

Supplementary information

Deletions (Cases 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 13, 14, 16, 19, and 25)

In Cases 1, 4, 5, and 13, although deletions were identified using OGM (Tables 2 and S4), the breakpoint determination of deletions using long-read sequencing was unsuccessful. In these cases, segmental duplications of approximately 857–930 kb (Case 1), 161–163 kb (Case 4), 218 kb (Case 5), and 39–42 kb (Case 13) were exactly mapped to deletion breakpoints (Fig. S4). Because N50 in the long-read sequencing data of Cases 1, 4, 5, and 13 were 4551, 6483, 7813, and 23790 bp, respectively, long-read sequencing cannot cover these segmental duplications. Using Blast Like Alignment Tool (BLAT; <https://genome.ucsc.edu/cgi-bin/hgBlat>)¹ to analyze the sequence homology of the pairs of segmental duplications mapped to the deletion breakpoints in Cases 1, 4, 5, and 13, high sequence homology (exceeding 98%) was identified. Therefore, although precise breakpoint determination was not possible, the involvement of non-allelic homologous recombination (NAHR) is strongly indicated as a mechanism for generating these deletions^{2,3}. OGM, which analyzes DNA molecules exceeding 150 kb, is therefore advantageous for mapping relatively large segmental duplications compared with long-read sequencing.

Duplication (Case 28)

We identified an ES-based CN gain with severe global developmental delay, abnormally large globus pallidus, and brain atrophy in Case 28. OGM detected the expected duplication at

the terminal region of chromosome 2q, as identified using ES (Fig. S5). However, we were unable to fully characterize the duplication using OGM (Table S4), even when referring to the telomere-to-telomere (T2T) reference genome (instead of GRCh38).

Long-read sequencing also failed to confirm a duplication at the 2q terminal region. This was likely because of the presence of many repetitive sequences such as DNA transposons, long terminal repeat (LTR) retrotransposons, and short interspersed nuclear elements (SINE), as well as long interspersed nuclear elements (LINE) of non-LTR retrotransposons at around the breakpoint, which hampered the mappability of the duplication breakpoint (Fig. S5). Notably, the OGM-detected duplication did not include any known triplosensitive genes associated with the clinical features of this case (Supplementary Data 1).

Unbalanced reciprocal translocations (Cases 15, 17, and 26)

In Case 17, characterized by Klippel–Feil syndrome, cleft lip and palate, dysphagia, hearing impairment, and esotropia, OGM with AS identified an unbalanced translocation with a 6.7-Mb deletion at 21q22.2–q22.3 and a 9.9-Mb duplication at 14q32.13–q32.33 (Fig. S6a–c and Table S4).

To investigate the origin of this unbalanced rearrangement, breakpoint-spanning PCR was performed using available maternal DNA (primers listed in Table S6). This analysis confirmed the absence of a maternal origin for this rearrangement (Fig. S6d). Furthermore, Sanger

sequencing of the breakpoint PCR products revealed a simple unbalanced translocation (Fig. S6f). No known genes associated with the clinical symptoms observed in Case 17 were disrupted, deleted, or duplicated by this translocation (Tables 2 and Supplementary Data 1).

In Case 26, presenting with seizures, global developmental delay, abnormal facial features, cerebral hemorrhage, laryngomalacia, low-set ears, atresia of the external auditory canal, and impaired ocular abduction, we identified an unbalanced reciprocal translocation using OGM with AS (Fig. S7a–c and Table S4). This translocation included a 4.6-Mb duplication at 5p15.33–p15.32 and a 7.4-Mb deletion at 18q22.3–q23.

Analysis of the long-read sequencing data using dnarrange (<https://github.com/mcfrith/dnarrange> ⁴) failed to detect the breakpoint. Manual inspection of soft-clipped reads at 5p15.32 and 18q22.3 using Integrative Genomics Viewer (IGV) revealed that soft-clipped reads originating from chromosomes 5 and 18 (arrows I–IV of Fig. S7d and e) mapped to chromosome 4 and another region of chromosome 5 that was unrelated to the translocation in the supplementary alignments. At the chromosome 18 breakpoint, several soft-clipped reads were also aligned to chromosomes 2, 9, and 14 (arrows 1–4 of Fig. S7e).

These regions contained repetitive sequences such as LTR retrotransposons and LINE of non-LTR retrotransposons, thereby causing mapping ambiguity despite good coverage, with

mapping quality scores ranging from 21–60⁵. Although the lengths of mismapped reads (11–37 kb) were much longer than those of the repetitive sequences (0.2–6 kb), precise breakpoint detection remained difficult. These results suggest that OGM, with its use of longer DNA molecules (approximately 150 kb), may achieve better breakpoint resolution in repeat-rich regions.

In Case 26, the unbalanced translocation caused CN changes that did not affect any genes linked to the clinical symptoms (Supplementary Data 1), and no genes were disrupted by the translocation itself (Table 2).

Direct duplications (Cases 7, 18, and 24)

OGM identified direct duplications in three cases (Figs. 1g, S8a, and S8d), which had previously been detected by ES. Long-read sequencing data confirmed the breakpoints in these cases and enabled the reconstruction of direct duplications through dot-plots (Figs. 1h, S8b, and S8e).

Case 7, who was clinically suspected of having Coffin–Siris syndrome, exhibited a 2.2-Mb direct duplication at 8p23.1 by OGM with AS (Fig. S8b, Tables 2 and S4). This CN gain, previously identified using ES, encompasses *SOX7* (pTriplo score = 0.421⁶, TS score = not recorded⁷) and *GATA4* (pTriplo score = 0.916⁶, TS score = 2⁷) and is consistent with 8p21.3 duplication syndrome (Supplementary Data 1), which is associated with developmental delay, dysmorphism,

hypotonia, visual impairment, and congenital heart disease ⁸ (Table 1). These findings align with the clinical presentation of the patient, and no additional aberrant genes relevant to the phenotype were detected using OGM. Although *MSRA* was disrupted by the direct duplication breakpoint (Table 2), it has not been associated with any of the clinical features observed in Case 7.

In Case 18, OGM with AS identified a 1.5-Mb direct duplication at 2q24.3 (Fig. 1h, Tables 2 and S4) that encompassed key sodium channel genes such as *SCN1A*, *SCN2A*, *SCN3A*, *SCN7A*, and *SCN9A*, all of which had been previously detected using ES (Tables 1 and Supplementary Data 1). The clustering of these sodium channel genes at 2q24.3 is strongly associated with early infantile epileptic spasms ⁹, in line with the clinical phenotype observed in this case. OGM did not reveal any additional aberrant genes that would further explain the patient's clinical symptoms. *SCN7A* was observed to be disrupted by the direct duplication (Table 2).

In Case 24, OGM detected a 606-kb direct duplication at 19q13.33 (Fig. S8d, Tables 2 and S4). The aberrant gene associated with the patient phenotype was unable to be identified by ES in this case, and no additional aberrant genes were identified using OGM. *TSKS* and *TRPM4* were disrupted by a direct duplication (Table 2) but neither was associated with the observed clinical features.

Intrachromosomal insertion events with inverted duplications (Cases 20–22)

Using OGM with AS, we identified novel intrachromosomal insertion events and inverted duplications in three cases (Cases 20–22) in whom CN gains or combined CN losses and CN gains had initially been detected by ES. OGM revealed a range of complex structural rearrangements in these cases.

Case 20 involved an inverted duplication, a deletion, and an intrachromosomal fusion (Fig. S10, Tables 2 and S4). In Case 21, an inverted duplication and two intrachromosomal fusions were identified (Fig. S11, Tables 2 and S4). Case 22 exhibited an inverted duplication and an intrachromosomal fusion (Fig. 3, Tables 2 and S4).

Regarding these chromosomal rearrangements, Case 20 included an 8.4-Mb duplication at Xq21.33–q22.3, whereas Case 21 harbored a 439-kb duplication at Xq22.2, as detected by OGM. Notably, the intrachromosomal fusions in these cases were positioned adjacent to their respective duplications, a feature that was missed by ES because of the copy-neutral nature of these SVs (Figs. S10a and S11a). Importantly, Case 20 had been previously diagnosed with *PCDH19*-related epilepsy syndrome, and *PCDH19* is located within the CN gain region detected by ES^{10,11} (Tables 1 and Supplementary Data 1). In Case 21, with spinal and bulbar muscular atrophy, *PLP1* (which causes Pelizaeus–Merzbacher disease) was identified within the Xq22.2 CN gain, consistent with earlier reports of complex *PLP1* rearrangements, including triplications and inverted duplications^{12,13} (Tables 1 and Supplementary Data 1). However, in both cases, the

newly resolved SVs identified by OGM and long-read sequencing did not reveal any additional pathogenic variants that might explain the patients' clinical features (Supplementary Data 1). Although *NRK* and *DIAPH2* in Case 20 and *ILIRAPL1* in Case 21 were disrupted by SVs (Table 2), none of the genes were associated with the observed clinical features of Cases 20 and 21.

References

- 1 Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome research* **12**, 656-664, doi:10.1101/gr.229202 (2002).
- 2 Burssed, B., Zamariolli, M., Bellucco, F. T. & Melaragno, M. I. Mechanisms of structural chromosomal rearrangement formation. *Molecular cytogenetics* **15**, 23, doi:10.1186/s13039-022-00600-6 (2022).
- 3 Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends in genetics : TIG* **18**, 74-82, doi:10.1016/s0168-9525(02)02592-1 (2002).
- 4 Frith, M. C. & Mitsuhashi, S. Finding Rearrangements in Nanopore DNA Reads with LAST and dnarrange. *Methods in molecular biology (Clifton, N.J.)* **2632**, 161-175, doi:10.1007/978-1-0716-2996-3_12 (2023).
- 5 Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research* **18**, 1851-1858, doi:10.1101/gr.078212.108 (2008).
- 6 Collins, R. L. *et al.* A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**, 3041-3055.e3025, doi:10.1016/j.cell.2022.06.036 (2022).
- 7 Riggs, E. R. *et al.* Towards an evidence-based process for the clinical interpretation of copy number variation. *Clinical genetics* **81**, 403-412, doi:10.1111/j.1399-0004.2011.01818.x (2012).
- 8 Barber, J. C. *et al.* Inside the 8p23.1 duplication syndrome; eight microduplications of likely or uncertain clinical significance. *American journal of medical genetics. Part A* **167a**, 2052-2064, doi:10.1002/ajmg.a.37120 (2015).
- 9 Yoshitomi, S. *et al.* Three patients manifesting early infantile epileptic spasms associated with 2q24.3 microduplications. *Brain & development* **37**, 874-879, doi:10.1016/j.braindev.2015.03.001 (2015).
- 10 Uchiyama, Y. *et al.* Efficient detection of copy-number variations using exome data: Batch- and sex-based analyses. *Human mutation* **42**, 50-65, doi:10.1002/humu.24129 (2021).
- 11 Kurian, M. *et al.* Focal cortical malformations in children with early infantile epilepsy and PCDH19 mutations: case report. *Developmental medicine and child neurology* **60**, 100-105, doi:10.1111/dmcn.13595 (2018).
- 12 Beck, C. R. *et al.* Complex genomic rearrangements at the PLP1 locus include triplication and quadruplication. *PLoS genetics* **11**, e1005050, doi:10.1371/journal.pgen.1005050 (2015).

- 13 Zhang, L. *et al.* Efficient CNV breakpoint analysis reveals unexpected structural complexity and correlation of dosage-sensitive genes with clinical severity in genomic disorders. *Human molecular genetics* **26**, 1927-1941, doi:10.1093/hmg/ddx102 (2017).

