

Article

Not peer-reviewed version

Characterization and Optimization of Multiomic Single-Cell Epigenomic Profiling

Leticia Sandoval , [Wazim Mohammed Ismail](#) , Amelia Mazzone , [Mihai Dumbrava](#) , Jenna Fernandez , Amik Munankarmy , Terra Lasho , [Mortiz Binder](#) , Vernadette Simon , Kwan Hyun Kim , Nicholas Chia , Jeong Heon Lee , [S. John Weroha](#) , Mrinal M. Patnaik , [Alexandre Gaspar-Maia](#) *

Posted Date: 8 May 2023

doi: 10.20944/preprints202305.0530.v1

Keywords: single-cell sequencing; epigenomic profiling; snATAC-seq; snRNA-seq; nuclei preparation



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Characterization and Optimization of Multiomic Single-Cell Epigenomic Profiling

Leticia Sandoval ^{1,2,‡}, Wazim Mohammed Ismail ^{1,2,‡}, Amelia Mazzone ^{1,2}, Mihai Dumbrava ^{1,2,3}, Jenna Fernandez ⁴, Amik Munankarmy ^{1,2}, Terra Lasho ⁴, Moritz Binder ^{2,4}, Vernadette Simon ⁵, Kwan Hyun Kim ², Nicholas Chia ⁶, Jeong-Heon Lee ^{1,2}, John Weroha ⁷, Mrinal Patnaik ^{2,4} and Alexandre Gaspar-Maia ^{1,2,*}

¹ Division of Experimental Pathology, Department of Lab Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

² Epigenomics Program, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA

³ Mayo Clinic Medical Scientist Training Program, Mayo Clinic Alix School of Medicine and Mayo Clinic Graduate School of Biomedical Sciences, Mayo Clinic, Rochester, MN, USA

⁴ Division of Hematology, Department of Internal Medicine, Mayo Clinic, Rochester, MN, USA

⁵ Medical Genome Facility, Genome Analysis Core, Mayo Clinic, Rochester, MN, USA

⁶ Microbiome Program, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA

⁷ Department of Medical Oncology, Mayo Clinic, Rochester, MN, USA

* Correspondence: maia.alexandre@mayo.edu

‡ These authors contributed equally to this work.

Abstract: The snATAC + snRNA platform allows epigenomic profiling of open chromatin and gene expression with single-cell resolution. The most critical assay step is to isolate high-quality nuclei to proceed with droplet-base single nuclei isolation and barcoding. With the increasing popularity of multiomic profiling in various fields, there is a need for optimized and reliable nuclei isolation methods, mainly for human tissue samples. Herein we compared different nuclei isolation methods for cell suspensions, such as peripheral blood mononuclear cells (PBMC, n=18) and a solid tumor type, ovarian cancer (OC) (n=18), derived from debulking surgery. Nuclei morphology and sequencing output parameters were used to evaluate the quality of preparation. Our results show that NP-40 detergent-based nuclei isolation yields better sequencing results than collagenase tissue dissociation for OC, significantly impacting cell identification and analysis. Given the utility of applying such techniques to frozen samples, we also tested frozen preparation and digestion (n=6). A paired comparison between frozen and fresh samples validated the quality of both specimens. Finally, we demonstrate the reproducibility of scRNA and snATAC + snRNA platform, by comparing the gene expression profiling of PBMC. Our results highlight how the choice of nuclei isolation methods is critical for obtaining quality data in multiomic assays. It also shows that the measurement of expression between scRNA and snRNA is comparable and effective for cell type identification.

Keywords: single-cell sequencing; epigenomic profiling; snATAC-seq; snRNA-seq; nuclei preparation

1. Introduction

Single-cell (sc) biology is a rapidly evolving field that interrogates various 'omics layers (genomics, epigenomics, proteomics) at SC resolution to define cellular heterogeneity in homeostasis and disease [1–3]. In order to ensure reliable results that provide the resolution and quality necessary to make meaningful scientific conclusions and to improve reproducibility, experimental protocols need to be optimized, evaluated, and standardized. This proves to be a challenge because several factors could affect the quality of the analyzed samples, leading to poor data quality. These include the type of tissue (PBMC, tumor, bone marrow, or cell line) which can impact the nuclear isolation protocol; the biological condition being studied (homeostasis, development, cancer, infection, or inflammation); the method of sample collection, handling, storage, and maintenance; cell viability and the number of cells

available. Moreover, most protocols are typically defined using ideal samples that do not reflect the clinical and translational research environment, the so-called “real world” samples.

Epigenomic profiling of DNA accessibility (open chromatin regions) helps create a genome-wide characterization of transcription factor (TF) binding events that orchestrate gene expression [4]. Open chromatin regions represent complementary information to gene expression that allows identifying TF and its targets, which are typically missed by transcriptome assays, given its low expression [5]. TF activity signatures, also known as TF footprinting, can be traced back to the identity of the cell [6,7], making it a valuable resource for characterizing tissue heterogeneity. Combining single-cell level data from both gene expression and open chromatin allows us to define sub-populations at higher resolution and study the complex interactions between the genes and the TFs that regulate them. The 10x Genomics Chromium single nucleus snATAC + snRNA platform allows us to profile both open chromatin using the assay for transposase-accessible chromatin (ATAC) [8] and gene expression (GEX) information from the same nuclei, bypassing the need for computational integration of this information retrieved from non-paired cells.

In this study, we characterize experimental protocol optimization for the snATAC + snRNA assay that has been utilized in our laboratory in a variety of studies, including mouse mammary stem cells [9], patient-derived xenografts (PDX) and patient samples from ovarian cancer, and blood samples from COVID-19 patients [10]. Performing simultaneous snATAC and snRNA-seq on the same cells requires careful sample preparation and sequencing involving several steps that will be discussed here: cell isolation (single-cell suspension from the tissue or sample of interest using appropriate methods, such as enzymatic digestion or mechanical dissociation); nuclei isolation (separate the nuclei from the cells using appropriate methods, such as lysis and centrifugation); data quality analysis (analyze the sequencing data to perform integrated analysis of snATAC-seq and snRNA-seq data using appropriate bioinformatics tools).

In contrast to the snATAC + snRNA assay that captures only nuclear transcripts (snRNA), scRNA-seq is an assay that captures transcripts from the whole cell (scRNA), which includes both the nucleus and the cytoplasm. Each method has specific advantages and disadvantages. While scRNA provides a complete transcriptomic profile about individual cells and allows for better identification of rare cell types, the requirement for enzymatic dissociation of samples which may disrupt gene expression profiles, and the time-sensitive nature of the technique makes obtaining high-quality RNA from individual cells challenging [11]. snRNA, on the other hand, can provide more consistent and reliable data since nuclei are easier to isolate and RNA is more stable inside of nuclei than in whole cells [12]. This is especially useful when working with tissues that cannot be easily dissociated [13,14]. The ability to work with frozen samples has also made snRNA a promising tool for the characterization of tissue atlases [15,16]. A disadvantage of this approach is that using nuclei instead of whole cells can result in the loss of information about cell-cell interactions and other critical cellular processes occurring outside the nucleus [17]. To better understand the differences between the transcriptomic signatures captured by the two assays, we performed both assays on matched PBMC samples originating from the same 18 patients. On analyzing the data obtained from both assays, we compare and characterize the power of each method in identifying immune cell subpopulations and their corresponding gene markers.

2. Materials and Methods

2.1. Single nuclei isolation

Unless alternatively stated, the composition of all the buffers used for all the protocols described was as stated in the cited 10x genomics demonstrated protocols. A schematic of the nuclei isolation protocols used is summarized in Figures 1 and S1.

