nature genetics

Perspective

https://doi.org/10.1038/s41588-025-02160-y

Toward clinical long-read genome sequencing for rare diseases

Received: 15 November 2024

Accepted: 11 March 2025

Published online: 7 May 2025

Check for updates

Jesper Eisfeldt[1,2,3], Marlene Ek[1,2], Magnus Nordenskjöld[1,2] & Anna Lindstrand[1,2]

Genetic diagnostics is driven by technological advances, forming a tight interface between research, clinic and industry, which enables rapid implementation of new technologies. Short-read genome and exome sequencing, the current state of the art in clinical genetics, can detect a broad spectrum of genetic variants across the genome. However, despite these advancements, more than half of individuals with rare diseases remain undiagnosed after genomic investigations. Long-read whole-genome sequencing (LR-WGS) is a promising technology that identifies previously difficult-to-detect variants while also enabling phasing and methylation analysis and has the potential of generating complete personal assemblies. To pave the way for clinical use of LR-WGS, the clinical genomic community must establish standardized protocols and quality parameters while also developing innovative tools for data analysis and interpretation. In this Perspective, we explore the key challenges and benefits in integrating LR-WGS into routine clinical diagnostics.

Genetic diagnostics is driven by technological advances that have a close interface between research, clinic and industry. The publication of the human genome in 2001 (refs. 1,2) started the modern era of clinical genetics, adding genome-wide screening with increasing resolution and quality to the rare disease analysis toolbox. Within 15 years, methods such as clinical microarray[3], exome sequencing[4,5] and short-read whole-genome sequencing (WGS)[6–8] were implemented, enabling more rare disease diagnoses. But even with the best available analyses, many individuals remain undiagnosed. Depending on inclusion criteria, WGS has shown diagnostic yields ranging from 20% to over 50%, leaving a substantial proportion of patients affected by rare disorders without definitive answers[6,8]. Although some may be explained by other factors, such as infections or environmental factors, the majority of the unsolved cases are likely to have genetic explanations that remain beyond the reach of current methodology and approaches. The low diagnostic yield highlights the limitations of short-read technologies, such as both balanced and complex structural variants (SVs), which are often missed by WGS[9–12]. Recent studies have demonstrated that LR-WGS can reliably detect clinically relevant variants that are challenging to identify using other technologies[12–15] and may increase the diagnostic yield by 5–20%[14,16] (Box 1).

This Perspective explores the challenges as well as possibilities of LR-WGS for rare disease diagnostics, highlighting crucial steps toward replacing the current clinical genetic methods.

Genetic variation in clinical diagnostics

The genetic landscape of rare diseases is complex, involving diverse genetic alterations ranging from single base pairs to whole chromosomes. Common disease variants include single-nucleotide variants (SNVs) and small insertions and deletions (indels), which can be disease causing when affecting coding regions or splicing. Many rare diseases also involve SVs, such as translocations, inversions, copy number variants (CNVs) and large chromosomal rearrangements, which may lead to disease through gene dosage, disruption, fusion or regulatory changes (Fig. 1). Short tandem repeat (STR) expansions, a subtype of SVs, are primarily linked to neuromuscular, neurodevelopmental and neurodegenerative disorders[17]. Some diseases, such as imprinting and certain X-linked disorders, arise from abnormal

[1]Department of Molecular Medicine and Surgery and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. [2]Department of Clinical Genetics and Genomics, Karolinska University Hospital, Stockholm, Sweden. [3]Science for Life Laboratory, Karolinska Institutet Science Park, Solna, Sweden. e-mail: anna.lindstrand@ki.se

BOX 1

Summary of LR-WGS platforms and technologies

PacBio SMRT HiFi sequencing

In the Pacific Biosciences (PacBio) single-molecule, real-time sequencing technology (SMRT) cell, DNA is circularized and sequenced in repeated passes to ensure high base-calling accuracy (>99%)[68]. Methylation is detected by a delay in the incorporation of methylated cytosines and adenines[69]. Each SMRT cell contains 25 million wells, called zero-mode wave guides, with each zero-mode wave guide capable of sequencing a single circularized DNA molecule per well. Currently, there is one PacBio sequencing platform available, the PacBio Revio, which supports sequencing of up to four SMRT cells simultaneously, generating approximately 90 Gb of data per cell with read lengths of 10–20 kbp (see the table)[11]. This makes it suitable for sequencing one human genome per SMRT cell. Additionally, various kits are available for targeted PacBio sequencing and multiplexing of multiple samples on a single SMRT cell. One such provider is Twist Bioscience.

Oxford Nanopore Technologies sequencing

The Oxford Nanopore Technologies (ONT) sequencer is unique in that it does not rely on DNA synthesis for sequencing. Instead, it sequences DNA fragments by passing them through nanopores. As DNA moves through the pore, it disrupts the flow of ions, and these changes are measured to determine the DNA sequence[70]. Methylation is detected based on variations in ion flow. The ONT sequencer is typically less accurate than PacBio HiFi, offering 90–99% accuracy depending on factors such as the choice of base caller[71] and use of error correction[72] as well as the quality and version of the pores (see the table). In addition, the quality can be improved further through so-called duplex sequencing, in which both DNA strands from a single molecule are read one after another[62] through the same pore. Unlike PacBio, ONT performs well with a wide range of molecule lengths, allowing for read lengths as small as a few hundred bases[73] or up to megabases in length[74]. ONT offers multiple platforms and flow cells. The PromethION flow cell has the highest capacity, generating approximately 120 Gb[53], making it suitable for whole human genome sequencing. By contrast, the Flongle is the smallest flow cell, producing less than 1 Gb, which is ideal for targeted sequencing. The PromethION 48 has the highest overall capacity, supporting up to 48 PromethION flow cells in a single experiment, while the MinION, with the smallest capacity, is compatible with various flow cells and can generate between 1 and 30 Gb depending on the configuration.

Emerging technologies

LR-WGS is a rapidly advancing field, with multiple novel and emerging technologies, including Roche's Sequencing by Expansion chemistry, which combines Xpandomer technology with nanopore sequencing to produce long reads with high base accuracy[75]. Other more experimental platforms include electron microscope-based long-read sequencing[76], graphene nanopores[77] and new approaches to linked and synthetic long-read sequencing[78], all of which offer exciting potential for improving read length and accuracy in genomic studies.

	PacBio Revio	ONT PromethION 48
Yield (per flow cell)	90 Gb, 120 Gb[a]	120 Gb
Maximum flow cells per run	4	48
Molecule lengths	10–20 kbp	1 kb–4 Mbp
Base accuracy	99%	90–99%
Run time	24 h	72 h

[a]SPRQ chemistry.

methylation rather than changes in the DNA sequence alone[18]. Finally, an increasing body of evidence highlights the importance of noncoding regions. Recent examples are genetic variants affecting noncoding RNAs such as *RNU4-2*, which has been found to underlie 0.41% of neurodevelopmental disorders (ReNU syndrome, MIM 620851)[19,20], and *CHASERR* (MIM 621012)[21].

