Fluent BioSciences

APPLICATION NOTE

# Validation of a DSP Fixation Protocol Compatible with PIPseq™ Technology for Single-Cell RNA Sequencing

## Introduction

Single-cell RNA sequencing (scRNA-seq) has made profound impacts in the study of cellular and molecular diversity in complex tissues[1-5]. However, specialized capital investment, high reagent costs, lack of accessibility and scalability are key factors limiting the wide scale adoption and use of single cell technologies. Fluent BioSciences has developed a breakthrough single-cell analysis technology that relies on Pre-templated Instant Partitions (PIPseq) that can scale easily from hundreds to millions of individual partitions in a single sample. This technology offers a low barrier to entry without the need for complex instrumentation or expensive consumables, and can be easily implemented in any molecular biology laboratory.

One complication to single-cell transcriptomic experimental design is timely acquisition, processing, and analysis of samples. Cell fixation, using a variety of chemical agents, has been utilized to preserve the transcriptional states of dissociated cells for scRNA-seq analysis. This method has enabled asynchronous collection of clinical samples, banking of samples, and time course of treatment studies with efficient, parallel sample processing of the preserved cells. Here, we demonstrate compatibility of cell fixation with PIPseq to evaluate cellular diversity in peripheral blood mononuclear cells (PBMCs), which have become an established cell type to resolve heterogeneity within complex biological systems, as they comprise a diverse collection of closely related immune cell populations.

In these studies, we use the reversible cross-linker 3,3-dithiodiprpionic acid di(N-hydroxysuccinimide ester) (DSP), which has been shown to preserve the quality and integrity of RNA and allows for a standard transcriptomics workflow post-fixation. We show that cell types from DSP-fixed PBMCs can be resolved using the PIPseq workflow with a high degree of accuracy, and the information is consistent with an established competing scRNA-seq technology.

## Methods

Here, we describe the PIPseq workflow and protocol for use with DSP-fixed PBMCs (see Figure 1 and Protocol 1). To begin with, we created a single cell suspension and then incubated the suspension in a DSP solution, followed by several washes. The DSP-fixed cells were then added into the PIPs-containing tubes provided in the Fluent 3' Single Cell RNA Kit following the standard PIPseq workflow.

The PIPseq workflow (Figure 1B) begins with prepared cell suspensions which are mixed with Fluent template particles and segregated into Pre-templated Instant Partitions (PIPs) by simple vortexing. Cells in PIPs are then lysed on a thermal device and mRNA is captured by barcoded oligonucleotides incorporated with the template particles. cDNA is generated from the captured mRNA via reverse transcription and amplified to create a cDNA library for each individual cell. The single-cell cDNA libraries are then processed into sequencing libraries using standard library preparation methods followed by next generation sequencing (NGS). From the sequencer, FASTQ files are output and uploaded into PIPseeker, Fluent's proprietary analysis platform, to process scRNA-seq data through barcode extraction, mapping, gene-barcode matrix generation, and secondary analysis (i.e., clustering and differential expression). PIPseeker is compatible with multiple species, cell types, and treatment approaches, including DSP fixation.

Pre-processed data can be transferred into multiple downstream processing environments, such as R-based Seurat[A] or Python-based ScanPy[B], to enable biological discoveries. Here, we utilized the Seurat infrastructure to analyze and visualize our results. Once loaded into Seurat, we removed cell barcodes that had less than 150 genes

[A] https://satijalab.org/seurat/

[B] https://scanpy.readthedocs.io/en/stable/

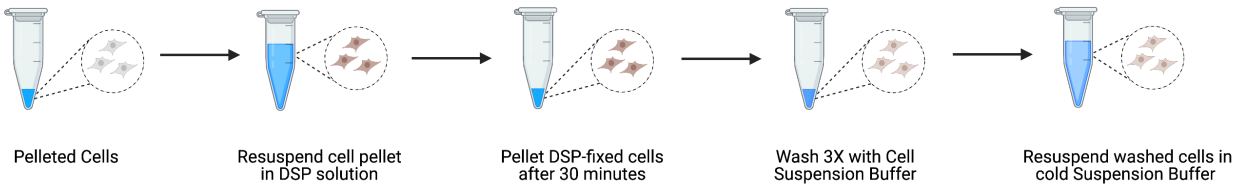
per cell. Barcodes that passed this filter were annotated by PBMC cell type using SingleR[6], a tool designed to compare each cell individually to a reference population. We used the annotated Seurat object from the Seurat PBMC Guided Clustering Vignette[c] as the reference, which is derived from a 10X Genomics v1 chemistry PBMC dataset. Then, the data were normalized using SCTransform[7], a method that utilizes variance stabilization to transform the data in a way that is most consistent with scRNA-seq data distributions. Principal component analysis (PCA) was performed on the transformed data to find axes with the most variance, and input as variables into the uniform manifold approximation and projection (UMAP) algorithm. Additionally, the data were clustered using a graph-based method[8], which uses nearest neighbors to identify local regions at different resolution levels. At our level of resolution (Seurat default, 0.5), we identified and removed four clusters that had significantly lower transcripts per cell and were associated with several SingleR-annotated cell types, indicative of lower data quality. Upon cleaning, data were re-clustered (using the graph-based method) and projected (using UMAP).

