Centromeric instability and chromoanasynthesis observed in nine supernumerary marker chromosomes resolved with long-read genome sequencing

Kristine Bilgrav Saether, Angelo Salazar Mantero, Marlene Ek, et al.

Genome Res. 2026 36: 661-670 originally published online March 25, 2026
Access the most recent version at doi:10.1101/gr.281175.125

Centromeric instability and chromoanasynthesis observed in nine supernumerary marker chromosomes resolved with long-read genome sequencing

Kristine Bilgrav Saether,[1,2,6] Angelo Salazar Mantero,[1,3,6] Marlene Ek,[1,3] Maria Pettersson,[1,3] Elisabeth Syk Lundberg,[1,3] Christopher M. Grochowski,[4] Claudia M.B. Carvalho,[5] Jesper Eisfeldt,[1,2,3] and Anna Lindstrand[1,2,3]

[1]Department of Molecular Medicine and Surgery, Karolinska Institutet, 17176 Stockholm, Sweden; [2]Science for Life Laboratory, Department of Molecular Medicine and Surgery, Karolinska Institutet, 17176 Stockholm, Sweden; [3]Department of Clinical Genetics and Genomics, Karolinska University Hospital, 17177 Stockholm, Sweden; [4]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; [5]Pacific Northwest Research Institute, Seattle, Washington 98122, USA

Small supernumerary marker chromosomes (sSMCs) remain a diagnostic challenge despite sequencing advances. As the field shifts toward cytogenomics, there is a need to establish methodologies to resolve these complex genetic variants at base pair resolution, as well as to identify their chromosomal origin and formation mechanism. Here, we apply long-read genome sequencing (lrGS) in combination with the telomere-to-telomere (T2T-CHM13) assembly to characterize the structure and genomic content of 10 clinically detected sSMCs. We use sequencing data to reconstruct the derivative chromosomes, identify breakpoint junctions (BPJs), and infer formation mechanisms. We resolve the BPJs of nine of the 10 sSMCs at base pair resolution. The analysis reveals six simple intrachromosomal rearrangements (one continuous and five discontinuous) with one to three BPJs, one complex three-way translocation with two BPJs, and two highly complex intrachromosomal rearrangements with five and nine BPJs, respectively. Breakpoint analysis reveals distinct mechanistic signatures: Simple sSMCs show features consistent with microhomology-mediated end joining (MMEJ) or microhomology-mediated break-induced replication (MMBIR), whereas complex sSMCs demonstrate evidence of translocation, chromoanasynthesis, and breakage–fusion–bridge (BFB) cycles. Haplotype analysis supports trisomy rescue in four cases, including all three complex sSMCs. In summary, our study demonstrates that lrGS combined with T2T-CHM13 enables detailed structural and mechanistic characterization of sSMCs, providing experimental support for disruption of trisomy rescue as a key formation mechanism. This work illustrates the feasibility of resolving highly challenging chromosomal abnormalities using long-read sequencing technologies.

[Supplemental material is available for this article.]

Small supernumerary marker chromosomes (sSMCs) are rare but potentially clinically significant when detected. These extra chromosomes may be linear, either as a centric minute or as an inverted duplicated structure, or ring-shaped; the latter are termed small supernumerary ring chromosomes (sSRCs). Their size and genomic content are highly variable, and determining their chromosomal origin is often challenging with G-banding owing to the small size. Although fluorescence in situ hybridization (FISH) has long been the standard tool for characterizing such events, genomic technologies are now emerging as powerful alternatives. However, resolving sSMCs and interpreting their clinical significance remains challenging, in part because they frequently are mosaic and often comprise repetitive or poorly annotated genomic regions (Reddy et al. 2013; Mostovoy et al. 2024).

sSMCs are found in ~0.044% of the human population, with ~30% of carriers exhibiting mild to severe phenotypes (Liehr and Weise 2007). They are classified based on shape, genomic content, and degree of mosaicism. About 20%–30% of small supernumerary chromosomes are sSRCs, whereas the remainder are linear sSMCs, often derived from inverted duplications. Approximately 70% of small supernumerary chromosomes originate from acrocentric chromosomes, with Chromosome 15 being the most common source (particularly inv dup(15)/idic(15)). To understand the clinical implications, each case requires careful molecular and clinical evaluation to determine the chromosomal origin and genomic content of the sSMC. Pathogenicity may result from extra copies of triplosensitive genes, as well as epigenetic effects (Anderlid et al. 2001; Mostovoy et al. 2024). Approximately 70% of sSMC carriers are asymptomatic, and a subset is discovered during infertility investigations, in which the increased risk of infertility could be explained by imbalances that arise during gametogenesis.

[6]These authors contributed equally to this work.
Corresponding author: jesper.eisfeldt@scilifelab.se
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.281175.125. Freely available online through the *Genome Research* Open Access option.

Overall, genomic characterization is relevant, especially for genetic counseling in prenatal diagnostics.

The majority of linear sSMCs possess a centric minute structure, meaning that they are linear with a centrally located centromere (Reddy et al. 2013). For some sSMCs, the mechanisms have been extensively studied, such as the isodicentric Chromosome 15 (idic(15)). In these cases, the extra chromosome consists of two copies of a segment from Chromosome 15q11–q13, arranged in an inverted mirror-image orientation with two centromeres (dicentric) and formed through nonallelic homologous recombination between inverted repeats. The mechanisms underlying other sSMCs are less understood but include postfertilization errors, as well as trisomy and monosomy rescue events following meiotic and mitotic nondisjunction (Grochowski et al. 2018; Kurtas et al. 2019).

Base pair–level resolution of sSMC structures has long been a challenge owing to breakpoints frequently occurring in repetitive regions such as centromeres, telomeres, and acrocentric p-arms, or within low copy repeats (LCRs) (Grochowski et al. 2018; Mostovoy et al. 2024; Schuy et al. 2024). This has limited our understanding of both their formation mechanisms and clinical consequences. Long-read genome sequencing (lrGS) enables high-accuracy analysis of the genome, allowing sSMCs to be resolved at nucleotide-level precision. In addition, the telomere-to-telomere (T2T-CHM13) assembly provides enhanced sequence completeness across satellite and other repetitive regions commonly involved in structural variant (SV) formation (Nurk et al. 2022; Bilgrav Saether et al. 2024).

This study aims to evaluate the capability of emerging genomic technologies, specifically lrGS and optical genome mapping (OGM) in combination with the T2T-CHM13 reference genome, to achieve high-resolution breakpoint mapping of sSMCs. We seek to determine whether these approaches can fully resolve complex chromosomal rearrangements and provide mechanistic insights into their formation, as well as assess their potential clinical relevance.

Results

Genetic characterization of sSMCs

Seven of the sSMCs were ring shaped, two were linear, and one remained unresolved. Nine were intrachromosomal events, consisting of genomic material from single chromosomes, whereas one involved material from three different chromosomes (Table 1).

