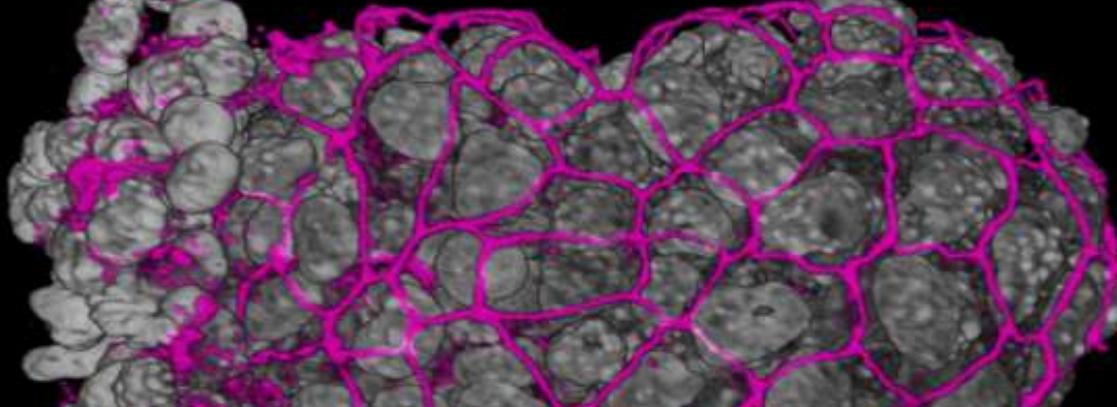


# eBook: Single-cell transcriptomics



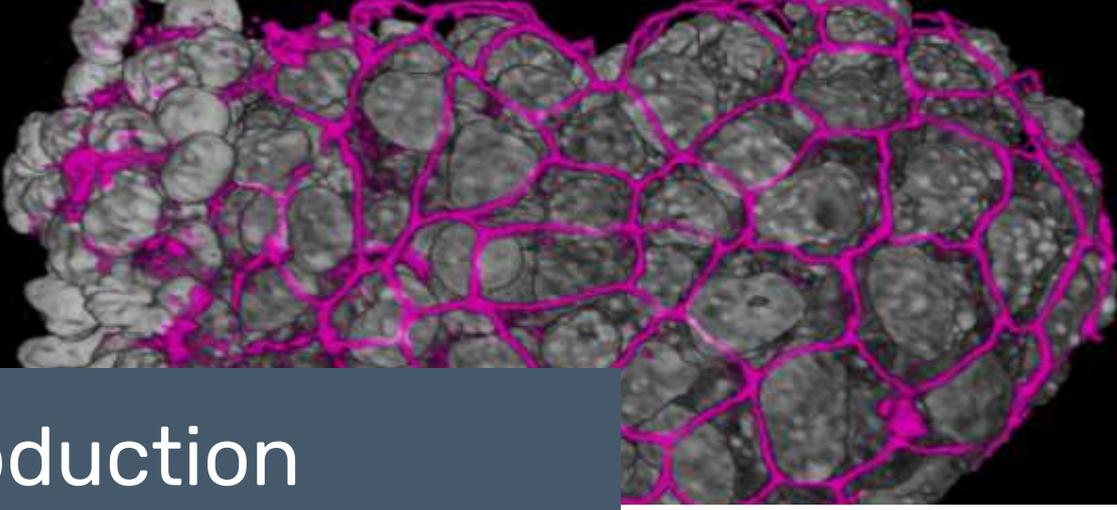


## eBook: Single-cell transcriptomics

# CONTENTS

- **Introduction**
- **News:** Single-cell sequencing provides new wisdom about our teeth
- **Tech News:** Frozen tissues and Tabula Sapiens: the latest discoveries using single-cell transcriptomics from the Human Cell Atlas
- **Infographic:** Single-cell analyses: top tips for sample preparation and informatics
- **Benchmark:** Comparison of optimized methodologies for isolating nuclei from esophageal tissue
- **Benchmark:** Single-cell RNA sequencing of intramedullary canal tissue to improve methods for studying fracture repair biology
- **Report:** Assigning immunoglobulin class from single-cell transcriptomes in IgA1-secreting versus membrane subpopulations
- **Webinar:** Using single-cell approaches to investigate early embryonic development





# Introduction

The advent of single-cell sequencing has revolutionized the transcriptomics space, allowing researchers to delve deeper into cell properties like gene expression than previous methods, which simply provided ensemble average measurements of all the cells in a sample. The technique has had such an impact that it has been recognized as Nature's Method of the Year twice in one decade; it was first awarded for its initial development in 2013 and then for the addition of multimodal omic capabilities in 2019.

Despite the meteoric rise of this technique, significant challenges remain: vast amounts of data are produced by these studies, which can be challenging to analyze and interpret. The collection of single-cell data; therefore, has to go hand-in-hand with the tailored use of machine learning methods that drive the generation of hypotheses and the design of validation experiments. Sample preparation for these single-cell studies can also be exacting and require sophisticated methods to isolate and extract RNA for analysis.

This eBook will provide an insight into the breadth of topics impacted by the development of single-cell transcriptomics, and provide critical tips for sample preparation for the generation of single-cell transcriptomic data and its subsequent analysis. Discover techniques for isolating individual cells suitable for single-cell RNA sequencing analyses from intramedullary canal tissue and learn about the analysis of single-cell transcriptome data for the identification of immunoglobulin classes.



Tristan Free  
Senior Editor,  
*BioTechniques*  
tfree@biotechniques.com



# Single-cell sequencing provides new wisdom about our teeth

*A new single-cell sequencing atlas of human teeth has provided a comprehensive understanding of their make up, and could take us one step further toward a reality of regenerative dental treatments.*

Teeth are a unique tissue, unique to each person. Underneath your enamel lies dentin and the dental pulp, a vascularized and innervated tissue lined by odontoblasts. The tooth is then anchored to the bone via the periodontium, which is crucial for tooth stability.

Both the pulp and the periodontium are subject to pathologies such as caries and periodontitis, which can often result in the need for dental treatments.

However, both of these areas have their own regenerative potential, owing to the presence of mesenchymal stem cells. These two stem cell populations are aptly named dental pulp stem cells and periodontal stem cells. Both types are multipotent and involved in tissue regeneration and thus have garnered interest for their potential use in dentistry. However, *in vivo* studies looking to harness their regenerative potential have not borne fruit.

The new single-cell sequencing atlas produced by Thimios Mitsiadis and his team at the University of Zurich (Switzerland), predominantly using single-cell RNA sequencing and analysis technologies, covers both these tissues and stem cell populations, providing the highest resolution data for human teeth to date.

Commenting on the reasons behind the study, Mitsiadis explained: "Single-cell approaches can help us understand the interactions of dental pulp and periodontal cells involved in immune responses upon bacterial insults. Therefore, single-cell analysis could be useful for diagnostic purposes to support the early detection of dental diseases."

Surprisingly, while the team identified great cellular heterogeneity in both tissues, the molecular signatures of the stem cell populations were similar. The team theorized this is due to their distinct microenvironments. If this theory stands true after further investigation, these findings could open new avenues for cell-based dental therapeutics.

#### Reference:

1. Pagella P, de Vargas Roditi L, Stadlinger B, Moor AE, Mitsiadis TA. A single-cell atlas of human teeth. *iScience* 24(5), 102405 (2021).



# Frozen tissues and Tabula Sapiens: the latest discoveries using single-cell transcriptomics from the Human Cell Atlas

The latest studies published by The Human Cell Atlas make further progress toward their goal of mapping every human cell type, and these four papers focus on multi-tissue cell analysis.

The Human Cell Atlas (HCA) was founded in 2016 and is an international consortium consisting of 2,300 members from 83 countries working towards charting every cell type in healthy human bodies. The HCA has created detailed maps of more than one million cells collected from 33 organs and systems and has mainly focused on individual organs and tissues, or smaller subsets of tissues. Now, they have developed methods to collect data needed for multi-tissue cell atlases. The resulting cell atlases are openly available meaning researchers can compare specific cell types and their functions across the body.

## Studying immune cells across tissues

Until now, the HCA focussed on immune cells that are transported in the blood; however, immune cells in tissues also play an important role in the immune system. Researchers from the HCA have created a catalog of immune cells after sequencing RNA from 330,000 single immune cells to understand their function in different tissues. [1] From this catalog, they developed a machine learning tool called CellTypist to automate cell identification. Using this tool, they identified around 100 different immune-cell types and their distribution across tissues, for example, T cells, B cells, and macrophages.

“By comparing particular immune cells in multiple tissues from the same donors we identified different flavors of memory T cells in different areas of the

body, which could have great implications for managing infections,” says Sarah Teichmann, who is the Head of Cellular Genetics at the Wellcome Sanger Institute (Cambridge, UK) and a co-author on the paper. “Our openly available data will contribute to the HCA and could serve as a framework for designing vaccines, or to improve the design of immune therapies to attack cancers.”

The second study published looks at the tissues involved in the formation of blood and immune cells and reveals the cell types lost from childhood to adulthood. This could inform *in vitro* cell engineering and research into regenerative medicine. [2]

## Freezing tissues for analysis

A single-cell atlas would be beneficial to identify and map out the specific cell types in which disease genes act. To create this, all the cell types need to be profiled, including those that are difficult to collect, for example, fat cells, or cells from skeletal muscle or neurons. Additionally, it is essential to profile cells from many different individuals, so freezing tissue before analysis is required.

Researchers from the HCA have developed a single-nucleus RNA sequencing method using frozen cells. [3] They then used this method to create a cross-tissue atlas and analyze 200,000 cells from a bank of frozen tissues with rare and common disease genes. A novel machine-learning algorithm was used to associate cells in the atlas with 6,000 single-gene diseases and 2,000 complex genetic diseases and traits to identify cell types and gene programs in disease. This could lead to novel starting points for



# Frozen tissues and Tabula Sapiens: the latest discoveries using single-cell transcriptomics from the Human Cell Atlas

health and disease studies in the future.

Aviv Regev (Genentech Research and Early Development; CA, USA), senior author of the paper explains: “Our single-nucleus HCA study demonstrates a powerful large-scale way to analyze cells from frozen tissue samples across the body with deep-learning computational advances and opens the way to studies of tissues from entire patient cohorts at the single-cell level. We were able to create a new roadmap for multiple diseases by directly relating cells to human disease biology and disease-risk genes across tissues.”

## The Tabula Sapiens dataset

The fourth and final paper being published in *Science* from this collection produced a cross-tissue atlas from live cells. [4] The resulting dataset is called ‘Tabula Sapiens’. This was done using single-cell RNA sequencing of live cells to analyze several organs from the same donors. The Tabula Sapiens has been used to characterize more than 400 specific cell types, distribution and variations in gene expression. This will provide researchers with a large resource of annotated cell types and the Tabula Sapiens enabled the first large-scale analysis of alternative gene splicing in a single-cell atlas.

“The Tabula Sapiens is a reference atlas that provides a molecular definition of hundreds of cell types across 24 organs in the human body,” said Stephen Quake, a senior author of this paper and a Professor at Stanford University (CA, USA). “It represents the efforts of more than 150 authors

across several institutions; the scientific community will be discovering new insights into human biology from this resource for many years to come.”

Together, these four studies contribute to the single Human Cell Atlas being created by the consortium and could have therapeutic implications like understanding common and rare diseases, vaccine development, and anti-tumor immunology.

## References:

1. Conde CD, Xu C, Jarvis LB *et al*. Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture in humans. *Science* 376(6594) eabl5197 (2022).
2. Suo C, Emma Dann, Goh I *et al*. Mapping the developing human immune system across organs. *Science* doi: 10.1126/10.1126/science.abo0510 (2022)(Epub ahead of print).
3. Eraslan G, Drokhlyansky E, Anand S *et al*. Single-nucleus cross-tissue molecular reference maps towards understanding disease gene function. *Science*. 376(6594) eabl4290 (2022).
4. The Tabula Sapiens Consortium. The Tabula Sapiens: a multiple organ single-cell transcriptomic atlas of humans. *Science* 376(6594) eabl4896 (2022).
5. Human Cell Atlas. Multi-tissue cell atlases lead to leap of understanding of immunity and disease. <https://www.humancellatlas.org/multi-tissue-cell-atlases-lead-to-leap-of-understanding-of-immunity-and-disease/> [Accessed 04 July 2022]



# Single-cell analyses

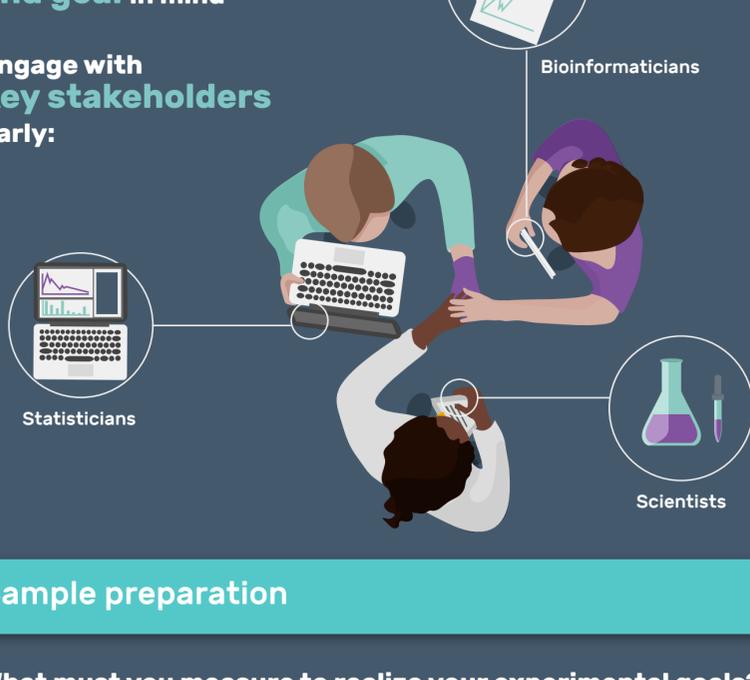
top tips for sample preparation and informatics

In association with 10x Genomics

## Robust single-cell analysis begins with thorough study design

Begin your design with the end goal in mind

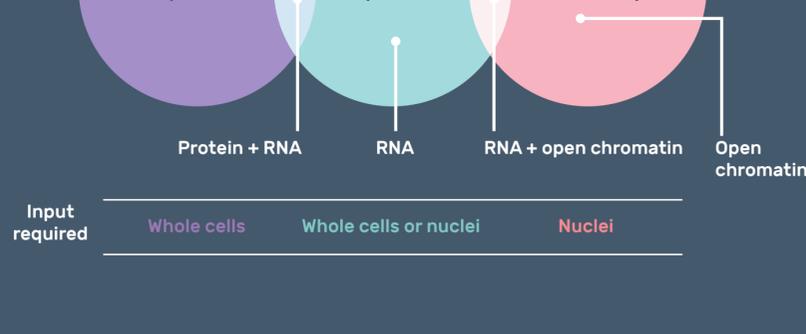
Engage with key stakeholders early:



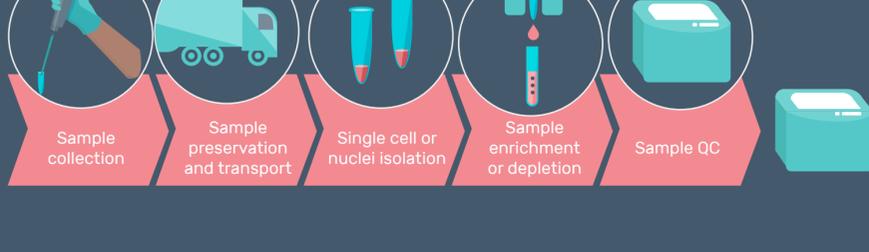
## Sample preparation

### What must you measure to realize your experimental goals?

What are you measuring?



