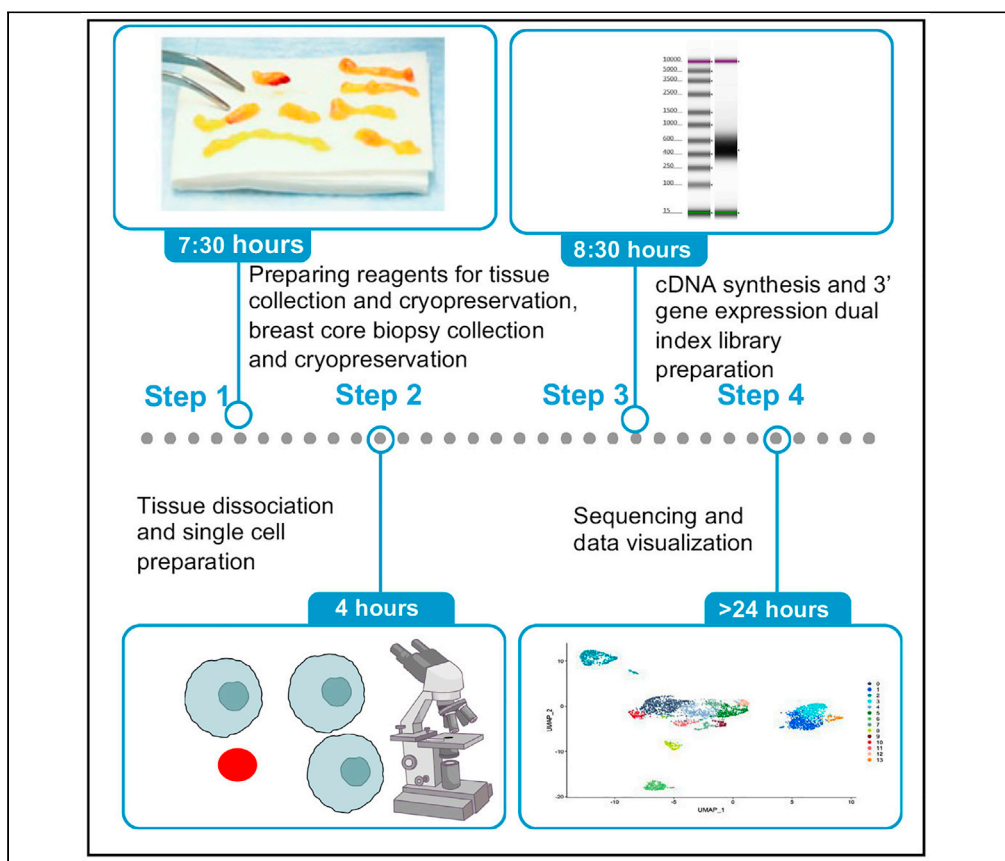


## Protocol

# Acquisition, processing, and single-cell analysis of normal human breast tissues from a biobank



The Komen Tissue Bank is the only biorepository in the world for normal breast tissues from women. Below we report the acquisition and processing of breast tissue from volunteer donors and describe an experimental and analysis pipeline to generate a single-cell atlas. This atlas is based on single-cell RNA-seq and is useful to derive breast epithelial cell subcluster-specific gene expression signatures, which can be applied to breast cancer gene expression data to identify putative cell-of-origin.

Poornima Bhat-Nakshatri, Natascia Marino, Hongyu Gao, Yunlong Liu, Anna Maria Storniolo, Harikrishna Nakshatri

pnakshat@iupui.edu (P.B.-N.)  
hnakshat@iu.edu (H.N.)

### Highlights

Breast core biopsies from healthy donors are suitable for single-cell RNA-seq

Cryopreserved biopsies are as good as fresh biopsies for the analysis

Biopsies from five donors can be combined to yield required cell number

Multiple cell types including epithelial, endothelial, and pericytes can be mapped

Bhat-Nakshatri et al., STAR Protocols 3, 101047  
March 18, 2022 © 2021 The Authors.  
<https://doi.org/10.1016/j.xpro.2021.101047>

## Protocol

## Acquisition, processing, and single-cell analysis of normal human breast tissues from a biobank

Poornima Bhat-Nakshatri,<sup>1,7,\*</sup> Natascia Marino,<sup>2</sup> Hongyu Gao,<sup>3</sup> Yunlong Liu,<sup>3,5</sup>  
Anna Maria Storniolo,<sup>2,5</sup> and Harikrishna Nakshatri<sup>1,4,5,6,8,\*</sup>

<sup>1</sup>Department of Surgery, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>2</sup>Department of Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>3</sup>Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>4</sup>Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>5</sup>Indiana University Melvin and Bren Simon Comprehensive Cancer Center, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>6</sup>Richard L. Roudebush VA Medical Center, Indianapolis, IN 46202, USA

<sup>7</sup>Technical contact

<sup>8</sup>Lead contact

\*Correspondence: [pnakshat@iupui.edu](mailto:pnakshat@iupui.edu) (P.B.-N.), [hnakshat@iu.edu](mailto:hnakshat@iu.edu) (H.N.)  
<https://doi.org/10.1016/j.xpro.2021.101047>

## SUMMARY

The Komen Tissue Bank is the only biorepository in the world for normal breast tissues from women. Below we report the acquisition and processing of breast tissue from volunteer donors and describe an experimental and analysis pipeline to generate a single-cell atlas. This atlas is based on single-cell RNA-seq and is useful to derive breast epithelial cell subcluster-specific gene expression signatures, which can be applied to breast cancer gene expression data to identify putative cell-of-origin.