By combining short-read genome sequencing (srGS) and lrGS, we resolved all breakpoint junctions (BPJs) of the sSMCs except for three cases with breakpoints in highly repetitive regions (RD_P272, RD_P166, and RD_P274). OGM analysis (performed in two cases lacking srGS data) could verify the ring structure of r(14) (RD_P550) and eight of 10 segments and their order in der(X) (RD_P586) (Supplemental Fig. S1A,B). Six of the resolved sSMCs were classified as simple events, involving one to three BPJs on a single chromosome (RD_P276, RD_P278, RD_P273, RD_P166, RD_P550, RD_P275), and three were complex rearrangements involving either more than three BPJs (RD_P272, RD_P586) or more than one chromosome (RD_P328) (Table 2). In traditional cytogenetics, a simple sSMC is defined as a continuous DNA fragment of a chromosome that has lost its distal ends, whereas a discontinuous sSMC consists of intrachromosomal material with

Table 1. Overview of the studied sSMCs

Case ID	Previous publication	Clinical summary	Gender	Cytogenomic description	Genome data	Tissue	Mosaicism degree (%)	Resolved
RD_P276	Anderlid et al. 2001	ID, DD, DF	Male	+ r(5)(p13.3 → p13.2::p12 → q11.1)	srGS, lrGS	Blood	13	Yes
RD_P278	Anderlid et al. 2001; Blennow et al. 1993	DF, neurological problems	Male	+ r(7)(p11.2 → p11.2::q11.22 → p11.2)	srGS, lrGS	Blood	100	Yes
RD_P273	Anderlid et al. 2001	DF, DD	Female	+ r(8)(q11.1 → q11.21::q21.13 → q11.22)	srGS, lrGS	LCL	40	Yes
RD_P272	Anderlid et al. 2001	ID, DD	Female	+ r(9)(p21.1 → p21.1::q21.31 → q21.31::p12→p21.1::p11.2 → ?::? → p21.1)	srGS, lrGS	Blood	36	Yes[a]
RD_P274	Unpublished	ID, DF	Male	+ r(10)(p12 → p11.2)	srGS, lrGS	LCL	66	No
RD_P166	Unpublished	DD, short stature	Female	+ r(13)(? → p11.2::q21.1 → q31.1:: q12.3 → ?)	srGS, lrGS	LCL	87	Yes[b]
RD_P550	Unpublished	Short stature	Female	+ r(14)(p11.2q11.2)	lrGS, OGM	Blood	46	Yes
RD_P275	Anderlid et al. 2001	ID, DD	Male	+ r(20)(p13 → p13::q13.31 → q13.33:: p11.21 → q12)	srGS, lrGS	LCL	48	Yes
RD_P328	Blennow et al. 1992	ID, DD	Female	+ der(7)(Xpter → Xp22.2::7q11.23 → 7p13::5q35.3 → 5qter)	srGS, lrGS	LCL	100	Yes
RD_P586	Unpublished	Mild ID, ataxia	Female	+ der(X)(qter → q27.3::q22.3 → q21.31::q21.31 → q21.2::q12 → q12:: p11.1 → q13.3::q27.3 → q27.2::q21.1 → q21.2::q13.3 → q21.1::q21.31 → q21.31::q27.3 → qter)	lrGS, OGM	Blood	100	Yes

(ID) Intellectual disability; (DD) developmental delay; (DF) dysmorphic features; (LCL) lymphoblastoid cell lines.
[a]Centromeric breakpoints are not mapped.
[b]Breakpoints on acrocentric p-arm are not mapped. Full ISCN and HGVS nomenclature is in Supplemental Table S3.

Table 2. Characteristics of the resolved breakpoint junctions (BPJs) of marker and ring chromosomes, with coordinates mapped to the T2T-CHM13v2.0 genome build

Case ID	sSMC	BPJ	Chr	PosA	PosB	Segment posA–posB	PosA: genes	PosB: genes	PosA: repeat	PosA: segmental duplication	PosB: segmental duplication	PosB: repeat	Microhomology (bp)	Insertion (bp)	Mechanism
RD_P276	r(5)	1	Chr 5	37,676,103	43,290,183	A–B	*WDR70, intron 5*	*AC025171.1, intron 2*	LTR: MLT1H2	—	—	Cent.	3	—	MMEJ
		2	Chr 5	32,320,809	**49,497,791**[a]	A–B	—	—	LTR: MLT1F2	—	—	Cent.	3	—	
RD_P278	r(7)	1	Chr 7	59,987,274	72,094,570	A–B	—	—	Cent.	*Yes*	—	*Alu*Sx1	—	20 NT	MMBIR
		2	Chr 7	59,978,281	59,987,351	A–B	–	–	Cent.	Yes	Yes	Cent.	3	—	
RD_P273	r(8)	1	Chr 8	50,587,595	77,860,383	A–B	*SNTG1*	*AC062004.1*	L1P1	—	—	—	2	—	MMEJ
		2	Chr 8	46,120,188[a]	51,426,338	A–B	—	—	Cent.	—	—	—	—	4 (templated)	
RD_P272	r(9)	1	Chr 9	30,592,557	31,057,050	B–C	—	—	—	—	—	L2c, L2a	—	28 NT	MMBIR
		2	Chr 9	28,144,043	31,568,556	A–C	*LINGO2, intron3*	—	L2d2	—	—	L1P1	—	224 NT	
		3	Chr 9	29,574,524	91,174,804	A–F	—	—	*Alu*Sx1	—	—	—	3	—	
		4	Chr 9	39,306,747	91,060,755	D–F	—	—	*AluY*k2, Satellite	Yes	—	MIRc	—	—	
		5	Chr 9	32,076,512	44,448,128	D–E	—	*ANKRD20A7P (pseudogene)*	L1PB3	—	Yes	Cent.	—	100 NT	
RD_P166	r(13)	1	Chr 13	**11,452,840**	56,358,138	A–C	—	*PRR20C*, intron 1	Cent.	Yes	Yes	—	—	42 (templated)	MMBIR
		2	Chr 13	29,045,867	84,773,620	B–C	*UBL3*, intron 1	LINC00351, intron 1	*Alu*Sp	—	—	—	—	—	
RD_P550	r(14)	1	Chr 14	**9,208,604**	16,434,843	A–A	—	*AC244502.1, TRA*	Cent.	Yes	—	Cent., LTR/ERV1	2	—	MMEJ
RD_P275	r(20)	1	Chr 20	4,979,434	59,369,410	A–C	*SLC23A2, intron 2*	LOC102724968	—	—	—	L2	—	205 (templated)	MMBIR
		2	Chr 20	24,892,355	65,776,662	B–C	—	*UCKL1, intron 1*	Cent.	—	—	—	7[b]	2[b]	
		3	Chr 20	3,306,020	42,368,945	A–B	*DNAAF9/ C20orf194/ MSTRG*	—	—	—	—	—	4	—	

