T-cell Receptor Profiling Using PacBio Sequencing of SMARTer® Libraries

Steve Oh[1], Sarah Taylor[2], Melissa Laird-Smith[3], Andrew Farmer[2], Cheryl Heiner[1]
[1]PacBio, Menlo Park, CA, [2]Takara Bio USA, Inc., Mountain View, CA, [3]Mt Sinai, New York, NY

Abstract

T-cells play a central part in the immune response in humans and related species. T-cell receptors (TCRs), heterodimers located on the T-cell surface, specifically bind foreign antigens displayed on the MHC complex of antigen-presenting cells. The wide spectrum of potential antigens is addressed by the diversity of TCRs created by V(D)J recombination. Profiling this repertoire of TCRs could be useful from, but not limited to, diagnosis, monitoring response to treatments, and examining T-cell development and diversification.

Due to the combinatorial generation of the TCRs, analyzing the full-length genes is more informative than probing partial regions when trying to distinguish unique variants. Accurate sequences are also required to confidently identify specific clonotypes that may be critical in patient diagnosis or finding optimal binding interactions for therapy. Sequencing from genomic DNA is useful for gaining information about the representation of different T-cells in a population. However, starting from mRNA provides additional benefits by enriching for the TCR sequence as well as giving a more direct view of the functional expression of the genes. Current methods may fall short if transcript length extends beyond the instrument limits and if read quality is not high.

By combining the Takara Bio SMARTer® Human TCR a/b Profiling Kit and the PacBio Sequel System, full-length TCR genes have been sequenced. Here we present results from Jurkat- and Peripheral Blood Mononuclear Cell (PBMC)-derived RNA showing high reference alignment and sequence accuracy along the entire lengths of the TCR genes. Starting with total RNA from Jurkat cells and peripheral blood leukocytes, 400-900 bp amplicons containing the entire variable region and a portion of the constant region of TCR-α and/or TCR-β were generated, taken directly into PacBio SMRTbell library prep and run on the Sequel System, producing >200,000 high-quality, full-length TCR gene sequences per SMRT Cell 1M.

SMARTer TCR a/b Profiling Kit and SMRTbell Library Prep Workflow

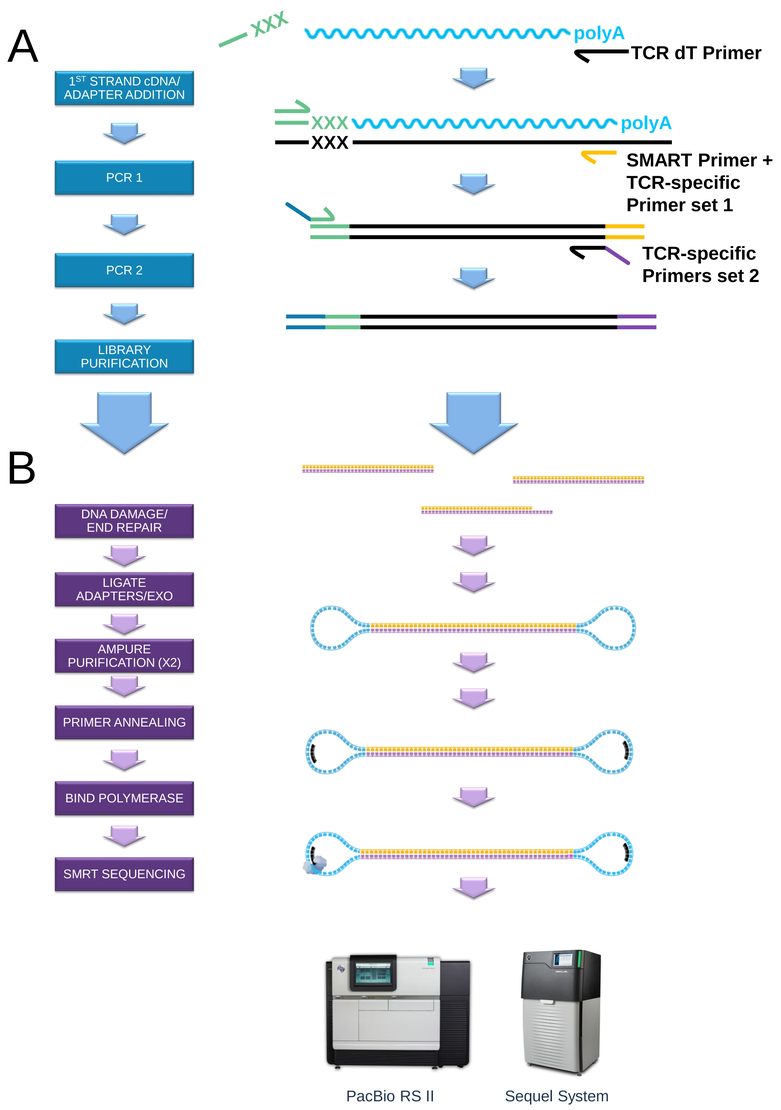
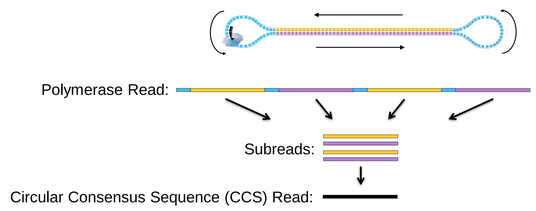


