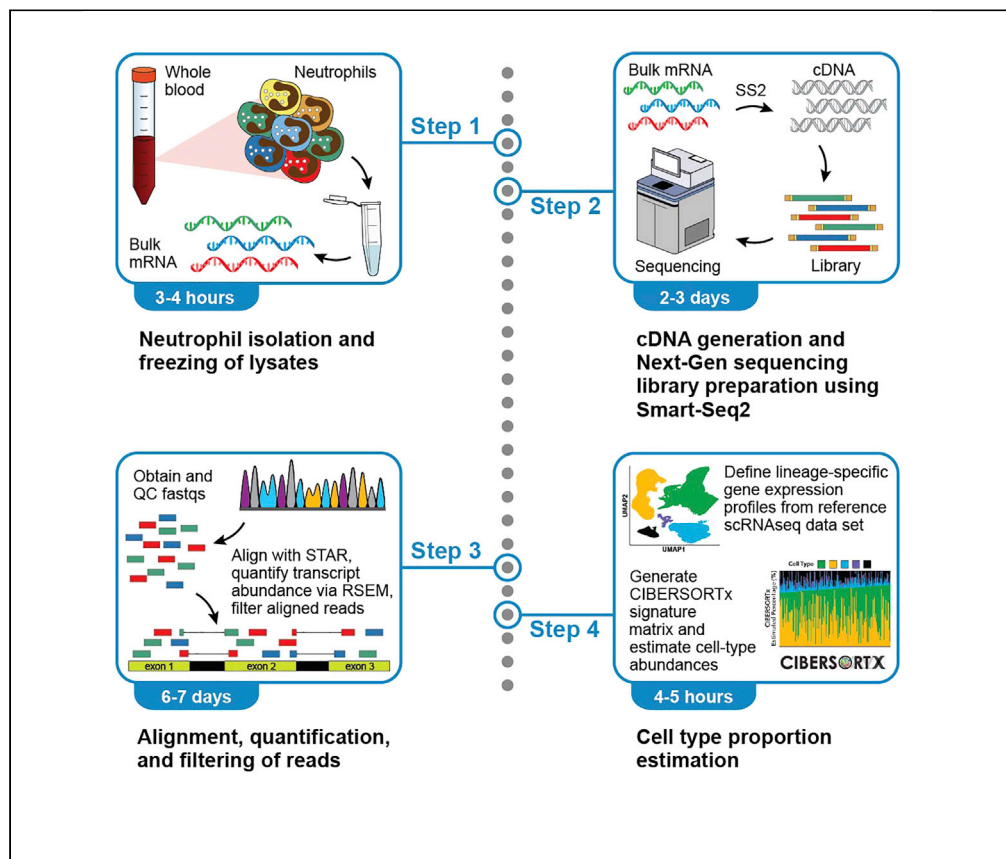## Protocol

Protocol for bulk RNA sequencing of enriched human neutrophils from whole blood and estimation of sample purity

Anna L.K. Gonye, Thomas J. LaSalle, Samuel S. Freeman, Miguel Reyes, Nir Hacohen, Alexandra-Chloé Villani, Moshe Sade-Feldman

agonye1@jhmi.edu (A.L.K.G.)
msade-feldman@mgh.harvard.edu (M.S.-F.)

Highlights

Streamlined negative selection enrichment of neutrophils from whole blood

Bulk purification of neutrophil RNA and high-quality, optimized library preparation

Quality control filtration and cell type abundance estimation

Although neutrophils are the most abundant leukocyte in healthy individuals and impact outcomes of diseases ranging from sepsis to cancer, they remain understudied due to technical constraints of isolation, preservation, and sequencing. We present a modified Smart-Seq2 protocol for bulk RNA sequencing of neutrophils enriched from whole blood. We describe steps for neutrophil isolation, cDNA generation, library preparation, and sample purity estimation via a bioinformatic approach. Our approach permits the collection of large cohorts and enables detection of neutrophil transcriptomic subtypes.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

STAR Protocols

Protocol

Protocol for bulk RNA sequencing of enriched human neutrophils from whole blood and estimation of sample purity

Anna L.K. Gonye,[1,2,3,8,10,*] Thomas J. LaSalle,[1,2,4,8] Samuel S. Freeman,[1,2] Miguel Reyes,[2,5,9] Nir Hacohen,[1,2] Alexandra-Chloé Villani,[1,2,6,7] and Moshe Sade-Feldman[1,2,7,11,*]

[1]Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Cellular and Molecular Medicine Graduate Program, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[4]Program in Health Sciences and Technology, Harvard Medical School & Massachusetts Institute of Technology, Boston, MA, USA

[5]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[6]Center for Immunology and Inflammatory Diseases, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[7]Department of Medicine, Harvard Medical School, Boston, MA, USA

[8]These authors contributed equally

[9]Present address: Department of Infectious Diseases, Genentech, South San Francisco, CA, USA

[10]Technical contact

[11]Lead contact

*Correspondence: agonye1@jhmi.edu (A.L.K.G.), msade-feldman@mgh.harvard.edu (M.S.-F.)
https://doi.org/10.1016/j.xpro.2023.102125

SUMMARY

Although neutrophils are the most abundant leukocyte in healthy individuals and impact outcomes of diseases ranging from sepsis to cancer, they remain understudied due to technical constraints of isolation, preservation, and sequencing. We present a modified Smart-Seq2 protocol for bulk RNA sequencing of neutrophils enriched from whole blood. We describe steps for neutrophil isolation, cDNA generation, library preparation, and sample purity estimation via a bioinformatic approach. Our approach permits the collection of large cohorts and enables detection of neutrophil transcriptomic subtypes.

For complete details on the use and execution of this protocol, please refer to LaSalle et al. (2022)[1] and Boribong et al. (2022).[2]

BEFORE YOU BEGIN

The protocol herein describes the specific steps for isolating neutrophils via negative selection for bulk RNA-sequencing. This technique could be extended to other cell types of interest by exchanging the commercially available negative selection cell type enrichment kit, though we have not tested the purity of other negative selection kits. It is important to consider that circulating human neutrophils have a typical half-life of 13–19 h, which may be further shortened in disease states.[3]

Therefore, it is critical to begin the isolation and storage steps as soon as possible following sample collection. All steps of the neutrophil isolation protocol should be performed at 25°C unless otherwise indicated. Finally, neutrophils may be activated both by surface binding of antibodies and centrifugation.[4,5] We avoid surface binding by using negative selection and ensure consistency in our centrifugation steps to avoid introducing artifacts between batches.

Institutional permissions

This protocol involves the use of human immune cells, and as such, the study must be approved by the governing institution. This study was approved by the Mass General Brigham Institutional Review Board under protocol 2017P001681 in Boston, Massachusetts, USA. As the study was initially designed for the isolation of neutrophils from COVID-19 patients during March-May 2020, samples were collected under a waiver of informed consent approved by the Massachusetts General Hospital governing institutional review board, in compliance with the Code of Federal Regulation (45CFR 46, 2018 Common Rule). Under all other circumstances, informed consent is required. Those intending to use the protocol will need to acquire permissions from the relevant institutions.

Prepare of Nextera index plates

⏱ Timing: 1 h

Prior to beginning any other wet-lab work in this protocol, researchers should first generate the four 96-well index plates (designated Q1-Q4) required for the construction of a 384-sample library.

1. Order index primers (listed in Table S1, sheet ''Primers_to_order'') through IDT using the following specifications:
   a. 100 nmole DNA oligo.
   b. Standard Desalting LabReady (normalized to 100 μM in IDTE pH 8.0).
2. Once received, index primers should be stored at −20°C.
3. Using the index plate layouts (Table S1, sheet ''i5_i7_index_and_layout,'' rows 119–161), label four 96-well plates Q1 through Q4.
4. Prepare index primers.
   a. Thaw index primers (100 μM).
   b. Vortex for 3 s and spin 10 s.
5. Arrange the index primer tubes in the appropriate order for allocation into the specified rows and columns for each index plate.
   a. Each index plate will have a different set of multiplexed index primers to create a total of 384 unique index combinations.
6. Distribute 5 μL of each index primer into each well of its corresponding row/column.
7. Add 40 μL of TE buffer to achieve a total volume of 50 μL per well (5 μL i5 primer, 5 μL i7 primer, 40 μL TE).
   a. This will result in a final concentration of 10 μM of each index primer.
8. Prepare index plates for storage.
   a. Seal the plates with foil plate seals.
   b. Vortex them for 3 s.
   c. Spin plates in plate centrifuge for 1 min at 800 × g at 25°C.
9. Store plates at −20°C until use.