(*continued*)

Table 2. *Continued*

Case ID	sSMC	BPJ	Chr	PosA	PosB	Segment posA–posB	PosA: genes	PosB: genes	PosA: repeat	PosA: segmental duplication	PosB: segmental duplication	PosB: repeat	Microhomology (bp)	Insertion (bp)	Mechanism
RD_P328	der(7)t (X;7;5)	1	Chr 7-Chr 5	Chr 7: 43,886,314	Chr 5: 177,119,928	A–B	*COA1, intron 1*	*CDHR2, exon 9*	L1MB8, *Alu*Yk2	—	—	—	—	—	MMEJ
		2	Chr X-Chr7	Chr X: 14,614,119	Chr 7: 75,313,264	C–A	—	*LIMK1, intron 10*	Simple repeats (AAAT)	—	—	*Alu*Jb	3	—	
RD_P586	der(X)	1	Chr X	103,977,823	145,148,123	F–J	*IL1RAPL2, intron 6*	—	L2d, L1M4a1	—	—	L1MA1	1	—	MMBIR
		2	Chr X	89,566,572	89,651,697	F–E	—	—	L1ME1	—	Yes	L1ME3, ERVL-MaLR	—	—	
		3	Chr X	66,181,028	84,675,390	B–E	—	*DACH2, intron 1*	L1ME3B	—	—	—	2	—	
		4	Chr X	**58,080,098**	66,184,481	A–B	NA	—	NA	—	—	LINE/L2a	2	—	
		5	Chr X	74,310,915	144,750,571	A–H	—	—	L1PA7, MLT2D, L1MEc	—	—	L1MA2, ERVL-MaLR	—	1	
		6	Chr X	83,792,179	141,176,418	D–H	*POF1B, intron 3*	–	L1PREC2	—	—	—	—	37 NT (L1)	
		7	Chr X	75,059,486	83,829,717	C–D	—	—	L1PBa	—	—	L1MEd, L1MA1	—	29 NT	
		8	Chr X	83,190,233	91,038,187	D–G	*TEX16P (pseudogene)*	*PCDH11X, intron 5*	—	—	Yes	—	—	41 NT	
		9	Chr X	91,089,607	144,762,724	G–I	—	—	L1MB7, L1PBb, L1PA13	Yes	—	LTR16B1	—	33 NT	

(Chr) Chromosome; (Pos) position; (bp) base pairs; (NT) nontemplated; (NA) not applicable; (MMBIR) microhomology-mediated break-induced replication; (MMEJ) microhomology-mediated end joining.
Positions marked in **bold** indicate the breakpoint in sequence is not present in GRCh38 (Bilgrav Saether et al. 2024). The size of each segment is available in Supplemental Table S2.
[a]Ambiguous position.
[b]Complex mechanism.

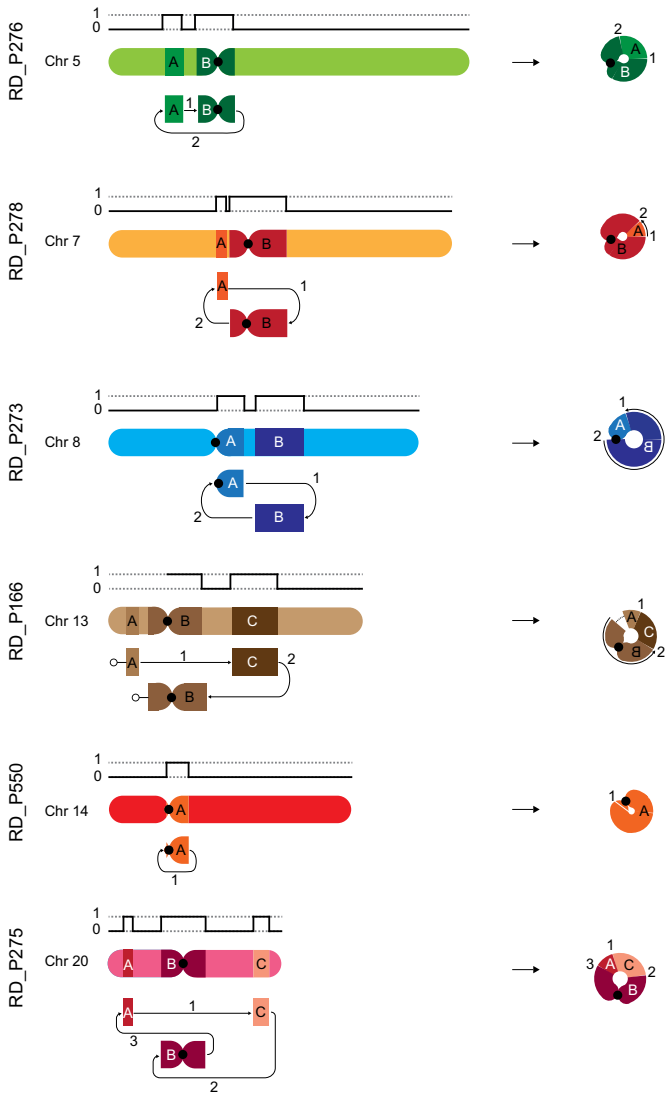

Figure 1. Six simple small supernumerary marker chromosomes (sSMCs). Schematic of the six resolved simple sSMCs and their chromosome copy-number profile. Letters indicate segments, and numbers indicate junctions. The centromere is shown as a black circle. For ring Chromosome 13, the two lines ending in circles represent a breakpoint junction (BPJ) between segments A and B (dotted line on the *right*) that was not identified.

more than one fragment, and complex sSMCs consist of material from more than one chromosome. Using this system, we found that five of the intrachromosomal sSMCs were discontinuous, and one was continuous. However, because the aim of this study is to classify sSMCs based on their characterized BPJs, we will hereafter use a different system that reflects their molecular complexity rather than their cytogenetic appearance.

Twenty-seven out of the 32 sSMC segments involved OMIM genes (range, 1–159). One hundred eighty-five were triplosensitive and 202 intellectual disability (ID) genes (Supplemental Table S2).

Characterization of simple sSMCs

The simple sSMCs consisted of one to three segments (Fig. 1; Supplemental Figs. S2–S7). The simplest sSRC (r(14); RD_P550) consisted of one segment sized 7.2 Mbp; although cytogenetically classified as a linear marker, lrGS revealed a ring structure, as the reads spanned over the BPJ fusing the end and start of the segment. Three sSRCs consisted of two segments (r(7), RD_P278; r(5), RD_P276; and r(8), RD_P273). Of those, r(7) and r(5) had similar structures, with one segment originating from the p-arm and one spanning the centromere; in contrast r(8) consisted of one segment with a centromeric breakpoint, and the other originated from the q-arm. Finally, two sSMCs consisted of three segments: the r(20) (RD_P275) and r(13) (RD_P166). The segment lengths ranged from 8 kbp to 28.4 Mbp, and three of 13 were inverted relative to the centromeric segment.

