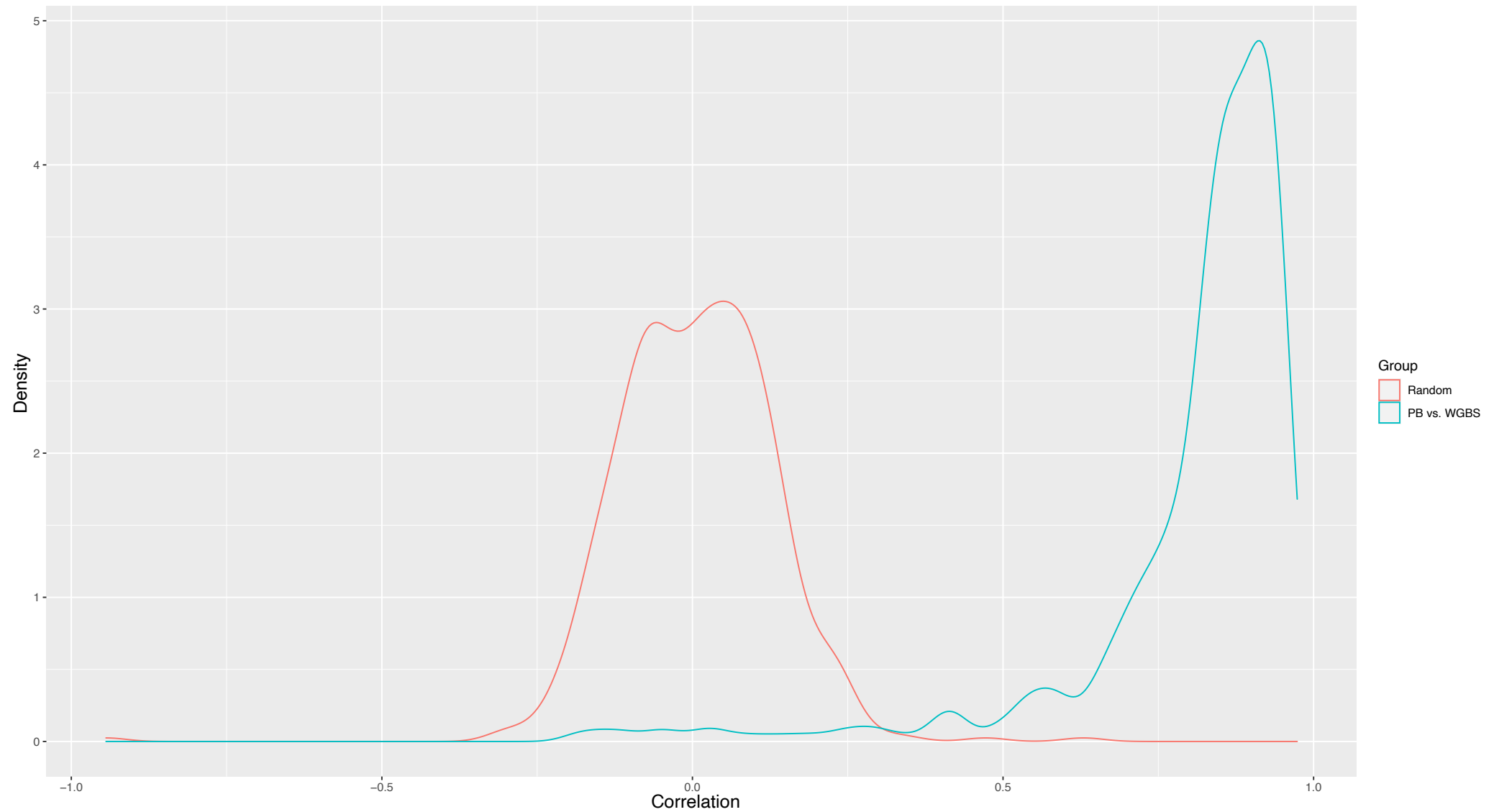


Extended Data

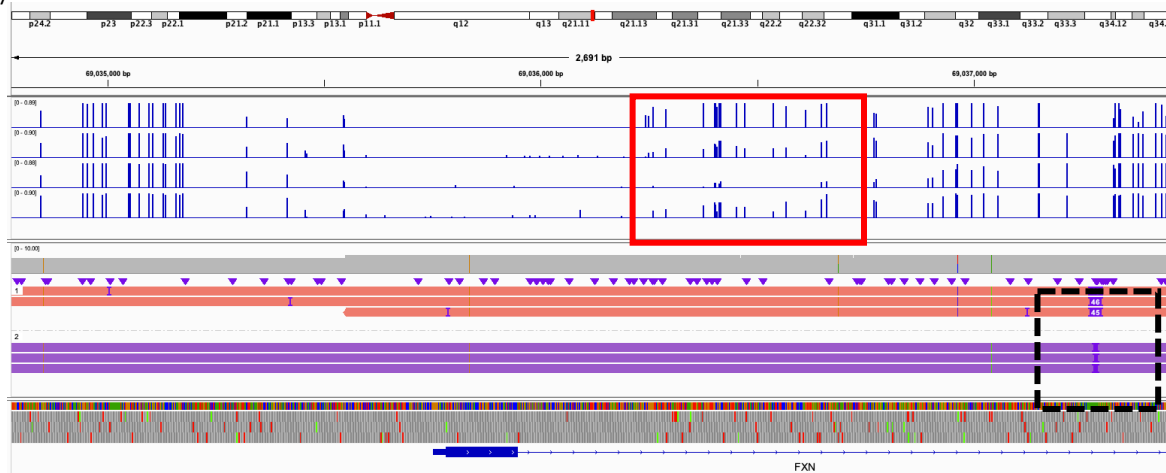
Direct haplotype-resolved 5-base HiFi sequencing for genome-wide profiling of hypermethylation outliers in a rare disease cohort

Extended Data 1: Density plot showing the distribution of correlation estimates (Pearson R, x-axis) of the top (N=500) variable CpGs obtained from 93 samples profiled by HiFi-GS and WGBS where the blue line represent CpG-based measures and red line permuted WGBS values, respectively .



Extended Data 2: Whole-genome bisulfite sequencing (WGBS) in patients with unstable repeat disorders. **A.** Genomic view of 2.7kb at the *FXN* locus showing an homozygote individual with Friedrich's Ataxia (FRDA1), two *FXN* expansion carriers and one individual without the expansion. Blue tracks shows mCpGs measured by WGBS (y-axis, 0-100%) with hypermethylation (red box) linked to repeat expansion. Black box (dashed line) shows location of repeat expansion by HiFi-GS. **B.** Genomic view of 3.7kb at the *FMR1* locus showing an hemizygote individual with Fragile X, one *FMR1* premutation carriers and two individuals without the *FMR1* expansion. Blue tracks shows mCpGs measured by WGBS (y-axis, 0-100%) with hypermethylation (red box) linked to repeat expansion. Black box (dashed line) shows location of repeat expansion by HiFi-GS.

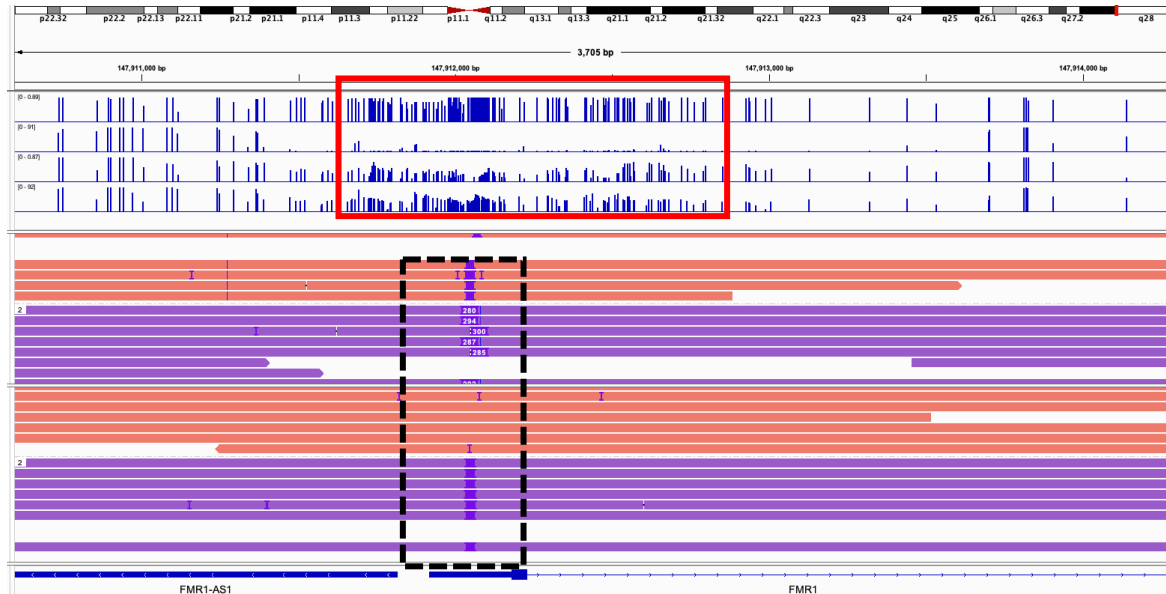
A



FXN expansion homozygote (FRDA1 patient)
FXN expansion carrier
FXN expansion negative
FXN expansion carrier

