Single-cell T-cell receptor profiling with SMART technology

Workflows for NGS library preparation from single cells isolated in 96-well plates or using the WaferGen ICELL8 Single-Cell System

Sarah Taylor[1*], Thomas Schaal[2], Nao Yasuyama[1], Jude Dunne[2], Maithreyan Srinivasan[2], and Andrew Farmer[1]

[1] Takara Bio USA, Inc. (formerly Clontech Laboratories, Inc.), [2] WaferGen Bio-Systems, Inc. * Presenting author

Abstract

Profiling T-cell receptor (TCR) diversity is critical for understanding the adaptive immune system and can provide valuable insights in studies involving immuno-oncology, immune deficiency, autoimmunity, and vaccine response. While the development of next-generation sequencing (NGS)-based methods for TCR profiling has dramatically expanded our understanding of immune repertoire dynamics, the majority of these approaches lack single-cell resolution and are limited to analysis of the TCR-β CDR3 region as a proxy for overall TCR diversity.

Using SMART® technology, we previously developed an NGS library-preparation method that employs a 5′ RACE-like approach to capture full-length variable regions of TCR-β and/or TCR-β subunits. In contrast with TCR profiling methods that use genomic DNA as starting material, our SMARTer® approach allows for analysis of expressed TCR sequences that are likely to encode functional receptors and utilizes single primer pairs for each subunit, thereby avoiding amplification biases associated with multiplex PCR.

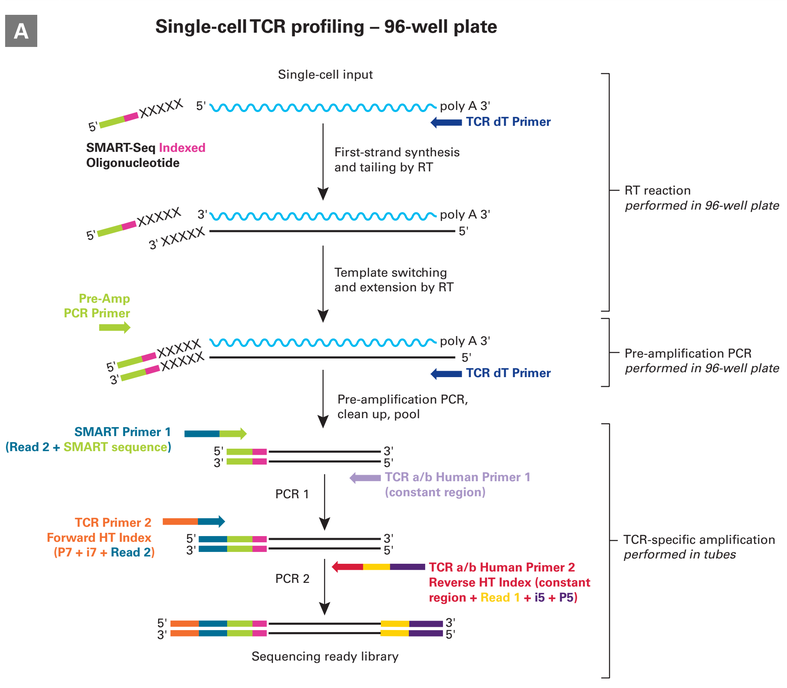
Here we present two high-throughput NGS library-preparation workflows based on our SMARTer approach—one for cells sorted into a 96-well plate, and another for cells isolated using the WaferGen ICELL8™ Single-Cell System—that provide full-length VDJ sequence information for both TCR subunits from single cells. Either of these approaches, in combination with phenotyping, enables researchers to identify specific pairings of alpha and beta chains that comprise functional receptors in individual cells, and provides a starting point for classifying these individual T cells on the basis of function, maturity, and other complex parameters such as the timing of cytokine secretion.

Our TCR profiling workflow for 96-well plates is ideal for smaller-scale studies involving cells that have been isolated by methods such as FACS. In preliminary experiments involving analysis of Jurkat cells with this approach, 92% of sequencing reads mapped to TCR sequences on average, while an average of 90% of reads identified the correct Jurkat clonotype.

