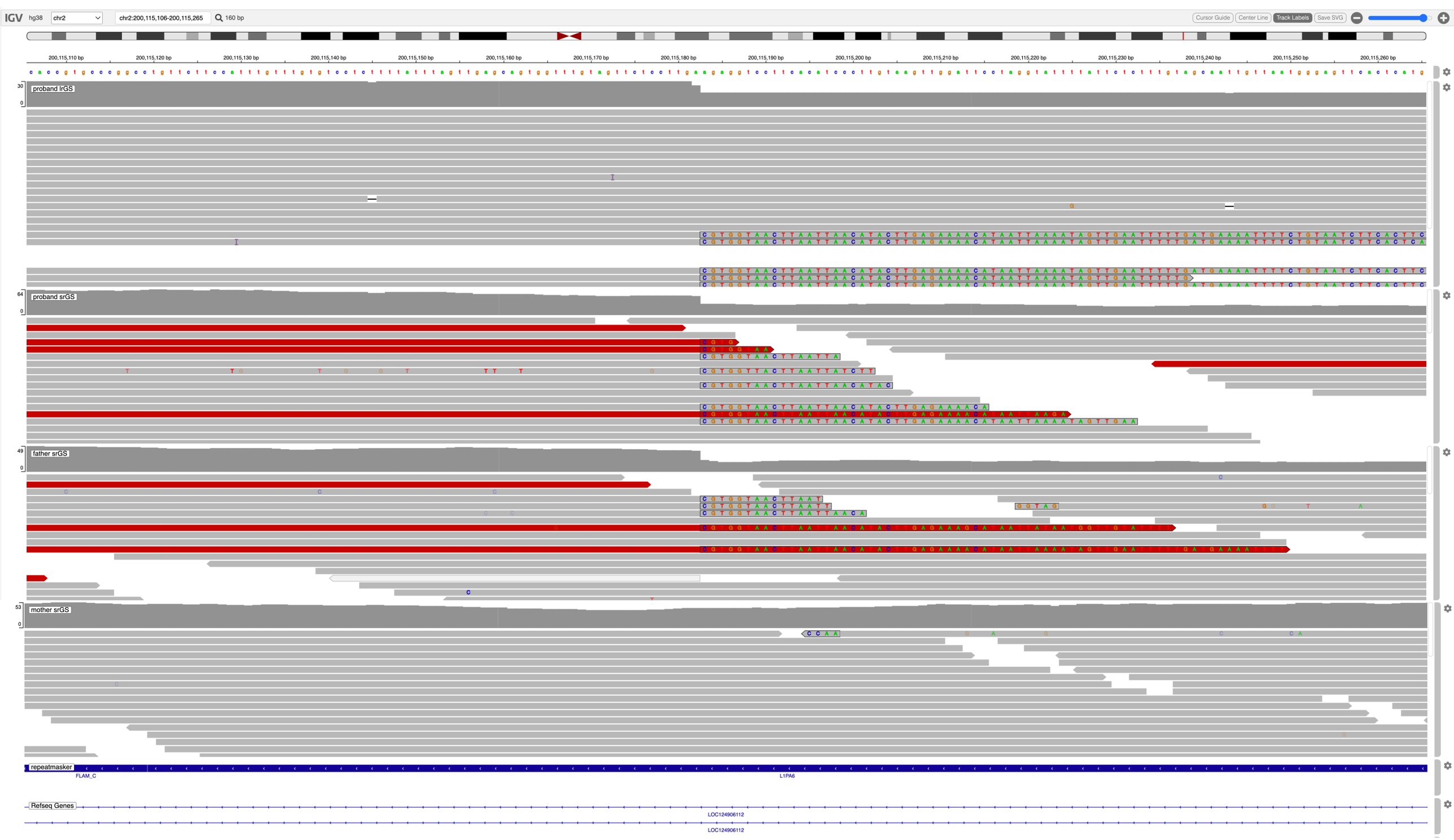
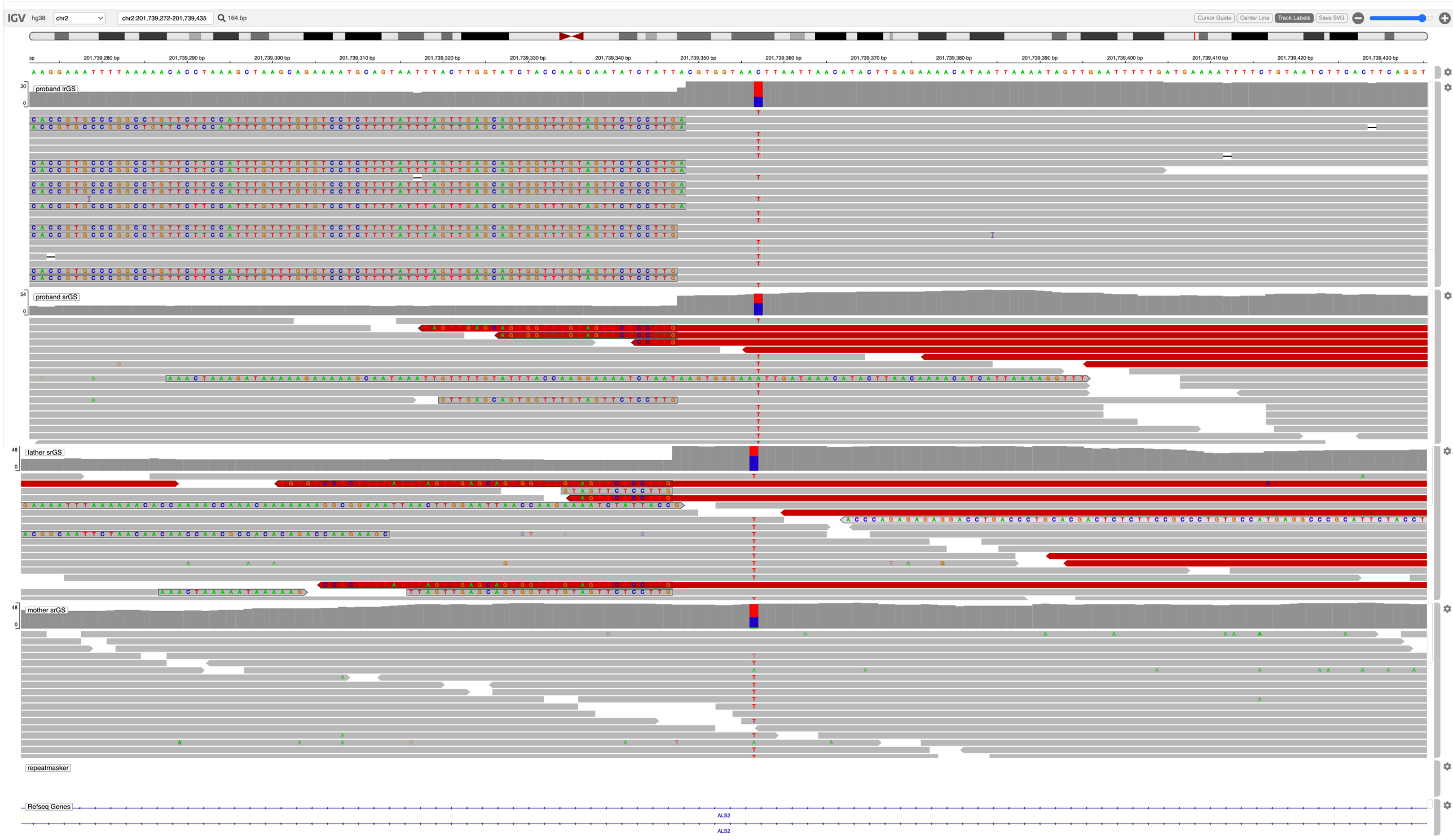




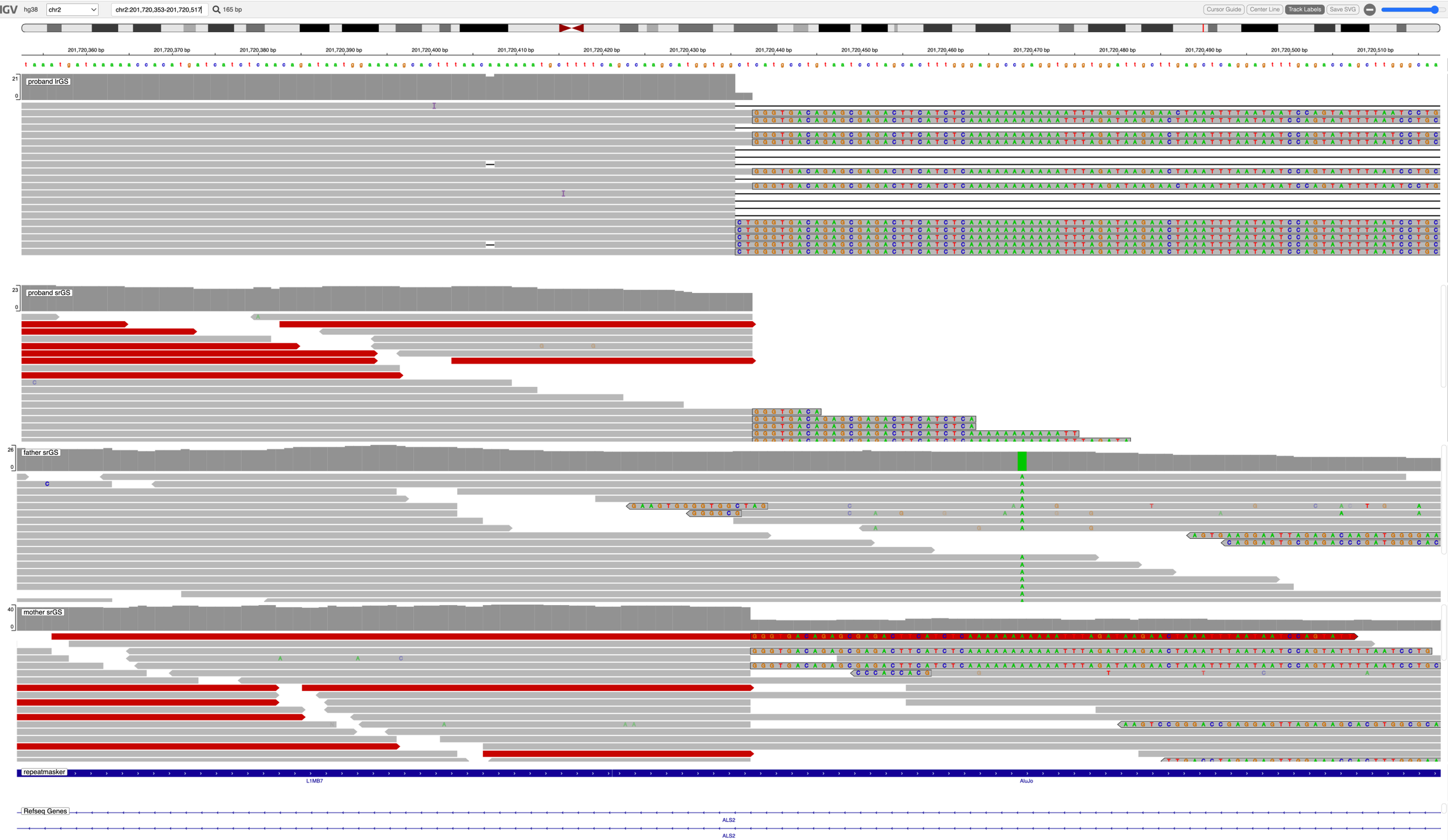
Supplemental Figure S1. Manta and pbsv calls in Proband 1. A. IGV tracks comparing breakend (BND) and inversion (INV) calls on Chromosome 3 between srGS (top track) and lrGS (second and third tracks). PASS variants are shown in dark blue; filtered variants are shown in light blue. The *ZBTB20* gene is marked in red below the RefSeq genes track. Top panel (A) shows the entire Chromosome 3, with many calls from manta in srGS. Bottom panel (B) shows the zoomed area around the inversion event. The inversion is clearly called by two pairs of breakends (*) at the same location in both the lrGS and srGS data.



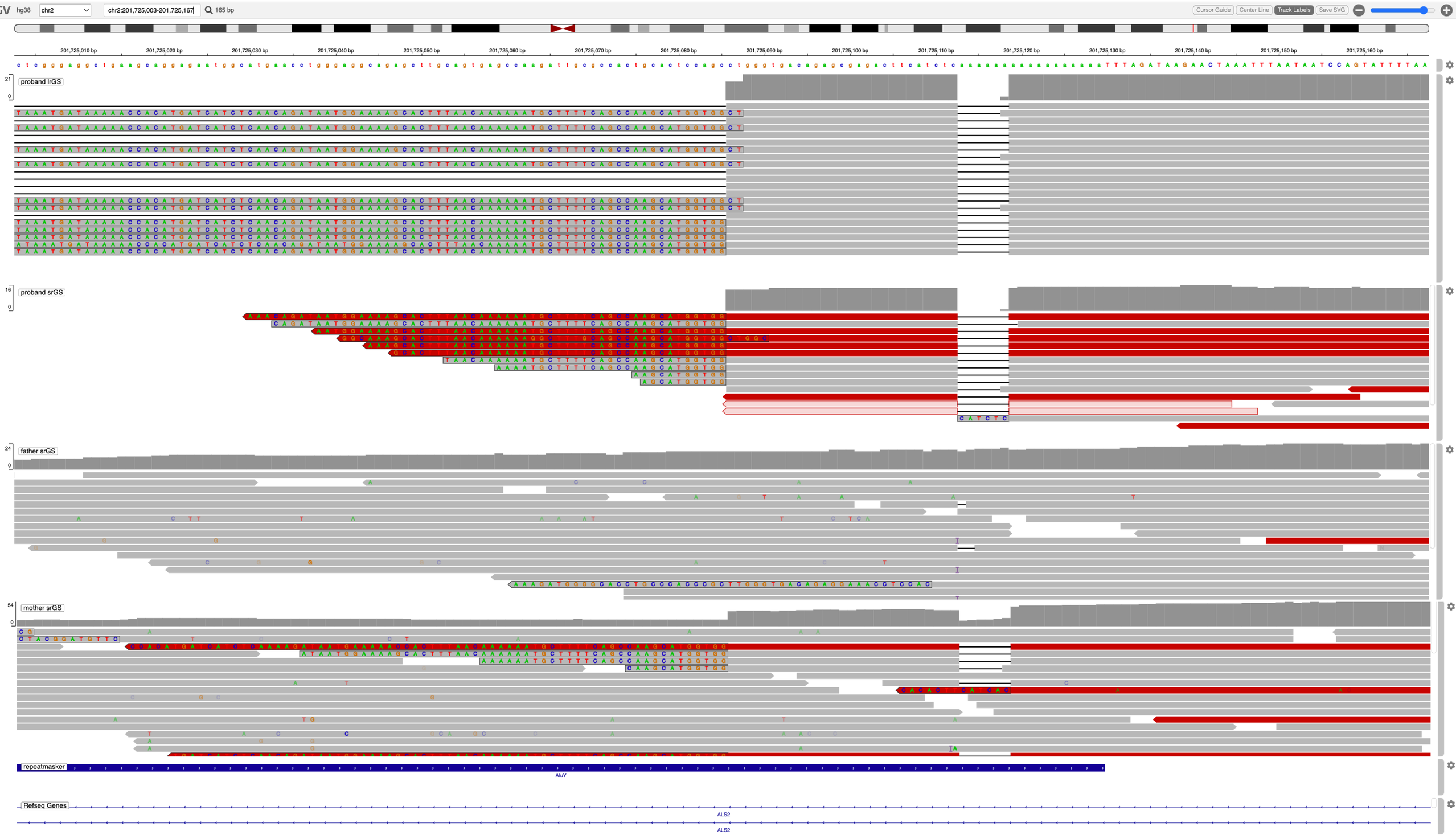
Supplemental Figure S2. Visualization of reads at one of the breakpoints of the *ALS2* SVs shown in Figure 2 (Chr2:200,115,106-200,115,265). A subset of reads for Proband 1 (both lrGS and srGS) and parent srGS in IGV are shown.



