Evaluation of the Performance of Twist Exome 2.0 kit on DNBSEQ-G400**

Introduction

Next-generation DNA sequencing has been a driving force to reduce sequencing cost and it empowers scientists to identify genetic variations associated with human diseases at higher resolution and greater sensitivity than before. For research groups or facilities providing genetic services, it is essential to balance the needs to generate high-quality data and reduce cost. Three approaches are commonly employed: targeted panels, exome sequencing and whole genome sequencing (WGS). Targeted panels are limited in scope and require frequent revision and validation to maintain their standard. WGS allows the sequencing laboratory to assay the majority of coding and non-coding regions of the genes and determine the structural variants, however the cost of WGS remains a significant barrier to routine testing for the patients. Exome is the coding region in the human genome constituting 1-2% of the entire genome. Approximately 85% of disease-causing variants are found in the exome, and hence exome sequencing is an efficient approach to determine the genetic basis of a disease.

This application note demonstrates the compatibility of Twist Library preparation and Targeted Enrichment products with MGI DNBSEQ™ sequencing by testing Twist Exome 2.0* which has been designed to include additional clinically relevant non-coding pathogenic and likely pathogenic variants.

Methods

DNA libraries were prepared from HapMap Reference gDNA NA12878 using Twist Library Preparation EF 2.0 and Twist Universal Adapter System. To assess the performance of Single-plex and Eight-plex libraries, duplicates of each library type were prepared. Eight-plex libraries were prepared by pooling eight pre-capture libraries into a single pool. Both single-plex library or eight-plex libraries were prepared and captured using Twist Target Enrichment Standard Hybridization v2 Protocol and Twist Exome 2.0 capture probes. Subsequently, MGIEasy Universal Library Conversion kit (App-A) was used to add phosphorylated end to the purified captured libraries by additional 4 PCR cycles, followed by the circularisation and digestion steps to remove nonspecific products and generate single stranded circular libraries. Final libraries were sequenced on the DNBSEQ-G400** (Figure 1) system using DNBSEQ-G400 High-Throughput Sequencing Kit (PE100)** and High-Throughput Sequencing Primer Kit (App-C) (refer to the Ordering information for more details on reagents used for library preparation and target enrichment).

Figure 1. DNBSEQ-G400** sequencer

Data was downsampled to 4Gb, 5Gb, 6Gb and 8Gb. Data QC was done using MegaBOLT, a field-programmable gate array (FPGA)-boosted data analysis platform which is capable of fast, accurate and cost-effective analysis of whole genome and whole exome sequencing data. The workflow is described in Figure 2.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

