varsomeclinical

WHITE PAPER

Clinical Genomic Interpretation in the Long-Read Era

CONTENTS

1. Introduction 3

2. Clinical Utility & Diagnostic Yield 4

3. Where Short Reads Remain Strong 6

4. Practical Guidance for Decision-Making 7

5. Interpretation 9

6. The Future of Clinical Genomics 12

7. References 13

LIST OF FIGURES

Figure 1. Long Read Mapping 5

Figure 2. Short Read Mapping 6

Figure 3. Sequencing Decision Framework Flowchart 8

Figure 4. VarSome Clinical Sample View 10

Figure 5. VarSome Clinical Region Browser 11

Figure 6. VarSome Clinical Germline Classification 11

Introduction

Short-read sequencing (SRS) has driven the rapid expansion of clinical genomics, powering diagnostics in rare disease, oncology, and reproductive health. Its high throughput, reproducibility, and relatively low cost have made it the standard technology for accredited laboratories worldwide. Yet, despite its success, its limitations are well recognized. Genomic regions dense with repeats, duplications, or pseudogenes remain difficult to resolve; structural variants (SVs) are often fragmented or miscalled, haplotype phasing across long distances is lost, and diagnostic yields plateau after comprehensive short-read analysis.

Long-read sequencing (LRS) has emerged as a way to address these gaps. By generating reads that span tens of kilobases, LRS enables continuous mapping across repetitive regions, accurate localization of breakpoints in SVs, and direct phasing of compound heterozygous alleles. Crucially, it allows the detection of epigenetic features such as DNA methylation natively, without the need for bisulfite conversion or separate assays[1].

The past few years have seen major technological advances that strengthen the case for clinical adoption. Pacific Biosciences' (PacBio) HiFi sequencing offers highly accurate consensus reads, reaching ~99.9% per-base accuracy, and shows reliable performance for both small variants and complex rearrangements[2]. Oxford Nanopore Technologies (ONT) has, in parallel, developed duplex sequencing chemistry, which has sharply improved raw read accuracy while retaining the advantages of long reads, flexible throughput, and native methylation detection. A recent population-scale evaluation within the All of Us Research Program highlighted that these improvements make LRS practical at scale, from portable MinION instruments to high-capacity PromethION devices[3].

Meanwhile, clinical studies are beginning to demonstrate the practical impact of LRS. Across different cohorts, LRS has been shown to increase diagnostic yields by 11-38% in cases that remained unsolved after exome or short-read genome sequencing[4–6]. In oncology, long reads are enabling detailed reconstruction of extrachromosomal DNA (ecDNA) and other complex tumor features that cannot be fully resolved with SRS, providing a more comprehensive view of cancer genome architecture[7].

These advances suggest a turning point. SRS continues to dominate many applications, thanks to its cost-efficiency, depth, and mature analysis pipelines. But long reads offer answers short reads cannot. As costs continue to fall and accuracy improves, the challenge will be to define where LRS provides unique clinical value, and how to ensure its outputs are interpreted consistently and meaningfully in patient care.

Clinical Utility and Diagnostic Yield

Peer-reviewed reports show that LRS can identify pathogenic variants that SRS fails to resolve, offering additional diagnostic potential in clinical genomics.

Rare Disease Diagnostics

Standard short-read whole-exome sequencing (WES) and whole-genome sequencing (WGS) solve a substantial number of cases, but diagnostic yield often plateaus. Studies applying PacBio HiFi sequencing to unsolved cohorts have shown that long reads can detect SVs, repeat expansions, and splice-disrupting variants missed by traditional approaches[4]. In exome-negative autosomal recessive cases, AlAbdi et al. (2023) used long-read WGS (lr-WGS) to find candidate variants, including noncoding, structural, and splice variants in 38% of their cohort, providing potential explanations for 13 families that had previously been undiagnosed[6].

