Data availability

All data are available from the corresponding author on reasonable request at the Genome Sequence Archive database (https://ngdc.cncb.ac.cn/,accession number: PRJCA063280).

Ethics Statement

The human study protocol was reviewed and approved by the Ethics Committee of the Obstetrics and Gynecology Hospital of Fudan University (Approval No. kyy2021-197). Prior to participation, written informed consent was obtained from the legal guardians of the 12-year-old minor proband, and assent was obtained from the minor participant. All animal experimental procedures in this study were approved by the Institutional Animal Care and Use Committee (IACUC) of the Obstetrics and Gynecology Hospital of Fudan University.

Supplementary Figure and Table Legends

Supplementary Figure S1. Results of SNP analysis in Case one.

Supplementary Figure S2. Long-read DNA-seq revealed that the ovotesticular DSD resulted from the dispermic fertilization of one oocyte and the PB2. Hap: Haplotype.

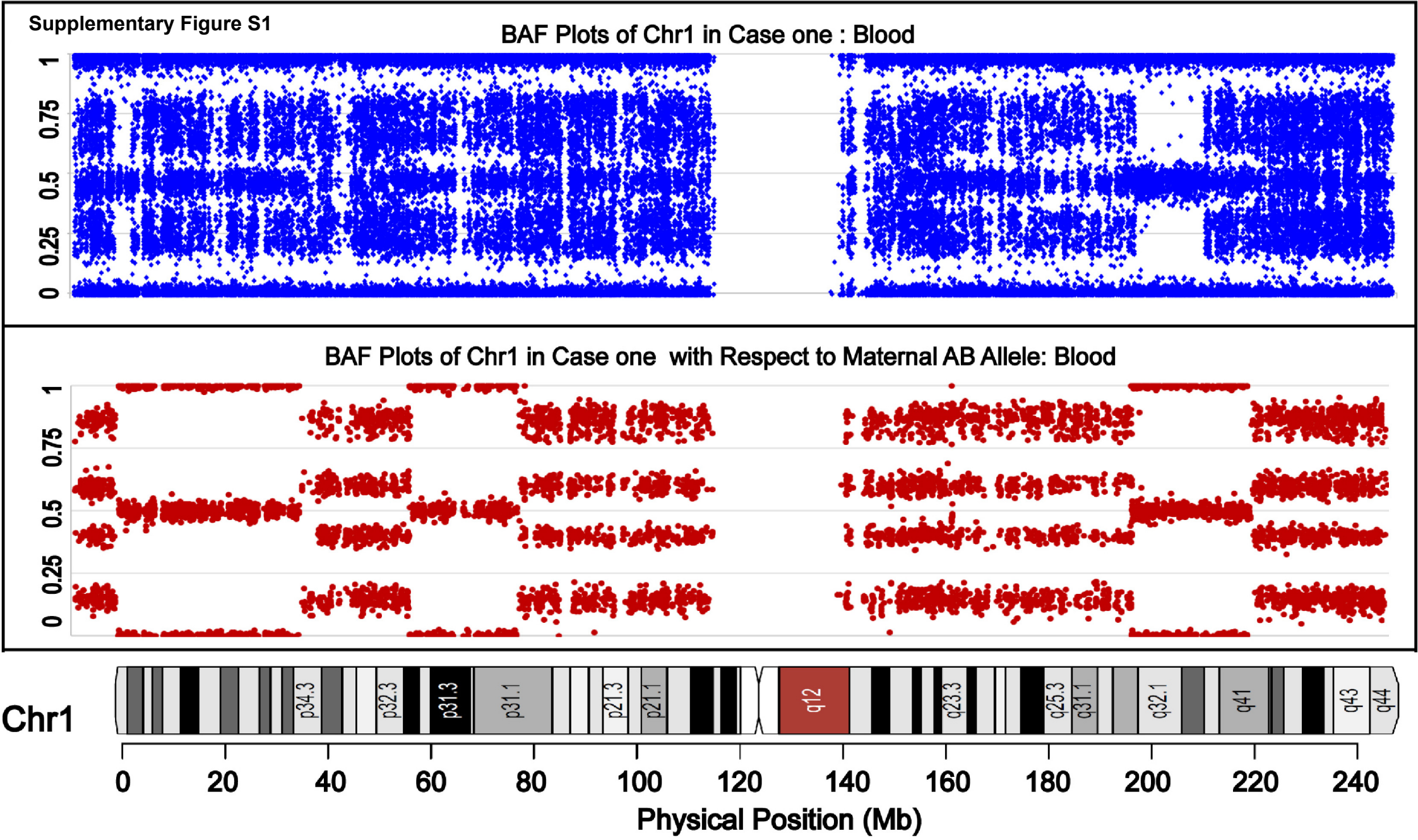
Supplementary Figure S3. Mice produced by injecting double sperm heads into a single oocyte. a, PB2 retention experimental flowchart. b, Representative blastocysts and a mouse pup generated by double sperm head injection. Scale bars, 100 μm for blastocysts and 1 cm for the pup. c, The qPCR results show inter-tissue heterogeneity in chimerism levels in Pup #1 and #4, as determined by Y/X chromosomal ratio quantification.

Supplementary Table S1. Results of STR analysis in Case one.

Supplementary Table S2. Results of SNP analysis in internal controls from artificially mixed twins.

Supplementary Table S3. Results of STR analysis in ovotesticular DSD proband. Supplementary Table S4. Systematic comparison of predicted and observed genomic signatures across alternative models of 46, XX/46, XY chimerism.

