TECH NOTE

SMARTer T-cell receptor profiling in single cells

Flexible workflow: Illumina-ready libraries from FACS or manually sorted single cells

Ease of use: Optimized indexing allows for pooling 96 cells into 12 libraries which can be further multiplexed for running in a single flow-cell lane

Sensitivity: RACE-based approach allows for the detection of low-abundance TCR variants

Specificity: Full-length reads, with a majority of reads on target and accurate pairing information

Introduction

T-cell receptor (TCR) profiling of bulk samples has improved our understanding of the TCR repertoire diversity, TCR-mediated antigen specificity, and mechanisms of adaptive immune response (Becattini et al. 2015). However, while bulk studies can indicate which clonotypes are expressed and their relative frequencies in the population, it is nearly impossible to determine the proper alpha-beta (αβ) pairing of specific receptor chains (Figure 1A) of the cells—with the exception of some very rare cell populations (Figure 1B). The lack of pairing information makes it difficult for these studies to elucidate how particular chain pairings contribute to antigenic specificity (Stubbington et al. 2016). Sequencing single T cells allows for the determination of αβ pairings on TCRs thus providing context for antigen specificity and insights for the efficient design of TCRs for targeted immunotherapy, and helping establish ancestral relationships to pinpoint the clonality of cells.
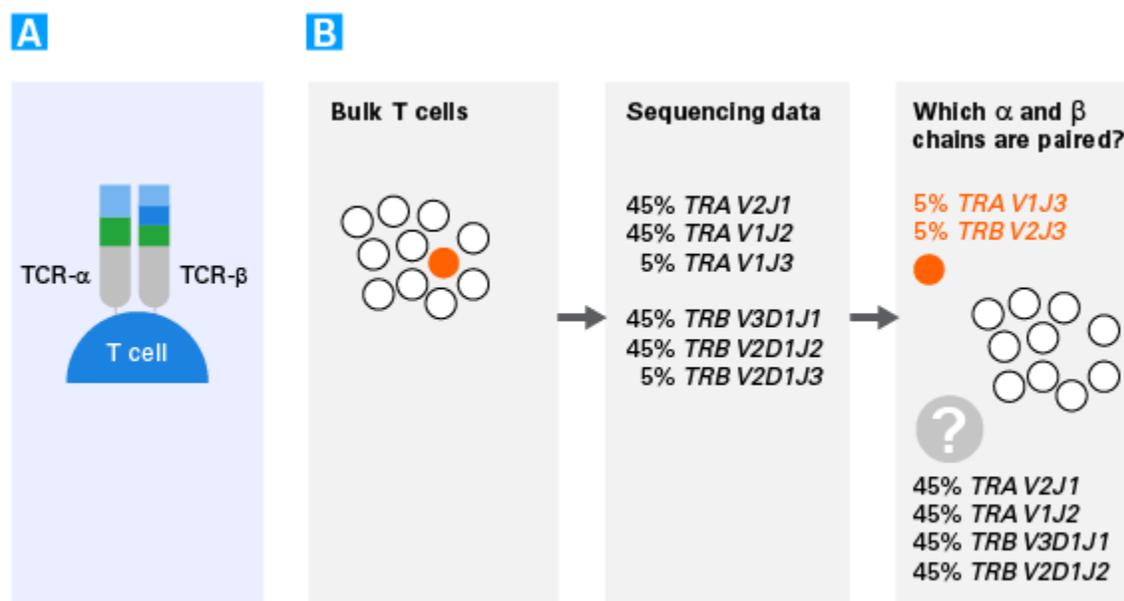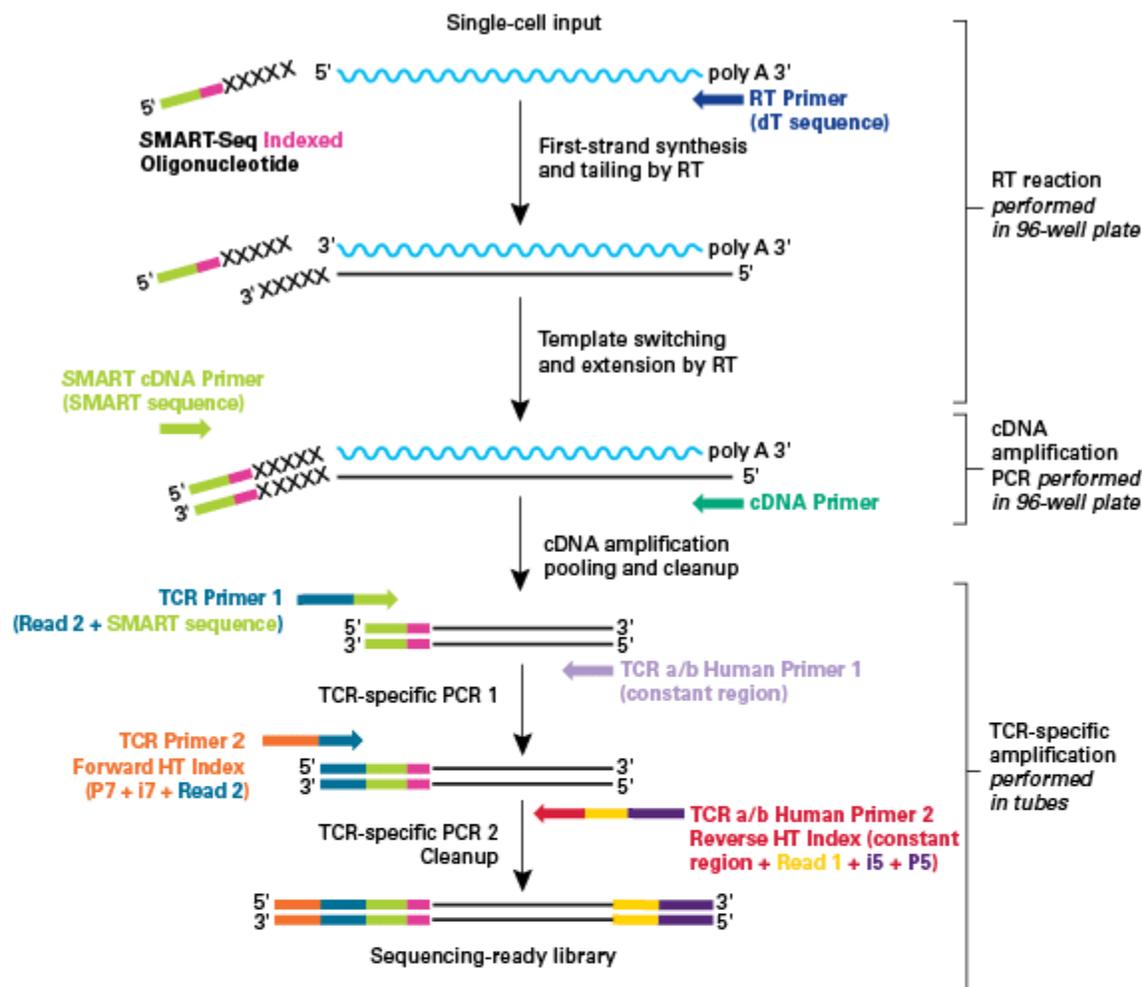