Repeat Expansion Disorders

Diseases such as Huntington's disease, fragile X syndrome, and myotonic dystrophy arise from unstable repeat tracts that are difficult or impossible to size accurately with short reads. Long reads can cover the full expansion, enabling precise measurement of repeat length, internal motif interruptions, and methylation status[8]. In targeted nanopore assays, long reads have resolved large and complex short tandem repeat (STR) expansions that confound conventional short-read methods[9]. In broader pipelines, lr-WGS has been employed to detect multiple repeat expansion disorders in parallel, consolidating what once required discrete locus-specific assays[10].

Structural Variants and Complex Rearrangements

LRS provides improved breakpoint resolution, allowing SVs to be mapped in their genomic context with base-pair precision. This is especially valuable in regions with high repeat content or segmental duplications, where short reads often fail[11]. Long reads also make it possible to phase SVs with nearby single-nucleotide variants (SNVs), distinguishing *cis* from *trans* configurations, and clarifying compound heterozygous states, an advantage that can directly influence variant interpretation and inheritance assessment in clinical settings. In tumor samples, LRS has resolved complex rearrangements, ecDNA, and viral insertion events that were not fully characterized by SRS[12].

Methylation and Imprinting Disorders

Unlike bisulfite-based assays, certain long-read platforms, such as ONT and PacBio, can infer methylation directly from native DNA molecules. This enables detection of imprinting disorders, conditions often driven by methylation changes rather than sequence variants. By integrating methylation signals with phased sequence data, clinicians can interpret pathogenic variants in their epigenetic context, which may enhance diagnostic precision in disorders involving allele-specific methylation. In a proof-of-concept study, long reads were used to detect epigenetic signatures in developmental disorder patients while concurrently calling SNVs, SVs, skewed X-inactivation, and imprinting in a single workflow[13].

Repeat Expansion Disorders

LRS provides insights into tumor genome complexity that SRS struggles with. In a cohort of advanced cancers, long reads enabled the reconstruction of ecDNA, viral integrations, and large rearrangements. They also revealed allele-specific methylation and haplotype-phased methylation landscapes[12]. These features are increasingly recognized as clinically relevant to prognosis and therapy response. While LRS is not yet standard in oncology, its ability to produce integrated structural, sequence, and methylation information in one assay positions it as a compelling tool for precision oncology.

Although these studies demonstrate the expanding potential of LRS, its integration into routine clinical practice requires balancing these advantages with cost, throughput, and regulatory readiness.

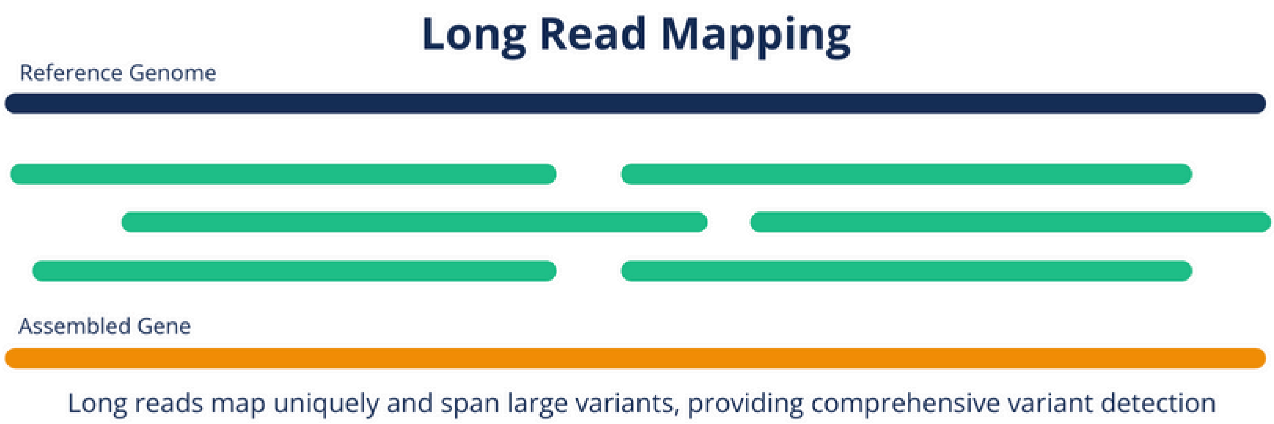

Figure 1. Long-read mapping. Long reads span large variants and align uniquely, providing comprehensive assessment of disease genes and clearer resolution of complex regions.