Fig. S1

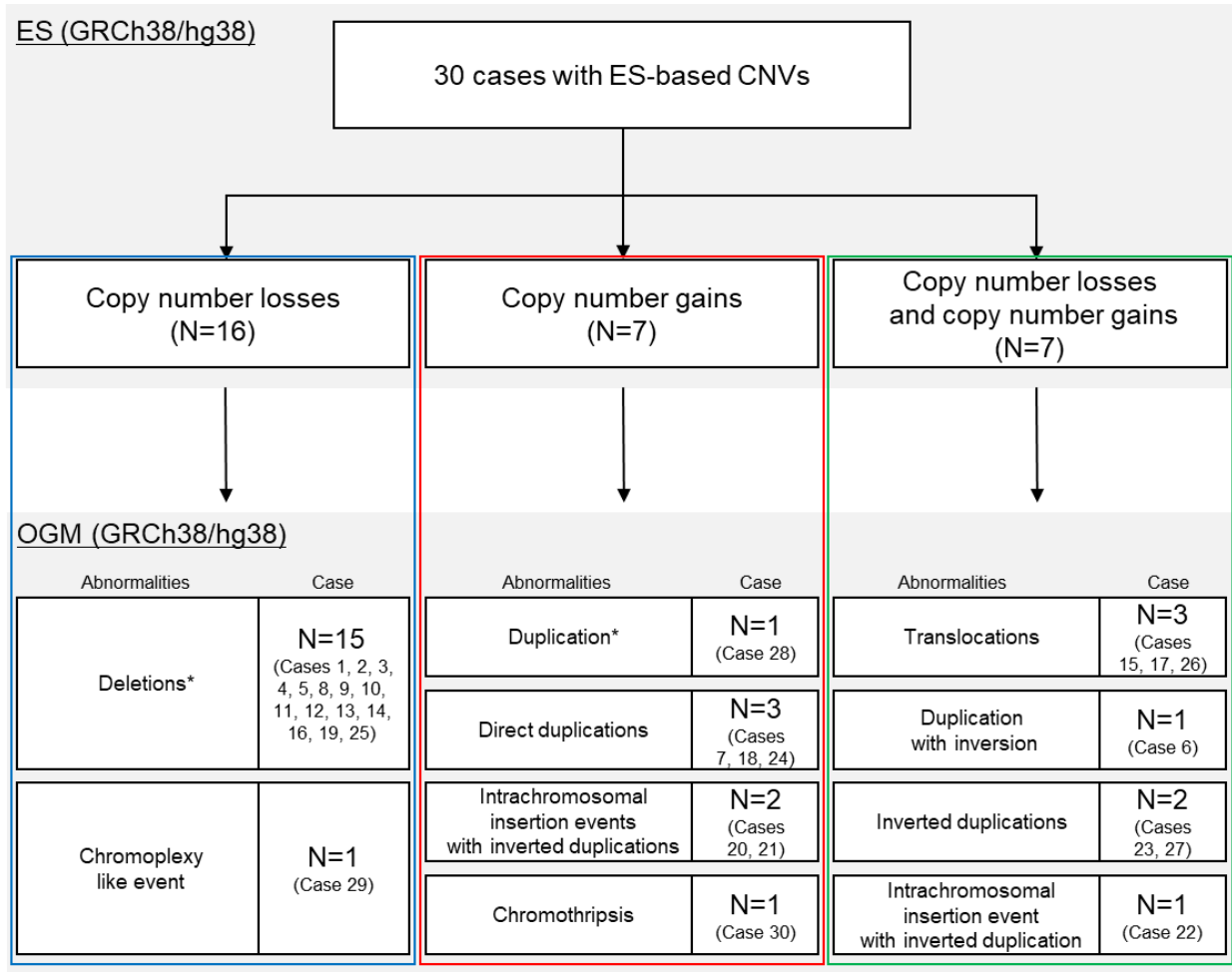
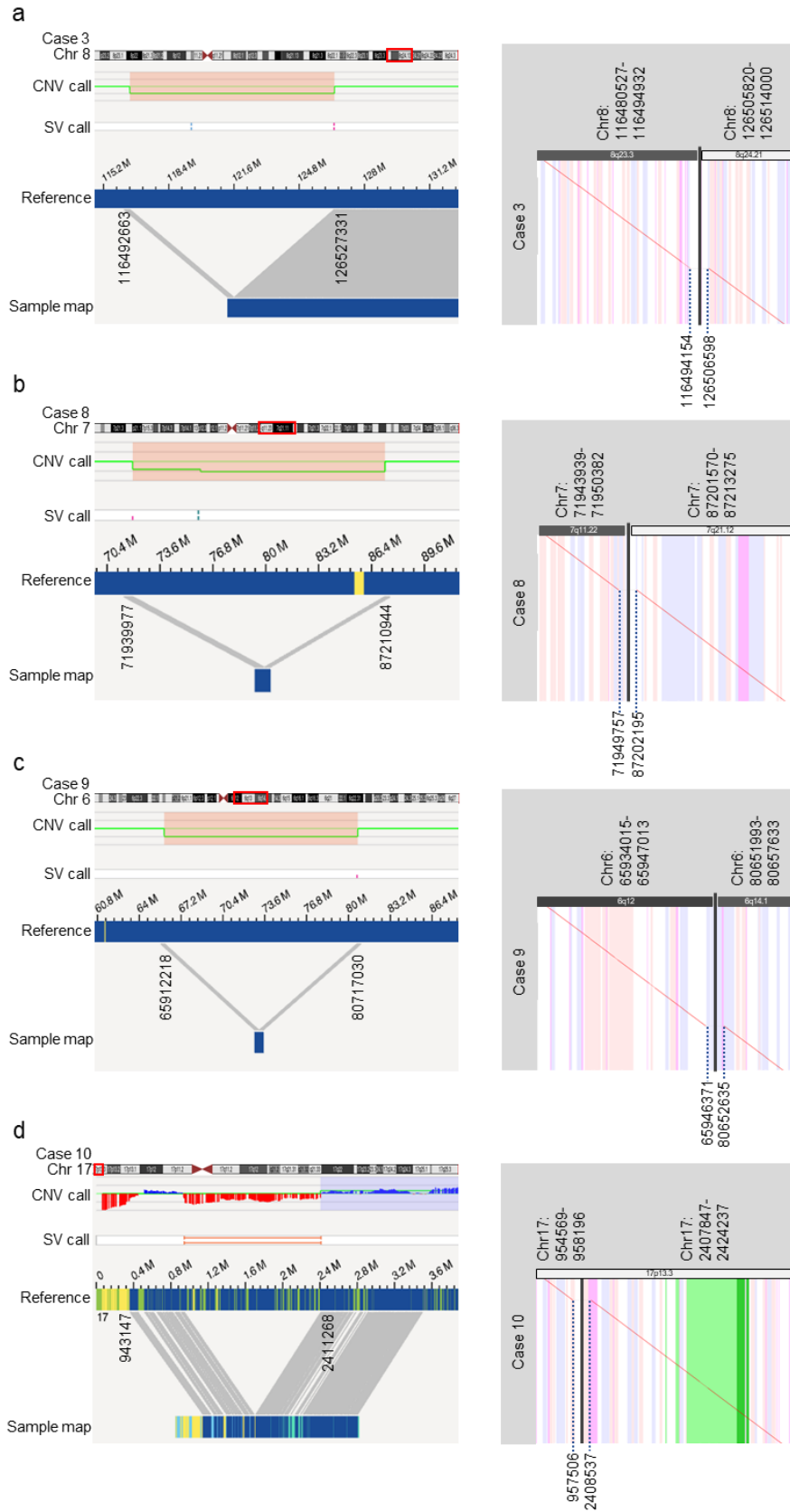
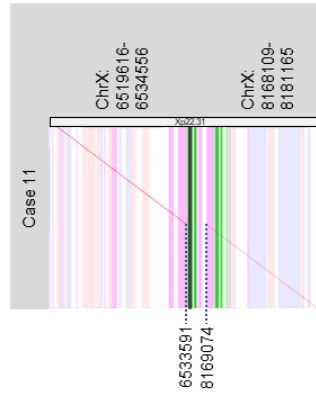
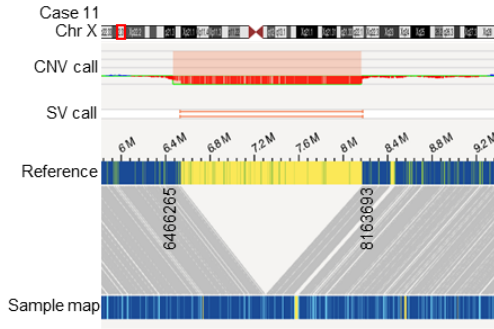


Fig. S1. Overview of the 30 cases verified by OGM with adaptive sampling. *Indicates expected structural variants that were previously detected using exome sequencing. CNV, CN variant; ES, exome screening; OGM, optical genome mapping.

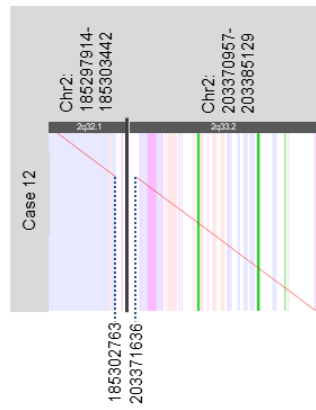
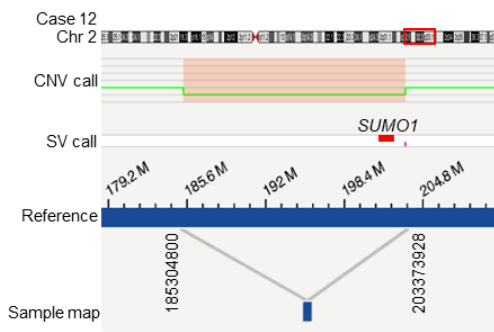
Fig. S2



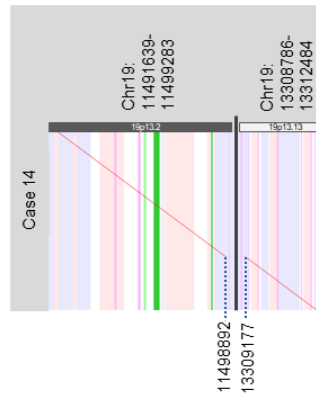
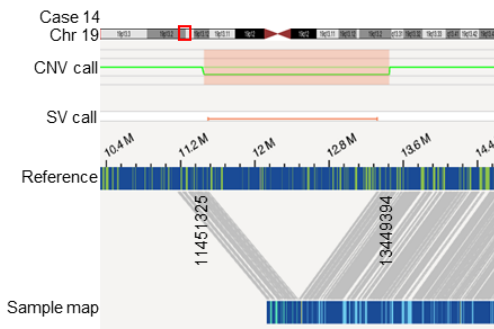
e



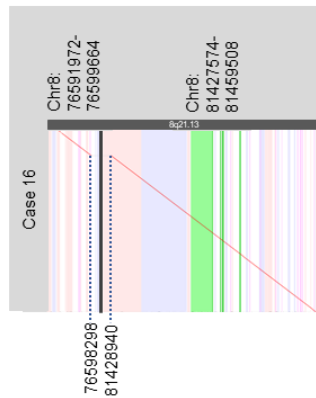
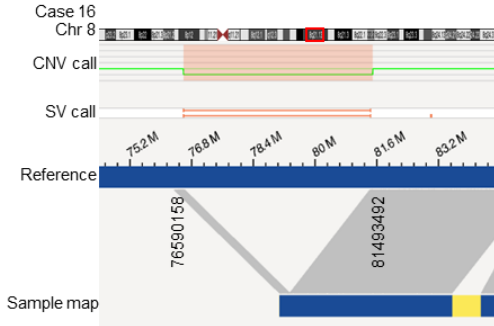
f



g



h



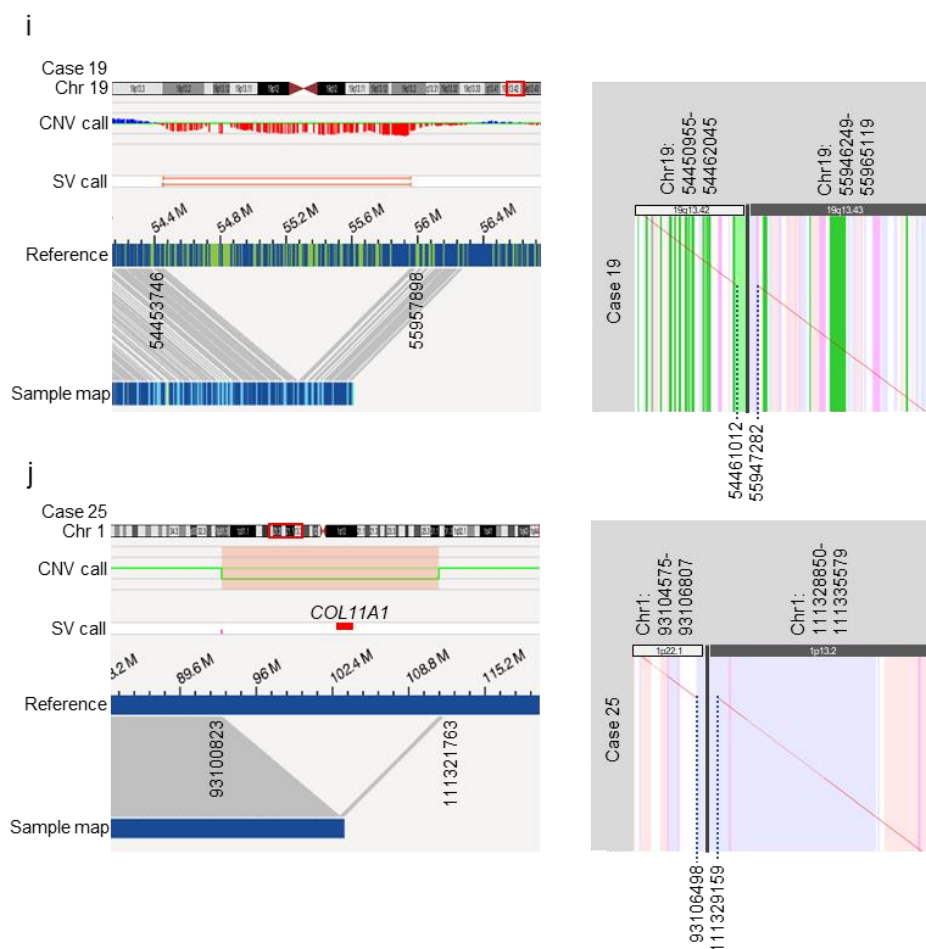


Fig. S2. Results of OGM and long-read sequencing in Cases 3 (a), 8 (b), 9 (c), 10 (d), 11 (e), 12 (f), 14 (g), 16 (h), 19 (i), and 25 (j). Left diagrams are OGM results in the genome browser, and large deletion is visualized as red bars in the CNV or SV call tracks. Right diagrams are dot-plots of the consensus sequence from long-read sequencing data. In the dot-plots, red lines show the forward orientation to the reference sequences, and vertical pale stripes show annotations in the reference genome with different colors: transposable elements (pale blue, reverse-strand; pale pink, forward-strand), tandem repeats (purple), protein-coding sequences (dark green), and exons (green). The numbers below the dot-plot are breakpoint positions. (f) *SUMO1* was located in the deletion. (j) *COL11A1* was located in the deletion. ISCN nomenclature in all cases is described in Table S4. The genome browsers were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Dot-plots were created by lamassemble 1.4.2 and dnarrange for targeted long-read sequencing. Chr, chromosome; CNV, CN variant; ISCN, International System for Human Cytogenomic Nomenclature; OGM, optical genome mapping; SV, structural variant.