Supplemental Figure S3. Visualization of reads at one of the breakpoints of the *ALS2* SVs shown in Figure 2 (Chr2:201,739,272-201,739,435). A subset of reads for Proband 1 (both lrGS and srGS) and parent srGS in IGV are shown.



Supplemental Figure S4. Visualization of reads at one of the breakpoints of the *ALS2* SVs shown in Figure 2 (Chr2:201,720,353-201,720,517). A subset of reads for Proband 1 (both lrGS and srGS) and parent srGS in IGV are shown.

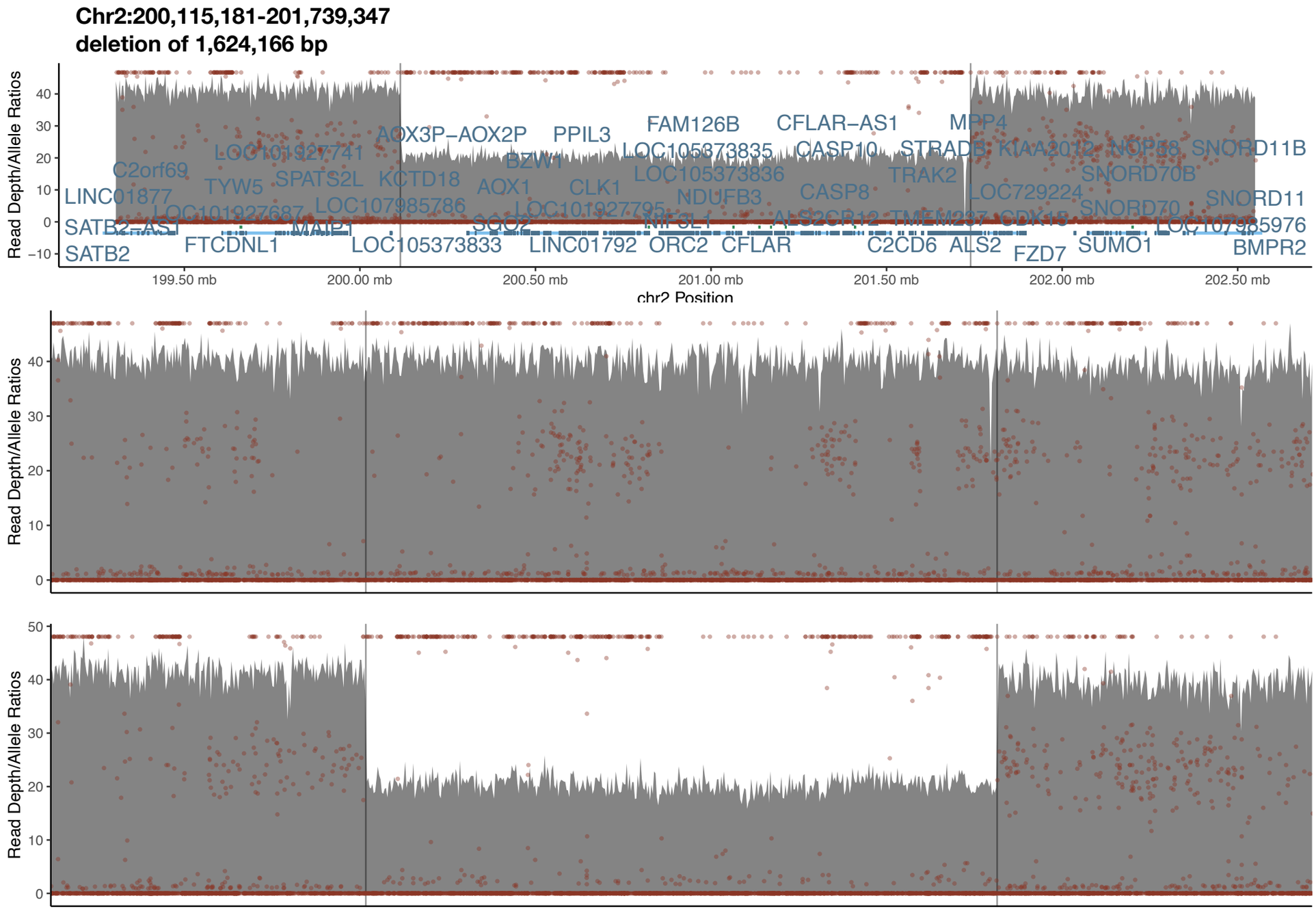


Supplemental Figure S5 Visualization of reads at one of the breakpoints of the *ALS2* SVs shown in Figure 2 (Chr2:201,725,003-201,725,167). A subset of reads for Proband 1 (both lrGS and srGS) and parent srGS in IGV are shown.

Proband srGS

Mother srGS

Father srGS

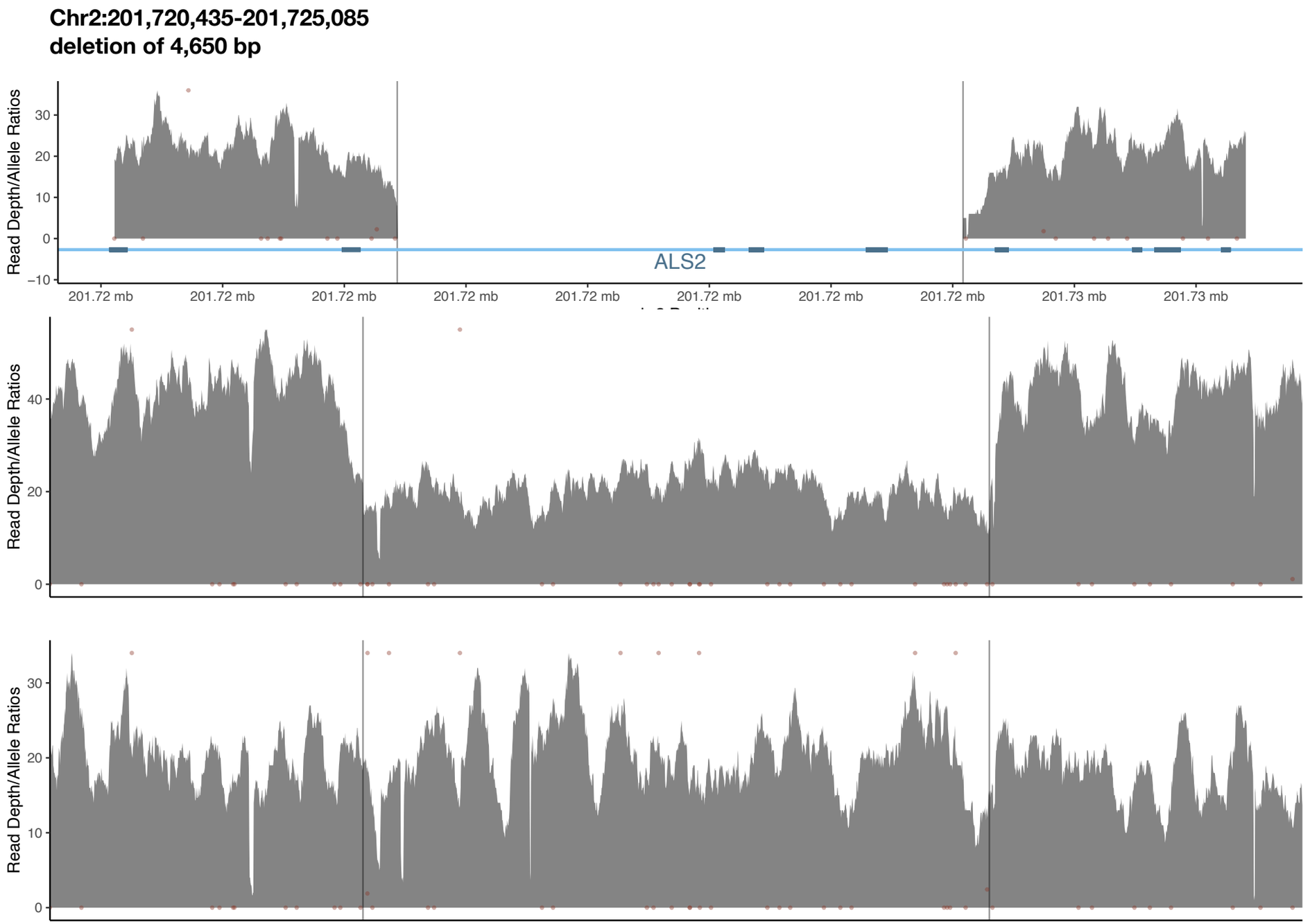


Supplemental Figure S6. Plot of CNV call in Proband 2 srGS. Panels show read depth (gray bars) and allele ratios (red dots) in Proband, Mother and Father's srGS data.

