Computational and Structural Biotechnology Journal 27 (2025) 912–919

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Benchmarking strategies for CNV calling from whole genome bisulfite data in humans

Shanghui Zhao [a,1], Dantong Xu [a], Jiali Cai [a], Qingpeng Shen [a], Mingran He [a], Xiangchun Pan [a], Yahui Gao [a], Jiaqi Li [a,b], Xiaolong Yuan [a,b,c,*]

[a] State Key Laboratory of Swine and Poultry Breeding Industry, National Engineering Research Center for Breeding Swine Industry, Guangdong Provincial Key Laboratory of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou, Guangdong 510642, China
[b] National Center of Technology Innovation for Pigs, Chongqing 402460, China
[c] Centre for Healthy Ageing, Health Futures Institute, Murdoch University, Murdoch, WA 6150, Australia

ARTICLE INFO

Keywords:
Copy number variations
DNA methylation
Whole genome bisulfite sequencing
Human genome

ABSTRACT

It's important to dissect the relationship between copy number variations (CNVs) and DNA methylation, because both greatly change the dosages of genes and are responsible for diverse human cancers. Although whole genome bisulfite sequencing (WGBS) informs CNVs and DNA methylation, no study has provided a systematic benchmark for detecting CNVs from WGBS data. Herein, based on simulated and real WGBS datasets of 84.62 billion reads, we undertook 714 CNV detections to comprehensively benchmark the performance of 35 strategies, 5 alignment algorithms (bismarkbt2, bsbolt, bsmap, bwameth, and walt) wrapping with 7 CNV detection applications (BreakDancer, cn.mops, CNVkit, CNVnator, DELLY, GASV and Pindel). The results highlighted a subset of strategies that accurately called CNVs depending on numbers, lengths, precision, recall, and F1 scores of CNV detections. We found that bwameth-DELLY and bwameth-BreakDancer were the best strategies for calling deletions, and walt-CNVnator and bismarkbt2-CNVnator were the best strategies for calling duplications. These works provided investigators with useful information to accurately explore CNVs from WGBS data in humans.

1. Introduction

Copy number variations (CNVs) are the genomic structural variations that are longer than 50 bp in the genomes, e.g., deletions (DELs) and duplications (DUPs) [1]. CNVs are widely presented in human genomes, accounting for about 12 % of the genome [2], and CNVs change gene dosages, disrupt coding sequences and modify non-coding regulatory elements such as enhancers, thereby to regulate the transcriptions of genes [3]. CNVs play a vital role in the susceptibility or resistance to human diseases, such as cancers [4], psychiatric disorders [5,6] and psoriasis [7]. Whole genome sequencing (WGS) provided and yielded uniform coverage that enables accurate detection of sequence variants, including single nucleotide polymorphisms (SNPs), indels, and CNVs [8].

DNA methylation refers to a methyl group to the 5th carbon position of cytosine residues on DNA double strands. Numerous studies have shown that DNA methylation causes the chromatin structural changes, alters DNA three-dimensional conformation, and modifies in the interactions between DNA and proteins, thereby to regulate gene expression [9–11]. Whole genome bisulfite sequencing (WGBS) is optimized for profiling DNA methylation at single-base resolution. The genomic DNA is treated with sodium bisulfite to deaminate unmethylated cytosines into uracils (which are subsequently read as thymines after PCR amplification) while leaving methylated cytosines intact [12].

Although WGS is traditionally used for CNV calling, some studies have demonstrated that WGBS holds potential for discovering CNVs by providing DNA-level sequencing data [13]. Exploring both CNVs and DNA methylation information from WGBS data not only reduces the costs and the requirement for sample quantity, but also ensures that the detected CNVs and DNA methylation information have the same biological origins, thus enabling more accurate research on the relationship between CNVs and DNA methylation [14]. In the past decade, plentiful

* Corresponding author at: State Key Laboratory of Swine and Poultry Breeding Industry, National Engineering Research Center for Breeding Swine Industry, Guangdong Provincial Key Laboratory of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou, Guangdong 510642, China.
E-mail address: yxl@scau.edu.cn (X. Yuan).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2025.02.040

Available online 3 March 2025

applications, e.g., BreakDancer [15], CNVnator [16] and Pindel [17], have provided a sensitive and accurate platform to detect CNVs from WGS data. However, there is no study to explore their performance of CNV call on WGBS data. Furthermore, different WGBS-based alignment algorithms, e.g., bsmap [18], abismal [19] and batmeth2 [20], showed the distinct alignment accuracy, which might cause false-positive detection of CNVs from WGBS data. Previously, we undertook 936 mappings to benchmark and evaluate 14 wildly utilized alignment algorithms from reads mapping to biological interpretation in WGBS data of humans, cattle and pigs [21]. Our works found that bwameth [22], bsbolt [23], bsmap, bismarkbt2 [24,25] and walt [26] exhibited higher uniquely mapped reads, mapped precision, recall and F1 score than other alignment algorithms [21].

In order to evaluate the effectiveness of detecting CNVs in WGBS data, we selected these aforementioned 5 alignment algorithms and wrapped them with 7 representative CNV detection applications (BreakDancer, cn.mops [27], CNVkit [28], CNVnator, DELLY [29], GASV [30] and Pindel) to form 35 CNV detection strategies. Here, we referred to each alignment algorithm as a "mapper" and the CNV detection application as a "caller". Moreover, for each combination of mapper and caller (such as "bismarkbt2-BreakDancer"), we referred it as a "strategy". We used these strategies to detect CNVs in simulated and real WGBS data of humans, and comprehensively evaluated multidimensional metrics including numbers, lengths, precision, recall, and F1 scores of CNVs detected. Moreover, to explore the differences in detecting CNVs between WGBS and WGS data, we also evaluated the performance of 7 strategies (bwa [31] wrapping with these 7 callers) in simulated and real WGS data as controls. Our study aimed to discuss the best strategies for CNV detections based on WGBS data and provide insights into genetic variation and epigenetic studies utilizing WGBS data.