Where Short Reads Remain Strong

SRS continues to deliver excellent sensitivity and specificity for SNVs and short indels. Its cost per sample is lower, throughput is higher, and pipelines are mature, widely accredited, and support regulatory frameworks. For population studies, carrier screening, and targeted panels, where depth of coverage and turnaround are critical, SRS remains the most practical choice.

Despite rapid improvements, LRS still faces challenges. Costs, while falling, remain higher per-sample, often several-fold higher than SRS in most diagnostic contexts, though the gap continues to narrow as chemistry and throughput improve. Achieving very high coverage across whole genomes with LRS can be more difficult than with short reads, limiting sensitivity in mosaic or low-frequency somatic contexts. Although raw accuracy has improved substantially, certain error profiles (such as homopolymers and extreme GC content) still require careful calibration.

Finally, bioinformatics pipelines for SVs, repeats, and methylation are less standardized than those for short reads, adding complexity for laboratories seeking accreditation. Regulatory validations and accreditation frameworks for LRS are still emerging, meaning laboratories face longer lead times and higher validation costs when establishing clinically accredited workflows.

The two technologies are best viewed as complementary rather than competitive. SRS provides cost-efficient, reliable detection of small variants at scale, while LRS expands diagnostic reach into SVs, repeat expansions, phasing, and methylation. As costs fall and bioinformatics tools mature, hybrid or sequential strategies are likely to become increasingly common.

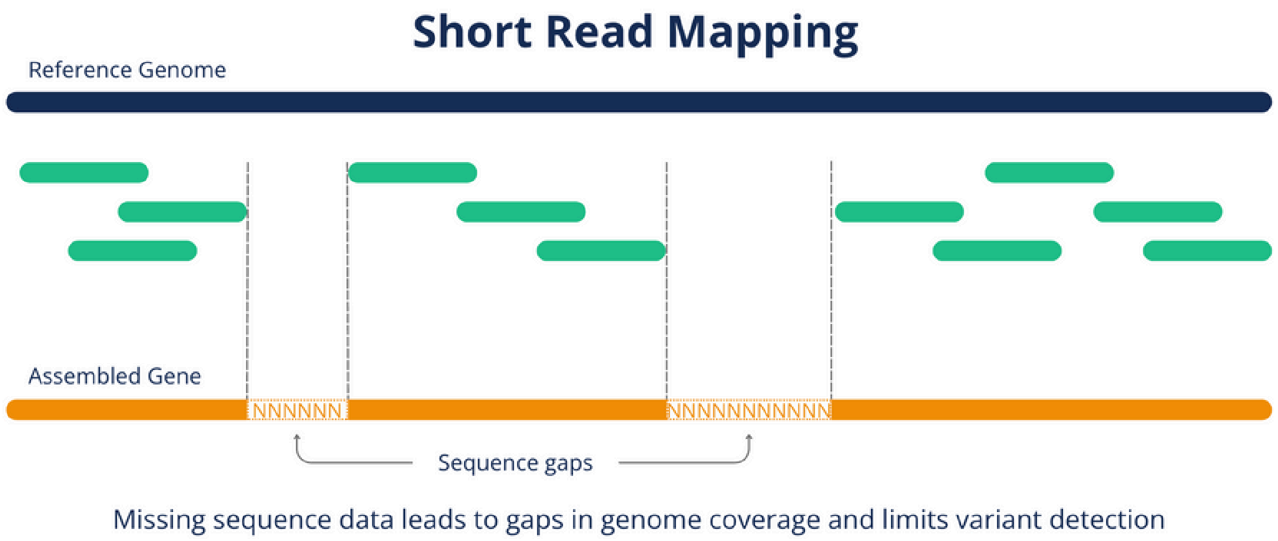

Figure 2. Short-read mapping: Incomplete coverage from short reads can create sequence gaps that obscure variants within disease genes.

