Investigating the Cellular Diversity of Breast Tissue 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, 2, 3]. Depending on the project objectives, tens of thousands of cells may need to be processed to capture the full diversity of cell types within these tissues. 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.

Breast tissue comprises a diverse mixture of epithelial, lymphatic, vascular, and immune cell populations, and the structure and composition of breast tissue remodels continuously throughout a woman's lifetime[4]. Previous scRNA-seq analysis of banked reduction mammoplasty tissue has revealed changes in cell type abundances in response to physiological events such as childbirth and the menstrual cycle[5]. However, current scRNA-seq platforms make it difficult and cost-prohibitive to process a large cohort of asynchronously obtained patient samples. In this White Paper, we apply Fluent's PIPseq T20 3' Single Cell RNA Kit v2.1 to process > 80,000 cells obtained from banked reduction mammoplasty patient samples.

PIPseq Benefits:

No complex instrumentation or consumables required

Quickly capture mRNA within minutes at point of collection

Flexible and scalable

Highly cost-effective

User-friendly bioinformatics portal

Methods

Fluent received four cryopreserved breast tissue cell suspensions from a collaborating laboratory. These samples were obtained from fresh reduction mammoplasty surgeries, subjected to mechanical and enzymatic dissociation, and filtered[5]. The sub-40 micron filtrate was frozen in media (RPMI 1640 with 30% FBS and 10% DMSO) and shipped to Fluent for PIPseq processing. (Figure 1A). Tissue aliquots were thawed and 30,000 input cells per reaction were processed using the Fluent BioSciences PIPseq T20 3' Single Cell RNA Kit v2.1, with each patient assayed in duplicate for a total of eight samples.

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 vortexing. Cells are then lysed on a thermal device and mRNA captured onto the PIPs. Subsequently, the first strand of 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), with primary data analysis performed with the Fluent Cloud Platform.

Figure 1:

Mammoplasty Patients → **Tissue Sample** → **Dissociate into Single Cells** → **PIPseq**
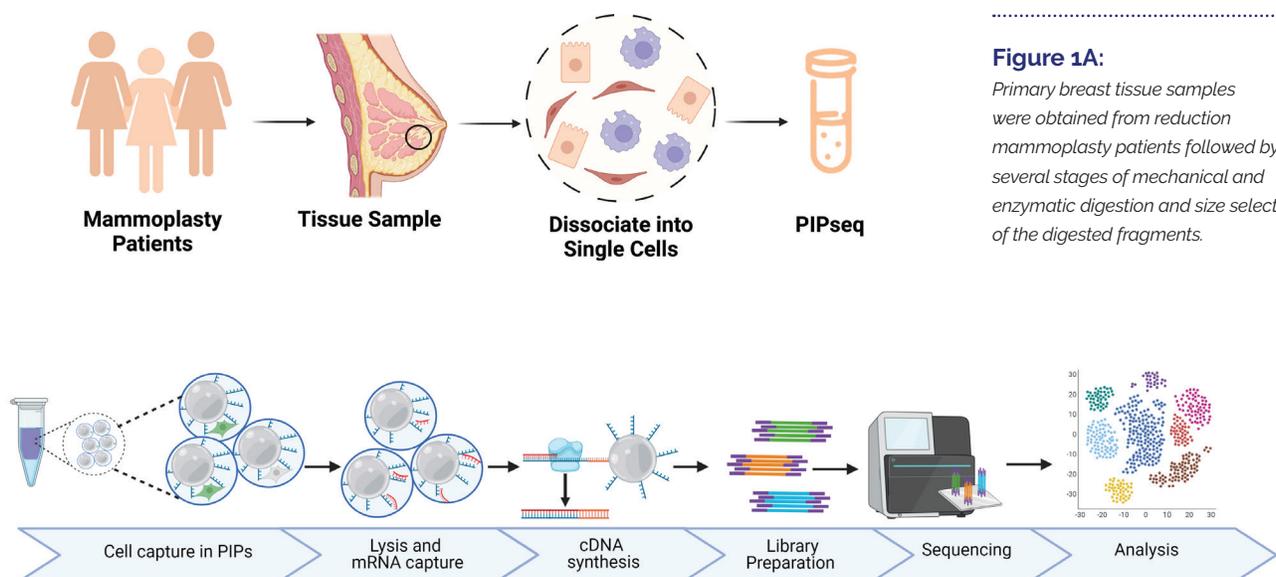
Figure 1A:
Primary breast tissue samples were obtained from reduction mammoplasty patients followed by several stages of mechanical and enzymatic digestion and size selection of the digested fragments.

Cell capture in PIPs → Lysis and mRNA capture → cDNA synthesis → Library Preparation → Sequencing → Analysis

Figure 1B:
Upon receipt of the samples, cells were isolated and input into the PIPseq workflow for processing.