HiFi-GS showing non carrier, polymorphic repeat

B

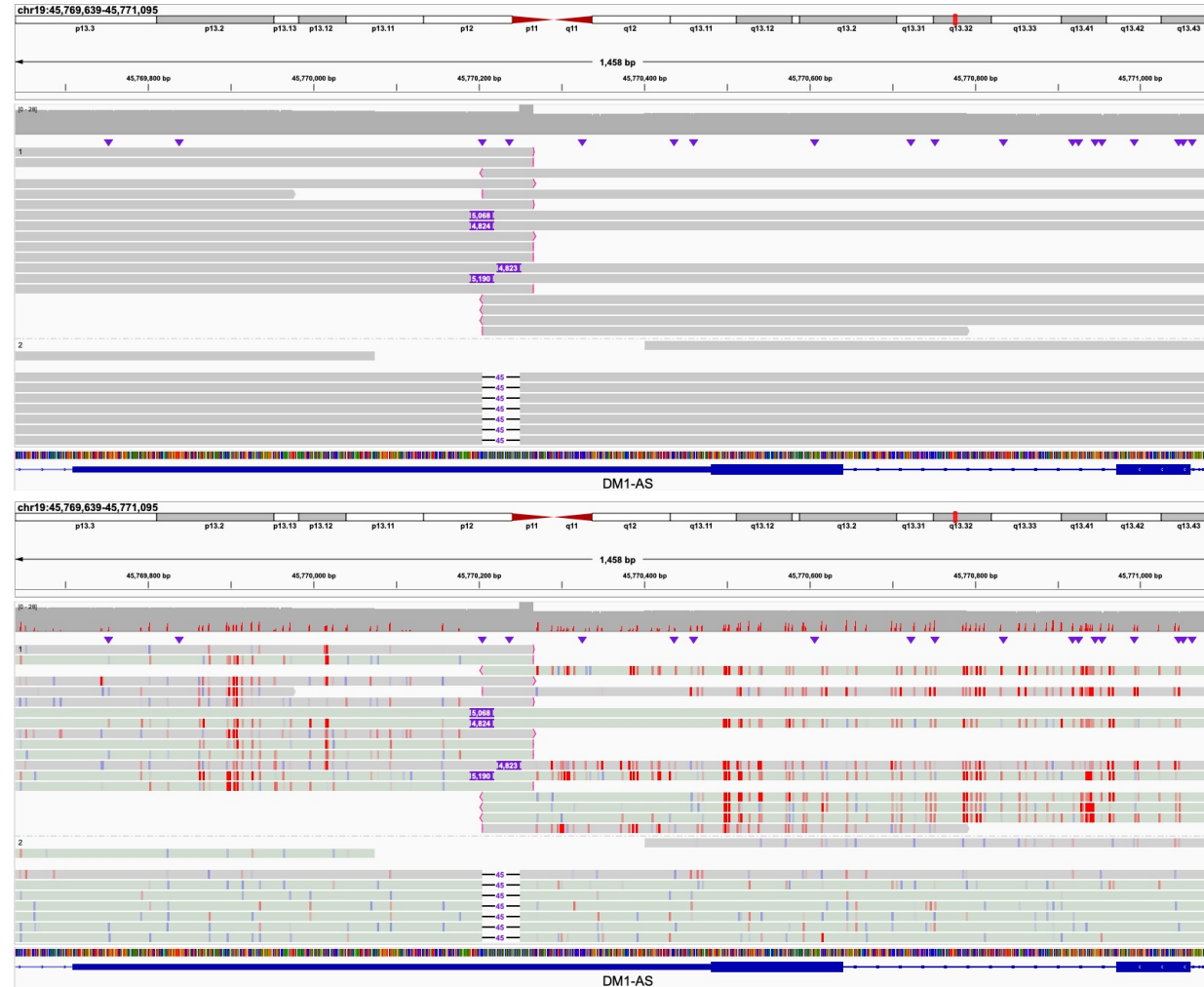


FMR1 full mutation hemizygote (Fragile-X patient)
FMR1 expansion negative male
FMR1 expansion premutation carrier (FXTAS) female
FMR1 expansion negative female

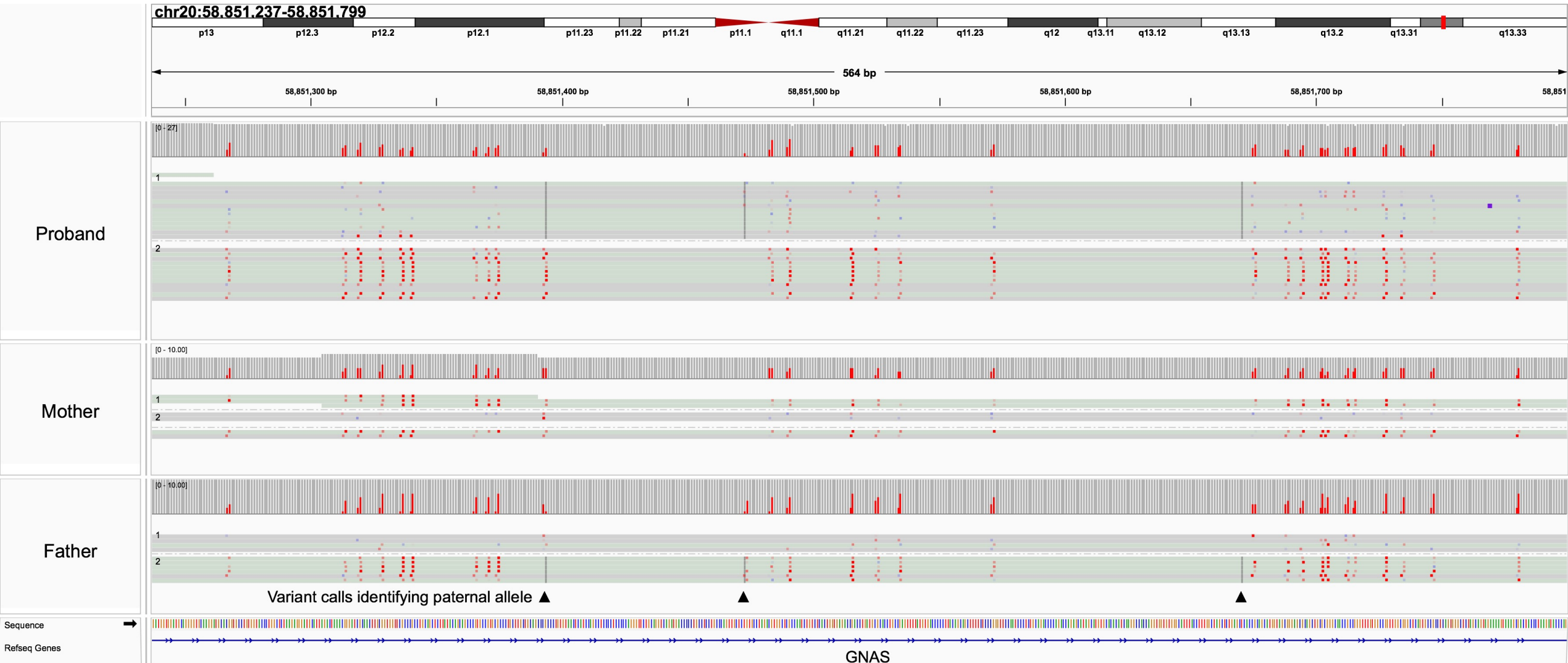
HiFi-GS showing the *FMR1* expansion premutation carrier (FXTAS) in female individual

HiFi-GS showing the *FMR1* expansion negative carrier in female individual

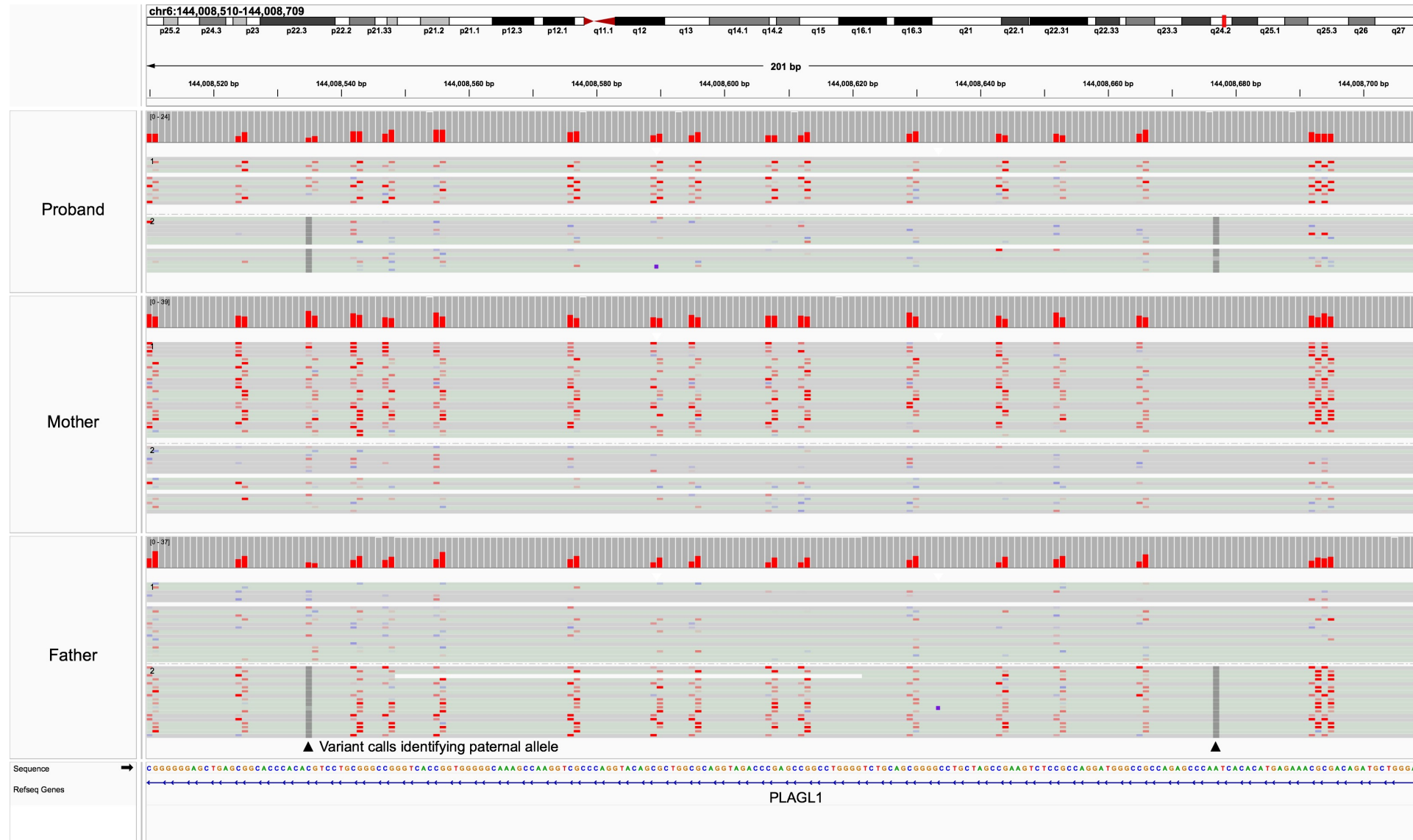
Extended Data 3: 5-base HiFi-GS in DM1 patient with DMPK repeat. Upper panel shows raw reads and variant calls identifying “soft-clipped” reads (i.e. reads with red ends), 45bp deletion and variable sized insertion (4,823bp–5,190) at repeat coordinates. Bottom panel illustrates the same reads now phased (maternal top, paternal bottom under the dotted line) and with CpG modification staining (blue indicating low methylation prediction and red indicating high methylation prediction from primrose algorithm). With phasing and CpG methylation staining the soft-clipped and insertion (expansion) spanning reads show consistent hypermethylation flanking the disease allele.



