

A national long-read sequencing study on chromosomal rearrangements uncovers hidden complexities

Jesper Eisfeldt,^{1,2,3,14} Adam Ameer,^{4,5,14} Felix Lenner,^{4,5} Esmee Ten Berk de Boer,^{1,2,3} Marlene Ek,^{1,2} Josephine Wincent,^{1,2} Raquel Vaz,¹ Jesper Ottosson,⁶ Tord Jonson,^{7,8} Sofie Ivarsson,^{7,8} Sofia Thunström,⁶ Alexandra Topa,⁶ Simon Stenberg,⁶ Anna Rohlin,^{6,9} Anna Sandestig,¹⁰ Margareta Nordling,^{10,11} Pia Palmebäck,¹⁰ Magnus Burstedt,¹² Frida Nordin,¹³ Eva-Lena Stattin,⁴ Maria Sobol,^{4,15} Panagiotis Baliakas,^{4,5} Marie-Louise Bondeson,^{4,5} Ida Höijer,^{4,5} Kristine Bilgrav Saether,^{1,3} Lovisa Lovmar,⁶ Hans Ehrencrona,^{7,8} Malin Melin,^{4,5} Lars Feuk,^{4,5,14} and Anna Lindstrand^{1,2,14}

¹Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, 171 77 Stockholm, Sweden; ²Department of Clinical Genetics and Genomics, Karolinska University Hospital, 171 76 Stockholm, Sweden; ³Science for Life Laboratory, Karolinska Institutet Science Park, 171 65 Solna, Sweden; ⁴Department of Immunology, Genetics and Pathology, Uppsala University, 751 85 Uppsala, Sweden; ⁵Science for Life Laboratory, Uppsala University, 752 37 Uppsala, Sweden; ⁶Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, 413 90 Gothenburg, Sweden; ⁷Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, 221 84 Lund, Sweden; ⁸Department of Clinical Genetics, Pathology and Molecular Diagnostics, Office for Medical Services, Region Skåne, 223 62 Lund, Sweden; ⁹Department of Laboratory Medicine, Institute for Biomedicine, Sahlgrenska Academy, University of Gothenburg, 405 30 Gothenburg, Sweden; ¹⁰Department of Clinical Genetics, Linköping University Hospital, 581 85 Linköping, Sweden; ¹¹Division of Cell and Neurobiology, Department of Biomedical and Clinical Sciences, Linköping University, 581 83 Linköping, Sweden; ¹²Department of Medical Bioscience, Medical and Clinical Genetics, ¹³Department of Pharmacology and Clinical Neurosciences, Umeå University, 901 87 Umeå, Sweden

Clinical genetic laboratories often require a comprehensive analysis of chromosomal rearrangements/structural variants (SVs), from large events like translocations and inversions to supernumerary ring/marker chromosomes and small deletions or duplications. Understanding the complexity of these events and their clinical consequences requires pinpointing breakpoint junctions and resolving the derivative chromosome structure. This task often surpasses the capabilities of short-read sequencing technologies. In contrast, long-read sequencing techniques present a compelling alternative for clinical diagnostics. Here, Genomic Medicine Sweden—Rare Diseases has explored the utility of HiFi Revio long-read genome sequencing (lrGS) for digital karyotyping of SVs nationwide. The 16 samples from 13 families were collected from all Swedish healthcare regions. Prior investigations had identified 16 SVs, ranging from simple to complex rearrangements, including inversions, translocations, and copy number variants. We have established a national pipeline and a shared variant database for variant calling and filtering. Using lrGS, 14 of the 16 known SVs are detected. Of these, 13 are mapped at nucleotide resolution, and one complex rearrangement is only visible by read depth. Two Chromosome 21 rearrangements, one mosaic, remain undetected. Average read lengths are 8.3–18.8 kb with coverage exceeding 20× for all samples. De novo assembly results in a limited number of phased contigs per individual (N50 6–86 Mb), enabling direct characterization of the chromosomal rearrangements. In a national pilot study, we demonstrate the utility of HiFi Revio lrGS for analyzing chromosomal rearrangements. Based on our results, we propose a 5-year plan to expand lrGS use for rare disease diagnostics in Sweden.

[Supplemental material is available for this article.]

Although short-read (sr) genomic analysis approaches such as exome and genome sequencing (GS) have been highly successful

in identifying disease-causing genetic variants for diagnostic purposes, the primary focus of analysis remains on single-nucleotide variations (SNVs) and small insertions–deletions (INDELs) (Stranneheim et al. 2021). In contrast, the calling and interpretation of structural chromosomal rearrangements is more

¹⁴These authors contributed equally to this work.

¹⁵Present address: SoboLAB Analytics and Consulting, 751 23 Uppsala, Sweden.

Corresponding authors: anna.lindstrand@ki.se; lars.feuk@igp.uu.se

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.279510.124>. Freely available online through the *Genome Research* Open Access option.

© 2024 Eisfeldt et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

challenging (Cameron et al. 2019; Kosugi et al. 2019). Such events, collectively called structural variants (SVs), are defined as genetic variants larger than 50 bp. When SVs involve a repositioning of genetic material either within a chromosome or between chromosomes, they are also referred to as chromosomal rearrangements. These encompass both recurrent and nonrecurrent copy number variations (CNVs, deletions, and duplications) and balanced chromosome abnormalities (translocations, inversions, and insertions). Furthermore, there is growing evidence that complex SVs, which contain multiple breakpoint junctions (BPJs) or consist of more than one simple structural variation in *cis*, are more prevalent than previously assumed (Schuy et al. 2022). A thorough analysis of such events detailing the DNA breakpoints at the nucleotide level is crucial for comprehending the disease-causing and rearrangement-generating mechanisms. This level of understanding is necessary for effective personalized clinical management and genetic counseling (Tesi et al. 2023).

While srGS is well established as a first-line diagnostic test, with the potential to capture pathogenic SVs previously identified by traditional methods (Lindstrand et al. 2019), it still has limitations. The existing srGS SV pipelines lack the ability to produce high-quality genome assemblies necessary to resolve complex disease-causing SVs. In this context, long-read genome sequencing (lrGS) has emerged as a promising alternative with the potential to capture the complete scope of structural genomic variation in the human genome (Audano et al. 2019; Eisfeldt et al. 2019, 2021). Indeed, lrGS provides a comprehensive approach for characterizing various types of SVs identified in a clinical genetic laboratory, whether discovered through traditional methods or srGS. With the introduction of lrGS using highly accurate consensus reads (HiFi reads), it has been demonstrated that SNV and INDEL calls have precision and recall rates that match or exceed what is obtained from srGS (Wenger et al. 2019; Mahmoud et al. 2024). In addition, HiFi reads dramatically increase the ability to detect SVs (Mahmoud et al. 2024). Due to the several kb long sequence reads, lrGS offers unique possibilities to study regions in the human genome that are not easily captured by other genomics technologies, such as highly repetitive or homologous regions. Furthermore, sequencing of the native DNA also provides methylation information. With decreasing cost, it is increasingly attractive to use lrGS in unsolved rare disease cases. Yet, lrGS differs markedly from short-read sequencing and various factors should be considered before its clinical implementation. These include the quality and quantity of DNA starting material, intricate specifics related to sequencing including read quality, total coverage, and read length, as well as the complexities tied to data analysis, visualization, and the eventual clinical interpretation.

Recognizing the promise and challenges of lrGS, we embarked on a national pilot study using the PacBio Revio system. We leveraged Genomic Medicine Sweden (GMS), a national initiative designed to implement precision medicine throughout the regionally organized healthcare system (Fioretos et al. 2022) with the aim to validate lrGS for clinical digital karyotyping of unresolved SVs nationwide. The included samples originated from all university hospital regions in Sweden and the DNA was not necessarily collected using lrGS-optimized protocols. This approach mirrored the typical sample quality and preparation found in most of Sweden's hospitals and biobanks today, thereby giving us a realistic view of the potential and challenges of lrGS in clinical diagnostics.

