Takara Bio solution on DNBSEQ sequencing platform facilitates human immune profiling

Performance Evaluation for SMARTer Human TCR a/b Profiling Kit v2 and SMART-Seq Human BCR (with UMIs) on DNBSEQ-G400 Genetic Sequencer

Takara has developed a variety of immune profiling kits based on its proprietary SMART technology. This study has chosen two representative kits: SMARTer Human TCR a/b Profiling Kit v2 and SMART-Seq Human BCR (with UMIs), to verify the compatibility of SMART technology on MGI DNBSEQ sequencing platform. The comprehensive analysis indicates that DNBSEQ-G400 is perfectly compatible with these kits and its performance is slightly better than Competitor A.

Recommended application: Immune Profiling

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

- **Best-in-class solution for immune profiling**

  Takara Bio kits provide sensitive and reproducible solutions for understanding the human TCR or BCR repertoire

- **Perfect compatibility with DNBSEQ sequencing platform**

  MGIEasy Universal Library Conversion kit enables Takara Bio TCR/BCR profiling kits to work on the DNBSEQ sequencing platform

- **Efficient and high-quality data output**

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

Background

TCR and BCR sequencing are predominantly employed in disease states where researchers investigate the expansion of specific B or T cell clone(s). This expansion could potentially indicate specificity towards a target linked to the disease. While primarily conducted in oncology studies, this comprehensive immune profiling data has the potential to be leveraged for any disease state[1-3].

Takara Bio's proprietary Switching Mechanism At the 5′ end of RNA Template (SMART®) technology uses a certain reverse transcriptase's template-switching property to integrate an adapter sequence during the first-strand cDNA synthesis process. Based on this feature, Takara Bio has invented a series of seamless, ligation-free library preparation methods and applied them to single-cell mRNA-seq, total RNA-seq, small RNA-seq, immune profiling, and ChIP-seq.

In this immune profiling study, two Takara Bio kits—the SMARTer Human TCR a/b profiling kit and the SMART-Seq Human BCR (with UMI) profiling kit—were chosen to comprehensively evaluate compatibility with MGI's DNBSEQ sequencing platform. Both kits adopt SMART technology and a 5'-RACE-based approach to capture full-length information from V(D)J variable regions of TCRs and BCRs, and are ideal tools for TCR and BCR profiling to gain insights into TCR and BCR repertoire diversity from bulk samples (total RNA or purified cells). Unique molecular identifiers (UMIs) are incorporated to facilitate PCR error correction and clonotype quantification during data analysis.

In this study, the TCR and BCR libraries were sequenced in parallel on either MGI's or Competitor A's platform using the PE300 long reading recipe. Then, the data generated from both platforms were downsampled to the same sequencing depth and compared for the total identified molecular identifier groups (MIGs; reads grouped using UMIs) and clonotype counts. The results are comparable on both sequencing platforms, demonstrating that the Takara Bio libraries a) can be converted with MGI's library conversion kit and b) showed excellent performance and compatibility on the MGI sequencing platform.

Materials and Methods

Sample preparation

Two group of samples were used in this report. The Group 1 samples included 13 cell lines coming from different human cell lines and/or human peripheral blood mononuclear cells (PBMCs). Total RNA was extracted using a Qiagen RNA extraction kit. Group 2 comprised of four total RNA samples extracted from normal human PBMCs (Human Blood, Peripheral Leukocytes Total RNA; Takara Bio, Cat. No. 636592) (Table 1).

Library preparation and sequencing

The Group 1 libraries were prepared with the SMART-Seq Human BCR (with UMIs) kit by Takara Bio technical staff, and the Group 2 libraries were prepared with the SMART-Seq Human BCR (with UMIs) and SMARTer Human TCR a/b Profiling Kit v2 by MGI technical staff (Table 1). Please refer to related instructions for detailed procedures[4,5]. The resulting products of these kits were purified double-stranded DNA libraries (dsDNA). Subsequently, these libraries were converted to single-stranded circular DNA (ssCircDNA) using the MGI Universal Library Conversion Kit (App-A). Afterwards, ssCircDNAs were used to generate DNA Nanoballs (DNBs) and were sequenced on DNBSEQ-G400 FCS in one lane with the PE300+8+8 sequencing strategy.