In total we detected 12 BPJs in six simple sSMCs (range, one to three BPJs per sSMCs). The BPJ architecture revealed seven of 12 with microhomology (2–7 nt) and five of 12 with insertions (range, 4–204 nt; three templated and two random). In eight of the 12 BPJs, one or both breakpoints were located within genes; four of the 12 were in segmental duplications, and 11 of the 12 were in repeat elements, eight of which involved centromeric sequences (Table 2).

Characterization of complex sSMCs

The three complex sSMCs (RD_P272, RD_P586, RD_P328) differed in structure and complexity. The interchromosomal sSMC, der(7)t(X;7;5) (RD_P328), consisted of three segments from three different chromosomes, joined by two BPJs (Fig. 2A–C). The three fragments originated from Chromosome 5 (2.9 Mbp), Chromosome 7 (31.4 Mbp), and Chromosome X (15 Mbp), with the smaller fragments flanking the Chromosome 7 fragment.

The two more complex sSMCs were both intrachromosomal (r(9), RD_P272, der(X); RD_P586) (Fig. 3; Supplemental Figs. S8, S9) consisting of six and 10 segments, respectively. Unlike the other sSMCs, these two incorporate triplicated segments (segments B, G, J in der(X) and C, D, E in r(9)) (Fig. 3).

Haplotype analysis

Analysis of SNV patterns within the sSMC revealed three distinct haplotypes in all three complex cases, der(7), der(X), and r(9), as well as in r(13), suggesting that the sSMCs originate from different chromatids (Fig. 3; Supplemental Fig. S10).

In total, we detected 16 BPJs in three complex sSMCs (two, five, and nine for der(7), r(9), and der(X), respectively). The BPJ architecture revealed that five of the 16 showed microhomology (1–3 nt) and eight of the 16 insertions (1–224 nt; all random). In eight of the 16 BPJs, one or both breakpoints were located within genes, five of the 16 in segmental duplications, and 15 of the 16 in repeat

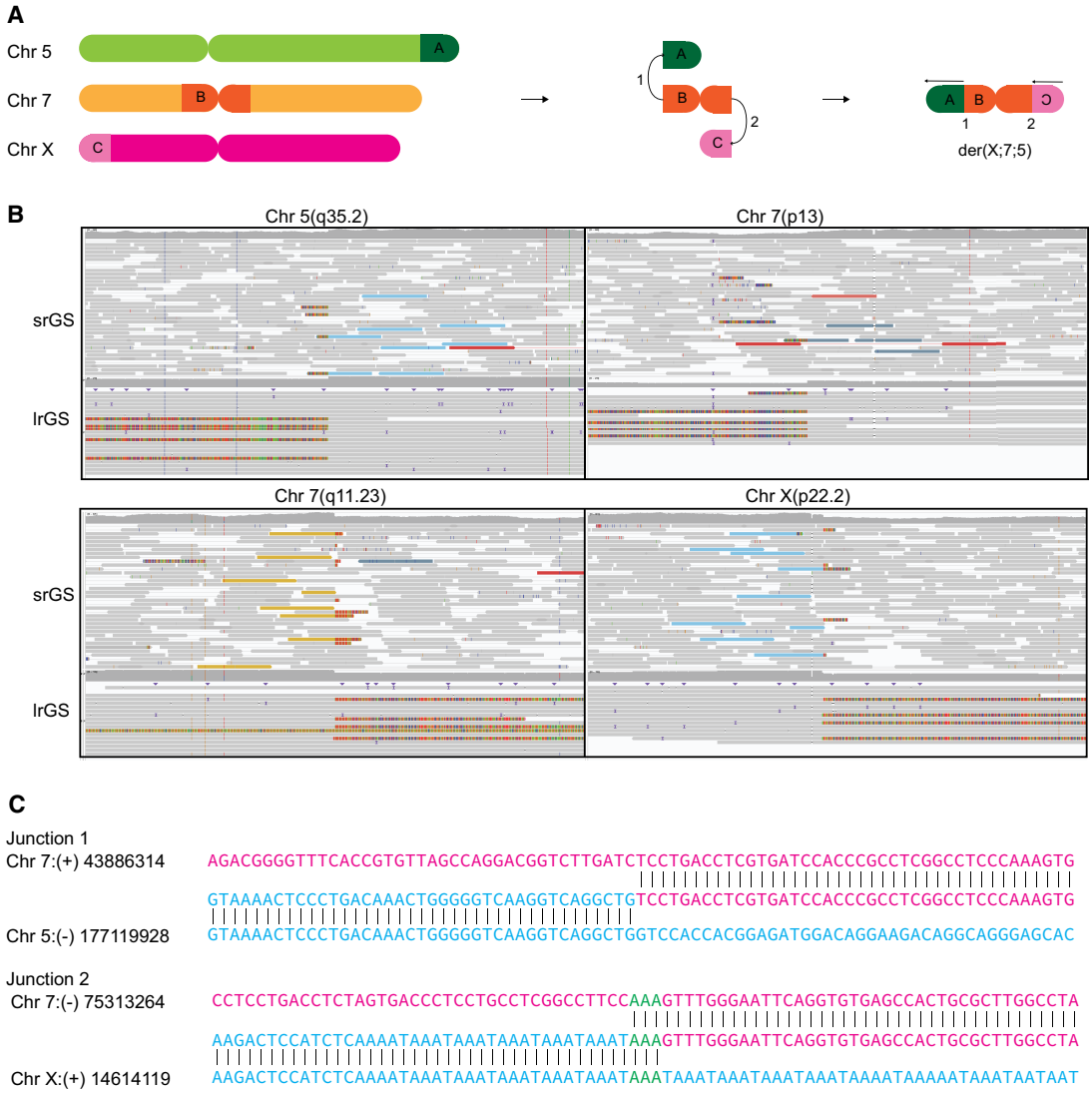

Figure 2. A complex sSMC involving Chromosomes X, 7, and 5. (*A*) Schematic of the chromosomes involved and the resulting structure of the derivative Chromosome 7 identified in RD_P328. (*B*) Integrative Genomics Viewer images of short-read genome sequencing (srGS) and long-read genome sequencing (lrGS) data for each BPJ. (*C*) Breakpoint sequences of the two BPJs with the chimeric sequence in the *center* aligned to the reference sequences shown in pink and blue. Microhomology is indicated in green. Coordinates according to genome build T2T-CHM13v2.0.

elements, none involving centromeric sequences. In der(X), matched repeats were observed in five out of nine BPJs (Table 2).