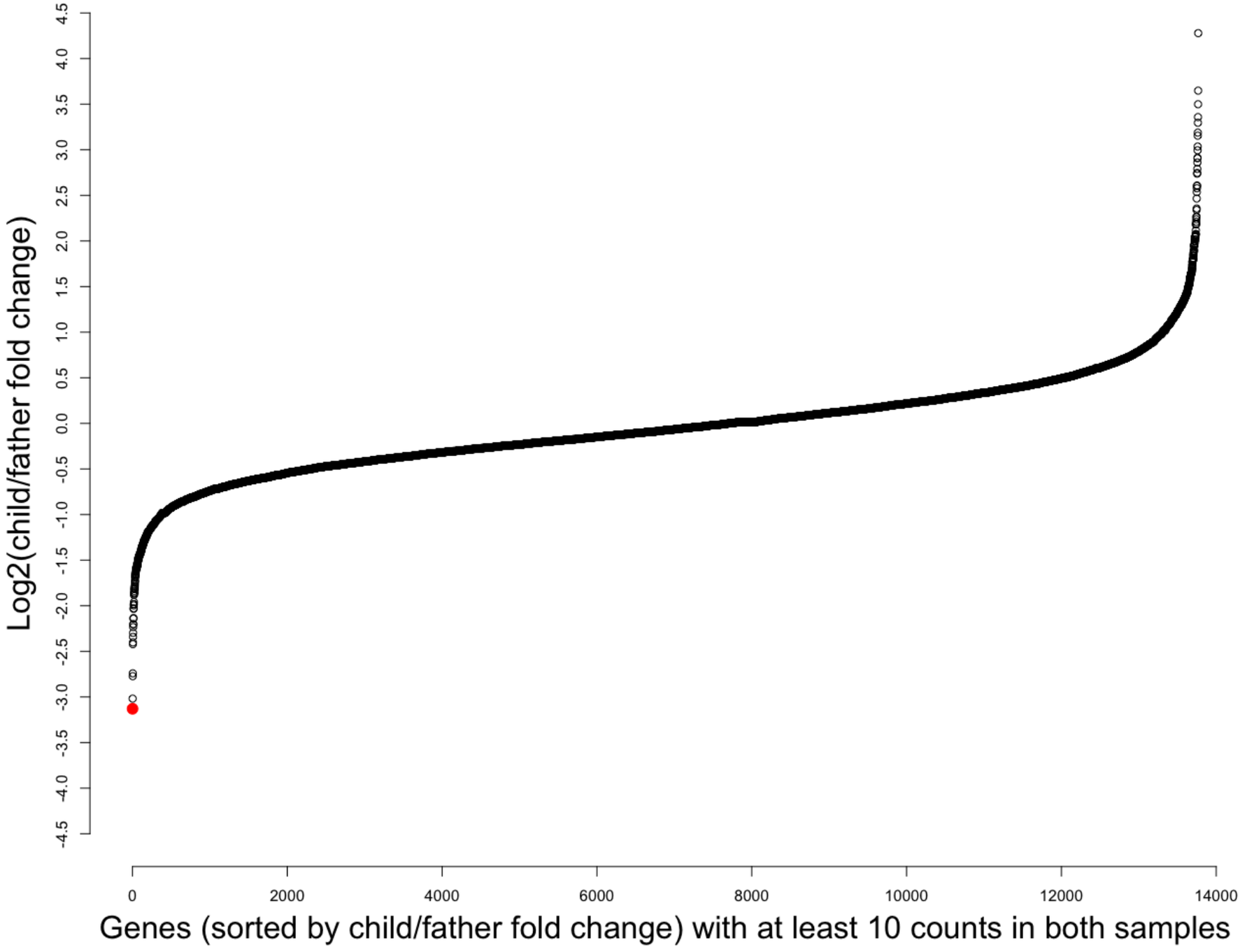
Proband srGS

Mother srGS

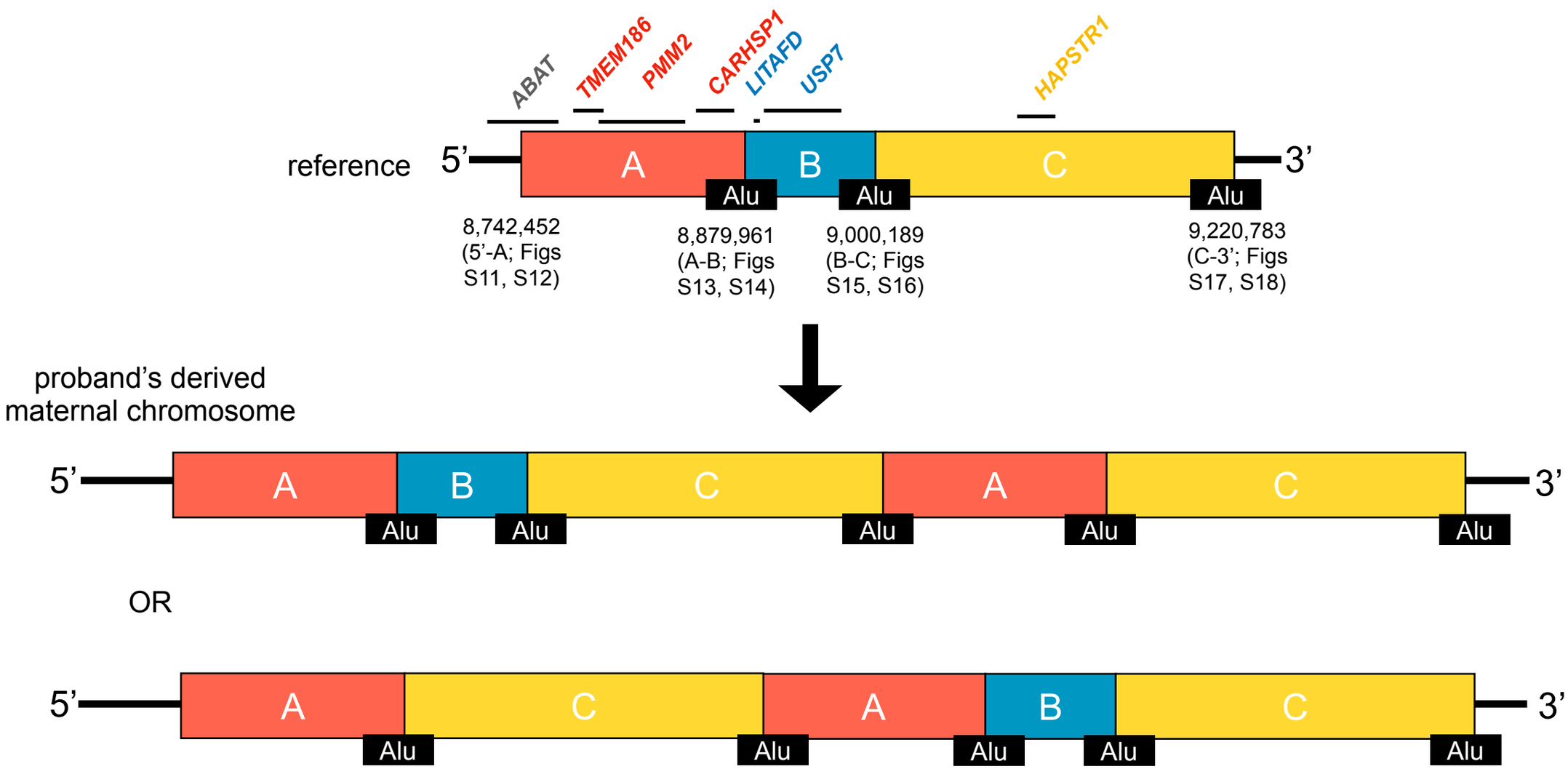
Father srGS



Supplemental Figure S7. Plot of CNV call in Proband 2 srGS. Panels show read depth (gray bars) and allele ratios (red dots) in Proband, Mother and Father's srGS data.



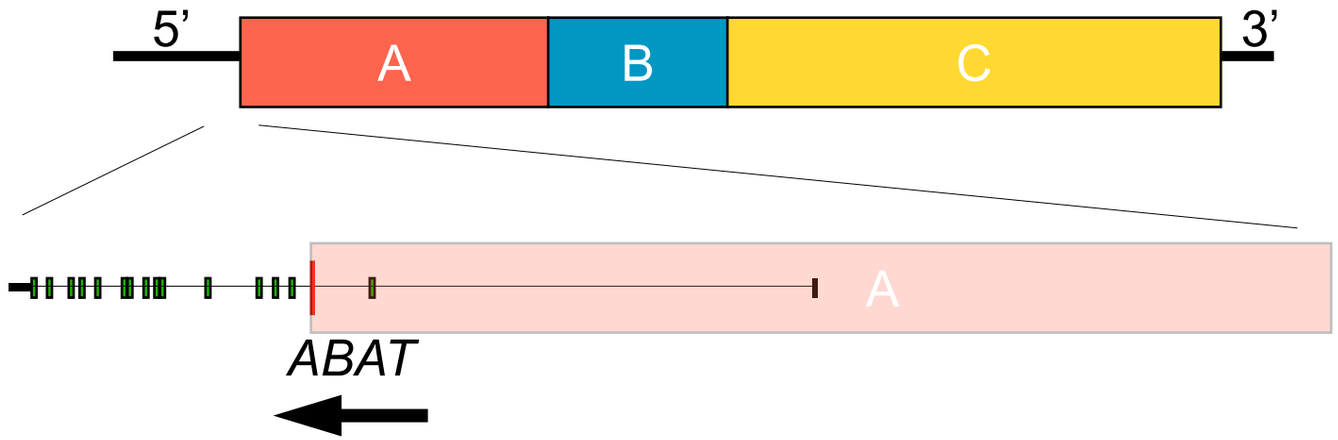
Supplemental Figure S8. Sorted, relative expression of transcripts in child and father using 3'-end RNA-seq. *HCFC1* (red dot) shows the greatest expression decrease in the proband relative to his father, an ~8.8-fold reduction. Ten counts in each sample were required.



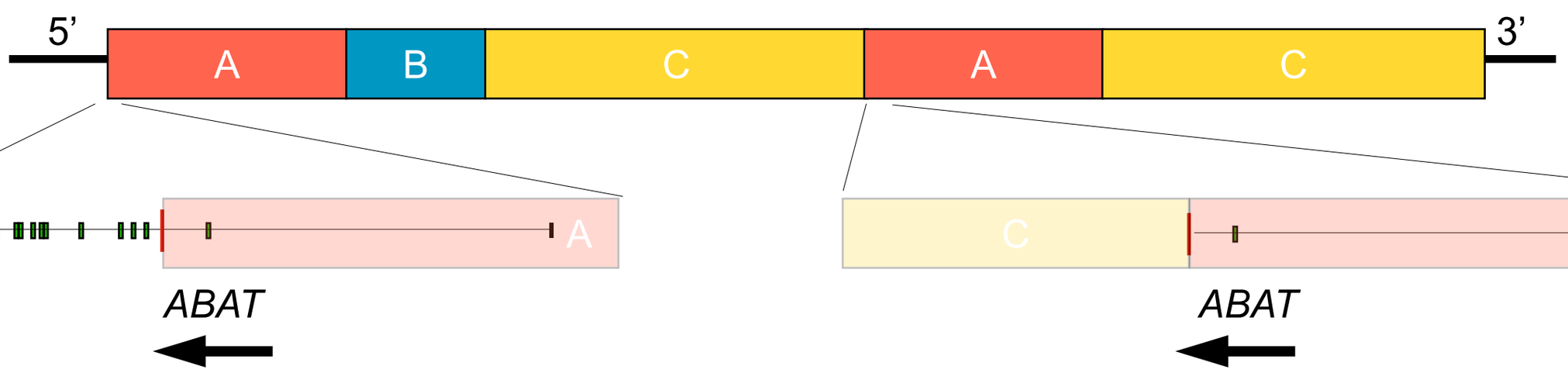
Gene	WT Copy Number	Proband Copy Number
ABAT	2	2 + 1 partial
TMEM186	2	3
PMM2	2	3
CARHSP1	2	3
LITAFD	2	2
USP7	2	2
HAPSTR1	2	3

Supplemental Figure S9. *De novo* Chr16 SV schematic in Proband 4. Breakpoints are named and defined with respect to the reference genome (5’-A, A-B, etc.). Regions of the reference genome shown in IGV images in Supplemental Figures S11-S18 are labeled next to names of the breakpoints. These structural variants result in overall copy numbers of three for *TMEM186, PMM2, CARHSP1* and *HAPSTR*. Total copy numbers for *LITAFD* and *USP7* remain at two. Note that the overall structure of the proband’s derived maternal chromosome could not be determined; two structural options are shown that are both supported by ccs reads. For simplicity, the top version is used in Supplemental Figures S10-S18.