For complete details on the use and execution of this protocol, please refer to Bhat-Nakshatri et al. (2021).

## BEFORE YOU BEGIN

This protocol describes specific steps for collection and processing of fresh breast tissue cores obtained with the ATEC® Breast Biopsy System for Stereotactic Biopsy vacuum-assisted biopsy system from healthy donors and their utilization to generate a single cell map of the breasts. Biopsies were collected with informed consent and with the approval from the Indiana University Institutional Review Board.

The Graphic abstract provides sequential steps involved in preparing single cells for RNA-seq to data analysis and number of hours required in each of the major steps.

## Preparation of stock solutions for 120 donor tissues

⌚ Timing: 1 h

1. Prepare Insulin (1 mg/mL) Stock (10 mL)
  - a. 1% BSA/PBS stock solution
    - i. In a 50 mL Conical tube dissolve 500 mg of BSA in 50 mL PBS
    - ii. Store at 4°C for up to 7 days



- b. 0.1% BSA/PBS stock solution
  - i. In a 50 mL Conical tube mix 45 mL of sterile PBS + 5 mL of 1% BSA/PBS stock
  - ii. Store at 4°C for up to 7 days
- c. Insulin stock
  - i. In a 15 mL Conical tube dissolve 10 mg of insulin in 10 mL of 0.1% BSA/PBS
  - ii. Store at 4°C for up to 7 days
2. Prepare Epidermal Growth Factor (EGF 1 µg/µL) Stock
  - a. Add 100 µL sterile water into 100 µg of EGF in EGF vial.
  - b. Dissolve and make 10 µL aliquots in 0.2 mL PCR tubes
  - c. Store at -80°C up to 1 year
3. Prepare Hydrocortisone (1 mg/mL) Stock
  - a. Dissolve 1 mg of hydrocortisone in 1 mL 100% EtOH
  - b. Make 200 µL aliquots in 1.5 mL microcentrifuge tubes
  - c. Store at -20°C for up to 1 month
4. Prepare ROCK Inhibitor (10 mM) Stock
  - a. Add 1.56 mL of sterile water to 5 mg ROCK inhibitor powder
  - b. Mix by pipetting
  - c. Aliquot into 300 µL aliquots in 1.5 mL brown Eppendorf tubes
  - d. Store at -20°C for up to 1 month

### Preparation of media for 120 donor tissues

⌚ Timing: 1 h

5. Growth Media for primary cells- make 532.5 mL in the following order ([Liu et al., 2012](#); [Nakshatri et al., 2015](#)). Adding Fetal Bovine Serum (FBS) first prevents Insulin and EGF from sticking to the plastic surface of the bottle.
  - a. F-12: 375 mL use as is
  - b. DMEM (low glucose): 125 mL use as is
  - c. FBS: 25 mL use as is
  - d. Pen/Strep: 5 mL use as is
  - e. Insulin Stock (1 mg/mL): 2.5 mL
  - f. Hydrocortisone Stock (1 mg/mL): 200 µL
  - g. EGF Stock (1 µg/µL): To ensure EGF is well distributed, add a little media from a. through f. mixture to 10 µL EGF stock aliquot and mix by pipetting. This helps in collecting all EGF from the aliquot.
  - h. Transfer EGF aliquot solution to media.
  - i. Filter mix through 0.22 µm Corning 500 mL filter system.
  - j. Store sterile media at 4°C for up to 1 month.

Reagent	Final concentration	Amount
F12	70%	375 mL
DMEM	30%	125 mL
FBS	5%	25 mL
Pen/Strep (10,000 units/mL; 10,000 µg/mL)	100 units/mL Penicillin 100 µg/mL Streptomycin	5 mL
Insulin	2.5 µg/mL	2.5 mL
Hydrocortisone	0.4 µg/mL	200 µL
EGF	20 ng/mL	10 µL

6. Freezing Media- 360 mL (make fresh same day as tissue collection)
  - a. Add 90  $\mu$ L of ROCK Inhibitor (10 mM stock) to 180 mL Lonza BioWhittaker Cryoprotective media in a 500 mL autoclaved bottle and mix
  - b. Add 180 mL Growth Media and mix.
  - c. Aliquot 1 mL into each 2.0 mL labeled cryotube (Prepare 3 cryotubes per tissue)

**Note:** ROCK inhibitors are suggested to prevent apoptosis of stem cells upon dissociation from extracellular matrix during tissue dissociation step (Liu et al., 2012). The growth media described above was originally described for growing primary epithelial cells and we had modified this media by substituting high glucose DMEM (4,500 mg/L glucose) with low glucose DMEM (1000 mg/L glucose), which we found extends replication potential of primary cells (Nakshatri et al., 2015). We have not compared the results if regular cell culture media such as MEM plus FBS or DMEM plus FBS used for growing cancer cell lines instead of specific media described above are used for tissue collection and cryopreservation.

