The High Compatibility of Twist Exome 2.0 on MGI's automated system and DNBSEQ sequencing platform facilitates WES research

In this study, MGI's automated system MGISP-960 & MGISP-100 and sequencing platform DNBSEQ-G400 & DNBSEQ-G99 were utilized for Twist WES library preparation and sequencing, respectively.

The performance was comprehensively evaluated and the results indicate that the Twist WES solution is perfectly compatible with MGI's automated system and DNBSEQ sequencing platform. Moreover, 8-plex hybridization strategy and 6Gb data for bioanalysis are suggested for the WES workflow.

Recommended application: Whole Exome Sequencing (WES)

Recommended models: DNBSEQ-G400RS, DNBSEQ-G99ARS, DNBSEQ-T7RS (Genetic Sequencer)
MGISP-960RS,MGISP-100RS (Automated Sample Preparation System)

- ## A best-in-class solution for WES application

  Twist Exome 2.0 exhibits high uniformity and low off-target rate, enabling high sequencing efficiency and data quality.

- ## Perfect compatibility with DNBSEQ sequencing platform

  MGIEasy Universal Library Conversion kit enables Twist Exome 2.0 perfectly work on DNBSEQ sequencing platform.

- ## Data output is efficient and high-quality

  DNBSEQ sequencing technology has excellent features such as high accuracy, low duplication rate and low index hopping rate.

- ## Automatic operation compatible

  MGI provides automated solutions for experimental processes, which can greatly save labor cost and improve efficiency.

Background

The advent of massively parallel sequencing (MPS) technology has greatly changed the diagnostic process of rare germline and somatic variants, improving the accuracy and speed of reporting and decreasing the cost[1]. WES is a popularly utilized MPS application for sequencing protein-coding regions of the genome[1]. Although the exonic region accounts for only about 2% of the entire genome, 85% of the reported pathogenic variants are located in it[1,2]. Therefore, compared with WGS (Whole genome sequencing), WES can effectively cover more than 95% of exonic regions with lower sequencing data and analysis cost[2]. WES can reveal more than 5000 germline and somatic variants through a single test[3-8], so tumor molecular diagnosis laboratories currently adopt WES as a regular detection method.

A satisfied WES solution should comprehensively cover the protein coding regions with high uniformity and on-target rate. Among the various commercial WES kits, Twist Exome 2.0 is one of the most popular WES target enrichment panels, exhibiting features like high uniformity and low off-target rate. The probes designed in Twist Exome 2.0 are proprietary double-stranded DNA that could target both sense and antisense chain and improve enrichment sensitivity. Moreover, the Twist Exome 2.0 panel also includes additional clinically relevant non-coding pathogenic and likely pathogenic variants, making the detection of rare germline and somatic variants to be fulfilled more easily.

The DNBSEQ sequencing technology developed by MGI displays attractive features such as high accuracy and sensitivity, ultra-low duplication rate and low index hopping rate[9]. Based on this platform, MGI has launched a series of genetic sequencers, such as DNBSEQ-G400 and DNBSEQ-G99, which can meet various MPS demands in life science research and clinical application. Additionally, MGI has also developed high-quality automated systems, which can improve library consistency by reducing sample-to-sample variation. This system can greatly increase library prep throughput and efficiency, speed up processing time, thus improving consistency of sequencing results and reducing cost for laboratories performing MPS[8].

However, the compatibility of Twist Exome 2.0 with MGI's automated system and sequencing platform remains elusive. Hence, this study was designed to systematically evaluate the performance of Twist Exome 2.0 target enrichment based on MGI's automated system (MGISP-960 & MGISP-100) followed by sequencing on MGI sequencing platform (DNBSEQ-G400 & DNBSEQ-G99).

Study description

In order to verify the compatibility of Twist Exome 2.0 with MGI's products, this study used MGI's automated system MGISP-960 or MGISP-100 for library preparation, DNBSEQ-G400 or DNBSEQ-G99 for genetic sequencing, matured WES bioinformatics workflow for data analysis. The comprehensive performance evaluation results showed that Twist Exome 2.0 performed perfectly on MGI's automated system and DNBSEQ sequencing platform, 8-plex is a recommended hybridization strategy, and intercepting 6Gb data per sample is cost-effective and efficient for bioinformatic analysis.

