A Protocol for Revealing Oral Neutrophil Heterogeneity by Single-Cell Immune Profiling in Human Saliva

Saibyasachi N. Choudhury
J. Craig Venter Institute https://orcid.org/0000-0002-7299-5540
Mark Novotny
J. Craig Venter Insitute
Brian D Aevermann
J. Craig Venter Insitute
Steven Lee
UCLA
Aishwarya Mandava
J. Craig Venter Institute
Yu Qian
J. Craig Venter Institute
Scheuermann, Richard
J. Craig Venter Institute
Marcelo Freire (mfreire@jcv.org)
J. Craig Venter Institute https://orcid.org/0000-0003-4906-7698

Method Article

Keywords: Single-cell, Transcriptomics, Saliva, Neutrophils, Heterogeneity

Posted Date: July 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.pex-953/v2

Abstract

Neutrophils are the most abundant white blood cells in the human body responsible for fighting viral, bacterial and fungi infections. Out of the 100 billion neutrophils produced daily, it is estimated that 10 % of these cells end up in oral biofluids. Because saliva is a fluid accessible through non-invasive techniques, it is an optimal source of cells and molecule surveillance in health and disease. While neutrophils are abundant in saliva, scientific advancements in neutrophil biology have been hampered likely due to their short life span, inability to divide once terminally differentiated, sensitivity to physical stress, and low RNA content. Here, we devise a protocol aiming to understand neutrophil heterogeneity by improving isolation methods, single-cell RNA extraction, sequencing and bioinformatic pipelines. Advanced flow cytometry 3D analysis, and machine learning validated our gating system model, by including positive neutrophil markers and excluding other immune cells and uncovered neutrophil heterogeneity. Considering specific cell markers, unique mitochondrial content, stringent and less stringent filtering strategies, our transcriptome single cell findings unraveled novel neutrophil subpopulations. Collectively, this methodology accelerates the discovery of salivary immune landscapes, with the promise of improving the understanding of diversification mechanisms, clinical diagnostics in health and disease, and guide targeted therapies.

Introduction

Rapid advances in mapping single-cell transcriptional states have been made through the Human Cell Atlas (HCA)[1] and the NIH BRAIN initiatives, providing important insights into human health[2]. By considering transcript and protein compositions, functional categorization of cells is reshaping our models of ontology and helping discover new cell populations and elucidating heterogeneity. Immunity works through orchestrated cellular actions with specialized cellular tasks. In an effort to catalogue immune repertoire, studies are surveying specific tissues and biofluids revealing unique signals, differentiation patterns, cell activation states and diversity. While a large number of immune cell populations, especially lymphocytes and macrophages, have been widely investigated by single cell analysis there is a lack of investigation in neutrophils, the most abundant myeloid cells [3–6]. Investigating neutrophil cell diversity is an emerging field with the potential to reveal novel cell functions and applications [7–10].

Neutrophils are produced at a rate of 10^{11} cells/day, comprising 50 to 70 % of all leukocytes [11] and highly active to respond to infection, injury, migrating to sites initiating inflammation [12,13]. Human neutrophils mature in the bone marrow from committed myeloid precursors and through subsequent stages of differentiation they develop into segmented adult mature cells [14]. Characterization of neutrophil diversity by transcriptomic data has been limited due to their complex nature including short lifespan, lower RNA

content, high levels of intracellular phosphatases and nucleases, and sensitivity to handling during experimentation, resulting in false negative and low heterogeneity in single cell studies. Neutrophils are very sensitive to physical stress such as higher centrifugation speeds, and cold temperatures commonly used in processing of cells for sorting and RNA-seq. Moreover, the levels of neutrophils in published datasets could have been artificially reduced by quality control steps in data processing of sc-RNAseq data which bias against their properties of neutrophils [15,16].

Conventional protocols in blood employ peripheral blood mononuclear layer PBMC as a sample source for single-cell transcriptomic studies of immune cells. In the context of neutrophils, this method allows the detection of homogenous normal density neutrophils (NDNs) found in the granulocyte layers at the interface between red blood cells and the gradient layers, but fails to detect other populations, including activated NDNs that are immunosuppressive with proinflammatory functions [17]. These cells are called low density neutrophils (LDN) and are only found in chronic diseases at the mononuclear layer (PBMCs) between gradient and plasma, which are not found in samples from healthy subjects[18]. Another level of difficulty is the fact that cells in PBMCs contain 10-20 times more RNA than neutrophils [19], thus sequencing data from neutrophils gets masked or filtered out through conventional PBMC protocols [19]. Consequently, in cell heterogeneity studies, neutrophil detection has not followed the trends of other immune cells with regard to diversity, which we believe is related to these technical issues. Since the results from neutrophil single cell transcriptomics reveal either low gene expression levels or decreased functional diversity [15,16,20,21], there is a need to overcome technical difficulties and further understand neutrophil heterogeneity.

In addition to blood, neutrophils are also found in high abundance in mucosal tissues and biofluids such as urine, tears, and saliva. The oral cavity is, however, an excellent and non-invasive source to access biological materials for sampling and single cell studies. Buccal swabs and saliva are the two most common oral sampling methods used for biomedical research. Saliva is rich in mucins, host cells, and most types of proteomic markers found in the body are also detected here. The immune salivary landscape is poorly defined, and leukocytes, lymphocytes and squamous epithelial cells are also present in the oral cavity [22]. The cellular content is derived from mucosal blood vessels, crevicular fluids, oral sweat, and salivary glands. It is estimated that 1/10 of all neutrophils produced daily in the bone marrow are destined for oral tissues [23]. The function, diversity and heterogeneity of neutrophils remains elusive, and while knowledge from bulk assays has been developed suggesting heterogeneity, unbiased single-cell investigations are needed to capture the full repertoire.

Processing single cells in biofluids presents unique challenges including preservation of activation states and relationship to the microbiome that require the development of specialized methods, including buffers, cell handling and protocol to successfully derive scientific insights to specific cells and their populations. In addition to biological limitations specific to neutrophils, there are no established protocols aiming to improve their detection rate and functional analysis. In this study, we establish that neutrophils require unique handling strategies to retain their RNA integrity during experimental procedure than those commonly used for preparation of other cell types in RNA-seq studies. We further demonstrate a protocol that uncovers the diversity of neutrophil heterogeneity. A modified Smart-seq2 method for cDNA synthesis, and bioinformatics pipeline for scRNA-seq data analysis of low mRNA-containing cells is shown, profiling of a large number of cell populations and genome-wide features by changes of buffers and cell handling of neutrophils. We detected novel neutrophil subpopulations and leveraged recent technological advancements in bioinformatics to design filtering criteria to define the signatures. In combination with machine learning profiling, and 3D flow cytometry analysis, we have validated our findings and explored clusters based on surface markers and density. Altogether, our protocol and feasibility studies demonstrated signatures of cells contributing to the human oral neutrophil heterogeneity. We suggest that our methods can be applied to heterogeneity in salivary immune cells, especially oral neutrophils, as an important feature to health and disease investigations, providing insights into possible mechanisms of cell diversification and their functions.

1.1. Development of the Protocol:

Our experimental workflow (Fig. 1) begins with harvesting primary cells by collection of saliva through oral rinse from three healthy donors to obtain cells present in the buccal cavity, using 0.9 % sterile saline solution, followed by a series of filtration using 40, 20, and 10 μm cell strainers to eliminate food particles and larger epithelial cells to obtain > 98 % enriched saliva neutrophil. After each filtration step the change in the cell population is verified by microscopic imaging of unstained cells by 3D- holotomographic microscope (Fig. 1A). Unstained/unlabeled neutrophils are then FACS sorted for high-FSC and high-SSC to obtain cell samples for unbiased transcriptomic gene-signature profiling and the possible discovery of new sub-population(s) of neutrophil (Fig. 1B), which is the primary goal of this study. Single cells were FACS sorted into separate wells containing lysis buffers in a 384 well FrameStar plate and stored at -80 °C for downstream processing. scRNA-seq was performed by a modified Smart-seq2 protocol following various quality-control (QC) steps between each step [24]. Single cell wells passing QC assay criteria were selected for NexteraXT library preparation and sequencing on an Illumina NovaSeq 6000 platform. Bioinformatics analysis was performed on the generated sequencing data to access sequence quality and batch effects, to map reads to the reference human transcriptome and quantify expression, to identify distinct transcriptional phenotypes using unsupervised clustering, and to evaluate expression of known neutrophil markers from the literature (Fig. 1D). Enriched saliva neutrophil samples were also used for FACS sorting to validate expression of known neutrophil markers including $CD66b^{+}CD11b^{+}CD14^{-}$ (Fig. 1C).

1.1.1. Sample Handling for Saliva Neutrophil Enrichment:

Fresh primary human saliva was harvested and prepared from healthy donors. Each donor rinsed five times with ~10 mL of 0.9 % NaCl solution (sterile) for 30-60 sec. with a gap of 3 min between each rinse to collect 50 mL total volume. The cells were then pelleted down through centrifugation at 160 x g at room temperature (RT) for 5-10 min. From 50 mL total volume, 40 mL of supernatant were discarded by aspirating carefully without disturbing the cell pellet. It is of utmost importance to maintain each step of primary saliva neutrophil isolation at RT as the cells tend to lose viability faster at lower temperature (e.g. at 4 °C). Cells were then resuspended in the remaining ~10 mL of 0.9 % saline solution and filtered through a 40-micron nylon mesh filter using gravity to remove food particles and eliminate mucus present in the collected oral rinse. 40 µm filtered oral rinse was sequentially filtered through 20 µm and 10 µm pluriStrainer nylon mesh filters to remove epithelial cells. All filtration steps were done by gravity flow or centrifugation at 160 x g at RT for 1 min. Vacuum suction for pluriStrainers should be avoided due to increased contaminants. Viability and cell density were determined through a trypan blue exclusion using the Countess automated cell counter. Cells obtained by this method from a single healthy donor in 50 mL oral rinse of resting saliva was ~1.8 x 10^5 cells. The purity of saliva neutrophils isolated by this enrichment method was verified by both microscopy (Fig. 2) and flow cytometry (Fig. 3). To compare the identification of neutrophil subpopulation by flow cytometry and scRNA-seq, we looked at the heterogeneity of the immune cell in whole saliva by computational analysis of flow cytometry data using FLOCK [25] unsupervised analysis for identifying both known and novel cell populations from the flow cytometry data (Fig. 4). The expression profile of neutrophil-specific genes from the scRNA-seq data also verified the purity of the FACS-sorted single cell as neutrophils (Fig. 5 and Fig. 6).

1.1.2. Flow Cytometry Sorting:

To ensure unbiased transcriptomic patterns, cells were not labeled with antibody/cell-surface-marker prior to sorting. The viability of cells was checked before each sorting and only samples containing > 70 % viable cells were sorted. Cells were sorted using BD FACS Aria II (BD Biosciences) with custom PMT2 using a 130-micron (10 PSI) nozzle size. A nozzle size of 100-micron is desirable due to a flow pressure of 20 PSI and would also be suitable to use. The flow pressure of 70 PSI generated by a 70-micron nozzle during sorting affects both the viability and RNA quality of sorted neutrophil cells. Cell populations were selected by software gating for high-FSC/high-SSC Scatter-gate followed by doublet-discrimination (DD) (FSC-DD and SSC-DD gates) to isolate single cells from any multicellular clusters/clump. Except for control wells, single cells were sorted into each well of a Framestar 384-well plate containing 2 µL of pre-dispensed lysis buffer having ERCC (External RNA Controls Consortium) spike-in RNA standards (final concentration of 2 % Triton X-100, 2 U/µL RNase inhibitor, 1:20,000 dil. ERCC and nuclease free water). Blank or ‘No Template Control (NTC)’ wells contained 2 µL nuclease free water only; ‘spike-in’ control well contained lysis/FACS sort buffer with ERCC but no cell; and ‘UHR’ control well contained 2 µL lysis/sort buffer + 10 pg of Universal Human RNA (UHR). Except for the sorting-sample tube, all other reagents and FACS sorting plates were maintained at 4 °C on ice by using chill blocks. The lysis/sort buffer must be prepared fresh. The sorted plates were then frozen on dry ice and stored at -80 °C for downstream

processes, including cell lysis, cDNA synthesis and PCR amplification according to the SMART-seq2 protocols. Out of three 384 wells, we generated 1084 single-cell cDNA wells and selected for QC-Pass criteria/parameters set to a single cDNA library plate (Supplementary Fig. 2).