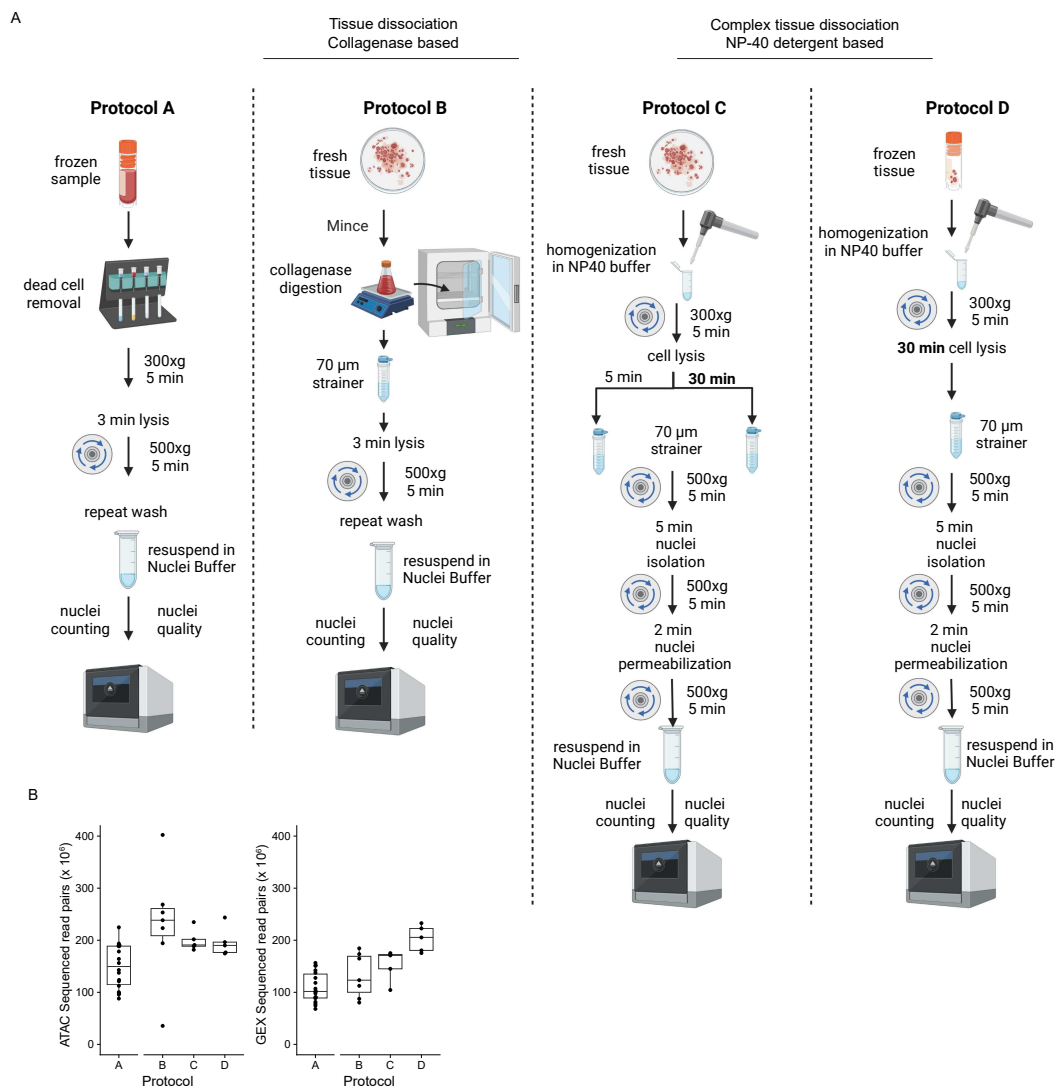


Figure 1. Overview of protocols and sequencing depth statistics. **(A)** Schematic of the main steps pertaining to the four single nuclei isolation protocols discussed. **(B)** Box plot showing the number of sequenced read pairs in ATAC and GEX, grouped by protocol. The middle line represents the median, the lower and upper edges of the rectangle represent the first and third quartiles, and the lower and upper whiskers represent the interquartile range (IQR) $\times 1.5$. Outliers beyond the end of the whiskers are plotted individually. Protocol A, $n = 18$; Protocol B, $n = 7$, Protocol C, $n = 5$; Protocol D, $n = 5$.

2.2. Protocol A

Frozen PBMC collected for the investigation of inflammatory cell profiles in patients diagnosed with symptomatic COVID-19 ($n=6$), clonal hematopoiesis of indeterminate potential (CHIP) ($n=6$), and with both CHIP and COVID-19 ($n=6$) were thawed in a 37°C water bath for 3–5 minutes until no ice was visible. Cells were washed twice with 1 milliliter (mL) phosphate-buffered saline (PBS) + 0.04% bovine serum albumin (BSA) (Sigma) and pelleted at 300 g for 5 min at 4°C . Dead cells were removed according to the 10x Genomics Protocol [18]. Briefly, using magnetic activated cell sorting (MACS) Dead Cell Removal Kit (Miltenyi Biotec), the pellet was resuspended in 100 mL Dead Cell Removal MicroBeads and incubated for 15 minutes at room temperature. After incubation, the cell suspension was diluted with 1X Binding Buffer and applied to an MS column. The dead cells were retained in the column, and the live cells passed through the column and were collected. After dead cell removal, the samples were washed twice with 1 mL PBS + 0.04% BSA. The cell concentration and viability were determined using acridine orange propidium iodide (AOPI) Staining solution (VitaStain) and

a Cellometer K2 cell counter (Nexcelom Biosciences). The nuclei isolation was performed following the 10x genomics suggested protocol [19]: About 1×10^6 cells were pelleted in a 2 mL microcentrifuge tube at 300 g for 5 min at 4°C and resuspended in 100 μ L chilled Lysis Buffer by pipetting 10 times. The cells were then incubated on ice for 3 min and, after the addition of 1 mL of wash buffer, they were centrifuged at 500 g for 5 mins at 4°C. The wash step was repeated one more time for a total of 2 washes. The pellet was resuspended in a chilled diluted Nuclei Buffer, and the nuclei concentration was assessed by propidium iodide (PI) staining (VitaStain) using a Cellometer K2 cell counter. Between 1000 to 5000 nuclei were targeted for capture and used for single nuclei snATAC + snRNA as described below. The quality of nuclei preparation was assessed visually (Figure 2) after observation with an Eclipse 50i microscope (Nikon) after staining with 0.4% trypan blue (Gibco).

2.3. Protocol B

Fresh tumor tissue was obtained from ovarian cancer patients after debulking surgery. Tissue was transferred in a sterile 10 centimeters (cm) culture dish and finely minced using scissors in a KREBS-ringer bicarbonate (KRB) buffer (20 mM sodium chloride, 5 mM potassium chloride, 25 mM sodium bicarbonate, 20 mM HEPES, and 5,5 mM D-glucose). The sample was centrifuged at 500 g for 5 minutes at 4°C, and the supernatant was gently removed without disrupting the pellet. Washes were repeated until the supernatant appeared clear. The final pellet was resuspended in 5 mL of KRB buffer with the addition of 2.5 mg/mL of Collagenase IV (Sigma) and Antibiotic/ Antimycotic (Gibco), and the tissue was digested in an orbital shaker (85-90 rpm) at 37°C for 1 hour. Following incubation, the dissociated solution was mixed by pipetting up and down 5 times, then strained using 100 μ m cell strainer to eliminate bigger aggregates. The flowthrough was centrifuged at 500 g for 5 mins at 4°C. The pellet was further digested with 0.05% trypsin-EDTA (Gibco) at 37°C for 5 min. After neutralizing the reaction with 4mL complete medium (DMEM [Gibco], 10% fetal bovine serum (FBS) [Sigma], 1% Antibiotic/ Antimycotic), the cells were centrifuged at 500 g for 5 mins at 4°C, and the pellet was resuspended in 1 mL of PBS + 0.04% BSA. The cell concentration and viability were determined using AOPI Staining solution and a Cellometer K2 cell counter. About 1×10^6 cells were used for nuclei isolation following the Protocol A procedures indicated above.

2.4. Protocol C

Fresh tumor tissue was obtained from ovarian cancer patients after debulking surgery. The nuclei isolation in protocol C was performed based on the 10x genomics suggested protocol for Complex Tissues for Single Cell snATAC + snRNA [20]. After mincing, a small piece of tissue the size of rice grain was transferred to a 2 mL microcentrifuge tube containing 300 μ L of NP-40 Lysis Buffer and homogenized on ice using a pellet pestle. After adding 1 mL of NP-40 Lysis Buffer, the samples were incubated on ice for either 5 or 30 minutes. After incubation, the samples were strained using a 70 μ m cell strainer, and the flowthrough was then centrifuged 500 g for 5 mins at 4°C. Nuclei isolation was achieved by resuspending the pellet in 1 mL PBS + 1% BSA + 1U/ μ L RNase Inhibitor (Roche) and incubation on ice for 5 minutes. After centrifugation at 500 g for 5 minutes at 4°C, the nuclei were permeabilized by resuspending in 100 μ L of 0.1X Lysis Buffer and incubating on ice for 2 minutes. After adding 1 mL of Wash Buffer, the nuclei were centrifuged at 500 g for 5 mins at 4°C, and the pellet was resuspended in Diluted nuclei buffer. The nuclei concentration was assessed by PI staining using a Cellometer K2 cell counter. As described below, five thousand nuclei were targeted for capture and used for single nuclei snATAC + snRNA.

2.5. Protocol D

Fresh tumor tissue obtained from ovarian cancer patients after debulking surgery was transferred in a sterile 10 cm culture dish and finely minced using scissors in KRB buffer. The sample was centrifuged at 500 g for 5 mins at 4°C, and the supernatant was gently removed without disrupting the pellet. Washes were repeated until the supernatant appeared clear. After the last centrifugation, tissue

was aliquoted in cryovials and frozen in liquid nitrogen. On the day of the experiment, frozen tissue was cut into small pieces without thawing and used for the nuclei isolations as described in protocol C.

2.6. Protocol E

Tumor tissue obtained from ovarian cancer patients after debulking surgery was frozen in liquid nitrogen as described in protocol D. The day of the experiment, frozen tissue was cut into small pieces without thawing and transferred to a 2 mL microcentrifuge tube containing 300 μ L of NP-40 Lysis Buffer and homogenized on ice by using a pellet pestle. After adding 1 mL of NP-40 Lysis Buffer, the samples were incubated on ice for 10 minutes. After incubation, the samples were strained using a 70 μ m cell strainer, and the flowthrough was then centrifuged at 500 g for 5 minutes at 4°C. Nuclei isolation was achieved by resuspending the pellet in 1 mL PBS + 1% BSA + 1U/ μ L RNase Inhibitor (Roche) and incubation on ice for 5 minutes. After centrifugation at 500 g for 5 minutes at 4°C the nuclei were resuspended in 1 mL of PBS + 1% BSA + 1U/ μ L RNase Inhibitor supplemented with 10 of 7-aminoactinomycin D (7-AAD) ready-made solution (Sigma) and incubated on ice for 5 minutes. After dissociation, single-nuclei suspensions were isolated into a 5 mL tube via Fluorescence Activated Cell Sorting (FACS) using FACSMelody (BD bioscience). Single nuclei were selected with forward scatter, and debris was excluded by size. A negative sample was loaded to determine the 7-AAD positive channel. Dye concentration and sorting buffer composition were as recommended by 10x genomics, protocol C. A total of 1,152,000 events were processed, and 801,344 positive nuclei were retrieved, with a flow rate of 5, sorting for purity. The viable FACS sorted nuclei were then centrifuged at 500 g for 5 minutes at 4°C and permeabilized by resuspending in 100 μ L of 0.1X Lysis Buffer and incubating on ice for 2 minutes. After adding 1 mL of Wash Buffer, the nuclei were centrifuged at 500 g for 5 minutes at 4°C, and the pellet was resuspended in diluted Nuclei Buffer. The nuclei concentration was assessed by PI staining using a Cellometer K2 cell counter. 5000 nuclei were targeted for capture and used for single nuclei snATAC + snRNA as described below.