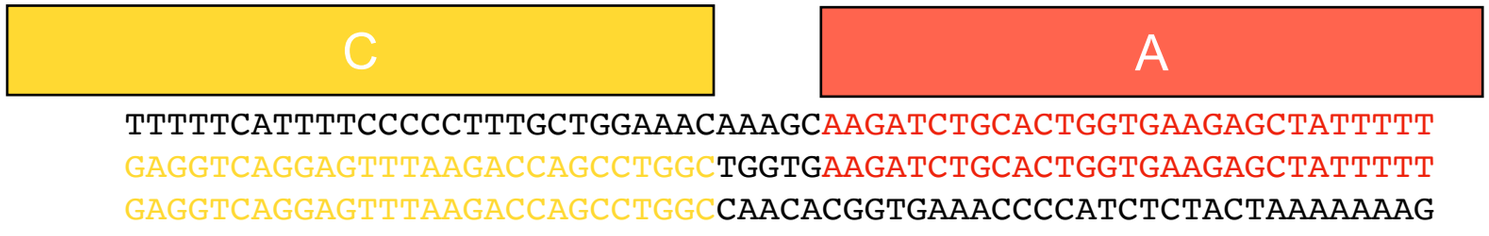
A reference



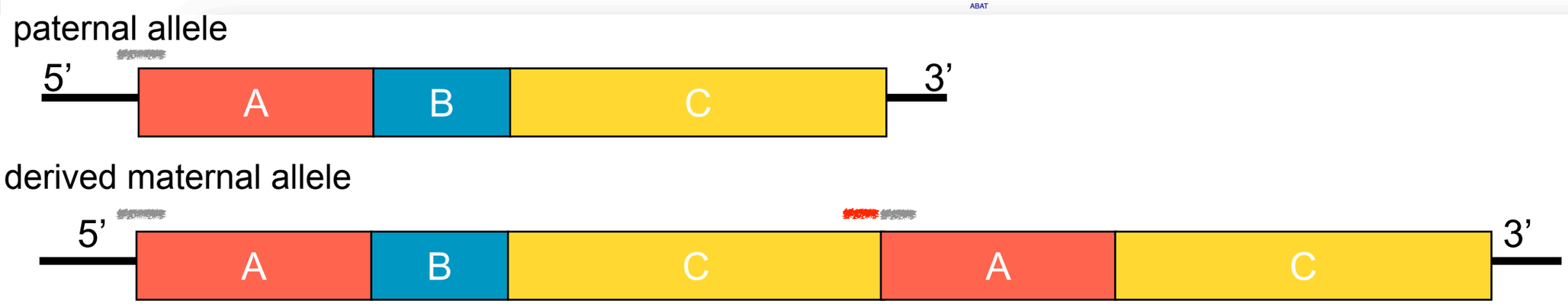
B proband’s derived maternal chromosome



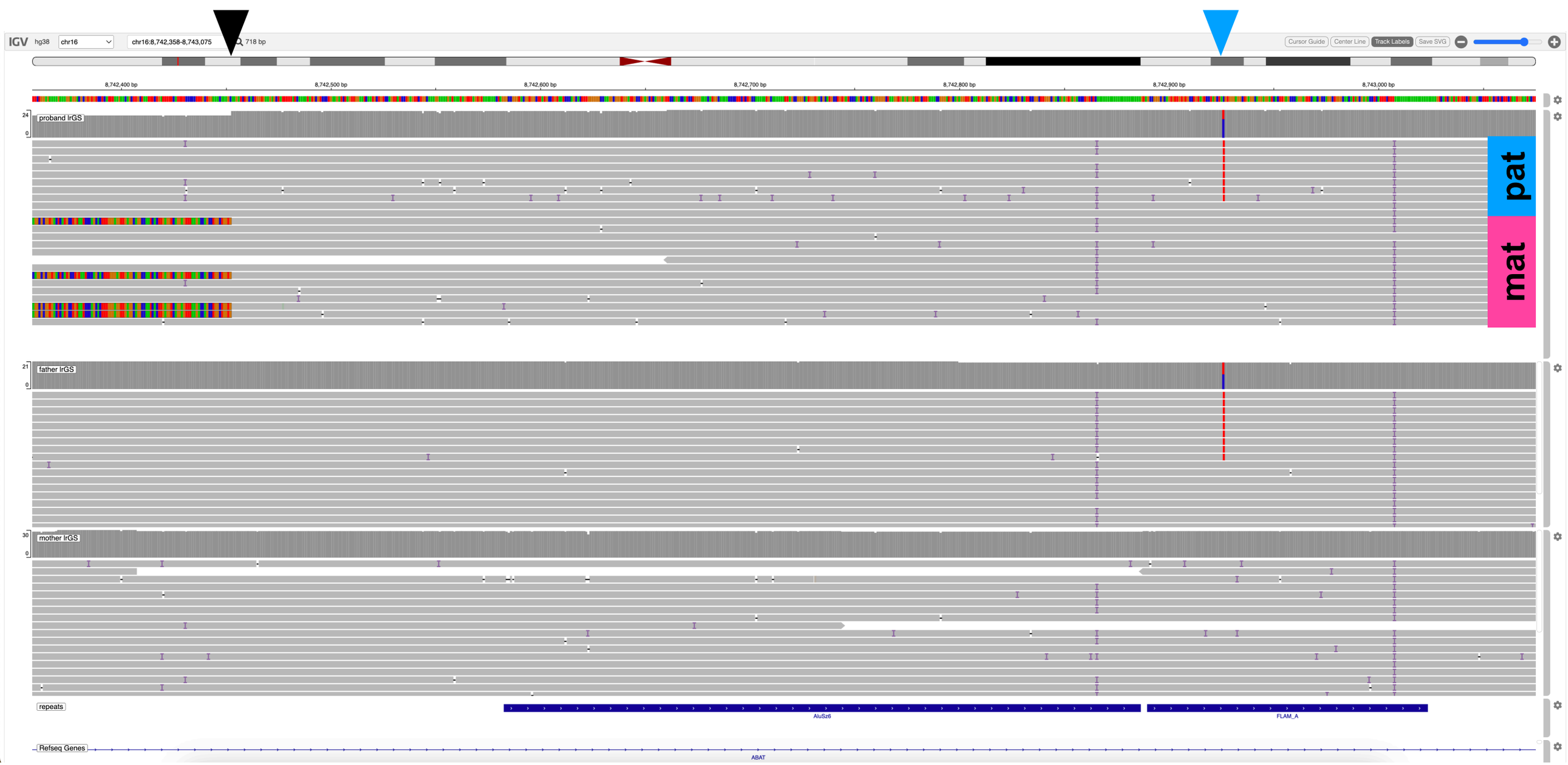
C



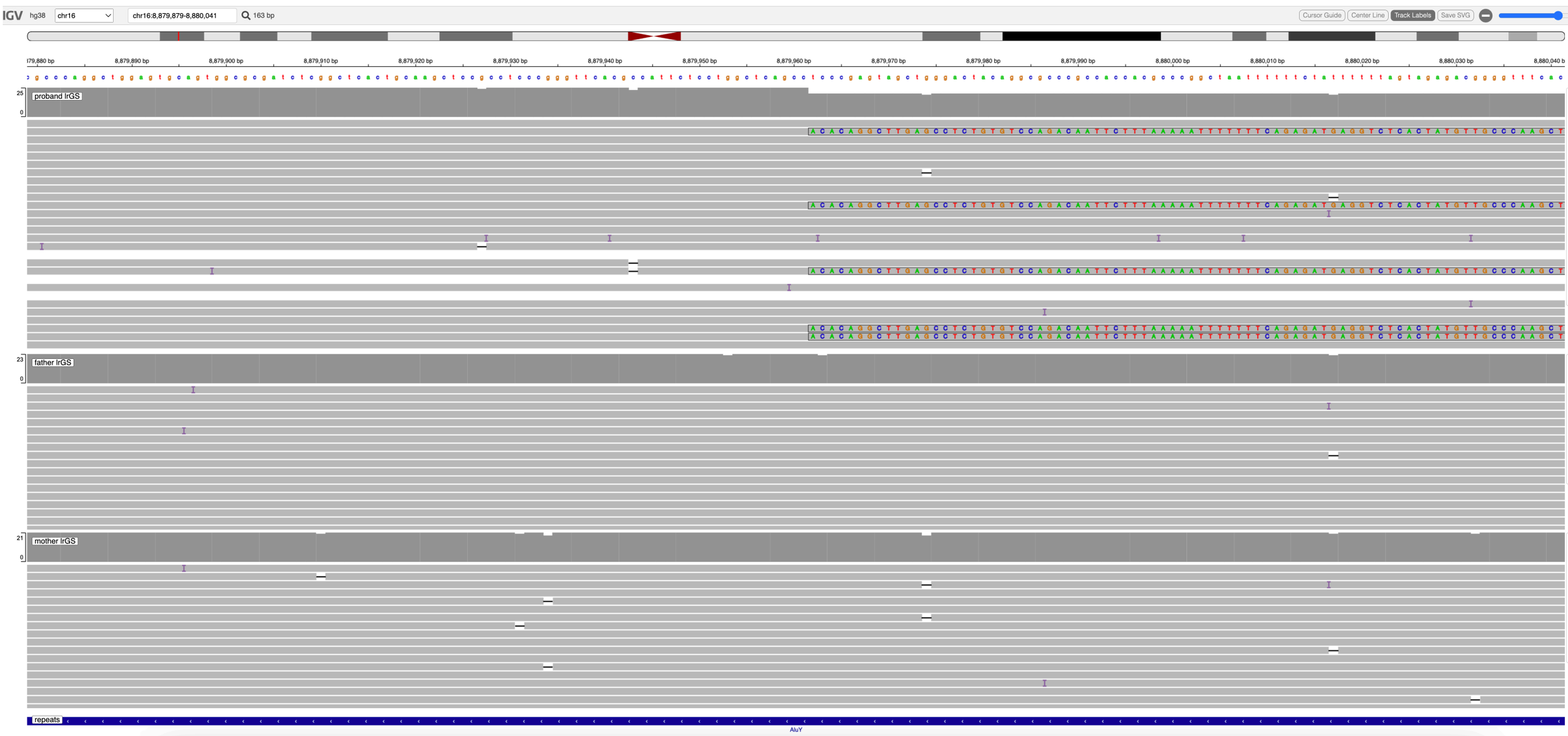
Supplemental Figure S10. *De novo* Chr16 SV schematic in Proband 4. A. The breakpoint at the 5’ end of A lies within an intron of *ABAT*. B. In the proband’s derived Chr16, there is a full copy of *ABAT* (left) and a partial copy of the two 5’ exons of *ABAT* (right). C. Nucleotide alignment of the breakpoint and fusion of regions C and A.



Supplemental Figure S11. Visualization of reads aligning to the 5’-A breakpoint region of the reference genome (Chr16:8,742,338-8,742,585). A subset of lrGS reads for proband and parents are shown in IGV at nt resolution. Note this breakpoint does not overlap any repeat sequences. Reads aligning here support both reference sequence (gray bars) and the proband’s novel C-A fusion (red/gray bar).



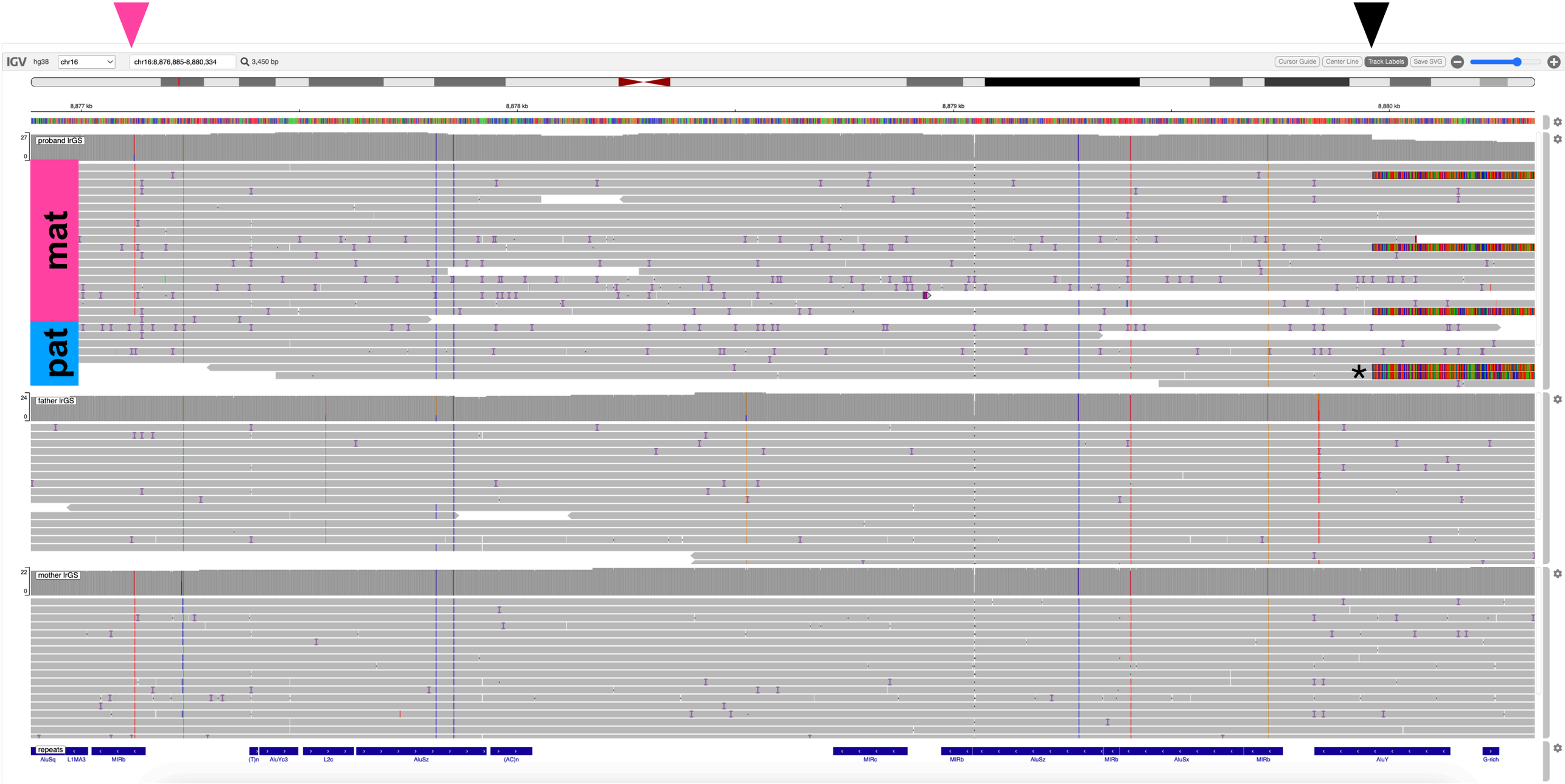
Supplemental Figure S12. Visualization of reads aligning to the 5'-A breakpoint of the reference genome (Chr16:8,742,338-8,742,585). A subset of lrGS reads for proband and parents are shown in IGV. The SV breakpoint (black arrow head) lies in *trans* to a paternally-inherited variant (blue arrowhead). Paternal (blue) and maternal (pink) reads are shown with colored boxes.



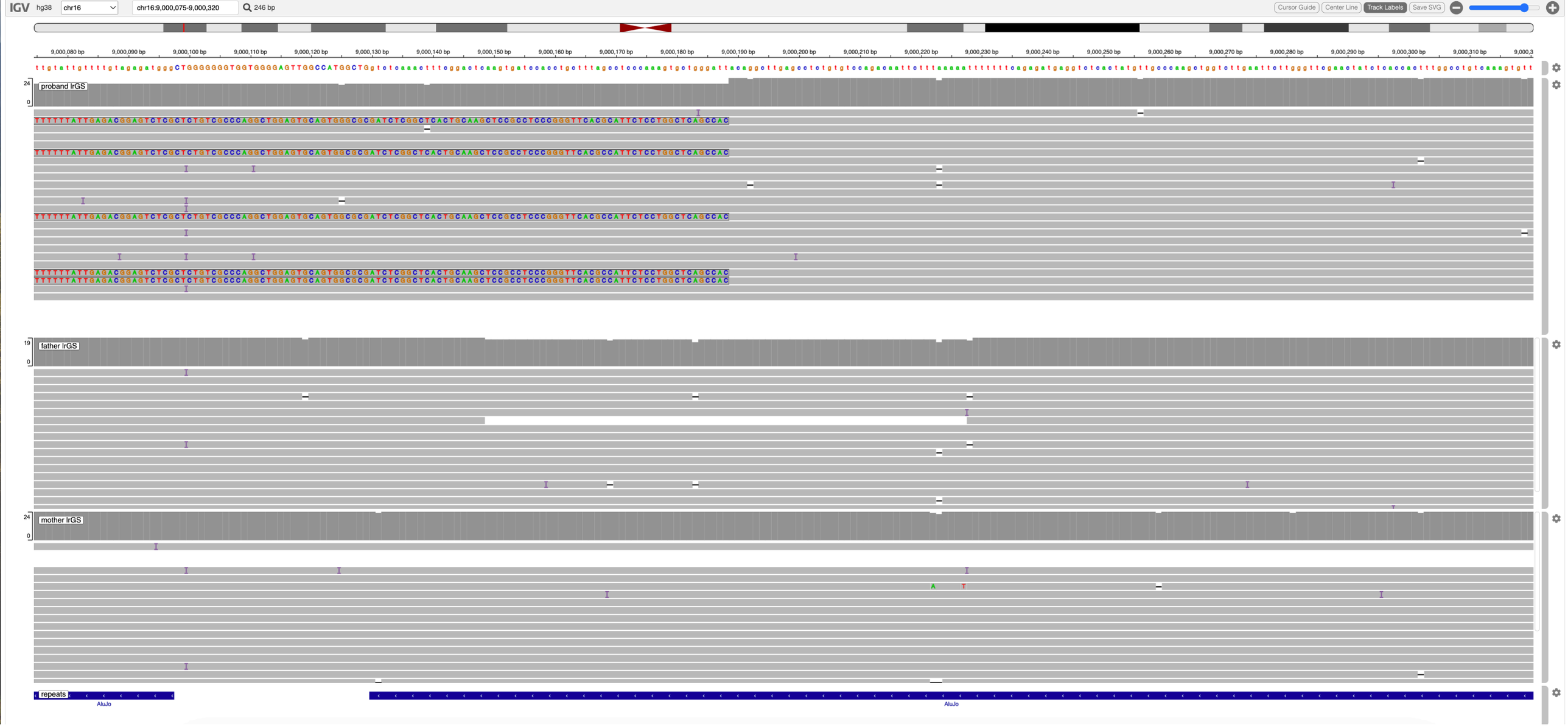
paternal allele

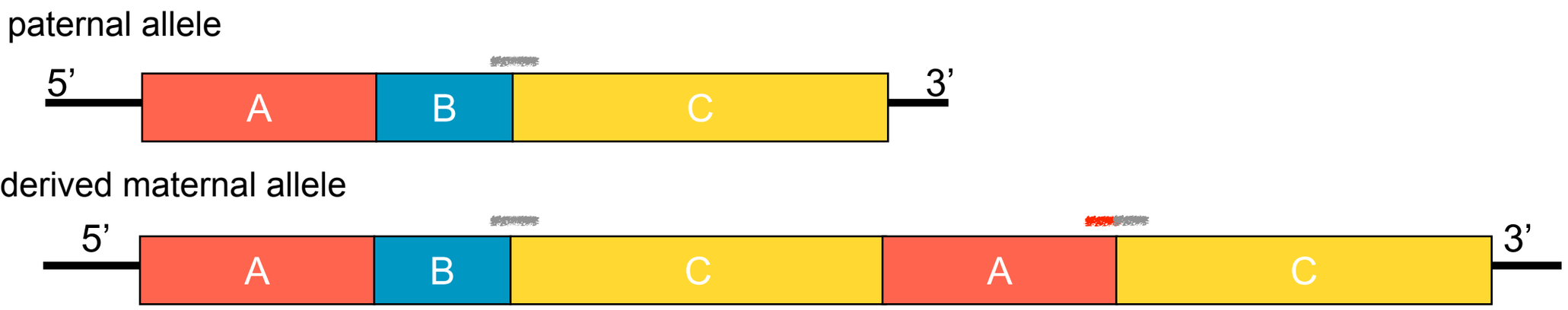
derived maternal allele

Supplemental Figure S13. Visualization of reads aligning to the A-B breakpoint region of the reference genome (Chr16:8,879,879-8,880,041). A subset of lrGS reads for proband and parents are shown in IGV at nt resolution. Note this breakpoint overlaps an Alu repeat. Reads aligning here support both reference sequence (gray bars) and the proband’s novel A-C fusion (red/gray bar).