Altogether, the BPJ architecture differed between the simple and complex sSMCs, suggesting distinct mechanisms of formation. In the simple sSMCs, we observed a high frequency of centromeric BPJs.

Discussion

Long-read genomic analysis using the T2T-CHM13 reference resolved nine of 10 sSMCs, identifying six simple intrachromosomal and three complex events, whereas one centromeric ring-shaped sSMC remained unresolved. The simple sSMCs consisted of one to three segments and BPJs, whereas the complex cases involved up to 10 segments, including duplicated fragments, and/or spanned multiple chromosomes. Notably, the resolved sSMCs displayed distinct breakpoint signatures and structural features, suggesting involvement of different mutational mechanisms reflecting diverse paths of sSMC formation.

Traditionally, sSRCs are believed to arise through terminal breaks on each chromosome arm, followed by direct end-to-end fusion via double-strand break repair mechanisms. However, recently it has been shown that sSRCs may form through mechanisms related to the formation of complex genomic rearrangements, such as breakage–fusion–bridge (BFB) cycles (Nazaryan-Petersen et al. 2018) and chromanasynthesis (Grochowski et al. 2018; Kurtas et al. 2018; Schuy et al. 2024).

The breakpoint characteristics of sSMCs studied here correlate with their structural complexity. All simple sSMCs (Fig. 1) had at least one BPJ located within a centromere (Table 2), and except for those centromeric repeats, no matched repeats were observed. Only two BPJs were blunt, both in r(13), and the remaining featured either insertions or microhomology consistent with formation via MMEJ or MMBIR. Whether these simple sSMCs arose in

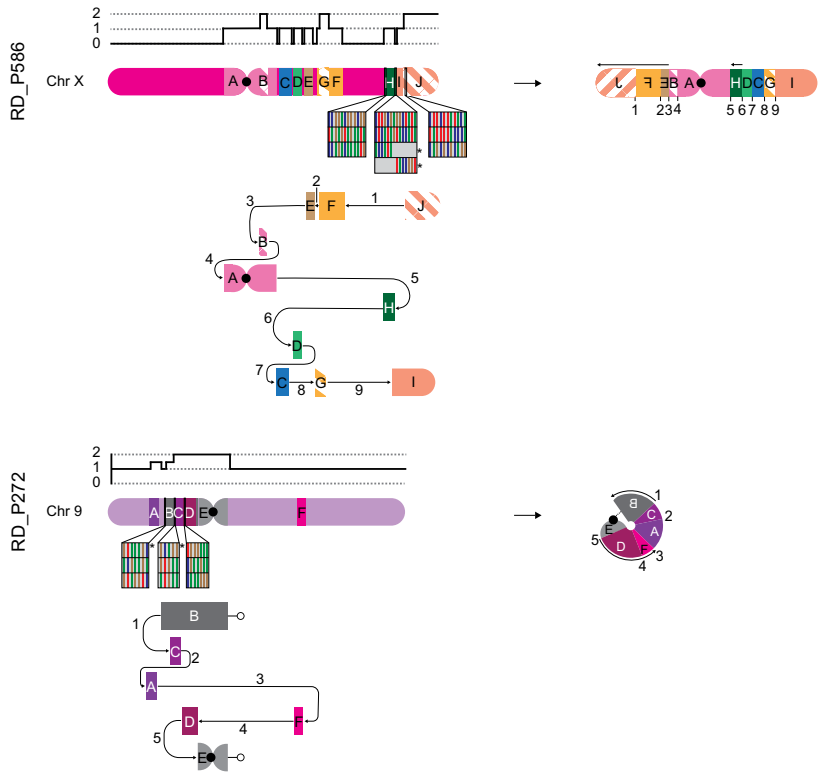

Figure 3. Two complex sSMCs with features of trisomy rescue. der(X) (RD_P586): On the *left* side, the copy-number profile and affected genomic regions A–J with triplications shown with striped patterns. For regions H, I, and J, the identified haplotypes' single-nucleotide variant pattern is displayed *below* with the sSMC haplotype indicated with *. In the *middle*, a subway plot shows how BPJs 1–9 connect segments A–J. On the *right* side, the resolved sSMC structure is given. r(9) (RD_P272): as in *A*. The two lines ending in circles represent a BPJ between segments *B* and *E* (dotted line on the *right*) that was not identified.

a single event or through a multistep process such as BFB cycles remains uncertain. The observation that all simple sSMCs consisted largely of centromeric material suggests that their formation may be driven by centromere instability. This might be a consequence of sSMCs containing centromeric fragments being more stable, as they can attach to the mitotic spindle. However, centromeric instability itself has been broadly implicated in pathological chromosomal rearrangements and human disease (Barra and Fachinetti 2018). Recent work on individuals with trisomy 21 (Down syndrome, OMIM 190685) has highlighted that centromeres may vary in size and structure and that extreme centromere size asymmetry is a frequent feature in trisomic cells (Mastrorosa et al. 2025). Future studies characterizing centromere size and composition in sSMCs may provide further insights into the exact contribution of centromeres to sSMC formation and stability. Segmental duplications were identified in nine of 28 junctions, but only one (junction 2, r(7)) had breakpoints within matching segmental duplications (Table 1), suggesting they played little role in sSMC formation.

In the complex group, blunt-ends and short microhomologies are present in the der(7) BPJs (Fig. 2; Table 2), suggesting that they likely formed through MMEJ.

The combined data indicate that this interchromosomal sSMC formed through a three-way translocation event. The other two complex sSMCs, r(9) and der(X), are intrachromosomal rearrangements consisting of five and 10 segments, respectively (Fig. 3). These segments include both duplicated and triplicated regions (present once or twice on the resolved sSMC structure). Although their overall structures appear similar, the BPJ features differ (Table 2). The r(9) BPJ characteristics, microhomology and random insertions, are consistent with chromoanasynthesis.

In contrast, der(X) consists of duplicated fragments, inversions, and telomere loss, suggesting that it may have formed through BFB cycles. Five out of nine BPJs in der(X) are located within matched repeat elements, indicating that the rearrangement was LINE-mediated. This is supported by our previous work showing that complex rearrangements clustering on the same chromosome can be mediated by both *Alu* and LINE elements (Nazaryan-Petersen et al. 2018). However, a highly complex neocentric sSMC that appears to have formed through chromothripsis highlights that chromoanagenesis of complex sSMCs may involve different underlying formation mechanisms (Weber et al. 2022).

