Long read sequencing on its way to the routine diagnostics of genetic diseases

Giulia Olivucci[1,2], Emanuela Iovino[1]*, Giovanni Innella[3,4], Daniela Turchetti[3,4], Tommaso Pippucci[1] and Pamela Magini[4]

[1]IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy, [2]Department of Surgical and Oncological Sciences, University of Palermo, Palermo, Italy, [3]Department of Medical and Surgical Sciences (DIMEC), University of Bologna, Bologna, Italy, [4]Medical Genetics Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

The clinical application of technological progress in the identification of DNA alterations has always led to improvements of diagnostic yields in genetic medicine. At chromosome side, from cytogenetic techniques evaluating number and gross structural defects to genomic microarrays detecting cryptic copy number variants, and at molecular level, from Sanger method studying the nucleotide sequence of single genes to the high-throughput next-generation sequencing (NGS) technologies, resolution and sensitivity progressively increased expanding considerably the range of detectable DNA anomalies and alongside of Mendelian disorders with known genetic causes. However, particular genomic regions (i.e., repetitive and GC-rich sequences) are inefficiently analyzed by standard genetic tests, still relying on laborious, time-consuming and low-sensitive approaches (i.e., southern-blot for repeat expansion or long-PCR for genes with highly homologous pseudogenes), accounting for at least part of the patients with undiagnosed genetic disorders. Third generation sequencing, generating long reads with improved mappability, is more suitable for the detection of structural alterations and defects in hardly accessible genomic regions. Although recently implemented and not yet clinically available, long read sequencing (LRS) technologies have already shown their potential in genetic medicine research that might greatly impact on diagnostic yield and reporting times, through their translation to clinical settings. The main investigated LRS application concerns the identification of structural variants and repeat expansions, probably because techniques for their detection have not evolved as rapidly as those dedicated to single nucleotide variants (SNV) identification: gold standard analyses are karyotyping and microarrays for balanced and unbalanced chromosome rearrangements, respectively, and southern blot and repeat-primed PCR for the amplification and sizing of expanded alleles, impaired by limited resolution and sensitivity that have not been significantly improved by the advent of NGS. Nevertheless, more recently, with the increased accuracy provided by the latest product releases, LRS has been tested also for SNV detection, especially in genes with highly homologous pseudogenes and for haplotype reconstruction to assess the parental origin of alleles with *de novo*

pathogenic variants. We provide a review of relevant recent scientific papers exploring LRS potential in the diagnosis of genetic diseases and its potential future applications in routine genetic testing.

1 Introduction

Molecular diagnosis has an important impact on clinical management of patients with genetic diseases and their families. It confers the certification of a chronic and often disabling condition with dedicated healthcare pathways, specific treatment options and, still in few cases, personalized therapies. The causative DNA alteration can be searched in future pregnancies, allowing informed reproductive choices, and, especially in the case of hereditary tumor predisposition syndromes, in healthy family members, who can benefit from the most appropriate clinical management based on the resulting risk. Reporting times are obviously important to ensure rapid communication and timely interventions, both in the postnatal and even more in the prenatal settings.

Genetic testing started in 1960 with karyotyping and later evolved rapidly with the introduction of molecular analysis using DNA as probe (i.e., fluorescence *in situ* hybridization) or target (i.e., Sanger sequencing). This first advancement from the study of all chromosomes to specific genomic regions brought a gain in resolution at the expense of the investigation scale, making the diagnosis dependent from target selection guided by clinical hypotheses. The next technological advancement was aimed at achieving high resolution in the context of genome-wide analysis, allowing clinically unbiased testing. The first techniques with these features were microarrays, that brought a huge revolution in the cytogenetic field, replacing karyotype analysis as first-tier test for neurodevelopmental disorders (NDDs) and congenital malformations with an increase in diagnostic yield from 3% to 15%–20% (Miller et al., 2010). The advent of second or next-generation sequencing technologies based on short reads marked the greatest progress in genetic medicine by the analysis of the entire exome or genome at nucleotide-level resolution, finding pathogenic variants in nearly 30%–38% of NDD cases or even higher percentages of patients with specific genetic conditions (Nurchis et al., 2023; van der Sanden et al., 2023). This technological evolution impacted not only on detection rates and diagnostic yields but also on laboratory routine. Faster and simpler wet lab protocols increased the number of patients that could be analyzed simultaneously and reduced turnaround times. However, all these techniques are far from being completely efficient, even if combined and applied sequentially, missing the molecular diagnosis in about half of the patients with genetic conditions (Haghighi et al., 2018; Boycott et al., 2019). Indeed, apart from interpretative and clinical issues, including the partial knowledge of the functions of genes and genomic regions or the incomplete phenotypic characterization of patients preventing accurate genotype-phenotype correlations, technical limitations contribute substantially to missing diagnoses (Mastrorosa et al., 2023). In particular, repetitive, highly homologous and GC-rich sequences are difficult to analyze with current genetic tests, and strategies specifically developed to reach these hard genomic regions are still laborious, time-consuming and low-sensitive (i.e., southern-blot for repeat expansion or long-PCR for genes with highly homologous pseudogenes) (Mandelker et al., 2016). In addition, the detection of structural variants (SVs) through short read sequencing (SRS) is still limited and requires advanced bioinformatic tools that have not yet been incorporated into the routine pipelines (Neerman et al., 2019).

The ability of third generation sequencing or long read sequencing (LRS) technologies to sequence long molecules of nucleic acids with improved mappability overcomes the technical limitations of SRS, potentially impacting on detection rate and diagnostic yield (Mastrorosa et al., 2023).

SRS technologies, particularly those based on sequencing-by-synthesis, have shown exceptional accuracy, reaching 99.9% base-call accuracy (Lin et al., 2021). However, reads are usually a few hundred base pairs long, limiting the ability to map reads, especially in low-complexity repetitive loci and duplicated regions, and to decipher structural variants or tandem repeats. LRS, on the other hand, enables sequencing of DNA fragments longer than 10,000 bp. At present, the two major LRS technologies, namely Pacific Biosciences and Oxford Nanopore Technology, are based on different approaches. The first is based on single molecule real-time sequencing and allows real-time detection of nucleotide incorporation events using a DNA polymerase, with high accuracy (up to 99.8% with circular consensus sequencing) (Wenger et al., 2019). One limitation of this technology is, however, the high cost of equipment and supplies. Nanopore-based sequencing detects changes in the ionic current during translocation of a single-stranded molecule of DNA/RNA through a biosensor (nanopore) to generate even longer reads (10–100 kb with standard long-reads, >100 kb with ultra-long reads, up to a described maximum length of about 2 Mb) (Payne et al., 2019), enabling access to long repetitive regions. This technology has been characterized by lower accuracy and particular high error rate in single nucleotide variant (SNV) and indel calling, which, however, is improving due to advances in chemistry and algorithms (Ni et al., 2023). The intent of this review is not to make a comparison between the two existing LRS technologies. For this reason, we do not make any further distinction between these different technological approaches, referring to them collectively as LRS. Descriptions of both technologies, their comparison and data analysis tools are reported in several scientific publications. We provide some of them to address readers for further information (Amarasinghe et al., 2020; Logsdon et al., 2020; Amarasinghe et al., 2021; Athanasopoulou et al., 2022; Mastrorosa et al., 2023; van Dijk et al., 2023).

TABLE 1 Comparison of current diagnostic techniques for the detection of structural variants. CNV, copy number variant; MLPA, multiplex ligation-dependent probe amplification; SRS, short read sequencing; SV, structural variant.

	Karyotyping	Microarrays	MLPA	SRS technologies
Chromosomal alterations detected	Polyploidies and aneuploidies Balanced SVs CNVs	Aneuploidies CNVs	CNVs	Balanced SVs CNVs
Resolution	>5–10 Mb	Variable according to array design, generally ≥10 kb	Exon level	All range of sizes from 50 bp
Analysis scale	Whole genome	Whole genome	Target regions	Whole genome
Limitations	Low resolution Time-consuming and labor-intensive	Inability to detect balanced SVs Low sensitivity for mosaic CNVs Inability to map the insertion position of duplicated regions	Targeted analysis Low sensitivity for mosaic CNVs	Low accuracy and reproducibility Inability to identify SVs in repeated regions or complex genomic regions

Human reference genome releases commonly used to align sequences (GRCh37 and GRCh38) are highly fragmented and contain unresolved portions. The advent of LRS technologies allowed the release of whole genome haplotype resolved assemblies (Nurk et al., 2020; Cheng et al., 2021), including pericentromeric and subtelomeric regions and the short arms of acrocentric chromosomes (Nurk et al., 2022), and the generation of benchmarks for variants in hard-to-assess disease-genes (Wagner et al., 2022), improving sequencing and pathogenic alteration identification.

Advantages of LRS clinical application might concern not only the enhanced detection of missing variants but also reduced turnaround times, from blood sampling to analysis results and report, being based on rapid wet lab protocols for sample processing. Moreover, LRS potentially allows the evaluation of different alteration types in the same analysis (i.e., methylation, SNVs and SVs), avoiding the sequential application of inconclusive tests and therefore bringing a cost advantage.

We provide here a review of the recent scientific literature exploring LRS potential applied to the identification of disease-related genetic and epigenetic alterations, focusing on diagnostic and clinical implications and highlighting the advantages over current genetic testing.

2 Structural variants

SVs are a large class of genomic variations that differ greatly in type and size. They usually range from 50 bp to well over megabases and include both balanced and unbalanced alterations. Balanced SVs are changes in DNA structure, such as inversions of a genetic fragment or translocations of genomic segments within or between chromosomes. Unbalanced SVs, also referred to as copy number variations (CNVs), include gains and deletions of genetic material. Due to their remarkable size, SVs are considered the largest source of genetic variation in terms of the number of nucleotides involved and they are implicated in several Mendelian disorders through different mechanisms, disrupting either regulatory or coding sequences or altering the copy number of dosage-sensitive genes (Carvalho and Lupski, 2016; Middelkamp et al., 2019; Ho et al., 2020). However, compared to other classes of variants such as SNVs, they are still largely understudied due to the technical

limitations of identification methods (Amarasinghe et al., 2020; Bolognini and Magi, 2021).

2.1 Current diagnostic analyses for SV detection

Various techniques were developed over the last 60 years to allow a genome-wide identification of different types of SVs in clinical diagnosis. These range from cytogenetic analyses (i.e., karyotyping), oligonucleotide-based technologies (i.e., Array Comparative Genomic Hybridization [aCGH] and Single Nucleotide Polymorphism [SNP] array) to short-reads whole genome sequencing (WGS) (Liu et al., 2022a). However, each of these techniques has several weaknesses, which are summarized in Table 1.

Karyotyping provides a complete overview of the human genome and enables the detection of large, clinically significant chromosomal rearrangements. It can detect balanced and unbalanced SVs with a limited resolution of 5–10 Mb, preventing the identification of smaller SVs. Array-based approaches or microarrays are extensively used in clinical genetics: with a diagnostic yield of 15%–20%, they represented the first-tier genetic test for NDD (Miller et al., 2010) for many years, until the clinical application of SRS pushed molecular diagnosis to at least 31% of NDD patients (Srivastava et al., 2019). Although they provide higher resolution than karyotyping, diagnostic microarrays generally detect CNVs larger than 10–50 kb, with low resolution for mosaic alterations, and miss all balanced or smaller SVs. The identification of SVs, together with SNVs, from SRS data would be the ideal solution to save time and money. Continuous advances in short-read technologies based on paired-end sequencing have led to an unprecedented development of algorithms and tools that can potentially detect all types of SVs with breakpoint resolution. However, despite the more than 80 bioinformatic tools that have been developed for the detection of SVs in short-read WGS data, several benchmark studies have shown that none of these tools has high sensitivity and accuracy for all SV types and sizes (Cameron et al., 2019; Kosugi et al., 2019). The identification of SVs spanning several hundred to thousands of base pairs with short reads (150–250 bp) represents the greatest SRS challenge. Reads are typically smaller than SVs, making the signal reconstruction

difficult to model with available methods. In addition, studies also showed a high number of false positives and the inability to detect SVs in low complexity regions consisting of tandem or simple repeat sequences or in genomic loci characterized by segmental duplications, which are hotspots for SVs (Alkan et al., 2011; Mills et al., 2011; Zook et al., 2020). These problems strongly limit the application of SRS to the identification of SVs in diagnostics, where microarrays remain the method of choice for the detection of CNVs and karyotyping is still the only reliable technique for balanced SVs, despite its low resolution.