2. Materials and methods

2.1. Construction of the real and simulated reference CNV datasets

For constructing the real reference CNV dataset, we downloaded structural variation datasets from the Database of Genomic Variants (DGV) (http://dgv.tcag.ca/dgv/docs/DGV.GS.hg38.gff3), and selected CNVs of individual NA12878, which has been considered a gold standard for validating the accuracy of CNV detections in multiple studies [32].

We used Python (v3.8) scripts to construct the simulated reference CNV dataset. Firstly, we obtained the real coordinates of CNVs from DGV. Secondly, we randomly sampled the coordinates of DELs and DUPs to build simulated CNVs.

2.2. Collection of real and simulated WGBS and WGS sequencing data

The real sequencing data of humans were downloaded from Sequence Read Archive (SRA) of National Center for Biotechnology Information (NCBI) (Table S1), including WGBS and WGS data, which came from the B lymphocyte of individual NA12878 in the 1000 Genomes Project [3,33]. There were totally 3.09 billion reads downloaded (read length: 150 bp, paired-end sequencing). FastQC program (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to control the quality of reads, and Fastp [34] was used to trim adaptor sequences and filter low-quality reads with the default parameters.

For the simulated sequencing data, we first constructed the genome by randomly inserting or deleting simulated DELs and DUPs into the reference genome hg38. Then we used Sherman (https://www.bioinformatics.babraham.ac.uk/projects/sherman/) to generate the response reads. We set 4 sequencing depths for simulating WGBS data: 10 ×, 15 ×, 20 ×, and 30 ×, with a bisulfite conversion rate of 99.60 %. Additionally, we set 15 × for WGS data. There were 4 samples for both WGBS and WGS data. All reads were set to 150 bp in length and were paired-end reads, totaling 81.53 billion.

2.3. Alignments of sequencing data and detections of CNVs

We selected five WGBS mappers (bismarkbt2, bsbolt, bsmap, bwameth and walt), as well as one WGS mapper (bwa) as controls (Table S2). Based on the human reference genome hg38, five WGBS mappers were used to align WGBS data, and one WGS mapper was used to align WGS data as controls. Seven callers (BreakDancer [15], cn.mops [27], CNVkit [35], CNVnator [16], DELLY [29], GASV [30] and Pindel [17]) were chosen for detection of CNVs (Table S3). For CNVnator, we set the window size to 100 bp, and other parameters to defaults. For CNVkit, regions of interest were generated at intervals of 5000 bp along each chromosome, using the reference genome to construct control sample files, with remaining parameters set to defaults. The remaining callers (cn.mops, BreakDancer, DELLY, GASV and Pindel) were used with default parameters. All the strategies were run on a computer with 64 GB of RAM and 24 CPU cores.

2.4. Calculation of precision, recall, F1 scores and relative F1 scores of CNV detections

We used BEDTools (v2.30.0) [36] to intersect the detected CNVs with the reference CNVs. We defined a true detected CNV as the detected CNV that overlapped with the reference CNVs, and we regarded it as the true positive (TP). Based on the TP, the true negatives (TN), false positives (FP), and false negatives (FN) were determined. The precision represented the proportion of true detected CNVs correctly detected out of all CNVs detected, and the recall represented the proportion of true detected CNVs correctly detected out of all reference CNVs. For each strategy, we calculated the average precision and recall for all samples. The precision, recall and F1 score of CNV detections were calculated by the following formula:

$$Precision = \frac{TP}{TP+FP}$$
$$Recall = \frac{TP}{TP+FN}$$
$$F1score = \frac{2*Precision*Recall}{Precision+Recall}$$

2.5. Statistical analysis

The significance differences of numbers, lengths, precision, recall and F1 scores of CNVs detected were tested by the Student's *t*-test using the function of "stats.ttest_ind" from "scipy" module in Python. The Pearson's correlation coefficient was calculated by the function of "pearsonr" from "scipy" module in Python.

3. Results and discussion

3.1. The numbers of CNVs detected by different strategies

We calculated the numbers of CNVs (DELs and DUPs) detected by 35 strategies using simulated WGBS data, and found that bismarkbt2-Pindel detected the most DELs (304,038) and DUPs (46,238) at 30 × ($P < 0.01$) (Fig. 1A and B). Bismarkbt2-BreakDancer, bsbolt-BreakDancer and bismarkbt2-DELLY did not detect DELs (Fig. 1A). Besides, BreakDancer and GASV did not identify DUPs (Fig. 1B). As shown in Fig. S1A, the number of DELs detected by BreakDancer, CNVnator, GASV and Pindel were positively correlated with sequencing depths (Pearson's correlation coefficients > 0.7, $P < 0.01$). The number of DUPs detected by Pindel, bismarkbt2-DELLY, walt-DELLY and bsmap-cn.mops were positively correlated to sequencing depths (Pearson's correlation coefficients > 0.7, $P < 0.01$). We also calculated the number of CNVs detected in real WGBS data and found that bismarkbt2-Pindel detected the most DELs (366,315) ($P < 0.05$), and bwameth-Pindel identified the most DUPs (82,775) ($P < 0.05$) (Fig. S1B). Similar to using simulated WGBS data, BreakDancer and GASV did not identify