For sequencing, it is recommended to use DNBSEQ-G400 with ECR 6.0, DNBSEQ-G400RS High-throughput Rapid Sequencing Set (FCS PE300), and High-Throughput Pair End Sequencing Primer Kit (App-A). 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 (ATOPlex E450 Dual Barcode Balanced Library Reagent was spiked in at 40% in this study) using the DNBSEQ-G400RS High-throughput Rapid Sequencing Set (FCS PE300) and High-Throughput Sequencing Primer Kit (App-D).

Bioinformatics analysis

FASTQ data were analyzed using Cogent™ NGS Immune Profiler, a software designed by Takara Bio for sequencing analysis of libraries prepared using SMART-Seq Human BCR (with UMIs) or SMARTer Human TCR a/b Profiling Kit v2. This pipeline consists of preprocessing, UMI-based analysis, clonotype calling, and reporting. The output CSV files from the pipeline were uploaded to the web-based Cogent NGS Immune Viewer for visualization and figure generation. The specific workflow can be referred to Cogent™ NGS Immune Profiler v1.5 User Manual on Takara Bio official website (https://www.takarabio.com/documents/User%20Guides/Cogent%20NGS%20Immune%20Profiler%20v1.5%20User%20Manual.pdf, P24~34).

Group	Sample ID	Cell lines	Library types	Library name
1 (Takara Bio)	U081	DAUDI cell line	BCR heavy chain	DAUDI-BH
	U082	CA46 cell line	BCR heavy chain	CA46-BH
	U083	DAUDI + CA46 cell line(ratio of 1:3)	BCR heavy chain	DAUDI/CA46-BH
	U084	Human PBMC	BCR heavy chain	PBMC-BH
	U085	Human PBMC + 10% of U083 cell line	BCR heavy chain	PBMC/DAUDI/CA46-10-BH
	U086	Human PBMC + 1% of U083	BCR heavy chain	PBMC/DAUDI/CA46-1-BH
	U089	DAUDI cell line	BCR light chain	DAUDI-BL
	U090	CA46 cell line	BCR light chain	CA46-BL
	U091	DAUDI + CA46 cell line (ratio of 1:3)	BCR light chain	DAUDI /CA46-BL
	U092	Human PBMC cell line	BCR light chain	PBMC-BL
	U093	Human PBMC + 10% of U091 cell line	BCR light chain	PBMC/DAUDI/CA46-10-BL
	U094	Human PBMC + 1% of U091 cell line	BCR light chain	PBMC/DAUDI/CA46-1-BL
2 (MGI)	U029	Human PMBC	BCR heavy chain-02	PBMC-BH02
	U030	Human PMBC	BCR light chain-02	PBMC-BL02
	U25	Human PMBC	TCR α/β-replicate 1	PBMC-Tα/β-1
	U26	Human PMBC	TCR α/β-replicate 2	PBMC-Tα/β-2

Table 1. Details about libraries prepared by Takara Bio and MGI.

Results

DNBSEQ-G400 generated high-quality data from Takara Bio libraries

Among all the samples, ~10 million paired-end (PE) reads were obtained with Q30 above 95%, which is significantly higher than accepted criteria (85%), suggesting that DNBSEQ-G400 offers unique advantages on the sequencing data quality. The percentage of data that can be used for the clonal identification of TCR or BCR (Determined %) is also higher than expected (93%), with values over 97% (Table 2). MIGEC were used in Cogent NGS Immune Profiler to group reads into MIGs based on UMI sequences. After MIG collapse, reads belonging to the same MIG are collapsed into a single read for the downstream determination of clonotype. As shown in Table 2, the number of reads after MIG collapse (Collapsed) and Total MIGs are also as expected.

