scientific reports

Check for updates

OPEN

Systematic performance evaluation and application validation of an end-to-end NGS workstation

Wenlong Xie[1,3], Chen Yang[1]✉ & Shibo Ren[2]✉

Next-generation sequencing (NGS) library preparation is a core component of precision genomics, but it is commonly constrained by inefficiency, variability, and low throughput of manual protocols. To address these limitations, we developed and systematically evaluated a fully automated NGS workstations and further validated its performance across representative application scenarios. The automated system reduced total processing time from 8 to 10 to 4–6 h. At the same time, it maintained similar performance in pre-library metric, including DNA yield and fragment size, as well as post-capture sequencing metrics (Q30 > 90%, mapping rates > 95%, on-target rates 85–90%). The duplication rate was reduced to 5–8%, compared with 10–15% for manual methods, indicating increased library complexity. Bioinformatic evaluation of inter-species read mapping showed minimal cross-contamination, with a maximum contamination ratio of 0.0003%, indicating effective sample isolation in the automated workflow. High concordance in variant detection was observed between automated and manual workflows. Overall, this automated workstation provides a standardized and reproducible workflow that supports scalable precision genomics applications.

Keywords Next-generation sequencing, Automated workstation, Library preparation, Hybridization capture, Cross-contamination

Next-generation sequencing (NGS)[1] has substantially advanced genomics by enabling high-throughput genetic analysis[2], with applications ranging from cancer diagnostics to infectious disease surveillance[3–8]. However, the complexity of NGS library preparation remains a major bottleneck[9,10]. The process involves multiple steps, including nucleic acid extraction, library preparation, and hybridization capture, each subject to variability and potential error when performed manually[9,11,12]. Manual workflows are labor-intensive and time-consuming, often requiring 8–10 h per batch, and are susceptible to human-induced inconsistencies, including pipetting accuracy and step timing[10,13,14]. These limitations are of particular concern in clinical settings[15,16]. For example, in tumor sequencing for precision oncology, batch-to-batch variability from manual library preparation has been reported to introduce up to 15% fluctuation in core metrics like library yield and fragment size, which may affect diagnostic performance through inconsistent mutation detection[15].

Although current semi-automated platforms address some aspects of manual NGS library preparation, they remain limited in scalability and flexibility[10,14,16]. Additionally, semi-automated systems require substantial hands-on time for setup and monitoring, limiting their applicability for large-scale genomic studies or time-sensitive clinical diagnostics workflows[14–16]. As a result, there is an increase in demand for fully-automated NGS solutions[10]. This gap is observed in high-throughput applications like pathogen detection, where rapid, standardized sequencing is required for both outbreak response and surveillance[10,15].

Here, we describe the development of a series of end-to-end fully automated NGS workstations that integrate library preparation and hybridization capture. The system consolidates robotic platforms, microfluidic liquid handling, and computer-controlled modules into a unified, modular workflow. This integration substantially reduces manual intervention and offers scalable throughput from 4 to 48 samples per run. This design supports diverse applications, including clinical tumor profiling and large-scale genomic research. In addition, the capability for short-turnaround, high-throughput sequencing addresses key operational requirements in pathogen surveillance, where rapid characterization of microbial genomes is essential during outbreaks of emerging infectious diseases[2,17–20]. Collectively, these workstations were developed to address current limitations

[1]Nanodigmbio (Nanjing) Biotechnology Co.Ltd., Floor 5-6, Tower 11, 71 XingHui Road, Jiangbei New District, Nanjing, Jiangsu, China. [2]Nanjing Institute of Metrological Supervision and Testing, No. 10, Maqun Avenue, Qixia District, Nanjing, Jiangsu, China. [3]School of Basic Medical Sciences, Yichun University, Yichun, China. ✉email: yangchen@nanodigmbio.com; 450997180@qq.com

in NGS automation by providing a standardized and reproducible solution suitable for both clinical and public health genomics.

Materials and methods

Evaluation of liquid handling accuracy and temperature performance

Accuracy and precision of liquid handling were assessed using a gravimetric method under controlled laboratory conditions (ambient temperature 20 ± 5 °C; relative humidity 40–60%). Purified water was dispensed at predefined target volumes (2 µL, 100 µL, and 200 µL) and weighed using an analytical balance. The dispensed volume was calculated based on the measured mass and the density of water corresponding to the recorded ambient temperature.

Temperature calibration was conducted using a calibrated temperature calibrator in conjunction with WINSCP software. The temperature probe was placed directly inside the reaction well to evaluate temperature accuracy, uniformity, and stability at setpoint temperatures of 45 °C, 72 °C, and 95 °C.

Library preparation

Human genomic DNA standards were used for performance assessment, including standard G1471 (Promega) and GW-OGTM800, GW-OGTM001, and GW-OGTM005 (GeneWell). Pre-libraries were prepared using the NadPrep EZ DNA Library Preparation Kit (Nanodigmbio) according to the manufacturer's instructions.