2.7. Single nuclei snATAC + snRNA-seq

Between 1000 and 5000 nuclei per sample were subjected to transposase assays (exposing buffered nuclei to Tn5 transposase) before proceeding to single-cell partitioning into gel beads in emulsion, barcoding, and pre-amplification. ATAC library construction and cDNA, followed by GEX library construction, were done following established 10x Genomics protocols [21]. The libraries concentration was measured using Qubit High Sensitivity assays (Thermo Fisher Scientific), and library profiles were assessed in a fragment analyzer (Advance Analytical) before sequencing.

2.8. Single cell RNA-seq

The cells were first counted and measured for viability using either the Vi-Cell XR Cell Viability Analyzer (Beckman-Coulter, Brea, CA) or a basic hemocytometer and light microscope. The cDNA master mix was prepared according to the manufacturer's instructions using a minimum concentration of 400,000 cells per milliliter per sample. Library constructions were done following established 10x Genomics protocols for Chromium Next GEM Single Cell 3' Kit [22]. All cDNA pools and resulting libraries were measured using Qubit High Sensitivity assays and Agilent Bioanalyzer High Sensitivity chips (Agilent).

2.9. Sequencing

The snATAC and snRNA libraries and scRNA libraries were sequenced on an HiSeq 4000 or NextSeq 2000 instrument (Illumina) before demultiplexing and alignment to the reference genome (GRCh38/hg38).

2.10. Data analysis

2.10.1. Single cell snATAC + snRNA

Sequenced reads from the gene expression (GEX) and DNA accessibility (ATAC) droplet libraries of the snATAC + snRNA assay were processed using 10x Genomics Cell Ranger ARC v2.0.0. The reads were aligned to the pre-built human reference genome GRCh38 - v2020-A-2.0.0 (May 3, 2021) provided by 10x Genomics. Read trimming, alignment, duplicate marking (ATAC), UMI counting (GEX), peak calling (ATAC), and joint cell calling were performed by Cell Ranger. Quality control statistics calculated by Cell Ranger were compiled, analyzed, and plotted using R [23] and ggplot [24]. Downstream processing, including the calculation of fragment length distribution and transcription start site (TSS) enrichment scores, was done using Signac v1.4.0 [25].

To compare single cell and single nucleus RNA-seq in the PBMC data, GEX counts matrices from all samples were integrated (batch-corrected) using Seurat v4.0.4 [26] integration method. Sample GEX count matrix from each sample was log-normalized, scaled to mean 0 and variance 1, and dimensionality reduced using PCA on the top 2000 variable genes across all samples. Default options (CCA and pairwise anchor-finding) were used in the integration anchor-finding step. Uniform manifold approximation and projection uniform manifold approximation and projections (UMAP) were calculated using principal components 1 to 50. The integrated data included 57,817 cells from 18 samples. Cells with more than 50% of reads mapped to mitochondrial genes, those with less than 200 unique genes detected (GEX), those with less than 200 unique peaks detected (ATAC), and those with TSS enrichment score (as calculated by Signac) less than one were removed. After quality control (QC) filtering, the number of cells was reduced to 43,160. Cell type identification was made using two methods for comparison, first by using SingleR v1.10.0 [27] with the immune data from celldex v1.6.0 [28] as a reference, and second by using Azimuth algorithm [29] to map the snATAC + snRNA data to the scRNA data generated from the same cohort of patients, then transferring the labels from the scRNA data to the snATAC + snRNA data. Pseudobulk profiles for the principal component analysis were calculated using muscat [30]. Marker genes were identified using the FindAllMarkers function from Seurat which finds all the differentially expressed genes for each of the cell types using Wilcoxon Rank Sum test ($p < 0.05$).

2.10.2. Single cell RNA-seq

Sequenced reads from the droplet libraries were processed using 10x Genomics Cell Ranger v6.0.1 [31]. The reads were aligned to the pre-built human reference transcriptome GRCh38 - v2020-A (July 7, 2020) provided by 10x Genomics. Read trimming, alignment, UMI counting, and cell calling were performed by Cell Ranger. Doublet prediction on scRNA-seq data was made using Scrublet v0.2.1 [32] with default parameters. Downstream processing was done using Seurat. Count matrices from all samples were integrated using Seurat v4 integration method. The count matrix from each sample was first normalized and dimensionality reduced as described for snATAC + snRNA. The reciprocal PCA and reference-based integration options were applied in the anchor-finding step. Six samples were chosen as references, and PCs 1-50 were used for the reference-based integration. Cells with more than 50% of reads mapped to mitochondrial genes, those with less than 200 unique genes detected, and those that were predicted as doublets by Scrublet were removed, resulting in 70,184 cells. UMAP was made using the top 50 PCs obtained by running PCA on the integrated (batch-corrected) gene expression matrix. Cell type identification was made with SingleR using the immune data from celldex as a reference.

3. Results

3.1. Nuclei isolation for snATAC + snRNA sequencing

PBMC samples (n=18) were processed according to protocol A (Figure 1A). Blood was derived from patients with symptomatic COVID-19, CHIP, or both (Methods). From cell suspensions, nuclei can be extracted without a dissociation step according to protocol A (Figure 1A), facilitating the acquisition of high-quality nuclei (Figure 2A), according to the standard (Figure 2B).

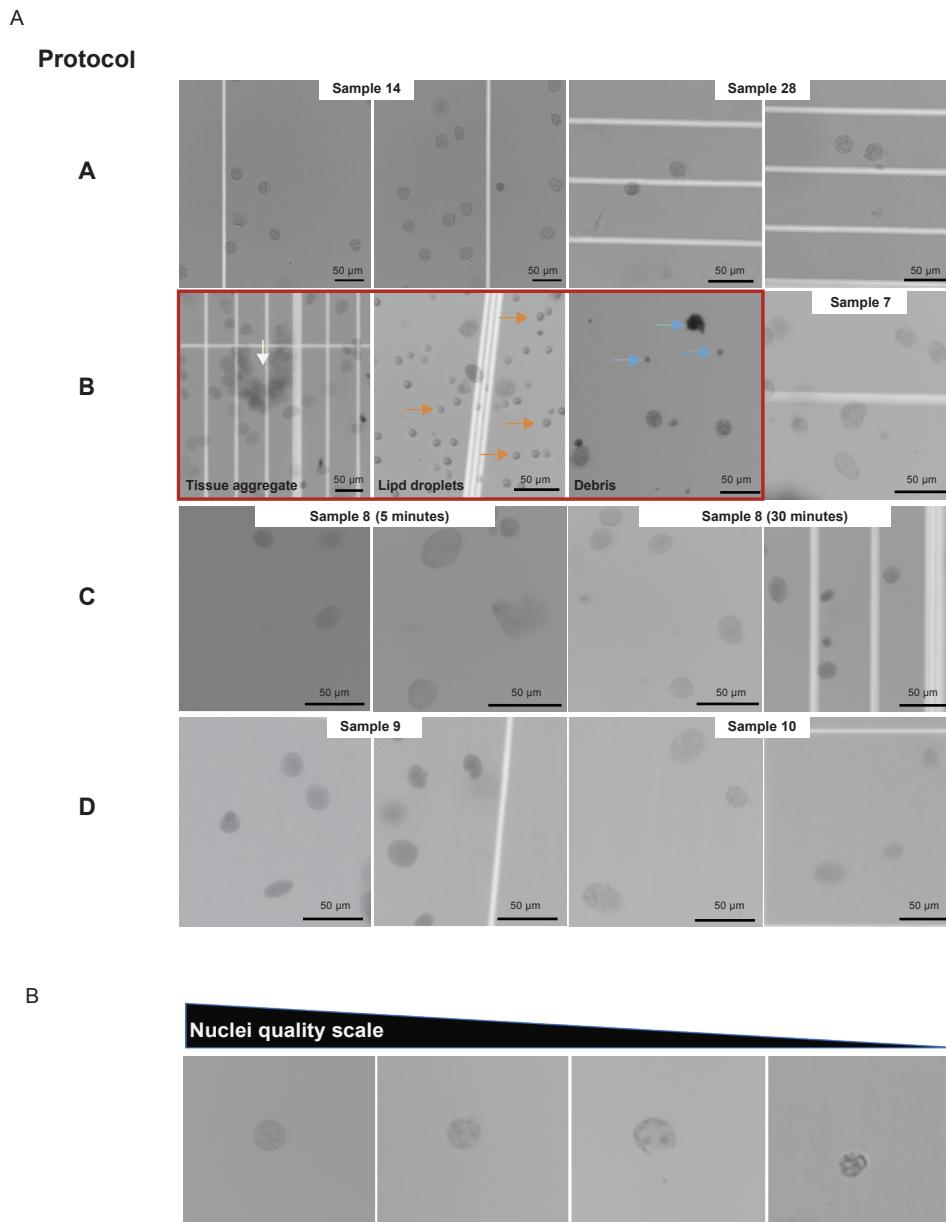


Figure 2. Representative images of single nuclei from the various protocols and samples examined in this study. (A) Protocol A shows nuclei isolated from PBMC (samples 14 and 28). Protocol B shows nuclei prepared from ovarian cancer tissues. Shown in order are examples of eliminated samples (inside red square) with undigested tissue aggregates (white arrow), lipid droplet contaminants (orange arrow), and debris (blue arrow), as well as a successful nuclei preparation (sample 7). Protocol C shows nuclei isolated with a 5 minutes incubation and 30 minutes incubation from sample 8. Protocol D shows nuclei preparation from samples 9 and 10. (B) Representative images of single nuclei organized in decreasing order of preparation quality were used as a standard. Scale bars throughout represent 50 μm .

snATAC + snRNA sequencing was also performed on 18 patients diagnosed with ovarian cancer. The tumor tissue was collected from debulking surgery. For solid tumor samples ensuring adequate tissue dissociation poses several challenges, especially with cancer samples, recognized for their complexity and significant inter-patient variability. A suitable protocol has to promote tissue dissociation, preserving nuclei membrane integrity across all the samples.

