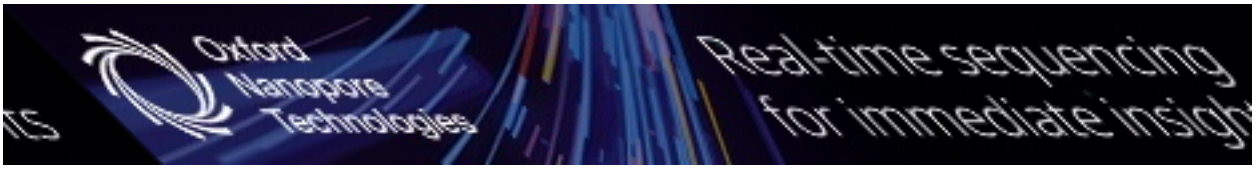
Assessing the readiness of Oxford Nanopore sequencing for clinical genomics applications

Judith Arres, Santosh Elavalli, Shalini Behl, et al.

Genome Res. 2026 36: 460-471 originally published online February 11, 2026
Access the most recent version at doi:10.1101/gr.280134.124



Assessing the readiness of Oxford Nanopore sequencing for clinical genomics applications

Judith Arres,[1,5] Santosh Elavalli,[1,5] Shalini Behl,[1] Daniel Matias Sanchez,[1] Ayesha Al Ali,[1] Abdelrahman Ahmed Yehia Abdelaziz Saad,[1] Azza Attia,[1] Cyla Minas,[1] Sharika Pariyachery,[1] Shariq Ahmed,[1] Fatmah Aldhuhoori,[1] Nitu Thulasidharan,[1] Gurunath Katagi,[1] Omar Soliman,[1] Shilp Purohit,[1] Vinay Kusuma,[1] Raony Cardenas,[1] Thyago Cardoso,[1] Luis F. Paulin,[2] Philippe Sanio,[2] Joseph Mafofo,[1] Haiguo Wu,[1] Val Zvereff,[1] Albarah El-Khani,[1] Fahed Al Marzooqi,[1] Tiago R. Magalhães,[1] Fritz J. Sedlazeck,[2,3,4] and Javier Quilez[1]

[1]M42, Abu Dhabi, United Arab Emirates; [2]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA; [3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; [4]Department of Computer Science, Rice University, Houston, Texas 77251-1892, USA

Long-read sequencing (LRS) technologies, namely, Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio), have emerged as promising solutions to overcome the limitations of short-read sequencing (SRS). Nevertheless, the still higher sequencing error rates compared with SRS, need for customized pipelines, rapidly updating software, and incipient scalability continue to present challenges for adopting ONT in standard clinical practice. Here we assess the performance of ONT (R9 and R10 chemistries) in comparison to Illumina and MGI across 17 well-characterized reference samples with 11 clinical variants representing nine different genetic diseases. To enable this, we have implemented a production-ready pipeline including SNV, indel, STR, SV, and CNV detection, alongside reporting key summary metrics to ensure high-quality data at the production sequencing level. Our results show high accuracy of ONT across SNVs (*F*-score 0.978–0.983) and SVs (*F*-score = 0.75) but still weaknesses across indels (*F*-score 0.659–0.758). However, we highlight that ONT accurately detected all four pathogenic indels as well as the performance improvement in exons and with the newer R10 chemistry. We further demonstrated the importance of long reads to detect clinically impactful variants such as a *FMR1* pathogenic expansion, often misclassified by SRS as being in the premutation range. Our multiplatform analysis and Sanger validation uncovered a 1 bp error in the Coriell annotation for a cystic fibrosis–causing indel in GM07829. This work underscores the growing readiness of ONT for clinical applications, highlighting both its advancements and its potential for broader adoption in clinical genomics and large-scale operations.

[Supplemental material is available for this article.]

The advent of long-read sequencing (LRS) enabled increasing read size from hundreds of base pairs (short reads) to multiple thousands or even millions of bases in one continuous read. These longer reads can resolve repetitive regions and improve the identification of structural variations (SVs). This has enabled a more comprehensive insight into the diversity of the human genome, such as tandem repeats, centromeres, and telomeres, which play critical roles in chromosome stability and aging (Nurk et al. 2022). LRS also provides haplotype phasing, crucial for understanding inheritance patterns (Logsdon et al. 2020).

Over the past decade, long reads have driven discoveries in evolution, population diversity, and medical research. In contrast, short-read sequencing (SRS) produces fragmented assemblies (e.g., SRS contig N50 ~ 50–100 kb vs. >1 Mb for long-read assemblies), underrepresents >70% of short tandem repeats (STRs), and fails to detect at least half of SVs >50 bp (De Coster et al. 2019; Mahmoud et al. 2019). Nevertheless, SRS remains the workhorse of genomics, producing many of the whole-genome and whole-exome data sets to study rare and complex diseases. Although SRS can readily identify single-nucleotide variants (SNVs) associated with disease, it often fails to resolve pathogenic alleles within homologous regions, SVs, or STRs. In many cases, the reported SNV is merely in linkage disequilibrium with a more complex causative variant—such as a copy number variant (CNV) or small insertion or deletion (indel) in a paralogous gene or deep intronic rearrangement—that short reads cannot resolve reliably (Flister et al. 2013). This might be overcome by improvements to LRS, which are expected to enhance indel accuracy, reduce costs, and streamline library preparation (Ni et al. 2023; Mahmoud et al. 2024).

The two main LRS technologies are Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), both advancing toward clinical applications (for review, see Oehler et al. 2023).

[5]These authors contributed equally to this work.
Corresponding author: jquilez@m42.ae
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.280134.124.

36:460–471 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/26; www.genome.org

With the growing adoption of LRS, several population-scale initiatives are under way (De Coster et al. 2021), and programs such as All of Us and Genomics England have implemented ONT-based whole-genome sequencing (WGS) for clinical genomics. This progress is partly motivated by LRS's ability to detect causative variants missed by short reads, such as pathogenic deletions in the *DMD* gene leading to Duchenne muscular dystrophy (DMD) (Geng et al. 2023) or repeat expansions in high GC content regions underlying disorders like fragile X syndrome (FXS) (Stevanovski et al. 2022). ONT also enables variant phasing for inheritance and de novo mutation analyses (Cretu Stancu et al. 2017) and can distinguish clinically relevant genes from pseudogenes, improving diagnostic accuracy (Leija-Salazar et al. 2019). Together, these developments demonstrate the increasing applicability of LRS technologies in clinical genomics.

ONT has evolved from early versions with >20% error rates to a robust production platform (Ni et al. 2023; Sanderson et al. 2024), supported by continuous improvements in chemistry and basecalling that enhance accuracy, read length, and methylation detection. These advances enable robust identification of causative alleles in complex or repetitive regions, including pseudogenes and other challenging medically relevant genes (CMRGs). Furthermore, the speed at which ONT can operate makes it an encouraging instrument for rapid clinical implementation (Gorzynski et al. 2024; Smolka et al. 2024). Despite recent advances, several factors still limit the adoption of ONT sequencing in genomics and clinical research. Homopolymer regions continue to increase error rates, particularly in LRS, although machine learning now mitigates these effects (Wick et al. 2019). ONT pipelines also require tailored optimizations and dedicated variant callers remain under active development (see Discussion) (Supplemental Note S2). Broader adoption will require rigorous benchmarking in clinically relevant regions, because most existing data sets, such as Genome in a Bottle (GIAB) (Zook et al. 2016; Krusche et al. 2019), focus mainly on SNVs and small indels, leaving performance on repeats and SVs largely untested.

In this work, we assess the readiness of ONT for clinical and population-scale genomics using an optimized analysis pipeline (https://github.com/juditharres/Nanopore4Clinics) for comprehensive variant detection. This pipeline balances comprehensiveness and efficiency, ensuring fast turnaround times for both clinical cases and large-scale population studies. To establish ONT's performance and validate our workflow, we sequenced 17 well-characterized Coriell reference samples using R9 and R10 chemistries and compared them with Illumina and MGI short-read platforms. Unlike GIAB and other benchmark data sets, our analysis evaluates not only genome-wide variant detection but also the performance on clinically pathogenic alleles, an essential step toward clinical validation. The Coriell samples were specifically selected to represent clinically relevant mutations, a scope not previously examined at this scale. In addition, we consolidated the ONT WGS analysis pipeline to detect key variant types (SNV, indel, SV, CNV, and STR) while integrating dedicated quality-control (QC) steps. Together, these efforts provide an unprecedented benchmark of ONT WGS performance relative to short-read technologies, offering a robust framework for its evaluation in clinical genomics.

Results

Performant analysis pipeline for ONT WGS data

ONT remains less mature than other sequencing platforms in the availability of standardized end-to-end analysis pipelines that integrate clinically relevant tools. In contrast, DRAGEN is the leading analysis platform for Illumina data, with version 4.2 onward including gene-specific callers relevant for pharmacogenomics (*CYP2D6* and *CYP2B6* in addition to the generic pharmacogenomics caller), RH blood type, and disease (*SMN1* for SMA, GBA for Gaucher's disease, *HB1/HB2* for alpha-thalassemia, *CYP21A2* for congenital adrenal hyperplasia, and *LPA* for cardiac disease risk) (Behera et al. 2025). Therefore, a scalable and accurate workflow must integrate variant-calling methodologies along with upstream basecalling, alignment steps, and quality metrics to ensure high-quality results (Supplemental Note S2).