To compare the effect of pressure and electrostatic charge on cells while sorting by BD FACS Aria II, we also tested single cell sorting using other microfluidics based cell sorters, including the On-chip Sort-microfluidic chip cell sorter (On-chip Biotechnologies Co., Ltd, Japan) and the WOLF Cell sorter (NanoCellect Biomedical Inc., CA, USA). With the WOLF cell sorter, the volume requirement was very high and not compatible for use with the Smart-Seq II reaction. On the other hand, the volume needed by On-chip Sort is minimal allowing for its use with the Smart-seq protocols. In addition, On-Chip-SPiS single-cell dispenser captures high quality images providing a visual confirmation of single deposition into the wells (Supplementary Fig. 1). We compared the purity of bulk sorted cells and quality of RNA of both systems and found that cell viability and cDNA obtained from On-Chip sorters were superior. However, due to its inherent pipetting system, the timing of this procedure was longer than neutrophil handling would allow. BD FACS Aria II workflow provided cell sorting with the volume needed, during a time frame compatible with maintaining good neutrophils viability and RNA quality.

1.1.3. Cell lysis, cDNA Synthesis and Quality Control:

Detailed procedure of cell lysis and cDNA for Smart-seq2 protocol was previously published [26]. In addition to modifications [24] e.g. increasing cDNA PCR-amplification cycle from 18 to 21 cycles to compensate for lower mRNA levels in neutrophil [15,16,20] and modifying Template Switch Oligo (TSO) primer by 5'-biotinylation to reduce non-specific amplification caused by TSO concatemers [24,27]. Before preparing the library for scRNA-seq, we carried out quality control assays of the cDNA samples by quantification using a PicoGreen dsDNA quantitation assay kit and qRT-PCR for ACTB expression using a TaqMan gene expression assay. A Hamilton Microlab STAR liquid handling system was used to select samples that passed the QC parameters for both PicoGreen dsDNA assay (RFU > 0.3) and qRT-PCR for ACTB expression (Ct < 35) (Supplementary Fig. 2).

1.1.4. Preparation of sequencing library and sequencing:

We employed the Illumina Nextera XT DNA library preparation kit and performed multiplexed paired-end sequencing of barcoded libraries using an Illumina MiSeq in order to determine the required sequencing parameters for sequencing on the pooled libraries on a NovaSeq 6000 sequencing system. Samples in each well were barcoded by using a unique combination of Nextera XT Index Kit v2 Set A and Set D to identify sequence data from each single cell. To ensure initial quality of the sequencing libraries, we performed a MiSeq run on a pooled library of 16 randomly selected Nextera XT libraries created from the selected cDNA plates from each donor in order to determine the sequence quality and coverage needed to discover subpopulations among saliva neutrophils. Prior to loading the MiSeq, an Agilent 2100 BioAnalyzer High Sensitivity DNA chip run was performed and insert size of the sequencing library was

determined to be between 250–500 bp. For the MiSeq-MO (medium output) run, 270 pM of 16-plex pool library was spiked with 1 % PhiX174. The 150 base paired-end MiSeq run of 16 single-cell data showed an average Q30 of Read 1 (75%), Read 2 (69%), i7 (83%) and i5 (79%) pass filter rate and a sequencing depth of 1.5–2.0 × 10^6 reads per cell. It has been previously shown that read-depth of 1.5–2.0 × 10^6 is adequate for the detection of saturating levels of RNA expression in single cells [28]. This information suggested that we could perform a sequence run of NovaSeq 6000 with a single pool of 384 Nextera XT libraries without over saturation of the RNA-seq read depth and gene counts from each single cell.

Based on the MiSeq run results, a total of 3 NovaSeq 6000 SP flow cell 2x150 XP workflow runs were performed on each of the 3 donors with 3 library pools consisting of 384 Nextera XT libraries from both single cells and controls. Each pool was loaded at 300 pM with a 1 % PhiX174 spike-in. Quality of 3 NovaSeq 384-plex pooled libraries were analyzed by both Agilent HS DNA chip and qPCR-based NGS library quantification using KAPA Library Quantification Kit -Illumina. FACS sorted single-cell plates were evaluated by PicoGreen dsDNA quantification assay (for cDNA concentration) and qRT-PCR TaqMan assay (for expression of housekeeping gene ACTB) for cDNA quality. Each selected cDNA library was used to generate an Illumina Nextera XT library and combined into a 384-plex pool for sequencing on the Illumina NovaSeq 6000 system.

1.1.5. Cell Morphology and Immunofluorescence 3D Holotomographic Microscopy:

Cell diversity of healthy human resting saliva and elimination of each cell type during each filtration steps of 40-, 20-, and 10-micron by our saliva neutrophil enrichment protocol was analyzed by 3D-holotomography imaging of unstained cells and cells stained with fluorophore-conjugated cell surface CD-markers using 3D Cell Explorer microscope (Nanolive's 3D Cell Explorer-*fluo*; Model CX-F). Cells were collected after each filtration step of neutrophil enrichment protocol. Unstained and stained cells suspended in 0.9 % saline solution and staining buffer respectively were imaged at 60X magnification using class 1 low power Laser (λ=520 nm, sample exposure 0.2 mW/mm^2) and a depth of field of 30 µm. For fluorescent imaging, cells were stained by a flow cytometry staining protocol and were imaged by the fluorescent module of 3D Cell Explorer-*fluo* in DAPI, GFP (or FITC), and OFP (or TRITC) fluorescent channels. Exposure of 100/100/100 ms, gains of 35/35/35 and intensities of 35/50/50 were used to capture images in DAPI/GFP/OFP-channels, respectively (Fig. 2A).

1.1.6. Validation of Neutrophil Morphology by Microscopy and Giemsa Staining:

To confirm the identity of unstained neutrophils done by 3D Holotomography microscope, based on their intracellular structure, we stained the same samples with Giemsa stain solution and imaged by Zeiss AxioVision microscope (Carl Zeiss Microscopy, LLC, NY, USA) at 40X objective using Zen Blue software. For staining, enriched neutrophils obtained after 10 µm filtration step were spread into a cell monolayer on the charged side of the slide by CytoSpin 4 centrifugation. Around 25,000 cells in 50-200 µL volumes of cell suspension are loaded in Cytofunnel with white filter cards and caps (Shandon EZ Single

Cytofunnel, cat. no. A78710003) for each slide and centrifuged at 700 RPM for 7 min at medium acceleration. Boundaries were drawn around the cell monolayer by using a hydrophobic pen and allowed to dry at RT inside the hood for ~30 mins. After drying, the cells were methanol fixed by incubating for 5-7 min at RT. The slides were carefully washed twice in $PBS^{(-)}$ and dried by draining the $PBS^{(-)}$ completely. At this step, the slides were dried and stored overnight at 4°C. After drying, Giemsa stain solution (1:20 dilution) was added and incubated at RT for 30-60 mins. After incubation, the slides were washed carefully by draining the stain and slowly dipping the slide in an angular position in deionized water to prevent cells from getting washed away. A second wash was performed by dipping the slide in fresh deionized water for 2 min. Slides were dried in the hood and then mounted with coverslip using aqueous based mounting media such as CytoSeal 60 and observed in brightfield under 40X objective of microscope and images. Images were captured and processed by using Zen Blue software (Fig. 2B) and Nanolive 3D-Cell explorer microscope (Fig. 2C; section 1.9.10).

1.1.7. Immunofluorescence for Microscopy and Flow Cytometric Analysis:

Cells from saliva samples were further processed for immunofluorescence. Briefly, cells were fixed by 4 % paraformaldehyde (PFA) on ice for 30 min. After incubation, fixed cells were washed at least twice by adding a staining buffer and centrifuged at 160 x g for 10 mins at RT to pellet cells. Because neutrophils don't form a clear pellet, the supernatant is carefully aspirated at slow speed from the top. After cell counting, 1 µL of Fc-block per million cells was added, followed by incubation for 15 mins at RT. The volume was reconstituted to 200 µL cell suspension/tube. Anti-human monoclonal antibodies were added to each tube according to manufacturer instructions. In the master mix, cell surface markers were added and incubated at RT for 1 hr in dark. Suggestive cell surface markers: PerCP/Cy5.5-CD11b (cat. no. 101227, BioLegend, CA), Brilliant Violet 650-CD15 (cat. no. 323033, clone W6D3, BioLegend, CA), Pacific Blue-CD66b (cat. no. 305111, clone G10F5, BioLegend, CA), APC/Cy7-CD3 (cat. no. 300425, clone UCHT1, BioLegend, CA), PE-CD19 (cat. no. 302207, clone HIB19, BioLegend, CA), APC-CD14 (cat. no. MHCD1405, Invitrogen, CA), FITC-CD18 (cat. no. MHCD1801, Invitrogen, CA). Cell counts were obtained by trypan blue exclusion in the Countess cell counter (Invitrogen, CA). Fixed-unstained/stained samples for flow analyses were run on a BD FACS Aria II (BD Biosciences) to obtain raw data in FCS format, which were later analyzed by FlowJo v10.6.1 (BD Biosciences) for 2D analysis (Fig. 3) and by FLOw Clustering without K (FLOCK v1) [25] for computational analysis using all markers simultaneously (Fig. 4), which is explained in detail in section 1.1.11. Each sample was analyzed for ‘ungated’ total cell populations and ‘high-SSC gated’ populations used for single-cell sorting of saliva neutrophil in this study.

1.1.8. scRNA-Seq data processing and Analysis:

Single-cell RNA-seq data were processed according to the procedure described in detail in Krishnaswami et. al. 2016 [24]. Briefly, raw sequencing files were demultiplexing using Illumina barcodes, while sequencing primers and low-quality bases were removed using the Trimmomatic package [29]. Trimmed reads were then aligned using HISAT [30] in two steps: first to a reference of ERCC sequences, and then the remaining reads were mapped to GRCh38 (GENCODE). StringTie [30] was then used to assemble the

resulting alignments into gene expression values (TPM) estimated using GENCODE v25 annotation (Ensembl 87; 10-2016).

Expression values for non-control cells were imported into Scanpy for PCA, UMAP, and cluster analysis [31]. Cells were filtered for quality using two different approaches. The first liberal filtering consisted of removing cells with less than 50 genes per cell or with greater than 10% of the total gene counts being from mitochondrial genes. Genes were required to be present in at least 4 cells, with greater than 50 total counts. From these genes, the top 2500 variable genes were selected (Fig. 5). The second stricter filtering consisted of removing cells with less than 400 genes per cell or with greater than 10% of the total gene count being from mitochondrial genes. In fact, genes were required to be present in at least 4 cells, with greater than 50 total counts. From these genes, the top 2000 variable genes were selected (Fig. 6).

Unsupervised clustering was performed by first performing PCA to determine principal components, then a neighborhood graph was constructed using those components. Next Louvain clustering was performed using the neighborhood graph. Using the Louvain clustering solution, marker gene determination was performed using logistic regression[31]. The outputs of this computational pipeline produce a set of unbiased cell type clusters, a gene expression matrix with the expression levels of genes in individual single cells grouped into cell type clusters, and a set of marker genes for each cell type cluster.