Tissue digestion status evaluation during sample preparation was critical. The presence of debris, nuclei aggregates, and membrane damage can negatively impact the assay and the sequencing quality. Over or under-digestion is equally detrimental to the data quality. It is important to obtain complete tissue dissociation of tissue and cell aggregates to avoid clogging the microfluidic system during the capture step while maintaining the integrity of the nuclear membrane. Residual peri-nuclear cytoplasm can also reduce the quality of GEX sequencing results.

Several measures can be taken if nuclei suspension contamination is recognized. Utilizing a fine needle syringe was helpful in dissociating nuclei aggregates; if needed, a strainer was used to remove unwanted debris and cell aggregates. Choosing a protocol tailored for sample characteristics minimizes several of these issues. Lastly, optimization of the duration of cell lysis, especially in the case of tissue dissociation, was also critical (see below).

In 9 OC samples, nuclei isolation for single cells followed collagenase-based tissue digestion (Figure 1A; protocol B). Two samples using protocol B were removed from further processing due to low-quality nuclei (Figure 2A; protocol B). The nuclei isolation protocol for complex tissues C, D (Figure 1), and E (Figure S1) was tested in 11 OC samples.

Protocol performance was compared during sample preparation by nuclei morphology, and its implications on downstream data quality are further discussed. Nuclei isolation following protocol B resulted in one sample clogging the microfluidic system. Under microscopic visualization, this sample had numerous debris (tissue remnants and lipid components), the most likely reason for sample clogging (Figure 2A; protocol B). Another specimen was discarded due to the amount of debris. Also, sample 7 (Figure 2A) exemplifies that some nuclei had intact membranes for protocol B, but frequently membrane blebs were noticed. Fewer aggregates were visualized with complex tissue dissociation (protocols C and D, Figure 2A), and no clogging events were observed after incubation time increment. In addition, protocols C and D showed better preservation of nuclear membranes (Figure 2A).

The incubation step for protocols C and D requires sample-based time optimization. Five and thirty minutes incubation times were tested using the same OC sample (sample 8). Both time periods yielded single nuclei suspension and minimal debris or aggregates (Figure 2A, protocol C). Interestingly, the shorter incubation time (5 minutes) failed the barcoding step. This resulted in less sample volume after the capturing step and uneven mixing compared to the 30 minutes incubation.

Targeting 5,000 cells was considered ideal for both samples, PBMC and OC. However, due to the lower viability of the PBMC samples, 4000 was the higher target number of cells. For OC samples, the majority were targeted at 5000 (n=16). The ratio between the number of cells retrieved and the goal, known as capture efficiency, is demonstrated in Figure S2A.

Ovarian cancer samples (n=18, 90%) that met nuclei preparation quality criteria from all protocols and PBMC samples (n=18) were sequenced according to the 10x Genomics Multiome platform (Methods).

3.2. Sequencing Data QC for PBMC (n=18) and fresh OC samples (n=11)

Sequenced reads (Figure 1B) from ATAC and GEX libraries were processed using 10x Genomics Cell Ranger ARC. Five representative QC statistics from the ATAC libraries and six from the gene expression (GEX) libraries, as reported by Cell Ranger ARC, were considered for quantitative sample quality comparison (Figure 3A). Downstream data analysis was performed using Signac, where QC statistics per cell were analyzed, allowing for filtering high-quality cells, if necessary. Here PBMC and OC samples QC statistics are compared according to their nuclei isolation protocol (Figure 3 A-D).

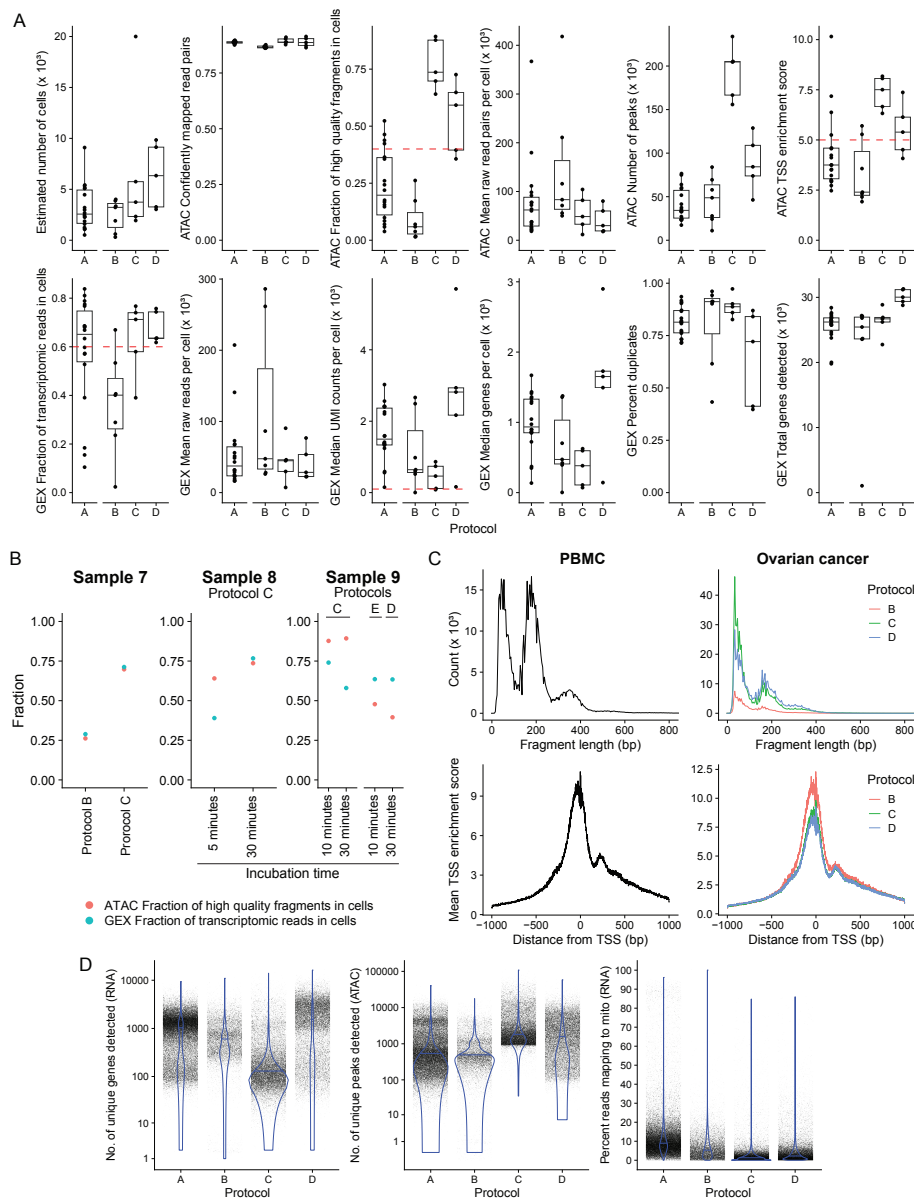


Figure 3. Quality control statistics comparing the protocols examined. **(A)** Estimated number of cells, confidently mapped read pairs (ATAC), a fraction of high-quality fragments in cells (ATAC), mean raw read pairs per cell (ATAC), number of peaks (ATAC), TSS enrichment score (ATAC), a fraction of transcriptomic reads in cells (GEX), mean raw reads per cell (GEX), median UMI counts per cell (GEX), median genes per cell (GEX), percent duplicates (GEX) and total genes detected (GEX) per sample grouped by protocol, shown as a boxplot where the middle line represents the median, the lower and upper edges of the rectangle represent the first and third quartiles and the lower and upper whiskers represent the interquartile range (IQR) $\times 1.5$. Outliers beyond the end of the whiskers are plotted individually. Protocol A, $n = 18$; Protocol B, $n = 7$, Protocol C, $n = 5$; Protocol D, $n = 5$. **(B)** The fraction of high-quality fragments in cells (ATAC) and a fraction of transcriptomic reads in cells (GEX) are shown for samples 7, 8, and 9, for which paired protocol and different incubation time comparisons were made. **(C)** Fragment length distribution and mean TSS enrichment score - the ratio of fragments centered at the TSS to fragments in TSS flanking regions from all samples grouped by tissue and protocol. **(D)** Violin plots showing the number of unique genes detected (RNA), number of unique peaks detected (ATAC), and the number of reads mapping to mitochondrial genes (RNA) calculated per cell in all samples and grouped by protocol. Protocol A, $n = 57817$; Protocol B, $n = 17277$; Protocol C, $n = 33759$; Protocol D, $n = 31623$.

Globally, the PBMC samples performed below the cut-off for ATAC fraction of high-quality reads and ATAC TSS enrichment score (Figure 3A). Cells under inflammatory stress are expected to produce poor-quality results, especially under stressors such as viral infection and CHIP (typically caused by clones carrying mutations in epigenetic regulators like *TET2*, *DNMT3A* and *ASXL1* [33]), contributing to a lower fraction of high-quality reads. In this situation, cell-level quality control analysis and filtering (Figure 3D) are crucial to ensure that only high-quality cells are used to make biological inferences. Despite that, the mean TSS fragment length distribution was adequate (Figure 3C).