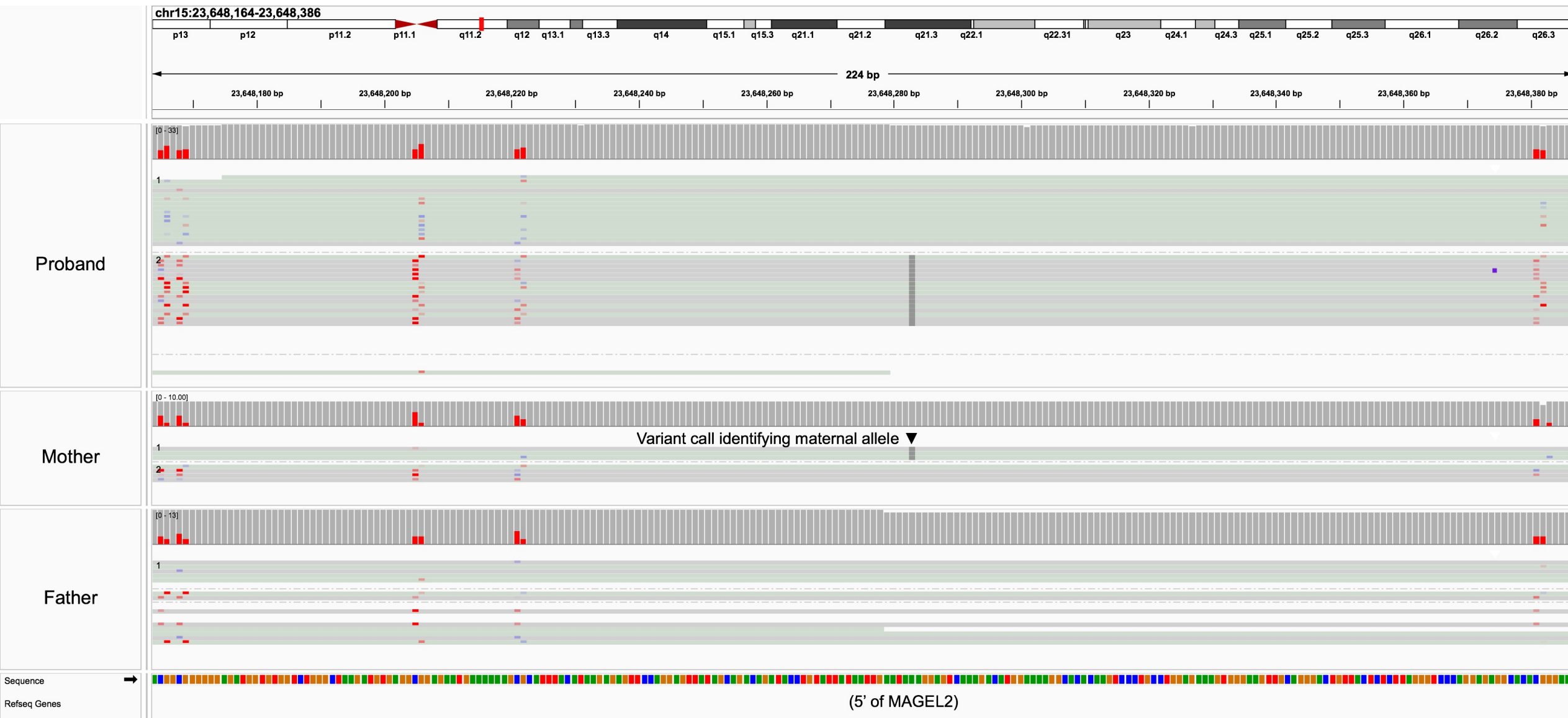
Extended Data 4: Haplotype-resolved HiFi-GS of an imprinted region (chr20: 58,851,237-58,851,799) at the *GNAS* locus associated with Albright hereditary osteodystrophy showing maternal allele-specific hypermethylation in proband. Top panel (red bars) shows mapped CpGs and individual CpG level is shown in proband, mother and father by blue<->red scale for low<->high methylation, respectively.



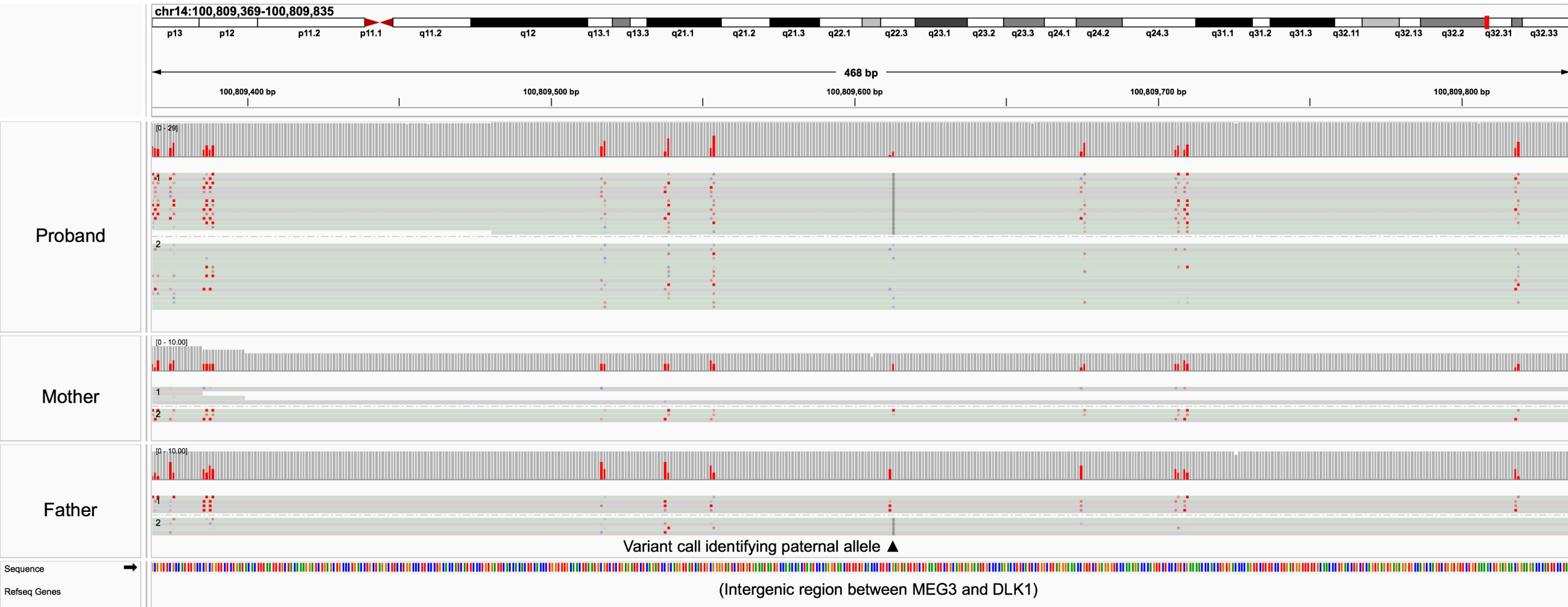
Extended Data 5: Haplotype-resolved HiFi-GS of an imprinted region (chr6:144,008,510-144,008,709) at the *PLAGL1* locus associated with transient neonatal diabetes mellitus showing maternal allele-specific hypermethylation in proband. Top panel (red bars) shows mapped CpGs and individual CpG level is shown in proband, mother and father by blue<->red scale for low<->high methylation, respectively.



Extended Data 6: Haplotype-resolved HiFi-GS of an imprinted region (chr15: 23,648,164-23,648,386) at the *MAGEL2* locus associated with Schaaf-Yang syndrome showing maternal allele-specific hypermethylation in proband. Top panel (red bars) shows mapped CpGs and individual CpG level is shown in proband, mother and father by blue<->red scale for low<->high methylation, respectively.

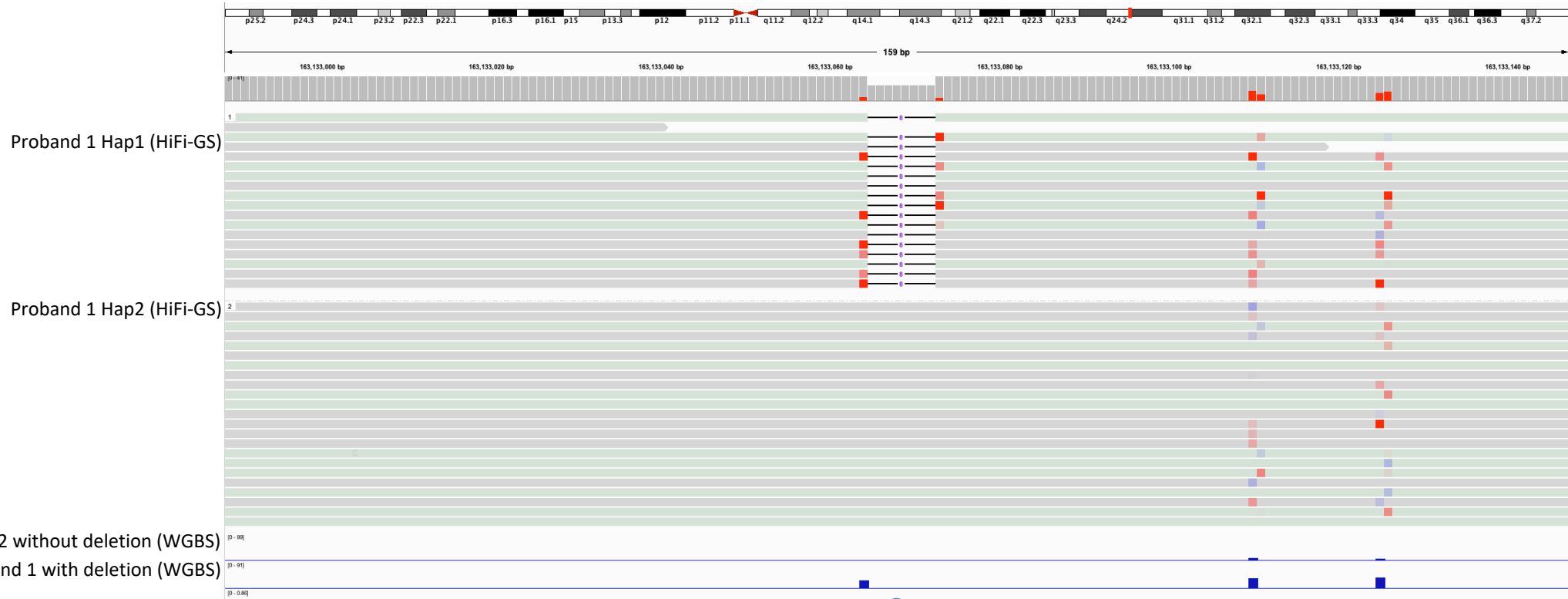


Extended Data 7: Haplotype-resolved HiFi-GS of an imprinted region (chr14: 100,809,369-100,809,835) at the *MEG3/DLM1* locus associated with Temple syndrome showing paternal allele-specific hypermethylation in proband. Top panel (red bars) shows mapped CpGs and individual CpG level is shown in proband, mother and father by blue<->red scale for low<->high methylation, respectively.