The rapid evolution of genetic diagnostics from early karyotyping and chromosome banding to the development of fluorescence-based sequencing and, ultimately, WGS has transformed genetic diagnostics (Box 2). In the genomic era, variant prioritization is largely driven by factors, such as predicted protein impact, evolutionary conservation and variant frequency within the general population. In-house databases containing variants observed in the local population are essential for comparing newly observed variants with those previously identified while also helping to filter out platform-specific sequencing artifacts[22,23]. Determining causality also requires the evaluation of phenotype–genotype correlations and inheritance patterns. Detailed and standardized phenotyping is crucial for accurate clinical interpretation. Human Phenotype Ontology[24], the PanelApp[25] and Online Mendelian Inheritance in Man[26] are commonly used resources to assess whether a gene is associated with a patient's phenotype and to generate in silico gene panels. Through joint analysis of affected individuals and their parents (that is, trio analysis), de novo variants (genetic changes that arise in the affected individual and are absent in both parents) can be identified. These variants are responsible for many rare diseases, particularly those with severe phenotypes that appear sporadically in families[27].

State-of-the-art sequencing in clinical diagnostics

Exome sequencing and WGS represent the current state of the art in genetic testing for rare diseases. However, there is no consensus on whether WGS offers a substantially higher diagnostic yield than exome sequencing. Both achieve diagnostic yields of 20–50%[8,28], depending on cohorts and availability of sequencing data from parents. While both techniques effectively detect coding SNVs and large CNVs, WGS additionally detects noncoding variation and a wider range of SVs. With specialized software, WGS can also genotype STR expansions and identify variants in genes with pseudogenes, such as *SMN1* and *SMN2* (ref. 29), which are associated with spinal muscular atrophy (MIM 253300)[30]. However, the diagnostic value of such analyses varies by cohort, and noncoding variants remain challenging to interpret in clinical settings. Furthermore, many diagnostic laboratories throughout the world still struggle with implementing WGS and exome sequencing, and even those that routinely perform WGS in the clinic seldom analyze all variant types and do not fully utilize the genomic data[31–33].

Therefore, despite these advancements, a substantial proportion of patients remain undiagnosed after exome sequencing or WGS[28]. This gap includes not only undiscovered monogenic disease genes but also combinations of pathogenic variants across multiple genes, which potentially cause digenic, oligogenic or polygenic diseases.

 https://doi.org/10.1038/s41588-025-02160-y

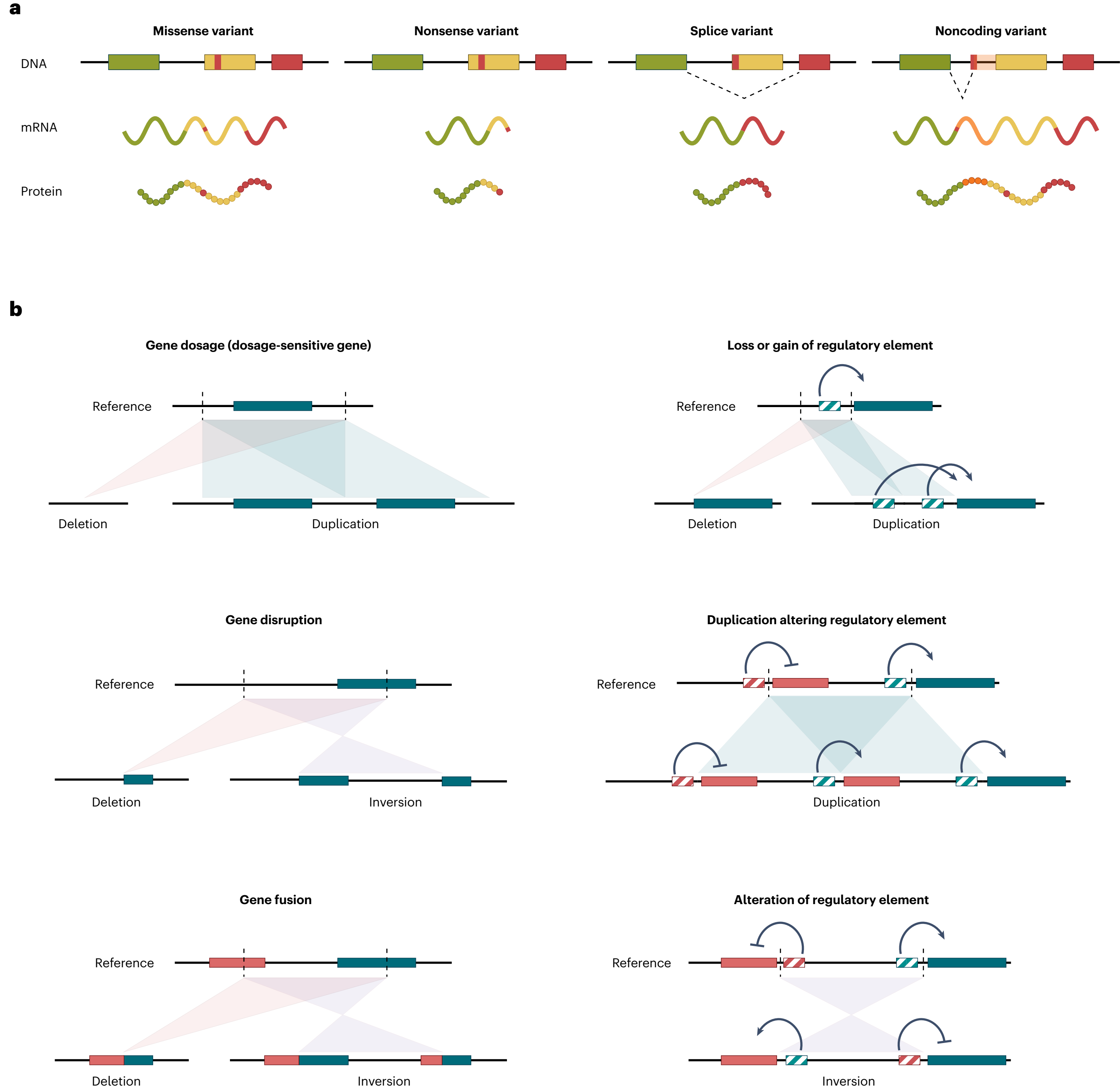


Fig. 1 | Coding and noncoding disease-causing variants. a, Schematic illustration of how small variants, that is, SNVs and indels, affect DNA, mRNA and protein. At the top, a gene is depicted with exons in colored boxes (green, yellow and red), and the variant is highlighted in red. Below, the expected RNA and protein products are shown in the same color scheme. Shown from left to right are a missense variant, a nonsense variant, a canonical splice variant (an SNV in the acceptor splice site resulting in exon skipping) and a noncoding deep intronic variant that generates a cryptic splice site, leading to the formation of a novel exon (orange). **b**, Schematic illustration of how SVs affecting both coding and noncoding regions cause disease. Left, in coding regions, deletions, duplications and inversions may cause disease through gene dosage imbalances, gene disruption or the creation of gene fusions. Right, in noncoding regions, SVs can contribute to disease by causing the gain or loss of regulatory elements or by rearranging regulatory elements, thereby altering gene expression.