Regarding gene expression, results from samples processed using protocol A were satisfactory. Overall, 55 percent of the samples (n=10) archived higher than 60% GEX fraction of transcriptomic read. Also, 78% of the samples (n=14) had higher than 1,000 GEX median unique molecules identified (UMI) counts per cell (Figure 3).

Significant heterogeneity among samples on ATAC and GEX components was observed for ovarian cancer samples (Figure 3A), possibly due to the high complexity of cancer tissue components. ATAC fraction of high-quality reads, ATAC TSS enrichment score, GEX median unique molecules identified (UMI) counts per cell, and GEX fraction of transcriptomic reads parameters can guide analysis decision-making since they have a recommended threshold regardless of the specimen.

ATAC fraction of high-quality fragments represents the percentage of high-quality fragments with a valid barcode associated with cell-containing partitions [Technical Note – Cell Ranger ARC Web Summary Files for Single Cell snATAC + snRNA ATAC + Gene Expression Assay • Rev A]. When this fraction is less than 40%, it indicates that a significant fraction of ATAC fragments were not associated with a barcoded cell. This can be explained by high ambient ATAC contamination, increased cell accessibility, or poor nuclei preparation. All samples processed according to protocol B (n=7) did not meet the minimum threshold (Figure 3A). Remarkably, all samples (n=10) with the nuclei extracted following protocol C had a higher than 40% threshold for ATAC fraction of high-quality fragments (Figure 3A; Protocol C mean = 76.9%; Protocol D mean = 54.4%).

The median UMI count represents gene expression elements (RNA molecules converted into cDNA) per cell, thus correlating with sequencing depth. Given that all the samples had an adequate number of reads (Figure 1B), read depth was further evaluated by the number of unique genes detected per cell, equally consistent between protocols (Figure 3D).

ATAC fragment size distribution also correlates with sample quality. To characterize DNA accessibility, samples must have fragments of the inter-nucleosome region around them. This distribution is only achieved with protocol C for fresh ovarian cancer samples (Figure 3C), suggesting that protocol B is not ideal for fresh cancer samples.

GEX fraction of transcriptomic reads translates the RNA percentage aligned with the genome and assigned to a barcoded cell [Technical Note – Cell Ranger ARC Web Summary Files for Single Cell snATAC + snRNA ATAC + Gene Expression Assay • Rev A]. Lower values can be due to environmental RNA contaminating the single-nuclei droplet. In accordance, the protocol C sample outperformed protocol B preparation. Considering the apparent superiority trend, a paired comparison was conducted to control for sample specificities.

The complex tissue dissociation protocol was better than protocol B for OC samples in the paired comparison (Figure 3B, sample 7). Protocol C had an ATAC fraction of high-quality reads of 70% compared to 26% obtained by the B protocol. Also, the complex tissue dissociation protocols' GEX fraction of transcriptomic reads was 2.4 times higher (Figure 3B). As was foreseeable, the quality differences impacted cell identification and clustering (Figure S3A, sample 7). Taken together, optimal nuclei preparation has been proven crucial to assay performance.

To further optimize protocol C, two incubation times were tested. Sample 8 (Figure 3B) was dissociated for 5 and 30 minutes. As mentioned, the shorter incubation period for protocol C failed the barcode step, resulting in empty droplets that were wrongly counted as cells (20000 against 5752 for the 30 minutes incubation), culminating in a poor GEX fraction of transcriptomic reads per cell and a higher percentage of filtered-out cells (Figure S2B). These technical issues culminated with inadequate

clustering (ATAC and GEX) for sample 8 when dissociated for only five minutes (Figure S3B), showing the same pattern observed in compromised sample examples in 10x genomics technical notes [34].

An intermediary incubation time point was tested to assess for possible tissue over-digestion within 30 minutes of incubation. Fresh preparation of sample 9 was incubated for 10 and 30 minutes. Both preparations yielded good QC values, yet the longer incubation had a lower GEX fraction of transcriptomic reads (Figure 3B). As previously mentioned, tissue over-digestion can impact GEX reads, even if membrane changes are not detected (Figure 2B, protocol C); this probably happened in the 30-minute incubation time. Also, the longer incubation had fewer cells retrieved (1,937, compared to 3,740 for 10 minutes). However, both targeted five thousand cells and had similar sample concentrations (4850 nuclei/ μL for 10 minutes, against 3520 nuclei/ μL), decreasing capture efficiency for the 30 minutes incubation (Figure S2A). Similar clustering was obtained for ATAC and GEX, demonstrating that, regardless of the absolute QC values, both incubation time points were feasible for the OC preparation (Figure S3C).

3.3. Frozen OC sample optimization

After optimizing nuclei extraction protocols for fresh tissue, the same process was performed for frozen tissue. The advantages of frozen tissue include allowing retrospective sample studies and curated specimen selection. Sample 9, previously discussed in the fresh sample preparation, was also processed after being frozen for paired comparisons.

The OC frozen preparation of sample 9 was sequenced before and after FACS sorting (protocols D and E, respectively) (Figures 1A and S1A). The rationale for sorting was to minimize the need for longer incubation and debris removal. This condition was tested only in frozen tissue since it has lower viability and higher nuclear remnants. FACS nuclei were selected based on 7-AAD signal (Figure S1B). A head-to-head comparison of nuclei, QC analysis, and clustering was performed.

Nuclei morphology was satisfactory throughout. Albeit the nuclei suspension for the FACS samples had no debris (Figure S1C), both samples performed similarly in the barcoding step. Furthermore, all QC values met the standard threshold and were consistent within samples (Figure 3B). Clustering for GEX and ATAC was also similar between both frozen preparations (Figure S3C).

Considering all the samples processed according to protocol C (n=5) and protocol D/E (n=6), DNA accessibility and RNA expression are consistent throughout (Figure 3A). Interestingly, frozen samples detected more unique genes per cell, ensuring greater sequencing depth and cell population representation (Figure 3D). These results suggest that frozen tissue recapitulates the quality and biology of fresh tissue. In that manner, frozen tissue preparation became the standard of our laboratory for OC samples.

3.4. Comparison of the transcriptomes profiled from snATAC + snRNA and scRNA-seq from a matched cohort

By performing both scRNA-seq and snATAC + snRNA assays on matched samples originating from the same individuals, we compared nuclear transcriptomic signatures against whole cell transcriptomes and their ability to capture immune cell subpopulations. scRNA profiles of 70,184 cells and snRNA profiles of 43,160 nuclei from matched PBMC samples from 18 patients were pooled. Dimensionality reduction using principal component analysis (PCA) and uniform manifold approximation and projection (UMAP), followed by cell type identification using SingleR, allowed us to identify the typical immune cell populations of B-cells, natural killer (NK) cells, CD4+ T-cells, CD8+ T-cells, CD14 monocytes, CD16 monocytes, and an intermediate monocytic population among others (Figure 4A). The cell type identification method used here is a reference-based method wherein the transcriptomic profiles of individual cells are mapped to reference profiles of known cell types which in this case come from the Monaco immune data - bulk RNA-seq samples of sorted immune cell populations from GSE107011 [28]. To evaluate the effect of the reference used for cell type identification, we compared two approaches for labeling cell types in the snRNA data - first using the Monaco reference and secondly by mapping the snRNA data onto the scRNA data and then

transferring labels using Azimuth. Since the snRNA and scRNA data come from matched samples, the latter approach should be able to make more biologically meaningful mapping of cell states. Both snRNA and scRNA captured a very similar distribution of cell types, with subtle differences between the two cell type identification methods (Figures 4A, B and S4A). We observed a decreased proportion of NK cells in snRNA when using scRNA as a reference compared to the Monaco reference, making the data more in line with scRNA proportions. However, when comparing the distribution of CD4+ T-cells and gamma delta T-cells, there was an increased proportion of both types in snRNA (using scRNA reference) that differed from both scRNA and snRNA (using Monaco reference). scRNA identified a larger proportion of CD8+ T-cells than snRNA in both approaches.

To better distinguish the two platforms in identifying cell type-specific transcriptomic signatures, we performed PCA on pseudo-bulk profiles obtained by aggregating the counts from each cell type from each platform. The PCA revealed that the first two principal components (PC) cluster the profiles by platform, while the third and the fourth PCs cluster transcriptomic signatures by major cell subpopulations (myeloid and lymphoid lineages) (Figure 4C). This suggests that the batch effect between scRNA and snRNA is captured by the first two PCs, so removing these PCs should facilitate pooled data analysis from both platforms. This effect is consistent regardless of the reference for snRNA cell type identification.

In figure 4D, the number of cells labeled differently by the two methods of cell type identification were compared. Overall there is a high concordance between the two methods. We observe some cross-labeling of cell types within the lymphoid and the myeloid lineages suggesting that the two methods perform similarly when identifying broad transcriptomic signatures. But they differ in the resolution of subtle signals between cell subtypes. We also compared the number of overlaps between the markers of each cell type - genes that are significantly differentially expressed (p -value < 0.05) between cells of each cell type and all other cells across the two platforms (Figure S4B) with snRNA cell types identified using Monaco reference on the left and using scRNA reference on the right. High concordance in marker genes within the lymphoid and the myeloid lineages cell type was observed, reinforcing the previous inference that both cell type identification methods and platforms perform equally in resolving broad cell lineages. At the same time, there are subtle differences in capturing subpopulations.

On analyzing the correlation of gene expression, averaged by cell type and sample, between scRNA and snRNA, the genes were both positively correlated and negatively correlated (Spearman's correlation test $p < 0.05$) (Figures 4E and S4C). But positively correlated genes were, on average, expressed higher than the negatively correlated genes (Figure 4F). This could mean that genes with a very low detection threshold are recognized with different sensitivity by the two platforms. All the marker genes were positively correlated (Figure 4G), implying that the cell type-specific genes are detected similarly by both scRNA and snRNA.