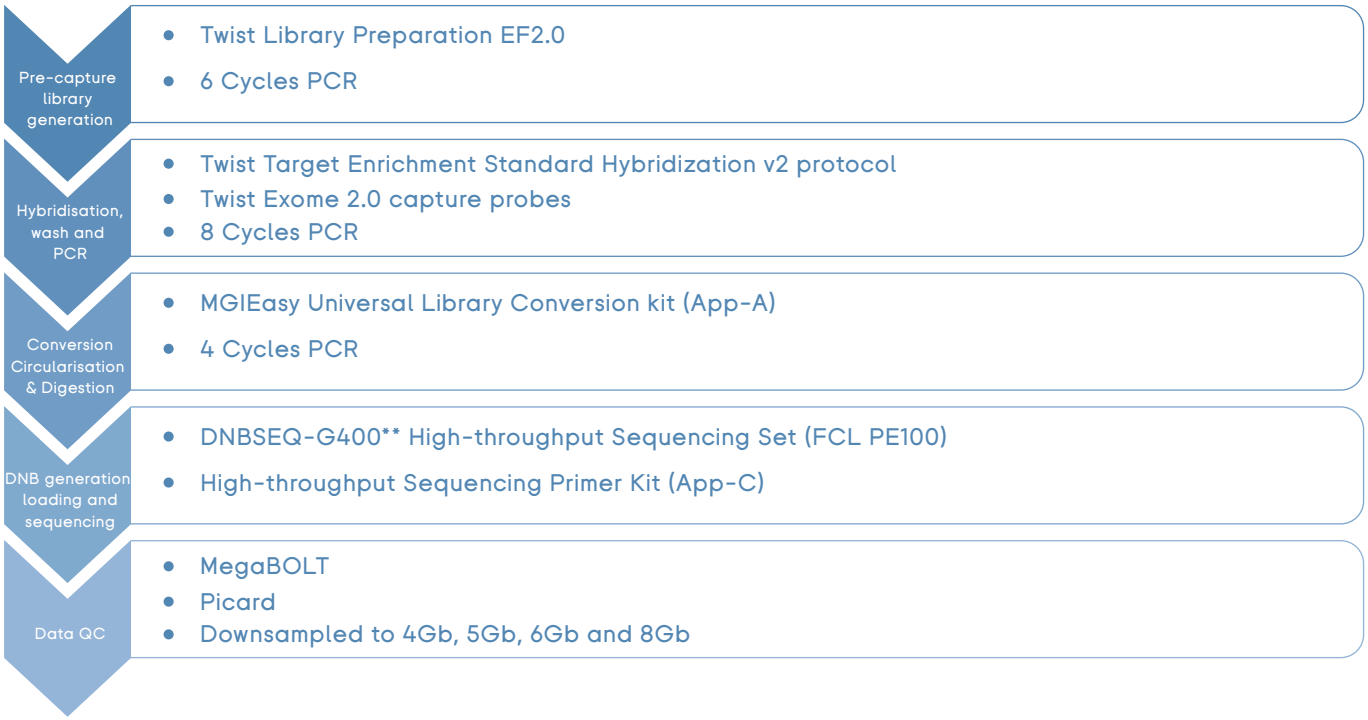


Figure 2. Twist Exome 2.0 Library Preparation and DNBSEQ-G400** sequencing workflow

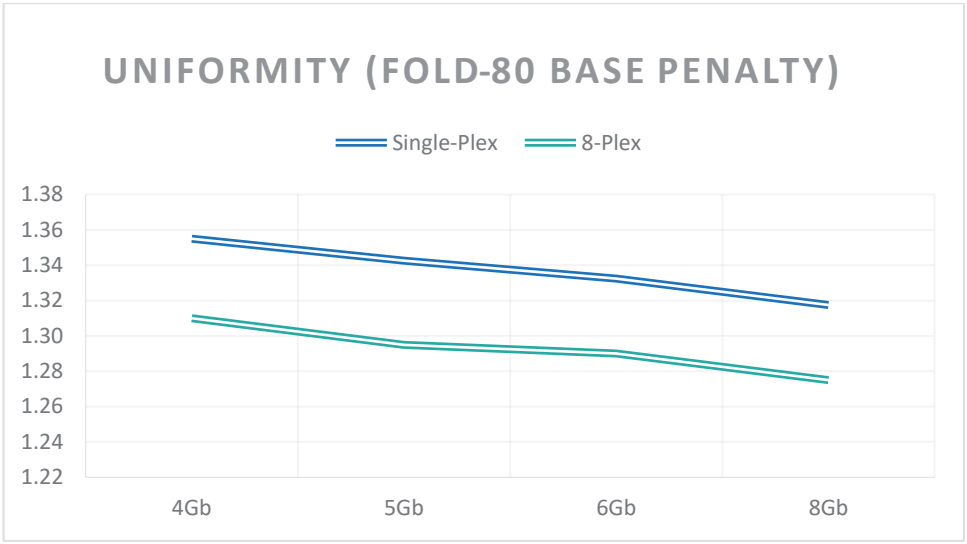
Results

Uniformity and On-target Rate

Targeted or exome capture experiment should ideally generate reads that are distributed uniformly across the target regions. However, off-target capture and non-uniform distribution of reads are sometimes present in real-life experiments. Increasing sequencing depth is sometimes used to ensure the sufficient coverage of most targeted regions. However, this strategy often leads to over-sequencing of the other regions which are otherwise adequately covered.

The performance of the target capture uniformity was assessed using the fold-80 base penalty generated using the Picard pipeline. Fold-80 penalty provides information on the additional sequencing required to ensure that 80% of the target bases achieve the required coverage. It was demonstrated that the fold-80 metric for Twist Exome 2.0 libraries sequenced on DNBSEQ-G400** for standard recommended data yield of 6 Gb was around 1.29 with minimal fold-80 values difference between single-plex and 8-plex libraries.

