

<https://doi.org/10.1038/s41525-026-00561-4>

Completely resolved structural variants by optical genome mapping with adaptive sampling from CNV discovery

Check for updates

Li Fu^{1,2}, Chong Ae Kim³, Masatoshi Tokita⁴, Yohane Miyata⁴, Nobuhiko Okamoto⁵, Yoshio Makita⁶, Hitoshi Osaka⁷, Ayataka Fujimoto⁸, Atsuro Daida⁹, Jun Nirei¹⁰, Noriko Udagawa¹¹, Seiichi Hayakawa¹², Kimiko Deguchi¹³, Mitsumasa Fukuda¹⁴, Hiroshi Matsumoto¹⁵, Manami Akasaka¹⁶, Junichiro Okada¹⁷, Yohei Misumi¹⁸, Jun Kido¹⁹, Toshifumi Suzuki^{2,20}, Hiromi Aoi², Rie Seyama², Haruka Hamanoue²¹, Satomi Mitsuhashi¹, Sachiko Ohori¹, Ken Saida¹, Yuta Inoue¹, Kohei Hamanaka¹, Naomi Tsuchida^{1,22}, Yuri Uchiyama^{1,22}, Eriko Koshimizu¹, Satoko Miyatake^{1,21}, Takeshi Mizuguchi¹, Atsuo Itakura², Naoki Harada²³, Atsushi Fujita¹ & Naomichi Matsumoto^{1,21,22} ✉

Structural variants (SVs), including duplications, deletions, inversions, translocations, and insertions, play major roles in human phenotypic diversity but remain difficult to detect because of variable size and structural complexity. Optical genome mapping (OGM) uses ultra-high molecular weight DNA (>150 kb) fluorescently labeled at a specific six-nucleotide sequence, enabling comprehensive SV detection by analyzing labeling patterns along long DNA molecules. This study aimed to fully characterize SVs using OGM. OGM was applied to 30 cases with exome sequencing-based copy number (CN) variants (16 CN losses, 7 CN gains, and 7 combined CN losses and gains). Additionally, targeted Oxford Nanopore long-read sequencing with adaptive sampling was used to determine breakpoints of SVs. This approach revealed undetected SVs in 14 cases (46.7%), and disclosed gene disruptions or CN alterations explaining clinical features in seven cases (23.3%). Even complex SVs involving numerous chromosomal segments and breakpoints were resolved efficiently, highlighting the power of combining OGM and long-read sequencing. Integrating OGM with long-read sequencing improves diagnostic resolution beyond sequencing alone and provides a robust framework for interpreting complex SVs. These findings highlight the potential clinical utility of combining OGM and long-read sequencing as a comprehensive diagnostic strategy for improved precision medicine in rare genetic diseases.

Structural variations (SVs)—defined as genomic alterations larger than 50 base pairs (bp) and including duplications, deletions, inversions, translocations, and insertions—constitute the majority of variable nucleotides in the human genome and contribute significantly to phenotypic diversity¹. Although exome sequencing (ES) enables the detection of copy number (CN) variants (CNVs)², it frequently fails to capture SV breakpoints, primarily because of the limited coverage of capture regions and technical limitations such as polymerase chain reaction (PCR) and hybridization biases in GC-rich regions, as well as poor alignment in low-complexity regions.

Even with genome sequencing, SV detection remains challenging because of the wide range of SV types and sizes and their frequent

localization within repetitive genomic regions. These complexities hinder both sensitivity and specificity in SV calling, especially for structurally complex and repetitive regions. As a result, many pathogenic SVs likely remain undetected, and the reliable identification of such variants from short-read sequencing data continues to be a major technical and computational challenge^{3,4}.

Several methodologies have been used for SV detection, including conventional karyotyping, fluorescence in situ hybridization, chromosomal microarray analysis, and long-read sequencing. More recently, optical genome mapping (OGM) has emerged as a promising technique for the comprehensive analysis of SVs. Unlike sequencing-based methods, OGM uses high-resolution fluorescence imaging of ultra-high molecular weight

A full list of affiliations appears at the end of the paper. ✉ e-mail: naomat@yokohama-cu.ac.jp

DNA to detect structural changes without nucleotide-level sequencing. Long-read sequencing produces reads typically ranging from 10 to 100 kilobases (kb), whereas OGM analyzes DNA molecules exceeding 150 kb, providing broader genome coverage (>80×), albeit at lower breakpoint resolution^{3,5,6}.

OGM identifies SVs through the de novo assembly of fluorescently labeled DNA molecules using a specific six-base sequence motif (CTTAAG), which occurs on average every 5 kb across the genome^{7–9}. This method enables the detection of a wide range of structural alterations, including insertions and deletions (>500 bp), duplications (>50 kb), inversions and translocations (>50–70 kb), chromosomal aneuploidy, absence of heterozygosity, triploidy, and copy-neutral events^{10,11}. In this regard, OGM may outperform traditional sequencing-based platforms for identifying large and complex SVs.

Recent genomic studies have demonstrated that ~0.5–4% of individuals with rare diseases harbor dual or triple molecular diagnoses, in which a single diagnosis explains only part of the observed phenotype¹². These findings emphasize the importance of comprehensive SV discovery beyond CNVs, because additional SVs may be responsible for the unexplained portion of a phenotype. It is therefore critical to identify full SV configurations, rather than mere CN changes, for accurate genotype–phenotype correlations and clinical diagnoses. In the present study, we aimed to use OGM to comprehensively characterize constitutional SVs that were initially identified as CNVs using ES. Furthermore, we used targeted Oxford Nanopore long-read sequencing with adaptive sampling (AS) to resolve the breakpoint structures of these newly discovered SVs at nucleotide resolution.

Results

We applied OGM to 30 cases (Table 1 and Fig. S1; case numbers are listed together with the Saphyr serial numbers assigned by our laboratory) and identified only the expected ES-identified CNVs in 16 cases (Cases 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 13, 14, 16, 19, 25, and 28). By contrast, previously undetected SVs were revealed in the remaining 14 cases (Fig. S1 and Tables 2 and 3). The newly discovered SVs in the latter 14 cases included: three translocations (Cases 15, 17, and 26), three direct duplications (Cases 7, 18, and 24), one duplication with inversion (Case 6), two inverted duplications (Cases 23 and 27), three intrachromosomal insertion events with inverted duplications (Cases 20, 21, and 22), one chromoplexy-like event (Case 29), and one chromothripsis (Case 30). The variant allele frequency (VAF), confidence, and supporting molecules of all SV calls identified by OGM are shown in Table S2. Confidence in OGM was defined as a normalized confidence score between 0 and 1; higher values indicate higher confidence. It has been reported that intrachromosomal fusions often exhibit lower confidence scores because of a higher probability of similar labeling patterns occurring by chance within the same chromosome¹³. Although intrachromosomal fusions also had low OGM confidence scores (of ~0.1–0.3) in the present study, SV calls were judged to be highly reliable from supporting molecules.

To further characterize the SVs in all 30 cases, we used Oxford Nanopore targeted long-read sequencing with AS (Table S3). This approach allowed us to precisely determine the SV breakpoints at the nucleotide level and generate dot plots to visualize the identified SVs.

OGM and long-read sequencing provide complementary and comprehensive detection of diverse SVs, including those invisible to ES. The representative cases demonstrated that integrating optical and sequence-level data enhances both the diagnostic yields and the mechanistic understanding of SV formation. Here, we present findings obtained through OGM analysis and long-read sequencing with AS. A detailed analysis of each case is provided in the Supplementary Information.

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

We performed OGM in 16 cases with ES-based CN losses, and detected only simple (expected) deletions in 15 cases; however, a chromoplexy-like event was identified in one case (Case 29) (Table 2). Breakpoints at nucleotide

resolution were successfully determined in 11 of the 15 cases with simple deletions: interstitial deletions in Cases 3, 8, 9, 10, 11, 12, 14, 16, 19, and 25, and a deletion involving 4q telomeric sequences (TTAGGG) in Case 2 (Figs. S2 and S3).

Among these 11 cases, although ES-based analysis did not identify haploinsufficient genes, we identified *SUMO1* (pHaplo score = 0.900¹⁴, ClinGen haploinsufficiency [HI] score = not recorded¹⁵) in Case 12 and *COL11A1* (pHaplo score = 0.894¹⁴, HI score = 1¹⁵) in Case 25 (Supplementary Data 1^{15,16}). *SUMO1* deletion causes orofacial cleft (OMIM# 613705) and *SUMO1* haploinsufficiency has been associated with cleft lip and palate, consistent with the phenotype observed in Case 12^{17,18}. In Case 25, ES had already identified *KCNA2* (OMIM# 616366) as a cause of seizures and global developmental delay^{19,20}. Our analysis additionally revealed *COL11A1*, whose haploinsufficiency causes Marshall (OMIM# 154780) and Stickler (OMIM# 604841) syndromes, both of which are associated with abnormal facial features, failure to thrive, and short stature, as observed in Case 25^{21,22}. Upon reviewing the ES data, we identified the same regions as simple CN losses in ES data; however, the corresponding genes had not been called by the eXome Hidden Markov Model (XHMM). These cases underscore the importance of OGM and long-read sequencing and of reviewing the ES data, thus demonstrating its limitations.

CTBP1 was disrupted by the deletion breakpoint in Case 2 (Table 2). *CTBP1* aberration is associated with hypotonia, ataxia, developmental delay, and tooth enamel defect syndrome (OMIM# 617915) in an autosomal dominant fashion, and may influence the neurological features observed in this case²³. Long-read analysis enabled us to determine the precise location at which the gene was disrupted.

In Cases 1, 4, 5, and 13, although deletions were identified using OGM (Table 2 and Fig. S4), the breakpoint determination of deletions using long-read sequencing was unsuccessful. Segmental duplications were exactly mapped to deletion breakpoints, and long-read sequencing cannot cover these segmental duplications (Fig. S4). Cases 1, 4, 5 and 13 are described in the Supplementary information.

Duplication (Case 28)

We identified an ES-based CN gain in Case 28. OGM detected the expected duplication at the terminal region of chromosome 2q (Fig. S5). However, we were unable to fully characterize the duplication using both OGM and long-read sequencing because of the presence of many repetitive sequences at the terminal region of the chromosome. Case 28 is described in detail in the Supplementary information.

Unbalanced translocations (Cases 15, 17, and 26)

In Cases 15, 17, and 26, unbalanced translocations—characterized by combined terminal CN gains and CN losses—were detected using ES. OGM also identified unbalanced translocations in these three cases.

Case 15 presented clinically with West syndrome, hypoplasia of the corpus callosum, and aortic regurgitation. In this case, an unbalanced translocation was detected using OGM and AS, with a 3.5-Mb deletion at 1p36.33–p36.32 and a 3.8-Mb duplication at 17p13.3–p13.2 (Fig. 1a, b, Table 2, and Table S4). The deletion encompassed *GABRD* (pHaplo score = 0.807¹⁴, HI score = 3¹⁵), which is associated with susceptibility to generalized epilepsy with febrile seizures plus, type 5 (GEFSP5) (OMIM# 613060) and has been reported to cause epileptic symptoms in 1p36 deletion syndrome²⁴ (Table 1 and Fig. 1b).