2.2 Clinical application of LRS for SV identification

LRS has been proposed as the sequencing alternative to overcome the limitations of existing technologies and increase the accuracy of SV detection (Branton et al., 2008; Eid et al., 2009). The reads generated by LRS can completely span SVs, generate highly reliable alignments at breakpoints, resolve complex structural variants (cxSVs) involving three or more breakpoint junctions and directly reconstruct haplotypes (Sedlazeck et al., 2018; Hiatt et al., 2021). Since the read length is usually longer than the most common repeated sequences, LRS technology significantly increases the sensitivity in detecting SVs and, according to a study by Ebert et al. can identify more than 68% of SVs that are undetectable with SRS (Sedlazeck et al., 2018; Ebert et al., 2021; Olson et al., 2022). Over the past few years, several studies have shown the potential of LRS in improving the diagnostic yield in Mendelian diseases, especially in cases that remain unsolved after SRS and may be carriers of SVs or complex events that are difficult to detect with current technologies because breakpoints are located in repetitive regions. In one of the first clinical applications of LRS Merker et al. identified a causative deletion of 2.2 kb in the PRKAR1 gene in a patient with Carney complex (Merker et al., 2018). In the same year, Miao et al., 2018 used LRS to detect a deletion of 7.1 kb in the G6PC gene missed by SRS and affecting the non-mutated allele, confirming the clinically suspected diagnosis of recessive glycogen storage disease type Ia syndrome. Another example of LRS utility in undiagnosed recessive disorders was provided by Sano et al., 2022, who studied a cohort of 15 patients with retinitis pigmentosa and carriers of pathogenic variants in one allele of the EYS gene: whole-genome LRS was able to detect all the causative SNVs previously identified and two large deletions (376 kb and 395 kb) encompassing the coding sequence of EYS in two patients. In a patient with Usher syndrome and a maternally inherited PCDH15 exonic deletion identified by CNV analysis on SRS data, LRS allowed a better characterization of the intragenic loss and the identification of a 4.6 Mb inversion on the paternal allele, generating two aberrant fusion transcripts, PCDH15-LINC00844 and BICC1-PCDH15 (Vaché et al., 2020).

Additional studies further confirmed the utility of LRS as an effective technology for the identification of clinically relevant SVs missed by standard genetic tests (Mizuguchi et al., 2019; Hiatt et al., 2021; Miller et al., 2021; Lecoquierre et al., 2023; Ohori et al., 2023). Damián et al., 2023 effectively characterized two cases of congenital aniridia after a decades-long diagnostic odyssey. In one of these cases, they found a cryptic 4.9 Mb de novo inversion disrupting

intron 7 of PAX6 while in the other they were able to reclassify a balanced t (6; 11) translocation cytogenetically identified years earlier and erroneously considered as non-pathogenic, by precisely mapping its breakpoint within a PAX6 enhancer.

The precise definition of SV breakpoints is one of the most important clinical applications of LRS. Examples include the accurate assessment of the genomic position with nucleotide-level resolution of a deletion in the BBS9 gene in a patient with Bardet-Biedl syndrome (Reiner et al., 2018) and of a balanced reciprocal t (X; 20) translocation, disrupting the X-linked ARHGEF9 gene, associated with developmental and epileptic encephalopathy, that was previously detected but not accurately mapped by karyotype analysis in a girl with intellectual disability and seizures (Dutta et al., 2019). In a family with three non-contiguous DMD duplications identified through standard testing, the precise breakpoint definition by LRS, showing the presence of an intact copy of the gene, contributed to the reclassification of the rearrangement as benign, with important implications for genetic counseling, especially in prenatal setting (He et al., 2022). Mariya et al., 2023 used a targeted LRS approach to sequence the breakpoint junctions of a PLP1 duplication and a MECP2 duplication/triplication causing Pelizaeus-Merzbacher disease and MECP2 duplication syndrome, respectively. In the latter case, the authors were able to map one of the MECP2 copies 45 Mb apart from the original position, avoiding a possible misinterpretation of the pathogenic allele by short tandem repeat haplotyping in pre-implantation genetic testing, in case of recombination events occurring within this 45 Mb interval. The inability of mapping the extra copy or copies of genomic gains is one of the main limitations of current diagnostic analyses widely applied to genetic conditions (i.e., microarrays and Multiplex Ligation-dependent Probe Amplification [MLPA]), impairing the correct clinical classification of these CNVs.

The impact of this clinical application of LRS has been studied also for cancer susceptibility genes (CSGs), in which CNVs are usually investigated by MLPA after inconclusive SRS results, but other types of SVs are not routinely evaluated. Thibodeau et al. evaluated the utility of LRS in resolving candidate germline SVs in CSGs detected through short-read genome sequencing in a cohort of 669 advanced cancer patients. They were able to confirm eight simple pathogenic or likely pathogenic SVs, resolve three additional variants whose impact could not be fully elucidated through SRS, and classify a recurrent alteration on chromosome 16p13 as a sequencing artifact and one complex rearrangement on chromosome 5q35 as likely benign, obviating the need for further clinical assessment (Thibodeau et al., 2020). These results demonstrated that LRS can improve the validation, resolution, and classification of germline SVs in CSGs. A similar approach was carried out by Dixon et al., 2023, who reported that LRS accurately detected all the CNVs (from single exons to whole genes, including a tandem duplication) in BRCA1, BRCA2, CHEK2 and PALB2 previously identified in nineteen patients, and defined the precise breakpoints of SVs in BRCA1 and CHEK2, revealing unforeseen allelic heterogeneity and demonstrating the potential for LRS as a powerful genetic testing assay in the hereditary cancer setting. Further confirmations of its diagnostic value are extended to very rare cancer predisposition syndromes. In two siblings both neonatally diagnosed with atypical teratoid rhabdoid tumor, Sabatella et al. applied LRS in combination with optical genome mapping and identified an insertion of a −2.8 kb retrotransposon element within the intron 2 of

SMARCB1, disrupting its correct splicing and present in mosaic state in the mother (Sabatella et al., 2021).

Also targeted approaches revealed their effectiveness and clinical utility in detecting SVs affecting CSGs. Filser et al. were able to classify a germline duplication event of exons 18–20 of *BRCA1* as pathogenic using adaptive sampling (Filser et al., 2023), while Yamaguchi et al. applied a hybridization-based enrichment to the LRS analysis of mismatch-repair genes responsible for Lynch syndrome, precisely mapping SV breakpoints and proposing their method as a possible substitute for MLPA assays (Yamaguchi et al., 2021).

An additional promising application of LRS is the resolution of cxSV configuration. The study by Sanchis-Juan et al., 2018 provides an example of how LRS can be used to elucidate the structure of a duplication-inversion-duplication in the *CDKL5* gene with breakpoints mapping in highly repeated regions inaccessible by SRS.

Rare cxSVs in the *DMD* gene contribute to the pathogenesis of dystrophinopathies but escape diagnosis by routine diagnostic testing (MLPA and targeted SRS). Gross rearrangements mediated by transposable elements could be missed also by targeted LRS if read length is insufficient to cover the SVs and their breakpoints. Conversely, whole-genome LRS was shown to correctly identify and characterize large-scale cxSVs, contributing to the diagnosis of unsolved cases when implemented as the fourth analysis in a stepwise genetic testing strategy for patients with dystrophinopathies (Xie et al., 2020; Xie et al., 2021). The alpha- and beta-globin genes map in an additional disease-associated locus involved in rare and complex rearrangements that are still underdiagnosed but can be detected by LRS, together with rare SNVs, increasing the diagnostic yield by approximately 7% compared to routine detection methods (variant-tailored Gap-PCR and Sanger sequencing) (Long et al., 2022; Xu et al., 2023; Zhuang et al., 2023).

Thanks to LRS, combined to additional genomic analyses in a multidimensional approach, Scharf et al., 2022 diagnosed a patient with an attenuated familial adenomatous polyposis who remained unsolved by Sanger sequencing and SRS multigene panel for years. They found a complex 3.9 Mb rearrangement involving 14 fragments from chromosome 5q22.1q22.3, which likely originated from chromothripsis events and separated the *APC* promoter 1B from the coding open reading frame, leading to allele-specific downregulation of *APC* mRNA.

All these studies revealed that, although underestimated because still unexplored in diagnostics, cryptic SVs could represent relevant pathogenic events in Mendelian disorders and efficient methods for their detection, such as LRS, should be rapidly translated from the research to the clinical setting. Routine laboratory activities could also benefit from more rapid and less laborious wet lab protocols, remarkably reducing sample processing times compared to conventional techniques. Additionally, the possibility of real-time analysis might significantly decrease the time to diagnosis, with a reported less-than-1-h timeframe for aneuploidies (Magini et al., 2022).

3 Tandem repeat-related diseases

Tandem repeats (TRs) are stretches of repeated nucleotide blocks, variable in size (from few to thousands of bases) and highly homologous, thus prone to rearrangements with length-dependent instability. Disorders mainly affecting the central nervous system manifest when the number of repeats exceeds a gene-specific threshold. At least 69 TR-related diseases are known and most of the expanded TRs map within both coding and non-coding regions, typically untranslated regions (UTR) and intronic sequences. Expansions affecting coding TRs generally generate proteins with long polyQ or polyA chains that accumulate within cells or lose their functions. Non-coding expansions have different pathogenic mechanisms, according to their location and length, including gene silencing through methylation for those mapping at the 5′-UTRs (Depienne and Mandel, 2021; Gall-Duncan et al., 2022). Technological limitations negatively affected both their discovery, which accelerated only in recent years, and their diagnosis.

3.1 Current diagnostic analyses for expanded TR detection

Three features of these repeated regions make their analysis particularly problematic, especially through PCR-based methods: size, homology, and high GC density.

Hybridization and correct mapping of PCR primers to target regions are hampered by homologous and GC-rich sequences and polymerase accuracy decreases with long amplicons. For these reasons, specific amplification strategies have been designed (i.e., repeat-primed PCR [RP-PCR]), combined with PCR-free hybridization approaches, such as Southern blot which is still the gold standard for sizing larger expansion. However, these techniques are slow, labor-intensive, not suited to the parallel analysis of multiple loci in a single assay and cannot accurately determine the number of repeats or the structure of the repeated region within very long alleles, which have usually clinical implications, affecting age of onset and disease severity and progression (Cumming et al., 2018; Chintalaphani et al., 2021). Moreover, TR instability may result in somatic mosaicism further challenging consensus size determination and structure definition (Figure 1). The diagnostic yield of these tests ranges from 10% to 30% and their efficacy is impaired for longer (>300 repeats) and complex TRs (Read et al., 2023).

The advent of SRS and the implementation of advanced bioinformatic tools improved the detection and the discovery of repeat expansions (Bahlo et al., 2018), but, although their clinical utility has been demonstrated (Ibañez et al., 2022), the low coverage due to library preparation issues and misalignment of short-reads related to the GC-rich and repetitive nature of these regions, prevented their use as routine diagnostic test (Chintalaphani et al., 2021).

