# Progress in Nanopore Sequencing Bioinformatics (2018–2023) and Hybrid Illumina Integration

Over the last five years, nanopore sequencing (a third-generation, single-molecule sequencing technology by Oxford Nanopore Technologies, ONT) has seen significant bioinformatics advancements. Many new tools and techniques have emerged to improve basecalling accuracy, error correction, hybrid assembly with Illumina data, structural variant detection, and epigenetic analysis. Notably, hybrid workflows that combine ONT’s ultra-long reads with Illumina’s high-accuracy short reads have become common to leverage the strengths of each platform. Below, we summarize these developments and their applications in cancer genomics, infectious disease surveillance, metagenomics, transcriptomics, and rare disease diagnosis. We also discuss technical limitations and ongoing challenges.

## Nanopore vs Illumina Sequencing: Capabilities Comparison

Nanopore sequencing differs fundamentally from Illumina short-read sequencing. The table below highlights key differences relevant to bioinformatics and analysis:

| **Feature** | **Illumina (Short-Read)** | **Nanopore (Long-Read)** |
| --- | --- | --- |
| Read length | 150–300 bp (paired-end) | Up to megabases (often 10–100 kb; record >4 Mb) ([Frontiers |
| Per-base accuracy | ~99.9% (Q30) raw; consensus very high | ~90–98% raw (Q15–Q20) ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance |
| Error profile | Mainly substitutions; low indel rate | More indels (esp. homopolymers); errors near mods ([ |

Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC

](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/#:~:text=quantification%20of%20full,to%20derive%20accurate%20consensus%20sequences)) |

| Throughput (per run) | High (e.g. ~300 Gb on NovaSeq flowcell) | Varies: ~10–50 Gb on MinION; >200 Gb on PromethION | | Run time | Requires batch completion (e.g. 1–2 days) | Real-time data streaming; can yield results in hours | | Biases | PCR/amplification bias; GC extremes undercovered | No amplification needed; better GC/repeat coverage ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Despite having a similar alignment,long reads of nanopore sequencing)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=It is crucial to produce,skeletons are located on chromosomes)) | | Structural variation | Detection indirect (split reads, assembly) | Long reads span SVs end-to-end (better SV resolution) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Structure variants ,allowing for precise characterization and)) ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=complex genetic changes%2C like structural,how TGS is used to) ) | | Epigenetic detection | Requires separate assay (e.g. bisulfite for 5mC) | Direct from raw signal (5mC, 6mA, m^6A RNA) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=traditional sequencing methods and methylation,detect DNA methylation in nanopore)) | | Portability | Large instruments (lab-bound) | Portable devices available (MinION, Flongle) | | Cost per Gb | Low for large projects (economy of scale) | Continual reagent usage; competitive for small projects | | Data processing | Mature pipelines (GATK, etc.) | Developing tools; often high computational demand |

Both platforms are complementary. Illumina’s reads offer high single-base accuracy and established analysis pipelines, while nanopore reads provide long-range information (resolving repeats, phasing haplotypes, and characterizing structural variants and modifications).

## Advances in Nanopore Sequencing Bioinformatics (2018–2023)

### Basecalling Improvements

Basecalling – converting nanopore raw electrical signals (“squiggles”) to DNA/RNA bases – has dramatically improved via novel algorithms and deep learning. Early neural network basecallers like Albacore (ONT’s official basecaller) were succeeded by Guppy, which introduced GPU acceleration for speed and accuracy ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=base calling analysis based on,of a slower base finding)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=years,in the table at the)). Guppy’s high-accuracy mode (HAC) and super-accuracy mode (SUP) now routinely yield raw read accuracies around 98–99% (Q20) using the latest R10.4 pore chemistry and “Q20+” kits. Third-party and research basecallers have also proliferated:

* Chiron – one of the first deep learning basecallers (RNN-based). It demonstrated feasibility but lagged in accuracy and speed (even with GPU) compared to Guppy ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=base calling analysis based on,of a slower base finding)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=accuracy metrics%2C ONT’s Albacore and,are interested in its development)).
* Bonito – an open-source ONT research project using convolutional models. It improved raw read accuracy further (sometimes at slower speeds) and allowed users to train custom models ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=fastest due to its GPU,are interested in its development)).
* Dorado – a more recent ONT tool (2022+) that implements improved neural network models and supports newer hardware, continuing the trend of accuracy gains.
* DeepNano-coral – a lightweight basecaller running on low-power devices (10 W), achieving ~Guppy-fast mode accuracy in real time ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=open,2021)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=correct the basecalling errors introduced,There are many)).
* Ravvent – an RNN encoder-decoder basecaller using joint raw signal and event processing ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=modification from dRNA sequence data,time base calling)).
* UNCALLED – not a basecaller per se, but an alignment tool that maps raw signal to a reference on-the-fly ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=nanopore signals for real,Readfish)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=performance in drawing yeast,a simple API has achieved)). It enables selective sequencing (deciding in real-time which reads to keep) and highlights the integration of basecalling with analysis.

Crucially, deep learning architectures have evolved: earlier recurrent neural networks are giving way to convolutional and transformer models (e.g., Temporal Convolutional Networks) for speed and accuracy gains ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=improvement is based on the,now train models with features)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=achieved competitive results ,data and raw FAST5 nanopore)). Custom hardware is also leveraged – one study demonstrated >100× speed-ups using an FPGA coprocessor for basecalling ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=UNCALLED is an open,Readfish)). These advances have reduced nanopore basecalling error rates from ~10–15% (2016–2017) to ~1–2% in 2023 for single-pass reads, greatly enhancing downstream analysis.

### Error Correction and Polishing

Even with better basecalling, nanopore reads still have higher error rates than Illumina. Bioinformatics methods for error correction are therefore critical, especially for assembly and variant calling. There are two main strategies: self-correction using overlapping long reads, and hybrid correction using short reads.

* Self-correction tools: Long reads can be corrected by consensus of multiple reads. For example, Racon and Medaka (from ONT) polish draft assemblies by aligning raw reads to contigs and improving consensus. NanoPolish (2017) uses signal-level data to call consensus and even detect methylation simultaneously. More recent algorithms like partial order alignment improvements (e.g., abPOA) accelerate consensus finding – abPOA achieves similar alignment accuracy as prior methods but ~15× faster ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=directed acyclic graph is common,As a result of)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=independently as an alignment tool,tools are available%2C and the)). QAlign (2021) is another preprocessor that improves long-read alignment accuracy, which in turn benefits error correction and SV calling ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=when used to correct errors,that uses adaptive banded dynamic)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=independently as an alignment tool,tools are available%2C and the)). New graph-based methods (directed acyclic graphs of reads) also help represent variants during correction ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=directed acyclic graph is common,As a result of)).
* Hybrid error correction: Illumina reads, with their low error rate, can correct nanopore reads. Tools like Nanocorr (2015) introduced this concept; more recently, pipelines often use an initial long-read assembly followed by Pilon or POLCA polishing with Illumina reads ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=developed a software tool called,quality)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=With the combination of long,and with less memory usage)). POLCA (2019) was shown to produce highly contiguous assemblies with low error rates by polishing long-read assemblies with short reads ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=developed a software tool called,quality)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=With the combination of long,and with less memory usage)). In assembly benchmarks, incorporating short reads either before assembly (hybrid assembly) or after (polishing) significantly improves base accuracy. For instance, in Clostridioides difficile genomes, polishing nanopore assemblies with short reads corrected hundreds of errors and fixed >180 misassigned alleles in MLST analysis ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11267-9" \l ":~:text=,provided comparable%2C satisfactory results for)) ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11267-9" \l ":~:text=(~ 0.015,frame deletions in tcdC)).