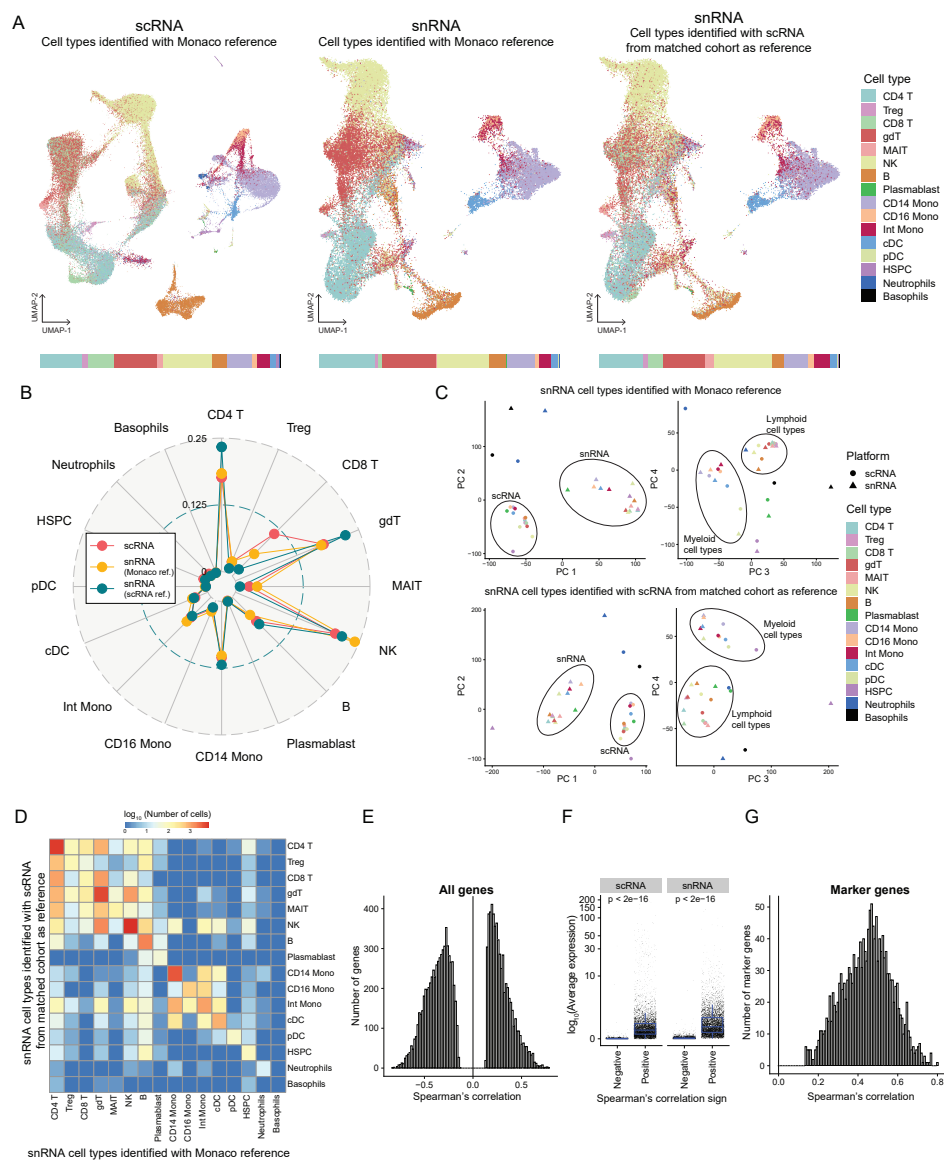


Figure 4. Comparison of the transcriptomes profiled from snATAC + snRNA (GEX) and scRNA-seq from a matched cohort. **(A)** UMAP projection showing 70184 cells from scRNA colored by cell types identified using SingleR with Monaco reference (left) and 43160 cells from snRNA, cell types identified using Monaco reference (middle) or using scRNA from the matched cohort as reference (right). The bar at the bottom shows the proportion of each cell type identified using each method. Treg, regulatory T-cells; gdT, gamma delta T-cells; MAIT, Mucosal-associated invariant T cells; NK, natural killer cells; Mono, monocytes; Int, intermediate; cDC, classical dendritic cells; pDC, plasmacytoid dendritic cells; HSPC, hematopoietic stem, and progenitor cells. **(B)** Radar plot showing the comparison of proportions of each cell type identified by each method. **(C)** Principal component analysis of the transcript counts aggregated by platform and cell type with snRNA cell types identified using Monaco reference (top) and scRNA from matched cohorts as reference (bottom). **(D)** Heatmap showing the log (base 10) of the number of cells that overlap cell type identification by the two references - Monaco and scRNA from matched cohort. **(E)** Histogram showing the distribution of genes significantly correlated (Spearman correlation, test for association $p < 0.05$) in expression (averaged over both sample and cell type) between scRNA and snRNA. **(F)** Log (base 10) average expression (averaged over all cells) of genes separated by positive and negative Spearman correlation coefficients. **(G)** Histogram showing the distribution of marker genes significantly correlated (Spearman correlation, test for association $p < 0.05$) in average expression between scRNA and snRNA.

4. Discussion

Performing ATAC and RNA sequencing from the same nuclei increases assay definition and ability to characterize cell states, including the activity of key transcription factors. As described here, it is critical to ensure good nuclei preparations, and different specimens will need specific dissociation steps. For example, nuclei could be directly extracted from PBMC (protocol A), but human tissue samples required a well-defined dissociation process.

Two different tissue digestion methods were tested in OC samples, one based on collagenase and the other using NP-40 as its major detergent (Methods). When performing the assay, the only step in evaluating sample quality is through visual evaluation of nuclei morphology. According to the standard (Figure 2B), both methods could yield good-quality nuclei when analyzed under the microscope. Although the microfluidic system capture step failed with both protocols (Figure 2) protocol B and sample 8 protocol C with 5 minutes incubation, it was more frequent in protocol B, resulting in two samples lost. Furthermore, the collagenase-based method (protocol B) had worse QC sequencing values than the NP-40-based method (protocols C, D, and E).

Poor performance in QC correlates with the inability to identify cells, hence poor sub-population characterization (Figure S3). This occurs mainly due to the extensive disposal of reads for improper barcoded nuclei or DNA and RNA fragments arising from contamination. Consequently, fewer cells achieve quality thresholds and can be used for downstream analysis (Figure S2B). Ensuring that the number of cells obtained is close to the target value (capture efficiency, Figure S2A) is also important for the optimal usage of sequenced reads. This refined protocol optimization is particularly important when dealing with scarce patient samples.

Beyond choosing the appropriate dissociation method, accessing optimal incubation time is needed. Sample 8 exemplifies that for the same specimen, the longer incubation time was decisive for the success of the experiment. Five minutes of incubation was not enough to dissociate the tissue (Figure 2A). Consequently, capturing a single-cell suspension and barcoding was not ideal, jeopardizing all the following steps.

The methods proved to be very different when post-sequencing quality control statistics for both ATAC and GEX were compared. All samples from protocol B had their high-quality ATAC fragments lower than recommended threshold. The open chromatin region was adequately assessed when OC samples were processed using the complex tissue dissociation technique (with fresh and frozen tissue, protocols C, D, and E). The previous protocols also exceeded method B in the GEX parameters.

Notwithstanding the importance of Cell Ranger ARC parameters, careful interpretation is needed since sample biology influences the results. For example, the PBMC cohort had inadequate high-quality ATAC fragments per cell, most likely due to the nature of the samples rather than poor sample preparation, emphasizing context-specific interpretation of these scores. In addition, the raw QC statistics do not translate directly to sequencing quality; multifactorial analysis is needed.

The single-cell definition can be used to favor sample quality assessment. All the mentioned criteria can be applied to each individual cell (Figure S3). This separation allows QC filtering by cell, ensuring that only high-quality data is used for downstream analysis.

The snATAC + snRNA platform is at its full potential when used to characterize different cell populations that compose a particular tissue, embryonic development, disease process, or even inter-patient variability. Ensuring adequate preparation for frozen tissue increases assay versatility. Sample 9 paired comparison assured results reproducibility with frozen preparations (Figure S3C). In addition, the high-quality QC values for protocols D and E endorse the feasibility of handling frozen cancer tissues.

Furthermore, the implication of FACS sorting for frozen samples was investigated. Sample 9 was sequenced before and after FACS. There were no significant differences in sequencing QC results (Figure 3), making FACS sorting an optional step. Nevertheless, the goal of FACS sorting in this paper was to increase sample purity. Yet, FACS is a powerful tool to separate different populations based on cell surface markers or ploidy [35].

Lastly, since the snATAC + snRNA platform quantifies gene expression only from within the nucleus, a comparative analysis of data obtained from scRNA and snRNA measurements from matched samples was done to help inform better decisions in study design. Although similar in their output, scRNA, and snRNA approaches vary in identifying different cell types and their respective proportions. We show that adjusting the reference dataset used for the cell type identification, such as using matched scRNA data as a reference for snRNA cell type identification, may further align the two modalities in a biologically meaningful sense. The choice between scRNA and snRNA sequencing should be driven by the biological question, sample composition, and preservation as fresh or frozen tissue.

Each high throughput sequencing strategy has advantages and limitations. scRNA-seq measures the expression profile of the whole cell and preserves an intact membrane allowing for further downstream applications such as CITE-Seq, which can profile cell surface protein interactions [36]. On the other hand, snRNA-seq is an especially beneficial approach when working with frozen samples, such as those from tissue biobanks and tissue samples that are difficult to dissociate. In a direct comparison of snRNA-seq and scRNA-seq of the adult kidney, Wu et al. showed that snRNA-seq achieves comparable gene detection and spatial information while eliminating dissociation-induced changes in gene expression [37]. Andrews et al. [38] showed that snRNA-seq could identify rare mesenchymal and precursor cell types in the healthy human liver but that certain subtypes of immune cells were only distinguishable in scRNA-seq. While the preparation process of working with scRNA-seq data is a significant source of variability [39], nuclei provide greater RNA stability and have a reproducible expression of transcripts when compared to whole cells [40].

