Achieving Unparalleled Sensitivity and Reproducibility in Single-Cell Transcriptomics

Rashmi Rajendra, Kanot Tel, Cynthia Chang, Yenygin Gindin, Magnolia Bostick, and Andrew Farmer

Abstract

Gaining the ability to identify and quantify the mRNA from a single cell has been a substantial benefit to many scientific fields, especially those where homogeneous populations are elusive, such as cancer research, developmental biology, neurobiology, and immunology. SMART-Seq® (Switching Mechanism at the 5’ end of the RNA Template) single-cell technology has emerged as the best solution for processing the small amounts of mRNA present in each cell. With the SMART-Seq® v4 Ultra Low RNA Input Kit for Sequencing, we have incorporated LNA technology in order to produce high-quality, reproducible sequencing data with superior identification of gene expression. Ideally, researchers working with single cells would like to analyze hundreds or thousands of individual cells. Unfortunately, many of the protocols for transcriptome library production from single cells is time consuming and not readily amenable to scaling experiments. By modifying adapters, adding cellular indexes, and pooling, we have developed a simplified protocol that enables parallel library preparation for high-throughput mRNA quantification utilizing 3’ end capture.

Here we present this new approach with multiple validation experiments. We first confirm the validity of pooling by mixing single reactions from different sample types with ERCC spike-in controls and testing for cross contamination. Additionally, samples within pools are compared to samples done completely independently. We show limited cross contamination within pooled samples (0.02%) which is not present in independent samples. We further validate the SMART-Seq® v4 Ultra Low RNA Input Kit for single-cell sequencing by pooling 12 individual cells. We observe a high number of transcripts identified and high correlations between the cells. This approach allows each pool of 12 cells to be tagged by one of the 96 Illumina HT barcode combinations, enabling up to 1,152 separate cells per run. Researchers using this kit can improve sensitivity while decreasing the cost and time required for discovery.

Introduction and Methods

End-capture methods are appealing as they can decrease the amount of reads necessary to determine differential expression between cells. Additionally, samples can be pooled prior to library preparation when barcodes are added, decreasing the work and resources required and potentially increasing the multiplexing capabilities of a flow cell. In this poster, we demonstrate the use of SMART-Seq® v4 to develop an end-capture method for low-input amounts down to the single-cell level. This method utilizes a modified RT primer which includes a barcode and a portion of the Illumina® read primer 2 sequence in order to accommodate a pooled library generation protocol.

Library preparation for differential expression

Overview of the end-capture method for analyzing differential expression. Oligonucleotides are synthesized with a blocked 3’ end and modified nucleotides, as indicated, with a 5′-end barcode sequence and a 3′-end transposon (transposase recognition) terminal oligonucleotide sequence (TnRP1 and TnRP2, orange and purple respectively). The 3’ ends of the original cDNA are captured by selective PCR with primers for the transposon complexes added (supression PCR). After cDNA amplification, the Illumina RP2 sequence is used as a priming site during library amplification, and the cell barcode is used for demultiplexing reads. This protocol is used to prepare mouse (I) and human (Iv) libraries.

Testing the end-capture method

Detailed experimental design to test the end-capture strategy with low-input amounts (10 fg RNA). To test the validity of the pooling strategy described in the Introduction and Methods (v4 pool), we designed an experiment to compare the results from samples derived from a single cell with those derived from three cells. In the experiment, 10 mouse Brain Total RNA (RNA in the figure, Conviron) prepared with General RNA Control Consortium (GRCC) spike-in controls to simulate mouse (Iv) and human (Iv) spike-in controls (Iv), targeting those sequences in the human brain (Ambion GRCC spike-in control) and priming oligon (dual indexing) i11 (i11) and i12 (i12). The samples were kept unpooled until they were sequenced on an Illumina MiSeq® (v4). The four unpooled libraries are labeled independent (Iv).

High library correlation:
K562 single cells

Pearson correlation heat map of K562 single cells. The heat map represents the Pearson correlation for each sample and represents the pairwise correlation for K562 single cells. For all comparisons, the correlation (R) is >0.9. The samples of the heat map scale from dark blue (correlation <0.9) to dark red (correlation >0.9). The intensity of the color is determined by the number of genes in the sample (2,183). The number of genes identified from K562 single cells was determined with an independent gene-containment test.

Conclusions

- End-capture technology based on SMART-Seq® v4 can be used to pool samples
- Pooled samples provide highly sensitive and reproducible data with low cross-contamination
- End-capture technology enables highly multiplexed experiments with single cells up to 1,152 separate cells per run

References