Investigating the performance of Oxford Nanopore long-read sequencing with respect to Illumina microarrays and short-read sequencing

Authors

Renato Santos[1] renato.santos@kcl.ac.uk https://orcid.org/0000-0003-1551-2511

Hyunah Lee[2] hyunah.1.lee@kcl.ac.uk https://orcid.org/0000-0002-3352-1908

Alexander Williams[2] alexander.c.e.williams@outlook.com https://orcid.org/0009-0006-7444-014X

Anastasia Baffour-Kyei[2] anastasia.baffour-kyei@kcl.ac.uk https://orcid.org/0000-0001-7103-4150

Gerome Breen[2] gerome.breen@kcl.ac.uk https://orcid.org/0000-0003-2053-1792

Alfredo Iacoangeli[1,3,4] alfredo.iacoangeli@kcl.ac.uk https://orcid.org/0000-0002-5280-5017

Affiliations

1 Department of Biostatistics & Health Informatics, Institute of Psychiatry Psychology & Neuroscience, King's College London, 16 De Crespigny Park, London SE5 8AB, United Kingdom

2 Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry Psychology & Neuroscience, King's College London, 16 De Crespigny Park, London SE5 8AB, United Kingdom

3 Maurice Wohl Clinical Neuroscience Institute, King's College London, 5 Cutcombe Rd, London SE5 9RX, United Kingdom

4 Perron Institute for Neurological and Translational Science, Ground RR Block QE II Medical Centre Ralph & Patricia Sarich Neuroscience Building, 8 Verdun St, Nedlands WA 6009, Australia

Abstract

Oxford Nanopore Technologies (ONT) long-read sequencing (LRS) has emerged as a promising tool for genomic analysis, but comprehensive comparisons with established platforms across diverse datasets remain limited. In this study, we present a comprehensive comparison of ONT long-read sequencing (LRS) against Illumina short-read sequencing (SRS) and microarray platforms across 14 human genomes.

We performed ONT sequencing using both multiplexed and singleplexed approaches and compared the results with matched Illumina microarray and SRS data. We assessed sequencing quality metrics, variant detection performance for single nucleotide variants (SNVs), insertions/deletions (indels), and structural variants (SVs), while examining the impact of sequencing depth, read length, and multiplexing.

ONT LRS demonstrated competitive performance with SRS for SNV detection, particularly in low complexity regions (F-measure: 0.763 vs 0.770), though with slightly lower performance in high complexity regions (F-measure: 0.954 vs 0.968). For indel detection, LRS showed robust performance in high complexity regions (F-measure: 0.850) which, however, decreased in low complexity regions (F-measure: 0.453). LRS identified 2.82 times more structural variants than SRS and detected variants across a broader size range (8 bp to 129 Mb vs 2 bp to 6 kb). Sequencing depth strongly correlated with variant calling performance across all variant types, with correlation coefficients of 0.80 for SNVs in high complexity regions, 0.84 for SNVs in low complexity regions, and exceeding 0.9 for indels. SV detection in LRS showed strong depth dependence ($r = 0.939$), while SRS SV calls remained stable across depths.

Our findings demonstrate that ONT LRS complements existing sequencing technologies, offering advantages in detecting structural variants and analysing low complexity regions, while maintaining competitive performance in standard variant detection. This study provides practical insights for optimising ONT sequencing strategies and highlights areas for future methodological improvement.

Introduction

Since the completion of the Human Genome Project (1), the field of genomics has continually evolved, with newly developed genotyping and sequencing technologies enabling the study of diverse types of genomic variation. The first published human reference genome assembly in April 2003 (2), provided the foundation for various variant calling methods, involving mapping sequenced reads to a reference and identifying nucleotides that differ between the reads and the reference. This advancement has facilitated the study of genomic variations associated with disease and the development of genetic diagnostics in clinical settings (3). Genomic variants can be characterised using several different methods, including DNA microarrays, short-read sequencing, and long-read sequencing, each with its own advantages and disadvantages.

SNP arrays, a type of DNA microarray, have emerged as a powerful and cost-effective tool for genotyping one of the most common types of variation, single nucleotide variants (SNVs) (4). These arrays consist of high-density oligonucleotide probes attached to a solid surface, designed to hybridise with specific DNA sequences containing SNVs of interest. The primary advantages of microarrays include their cost-effectiveness for large-scale studies, high-throughput capabilities, and ability to detect SNVs across the genome with high accuracy (5), leading to their widespread adoption in large-scale population studies and genome-wide association studies (GWAS) (6).

However, microarrays are limited to detecting known variants for which probes have been designed, making them less suitable for discovering novel or rare variants, or structural variation, such as balanced chromosomal rearrangements (5). Despite these limitations, microarrays remain a valuable tool in genomics research, particularly for large-scale genotyping projects where cost and high-resolution throughput are the primary considerations.

Short-read sequencing (SRS) has become a cornerstone technology in genomics research and clinical diagnostics, with widespread adoption due to its high accuracy, throughput, and cost-effectiveness (7). The typical error rates of approximately 0.1% (Q30) (8), allow for confident detection SNVs and small insertions/deletions (indels) in high-confidence regions of the genome, where reads can be accurately mapped (9). Furthermore, recent developments in versatile and scalable bioinformatics analysis pipelines have made it possible to efficiently process and analyse SRS data (10-12). These key advantages have made SRS the go-to technology for both large-scale whole genome studies (13, 14) and clinical applications (15, 16).

Conversely, the relatively short read lengths (typically around 100–250 bp) pose challenges when performing variant calling in repetitive genomic regions, such as segmental duplications (17), transposable elements (TEs) (18), short tandem repeats (STRs) (19), homopolymers (20), and telomeres (9). In these regions, short reads map to

the reference ambiguously, or fail to map entirely, leading to incomplete coverage and potential misidentification of variants. Although various computational approaches to improve variant calling in challenging regions have been developed to address these limitations, including methods for local assembly, split-read analysis, and paired-end mapping strategies (21), many dark regions of the genome still remain inaccessible or poorly characterised using SRS alone (22).

In recent years, long-read sequencing (LRS) technologies were developed as powerful complements to short-read sequencing, addressing many of the limitations associated with short reads in repetitive and complex genomic regions. Among these, Oxford Nanopore Technologies (ONT) has gained significant attention due to its unique advantages and potential applications. Introduced commercially in 2014, ONT utilises a fundamentally different approach to sequencing, by passing a single strand of DNA through a protein nanopore embedded in an electrically resistant membrane (23). As the DNA molecule traverses the pore, it causes characteristic disruptions in the ionic current flowing through the pore, which are then measured and translated into base calls (24). This approach allows for the generation of exceptionally long reads, typically ranging from 10 to 100 kb (25), with maximum length reads demonstrated to exceed 2 Mb in length (26). These can span entire repetitive regions, resolve complex structural variants, and improve de novo genome assemblies (22), and allow the detection of epigenetic marks, such as methylation, without the requirement of additional library preparation methods (27).

While initial iterations of ONT LRS had error rates of 38% for R6 nanopore chemistry, based on read-to-reference alignment metrics (28), recent advancements in base-calling algorithms, improved pore chemistry, and sequencing protocols have substantially improved the accuracy of ONT sequencing. Indeed, the latest R10.4 nanopore chemistry has significantly reduced error rates, achieving a modal read accuracy of over 99% (Q20) (29) and an average accuracy of 96.8% (Q15) (30) on read-to-reference alignment. In addition, the development of highly accurate consensus approaches, such as duplex sequencing, has further improved accuracy, albeit at the cost of reduced throughput (31). The ability to generate long reads at increasingly competitive costs and accuracy positions ONT as a valuable tool for comprehensive genome analysis, particularly in regions of low-complexity where short-read technologies fall short.

While previous LRS benchmarking efforts have largely relied on well-characterised reference samples from the Genome in a Bottle (GIAB) consortium (32-35), there remains a need for comprehensive comparisons using diverse sample datasets. To address this gap, we present a new benchmark of LRS against other technologies using a cohort of 14 in-house sequenced genomes, while also thoroughly characterising this dataset. Our design utilises a multi-platform strategy, combining data from Illumina microarrays, Illumina SRS, and ONT LRS thereby allowing for a thorough evaluation of ONT's

capabilities in calling SNVs, indels, and structural variants (SVs). Thus, by comparing LRS results to those obtained from well-established technologies, we can gain insights into the strengths and limitations of each platform across different genomic contexts.

Results

Sequencing Quality Control

To evaluate the performance of our LRS dataset, we first conducted an analysis of the sequencing data quality and characteristics. We examined several key metrics, including sequencing yields, read length and quality distributions, alignment quality, sequencing depth, and flowcell and barcoding performance, to identify potential factors that may influence variant detection accuracy (**Figure 1**).

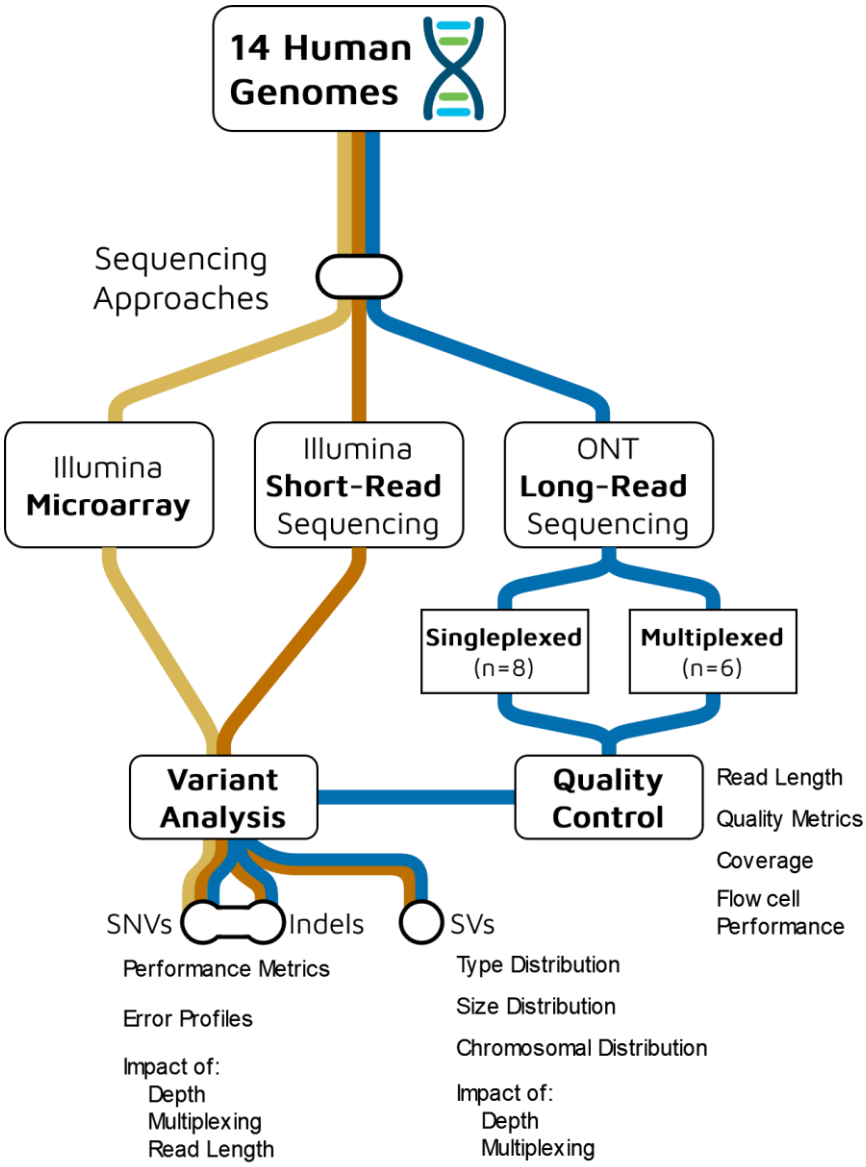

Figure 1 – Experimental design and analysis workflow for comparing Oxford Nanopore Technologies (ONT) long-read sequencing with Illumina short-read sequencing and microarray platforms across 14 human genomes.

The yellow, orange, and blue paths indicate the different sequencing technologies, microarray, short-read sequencing and long-read sequencing, respectively, and their respective analytical routes.

Sequencing Yields

Our analysis of multiplexed (two barcoded samples per flowcell) and singleplexed (one sample per flowcell) nanopore sequencing runs revealed significant differences in yield and read length characteristics between the two approaches.

In terms of sequencing yield, singleplexed samples consistently produced a higher number of reads and bases compared to multiplexed samples (**Figure 2A**, **B**). The mean number of reads for singleplexed samples was 9,852,491, which is 119.41% higher than the mean of 4,490,539 reads for multiplexed samples. Despite only producing approximately 20% more reads per flowcell, the mean number of bases for singleplexed samples (63,738,789,433) was 100.15% higher than that of multiplexed samples (31,845,929,048).

Interestingly, the read length distributions showed a different pattern (**Figure 2C**, **D**). While the mean read length for multiplexed samples (7,091.78 bp) was 8.78% higher than that of singleplexed samples (6,469.31 bp), the median read length for singleplexed samples (5,341 bp) was 26.44% higher than for multiplexed samples (4,224 bp).

The distribution of read lengths across all samples (**Figure 2E**) reveals that while both approaches produced reads spanning a wide range of lengths, there were notable differences in the distribution patterns. Notably, all singleplexed samples exhibited a clear bimodal distribution of read lengths, characterised by two distinct peaks: one at shorter read lengths (approximately 400 bp) and another at longer read lengths (approximately 10,000 bp). In contrast, the multiplexed samples showed more variable distribution patterns. While some multiplexed samples displayed a bimodal distribution similar to the singleplexed samples, although with less pronounced peaks, other multiplexed samples exhibited only a single peak, typically at shorter read lengths (around 1,000-5,000 bp).

This difference in distribution patterns suggests that singleplexing consistently produces two distinct populations of reads: a shorter fraction and a longer fraction. Multiplexing, on the other hand, appears to be less consistent in generating the longer read fraction, with some samples failing to produce the second peak at higher read lengths. This might explain why the larger number of reads per flowcell produced in singleplex did not translate into a larger number of genotyped bases.