The trend is toward integrated polishing pipelines that apply multiple rounds of long-read consensus and short-read polishing. Care must be taken not to over-polish: too many rounds can sometimes re-introduce minor errors ([Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes | Microbiome | Full Text](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8" \l ":~:text=,and one round of medaka)) ([Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes | Microbiome | Full Text](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8" \l ":~:text=with two rounds of Racon,Only MAGs with the)). Nonetheless, these error-correction advancements now enable high-accuracy genome assemblies and variant calls from nanopore data, especially when Illumina data is available for polishing ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=developed a software tool called,quality)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=With the combination of long,and with less memory usage)).

### Hybrid De Novo Assembly

Nanopore reads, with read lengths tens of kilobases, dramatically improve assembly contiguity by spanning repeats and structural variants that short reads cannot resolve ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Structure variants ,allowing for precise characterization and)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=assemblies may result%2C while PCR,repeat expansions and repetitive regions)). However, due to residual errors, a purely nanopore-based assembly may have lower base accuracy and some misassemblies. Hybrid assembly approaches combine ONT and Illumina data to produce assemblies that are both continuous and accurate. Key developments include:

* New hybrid assemblers: Assemblers like Unicycler, MaSuRCA, and SPAdes (with hybridSPAdes mode) were benchmarked. In a study of bacterial pathogens, Unicycler gave the most contiguous genomes, followed closely by MaSuRCA, whereas short-read-only assemblies were more fragmented ([Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07041-8" \l ":~:text=Results)) ([Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07041-8" \l ":~:text=Unicycler performed the best for,The numbers of)). Unicycler’s pipeline (which maps long reads to short-read assembly and iteratively bridges gaps) was deemed “superior” for combining Illumina and ONT data ([Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07041-8" \l ":~:text=Conclusions)). OPERA-MS (2019) is another hybrid assembler optimized for metagenomes; it efficiently scaffolds long reads with short-read assemblies and was shown to recover complete genomes at even ~9× long-read coverage ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=Theoretically%2C long,and there are only a)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=complex microbial community in a,quality MAGs)).
* Assembly polishing integration: Modern pipelines often incorporate polishing during assembly. For example, Flye can assemble long reads and then polish with long-read consensus. A hybrid pipeline may run Flye (ONT only) for contigs, then Pilon/Polca with Illumina for polishing ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=developed a software tool called,quality)). Alternatively, MaSuRCA performs an initial error correction of nanopore reads with short reads, then assembles corrected reads, and finally polishes.
* Gap closing and circularization: Tools like TGS-GapCloser (2019) specifically address long-read assemblies by filling gaps without needing additional error correction ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=automatically,synteny between draft assemblies and)). It improved draft assembly N50 by ~25%. CCBGPipe (2019) was developed to fully circularize bacterial genomes using long reads ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=With the continuous maturity of,on average by 25 percent)).
* Large genome assembly: The Telomere-to-Telomere (T2T) consortium demonstrated that ONT ultra-long reads can assemble entire human chromosomes end-to-end, a feat impossible with short reads. By combining PacBio HiFi long reads for base accuracy and ONT reads for ultra-long contiguity, the first complete human genome (including centromeres and repeat arrays) was achieved in 2022. This showcased how hybrid strategies can conquer even the largest genomes.

Result: Hybrid assembly workflows yield high-contiguity, high-accuracy genomes. A human gut microbiome study highlighted that combining Illumina and ONT reads significantly increased contig N50 and genome completeness. Hybrid assemblies had contigs 3–6× longer than Illumina-only, enabling better gene and operon annotation and recovery of plasmids ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=of folds of data,average%2C ∼3× to ∼6× larger)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=We presented human gut microbiome,which could serve as a)). In Staphylococcus aureus isolates, a 2024 study found hybrid assemblies were more complete than either technology alone and correctly resolved all MLST strain types and antibiotic resistance genes ( [Hybrid Illumina-Nanopore assembly improves identification of multilocus sequence types and antimicrobial resistance genes of Staphylococcus aureus isolated from Vermont dairy farms: comparison to Illumina-only and R9.4.1 nanopore-only assemblies - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11077346/" \l ":~:text=Our analyses showcased the hybrid,the hybrid assembly approach emerged) ) ( [Hybrid Illumina-Nanopore assembly improves identification of multilocus sequence types and antimicrobial resistance genes of Staphylococcus aureus isolated from Vermont dairy farms: comparison to Illumina-only and R9.4.1 nanopore-only assemblies - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11077346/" \l ":~:text=Furthermore%2C the hybrid assembly accurately,AMR management strategies crucial for) ). In summary, hybrid assembly capitalizes on Illumina’s accuracy and ONT’s length, producing near-finished genomes in many cases.

### Structural Variant Detection

Long-read sequencing has been transformative for structural variant (SV) detection. SVs include large insertions, deletions, inversions, translocations, and repeat expansions (typically >50 bp) that short-read methods often miss or misassemble ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=,Variants) ). Several advancements and tools stand out:

* SV Callers for long reads: Early on, tools like Sniffles (2017) enabled SV calling from long reads. In the last five years, new callers have improved speed and sensitivity: cuteSV (2020) uses a streamlined approach to gather SV signatures from noisy reads, achieving high sensitivity and scalability ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Furthermore%2C in addition to the,at frequencies as low as)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=proposes cuteSV%2C a sensitive%2C fast%2C,fails to detect pathogenic SV)). SVIM (2019) and NanoVar (2018) introduced robust algorithms for ONT/PacBio data, and Sniffles2 (2021) improved on its predecessor, boosting speed and adding better genotyping. A recent entrant, SEN-SV (2022), was designed for cancer genomes and claims high sensitivity to all SV types with breakpoint precision ~100 bp ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=ONT WGS data%2C SV calling,tools below for your convenience)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=duplications%2C and unbalanced translocations,tools below for your convenience)).
* Coverage and signal integration: Because ONT reads have residual errors, algorithms often require high coverage or employ statistical methods to distinguish true SV signals from noise. For example, Igor (2021, referred to as “igda” in one study) phases SNVs at 0.2% frequency to assist detecting low-allele-fraction SVs in heterogeneous samples ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Furthermore%2C in addition to the,at frequencies as low as)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=the,(2022)). This is particularly relevant in cancer, where subclonal SVs may be present at low frequency.
* Hybrid SV analysis: While long reads can directly reveal SVs by spanning breakpoints, Illumina data is sometimes used to supplement SV discovery – e.g., for precise breakpoint refinement or for validating SVs. One strategy is to do a long-read assembly and then align short reads to it to confirm and quantify structural differences. Another is to run short-read SV callers in parallel and combine results, although differences in variant representations make direct integration non-trivial.

Impact: Nanopore sequencing can detect orders of magnitude more SVs than short-read sequencing. Studies show that short reads (<300 bp) miss >70% of human genome SVs >50 bp ([Long-read human genome sequencing and its applications - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC7877196/" \l ":~:text=Long,50 bp)%2C with)). In cancer, this is critical – roughly 55% of driver mutations in cancers are structural variants, outnumbering point mutations ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=encompass structural variants within smaller,outnumbering point) ). For example, long-read sequencing of a colorectal cancer cohort (2021) identified 5,200 novel somatic SVs across 21 tumors – variants undetectable by short-read sequencing. Long reads have pinpointed complex rearrangements in cancer cell lines (e.g., CDKN2A and SMAD4 deletions in pancreatic cancer, large duplications in breast cancer) that were previously ambiguous ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=match at L575 technology to,breast cancer cell line were) ). They also allow phasing of SVs with other variants and even with epigenetic marks in the same reads ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=In this way%2C intact modified,is not without its challenges)), giving a more comprehensive view of structural alterations and their regulatory context.