Library name	Library Type	Total Reads	GC%	Q30%	Determined%	Total MIGs	Collapsed
DAUDI-BH	BCR	9,620,845	53.09	95.87	97.9	537,764	177,112
CA46-BH	BCR	8,987,764	59.93	97.12	97.8	362,526	125,210
DAUDI/CA46-BH	BCR	11,453,613	54.99	96.54	97.9	473,711	168,055
PBMC-BH	BCR	12,967,353	56.49	96.51	97.9	287,158	44,895
PBMC/DAUDI/CA46-10-BH	BCR	12,245,364	55.38	96.93	97.9	431,219	121,196
PBMC/DAUDI/CA46-1-BH	BCR	11,283,318	56.27	96.62	97.9	280,605	48,578
DAUDI-BL	BCR	9,408,528	49.37	96.86	98.3	420,193	148,720
CA46-BL	BCR	11,494,906	55.14	96.52	98.4	816,883	324,335
DAUDI /CA46 -BL	BCR	11,458,391	51.76	96.9	98.4	530,379	206,286
PBMC-BL	BCR	12,778,326	53.01	96.88	98.0	215,465	85,796
PBMC/DAUDI/CA46-10-BL	BCR	11,371,160	52.11	96.76	98.3	427,488	171,335
PBMC/DAUDI/CA46-1-BL	BCR	10,248,519	52.62	96.7	98.1	230,569	87,923
PBMC-BH02	BCR	9,171,213	55.75	96.52	99.2	449,988	108,118
PBMC-BL02	BCR	8,780,848	55.35	96.85	99.1	351,640	216,907
PBMC-Tα/β-1	TCR	11,172,773	49.54	95.85	99.2	369,646	145,040
PBMC-Tα/β-2	TCR	10,020,030	49.43	96.6	99.1	366,982	160,961

Table 2. DNBSEQ-G400 performance for TCR/BCR libraries. Determined%: percentage of reads determined to be a TCR or BCR clonotype. Total MIGs: total molecular identifier groups. Collapsed: number of reads after MIG collapse.

DNBSEQ-G400 showed accurate clonotype detecting ability in TCR and BCR libraries

To obtain more clonotype information, the proportion of reads determined to be different clonotypes were calculated. Specifically, DAUDI-BH, CA46-BH, and DAUDI/CA46-BH all exhibit a high proportion of IgM content. Similarly, DAUDI-BL, CA46-BL, and DAUDI/CA46 -BL show a high proportion of IgK content. The other BCR libraries are from PBMC. PBMC-BH02, PBMC-BH, PBMC/DAUDI/CA46-10-BH, and PBMC/DAUDI/CA46-1-BH are heavy chains of BCR, and their compositions are largely similar, with the main components being IgM, IgA, and IgG. PBMC-BL02, PBMC-BL, PBMC/DAUDI/-CA46-10-BL, and PBMC/DAUDI/CA46-1-BL are light chains of BCR. The primary constituents of their compositions are IgK and IgL. Finally, PBMC-Tα/β-1 and PBMC-Tα/β-2 are TCR samples, characterized by T-cell receptor alpha and beta (TRA and TRB) genes (Figure 1). All clones identified above are consistent with our expectations.