1.1.9. Publicly available healthy control data

In addition to salivary neutrophils, RNA-seq data of immune-cell types and PBMC obtained from blood were acquired from the Gene Expression Omnibus (GEO) database under accession code GSE64655[32]. This data was analyzed by Scanpy in a similar fashion described above. Expression patterns for targeted marker genes, those with known differential expression patterns in the oral and blood neutrophils, were visualized (Supplementary Fig.6).

1.1.10. Microscopic Analysis:

Raw images obtained for both unstained and stained cells were processed by software (STEVE software v1.6.3496, Nanolive) with similar parameters for the brightfield images which was also used to obtain digital acid-stained and RI 3D-rendering images. Parameters for processing of raw images captures in each fluorescent channel (i.e. DAPI, FITC and TRITC) to determine the background cutoff pixel was determined by comparing with images of unstained negative control samples. These parameters were used consistently for all images processed for each fluorescent channel.

For unbiased identification of cell types based on CD-markers, unstained cells were imaged and stained via digital acid-staining (STEVE software v1.6.3496, Nanolive) according to the refractive index (RI) of the intracellular structures (Fig. 2A). Four major cell types i.e. epithelial cells, neutrophils, monocytes, and

lymphocytes were identified based on their size and nuclear structure and quantified by manually counting from multiple images obtained from three healthy donors (Fig. 2C). To further verify the neutrophil enrichment and confirm the elimination of other cell types, we did Giemsa staining of 10 µm filtered unstained samples (Fig. 2B insert).

To rule out the biased staining of cell types with cell surface markers, stained samples used in flow cytometry analyses were further imaged for selected CD-markers conjugated with fluorophores detectable in the three fluorescent channels (i.e. DAPI, FITC, and TRITC channels) available in Nanolive's 3D Cell Explorer microscope. Classical neutrophil positive marker CD-66b conjugated with pacific blue was imaged in the DAPI channel, whereas classical neutrophil negative markers CD14 and CD19 conjugated with FITC and PE, respectively were imaged in the FITC and TRITC channels (Fig. 2A). Four major CD-marker positive cells i.e. CD-66b-PacBlue for neutrophils, CD14-FITC for monocytes, and CD19-PE for B-lymphocytes were identified based on obtaining positive signals in the respective channels.

1.1.11. Flow cytometry Data Analysis:

The FCS files from BD FACS Aria II flow cytometry were transformed using FCSTrans[33] on R programming language. FCSTrans applies a logical transformation on fluorescence channels and a linear transformation on scattering parameters followed by a min-max linear rescaling applied across all the channels to scale the range to [0, 4095]. After FCSTrans transformation, area parameters of measured channels were selected and saved to tab-delimited text files for downstream data clustering analysis. Auto-gating is applied using an unsupervised clustering approach – FLOCK[24] (Flow Clustering Without K). FLOCK is publicly available on ImmPort Galaxy (https://immportgalaxy.org/). FLOCK identified 23 cell subsets across the 6 stained samples (two replicates for each of the three filtrations: 40 µm, 20 µm and 10 µm) by using all measured parameters simultaneously in density-based clustering analysis. Then each identified cell subset was visually examined on dot plots of all 2D marker combinations for manually annotating the phenotype (e.g., $CD3^{-}CD19^{-}CD14^{-}CD11b^{+}CD15^{+}CD66^{+}$ neutrophils).

Frequencies of the 23 FLOCK-identified cell subsets were quantified by their percentages (with the total number of cells as the parent). Mean fluorescent intensity (MFI) values of each cell population for each marker were also calculated. Figure 4A visualizes the MFI values of a 40 µm sample (Tube-10_Specimen_001_002) using a heatmap for indicating the phenotype of each identified cell population. Based on visual examination of the 2D dot plots, 5 of the 23 FLOCK-identified cell subsets are salivary neutrophils. For each neutrophil subpopulation identified by FLOCK, mean percentage is calculated for duplicate files in each filtration step. Figure 4B shows the bar graph for mean percentages across the five FLOCK-identified neutrophil subsets, indicating that the major/abundant neutrophil subpopulations consistently increased their percentages as filtration size decreased. For both manual annotation and interpretation of the 5 neutrophil subpopulations, 2D dot plots (Fig. 4C) of the 40 µm sample with each neutrophil subpopulation highlighted in a different color were generated to visualize the phenotype differences of these subpopulations.

1.2. Comparison to other methods:

- The key strength of our protocol are as follows:

- The protocol overcomes the methodological limitations that produced the false-negative expression of neutrophils in many published studies due to low mRNA content of neutrophils and the different experimentation conditions needed in comparison with PBMCs.

- Saliva is an easily accessible and readily available clinical sample which makes this protocol non-invasive to patients requiring deep sequencing for diagnosis.

- We suggested important experimental conditions for neutrophil that are overlooked by researchers, different from processing of PBMCs. Neutrophils prefer room temperature and lower centrifugation speeds of 160 x g to 300 x g.

- This study presented processing time of primary neutrophils and the rate of RNA degradation. This information is important to consider a sample for further downstream steps of an NGS workflow.

- For deep sequencing, the samples should be collected fresh and processed within 4 hours. Storage of samples at -150 ℃ (liquid N2) or -20 ℃ after collection by DMSO-cryopreservation and methanol-fixation protocols established in other immune cells are not suitable for deep sequencing of neutrophils [16,34].

- This protocol demonstrates stringency levels needed for filtering raw sequence data during bioinformatics analysis. It suggests inclusion of mitochondrial gene expression data during analysis that is considered an exclusion criterion for single cell sequence data processing.

1.3. Limitations of this protocol:

- This study is aimed for unbiased sequencing of saliva neutrophils in which unlabeled morphology was the gating determinant. While we coupled with specific pipelines able to detect other cells, it may lead to lack of specificity in the sorting scheme.

- Neutrophils are complex cells that are fragile and get easily activated during handling 15. Thus, personnel need to be trained, and reagents prepared carefully to yield replicable results.

- Neutrophils contain 10-20 times lower amounts of RNA per cell than PBMC 19. Therefore, if another PBMC cell is sorted along with neutrophil by chance during FACS sorting, the probability of amplification of non-targeted cell mRNA during cDNA synthesis is higher, which may lead to erroneous sequencing results.

- Due to the small amount of mRNA in neutrophils it may be necessary to optimize the PCR cycles required to obtain sufficient cDNA for the NGS-workflow. We amplified the single cell neutrophil cDNA with 21 cycles, similar to single nucleus protocols24 because of low amounts of RNA, compared with 18

cycles for other cell types 26. Some low-copy number transcripts may still be difficult to detect in neutrophils. However, increasing the number of PCR cycles could introduce some amplification bias in the library by compressing expression values for high-copy number transcripts.

- Small noncoding RNAs (ncRNAs) and other short sequence mRNAs lacking polyA tails would not be detected. The low amounts of RNA contained in the neutrophils may also prevent the detection of some ncRNAs.

- Since the neutrophils are found to have generally lower RIN values compared to other cell types, it is possible that the RNA of the samples are partly degraded and may result in an increasingly 3' biased library preparation as suggested by Chen et al., and therefore losing valuable reads from your data is a possibility 35.

1.4. Future Applications:

Comparing Health and Disease. Viral, autoimmune, metabolic, and chronic inflammatory diseases require novel and non-invasive methods to monitor cellular phenotypes from humans, comparing health versus disease states. This protocol provides the experimental conditions and time needed for processing of neutrophils for NGS-workflow to obtain their transcriptomic signatures.

Revealing Oral-Systemic Axis. Emerging evidence demonstrates that markers expressed in biofluids such as saliva are representative of systemic changes. A protocol for the unbiased evaluation of single cells in saliva could yield a better understanding of systemic health through oral sampling.

Longitudinal Monitoring. Sampling saliva is non-invasive and easy to perform. Thus, continuous monitoring of cells, biomarkers and gene expression patterns in saliva provides an effective system for longitudinal survey. In addition to research studies, this system would also be optimal for the development of novel diagnostic systems and drug delivery.

Reagents

2.1. Reagents:

- *Cell sample*: We have successfully isolated human saliva neutrophils by series of the filtration processes and obtained enrichment of neutrophils. We have accessed the purity of the enriched neutrophil by microscopic and flow analysis. CAUTION: An Institutional Review Board approval or patient consent form may be required for sample collection from healthy donors.

- HL-60 cell line (ATCC, cat. no. CCL-240)

- Iscove's Modified Dulbecco's Medium (IMDM) (Gibco, cat. no. 12440-061)
- Roswell Park Memorial Institute (RPMI) 1640 (Gibco, cat. no. 11875-093)
- Fetal Bovine Serum (FBS) (Gibco, cat. no. 26140-095)
- Penicillin-Streptomycin (10,000 U/mL; Gibco, cat. no. 15-140-122)
- Dimethyl sulfoxide (DMSO) (ATCC, cat. no. 4-X)
- Centrifuge tubes (50 mL; Denville Scientific Inc., cat.no. C1060-P)
- 0.9% Saline solution sterile (Teknova, cat. no. 50-843-140)
- pluriStrainer 20 µm-sterile cell strainer (pluriSelect, cat. no. 43-50020-03)
- pluriStrainer 10 µm-sterile cell strainer (pluriSelect, cat. no. 43-50010-03)
- 40 µm sterile cell strainer (Fisher Scientific, cat. no. 22363547)
- Connector ring (pluriSelect, cat. no. 41-50000-03)
- RNaseZap RNase decontamination solution (Ambion, cat. no. AM9780)
- Nuclease free water (Ambion, cat. no. AM9932)
- β-Mercaptoethanol (14.3 M; Sigma, cat. no. M6250-100 mL) CAUTION: This is a combustible liquid. Avoid contact with skin and eyes. Avoid inhalation of vapor or mist and handle it while you are wearing appropriate personal protective equipment (PPE). It is toxic if swallowed or if inhaled. It is very hazardous in case of skin contact (permeator) and ingestion. Severe overexposure can result in death. It causes skin irritation, and it may cause an allergic skin reaction. It also causes serious eye damage.
- Tris buffer (pH 7.0, 1 M; Ambion buffer kit, cat. no. 9010)
- Magnesium chloride ($MgCl_2$) (1 M; Ambion buffer kit, cat. no. 9010)
- EDTA (pH 8.0, 0.5 M; Ambion buffer kit, cat. no. 9010)
- RNase inhibitor, cloned (40 U/µl; Ambion, cat. no. AM2682)
- Triton X-100 (Sigma-Aldrich, cat. no. T8787-100ML). CAUTION: Harmful if swallowed and can cause serious eye damage. Handle it while wearing appropriate PPE.
- dNTP mix (10 mM each; Thermo Fisher, cat. no. 18427-088)
- ProtoScript II reverse transcriptase (New England Biolabs, cat. no. M0368X). Includes:

- ProtoScript II buffer (5X)
- 1,4-Dithiothreitol (DTT) 0.1 M
- ProtoScript II Reverse Transcriptase (200 U/µl)
- Betaine (BioUltra, ≥ 99.0 %; Sigma-Aldrich, cat. no. 61962)
- PicoGreen dsDNA quantitation assay kit (Invitrogen, cat. no. P7589)
- KAPA HiFi HotStart ReadyMix (2X; KAPA Biosystems, cat. no. KK2602)
- KAPA Library Quantification Kit - Illumina (KAPA Biosystems, cat. no. KK4835)
- Ethanol- molecular biology grade (Sigma-Aldrich, cat. no. E7023-500 ml)
- Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63881)
- Adapter oligos (See cDNA Synthesis on section 3.4). Locked Nucleic Acid (LNA)-modified TSO were ordered from QIAGEN (https://www.qiagen.com/). All other oligos were ordered from IDT (https://www.idtdna.com). All oligos were HPLC-purified. The identity of LNA-modified TSO compounds is also confirmed by MS.
- Biotin TSO Custom LNA Oligonucleotide (1 µmole synthesis; QIAGEN, cat. no. 339413; GeneGlobe ID: YCO0078131; 5’-biotin-AAGCAGTGGTATCAACGCAGAGTACrGrG+G-3’)
- oligo-dT (1 µmole, 57 bases; 5'-AAG CAG TGG TAT CAA CGC AGA GTA CTT TTT TTT TTT TTT TTT TTT TTT TTT TVN-3')
- ISPCR oligo (1 µmole, 23 bases; 5’-AAG CAG TGG TAT CAA CGC AGA GT -3’)
- UltraPure BSA (50 mg/ml; Ambion, cat. no. AM2616)
- Trypan blue (0.4 %; Sigma-Aldrich, cat. no. T8154)
- ERCC RNA spike-in mix 1 (Ambion, cat. no. 4456740)
- Universal Human RNA (UHR) control (Takara/CloneTech, cat. no. 636538, discontinued product)
- PhiX Control v3 Library (Illumina, cat. No. FC-110-3001)
- RNase-free PBS, pH 7.4 (Ambion, cat. no. AM9625)
- 0.5 % RNase-free BSA (Ambion, cat. no. AM2616)
- Mouse IgG1κ (BD Pharmingen, cat. no. 554121)

- Fluorochrome conjugated anti-human monoclonal antibodies:
- PerCP/Cy5.5-CD11b (BioLegend, cat. no. 101227)
- Brilliant Violet 650-CD15 (clone W6D3; BioLegend, cat. no. 323033)
- Pacific Blue-CD66b (clone G10F5; BioLegend, cat. no. 305111)
- APC/Cy7-CD3 (BioLegend, cat. no. 300425, clone UCHT1)
- PE-CD19 (clone HIB19; BioLegend, cat. no. 302207)
- APC-CD14 (Invitrogen, cat. no. MHCD1405)
- FITC-CD18 (Invitrogen, cat. no. MHCD1801)
- LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Invitrogen, cat. no. L34965)
- Cell Staining Buffer (BioLegend, cat. no. 420201)
- Paraformaldehyde (PFA) (16 % w/v; Electron Microscopy Sciences, cat. no. 15710)
- Giemsa stain solution (LabCam, cat. no. LC148407)
- CytoSeal 60 (Thermo Scientific. cat. no. 8310-4)
- Yellow fluorescent polystyrene microspheres, 10 µm (Spherotech, cat. no. FP-10052-2)
- PerfeCTa qPCR FastMix II, ROX (2X, Quanta Biosciences, cat. no. 95119-05K)
- TaqMan gene expression assay, ACTB (60X, Applied Biosystems, cat. no. 4351368, assay ID-Hs01060665_g1).
- RNeasy Mini Kit (50) (Qiagen, cat. no. 74104)
- AllPrep DNA/RNA/Protein Mini Kit (50) (Qiagen, cat. no. 80004)
- Quant-iT PicoGreen dsDNA assay kit (Invitrogen, cat. no. P11496)
- Agilent RNA 6000 pico kit (Agilent Technologies, cat. no. 5067-1513)
- Agilent high sensitivity DNA kit (Agilent Technologies, cat. no. 5067-4626)
- Nextera XT DNA library preparation kit, 96 samples (Illumina, cat. no. FC-131-1096)
- Nextera XT Index Kit v2 Set A (Illumina, cat. no. FC-131-2001)
- Nextera XT Index Kit v2 Set D (Illumina, cat. no. FC-131-2004)

- MiSeq Reagent Kit v2 (300-cycles) (Illumina, cat. no. MS-102-2002)
- CytoOne T75 filter cap TC flask (USA Scientific, cat. no. CC7682-4875)
- E-Gel General Purpose Agarose Gels, 1.2 % (Invitrogen, cat. no. A03076)
- 1 Kb Plus DNA Ladder (Invitrogen, cat. no. 10787018)
- Blue/Orange Loading Dye (6X, Promega, cat. no. G190A)
- Countess cell counting chamber slides (Invitrogen, cat.no. C10283)

Equipment

- On-chip Sort the microfluidic chip cell sorter (Model: HSG; On-chip Biotechnologies Co., Ltd, Tokyo, Japan; cat.no. 362S2001G)
- Single Particle isolation System “On-chip SpiS” (On-chip Biotechnologies Co., Ltd, Tokyo, Japan)
- Microfluidic chip for sorting (2D Chip-Z1001; On-chip Biotechnologies Co., cat. no. 1002004)
- BD FACS-ARIA II Flow sorter with an automated cell deposit unit (BD Biosciences)
- BD Falcon tube with a cell strainer cap (Becton Dickinson, cat. no. 352235)
- Falcon polystyrene conical tube (50 mL, BD Biosciences, cat. no. 352095)
- Inverted fluorescence microscope Olympus IX70 (Olympus Corporation)
- Zeiss AxioVision microscope (Carl Zeiss Microscopy, LLC, NY, USA)
- 3D Cell Explorer microscope (3D Explorer-fluo; Model CX-F, Nanolive SA, Switzerland)
- Hemocytometer (Hausser Scientific, cat. no. 1483)
- Countess automated cell counter (Invitrogen, cat. no. C10281)
- 96-well black Fluortrac microplate (VWR, cat. no. 82050728)
- FrameStar Clear 384-well Skirted PCR Plates (Phenix Research Products, cat. no. MPC-384HS4NH-C)
- FrameStar 384 Well Skirted PCR Plates- black frame with white well (Phenix Research Products, cat. no. MPC-384HS4-WW)
- 96-well plates (twin.tec PCR plate 96 LoBind, skirted, colorless; Eppendorf, cat. no. 0030129512)
- 8-strip, nuclease-free, 0.2 mL, thin-walled PCR tubes with caps (Eppendorf, cat. no. 951010022)

- Microcentrifuge DNA LoBind Safe-Lock tubes (1.5 mL; Eppendorf, cat. no. 022431021)
- ErgoOne Multichannel pipette (USA Scientific)
- ErgoOne single channel pipettes set, 1–10 µL; 2–20 µL; 20-200 µL (USA Scientific)
- accu-jet pro Pipette Controller (BrandTech, cat. no. 26332)
- TipOne ultra low retention filter tips- 10 µL; 200 µL; 1250 µL (USA Scientific)
- DynaMag-96 side skirted magnetic rack (Thermo Fisher, cat. no. 12027)
- MicroAmp clear adhesive film (Applied Biosystems, cat. no. 4306311)
- MicroAmp optical adhesive film (Applied Biosystems, cat. no. 4311971)
- GeneAmp PCR System 9700 (Applied Biosystems, cat. no. 4307808)
- QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems, cat. no. 4485694)
- FlexStation 3 multi-mode microplate reader (Molecular Devices, San Jose, CA, USA)
- NanoDrop ND-1000 Spectrophotometer (Thermo Fisher)
- Cytospin 4 Cytocentrifuge (Thermo Scientific, cat. no. A78300003)
- Microlab STAR Liquid Handling System (Hamilton Company, Reno, NV, USA)
- Agilent BioCel 1200 system (Agilent Technologies, Santa Clara, CA, USA, cat. No. G5500-90011)

Integrates following devices:

- Bravo Automated Liquid Handling Platform (Agilent Technologies)
- Direct Drive Robot (DDR)
- Microplate Centrifuge (Eppendorf)
- BioRAPTR FRD Microfluidic Workstation (Beckman Coulter)
- Thermo CUBE 300A Thermoelectric Recirculating Chiller (Solid State Cooling System, NY, USA)
- Auxiliary Barcode reader
- Lid Hotel Station
- Labware MiniHub
- Labware Stacker

- PlateLoc Thermal Microplate Sealer
- Microplate Labeler
- Agilent 2100 Bioanalyzer (Agilent Technologies)
- Refrigerated centrifuge (Eppendorf, Model: Centrifuge 5804 R)
- DNA sequencing instrument. CRITICAL: A compatible Illumina DNA sequencing instrument (MiSeq, NextGen 500, HiSeq 2000, HiSeq 2500, NovaSeq 6000) is necessary to complete sequencing of the Nextera XT libraries, as the barcodes and sequencing adapters are designed for the Illumina sequencing platform.
- 64-bit computer running Linux with 4 GB of RAM (16 GB preferred)

2.3. Equipment software:

- VENUS (for Microlab STAR Liquid Handling System)
- VWorks Automation Control software (for Agilent BioCel 1200 system)
- BioRapTR 3.3.2 (for BioRAPTR FRD Microfluidic Workstation)
- SoftMax Pro (for Flexstation 3)
- BD FACSDiva Software v8.0.2 (for BD FACS-ARIA II Flow sorter)
- FlowJo v10.6.1 (BD Biosciences)
- 2100 Expert Software (For Agilent 2100 Bioanalyzer System)
- STEVE FULL v1.6.3496 (for Nanolive 3D Cell Explorer microscope; Nanolive SA, Switzerland)
- Zen 3.0 -blue edition (for Zeiss AxioVision microscope; Carl Zeiss Microscopy, LLC, NY, USA)

2.4. Sequence data analysis software:

Software function

Software name

Web link

References

For sequence quality assessment

FASTX

http://hannonlab.cshl.edu/fastx_toolkit/download.html

fastQC

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

RSeQC

(alternative to FASTX and fastQC)

http://rseqc.sourceforge.net/

[36,37]

For sequence trimming

Trimmomatic

http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/Trimmomatic-0.33.zip

Cutadapt

(alternative to Trimmomatic)

https://cutadapt.readthedocs.org/en/stable/

[38]

For sequence alignment

HISAT2

http://daehwankimlab.github.io/hisat2/download/

[30]

SAM tools

http://sourceforge.net/projects/samtools/files/samtools/

For RNA expression analysis

StringTie

https://ccb.jhu.edu/software/stringtie/#install

RSEM (alternative to HISAT2/ Stringtie)

http://deweylab.biostat.wisc.edu/rsem/

For data analysis

R

https://cran.r-project.org/

Python

https://www.python.org/

Juptyer

https://jupyter.org/install

Pandas

http://pandas.pydata.org/

Matplotlib

http://matplotlib.org/

Scanpy

https://pypi.org/project/scanpy/

Bedtools

http://bedtools.readthedocs.org/en/latest/

IGV

http://www.broadinstitute.org/igv/

Procedure

3. Procedure:

3.1. Oral rinse collection • TIMING: 1-1.5 hrs

CRITICAL: Maintain each step of primary saliva neutrophil isolation at room temperature (RT) as the cells tend to lose viability faster at a colder temperature (e.g. 4 ℃) that is generally used for PBMCs or any other primary cell type isolation.

1) Label the sterile 50 mL centrifuge tubes for collection of oral rinse and set the centrifuge to RT.

2) Before collecting the first oral rinse, each donor must clean their oral cavity and wait for 3 min.

3) Each donor must rinse their oral cavity five times with ~10 mL of 0.9 % NaCl solution (sterile) for 30-60 sec each time with a gap of 3 min between each rinse to collect 50 mL total volume. (CRITICAL: The collection tube should be immediately proceeded for the enrichment process, as neutrophil RNA keeps degrading over time.)