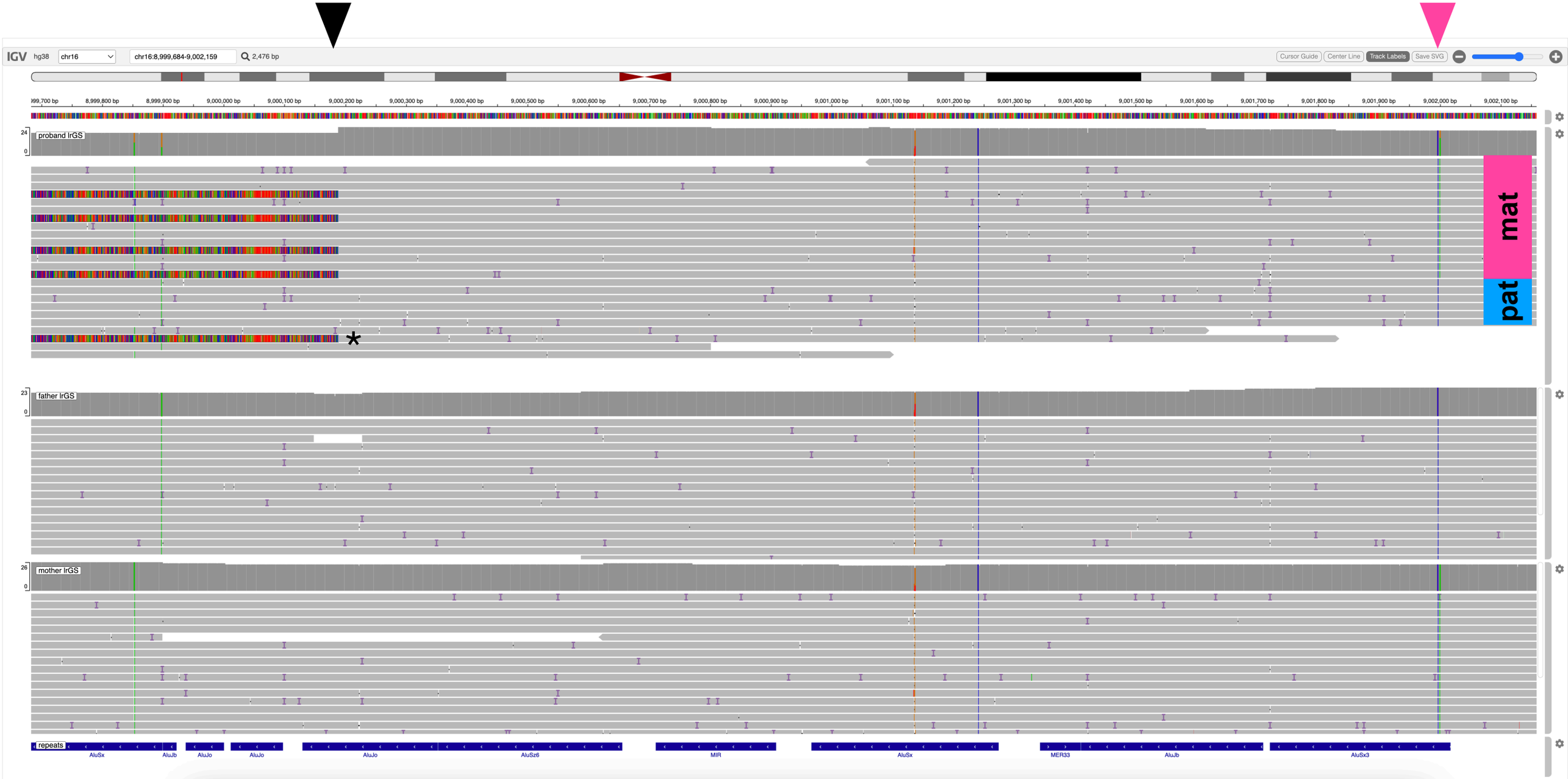


Supplemental Figure S14. Visualization of reads aligning to the A-B breakpoint region of the reference genome (Chr16:8,879,879-8,880,041). A subset of lrGS reads for proband and parents are shown in IGV. The SV breakpoint (black arrow head) lies in *cis* with a maternally-inherited variant (pink arrowhead). Note that two reads (marked with an asterisk, *) are not long enough to include both the breakpoint and the maternally-inherited snp. Paternal (blue) and maternal (pink) reads are shown with colored boxes.

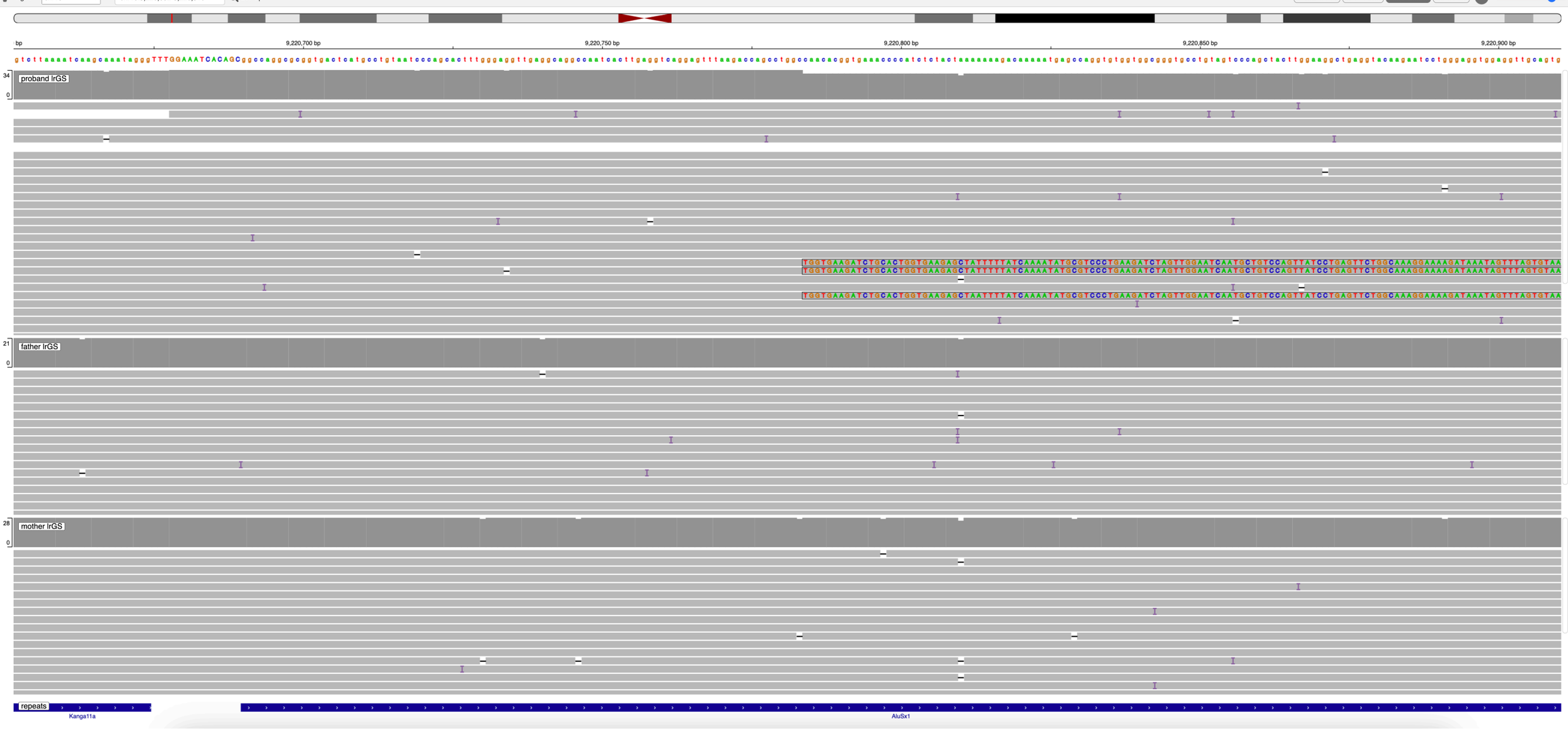


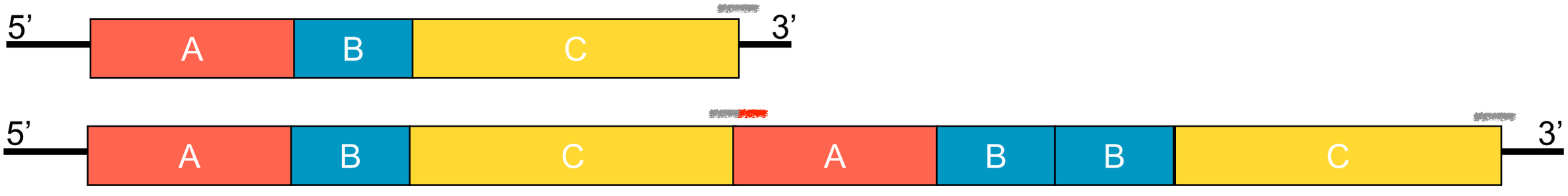


Supplemental Figure S15. Visualization of reads aligning to the B-C breakpoint region of the reference genome (Chr16:9,000,075-9,000,320). A subset of lrGS reads for proband and parents are shown in IGV at nt resolution. Note this breakpoint overlaps and Alu repeat. Reads aligning here support both reference sequence (gray bars) and the proband's novel A-C fusion (red/gray bar).

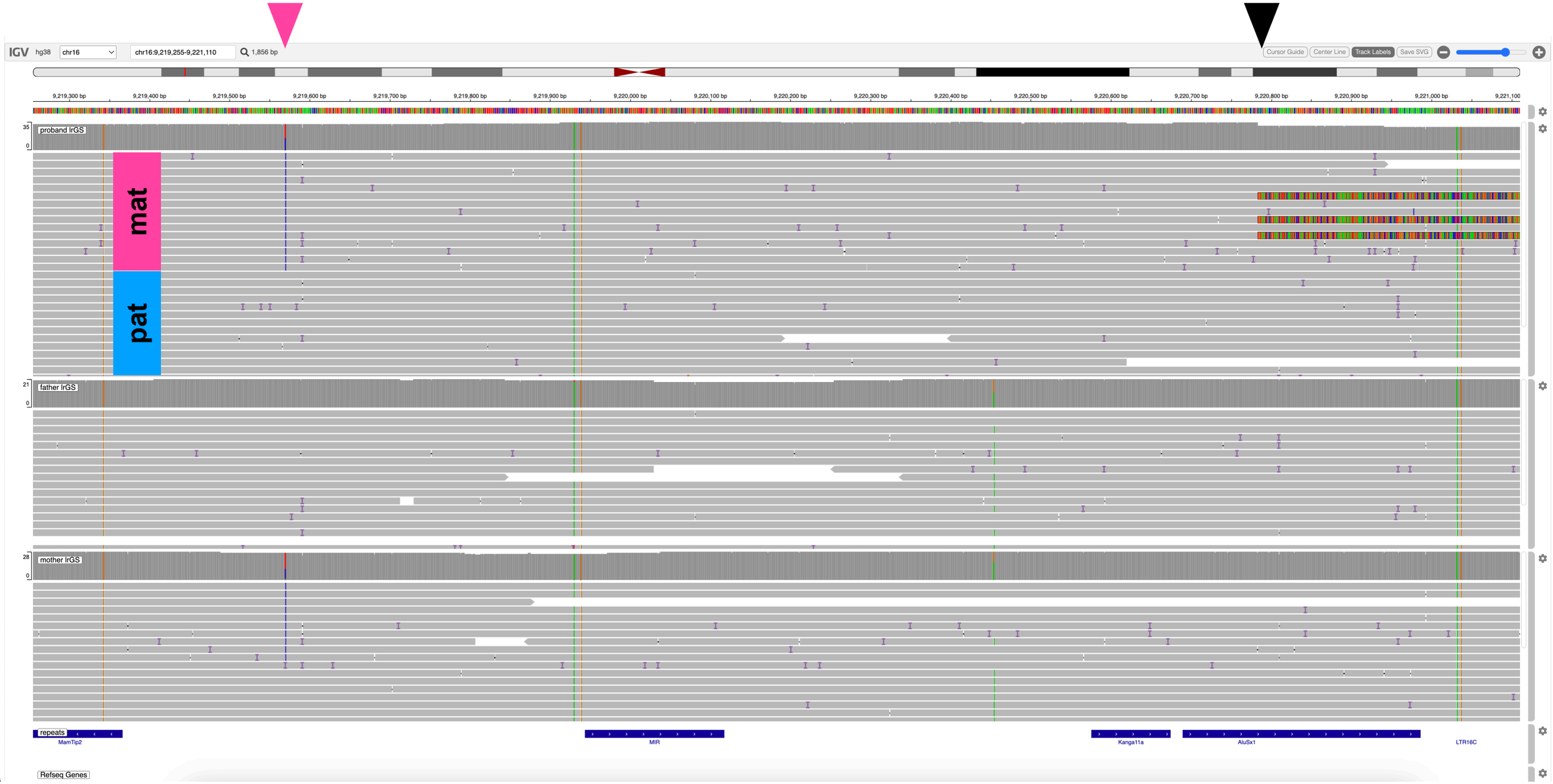


Supplemental Figure S16. Visualization of reads aligning to the B-C breakpoint region of the reference genome (Chr16:8,999,684-9,002,159). A subset of lrGS reads for proband and parents are shown in IGV. The SV (black arrow head) lies in *cis* with a maternally-inherited variant (pink arrowhead; green-colored base). Note that one read (marked with an asterisk, *) is not long enough to include both the breakpoint and the maternally-inherited snp. Paternal (blue) and maternal (pink) reads are shown with colored boxes.



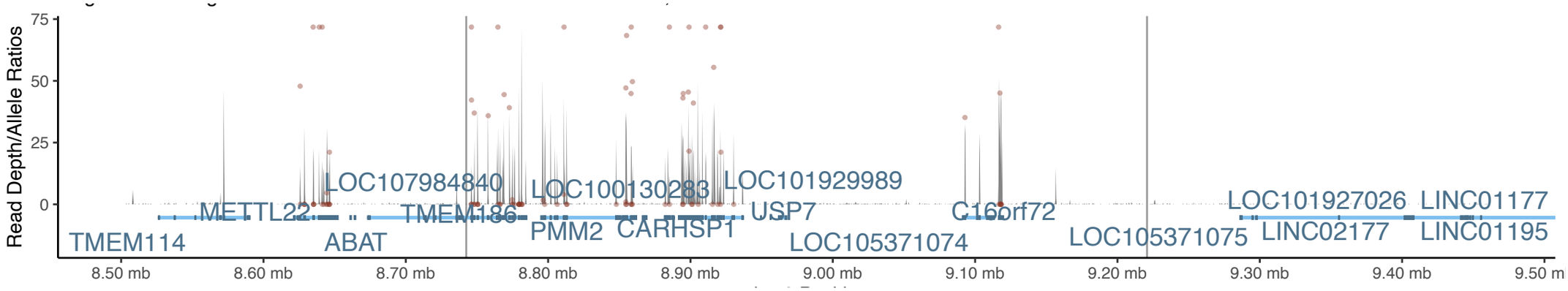


Supplemental Figure S17. Visualization of reads aligning to the C-3’ breakpoint region of the reference genome (Chr16:9,220,652-9,220,910). A subset of lrGS reads for proband and parents are shown in IGV at nt resolution. Note this breakpoint overlaps and Alu repeat. Reads aligning here support both reference sequence (gray bars) and proband’s novel C-A fusion (red/gray bar).