Hybridization capture

Hybridization capture was performed using NadPrep ES Hybrid Capture reagents in combination with the NanOnco Plus Panel v3.0 (Nanodigmbio) or LungCancer Panel v1.0 (Nanodigmbio), according to the manufacturer's instructions.

Next-generation sequencing and data analysis

We utilized fastp (v0.20.0) for basic filtering and adapter removal of raw sequencing data. A Python (v2.7.5) script was employed to assess the Q30 score of the data. FastQC (v0.11.3) was used to evaluate quality metrics such as GC content distribution and base quality scores. The data were then aligned to the reference genome (hg38) using BWA (v0.7.17). GATK (v3.8) was applied for the detection of SNPs and indels, while CNVkit (v0.9.6) was used for CNV detection.

Results

Instrument overview

Nanodigmbio (Nanjing) Biotechnology Co., Ltd. has developed a series of fully automated workstations for next-generation sequencing (NGS) library preparation. The NadAuto-48R model evaluated in this study is one of these systems. The NadAuto-48R workstation supports automated next-generation sequencing (NGS) library preparation with a maximum batch size of 48 samples per run. As illustrated in Fig. 1a, the system is equipped with a fully enclosed structure with a transparent sliding door. The external hardware configuration includes a three-dimensional motion system with a multi-axis robotic arm, a configurable deck accommodating up to 24 SBS-standard plate positions, and control components including an emergency stop button, status indicator lights, and a touchscreen interface. The enclosure is constructed from UV-resistant materials and incorporates a UV sterilization unit and HEPA filtration.

The internal layout of the workstation is shown in Fig. 1b. The deck positions accommodate modules including a thermal cycler for PCR amplification, temperature-controlled modules for reagent storage, a magnetic stand for bead-based purification, a mixing module for consistent agitation, and a quantitative module for fluorometric nucleic acid quantification, which enables automated measurements at defined steps of the workflow.

The software control system, shown in Fig. 1c, includes a graphical workflow editor, run status monitoring, and a pre-run workflow simulation function. Role-based user access and electronic operation logs are implemented to support compliance with Good Laboratory Practice (GLP) requirements. Ethernet and serial interfaces are used for communication between software and hardware modules. Together, these components enable execution of automated workflows for genomic applications, including tumor profiling and infectious disease detection.

In some automated workflows, reagents are supplied in individual tubes and require manual transfer prior to automated processing, which increases the likelihood of handling errors. During prolonged operations, unsealed reagent containers may be subject to evaporation or volume changes. In this system, a plate-based reagent format is used. As shown in Fig. 1d, the NadPrep® EZ DNA Library Preparation Kit Plate consists of pre-dispensed, single-reaction aliquots sealed with aluminum film. The plate format is compatible with automated handling during the library preparation workflow. At the same time, plate-based reagent formats can also increase per-run sample throughput during manual operation.

Cross-contamination performance evaluation

Cross-contamination, defined as the unintended transfer of biological material between samples within a single run, represents a critical concern in high-throughput automated workflows, as it may lead to biased results or false-positive signals. To minimize this risk, the system integrates multiple anti-contamination measures at both structural and operational levels, including UV sterilization and HEPA filtration to reduce aerosol dispersion, a fully enclosed automated PCR module enabling closed-lid amplification, and the use of single-use filtered pipette tips to prevent aerosol backflow during liquid handling. To evaluate the anti-cross-contamination performance of the system, a series of dedicated experiments was conducted. Anti-cross-contamination performance was assessed using a checkerboard experimental design (Fig. 2).

Fig. 1. Overview of the automated NGS workstation. (**a**) Front view of the automated workstation used in this study. (**b**) Top view of the configurable deck layout. (**c**) Screenshot of the software control interface. (**d**) Plate-based reagent format used for automated DNA library preparation.

PCR amplification was carried out using a two-stage protocol (6 + 12 cycles). To simulate potential aerosol exposure during automated operation, the PCR sealing lid was intentionally opened and closed once after completion of the first stage. Following amplification, all samples were subjected to magnetic bead purification and subsequent quantification. To control for background signals resulting from residual primers retained by magnetic beads, appropriate negative controls were included, in which nuclease-free water was used as the amplification template. Quantitative analysis showed that background signal concentrations ranges from 0.01 to 0.03. The experimental group consisted of alternating wells containing nuclease-free water and DNA libraries (5 ng input). Quantitative analysis (Table 1) showed no statistically significant increase in signal in water-only wells compared with controls.

Performance of automated pre-library construction: yield, variability, and fragment size distribution

To evaluate the performance of the fully-automated NGS workstations (NadAuto-16R and NadAuto-48R) in pre-library construction, a key step in precision genomics, library yield, coefficient of variation (CV), and fragment size distribution were assessed across multiple experimental batches. All libraries were prepared from human genomic DNA (Promega, G1471; 50 ng per sample) using the NadPrep® EZ DNA Library Preparation Kit Plate (for Illumina®) with 7 PCR cycles, performed using integrated instrument scripts.