Our workflow for the WaferGen ICELL8 system is intended for higher-throughput applications, and can generate sequencing libraries for ~1,000 cells at a time. These libraries can be pooled together and analyzed on a single MiSeq® run. Analysis of individual Jurkat cells using this approach identified >60% of reads mapping to TCR sequences in most cells, with ~70–80% of these reads being used in clonotype identification.

1 TCR sequencing library preparation from single cells sorted into a 96-well plate

A



B

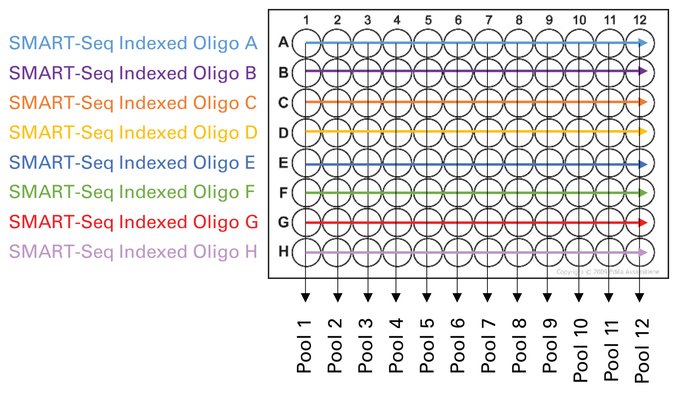


Figure 1. SMARTer library preparation workflow and pooling strategy for single-cell TCR profiling using a 96-well format. Panel A. First-strand cDNA synthesis is primed by the TCR dT Primer and performed by an MMLV-derived reverse transcriptase (RT). Upon reaching the 5′ end of each mRNA molecule, the RT adds non-templated nucleotides to the first-strand cDNA. The SMART-Seq® Indexed Oligonucleotide includes a sequence that is complementary to the non-templated nucleotides added by the RT in addition to a row-specific index sequence (pink), and hybridizes to the first-strand cDNA. In the template-switching step, the RT uses the remainder of the SMART-Seq Indexed Oligonucleotide as a template for the incorporation of an additional sequence on the end of the first-strand cDNA (green). A pre-amplification PCR step is then performed in each well to generate double-stranded cDNA, which will serve as starting material for PCR amplification and library construction. Following a cleanup step, cDNA from each column of the 96-well plate are pooled in separate tubes (see Panel B), and TCR-specific sequences are 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 Forward HT Index and TCR a/b Human Primer 2 Reverse HT Index. 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.** Single cells are sorted into lysis buffer in individual wells of a 96-well plate. Reverse transcription reagents are then added to the plate, including a uniquely-indexed SMART-Seq oligonucleotide for each row (A–H). RT and pre-amplification PCR steps are performed in each well of the plate, and products are pooled by column, yielding 12 pools that each contain a mixture of cDNA molecules that are barcoded according to the individual cells from which they were derived. Subsequent library construction steps are performed on each of the 12 pools, reducing the number of libraries that need to be prepared from 96 to 12. Inclusion of the barcodes allows for demultiplexing of sequencing data for each pool, thereby allowing for single-cell resolution and pairing of sequence information for TCR-α and TCR-β subunits.

