P2233 **Clontech Laboratories, Inc.**
A Takara Bio Company

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

Sarah Taylor, Nao Yasuyama and Andrew Farmer[1]

Clontech Laboratories, Inc., 1290 Terra Bella Ave., Mountain View, CA 94043

[1]Corresponding Author: Andrew_Farmer@clontech.com

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. While low-throughput approaches have yielded important insights concerning TCR repertoire dynamics, development of next-generation sequencing (NGS) technologies has dramatically expanded research prospects.

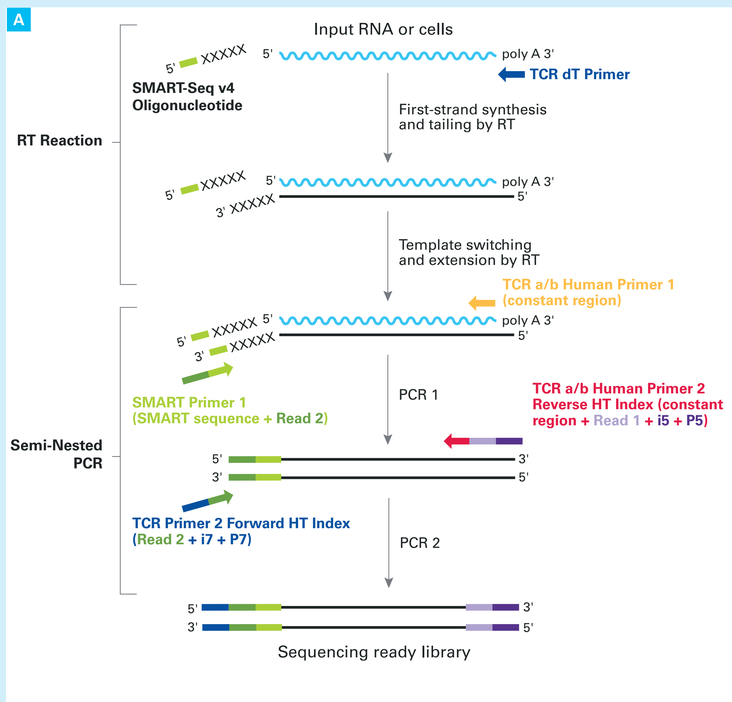
Using SMART® (**S**witching **M**echanism **a**t 5′ End of **R**NA **T**emplate) technology, we have developed an NGS library preparation kit for TCR profiling that employs a 5′ RACE-like approach to capture full-length variable regions of TCR-α and/or TCR-β subunits. With this approach, 10 ng–3 µg of input RNA or 50–10,000 purified T cells undergo reverse transcription 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 peripheral blood RNA, libraries containing TCR-α and TCR-β sequences were generated and sequenced. For each replicate >70% of reads were on-target, and the most highly represented clonotypes remained consistent across a range of input amounts. A separate sensitivity assay demonstrated that 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 RNA in ~2.5 hours of hands-on time. Whereas approaches that utilize genomic DNA as starting material require multiplexed PCR strategies, RNA-based amplification of TCR sequences uses single primer pairs for each subunit. In this way, our approach minimizes the likelihood of sample misrepresentation due to amplification biases.

