10× Genomics Solution based on DNBSEQ Sequencing Platform Facilitates Single Cell Immune Profiling

Performance Evaluation for 10× Genomics Chromium Next GEM Single Cell 5' Kit v2 on DNBSEQ-G400 and Brand A Genetic Sequencer

10x Genomics developed the Chromium Next GEM Single Cell 5' Kit v2 that enables simultaneous profiling of the V(D)J repertoire, cell surface protein, and gene expression (GEX) data. In this study, the technical staff prepared the 5' gene expression, antibody capture and V(D)J libraries ,then sequenced on both MGI DNBSEQ-G400 and Brand A genetic sequencer. A comprehensive performance comparison indicates that DNBSEQ-G400 is perfectly compatible with this kit and its performance is comparable to Brand A.

Recommended application: Single-cell genomics

Recommended models: DNBSEQ-G400RS, DNBSEQ-G99ARS, DNBSEQ-T7RS

- ### A best-in-class solution for single cell immune profiling

  The relevant kit developed by 10×genomics exhibited high resolution profiling of adaptive immune repertoire and cellular context of individual cells

- ### Perfect compatibility with DNBSEQ sequencing platform

  MGIEasy Universal Library Conversion kit enables Chromium Next GEM Single Cell 5' Kit from 10×genomics 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

Background

V(D)J recombination is a critical process that takes place during lymphocyte maturation[1]. It involves the rearrangement of variable (V), diversity (D), and joining (J) gene segments in a nearly random manner which creates an enormous degree of diversity in both immunoglobulins (Ig) and T cell receptors (TCR)[2]. This process holds significant importance in the study of the adaptive immune system, including the clinical diagnosis, drug development, and health management[2].

10× Genomics is a well-recognized company that primarily focuses on single-cell sequencing technologies. Its RNA profiling kits have found widespread applications in clinical diagnosis, drug development, and health management. The company's V(D)J libraries preparation kit has also emerged as a popular solution for many researchers studying immune profiling.

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[3]. With years of budget and intelligence investment, MGI has developed and launched a series of DNBSEQ based genetic sequencers, like DNBSEQ-G400RS, DNBSEQ-T7RS and DNBSEQ-G99ARS, meeting multiple MPS demands in life science research and clinical application. However, there are limited reports on the sequencing data of 10× Genomics libraries running on the MGI DNBSEQ sequencing platform.

In this study, DNBSEQ-G400 genetic sequencer from MGI was utilized to sequence libraries generated from the Chromium Next GEM Single Cell 5' Kit v2, a 10× Genomics product for combined library preparation for RNA transcriptomic, V(D)J, and antibody capture (feature barcode) (Figure 1). The purpose was to evaluate the sequencing quality on DNBSEQ platform and the result of data analysing. Additionally, we obtained similar results when sequencing data from another company, Brand A, was incorporated for comparison.

Materials and Methods

Sample preparation

In this study, two types of human peripheral blood mononuclear cells (PBMCs) from clinical patients, named D01 and D02, were utilized to evaluate the compatibility of Chromium Next GEM Single Cell 5' Kit v2 with DNBSEQ sequencing platform.

Library preparation and sequencing

Single cell suspensions were loaded to Chomium Chip with the right concentration produced with the GEM beads in each single cell droplet. 5' gene expression (GEX), feature barcode (FB, including cell surface protein, antibody/antigen, or CRISPR, this study only including antibody) and V(D)J libraries were produced within the same droplet.

For the preparation of the libraries, the Chromium Next GEM Single Cell 5' Reagent Kits v2 (Dual Index) were utilized, with each sample consisting of three distinct libraries: Gene expression,V(D)J, and antibody capture. Detail protocol can be found in the user manual (CG000330_ChromiumNextGEMSingleCell5_v2_CellSurfaceProtein_UserGuide_RevF, P16~21).