Extended Data 8: Validated rare deletion (8bp) in *cis*-regulatory element (CRE) mapped by ENCODE that cause proximal hypermethylation as shown by WGBS and HiFi-GS. **A.** Genomics view of 159 bp in an intergenic region comprising a 8 bp deletion (Proband 1 Hap1) resolved by HiFi-GS that results in proximal hypermethylation. Top panel (red bars) shows mapped CpGs and individual CpG level is shown in proband by blue<->red scale for low<->high methylation, respectively. **B.** Same 159bp region profiled by WGBS validating the hypermethylation in Proband 1 but not in Proband 2 who lacks the deletion. **C.** Zoomed in genomics view of the deletion showing overlap with CRE.

A

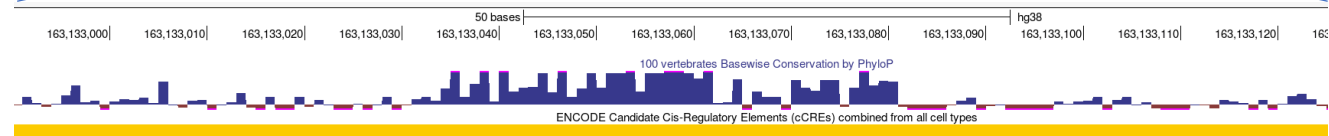


B

Proband 2 without deletion (WGBS)
Proband 1 with deletion (WGBS)

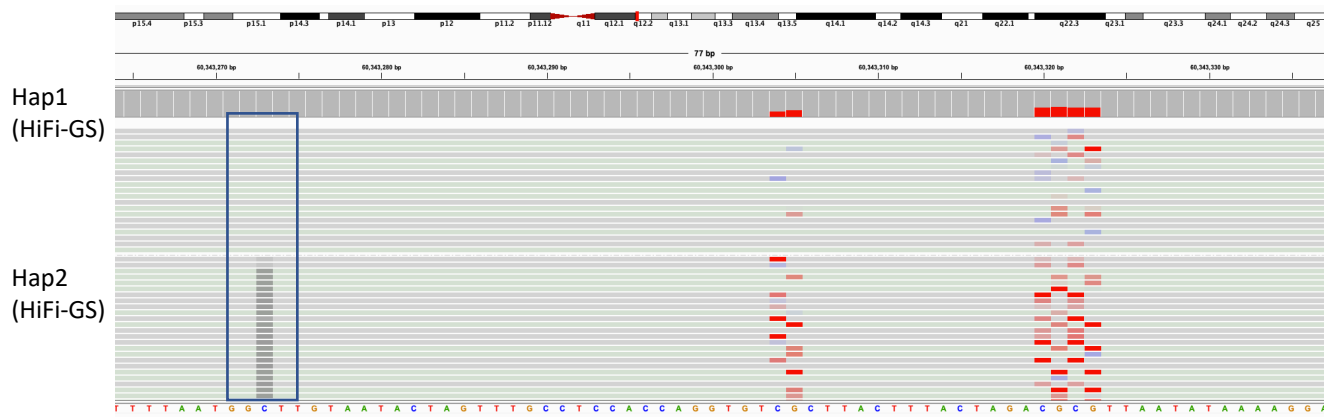


C

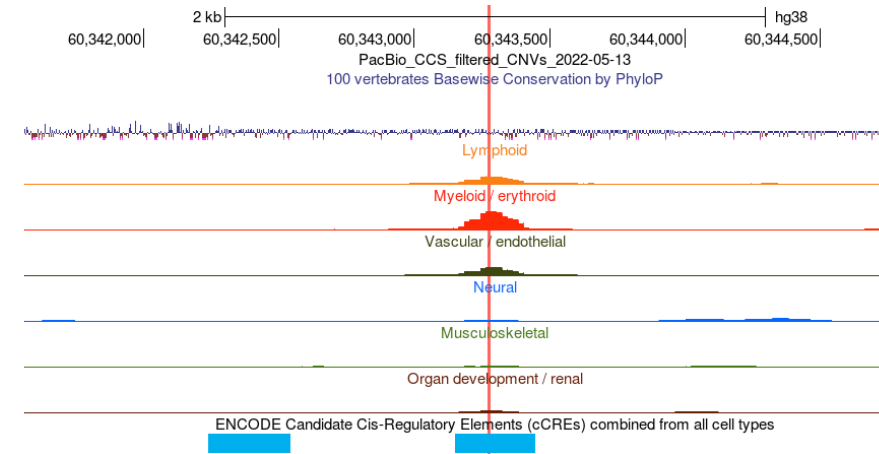


Extended Data 9: A rare single nucleotide variant (SNV) in *cis*-regulatory element (CRE) causing hypermethylation. **A.** Genomics view of 77 bp at an intergenic locus showing a SNV (blue box) resolved by HiFi-GS that results in proximal hypermethylation in one allele (Hap2). Red bars shows mapped CpGs and individual CpG level is shown in proband by blue<->red scale for low<->high methylation, respectively. **B.** Zoomed in region showing overlap of SNV (red line) with CRE as mapped by ENCODE.

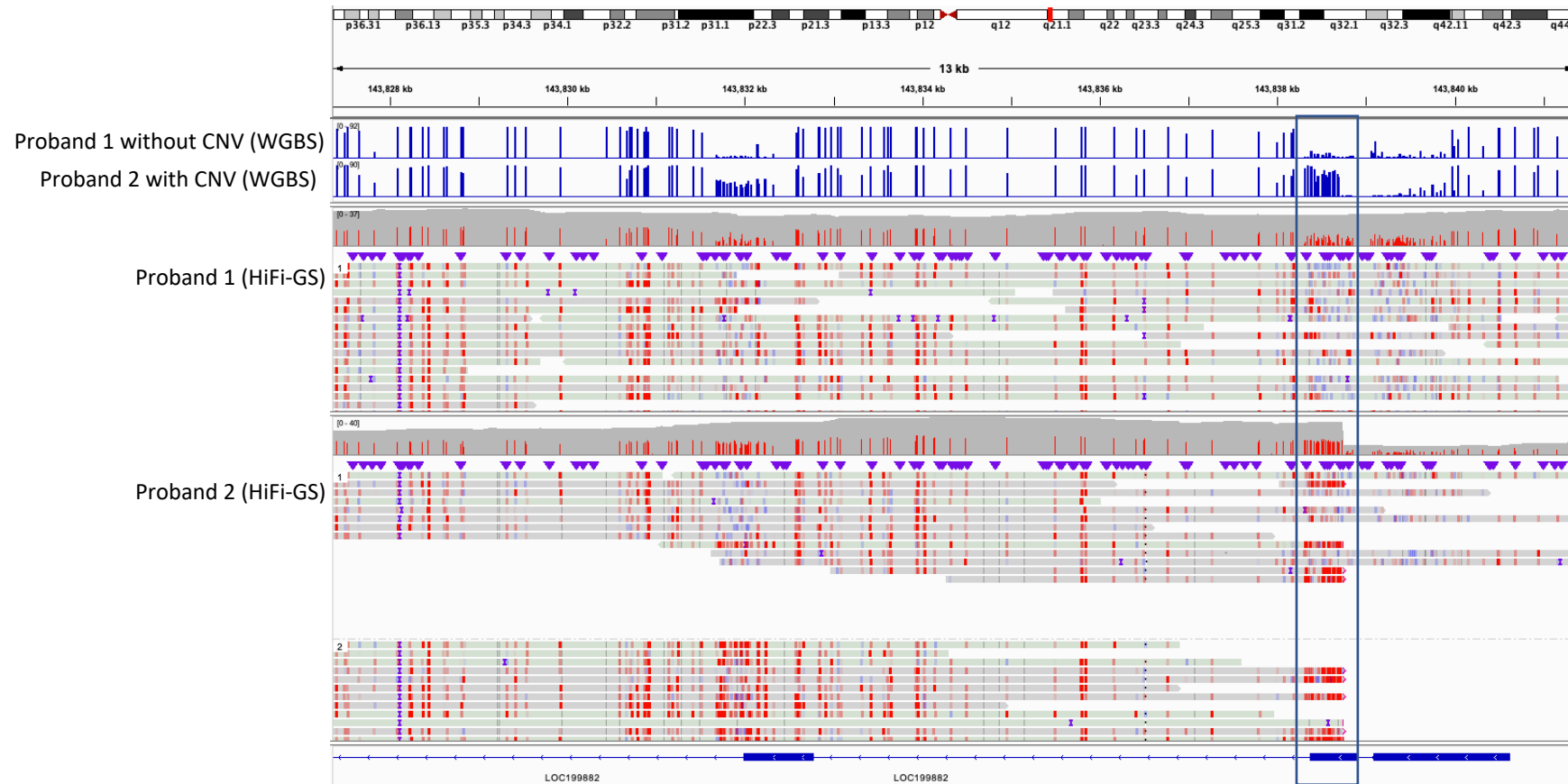
A



B



Extended Data 10: Structural variant (SV) that cause hypermethylation as shown by WGBS and HiFi-GS. Genomics view of 13kb at the *LOC199882* locus comprising a breakpoint of a SV (blue box, Proband 2) that results in proximal hypermethylation by HiFi-GS (red) and WGBS (blue tracks). Red bars in HiFi-GS tracks shows mapped CpGs and individual CpG level is shown in proband by blue<->red scale for low<->high methylation, respectively. Blue bars in WGBS tracks shows mCpG level (0-100%, y axis).



Extended Data 11: Previously uncharacterized repeat expansion at *ELF1* locus causing hypermethylation. **A.** Haplotype-resolved HiFi-GS reads for proband with repeat expansion (red box). **B.** Corresponding HiFi-GS reads with mCpG profiles showing hyper-mCpG linked to repeat expansion for Hap2 only. Individual CpG level by HiFi-GS is shown in proband by blue<->red scale for low<->high methylation, respectively.

