

Abstract

Profiling T-cell receptor (TCR) repertoires involves characterizing the diversity of TCR nucleotide sequences in a sample, and is an increasingly popular approach for analyzing the composition of the adaptive immune system in the context of human development and disease. While low-throughput approaches have yielded important insights concerning TCR repertoire dynamics, the development of next-generation sequencing (NGS) technologies has dramatically expanded the prospects for this research area.

Using SMART[®] technology, we have developed an NGS library preparation kit for TCR profiling that employs a 5' RACE-like approach to capture full-length variable regions of TCR- α and/or TCR- β subunits. With this approach, input RNA is reverse transcribed using dT priming. Following first-strand cDNA synthesis, semi-nested PCR is used to amplify TCR-specific sequences and add Illumina[®] sequencing adapters.

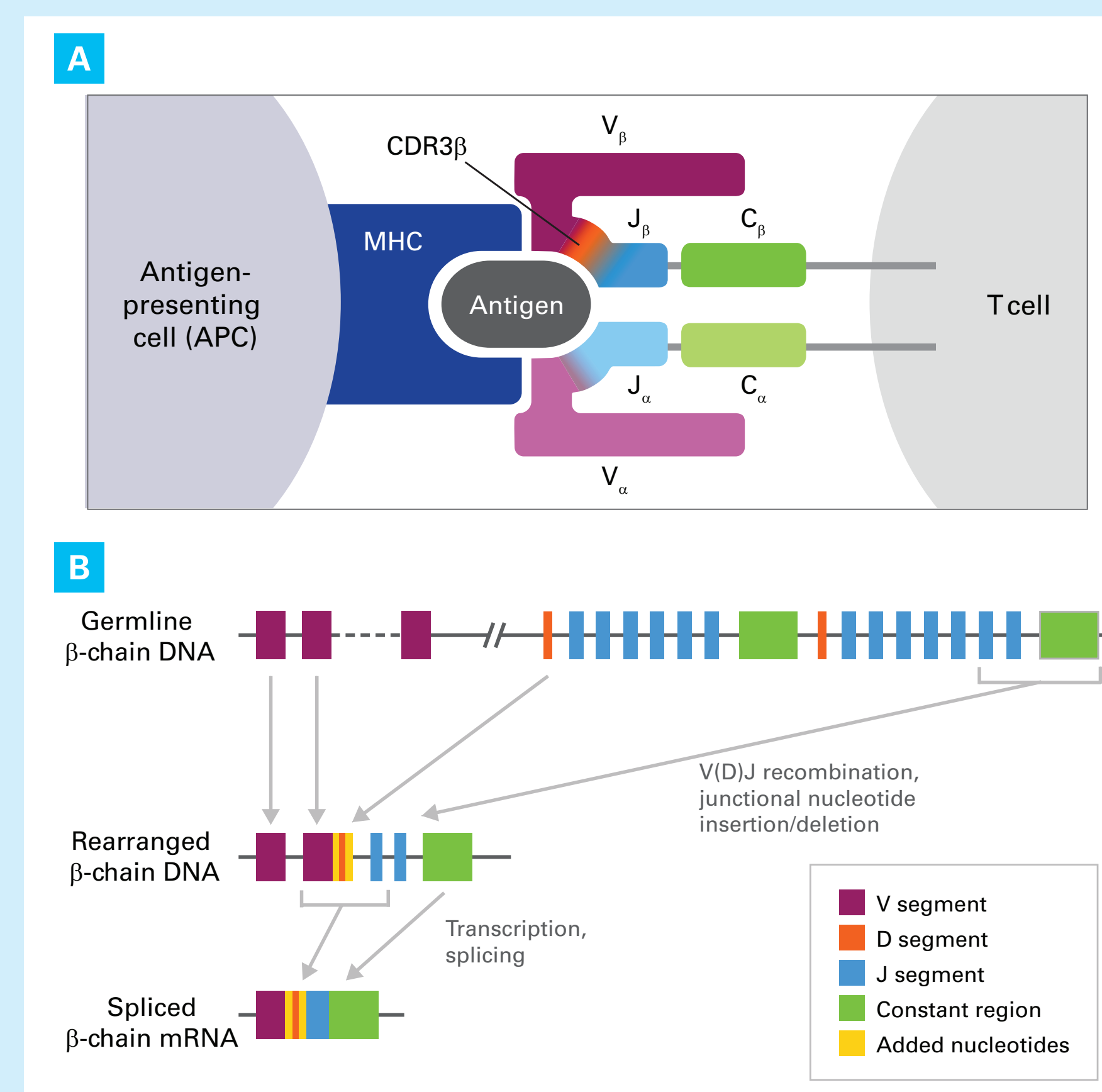
Starting with RNA obtained from human peripheral blood, libraries containing TCR- α and TCR- β sequences were generated with the kit and analyzed on an Illumina MiSeq[®] using 300-bp paired-end reads. For each experimental replicate, >70% of sequencing reads mapped to TCR variable regions, and the most highly represented clonotypes remained consistent across a range of input amounts. A separate sensitivity assay demonstrated that Jurkat RNA corresponding to a single clonotype could be detected above background levels when spiked into input RNA at a relative concentration of 0.1%.

