Efficient and sensitive high-throughput human B-cell receptor repertoire profiling using SMART® technology

Shaveta Goyal*, Asami Sugihara, Wenwen Xiang, Ilona Holcomb, Karthikeyan Swaminathan, Bryan Bell, Yue Yun & Andrew Farmer

Takara Bio USA, Inc., San Jose, CA 94043, USA   *Corresponding Author

TaKaRa

Abstract

OBJECTIVE: B-cell receptor (BCR) repertoire profiling is increasingly used in health and pathogenic contexts with the goal of biomarker discovery. However, current sequencing technologies are limited in their ability to generate data accurately and reproducibly for all BCR isotypes. To overcome these limitations, we have developed a new kit to accurately profile all heavy (A, D, E, G, M) and light-chain (K, L) isotypes—an end-to-end solution, from library preparation to streamlined data analysis. Here we present data on an updated approach for efficient and high-throughput BCR repertoire profiling of human samples.

METHODS: Libraries were prepared from human peripheral blood cells (10 ng–1 µg total RNA) or from B-cells (1 ng–100 ng total RNA) using our new human BCR repertoire profiling kit (~2.5 hours hands-on time). Prepared libraries were then analyzed on the Illumina® Miseq® benchtop sequencer using 300-bp paired-end reads.

RESULTS: For each library, >90% of sequencing reads were on-target while the most highly represented clonotype was found to remain consistent among technical duplicates across a range of input amounts. In comparison to the previous version of our BCR-sequencing kit, the new approach enabled a ~4x increase in total clonotype count observed across various RNA inputs. Furthermore, a sensitivity assay demonstrated that B-cell RNA corresponding to a single clonotype could be detected above background levels when spiked into input total RNA at a relative concentration of 0.001%.

CONCLUSION: Our new human BCR repertoire profiling kit was found to accurately and reproducibly profile B-cell clones and provide information on the diversity of BCR repertoire in human samples.

Methods

BCR libraries were constructed from human PBMC total RNA (Takara Bio, cat. #636592) and human B-Cell (CD19+) total RNA (Miltenyi Biotech, cat. # 130-0930169). For sensitivity test, total RNA from TIB-190 B-cell carcinoma cell line (ATCC) was spiked in 10ng of PBMC RNA at the listed concentrations - 100pg, 10pg, 1pg and 0.1pg. Total RNA from TIB-190 cells was extracted using the Nucleospin RNA plus kit (Macherey-Nagel, cat. # 740984.50).

Libraries were produced using first-strand cDNA as a template in a single PCR reaction for all the isotypes. The PCR product was used as template for nested PCR for heavy chain isotypes (IgA/D/E/G/M) and light chain isotypes (IgK/L) in two separate PCRs. Following purification and size selection, libraries were validated using the Agilent 2100 Bioanalyzer. Libraries were spiked in with 20% PhiX to increase sequence diversity and sequenced on an Illumina MiSeq platform with 600-cycle V3 cartridges (cat. # MS-102-3003).