### Epigenetic and Modification Analysis

Nanopore sequencing directly senses DNA or RNA modifications as changes in ionic current, enabling epigenetic analysis concurrently with basecalling. This has opened new analysis avenues:

* 5-methylcytosine (5mC) in DNA: ONT’s platform can detect 5mC (and 5-hydroxymethylcytosine) without bisulfite conversion. Tools like Nanopolish and Tombo (2018) compare measured signal against expected for unmodified bases to call methylation. ONT’s newer basecalling models (e.g., Remora in Guppy/Dorado) can output modification probabilities alongside bases. The approach yields methylation profiles with lower bias and higher mapping rates than bisulfite sequencing ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=match at L1496 traditional sequencing,detect DNA methylation in nanopore)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=traditional sequencing methods and methylation,detect DNA methylation in nanopore)). Notably, a recent Science (2022) study generated a complete human methylome from nanopore reads, correlating DNA methylation with chromatin state across centromeres and repeats ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=match at L1200 S,Google Scholar) ).
* N6-methyladenosine (m6A) in RNA: Direct RNA sequencing (DRS) on nanopore is unique – it sequences native RNA strands (no reverse transcription), thus retaining modifications like m6A. The algorithm EpiNano (2021) uses features from raw DRS signals to predict m6A sites in transcripts ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=networks ,NanoReviser%2C an)). Although detecting RNA mods is challenging (signal differences subtle), progress is being made. For instance, direct RNA nanopore data helped map m6A in Arabidopsis mRNAs, capturing isoform-specific modification patterns ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=(Choi et al,aspects play an important role)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Arabidopsis wild,Some)).
* Histone modifications & protein binding: While primarily DNA/RNA focused, nanopore’s single-molecule signals can be adapted (with special protocols) to study DNA-protein interactions (e.g., ATAC-see for chromatin accessibility) or even protein sequencing (in early research).
* Computational tools: The growing toolkit includes MegaMod (multi-modification caller), Megalodon (ONT’s framework for per-read mod calling at scale), and academic tools like DeepMod and MRCanu for integrating methylation into assembly polishing. As modifications can cause systematic basecalling errors if unmodeled, one Nature Communications study proposed methods to correct modification-induced errors to improve assembly quality (2023) – highlighting the interplay between basecalling and epigenetics.

Application: In cancer and other diseases, these epigenetic insights are powerful. For example, methylation changes (like CpG island hypermethylation in tumor suppressor genes) can be directly observed. Nanopore sequencing can simultaneously call an SV and note if the region carried certain methylation marks or not – all in one experiment ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=In this way%2C intact modified,is not without its challenges)). This combined genetic–epigenetic reading is not possible with other sequencing platforms without special sample preparations.

## Hybrid Nanopore–Illumina Workflows

Combining nanopore and Illumina data is now a standard strategy in many projects. Hybrid workflows aim to achieve: (a) high consensus accuracy (from Illumina) and (b) long-range resolution (from ONT). Key integration methods include:

* Hybrid Assembly & Polishing: As discussed, assembly pipelines like Unicycler, hybridSPAdes, OPERA-MS, MaSuRCA etc. use both data types. Typically, short reads are assembled first, then long reads scaffold the assembly and resolve repeats ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=Theoretically%2C long,and there are only a)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=genome assembly approach that employs,biogas reactor by using the)). The result is contiguous assemblies with accuracy boosted by polishing. For bacterial genomes, a finished circular chromosome often requires just one hybrid assembly pipeline run. In metagenomics, one study recovered 58 high-quality metagenome-assembled genomes (MAGs) from a human gut sample with a hybrid approach, versus only 11 with short reads alone ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=algorithms%2C particularly hybrid assembly ,In the analysis of)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=We presented human gut microbiome,which could serve as a)). Hybrid assembly also increased plasmid recovery – e.g. identifying more antibiotic resistance plasmids than short-read or long-read-only methods ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=respectively (Supplementary File 2),These findings indicated the)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=AMR genes compared to single,These findings indicated the)).
* Hybrid Error Correction: Instead of assembling first, some workflows first error-correct long reads with short reads. For instance, LoRDEC (2014) and FMLRC (2019) correct long-read sequences by building k-mer maps from short reads. Corrected long reads are then assembled or analyzed. This can reduce base error rates in raw ONT reads substantially (often from ~5% to <1%). However, it requires deep short-read coverage and adds computational steps.
* Integrated Variant Calling: For variant analysis (SNVs and small indels), Illumina data is still gold-standard, but long reads aid in phasing variants and discovering complex variants. A practical approach is to do parallel variant calling on both data types and combine results: use Illumina (GATK or DeepVariant) calls for single-nucleotide variants, and ONT calls (Clair3, etc.) for indels/SVs, then merge. Some pipelines attempt joint calling. For example, Google’s PEPPER-DeepVariant pipeline trains a model to polish ONT read alignments and then call variants, essentially injecting some of the short-read variant calling rigor into long-read data. There are also cases of “shallow hybrid sequencing”: a study showed that 15× ONT + 15× Illumina (rather than 30× of one alone) sufficed for near-complete human genome variant detection ([[PDF] Deep-leaning and shallow hybrid Nanopore-Illumina sequencing for ...](https://www.biorxiv.org/content/10.1101/2024.11.26.625492v1.full.pdf" \l ":~:text=,achieve a high germline)), suggesting cost-effective hybrid WGS is viable.
* Short-read confirmation: In critical applications (like clinical diagnostics), variants (especially novel point mutations) detected in error-prone ONT reads are often validated by Illumina or Sanger sequencing. Conversely, ONT can validate and scaffold Illumina-detected SVs. Tools such as NanoVar can take input from Illumina-called SVs to look for supporting evidence in ONT reads.

Pipeline example: A recent pipeline SUMMER (2023) illustrates a comprehensive hybrid approach for human genomes ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=We present SUMMER%2C a comprehensive%2C,to interface with Docker or) ) ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=software,of LRS in clinical settings) ). SUMMER processes ONT reads through multiple variant callers: it uses Illumina-trained algorithms (Clair3) for SNVs on ONT data, merges multiple ONT SV callers (Sniffles2, SVIM, cuteSV) for robust SV detection, and even calls tandem repeat expansions (via Straglr) ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=tools%2C including ,SUMMER is also) ) ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=that sequenced at a certain,BAM file%2C to meet specific) ). It then outputs a unified variant set with clinical annotations. By benchmarking against the Genome-In-a-Bottle truth set HG002, SUMMER achieved >90% precision/recall for SVs ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=expertise%2C SUMMER employs Python ,of LRS in clinical settings) ) ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=software,of LRS in clinical settings) ) and correctly genotyped known repeat expansions. The developers note this “significant advancement…has the potential to enhance diagnostic efficiency and broaden the adoption of [long-read sequencing] in clinical settings.” ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=software,of LRS in clinical settings) ) ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=successfully identifies the number of,of LRS in clinical settings) ).

Overall, hybrid workflows have become essential to maximize accuracy and completeness. They are prominently used in high-stakes projects, as seen in the applications below.

## Applications in Various Scenarios

### Cancer Genomics

Cancer genomes are riddled with complex mutations – large rearrangements, copy-number variations, gene fusions, and epigenetic alterations – which long reads are uniquely positioned to resolve ([The Promising Role of Nanopore Sequencing in Cancer Diagnostics ...](https://www.sciencedirect.com/science/article/pii/S2772892725000033" \l ":~:text=The Promising Role of Nanopore,CNVs)%2C gene)). Recent years have seen long-read sequencing (including nanopore) applied to cancer research and even clinical diagnostics:

* Structural Variants and Fusions: Many driver alterations in cancer (e.g., gene amplifications, deletions of tumor suppressors, chromosomal translocations like BCR-ABL) are SVs. Short reads often fail to map these or assemble them correctly. Nanopore reads can span entire SVs, allowing direct detection. For instance, nanopore sequencing identified intragenic deletions in CDKN2A in pancreatic cancer that were missed by prior methods ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=match at L575 technology to,breast cancer cell line were) ). In a breast cancer cell line, ONT reads resolved complex rearrangements involving tandem duplications and inversions that short-read assemblies had fragmented ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=technology to detect structural variants,breast cancer cell line were) ). Long reads also facilitate fusion gene discovery by sequencing full-length cDNA; this has helped find novel oncogenic fusions in leukemias and solid tumors (complementing short-read RNA-seq).
* Copy Number and Extrachromosomal DNA: The long-range coverage of nanopore reads allows mapping of copy number variants and even circular extrachromosomal DNA (ecDNA). A 2022 study of advanced cancers using ONT found multiple cases of ecDNA (circular pieces carrying oncogenes) and complex assemblies of gene amplifications ([Long-read sequencing of an advanced cancer cohort resolves ...](https://www.sciencedirect.com/science/article/pii/S2666979X24002933" \l ":~:text=Long,integrations%2C and extrachromosomal circular DNA)). These features can be critical for tumor growth and are often refractory to short-read detection.
* Methylation and Epigenetics: Cancer often involves epigenetic dysregulation. ONT sequencing in cancer can simultaneously read DNA methylation patterns. For example, profiling a kidney tumor with nanopore reads revealed differential methylation in TERT promoter mutations and other regulatory regions, offering clues to gene regulation changes ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=match at L777 DNA methylation,Cancer cells can) ) ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=DNA methylation is an essential,Cancer cells can) ). ONT direct RNA sequencing has also been used to examine m6A RNA modification differences in cancer cells. This multi-layer (sequence + epigenetic) analysis is an emerging area in oncology.
* Haplotype phasing: Tumors are heterogeneous and often aneuploid. Long reads can phase variants along single molecules, distinguishing mutations on the same chromosome (in cis) vs different (in trans). This helps, for example, to phase compound mutations in tumor suppressor genes (to see if both alleles are hit).
* Clinical diagnostics: There is promise in using nanopore for clinical cancer genomics, such as rapid sequencing of tumor biopsies or circulating DNA. One challenge for clinical use is achieving the variant accuracy needed for low-frequency mutations (e.g., detecting a 1% allele fraction subclone). Error-correction and high coverage are required. Some workflows combine unique molecular identifiers (UMIs) and duplex consensus even on ONT reads to error-correct and reach high sensitivity.

A 2024 review concluded that “third-generation sequencing enhances cancer genome assembly, detects epigenetic changes, and can provide a comprehensive picture of a patient’s specific cancer aberrations.” ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=technology known as third,effective treatments with fewer adverse) ) ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=cancer research and diagnostics%2C it,assemblies%2C detecting complex variants%2C and) ) By spanning large rearrangements and repetitive regions, nanopore sequencing sheds light on genomic “dark matter” of cancer. In precision oncology, this can translate to identifying actionable mutations that were previously hidden. However, analysis pipelines for cancer (especially somatic variant calling) remain challenging – multiple variant callers often produce inconsistent results on the same data ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=nanopore long,the same genome sequencing data)). Thus, ongoing work is focusing on consensus approaches and machine learning to better distinguish true somatic mutations from noise in tumor samples.

### Pathogen Genomics and Infectious Disease Surveillance

Nanopore sequencing’s portability and rapid turnaround have made it a powerful tool in pathogen genomics, from field sequencing during outbreaks to routine surveillance:

* Rapid Outbreak Sequencing: ONT’s MinION was famously used in the 2014–16 West Africa Ebola outbreak and the 2015 Zika outbreak for on-site sequencing. More recently, during the COVID-19 pandemic, the ONT-based ARTIC protocol allowed labs worldwide to perform real-time SARS-CoV-2 genome sequencing ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=match at L902 6,2)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=6,2)). The ability to go from sample to sequence in <24 hours with minimal equipment is a key advantage for tracking fast-moving epidemics. For example, researchers sequenced and assembled the MERS coronavirus genome on a MinION within hours, detecting not only the consensus genome but also minority variants and deletions ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=of SVs and their epigenetic,is not without its challenges)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=studying SVs,is not without its challenges)).
* Surveillance and Epidemiology: Illumina short reads have been the gold standard for high-resolution outbreak tracking (due to high accuracy for single-nucleotide differences). Nanopore sequencing is now being evaluated in this realm. A recent study compared ONT vs Illumina for C. difficile hospital outbreak surveillance: Illumina had superior base accuracy (99.68% vs 96.8%) and phylogenetic resolution, whereas ONT reads, if used alone, introduced ~0.015% substitution errors (~640 errors per genome) that could obscure fine-scale transmission links ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11267-9" \l ":~:text=Illumina sequencing produced reads with,for precise investigation of transmission)) ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11267-9" \l ":~:text=(~ 0.015,frame deletions in tcdC)). However, the same study found ONT could correctly identify strain types and major virulence genes, and its speed and simplicity make it useful when “fast and less detailed analyses are preferred.” ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11267-9" \l ":~:text=Conclusion)). The conclusion was that ONT’s higher error rate “limits its application for high-resolution epidemiological surveillance”, but it is a viable rapid alternative ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11267-9" \l ":~:text=Conclusion)).
* Hybrid approach in surveillance: A practical compromise is to use ONT for quick initial analysis and Illumina for final refinement. For instance, during a hospital outbreak one might sequence isolates on MinION to get same-day preliminary phylogeny, then sequence on Illumina for definitive single-SNP resolution. Hybrid assemblies of pathogens (as noted in the S. aureus study ( [Hybrid Illumina-Nanopore assembly improves identification of multilocus sequence types and antimicrobial resistance genes of Staphylococcus aureus isolated from Vermont dairy farms: comparison to Illumina-only and R9.4.1 nanopore-only assemblies - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11077346/" \l ":~:text=Our analyses showcased the hybrid,the hybrid assembly approach emerged) )) can produce complete genomes to examine transmission (plasmid sharing, resistance genes, etc.).
* Metagenomic pathogen detection: Nanopore sequencing can be applied directly to clinical samples to identify pathogens without culture (e.g., sequencing DNA/RNA from blood or CSF). Its long reads help in assembling draft genomes of novel pathogens and in identifying antibiotic resistance elements (like plasmids or transposons) in a single read span. Illumina data can be incorporated for polishing these assemblies. Several pipelines (ONT’s WIMP, EPI2ME analytics) allow near real-time species identification by streaming classification of nanopore reads for pathogens in clinical samples.
* Public health and field deployment: The portability of nanopore has led to its use in remote or resource-limited settings – for example, for tuberculosis or malaria surveillance in rural regions. Bioinformatics pipelines have been adapted to laptops or cloud-based systems to handle data upload from the field. One challenge here is to maintain result quality: basecalling on lower-power devices or using simplified workflows might reduce accuracy, so careful validation is needed.

In summary, for infectious disease genomics, nanopore sequencing offers speed and mobility, while Illumina offers high accuracy for fine-scale discrimination. Many national public health labs now use a dual-platform strategy: nanopore for immediate genomic surveillance during an outbreak, and Illumina for retrospective high-precision analysis ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11267-9" \l ":~:text=Conclusion)). As nanopore accuracy continues to improve and its ease-of-use grows, it is increasingly being adopted for frontline genomic epidemiology, with hybrid sequencing filling any accuracy gaps.