Introduction and Methods

TCR sequencing library preparation

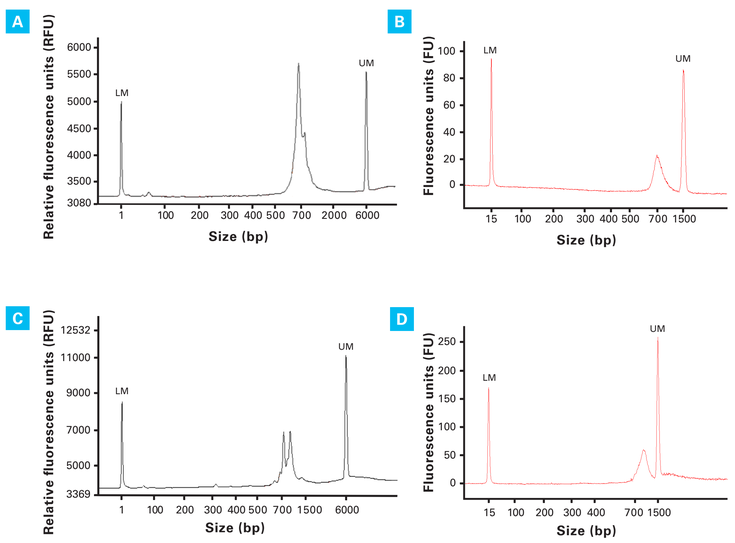


Library preparation workflow and PCR strategy for TCR profiling using the SMARTer® Human TCR a/b Profiling Kit. Panel A. Reverse transcription and PCR amplification of TCR subunit mRNA sequences. First-strand cDNA synthesis is primed by the TCR dT Primer and performed by an MMLV-derived reverse transcriptase (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. 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 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 ~700–800 bp.

References

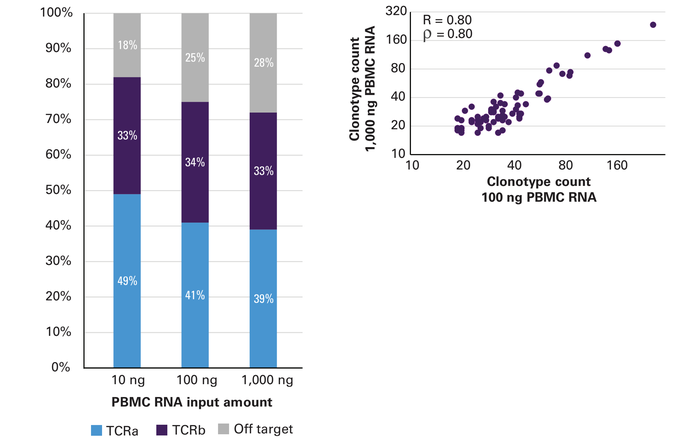
Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., & Chudakov, D. M. (2015) MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**(5):380–381.

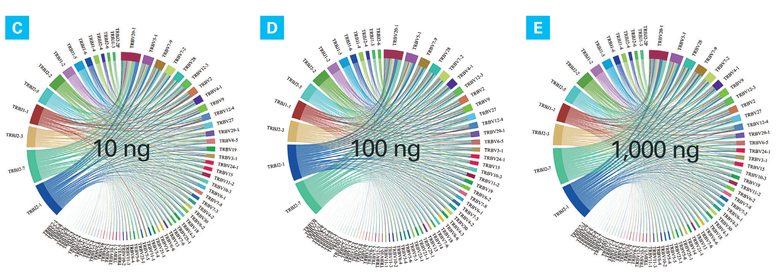
1 Sequencing library validation and quality analysis



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.

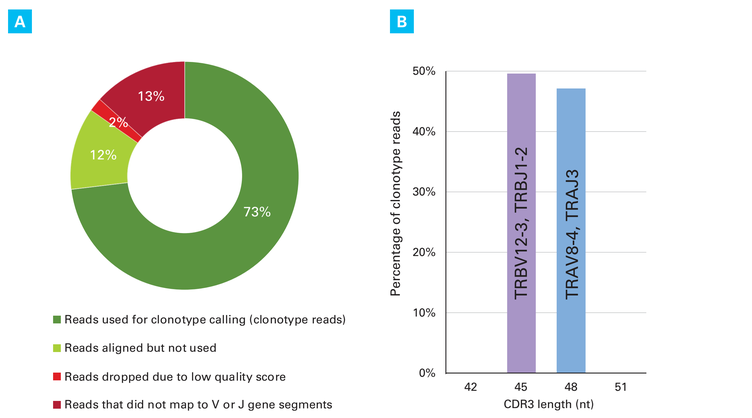
2 Reads on target, correlation of clonotype count data, and TCR-β clonotype distributions