Examples include Bardet–Biedl syndrome (MIM 209900), for which triallelic inheritance was first described[34,35], and facioscapulohumeral muscular dystrophy type 2 (MIM 158901), which requires both a *DUX4*-permissive haplotype and a *SMCHD1* variant[36]. Discovering more gene interactions will require large, clinically well-characterized cohorts. Noncoding variation is another reason for rare diseases remaining undiagnosed, as several recent studies highlighted the importance of variants in noncoding RNAs[19,20,37]. Other noncoding elements with potential clinical relevance include topologically associated domains, enhancers and regulatory regions that influence the expression of protein-coding genes[38,39] (Fig. 1).

To enhance diagnostic yields, the potential of multiomic approaches is being considered. A major challenge is clinical interpretation of variants of uncertain significance (VUS), such as missense variants and noncoding variants. In such cases, RNA sequencing may offer an additional 8–36% diagnostic rate after WGS and has been

 https://doi.org/10.1038/s41588-025-02160-y

BOX 2

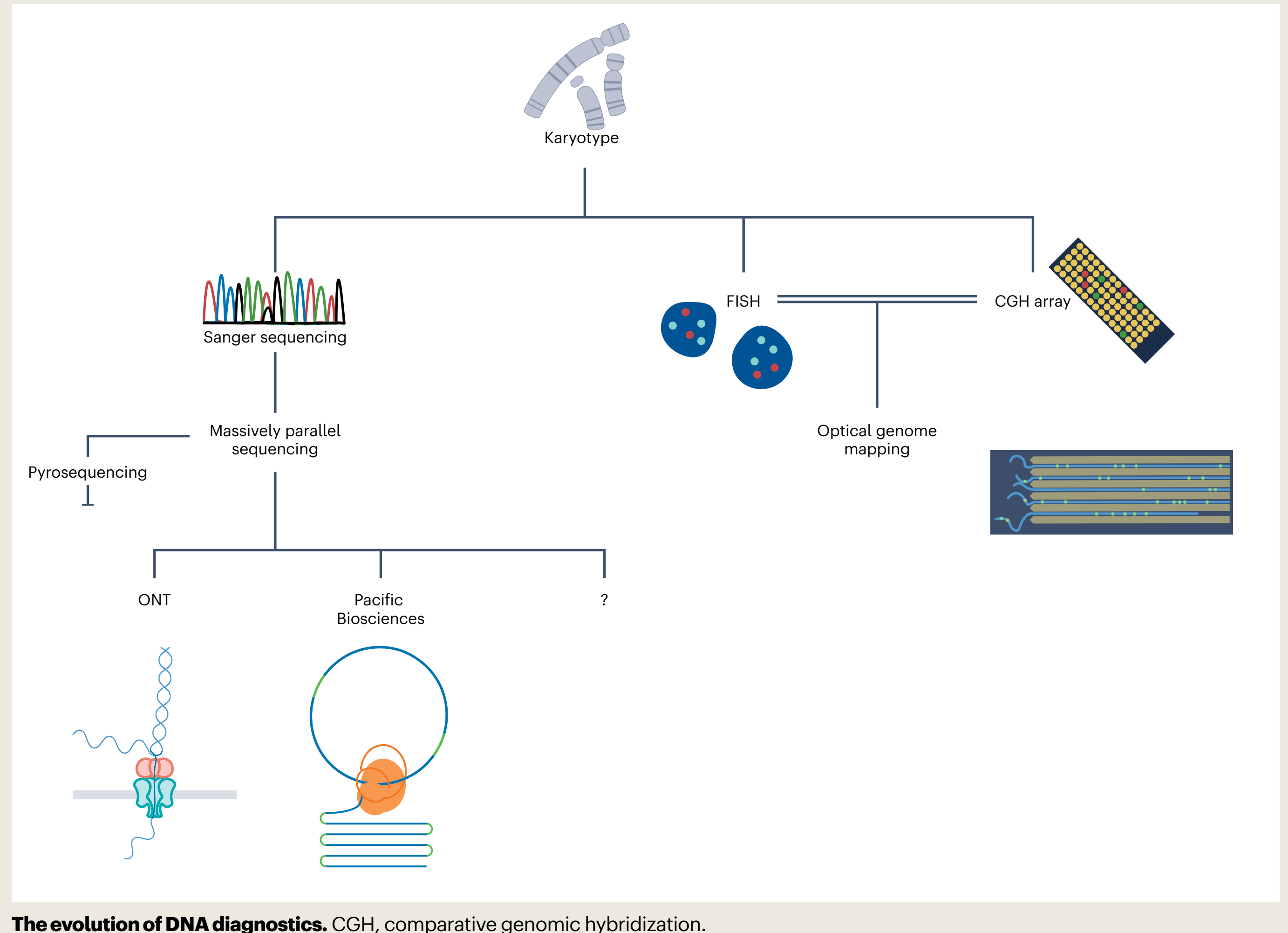
Historical perspective of genetic diagnosis

Few, if any, fields in science history have undergone such a rapid development as medical genetics since the 3D structure of DNA was discovered by Watson and Crick in 1953 (ref. 79). The genetic revolution reached healthcare rapidly, and clinical testing for genetic disorders has been part of medical service for many years. Genetic testing has evolved in three major steps, initiated by new technologies.

Soon after chromosome number in humans was established by Tjio and Levan in 1956 (ref. 80), Lejeune, Turpin and Gautier found that Down syndrome (MIM 190685) was caused by an extra chromosome[81]. This transformed medical genetics from a science based on statistics to clinical diagnosis by robust laboratory tests. During the first decade that followed, the link between aneuploidy and some additional severe developmental disorders was established. However, cytogenetics improved dramatically after chromosome banding was developed in 1970 (ref. 82). With this, structural chromosome aberrations could be linked to constitutional disorders and reach cancer research as well as clinical pathology. Although many rare diseases are caused by chromosome aberrations, karyotyping is usually limited to a resolution of 5–10 Mbp, which means that it can only detect disorders involving large-scale gene dose alterations affecting many adjacent genes.