Supplementary Figure S1



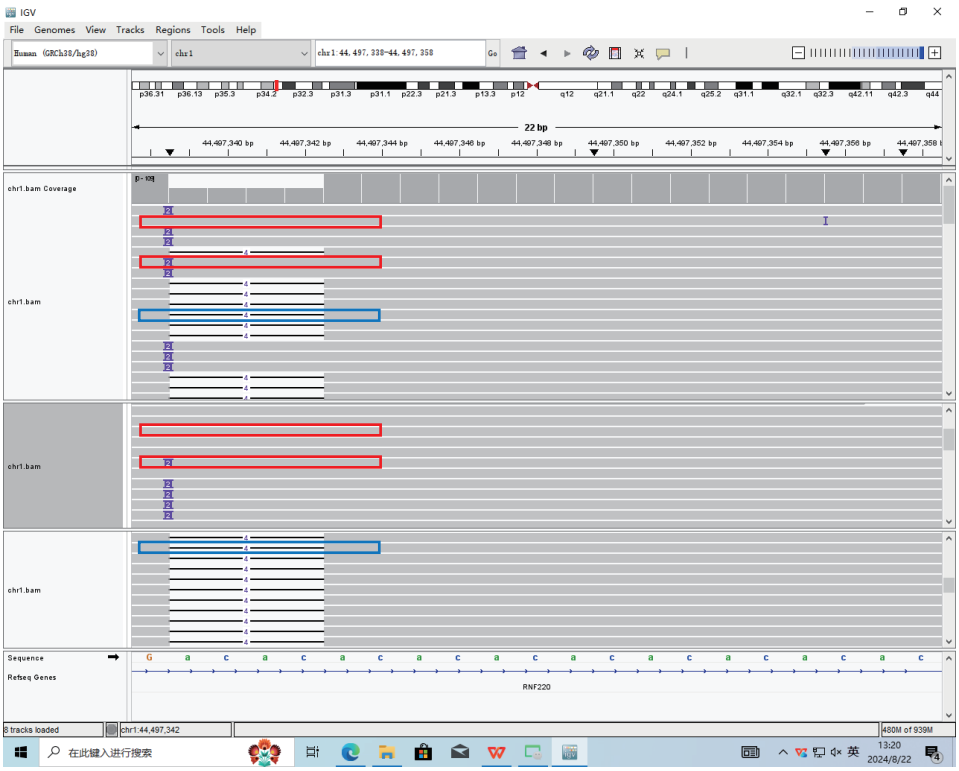
Supplementary Figure S2

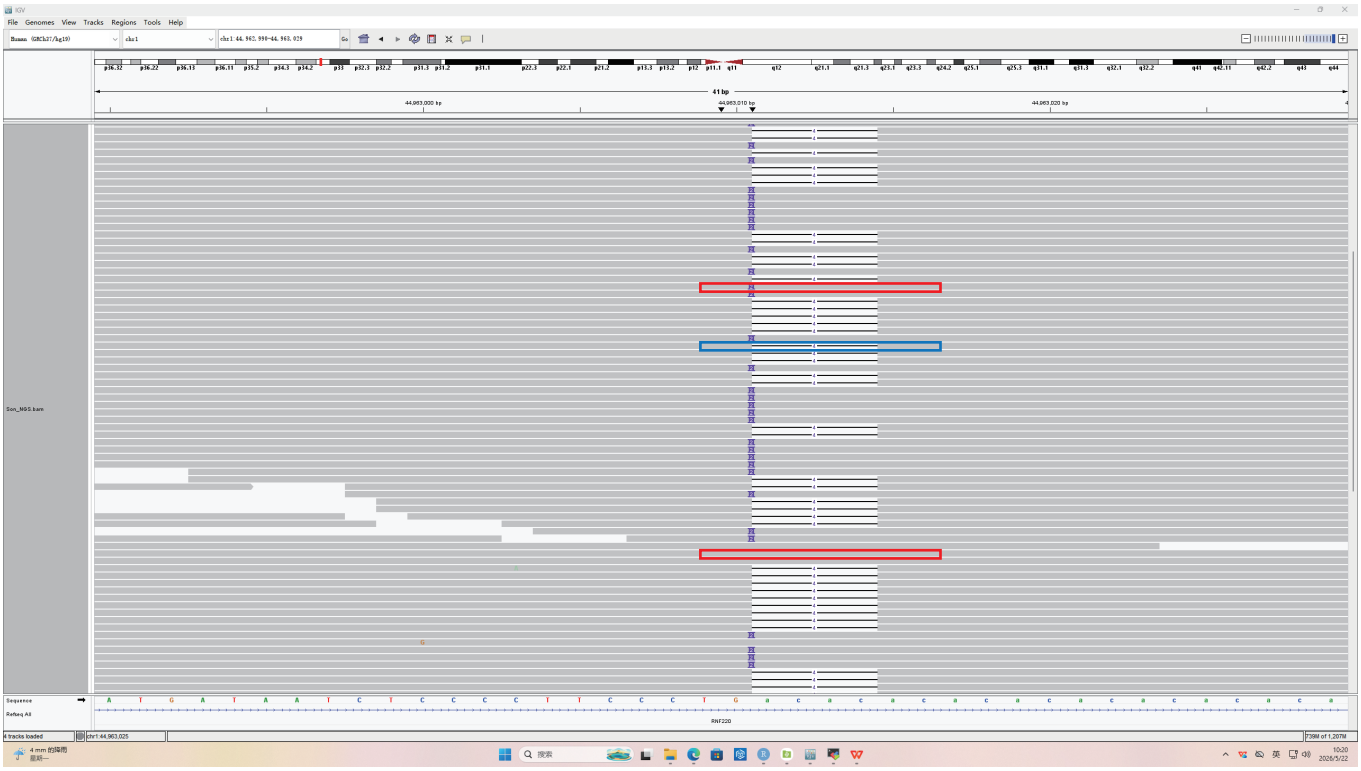
chr1

Case Two:
WT/Hap1/2

Maternal :
WT/Hap1

Paternal:
Hap2





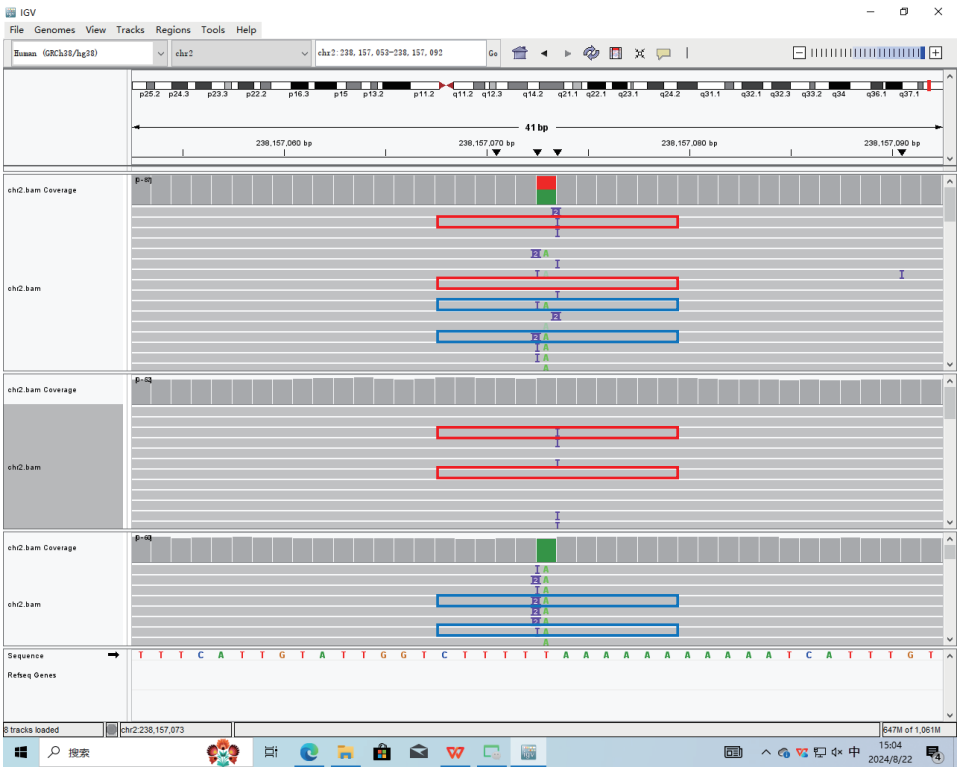
Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2

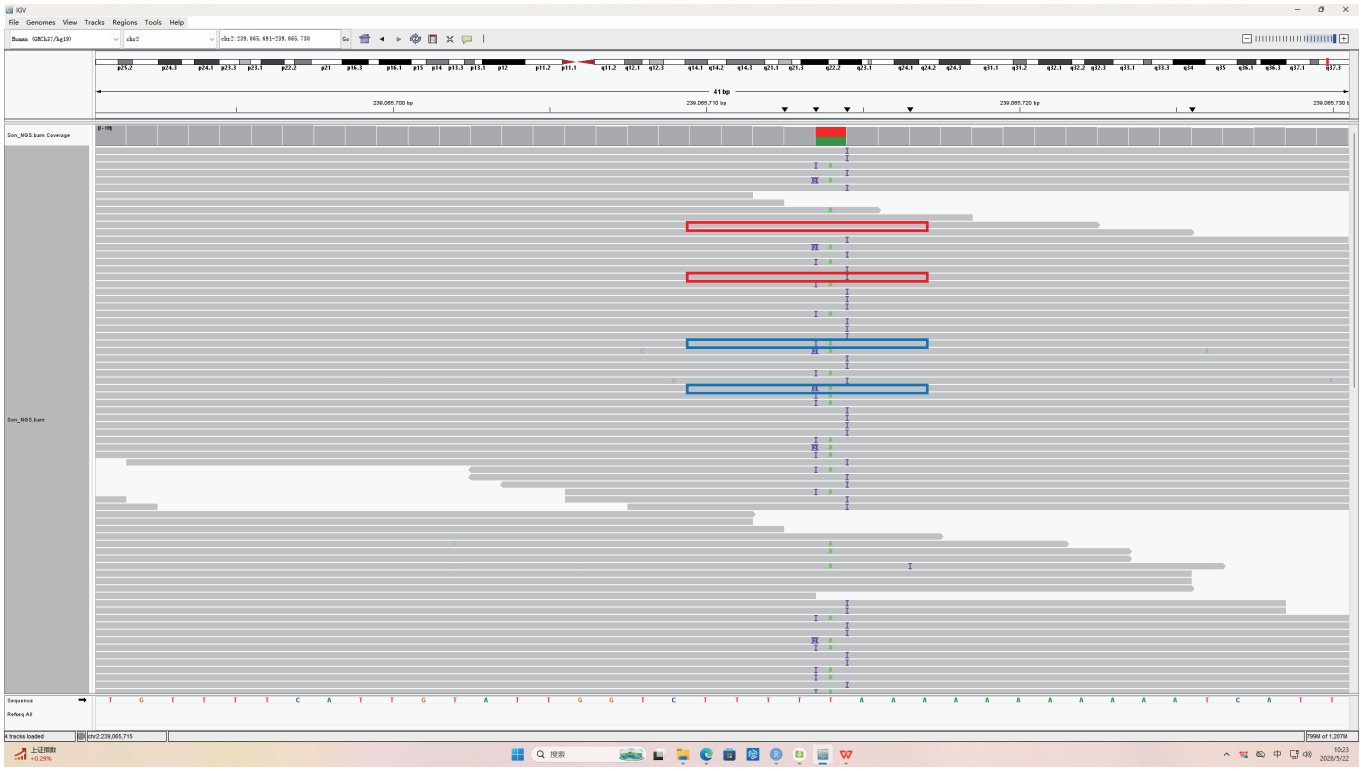
chr2

Case Two:
WT/Hap1/Hap2/Hap3

Maternal :
WT/Hap1

Paternal:
Hap2/Hap3





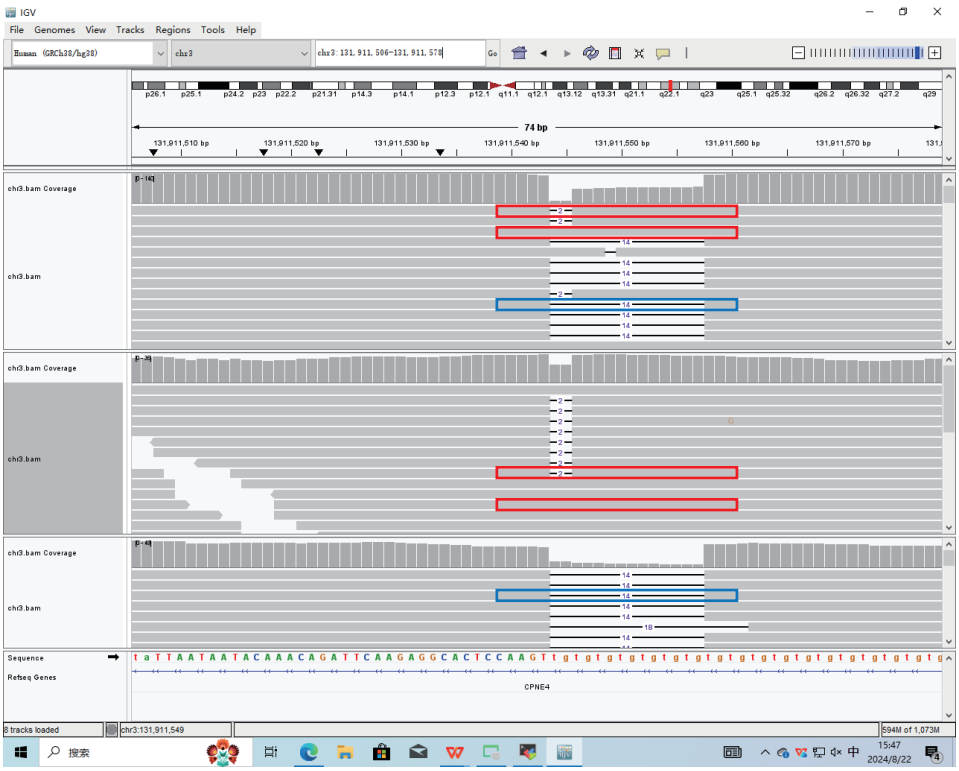
Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2/Paternal_Hap3