### Key steps for sample preparation



### Sample type: whole cells or nuclei?

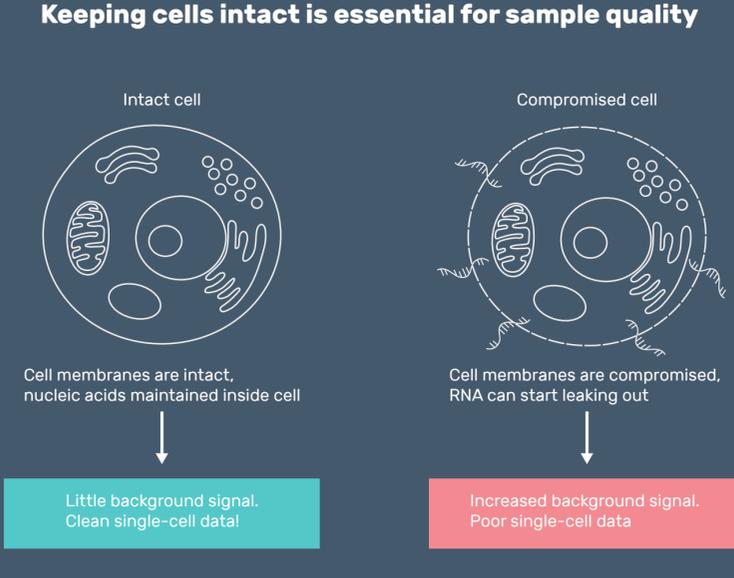
	Cells	Nuclei
<b>Assay compatibility</b>	• All gene-expression and immune-profiling solutions	• All gene-expression and epigenomics assays
<b>Sample type</b>	• Requires fresh tissue that can be easily dissociated	• Good for flash-frozen tissue or hard to dissociate tissue • Required input for epigenomics assays
<b>Analytes obtained</b>	• Spliced mRNA and lots of it • Cell surface proteins	• Unspliced RNAs with lots of introns • Chromatin (epigenomics assays) • No cell surface proteins, no cell membrane
<b>Storage</b>	• Cell suspensions can be cryopreserved • Fresh tissue must be dissociated	• Frozen tissue can be transported and stored
<b>Protocols available</b>	• Lots of dissociation protocols, some optimization may be required	• Protocols available, optimization required

### Sample quality is critical

To secure a good quality sample, with cell viability over 70%:



### Keeping cells intact is essential for sample quality



### Selecting sample dissociation methods

Enzymatic dissociation



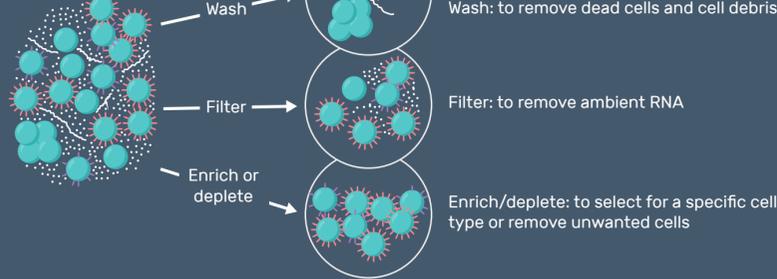
- Preserves cell integrity
- Some enzymes impact cell surface proteins

Mechanical dissociation



- Required for frozen samples
- More likely to damage cells

### Improving poor quality samples to reach 70% cell viability



## Informatics

### Consider how your data will be analyzed

Department/institute bioinformatician



- Contact early for tips on how to provide them with the best data
- Must clearly communicate needs and goals
- Typically requires more back and forth communication

External company/data analysis service provider



- Ensure your data is produced and formatted to the provider specifications
- Must clearly communicate needs and goals
- Typically requires more back and forth communication

Yourself



- Save budget, but requires expertise, IT infrastructure and time
- Opportunity to develop new skills in the lab
- If expertise available, can be faster and require less communications

### Analyzing yourself?

Select your tools wisely

Command line interface tools



- Flexible
- Requires command coding knowledge
- Can be the only option in some situations

Graphical user interface tools



- More user friendly
- More constrained
- Can be faster and more convenient for the initial stages of analysis

Commercial



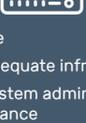
- More reliable
- Support available to fix occasional bugs, answer questions and resolve issues
- Require budgets
- Occasionally free

Open source



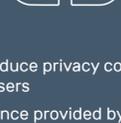
- More prone to bugs
- Many options
- Free
- Support is inconsistent between platforms

Install locally



- More secure
- Requires adequate infrastructure
- Requires system administrators for maintenance

Store in the cloud



- Scalable
- Can introduce privacy concerns for certain users
- Maintenance provided by the cloud

# Comparison of optimized methodologies for isolating nuclei from esophageal tissue

Lucy M Kimbley<sup>1</sup> , Rachel Parker<sup>1</sup>, Jack Harrington<sup>1</sup> , Robert C Walker<sup>1</sup> , Ben Grace<sup>1</sup>, Jonathan J West<sup>1,2,3</sup> , Tim J Underwood<sup>1,3</sup>  & Matthew JJ Rose-Zerilli<sup>\*,1,3</sup> 

<sup>1</sup>Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK; <sup>2</sup>Centre for Hybrid Biodevices, University of Southampton, Southampton, UK;

<sup>3</sup>Institute for Life Sciences, University of Southampton, Southampton, UK; \*Author for correspondence: [mjrz@soton.ac.uk](mailto:mjrz@soton.ac.uk)

BioTechniques 72: 105–109 (March 2022) 10.2144/btn-2021-0036

First draft submitted: 16 April 2021; Accepted for publication: 17 January 2022; Published online: 14 February 2022

## ABSTRACT

Single-nuclei RNA sequencing allows single cell-based analysis in frozen tissue, ameliorating cell recovery biases associated with enzymatic dissociation methods. The authors present two optimized methods for isolating and sequencing nuclei from esophageal tissue using a commercial EZ and citric acid (CA)-based method. Despite high endogenous RNase activity, these protocols produced libraries of expected fragment length (average length EZ: 745 bp; CA: 1232 bp) with comparable complexity (median Transcript/Gene number, EZ: 496/254; CA: 483/256). CA nuclei showed a higher proportion of ribosomal gene reads, potentially reflecting co-isolation of nuclei and adherent ribosomes. The authors identified 11 cell lineages in the combined datasets, with differences in cell type recovery between the two methods, providing utility dependent on experimental needs.

## METHOD SUMMARY

The authors present a method for isolating intact nuclei from frozen esophageal tissue for high throughput single-nuclei RNA sequencing using microfluidic-based droplet partitioning. This protocol overcomes high levels of endogenous RNase activity in this barrier epithelium and would therefore be applicable to other tissue types affected by similar issues.

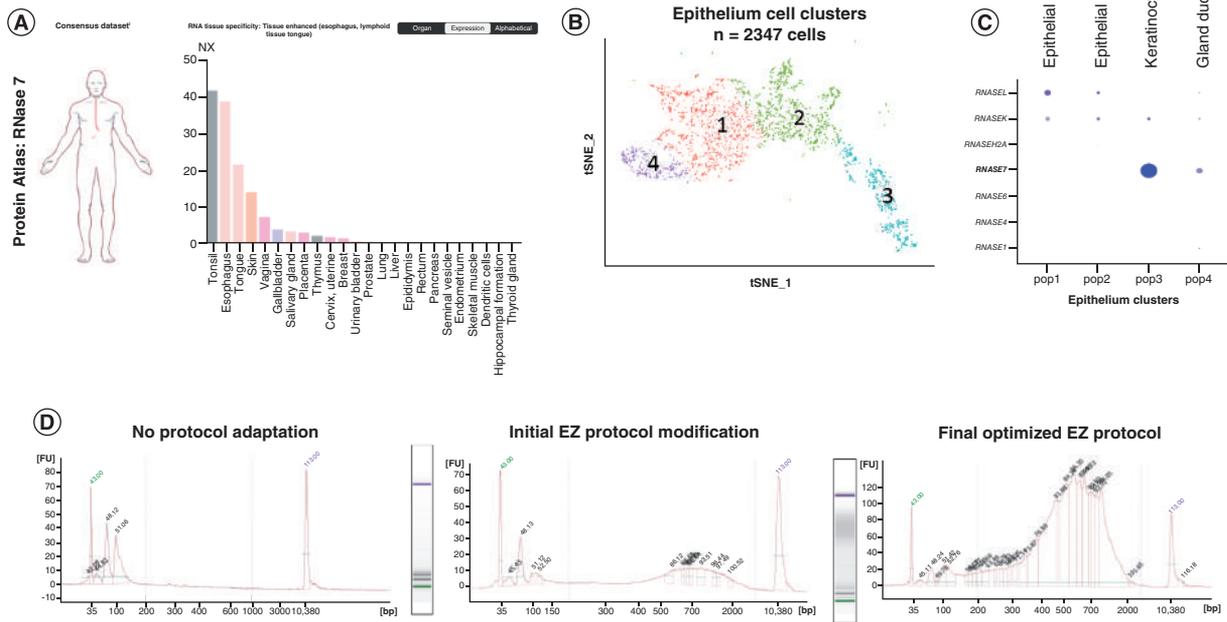
## KEYWORDS:

esophagus • nuclei • RNA • sequencing • single cell • tissue sampling • transcriptome

Single-nuclei RNA sequencing (snRNA-seq) allows the identification of cell types in heterogeneous tissues that have been frozen before downstream processing and analysis. Commercial buffers exist to isolate nuclei, as well as traditional salt and detergent-based methods [1]. These buffers must be gentle enough to lyse cell membranes while preserving nuclear membrane integrity. Nuclei can then be purified from cellular debris and contaminating cell-free RNA through multiple centrifugation steps, fluorescence-activated cell sorting (FACS) or density gradients (e.g., sucrose or iodixanol) before proceeding with microfluidic encapsulation and RNA barcoding [1,2].

As a barrier epithelial tissue, the esophagus contains RNases, which may play a role in mediating antiviral defense [3]. From the authors' single-cell RNA sequencing datasets, they have identified the cellular source of the expression of these RNases (Figure 1A–C). The authors identified high levels of endogenous RNases in outer keratinocyte populations, which can contribute to the loss of RNA integrity. This is in keeping with the role for RNase 7 in antiviral activity in keratinized cells [4,5]. Upon disruption of cell membranes during freezing and subsequent nuclei extraction, there is significant and rapid degradation of nuclear RNA (Figure 1D), which is likely to be associated with the release of reactivated RNases [6]. The authors have optimized two nuclei isolation methods to protect nuclear RNA integrity in frozen esophageal tissue (Figure 2) and demonstrate that both methods can recover a high diversity of cell populations following snRNA-seq.

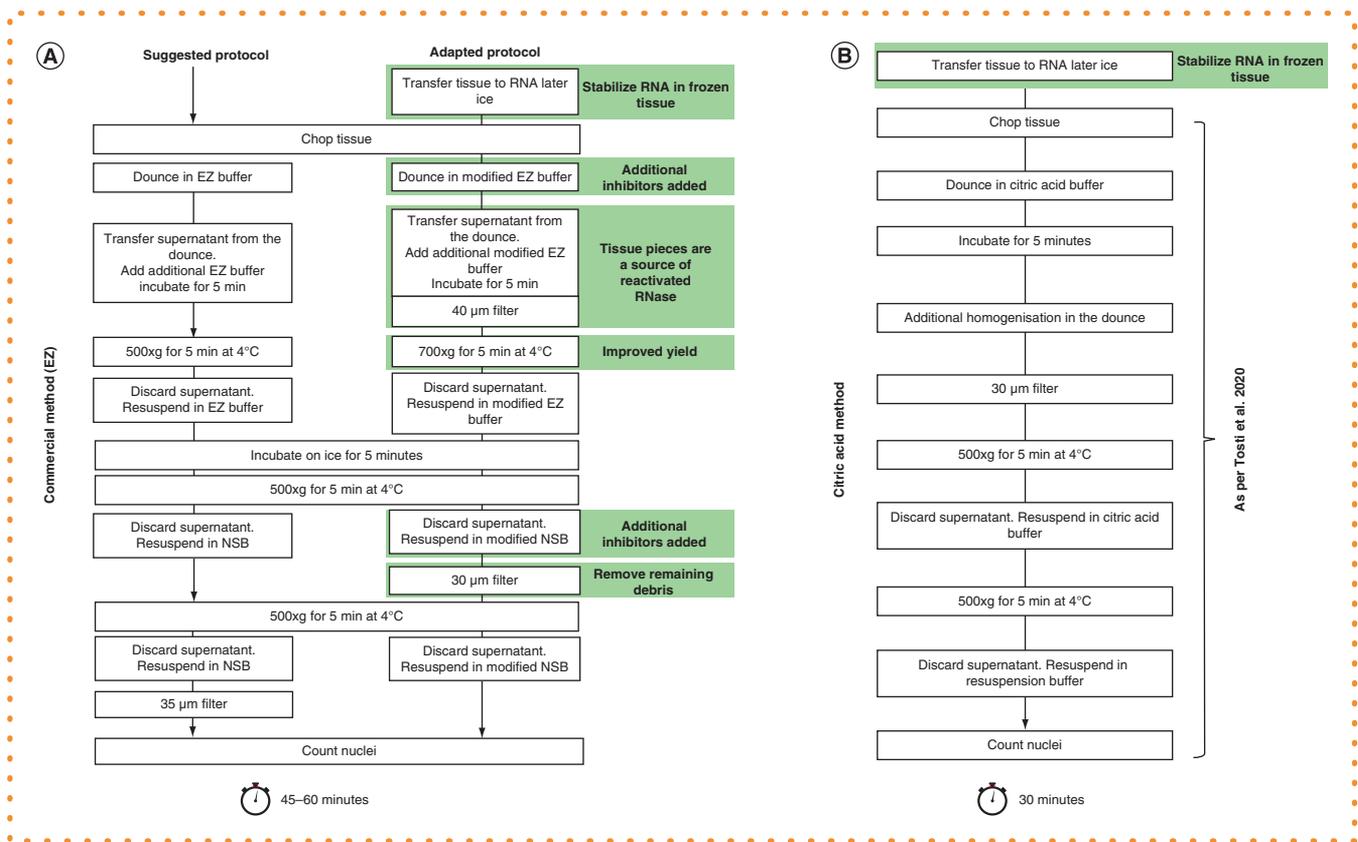
To assess these methods, the authors obtained sequencing data from nuclei from a single patient, using one sample of snap-frozen normal human esophagus obtained by surgical resection (see ethical disclosure statement). Using the Nuclei EZ prep nuclei isolation kit (Sigma), according to published methods used for isolating nuclei for snRNA-seq by Dronc-seq [7], the authors found that nuclei from frozen esophagus lacked sufficient quality to generate cDNA libraries (Figure 1D). To overcome the effects of endogenous enzymes, they initially added an RNase inhibitor to the nuclei isolation buffer (0.2% v/v Lucigen NxGen RNase inhibitor) and transferred the frozen tissue to an RNA rescue medium (RNALater-ICE) without defrosting and then maintained it in the solution at -20°C until nuclei extraction. This enabled stabilization of the frozen tissue, providing protection during immediate removal from cold storage, allowing the inhibitors in the lysis buffer to take effect. This method successfully yielded cDNA libraries, and with additional optimization (0.2% v/v Lucigen NxGen RNase inhibitor, 0.2% v/v SUPERase-in, 1x Roche Complete protease inhibitor cocktail, 4 mM DTT) (Figure 1D & Figure 2A) the final optimized EZ protocol resulted in cDNA libraries of sufficient quality.



**Figure 1. Identification of RNA inhibitors in esophageal tissue.** (A) RNase 7 expression (NX: normalized expression) across tissue types (consensus dataset from the Protein Atlas; <https://www.proteinatlas.org/>). Image credit: Human Protein Atlas. (B) t-Distributed Stochastic Neighbor Embedding of epithelial cell populations from esophageal tissue analyzed by single-cell RNA sequencing. (C) RNase gene expression in epithelial cell populations from esophageal tissue analyzed by single-cell RNA sequencing. The size and color of the dot denote the percentage of cells and the average expression level, respectively. (D) Optimized nuclei isolation protocol for frozen esophageal tissue to mitigate reactivated endogenous RNase activity. cDNA library electrophoresis traces (high sensitivity DNA chip [Agilent Bioanalyzer] cDNA concentration quantified in fluorescence units) for pooled nuclei from a sample before, during and after successful protocol optimization.

Using a portion of the same sample stored in RNALater-ICE, the authors evaluated a citric acid (CA)-based isolation method, as described in Tosti *et al.* [8,9], where nuclei were isolated from pancreatic specimens, a tissue type with high RNase activity similar to esophageal tissue [10]. CA has historically been used for nuclei isolation and can improve the separation of nuclei from cytoplasmic contaminants [11] and yield greater quality RNA owing to its acidic pH and activity as an RNase chelator [12,13]. The tissue was removed from RNALater-ICE and homogenized in 25 mM citric acid and 0.25 M sucrose, following the protocol outlined in Figure 2B. This CA nuclei isolation was performed separately (+3 months) from the EZ isolation protocol. However, tissue stabilization using RNALater-ICE is a critical step in stabilizing nuclear RNA, and this step was not varied between methods. Again, the authors found that RNALater-ICE use yielded larger cDNA libraries with longer fragment lengths, suggesting a higher quality of recovered RNA (Figure 3A & B). Following sequencing, the resulting libraries from both isolation methods showed similar complexity (EZ median nTranscript/nGene: 496/254; CA: 483/256) (Figure 3C). The proportion of mitochondrial genes was higher than expected (EZ median: 25.1%; CA: 12.7%), given that these should be pure nuclei fractions. Previously published nuclei datasets have observed elevated mitochondrial gene expression, hypothesizing that mitochondria may associate with nuclear membranes [14]. The EZ method displayed approximately the anticipated proportion of intronic reads (median: 40.9%), given that nuclei contain pre-spliced transcripts. For the CA method, the proportion of intronic reads was lower than the EZ method (median: 28.6%), suggesting the presence of mature RNA. It has previously been found that CA-based methods may result in nuclei co-isolated with ribosomes or outer nuclear membrane fragments (Figure 3D) [15]. The CA nuclei showed a significantly higher percentage of reads originating from ribosomal genes (median: 7.6%) than the EZ method (median: 1.7%;  $p < 0.001$  Mann-Whitney U test), potentially reflecting co-isolation with adherent ribosomes. This is further supported by stratification of the CA nuclei by percentage of ribosomal reads (Figure 3E & F), identifying two subpopulations of nuclei: those with a ribosomal gene proportion  $>8\%$  showing a lower median intronic proportion (14%) and those with  $<8\%$  ribosomal genes having a higher intronic proportion (37%). These two populations could reflect nuclei isolated with adherent ribosomes and pure nuclei, respectively.