The second step in gene testing came after discoveries during the 1980s. Most important for clinical testing was the development of labeling nucleic acids with fluorophores[83], which was the basis for automatic DNA sequencing with chain termination inhibitors[84]. With the development of fluorescent in situ hybridization (FISH), cytogenetics reached a new era in the early 1990s[85]. FISH enabled the detection of specific gene loci and gene dose in fixed tissue slices, adding a new dimension to clinical pathology by allowing spatial visualization of chromosomal abnormalities directly within tissue samples. After the identification of *CFTR*, the disease gene underlying the not-so-rare autosomal recessive disease cystic fibrosis (MIM 219700)[86], an avalanche of new Mendelian gene discoveries followed. Polymerase chain reaction and automatic DNA sequencing were the basis for positional cloning exploited in the Human Genome Project and were the backbone of the first drafts of the human genome in 2001. Closely linked to these technologies was the development of array-based methods, in which oligonucleotides are 'printed' on a solid surface to detect deletions, duplications and higher-order amplifications[87].

The third step in clinical gene testing came from advancement in computational science and the development of massively parallel sequencing around 2007. This technology uses fluorescently labeled



The evolution of DNA diagnostics. CGH, comparative genomic hybridization.

(continued from previous page)

nucleotides, polymerase chain reaction and the data from the Human Genome Project but has taken DNA sequencing to a completely new level. While the first complete human genome sequence published in 2001 was based on a handful of individuals analyzed by Sanger sequencing[1,2], today, WGS is an everyday tool in modern medicine[6–8]. The clinical impact of these methods is evident from millions of pregnant women undergoing non-invasive prenatal tests, which use low-coverage WGS to find or exclude chromosome abnormalities in the fetus. These methods are also the backbone of precision medicine for genetic disorders.

shown to be particularly effective for analyzing splice variants and deep intronic variants[40]. Proteomics is also emerging as a valuable tool in these efforts, supporting pathogenicity of specific variants or pinpointing affected pathways[41–43]. However, a clear and critical step in improving rare disease diagnostics is the implementation of LR-WGS, which provides both epigenetic and genetic information and has the potential to replace current genetic methodology[11].

Analysis of LR-WGS data

LR-WGS pipelines require specialized software for each variant type, including SNV, SV, CNV, STRs, paralogous regions and methylation (Fig. 2). Some of these analyses, such as SNV and indel calling, are relatively mature, while the detection of other variant types, such as mobile elements or variants in paralogous regions, is more challenging[44]. Variant calling is typically followed by haplotype phasing, a distinctive feature of LR-WGS data processing. All current phasing tools can phase SNVs; however, they differ in their capability to phase other variant types, such as SVs and methylation. Methylation calling is achieved alongside base calling, and methylation signatures can be manually inspected in visualization programs, such as the Integrative Genomics Viewer[45]. After variant calling and phasing, the annotation of LR-WGS variants differs depending on variant type. Most WGS annotation and visualization tools are compatible with the output from LR-WGS SNV detection tools. Furthermore, because phasing and methylation are novel features, the downstream analysis software packages, such as rank–score software and visualization tools, must be updated to fully use that information. Although there is progress in these areas, only a few tools are fully developed for clinical use[46].

Benefits of LR-WGS in clinical diagnostics

The advantages of implementing LR-WGS depend on the current setup of the diagnostic laboratory. For laboratories that are currently performing only exome sequencing, transitioning to LR-WGS may result in substantial additional benefits. By contrast, laboratories already using WGS with comprehensive variant calling may only see minimal to no additional gains in variant discovery rates. In these cases, the value of LR-WGS lies in its more thorough variant characterization[11]. Notably, certain variant types, such as STR expansions, mobile elements, inversions and translocations, often cluster in difficult-to-sequence regions, such as variable number tandem repeats and low-copy repeats, and are challenging to detect with WGS. Although the exact percentage of these variants that can be detected by LR-WGS remains unclear, LR-WGS has great promise for improving variant detection. Importantly, for clinical utility, understanding the genomic context is crucial, for example, determining how a specific Alu element is inserted and whether it disrupts a gene. Additionally, the high accuracy of LR-WGS enables a more comprehensive assessment of de novo mutations. Recent work has shown that LR-WGS detects more than twice as many de novo mutations per generation than WGS, largely due to its ability to capture all variant types[47]. This more complete analysis strengthens the evidence that a specific variant is the most likely cause of disease. Below, we highlight four areas in which the current long-read technologies have an advantage compared to short reads: phasing, SV detection, STR expansion detection and methylation profiling (Table 1).

Haplotype phasing

A key advantage of LR-WGS is its ability to phase haplotypes, which is beneficial in various clinical scenarios. For instance, when two pathogenic variants are identified in an autosomal recessive gene, phasing helps determine whether the variants are in *cis* or *trans*. This is especially important in cases where parental samples are unavailable or when one of the variants has occurred de novo. In addition, phased variants can help to distinguish between genes and their pseudogenes, facilitating the detection of both copy number alterations and SNVs. A notable example is the survival motor neuron (SMN) locus on chromosome 5q13, which includes both *SMN1* (associated with spinal muscular atrophy) and its pseudogene *SMN2*. Accurate assessment of the copy number of *SMN2* is important, as this affects disease severity and influences the treatment response to SMN-targeted therapies[48]. Furthermore, LR-WGS may also help identify silent carriers who exhibit one deleted allele and one allele with two copies of *SMN1* (ref. 48).

SV detection

Another area in which LR-WGS is particularly effective is SV calling. The longer reads obtained can bridge repetitive regions in the genome, which enables a more accurate identification and resolution of the variant structures. From historical diagnostics using karyotyping, we know that translocations and inversions are more common in individuals with intellectual disability[49,50]. However, their detection has declined with the shift toward sequencing-based diagnostics. This may be due, in part, to the widespread use of exome sequencing, which lacks the ability to detect such balanced SVs (Table 1), as well as the limited availability of bioinformatic tools for their identification. But this also highlights a limitation of short reads, as they may not span across breakpoints, an issue that can be overcome with LR-WGS[51]. Finally, LR-WGS detects substantially more SVs: approximately 20,000 SVs per human genome, compared to 6,000 with WGS[52–55]. The majority of these SVs are small deletions (<1,000 bp), mobile element indels and contractions or expansions of repeat sequences. Interpretation in a clinical setting will depend on SV location and potential impact on disease-causing genes (Fig. 1).

STR expansion detection

The ability of LR-WGS to span repetitive regions, such as variable number tandem repeats and STRs, enables detailed characterization of disease-causing STRs[46]. Specifically, it allows for haplotype-resolved sizing of expansions, identification of interrupting repeats and exploration of associated methylation patterns (primarily in noncoding STR expansions). Such an approach has already resulted in novel biological insights. One example is that pathogenic repeats are more polymorphic than previously thought, many times exhibiting variable expansion size within one individual[56]; however, the clinical relevance of this is not yet known.