Practical Guidance For Decision-Making

Choosing between LRS and SRS depends on the clinical context, variant classes of interest, and resource constraints. The following points outline when each approach is most appropriate and how hybrid strategies may be deployed.

When to use long reads

- **Repeat expansion disorders:** Direct sizing of tracts, detection of interruptions, and methylation analysis.
- **Unsolved cases after WES/WGS:** LRS adds diagnostic yield by capturing SVs, splice variants, and other events missed by SRS.
- **Pseudogene-rich or repetitive regions:** LRS reduces misalignment and clarifies variant calls in complex loci.
- **Methylation or imprinting disorders:** Native detection of allele-specific methylation without bisulfite conversion.
- **Oncology applications:** Reconstruction of ecDNA, viral integrations, and complex rearrangements.

When to use short reads

- **Targeted panels:** Cost-effective for defined gene sets with well-characterized variants.
- **High-throughput screening:** Carrier testing, newborn screening, or population-scale projects where scale and cost dominate.
- **Time- or cost-constrained settings:** Faster turnaround and lower per-sample cost remain practical advantages of SRS.

Hybrid or stepwise models

- **Sequential testing:** Perform SRS first, then apply LRS to unresolved cases for additional diagnostic yield.
- **Complementary use:** Combine short- and long-read data in the same patient, using SRS for small variants and LRS for SVs and repeats.
- **First-line LRS pilots:** Some centers are trialing LRS as a primary assay in high-yield contexts (e.g., rare disease cohorts), though this remains largely experimental.

The two technologies are best viewed as complementary rather than competitive. SRS provides cost-efficient, reliable detection of small variants at scale, while LRS expands diagnostic reach.

Decision framework

1. Phenotype assessment → identify if repeat expansions, imprinting disorders, or complex SVs are likely.
2. Select assay
   - Routine coding variant screening, high-throughput, or cost-limited → SRS
   - Repeat expansion, methylation, complex SVs suspicion, or unresolved exome/genome → LRS
3. Analysis and interpretation → integrate results into a unified framework (classifiers, databases, community standards).