Results

High-quality long-read genome data are obtained from clinical DNA samples

All individuals were sequenced on one Revio 25M ZMW SMRT Cell, generating at least 20× coverage of high-quality (HiFi) reads for all samples (range 19.9–35.5×), with mean read length ranging from 8.3 to 18.8 kb (Fig. 1A,B; Supplemental Table S1). Two samples (P8.1 and P8.2) were obtained from high-molecular-weight DNA extraction of fresh blood, enabling a more efficient size selection during library preparation. Gel-based size selection was performed not only on those two samples but also on three regular DNA samples (P11, P12, and P13). Contrary to expectations, all five samples achieved similar levels of high coverage and longer read lengths, with P13 reaching the highest overall throughput of >35× coverage and the best assembly results. This indicates that the regular DNA extractions are suitable for HiFi sequencing and can be used to generate the best possible data from a single SMRT Cell. Variant calling resulted in an average of 4.9 million SNVs and small INDELS per individual (Supplemental Tables S2, S3). Of the 4.7 M SNVs with a variant quality above 20, 96% are present in gnomAD (Supplemental Table S4; Chen et al. 2024). We further performed *de novo* assembly for all individuals and the contig N50 values are shown in Figure 1C.

Identification of chromosomal rearrangements

The lrGS analysis could identify 14 of the 16 unique chromosomal rearrangements that were present in the 16 individuals sequenced. Of the two rearrangements that fully eluded detection, one was mosaic (P10) and the second affected the acrocentric p-arm of Chromosome 21, a known repetitive genomic region (P13) (Table 1). The mosaic tri- and tetrasomy on Chromosome 15 in individual P2 was characterized as a copy number gain (between 2× and 5×) in the long-read CNV-calls, although a clear distinction between the tri- and tetrasomy could not be made. While the starting point could not be distinguished from the centromere, the location of the end point was approximated between Chr15:32176000–32346000, with the exact breakpoint located within a segmental duplication (Supplemental Fig. S1). The remaining samples were then subdivided depending on whether they carried an interchromosomal or intrachromosomal rearrangement.

Translocations and inversions

Three translocations and one inversion were characterized in four individuals (Fig. 2; Table 1) with a mother and daughter both carrying the same two balanced events (P7.1 and P7.3). All variants were fully resolved, but the translocation between Chromosome X and Chromosome 9 (P9) was only detectable using the Telomere-to-Telomere (T2T) assembly for analysis. A secondary analysis in hg38 revealed a call of a t(X;5) in that sample, indicating that the breakpoint 9 region is missing in GRCh38. In individual P1 the three SVs (two deletions and one duplication) detected before lrGS were in fact two separate rearrangements; one recurrent 16p deletion and one unbalanced translocation between Chromosome 4 and Chromosome 9 (Fig. 2). Finally, we were able to delineate a complex chromosome translocation involving Chromosomes 1, 4, and 6 (P12). The analysis detected four BPJs, a 13.8 Mb deletion on Chromosome 4 as well as an independent

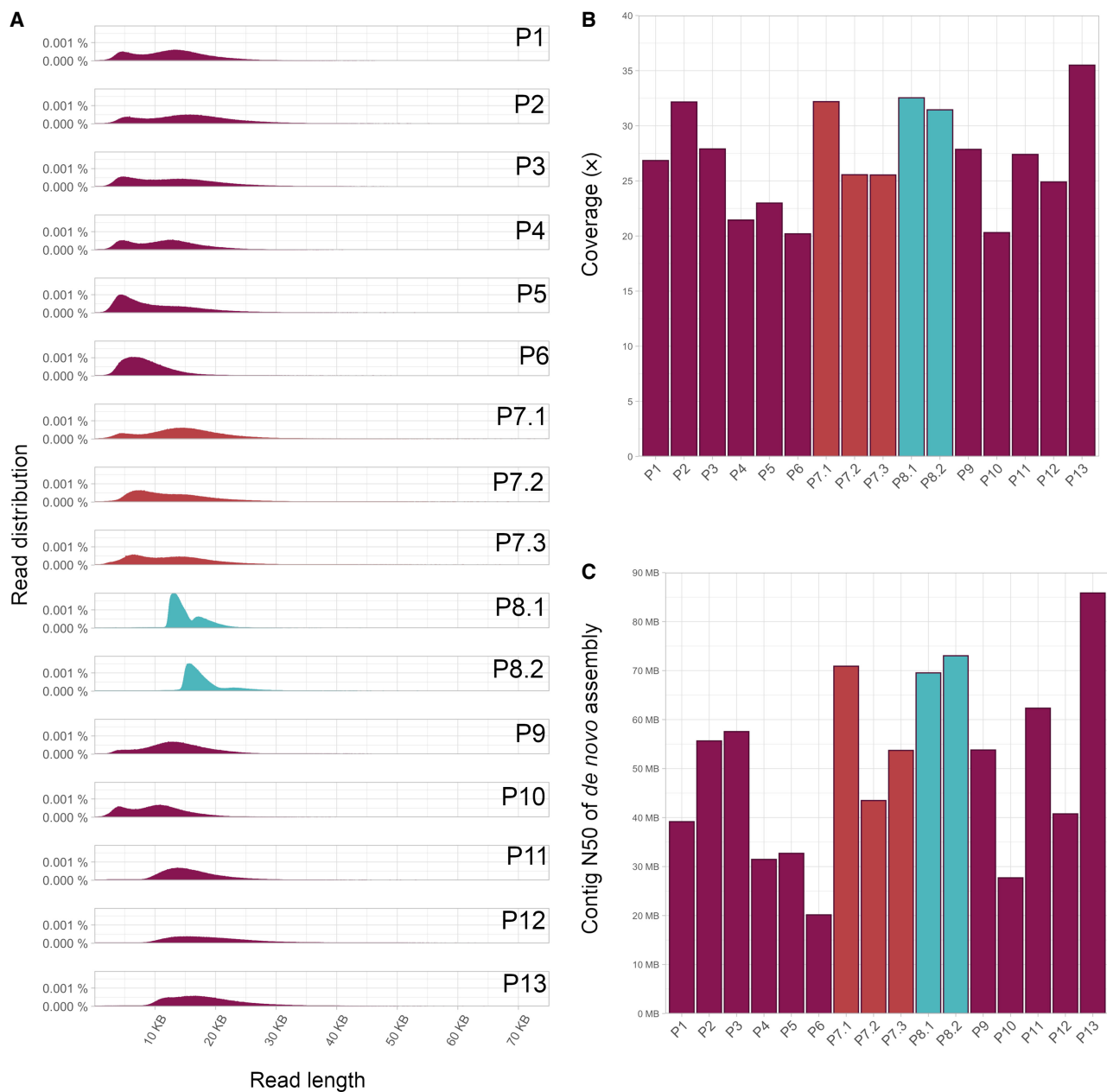


Figure 1. Quality measures of IrGS. (A) Read length distribution, (B) coverage, and (C) N50 of de novo assembly.

1.2 Mb deletion on Chromosome 6 located centromeric of the Chromosome 6 breakpoint (Fig. 2; Table 1).

Complex intrachromosomal rearrangements

In addition to the complex translocation discussed above, six intrachromosomal complex rearrangements were detected and resolved in five unrelated individuals, all of which were unbalanced (Table 1). Notably, the same inv(X) was detailed in a mother and daughter pair (P8.1 and P8.2). The highest complexity level was identified in P4 where two complex events were identified, one DUP-TRIP-QUAD-TRIP-DUP-DEL and one DEL-INV-DEL (Fig. 3A; Table 1). The complexity found in P4 was tightly followed by P5, who carries a complex rearrangement consisting of two duplica-

tions, one deletion, as well as an inverted copy number neutral segment (Fig. 3B; Table 1). In contrast, the simplest rearrangements consisted of DEL-INV-DEL (P3 and P4) or DEL-INS-DEL (P6) (Supplemental Fig. S2). Finally, we resolved a terminal invDUP-NML-DEL on Chromosome X (P11), mediated by matching *AluY* elements and harboring a 291 long stretch of microhomology (Table 2).