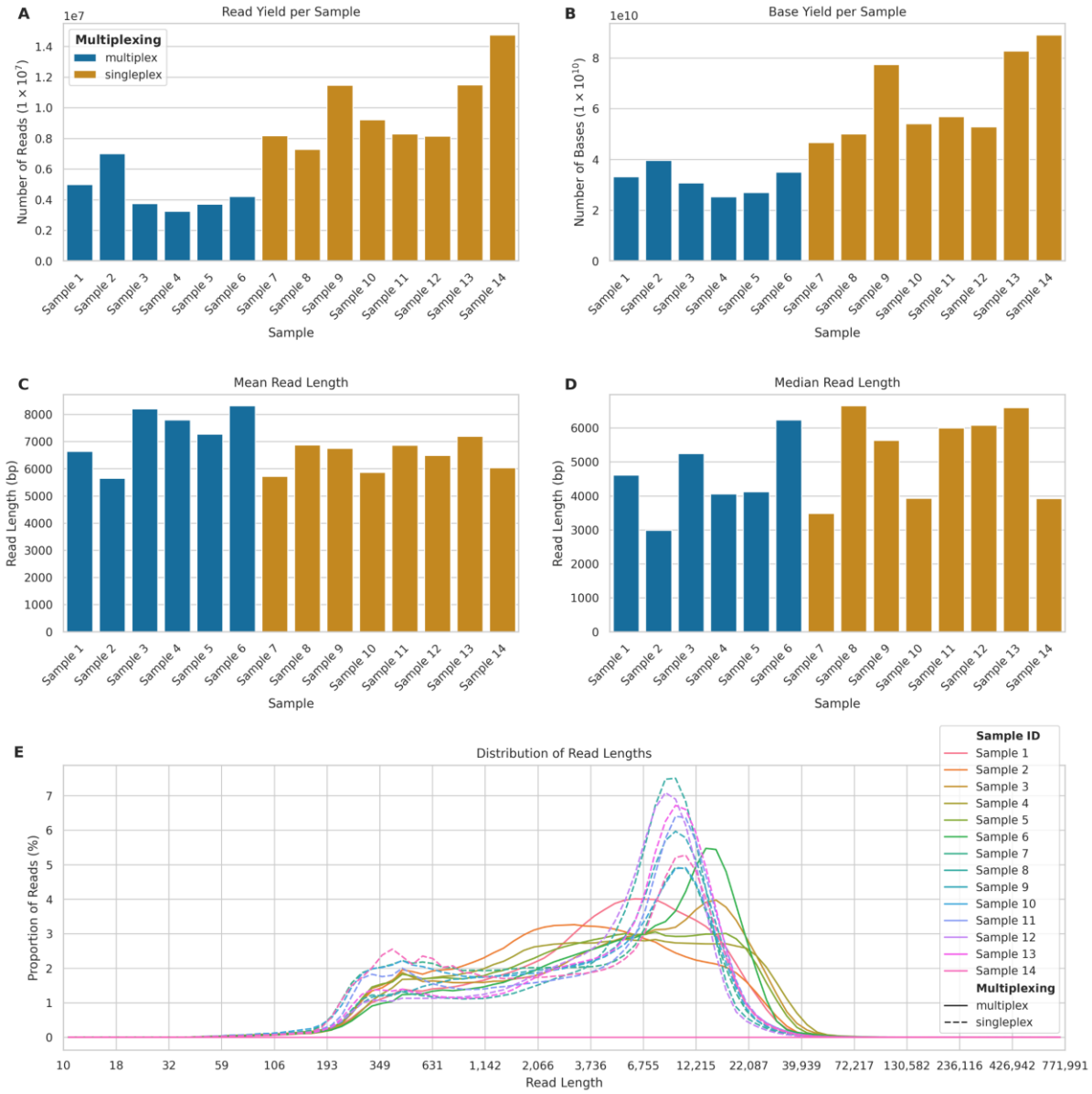

Figure 2 – Comparison of read yields and lengths between multiplexed and singleplexed ONT sequencing runs.

(A) Number of reads per sample. Each bar represents the total number of reads obtained for an individual sample. Blue bars indicate multiplexed samples (Samples 1-6), while orange bars represent singleplexed samples (Samples 7-14). The y-axis shows the number of reads in millions (1×10^7).

(B) Number of bases per sample. Blue bars indicate multiplexed samples and orange bars represent singleplexed samples. The y-axis displays the number of bases in billions (1×10^{10}).

(C) Mean read length per sample. Blue bars represent multiplexed samples, and orange bars show singleplexed samples. The y-axis indicates the mean read length in base pairs (bp).

(D) Median read length per sample. Blue bars are for multiplexed samples, and orange bars represent singleplexed samples. The y-axis shows the median read length in base pairs (bp).

(E) Distribution of read lengths across all samples. Each line represents a different sample, with solid lines indicating multiplexed samples and dashed lines indicating singleplexed samples. The x-axis shows read length for 100 logarithmically spaced bins between 10 bp and 729,335 bp (the maximum read length in the dataset), while the y-axis shows the proportion of reads for each bin, as a percentage of the total number of reads.

Read and Alignment Quality

Both read quality and alignment metrics showed minimal differences between multiplexed and singleplexed samples. Base quality distribution (**Figure 3A**) showed a similar pattern across all samples, with peaks at quality scores 17-20. The mean base quality for singleplexed samples (18.03) was marginally higher than that of multiplexed samples (17.96), representing a small 0.43% increase. Similarly, the median base quality for singleplexed samples (18.14) was slightly higher than multiplexed samples (18.00), a 0.75% increase.

The mapping quality distribution (**Figure 3B**) demonstrated that the vast majority of reads across all samples mapped with a quality score of 60. This pattern was nearly identical between singleplexed and multiplexed samples, with mean mapping qualities of 56.99 and 56.98, respectively, representing a negligible 0.01% difference. The median mapping quality was 60 for both approaches, indicating no practical difference in mapping performance.

Both multiplexed and singleplexed samples showed a similar trend on the proportion of reads exceeding various quality score thresholds (Q5, Q7, Q10, Q12, and Q15) (**Figure 3C**, **D**). While the percentage of reads decreased as the quality threshold increased, at the highest threshold of Q15 all samples retained over 70% of their reads, suggesting multiplexing can be employed without significantly compromising data quality.

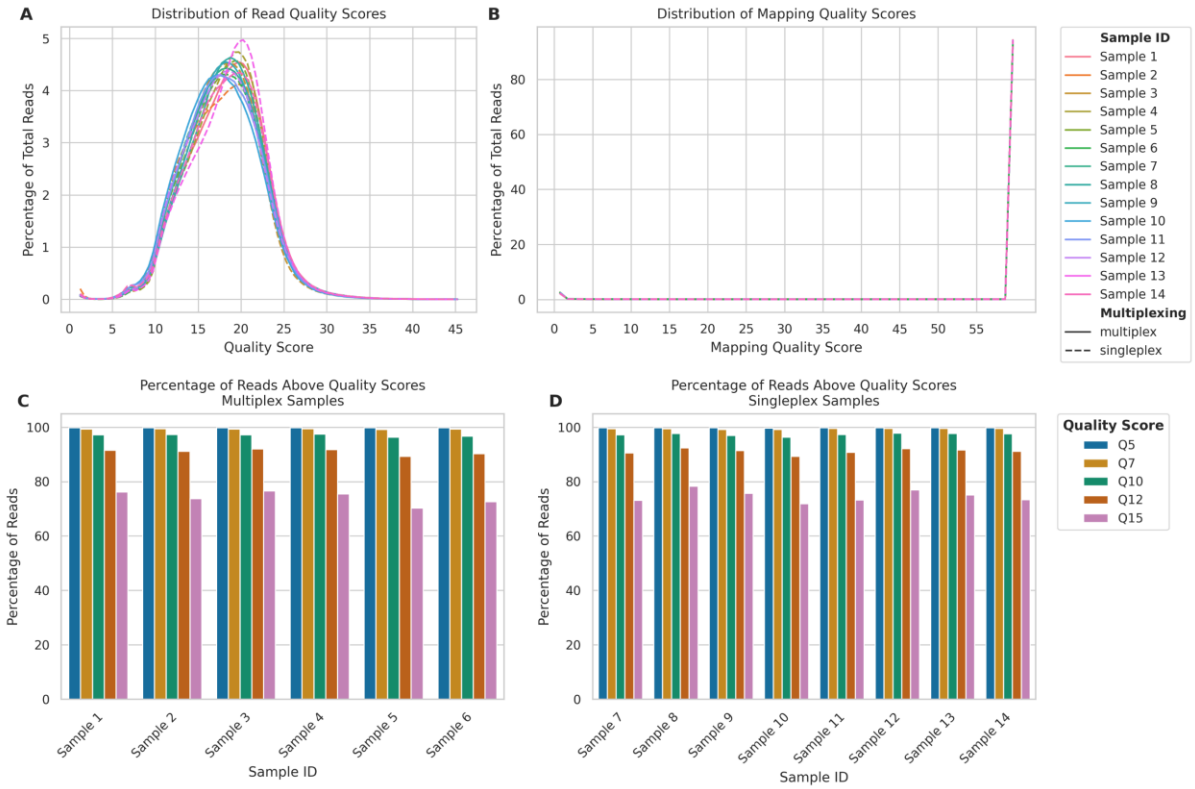

Figure 3 – Comparison of sequencing quality metrics between multiplexed and singleplexed ONT sequencing runs.

(A) Distribution of read quality scores. Each line represents an individual sample, with solid lines indicating multiplexed samples (Samples 1-6) and dashed lines indicating singleplexed samples (Samples 7-14). The x-axis shows the quality score calculated by the Dorado basecaller. The y-axis displays the percentage of total reads for each 0.5 quality score bin.

(B) Distribution of mapping quality scores. The x-axis shows the mapping quality score calculated by minimap2. The y-axis represents the percentage of total reads for each mapping quality score bin.

(C) Percentage of reads above quality score thresholds for multiplexed samples. Each group of five bars represents one sample (Samples 1-6). Colours indicate different quality thresholds: Q5 (blue), Q7 (orange), Q10 (green), Q12 (red), and Q15 (purple). The y-axis shows the percentage of reads above each quality threshold.

(D) Percentage of reads above quality score thresholds for singleplexed samples. Bar groups represent Samples 7-14, with colours indicating quality thresholds as in panel C. The y-axis shows the percentage of reads above each quality threshold.

Depth of Coverage

Focusing on sequencing depth of coverage per chromosome and across the whole genome, we observed differences between singleplexed and multiplexed samples in line with our expectations. Firstly, per-chromosome sequencing depth (**Figure 4A**) showed consistent patterns across all chromosomes, with singleplexed samples exhibiting higher depth than multiplexed samples. The mean depth for singleplexed samples (19.02X) was 101.33% higher than multiplexed samples (9.45X), even though singleplexed samples demonstrated greater variability in depth (SD: 5.98) compared to multiplexed samples (SD: 2.33).

The mean whole genome depth (**Figure 4B**) mirrored the per-chromosome trends. Singleplexed samples achieved a mean depth of 19.68X, while multiplexed samples reached 9.76X, indicating a 101.75% increase in the mean depth of singleplexed samples. The median depths were 17.05X and 9.82X for singleplexed and multiplexed samples, respectively, showing a 73.54% increase.

The consistent depth patterns across chromosomes show that ONT LRS maintains relatively uniform coverage across the genome in both settings, with expected drops in depth for sex chromosomes.

Flowcell and Barcoding Quality

We analysed the initial nanopore availability in the flowcells we used for sequencing to investigate its impact on sequencing performance and depth of coverage obtained. There were distinct differences in the number of pores available at the start of sequencing (**Figure 4C**), with multiplexed flowcells consistently having a higher number of starting pores (mean: 8,116.00; SD: 100.34) compared to singleplexed flowcells (mean: 7,008.75; SD: 1,190.38), which exhibited greater variability in starting pore numbers. This occurred because the best quality flowcells were selected for multiplexing to yield the best results possible for this approach.

The relationship between initial pore count and total mean whole genome depth (**Figure 4D**) revealed a strong positive correlation (Pearson's R-value: 0.64, p-value: 0.035). The linear regression analysis yielded a slope of 0.0025 and an intercept of 1.61, which indicates that, for every additional 1,000 starting pores, we can expect an increase of approximately 2.5× in mean whole genome depth. This suggests that initial pore availability is a significant factor in determining sequencing depth yields.

Examining barcoded read distribution in multiplexed samples allowed us to assess the efficiency of the barcoding process and explain the differences in depth yields between multiplexed and singleplexed flowcells with similar number of nanopore availability (**Figure 4E**). Across the three multiplexed flowcells, we observed some variation in the number of reads assigned to the barcodes of interest and unclassified reads. To quantify this variation, we calculated the Coefficient of Variation (CV) for each flowcell: Multiplex

Flowcell 1 showed moderate variation (CV: 14.64%), while Multiplex Flowcell 2 exhibited the highest variability (CV: 34.65%); in contrast, Multiplex Flowcell 3 demonstrated the most consistent read distribution (CV: 7.82%). Importantly, an average of 19.03% of reads remained unclassified across all multiplexed flowcells. This level of variation within flowcells, and substantial proportion of unclassified reads indicates there is room for efficiency improvement in the barcoding process and, if not optimised, multiplexing might result in a lower total number of genotyped bases per flowcell.

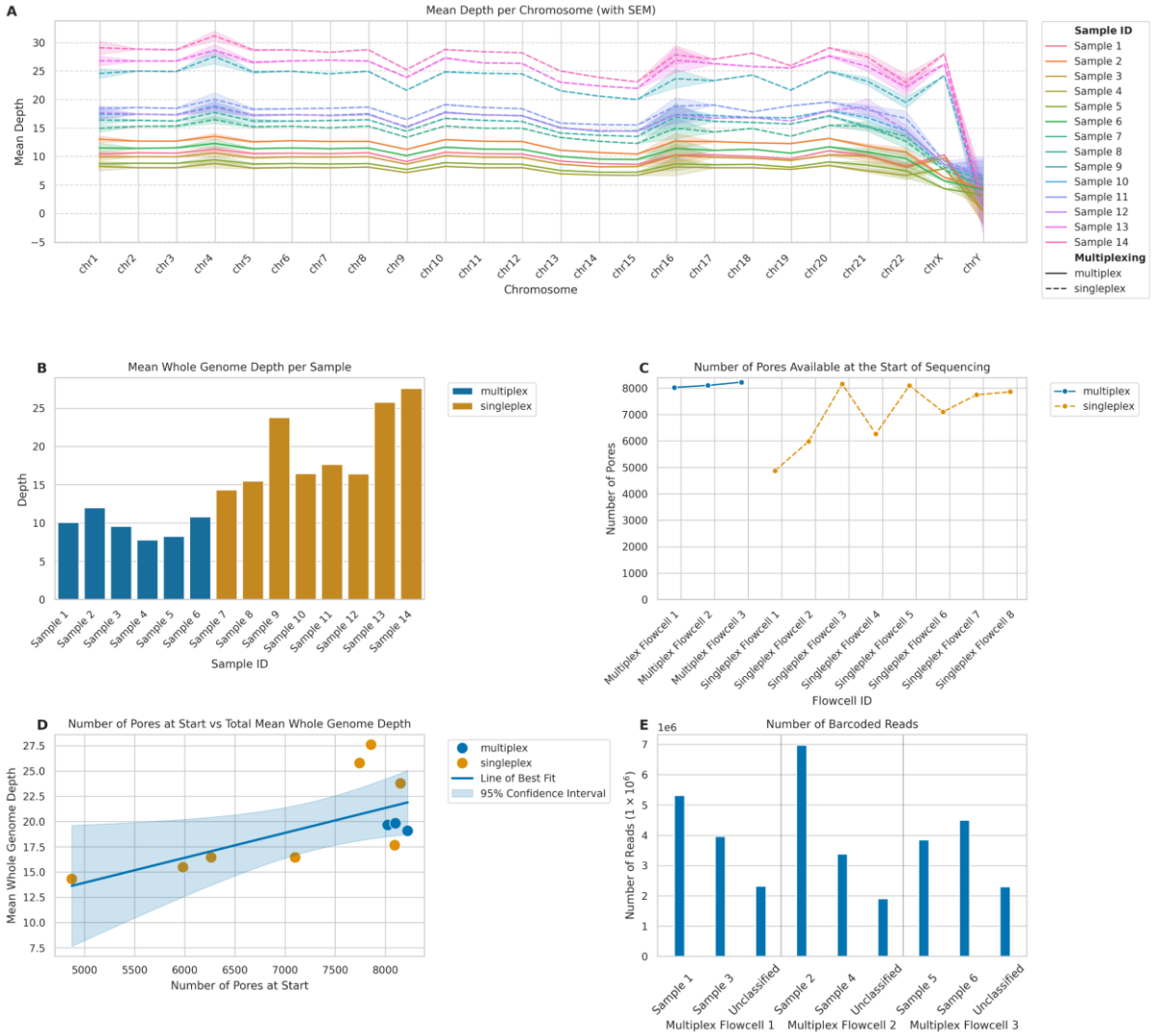

Figure 4 – Sequencing depth, flowcell pore availability, and barcoding analysis for multiplexed and singleplexed ONT sequencing runs.