Across four sSMCs, one simple and all three complex (r(7), r(9), r(13), and der(X)), three haplotypes were detected, indicating that the sSMC is likely a remnant of a trisomy rescue event. In der(X), however, only three of the 10 segments showed such features, indicating partial trisomy rescue. Previous studies have reported chromothripsis in micronuclei during incomplete trisomy rescue as a formation mechanism for sSMCs (Kurtas et al. 2019; Matsubara et al. 2020), which resembles the patterns observed in r(9), r(13), and der(X). Micronucleus formation and incomplete trisomy rescue involving multiple chromosomes have also been described, leading to complex chromosomal rearrangements (Storchová and Kloosterman 2016), which may explain the r(7) and the presence of three haplotypes. This likely represents a peri-zygotic event, with the mosaic nature of the sSMCs attributable to mitotic instability. This is a similar phenomenon to that described for multiple de novo copy-number variants (CNVs) (Du et al. 2022).

The 10 individuals with sSMCs were initially referred for genetic investigation owing to neurodevelopmental and neurological symptoms, as well as short stature (Table 1). These features are commonly seen in patients undergoing chromosome analysis and consistent with previously reported clinical presentations in individuals with sSMCs (Conlin et al. 2011; Guilherme et al. 2011; Pristyazhnyuk and Menzorov 2018; Kurtas et al. 2019; Li et al. 2022). Even though we have detailed the exact coordinates of involved genomic segments and genes (Supplemental Table S2) whether the duplication or triplication of specific genes contribute to the phenotype remains unclear, as transcriptional activity and downstream effects were not assessed.

We and others have previously shown that lrGS is an effective genomic technique for identifying and resolving large chromosomal rearrangements within highly repetitive regions (Porubsky et al. 2022; Bilgrav Saether et al. 2024; Eisfeldt et al. 2024;

Mostovoy et al. 2024; Schuy et al. 2024; Ten Berk de Boer et al. 2024). In this study, lrGS was critical for resolving the breakpoints of sSMCs with BPJs located in highly repetitive and complex regions. In contrast, srGS, performed in eight of the 10 sSMCs, detected and even resolved two events, r(7) and the complex interchromosomal der(7), but could not phase the remaining sSMC structures with BPJs in large repeats (Table 2). Therefore, in this study, srGS data were primarily useful for CNV analysis. OGM confirmed eight of the 10 involved segments and the overall structure of the most complex sSMC (der(X)) and was a valuable tool for validation in this case.

The T2T-CHM13 assembly added >250 Mbp of new sequence to the reference genome compared with GRCh38 (Nurk et al. 2022; Bilgrav Saether et al. 2024). We and others have shown that these sequences can be clinically relevant, especially for SV analysis (Porubsky et al. 2023; Bilgrav Saether et al. 2024; Ten Berk de Boer et al. 2024; Paulin et al. 2025). Of note, in four of the 28 sSMC BPJs reported here, one breakpoint was located in genomic sequence missing from GRCh38 but present in T2T-CHM13 (Table 2). The specific sSMCs (r(5), r(13), r(14), der(X)) could therefore not have been resolved with lrGS aligned to GRCh38; hence, T2T-CHM13 was essential. However, for r(13) and r(14), with breakpoints in the p-arm of acrocentric chromosomes, T2T-CHM13 still did not enable full characterization. Specifically, although lrGS allowed us to identify the breakpoints in these regions, we were not able to fully characterize the BPJs. Similarly, for r(9), two segments remain unplaced in the final sSMC structure, most likely owing to polymorphic repetitive sequences at the breakpoints. Although one BPJ cannot be resolved with lrGS aligned to T2T-CHM13, cytogenetic investigations have shown that it has a ring structure. This highlights the continued importance of cytogenetics as a complementary method to genomics for resolving highly complex rearrangements, particularly when breakpoints are located in polymorphic or repetitive regions. Some of these limitations may be addressed using pangenomes, which can represent polymorphic sequences more comprehensively (Liao et al. 2023).

In summary, we demonstrate the power of lrGS, in combination with the T2T-CHM13 assembly, to resolve complex and repetitive breakpoint regions and capture variation within centromeres and acrocentric p-arms, allowing a detailed molecular characterization of sSMCs. This approach delineates distinct formation mechanisms for simple and complex markers and reveals evidence of trisomy rescue in both groups. It also points toward improved diagnostics by enabling more complete capture and interpretation of cytogenetic aberrations with higher resolution and accuracy compared with traditional karyotyping and array-based methods.

Methods

Ethics approval and consent

Ethical approval was given by the regional ethical review board in Stockholm, Sweden (ethical permit nos. 2012/222-31/3 and 2024-05341-01). These ethics permits allow the use of clinical samples for analysis of scientific importance as part of clinical development. The IRB approval does not require us to get written consent for clinical testing. The research conformed to the principles of the Helsinki Declaration.

Short-read genome sequencing

srGS was performed using the HiSeq Xten platform. The samples were sequenced using 2 × 150 bp paired-end reads, on one flow cell for each sample, which resulted in 30× median coverage. The data were aligned to T2T-CHM13 (chm13v2.0) (Nurk et al. 2022) using BWA-MEM (v0.7.17) (Li and Durbin 2009). SVs were detected using CNVnator (v0.4.1) (Abyzov et al. 2011) and TIDDIT (v2.8.1) (Eisfeldt et al. 2017). The variant calls were annotated and filtered based on allele frequency using SVDB (2.8.2), and relevant SV calls were visualized and quality-controlled using the Integrative Genomics Viewer (IGV) (Robinson et al. 2011).

Long-read genome sequencing

lrGS was conducted using the Pacific Biosciences (PacBio) Revio system, with average read lengths ranging from 10 to 19 kbp and average coverage between 22× and 67× (Supplemental Table S1). In brief, reads were aligned to T2T-CHM13 (chm13v2.0) using minimap2 (2.24-r1122) (Li 2018); SNVs were called using DeepVariant (1.5.0) (Poplin et al. 2018); SVs were called using Sniffles (1.0.12-1) (Sedlazeck et al. 2018); and CNVs were called using HiFiCNV (v1.0.1-4de7d40) (https://github.com/PacificBiosciences/HiFiCNV).

A variant frequency database consisting of nine samples was used to annotate each variant with population frequency. WhatsHap (1.7) (Martin et al. 2023) was used to phase variants and haplotag the BAM file. All commands are given in the Supplemental Code. De novo assembly was performed using Hifiasm (0.23.0–h5ca1c30_0) (Cheng et al. 2021), and SVs were called using SVIM-asm (v1.0.3) (Supplemental Code; Heller and Vingron 2021).

Optical Genome Mapping

Optical Genome Mapping (OGM) was performed as described previously (Bilgrav Saether et al. 2024). In brief, ultra-high-molecular-weight genomic DNA from cases RD_P550 and RD_P586 were loaded onto a flow cell on the Saphyr optical mapping system (Bionano Genomics). Data were aligned to an in silico T2T-CHM13 reference genome using the Bionano Solve v3.7 RefAligner module.