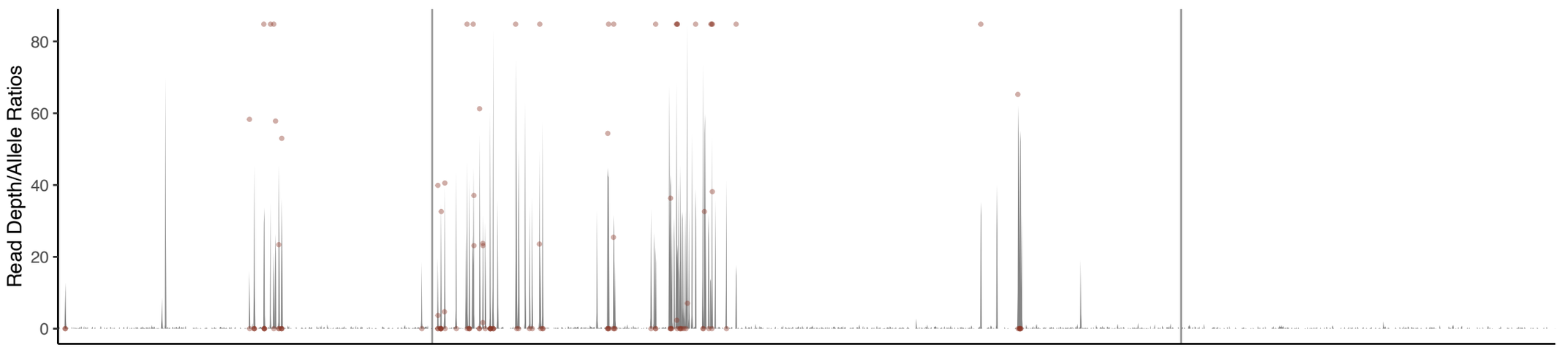


Supplemental Figure S18. Visualization of reads aligning to the C-3' breakpoint region of the reference genome (Chr16:9,219,255-9,221,110). A subset of lrGS reads for proband and parents are shown in IGV. The duplication (black arrow head) lies in cis with a maternally-inherited variant (pink arrowhead). Paternal (blue) and maternal (pink) reads are shown with colored boxes.

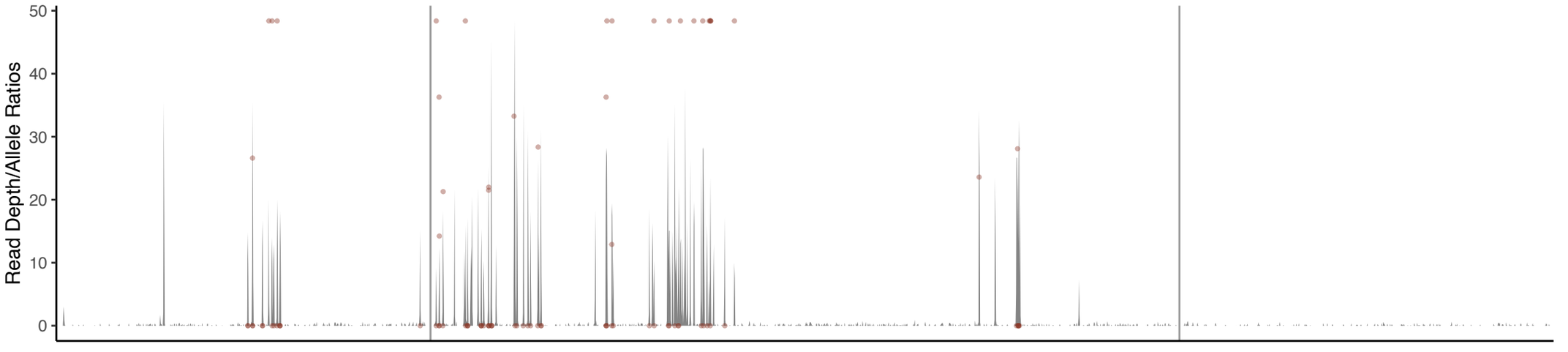
Proband ES



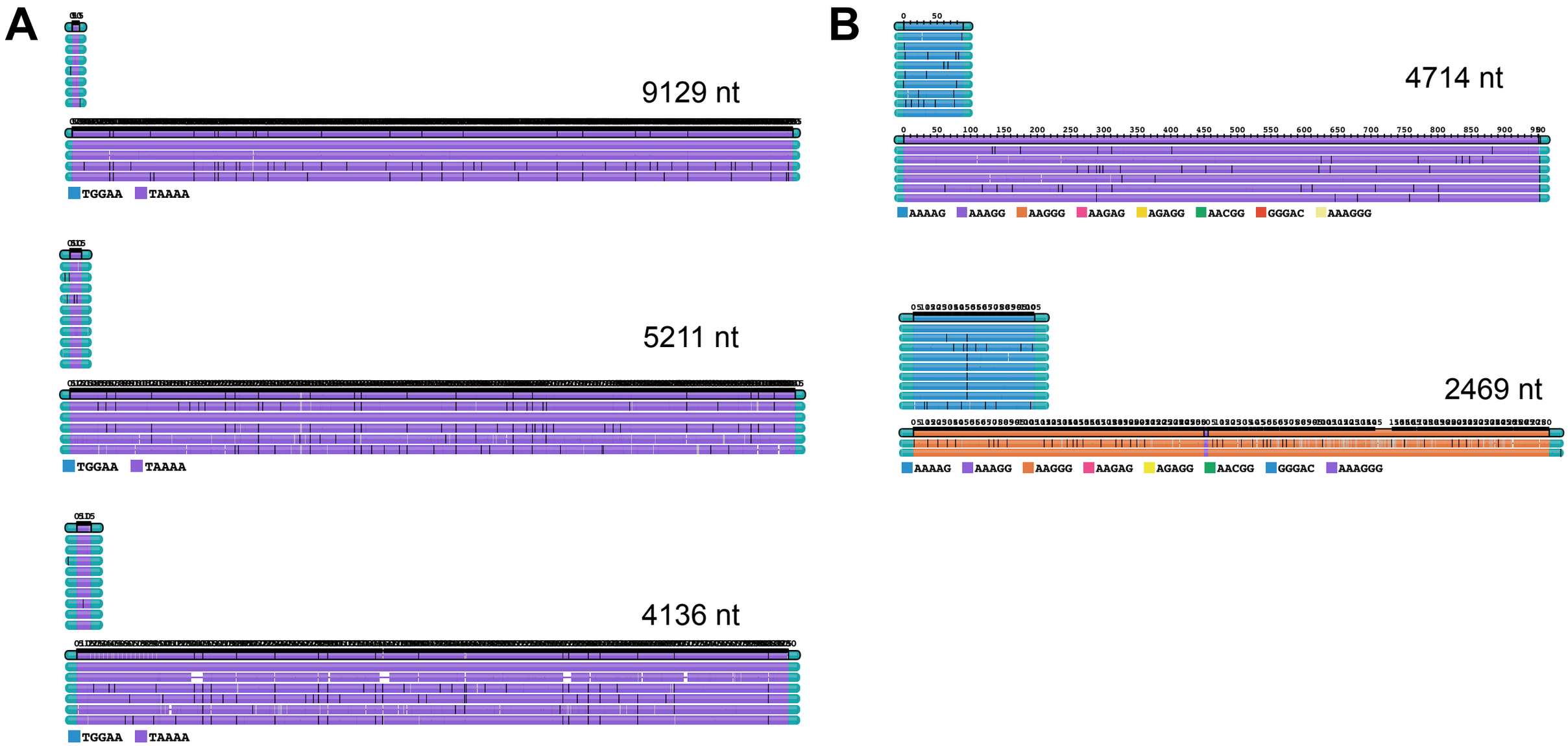
Father ES



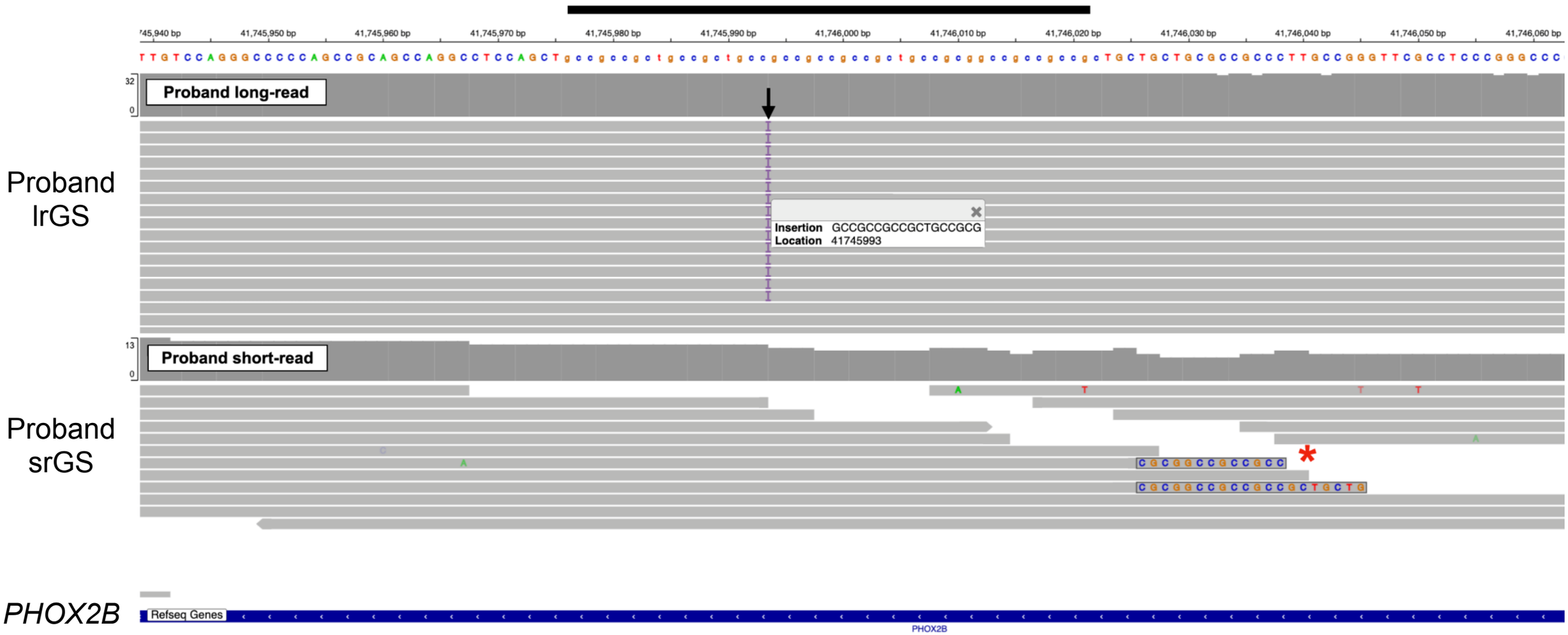
Mother ES



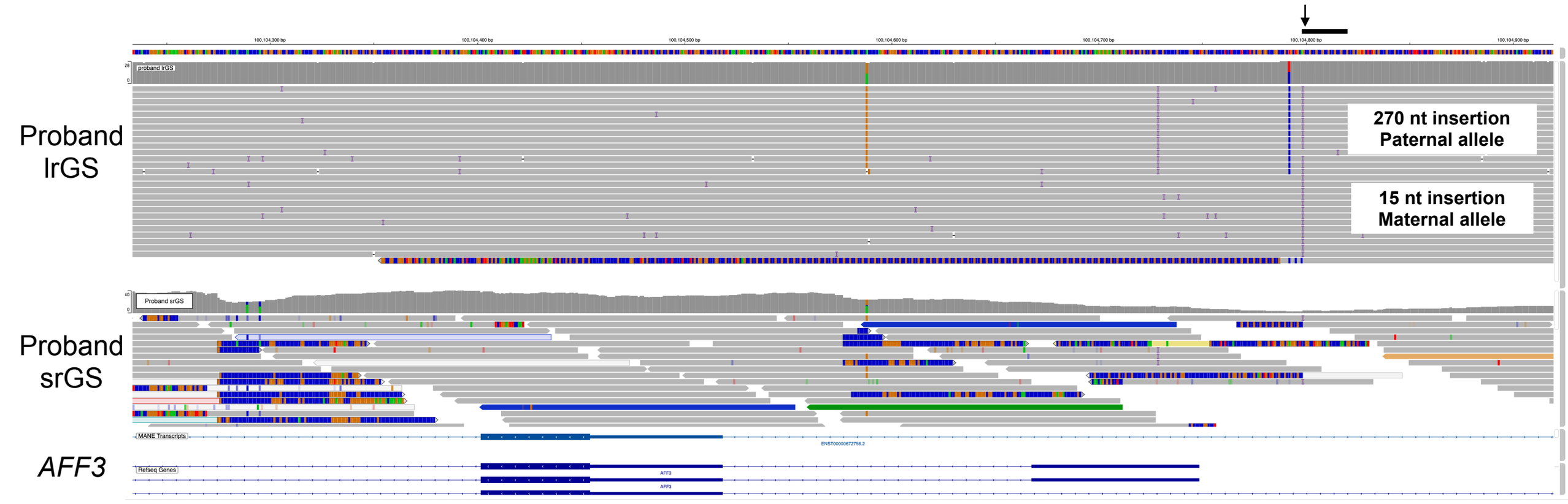
Supplemental Figure S19. Plot of the region of interest in Proband 4 and both parents. Panels show read depth of variant calls (gray lines) and allele ratios (red dots) from exome data in Proband, Father and Mother. Genes are also shown in blue in the top panel.