Although no known triplosensitive genes were initially identified within the 17p13.3–p13.2 duplication, precise mapping using OGM identified that this duplication also encompassed *PAFAH1B1*, with pTripto score = 0.998¹⁴ and ClinGen triplosensitivity [TS] score = 0¹⁵ (Supplementary Data 1). Duplications of *PAFAH1B1* have been linked to West syndrome²⁵, seizures, and a range of structural brain anomalies, including abnormalities of the corpus callosum as a part of 17p13.3 duplication syndrome²⁶. Thus, the duplicated *PAFAH1B1* may account for the hypoplasia of the corpus callosum that was observed in this patient. To confirm the origin of this unbalanced translocation, breakpoint-spanning PCR was

Table 1 | Summary of the phenotypes and CNVs of patients in this study

Case	Saphyr	Primary phenotype	Sex	XHMM		Key genes	Inheritance	Confirmation
				CNV	Cytoband			
1	109	Intellectual disability, microcephaly, seizures, palatopharyngeal incompetence	Male	Copy number loss	1q21.1-q21.2	-	Paternal	qPCR
2	110	Spinocerebellar degeneration	Male	Copy number loss	4pter-4p16.3	-	Assumed de novo	qPCR
3	111	Global developmental delay	Male	Copy number loss	8q23.3-q24.13	RAD21	Assumed de novo	qPCR
4	112	Congenital heart defects (atrial septal defect/ventricular septal defect), global developmental delay, postaxial polydactyly, dysmorphic features	Female	Copy number loss	22q11.21	TBX1	Assumed de novo	qPCR
5	115	Autism spectrum disorder, intellectual disability, epilepsy, and amblyopia	Female	Copy number loss	15q13.2-q13.3	-	Assumed de novo	qPCR
6	116	Hereditary spastic paraplegia, autism	Male	Copy number gain	21q22.2-q22.3	-	Assumed de novo	qPCR
7	117	Coffin-Siris syndrome (growth retardation, coarse facial features, generalized hirsutism, muscular hypotonia, visual impairment, hypoplastic left heart)	Male	Copy number gain	8p23.1	SOX7, GATA4	Assumed de novo	qPCR
8	118	Severe global developmental delay, short stature, microcephaly, epileptic spasms, West syndrome	Male	Copy number loss	7q11.22-q21.12	-	Assumed de novo	qPCR
9	119	Focal-onset seizure, cardiomyopathy, laryngomalacia, talipes equinovarus, low-set ears	Male	Copy number loss	6q12-q14.1	PHIP, KCNQ5, COL12A1	Assumed de novo	qPCR
10	120	Leukodystrophy, global developmental delay, short stature	Female	Copy number loss	17p13.3	YWHAE	Assumed de novo	qPCR
11	121	Cornelia de Lange syndrome, ichthyosis	Male	Copy number loss	Xp22.31	STS	Maternal	qPCR
12	122	Dyspnea, stridor, upper airway stenosis, anterior segment dysgenesis, cleft palate, anal atresia, talipes equinovarus, aortic root aneurysm, aortic regurgitation	Female	Copy number loss	2q32.1-q33.2	STAB2, COL3A1, BMPR2	Assumed de novo	qPCR
13	123	Seizures	Male	Copy number loss	16p13.11	-	Assumed de novo	qPCR
14	124	Sotos syndrome, retinal detachment, vascular remnant arising from the disc	Male	Copy number loss	19p13.2-p13.13	NFIX, CACNA1A	Assumed de novo	qPCR
15	125	West syndrome, hypoplasia of the corpus callosum, aortic regurgitation	Female	Copy number loss	1p36	GABRD	Assumed de novo	qPCR
16	133	Global developmental delay	Male	Copy number loss	8q21.13	ZFXH4	Assumed de novo	qPCR
17	134	Klippel-Feil syndrome, cleft lip and cleft palate, dysphagia, hearing impairment, esotropia	Male	Copy number gain	14q32.2-q32.33	-	Assumed de novo	qPCR
18	135	Early infantile epileptic encephalopathy with burst-suppression	Female	Copy number gain	2q24.3	SCN1A, SCN2A, SCN3A, SCN7A, SCN9A	Assumed de novo	qPCR
19	136	Mild global developmental delay, laryngeal stridor, facial palsy, cervical dimple, pelvic kidney	Male	Copy number loss	19q13.42-q13.43	U2AF2	Assumed de novo	qPCR

Table 1 (continued) | Summary of the phenotypes and CNVs of patients in this study

Case	Saphyr	Primary phenotype	Sex	XHMM	CNV		Key genes	Inheritance	Confirmation
					Cytoband	Copy number			
20	137	PCDH19-related epilepsy syndrome	Female			Xq22.1-q22.3	PCDH19	Assumed de novo	qPCR
21	138	Spinal and bulbar muscular atrophy	Male			Xq22.2	PLP1	Assumed de novo	qPCR
22	139	Cornelia de Lange syndrome	Male			18p11.32-p11.22	-	Assumed de novo	qPCR
						18q21.31-q21.32	-	Assumed de novo	qPCR
						18q21.33-q23	TSHZ1, NFATC1	Assumed de novo	qPCR
23	140	Cornelia de Lange syndrome	Female			9p24.3-p22.3	-	Assumed de novo	qPCR
						9p22.3	-	Assumed de novo	qPCR
24	141	Multiple congenital malformations, global developmental delay, microcephaly, cerebral palsy, abnormal facial shape, abnormal heart morphology	Female			19q13.33	-	Assumed de novo	qPCR
25	142	Global developmental delay, abnormal facial shape, autism, failure to thrive, short stature, seizures	Female			1p22.1-p13.2	KCMA2	Assumed de novo	qPCR
26	143	Seizures, global developmental delay, abnormal facial shape, cerebral hemorrhage, laryngomalacia, low-set ears, atresia of the external auditory canal, impaired ocular abduction	Male			18q22.3-q23	TSHZ1	Assumed de novo	qPCR
						5p15.33	-	Assumed de novo	qPCR
27	144	Omphalocele, mental deterioration, seizures, abnormality of the face, high palate	Female			9p24.3-p24.1	-	Assumed de novo	qPCR
						9p23-p22.1	-	Assumed de novo	qPCR
						9p22.1-p13.1	-	Assumed de novo	qPCR
28	145	Severe global developmental delay, abnormally large globus pallidus, brain atrophy	Male			2q36.3-q37.3	-	Assumed de novo	qPCR
29	147	Ehlers-Danlos syndrome, joint laxity, arthritis, mild intellectual disability	Female			12q21.1-q21.2	-	Assumed de novo	qPCR
30	14	Vitreoretinal degeneration, short stature, growth delay, abnormal facial shape, calcified skin lesions, ventricular septal defect	Male			22q12.1-q12.2	-	Assumed de novo	qPCR
						22q12.2	-	Assumed de novo	qPCR
						22q12.3	-	Assumed de novo	qPCR
						22q13.1 (3 SVs)	-	Assumed de novo	qPCR
						22q13.31	-	Assumed de novo	qPCR
						22q13.33	-	Assumed de novo	qPCR

CNV/CN variant, qPCR quantitative polymerase chain reaction, SVs structural variants, XHMM whole exome sequencing data analyzed using the eXome hidden Markov model.

Table 2 | Overview of structural variants identified by OGM in this study

Case	Saphyr	Chr	XHMM (GRCh38/hg38)		OGM (GRCh38/hg38)		Breakpoint (GRCh38/hg38)	Long-read sequencing	Differences (bp)	Figure	Gene disruption (gene disruption identified by breakpoints are in bold, genes already identified in ES are underlined)		
			Genomic positions	Size (bp)	Cytoband	CNV						Structural variant call	Size (bp)
1	109	1	146066406-148125101	2058695	1q21.1-q21.2	Copy number loss	Deletion	2542691	-	146054918 146180245	85082-125327	S4a	Unknown
2	110	4	60107-1230212	1170105	4pter-4p16.3	Copy number loss	Deletion	1196612	-	12985	12985	S3	-
3	111	8	116644905-125436571	8791666	8q23.3-q24.13	Copy number loss	Deletion	10034669	-	1209596 1225480	15684	S2a	CTBP1
4	112	22	19038764-21061835	2023071	22q11.21	Copy number loss	Deletion	2323551	-	116492663 126527331	1491 20733	S4b	Unknown
5	115	15	30140741-32401575	2260834	15q13.2-q13.3	Copy number loss	Deletion	1989617	-	18876658 18859569	85307-95306	S4c	Unknown
6	116	21	39293710-41466253	2172543	21q22.2-q22.3	Copy number gain	Duplication	2195417	-	30434571 30410000	24571-44571	S4c	Unknown
7	117	8	10525466-11996371	1470905	8p23.1	Copy number gain	Direct duplication	2167127	-	32424187 32431328-32451372	7141-27185	S8a-c	Unknown
8	118	7	72106142-85081670	12975528	7q11.22-q21.12	Copy number loss	Deletion	15269269	-	41463212 39267796	4618	S2b	TMPRSS2
9	119	6	68638575-80343904	11705329	6q12-q14.1	Copy number loss	Deletion	14804813	-	45321003 45271632	11420 172269	S2c	-
10	120	17	979326-2407279	1427953	17p13.3	Copy number loss	Deletion	1448734	-	46697230	12753	S2d	COL18A1
11	121	X	7050340-8170199	1119859	Xp22.31	Copy number loss	Deletion	1697429	-	12328822 10361696	43268	S2e	-
12	122	2	185738462-202560043	16621581	2q32.1-q33.2	Copy number loss	Deletion	18069129	SUMO1	71939977 87210944	3836	S2f	MSRA
13	123	16	15032824-16221753	1188929	16p13.11	Copy number loss	Deletion	1691631	-	65912218 80717030	34153	S4d	Unknown
14	124	19	11505424-13307751	1802327	19p13.2-p13.13	Copy number loss	Deletion	1998070	-	80652635 943147	64395	S2g	Unknown
										957506 2408637	14359	S2g	ZNF653
										241268	2731	S2g	CAGNA1A
										6466265 8163693	5381	S2g	
										185304800 203373928	2037	S2g	
										14681564 14954790	2292	S2g	
										16373394	18758-20428	S2g	
										11451325 13449394	47567	S2g	
										13309177	140217	S2g	

Table 2 (continued) | Overview of structural variants identified by OGM in this study

Case	Saphyr	Chr	XHMM (GRCh38/hg38)			OGM (GRCh38/hg38)			Breakpoint (GRCh38/hg38)		Gene disruption (gene disruption identified by breakpoints are in bold, genes already identified in ES are underlined)			
			Genomic positions	Size (bp)	Cytoband	CNV	Structural variant call	Size (bp)	Additional findings (aberrant gene, etc.)	OGM		Long-read sequencing	Differences (bp)	Figure
15	125	17	156112-3801270	3645158	17p13.3-p13.2	Copy number gain	Translocation	-	PAPAHTB1	3831287	3847652	16365	1a-f	-
		1	925786-3691130	2765344	1p36	Copy number loss	-	-	-	3706717	3699571	7146	-	TP73
16	133	8	76704036-81283570	4579534	8q21.13	Copy number loss	Deletion	4903335	-	76590158	76598298	8140	S2h	-
						Copy number loss	-	-	-	81493492	81428940	64552	-	-
17	134	21	40042270-46664474	6622204	21q22.2-q22.3	Copy number loss	Translocation	-	-	40024754	40025712	958	S6	DSCAM
		14	95686367-106874462	11188115	14q32.2-q32.33	Copy number gain	-	-	-	95650739	95642831	7908	-	TC16
18	135	2	165090176-166441741	1351565	2q24.3	Copy number gain	Direct duplication	1487936	-	166464506	166468854	2348	1g-i	SCN7A
						Copy number gain	-	-	-	164976570	164975946	624	-	-
19	136	19	54461133-55932399	1471266	19q13.42-q13.43	Copy number loss	Deletion	1486096	-	54453746	54461012	7266	S2i	LENG8
						Copy number loss	-	-	-	55957898	55947282	10616	-	-
20	137	X	100298247-104040976	3744729	Xq22.1-q22.3	Copy number gain	Inverted duplication	-	-	108269338	105883789	385549	S10	NRK
						Copy number gain	Intrachromosomal fusion	-	-	97354335	97351663	2672	-	DIAPH2
						Copy number gain	-	-	-	108285726	106277461	8265	-	-
21	138	X	103629963-104040976	411013	Xq22.2	Copy number gain	Inverted duplication	-	-	103986681	103853617	133064	S11	-
						Copy number gain	Intrachromosomal fusion	-	-	103820569	103808152	12417	-	-
						Copy number gain	-	-	-	29092453	29087541	4912	-	IL1RAPL1
						Copy number gain	Intrachromosomal fusion	-	-	29323094	29331011	7917	-	-
						Copy number gain	-	-	-	103619503	103618581	922	-	-
22	139	18	47395-10785024	10737629	18p11.32-p11.22	Copy number gain	Intrachromosomal fusion	-	-	10780464	10785166	4702	3	PIEZO2
			58578800-60643598	2064798	18q21.31-q21.32	Copy number loss	Inverted duplication	-	-	56575606	58557581	18025	-	ALPK2
						Copy number loss	-	-	-	61372583	61373711	1128	-	CDH20
23	140	9	116864-14313625	14196761	9p24.3-p22.3	Copy number loss	Inverted duplication	-	NFIB	14373459	14369157	4302	2d-f	NFIB-AS1
			14313625-15211453	897828	9p22.3	Copy number gain	-	-	-	14642814	14397270	-	-	-
24	141	19	49145811-49762265	616454	19q13.33	Copy number gain	Direct duplication	552443	-	49740665	49753101	12436	S8d-f	TSKS
						Copy number gain	-	-	-	49160094	49165336	5242	-	TRPM4
25	142	1	93110129-111324983	18214854	1p22.1-p13.2	Copy number loss	Deletion	18220941	COL11A1	93100823	93106498	5675	S2j	MTF2
						Copy number loss	-	-	-	111321763	111329159	7396	-	-
26	143	5	92131-3601140	3509009	5p15.33	Copy number gain	Translocation	-	-	4619564	4621199	1635	S7	-
						Copy number gain	-	-	-	72972097	72972554	457	-	-