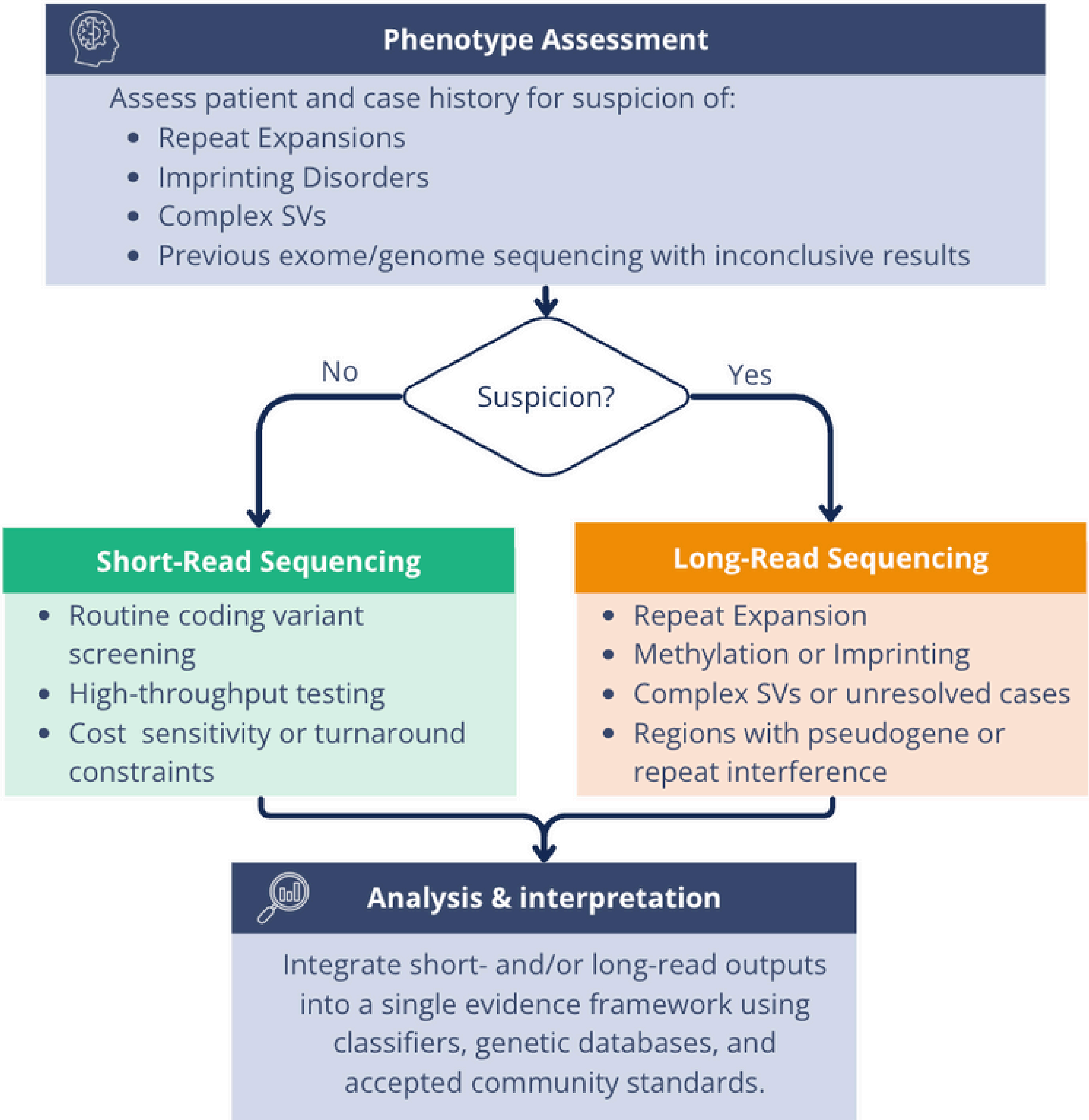

Figure 3. Flowchart showing how phenotype assessment guides the choice between short- and long-read sequencing. Suspected repeat expansion, imprinting, or complex SVs direct users to long reads; short reads suit routine, high-throughput testing.

Interpretation

As sequencing technologies advance, the limiting factor in clinical genomics is often interpretation. LRS adds new biological dimensions that extend beyond the output of traditional short-read pipelines. These data offer unprecedented insights but also introduce substantial interpretive complexity. The challenge for clinical laboratories is to translate these signals into clear, reproducible variant classifications and patient reports.

Existing frameworks built for SNVs and short indels are not always equipped to handle the richer variant classes revealed by LRS. SVs can span multiple genes or regulatory regions, and some repeat expansions influence, or are influenced by, local chromatin and methylation patterns, which makes the boundary between genetic and epigenetic effects less clear. Without standardized methods for integrating these data, there is a risk of inconsistent results across laboratories and uncertainty for clinicians.

VarSome Clinical helps experts connect sequencing output to clinical decision-making. With both SRS and LRS compatibility, VarSome Clinical helps harmonize results in a sequencer-agnostic platform. Within this framework, variants are evaluated against over 140 data sources, including population databases, gene- and region-level annotations, and curated clinical evidence. VarSome's ACMG/AMP-based classifiers support SNVs, indels, and CNVs, with work ongoing in the wider community to extend standardized frameworks to other SV classes. VarSome's classifiers help ensure each variant class is assessed using transparent, reproducible criteria, regardless of how it was detected.

For laboratories beginning to incorporate LRS, this consistency is essential. VarSome Clinical enables results from different assays to be interpreted through a unified evidence model, providing a single environment where geneticists, bioinformaticians, and clinicians can review the same data and reasoning. VarSome Clinical provides a clinician-friendly summary that consolidates population frequencies, literature evidence, and functional annotations into a transparent and interpretable classification aligned with international standards.