3.2. Saliva Neutrophil Isolation/enrichment • TIMING: 1-1.5 hrs

4) Pellet down the cells by centrifuging at 160 x g at RT for 5-10 min. (Alternatively, if there is high mucus content in the sample, the cell suspension can be passed through 40 µm sterile cell strainer before centrifugation).? TROUBLESHOOTING

5) Discard 40 mL of supernatant from 50 mL total vol. by aspirating carefully without disturbing the cell pellet. (CRITICAL: Neutrophils do not form a solid pellet)

6) Resuspended the cells in the remaining ~10 mL of 0.9 % saline solution and passed through 40 µm sterile cell strainer using gravity to remove any food particles or mucus present in the collected oral rinse.? TROUBLESHOOTING

7) Then sequentially filter the 40 µm filtered oral rinse through pluriStrainer nylon mesh filters 20 µm and 10 µm to remove epithelial cells. (CRITICAL: All filtration steps are to be done by gravity flow or centrifuge at 160 x g at RT for 1 min. Avoid using vacuum suction as this suck's in smaller epithelial cells through the 20 µm & 10 µm pluriStrainer. Check under an optical microscope after each filtration process to ensure removal of larger cell and food particles. If necessary, filter again. Check cell viability and density after each filtration step. The cell viability tends to increase with reduction of epithelial cells, which are mostly dead and add to reduced cell viability before 20 and 10 µm filtration).

8) Count the cells to check the viability and cell density by staining 10 µL aliquot of filtered cell suspension with trypan blue (10 µL) and loading into Countess cell counting chamber slides and using Countess automated cell counter. (Expected total number of cells obtained from a healthy donor in 50 mL oral rinse of resting saliva is ~1.8 x 10^5 cells having viability of 70-80 %. This procedure provides > 95 % enriched saliva neutrophils). (CRITICAL: Proceed immediately to the downstream experimental step requiring unfixed cells. Otherwise, for downstream experiments requiring fixed cells, proceed with cell fixation by 4 % PFA).? TROUBLESHOOTING

3.3. FACS sorting • TIMING: 2-3 hrs

(CRITICAL: The FACS equipment for sorting should be prepared and kept ready beforehand to minimize the exposure of saliva neutrophil in its non-native environment.) [**NOTE**: All provided timing is for single 384-well plate]

On verifying the desired purity (i.e. > 95 %) by observing under optical microscope (Fig. 2) and determining > 70 % cell viability by trypan blue exclusion method (using Countess cell counter) of enriched neutrophils (Supplementary Fig. 5), proceed immediately with FACS sorting. As we performed unbiased cell sorting, we did not stain the cells with any antibody/marker and used the cell suspension in saline solution for FACS sorting. Alternatively, cell pellets can be resuspended in $PBS^{(-)}$. (OPTIONAL: Cell viability and purity can be checked by flow cytometry using viability dye.)

9) In a 1.5-ml Eppendorf tube prepare the desired volume of lysis buffer by adding the reagents as in **section 2.5.2.** and then place it on ice. (CRITICAL: For each experiment, the lysis buffer should be made fresh.)

10) Prepare 384-well thin-walled PCR plates for sorting by adding 2 µL of lysis/FACS sort buffer to each well, except NTC-control well(s). (CRITICAL: Maintain the lysis buffer added plates at 4 ℃ using chill blocks)

11) Prepare the FACS instrument with 100/130-micron nozzle size for daily FACS setup, testing, and droplet delay optimization, plate targeting. (CRITICAL: Failure to optimize the droplet breakoff may sort satellite droplet instead of the droplet of interest by placing a charge on the satellite. Follow the FACS manufacturer's instructions for the droplet stream optimization for timing delay.)

12) Using the FACSDiva software, prepare the gating strategy for doublet discrimination gating to prevent the sorting of cell doublets or multiple cell clumps. Load a small amount of sample into the instrument to confirm the set gating and rearrange the gates if needed. Adjust the voltage of the instrument for each channel if needed.

13) Confirm the FACS setting parameters for single-cell sorting by targeting the plate using 10-µm yellow fluorescent polystyrene microspheres or similar fluorescent beads and observing under a fluorescent microscope for accurate targeting. (CRITICAL: Achieving an accuracy of a minimum 95 % single microsphere sorting is recommended. To obtain this we suggest practice sorts for single microsphere before actual sort day.)

14) Proceed with FACS sorting for a single cell. The overall event rate for 100-micron nozzle is kept at 1000–2,000 events per second on the FACS instrument (minimum 1200 events for 130-micron nozzle setup). Sort 1 cell in each except the Control wells (CRITICAL/OPTIONAL: Final confirmation of single-cell sorting can be performed by sorting a single cell in a slide and observing under microscope.)? TROUBLESHOOTING

15) Seal the sorted plates with MicroAmp Thermo-Seal lid, and immediately proceed with lysis and reverse transcription. **PAUSE POINT**: Otherwise, immediately freeze the plate on dry ice for storage at −80 °C.

3.4. cDNA synthesis by Smart-seq2 • TIMING: 1 day all steps

We performed cDNA synthesis by using modified Smart-seq2 protocol, previously published by our team. Cell lysis, cDNA synthesis and Nextera XT library preparation can be performed using any of the currently available methods for single cells [39–41]. All liquid dispensing steps are performed using the BioCel-1200 system incorporated with Bravo and BioRapTR fluidics systems. For dispensing of master mixes by BioCel we used Half-reaction volumes that have been shown below.

3.4.1. Single cell lysis • TIMING: ~15 min

16) Perform cell lysis on each single cell well by adding lysis mix containing 25 µM Oligo dT (0.25 µL) and 25 mM dNTP mix (0.25 µL) to each reaction mix (see **section 2.5.3**). To the control wells, add 10 pg/µL UHR (1 µL) to UHR-Control well; nuclease-free water (2 µL) to each NTC-control well; nothing to ERCC spike-in control wells. (CRITICAL: If FACS sorted plates are taken out of -80 ℃ storage, thaw the plates on ice in chill blocks.)

17) Denature by incubating at 72 ℃ for 3 min and immediately putting the plate on ice.

18) Centrifuge at 700 x g for 10 sec at RT to spin down the samples to the bottom of the well. Immediately put the plate back on ice. At this step, the oligo-dT primer is hybridized to the poly(A) tail of mRNA strands.

3.4.2. Reverse transcriptase (RT) reaction • TIMING: 3 hrs

19) Add reagents for RT-reaction as on **section 2.5.4** to each well containing 2.50 µL of lysed cell soup (STEP 18) by adding 3 µL of RT-master mix.

20) Perform first-strand cDNA synthesis of RT-reaction in Thermocycler by following reaction cycle:

42 ℃ for 90 min : RT and template-switching

10 cycles of

50 ℃ for 2 min : RNA-secondary structure unfolding

42 ℃ for 2 min : Completion of RT and template-switching

70 ℃ for 15 min : Final heat inactivation of enzyme

4 ℃ hold : Temperature for safe storage.

3.4.3. PCR-preamplification • TIMING: 3 hrs

21) Prepare PCR-preamplification master mix for cDNA synthesis by addition of ISPCR primer as in **section 2.5.5.**

22) Run PCR-preamplification reaction in a thermocycler by using the following reaction cycle:

98 ℃ for 3 min : Denaturation

21 cycles of

98 ℃ for 20 min : Denaturation

67 ℃ for 15 sec : Annealing

72 ℃ for 6 min : Extension

72 ℃ for 5 min : Final extension

4 ℃ for Infinite : HOLD temperature

PAUSE POINT: PCR run plate can be stored at -20 ℃ for short term or at -80 ℃ for long term storage

3.4.4. PCR Purification of cDNA synthesis product • TIMING: ~45 min

23) Using BRAVO protocol, perform the purification of the cDNA synthesis product by adding AMPure XP beads (1:1 ratio) to the RT-reaction mix from above. Incubate the mix for 5 min at RT and then place it on a magnetic rack for 2 min. Carefully remove the supernatant by pipetting and wash the beads twice with 80 % ethanol (molecular biology grade) for 30 sec. Dry the beads on a magnetic rack for 10 sec. Finally, elute the biotinylated-cDNA with 12.5 µL of Low Tris-EDTA (TE) buffer (10 mM Tris + 0.1 mM EDTA) by incubating for 10 min at RT followed by 2 min on a magnetic rack. (CRITICAL: Low TE buffer and AMPure XP beads should be at RT.)

24) Collect purified cDNA by pulling the supernatant onto a newly labeled thin-wall PCR plate. **PAUSE POINT**: Seal the plate and store it at -80 ℃. Otherwise, proceed with the QC step for analysis of cDNA quality as in the next step.

3.5. Quality control analysis of purified cDNA • TIMING: 1-4 hrs

(CRITICAL: Quality of purified cDNA library can be analyzed by four methods: Agilent high-sensitivity DNA chips, PicoGreen dsDNA assay, qRT-PCR for expression of housekeeping gene, and qRT-PCR for expression of cell-type specific gene.) Agilent high-sensitivity DNA chips are used for a few randomly picked cDNA samples and analyzed on the 2100-Bioanalyzer system to check the fragment size of DNA. PicoGreen dsDNA analysis is performed on the whole plate to accurately quantify the cDNA concentration in each well. Finally, TaqMan assay is performed in qRT-PCR to check the expression of housekeeping genes such as β-actin (ACTB) to make sure that each well has an eukaryotic cell. (OPTIONAL: TaqMan assay for a known cell-type specific marker gene(s) can be used to confirm the target cell sorted into each well. We didn’t use any neutrophil specific marker as we performed unbiased sorting.)

3.5.1. QC1: Quality check of cDNA library by Agilent high-sensitivity DNA kit • TIMING: 1 hr

25) cDNA library size distribution and quality are checked by Agilent high-sensitivity DNA chips on 2100-Bioanalyzer system for randomly picked cDNA samples from the plate by following the manufacturer's protocol for Agilent high-sensitivity DNA kit.

26) Undiluted cDNA sample (1.0 µL) is loaded on each of the Agilent high-sensitivity DNA chip and then run on the 2100-Bioanalyzer system to obtain the raw data.

27) The sample run data of each chip was analyzed by 2100 Expert software to obtain the electropherogram. Sample free of < 500 bp fragments and showing a peak at ~1.5-2 kb is considered a good library.? TROUBLESHOOTING

3.5.2. QC2: Picogreen dsDNA quantitation assay • TIMING: 2 hrs

28) Quant-iT PicoGreen dsDNA assay kit (Invitrogen, cat. no. P11496) is used for quantification of dsDNA following the manufacturer's protocol. Working concentration of PicoGreen solution from stock concentration is prepared by 200-fold dilution in TE-buffer (1X) as in **section 2.5.6** and 24.50 µL is dispensed in each well of a 384 Black flat bottom plate using BioRapTR.

29) On the other hand, prepare Lambda (λ) gDNA standard of 10 ng/µL working concentration from the provided stock solution in TE-buffer (1X). Prepare λ-gDNA of varying concentrations by serial dilution (10.00 ng/µL, 5.00 ng/µL, 2.50 ng/µL, 1.25 ng/µL, 0.625 ng/µL, 0.3125 ng/µL, 0.15625 ng/µL, and 0.00 ng/µL) to obtain the standard curve, which is used to determine the cDNA concentration of each sample well by plotting the RFU (Relative Fluorescence Unit) value of each sample in the standard curve.

30) To each of the reaction well transfer 0.50 µL of purified cDNA samples using Bravo to obtain a final concentration of 1:50 dilution. To the standard wells, by using a pipette manually load 0.50 µL of Lambda (λ) gDNA standard of varying concentration prepared by serial dilution (see **section 2.5.7**).

31) Seal the plate, mix the reaction components and centrifuge briefly to bring the reaction mix to the bottom.

32) Incubate for 2-5 min at RT by protecting from light.

33) Read plates in FlexStation 3 or any available Fluorescent microplate reader using standard Fluorescein wavelengths of Ex/Em of 480/520 nm

34) The RFU-values obtained from each sample well is plotted against the standard curve generated from the lambda-DNA standards (obtained by serial dilutions) to obtain the cDNA concentration of each single-cell well and the control wells (i.e. NTC, ERCC and UHR controls) by using the SoftMax Pro Software for FlexStation 3 (Molecular Devices).