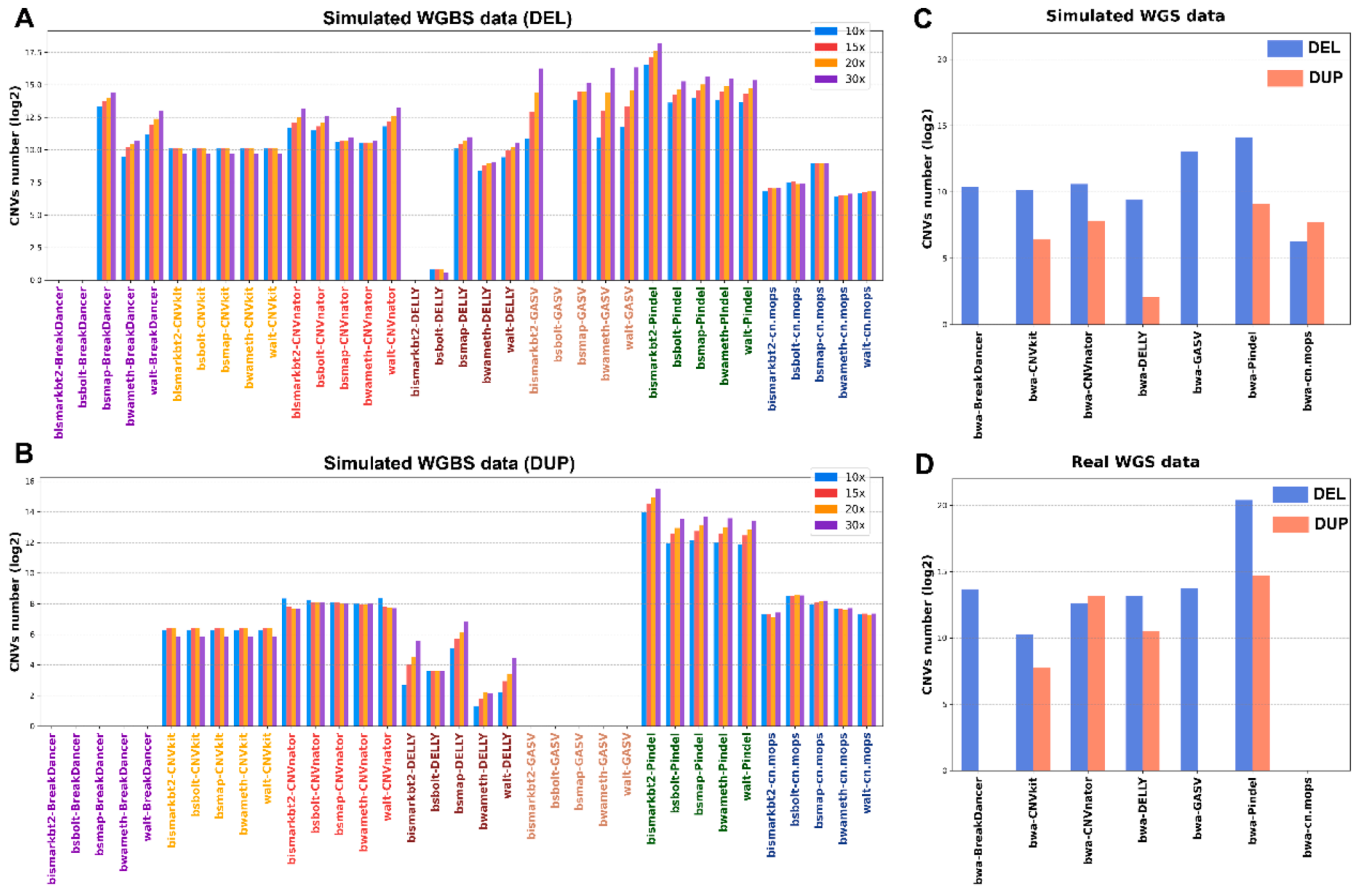

Fig. 1. The numbers of CNVs detected by 35 strategies in simulated WGBS data, simulated WGS data and real WGS data. **A:** The number of DELs detected in simulated WGBS data. **B:** The number of DUPs detected in simulated WGBS data. **C:** The number of CNVs detected in simulated WGS data. **D:** The number of CNVs detected in real WGS data.

DUPs.

To benchmark the differences in the number of CNV detections between WGBS and WGS data, we counted the results of 7 WGS-based strategies. We found that in simulated WGS data, bwa-Pindel detected the most DELs (34,821) and DUPs (39,820) ($P < 0.05$), but bwa-BreakDancer and bwa-GASV did not detect DUPs (Fig. 1C). In real WGS data, bwa-Pindel detected the most DELs (1385,000) and DUPs (26,792) ($P < 0.05$), but bwa-cn.mops did not identify any CNV. Additionally, bwa-BreakDancer and bwa-GASV only detected DELs (Fig. 1D).

3.2. The lengths of CNVs detected by different strategies

We categorized the lengths of CNVs into 6 levels (I: 1–100 bp, II: 100–1000 bp, III: 1000–10,000 bp, IV: 10,000–100,000 bp, V: 100,000–1000,000 bp and VI: 1000,000–10,000,000 bp). In simulated WGBS data, we first discussed the effect of sequencing depths on the lengths of CNVs detected, and the results showed that they were not affected by the change of sequencing depths ($P < 0.05$). Then, as shown in Fig. 2A, we found that CNVs detected by Pindel were mostly in length I (> 90 %), and CNVs detected by BreakDancer and GASV were in length II (> 90 %). CNVs detected by DELLY were also mainly in length II (> 85 %). For CNVnator, there were 41.12 % CNVs in length III and 27.41 % CNVs in length IV. For CNVkit, CNVs were mostly in length IV (50.19 %) and V (43.94 %). Cn.mops tended to detect CNVs that were in length V (77.20 %). As shown in Fig. 2B, in real data, the lengths of CNVs detected by strategies with Pindel, BreakDancer, GASV, CNVkit and CNVnator were similar to using simulated WGBS data.

We also benchmarked the differences in the lengths of CNVs between WGBS and WGS data. We found that in WGS data, CNVs detected by CNVkit, CNVnator, GASV, Pindel and cn.mops were similar to those detected in WGBS data. However, in simulated WGS data, CNVs detected by bwa-BreakDancer were mainly in length III (44.18 %) and IV (30.94 %). Besides, bwa-DELLY detected DELs that were in length II (23.97 %) and III (40.33 %), but > 50 % of DUPs were in length VI (Fig. 2A and B). In real WGS data, bwa-DELLY tended to detect CNVs of length II (Fig. 2 A and 2B).