2 Validation and quality analysis of libraries generated from single cells and equivalent amounts of RNA

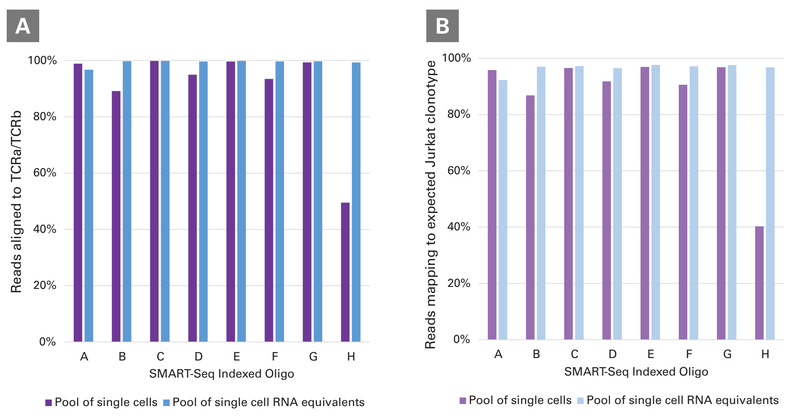


Figure 2. Sequencing reads on target and reads that align to the correct Jurkat clonotype. To assess the performance of the workflow, libraries containing both TCR-α and TCR-β sequences were generated either from single Jurkat cells or from single cell-equivalent amounts of Jurkat RNA (2.5 pg of RNA). Eight cells and eight RNA samples were processed individually in a 96-well plate, using a different SMART-Seq Indexed Oligo for each input of the same type. RT and pre-amplification PCR reactions were performed in each well, and cDNA products derived from inputs of the same type were pooled together prior to subsequent rounds of PCR, as outlined in Figure 1. Final libraries were sequenced on a MiSeq with 300 bp paired-end reads. Following sequencing, barcode sequences derived from the SMART-Seq Indexed Oligos were used to demultiplex the sequencing data, which was then analyzed using MiXCR (Bolotin *et al.*, 2015). **Panel A.** The percentages of sequencing reads that map to CDR3 regions in TCR-α or TCR-β from each cell or RNA sample. For each RNA sample, >96% of reads mapped to either TCR subunit. For seven out of the eight cells analyzed, >89% of reads mapped to either TCR subunit. **Panel B.** The percentages of sequencing reads that map to the expected Jurkat clonotype (TRAV8-4, TRAJ3/TRBV12-3, TRBJ1-2) from each cell or RNA sample. For each RNA sample, >92% of reads identified the correct Jurkat clonotype. For seven out of the eight cells analyzed, >86% of reads identified the correct Jurkat clonotype, in agreement with the read-alignment data.

3 Overview of the WaferGen ICELL8 Single-Cell System

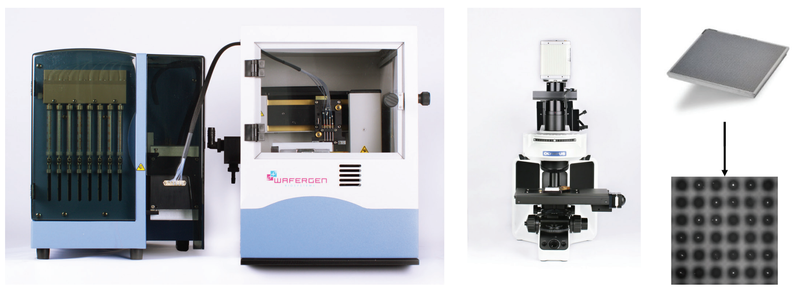


Figure 3. Overview of the WaferGen ICELL8 System: ICELL8 System Components – Multi-Sample Nano Dispenser (MSND) dispenses cells and reagents in nanoliter volumes into the wells of the ICELL8 chip, Imaging Station captures images of fluorescently labeled cells in the ICELL8 chip, 36 wells at a time (representative image shown), and CellSelect™ software analyzes images and identifies single cell-containing wells for further processing.