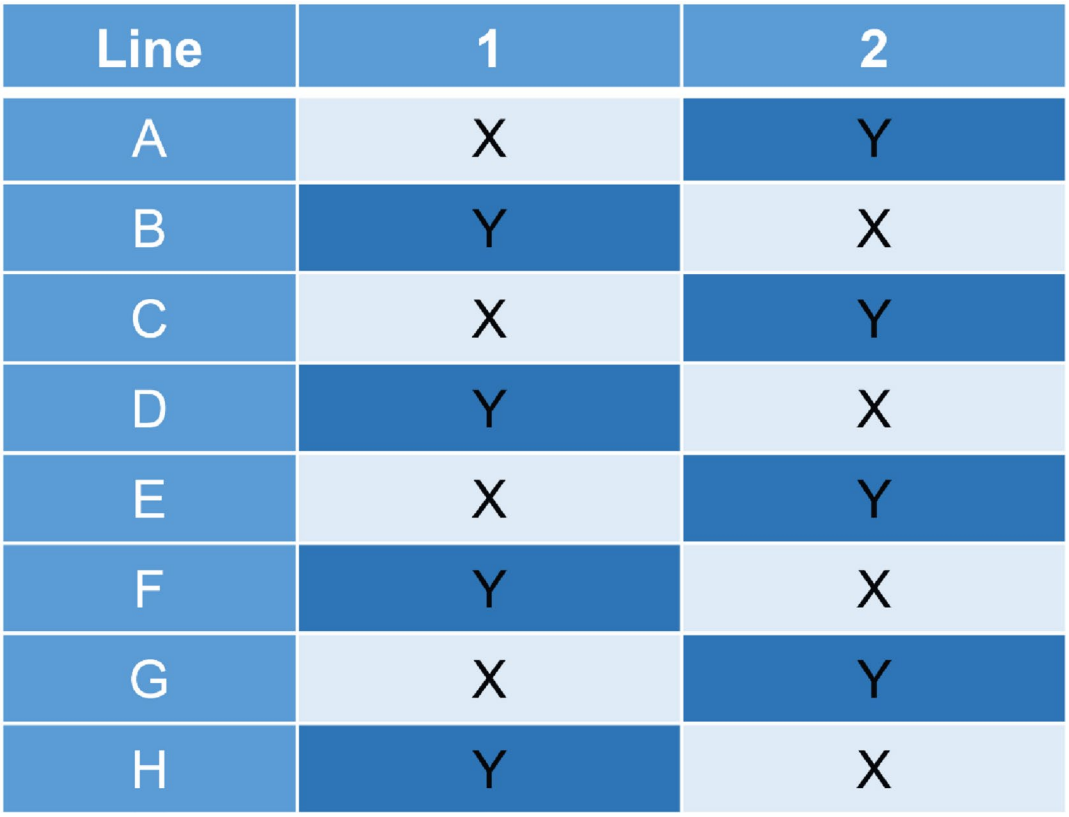

Fig. 2. Checkerboard experimental design for assessment of inter-well cross-contamination. X and Y indicate libraries derived from different species or sample types.

Sample type	Number of wells	Library yield range (ng/μl)
gDNA	8	32.8–36.6
NFW	8	0.01

Table 1. Library yield of checkerboard-based cross-contamination assay. NFW, nuclease-free water In addition to quantitative background signal measurements, the sequencing experiment was conducted to further evaluate inter-well cross-contamination. Human genomic DNA and *Escherichia coli* nucleic acid samples were alternately arranged across the plate and processed in parallel using distinct index adapters. Each resulting library was sequenced to an average sequencing data volume of approximately 10 Gb. Sequencing reads from individual wells were independently aligned to the corresponding reference genomes, and mapping specificity was assessed on a per-index basis. Inter-well cross-contamination was evaluated by identifying index-consistent reads exhibiting unexpected cross-species alignment (Table 2). Bioinformatic analysis demonstrated the absence of aberrant cross-species read mapping within libraries sharing the same index, with a maximum observed contamination ratio of 0.0003% across all libraries.

The NadAuto-48R workstation was evaluated under 24- and 48-sample throughput modes. Library yields exceeded 2500 ng in both configurations, with inter-batch CV values below 8% (Fig. 3a,b). Fragment size distribution analysis of 8 randomly selected samples from each workflow showed comparable profiles centered within the Illumina®-compatible range, with no detectable primer dimer signals or enrichment of large fragments (Fig. 3c,d). These results reflect consistent enzymatic fragmentation and library construction across runs.