Our approach enables the user to obtain sequencing-ready TCR libraries from 10 ng–3 μ g of total RNA or directly from 50–10,000 purified T cells in ~2.5 hours of hands-on time. The incorporation of Illumina indexes during PCR amplification allows for multiplexing of up to 96 samples in a single flow-cell lane. Whereas approaches that utilize genomic DNA as starting material require multiplexed PCR strategies, amplification of TCR sequences derived from RNA can be accomplished using single primer pairs for each subunit. In this way, our approach minimizes the likelihood of sample misrepresentation due to amplification biases.

Background

In humans and closely related species, cellular immunity is mediated by T cells, which participate directly in the detection and neutralization of pathogenic threats. Essential to T-cell function are highly specialized extracellular receptors (T-cell receptors or TCRs) that selectively bind specific antigens displayed by major histocompatibility complex (MHC) molecules on the surface of antigen-presenting cells (APCs). The vast majority of TCRs are heterodimers composed of two distinct subunit chains (α - and β -), which both contain variable domains and, in humans, are encoded by single-copy genes.

Given the relative specificity of TCR-antigen interactions, a tremendous diversity of TCRs are required to recognize the wide assortment of pathogenic agents one might encounter. To this end, the adaptive immune system has evolved a system for somatic diversification of TCRs that is unrivaled in all of biology. Ultimately, this process—commonly referred to as “(V)DJ recombination”—yields a population of T cells with sufficient TCR diversity to collectively recognize any peptide imaginable.



T-cell receptor structure and diversification. Panel A. A functional $\alpha\beta$ TCR heterodimer consisting of α - and β -subunit chains. TCR α -subunit chains consist of “variable” (V), “joining” (J), and “constant” (C) segments depicted in magenta, blue, and green, respectively, while TCR β -subunit chains include these and an additional “diversity” (D) segment, depicted in orange. The CDR3 region of the TCR β -subunit is labeled. The TCR is depicted on the T-cell surface, bound to an antigen associated with an MHC molecule on the surface of an APC. Panel B. V(D)J recombination and post-transcriptional processing of a TCR β -subunit chain. The TCR β locus includes over 50 V segments (magenta), 2 D segments (orange), and 13 J segments (blue). During somatic diversification, at least one of each segment type is randomly selected and further variability is introduced through the incorporation and/or deletion of additional nucleotides (yellow). Splicing of TCR mRNA combines a subset of the respective segments (along with a constant region) into a continuous unit. TCR α -subunit chains are generated via analogous mechanisms.

The seemingly endless number of potential TCR clonotypes poses significant challenges for researchers seeking to characterize T-cell repertoires in the context of human development and disease. The vast majority of TCR-profiling experiments performed thus far have focused on capturing genomic DNA or mRNA sequences that correspond to the complementarity-determining region 3 (CDR3) of the TCR- β subunit chain (Calis and Rosenberg, 2014; Woodsworth *et al.*, 2013). Given that the CDR3 region is thought to be unique to each TCR- β variant, sequence variation in this region has served as a useful proxy for overall T-cell repertoire diversity.

While sequencing genomic DNA may be preferable for certain TCR-profiling applications, mRNA-seq carries several important advantages. TCR mRNA templates are likely to be more highly represented than DNA templates in any one T cell, such that mRNA-seq affords greater sensitivity and allows for identification of low-abundance TCR variants. Furthermore, amplification of genomic DNA corresponding to TCR variable regions (including CDR3) requires multiplex PCR and is potentially susceptible to biases imposed by the various primer pairs. In contrast, the relatively shorter length of TCR mRNA templates allows for simpler amplification schemes in which entire TCR- α and TCR- β variable regions are captured with single primer pairs, minimizing the potential for amplification biases and allowing for analysis of both subunit chains in the same experiment. Additionally, for those interested in exploring TCR function, RNA-based approaches specifically capture expressed TCR sequences that are more likely to be functionally relevant.

References

- Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., & Chudakov, D. M. (2015) MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* 12(5):380–381.
- Calis, J.J.A. & Rosenberg, B.R. (2014) Characterizing immune repertoires by high throughput sequencing: strategies and applications. *Trends Immunol.* 35(12):581–590.
- Woodsworth, D.J., Castellarin, M., & Holt, R.A. (2013) Sequence analysis of T-cell repertoires in health and disease. *Genome Med.* 5(10):98.