[c] https://satijalab.org/seurat/articles/pbmc3k_tutorial.html

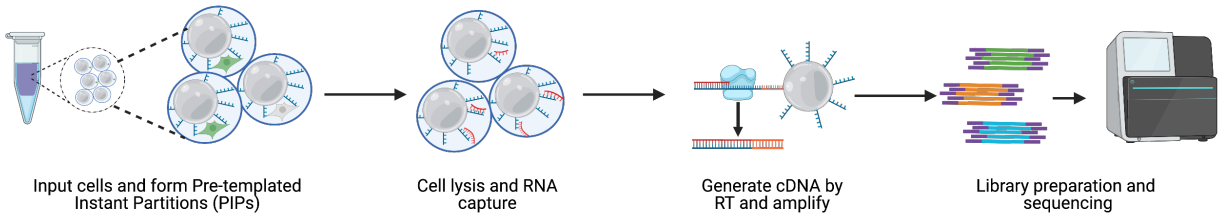
Further analysis was performed on the cleaned dataset, including calculating a percentage of each cell type from the SingleR-annotated cell type counts and a Pearson correlation between these percentages. The 4 example DSP-fixed PIPseq replicates and 2 replicates from a competitor technology were subjected to a transcript length analysis. Mapped transcripts were normalized by transcripts per million (TPM), and matched to their associated reference transcript on the reference genome. Reference transcripts were grouped into 5,000 bins based on their length (in linear space, although the plot axis is $\log_{10}$), and the mean and standard deviation of the TPM was calculated for each bin. Finally, a gene ontology (GO) semantic similarity analysis was performed to quantify the data similarity between DSP-fixed PIPseq and an unfixed competitor sample replicate. Specifically, the top 500 most variable genes in each assay (using SCTransform) were subjected to a GO over-representation analysis to identify significant GO terms. A pairwise similarity matrix was generated for each term in two comparison GO term lists, and combined into a single score using the best-match average (BMA) approach in GOSemSim[9], an established package to calculate semantic similarity. GO terms that were shared in the over-representation analysis independently for DSP-fixed PIPseq and the competitor assay were identified.

## Figure 1:

### Figure 1A:




### Figure 1B:




***PIPseq workflow for DSP-fixed PBMCs. (A)*** *DSP fixation cell preparation workflow for PIPseq (see Protocol 1).*. ***(B)*** *Standard PIPseq workflow. DSP-fixed cells can be input into the same PIPseq assay as unfixed cells.*

## Protocol 1: DSP Fixation of PBMCs

**Required reagents and equipment:**

1. Pierce DSP, no-weigh format
2. Anhydrous DMSO
3. Fluent Cell Suspension Buffer
4. Swinging bucket centrifuge with cooling capability

**Procedure:**

1. Equilibrate DSP from 4C to room temperature for 20 minutes
2. Make 50 mg/mL DSP in anhydrous DMSO. This is the 50X stock.
3. Make 1X DSP solution in Cell Suspension Buffer
4. Resuspend cell pellet in 500 uL 1X DSP and let stand for 30 minutes at room temperature
5. Centrifuge cells at 200 x g for 3 minutes to pellet cells
6. Resuspend cells in 6 mL Cell Suspension Buffer
7. Using a 4C pre-chilled centrifuge, pellet cells at 200 x g for 3 minutes
8. Aspirate the supernatant without disturbing the cell pellet
9. Repeat steps 6-8 two more times for a total of three washes
10. Resuspend cells in 400 uL 4C pre-chilled Cell Resuspension Buffer

## Results

We processed four replicate reactions from one sample of DSP-fixed PBMCs through the PIPseq T2 3' Single Cell RNA Kit and sequenced them on a NextSeq 2000 at approximately 95 million reads per replicate. In total, four replicates captured a total of 3,565 cells at ~108,000 reads per captured cell. Each replicate achieved similar metrics, with an average of ~63% mapping, ~777 median genes/cell, and ~1,082 median transcripts/cell (Table 1). In order to resolve the PBMC cell type distribution, we ran the captured cell count matrix through downstream processing in Seurat and performed automated cell type annotation. The preprocessed gene expression matrix was normalized and projected in two-dimensional space using the Uniform Manifold Approximation and Projection (UMAP) algorithm (FIG. 2). PIPseq resolves all of the major PBMC cell types, including CD4+ T cells (naive and memory), CD8+ T cells, natural killer (NK) cells (also CD8+), B cells, monocytes (CD14+ and CD16+), dendritic cells (DC), and plasma cells.