Materials and Methods

Sample preparation

In this study, commercial standard Human Genomic DNA (Takara PN: 636401) was adopted for performance evaluation experiments.

Library preparation and sequencing

50 ng of genomic DNA was used as input for library preparation. In the pre-hybridization step, libraries were prepared with the Twist Library Preparation EF Kit 2.0 and Twist Universal Adapter System based on MGI's automated system MGISP-960 or MGISP-100 with 8 PCR cycles. The detailed operation procedures can be referred to related instructional manuals. During the manual hybridization step, 1500 ng of the pre-hybridization product was aliquoted(1500 ng/sample for 1-plex, 187.5 ng/sample for 8-plex) and enriched with Twist Exome 2.0. In the post-hybridization step (including washing, Post-PCR and purification), these libraries were amplified with AC-primers from MGIEasy Universal Library Conversion kit (App-A) through 8 post-PCR cycles and this process was performed either by automated system (MGISP-960 or MGISP-100) or manual. Then, the final libraries were prepared into DNA nanoballs (DNBs) with the above App-A kit through single-strand DNA circularization.

Subsequent sequencing was carried out on DNBSEQ-G400 or DNBSEQ-G99 with paired-end 150 bp (PE150) recipe using App-C as the sequencing primer. Specifically, a portion of 1-plex libraries were sequenced on DNBSEQ-G99, the rest of 1-plex and all 8-plex libraries were sequenced on DNBSEQ-G400. In this study, the hybridization steps were all carried out manually; the pre-hybridization step was performed on either MGISP-960 or MGISP-100; the post-hybridization step was performed on MGISP-960 or manually. G400-8plex (960, Manual) refers to a 8-plex library prepared on MGISP-960 in the pre-hybridization step, and manually in the post-hybridization step; followed by sequencing on DNBSEQ-G400. The meanings of other sample names follow this analogy.

Bioinformatics analysis

In this study, the WES bioinformatics analysis was completed on the MegaBOLT Bioinformatics analysis accelerator. The brief process is as follows: 1. the raw reads obtained from sequencing were filtered through the data quality control process to obtain clean reads;2. these clean reads were mapped to the human reference genome (hg38) by Burrows-Wheeler Aligner (BWA, http://bio-bwa.sourceforge.net/); 3.the mapping results were recorded in SAM files, and after position sorting and duplicate marking, they were compressed into BAM files, followed

by base quality score recalibration (BQSR) and BAM/VCF statistics; 4. Finally, a visual report was generated, which includes duplication rate , reads clean rate, reads mapping rate, Q30, average depth (rmdup), reads on target rate, fold 80 base penalty, target coverage (%) and other key metrics.

Sample collection	Library preparation and sequencing	Bioinformatics analysis	Result analysis
Commercial standard Human Genomic DNA from Takara	Twist Library Preparation EF Kit 2.0 Twist Universal Adapter System Twist Exome 2.0 Twist Standard Hyb and Wash Kit v2  MGIEasy Universal Library Conversion kit (App-A)  MGISP-960 or MGISP-100  DNBSEQ-G400 or DNBSEQ-G99	WES analysis integrated in MegoBOLT	Systematic evaluation of Twist Exome 2.0 based on MGI's automated system and DNBSEQ sequencing platform

Results

High-quality WES libraries prepared by MGI's MGISP automated system

In this study, Twist Library Preparation EF Kit 2.0 and Twist Exome 2.0 were utilized for WES library preparation in the pre-hybridization and target enrichment steps. The performance of MGISP automated system was comprehensively evaluated by comparing manual and automated solutions (MGISP-960 or MGISP-100). Figure 1 shows the basic library information generated from automated (sample No. 9) or manual (sample No. 11) solutions. The pre-hybridization step of both samples was carried out on MGISP-960 (Figure 1A,1B left), and the post-hybridization step was performed on MGISP-960 or manually (Figure 1A,1B right). Under the same input of 50ng DNA standard, these libraries prepared with both solutions could all meet the requirements for subsequent DNB making (Figure 1A). Meanwhile, the DNA size distribution determined with the Agilent TapeStation system suggest both libraries are high-quality: the fragment size are as expected, no significant difference between those two solutions and no primer dimers or other non-specific peaks are observed (Figure 1B).