We have consolidated a comprehensive ONT WGS pipeline (Supplemental Fig. S1) that integrates improvements from our primary and secondary analysis workflows, along with additional callers for CNV and STR detection. We implemented concurrent real-time basecalling using a NVIDIA A100 tower connected to a PromethION 48 ONT sequencing instrument and the latest Dorado base-caller (https://github.com/nanoporetech/dorado), extracting both DNA sequence and 5mC methylation signals during basecalling. We then perform, in the cloud (a high-performance computing platform), mapping to the reference genome sequence using Sentieon-accelerated minimap2 (https://www.sentieon.com/), followed by variant calling with Clair3 for SNVs and indels (Zheng et al. 2022), Sniffles2 for SVs (Smolka et al. 2024), Spectre for CNVs (https://github.com/fritzsedlazeck/Spectre), and Straglr for STRs (see Methods; Supplemental Note S2; Chiu et al. 2021).

A high-quality diverse set of 17 WGS reference samples sequenced on multiple HTS technologies

We analyzed 17 Coriell reference samples to evaluate the performance of high-throughput sequencing (HTS) technologies in accurately detecting genetic variants. The samples were ordered from the Coriell Institute for Medical Research, which provides well-characterized reference materials with publicly available genetic truth sets. Cell lines were used instead of extracted DNA to minimize fragmentation and ensure high-quality input for long-read ONT sequencing (see Supplemental Note S1).

We selected two types of Coriell reference samples. The first comprised a well-characterized parent–offspring trio (GM24143, GM24149, and GM24385) that has been extensively whole-genome-sequenced by the community, with publicly available truth sets for SNVs and indels across the genome (Supplemental Table S2; Zook et al. 2019). The second type consisted of 14 samples from individuals affected by specific diseases, for which the pathogenic genetic variants are well documented (Supplemental Table S2). In each case, the disease-causing variant was expected to be present only in the corresponding sample. This design allowed us to evaluate the ability of each sequencing platform to detect not only small variants (SNVs and indels) but also larger forms of variation such as CNVs and STRs.

We performed ~30× WGS for each of the 17 Coriell cell lines using long-read ONT (R9 and R10 chemistries) and short-read technologies (Illumina and MGI). The R9 chemistry, used since 2016, was discontinued in 2024 and replaced by R10, which provides improved accuracy and read length (Ni et al. 2023). This design resulted in 68 unique ID samples across platforms (17 samples × 4 platforms). To further assess intra- and interrun variability, we sequenced two samples (GM27631 and GM03620) two additional times on ONT, resulting in triplets: GM27631a, GM27631b, GM27631c and GM03620a, GM03620b, GM03620c. Thus,

although Illumina and MGI each produced 17 runs (one per sample), the additional replicates on the ONT platforms increased their total to 21 runs corresponding to 17 unique IDs. Altogether, this yields a total of 76 runs/samples: 17 (Illumina) + 17 (MGI) + 21 (ONT R9) + 21 (ONT R10).

We evaluated the quality of all WGS data set and confirmed that all 76 samples met high-quality standards (Supplemental Tables S1, S3). These thresholds were defined based on the UK Biobank (Sudlow et al. 2015), The 1000 Genomes Project (1000 Genomes Project Consortium et al. 2015), and GATK Best Practices (DePristo et al. 2011). Overall, we generated >100 Gb of WGS data per sample across all three platforms (Supplemental Fig. S2; Supplemental Table S3).

The average genome coverage in ONT R9 samples (~45×) was approximately one-third higher than that in R10 (~33×) (Supplemental Fig. S3A), which reflects the known lower throughput of R10 flow cells (Ni et al. 2023) evident in our data (median yield: 103 Gb for R10 vs. 141 Gb for R9). Although ONT achieved higher mean coverage (~40×) than SRS (~35×), the proportion of bases with >10× coverage was slightly lower (92.0–94.0% for ONT vs. 95.8–99.3% for SRS), reflecting greater local variability. This variability stems from ONT reads extending into repetitive, subtelomeric, and pericentromeric regions where SRS often shows no coverage. As a result, ONT yields low (but nonzero) coverage in these areas, whereas SRS exhibits complete dropouts, inflating apparent dispersion when averaged genome-wide. Regions showing the largest platform differences were enriched near centromeres and telomeres (Supplemental Fig. S3D).

To assess whether this variability affects functionally important regions, we analyzed the well-characterized GM24385 sample and quantified the proportion of bases with >10× coverage across medically relevant genes (MRGs), protein-coding regions, and nonrepetitive regions. Across all subsets, ONT coverage was uniform and comparable to SRS (Supplemental Table S4). In MRG, ONT R9 and R10 achieved 99.8% and 99.2% of bases covered at >10×, closely matching Illumina (99.7%) and MGI (99.3%). Similar uniformity was observed genome-wide, confirming that ONT's variability arises mainly from repetitive regions, without affecting coverage in clinically relevant loci.

We also compared yield and read quality across platforms. Illumina and ONT R9 showed the smallest yield variation (93–131 Gb and 122–163 Gb, respectively), whereas MGI (103–171 Gb) and ONT R10 (72–198 Gb) were more variable. In Supplemental Table S3, we can also see how read accuracy is still much lower in ONT compared with SRS, as previously reported (Amarasinghe et al. 2020). Although we observed close to 90% of SRS sequencing bases with quality scores >Q30, a similar proportion of ONT sequencing bases only achieved >Q10. We could detect the improvement in read accuracy in R10 relative to R9 (median: ~86% and ~80% bases >Q10, respectively). This improvement is consistent at increasing quality thresholds as well as when comparing average values (Supplemental Fig. S4). In addition to increased read accuracy in R10, we also detected larger read lengths in that chemistry compared with R9 (Supplemental Fig. S5). Although we found >95% of mapping rate across all platforms, both ONT chemistries achieved the highest (>99%). Given these very high mapping rates across samples, yield translated into the proportionally expected effective mapping coverage values.

In summary, compared with SRS, ONT (R9 and R10) shows (1) lower read accuracy, (2) slightly higher variation in coverage driven by repetitive and pericentromeric regions, and (3) higher mapping rate; between the two ONT chemistries, R10 shows (1) lower yield, (2) improved read quality, and (3) longer reads.

In terms of the number of variants called per sample, we observed platform-specific differences (see Supplemental Fig. S6). Both SRS technologies resulted in about 5 million variants per Coriell sample (about 4 million and 1 million SNVs and indels, respectively). Both ONT chemistries resulted in a higher number of variants (R9: total = 5.5 million, SNV = 4.5 million, indel = 1 million; R10: total = 5.7 million, SNV = 4.5 million, indel = 1 million). These differences reflect intrinsic platform characteristics and the higher sensitivity of LRS, which can also increase false-positive (FP) rates and thus require stricter postcalling filtering.

We next performed principal component analysis (PCA) to the SNV/indel call sets to assess data consistency across samples and detect potential outliers or batch effects. The first two principal components (PC1–PC2, 20.58% variance) reflected the ancestry structure of the 17 Coriell reference samples. As shown in Supplemental Figure S7A, the single African American sample and the South American family trio (left and bottom data points, respectively) separated from the remaining European-descent samples. Each individual was represented by four data points, corresponding to sequencing performed with ONT (R9 and R10), Illumina, and MGI, which clustered closely together (Supplemental Fig. S7B). PC1 and PC2 also grouped family members, whereas PC4 distinguished short- and long-read data sets (6.47% variance), with minimal separation between ONT R9 and R10 (Supplemental Fig. S8). Although this shows the platform's influence on genotype calls, it remains minor and has no impact on the subsequent analyses. Additionally, no outliers were observed in any of the PCs, confirming the absence of problematic samples and the robustness of the data set.

Performance of SNV, indel, and SV detection across the genome

We used the parent-offspring trio (GM24143, GM24149, and GM24385) to assess SNV/indel variant-calling performance genome-wide (Supplemental Table S2). For each ONT (R9 and R10), Illumina, and MGI sample, we compared variant call sets to publicly available golden-truth variants to calculate recall, precision, and their harmonic mean or *F*-score (see Methods). These performance metrics were evaluated genome-wide, in exonic regions and MRGs. Across all comparisons, values were highly consistent within each sequencing platform, as shown by the narrow interquartile ranges in Figure 1, indicating robust and reproducible performance.

In all three sequencing platforms, we achieved high genome-wide performance for SNV calling, with *F*-scores of 0.975–0.983 across all samples (Fig. 1; Table 1A; Supplemental Table S5). As expected, SNV precision (median = 0.997) exceeded recall (median = 0.963) regardless of the sequencing platform. Although Illumina exhibited slightly higher SNV precision, ONT (R9 and R10) achieved better recall, especially compared with MGI. Regardless, these values confirm robust and consistent SNV detection across platforms.

