OXFORD

*Briefings in Bioinformatics*, 2024, **25(5)**, bbae441
https://doi.org/10.1093/bib/bbae441
**Review**

# Comprehensive assessment of long-read sequencing platforms and calling algorithms for detection of copy number variation

Na Yuan [1,2] and Peilin Jia [1,2,*]

[1]National Genomics Data Center, China National Center for Bioinformation, Beichen West Road, Chaoyang District, Beijing 100101, China
[2]Beijing Institute of Genomics, Chinese Academy of Sciences, Beichen West Road, Chaoyang District, Beijing 100101, China
[*]Corresponding author. Peilin Jia, National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China. E-mail: pjia@big.ac.cn

## Abstract

Copy number variations (CNVs) play pivotal roles in disease susceptibility and have been intensively investigated in human disease studies. Long-read sequencing technologies offer opportunities for comprehensive structural variation (SV) detection, and numerous methodologies have been developed recently. Consequently, there is a pressing need to assess these methods and aid researchers in selecting appropriate techniques for CNV detection using long-read sequencing. Hence, we conducted an evaluation of eight CNV calling methods across 22 datasets from nine publicly available samples and 15 simulated datasets, covering multiple sequencing platforms. The overall performance of CNV callers varied substantially and was influenced by the input dataset type, sequencing depth, and CNV type, among others. Specifically, the PacBio CCS sequencing platform outperformed PacBio CLR and Nanopore platforms regarding CNV detection recall rates. A sequencing depth of 10x demonstrated the capability to identify 85% of the CNVs detected in a 50x dataset. Moreover, deletions were more generally detectable than duplications. Among the eight benchmarked methods, cuteSV, Delly, pbsv, and Sniffles2 demonstrated superior accuracy, while SVIM exhibited high recall rates.

**Keywords**: long-read sequencing; CNV calling; benchmark

## Introduction

Structural variation (SV) represents a vital component of genetic diversity. It is characterized by genomic variations typically exceeding 50 bp, encompassing insertions, deletions, duplications, inversions, translocations, and other complex rearrangements [1–3]. Among the various types of structural variations, copy number variation (CNV) plays a prominent role, accounting for ~4.8%–9.5% of the human genome [4]. CNVs include relatively large deletions or duplications of DNA fragments and are typically defined as events longer than 1 kbp, often ranging up to 5 Mbp. CNVs exert a profound influence on gene phenotypes, e.g. through alterations in gene copy numbers or coding sequences. Thus, they can lead to the onset of diseases or increase susceptibility to diseases [5–10]. Notably, the frequency of CNVs surpasses that of single-nucleotide polymorphisms (SNPs), making CNVs a significant contributor to variations in disease susceptibility [11].

High-throughput sequencing technology has significantly advanced the development of novel methodologies and approaches for investigating genomic structural variation. Second-generation sequencing with short reads offers high throughput and cost-effectiveness. Yet, its capacity to detect tandem repeats, GC-rich regions, highly polymorphic regions, highly homologous segments, or extensive SVs is notably constrained [12–15]. This limitation impedes the comprehensive characterization of genomic alterations using data generated by short-read sequencing platforms. Recently, long-read sequencing platforms [16], such as the Pacific Biosciences (PacBio) and Oxford Nanopore Technologies, have been well developed and applied to generate long sequencing reads. These platforms implemented methods such as single-molecule real-time (SMRT) sequencing for PacBio and nanopore-based sequencing for Oxford Nanopore. These techniques do not require amplification or fragmentation, thus providing a more comprehensive view of genomic regions. Long read length facilitates SV detection [2, 13, 17] by increasing the likelihood of capturing an entire SV event within a single read. This reduces the reliance on inference or assembly-based methods for SV detection and enables the spanning of repetitive sequences and complex genomic regions. Additionally, longer reads reduce the ambiguity associated with mapping short reads, thereby improving the accuracy and sensitivity of SV detection using long-read sequencing platforms. Indeed, some studies have already leveraged long-read sequencing technologies to identify SVs in large population cohorts [18–20].

Several methods have been developed to detect SVs based on long-read sequencing platforms, as summarized in Table 1. Among these, cuteSV [21] employs tailored techniques to collect signatures representing various types of SVs and utilizes a clustering-and-refinement methodology to sensitively detect

**Received:** February 1, 2024. **Revised:** July 9, 2024. **Accepted:** August 25, 2024

Table 1. Summary of copy number variation detection tools used in this benchmark

| Tool | Version | Language | Latest update | Description | SVs detected | Citations | URL |
|---|---|---|---|---|---|---|---|
| cuteSV | 1.0.13 | Python | 2021–1 | Split-read and intra-read signatures | DEL, DUP, INS, INV | 125 | https://github.com/tjiangHIT/cuteSV |
| Delly | 0.8.5 | C++ | 2022–4 | Split-reads and read-depth evidence | DEL, DUP, INS, INV | 1,421 | https://github.com/dellytools/delly |
| NanoSV | 1.2.4 | Python | 2019–9 | Split-read signatures and breakpoint evidence | DEL, DUP, INS | 222 | https://github.com/mroosmalen/nanosv |
| NanoVar | 1.4.1 | Python | 2021–10 | Spit-read signatures and coverage evidence | DEL, DUP, INS, INV | 59 | https://github.com/cytham/nanovar |
| PBHoney | 15.8.24 | Python | 2020–10 | Unmapped split-read tails and intra-read discordance | DEL, INS | 122 | http://sourceforge.net/projects/pb-jelly/ |
| pbsv | 2.4.0 | Python,C++ | 2020–10 | Split-read and intra-read signatures | DEL, DUP, INS, INV, CNV | 787 | https://github.com/PacificBiosciences/pbsv |
| Sniffles2 | 2.2.0 | C++ | 2023–7 | Split-read signatures and coverage evidence | DEL, DUP, INS, INV, CPX | 841 | https://github.com/fritzsedlazeck/Sniffles |
| SVIM | 1.4.2 | Python | 2021–1 | Split-read and intra-read signatures | DEL, DUP, INS, INV | 142 | https://github.com/eldariont/svim/ |

Note: The citation count data is from the Web of Science platform as of March 7, 2024.

SVs based on these signatures. Delly [22] leverages split-reads and read-depth information to achieve sensitive and accurate delineation of genomic rearrangements throughout the genome. NanoVar [23] employs a neural network–based algorithm to achieve high-confidence SV detection and SV zygosity estimation for all classes of SVs. Specifically, NanoSV [24] identifies split and gapped-aligned reads and clusters the reads according to the orientations and genomic positions of the read segments to define breakpoint junctions of candidate SVs. It is further enhanced by filtering out SVs by clustering the alignment segments that align to the same genomic regions based on several key characteristics, including coverage depth, read length, orientation, and size. PBHoney [25] utilizes intraread discordance and soft-clipped tails of long reads to identify SVs with precision. pbsv [16] is one of the most widely used software programs for calling and analyzing SVs. It gathers signatures from sequences aligned to the reference genome per subreads. Sniffles2 [26] exerts repeat-aware clustering coupled with a fast consensus sequence and coverage-adaptive filtering to improve the accuracy of the germline SV calls. SVIM [27] takes an integrated approach, utilizing information from across the genome to precisely distinguish similar events, including tandem and interspersed duplications and simple insertions.

Previous investigations have assessed SV detection tools tailored for long-read sequencing, focusing on parameters such as specificity, sensitivity, and computational demands [21, 28–31]. These studies have diligently explored the merits and limitations of these tools. However, it is important to note that many of these studies have primarily relied on simulated datasets or have been limited to datasets from single sequencing platforms. Consequently, research on real sample datasets encompassing various sequencing data types remains relatively scarce. Furthermore, it is worth noting that some tools have been updated, and new methods have emerged, including specialized long-read sequencing tools such as Sniffles2 [26]. These newly developed tools have not yet been systematically evaluated, especially in the long-read sequencing landscape. In addition, there has been a notable deficiency in the systematic evaluation of CNVs pertaining to the gain or loss of DNA segments in the context of long-read sequencing platforms in previous research endeavors. This underscores the need for comprehensive evaluations to enhance our understanding of CNV detection capabilities associated with long-read sequencing technologies.

In this study, we benchmarked eight widely used SV detection tools for their performance in CNV detection. The assessment was conducted using a comprehensive resource including 15 simulation datasets and 22 datasets from nine publicly available samples representing three long-read sequencing data types. To ensure the robustness of our analyses, we ensembled high-quality CNV reference datasets from multiple software programs using real sequencing data to serve as gold-standard datasets. Our evaluation encompassed a series of comprehensive assessments designed to measure CNV detection performance across real sample datasets and simulated datasets. The various long-read sequencing platforms involved in this study also enabled us to systematically analyze and compare the CNV detection capabilities of the selected SV detection tools under different conditions.

## Materials and Methods

### Public long-read sequencing datasets collection

We downloaded the long-read sequencing datasets from nine publicly available samples (Additional file 1: Table S1). For sequencing files in bas.h5 format, we filtered out low-quality reads (i.e. those with read score <0.8) and converted these files to FASTA format using bash5tools.py from pbh5tools (https://github.com/edawson/pbh5tools, v0.1.0). For sequencing files in FASTQ format, we filtered out low-quality reads using Filtlong (https://github.com/rrwick/Filtlong, v0.2.1). For all sequencing files, we filtered out reads with a length <1 kbp and converted them to FASTA format. Notably, due to varying sequencing depths across different samples on the same platform from the original study design, we performed downsampling for each sample to match the minimum sequencing depth observed across all samples. This approach ensured that all original samples were included in subsequent benchmark analyses, resulting in uniformly downscaled datasets. Seqtk (https://github.com/lh3/seqtk, version 1.2-r94) was utilized for downsampling based on FASTA files.