A

Sample number	Input amount (ng)	Index name	Pre-hybridizatior library prep	Hybridization	Post-hybridization library prep	Pre-hybridization library yield (ng)	Post-hybridization library yield (ng)
9	50	257	MGISP-960	Manual	MGISP-960 (1-plex)	3800	495
11	50	273	MGISP-960	Manual	Manual(1-plex)	3660	702

B

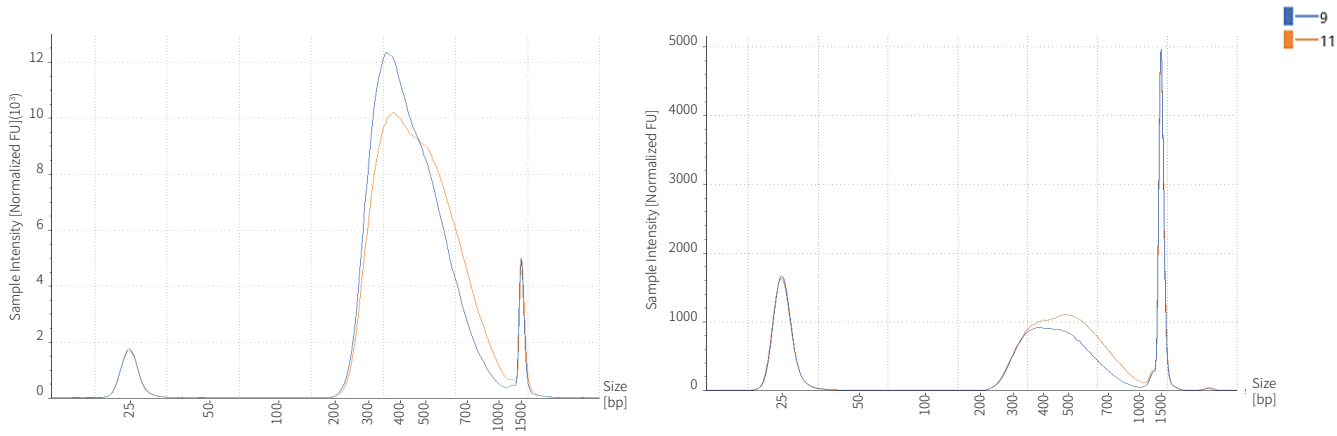

Figure 1. A preliminary comparison of Twist WES libraries prepared based on manual or automated solutions. (A) The comparison of libraries obtained with automated (sample 9) and manual (sample 11) solutions. (B) The size distribution of pre-hybridization(left) and post-hybridization(right) libraries obtained from automated (sample 9, blue line) and manual (sample 11, orange line) solutions were analyzed with Agilent TapeStation system. The DNA fragment distribution from both solutions are as expected and comparable, and no non-specific peaks are observed.

6Gb is a recommended data size for WES analysis

To assess the amount of sequencing data required for WES analysis, this study independently intercepted 2, 4, 6, and 8 Gb sequencing data from the G99-1plex (960, 960) group and compared their results quality. The results show that the duplication rate, average depth (rmdup) and target coverage increase along with the data size (Figure 2A, B, E). However, reads on target rate and fold 80 base penalty are comparable with each other. Reads on target rate are higher than 77.4% (Figure 2C), and fold 80 base penalty are less than 1.5. However, fold 80 base penalty is 1.41 when the data size is 2Gb, which is slightly higher than 4, 6 and 8 Gb (Figure 2D). In conclusion, it's suggested that the more sequencing data, the more accurate the analysis results.

Considering Twist Exome 2.0 probes covers 36.5 Mb of human protein coding region, the sequencing depth of ~150× can be achieved with only 6Gb data. In WES research, a sequencing depth of 150× could already obtain relatively accurate results. In addition, all metrics are good enough with 6Gb data. Thus, it is recommended to intercept 6GB data for WES bioinformatics analysis based on the comprehensive consideration of sequencing cost and time,.