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological Samples</b>		
Breast biopsies, human, female (age range- 21–60)	Volunteer donors	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
Lonza BioWhittaker Cryoprotective™ Media	Fisher Scientific	BW12-132-A
Y-27632 Inhibitor (ROCK Inhibitor)	Fisher Scientific	50-103-1737
F-12	Life Technologies	11765-054
DMEM low glucose	Life Technologies	12320-032
Fetal Bovine Serum (FBS)	Sigma-Aldrich	F6178
Hydrocortisone	Sigma-Aldrich	H0888
Penicillin-Streptomycin	Sigma-Aldrich	P4333
Insulin	Sigma-Aldrich	I6634
Bovine Serum Albumin (BSA)	Sigma-Aldrich	A4503
HyClone™ Phosphate Buffered Saline (PBS) 1x	Fisher Scientific	SH3025601
Epidermal growth factor (EGF), human recombinant	Merck Millipore	01-107
Ethanol, pure (200 proof, anhydrous)	Millipore Sigma	E7023-500 mL
10% Tween 20	Bio-Rad Laboratories	1662404
Glycerin (glycerol), 50% (v/v) Aqueous solution	RICCA Chemical Company	3290-32
<b>Critical Commercial Assays</b>		
Low TE Buffer	Thermo Fisher Scientific	12090-015
SPRIselect Reagent Kit	Beckman Coulter	B23318
Buffer EB	Qiagen	19086
Red blood cell lysis buffer (10x)	Miltenyi Biotec	130-094-183
Dead cell removal kit	Miltenyi Biotec	130-042-303
Tumor dissociation kit	Miltenyi Biotec	130-094-183
Agilent High Sensitivity DNA kit	Agilent	5067-4626
Agilent High Sensitivity DNA Reagents	Agilent	5067-4627
Chromium next GEM single cell 3' Reagent kits v3.1 kits 3.1	10x Genomics	PN-1000121
Chromium Next GEM chip G single cell kit	10x Genomics	PN-1000127
Single index kit T set A	10x Genomics	PN-1000213
NovaSeq 6000	Illumina	<a href="https://www.illumina.com/systems/sequencing-platforms/novaseq.html">https://www.illumina.com/systems/sequencing-platforms/novaseq.html</a>

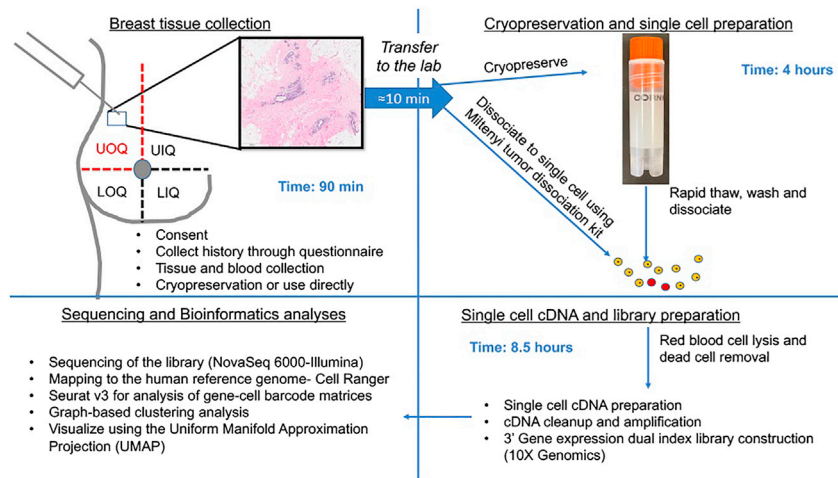
(Continued on next page)

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Software and Algorithms</b>		
Cell Ranger mkfastq	10x Genomics	<a href="https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/mkfastq">https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/mkfastq</a>
cellranger count	10x Genomics	<a href="https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/count">https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/count</a>
human reference genome GRCh38	UCSC Genome Browser and 10x Genomics	<a href="https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation">https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation</a>
Seurat v3	Open source	<a href="https://satijalab.org/seurat/index.html">https://satijalab.org/seurat/index.html</a>
isOutlier function from the R package scatter	Open source	<a href="https://github.com/jimhester/scater">https://github.com/jimhester/scater</a>
<b>Deposited Data</b>		
Dataset	This paper	<a href="https://data.humancellatlas.org/explore/projects/a004b150-1c36-4af6-9bbd-070c06dbc17d">https://data.humancellatlas.org/explore/projects/a004b150-1c36-4af6-9bbd-070c06dbc17d</a>
Dataset	This paper	<a href="https://cellxgene.cziscience.com/collections/c9706a92-0e5f-46c1-96d8-20e42467f287">https://cellxgene.cziscience.com/collections/c9706a92-0e5f-46c1-96d8-20e42467f287</a>
<b>Other</b>		
ATEC® Breast Biopsy System Console	Hologic	N/A
ATEC 9 Gauge x 12 cm Biopsy Needle	Hologic	09612-20
Corning 500 mL Filter System 0.22 µm	Corning	430186
CoolCell FTS30	Corning	CLS-CC-097
50 mL Conical Centrifuge Tubes	Fisher Scientific	05-539-8
15 mL Conical Centrifuge Tubes	Fisher Scientific	05-539-5
2.0 mL Cryovial	Corning	430488
1.5 mL Microcentrifuge tube	Fisher Scientific	05-408-129
1.5 mL Microcentrifuge tube Brown	Fisher Scientific	05-408-134
0.2 mL PCR tubes	Fisher Scientific	14-230-225
Telfa Cotton 2 x 3 Sterile Dressing	McKesson	9962
Wound Wash Saline	McKesson	636352
Specimen Cups	Fisher Scientific	14-955-103
Petri Dishes	Fisher Scientific	08-772E
Sterile Forceps	Fisher Scientific	NC9812170
Scalpels #11	Fisher Scientific	08-927-5B
10x genomics Chromium Controller	10x Genomics	110211
Thermal Cycler	Bio-Rad Laboratories	T100
Vortexer	Thermo Fisher Scientific	12-812
8-tube strips	Eppendorf	951010022
DNA LoBind tubes (1.5 mL)	Eppendorf	022431021
DNA LoBind tubes (2.0 mL)	Eppendorf	022431048
Tips LTS 200 µL Filter RT-L200FLR	Rainin	30389240
Tips LTS 1 mL Filter RT-L1000FLR	Rainin	30389213
Tips LTS 20 µL Filter RT-L10FLR	Rainin	30389226
Nuclease-free water	Thermo Fisher Scientific	AM9937
Debris removal filter	CellTrics	04-004-2326
Cell Counter	DeNovix	CellDrop