(A) Mean sequencing depth per chromosome. Lines represent individual samples, with solid lines indicating multiplexed samples and dashed lines representing singleplexed samples. Shaded areas depict the standard error of the mean (SEM) for the depth across each chromosome. Chromosomes are arranged along the x-axis, with depth on the y-axis.

(B) Mean whole genome sequencing depth per sample. Bar plot showing the average sequencing depth across the entire genome for each sample. Blue bars represent multiplexed samples, while orange bars indicate singleplexed samples. Sample IDs are listed along the x-axis, with depth on the y-axis.

(C) Number of pores available at the start of sequencing for each flowcell. The line plot shows the initial pore count for multiplexed (blue) and singleplexed (orange) flowcells. Flowcell IDs are arranged along the x-axis, with the number of pores on the y-axis.

(D) Scatter plot of the number of pores at the start of sequencing versus total mean whole genome depth. Each point represents a flowcell, with colours indicating multiplexing status (blue for multiplexed, orange for singleplexed). The solid line represents the linear regression fit, with the shaded area showing the 95% confidence interval.

(E) Bar plot showing the number of barcoded reads for each sample in multiplexed flowcells. Each group of bars represents a multiplexed flowcell, with individual bars showing read counts for different samples and unclassified reads. The y-axis displays the number of reads, while sample IDs and flowcell information are shown on the x-axis.

Single Nucleotide Variants Benchmark

We compared the performance of the ONT and Illumina sequencing platforms for SNV detection with the VCFeval tool of RTG tools (36, 37). Given their high accuracy in genotyping SNVs, we used the Illumina microarray datasets as the gold standard, stratified by high complexity regions (excluding the Tandem Repeats and Homopolymers regions of GIAB v3.4 (32)) and low complexity genomic regions (only including the aforementioned sites). In addition, only the sites present in the SNP arrays were retained in SRS and LRS data for comparison. For ONT, we analysed 9,553,057 variants in high complexity regions and 490,420 variants in low complexity regions. For Illumina, we analysed 9,467,531 variants in high complexity regions and 482,390 variants in low complexity regions. **Figure 5A**, **B** shows the precision, sensitivity, and F-measure for both technologies in high- and low-complexity regions of the genome.

In high complexity regions, SRS maintains an advantage over LRS technology, with statistically significant differences ($p < 0.01$) in mean precision (SRS: 0.963, SD = 0.004; LRS: 0.953, SD = 0.010), sensitivity (SRS: 0.972, SD = 0.001; LRS: 0.955, SD = 0.018), and F-measure (SRS: 0.968, SD = 0.002; LRS: 0.954, SD = 0.014) compared to LRS. However, the performance slightly shifted in low complexity regions. While SRS still exhibited higher mean precision (SRS: 0.799, SD = 0.005; LRS: 0.783, SD = 0.013, $p < 0.001$), the gap in performance narrowed considerably for mean sensitivity (SRS: 0.743, SD = 0.002; LRS: 0.745, SD = 0.017) and F-measure (SRS: 0.770, SD = 0.003; LRS: 0.763, SD = 0.015), with no statistically significant difference ($p > 0.05$) between platforms in the latter two metrics. On the whole, the performance of both technologies degrades in low complexity regions, but to different degrees, with a more pronounced performance drop for SRS.

To further elucidate the performance of each technology, we conducted an in-depth analysis of SNV error rates, stratified by complexity and error type (**Figure 5C**-**F**), revealing distinct error profiles for SRS and LRS across different SNV types. In high complexity regions, LRS exhibited significantly higher error rates across all SNV types for both false positives (FP) and false negatives (FN) ($p < 0.01$ for all comparisons, **Figure 5C**, **D**). The most pronounced differences were observed in transitions (A>G, G>A, C>T, and T>C). The error profile substantially shifted in low complexity regions (**Figure 5E**, **F**). While for FPs, LRS maintained significantly higher error rates across all SNV types ($p < 0.001$), with the largest differences seen again in transitions, for FNs LRS demonstrated significantly lower FN rates only for A>T transversions ($p < 0.001$), while most other SNV types showed no significant differences between technologies.

While SRS generally maintains an advantage in high complexity regions and in minimising false positives, LRS shows competitive performance in detecting certain SNV types within low complexity regions, particularly in terms of false negative rates. The observed differences in error profiles between the two technologies likely stem from their fundamental sequencing approaches. SRS's short-read methodology may struggle with

repetitive regions due to mapping ambiguities, while LRS's long-read approach can span these regions but introduces more base-calling errors, particularly for transitions.

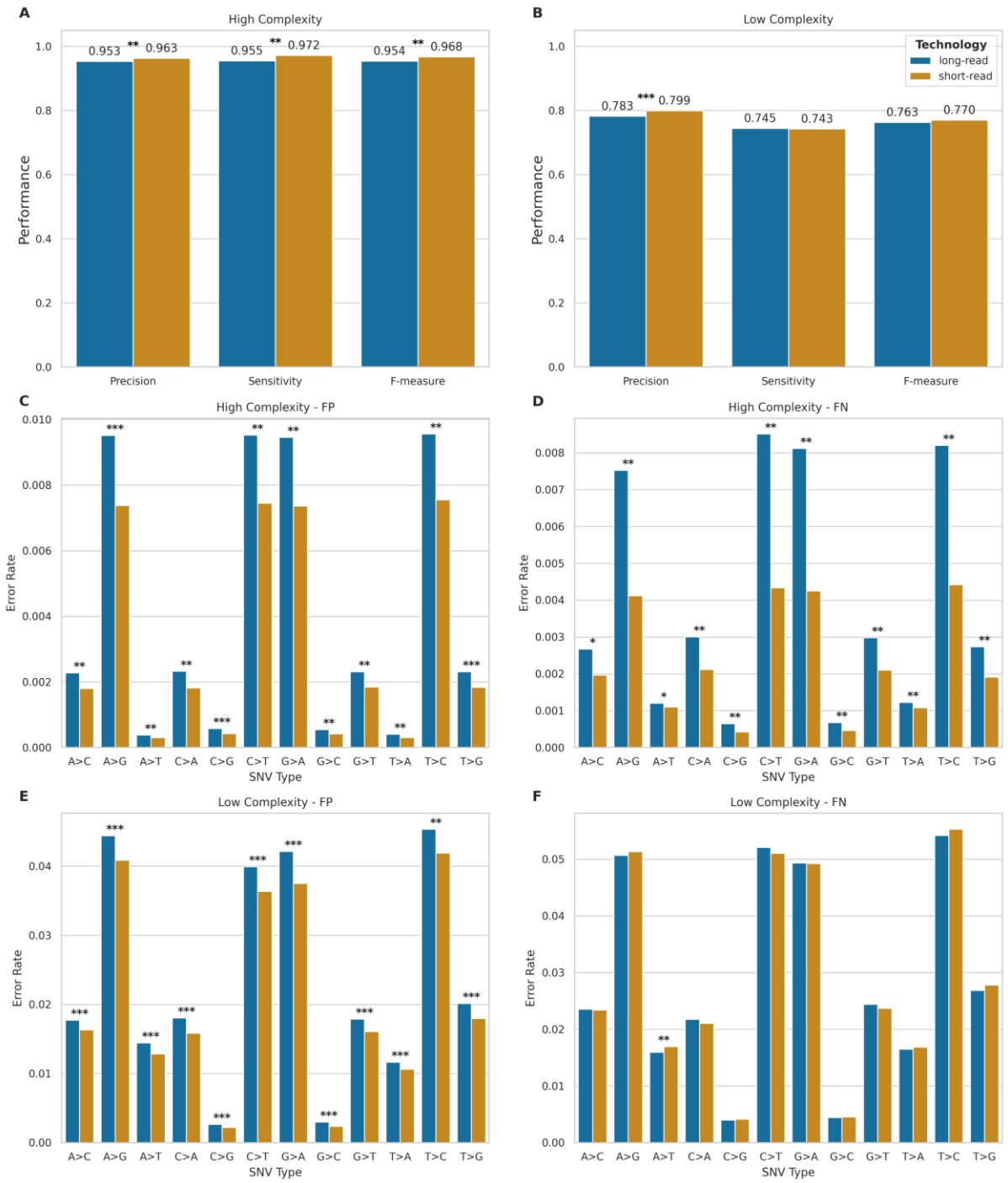

Figure 5 – Performance comparison of Oxford Nanopore (long-read) and Illumina (short-read) sequencing platforms for SNV detection.

(A, B) Performance metrics in high complexity (A; ONT: n=9,553,043; Illumina: n=9,467,517) and low complexity (B; ONT: n=490,420; Illumina: n=482,390) genomic regions. Bars represent precision, sensitivity, and F-measure for long-read (blue) and short-read (orange) sequencing technologies. The y-axis shows the performance score from 0 to 1. Asterisks indicate statistically significant differences between sequencing platforms based on paired t-tests with FDR correction (**$p < 0.01$, ***$p < 0.001$).

(C-F) Error rates for different SNV types in high complexity (C, D) and low complexity (E, F) regions, showing false positives (C, E) and false negatives (D, F). The x-axis shows different SNV types, while the y-axis represents the error rate. Blue bars indicate long-read, and orange bars indicate short-read sequencing. Asterisks indicate statistically significant differences between sequencing platforms based on paired t-tests with FDR correction (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

SNV types on the x-axis of panels C-F represent specific nucleotide changes (e.g., A>C represents adenine to cytosine substitution).

Indels Benchmark

To evaluate the performance of LRS for indel detection using VCFeval, we utilised the SRS dataset as the gold-standard, as SRS has demonstrated superior performance in indel detection with F1 scores of up to 0.99 for state-of-the-art variant calling pipelines (38).

Our results showed a marked disparity in performance between high complexity and low complexity genomic regions (**Figure 6A**, **B**). On one hand, in high complexity areas, LRS exhibited robust performance with a mean precision of 0.807 (SD = 0.054), sensitivity of 0.899 (SD = 0.039), and F-measure of 0.850 (SD = 0.048). Conversely, in low complexity regions, we observed a significant decline in performance metrics, with mean precision, sensitivity, and F-measure dropping to 0.517 (SD = 0.047), 0.403 (SD = 0.051), and 0.453 (SD = 0.050), respectively.

We further analysed the size distribution of insertions and deletions detected by both LRS and SRS platforms in high and low complexity genomic regions (**Figure 6C**-**F**), to ascertain potential biases in the detection of indels of various sizes. We classified the variants into insertions and deletions, within their respective complexity regions. Indeed, both technologies performed comparably in detecting small indels within high complexity genomic regions (**Figure 6C**, **E**), with similar size distributions peaking at 1-2 bp and rapidly decreasing for larger sizes.

To quantify these observations, we calculated summary statistics for indel sizes detected by each platform. The Kolmogorov-Smirnov test revealed minimal differences between platforms in these regions (KS statistic: 0.0060 for insertions, 0.0103 for deletions). However, in low complexity regions (**Figure 6D**, **F**) LRS demonstrated a higher propensity for detecting larger indels, particularly around 8-12 bp. This disparity was more pronounced, with moderate effect sizes (KS statistic: 0.1095 for insertions, 0.0759 for deletions). All differences were statistically significant (p-values $< 2.1\times10^{-25}$ after FDR correction).

In high complexity regions, both technologies showed similar mean indel sizes (LRS: 3.17 bp; SD = 4.85 for insertions, 3.01 bp; SD = 4.12 for deletions; SRS: 3.11 bp; SD = 4.69 for insertions, 3.02 bp; SD = 4.08 for deletions). However, in low complexity regions, LRS detected significantly larger indels on average (mean sizes 4.93 bp; SD = 7.01 for insertions, 4.70 bp; SD = 6.81 for deletions) compared to SRS (3.63 bp; SD = 5.53 for insertions, 3.80 bp; SD = 5.72 for deletions).

To further investigate the performance characteristics of LRS across different indel lengths, we stratified the analysis by indel size categories and complexity regions (**Figure 6G** - I). In high complexity regions, LRS demonstrated robust performance for smaller indels (1-5bp), which represented the majority of variants (89.3%), achieving high precision (median=0.838, IQR=0.061), sensitivity (median=0.913, IQR=0.046), and F-measure (median=0.874, IQR=0.058). Performance remained stable for medium-sized

indels (6-20bp), with only modest decreases in precision (median=0.781-0.800) and sensitivity (median=0.893-0.927). However, there was a notable decline in performance for larger indels (21-50bp), particularly in precision (median=0.625, IQR=0.179) while maintaining relatively high sensitivity (median=0.868, IQR=0.100).

In low complexity regions, we observed substantially lower performance across all size categories. For the predominant 1-5bp small indels (81.4% of variants), precision (median=0.566, IQR=0.081) and sensitivity (median=0.403, IQR=0.169) were markedly reduced compared to high complexity regions. Performance further deteriorated with increasing indel size, with the largest indels (21-50bp) showing the lowest precision (median=0.265, IQR=0.090), maintaining moderate sensitivity (median=0.541, IQR=0.085). Regression analysis revealed strong relationships between indel size and performance metrics in high complexity regions (R^2 = 0.838, 0.695, and 0.912 for precision, sensitivity, and F-measure respectively), while relationships were weaker or absent in low complexity regions (R^2 = 0.770, 0.000, and 0.532 respectively). This confirms that indel size significantly impacts variant calling accuracy, with the effect being more pronounced in high complexity regions.