3.2 Clinical application of LRS to TR identification

LRS contributed to the discovery of novel disease-associated TRs (Gall-Duncan et al., 2022) and appears as a promising diagnostic method because its whole-genome pre-sequencing protocol avoids PCR-based amplification steps and long reads, sizing over 10 kb, can

RP-PCR	Southern Blot	SRS*
• Reduced sensitivity for larger expansion	• Time-consuming, labor-intensive and expensive	• Low coverage due to library preparation issues
• Low sensitivity for mosaicism	• Low sensitivity for mosaicism	• Misalignment of short-reads in GC-rich and repetitive regions
• Possible allele dropout events	• High amount of input DNA required	• No information about methylation status
• Limited multiplexing possibility	• Limited multiplexing possibility	
• No information about methylation status		
• Low accuracy for length determination		

FIGURE 1
Limitations of current diagnostic tests for tandem-repeat-related diseases. SRS: short read sequencing; RT-PCR: repeat primed-polymerase chain reaction; *its clinical application has been evaluated but technical issues prevented its transition to the diagnostic routine.

entirely cover most known pathogenic expanded repeats reaching unique flanking regions, giving specificity to sequence alignment (Mitsuhashi and Matsumoto, 2020). Furthermore, the methylation status of loci can be simultaneously determined, providing an additional relevant disease marker for some conditions.

Although whole-genome LRS is feasible and has been helpful in the discovery of new disease-causing expansions (Corbett et al., 2019; Chintalaphani et al., 2021), target approaches will probably represent more cost-effective strategies to be applied for diagnostic purposes.

Liang et al., 2022 developed and evaluated a long-range PCR and LRS-based assay to analyze the FMR1 locus on clinical samples. Compared to conventional techniques, target LRS showed higher sensitivity for mosaic expansions, accurate determination of the CGG repeat number of any length, and of AGG interruptions, even in mosaic alleles. This information is extremely important to assess the risk of recurrence and guide reproductive choices. Additionally, the novel approach was able to simultaneously identify intragenic SNVs and repeat-flanking microdeletions potentially disrupting primer annealing and leading to false or negative results through RP-PCR. The turnaround time of the target LRS assay was 8 days and the cost per sample ranged from 40 to 80 $, according to the number of pooled samples. The possibility to combine and sequence together different target assays will further improve the cost-benefit ratio.

Although generally getting a small amount of target DNA, amplification-free enrichment methods, amenable to multiplexing, have been designed by using clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9, avoiding PCR biases, such as preferential amplification of shorter alleles, and issues with repetitive and GC-rich regions.

The combination of CRISPR-Cas9 enrichment and LRS allowed to exclude the presence of interruptions in very long expanded TCF4 alleles as a penetrance reduction mechanism for Fuchs endothelial corneal dystrophy, at least in some individuals. In addition, it detected novel sequence variants flanking the expansions with a potential role in sizing borderline repeats and confirmed that the heterogeneity in repeat length, previously observed in leukocytes through conventional methods, has a likely biological origin and is

not due to PCR artifacts (Wieben et al., 2019). Unamplified DNA enriched for the DMPK gene through the CRISPR-Cas9 approach was sequenced by LRS in four patients with myotonic dystrophy type 1 (DM1), revealing more interruptions than those detected by RP-PCR and providing sensitive expanded allele-specific characterization of methylation levels, proposed as more reliable markers of the severe congenital form compared to repeat length. Assessing the length of individual alleles, LRS might also improve the estimation of the progenitor allele length that has been suggested to have a stronger correlation with disease severity and age of onset compared to the modal repeat length of the heterogeneous population of alleles generated by somatic instability (Rasmussen et al., 2022). Compared to amplicon-based enrichment, successfully used by Mangin et al., 2021, the CRISPR-Cas9 approach improved the measurements of repeat length and somatic mosaicism in DM1 patients and the detection of interruptions even within very long expanded alleles. These three clinically relevant parameters (repeat length, somatic mosaicism and interruptions) could be simultaneously analyzed in less than 1 month (Tsai et al., 2022). In addition to the identification of the number of very large repeated units (3.3 kb/unit) in patients with facioscapulohumeral muscular dystrophy, LRS of CRISPR-Cas9-enriched DUX4 gene detected an atypical pathogenic deletion and provided a methylation profile of the entire D4Z4 locus at nucleotide-level resolution, revealing hypomethylation of the causative contracted allele (Hiramuki et al., 2022). An increased diagnostic yield was obtained by the application of the CRISPR-Cas9 enrichment-LRS approach to a cohort of patients with benign adult familial myoclonic epilepsy, resolving two cases with ambiguous results from RP-PCR/Southern blot on SAMD12 repeats. Furthermore, whole-genome LRS revealed infrequent, disease-associated repeat expansions in SAMD12-negative patients (Mizuguchi et al., 2021). The development of an alignment-independent analysis pipeline simplified the definition of HTT repeat sequence in Huntington patients and provided a useful method for the investigation of TRs with variable or unknown length in the reference genome. Although multiplexing with a consequent cost reduction is possible, demanding laboratory protocols, the relevant Cas9-derived off-target issue and the huge input DNA required represent important limitations for the implementation of the CRISPR-Cas9-based LRS in clinical routine (Höijer et al., 2018).

The bioinformatic selection of target regions during the sequencing process through adaptive sampling tools avoids wet-lab-based enrichment procedures and experimental optimizations, reducing DNA amount requirements, costs and analysis times, and allows the simultaneous investigation of hundreds of loci, with easy customization (Chintalaphani et al., 2021).

Stevanovski et al., 2022 tested a custom bioinformatically enriched panel of 37 TR loci associated to neurological or neuromuscular diseases, showing accurate haplotype-resolved sizing, sequence determination and DNA methylation profiling of neuropathogenic short TRs. They also demonstrated the utility of LRS for genotyping very large and complex TRs, with different motif conformations, discriminating between known disease-causing and non-clinically relevant alleles. Fifty-nine disease-associated TR loci have been simultaneously tested through the adaptive sampling method by Miyatake et al. (2022a), that showed higher speed, accuracy and comprehensiveness compared to conventional

diagnostic methods, providing suggestions to improve coverage depth and reduce costs.

Additional proof-of-principle studies, small cohort analyses or case reports showed feasibility of LRS application to the characterization of TRs associated to several diseases, including amyotrophic lateral sclerosis and frontotemporal dementia (Ebbert et al., 2018), familial cortical myoclonic tremor with epilepsy (Zeng et al., 2019), neuronal intranuclear inclusion disease (Deng et al., 2019), CANVAS syndrome (Miyatake et al., 2022b), autosomal dominant tubulointerstitial kidney disease (Okada et al., 2022) and synpolydactily (Melas et al., 2022).

The clinical application of an LRS-based comprehensive analysis of expanded TRs, accurately defining repeat length and structure, interruptions, somatic levels and methylation in a single test will improve turnaround times, genotype-phenotype correlations and genetic counseling. The relevance of accurate diagnosis is even more evident in light of the emerging target therapies for some expansion-related diseases, where LRS could be also applied as an effective method to define *in vivo* editing outcomes (Simpson et al., 2023).

However, improvements are needed to make LRS the first-tier test for the diagnosis of TR-related diseases in the future. Despite recent efforts (Mousavi et al., 2019; Liu et al., 2020; Mitsuhashi et al., 2021; Liu et al., 2022b), current population catalogs of repeats are still incomplete, resulting in a hard discrimination between pathogenic and benign repeated tracts (Mousavi et al., 2019; Depienne and Mandel, 2021). Discordant results are generated by different bioinformatic tools, needing more accurate benchmarking (Chintalaphani et al., 2021). Finally, the development of advanced strategies and algorithms is required to improve read depth and base-calling accuracy as high sequencing error rate, low coverage and significant off-target issues observed with some approaches (i.e., CRISPR-Cas9) could impair clinical applications (Ebbert et al., 2018; Tsai et al., 2022).

4 Single nucleotide variants

After several decades of Sanger sequencing dominance in the diagnostic field, SRS has revolutionized the detection of SNVs allowing the parallel analysis of many samples and genes. The accuracy and the clinical application of SRS techniques have laid the groundwork for reaching genetic diagnosis in a large proportion of monogenic disorders with a single analysis. However, some pathogenic variants remain undetected, especially if located within difficult-to-access genomic regions. Introducing LRS into clinical practice can overcome these limitations and increase diagnostic yields. Although LRS accuracy in detecting SNVs is not yet comparable to that of the state-of-the-art SRS techniques, recent technological advancements could fill this gap and further foster its diagnostic application, that can help clarify undiagnosed cases and elucidate findings with uncertain clinical significance (Lin et al., 2021; Ni et al., 2023).

4.1 Pseudogenes

As previously mentioned, one of the most important limitations of SRS is the inability to read repetitive or duplicated regions of the genome. Mapping highly homologous sequences is complicated and the incorrect alignment of reads can lead to either false negative or positive results. The analysis of these loci usually requires more than one technique to obtain a reliable result, and the process is time-consuming and highly dependent on the expertise of the center where the analysis is performed. Conversely, LRS can access and accurately map these genomic regions and successfully detect variants in genes with a high degree of homology with one or more pseudogenes, providing an efficient single-test tool.

The classic example of a difficult-to-test gene is *PKD1*, which is characterized by high GC-content, high homology with six pseudogenes and a heterogeneous mutational spectrum. The SRS analysis of this gene requires a long-range PCR-based amplification of coding exons through specific primers anchored to mismatched sequences between *PKD1* and pseudogenes. To further increase mapping accuracy, reads are aligned against a modified reference genome masking nucleotides outside the *PKD1* locus. Compared to Sanger sequencing, this strategy showed 100% sensitivity, but 63% specificity and 92% accuracy, with false positive calls in GC-rich regions and homopolymers, and allele dropout issues (Mantovani et al., 2020). MLPA should be applied on patients with wild type *PKD1* sequence to identify possible pathogenic CNVs.

Borràs et al. applied LRS on pooled long-range-PCR amplicons to screen for *PKD1* mutations in 19 patients with autosomal dominant polycystic kidney disease, enabling to distinguish between *PKD1* and its pseudogenes and reducing the interference of homologous sequences, high GC-content or repetitive elements (Borràs et al., 2017). The technique allowed the identification of 20 out of 21 pathogenic variants including substitutions, single nucleotide deletions, large deletions, one deletion-insertion and three out of four insertions or duplications, providing a correct diagnosis in 18/19 patients. One pathogenic insertion was not detected by LRS despite high read depth, most likely due to noise in homopolymer sequencing. The authors suggest that LRS could be a possible alternative technique to those currently used to enable *PKD1* screening with a single diagnostic test, although further studies on larger cohorts are required before implementing this strategy into the clinic.

Several authors have applied LRS to solve sequencing difficulties related to *CYP21A2*. Biallelic mutations in *CYP21A2* are responsible for autosomal recessive congenital adrenal hyperplasia. This gene has a high degree of homology with its nonfunctional pseudogene, *CYP21A1P*, which is located in tandem with the functional gene. Together with the adjacent genes and relative pseudogenes, they form a genetic module called RCCX (*RP1-C4A-CYP21A1P-TNXA-RP2-C4B-CYP21A2-TNXB*), where the high frequency of misalignments leads to gene conversions, structural rearrangements and formation of chimeric genes. Currently, diagnostic testing mostly relies on MLPA and Sanger sequencing, both with relevant limitations preventing an accurate analysis of this complex locus and often leading to ambiguous results. Long-range locus-specific PCR combined with LRS successfully identified SNVs, chimeric genes and complex rearrangements, providing a comprehensive analysis of *CYP21A2* mutations with 100% sensitivity and specificity when compared to traditional methods (Liu Y. et al., 2022; Tantirukdham et al., 2022; Li et al., 2023). Moreover, LRS allowed the direct detection of copy number

and cis-trans configuration of *CYP21A2* and *CYP21A1P* without the need to test other family members. However, as long-range-PCR is needed, there is a residual risk of allele dropout during amplification. All things considered, LRS was proved to be a promising, cost and time effective technique for carrier screening and diagnosis of congenital adrenal hyperplasia.