Prepare the software and computing environment

⏱ Timing: 2 h

Note: Our analyses requiring high computational power (e.g., alignment and expression quantification) are conducted on Terra, the Broad Institute's cloud computing platform. The user should ensure they have sufficient computational resources for the scale of their project.

Note: To ensure uniformity of results, we utilize the Docker image "gcr.io/broad-cga-francois-gtex/gtex_rnaseq:V10" when performing computational analyses and follow the steps in the Broad Institute GTEx/TOPMed RNA-Seq pipeline with custom modifications. The Docker image contains all the code, libraries, and other files needed to run the pipelines described in the

protocol and can be built and run as a Docker container to seamlessly perform the necessary analyses. The user should ensure that they can run code in this environment. For more information on using Docker, see https://docs.docker.com/get-started/.

Prepare the alignment and quantification reference files

🕐 Timing: 40 min

Various reference files are needed to align the FASTQ files containing the sequenced reads to the genome and quantify transcript abundance. Some can be downloaded from online sources, and others will need to be generated based on the parameters of the experiment.

10. Download a copy of GRCh38 and remove the ALT, HLA, and Decoy contigs from the FASTA, as the pipeline is not well-suited for these regions.
    a. Utilize the following python code from the GTEx/TOPMed RNA-seq pipeline.

```
>with open('Homo_sapiens_assembly38.fasta', 'r') as fasta:

> contigs = fasta.read()

>contigs = contigs.split('>')

>contig_ids = [i.split(' ', 1)[0] for i in contigs]

>

>filtered_fasta = '>'.join([c for i,c in zip(contig_ids, contigs)

>  if not (i[-4:]=='_alt' or i[:3]=='HLA' or i[-6:]=='_decoy')])

>

>with open('Homo_sapiens_assembly38_noALT_noHLA_noDecoy.fasta', 'w') as fasta:

>  fasta.write(filtered_fasta)
```

11. Generate the FASTA index and dictionary from the FASTA file with the exclusions from the previous substep.

```
>samtools faidx Homo_sapiens_assembly38_noALT_noHLA_noDecoy.fasta

>java -jar picard.jar \

>  CreateSequenceDictionary \

>  R=Homo_sapiens_assembly38_noALT_noHLA_noDecoy.fasta \

>  O=Homo_sapiens_assembly38_noALT_noHLA_noDecoy.dict
```

12. Download the gtf file for the reference annotation, GENCODE 34, from the GENCODE ftp (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_34/gencode.v34.annotation.gtf.gz).

    Note: Researchers may also choose to use more recent GRCh38 annotation sets.

13. Generate a collapsed reference annotation for gene-level quantification from the reference gtf file, as the previously-downloaded reference gtf is separated into transcript isoforms.

    a. Utilize the "collapse_annotation.py" script from the GTEx/TOPMed RNA-seq pipeline available at https://github.com/broadinstitute/gtex-pipeline/blob/master/gene_model/collapse_annotation.py.

```
>python3 collapse_annotation.py \
>  -collapse_only gencode.v34.GRCh38.annotation.gtf \
>  gencode.v34.GRCh38.genes.collapsed_only.gtf
```

14. Generate the STAR index using overhang length of (read length - 1).
    a. For a read length of 150 base pairs, choose an overhang of 149, as shown below.

```
>STAR -runMode genomeGenerate \
>  -genomeDir >STARv275a_genome_GRCh38_noALT_noHLA_noDecoy_v34_oh149 \
>  -genomeFastaFiles Homo_sapiens_assembly38_noALT_noHLA_noDecoy.fasta \
>  -sjdbGTFfile gencode.v34.GRCh38.annotation.gtf \
>  -sjdbOverhang 149 -runThreadN 10
```

    ⚠ CRITICAL: The STAR index must be generated using the correct overhang length. Since the read length varies between sequencers, this must be correctly specified.

15. Generate the RSEM reference.

```
>rsem-prepare-reference -num-threads 10 \
>  -gtf gencode.v34.GRCh38.annotation.gtf \
>Homo_sapiens_assembly38_noALT_noHLA_noDecoy.fasta \
>  rsem_reference
```

16. Extract the transcript/gene symbol to Ensembl ID mapping from BioMart.

    Note: Depending on the research question, it may be of interest to search for the presence of RNA from viruses or other microorganisms in the sample (e.g., SARS-CoV-2 RNA in studies of COVID-19). The genome FASTA of interest may be appended to the human genome and reference annotation. When collapsing the reference annotation for gene-level quantification, this should be done separately for each reference, then merged as below.

```
>cat gencode.v34.GRCh38.annotation.gtf SARSCoV2.gtf \
>  > gencode.v34.GRCh38.annotation.SARSCoV2.gtf
>cat gencode.v34.GRCh38.genes.gtf SARSCoV2.genes.gtf \
>  > gencode.v34.GRCh38.SARSCoV2.genes.gtf
```

Prepare the deconvolution reference dataset and signature matrix
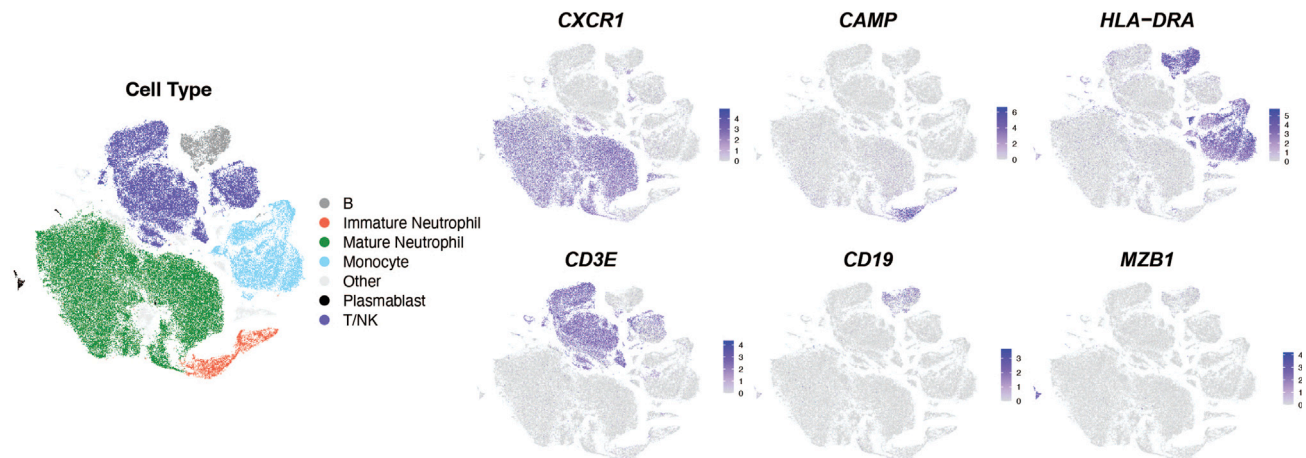
⏱ Timing: 2–6 h

Figure 1. Uniform Manifold Approximation and Projection (UMAP) plots of Schulte-Schrepping et al. Cohort 2 fresh whole blood scRNA-seq data
Shown on the left are suggested cell lineage assignments, and key marker genes of each lineage are shown to the right.

In this protocol, CIBERSORTx,[6] a digital cytometry method, is used to estimate the cell type proportions in the enriched neutrophil samples. The deconvolution algorithm uses a cell type-specific reference gene expression matrix from single-cell RNA-sequencing (scRNA-seq) data of individual cell types to infer the composition of a bulk sample based on weighted scores of expressed genes from the reference matrix. Therefore, when choosing or generating a reference scRNA-seq dataset, the reference must contain all possible cell types that may be present in the bulk RNA-Seq sample being deconvoluted. If there are cell types in the bulk RNA-Seq samples that are not present in the reference dataset, then the deconvolution will not be accurate.

> *Note:* Context matters; for example, severe COVID-19 is characterized by an increase in immature neutrophils,[7] so immature neutrophils must be present in the reference scRNA data. If they are missing from the reference, their contributions to the bulk expression profile will be misattributed to other cell types. As another example, eosinophil counts in circulating whole blood are quite low (typically less than 1%), so the algorithm will not be able to detect meaningful variation and they may be excluded from the reference at the expense of potentially missing samples with rare eosinophilia.

> ⚠ CRITICAL: Gene expression profiles can change drastically from one tissue to another and between diseases and healthy states. Therefore it is critical that the scRNA-seq data includes healthy controls and samples across the spectrum of disease anticipated in the study.