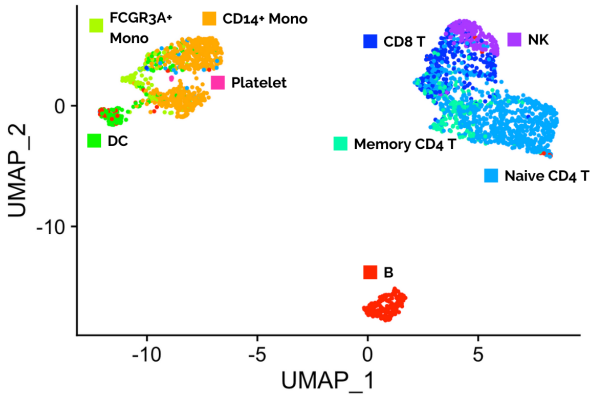
## Table 1:

| | CAPTURED CELLS | READS / CAPTURED CELL | MAPPING RATE (%) | MEDIAN TRANSCRIPTS / CAPTURED CELL | MEDIAN GENES / CAPTURED CELL |
|---|---|---|---|---|---|
| Replicate 1 | 814 | 115,208 | 58.5 | 962 | 737 |
| Replicate 2 | 654 | 116,474 | 57.8 | 1,184 | 864 |
| Replicate 3 | 978 | 99,989 | 66.9 | 1,225 | 828 |
| Replicate 4 | 1,119 | 99,545 | 68.2 | 976 | 707 |
| **Average** | **891** | **107,804** | **62.8** | **1,082** | **777** |

*Key sequencing quality metrics for DSP-fixed PBMCs processed using the Fluent PIPseq T2 3' Single Cell RNA Kit.*

## Figure 2:

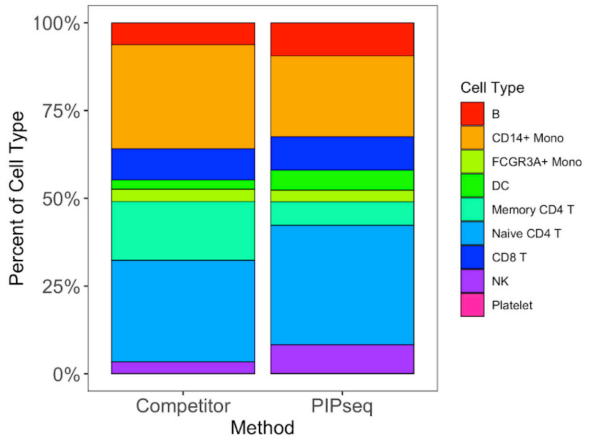
***PIPseq enables immune cell type determination in DSP-fixed PBMCs.***
*After preprocessing, approximately 3,565 were captured. Further quality filtering resolved 2,521 cells, which were annotated by immune cell type and projected in two dimensional space with UMAP. Cell type annotation was performed using SingleR with a Seurat-annotated 10X Genomics v1 PBMC reference.*



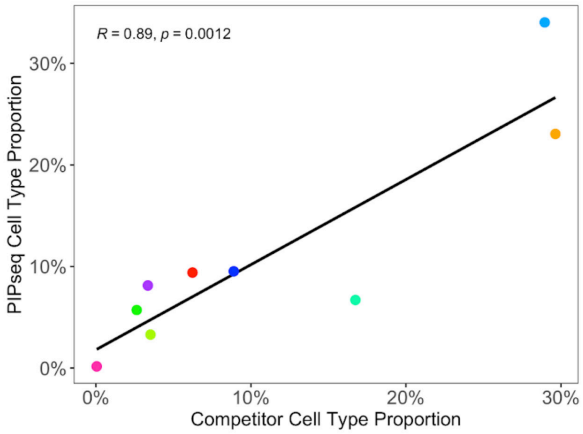
The cell type distribution captured by PIPseq with DSP-fixed cells resembles a leading competitor kit processed using the same cell batch (Figure 3A), with a strong positive correlation (Figure 3B). Specifically, CD4+ T cells were observed in highest abundance, followed by monocytes, CD8+ cells, B cells, DCs, and platelets. PIPseq has a comparable transcript length distribution to the competition, with a slight bias towards shorter transcripts (Figure 3C). In terms of transcript diversity, PIPseq captures similar classes of transcripts to the competitor assay, as seen by the gene ontology (GO) semantic similarity scores (Figure 3D). The competitor-PIPseq comparison (green) yielded a similar score to the positive controls (red/blue), and is significantly different from the negative control (grey). Together, we can see that PIPseq matches the resolution of a key benchmark technology at the level of true biology.