## Read mapping and copy number variation identification

All 22 datasets were aligned to the human reference genome hg19 (https://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/latest/) using minimap2 (version 2.20-r1061) or pbmm2 (https://github.com/PacificBiosciences/pbmm2, version 1.7.0). Default parameters were employed for both tools, except for those specific to sequencing platforms. Specifically, for minimap2, the parameter "-ax" was set to "map-pb" for PacBio CLR datasets, "map-hifi" for PacBio HiFi/CCS, and "map-ont" for Nanopore datasets, respectively. Similarly, pbmm2 used the "–preset" parameter adjusted accordingly. BAM files were sorted by genomic coordinates using Samtools [32] (version 1.9).

Throughout this work, we assessed eight CNV callers. They were cuteSV (v1.0.13), Delly (v0.8.5), NanoSV (v1.2.4), NanoVar (v1.4.1), PBHoney (15.8.24), pbsv (v2.4.0), Sniffles2 (v2.2.0), and SVIM (v1.4.2). All tools identified CNVs from the BAM files. Notably, pbsv required BAM files generated by pbmm2, while the other seven CNV callers utilized BAM files from minimap2, a widely adopted alignment tool.

For each tool, specific configurations were applied: cuteSV utilized parameters "–max_cluster_bias_INS," "–diff_ratio_merging_INS," "–max_cluster_bias_DEL," "–diff_ratio_merging_DEL," and "–genotype -s 3" adjusted according to the input dataset types. Delly filtered CNVs based on the following criteria: selecting SVTYPE "DEL" or "DUP," setting FILTER to "LowQual," and supporting reads more than three. NanoSV used default settings from config.ini, while NanoVar adjusted parameter "-x" based on dataset types. PBHoney executed commands "tails" and "spots" universally. pbsv executed "discover" and "call" commands universally. Sniffles2 utilized configuration "-t 10 –minsvlen 1000" across all datasets. SVIM employed alignment mode across all datasets. All methods retained only CNVs exceeding 1 kbp in length.

Runtime performance was assessed using the shell command time [GNU bash, version 4.2.46 (1) - release (x86_64-redhat-linux-gnu)] on CNV calling shell scripts. The analyses were conducted on a CentOS Linux release 7.2.1511 platform, equipped with Intel Xeon E7-4850 v3 Central Processing Unit (CPUs) with 10 cores operating at a memory frequency of 1.6 GHz.

## Evaluation of copy number variation callers using simulated long reads

To validate the CNV callers effectively, we simulated data using Sim-it (version 1.3.3) [28]. We employed a catalog of 24 600 SVs, including 10 031 deletions, 857 duplications, 10 469 insertions, 170 inversions, and 3073 complex substitutions, originally detected from the NA19240 sample and accessible through dbVAR (nstd152: https://www.ncbi.nlm.nih.gov/dbvar/studies/nstd152/). The simulations involved generating PacBio CCS, PacBio CLR, and Nanopore reads for GRCh38 (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/latest/) at a sequencing depth of 50x. These long reads were aligned to GRCh38 using pbmm2 (for pbsv input) or minimap2 (for the other seven callers). Subsequently, we performed downsampling using Samtools on the aligned BAM files to generate datasets at sequencing depths of 10x, 20x, 30x, and 40x for PacBio CCS, PacBio CLR, and Nanopore platforms, respectively. These datasets were then used to evaluate the impact of sequencing depths on various CNV calling tools.

We also generated three replicates of simulated datasets at 50x depth for each platform using VCF [33] files from chromosome 1 of real samples (HG00514, HG00733, and NA19240) via Sim-it. Furthermore, SV-inserted haplotype sequences were generated using Sim-it, and pbsim3 (version 3.0.0) [34] was utilized to simulate 50x sequencing data with varying error rates: 0.1%, 0.2%, 0.5%, 1%, 2.5%, 5%, 7.5%, and 15%.

To explore the potential advantages of longer read lengths offered by Nanopore and PacBio CLR sequencing platforms in CNV detection, we employed pbsim3 to generate three Nanopore samples at 50x coverage with average read lengths of 40, 45, and 50 kbp, as well as three PacBio CLR samples at 50x coverage with average read lengths of 15, 20, and 25 kbp. These were compared with PacBio CCS samples at 50x coverage using the default average length of 9 kbp.

Furthermore, we evaluated cuteSV and Sniffles2 with varying configurations of critical parameters [–minsupport (-s) and –minsvlen (-l)] across different depths of PacBio CCS, PacBio CLR, and Nanopore datasets. True positive CNV calls were defined as those with ≥50% reciprocal overlaps with the reference SVs in the simulated data.

## Evaluation of copy number variation callers using the NA12878 copy number variation set as the gold-standard set

To construct a high-confidence set of CNVs using the NA12878 sample, we curated four CNV datasets including verified deletions (DELs) and duplications (DUPs) sourced from various references. First, we retrieved CNVs for NA12878 from the DGV Gold Standard Variants via the Genome Variation Database (http://dgv.tcag.ca/dgv/app/home) (#DELs: 5805, #DUPs: 1710). Second, we downloaded CNVs from the DNBSEQ platform (#DELs: 3168, #DUPs: 344) [32]. Third, we collected a previously published SV benchmark set for NA12878 from the File Transfer Protocol (FTP) site with Polymerase Chain Reaction (PCR) validation [33], which only included deletions (#DELs: 2676, #DUP: 0). Fourth, we incorporated long-read SV data for NA12878 with PCR validation (#DEL: 28, #DUP: 0) [34]. For each dataset, CNVs with lengths <1 kbp were excluded. To ensure data robustness [29], we collected all deletions from all four datasets and merged those overlapping sequences exceeding 80%. Similarly, we collected all duplications from two datasets and merged under the same criteria. The final gold-standard set included 1174 deletions and 605 duplications (Additional file 2).

We downloaded BAM files for NA12878 generated from three long-read sequencing platforms, i.e. PacBio CCS, PacBio CLR, and Nanopore, and applied all eight CNV callers to detect CNVs in this sample. For systematic evaluation, two analytical approaches were conducted. First, the whole human genome was divided into 10 Mbp sliding windows, with the number of detected deletions and duplications by each caller in each window computed. Additionally, the length of regions covered by detected deletions or duplications within each window was summarized. These analyses facilitated comparison of CNVs detected by different tools within predefined genomic windows. Second, detected CNVs from each tool were compared against the gold-standard CNV sets. Following the criteria used in previous studies [29, 38], we considered a detected CNV as positive if it overlapped with those in the gold-standard set for >50% of either the detected CNV or the gold-standard CNV. Deletions and duplications were evaluated separately.

Furthermore, to assess the impact of sequencing depth on various CNV calling tools, BAM files of NA12878 were subsampled at depths of 5x, 10x, 15x, 20x, 25x, and 30x using Samtools. CNV calling processes were repeated for each tool across BAM files generated from different sequencing platforms at each depth. For

evaluating CNV detection performance based on length distribution, CNV calling tools were compared across datasets from different sequencing platforms using five consecutive length intervals: 1–5 kbp, 5–10 kbp, 10–50 kbp, 50–100 kbp, and 100 kbp–1 Mbp.

### Evaluation using other public samples without gold-standard copy number variation data

In addition to the NA12878 sample, we also collected datasets from eight other samples for evaluating CNV callers, obtained from various sequencing platforms. We employed two distinct alignment tools and eight CNV calling tools, resulting in diverse candidate CNV sets.

Due to the absence of a gold-standard CNV set for these samples, we adopted an alternative strategy to define consensus CNVs detected by different tools across multiple sequencing platforms. Specifically, CNVs overlapping by >80% in length, identified by three or more tools, were considered as true CNVs, with deletions and duplications processed separately. This approach yielded highly reliable CNV sets for each sample across different sequencing platforms (Additional file 3).

To evaluate tool performance for each dataset, we used these highly reliable CNV sets as references. Specifically, we compared the chromosome locations and types of CNVs detected by each tool with the highly reliable CNV set for the respective sample. CNVs were validated if they exhibited an overlapping length accounting for at least 50% of the total length and if their types were consistent with those in the highly reliable CNV set.

### Evaluation of CNV callers

To evaluate the performance of various tools on different datasets, we calculated the precision, recall, and F1 score for each tool detected in each dataset. These measurements were calculated as follows, respectively.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP indicates true positive, FP indicates false positive, and FN indicates false negative.

## Results

### Overview of the study design

This study aimed to assess the performance of existing SV tools in calling CNVs from long-read data generated by various long-read sequencing platforms. We formulated our benchmark pipeline as a five-step process, outlined below.

- Step 1: We collected long-read sequencing datasets from multiple sequencing platforms, including PacBio CCS, PacBio CLR, Nanopore, and Ultra-long Nanopore, ensuring comprehensive representation of current sequencing technologies.
- Step 2: Rigorous quality control measures and carefully designed dataset sampling yielded datasets for eight samples from PacBio CCS, seven from PacBio CLR, four from Nanopore, and three from Ultra-long Nanopore.
- Step 3: Simulation datasets were generated for three platforms to provide a complementary assessment to those from the real datasets.
- Step 4: Read alignment was performed using two distinct methods, followed by CNV identification using eight CNV callers.
- Step 5: We systematically evaluated the performance of CNV calling tools across all collected samples.

The complete workflow of this study is illustrated in Fig. 1, with further details provided in the MATERIALS AND METHODS section.

### Generation of benchmark datasets

We obtained long-read sequencing datasets from nine publicly available samples across four platforms (Additional file 1: Table S1 and Table S2). These samples included four of European (EUR: HG002, NA12878, CHM13, and CHM1), three of East Asian (EAS: HG00514, HX1, and HG005), one of Ad Mixed American (AMR: HG00733), and one of African (AFR: NA19240). Except for CHM1 and HG005, all other samples were sequenced using two or more platforms (Table S1).

To prepare the raw sequencing datasets for benchmarking, we conducted a series of data processing steps, including quality control, filtering, format conversion, and standardizing sampling depth across samples from the same sequencing platform. Ultimately, we curated eight datasets for PacBio CCS with a sequencing depth of 25x, excluding the HX1 sample, which lacked PacBio CCS data. Additionally, we curated seven datasets for PacBio CLR at a sequencing depth of 20x, excluding CHM1 and HG005 as they had no PacBio CLR sequencing datasets. Furthermore, four datasets were curated for Nanopore at a sequencing depth of 25x, and three datasets for Ultra-long Nanopore at a sequencing depth of 30x.