Sequencing data analysis was completed by the Takara Bio Cogent™ NGS Immune Profiler Software, which features MIGEC (Shugay, 2014) and MiXCR (Bolotin, 2015) software.
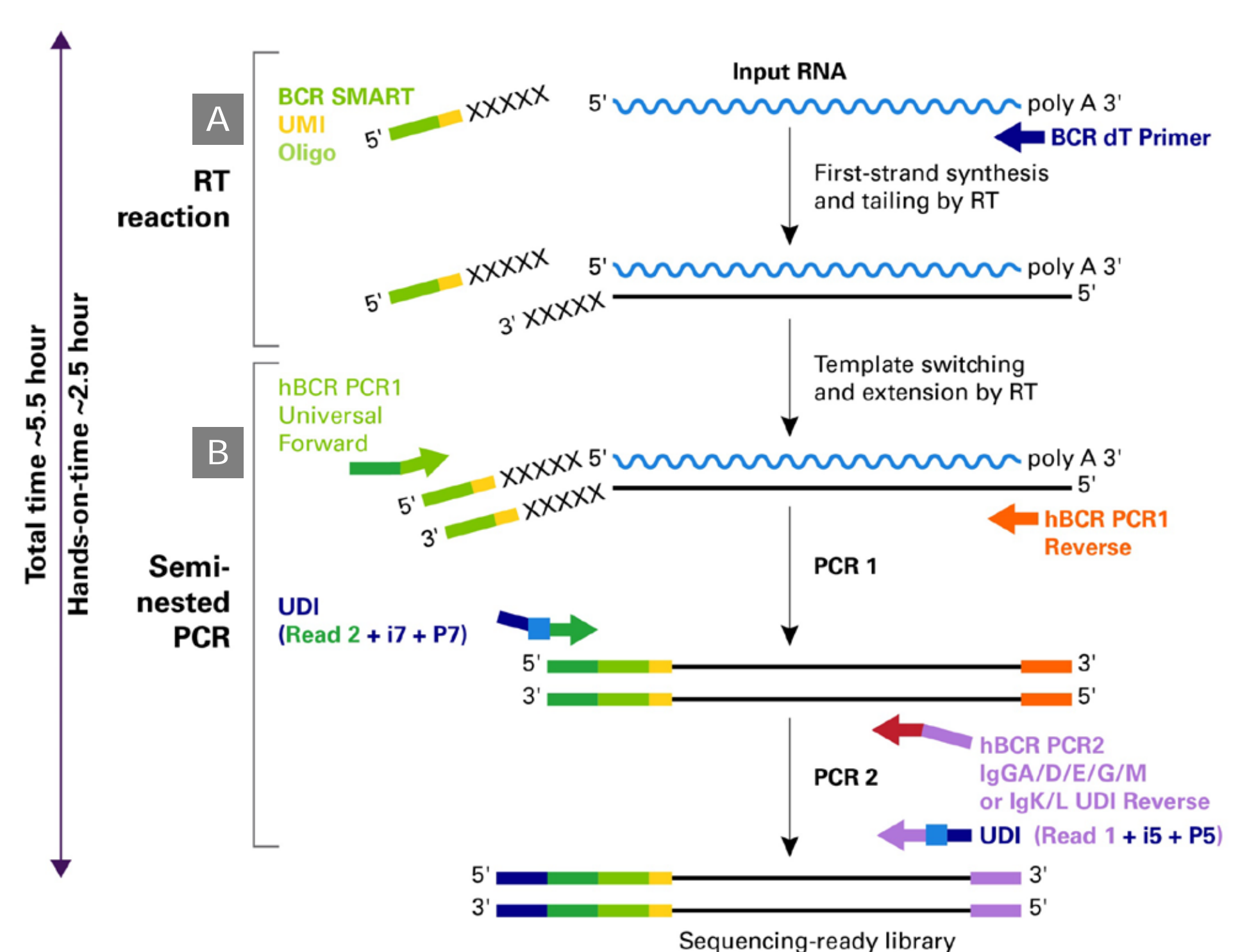
1 BCR library preparation workflow

Figure 1. Library preparation workflow and PCR strategy for BCR repertoire profiling using the updated version of the SMARTer® Human BCR IgG IgM H/K/L Profiling Kit (BCRv2). Panel A. First-strand cDNA synthesis is primed by the BCR dT Primer and performed by an MMLV-derived reverse transcriptase (RT). Upon reaching the 5' end of each mRNA molecule, the RT adds non-templated nucleotides to the first-strand cDNA. The SMART-Seq® v4 Oligonucleotide contains a sequence that is complementary to the non-templated nucleotides added by the RT and hybridizes to the first-strand cDNA. In the template-switching step, the RT uses the remainder of the SMART-Seq v4 Oligonucleotide as a template for the incorporation of an additional sequence on the end of the first-strand cDNA. Panel B. Full-length variable regions of BCR cDNA are selectively amplified by PCR using primers that are complementary to the oligonucleotide-templated sequence (SMART Primer 1) and the constant region(s) of BCR IgA/D/E/G/M/K/L isotypes (BCR Human Primer 1). A subsequent round of PCR is performed to further amplify variable regions of BCR-IgA/D/E/G/M and/or BCR- IgK/L subunits and incorporate adapter sequences, using BCR Human Primer 2 and UDI primer. Included in the primers are adapter and index sequences (read 2 + i7 + P7 and read 1 + i5 + P5, respectively) that are compatible with the Illumina sequencing platform. Following purification, size selection, and quality analysis, BCR cDNA libraries are sequenced on the Illumina platform using 300 bp paired-end reads.