Accurately detecting indels is more challenging compared with SNVs (i.e., small indels <50 bp compared with single-point sequence changes). As expected, indel calling showed lower precision and worse recall across all platforms (Fig. 1; Table 1A; Supplemental Table S5). We found that the drop in the indel performance is very pronounced in ONT, with an *F*-score of 0.659–0.758, precision of 0.745–0.832, and recall of 0.592–0.696 in all samples. In contrast, the same metrics exceeded >0.85 in the SRS

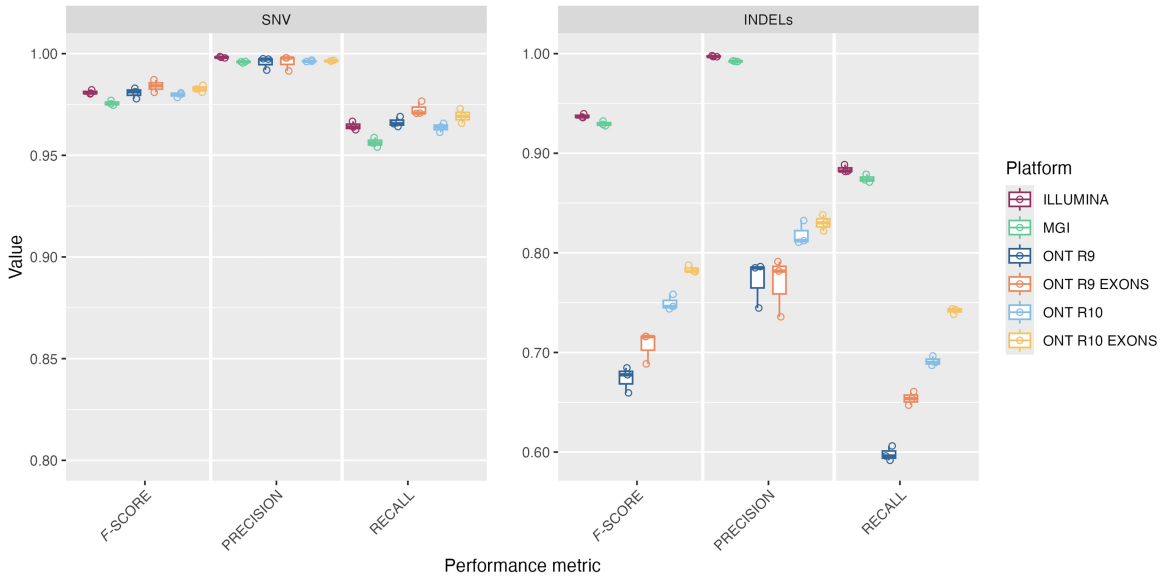


Figure 1. SNV- and indel-calling performance across sequencing platforms. *F*-score, precision, and recall distribution for SNVs (*left*) and indels (*right*) across the different sequencing platforms. For Illumina, MGI, and ONT (both R9 and R10), displayed are performance metrics in high-confidence regions genome-wide. In addition, for ONT R9 and R10, shown are also performance metrics in exonic regions. Each boxplot encapsulates the corresponding metrics for the three Coriell samples with truth sets available (GM24143, GM24149, and GM24385).

Table 1. Summary of genome-wide SNV/indel-calling performance and detection of positive controls across sequencing platforms

A

Sample	Description	Variant Type	Illumina	MGI	ONT (R9)	ONT (R10)
			F-score			
GM24143	Trio (Mother)	SNV	0.980	0.975	0.981	0.978
		INDEL	0.936	0.928	0.678	0.744
GM24149	Trio (Father)	SNV	0.981	0.975	0.978	0.980
		INDEL	0.936	0.929	0.659	0.758
GM24385	Trio (Child)	SNV	0.982	0.977	0.983	0.981
		INDEL	0.940	0.932	0.685	0.746

B

GM13708	BC	SNV	✓	✓	✓	✓
GM27630	MRD40	SNV	✓	✓	✓	✓
GM27631*	MRD40	SNV	✓	✓	✓	✓
GM27632*	MRD40	SNV	✓	✓	✓	✓
GM10354	USH1C	SNV	✓	✓	✓	✓
GM07828	CF	INDEL	✓	✓	✓	✓
GM07829	CF	INDEL	✓	✓	✓	✓
GM07830	CF	INDEL	✓	✓	✓	✓
GM08211	CF	INDEL	✓	✓	✓	✓
GM04099	DMD	CNV (Deletion)	✓	✓	✓	✓
GM10684	SMA1	CNV (Deletion)	✓	✓	✓	✓
GM03990	DM1	STR	✓	✓	✓	✓
GM09145	FMR1	STR	✗	✗	✓	✓
GM03620	HD	STR	✓	✓	✓	✓

(*A*) *F*-score performance metric values for SNV and indel calls separately for each of the samples in the Coriell trio with truth sets available. The gray scale shading indicates the *F*-score values for each sample and variant type (SNV and INDEL) across sequencing platforms, with darker shading corresponding to higher *F*-scores and lighter shading to lower *F*-scores. (*B*) Detection (check mark) or not (x) of the expected genotype for each of the 14 positive control Coriell reference samples.
[a]For GM27631 and GM27632, detection of the precise genotype is the absence of the MRD40-causing de novo present in their child (GM27630). (BC) Breast cancer; (MRD40) mental retardation autosomal-dominant 40; (USH1C) Usher syndrome type IC; (CF) cystic fibrosis; (DMD) muscular dystrophy, Duchenne type; (SMA1) spinal muscular atrophy I; (DM1) myotonic dystrophy type 1; (FMR1) fragile X syndrome; and (HD) Huntington disease.

samples and were consistently higher in Illumina compared with MGI. Although ONT performed worse than Illumina and MGI, we observed a substantial improvement in indels detection with the newer chemistry compared with the previous R9.

Clinical genomic applications primarily target genetic variants with functional impact, typically restricting analyses to coding regions, namely, exons. Variants within exons are more likely to produce interpretable functional and phenotypic effects, whereas those in noncoding regions often have uncertain consequences. We therefore assessed whether ONT SNV/indel-calling performance improved in exons relative to genome-wide results. ONT showed higher indel performance in exons compared with genome-wide (Fig. 1; Supplemental Table S5), with recall increasing by 5.6% in R9 and 5.1% in R10. R10 also outperformed R9 in exons in terms of indel recall. This improvement is particularly important, as it reduces the proportion of missed true variants or false negatives (FNs), more frequent at the genome-wide level. Restricting our analysis to exons did not alter indel precision in R9, but it slightly increased that metric in R10, resulting in a lower FP rate. Nevertheless, ONT continued to miss some intergenic indels genome-wide. In contrast, restricting analyses to exons had minimal effect on the high SNV performance reported above (Fig. 1; Supplemental Table S5), with marginal recall gains and consistently high precision across both R9 and R10 data sets.

In the previous analysis, we focused on exonic regions, as these are most commonly examined in clinical genomics. Similarly, prior benchmarks have concentrated on medically important genes, such as the 73 actionable genes defined by the American College of Medical Genetics and Genomics (ACMG) (Mandelker et al. 2016; Miller et al. 2021; Mahmoud et al. 2024). Others have particularly focused on MRGs that are difficult or even impossible to characterize by SRS to highlight the potential of LRS to resolve these loci (Mandelker et al. 2016; Miller et al. 2021; Mahmoud et al. 2024). Thus, we evaluated all three sequencing technologies in a larger set of approximately 5000 MRGs (Wagner et al. 2022). SNV calling remained highly consistent across platforms (Supplemental Fig. S9, top left), underscoring their reliability for single-base variant identification. In contrast, ONT showed notable gains in both recall and precision relative to its genome-wide performance, particularly for indels (Supplemental Fig. S9, top right). Specifically, R9 and R10 each demonstrated an increase of ∼3% in *F*-score compared with their genome-wide benchmarks, with R10 exhibiting the greater improvement owing to its higher baseline performance. Illumina and MGI also performed robustly in this gene set, although their gains were more modest, with *F*-score increases of ∼1%. These findings underscore the impact of ongoing ONT refinements on variant detection in clinically relevant regions.

To better understand the source of ONT-specific indel errors, we examined whether homopolymer-associated artifacts contributed to the performance gap. Excluding homopolymer regions, both genome-wide and within the 5000 MRG set, led to mean *F*-score improvements of 24.9% and 18.9% for R9 and R10 genome-wide and 27.4% for R9 and 17.9% for R10 within MRGs (Supplemental Fig. S10). In contrast, Illumina and MGI exhibited minor variation, maintaining *F*-scores above 0.90 regardless of homopolymer inclusion. These results suggest that a substantial fraction of ONT-specific indel errors are attributable to homopolymer-associated artifacts, consistent with previous reports of signal compression effects (Lang et al. 2020; Delahaye and Nicolas 2021), and demonstrate that excluding such regions can improve variant detection in clinical contexts.

Building on this, we examined GM24385 across 273 clinically CMRGs with established truth sets to evaluate performance in complex loci. Consistent with the genome- and exome-wide results (Fig. 1), ONT (R9 and R10) performed comparably to Illumina and outperformed MGI when calling SNVs (Supplemental Fig. S9). Although ONT showed slightly lower precision than the two SRS platforms, recall was higher, even relative to Illumina. As previously reported (Wagner et al. 2022; Mahmoud et al. 2024), detecting indels within CMRGs is particularly challenging for SRS, as denoted by the decrease of 0.07 in precision relative to the genome-wide performance. In contrast, the impact in ONT is lower, with only a drop of 0.03 in precision. Again, the improvement of R10 relative to R9, especially in recall, is also noticeable in CMRG.