Within VarSome Clinical, several tools support this interpretive layer. The Sample View provides an integrated environment for reviewing SNVs, indels, and SVs at the sample level. Long reads produce large, phased, and structurally complex calls. The Sample View helps users inspect the events found in their sample in the proper genomic context. It displays variant positions alongside transcripts, conservation scores, ROH segments, and phasing information, which supports verification of breakpoints and helps clarify how small and large variants interact across a locus. This allows laboratories to examine the structure and local context of SVs, assess phased or compound configurations, and interpret variant patterns that may influence inheritance or pathogenicity.

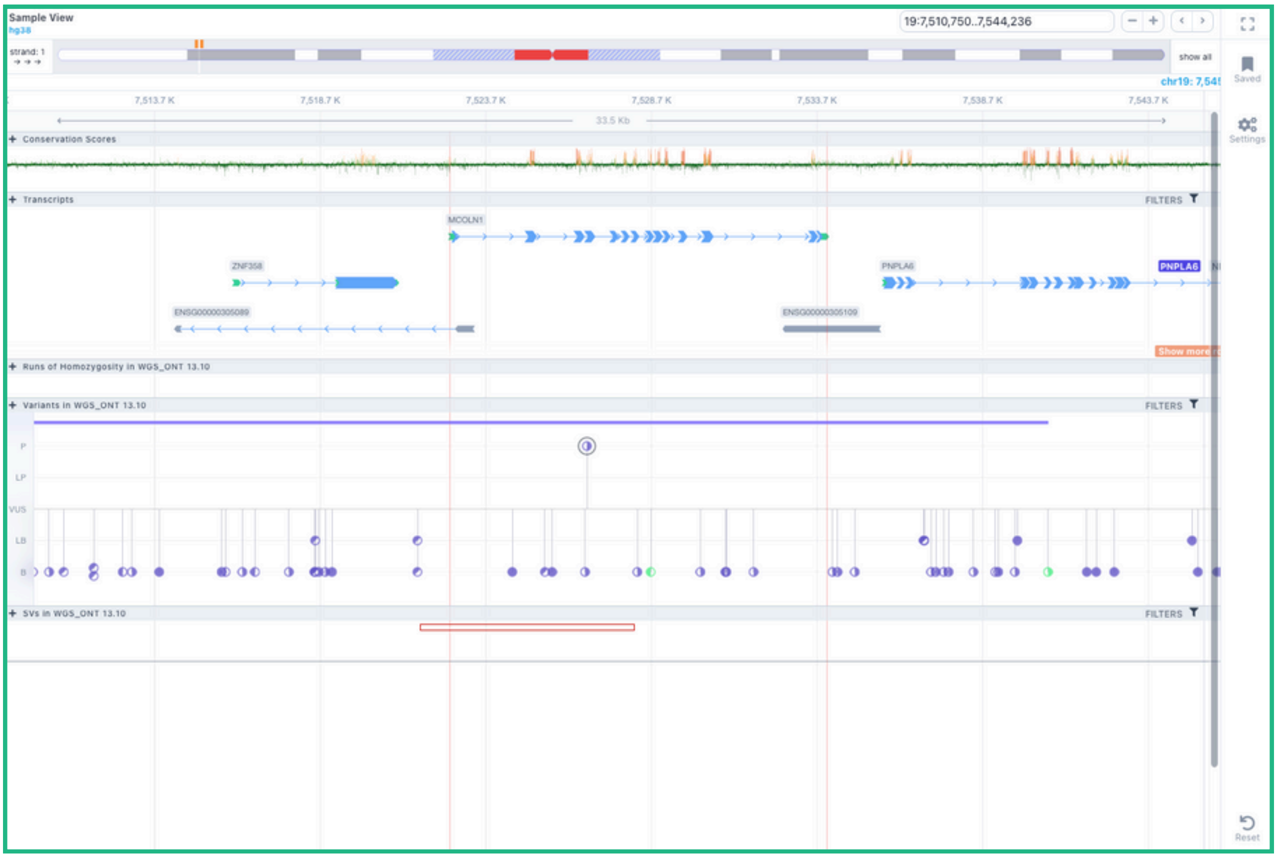

Figure 4. VarSome Clinical Sample View: Integrated tracks show variant positions, transcripts, conservation, ROH, and phasing to help assess small and structural variants across a locus.