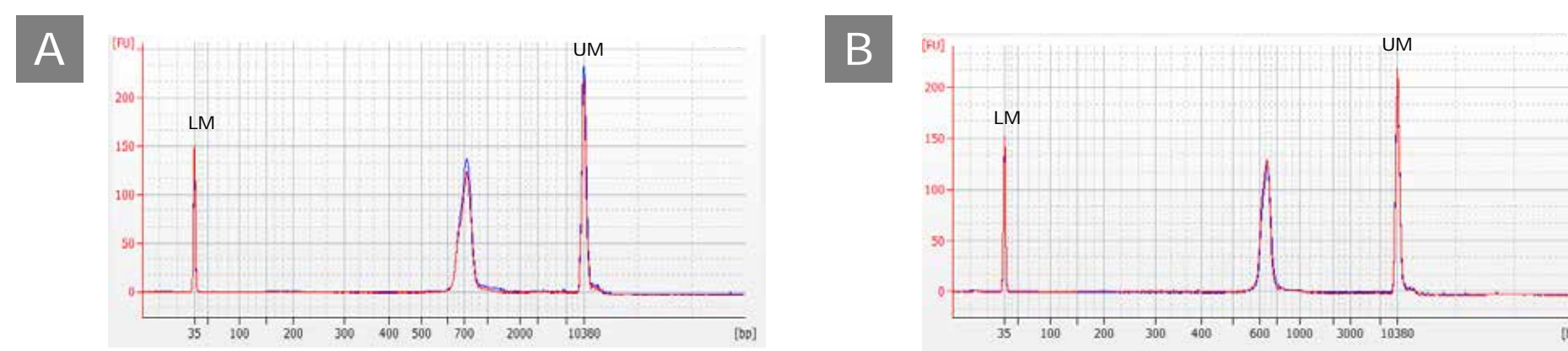## 2 Sequencing library validation and quality analysis

Figure 2. Electropherogram profiles of BCRv2 sequencing libraries. Libraries containing both heavy-chain and light-chain isotypes were generated using 10 ng of RNA obtained from a heterogeneous population of peripheral blood leukocytes. Electropherogram profiles of the final libraries were obtained on an Agilent 2100 Bioanalyzer. Peaks situated at the far left and right ends of each electropherogram correspond to DNA reference markers included in each analysis. Panel A. Typical Bioanalyzer profile of sequencing library for heavy-chain isotypes, obtained from peripheral blood leukocyte RNA (same library as in Panel A). Panel B. Typical Bioanalyzer profile of sequencing library for light-chain isotypes, obtained from peripheral blood leukocyte RNA. The library profiles from the Bioanalyzer displayed a peak ~750 bp for heavy-chain isotypes and a peak ~650 bp for light-chain isotypes, in line with the predicted peak sizes for those sequence fragments.

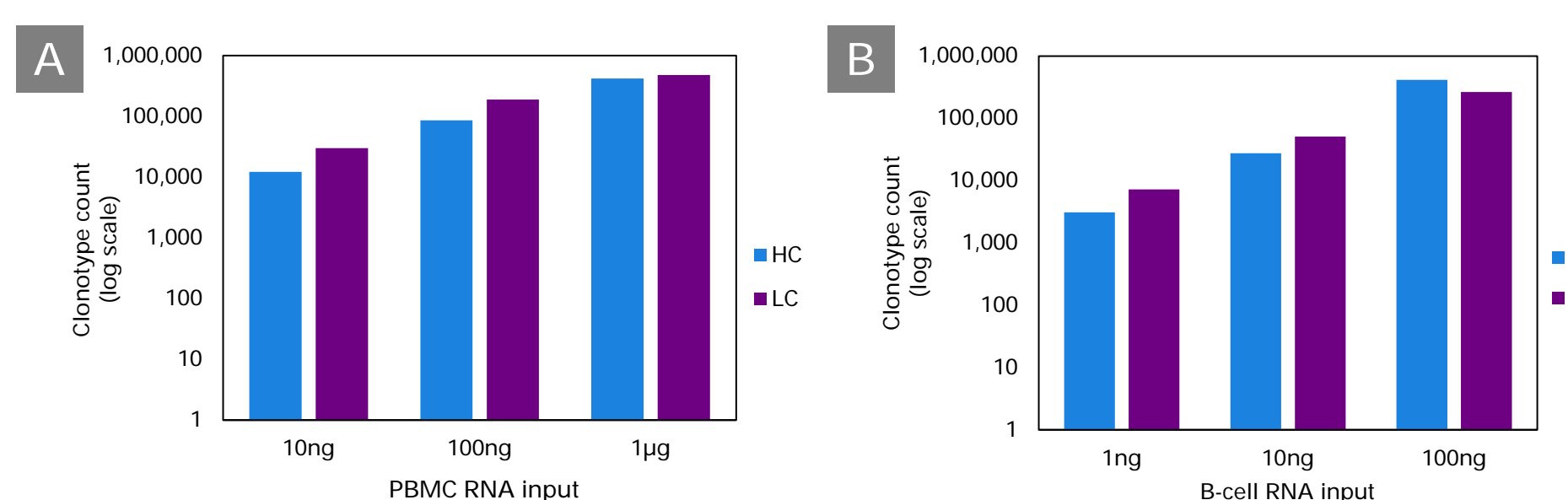## 3 Clonotype count across various RNA inputs