We also used the well-characterized GM24385 sample to compare SVs between LRS and SRS, for which SV truth sets exist. ONT achieved high average precision genome-wide (0.910 for R9 and 0.906 for R10), comparable to Illumina (0.916) and MGI (0.953) (Supplemental Table S6A). On the other side, the average recall in ONT was more than twice that of Illumina and nearly five times that of MGI (0.627, 0.644, 0.258, and 0.132 for ONT R9, ONT R10, Illumina, and MGI respectively). Consequently, the *F*-scores for ONT (0.743 and 0.753 for R9 and R10) were nearly double that of Illumina (0.403) and about three times higher than that MGI (0.231). The advantage of ONT became even more pronounced in CMRGs (Supplemental Table S6B). Precision for both LRS and SRS, along with the recall values for Illumina and MGI, exhibited consistency with genome-wide performance (Supplemental Table S6B), in which recall and *F*1-values approached 0.90, whereas precision across all platforms remained consistent with genome-wide results. Performance metrics were highly similar between ONT R9 and R10.

Performance detection of disease-causing mutations

We extended the evaluation of the performance of the three sequencing technologies to detect disease-causing mutations in 14 Coriell reference samples. These samples contained 11 mutations (three SNVs, three indels, two CNVs, and three STRs) linked to nine different diseases of high clinical relevance, for example, breast cancer, cystic fibrosis (CF), or DMD (Supplemental Table S2). In each of these samples, we assessed the presence or absence of the associated known mutation, for which genomic coordinates and changes in the genome reference sequence are publicly available.

Both SRS and LRS precisely detected the expected three disease-causing SNVs evaluated (Table 1B; Supplemental Table S7). In GM27630, mental retardation autosomal-dominant (MRD40) is caused by a de novo c.2127T>G (p.Tyr709*) pathogenic dominant mutation, which should therefore be absent in the parents (GM27631 and GM27632). By analyzing the trio samples from each sequencing platform, we confirmed, as expected, a single copy of the pathogenic variant in the proband and its absence in both parents.

We also evaluated three indels causing CF (556delA, del508, and 557delT), with delF508 present in three Coriell samples and GM07830 carrying two CF-associated mutations (Supplemental Table S2). We confidently detected the three phenotype-causing indels in all four sequencing technologies/chemistries. Our multiplatform high-quality sequencing approach detected a likely 1 bp annotation error in the GM07829 reference sample. Although the Coriell Institute for Medical Research reports a CF-causing adenine

(A) deletion at Chr 7: 117,531,049 bp, corresponding to NM_000492.3(CFTR):c.424delA (p.Ile142fs) (Zielenski et al. 1991), we detected a deletion of the immediately downstream thymine (T) in all four WGS samples for GM07829 (Chr 7: 117,531,050 bp) (Supplemental Table S7). We ruled out mapping or variant-calling artifacts and confirmed the T deletion through visual inspection in the Integrative Genomics Viewer (IGV) (Robinson et al. 2011) (Fig. 2A). Sanger sequencing further validated the T deletion at Chr 7: 117,531,050 bp (Fig. 2B), providing orthogonal confirmation of the genotypes predicted by all HTS technologies. Of note, the detected indel has the same functional impact on the CFTR protein, fully consistent with the CF phenotype of the GM07829 sample.

We accurately genotyped the CNV responsible for DMD in GM04099. Both SRS and LRS detected a heterozygous deletion of exons 49–52 in the *DMD* gene. For LRS, we assessed CNV detection with two different variant callers. The Sniffles2 v2.2 SV caller detected the expected drop in coverage corresponding to the deletion but incorrectly classified it as homozygous rather than the expected heterozygous genotype. In contrast, the updated Sniffles2 v2.5.3 correctly identified the deletion as heterozygous. Similarly, the latest version of the Spectre CNV caller implemented in our pipeline successfully identified this pathogenic deletion (Supplemental Table S7). Sequencing read alignments for GM04099 in both ONT R9 and R10 further supported the expected heterozygous genotype, with normal coverage in the flanking regions (~30× for ONT R10) and approximately half coverage within the deletion region (Supplemental Fig. S11). These findings ruled out abnormal coverage as a confounding factor in the genotype call and indicate that the discrepancies observed with Sniffles2 reflect algorithmic rather than sequencing limitations.

Furthermore, the deletion of exons 7 and 8 in the *SMN1* gene illustrates how ONT is catching up in sequencing accuracy and software developments with SRS. This variant is particularly relevant in the clinical context owing to its association with spinal muscular atrophy (SMA), an autosomal recessive neuromuscular disorder characterized by degeneration of α-motor neurons in the spinal cord, leading to progressive muscle weakness and atrophy (Wang and Lunn 2008). In most cases, SMA results from homozygous deletions in *SMN1*; however, accurate detection is complicated by the nearly identical paralog *SMN2*, which differs by a single nucleotide in exon 7 and modulates disease severity (Feldkötter et al. 2002). Illumina's DRAGEN platform includes a

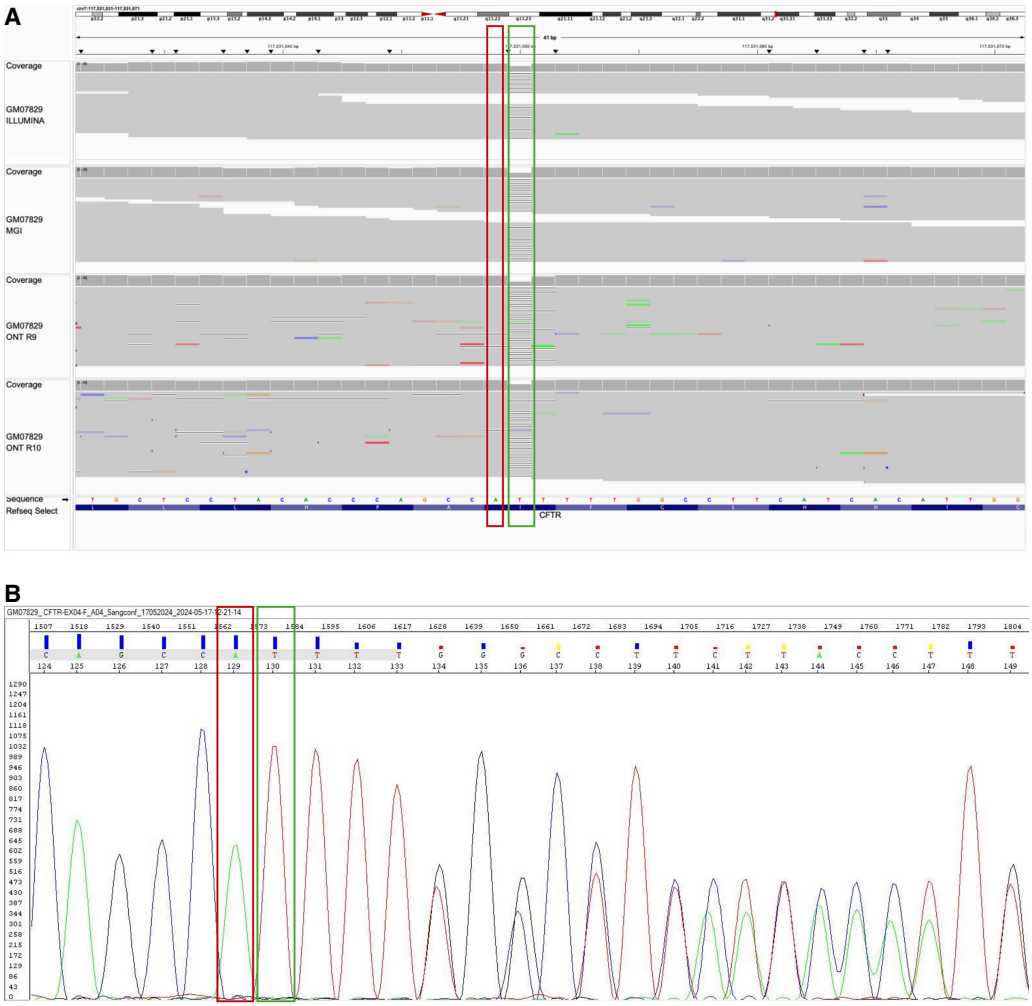


Figure 2. Across-platforms HTS and Sanger orthogonal validation uncover a 1 bp annotation error in GM07829. (*A*) IGV snapshots supporting the absence of the 556delA indel (Chr 7: 117,531,049 bp) annotated in Coriell for GM07829 but instead a 1 bp downstream T deletion (Chr 7: 117,531,050 bp) with the same detrimental functional effect. (*B*) Orthogonal validation of the HTS results through Sanger sequencing.

dedicated caller to resolve *SMN1/SMN2* variation, which we also implemented in its open-source version (Chen et al. 2020) for MGI. Additionally, we introduced an alpha version of an *SMN1* and *SMN2* caller, named *Sillago*, specifically developed by ONT (Oxford Nanopore Technologies 2024). As expected, across all sequencing technologies evaluated, we correctly genotyped the SMA GM10684 sample as homozygous for a deletion of exons 7 and 8 in *SMN1*, while identifying two normal copies of *SMN2* (Supplemental Table S7). Notably, *Sillago* provided concordant results for this sample on both R9 and R10 flow cells, suggesting promising performance in resolving this complex locus.