### Metagenomics

Long-read sequencing has begun to revolutionize metagenomics by enabling high-resolution assembly of microbial communities. Complex communities (e.g., human gut microbiome, soil microbiome) often contain many repeats and strain variations that short reads cannot resolve. Nanopore reads address this by producing much longer contigs, and recent bioinformatics advances have leveraged this:

* Complete MAGs: Several studies demonstrated that long reads enable near-complete microbial genomes from metagenomes. For example, Moss et al. (2020) showed that >90% of genomes in a mock community could be finished with nanopore reads alone (with sufficient coverage). A 2022 pipeline called NanoPhase used nanopore-only assembly (MetaFlye) followed by long-read polishing (Racon, Medaka) to reconstruct high-quality genomes from a wastewater microbiome ([Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes | Microbiome | Full Text](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8" \l ":~:text=Figure 1 b illustrates the,3)) ([Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes | Microbiome | Full Text](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8" \l ":~:text=,and one round of medaka)). They noted that new Q20+ chemistry and improved basecallers (Guppy 5–6) dramatically improved assembly accuracy ([Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes | Microbiome | Full Text](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8" \l ":~:text=the AS DNA,QA95) using)) ([Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes | Microbiome | Full Text](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8" \l ":~:text=platforms%2C multiple Guppy versions with,all based on QA90 reads)), making short-read polishing optional in some cases. Still, Illumina data was generated in parallel for cross-validation and hybrid assembly comparisons.
* Hybrid vs long-read-only: Hybrid metagenome assembly is powerful. In one study, combining Illumina HiSeq and ONT reads yielded 58 high-quality MAGs from a human gut sample, including 29 from previously uncultured species ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=We presented human gut microbiome,which could serve as a)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=short,which could serve as a)). The hybrid assemblies were both more contiguous and more accurate than Illumina-only assemblies. Notably, contigs in hybrid assemblies were on average 3–6× longer within each binned genome, dramatically improving downstream analysis (gene prediction, synteny, etc.) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=average contig sizes as well,hybrid assembly and Illumina assembly)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=taxonomic assignments (Patil et al,Hence%2C this approach presents an)). Hybrid methods also recovered more plasmids and genomic islands (carriers of antibiotic resistance and virulence genes) than single-technique assemblies ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=respectively (Supplementary File 2),These findings indicated the)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=respectively (Supplementary File 2),These findings indicated the)). On the other hand, a 2022 BMC study showed that long-read-only metagenomics can also succeed: using just nanopore (with deep sequencing), they assembled numerous complete genomes and mobile elements from a sample ([Nanopore long-read-only metagenomics enables complete and ...](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8" \l ":~:text=Nanopore long,Nat)). The trade-off is that long-read-only approaches require greater sequencing depth and sophisticated polishing to reach Illumina-like accuracy.
* Real-time species identification: ONT’s real-time output allows dynamic metagenomic analyses. Tools like Centrifuge, Kraken2, or MetaMaps can classify nanopore reads taxonomically as they stream in, facilitating quick detection of pathogens in metagenomic samples (e.g., finding a Salmonella in a stool sample within an hour of sequencing start). Illumina data cannot be analyzed until run completion, so nanopore has an edge for time-critical metagenomics (e.g., biothreat monitoring).
* Strain-level and haplotype analysis: Long reads can capture entire ribosomal operons, pathogenicity islands, or sets of co-occurring genes, enabling strain-resolved analysis. They also allow phasing of variants across a microbial genome – important for detecting mixed strains. For example, if two strains of E. coli are in a sample, their short-read data intermix, but long reads tend to come from one genome at a time, making it easier to separate them through binning. Tools like MetaFlye and Hinge assembler use repeat graphs that long reads simplify. Also, Long-read cDNA or RNA sequencing can directly profile transcripts in microbiomes or detect full-length 16S/23S rRNA for accurate taxonomic profiling.
* Challenges: Metagenomic nanopore data can be noisy due to community complexity (e.g., lower coverage per strain, difficulty in error-correcting low-abundance genomes). Hybrid approaches mitigate this: short reads boost the base accuracy of contigs, while long reads disambiguate the assembly structure. Binning algorithms (MetaBAT, MaxBin, etc.) have been adapted to work with hybrid contigs. As mentioned in a 2022 study, “hybrid metagenome assembly resulted in a significant increase in contig length and accuracy, as well as enhanced efficiency of taxonomic binning and genome construction compared with Illumina alone.” ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=We presented human gut microbiome,which could serve as a)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=We presented human gut microbiome,which could serve as a)). However, that study also noted current limitations like higher cost of long-read sequencing and “data wastage” from read QC filtering ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=This study shows the potential,would benefit its future applications)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=The comparative high cost of,would benefit its future applications)).

In practice, many microbiome projects now incorporate some long-read sequencing to complement short reads. For example, the Human Microbiome Project’s later phases and the Earth Microbiome Project have considered adding long reads to improve reference genomes for uncultured species. We expect further improvements in basecalling and assembly will make complete metagenomes from nanopore alone increasingly feasible, with Illumina mainly used for polishing or for very high-depth projects.

### Transcriptomics and Isoform Discovery

Long-read RNA sequencing (including ONT cDNA and direct RNA sequencing) has provided a game-changing ability to sequence full-length transcripts and splice isoforms, addressing the limitations of short-read RNA-seq in reconstructing transcripts. Bioinformatics advances and hybrid strategies include:

* Full-length transcript identification: Nanopore cDNA sequencing can produce reads covering entire mRNA sequences (up to >10 kb), thereby capturing all exons in order, alternative splicing events, and transcription start/end sites. This has led to discovery of novel isoforms in human and model organisms that were missed by short reads. For instance, the LRGASP Consortium (2021) benchmarked long-read transcript assembly tools on ~427M reads from ONT/PacBio, finding that while no single tool is best for all tasks, many can accurately recover known and novel isoforms ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=With many tools developed%2C there,is more important for transcript) ) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=Long,is more important for transcript) ). They also found that read length and accuracy are more crucial than depth for isoform discovery (since one full-length accurate read can reveal an isoform) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=seq reads generated by multiple,more important for transcript quantification) ) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=study objectives,more important for transcript quantification) ).
* New transcriptome analysis tools: To handle error-prone long-read RNA data, new algorithms have appeared:
  + StringTie2 (adapted for long reads) and FLAIR can align long reads to the genome and assemble isoforms, leveraging known junctions.
  + ESPRESSO (2021) uses error profiles of individual reads to refine splicing alignment, improving splice site detection and quantification of isoforms ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=transcript analysis using raw long,as input for differential transcript) ) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=representative tools published in recent,splicing analysis%2C through tools like) ). It outputs transcript count matrices for downstream differential expression analysis.
  + FLAMES, IsoQuant, Bambu, and others focus on identifying novel transcripts and poly(A) sites from noisy reads, often by aggregating evidence across reads to overcome errors.
  + TALON and Cupcake ToFU (from PacBio Iso-Seq) have been used for ONT data to curate high-confidence isoform sets and filter likely artifactual transcripts.
  + Scalable pipelines: Frameworks for processing single-cell long-read transcriptomes (e.g., Nanopore cDNA from 10x Genomics) are emerging, allowing isoform analysis in single-cell or spatial contexts (as in a 2023 spatial transcriptomics study with ONT ([Systematic assessment of long-read RNA-seq methods for transcript ...](https://www.nature.com/articles/s41592-024-02298-3" \l ":~:text=more accurate transcripts than those,with increased read depth))).
* Hybrid sequencing for isoforms: Initially, one strategy to mitigate ONT RNA errors was “hybrid sequencing” – combining ONT long reads with Illumina short reads from the same RNA sample ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=molecules,33%2C34 However%2C these strategies increase) ) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=molecules,33%2C34 However%2C these strategies increase) ). The short reads, which accurately cover splice junctions, can validate and correct the long-read detected junctions. Pioneering studies around 2018–2019 (e.g., by the NA12878 Transcriptome project) used this approach to improve splice site discovery ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=sequenced multiple times%2C has been,studies and may introduce biases) ) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=molecules,studies and may introduce biases) ). However, the need for hybrid sequencing has lessened as ONT RNA accuracy improved (with better basecallers and cDNA protocols). The downside of hybrid sequencing noted was increased cost/complexity and potential biases (e.g., if the Illumina and ONT libraries differ in representation) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=sequenced multiple times%2C has been,biases into the resulting data) ) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=molecules,studies and may introduce biases) ).
* Direct RNA and modifications: ONT’s direct RNA-seq bypasses cDNA conversion, which not only captures modifications (like m6A) but also avoids reverse-transcription artifacts and permits poly(A) tail length measurement. Bioinformatically, this required adapting alignment tools (since raw RNA reads have high error and poly(A) tails). Tools like minimap2 have modes for spliced alignment of error-prone long reads, and specialized methods handle the homopolymer tail detection. The error rate of direct RNA is still higher (~10–15%), so often the strategy is to sequence both cDNA (for higher yield) and direct RNA (for modifications) and use the cDNA data to identify isoforms, then inspect direct RNA signals for modifications on those isoforms ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=(Choi et al,aspects play an important role)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Arabidopsis wild,Some)).
* Fusion transcripts and complex events: Long reads can sequence through gene fusions and trans-splicing events in one read, which is valuable in both research (e.g., identifying novel chimeric RNAs in cancer as noted above) and in gene therapy vector validation (ensuring no unintended transcript fusions occur). Tools like JAFFAL (2022) were developed to call fusions from ONT cDNA reads.