### Selection of read alignment tools and copy number variation callers

We performed read alignment using minimap2 [35] and pbmm2 (Additional file 1: Table S3). For CNV calling from long-read sequencing data, we employed eight widely used and recently updated tools, including cuteSV, Delly, NanoSV, NanoVar, PBHoney, pbsv, Sniffles2, and SVIM (Table 1). Specifically, due to the specific requirement by the CNV caller pbsv, we used the alignment output from pbmm2 as the input for pbsv. For all the other tools, we used the alignment output from minimap2, as minimap2 had been widely employed for long-read alignment.

### Evaluation of copy number variation callers using simulated long reads

To evaluate the performance of the CNV callers, we generated simulated datasets at sequencing depths of 10x, 20x, 30x, 40x, and 50x for PacBio CCS, PacBio CLR, and Nanopore datasets, respectively. Notably, for simulation data, we did not use Ultra-long Nanopore. For the number of detected CNVs (Fig. S1 and Additional file 1: Table S4), most tools exhibited an increased number of CNVs with higher sequencing depth, except NanoSV and PBHoney. For example, Sniffles2 detected 1827 CNVs at 10x (accounting for 85% of the total CNVs detected at 50x), 2008 at 20x (93% of 50x), 2091 at 30x (97% of 50x), 2123 at 40x (99% of 50x), and 2141 at 50x in PacBio CCS dataset.

Unexpectedly, pbsv identified tens of thousands of CNVs at 40x and 50x sequencing depths on both PacBio CCS and Nanopore datasets, resulting in extremely lower accuracy. Conversely, NanoSV detected only a minimal number of duplications and none were overlapping with true CNVs, resulting in nearly 0%

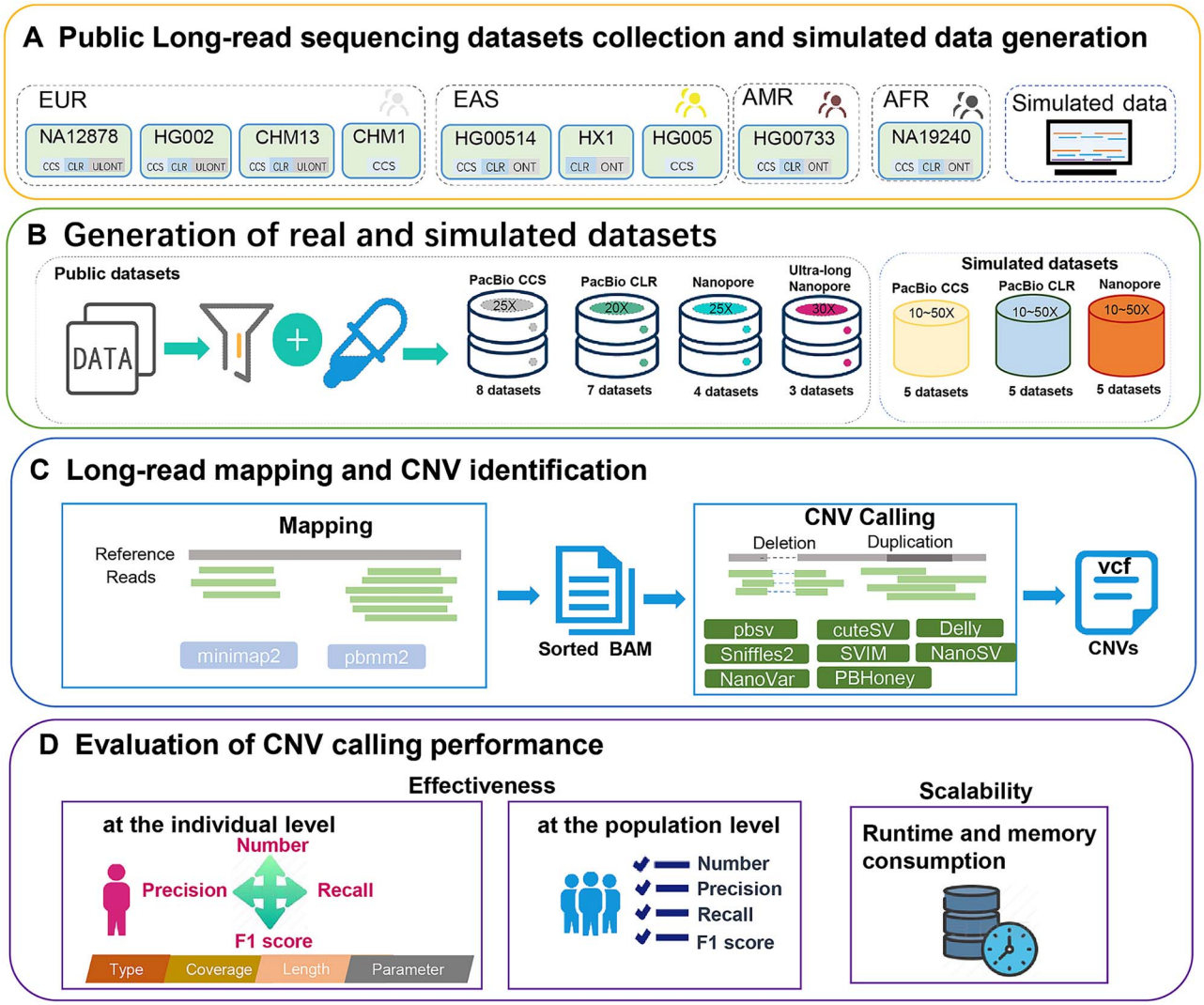


Figure 1. Flowchart of the analysis procedure in this study. (A) Collection of publicly available long-read datasets and generation of simulated datasets. (B) Generation of real and simulated datasets on different sequencing platforms. (C) Mapping of long-reads and detection of CNVs using multiple tools. (D) Evaluation of different tools on various platform datasets.

accuracy. PBHoney also detected fewer CNVs compared to other tools, leading to lower accuracy and recall.

These observations may be attributed to potential systematic biases in simulated long-read datasets, the inherent limitations of the CNV calling algorithms, or challenges related to the specific characteristics of the CNVs present in simulated datasets.

The results in Fig. 2 showed that in terms of accuracy, Sniffles2 outperformed other tools across all sequencing depths for three sequencing platforms, followed by cuteSV and Delly. As for the recall rate, except for the poor performance of NanoSV and PBHoney, the performance of other tools was relatively consistent. Regarding the F1 score, all tools performed similarly except for NanoSV, PBHoney, and pbsv.

To further evaluate tools performance, we simulated three replicate datasets of chromosome 1 at a sequencing depth of 50x for each platform. As shown in Fig. S2, Sniffles2 exhibited slightly higher accuracy across the datasets from all three platforms. PBHoney showed strong performance on PacBio CCS and CLR datasets, while other tools performed comparably. Regarding the recall rate, PBHoney demonstrated a lower recall rate, while the performance of the other tools was comparable. For the F1 score, Sniffles2 consistently outperformed other tools across all three datasets.

Additionally, to further investigate the impact of varying error rates on CNV detection, we generated multiple simulated sequencing datasets at the depth of 50x with error rates ranging from 0.1% to 15%. Figure S3 illustrated the relationship between the error rate and precision, recall, and F1 score, respectively. We observed a negative correlation between precision, recall, and F1 score with the error rate across most tools. Among them, SVIM showed the strongest correlation coefficient between error rate and F1 score (Pearson's $r = -0.786$, $p = 0.02$). Other tools, such as cuteSV and Sniffles2 also showed significant negative correlations with error rates (correlation coefficients below −0.733 and p-values less than 0.05).

In terms of CNV types (Fig. S4 and Additional file 1: Table S4), results across all three sequencing platforms showed that deletions were more detectable than duplications. Specifically, for deletions, the recall rate mostly exceeded 0.8, while the precision was around 0.5. Sniffles2 showed the highest accuracy in detecting deletions, achieving a precision of 0.593 on the PacBio CCS dataset. NanoVar exhibited the highest recall rate for deletions,