This size-dependent performance pattern indicates that while LRS technology maintains reliable detection of small to medium-sized indels in high complexity regions, its accuracy is substantially compromised in low complexity regions regardless of indel size. The particularly poor performance for larger indels in low complexity regions highlights the ongoing challenges in accurately sequencing and analysing repetitive or structurally complex genomic areas. However, when considering the inherent limitations of our benchmarking approach, it is important to interpret these results with caution. While we used Illumina SRS as the gold standard for this comparison, it is important to note that SRS itself has limitations, particularly in sequencing repetitive regions. The increasing error rates with increasing indel size could partially reflect the limitations of SRS in detecting larger indels, especially in repetitive regions. Indeed, some of the false positives in LRS, particularly for larger indels in low complexity regions, may actually represent true variants that were missed by SRS.

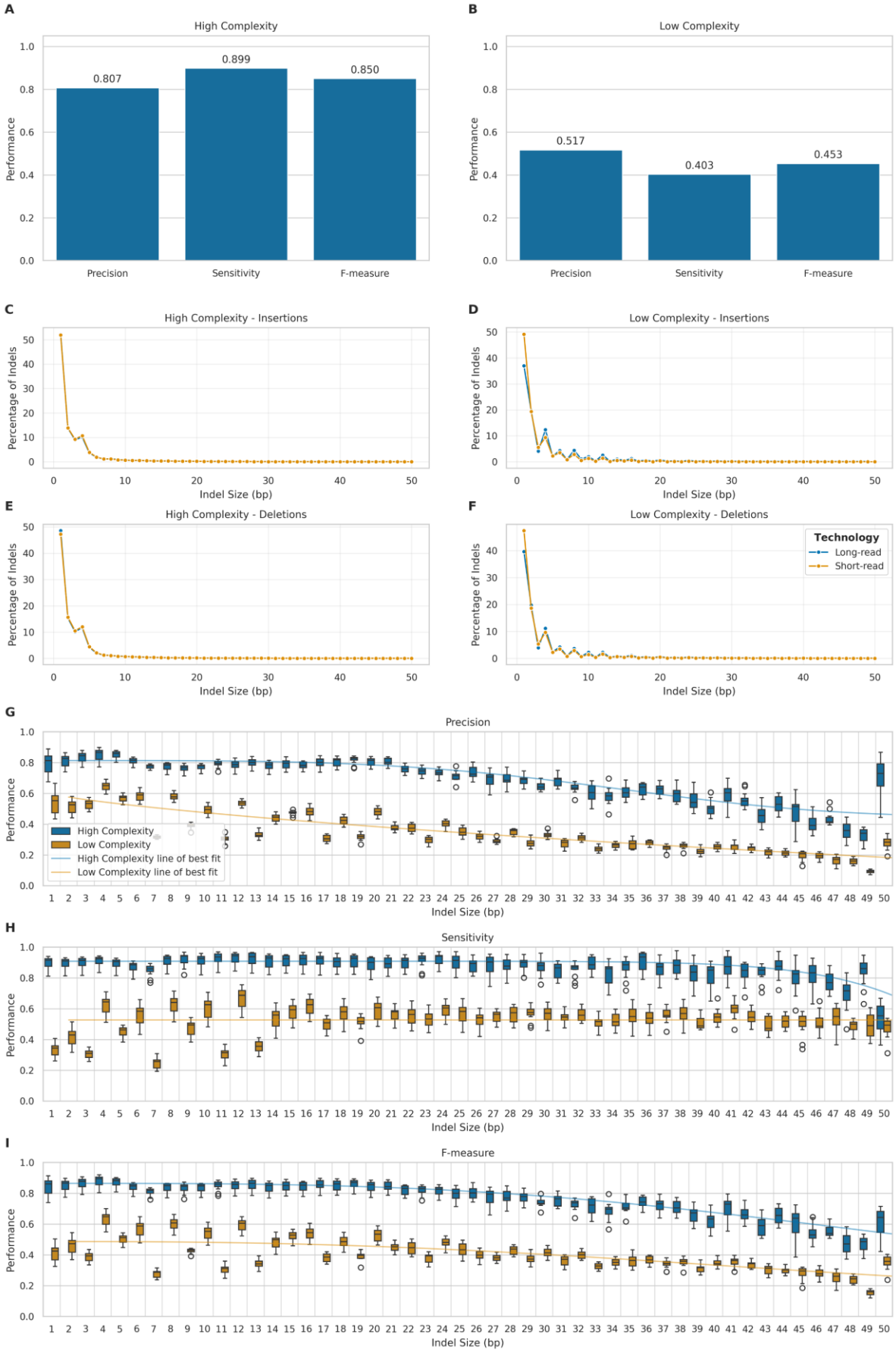
A
High Complexity
Performance
0.807
0.899
0.850
Precision
Sensitivity
F-measure
B
Low Complexity
0.517
0.403
0.453
C
High Complexity - Insertions
Percentage of Indels
Indel Size (bp)
D
Low Complexity - Insertions
E
High Complexity - Deletions
F
Low Complexity - Deletions
Technology
Long-read
Short-read
G
Precision
High Complexity
Low Complexity
High Complexity line of best fit
Low Complexity line of best fit
H
Sensitivity
I
F-measure

Figure 6 – Oxford Nanopore (long-read) sequencing performance analysis for Indel detection.

(A-B) Bar plots showing ONT (long-read) indel detection performance metrics (Precision, Sensitivity, F-measure) in high complexity (A) and low complexity (B) genomic regions. The y-axis represents the performance score from 0 to 1.

(C-F) Line plots depicting indel size distribution for ONT (long-read) and Illumina (short-read) technologies. Plots show the percentage of indels (y-axis) detected for each indel size (x-axis) up to 50 base pairs. (C) Insertions in high complexity regions. (D) Insertions in low complexity regions. (E) Deletions in high complexity regions. (F) Deletions in low complexity regions.

(G-I) Box plots showing ONT (long-read) indel calling performance metrics stratified by indel length (1-50 bp). The y-axis shows the performance score from 0 to 1, while the x-axis represents indel lengths. Results are shown separately for high complexity (blue) and low complexity (orange) genomic regions for: (G) Precision, (H) Sensitivity, and (I) F-measure.

Impact of multiplexing on variant calling

We compared the performance of variant calling, for both SNVs and indels in high complexity and low complexity regions, for multiplexed and singleplexed samples, to determine the effect of multiplexing on the accuracy of variant calling (**Figure 7**).

For SNVs in high complexity regions (**Figure 7A**), singleplexed samples demonstrated marginally superior performance across all metrics. Indeed, mean precision (singleplex: 0.960, SD = 0.003; multiplex: 0.944, SD = 0.008), sensitivity (singleplex: 0.967, SD = 0.003; multiplex: 0.939, SD = 0.016), and F-measure (singleplex: 0.964, SD = 0.003; multiplex: 0.941, SD = 0.012) were all higher for singleplexed samples. A similar trend was evident in low complexity regions (**Figure 7B**), with higher mean precision (singleplex: 0.792, SD = 0.005; multiplex: 0.771, SD = 0.009), sensitivity (singleplex: 0.755, SD = 0.004; multiplex: 0.730, SD = 0.015), and F-measure (singleplex: 0.773, SD = 0.004; multiplex: 0.750, SD = 0.012) for singleplexed samples.

Indel calling exhibited more pronounced differences between multiplexing conditions. In high complexity regions (**Figure 7C**), singleplexed samples showed substantial improvements in mean precision (singleplex: 0.848, SD = 0.021; multiplex: 0.753, SD = 0.029), sensitivity (singleplex: 0.926, SD = 0.009; multiplex: 0.862, SD = 0.032), and F-measure (singleplex: 0.885, SD = 0.016; multiplex: 0.804, SD = 0.030). The disparity was even more evident in low complexity regions (**Figure 7D**), with higher mean precision (singleplex: 0.551, SD = 0.025; multiplex: 0.471, SD = 0.019), sensitivity (singleplex: 0.440, SD = 0.025; multiplex: 0.355, SD = 0.025), and F-measure (singleplex: 0.489, SD = 0.025; multiplex: 0.404, SD = 0.023) for singleplexed samples.

To separate the effects of multiplexing and sequencing depth, we performed an analysis of covariance (ANCOVA) (**Figure 8**). This analysis revealed that sequencing depth had a significant positive effect on all performance metrics ($p < 0.001$), with a particularly strong impact on indel calling ($R^2 > 0.95$). The independent effect of multiplexing, after controlling for depth, was considerably smaller and reached statistical significance only for indel precision in both high complexity (effect size = -0.021, $p = 0.013$) and low complexity regions (effect size = -0.008, $p = 0.010$). For SNV calling, the multiplexing effect was not statistically significant across any metric ($p > 0.15$), suggesting that the apparent performance differences were primarily driven by variations in sequencing depth rather than multiplexing itself.

Thus, while initial comparisons suggested that multiplexing has a modest impact on SNV calling performance and more significantly affects indel calling accuracy, particularly in low complexity regions, our subsequent analysis indicates that these effects are largely attributable to differences in sequencing depth rather than multiplexing per se. These findings suggest that maintaining adequate sequencing depth is more critical for variant calling performance than the choice between multiplexed and singleplexed approaches.

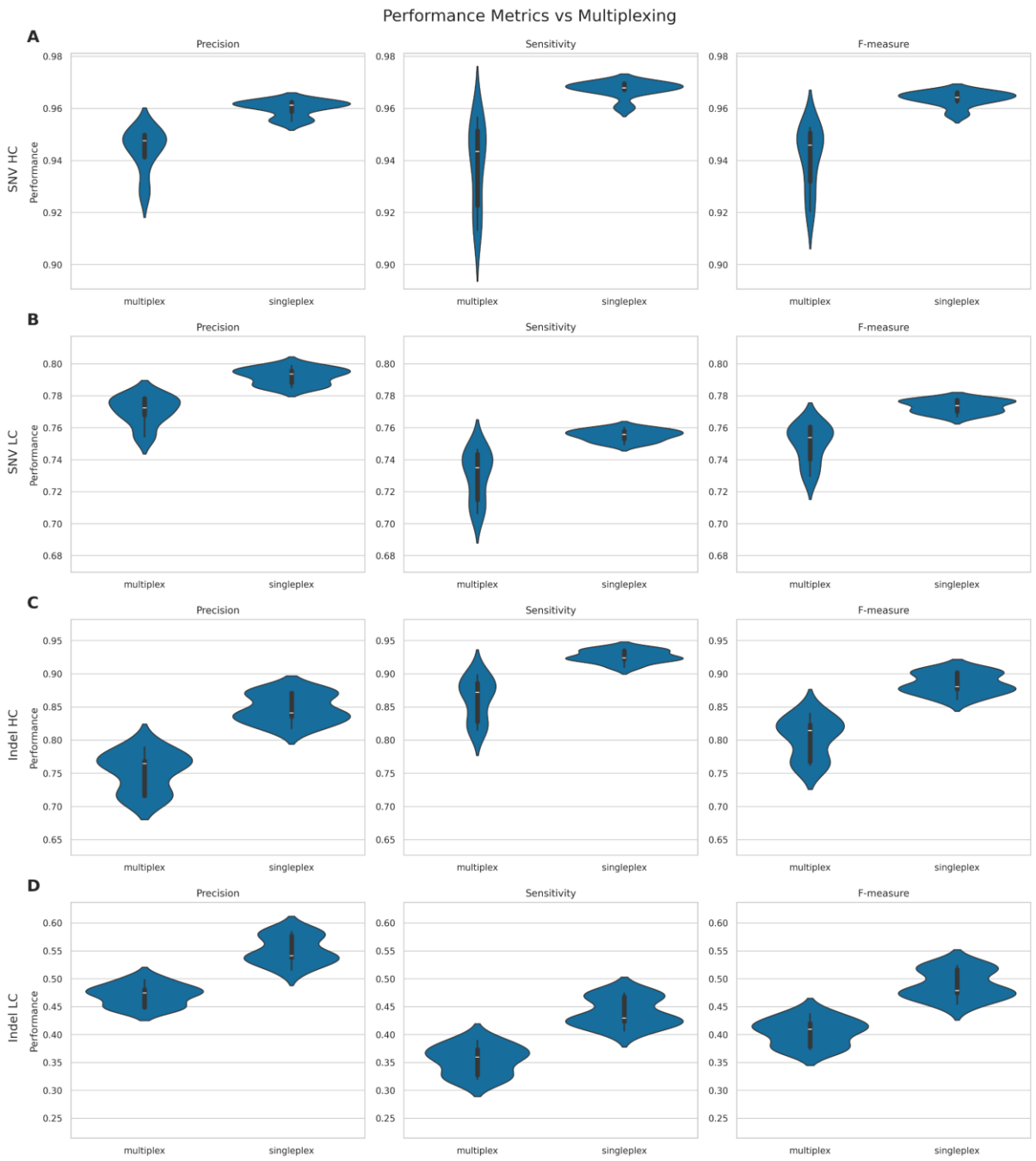

Figure 7 – Performance metrics comparison between multiplexed and singleplexed samples for SNV and Indel detection.

(A-B) Violin plots showing performance metrics (Precision, Sensitivity, F-measure) for Single Nucleotide Variant (SNV) detection in high complexity (A) and low complexity (B) genomic regions. The y-axis represents the performance scores.

(C-D) Violin plots showing performance metrics (Precision, Sensitivity, F-measure) for Indel detection in high complexity (C) and low complexity (D) genomic regions. The y-axis represents the performance scores.

For all plots (A-D), the x-axis compares multiplexed (left) and singleplexed (right) samples. The width of each plot represents the density of data points at different performance levels, with the white bar indicating the median value.