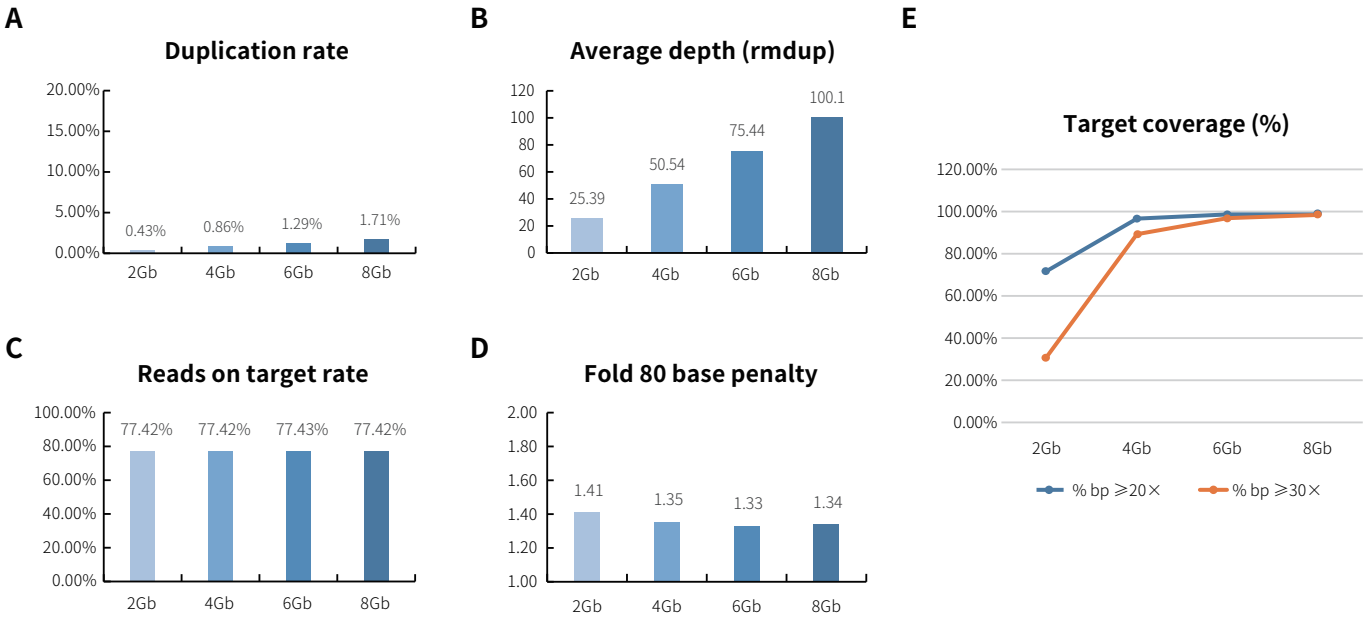

Figure 2. To determine the amount of data required for WES analysis, sequencing results of G99-1plex (960,960) were intercepted to 2,4,6 and 8Gb, respectively. Although the duplication rate increases along with the data size, it's still low enough to meet the requirements (A), similar result is also observed with the average depth (rmdup) (B). Reads on target rate and fold 80 base penalty remain basically unchanged, indicating that Twist Exome 2.0 has high specific enrichment rate (C) and coverage uniformity (D) for exons. The target coverage(≥20× and ≥30×) also increase along with the data size and reach almost saturation at 6Gb (E).

8-plex hybridization strategy is recommended in WES

In this study, sequencing data from G400-1plex (960, 960) and G400-8plex (960, 960) were intercepted to 6Gb to explore which hybridization strategy is better (1-plex or 8-plex). The duplication rate of G400-8plex (960, 960) is 1.71%, slightly higher than G400-1plex (960, 960). However, the reads on target rate, target coverage (% bp ≥ 30×), average depth (rmdup) and fold 80 base penalty of the former are 80.18%, 97.24%,79.56 and 1.28, slightly better than the latter group (Figure 3). Results for G400-1plex(100,960) and G400-8plex (100,960) are not presented in this article.

Conclusively, 8-plex hybridization strategy can still gain high-quality results while saving costs and improving efficiency. Therefore, the 8-plex strategy is recommended for hybridization when conducting WES studies based on MGI's automated system and DNBSEQ sequencing platform.