Fig. S3

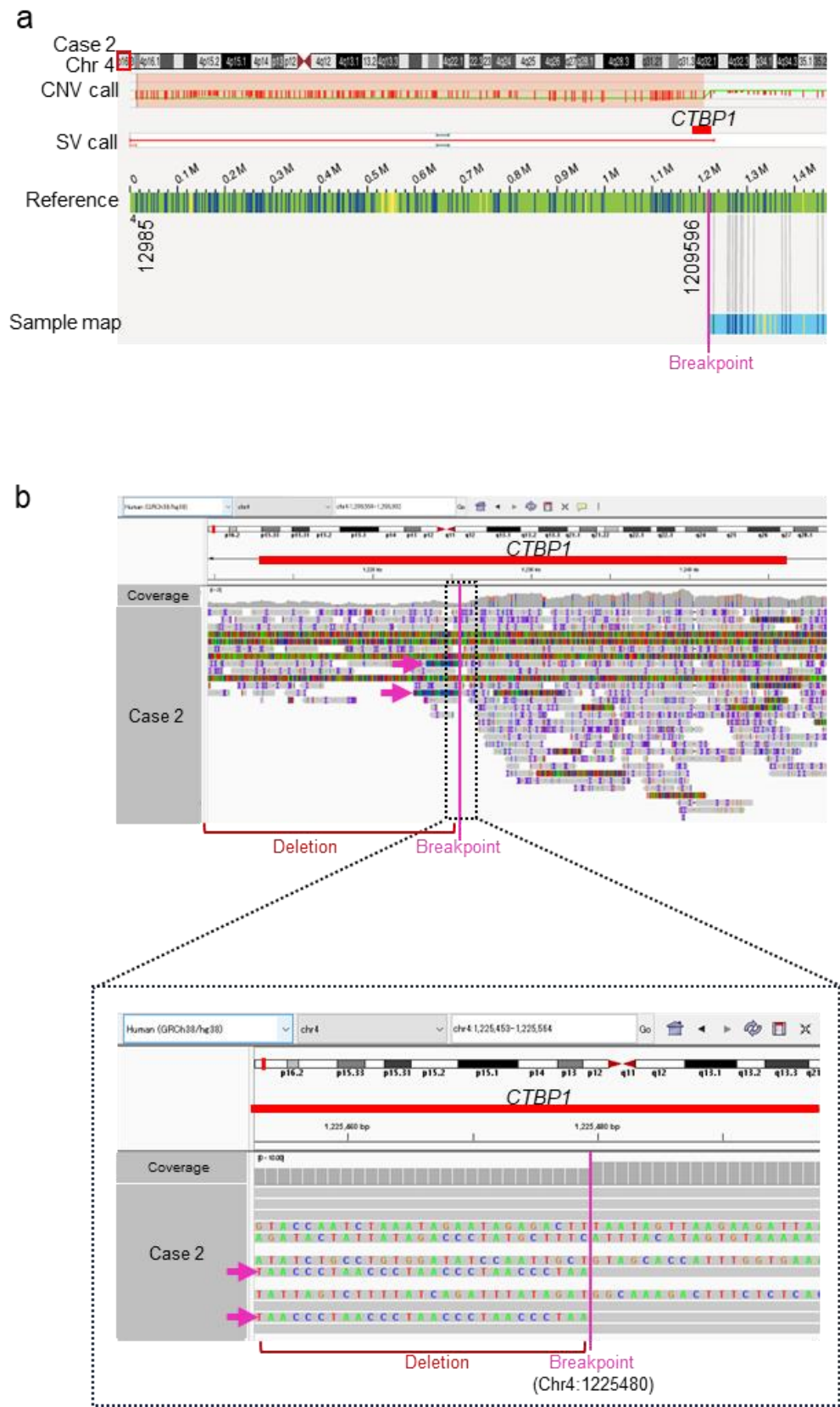


Fig. S3. Results of OGM and long-read sequencing in Case 2. (a) OGM results in the genome browser. A large deletion in Chr 4 is visualized as a red bar in the CNV and SV call tracks. The genome browser was visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. (b) Upper diagram is IGV of long-read sequencing data in Chr 4. The breakpoint of a deletion downstream is shown as a pink line. *CTBPI* is located across the breakpoint. Two soft-clipped reads from a deletion upstream are shown by pink arrows. A deletion in Chr 4 is illustrated as lower coverage upstream of the breakpoint. Lower diagram is an enlarged view of the breakpoint at Chr 4 in IGV. Two soft-clipped reads (pink arrows) show repetitive telomeric sequences (CCCTAA), indicating that this deletion is located at the terminal of the short arm of Chr 4. ISCN nomenclature is described in Table S4. Chr, chromosome; CNV, CN variant; IGV, Integrative Genomics Viewer; ISCN, International System for Human Cytogenomic Nomenclature; OGM, optical genome mapping; SV, structural variant.

Fig. S4

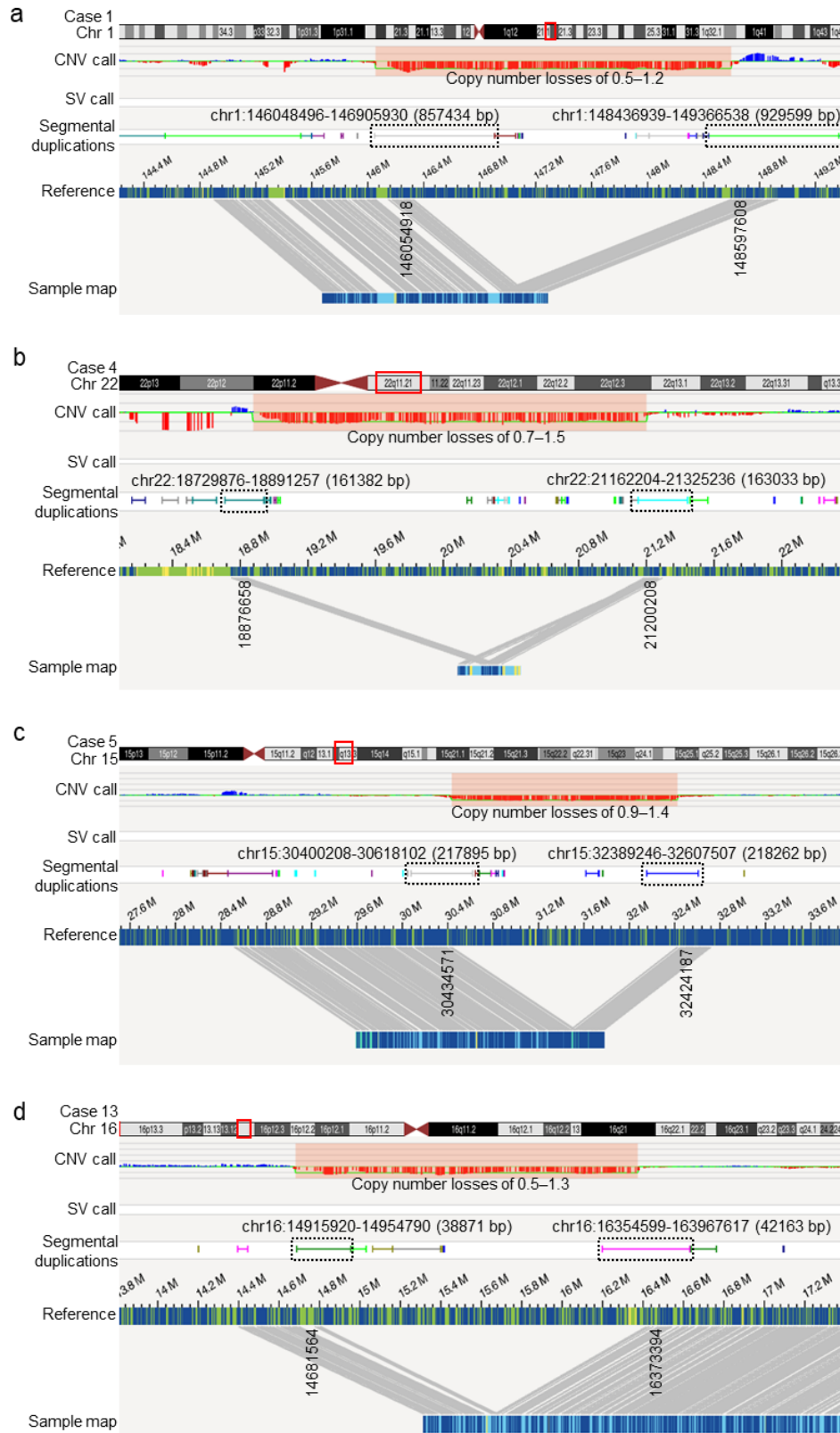


Fig. S4. Results of OGM in Cases 1 (a), 4 (b), 5 (c), and 13 (d). (a–d) OGM results in the genome browser. Large deletions are visualized as red bars in the CNV call track. Both ends of each deletion overlap with segmental duplications shown in the segmental duplications call track. The regions and sizes of segmental duplications are shown in the segmental duplications call track, and segmental duplications in each case are directly oriented within the same chromosomes. ISCN nomenclature in the four cases is described in Table S4. The genome browsers were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Chr, chromosome; CNV, CN variant; ISCN, International System for Human Cytogenomic Nomenclature; OGM, optical genome mapping; SV, structural variant.

Fig. S5

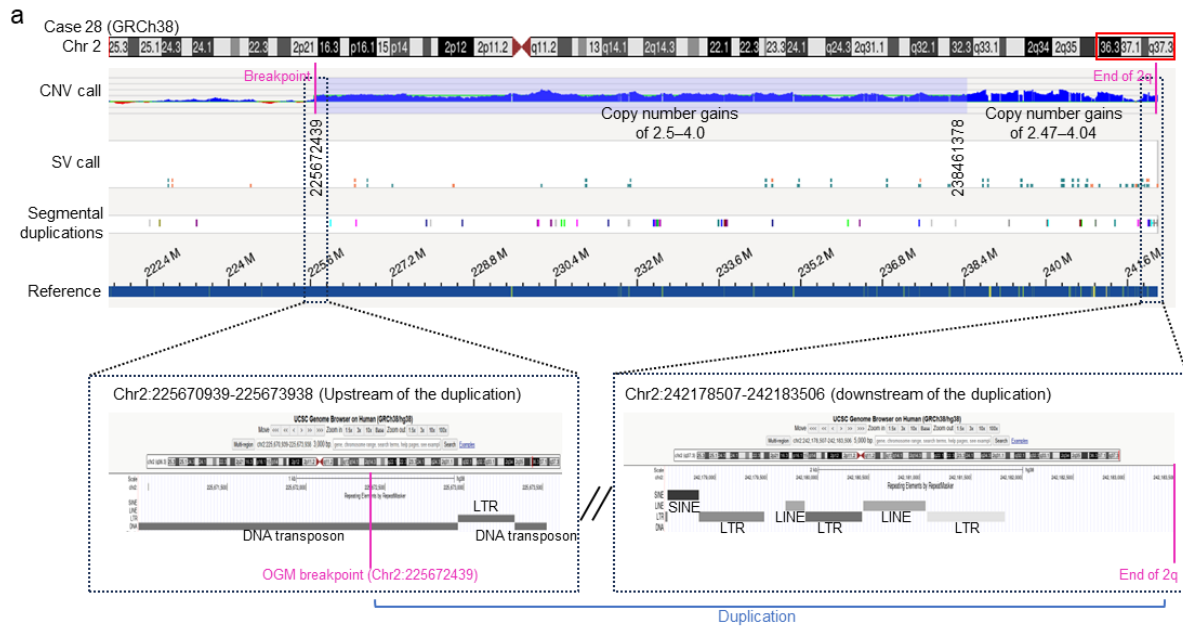


Fig. S5. Results of OGM in Case 28. (a) Upper diagram shows the OGM result in the genome browser. A large duplication is visualized as a blue bar in the CNV call track; the CNV call of the telomere region at the long arm of Chr 2 is not shown despite CN gains of 2.47–4.04. The genome browser was visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. In the lower diagram, the types and sizes of repeat sequences at the breakpoint and its surrounding regions are shown in the UCSC Genome Browser. Both end regions of a duplication contain repetitive sequences. ISCN nomenclature is described in Table S4. Chr, chromosome; CNV, CN variant; ISCN, International System for Human Cytogenomic Nomenclature; LINE, long interspersed nuclear elements; LTR, long terminal repeat; OGM, optical genome mapping; SINE, short interspersed nuclear elements; SV, structural variant.