Methylation profiling

In addition to base calling, LR-WGS data include direct detection of DNA methylation. This has clinical utility, particularly for diagnosing imprinting and X-linked disorders. For example, LR-WGS can detect differentially methylated regions in individuals with the known imprinting disorders Beckwith–Wiedemann syndrome (MIM 130650), Prader–Willi

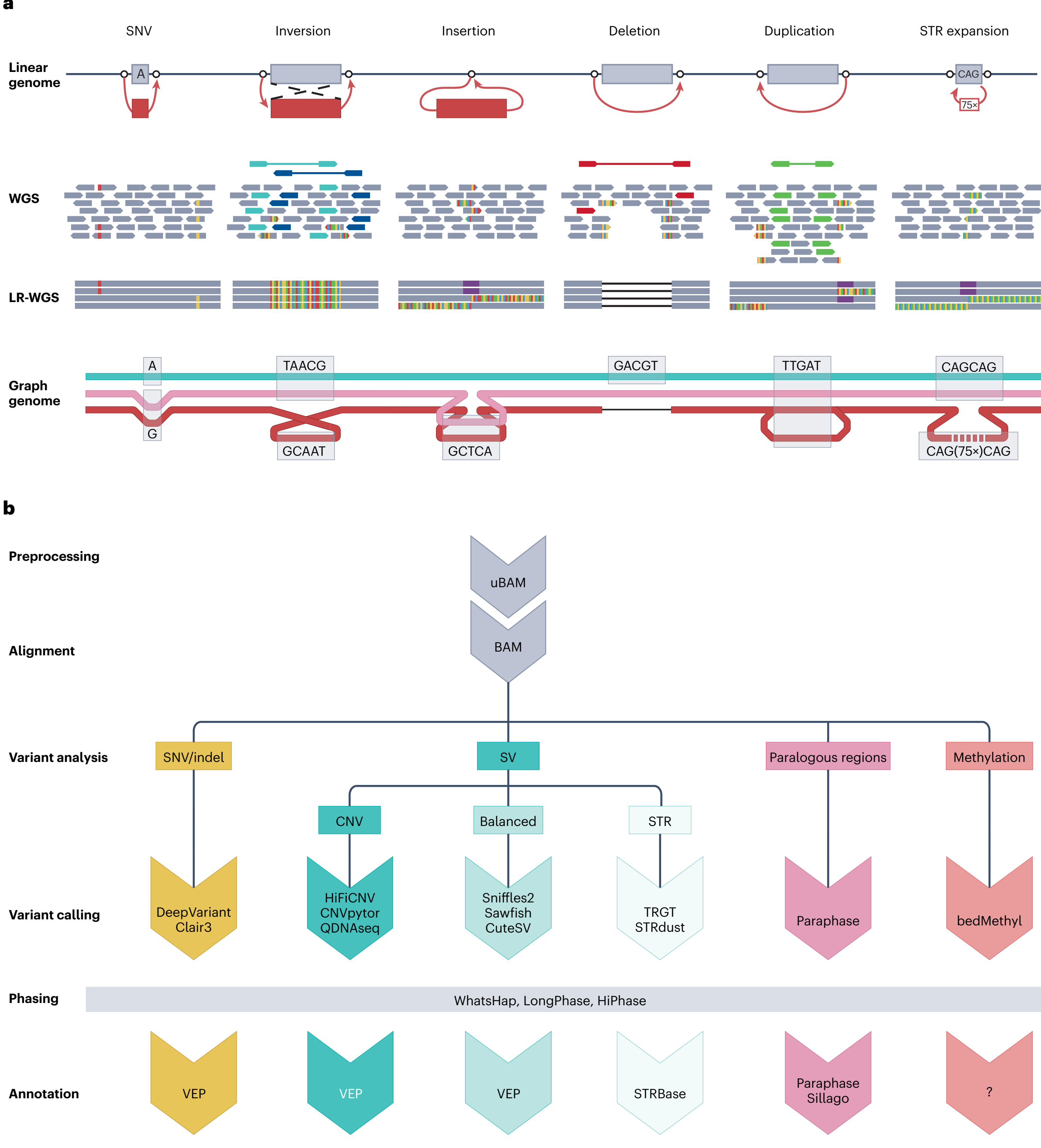


Fig. 2 | Analysis of various genetic variants. a, Illustration of different types of genetic variants, an SNV, an inversion, an insertion, a deletion, a duplication and an STR expansion, on a linear genome (top). Below, variant signatures in both short-read genome sequencing (WGS) and LR-WGS are shown. The graph genome representation at the bottom depicts the different variants on separate haplotypes (teal, pink and red). **b**, Flowchart of an LR-WGS bioinformatic pipeline for comprehensive genome analysis. The input reads, stored in unaligned binary alignment maps (uBAMs) are first aligned to the reference genome, followed by variant calling, phasing and annotation. Variant classes are color coded (yellow, teal, pink and red). See Supplementary Table 1 for currently used software packages.

syndrome (MIM 176270) and Angelman syndrome (MIM 105830)[57]. Additionally, methylation profiling can aid in the interpretation of both coding and noncoding VUS[58]. Episignatures, developed from methylation arrays, have provided promising results in identifying disease-specific epigenetic patterns[59]. These methylation changes can also be assessed using LR-WGS[60].

Table 1 | Detection capabilities of genetic methods

		Aneuploidy	Balanced SV	CNV	STR expansion	SNV/indel	Methylation
Genome wide	Karyotype	✓	✓	>5–10 Mb	–	–	–
	CMA	✓	–	>50 kb	–	–	–
	OGM	✓	✓	✓	(✓)	–	–
Targeted	FISH	✓	✓	✓	–	–	–
	Sanger sequencing	–	–	–	–	✓	–
	Fragment length	✓	–	–	✓	–	–
	MLPA	✓	–	✓	–	–	✓
MPS	Panel	–	–	(✓)	–	✓	–
	Exome sequencing	✓	–	(✓)	–	✓	–
	WGS	✓	(✓)	✓	(✓)	✓	–
LR-WGS		✓	✓	✓	✓	✓	✓

Brackets, detects in specific cases. CMA, chromosomal microarray; OGM, optical genome mapping; MLPA, multiplex ligation-dependent probe amplification; MPS, massively parallel sequencing.

Limitations of LR-WGS in clinical diagnostics

To transition toward the use of LR-WGS in routine clinical diagnostics, both technological advancements and standardization are needed across key areas, such as laboratory protocols, bioinformatic processing and variant interpretation. However, because LR-WGS data will be used for a wide range of clinical questions, the field must establish standardized quality criteria that focus on clinical outcomes and patient benefit rather than relying solely on overall parameters, such as average coverage or read length. Specific criteria for different analyses and genomic regions will be necessary, enabling maximized use of limited patient material. Even data that have low quality may be useful for validation of already detected variants[61], variant phasing or variant detection[16].