3.3. The precision and recall of different strategies

We averaged the precision and recall of CNV detections of these strategies using 10 × , 15 × , 20 × and 30 × simulated WGBS data. We found that bsbolt-DELLY showed the highest precision (100 %) ($P < 0.01$) (Fig. 3 A) for DELs, and it was affected by sequencing depths (Fig. S2A) (Pearson's correlation coefficients = 0.07, P = -0.46). Walt-CNVnator showed the highest recall (85.15 %) ($P < 0.01$) (Fig. 3A), and the recall was positively correlated to the sequencing depths (Fig. S2B) (Pearson's correlation coefficients > 0.7, $P < 0.01$). For DUPs, bwameth-CNVkit showed the highest precision (86.32 %) ($P < 0.01$) (Fig. 3A), and the correlation between its precision and sequencing depths was not significant (Fig. S2A) (Pearson's correlation coefficients = 0.10, $P = 0.71$). Bwameth-CNVnator showed the highest recall ($P < 0.01$) (81.64 %) (Fig. 3A), and its recall was not significant correlated with sequencing depths (Fig. S2B) (Pearson's correlation coefficients = 0.33, $P = 0.21$). We also calculated the precision and recall of these strategies using real WGBS data. For DELs, bwameth-BreakDancer showed the highest precision (54.72 %) and walt-CNVnator exhibited the highest recall (36.43 %) (Fig. 3B). For DUPs, bismarkbt2-CNVkit showed the highest precision (42.6 %) and bsbolt-

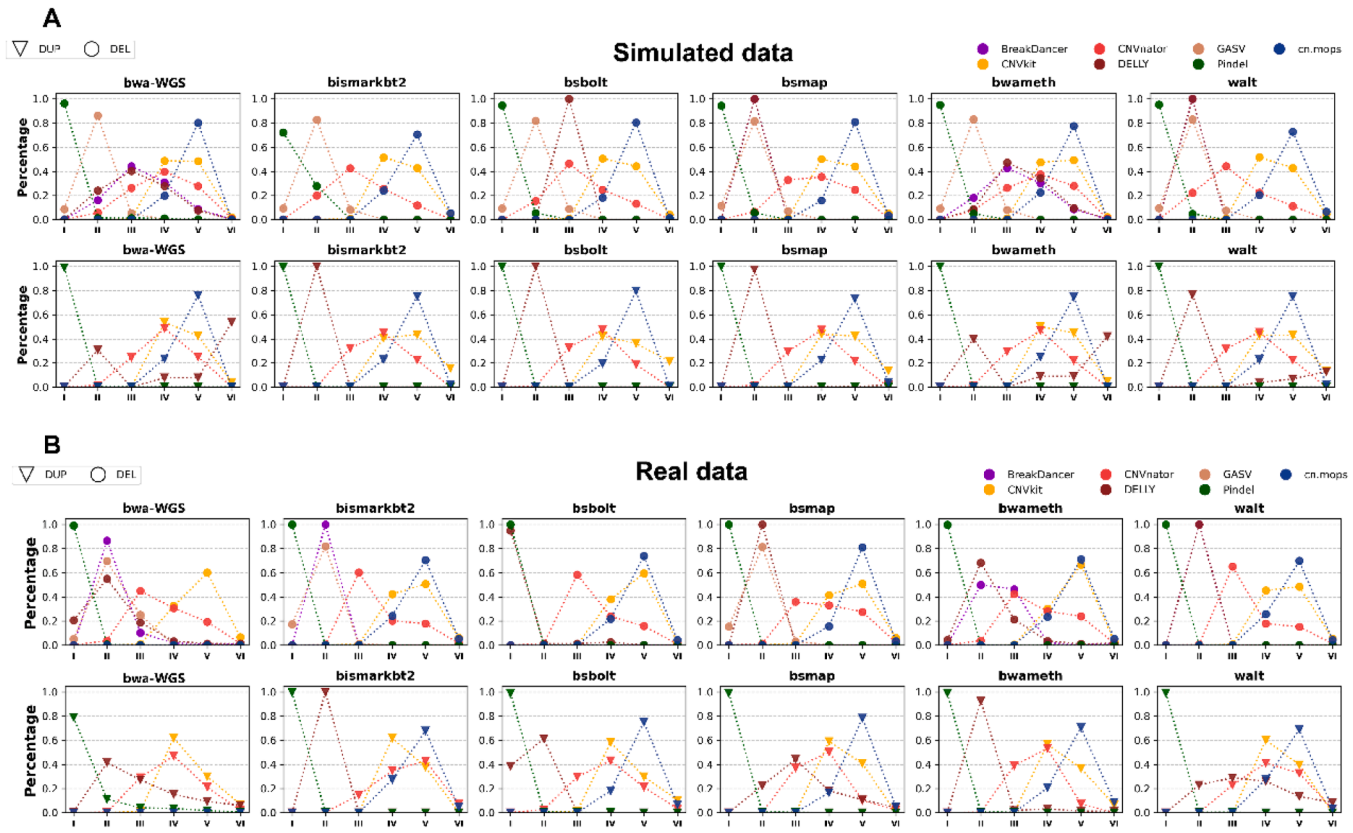

Fig. 2. The length of CNVs detected by different strategies. **A:** The length of CNVs detected using simulated WGS and WGBS data. **B:** The length of CNVs detected using real WGS and WGBS data.