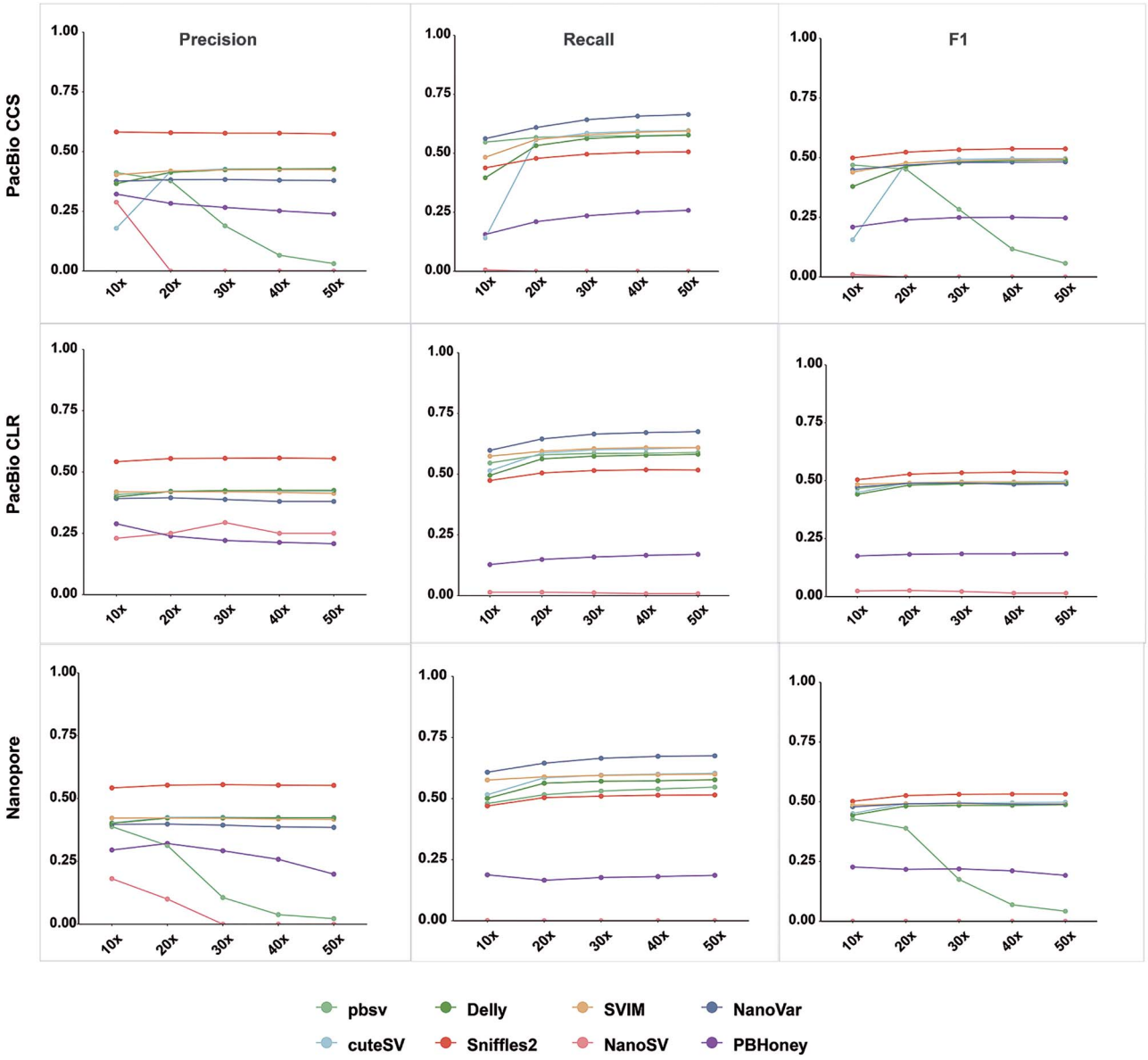


Figure 2. Benchmark results for CNV detection on simulated datasets at different depths, with PacBio CCS (top), PacBio CLR (middle), and Nanopore datasets (bottom).

reaching 0.931 on the PacBio CLR dataset. In contrast, detection of duplications across all tools yielded a precision around 0.3, with generally low recall rates, mostly below 0.2.

In calling CNVs, two parameters played critical roles: the minimal number of supporting reads ("min_support" or "-s") and the minimum size for detecting CNVs ("–min_size" or "-l"). Evaluation using PacBio CCS datasets (Fig. S5 and Additional file 1: Table S5) across various sequencing depths demonstrated that Sniffles2 consistently achieved accuracy above 0.55 at 20x (-s 3), 30x (-s 3 and -s 5), 40x (-s 5), and 50x (-s 5 and -s 10). In contrast, cuteSV exhibited relatively lower accuracy, achieving only 0.419 at 50x (-s 10). For the recall rate, cuteSV occasionally showed slightly higher performance compared to Sniffles2. For example, at 30x (-s 3), cuteSV exhibited a recall rate of 0.585 versus 0.551 by Sniffles2. Further analysis indicated that adjusting the "-l" parameter to larger values in cuteSV resulted in decreased accuracy and recall rates for the same dataset. Conversely, Sniffles2 maintained nearly constant accuracy with changes in "-l," although with a decrease in recall rates. Notably, Sniffles2 consistently achieved an accuracy around 0.6, which was higher compared to cuteSV under similar parameter settings, with slight differences in recall rates.

## Evaluation of NA12878 copy number variation detection based on long-read whole-genome sequencing by the PacBio and Nanopore platforms

We further evaluated CNV detection tools across different long-read sequencing platforms using the gold-standard CNVs assembled for NA12878, encompassing 1174 deletions and 605 duplications. Previous studies utilized three types of long-read sequencing platforms to generate NA12878 data (Additional file 1: Table S6): PacBio CCS (depth: 25x), PacBio CLR (depth: 20x), and Ultra-long Nanopore (depth: 30x). All eight CNV callers were applied to these datasets.

For the PacBio CCS dataset, seven tools (excluding NanoSV) detected a comparable number of CNVs, ranging from 1055 (by pbsv) to 2016 (by PBHoney), averaging 1413 CNVs. In contrast,

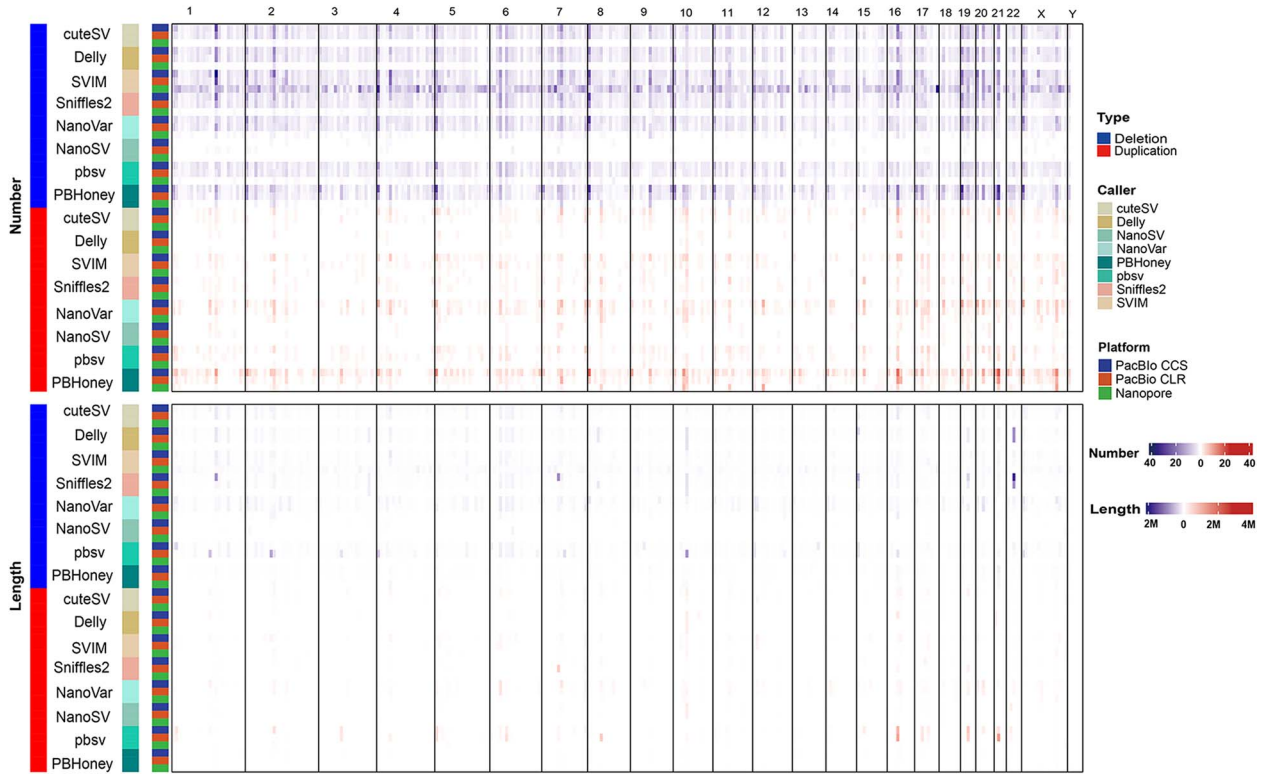


Figure 3. Heatmap of CNVs detected by each tool across the whole genome. The numbers and lengths of deletions and duplications detected by eight tools across three long-read sequencing datasets were analyzed per 10 Mb bin across the whole genome. The top section represents the number distribution, while the bottom section shows the length distribution. The left annotation bar in the panel denotes the deletion and duplication types, followed by eight tools and three platform data types, respectively.

NanoSV identified only 65 CNVs. Similarly, for the PacBio CLR dataset, seven tools detected 859 (by pbsv) to 1268 (by SVIM) CNVs, averaging 1078 CNVs (Additional file 1: Table S7), whereas NanoSV identified 69 CNVs. The difference between results from the two platforms was likely influenced by their respective sequencing depths, which were 25x for PacBio CCS and 20x for PacBio CLR. The number of CNVs detected in the Ultra-long Nanopore dataset was generally lower, where seven tools detected up to 306 CNVs, except SVIM, which identified 2279 CNVs. Variations among tools likely stemmed from their algorithmic designs. Notably, SVIM employed a reference-guided assembly approach, split-read alignment, and graph-based representation to effectively detect SVs, potentially explaining its higher CNV count. The average CNV length detected by pbsv and NanoSV on PacBio CCS and CLR datasets were notably longer compared to other tools (Additional file 1: Table S8). Specifically, on the PacBio CCS dataset, pbsv detected CNVs averaging 11.9 kbp, while NanoSV detected CNVs averaging 36.7 kbp, whereas other tools detected CNVs ranging from 2 to 10 kbp.

To further examine the characteristics of CNVs detected by various tools, we calculated the distribution of CNV numbers and lengths using 10 Mbp sliding windows across the whole genome. As shown in Fig. 3, the number and length distribution of CNVs detected by the same tool exhibited similar trends in PacBio CCS and CLR, but were much different from those detected by Ultra-long Nanopore. Notably, cuteSV, Delly, NanoVar, pbsv, Sniffles2, and SVIM consistently detected more CNVs in each window compared to other tools. These differences were statistically significant ($P < 2.2 \times 10^{-16}$ by the Kruskal–Wallis test, Fig. S6 and Additional file 1: Tables S9 and S10), even based on datasets from the same sequencing platform. Therefore, conducting a comprehensive analysis of each tool's detection performance characteristics is essential.