Fig. S6

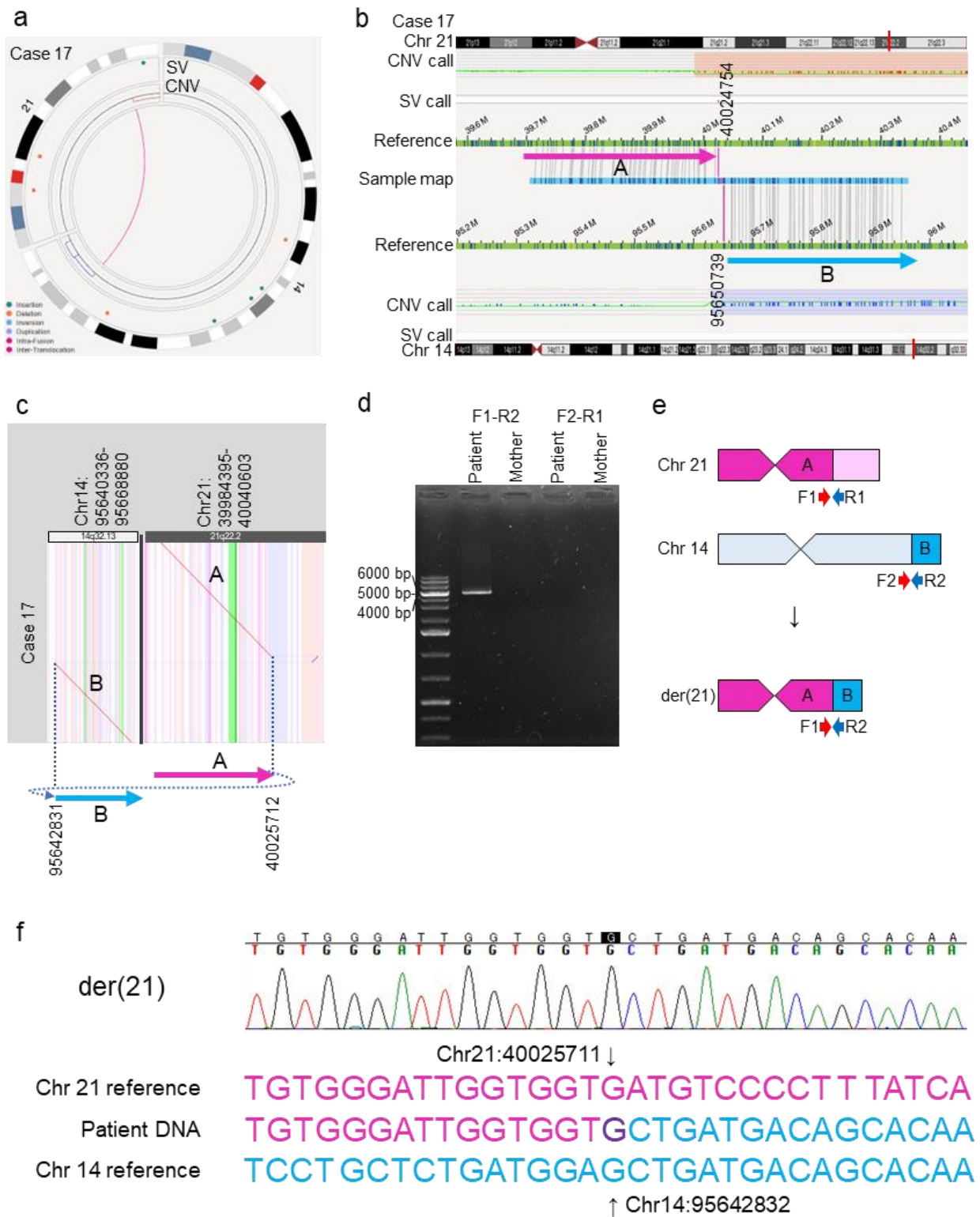


Fig. S6. Results of OGM and long-read sequencing in Case 17. (a) Circos plot of OGM shows unbalanced translocation between Chr 21 and 14. (b) OGM results in the genome browser. A large deletion in Chr 21 is visualized as a red bar and a large duplication in Chr 14 is visualized as a blue bar in the CNV call track. Translocation breakpoints are shown in purple on the sample map between two matching chromosomes. (c) Dot-plot of the consensus sequence from long-read sequencing data. (d) Electrophoresis image of PCR products. (e) Schematic presentation of translocation and PCR primers. (f) Sanger sequencing electropherograms of breakpoints. Translocation breakpoint sequence (middle line) and matching reference sequences (top and bottom lines) are shown in pink (Chr 21) and sky blue (Chr 14). A common single nucleotide G between Chr 14 and 21 at breakpoint is shown in purple. ISCN nomenclature is described in Table S4. Circos plot and the genome browser were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Dot-plot was created by lamassemble 1.4.2 and dnarrange for targeted long-read sequencing. Chr, chromosome; CNV, CN variant; ISCN, International System for Human Cytogenomic Nomenclature; OGM, optical genome mapping; PCR, polymerase chain reaction; SV, structural variant.

Fig. S7

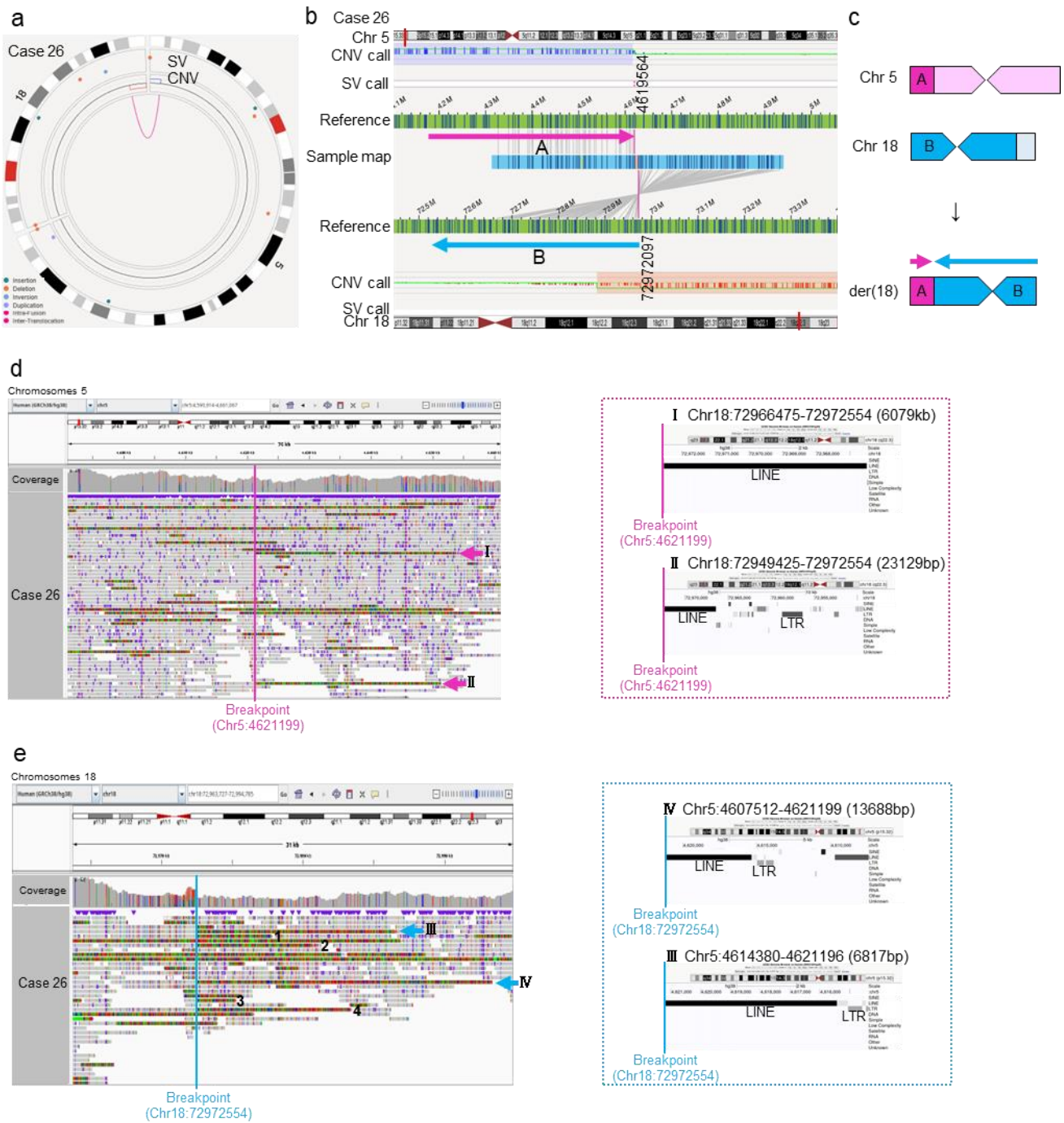


Fig. S7. Results of OGM and long-read sequencing in Case 26. (a) Circos plot of OGM shows unbalanced translocation between Chr 5 and 18. (b) OGM results in the genome browser. A large duplication in Chr 5 is shown as a blue bar and a large deletion in Chr 18 is visualized as a red bar in the CNV call track. Translocation breakpoint is shown in purple on the sample map between two matching chromosomes. (c) Scheme of translocation. (d) Left diagram is IGV of

long-read sequencing data in Chr 5. The breakpoint of translocation is shown as a pink line. Two soft-clipped reads (pink arrows; I and II) show translocation to Chr 18. Duplication in Chr 5 is illustrated as slightly higher coverage upstream of the breakpoint. In the right diagram, the types and sizes of repeat sequences in the regions of supplementary alignments in translocation to Chr 18 are shown in the UCSC Genome Browser. This region contained repeat sequences such as LTR and LINE. (e) Left diagram is IGV of long-read sequencing data in Chr 18. The breakpoint of translocation is shown as a sky blue line. Two soft-clipped reads (blue arrows; III and IV) show translocation to Chr 5 and four soft-clipped reads (1–4) are aligned to other chromosomes unrelated to the translocation. A deletion in Chr 18 is illustrated as lower coverage downstream of the breakpoint. In the right diagram, the types and sizes of repeat sequences in the regions of supplementary alignments in translocation to Chr 5 are shown in the UCSC Genome Browser. This region also contained repeat sequences such as LTR and LINE. ISCN nomenclature is described in Table S4. Circos plot and the genome browses were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Chr, chromosome; CNV, CN variant; IGV, Integrative Genomics Viewer; ISCN, International System for Human Cytogenomic Nomenclature; LINE, long interspersed nuclear elements; LTR, long terminal repeat; OGM, optical genome mapping; SV, structural variant.