17. Download an appropriate scRNA-seq reference dataset.
    a. In this case, we use the Schulte-Schrepping Cohort 2 fresh whole blood scRNA-seq dataset,[7] which can be downloaded from https://beta.fastgenomics.org/datasets/detail-dataset-1ad2967be372494a9fdba621610ad3f3#Files.
    b. This dataset includes all relevant circulating blood cell types (including immature neutrophils) from healthy controls and COVID-19 patients (Figure 1).
18. Using the cluster annotations from the publication dataset, assign single cells to their broad lineage.

> *Note:* We reassign the more specific cluster labels to the following categories: Mature neutrophil, Immature neutrophil, Monocyte, T/NK (combined), B, Plasmablast, Other. "Other"

encompasses cell types with lower frequencies that are not expected to significantly contribute to contamination of neutrophil samples due to their low frequency in whole blood samples.

```
>prefix <- "~/Downloads/COVID19_Neutrophil_Code/" #Adapt as necessary

>seuratwb <- readRDS(paste0(prefix,"seurat_COVID19_freshWB-PBMC_cohort2_rhapsody_jonas_FG_2020-08-18.rds"))

>barcodecluster <- as.character(seuratwb$RNA_snn_res.0.8)

>temp <- seuratwb$RNA_snn_res.0.8

>names(barcodecluster) <- names(temp)

>barcodecluster <- cbind(rownames(barcodecluster),barcodecluster)

>barcodecluster <- as.data.frame(barcodecluster)

>barcodecluster$barcode <- as.character(rownames(barcodecluster))

>barcodecluster$barcode[barcodecluster$barcodecluster %in% c("0","1","4","8")] <- "Mature_Neutrophil"

>barcodecluster$barcode[barcodecluster$barcodecluster %in% c("12","15")] <- "Immature_Neutrophil"

>barcodecluster$barcode[barcodecluster$barcodecluster %in% c("3","10","14","18")] <- "Monocyte"

>barcodecluster$barcode[barcodecluster$barcodecluster %in% c("2","5","6","13","16")] <- "T_NK"

>barcodecluster$barcode[barcodecluster$barcodecluster %in% c("7","22")] <- "B"

>barcodecluster$barcode[barcodecluster$barcodecluster %in% c("19")] <- "Plasmablast"

>barcodecluster$barcode[barcodecluster$barcodecluster %in% c("9","11","17","20","21","23","24")] <- "Other"

>seuratwb <- AddMetaData(seuratwb, metadata = subset(barcodecluster, select = c("barcode")), col.name = "Cell.Type")
```

Note: For CIBERSORTx to be able to extract cell type-specific signatures from the single cell data, all the counts from a given cell type for each separate patient must be summed. This is known as "pseudobulking" because it is similar to a bulk RNA-seq sample of that given cell type.

19. To begin pseudobulking, first initialize a matrix with one row per gene and one column per cell type per donor (in this case, we have six cell types and thus there will be six times the number of patients as columns).

```
>pseudobulk <- matrix(0L, nrow = nrow(seuratwb@assays$RNA@counts), ncol = (length(levels
(factor(seuratwb$donor))))*6)

>rownames(pseudobulk) <- (seuratwb@assays$RNA@data@Dimnames[[1]])

>idx <- (levels(factor(seuratwb$donor)))

>pseudobulk <- as.data.frame(pseudobulk)

>for (i in 0:(ncol(pseudobulk)/6 - 1)){

>   k <- i*6+1

>   colnames(pseudobulk)[k] <- paste(as.character(idx[i+1]),"_Mature_Neu",sep = "")

>   colnames(pseudobulk)[k+1] <- paste(as.character(idx[i+1]),"_Immature_Neu",sep = "")

>   colnames(pseudobulk)[k+2] <- paste(as.character(idx[i+1]),"_Mono",sep = "")

>   colnames(pseudobulk)[k+3] <- paste(as.character(idx[i+1]),"_TNK",sep = "")
```

```
```

```
> pseudobulk[,k+4] <- rowSums(temp@assays$RNA@counts)

> temp <- subset(seuratwb, subset = donor == as.character(idx[i+1]) & Cell.Type ==
"Plasmablast")

> pseudobulk[,k+5] <- rowSums(temp@assays$RNA@counts)

>}

>pseudobulk <- pseudobulk[rowSums(pseudobulk) > 0,]

>pseudobulk <- pseudobulk[-grep("\\.",rownames(pseudobulk)),]

>pseudobulk <- pseudobulk[-grep("^MT",rownames(pseudobulk)),]
```

21. Convert the Ensembl gene IDs to the gene symbols.

   Note: In this code example, "genepc" is the lookup table for Ensembl gene IDs and gene symbols, extracted from the annotation gtf used during the alignment. Alternatively, the user may use the function "getBM" from the R package biomaRt.

```
>genepctemp <- genepc[genepc$Gene.stable.ID %in% rownames(Count_filtered),]

>Count_temp <- Count

>for (i in 1:nrow(Count_temp)){

>  if (length(which(genepctemp$Gene.stable.ID == rownames(Count_temp)[i])) == 1){

>    if (!(genepctemp$Gene.name[which(genepctemp$Gene.stable.ID == rownames(Count_temp)
[i])] %in% rownames(Count_temp))){

>    rownames(Count_temp)[i] <- genepctemp$Gene.name[which(genepctemp$Gene.stable.ID ==
rownames(Count_temp)[i])]

>  }

>  }

>}

>pseudobulk_temp <- pseudobulk[rownames(pseudobulk) %in% rownames(Count_temp),]
```

22. Exclude columns with low gene counts in cases where samples do not contain enough of specific cell types.

   Note: In our dataset, we exclude columns with less than 5,000 counts.

23. Edit the column names such that they only represent the cell type.
    a. In this case, column names should only read "Mature_Neutrophil", "Immature_Neutrophil", "Monocyte", "T_NK", "B", "Plasmablast".
24. Generate a class labels file; this is a separate .txt file with rows corresponding to cell type and columns corresponding to the columns of the expression file.
    a. Place a "1" in rows where that column of the expression matrix matches the cell type, and place a "2" otherwise.

   Note: A portion of an example class labels table is shown below, which would correspond to an expression matrix with columns "Mature_Neutrophil", "Immature_Neutrophil", "Monocyte", "T_NK", "B", "Plasmablast", "Mature_Neutrophil".

Mature_Neutrophil	1	2	2	2	2	2	1
Immature_Neutrophil	2	1	2	2	2	2	2
Monocyte	2	2	1	2	2	2	2
T_NK	2	2	2	1	2	2	2
B	2	2	2	2	1	2	2
Plasmablast	2	2	2	2	2	1	2

25. Create an account at https://cibersortx.stanford.edu/ and log in.
    a. Upload the reference expression matrix as a "Mixture", and upload the class labels file as "Class Labels".
26. To generate the CIBERSORTx signature matrix, select "1. Create Signature Matrix", choose "Custom", and select "RNA-seq" (not "scRNA-seq", as the counts have been pseudobulked).
    a. Select the Reference sample and Phenotype classes files that have just been uploaded.
    b. Under Additional Options, set limits of 50–100 marker genes per cell type, and filter for only hematopoietic genes.
    c. Press "Run", and the signature matrix will be generated.