A

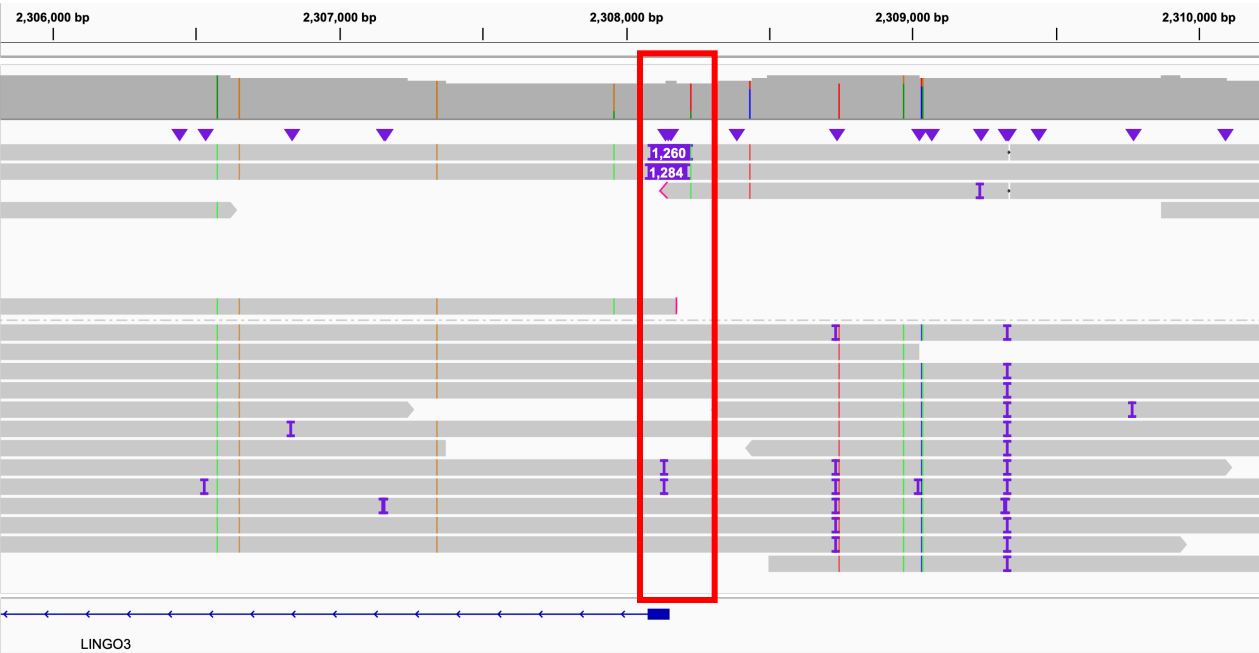


B

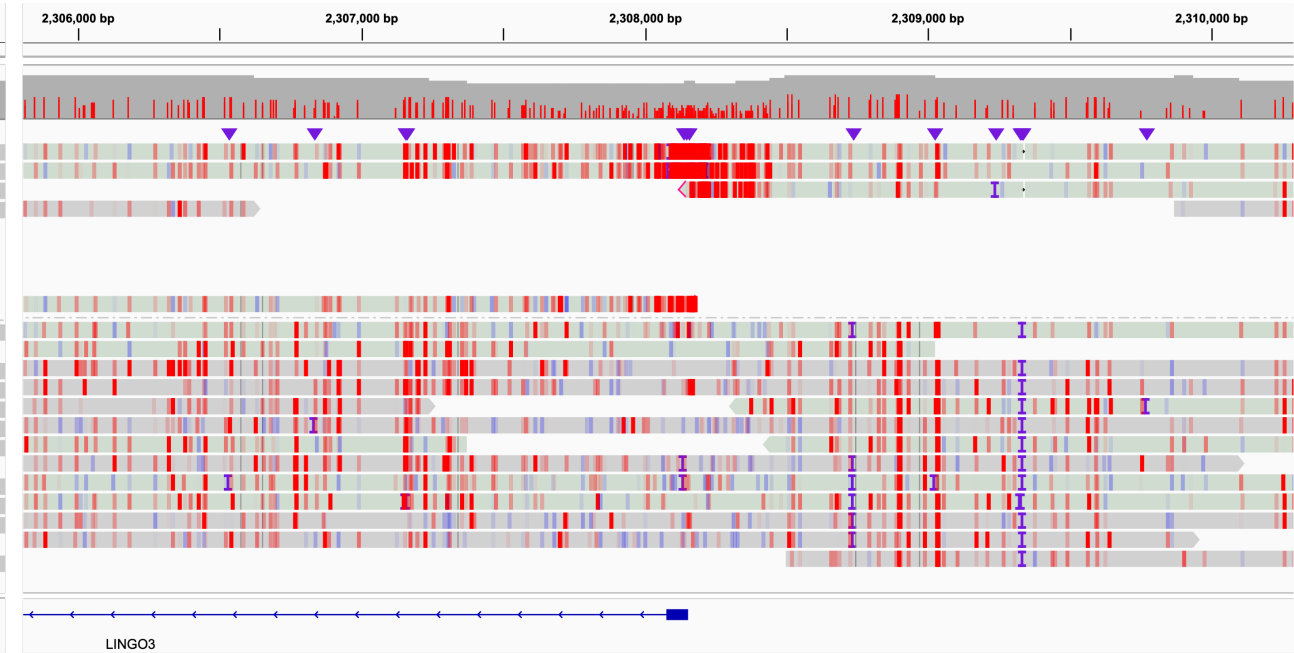


Extended Data 12: Previously uncharacterized repeat expansion at *LINGO3* locus causing promoter hypermethylation. **A.** Haplotype-resolved HiFi-GS reads for proband with repeat expansion (red box). **B.** Corresponding HiFi-GS reads with mCpG profiles showing hyper-mCpG linked to repeat expansion for Hap1 only. Individual CpG level by HiFi-GS is shown in proband by blue<->red scale for low<->high methylation, respectively.

