Single Cell Genomics 2018; Poster # 74

Single-Cell Stranded Total RNA Sequencing Provides a More Comprehensive Transcriptome Analysis than Oligo(dT)-Based Method

Aik Ooi, Ben Lacar, Chad Sanada, Mark Lynch, Naveen Ramalingam
Fluidigm Corporation, 7000 Shoreline Court, Suite 100, South San Francisco, CA 94080 USA

Introduction

Most single-cell RNA sequencing (RNA-seq) methods rely on capturing polyadenylated [poly(A)] mRNAs while excluding non-poly(A) RNAs such as some long non-coding RNAs and histone mRNAs. The expression status of these RNAs could broaden insights into each cell since many non-poly(A) RNAs take regulatory roles in cellular processes. Accordingly, a single-cell total RNA-seq method that retains both poly(A) and non-poly(A) RNAs is needed. We successfully adapted a total RNA-seq (TR-Seq) protocol for single cells on the Fluidigm C1™ system. This TR-Seq method also preserves strand information, allowing for improved alignment and the ability to distinguish genes transcribed from opposite strands of the same genomic loci.

Methods

We implemented the SMARTer® Stranded Total RNA-Seq Kit v2—Pico Input Mammalian (Takara Bio, Cat. No. 634413) and its updated version SMART-Seq® Stranded Kit (Takara, Cat. No. 634444) on the C1 system for single cells. To evaluate single-cell TR-Seq on C1, we performed a side-by-side comparison with an oligo(dT)-based method using SMART-Seq v4 Ultra Low Input RNA Kit (Takara, Cat. No. 635025) across three different cell types.

Protocol Name	Abbreviation	Type of RNA
C1 with SMARTer Stranded Total RNA-Seq Kit v2 —Pico Input Mammalian	TRv1	Total RNA by random priming
C1 with SMART-Seq Stranded Kit	TRv2	Total RNA by random priming
C1 with SMART-Seq v4 Ultra Low Input RNA Kit	SSv4	Poly(A) RNA by oligo(dT) priming

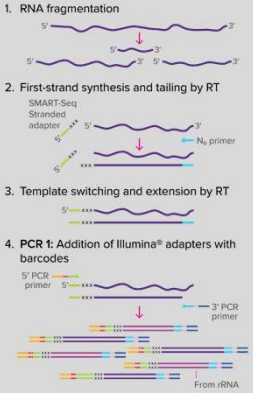
TRv1 was first developed and tested in conjunction with SSv4 on K562 and HL60 cells. The improved version, TRv2, was compared with SSv4 on activated T cells*. TRv2 was tested again on K562 and HL60 cells to validate performance. Full protocols and C1 scripts for the SMART-Seq v4 and TR-Seq applications can be found on the Fluidigm C1 Script Hub™ website (fluidigm.com/c1openapp/scripthub). Libraries were sequenced with the MiSeq® Reagent Kit v3 at 72 bp PE. FASTQ files were processed using Kallisto (v0.44). Ensembl cDNA and ncRNA FASTA files (release 92) were used to build the Kallisto index. SSv4 datasets were processed in unstranded mode while TRv1 and TRv2 datasets utilized stranded mode. The R package tximport was used to generate gene-level expression tables from Kallisto transcript-level output. Custom Python® and R scripts were created for figure visualizations.

* Frozen human pan T cells were thawed and stimulated for three days with anti-CD3/28 beads in the presence of 30 U/mL IL-2.

Overview of C1 total RNA-seq

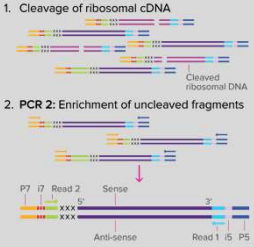
Automated steps on C1
Cell lysis and RNA fragmentation occur, followed by reverse transcription and template switching extension. Indexing primers are added to cDNA fragments by PCR, retaining the strand information.