Breakpoint junction analysis

In 13 of the rearrangements, the BPJs were characterized at nucleotide resolution, allowing for a detailed analysis of insertions, repeat elements and microhomology, and disrupted genes (Table 2; Supplemental Table S5). Comparing the Sniffles SV calls

Table 1. Rearrangement details and clinical features of patients included in the present study

Case ID	SV structure	ISCN 2020/HGVS nomenclature	Main phenotype	Clinical assessment
P1	Unbalanced translocation, DEL	seq[GRCh38] der(4)t(4;9)(q34.3;p23),del(16)(p12.1p11.2) NC_000004.12:g.181414917_qterdelins [NC_000009.12:g.pter_13043905] NC_000016.10:g.28455071_29508000del	Mild ID and obesity. Strabismus (exotropia), hypermetropia, and astigmatism. Features include a large forehead, narrow eye slits, prominent labial furrows, and skin hyperpigmentation.	Pathogenic/ pathogenic
P2	Isodicentric chromosome	seq[GRCh38] 15q11.1q13.2 (19799420_30095350x3-4)dn,15q13.2q13.3 (30521460_32201830x2-3)dn	Suspected global developmental delay. Epilepsy, spasms.	Pathogenic
P3	DEL-INV-DEL	seq[GRCh38] 22q13.33q13.33 (50624868_50626276)x0 NC_000022.11:g.50618055_50626277delins [NC_000022.11:g.50624361_50624868inv] NM_000487.6(ARSA):c.[857_*277del]; [857_*277del] p.(Pro286_Ala509delins8)	Normal development in the first year. Progressive loss of functions, muscle weakness, areflexia, ptosis, and swallowing difficulties in the second year.	Pathogenic
P4	DUP-TRIP-QUAD-TRIP-DUP-NML-DEL, DEL-INV-DEL	seq[GRCh38] der(2)ins(2)(p25.1p25.2p25.2)ins(2)(p25.1p25.1p25.2)ins(2)(p25.1p25.2p25.2)del(p25.1p25.1),del(q21.2)ins(2)(q21.2q22.1q21.2) NC_000002.12:g.7246507_7734100delins [g.5620671_6248808inv;g.5549092_7241289inv; g.5858296_6246303], g.132954122_141163928delins [g.133434126_137703180inv]	Syndromic craniofacial condition	Likely pathogenic/ likely pathogenic
P5	DEL-INV-NML-DUP-NML-DUP	seq[GRCh38] der(3)del(3)(p14.3p14.2)inv(3)(p14.2p14.2)ins(3)(p14.2q22.2q22.2)ins(3)(p12.2q24q24) NC_000003.12:g.54518907_80981660delins [g.59407900_63614153inv; g.135767535_135891958; g.63614153_80981660inv]	Motor delay	Pathogenic
P6	Translocated insertion	seq[GRCh38] der(X)del(X)(q28q28)ins(X;9)(q28;q22.33q22.33) NC_000023.11:g.153724706_153779639delins [NC_000009.12:g.97596764_97598236inv]	Adrenoleukodystrophy	Pathogenic
P7.1	Translocation, INV	seq[GRCh38] t(1;10)(p36.2;q24),inv(2)(q32.2q33.2) NC_000001.11:g.19783172_qterdelins [NC_000010.11:g.pter_95395335] NC_000010.11:g.pter_95395328delins [NC_000001.11:g.19783105_qter] NC_000002.12:g.189694535_202576083inv	Recurrent miscarriages	Pathogenic Pathogenic
P7.2	NA	NA	Healthy father of P7.1	NA
P7.3	NA	Same as P7.1	Healthy mother of P7.1	NA
P8.1	DUP-INV-DUP	seq[GRCh38] der(X)ins(X)(p22.2q28q28)inv(X)(p22.2q28)ins(X)(q28p22.2p22.31) NC_000023.11:g.9768910_154113036delins [g.9420014_154208530inv]	Short stature	Pathogenic
P8.2	NA	Same as P8.1	Healthy, normal height	NA
P9	Translocation	47,XX,t(X;9)(p22;q12)[27]/46,X,t(X;9)(p22;q12)[3].seq[GRCh38] t(X;9)(p22.33;q21.13) NC_000009.12:g.pter_(75862011)delins [NC_000023.11:g.3044233_qter] NC_000023.11:g.3044228_qterdelins [NC_000009.12:g.pter_(75862011)]	Thrombocytopenia, early menopause, learning difficulties	Likely pathogenic
P10	Ring chromosome, DEL	46,XY,r(21)(p11q22)[9]/46,XY,del(21)(q22.3)[4]/46,XY[12]	Infertility, oligospermia	VUS
P11	DUP-NML-DEL	seq[GRCh38] der(X)ins(X)(q22p11.21p22.33)del(X)(q22q28) NC_000023.11:g.101431832_qterdelins [g.pter_55349282inv]	ID, delayed puberty	Likely pathogenic

(continued)

Table 1. Continued

Case ID	SV structure	ISCN 2020/HGVS nomenclature	Main phenotype	Clinical assessment
P12	Complex translocation	seq[GRCh38] t(1;4;6;4)(p32.2;q21.1;p22.3;q22.3q24),del(6)(p12.3p12.3) NC_000001.11:g.pter_58205785delins [NC_000004.12:g.106505092_qterinv] NC_000004.12:g.80390384_qterdelins [NC_000006.12:g.pter_20052376inv] NC_000006.12:g.20052374_pterdelins [NC_000004.12:g.94218568_106505089inv; NC_000001.11:g.20052374_pter] NC_000006.12:g.48002694_49160076 del	Neonatal hypotonia, ID, short stature	Pathogenic
P13	Unknown insertion	46,XY,add(21)(p1?3)	Infertility, oligoasthenozoospermia	VUS

(SV) structural variant, (DEL) deletion, (INV) inversion, (DUP) duplication, (TRIP) triplication, (QUAD) quadruplication, (NML) normal (copy number), (NA) not applicable, (ISCN) the international system of for human cytogenomic nomenclature, (HGVS) human genome variation society nomenclature, (DEL) deletion, (INV) inversion, (DUP) duplication, (TRIP) triplication, (QUAD) quadruplication, (NML) normal copy number, (ID) intellectual disability, (VUS) variant of uncertain significance. Unresolved rearrangements are shown in gray.

to the calls from our de novo assembly pipeline, we find that the de novo workflow detects 30 of the 31 BPJs detected by Sniffles (Table 2; Supplemental Table S5). Notably, we find a great diversity of such genetic signatures in the analyzed cases. In three rearrangements, the breakpoints contain matched repeat elements: L1 in the t(4;9) (P1) and various *Alu* elements in the inv(X) (P8.1 and P8.2) as well as in P11. In individual P4 (Table 2) microhomology of 1–11 nt was observed in five of the six BPJs and the final BPJ contains a 9 nt insertion. Insertions larger than 2 nt were detected in three additional rearrangements (Table 2); a complex translocational insertion (P6) (Supplemental Fig. S2), a t(1;10) (P7.1 and P7.3) (Fig. 2) and a t(X;9) (P9) (Fig. 2). The final two rearrangements, a DEL-INV-DEL on Chromosome 22 (P3) (Supplemental Fig. S2) and a complex DEL-INV-NML-DUP-NML-DUP on Chromosome 3 (P5) (Fig. 3B), contained neither microhomologies, matched repeats, nor larger insertions (Table 2).