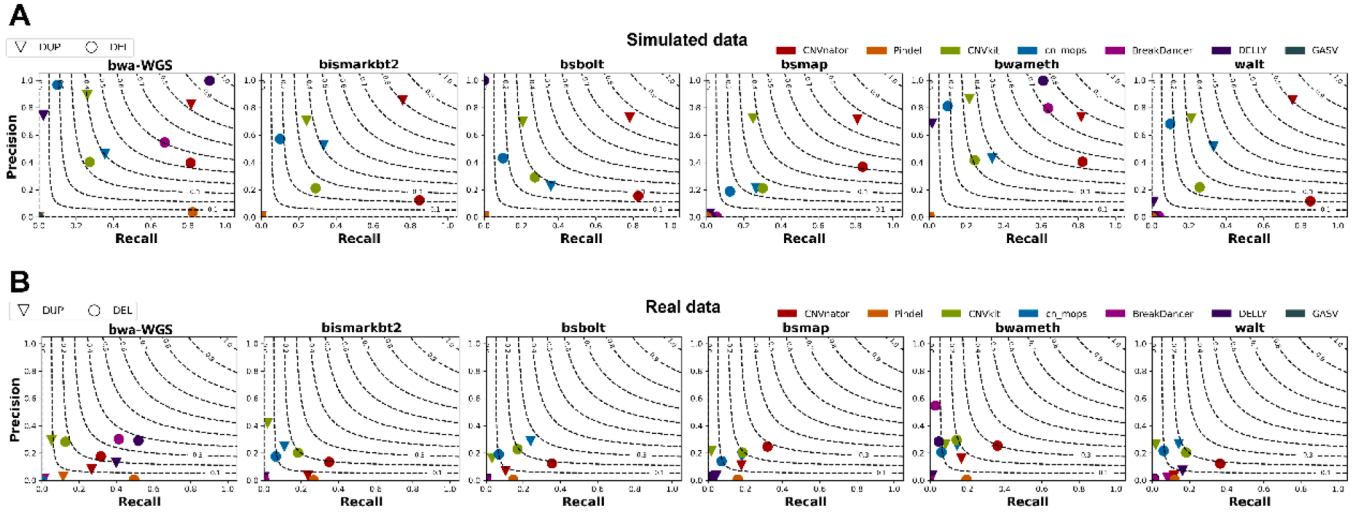

Fig. 3. The precision and recall of CNVs detected in simulated and real data. **A:** The average precision and recall of CNVs detected using simulated WGBS and WGS data. **B:** The precision and recall of CNVs detected using real WGBS and WGS data.

cn.mops exhibited the highest recall (23.95 %) ($P < 0.01$) (Fig. 3B).

As controls, we benchmarked the precision and recall of 7 WGS-based strategies. In simulated data, for DELs, bwa-DELLY showed the highest precision (99.82 %) ($P < 0.05$) and recall (91.61 %) ($P < 0.05$) (Fig. 3A). For DUPs, bwa-CNVkit showed the highest precision (89.19 %) ($P < 0.05$) and bwa-CNVnator showed the highest recall (81.68 %) ($P < 0.05$) (Fig. 3A). In real data, for DELs, bwa-BreakDancer showed the highest precision (30.27 %) ($P < 0.05$), and bwa-DELLY showed the highest recall (51.31 %) ($P < 0.05$) (Fig. 3B). In detection of DUPs, bwa-CNVkit showed the highest precision (29.36 %) ($P < 0.05$) and bwa-DELLY showed the highest recall (30.90 %) ($P < 0.05$) (Fig. 3B).

3.4. *The F1 scores of different strategies*

To comprehensively assess the performance of different strategies, we further calculated F1 scores through their precision and recall. In simulated WGBS data, for DELs, bwameth-DELLY, bwameth-

BreakDancer, bwameth-CNVnator, and bsmap-CNVnator had higher F1 scores than other strategies (Fig. 4A). Bwameth-DELLY showed the highest F1 score at 30 × data (0.84) ($P < 0.05$), followed by bwameth-BreakDancer (0.82) (Fig. 4A), and their F1 scores were positively correlated to the sequencing depths (Pearson's correlation coefficients > 0.7, $P < 0.01$) (Fig. S3A). For DUPs, CNVnator performed better than other callers (Fig. 4B). Both bismarkbt2-CNVnator and walt-CNVnator had the highest F1 score at 30 × data (0.86) (Fig. 4B). In addition, their F1 scores showed a significant positive correlation with sequencing depths (Pearson's correlation coefficients > 0.7, $P < 0.01$) (Fig. S3A).

In real WGBS data, for DELs, CNVnator and CNVkit showed higher F1 scores than other callers, and bwameth-CNVnator had the highest F1 score (0.30) ($P < 0.05$) (Fig. S3B). For DUPs, bsbolt-cn.mops showed the highest F1 scores (0.26) ($P < 0.05$) (Fig. S3B). As controls, in simulated WGS data, bwa-DELLY showed the highest F1 score in detection of DELs (0.35) ($P < 0.05$) and bwa-CNVnator had the highest F1 score in detection of DUPs (0.19) ($P < 0.05$) (Fig. 4 C). In real WGS data, bwa-DELLY showed the highest F1 scores in detection of DELs and DUPs (Fig. 4D).

3.5. *The relative F1 scores of different strategies between WGBS and WGS data*

We further compared the performance of strategies in WGBS and WGS data with the relative F1 scores. Using simulated WGBS data, in detection of DELs, bwameth-BreakDancer and bwameth-CNVnator showed relative F1 scores > 1 at 15 × , 20 × and 30 × depths, among which bwameth-BreakDancer showed the highest score at 30 × data (1.36) ($P < 0.05$) (Fig. 5 A). In detection of DUPs, bismarkbt2-CNVnator and walt-CNVnator showed relative F1 scores > 1 at 20 × and 30 × depths, among which walt-CNVnator had the highest score at 30 × the depth (1.05). Additionally, bismarkbt-cn.mops and walt-cn.mops showed relative F1 scores > 1 at 15 × and 20 × depths (Fig. 5B).

In real WGBS data, we found that strategies with cn.mops had relative F1 scores > 1 both for DELs and DUPs. Additionally, bwameth-CNVnator and bsmap-CNVnator also had relative F1 scores > 1 in detection of DELs and DUPs (Fig. S4A and S4B). Among these strategies, bsbolt-cn.mops showed the highest relative F1 score (2.02) for DELs (Fig. S4A), and bsbolt-cn.mops showed the highest relative F1 score (5.24) in detection of DUPs (Fig. S4B). Considering simulated and real data comprehensively, we found that BreakDancer, cn.mops, CNVnator and CNVkit performed better using WGBS data than WGS data.