4 TCR sequencing library preparation from single cells isolated with the WaferGen ICELL8 system

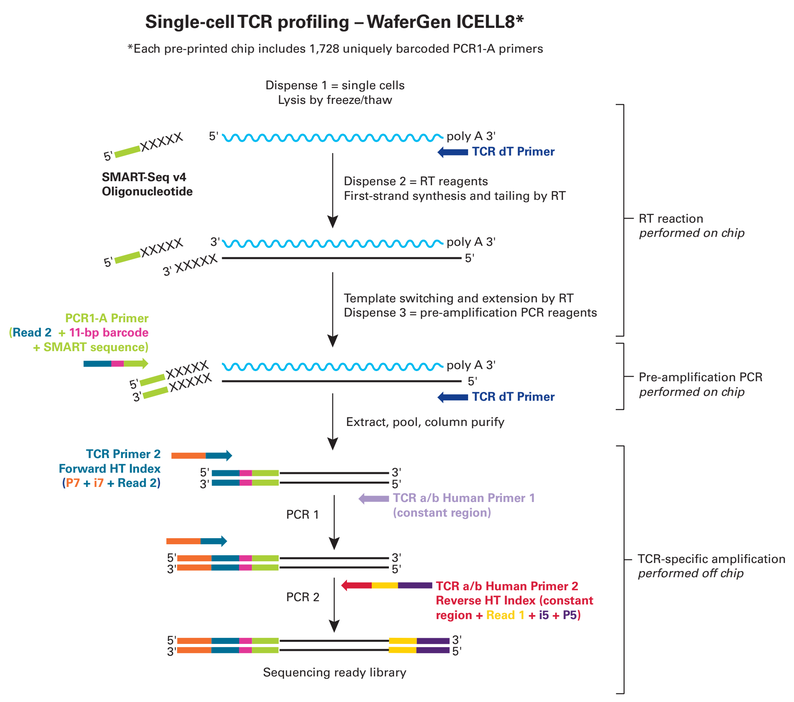


Figure 4. SMARTer library preparation workflow for single-cell TCR profiling using the ICELL8 System. On-chip reactions: Dispense #1 (single-cell solution): T cells are dispensed using MSND into WaferGen 72 x 72 chips with (or without) pre-printed, barcoded PCR primers, using methods designed to maximize the single-cell yield as dictated by Poisson statistics. Automated imaging of the cells is performed using CellSelect software. Single cell-containing wells are down-selected so each barcode is used only once. There are three copies of each barcode (n=1,728) on the chip that provide well-specific addresses. Cells are lysed on-chip by freeze-thaw and immediately processed. **Dispense #2 (RT mix):** First-strand cDNA synthesis is primed by the TCR dT Primer and performed by an MMLV-derived reverse transcriptase (RT) in the presence of cell lysis buffer. 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 template-switching oligo as a template for the incorporation of an additional sequence on the end of the first-strand cDNA. **Dispense #3 (pre-amp mix):** Ten cycles of pre-amplification are performed to incorporate the pre-printed, PCR1-A Primer (used in concert with the TCR dT Primer). The contents of the chip are extracted by centrifugation using a fixture, followed by column purification. **In-tube PCR amplification reactions:** Full-length variable regions of TCR cDNA are selectively amplified by PCR using indexed primers that are complementary to the oligonucleotide-templated sequence (TCR primer 2 Forward HT Index), 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 Forward HT Index and TCR a/b Human Primer 2 Reverse HT Index. 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.

5 Sequencing metrics, read mapping, and clonotype calling for libraries generated from single cells on the ICELL8 system

	Single barcode (1)	Single barcode (2)	1,728 barcodes
Number of single-cell wells (Jurkat cells)	1,471	1,471	824
Number of control wells (RNA)	6	48	10 + 10
Total sequencing reads	734,613	1,130,907	4,168,990
Successfully aligned reads	614,886	778,300	2,790,532
Successfully aligned reads (percent)	**83.7%**	**68.8%**	**66.9%**
Number of reads used in clonotype calling	594,173	716,531	2,663,900
Reads used (percent of total)	**80.9%**	**63.4%**	**63.9%**
High-quality reads (percent of reads used)	69.9%	72.0%	61.8%
Low-quality reads (percent of reads used)	30.1%	28.0%	38.2%
Reads mapping to Jurkat clonotype:			
TRAV8-4, TRAJ3	50.3%	80.9%	35.2%
TRBV12-3, TRBJ1-2	49.7%	13.7%	64.6%
TOTAL	**99.9%**	**94.6%**	**99.8%**

Table 1. Sequencing data analysis for three independent library preparation runs using the ICELL8 System. To evaluate the performance of the SMARTer workflow on the ICELL8, the protocol was initially performed on Jurkat cells and RNA control samples, and libraries were generated using a single barcode ((1) and (2)). Validation was then performed on pre-printed chips containing 1,728 barcodes. The resulting cDNA libraries were sequenced and then analyzed using MiXCR (Bolotin *et al.*, 2015).