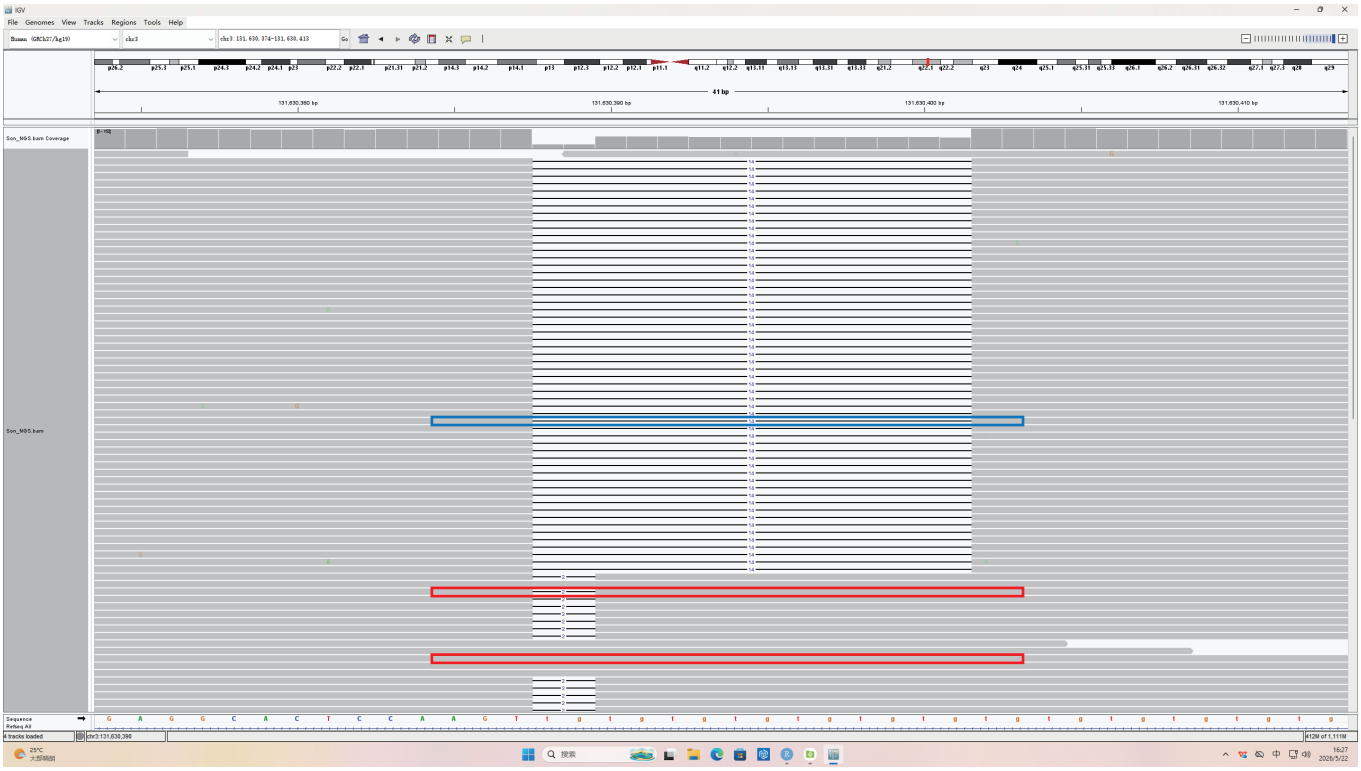
chr3

Case Two:
WT/Hap1/Hap2

Maternal :
WT/Hap1

Paternal:
Hap2





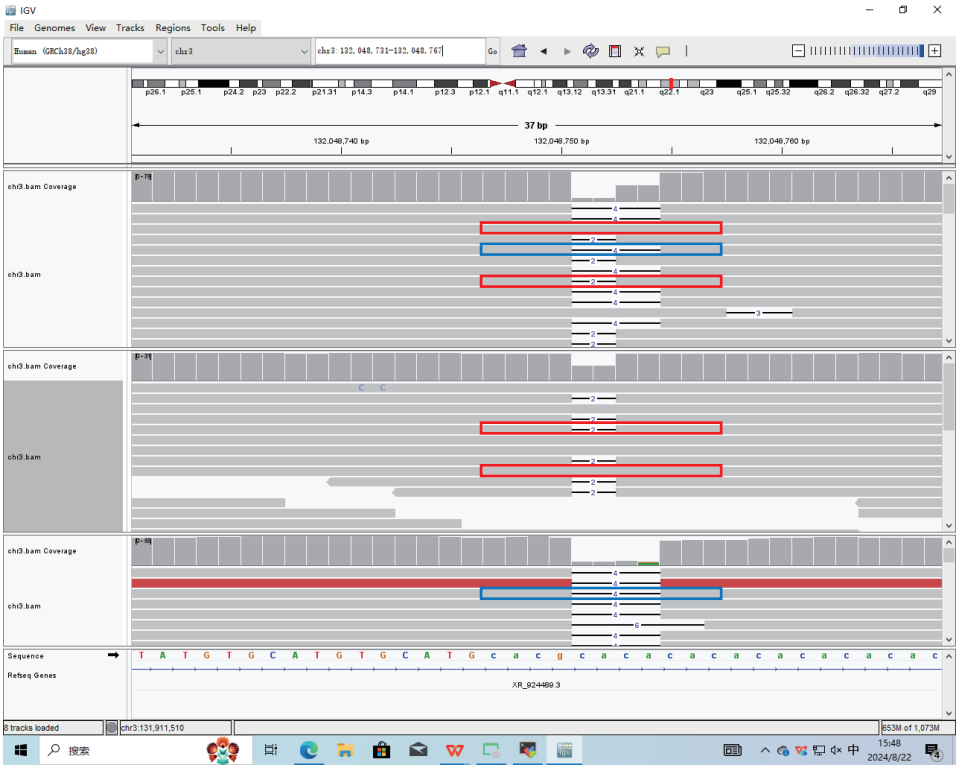
Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2

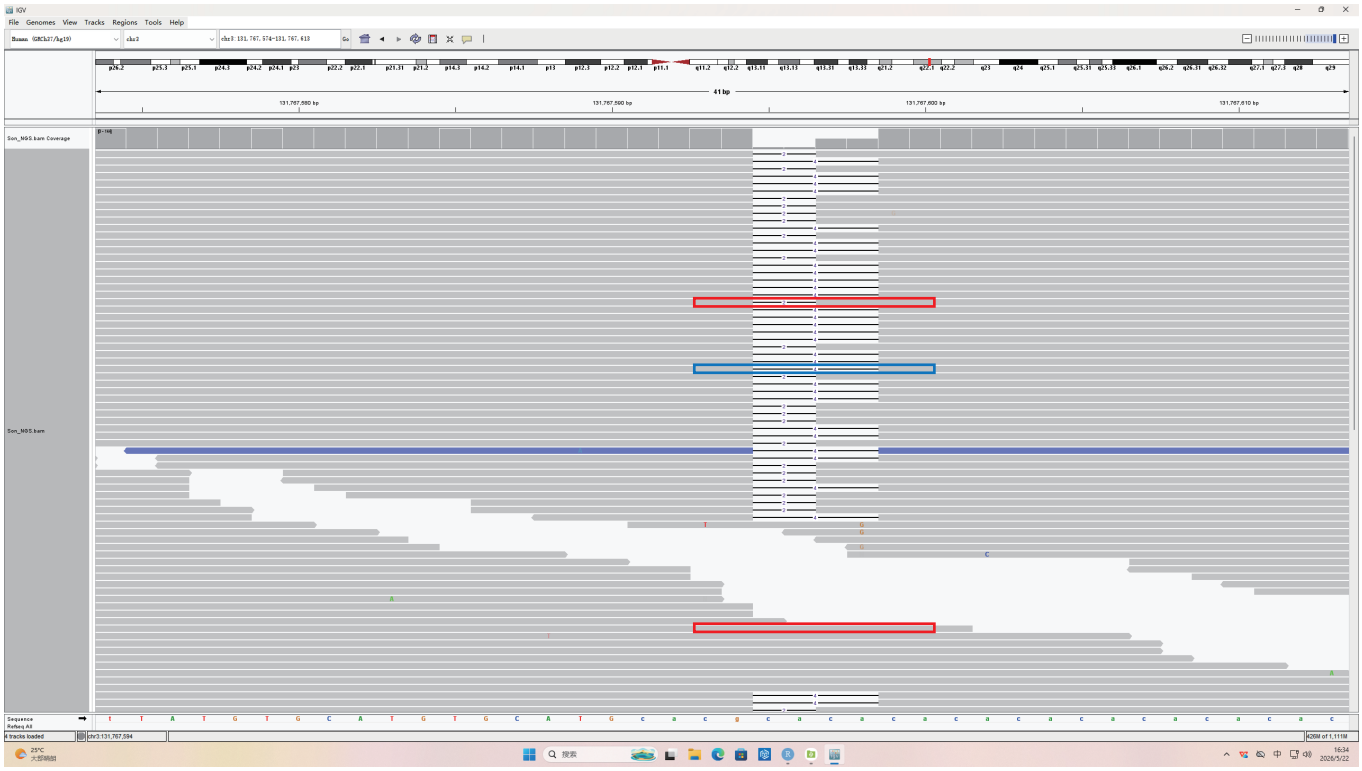
chr3

Case Two:
WT/Hap1/Hap2

Maternal :
WT/Hap1

Paternal:
Hap2





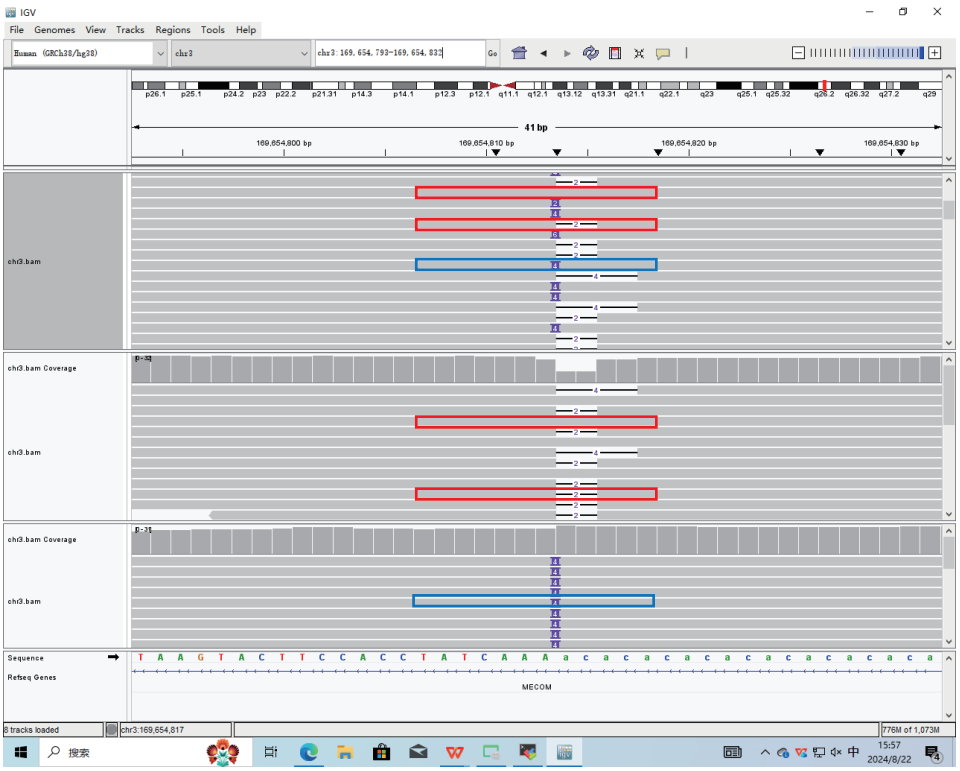
Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2