Similarly, the NadAuto-16R workstation was evaluated in both 8- and 16-sample workflows across independent batches. Both workflows yielded around 1600 ng per library (Supplementary Fig. S1a,b). The inter-batch CV was below 7% for the 8-sample workflow (Supplementary Fig. S1a) and below 4% for the 16-sample workflow (Supplementary Fig. S1b), indicating comparable variability across throughput configurations.

In summary, both NadAuto systems generated pre-libraries with comparable yields, low inter-batch coefficients of variation, and fragment size distributions within the expected range.

Number	Sample type	Species alignment signals (human)	Species alignment signals (*E. coli*)
A1	Human gDNA	100.000%	0.000%
A2	*E. coli* gDNA	0.000162%	99.99985%
B1	*E. coli* gDNA	0.000301%	99.99988%
B2	Human gDNA	100.000%	0.000%
C1	Human gDNA	100.000%	0.000%
C2	*E. coli* gDNA	0.000036%	99.99986%
D1	*E. coli* gDNA	0.000042%	99.99980%
D2	Human gDNA	100.000%	0.000%
E1	Human gDNA	100.000%	0.000%
E2	*E. coli* gDNA	0.000076%	99.99995%
F1	*E. coli* gDNA	0.000157%	99.99962%
F2	Human gDNA	100.000%	0.000%
G1	Human gDNA	100.000%	0.000%
G2	*E. coli* gDNA	0.000189%	99.99982%
H1	*E. coli* gDNA	0.000100%	99.99987%
H2	Human gDNA	100.000%	0.000%

Table 2. Mapping specificity of checkerboard-based cross-contamination assay. *E. coli*, *Escherichia coli*.

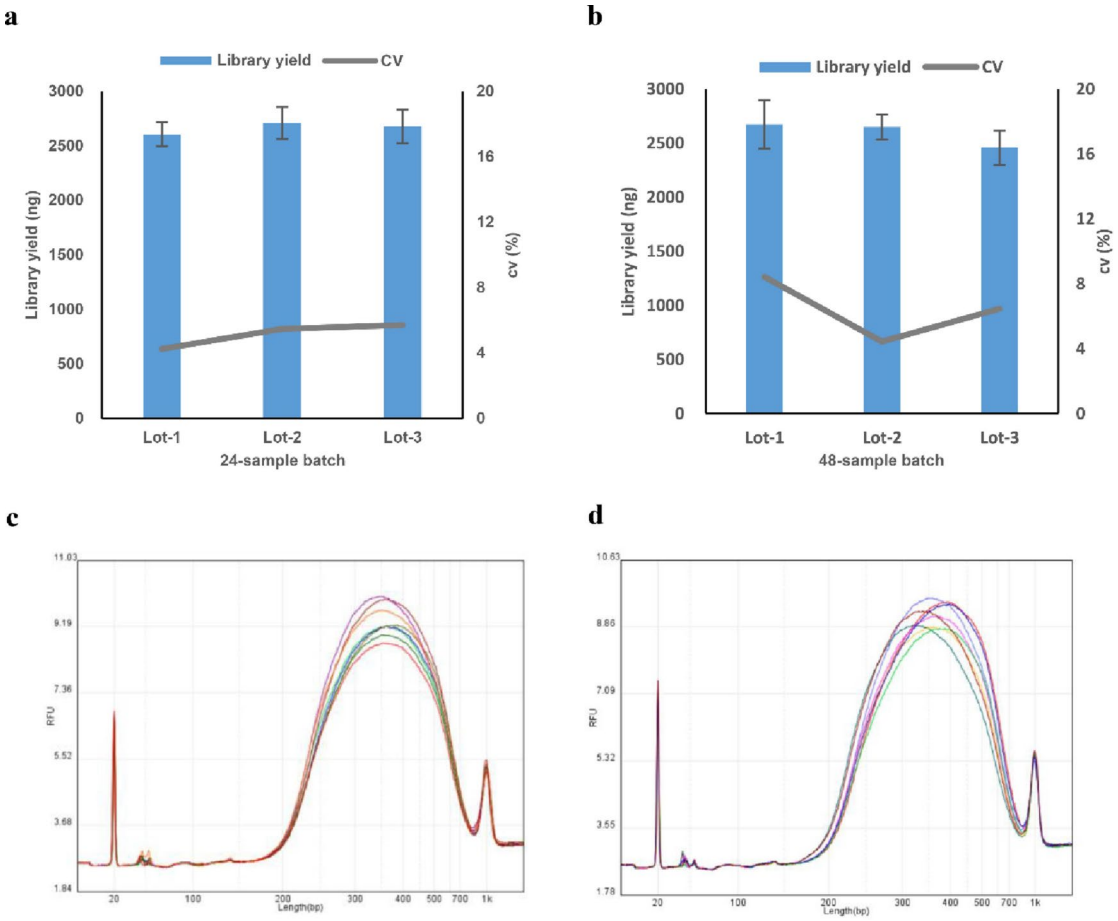

Fig. 3. Pre-library construction performance of the automated NGS workstation. (**a**) Library yield across 24-sample batches ($n = 3$). (**b**) Library yield across 48-sample batches ($n = 3$). (**c**) Fragment size distributions of eight randomly selected libraries from 24-sample batches. (**d**) Fragment size distributions of eight randomly selected libraries from 48-sample batches. Libraries were prepared from human genomic DNA (Promega G1471; 50 ng per sample) using a plate-based DNA library preparation kit compatible with Illumina® sequencing, with seven PCR cycles.