For analysis, CA and EZ nuclei datasets from the same patient sample were combined (pre-combined analysis in Supplemental Figure 1). Following clustering with the Seurat package (v3.2.2) [16] using the first 15 principal components of variable gene expression, cell

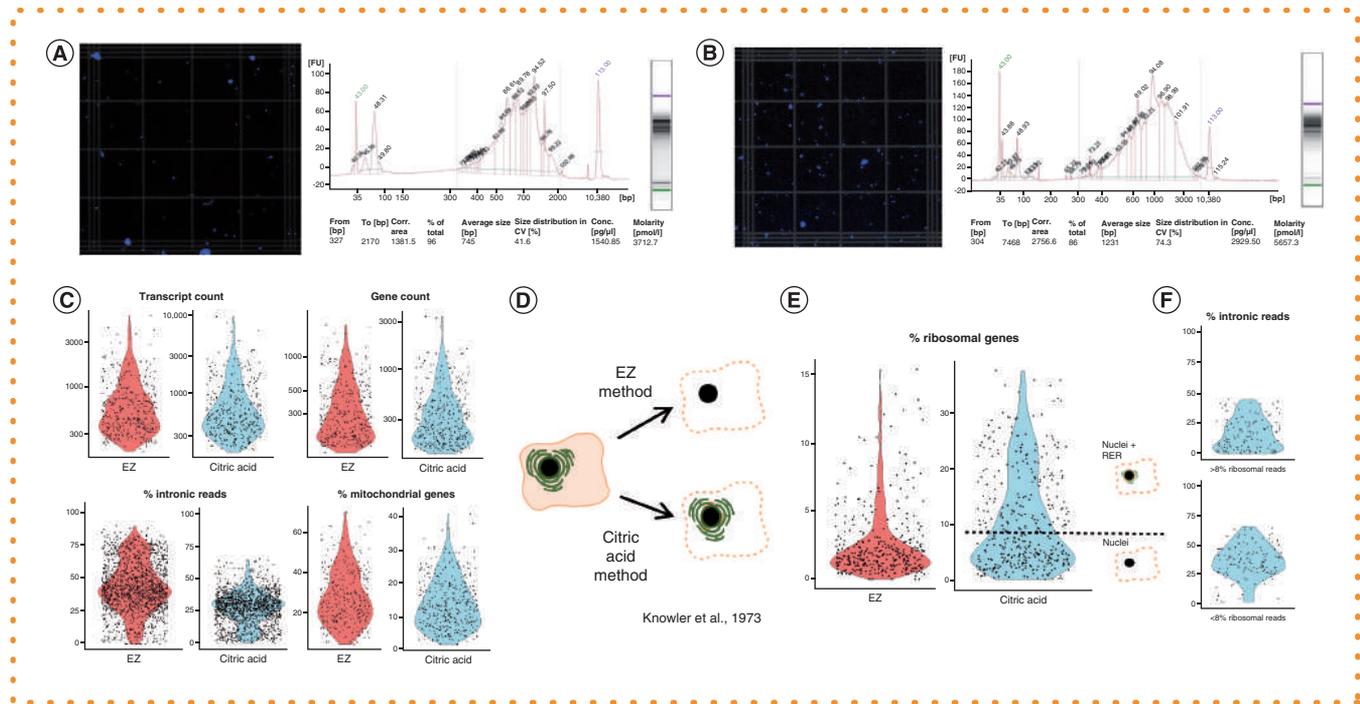


**Figure 2. Optimization of nuclei isolation protocols. (A)** Flowchart of adaptations to commercial protocol using modified EZ lysis buffer (EZ nuclei lysis buffer, 0.2% v/v Lucigen NxGen RNase inhibitor, 0.2% v/v SUPERase-in, 1x Complete protease inhibitor cocktail, 4 mM DTT) and modified nuclei suspension buffer (1xPBS, 0.01% BSA, 0.2% v/v RNase inhibitor, 1 mM DTT). **(B)** Citric acid protocol workflow. Citric acid buffer (0.25M sucrose, 25 mM citric acid), resuspension buffer (25 mM KCl, 3 mM MgCl<sub>2</sub>, 50 mM Tris-buffer, 0.4 U/µl NxGen RNase inhibitor, 1 mM DTT, 0.4 U/µl SUPERasin).

types were identified using lineage marker genes from previous large-scale scRNA-seq esophageal datasets (Figure 4) [17]. Samples were sequenced separately, and batch effects were controlled for by regression of nCount and mitochondrial percentage. The authors were able to identify 11 cell clusters and one cluster of unassignable low-informative nuclei captures (Figure 4). Two clusters were uniquely observed in the CA dataset and marked by high ribosomal gene expression (stratified keratinocyte ribo-high and basal epithelial ribo-high clusters). The stratified keratinocyte and ribo-high populations clustered closely, with a highly correlated average gene expression profile between the populations (Pearson's correlation: 0.89). While there was co-clustering of the EZ- and CA-derived stratified keratinocytes, the CA-derived basal epithelial cells did not cluster with the primary basal epithelial cluster because of high expression of ribosomal genes (Figure 4C).

More keratinocytes were recovered by the CA method, particularly stratified keratinocytes (~14-fold increase in the percentage recovered) (Figure 4B), distinguished by their expression of *KRT4* (Figure 4C & D). Interestingly, the single-cell dataset shows that *RNase7* expression was highest in outer keratinocytes, suggesting that the CA method may result in improved keratinocyte membrane lysis and/or RNA integrity from nuclei compared with the EZ method. In contrast, the EZ method yielded a greater recovery of fibroblasts (sixfold) and immune cells (2.5-fold for combined immune populations), with populations expressing marker genes of natural killer (*PTPRC*, *KLRD1*) and macrophage (*CD163*) lineages (Figure 4C & D). Both methods identified two clusters of vascular smooth muscle cells (VSMCs) marked by both *ACTA2* and *MYH11* and discriminated by *A2M* expression, potentially suggesting an activated subset involved in an inflammatory VSMC response (Figure 4C & D) [18]. Nuclei doublet estimates by DoubletFinder [19] were generally less than 4.8% across most cell populations, bar the outer keratinocyte cluster, where doublet estimates were 77.8% (Figure 4E). This finding was corroborated by a lower genes/transcript ratio, clearly identifying them as nuclei doublet captures, and the absence of mixed lineage marker expression suggests these are homotypic doublets. Alternatively, this could be due to keratinocytes being more resistant to lysis, or greater levels of adherent cytoplasm. Both methods provided representation of stromal, immune and vascular cell types known to make up the architecture of the normal esophagus.

In the datasets, the authors have corroborated identification of the cell types they recovered using markers established in large-scale scRNA-seq studies of the esophagus. In their hands, the EZ method recovered a greater number of immune and fibroblast cells, while

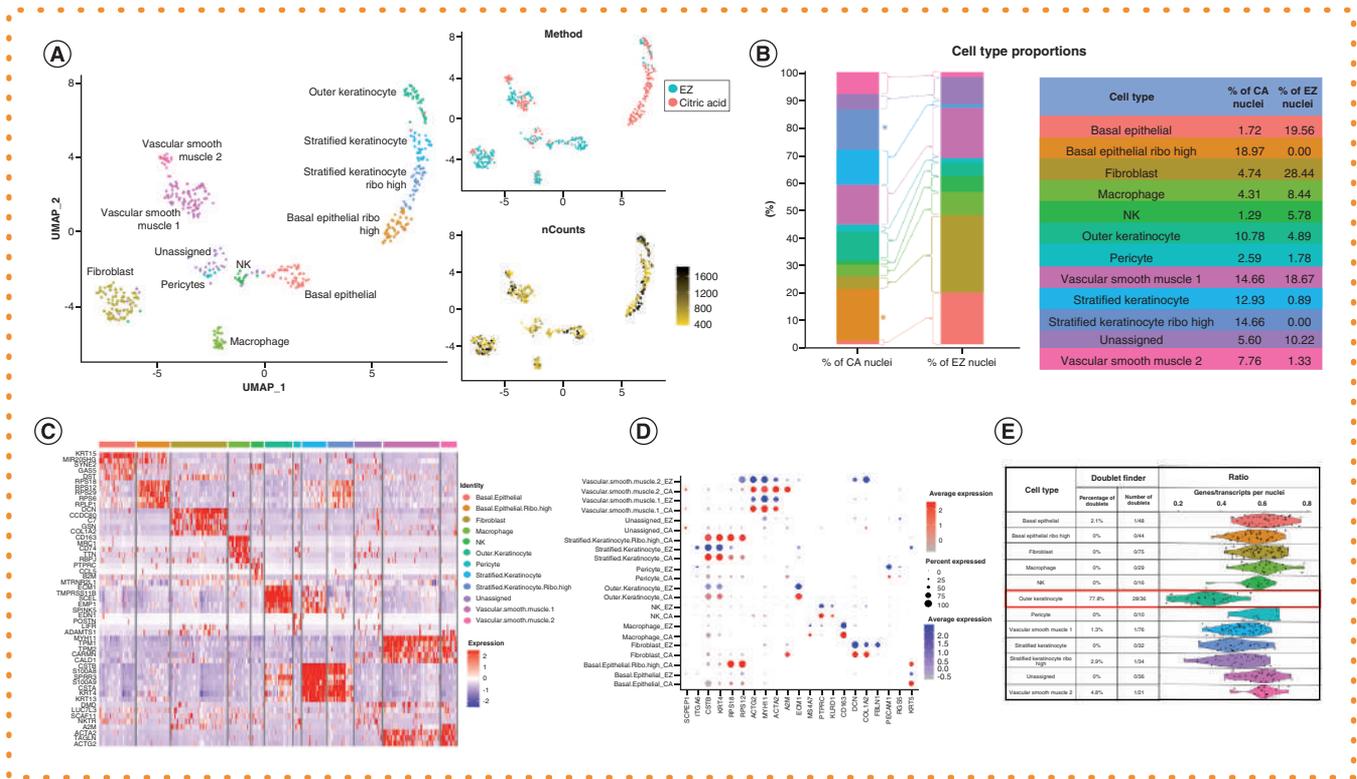


**Figure 3. Comparison of EZ and citric acid (CA)-isolated nuclei preparations and library quality. (A & B)** Representative images (4x magnification) of DAPI-stained nuclei and cDNA electrophoresis fragment traces for recovered nuclei pools for the (A) modified EZ method and (B) CA method, respectively. In both nuclei images, a 0.2  $\mu$ l volume region is shown in both, but with different nuclei loading concentrations. (C) Percentage reads originating from ribosomal genes for single nuclei isolated by the EZ (median: 1.7%) or CA method (median: 7.6%). Dotted line at 8% representing a potential divide between nuclei isolated with adherent ribosomes and nuclei isolated alone. (D) Intronic read percentage for populations with  $>$  or  $<$  8% ribosomal reads. (E) Diagram demonstrating potential retention of endoplasmic reticulum and ribosomes when nuclei are isolated by the CA method compared with the EZ method. (F) Transcript count and feature count for single nuclei (points) isolated by the EZ (red violin) or CA method (blue violin). Log scale. EZ (median transcript count: 496; median gene count: 254). CA (median transcript count: 483; median gene count: 256). Percentage of intronic reads (median EZ: 40.9%; median CA: 28.6%) and mitochondrial genes (median EZ: 25.1%; median CA: 12.7%) for single nuclei isolated by the EZ or CA method.

the CA method also provided representation of keratinocyte populations, which will be useful for researchers studying these specific cell types. Further validation studies using additional patient samples to determine the extent of interpatient variability in cell type recoveries are required, but the results presented here allow direct comparison of cell type recoveries between methods using one patient sample.

The authors' observation of high mitochondrial genes expression has been observed previously in snRNA-seq datasets [14], supporting the assertion that mitochondria, and organelles including ribosomes, may become associated with nuclear membranes. However, cell viability and sample handling, preservation and differences in storage conditions or time may also be contributing factors, which the authors were unable to address in full. The localization of organelles may have additional benefits for the CA method. First, nuclei and attached ribosome sequencing has been shown to facilitate the study of rare cell types that are challenging to analyze by nuclei sequencing methods alone [20]. Second, the CA method may have utility for studies of RNA transcripts actively undergoing translation.

To facilitate analysis of primary tissues such as the esophagus, it is essential that dissociation methods are validated, so that cellular heterogeneity captured across studies can be corroborated. Validated methods that can be used with frozen tissue have advantages for the retrospective study of archived samples and for facilitating work in laboratories without access to fresh samples. The authors' protocols have immediate utility in studies of esophageal cancer, a disease with significant unmet clinical need that was previously challenging to study at the single-cell level. The authors have provided detailed laboratory protocols here: [www.protocols.io](http://www.protocols.io) (dx.doi.org/10.17504/protocols.io.btm6nk9e; dx.doi.org/10.17504/protocols.io.t9wer7e).



**Figure 4. Cell type identification of nuclei.** (A) Uniform Manifold Approximation and Projection (UMAP) clustering of nuclei isolated by both citric acid (CA) and EZ methods with cell types identified. UMAPs showing method of origin and transcript count for each nucleus. (B) Proportions of each cell type isolated by each method. \*A unique population not represented in both methods. (C) Heatmap of top five genes expressed by each cell type. Cell type indicated by bar. (D) Dot plot showing expression of cell type marker genes in nuclei clusters, where color shows average expression and size of dot indicates percentage of cells expressing the marker gene. (E) Ratio of genes to transcript counts by cell type. Proportions of doublets in each cell type cluster are listed as determined by DoubletFinder assuming a doublet rate of 7%.

### Author contributions

Conceptualization: MJJ Rose-Zerilli, TJ Underwood; methodology: R Parker, J Harrington, RC Walker, MJJ Rose-Zerilli, LM Kimbley; formal analysis: MJJ Rose-Zerilli, LM Kimbley; investigation: R Parker, J Harrington, RC Walker, B Grace, LM Kimbley; resources: JJ West; data curation: LM Kimbley; writing: original draft, LM Kimbley; writing: review and editing, MJJ Rose-Zerilli, TJ Underwood; visualization: LM Kimbley; supervision, MJJ Rose-Zerilli; project administration: MJJ Rose-Zerilli, TJ Underwood; funding acquisition: MJJ Rose-Zerilli, TJ Underwood.

### Acknowledgments

The authors thank University Hospital Southampton patients for their participation in this study and the Southampton CR UK Tissue Bank. The authors would also like to thank the Wessex Investigation Sciences Hub Laboratory for access to their facilities and their technical support.

### Financial & competing interests disclosure

TJ Underwood is supported by a Royal College of Surgeons of England and Cancer Research UK Advanced Clinician Scientist Fellowship (A23924). RC Walker is supported by a Cancer Research UK Clinical Research Training Fellowship (A25162). Development of Drop-seq technology in Southampton was funded by MRC grant (MC\_PC\_15078). This work was supported by a grant from the Chan Zuckerberg Initiative (Human Cell Atlas Pilot project: START: Standardization of Single-Cell and Single-Nucleus RNA-Seq Protocols for Tumors) to JJ West, TJ Underwood and MJJ Rose-Zerilli, and a grant from the Southampton CR UK Development Fund to JJ West, TJ Underwood and MJJ Rose-Zerilli. This investigation was supported by a collaboration with The General Hospital Corporation (Massachusetts General Hospital) under an award from The Chan Zuckerberg Initiative, an advised fund of Silicon Valley Community Foundation. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

## Ethical conduct of research

Human tissue samples were obtained from patients through University Hospitals Southampton NHS trust after informed consent. This study was approved by The Proportionate Review Sub-Committee of the North East – Newcastle & North Tyneside 1 Research Ethics Committee (REC no: 18/NE/0234).

## Open access

This article is distributed under the terms of the Creative Commons Attribution License 4.0 which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited. To view a copy of the license, visit <http://creativecommons.org/licenses/by/4.0/>

## References

1. Slyper M, Porter CBM, Ashenberg O *et al.* A single-cell and single-nucleus RNA-seq toolbox for fresh and frozen human tumors. *Nat. Med.* 26(5), 792–802 (2020).
2. Blobel G, Potter VR. Nuclei from rat liver: isolation method that combines purity with high yield. *Science* 154(3757), 1662–1665 (1966).
3. Ilinskaya ON, Mahmud RS. Ribonucleases as antiviral agents. *Mol. Biol.* 48(5), 615–623 (2014).
4. Harder J, Schröder J-M. RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin. *J. Biol. Chem.* 277(48), 46779–46784 (2002).
5. Rademacher F, Dreyer S, Kopfnagel V, Gläser R, Werfel T, Harder J. The antimicrobial and immunomodulatory function of RNase 7 in skin. *Front. Immunol.* 10, 2553 (2019).
6. Butler AE, Matveyenko AV, Kirakossian D, Park J, Gurlo T, Butler PC. Recovery of high-quality RNA from laser capture microdissected human and rodent pancreas. *J. Histotechnol.* 39(2), 59–65 (2016).
7. Habib N, Avraham-Davidi I, Basu A *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* 14(10), 955–958 (2017).
8. Tosti L, Conrad C. Nuclei isolation from snap frozen human pancreatic tissue using a citric acid buffer. *io.protocols* (2020).
9. Tosti L, Hang Y, Debnath O *et al.* Single-nucleus and *in situ* RNA-sequencing reveal cell topographies in the human pancreas. *Gastroenterology* 160(4), 1330–1344.e1311 (2021).
10. Farrell RE (Ed.). Chapter 7: Resilient ribonucleases. In: *RNA Methodologies (4th Edition)*. Academic Press, CA, USA, 155–172 (2010).
11. Higashi K, Narayanan KS, Adams HR, Busch H. Utilization of the citric acid procedure and zonal ultracentrifugation for mass isolation of nuclear RNA from Walker 256 carcinosarcoma. *Cancer Res.* 26(7 Part 1), 1582 (1966).
12. Adamala K, Szostak JW. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* 342(6162), 1098–1100 (2013).
13. Nowotny M, Yang W. Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J.* 25(9), 1924–1933 (2006).
14. Lake BB, Chen S, Hoshi M *et al.* A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat. Commun.* 10(1), 2832 (2019).
15. Knowler JT, Moses HL, Spelsberg TC. Comparison and characterization of nuclear isolation procedures as applied to chick oviduct. *J. Cell Biol.* 59(3), 685–695 (1973).
16. Stuart T, Butler A, Hoffman P *et al.* Comprehensive integration of single-cell data. *Cell* 177(7), 1888–1902.e1821 (2019).
17. Madissoon E, Wilbrey-Clark A, Miragaia RJ *et al.* scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol.* 21(1), 1 (2019).
18. Rodríguez-Calvo R, Ferrán B, Alonso J *et al.* NR4A receptors up-regulate the antiproteinase alpha-2 macroglobulin (A2M) and modulate MMP-2 and MMP-9 in vascular smooth muscle cells. *Thromb. Haemost.* 113(6), 1323–1334 (2015).
19. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 8(4), 329–337.e324 (2019).
20. Drokhllyansky E, Smillie CS, Van Wittenberghe N *et al.* The human and mouse enteric nervous system at single-cell resolution. *Cell* 182(6), 1606–1622.e1623 (2020).