Figure 1. Benefits of studying T cells at the single-cell level. Panel A Schematic of a T cell receptor comprised of an alpha chain (TCR-α) and a beta chain (TCR-β). **Panel B.** Schematic representing the difficulties of obtaining pairing information for alpha and beta chains from bulk sequencing data. Rare clonotypes can allow for assessing the pairing of some cells (clonotypes in orange). However, for the more highly expressed clonotypes (TCRA V2J1, TCRA V2J2, TCRB V3D1J1, TCRB V2D1J2) there are four possible pairing combinations.

While single-cell sequencing is a powerful approach, most methods require sequencing of a large number of cells, making the experiments prohibitively expensive and the analyses difficult. The SMARTer Human scTCR a/b Profiling Kit (referred to below as the "human scTCR kit") tackles this problem by combining SMART cDNA synthesis and a RACE-based gene-specific priming followed by TCR-specific PCR to fully capture and amplify TCR-α and TCR-β variable regions and generate Illumina-ready libraries that provide a highly sensitive approach to sequencing TCRs (Figure 2A).

A

Single-cell TCR profiling – 96-well plate

Single-cell input

5' ~~~~~~~~~~~~~~~~~~~~~ poly A 3'

◄── **RT Primer
(dT sequence)**

**SMART-Seq Indexed
Oligonucleotide**

First-strand synthesis
and tailing by RT

| RT reaction
*performed
in 96-well plate*

3' ~~~~~~~~~~~~~~~~ poly A 3'
3' XXXXX ─────────────────── 5'

Template switching
and extension by RT

**SMART cDNA Primer
(SMART sequence)**

5' ~~~~~~~~~~~~~~~~~~~ poly A 3'
────────────────────── 5'

◄── **cDNA Primer**

| cDNA
amplification
PCR *performed
in 96-well plate*

cDNA amplification
pooling and cleanup

TCR Primer 1
(Read 2 + SMART sequence)

5' ──────────── 3'
3' ──────────── 5'

◄── **TCR a/b Human Primer 1
(constant region)**

TCR-specific PCR 1

TCR Primer 2
**Forward HT Index
(P7 + i7 + Read 2)**

5' ──────────── 3'
3' ──────────── 5'

◄── **TCR a/b Human Primer 2
Reverse HT Index (constant
region + Read 1 + i5 + P5)**

**TCR-specific PCR 2
Cleanup**

| TCR-specific
amplification
*performed
in tubes*

5' ──────────── 3'
3' ──────────── 5'

Sequencing-ready library

B

SMART-Seq Indexed Oligo A
SMART-Seq Indexed Oligo B
SMART-Seq Indexed Oligo C
SMART-Seq Indexed Oligo D
SMART-Seq Indexed Oligo E
SMART-Seq Indexed Oligo F
SMART-Seq Indexed Oligo G
SMART-Seq Indexed Oligo H

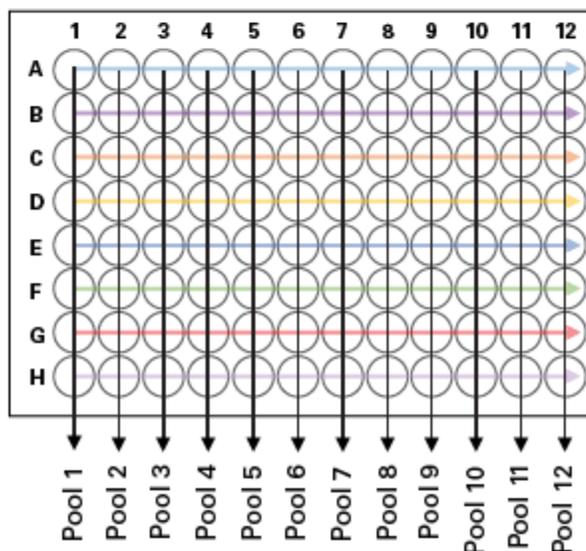96-well plate grid with rows A–H and columns 1–12; columns feed into Pool 1 through Pool 12.