Performance of automated targeted capture sequencing: quality and inter-batch reproducibility

To assess the targeted capture performance of the automated NadAuto workstations (NadAuto-48R and NadAuto-16R), hybridization capture experiments were conducted using pre-constructed libraries from human genomic DNA standards (Promega, G1471) and NadPrep ES Hybrid Capture Reagents Kit under standardized conditions. Key metrics, including sequencing quality, on-target rates, coverage uniformity, and GC bias, were subsequently analyzed.

The NadAuto-48R workstation was evaluated under high-throughput conditions. Hybridization capture was performed using 500 ng of pre-library and the LungCancer Panel v1.0, with hybridization times of 2–16 h. Across independent batches, on-target rates ranged from 88.5% to 91.2%, Q30 values exceeded 93.2% and duplication rates were ≤ 12.5%. Coverage analysis showed a mean depth ≥ 0.2 × (0.2× mean) for at least 99.3% of target regions, with Fold 80 values ranging from 1.44 to 1.49 (Fig. 4a,b). GC coverage ranged from 0.97× to 1.04× within the 40–60% GC content range, and coverage across regions with extreme GC content was observed (Fig. 4c,d).

The NadAuto-16R system was evaluated in medium-throughput workflows using 4- or 8-sample formats. Sequencing data showed Q30 ≥ 92.6%, mapping rates > 99.5%, and duplication rates ≤ 13.1%. The on-target rates were ≥ 72% (Supplementary Fig. S2a). Coverage uniformity was high, with ≥ 99.0% of targets covered at 0.2× mean depth and Fold 80 values consistently below 1.5 (Supplementary Fig. S2b). GC coverage ranged from 0.95× to 1.05× of the mean within the 40–60% GC content range (Supplementary Fig. S2c,d).

Comparative analysis of detection performance: fully-automated vs. manual workflows

A comparative study was conducted using three human genomic DNA standards with defined genetic backgrounds, including GW-OGTM800, GW-OGTM001, a negative control (GW-OGTM005, wild-type). Library preparation began with 50 ng of DNA, which was enzymatically fragmented for 20 min and pre-amplified with 6 PCR cycles. 500 ng of each pre-library was vacuum-concentrated and hybridized with the NanOnco Plus Panel v3.0 for 16 h.

The primary endpoints for this evaluation included mutation detection accuracy, consistency, and key sequencing quality metrics such as coverage uniformity. Comparative analysis showed no statistically significant differences between the automated and manual workflows with respect to base quality (Q30), mapping rates, or