# Single-cell RNA sequencing of intramedullary canal tissue to improve methods for studying fracture repair biology

James M Dominguez II<sup>1</sup> , Sharon M Moe<sup>1</sup>, Neal X Chen<sup>1</sup>, Todd O McKinley<sup>2</sup>, Krista M Brown<sup>2</sup>, Yunlong Liu<sup>3</sup>, Hongyu Gao<sup>3</sup> & Roman M Natoli<sup>\*,2</sup> 

<sup>1</sup>Department of Medicine, Division of Nephrology, Indiana University School of Medicine, Indianapolis, IN 46202, USA; <sup>2</sup>Department of Orthopaedic Surgery, Indiana University Health Methodist Hospital, Indianapolis, IN 46202, USA; <sup>3</sup>Department of Medical & Molecular Genetics, Center for Medical Genomics, Indiana University School of Medicine, Indianapolis, IN 46202, USA; \*Author for correspondence: [rnatoli@iuhealth.org](mailto:rnatoli@iuhealth.org)

BioTechniques 71: 431–438 (August 2021) 10.2144/btn-2021-0002

First draft submitted: 8 January 2021; Accepted for publication: 23 June 2021; Published online: 10 August 2021

## ABSTRACT

The ability to study the bone microenvironment of failed fracture healing may lead to biomarkers for fracture nonunion. Herein the authors describe a technique for isolating individual cells suitable for single-cell RNA sequencing analyses from intramedullary canal tissue collected by reaming during surgery. The purpose was to detail challenges and solutions inherent to the collection and processing of intramedullary canal tissue samples. The authors then examined single-cell RNA sequencing data from fresh and reanimated samples to demonstrate the feasibility of this approach for prospective studies.

## METHOD SUMMARY

Intramedullary canal tissue is challenging to study directly because of its inaccessibility and heterogeneous composition. In addition, single-cell RNA sequencing requires high sample purity and cell viability. The authors determined that the critical step required for producing usable samples for single-cell RNA sequencing from intramedullary canal tissue was collagenase digestion, followed by centrifugation with density gradient medium (Ficoll).

## KEYWORDS:

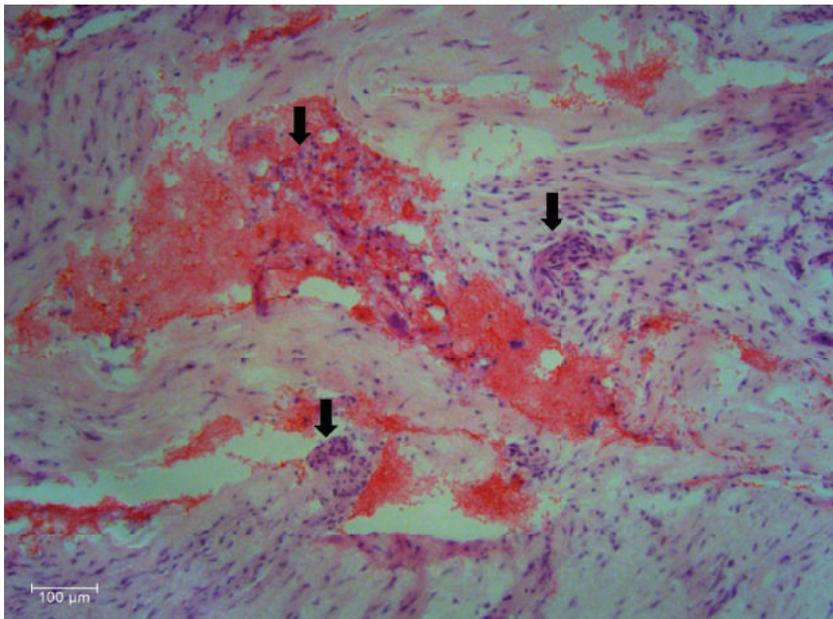
cell isolation • cryopreservation • fracture • nonunion • single-cell RNA sequencing

Nonunion of lower extremity fractures is a debilitating orthopedic condition. Patients experience prolonged pain, physical disability and decreased quality of life of the same magnitude as congestive heart failure [1]. Implicit to lower limb nonunions are long hospital stays and long-term unemployment, outcomes that are relevant to the public economy [2,3]. A practical definition of nonunion is when a fracture, in the opinion of the treating physician, has no possibility of healing without further intervention [4]. The overall risk estimate for nonunion in adults is 1.9% of all fractures, or 500,000 cases annually [5,6]. Clinical risk factors include injury characteristics and host factors, including age [7], fracture site (e.g., tibia/fibula [5.4%], femur/pelvis [1.3%]) [6], type of injury, open versus closed fracture [8], smoking status [9] and patient comorbidities [8]. However, the cellular biology of impaired fracture healing remains only marginally understood.

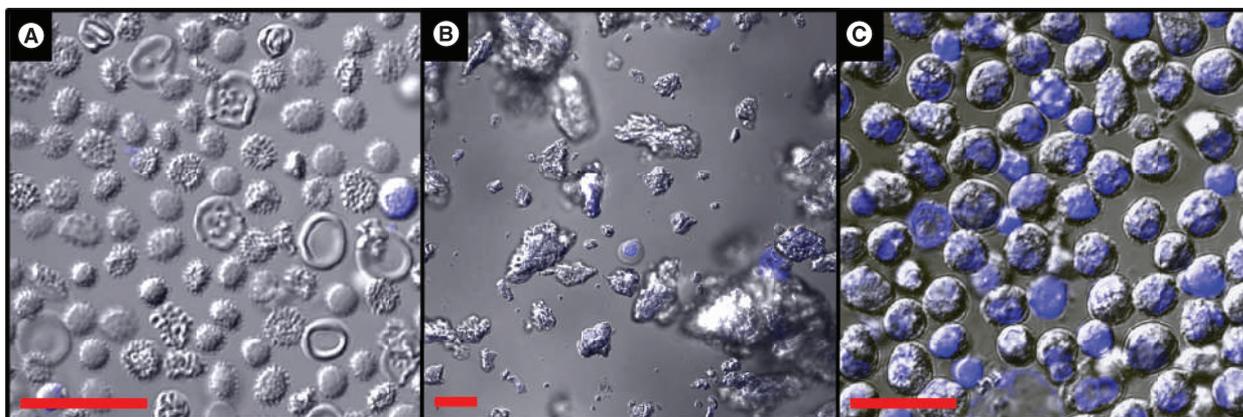
Once identified, nonunion is largely a surgical disease. Frequently, the procedure for nonunion repair involves accessing the intramedullary canal and reaming for implant removal, exchange nailing or obtaining autogenous bone graft. The primary goal of the present study was to determine the feasibility of collecting human intramedullary canal tissue (ICT) for single-cell RNA sequencing (scRNA-seq), looking toward improved understanding of cellular mechanisms in fracture repair or biomarkers predictive of failed healing. To demonstrate the applicability of this technique for downstream analyses, the authors first defined a suitable cell isolation protocol, determined cell viability and then performed scRNA-seq analyses. This is a time-intensive endeavor that involves coordination between multiple entities (surgeons, laboratory personnel and scRNA-seq core facility). Therefore, a secondary goal was to investigate whether freezing and subsequent reanimation of samples would allow for scRNA-seq. Taken together, these two goals have allowed the authors to collect, store and process ICT samples for future research. The authors have successfully completed seven additional sample acquisitions and analyses using the protocol described in this brief report.

## Human samples

Deidentified samples of ICT were obtained from three patients. The three patient samples consisted of the following: femur nonunion sample was ICT from a femoral nonunion undergoing exchange intramedullary nailing for treatment of the nonunion; femur control



**Figure 1. Histology of intramedullary canal tissue.** Intramedullary canal tissue (ICT) samples were fixed for 24 h in neutral buffered formalin (10%) and serially dehydrated in 5, 15 and 30% sucrose–phosphate-buffered saline for 24 h. Samples were embedded in optimal cutting temperature medium, frozen on dry ice and ultimately sectioned via cryostat at 10- $\mu$ m thickness. After sectioning, slides were stained with hematoxylin and eosin. Images were obtained using a Leica DM5000 B microscope, FLUOTAR objectives, DFC310 FX camera and LAS X software. Cells (blue) are surrounded by both soft tissue (pink) and hard tissue (red), indicating that the ICT obtained from reaming is a complex matrix. Black arrows indicate dense collections of cells.



**Figure 2. Live imaging of cell release by fluorescence microscopy.** To troubleshoot and validate the cell isolation procedure, the products were pelleted and then resuspended and incubated in Hoechst 33342 at 8 mM in Dulbecco's modified Eagle medium (DMEM) for 5 min. Cells were thereafter washed once, resuspended in DMEM and examined by microscopy. Images were obtained using a Leica DM5000 B microscope, FLUOTAR objectives, DFC310 FX camera and LAS X software. (A) Representative image of Hoechst-labeled live cells from intramedullary canal tissue sample filtrate without enzymatic digestion, revealing mostly erythrocytes. (B) After enzymatic digestion, demonstrating the dense mineral debris that contaminates the live cell pellet. (C) After utilization of both enzymatic digestion and subsequent Ficoll density gradient medium separation. The latter combination successfully and consistently partitioned the desired fraction of viable cells away from the mineral debris. Scale bars = 20  $\mu$ m.

sample was from a normal, previously unoperated femur from which autogenous bone graft was obtained for a contralateral femur nonunion site; and acute tibia fracture sample was collected at the time of intramedullary nail (IMN) fixation. Reaming was performed using flexible reamers (Zimmer Biomet, IN, USA). For the nonunion sample, the previously placed IMN was removed, and the canal

was then reamed in preparation for receipt of a new, larger IMN. For the control sample, a standard reamer was passed prior to bone graft harvest with the Synthes reamer–irrigator–aspirator (DePuy Synthes, PA, USA), with the initial reamings from the standard reamer constituting the ICT sample. Finally, the acute tibia fracture underwent reaming for standard IMN placement for fracture treatment. In all cases, ICT was collected from the reamings, immediately placed on ice in 50-ml conical tubes and transferred to the laboratory for tissue processing. All specimens were from samples routinely collected and discarded at the time of surgery, and all were deidentified for this investigation.

## Sample preparation

The necessary supplies and reagents are shown in the reagents and materials in the protocol template. The cell isolation protocol from the ICT went through a series of trial and error revisions to obtain suitable live cell yield and minimize contaminating mineral debris. The authors' goal was to obtain a protocol that would allow cell isolation from ICT and pass scRNA-seq quality control (QC). The primary error in the design of the initial protocol was due to the assumption that cells would be organized in an easily dispersible matrix, as one would encounter with a typical bone marrow aspiration sample [10]. However, after fixing, staining and observing the ICT samples via microscopy (Figure 1), it became evident that the majority of cells were unable to be extracted by physical manipulation alone. The second major obstacle was partitioning viable cells away from contaminating mineral debris, which is voluminous and an inherent byproduct of the reaming procedure. Ultimately, a collagenase digestion followed by Ficoll density gradient separation liberated single cells and allowed for the separation of mineral debris from viable cells. Figure 2A & B displays the cellular endpoints from early failed isolations, and Figure 2C reveals the product of the final successful protocol. The final protocol, as described, was used to isolate cells from ICT for all three patient samples. Each of the three samples (acute tibia fracture, femur control and femur nonunion) was divided into samples that were freshly analyzed and samples frozen at -80°C for greater than 3 weeks. The frozen samples were then thawed, reanimated and analyzed again. Thus, the same isolation protocol was used for each of the three samples (both fresh and frozen).

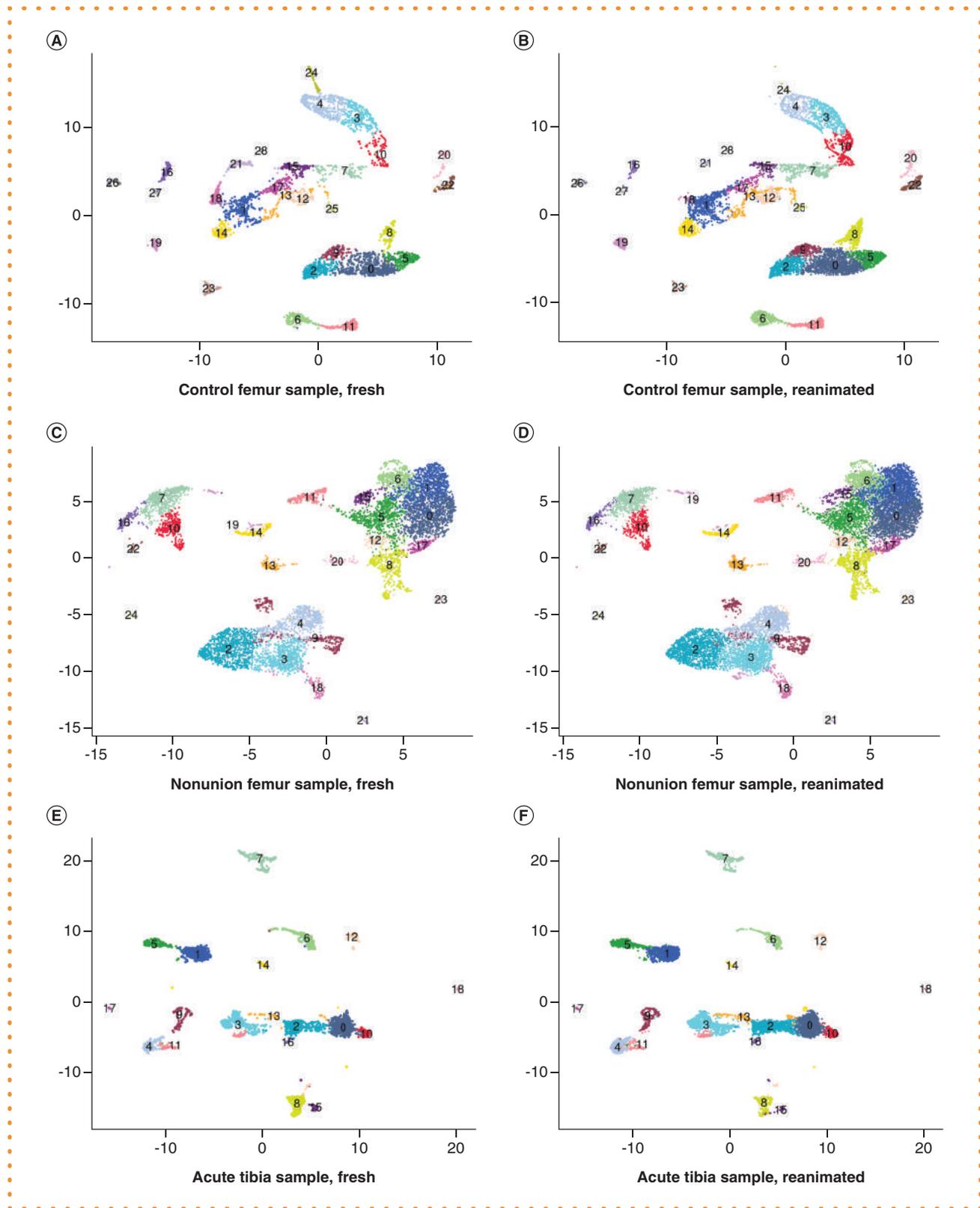
The number of cells expected for the 10x Chromium system (10x Genomics, CA, USA) varies from 100 to 10,000, depending on cell availability and the need for single-cell gene expression analysis. The Chromium system requires 60 µl of a clean single-cell suspension at 700–1200 cells/µl. The definition of clean single-cell suspension is as follows: approximately 90% viability; zero to minimal cell aggregation, particularly doublets and triplets; minimal cell debris; and final cell suspension washed at least three times. Often, when obtaining a sample from a reaming procedure, the initial pellet is very large but consists almost entirely of mineral debris. In the later steps, the pellet is often very small and barely visible. Thus, pellet size (or absence thereof) is not a criterion for successful isolation. If the cell pellet is not visible, the final cell count by hemocytometer is used to ensure adequate cell number. In the fresh state, the number of cells that passed QC was 3156, 4515 and 7770 for the acute tibia fracture, femur control and femur nonunion samples, respectively. Following reanimation, the number of cells that passed QC was 6201, 7204 and 13,006 for the acute tibia fracture, femur control and femur nonunion samples, respectively.