The Region Browser provides the broader genomic context needed to interpret complex events. It presents gene structures, transcripts, conservation scores, and population frequencies across a genomic region, and clinical annotations across the genomic interval, allowing users to see how a variant intersects with exons, promoters, and regulatory elements. Users can refine the view with filters, explore variants by source or coding impact, and visually assess clinically meaningful patterns; for example, if pathogenic missense variants cluster within functional domains or known hotspots, or if they concentrate in constrained regions that show low tolerance to variation. This helps clinical scientists understand the structural footprint of an event, assess which transcripts or functional domains are affected, and evaluate how the breakpoint fits within the wider architecture of the locus.

Figure 5. VarSome Clinical Region Browser: The browser shows transcripts, conservation, population data, and clinical annotations across a locus to help assess how variants intersect with functional elements.

VarSome Clinical also includes many dynamic and algorithmic filters that streamline triage by narrowing the search space to variants that match defined pathogenicity, phenotype, or inheritance criteria. These filters use gene–phenotype links, mode-of-inheritance rules, and curated annotations to surface variants that are more likely to be relevant to the case. They support consistent review across different assays by reducing manual inspection and helping users focus on the signals that matter most in both short- and long-read datasets.

As LRS expands the scope of genomic diagnostics, interpretation frameworks must advance in parallel. Tools that combine transparency, regulatory readiness, and cross-platform compatibility will be critical to ensuring that new sequencing capabilities translate into confident, actionable insights for patient care.

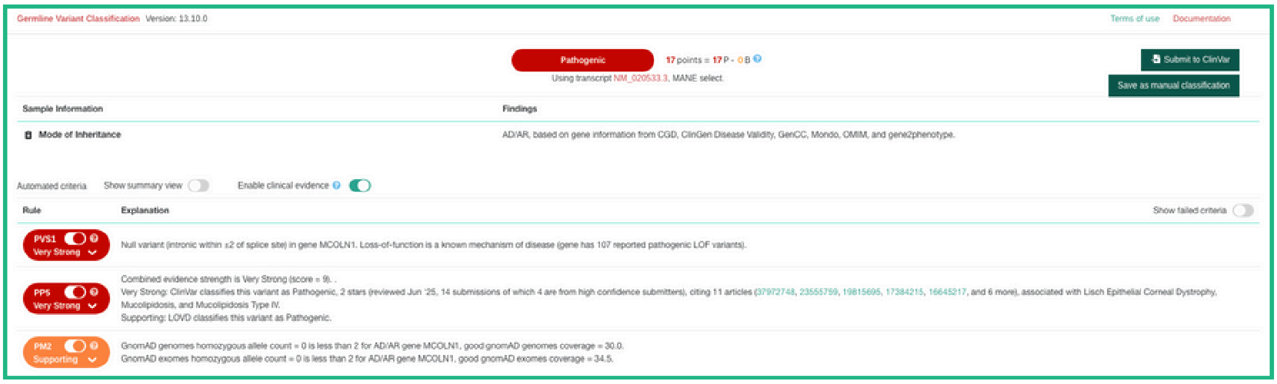

Figure 6. VarSome Clinical Germline Classification: The classifier displays the pathogenic CNV call, along with triggered rules, evidence summaries, and links to supporting data.

The Future of Clinical Genomics

The boundaries between short- and long-read sequencing are becoming less distinct as each technology evolves and costs converge. Short reads continue to anchor routine diagnostics through established pipelines and regulatory maturity, while long reads expand the variant classes that can be detected.

Over the next few years, hybrid and stepwise approaches are likely to dominate. Many laboratories will continue to deploy SRS as a first-line test, reserving long reads for unresolved cases or phenotypes strongly suggestive of complex variation. Others will integrate both from the outset, leveraging long-read context to improve short-read interpretation or confirm challenging regions. As sequencing platforms improve accuracy and reduce costs, first-line LRS will become more realistic for a growing number of indications.