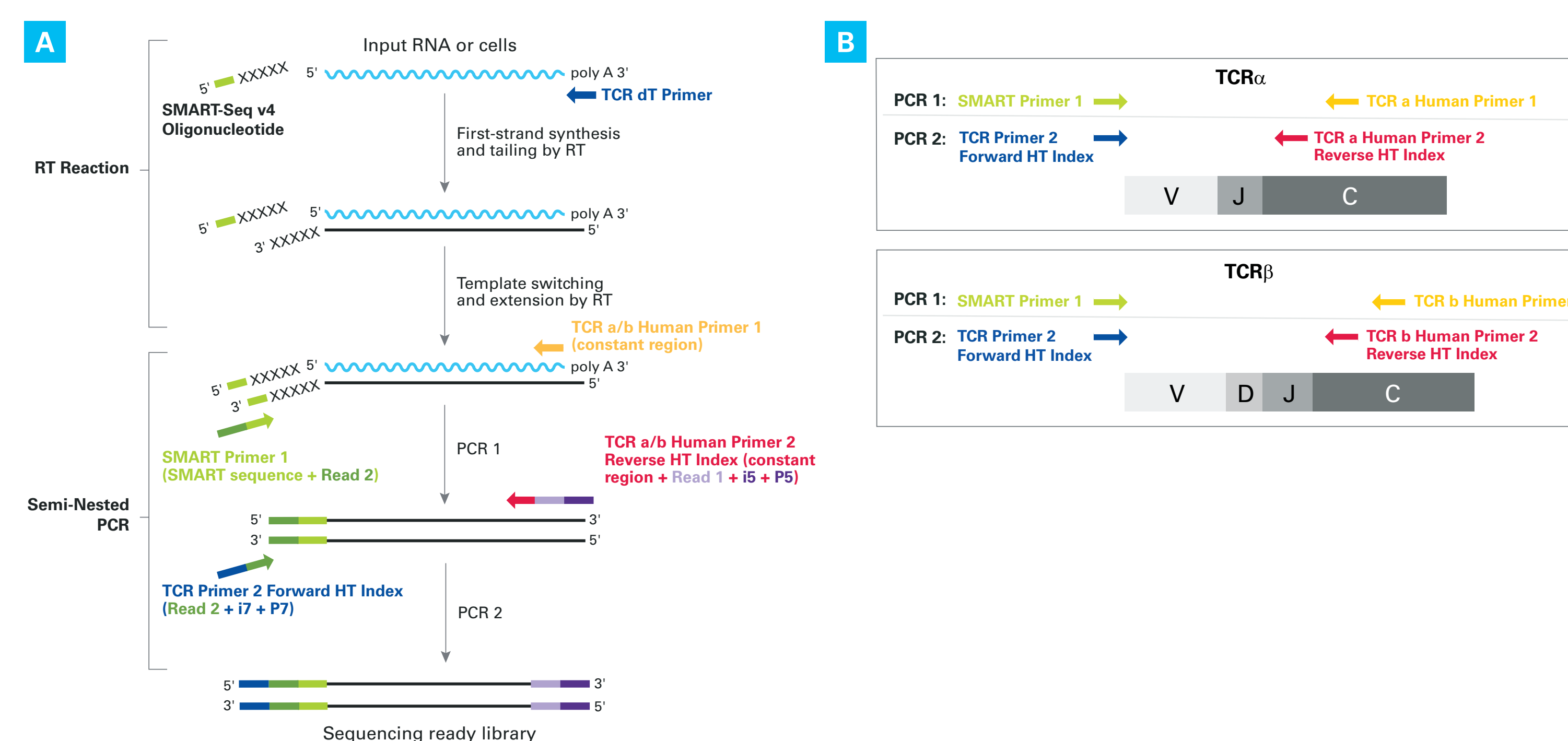
A SMARTer Approach to Profiling the Human T-Cell Receptor Repertoire