Figure 1. (A) SMARTer TCR transcript enrichment and (B) SMRTbell library prep and sequencing

Highly Accurate Single Molecule Sequencing

Figure 2. Multiple reads from a single molecule. As a function of the SMRTbell adapters, multiple single-pass reads are generated from an individual molecule. Combining these subreads corrects for random errors and results in a highly accurate single molecule consensus sequence. Data can be filtered to an accuracy of 99.99%.



SMRTbell Library Prep and Sequencing

Sample	cDNA Input	SMRTbell Yield
Jurkat	247.5 ng	66 ng (26.7%)
PBMC	211.7 ng	54 ng (25.5%)

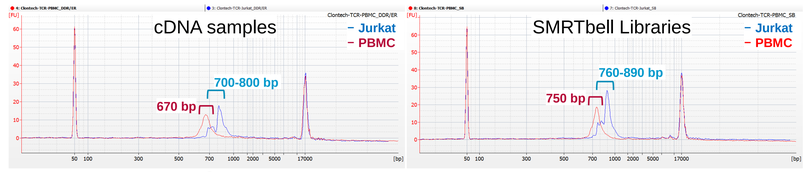
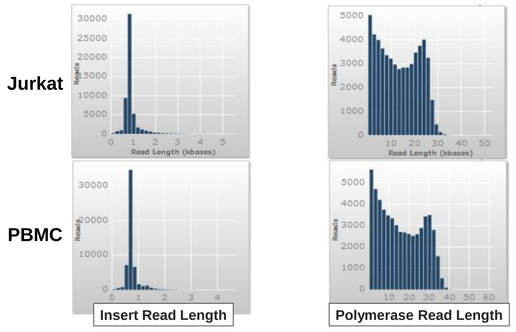


Figure 3. Bioanalyzer® traces of samples before and after SMRTbell library construction.

Sample	# PacBio RS II SMRT Cells	Loading Conc	P1* Reads	Reads at CCS 99%**	Total Bases (Mb)
Jurkat	1	20 pM (≈0.7 ng)	54,433	29,760	751
PBMC	1	30 pM (≈0.6 ng)	56,012	31,196	915



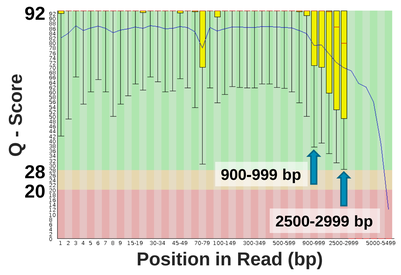


Table 1 / Figure 4. Sequencing and CCS results on the PacBio RS II system. (***above***) Primary results show high quality reads for full-length Jurkat and PBMC amplicons using DNA Sequencing kit 4.0 v2. (***left***) Read quality scores were determined for the Jurkat SMRTbell library by FastQC[1] analysis. A high quality score is maintained at each base along the length of the reads.
**P1 reads represent ZMW wells that are productive and sequencing*
***Reads with a minimum predicted accuracy of 99%*

Sequel System Sequencing Results

Sample	# Sequel SMRT Cell 1Ms	Loading Conc	P1* Reads	Reads at CCS 99%**	Total Bases (Gb)
Jurkat	1	125 pM (≈4.4 ng)	240,604	145,432	3
PBMC	1	130 pM (≈5.4 ng)	509,312	333,795	7

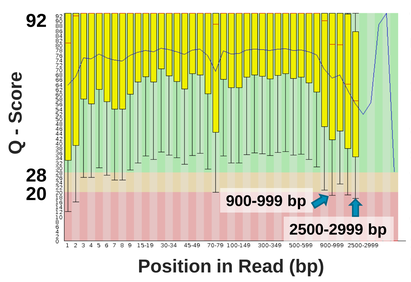


Table 2 / Figure 5. Sequencing and CCS results on the Sequel System. The number of reads is increased on the Sequel System, using Sequel Sequencing kit 1.2, compared to the PacBio RS II system. (***left***) FastQC analysis was performed on the Jurkat sample. Mean quality score ranges from 64-76 per base along the length of the TCR amplicon. Loading requirements were decreased and read quality was improved for other samples when Sequel Sequencing kit v1.2.1 was used. Kit v1.2.1 is to be tested for TCR sequencing.