**Note:** Due to high cost associated with the experiments and precious nature of tissues used for the study, we did not test reagents from sources other than listed in the table and recommended by 10x Genomics.

## STEP-BY-STEP METHOD DETAILS

Figure 1 provides a schematic view of major steps associated with this protocol.



**Figure 1. Complete overview of the study starting with tissue collection to data analysis**

LIQ, Lower Inner Quadrant; LOQ, Lower Outer Quadrant; UIQ, Upper Inner Quadrant; UOQ, Upper Outer Quadrant.

## Breast tissue collection

⌚ Timing: 30 min

Breast Tissue is acquired following proper consent, medical questionnaire, risk review and eligibility confirmation from the healthy volunteer donors during a Komen Tissue Bank collection event. <https://komentissuebank.iu.edu/donate-tissue/donation-procedure.php>. Necessary tubes, labeling material for cryopreservation, and growth media for cryopreservation are prepared prior to the tissue collection event. However, cryopreservation media (50% growth media plus 50% Bio-Whittaker cryoprotective media) is made fresh the day of tissue collection.

1. Take a predetermined number of core samples from the upper outer quadrant of the normal breast, (donor and/or MD choice left or right), using the ATEC® Breast Biopsy System (Hologic Inc.) and ATEC 9G × 12 cm biopsy needle (Figure 1). Breast tissue volume from this Bx system averages 120 mg/core based on our lab measurements.
2. Remove the cores from the biopsy hand piece, place them on telfa pad, spray with wound wash saline and place into a specimen cup. Then, transport the cores to the tissue processing area within 10 min (or less) of procurement.
3. Immediately place one core in a 50 mL conical tube prepared with 5 mL Lonza BioWhittaker Cryoprotective Media for cryopreservation processing.
4. If a fresh core is requested, place one core in a 50 mL conical tube prepared with 5 mL researcher provided growth media described above.

## Cryopreservation of normal breast tissue

⌚ Timing: 5 h

Cryopreservation of a single approximately 120 mg core of normal breast tissue into 3 aliquots.

5. In a sterile environment, pour the growth media with a single core sample into a fresh sterile petri dish. Transfer the core to the petri dish lid or a new dish to separate the core from the media.
6. With a sterile scalpel, cut the core into multiple small pieces (<2 mm).

7. Using sterile forceps, divide the cut pieces into 3 equal aliquots. Add each tissue aliquot to one of three labelled cryotubes prepared with 1 mL Freezing Media.
8. Discard media remaining in the petri dish into a media waste bottle.
9. Place the 3 cryotubes with tissue and freezing media into slots in a Corning® CoolCell FTS30 Cryo Container following manufacturer instructions for controlled-rate freezing. Once the cryo-container is full (30 tubes/10 tissue samples) it is placed into dry ice until it can be transferred to the  $-80^{\circ}\text{C}$  freezer. If less than 30 vials are to be frozen, fill the empty positions with blanks containing 1 mL freezing media. Our collection event typically collects tissues from  $\sim 100$  donors in six hours and all can be cryopreserved.
10. After at least 4 h in  $-80^{\circ}\text{C}$ , or within one week, transfer cryotubes to Liquid Nitrogen vapor phase freezer for long term storage ( $-140^{\circ}\text{C}$  to  $-190^{\circ}\text{C}$ ).

### Preparing single cells from cryopreserved or freshly collected breast biopsies

⌚ Timing: 4 h

11. Thaw cryopreserved tissue in a  $37^{\circ}\text{C}$  water bath with rigorous shaking for 3–5 min. As with cell lines, thawing has to be quick.
12. Transfer tissue to petri dish with 30 mL of PBS using a forceps. Wash tissues by shaking in PBS for 10 s. Transfer tissues to the petri dish lid or a new petri dish.
13. Chop tissues into tiny pieces using a scalpel or blade.
14. Use human tumor dissociation kit (Miltenyi Biotec) to dissociate the tissue.
15. To remove erythrocytes, add 10 mL  $1\times$  red blood cell lysis solution (Miltenyi Biotec) and follow the protocol from Miltenyi Biotec for red blood cell lysis (<https://www.miltenyibiotec.com/upload/assets/IM0022139.PDF>).
16. Examine cells under the microscope for live cells and debris. We use DeNovix cell drop FL (Fluorescence cell counter) for both live cell counting and debris visualization. If a lot of debris is still left, filter the sample through  $30\ \mu\text{m}$  filters from Celltrics. If still dead cells are left, use the dead cell removal kit from Miltenyi Biotec and follow their protocol (<https://www.miltenyibiotec.com/upload/assets/IM0013462.PDF>).
17. Count the cell number. Resuspend cells at 700–1200 cells per microliter.
18. Follow the  $10\times$  Genomics protocol for chromium next GEM (Gel Bead-in-Emulsion) single cell 3' kit (<https://support.10xgenomics.com/single-cell-gene-expression/library-prep/doc/user-guide-chromium-single-cell-3-reagent-kits-user-guide-v3-chemistry>).