Further studies have shown that LRS is suitable for the analysis of genes with important epidemiological significance and whose sequencing is complicated by the presence of pseudogenes, such as for spinal muscular atrophy (SMA). SMA diagnosis deserves particular attention because of recent great progress in the implementation of new therapies with outcomes critically related to early administration and consequently to the availability of efficient newborn screening tests for the rapid identification of affected children (Schorling et al., 2020). SMA is an autosomal recessive disease caused by biallelic mutations in *SMN1*, that has a high degree of homology with neighboring *SMN2* gene. The number of copies of *SMN2* is variable in the population and is inversely correlated with clinical severity (Kolb and Kissel, 2015). Most patients (95%) carry homozygous deletion of *SMN1*; less frequently, the mutation is represented by a SNV/indel (Bai et al., 2024). Carrier screening in parents or in healthy at-risk population is needed to assess recurrence risk in offspring. Current methods for SMA diagnosis involve a multiple-step, time-consuming process, making use of *SMN1* dosage tests (MLPA or quantitative PCR for copy number detection), followed by *SMN1* sequencing when a heterozygous deletion is identified (Bai et al., 2024). Moreover, conventional dosage tests for *SMN1* fail to identify the *SMN1* 2 + 0 genotype, a carrier status characterized by two copies of *SMN1* on the same chromosome. To overcome these limitations, LRS is being tested as a single-step method for SMA diagnosis and carrier screening. It correctly identified CNVs, SNVs/indels and large deletions in *SMN1*, *SMN2* copy number and modifying SNVs (Bai et al., 2024) and detected *SMN1* 2 + 0 carriers (Chen et al., 2023). Additional large-scale studies are needed to safely implement the technique as a single-test screening for this condition, but the results to date are promising.

4.2 Haplotype phasing

In autosomal recessive diseases, haplotype phasing of variants is essential to determine the genetic diagnosis in compound heterozygote cases. According to the guidelines of the American College of Medical Genetics (ACMG), the detection of a genetic variant in *trans* with a second pathogenic mutation is moderate evidence of pathogenicity (PM3) and this criterion can help redefine variants of unknown significance (VUS) as likely pathogenic/ pathogenic (Richards et al., 2015). However, variant phasing by short-read or Sanger sequencing requires the analysis of patients' parents or other family members and can be complicated by their unavailability. This situation is unfortunately frequently encountered in clinical practice. Recently, Gupta et al. reported on a proband affected by visual and hearing impairment in whom genetic testing revealed two variants in the *USH2A* gene, a pathogenic variant and a VUS. Parental testing was not possible, but a single LRS read covering the genomic positions of both

variants demonstrated their localization in *trans*. This allowed the reclassification of the VUS as likely pathogenic and confirmed the diagnosis of Usher syndrome in the proband (Gupta et al., 2023).

Diagnostic uncertainty in autosomal recessive diseases also remains when parental samples are available, but one of the two variants occurred *de novo* in the proband. Again, LRS can resolve the phasing of variants and define their clinical relevance (Zhao et al., 2021).

The identification of the parental allele affected by a *de novo* variant, can have important implications not only for autosomal recessive genes. Indeed, in imprinting disorders, the parental origin of the mutated allele determines its clinical significance. For instance, Watson and colleagues reported on a proband with a clinical diagnosis of Angelman syndrome (AS) and a potentially pathogenic heterozygous variant in *UBE3A*. To confirm the diagnosis, the maternal origin of the mutation should have been verified, but DNA analysis of both parents was negative for the specific variant. LRS allowed the reconstruction of the parental haplotypes and demonstrated the maternal origin of the mutated allele (Watson et al., 2022). This example illustrates the usefulness of LRS in imprinting disorders besides its ability of detecting methylation alterations (see below).

4.3 Lost in SRS

Most SRS assays just study the coding sequence and the canonical splice sites through whole exome sequencing (WES). Besides the identification of missed coding variants in difficult-to-sequence regions, several case reports and small-cohort studies showed how LRS helped the detection of intronic variants located outside of the canonical splice-site that weren't previously identified by standard genetic techniques (Bruels et al., 2022; Miller et al., 2022; Viering et al., 2023). As a matter of fact, most of these variants would likely be identified with short-read WGS and limitations remain in computational prediction algorithms for deep intronic variant interpretation. Nevertheless, there are advantages in LRS over WGS, such as the ability to simultaneously detect potential SVs and phasing the different variants, in addition to a possible cost-effectiveness when targeted approaches are applied. We suggest that LRS might transiently become a valid second-line approach for the diagnosis of patients with inconclusive findings after standard analyses, such as SRS gene-panels or WES. However, its application as first-tier diagnostic test will have to wait for technological advances solving technical limitations and the development of analysis software easily manageable by biologists without bioinformatic expertise.

5 Methylation changes at imprinted genomic regions

Detecting methylation differences can be useful in defining the diagnostic classification of DNA sequence changes, such as in repeat expansion disorders, where LRS can simultaneously detect both DNA sequence alterations and base modifications (Miller et al., 2022). Nevertheless, at present the main diagnostic application of

methylation analyses concerns the identification of defects responsible for imprinting disorders. Imprinting disorders are caused by genetic or epigenetic alterations affecting differentially methylated regions that regulate the monoallelic and parent-of-origin-dependent expression of imprinted genes. The underlying molecular defects are heterogeneous (uniparental disomy [UPD], CNVs, SNVs and methylation changes), contribute with different frequencies to the occurrence of the different clinical conditions, and sometimes arise in postzygotic cells. Molecular heterogeneity, mosaicism and the frequent absence of specific clinical manifestations make diagnosis challenging. Identifying the molecular cause is important not only for better clinical management, as for other genetic diseases, but also for surveillance of tumor risk occurring in some imprinting disorders (Beygo et al., 2020).

5.1 Current diagnostic analyses for imprinting disorders

Accurate estimates of diagnostic yield for individual imprinting disorders are not available as this is influenced by the stringency level of referral clinical indications resulting in lower diagnostic rates in patients routinely analyzed by genetic laboratories compared to those achieved in clinically well-characterized cohorts (Mackay et al., 2022; Eggermann et al., 2023). The diagnostic workflow of a specific disorder depends on the frequencies of the different molecular mechanisms and on the expertise of laboratories, but generally, for most conditions, such as AS and Prader-Willi (PWS) or Beckwith-Wiedemann (BWS) and Silver Russell syndrome, mainly due to UPD, CNVs or imprinting defects, all leading to changes in the methylation pattern at specific genomic loci, it starts with tests targeting these epigenetic modifications.

Epigenetic modifications are often studied using bisulfite conversion of unmethylated cytosines to uracil, that are eventually converted to thymidines during PCR amplification, while 5-methylated cytosines (5 mC) are protected from conversion. These methods allow genome-wide methylation profiling at a single-base level, although C>T conversion creates mapping difficulties and bisulfite treatment can alter DNA integrity.

Methylation-Sensitive (MS)-MLPA, which simultaneously detects altered methylation and CNVs, has replaced MS bisulfite conversion-dependent methods and is now routinely applied as a first-tier test for clinically suspected imprinting disorders. This technique analyzes native DNA through methylation-sensitive restriction enzyme digestion, but cannot distinguish between UPD and imprinting defects with few exceptions, needing second-line assays to identify the exact underlying defect and thus enabling adequate genetic counseling with accurate estimation of recurrence risks (Beygo et al., 2020; Eggermann et al., 2023). Reduced sensitivity for low-level mosaic alterations and allele dropout are major limitations, with the latter mitigated by using more probes for single loci. Additionally, commercial MS-MLPA kits generally test for single or few chromosomes requiring the sequential application of more kits in patients with phenotypic features compatible with clinically overlapping imprinting disorders and are unsuitable for multi-locus imprinting disturbances. The

adoption of a multi-locus imprinting test might be useful to increase diagnostic yield (Mackay et al., 2022). Microsatellite analysis is the gold standard test for UPD involving either whole or partial chromosomes (segmental UPD). Probands and both parents must be tested in parallel to establish the parental origin of all the markers analyzed. The same trio strategy should be applied when using SNP-arrays to be able to detect heterodisomy in addition to isodisomy (Beygo et al., 2020).

5.2 Application of LRS to the diagnosis of imprinting disorders

A single assay investigating all the possible molecular pathogenic mechanisms simultaneously will speed up diagnosis, prognosis and therapy, through the rapid definition of (epi)genotype-phenotype correlations. LRS allows the detection of modified or methylated nucleotides (including N6-methyladenine, 5mC and 5-hydroxymethylcytosine) directly in the native genomic DNA during canonical DNA sequencing, without bisulfite treatment. In real-time single molecule sequencing approaches, methylation impacts the kinetics of the polymerase while incorporating fluorescently labeled nucleotides into nucleic acid strands during DNA sequencing: base modifications alter both pulse width and interpulse duration, allowing the detection of epigenetics changes (Flusberg et al., 2010). In nanopore-based technologies, the electric current pattern is different for modified and non-modified bases and methylation state can be predicted with specific tools (Liu et al., 2021).

Few studies tested the clinical application of LRS for the diagnosis of imprinting disorders, probably due to the still high costs of the technology. Long-read genome sequencing has been recently used to analyze two patients with imprinting disorders, identifying a loss of methylation at IC2 locus at 11p15 responsible for BWS in one case and maternal UPD at 15q11 causing PWS in the other case (Dixon et al., 2022). The genome-wide approach also allowed researchers to investigate the additive contribution of possibly pathogenic variants in different genes to the clinical picture of patients. More patients (18) with a clinical diagnosis of AS or PWS were analyzed through LRS and standard clinical testing by Paschal et al., 2023, observing a 100% concordance between methods, without specifying the molecular causes.

As mentioned before, a possible solution to decrease costs is the adoption of a target strategy.

Yamada et al., 2023 demonstrated that targeted LRS through adaptive sampling might be an efficient one-step assay for the diagnosis of AS and PWS and potentially other imprinting disorders. They correctly identified alleles with altered methylation, which is sufficient for diagnosis, and were allowed to discriminate large and very small (imprinting center) deletions, uniparental isodisomy and heterodisomy, which is useful for genetic counseling and clinical management. Costs could be further reduced by including additional clinically relevant regions in the target analysis, as also technically recommended for the adaptive sampling approach that requires relatively large target size.

More data about sensitivity in additional disease-causing differentially methylated regions and for mosaic alterations of methylation are needed to confirm LRS clinical utility for the diagnosis of imprinted disorders.

6 Alterations of gene expression: to be included within diagnostic workflows

RNA sequencing (RNAseq) technologies allow the quantitative and qualitative analysis of gene expression in different cells and tissues, enabling the reconstruction of the transcriptome, that is, the set of coding and noncoding transcripts. Among its potential applications, RNAseq can be of use in the diagnosis of rare genetic disorders to identify differentially expressed isoforms and alternative splicing, detect gene fusions and predict the functional impact of genetic variants (especially non-coding variants), increasing the diagnostic yield by up to 15% compared with WES alone (Peymani et al., 2022).

The presence of diverse transcripts for a single genomic locus can raise problems in the assembly process, complicating transcript identification from short reads (de Jong et al., 2017). LRS can overcome these difficulties and simultaneously retrieve phasing information, although there are still few examples of its application in a clinical setting, just as the implementation of RNA analysis in diagnostic workflows is still limited.

Dainis et al., 2019 reported on a patient with severe hypertrophic cardiomyopathy (HCM) in whom a splice-site variant was identified in *MYBPC3*, the most frequently mutated gene in HCM. The variant was predicted to disrupt exon 19 to 20 splice junction. With targeted LRS, the authors identified a novel, highly expressed, alternatively spliced transcript (missing exon 20) in the patient's cardiomyocytes and, with phasing information, they reconstructed that the alternative isoform was generated by the mutant allele. In this case, RNAseq allowed the characterization of the variant and provided evidence of pathogenicity.

For an effective application of long-read RNAseq in a clinical setting, automated workflows are required. For example, a capture and ultradeep long-read RNA sequencing (CAPLRseq) workflow was developed and validated for the diagnosis of hereditary non-polyposis colorectal cancer/Lynch syndrome that enables the interpretation of variants affecting mRNA structure or expression in 123 hereditary cancer gene transcripts (Schwenk et al., 2023). This method allowed the reclassification of VUS in two patients (one likely benign and one likely pathogenic variant) and confirmed the presence of splicing defects or alterations in the expression of mismatch repair genes in 17 more patients. The variants detected included coding and non-coding SNVs and SVs that resulted in premature termination of transcript, splicing defects, monoallelic loss of expression, insertion of retrotransposons or formation of fusion transcripts.