Note: It is important to evaluate the signature matrix by examining the heatmap of marker genes for each cell type generated by CIBERSORTx. Depending on the reference dataset or the application, it may be beneficial to adjust the number of marker genes per cell type or filter the input samples based on the results.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Whole blood healthy control samples	StemExpress	CAT# PBEDT020F
Whole blood experimental samples	N/A	N/A
Critical commercial assays		
EasySep™ Direct Human Neutrophil Isolation Kit	STEMCELL Technologies	CAT# 19666
Qubit dsDNA High Sensitivity Assay Kit	Invitrogen	CAT# Q32854
High-Sensitivity DNA Bioanalyzer Kit	Agilent	CAT# 5067-4626
Nextera XT Library Prep kit	Illumina	CAT# FC-131-1024
Oligonucleotides		
RT primer (DNA oligo)	IDT	5′-AAGCAGTGGTATCAACGCAGAGTACT30VN-3′
TSO primer (RNA oligo with LNA)	Qiagen	5′-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3′
ISPCR (DNA oligo)	IDT	5′-AAGCAGTGGTATCAACGCAGAGT-3′
Software and algorithms		
FlowJo (v10.7.1)	FlowJo, LLC	https://www.flowjo.com/solutions/flowjo
Prism 9.2.0 (283)	GraphPad	https://www.graphpad.com/scientific-software/prism/
GTEx-TOPMed RNA-Seq pipeline	Broad Institute	https://github.com/broadinstitute/gtex-pipeline
STAR v2.7.5a	Dobin et al. 2013[7]	https://github.com/alexdobin/STAR/releases/tag/2.7.5a
RSEM v1.3.0	Li et al. 2011[8]	https://github.com/deweylab/RSEM/releases/tag/v1.3.0
RNA-SeQC 2	Graubert et al. 2021[9]	https://github.com/getzlab/rnaseqc
CIBERSORTx	Newman et al. 2019[6]	https://cibersortx.stanford.edu
Seurat v4.0.4	Hao and Hao et al. 2021[10]	https://cran.r-project.org/web/packages/Seurat/index.html
Other		
EasyEights™ EasySep™ Magnet	STEMCELL Technologies	CAT# 18103
Invitrogen™ DynaMag™-96 Side Skirted Magnet	Thermo Fisher Scientific	CAT# 12027
DynaMag™-2 Magnet (1.5–2 mL tubes)	Thermo Fisher Scientific	CAT# 12321D
1.5 mL cryopreservation tubes	VWR	CAT# 66008-710

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Eppendorf centrifuge 5910R	Eppendorf	CAT# 2231000657
Eppendorf centrifuge 5430	Eppendorf	CAT# 022620513
Eppendorf Mastercycler w/ 96-well plate adapter X50s	Eppendorf	CAT# 6311000010
Eppendorf Mastercycler w/ 384-well plate adapter X50t	Eppendorf	CAT# 6306000010
Cytation 5 Plate-based spectrophotometer/fluorometer	BioTek/Agilent	CAT# CYT5M
Qubit 4 fluorometer	Thermo Fisher Scientific	CAT# Q33226
TC20 Automated Cell Counter	Bio-Rad	CAT# 1450102
2100 Bioanalyzer & chip vortex	Agilent	CAT# G2939BA
12-channel multichannel P10 pipette	Rainin	CAT# 17013807
12-channel multichannel P20 pipette	Rainin	CAT# 17013808
12-channel multichannel P200 pipette	Rainin	CAT# 17013810
accu-jet® S pipet-aid	BrandTech	CAT# 26350
DNA LoBind microcentrifuge tubes, 1.5 mL	Eppendorf	CAT# 022431021
DNA LoBind microcentrifuge tubes, 2 mL	Eppendorf	CAT# 022431048
Qubit assay tubes	Invitrogen	CAT# Q32856
20 μL low-retention pipette tips	Rainin	CAT# 30389226
200 μL low-retention pipette tips	Rainin	CAT# 30389240
20 μL pipette tips	Rainin	CAT# 30389225
200 μL pipette tips	Rainin	CAT# 30389239
1000 μL pipette tips	Rainin	CAT# 30389213
Armadillo High Performance 96-well PCR plate	Thermo Fisher Scientific	CAT# AB-2396
Clear-bottom black sided 96-well plate	Corning	CAT# 3904
Armadillo High Performance 384-well PCR plate	Thermo Fisher Scientific	CAT# AB-2384
10 mL serological pipette	Corning	CAT# 357551
2 mL aspirating pipette	Corning	CAT# 357558
Alcohol-resistant markers	VWR	CAT# 52877-355
ACK Lysing Buffer	Gibco	CAT# A10492-1
TCL buffer	Qiagen	CAT# 1031576
2-Mercaptoethanol	Gibco	CAT# 21985023
Pierceable aluminum PCR plate sealing film	Axygen	CAT# PCR-AS-200
MicroAmp Clear Adhesive Film	Applied Biosystems	CAT# 4306311
100% ethanol	VWR	CAT# 89125-170
15 mL conical tubes	Corning	CAT# 430052
UltraPure™ 0.5 M EDTA	Thermo Fisher Scientific	CAT# 15575020
1× PBS	Corning	CAT# 21-040-CV
Trypan blue stain 0.4%	Thermo Fisher Scientific	CAT# T10282
Cell counting slides for TC20 cell counter	Bio-Rad	CAT# 1450015
Agencourt RNAClean XP SPRI Beads	Beckman Coulter	CAT# A63987
Nuclease-free water	Thermo Fisher Scientific	CAT# 1097703
Recombinant RNase inhibitor	Clontech	CAT# 2313B
Magnesium chloride (100 mM)	Sigma-Aldrich	CAT# M1028
dNTPs	Thermo Fisher Scientific	CAT# R0192
5× reverse transcription (RT) buffer	Thermo Fisher Scientific	CAT# EP0753
Betaine	Sigma-Aldrich	CAT# B0300-1VL
Maxima H Minus Reverse Transcriptase	Thermo Fisher Scientific	CAT# EP0753
2× KAPA HiFi HotStart ReadyMix	KAPA Biosystems	CAT# KK2602
Agencourt AMPureXP SPRI Beads	Beckman Coulter	CAT# A63881
TE buffer	Tekanova	CAT# T0228

MATERIALS AND EQUIPMENT

Reagents/materials/equipment required for each major step

Due to the length of the protocol herein, below we have specified which elements of the key resources table are necessary for each major wet-lab element in order to streamline and ease the reader's utilization of this protocol.

Red blood cell lysis; neutrophil enrichment

The following are required for the neutrophil enrichment portion of this protocol:

- Reagents and materials required per draw.
  - 4 15 mL conical tubes.
  - Serological pipets: aspirating pipets, 10 mL.
  - 2 1.5 mL cryopreservation tubes.
  - 8 mL ACK Lysing Buffer.
  - EasySep™ Direct Human Neutrophil Isolation Kit.
  - Trypan Blue.
  - 1× PBS.
  - UltraPure 0.5 M EDTA, pH 8.
  - 200 μL TCL Buffer.
  - 2 μL 2-Mercaptoethanol.
- Required equipment.
  - Pipet-aid and serological pipets.
  - 1,000 μL pipette and tips.
  - 200 μL pipette and tips.
  - 20 μL pipette and tips.
  - EasyEights™ EasySep™ Magnet.
  - Phase-contrast microscope.
  - Hemocytometer.
  - Tally counter (hand clicker).
  - Variable temperature centrifuge capable of accommodating at least two 15 mL conical tubes, set to 25°C (25°C) (Eppendorf centrifuge 5910R).
  - Alcohol-resistant markers.

Modified Smart-Seq2 cDNA generation

The following are required for the Smart-Seq2 cDNA generation portion of this protocol:

- Reagents and materials.
  - 96-well PCR plates.
  - Low-retention 20 μL tips.
  - 200 μL tips (for ethanol washes).
  - Aluminum PCR sealing film.
  - Agencourt RNAClean XP SPRI beads.
  - Agencourt AMPureXP SPRI beads.
  - 100% Ethanol.
  - TE buffer.
  - RT primer (DNA oligo).
  - dNTPs.
  - Recombinant RNase Inhibitor.
  - Nuclease-free water.
  - 5× RT buffer.
  - Betaine.
  - $MgCl_2$.
  - TSO primer (RNA oligo).
  - Maxima H Minus Reverse Transcriptase.
  - ISPCR primer (DNA oligo).
  - 2× KAPA HiFi HotStart ReadyMix.

- ○ Qubit dsDNA High Sensitivity Assay Kit.
- ○ High-Sensitivity DNA Bioanalyzer Kit.
- Required equipment.
  - ○ 12-channel multichannel P10 pipette.
  - ○ 12-channel multichannel P20 pipette.
  - ○ 12-channel multichannel P200 pipette.
  - ○ Invitrogen™ DynaMag™-96 Side Skirted Magnet.
  - ○ Plate centrifuge (Eppendorf centrifuge 5430).
  - ○ Thermocycler w/ 96-well plate adapter.
  - ○ Reagent reservoirs.
  - ○ 2100 Bioanalyzer & chip vortex.
  - ○ Plate-based spectrophotometer/fluorometer.