We compared the sensitivity of all tools in detecting CNVs across different sequencing platforms, calculating the percentage of detected CNVs that overlapped with the gold-standard dataset (overlapping ≥50%) out of the total number of detected CNVs. Overall, most tools detected more CNVs using the PacBio CCS and CLR datasets compared to the Ultra-long Nanopore dataset. Specifically, in the PacBio datasets, pbsv, Delly, and Sniffles2 were the top three tools with the largest share (each accounting for ~75%), followed by cuteSV, SVIM, NanoVar, PBHoney, and NanoSV. For the Ultra-long Nanopore dataset, Sniffles2 identified the highest number of gold-standard CNVs (around 75%), followed by pbsv (~ 60%), NanoVar (~50%), and Delly (~ 45%) (Fig. S7).

We also compared the precision, recall, F1 score, and intersection count of eight tools across datasets from three long-read sequencing platforms. For both the measurements of precision and recall, all tools exhibited superior performance on datasets from the PacBio platforms (CCS and CLR) compared to the Ultra-long Nanopore dataset (Fig. 4A). Specifically, for precision, pbsv, Delly, and Sniffles2 demonstrated the highest precision on PacBio datasets, each achieving ~0.75 (Fig. 4A). cuteSV, SVIM, and NanoVar attained around 0.6 precision, while NanoSV and PBHoney yielded <0.4. On the Ultra-long Nanopore platform, Sniffles2 performed best with a precision of 0.696, followed by pbsv at 0.582. The recall rates of all tools were relatively consistent, none exceeding 0.5, with PBHoney and NanoSV below 0.3. The recall rates of Ultra-long Nanopore datasets were lower compared to PacBio, with SVIM achieving the highest recall at 0.182. Comparing the results across different sequencing platforms, PacBio CCS identified slightly more CNVs than PacBio CLR due to a slightly

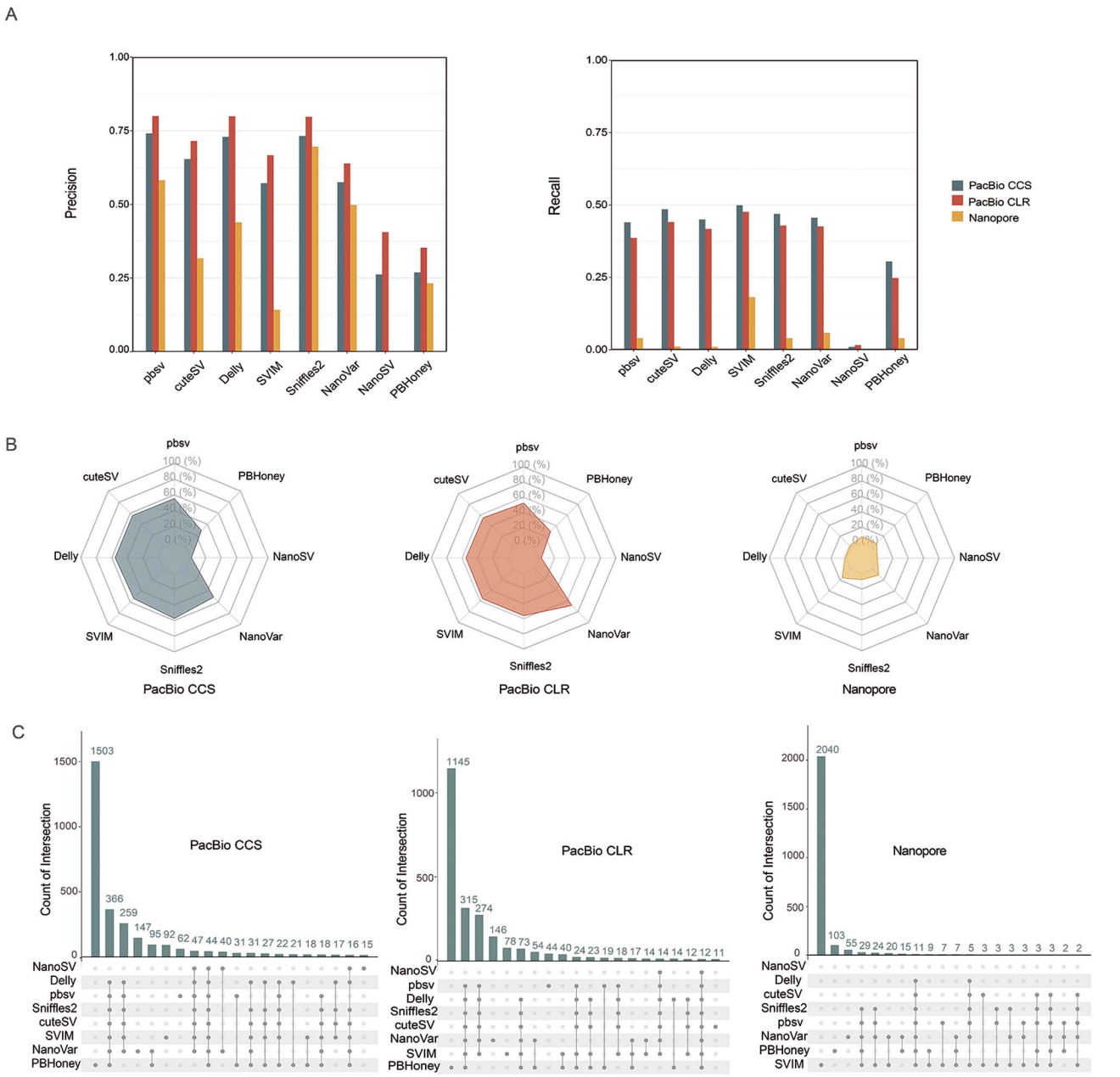


Figure 4. Comparison of the CNV calling performance on various NA12878 long-read sequencing datasets. (A, B) Precision, recall, and F1-score of each tool. (C) The top 20 intersection of CNV calls produced by different tools on diverse sequencing datasets.

higher sequencing depth (Additional file 1: Table S7), resulting in reduced precision and increased recall. Moreover, the higher accuracy observed in PacBio CCS and CLR datasets compared to Ultra-long Nanopore may be attributed to factors related to the error rates (Additional file 1: Table S6), as prior studies have reported higher error rates with Ultra-long Nanopore compared to PacBio [35].

According to the F1 score (Fig. 4B), most tools exhibited comparable performance on PacBio CCS and CLR datasets, generally around 0.5, except that NanoVar was slightly higher on PacBio CLR than CCS. The overall performance of all eight tools on Ultra-long Nanopore was notably poor.

For the overlapping CNVs called by multiple tools (Fig. 4C), CNVs detected by seven tools accounted for 366 and 315 on PacBio CCS and CLR, respectively, while Ultra-long Nanopore only accounted for 11. CNVs detected by two tools were also relatively small on Ultra-long Nanopore. Notably, cuteSV, Delly, pbsv, and Sniffles2 identified fewer unique CNVs, indicating that the CNVs identified by these tools were replicable by other tools. In contrast, on the Ultra-long Nanopore platform, SVIM detected a substantial number of unique CNVs that were not detected by other tools, with 2040 on the Ultra-long Nanopore platform.

The comparison of deletions and duplications detected by various tools across different platforms (Fig. S8) indicated that detecting deletions was generally more robust than duplications across most tools, likely due to fewer duplications in the reference set. Overall, the performance of tools in detecting deletions and duplications remained consistent across PacBio CCS and CLR datasets. Specifically, cuteSV, Delly, pbsv, and Sniffles2 achieved high accuracy approaching 0.8 (Additional file 1: Table S7), while NanoSV and PBHoney showed relatively poorer performance. On the Ultra-long Nanopore dataset, only Sniffles2 showed