Fig. S8

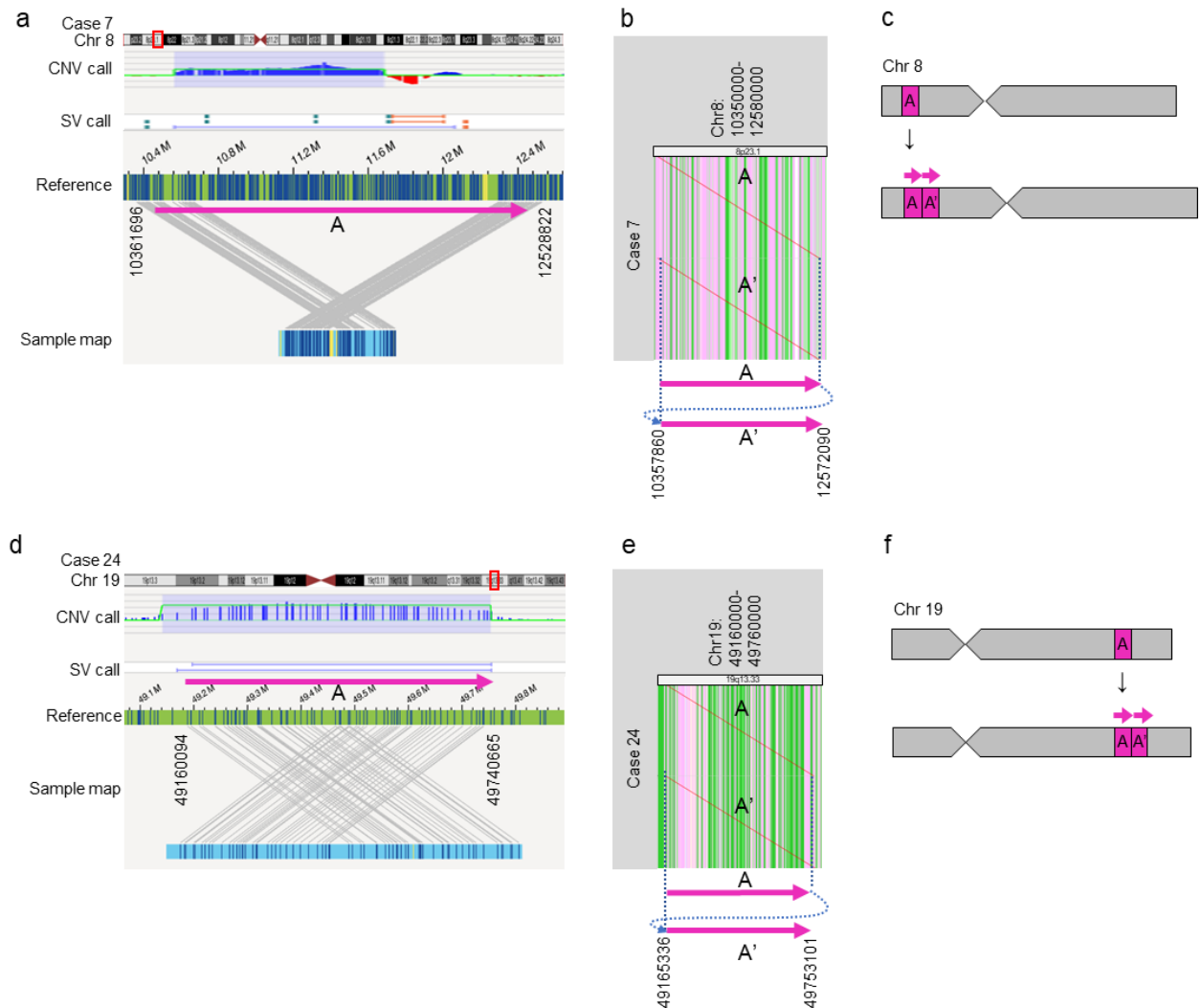


Fig. S8. Results of OGM and long-read sequencing in Cases 7 (a–c) and 24 (d–f) with direct duplications. (a) A large duplication is visualized as a blue bar in the CNV call track and a blue line in the SV call track in the genome browser. Arrow A indicates that the map has a forward orientation to the reference genome. CN loss located immediately downstream of a direct duplication is shown by a red line in the SV call track and was confirmed to be different from the duplication allele using Bionano Access. (b) Dot-plot of the consensus sequence containing a direct duplication region created from long-read sequencing data using dnarrange. (c) Scheme of direct duplication. (d) A large duplication is visualized as a blue bar in the CNV call track and a blue line in the SV call track in the genome browser. (e) Dot-plot of the consensus sequence containing a direct duplication region created from long-read sequencing data using dnarrange. (f) Scheme of direct duplication. ISCN nomenclature in the two cases is

described in Table S4. The genome browsers were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Dot-plots were created by lamassemble 1.4.2 and dnarrange for targeted long-read sequencing. Chr, chromosome; CN, copy number; CNV, CN variant; ISCN, International System for Human Cytogenomic Nomenclature; OGM, optical genome mapping; SV, structural variant.

Fig. S9

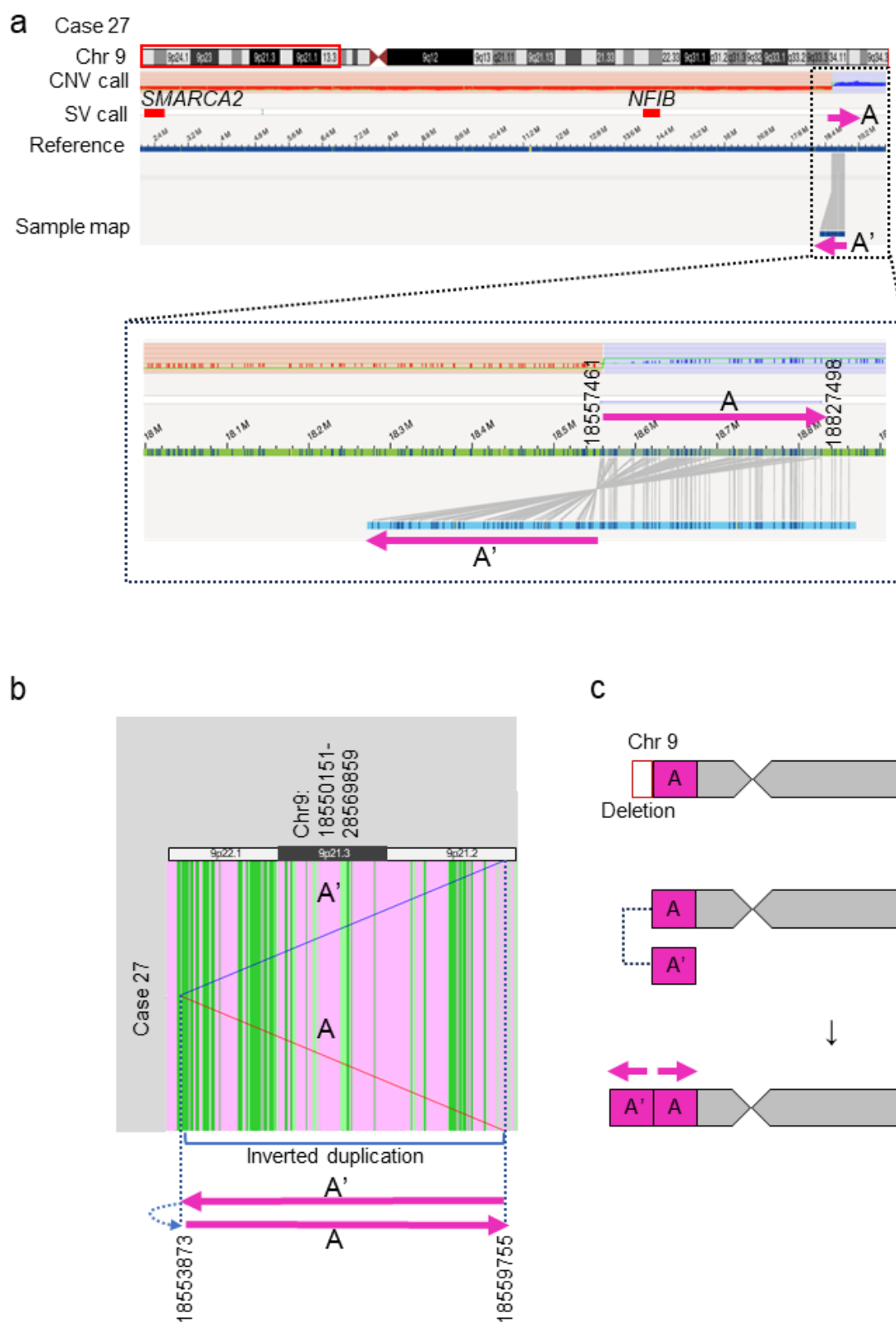


Fig. S9. Results of OGM and long-read sequencing in Case 27. (a) The entire Chr 9 (upper diagram) and its enlargement within the dotted square (lower diagram) in the genome browser. Large deletions and duplications are visualized as red and blue bars, respectively, in the CNV call track. An inverted duplication is indicated by a blue line in the SV call track and is shown in the sample map below. *NFIB* and *SMARCA2* were located in the deletion. (b) Dot-plot of the consensus sequence containing the inverted duplication region created from long-read sequencing data using dnarrange. (c) Scheme of inverted duplication. ISCN nomenclature is described in Table S4. The genome browsers were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Dot-plot was created by lamassemble 1.4.2 and dnarrange for targeted long-read sequencing. Chr, chromosome; CNV, CN variant; ISCN, International System for Human Cytogenomic Nomenclature; OGM, optical genome mapping; SV, structural variant.

Fig. S10

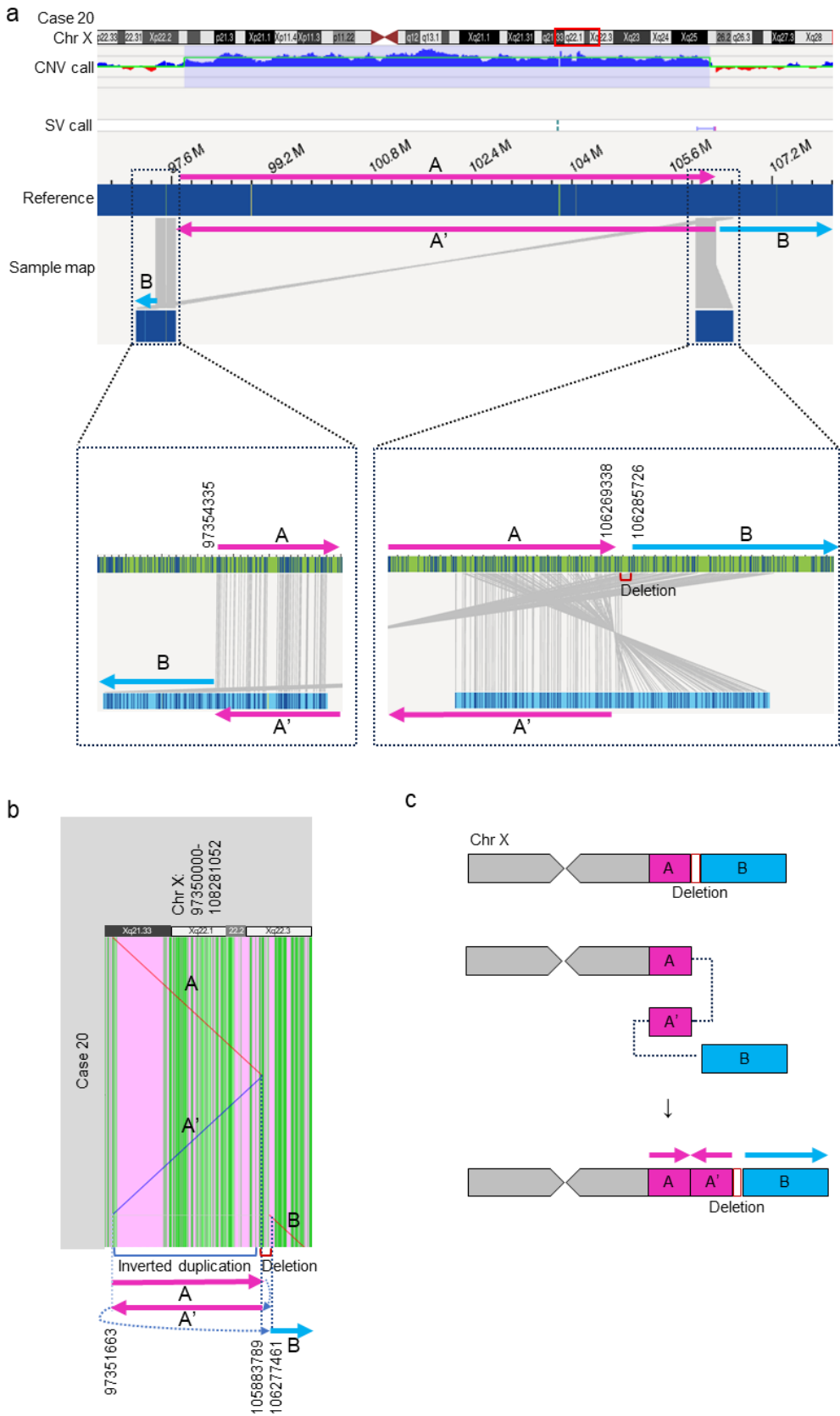


Fig. S10. Results of OGM and long-read sequencing in Case 20. (a) The entire SV of Chr X (upper diagram) and its enlargement within the dotted square (lower diagram) in the genome browser. Large duplication is visualized as a blue bar in the CNV call track and inverted duplication is visualized as a blue line in the SV call track of the genome browser. The numbers are the breakpoint positions. Arrow A' indicates inverted duplication and arrow A' to B indicates intrachromosomal fusion. (b) Dot-plot of the consensus sequence containing inverted duplication and intrachromosomal fusion regions created from long-read sequencing data using dnarrange. (c) Scheme of SVs. ISCN nomenclature is described in Table S4. The genome browsers were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Dot-plot was created by lamassemble 1.4.2 and dnarrange for targeted long-read sequencing. Chr, chromosome; CNV, CN variant; ISCN, International System for Human Cytogenomic Nomenclature; SV, structural variant.