### Single cell cDNA synthesis

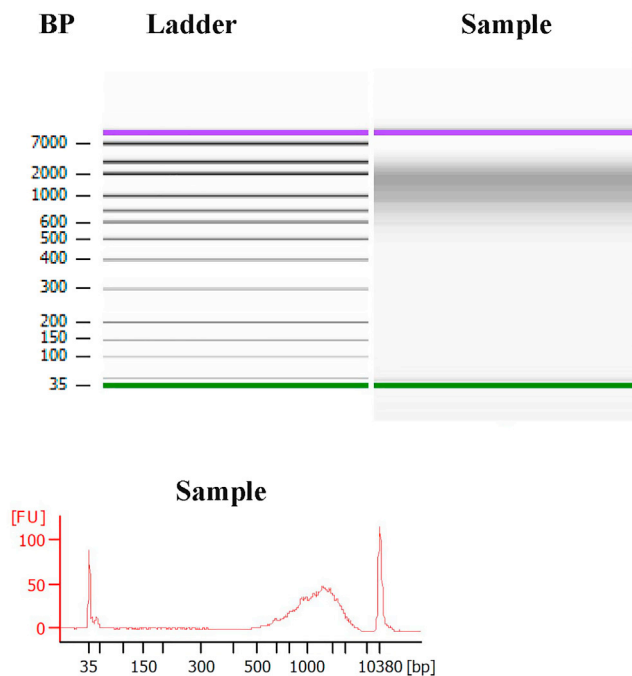
⌚ Timing: 2 h, 8 samples can be run at a time, but our experience is up to four samples at a time

Single cell cDNA preparation is performed as per protocol described in the  $10\times$  Genomics protocol for chromium next GEM single cell 3' kit without any modifications. Although the protocol suggests using  $\sim 1200$  cells per microliter for targeted recovery of 10,000 cells, we have performed the assays with cells as low as 100 cells per microliter for targeted cell recovery of 2000 cells.

**⚠ CRITICAL:** The protocol describes with figures where this process can go wrong. Reagents need to be added in a specified row and in a specific order for GEM generation. After the run, check the volume in row 3. Less than  $100\ \mu\text{L}$  of GEMs in row 3 indicates a clog in wells of row 1 or row 2.

### GEM-RT cleanup and cDNA amplification

⌚ Timing: 1.5 h



**Figure 2. Quantification and quality assessment of single cell cDNA**

The majority of cDNAs were 500–2000 bp long. The bottom panel shows distribution patterns of cDNAs based on size. BP, base pair.

Perform GEM-RT Cleanup and cDNA amplification as described in the 10× Genomics protocol for chromium next GEM single cell kit (<https://support.10xgenomics.com/single-cell-gene-expression/library-prep/doc/user-guide-chromium-single-cell-3-reagent-kits-user-guide-v3-chemistry>).

### cDNA cleanup, QC and quantification

⌚ Timing: 1 h

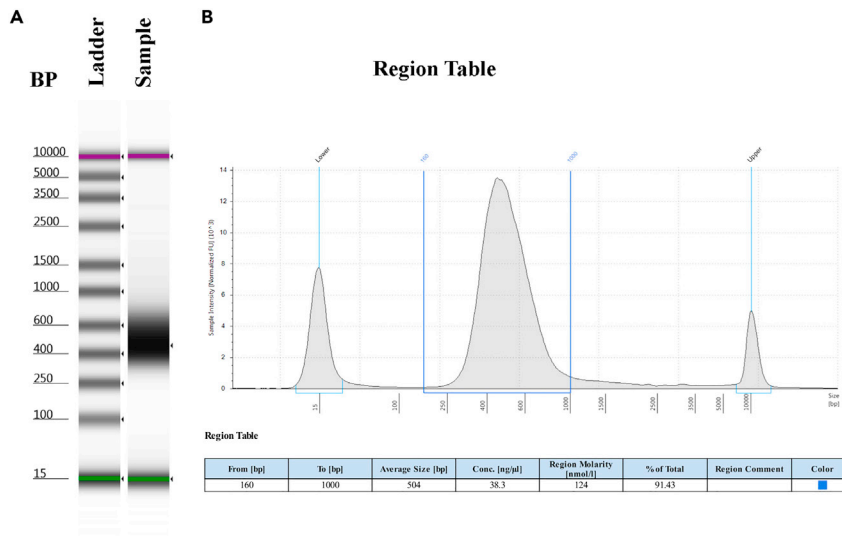
19. Perform cDNA cleanup as per the protocol described in the 10× Genomics protocol for chromium next GEM single cell kit.
20. Perform Quality Control (QC) and quantification using the Agilent Bioanalyzer High Sensitivity kit and reagents (<https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-dna-kits-reagents/bioanalyzer-high-sensitivity-dna-analysis-228249>). Assays are done as per manufacturer’s instructions ([https://www.agilent.com/cs/library/usermanuals/public/HighSensitivity\\_DNA\\_KG.pdf.pdf](https://www.agilent.com/cs/library/usermanuals/public/HighSensitivity_DNA_KG.pdf.pdf)). Representative cDNA prep quality is shown in Figure 2. Size of cDNA ranged from 500 bp to 10,000 bp.