The challenge will not lie in data generation but in interpretation and integration. Each new data layer–structural, epigenetic, haplotypic–adds richness but also complexity. This richness also increases storage demands, requiring new approaches to data management, privacy, and secure sharing, particularly as long-read datasets include methylation and phasing information that go beyond conventional variant files. Ensuring that laboratories apply consistent, transparent standards to these results will be essential for maintaining clinical confidence and regulatory compliance. Interpretation frameworks, such as VarSome Clinical, will play a key role here, providing the structured evidence models needed to convert raw data into reproducible, clinically meaningful findings.

Ultimately, the goal is not to choose one sequencing technology over another, but to ensure that every patient benefits from the approach most likely to reveal a diagnosis. The emerging landscape of clinical genomics will therefore be defined by interoperability between technologies, laboratories, and interpretation systems. Long- and short-read sequencing will work in concert, and interpretation platforms will provide the clarity needed to transform complex data into confident clinical action.

References

1.Gigante S, Gouil Q, Lucattini A, et al. Using long-read sequencing to detect imprinted DNA methylation. Nucleic Acids Res. 2019;47(8):e46-e46. doi:10.1093/nar/gkz107

2. Pacific Biosciences. How HiFi sequencing works. PacBio. Accessed December 1, 2025. https://www.pacb.com/technology/hifi-sequencing/how-it-works/

3. Mahmoud M, Huang Y, Garimella K, et al. Utility of long-read sequencing for All of Us. Nat Commun. 2024;15(1):837. doi:10.1038/s41467-024-44804-3

4. Steyaert W, Sagath L, Demidov G, et al. Unraveling undiagnosed rare disease cases by HiFi long-read genome sequencing. Genome Res. 2025;35(4):755-768. doi:10.1101/gr.279414.124

5. Del Gobbo GF, Boycott KM. The additional diagnostic yield of long-read sequencing in undiagnosed rare diseases. Genome Res. 2025;35(4):559-571. doi:10.1101/gr.279970.124

6. AlAbdi L, Shamseldin HE, Khouj E, et al. Beyond the exome: utility of long-read whole genome sequencing in exome-negative autosomal recessive diseases. Genome Med. 2023;15(1):114. doi:10.1186/s13073-023-01270-8

7. Sun R, Li Z, Liu Y, et al. Landscape of extrachromosomal DNA characteristics in high-grade serous ovarian cancer via long-read sequencing. Cell Rep. 2025;44(10):116343. doi:10.1016/j.celrep.2025.116343

8. Kaplun L, Krautz-Peterson G, Neerman N, et al. ONT in Clinical Diagnostics of Repeat Expansion Disorders: Detection and Reporting Challenges. Int J Mol Sci. 2025;26(6):2725. doi:10.3390/ijms26062725

9. Stevanovski I, Chintalaphani SR, Gamaarachchi H, et al. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. Sci Adv. 2022;8(9):eabm5386. doi:10.1126/sciadv.abm5386

10. Berlyoung A, Noya J, Johnson A, et al. P635: Long read sequencing for the analysis of repeat expansion disorders. Genet Med Open. 2024;2:101541. doi:10.1016/j.gimo.2024.101541

11. Moustakli E, Christopoulos P, Potiris A, et al. Long-Read Sequencing and Structural Variant Detection: Unlocking the Hidden Genome in Rare Genetic Disorders. Diagnostics. 2025;15(14):1803. doi:10.3390/diagnostics15141803

12. O'Neill K, Pleasance E, Fan J, et al. Long-read sequencing of an advanced cancer cohort resolves rearrangements, unravels haplotypes, and reveals methylation landscapes. Cell Genomics. 2024;4(11):100674. doi:10.1016/j.xgen.2024.100674

13. Geysens M, Huremagic B, Souche E, et al. Clinical evaluation of long-read sequencing-based episignature detection in developmental disorders. Genome Med. 2025;17(1):1. doi:10.1186/s13073-024-01419-z

VSC-WP-004-V01

Built with Swiss precision by

SAPHETOR

EPFL Innovation Park - C
1015 Lausanne, Switzerland