Table 2 (continued) | Overview of structural variants identified by OGM in this study

Case	Saphyr	Chr	XHMM (GRCh38/hg38)		OGM (GRCh38/hg38)		Structural variant call	Size (bp)	Additional findings (aberrant gene, etc.)	OGM	Long-read sequencing	Differences (bp)	Figure	Gene disruption (gene disruption identified by breakpoints are in bold, genes already identified in ES are underlined)		
			Genomic positions	Size (bp)	Cytoband	CNV										
27	144	9	116689-8733943	8617244	9p24.3-p24.1			Inverted duplication		18557461	18553873	3588	S9	ADAM/TS1		
			12693896-18533392	5839496	9p23-p22.1											
			18573929-38068745	19494816	9p22.1-p13.1						18827498	18559755				
28	145	2	226731317-242160331	15429014	2q36.3-q37.3			Duplication		225672439	Unknown	Unknown	S5	Unknown		
										238461378	Unknown	Unknown			Unknown	
29	147	12	72542190-77064735	4522545	12q21.1-q21.2			Inversion	Chromoplexy-like event t(12;14)	70920270	70917324	2946	4	PTPARR		
										72506564	72511491	4927		TRHDE		
									Intrachromosomal fusion		70920270	70917440	2830		PTPARR	
									Translocation		78307029	78304150	2879		-	
									Translocation		78386157	78382960	3197		-	
									Inversion		Chr14: 86401214	Chr14: 86404434	3220		-	
										Chr14: 84178337	Chr14: 84172309	6028				
									Translocation		Chr14: 84132163	Chr14: 84135076	2913			
										78307029	78382967	75938				
									Intrachromosomal fusion		78058671	78304148	245477			
						Inversion		77959302	77696122	263180						
								77959302	78304147	344845						
								257744								
									77701558	77696122	5436					

Chr chromosome, CNV CN variant, ES exome sequencing, OGM optical genome mapping, XHMM whole exome sequencing data analyzed using the exome hidden Markov model.

Table 3 | Overview of chromothripsis in Case 30

Case	Saphyr	Chr	OGM		Map name	Breakpoint (GRCh38/hg38)			Gene disruption	
			Structural variant call	Size (bp)		OGM	Long-read sequencing	Differences (bp)		
30	14	22	Intrachromosomal fusion	-	1	A	21590708	21592378	1670	<i>UBE2L3</i>
						B	37788649	37786951	1698	<i>PLA2G6</i>
			Deletion	670172	C	38118769	38122247	3478	-	
						38794317	38792466	1851	<i>DNAL4</i>	
			Intrachromosomal fusion	-	2	D	45862244	45868639	6395	-
						E	31676310	31673907	2403	-
			-	-	3		35489164	35725198	236034	<i>APOL5</i>
			Inversion	580782	F	39249596	39258460	8864	-	
			Intrachromosomal fusion	-	4		37229398	37216272	13126	-
					G	29017395	29015909	1486	<i>ZNRF3</i>	
			Inverted duplication	270381	5		31663482	31673909	10427	-
			Intrachromosomal fusion	-	6	G'	31215599	31214183	1416	<i>LIMK2</i>
		H	50594574	50616489	21915	-				
Intrachromosomal fusion	-	7		45261412	45243553	17859	-			
		I	21221283	21202844	18439	-				
Intrachromosomal fusion	-	8		29430946	29435389	4443	-			
		J	35489164	35486947	2217	-				

Chr chromosome, OGM optical genome mapping.

performed using long-read sequencing data (primers listed in Table S5); this confirmed its paternal origin (Fig. 1d, e). Sanger sequencing of the translocation breakpoints further identified a 3-bp insertion on der(1) and a 2-bp insertion on der(17) (Fig. 1f). Although *TP73* was disrupted by a translocation (Table 2), none of the disrupted genes were associated with the clinical features observed in Case 15.

The same analyses using OGM were also performed for Cases 17 (Fig. S6) and 26 (Fig. S7); however, no additional gene aberrations relevant to the phenotype were detected using OGM. Cases 17 and 26 are described in the Supplementary information.

Direct duplications (Cases 7, 18, and 24)

OGM identified direct duplications in three cases (Fig. 1g and Fig. S8a, d), 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 (Fig. 1h and Fig. S8b, e). This approach revealed no additional aberrant genes or gene disruption relevant to the phenotype. Cases 7, 18, and 24 are described in detail in the Supplementary information.

Duplication with inversion (Case 6)

In Case 6, we newly identified a 6.2-Mb inversion at 21q22.2–q22.3 and a 218-kb duplication at 21q22.2 using OGM with AS. The CN gain had previously been detected using ES (Tables 2 and S4). OGM analysis revealed that the inversion was located immediately downstream of the duplication (Fig. 2a), thus highlighting a limitation of ES, which typically fails to detect breakpoints outside coding regions. Furthermore, OGM analysis revealed that a CN loss previously identified by ES was the deletion from 21q22.3 to 21qter, located downstream of this inversion, with the 3' end of the inversion corresponding to the end of the long arm of chromosome 21 (Fig. 2a).

Long-read sequencing data further illustrated these SVs through dot-plot reconstruction (Fig. 2b). The CN loss, previously identified by ES, encompassed *COL6A1* (pHaplo score = 0.942¹⁴, HI score = not registered¹⁵), whose haploinsufficiency causes Bethlem myopathy 1A (OMIM# 158810) or Ullrich congenital muscular dystrophy 1A (OMIM# 254090), and *COL6A2* (pHaplo score = 0.593¹⁴, HI score = not registered¹⁵), whose

deletion results in Bethlem myopathy 1B (OMIM# 620725) or Ullrich congenital muscular dystrophy 1B (OMIM# 620727). In addition, the deletion of *DIP2A* (pHaplo score = 0.916¹⁴, HI score = 1¹⁵) is associated with neurodevelopmental disorder²⁷ and autism²⁸ (Table 1 and S3). However, no additional aberrant haploinsufficiency or triplosensitive genes associated with hereditary spastic paraplegia or autism were identified within these regions using OGM (Supplementary Data 1). Although *TMPRSS2* and *COL18A1* were disrupted by the SVs (Table 2), neither has been associated with hereditary spastic paraplegia or autism.

Inverted duplications (Cases 23 and 27)

OGM identified inverted duplications in Cases 23 and 27, involving complex rearrangements that combined deletions and duplications (Fig. 2d and Fig. S9a). Long-read sequencing successfully confirmed the breakpoints in both cases (Fig. 2e and Fig. S9b).

In Case 23, who was clinically suspected to have Cornelia de Lange syndrome, OGM revealed a 14.3-Mb deletion and an 851-kb duplication at 9p22.3, resulting in inverted duplication (Table 2 and Table S4). Notably, *NFIB* (pHaplo score = 0.999¹⁴, HI score = 3¹⁵) was disrupted at intron 1 by the SV (Fig. 2d, Table 2, and Supplementary Data 1). Haploinsufficiency of *NFIB* causes macrocephaly with impaired intellectual development (OMIM# 618286) as well as intellectual disability and dysmorphic features^{29,30}. These clinical features are consistent with those observed in Case 23, including intellectual disability and dysmorphic features.

Case 27 involved an inverted duplication rearrangement detected by OGM with AS (Table 2 and Supplementary Data 1), similar to a previously reported case³¹. Within the 18.5-Mb deletion at 9p22.1–24.3, both *NFIB* and *SMARCA2* were identified (Fig. S9b and Supplementary Data 1). As previously mentioned, *NFIB* haploinsufficiency can contribute to intellectual impairment, dysmorphic facial features, and a high palate (Table 1). *SMARCA2* (pHaplo score = 0.999¹⁴, HI score = 0¹⁵) is associated with blepharophimosis-impaired intellectual development syndrome (OMIM# 619293) and Nicolaides–Baraitser syndrome (OMIM# 601358) in an autosomal dominant manner, and its intragenic deletions reportedly cause intellectual disability, seizures, and dysmorphic features^{32,33}. Although *SMARCA2* deletion has not been clearly established as a distinct deletion