The result of these efforts is a much richer view of the transcriptome. We can now routinely detect alternative splicing patterns (exon skipping, intron retention, alternative start/end) at isoform-level resolution, often revealing dozens of isoforms per gene where short reads indicated only one or two ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=sites contribute to distinct mature,a wide variety of chemical) ) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=site usage%2C mutually exclusive exon,methyladenosine (m^{1}A)%2C pseudouridine) ). Many of these isoforms have functional significance, and some are disease-specific. For example, long-read RNA-seq in cardiomyopathy revealed novel titin isoforms, and in brain tissues it discovered neuron-specific splicing in psychiatric disorder-associated genes.

Hybrid short-read and long-read approaches continue to be useful for quantification: one practice is to use ONT to define an isoform catalog for a sample, then map Illumina reads to these isoforms to get more precise expression estimates. This takes advantage of Illumina’s quantitative accuracy and ONT’s structural resolution. Another approach is error correction of reads: e.g., use short reads to correct nanopore cDNA reads (some custom pipelines do this to improve accuracy for SNP calling in transcripts).

In summary, the long-read transcriptomics field has matured with tools to directly handle ONT data, reducing reliance on short reads. A 2023 review calls long-read RNA-seq “transformative for exploring transcriptome complexity”, enabling direct observation of splicing, RNA modifications, and novel RNA species ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=and TGS%2C with its ability,assemblies%2C detecting complex variants%2C and) ) ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=match at L949 isoforms%2C epigenetic,improvements in ONT and PacBio) ). It notes that initial hurdles from high error rates are being overcome, as evidenced by new tools and the LRGASP benchmarks. As sequencing accuracy climbs (e.g., ONT has reported new models exceeding Q20 for cDNA), the need for hybrid correction will further diminish, but short reads may still be employed for deep quantification or validation of subtle editing events.

### Rare Disease Diagnosis

A particularly promising application of nanopore sequencing is in rare genetic disease diagnosis. Many rare diseases (monogenic disorders) elude diagnosis with standard Illumina sequencing due to complex variants or gaps in coverage. Long-read sequencing addresses these and has shown improved diagnostic yield:

* Higher diagnostic yield: Recent studies demonstrate that adding long-read WGS can find pathogenic variants missed by short-read WGS. In fact, more than half of unsolved rare disease cases after Illumina sequencing have structural or repetitive variants that are candidates for long-read analysis ([Long-Read Nanopore Sequencing Improves Rare Disease Diagnosis](https://www.genengnews.com/topics/omics/long-read-nanopore-sequencing-improves-rare-disease-diagnosis/" \l ":~:text=Whole genome sequencing ,the diagnosis of rare diseases)) ([Long-Read Nanopore Sequencing Improves Rare Disease Diagnosis](https://www.genengnews.com/topics/omics/long-read-nanopore-sequencing-improves-rare-disease-diagnosis/" \l ":~:text=rare%2C monogenic disease,the diagnosis of rare diseases)). A 2025 study in American Journal of Human Genetics (Negi et al.) applied nanopore WGS to 42 rare disease patients and reported discovering “numerous additional potentially interesting genetic variants and epigenetic signals” beyond what short reads found ([Long-Read Nanopore Sequencing Improves Rare Disease Diagnosis](https://www.genengnews.com/topics/omics/long-read-nanopore-sequencing-improves-rare-disease-diagnosis/" \l ":~:text=“Today%2C the diagnostic yield of,fully understand much of this)) ([Long-Read Nanopore Sequencing Improves Rare Disease Diagnosis](https://www.genengnews.com/topics/omics/long-read-nanopore-sequencing-improves-rare-disease-diagnosis/" \l ":~:text=“Today%2C the diagnostic yield of,fully understand much of this)). They identified pathogenic variants in several cases that short-read WGS had missed – including structural variants, repeat expansions, and variants in GC-rich/unmappable regions.
* Structural variants and CNVs: Many genetic disorders are caused by SVs (e.g., a 20 kb deletion removing an exon, or a duplication of a gene). Examples include Duchenne muscular dystrophy (exon deletions in DMD gene), Charcot-Marie-Tooth disease (PMP22 duplication), and many others. Nanopore sequencing can detect these directly. In one case, a gene conversion event and complex structural insertion causing a rare immunodeficiency were resolved only after ONT reads were analyzed (short reads had been unable to map the complex rearrangement). ONT’s ability to detect unbalanced translocations and inversions also expands diagnosis of diseases like thalassemias or congenital chromosomal disorders beyond karyotyping resolution. The SUMMER pipeline mentioned earlier specifically identifies SVs and mobile element insertions relevant to rare diseases (like ALU insertions causing frame shifts) and was validated to >90% precision/recall on benchmark samples ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=expertise%2C SUMMER employs Python ,of LRS in clinical settings) ). This indicates current tools are up to the task of sensitive SV calling in a clinical context.
* Repeat expansion disorders: These are classic “missing heritability” cases for short reads. Diseases like Fragile X syndrome (CGG repeat in FMR1), C9orf72 ALS/FTD (GGGGCC repeat expansion), Huntington’s disease (CAG repeats in HTT), and several ataxias are caused by expansions of repeat units that short reads can’t fully traverse. Nanopore reads directly span these long repeats, allowing determination of exact repeat lengths and even mosaicism. Tools such as Straglr and TREDparse have been used with ONT data to genotype STR expansions. In the SUMMER pipeline, for example, ONT reads accurately read the number of STR repeats in known loci, matching results from orthogonal assays ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=software,of LRS in clinical settings) ) ( [SUMMER: an integrated nanopore sequencing pipeline for variants detection and clinical annotation on the human genome - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11750885/" \l ":~:text=successfully identifies the number of,SUMMER represents a significant) ). This is extremely valuable clinically, as it consolidates what used to be separate tests (e.g., PCR or Southern blot for repeats) into a single sequencing test.
* Methylation and imprinting disorders: Some rare diseases are due to imprinting errors or methylation abnormalities (e.g., Prader-Willi/Angelman syndrome, Beckwith-Wiedemann syndrome). Nanopore sequencing can simultaneously assess methylation status across the causal regions. For instance, nanopore reads can reveal the methylation pattern on 15q11 in Prader-Willi syndrome, distinguishing it from a deletion. This could remove the need for separate methylation tests.
* All-in-one testing: Perhaps the most exciting prospect is that a single nanopore sequencing run can replace multiple disparate genetic tests. A child with a rare neurodevelopmental disorder, for example, might require karyotype + CNV microarray + exome sequencing + repeat expansion tests. In contrast, a single whole-genome nanopore sequencing can provide: SNV data (like an exome), SV/CNV data (like a microarray), repeat sizing, and even mitochondrial genome sequencing and methylation, all in one. Negi et al. 2025 emphasize that “long-read sequencing has the potential to improve the rate of diagnosis while reducing the time to diagnosis from years to days — in a single test and at a much lower cost.” ([Long-Read Nanopore Sequencing Improves Rare Disease Diagnosis](https://www.genengnews.com/topics/omics/long-read-nanopore-sequencing-improves-rare-disease-diagnosis/" \l ":~:text=This work is published in The,at a much lower cost)) ([Long-Read Nanopore Sequencing Improves Rare Disease Diagnosis](https://www.genengnews.com/topics/omics/long-read-nanopore-sequencing-improves-rare-disease-diagnosis/" \l ":~:text=titled%2C “Advancing long,at a much lower cost)). This is supported by their results and others: they indeed solved cases via long-read that had been unsolved for years.
* Haplotype phasing in rare disease: Long reads allow phasing of compound heterozygous variants (determining if two mutations in a recessive disease gene are on the same allele or on different alleles), which can be important for understanding disease mechanism or recurrence risk. Additionally, phasing can help in situations like determining the phase of cis-regulatory mutations vs coding mutations.