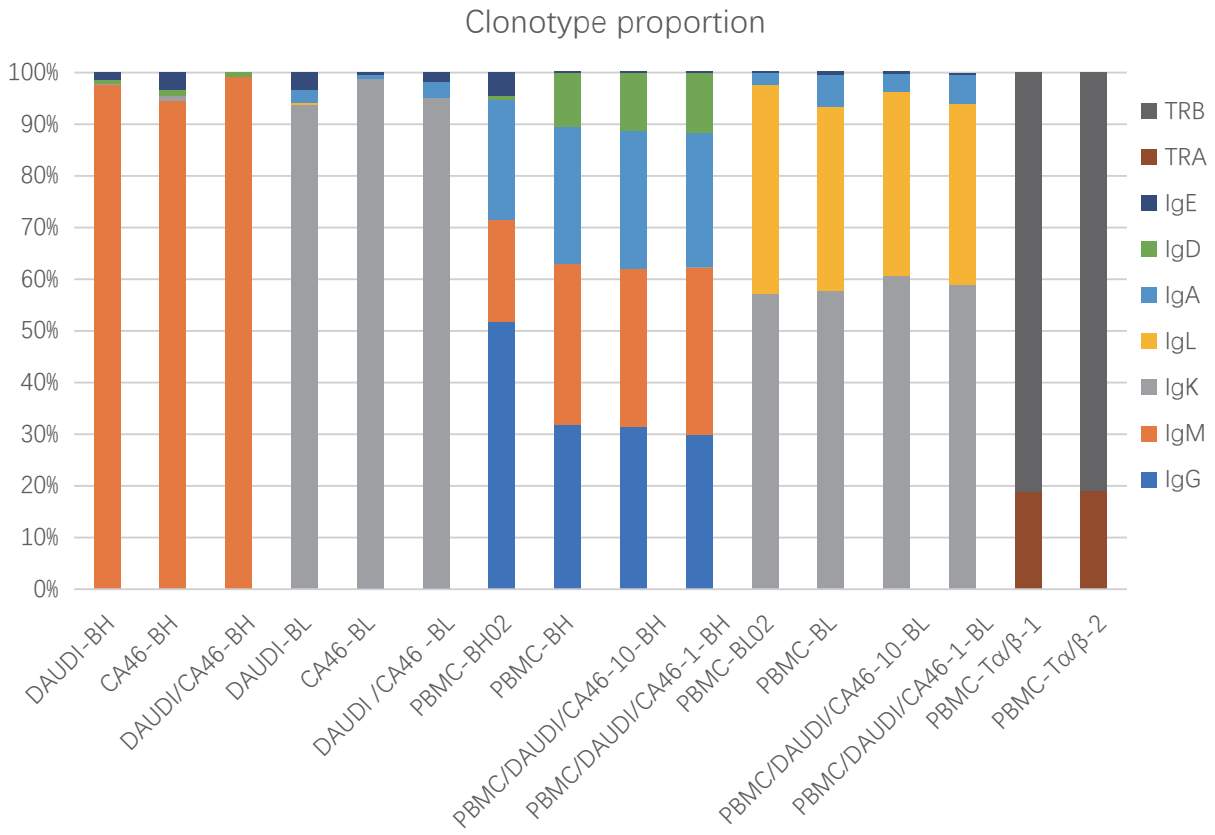

Figure 1. Proportion of TCR and BCR clonotypes discovered using DNBSEQ-G400. Proportion: The observed fraction of cell barcodes with a particular clonotype.

Each TCR or BCR has a variable region and constant region, and the variable regions are assembled through V(D)J recombination. The third complementarity determining region (CDR3) is the core region of V(D)J, so the CDR3 length distribution was calculated for an overview of the V(D)J combinations. The results in Figure 2 show that most of the samples shared an enrichment near 30 to 40 bases, and the overall distribution is from 30 to 80 bases, which meets the criteria established in previous studies[6].

The clonotype distribution analysis revealed that some samples display a similar proportion of subtypes (Figure 1). To investigate this further, the researchers conducted a detailed observation of V(D)J combination for each of the CDR3 genes using Cogent NGS Immune Profiler. As D segments are very short and highly altered during the rearrangement process, V and J segment assignments are much more reliable. V/J usage plots and chord diagrams for V-J pairings are showed in Figure 3. Figure 3A shows the V/J distribution of DAUDI/CA46-BL, PBMC/DAUDI/-CA46-10-BL, and PBMC-Tα/β-1, which correspond to the chord diagrams in Figure 3 B, D, E.

The chord diagrams indicate that each library has distinct V-J pairings. For example, DAUDI-BH and CA46-BH exhibited significantly different V-J pairings corresponding to the respective cell line, while DAUDI/CA46-BH utilized V-J pairings from both DAUDI-BH and CA46-BH, as expected (Figure 3B, left). A similar pattern in DAUDI-BL, CA46-BL, and DAUDI/-CA46-BL is also observed (Figure 3B, right).