Nextera library generation

The following are required for the Nextera library generation portion of this protocol:

- Reagents and materials.
  - ○ 22 boxes (plus 1 extra) of regular 20 μL tips.
  - ○ 4 boxes of low-retention 20 μL tips.
  - ○ 1 box of low-retention 200 μL low retention tips (for aliquoting reagents into reagent plate).
  - ○ Nextera XT Library Prep kit (96 samples).
  - ○ 2 96-well PCR plates.
  - ○ 384-well PCR plate.
  - ○ Aluminum PCR sealing film.
  - ○ 100% Ethanol.
  - ○ DNA LoBind tube 1.5 mL.
  - ○ DNA Lobind tube 2 mL.
  - ○ TE buffer.
  - ○ Agencourt AMPureXP SPRI Beads.
  - ○ Qubit dsDNA High Sensitivity Assay Kit.
  - ○ Qubit assay tubes.
  - ○ High-Sensitivity DNA Bioanalyzer Kit.
- Required equipment.
  - ○ 12-channel multichannel P10 pipette.
  - ○ Thermocycler w/ 96-well plate adapter.
  - ○ Thermocycler w/ 384-well plate adapter.
  - ○ Plate centrifuge (Eppendorf centrifuge 5430).
  - ○ 2100 Bioanalyzer & chip vortex.
  - ○ Qubit 4 fluorometer.
  - ○ DynaMag™-2 Magnet.

Recipes for solutions used in this protocol

Short recipes:

- **1× PBS+ 1 mM EDTA:** add 1 mL of 0.5 M EDTA to 499 mL of 1× PBS.

Store at 25°C. Can be stored indefinitely.

- **TCL+2-mercaptoethanol lysis buffer:** add 50 μL of 1% 2-mercaptoethanol to 4.95 mL TCL buffer.

Store at 25°C prior to using. Discard after 24 h.

Recipes with ≥ 3 ingredients

Smart-Seq2 Mix-1

Reagent	Final concentration	Amount (per sample)
RT primer (10 μM)	2.5 μM	1 μL
dNTPs (10 μM)	2.5 μM	1 μL
recombinant RNase inhibitor (4 U/μL)	4 U/μL	1 μL
nuclease-free water	N/A	1 μL
Total	N/A	4 μL

Prepare fresh and store on ice, maximum storage time 20 min.

Smart-Seq2 Mix-2

Reagent	Final concentration	Amount (per sample)
TSO primer (10 μM)	1.43 μM	1 μL
RT Buffer (5×)	1.43×	2 μL
betaine (5 M)	1.43 M	2 μL
MgCl$_2$ (100 mM)	12.86 mM	0.9
recombinant RNase inhibitor (4 U/μL)	0.14 U/μL	0.25 μL
Maxima H Minus reverse transcriptase (200 U/μL)	2.86 U/μL	0.1 μL
nuclease-free water	N/A	0.75 μL
Total	N/A	7 μL

Prepare fresh and store on ice, maximum storage time 10 min.

Smart-Seq2 Mix-3

Reagent	Final concentration	Amount (per sample)
ISPCR primer (10 μM)	0.36 μM	0.5 μL
KAPA HiFi HotStart ReadyMix (2×)	1.79×	12.5 μL
nuclease-free water	N/A	1 μL
Total	N/A	14 μL

Prepare fresh and store on ice, maximum storage time 30 min.

STEP-BY-STEP METHOD DETAILS
Red blood cell lysis

⏱ Timing: 10 min

A large proportion of whole blood is red blood cells (RBCs), thus the first step in isolating neutrophils is an RBC lysis (Figure 2).

1. To begin, if the anticoagulant in the vacutainer is not EDTA, add 1 μL of 0.5 M EDTA to the whole blood.
2. Prepare the following solutions (see "materials and equipment"):
   a. 1× PBS+EDTA.
   b. TCL+2-mercaptoethanol lysis buffer.
3. For each individual sample, gently open the vacutainer tube and move 500 μL of whole blood to a 15 mL conical tube.
4. Add 8 mL of ACK lysis buffer to each tube using a 10 mL serological pipet, pipetting up and down 5 times to mix the sample and lyse the RBCs.
5. Centrifuge the samples at 300 × *g* for 5 min at 25°C with max acceleration and max deceleration.
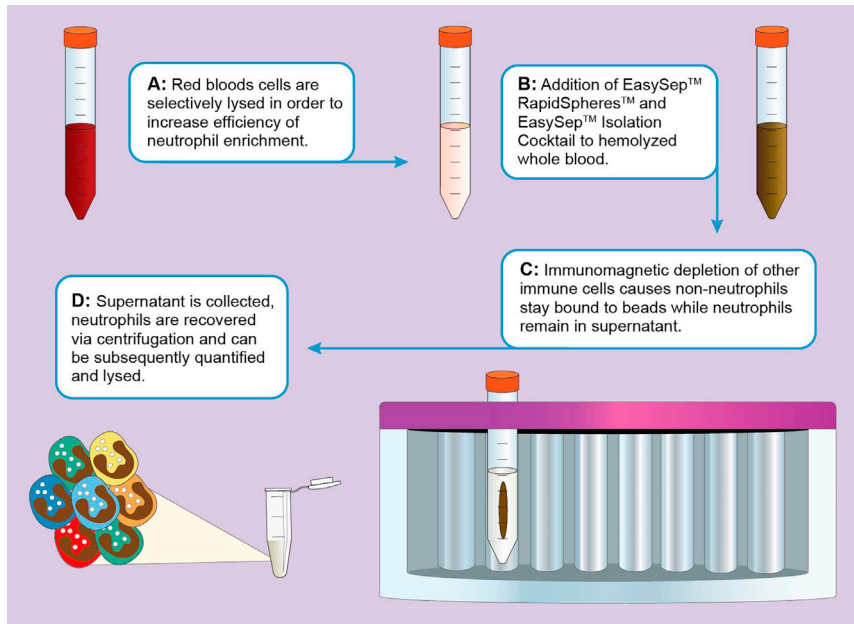6. Aspirate the supernatant, being careful not to disturb the pellet.

Figure 2. Schematic overview of the general steps involved in the enrichment of neutrophils from fresh whole blood samples

    a. The pellet contains the neutrophils and all other PMBCs.
7. Resuspend the pellet of cells in 250 μL of 25°C 1× PBS+EDTA by pipetting up and down gently 5 times.

Neutrophil enrichment

⏱ Timing: 40–60 min

Once most RBCs have been removed and the pellet is resuspended, enrichment for neutrophils via EasySep isolation can be performed (Figure 2).

8. Vortex EasySep™ RapidSpheres™ for 30 s.
9. Add 50 μL of EasySep™ Isolation Cocktail to each sample.
10. Add 50 μL of RapidSpheres to each sample.
11. Mix gently by pipetting up and down 10 times.
12. Incubate samples for 5 min at 25°C.
13. Top up the volume of each tube to 4 mL with 1× PBS+EDTA.
14. Cap all of the 15 mL conical tubes tightly.
15. Mix gently by tilting the samples back and forth 5 times.
16. Remove the lid from each 15 mL conical tube.
17. Place samples in the EasyEights™ EasySep™ Magnet, ensuring that all tubes sit all the way down.
18. Incubate samples on the magnet for 5 min at 25°C.
19. Use a 5 mL serological pipet to transfer the supernatant of the samples into a new set of labeled 15 mL conical tubes.
    a. Discard tubes containing the RapidSpheres.

Note: Angle the pipet such that the tip is opposite the magnet to avoid touching the magnetic beads. The procedure is a negative selection, meaning that neutrophils remain in the supernatant while PBMCs are captured by the RapidSpheres.

20. Add 25 μL of RapidSpheres to the tubes containing the supernatant from the first selection step.
21. Cap samples tightly and mix gently by tilting the samples back and forth 5 times.
22. Incubate samples for 5 min at 25°C.
23. Following the incubation, place tubes on the EasySep magnet and incubate for 5 min.
24. Use a 5 mL serological pipet to transfer the supernatant of the samples into a new set of labeled 15 mL conical tubes.
    a. Discard tubes containing the RapidSpheres pellet.
25. Place the samples (supernatant) back on the EasySep magnet and incubate for 5 min.

    Note: This additional incubation on the magnet is necessary in order to pull any remaining RapidSpheres from the suspension.

26. Use a 5 mL serological pipet to carefully transfer 3.5 mL of the supernatant of the samples into a new set of labeled 15 mL conical tubes.
    a. Discard tubes containing the RapidSpheres pellet.
27. Cap all samples and centrifuge at 300 × *g* for 5 min at 25°C.
28. Aspirate the supernatant, being careful not to disturb the pellet.

    Note: The cell pellet should now be highly enriched for neutrophils with minimal other cell species present.

29. Resuspend the isolated neutrophils in 1 mL 1× PBS + EDTA.
30. Mix thoroughly by pipetting up and down.

Cell counting and lysis

⏱ Timing: 5–20 min

Once isolated, neutrophils should be counted in order to determine the yield and to determine the correct volume of lysis buffer to use so that all samples are at the same concentration for bulk RNA sequencing library preparation.

