Chinese Medical Journal®

Targeted next-generation sequencing for copy number variation detection: Recent progress in clinical oncology

Zhao Du[1], Jie Hu[1], Guannan Chen[1], Jinqiao Tang[1], Yongchao Yao[1], Binwu Ying[1,2], Wenchuang Hu[1]

[1]Department of Laboratory Medicine, Precision Medicine Translational Research Center (PMTRC), West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China; [2]Clinical Laboratory Medicine Research Center, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China.

Cancer development is a complex, multistep process fundamentally driven by diverse factors that ultimately arise from somatic genetic alterations. Copy number variations (CNVs) are defined as the gains or losses of large genomic segments exceeding 50 base pairs that arise from structural rearrangements of the genome.[1] CNVs significantly influence disease recurrence and patient outcomes, particularly overall survival and progression-free survival. For example, detecting CNVs uncovers actionable genomic alterations that predict therapeutic response, thus facilitating the selection of optimal treatment strategies.

Conventional CNV detection methods include fluorescence *in situ* hybridization, polymerase chain reaction, multiplex ligation-dependent probe amplification, and immunohistochemistry. Although these techniques are robust, their application is constrained by their inability to simultaneously assess more than a few CNV loci. In contrast, high-throughput sequencing technologies, such as next-generation sequencing (NGS), enable the simultaneous detection of multiple types of genomic alterations, thereby enhancing both efficiency and accuracy. One of the key advantages of NGS-based diagnostics is their ability to identify, within a single assay, a broad spectrum of genomic changes, including single nucleotide variants, insertions and deletions, gene fusions, CNVs, and chromosomal translocations.[2] Consequently, NGS has been widely adopted for comprehensive genomic profiling in clinical settings.

NGS-based methods for detecting CNVs are generally categorized into four main strategies: Read-pair (RP), split-read (SR), assembly-based (AS), and read-depth (RD) approaches [Supplementary Figure 1, http://links.lww.com/CM9/C787].[3,4] The RD method is currently the predominant approach utilized for CNV detection, particularly for data derived from targeted next-generation sequencing (tNGS). This method is based on the principle that the CNV in a genomic region directly correlates with the variation in the number of sequencing reads mapped to that region. RD-based methods comprise two main categories. Reference-based methods use normal samples to establish an RD baseline and are robust to technical biases; however, they require matched controls. Conversely, self-referencing methods infer the baseline from the test sample itself, assuming genome-wide diploidy, which offers flexibility when normal tissue is unavailable but may result in reduced accuracy in highly aneuploid or low-purity tumors. To date, numerous RD-based CNV detection tools have been developed. Table 1 summarizes several representative tools, their publication year, underlying algorithm or statistical model, control sample requirement, input format, applicable data type, and their key strengths and limitations.

Although numerous methods for CNV detection in tNGS data have been developed, several challenges persist: (1) Short read lengths in NGS hinder accurate alignment within repetitive or complex genomic regions, thereby limiting the sensitivity and precision of CNV detection in these areas. (2) Tumor heterogeneity, including variable tumor purity and subclonal populations, can dilute the apparent copy number signal from tumor cells, making it difficult to distinguish subtle or borderline CNVs from background noise. (3) Inherent technical variability in the tNGS data further complicates the analysis. Factors such as differences in library preparation protocols, base quality scores, guanine–cytosine content bias, alignment artifacts, uneven coverage, poor uniformity, and limited reproducibility can introduce noise that obscures true CNV signals. Given these limitations, there is an urgent need for developing and refining computational and experimental strategies to address these issues effectively.

Access this article online

Quick Response Code:

Website: www.cmj.org

DOI: 10.1097/CM9.0000000000004001

Correspondence to: Yongchao Yao, Department of Laboratory Medicine, Precision Medicine Translational Research Center (PMTRC), West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China
E-Mail: hatuu@wchscu.edu.cn

Chinese Medical Journal 2026;139(8)

Received: 16-01-2025; **Online:** 06-03-2026 **Edited by:** Xuehong Zhang

Table 1: Representative read-depth-based CNV detection tools for WES/tNGS data.