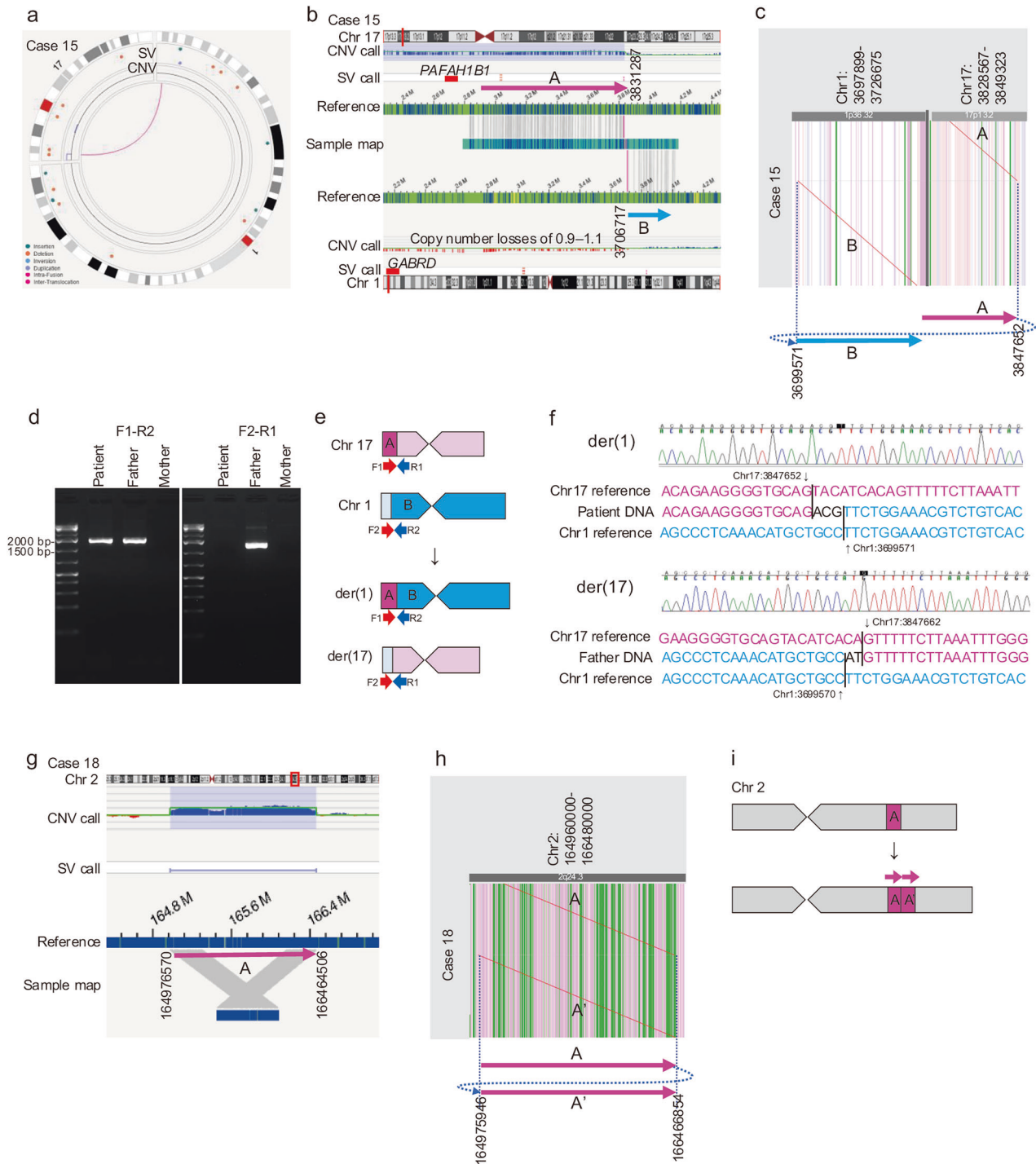
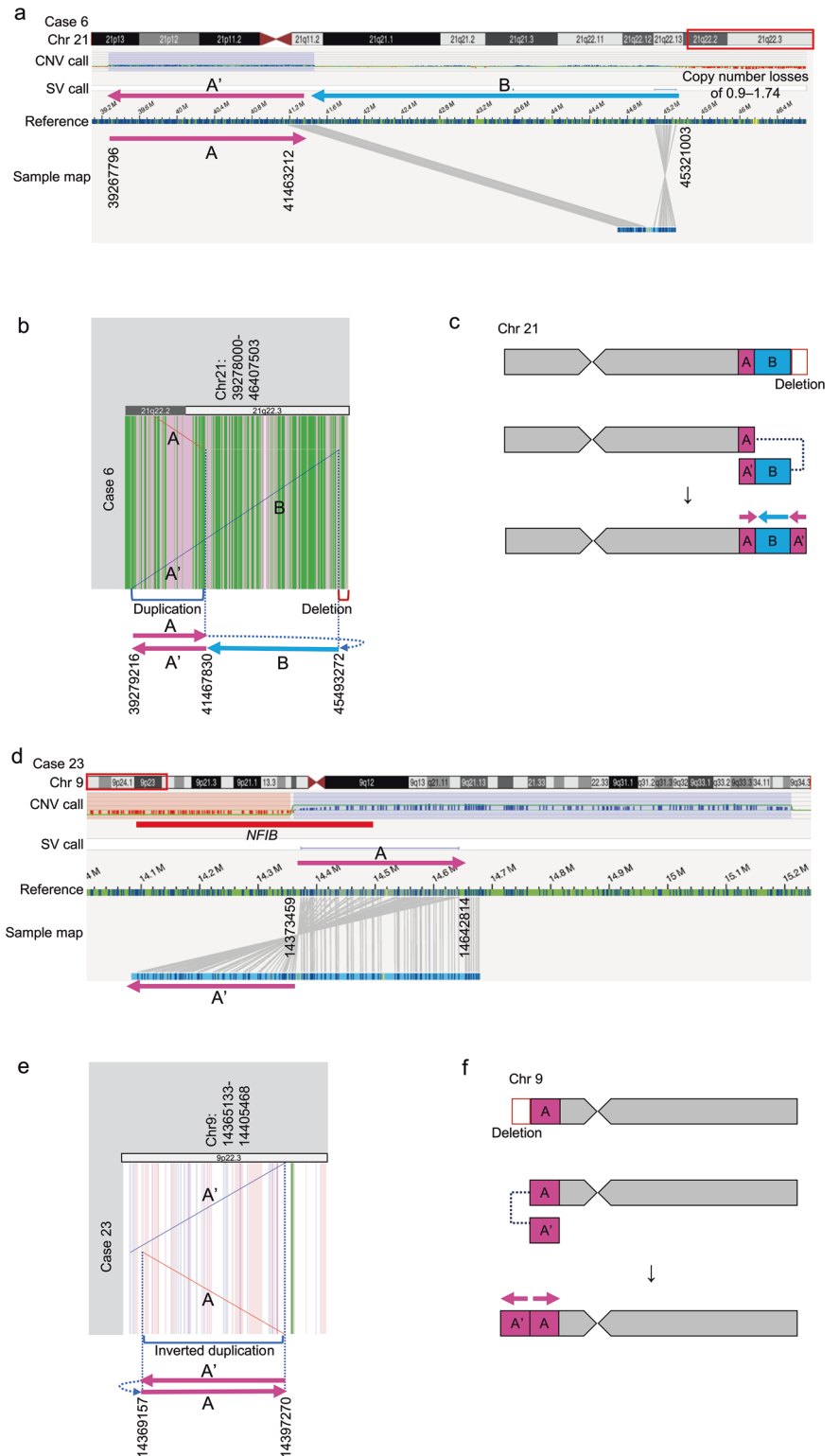


Fig. 1 | Results of OGM and long-read sequencing in Cases 15 and 18. a–f and g–i are generated from Cases 15 and 18, respectively. a Circos plot of OGM shows unbalanced translocation involving Chr 17 and 1. b OGM results in the genome browser. In Chr 17, a large duplication is visualized as a blue bar in the CNV call track; *PFAH1B1* was located in the duplication. Deletion of Chr 1 should have been shown as red bars in the CNV call track, but is not shown in this case despite CN losses of 0.9–1.1; *GABRD* was located within the deletion. Translocation breakpoints are shown in purple on the sample map between two matching chromosomes. The numbers are the breakpoint positions. Arrows A and B indicate that these maps are the forward orientations to the reference genome. c Dot-plot of the consensus sequence from long-read sequencing data. Red lines show the forward orientation to the reference sequences. 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.

d Electrophoresis image of polymerase chain reaction products. e Scheme of translocation and primers. f Upper and lower diagrams show Sanger sequencing electropherograms of der(1) and der(17), respectively, at breakpoints. Rearrangement junction sequence (middle line) and matching reference sequence (top and bottom lines) are shown in pink and sky blue. der(1) includes a 3-bp insertion and der(17) includes a 2-bp insertion, as shown in black. g Large duplication is visualized as blue bars in the CNV call track and a blue line in the SV call track in the genome browser. h Dot-plot of the consensus sequence containing the direct duplication region created from long-read sequencing data using dnarrange. i Scheme of direct duplication. ISCN nomenclature in the two cases is described in Table S4. Circos plots and 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, SV structural variant.

Fig. 2 | Results of OGM and long-read sequencing in Cases 6 and 23. a–c and d–f are generated from Cases 6 and 23, respectively. a OGM results in the genome browser. Large duplication is visualized as a blue bar in the CNV call track and inversion is visualized at the end of arrow B. Downstream of the inversion (21q22.3–qter) shows CN losses of 0.9–1.74 in the genome browser, and the 3' end of the inversion is the end of the long arm of Chr 21. b Dot-plot of duplication and inversion regions created from long-read sequencing data using dnarrange. c Scheme of SVs. d OGM results in the genome browser. Large deletion and duplication are visualized as red and blue bars, respectively, in the CNV call track. Inverted duplication is indicated by a blue line in the SV call track and is shown in the sample map below. *NFIB* is located at the boundary between deletion and duplication. e Dot-plot of the consensus sequence from long-read sequencing data. f Scheme of the inverted duplication. ISCN nomenclature of 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, SV structural variant.



syndrome, it has been previously reported with limited clinical information³⁴. By contrast, the adjacent 19.7-Mb duplication did not contain any known genes associated with human diseases (Supplementary Data 1). Notably, although *NFIB-AS1* and *ADAMTSL1* were disrupted by inverted duplication in each of Cases 23 and 27 (Table 2), neither gene was associated with the clinical features observed in these cases.

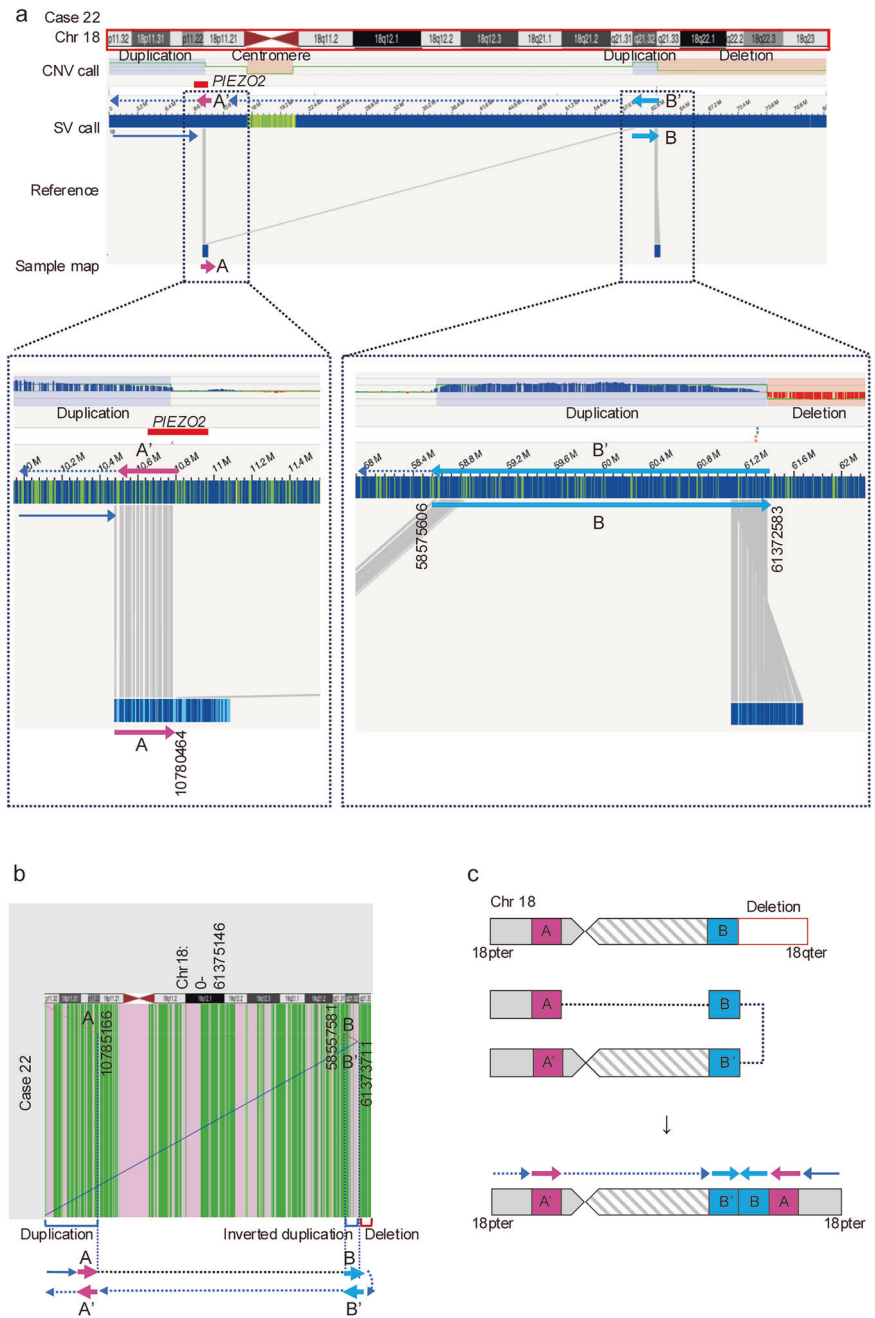
Except in rare cases where one breakpoint coincides with an inverted duplication and sequencing reads spanning two different exons are present, ES is generally incapable of identifying inverted duplications and their

breakpoints, because orientation information is not available. By contrast, the present approach yielded meaningful findings that were linked to clinical symptoms in these two cases.

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

Fig. 3 | Results of OGM and long-read sequencing in Case 22. **a** The entire SV of Chr 18 (upper diagram) and its enlargement within the dotted square (lower diagram) in the genome browser. Large duplications and deletions are visualized as blue and red squares, respectively, in the CNV call track, and an inverted duplication is visualized as a blue line in the SV call track of the genome browser. Arrows A and B indicate intrachromosomal fusion, and arrows B and B' indicate inverted duplication. **b** Dot-plot of the consensus sequence containing the 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 browser was 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.



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, Table 2, and Table S4). In Case 21, an inverted duplication and two intrachromosomal fusions were identified (Fig. S11, Table 2, and Table S4). Case 22 exhibited an inverted duplication and an intrachromosomal fusion (Fig. 3, Table 2, and Table 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. 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). Cases 20 and 21 are described in detail in the Supplementary information.