31. Mix 10 μL of neutrophil suspension with 10 μL of trypan blue by pipetting.
32. Add 10 μL to a counting chamber of a hemocytometer.
33. Using a hemocytometer or an automated cell counter, determine the total number of neutrophils isolated from each sample.

    Note: if an automated cell counter is used, be sure to adjust the detection gates to count only the smaller population of cells. Larger leukocytes may still be present at low levels in the sample, and the goal is to determine the neutrophil count, not the absolute cell count of each sample.

34. Move 2×10^5 cells to a 1.5 mL Eppendorf tube by transferring an appropriate volume of the cell suspension based on the determined count/concentration.
35. Top up the tube of cells to 1.5 mL total volume by adding additional 1× PBS+EDTA.
36. Centrifuge at 300 × *g* for 5 min at 25°C.
37. Aspirate the supernatant, being careful not to disturb the pellet.
    a. Remove the last bit of supernatant with a 20–200 μL pipette.
38. Resuspend the neutrophils in 200 μL of the prepared TCL + 1% 2-mercaptoethanol lysis buffer.
    a. If a lower number of cells than desired were retrieved, resuspend the cells in a lower volume such that the final lysis concentration is 1,000 cells/μL.
39. Mix well to fully lyse all cells.
40. Transfer 100 μL per tube to two labeled cryopreservation tubes.

    a. Place tubes on dry ice immediately.

41. After 5 min, the aliquoted lysates can be moved to storage at −80°C.

    Optional: If enough of the original sample remains, it may be helpful to check for successful enrichment by flow cytometry, comparing the unenriched whole blood to the enriched neutrophil product. A suggested antibody panel for discriminating between the various PBMCs and neutrophils is suggested in LaSalle et al.[1] and includes CD3, CD14, CD16, CD19, CD20, CD56, CD66b.

Modified Smart-Seq2 cDNA generation

  ⏱ Timing: 6–8 h

Once neutrophils have been enriched, cDNA can now be prepared via a modified version of the Smart-Seq2 protocol[11] and then used for downstream library preparation. These steps involve purifying RNA from the lysed cells, reverse transcription of mRNA transcripts to cDNA, whole-transcriptome amplification, and purification of amplified cDNA.

42. Prepare neutrophil samples for cDNA generation.
    a. Bring RNA-SPRI beads to 25°C.
    b. Thaw lysed neutrophils on ice.
    c. Pipette 20 μL (at a concentration of 1,000 cells/μL) of each neutrophil lysis sample into the well of a 96-well plate on ice.
    d. Centrifuge plates at 1,500 rpm for 30 s at 4°C.
    e. Place plate(s) back on ice.
    f. Cover with the lid of a pipette tip box, until beginning the following step.
43. Prepare RNA-SPRI beads.
    a. Vortex the beads for 30 s.
    b. Use a serological pipet to transfer the necessary volume (44 μL per sample, plus excess) into a reagent reservoir.
44. Using a multichannel pipette, add 44 μL of RNA-SPRI beads (2.2 times the sample volume) to samples.
45. Mix thoroughly by pipetting up and down.
46. Cover the plate(s), then incubate for 10 min at 25°C.
    a. During this incubation, prepare 50 mL of 80% EtOH: 40 mL EtOH and 10 mL ddH$_2$O.
    b. Also prepare the appropriate volume of **Smart-Seq2 Mix-1** (see "materials and equipment"), plus excess (10%), and set aside on ice.
47. Following the incubation, place the plate on the DynaMag™-96 Side Skirted Magnet and incubate for 5 min at 25°C.
48. Remove the supernatant.
    a. Aspirate the liquid by angling the tips of the pipette opposite to the beads.

  ⚠ CRITICAL: Do not disturb the beads. RNA from the neutrophils is bound to them, and any loss of beads at this step will result in loss of genetic material input for subsequent steps.

49. Use a serological pipet to fill a reagent reservoir with 80% EtOH.
50. Wash the beads in 100 μL of 80% EtOH twice.
    a. Add the EtOH, ensuring the beads are covered.
    b. Agitate the beads, moving the plate side to side on the magnet such that the beads move from one side of the well to the other side (Figure 3).
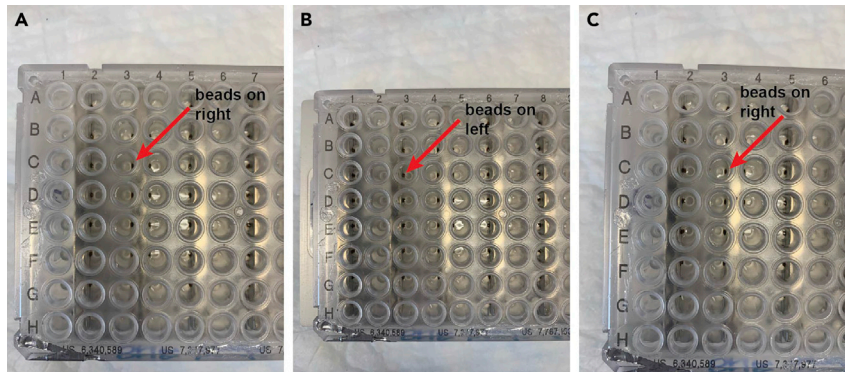
Figure 3. Smart-Seq2 cDNA generation bead washing
(A–C) As the 96-well plate is moved side to side on the side-skirted plate magnet as depicted in A–C, note how the pellets of beads move from the right side of the well to the left, and then back to the right. This movement of the beads through the ethanol allows for superior removal of non-nucleic acids and better retrieval of RNA.

> *Note:* Make sure to move in both directions so that samples in columns A and H are washed properly.

    c. Place the plate on the magnet and incubate for 30–60 s to allow all beads to come out of solution.

    d. Use a P200 pipette to remove the ethanol.

    e. Repeat steps a-d (total 2 washes).

    f. After the final removal of the EtOH with the P200, use a P20 to remove any additional EtOH left behind in the wells.

51. After complete removal of the EtOH, keep the plate on the magnet and allow the RNA-SPRI beads to dry for around 10 min, until cracks in the beads are visible.

    a. Cover the plate to keep out contaminants but skew the lid to allow for more airflow/easier evaporation, as shown in Figure 4.

> ⚠ CRITICAL: Once cracks appear, proceed to the next step. Overdrying of the RNA-SPRI beads can negatively impact sample yields.

    b. Dilute 100 μL of 1 M MgCl$_2$ in 900 μL nuclease-free water to make a 100 mM stock solution.

    c. During this incubation, prepare the appropriate volume of **Smart-Seq2 Mix-2** (see "materials and equipment"), plus excess, *without adding* Maxima H Minus reverse transcriptase.

      i. Set aside on ice.

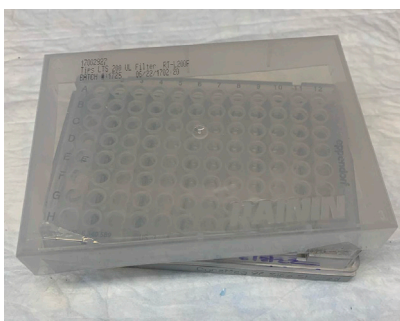   d. Start the Denaturing protocol on the thermocycler.

Figure 4. Appropriate way to cover SPRI beads during drying
The lid must be slightly skewed in order to increase airflow. If the plate is covered too tightly, the beads will dry unevenly and take too long, enabling RNA degradation.

i. Immediately press pause so that the lid and block get up to temperature.

Denaturing protocol

Step	Temperature	Time	Cycles
Denaturing of RNA	72°C	3 min	1
Hold	4°C	∞	1

52. Elute RNA from the dry RNA-SPRI beads by adding 4 μL of Smart-Seq2 Mix-1 and resuspending the samples thoroughly.
53. Seal the plate with a foil plate seal.
54. Centrifuge plate to get rid of any bubbles.
    a. Set the centrifuge to run at 300 × *g* for 1 min.
    b. Start the spin but stop it once the speed has risen to 300 × *g* to eliminate bubbles but prevent pelleting of the magnetic beads.
55. Place the samples in the thermocycler and resume the Denaturing protocol.
56. As soon as the 72°C step completes, immediately place samples on ice.
    a. Do not wait for the thermocycler to ramp down to 4°C.
    b. After placing the samples on ice, start the Reverse Transcription (RT) protocol on the thermocycler and immediately press pause to allow the block and the lid to come up to temperature.