A

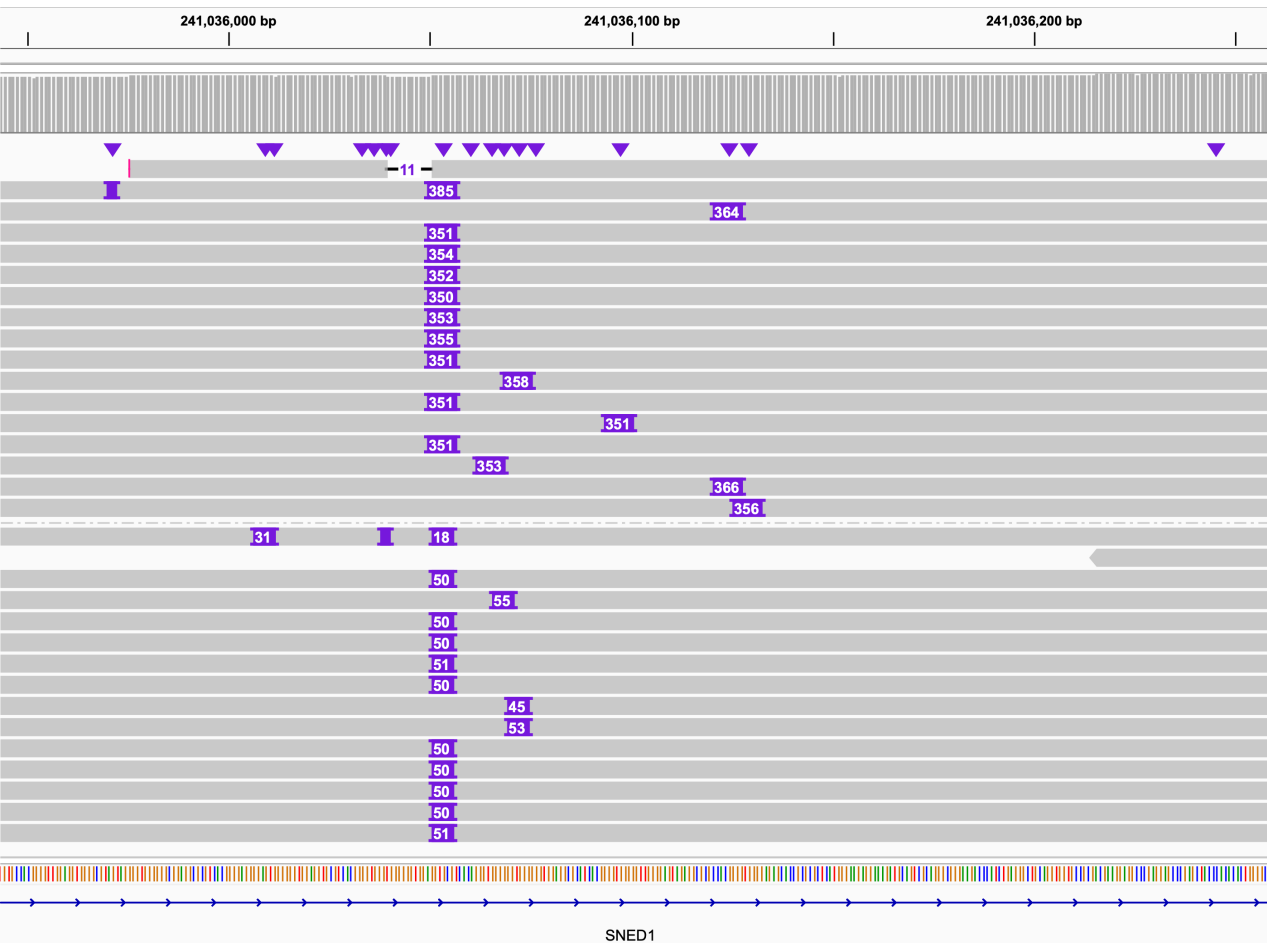


B

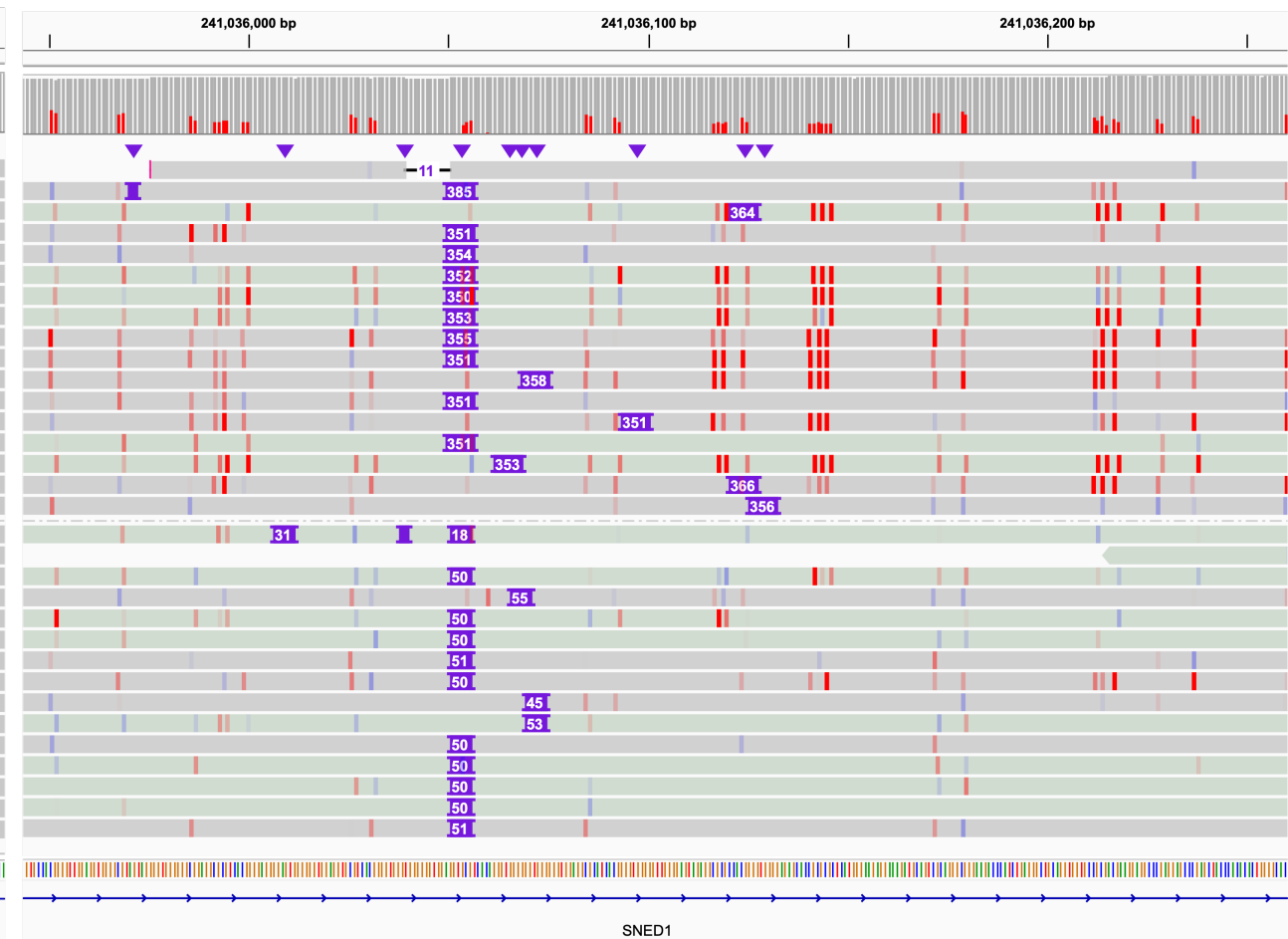


Extended Data 13: Previously uncharacterized repeat expansion at *SNED1* locus causing hypermethylation. **A.** Haplotype-resolved HiFi-GS reads for proband with repeat expansion. **B.** Corresponding HiFi-GS reads with mCpG profiles showing hyper-mCpG linked to repeat expansion for Hap1 only. Individual CpG level by HiFi-GS is shown in proband by blue<->red scale for low<->high methylation, respectively.

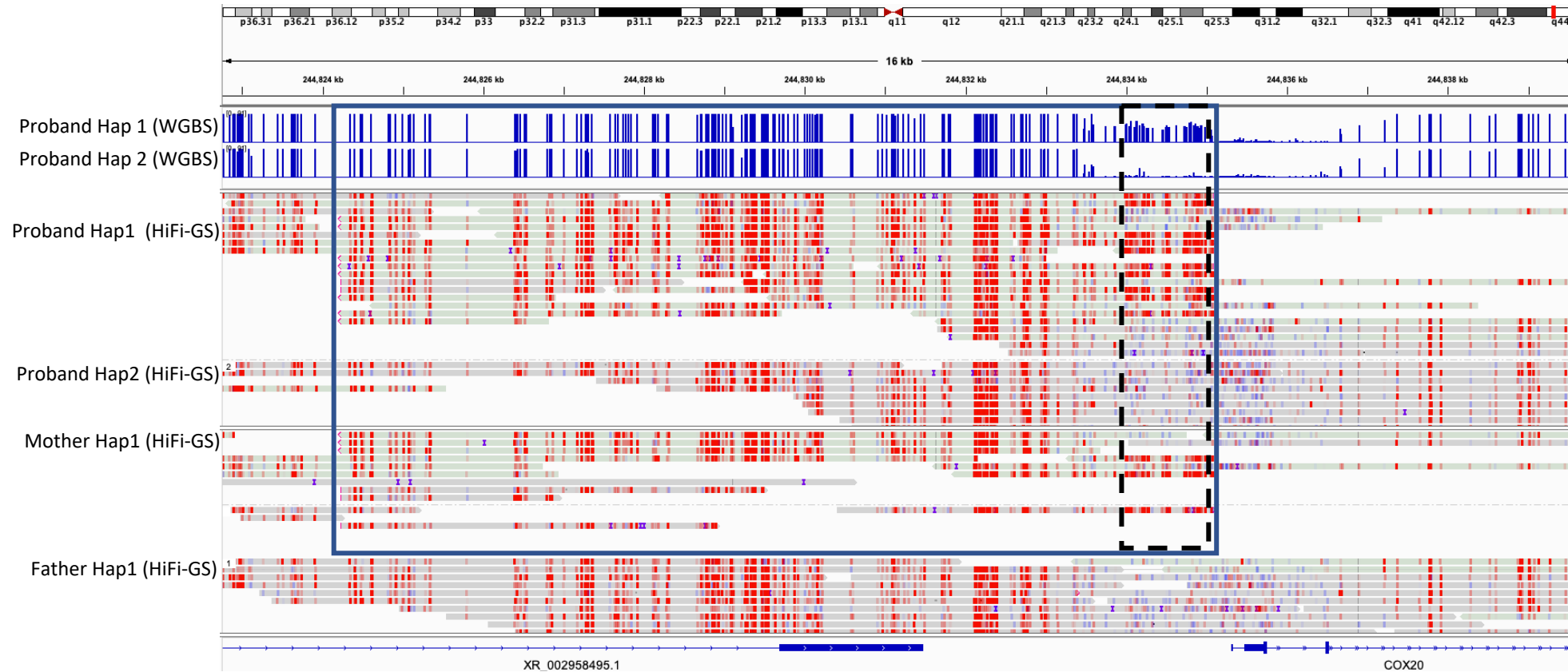
A



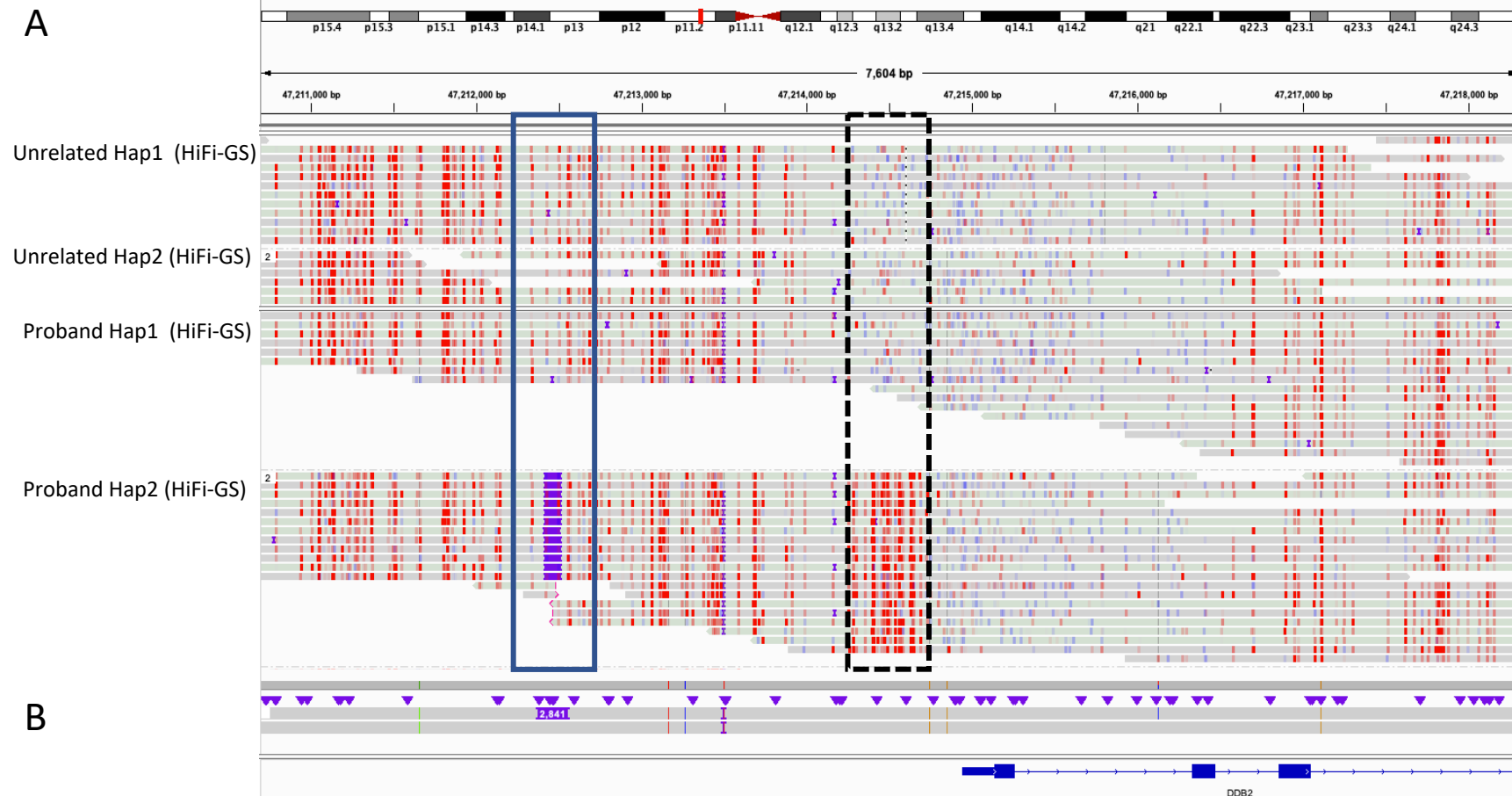
B



Extended Data 14: Rare duplication (10.6kb) causing “compensatory” promoter hypermethylation. Genomics view of 16kb at the *COX20* disease locus comprising an maternally inherited duplication (Proband Hap1 and Mother Hap1, blue box) resolved by HiFi-GS that results in a compensatory promoter hypermethylation (black box, dashed line). Individual CpG level by HiFi-GS is shown in proband, mother and father by blue<->red scale for low<->high methylation, respectively. Phased WGBS reads (blue tracks) in proband confirms allele-specific hypermethylation at duplicated locus.