By contrast, Case 22 exhibited a more complex pattern involving a 10.8-Mb duplication at 18p11.32–p11.22, a 2.8-Mb duplication at

18q21.31–21.33, and an 18.9-Mb deletion at 18q21.33–q23 (Fig. 3a). Within the deleted 18q21.33–q23 region, ES identified *TSHZ1* and *NFATC1*, which have been associated with auricular deformities, hearing impairment^{35,36}, and congenital heart disease^{37,38} (Table 1 and Supplementary Data 1). Nevertheless, in the newly characterized SVs using OGM and long-read sequencing, no additional candidate pathogenic genes were identified that might account for the patient's clinical presentation, which included features potentially resembling Cornelia de Lange syndrome (Supplementary Data 1). Additionally, *PIEZO2*, which is associated with Marden–Walker syndrome (OMIM# 248700) through an autosomal dominant mechanism, was disrupted by SVs in Case 22 (Table 2). *PIEZO2* has also been reported as associated with Cornelia de Lange syndrome³⁹, and may be involved in the phenotype observed in this case.

Notably, the intrachromosomal fusions in these cases were positioned adjacent to their respective CN gains—a feature that was missed by ES because of the copy-neutral nature of these SVs. Furthermore, three cases

exhibited complex multi-breakpoint events, meaning that it was difficult to understand the structural reconstruction using long-read sequencing alone; OGM, which enables the visualization and overview of complex SVs, was therefore extremely helpful.

Chromoplexy-like event in Case 29

In Case 29, who presented with Ehlers–Danlos syndrome, joint laxity, arthritis, and mild intellectual disability, ES previously identified a 5.2-Mb CN loss at 12q21.1–q21.2. OGM further revealed an intrachromosomal insertion event and inversion within chromosome 12, an inversion within chromosome 14, and a translocation between chromosomes 12 and 14 (Fig. 4a, b). Long-read sequencing precisely determined all of the breakpoints (Fig. 4c, d), confirming a chromoplexy-like event between chromosomes 12 and 14 (Fig. 4l and Table S4). Chromoplexy refers to a large-scale and complex formation of chain rearrangements that affects multiple chromosomes. This phenomenon is characterized by the interdependent occurrence of multiple inter- and intrachromosomal translocations and deletions⁴⁰. Chromoplexy leads to the generation of derivative chromosomes that present little or no CN alterations, and Case 29, with CN loss, presented a chromoplexy-like event. Conversely, chromothripsis is defined as a relatively clustered chromosomal shattering followed by a random restitching of chromosomal fragments, which results in the formation of complex genomic rearrangements⁴⁰. Chromothripsis was observed in Case 30.

The chromoplexy-like event involved chromosomes 12 and 14, with reciprocal inverted insertions, partial deletions, and inversion of multiple segments. Despite the complex rearrangement, no clearly haploinsufficient gene associated with the patient's phenotype was identified within the deleted region (Supplementary Data 1). *PTPRR* and *TRHDE* were disrupted by a reciprocal translocation (Table 2), but neither was associated with the clinical features observed in Case 29. The derivative chromosome 12 harbored the 5.2-Mb deletion at 12q21.1–q21.2, which was joined to chromosome 14 without a CN change. Therefore, although the CN loss was detectable by ES-based CNV analysis, the reciprocal translocation itself was not identified. Because the deleted region did not encompass any previously known disease-associated genes that would explain the clinical features, this case clearly demonstrates the limitations of ES for detecting complex SVs and highlights the added value of OGM and long-read sequencing in cases with no apparent clinical features.

Chromothripsis (Case 30)

In Case 30, who presented with vitreoretinal degeneration, short stature, growth delay, abnormal facial shape, calcified skin lesions, and ventricular septal defect, ES detected eight CN gains in chromosome 22, implying chromothripsis (Fig. 5a). The chromothripsis is thought to arise when one homolog of chromosome 22 undergoes aberrant segregation due to anaphase-lagging and becomes encapsulated within a micronucleus. Extensive DNA damage incurred within the micronucleus is followed by erroneous reassembly upon reincorporation into the primary nucleus, resulting in abnormal chromosomal rearrangements⁴¹. Based on ES data, the CNV sizes of the eight CN gains at 22q12.1–q13.33 ranged from 107–437 kb. OGM analysis revealed six intrachromosomal fusions, one deletion, one inversion, and one inverted duplication with ogm (22q)cth at 22q11.21–q13.33 (Fig. 5b, Fig. S11, and Table 3). Although we attempted to reconstruct the full spectrum of SVs based on the OGM results, the SVs identified by OGM did not completely match those detected by dnarrange using long-read sequencing data. We therefore manually evaluated the breakpoint sequences using long-read sequencing data, considering the orientations of the SVs and CNVs called using OGM (Fig. 5c). The CNV sizes of the eight duplications in a dot-plot of long-read sequencing data were 390 kb at 22q11.21, 419 kb at 22q12.1–q12.2, 460 kb at 22q12.2, 238 kb at 22q12.3, 884 kb at 22q13.1, 335 kb at 22q13.1, 466 kb at 22q13.1, and 625 kb at 22q13.31; all eight of these were revealed to be three CN gains. These regions included 6, 9, 7, 4, 31, 15, 14, and 7 genes, respectively (Supplementary Data 1); however, none of the genes were associated with

known phenotypes. In addition, six genes were disrupted by the chromothripsis (Table 3), but none of these were associated with the clinical features observed in Case 30. We also investigated the pathogenicity of triplosensitive regions within all eight duplications in case 30 by using ClinGen. Duplications of the 22q11.2 region (TS score = 1¹⁵) have been reported as causing developmental delay, dysmorphism, and multiple congenital abnormalities⁴², suggesting they may cause the growth delay and abnormal facial shape observed in case 30. However, no cause for other phenotypes in this case could be identified. Ultimately, the combined analysis of OGM and targeted long-read sequencing provided a comprehensive and accurate identification of the chromothripsis (Table S4).

Discussion

We identified novel SVs using OGM with AS in 14 of 30 cases (Cases 6, 7, 15, 17, 18, 20–24, 26, 27, 29, and 30) in whom CNVs had previously been detected using ES. These SVs were undetectable by ES alone because of unstable genomic coverage, limited fragment size, and analytical limitations. ES typically misses SVs with breakpoints outside coding regions and cannot resolve the configuration of CN gains or unbalanced translocations inferred from telomeric CNVs. By contrast, OGM uses ultra-high molecular weight DNA and genome-wide mapping to detect large and complex SVs—including inversions, inverted duplications, and intrachromosomal fusions—outside ES target regions (Cases 6 and 20–22) and within repetitive regions (Case 26). Importantly, OGM with AS revealed disruptions (Cases 2 and 22) and CN changes (Cases 12, 15, 23, 25, and 27) in genes that are likely associated with clinical features (Table 2 and Supplementary Data 1) (7/30 cases, 23.3%), clearly demonstrating the utility of OGM with AS following CNV detection.

Most novel SVs were observed in cases involving CN gains (Cases 6, 7, 15, 17, 18, 20–24, 26, 27, and 30), with only one additional SV found in a case of CN loss (Case 29). Notably, in cases in which ES detected both CN gain and CN loss, OGM clarified the underlying events as unbalanced translocations (Cases 15, 17, and 26) or inverted duplications (Cases 23 and 27), which were undetected by ES. These findings suggest that such CNV patterns should raise a suspicion of hidden SVs, warranting further investigation with OGM.

Compared with ES, OGM achieved a superior detection of balanced (copy neutral) and complex (inversions, intrachromosomal fusions, chromoplexy, or chromothripsis) SVs that lacked CN changes, whereas ES remained sensitive to CNVs involving coding regions. Long-read sequencing complemented OGM by defining nucleotide-level breakpoints, although its limited read length sometimes prevented full structural reconstruction. Thus, OGM provides a genome-wide structural overview, whereas long-read sequencing ensures breakpoint precision; together, they offer comprehensive SV characterization.

OGM also enables CNV detection through molecule coverage-based analysis, to identify CNVs larger than 500 bp⁴³. In the present study, OGM detected CNVs ranging from 238 kb to 18.9 Mb. Although pathogenic genes were identified within these CNVs in six cases (Cases 12, 15, 23, 25, and 27), they had already been included within regions detected by ES, suggesting that ES can capture certain pathogenic CNVs if they are covered by baits.

Several factors may explain why some haploinsufficient genes were not captured by ES. Large CNVs spanning multiple genes can dilute read-depth changes, and capture efficiency varies across exons because of GC-rich or probe hybridization biases. Although the affected genes are generally covered by baits, exonic read-depth variation or pipeline filtering thresholds can lead to missed calls. In our laboratory, CNV detection was performed using an XHMM-based pipeline that was optimized for batch- and sex-specific analysis to minimize sequencing bias; however, this process was not perfect^{2,44}.

ES was performed between 2015 and 2022 using standardized pipelines from our laboratory. Although the retrospective visual inspection of BAM alignments was conducted for selected loci, no clear breakpoint-supporting reads were observed. The undetected SVs are therefore likely attributable to the limitations of read depth-based algorithms rather than sequencing