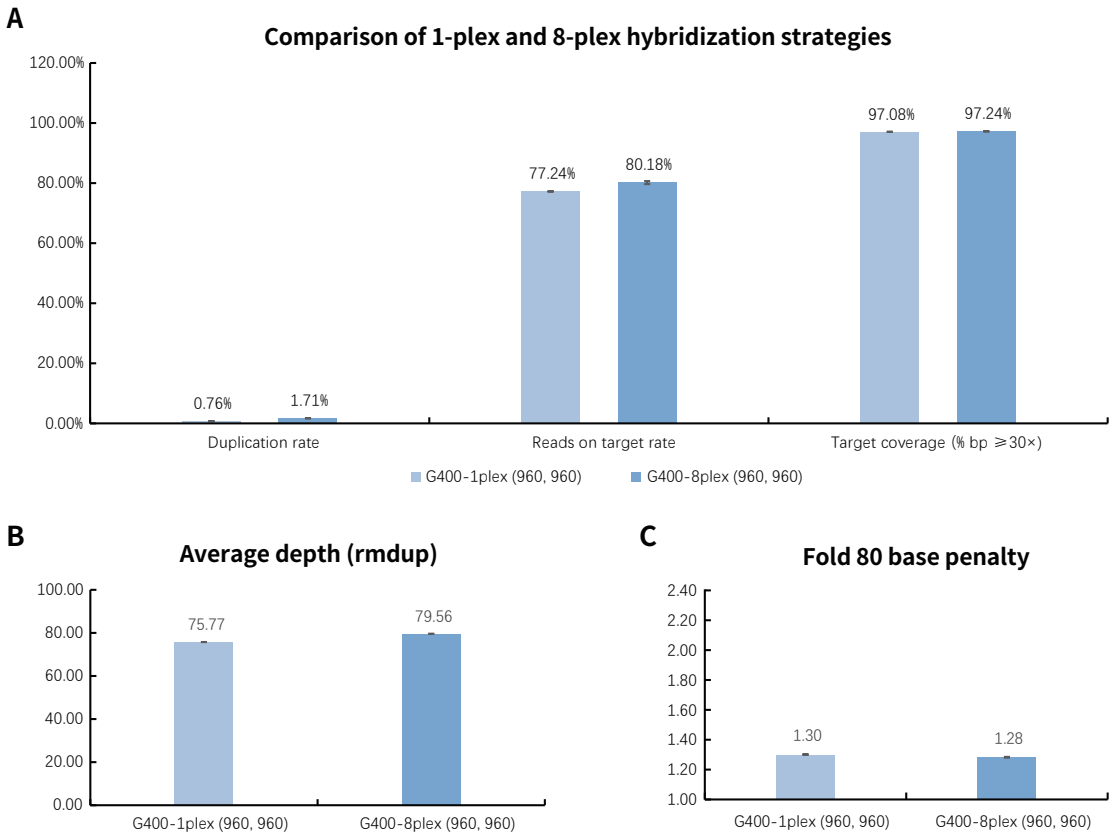

Figure 3. The comparison of hybridization strategies between 1-plex and 8-plex based on automated solution. G400-1plex (960, 960) and G400-8plex (960, 960) were intercepted to 6Gb for analysis. Considering duplication rate, G400-8plex (960, 960) is slightly worse than G400-1plex (960, 960)(A). However, in terms of reads on target rate, target coverage (% bp ≥30×), average depth (rmdup) and fold 80 base penalty, the 8-plex strategy is slightly better than 1-plex (A,B,C).

The sequencing quality of WES library generated from MGI's automated system is high

To explore if there is any quality difference in the sequencing data obtained by automated and manual solutions, sequencing data from G400-8plex (960, Manual), G400-8plex (960, 960), G99-1plex (960,Manual) and G99-1plex (960,960) were intercepted to 6Gb and compared. It is found that duplication rate of these four groups are below 2%, reads clean rate and reads mapping rate are higher than 99%, and Q30 are around 95% (Figure 4A). Further analysis shows that the average depth (rmdup) are higher than 65×, reads on target rate are higher than 70%, and fold 80 base penalty are lower than 1.4, while the metrics of automated solutions are slightly better than manual groups(Figure 4B,C,E). Similarly, Target coverage (% bp ≥30×) are higher than 96%, which is sufficient to support high-confidence variant calling (Figure 4D). In addition, the above results are comparable to the public data on the Twist website.

Overall, these results demonstrate that the Twist WES solution perfectly match with MGI's automated system and DNBSEQ sequencing platform when conducting WES research. Moreover, the sequencing data is of high quality and comparable to manual solution.