Reverse Transcription protocol

Step	Temperature	Time	Cycles
Reverse transcription	50°C	90 min	1
Heat inactivation	85°C	5 min	1
Hold	4°C	∞	1

57. Add Maxima H Minus reverse transcriptase to Smart-Seq2 Mix-2, pipette to mix.
58. Spin down for 5 s in a benchtop microcentrifuge.
59. With the plate of samples on ice, carefully remove the plate seal and add 7 μL of Smart-Seq2 Mix-2 to each well.
    a. Mix by pipetting up and down 10 times or until RNA-SPRI beads are resuspended.
60. Re-seal the plate with a fresh foil seal.
61. Centrifuge plate to get rid of any bubbles.
    a. Set the centrifuge to run at 300 × *g* for 1 min.
    b. Start the spin but stop it once the speed has risen to 300 × *g* to eliminate bubbles but prevent pelleting of the magnetic beads.
62. Place the plate in the thermocycler and resume the RT protocol.
    a. 15–20 min before the program ends, take the KAPA HiFi HotStart ReadyMix and the ISPCR primers out of the freezer to thaw on ice.
63. When the RT program has ended, remove the samples from the thermocycler and place the plate on ice.
    a. After placing the samples on ice, start the PCR Preamplification program on the thermocycler and immediately press pause to allow the block and the lid to come up to temperature.

PCR Preamplification protocol

Step	Temperature	Time	Cycles
Initial denaturation	98°C	3 min	1
Denaturation	98°C	15 s	16
Annealing	67°C	20 s	
Extension	72°C	6 min	

(Continued on next page)

Continued

Step	Temperature	Time	Cycles
Final Extension	72°C	5 min	1
Hold	4°C	∞	1

    b. Prepare **Smart-Seq2 Mix-3** (see "materials and equipment") and place on ice.

64. Carefully remove the foil plate seal.
65. Add 14 μL of Mix-3 to samples and mix by pipetting up and down 10 times or until RNA-SPRI beads are resuspended.
66. Re-seal the plate with a fresh foil seal.
67. Centrifuge plate to get rid of any bubbles.
    a. Set the centrifuge to run at 300 × *g* for 1 min.
    b. Start the spin but stop it once the speed has risen to 300 × *g* to eliminate bubbles but prevent pelleting of the magnetic beads.
68. Place the samples in the thermocycler and resume the PCR Preamplification program.

    ⏸ **Pause point:** Samples can be held at 4°C for up to 24 h once the PCR Preamplification is completed.

    △ CRITICAL: Before moving to the next step, make sure to equilibrate the AMPure XP SPRI beads to 25°C (∼30 min).

69. Prepare AMPure XP SPRI beads.
    a. Vortex the beads for 30 s.
    b. Use a serological pipet to transfer the necessary volume (40 μL total per sample, plus excess) into a reagent reservoir.
70. Purify PCR products.
    a. Using a multichannel pipette, add 20 μL of AMPure XP SPRI beads (0.8 times the sample volume) to samples.
    b. Mix thoroughly by pipetting up and down.
71. Cover the plate(s), then incubate for 5 min at 25°C.
72. Following the incubation, place the plate on the magnet and incubate for 5 min at 25°C.
    a. During this incubation, prepare 50 mL of 70% EtOH: 35 mL EtOH and 15 mL ddH$_2$O.
73. Remove the supernatant.
    a. Aspirate the liquid by angling the tips of the pipette opposite to the beads.
74. Use a serological pipet to fill a reagent reservoir with 70% EtOH.
75. Wash the beads in 100 μL of 70% EtOH twice.
    a. Add the EtOH, ensuring the beads are covered.
    b. Agitate the beads, moving the plate side to side on the magnet such that the beads move from one side of the well to the other side.

    Note: Make sure to move in both directions so that samples in columns A and H are washed properly.

    c. Place the plate on the magnet and incubate for 30–60 s to allow all beads to come out of solution.
    d. Use a P200 pipette to remove the ethanol.
    e. Repeat steps a-d.
    f. After removing the EtOH with the P200, use a P20 to remove any additional EtOH left behind in the wells.

76. After complete removal of the EtOH, keep the plate on the magnet and allow the AMPure XP SPRI beads to dry for around 10 min, until cracks in the beads are visible.
    a. While the beads are drying, use a serological pipet to transfer an excess of TE buffer to a reagent reservoir.
    b. Additionally, prepare/label a new 96-well plate which will be used for elution.
77. Elute amplified cDNA from the dry AMPure XP SPRI beads by adding 25 µL of TE buffer and resuspending the samples thoroughly.
78. Incubate for 1 min at 25°C.
79. Place the plate on the magnet and wait until all magnetic beads have come out of solution.
    a. In most cases, this will take approximately 1 min.
80. With a pipette set to 25 µL, transfer the supernatants (containing the cDNA) to the new 96-well plate.
81. Perform a second round of purification, repeating steps 70–80, for a total of 2 cleanup cycles.
82. Store samples at 4°C for 1–2 days, or at −20°C for longer durations.

⏸️ **Pause point:** Samples can be held at 4°C for 1–2 days (or at −20°C for up to 4 weeks, without freeze/thaw) following cDNA purification and before moving on to cDNA quality control.

cDNA quality control

⏱ Timing: 1–2 h

Prior to taking the cDNA generated via this modified Smart-Seq2 protocol forward into library preparation for bulk RNA sequencing, the quality of the cDNA must be determined. We use two distinct metrics to do so: a measurement of the concentration of each individual sample using a plate-based Qubit dsDNA HS assay, and a measurement of the size distribution of cDNAs in representative samples via generation of BioAnalyzer traces.

83. Perform a Qubit assay using a Qubit™ dsDNA HS kit.
    a. Centrifuge plates of cDNA for 5 s and then place on ice.
    b. Prepare enough Qubit working solution for all of the samples you will be testing and for 8 standards (198 µL per sample/standard), plus excess.
       i. Dilute the appropriate volume of Qubit™ dsDNA HS Reagent 1:200 in dsDNA HS Buffer.
       ii. Wrap the tube in aluminum foil to protect from light, and set aside.
    c. Prepare standards.
       i. Prepare an 8-well PCR strip tube and label wells 1–8.
       ii. Add 25 µL of standard-1 (provided with the kit) into all wells except the first one.
       iii. Add 50 µL of standard-2 to the first empty well, take 25 µL, add to the second well containing 25 µL of standard-1 and mix.
       iv. Continue with the serial dilution until the 7th well, leaving the 8th well with only 25 µL of standard-1.
    d. Obtain one black-sided spectrophotometer assay plate per plate of neutrophils cDNA as well as an additional plate for the standards.
       i. The standards should be prepared in triplicate, so 24 wells (columns A-C) should be prepped.
    e. Pour the prepared Qubit working solution into a reagent reservoir.
       i. Use a multichannel pipette to add 198 µL of Qubit solution into each well that will be used to measure samples/standards.
    f. Place the plate of cDNA on the side-skirted magnet.
       i. Use a multichannel pipet to transfer 2 µL of sample into the corresponding wells in the plate containing Qubit working reagent.
    g. Transfer 2 µL of standards to the appropriate wells on the standard plate.
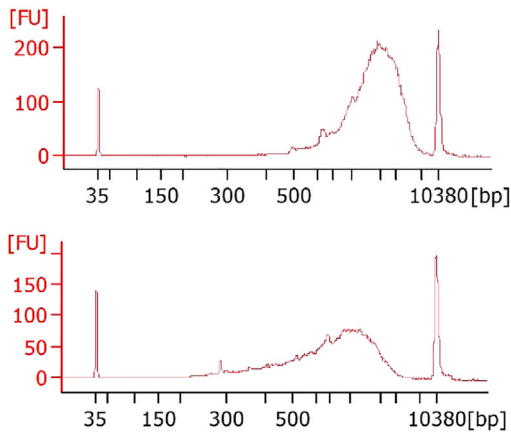   h. Seal all plates thoroughly with a plastic plate seal, vortex for 5–10 s, then spin down.

Figure 5. Example Bioanalyzer traces from two different bulk neutrophil cDNA samples following preparation via SmartSeq2 and clean-ups

Note the size distributions and peaks of both traces. The upper represents an ideal cDNA trace, with all fragments larger than 500 bp. The lower trace is still acceptable to take forward into library preparation, but the presence of cDNA fragments lower than 500 bp suggests that some mRNA degradation occurred prior to the reverse transcription reaction.

    i. Remove plate seals and read on a plate-based spectrophotometer/fluorometer.
        i. Excitation: 485 nm; emission: 530 nm.
    j. Use the standards to create a standard curve and linear formula to relate fluorescence to dsDNA concentration.
        i. Use this formula to determine the concentration of each sample.