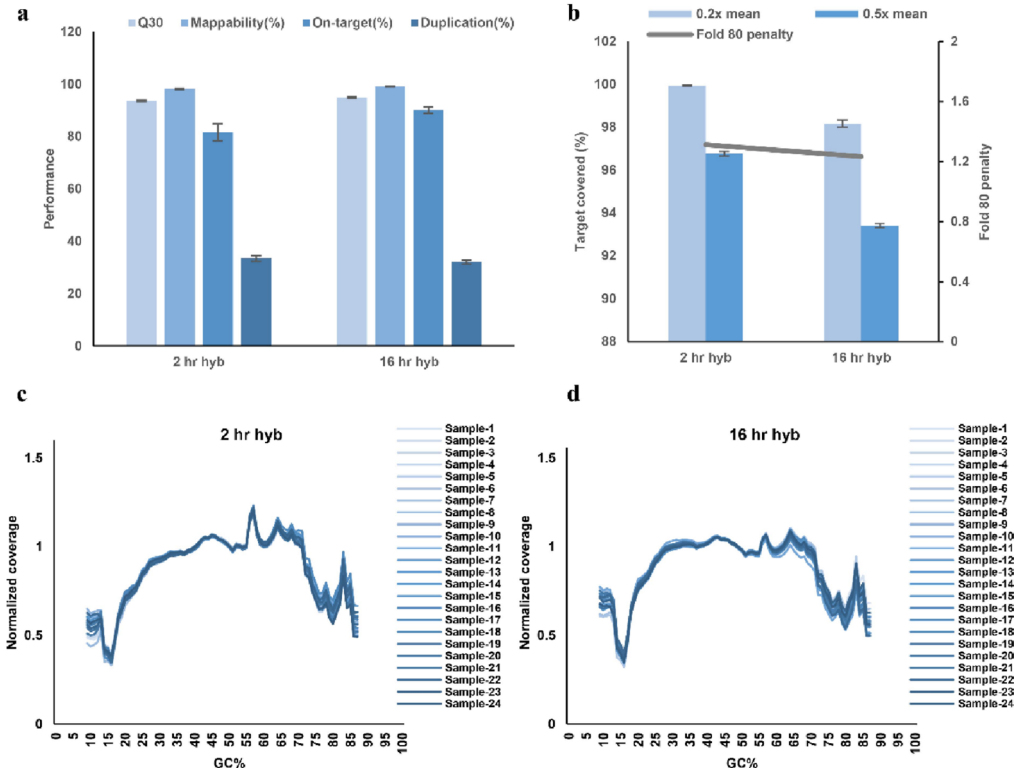

Fig. 4. Targeted capture sequencing performance of the automated NGS workstation. (**a**) Sequencing quality metrics obtained using hybridization times of 2 h and 16 h ($n = 3$). (**b**) Coverage uniformity for 2 h and 16 h hybridization conditions ($n = 3$). (**c**,**d**) GC bias profiles under 2 h and 16 h hybridization conditions ($n = 3$). Pre-libraries were prepared from human genomic DNA (Promega G1471; 50 ng per sample) using a plate-based DNA library preparation kit compatible with Illumina® sequencing, with seven PCR cycles. For hybrid capture, 500 ng of pre-library was used with a lung cancer–targeted gene panel and compatible hybrid capture reagents, following a single-plex hybridization protocol. "Hyb" denotes hybridization.

coverage uniformity (Supplementary Fig. S3). A lower on-target rate was observed for the automated workflow relative to the manual process. The only minor deviation was a slightly reduced on-target rates, which was attributed to standardized reagent delivery and, crucially, did not compromise the accuracy of variant detection.

For single-nucleotide variant (SNV) and insertion/deletion (InDel) detection, both workflows demonstrated diagnostic concordance (100%) across all samples, with all variant allele frequencies (VAFs) detected showed deviations of less than 5%. For the complex variant profile of GW-OGTM800 (contains 14 variants), the two workflows produced highly correlated VAFs (R^2 = 0.99). Low-frequency variants down to ~1% (e.g., *EGFR* T790M) were detected using the applied variant-calling pipeline based on Vardict (Fig. 5).

Furthermore, both workflows detected oncogenic amplifications, with all copy number estimates for all targets showed relative errors of less than 5%. For *MET* amplification (~4 copies) in sample GW-OGTM001, both manual and automated measurements showed close agreement with the reference values, exhibiting a relative difference of < 1% after outlier exclusion (manual: 4.19–4.23; automated: 3.62–4.20) (Fig. 6a). Similarly, for *ERBB2* amplifications (8 copies in GW-OGTM001; 5.5 copies in GW-OGTM800), relative differences of < 3% and < 2%, respectively, and not statistically significant (Fig. 6a,b).

Discussion

The increasing adoption of next-generation sequencing (NGS) in precision genomics has underscored persistent limitations in upstream processes, particularly library preparation and hybridization capture. Manual workflows remain labor-intensive and are prone to operator-dependent variability, which can affect reproducibility and consistency. In this study, we assessed the performance of a fully automated NGS workstation, focusing on its ability to address these limitations through standardized and reproducible processing.

Across multiple pre-library construction batches, the automated system demonstrated relatively low inter-batch variability, with coefficients of variation consistently below 8%. This level of variability is lower than values commonly reported for manual workflows. This improvement is attributable to the system's high liquid-handling precision (CV < 5%) and stable incubation control, which together reduce v fluctuations during enzymatic reactions and bead-based cleanup steps. Such improvements are particularly relevant in clinical sequencing contexts, where variability in library yield or fragment size distribution can influence downstream analytical sensitivity.