### 3' gene expression dual index library construction

⌚ Timing: 4 h

21. Prepare Index library using the protocol suggested by 10× Genomics without any modifications (<https://support.10xgenomics.com/single-cell-gene-expression/library-prep/doc/user-guide-chromium-single-cell-3-reagent-kits-user-guide-v3-chemistry>).
22. Determine fragment size using the Agilent Bioanalyzer High Sensitivity kit and reagents. Representative library size is shown in Figure 3 (400–600 base pair range).





**Figure 3. Quantification and quality assessment of 3' Gene Expression dual index library before sequencing**

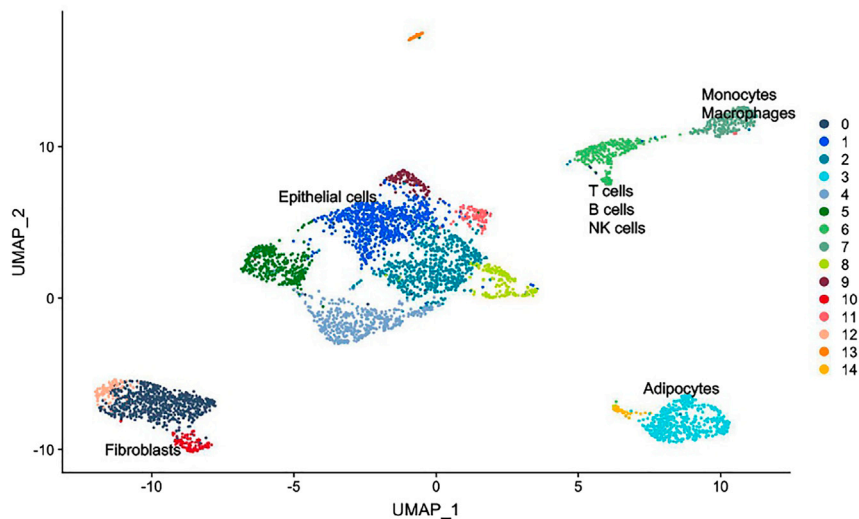
(A) Gel electrophoresis of the library.

(B) Size distribution of DNA in the library. Peak size was 500 bp.

## Sequencing, mapping and data analysis

### Sequencing and data analysis

23. Sequence the final libraries on a NovaSeq 6000 (Illumina) with the sequencing configuration: read 1, 26 or 28 cycles; i7 index read, 8 cycles; read 2, 91 cycles.
24. After the sequence data are generated, use Cell Ranger mkfastq (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/mkfastq>) to demultiplex the raw base sequence calls into sample-specific FASTQ files. Apply cellranger count (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/count>) to align the data to the human reference genome GRCh38 and generate gene-cell barcode count matrices. Cellranger count was run on the Indiana University's high-throughput computing clusters with 8 cores and memory of 64 G. In most cases, the default parameters are sufficient. To download and install Cell Ranger, there are detailed step-by-step instructions at the 10x genomics website (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation>).
25. Use Seurat v3 (Butler et al., 2018; Stuart et al., 2019) (<https://satijalab.org/seurat/index.html>) to further analyze the filtered gene-cell barcode matrices generated with Cell Ranger. The following steps are performed for each dataset:
  - a. First step is to implement quality control (QC) of the data using the isOutlier function from the R package scatter (McCarthy et al., 2017) (<https://github.com/jimhester/scater>) together with visual inspection of the distributions of number of genes, Unique Molecular Identifiers (UMIs), and mitochondrial gene content. One can follow the Seurat tutorial for single sample analysis at [https://satijalab.org/seurat/articles/pbmc3k\\_tutorial.html](https://satijalab.org/seurat/articles/pbmc3k_tutorial.html).
  - b. Normalize the gene expression levels for each cell with the Normalize Data function (normalization.method = "LogNormalize") in Seurat. Default parameters can be applied.
  - c. Identify highly variable features using the FindVariableFeatures function in Seurat. Default parameters can be applied.
26. To integrate the single cell data from individual donor samples, the following steps are performed (Seurat tutorial for integration analysis: [https://satijalab.org/seurat/articles/integration\\_introduction.html](https://satijalab.org/seurat/articles/integration_introduction.html)):
  - a. Apply functions FindIntegrationAnchors and IntegrateData from Seurat.