Characterization of background structural variants

Next, we assessed the SV burden in the 14 unrelated samples of the GMS Rare Disease (GMS-RD) IrGS cohort. On average, we detected 16 905 SVs larger than 100 bp per individual that were subdivided into complex (0.05%), deletions (41%), duplications (1%), insertions (51%), inversions (1%), and break end (translocations, mobile elements, transpositions, and reference misalignments) (5%) (Fig. 4A; Supplemental Table S6). In general, the detected SVs were small, with 75% of SV below 1000 bp and a peak at 300 bp representing the *Alu* insertions and deletions (Fig. 4B; Supplemental Table S6). Most variants are present in more than one of the 14 unrelated individuals and only 9% are unique (Fig. 4C; Supplemental Table S6). Furthermore, very few unique SVs affect exons of 1467 genes in the panel app intellectual disability gene panel v6.0 (average 8 per individual, range 1–14). Finally, when comparing to the SweGen SV database (Ameur et al. 2017) containing background SVs in 1000 Swedish individuals assessed from srGS data, ~70% of the IrGS SV calls are novel and not present in the SweGen data set (Fig. 4D).

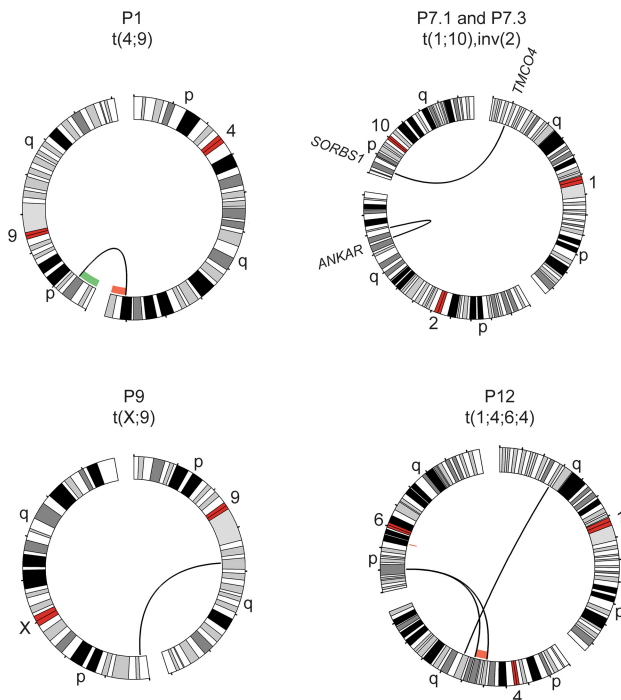


Figure 2. Translocations and inversions identified with IrGS. Circos plots of rearrangements detected in four cases using IrGS: a t(4;9) in P1, a t(1;10) and inv(2) (P7.1 and P7.3), a t(X;9) (P9), and a t(1;4;6;4) (P12). A green/red line indicates copy number gain/loss, respectively. Genes disrupted are indicated at the breakpoint.

Discussion

Discussion

In this national collaborative project, we not only demonstrated the usefulness of IrGS as a clinical follow-up analysis of SVs detected by other methods, but we have also taken the first steps toward clinical long-read diagnostics of rare diseases in our nation. By utilizing the expertise in our different regions, we have developed a comprehensive workflow covering sample collection, DNA preparation as well as sequencing, and data analysis.

We fully characterized 13 of the 16 unique chromosomal rearrangements in the GMS-RD cohort. For all cases, both an SV and a CNV (read depth) analysis were done. Specifically, IrGS missed two Chromosome 21 rearrangements (P10 and P13) and one case of mosaic partial tri- and tetrasomy on Chromosome 15

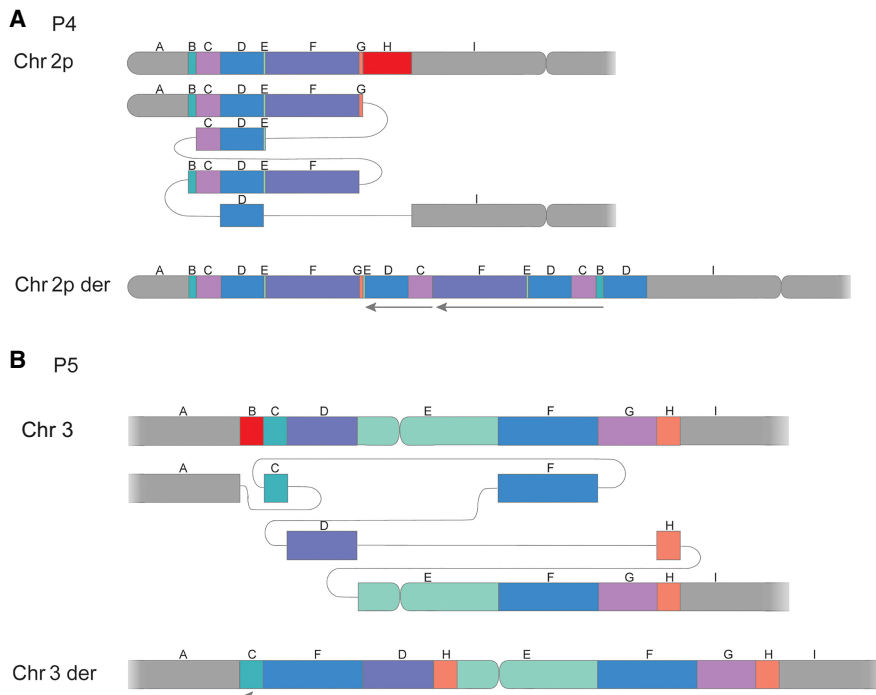


Figure 3. Subway plots of two complex rearrangements: (A) a complex dup-trip-quad-trip-dup-del on Chromosome 2 observed in P4 and (B) a clustered CNV on Chromosome 3 detected in P5. Deleted segments are shown in red, and arrows mark inverted segments.

(P2), was only revealed through read depth analysis. This underscores the necessity of including read depth callers for a comprehensive IrGS analysis. The P2 rearrangement was first detected in a clinical diagnostic laboratory by srGS, using read depth analysis, and no cytogenetic analysis was performed. However, the read depth pattern suggested an isodicentric Chromosome 15. The coverage in the two undetected events (20× and 35×, respectively, for P10 and P13) might have affected our sensitivity. In P10, the 21q deletion is estimated to be present in <20% of the cells from the array analysis and in theory should have been supported by two reads and requiring a total of three flow cells to obtain >5 reads. However, the mappability of the data may also influence the result, and other analytical approaches such as a pangenome analysis or optical genome mapping and hybrid assembly may be required to resolve the derivative chromosome structure (Eisfeldt et al. 2021; Schuy et al. 2024). This might be especially relevant in P13 (where the coverage was good) since the added genetic material is located in a repetitive region longer than the read length.

By resolving the structure of the BPJs in 13 rearrangements (Table 2), we uncovered additional complexities compared to the original analysis important for a complete clinical interpretation. In addition, we identified diverse mechanisms of SV formation. Matched *Alu* repeats were present at the BPJs in two rearrangements (P8 and P11), a pattern indicative of *Alu/Alu*-mediated rearrangements (Nazaryan-Petersen et al. 2016, 2018; Song et al. 2018; Pettersson et al. 2020). The Chromosome 16 deletion observed in P1 has both BPJs in segmental duplications as expected in a recurrent nonallelic homologous recombination (NAHR) mediated event (Carvalho and Lupski 2016). In five complex SVs from four individuals (P3, P4, P6, and P7) both microhomology and insertions were observed at the breakpoints (Table 2), features indicative of

replicative mechanisms such as Fork Stalling and Template Switching (FoSTeS)/microhomology-mediated break-induced replication (MMBIR) (Hastings et al. 2009; Carvalho and Lupski 2016). However, the largest insertion, 990 nt present at the BPJ of the t(X;9) in P9, consists mainly of TTCCA repeats that may be due to background variation in this repeat, and not caused through the formation of the translocation.