Figure 3. Clonotype count for varying sample input amount. To evaluate the performance of the kit for a range of input amounts, the BCRv2 workflow was performed on three different amounts of peripheral blood RNA (10 ng, 100 ng, and 1,000 ng) and three different amount of CD19+ B-cell RNA (1ng, 10ng and 100ng). The resulting cDNA libraries were sequenced on Illumina platform. Sequencing outputs were down sampled to ~2 million, 6 million and 12 million reads for PBMC RNA and ~1 million, 5 million and 12 million reads for B-cell RNA., Data was processed using Takara Bio Immune profiling software. Panel A. Bar plot showing clonotype counts for PBMC RNA at various inputs. Panel B. Bar plot showing clonotype count for B-cell RNA at various inputs. Comparison across various RNA inputs show consistent increase in clonotype count with increase in RNA input. HC=heavy chain, LC=light chain.

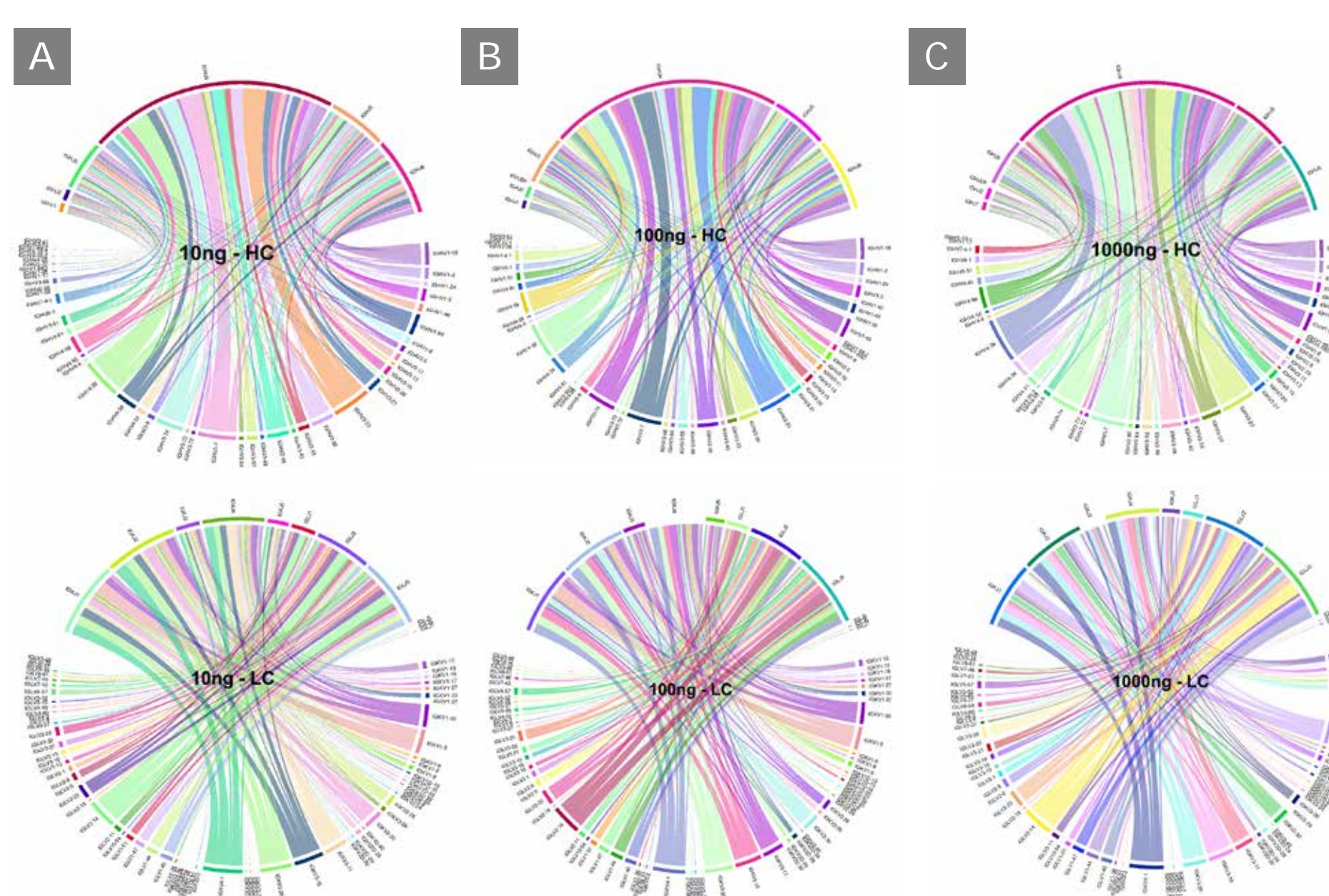## 4 Visual representation of clonotype distribution