**Figure 4. A representative UMAP of breast tissues obtained from a 35 years-old woman of European Ancestry.**

- b. Scale the integrated data and run principal component analysis (PCA) with the Seurat functions `ScaleData` and `RunPCA`. Determine the appropriate number of PCs to use.
- c. Carry out graph-based clustering analysis. Identify clusters with the Seurat functions `FindNeighbors` and `FindClusters` using a resolution of 0.8 and 36 PCs. Don't want to over-cluster or under-cluster the cells. However, there is no easy way to determine the number of dimensions to use. Different resolutions and number of PCs can be tried. There are few recommendations in the Seurat website ([https://satijalab.org/seurat/articles/pbmc3k\\_tutorial.html](https://satijalab.org/seurat/articles/pbmc3k_tutorial.html)).
- d. Detect cell type marker genes with the Seurat function `FindConservedMarkers`.
- e. Visualize cell clusters using the Uniform Manifold Approximation and Projection (UMAP). A representative UMAP is shown in [Figure 4](#). Please also refer to [Figures 1A, 3C, and 4](#) of [Bhat-Nakshatri et al.](#) for detailed UMAP of breast tissue ([Bhat-Nakshatri et al., 2021](#)).
- f. Annotate cell clusters using the cell-type specific marker genes identified and known canonical marker genes. For example, for breast epithelial cells, *CD49f/EpCAM* expression pattern can classify them into basal (*CD49f+/EpCAM-*), luminal progenitor (*CD49f+/EpCAM+*), and mature luminal (*CD49f-/EpCAM+*) cells ([Lim et al., 2009](#); [Visvader and Stingl, 2014](#)). Further validation of this classification could be achieved based on the expression patterns of *ESR1* and *FOXA1* (mature luminal), *CD117*, *ELF5*, and *EHF* (luminal progenitor) and *TP63* and *NFIB* (basal cells) ([Lim et al., 2009, 2010](#); [Pellacani et al., 2016](#)). Similarly endothelial cells can be annotated using *PECAM*, *CDH5* and *VWF* expression pattern ([Zhao et al., 2021](#)). Fibroblasts can be annotated using *Collagen 1A2* (*COL1A2*) and *COL3A1* expression pattern ([Muhl et al., 2020](#)). Adipocytes express *ADIPOQ* and *Apolipoprotein E* (*APOE*) ([Zhong et al., 2020](#)).
- g. Use R packages `ggplot2` ([Wickham, 2016](#)) and `Seurat FeaturePlot` to generate plots of specific genes.

## EXPECTED OUTCOMES

This method describes steps involved in the development of human breast atlas at single cell level (mRNA-based) using breast tissues from healthy donors. Quality of single cell cDNA and library for sequencing is shown in [Figures 2 and 3](#). Breast tissue is composed of basal, luminal progenitor and mature luminal breast epithelial cells, endothelial cells, fibroblasts, adipocytes and immune cells. The assay also allowed further refinement of basal, luminal progenitor and mature luminal epithelial cells into various subclusters and gene expression signatures of these subclusters are useful to assess different subpopulation of breast epithelial cells in intact breast tissues using spatial proteomics techniques in future. Unlike the commonly used abundant breast tissues from reduction

mammoplasty or normal tissues adjacent to tumors, which are often histologically and molecularly abnormal (Degnim et al., 2012; Teschendorff et al., 2016), this method describes feasibility of using breast biopsies from healthy donors. The method described here also allows tissue analysis utilizing standard operating procedure for tissue collection, which reduces variability in sites of tissue collection and collection to analysis time. Furthermore, cryopreserved samples and freshly collected samples provided similar results. These results have important implications as samples could be collected at institutions with no facilities for single cell analysis using standard operating procedure for tissue collections and cryopreserved samples can then be transferred to facilities that can perform single cell analysis.

## LIMITATIONS

Because of limited tissue availability and epithelial cellularity can vary from sample to sample, not all preparations yield enough cells to continue single cell analysis. This limitation can be overcome by combining breast tissues from multiple donors. However, if the research focus is in developing a single cell map at individual level, an additional step of deconvoluting data using genomic sequence information of the donor is needed. That step adds additional cost for sequencing and bioinformatics analysis. Since breast biopsies are collected from a specified region of the breast, the single cell atlas may correspond to the specific region of the breast instead of representing the entire breast.

## TROUBLESHOOTING

### Problem 1

Epithelial cellularity varies with a significant number of samples enriched for adipocytes or with excessive extracellular matrix deposits. The example shown [Figure 4](#) represents a sample with good epithelial cellularity. Although all cryopreserved samples for the single cell analysis were selected based on the results of Hematoxylin/Eosin (H&E) staining that indicated epithelial cellularity as high, moderate, or low (those with high and moderate cellularity were selected), there were discrepancies with respect to epithelial cellularity assessment by H&E staining and UMAP (step: [breast tissue collection](#) and [preparing single cells from cryopreserved or freshly collected breast biopsies](#)).

### Potential solution

Tissues from five donors (two cores each) can be combined to obtain a sufficient number of cells and to increase epithelial cellularity. Individual samples can be identified through sequence information. Cryopreserved samples work as good as freshly collected samples, which give tremendous flexibility for initiating single cell analysis and would allow initial quality assessment of samples by H&E staining. Our studies utilized tissues from healthy donors. However, in studies that involve breast tumors, mixing samples may not be a better approach as it may result in loss of information regarding tumor heterogeneity.

### Problem 2

Too many red blood cells (step: [preparing single cells from cryopreserved or freshly collected breast biopsies](#))

### Potential solution

Although total cell yield may come down, it is better to repeat red blood cell lysis and dead cell removal before proceeding to the next step.

### Problem 3

Too many dead cells (step: [preparing single cells from cryopreserved or freshly collected breast biopsies](#)).

### Potential solution

Repeat the debris removal step. The protocol provides a table that describes optimal range of cell loading for achieving the desired cell recovery target of 10,000. We did not proceed further if debris still persisted and red blood cells could not be removed after two attempts and live cells were less than 100 cells per microliter.

### Problem 4

cDNA synthesis step did not appear to work as tubes in row 3, which are expected to contain cDNAs in 100  $\mu$ L of volume, had lower volume (step: [single cell cDNA synthesis](#)).

### Potential solution

Cells >1200 per microliter or manufacturing defects in chips used may have caused a clog and prevented partition. Recount cells and load as suggested in the protocol. If the problem persists, contact 10 $\times$  Genomics for trouble shooting with the machine or chips provided for the analysis.