## Figure 3:

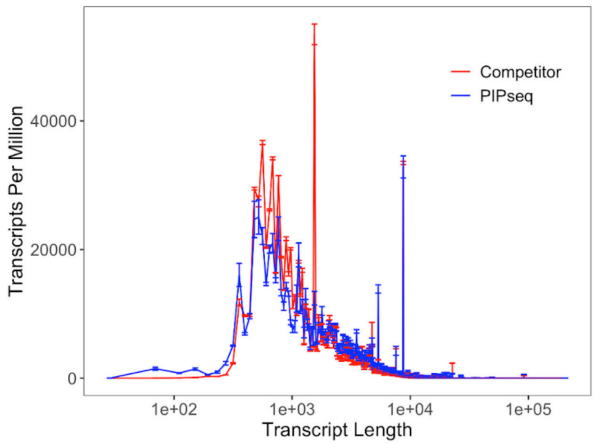
### Figure 3A:



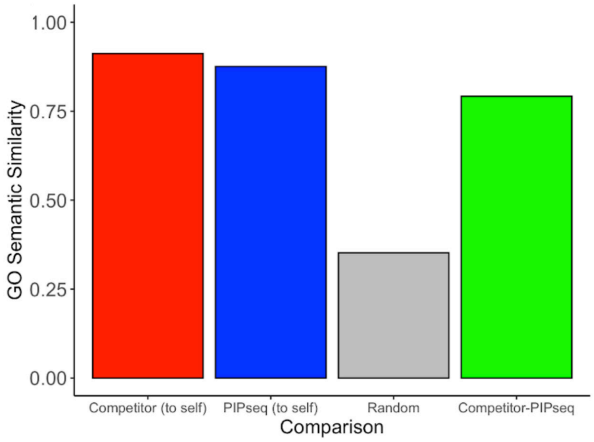

### Figure 3B:




### Figure 3C:




### Figure 3D:




***DSP-fixed PBMCs exhibit similar cell type and transcript distributions to a competitor technology. (A)*** *Cell type distribution for competitor technology (left) vs DSP-fixed PIPseq PBMCs (right).* ***(B)*** *Correlation of PBMC cell type proportions for the competitor technology and DSP-fixed PIPseq PBMCs.* ***(C)*** *Transcript length distribution for competitor technology vs. DSP-fixed PIPseq. Mapped transcripts lengths (in bp) were grouped into 5000 bins and the average TPM across that bin are plotted at the center of that bin with a standard deviation.* ***(D)*** *Semantic similarity score (scale: 0 - 1) of GO terms derived from top 500 variable genes from the competitor assay and DSP-fixed PIPseq assay. Random samplings of 50 GO terms were used for each comparison in order not to bias the score. Self scores (red/blue) were calculated by comparing 2 random samplings (with replacement) of the GO term list from that technology (positive controls), while the random score (grey) was calculated by comparing 2 random samplings (with replacement) of the entire list of possible GO terms (negative control). The direct technology comparison (green) was generated by comparing the random sampling of the competitor and PIPseq GO term lists.*

## Conclusion

We demonstrate that the Fluent BioSciences PIPseq T2 3' Single Cell RNA kit can generate high quality data to resolve immune cell types in DSP-fixed PBMCs. By generating this data, we show our ability to differentiate complex, heterogeneous cell populations with high resolution, a key indicator of success in single-cell technologies. Additionally, we provide evidence that our DSP-fixed cells provide similar information to standard unfixed PBMCs using a leading competitor kit. Our ability to provide high-level biological insights to users from different cell preparation methods will enable new discoveries and make single-cell biology accessible to every researcher.

## References


1. Ding, J., Adiconis, X., Simmons, S.K. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat Biotechnol 38, 737–746 (2020). https://doi.org/10.1038/s41587-020-0465-8
2. Zheng, G., Terry, J., Belgrader, P. et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun 8, 14049 (2017). https://doi.org/10.1038/ncomms14049
3. Macosko, E., Basu, A.. Satija, R. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 5, 161 (2015). http://dx.doi.org/10.1016/j.cell.2015.05.002
4. Klein, A., Mazutis, L., Akartuna, I. et al. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell 5, 161 (2015). https://doi.org/10.1016/j.cell.2015.04.044
5. Rosenberg A., Roco C., Muscat R. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 13, 360 (2018). https://doi.org/10.1126/science.aam8999
6. Aran, D., Looney, A.P., Liu, L. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 20, 163–172 (2019). https://doi.org/10.1038/s41590-018-0276-y
7. Hafemeister, C., Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol 20, 296 (2019). https://doi.org/10.1186/s13059-019-1874-1
8. Traag, V.A., Waltman, L. & van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9, 5233 (2019). https://doi.org/10.1038/s41598-019-41695-z
9. Yu G., Li, F., Qin, Y. et al. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 7, 26 (2010). https://doi.org/10.1093/bioinformatics/btq064