Identification and characterization of sSMCs

SV calls (srGS, lrGS, OGM) were intersected with CNV patterns (srGS) and manually searched for informative signals involving the chromosomal regions of interest. Their obtained list was visualized in IGV (Robinson et al. 2011).

To identify regions present in T2T-CHM13 but absent from GRCh38, breakpoint positions were intersected using BEDTools (Quinlan and Hall 2010) with a list of regions not present in GRCh38 (Bilgrav Saether et al. 2024).

To understand the function and biological consequences of the sSMCs, lists of disease-associated genes were downloaded on November 14, 2024, from OMIM (Amberger et al. 2015) (Morbid OMIM version 31.0, 4629 genes), ClinGen (1281 triplosensitivity genes) (Rehm et al. 2015), and Genomics England PanelApp (2792 ID genes) (Martin et al. 2019). Gene positions in GRCh38 were converted to T2T-CHM13 coordinates by extracting gene information from a gene annotation file downloaded from the UCSC Table Browser (Perez et al. 2025). Finally, the coordinates were intersected with the genomic locations of the segments in the sSMCs.

Haplotype analysis was performed by manual inspection of informative SNVs using IGV. Trisomy rescue was considered when heterozygous SNVs were observed that did not align with either of the two haplotypes defined by WhatsHap, suggesting the presence of a third haplotype.

Data access

The long-read sequencing data generated in this study have been submitted to the European Genome-phenome Archive (EGA; https://ega-archive.org) under accession number EGAS5000000 1466.

Competing interest statement

A.L. has received speakers' honoraria from Pacific Biosciences and Illumina. All other authors declare no competing interests.

Acknowledgments

We acknowledge the support from the National Genomics Infrastructure (NGI) at Science for Life Laboratory in providing assistance in massive parallel sequencing. The computations were performed on resources provided by SNIC through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project SNIC sens2017106, sens2020021, and sens2023005. An author of this publication is a member of the European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability ERN-ITHACA (EU framework partnership agreement ID: 3HP-HP-FPA ERN-01-2016/739516). A.L. was supported by grants from the Swedish Research Council (2019-02078), Region Stockholm (FoUI-1000468 and FoUI-978581), the Rare Diseases Research Foundation (Sällsynta Fonden), the Swedish Brain Foundation (FO2024-0128-HK-44), and the Swedish Cancer Society (24 3504 Pj).

Author contributions: A.L. and J.E. conceptualized the study, administered the project, and provided resources, supervision, and funding. K.B.S. performed bioinformatic analysis. K.B.S. and A.S.M. performed initial data curation and formal analysis, and M.E., A.L., J.E. C.M.B.C., C.M.G., and E.S.L. helped with clinical and genomic interpretations of results. M.E. did wet-laboratory experiments. K.B.S., A.L., and J.E. wrote the original draft. K.B.S. made the figures supported by A.S.M. A.S.M, M.E., M.P., E.S.L., C.M.G., and C.M.B.C. reviewed and edited the original draft.

References

Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* **21:** 974–984. doi:10.1101/gr.114876.110

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. 2015. OMIM.org: online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* **43:** D789–D798. doi:10.1093/nar/gku1205

Anderlid BM, Sahlén S, Schoumans J, Holmberg E, Åhsgren I, Mortier G, Speleman F, Blennow E. 2001. Detailed characterization of 12 supernumerary ring chromosomes using micro-FISH and search for uniparental disomy. *Am J Med Genet* **99:** 223–233. doi:10.1002/1096-8628(2001)9999:9999<::AID-AJMG1146>3.0.CO;2-W

Barra V, Fachinetti D. 2018. The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nat Commun* **9:** 4340. doi:10.1038/s41467-018-06545-y

Bilgrav Saether K, Eisfeldt J, Bengtsson JD, Lun MY, Grochowski CM, Mahmoud M, Chao HT, Rosenfeld JA, Liu P, Ek M, et al. 2024. Leveraging the T2T assembly to resolve rare and pathogenic inversions in reference genome gaps. *Genome Res* **34:** 1785–1797. doi:10.1101/gr.279346.124

Blennow E, Telenius H, Larsson C, de Vos D, Bajalica S, Ponder BA, Nordenskjöld M. 1992. Complete characterization of a large marker chromosome by reverse and forward chromosome painting. *Hum Genet* **90:** 371–374. doi:10.1007/BF00220461

Blennow E, Annerén G, Bui TH, Berggren E, Asadi E, Nordenskjöld M. 1993. Characterization of supernumerary ring marker chromosomes by fluorescence in situ hybridization (FISH). *Am J Hum Genet* **53:** 433–442.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18:** 170–175. doi:10.1038/s41592-020-01056-5

Conlin LK, Kramer W, Hutchinson AL, Li X, Riethman H, Hakonarson H, Mulley JC, Scheffer IE, Berkovic SF, Hosain SA, et al. 2011. Molecular analysis of ring chromosome 20 syndrome reveals two distinct groups of patients. *J Med Genet* **48:** 1–9. doi:10.1136/jmg.2010.080382

Du H, Jolly A, Grochowski CM, Yuan B, Dawood M, Jhangiani SN, Li H, Muzny D, Fatih JM, Coban-Akdemir Z, et al. 2022. The multiple de novo copy number variant (MdnCNV) phenomenon presents with peri-zygotic DNA mutational signatures and multilocus pathogenic variation. *Genome Med* **14:** 122. doi:10.1186/s13073-022-01123-w

Eisfeldt J, Vezzi F, Olason P, Nilsson D, Lindstrand A. 2017. TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000Res* **6:** 664. doi:10.12688/f1000research.11168.1

Eisfeldt J, Ameur A, Lenner F, Ten Berk de Boer E, Ek M, Wincent J, Vaz R, Ottosson J, Jonson T, Ivarsson S, et al. 2024. A national long-read sequencing study on chromosomal rearrangements uncovers hidden complexities. *Genome Res* **34:** 1774–1784. doi:10.1101/gr.279510.124

Grochowski CM, Gu S, Yuan B, Tcw J, Brennand KJ, Sebat J, Malhotra D, McCarthy S, Rudolph U, Lindstrand A, et al. 2018. Marker chromosome genomic structure and temporal origin implicate a chromoanasynthesis event in a family with pleiotropic psychiatric phenotypes. *Hum Mutat* **39:** 939–946. doi:10.1002/humu.23537

Guilherme RS, Ayres Meloni VF, Kim CA, Pellegrino R, Takeno SS, Spinner NB, Conlin LK, Christofolini DM, Kulikowski LD, Melaragno MI. 2011. Mechanisms of ring chromosome formation, ring instability and clinical consequences. *BMC Med Genet* **12:** 171. doi:10.1186/1471-2350-12-171

Heller D, Vingron M. 2021. SVIM-asm: structural variant detection from haploid and diploid genome assemblies. *Bioinformatics* **36:** 5519–5521. doi:10.1093/bioinformatics/btaa1034