Then, the MGI Universal Library Conversion Kit (App-A) was used to transform the double-stranded DNA (dsDNA) libraries into single-stranded circular DNA (ssCirDNA), where 50 ng of dsDNA was employed as input for 5 rounds of PCR cycles during the AC-PCR conversion phase. While the gene expression and V(D)J libraries were processed using the standard protocol, the input of AC-PCR product for the antibody capture library needed to be increased to around 3 pmol for circularization. Circularized single stranded DNA libraries were used to generate DNA nanoballs (DNBs) prior to sequencing. The gene expression, V(D)J, and antibody capture DNBs were combined in a ratio of 4:1:1, as recommended in the Chromium Next GEM Single Cell 5' Reagent Kits v2 (Dual Index) user manual. As the antibody capture library had a short insert size, only index i7 of the pooled library could be sequenced, despite having a dual index. Due to the short insert size of the antibody capture library, the beads ratio during the purification stage after circularization had to be increased to 4 times.

For sequencing, it is recommended to use DNBSEQ-G400 with ECR 6.0 and the DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150) and High-Throughput Pair End Sequencing Primer Kit (App-A) with a sequencing strategy of PE150+10. If the ECR version is lower than 5.2, the sequencing quality could be improved by spiking in 20% - 40% MGI standard libraries or other balanced libraries, using the DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150) and High-Throughput Sequencing Primer Kit (App-D).

At the same time, a specific genetic sequencer from Brand A was also parallelly utilized to sequence the above mentioned libraries to comprehensively evaluate the performance of DNBSEQ-G400.

Bioinformatics analysis

FASTQ data were uploaded and analyzed with the 10× Genomics Cellranger Multi Pipeline. In the 10× Genomics library, both the barcode and the unique molecular identifier (UMI) are positioned within the initial 26 base pairs (bp) of read 1. Therefore, in order to improve the analysis efficiency, any sequence beyond 26 cycles in

read 2 was excluded and not considered during the analysis.

The Cellranger Multi Pipeline adopts a CSV configuration containing the path to the FASTQ file from cellranger mkfastq, which can be used to analyze any combination libraries of 5' gene expression (GEX), features barcode (FB) and V(D)J from a single GEM well. It can perform alignment, filtering, barcode counting, and UMI counting on the gene expression and/or feature barcode libraries, and also performs sequence assembly and paired clonotype calling on the V(D)J libraries. It can selectively select cells that can be detected in both V(D)J library and 5' GEX library data to obtain more accurate results. After the FASTQ file is input, a count matrix will be generated. This count matrix can be used as an input for many downstream analysis tools such as Seuratpackages. For more information, please refer to the 10× Genomics manual page: https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/using/multi.