We next assessed trinucleotide repeat expansions leading to disease phenotypes (Supplemental Table S2) using dedicated STR callers implemented in our analysis pipeline. Expansion Hunter (Dolzhenko et al. 2017) was employed for Illumina and MGI, whereas Straglr (Chiu et al. 2021) was applied for ONT data. Our data set included a sample affected by myotonic dystrophy type 1 (DM1), an autosomal-dominant disorder marked by muscle weakness, myotonia, cataracts, and cardiac and endocrine involvement, typically appearing in the 20 sec to 40 sec (Brook et al. 1992; Udd and Krahe 2012). DM1 is caused by expansion of more than 50 CTG repeats in the 3′ UTR of the *DMPK* gene (Brook et al. 1992). Specifically, the GM03990 sample was reported to have between 50 and 80 CTG repeats. In our analysis, we observed 64, 57, 78, and 79 copies using Illumina, MGI, ONT R9, and ONT R10, respectively (see Fig. 3). Similarly, Huntington disease (HD) is an autosomal-dominant, progressive neurodegenerative disorder characterized by chorea, cognitive decline, and psychiatric symptoms, with onset typically between 30 and 50 years of age (Walker 2007). HD results from an expansion of more than 36 CAG repeats in exon 1 of the *HTT* gene. The GM03620 sample in our study is expected to present 60 copies of such repeat. Illumina detected 55 repeats and MGI identified 68, whereas ONT R9 and R10 found 58 and 60 copies (Fig. 3). Altogether, for both GM03990 and GM03620, in all sequencing technologies we predicted the STR copy number in the pathogenic range for DM1 and HD, respectively (Supplemental Table S7).

The third STR evaluated was a CGG expansion in the *FMR1* gene that causes FXS when exceeding 200 copies (Supplemental Table S2; Hagerman et al. 2017). SRS failed to precisely call the pathogenic expansion in the GM09145 sample. Illumina and MGI detected 176 and 111 copies of the CGG repeat, respectively (Fig. 3; Supplemental Table S7), both below the greater-than-200 threshold defining the disease state. This highlights the inherent limitations of SRS in accurately detecting STR, mainly owing to average read length and the extensive size of these specific repeat sequences. In contrast, ONT detected more than 200 copies of this CGG repeat in GM09145 using both R9 and R10 chemistries (Fig. 3), although considerable variation was observed between them (864 and 573 copies, respectively) (Fig. 3).

Because the FXS-linked CGG repeat is relatively long, we hypothesized it could also be detected by Sniffles2 (Smolka et al. 2024), the SV caller in our ONT pipeline. Indeed, Sniffles2 identified the CGG repeat in *FMR1* as an CGG-based insertion, consistent with Straglr estimates, detecting 842 repeats for ONT R9 and 772 for ONT R10 (Fig. 3). This approach using different callers for variant detection, highlighted the robustness of LRS in identifying variants confidently.

In summary, we correctly identified the expected genotype in 13 out of the 14 (92.86%) Coriell positive control samples across all four sequencing platforms, corresponding to 54 out of all the 56 genotypes (96.43%) overall. Moreover, each of the 11 disease-causing mutations we evaluated (Supplemental Table S2) was detected exclusively in its corresponding positive control and absent from all others, yielding a genotyping accuracy near 100% (Supplemental Table S7).

Finally, for ONT only we tested the consistency within (intra) and between (inter) sequencing runs using two of the Coriell positive controls: GM27631 (c.2127T>G in MRD40) and GM03620 (CAG repeat in *HTT*. Intrarun tests assess the stability of sequencing results within a single batch, detecting potential variability during the process. Interrun tests, on the other hand, extend this quality assessment across different runs, reagents, or instruments to ensure reproducibility under varying conditions. In the intrarun tests, we sequenced each of the two samples twice within the same run. The c.2127T>G mutation in GM27631 was successfully detected in both intrarun replicates for both R9 and R10. Similarly, the predicted CAG repeat numbers in GM03620 were consistent

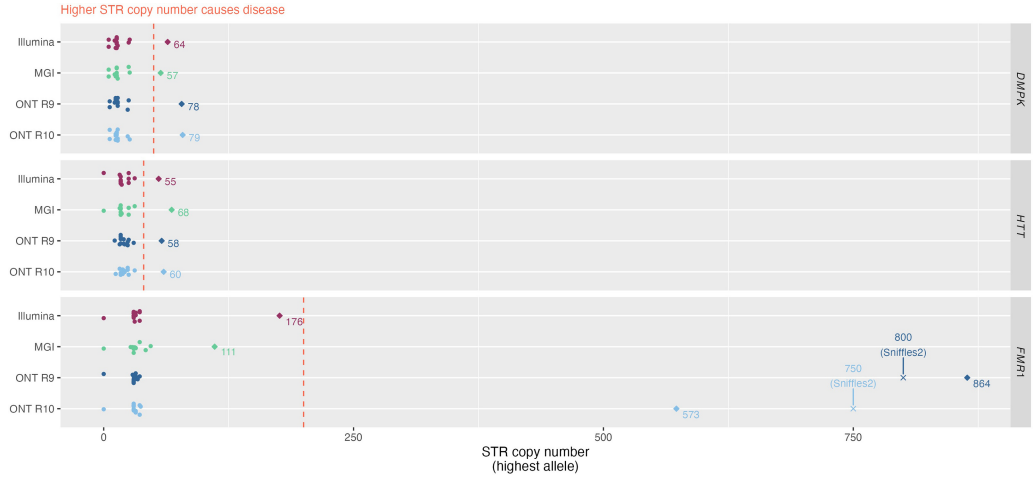


Figure 3. Detection of STR expansions across sequencing platforms for genes *FMR1*, *HTT*, and *DMPK* in the 14 Coriell samples. The *top* panel shows the detection of CTG repeats in the *DMPK* gene in which expansions greater than 50 repeats are associated with myotonic dystrophy type 1. The *middle* panel displays the detection of CAG repeats in the *HTT* gene, with expansions beyond 36 repeats, characteristic of Huntington's disease. And finally, the *bottom* panel illustrates the detection of CGG repeats in the *FMR1* gene, in which expansions exceeding 200 repeats are indicative of fragile X syndrome.

between the intrarun replicates for R9 (58 and 58 repeats) and R10 (59 and 60 repeats). For the interrun analysis, we sequenced the same two samples in two separate sequencing runs. The c.2127T > G mutation was detected in all interrun replicates, and the STR repeat numbers were again consistent across runs (R9: 58 and 58; R10: 59 and 60). Altogether, our results demonstrated satisfactory intra- and interrun replicability of ONT sequencing.

Across-platforms concordance

Beyond comparisons to reference truth sets, we also examined systematic differences in variant calling between sequencing platforms. We noticed that for ONT, both R9 and R10, an excess of SNPs/indels around the centromeres is called compared with other parts of the chromosome (Supplemental Fig. S12). That is also visible for SRS to a lesser extent and only in certain chromosomes. ONT also shows a higher proportion of variants in the region of Chromosome 6 that harbors the human leukocyte antigen (HLA). These trends likely reflect inaccuracies in the GRCh38 reference genome within repetitive and structurally complex regions, for which read misalignment is frequent. Although SRS typically filters these reads using stringent mapping quality thresholds, long-read technologies retain more of them, potentially contributing to the observed increase in variant calls.

We then assessed genotype concordance across sequencing platforms by calculating the Jaccard index between variant call sets derived from the same Coriell sample, separately for SNVs and indels (Methods) (Supplemental Fig. S13). In SNVs, we detected 78.5%–92.0% genotype sharing between any platform and chemistry compared (median = 82.4%); we evaluated about 4 million sites genome-wide. The two SRS technologies showed the highest concordance (90.1%) followed by 87.2% between ONT chemistries (Supplemental Fig. S12; Supplemental Table S8). The overall concordance between LRS and SRS was the lowest, ranging from 79.2% to 83.0%. Similarity, the concordance between ONT and Illumina was higher compared with between ONT and MGI. Unexpected similarity values between the newer ONT chemistry R10 and SRS were lower compared with R9 and SRS. As expected, similarity patterns in indels were lower across all platform comparisons (median: 59.9%; range: 50.8%–82.6%). Intra-SRS comparisons still showed the highest similarity (81.2%) but were ~10% lower compared with SNVs (90.6%). Any ONT-based comparison fell below 70% similarity, even the R9 versus R10, which for SNVs was >86% in any sample. Contrary to the SNV results, in which R9 demonstrated better concordance with SRS than R10, the R10–SRS comparison for indels exhibited higher similarity (57.8%–61.7%) relative to the R9–SRS comparison (55.2% in both cases). Altogether, these results reinforce that accurate calling of indels is more challenging than for SNVs, especially for ONT.

Discussion

In this study, we evaluated and contributed to the readiness of ONT for clinical genomics and population studies, addressing key questions about its performance and clinical applicability.

Improvements in this LRS technology over recent years and its advantages over SRS have resulted in increasing clinical and research work leveraging ONT (Wick et al. 2019; Oehler et al. 2023; Mahmoud et al. 2024). Despite these advancements, some key questions remain and hinder the broader and faster adoption of ONT: Is the lower read accuracy of ONT relative to SRS still translating into less trustworthy variant calls? Is the analysis of ONT production-ready for fast turnaround times and large volumes of samples, and it is capable of targeting important genetic variants?