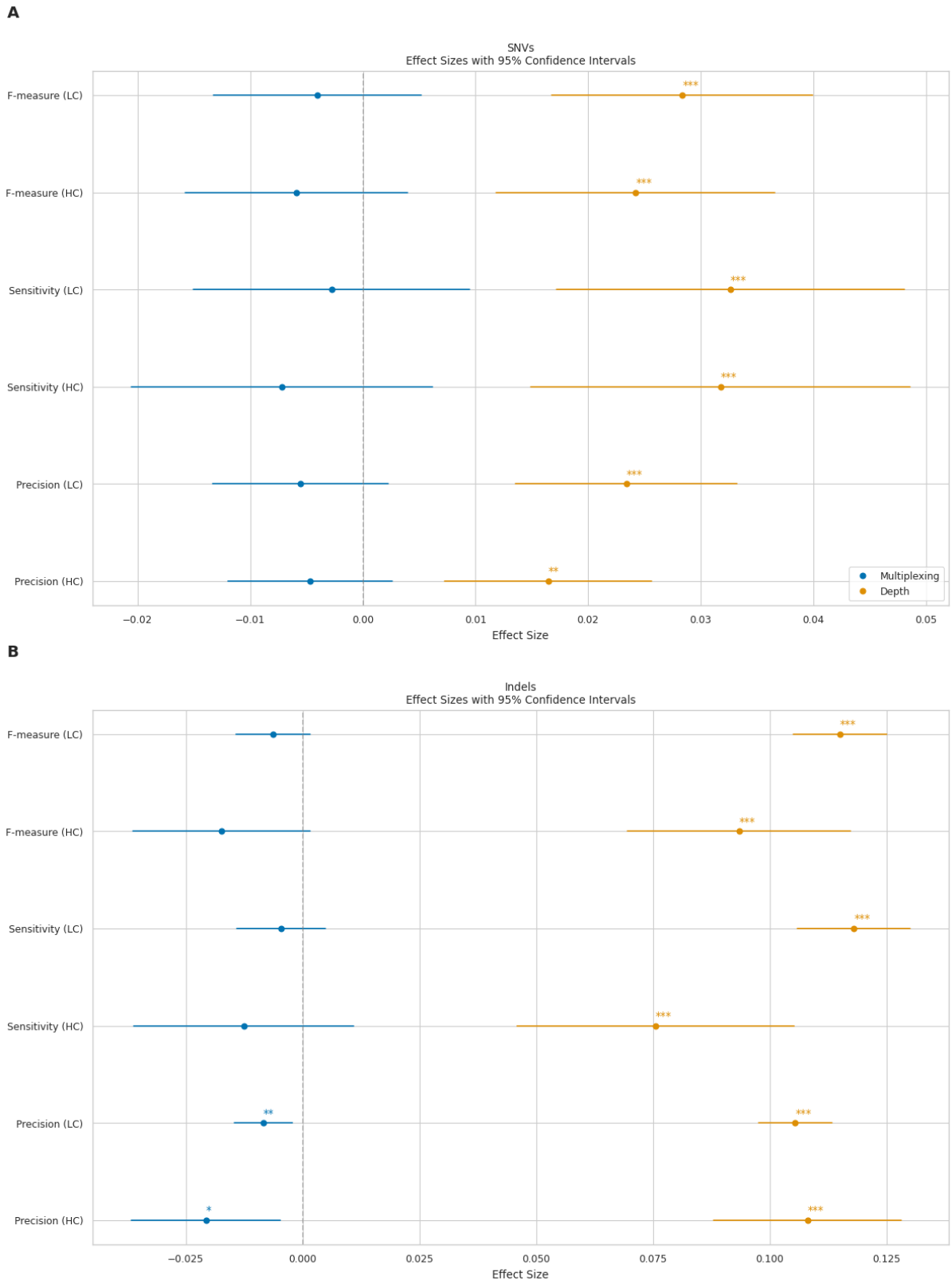

Figure 8 – Effect sizes of multiplexing and sequencing depth on variant calling performance metrics across genomic complexity regions.

(A) Forest plot showing the effect sizes with 95% confidence intervals for Single Nucleotide Variant (SNV) detection performance metrics (Precision, Sensitivity, F-measure) in both high complexity (HC) and low complexity (LC) genomic regions. Blue points and intervals represent multiplexing effects (relative to singleplex), while orange points and intervals represent the effects of log-transformed sequencing depth.

(B) Forest plot showing the corresponding effect sizes for Indel detection performance metrics. The x-axis represents the magnitude and direction of effects, with values below zero indicating negative impacts on performance and values above zero indicating positive impacts.

Statistical significance is indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Impact of sequencing depth on variant calling

To assess the influence of sequencing depth on variant calling performance, we analysed the relationship between whole genome mean depth and the variant calling performance metrics (precision, sensitivity, and F-measure) for both SNVs and indels in high complexity and low complexity regions (**Figure 9**).

For SNVs in high complexity regions (**Figure 9A**), we observed strong positive correlations between sequencing depth and all performance metrics (precision: $r = 0.802$, $p = 2.86 \times 10^{-7}$; sensitivity: $r = 0.785$, $p = 7.50 \times 10^{-7}$; F-measure: $r = 0.800$, $p = 3.31 \times 10^{-7}$). Similar trends were observed for SNVs in low complexity regions (**Figure 9B**), with slightly stronger correlations (precision: $r = 0.863$, $p = 3.40 \times 10^{-9}$; sensitivity: $r = 0.790$, $p = 5.73 \times 10^{-7}$; F-measure: $r = 0.836$, $p = 3.04 \times 10^{-8}$).

Indel calling performance showed even stronger correlations with sequencing depth. In high complexity regions (**Figure 9C**), the correlations were highly significant for all metrics (precision: $r = 0.925$, $p = 1.89 \times 10^{-12}$; sensitivity: $r = 0.842$, $p = 2.02 \times 10^{-8}$; F-measure: $r = 0.899$, $p = 7.78 \times 10^{-11}$), but the strongest correlations were observed for indels in low complexity regions (**Figure 9D**) (precision: $r = 0.968$, $p = 3.82 \times 10^{-17}$; sensitivity: $r = 0.957$, $p = 1.90 \times 10^{-15}$; F-measure: $r = 0.961$, $p = 4.95 \times 10^{-16}$).

On the whole, the relationship between sequencing depth and performance metrics followed an asymptotic trend for all variant types and genomic regions, therefore suggesting that while increasing sequencing depth generally improves variant calling performance, the rate of improvement plateaus at higher depths. Additionally, the impact of sequencing depth is more pronounced for indel calling compared to SNV calling, particularly in low complexity regions.

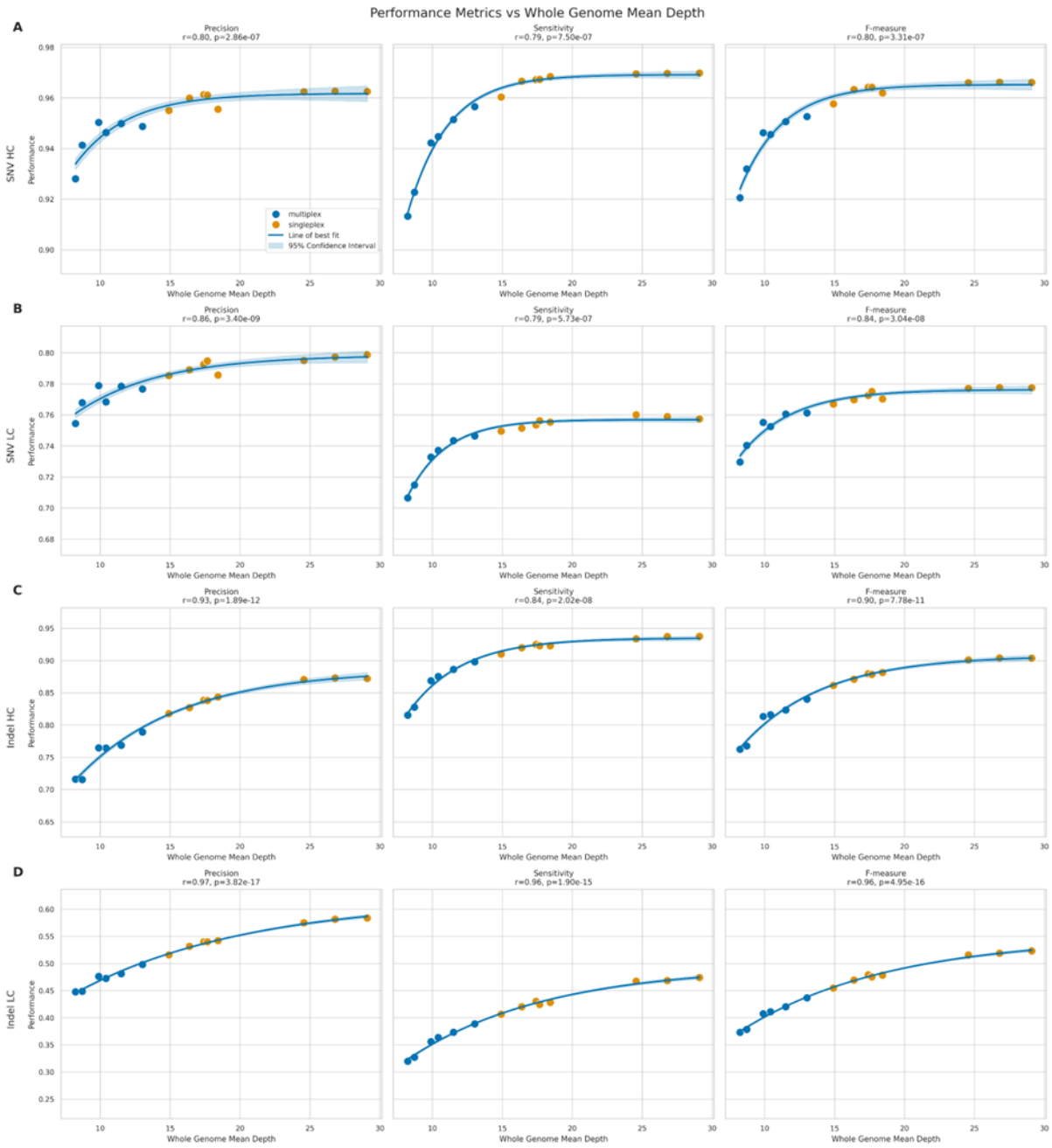

Figure 9 – Impact of whole genome mean depth on performance metrics for SNV and Indel detection in high and low complexity regions

(A-B) Scatter plots showing the relationship between whole genome mean depth and performance metrics (Precision, Sensitivity, F-measure) for SNV detection in high complexity (A) and low complexity (B) genomic regions.

(C-D) Scatter plots showing the relationship between whole genome mean depth and performance metrics (Precision, Sensitivity, F-measure) for indel detection in high complexity (C) and low complexity (D) genomic regions.

For all plots (A-D), the x-axis represents the whole genome mean depth, while the y-axis shows the performance scores. Blue dots represent multiplexed samples, and orange dots represent singleplexed samples. The blue line indicates the line of best fit using an asymptotic function fitted with non-linear least squares regression, with the light blue shaded area representing the 95% confidence interval. Each subplot includes the Pearson correlation coefficient (r) and its associated p-value.

Impact of read length on calling small variants

To investigate the relationship between read length and variant calling performance, we analysed the correlation between mean read length and the three performance metrics: precision, sensitivity, and F-measure, for both SNVs and indels, in high complexity and low complexity regions (**Figure 10**).

For SNVs in high complexity regions (**Figure 10A**), we observed moderate negative correlations between mean read length and all three performance metrics (precision: $r = -0.44$, $p = 0.021$; sensitivity: $r = -0.52$, $p = 0.005$; F-measure: $r = -0.49$, $p = 0.008$). Similar negative correlations were observed for SNVs in low complexity regions (**Figure 10B**) (precision: $r = -0.41$, $p = 0.029$; sensitivity: $r = -0.50$, $p = 0.007$; F-measure: $r = -0.47$, $p = 0.012$).

Moreover, indel calling performance also showed comparable negative correlations with mean read length. In high complexity regions (**Figure 10C**), the correlations were slightly stronger than those observed for SNVs (precision: $r = -0.49$, $p = 0.008$; sensitivity: $r = -0.52$, $p = 0.005$; F-measure: $r = -0.50$, $p = 0.006$), and, for indels in low complexity regions (**Figure 10D**), the correlations were also negative and statistically significant (precision: $r = -0.45$, $p = 0.016$; sensitivity: $r = -0.48$, $p = 0.010$; F-measure: $r = -0.47$, $p = 0.011$).

Therefore, these results suggest that longer read lengths are associated with a moderate decrease in variant calling performance for both SNVs and indels, across both high and low complexity regions.

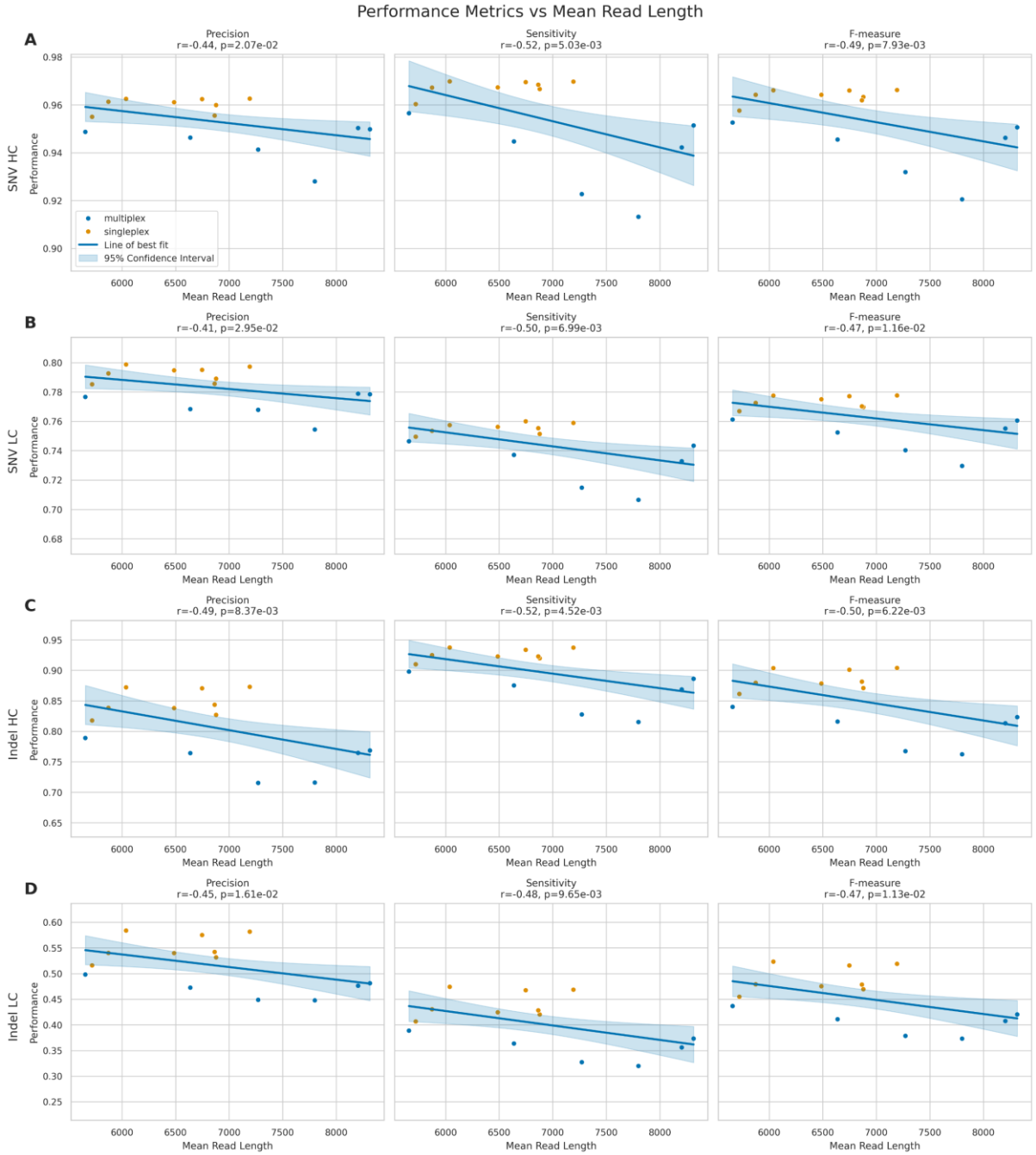

Figure 10 – Impact of mean read length on variant calling performance metrics for SNVs and Indels in high and low complexity regions.