In the BCR heavy chain subgroup (PBMC-BH02, PBMC-BH, PBMC/DAUDI/CA46-10-BH, and PBMC/DAUDI/CA46-1-BH) and light chain subgroup (PBMC-BL02, PBMC-BL, PBMC/DAUDI/-CA46-10-BL, and PBMC/DAUDI/CA46-1-BL), some samples had complex combinations of V/J genes, which correlated to BCR diversity in PBMC samples (Figure 3C, D). Some samples like PBMC/DAUDI/CA46-10-BH had a high proportion of IGHJ4, IGHV3-74, and IGHV5-51, which was consistent with DAUDI/CA46-BH (Figure 3B, C). Additionally, two technical replicates of TCR samples (PBMC-Tα/β-1 and PBMC-Tα/β-2) had nearly identical V/J gene usage (Figure 3E), which contrasted with PBMC-BH02 vs. PBMC-BL02 (Figure 3C, D).

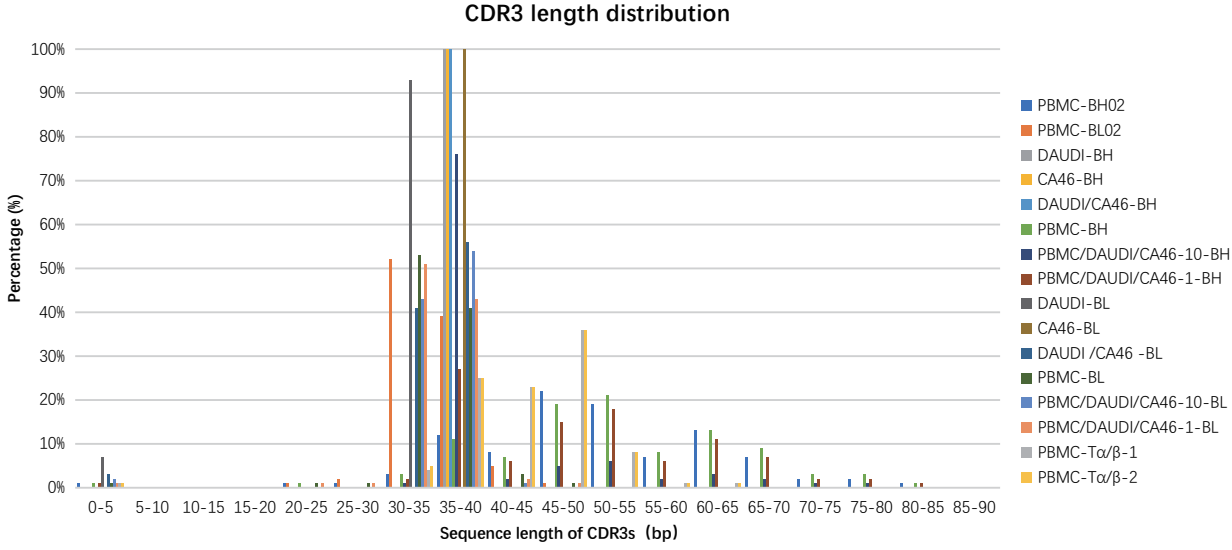

Figure 2. CDR3 length distribution. CDR3 sequences with lengths over 90 bases were not included in the comparison. The above results were obtained from DNBSEQ-G400.

Figure 3. V/J usage plots and chord diagrams for V-J pairings. (A) V/J usage plots—shown for DAUDI/CA46-BL, PBMC/DAUDI/CA46-10-BL, and PBMC-Tα/β-1—provide the distribution of V/J gene usage across clonotypes. This information allows evaluation of the distribution and co-occurrence of different V/J germline gene segments. (B, C, D ,E) Chord diagrams provide a visual representation of clonotype distribution and show pairing of V and J segments. Chord widths represent the pairing frequency. The above results were obtained from DNBSEQ-G400.

The researchers also determined the top 10 clones for BCR or TCR from PBMCs and a mixed cell line (Table 3).