Figure 4. The distribution of BCR clonotypes identified in the sequencing data depicted by chord diagrams. Each arc (on the periphery of each diagram) represents a V, D or J gene segment and is scaled lengthwise according to the relative proportion at which the gene segment is represented in the dataset. Each chord (connecting the arcs) represents a set of clonotypes including the indicated V-J combination and is weighted according to the relative abundance of that combination in the dataset. Panel A. Chord diagram for heavy chain (HC) and light chain (LC) isotypes of 10 ng input of PBMC RNA. Panel B. Chord diagram for heavy chain (HC) and light chain (LC) isotypes of 100 ng input of PBMC RNA. Panel C. Chord diagram for heavy chain (HC) and light chain (LC) isotypes of 1000 ng input of PBMC RNA. Comparison of the three diagrams suggests that the indicated clonotypes are identified at similar proportions for each RNA input amount.
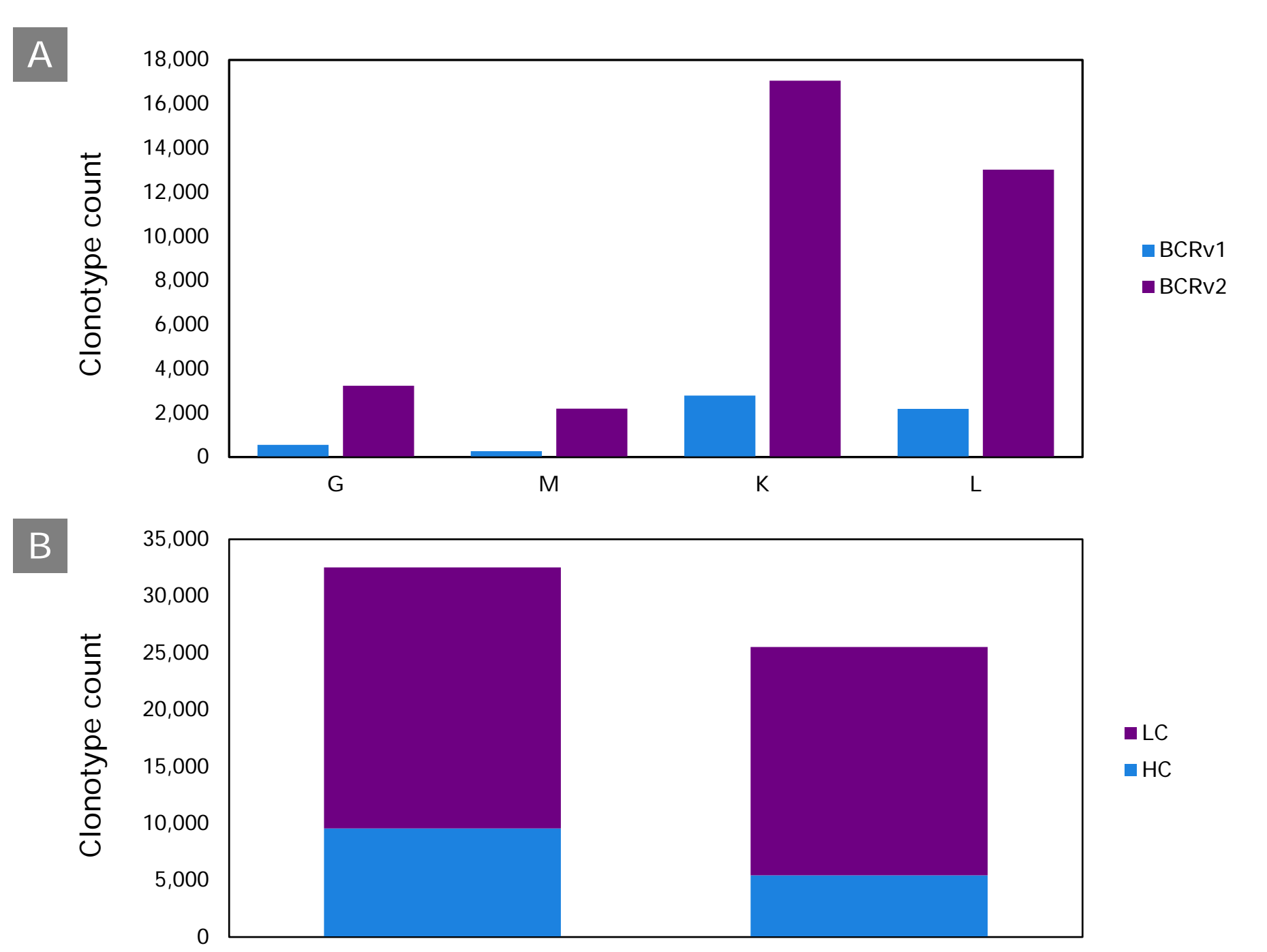
5 Performance: clonotype count