### Problem 5

Sequencing reveals ambient RNA contamination or doublet contamination (step: [sequencing and data analysis](#)).

### Potential solution

Ambient RNA contamination is due to cell lysis before partitioning. Doublet contamination is due to incomplete dissociation. Data from such experiments are no longer ideal for analysis. Reducing time between tissue collection and processing can reduce cell lysis. Doublet issue can be resolved by flow cytometry based sorting of cells although sorting increases the time gap between tissue collection and cDNA synthesis. Expression of stress-associated genes increases with any increase in time gap and additional steps. A recent review by Lim et al. provides tips on what to avoid in the tissue processing step for obtaining best quality results ([Lim et al., 2020](#)).

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, [Harikrishna Nakshatri] ([hnaakshat@iupui.edu](mailto:hnaakshat@iupui.edu)).

### Materials availability

This study did not generate any new reagents.

### Data and code availability

No new codes have been generated. Data related to this project are available at <https://data.humancellatlas.org/explore/projects/a004b150-1c36-4af6-9bbd-070c06dbc17d> and <https://cellxgene.cziscience.com/collections/c9706a92-0e5f-46c1-96d8-20e42467f287>.

## ACKNOWLEDGMENTS

Chan-Zuckerberg Initiative and Catherine Peachy funds at the Heroes Foundation supported this work. Susan G. Komen for the Cure, USA, provided funding support for the Susan G. Komen Normal tissue bank. We thank members of Center for Medical Genomics for sequencing, staff of the Susan G. Komen Normal tissue bank, particularly Pam Rockey and Jill Henry, volunteers for enabling tissue collection and the countless number of women who donated breast tissue for research.

## AUTHOR CONTRIBUTIONS

Methodology, P.N. and H.G.; Tissue acquisition, N.M. and A.M.S.; Investigation, H.N.; Writing, P.N., H.G., and H.N.; Funding Acquisition, A.M.S., Y.L., and H.N.; Supervision, A.M.S., Y.L., and H.N.

## DECLARATION OF INTERESTS

Authors have no conflict of interest to disclose.

## REFERENCES

- Bhat-Nakshatri, P., Gao, H., Sheng, L., McGuire, P.C., Xuei, X., Wan, J., Liu, Y., Althouse, S.K., Colter, A., Sandusky, G., et al. (2021). A single-cell atlas of the healthy breast tissues reveals clinically relevant clusters of breast epithelial cells. *Cell Rep. Med.* **2**, 100219.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420.
- Degnim, A.C., Visscher, D.W., Hoskin, T.L., Frost, M.H., Vierkant, R.A., Vachon, C.M., Shane Pankratz, V., Radisky, D.C., and Hartmann, L.C. (2012). Histologic findings in normal breast tissues: comparison to reduction mammoplasty and benign breast disease tissues. *Breast Cancer Res. Treat.* **133**, 169–177.
- Lim, B., Lin, Y., and Navin, N. (2020). Advancing cancer research and medicine with single-cell genomics. *Cancer Cell* **37**, 456–470.
- Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A., et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913.
- Lim, E., Wu, D., Pal, B., Bouras, T., Asselin-Labat, M.L., Vaillant, F., Yagita, H., Lindeman, G.J., Smyth, G.K., and Visvader, J.E. (2010). Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* **12**, R21.
- Liu, X., Ory, V., Chapman, S., Yuan, H., Albanese, C., Kallakury, B., Timofeeva, O.A., Nealon, C., Dakic, A., Simic, V., et al. (2012). ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am. J. Pathol.* **180**, 599–607.
- McCarthy, D.J., Campbell, K.R., Lun, A.T., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186.
- Muhl, L., Genove, G., Leptidis, S., Liu, J., He, L., Mucci, G., Sun, Y., Gustafsson, S., Buyandelger, B., Chivukula, I.V., et al. (2020). Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat. Commun.* **11**, 3953.
- Nakshatri, H., Anjanappa, M., and Bhat-Nakshatri, P. (2015). Ethnicity-dependent and -independent heterogeneity in healthy normal breast hierarchy impacts tumor characterization. *Sci. Rep.* **5**, 13526.
- Pellacani, D., Bilenky, M., Kannan, N., Heravi-Moussavi, A., Knapp, D.J.H.F., Gakkhar, S., Moksa, M., Carles, A., Moore, R., Mungall, A.J., et al. (2016). Analysis of normal human mammary epigenomes reveal cell-specific active enhancer states and associated transcription factor networks. *Cell Rep.* **17**, 2060–2074.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21.
- Teschendorff, A.E., Gao, Y., Jones, A., Ruebner, M., Beckmann, M.W., Wachter, D.L., Fasching, P.A., and Widschwendter, M. (2016). DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat. Commun.* **7**, 10478.
- Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev.* **28**, 1143–1158.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag)978-3-319-24277-4.
- Zhao, G., Lu, H., Liu, Y., Zhao, Y., Zhu, T., Garcia-Barrio, M.T., Chen, Y.E., and Zhang, J. (2021). Single-cell transcriptomics reveals endothelial plasticity during diabetic atherogenesis. *Front. Cell Dev. Biol.* **9**, 689469.
- Zhong, L., Yao, L., Tower, R.J., Wei, Y., Miao, Z., Park, J., Shrestha, R., Wang, L., Yu, W., Holdreith, N., et al. (2020). Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *Elife* **9**, e54695.