Library type	No.	CDR3s_nt	Clonotype	Proportion (%)
BCRs in PBMCs	1	TGTCAGCAGCGTAGCAACAAGTCTTTT	IGK	2.12
	2	TGTGCGAGACAATACGTGGCCGTCTGG	IGG	0.58
	3	TGTGCGAAGCAGTACCTGGTTGGGCGGGGCTACTGG	IGG	0.47
	4	TGTCAGCAATATTACAGTACTCCGTGGACGTTC	IGK	0.44
	5	TGTCAGCAGCGTAGCAACTGGCCTCTCACTTTC	IGK	0.4
	6	TGTCAACAGTATGTTACTTATCCGTGGACGTTC	IGK	0.38
	7	TGTCAGCAGCGTAGCAAGAAGTCTTTT	IGK	0.32
	8	TGTCAGCAATATTATAGTACTCCGTACACTTTT	IGK	0.27
	9	TGTCAGCAGTATAATAACTGGCCTCGGACGTTC	IGK	0.23
	10	TGTCTACAAGATTACAATTACCCTCGGACGTTC	IGK	0.22
BCRs in a mixed cell line	1	TGTCAGCAGTATGGTAGTTCACCTCCGTGGACGTTC	IGK	47.58
	2	TGTCAACACAATTACAATTTCTCGTTCACTTTC	IGK	30.27
	3	GAGAACACTAGGAATTTACTCAGCCAGTG	IGK	1.67
	4	TGTCAATCAGCAGACAGCAGTGGTAGTTATGTCTTC	IGL	0.8
	5	GACATCCAGATGACCCAGTCTCCATCCTC	IGK	0.16
	6	TGTCAGCAGTATAATAACTGGCCTCCGTACACTTTT	IGK	0.08
	7	TGTCAGCAATATTATAGTACTCCGTACACTTTT	IGK	0.08
	8	TGTCAACAGAGTTACAGTACCCCGTACACTTTT	IGK	0.06
	9	TGTCAGCAATATCTTAACTGGCTCACGATCACCTTC	IGK	0.06
	10	TGTCAGCAATCTTATAATCATCCTCGCACTTTC	IGK	0.06
TCRs in PBMCs	1	TGTGTGGTGAACCCGCCGAACACAGGCTTTCAGAAACTTGTATTT	TRA	4.14
	2	TGTGCCACCAGCAGAGATACGGGGTTCTACGAGCAGTACTTC	TRB	1.51
	3	TGTGCCAGCAGCGGCCAGGGGGACCTGGGAGAGACCCAGTACTTC	TRB	0.74
	4	TGTGCCAGCAGCCCCCGAGACAGGGAGTTCTACAATGAGCAGTTCTTC	TRB	0.71
	5	TGTGCCAGCAAGGTAGCGCTCAATACCTACGAGCAGTACTTC	TRB	0.67
	6	TGTGCCAGCAGTTTGGGGCAGGCCAATGAGCAGTTCTTC	TRB	0.66
	7	TGCGCCAGCAGCCAAGTTATGCGGGGCGCGAACACTGAAGCTTTCTTT	TRB	0.49
	8	TGTGCCAGCAGTTTGGGCCTGCACTACGAGCAGTACTTC	TRB	0.49
	9	TGTGCCAGCAGTTACCAGGGCCATCAGCCCCAGCATTTT	TRB	0.45
	10	TGTGCCAGCAGCCCGACTACGTGGGAGGCCCAGTACTTC	TRB	0.4

Table 3. Top 10 clonotypes for BCRs in PBMCs (PBMC-BH02 and PBMC-BL02), BCRs in a mixed cell line (PBMC/DAUDI/-CA46-10-BL), and TCRs in PBMCs (PBMC-Tα/β-1 and PBMC-Tα/β-2). The above are the sequencing results on DNBSEQ-G400.

DNBSEQ-G400 showed high sequencing quality comparable to Competitor A in both PBMC and monoclonal/mixed cell line libraries

To further verify the performance of DBSEQ-G400, the immune profiling results from the same libraries sequenced on DNBSEQ-G400 and on Competitor A's sequencer were compared. All samples were downsampled to 1 x 106 reads for fair comparisons. Compared to Competitor A, mapping rates (aligned %) of DNBSEQ-G400 sequencing data are slightly higher(Figure 4 A), and the rates of undetermined reads (undetermined %) and failed linker-based correction (flc %) are slightly lower (lower is better) (Figure 4B, C). The reads assigned to chains shorter than 30 bp in length (short %) on the two platforms are 0% (data not shown). Taken together, DNBSEQ-G400 produced slightly higher quality data than Competitor A's sequencer.