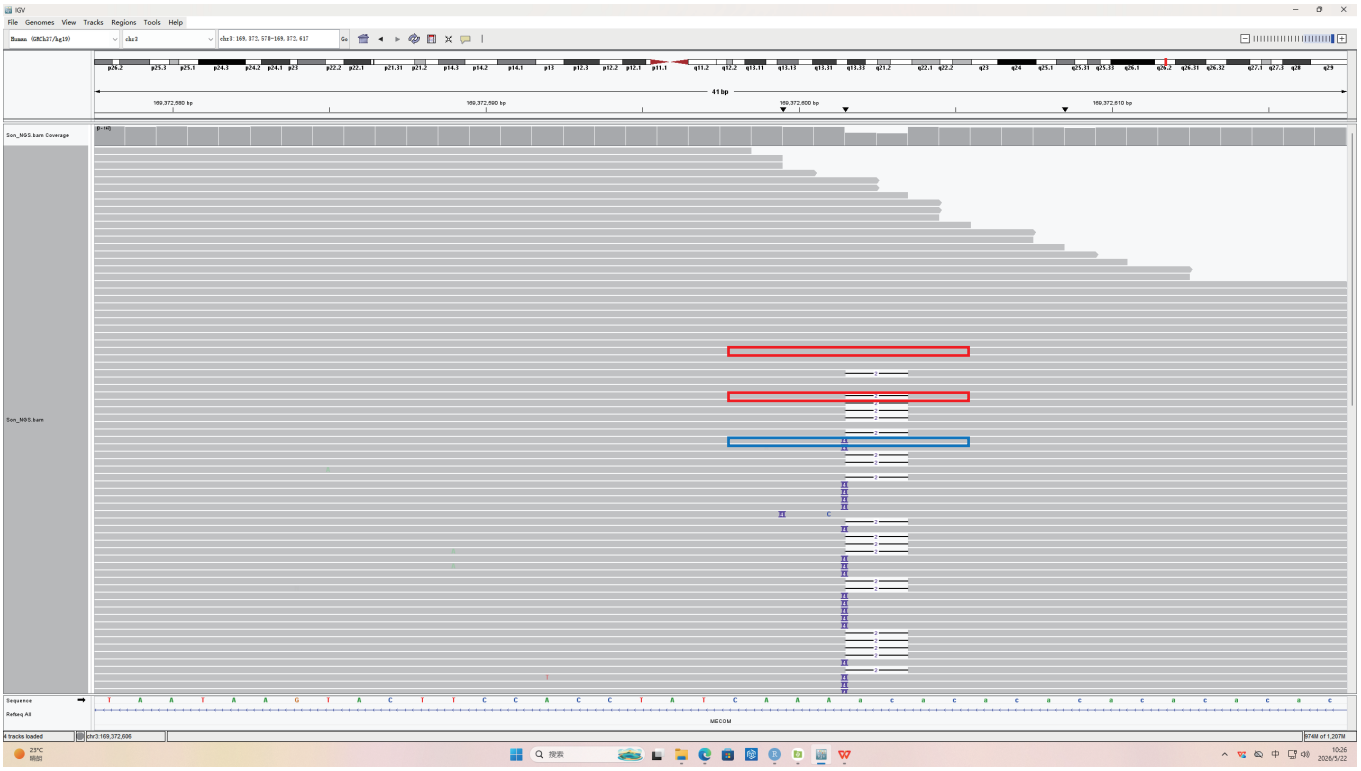
chr3

Case Two:
WT/Hap1/Hap2

Maternal :
WT/Hap1

Paternal:
Hap2





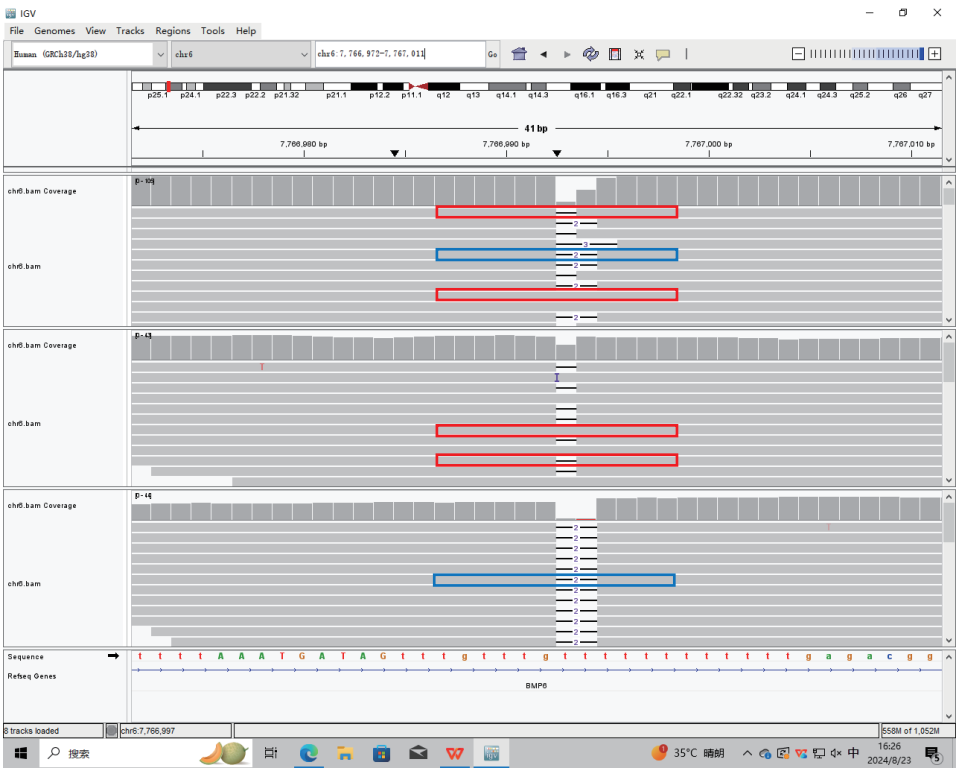
Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2

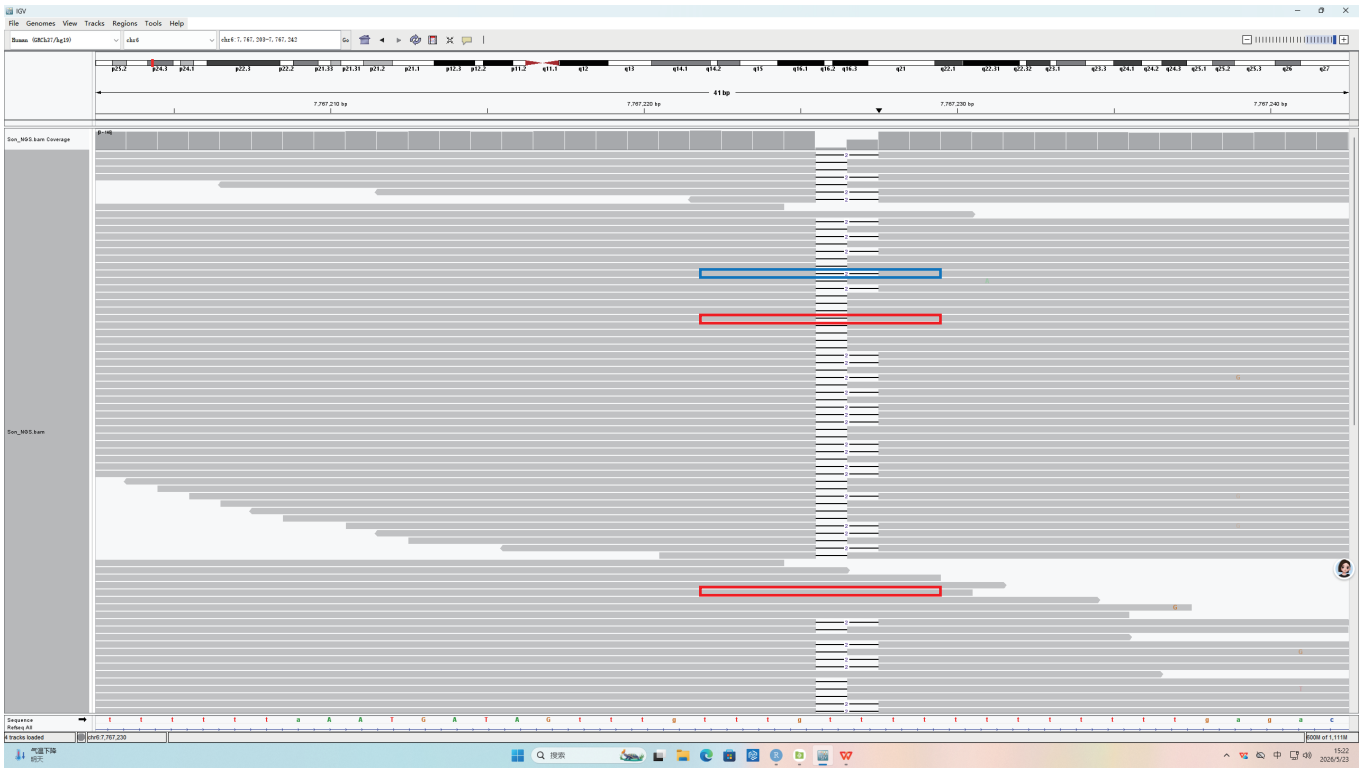
chr6

Case Two:
WT/Hap1/Hap2

Maternal :
WT/Hap1

Paternal:
Hap2





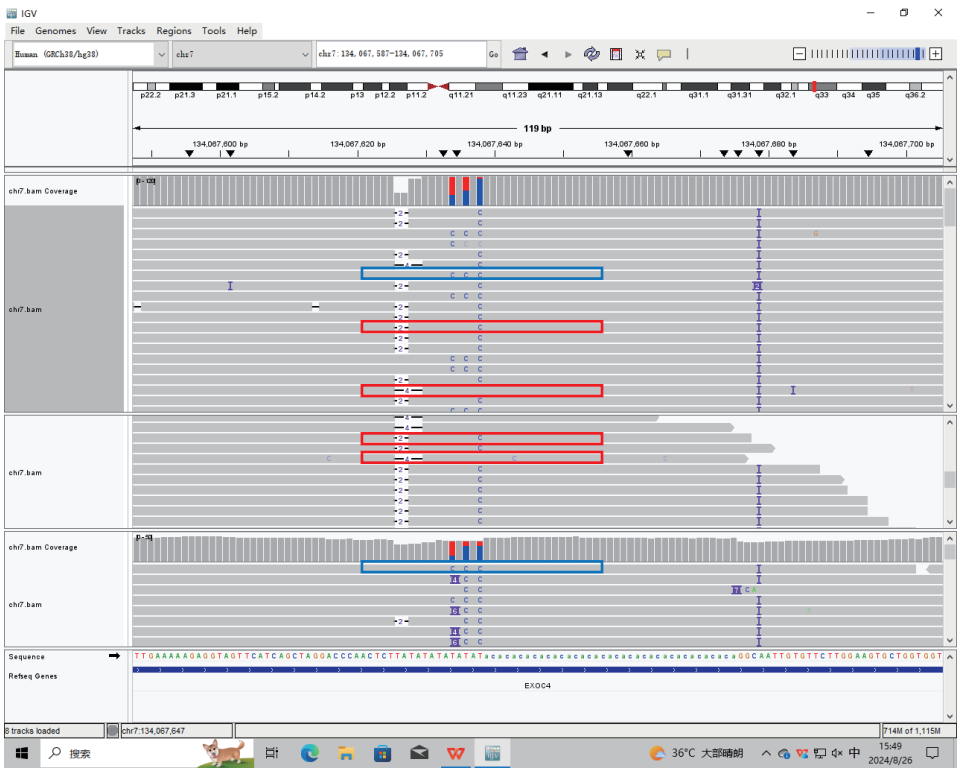
Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2