1. RNA fragmentation
2. First-strand synthesis and tailing by RT (SMART-Seq Stranded adapter; N_6 primer)
3. Template switching and extension by RT
4. **PCR 1:** Addition of Illumina® adapters with barcodes (5' PCR primer; 3' PCR primer; From rRNA)



Single-tube library prep after C1
Indexed cDNA samples are pooled and subjected to cleanup by AMPure® XP beads. Ribosomal cDNA are cleaved and the remaining uncleaved cDNA are enriched. Stranded TR-Seq libraries are ready for sequencing after purification by AMPure XP beads.

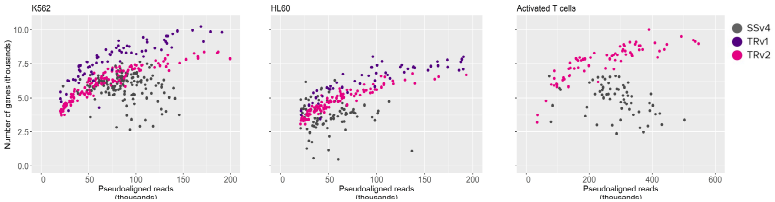
1. Cleavage of ribosomal cDNA (Cleaved ribosomal DNA)
2. **PCR 2:** Enrichment of uncleaved fragments (P7 i7 Read 2 Sense; Anti-sense; Read 1 i5 P5)



Schematic diagram adapted from SMART-Seq Stranded Kit User Manual (Takara Bio Cat. No. 070518)

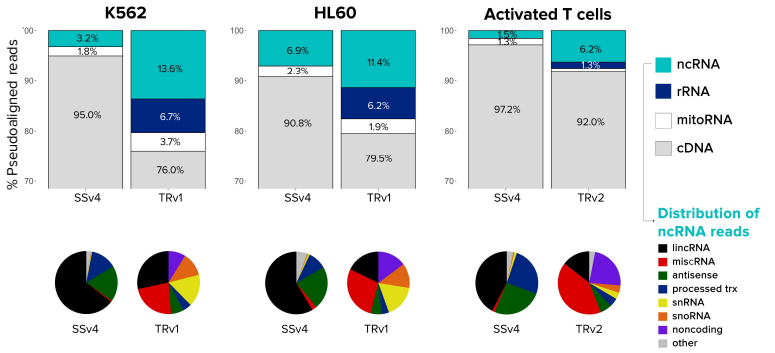
Results

TR-Seq detects more genes than SSv4



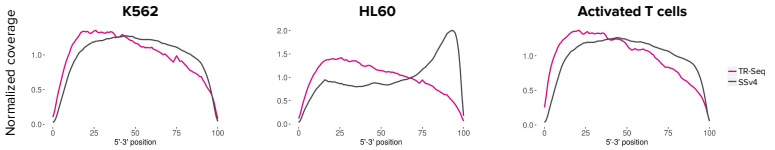
TR-Seq outperforms SSv4 in the number of genes detected per single cell at a varying number of pseudoaligned reads for all three cell types tested. TRv2 shows improvement over TRv1, as shown in K562 and HL60 data. Samples that were visually scored as single and live with >20,000 pseudoaligned reads were analyzed further. Pseudoaligned reads for each sample were derived directly from the Kallisto output. A gene is counted if it shows an expression level TPM >1.

TR-Seq captures greater proportion and diversity of noncoding RNA



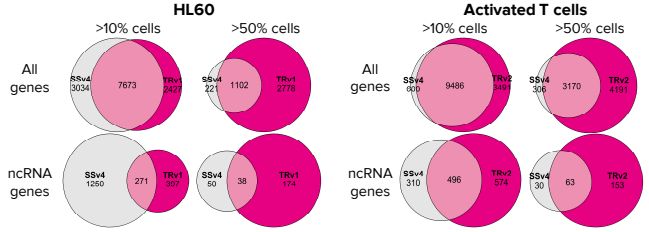
A higher percentage of reads pseudoaligned to noncoding RNA (ncRNA) is observed in TR-Seq compared to SSv4. Within reads attributed to ncRNA, a greater diversity of transcript biotypes is observed in the TR-Seq data. TRv2 detects a lower percentage of ribosomal RNA thanTRv1, indicating improvement on ribosomal cDNA depletion step (data not shown). Biotypes for genes extracted from FASTA file headers were used to quantify biotype percentages for each cell type. The biotype percentage was defined as the sum of all counts for that biotype divided by the total counts for that sample.