Downstream analysis suggested that the data from DNBSEQ-G400 have slightly higher numbers of total MIGs in PBMC or monoclonal/mixed cell lines (Figure 5A, B) and higher clonotype counts in PBMCs (Figure 5C) than Competitor A. Again prove that DNBSEQ-G400's performance is slightly better than the Competitor A platform.

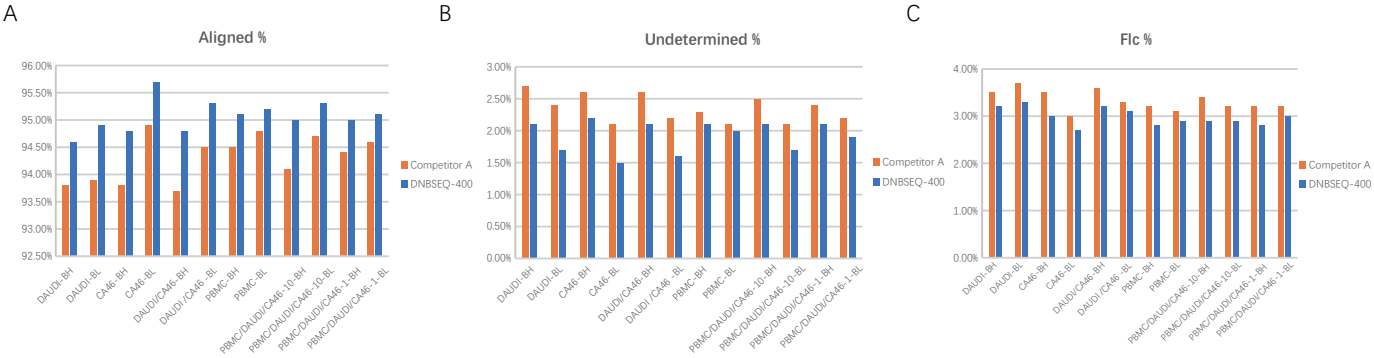

Figure 4. QC statistics comparison of DNBSEQ-G400 and Competitor A's sequencer.(A). Aligned %: The proportion of mapping reads.(B) Undetermined %: Percentage of reads from undetermined chains. (C) Flc % :Percentage of reads assigned to chains failed linker-based correction.

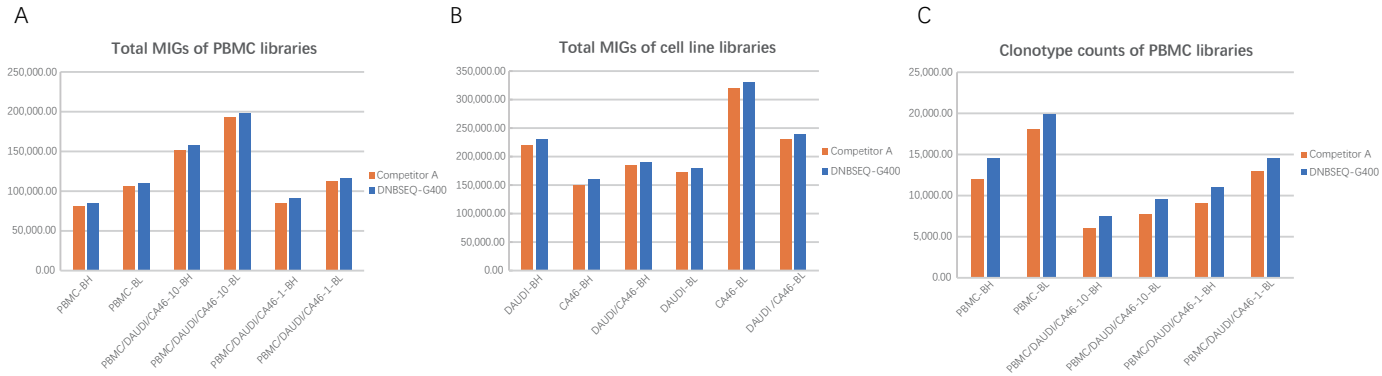

Figure 5. Comparison of DNBSEQ-G400 and Competitor A with regards to total MIGs and clonotype counts. (A, B) Total MIGs in PBMC and monoclonal/mixed cell lines, respectively. (C) Clonotype counts in PBMC samples.