4. Discussion

Many studies have revealed the relevance between CNVs and DNA methylation. In 2022, 1000 Genomes Project has identified 851 significant CNV-CpG associations in the human genome, involving 656 CNVs, 715 genes, and 738 CpG loci [37]. In humans, DUPs in *CTCF* gene reduces methylation levels of 134 CpG sites and may result in the appearance of breast cancer [38]. Large fragment of DELs in *BRD1* gene region leads to increased methylation levels [39]. Additionally, the activity of transposable elements is to be suppressed by CpG methylation thus DNA methylation could play a critical role in CNV formation and genome stability [40]. These studies showed that abnormal DNA methylation leads to the occurrence of CNVs, and the occurrence of

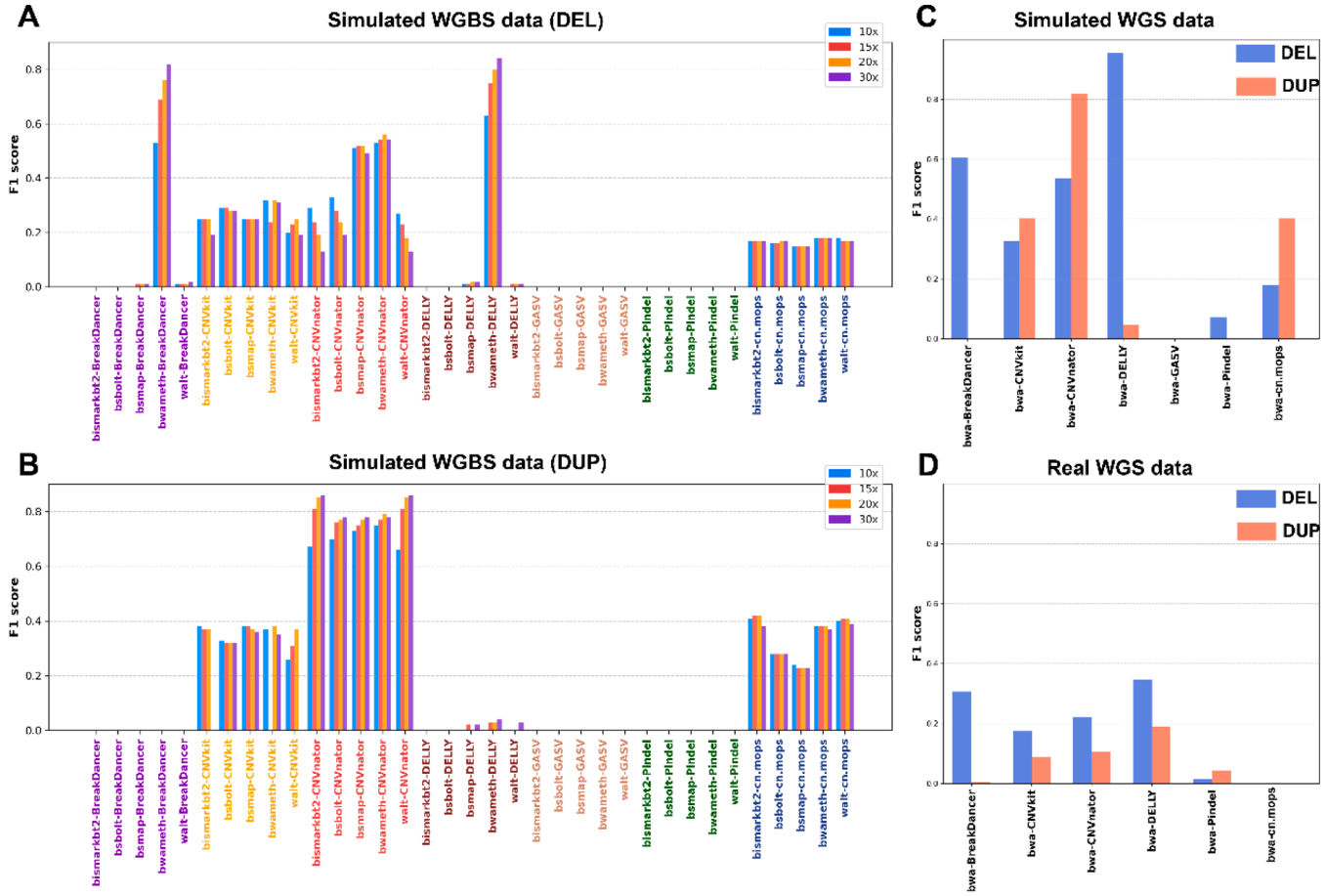

Fig. 4. The F1 scores of CNVs detected by 35 strategies in simulated WGBS data, simulated WGS data and real WGS data. **A:** The F1 scores of DELs detected by 35 strategies. **B:** The F1 scores of DUPs detected by 35 strategies. **C:** The F1 scores of CNV detections in simulated WGS data. **D:** The F1 scores of CNV detections in real WGS data.