DNA extraction

To achieve optimal read lengths, high-molecular-weight DNA is generally recommended. Although clinical laboratories often favor automated methods, manual extraction may initially be required, making the process more costly and time consuming than standard automated DNA extraction. Protocols exist to isolate high-molecular-weight DNA from regular DNA by removing smaller fragments (for example, Megaruptor 3), and these are likely to become the preferred approach for clinical laboratories during the early stages of implementation[11], until automated solutions are introduced.

Cost and throughput

Although reagent and flow cell costs vary between centers, LR-WGS remains more expensive than WGS, both in terms of cost per base and cost per genome. This high cost of LR-WGS is partly due to its complex library preparation and sequencing processes, which require more manual labor than WGS, for which most steps are automated using liquid handling systems. Additionally, the low throughput (Box 1), combined with the high price of reagents and flow cells, further adds to the cost. The low throughput is also a limiting factor for clinical implementation. A facility aiming to sequence 10,000 human genomes per year would need to invest in multiple LR-WGS instruments before being able to fully transition from WGS, covering costs for both instrument procurement and maintenance.

Accuracy

LR-WGS accuracy was initially poor but has steadily improved over the years. Currently, base-calling accuracy ranges from 95% to 99% for ONT[62] and 99% for PacBio[63] (Box 1). These figures are comparable to those of WGS but slightly lower than those of newer short-read sequencers, such as the Element Biosciences AVITI and PacBio Onso, both of which achieve 99.99% base accuracy[64]. While LR-WGS accuracy is generally sufficient for analyzing germline variants, it may pose limitations in specific contexts, such as detecting mosaicism or somatic variation.

Coverage

Currently, a PacBio Revio SMRT cell or an ONT PromethION flow cell can generate approximately 40× coverage[11,53]. However, coverage will depend on factors such as multiplexing, sequencing time and DNA quality. While a coverage of 30×, or even as low as 20×, is typically sufficient for the reliable detection of germline variants, uneven coverage and coverage dropouts are commonly observed in both platforms. Notably, consistent dropouts have been reported even with high-quality DNA, including in clinically relevant regions, such as the biallelic pentanucleotide repeat expansion in *RFC1* that causes the cerebellar ataxia CANVAS (MIM 614575)[65].

Data analysis and interpretation

The analysis of LR-WGS data remains complex, and, compared to WGS, there is still a lack of validated bioinformatic pipelines for clinical-grade analysis. Additionally, well-established datasets for validation and comparing pipelines, callers and interpretation software are limited. While LR-WGS performs exceptionally well in detecting variant types, such as SVs and STR expansions[14,15], there are currently very few solutions for effective visualization and interpretation of these variants[46]. The lack of comprehensive allele frequency databases specific to LR-WGS limits the annotation and filtering of SNVs and SVs, making it challenging to reliably interpret variants from previously inaccessible genomic regions and to distinguish between sequencing artifacts and true variation. Regarding methylation, it is also important to exercise caution when interpreting signals obtained from emerging technologies. As a relatively new technology, there is still a risk of misinterpreting the resulting data. This is exemplified by the initial misinterpretation of an N^6-methyladenine signal in LR-WGS, later identified as an artifact originating from cytosine methylation rather than adenine[66].

Cost-effectiveness and funding

Today, WGS and exome sequencing are well-established clinical tests with fully developed workflows, including gene panels, defined turnaround times based on clinical urgency and established reimbursement procedures (which vary between healthcare systems). While many of these workflows could be directly adapted to LR-WGS, WGS currently remains the more cost-effective option for most rare disease applications. As sequencing efficiency and bioinformatics continue to advance, the cost-effectiveness of LR-WGS is expected to improve.

Additionally, in-depth studies on its added diagnostic yield and fully validated workflows will be essential for broader clinical adoption.

The path toward LR-WGS in routine clinical diagnostics

Despite its potential, the clinical utility of LR-WGS remains limited. Some clinical laboratories have started using targeted long-read sequencing, and, in 2024, a few genetic centers announced plans to transition toward clinical LR-WGS. Which specific tests and which patient groups to focus on initially will depend on the specific needs of the local genetic laboratory. Cost considerations must be addressed through health economic studies to identify patient groups that would benefit the most from LR-WGS. Rare diseases and especially those caused by diverse pathogenic variation such as neurological diseases are likely to be first in line.

In Sweden, to advance the clinical implementation of LR-WGS, Genomic Medicine Sweden Rare Diseases has conducted several pilot studies. In 2023, a national project demonstrated the potential of LR-WGS for resolving complex SVs[11]. Subsequently, over 500 individuals with unresolved rare diseases as well as challenging variants have been analyzed by LR-WGS (Box 1). A national LR-WGS pipeline performs comprehensive variant calling, annotation and ranking[11,67]. Local LR-WGS frequency databases are continuously populated, and ranked variants are uploaded into the open-source interpretation tool Scout, developed at the Genomic Medicine Center Karolinska[8]. In 2025, we will use this workflow and prospectively include 1,000 individuals with neurological disorders newly referred for genetic investigation who will be sequenced using long-read technologies. We will standardize both laboratory techniques and bioinformatics while providing financial resources to facilitate cross-country participation.

This initiative outlines one potential path forward toward clinical LR-WGS. This path will however vary between countries and healthcare systems. A number of clinical genetic centers are currently working toward automated processes for DNA extraction, library preparation and standardized bioinformatics. Of note, the field of LR-WGS bioinformatics is rapidly evolving, with some components already standardized, while new tools for detecting various variant types continue to be developed. Both ONT and PacBio provide pipelines for analyzing their respective data, alongside a wide range of publicly available or custom-built pipelines (Table 1 and Supplementary Table 1).

Overall, we anticipate future developments toward more accessible, cost-effective, personal genomes. However, longer reads alone will not solve the complexities of genome interpretation; progress will rely on improved analytical approaches. The use of pangenome graphs could improve SV analysis, partly by including haplotypes and sequences that may be absent from the linear reference54 (Fig. 2). De novo assembly is another promising approach for LR-WGS, allowing for improved characterization of repetitive regions[51]. While LR-WGS enables the reliable detection of complex rearrangements and noncoding variants, the lack of established disease associations poses challenges for clinical interpretation. Furthermore, integrating LR-WGS with methylation analysis and episignatures has the potential to improve variant interpretation and increase diagnostic rates for individuals with rare diseases. Nonetheless, ongoing discussions continue to explore the optimal strategies for implementing LR-WGS in genetic diagnostics of rare diseases and maximizing the benefits of this revolutionary technology.

Concluding remarks

By rapidly implementing new technologies into clinical practice, the diagnostic yields for rare diseases are steadily improving. However, even with WGS, at least half of patients with rare diseases remain undiagnosed. To improve these numbers, we need to (1) implement new methods to analyze longer reads and multiomic data and (2) bridge the gap between rare and common disorders to better understand the impact the entire genetic landscape has on disease occurrence. Large-scale collaborations to standardize interpretation and share clinical data in a findable, accessible, interoperable and reusable (FAIR) and inclusive manner will pave the way to a more rapid diagnostic odyssey for millions of patients affected by rare diseases worldwide. The ultimate goal is to rank and interpret all omics data collectively.