Kurtas N, Arrigoni F, Errichiello E, Zucca C, Maghini C, D'Angelo MG, Beri S, Giorda R, Bertuzzo S, Delledonne M, et al. 2018. Chromothripsis and ring chromosome 22: a paradigm of genomic complexity in the Phelan-McDermid syndrome (22q13 deletion syndrome). *J Med Genet* **55:** 269–277. doi:10.1136/jmedgenet-2017-105125

Kurtas NE, Xumerle L, Leonardelli L, Delledonne M, Brusco A, Chrzanowska K, Schinzel A, Larizza D, Guerneri S, Natacci F, et al. 2019. Small supernumerary marker chromosomes: a legacy of trisomy rescue? *Hum Mutat* **40:** 193–200. doi:10.1002/humu.23683

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34:** 3094–3100. doi:10.1093/bioinformatics/bty191

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760. doi:10.1093/bioinformatics/btp324

Li P, Dupont B, Hu Q, Crimi M, Shen Y, Lebedev I, Liehr T. 2022. The past, present, and future for constitutional ring chromosomes: a report of the international consortium for human ring chromosomes. *HGG Adv* **3:** 100139. doi:10.1016/j.xhgg.2022.100139

Liao WW, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, et al. 2023. A draft human pangenome reference. *Nature* **617:** 312–324. doi:10.1038/s41586-023-05896-x

Liehr T, Weise A. 2007. Frequency of small supernumerary marker chromosomes in prenatal, newborn, developmentally retarded and infertility diagnostics. *Int J Mol Med* **19:** 719–731. doi:10.3892/ijmm.19.5.719

Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, Leong IUS, Smith KR, Gerasimenko O, Haraldsdottir E, et al. 2019. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet* **51:** 1560–1565. doi:10.1038/s41588-019-0528-2

Martin M, Ebert P, Marschall T. 2023. Read-based phasing and analysis of phased variants with WhatsHap. *Methods Mol Biol* **2590:** 127–138. doi:10.1007/978-1-0716-2819-5_8

Mastrorosa FK, Daponte A, Wertz J, Rozanski AN, Harvey WT, Porubsky D, Knuth J, Garcia G, Ayllon M, Munson KM, et al. 2025. Complete chromosome 21 centromere sequencing of families with down syndrome reveals centromere size asymmetry. bioRxiv doi:10.1101/2024.02.25.581464

Matsubara K, Yanagida K, Nagai T, Kagami M, Fukami M. 2020. De novo small supernumerary marker chromosomes arising from partial trisomy rescue. *Front Genet* **11:** 132. doi:10.3389/fgene.2020.00132

Mostovoy Y, Boone PM, Huang Y, Garimella KV, Tan KT, Russell BE, Salani M, de Esch CEF, Lemanski J, Curall B, et al. 2024. 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. doi:10.1016/j.ajhg.2024.10.006

Nazaryan-Petersen L, Eisfeldt J, Pettersson M, Lundin J, Nilsson D, Wincent J, Lieden A, Lovmar L, Ottosson J, Gacic J, et al. 2018. Replicative and non-replicative mechanisms in the formation of clustered CNVs are indicated by whole genome characterization. *PLoS Genet* **14:** e1007780. doi:10.1371/journal.pgen.1007780

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. 2022. The complete sequence of a human genome. *Science* **376:** 44–53. doi:10.1126/science.abj6987

Paulin LF, Fan J, O'Neill K, Pleasance E, Porter VL, Jones SJM, Sedlazeck FJ. 2025. Closing the gaps, and improving somatic structural variant analysis and benchmarking using CHM13-T2 T. *Genome Res* **35:** 621–631. doi:10.1101/gr.279352.124

Perez G, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee CM, et al. 2025. The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res* **53:** D1243–D1249. doi:10.1093/nar/gkae974

Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT, et al. 2018. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* **36:** 983–987. doi:10.1038/nbt.4235

Porubsky D, Höps W, Ashraf H, Hsieh P, Rodriguez-Martin B, Yilmaz F, Ebler J, Hallast P, Maggiolini M, Harvey FA, et al. 2022. Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell* **185:** 1986–2005.e26. doi:10.1016/j.cell.2022.04.017

Porubsky D, Harvey WT, Rozanski AN, Ebler J, Höps W, Ashraf H, Hasenfeld P, Human Pangenome Reference Consortium (HPRC), and , Human Genome Structural Variation Consortium (HGSVC), Paten B, et al. 2023. Inversion polymorphism in a complete human genome assembly. *Genome Biol* **24:** 100. doi:10.1186/s13059-023-02919-8

Pristyazhnyuk IE, Menzorov AG. 2018. Ring chromosomes: from formation to clinical potential. *Protoplasma* **255:** 439–449. doi:10.1007/s00709-017-1165-1

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842. doi:10.1093/bioinformatics/btq033

Reddy KS, Aradhya S, Meck J, Tiller G, Abboy S, Bass H. 2013. A systematic analysis of small supernumerary marker chromosomes using array CGH exposes unexpected complexity. *Genet Med* **15:** 3–13. doi:10.1038/gim.2012.78

Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, et al. 2015. Clingen: the clinical genome resource. *N Engl J Med* **372:** 2235–2242. doi:10.1056/NEJMsr1406261

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29:** 24–26. doi:10.1038/nbt.1754

Schuy J, Sæther KB, Lisfeld J, Ek M, Grochowski CM, Lun MY, Hastie A, Rudolph S, Fuchs S, Neveling K, et al. 2024. A combination of long- and short-read genomics reveals frequent p-arm breakpoints within chromosome 21 complex genomic rearrangements. *Genet Med Open* **2:** 101863. doi:10.1016/j.gimo.2024.101863

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz CM. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* **15:** 461–468. doi:10.1038/s41592-018-0001-7

Storchová Z, Kloosterman WP. 2016. The genomic characteristics and cellular origin of chromothripsis. *Curr Opin Cell Biol* **40:** 106–113. doi:10.1016/j.ceb.2016.03.003

Ten Berk de Boer E, Ameur A, Bunikis I, Ek M, Stattin EL, Feuk L, Eisfeldt J, Lindstrand A. 2024. Long-read sequencing and optical mapping generates near T2T assemblies that resolves a centromeric translocation. *Sci Rep* **14:** 9000. doi:10.1038/s41598-024-59683-3

Weber A, Liehr T, Al-Rikabi A, Bilgen S, Heinrich U, Schiller J, Stumm M. 2022. The first neocentric, discontinuous, and complex small supernumerary marker chromosome composed of seven euchromatic blocks derived from five different chromosomes. *Biomedicines* **10:** 1102. doi:10.3390/biomedicines10051102

Received July 9, 2025; accepted in revised form March 2, 2026.