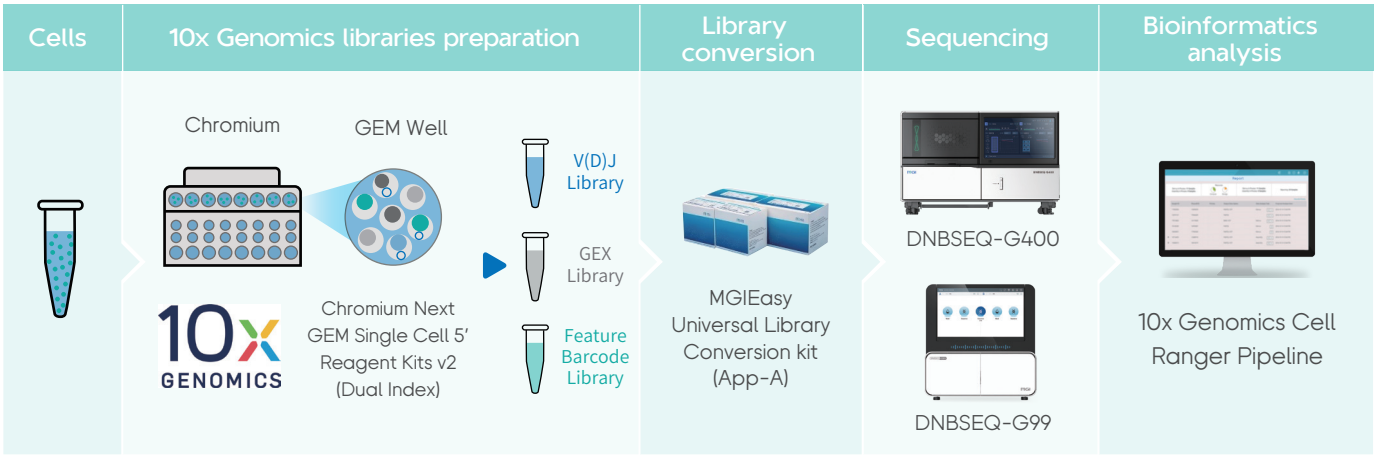

Figure 1. 10× Genomics library preparation and MGI sequencing workflow.

Results

The DNBSEQ-G400 generated high-quality data from 10× Genomics libraries

Following sequencing on DNBSEQ-G400 and Brand A, 30-270 million paired-end (PE) reads were obtained, and trimmed to the same amount for downstream analysis. The GEX library had a higher number of reads compared to the other libraries, and this trend was consistent across the two test samples. The estimated number of cells from DNBSEQ-G400 is slightly higher than Brand A. Notably, all libraries had an extremely high proportion of valid UMIs and barcodes (Table 1), which ensures a high level of confidence in the quantification during subsequent steps.

In conclusion, the estimated number of cells, valid UMI rate and valid barcode rate shows the parameters for valid counting via different parts of the library are highly comparable between sequencing data from DNBSEQ-G400 and Brand A sequencing platform.

Cell ID	Library types	Number of reads	Genetic Sequencer	Estimated number of cells	Valid UMIs	Valid barcodes
D01	Gene Expression（GEX）	274,429,144	Brand A	8,307	99.78%	93.03%
			DNBSEQ-G400	8,344	99.65%	92.98%
	Antibody Capture	32,852,214	Brand A	8,307	100.00%	98.24%
			DNBSEQ-G400	8,344	100.00%	98.13%
	V(D)J	53,617,427	Brand A	6,808	--	96.34%
			DNBSEQ-G400	6,816	--	96.08%
D02	Gene Expression（GEX）	255,343,816	Brand A	9,792	99.64%	90.81%
			DNBSEQ-G400	9,852	99.57%	90.81%
	Antibody Capture	34,404,200	Brand A	9,792	99.99%	98.19%
			DNBSEQ-G400	9,852	100.00%	98.08%
	V(D)J	48,448,059	Brand A	7,672	--	94.15%
			DNBSEQ-G400	7,672	--	94.16%

Table 1. The comparison of primary QC (quality control) metrics obtained from DNBSEQ-G400 and Brand A.

To further evaluate the sequencing quality, the Q30 score for RNA read, UMI and barcodes was calculated for different library types. The results show that sequencing data generated from both DNBSEQ-G400 and Brand A have a Q30 score of over 85%, which exceeds the generally accepted criteria. Importantly, the Q30 score for UMI and barcodes were around 95%, which is high enough to ensure accurate detection and counting of molecular expressions (Figure 2A and B). Both sequencing platforms exhibited similar saturation for transcriptomes (Figure 2C), with an endpoint of ~80%, which meets the criteria for single cell studies. A similar trend is also observed for gene saturation (Figure 2D). The Sequencing Saturation and the Median Gene counts are more simliar between the 2 platform results compared to the differences in the 2 cell samples-D01 vs D02.