Despite its usefulness in resolving the clinical role of uncertain variants, the implementation of long read RNAseq in clinical practice is still limited by the high cost, the practical difficulties in obtaining suitable tissues for analysis and the need for more automated workflows.

7 Discussion

After SRS, the last huge advancement in the field of genomic and genetic research is represented by the advent of LRS, which proved to be a powerful tool to rapidly analyze genomes, epigenomes and transcriptomes. Its ability to accurately sequence complex genomic regions allowed to sequence the first telomere-to-telomere complete human genome, which earned it the nomination as method of the year 2022. Although LRS is increasingly applied in research, its use in diagnostics is not yet part of routine clinical practice. In this review, we focused on the potential diagnostic applications of the technique, highlighting its advantages and limitations (Table 2).

One of the main benefits of LRS, which is clearly reflected in the available literature, is the detection of balanced and unbalanced SVs, including complex rearrangements, with high sensitivity and accuracy, through reliable alignment and precise breakpoint definition. Moreover, for the detection of expanded TRs, LRS outperforms current diagnostic techniques in both accuracy and practicality. Considering also the availability of efficient targeted solutions, we suggest that LRS could become the diagnostic method of choice for these types of genetic alterations in the next few years, providing significant improvements in terms of diagnostic yield and time to diagnosis. In line with these highlighted capabilities, many authors have reported the usefulness of LRS in a diagnostic setting, especially in cases previously unsolved by standard techniques and with a strongly suspected genetic etiology. However, at present LRS performances for SNV detection are not yet comparable to those of SRS, preventing its application as a first-tier diagnostic test. Although technical advancements have led over time to the release of new kits and protocols providing greatly improved sensitivity and accuracy, extensive studies are still needed to ascertain their clinical potential. By accessing difficult-to-sequence regions, an additional feasible diagnostic application, replacing current genetic tests, might be the analysis of disease-associated genes highly homologous to pseudogenes (e.g., *PKD1*, *CYP21A2*, *SMN1*), in which the high coverage achieved through targeted approaches might reduce the impact of sequencing errors and allow the identification of SNVs along with SVs, in a comprehensive one-step assay.

The ability to detect methylation changes gives LRS a huge advantage over previous sequencing techniques, but with a currently limited clinical impact on medical genetics. Indeed, methylation evaluation in genetic laboratories is aimed only at the diagnosis of imprinted disorders. The method mainly used in diagnostic routine, MS-MLPA, is sufficiently sensitive, robust and cost-effective and is based on easy wet lab protocols and very user-friendly analysis software. These features make the replacement of MS-MLPA with LRS highly unlikely at least in the next few years, despite LRS greater sensitivity for mosaic alterations and its ability to screen the whole genome and give complete information regarding the pathogenetic molecular mechanisms that have implications for the clinical management of patients and families. The emerging clinical relevance of altered methylation patterns, representing episignatures of specific genetic disorders (Aref-Eshghi et al., 2019), will probably open a new way for LRS to be useful in the field of medical epigenetics.

The present review focuses on diagnostic tests provided by genetic laboratories for the identification of rare genetic variants with strong effects on constitutional phenotypes or hereditary tumor risk. However, further future expansion of LRS clinical applications may include the dissection of cancer cellular heterogeneity to improve diagnosis and targeted therapies

TABLE 2 Advantages and limitations of long read sequencing clinical application for the detection of different genetic/epigenetic anomalies. SRS, short read sequencing; SNV, single nucleotide variant; SV, structural variant; TR: tandem repeat.

	SVs	TR expansion	SNVs	Methylation changes
ADVANTAGES	Accurate definition of breakpoints	Precise determination of the repeat number in individual alleles, regardless the length	Identification and correct mapping of variants in genes highly homologous to pseudogenes	Definition of the underlying pathogenic mechanism
	Mapping of duplicated regions	Improved estimation of the progenitor allele length	Direct haplotype phasing without the need of parental testing	Allele discrimination
	Resolution of complex rearrangements	Accurate identification of interruptions		High sensitivity for mosaic changes
	Identification of SVs involving repetitive regions	High sensitivity for mosaic expansions		
	High sensitivity for mosaic alterations	Parallel detection of methylation changes		
	Identification of missed alterations in the second allele of recessive genes			
	Rapid wet-lab and analysis protocols			
	Possibility of targeted analysis with easy multiplexing and customization (adaptive sampling)			
LIMITATIONS	Reduced accuracy for SNV detection compared to SRS			
	Interpretation difficulties			
	High costs (containable with targeted approaches)			
	Bioinformatic expertise required			

through single-cell genomics and the genotyping of common variants for polygenic risk score (PRS) estimation in complex diseases. Currently, these approaches do not have a clearly assessed clinical relevance and are affected by technical and methodological limitations preventing their implementation within diagnostic workflows, despite intense research and evidence provided on their potential.

Indeed, single-cell analysis, even combined to LRS, is mainly used to investigate the biological complexity of cell types and tissues, especially in cancer, with studies primarily aimed at developing efficient and accurate analytical methods (Hård et al., 2023; Shiau et al., 2023; Wu and Schmitz, 2023; You et al., 2023). At present, published work in single-cell genomics remains firmly in the research space and exploration of clinical relevance, especially in the fields of oncology, immunology, and hematology, is still at an early stage (Kim N. et al., 2021; Lim et al., 2023).

Common variants can be combined to define a PRS, estimating the individual susceptibility to complex diseases. Since the clinical utility of this score has not been fully established, due to disease risk prediction challenges, low power in the general population and weak evidence in non-European ancestry, the clinical interest is limited to use PRS in specific cohorts with high prior probability of disease for risk stratification and identification of individuals to be addressed to screening programs, lifestyle modifications, or preventive treatment, when available and appropriate (Lewis and Vassos, 2020). Microarrays have generally been used to detect known common variants associated to diseases and clinically tested to infer PRS (Hao et al., 2022), and whole genome sequencing, extending the analysis to rare variants, may become an appealing alternative, especially through the cost-effective low-pass sequencing strategy (Kim S. et al., 2021;
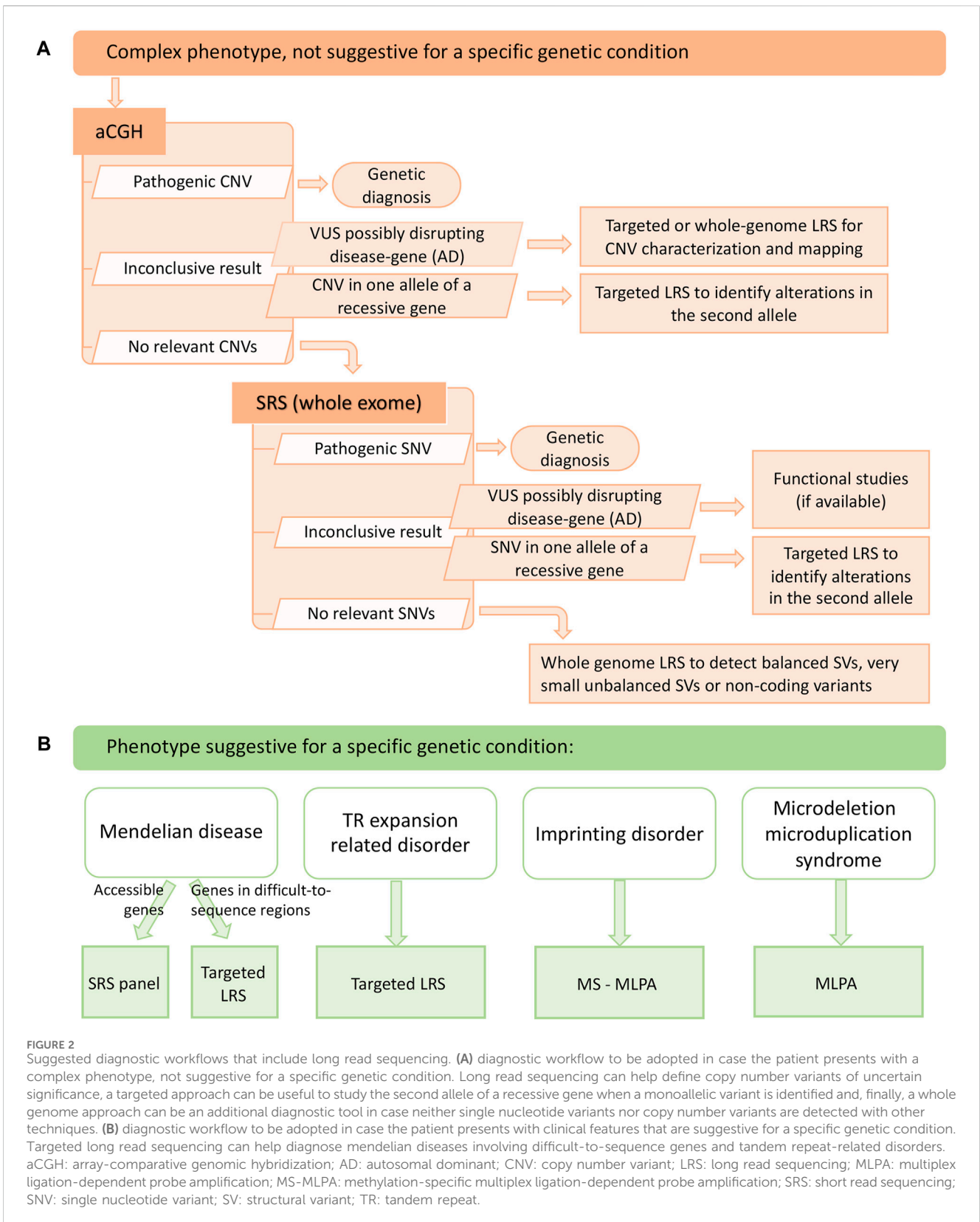
Bagger et al., 2024). An interesting saving-money approach is the analysis of discarded off-target reads from SRS gene panels to reconstruct genome-wide germline genotypes (Gusev et al., 2021). In these perspectives also LRS may give its contribution and a very recent study about targeted LRS on patients with suspected hereditary cancer shows that off-target reads can be effectively used to accurately genotype common SNPs across the entire genome, enabling PRS calculation (Nakamura et al., 2024).

Despite strong evidence of great diagnostic potential, for many laboratories the access to LRS is still complicated by the high cost of both instrumentation and reagents.

However, it should be considered that the high economic commitment for LRS analysis is mitigated by the feasibility of performing different levels of analysis (e.g., SNVs, SVs, methylation) in the same experiment, avoiding the use of multiple and sequential unnecessary tests. Moreover, in cases where a whole-genome approach is not necessary, the cost can be lowered by targeted strategies.

Besides costs, the widespread adoption of LRS in diagnostic laboratories could be slowed down by the high level of bioinformatic expertise required for data analysis, generally available only in big diagnostic centers, and by interpretation difficulties of the huge number and variety of detectable variants, especially due to the incomplete knowledge about genomic regions that have been inaccessible so far.

Considering advantages and limitations, we propose two diagnostic workflows for suspected genetic conditions with or without specific clinical hypotheses, where LRS could be implemented as a second-level test for completing results from previous investigations and as a first-tier test for TR-related disorders and diseases associated to genes in repetitive or GC-rich regions, especially those known to be frequently affected by SVs. Its remarkable ability to identify and characterize SVs should also promote the use of LRS in cases where aCGH and SRS are negative (Figure 2).