In our validation experiments where a single barcode was used, 1,471 Jurkat cells and either 6 or 48 PBMC RNA control wells were included in the final library. Both of these experiments gave good numbers of reads mapping to CDR3 regions in TCR-α or TCR-β (~84% and ~69% for run 1 (1) and run 2 (2), respectively), with the vast majority of these reads being used for clonotype calling. In (1), over 99% of the reads used for clonotype calling identified the correct Jurkat clonotype (TRAV8-4,TRAJ3/TRBV12-3,TRBJ1-2). In (2), where more PBMC RNA control samples were used, this number was ~95% (the remaining ~5% of reads identified alternative clonotypes present in the PBMC RNA controls).

We then tested the workflow on ICELL8 chips pre-printed with 1,728 barcodes. In this experiment, 824 Jurkat cells, 10 Jurkat RNA controls, and 10 PBMC controls were included in the final library. In this case, ~67% of reads mapped to CDR3 regions in TCR-α or TCR-β. Over 99% of the reads used for clonotype calling identified the correct Jurkat clonotype (TRAV8-4,TRAJ3/TRBV12-3,TRBJ1-2), with good representation of both the alpha and beta chains.

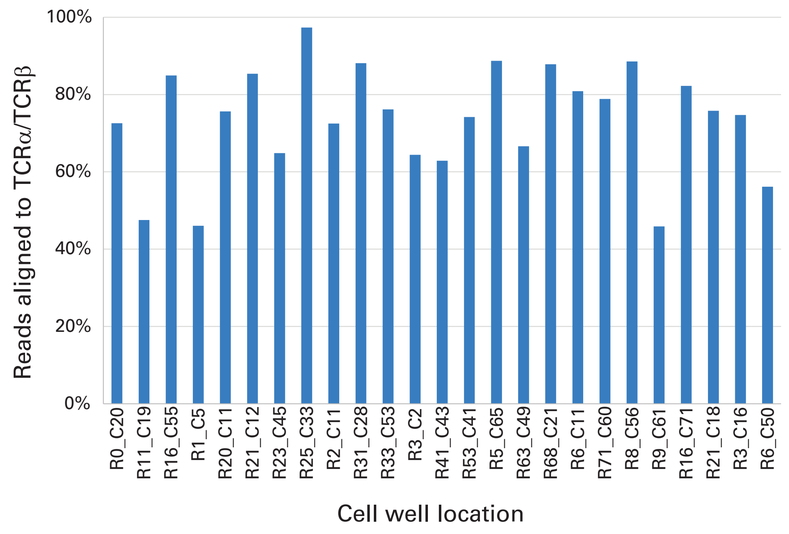


Figure 5. Alignment of sequencing reads from single Jurkat cells. Sequencing data for 25 randomly selected cells included in the experiment described for Table 1 was analyzed to determine the percentages of sequencing reads that map to CDR3 regions in TCR-α or TCR-β. "R" and "C" refer to row and column positions, respectively, for each cell. In the majority of cells (21/25), >60% of reads mapped to TCR-α or TCR-β sequences. The correct Jurkat clonotype (TRAV8-4,TRAJ3/TRBV12-3,TRBJ1-2) was identified for all cells. Data from control wells yielded a similar range of alignment rates.

Conclusions

- Paired identification of TCR-α and TCR-β chains in single cells, and matching of these pairings with their cognate antigens (neoantigens) is an important analytical step in guiding immunotherapy strategies.
- Our SMARTer method provides a streamlined workflow for the generation of Illumina-ready TCR sequencing libraries that capture full-length sequence information for TCR-α and TCR-β variable regions in a highly reproducible manner.
- By coupling well-specific barcoding of sequencing libraries generated from single cells with pool- or chip-specific indexing of TCR α/β amplicons, we have adapted our SMARTer method to enable high-throughput resolution of TCR-α + TCR-β pairings from single cells.
- Our SMARTer method can be applied effectively on the WaferGen ICELL8 system or in a 96-well format, providing solutions for a variety of applications involving large-scale studies of T-cell receptor diversity.

References

1. 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.

Scan to download your copy of this poster, or visit http://info.clontech.com/single-cell-TCR.html

Visit us at www.clontech.com/ngs or www.wafergen.com/products/icell8-single-cell-system

09.16 US (633837)