Rachel Fish, Sarah Taylor, Nao Yasuyama and Andrew Farmer¹

Corresponding Author: Andrew_Farmer@clontech.com

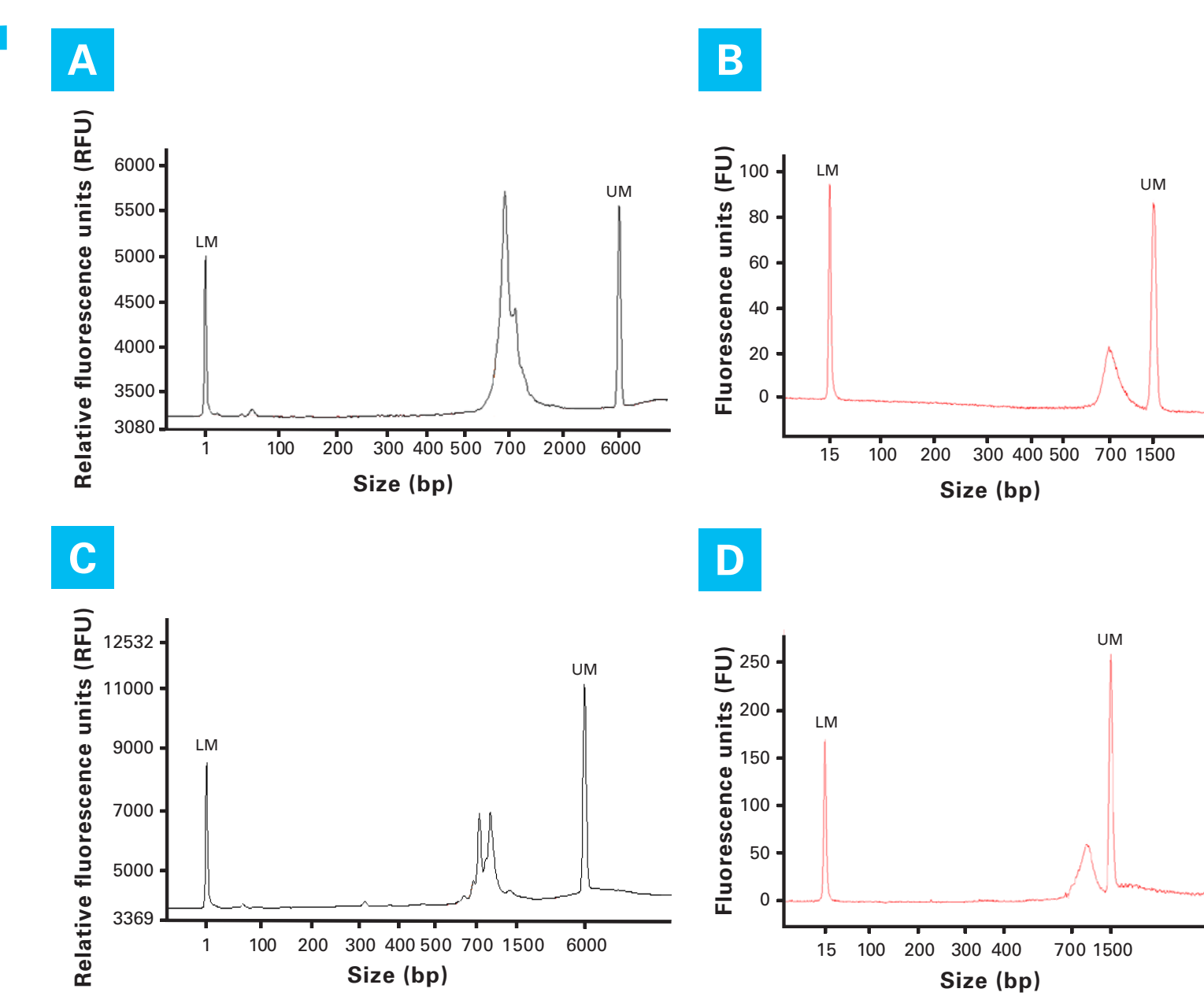
Methods and Data

1 TCR sequencing library preparation



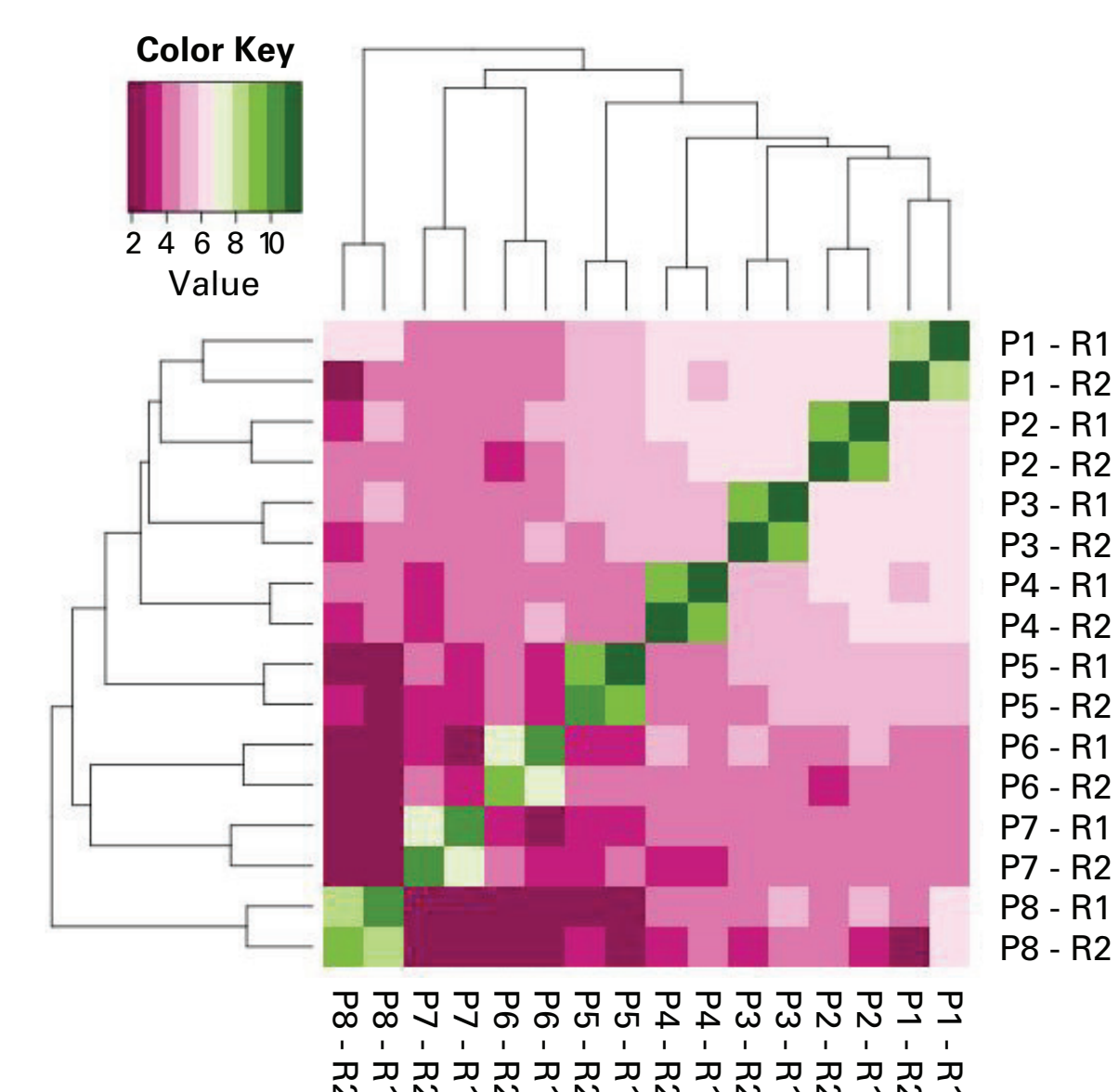
Library preparation workflow and PCR strategy for TCR profiling using the SMARTer[®] Human TCR a/b Profiling Kit. Panel A. Reverse transcription and PCR amplification of TCR subunit mRNA sequences. First-strand cDNA synthesis is primed by the TCR dT Primer and performed by a MMLV-derived reverse transcriptase (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. Full-length variable regions of TCR 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 TCR- α and/or TCR- β subunits (TCR a/b Human Primer 1). A subsequent round of PCR is performed to further amplify variable regions of TCR- α and/or TCR- β subunits and incorporate adapter sequences, using TCR Primer 2 and TCR a/b Human Primer 2. 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, TCR cDNA libraries are sequenced on the Illumina platform using 300 bp paired-end reads. Panel B. Semi-nested PCR approach for amplification of TCR- α and/or TCR- β subunits. The primer pairs used for the first round of amplification capture the entire variable region(s) and most of the constant region(s) of TCR- α and/or TCR- β cDNA. The second round of amplification retains the entire variable region(s) of TCR- α and/or TCR- β cDNA, and a smaller portion of the constant region(s). The anticipated size of final TCR library cDNA (inserts + adapters) is ~700–800bp.