In a broader context, the current case series showcases how SVs represent a promising start for the introduction of IrGS into clinical diagnostics. However, a more complete transition to IrGS will take some time. As such, the GMS Rare Disease (GMS-RD) group has drafted a 5-year roadmap, during which new IrGS diagnostic tests successively will be implemented in Sweden (Fig. 5). To accomplish this, there will be a need to streamline the workflows for DNA extraction, library preparation, sequencing, bioinformatic analysis, and variant interpretation. Based on this pilot study, we here highlight some of the main challenges and potential solutions for the wider implementation of IrGS into clinical routine.

Starting with DNA extraction, our results show that IrGS with PacBio Revio performs well with blood samples extracted in routine diagnostics. We obtained >20× coverage for all the samples, which is likely enough for calling of both small and large variants (Harvey et al. 2023; Kucuk et al. 2023). The two samples where HMW DNA was extracted from fresh blood (P8.1 and P8.2) gave slightly more data on average, but we also obtained high yields from the three routine DNA samples (P11, P12, and P13) where we had enough DNA to perform gel-based size selection (Supplemental Table S3). For routine clinical implementation, we, therefore, propose to use the regular DNA extraction protocols that are already implemented at the hospitals, which will reduce the turn-around time and cost.

The second step is library preparation, which was performed manually. Although feasible for a small number of samples, it will not scale in clinical routine, where thousands of rare disease patients and relatives are expected to be tested yearly in Sweden (Fioretos et al. 2022). We, therefore, face the need to implement automated library preparation on liquid handling systems. This will be challenging for certain library preparation steps, such as the shearing and size selection that are not part of a typical srGS library preparation.

Regarding the long-read sequencing, the main limitations are capacity and costs. The PacBio Revio system used in this study has a theoretical capacity of ~1300 human genomes per year which is ~15 times less than the NovaSeq X platform. Nevertheless, advancements in technology may lower costs and increase throughput, allowing for more patient testing. Introducing barcoding could also enable sequencing multiple samples per run, improving efficiency. Particularly, for certain rare disease patient groups, such as those with neurological disorders caused by repeat expansions,

Table 2. Junction characteristics of identified rearrangements, one case was only detected in T2T

Sample ID	BPIs	ChrA	posA	ChrB	posB	Strand	Gene posA	Gene posB	Breakpoint insertions (nt)	Repeat A	Repeat B	Microhomology (nt)
GRCh38 (hg38)												
P1	1	4	181414917	9	13043905	+/-	-	-	-	L1	L1	-
	1	16	28455071	16	29476612	+/+	-	-	-	AluJo	SVA_D	-
P3	2	22	50618055	22	50624868	+/-	-	ARSA	2	tigger1	MIR	-
	22	50624361	22	50626277	+/-	-	ARSA	ARSA	-	Alusp	-	2
P4	4	2	5549092	2	5858296	+/-	XR_001739261.1	-	-	-	L1PA7	4
	2	5620671	2	7241289	+/+	-	-	-	9	-	-	-
	2	6246303	2	7734100	+/+	-	-	-	-	L1M5	-	2
	2	6248808	2	7246507	+/-	-	-	-	-	L1ME4b	-	3
	2	132954122	2	137703180	+/-	-	NCKAP5	-	-	MLTK	L1PA16	1
	2	133434126	2	141163928	+/-	-	NCKAP5	LRP1B	-	-	-	11
P5	5	3	54518907	3	63614153	+/-	CACNA2D3	SYNPR	-	-	MamRep38	-
	3	59407900	3	135767535	+/-	-	XR_002959675.1	-	-	-	MER101	1
	3	63614153	3	145600155	+/-	-	SYNPR	-	-	MamRep38	L2A	-
	3	80981660	3	135891958	+/-	-	-	-	-	MSTA-int	L2a	1
	3	80981660	3	148110490	+/+	-	-	-	1	MSTA-int	MERS8A	-
P6	2	9	97596764	X	153779639	+/-	TMOD1	-	-	-	-	1
	9	97598236	X	153724706	+/-	-	TMOD1	BCAP31	12	AluSX1	-	-
P7.1	4	1	19783172	10	95395335	+/-	TMCO4	SORBS1	7	MIRb	-	-
P7.3	1	19783105	10	95395328	+/-	-	TMCO4	SORBS1	5	MIRb	-	-
	2	189694535	2	202576085	+/-	-	-	ANKAR	1	-	L1PA17	-
	2	189694533	2	202576083	+/-	-	-	ANKAR	-	-	L1PA17	1
P8.1	2	X	9420014	X	154113037	+/-	XR_001755796.1	-	28	AluJr	AluSx1	-
P8.2	X	9768909	X	154208530	+/-	-	GPR143	-	31	AluSz6	AluJo	-
P11	1	X	55349282	X	101431831	+/-	-	ARMCX4	-	AluY	AluY	290
P12	5	4	80390384	6	20052376	+/-	CFAP299	-	-	L1MA9	AluSsq2	1
	4	94218568	6	20052374	+/-	-	SMARCAD1	-	-	AluSz	AluSsq2	1
	1	58205787	4	106505089	+/-	-	DAB1	-	-	-	THE1D	-
	1	58205785	4	106505092	+/-	-	DAB1	-	-	-	THE1D	2
	6	48002694	6	49160076	+/+	-	PTCHD4	-	-	L1PA4	-	3
T2T												
P9	2	9	75862011 ^a	X	3044233	+/-	-	-	-	(AATGG)n	ERVL-MaLR	2
	9	75862011 ^a	X	3044228	+/-	-	-	-	990	(TTCCA)n	ERVL-MaLR	-

(BPI) breakpoint junction, (Chr) Chromosome, (pos) position, (nt) nucleotide; bold text indicates a disease-causing gene.
^aEstimated location.