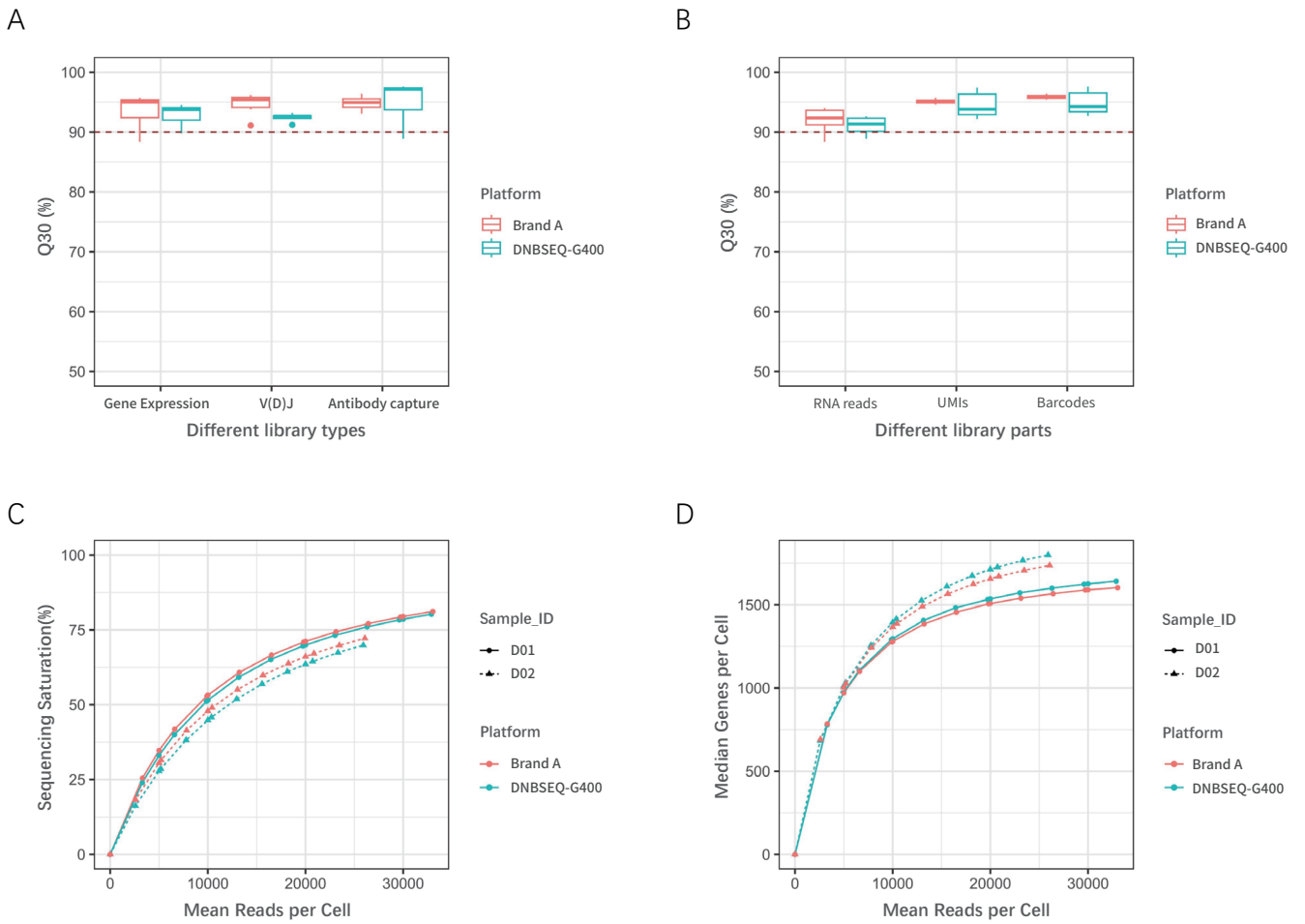

Figure 2. The comparison of primary QC performance for DNBSEQ-G400 and Brand A. A) Q30 value for different library types. Red dashed line is for 90%. B) Q30 value for different sequencing parts. Red dashed line is for 90%. C) Saturation for 2 samples. Sequencing Saturation is a measure of the observed library complexity and approaches 1.0 (100%) when all converted mRNA transcripts (or ligation products in the context of fixed RNA profiling libraries) have been sequenced. D) Median genes per cell on different platforms. This plot shows the median genes per cell as a function of down sampled sequencing depth in mean reads per cell, up to the observed sequencing depth.