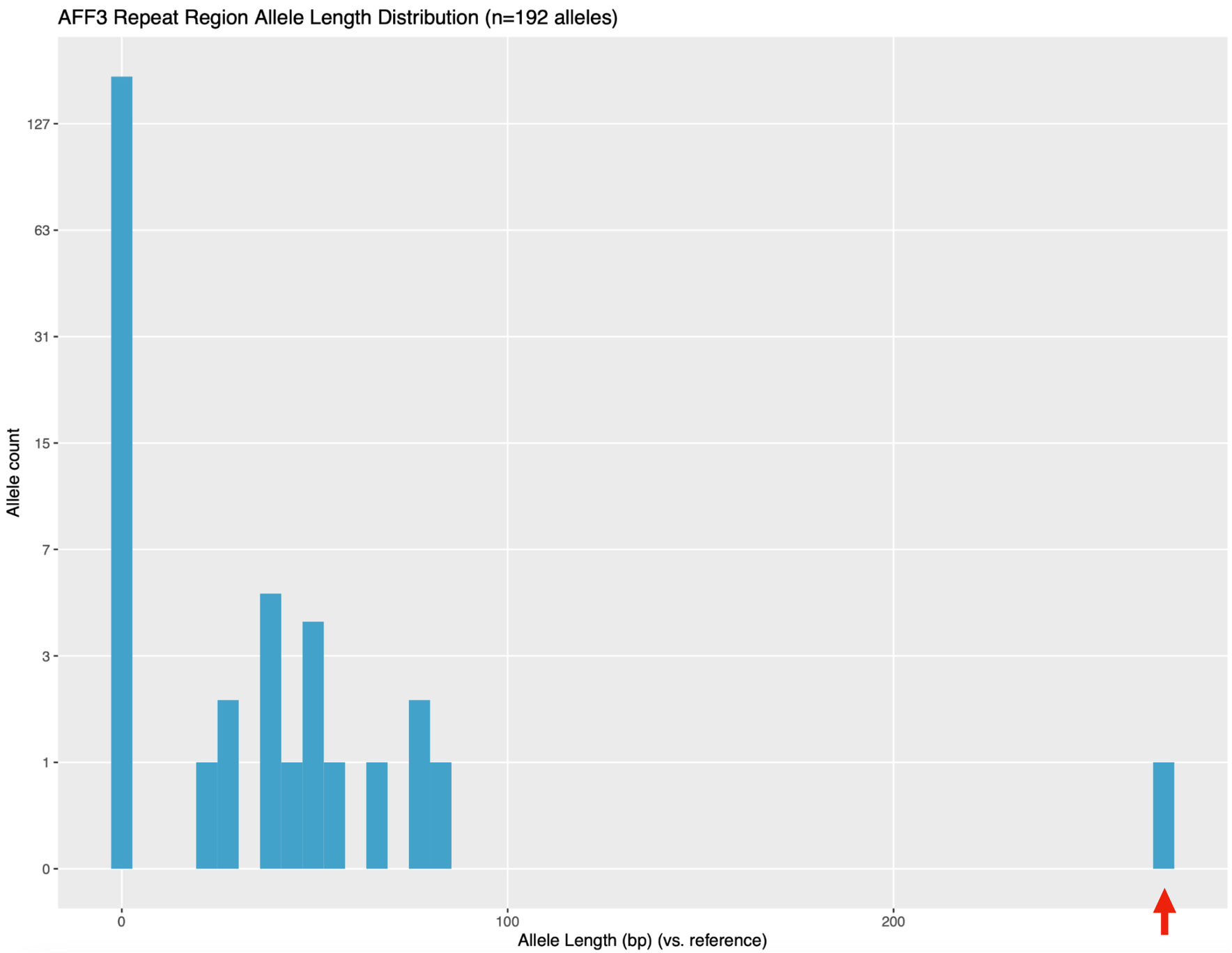
Supplemental Figure S20. Visualization of repeat expansions using TRGT and TRVZ (see Methods). Repeats are separated by haplotype. **A.** Three large heterozygous insertions in *BEAN1* in three individuals were determined to be benign based on sequence content (TAAAA). **B.** Heterozygous expansions in two individuals in *RFC1* were determined to be benign (AAAGG) and potentially clinically significant (AAGGG). Note that only biallelic expansions in *RFC1* have been reported to be associated with disease, and further, this AAGGG repeat is interrupted by a small stretch of AAAGG.



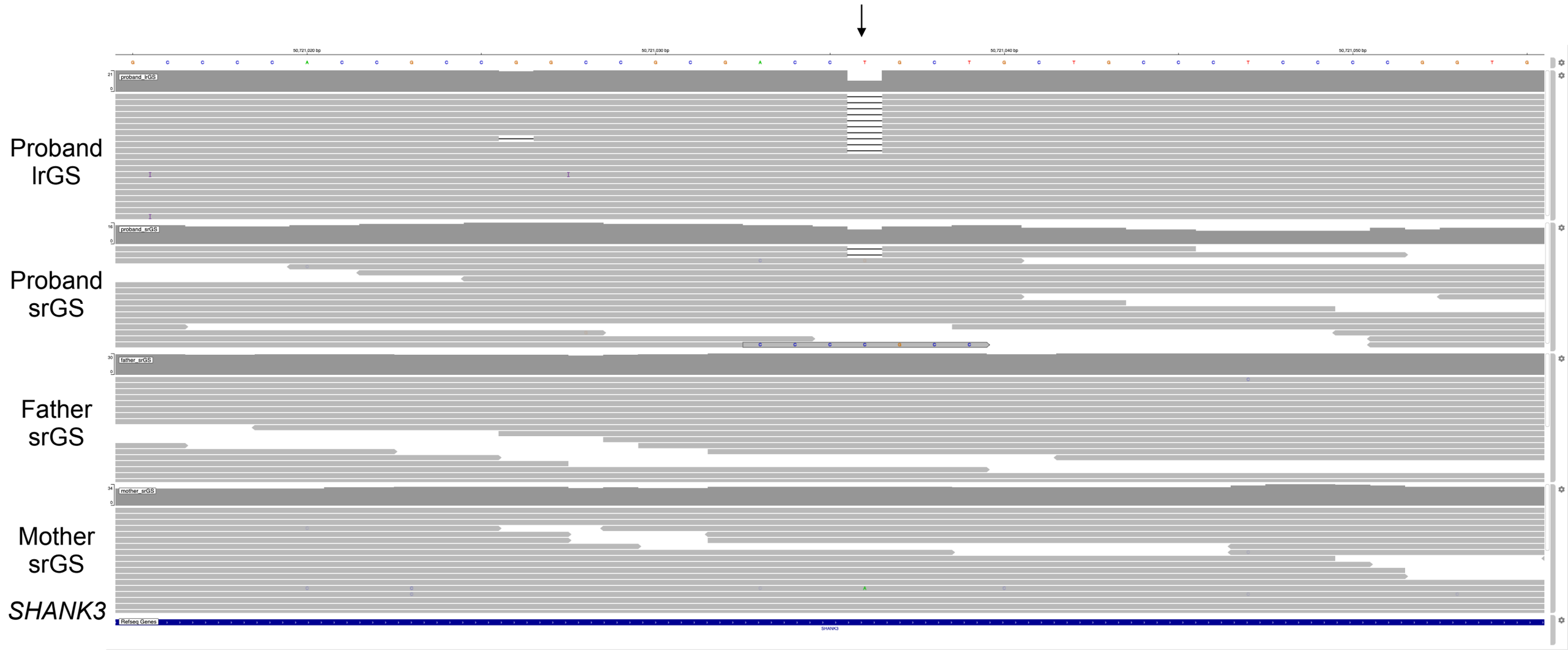
Supplemental Figure S21. A pathogenic 18 bp alanine tract expansion in *PHOX2B* (black arrow) was observed in a repeat region (black bar) in Proband 5. Some srGS reads have evidence of this insertion (red asterisk) but the variant was not called.



Supplemental Figure S22. A 270 bp insertion in the promoter of *AFF3* (black arrow) was observed in a repeat region (black bar) in Proband 6. Segregation of a downstream SNP determined that the 270 bp insertion is on the paternal allele.

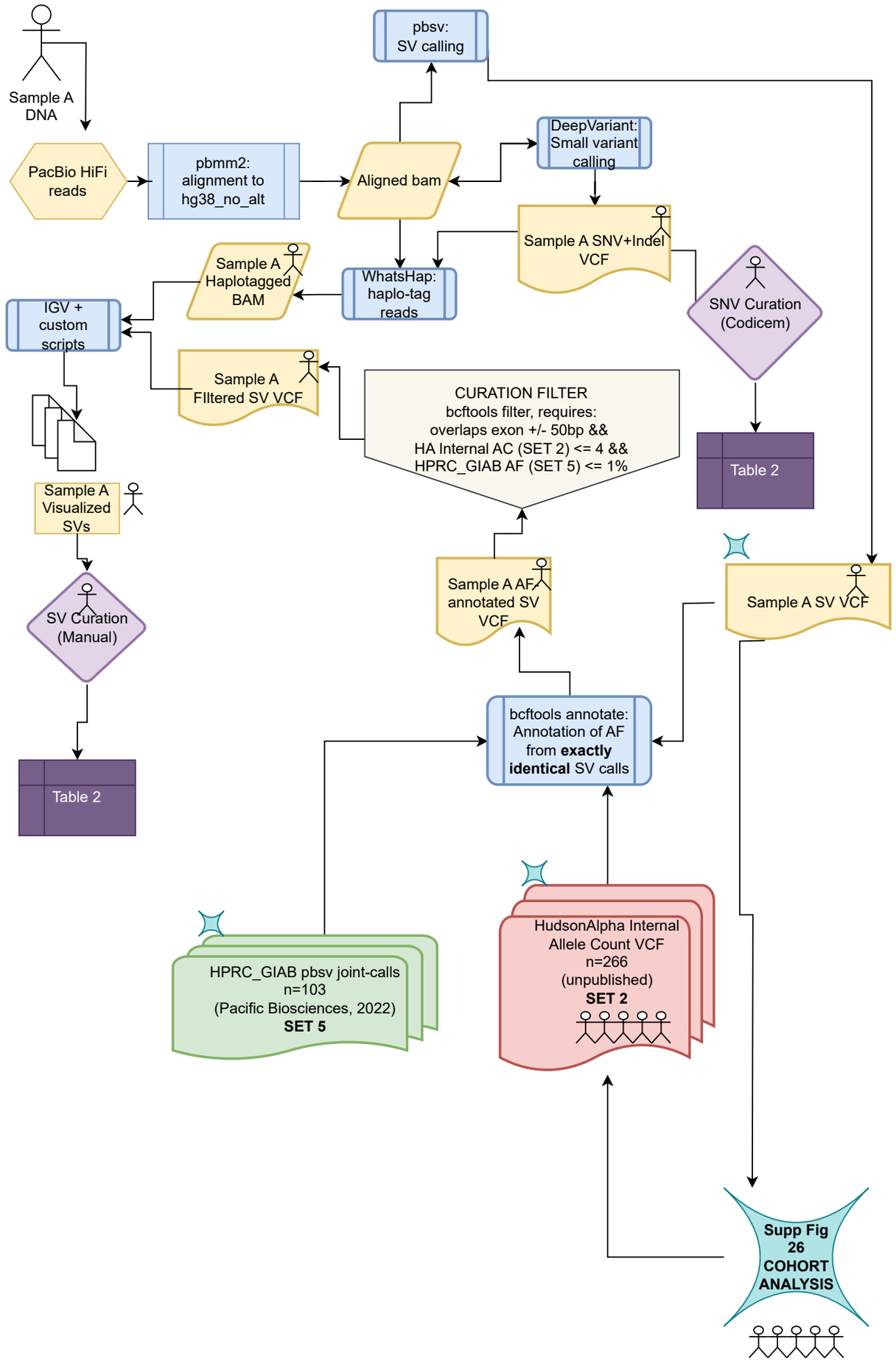


Supplemental Figure S23. Distribution of lengths of variant calls in the promoter of *AFF3* (Chr2:100104769-100104854) in 96 probands (192 alleles). The 270 nt insertion in *AFF3* in Proband 6 (red arrow) is a clear outlier compared to other calls in long-read data in the region.