TR-Seq shows consistent transcript coverage



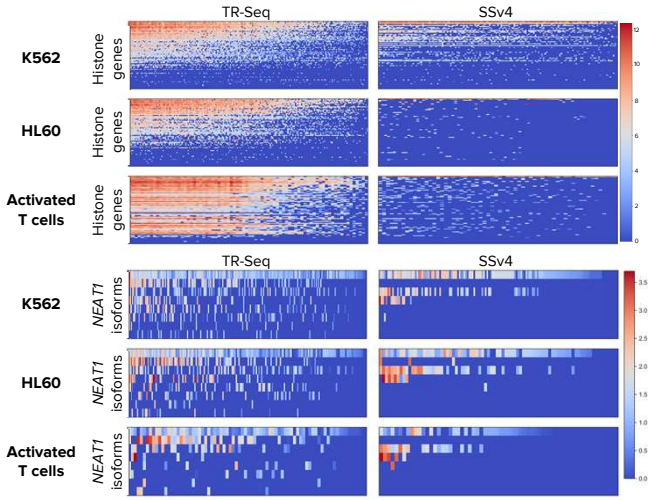
TR-Seq exhibits a slight 5' preference in transcript coverage, but its performance is consistent across cell types. In HL60 cells where RNAs are more fragmented, SSv4 shows high 3' bias due to its poly(A) priming while TR-Seq provides better coverage across transcripts. The TR-Seq versions were TRv1 for K562 and HL60 and TRv2 for activated T cells. The transcript coverage plots were derived from Picard output after FASTQ files were processed using Tophat. The coverages shown are the normalized averages across all samples for each cell type based on top 1,000 genes.

TR-Seq returns most genes that are detected in SSv4 with less variable data across single cells



Venn diagrams showing number of genes (TPM >1) detected in >10% and >50% of single cells for each protocol. The majority of genes uniquely detected in SSv4 appear in fewer than 50% of cells, indicating technical variability of data. In comparison, TR-Seq shows more consistent and reproducible data across single cells. The R package EulerR was used to generate Venn diagrams.

TR-Seq detects non-poly(A) genes and transcript isoforms



Many histone genes and some Nuclear Enriched Abundant Transcript 1 (*NEAT1*) isoforms are known to lack a poly(A) tail. TR-Seq demonstrates better detection of histone genes and *Neat1* isoforms than SSv4. Enabling a more comprehensive survey of histone genes is valuable in the study of cell cycle. For each plot, the TPM values were log2-transformed. The histone genes or *NEAT1* transcripts are arrayed from top to bottom in order of descending average expression level across all cell types. Within each method, the samples are positioned from left to right in descending average gene/transcript expression. Both TRv1 and TRv2 samples for K562 and HL60 are included in the heat maps.

Conclusions

We demonstrate that TR-Seq provides a more comprehensive transcriptome for single cells than an oligo(dT)-based method. More noncoding RNA and a greater diversity of ncRNA biotypes are captured with TR-Seq. In addition to finding that more genes are detected with TR-Seq, we observe that the genes detected are more consistent across individual cells within a given cell type. TR-Seq also generates more even transcript coverage. Finally, non-poly(A) genes and transcripts show significantly better detection rates with TR-Seq. In summary, we demonstrate that TR-Seq has the potential to provide further insight into the regulatory roles of ncRNA and non-poly(A) genes in single-cell gene expression.

Fluidigm Corporation
7000 Shoreline Court, Suite 100
South San Francisco, CA 94080, USA
+1 650 266 6100 • Toll-free: 866 359 4354 in the US and Canada
fluidigm.com

©2018 Fluidigm Corporation. All rights reserved. Fluidigm, the Fluidigm logo, and C1 are trademarks and/or registered trademarks of Fluidigm Corporation in the United States and/or other countries. All other trademarks are the sole property of their respective owners. 10/2018