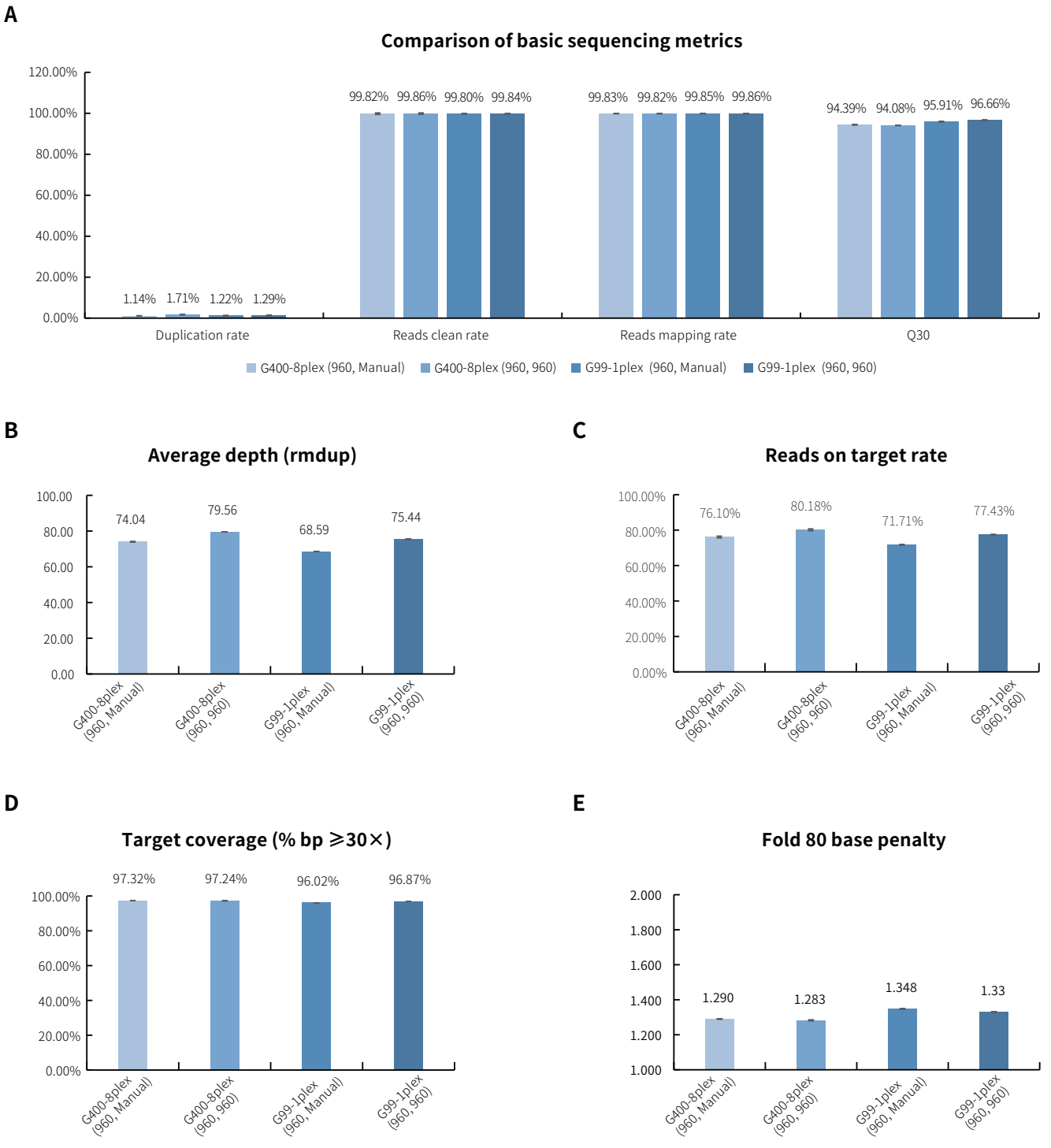

Figure 4. The comparison of key sequencing metrics of libraries obtained by manual and automatic solutions. (A) The primary sequencing metrics show that there is almost no difference between manual and automated library preparation. The duplication rate are < 2%, reads clean rate and reads mapping rate are >99%, and the Q30 are ~ 95%. (B, C, E) When concerned about average depth (rmdup), reads on target rate and fold 80 base penalty, the automatic solution is slightly better than manual. (D) For Target coverage (% bp ≥30×), automation is comparable to manual solution.