Figure 2. SMARTer Human scTCR a/b Profiling Kit workflow and pooling strategy. Panel A. First-strand cDNA synthesis is dT-primed (RT Primer) and performed by an MMLV-derived reverse transcriptase (RT), which adds nontemplated nucleotides to the 5' end of each mRNA template. The SMART-Seq Indexed Oligos anneal to these nontemplated nucleotides and serve as a template for the incorporation of an additional sequence of nucleotides into the first-strand cDNA by the RT (this is the template-switching step). Each of the eight different SMART-Seq Indexed Oligos provided in the kit contains a unique six-base in-line index that serves as a cell barcode to allow downstream cell identification after pooling. The additional sequence added to the cDNA by the RT—referred to as the "SMART sequence"—serves as a primer-annealing site for subsequent rounds of PCR, ensuring that only sequences from full-length cDNAs undergo amplification. After pooling (described in Panel B) and a cleanup step, two rounds of gene-specific PCR are performed in succession to amplify cDNA sequences corresponding to variable regions of TCR-α and/or TCR-β transcripts. The first gene-specific PCR uses the amplified double-stranded cDNA as a template and includes a forward primer with complementarity to the SMART sequence—which also incorporates the Illumina Read 2 sequence (TCR Primer 1)— and reverse primers that are complementary to the constant (i.e., nonvariable) region of TCR-α and TCR-β (TCR a/b Human Primer 1). The second round of PCR takes the product from the first PCR as a template and uses a forward primer that binds to the Read 2 sequence added by the previous PCR step. The reverse primers bind in the constant region, internal to the PCR1 primers (TCR a/b Human Primer 2 Reverse HT Index), allowing amplification of the entire variable region and a portion of the constant region of TCR-α and TCR-β cDNA. The forward and reverse primers include adapter and index sequences that are compatible with the Illumina sequencing platform and allow for multiplexing of up to 96 samples in a single flow-cell lane. **Panel B.** Samples are pooled by column, such that each pool contains eight cells, each with a differently indexed SMART-Seq Indexed Oligo. Different combinations of the Forward and Reverse HT indexes are used during PCR 2 to allow multiplexing of the samples into a single flow-cell lane (see the User Manual for more details).