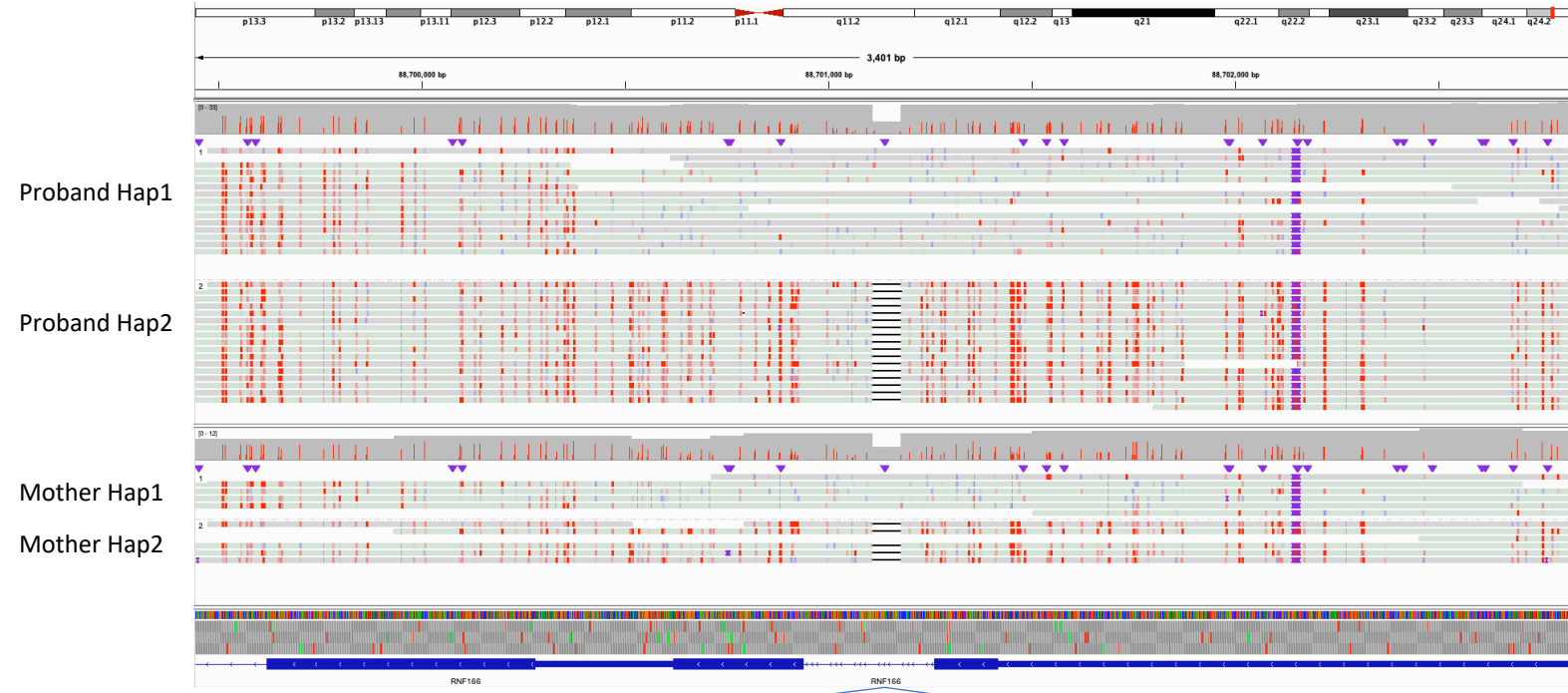


Extended Data 15: Rare insertion (2.6kb) in *cis* causing proximal regulatory element hypermethylation. **A.** Genomics view of 7.6kb at the *DDB2* disease locus comprising an insertion (Proband Hap2, blue box) resolved by HiFi-GS that results in 1kb proximal promoter hypermethylation (black box, dashed line). Individual CpG level by HiFi-GS is shown in proband and unrelated sample by blue<->red scale for low<->high methylation, respectively. **B.** Genome assembly of proband Hap1 and Hap2 shows the size of the insertion.

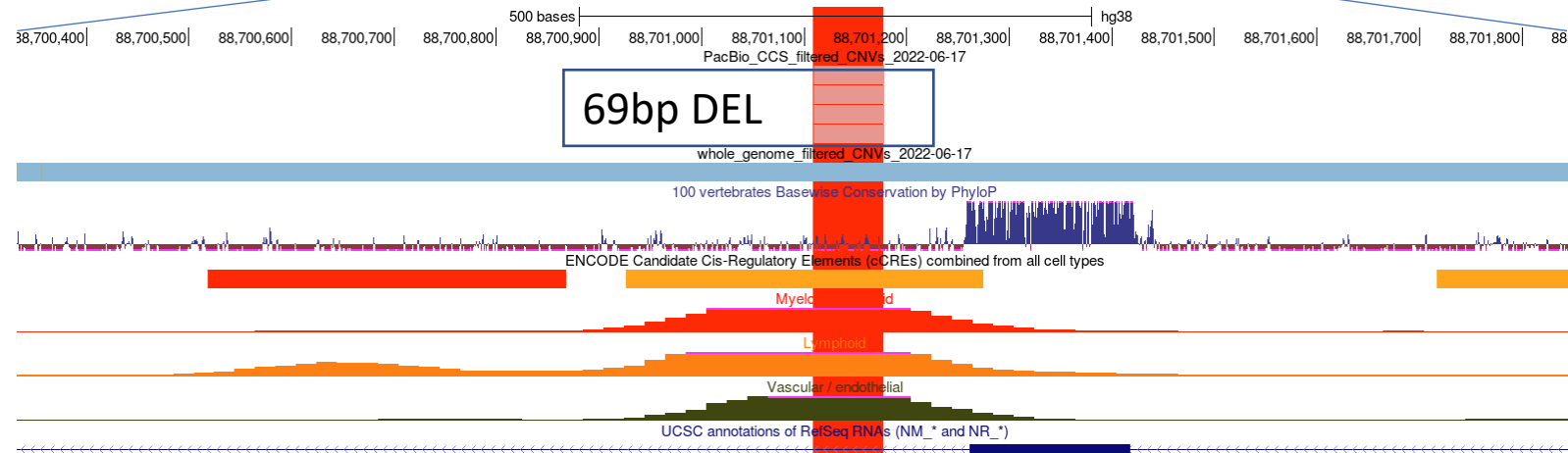


Extended Data 16: Rare (GnomAD 0.0351%) deletion (69bp) in *cis*-regulatory element (CRE) mapped by ENCODE that cause 2kb hypermethylation. **A.** Genomics view of 3401 bp at the *RNF166* locus comprising an inherited deletion (Proband Hap2 and Mother Hap2) resolved by HiFi-GS that results in 2kb hypermethylation surrounding the deletion. Top panel (red bars) shows mapped CpGs and individual CpG level is shown in proband and mother by blue<->red scale for low<->high methylation, respectively. **B.** Zoomed in genomics view of the deletion showing overlap with CRE.

A

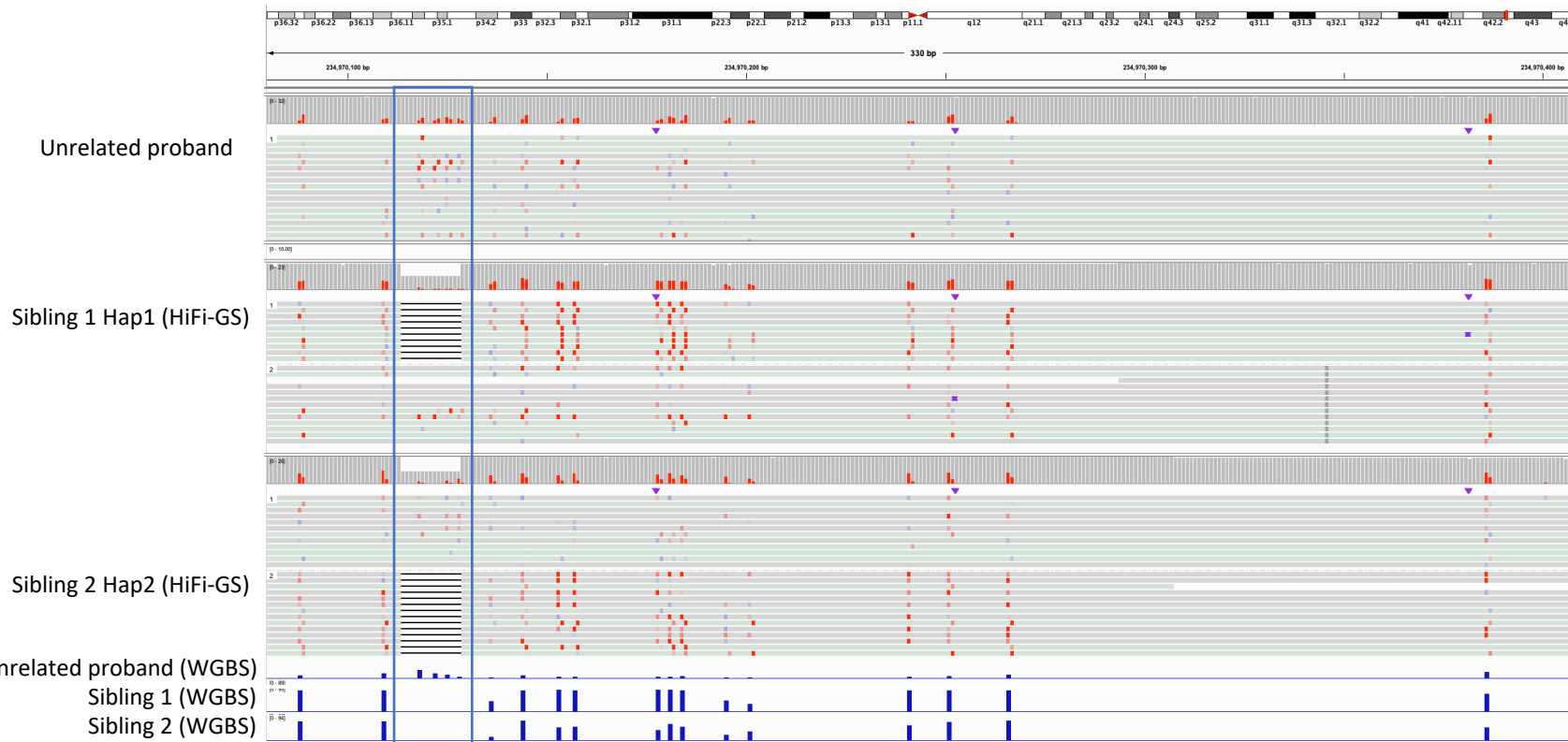


B



Extended Data 17: A 13bp deletion shared in siblings maps to constrained cis-regulatory element (CRE) and cause hypermethylation. **A.** Genomics view of 330 bp in an intergenic region comprising a 13 bp deletion (blue box) shared in sibling (Sibling 1; Hap1 and Sibling 2; Hap2) resolved by HiFi-GS that results in proximal hypermethylation. Unrelated proband lacks the deletion and show hypomethylation of nearby region. Red bars shows mapped CpGs and individual CpG level is shown in proband by blue<->red scale for low<->high methylation, respectively. **B.** Same 330bp region profiled by WGBS validating the hypermethylation in Siblings but not in Proband 2 who lacks the deletion.