TCR Representation

Alignment to TCRα and TCRβ loci				
Input	Jurkat RNA		PBMC RNA	
Run	1 (04/16)	2 (07/16)	1 (04/16)	2 (07/16)
# PacBio RS II SMRT Cells	4	1	4	1
Total reads	215,433	32,874	105,336	35,664
Aligned reads	210,012	31,145	89,948	27,838
% of total reads	**97.5%**	**94.7%**	**85.4%**	**78.1%**

Clonotype Mapping				
Clonotypes identified	88	13	49,002	18,354
Reads used in clonotypes	204,487	30,122	77,157	22,718
% of aligned reads	**97.4%**	**96.7%**	**85.8%**	**81.6%**
% of total reads	**94.9%**	**91.6%**	**73.3%**	**63.7%**
Reads mapping to correct, dominant clonotype	192,042	28,629		
% of reads mapping to clonotypes	**93.9%**	**95.0%**		
% of total reads	**89.1%**	**87.1%**		

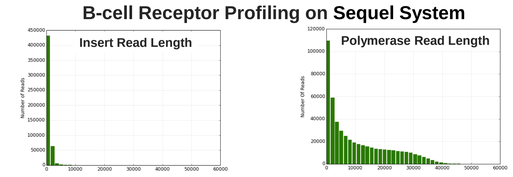
Table 1. Analysis of TCR repertoire sequenced on the PacBio RS II System. Datasets from two different experiments are compared for Jurkat and PBMC samples. TCR sequences were extracted from using MiXCR[2]. (***top***) >94% and >78% of reads mapped to the TCR loci for Jurkat and PBMC libraries, respectively. (***bottom***) Compared to the low number of clonotypes detected from Jurkat, >49,000 and >18,000 clonotypes were identified in the complex PBMC sample sequenced from 4 and 1 SMRT Cells, respectively. Further, 94-95% of reads used to identify clonotypes in the Jurkat sample map to the expected, dominant clonotype.

Expanding Applications in Immunology

Beyond TCR profiling, we are investigating how PacBio sequencing may provide advantages in other applications in the immunology space. Long-read sequencing for B-cell receptor related constructs as well as synthetic short chain variable fragments are presented below.

B-cell Receptor Profiling on Sequel System

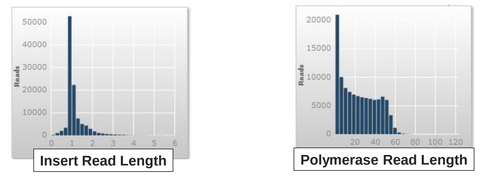
A



# Sequel SMRT Cells	Loading Conc	P1* Reads	Reads at CCS 99%**	Total Bases (Gb)
1	25 pM (≈1 ng)	507,141	268,926	5.5

Short Chain Variable Fragment Screening on PacBio RS II

B



# PacBio RS II SMRT Cells	Loading Conc	P1* Reads	Reads at CCS 99%**	Total Bases (Gb)
1	30 pM (≈0.9 ng)	49,931	24,121	2.2

Figure 6. (**A**) B-cell receptor profiling library was sequenced on the Sequel System**.** Using Sequel Sequencing chemistry v1.2.1, amplicons 700-900 bp were loaded at a concentration (25 pM), ≈ 5-fold lower than required with Sequencing chemistry v1.2 (125 pM). (**B**) Short chain variable fragment (ScFv) candidates were sequenced using DNA Sequencing kit 4.0 v2 for the PacBio RS II system. Results from both B-cell receptor and ScFv experiments are currently being analyzed.

Results and Conclusions

- Combining the Takara Bio TCR a/b Profiling Kit and PacBio sequencing, full-length (650-900 bp) amplicons were sequenced from Jurkat and PBMC total RNA.
- With high-quality CCS reads, nearly all reads (97.5%) align to TCR loci from Jurkat, and a wide range (>18,000) of PBMC clonotypes were detected on a single PacBio RS II SMRT Cell.
- The library preparation yield of SMRTbell DNA exceeded the amount required for a single Sequel SMRT Cell 1M, starting with 200-250 ng input cDNA. Investigation into lower DNA input levels for library preparation is currently underway.
- PacBio sequencing is well-suited for immunological projects requiring high accuracy sequences >600 bp to several kb.

References

[1]FastQC http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[2]Bolotin D, et al. (2015) MiXCR: software for comprehensive adaptive immunity profiling. *Nature Methods*. 12(5), 380-381.