Our findings show that ONT's performance is rapidly catching up to SRS in detecting SNVs, and it outperforms SRS for large and repetitive variants, including SVs and STRs. Although indel detection remains the main weakness inherent to the Nanopore sequencing technology (Amarasinghe et al. 2020), the new R10 chemistry marks a significant improvement, particularly in loci that are medically relevant and traditionally challenging for SRS. This improvement compounds with ONT's enhanced exon-level performance, in which variants are more likely to translate into phenotype changes. Notably, ONT was able to accurately detect all four pathogenic indels across 14 samples, which underscores its potential for identifying critical disease-causing variants. This improvement suggests that ONT's indel-calling capability, although still behind SRS, may become a viable option for clinical applications as the technology evolves.

To further facilitate the integration of ONT into clinical and population genomics workflows, we implemented a robust analysis pipeline that highlights the improvements of the ONT toward a scalable end-to-end sequencing solution. Upgrading from NVIDIA V100 to A100 GPUs and from Guppy to Dorado enabled real-time basecalling, streamlining our workflow and reducing storage needs. Our pipeline operates directly on unaligned BAM (uBAM) files, eliminating the need to retain large raw files (~700 Gb per ~30× human WGS). Continuous improvements of the basecalling models and incorporating error-correction tools (Salmela and Rivals 2014; Wang et al. 2018) may not only improve computational efficiency but also compensate for sequencing errors of ONT that are not addressed through upgrades of their sequencing technology. We achieved shorter pipeline runtimes by choosing speed-optimized software for ONT, namely, Sentieon's minimap2 (Li 2018), about twice as fast as the open-source version, which is included in ONT's human variation workflow (https://github.com/epi2me-labs/wf-human-variation). ONT has also benefited by the active development of variant callers specific to their technology and targeting key genetic variant types: from small variants (Clair3; used in our pipeline) (Zheng et al. 2022), DeepVariant (Poplin et al. 2018), and Medaka (https://github.com/nanoporetech/ont-spectre), to larger and more complex forms such as STRs (Straglr) (Chiu et al. 2021), CNVs (Spectre, https://github.com/nanoporetech/ont-spectre), and SVs (Sniffles2) (Smolka et al. 2024), including challenging medically relevant loci like the *SMN1/2* deletion (Sillago) (Oxford Nanopore Technologies 2024). We include most of those tools in our analysis to produce a comprehensive catalog of genetic variation with a single run.

Indeed, one important milestone missing in genomics and genetics is the simultaneous detection of all variant types. Current genetic studies often focus on either SNVs/indels or other forms of variation, such as STRs, separately. However, these variants coexist on the same DNA molecules and jointly contribute to phenotypes. For instance, although SVs are less frequent than SNVs and indels (approximately 23,000 SVs vs. 4 million–5 million SNVs/indels), they typically impact a larger number of nucleotides across the genome. We primarily focus on SNVs because their annotation and impact prediction are better established, partly owing to the detection biases and analytical challenges associated with more complex variants. Improving SV and STR detection at the population scale is essential to develop better functional models of genome variation, which will require large-scale LRS efforts now underway in several consortia. ONT's ability to efficiently capture these complex variants at scale, as

demonstrated by our pipeline, highlights its value in comprehensive genomic studies.

Despite these advancements, ONT sequencing still faces challenges, particularly in indel detection. Our study confirmed that although ONT shows comparable accuracy to Illumina and MGI in detecting SNVs ($F1 = 0.978$ for ONT vs. 0.980 for Illumina and 0.975 for MGI), its indel accuracy lags behind. This was also reported previously and seems to improve over the different generations of base callers and variant callers (Wenger et al. 2019; Logsdon et al. 2020). We demonstrated that improvements in the R10 chemistry resulted in 5%–7% increase in accuracy and that exonic regions, easier to interpret clinically, are less prone to repeat-induced errors. Still, SRSs show a higher accuracy for indels compared with ONT both genome- and exon-wide. Nonetheless, ONT correctly identified all four pathogenic indels across 14 samples, demonstrating reliable detection of disease-causing variants even in difficult regions.

Furthermore, ONT's strengths in detecting SVs and STRs should not be underestimated. Our findings revealed that Illumina and MGI struggle to fully characterize repeat expansions, as seen in our example of the GM09145 sample with a pathogenic *FMR1* repeat expansion (more than 200 copies) associated with FXS. Both Illumina and MGI misclassified this expansion as a premutation range (50–200 repeats), whereas ONT accurately identified it as pathogenic, showcasing its ability to handle challenging genomic regions that are often problematic for SRS technologies. This highlights ONT's potential to enhance diagnostic accuracy for repeat expansion disorders.

Although our study presents promising results, there are some limitations to consider. First, we used cell lines rather than blood or saliva, which may not fully reflect the challenges of working with clinical samples. LRS requires high-molecular-weight DNA, often more difficult to extract from blood or saliva than from cultured cells. However, our goal was to compare the three sequencing technologies irrespective of factors upstream of the sequencing process. Second, we could have included a larger number of reference samples, especially for phenotypes caused by indel given the lower performance of ONT compared to SRS. Moreover, all the indel we evaluated are within the same gene (*CFTR*). We argue we allocated our resources to achieve a benchmark including multiple sequencing platforms and both R9 and R10 as opposed to exclusively evaluating ONT on a larger cohort of samples. Notably, our data set of 17 well-characterized reference samples, including genome-wide and pathogenic variants truth sets, sequenced across two SRS (Illumina and MGI) and two chemistries of the ONT LRS has no precedent to our knowledge. We highlight that we evaluated both ONT chemistries, even though R10 is now the default chemistry and R9 is deprecated, which showed that both performed similarly and successfully. More than 100,000 human genomes have been sequenced using R9, and key control data sets, such as those used in The 1000 Genomes Project for population allele frequency annotations, rely on R9 data.

We agree that laboratories intending to use ONT for clinical applications must conduct rigorous clinical validations of their end-to-end workflow with a larger number of samples and diverse types of genetic variants, as is required for any molecular test in a clinical setting. Beyond analytical performance, cost will also influence adoption. The easiest component to estimate is the sequencing instrument and reagent cost, although these vary globally and with new releases. Harder to estimate upfront and compare across platforms is the computing cost as it depends on several factors, for example, using commercial solutions (Illumina's DRAGEN, Sentieon) versus open-source ones, the need to keep up the basecalling of ONT data with the sequencing depending on the required throughput, the basecalling model (ONT's SUP requires more GPU than HAC), and the inclusion of methylation calling, etc.

In summary, our work shows that ONT identifies SNVs genome-wide as accurately as SRS. Although the Achilles heel of ONT continues to be indel detection, this is less so in exons and in regions that are challenging for SRS whatsoever; plus, we showed the improvement in the new R10 ONT chemistry. ONT accurately detected all four disease-causing indel evaluated here. Indeed, ONT performed similarly well as SRS in detecting other disease-causing variants we interrogated. Altogether, our results will provide guidance for organizations aspiring to incorporate ONT into their clinical workflows. To additionally help that process, here we share practical advice for the implementation of bioinformatics pipelines for the analysis of ONT WGS data.

Methods

Analysis pipeline for each HTS platform

Illumina

The BCL file is the native output format of Illumina sequencing systems. We used the on-premise Illumina DRAGEN germline pipeline host software version 4.1.7 (DRAGEN Bio-IT Platform developed by Illumina) (Behera et al. 2025) to demultiplex and basecall BCL files into per-sample FASTQ files. DRAGEN germline pipeline accelerates the secondary analysis of NGS data. For example, the time taken to process an entire human genome variant calling at 30× takes ~20 min. DRAGEN is used to preprocess the FASTQ files (adapter trimming, quality filtering), and the resulting reads are aligned to the GRCh38 human reference genome. Further, after alignment, the identification of various genetic variations such as single-nucleotide polymorphisms, indels, CNVs, STRs, and HLAs are performed by the DRAGEN variant caller with high accuracy and speed. The alignments and genetic variant calls are stored in CRAM and VCF/gVCF format, respectively, for any tertiary analysis. Summary statistics of alignment/variant calls and various logs of tools from DRAGEN are stored.

MGI

The CAL file is the native output of MGI sequencing systems. We used the on-premise MGI Ztron Pro to demultiplex and basecall CAL files into per-sample FASTQ files. MGI's FPGA-based hardware acceleration, ZBOLT Pro, is a rack server that provides higher analysis capacity to perform bioinformatic analysis on data generated from MGI's sequencers. ZBOLT Pro was used to preprocess the FASTQ files (adapter trimming, quality filtering), and the reads are aligned to the GRCh38 human reference genome. After mapping to the host genome, the mutation detection is performed by ZTRON Pro, which yields genetic variations such as single-nucleotide polymorphisms, and indels. The alignments and genetic variant calls are stored in CRAM and VCF/gVCF format, respectively, for any tertiary analysis. A custom in-house pipeline on G42 Cloud was used for performing the analysis of CNVs (using CANVAS v1.40.0.1613) (Roller et al. 2016), STRs (using Expansion Hunter v5.0.0) (Dolzhenko et al. 2017), and HLA (using HLA-LA v1.0.3) (Dilthey et al. 2019) using the CRAM and VCF outputs from ZBOLT. Summary statistics of alignment/variant calls and various logs of tools from ZBOLT are stored.