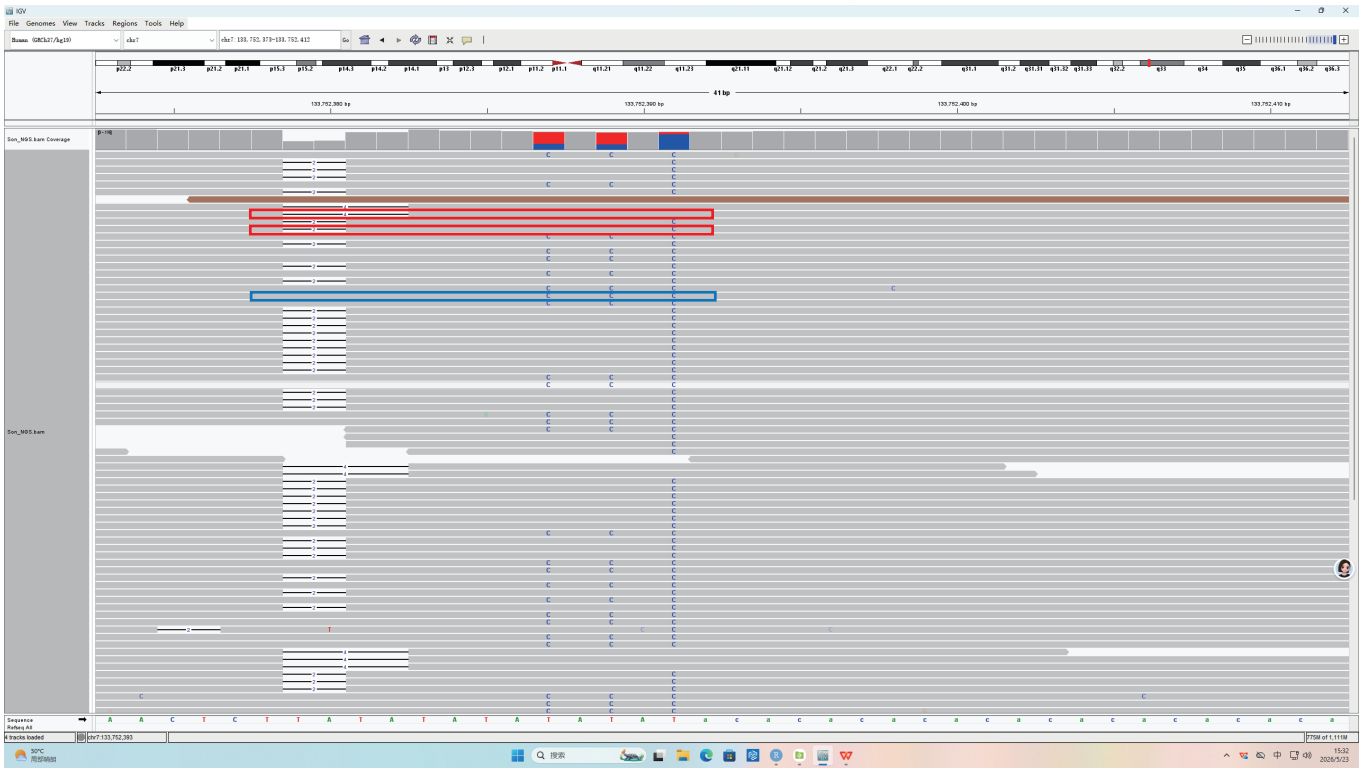
chr7

Case Two:
Hap1/Hap2/Hap3

Maternal :
Hap1/Hap2

Paternal:
Hap3





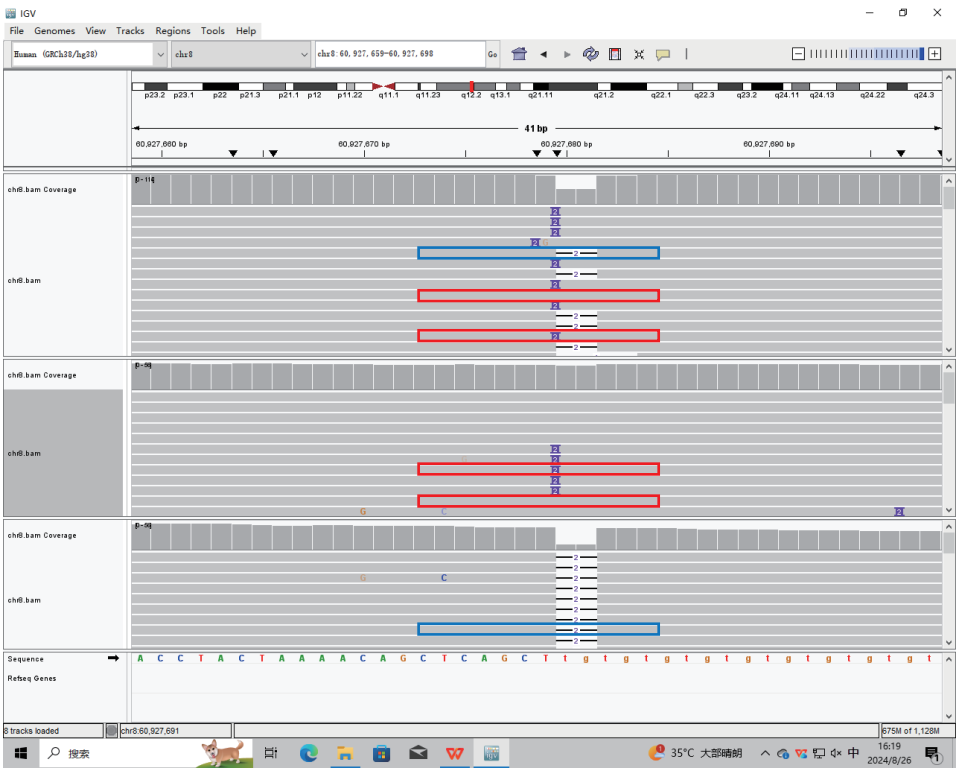
Case Two (PCR-free WGS):
Maternal_Hap1/Maternal_Hap2/Paternal_Hap3