2 Sequencing library validation and quality analysis



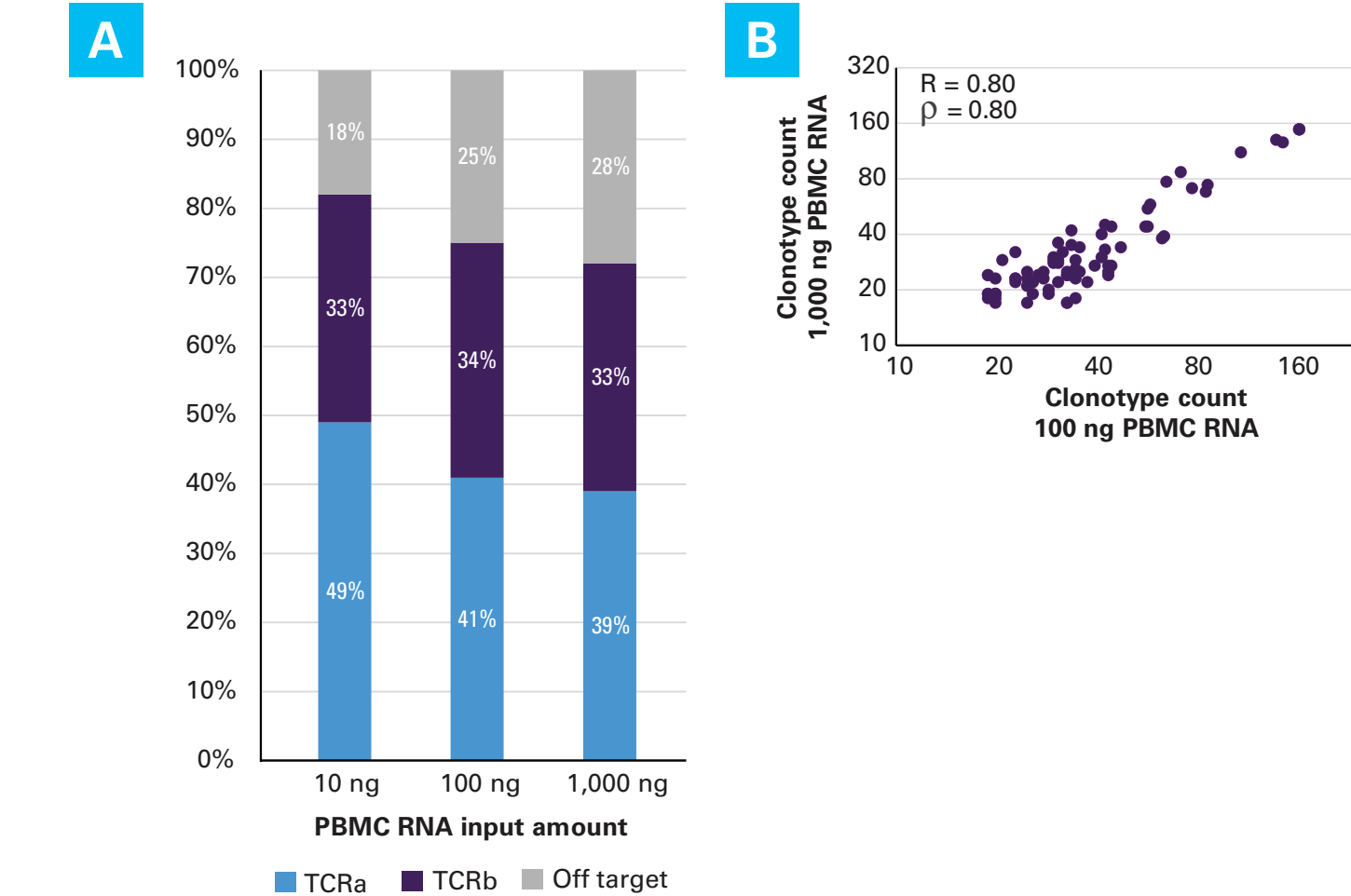
Electropherogram profiles of TCR sequencing libraries. Libraries containing both TCR- α and TCR- β sequences were generated using 10 ng of RNA obtained from either a heterogeneous population of peripheral blood leukocytes or a Jurkat cell line consisting of a single T-cell clonotype. Electropherogram profiles of the final libraries were obtained on both an Advanced Analytical Fragment Analyzer and 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 Fragment Analyzer profile of sequencing library for TCR- α and TCR- β , obtained from peripheral blood leukocyte RNA. Panel B. Typical Bioanalyzer profile of sequencing library for TCR- α and TCR- β , obtained from peripheral blood leukocyte RNA (same library as Panel A). The library profiles from the Fragment Analyzer and the Bioanalyzer both show a broad peak between ~650–1150 bp and a maximal peak in the range of ~700–800 bp for the library obtained from peripheral blood leukocyte RNA. Panel C. Typical Fragment Analyzer profile of sequencing library for TCR- α and TCR- β , obtained from monoclonal Jurkat T-cell RNA. The Fragment Analyzer profile for the library obtained from Jurkat RNA shows distinct peaks at approximately 700 bp and 800 bp, which correspond to predicted sizes of TCR- β and TCR- α sequence fragments, respectively. Panel D. Typical Bioanalyzer profile of sequencing library for TCR- α and TCR- β , obtained from Jurkat T-cell RNA (same library as Panel C). The Bioanalyzer profile for the library obtained from Jurkat RNA shows a peak similar in shape and position to the peaks shown in Panel A and Panel B.