ONT

The PromethION (P48) sequencing system generates raw signal data that are processed using an NVIDIA A100 Tensor GPU (https://www.nvidia.com/en-us/data-center/a100/) for demultiplexing and basecalling into per-sample FASTQ/uBAM files. G42 clouds host an in-house custom pipeline including ONT-recommended tools for processing the uBAM files. Multiple uBAM files are merged into a single uBAM file per sample using SAMtools v1.19 (Danecek et al. 2021). On each sample, fastp (v0.23.4) (Chen et al. 2018) is performed for initial sequencing QC and for checking if the target total number of gigabases was achieved during the sequencing. We aligned uBAM file to the GRCh38 human reference genome using Sentieon's acceleration of minimap2 (v2.22) (Li 2018), and we used Alfred (v0.2.6) (Rausch et al. 2019) to check the quality and alignment QC for each sample. The alignments generated by minimap2 were stored in CRAM format. We used the alignments to call SNV, indels, and SVs using Clair3 v1.0.4 (Zheng et al. 2022) and Sniffles2 v2.2 (Smolka et al. 2024), respectively. CNVs (using Spectre v0.2.1-alpha), STRs (using Straglr v0.2.4) Chiu et al. 2021), HLAs (using HLA-LA v1.0.3) (Dilthey et al. 2019), and survival motor neutron (*SMN1/2*) using Sillago (Oxford Nanopore Technologies 2024) were also part the in-house pipeline. We used VariantQC (Yan et al. 2019) for performing the quality checks on the variants called and reporting the statistics for each sample. We primarily kept alignments (in CRAM format) and genetic variants (VCF and/or gVCF) as well as software tools logs and summary statistics.

Assessment of genome-wide and regional coverage uniformity

Coverage metrics were calculated from the aligned CRAM files using mosdepth v0.3.10 (Pedersen and Quinlan 2018). Coverage was computed at single-base resolution (1 bp bins), excluding secondary and supplemental alignments to ensure that each read contributes only once to the depth calculation.

To further evaluate coverage uniformity across functionally relevant portions of the genome, we quantified the proportion of bases with >10× coverage within three genomic subsets, MRGs, protein-coding regions, and nonrepetitive regions, using the GM24385 reference sample. This targeted analysis allowed us to determine whether genome-wide coverage variability extended into clinically or functionally important regions.

PCA

Beyond the standard QC metrics, PCA emerges as a powerful tool for uncovering underlying patterns and discrepancies in sequencing data. This step presents an additional layer of QC by assessing consistency across samples sequenced on multiple platforms, including Illumina, MGI, ONT R9, and ONT R10. PCA is a statistical procedure that transforms a set of possibly correlated variables into a set of linearly uncorrelated variables called principal components, in which the first component captures the greatest variance, and each subsequent one explains progressively less, while remaining orthogonal to the previous. In this study, PCA was applied as an additional QC measure to assess the consistency of sequencing data across sequencing platforms using 17 Coriell samples sequenced on each. After the analysis, the two main components were visually summarized in Supplemental Figure S7.

SNV/indel variant-calling performance evaluation

Evaluating variant-calling accuracy is essential in genomic research and clinical diagnostics to ensure reliable identification of genetic variants. This involves comparing the variants identified by our caller (test set) against a reference or "golden truth" for a given set of samples. The metrics commonly used for this assessment are recall, precision, and *F*-score.

Recall (or sensitivity)

Recall measures the variant caller's ability to correctly identify variants present in the golden truth data set. It is calculated as the number of true-positive (TP) variants detected divided by the sum of TP and FN variants in the reference set.

Precision

Precision assesses the proportion of identified variants that are TPs. It is determined by dividing the number of TP variants by the total number of variants called (the sum of TPs and FPs).

F-score

F-score is a harmonized metric that combines recall and precision into a single value, providing a balanced measure of the variant caller's overall performance.

To evaluate SNV and indel calling, we analyzed samples GM24149, GM24143, and GM24385, sequenced on each of the presented HTS platforms. The resulting VCFs were compared to their truth sets, filtered according to whether the performance was to be assessed genome-wide or at the MRG level. Genome-wide and MRG variant-calling performance was evaluated using Illumina's hap.py script (https://github.com/Illumina/hap.py), whereas assessments at CMRG level employed the Real Time Genomics (RTG) tool (https://github.com/RealTimeGenomics/rtg-tools). Both tools reported FPs, TPs,and FNs.

To further investigate the influence of sequence context on indel detection, we repeated the hap.py benchmarking after excluding homopolymer regions, defined following the University of California, Santa Cruz (UCSC) Genome Browser terminology as perfect tandem repeats of a single nucleotide (period = 1) spanning four to six bases, extended by 5 bp on each side to include flanking context. This analysis enabled the quantification of homopolymer-associated artifacts, particularly in ONT data sets.

For the SV analysis, sample GM24385 was used across Illumina, MGI, ONT R9, and ONT R10. The resulting VCFs were compared against the truth set both genome-wide and at the CMRG level, using the bench function of Truvari.

Pairwise genotype concordance between platforms

We measured genotype concordance between sequencing platforms using the Jaccard index, a statistical measure for comparing the similarity and diversity of sample sets. The Jaccard index was calculated from the variant call sets in pairs of sequencing samples from the same Coriell specimen. This provided a quantitative measure of the proportion of shared variants relative to the total number of unique variants, thereby reflecting the similarity and concordance of the variant call sets across platforms.

To ensure high-confidence data, variants that passed all QC filters from the VCF were first retained. These filtered variants were then categorized into SNVs and indels to facilitate independent analysis of each type. The Jaccard index was subsequently computed for the specified pair-sets using BEDTools (Quinlan and Hall 2010).

Data access

The sequencing data generated in this study have been submitted to the database of Genotypes and Phenotypes

(dbGaP; https://dbgap.ncbi.nlm.nih.gov/home/) under acccession number phs004346. The ONT analysis pipeline is provided as Supplemental Code and is available at GitHub (https://github.com/juditharres/Nanopore4Clinics).

Competing interest statement

Several authors are affiliated with M42, which supported the sequencing and processing of the Coriell samples in this study. The authors declare no additional competing interests related to this work. F.J.S. receives research support from PacBio, Illumina, Genetech, and ONT. L.F.P. received research support from Genetech until September 2023 and travel support from ONT in 2023. J.Q. received travel support from ONT in 2024 and 2025.

Acknowledgments

This study used samples (GM04099, GM10684, GM07828, GM07829, GM07830, GM08211, GM10354, GM13708, GM27630, GM27631, GM27632, GM03620, GM03990, GM09145, GM24143, GM24149, GM24385) from the Coriell Institute for Medical Research. Funding for F.J.S. and L.F.P. was provided in part by National Institutes of Health (NIH, 1U01HG011758). We also thank the supporting teams and colleagues at M42 who contributed to this work. M42 is a global healthcare and life sciences organization headquartered in Abu Dhabi, United Arab Emirates (https://m42.ae), and leverages artificial intelligence, technology, and genomics to advance precise, preventive, and predictive care. In this study, M42 supported the sequencing and processing of the Coriell samples used for analysis.

Authors contributions: Sponsor and supervision were by A.E.-K., T.R.M., V.Z., F.J.S., and J.Q. Study design was by S.E., S.B., T.R.M., V.Z., and J.Q. Data generation was by S.B., F.A., T.C., and J.M. Data preprocessing was by S.E., A.A.Y.A.S., G.K., O.S., S.Pu., A.A.A., V.K., R.C., and H.W. Data analysis was by J.A., S.E., D.M.S., L.F.P., J.Q., and P.S. Manuscript writing was by J.A., S.E., D.M.S., S.B., F.J.S., and J.Q. Critical feedback and manuscript revision were by all authors.

References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. *Nature* **526:** 68–74. doi:10.1038/nature15393

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

Behera S, Catreux S, Rossi M, Truong S, Huang Z, Ruehle M, Visvanath A, Parnaby G, Roddey C, Onuchic V, et al. 2025. Comprehensive genome analysis and variant detection at scale using DRAGEN. *Nat Biotechnol* **43:** 1177–1191. doi:10.1038/s41587-024-02382-1

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, et al. 1992. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell* **68:** 799–808. doi:10.1016/0092-8674(92)90154-5

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34:** i884–i890. doi:10.1093/bioinformatics/bty560

Chen X, Sanchis-Juan A, French CE, Connell AJ, Delon I, Kingsbury Z, Chawla A, Halpern AL, Taft RJ, NIHR BioResource, et al. 2020. Spinal muscular atrophy diagnosis and carrier screening from genome sequencing data. *Genet Med* **22:** 945–953. doi:10.1038/s41436-020-0754-0

Chiu R, Rajan-Babu I-S, Friedman JM, Birol I. 2021. Straglr: discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biol* **22:** 224. doi:10.1186/s13059-021-02447-3

Cretu Stancu M, van Roosmalen MJ, Renkens I, Nieboer MM, Middelkamp S, de Ligt J, Pregno G, Giachino D, Mandrile G, Espejo Valle-Inclan J, et al. 2017. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun* **8:** 1326. doi:10.1038/s41467-017-01343-4