Both single-plex and 8-plex libraries showed comparable performance with greater than 90% of bases covered over 30× at 4Gb sequencing depth and more than 95% of bases covered over 30× at 6Gb sequencing depth. The result obtained was very similar to the metric provided on Twist Exome 2.0 datasheet [1].



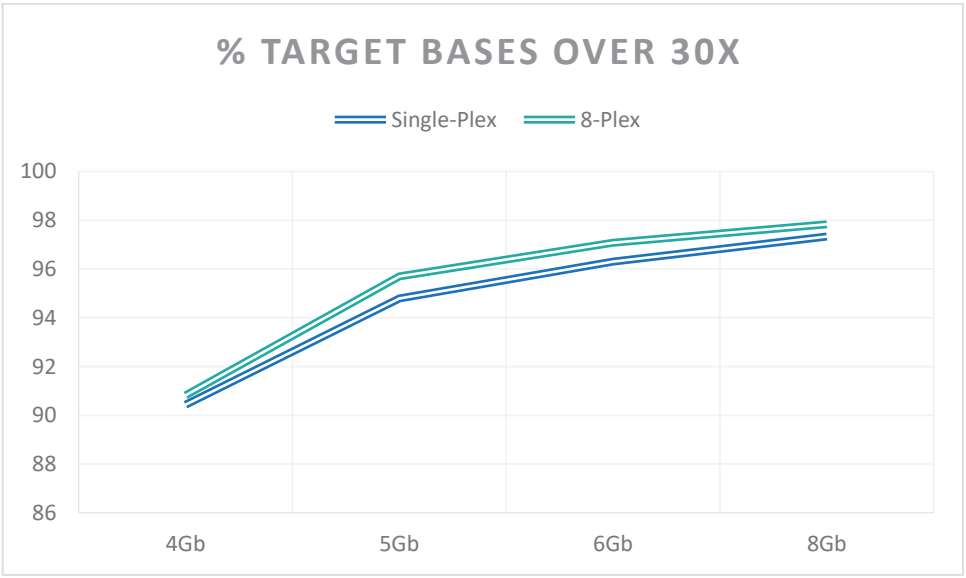


Figure 3. (a) Uniformity of Twist Exome 2.0 across different sequencing depths (reported as fold-80 base penalty score) for single-plex and 8-plex libraries. (b) Target bases over 30× for down sampled data 4Gb, 5Gb, 6Gb, 8Gb.

On-target Rate

The on-target rate was calculated by deducting the off-target rate from 100% effective target (1-PCT_OFF_BAIT), using the bed files provided on the Twist website "https://www.twistbioscience.com/resources/data-files/twist-exome-20-bed-files ". Some off-target sequencings can occur due to lower specificity of the probes or due to probes sharing sequence similarity with non-coding sequences resulting in non-specific hybridisation.

On-target rate from Twist Exome 2.0 libraries sequenced on DNBSEQ-G400** was observed to be on par with the Twist Exome 2.0 data published on Twist official website [1].