Conclusions

The performance evaluation of SMARTer Human TCR a/b Profiling Kit v2 and SMART-Seq Human BCR (with UMIs) on DNBSEQ-G400 and Competitor A sequencing platforms shows that SMART technology is adaptable to the DNBSEQ platform. In this study, the constructed BCR or TCR libraries prepared from PBMC, monoclonal, or mixed cell lines can all achieve quality results on the DNBSEQ platform. The diversity of TCR and BCR is achieved through the rearrangement of variable (V), diversity (D), joining (J) genes. The DNBSEQ-G400 platform combined with Takara Bio TCR/BCR profiling kits has an excellent ability to determine TCR/BCR clonotypes, and this combination performs slightly better than with Competitor A's platform.

DNBSEQ-G400 is built with a new flow cell system that can flexibly support a variety of different sequencing modes. It adopts optimized optical and biochemical systems, which can complete the sequencing process rapidly. PE300 (FCS) sequencing at full capacity takes only about 98 hours.

DNBSEQ-G400 combined with Takara Bio immune profiling library preparation kits can help researchers further understand the V(D)J rearrangement process, identify disease-related TCRs and BCRs, and provide new insights into understanding disease status and discovering therapeutic targets.

DNBSEQ-G400 Genetic Sequencer

References

1. Friedensohn S, Khan TA, Reddy ST. Advanced Methodologies in High-Throughput Sequencing of Immune Repertoires. Trends in Biotechnology 2017. 35(3):203-14.
2. McDaniel JR, Brandon J DeKosky, Tanno H, Ellington AD & Georgiou G. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. Nature Protocols 2016. 11:429-42.
3. Hou D, Chen C, Seely EJ, Chen S and Song Y. High-Throughput Sequencing-Based Immune Repertoire Study during Infectious Disease. Front. Immunol. 2016. 7
4. SMART-Seq® Human BCR (with UMIs) User Manual: www.takarabio.com/documents/User%20Manual/SMART/SMART-Seq%20Human%20BCR%20%28with%20UMIs%29%20User%20Manual.pdf
5. SMARTer® Human TCR a/b Profiling Kit v2 User Manual: www.takarabio.com/assets/a/117462
6. Rock EP, Sibbald PR, Davis MM, Chien YH. CDR3 length in antigen-specific immune receptors. J Exp Med 1994. 179:323–328.

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	DNBSEQ-G400 Genetic Sequencer (Configuring A)	900-000168-00
Software	Cogent NGS Immune Profiler	https://www.takara-bio.com/products/next-generation-sequencing/bioinformatics-tools/cogent-ngs-immune-profiler*
	Cogent NGS Immune Viewer	https://www.takara-bio.com/products/next-generation-sequencing/bioinformatics-tools/cogent-ngs-immune-viewer*
Library Prep	SMARTer Human TCR a/b Profiling Kit v2	634478*
	SMART-Seq Human BCR (with UMIs)	634777*
	MGI Universal Library Conversion Kit (App-A)	1000004155
Sequencing Reagents	a) ATOPlex E450 Dual Barcode Balanced Library Reagent	940-000637-00**
	b) DNBSEQ-G400RS High-throughput Rapid Sequencing Set (FCS PE300)	940-000152-00**
	c) High-Throughput Pair End Sequencing Primer Kit (App-A)	1000020832**
	d) High-Throughput Sequencing Primer Kit (App-D)	1000028550**

*The relevant products are available and can be ordered on the Takara Bio official website.

**Use either b) + c) when DNBSEQ-G400 with ECR 6.0 is chosen, or a) + b) + d) when DNBSEQ-G400's ECR version lower than 5.2 is chosen.

MGI Tech Co.,Ltd

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

Version: October 2023

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Author: Lian Ying, HoYingyee, Lim Tingsen Benson, Likai Miao, 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.