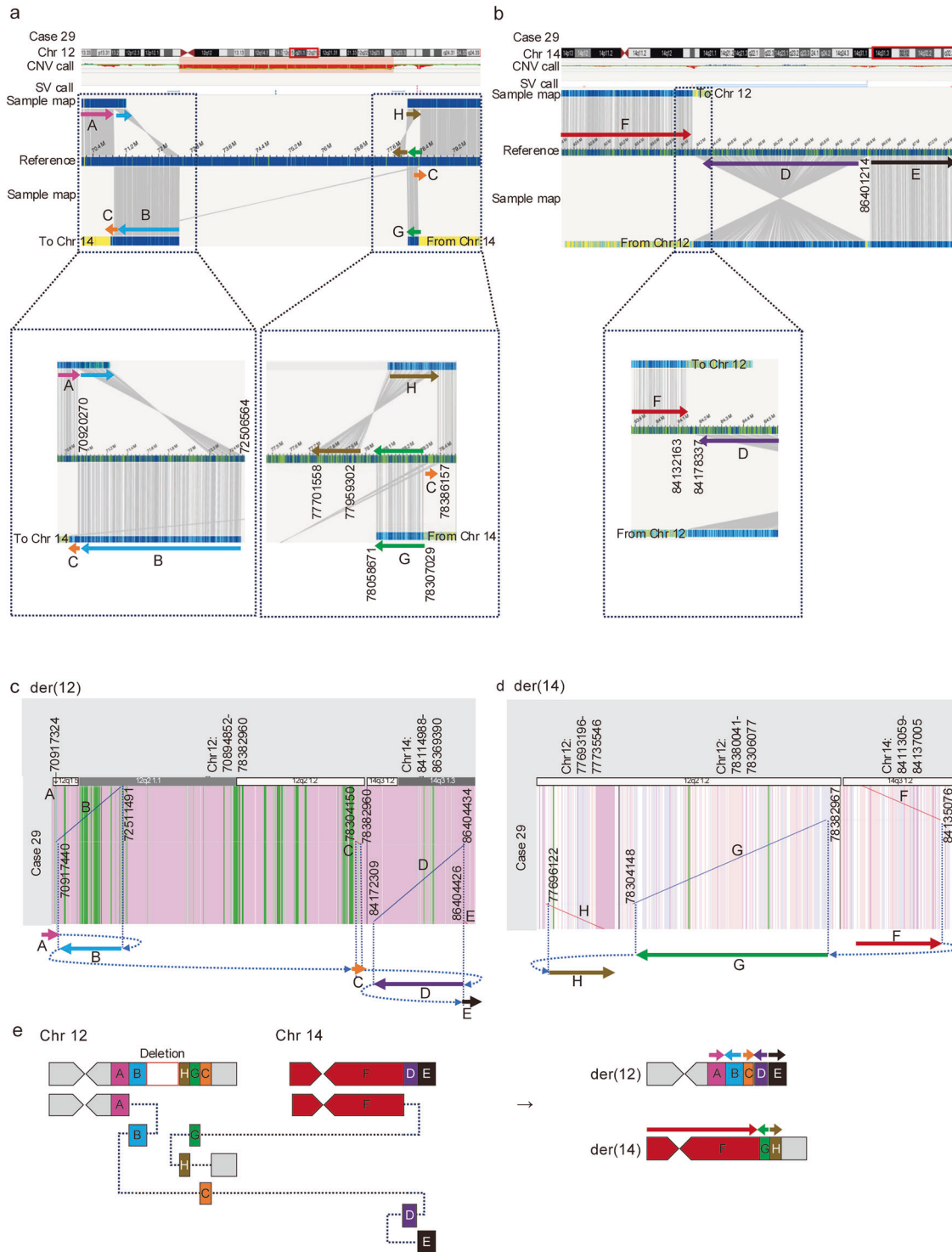
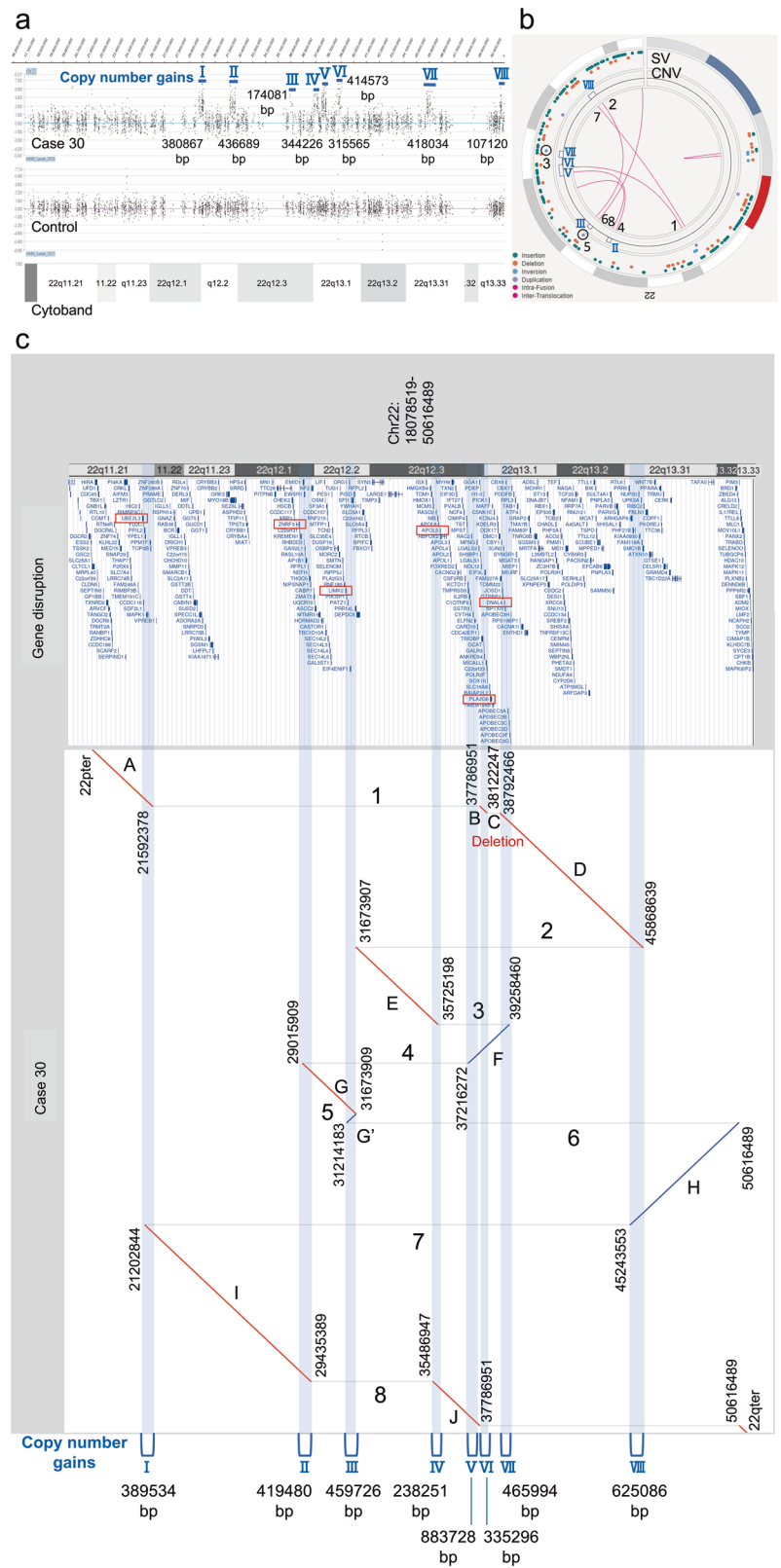


Fig. 4 | Results of OGM and long-read sequencing in Case 29. a The entire SV of Chr 12 (upper diagram) and its enlargement within the dotted square (lower diagram) in the genome browser. Large deletion is visualized as a red bar in the CNV call track, and inversion is visualized as blue lines in the SV call track at both ends of the deletion. Arrow C' in the sample map is a fragment of arrow C in the reference. Arrow B indicates inversion and arrows B and C indicate an intrachromosomal insertion event. Yellow sample maps lead to maps of Chr14 (shown in Fig. 4b). **b** The entire SV of Chr14 (upper diagram) and its enlargement within the dotted square (lower diagram) in the genome browser. Arrow D indicates inversion, shown as blue

lines in the SV call track. Yellow sample maps lead to maps of Chr 12 (shown in Fig. 4a). **c, d** Dot plots of the consensus sequence in derivative Chr 12 (c) or 14 (d) created from long-read sequencing data using dnarrange. Arrows C–D and F–G indicate reciprocal translocation between Chr 12 and 14. **e** Scheme of reciprocal translocation. 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 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. 5 | Results of the CNV analysis using ES data, OGM, and long-read sequencing in Case 30.

a CNV analysis of XHMM using ES data detected eight CN gains in Chr 22, indicated by blue bars (I–VIII). The x-axis shows physical positions, and the y-axis shows z-scores. **b** Circos plot of Chr 22 in OGM. Intrachromosomal fusions are indicated by magenta lines with numbers 1, 2, 4, and 6–8, corresponding to Maps 1, 2, 4, and 6–8, respectively, in Fig. S12. An inversion is indicated as a light blue dot with the number 3 (corresponding to Map 3 in Fig. S12), and an inverted duplication is indicated as a purple dot with the number 5 (corresponding to Map 5 in Fig. S12) in the SV track. Blue lines in the CNV track indicate five duplications (II, III, V, VII, and VIII), one of which contains a different duplication (VI), making a total of six duplications that correspond to those in (a). Two duplications (I and IV) should have been originally shown as blue bars in the CNV call track, but these two duplications are not shown in this case despite CN gains of 2.58–2.94 and 2.36–2.62, respectively. **c** Dot-plot of chromothripsis manually connected from long-read sequencing data using dnarrange. This dot-plot was created by linking each dot-plot from Maps 1–8 of Fig. S12; the numbers (1–8) and letters (A–J) in the dot-plot correspond to the map numbers (Maps 1–8) and letters (A–J) in Fig. S12. The blue areas (I–VIII) show eight CN gains that correspond to those in (a). The sizes of the eight CN gains from the long-read sequencing data are shown below the dot-plot. The top of the dot-plot shows genes located in this region, and the six genes in red squares were disrupted by the chromothripsis. ISCN nomenclature is described in Table S4. Circos plot was 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, ES exome sequencing, ISCN international system for human cytogenomic nomenclature, OGM optical genome mapping, SV structural variant.



artifacts. Previous studies have revealed that incorporating one-end-anchored split-read mappings into ES analysis can reveal SVs that are otherwise missed by standard read-depth approaches⁴⁵. Furthermore, CNV and SV calling software increasingly combines multiple data signatures—including read-pair, split-read, read-depth, and assembly-based methods—for improved detection accuracy and breakpoint precision⁴⁶. Together, these

findings suggest that further improvements in ES pipelines, such as the integration of read-pair or split-read analysis, may enhance SV detection sensitivity.

OGM and long-read sequencing detect SVs through fundamentally different mechanisms. Although long-read sequencing provides single-nucleotide resolution and accurate breakpoint determination, it is

constrained by physical read lengths (10–100 kb), often requiring multiple reads to reconstruct complex SVs⁶. By contrast, OGM analyzes molecules spanning hundreds of kilobases, enabling the better characterization of biallelic CNVs and phasing of CN changes⁴⁷.

For complex rearrangements with multiple breakpoints—such as intrachromosomal insertion events and inverted duplications (Cases 20–22), chromoplexy-like events (Case 29), or chromothripsis (Case 30)—OGM was valuable for structural delineation, with long-read sequencing used for precise breakpoint resolution. In Case 30, discrepancies in breakpoint orientation and CN between OGM and long-read sequencing illustrated the difficulty of resolving such events with sequencing alone. The genome-wide view of OGM and its ability to detect both CNVs and SVs are particularly advantageous for analyzing chromoanagenesis, including chromothripsis⁴⁸.

In the present study, OGM was particularly advantageous for delineating complex multi-breakpoint events (Cases 20–22, 29, and 30) in which complete reconstruction was difficult to perform using long-read sequencing alone. However, for translocation, direct duplications, or inverted duplications, long-read sequencing alone may suffice. OGM and long-read sequencing can also be complementary: the former provides superior structural mapping, whereas the latter is better at refining breakpoints.

Compared with ES, which primarily targets coding regions and relies on read depth-based CNV detection, combined OGM and long-read sequencing offers genome-wide coverage and direct breakpoint visualization, thereby enabling the accurate identification of balanced or complex SVs that are invisible to ES. In contrast to short-read whole genome sequencing (srWGS), which can capture small CNVs and SNVs genome-wide but often struggles to resolve long-distance rearranged regions, the integration of OGM and long-read sequencing overcomes these limitations by providing both optical- and nucleotide-level mapping continuity.

All 14 SVs detected using OGM were validated with high accuracy. However, the resolution limit (~5 kb) of OGM means that smaller variants may be missed^{6,49}. Breakpoint positions between OGM and long-read sequencing differed by 0.5–385.5 kb (Tables 2 and 3), often because of unmapped regions in OGM data. Larger unmapped regions correlated with greater discrepancies. Conversely, smaller gaps allowed near-precise breakpoint localization.

In Case 26, in which all reads included repetitive sequences, OGM accurately detected SVs, whereas long-read sequencing yielded mismatched reads, failing to detect the true breakpoints. This highlights the strength of OGM in repetitive and complex regions, leveraging long-range DNA molecules without sequence-based alignment^{11,50}.

Despite its advantages, OGM cannot fully resolve SVs in centromeres and telomeres¹¹. Whole-arm rearrangements such as Robertsonian translocations or isodicentric chromosomes remain challenging because of repetitive elements and gaps in the reference genome^{3,43,48}. Additionally, telomere fusions may obscure insertion sites⁴³. However, complete chromosome assemblies (e.g., of the X chromosome) are becoming feasible via the integration of T2T reference genomes and OGM^{51,52}.

Recent advances in genomic technologies have exhibited a clear trajectory, from ES to srWGS and onward to long-read sequencing and OGM. For example, in a large cohort of 722 undiagnosed patients with inherited retinal disease, a comparison of ES and srWGS demonstrated that srWGS using SV scanning identified molecular diagnoses in 4.29% of cases⁵³, demonstrating an added—but modest—gain over ES. More strikingly, a recent long-read whole genome sequencing (lrWGS) study of 34 exome-negative autosomal recessive families achieved a 38% diagnostic yield⁵⁴, thus underscoring the major improvement in SV detection afforded by long reads. Our study built on this idea by integrating OGM and long-read sequencing, and demonstrated that this combined approach was able to resolve SVs—including balanced and complex rearrangements—that both ES and srWGS often miss. Specifically, in our cohort, the combined method delineated multi-breakpoint events that long-read sequencing alone may not fully reconstruct. Our findings indicate that integrating OGM and AS

after ES-based CNV discovery increases diagnostic resolution and clarifies complex rearrangements. This ES–OGM–long-read sequential approach refines SV interpretation and establishes a framework for integrating genomic and sequencing data in rare disease diagnostics.