Tool	Year	Algorithm or model	Control data	Input	Data type	Strengths	Limitations
Exome CNV	2011	CBS algorithm	Yes	GTF or Plain text	WES	One of the earliest WES-CNV callers; stable and well-documented	Requires matched normal samples; sensitive to low-coverage regions
Exome Copy	2011	HMM	Yes	BAM files	WES	Capable of detecting small CNVs; robust to noise	Computationally intensive; requires a cohort of control samples
Exome Depth	2012	HMM	Yes	BAM files	WES	High sensitivity and specificity; widely used in clinical diagnostics	Requires a reference set of normal samples for baseline modeling
CoNIFER	2012	SVD + normalization model	No	BAM files, RPKM files	WES	No need for matched controls; effectively removes batch effects	Performance improves with larger sample sizes (>20 recommended)
CANOES	2014	Negative binomial distribution model	Yes	Plain text	WES	Accounts for over-dispersion in sequencing depth; good for capture bias	Only works at exon-level resolution
CODEX	2015	Poisson log-linear model	Yes	BAM files	WES	Effectively corrects GC bias and captures efficiency variation	Original version requires manual parameter tuning
CoNVaD-ING	2015	Z-score	Yes	BAM files	WES/tNGS	Lightweight and fast; ideal for small panels or single-gene disorders	Lower sensitivity; struggles with single-exon deletions/duplications
DECoN	2016	HMM	Yes	BAM files	WES/tNGS	Clinically validated; suitable for diagnostic reporting	Optimized mainly for Ion Torrent platforms
CNVkit	2016	CBS/HMM	Yes	BAM files	WES/tNGS	Open-source and flexible; supports hybrid-capture and amplicon panels; outputs VCF	Reference pool recommended for best performance
CNV-RF	2016	Random forest algorithm	No	BAM files, VCF files	WES/tNGS	No control required; integrates multiple features using machine learning	Performance depends on training dataset quality
SeqCNV	2017	MPLE model	Yes	BAM files	WES/tNGS	High-resolution breakpoint detection; statistically rigorous	High computational cost
MFCNV	2017	Neural network algorithm	No	BAM files	WES	Uses deep learning for pattern recognition; no parametric assumptions	Requires platform-specific training; limited generalizability
DeepSV	2018	Deep learning (CNN-like architecture)	No	BAM files	WES	Early application of deep learning to CNV detection	Limited validation and generalizability across platforms
KNNCNV	2021	KNN	No	BAM files, GTF	WES/tNGS	Non-parametric; robust to outliers; easy to interpret	Sensitive to reference cohort composition
ClearCNV	2022	Linear regression + outlier detection	Yes	BAM files	WES/tNGS	Optimized for low-input DNA and FFPE samples; high accuracy	Requires reference samples; not fully automated
ifCNV	2022	Isolation Forest	Yes	BAM files	WES/tNGS	Detects outliers in coverage patterns using unsupervised ML	Requires training on negative samples; limited sensitivity for subtle CNVs
varAm-pliCNV	2023	PCA (principal component analysis) and multidimensional scaling	No	BAM files	tNGS	Specifically designed for amplicon-based panels; handles amplification bias	Limited to targeted NGS with uniform design
CNV-FB	2023	Random sampling with bootstrap-like confidence estimation	No	BAM files	WES/tNGS	Provides confidence scores for calls; fast and lightweight	Lower resolution compared to model-based methods
LDCNV	2023	KNN	No	BAM files	WES/tNGS	Incorporates genetic LD structure for improved calling accuracy	Best suited for germline variants in population datasets
LMADCNV	2025	Median Absolute Deviation based anomaly scoring	No	BAM files	WES/tNGS	Robust to local noise; detects subtle CNVs using localized statistics	Novel method; limited real-world validation so far

BAM: Binary alignment/map; CBS: Circular binary segmentation; CNV: Copy number variation; FFPE: Formalin-fixed paraffin-embedded; high accuracy GC: Guanine-cytosine; GTF: Gene transfer format; HMM: Hidden Markov model; KNN: K-nearest neighbor; LD: Linkage disequilibrium; ML: Machine learning; MPLE: Maximum penalized likelihood estimation; RPKM: Reads Per Kilobase per million mapped reads; SVD: Singular value decomposition; tNGS: Targeted next-generation sequencing data; VCF: Variant call format; WES: Whole exome sequencing.