Figure 5. Superior data quality. Libraries were generated using 10 ng of human PBMC total RNA according to the manufacturer's instructions. The resulting cDNA libraries were sequenced on Illumina platform. Through these data, it is evident that the updated BCR profiling kit (BCRv2) was found to have better performance as compared to both the previous version SMARTer Human BCR IgG IgM H/K/L Profiling Kit (BCRv1) and the competitor's kit (Company N). Panel A. Sequencing outputs were downsampled to ~1 million for all the libraries – IgG/M/K/L for BCRv1; heavy and light chain libraries for BCRv2. In comparison to BCRv1, BCRv2 generated approximately 550% more than previous version. Panel B. Sequencing outputs were downsampled to ~1 million for Company N libraries and 500,000 each for heavy and light chain libraries for BCRv2. In comparison to Company N, BCRv2 generated approximately 9.5K and 22.9K clonotypes for heavy chain (HC) and light chain (LC) isotypes respectively, representing a 127% increase against Company N. These results demonstrate the superior data quality in comparison to Takara Bio's previous version of BCR profiling kit and to the Company N's immune profiling method.

6 Performance: sensitivity

10ng PBMC RNA + spike-in RNA	% spike-in RNA	Detected reads (after UMI collapse) TIB190	Total reads
100 pg spike-in	1%	3167	200,000
10 pg spike-in	0.10%	231	200,000
1 pg spike-in	0.01%	18	200,000
0.1 pg spike-in	0.001%	3	200,000