Fig. S11

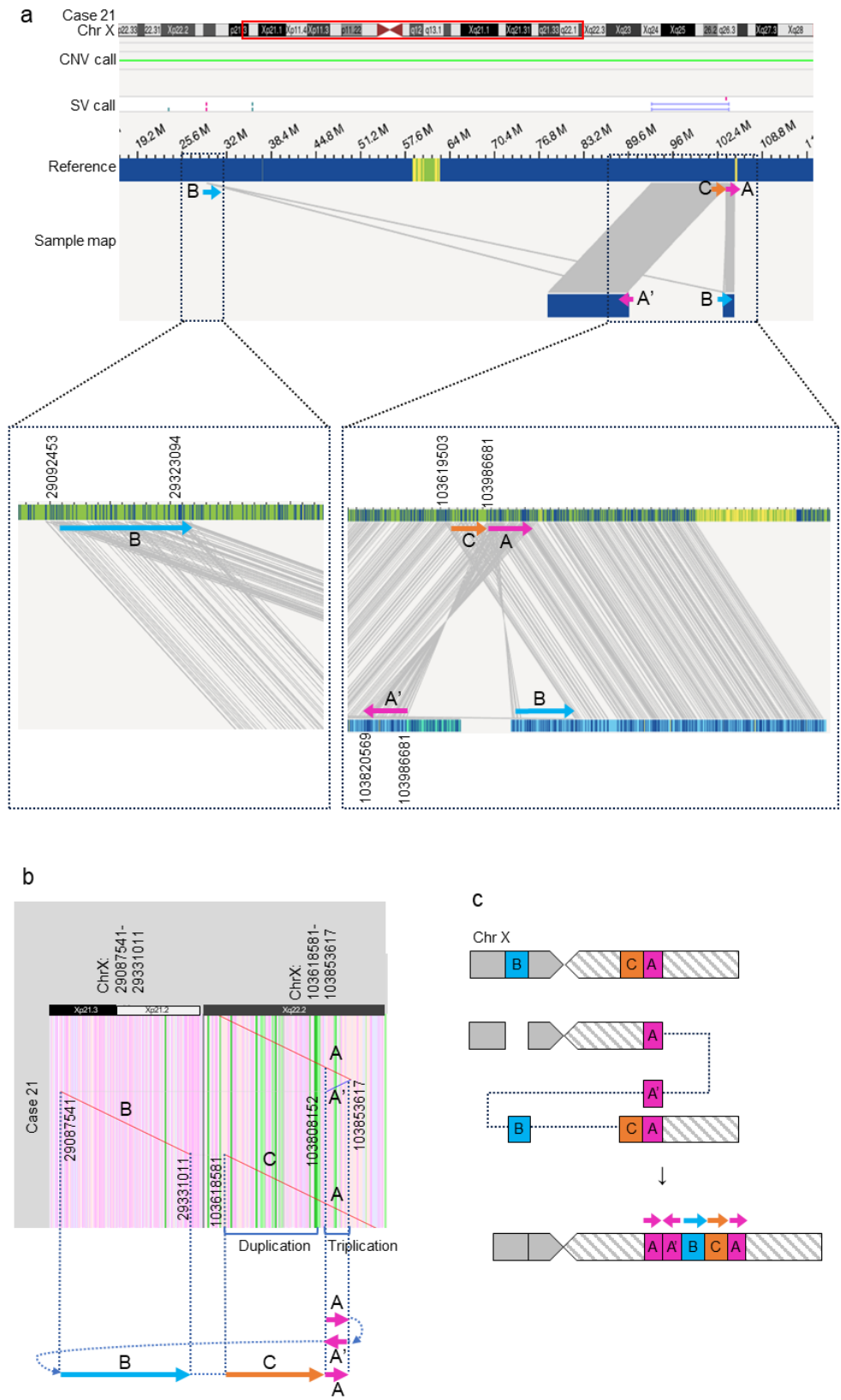
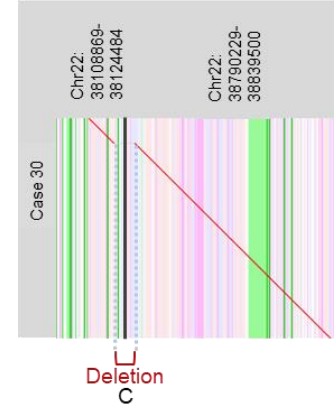
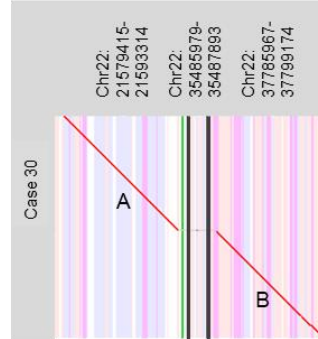
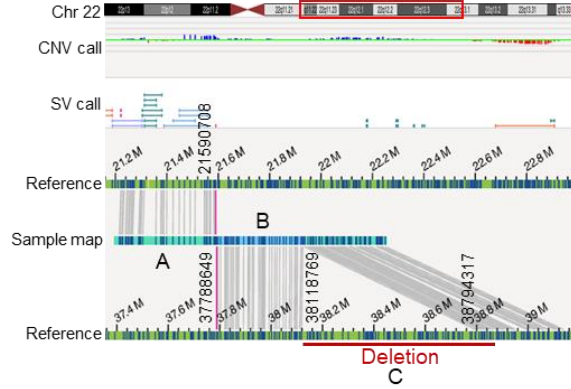


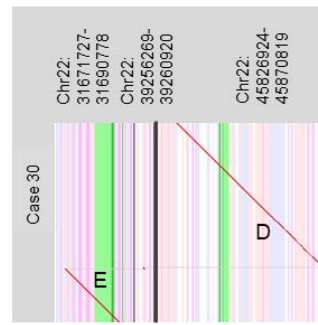
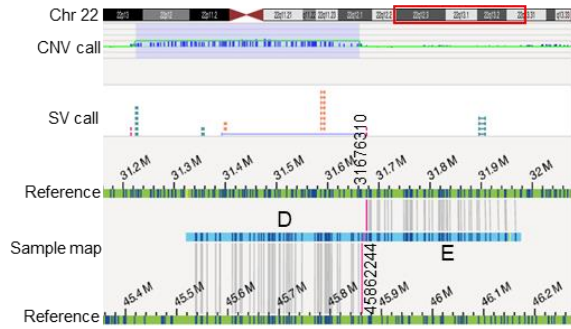
Fig. S11. Results of OGM and long-read sequencing in Case 21. (a) The entire SV of Chr X (upper diagram) and its enlargement within the dotted square (lower diagram) in the genome browser. Inverted duplication is visualized as a blue line in the SV call track of the genome browser. Arrows A and A' indicate inverted duplication, and arrows A', B, and C indicate two intrachromosomal fusions. (b) Dot-plot of the consensus sequence containing inverted duplication and intrachromosomal fusion regions created from long-read sequencing data using dnarrange. (c) Scheme of SVs. ISCN nomenclature is described in Table S4. The genome browsers were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Dot-plot was created by lamassemble 1.4.2 and dnarrange for targeted long-read sequencing. Chr, chromosome; ISCN, International System for Human Cytogenomic Nomenclature; OGM, optical genome mapping; SV, structural variant.

Fig. S12

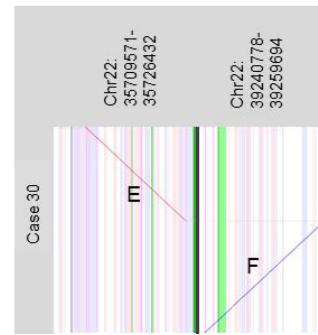
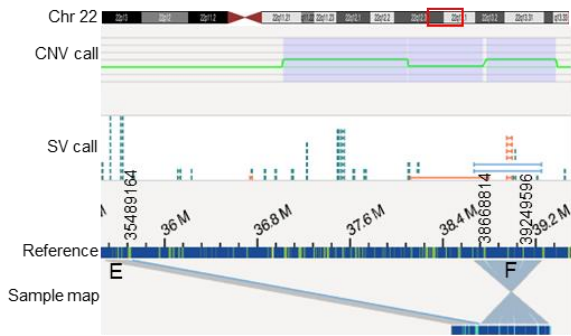
a. Map 1



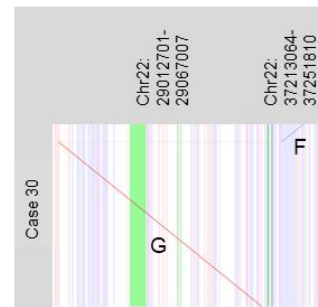
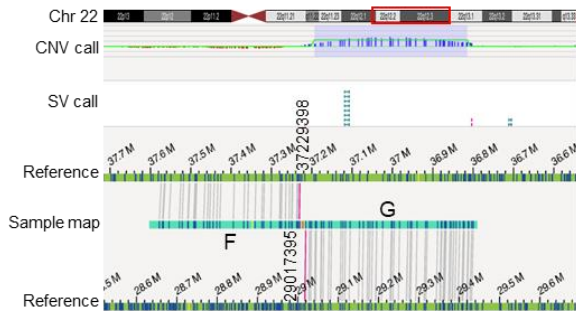
b. Map 2



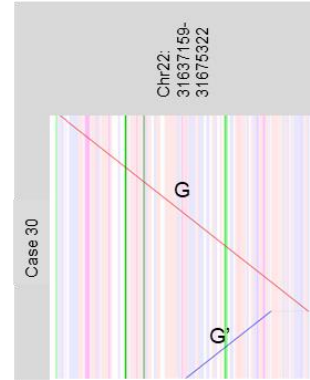
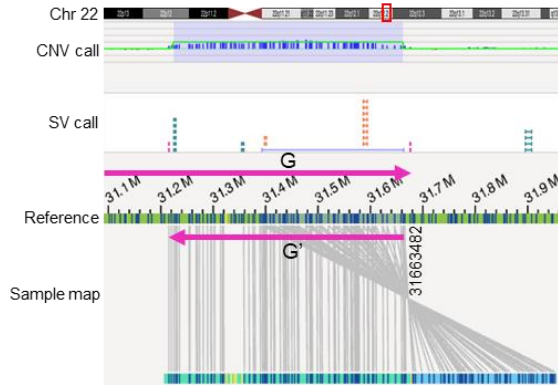
c. Map 3



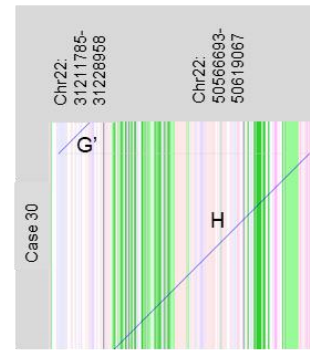
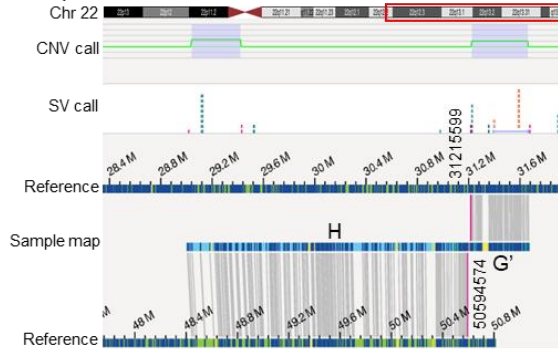
d. Map 4



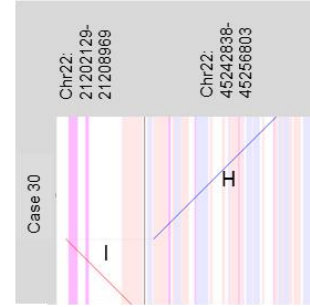
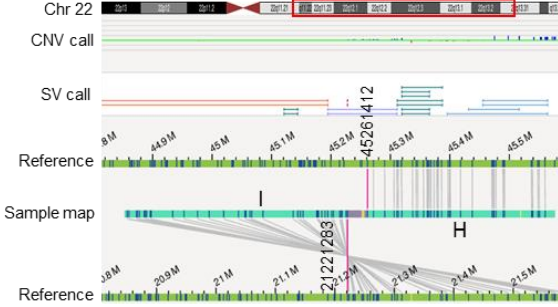
e. Map 5



f. Map 6



g. Map 7



h. Map 8

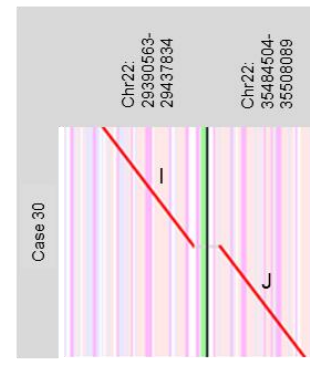
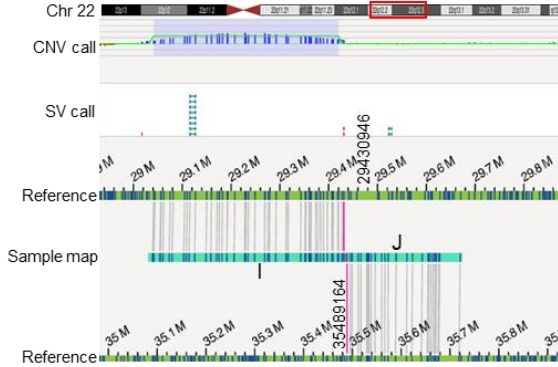


Fig. S12. Results of OGM and long-read sequencing in Case 30. Left diagrams are the genome browser in Chr 22; large duplications are visualized as blue bars in the CNV call track. Right diagrams are dot-plots of the consensus sequence from long-read sequencing data. The map numbers (Map 1–8) and letters (A–J) in each figure correspond to those in Fig. 5b and 5c and Table 3, with Map 1–8 sequentially connecting to the results in Fig. 5c. (a, b, d, f, g, and h). At breakpoints 1, 2, 4, 6, 7, and 8 in the genome browsers, intrachromosomal fusion breakpoints are shown in the region on the sample map between two matching references of Chr 22. (a) This region includes the deletion shown in C. (c) Inversion is visualized as blue lines in the SV call track. (e) Inverted duplication is indicated by a blue line in the SV call track and is shown below in the sample map. The genome browsers were visualized by Bionano Solve 3.8.1 and Bionano Access 1.8.1 for OGM. Dot-plots were created by lamassemble 1.4.2 and dnarrange for targeted long-read sequencing. Chr, chromosome; CNV, CN variant; OGM, optical genome mapping; SV, structural variant.