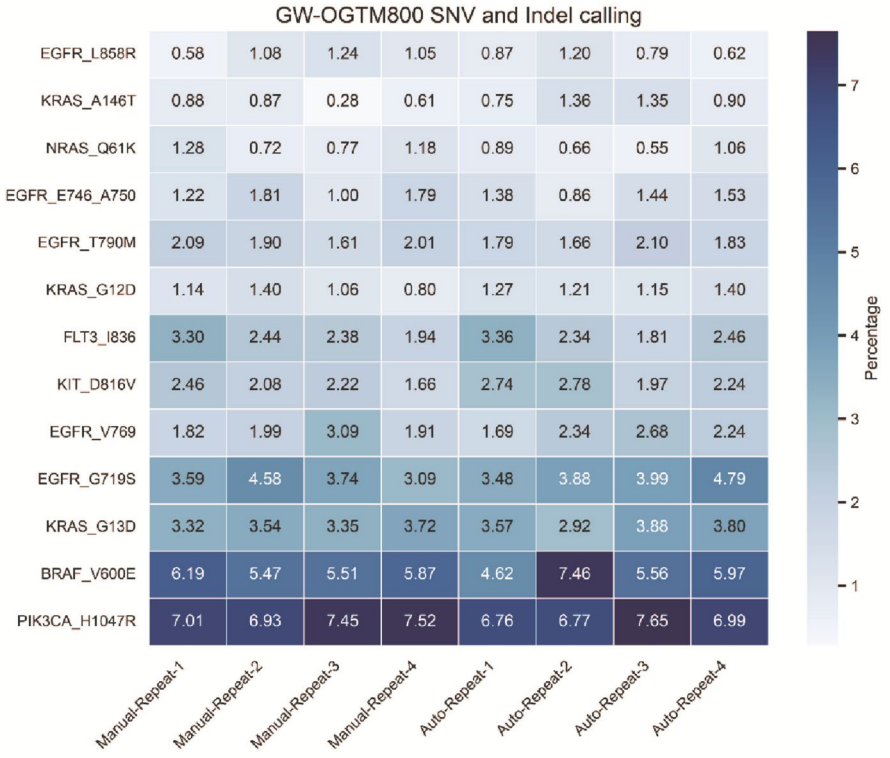

Fig. 5. Performance of targeted capture sequencing for SNV and indel detection using the automated NGS workstation. SNV and indel detection was evaluated using sample GW-OGTM800, with GW-OGTM005 included as a negative control. All reactions were performed using a four-plex hybridization protocol. Pre-libraries were prepared using a plate-based DNA library preparation kit compatible with Illumina® sequencing with six PCR cycles. A total of 500 ng of pre-library was vacuum-concentrated and subjected to 16 h hybridization with a targeted oncology gene panel using compatible hybrid capture reagents.

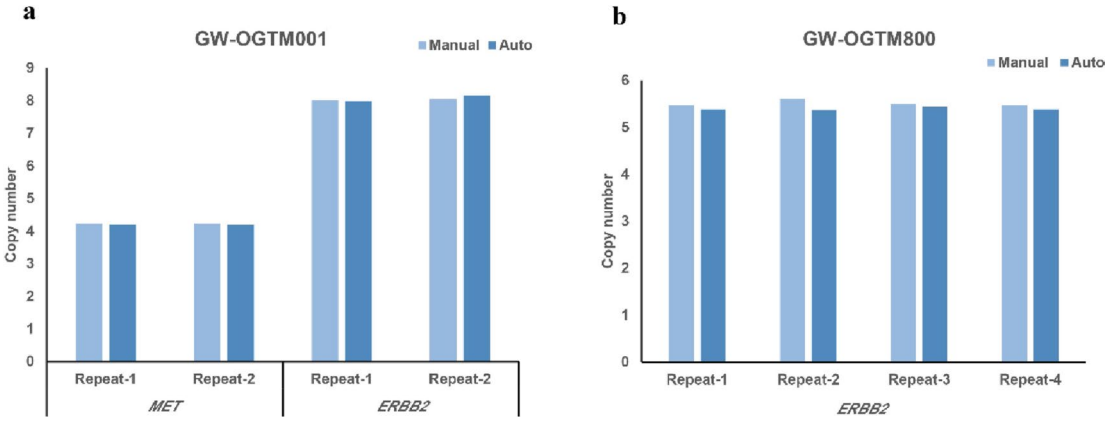

Fig. 6. Performance of targeted capture sequencing for CNV detection using the automated NGS workstation. (**a**) Copy number variation (CNV) analysis of *MET* and *ERBB2* in sample GW-OGTM001, with GW-OGTM005 included as a negative control. All reactions were performed using a single-plex hybridization protocol. (**b**) CNV analysis of *ERBB2* in sample GW-OGTM800, with GW-OGTM005 included as a negative control, using a four-plex hybridization protocol. Pre-libraries were prepared using a plate-based DNA library preparation kit compatible with Illumina® sequencing with six PCR cycles. A total of 500 ng of pre-library was vacuum-concentrated and hybridized for 16 h with a targeted oncology gene panel using compatible hybrid capture reagents.

Targeted capture experiments further demonstrated stable sequencing performance across different hybridization conditions. Temperature stability was associated with uniform coverage profiles and limited GC-dependent bias. These findings suggest that precise thermal control may help maintain consistent hybridization efficiency across regions with variable GC content.

In addition to analytical performance, the automated workflow reduced hands-on time relative to manual preparation. Although increased throughput is a general characteristic of automated systems, reduced manual intervention may also help limit operator-induced variability and facilitate more standardized implementation across different laboratories and users.

Several limitations of this study should be acknowledged. Performance evaluation was conducted using high-quality human genomic DNA standards, providing a controlled experimental framework but not fully reflecting the diversity of clinical sample types. Future work should extend validation to more complex and clinically relevant sample types, such as FFPE-derived DNA and RNA specimens, which present additional challenges related to input quality and fragmentation. Comparative studies with other automated platforms and long-term operational assessments will further clarify performance characteristics under routine use. Integration of upstream extraction and implementation of real-time quality control modules represent logical extensions toward fully automated, sample-to-sequence workflows.