35) The SoftMax Pro software generated file is saved in .txt format. The cDNA concentration for each sample well is then incorporated in the final project template file in .xlsx format. Samples with DNA concentration > 0.30 ng/µL are considered good suitable for downstream processes.? TROUBLESHOOTING

3.5.3. QC3: qRT-PCR for housekeeping gene expression • TIMING: 4 hrs

36) For qRT-PCR TaqMan assay, first dilute the cDNA sample to 1:10 in a new 384 well FrameStar plate by adding 9 µL Low TE buffer or nuclease free water to 1 µL of cDNA using Bravo.

37) Prepare required volume of qRT-PCR master mix as in **section 2.5.8** to be dispensed to each well of the PCR-plate by BioRapTR. (CRITICAL: Consider the “dead volume” of BioRapTR while calculating the total required volume of qRT-PCR Master mix.)

38) Dispense 7.50 µL RT-Master mix with BioRPTR using designated Reservoir and Tip to each well.

39) Using BRAVO, add 2.5 µL of diluted cDNA template of each sample for a total reaction volume of 10 µL/well. For NTC-control wells, added 2.5 µL of nuclease-free water.

40) Load the reaction plate on qRT-PCR machine (QuantStudio 6 Flex Real-Time PCR system) and run the following reaction cycle:

95 ℃ for 2 min : Denaturation

50 cycles of

95 ℃ for 10 sec : Denaturation

60 ℃ for 30 sec : Annealing

4 ℃ for Infinite : HOLD temperature

41) On completion of the qRT-PCR reaction cycles, analysis of the generated raw data is done by QuantStudio Real Time PCR Software V1.3. The "CT Settings” under the "Analysis” is changed from its “Default Settings” by changing the ‘Threshold' to 0.01 and 'Baseline Start and End' to 2 and 10 respectively. Then “Analysis Settings" is applied to obtain the final Ct-values.

42) The Ct-values are then exported in the .XLS format to be incorporated in the final project template file. Samples with Ct-values < 35 is considered good quality and suitable for downstream processes.

3.5.4. QC4: qRT-PCR for target cell specific gene expression • TIMING: 4 hrs

(OPTIONAL: Expression of genes known to be expressed in the target cell sorted can be checked by qRT-PCR to confirm the single cell sorted in each well.)

43) Taqman assay or preferred qRT-PCR method targeting the marker gene specific to the target cell sorted for the study can be performed to verify the single cell sorted in each well

3.5.5. cDNA-library plate preparing by HitPicking of QC-Pass wells • TIMING: 2 hrs

44) The cDNA concentration obtained from PicoGreen assay (From STEP 35) along with the Ct-values for ACTB expression obtained from qRT-PCR Taqman Assay (From STEP 42) for each sample well in sample plate is pasted on the project template file. Project template file is an excel file prepared to keep track of each sample well of each sort plate.

45) The samples having cDNA concentration > 0.3 ng/µL and ACTB Ct-values < 35 are considered double QC-Pass and selected for Hamilton transfer to a new cDNA-library plate.

46) For Hamilton transfer, a “Hamilton input file” in .CSV format is generated for each library plate of each donor (viz. H-SN1_Lib#1 etc.). This file is loaded in the “VENUS” software and Hamilton transfer protocol is run after placing the desired tips and plates at their designated location set on the protocol.

47) On completion of the Hamilton transfer to combine HitPicked cDNA-library plate from two/three single-cell cDNA plates, the plates are sealed, barcoded and stored at -80 ℃ for downstream processes.

3.6. Illumina Nextera XT Library preparation of HitPicked cDNA-library plates:

Illumina Nextera XT library is prepared for the HitPicked cDNA library plate by using ‘Nextera XT DNA library preparation kit’ and each sample is barcoded by using ‘Nextera XT index kit Set A and Set D’ following the manufacturers protocol. We used 1/8th reaction protocol for automated/robotic dispensing system, where 1/8th the volume of each reagent is used as mentioned in manufacturers protocol for 96-well reaction plate. The required target DNA quality for Nextera Library preparation is 1 ng of input DNA with 260/280 ratio of 2.0 - 2.2

3.6.1. Normalization of cDNA library plate • TIMING: ~1 hr

48) Before starting the Nextera XT 1/8th reaction protocol, ‘Normalized cDNA Library’ plate is prepared to obtain 0.2 ng/µL cDNA concentration in all wells.

49) cDNA library plate stored at -80 ℃ is taken out and thawed on a chill block in ice.

50) From each sample well 1 µL of cDNA is transferred to a new Framestar 384-plate by BRAVO and the desired volume of low TE-buffer is dispensed by BioRapTR to obtain 0.2 ng/µL cDNA concentration in each well. Care is taken so that total volume per well should not exceed 100 µL. If the calculated volume for any well exceeds 100 µL, the final dispensing volume is calculated for 100 µL.

3.6.2. Nextera XT Tagmentation reaction • TIMING: 10 min

51) 0.625 µL of diluted cDNA (0.2 ng/µL) from cDNA normalization plate is added to 1.250 µL of Tagment DNA Buffer (TD, 2X) and 0.625 µL of Amplification Tagment Mix (ATM) in a Framestar 384-well

microplate to obtain 2.5 µL total tagmentation reaction mix volume.

52) The plate is sealed, mixed by brief centrifuge and loaded on the thermocycler to run the tagmentation reaction by incubation at 55 °C for 10 min.

53) On completion of the reaction, immediately add 0.625 µL of NT buffer to neutralize the Tagmentation reaction to obtain 3.125 µL Total Neutralized Tagmentation volume/well.

3.6.3. Nextera XT PCR reaction with Set A and Set D barcoding kits • TIMING: 1 hr

54) To the 3.125 µL of Tagmentation volume, 1.875 µL NPM PCR master mix and 1.250 µL of Index Primer mix (0.625 µL Index Primer i5 + 0.625 µL Index Primer i7) is added to obtain total volume of 6.25 µL Nextera PCR reaction/well.

55) Seal and centrifuge FrameStar Plate at 4 °C, 500 x g (2,000 RPM) for 30 sec to mix, keep on ice till running the thermocycler reaction.

72 °C for 3 min : Extension

95 °C for 30 sec : Denaturation

16 cycles of

95 °C for 10 sec : Denaturation

55 °C for 30 sec : Annealing

72 °C for 60 sec : Extension

72 °C for 5 min : Final extension

4 °C for Infinite : HOLD

PAUSE POINT: Seal the plate and store at -80 °C until ready for library purification and cleanup.

3.6.4. Nextera XT Library purification and cleanup • TIMING: ~45 min

56) Purify each sample individually as is Step 23-24 but use 0.9:1 ratio of AMPure XP beads to Nextera library (i.e. 5.625 µL beads + 6.25 µL of Nextera Library). Elute with 6.25 µL of Low TE buffer into a new FrameStar plate "Purified Nextera XT".

PAUSE POINT: Seal the plate and store at -80 °C until ready for PicoGreen QC or normalization of the purified library.

3.6.5. QC5: Nextera XT Library QC by PicoGreen assay • TIMING: 2 hrs

57) Use 1 µL of the purified Nextera XT reactions for Picogreen dsDNA assay as in Step 28-35.

3.6.6. Normalization of Nextera XT Library plate • TIMING: ~1 hr

58) Prepare 1.0 ng/µL normalization plate of purified Nextera library based on the PicoGreen quantification above.

59) Using BioRaPTR dispense the desired amount of Low-TE buffer to a new Framestar plate and then using BRAVO to add 1 µL of purified Nextera XT library sample to obtain 1.0 ng/µL purified Nextera XT sample in each well.

3.6.7. 16 sample pooling of NexteraXT samples for MiSeq run • TIMING: 30 min

(OPTIONAL: As the sequencing of NovaSeq run for 384-plex pool library is expensive, we sequenced randomly picked 16-plex pool library from each NexteraXT library plate in MiSeq-Nano to confirm the sequence quality, determine the coverage needed and required depth of the transcriptome. We used Illumina MiSeq Reagent kit v2 (300 cycle) for the MiSeq-Nano low output run.)

60) Pool 16 samples (from Step- 59) by pipetting 3 µL of all normalized NexteraXT samples (1.0 ng/µL) for a 3 ng pool into a 1.5 mL Eppendorf LoBind tube.

61) Reverse pipette to determine the total volume and then add 90 % of that volume AMPure beads to purify the MiSeq-pooled library as in Nextera XT library cleanup on Step 56.

62) Elute with Low-TE buffer using 10-fold lower volume than the original volume of pooled library determined by reverse pipetting.

63) For QC check on Agilent high-sensitivity DNA chips, prepare 1:1 dilution (take out 1 µL), 1:10 dilution (1 µL library + 9 µL Low TE), and 1:20 dilution (2 µL of 1:10 dilution + 2 µL of Low-TE)

64) Run each of the 3 samples in triplicate on Agilent high-sensitivity DNA chip. Calculate the average fragment size and the average pool concentration in pmol/L (pM) and nM of all replicates from DNA chip run report.

PAUSE POINT: Store the MiSeq-pooled library at -20 °C until ready for sequencing run.

3.6.8. 384 sample pooling of NexteraXT samples for NovaSeq run • TIMING: 30 min

65) For each 384-well library plate, combine all 384 normalized Nextera XT libraries from each well to a single well by using BRAVO. Pipette the 384-plex pool to an Eppendorf tube and label the tube with

sample and library name (e.g. H-SN1_Lib#1, H-SN2_Lib#1, and H-SN3_Lib#1 in this study).

3.6.9. Pooled Nextera XT Library Cleanup • TIMING: ~45 min

66) Clean the 384-plex pool Nextera XT library by AMPure XP bead purification as in Step 23-24 using manual protocol for fewer pooled library sample tubes. BRAVO protocol for AMPure XP bead purification can be used for more samples.

3.6.10. QC6: QC check of 384-plex Nextera XT Library (Agilent DNA Chip) • TIMING: 1 hr

67) Use 1 µL of the 384-plex pool NexteraXT Library samples to check the quality on 2100-Bioanalyzer using Agilent high-sensitivity DNA chips. The DNA concentration of the sample is assumed to be within the range recommended by Agilent for the high-sensitivity DNA chips.

68) The Bioanalyzer report for each 384-plex pool library is saved and needed to submit samples to the sequencing core.

PAUSE POINT: The Pooled library can be stored at -20 °C for short term and at -80 °C for long term till ready for sequencing run.

3.6.11. QC7: KAPA Library Quantification Kits - Illumina • TIMING: 2-3 hrs

69) Calculate and prepare volumes enough for 3 replicates of NTC, each sample dilutions, and each standard.

70) Six pre-diluted DNA Standards of concentration 20, 2, 0.2, 0.02, 0.002, 0.0002 pM respectively are provided in the kit.

71) Before starting, prepare four different dilution sets of the pooled NexteraXT library (from Step 66) by adding required volume of dilution solution (10 mM Tris-HCL + 0.05 % Tween 20) as in **section 2.5.9**.

72) Prepare qPCR master mix as in **section 2.5.10** by combining the 1 mL of Illumina Primer Premix (10X) and the 5 mL bottle of KAPA SYBR Fast qPCR master mix (2X) provided in the kit after thawing properly. Vortex briefly to mix well and store at -20 °C till ready to use.

73) For half reaction volume, add 6 µL of qPCR master mix and 4 µL of sample or standards to appropriate well. Add 4 µL nuclease free water to NTC wells

74) Seal plate, gently vortex to mix and spin down sample to bottom of the well. Protect plate from light until ready to run.

75) Run the sample plate on QuantStudio 6 Flex or any other qPCR machine by selecting the “Standard Curve” experimental method and “SYBR Green” detector. Run the instrument in “Fast” mode using the

following thermocycler protocol:

95 °C for 5 min	: Denaturation
35 cycles of	
95 °C for 20 sec	: Denaturation
60°C for 45 sec	: Annealing
4 °C for Infinite	: HOLD

76) On completion of qPCR run, analyze the data and evaluate the Slope and R2. If the auto set Ct-threshold if the acceptable range of slope (-3.58 to -3.10) and/or R2 (~0.99) is not obtained, manually adjust Ct by setting threshold of 0.2 and set start cycle to 2 and end cycle to 3.