Results

Optimal index design allows for unbiased library construction and accurate multiplexing

In order to streamline the workflow and processing of samples, we devised a strategy for using in-line indexes (SMART-Seq Indexed Oligos) that act as cell barcodes which allow pooling the 96 cells into twelve sample pools following the cDNA amplification step (Figure 2B). Before proceeding with an analysis of large data sets, we wanted to validate two key aspects of our assay:

- Do all the SMART-Seq Indexed Oligos behave equivalently?

- How many errors should we allow when demultiplexing the data?

Control Jurkat Total RNA was processed with the human scTCR kit to generate libraries for each SMART-Seq Indexed Oligo. The libraries were then sequenced on an Illumina MiSeq® sequencer (see Methods section for details) to assess the performance of the barcodes. In particular, we wanted to assess whether any barcode introduced bias (resulting from sequence-specific differences in template-switching or amplification). The sequencing results from the final libraries showed that all of the different SMART-Seq Indexed Oligos performed equivalently and showed no bias, as evidenced by the similar percentage of total reads obtained from each indexed library across the different barcodes (Figure 3A). Additionally, the data indicate that minimal reads (~3%) cannot be assigned to any of the index sequences (Figure 3, Bad barcode).

We next wanted to examine the ability to demultiplex the data. Data for one of the pools sequenced for Panel A was run through the SMARTer Human scTCR Demultiplexer software allowing for either zero or one errors in the assignment of reads to the SMART-Seq Indexed Oligo barcodes. When zero errors were allowed, 6.4% of reads for the pool were unassignable to a specific index (Figure 3, Bad barcode). When one error was allowed the unassignable reads were reduced to 2.9% with a parallel, even, increase in the percent of reads mapping to each of the indexes (Figure 3B). Furthermore, analysis of the clonotype calls for the corresponding reads with zero or one errors, revealed identification of the correct TCR-α and -β clonotype for all Control Jurkat Total RNA wells at the same frequency, independent of the allowed errors (Figure 3C). Since allowing for one error in the demultiplexing process did not have any adverse effects on the data, we performed our data analysis using this setting. Taken together, these results speak to the robustness of the design of the in-line indexes in the SMART-Seq Indexed Oligos.
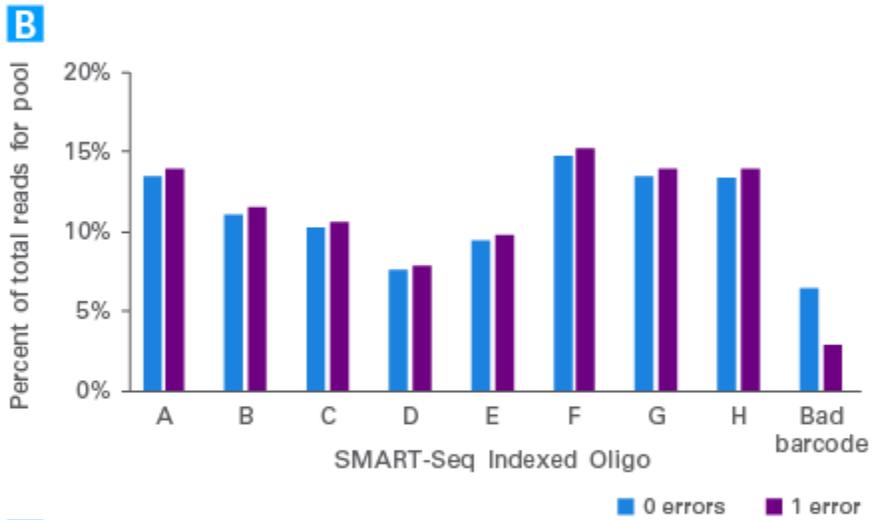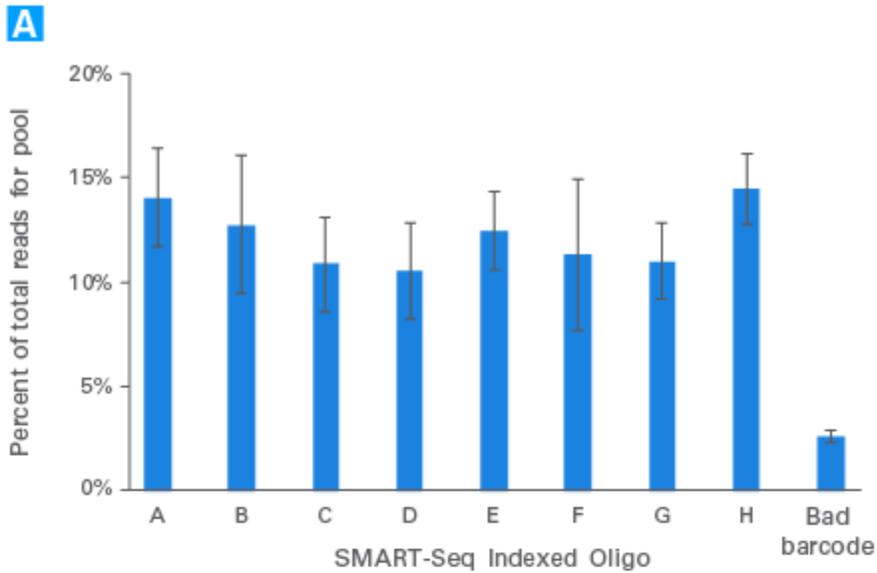
A