FIGURE 2
Suggested diagnostic workflows that include long read sequencing. **(A)** diagnostic workflow to be adopted in case the patient presents with a complex phenotype, not suggestive for a specific genetic condition. Long read sequencing can help define copy number variants of uncertain significance, a targeted approach can be useful to study the second allele of a recessive gene when a monoallelic variant is identified and, finally, a whole genome approach can be an additional diagnostic tool in case neither single nucleotide variants nor copy number variants are detected with other techniques. **(B)** diagnostic workflow to be adopted in case the patient presents with clinical features that are suggestive for a specific genetic condition. Targeted long read sequencing can help diagnose mendelian diseases involving difficult-to-sequence genes and tandem repeat-related disorders. aCGH: array-comparative genomic hybridization; AD: autosomal dominant; CNV: copy number variant; LRS: long read sequencing; MLPA: multiplex ligation-dependent probe amplification; MS-MLPA: methylation-specific multiplex ligation-dependent probe amplification; SRS: short read sequencing; SNV: single nucleotide variant; SV: structural variant; TR: tandem repeat.

Improvements in SNV detection and studies confirming their clinical accuracy are particularly awaited for the possibility of fully characterizing the genome in a single assay. Efforts are being made to overcome the limitations of LRS and make it more widely available.

Author contributions

GO: Writing–original draft, Writing–review and editing, Investigation, Visualization. EI: Writing–original draft,

Writing–review and editing, Investigation, Visualization. GI: Writing–original draft, Writing–review and editing, Investigation. DT: Writing–original draft, Writing–review and editing, Supervision. TP: Conceptualization, Writing–original draft, Writing–review and editing, Funding acquisition. PM: Supervision, Writing–original draft, Writing–review and editing, Investigation, Visualization.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alkan, C., Sajjadian, S., and Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nat. Methods* 8, 61–65. doi:10.1038/nmeth.1527

Amarasinghe, S. L., Ritchie, M. E., and Gouil, Q. (2021). long-read-tools.org: an interactive catalogue of analysis methods for long-read sequencing data. *GigaScience* 10, giab003. doi:10.1093/gigascience/giab003

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21, 30. doi:10.1186/s13059-020-1935-5

Aref-Eshghi, E., Bend, E. G., Colaiacovo, S., Caudle, M., Chakrabarti, R., Napier, M., et al. (2019). Diagnostic utility of genome-wide DNA methylation testing in genetically unsolved individuals with suspected hereditary conditions. *Am. J. Hum. Genet.* 104, 685–700. doi:10.1016/j.ajhg.2019.03.008

Athanasopoulou, K., Boti, M. A., Adamopoulos, P. G., Skourou, P. C., and Scorilas, A. (2022). Third-generation sequencing: the spearhead towards the radical transformation of modern genomics. *Life* 12, 30. doi:10.3390/life12010030

Bagger, F. O., Borgwardt, L., Jespersen, A. S., Hansen, A. R., Bertelsen, B., Kodama, M., et al. (2024). Whole genome sequencing in clinical practice. *BMC Med. Genomics* 17, 39. doi:10.1186/s12920-024-01795-w

Bahlo, M., Bennett, M. F., Degorski, P., Tankard, R. M., Delatycki, M. B., and Lockhart, P. J. (2018). Recent advances in the detection of repeat expansions with short-read next-generation sequencing. *F1000Res* 7, F1000 Faculty Rev-736. doi:10.12688/f1000research.13980.1

Bai, J., Qu, Y., Huang, W., Meng, W., Zhan, J., Wang, H., et al. (2024). A high-fidelity long-read sequencing-based approach enables accurate and effective genetic diagnosis of spinal muscular atrophy. *Clin. Chim. Acta Int. J. Clin. Chem.* 553, 117743. doi:10.1016/j.cca.2023.117743

Beygo, J., Kanber, D., Eggermann, T., and Begemann, M. (2020). Molecular testing for imprinting disorders. *Med. Genet.* 32, 305–319. doi:10.1515/medgen-2020-2048

Bolognini, D., and Magi, A. (2021). Evaluation of germline structural variant calling methods for nanopore sequencing data. *Front. Genet.* 12, 761791. doi:10.3389/fgene.2021.761791

Borràs, D. M., Vossen, RHAM, Liem, M., Buermans, H. P. J., Dauwerse, H., van Heusden, D., et al. (2017). Detecting PKD1 variants in polycystic kidney disease patients by single-molecule long-read sequencing. *Hum. Mutat.* 38, 870–879. doi:10.1002/humu.23223

Boycott, K. M., Hartley, T., Biesecker, L. G., Gibbs, R. A., Innes, A. M., Riess, O., et al. (2019). A diagnosis for all rare genetic diseases: the horizon and the next Frontiers. *Cell* 177, 32–37. doi:10.1016/j.cell.2019.02.040

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., et al. (2008). The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* 26, 1146–1153. doi:10.1038/nbt.1495

Bruels, C. C., Littel, H. R., Daugherty, A. L., Stafki, S., Estrella, E. A., McGaughy, E. S., et al. (2022). Diagnostic capabilities of nanopore long-read sequencing in muscular dystrophy. *Ann. Clin. Transl. Neurol.* 9, 1302–1309. doi:10.1002/acn3.51612

Cameron, D. L., Di Stefano, L., and Papenfuss, A. T. (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, C. M., and Lupski, J. R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17, 224–238. doi:10.1038/nrg.2015.25

Chen, X., Harting, J., Farrow, E., Thiffault, I, Kasperaviciute, D., Genomics England Research Consortium, , et al. (2023). Comprehensive SMN1 and SMN2 profiling for spinal muscular atrophy analysis using long-read PacBio HiFi sequencing. *Am. J. Hum. Genet.* 110, 240–250. doi:10.1016/j.ajhg.2023.01.001

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and 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

Chintalaphani, S. R., Pineda, S. S., Deveson, I. W., and Kumar, K. R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol. Commun.* 9, 98. doi:10.1186/s40478-021-01201-x

Corbett, M. A., Kroes, T., Veneziano, L., Bennett, M. F., Florian, R., Schneider, A. L., et al. (2019). Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. *Nat. Commun.* 10, 4920. doi:10.1038/s41467-019-12671-y

Cumming, S. A., Hamilton, M. J., Robb, Y., Gregory, H., McWilliam, C., Cooper, A., et al. (2018). *De novo* repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur. J. Hum. Genet.* 26, 1635–1647. doi:10.1038/s41431-018-0156-9

Dainis, A., Tseng, E., Clark, T. A., Hon, T., Wheeler, M., and Ashley, E. (2019). Targeted long-read RNA sequencing demonstrates transcriptional diversity driven by splice-site variation in MYBPC3. *Circ. Genom Precis. Med.* 12, e002464. doi:10.1161/CIRCGEN.119.002464

Damián, A., Núñez-Moreno, G., Jubin, C., Tamayo, A., de Alba, M. R., Villaverde, C., et al. (2023). Long-read genome sequencing identifies cryptic structural variants in congenital aniridia cases. *Hum. Genomics* 17, 45. doi:10.1186/s40246-023-00490-8

de Jong, L. C., Cree, S., Lattimore, V., Wiggins, G. A. R., Spurdle, A. B., Miller, A., et al. (2017). Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events. *Breast Cancer Res.* 19, 127. doi:10.1186/s13058-017-0919-1

Deng, J., Gu, M., Miao, Y., Yao, S., Zhu, M., Fang, P., et al. (2019). Long-read sequencing identified repeat expansions in the 5'UTR of the NOTCH2NLC gene from Chinese patients with neuronal intranuclear inclusion disease. *J. Med. Genet.* 56, 758–764. doi:10.1136/jmedgenet-2019-106268

Depienne, C., and Mandel, J. L. (2021). 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* 108, 764–785. doi:10.1016/j.ajhg.2021.03.011

Dixon, K., Shen, Y., Chin, H.-L., Gazzaz, N., Huynh, S., Chan, S., et al. (2022). eP343: long-read genome sequencing informs the molecular etiology of imprinting disorders. *Genet. Med.* 24, S214–S215. doi:10.1016/j.gim.2022.01.378

Dixon, K., Shen, Y., O'Neill, K., Mungall, K. L., Chan, S., Bilobram, S., et al. (2023). Defining the heterogeneity of unbalanced structural variation underlying breast cancer susceptibility by nanopore genome sequencing. *Eur. J. Hum. Genet. EJHG* 31, 602–606. doi:10.1038/s41431-023-01284-1

Dutta, U. R., Rao, S. N., Pidugu, V. K., V. S., V., Bhattacherjee, A., Bhowmik, A. D., et al. (2019). Breakpoint mapping of a novel *de novo* translocation t(X;20)(q11.1;p13) by positional cloning and long read sequencing. *Genomics* 111, 1108–1114. doi:10.1016/j.ygeno.2018.07.005

Ebbert, M. T. W., Farrugia, S. L., Sens, J. P., Jansen-West, K., Gendron, T. F., Prudencio, M., et al. (2018). Long-read sequencing across the C9orf72 "GGGGCC" repeat expansion: implications for clinical use and genetic discovery efforts in human disease. *Mol. Neurodegener.* 13, 46. doi:10.1186/s13024-018-0274-4

Ebert, P., Audano, P. A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M. J., et al. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* 372, eabf7117. doi:10.1126/science.abf7117

Eggermann, T., Monk, D., De Nanclares, G. P., Kagami, M., Giabicani, E., Riccio, A., et al. (2023). Imprinting disorders. *Nat. Rev. Dis. Primer* 9, 33. doi:10.1038/s41572-023-00443-4

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi:10.1126/science.1162986

Filser, M., Schwartz, M., Merchadou, K., Hamza, A., Villy, M.-C., Decees, A., et al. (2023). Adaptive nanopore sequencing to determine pathogenicity of BRCA1 exonic duplication. *J. Med. Genet.* 60, 1206–1209. doi:10.1136/jmg-2023-109155

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., et al. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465. doi:10.1038/nmeth.1459

Gall-Duncan, T., Sato, N., Yuen, R. K. C., and Pearson, C. E. (2022). Advancing genomic technologies and clinical awareness accelerates discovery of disease-associated tandem repeat sequences. *Genome Res.* 32, 1–27. doi:10.1101/gr.269530.120

Gupta, P., Nakamichi, K., Bonnell, A. C., Yanagihara, R., Radulovich, N., Hisama, F. M., et al. (2023). Familial co-segregation and the emerging role of long-read sequencing to re-classify variants of uncertain significance in inherited retinal diseases. *NPJ Genom Med.* 8, 20. doi:10.1038/s41525-023-00366-9

Gusev, A., Groha, S., Taraszka, K., Semenov, Y. R., and Zaitlen, N. (2021). Constructing germline research cohorts from the discarded reads of clinical tumor sequences. *Genome Med.* 13, 179. doi:10.1186/s13073-021-00999-4

Haghighi, A., Krier, J. B., Toth-Petroczy, A., Cassa, C. A., Frank, N. Y., Carmichael, N., et al. (2018). An integrated clinical program and crowdsourcing strategy for genomic sequencing and Mendelian disease gene discovery. *NPJ Genom Med.* 3, 21. doi:10.1038/s41525-018-0060-9

Hao, L., Kraft, P., Berriz, G. F., Hynes, E. D., Koch, C., Korategere, V., et al. (2022). Development of a clinical polygenic risk score assay and reporting workflow. *Nat. Med.* 28, 1006–1013. doi:10.1038/s41591-022-01767-6

Hård, J., Mold, J. E., Eisfeldt, J., Tellgren-Roth, C., Häggqvist, S., Bunikis, I., et al. (2023). Long-read whole-genome analysis of human single cells. *Nat. Commun.* 14, 5164. doi:10.1038/s41467-023-40898-3

He, W., Meng, G., Hu, X., Dai, J., Liu, J., Li, X., et al. (2022). Reclassification of DMD duplications as benign: recommendations for cautious interpretation of variants identified in prenatal screening. *Genes* 13, 1972. doi:10.3390/genes13111972

Hiatt, S. M., Lawlor, J. M. J., Handley, L. H., Ramaker, R. C., Rogers, B. B., Partridge, E. C., et al. (2021). Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders. *HGG Adv.* 2, 100023. doi:10.1016/j.xhgg.2021.100023