77) The average Ct value of each DNA Standard is plotted against its known concentration (pM) to generate a standard curve which is used to determine the concentration (pM) of diluted libraries. Finally, the working concentration of each library is calculated from the concentration of diluted libraries.

78) To further check the fragment size, the qPCR amplified product from three replicate wells were combined and ran on 1.2 % E-Gel (Invitrogen, cat. no. A03076) for 30 min using 1 kb plus ladder (Invitrogen, cat. no. 10787018) and 6X loading dye (Promega, cat. no. G190A).

79) On verification of the quality and fragment size of the pooled NexteraXT library samples, proceed to the cDNA sequencing step.

3.7. cDNA Sequencing

3.7.1 cDNA Sequencing: kit selection, run parameters, and yield • TIMING: ~24 hrs

80) The purified pooled-NexteraXT library is subjected to paired-end sequencing on a suitable Illumina NGS platform (MiSeq, HiSeq 2500, NextSeq 500, and NovaSeq 6000) with the aim to generate 1-2 million reads per sample having a read length of 100-150 bases. The sequencing data generated by the HGS platform is in fastq format.

The following sections have been briefly explained in this paper. For details on “RNA-seq analysis” of the fastq data files, please refer to Step 26 of previous publication from our group [24].

3.7.2. RNA-seq analysis: sequence quality assessment and preprocessing • TIMING: Variable

81) *Sequence quality assessment*: Sequence quality is assessed by evaluating the fastq sequence files (from Step 80) from each cell (i.e. single cell saliva neutrophil) using the fastQC tool for sequence yield, base quality, GC profile, k-mer distribution, contamination and other desired parameters.

82) *Sequence duplication*: Sequence duplication is determined in the input data. Tools such as fastx_collapser are used to calculate the absolute number of identical reads (i.e. duplicates) in the input sample fastq sequences (from Step 80). Use correct base quality score offset (-Q). Process multiple files by repeating each sequence file at a time, as the program accepts only one sequence file as input.

83) *Sequence trimming*: Sequence trimming of input paired-end fastq reads (from Step 80) is performed by Trimmomatic program to remove adapter/primer sequences and low-quality end bases.

3.7.3. RNA-seq analysis: sequence mapping and gene expression analysis • TIMING: Variable

84) *Prepare the reference genome*: Prepare the reference genome index for alignment using the build function in HISAT2 program and the reference genome fasta file. Here we used GRCh38 downloaded from Ensembl /Gencode.

85) *Calculating expression values*: Calculate the gene expression values (transcripts per million or TPM) by mapping the paired-end reads that passed trimming (from Step 83) to the reference index using HISAT2 and then evaluating the alignments using StringTie to estimate levels of expression per gene models in the annotation file. Here we used gencode.v25.annotation.gtf.

86) *Calculate and plot overall mapping statistics*: Calculate the number of reads that are mapped to the genome, to the ERCC spike-in transcripts as well as that remained unmapped using SAM tools.

Troubleshooting

Time Taken

Steps 1–3, Oral rinse collection: 1–1.5 hr

Steps 4–8, Saliva Neutrophil Isolation/enrichment: 1–1.5 hr

Steps 9–15, FACS sorting: 2–3 hr

Steps 16-18, Single cell lysis: ~15 min

Steps 19-20, Reverse transcriptase (RT) reaction: 3 hr

Steps 21-22, PCR-preamplification: 3 hr

Steps 23-24, PCR Purification of cDNA synthesis product: ~45 min

Steps 25–43, QC analysis of purified cDNA: 1-4 hrs depending on QC method(s) chosen

Steps 25–27, QC1 (OPTIONAL) - Agilent high-sensitivity DNA kit: 1 hr

Steps 28–35, QC2- Picogreen dsDNA quantitation assay: 1 hr

Steps 36–42, QC3- qRT-PCR for housekeeping gene expression: 4 hr

Step 43, QC4 (OPTIONAL) - qRT-PCR for target cell specific gene expression: 4 hr

Steps 44–47, cDNA-library plate preparing by HitPicking of QC-Pass well: 2 hr

Steps 48–56, Illumina Nextera XT Library preparation of HitPicked cDNA-library plates: ~3 hr

Steps 48–50, Normalization of cDNA library plate: 1 hr

Steps 51-53, Nextera XT Tagmentation reaction: ~10 min

Steps 54-55, Nextera XT PCR reaction: 1 hr

Steps 56, Nextera XT Library purification and cleanup: ~45 min

Steps 57, QC5: Nextera XT Library QC by PicoGreen assay: 2 hr

Steps 58-59, Normalization of Nextera XT Library plate: 1 hr

Steps 60-65, (OPTIONAL) 16 sample pooling of NexteraXT samples for MiSeq run: 30 min

Step 66, Pooled Nextera XT Library Cleanup: ~45 min

Steps 67-68, QC6: 384-plex Nextera XT Library QC by Agilent DNA Chip: 1 hr

Steps 69-79, QC7: KAPA Library Quantification Kits - Illumina: 2-3 hr

Steps 80, cDNA Sequencing: kit selection, run parameters, and yield: Variable

Steps 81-83, RNA-seq analysis: sequence quality assessment and preprocessing: variable

Steps 84-87, RNA-seq analysis: sequence mapping and gene expression analysis: variable

Anticipated Results

6. Anticipated Results

This protocol enables isolation of immune cells and enrichment of human primary salivary neutrophils, for cell isolation, flow cytometry analysis, sorting and scRNA-seq workflow (Fig. 1). Our protocol shows that the repertoire of a myeloid derived cell can be evaluated at a single cell level after saliva collection. The use size exclusion allowed > 98 % pure enriched neutrophils with viability compatible with the protocols. In order to develop the protocol that is consistent, oral samples were collected from healthy subjects throughout the project for all the experiments including microscopy (Fig. 2), flow cytometry (Fig. 3 and Fig. 4), and scRNA-seq (Fig. 5 and 6). We found that neutrophils were sensitive to cold temperature and physical stress employed in cell isolation procedures for NGS workflow. Chen et. al has shown that the integrity of total RNA is a critical parameter for RNA-seq analysis and degraded RNA heavily influences the gene expression profiles [35]. Here we have shown that neutrophil's RIN (RNA Integrity Number) values [42] decreases by one with every passing hour and after 4 hours of harvest the RNA is degraded enough (RIN < 3) which is not suitable for transcriptomic profiling, though they do not lose their viability completely in the time period (Supplementary Fig. 4). After RNA sequencing, our liberal filtration criteria of raw data allowed to obtain transcriptomic signature of neutrophils similar to other cell types and identification of eight novel sub-populations of neutrophils from healthy human saliva as compared to four sub-populations identified by stringent filter criteria typically used for analysis of RNA-seq data (Fig. 5 and Fig. 6). In addition, both manual gating analysis and the FLOCK-based automated gating analysis of the flow cytometry data confirmed diverse neutrophil subpopulations and a more stringent analysis revealed five sub-populations based on markers and density (Fig. 4).

Through microscopy, saliva presented four major different cell types (i.e. epithelial cells, neutrophils, monocytes, and lymphocytes) were initially identified based on histochemical morphology by cytospin and GIEMSA staining. For verification of the purity of neutrophils the enriched samples from healthy donors were also stained (Fig. 2B) and quantified from a minimum of 10 slides per donor which shows ~98 % neutrophil purity. This was confirmed by immunofluorescence of CD66b, CD14, CD19 markers (Fig. 2A and 2B). Our robust holotomography microscopy preserves live cells while imaging. At least 45 images per sample were taken, and cell count was represented by percentage of the total count. This morphological data was also compared through flow cytometry. While epithelial cells are quite abundant, their morphology was distinct from immune cells. In our serial filtration strategies, epithelial cells did not show significant reduction after 40 or 20 µm filtration (p = ns), but were significantly reduced after 10 µm filtration (40 versus 20 µm, p = 0.0037; 40 versus 10 µm, p = 0.0197). Similarly, neutrophil enrichment was feasible after 40 µm and 20 µm filtered samples (p = ns), reaching significant increase abundance after 10 µm filtration (p = 0.0025). This was also true for 10 µm filtered samples when compared to 40 µm

filtered samples (p = 0.0233). No significant increase or decrease has been observed in monocyte, lymphocyte, and unidentified cell types (Fig. 2C). Morphological analysis was also confirmed by immunofluorescence analysis.

We further compared gating strategies through viability, expression marker and cell size. When utilizing gating for Aqua LIVE/DEAD fixable dye (Fig. 3A and 3C) neutrophils were positive in high abundance, but minimum monocyte marker expression is detachable. We hypothesized that by gating the cells on their size we would further exclude the other immune cell. High-SSC gating demonstrated high purity of neutrophils and exclusion of other immune cells (Fig. 3B and 3D). Comparatively, live-cell gating and high-SSC gating were quantified and plotted for the type of cells (Fig. 3E and 3F). Thus, the employment of high-SSC high-FSC gating to sort unlabeled cells allows for selection of viable cells, prevents activation and cell death of oral neutrophils, minimizing the possibility of sorting monocytes or lymphocytes.

Flow cytometry analysis (Fig. 3) revealed that oral neutrophils were positive for specific neutrophil markers (CD15, CD66b, CD11b) and negative for monocyte and lymphocytes markers (CD14, CD19, CD3). In fact, when staining for most common blood immune-cell markers i.e. CD11b, CD66b, CD15, CD14, CD19, and CD18 oral neutrophils were positive with different levels of expression. Oral neutrophils were identified on $CD14^{-}CD19^{-}CD3^{-}CD15^{+}CD66b^{+}CD11b^{+}$Aqua- (Fig. 4) whereas monocytes and lymphocytes were identified for $CD14^{+}$CD11b+ and CD19+CD3+ respectively in live-cell and high-SSC gated populations (Fig. 3C and 3D). $CD15^{+}$ neutrophil density increases from 38.30 % of live-cell gated population for 40 µm filtered un-enriched samples to 60.72 % in 10 µm filtered enriched neutrophil. In addition, this increase in neutrophil was also seen to be associated with more $CD14^{+}$ monocytes in 10 µm filtered samples (Fig. 3C). When using the high-SSC gating (instead of live-cell gating), an increase of $CD15^{+}$ neutrophil density from 30.10 % of high-SSC gated population in 40 µm filtered sample to 47.35 % in 10 µm filtered enriched sample. Though this increase in density of neutrophil is not very high, the $CD14^{+}$ monocytes were completely eliminated on 10 µm filtered enriched samples (Fig. 3D). Similar to monocytes, $CD3^{+}$ T cells and $CD19^{+}$ B cells were totally eliminated even in 40 µm filtered samples when selected for high-SSC gate instead of live-cell gate. Similarly, when the frequency of each immune-cell types in the live-cell gated population (Fig. 3E) and high-SSC gated population (Fig. 3F) compared to the total ungated cell population, we see an increase in the neutrophil population and decrease in monocytes, B cells and T cells.