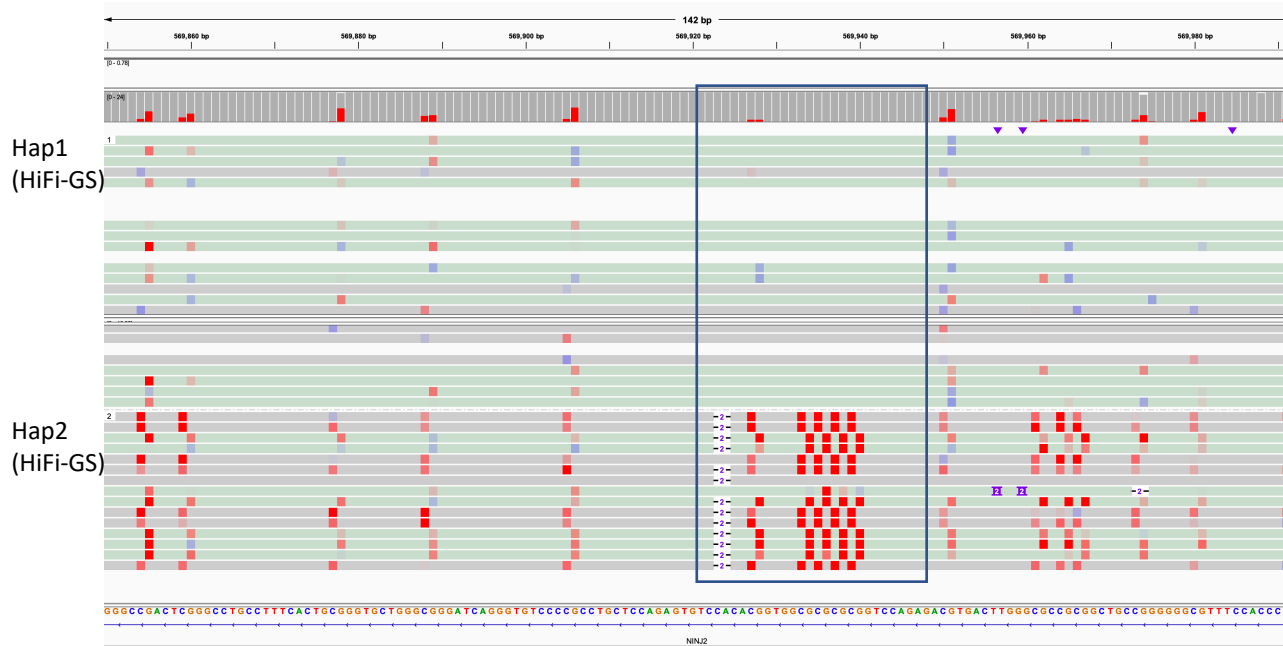
A



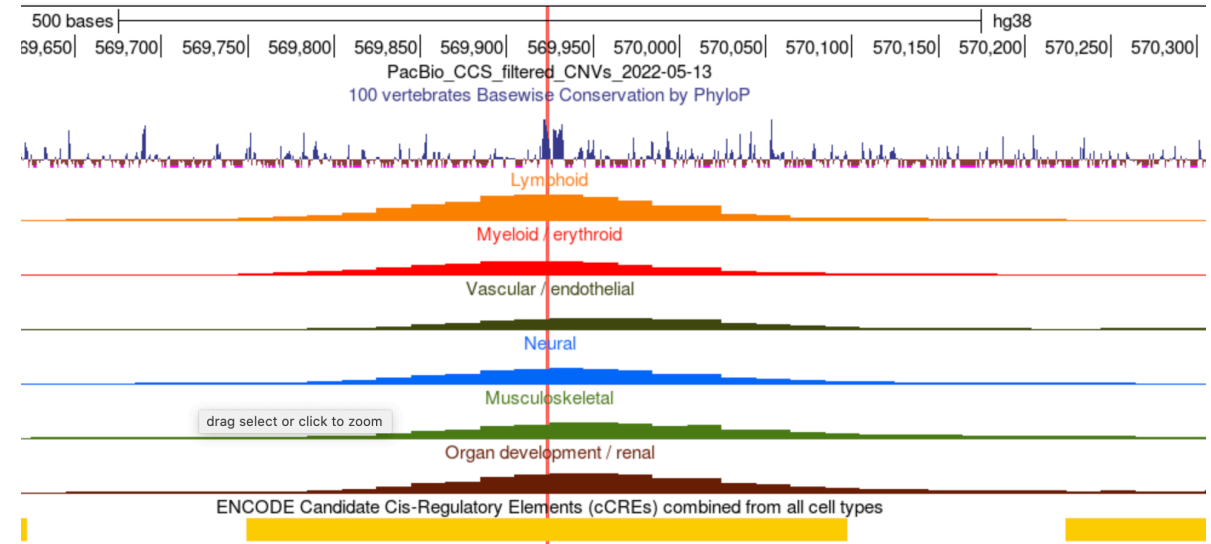
B

Extended Data 18: A conserved 2bp deletion in cis-regulatory element (CRE) causing hypermethylation. **A.** Genomics view of 142 bp at the *NINJ2* locus deletion (blue box) comprising a 2bp deletion resolved by HiFi-GS that results in proximal hypermethylation (blue box) in one allele (Hap2). Red bars shows mapped CpGs and individual CpG level is shown in proband by blue<->red scale for low<->high methylation, respectively. **B.** Zoomed in region of the deletion showing overlap with CRE as mapped by ENCODE.

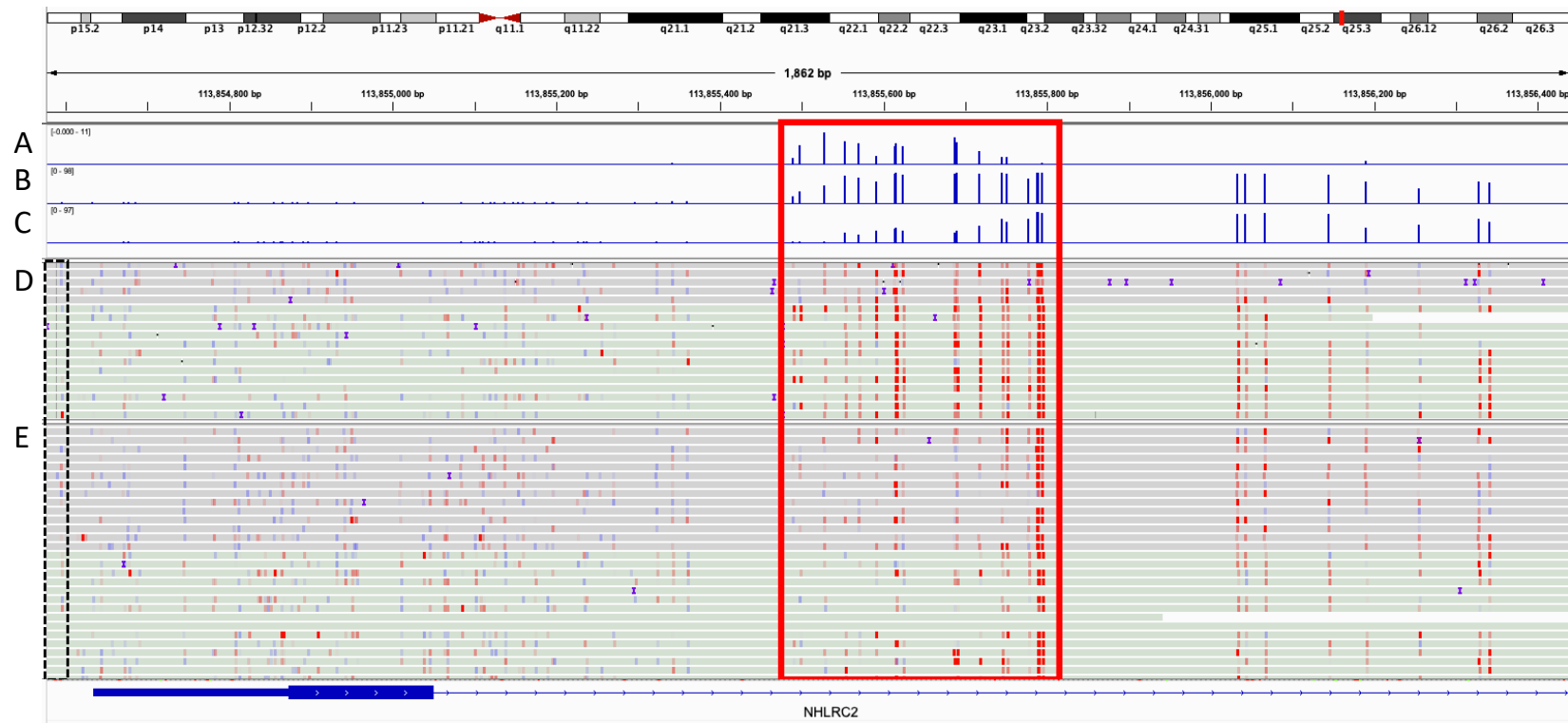
A



B



Extended Data 19: Statistical Determination of the Extent of the Hypermethylation Landscape. To assess the size of the extreme CpG methylation window linked to a rare, local (cis) variant, 5-base HiFi-GS reads overlapping a candidate cis-variant were fetched. For the autosomal recessive disease gene NHLRC2, an A to C transversion (chr10:113,854,859, black box with dashed line) maps 900bp upstream of a rare hyper-mCpG tile (red box). **A.** P-values (-log10) from 2-by-2 Fisher's exact test testing each CpG state (methylated vs. unmethylated) in reads with rare C versus common A-allele carrying reads, respectively. **B.** Average CpG methylation level (y-axis, 0-100%) across all C carrying reads (rare). **C.** Average CpG methylation level (y-axis, 0-100%) from the population prevalent A-allele. **D.** Hypermethylation shows statistical significance for the phased C-allele from haplotype-resolved HiFi-GS data across ~200-300bp – relatively typical SNV-linked hyper-mCpG landscape often corresponding to single hypomethylated regulatory element. **E.** Haplotype-resolved HiFi-GS data showing hypomethylation for the common A allele in the same reads. In D and E, blue indicates low methylation prediction and red indicating high methylation prediction based on the primrose algorithm.



Extended Data 20: Validation of Hyper-mCpG events by allelic expression profiling. Hyper-mCpGs tiles mapping within 1kb of a transcription start site (TSS) and present in an individual where long-read isoform sequencing (IsoSeq) was performed on proband-specific iPSCs were selected for follow-up analysis. The TSS was linked to the closest consensus coding sequence (CCDS) region and Isoseq reads informative for a heterozygous SNV were selected if covered at more than 4 reads for case and controls. The latter corresponded to all other probands not 'carrying' the hyper-mCpG tile. Fisher's exact test was performed on reference and alternative allele for case and control IsoSeq reads. In total of 44 transcripts were tested of which N=3 showed Fisher's $p < 0.05$ and consequently allele-specific expression. Top significant genes are shown at **A. *NUP153* locus** for differential expression of A and G alleles (red box; $P=0.005$) and **B. *FES* locus** for differential expression of T and G alleles (red box $p=0.008$).