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

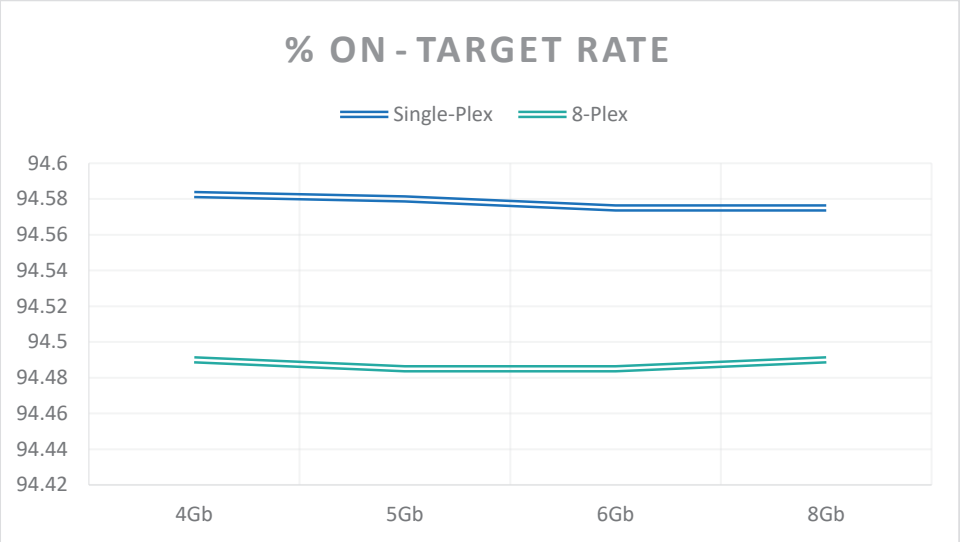


Figure 4. On-target rate of single-plex and 8-plex Twist Exome 2.0 libraries sequenced on DNBSEQ-G400 sequencer**. On-target rate calculated as 1- "PCT_OFF_BAIT.

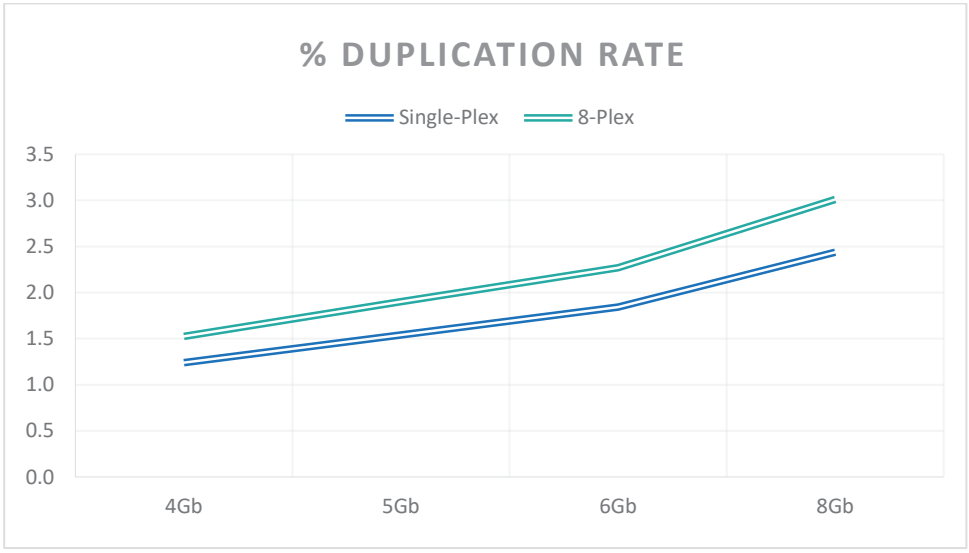
Duplication Rate and Library Complexity

There are several types of duplications that are normally observed in Next Generation Sequencing experiments, including PCR duplicates, Optical duplicates and ExAmp duplicates. PCR duplicates are generated during the library preparation step from PCR amplification of the original DNA molecules. Optical duplicates are normally observed in the instrument utilising cluster generation approach and non-patterned flow cells due to large clusters being interpreted as two separate clusters. ExAmp duplicates occur in the patterned flow cells utilising cluster generation method where re-clustering of the original library molecules after released from polymerase copying results in the seeding of neighbouring nanowells with identical fragments during amplification on the flow cell.

DNBSEQ™ technology, based on DNA nanoballs, works by generating multiple copies of the same circular template using rolling circle amplification. DNA nanoballs are generated in a tube before loading into patterned flow cells, which prevents issues with optical duplicates or ExAmp duplicates observed in other sequencing platform utilising cluster generation approach.

On the other hand, Twist has developed a proprietary method that allows the use of double-stranded DNA probes without the risk of the double strands binding with each other and compromising target capture efficiency. Capturing both strands of the target allows an increased number of unique molecules to be sequenced, and results in higher complexity libraries and lower duplicates.

Despite the additional four (4) cycles applied during the library conversion step, single-plex and 8-plex Twist Exome 2.0 libraries sequenced on DNBSEQ-G400** generated very low duplication rate and high estimated library complexity, with around 1.8% and 2.3% duplication rate for Single-plex and 8-plex libraries respectively at 6Gb sequencing depth (Figure 5).