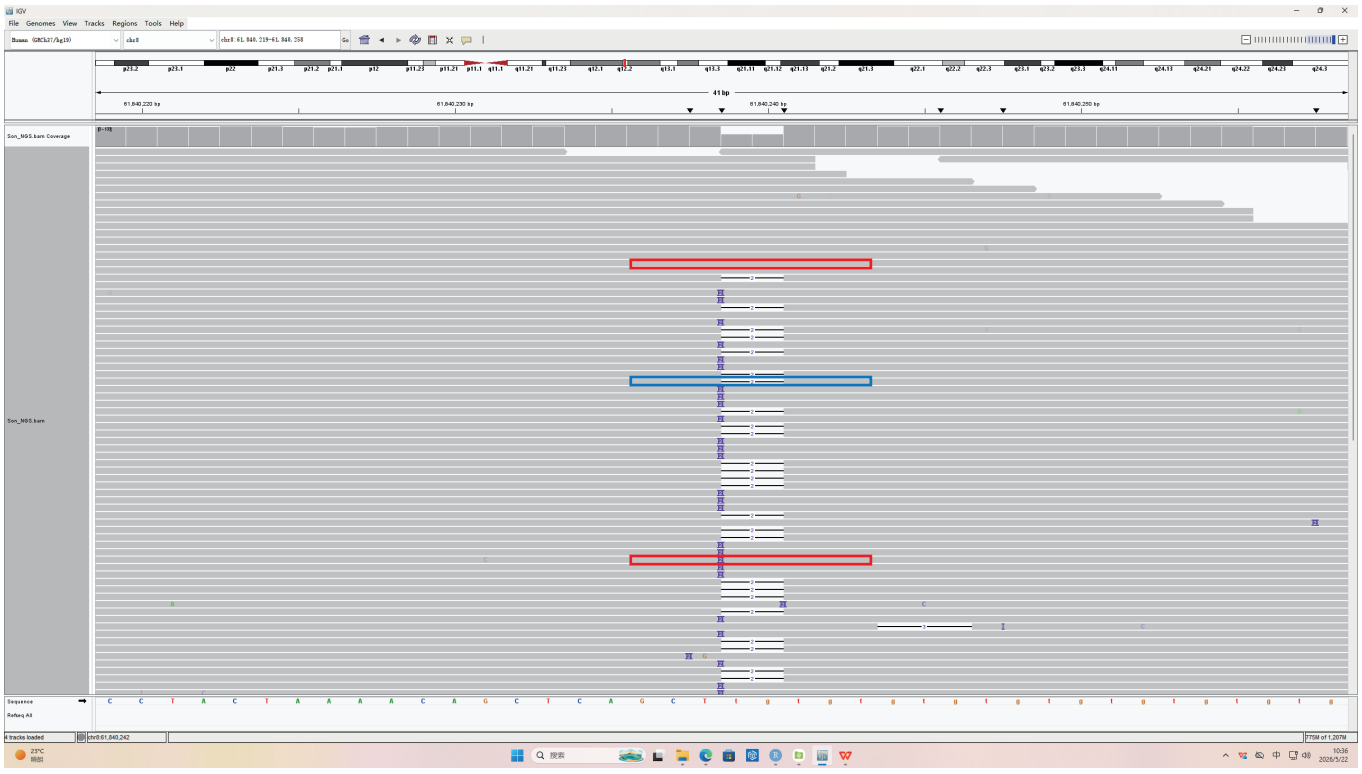
chr8

Case Two:
WT/Hap1/Hap2

Maternal :
WT/Hap1

Paternal:
Hap2





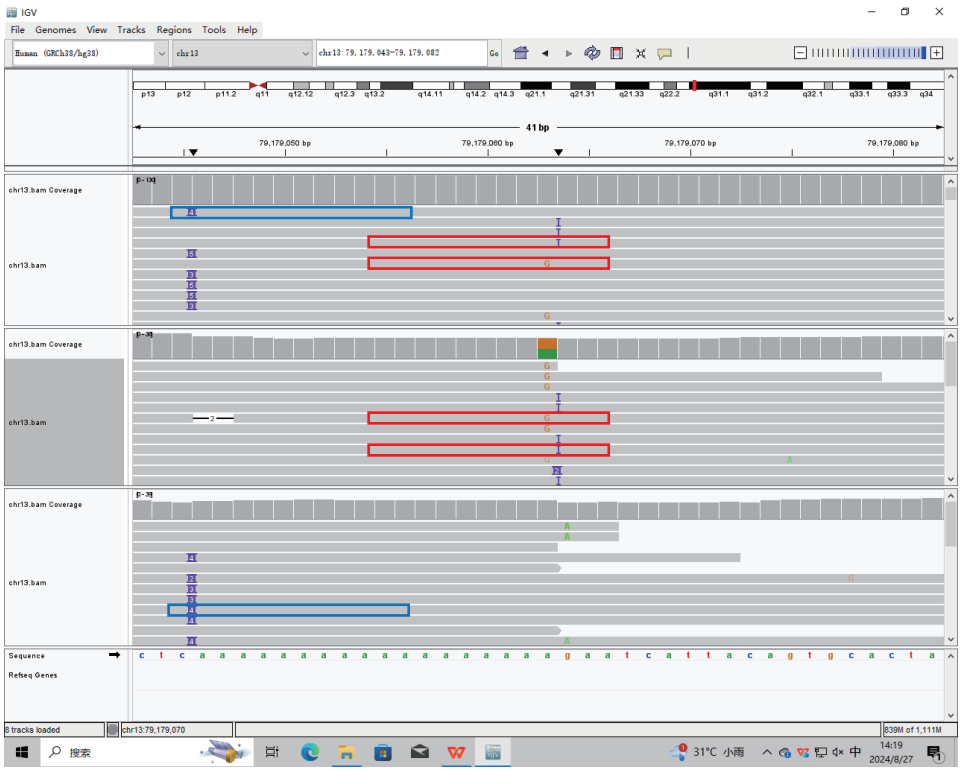
Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2

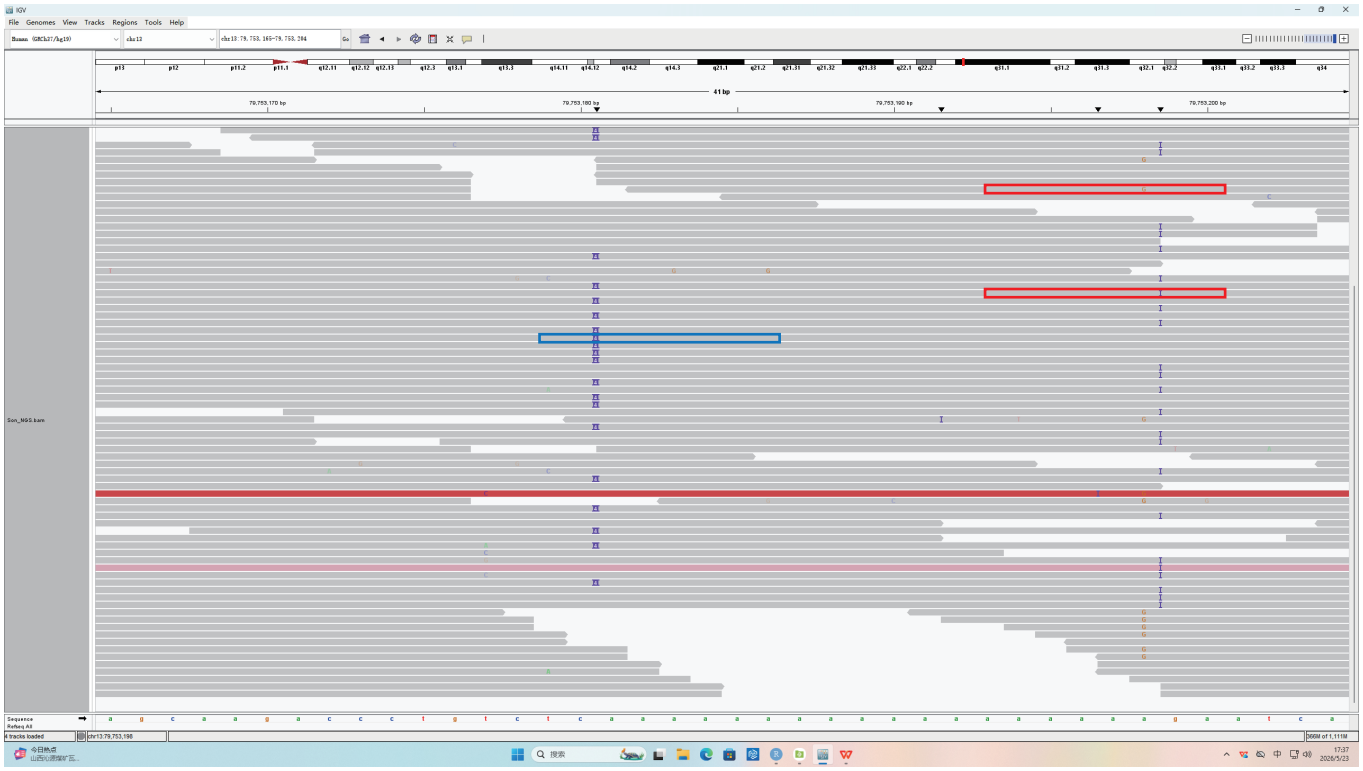
chr13

Case Two:
Hap1/Hap2/Hap3

Maternal :
Hap1/Hap2

Paternal:
Hap3





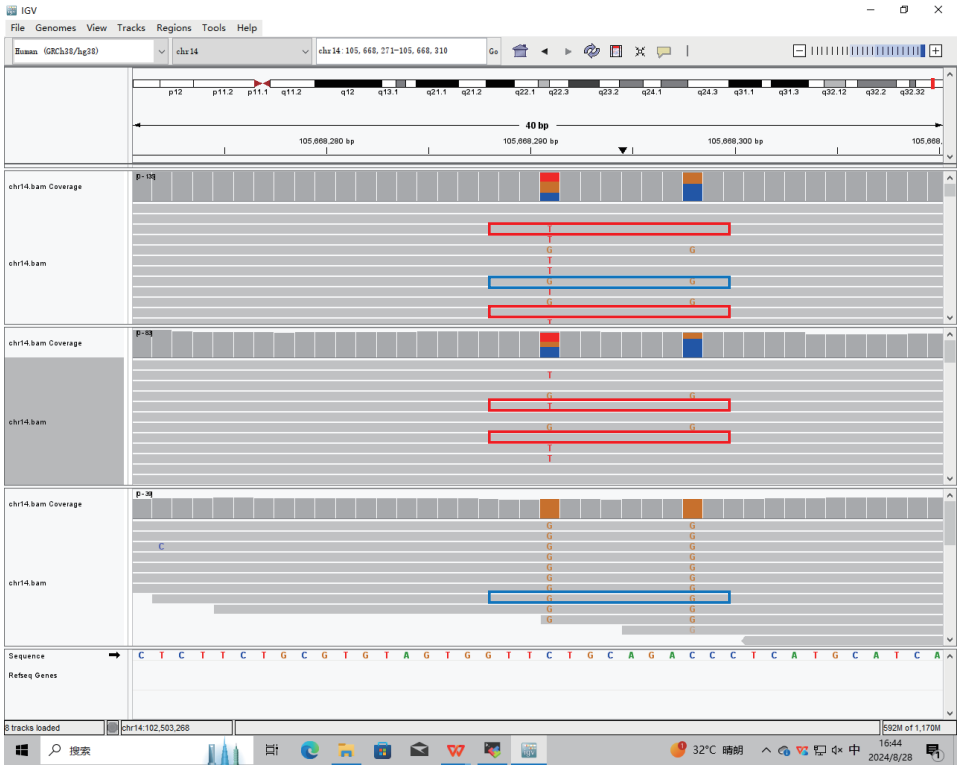
Case Two (PCR-free WGS):
Maternal_Hap1/Maternal_Hap2/Paternal_Hap3