B

0 errors 1 error

C

SMART-Seq Indexed Oligo:
errors allowed in demultiplexing

TCR-α – 0 errors   TCR-α – 1 error
TCR-β – 0 errors   TCR-β – 1 error

Figure 3. Testing performance of indexed oligos and allowable errors for demultiplexing with the in-line indexes. Panel A. The performance of the in-line indexes was assessed based on the percent of reads mapping to each in-line index after demultiplexing allowing for one error in the barcode assignment. Seven replicate pools containing 8 x 5 pg Control Jurkat Total RNA were run across three experiments. Bad barcode = unidentifiable reads. **Panel B.** Data for one of the seven pools of Control Jurkat Total RNA was demultiplexed allowing for either zero or one errors in the in-line index. **Panel C.** Analysis of the clonotype calls for the data demultiplexed with zero or one errors.

Specific amplification of TCR a/b chains in the final libraries

We next moved to evaluating the human scTCR kit's performance using cells. To assess the performance, we mixed Jurkat and CCRF-CEM cells at a 1:1 ratio and sorted the cells via FACS. Cells were sorted into a 96-well plate, and TCR libraries were generated per the kit protocol (see Methods section). These cell lines were chosen because they are known to express different clonotypes that can be easily distinguished. To confirm the success of library amplification and purification, samples were run on a Bioanalyzer. The electropherograms show distinct peaks at ~700 bp and 900 bp, the expected positions for TCR-β and TCR-α sequences, respectively. The variations in the peak profiles, however, suggest pool-to-pool differences in the abundances of the two different chains.
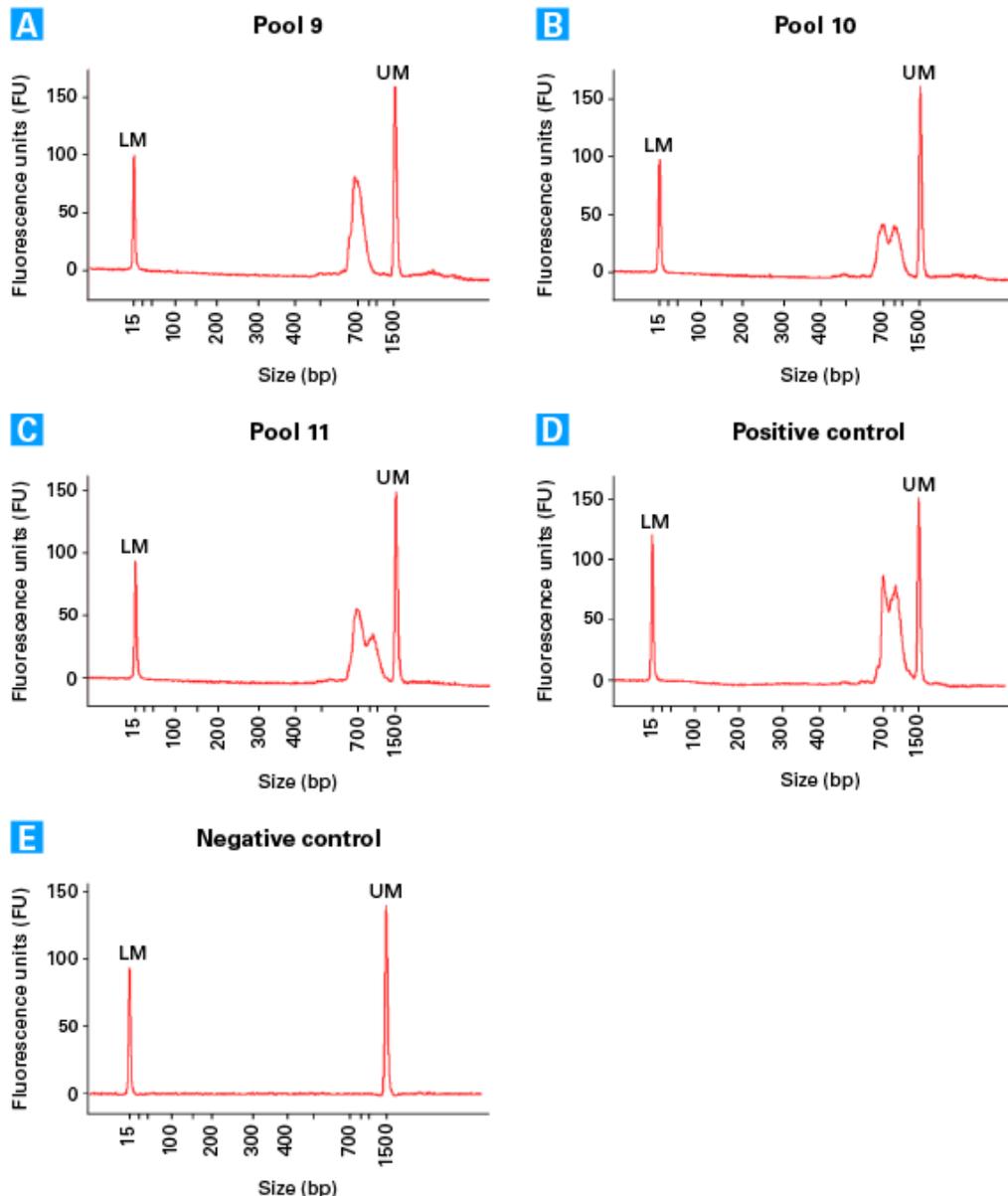
Figure 4. Electropherogram profiles of TCR sequencing libraries. Electropherograms from three different cell pools shown in Figure 6: Pool 9 (**Panel A**), 10 (**Panel B**), and 11 (**Panel C**). The variation in the profiles reflects the abundance of amplified TCR-α (peak at ~900 bp) and TCR-β (peak at ~700 bp). **Panel D.** Positive Control RNA TCR library (generated with 8 x 5 pg Control Jurkat Total RNA). **Panel E.** Negative Control TCR library shows no library is produced.

Making accurate clonotype calls in cells

We next set out to establish analysis criteria for the confident calling of clonotypes. This was done by examining the number of reads mapping to the top TCR-α or the top TCR-β clonotype in negative control wells (wells with no cells).