In summary, the findings demonstrate that automated NGS library preparation and targeted capture can achieve reproducible and consistent performance while reducing workflow variability. Such systems are expected to play an increasingly important role in supporting standardized and scalable genomic analyses as NGS continues to expand in clinical and translational settings.

Data availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive[21] in National Genomics Data Center[22], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA017097) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa-human.

Received: 19 November 2025; Accepted: 9 March 2026
Published online: 11 March 2026

References

1. Metzker, M. L. Sequencing technologies—the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
2. Mandlik, J. S., Patil, A. S. & Singh, S. Next-generation sequencing (NGS): Platforms and applications. *J. Pharm. Bioallied Sci.* **16**, S41–S45 (2024).
3. Chougule, A. et al. Comprehensive development and implementation of good laboratory practice for NGS based targeted panel on solid tumor FFPE tissues in diagnostics. *Diagn. Basel Switz.* **12**, 1291 (2022).
4. Lan, J. H. & Zhang, Q. Clinical applications of next-generation sequencing in histocompatibility and transplantation. *Curr. Opin. Organ. Transpl.* **20**, 461–467 (2015).
5. Najmabadi, H. et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57–63 (2011).
6. Calvo, S. E. et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci. Transl Med.* **4**, 118ra10 (2012).
7. Baum, A., Sachidanandam, R. & García-Sastre, A. Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16303–16308 (2010).

8. Alvim, L. B., Marinho, F. L., Freire, M. C. & Zauli, D. A. B-391 is comprehensive cancer panel by next-generation sequencing (NGS) more efficient than cancer-specific NGS panel in the management of non-small cell lung cancer patients? *Clin. Chem.* **69**, hvad097.700 (2023).
9. Head, S. R. et al. Library construction for next-generation sequencing: overviews and challenges. *BioTechniques* **56**, 61–64 (2014).
10. Socea, J. N., Stone, V. N., Qian, X., Gibbs, P. L. & Levinson, K. J. Implementing laboratory automation for next-generation sequencing: Benefits and challenges for library preparation. *Front. Public. Health.* **11**, 1195581 (2023).
11. van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: tone down the bias. *Exp. Cell. Res.* **322**, 12–20 (2014).
12. Chao, T., Chen, P. & Huang, W. B-234 the quality in NGS: Do you know how much we lost? *Clin. Chem.* **69**, hvad097560 (2023).
13. Feng, K., Costa, J. & Edwards, J. S. Next-generation sequencing library construction on a surface. *BMC Genom.* **19**, 416 (2018).
14. Zimmerman Zuckerman, E. et al. Automation of hybridization and capture based next generation sequencing library preparation requires reduction of on-deck bead binding and heated wash temperatures. *SLAS Technol.* **27**, 214–218 (2022).
15. Shi, C. et al. Development and clinical applications of an enclosed automated targeted NGS library preparation system. *Clin. Chim. Acta Int. J. Clin. Chem.* **540**, 117224 (2023).
16. Antonios, K., Croxatto, A. & Culbreath, K. Current state of laboratory automation in clinical microbiology laboratory. *Clin. Chem.* **68**, 99–114 (2021).
17. Bendlin, A. et al. Evaluation of a commercial NGS service for detection of bacterial and fungal pathogens in infectious ulcerative keratitis. *Vet. Ophthalmol.* **26**, 500–513 (2023).
18. Tan, B. et al. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. *Front. Microbiol.* **6**, 1027 (2015).
19. Shen-Gunther, J. & Easley, A. H. P. V. HBV, and HIV-1 viral integration site mapping: A streamlined workflow from NGS to genomic insights of carcinogenesis. *Viruses* **16**, 975 (2024).
20. Döring, M. et al. geno2pheno[ngs-freq]: a genotypic interpretation system for identifying viral drug resistance using next-generation sequencing data. *Nucleic Acids Res.* **46**, W271–W277 (2018).
21. Zhang, S. et al. The GSA Family in 2025: A broadened sharing platform for multi-omics and multimodal data. *Genom. Proteom. Bioinf.* **23**, qzaf072 (2025).
22. CNCB-NGDC Members and Partners. Database Resources of the National Genomics Data Center. China National Center for Bioinformation in 2025. *Nucleic Acids Res.* **53**, D30–D44 (2025).

Author contributions

Conceived and designed the experiments: Y.C., R.S., X.W. Performed the experiments: Y.C., X.W. Analyzed the data: Y.C., X.W. Wrote the paper: Y.C., R.S., X.W.

Funding

This study was supported by Nanodigmbio (Nanjing) Biotechnology Co.Ltd.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-43941-7.

Correspondence and requests for materials should be addressed to C.Y. or S.R.