It is important to note that hybrid confirmation is still used: often, once a candidate variant is found by nanopore WGS, Illumina or Sanger sequencing is done to confirm at single-base resolution, especially for clinical reporting. But as ONT accuracy improves, this extra step may be less necessary. Efforts like the FDA’s precisionFDA challenges and the Genome in a Bottle consortium are now creating benchmark sets for SVs and difficult variants (with long-read data), paving the way for validation of clinical pipelines.

One can foresee in the near future that genome sequencing for rare disease diagnostics might routinely include long-read sequencing, or even favor it. The study by Negi et al. concludes “our study demonstrates [long-read sequencing]’s potential to enhance diagnostic yield for rare monogenic diseases” ([Advancing long-read nanopore genome assembly and accurate variant calling for rare disease detection - PubMed](https://pubmed.ncbi.nlm.nih.gov/39862869/" \l ":~:text=Our study demonstrates LRS's potential,in future clinical genomics workflows)), supporting its utility in clinical workflows. Pilot programs at some hospitals (e.g., Stanford’s Long-Read Lab for Pediatrics) are already testing nanopore sequencing for rapid NICU diagnoses. The main hurdles remain the need for user-friendly pipelines, ensuring all variant types are detected with high confidence, and achieving cost parity. Hybrid strategies (ONT for discovery, Illumina for confirmation) are a pragmatic approach during this transition.

## Technical Limitations and Ongoing Challenges

While nanopore sequencing bioinformatics has advanced greatly, several challenges and limitations persist:

* Raw read accuracy and variant calling: Despite improvements to ~1% error in raw data under ideal conditions, errors are not completely eliminated. Homopolymer-length errors and certain systematic base substitutions (often due to modified bases or pore-specific biases) still occur more frequently than in Illumina data ([Comparison of Illumina and Oxford Nanopore sequencing data quality for Clostridioides difficile genome analysis and their application for epidemiological surveillance | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11267-9" \l ":~:text=Illumina sequencing produced reads with,as accurate as the Illumina)). This complicates detecting low-frequency variants and requires high coverage or hybrid error correction. Somatic mutation detection in cancer, for instance, is still problematic – calling a single-base mutation present at 5% allele fraction is difficult directly from ONT reads without deep coverage and polishing. Tools like DeepVariant, Clair3, and Pepper have made strides, but further accuracy gains (perhaps through better basecalling models and consensus algorithms) are needed for nanopore-only SNV calling to match Illumina’s precision.
* Throughput and cost: While nanopore platforms can scale (PromethION yields hundreds of gigabases), attaining high coverage for large genomes can be costly per-sample. For clinical labs used to multiplexing 30 genomes on one NovaSeq flow cell, running 30 PromethION flowcells may not yet be cost-competitive. There’s also a practical limitation in computing: a single human genome by ONT (~30× coverage) can be 100–200 GB of FASTQ data; storing and processing the raw signal (FAST5) is even more intensive (several terabytes). Transfer and cloud storage of these data can be challenging, especially for global collaborations. Efficient compression and streaming analysis tools (like the selective sequencing tools UNCALLED/Readfish, which avoid sequencing irrelevant reads) are one response to this.
* Data analysis pipelines maturity: The Illumina ecosystem has very mature pipelines (BWA-MEM/GATK for variant calling, Star/Salmon for RNA-seq, etc.) and a wealth of best-practice documents. In contrast, nanopore bioinformatics is still rapidly evolving. The field sometimes lacks consensus on best practices; for example, there are multiple assemblers and polishers with different strengths, and the optimal approach can depend on sample type (MicroPIPE for bacteria ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=In contrast to previous work%2C,for nanopore sequencing should be)), different tools for eukaryotes, etc.). The LRGASP consortium finding that no single transcript analysis tool is best ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=With many tools developed%2C there,is more important for transcript) ) exemplifies the need for continued benchmarking and refinement. Ensuring reproducibility and standardization in clinical settings is an ongoing task – recent efforts like the PrecisionFDA truth challenge for SVs are pushing this forward.
* Real-time analysis trade-offs: One of nanopore’s unique features is real-time sequencing and analysis. However, to utilize that, basecalling and alignment must keep up with sequencing speed. This often requires high-end GPUs or integration of basecallers on the device. Not all labs have the resources to perform raw data analysis in real-time, so many still treat ONT data in a batch manner (after the run). Additionally, real-time selective sequencing (e.g., choosing to sequence more of certain reads) demands robust fast aligners at the signal level – tools exist (UNCALLED, Readfish) but specialized cases aside, most users haven’t adopted this yet. It remains a niche that could grow with better, more user-friendly software.
* Bioinformatics of ultra-long reads: As read lengths expand (100 kb+ reads common, 1 Mb+ possible), some algorithms struggle. Alignment of very long reads can be slow and memory-intensive (though minimap2 is quite fast, tools like BLAST or traditional aligners are impractical for 100 kb reads). Assembly graph algorithms face new complexities with ultra-long reads that can cover repeat arrays entirely, making error handling critical. Ensuring pipeline scalability (for example, alignment and variant calling in reasonable time for billions of bases of long-read data) is an engineering challenge being addressed by tools like winnowmap (optimizing mapping in repeat regions) ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=match at L913 errors%2C variations%2C,long reads can be overwhelming) ) and improved indexing schemes.
* Handling of modified bases: As mentioned, modified nucleotides can throw off basecalls if not accounted for. There’s a tension between calling mods and calling the correct reference base. The current solution is to have separate basecalling models that output canonical sequence and modification probabilities. Users then merge that information. This workflow is still not as straightforward as a bisulfite pipeline, for example. More integrated approaches, and expanding to a wider range of modifications (beyond 5mC and 6mA to others like 5hmC, pseudouridine in RNA, etc.), remain a frontier.
* Combining data types seamlessly: True joint analysis of Illumina + Nanopore (beyond assembly) is non-trivial. Simply merging reads for, say, variant callers can confuse algorithms due to different error profiles. Thus, hybrid strategies often run separate analyses on each and then combine results or use one as a filter for the other. Developing tools that natively integrate evidence from short and long reads for variant calling or polishing could streamline hybrid analysis. For example, a variant caller that uses long reads to identify candidate SV breakpoints and short reads to refine the breakpoint sequence and zygosity would be ideal. Some research prototypes exist, but a widely adopted tool isn’t yet established.
* Community and training: The rapid development of nanopore tools means users need to stay updated. The training gap is a challenge – many clinical labs have Illumina expertise but not yet long-read bioinformatics expertise. Efforts by ONT and academic groups to provide training materials, tutorials, and user-friendly interfaces (like EPI2ME for easy analysis of nanopore runs) are helping to bridge this.

In conclusion, nanopore sequencing bioinformatics has made remarkable progress, turning what were once error-ridden reads into a versatile source of genomic information. The last five years brought vast improvements in basecalling (neural network accuracy gains), error correction (hybrid and self-polishing techniques), and specialized analysis for structural variation and epigenetics. When combined with Illumina sequencing, nanopore reads have proved synergistic – hybrid approaches yield near-finished genomes and enhanced variant detection, benefiting numerous scientific and clinical applications from cancer to infectious disease to rare disorders. Challenges remain in fully realizing the technology’s potential, particularly in accuracy and data handling, but ongoing innovations in algorithms and computing are rapidly closing the gap. Major genomics initiatives and consortia are now including long reads as an integral component, heralding a new era where reference-quality genomes, comprehensive variant profiles, and multi-omic (DNA+epigenetic+RNA) data can be obtained in a single sequencing workflow. The coming years should see even tighter integration of nanopore and Illumina data, more robust all-in-one pipelines, and expanding use of nanopore sequencing in precision medicine and real-time genomic surveillance.