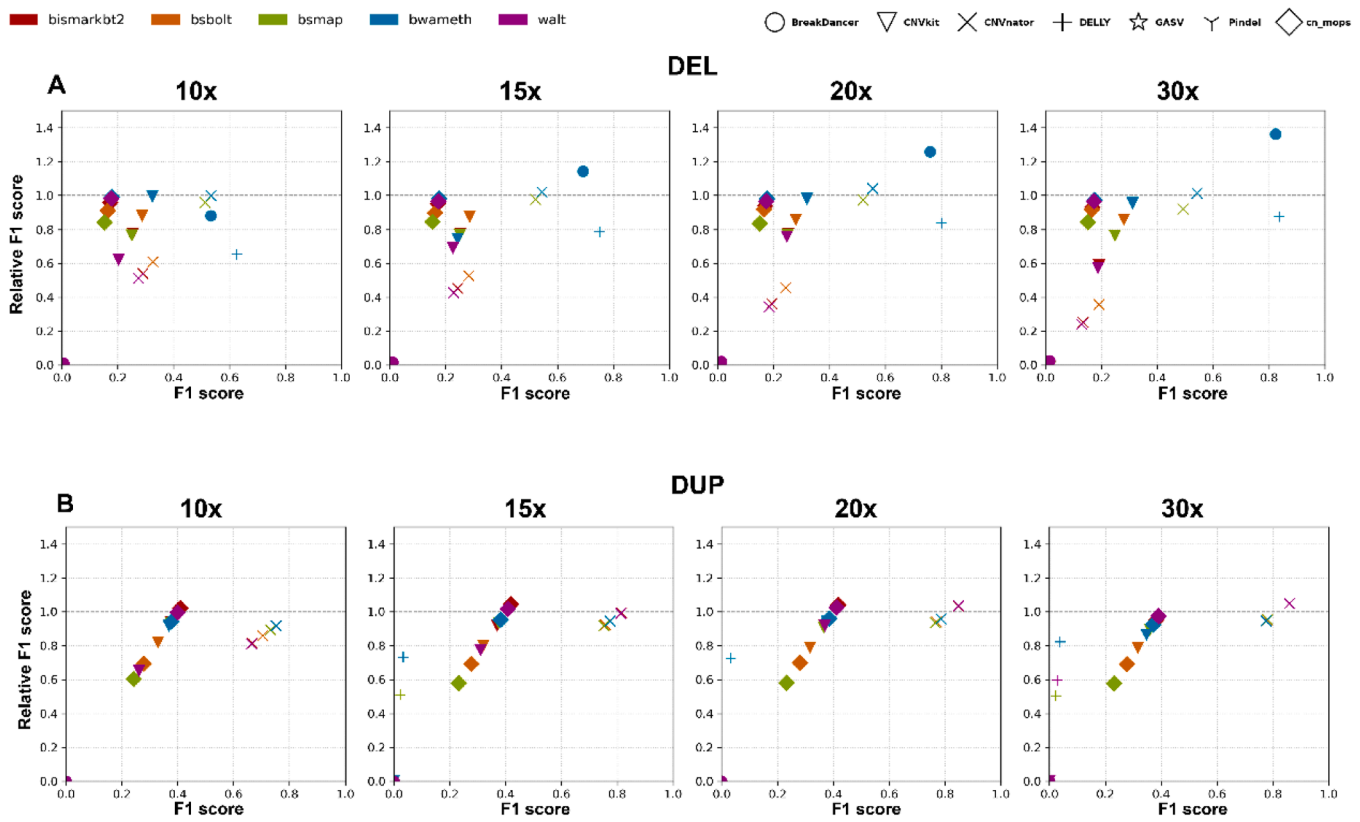

Fig. 5. The relative F1 scores of CNVs detected by 35 strategies using simulated WGBS data. **A:** The relative F1 scores of DELs and **B:** DUPs detected at 10 × , 15 × , 20 × and 30 × depths.

CNVs alters the levels of DNA methylation. Therefore, it is important to detect both CNVs and DNA methylation accurately, and to perform comprehensive correlation analysis and mechanistic studies.

In studies of human diseases, such as cancers, integrating analyses of copy number variations (CNVs) and DNA methylation has proven to be an effective strategy. For instance, a study demonstrated that specific CNVs are associated with DNA methylation changes, which may modulate gene expression and impact cancer risk [37]. Furthermore, integrating DNA methylation and CNVs from plasma cell-free DNA has been shown to enhance accuracy of early cancer detection [41]. Our benchmark provides valuable insights for simultaneously investigating the mechanisms of CNVs and DNA methylation, thereby improving the accuracy of study on human diseases and cancers.

The approaches employed by different CNV callers are mainly defined into 4 categories, including Read Depth (RD) [42,43], Read Pair (RP) [44], Split Read (SR) [45], and De novo Assembly (AS) [46]. In our benchmark, bwameth-DELLY and bwameth-BreakDancer performed better for DELs than other strategies, and this result supports the study by Whitney et al., who found that BreakDancer and DELLY were the most accurate methods to call DELs from WGS data [47]. Bismarkbt2-CNVnator and walt-CNVnator showed the highest F1 scores according to our study which was based on human genome, however, a study to detect copy number variation in plant genomes showed that CNVnator performed poorly during evaluations [48].

The reference CNVs for NA12878 individual was sourced from DGV and was considered a gold standard for validating the accuracy of CNVs detection results in multiple studies [32]. While this individual is a widely used reference in genomic studies, our findings may not fully capture the genetic diversity present in the broader human population. Future studies should aim to expand the scope of analysis to include multiple individuals from diverse populations to better assess the generalizability of CNV detection strategies. Moreover, the robust population-level validation is necessary to assess the applicability and accuracy of these strategies across diverse genomic backgrounds in the future. Population-level validation helps determine how well these strategies perform in larger, more heterogeneous datasets, providing a clearer understanding of their real-world utility. Besides, the detection strategies we evaluated may perform differently when applied to the genomes of other organisms, which may have distinct structural variation patterns, genomic architectures, and epigenetic features. Comparative analyses across species could provide valuable insights into the robustness and versatility of these CNV detection methods in different biological contexts.

5. Conclusion

We comprehensively evaluated the performance of a total of 35 strategies in detecting CNVs based on simulated and real WGBS data in humans, with various metrics including numbers, lengths, precision, recall, and F1 scores. Our evaluation test reveals that bwameth-DELLY and bwameth-BreakDancer were the best strategies for detecting DELs, and walt-CNVnator and bismarkbt2-CNVnator were the best strategies in detection of DUPs in WGBS data.

Funding

This research was funded by National Key Research and Development Program of China (2022YFF1000900), Guangdong Basic and Applied Basic Research Foundation (2024B1515020112 and 2024A1515012999), Science and Technology Program of Guangzhou (2024B03J1305), the earmarked fund for China Agriculture Research System (CARS-35), and National Center of Technology Innovation for Pigs (NCTIP-XDB14).

CRediT authorship contribution statement

Zhao Shanghui: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Yuan Xiaolong:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Shen Qingpeng:** Methodology, Formal analysis, Data curation. **Cai Jiali:** Methodology, Formal analysis, Data curation. **Xu Dantong:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Li Jiaqi:** Supervision, Resources, Funding acquisition. **Gao Yahui:** Supervision, Resources, Funding acquisition. **Pan Xiangchun:** Methodology, Formal analysis, Data curation. **He Mingran:** Methodology, Formal analysis, Data curation.

Declaration of Competing Interest

The authors have declared no conflict of interest.

Acknowledgements

We thank the National Supercomputer Center in Guangzhou for its computing platform.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.02.040.