## scRNA-seq & bioinformatics

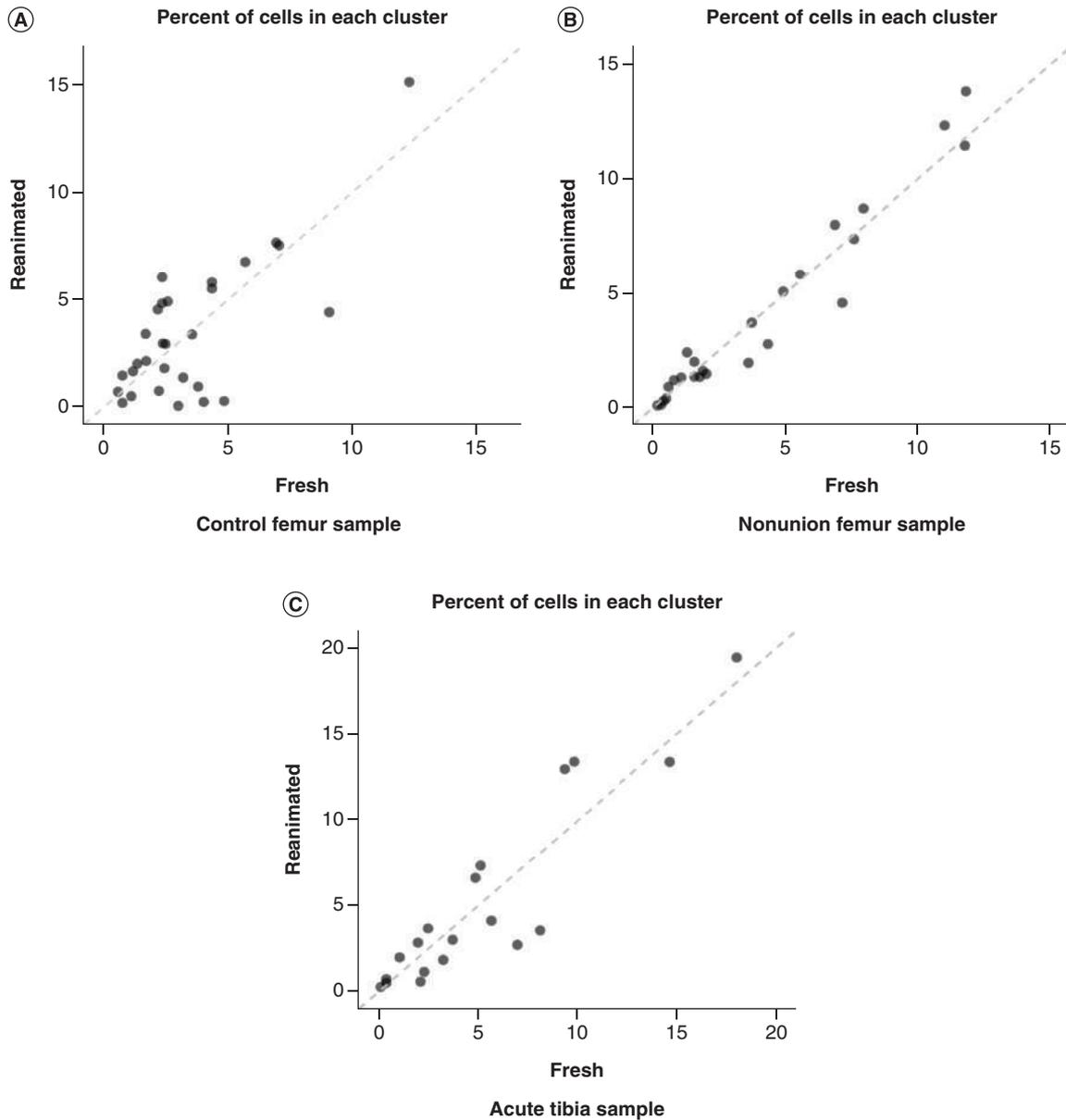
To coordinate the services of a multidisciplinary core facility with the time-intensive sample preparation and challenges inherent to surgical scheduling, it is ideal to be able to freeze samples after isolation for later analysis. To investigate whether the freeze/thaw process compromised cell viability or impaired downstream analyses, the authors compared samples analyzed on the day of collection (never frozen) with the same sample stored for greater than 3 weeks at -80°C and subsequently thawed per the recommendations of 10x Genomics [11]. Cell viability was determined manually by the core facility via trypan blue and hemocytometer prior to downstream analysis. The fresh versus reanimated cell viability was 84 versus 78% for femur nonunion, 91 versus 90% for the acute fracture and 92 versus 61% for the control sample. Of note, because of the lower percentage of viable cells after freezing the control sample, the core utilized a dead cell removal kit (Miltenyi Biotec, Auburn, CA, USA), which increased viability to 93% in the control sample used for scRNA-seq. Each sample/cell suspension was processed according to the manufacturer's instructions, as previously described [12,13]. Briefly, each sample was loaded onto a Chromium chip (10x Genomics). The 10x V3 single-cell reagent kit (10x Genomics) was used to generate single-cell gel beads for cDNA synthesis. An Agilent 2100 Bioanalyzer (Agilent, CA, USA) was used to assess the subsequent Illumina libraries, which were then sequenced via an Illumina NovaSeq 6000 (Illumina, CA, USA).

The data were analyzed with Cell Ranger 3.0.2 [14], with slight modifications from previous reports [12,13]. Briefly, FASTQ files were aligned to the human reference genome GRCh38. The aligned reads were traced back to individual cells, and gene expression was quantified based on the number of unique molecular indexes detected in each cell. R package Seurat 3.1.0 [15] with RStudio 1.1.453 and R 3.5.1 was used to further analyze the filtered gene–cell barcode matrices. Exclusion criteria were genes detected in less than five cells, cells with less than 200 genes, cells with extremely high or low numbers of detected genes/unique molecular identifiers and cells with high percentages of mitochondrial reads. The function 'is Outlier' from R package scater [16] was used in this process of data clean-up. Following exclusion as described, the resultant data were normalized with the NormalizeData function in Seurat. FindIntegrationAnchors and IntegrateData from Seurat 3.1.0 integrated each pair of fresh and reanimated samples. The integrated data were scaled and principal component analysis was performed. FindNeighbors and FindClusters were used to identify clusters. FindConservedMarkers was subsequently used to identify canonical cell type marker genes. The cell clusters were visualized using the Uniform Manifold Approximation and Projection plots. R package ggplot2 [17] was used to plot the percentage of cells in each cluster.

# Benchmark



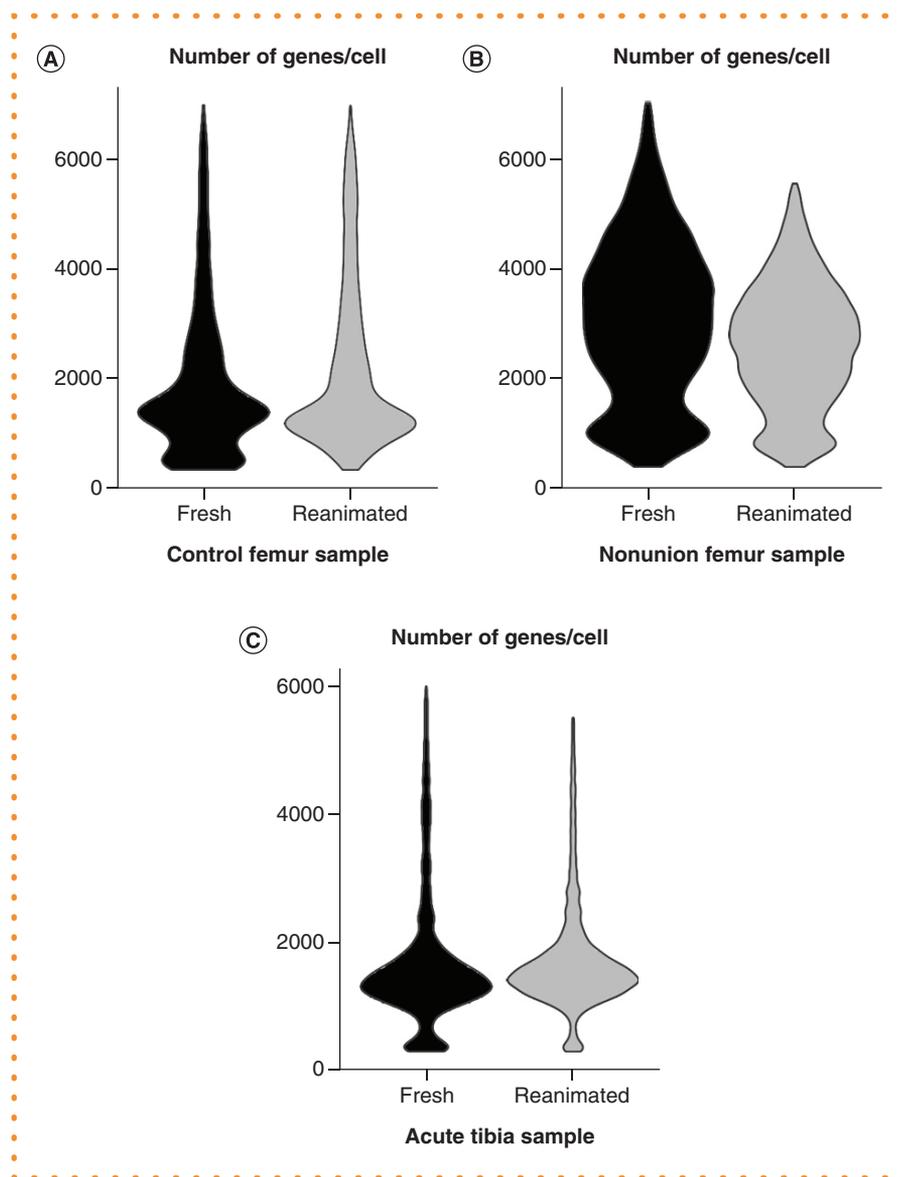
**Figure 3. Comparison of single-cell RNA sequencing cell clusters in fresh versus reanimated samples.** Uniform manifold approximation and projection comparisons of scRNA-seq analyses of the three pairs of fresh and reanimated samples. Each dot represents a cell, and each number/color represents clusters of similar cell types determined by gene expression profile, referenced to a standard data set. The figures demonstrate similar clustering in the samples when analyzed fresh or after reanimation. (A) Control femur sample, fresh; (B) control femur sample, reanimated; (C) nonunion femur sample, fresh; (D) nonunion femur sample, reanimated; (E) acute tibia sample, fresh; (F) acute tibia sample, reanimated.



**Figure 4. Comparison of cell percentages in each cluster in fresh versus reanimated samples.** The values along the x- and y-axes are the percentages of either fresh (x-axis) or reanimated (y-axis) samples that a particular cluster represents. A cluster that has an identical percentage of cells (comparing frozen and reanimated samples) would be located along the diagonal hashed line. A cluster with a large deviation of percentages between the two samples would appear as farther away from the diagonal line. Most of the dots are distributed close to the line, suggesting that freezing had minimal effect on the proportion of cells from a cluster in the overall population. (A) Control femur sample, (B) nonunion femur sample and (C) acute tibia sample.

Figure 3 compares the Uniform Manifold Approximation and Projection plots for clusters between the fresh and reanimated samples, demonstrating relative stability of the clusters. Figure 4 shows that the percentage of cells from each cluster in the overall cell population was minimally affected by reanimation and close to the line of identity. Figure 5 uses violin plots to illustrate frequency distributions for the number of genes expressed in each cell. The black (fresh) and gray (reanimated) plots illustrate the distribution within each condition. The main finding of these data was that reanimation passed scRNA-seq QC and did not prevent subsequent analyses.

Improving our understanding of failed fracture healing will have a profound effect on the diagnosis and treatment of nonunion. One potential avenue of inquiry is studying scRNA-seq of ICT. In this article, the authors have described a novel, successful and repeatable operating room-to-benchtop process for the isolation of single cells from ICT collected at the time of reaming during nonunion repair



**Figure 5. Violin plots comparing the number of genes per cell in fresh versus reanimated samples. The black (fresh) and gray (reanimated) plots represent the proportion of cells that express a given number of genes (y-axis). These figures show minimal change when comparing the fresh samples with their reanimated counterparts. (A) Control femur sample, (B) nonunion femur sample and (C) acute tibia sample.**

and acute fracture treatment. Through a series of adjustments, the authors successfully isolated single cells of adequate quality and quantity for scRNA-seq analyses. Most importantly, the authors showed that cryogenic storage of samples is possible, thereby facilitating the feasibility of this translational research technique. To the authors' knowledge, there are no published protocols describing the isolation of cells from ICT obtained by reaming that are suitable for scRNA-seq analyses. The authors acknowledge that this protocol may not yet be optimal and welcome further investigation. Additionally, there remain aspects of the comparison between fresh and reanimated samples that are underdeveloped. However, this technique article demonstrates the ability to isolate cells from ICT suitable for scRNA-seq. The authors also acknowledge that reanimated samples may not fully reflect fresh samples; however, the differences appear minimal. Pragmatism in doing these types of experiments is a relevant concern. Although some scientific questions using this protocol may require fresh samples, others may be answerable with frozen samples. The authors' protocol shows that cells passing scRNA-seq QC can be obtained from both types of samples.

## Future perspective

This communication provides a protocol for isolating cells from ICT suitable for scRNA-seq analyses. Using this protocol, investigation of cell types in each cluster or gene expression differences between nonunions and controls may yield insight into, or biomarkers for,

failed fracture healing. Ongoing research in our laboratory using this methodology is examining differences between multiple control and nonunion patients.

#### Executive summary

- Analysis of unique cellular characteristics of intramedullary canal tissue from fracture nonunion patients may provide critical insight into the etiology of, or biomarkers for, failed fracture healing.
- The authors report the feasibility of collecting, processing and analyzing intramedullary canal tissue specimens using single-cell RNA sequencing in both fresh and reanimated conditions.

#### Author contributions

The study was conceived by RM Natoli, TO McKinley, Y Liu and SM Moe. Logistics were provided by KM Brown and NX Chen. Experiments were performed by JM Dominguez. Bioinformatics analysis was generated by H Gao. The manuscript was written by JM Dominguez, SM Moe and RM Natoli. All authors participated in data interpretation and revising the manuscript and approved the final version.

#### Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

#### Financial & competing interests disclosure

This work was supported by grants from the Indiana Center for Musculoskeletal Health (multi-investigator pilot P30 award number AR072581); Indiana Clinical and Translational Sciences Institute (award number UL1TR002529); NIH, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award; Indiana Center for Musculoskeletal Health Collaborative Pilot Project Grant; and Indiana University Health Values Fund Grant (VFR-478). SM Moe is supported by P30AR072581, UL1TR002529 and R01DK110871. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

#### Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. Samples obtained were discarded products of surgery and did not contain human subject identifiers.

#### Data sharing statement

RNA sequencing data have been deposited in the National Center for Biotechnology Information Sequence Read Archive (accession number PRJNA540413).

#### Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

#### References

1. Brinker MR, Hanus BD, Sen M, O'connor DP. The devastating effects of tibial nonunion on health-related quality of life. *J. Bone Joint Surg. Am.* 95(24), 2170–2176 (2013).
2. Antonova E, Le TK, Burge R, Mershon J. Tibia shaft fractures: costly burden of nonunions. *BMC Musculoskelet. Disord.* 14, 42 (2013).
3. Kanakaris NK, Giannoudis PV. The health economics of the treatment of long-bone non-unions. *Injury* 38(Suppl. 2), S77–S84 (2007).
4. Calori GM, Mazza EL, Mazzola S *et al.* Non-unions. *Clin. Cases Miner. Bone Metab.* 14(2), 186–188 (2017).
5. Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005–2025. *J. Bone Miner. Res.* 22(3), 465–475 (2007).
6. Mills LA, Aitken SA, Simpson A. The risk of non-union per fracture: current myths and revised figures from a population of over 4 million adults. *Acta Orthop.* 88(4), 434–439 (2017).
7. Mills LA, Simpson AH. The risk of non-union per fracture in children. *J. Child Orthop.* 7(4), 317–322 (2013).
8. Zura R, Xiong Z, Einhorn T *et al.* Epidemiology of fracture nonunion in 18 human bones. *JAMA Surg.* 151(11), e162775 (2016).
9. Pearson RG, Clement RG, Edwards KL, Scammell BE. Do smokers have greater risk of delayed and non-union after fracture, osteotomy and arthrodesis? A systematic review with meta-analysis. *BMJ Open* 6(11), e010303 (2016).
10. Oetjen KA, Lindblad KE, Goswami M *et al.* Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight* 3(23), e124928 (2018).
11. 10x Genomics. Nuclei isolation for single cell ATAC sequencing (2021). <https://assets.ctfassets.net/an68im79xiti/5g035d2ngCW1aB9DFqPphO/71445a59fb282ea273a866c26cb5d319/CG000169.DemonstratedProtocol.NucleiIsolation.ATAC.Sequencing.RevD.pdf>
12. Lee J, Rabbani CC, Gao H *et al.* Hair-bearing human skin generated entirely from pluripotent stem cells. *Nature* 582(7812), 399–404 (2020).
13. Piñeros AR, Gao H, Wu W, Liu Y, Tersey SA, Mirmira RG. Single-cell transcriptional profiling of mouse islets following short-term obesogenic dietary intervention. *Metabolites* 10(12), 513 (2020).