Hiramuki, Y., Kure, Y., Saito, Y., Ogawa, M., Ishikawa, K., Mori-Yoshimura, M., et al. (2022). Simultaneous measurement of the size and methylation of chromosome 4qA-D4Z4 repeats in facioscapulohumeral muscular dystrophy by long-read sequencing. *J. Transl. Med.* 20, 517. doi:10.1186/s12967-022-03743-7

Ho, S. S., Urban, A. E., and Mills, R. E. (2020). Structural variation in the sequencing era. *Nat. Rev. Genet.* 21, 171–189. doi:10.1038/s41576-019-0180-9

Höijer, I., Tsai, Y. C., Clark, T. A., Kotturi, P., Dahl, N., Stattin, E. L., et al. (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

Ibañez, K., Polke, J., Hagelstrom, R. T., Dolzhenko, E., Pasko, D., Thomas, E. R. A., et al. (2022). Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK: a retrospective diagnostic accuracy and prospective clinical validation study. *Lancet Neurol.* 21, 234–245. doi:10.1016/S1474-4422(21)00462-2

Kim, N., Eum, H. H., and Lee, H.-O. (2021a). Clinical perspectives of single-cell RNA sequencing. *Biomolecules* 11, 1161. doi:10.3390/biom11081161

Kim, S., Shin, J.-Y., Kwon, N.-J., Kim, C.-U., Kim, C., Lee, C. S., et al. (2021b). Evaluation of low-pass genome sequencing in polygenic risk score calculation for Parkinson's disease. *Hum. Genomics* 15, 58. doi:10.1186/s40246-021-00357-w

Kolb, S. J., and Kissel, J. T. (2015). Spinal muscular atrophy. *Neurol. Clin.* 33, 831–846. doi:10.1016/j.ncl.2015.07.004

Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., and 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

Lecoquierre, F., Quenez, O., Fourneaux, S., Coutant, S., Vezain, M., Rolain, M., et al. (2023). High diagnostic potential of short and long read genome sequencing with transcriptome analysis in exome-negative developmental disorders. *Hum. Genet.* 142, 773–783. doi:10.1007/s00439-023-02553-1

Lewis, C. M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* 12, 44. doi:10.1186/s13073-020-00742-5

Li, H., Zhu, X., Yang, Y., Wang, W., Mao, A., Li, J., et al. (2023). Long-read sequencing: an effective method for genetic analysis of CYP21A2 variation in congenital adrenal hyperplasia. *Clin. Chim. Acta* 547, 117419. doi:10.1016/j.cca.2023.117419

Liang, Q., Liu, Y., Duan, R., Meng, W., Zhan, J., Xia, J., et al. (2022). Comprehensive analysis of fragile X syndrome: full characterization of the FMR1 locus by long-read sequencing. *Clin. Chem.* 68, 1529–1540. doi:10.1093/clinchem/hvac154

Lim, J., Chin, V., Fairfax, K., Moutinho, C., Suan, D., Ji, H., et al. (2023). Transitioning single-cell genomics into the clinic. *Nat. Rev. Genet.* 24, 573–584. doi:10.1038/s41576-023-00613-w

Lin, B., Hui, J., and Mao, H. (2021). Nanopore technology and its applications in gene sequencing. *Biosensors* 11, 214. doi:10.3390/bios11070214

Liu, Q., Tong, Y., and Wang, K. (2020). Genome-wide detection of short tandem repeat expansions by long-read sequencing. *BMC Bioinforma.* 21, 542. doi:10.1186/s12859-020-03876-w

Liu, Y., Chen, M., Liu, J., Mao, A., Teng, Y., Yan, H., et al. (2022c). Comprehensive analysis of congenital adrenal hyperplasia using long-read sequencing. *Clin. Chem.* 68, 927–939. doi:10.1093/clinchem/hvac046

Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., et al. (2021). DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* 22, 295. doi:10.1186/s13059-021-02510-z

Liu, Z., Roberts, R., Mercer, T. R., Xu, J., Sedlazeck, F. J., and Tong, W. (2022a). Towards accurate and reliable resolution of structural variants for clinical diagnosis. *Genome Biol.* 23, 68. doi:10.1186/s13059-022-02636-8

Liu, Z., Zhao, G., Xiao, Y., Zeng, S., Yuan, Y., Zhou, X., et al. (2022b). Profiling the genome-wide landscape of short tandem repeats by long-read sequencing. *Front. Genet.* 13, 810595. doi:10.3389/fgene.2022.810595

Logsdon, G. A., Vollger, M. R., and Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* 21, 597–614. doi:10.1038/s41576-020-0236-x

Long, J., Sun, L., Gong, F., Zhang, C., Mao, A., Lu, Y., et al. (2022). Third-generation sequencing: a novel tool detects complex variants in the α-thalassemia gene. *Gene* 822, 146332. doi:10.1016/j.gene.2022.146332

Mackay, D., Bliek, J., Kagami, M., Tenorio-Castano, J., Pereda, A., Brioude, F., et al. (2022). First step towards a consensus strategy for multi-locus diagnostic testing of imprinting disorders. *Clin. Epigenetics* 14, 143. doi:10.1186/s13148-022-01358-9

Magini, P., Mingrino, A., Gega, B., Mattei, G., Semeraro, R., Bolognini, D., et al. (2022). Third-generation cytogenetic analysis: diagnostic application of long-read sequencing. *J. Mol. Diagn JMD* 24, 711–718. doi:10.1016/j.jmoldx.2022.03.013

Mandelker, D., Schmidt, R. J., Ankala, A., McDonald Gibson, K., Bowser, M., Sharma, H., et al. (2016). Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet. Med.* 18, 1282–1289. doi:10.1038/gim.2016.58

Mangin, A., de Pontual, L., Tsai, Y. C., Monteil, L., Nizon, M., Boisseau, P., et al. (2021). Robust detection of somatic mosaicism and repeat interruptions by long-read targeted sequencing in myotonic dystrophy type 1. *Int. J. Mol. Sci.* 22, 2616. doi:10.3390/ijms22052616

Mantovani, V., Bin, S., Graziano, C., Capelli, I., Minardi, R., Aiello, V., et al. (2020). Gene panel analysis in a large cohort of patients with autosomal dominant polycystic kidney disease allows the identification of 80 potentially causative novel variants and the characterization of a complex genetic architecture in a subset of families. *Front. Genet.* 11, 464. doi:10.3389/fgene.2020.00464

Mariya, T., Shichiri, Y., Sugimoto, T., Kawamura, R., Miyai, S., Inagaki, H., et al. (2023). Clinical application of long-read nanopore sequencing in a preimplantation genetic testing pre-clinical workup to identify the junction for complex Xq chromosome rearrangement-related disease. *Prenat. Diagn* 43, 304–313. doi:10.1002/pd.6334

Mastrorosa, F. K., Miller, D. E., and Eichler, E. E. (2023). Applications of long-read sequencing to Mendelian genetics. *Genome Med.* 15, 42. doi:10.1186/s13073-023-01194-3

Melas, M., Kautto, E. A., Franklin, S. J., Mori, M., McBride, K. L., Mosher, T. M., et al. (2022). Long-read whole genome sequencing reveals HOXD13 alterations in synpolydactyly. *Hum. Mutat.* 43, 189–199. doi:10.1002/humu.24304

Merker, J. D., Wenger, A. M., Sneddon, T., Grove, M., Zappala, Z., Fresard, L., et al. (2018). Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet. Med.* 20, 159–163. doi:10.1038/gim.2017.86

Miao, H., Zhou, J., Yang, Q., Liang, F., Wang, D., Ma, N., et al. (2018). Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis. *Hereditas* 155, 32. doi:10.1186/s41065-018-0069-1

Middelkamp, S., Vlaar, J. M., Giltay, J., Korzelius, J., Besselink, N., Boymans, S., et al. (2019). Prioritization of genes driving congenital phenotypes of patients with *de novo* genomic structural variants. *Genome Med.* 11, 79. doi:10.1186/s13073-019-0692-0

Miller, D. E., Lee, L., Galey, M., Kandhaya-Pillai, R., Tischkowitz, M., Amalnath, D., et al. (2022). Targeted long-read sequencing identifies missing pathogenic variants in unsolved Werner syndrome cases. *J. Med. Genet.* 59, 1087–1094. doi:10.1136/jmedgenet-2022-108485

Miller, D. E., Sulovari, A., Wang, T., Loucks, H., Hoekzema, K., Munson, K. M., et al. (2021). Targeted long-read sequencing identifies missing disease-causing variation. *Am. J. Hum. Genet.* 108, 1436–1449. doi:10.1016/j.ajhg.2021.06.006

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., et al. (2010). 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. doi:10.1016/j.ajhg.2010.04.006

Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., et al. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature* 470, 59–65. doi:10.1038/nature09708

Mitsuhashi, S., Frith, M. C., and Matsumoto, N. (2021). 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. doi:10.1186/s12920-020-00853-3

Mitsuhashi, S., and Matsumoto, N. (2020). Long-read sequencing for rare human genetic diseases. *J. Hum. Genet.* 65, 11–19. doi:10.1038/s10038-019-0671-8

Miyatake, S., Koshimizu, E., Fujita, A., Doi, H., Okubo, M., Wada, T., et al. (2022a). Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. *NPJ Genom Med.* 7, 62. doi:10.1038/s41525-022-00331-y

Miyatake, S., Yoshida, K., Koshimizu, E., Doi, H., Yamada, M., Miyaji, Y., et al. (2022b). Repeat conformation heterogeneity in cerebellar ataxia, neuropathy, vestibular areflexia syndrome. *Brain* 145, 1139–1150. doi:10.1093/brain/awab363

Mizuguchi, T., Suzuki, T., Abe, C., Umemura, A., Tokunaga, K., Kawai, Y., et al. (2019). A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. *J. Hum. Genet.* 64, 359–368. doi:10.1038/s10038-019-0569-5

Mizuguchi, T., Toyota, T., Miyatake, S., Mitsuhashi, S., Doi, H., Kudo, Y., et al. (2021). Complete sequencing of expanded SAMD12 repeats by long-read sequencing and Cas9-mediated enrichment. *Brain J. Neurol.* 144, 1103–1117. doi:10.1093/brain/awab021

Mousavi, N., Shleizer-Burko, S., Yanicky, R., and Gymrek, M. (2019). Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* 47, e90. doi:10.1093/nar/gkz501

Nakamura, W., Hirata, M., Oda, S., Chiba, K., Okada, A., Mateos, R. N., et al. (2024). Assessing the efficacy of target adaptive sampling long-read sequencing through hereditary cancer patient genomes. *NPJ Genomic Med.* 9, 11. doi:10.1038/s41525-024-00394-z

Neerman, N., Faust, G., Meeks, N., Modai, S., Kalfon, L., Falik-Zaccai, T., et al. (2019). A clinically validated whole genome pipeline for structural variant detection and analysis. *BMC Genomics* 20, 545. doi:10.1186/s12864-019-5866-z

Ni, Y., Liu, X., Simeneh, Z. M., Yang, M., and Li, R. (2023). Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Comput. Struct. Biotechnol. J.* 21, 2352–2364. doi:10.1016/j.csbj.2023.03.038

Nurchis, M. C., Altamura, G., Riccardi, M. T., Radio, F. C., Chillemi, G., Bertini, E. S., et al. (2023). Whole genome sequencing diagnostic yield for paediatric patients with suspected genetic disorders: systematic review, meta-analysis and GRADE assessment. *Arch. Public Health* 81, 93. doi:10.1186/s13690-023-01112-4

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., et al. (2022). The complete sequence of a human genome. *Science* 376, 44–53. doi:10.1126/science.abj6987

Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., et al. (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* 30, 1291–1305. doi:10.1101/gr.263566.120

Ohori, S., Miyauchi, A., Osaka, H., Lourenco, C. M., Arakaki, N., Sengoku, T., et al. (2023). Biallelic structural variations within FGF12 detected by long-read sequencing in epilepsy. *Life Sci. Alliance* 6, e202302025. doi:10.26508/lsa.202302025