Since these wells are truly empty, any reads assigned to them can be considered to be background. For this experiment, libraries were generated and sequenced for three pools (each containing seven wells with cells and one empty well as a negative control). The FASTQ files for each pool were down-sampled to 300,000 reads, demultiplexed with one error allowed, and individual wells were analyzed using MiXCR2.0.2. To ensure that our data did not contain any clonotype calls that were within the background range, the mean number of clonotype reads (mean = 74 reads) for the negative wells was established and any clonotype call for a cell-containing well that fell within three standard deviations (mean + 3SD = 238 reads) of the negative-well mean was omitted from further analysis. The data of cell-containing wells (Figure 5, blue dots on the left) was evaluated using this threshold of 238 reads to make a clonotype call. With the threshold applied (shown by the dashed red line), low-read clonotype calls were eliminated, leaving us with data that was out of the background range (purple dots, right).
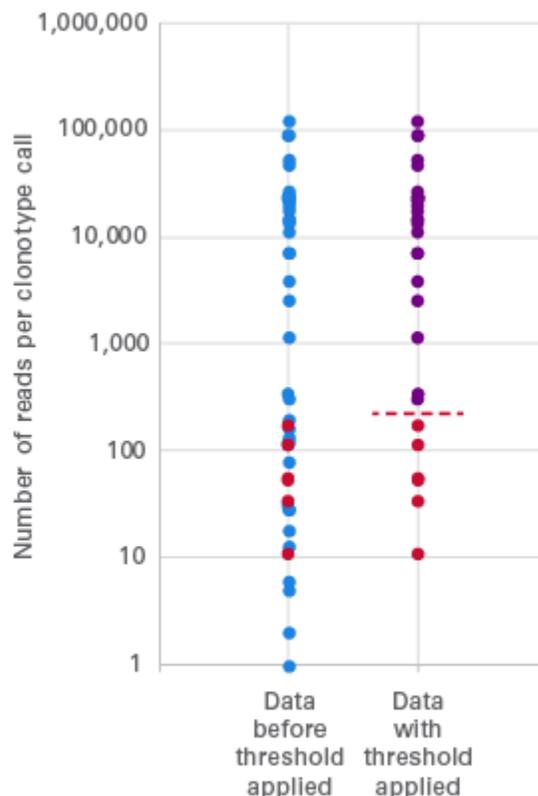

Figure 5. Establishing a threshold for clonotype calling using negative control wells. The number of reads mapping to the top TCR-α and the top TCR-β clonotype identified for each cell before the established threshold was applied are shown on the left (blue dots). The clonotype calls that meet the threshold criteria are shown on the right (purple dots). Negative control wells are represented by red dots (negative control well clonotype read number mean = 74 reads, SD = 55 reads, mean + 3SD = 238 reads). The dashed line represents the established 238-read cutoff.

Analysis of a mixed population of single cells reveals pairing information

Following sequencing of the mixed Jurkat-CCRF pools, we sought to demultiplex the data and determine which cell type was present in each well of the pool, and subsequently which TCR clonotypes were present in each of the cells. The FASTQ files for each pool were down-sampled to 300,000 reads, demultiplexed with one error allowed for barcode assignment, and individual cells were anlyzed using MiXCR2.0.2. The threshold of 238 reads for clonotype calling that was established in previous experiments (Figure 5) was applied to this dataset, and clonotype calls that fell below the threshold were omitted.

When the clonotype sequences were examined to determine which cell type they were characteristic of, 43 wells were assigned as Jurkat while 46 cells were deemed CCRF-CEM (Figure 6A). This observed distribution corresponds well with 1:1 cell mixture prepared for these experiments. A total of seven wells could not be assigned a cell type because they did not have clonotype calls for either TCR-α or TCR-β chains with read numbers that were above the cutoff threshold.

Digging further into the data, we next sought to examine which chains were contained in each cell (Figure 6B, 6C). A total of 38% of the cells were assigned an αβ pairing, 16% of the cells had only an α chain, and 46% of the cells showed only a β chain. A careful comparison of the electropherograms in Figure 4 with the data in Pools 9, 10, and 11 shows a clear correlation between the observed peaks on the trace and the relative abundance of TCR-α or TCR-β clonotypes in the resulting data. For example, Pool 9 was largely comprised of cells for which only a β chain was sequenced, and the pool's electropherogram had a very predominant peak at ~700 bp, a characteristic of TCR-β chains. Meanwhile, Pool 11's electropherogram, shows peaks for both TCR-α and TCR-β chains with the TCR-β peak slightly more prominent as expected for this pool where eight cells had a sequenced β chain while only five contained an α chain.