14. 10x Genomics. Support. <http://support.10xgenomics.com/>
  15. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36(5), 411–420 (2018).
  16. Mccarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33(8), 1179–1186 (2017).
  17. Wickham H. *ggplot2*. Springer, NY, USA. doi: 10.1007/978-0-387-98141-3 (2009).
-

## Assigning immunoglobulin class from single-cell transcriptomes in IgA1-secreting versus membrane subpopulations

Colin Reilly<sup>\*1</sup>, Nuo Xu<sup>2</sup> & David K Crossman<sup>3</sup>

<sup>1</sup>Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA; <sup>2</sup>Department of Management, Information Systems & Quantitative Methods, University of Alabama at Birmingham, Birmingham, AL 35294, USA; <sup>3</sup>Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA; \*Author for correspondence: creilly@uab.edu

BioTechniques 70: 89–99 (February 2021) 10.2144/btn-2020-0044

First draft submitted: 9 April 2020; Accepted for publication: 29 October 2020; Published online: 14 December 2020

### ABSTRACT

IgA nephropathy (IgAN) is an autoimmune disease characterized by renal glomerular immunodeposits enriched for galactose-deficient IgA1 (Gd-IgA1; autoantigen) with the corresponding IgG autoantibodies. Despite the known contribution of Gd-IgA1 to IgAN, little is known concerning IgA1-secreting subpopulations responsible for autoantigen production. The goal of this study is to identify IgA1-secreting and membrane subpopulations from single-cell transcriptomic analysis. We developed a novel single-cell analytics workflow to discern cells expressing IgA1 secreted isoform or membrane-bound isoform. Multiple approaches were compared to assess immunoglobulin-isotype identity in single cells, and multiple immunoglobulin heavy-chain genes expressed in the same cells were found. To better identify specific immunoglobulin heavy-chain transcripts, we merged a software platform called Alteryx with the existing single-cell R toolkit program Seurat. This process allowed for improved calls on IgA1-secreting subpopulations based on secreting versus membrane splice-variant expression levels.

### KEYWORDS:

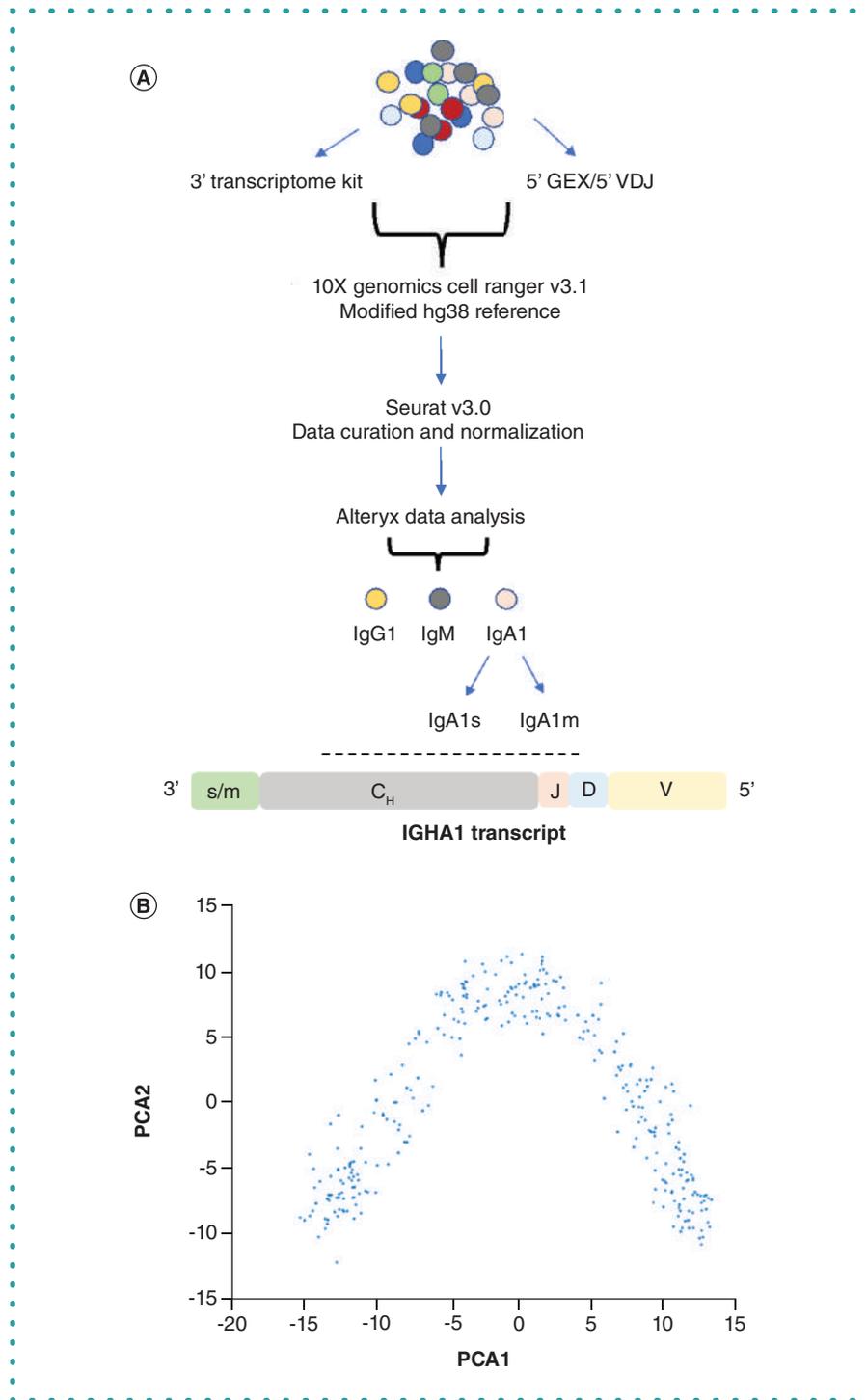
Alteryx secreting • bioinformatics • IgA nephropathy • immunoglobulin • membrane • scRNA-seq • seurat

IgA nephropathy (IgAN) is the most common primary glomerulonephritis worldwide [1]. This autoimmune disease is characterized by elevated blood levels of abnormally glycosylated IgA1 (galactose-deficient in some O-glycans; Gd-IgA1). Identifying IgA1-secreting subpopulations responsible for production of Gd-IgA1, and their unique transcriptional profile, would help identify critical mechanisms in the pathobiology of IgAN. To identify transcriptional mechanisms of Gd-IgA1 production in IgA1-secreting subpopulations, we used a previously generated biobank of Epstein–Barr virus (EBV)-immortalized B cells from peripheral blood of IgAN patients and healthy controls [2–6].

The aberrant glycosylation of Gd-IgA1 is related to abnormal expression and activity of key glycosyltransferases [2,3,5]. To assess mechanisms that drive abnormal glycosylation of IgA1 in cells from IgAN patients versus controls, we used high-throughput single-cell RNA sequencing (scRNA-seq) to map differential transcriptional responses in IgA1-secreting subpopulations. To identify IgA1-secreting cells, we used the *IGHA1* secreted splice variant. The splice variant is located on the 3' end of the transcript, but the VDJ clonality – also of interest when discerning immunoglobulin function – is on the 5' end of the gene. Thus we approached the analysis using scRNA-seq kits that targeted either the 3' or the 5' end (Figure 1). The purpose of this study was to develop a bioinformatic process to identify individual cells that have the IgA1-secreting isoform, the IgA1 membrane isoform or both isoforms; however, during our analysis of these IgA1 subpopulations, we found significant expression of other immunoglobulin heavy chains in the same cells, necessitating a process to identify which isotype class each cell should be called for.

This problem of needing sequence identification at both ends of the transcript is well recognized, and a recent publication has looked at using 3' end reverse transcription with targeted capture and nanopore sequencing [7]. This study used both whole-transcriptome analysis from the 3' end and long-read sequencing via Oxford Nanopore technology. The advantage was the availability of both whole-cell transcriptome and VDJ sequences; however, the process required significantly more PCR amplification upstream for hybridization capture, which can bias expression levels, and the nanopore sequencing technology generally has low recovery rates [7].

For our studies we needed an analysis process that could be hypothesis-driven, allowing for specific subpopulations determined *a priori*, and then we needed to perform both hypothesis testing and unbiased analysis on these subpopulations. For this, we turned to Alteryx, a software package used to handle large databases that allows users to generate unique workflows to quickly analyze data for gene expression profiles across all their datasets. Coupled with the R single-cell package 'Seurat', we generated the curated data sets from the EBV-immortalized cells derived from human peripheral blood B cells [8].



**Figure 1. Overall scheme for data curation and analysis.** (A) Multiple kits targeting the 3' or 5' end of mRNA transcripts were used. The 5' VDJ kit was used as well to selectively amplify transcripts for sequencing the VDJ region. Sequencing data were aligned in Cell Ranger 3.1, curated and normalized in Seurat, then subgrouped in Alteryx. Cells were grouped by isotype heavy chain, followed by IGHA1 secretory or membrane form (s/m). The bottom diagram depicts regions of the IgA1 transcript, from the 3' to the 5' end, with the secretory and membrane splice variant on the 3' end. (B) IGHA1s-expressing (IgA1s-secreting isoform) cells identified from the workflow were subject to PCA analysis.

Analysis of the raw and curated data from Seurat and Alteryx indicated that EBV-immortalized cells express multiple isotype classes of heavy chain genes; however, due to allelic exclusion during isotype switching, this is not biologically possible [9]. After analyzing the distribution profiles of expression of heavy chain isotypes in each cell, we defined an approach to make more appropriate calls of isotype heavy chains. With this process, we modified the hg38 reference database to include *IGHA1* splice variants for secretory and membrane-bound antibodies. This approach enabled a more accurate assessment of the immunoglobulin isotype calls and identification of critical IgA1-secreting subpopulations.

## Materials & methods

A previously established biobank of EBV-immortalized peripheral blood mononuclear cells (PBMCs) was used [2,4–6]. Briefly, PBMCs isolated from patients with IgAN or other renal disease and healthy controls underwent EBV immortalization, a process that only targets B cells. For the purposes of this study, we only used IgAN donors, but immortalized B cells from healthy controls exhibit similar immunoglobulin heavy chain patterns [2]. Heterogenic mixtures (populations secreting multiple isotypes of immunoglobulin heavy chain) of B cells were grown in RPMI 1640 medium with 10% fetal bovine serum at 5% CO<sub>2</sub> [2,4,6]. Cells were centrifuged at 4°C, stored on ice for 30 min and, for the purpose of removing clumped and dead cells, were isolated as single live cells by using forward and side scatter in an Aria II flow cytometer before single-cell transcriptomic analysis. B cells were assessed using 10× Genomics 3′ transcriptome (v2.0, n = 4) and 5′ VDJ and GEX transcriptome kits (v1.1, n = 5), with target cell numbers of 3000 [10,11]. The target cell number of 3000 was used as recommended by 10× Genomics and See *et al.* who found that approximately 2000 cells were sufficient to target isotype subpopulations [12]. Transcripts were sequenced on an Illumina NextSeq 500, at 50,000 reads per cell. The sequences were called using 10× Genomics Cell Ranger v3.1. In addition to using the hg38 reference genome, splice variants for *IGHA1* secreted (ENSG00000282633.1) and membrane-bound (ENSG00000211895.5) (*IgHA1s* and *IgHA1m*) were added to the reference genome. Raw data were curated and normalized using Seurat v3.0 and analyzed using Alteryx Designer x64<sup>®</sup> in the following manner (Figure 1; example workflow in Supplementary Figure 1) [13]. Using the 5′ VDJ and GEX 10× Genomics kit, we identified antibody isotypes from the 5′ VDJ results and imported those barcodes into the 5′ GEX data for further comparison of immunoglobulin heavy chain calls.

Matrix data from Cell Ranger were read into Seurat v3.0 using the open source RStudio (v1.2.1335) [14,15]. Seurat was used to curate and normalize the data, using the instructions listed on the website (<https://satijalab.org/seurat/v3.0/pbmc3k.tutorial.html>):

- The Seurat object was created, and mitochondrial percentage assessed for each cell.
- Violin plots were created to visualize outliers and distribution characteristics. The EBV-immortalized cells are larger than primary B cells and have a higher mitochondrial content and gene set than primary cells (for visualization of primary PBMCs, see Vignettes in [14]). Thus for quality control (QC), minimum and maximum gene features were set at 1000/5000, and mitochondrial percentage at <20% (Figure 2 & Supplementary Figure 2).
- Data were then normalized using the LogNormalize function at 10,000.
- Normalized data from the Seurat object were copied into a new matrix and saved as an .rds file for future analysis. The following commands allow pulling out the data from the Seurat object and into an .rds file for further analysis by Alteryx:
  - `normdataset <- pbmc[["RNA"]]`@data
  - `normnames <- normdataset@Dimnames`
  - `normdata <- normdataset@x`
  - `normmatrix <- matrix(normdata, ncol = "number of cells in object", nrow = "number of gene in object", dimnames = list(normnames[[1]], normnames[[2]]), byrow = TRUE)`
  - `normgenenames <- normnames [[1]].`
- Both 'normmatrix' and 'normgenenames' were saved as .rds files.

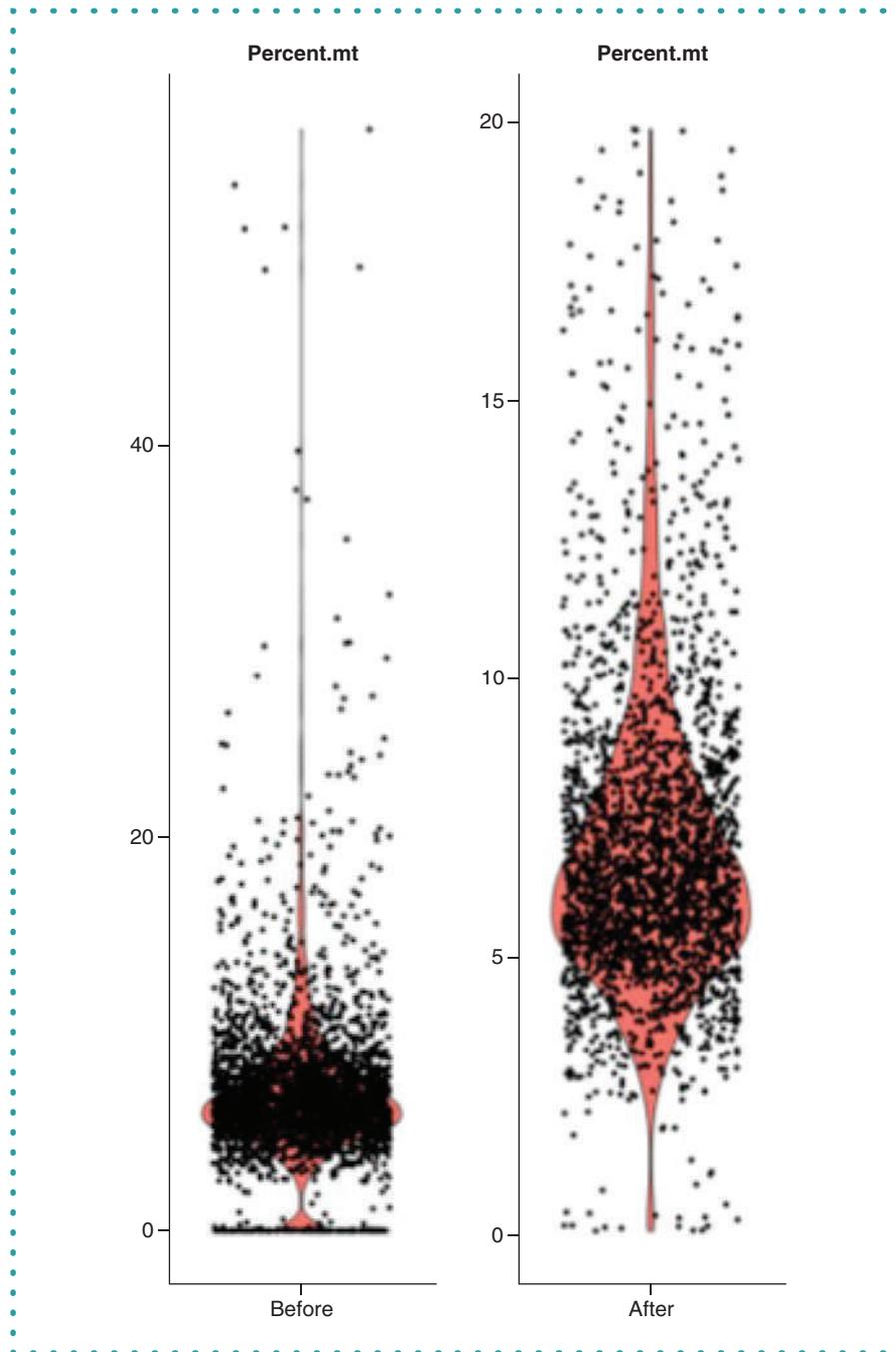
To compare isotype class calls in Seurat and in our workflow, we identified a subset of *IGHA1s*-containing cells in Seurat using the following commands only after the Seurat object had been normalized and scaled using either the standard workflow or 'SCTransform' (for SCTransform we also regressed out mitochondrial percentage as suggested):

- `IGHA1s_sub <- subset('SeuratObject', subset = IGH1s > 0)`
- `WhichCells(IGHA1s_sub)`

The cell IDs were then imported into our Alteryx workflow and used to isolate those specific cells identified by Seurat as *IGHA1s* expressers. All immunoglobulin heavy chain isotypes for each cell were assessed.