For this project, Fluent sequenced the 8 samples on two P3 100 cycle Illumina NextSeq 2000 cartridges. Following sequencing, samples were processed using the Fluent Cloud Platform to generate gene expression matrices. The Seurat package in R was used to perform downstream bioinformatic analyses.

Table 1:

SAMPLE NAME	CAPTURED CELLS	SEQUENCING DEPTH (READS/CELL)	MEDIAN GENES IN CELLS	MEDIAN UMI IN CELLS
Patient A, Replicate 1	11116	7,300	692	1357
Patient A, Replicate 2	11049	7,300	678	1330
Patient B, Replicate 1	10644	7,300	640	1252
Patient B, Replicate 2	11027	7,300	678	1347
Patient C, Replicate 1	12560	8,200	778	1531
Patient C, Replicate 2	13215	8,200	798	1595
Patient D, Replicate 1	6173	5,700	224	303
Patient D, Replicate 2	6921	5,700	337	489

Summary of sequencing metrics from processed samples, broken down by patient and replicate.

Results & Discussion

Table 1 summarizes the key sequencing metrics obtained after processing through the Fluent Cloud Platform for Data Analysis. At an average sequencing depth of ~8,000 reads per input cell, > 10,000 cells per replicate were obtained for Patients A, B, and C, and > 6,000 cells per replicate from Patient D. For Patients A, B, and C, > 1,000 median transcripts per cell were obtained. In total, > 82,000 cells were captured from the 8 samples. UMAP clustering[6], a commonly used downstream analysis approach for scRNA-seq data, was performed on all cells obtained from the samples. After clustering, major cell types were annotated based on known marker genes based on previous breast tissue[6, 7] and immune cell[8] scRNA-seq datasets (Figure 2A). Analysis of the proportion of these different cell types in the entire dataset and by patient showed expected patient to patient variation in cell type proportions (Figure 2B). It was also observed that Patient C and D have more similar tissue expression profiles when compared to either Patient A or B.

Figure 2:

Figure 2A:
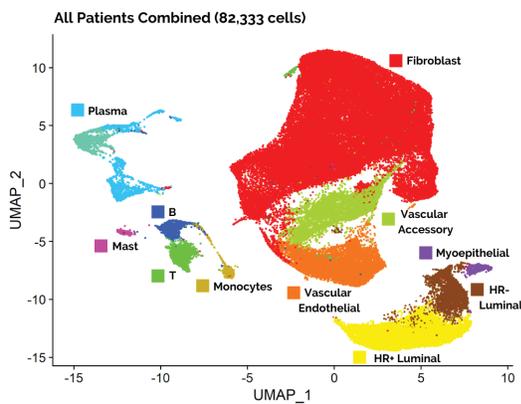
Figure 2B:
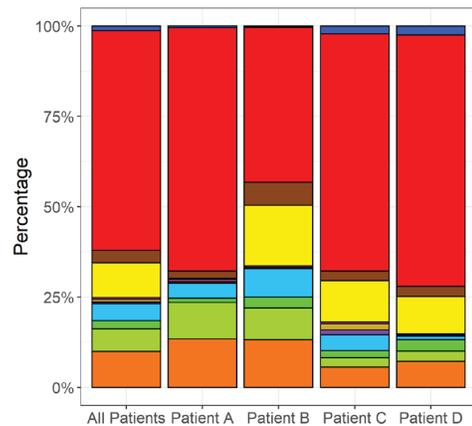
*Gene expression analysis of 82,333 captured cells. **(A)** UMAP analysis was performed on the cells and major cell types identified by cluster-based annotation. **(B)** Proportions of cell types across the entire dataset and broken down by patient.*

Figure 3:

Figure 3A:
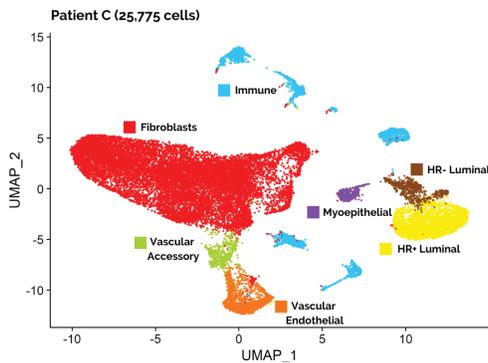
Figure 3B:
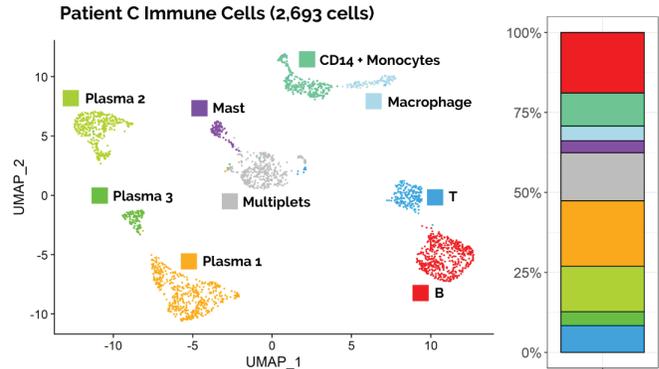
Figure 3C:
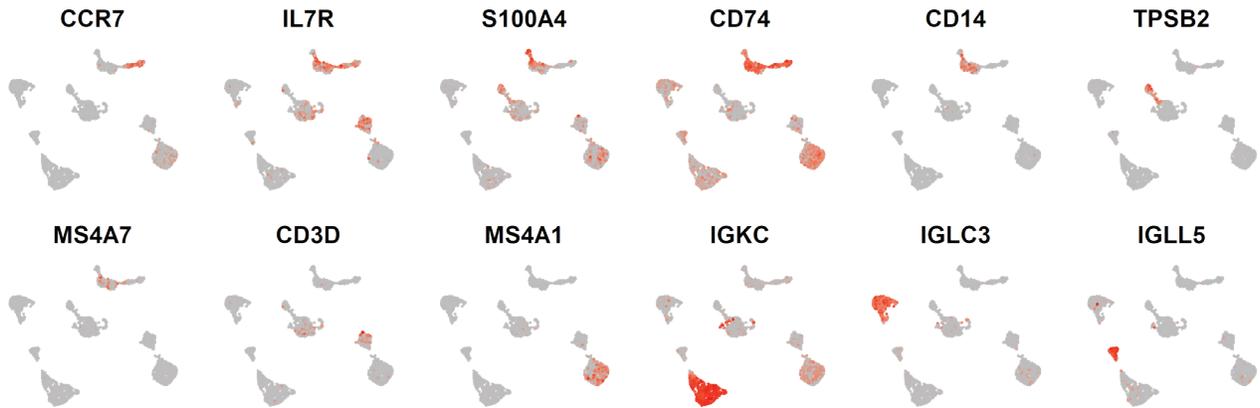
*Focused analysis on 25,775 cells from Patient C. **(A)** UMAP analysis was performed on only those cells from Patient C and cell type identified by cluster-based annotation. **(B)** Re-clustering of just the immune cells that were identified in panel A, with cell type proportions shown on the right. **(C)** Expression of marker genes overlaid on the clustering plot, with grey indicating low expression and red indicating high expression.*

The highest diversity of cells was obtained from Patient C which were subsequently isolated for a deeper analysis. In particular, when clustering the ~25,000 cells from Patient C alone, several well-defined immune cell clusters are observed (Figure 3A). These ~2,600 immune cells were then isolated and re-clustered (Figure 3B). From this secondary clustering, eight distinct immune cell clusters were identified and annotated based on known marker genes[8]. By overlaying the expression of selected marker genes onto the UMAP plot, the expected enrichment of cell-type specific gene signatures was detected (Figure 3C).

Previous work has shown correlations between the cell abundance and state of the epithelial cells within breast tissue to physiological factors[5]. The batch correction algorithm, Harmony[9], was used prior to clustering the > 11,000 observed epithelial cells across eight samples. After clustering, the three expected epithelial cell types are observed along with some additional subtypes (Figure 4A).

When focusing on the expression of selected genes across these cells, significant concordance with the expression patterns of previously identified marker genes[5, 7] across the 3 major epithelial cell types is observed (Figure 4B). Marker genes in high concordance with the following epithelial subtypes were also observed; HR+ Luminal Subtype A correlates with Marker Gene D, HR+ Luminal Subtype B correlates with Marker C and the HR- Luminal Subtype correlates with Marker B. The proportions of these luminal cell subtypes across the four patients (Figure 4C) reveal patient to patient variation. Notably, Patient C and D have similar luminal subtype proportions when compared to Patient A or B.

Figure 4:

Figure 4A:
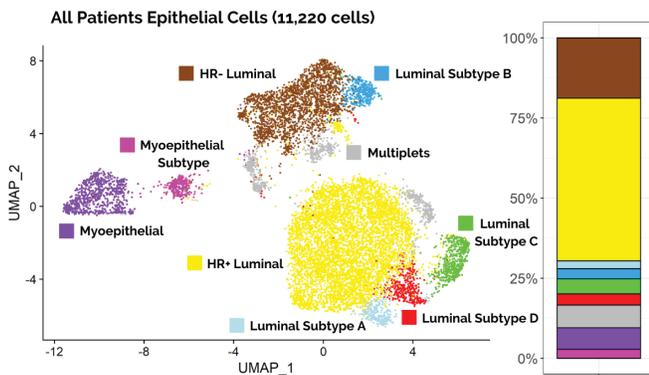
Figure 4B:
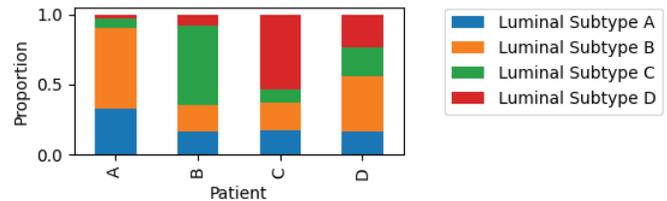
Figure 4C:
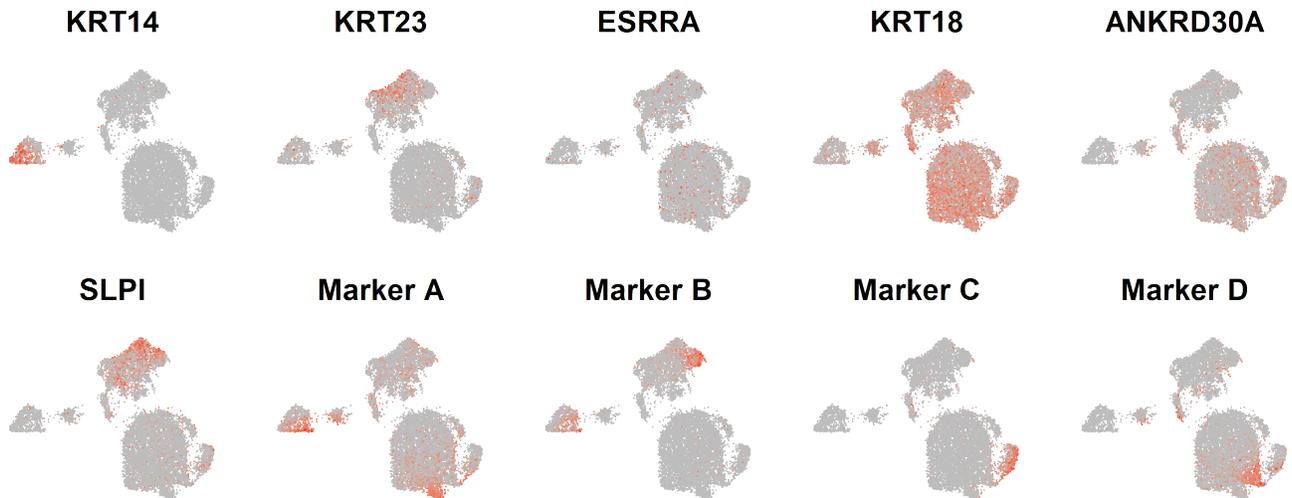
Analysis of 11,220 epithelial cells across the eight sample dataset (A) UMAP clustering was performed on epithelial cells following batch correction using Harmony. Major cell types were identified through cluster-based annotation. (B) Proportion of luminal cell subtypes across the four patients (C) Marker gene expression overlaid on the UMAP plot, with grey indicating low expression and red indicating high expression.

Conclusion

In this White Paper, Fluent has demonstrated the capability of the PIPseq T20 3' Single Cell RNA Kit v2.1 to capture tens of thousands of cells and to clearly resolve cell populations within a complex tissue in a simple, streamlined workflow. In particular, the kit has the ability to provide this resolution at a reasonably shallow sequencing depth if necessary. Further, the product revealed novel biology in these samples, highlighted by patient-specific cell states. Overall, the Fluent BioSciences technology offers a novel instrument-free, easy-to-use and accessible platform that can be implemented in any molecular biology laboratory, and in particular to users interested in performing scRNA-seq analysis on asynchronously obtained primary samples.

References

1. Klein, Allon M., et al. "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells." Cell 161.5 (2015): 1187-1201.

2. Macosko, Evan Z., et al. "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets." Cell 161.5 (2015): 1202-1214.

3. Han, Xiaoping, et al. "Mapping the mouse cell atlas by microwell-seq." Cell 172.5 (2018): 1091-1107.

4. Russo, J., R. Rivera, and I. H. Russo. "Influence of age and parity on the development of the human breast." Breast Cancer Research and Treatment 23.3 (1992): 211-218.

5. Murrow, Lyndsay M., et al. "Changes in epithelial proportions and transcriptional state underlie major premenopausal breast cancer risks." Biorxiv (2020).

6. McInnes, Leland, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction." arXiv:1802.03426 (2018).

7. Nguyen, Quy H., et al. "Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity." Nature Communications 9.1 (2018): 1-12.

8. Zhang, Xinxin, et al. "CellMarker: a manually curated resource of cell markers in human and mouse." Nucleic Acids Research 47.D1 (2019): D721-D728.

9. Korsunsky, Ilya, et al. "Fast, sensitive and accurate integration of single-cell data with Harmony." Nature Methods 16.12 (2019): 1289-1296.

Visit Fluentbio.com to Learn More