3 Cross-sample clonotype comparisons



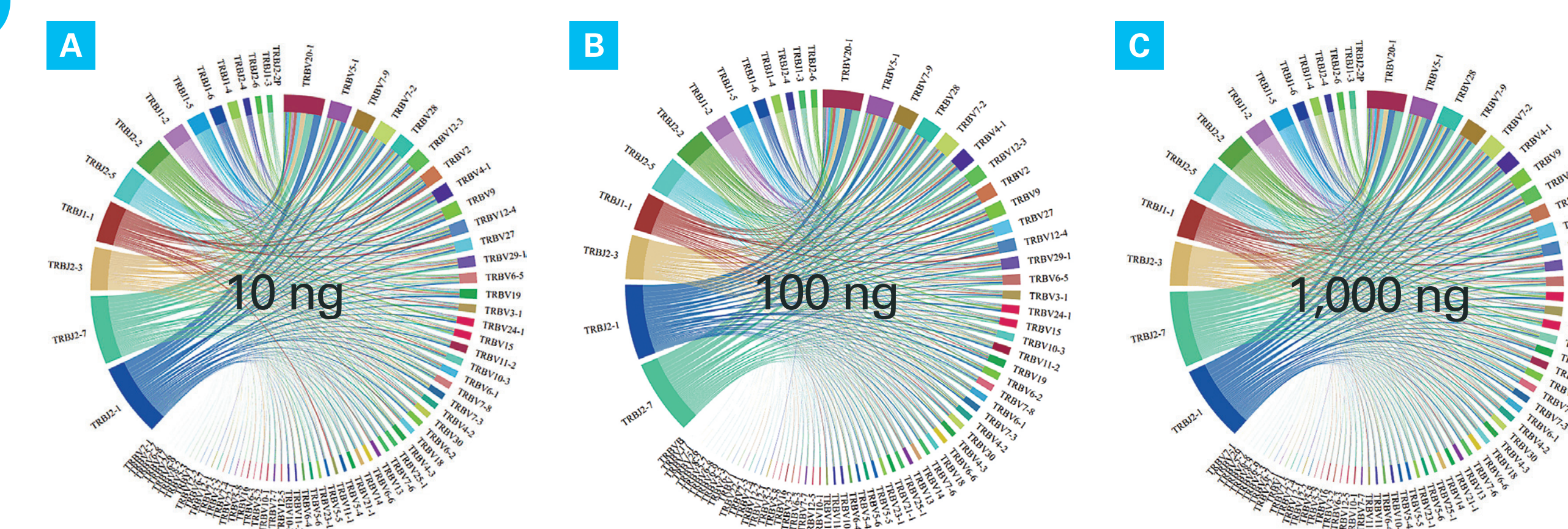
Heat map of cross-sample clonotype comparisons. The experimental protocol was performed on RNA extracted from replicate peripheral blood mononuclear cell (PBMC) samples (R1, R2) obtained from eight blood cancer patients (P1–P8). Prior to sequencing, libraries were purified and size selected using Solid Phase Reversible Immobilization (SPRI) beads, and analyzed as in Figure 2 (above). Libraries were sequenced on an Illumina MiSeq using 300 bp paired-end reads, which fully capture the TCR sequence included in each cDNA molecule. Resulting sequencing reads were mapped to CDR3 regions of TCR- α or TCR- β using an analysis platform provided by Distributed Bio, Inc. to identify clonotypes present in each sample. For each pairwise comparison indicated by the labels along the bottom and right-hand sides of the heat map, the tile color indicates the Log_2 -transformed value for the number of clonotypes that are common to both samples. Dendrograms on both axes indicate relative similarities between samples. Not surprisingly, the greatest level of overlap is observed between replicate PBMC samples from the same patient. In contrast, the level of overlap is dramatically lower for PBMC samples from different patients. These results speak to both the reproducibility and sensitivity of the SMARTer approach, in that it generates similar TCR profile data for replicate samples, and markedly divergent TCR profile data for samples obtained from different patients.