In addition, using nuclei instead of whole cells can result in loss of information about important cellular processes that occur outside the nucleus. While there are differences between the transcripts obtained from the two platforms, our analysis suggests that both platforms equally capture the underlying cell type signatures, and it is possible to pool information from both platforms removing the batch effect by removing the principal components that correlate with the sequencing platforms. Moreover, we observe that highly expressed genes and marker genes are positively correlated between the two platforms. In contrast, genes that are at low detection levels are negatively correlated, suggesting that the two platforms vary in detecting specific low-expression genes. Careful consideration is necessary when choosing the platforms when the genes of interest are lowly expressed. In summary, both scRNA-seq and snRNA-seq are powerful tools for analyzing gene expression, and the choice between them depends on the specific research question and the type of cells being analyzed.

In summary, we analyzed how different nuclei preparation can impact the output of snATAC + snRNA sequencing for PBMC and OC samples. To compare methodologies, we utilized several parameters, such as nuclei morphology, sequencing depth, QC values per sample and for each cell, capture efficiency, and dimensionality reduction techniques, such as UMAP. To control for sample specificity, several paired comparisons between protocols were performed. Also, the PBMC cohort had both snRNA and scRNA sequencing results compared for differences in cell identification. Our results show how careful selection of an appropriate nuclei isolation protocol is needed for good sequencing results and the importance of context-based analysis of QC results. Finally, snRNA and scRNA results were consistent, especially considering cell specific markers and abundant transcripts.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on [Preprints.org](https://www.preprints.org)

Author Contributions: A.G.M. conceived and designed the study with the help of W.M.I. and L.A.M.S.; methodology and experimentation were performed by L.A.M.S. and A. Mazzone, A. Munankarmy, J.F., T.L., M.D., V.S., K.H.K.; software and data analysis, W.M.I., M.B.; writing L.A.M.S., W.M.I., M.D., A. Mazzone, and A.G.M.; supervision, J.H.L., M.P., J.W. and N.C.; funding acquisition, A.G.M., M.P. and N.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Mayo Clinic Center for Individualized Medicine to A.G.M, M.P. and N.C., and by the Center for Biomedical Discovery and the Mayo Clinic National Cancer Institute-designated Comprehensive Cancer Center Ovarian SPORE grant (Career Development Award P50 CA136393) to A.G.M.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Mayo Clinic protocol code 20-005400 for the PBMC samples approved on June 2020 and protocol code 09-008768 for the ovarian cancer samples approved on January 2010.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data used in this article is available on request.

Acknowledgments: The authors thank Fariborz Rakhshanrohakhtar, Jon Harrington, Julie Lau, Eric Weiben and the Epigenomics Development Lab staff (Mayo Clinic Rochester) for their support with the sample collection, processing and sequencing. Raj Kannan, Musheer Aalam, Megan Ritting (Mayo Clinic Rochester) and Michael Barrett (Mayo Clinic Arizona) for technical help with FACS. Illustration in Fig. 1A was created with BioRender.com

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ATAC	Assay for transposase-accessible chromatin
CHIP	Clonal hematopoiesis of indeterminate potential
CITE-seq	Cellular indexing of transcriptomes and epitopes sequencing
COVID-19	Coronavirus disease 2019
DNA	Deoxyribonucleic acid
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
GEX	Gene expression
IQR	Interquartile range
KRB	KREBS-ringer bicarbonate
log	Logarithmic
MACS	Magnetic activated cell sorting
mg	Milligram
ml	Milliliter
mM	Millimolar
NK	Natural killer
OC	Ovarian cancer
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PDX	Patient-derived xenograft
PI	Propidium iodide
QC	Quality Control
RNA	Ribonucleic acid
sc	Single cell
sn	Single nucleus
TF	Transcription factor
TSS	Transcription start sites
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecules identified
μL	Microliter
μm	Micrometer

References

1. Filbin, M.G.; Tirosh, I.; Hovestadt, V.; Shaw, M.L.; Escalante, L.E.; Mathewson, N.D.; Neftel, C.; Frank, N.; Pelton, K.; Hebert, C.M.; Haberler, C.; Yizhak, K.; Gojo, J.; Egervari, K.; Mount, C.; van Galen, P.; Bonal, D.M.; Nguyen, Q.D.; Beck, A.; Sinai, C.; Czech, T.; Dorfer, C.; Goumnerova, L.; Lavarino, C.; Carcaboso, A.M.; Mora, J.; Mylvaganam, R.; Luo, C.C.; Peyrl, A.; Popović, M.; Azizi, A.; Batchelor, T.T.; Frosch, M.P.; Martinez-Lage, M.; Kieran, M.W.; Bandopadhyay, P.; Beroukhi, R.; Fritsch, G.; Getz, G.; Rozenblatt-Rosen, O.; Wucherpfennig, K.W.; Louis, D.N.; Monje, M.; Slavc, I.; Ligon, K.L.; Golub, T.R.; Regev, A.; Bernstein, B.E.;

- Suvà, M.L. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **2018**, *360*, 331–335. doi:10.1126/science.aao4750.
2. Cieslik, M.; Chinnaiyan, A.M. Cancer transcriptome profiling at the juncture of clinical translation. *Nature Reviews Genetics* **2018**, *19*, 93–109. doi:10.1038/nrg.2017.96.
 3. Jerby-Arnon, L.; Shah, P.; Cuoco, M.S.; Rodman, C.; Su, M.J.; Melms, J.C.; Leeson, R.; Kanodia, A.; Mei, S.; Lin, J.R.; Wang, S.; Rabasha, B.; Liu, D.; Zhang, G.; Margolais, C.; Ashenberg, O.; Ott, P.A.; Buchbinder, E.I.; Haq, R.; Hodi, F.S.; Boland, G.M.; Sullivan, R.J.; Frederick, D.T.; Miao, B.; Moll, T.; Flaherty, K.T.; Herlyn, M.; Jenkins, R.W.; Thummalapalli, R.; Kowalczyk, M.S.; Cañadas, I.; Schilling, B.; Cartwright, A.N.R.; Luoma, A.M.; Malu, S.; Hwu, P.; Bernatchez, C.; Forget, M.A.; Barbie, D.A.; Shalek, A.K.; Tirosh, I.; Sorger, P.K.; Wucherpfennig, K.; Van Allen, E.M.; Schadendorf, D.; Johnson, B.E.; Rotem, A.; Rozenblatt-Rosen, O.; Garraway, L.A.; Yoon, C.H.; Izar, B.; Regev, A. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* **2018**, *175*, 984–997.e24. doi:10.1016/j.cell.2018.09.006.
 4. Hammelman, J.; Patel, T.; Closser, M.; Wichterle, H.; Gifford, D. Ranking reprogramming factors for cell differentiation. *Nat Methods* **2022**, *19*, 812–822. doi:10.1038/s41592-022-01522-2.
 5. Tsompana, M.; Buck, M.J. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* **2014**, *7*, 33. doi:10.1186/1756-8935-7-33.
 6. Bentsen, M.; Goymann, P.; Schultheis, H.; Klee, K.; Petrova, A.; Wiegandt, R.; Fust, A.; Preussner, J.; Kuenne, C.; Braun, T.; Kim, J.; Looso, M. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nature Communications* **2020**, *11*, 4267. doi:10.1038/s41467-020-18035-1.
 7. Ma, C.Z.; Brent, M.R. Inferring TF activities and activity regulators from gene expression data with constraints from TF perturbation data. *Bioinformatics* **2021**, *37*, 1234–1245. doi:10.1093/bioinformatics/btaa947.
 8. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology* **2015**, *109*, 21.29.1–21.29.9. doi:https://doi.org/10.1002/0471142727.mb2129s109.
 9. Mohammed Ismail, W.; Mazzone, A.; Ghiraldini, F.G.; Kaur, J.; Bains, M.; Munankarmy, A.; Bagwell, M.S.; Safgren, S.L.; Moore-Weiss, J.; Buciu, M.; Shimp, L.; Leach, K.A.; Duarte, L.F.; Nagi, C.S.; Carcamo, S.; Chung, C.Y.; Hasson, D.; Dadgar, N.; Zhong, J.; Lee, J.H.; Couch, F.J.; Revzin, A.; Ordog, T.; Bernstein, E.; Gaspar-Maia, A. MacroH2A histone variants modulate enhancer activity to repress oncogenic programs and cellular reprogramming. *Commun Biol* **2023**, *6*, 215. doi:10.1038/s42003-023-04571-1.
 10. Binder, M.; Carr, R.M.; Lasho, T.L.; Finke, C.M.; Mangaonkar, A.A.; Pin, C.L.; Berger, K.R.; Mazzone, A.; Potluri, S.; Ordog, T.; Robertson, K.D.; Marks, D.L.; Fernandez-Zapico, M.E.; Gaspar-Maia, A.; Patnaik, M.M. Oncogenic gene expression and epigenetic remodeling of cis-regulatory elements in ASXL1-mutant chronic myelomonocytic leukemia. *Nat Commun* **2022**, *13*, 1434. doi:10.1038/s41467-022-29142-6.
 11. Slyper, M.; Porter, C.B.M.; Ashenberg, O.; Waldman, J.; Drokhlyansky, E.; Wakiro, I.; Smillie, C.; Smith-Rosario, G.; Wu, J.; Dionne, D.; Vigneau, S.; Jané-Valbuena, J.; Tickle, T.L.; Napolitano, S.; Su, M.J.; Patel, A.G.; Karlstrom, A.; Gritsch, S.; Nomura, M.; Waghay, A.; Gohil, S.H.; Tsankov, A.M.; Jerby-Arnon, L.; Cohen, O.; Klughammer, J.; Rosen, Y.; Gould, J.; Nguyen, L.; Hofree, M.; Tramontozzi, P.J.; Li, B.; Wu, C.J.; Izar, B.; Haq, R.; Hodi, F.S.; Yoon, C.H.; Hata, A.N.; Baker, S.J.; Suvà, M.L.; Bueno, R.; Stover, E.H.; Clay, M.R.; Dyer, M.A.; Collins, N.B.; Matulonis, U.A.; Wagle, N.; Johnson, B.E.; Rotem, A.; Rozenblatt-Rosen, O.; Regev, A. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **2020**, *26*, 792–802. doi:10.1038/s41591-020-0844-1.
 12. Habib, N.; Li, Y.; Heidenreich, M.; Swiech, L.; Avraham-Davidi, I.; Trombetta, J.J.; Hession, C.; Zhang, F.; Regev, A. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **2016**, *353*, 925–8. doi:10.1126/science.aad7038.
 13. Habib, N.; Avraham-Davidi, I.; Basu, A.; Burks, T.; Shekhar, K.; Hofree, M.; Choudhury, S.R.; Aguet, F.; Gelfand, E.; Ardlie, K.; Weitz, D.A.; Rozenblatt-Rosen, O.; Zhang, F.; Regev, A. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **2017**, *14*, 955–958. doi:10.1038/nmeth.4407.
 14. Nagy, C.; Maitra, M.; Tanti, A.; Suderman, M.; Thérout, J.F.; Davoli, M.A.; Perlman, K.; Yerko, V.; Wang, Y.C.; Tripathy, S.J.; Pavlidis, P.; Mechawar, N.; Ragoussis, J.; Turecki, G. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nature Neuroscience* **2020**, *23*, 771–781. doi:10.1038/s41593-020-0621-y.