The present study has some limitations. First, because all patients were referred from tertiary or specialized clinical centers, a potential bias toward severe or atypical cases cannot be excluded. Moreover, the small sample size (30 cases were selected according to the selection criteria from 281 cases with CNVs identified byXHMM analysis in ES, out of a total of 8622 cases with rare genetic disorders who underwent ES analysis in our laboratory between 2015 and 2022; please refer to the Methods for the selection criteria) and technical differences/improvements in ES collected over multiple years may affect the generalizability of our results. Further large-scale prospective studies are needed to evaluate the true diagnostic yield and reproducibility of the OGM and long-read sequencing pipeline. Second, although the combined OGM and long-read sequencing approach provides maximal structural resolution, its high cost currently limits its routine clinical application. Although 13 of 14 cases (excluding Case 30) may have been resolved by long-read sequencing alone, complete reconstruction using only long-read sequencing is likely to be considerably difficult, especially considering complex multi-breakpoint events such as those of Cases 20–22 and 29. A tiered strategy—for example, starting with ES, followed by OGM or long-read sequencing for unresolved or structurally complex cases—may allow rapid solutions to be achieved using a relatively cost-effective and diagnostically efficient workflow. Such a layered approach balances cost, resolution, and interpretability, thereby offering a practical framework for clinical implementation.

Collectively, our results highlight the potential of OGM and long-read sequencing for bridging the gap between CNV detection and full SV interpretation. In clinical practice, this approach may enhance diagnostic accuracy, inform personalized management, and guide genetic counseling for patients with unexplained rare genetic/genomic disorders by sequencing alone, and for those with dual or triple molecular diagnoses. Future research should systematically assess the pathogenic importance of newly identified SVs and explore the integration of OGM and long-read sequencing data into clinical pipelines.

In conclusion, CN gains or CN gain–CN loss patterns detected using ES should prompt the suspicion of hidden SVs, particularly when CNVs do not fully explain the phenotype. Highly complex SVs, which are often unresolvable by sequencing alone, can be effectively characterized by combining OGM and long-read sequencing. OGM is thus a powerful tool for identifying complex chromosomal rearrangements within its resolution range. Combining OGM with targeted long-read sequencing provides a robust and scalable strategy for comprehensive SV characterization, with direct implications for improving genetic diagnosis and personalized medicine in rare disorders.

Methods

Patients

The study was performed on 30 patients with different disease profiles (Table 1). All patients were clinically evaluated and referred to our laboratory from July 2015 to August 2022. ES had been performed on the patients using the method described in the following section, and all 30 patients had CNVs identified using ES. The ethnicities of patients were Japanese (25 patients) and Brazilian (five patients). The study protocol was approved by the institutional review boards, and written informed consent was obtained from patients or their guardians (IRB of Yokohama City University Faculty of Medicine, Approval number: A170525011/RevB2602000031). The cohort consisted of patients with congenital multiple anomalies who were referred to our laboratory from the pediatrics, neurology, clinical genetics, and obstetrics departments of collaborating institutions/hospitals. To elucidate the causes of rare genetic disorders, we selected 30 cases from 281 cases with CNVs identified byXHMM analysis in ES, out of a total of 8622 cases with rare genetic disorders that underwent ES analysis in our laboratory between 2015 and 2022. Inclusion criteria were as follows: (1)

cases with non-recurrent CNVs identified by ES, in which the causative genes within CNVs did not fully explain the phenotypes, and (2) cases with lymphoblastoid cell lines available in our laboratory. Within this cohort, cases were selected equally from three groups: CN losses, CN gains, and CN gain–CN loss patterns. Exclusion criteria included: (1) cases with CNVs that were already reported to cause recurrent deletion or duplication syndromes, and (2) cases with unavailable lymphoblastoid cell lines. An overview of the study workflow, from ES-based CNV detection through OGM analysis to targeted long-read sequencing using AS, is shown in Fig. S13.

CNV analysis using whole ES

Genomic DNA was extracted from peripheral blood leukocytes using a QuickGene-610L kit (Fujifilm, Tokyo, Japan), according to the manufacturer's instructions. Genomic DNA samples were sequenced on either the NovaSeq 6000 (Illumina, San Diego, CA, USA) using 150-bp paired-end reads after the enrichment of exonic regions using the Twist Human Comprehensive Exome (Twist BioScience, San Francisco, CA, USA) or the HiSeq2000 or HiSeq2500 (Illumina) with 101-bp paired-end reads after capture using the SureSelect Human All Exon kit (Agilent Technologies, Santa Clara, CA, USA). Exome data processing was performed as previously described⁵⁵. Reads were aligned to GRCh38 using Novoalign (<http://www.novocraft.com/>), and PCR duplicates were eliminated using Picard (<https://broadinstitute.github.io/picard/>). The Genome Analysis Toolkit (GATK) (<https://gatk.broadinstitute.org/hc/en-us>) was used to realign indels and recalibrate base quality scores. With 20 reads covering >90% of the complete coding sequence, the mean depth of coverage of ES against the RefSeq coding sequence was 70–80×.

CNVs were detected from the ES data using XHMM, a statistical tool for CN analysis^{2,44,56}. In brief, XHMM detects CNVs by analyzing normalized raw exome read depth data with principal component analysis of the complete coding regions. In Case 30, principal component analysis-normalized and filtered z-scores for the whole read depth were obtained, and SignalMap Version 1.9.0.05 (Roche Nimblegen, Madison, WI, USA) was used for visualization purposes. All CNVs were confirmed using quantitative PCR analysis of patients and their unaffected parents, and inheritance was assumed to be de novo or uniparental (Table 1).

OGM

Ultra-high molecular weight DNA was extracted from 1.5 million lymphoblastoid cell lines using the SP-G2 Blood and Cell Culture DNA Isolation Kit (Bionano Genomics, San Diego, CA, USA) following the manufacturer's instructions⁵⁷. DNA was labeled using the DLS-G2 Labeling Kit (Bionano Genomics) for the attachment of fluorophores to a specific six-nucleotide sequence (CTTAAG). The labeled DNA was loaded on a Saphyr chip G3.2 and run on a Saphyr instrument (Bionano Genomics) for an output of 400 Gb, targeting 80× effective coverage, with >70% of molecules >150 kb aligning (map rate) at N50 (N50 is defined as the length of the shortest molecule for which equal and longer molecules make up 50% of the total data) of >230 kb⁸ (Table S1). De novo assembly, alignment to the GRCh38 reference, and variant annotation of OGM data were performed using Bionano Solve 3.8.1. Visualization by Circos plot and genome browser, filtering, and interpretation of the identified SVs were conducted using Bionano Access 1.8.1. Standard filter settings were adjusted to select only the variants present in the regions of interest identified by previous standard diagnostic tests and in <1% of control samples. In Case 28, OGM data were analyzed using the T2T reference genome in addition to GRCh38 to observe any differences.

Targeted long-read sequencing using AS with GridION

Here, 3 µg of the ultra-high molecular weight DNA used in OGM was fragmented to a target size of 50 kb using a Megaruptor 3 DNAFluid Kit (Diagenode, Seraing, Belgium). In cases in which the remaining amount of ultra-high molecular weight DNA was low, 5 µg of genomic DNA extracted from peripheral blood leukocytes was fragmented to a target size

of 40 kb using a Megaruptor 3 Shearing Kit (Diagenode). The sheared DNA was used to construct sequencing libraries using an Oxford Nanopore Ligation Sequencing Kit (SQK-LSK114) (Oxford Nanopore Technologies, Oxford, UK), largely following the manufacturer's instructions^{55,58} except that enzyme incubation times were doubled, with the final AMPure purification incubation for 10 min at 37 °C. Approximately 25 fmol of the library was loaded onto a flow cell (FLO-MIN114, R10.4.1) on a GridION (Oxford Nanopore Technologies). Target regions comprising 0.24%–0.98% of the whole genome were enriched using the AS option⁵⁹ of GridION Mk1 with a BED file that assigned regions of CNVs detected by ES and each of the surrounding total 7.3–29.5 Mb regions (Table S3). Sequencing was performed for ~3 days, with one additional library loading after nuclease flushing of a flow cell. In cases in which the amount of data was insufficient, new flow cells were used for additional analysis as appropriate.

The data were base-called with super accuracy mode and processed into a BAM file using Dorado v.0.7.2 (Oxford Nanopore Technologies) with GRCh38 as a reference. The depth of coverage was calculated using mosdepth v0.3.1 (<https://github.com/brentp/mosdepth>). The mean depth of the target region was 11.72×–47.38× (Table S3). SVs and CNVs were called using pbsv v.2.9.0 (PacBio SV calling and analysis tools) (<https://github.com/PacificBiosciences/pbsv>) together with breakpoint identification. dnarrange was used to detect SVs and extract patient-specific SVs by referencing 29 control datasets^{57,60}. Control datasets were not used in the analysis if dnarrange excluded breakpoints of interest. Subsequently, using lamassemble (<https://gitlab.com/mcfrith/lamassemble>), each group of overlapping SV reads was merged into a consensus sequence and realigned to the reference genome. Next, dnarrange created a dot-plot representation of SV breakpoints. In dot plots, lines show the alignment to the reference sequence and are drawn in red and blue to denote forward and reverse orientations to the reference genome, respectively. To understand entire SVs, we used the algorithm dnarrange-link to infer the order and orientation of multiple rearrangements⁶¹.

Data availability

The datasets for this article are not publicly available because of concerns regarding patients' anonymity. Requests to access the datasets from qualified researchers should be directed to the corresponding author. There are restrictions on a qualified researcher accessing the data (non-commercial use only and requiring a Data Usage Agreement).

Code availability

All computational tools used in this study are available as open-source software, and their download links are presented in the Methods. Software used are: Dorado v.0.7.2 (<https://github.com/nanoporetech/dorado>), mosdepth v0.3.1 (<https://github.com/brentp/mosdepth>), lamassemble 1.4.2 (<https://gitlab.com/mcfrith/lamassemble>), and dnarrange (<https://github.com/mcfrith/dnarrange>) for targeted long-read sequencing using AS with GridION.

Received: 14 August 2025; Accepted: 6 March 2026;

Published online: 26 March 2026

References

1. Sudmant, P. H. et al. An integrated map of structural variation in 2504 human genomes. *Nature* **526**, 75–81 (2015).
2. Uchiyama, Y. et al. Efficient detection of copy-number variations using exome data: batch- and sex-based analyses. *Hum. Mutat.* **42**, 50–65 (2021).
3. Mantere, T. et al. Optical genome mapping enables constitutional chromosomal aberration detection. *Am. J. Hum. Genet.* **108**, 1409–1422 (2021).
4. Xu, P. et al. OGM and WES identifies translocation breakpoints in PKD1 gene in a polycystic kidney patient and healthy baby delivered using PGT. *BMC Med. Genomics* **16**, 285 (2023).

5. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
6. De Clercq, G. et al. Full characterization of unresolved structural variation through long-read sequencing and optical genome mapping. *Sci. Rep.* **14**, 29142 (2024).
7. Sahajpal, N. S., Barseghyan, H., Kolhe, R., Hastie, A. & Chaubey, A. Optical genome mapping as a next-generation cytogenomic tool for detection of structural and copy number variations for prenatal genomic analyses. *Genes* <https://doi.org/10.3390/genes12030398> (2021).
8. Broeckel, U. et al. Detection of constitutional structural variants by optical genome mapping: a multisite study of postnatal samples. *J. Mol. Diagn.* **26**, 213–226 (2024).
9. Zhang, S. et al. Detection of cryptic balanced chromosomal rearrangements using high-resolution optical genome mapping. *J. Med. Genet.* **60**, 274–284 (2023).
10. Chan, S. et al. Structural variation detection and analysis using bionano optical mapping. *Methods Mol. Biol.* **1833**, 193–203 (2018).
11. Neveling, K. et al. Next-generation cytogenetics: Comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping. *Am. J. Hum. Genet.* **108**, 1423–1435 (2021).
12. Gimeno, A. F., Tinker, R. J., Furuta, Y. & Phillips, J. A. 3rd. Prevalence of individuals with multiple diagnosed genetic diseases in the undiagnosed diseases network. *Am. J. Med. Genet. A* **197**, e63888 (2025).
13. Iqbal, M. A. et al. Multisite assessment of optical genome mapping for analysis of structural variants in constitutional postnatal cases. *J. Mol. Diagn.* **25**, 175–188 (2023).
14. Collins, R. L. et al. A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**, 3041–3055.e3025 (2022).
15. Riggs, E. R. et al. Towards an evidence-based process for the clinical interpretation of copy number variation. *Clin. Genet.* **81**, 403–412 (2012).
16. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
17. Shi, M. et al. Identification of microdeletions in candidate genes for cleft lip and/or palate. *Birth Defects Res. A Clin. Mol. Teratol.* **85**, 42–51 (2009).
18. Alkuraya, F. S. et al. SUMO1 haploinsufficiency leads to cleft lip and palate. *Science* **313**, 1751 (2006).
19. McKnight, D. et al. Multigene panel testing in a large cohort of adults with epilepsy: diagnostic yield and clinically actionable genetic findings. *Neurol. Genet.* **8**, e650 (2022).
20. Truty, R. et al. Possible precision medicine implications from genetic testing using combined detection of sequence and intragenic copy number variants in a large cohort with childhood epilepsy. *Epilepsia Open* **4**, 397–408 (2019).
21. Richards, A. J. et al. Stickler syndrome and the vitreous phenotype: mutations in COL2A1 and COL11A1. *Hum. Mutat.* **31**, E1461–E1471 (2010).
22. Annunen, S. et al. Splicing mutations of 54-bp exons in the COL11A1 gene cause Marshall syndrome, but other mutations cause overlapping Marshall/Stickler phenotypes. *Am. J. Hum. Genet.* **65**, 974–983 (1999).
23. Danecek, P. et al. Detection and characterization of copy-number variants from exome sequencing in the DDD study. *Genet. Med. Open* **2**, 101818 (2024).
24. Cai, M. et al. Ultrasound phenotype, genetic analysis, and pregnancy outcomes of fetuses with 1p36 deletion syndrome. *Mol. Genet. Genom. Med.* **13**, e70104 (2025).
25. Shimojima, K. et al. Genomic copy number variations at 17p13.3 and epileptogenesis. *Epilepsy Res.* **89**, 303–309 (2010).
26. Hyon, C. et al. A new 17p13.3 microduplication including the PAFAH1B1 and YWHAE genes resulting from an unbalanced X;17 translocation. *Eur. J. Med. Genet.* **54**, 287–291 (2011).
27. Stessman, H. A. et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49**, 515–526 (2017).
28. Wilfert, A. B. et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat. Genet.* **53**, 1125–1134 (2021).
29. Schanze, I. et al. NFIB haploinsufficiency is associated with intellectual disability and macrocephaly. *Am. J. Hum. Genet.* **103**, 752–768 (2018).
30. Marinella, G. et al. Further characterization of NFIB-associated phenotypes: report of two new individuals. *Am. J. Med. Genet. A* **191**, 540–545 (2023).
31. Recalcati, M. P. et al. Complex rearrangement involving 9p deletion and duplication in a syndromic patient: genotype/phenotype correlation and review of the literature. *Gene* **502**, 40–45 (2012).
32. Wolff, D. et al. In-frame deletion and missense mutations of the c-terminal helicase domain of SMARCA2 in three patients with Nicolaides-Baraitser syndrome. *Mol. Syndromol.* **2**, 237–244 (2012).
33. Foley, R. et al. Nicolaides-Baraitser syndrome in a patient with hypertrophic cardiomyopathy and SMARCA2 gene deletion. *Cardiol. Young-* **32**, 821–823 (2022).
34. Perovic, D. et al. Chromosomal microarray in postnatal diagnosis of congenital anomalies and neurodevelopmental disorders in Serbian patients. *J. Clin. Lab. Anal.* **36**, e24441 (2022).
35. Tassano, E. et al. Interstitial de novo 18q22.3q23 deletion: clinical, neuroradiological and molecular characterization of a new case and review of the literature. *Mol. Cytogenet.* **9**, 78 (2016).
36. Xiao, F. et al. Genetic spectrums and clinical profiles of critically ill neonates with congenital auricular deformity in the China Neonatal Genomes Project. *Hum. Genet.* **142**, 1737–1745 (2023).
37. Silversides, C. K. et al. Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS Genet.* **8**, e1002843 (2012).
38. Glessner, J. T. et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ. Res.* **115**, 884–896 (2014).
39. Wu, T. & Chen, J. Prenatal diagnosis of Cornelia de Lange syndrome from 12 to 17 weeks' gestation. *Prenat. Diagn.* **42**, 1511–1513 (2022).
40. Pellestor, F., Gaillard, J. B., Schneider, A., Puechberty, J. & Gatinois, V. Chromoanagenesis, the mechanisms of a genomic chaos. *Semin. Cell Dev. Biol.* **123**, 90–99 (2022).
41. Zhang, C. Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
42. Ou, Z. et al. Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes. *Genet. Med.* **10**, 267–277 (2008).
43. Barseghyan, H. et al. Comparative benchmarking of optical genome mapping and chromosomal microarray reveals high technological concordance in cnv identification and additional structural variant refinement. *Genes* <https://doi.org/10.3390/genes14101868> (2023).
44. Miyatake, S. et al. Detecting copy-number variations in whole-exome sequencing data using the eXome Hidden Markov model: an 'exome-first' approach. *J. Hum. Genet.* **60**, 175–182 (2015).
45. Karakoc, E. et al. Detection of structural variants and indels within exome data. *Nat. Methods* **9**, 176–178 (2011).
46. Whitford, W., Lehnert, K., Snell, R. G. & Jacobsen, J. C. Evaluation of the performance of copy number variant prediction tools for the detection of deletions from whole genome sequencing data. *J. Biomed. Inform.* **94**, 103174 (2019).
47. Otori, S. et al. Biallelic structural variations within FGF12 detected by long-read sequencing in epilepsy. *Life Sci. Alliance* <https://doi.org/10.26508/lsa.202302025> (2023).

48. Coccaro, N. et al. Feasibility of optical genome mapping in cytogenetic diagnostics of hematological neoplasms: a new way to look at DNA. *Diagnostics* <https://doi.org/10.3390/diagnostics13111841> (2023).
49. Pei, Y. et al. A comparison of structural variant calling from short-read and nanopore-based whole-genome sequencing using optical genome mapping as a benchmark. *Genes* <https://doi.org/10.3390/genes15070925> (2024).
50. Chaisson, M. J. P. et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* **10**, 1784 (2019).
51. Banu, S. et al. Enhanced resolution of optical genome mapping utilizing telomere-to-telomere reference in genetic disorders. *Eur. J. Hum. Genet.* **33**, 956–959 (2025).
52. Mostovoy, Y. et al. Resolution of ring chromosomes, Robertsonian translocations, and complex structural variants from long-read sequencing and telomere-to-telomere assembly. *Am. J. Hum. Genet.* **111**, 2693–2706 (2024).
53. Carss, K. J. et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am. J. Hum. Genet.* **100**, 75–90 (2017).
54. AlAbdi, L. et al. Beyond the exome: utility of long-read whole genome sequencing in exome-negative autosomal recessive diseases. *Genome Med.* **15**, 114 (2023).
55. Fu, L. et al. Biallelic missense CEP55 variants cause prenatal MARCH syndrome. *J. Hum. Genet.* **70**, 63–66 (2025).
56. Fromer, M. et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am. J. Hum. Genet.* **91**, 597–607 (2012).
57. Ohori, S. et al. Complex chromosomal 6q rearrangements revealed by combined long-molecule genomics technologies. *Genomics* **116**, 110894 (2024).
58. Miyatake, S. et al. Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. *NPJ Genom. Med.* **7**, 62 (2022).
59. Loose, M., Malla, S. & Stout, M. Real-time selective sequencing using nanopore technology. *Nat. Methods* **13**, 751–754 (2016).
60. Mitsuhashi, S., Ohori, S., Katoh, K., Frith, M. C. & Matsumoto, N. A pipeline for complete characterization of complex germline rearrangements from long DNA reads. *Genome Med.* **12**, 67 (2020).
61. Lei, M. et al. Long-read DNA sequencing fully characterized chromothripsis in a patient with Langer-Giedion syndrome and Cornelia de Lange syndrome-4. *J. Hum. Genet.* **65**, 667–674 (2020).
- (JSPS) KAKENHI Grant-in-Aid for Scientific Research [grant numbers JP22K15901 (A.F.), JP23H02829 (S.M.), JP23H02877 (T.M.), JP23K07229 (Y.U.), JP23K15353 (N.T.), and JP24K02230 (N.M.)]; the Takeda Science Foundation (T.M. and N.M.); the Japan Agency for Medical Research and Development (AMED) [grant numbers JP25ek0109674, JP25ek0109760, JP25ek0109617, JP25ek0109648 and JP25ek0109677 (N.M.)]; and Kawano Masanori Memorial Public Interest Incorporated Foundation for Promotion of Pediatrics (S.M.) provided support for this work.

Author contributions

L.F. conceptualized and designed the study, reviewed the literature, analyzed the data, and drafted the manuscript; C.A.K., M.T., Y.Miy., N.O., Y.Ma., H.O., A.F., A.D., J.N., N.U., S.H., K.D., M.F., H.M., M.A., J.O., Y.Mis, J.K., T.S., H.A., R.S., H.H., S.Mit., S.O., K.S., Y.I., K.H., N.T., Y.U., E.K., S.Miy., T.M., A.I., and N.H. analyzed the data and revised the manuscript; A.F. and N.M. supervised all aspects of the study and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-026-00561-4>.

Correspondence and requests for materials should be addressed to Naomichi Matsumoto.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

Acknowledgements

We express our gratitude to the families for their involvement in this research. We further acknowledge the great technical assistance provided by Mr. Takafumi Miyama, Ms. Sayaka Sugimoto, Ms. Mai Sato, Ms. Nobuko Watanabe, and Ms. Kaori Takabe at the Department of Human Genetics, Yokohama City University Graduate School of Medicine. We also thank Bronwen Gardner, PhD, from Edanz (<https://jp.edanz.com/ac>) for editing a draft of this manuscript. The Japan Society for the Promotion of Science

¹Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. ²Department of Obstetrics and Gynecology, Juntendo University Faculty of Medicine, Tokyo, Japan. ³Genetics Unit, Instituto da Crianca, Hospital das Clinicas, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil. ⁴Department of Pediatrics, Kyorin University School of Medicine, Tokyo, Japan. ⁵Department of Medical Genetics, Osaka Women's and Children's Hospital, Osaka, Japan. ⁶Department of Genetic Counseling, Asahikawa Medical University Hospital, Asahikawa, Japan. ⁷Department of Pediatrics, Jichi Medical University, Shimotsuke, Japan. ⁸Comprehensive Epilepsy Center, Seirei Hamamatsu General Hospital, Hamamatsu, Japan. ⁹Division of Neurology, Saitama Children's Medical Center, Saitama, Japan. ¹⁰Department of Pediatrics, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan. ¹¹Department of Pediatrics, St. Marianna University School of Medicine, Kawasaki, Japan. ¹²Department of Pediatrics, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan. ¹³Deguchi Pediatric Clinic, Omura, Japan. ¹⁴Department of Neuropediatrics, Tokyo Metropolitan Neurological Hospital, Tokyo, Japan. ¹⁵Department of Pediatrics, Saitama Medical University Hospital, Saitama, Japan. ¹⁶Department of Pediatrics, Iwate Medical University School of Medicine, Yahaba, Japan. ¹⁷Division of Neonatology, St. Mary's Hospital, Fukuoka, Japan. ¹⁸Department of Neurology, Graduate School of Medical Sciences, Kumamoto

University, Kumamoto, Japan. ¹⁹Department of Pediatrics, Kumamoto University Hospital, Kumamoto, Japan. ²⁰Department of Obstetrics and Gynecology, Keiai Hospital, Saitama, Japan. ²¹Department of Clinical Genetics, Yokohama City University Hospital, Yokohama, Japan. ²²Department of Rare Disease Genomics, Yokohama City University Hospital, Yokohama, Japan. ²³Department of Fundamental Cell Technology, Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan. ✉ e-mail: naomat@yokohama-cu.ac.jp