Fig. S13

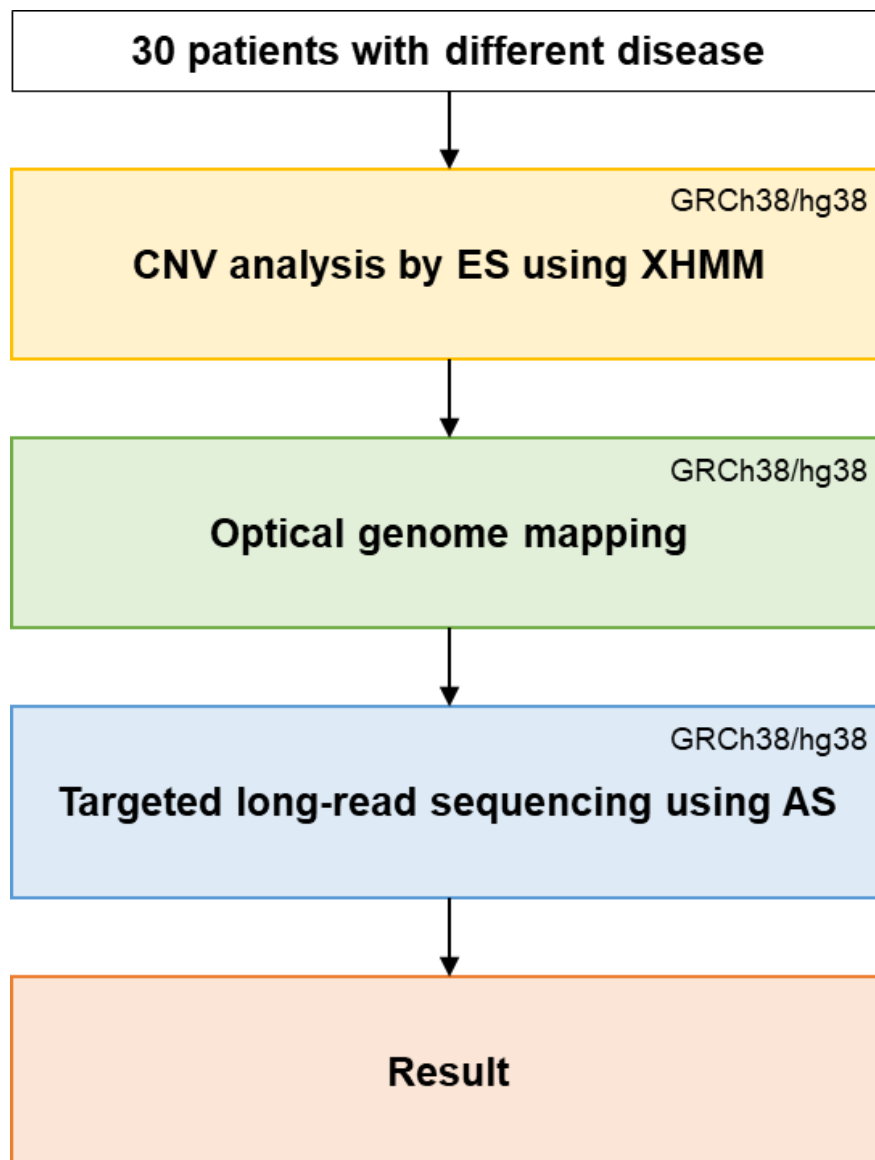


Fig. S13. Overview of the study workflow.

Fig. S14

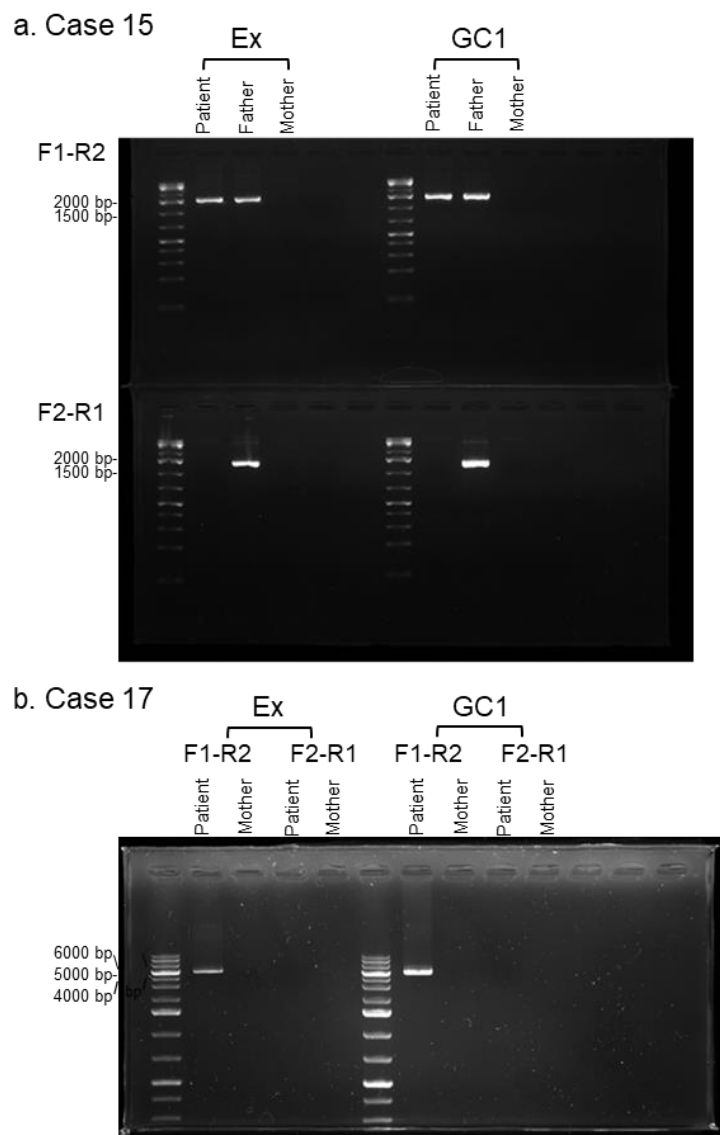


Fig. S14. Original electrophoresis image of polymerase chain reaction products (PCR) in Cases 15 and 17. (a) The upper panel shows the original electrophoresis image of PCR performed using Forward 1 and Reverse 2 primers listed in Table S5. The lower panel shows the original electrophoresis image of PCR performed using Forward 2 and Reverse 1 primers listed in Table S5. For both reactions, PCR conditions were optimized using Ex and GC1 polymerases, and the GC1 polymerase images were used in Fig. 1d which produced clearer bands. (b) The panel shows the original electrophoresis image of PCR performed using primers listed in Table S6. PCR conditions were optimized using Ex and GC1 polymerases, and the Ex polymerase image was used in Fig. S6d which produced clearer bands.

Supplementary Data 1. Gene list in SVs detected by OGM.

Gene list is shown separately in the Supplementary Data 1. pHaplo scores ≥ 0.86 or pTriplo scores ≥ 0.94 denote an odds ratio ≥ 2.7 , which indicates that the average effect sizes of the deletion or duplication are equivalent in strength to the loss-of-function of genes constrained by protein-truncating variants (Karczewski et al., 2020). HI and TS scores range from 0–3 (3: sufficient evidence, 2: emerging evidence, 1: little evidence, 0: no evidence), with increasing levels of evidence suggesting that dosage-sensitivity results in the clinical phenotype (Riggs et al., 2012). Other than scores 0–3, the following ratings were available: 30: gene associated with autosomal recessive phenotype, and 40: dosage sensitivity unlikely. Genes with copy number aberrations identified by OGM that were associated with clinical features in the respective cases are colored in red, and those previously identified by ES are underlined. CNV, CN variant; ES, exome sequencing; HI score, ClinGen haploinsufficiency score; OGM, optical genome mapping; SV, structural variant, TS score; ClinGen triplosensitivity score.

Table S1. DNA molecule quality of OGM.

Case	Saphyr	Molecule N50 (kbp) [value: >230]	Label density (/100kb) [value: 14–17]	Effective coverage of reference (×) [value: >80]
1	109	267.422	16.391	105.918
2	110	286.442	16.04	112.024
3	111	278.008	16.216	110.273
4	112	273.603	16.008	96.439
5	115	369.813	15.645	106.249
6	116	390.251	16.213	117.724
7	117	448.098	16.878	121.303
8	118	440.497	16.983	121.454
9	119	418.941	16.953	121.765
10	120	452.298	16.901	123.217
11	121	396.177	15.502	109.128
12	122	403.824	15.866	109.483
13	123	404.708	16.597	113.798
14	124	402.528	16.19	117.646
15	125	469.351	16.51	117.494
16	133	452.828	16.256	112.756
17	134	436.898	16.426	130.299
18	135	410.239	16.594	119.932
19	136	301.874	15.993	103.262
20	137	267.422	16.391	105.918
21	138	286.442	16.04	112.024
22	139	278.008	16.216	110.273
23	140	273.603	16.008	96.439
24	141	369.813	15.645	106.249
25	142	390.251	16.213	117.724
26	143	448.098	16.878	121.303
27	144	440.497	16.983	121.454
28	145	418.941	16.953	121.765
29	147	396.177	15.502	109.128
30	14	403.824	15.866	109.483

OGM, optical genome mapping.

Table S2. VAF and confidence with SV calls in OGM.

Case	Saphyr	OGM			
		SV call	VAF	Confidence	Supporting molecules ≥ 5
1	109	Deletion	0.489	1.0	Yes
2	110	Deletion	0.450	1.0	Yes
3	111	Deletion	0.485	1.0	Yes
4	112	Deletion	0.495	1.0	Yes
5	115	Deletion	0.430	1.0	Yes
6	116	Duplication	0.245	1.0	Yes
		Inversion	0.160	0.99	Yes
7	117	Direct duplication	0.270	-1	Yes
8	118	Deletion	0.531	1.0	Yes
9	119	Deletion	0.502	1.0	Yes
10	120	Deletion	0.620	0.99	Yes
11	121	Deletion	1.000	1.0	Yes
12	122	Deletion	0.489	1.0	Yes
13	123	Deletion	0.524	1.0	Yes
14	124	Deletion	0.421	1.0	Yes
15	125	Translocation	0.320	0.89	Yes
16	133	Deletion	0.486	1.0	Yes
17	134	Translocation	0.290	0.47	Yes
18	135	Direct duplication	0.350	-1	Yes
19	136	Deletion	0.540	0.99	Yes
20	137	Inverted duplication	0.310	-1	Yes
		Intrachromosomal fusion	0.240	0.23	Yes
21	138	Inverted duplication	0.270	-1	Yes
		Intrachromosomal fusion	0.660	0.46	Yes
22	139	Intrachromosomal fusion	0.290	0.3	Yes
		Inverted duplication	0.350	-1	Yes
23	140	Inverted duplication	0.170	-1	Yes
24	141	Direct duplication	0.210	-1	Yes
25	142	Deletion	0.506	1.0	Yes
26	143	Translocation	0.360	0.98	Yes
27	144	Inverted duplication	0.190	-1	Yes

28	145	Duplication	0.515	1.0	Yes
		Inversion	0.280	0.99	Yes
		Intrachromosomal fusion	0.590	0.32	Yes
29	147	Translocation	0.560	0.58	Yes
		Inversion	0.930	0.99	Yes
		Translocation	0.460	0.82	Yes
		Intrachromosomal fusion	0.590	0.23	Yes
		Intrachromosomal fusion	0.270	0.60	Yes
		Deletion	0.260	0.99	Yes
		Intrachromosomal fusion	0.400	0.13	Yes
30	14	Inversion	0.270	1.0	Yes
		Intrachromosomal fusion	0.180	0.14	Yes
		Inverted duplication	0.250	−1	Yes
		Intrachromosomal fusion	0.450	0.10	Yes
		Intrachromosomal fusion	0.250	0.10	Yes
		Intrachromosomal fusion	0.300	0.11	Yes

Confidence in OGM is defined as the normalized confidence score between 0 and 1 (each SV is assigned a confidence score derived from the probability [p-value] that the observed molecule patterns could occur by random chance. A confidence score was defined as $-\log_{10}(p)$), and higher values indicate higher confidence. Confidence for duplications called using the SV calling pipeline of OGM was undefined and set to −1. Supporting molecules in OGM were defined as plausible SV calls when at least five molecules were required around SV breakpoints. OGM, optical genome mapping; SV, structural variant; VAF, variant allele frequency.

Table S3. Read coverage of targeted long-read sequencing using adaptive sampling with GridION.