DNBSEQ-G400 is comparable to Brand A in GEX and antibody libraries sequencing

To assess feature selection, the distribution of features vs RNA counts was evaluated, and highly similar results were observed between the two sequencing platforms, indicating a significant level of consistency and stability between the same library sequenced by both platforms (Figure 3A and Figure 3B). Furthermore, UMAP analysis demonstrated that the Brand A and DNBSEQ-G400 sequencing data produced near identical cluster results for both transcriptome and antibody capture expression (Figure 3C and Figure 3D). These findings suggest that the two platforms are compatible and can be confidently applied in single cell studies.

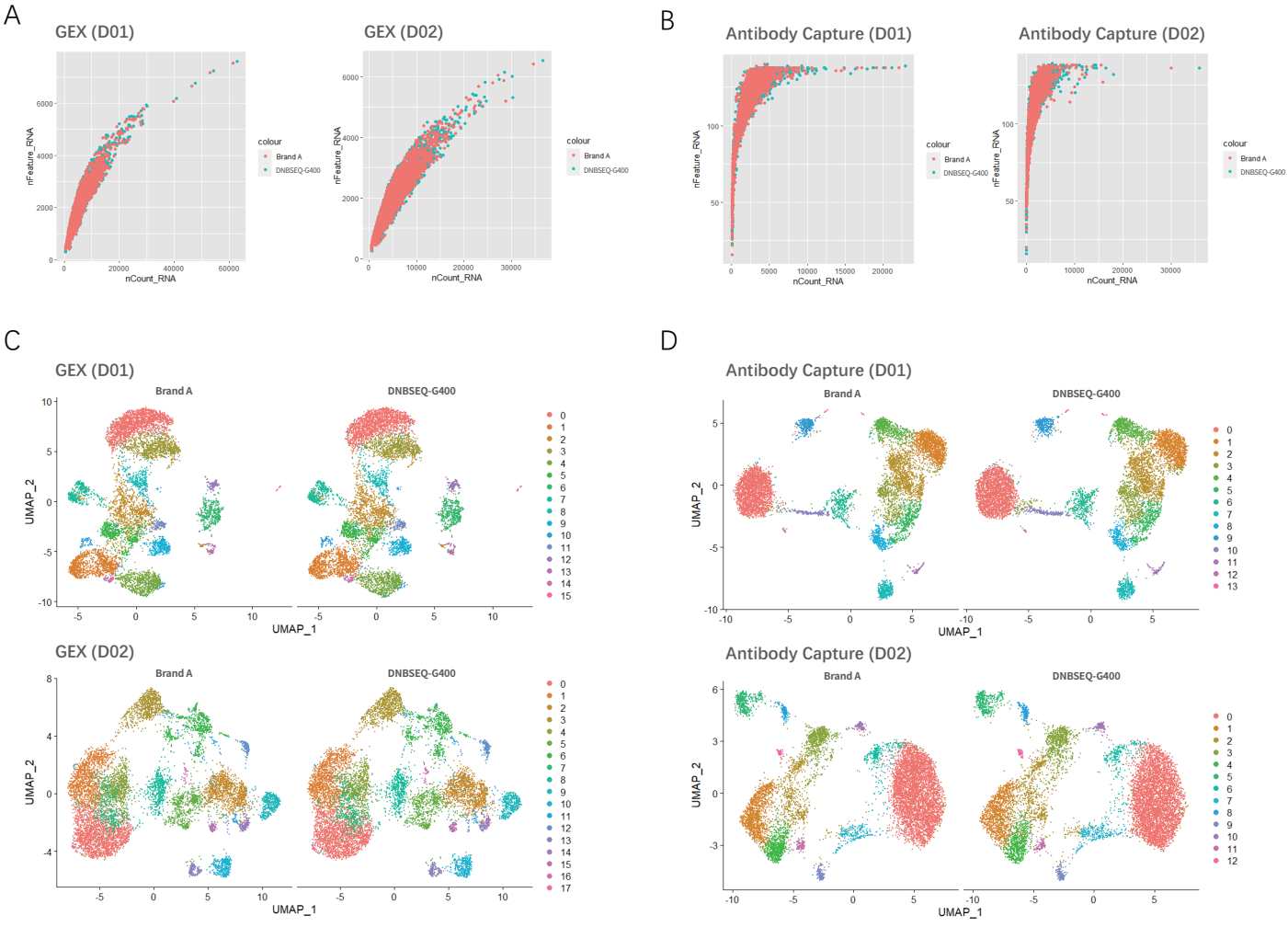

Figure 3. The comparison of secondary metrics for GEX and antibody libraries between DNBSEQ-G400 and Brand A. A) the distribution of feature vs RNA counts for sample D01 and D02, for the transcriptome library (GEX). B) the distribution of feature vs RNA counts for sample D01 and D02, for the antibody capture library. C) UMAP analysis for data from DNBSEQ-G400 and Brand A platforms, for the transcriptome library (GEX). D) UMAP analysis for data from DNBSEQ-G400 and Brand A platforms, for the antibody capture library.

DNBSEQ-G400 showed highly consistent clonotypes detecting ability with Brand A in V(D)J libraries

To better understand T cell and B cell detection, clonotypes proportion were calculated from two samples. Before the analysis, clonotypes supported by less than 3 reads were eliminated to improve efficiency. First, a correlation analysis was conducted for the two platforms. The results show that the clonotypes proportion about D01 of DNBSEQ-G400 are good accordance with those of Brand A with R^2 of 0.9901, while D02 is slightly lower but still within a good range (R^2 = 0.9105). Since R^2 is close to 1, the clonotype detection performance of MGI is comparable to Brand A (Figure 4A, B). Furthermore, the top 10 clonotypes were assessed, and the results show that D01 has slightly more consistent results than D02 (Figure 4C,D), which may due to different sequencing strategies. When sequencing D01 on DNBSEQ-G400, the GEX, antibody capture and V(D)J libraries were sequencing together in the same flow cell. While for D02 they were sequenced individually. These together suggest us that, to achieve high quality and accuracy sequencing result, the three libraries preferred to be sequenced in the same batch. Overall, these results suggested that DNBSEQ-G400 could confidently be used for single-cell T/B cell studies and comparable to Brand A.