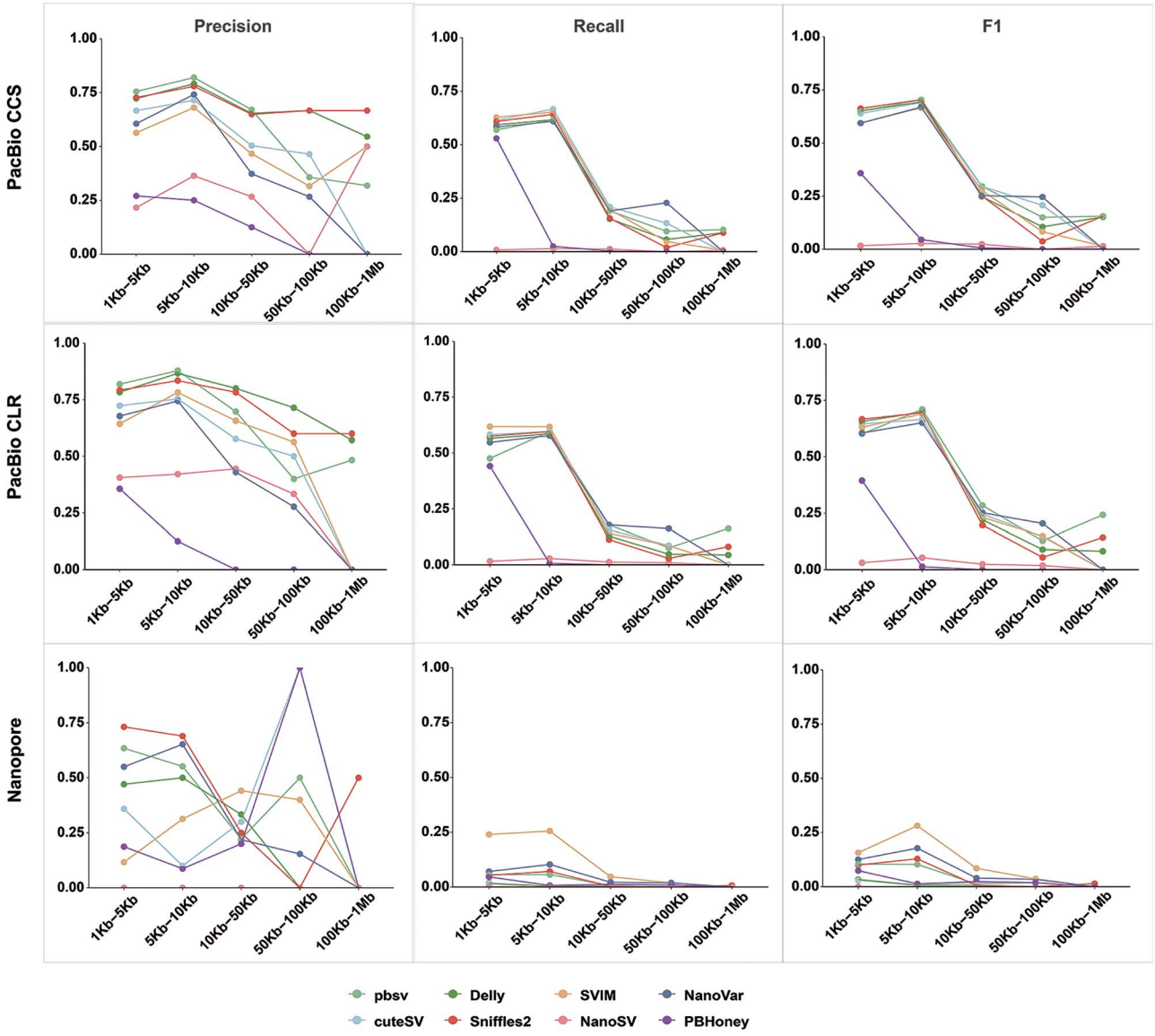


Figure 5. Performance of CNV detection across different size ranges on diverse long-read sequencing platforms.

comparable performance to PacBio CCS and CLR, with other tools experiencing reduced recall rates and slightly lower accuracy compared to PacBio CCS and CLR.

The comparison of deletions and duplications detected by various tools across different platforms (Fig. S8) indicated that deletions were easier to detect than duplications by most tools, likely due to fewer duplications in the reference set. Overall, the performance of various tools in detecting deletions and duplications remained consistent across PacBio CCS and CLR datasets. Specifically, cuteSV, Delly, pbsv, and Sniffles2 exhibited high accuracy approaching 0.8 (Additional file 1: Table S7), while NanoSV and PBHoney showed relatively poorer performance compared to other tools. On the Ultra-long Nanopore dataset, only Sniffles2 showed comparable performance compared to PacBio CCS and CLR, with other tools observing reduced recall rates and slightly lower accuracy compared to PacBio CCS and CLR.

Figure S9 showed the maximum memory usage and computation time of each tool running on the 25x PacBio CCS dataset. Among tools running on a single CPU without multithread support, pbsv and SVIM demonstrated efficient and resource-friendly performance. In contrast, Delly demonstrated high accuracy in CNV calling but required extended computation time primarily due to its reliance on the read depth method. Moreover, NanoSV exhibited a higher memory consumption compared to other tools. Both cuteSV and Sniffles2 are multiprocessing tools. When utilizing 10 threads, they demonstrated notably rapid processing speeds and reduced memory usage compared to alternative tools. Notably, Sniffles2 demonstrated the fastest performance, while NanoVar and NanoSV required slightly higher memory resources.

To assess the tools' performance in detecting CNVs of varying lengths, we categorized CNVs into five intervals from 1 kbp to 1 Mbp (Fig. 5 and Additional file 1: Table S11). On the PacBio CCS dataset, both Delly and Sniffles2 exhibited strong accuracy in detecting CNVs ranging from 1 to 100 kbp, followed closely by pbsv and cuteSV, which showed decreased detection rates for CNVs larger than 50 kbp. Analysis of recall rates indicated similar performance among tools for CNVs under 50 kbp, with differences of <0.15 (Additional file 1: Table S11), except for notably poor performance by PBHoney and NanoSV. Due to fewer detections of CNVs larger than 50 kbp, slight fluctuations in CNV counts led to noticeable changes in F1 scores within the 50–100 kbp and 100 kbp–1 Mbp intervals. On the PacBio CLR dataset, the

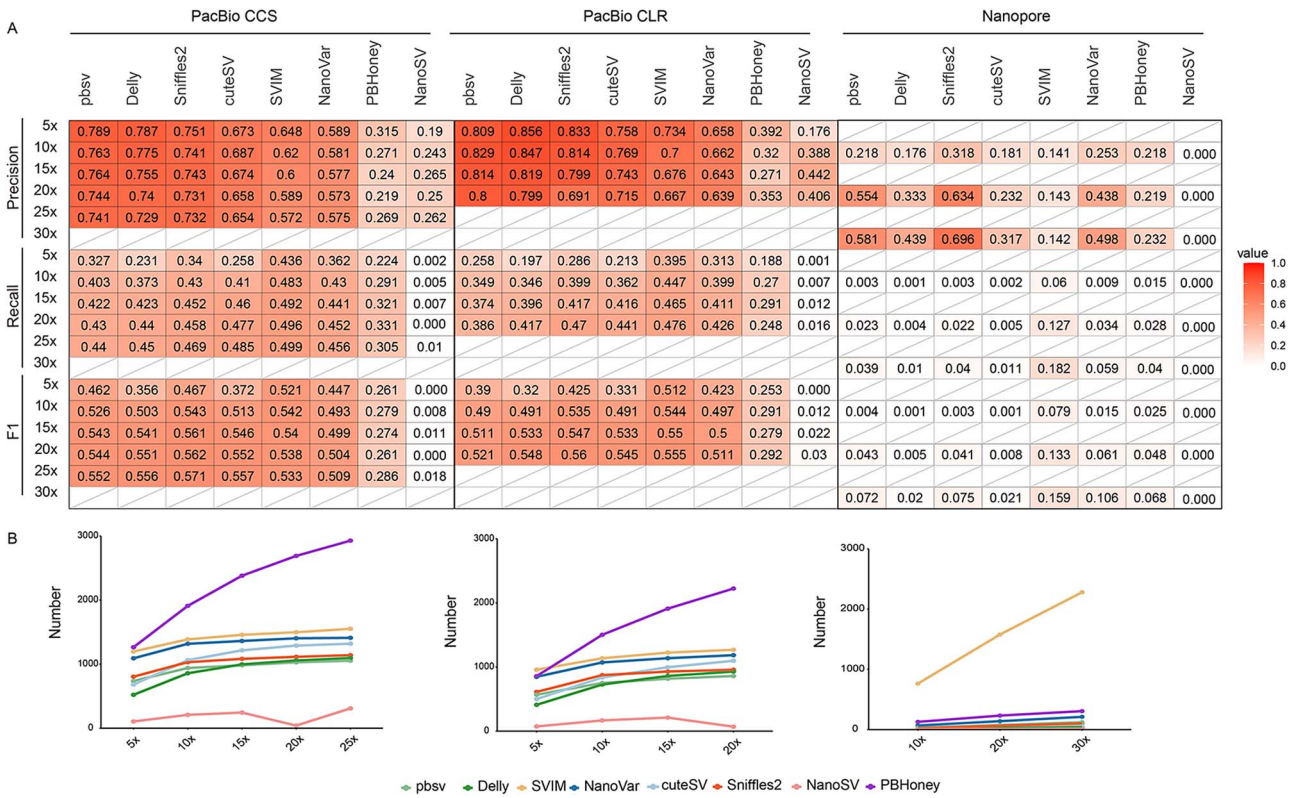


Figure 6. Comparison of CNV detection across different depths on various long-read sequencing platforms. The sampling depths of PacBio CCS datasets are 5x, 10x, 15x, 20x, and 25x; the sampling depths of PacBio CLR are 5x, 10x, 15x, and 20x; and the sampling depths of nanopore are 10x, 20x, and 30x. (A) Precision, recall, and F1-score of each tool at different sequencing depths. The slashed box indicates that the dataset of the current sequencing platform was not sampled at this depth. (B) Number of CNVs detected by each tool at different sequencing depths.

overall performance of each tool was basically consistent with that of the CCS platform, except for PBHoney and SVIM, which showed decreased accuracy in detecting large CNVs. On the Ultra-long Nanopore dataset, all tools exhibited reduced performance. However, Sniffles2 exhibited relatively higher accuracy in detecting ultra-long CNVs compared to other tools within the 1 to 50 kbp range, despite detecting a small total number of 102 CNVs averaging 7 kbp in length (Additional file 1: Tables S7 and S8).

It is widely acknowledged that increasing sequencing depth can improve the CNV detection rate [35]. To assess the impact of sequencing depth on CNV detection, we downsampled datasets from three sequencing platforms to generate new datasets at 5x, 10x, 15x, 20x, 25x, and 30x sequencing depths. We applied identical CNV calling procedures to these downsampled datasets and evaluated the performance using the accuracy, recall, and F1 score (Fig. 6A), and the number of detected CNVs (Fig. 6B) for each dataset. Across all sequencing depths, tools generally exhibited superior performance on PacBio CCS and CLR datasets compared to Ultra-long Nanopore datasets. Accuracy for PacBio CCS and CLR datasets ranged between 0.7 and 0.8, while the highest accuracy of Ultra-long Nanopore was only 0.696 at 30x. For the PacBio CCS dataset, increasing sequencing depth had minimal effect on accuracy, with a slight increase in identified CNVs. The comparison between different tools revealed that pbsv, Delly, and Sniffles2 consistently demonstrated the highest accuracy (darkest color), followed closely by cuteSV, SVIM, NanoVar, PBHoney, and NanoSV. In terms of recall rate and F1 score, cuteSV, NanoVar, pbsv, Sniffles2, and SVIM generally outperformed the other tools. For the PacBio CLR dataset, the sampling results from 5x to 20x indicated that all tools performed better than that of CCS dataset. However, due to the number of CNVs identified by PacBio CLR being less than that of CCS, the recall rate and F1 score slightly decreased. Conversely, most tools performed relatively poorly on Ultra-long Nanopore datasets, even at 30x sequencing depth where accuracy remained below 0.7 for all tools, and the recall rate was extremely low, below 0.2. This was primarily attributed to the limited number of CNVs detected, with most tools identifying very few CNVs, except SVIM, which detected >500 (Fig. 6B). Overall, the performance comparison across different sequencing datasets showed that Delly, pbsv, and Sniffles2 consistently exhibited the highest accuracy across all sampling depths on PacBio CLR datasets, while SVIM and Sniffle2 achieved the highest recall rates and F1 scores on PacBio CCS datasets.