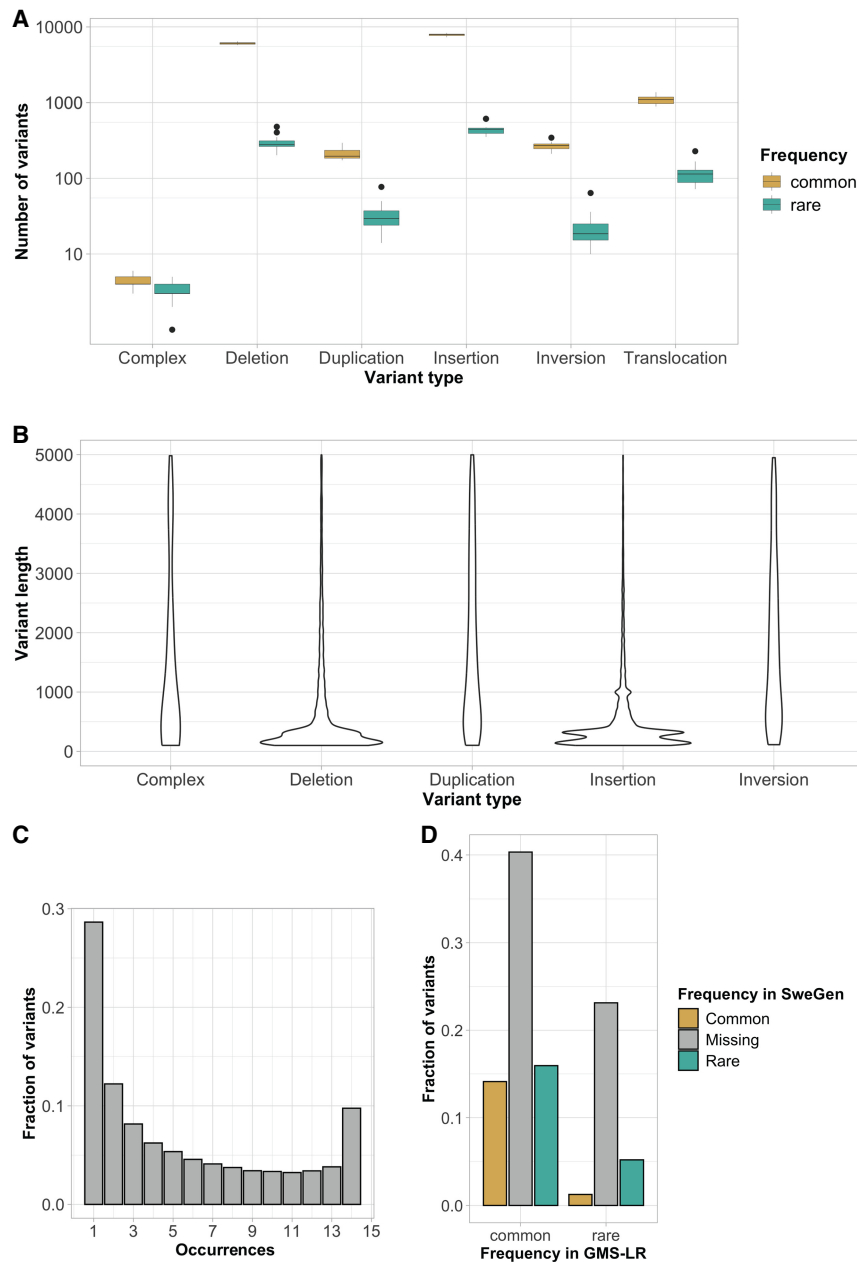


Figure 4. Characterization of background SVs. (A) Boxplot illustrating the number of SVs per type (common in yellow and rare in green). (B) Violin plot of SV length per SV type (excluding SV > 5 kbp). (C) Allele frequency histogram. (D) Comparison of allele frequencies between the GMS long-read cohort and SweGen srGS SV database (common in yellow, missing in gray, and rare in green).

targeted gene panels might be a more cost-effective approach (Höijer et al. 2018; Stevanovski et al. 2022).

Long-read sequencing generates large amounts of data that need to be processed with streamlined bioinformatics pipelines. For clinical implementation, we need robust and efficient pipelines. Moreover, since bioinformatics is a rapidly evolving research area, it is crucial that the pipelines can be modified as new software emerges. We see great value in processing samples from different hospitals with a similar analysis pipeline, since this will enable joint downstream analyses and comparisons between

larger patient groups. For this purpose, we have established a common pipeline (<https://github.com/genomic-medicine-sweden/nallo>) for the lrGS analysis of rare disease patients in Sweden. All regions in the country have access to the pipeline and can contribute to its continuous development.

Finally, the called variants undergo prioritization and interpretation. In this pilot, the focus was on SVs, and for all cases, there was prior information about the chromosomal rearrangement. In the future, when lrGS is applied as a first-line test for rare disease diagnostics the data analysis will be more challenging. In particular, there is a need to establish a national lrGS reference data set for effective filtering of nonpathogenic variants. However, even with a database containing 14 unrelated long-read genomes, we can reduce the number of unique SVs by 91% (Supplemental Table S4), but for clinical implementation, larger databases will be needed. A few lrGS databases, including the 1000 Genomes Project ONT Sequencing Consortium (IKGP-ONT) (Gustafson et al. 2024) and the Consortium of Long Read Sequencing Database (ColoRSdb, HiFi data) (<https://doi.org/10.5281/ZENODO.11511513>), are available; however, an in-house database is likely necessary for optimal filtering. To this end, we are planning to sequence a cross section of the Swedish population. The initial reference data set will consist of at least 100 individuals capturing most of the common variants in the population. Ideally, ethnic minority groups should also be included although it might be possible to obtain such data through international collaborations. In addition to a national reference data set, other lrGS resources such as HPRC (Liao et al. 2023) and 1KGP (Gustafson et al. 2024) will be valuable. It may also be necessary to utilize multiple reference genomes, or even a human pangenome graph (Liao et al. 2023), since we noticed here that T2T-CHM13 (Nurk et al. 2022) enables the detection of SV breakpoints not seen in GRCh38.

Variant calling is followed by variant prioritization, enabling the identification of the pathogenic variant(s). During this process, visualization is an important tool. The Integrative Genomics Viewer (IGV) (Robinson et al. 2011), often used for short reads, is mainly designed for the visualization of small variants (INDELs and SNVs), and for SVs identified by lrGS, alternative methods will be required. In this study, we generated subway and Circos (Krzyszowski et al. 2009) plots for this purpose, but these might not be suitable for all types of variants. Particularly, specialized tools should be implemented to

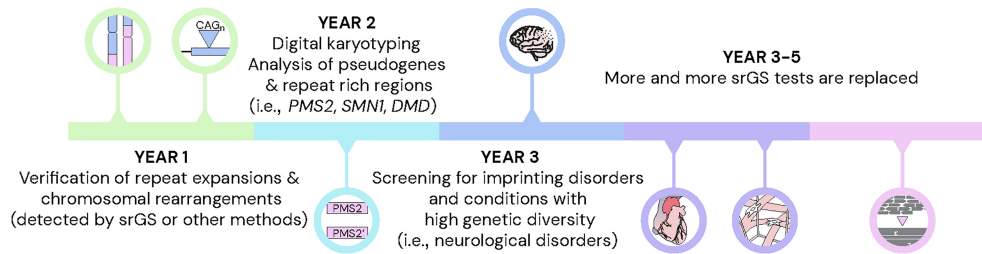


Figure 5. Toward lrGS as a first tier diagnostic test in rare disease. A 5-year time line with the expected development of long-read sequencing in the clinical setting.

visualize tandem repeats (De Coster et al. 2024; Dolzhenko et al. 2024).

When a new candidate variant has been detected and visually inspected, its role in the disease needs to be established. We expect that many of the novel variants detected through lrGS will be of unknown significance, since they have not been observed with previous genome technologies, and therefore it will be important to obtain good annotations and functional predictions. To some extent, the same tools as for srGS can be used for functional prediction, but it may be a challenge to understand the potential consequences of complex SVs, especially those located within noncoding regions. For that reason, it may be necessary to focus on the most obvious results and leave some of the uncertain diagnoses for later. As the databases and annotations grow, we can then revisit those patients and hopefully give a correct diagnosis.

Altogether, for the 14 detected rearrangements, the lrGS analysis has not yet led to a clinical reclassification (10 pathogenic, 4 likely pathogenic) (Table 1). In two cases, the SVs caused monogenic diseases; the homozygous 22q13.33 deletion in P3 results in the loss of *ARSA* (Metachromatic Leukodystrophy, MIM 250100), and the hemizygous Xq28 deletion in the loss of *ABCD1* in P6 (Adrenoleukodystrophy, MIM 300100); however, this was known before lrGS. For P10 and P13, whose Chromosome 21 rearrangements remain unresolved, it is still unclear if the clinical symptoms (Infertility and oligospermia/oligoasthenozoospermia) are a result of the cytogenetic rearrangements.

Another potential benefit of lrGS is the ability to call methylation. However, currently, it is challenging to go beyond targeted analysis of regions of interest. The horizon to replace methylation-sensitive MLPA analysis, used in imprinting disorders such as Prader–Willi syndrome (MIM 176270) or to assess skewed X-inactivation, is likely short (Johansson et al. 2023; Yamada et al. 2023; Geysens et al. 2024). However, for a more comprehensive analysis, a large control data set is required. The rare disease community will need to collaborate and share data from methylation analyses conducted with lrGS to establish informative methylation sites, as well as analysis protocols and standard operating procedures. Ultimately, the diagnostic value of methylation analysis needs to be assessed across different rare disease cohorts.

In summary, we demonstrate that by coordinating our local efforts and working together in the GMS Rare Diseases consortium, we were able to build the tools and workflows necessary to validate lrGS for digital karyotyping in the entire nation. Even though there is still quite some work to be done for the full clinical utility of lrGS, our preliminary results show that there is no question that lrGS will provide benefits for rare disease diagnostics. With the plan we have laid out, we hope to have the

methods set up and running at scale for rare diseases within a 5-year period.