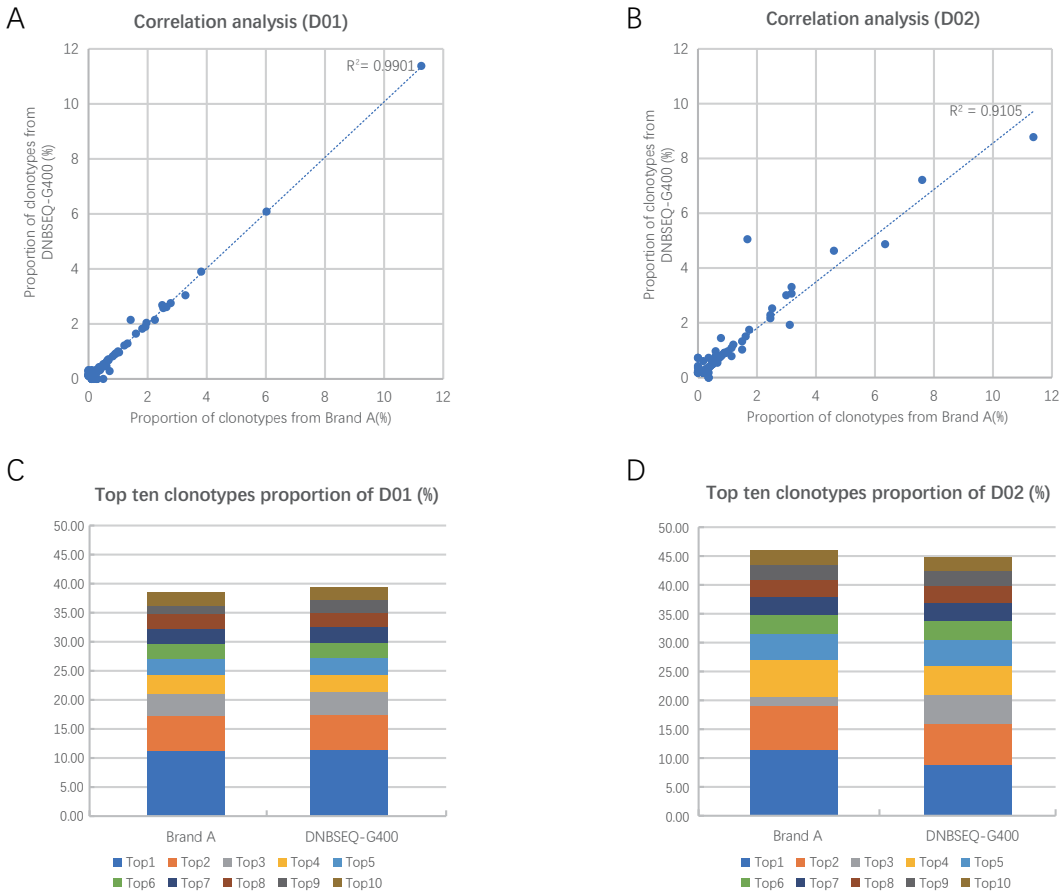

Figure 4. Comparison of V(D)J libraries for 2 samples from 2 platforms, clonotypes supported by fewer than 3 reads were eliminated prior to analysis. A,B) Proportion correlation analysis for clonotypes detected by 2 platforms. Proportion: the observed fraction of cell barcodes with this clonotype. C,D) Top ten clonotype distribution. The clonotype distribution was plotted with the top ten clonotypes from different sequencing platforms on the x-axis and the proportion of each clonotype in the whole library on the y-axis.

Summary

The performance evaluation results of Chromium Next GEM Single Cell 5' Kit v2 on the DNBSEQ-G400 and Brand A platforms indicates that DNBSEQ-G400 is perfectly compatible with 10× Genomics single cell solution.Whether gene expression library, V(D)J or antibody capture libraries, high-quality and reliable results can be obtained and DNBSEQ-G400's performance is comparable to Brand A.

DNBSEQ-G400 sequencer, with a new flow cell system, can flexibly support a variety of different sequencing applications. It adopts optimized optical and biochemical systems, making PE150 (FCL) sequencing at full capacity only at 56 hours.

This study demonstrated that DNBSEQ-G400 combined with the 10× Genomics solution can fully facilitate immune profiling such as cell surface anetigen and immune receptors profiling and TCR and BCR-related research in diverse domains. Oncology, autoimmune disease, infectious disease, transplantation, and other related fields will greatly benefit from this cutting-edge single cell tools, which provides high performance data and cost-effectiveness as these tools are made available.

DNBSEQ-G400 Genetic Sequencer

References

1. Bassing, C. H., Swat, W. & Alt, F. W. The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109**, S45-S55, doi:Doi 10.1016/S0092-8674(02)00675-X (2002).
2. Lieber, M. R. Transposons to V(D)J Recombination: Evolution of the RAG Reaction. *Trends Immunol* **40**, 668-670, doi:10.1016/j.it.2019.06.007 (2019).
3. Porreca, G. J. Genome sequencing on nanoballs. *Nature Biotechnology* **28**, 43-44, doi:10.1038/ nbt 0110-43 (2010).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-G400RS	900-000170-00
Software	cellranger multi	https://support.10xgenomics.com/single-cell-vdj/software/downloads/latest*
Library Prep	Chromium Next GEM Single Cell 5' Kit v2(16 rxn)	1000263*
	MGI Universal Library Conversion Kit (App-A)(16 rxn)	1000004155
	Standard library reagent V3	1000005033[a]
Sequencing Reagents	DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150)	1000016952[b]
	High-Throughput Pair End Sequencing Primer Kit (App-A)	1000020832[c]
	High-Throughput Sequencing Primer Kit (App-D)	1000028550[d]

*The relevant products are available and can be ordered on the 10× Genomics official website

**Either b+c or b+d+a could be chosen

MGI Tech Co.,Ltd

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

Version: September 2023

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Author: Wei Tian, Ningxin Gu, Weiyuan Yu, Hanfei Zhang

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.