(A) Scatter plots showing the correlation between mean read length (x-axis) and performance metrics (Precision, Sensitivity, F-measure; y-axis) for SNVs in high complexity (HC) genomic regions.

(B) Scatter plots depicting the correlation between mean read length and performance metrics for SNVs in low complexity (LC) genomic regions.

(C) Scatter plots illustrating the correlation between mean read length and performance metrics for indels in high complexity (HC) genomic regions.

(D) Scatter plots showing the correlation between mean read length and performance metrics for indels in low complexity (LC) genomic regions.

In all plots, each point represents a sample, with blue dots indicating multiplex samples and orange dots representing singleplex samples. The blue line shows the line of best fit, with the shaded area representing the 95% confidence

interval. Pearson correlation coefficients (r) and p-values are displayed for each metric. The x-axis represents mean read length in base pairs, while the y-axis shows the performance score from 0 to 1.

Structural Variants overview

To evaluate the performance of ONT LRS and Illumina SRS platforms in detecting structural variants (SVs), we conducted a consensus analysis with SURVIVOR (39) to merge consensus SV calls from both technologies. SV calls were considered in consensus if they met the following criteria: (1) located within 500 base pairs of each other, (2) detected in both VCF files for a given sample, (3) classified as the same SV type, (4) identified on the same strand, and (5) exhibiting a size difference of no more than 30% between platforms.

Overall, ONT LRS identified 2.82 times more SVs compared to Illumina SRS across all samples (**Figure 11A**). However, the consensus between the two platforms was relatively low, with an average of 20.87% (SD = 1.56%) of LRS calls and 57.79% (SD = 9.73%) of SRS calls being concordant. Furthermore, regarding SV size distribution (**Figure 11B**), LRS demonstrated a broader range of SV sizes, particularly in detecting larger variants, while Illumina SRS showed a bias towards detecting smaller SVs. In fact, LRS detected SVs ranging from 8 bp to 129,371,498 bp, with a mean size of 3,899.31 bp (SD = 509,051.75 bp) and a median of 81 bp. In contrast, SRS identified SVs between 2 bp and 6,063 bp, with a mean size of 165.59 bp (SD = 164.37 bp) and a median of 96 bp. The markedly larger maximum SV size (129 Mb for LRS vs 6 Kb for SRS) and higher SD observed in LRS data highlight its superior capability in detecting large-scale genomic rearrangements, which are often challenging to identify with SRS.

To further investigate the nature of these SV calls, we analysed the distribution of SV types for each sequencing platform (**Figure 11C**, **D**). For LRS, insertions were the most frequently detected (mean = 14,538.64, SD = 3,287.51, per sample), followed closely by deletions (mean = 11,945.57, SD = 2,656.37). SRS, however, detected deletions most frequently (mean = 5,491.57, SD = 139.84), followed by insertions (mean = 3,230.07, SD = 127.40). Notably, SRS detected a higher number of breakends (BND) (mean = 741.21, SD = 44.07) compared to LRS (mean = 50.71, SD = 35.10), while LRS showed the capability to detect STRs (mean = 27.14, SD = 7.84).

The size distribution of different SV types further showed distinct patterns between LRS and SRS platforms (**Figure 12**). For insertions (**Figure 12A**), both technologies showed multimodal distributions, with LRS capturing a broader range of sizes. SRS exhibited a trimodal distribution with a sharp peak at smaller insertion sizes (< 10 bp), a second peak around 100 bp, and a third, broader peak centred approximately at 300 bp. In contrast, LRS demonstrated a bimodal distribution with peaks at approximately 300 bp and 6 kb. Deletion size distributions (**Figure 12B**) showed similar bimodal patterns for both platforms, with LRS again exhibiting a wider range. SRS deletions were predominantly clustered below 1 kb, while LRS effectively captured deletions spanning several orders of magnitude, from 50 bp to 10 Mb.

Duplication events (**Figure 12C**) were primarily detected by LRS, with sizes ranging from 100 bp to 100 Mb and a peak around 10 kb. The absence of a substantial SRS signal for duplications highlights the advantage of LRS in identifying these complex structural variants. Inversion size distributions (**Figure 12D**) showed striking differences between the platforms. LRS detected inversions across a broad size range (10^2 to 10^8 bp), while SRS mainly identified smaller inversions (< 10^3 bp).

Breakend (BND) size distributions (**Figure 12E**) were similar between platforms, with both showing a peak around 20 bp. However, SRS demonstrated a slightly broader range, extending to approximately 100 bp. Short Tandem Repeat (STR) size distributions (**Figure 12F**) were exclusively captured by LRS, with sizes ranging up to approximately 700 bp and a peak around 50 bp.

To investigate the genomic distribution of SVs, we analysed their occurrence across chromosomes, normalizing for chromosome length (**Figure 13A**). For both LRS and SRS platforms, chromosomes 19, 20, 21 exhibited notably higher percentages of SVs per megabase compared to other chromosomes, while, conversely, chromosomes 14, 15, X, and Y showed lower SV densities for both platforms. Interestingly, despite their technological differences, both LRS and SRS platforms showed similar trends in chromosomal SV distribution.

A further correlation analysis between chromosome length and SV occurrence (**Figure 13B**) showed strong positive correlations between chromosome length and SV counts for both LRS ($r = 0.891$, $p = 5.360\times10^{-9}$) and SRS ($r = 0.907$, $p = 1.039\times10^{-9}$) platforms.

The elevated SV density observed in chromosomes 19, 20, and 21 could be attributed to various factors, including higher gene density, increased recombination rates, or the presence of specific genomic features that promote structural rearrangements (40). For instance, chromosome 19, is known for its high gene density and elevated GC content, which may contribute to its increased propensity for structural variations (41).

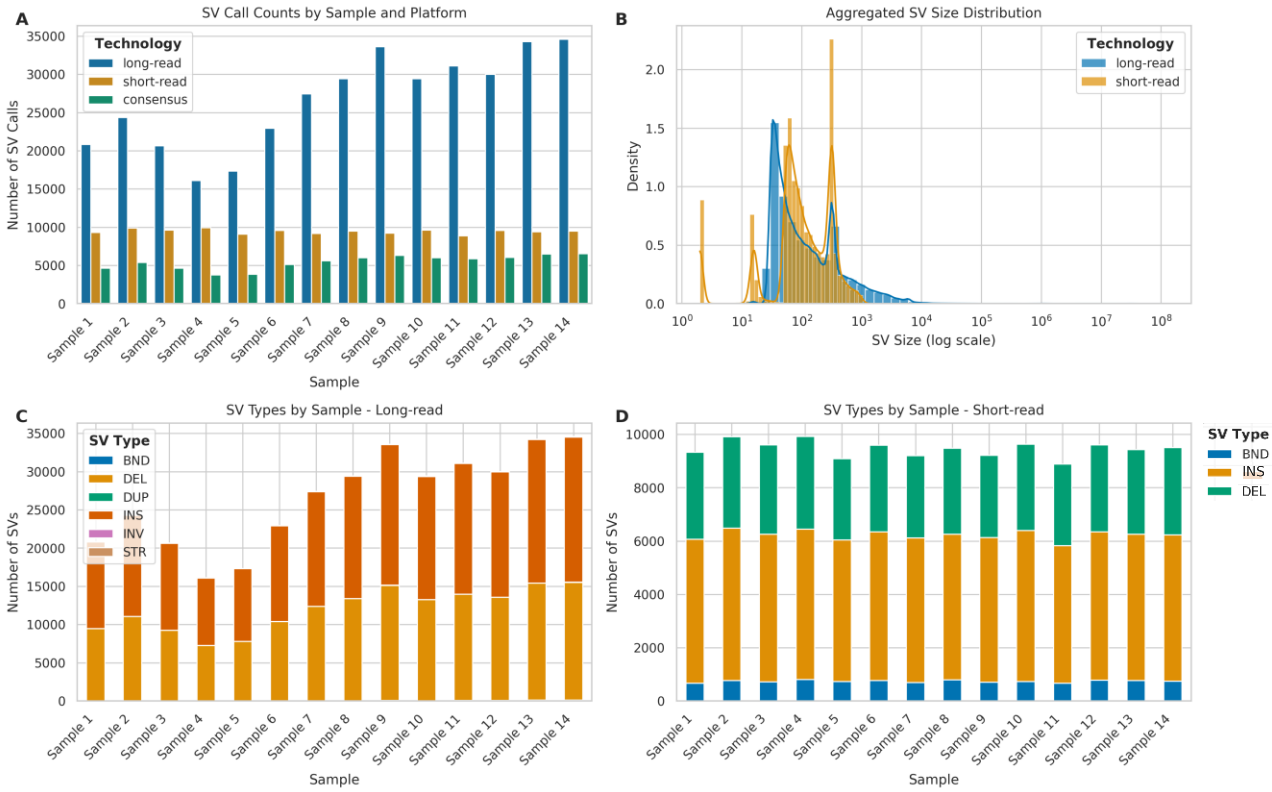

Figure 11 – Comparative analysis of structural variant (SV) detection using long-read (ONT) and short-read (Illumina) sequencing platforms.

(A) Bar plot showing the number of SV calls by sample and platform. Each group of bars represents a sample, with blue bars indicating long-read (ONT) calls, orange bars showing short-read (Illumina) calls, and green bars representing consensus calls between the two platforms.

(B) Histogram depicting the aggregated SV size distribution for long-read (blue) and short-read (orange) platforms. The x-axis shows SV size on a logarithmic scale, while the y-axis represents density of SV calls.

(C) Stacked bar plot illustrating the distribution of SV types detected by the long-read (ONT) platform across all samples. Each bar represents a sample, with different colours indicating various SV types; BND (Breakend), DEL (Deletion), DUP (Duplication), INS (Insertion), INV (Inversion), STR (Short Tandem Repeat).

(D) Stacked bar plot showing the distribution of SV types detected by the short-read (Illumina) platform across all samples. Each bar represents a sample, with different colours indicating various SV types; BND (Breakend), DEL (Deletion), INS (Insertion).

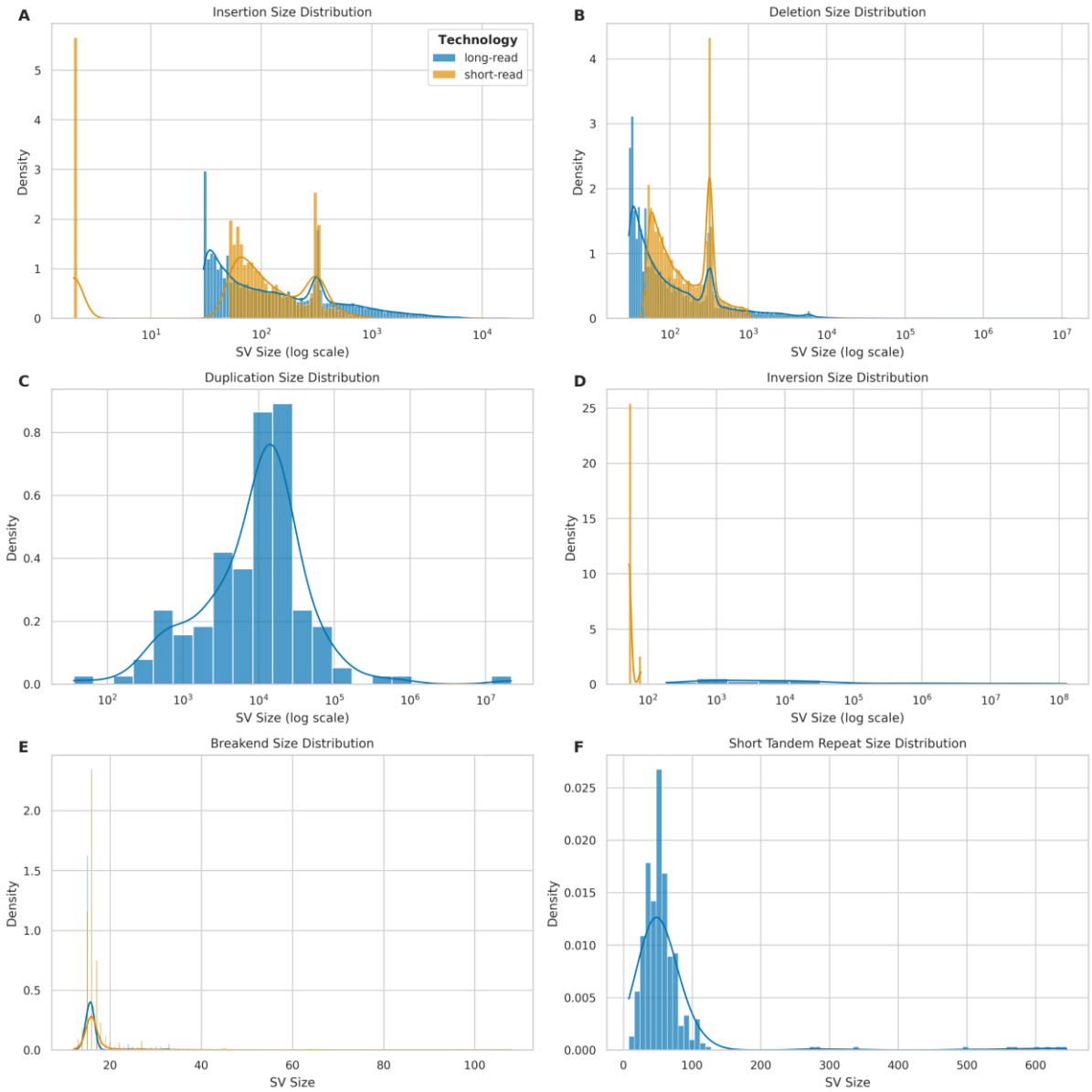

Figure 12 – Size distribution of different structural variant (SV) types detected by long-read (ONT) and short-read (Illumina) sequencing platforms.

(A) Histogram showing the insertion size distribution for long-read (blue) and short-read (orange) platforms. The x-axis displays SV size on a logarithmic scale, while the y-axis represents density.

(B) Histogram depicting the deletion size distribution for long-read (blue) and short-read (orange) platforms. The x-axis shows SV size on a logarithmic scale, and the y-axis represents density.

(C) Histogram illustrating the duplication size distribution, primarily detected by the long-read platform (blue). The x-axis displays SV size on a logarithmic scale, and the y-axis represents density.

(D) Histogram showing the inversion size distribution for long-read (blue) and short-read (orange) platforms. The x-axis displays SV size on a logarithmic scale, and the y-axis represents density.

(E) Histogram depicting the breakend (BND) size distribution for long-read (blue) and short-read (orange) platforms. The x-axis shows SV size, and the y-axis represents density.

(F) Histogram illustrating the Short Tandem Repeat (STR) size distribution, exclusively detected by the long-read platform (blue). The x-axis displays SV size, and the y-axis represents density.