To input the .rds files into Alteryx, we set up a workflow using the R script tool, with the following commands code:

- `dat <- as.data.frame(readRDS("C:/filename/normmatrix.rds"))`
- `write.Alteryx(dat, 1)`



**Figure 2. Mitochondrial percentage in EBV-immortalized cells before and after curation.** Generally, these cells have higher expression of mitochondrial genes than their primary counterparts.

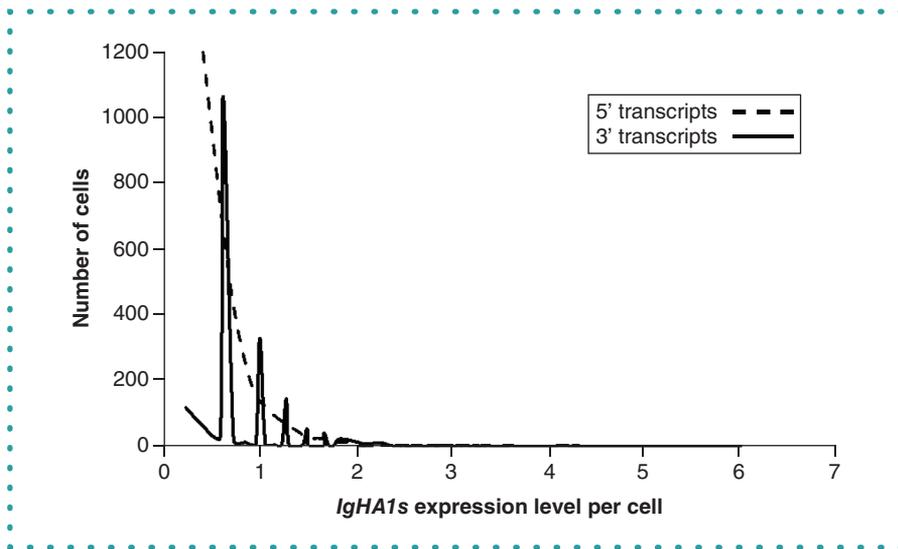
```

● dat2 <- as.data.frame(readRDS("C:/filename/normgenenames.rds"))
● write.Alteryx(dat2, 2)

```

Using the Output Data tool, we copied the new files in Alteryx database format (.yxd) to the appropriate file.

Using Alteryx, we can query and categorize the data matrices from Seurat to find and analyze subpopulations of cells based on their gene expression profiles. This is done by creating workflows that employ a suite of tools designed to manage large datasets. Single-isotype heavy chain-expressing cells were subgrouped based on their expression isotype (*IGHM*, *IGHG1* etc.), and *IGHA1* specifically into *IGHA1s* and *IGHA1m* subpopulations. Once specific subpopulations were grouped, either analysis was performed in Alteryx or the data were exported into a matrix table for further processing in R or other statistics packages.



**Figure 3.** *IGHA1s* expression across all cells in two different samples. One sample of immortalized B cells was analyzed using a 5' transcriptome kit (dashed line, 2326 cells), and the other using a 3' transcriptome kit (solid line, 3233 cells). The Y-axis represents the number of cells that, based on the x-axis *IgHA1s* expression level, expressed a specific number of transcripts.

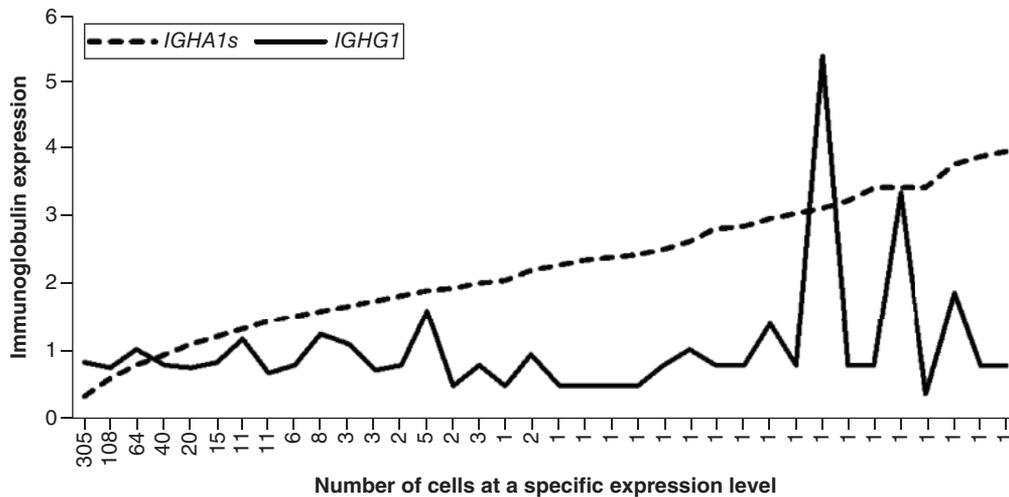
The relative homology of *IGHA1s* and *IGHG1* was assessed using cDNA transcripts from Ensembl ([www.ensembl.org](http://www.ensembl.org)) of *IGHA1* secretory (ENST00000633714.1/ ENSG00000282633.1) and *IGHG1*-201/202/203 (ENST00000390542.6/ ENST00000390548.6/ ENST00000390549.6). Additional cDNA sequences used for comparison are: *IGHG2*-202 (ENST00000641095.1), *IGHG3*-202 (ENST00000641136.1), *IGHG4*-202 (ENST00000641978.1) and *IGHM*-202 (ENST00000637539.2).

## Results & discussion

Immortalized B cells and IgA1-secreting cell lines derived from peripheral blood of patients with IgAN and healthy controls were used as model systems for mechanistic studies of Gd-IgA1 production [2]. EBV immortalization targets cells with CD21 receptors and thus creates a heterogenic mixture of B cells and immunoglobulin-secreting cells [16]. A mature B cell can be activated by a specific antigen and various cytokines to undergo a class switch from IgM/IgD to express other antibodies (IgA1/IgA2/IgG1-4/IgE). Once this occurs, the plasma cell or memory B cell does not produce any other antibody isotype [17]. With this in mind, we compared the expression of *IGHA1s* across all cells in two tested samples (Figure 3). As expected, we saw a range of expression levels, but we did not expect to see the *IGHA1s* gene expressed in all cells. As both samples were from a mixture of B cells that also secrete IgG and/or IgM (data not shown), we can assume that some or even a majority of the gene calls for *IGHA1s* were incorrect, and that most of the *IGHA1s* calls constituted background noise. These observations appear contradictory to the conventional doctrine of 'one B cell, one antibody'; however, another publication found similar results [18]. Using the same 10× Genomics 5' VDJ and GEX platform, Shi *et al.* found that a significant number of B cells (~10–40%) at various stages of differentiation express multiple immunoglobulin heavy chain transcripts [18]. Although the study did not assess the secretory sequence of the immunoglobulin heavy chain transcripts, the results were consistent with our findings.

To better elucidate the potential for confounding isotype heavy chain calls in the same cell, we compared *IGHA1s* expression across all cells with the ratio of other heavy chain isotypes (Figure 4). Figure 4A shows the relationship between *IGHA1s* and *IGHG1* expression in individual cells from a heterogenic B cell population assessed with a 5' kit. This figure shows that, while there is some variability, as cells increase *IGHA1s* expression the relative amount of *IGHG1* also expressed appears stable. Figure 4B, a close-up panel of Figure 4A, highlights that the majority of cells expressing *IGHA1s* are low expressers and likely constitute background noise. Figure 4C shows the exact same cells as in Figure 4A & B, but includes information on expression of all other heavy chain isotypes for every cell. These data indicate that across all cells, there will be some level of expression of multiple heavy chain isotypes, confounding the ability to make an isotype-specific designation because the ratio of other isotype heavy chains is variable. This phenomenon is consistent across multiple cell lines, the only difference being the variability of contaminating heavy chain isotypes in different samples (Supplementary Figure 3A–C). We employed the 5' VDJ-GEX kit from 10× Genomics, which sequences the VDJ transcripts in the same cells along with 5' GEX transcriptomes. Using the Loupe Browser from 10× Genomics, we incorporated the 5' VDJ data into the 5' GEX transcriptome data and exported the cell barcode IDs specific to *IGHA1*-positive cell calls based on the VDJ analysis [19]. We used this subpopulation of cells to compare *IGHA1s* and *IGHG1* expression (Figure 5). Although the pattern of *IGHG1* expression to *IGHA1s* was slightly different (the cells in Figure 5 are the same cells as in Figure 4), we still found significant *IGHG1* expression across a significant number of moderate- to high-expressing *IGHA1s* cells using the 5' VDJ scRNA-seq kit. While the power of this approach can yield sequence data on the VDJ





**Figure 5.** *IGHA1s* expression level compared with *IGHG1* expression level. Single-cell 5' VDJ PCR was used to identify *IGHA1*-positive cells. These data were added to the 5' GEX transcriptome in 10× Genomics Cell Loupe browser on 2218 immortalized B cells. *IGHA1s* and *IGHG1* expression levels were derived from the 5' GEX data. The dotted line shows the *IGHA1s* expression level and the solid line shows the *IGHG1* expression level.

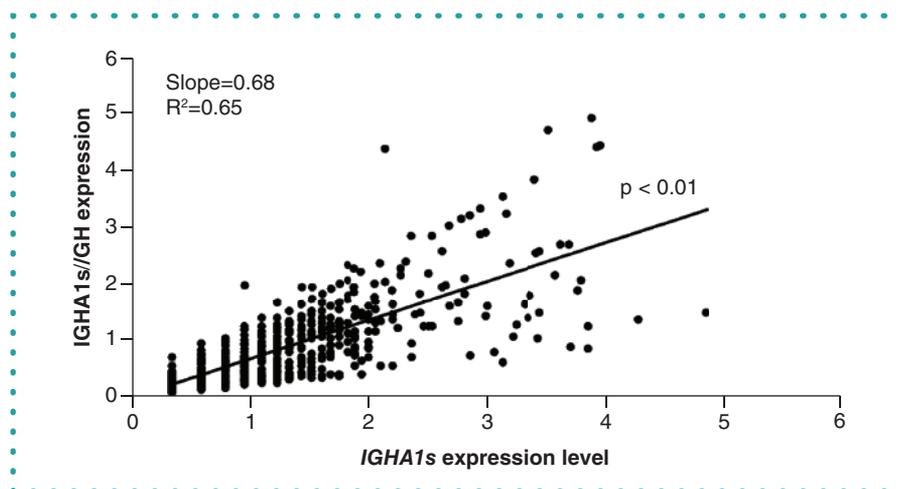
regions when combined with transcriptome profiling, the limitation is the poor coverage at the 3' end where the secreted/membrane sequence is located.

Structural homology can be an issue in overlapping calls of similar sequences, but there is limited alignment similarity in the immunoglobulin heavy chains [20]. Given that we were observing expression of multiple heavy chain isotypes in the same cell, the simplest answer is that we could be seeing doublets in our capture (i.e., two cells in one droplet), or another mRNA capture overlap from lysed cells. Therefore we checked whether the initial QC steps performed in Seurat to remove data outliers were sufficient. We compared our normal QC workflow with another one with a lower nFeature (3500) cutoff, a lower nFeature and nCount (15,000) and the new SCTransform function in Seurat (VlnPlots shown in Supplementary Figure 4). We found that these modified workflows did not affect the multiple isotype *IGH* calls in our cells (Supplementary Figure 5).

Given the overlap in calls among the different heavy chain isotypes, we needed to find the best route to assign the appropriate antibody isotype. A previous publication analyzing transcriptome and VDJ clonality in B cells from breast cancer tissue found similar issues with assigning isotype class and used a cut-off value of tenfold higher than the next highest expressing heavy chain isotype in that cell [21]. However, because each cell line likely has a unique proportion of immunoglobulin producers in the total population, using a predetermined expression cutoff would not work well. Using *IGHA1s*, we mapped the relative ratio of its expression to the next highest expressed heavy chain isotype (*IGH*) for each cell (Figure 6). We can calculate this relationship using the slope of the *IGHA1s/IGH*, and this can help determine whether the informatic calls for a positive *IGHA1s* cell are improving. It is important to note that the expression levels have been log-normalized, so biological expression differences are much larger. Although we found a significant positive correlation between cells with higher *IGHA1s* expression and an increased ratio of *IGHA1s/IGH* expression, we do not believe we should use a simple expression cut-off measurement to make appropriate isotype calls, because some cells with *IGHA1s* expression had low *IGHA1s/IGH* expression ratios (Figure 6). This phenomenon held true for three other cell lines assessed (Supplementary Figure 6).

Because we found consistent overlap in the heavy chain isotype calls in the same cell, we assessed expression across multiple heterogenic cell lines from various donors to determine what may be an appropriate ratio to make for a cell heavy chain isotype call. The ratio of *IGHA1s*/next highest *IGH* expressed was calculated for each cell across multiple cell lines and transcriptome PCR kits (3' transcriptome, 5' GEX and 5' VDJ). The number of cells that met specific *IGHA1s/IGH* ratio minimums were tabulated and presented as a percentage of total population (ratios >2, >3, >4; Table 1). Additionally, the average *IGHA1s/IGH* ratio and standard deviation (SD) were calculated for the total population and used to find cells that fell within one or two SD higher than the average ratio; these were presented as a percentage of the total population (1SD, 2SD; Table 1). As shown in Table 1, there were substantial differences between cell lines in the amount of *IGHA1s* expression, but also in the relationship between the SD percentage and the hard ratio cutoff numbers. We did not see a consistent relationship between 1SD or 2SD cell numbers and the hard cutoff ratio numbers (>2, >3, >4), suggesting that biological variability may confound any attempt to assign isotype based on SD.

To assess the overlap of *IGHA1* secreted and membrane splice variants, ratios for *IGHA1s/IGHA1m* were calculated for every cell (Figure 7). A pattern similar to that seen in Figure 6 was observed, but with higher ratios associating with higher expression of *IGHA1s*. A significant difference between *IGHA1s/m* splice variants and *IGHA1s* and other *IGH* expression is the substantially higher slope value



**Figure 6. Plot of *IGHA1s/IGH* versus *IGHA1s* expression.** To determine the relationship of *IGHA1s* expression to other immunoglobulin heavy chains in the same cell, we calculated the slope of *IGHA1s* expression versus the ratio of *IGHA1s*/next highest isotype heavy chain expression in the same cell using the 5' GEX transcript data. For each cell, all immunoglobulin heavy chain expressions were assessed, and the ratio of *IGHA1s* to the highest or next-highest immunoglobulin heavy chain (*IGHG1/2/3/4*, *IGHM*, *IGHD*, *IGHD*) was calculated. Total number of cells = 2218. Statistical significance was calculated using analysis of variance regression.

**Table 1. Percentage of cells out of the total population falling within a specific *IGHA1s*/next-highest isotype heavy chain expression ratio.**

5' GEX	Percentage of total population, <i>IGHA1s/IGH</i> ratio					<i>IGHA1s/IGH</i> ratio	
	>2	>3	>4	1SD <sup>†</sup>	2SD <sup>†</sup>	Avg	SD
<b>Cell line A</b>	1.76	0.54	0.23	11.45	4.24	0.48	0.49
<b>Cell line B</b>	4.26	1.60	0.70	9.46	4.46	0.63	0.67
<b>Cell line C</b>	6.96	3.19	1.84	8.21	3.77	0.86	0.91
<b>5' VDJ</b>							
<b>Cell line A</b>	0.45	0.09	0.09	2.93	1.26	0.46	0.48
<b>Cell line B</b>	3.36	1.31	0.66	7.41	3.48	0.62	0.67
<b>Cell line C</b>	0.48	0.14	0.00	1.21	0.48	0.80	0.64
<b>3' GEX</b>							
<b>Cell line D</b>	8.02	3.66	1.68	9.93	4.50	0.85	0.87
<b>Cell line E</b>	4.06	1.59	0.77	9.72	4.26	0.69	0.63
<b>Cell line F</b>	9.84	4.95	2.72	9.84	4.58	0.93	1.07
<b>Cell line G</b>	3.23	1.44	0.58	9.09	3.69	0.60	0.63

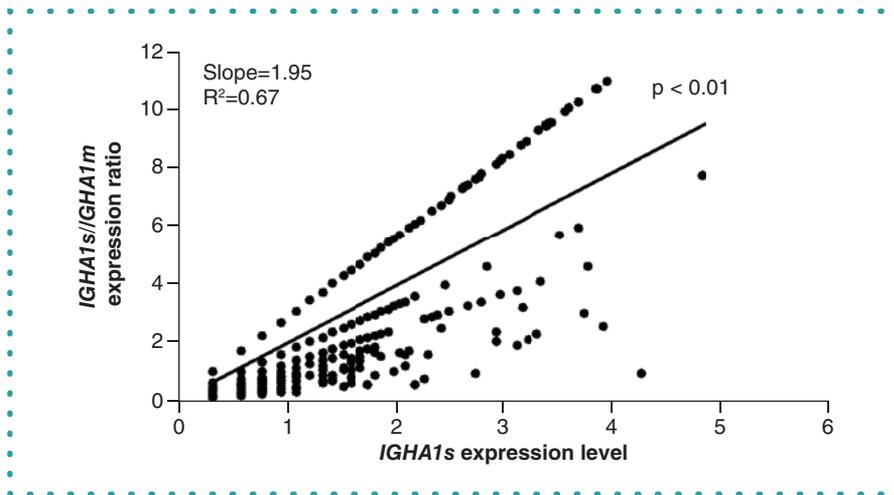
Three cell lines (A, B and C) were analyzed using two different kits: 5' GEX and 5' VDJ. Four cell lines (D, E, F and G) were analyzed using the 3' GEX kit. Each cell was analyzed for its *IGHA1s/IGH* ratio using the highest other expressing immunoglobulin heavy chain in that cell. The total number of cells meeting the ratio criteria of >2, >3 and >4 were added up and presented as a percentage of the total population. The average *IGHA1s/IGH* ratio for all the cells in the population was calculated, as well as the SD of the total population (right-hand column). These were used to calculate how many cells had *IGHA1s/IGH* ratios higher than one or two SD above the average.