Conclusion

The performance evaluation of Twist Exome 2.0 on DNBSEQ platform show that the Twist WES solution is perfectly compatible with MGI's automated system and DNBSEQ platform, which can greatly save labor cost while ensuring high-quality data output. Meanwhile, it is recommended to adopt 8-plex hybridization strategy and intercept 6Gb data for bioanalysis when conducting WES research based on this solution.

DNBSEQ-G400, with a new flow cell system, can flexibly support a variety of different sequencing modes. It adopts optimized optical and biochemical systems, making PE150 (FCL) sequencing at full capacity at 56 hours become possible. DNBSEQ-G99 is one of the fastest genetic sequencers among mid-low throughput in the world. Its effective reads is 80M per flow cell and PE150 sequencing takes only 12 hours, maximizing sequencing efficiency. The built-in computing module integrates sequencing and bioinformatics, resulting in efficient and high-quality data output.

This study demonstrated that the DNBSEQ-G400 and DNBSEQ-G99 combined with the Twist solution can fully facilitate WES, supporting sequencing applications and data analysis in scientific research, clinical medicine, justice, agriculture and other fields.

MGISP-960 Automated
Sample Preparation System

DNBSEQ-G400 Genetic Sequencer

References

1. Rabbani, B., Tekin, M. & Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *J Hum Genet* **59**, 5-15, doi:10.1038/jhg.2013.114 (2014).
2. Wang, Q., Shashikant, C. S., Jensen, M., Altman, N. S. & Girirajan, S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep* **7**, 885, doi:10.1038/s41598-017-01005-x (2017).
3. Clark, M. M. et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med* **3**, 16, doi:10.1038/s41525-018-0053-8 (2018).
4. Wright, C. F., FitzPatrick, D. R. & Firth, H. V. Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet* **19**, 325, doi:10.1038/nrg.2018.12 (2018).
5. Lalonde, E. et al. Genomic Diagnosis for Pediatric Disorders: Revolution and Evolution. *Front Pediatr* **8**, 373, doi:10.3389/fped.2020.00373 (2020).
6. Yuan, J. et al. Novel technologies and emerging biomarkers for personalized cancer immunotherapy. *J Immunother Cancer* **4**, 3, doi:10.1186/s40425-016-0107-3 (2016).
7. Samstein, R. M. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* **51**, 202-206, doi:10.1038/s41588-018-0312-8 (2019).
8. 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, doi:10.1016/j.slast.2021.10.016 (2022).
9. Porreca, G. J. Genome sequencing on nanoballs. *Nat Biotechnol* **28**, 43-44, doi:10.1038/nbt0110-43 (2010).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-G400RS	900-000170-00
	Genetic Sequencer DNBSEQ-G99ARS	900-000609-00
	MGISP-100RS Automated Sample Preparation System	900-000206-00
	MGISP-960RS Automated Sample Preparation System	900-000147-00
Software	MegaBOLT Bioinformatics analysis accelerator	900-000555-00
Library Prep	Twist Library Preparation EF Kit 2.0 (96 RXN)	104207*
	Twist Universal Adapter System- TruSeq Compatible, 96 Samples Plate C	101310*
	Twist Exome 2.0, 12 Reactions, Kit	104134*
	Twist Standard Hyb and Wash Kit v2, 12 Reactions	104446*
	MGIEasy Universal Library Conversion kit (App-A) (16RXN)	1000004155
Sequencing Reagents	DNBSEQ-G400RS High-throughput Sequencing Set (App-A FCL PE150)	1000016995
	High-throughput Sequencing Set (G99 SM App-C FCL PE150)	940-000413-00

*The relevant products are available and can be ordered on the Twist official website

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Author: Hanfei Zhang, Qiwei Wang

Editor-in-Charge: Qiwei Wang

Reviewer: Yao Jiang

1. For StandardMPS and CoolMPS: Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, Spain, UK, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland, Portugal, Austria and Romania. Unless otherwise informed, StandardMPS sequencing reagents, and sequencers for use with such reagents are not available in Hong Kong. No purchase orders for StandardMPS products will be accepted in the USA until after January 1, 2023.
2. For HotMPS sequencers: This sequencer is only available in selected countries, and its software has been specially configured to be used in conjunction with MGI's HotMPS sequencing reagents exclusively.
3. For HotMPS reagents: This sequencing reagent is only available in selected countries.