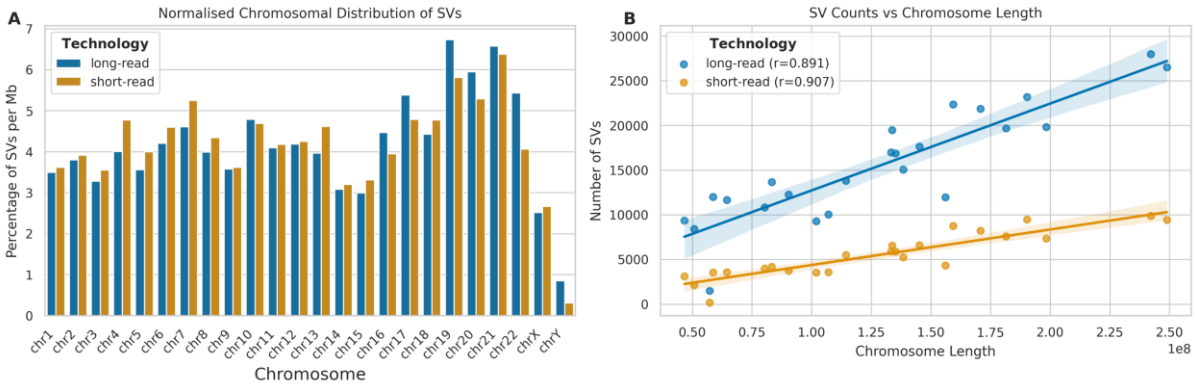

Figure 13 – Chromosomal distribution and correlation analysis of structural variants (SVs) detected by long-read (ONT) and short-read (Illumina) sequencing platforms.

(A) Bar plot showing the normalized chromosomal distribution of SVs for long-read (blue) and short-read (orange) platforms. The x-axis displays chromosomes, while the y-axis represents the percentage of SVs per megabase (Mb).

(B) Scatter plot illustrating the correlation between SV counts and chromosome length for long-read (blue) and short-read (orange) platforms. The x-axis shows chromosome length in base pairs, and the y-axis represents the number of SVs. Solid lines represent the lines of best fit for each platform, with shaded areas indicating the 95% confidence intervals. Pearson correlation coefficients (r) are provided for each platform.

Impact of sequencing depth on calling structural variants

To evaluate the influence of sequencing depth on structural variant (SV) detection, we analysed the relationship between whole genome mean depth and both the number and size distribution of SV calls across LRS and SRS platforms.

For LRS, we observed a strong positive correlation between sequencing depth and the number of SV calls (**Figure 14A**) ($r = 0.9388$, $p = 6.65\times10^{-7}$). The relationship followed an asymptotic trend, with the number of SV calls increasing from approximately 16,000 at 10X coverage to around 35,000 at 30X coverage, though the rate of increase diminished at higher depths. In contrast, short-read sequencing showed no significant correlation between depth and SV calls ($r = -0.2137$, $p = 0.4633$), with the number of calls remaining relatively stable around 9,450 across all sequencing depths. The consensus calls between the two platforms demonstrated a strong positive correlation with sequencing depth ($r = 0.8897$, $p = 2.04\times10^{-5}$), suggesting that increased depth primarily improves the detection of true structural variants.

Analysis of SV size distributions across sequencing depths revealed distinct patterns between the two platforms (**Figure 14B**). For LRS, we observed consistent detection of SVs across a broad size range (10^1 to 10^8 bp) at all sequencing depths, with no significant trend in median SV size (slope = -62.23, $p = 0.508$). The mean SV size showed considerable variation (range: 412 - 9,848 bp), likely due to improved detection of larger variants at higher depths. On the other hand, SRS demonstrated a more constrained size distribution (10^1 to 10^3 bp) that remained stable across all depths, with minimal variation in both median (96 - 99 bp) and mean (163 - 168 bp) SV sizes (slope = -0.04, $p = 0.434$).

These findings suggest that while increased sequencing depth substantially improves SV detection in long-read sequencing, particularly for larger variants, short-read SV calling remains relatively consistent across different depths. The data indicate that a minimum of 15-20X coverage may be optimal for long-read SV detection, as the rate of novel SV discovery plateaus beyond this depth.

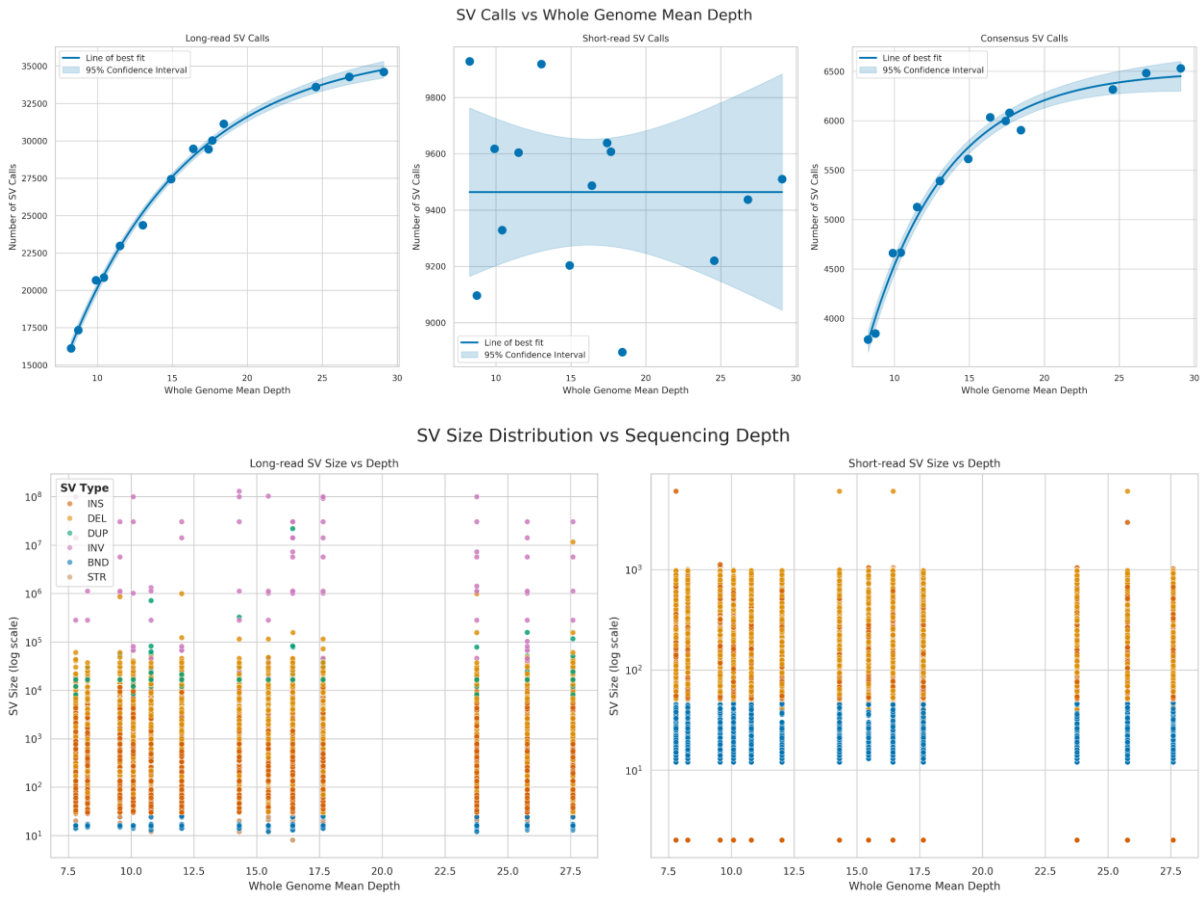

Figure 14 – Impact of sequencing depth on structural variant (SV) detection using long-read (ONT) and short-read (Illumina) sequencing platforms.

(A) Scatter plots showing the relationship between whole genome mean depth and number of SV calls for long-read (left), short-read (middle), and consensus (right) calls. The x-axis shows whole genome mean depth, while the y-axis represents the number of SV calls. Solid blue lines represent the lines of best fit, with shaded areas indicating the 95% confidence intervals. Pearson correlation coefficients (r) are provided for each analysis.

(B) Scatter plots depicting the distribution of SV sizes across different sequencing depths for long-read (left) and short-read (right) platforms. The x-axis shows whole genome mean depth, while the y-axis represents SV size on a logarithmic scale. Different colors indicate various SV types: INS (insertions), DEL (deletions), DUP (duplications), INV (inversions), BND (breakends), and STR (short tandem repeats).

Discussion

This study provides a comprehensive comparison of Oxford Nanopore Technologies (ONT) long-read sequencing (LRS) and Illumina short-read sequencing (SRS) platforms, focusing on their performance in variant detection across different genomic contexts. Our findings provide a comprehensive understanding of their performance in variant detection, particularly in challenging genomic regions, and highlight the impact of experimental factors on sequencing outcomes.

Sequencing Quality and Yield

The comparison between multiplexed and singleplexed ONT samples revealed significant differences in sequencing yield and read length characteristics. Singleplexed samples consistently produced higher numbers of reads and bases, with a 119.41% increase in mean read count and a 100.15% increase in mean base count compared to multiplexed samples. This substantial difference in yield has important implications for experimental design and cost-effectiveness in genomic studies. While multiplexing offers cost savings and increased throughput, our results suggest that it may come at the expense of reduced per-flowcell sequencing depth and potentially slightly lower variant detection accuracy, particularly for structural variants and in low complexity regions, although this can be mitigated by optimising barcoding.

Interestingly, we observed distinct read length distribution patterns between multiplexed and singleplexed samples. The consistent bimodal distribution in singleplexed samples, with peaks at shorter (~400 bp) and longer (~10,000 bp) read lengths, contrasts with the more variable patterns seen in multiplexed samples. This difference may reflect variations in library preparation or sequencing efficiency between the two approaches and warrants further investigations to optimise multiplexing protocols for long-read sequencing.

Single Nucleotide Variant Detection

Our analysis of SNV detection performance revealed that while SRS maintains an overall advantage, particularly in high complexity regions, LRS demonstrates competitive performance, especially in low complexity areas. In high complexity regions, SRS showed statistically significant higher precision, sensitivity, and F-measure compared to LRS. However, the performance gap narrowed considerably in low complexity regions, with LRS showing comparable sensitivity and F-measure to SRS.

The distinct error profiles observed for LRS and SRS across different SNV types provide valuable insights into the strengths and weaknesses of each technology. The higher error rates for LRS, particularly for transitions in high complexity regions, suggest that improvements in base-calling accuracy should remain a priority for ONT. However, the competitive performance of LRS in detecting certain SNV types within low complexity

regions, particularly in terms of false negative rates, highlights its potential advantages in analysing repetitive genomic areas where traditional short-read sequencing faces inherent limitations (22).

Indel Detection

The performance of LRS in indel detection showed marked differences between high and low complexity regions, with a clear dependency on indel size. In high complexity areas, LRS demonstrated robust performance particularly for small (1-5bp) and medium-sized indels (6-20bp), which represent the vast majority of variants (>95%), achieving F-measures of up to 0.874. However, there was a significant decline in performance in low complexity regions, where even small indels showed markedly reduced F-measures of 0.481. Performance deteriorated further with increasing indel size in these regions, with F-measures dropping to 0.356 for indels of 21-50bp.

This size-dependent pattern suggests that while current LRS technology can reliably detect small to medium-sized indels in non-repetitive regions, significant challenges remain in accurately calling indels of any size in repetitive areas. The particularly poor performance for larger indels in low complexity regions could partially reflect limitations in our SRS-based benchmarking approach, as short-read technologies face inherent challenges in detecting larger variants in repetitive regions. Future benchmarking studies incorporating more robust validation methods may help distinguish true technological limitations from artifacts of the comparison methodology.

Impact of Experimental Factors on Variant Calling

Our investigation into the effects of multiplexing, sequencing depth, and read length on variant calling performance yielded several important insights. Initially, multiplexing appeared to have a modest impact on SNV calling performance and a more substantial effect on indel calling accuracy. However, our analysis revealed that these apparent effects were largely attributable to differences in sequencing depth rather than multiplexing status itself. After controlling for depth, multiplexing showed minimal impact on SNV calling and only modestly affected indel precision. This finding suggests that researchers can confidently employ multiplexing strategies without significantly compromising variant calling accuracy, provided adequate sequencing depth is maintained. The ANCOVA results demonstrated particularly strong relationships between sequencing depth and variant calling performance metrics, especially for indel detection ($r^2 > 0.95$). This finding highlights the importance of selecting adequate coverage in genomic studies. Furthermore, the asymptotic trend in this relationship suggests that while increasing sequencing depth generally improves variant calling performance, there may be diminishing returns at very high depths. This information can guide researchers in optimising their sequencing strategies to balance cost and variant calling accuracy.

Notably, we also found a moderate negative correlation between mean read length and variant calling performance for both SNVs and indels. This unexpected result warrants further investigation and may reflect interactions between read length, sequencing errors, and alignment challenges. It also underscores the need for continued refinement of bioinformatics tools to fully leverage the potential of the read lengths generated by LRS.

Structural Variant (SV) Detection

Our analysis revealed significant differences in SV detection capabilities between LRS and SRS platforms. LRS identified 2.82 times more SVs compared to SRS across all samples, demonstrating its enhanced ability to capture a broader spectrum of structural variations. This advantage was particularly evident at higher sequencing depths, with our analysis revealing a strong positive correlation between depth and SV detection in LRS ($r = 0.9388$, $p = 6.65\times10^{-7}$), while SRS showed no significant depth dependence ($r = -0.2137$, $p = 0.4633$). The relationship in LRS followed an asymptotic trend, suggesting optimal SV detection might be achieved around 15-20X coverage. Furthermore, the superior capability of LRS in detecting large-scale genomic rearrangements, as evidenced by the markedly larger maximum SV size (129 Mb for LRS vs 6 Kb for SRS), highlights its potential in uncovering previously undetectable genomic events. Importantly, this broad size range detection capability remained consistent across all sequencing depths in LRS (10^1 to 10^8 bp), while SRS maintained a more constrained size distribution (10^1 to 10^3 bp) regardless of coverage depth.

The distinct patterns observed in SV size distributions and types between LRS and SRS platforms further emphasise their complementary nature. While SRS showed a bias towards detecting smaller SVs, LRS demonstrated a broader range of detectable SV sizes and types, including improved detection of large insertions, deletions, and inversions. The exclusive detection of Short Tandem Repeats (STRs) by LRS underscores its unique advantages in detecting certain types of genomic variation. However, it is important to note that these differences are not solely attributable to the sequencing technologies themselves, but also to the specific variant callers used for each platform in this study. Indeed, the choice of variant caller can significantly impact the number and types of structural variants discovered, and utilising different callers or a combination of multiple callers on the same data could potentially increase the number and diversity of structural variants identified on both platforms. Future studies comparing LRS and SRS technologies should consider evaluating and comparing the performance of multiple variant callers to provide a more comprehensive assessment of each platform's capabilities in structural variant detection.

In addition, the analysis of chromosomal distribution of SVs revealed interesting patterns, with chromosomes 19, 20, and 21 exhibiting notably higher SV densities, and the strong positive correlation between chromosome length and SV counts for both platforms provides further insights into the genomic landscape of structural variations.