<sup>†</sup> 1 or 2 SD higher than the average ratio (Avg + 1\*SD or Avg + 2\*SD).

Avg: Average; IGH: Immunoglobulin heavy chain isotype; *IGHA1s*: IgA1-secreting; SD: Standard deviation.

for *IGHA1s/IGHA1m* ratios (1.95 vs 0.68; Figures 6 & 7). It is unclear whether this greater specificity in calls for *IGHA1s* versus *IGHA1m* versus delineation from other *IGH* has a technical or biological underpinning. As shown in Table 2, the percentage of cells with high ratios of *IGHA1s/IGHA1m* was substantially greater than the comparisons shown in Table 1. This observation further supports the idea that making the call between *IGHA1s* and *IGHA1m* splice variants is easier than for alternative *IGH* expressers when comparing with *IGHA1s*. However, while most antibody-secreting cells do not express cell-surface B cell receptor, there are data suggesting that some antibody-secreting cells have an active B cell receptor [22]. This means that there will be subpopulations of cells that are only IgA1 secretors, some that are only IgA1 presenters and some that are both.

Screening for only cells that have a ratio >2 of *IGHA1s/IGH*, we calculated the *IGHA1s/IGHA1m* ratio and plotted against *IGHA1s* expression level. This approach provided a mechanism to screen cells for *IGHA1s* specificity versus other *IGH* expressers and determine the splice variants for *IGHA1* secretion, membrane or both within a cell. Figure 8A shows the relationship between *IGHA1s/IGHA1m* in cells where the *IGHA1s* expression ratio was >2 compared with expression of all other *IGH* isotypes. The slope (and ratios) in Figure 8A



**Figure 7.** Plot of *IGHA1s/IGHA1m* versus *IGHA1s* expression. To determine the relationship of *IGHA1s* expression to *IGHA1m* in the same cell we calculated the slope of *IGHA1s/IGHA1m* versus *IGHA1s* expression. *IGHA1s* and *IGHA1m* expression and their ratios in the same cell were calculated from 5' GEX transcript data. For each cell, *IGHA1s* and *IGHA1m* expressions were assessed, and the ratio of the two was calculated. Total number of cells = 2218. Statistical significance was calculated using analysis of variance regression.

**Table 2.** Percentage of cells out of the total population falling within a specific *IGHA1s/IGHA1m* ratio.

5' GEX	Percentage of total population, <i>IGHA1s/IGHA1m</i>					<i>IGHA1s/IGHA1m</i> ratio	
	>2	>3	>4	1SD <sup>†</sup>	2SD <sup>†</sup>	Avg	SD
<b>Cell line A</b>	23.22	11.99	4.82	12.17	4.10	0.48	0.49
<b>Cell line B</b>	16.99	10.81	5.94	9.05	4.50	0.63	0.67
<b>Cell line C</b>	56.62	47.20	20.24	9.95	3.29	0.86	0.91
<b>5' VDJ</b>							
<b>Cell line A</b>	6.45	3.25	1.35	3.43	1.22	0.46	0.48
<b>Cell line B</b>	13.55	8.56	4.67	8.56	3.81	0.62	0.67
<b>Cell line C</b>	3.96	3.19	1.35	0.72	0.24	0.80	0.64
<b>3' GEX</b>							
<b>Cell line D</b>	20.09	9.58	5.14	33.71	4.59	0.85	0.87
<b>Cell line E</b>	19.07	7.42	3.92	10.74	4.55	0.69	0.63
<b>Cell line F</b>	16.08	8.23	4.83	9.93	4.83	0.93	1.07
<b>Cell line G</b>	19.94	10.97	6.14	9.92	4.30	0.60	0.63

Three cell lines (A, B and C) were analyzed from two different kits, 5' GEX and 5' VDJ. Four cell lines (D, E, F and G) were analyzed from the 3' GEX kit. Each cell was analyzed for its *IGHA1s/IGHA1m* ratio. The total number of cells meeting the ratio criteria of >2, >3 and >4 were added up and presented as a percentage of the total population. The average *IGHA1s/IGHA1m* ratio for all the cells in the population was calculated, as well as the SD of the total population (right-hand column). These were used to calculate how many cells had an *IGHA1s/IGHA1m* ratio higher than one or two SD above the average.

<sup>†</sup> 1 or 2 SD higher than the average ratio (Avg + 1\*SD or Avg + 2\*SD).

Avg: Average; IGHA1m: IgA1 membrane-bound; IGHA1s: IgA1-secreting; SD: Standard deviation.

is significantly higher than the slope in the total cell population in *IGHA1s/IGHA1m* analysis (Figure 7A), indicating that selecting for lower background of other non-*IGHA1* *IGH* transcripts can increase the specificity for *IGHA1s*-expressing cells. Figure 8B shows the relationship of *IGHA1s* to all other *IGH* isotypes expressed in the same cell after screening for a ratio >2. This workflow compares well for isotype specificity (*IGHA1s* in this case) when viewed against the overall isotype overlap in Figure 4C & Figure 6.

An alternate workflow to assess splice variants at the 3' end using nanopore technology has also been employed with the 10x Genomics 3' kit. The cDNA was split for gene expression analysis and whole sequencing of antibody genes, providing significant coverage on the 3' end for secretion versus membrane delineation. This approach requires both nanopore and next-generation sequencing technologies and has a lower recovery of cell barcodes, but could potentially be used alongside a standard 3' gene expression workflow [7].

Using a combination of R tools to curate the data and Alteryx to visualize and subcategorize cell populations, we can better identify critical IgA1-producing cells with more accurate isotype calls. With this bioinformatic and single-cell transcriptome approach, we can cast a broader net to assess heterogenic populations of IgA1-secreting cells using complex mixtures of immortalized as well as primary B cells.



## Author contributions

C Reily designed the study, performed data analysis and drafted the manuscript. N Xu performed data analysis. D Crossman designed the 10× Ranger pipeline and performed data analysis. All authors were responsible for manuscript editing.

## Acknowledgments

The authors thank J Novak and B Julian for providing the immortalized cell lines (with IRB approval) and for their advice on this project, and L Shanrun for his support in running and executing the single-cell transcriptomics core at the University of Alabama at Birmingham. C Reily appreciates J Novak and A Agarwal for their mentoring during the K01 award.

## Financial & competing interests disclosure

This work was supported by grants from the National Institutes of Health (R03DK122194, K01DK106341) and University of Alabama at Birmingham Center for Clinical and Translational Science (UL1TR003096). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

## Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval through the University of Alabama at Birmingham IRB (#070413006). In addition, informed consent was obtained from all donors and participants.

## Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## References

- Wyatt RJ, Julian BA. IgA nephropathy. *N. Engl. J. Med.* 368(25), 2402–2414 (2013).
- Suzuki H, Moldoveanu Z, Hall S *et al.* IgA1-secreting cell lines from patients with IgA nephropathy produce aberrantly glycosylated IgA1. *J. Clin. Invest.* 118(2), 629–639 (2008).
- Suzuki H, Raska M, Yamada K *et al.* Cytokines alter IgA1 O-glycosylation by dysregulating C1GalT1 and ST6GalNAc-II enzymes. *J. Biol. Chem.* 289, 5330–5339 (2014).
- Yamada K, Huang ZQ, Raska M *et al.* Inhibition of STAT3 signaling reduces IgA1 autoantigen production in IgA nephropathy. *Kidney Int. Rep.* 2(6), 1194–1207 (2017).
- Kiryuk K, Li Y, Moldoveanu Z *et al.* GWAS for serum galactose-deficient IgA1 implicates critical genes of the O-glycosylation pathway. *PLoS Genet.* 13(2), e1006609 (2017).
- Yamada K, Huang Z, Raska M, Reily CR, Aderson J, Suzuki H *et al.* LIF signaling enhances production of galactose-deficient IgA1 in IgA nephropathy. *Kidney Dis.* 6(3), 168–180 (2020).
- Singh M, Al-Eryani G, Carswell S *et al.* High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* 10(1), 3120 (2019).
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36(5), 411–420 (2018).
- Early P, Hood L. Allelic exclusion and nonproductive immunoglobulin gene rearrangements. *Cell* 24(1), 1–3 (1981).
- Zheng GX, Lau BT, Schnell-Levin M *et al.* Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* 34(3), 303–311 (2016).
- Lindeman I, Emerton G, Mamanova L *et al.* BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. *Nat. Methods* 15(8), 563–565 (2018).
- See P, Lum J, Chen J, Ginhoux F. A single-cell sequencing guide for immunologists. *Front. Immunol.* 9, 2425 (2018).
- Alteryx, Inc. <http://www.alteryx.com>
- Satija Lab. Seurat software. <https://satijalab.org/seurat/>
- RStudio. <https://rstudio.com>
- Sinclair AJ, Farrell PJ. Host cell requirements for efficient infection of quiescent primary B lymphocytes by Epstein–Barr virus. *J. Virol.* 69(9), 5461–5468 (1995).
- Owen JA, Punt J, Stranford SA, Jones PP, Kuby J. *Kuby Immunology (7th Edition)*. WH Freeman, NY, USA (2013).
- Shi Z, Zhang Q, Yan H *et al.* More than one antibody of individual B cells revealed by single-cell immune profiling. *Cell Discov.* 5, 64 (2019).
- 10× Genomics. What is Loupe browser? <https://support.10xgenomics.com/single-cell-gene-expression/software/visualization/latest/what-is-loupe-cell-browser>
- Janeway CA, Travers P, Walport M, Schlomchik MJ. *Structural Variation in Immunoglobulin Constant Regions*. Garland Science, NY, USA (2001).
- Hu Q, Hong Y, Qi P *et al.* An atlas of infiltrated B-lymphocytes in breast cancer revealed by paired single-cell RNA-sequencing and antigen receptor profiling. *bioRxiv* doi:10.1101/695601 (2019). [Epub ahead of print]
- Pinto D, Montani E, Bolli M *et al.* A functional BCR in human IgA and IgM plasma cells. *Blood* 121(20), 4110–4114 (2013).
- Stoeckius M, Hafemeister C, Stephenson W *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14(9), 865–868 (2017).

# Using single-cell approaches to investigate early embryonic development

There have been rapid developments in single-cell sequencing and imaging techniques over recent years. Such approaches are particularly well suited to studying embryonic development during the earliest cell specification events. The collection of single-cell data has to go hand-in-hand with using tailored machine learning methods that drive the generation of hypotheses and the design of validation experiments.

In this webinar, we will discuss how the combination of single-cell techniques, machine learning, and developmental biology, is delivering new insights into the early events controlling embryonic development in both mice and humans.

## What will you learn?

- How to analyze single-cell transcriptomic datasets taking advantage of the most recent machine learning methods
- How to generate hypotheses from sequencing data and test them with orthogonal single-cell approaches

## Speakers



Antonio Scialdone (Group Leader, Helmholtz Zentrum München, Germany; left) studied at the University of Naples “Federico II” (Italy), where he received a PhD in physics. He then worked as a postdoc at the John Innes Centre (Norwich, UK) in the lab of Martin Howard, and at the EMBL-EBI (Cambridge, UK) in John Marioni’s lab. Antonio established his own independent lab in 2017 at the Helmholtz Zentrum München. In his lab, he combines machine learning and physical modeling to understand cellular fate decision starting from single-cell data.

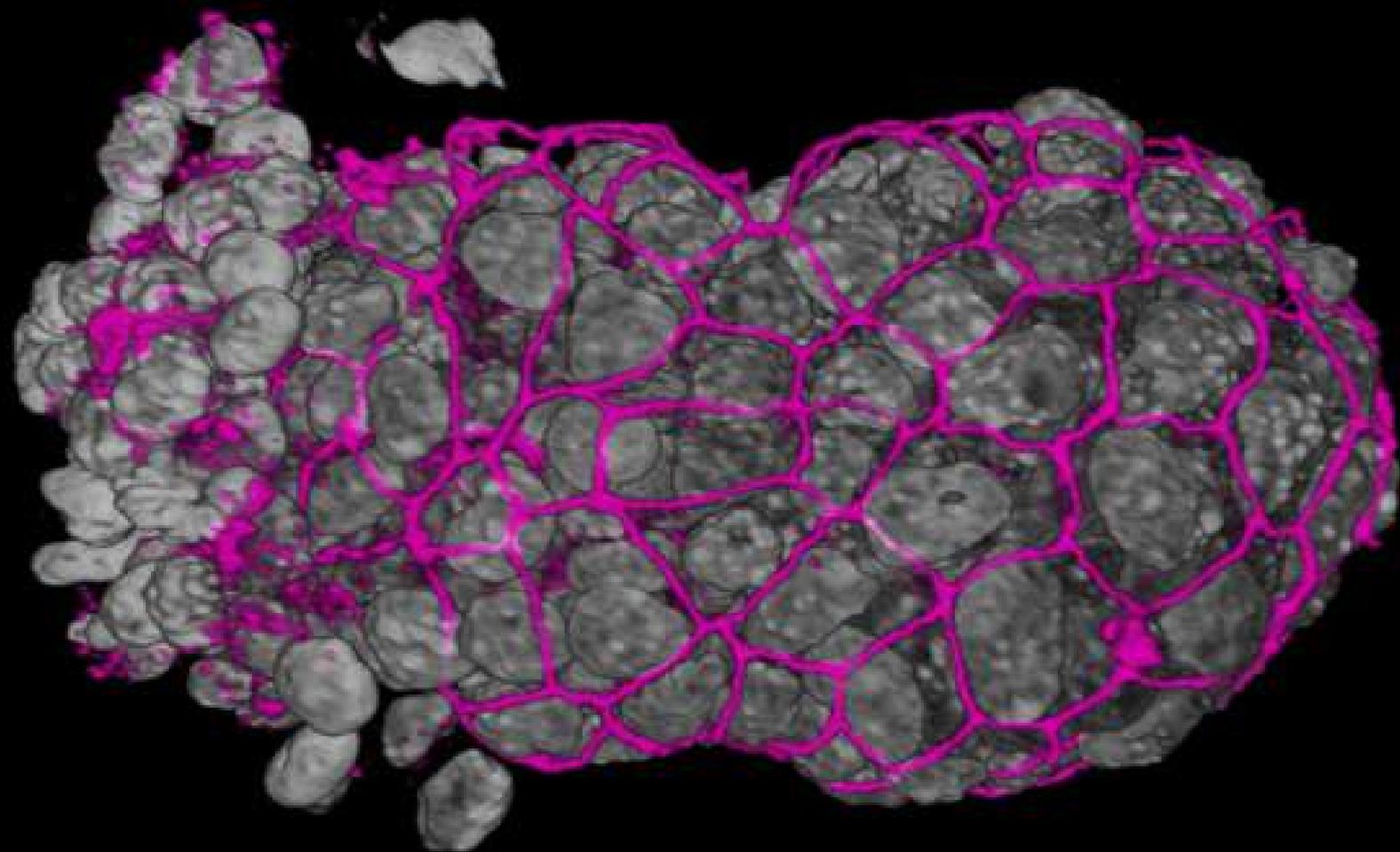


Shankar Srinivas (Professor of Developmental Biology, University of Oxford, UK; left) completed a BSc in Nizam College (Hyderabad, India) before joining the group of Frank Costantini at Columbia University (NY, USA), where he received a PhD for work on kidney development. He then moved to the National Institute for Medical Research (London, UK), as an HFSPO fellow in the groups of Rosa Beddington and Jim Smith. Here, he pioneered the use of time-lapse microscopy of early mouse embryos to study anterior patterning. He established his own independent group at the University of Oxford in 2004, where he now uses mouse and human embryos to understand the formation of the anterior-posterior axis, gastrulation and early cardiogenesis.



Follow the QR code to  
access the webinar:





## Contact us

### Editorial Department

Senior Editor

Tristan Free

[tfree@biotechniques.com](mailto:tfree@biotechniques.com)

### Business Development and Support

Commercial Director

Sarah Mayes

[s.mayes@future-science-group.com](mailto:s.mayes@future-science-group.com)

This supplement is brought to you by *BioTechniques* in association with:

**10x**  
GENOMICS