Okada, E., Morisada, N., Horinouchi, T., Fujii, H., Tsuji, T., Miura, M., et al. (2022). Detecting MUC1 variants in patients clinicopathologically diagnosed with having autosomal dominant tubulointerstitial kidney disease. *Kidney Int. Rep.* 7, 857–866. doi:10.1016/j.ekir.2021.12.037

Olson, N. D., Wagner, J., McDaniel, J., Stephens, S. H., Westreich, S. T., Prasanna, A. G., et al. (2022). PrecisionFDA Truth Challenge V2: calling variants from short and long reads in difficult-to-map regions. *Cell Genom* 2, 100129. doi:10.1016/j.xgen.2022.100129

Paschal, C., Beck, A., Gillentine, M., Narayanan, J., Reed, P., Galey, M., et al. (2023). O33: concordance of long-read genome sequencing with methylation calling with clinical testing for individuals with Prader-Willi or Angelman Syndrome. *Genet. Med. Open* 1, 100459. doi:10.1016/j.gimo.2023.100459

Payne, A., Holmes, N., Rakyan, V., and Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35, 2193–2198. doi:10.1093/bioinformatics/bty841

Peymani, F., Farzeen, A., and Prokisch, H. (2022). RNA sequencing role and application in clinical diagnostic. *Pediatr. Investig.* 6, 29–35. doi:10.1002/ped4.12314

Rasmussen, A., Hildonen, M., Vissing, J., Duno, M., Tümer, Z., and Birkedal, U. (2022). High resolution analysis of DMPK hypermethylation and repeat interruptions in myotonic dystrophy type 1. *Genes Basel* 13, 970. doi:10.3390/genes13060970

Read, J. L., Davies, K. C., Thompson, G. C., Delatycki, M. B., and Lockhart, P. J. (2023). Challenges facing repeat expansion identification, characterisation, and the pathway to discovery. *Emerg. Top. Life Sci.* 7, 339–348. doi:10.1042/ETLS20230019

Reiner, J., Pisani, L., Qiao, W., Singh, R., Yang, Y., Shi, L., et al. (2018). Cytogenomic identification and long-read single molecule real-time (SMRT) sequencing of a Bardet-Biedl Syndrome 9 (BBS9) deletion. *NPJ Genom Med.* 3, 3. doi:10.1038/s41525-017-0042-3

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* 17, 405–424. doi:10.1038/gim.2015.30

Sabatella, M., Mantere, T., Waanders, E., Neveling, K., Mensenkamp, A. R., van Dijk, F., et al. (2021). Optical genome mapping identifies a germline retrotransposon insertion in SMARCB1 in two siblings with atypical teratoid rhabdoid tumors. *J. Pathol.* 255, 202–211. doi:10.1002/path.5755

Sanchis-Juan, A., Stephens, J., French, C. E., Gleadall, N., Mégy, K., Penkett, C., et al. (2018). Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Med.* 10, 95. doi:10.1186/s13073-018-0606-6

Sano, Y., Koyanagi, Y., Wong, J. H., Murakami, Y., Fujiwara, K., Endo, M., et al. (2022). Likely pathogenic structural variants in genetically unsolved patients with retinitis pigmentosa revealed by long-read sequencing. *J. Med. Genet.* 59, 1133–1138. doi:10.1136/jmedgenet-2022-108428

Scharf, F., Leal Silva, R. M., Morak, M., Hastie, A., Pickl, J. M. A., Sendelbach, K., et al. (2022). Constitutional chromothripsis of the APC locus as a cause of genetic predisposition to colon cancer. *J. Med. Genet.* 59, 976–983. doi:10.1136/jmedgenet-2021-108147

Schorling, D. C., Pechmann, A., and Kirschner, J. (2020). Advances in treatment of spinal muscular atrophy - new phenotypes, new challenges, new implications for care. *J. Neuromuscul. Dis.* 7, 1–13. doi:10.3233/JND-190424

Schwenk, V., Leal Silva, R. M., Scharf, F., Knaust, K., Wendlandt, M., Häusser, T., et al. (2023). Transcript capture and ultradeep long-read RNA sequencing (CAPLRseq) to diagnose HNPCC/Lynch syndrome. *J. Med. Genet.* 60, 747–759. doi:10.1136/jmg-2022-108931

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

Shiau, C.-K., Lu, L., Kieser, R., Fukumura, K., Pan, T., Lin, H.-Y., et al. (2023). High throughput single cell long-read sequencing analyses of same-cell genotypes and phenotypes in human tumors. *Nat. Commun.* 14, 4124. doi:10.1038/s41467-023-39813-7

Simpson, B. P., Yrigollen, C. M., Izda, A., and Davidson, B. L. (2023). Targeted long-read sequencing captures CRISPR editing and AAV integration outcomes in brain. *Mol. Ther. J. Am. Soc. Gene Ther.* 31, 760–773. doi:10.1016/j.ymthe.2023.01.004

Srivastava, S., Love-Nichols, J. A., Dies, K. A., Ledbetter, D. H., Martin, C. L., Chung, W. K., et al. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet. Med.* 21, 2413–2421. doi:10.1038/s41436-019-0554-6

Stevanovski, I., Chintalaphani, S. R., Gamaarachchi, H., Ferguson, J. M., Pineda, S. S., Scriba, C. K., 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

Tantirukdham, N., Sahakitrungruang, T., Chaisiwamongkol, R., Pongpanich, M., Srichomthong, C., Assawapitaksakul, A., et al. (2022). Long-read amplicon sequencing of the CYP21A2 in 48 Thai patients with steroid 21-hydroxylase deficiency. *J. Clin. Endocrinol. Metab.* 107, 1939–1947. doi:10.1210/clinem/dgac187

Thibodeau, M. L., O'Neill, K., Dixon, K., Reisle, C., Mungall, K. L., Krzywinski, M., et al. (2020). Improved structural variant interpretation for hereditary cancer susceptibility using long-read sequencing. *Genet. Med. Off. J. Am. Coll. Med. Genet.* 22, 1892–1897. –7. doi:10.1038/s41436-020-0880-8

Tsai, Y. C., de Pontual, L., Heiner, C., Stojkovic, T., Furling, D., Bassez, G., et al. (2022). Identification of a CCG-enriched expanded allele in patients with myotonic dystrophy type 1 using amplification-free long-read sequencing. *J. Mol. Diagn* 24, 1143–1154. doi:10.1016/j.jmoldx.2022.08.003

Vaché, C., Puechberty, J., Faugère, V., Darmaisin, F., Liquori, A., Baux, D., et al. (2020). A 4.6 Mb inversion leading to PCDH15-linc00844 and BICC1-PCDH15 fusion

transcripts as a new pathogenic mechanism implicated in usher syndrome type 1. *Front. Genet.* 11, 623. doi:10.3389/fgene.2020.00623

van der Sanden, BPGH, Schobers, G., Corominas Galbany, J., Koolen, D. A., Sinnema, M., van Reeuwijk, J., et al. (2023). The performance of genome sequencing as a first-tier test for neurodevelopmental disorders. *Eur. J. Hum. Genet.* 31, 81–88. doi:10.1038/s41431-022-01185-9

van Dijk, E. L., Naquin, D., Gorrichon, K., Jaszczyszyn, Y., Ouazahrou, R., Thermes, C., et al. (2023). Genomics in the long-read sequencing era. *Trends Genet.* 39, 649–671. doi:10.1016/j.tig.2023.04.006

Viering, DHHM, Hureaux, M., Neveling, K., Latta, F., Kwint, M., Blanchard, A., et al. (2023). Long-read sequencing identifies novel pathogenic intronic variants in gitelman syndrome. *J. Am. Soc. Nephrol.* 34, 333–345. doi:10.1681/ASN.2022050627

Wagner, J., Olson, N. D., Harris, L., McDaniel, J., Cheng, H., Fungtammasan, A., et al. (2022). Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat. Biotechnol.* 40, 672–680. doi:10.1038/s41587-021-01158-1

Watson, C. M., Jackson, L., Crinnion, L. A., Bonthron, D. T., and Sheridan, E. (2022). Long-read sequencing to resolve the parent of origin of a *de novo* pathogenic *UBE3A* variant. *J. Med. Genet.* 59, 1082–1086. doi:10.1136/jmedgenet-2021-108314

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., 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

Wieben, E. D., Aleff, R. A., Basu, S., Sarangi, V., Bowman, B., McLaughlin, I. J., et al. (2019). Amplification-free long-read sequencing of TCF4 expanded trinucleotide repeats in Fuchs Endothelial Corneal Dystrophy. *PLoS One* 14, e0219446. doi:10.1371/journal.pone.0219446

Wu, S., and Schmitz, U. (2023). Single-cell and long-read sequencing to enhance modelling of splicing and cell-fate determination. *Comput. Struct. Biotechnol. J.* 21, 2373–2380. doi:10.1016/j.csbj.2023.03.023

Xie, Z., Sun, C., Liu, Y., Yu, M., Zheng, Y., Meng, L., et al. (2021). Practical approach to the genetic diagnosis of unsolved dystrophinopathies: a stepwise strategy in the genomic era. *J. Med. Genet.* 58, 743–751. doi:10.1136/jmedgenet-2020-107113

Xie, Z., Sun, C., Zhang, S., Liu, Y., Yu, M., Zheng, Y., et al. (2020). Long-read whole-genome sequencing for the genetic diagnosis of dystrophinopathies. *Ann. Clin. Transl. Neurol.* 7, 2041–2046. doi:10.1002/acn3.51201

Xu, R., Li, H., Yi, S., Du, J., Jin, J., Qin, Y., et al. (2023). Identification of a novel 10.3 kb deletion causing α0-thalassemia by third-generation sequencing: pedigree analysis and genetic diagnosis. *Clin. Biochem.* 113, 64–69. doi:10.1016/j.clinbiochem.2022.12.018

Yamada, M., Okuno, H., Okamoto, N., Suzuki, H., Miya, F., Takenouchi, T., et al. (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

Yamaguchi, K., Kasajima, R., Takane, K., Hatakeyama, S., Shimizu, E., Yamaguchi, R., et al. (2021). Application of targeted nanopore sequencing for the screening and determination of structural variants in patients with Lynch syndrome. *J. Hum. Genet.* 66, 1053–1060. doi:10.1038/s10038-021-00927-9

You, Y., Prawer, Y. D. J., De Paoli-Iseppi, R., Hunt, C. P. J., Parish, C. L., Shim, H., et al. (2023). Identification of cell barcodes from long-read single-cell RNA-seq with BLAZE. *Genome Biol.* 24, 66. doi:10.1186/s13059-023-02907-y

Zeng, S., Zhang, M. Y., Wang, X. J., Hu, Z. M., Li, J. C., Li, N., et al. (2019). Long-read sequencing identified intronic repeat expansions in SAMD12 from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *J. Med. Genet.* 56, 265–270. doi:10.1136/jmedgenet-2018-105484

Zhao, A., Shen, J., Ding, Y., Sheng, M., Zuo, M., Lv, H., et al. (2021). Long-read sequencing identified a novel nonsense and a *de novo* missense of PPA2 in trans in a Chinese patient with autosomal recessive infantile sudden cardiac failure. *Clin. Chim. Acta* 519, 163–171. doi:10.1016/j.cca.2021.03.029

Zhuang, J., Chen, C., Fu, W., Wang, Y., Zhuang, Q., Lu, Y., et al. (2023). Third-generation sequencing as a new comprehensive technology for identifying rare α- and β-globin gene variants in thalassemia alleles in the Chinese population. *Arch. Pathol. Lab. Med.* 147, 208–214. doi:10.5858/arpa.2021-0510-OA

Zook, J. M., Hansen, N. F., Olson, N. D., Chapman, L., Mullikin, J. C., Xiao, C., et al. (2020). A robust benchmark for detection of germline large deletions and insertions. *Nat. Biotechnol.* 38, 1347–1355. doi:10.1038/s41587-020-0538-8