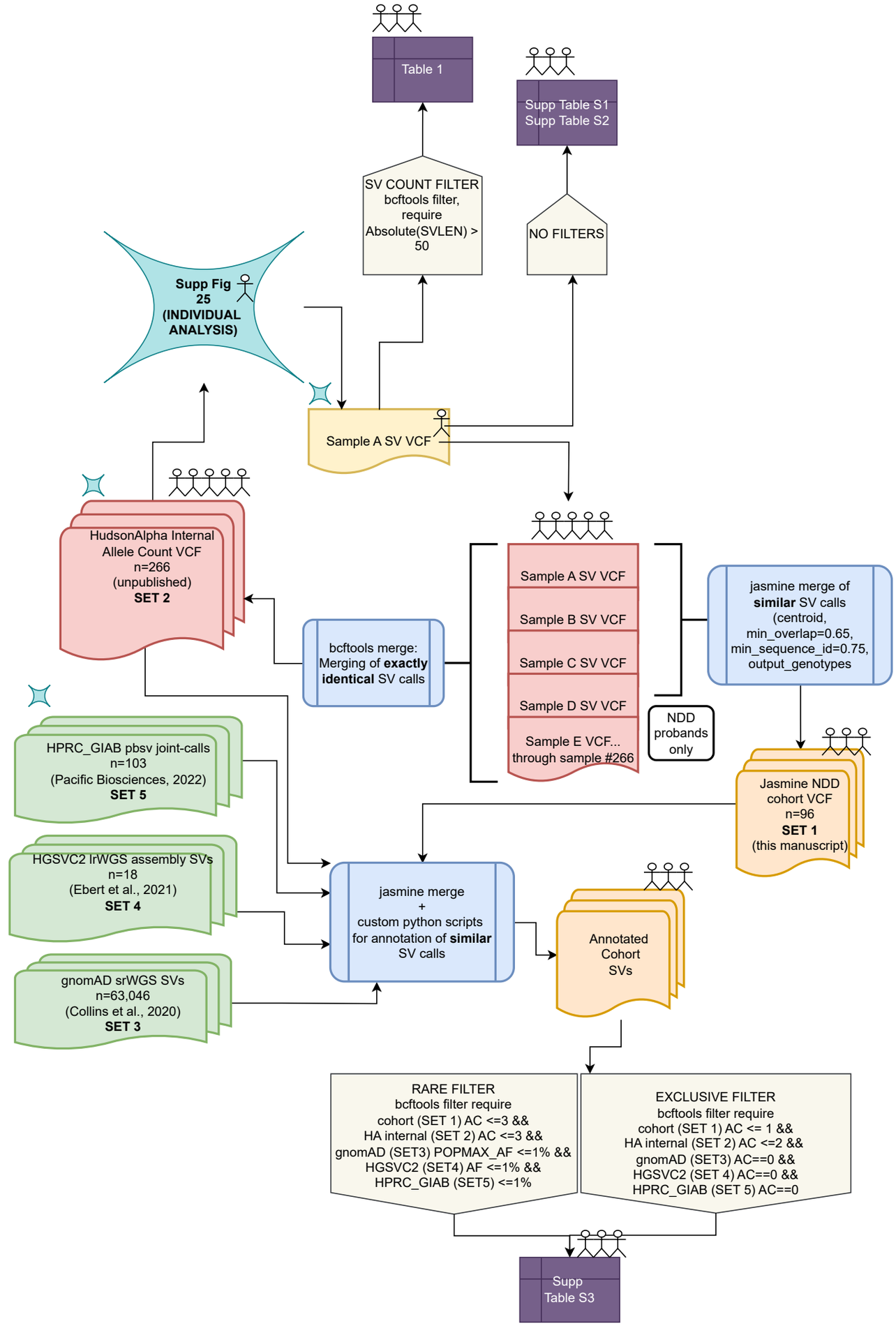


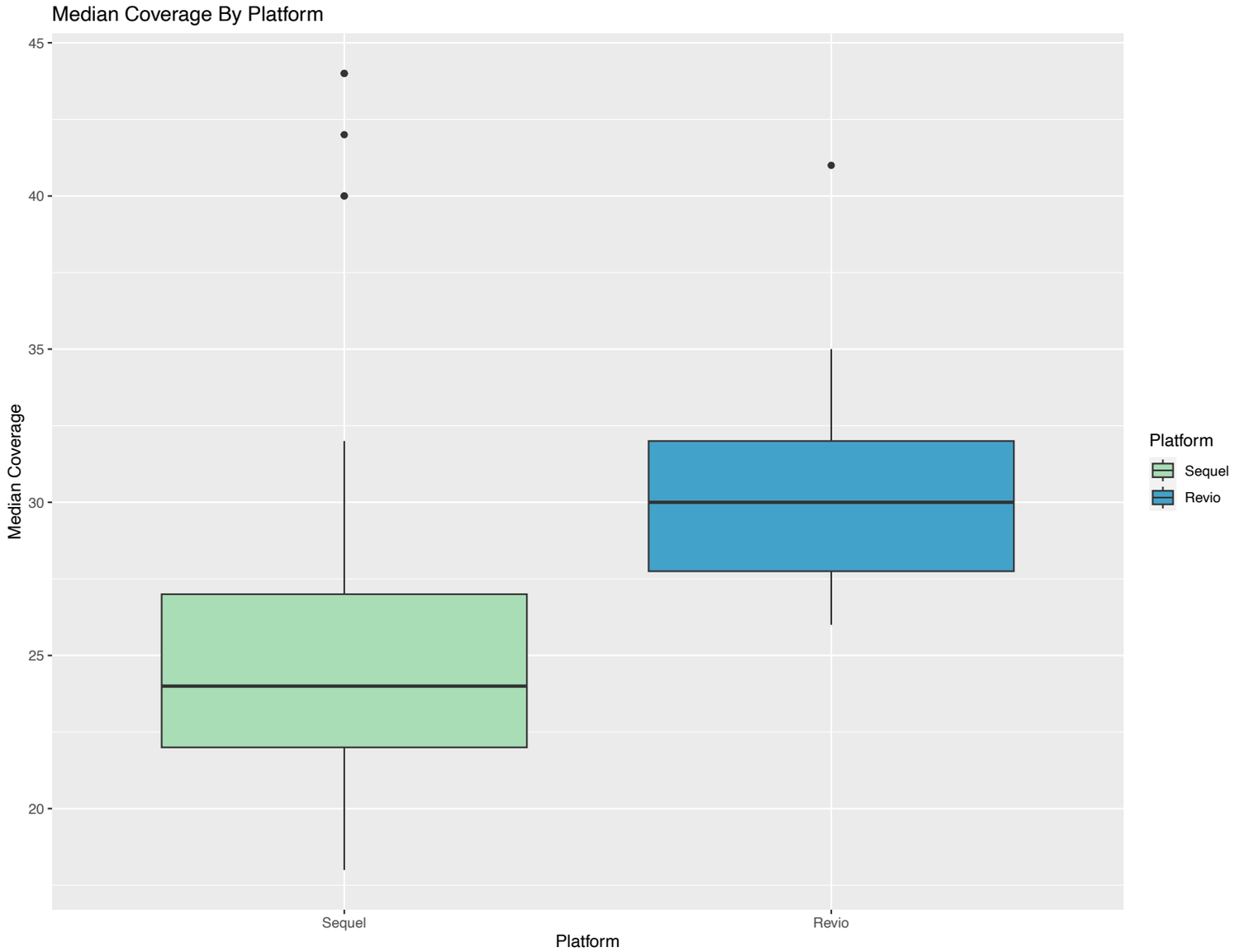
Supplemental Figure S24. A 1 bp deletion in *SHANK3* (black arrow) was observed in lrGS data and srGS data in the probing but was not called in srGS in Proband 7. The 1 bp deletion was not observed in parental reads.

Supplementary Figure S25. Overview of the structural variant calling, annotation, filtering, and curation pipeline used for individual case analysis. The teal star represents where the process links to and from Supplementary Figure S26 (SV cohort analysis). Datasets that appear on both figures are marked with a small teal star. Single sample datasets are represented in gold and marked with one person icon; BAMs are represented by parallelograms, single-sample VCFs are represented by curved-bottom rectangles, multi-sample VCFs are represented by a stack of curve-bottom rectangles, HiFi reads are represented by a hexagon, and images are represented by file icons. Software packages are represented by blue barred rectangles. Filtering steps are represented by pentagons and shaded tan. Among multi-sample VCFs: red represents the HudsonAlpha internal allele count dataset which contains samples from other projects; orange represents the proband cohort dataset described in this manuscript; and green represents public data. Sample curation is represented as a purple diamond. Manuscript tables are represented as dark purple table icons. As an additional signifier, proband single-sample data and analyses are marked with a singular person icon; this manuscript's cohort data and analyses are marked with a three-person icon; and general HudsonAlpha internal data (including individuals beyond the scope of this manuscript) are represented with a five-person icon. Figure created with draw.io.

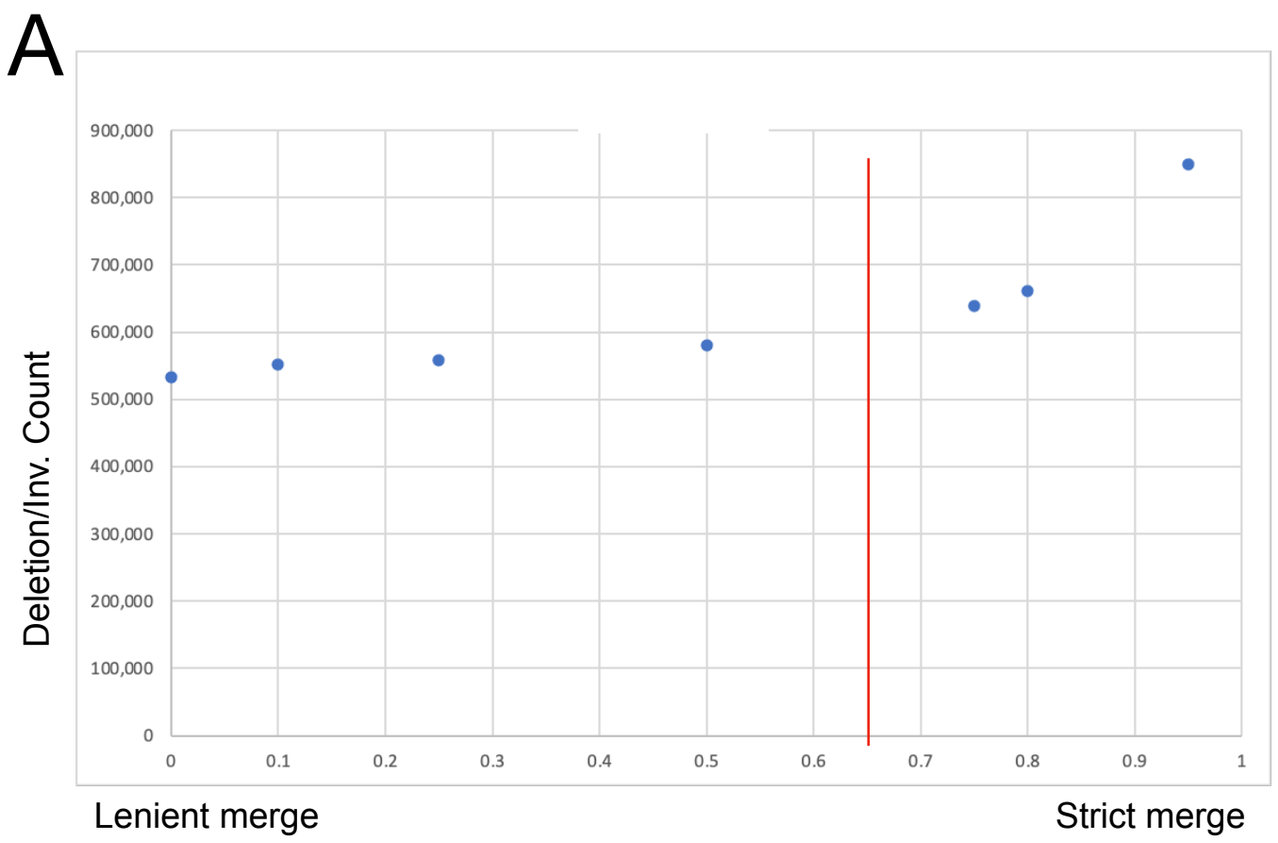


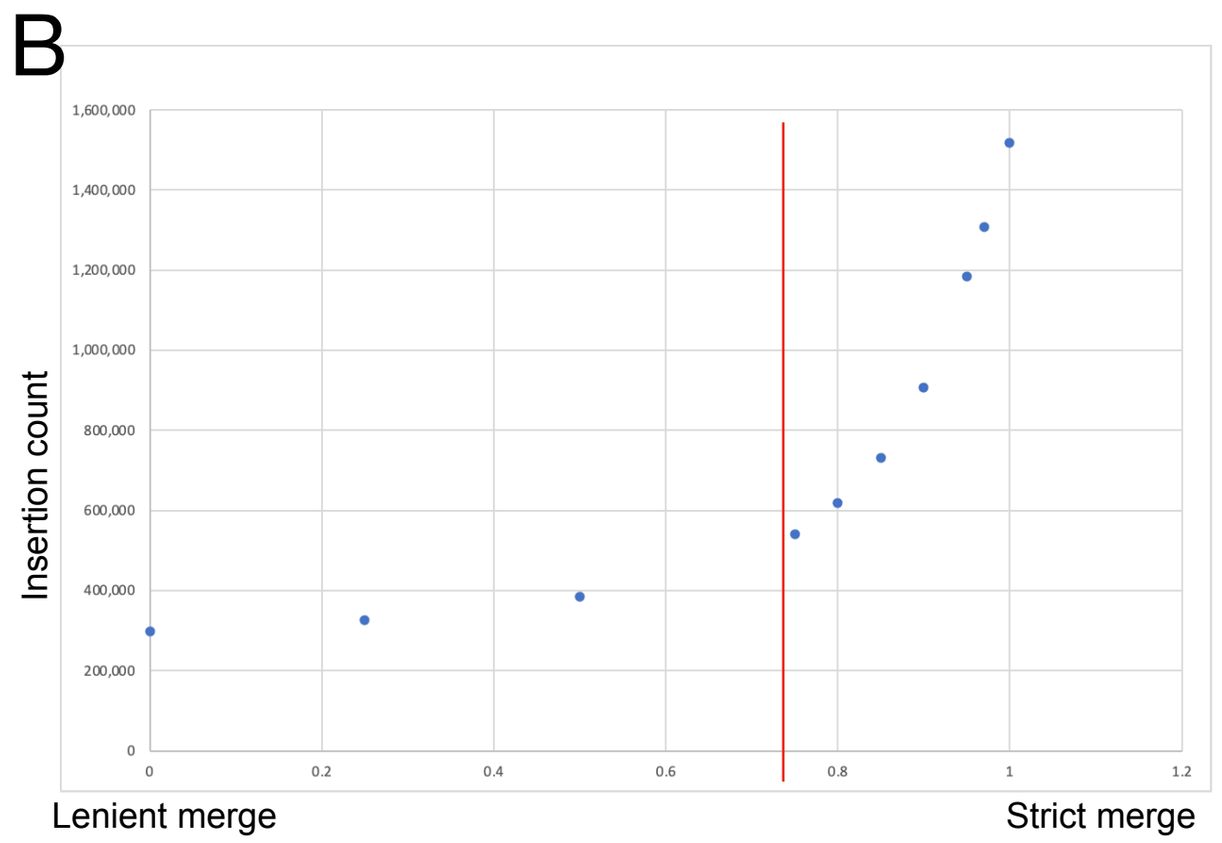
Supplementary Figure S26. Overview of the cohort-level structural variant structural variant analysis pipeline and HudsonAlpha internal SV allele count set. The teal star represents where the process links to and from Supplementary Figure 25 (individual case analysis). Datasets that appear on both figures are marked with a small teal star. Single sample datasets are represented in gold and marked with one person icon; BAMs are represented by parallelograms, single-sample VCFs are represented by curved-bottom rectangles, multi-sample VCFs are represented by a stack of curve-bottom rectangles, HiFi reads are represented by a hexagon, and images are represented by file icons. Software packages are represented by blue barred rectangles. Filtering steps are represented by pentagons and shaded tan. Among multi-sample VCFs: red represents the HudsonAlpha internal allele count dataset which contains samples from other projects; orange represents the proband cohort dataset described in this manuscript; and green represents public data. Sample curation is represented as a purple diamond. Manuscript tables are represented as dark purple table icons. As an additional signifier, proband single-sample data and analyses are marked with a singular person icon; this manuscript's cohort data and analyses are marked with a three-person icon; and general HudsonAlpha internal data (including individuals beyond the scope of this manuscript) are represented with a five-person icon. Figure created with draw.io.





Supplemental Figure S27. Median coverage by platform.





Supplemental Figure S28. Effect of Jasmine Merging Settings on Unique SV Count. A. Unique merged DEL/INV count as a function of the minimum reciprocal overlap required to merge SVs. B. Unique insertion count as a function of the minimum jaccard sequence similarity required to merge insertions. Duplication events were converted to insertions. Lenient merges (fewer resulting unique SVs) are toward the left on the axes, strict merges (more unique SVs) are toward the right. The selected settings are represented with a red vertical line. Selected settings were chosen approximately at the inflection points where unique SV counts began to increase rapidly with additional strictness.