## Evaluation of long-read sequencing copy number variation detection tools for all publicly available samples

We implemented all eight CNV detection tools across a total of 22 datasets derived from nine individuals spanning three sequencing platforms: PacBio CCS (eight datasets), PacBio CLR (seven datasets), and Nanopore (seven datasets, including four Nanopore and three ultra-long Nanopore datasets). Our evaluation focused on assessing the number, precision, recall, and F1 score of CNVs detected by each tool across all individuals (Fig. 7, Additional file 1: Tables S12 and S13). The number of CNVs detected by most tools across datasets from the three sequencing platforms remained consistent, while SVIM detected a large number of CNVs across multiple samples on Nanopore and CLR datasets (Fig. S10), reaching over 8000, resulting in a higher false positive rate in CNV

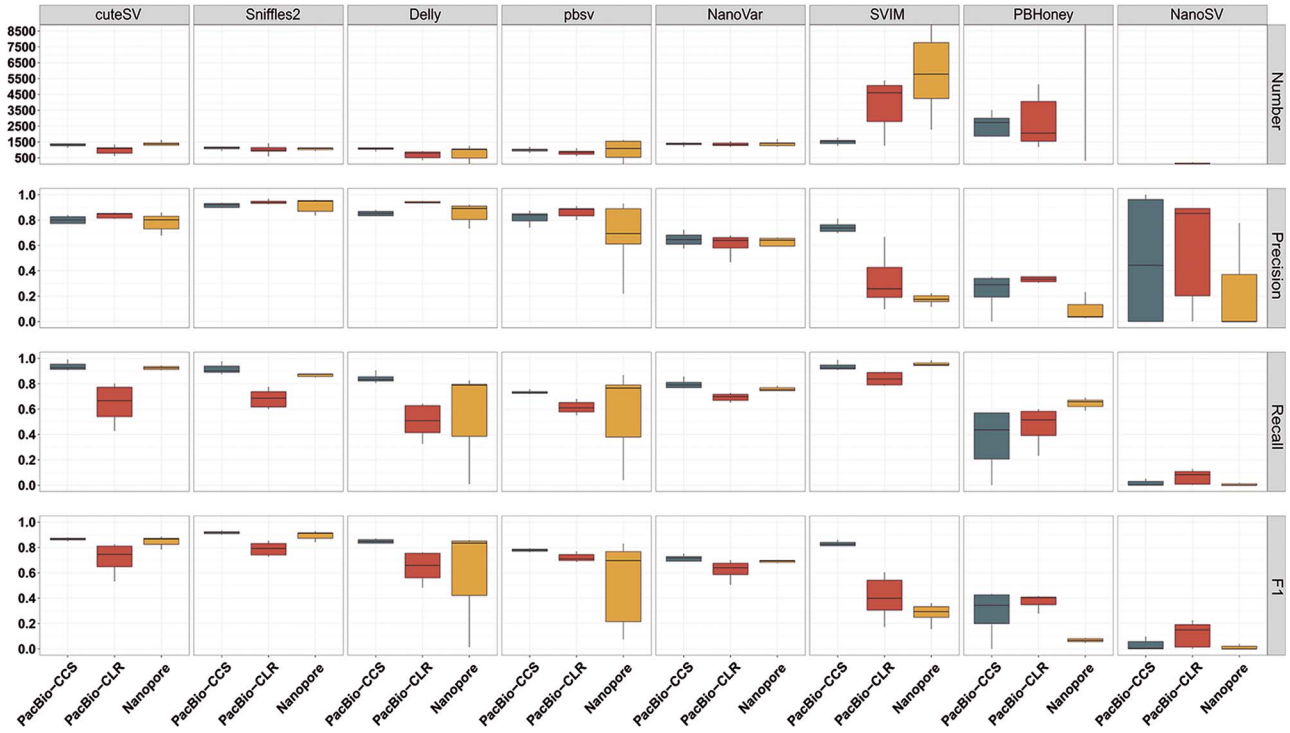


Figure 7. Comparison of CNVs detected by various tools and platforms across all nine samples. Box plot shows CNV numbers, precision, recall, and F1 scores across platforms by different tools. The evaluation of precision, recall, and F1-score is based on the high-confidence sets of CNVs identified jointly by at least three tools across each platform dataset.

detection, and the number of CNVs detected by other tools was relatively stable, generally within 2000.

cuteSV, Delly, pbsv, and Sniffles2, achieved precision levels exceeding 0.8 on both PacBio CCS and CLR datasets. Moreover, their precision was slightly higher on the CLR dataset compared to CCS and Nanopore datasets. SVIM exhibited the highest precision on CCS dataset (~0.8), much lower on CLR dataset (~0.3), and even lower on Nanopore dataset (only 0.1). In contrast, PBHoney and NanoSV showed relatively poor detection performance compared to the other six tools.

The recall rates of NanoVar, pbsv, Sniffles2, and SVIM were generally consistent across all three platform datasets, while cuteSV had a decrease on CLR dataset (~0.2) and Delly showed a decrease and fluctuation on both CLR and Nanopore datasets (fluctuation around 0.3). Combinations with relatively good recall performance included cuteSV on CCS and Nanopore datasets, pbsv and Sniffles2 on Nanopore datasets, and SVIM on CCS and Nanopore datasets, all achieving around 0.8.

All tools, except PBHoney and NanoSV, demonstrated high F1 scores on CCS datasets. cuteSV, NanoVar, pbsv, and Sniffles2 achieved F1 scores surpassing 0.65 on CLR datasets, while Delly and SVIM showed slightly lower and fluctuating scores. Particularly, cuteSV and Sniffles2 stood out as the only tools reaching F1 scores of around 0.8 on Nanopore datasets.

We conducted a comprehensive analysis of CNV calling performance across a spectrum of error rates present in all samples for PacBio CCS, PacBio CLR, and Nanopore (except for Ultra-long Nanopore) datasets. Figure S11 illustrated the relationship between the error rate and precision, recall, and F1 score, respectively. We observed a negative correlation between recall and error rate, with Delly showing a significant correlation (Pearson's $r = -0.578$, $P = .012$). However, we noted a positive correlation between accuracy and error rate, potentially attributed to the advantages of PacBio CLR and Nanopore sequencing technologies in long read lengths for CNV detection. To investigate further, we conducted CNV detection using simulated and real datasets from PacBio CCS, PacBio CLR, and Nanopore. Our findings indicate that PacBio CLR and Nanopore datasets generally exhibited higher accuracy rates compared to PacBio CCS (Fig. S12).

In summary, cuteSV and Sniffles2 achieved high precision and recall across PacBio CCS and CLR datasets. Sniffles2 had the highest precision and SVIM had the highest recall on Nanopore datasets.

## Discussion

To the best of our knowledge, this study represents the first comprehensive comparison of existing CNV detection algorithms tailored to long-read data. Following a careful design, we selected eight CNV callers employing diverse strategies for CNV detection and benchmarked their efficacy. We assessed the performance of these tools using both simulation data and real sequencing data obtained from previous studies. The latter included raw long-read sequencing data from nine publicly available samples for CNV detection. Furthermore, our assessment encompassed datasets generated across multiple sequencing platforms, including PacBio CCS and CLR, Nanopore, and Ultra-long Nanopore.

Several features of the input sequencing datasets significantly influence the accuracy of CNV detection. The sequencing platform notably impacts CNV detection precision and recall. According to our evaluation, datasets with lower sequencing error rates generally exhibited superior CNV detection recall compared to datasets with higher error rates at similar sequencing depths. Specifically, datasets from the PacBio CCS platform demonstrated higher CNV detection recall rates than those from PacBio CLR and

Nanopore platforms. The relatively lower performance observed in the Nanopore dataset compared to others may be attributed to its comparatively lower sequencing quality and higher error rates relative to PacBio platforms. Future studies should consider further validation at a population level using a broader sample base. Additionally, sequencing depth emerged as a crucial factor influencing the accuracy of CNV detection. Our findings indicate that CNVs detected using 10x datasets could account for 85% of the CNVs detected using 50x datasets, while 20x datasets accounted for 90%–95%, and 30x datasets exceeded 95%.

The evaluation of CNV detection tools revealed that cuteSV, Delly, pbsv, and Sniffles2 consistently demonstrated higher accuracy compared to other tools. Notably, on the PacBio CLR platform, these tools detected a higher number of CNVs, resulting in a decrease in recall rates compared to the PacBio CCS platform. Conversely, fewer CNVs detected on the Nanopore platform led to decreased accuracy and recall rates. SVIM exhibited significant platform-dependent differences in detected CNVs, notably showing a decreased F1 score on PacBio CLR and Nanopore platforms relative to PacBio CCS. NanoVar's performance across platforms was relatively moderate. Furthermore, individual sample analysis revealed that CNVs detected by cuteSV, Delly, pbsv, and Sniffles2 were more repeatedly detected by other tools, whereas SVIM identified a substantial number of unique CNVs not reported by others. However, compared with other tools, SVIM had a higher recall rate, which may be due to its extensive detections. This emphasizes the necessity of employing multiple detection tools to comprehensively detect CNVs in practical applications.

The lower recall rates observed across various tools for CNV detection on NA12878 sequencing datasets from different platforms may stem from the incompleteness of the standard dataset, which may not include all true CNVs in NA12878 and may contain some incorrect CNVs. Among the CNVs in the standard dataset, 645 CNVs (428 duplications and 217 deletions) were undetected by any of the eight methods, suggesting these CNVs might be false positives requiring further validation in future studies.