The integration of long-read sequencing and tNGS is emerging as a powerful strategy for addressing the limitations of short-read technologies for resolving complex CNVs.[5] Current tNGS panels rely on short reads. However, long-read sequencing spans repetitive regions and entire structural variants in a single read, enabling precise CNV breakpoint mapping. This allows the direct capture of exact junctions, thereby avoiding indirect inferences from discordant pairs or depth changes required by short-read methods. By contrast, short-read methods only localize breakpoints to windows spanning from hundreds to thousands of base pairs. Thus, long reads offer improved resolution of intermediate-to-large CNVs. Conversely, complex rearrangements are often missed or misassembled by short reads. The integration of long-read data into routine tNGS remains a key challenge. However, emerging hybrid bioinformatics approaches, which combine short-read sensitivity with long-read precision, hold significant

promise for a more accurate and comprehensive CNV detection in clinical genomics [Supplementary Figure 2, http://links.lww.com/CM9/C787].

Furthermore, advances in single-cell DNA sequencing (scDNA-seq) have transformed our ability to dissect tumor heterogeneity and reconstruct clonal architecture at an unprecedented resolution. In contrast to bulk sequencing, which yields an averaged genomic signal, scDNA-seq enables the direct detection of CNVs in individual tumor cells, revealing subclonal populations, tracing evolutionary trajectories, and identifying the emergence of therapy-resistant clones.[6] Although single-cell RNA sequencing (scRNA-seq) can indirectly detect large-scale CNVs by analyzing aberrant gene expression patterns, this approach is limited by transcriptional noise, variable gene expression, and low sensitivity to focal or small CNVs.[7] Consequently, DNA-based methods remain the established benchmark for accurate CNV profiling at single-cell resolution. Although challenges regarding cost, scalability, and analytical standardization currently constrain routine clinical adoption, ongoing technological improvements are facilitating the integration of single-cell genomics into precision oncology frameworks.

Moreover, artificial intelligence and machine learning (ML) are expected to play increasingly important roles in refining CNV detection using tNGS data. Emerging algorithms aim to improve robustness in challenging scenarios, such as low tumor purity, intratumoral heterogeneity, or stromal contamination, by utilizing patterns in sequencing depth, B-allele frequency, and other complementary genomic features. The integration of multi-omics and clinical data is anticipated to contextualize CNVs. Nevertheless, current efforts remain largely focused on enhancing signal-to-noise ratios and reducing false positives in real-world samples. Beyond detection, robustly validated ML models may eventually support clinical interpretation by linking specific CNV profiles to drug responses or prognoses, consequently aiding biomarker discovery and therapeutic decision-making.

In conclusion, the clinical utility of CNV detection in oncology is steadily progressing through synergies between improved wet lab protocols, sophisticated bioinformatics tools, and interdisciplinary collaborations. As these methods become more standardized and rigorously validated, tNGS-based CNV analysis is likely to transition from a research tool to a routine component of molecular diagnostics, enabling more precise tumor characterization, identifying actionable alterations, and guiding personalized treatment strategies.

Funding

This study was supported by a grant from the 1.3.5 Project for Disciplines of Excellence from West China Hospital of Sichuan University (No. ZYGD24007).

Conflicts of interest

None.

References

1. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet 2015;16:172–183. doi: 10.1038/nrg3871.
2. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, *et al*. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013;31:1023–1031. doi: 10.1038/nbt.2696.
3. Gordeeva V, Sharova E, Arapidi G. Progress in methods for copy number variation profiling. Int J Mol Sci 2022;23. doi: 10.3390/ijms23042143.
4. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: Features and perspectives. BMC Bioinformatics 2013;14. doi: 10.1186/1471-2105-14-S11-S1.
5. Zheng Y, Shang X. SVvalidation: A long-read-based validation method for genomic structural variation. PLoS One 2024;19:e0291741. doi: 10.1371/journal.pone.0291741.
6. Guo L, Yi X, Chen L, Zhang T, Guo H, Chen Z, *et al*. Single-cell DNA sequencing reveals punctuated and gradual clonal evolution in hepatocellular carcinoma. Gastroenterology 2022;162:238–252. doi: 10.1053/j.gastro.2021.08.052.
7. Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y, *et al*. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat Biotechnol 2021;39:599–608. doi: 10.1038/s41587-020-00795-2.

How to cite this article: Du Z, Hu J, Chen GN, Tang JQ, Yao YC, Ying BW, Hu WC. Targeted next-generation sequencing for copy number variation detection: Recent progress in clinical oncology. Chin Med J 2026;139:1107–1109. doi: 10.1097/CM9.0000000000004001