Case	Saphyr	Target region of adaptive sampling (GRCh38/hg38)				Mean depth
		Chr	Start position	End position	Target region size (bp)	
1	109	1	144179918	147929918	7500000	12.62×
			146722608	150472608		
2	110	4	1	7500000	7500000	15.87×
3	111	8	114617663	118367663	7500000	14.21×
			124652331	128402331		
4	112	22	17001658	20751658	7500000	27.51×
			19325208	23075208		
5	115	15	28559571	32309571	7500000	20.39×
			30549187	34299187		
6	116	21	38544399	46044399	7500000	23.13×
7	117	8	7695259	15195259	7500000	11.72×
8	118	7	70064977	73814977	7500000	16.31×
			85335944	89085944		
9	119	6	64037218	67787218	7500000	20.83×
			78842030	82592030		
10	120	17	1	3750000	7500000	13.01×
			536268	4286268		
11	121	X	4591265	8341265	7500000	6.10×
			6288693	10038693		
12	122	2	183429800	187179800	7500000	18.87×
			201498928	205248928		
13	123	16	12806564	16556564	7500000	18.13×
			14498394	18248394		
14	124	19	9576325	13326325	7500000	10.59×
			11574394	15324394		
15	125	1	1956717	5456717	7500000	28×
		17	2081287	5581287		41.14×
16	133	8	74715158	78465158	7500000	15.72×
			79618492	83368492		
17	134	14	93900739	97400739	7500000	28.58×
		21	38274754	41774754		16.65×

18	135	2	161970538	169470538	7500000	24.53×
			52578746	56328746		
19	136	19	54082898	57832898	7500000	7.55×
			95604335	99104335		
20	137	X	104251620	108035726	7284106	34.47×
			28123094	30523094		
21	138	X	91736405	94136405	7567178	26.07×
			102419503	105186681		
			57375606	62572583		
22	139	18	9580464	11980464	7596977	35.79×
23	140	9	10623459	18392814	7769355	30.48×
24	141	19	45560094	53340665	7780571	31.60×
			91225823	94975823		
25	142	1	109446763	113196763	7500000	9.14×
			71122097	74822097		
26	143	18			7400000	30.21×
		5	2769564	6469564		47.38×
27	144	9	14807461	22577498	7770037	33.51×
			223797439	227547439		
28	145	2	236586378	240336378	7500000	54.42×
			70420270	73006564		
29	147	12	77201558	78886157	7539944	31.83×
			83632163	86901214		
		14				34.14×
30	14	22	21161283	50654574	29493291	26.97×

Chr, chromosome.

Table S4. ISCN nomenclature list identified using OGM with long-read sequencing.

Case	ISCN nomenclature
1	ogm[GRCh38] del(1)(q21.1q21.2)(146,054,918_148,597,608)
2	ogm[GRCh38] del(4)(p16.3)(12,985_11,209,596).seq[GRCh38] del(4)(p16.3)dn NC_000004.12:g.pter_1,225,480del
3	ogm[GRCh38] del(8)(q23.3q24.21)(116,492,663_126,527,331).seq[GRCh38] del(8)(q23.3q24.21)dn NC_000008.11:g.116,494,154_126,506,598del
4	ogm[GRCh38] del(22)(q11.21q11.21)(18,876,658_21,200,208)
5	ogm[GRCh38] del(15)(q13.2q13.3)(30,434,571_32,424,187)
6	ogm[GRCh38] der(21)(pter→q22.3::q22.3→q22.3::q22.3→q22.2:)(pter_41,463,212::45,321,003_41,463,213::41,463,212_39,267,796:).seq[GRCh38] der(21)(pter→q22.3::q22.3→q22.3::q22.3→q22.2:)dn NC_000021.9:g.[41,467,831_45,493,272inv; 45,493,272_45,493,273ins39,279,216_41,467,830inv;45,493,273_qterdel]
7	ogm[GRCh38] dup(8)(p23.1p23.1)(10,361,696_12,528,822).seq[GRCh38] dup(8)(p23.1p23.1)dn NC_000008.11:g.10,357,860_12,572,090dup
8	ogm[GRCh38] del(7)(q11.22q21.12)(71,939,977_87,210,944).seq[GRCh38] del(7)(q11.22q21.12)dn NC_000007.14:g.71,949,757_87,202,195del
9	ogm[GRCh38] del(6)(q12q14.1)(65,912,218_80,717,030).seq[GRCh38] del(6)(q12q14.1)dn NC_000006.12:g.65,946,371_80,652,635del
10	ogm[GRCh38] del(17)(p13.3p13.3)(943,147_2,411,268).seq[GRCh38] del(17)(p13.3p13.3)dn NC_000017.11:g.957,506_2,408,537del
11	ogm[GRCh38] del(X)(p22.31p22.31)(6,466,265_8,163,693).seq[GRCh38] del(X)(p22.31p22.31)mat NC_000023.11:g.6,533,591_8,169,074del
12	ogm[GRCh38] del(2)(q32.1q33.2)(185,304,800_203,373,928).seq[GRCh38] del(2)(q32.1q33.2)dn NC_000002.12:g.185,302,763_203,371,636del
13	ogm[GRCh38] del(16)(p13.12p13.11)(14,681,564_16,373,394)

14	ogm[GRCh38] del(19)(p13.2p13.13)(11,451,325_13,449,394).seq[GRCh38] del(19)(p13.2p13.13)dn NC_000019.10:g.11,498,892_13,309,177del
15	ogm[GRCh38] der(1)(17pter→17p13.2::1p36.32→1qter)(17pter_3831287::3,706,717_1qter).seq[GRCh38] der(1)t(1;17)(p36.32;p13.2)pat NC_000001.11:g.pter_3,699,571delins[NC_000017.11:g.pter_3,847,652;ACG]
16	ogm[GRCh38] del(8)(q21.13q21.13)(76,590,158_81,493,492).seq[GRCh38] del(8)(q21.13q21.13)dn NC_000008.11:g.76,598,298_81,428,940del
17	ogm[GRCh38] der(21)(21pter→q22.2::14q32.13→14qter)(21pter_40,024,754::95,650,739_14qter).seq[GRCh38] der(21)t(14;21)(q32.13;q22.2) NC_000021.9:g.40,025,712_qterdelins[NC_000014.9:g.95,642,831_qter]
18	ogm[GRCh38] dup(2)(q24.3q24.3)(164,976,570_166,464,506).seq[GRCh38] dup(2)(q24.3q24.3)dn NC_000002.12:g.164,975,946_166,466,854dup
19	ogm[GRCh38] del(19)(q13.42q13.43)(54,453,746_55,957,898).seq[GRCh38] del(19)(q13.42q13.43)dn NC_000019.10:g.54,461,012_55,947,282del
20	ogm[GRCh38] der(X)(pter→q22.3::q22.3→q21.33::q22.3→qter)(pter_106,229,338::106,229,338_97,354,335::106,285,726_qter).seq[GRCh38] der(X)(pter→q22.3::q22.3→q21.33::q22.3→qter)dn NC_000023.11:g.[pter_105,883,789;105,883,789_105,883,790ins97,351,663_105,883,789inv;105,883,790_106,277,460del;106,277,461_qter]
21	ogm[GRCh38] der(X)(pter→p21.3::p21.2→q22.2::q22.2→q22.2::p21.3→p21.2::q22.2→qter)(pter_29,092453::29,323,094_103,986,681::103,986,681_103,820,569::29,092453_29,323,094::103,619,503_qter).seq[GRCh38] der(X)(pter→p21.3::p21.2→q22.2::q22.2→q22.2::p21.3→p21.2::q22.2→qter)dn NC_000023.11:g.[29,087,541_29,331,011del;103,853,617_103,853,618ins103,808,152_103,853,617inv;29,087,541_29,331,011;103,618,581_qter]

22	ogm[GRCh38] der(18)(pter→p11.22::q21.31→q21.33::q21.33→q21.31::p11.22→pter)(pter_10,780,464::58,575,606_61,372,583::61,372,583_58,575,606::10,780,464→pter).seq[GRCh38] der(18)(pter→p11.22::q21.31→q21.33::q21.33→q21.31::p11.22→pter)dn NC_000018.10:g.[pter_10,785,166;58,557,581_61,373,711;61,373,711_61,373712ins58,557,581_61,373,711inv;61,373712_qterdelins10,785,166_pterinv]
23	ogm[GRCh38] inv dup del(9)(:p22.3→p22.3::p22.3→qter)(14,642,814_14,373,459::14,373,459_qter).seq[GRCh38] inv dup del(9)(:p22.3→p22.3::p22.3→qter)dn NC_000009.12:g.[pter_14,369,156delins14,369,157_14,397,270inv;14,369,157_qter]
24	ogm[GRCh38] dup(19)(q13.33q13.33)(49,160,094_49,740,665).seq[GRCh38] dup(19)(q13.33q13.33)dn NC_000019.10:g.49,165,336_49,753,101dup
25	ogm[GRCh38] del(1)(p22.1p13.2)(931,008,23_111,321,763).seq[GRCh38] del(1)(p22.1p13.2)dn NC_000001.11:g.93,106,498_111,329,159del
26	ogm[GRCh38] der(18)(18pter→18q22.3::5p15.32→5pter)(18pter_72,972,097::4,619,564_5pter).seq[GRCh38] der(18)t(5;18)(p15.32;q22.3) NC_000018.10:g.72,972,554_qterdelins[NC_000005.10:g.pter_4,621,199inv]
27	ogm[GRCh38] inv dup del(9)(:p21.2→p22.1::p22.1→qter)(18,827,498_18,557,461::18,557,461_qter).seq[GRCh38] inv dup del(9)(:p21.2→p22.1::p22.1→qter)dn NC_000009.12:g.[pter_18,553,872delins18,553,873_18,559,755inv;18,553,873_qter]
28	ogm[GRCh38] dup(2)(q36.3q37.3)(225,672,439_238,461,378)
29	ogm[GRCh38] der(12)(12pter→q15::12q21.1→q15::12q21.2→q21.2::14q31.3→q31.2::14q31.3→qter)(12pter_70,920,269::72,506,564_70,920,270::78,307,029_78,386,157;86,401,214_84,178,337::86,401,215_14qter),der(14)(14pter→14q31.2::12q21.2→12q21.2::12q21.2→12q21.2::12q21.2→12qter)(14pter_84,132,163;78,307,029_78,058,671::77,959,302_77,701,558::78,386,158_12qter).seq[GRCh38] der(12)(12pter→q15::12q21.1→q15::12q21.2→q21.2::14q31.3→q31.2::14q31.3→qter)dn NC_000012.12:g.[pter_70,917,323;70,917,324_72,511,491inv;78,304,150_78,382,960;78,382,961_qterdelins[NC_000014.9:g.84,172,309_86,404,434inv;86,404,435_qter]],der(14)(14pter→14q31.2::12q21.2→12q21.2::12q21.2→12q21.2::12q21.2→12qter)dn

NC_000014.9:g.[pter_84,135,076;84,135,077_qterdelins[NC_000012.12:g.78,304,148_78,382,967inv;77,696,122_78,304,147;78,382,967_qter]]

30 ogm[GRCh38]
der(22)(pter→q11.21::q13.1→q13.1::q13.1→q13.31::q12.2→q12.3::q13.1→q13.1::q12.1→q12.2→q12.2::q13.33→q13.31::q11.21→q12.2::q12.3→q13.1::q13.33→qter)(pter_21,590,708::37,788,649_38,118,769::38,794,317_45,862,244::31,676,310_35,489,164::39,249,596_37,229,398::29,017,395_31,663,482_31,215,599::50,594,574_45,261,412::21,221,283_29,430,946::35,489,164_37,788,649::50,594,575_qter).seq[GRCh38]
der(22)(pter→q11.21::q13.1→q13.1::q13.1→q13.31::q12.2→q12.3::q13.1→q13.1::q12.1→q12.2→q12.2::q13.33→q13.31::q11.21→q12.2::q12.3→q13.1::q13.33→qter)dn
NC_000022.11:g.[pter_21,592,378;37,786,951_38,122,247;38,122,248_38,792,465del;38,792,466_45,868,639;31,673,907_35,725,198;39,258,460_37,216,272inv;29,015,909_31,673,909;31,673,908_31,673,909ins31,214,183_31,673,908inv;45,243,553_50,616,489inv;21,202,844_29,435,389;35,486,947_37,786,951;50,616,489_qter]

ISCN, International System for Human Cytogenetic Nomenclature; OGM, optical genome mapping.

Table S5. Primer sequences used in Case 15.

Primer name	Location		Primer sequence (forward)	Primer name	Location		Primer sequence (reverse)	Product size (bp)
	Start	End			Start	End		
F1	Chr17: 3846937	Chr17: 3846946	AGTTTCCGGCGGATATTCT	R1	Chr17: 3848331	Chr17: 3848350	TTCTGGATGGCACTGAACTG	1414
F2	Chr1: 3698911	Chr1: 3698930	CAGTCCTGCAGACACACGTC	R2	Chr1: 3700269	Chr1: 3700288	AGAGAAACCTGCCAGCAAAA	1378

Chr, chromosome.

Table S6. Primer sequences used in Case 17.

Primer name	Location		Primer sequence (forward)	Primer name	Location		Primer sequence (reverse)	Product size (bp)
	Start	End			Start	End		
F1	Chr21: 40022468	Chr21: 40022495	AACCAGGATAAGCAGTTGTACCCCTCAG	R1	Chr21: 40028934	Chr21: 40028961	ACAGCTAGATAATGCTGCCTGCTTTGTG	6494
F2	Chr14: 95641353	Chr14: 95641379	CCTTGTAACCTGTTGTCTGGGCTTGT	R2	Chr14: 95644289	Chr14: 95644315	CTCTTGGGGTTCCCCACTTCTATTGAG	2963

Chr, chromosome.