Methods

Study subjects

All Swedish university regions are affiliated with the GMS working group for rare diseases (GMS-RD). In a nationwide lrGS pilot study samples from individuals with seemingly complex SVs were recruited from all healthcare regions. The multicenter study was approved by the Swedish Ethical Review Authority (2019-04746) and written informed consent was obtained from each participating individual or their respective legal guardians. Altogether, 16 samples were sequenced including 11 proband singletons, one duo, and one trio. The duo was a mother and daughter from a previously published family with several carriers of the balanced and unbalanced versions of a complex inversion (Pettersson et al. 2020). The two samples (P8.1 and P8.2) both had the balanced inversion but presented with variable severity (short stature vs. normal height). The trio included a proband with multiple miscarriages that was a carrier of an inversion and a translocation. Her healthy parents, not previously analyzed, were included. For the duo and trio high-molecular preparation had been performed before lrGS, using the Nanobind CBB kit (Pacific Biosciences) while following protocol Nanobind HMW DNA extraction—mammalian whole blood (PN 102-573-500, REV01) and the SP blood and cell culture DNA isolation kit (Bionano Genomics) following SP Blood and Cell Culture DNA Isolation Protocol v2 (document no. 30398, revision B), respectively. The remaining DNA samples were retrieved from clinical biobanks and obtained using standard extraction protocols (Supplemental Table S7). The clinical relevance of the SVs was classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines (Riggs et al. 2020). Details on the included individuals are given in Table 1.

Ethics approval and consent

The Ethical Review Board in Sweden approved the study (ethics permit number 2019-04746). Written consent to participate was provided by the subject or their legal guardians. The research conformed to the principles of the Helsinki Declaration. Written informed consent was obtained to publish.

Long-read genome sequencing

The DNA samples were fragmented to 15–20 kb using Megaruptor 3 (Diagenode). PacBio SMRTbell library construction was performed using the SMRTbell Template prep kit 3.0. SMRTbell libraries were size-selected either using AMPure beads or by the gel-based systems pippinHT (Sage Science) or SageElf (Sage Science).

The library preparation procedure is described in the protocol “Preparing whole genome and metagenome sequencing libraries using SMRTbell prep kit 3.0” from PacBio. The SMRTbell library sizes and profiles were evaluated using Fragment Analyzer (Agilent Technologies). PacBio sequencing was performed on the PacBio Revo system with 24 h movie time. Each SMRTbell library was sequenced on a 25M SMRT cell.

Analysis of long-read genome data

The samples were run using tools available in the Nallo pipeline (<https://github.com/genomic-medicine-sweden/nallo>). Additional analyses were then carried out on top of these results. Briefly, the PacBio HiFi reads were aligned to both GRCh38 and the T2T-CHM13v2.0 reference using minimap2 (version 2.26) (Li 2018) SAMtools (version 1.17) (Danecek et al. 2021). SNVs and INDELS were called with DeepVariant 1.5.0 (Poplin et al. 2018) and phased using WhatsHap 1.7 (Martin et al. 2016). SVs were called using Sniffles 1.1.12 (Sedlazeck et al. 2018), which was run according to previously published parameters (Eisfeldt et al. 2021), and CNVs were called using HiFiCNV (0.1.6b or 0.1.7) (<https://github.com/PacificBiosciences/HiFiCNV>). Quality metrics were gathered using FastQC 0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), cramino 0.9.7 (De Coster and Rademakers 2023), and mosdepth 0.3.3 (Pedersen and Quinlan 2018). Phased assemblies were generated using hifiasm (version 0.19.5-r587) (Cheng et al. 2021). The resulting de novo assemblies were aligned to hg38 and T2T-CHM13 reference using minimap2 (version 2.26) (Li 2018), SNVs were called using HTSBOX (version r345) (Danecek et al. 2021) and counted using BCFTOOLS (version 1.17) (Danecek et al. 2021), SVs were called using SVIM-asm using diploid mode for phased assemblies and haploid mode for non-phased assemblies (Heller and Vingron 2021), and quality control was performed using QUASt (version 5.0.2) (Mikheenko et al. 2018).

Identification and characterization of structural variants

Before SV characterization, the Sniffles SV calls were filtered based on size and allele frequencies. The size filtering was performed using BCFTOOLS (Danecek et al. 2021), excluding all SV calls smaller than 2 kbp. The frequency filtering was performed using SVDB (Eisfeldt et al. 2017). First, a frequency database was constructed using the calls from all 16 individuals sequenced (Table 1). Next, we annotated the remaining calls using that database, and variants present in more than one unrelated individual were removed. Lastly, the remaining SV calls were inspected using IGV (Robinson et al. 2011) and characterized as described previously (Eisfeldt et al. 2021).

Data access

All lrGS data generated in this study have been submitted to the European Genome-phenome Archive (EGA; <https://ega-archive.org>) under accession number EGAD50000000676. Code for running the lrGS analysis is available at GitHub (<https://github.com/genomic-medicine-sweden/nallo>) and as Supplemental Code.

Competing interest statement

A.L. has received honoraria from Pacific Biosciences.

Acknowledgments

We are very grateful to the participating families. We would also like to thank UPPMAX for the use of computer infrastructure resources (project sens2022541) and the support from the National Genomics Infrastructure (NGI) Uppsala and Clinical Genomics Uppsala. Several authors of this publication are members 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). We also acknowledge the Uppsala ALF office. Funding was provided by Genomic Medicine Sweden as well as grants to A.L. from the Swedish Research Council (2017-02936 and 2019-02078), the Stockholm Regional Council, the Swedish Brain Foundation/Hjärnfonden (FO2022-0256), and the Swedish Rare Diseases Foundation, and to L.F. from the Swedish Research Council (2017-01861) and the Swedish Brain Foundation/Hjärnfonden (FO2022-0207).

Authors contributions: J.E., A.A., M.M., L.L., H.E., L.F., and A.L. conceptualized the study. J.E., A.A., F.L., E.T.B.d.B., R.V., L.F., and A.L. were major contributors to writing the manuscript. M.E. and I.H. performed extraction and QC analysis of DNA. J.W., J.O., T.J., S.L., S.T., A.T., S.S., A.R., A.S., M.N., P.P., M.B., F.N., E.-L.S., M.S., P.B., M.-L.B., L.L., H.E., and A.L. provided patient samples and clinical information of patients. J.E., A.A., F.L., E.T.B.d.B., K.B.S., and L.F. performed bioinformatics analysis and development. All authors have read, edited, and approved the final manuscript.

References

- Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, Viklund J, Kähäri AK, Lundin P, Che H, et al. 2017. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. *Eur J Hum Genet* **25**: 1253–1260. doi:10.1038/ejhg.2017.130
- Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, Dougherty ML, Nelson BJ, Shah A, Dutcher SK, et al. 2019. Characterizing the major structural variant alleles of the human genome. *Cell* **176**: 663–675.e19. doi:10.1016/j.cell.2018.12.019
- Cameron DL, Di Stefano L, Papenfuss AT. 2019. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun* **10**: 3240. doi:10.1038/s41467-019-11146-4
- Carvalho CM, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* **17**: 224–238. doi:10.1038/nrg.2015.25
- Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, et al. 2024. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**: 92–100. doi:10.1038/s41586-023-06045-0
- 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
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFTOOLS. *Gigascience* **10**: giab008. doi:10.1093/gigascience/giab008
- De Coster W, Rademakers R. 2023. Nanopack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**: btad311. doi:10.1093/bioinformatics/btad311
- De Coster W, Höijer I, Bruggeman I, D’Hert S, Melin M, Ameur A, Rademakers R. 2024. Visualization and analysis of medically relevant tandem repeats in nanopore sequencing of control cohorts with pathSTR. *Genome Res* (this issue) **34**: 2074–2080. doi:10.1101/gr.279265.124
- Dolzhenko E, English A, Dashnow H, De Sena Brandine G, Mokveld T, Rowell WJ, Karniski C, Kronenberg Z, Danzi MC, Cheung WA, et al. 2024. Characterization and visualization of tandem repeats at genome scale. *Nat Biotechnol* doi:10.1038/s41587-023-02057-3
- 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, Pettersson M, Vezzi F, Wincent J, Käller M, Gruselius J, Nilsson D, Syk Lundberg E, Carvalho CMB, Lindstrand A. 2019. Comprehensive