Limitations and Future Directions

While our study provides insights into the performance of ONT LRS technology, several limitations should be acknowledged. Our deliberate selection of the highest quality flowcells for multiplexing introduced a methodological bias that confounds the direct comparison between multiplexed and singleplexed approaches. This bias, while potentially minimising the performance impact of multiplexing, makes it challenging to fully assess the true effects of multiplexing on sequencing quality and variant detection. Moreover, our relatively small sample size (n=14) limits the statistical power and generalisability of our findings, and the use of SRS as a gold standard for indel comparisons may introduce biases, particularly in regions where SRS itself faces challenges. Additionally, the potential biases in SV detection methods and the limited sample size and diversity in our study call for further investigations with larger, more diverse cohorts.

To address these limitations, future studies should employ randomized flowcell allocation for multiplexing experiments and significantly larger sample sizes to provide more robust statistical comparisons. A larger comprehensive study, using randomly assigned flowcells of varying quality, would better elucidate the true impact of multiplexing and flowcell quality on sequencing performance. Such a study could also incorporate technical replicates to assess the reproducibility of findings across different flowcell qualities.

Looking ahead, the integration of LRS and SRS technologies presents exciting opportunities for comprehensive genomic analysis. The complementary strengths of these platforms could enable more accurate and complete characterisation of genomic variations across different scales and complexities. Continued improvements in base-calling accuracy and variant detection algorithms for LRS will be vital in fully realising its potential. Future research should explore the applications of LRS in clinical genomics and complex trait studies, where its ability to detect a wider range of structural variations could uncover novel disease-associated variants. Additionally, further optimisation of multiplexing protocols and sequencing strategies for LRS could enhance its cost-effectiveness and accessibility for large-scale genomic studies.

Methods

Sample Selection and Preparation

Samples used in this study were obtained from Project MinE, an international collaboration focused on whole-genome sequencing of amyotrophic lateral sclerosis (ALS) patients and controls (42). From this cohort, we selected 14 samples for analysis. Venous blood was drawn from participants, and genomic DNA was isolated using standard methods. DNA concentrations were standardized to 100 ng/µl as measured by fluorometry using the Quant-iT™ PicoGreen™ dsDNA quantitation assay (Thermo Fisher Scientific). DNA integrity was assessed using gel electrophoresis.

Microarray Genotyping

DNA samples were genotyped using the Infinium Omni2.5-8 v1.4 array (Illumina, Cambridge, United Kingdom) according to the manufacturer's protocol. To ensure consistency across platforms, the genotyped variants were lifted over to the GRCh38 reference genome using GATK LiftoverVcf (43). Only variants that were successfully lifted over and had matching alleles in the reference genome were retained for subsequent analyses.

Illumina Short-Read Sequencing

DNA samples were sequenced using Illumina's FastTrack services (Illumina, San Diego, CA, USA) on the Illumina HiSeq 2000 platform. Sequencing was performed using 100 bp paired-end reads with PCR-free library preparation, yielding approximately 35× coverage across each sample. The Isaac pipeline (44) was used for alignment to the hg19 reference genome and to call single nucleotide variants (SNVs), insertions and deletions (indels), and larger structural variants (SVs). To ensure consistency across platforms, the called variants were lifted over to the GRCh38 reference genome using GATK LiftoverVcf (43). Only variants that were successfully lifted over and had matching alleles in the reference genome were retained for subsequent analyses.

Oxford Nanopore Long-Read Sequencing

For ONT sequencing, DNA libraries were extracted with the Monarch® Spin gDNA Extraction Kit T3010S (New England Biolabs) and prepared using the SQK-LSK114 Ligation Sequencing Kit (Oxford Nanopore Technologies), according to the respective manufacturers' protocols. The prepared libraries were quantified using a Qubit fluorometer (Thermo Fisher Scientific). For multiplexed runs, barcoding was performed using the SQK-NBD114.24 Native Barcoding Kit (Oxford Nanopore Technologies). Sequencing was performed on PromethION 24 instruments using R10.4.1 flowcells. Two samples were pooled and sequenced on a single flow cell for multiplexed runs, while

individual samples were sequenced on dedicated flow cells for singleplexed runs. Sequencing runs were performed for approximately 72 hours.

Raw signal data were base-called with the EPI2ME wf-basecalling v1.1.7 pipeline (45), using the Dorado model dna r10.4.1 e8.2 400bps sup v4.3.0. The unaligned BAM output was assessed for quality with NanoPlot (46). Using the EPI2ME wf-human-variation v2.1.0 pipeline (47-55), reads were aligned to the GRCh38 reference genome; this pipeline uses Clair3 (50) to call SNVs and indels, Sniffles2 (51) to detect SVs, and Straglr (55) for STRs.

Variant Comparison and Benchmarking

All variant files were filtered and compared using a custom Nextflow pipeline, which is available at https://github.com/renatosantos98/ont-benchmark and can be executed using the command *nextflow run renatosantos98/ont-benchmark*.

Single Nucleotide Variants and Indels

For SNVs, variants from all three platforms (ONT LRS, Illumina SRS, and Illumina microarray) were intersected with regions excluding and including the Tandem Repeats and Homopolymers regions defined by the GIAB v3.4 benchmark (32), to form the high complexity and low complexity regions, respectively. Only variants present in the microarray dataset were retained for the LRS and SRS datasets to ensure consistent comparison across platforms.

For indels, variants from LRS and SRS platforms were also filtered, based on the same high-confidence regions, into high-complexity and low-complexity sets based on their genomic location.

Comparison of SNVs and indels was performed using RTG Tools vcfeval v3.12 (36, 37), to generate true positive, false positive, and false negative variant sets, as well as performance metrics including precision, recall, and F-measure. For SNV comparisons, the microarray data served as the truth set, while for indel comparisons, the Illumina SRS data was used as the truth set.

Structural Variants

SVs from LRS and SRS were filtered for quality, whereby for LRS variants with at least 5 supporting reads and STRs with a minimum of 10 spanning reads were retained, and for SRS variants with at least 5 supporting paired-end or split reads were retained.

The filtered callsets were compared using SURVIVOR (39), by generating a merged consensus set of variants from both platforms which included only variants within 500 base pairs of each other, of the same type, on the same strand, and with a size difference of no more than 30%.

Statistical Analysis

Statistical analyses were performed using jupyterlab 4.2.4 (56), Python 3.12 (57), numpy 2.0.0 (58), pandas 2.2.2 (59), matplotlib 3.9.1(60), seaborn 0.13.2 (61), SciPy 1.14.0 (62), scikit-learn 1.5.1 (63), and statsmodels 0.14.2 (64). This work was performed on the King's Computational Research, Engineering and Technology Environment (CREATE) at King's College London (65).

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
2. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.
3. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
4. Balague-Dobon L, Caceres A, Gonzalez JR. Fully exploiting SNP arrays: a systematic review on the tools to extract underlying genomic structure. Brief Bioinform. 2022;23(2).
5. Cooley LD, Lebo M, Li MM, Slovak ML, Wolff DJ, Working Group of the American College of Medical G, et al. American College of Medical Genetics and Genomics technical standards and guidelines: microarray analysis for chromosome abnormalities in neoplastic disorders. Genet Med. 2013;15(6):484-94.
6. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet. 2013;93(5):779-97.
7. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol. 2021;82(11):801-11.
8. Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. NAR Genom Bioinform. 2021;3(1):lqab019.
9. Olson ND, Wagner J, McDaniel J, Stephens SH, Westreich ST, Prasanna AG, et al. PrecisionFDA Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions. Cell Genom. 2022;2(5).
10. Iacoangeli A, Al Khleifat A, Sproviero W, Shatunov A, Jones AR, Morgan SL, et al. DNAscan: personal computer compatible NGS analysis, annotation and visualisation. BMC Bioinformatics. 2019;20(1):213.
11. Marriott H, Kabiljo R, Al Khleifat A, Dobson RJ, Al-Chalabi A, Iacoangeli A. DNAscan2: a versatile, scalable, and user-friendly analysis pipeline for human next-generation sequencing data. Bioinformatics. 2023;39(4).
12. Garcia M, Juhos S, Larsson M, Olason PI, Martin M, Eisfeldt J, et al. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. F1000Res. 2020;9(63):63.
13. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
14. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
15. Iacoangeli A, Al Khleifat A, Sproviero W, Shatunov A, Jones AR, Opie-Martin S, et al. ALSgeneScanner: a pipeline for the analysis and interpretation of DNA sequencing data of ALS patients. Amyotroph Lateral Scler Frontotemporal Degener. 2019;20(3-4):207-15.
16. Investigators GPP, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. N Engl J Med. 2021;385(20):1868-80.

17. Vollger MR, Guitart X, Dishuck PC, Mercuri L, Harvey WT, Gershman A, et al. Segmental duplications and their variation in a complete human genome. Science. 2022;376(6588):eabj6965.
18. Bowles H, Kabiljo R, Al Khleifat A, Jones A, Quinn JP, Dobson RJB, et al. An assessment of bioinformatics tools for the detection of human endogenous retroviral insertions in short-read genome sequencing data. Front Bioinform. 2022;2:1062328.
19. Halman A, Oshlack A. Accuracy of short tandem repeats genotyping tools in whole exome sequencing data. F1000Res. 2020;9(200):200.
20. Pfeiffer F, Grober C, Blank M, Handler K, Beyer M, Schultze JL, et al. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. Sci Rep. 2018;8(1):10950.
21. Olson ND, Wagner J, Dwarshuis N, Miga KH, Sedlazeck FJ, Salit M, et al. Variant calling and benchmarking in an era of complete human genome sequences. Nat Rev Genet. 2023;24(7):464-83.
22. Ebbert MTW, Jensen TD, Jansen-West K, Sens JP, Reddy JS, Ridge PG, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. Genome Biol. 2019;20(1):97.
23. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016;17(1):239.
24. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348-65.
25. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nat Rev Genet. 2020;21(10):597-614.
26. Payne A, Holmes N, Rakyan V, Loose M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics. 2019;35(13):2193-8.
27. Ahsan MU, Gouru A, Chan J, Zhou W, Wang K. A signal processing and deep learning framework for methylation detection using Oxford Nanopore sequencing. Nat Commun. 2024;15(1):1448.
28. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif. 2015;3:1-8.
29. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sorensen EA, Wollenberg RD, et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. Nat Methods. 2022;19(7):823-6.
30. Ni Y, Liu X, Simeneh ZM, Yang M, Li R. Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. Comput Struct Biotechnol J. 2023;21:2352-64.
31. Stoeck T, Katzenmeier SN, Breiner H-W, Rubel V. Nanopore duplex sequencing as an alternative to Illumina MiSeq sequencing for eDNA-based biomonitoring of coastal aquaculture impacts. Metabarcoding and Metagenomics. 2024;8:e121817.
32. Krusche P, Trigg L, Boutros PC, Mason CE, De La Vega FM, Moore BL, et al. Best practices for benchmarking germline small-variant calls in human genomes. Nat Biotechnol. 2019;37(5):555-60.
33. Møller PL, Holley G, Beyter D, Nyegaard M, Halldórsson BV. Benchmarking small variant detection with ONT reveals high performance in challenging regions. bioRxiv. 2020:2020.10.22.350009.

34. Pei S, Liu T, Ren X, Li W, Chen C, Xie Z. Benchmarking variant callers in next-generation and third-generation sequencing analysis. Brief Bioinform. 2021;22(3):bbaa148.
35. Ahsan MU, Liu Q, Fang L, Wang K. NanoCaller for accurate detection of SNPs and indels in difficult-to-map regions from long-read sequencing by haplotype-aware deep neural networks. Genome Biol. 2021;22(1):261.
36. Cleary JG, Braithwaite R, Gaastra K, Hilbush BS, Inglis S, Irvine SA, et al. Joint variant and de novo mutation identification on pedigrees from high-throughput sequencing data. J Comput Biol. 2014;21(6):405-19.
37. Cleary JG, Braithwaite R, Gaastra K, Hilbush BS, Inglis S, Irvine SA, et al. Comparing Variant Call Files for Performance Benchmarking of Next-Generation Sequencing Variant Calling Pipelines. bioRxiv. 2015:023754.
38. Barbitoff YA, Abasov R, Tvorogova VE, Glotov AS, Predeus AV. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. BMC Genomics. 2022;23(1):155.
39. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun. 2017;8(1):14061.
40. Liao X, Zhu W, Zhou J, Li H, Xu X, Zhang B, et al. Repetitive DNA sequence detection and its role in the human genome. Commun Biol. 2023;6(1):954.
41. Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J, et al. The DNA sequence and biology of human chromosome 19. Nature. 2004;428(6982):529-35.
42. Consortium PMAS. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. Eur J Hum Genet. 2018;26(10):1537-46.
43. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.
44. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics. 2013;29(16):2041-3.
45. Labs EM. Basecalling workflow 2024 [Available from: https://github.com/epi2me-labs/wf-basecalling.
46. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 2023;39(5).
47. Labs EM. Human variation workflow 2024 [Available from: https://github.com/epi2me-labs/wf-human-variation.
48. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094-100.
49. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 2018;34(5):867-8.
50. Su J, Zheng Z, Ahmed SS, Lam TW, Luo R. Clair3-trio: high-performance Nanopore long-read variant calling in family trios with trio-to-trio deep neural networks. Brief Bioinform. 2022;23(5).
51. Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, et al. Detection of mosaic and population-level structural variants with Sniffles2. Nat Biotechnol. 2024;42(10):1571-80.

52. Bonfield JK, Marshall J, Danecek P, Li H, Ohan V, Whitwham A, et al. HTSlib: C library for reading/writing high-throughput sequencing data. Gigascience. 2021;10(2).
53. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2).
54. English AC, Menon VK, Gibbs RA, Metcalf GA, Sedlazeck FJ. Truvari: refined structural variant comparison preserves allelic diversity. Genome Biol. 2022;23(1):271.
55. Chiu R, Rajan-Babu IS, Friedman JM, Birol I. Straglr: discovering and genotyping tandem repeat expansions using whole genome long-read sequences. Genome Biol. 2021;22(1):224.
56. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks–a publishing format for reproducible computational workflows. Positioning and power in academic publishing: Players, agents and agendas: IOS press; 2016. p. 87-90.
57. Van Rossum G, Drake FL. Introduction to python 3: python documentation manual part 1: CreateSpace; 2009.
58. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357-62.
59. McKinney W, editor Data structures for statistical computing in Python. SciPy; 2010.
60. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007;9(3):90-5.
61. Waskom ML. Seaborn: statistical data visualization. Journal of Open Source Software. 2021;6(60):3021.
62. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261-72.
63. Pedregosa F. Scikit-learn: Machine learning in python Fabian. Journal of Machine Learning Research2011. p. 2825.
64. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. 2010.
65. London KsC. King's Computational Research, Engineering and Technology Environment (CREATE) 2022 [Available from: https://doi.org/10.18742/rnvf-m076.