4 Reads on target and correlation of clonotype count data



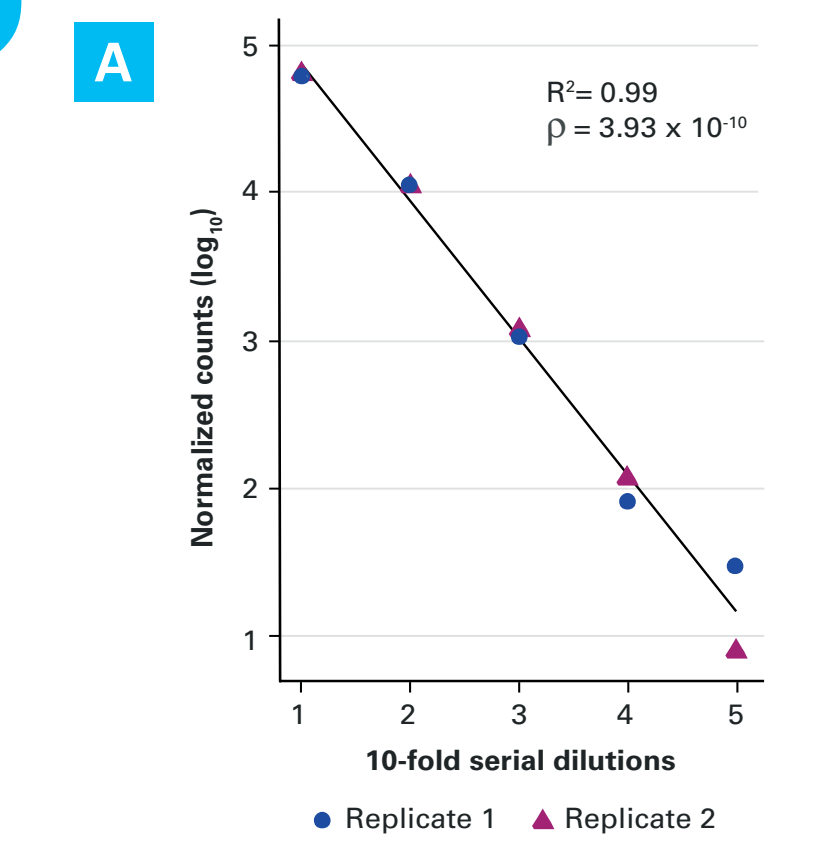
Sequencing reads on target and correlation of clonotype count data for varying sample input amounts. To evaluate the performance of the kit for a range of input amounts, the SMARTer workflow was performed on three different amounts of peripheral blood RNA (10 ng, 100 ng, and 1,000 ng) and the resulting cDNA libraries were sequenced as before. Sequencing outputs were downsampled to either ~260,000 or ~275,000 reads, depending on the analysis method, and processed using an application provided by Illumina via the BaseSpace website (Bolotin *et al.*, 2015). Panel A. Percentages of sequencing reads that map to CDR3 regions in either TCR- α (blue) or TCR- β (purple) or that represent off-target reads (gray) for each indicated sample input amount. For each sample amount analyzed, ~270% of sequencing reads mapped to a CDR3 region in either TCR- α or TCR- β , with the 10 ng sample amount yielding the highest percentage of on-target reads. These results demonstrate that the SMARTer approach can capture and amplify TCR sequences from total RNA with considerable specificity across a wide range of sample input amounts. Panel B. Correlation of clonotype count data for 100 ng input RNA vs. 1,000 ng input RNA. Comparison of clonotype count data yielded a Pearson correlation coefficient (R) of 0.80 and a Spearman coefficient (ρ) of 0.88, a result that speaks to the robustness of the SMARTer approach for input sample amounts that vary by at least one order of magnitude.