- structural variation genome map of individuals carrying complex chromosomal rearrangements. *PLoS Genet* **15**: e1007858. doi:10.1371/journal.pgen.1007858
- Eisfeldt J, Pettersson M, Petri A, Nilsson D, Feuk L, Lindstrand A. 2021. Hybrid sequencing resolves two germline ultra-complex chromosomal rearrangements consisting of 137 breakpoint junctions in a single carrier. *Hum Genet* **140**: 775–790. doi:10.1007/s00439-020-02242-3
- Fioretos T, Wirta V, Cavellier L, Berglund E, Friedman M, Akhras M, Botling J, Ehrencrona H, Engstrand L, Helenius G, et al. 2022. Implementing precision medicine in a regionally organized healthcare system in Sweden. *Nat Med* **28**: 1980–1982. doi:10.1038/s41591-022-01963-4
- Geysens M, Huremagic B, Souche E, Breckpot J, Devriendt K, Peeters H, Buggenhout GV, Esch HV, Bogaert KVD, Vermeesch JR. 2024. Nanopore sequencing-based episinature detection. medRxiv doi:10.1101/2024.04.19.24305959
- Gustafson JA, Gibson SB, Damaraju N, Zalusky MPG, Hoekzema K, Twesigomwe D, Yang L, Snead AA, Richmond PA, De Coster W, et al. 2024. High-coverage nanopore sequencing of samples from the 1000 Genomes Project to build a comprehensive catalog of human genetic variation. *Genome Res* (this issue) **34**: 2061–2073. doi:10.1101/gr.279273.124
- Harvey WT, Ebert P, Ebler J, Audano PA, Munson KM, Hoekzema K, Porubsky D, Beck CR, Marschall T, Garimella K, et al. 2023. Whole-genome long-read sequencing downsampling and its effect on variant-calling precision and recall. *Genome Res* **33**: 2029–2040. doi:10.1101/gr.278070.123
- Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. *Nat Rev Genet* **10**: 551–564. doi:10.1038/nrg2593
- 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
- Höijer I, Tsai YC, Clark TA, Kotturi P, Dahl N, Stattin EL, Bondeson ML, Feuk L, Gyllenstein U, Ameer A. 2018. Detailed analysis of *HTT* repeat elements in human blood using targeted amplification-free long-read sequencing. *Hum Mutat* **39**: 1262–1272. doi:10.1002/humu.23580
- Johansson J, Lidéus S, Höijer I, Ameer A, Gudmundsson S, Annerén G, Bondeson ML, Wilbe M. 2023. A novel quantitative targeted analysis of X-chromosome inactivation (XCI) using nanopore sequencing. *Sci Rep* **13**: 12856. doi:10.1038/s41598-023-34413-3
- Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. 2019. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol* **20**: 117. doi:10.1186/s13059-019-1720-5
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–1645. doi:10.1101/gr.092759.109
- Kucuk E, van der Sanden B, O’Gorman L, Kwint M, Derks R, Wenger AM, Lambert C, Chakraborty S, Baybayan P, Rowell WJ, et al. 2023. Comprehensive de novo mutation discovery with HiFi long-read sequencing. *Genome Med* **15**: 34. doi:10.1186/s13073-023-01183-6
- Lake JA. 2024. Consortium of Long Read Sequencing database (CoLoRSdb). doi:10.5281/zenodo.11511512
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100. doi:10.1093/bioinformatics/bty191
- 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
- Lindstrand A, Eisfeldt J, Pettersson M, Carvalho CMB, Kvarnung M, Grigelioniene G, Anderlid BM, Bjerin O, Gustavsson P, Hammarsjö A, et al. 2019. From cytogenetics to cytogenomics: whole-genome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability. *Genome Med* **11**: 68. doi:10.1186/s13073-019-0675-1
- Mahmoud M, Huang Y, Garimella K, Audano PA, Wan W, Prasad N, Handsaker RE, Hall S, Pionzio A, Schatz MC, et al. 2024. Utility of long-read sequencing for all of us. *Nat Commun* **15**: 837. doi:10.1038/s41467-024-44804-3
- Martin M, Patterson M, Garg S, Fischer SO, Pisanti N, Klau GW, Schöenhuth A, Marschall T. 2016. WhatsHap: fast and accurate read-based phasing. bioRxiv doi:10.1101/085050:085050
- Mikheenko A, Pribelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**: i142–i150. doi:10.1093/bioinformatics/bty266
- Nazaryan-Petersen L, Bertelsen B, Bak M, Jonson L, Tommerup N, Hancks DC, Tumer Z. 2016. Germline chromothripsis driven by L1-mediated retrotransposition and Alu/Alu homologous recombination. *Hum Mutat* **37**: 385–395. doi:10.1002/humu.22953
- 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, Bizkade 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
- Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**: 867–868. doi:10.1093/bioinformatics/btx699
- Pettersson M, Grochowski CM, Wincent J, Eisfeldt J, Breman AM, Cheung SW, Krepischi ACV, Rosenberg C, Lupski JR, Ottosson J, et al. 2020. Cytogenetically visible inversions are formed by multiple molecular mechanisms. *Hum Mutat* **41**: 1979–1998. doi:10.1002/humu.24106
- Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Djamco 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
- Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, Raca G, Ritter DI, South ST, Thorland EC, et al. 2020. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med* **22**: 245–257. doi:10.1038/s41436-019-0686-8
- 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, Grochowski CM, Carvalho CMB, Lindstrand A. 2022. Complex genomic rearrangements: an underestimated cause of rare diseases. *Trends Genet* **38**: 1134–1146. doi:10.1016/j.tig.2022.06.003
- 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
- Sedlaczek FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* **15**: 461–468. doi:10.1038/s41592-018-0001-7
- Song X, Beck CR, Du R, Campbell IM, Coban-Akdemir Z, Gu S, Breman AM, Stankiewicz P, Ira G, Shaw CA, et al. 2018. Predicting human genes susceptible to genomic instability associated with *Alu/Alu*-mediated rearrangements. *Genome Res* **28**: 1228–1242. doi:10.1101/gr.229401.117
- Stevanovski I, Chintalaphani SR, Gamaarachchi H, Ferguson JM, Pineda SS, Scriba CK, Tchan M, Fung V, Ng K, Cortese A, et al. 2022. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci Adv* **8**: eabm5386. doi:10.1126/sciadv.abm5386
- Stranneheim H, Lagerstedt-Robinson K, Magnusson M, Kvarnung M, Nilsson D, Lesko N, Engvall M, Anderlid BM, Arnell H, Johansson CB, et al. 2021. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med* **13**: 40. doi:10.1186/s13073-021-00855-5
- Tesi B, Boileau C, Boycott KM, Canaud G, Caulfield M, Choukair D, Hill S, Spielmann M, Wedell A, Wirta V, et al. 2023. Precision medicine in rare diseases: what is next? *J Intern Med* **294**: 397–412. doi:10.1111/joim.13655
- Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* **37**: 1155–1162. doi:10.1038/s41587-019-0217-9
- Yamada M, Okuno H, Okamoto N, Suzuki H, Miya F, Takenouchi T, Kosaki K. 2023. Diagnosis of Prader-Willi syndrome and Angelman syndrome by targeted nanopore long-read sequencing. *Eur J Med Genet* **66**: 104690. doi:10.1016/j.ejmg.2022.104690

Received April 25, 2024; accepted in revised form September 25, 2024.