Data Availability

The real WGBS data was downloaded from SRA of project ID PRJNA860202. The real WGS data was downloaded from SRA of project ID PRJNA186949, PRJEB31736 and PRJEB3381. The detailed numbers of SRA are available in Supplemental Table S1. The gold standard CNV datasets were downloaded from DGV (http://dgv.tcag.ca/dgv/docs/DGV.GS.hg38.gff3). The source code supporting the generation of simulated datasets and statistical analysis in this study is available at GitHub (http://github.com/xdtabc/bisCNV).

References

[1] Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet 2015;16:172–83.
[2] Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007;315:848–53.
[3] Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, et al. An integrated map of structural variation in 2,504 human genomes. Nature 2015;526:75–81.
[4] de Pagter MS, Kloosterman WP. The diverse effects of complex chromosome rearrangements and chromothripsis in cancer development. Recent Results Cancer Res 2015;200:165–93.
[5] Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 2006;38:24–6.
[6] Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al. Strong association of de novo copy number mutations with autism. Science 2007;316:445–9.
[7] de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, et al. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. Nat Genet 2009;41:211–5.
[8] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53–9.
[9] Espada J, Esteller M. DNA methylation and the functional organization of the nuclear compartment. Semin Cell Dev Biol 2010;21:238–46.
[10] Delpu Y, Cordelier P, Cho WC, Torrisani J. DNA methylation and cancer diagnosis. Int J Mol Sci 2013;14:15029–58.
[11] Jeltsch A, Jurkowska RZ. New concepts in DNA methylation. Trends Biochem Sci 2014;39:310–8.
[12] Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009;462:315–22.
[13] Liao WW, Yen MR, Ju E, Hsu FM, Lam L, et al. MethGo: a comprehensive tool for analyzing whole-genome bisulfite sequencing data. BMC Genom 2015;16(12):S11.
[14] Teng H, Xue M, Liang J, Wang X, Wang L, et al. Inter- and intratumor DNA methylation heterogeneity associated with lymph node metastasis and prognosis of esophageal squamous cell carcinoma. Theranostics 2020;10:3035–48.
[15] Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 2009;6:677–81.
[16] Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res 2011;21:974–84.
[17] Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009;25:2865–71.
[18] Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinforma 2009;10:232.
[19] de Sena BG, Smith AD. Fast and memory-efficient mapping of short bisulfite sequencing reads using a two-letter alphabet. NAR Genom Bioinform 2021:3.
[20] Zhou Q, Lim JQ, Sung WK, Li G. An integrated package for bisulfite DNA methylation data analysis with Indel-sensitive mapping. BMC Bioinforma 2019;20:47.
[21] Gong W, Pan X, Xu D, Ji G, Wang Y, et al. Benchmarking DNA methylation analysis of 14 alignment algorithms for whole genome bisulfite sequencing in mammals. Comput Struct Biotechnol J 2022;20:4704–16.
[22] Foschini MP, Morandi L, Sanchez AM, Santoro A, Mule A, et al. Methylation profile of X-chromosome-related genes in male breast cancer. Front Oncol 2020;10:784.
[23] Farrell C, Thompson M, Tosevska A, Oyetunde A, Pellegrini M. BiSulfite bolt: a bisulfite sequencing analysis platform. Gigascience 2021:10.
[24] Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 2011;27:1571–2.
[25] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9.
[26] Chen H, Smith AD, Chen T. WALT: fast and accurate read mapping for bisulfite sequencing. Bioinforma (Oxf, Engl) 2016;32:3507–9.
[27] Klambauer G, Schwarzbauer K, Mayr A, Clevert DA, Mitterecker A, et al. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res 2012;40. e69.
[28] Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol 2016;12:e1004873.
[29] Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 2012;28:i333–9.
[30] Sindi S, Helman E, Bashir A, Raphael BJ. A geometric approach for classification and comparison of structural variants. Bioinformatics 2009;25:i222–30.
[31] Li H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. Ithaca: Cornell University Library, arXiv.org; 2013.
[32] Pirooznia M, Goes FS, Zandi PP. Whole-genome CNV analysis: advances in computational approaches. Front Genet 2015;6:138.
[33] Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
[34] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90.
[35] Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol 2016;12:e1004873.
[36] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2.
[37] Shi X, Radhakrishnan S, Wen J, Chen JY, Chen J, et al. Association of CNVs with methylation variation. NPJ Genom Med 2020;5:41.
[38] Sun W, Bunn P, Jin C, Little P, Zhabotynsky V, et al. The association between copy number aberration, DNA methylation and gene expression in tumor samples. Nucleic Acids Res 2018;46:3009–18.
[39] Schenkel LC, Aref-Eshghi E, Rooney K, Kerkhof J, Levy MA, et al. DNA methylation epi-signature is associated with two molecularly and phenotypically distinct clinical subtypes of Phelan-McDermid syndrome. Clin Epigen 2021;13:2.
[40] Luo Y, Lu X, Xie H. Dynamic Alu methylation during normal development, aging, and tumorigenesis. Biomed Res Int 2014;2014:784706.
[41] Hua X, Zhou H, Wu HC, Furnari J, Kotidis CP, et al. Tumor detection by analysis of both symmetric- and hemi-methylation of plasma cell-free DNA. Nat Commun 2024;15:6113.
[42] Kadalayil L, Rafiq S, Rose-Zerilli MJ, Pengelly RJ, Parker H, et al. Exome sequence read depth methods for identifying copy number changes. Brief Bioinform 2015;16:380–92.
[43] Teo SM, Pawitan Y, Ku CS, Chia KS, Salim A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. Bioinformatics 2012;28:2711–8.
[44] Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science 2007;318:420–6.
[45] Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, et al. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 2006;16:1182–90.

[46] Eichler EE, Alkan C, Coe BP. Genome structural variation discovery and genotyping. Nat Rev Genet 2011;12:363–76.
[47] Whitford W, Lehnert K, Snell RG, Jacobsen JC. Evaluation of the performance of copy number variant prediction tools for the detection of deletions from whole genome sequencing data. J Biomed Inf 2019;94:103174.
[48] Wijfjes RY, Smit S, de Ridder D. Hecaton: reliably detecting copy number variation in plant genomes using short read sequencing data. BMC Genom 2019;20:818.