For computational resources, Delly, NanoSV, and NanoVar required slightly longer computing times, whereas other tools completed CNV detection for a sample with a sequencing depth of 30x within 30 min using a single CPU. Notably, Sniffles2 stood out as the fastest tool currently available. For the length distribution of detected CNVs, CNVs with lengths below 100 kbp can be detected by most tools, but there is still a lack of effective detection tools for CNVs larger than 100 kbp. For different CNV types, deletions were generally easier to detect than duplications. Specifically, pbsv achieved nearly 90% accuracy in the PacBio CCS and CLR datasets when analyzing real sample data. However, all tools exhibited lower accuracy in detecting duplications, which was related to the complex genome structure and long variation length. This requires the development of more precise specialized tools to explore more duplications. In our future research, we will enlarge the sample size to identify large duplications. As for the gold-standard CNV set, we currently map them to the reference genome hg19. In the future, we will assemble updated sets of CNVs on GRCh38 to increase the data usability of this research.

This study utilized real sequencing data from publicly available samples, providing a valuable resource of identified CNVs and reliable CNV benchmarks for each sample. Our comprehensive evaluation across nine samples offers insights crucial for future population-level CNV detection using long-read sequencing platforms. In conclusion, our study underscores the significance of considering sequencing data characteristics, platform selection, and tool efficacy when conducting CNV analyses. The resources generated by this research will support advancements in CNV detection within long-read sequencing data.

## Conclusion

Our benchmark results provide a comprehensive performance evaluation for researchers conducting CNV detection using long-read sequencing data. Additionally, the CNVs identified by each tool, along with the highly reliable CNVs for each publicly available sample, serve as valuable data sources for related research.

**Key Points**

- We conducted a comprehensive evaluation of CNV detection tools for long-read sequencing data, encompassing both simulated and real datasets.
- For the sequencing platforms, CNV detection tools showed better recall rates using the data generated from the PacBio CCS platform than those from the PacBio CLR and Nanopore platforms.
- In terms of sequencing depth, a sequencing depth of 10x demonstrated the capability to identify 85% of the CNVs detected in a 50x dataset.
- Among the benchmarked methods, cuteSV, Delly, pbsv, and Sniffles2 demonstrated superior accuracy, while SVIM exhibited high recall rates.
- We collected and processed a comprehensive list of publicly available long-read sequencing data. The CNVs identified by each tool, along with the highly reliable CNVs generated for each sample, represent valuable resources that can significantly contribute to future research endeavors.

## Supplementary data

Supplementary data are available at *Briefings in Bioinformatics* online.

## Author contributions

Na Yuan: Conceptualization, Data curation, Methodology, Software, Investigation, Formal Analysis, Funding Acquisition, Writing - Original Draft. Peilin Jia: Conceptualization, Methodology, Funding Acquisition, Resources, Supervision, Writing - Review & Editing.

Conflict of interest: None of the authors reported financial interests or potential conflicts of interest.

## Funding

This work was supported by Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38010400], National Natural Science Foundation of China [32000460], Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (LCNBI), and ZJLab.

## Data availability

All statistical information for the datasets is included in the additional files. All data and codes are available at Zenodo (https://zenodo.org/records/11257602) and GitHub (https://github.com/Na-Yuan-BIG/long-read-SVcaller-evaluation) for easy access and reproducibility of the evaluation results.

## References


1. Sudmant PH, Rausch T, Gardner EJ. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* 2015;**526**:75–81. https://doi.org/10.1038/nature15394.
2. Chaisson MJP, Sanders A, Zhao X. *et al.* Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun* 2019;**10**:1784. https://doi.org/10.1038/s41467-018-08148-z.
3. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. *Nat Rev Genet* 2020;**21**:171–89. https://doi.org/10.1038/s41576-019-0180-9.
4. Zarrei M, MacDonald J, Merico D, Scherer S. A copy number variation map of the human genome. *Nat Rev Genet* 2015;**16**:172–83. https://doi.org/10.1038/nrg3871.
5. Sekar S, Tomasini L, Proukakis C. *et al.* Complex mosaic structural variations in human fetal brains. *Genome Res* 2020;**30**:1695–704. https://doi.org/10.1101/gr.262667.120.
6. Polley S, Prescott N, Nimmo E. *et al.* Copy number variation of scavenger-receptor cysteine-rich domains within DMBT1 and Crohn's disease. *Eur J Hum Genet* 2016;**24**:1294–300. https://doi.org/10.1038/ejhg.2015.280.
7. Freeman JL, Perry G, Feuk L. *et al.* Copy number variation: new insights in genome diversity. *Genome Res* 2006;**16**:949–61. https://doi.org/10.1101/gr.3677206.
8. Leija-Salazar M, Sedlazeck F, Toffoli M. *et al.* Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. *Mol Genet Genomic Med* 2019;**7**:e564–e575. https://doi.org/10.1002/mgg3.564.
9. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* 2016;**17**:224–38. https://doi.org/10.1038/nrg.2015.25.
10. Beck CR, Carvalho C, Akdemir Z. *et al.* Megabase length hypermutation accompanies human structural variation at 17p11.2. *Cell* 2019;**176**:1310–1324 e10. https://doi.org/10.1016/j.cell.2019.01.045.
11. Handsaker RE, van V, Berman J. *et al.* Large multiallelic copy number variations in humans. *Nat Genet* 2015;**47**:296–303. https://doi.org/10.1038/ng.3200.
12. Song JHT, Lowe CB, Kingsley DM. Characterization of a human-specific tandem repeat associated with bipolar disorder and schizophrenia. *Am J Hum Genet* 2018;**103**:421–30. https://doi.org/10.1016/j.ajhg.2018.07.011.
13. Merker JD, Wenger A, Sneddon T. *et al.* Long-read genome sequencing identifies causal structural variation in a mendelian disease. *Genet Med* 2018;**20**:159–63. https://doi.org/10.1038/gim.2017.86.
14. Zeng S, Zhang MY, Wang XJ. *et al.* Long-read sequencing identified intronic repeat expansions in SAMD12 from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *J Med Genet* 2019;**56**:265–70. https://doi.org/10.1136/jmedgenet-2018-105484.
15. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 2015;**13**:278–89. https://doi.org/10.1016/j.gpb.2015.08.002.
16. Wenger AM, Peluso P, Rowell W. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 2019;**37**:1155–62. https://doi.org/10.1038/s41587-019-0217-9.
17. Mizuguchi T, Suzuki T, Abe C. *et al.* A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. *J Hum Genet* 2019;**64**:359–68. https://doi.org/10.1038/s10038-019-0569-5.
18. Beyter D, Ingimundardottir H, Oddsson A. *et al.* Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. *Nat Genet* 2021;**53**:779–86. https://doi.org/10.1038/s41588-021-00865-4.
19. Wu Z, Jiang Z, Li T. *et al.* Structural variants in the Chinese population and their impact on phenotypes, diseases and population adaptation. *Nat Commun* 2021;**12**:6501. https://doi.org/10.1038/s41467-021-26856-x.
20. De Coster W, Weissensteiner MH, Sedlazeck FJ. Towards population-scale long-read sequencing. *Nat Rev Genet* 2021;**22**:572–87. https://doi.org/10.1038/s41576-021-00367-3.
21. Jiang T, Liu Y, Jiang Y. *et al.* Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol* 2020;**21**:189. https://doi.org/10.1186/s13059-020-02107-y.
22. Rausch T, Zichner T, Schlattl A. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 2012;**28**:i333–9. https://doi.org/10.1093/bioinformatics/bts378.
23. Tham CY, Tirado-Magallanes R, Goh Y. *et al.* NanoVar: accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. *Genome Biol* 2020;**21**:56. https://doi.org/10.1186/s13059-020-01968-7.
24. Cretu SM, van M, Renkens I. *et al.* Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun* 2017;**8**:1326. https://doi.org/10.1038/s41467-017-01343-4.
25. English AC, Salerno WJ, Reid JG. PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics* 2014;**15**:180. https://doi.org/10.1186/1471-2105-15-180.
26. Smolka M, Paulin L, Grochowski C. *et al.* Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol* 2024:1–10. https://doi.org/10.1038/s41587-023-02024-y.
27. Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. *Bioinformatics* 2019;**35**:2907–15. https://doi.org/10.1093/bioinformatics/btz041.
28. Dierckxsens N, Li T, Vermeesch J, Xie Z. A benchmark of structural variation detection by long reads through a realistic simulated model. *Genome Biol* 2021;**22**:342. https://doi.org/10.1186/s13059-021-02551-4.
29. Kosugi S, Momozawa Y, Liu X. *et al.* Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol* 2019;**20**:117. https://doi.org/10.1186/s13059-019-1720-5.
30. Zhou A, Lin T, Xing J. Evaluating nanopore sequencing data processing pipelines for structural variation identification. *Genome Biol* 2019;**20**:237. https://doi.org/10.1186/s13059-019-1858-1.
31. Jiang T, Liu S, Cao S. *et al.* Long-read sequencing settings for efficient structural variation detection based on comprehensive evaluation. *BMC Bioinformatics* 2021;**22**:552. https://doi.org/10.1186/s12859-021-04422-y.
32. Rao J, Peng L, Liang X. *et al.* Performance of copy number variants detection based on whole-genome sequencing by DNBSEQ platforms. *BMC Bioinformatics* 2020;**21**:518. https://doi.org/10.1186/s12859-020-03859-x.

33. Parikh H, Mohiyuddin M, Lam H. *et al.* svclassify: a method to establish benchmark structural variant calls. *BMC Genomics* 2016;**17**:64. https://doi.org/10.1186/s12864-016-2366-2.
34. Pendleton M, Sebra R, Pang A. *et al.* Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat Methods* 2015;**12**:780–6. https://doi.org/10.1038/nmeth.3454.
35. Singh AK, Olsen M, Lavik L. *et al.* Detecting copy number variation in next generation sequencing data from diagnostic gene panels. *BMC Med Genomics* 2021;**14**:214. https://doi.org/10.1186/s12920-021-01059-x.