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10:** giab008. doi:10.1093/gigascience/giab008

De Coster W, De Rijk P, De Roeck A, De Pooter T, D'Hert S, Strazisar M, Sleegers K, Van Broeckhoven C. 2019. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res* **29:** 1178–1187. doi:10.1101/gr.244939.118

De Coster W, Weissensteiner MH, Sedlazeck FJ. 2021. Towards population-scale long-read sequencing. *Nat Rev Genet* **22:** 572–587. doi:10.1038/s41576-021-00367-3

Delahaye C, Nicolas J. 2021. Sequencing DNA with nanopores: troubles and biases. *PLoS One* **16:** e0257521. doi:10.1371/journal.pone.0257521

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43:** 491–498. doi:10.1038/ng.806

Dilthey AT, Mentzer AJ, Carapito R, Cutland C, Cereb N, Madhi SA, Rhie A, Koren S, Bahram S, McVean G, et al. 2019. HLA*LA-HLA typing from linearly projected graph alignments. *Bioinformatics* **35:** 4394–4396. doi:10.1093/bioinformatics/btz235

Dolzhenko E, van Vugt JJFA, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, Ajay SS, Rajan V, Lajoie BR, Johnson NH, et al. 2017. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res* **27:** 1895–1903. doi:10.1101/gr.225672.117

Feldkötter M, Schwarzer V, Wirth R, Wienker TF, Wirth B. 2002. Quantitative analyses of SMN1 and SMN2 based on real-time lightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. *Am J Hum Genet* **70:** 358–368. doi:10.1086/338627

Flister MJ, Tsaih S-W, O'Meara CC, Endres B, Hoffman MJ, Geurts AM, Dwinell MR, Lazar J, Jacob HJ, Moreno C. 2013. Identifying multiple causative genes at a single GWAS locus. *Genome Res* **23:** 1996–2002. doi:10.1101/gr.160283.113

Geng C, Zhang C, Li P, Tong Y, Zhu B, He J, Zhao Y, Yao F, Cui L-Y, Liang F, et al. 2023. Identification and characterization of two DMD pedigrees with large inversion mutations based on a long-read sequencing pipeline. *Eur J Hum Genet* **31:** 504–511. doi:10.1038/s41431-022-01190-y

Gorzynski JE, Marwaha S, Reuter CM, Jensen T, Ferrasse A, Raja A, Fernandez L, Kravets E, Carter J, Bonner D, et al. 2024. Clinical application of complete long read genome sequencing identifies a 16kb intragenic duplication in EHMT1 in a patient with suspected Kleefstra syndrome. medRxiv doi:10.1101/2024.03.28.24304304

Hagerman RJ, Berry-Kravis E, Hazlett HC, Bailey DB, Moine H, Kooy RF, Tassone F, Gantois I, Sonenberg N, Mandel JL, et al. 2017. Fragile X syndrome. *Nat Rev Dis Primers* **3:** 17065. doi:10.1038/nrdp.2017.65

Krusche P, Trigg L, Boutros PC, Mason CE, De La Vega FM, Moore BL, Gonzalez-Porta M, Eberle MA, Tezak Z, Lababidi S, et al. 2019. Best practices for benchmarking germline small-variant calls in human genomes. *Nat Biotechnol* **37:** 555–560. doi:10.1038/s41587-019-0054-x

Lang D, Zhang S, Ren P, Liang F, Sun Z, Meng G, Tan Y, Li X, Lai Q, Han L, et al. 2020. Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences sequel II system and ultralong reads of Oxford Nanopore. *GigaScience* **9:** giaa123. doi:10.1093/gigascience/giaa123

Leija-Salazar M, Sedlazeck FJ, Toffoli M, Mullin S, Mokretar K, Athanasopoulou M, Donald A, Sharma R, Hughes D, Schapira AHV, et al. 2019. Evaluation of the detection of *GBA* missense mutations and other variants using the Oxford nanopore MinION. *Mol Genet Genomic Med* **7:** e564. doi:10.1002/mgg3.564

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

Logsdon GA, Vollger MR, Eichler EE. 2020. Long-read human genome sequencing and its applications. *Nat Rev Genet* **21:** 597–614. doi:10.1038/s41576-020-0236-x

Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. 2019. Structural variant calling: the long and the short of it. *Genome Biol* **20:** 246. doi:10.1186/s13059-019-1828-7

Mahmoud M, Huang Y, Garimella K, Audano PA, Wan W, Prasad N, Handsaker RE, Hall S, Pionzio A, Schatz MC, et al. 2024. Utility of long-read sequencing for all of us. *Nat Commun* **15:** 837. doi:10.1038/s41467-024-44804-3

Mandelker D, Schmidt RJ, Ankala A, McDonald Gibson K, Bowser M, Sharma H, Duffy E, Hegde M, Santani A, Lebo M, 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

Miller DT, Lee K, Chung WK, Gordon AS, Herman GE, Klein TE, Stewart DR, Amendola LM, Adelman K, Bale SJ, et al. 2021. ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American college of medical genetics and genomics (ACMG). *Genet Med* **23:** 1381–1390. doi:10.1038/s41436-021-01172-3

Ni Y, Liu X, Simeneh ZM, Yang M, 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

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

Oehler JB, Wright H, Stark Z, Mallett AJ, Schmitz U. 2023. The application of long-read sequencing in clinical settings. *Hum Genomics* **17:** 73. doi:10.1186/s40246-023-00522-3

Oxford Nanopore Technologies. 2024. *Sillago*. Oxford Nanopore Technologies, Private GitHub Repository, Oxford, UK.

Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34:** 867–868. doi:10.1093/bioinformatics/btx699

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

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

Rausch T, Hsi-Yang Fritz M, Korbel JO, Benes V. 2019. Alfred: interactive multi-sample BAM alignment statistics, feature counting and feature annotation for long- and short-read sequencing. *Bioinformatics* **35:** 2489–2491. doi:10.1093/bioinformatics/bty1007

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

Roller E, Ivakhno S, Lee S, Royce T, Tanner S. 2016. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* **32:** 2375–2377. doi:10.1093/bioinformatics/btw163

Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* **30:** 3506–3514. doi:10.1093/bioinformatics/btu538

Sanderson ND, Hopkins KMV, Colpus M, Parker M, Lipworth S, Crook D, Stoesser N. 2024. Evaluation of the accuracy of bacterial genome reconstruction with Oxford nanopore R10.4.1 long-read-only sequencing. *Microb Genom* **10:** 001246. doi:10.1099/mgen.0.001246

Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, Kalef-Ezra E, Gandhi M, Hong K, Pehlivan D, et al. 2024. Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol* **42:** 1571–1580. doi:10.1038/s41587-023-02024-y

Stevanovski I, Chintalaphani SR, Gamaarachchi H, Ferguson JM, Pineda SS, Scriba CK, Tchan M, Fung V, Ng K, Cortese A, et al. 2022. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci Adv* **8:** eabm5386. doi:10.1126/sciadv.abm5386

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12:** e1001779. doi:10.1371/journal.pmed.1001779

Udd B, Krahe R. 2012. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. *Lancet Neurol* **11:** 891–905. doi:10.1016/S1474-4422(12)70204-1

Wagner J, Olson ND, Harris L, McDaniel J, Cheng H, Fungtammasan A, Hwang Y-C, Gupta R, Wenger AM, Rowell WJ, et al. 2022. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol* **40:** 672–680. doi:10.1038/s41587-021-01158-1

Walker FO. 2007. Huntington's disease. *Semin Neurol* **27:** 143–150. doi:10.1055/s-2007-971176

Wang CH, Lunn MR. 2008. Spinal muscular atrophy: advances in research and consensus on care of patients. *Curr Treat Options Neurol* **10:** 420–428. doi:10.1007/s11940-008-0044-7

Wang JR, Holt J, McMillan L, Jones CD. 2018. FMLRC: hybrid long read error correction using an FM-index. *BMC Bioinformatics* **19:** 50. doi:10.1186/s12859-018-2051-3

Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* **37:** 1155–1162. doi:10.1038/s41587-019-0217-9

Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* **20:** 129. doi:10.1186/s13059-019-1727-y

Yan MY, Ferguson B, Bimber BN. 2019. VariantQC: a visual quality control report for variant evaluation. *Bioinformatics* **35:** 5370–5371. doi:10.1093/bioinformatics/btz560

Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, Luo R. 2022. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci* **2:** 797–803. doi:10.1038/s43588-022-00387-x

Zielenski J, Bozon D, Kerem B, Markiewicz D, Durie P, Rommens JM, Tsui L-C. 1991. Identification of mutations in exons 1 through 8 of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Genomics* **10:** 229–235. doi:10.1016/0888-7543(91)90504-8

Zook JM, Catoe D, McDaniel J, Vang L, Spies N, Sidow A, Weng Z, Liu Y, Mason CE, Alexander N, et al. 2016. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci Data* **3:** 160025. doi:10.1038/sdata.2016.25

Zook JM, McDaniel J, Olson ND, Wagner J, Parikh H, Heaton H, Irvine SA, Trigg L, Truty R, McLean CY, et al. 2019. An open resource for accurately benchmarking small variant and reference calls. *Nat Biotechnol* **37:** 561–566. doi:10.1038/s41587-019-0074-6

Received November 5, 2024; accepted in revised form December 17, 2025.