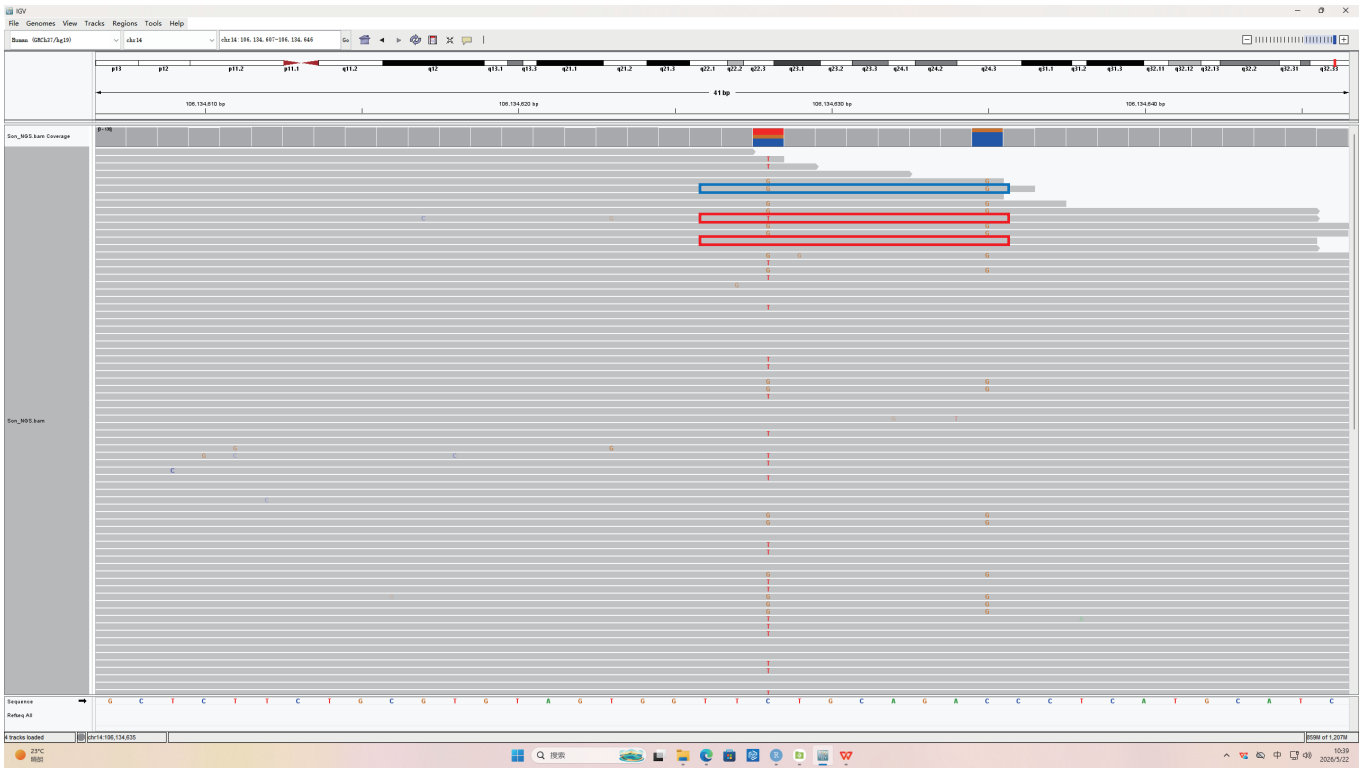
chr14

Case Two:
WT/Hap1/Hap2

Maternal :
WT/Hap1

Paternal:
Hap2





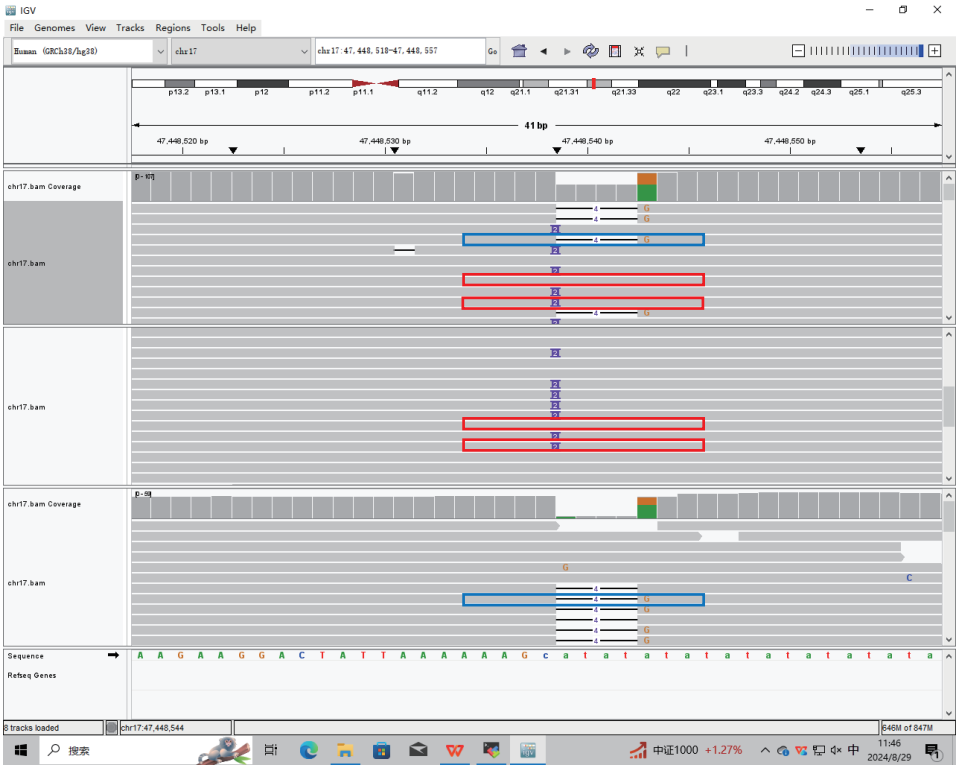
Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2

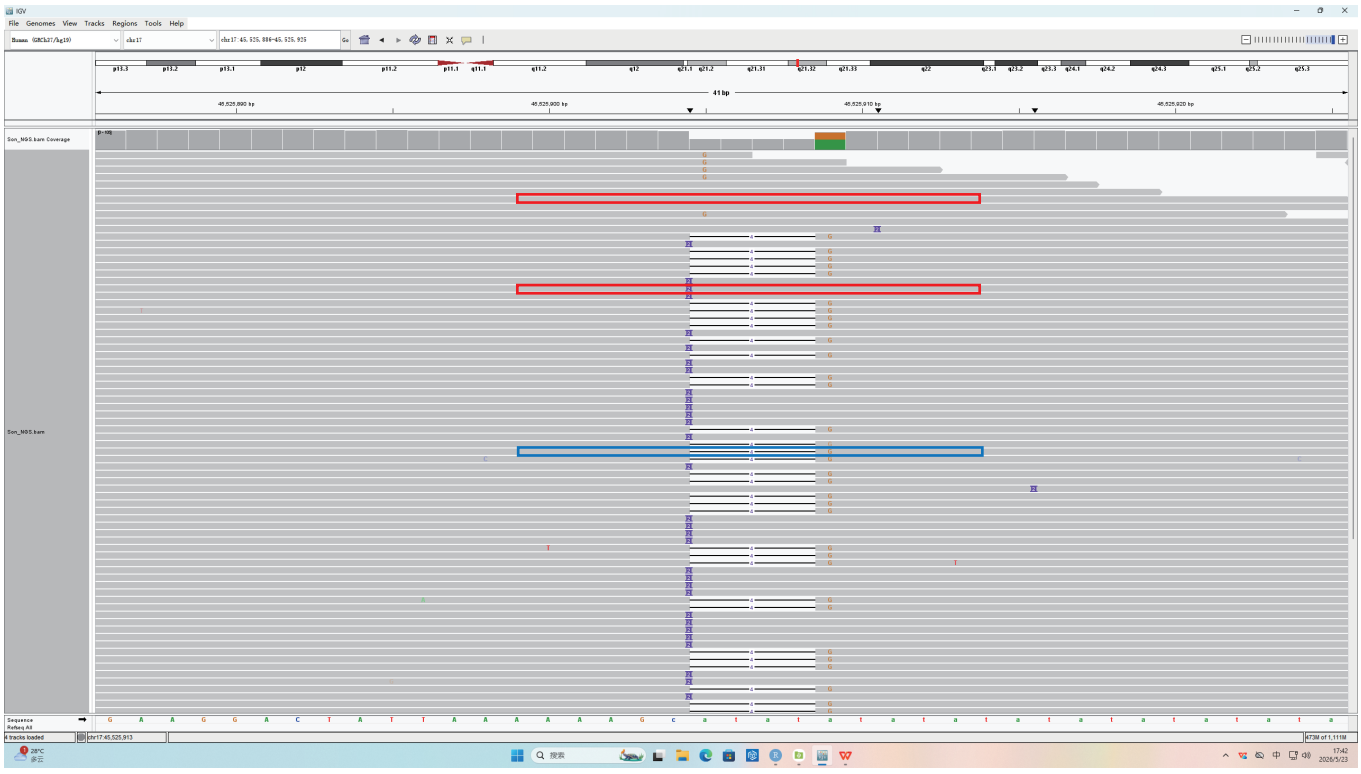
chr17

Case Two:
WT/Hap1/Hap2

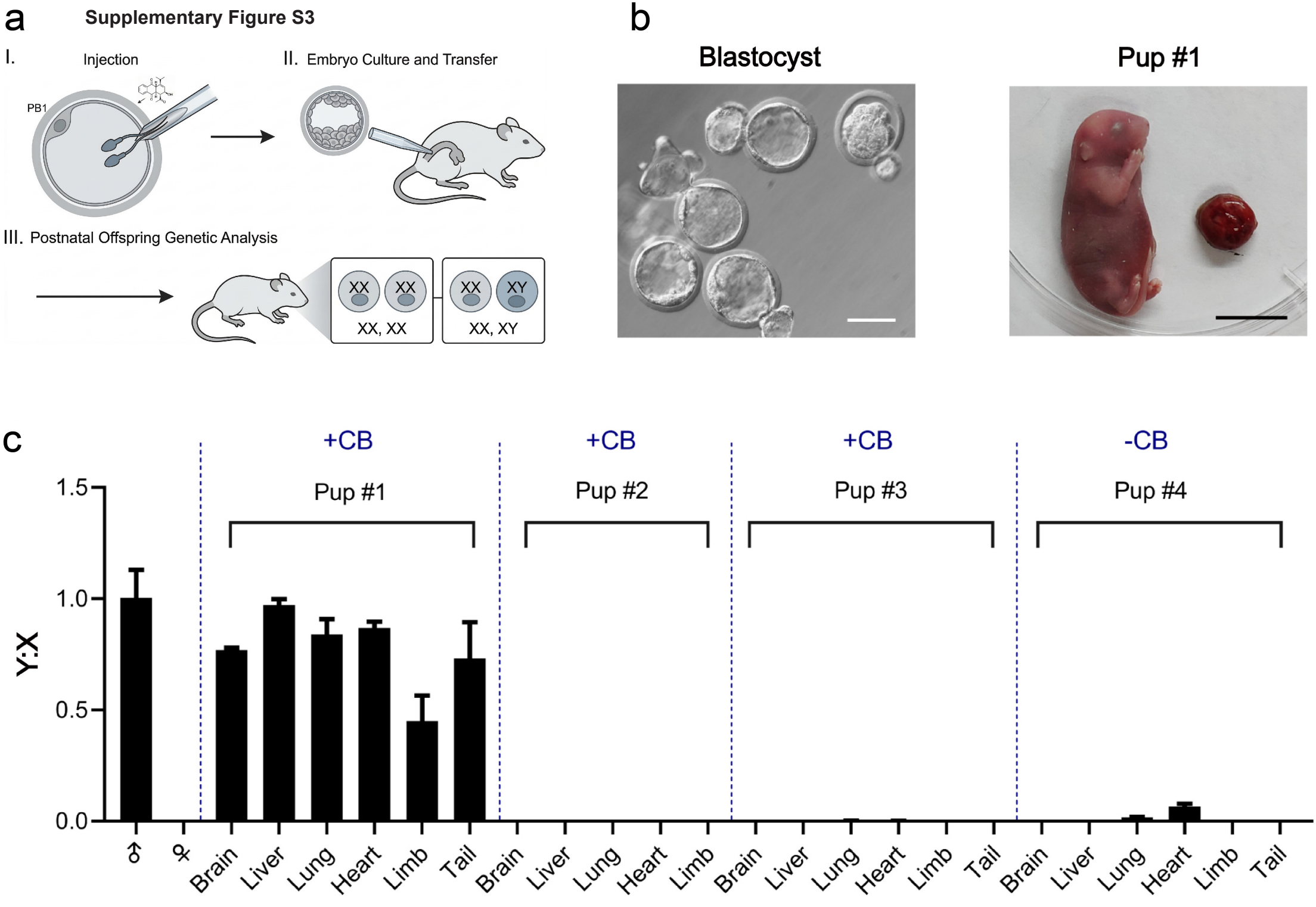
Maternal :
WT/Hap1

Paternal:
Hap2





Case Two (PCR-free WGS):
WT/Maternal_Hap1/Paternal_Hap2


a
Supplementary Figure S3
I. Injection
II. Embryo Culture and Transfer
PB1
III. Postnatal Offspring Genetic Analysis
XX
XX
XX, XX
XX
XY
XX, XY
b
Blastocyst
Pup #1
c
+CB
+CB
+CB
-CB
Pup #1
Pup #2
Pup #3
Pup #4
1.5
1.0
0.5
0.0
Y:X
♂
♀
Brain
Liver
Lung
Heart
Limb
Tail

Supplementary Table S1

Table 1. Results of STR analysis in case one family.