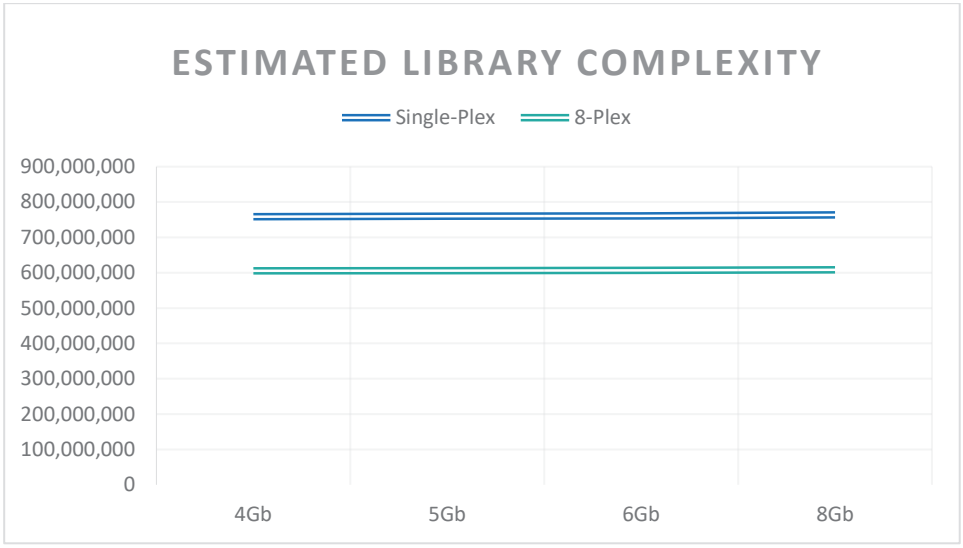


Figure 5. (a) Estimated duplication rate from single-plex and 8-plex Twist Exome 2.0 libraries across different sequencing depths. (b) Estimated library complexity for Single-plex and 8-plex Twist Exome 2.0 libraries across different sequencing depths.

Conclusion

In conclusion, we demonstrated that the Twist Exome Capture Library Preparation workflow pairs seamlessly with the DNBSEQ™ sequencing platform. The combination of DNBSEQ-G400** and Twist Exome chemistry results in an extremely low duplication rate and has been proven to be an effective strategy to enable effective utilisation of sequencing reads, and generation of quality data with less sequencing.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Ordering Information

Procedure	Vendor	Kit	Catalog Number
Library Preparation	Twist	Twist Library Preparation EF kit v2.0*	104207 (96 samples)
Target Enrichment	Twist	Twist Exome 2.0*	104140 (12 rxns)
Conversion, Circularisation and Digestion	MGI	MGIEasy Universal Library Conversion Kit (App-A)	1000004155 (16 rxns)
DNA Nanoball Generation and Sequencing	MGI	Genetic Sequencer DNBSEQ-G400RS**	900-000170-00
		DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE100)**	1000016950
		High-Throughput Sequencing Primer Kit (App-C)	1000027472

Reference

1. Twist. Twist Exome 2.0 data sheet. Retrieved 10 March, 2022. From Exome 2.0 | Twist Bioscience
2. Twist. Twist Exome 2.0 bed files. Retrieved 23 February 2022. From Twist Exome 2.0 .bed files | Twist Bioscience

MGI Tech Co.,Ltd | Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083
en.mgi-tech.com | MGI-service@mgi-tech.com

The copyright of this brochure is solely owned by MGI Tech Co. Ltd.. The information included in this brochure or part of, including but not limited to interior design, cover design and icons, is strictly forbidden to be reproduced or transmitted in any form, by any means (e.g. electronic, photocopying, recording, translating or otherwise) without the prior written permission by MGI Tech Co., Ltd.. All the trademarks or icons in the brochure are the intellectual property of MGI Tech Co., Ltd. and their respective producers.

* FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES. NOT FOR USE IN USA

** Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, USA, Spain, UK, Hong Kong, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland and Portugal.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.