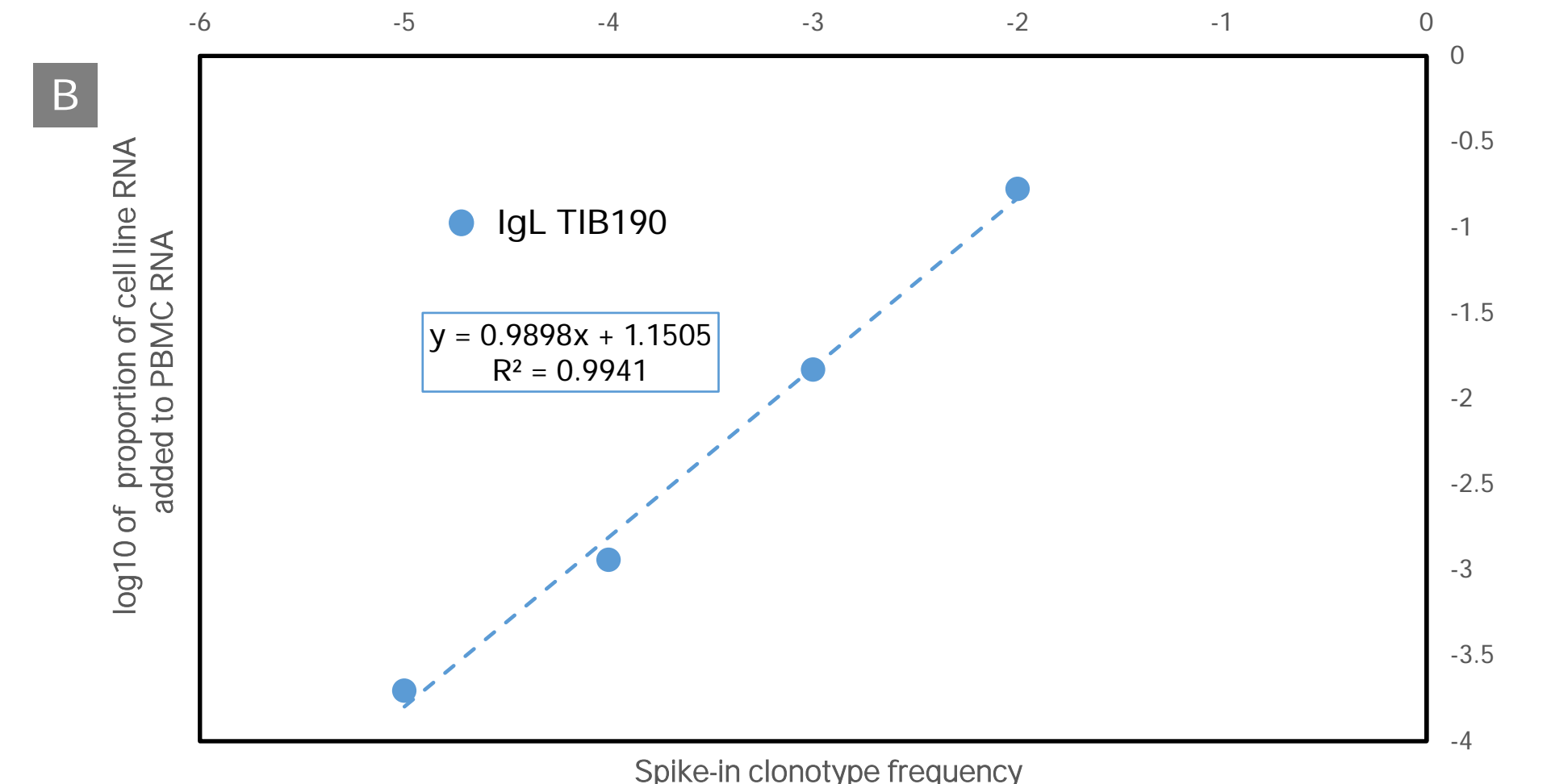$$y = 0.9898x + 1.1505$$
$$R^2 = 0.9941$$