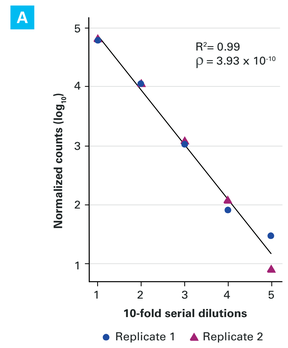
Sequencing reads on target, correlation of clonotype count data, and chord diagrams of TCR-β clonotype distributions observed for varying sample input amounts. To evaluate the performance of the kit for a range of input amounts, the SMARTer workflow was performed on three different amounts of peripheral blood RNA (10 ng, 100 ng, and 1,000 ng) and the resulting cDNA libraries were sequenced. Sequencing outputs were downsampled to ~260,000 reads, 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) for each indicated sample input amount. For each sample amount analyzed, >70% of sequencing reads mapped to a CDR3 region in either TCR-α or TCR-β. 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 (R) of 0.80, a result that speaks to the robustness of the SMARTer approach for input sample amounts that vary by at least one order of magnitude. **Panels C–E.** The distribution of TCR-β clonotypes identified in the sequencing data depicted by chord diagrams. Each arc (on the periphery of each diagram) represents a V 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 which include the indicated V-J combination, and is weighted according to the relative abundance of that combination in the dataset. **Panel C.** Chord diagram for 10 ng input of PBMC RNA. **Panel D.** Chord diagram for 100 ng input of PBMC RNA. **Panel E.** 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.

3 Sequencing reads on target and spectratype analysis for libraries generated directly from cells



Sequencing reads on target and spectratype analysis for cell inputs. To evaluate the performance of the kit using cells as the input material, the SMARTer workflow was performed on Jurkat cells and the resulting cDNA libraries were sequenced and processed using an application provided by Illumina via the BaseSpace website (Bolotin *et al.*, 2015). **Panel A.** The percentage of sequencing reads that map to CDR3 regions in TCR-α or TCR-β. While 85% of the reads align to TCR sequences, 73% of the reads are used in the downstream clonotype calling by the software. Two percent of the reads were dropped due to low quality scores and 13% did not map to either V or J gene segments. **Panel B.** The histogram shows the distribution of reads that were mapped to clonotypes by CDR3 length. Two distinct CDR3 lengths which correspond to the correct Jurkat clonotype were identified. Of the reads that mapped to clonotypes, 49.6% had a CDR3 length of 45 nucleotides and were identified as the TRBV12-3, TRBJ1-2 clonotype, and 47.1% had a CDR3 length of 48 nucleotides and were identified as being the TRAV8-4, TRAJ3 clonotype. These data attest to the robustness of the SMARTer approach when working with inputs consisting of intact cells.

4 Assessing sensitivity and reproducibility



B

Amount of Jurkat RNA	% spike-in (% Jurkat RNA in 10 ng PBMC RNA)	# TRBV12-3-TRBJ1-2 clonotypes identified		Signal:noise	Two-tailed, Student's t-test (p value)	p<0.005?	p<0.001?
		Replicate 1	Replicate 2	Ratio (x̄ spike-in/x̄ control (0%))			
0	0.000%	619	480	1.0000	—	—	—
100 fg	0.001%	649	488	1.0346	0.87498	False	False
1 pg	0.010%	702	597	1.1820	0.36972	False	False
10 pg	0.100%	1,695	1,673	3.0646	0.00382	True	False
100 pg	1.000%	12,074	11,206	21.1829	0.00157	True	False
1 ng	10.000%	62,395	63,006	114.1592	0.00003	True	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_{10} transformed. Circles and triangles correspond to experimental replicates for each sample concentration. Linear regression analysis revealed a statistically significant correlation ($p = 3.93 \times 10^{-10}$, $R^2 = 0.99$) 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 more information, please visit
www.clontech.com/SMARTer-Human-TCR-AAI-2016