To characterize heterogeneity of neutrophils in a data-driven way, we employed an unsupervised data clustering method[24] (FLOCK, http://immportgalaxy.org) to understand cell phenotype differences and cell surface markers expression levels (Fig. 4). We first applied the FLOCK method to identify cell populations

from the 40 μm filtered sample, before applying the identified cluster centroids across all 6 samples for cross-sample identification and comparison of the 23 cell populations. The advantage of unsupervised clustering analysis is being data-driven, without requiring or being limited by predefined cellular phenotype. Therefore, manual annotation of each identified data cluster for identifying the cellular phenotype and interpreting the phenotype difference between the identified cell populations is usually required. Percentages of identified cell populations, mean fluorescence intensities (MFI) for each marker, expression profiles across the markers (levels 1 to 4, from negative, low, positive to high) as well as 2D dot plots of samples with cell populations highlighted in different colors are automatically generated or calculated by the FLOCK procedure. Heatmap of MFI (Fig. 4A) and bar graphs of population percentages (Fig. 4B) are commonly used to visualize the characteristics of the FLOCK-identified cell populations. Based on the MFI heatmap in figure 4A, one can easily identify the 5 salivary neutrophil subpopulations by selecting those with the phenotype of high-FSC/high-SSC/ $CD11b^{+}CD14^{-}CD15^{+}CD66^{+}CD3^{-}CD19^{-}$. From the FACS Aria II data, the FLOCK-identified neutrophil phenotype included:

- Pop10: $CD14^{-}CD19^{-}CD3^{-}$CD15+$CD66b^{+}CD11b^{+}$Aqua-$CD18^{-}$
- Pop12: CD14intCD19-CD3-$CD15^{+}CD66b^{+}CD11b^{+}$Aqua-$CD18^{-}$
- Pop14 and Pop15: CD14intCD19-$CD3^{-}$CD15hiCD66bhi$CD11b^{+}$Aqua-CD18int
- Pop17: CD14intCD19intCD3-CD15hiCD66bhiCD11b+Aqua-CD18int

When examining the percentage values of the 5 neutrophil subpopulations, we found that the abundant/known neutrophil subpopulations (Pop10 and Pop12, with different expression levels of CD14) increased frequency as the filtration size decreased, which confirmed the finding in Fig. 3. However, FLOCK also identified three other rare salivary neutrophil subpopulations (Pop14, Pop15, and Pop17) that were not in the region of the “classical” neutrophils. These rare neutrophil subpopulations have larger size and complexity (based on FSC.A and SSC.A, Fig. 4C) as well as slightly higher expression on CD15/C66b/CD18 than the classical neutrophils (Fig. 4C). The frequencies of these rare neutrophil cell populations did not increase as the filtration size decreased. Limited by the small number of markers measured in flow cytometry, this finding further emphasizes the necessity of performing a single cell RNA-seq assay to elucidate transcriptional profiles of these interesting neutrophil subpopulations.

Because we determined the optimal gating for oral neutrophils that maintain their viability and purity, we pursued sorting single neutrophils for RNA-seq. After raw sequencing files were demultiplexing using Illumina barcodes, and processed[29], we further evaluated expression values for non-control cells were

imported into Scanpy for PCA, UMAP, and cluster analysis [31]. The liberal filtering 8 neutrophil subpopulations and top 2500 variable genes were selected (Fig. 5). In fact, half of the subpopulations presented differential expression of surface markers when compared to other clusters. In contrast the strict filtering showed 4 subpopulations and top 2000 variable genes were selected (Fig. 6). Unsupervised clustering determined the differential expression of each cell population provided individual single cells grouped into cell type clusters, and a set of sensitive and specific marker genes. Logistic regression showed the comparison of gene levels in each cluster (Fig.5B and 6B). Oral neutrophils highly expressed the following gene markers CD11c, CD14, CD16a, CD16b, CD32, CD55, CD62L, CD141, Lysozyme, BEST1, FTH1, with moderately levels of CD10, CD11b, CD18, CD31, CD50, CD63, CD85D and low levels of CD11a, CD13, CD19, CD43, CD170, CD172A, CHEMR23 (Fig. 5E and 6E). These gene signatures are important to understand future neutrophil functions.

Previous single-cell sequencing of immune cells was unsuccessful in generating desirable transcriptomic profile/signature of neutrophils among PBMCs [43]. We have evaluated a dataset that is publicly available (Supplementary Fig. 6). Evaluation following our oral immune-cell pipeline applied to the publicly available dataset accession number GSE64655**.** Similar to oral neutrophils, blood derived neutrophils presented strong RNA expression levels of CD55, CD16b, CD15, and not CD66b. Blood neutrophils however presented much higher expression levels on CD10, CD11a and CD11b. We believe that the cell collection procedures of PBMCs such as the Ficoll-Hypaque method are not suitable for neutrophils as the NDNs found in the granulocyte layers at the interface of red blood cells and the gradient layer are excluded during the sample collection. In health and disease this is extremely important. SLE is characterized by neutrophil subsets known as low-density granulocytes (LDGs). When compared to other immune cell subtypes, LDGs showed the highest expression of interferon gamma genes, CD10 with subpopulations that were specifically positive correlation with disease severity and coronary patterns [44].. In addition, neutrophils patterns were also described in the emerging COVID-19 infectious disease. Bronchoalveolar single cell analysis demonstrated that myeloid cells such as macrophages were highly inflamed expressing high levels of CD68, while neutrophils from infected lungs were highly expressed (FCGR3B or CD16b) [45]. Interestingly, similar to lung neutrophils, oral neutrophils highly expressed Cd16b as part of our cell clustering system. Whereas similar to SLE cells, oral cells expressed moderate levels of CD10, revealing that oral neutrophils present shared markers with systemic cells. Future studies should implement similar protocols for systemic neutrophils compared to oral cells to uncover methods and heterogeneity.

Altogether, we developed a protocol to survey neutrophil cell heterogeneity by improving isolation methods, flow cytometry evaluation, single-cell RNA extraction, sequencing and bioinformatic pipeline. Our findings suggest novel transcriptomic signatures with identification of novel sub-populations. By combining flow cytometry with machine learning, we validated our model and gating strategies which led

to exclusion of other immune cells, while enriching for neutrophils. Ultimately, this methodology advances methods to understand oral immune cell landscape and heterogeneity.

References

14. References:

1. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, (2017).

2. Mott, M. C., Gordon, J. A. & Koroshetz, W. J. The NIH BRAIN Initiative: Advancing neurotechnologies, integrating disciplines. *PLoS Biol.* **16**, e3000066 (2018).

3. Papalexi, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2018).

4. Adlung, L. & Amit, I. From the Human Cell Atlas to dynamic immune maps in human disease. *Nat. Rev. Immunol.* **18**, 597–598 (2018).

5. Chen, H., Ye, F. & Guo, G. Revolutionizing immunology with single-cell RNA sequencing. *Cell. Mol. Immunol.* **16**, 242–249 (2019).

6. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, (2017).

7. Zhu, Y. P. *et al.* Identification of an Early Unipotent Neutrophil Progenitor with Pro-tumoral Activity in Mouse and Human Bone Marrow. *Cell Rep.* **24**, 2329–2341.e8 (2018).

8. Rincón, E., Rocha-Gregg, B. L. & Collins, S. R. A map of gene expression in neutrophil-like cell lines. *BMC Genomics* **19**, 573 (2018).

9. Lavin, Y. *et al.* Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).

10. Lakschevitz, F. S. *et al.* Identification of neutrophil surface marker changes in health and inflammation using high-throughput screening flow cytometry. *Exp. Cell Res.* **342**, 200–209 (2016).

11. Kolaczkowska, E. & Kubes, P. Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* **13**, 159–175 (2013).

12. Kubes, P. The enigmatic neutrophil: what we do not know. *Cell Tissue Res.* **371**, 399–406 (2018).

13. Hesselink, L. *et al.* Neutrophil heterogeneity and its role in infectious complications after severe trauma. *World J. Emerg. Surg.* **14**, 24 (2019).

14. Hidalgo, A., Chilvers, E. R., Summers, C. & Koenderman, L. The Neutrophil Life Cycle. *Trends Immunol.* **40**, 584–597 (2019).

15. Zilionis, R. *et al.* Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity* **50**, 1317–1334.e10 (2019).

16. Chen, J. *et al.* PBMC fixation and processing for Chromium single-cell RNA sequencing. *J. Transl. Med.* **16**, 198 (2018).

17. Ley, K. *et al.* Neutrophils: New insights and open questions. *Science Immunology* vol. 3 eaat4579 (2018).

18. Silvestre-Roig, C., Fridlender, Z. G., Glogauer, M. & Scapini, P. Neutrophil Diversity in Health and Disease. *Trends Immunol.* **40**, 565–583 (2019).

19. Scapini, P., Calzetti, F. & Cassatella, M. A. On the detection of neutrophil-derived vascular endothelial growth factor (VEGF). *J. Immunol. Methods* **232**, 121–129 (1999).

20. Monaco, G. *et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep.* **26**, 1627–1640.e7 (2019).

21. Silvestre-Roig, C., Hidalgo, A. & Soehnlein, O. Neutrophil heterogeneity: implications for homeostasis and pathogenesis. *Blood* **127**, 2173–2181 (2016).

22. Lakschevitz, F. S., Aboodi, G. M. & Glogauer, M. Oral neutrophils display a site-specific phenotype characterized by expression of T-cell receptors. *J. Periodontol.* **84**, 1493–1503 (2013).

23. Cianga, C. M., Antohe, I., Zlei, M., Constantinescu, D. & Cianga, P. Saliva leukocytes rather than saliva epithelial cells represent the main source of DNA. *Revista Romana de Medicina de Laborator* vol. 24 31–44 (2016).

24. Krishnaswami, S. R. *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016).

25. Qian, Y. *et al.* Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data. *Cytometry B Clin. Cytom.* **78 Suppl 1**, S69–82 (2010).

26. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

27. Aevermann, B. *et al.* PRODUCTION OF A PRELIMINARY QUALITY CONTROL PIPELINE FOR SINGLE NUCLEI RNA-SEQ AND ITS APPLICATION IN THE ANALYSIS OF CELL TYPE DIVERSITY OF POST-MORTEM HUMAN BRAIN NEOCORTEX. *Pac. Symp. Biocomput.* **22**, 564–575 (2017).

28. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11**, 41–46 (2014).

29. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

30. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).

31. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

32. Hoek, K. L. *et al.* A cell-based systems biology assessment of human blood to monitor immune responses after influenza vaccination. *PLoS One* **10**, e0118528 (2015).

33. Qian, Y. *et al.* FCSTrans: an open source software system for FCS file conversion and data transformation. *Cytometry A* **81**, 353–356 (2012).

34. Wohnhaas, C. T. *et al.* DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. *Sci. Rep.* **9**, 10699 (2019).

35. Chen, E. A. *et al.* Effect of RNA integrity on uniquely mapped reads in RNA-Seq. *BMC Res. Notes* **7**, 753 (2014).

36. DeLuca, D. S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).

37. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).

38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

39. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).

40. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).

41. Kim, J. & Daadi, M. M. Single-Cell Library Preparation of iPSC-Derived Neural Stem Cells. *Methods Mol. Biol.* **1919**, 129–143 (2019).

42. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**, 3 (2006).

43. Ng, L. G., Ostuni, R. & Hidalgo, A. Heterogeneity of neutrophils. *Nat. Rev. Immunol.* **19**, 255–265 (2019).

44. Mistry, P. *et al.* Transcriptomic, epigenetic, and functional analyses implicate neutrophil diversity in the pathogenesis of systemic lupus erythematosus. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25222–25228 (2019).

45. Liao, M. *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* (2020) doi:10.1038/s41591-020-0901-9.

Acknowledgements

12. Acknowledgments:

This work was supported in part by US Public Health Service grant DE025383 from the National Institutes of Dental and Craniofacial Research and by the J. Craig Venter Institute. This work was supported in part by the Chan–Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation (2018-182730) and the National Center for Advancing Translational Sciences (U01TR001801).

13. Competing interests:

The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ManuscriptNeutrophilSingleCellSequencingNPFreire1.docx
- ManuscriptNeutrophilSingleCellSequencingNPFreire1.docx
- ManuscriptNeutrophilSingleCellSequencingNPFreire31.pdf
- ManuscriptNeutrophilSingleCellSequencingNPFreire31.pdf