The field of LR-WGS is still in its infancy, and we anticipate continued rapid advancements in both laboratory techniques and bioinformatic methods in the near future. As the technology matures, we will see an increasing number of diagnostic tests transition to LR-WGS, paving the way for a future in which complete, haplotyped personal genomes replace the full range of cytogenetic and molecular genetic investigations with a single test. This will not only enhance diagnostic yield but also enable truly comprehensive genomic testing, allowing for the assessment of all variants across the entire genome.

References

1. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Miller, D. T. et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86**, 749–764 (2010).
4. Choi, M. et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl Acad. Sci. USA* **106**, 19096–19101 (2009).
5. Ng, S. B. et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
6. Gilissen, C. et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).
7. Investigators, G. P. P. et al. 100,000 genomes pilot on rare-disease diagnosis in health care — preliminary report. *N. Engl. J. Med.* **385**, 1868–1880 (2021).
8. Stranneheim, H. et al. 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 (2021).
9. Ebert, P. et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**, eabf7117 (2021).
10. Hiatt, S. M. et al. Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders. *HGG Adv.* **2**, 100023 (2021).
11. Eisfeldt, J. et al. A national long-read sequencing study on chromosomal rearrangements uncovers hidden complexities. *Genome Res.* **34**, 1774–1784 (2024).
12. Bilgrav Saether, K. et al. Leveraging the T2T assembly to resolve rare and pathogenic inversions in reference genome gaps. *Genome Res.* **34**, 1785–1797 (2024).
13. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
14. Hiatt, S. M. et al. Long-read genome sequencing and variant reanalysis increase diagnostic yield in neurodevelopmental disorders. *Genome Res.* **34**, 1747–1762 (2024).
15. Hops, W. et al. HiFi long-read genomes for difficult-to-detect, clinically relevant variants. *Am. J. Hum. Genet.* **112**, 450–456 (2025).
16. Steyaert, W. et al. Unravelling undiagnosed rare disease cases by HiFi long-read genome sequencing. *Genome Res.* https://doi.org/10.1101/gr.279414.124 (2025).
17. Leitao, E., Schroder, C. & Depienne, C. Identification and characterization of repeat expansions in neurological disorders: methodologies, tools, and strategies. *Rev. Neurol.* **180**, 383–392 (2024).

18. Eggermann, T. et al. Imprinting disorders. *Nat. Rev. Dis. Primers* **9**, 33 (2023).
19. Greene, D. et al. Mutations in the U4 snRNA gene *RNU4-2* cause one of the most prevalent monogenic neurodevelopmental disorders. *Nat. Med.* **30**, 2165–2169 (2024).
20. Chen, Y. et al. De novo variants in the *RNU4-2* snRNA cause a frequent neurodevelopmental syndrome. *Nature* **632**, 832–840 (2024).
21. Ganesh, V. S. et al. Neurodevelopmental disorder caused by deletion of *CHASERR*, a lncRNA gene. *N. Engl. J. Med.* **391**, 1511–1518 (2024).
22. Magnusson, M. et al. Loqusdb: added value of an observations database of local genomic variation. *BMC Bioinformatics* **21**, 273 (2020).
23. Gonzalez-Del Pozo, M. et al. A comprehensive WGS-based pipeline for the identification of new candidate genes in inherited retinal dystrophies. *npj Genom. Med.* **7**, 17 (2022).
24. Kohler, S. et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **49**, D1207–D1217 (2021).
25. Martin, A. R. et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **51**, 1560–1565 (2019).
26. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
27. Acuna-Hidalgo, R., Veltman, J. A. & Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* **17**, 241 (2016).
28. Tesi, B. et al. Precision medicine in rare diseases: what is next? *J. Intern. Med.* **294**, 397–412 (2023).
29. Chen, X. et al. Spinal muscular atrophy diagnosis and carrier screening from genome sequencing data. *Genet. Med.* **22**, 945–953 (2020).
30. Dolzhenko, E. et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* **35**, 4754–4756 (2019).
31. Maggi, J. et al. Limited added diagnostic value of whole genome sequencing in genetic testing of inherited retinal diseases in a Swiss patient cohort. *Int. J. Mol. Sci.* **25**, 6540 (2024).
32. Yang, M. et al. Diagnostic utility of whole genome sequencing after negative karyotyping/chromosomal microarray in infants born with multiple congenital anomalies. *J. Korean Med. Sci.* **39**, e250 (2024).
33. Lionel, A. C. et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet. Med.* **20**, 435–443 (2018).
34. Katsanis, N. et al. Triallelic inheritance in Bardet–Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256–2259 (2001).
35. Smail, C. et al. Complex trait associations in rare diseases and impacts on Mendelian variant interpretation. *Nat. Commun.* **15**, 8196 (2024).
36. Lemmers, R. J. et al. Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 (2012).
37. Grigelioniene, G. et al. Gain-of-function mutation of microRNA-140 in human skeletal dysplasia. *Nat. Med.* **25**, 583–590 (2019).
38. Spielmann, M. & Mundlos, S. Looking beyond the genes: the role of non-coding variants in human disease. *Hum. Mol. Genet.* **25**, R157–R165 (2016).
39. Weischenfeldt, J. & Ibrahim, D. M. When 3D genome changes cause disease: the impact of structural variations in congenital disease and cancer. *Curr. Opin. Genet. Dev.* **80**, 102048 (2023).
40. Yepez, V. A. et al. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med.* **14**, 38 (2022).
41. Carrasco-Zanini, J. et al. Proteomic signatures improve risk prediction for common and rare diseases. *Nat. Med.* **30**, 2489–2498 (2024).
42. Chantada-Vazquez, M. D. P., Bravo, S. B., Barbosa-Gouveia, S., Alvarez, J. V. & Couce, M. L. Proteomics in inherited metabolic disorders. *Int. J. Mol. Sci.* **23**, 6540 (2022).
43. Lunke, S. et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat. Med.* **29**, 1681–1691 (2023).
44. Ewing, A. D. et al. Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol. Cell* **80**, 915–928 (2020).
45. Robinson, J. T. et al. Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
46. Chaisson, M. J. P., Sulovari, A., Valdmanis, P. N., Miller, D. E. & Eichler, E. E. Advances in the discovery and analyses of human tandem repeats. *Emerg. Top. Life Sci.* **7**, 361–381 (2023).
47. Porubsky, D. et al. A familial, telomere-to-telomere reference for human de novo mutation and recombination from a four-generation pedigree. Preprint at *bioRxiv* https://doi.org/10.1101/2024.08.05.606142 (2024).
48. Chen, X. et al. Comprehensive *SMN1* and *SMN2* profiling for spinal muscular atrophy analysis using long-read PacBio HiFi sequencing. *Am. J. Hum. Genet.* **110**, 240–250 (2023).
49. Funderburk, S. J., Spence, M. A. & Sparkes, R. S. Mental retardation associated with 'balanced' chromosome rearrangements. *Am. J. Hum. Genet.* **29**, 136–141 (1977).
50. Jacobs, P. A. Correlation between euploid structural chromosome rearrangements and mental subnormality in humans. *Nature* **249**, 164–165 (1974).
51. Ten Berk de Boer, E. et al. Long-read sequencing and optical mapping generates near T2T assemblies that resolves a centromeric translocation. *Sci. Rep.* **14**, 9000 (2024).
52. Abel, H. J. et al. Mapping and characterization of structural variation in 17,795 human genomes. *Nature* **583**, 83–89 (2020).
53. Gustafson, J. A. et al. High-coverage nanopore sequencing of samples from the 1000 Genomes Project to build a comprehensive catalog of human genetic variation. *Genome Res.* **34**, 2061–2073 (2024).
54. Liao, W. W. et al. A draft human pangenome reference. *Nature* **617**, 312–324 (2023).
55. Kronenberg, Z. et al. The Platinum Pedigree: a long-read benchmark for genetic variants. Preprint at *bioRxiv* https://doi.org/10.1101/2024.10.02.616333 (2024).
56. Mitsuhashi, S., Frith, M. C. & Matsumoto, N. Genome-wide survey of tandem repeats by nanopore sequencing shows that disease-associated repeats are more polymorphic in the general population. *BMC Med. Genomics* **14**, 17 (2021).
57. Baekgaard, C. H., Lester, E. B., Moller-Larsen, S., Lauridsen, M. F. & Larsen, M. J. NanoImprint: a DNA methylation tool for clinical interpretation and diagnosis of common imprinting disorders using nanopore long-read sequencing. *Ann. Hum. Genet.* **88**, 392–398 (2024).
58. Eisfeldt, J. et al. Resolving complex duplication variants in autism spectrum disorder using long-read genome sequencing. *Genome Res.* **34**, 1763–1773 (2024).
59. Levy, M. A. et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. *HGG Adv.* **3**, 100075 (2022).
60. Geysens, M. et al. Clinical evaluation of long-read sequencing-based episignature detection in developmental disorders. *Genome Med.* **17**, 1 (2025).