Chromosome	Locus	Marker	Father	Mother	Twin-Boy	Twin- Girl
3	3p21.31	D3S1358	15/16	17	15/17	15/17
1	1q42	D1S1656	13/15	15	13/15	13/15
6	6q15	D6S1043	18/20	11/12	11/12/18/20	11/12/18/20
13	13q31.1	D13S317	10/12	11/12	12	12
15	15q26.2	Penta E	12/18	10/16	12/16	12/16
16	16q24.1	D16S539	11/12	9/10	10/11	10/11
18	18q21.33	D18S51	13/14	15/16	13/15	14/15
2	2q35	D2S1338	23/24	19/23	23	23/24
5	5q33.1	CSF1PO	11	10/12	11/12	11/12
21	21q22.3	Penta D	11/14	9	9/11/14	9/11/14
11	11p15.5	TH01	6/7	7	6/7	6/7
12	12p13.31	vWA	16/18	17/19	16/19	16/19
21	21q21.1	D21S11	30	29/32	29/30	29/30
7	7q21.11	D7S820	11	10/11	10/11	10/11
5	5q23.2	D5S818	11/13	7/12	11/12/13	11/12/13
2	2p25.3	TPOX	11	8	8/11	8/11
8	8q24.13	D8S1179	13/14	13/14	13/14	13/14
12	12p	D12S391	18	19/23	18/19/23	18/19/23
19	19q12	D19S433	13/15	13	13/15	13/15
4	4q28	FGA	19/23	20/21	21/23	21/23
X/Y	Xp22.1-Xp22.3Yp11.2	AMEL	Y	X	X/Y	X/Y

Supplementary Table S2

Supplementary Table S2a: The number of SNPs in Artificial fraternal twin chimeric samples

	Family one	Family two	Family three
Offsets SNPs	24.5%(29990/122461)	41.0%(50782/123777)	32.1%(39278/122265)
Background SNPs	25.0%(31229/124788)	38.2%(46988/122965)	30.6%(37658/123070)

Supplementary Table S2b: The number of SNPs in Artificial non-fraternal twin chimeric samples

	Family one	Family two	Family three
Offsets SNPs	6.5e-5(8/123411)	1.1e-4(14/124668)	5.0e-4(62/123934)
Background SNPs	8.9e-5(11/123273)	1.1e-4(14/123435)	4.6e-4(56/122816)

Table 3. Results of STR analysis in case two family.

Chromosome	Locus	Marker	Father	Mother	Ovotesticular DSD
3	3p21.31	D3S1358	15/17	15/16	15/16/17
1	1q42	D1S1656	16	12/15	12/16
6	6q15	D6S1043	11/18	18/19	11/18
13	13q31.1	D13S317	8/11	9/11	8/9
15	15q26.2	Penta E	11	16/21	11/16
16	16q24.1	D16S539	10/11	9/12	10/11/12
18	18q21.33	D18S51	13/16	15/16	16
2	2q35	D2S1338	19/23	21/23	19/23
5	5q33.1	CSF1PO	10/12	10/12	10/12
21	21q22.3	Penta D	9/12	9/10	9/12
11	11p15.5	TH01	7/9	9	9
12	12p13.31	vWA	18/19	14	14/18/19
21	21q21.1	D21S11	28/29	32/32.2	28/29/32.2
7	7q21.11	D7S820	9/11	8/10	8/11
5	5q23.2	D5S818	7/11	9/11	11
2	2p25.3	TPOX	8/11	8/10	8/11
8	8q24.13	D8S1179	13/14	12/16	12/13/14
12	12p	D12S391	20/22	19/21	20/21/22
19	19q12	D19S433	14.2	15.2	14.2/15.2
4	4q28	FGA	24/25	21/22	21/25
X/Y	Xp22.1-Xp22.3Yp11.2	AMEL	Y	X	X/Y

Glossary of Key Terminology and Step-by-Step Analytical Workflow

Part A: Glossary of Key Terminology

- **Informative sites (STRs/SNPs):** Polymorphic genetic loci where the parents possess distinct, non-overlapping alleles. These sites enable the unambiguous tracing and assignment of the paternal or maternal origins for each allele detected in the offspring, helping to determine the precise number of gametes involved in fertilization.
- **Candidate SNPs:** Specific genomic loci selected for tracing maternal inheritance. In our analysis, these are strictly defined as loci where the paternal genotype is uniformly homozygous (e.g., *ApAp* or *BpBp*) and the maternal genotype is heterozygous (e.g., *AmBm*).
- **Offset SNPs:** A subset of "candidate SNPs" where both distinct maternal alleles (*Am* and *Bm*) are successfully detected in the chimeric offspring. The ratio of offset SNPs to total candidate SNPs uniquely quantifies the sequence divergence between the contributing maternal genomes (e.g., ~50% indicates fertilization of two distinct oocytes, while ~0.1% indicates nearly identical maternal sources, such as a single oocyte and its second polar body [PB2]).
- **Haplotype phasing and origin assignment:** The computational and sequencing process of determining which SNP alleles are physically linked on the same continuous DNA molecule (haplotype). By utilizing PCR-free short-read sequencing of the parents to establish baseline reference haplotypes, and third-generation long-read sequencing of the proband to capture intact DNA fragments (~15 kb), contiguous sequences are definitively assigned to their specific maternal or paternal origins at single-molecule resolution.

Part B: Step-by-Step Analytical Workflow

To systematically distinguish PB2-derived fertilization from parthenogenetic endoreplication or classical tetragametic chimerism, we implemented the following multi-tiered pipeline:

- **Step 1: Preliminary Chimerism Screening via STR Profiling.** We perform multi-tissue Short Tandem Repeat (STR) profiling to confirm biological parentage. The detection of $\geqslant 3$ distinct parental alleles at informative loci confirms chimerism and delineates the number of contributing sperm and oocytes.
- **Step 2: Quantification of Maternal Genomic Divergence.** High-density SNP arrays are used to interrogate "candidate SNPs." By calculating the proportion of "offset SNPs" in the proband, we estimate global maternal sequence divergence to distinguish between classical dizygotic chimerism (~50% divergence) and PB2/parthenogenetic

origins (<0.1% divergence).

- **Step 3: Background Noise Assessment.** Artificially mixed DNA (1:1) from confirmed monozygotic twins is analyzed using the identical SNP array strategy to establish the baseline technical noise, verifying that the minimal ~0.1% divergence is not merely a systemic array artifact.
- **Step 4: Parental Reference Phasing.** PCR-free whole-genome short-read sequencing (50×) is performed on both parents to precisely phase adjacent SNPs and construct definitive maternal and paternal reference haplotypes.
- **Step 5: Single-Molecule Resolution via Long-Read Sequencing.** The proband's DNA is subjected to third-generation PacBio long-read sequencing (120× depth). By aligning these long reads to the parental references, we identify the concurrent presence of one paternal wild-type haplotype and two distinct maternal haplotypes at specific recombined loci, providing definitive single-molecule proof of PB2-participated dispermic fertilization.
- **Step 6: *In Vivo* Biological Validation.** We utilize a mouse model where PB2 extrusion is artificially blocked (via cytochalasin B), coupled with double-sperm injection, to test the possibility of dispermic fertilization.

Supplementary Methods: Detailed Genomic and Bioinformatic Analysis

1. Sample Selection and DNA Extraction

To account for potential inter-tissue heterogeneity of chimerism in the hermaphrodite proband (Case 2), multiple biological samples including peripheral blood, buccal cells, and urine cells were collected. For the high-density single-nucleotide polymorphism (SNP) array genotyping (Illumina, OmniZhongHua-8), genomic DNA extracted from urine cells was utilized. The corresponding Circos plot (Fig. 1A) reflects the SNP array results derived from these urine cells, which effectively captured the chimeric distribution. Conversely, for the third-generation long-read DNA sequencing, genomic DNA extracted from peripheral blood was strictly utilized. Blood was chosen for this assay because it consistently provided the necessary high-molecular-weight DNA integrity required for constructing ~15 kb long-read sequencing libraries. We performed short paired-end, PCR-free whole-genome sequencing for the family of case two at BGI.

2. PacBio Long-Read Sequencing and Quality Metrics

To achieve single-molecule resolution and accurately distinguish highly homologous maternal haplotypes, we utilized the PacBio High-Fidelity (HiFi) sequencing mode based on Circular Consensus Sequencing (CCS) technology at Grandomics. The genomic DNA from the proband's blood was sheared to a target length of ~15 kb. The HiFi/CCS mode generated highly accurate long reads with a per-read consensus accuracy of >99.9% (> Q30), which is fundamentally crucial for distinguishing true biological homologous recombination microheterogeneity from stochastic sequencing errors. The long-read sequencing achieved a massive average depth of 120× across the genome. Concurrently, paired-end, PCR-free whole-genome short-read sequencing (WGS) was performed on the parents' blood samples using the Illumina platform at an average depth of 50× to establish precise reference genotypes.

3. Computational Pipeline for Haplotype Phasing

To exclude alignment artifacts or stochastic errors, we implemented a rigorous bioinformatics pipeline:

Sequencing reads from the DNBSEQ-T7 platform were aligned to the human reference

genome (GRCh38) using the Burrows-Wheeler Aligner[1] (v0.7.17) with default parameter settings. The PCR duplicates were eliminated using GATK. CCS reads from PacBio SeqIIe platform were aligned to the human reference genome (GRCh38) using minimap2[2] with parameter "-a -x map-hifi". Mutation calling for each sample was done using Haplotyper function from sentieon[3] script. For parental samples and kid's sample, the parameter of ploidy was set to 2 and 4, respectively. The generated GVCF files were then used for final family mutation calling using GVCFtyper function from sentieon script. The raw SNP were filtered using the GATK VariantFiltration function with parameter "-filter-name QD_filter -filter QD < 2.0 -filter-name FS_filter -filter FS > 30.0 -filter-name MQ_filter -filter MQ < 30.0 -filter-name SOR_filter -filter SOR > 4.0 -filter-name DP_filter -filter DP < 50.0". SNP sites were removed if the sequencing depth were lower than 10 in any samples. For each SNP site, if the paternal allelic sites were A/B, the maternal allelic sites were C/D, the C/D allelic should be heterozygous and the variant bases of A/B should not overlap with C/D, and the kid's allelic sites should be A/B/C/D then those left SNPs were used for feature analysis and SNP phasing. The SNP phasing for each sample was done using whatshap[4] script.

References:

1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.
2. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572-4574.
3. Freed, D.*, et al.* DNAscope: High accuracy small variant calling using machine learning. 2022.2005.2020.492556 (2022).
4. Martin, M.*, et al.* WhatsHap: fast and accurate read-based phasing. 085050 (2016).