15. Sunkin, S.M.; Ng, L.; Lau, C.; Dolbeare, T.; Gilbert, T.L.; Thompson, C.L.; Hawrylycz, M.; Dang, C. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* **2013**, *41*, D996–d1008. doi:10.1093/nar/gks1042.
16. Eraslan, G.; Drokhyansky, E.; Anand, S.; Fiskin, E.; Subramanian, A.; Slyper, M.; Wang, J.; Van Wittenberghe, N.; Rouhana, J.M.; Waldman, J.; Ashenberg, O.; Lek, M.; Dionne, D.; Win, T.S.; Cuoco, M.S.; Kuksenko, O.; Tsankov, A.M.; Branton, P.A.; Marshall, J.L.; Greka, A.; Getz, G.; Segrè, A.V.; Aguet, F.; Rozenblatt-Rosen, O.; Ardlie, K.G.; Regev, A. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science* **2022**, *376*, eabl4290. doi:10.1126/science.abl4290.
17. Ding, J.; Adiconis, X.; Simmons, S.K.; Kowalczyk, M.S.; Hession, C.C.; Marjanovic, N.D.; Hughes, T.K.; Wadsworth, M.H.; Burks, T.; Nguyen, L.T.; Kwon, J.Y.H.; Barak, B.; Ge, W.; Kedaigle, A.J.; Carroll, S.; Li, S.; Hacohen, N.; Rozenblatt-Rosen, O.; Shalek, A.K.; Villani, A.C.; Regev, A.; Levin, J.Z. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nature Biotechnology* **2020**, *38*, 737–746. doi:10.1038/s41587-020-0465-8.
18. 10x Genomics. Removal of Dead Cells from Single Cell Suspensions for Single Cell RNA Sequencing. Document Number CG000093 Rev C, 2022.
19. 10x Genomics. Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing. Document Number CG000365 Rev C, 2022.
20. 10x Genomics. Nuclei Isolation from Complex Tissues for Single Cell Multiome ATAC + Gene Expression Sequencing. Document Number CG000375 Rev C, 2022.
21. 10x Genomics. Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Kits User Guide. Document Number CG000338 Rev F, 2022.
22. 10x Genomics. Chromium Single Cell 3' Reagent Kits User Guide (v3.1 Chemistry). Document Number CG000204 Rev D, 2022.
23. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.
24. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag New York, 2016.
25. Stuart, T.; Srivastava, A.; Madad, S.; Lareau, C.A.; Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **2021**, *18*, 1333–1341. doi:10.1038/s41592-021-01282-5.
26. Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W. M., r.; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; Hoffman, P.; Stoeckius, M.; Papalexi, E.; Mimitou, E.P.; Jain, J.; Srivastava, A.; Stuart, T.; Fleming, L.M.; Yeung, B.; Rogers, A.J.; McElrath, J.M.; Blish, C.A.; Gottardo, R.; Smibert, P.; Satija, R. Integrated analysis of multimodal single-cell data. *Cell* **2021**, *184*, 3573–3587.e29. doi:10.1016/j.cell.2021.04.048.
27. Aran, D.; Looney, A.P.; Liu, L.; Wu, E.; Fong, V.; Hsu, A.; Chak, S.; Naikawadi, R.P.; Wolters, P.J.; Abate, A.R.; Butte, A.J.; Bhattacharya, M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **2019**, *20*, 163–172. doi:10.1038/s41590-018-0276-y.
28. Monaco, G.; Lee, B.; Xu, W.; Mustafah, S.; Hwang, Y.Y.; Carré, C.; Burdin, N.; Visan, L.; Ceccarelli, M.; Poidinger, M.; Zippelius, A.; Pedro de Magalhães, J.; Larbi, A. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep* **2019**, *26*, 1627–1640.e7. doi:10.1016/j.celrep.2019.01.041.
29. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W. M., r.; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e21. doi:10.1016/j.cell.2019.05.031.
30. Crowell, H.L.; Soneson, C.; Germain, P.L.; Calini, D.; Collin, L.; Raposo, C.; Malhotra, D.; Robinson, M.D. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nature Communications* **2020**, *11*, 6077.
31. Zheng, G.X.Y.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; Gregory, M.T.; Shuga, J.; Montesclaros, L.; Underwood, J.G.; Masquelier, D.A.; Nishimura, S.Y.; Schnall-Levin, M.; Wyatt, P.W.; Hindson, C.M.; Bharadwaj, R.; Wong, A.; Ness, K.D.; Beppu, L.W.; Deeg, H.J.; McFarland, C.; Loeb, K.R.; Valente, W.J.; Ericson, N.G.; Stevens, E.A.; Radich, J.P.; Mikkelsen, T.S.; Hindson, B.J.; Bielas, J.H. Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **2017**, *8*, 14049. doi:10.1038/ncomms14049.
32. Wolock, S.L.; Lopez, R.; Klein, A.M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **2019**, *8*, 281–291.e9. doi:10.1016/j.cels.2018.11.005.

33. Jaiswal, S.; Ebert, B.L. Clonal hematopoiesis in human aging and disease. *Science* **2019**, *366*. doi:10.1126/science.aan4673.
34. 10x Genomics. Interpreting Cell Ranger ATAC Web Summary Files for Single Cell ATAC Assay. Document Number CG000202 Rev B, 2022.
35. Lorber, T.; Andor, N.; Dietsche, T.; Perrina, V.; Juskevicius, D.; Pereira, K.; Greer, S.U.; Krause, A.; Müller, D.C.; Savic Prince, S.; Lardinois, D.; Barrett, M.T.; Ruiz, C.; Bubendorf, L. Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach. *J Pathol* **2019**, *247*, 199–213. doi:10.1002/path.5183.
36. Stoeckius, M.; Hafemeister, C.; Stephenson, W.; Houck-Loomis, B.; Chattopadhyay, P.K.; Swerdlow, H.; Satija, R.; Smibert, P. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **2017**, *14*, 865–868. doi:10.1038/nmeth.4380.
37. Wu, H.; Kirita, Y.; Donnelly, E.L.; Humphreys, B.D. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *Journal of the American Society of Nephrology* **2019**, *30*.
38. Andrews, T.S.; Atif, J.; Liu, J.C.; Perciani, C.T.; Ma, X.Z.; Thoeni, C.; Slyper, M.; Eraslan, G.; Segerstolpe, A.; Manuel, J.; Chung, S.; Winter, E.; Cirlan, I.; Khuu, N.; Fischer, S.; Rozenblatt-Rosen, O.; Regev, A.; McGilvray, I.D.; Bader, G.D.; MacParland, S.A. Single-Cell, Single-Nucleus, and Spatial RNA Sequencing of the Human Liver Identifies Cholangiocyte and Mesenchymal Heterogeneity. *Hepatology Communications* **2022**, *6*.
39. Tung, P.Y.; Blischak, J.D.; Hsiao, C.J.; Knowles, D.A.; Burnett, J.E.; Pritchard, J.K.; Gilad, Y. Batch effects and the effective design of single-cell gene expression studies. *Scientific Reports* **2017**, *7*, 39921. doi:10.1038/srep39921.
40. Grindberg, R.V.; Yee-Greenbaum, J.L.; McConnell, M.J.; Novotny, M.; O’Shaughnessy, A.L.; Lambert, G.M.; Araúzo-Bravo, M.J.; Lee, J.; Fishman, M.; Robbins, G.E.; Lin, X.; Venepally, P.; Badger, J.H.; Galbraith, D.W.; Gage, F.H.; Lasken, R.S. RNA-sequencing from single nuclei. *Proceedings of the National Academy of Sciences* **2013**, *110*, 19802–19807. doi:10.1073/pnas.1319700110.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.