61. De Clercq, G. et al. Full characterization of unresolved structural variation through long-read sequencing and optical genome mapping. *Sci. Rep.* **14**, 29142 (2024).
62. Koren, S. et al. Gapless assembly of complete human and plant chromosomes using only nanopore sequencing. *Genome Res.* **34**, 1919–1930 (2024).
63. Wenger, A. M. et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162 (2019).
64. Kruglyak, S. et al. Characterizing and addressing error modes to improve sequencing accuracy. Preprint at *bioRxiv* https://doi.org/10.1101/2024.02.01.578321 (2024).
65. Tesi, N. et al. Characterizing tandem repeat complexities across long-read sequencing platforms with TREAT and otter. *Genome Res.* **34**, 1942–1953 (2024).
66. Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B. & Nestor, C. E. No evidence for DNA N^6-methyladenine in mammals. *Sci. Adv.* **6**, eaay3335 (2020).
67. Lenner, F. et al. Nallo: a comprehensive Nextflow pipeline for long-read rare disease analysis. *GitHub* https://github.com/genomic-medicine-sweden/nallo (2025).
68. Hon, T. et al. Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci. Data* **7**, 399 (2020).
69. Ni, P. et al. DNA 5-methylcytosine detection and methylation phasing using PacBio circular consensus sequencing. *Nat. Commun.* **14**, 4054 (2023).
70. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
71. Pages-Gallego, M. & de Ridder, J. Comprehensive benchmark and architectural analysis of deep learning models for nanopore sequencing basecalling. *Genome Biol.* **24**, 71 (2023).
72. Stanojević, D., Lin, D., Florez de Sessions, P. & Šikić, M. Telomere-to-telomere phased genome assembly using error-corrected Simplex nanopore reads. Preprint at *bioRxiv* https://doi.org/10.1101/2024.05.18.594796 (2024).
73. Jorgensen, M. W., Miceikaite, I. & Larsen, M. J. nanoNIPT: short-fragment nanopore sequencing of cell-free DNA for non-invasive prenatal testing of fetal aneuploidies and sex chromosome aberrations. *Prenat. Diagn.* **43**, 314–317 (2023).
74. Jain, M. et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338–345 (2018).
75. Dorey, A. & Howorka, S. Nanopore DNA sequencing technologies and their applications towards single-molecule proteomics. *Nat. Chem.* **16**, 314–334 (2024).
76. Loukanov, A., Filipov, C., Mladenova, P., Toshev, S. & Emin, S. Electron microscopic visualization of complementary labeled DNA with platinum-containing guanine derivative. *Microsc. Res. Tech.* **79**, 280–284 (2016).
77. Mittal, S. & Pathak, B. Towards a graphene semi/hybrid-nanogap: a new architecture for ultrafast DNA sequencing. *Nanoscale* **15**, 757–767 (2023).
78. Liu, S. et al. Targeted transcriptome analysis using synthetic long read sequencing uncovers isoform reprograming in the progression of colon cancer. *Commun. Biol.* **4**, 506 (2021).
79. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
80. Tjio, J. H. & Levan, A. The chromosome number of man. *Hereditas* **42**, U1–U6 (1956).
81. Lejeune, J., Turpin, R. & Gautier, M. [Chromosomic diagnosis of mongolism]. *Arch. Fr. Pediatr.* **16**, 962–963 (1959).
82. Caspersson, T., Zech, L., Johansson, C. & Modest, E. J. Identification of human chromosomes by DNA-binding fluorescent agents. *Chromosoma* **30**, 215–227 (1970).
83. Smith, L. M., Fung, S., Hunkapiller, M. W., Hunkapiller, T. J. & Hood, L. E. The synthesis of oligonucleotides containing an aliphatic amino group at the 5′ terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res.* **13**, 2399–2412 (1985).
84. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
85. Pinkel, D. et al. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc. Natl Acad. Sci. USA* **85**, 9138–9142 (1988).
86. Riordan, J. R. et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066–1073 (1989).
87. Solinas-Toldo, S. et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**, 399–407 (1997).

Acknowledgements

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 (Sällsyntafonden), the Swedish Brain Foundation (FO2024-0128-HK-44) and the Swedish Cancer Society (24 3504 Pj).

Author contributions

Conceptualization: J.E. and A.L.; writing of manuscript: J.E., M.E., M.N. and A.L.; figures: J.E., A.L. and M.E.

Competing interests

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

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02160-y.

Correspondence should be addressed to Anna Lindstrand.

Peer review information *Nature Genetics* thanks Alexander Hoischen and Bekim Sadikovic for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2025