Figure 6. Successful identification of low-abundance clones. 100pg, 10pg, 1pg and 0.1pg of RNA extracted from TIB-190 cell line was spiked into 10 ng of PBMC RNA. Panel A. Clone counts at different spike-in levels are listed in the table. Libraries were normalized to 200,000 reads and all counts were measured after UMI-based consensus collapse. Panel B. Calculated correlation between spike-in RNA proportions and detected clonotype frequencies. (both axes logarithmically transformed).
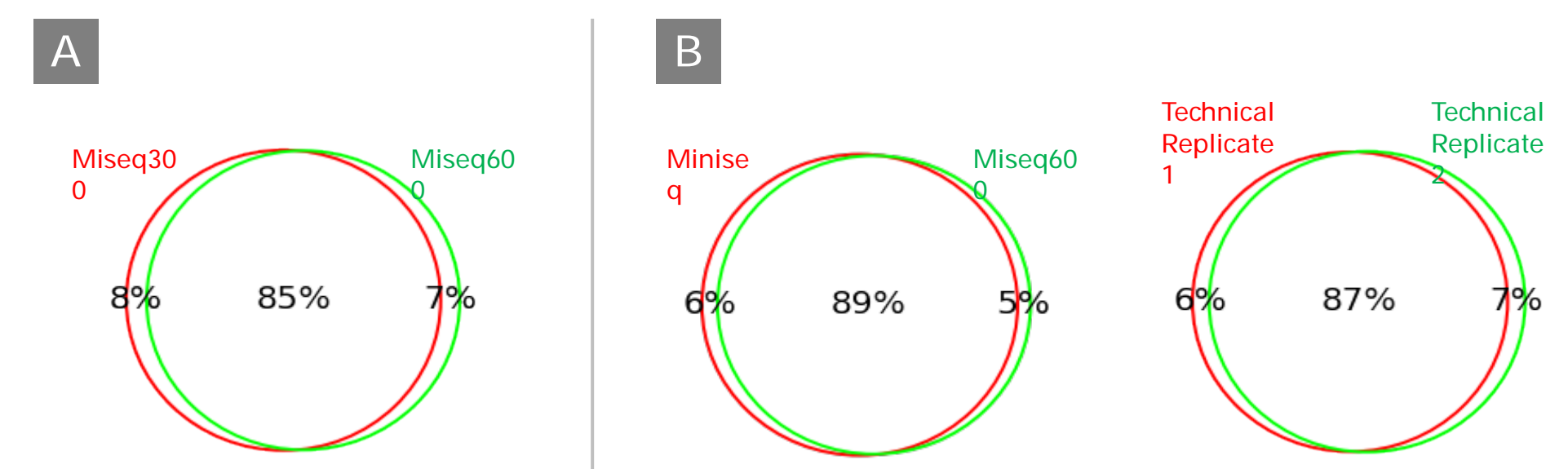
7 Performance: reproducibility

Figure 7. Evaluation of clonotype reproducibility. BCR profiling libraries from 10 ng of PBMC RNA and 1 ng B-cell RNA were prepared using the BCRv2 workflow. Panel A. Technical replicates prepared with 10 ng PBMC RNA were sequenced on the Illumina Miseq platform. Data generated were downsampled to 1,000,000 reads and analyzed using Immune Profiler software. Here, the venn diagram illustrates 87% clonotype overlap between technical replicates libraries. Panel B. Libraries prepared using 1 ng B-cell RNA were sequenced on the same Miseq platform as mentioned previously with both a 600-cycle V3 cartridge and a 300-cycle V2 cartridge, as well as on the Illumina Miniseq™ platform. Data generated were downsampled to 1,000,000 reads and is analyzed using Immune Profiler software. Venn diagrams show ≥85% clonotype overlap between the libraries sequenced on different platforms.

8 Updated analysis tools

Preprocess	• Split reads by matching read sequence to different receptor chains (allows one mismatch) • Exclude short reads (<30 bp) and reads ambiguously matched to multiple receptor chains • If linker-based correction is enabled, exclude read failed correction
UMI-based Analysis	• Group reads into molecular identifier groups (MIG) using UMI • Conduct sequencing error correction and exclude reads that fail abundance check • Generate collapsed-read FASTQs for downstream analysis
Clonotype Calling	• Align reads to V(D)J sequences • Assemble alignments • Define and report clonotypes
Summary Report	• Summarize read QC statistics: chain-specific, short, undetermined, flc • Summarize mapping statistics: aligned, overlapped, mapped • Summarize clonotype details: numbers, percentage, nucleotide, and amino acid sequence
Cleanup (Optional)	• Configuration options available to user on a per run basis • Deletes intermediate files generated during processing

Figure 8. Cogent NGS Immune Profiler Software is a bioinformatic tool designed to analyze sequence data stored in FASTQ files generated from the BCRv2 kit. The Immune Profiler software incorporates two third-party software packages: MIGEC and MiXCR. Output includes UMI number, UMI threshold counts, QC results, and sequences.

Conclusions

- BCRv2 can capture entire BCR variable regions, allowing for the analysis of all heavy chain isotypes and/or light chain isotypes, all within the same experiment

- BCRv2 provides highly sensitive, reproducible, and efficient detection of low abundance clones for a range of total RNA input amounts

- BCRv2 also includes unique dual indexes (UDIs) to prepare libraries for multiplexing, has the flexibility to be sequenced on all Illumina instruments, and features an automation-friendly workflow

References

Shugay, M. et al. Towards error-free profiling of immune repertoires. Nat. Methods 11, 653–655 (2014).

Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods 12, 380–381 (2015).

Scan to download your copy of this poster and view additional posters, or visit
www.takarabio.com/abrf-2022

800.662.2566
www.takarabio.com