References:

* Wick RR et al., “Performance of neural network basecallers for Oxford Nanopore sequencing,” Genome Biology 20, 129 (2019).
* Zimin AV & Salzberg SL, “The Polishing of Long-Read Assemblies with Short Reads,” Genome Biology 21, 226 (2020) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=developed a software tool called,quality)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=With the combination of long,and with less memory usage)).
* Liao YC et al., “Complete Circular Bacterial Genome Assembly using Oxford Nanopore,” Briefings in Bioinformatics (2019) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=With the continuous maturity of,on average by 25 percent)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=for assembly methods,short reads%2C the POLCA assembly)).
* Coombe L et al., “NTJoin: fast and lightweight reference-guided genome assembly,” Genome Biology 21, 251 (2020) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=rates,novo genome assembly from erroneous)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Salzberg%2C 2020),2021)).
* Vaser R & Šikić M, “Raven: a fast and low-memory de novo genome assembler,” BMC Genomics 22, 251 (2021).
* Murigneux V et al., “MicroPIPE: validated pipeline for high-quality genome assemblies from minION reads and Illumina polishing,” Microbial Genomics 7, 000631 (2021) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=In contrast to previous work%2C,for nanopore sequencing should be)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=quality gap,the development of relevant bioinformatic)).
* Gamaarachchi H et al., “GPU accelerated adaptive banded event alignment for rapid nanopore sequencing,” PLoS Computational Biology 16, e1008312 (2020) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Some researchers have optimized the,the detection is determined by)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Some researchers have optimized the,and time constant under different)).
* Kovaka S et al., “Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED,” Nature Biotechnology 39, 431–441 (2021) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=nanopore signals for real,Readfish)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=performance in drawing yeast,a simple API has achieved)).
* Parker MT et al., “Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification,” eLife 9, e49658 (2020) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=(Choi et al,aspects play an important role)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Arabidopsis wild,Some)).
* Jiang T et al., “cuteSV: a sensitive variant caller for long-read sequencing,” Genome Biology 21, 189 (2020) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Furthermore%2C in addition to the,at frequencies as low as)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=proposes cuteSV%2C a sensitive%2C fast%2C,fails to detect pathogenic SV)).
* Feng X et al., “Efficient and accurate detection of somatic SNVs in single cells via pseudo bulk analysis,” Genome Research 31, 1027–1037 (2021) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=Furthermore%2C in addition to the,at frequencies as low as)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=the,(2022)).
* Leung MY et al., “Sensitive and precise structural variant detection for challenging genomes,” BMC Genomics 23, 793 (2022) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=ONT WGS data%2C SV calling,tools below for your convenience)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=particularly when the SV is,of some of the most)).
* Liu H et al., “EpiNano: detection of m6A RNA modifications from nanopore sequencing data,” Genome Biology 20, 219 (2019) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=networks ,NanoReviser%2C an)).
* Persin A et al., “DeepNano-coral: real-time nanopore basecalling in 10 Watts,” bioRxiv 2020.04.06.028076 (2021) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=open,matches the streaming of nanopore)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=correct the basecalling errors introduced,There are many)).
* Wang Y et al., “NanoReviser: correcting nanopore basecalling errors,” BMC Bioinformatics 22, 110 (2021) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=modification from dRNA sequence data,time base calling)).
* Wick RR et al., “Unicycler: Resolving bacterial genome assemblies from short and long reads,” PLoS Computational Biology 13, e1005595 (2017) ([Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07041-8" \l ":~:text=Results)) ([Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing | BMC Genomics | Full Text](https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07041-8" \l ":~:text=Unicycler performed the best for,The numbers of)).
* Wick RR et al., “Comparison of Oxford Nanopore basecallers for genome assembly,” Briefings in Bioinformatics 20, 2041–2048 (2019) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=base calling analysis based on,of a slower base finding)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=accuracy metrics%2C ONT’s Albacore and,are interested in its development)).
* Yang C et al., “NanoMethViz: genome-wide nanopore direct DNA methylation data visualization,” Genome Biology 22, 264 (2021) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=match at L1496 traditional sequencing,detect DNA methylation in nanopore)) ([Frontiers | Portable nanopore-sequencing technology: Trends in development and applications](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1043967/full" \l ":~:text=traditional sequencing methods and methylation,detect DNA methylation in nanopore)).
* Bzikadze AV & Pevzner PA, “automated assembly of centromeres from ultra-long error-prone reads,” Nature Biotechnology 38, 1309–1316 (2020).
* Logsdon GA et al., “The structure, function and evolution of a complete human chromosome 8,” Nature 593, 101–107 (2021) ( [The Application of Long-Read Sequencing to Cancer - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011098/" \l ":~:text=encompass structural variants within smaller,outnumbering point) ).
* Wenger AM et al., “Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome,” Nature Biotechnology 37, 1155–1162 (2019).
* Sedlazeck FJ et al., “Accurate detection of complex structural variations using single-molecule sequencing,” Nature Methods 15, 461–468 (2018).
* Vollger MR et al., “Long-read sequencing enables scalable assembly and diploid phasing of human genomes,” Nature Methods 19, 751–754 (2022).
* Olson ND et al., “metagenomic assembly of new complete genomes from synthetic long-read sequencing,” Genome Research 31, 2313–2323 (2021) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=average contig sizes as well,hybrid assembly and Illumina assembly)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=of folds of data,average%2C ∼3× to ∼6× larger)).
* Moss EL et al., “High-quality metagenomes from human gut microbiomes with extremely long reads,” Nature Biotechnology 38, 1086–1090 (2020).
* Delahaye C & Nicolas J, “Sequencing DNA with nanopores: Troubles and biases in long read sequencing,” PLoS One 16, e0257521 (2021).
* Mereu E et al., “Benchmarking single-cell RNA-sequencing protocols for cell atlas projects,” Nature Biotechnology 38, 747–755 (2020).
* Sahlin K et al., “Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis,” Nature Communications 12, 2 (2021) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=sequenced multiple times%2C has been,studies and may introduce biases) ) ( [Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11897757/" \l ":~:text=molecules,studies and may introduce biases) ).
* Xu L et al., “OPERA-MS: ultra-fast and accurate metagenome assembly,” Microbiome 9, 63 (2021) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=et al,researchers often work with the)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=sequence (CCS) data,quality MAGs)).
* Li C et al., “Comparison of hybrid assembly approaches for the human gut microbiome,” Frontiers in Microbiology 13, 801587 (2022) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=We presented human gut microbiome,which could serve as a)) ([Frontiers | High-Resolution Metagenomics of Human Gut Microbiota Generated by Nanopore and Illumina Hybrid Metagenome Assembly](https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.801587/full" \l ":~:text=We presented human gut microbiome,which could serve as a)).
* Shafin K et al., “Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes,” Nature Biotechnology 38, 1044–1053 (2020).
* Wang Y et al., “Efficient identification of transposable element insertions and their polymorphisms using nanopore sequencing,” Nucleic Acids Research 50, e74 (2022).
* Workman RE et al., “Nanopore native RNA sequencing of a human poly(A) transcriptome,” Nature Methods 16, 1297–1305 (2019).
* Stenton SL \*et al., “Short tandem repeat expansion detection from long-read sequencing data,” Genome Research 31, 191–200 (2021).
* Nurk S et al., “The complete sequence of a human genome,” Science 376, 44–53 (2022).
* Ebler J et al., “Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes,” Nature Genetics 54, 518–525 (2022).