It should be noted that for cells where only one chain was detected the data does not rule out the complete absence of the opposite chain. One possibility is that since the cells were not activated the missing chain was not present at high enough levels to be detected. Another is that there were not enough reads (i.e., reads fell below the threshold) to confidently call the clonotype for the opposing chain. The low input concentration for these single cells may be another contributing factor. The observation of a high proportion of cells containing only TCR-β reads is not unexpected since literature indicates that beta chains are more highly expressed.
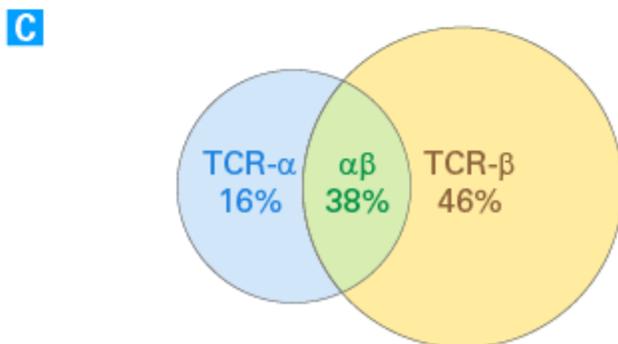
Figure 6. Analysis of a mixed cell population from a 96-well plate. Panel A. Cell-type calling based on the identified clonotypes for each well. The seven omitted cells did not have clonotype calls for either TCR-α or TCR-β with read numbers that were above the threshold. **Panel B.** Analysis of pairing information. Paired TCR-αβ clonotypes were obtained for 34 cells in the plate. **Panel C.** The alpha, beta, or alpha-beta pairing information represented as a percent of cells analyzed. Omitted cells were not included in this analysis.

Conclusions

The SMARTer Human scTCR a/b Profiling Kit provides a powerful solution for sensitive profiling of TCR chains from single cells, allowing the identification of alpha-beta chain pairings in each cell. Our validated assay design is optimized to pool 96 samples into 12 sequencing libraries, making it easy to use. In addition, the 12 libraries can be further multiplexed to run in a single flow-cell lane. Sequencing data can be demultiplexed using the SMARTer Human scTCR Demultiplexer software to assign reads and make accurate clonotype calls for each cell. The sensitivity of the kit allowed identification of alpha-beta pairings in Jurkat and CCRF-CEM cells that are in agreement with single-cell sequencing reports by other groups.

Methods

Libraries containing TCR-α and TCR-β sequences were generated using the SMARTer Human scTCR a/b Profiling Kit per the protocol given in the user manual. For assay validation experiments, libraries were generated from Control Jurkat Total RNA (Takara Bio). Each pool contained 8 x 5 pg of the Control Jurkat Total RNA, and seven replicate pools were run across three experiments. For the mixed-cell population studies, Jurkat and CCRF-CEM cells (ATCC) were mixed at a 1:1 ratio. Cells were sorted into a 96-well plate using FACS prior to processing with the human scTCR kit.

Samples were pooled at a final concentration of 4 nM. The final library pool was diluted to 13.5 pM, including a 5–10% PhiX Control v3 (Illumina) spike-in for sequencing. While not essential, the addition of the PhiX control allows for detection of sequencing errors and increases the nucleotide diversity and thus aids in high-quality data generation. Sequencing was performed on an Illumina MiSeq sequencer using the 600-cycle MiSeq Reagent Kit v3 (Illumina) with paired-end, 2 x 300 base pair reads.

After sequencing, the FASTQ files for each pool were demultiplexed using the SMARTer Human scTCR Demultiplexer available on the Takara Bio website and reads were assigned to each in-line index/sample well. Unless otherwise stated, demultiplexing was performed allowing for one error in the in-line index. Repertoire analysis for each sample well was performed using MiXCR 2.0.2 (Bolotin et al. 2015).

References

Becattini, S. *et al.* Functional heterogeneity of human memory CD4+ T cell clones primed by pathogens or vaccines. *Science* **347**, 400–406 (2015).

Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–1 (2015).

Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–32 (2016).

View web page >>

http://www.takarabio.com/US/Products/cDNA_Synthesis_and_Library_Construction/NGS_Learning_Resources/Technical_Notes/Human_Single_Cell_TCR_Profiling