5 Visual representation of clonotype distributions



Chord diagrams of TCR- β clonotype distributions observed for varying sample input amounts. The distribution of TCR clonotypes identified in the sequencing data can also be depicted visually using chord diagrams. Each chord diagram depicts the distribution of the indicated TCR- β Variable-Joining (V-J) segment combinations for the indicated RNA input amount. Each arc (along the circumference of each diagram) represents a V or J segment and is scaled lengthwise according to the relative proportion at which the segment is represented in the dataset. Each chord (connecting the arcs) represents a set of clonotypes which include the indicated V-J combination, and is weighted according to the relative abundance of that combination in the dataset. Panel A. Chord diagram for 10 ng input of PBMC RNA. Panel B. Chord diagram for 100 ng input of PBMC RNA. Panel C. Chord diagram for 1,000 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.

6 Assessing sensitivity and reproducibility



| Amount of Jurkat RNA | % spike-in (% Jurkat RNA in 10 ng PBMC RNA) | # TRBV12-3-TRBJ1-2 clonotypes identified | Signal:noise Ratio (x spike-in/control (%)) | Two-tailed, Student's t-test (p-value) | p < 0.005? | p < 0.001? |
|----------------------|---|--|---|--|------------|------------|
| 0 | 0.000% | 619 | 480 | 1.0040 | — | — |
| 100 fg | 0.001% | 649 | 488 | 1.0346 | 0.8798 | False |
| 1 pg | 0.010% | 702 | 597 | 1.1820 | 0.36972 | False |
| 10 pg | 0.100% | 1,695 | 1,673 | 3.0646 | 0.00382 | True |
| 100 pg | 1.000% | 12,074 | 11,206 | 2.11829 | 0.00157 | True |
| 1 ng | 10.000% | 62,395 | 63,006 | 114.1592 | 0.00003 | True |

Assessing the sensitivity and reproducibility of the SMARTer approach. In order to assess the sensitivity of the SMARTer approach, the protocol was performed in replicate on PBMC RNA samples spiked at varying concentrations (10%, 1%, 0.1%, 0.01%, and 0.001%) with RNA obtained from a homogenous population of Leukemic Jurkat T cells (TRAV8-4-TRAJ3, TRBV12-3-TRBJ1-2 clonotype). Panel A. Correlation between concentration of spiked-in Jurkat RNA and number of TRBV12-3-TRBJ1-2-specific sequence reads. Numbers along the X-axis indicate serial-diluted concentrations of spiked-in Jurkat RNA (by mass): 1 = 10%; 2 = 1%; 3 = 0.1%; 4 = 0.01%; 5 = 0.001%. Count data for TRBV12-3-TRBJ1-2-specific sequence reads were normalized by subtracting the number of corresponding reads obtained for negative control samples consisting of unspiked PBMC RNA. Normalized count data were then Log_2 transformed. Circles and triangles correspond to experimental replicates for each sample concentration. Linear regression analysis revealed a statistically significant correlation ($p = 3.93 \times 10^{-11}$; $R^2 = 0.99$) between the amount of spiked-in Jurkat RNA and the number of TRBV12-3-TRBJ1-2-specific sequence reads. These results demonstrate that differences in the relative abundance of transcripts for a particular TCR clonotype are faithfully and reproducibly represented in sequencing libraries generated using the SMARTer approach. Panel B. Count data, signal-to-noise ratios, and statistical analysis for TRBV12-3-TRBJ1-2-specific sequence reads obtained from spiked RNA samples. Signal-to-noise ratios were generated using the mean counts of TRBV12-3-TRBJ1-2-specific sequence reads for each pair of experimental replicates. Rows highlighted in yellow include concentrations of spiked-in Jurkat RNA for which statistically significant elevations in TRBV12-3-TRBJ1-2-specific sequence reads were detected relative to background counts observed for unspiked negative control RNA samples. Added Jurkat RNA at a concentration of 0.1% was detectable above background in the sequencing output ($p < 0.005$) at a depth of ~275,000 reads, evidence of the sensitivity afforded by the SMARTer approach.

Conclusions

- Our SMARTer Human TCR a/b Profiling Kit provides a streamlined workflow for the generation of Illumina-ready TCR sequencing libraries that minimizes the likelihood of amplification biases by avoiding multiplex PCR
- The SMARTer approach captures entire TCR variable regions and allows for analysis of both α and β chains in the same experiment
- The kit performs reliably for a range of total RNA or T cell input amounts, and the data generated consists largely of on-target reads
- The sensitivity afforded by the SMARTer approach is such that clonotype-specific TCR RNA added at a concentration of 0.1% is detectable above background and reproducibly represented in sequencing libraries produced with the kit

1290 Terra Bella Ave., Mountain View, CA 94043
Orders and Customer/Technical Service: 800.662.2566
Visit us at www.clontech.com

For Research Use Only. Not for use in diagnostic or therapeutic procedures. Not for resale. Illumina[®] and MiSeq[®] are registered trademarks of Illumina, Inc. Takara[®] and the Takara logo are trademarks of Takara Holdings, Inc. Clontech[®], the Clontech logo, SMART[®], SMARTer[®] and SMART-Seq[™] are trademarks of Clontech Laboratories, Inc. All other trademarks are the property of their respective owners. Certain trademarks may not be registered in all jurisdictions. ©2016 Clontech Laboratories, Inc.