84. Once neutrophil cDNA concentrations have been determined, select 11 samples from each plate to run on an Agilent Bioanalyzer using the High-Sensitivity DNA Bioanalyzer Kit.
    a. Choose the samples such that you will be able to observe the DNA quality over the range of concentrations measured via the Qubit assay.
    b. Prepare the Bioanalyzer chip and load the 11 samples according to the manufacturer's protocol.
    c. Example traces are shown below in Figure 5.

⚏ **Pause point:** Samples can be held at 4°C overnight or at −20°C for up to 4 weeks (without freeze/thaw) following cDNA purification and before moving on to library preparation.

Modified Nextera library generation

⏱ **Timing: 2–3 h**

Illumina's Nextera XT Library Prep Kit is a fast and easy to use library preparation method. We have modified and optimized the protocol to work with ultra-low DNA inputs of 0.1–0.2 ng in a 384-well plate format to allow for easy multiplexing. This is ideal for use with cell-types such as neutrophils that have relatively small transcriptomes and enables a low-effort way to process many samples for RNAseq library generation at one time.

85. Following library preparation, all samples should be diluted to 0.1–0.2 ng/μL.

Note: We have found that the easiest way to do this is to create a 96-well "dilution plate" that corresponds to each plate of cDNA generated via SmartSeq2.

    a. First, 10 μL of each sample determined to already have concentrations in the range of 0.1–0.2 ng/μL should be transferred to the corresponding sample wells in the dilution plate.
        i. 10 μL of any samples below the desired concentration should also be transferred at this time.
    b. All remaining samples should be diluted using ultra-pure, double-distilled water.

    c. At least 1 μL of original cDNA should be used to create the diluted sample, and the total dilution volume should be no less than 10 μL.

86. Prepare the set-up for library generation.
    a. Obtain two large ice buckets, and place a metal 384-well plate cooling block on one of them.
    b. Retrieve Tagmentation DNA Buffer (TD) and Amplicon Tagment Mix (ATM) from storage at −20°C and allow to thaw on ice.
    c. Retrieve the prepared index plates (Q1, Q2, Q3, Q4) and thaw at 25°C before placing on ice.
    d. If diluted plates were held at 4°C, place on ice.
       i. If they were stored at −20°C, allow them to thaw at 25°C and then place on ice.
    e. Retrieve a fresh 384-well PCR plate for the library generation.
    f. Obtain a 96-well plate for use as a reagent plate.
87. Add TD, ATM, and diluted samples to the 384-well PCR plate.
    a. Aliquot 55 μL of TD into all 12 wells of row A of the reagent plate.
    b. Use two tip boxes (one row of tips per row of the 384-well plate) to transfer 1.25 μL of TD into each well of the 384-well plate.
    c. Aliquot 30 μL of ATM into all 12 wells of row B of the reagent plate.
    d. Use four tip boxes (two rows of tips per row of the 384-well plate) to transfer 0.625 μL of ATM to each well of the 384-well plate.
    e. Use four tip boxes to transfer 0.625 μL of diluted samples into the 384-well plate, keeping track of the positions of samples in order to match them with the unique indices that will be added in step 93.
       i. Seal the 384-well plate with a foil plate seal and mix samples by tapping gently on the benchtop.
    f. Spin the plate at 20,000 × g for 3 min to ensure that there are no bubbles.

   ⚠ CRITICAL: Make sure that the library plate is held on ice while samples are added, and place the plate immediately back on ice after centrifugation. This will limit premature tagmentation of the cDNA.

88. Perform the tagmentation reaction.
    a. Put the plate in a thermocycler equipped with a 384-well plate block, and run the Make NTA protocol.

Make NTA

Step	Temperature	Time	Cycles
Tagmentation	55°C	10 min	1
Hold	10°C	∞	1

89. While the tagmentation reaction is occurring, prepare for the next steps.
    a. Retrieve the NT buffer (NT) from storage at 25°C, and aliquot 30 μL of TD into all 12 wells of row C of the reagent plate.
    b. Replace the TD and ATM at −20°C.
    c. Reseal the cDNA plates and store at −20°C.
90. As soon as the 55°C step of the Make NTA program finishes and the thermocycler begins to ramp down to 10°C, take the plate out and immediately place it on ice.
91. Carefully remove the foil plate seal and add NT to the tagmented samples.
    a. Use four tip boxes to transfer 0.625 μL of NT to each well of the 384-well plate.
    b. Seal the plate and mix samples by tapping gently on the benchtop.
    c. Spin the plate at 20,000 × g for 3 min to ensure that there are no bubbles.
    d. Incubate the plate for 10 min at 25°C.

    i. Begin the timer for this incubation as soon as you take the plate out of the centrifuge.

  ⚠ CRITICAL: Place the library plate back on the metal block on ice as soon as possible after the tagmentation reaction (55°C step) is complete and add the NT carefully but efficiently to prevent over-tagmentation of the cDNA.

92. While the samples are incubating with NT, prepare for the next steps.
    a. Retrieve the NPM reagent from storage at −20°C and place on ice.
    b. Ensure that the index plates have thawed.
    c. Gently vortex the index plates and then spin them down by allowing the centrifuge to reach 20,000 × g before stopping it.
93. Remove the foil plate seal, then add NPM and indices.
    a. Aliquot 80 μL of NPM into all 12 wells of row D of the reagent plate.
    b. Use four tip boxes to transfer 1.875 μL of NPM into each well of the 384-well plate.
    c. Use four tip boxes to add 1.25 μL of indices to each well of the 384-well plate. Q1 indices will be added to rows 1–4 of the 384-well plate, Q2 to rows 5–8, etc.
       i. This layout is visualized in Table S1 (sheet "i5_i7_index_and_layout").
    d. Seal the plate and mix the samples by tapping gently on the benchtop.
    e. Spin the plate at 20,000 × g for 3 min.
    f. Put the plate in a thermocycler equipped with a 384-well plate block, and run the Nextera PCR protocol.

Nextera PCR

Step	Temperature	Time	Cycles
Pre-denaturation	72°C	3 min	1
Initial Denaturation	95°C	30 s	1
Denaturation	98°C	10 s	12
Annealing	67°C	30 s	
Extension	72°C	60 s	
Final Extension	72°C	5 min	1
Hold	4°C	∞	1

94. Pool the individual samples.
    a. Once the Nextera PCR program has finished and the thermocycler is holding at 4°C, remove the place and spin it at 20,000 × g for 3 min.
    b. Using the four boxes of low-retention tips, pool together 2.5 μL of each sample from the 384-well plate into row H of the reagent plate.
    c. Seal the plate with a foil seal.
    d. Spin the plate at 20,000 × g for 1 min to eliminate any bubbles.
    e. Form the final pooled library by collecting all of the contents of row H into a 2 mL DNA LoBind Eppendorf tube.
       i. Set the pipette to 80 μL to ensure that you are collecting all of the liquid from each well of row H.
95. Perform the first purification of the cDNA library.
    a. Ensure that the AMPure XP SPRI beads are at 25°C and vortex them for 30 s to resuspend.
    b. Prepare fresh 80% EtOH.
    c. Add 0.9 volumes of AMPure XP SPRI beads to the sample (for a library generated from 384 samples the final volume will be 960 μL so 864 μL of beads would be added, for example).
    d. Mix well by pipetting up and down.
    e. Incubate bead/library mixture at 25°C for 5 min.

Note: This mixture will be viscous. Ensure that the total volume of the mixture is dispensed by pipetting slowly.

f. Incubate the sample on a DynaMag™-2 Magnet for 5 min.
    i. At the end of the 5 min, tighten the cluster of beads using the 96-well plate magnet.
g. Remove the supernatant, being careful not to disturb the beads.
h. Wash the beads twice with 1 mL of 80% EtOH, making sure that the beads are covered.
i. For each wash, rotate the Eppendorf tube at least twice on the magnet.

Note: Due to the size of the pellet, not all of the beads will move when the tube is turned. This is okay.

j. After the second wash, remove all of the EtOH. Ensure that all EtOH is removed by aspirating any final drops left in the tube with a 20 μL tip.
k. Dry the beads for 10 min, or until cracks start to appear in the cluster. During the drying period, cover the tube loosely with the lid of a pipette tip box.
l. Elute the library in 200 μL of TE buffer, remove the sample from the magnet, and pipette up and down to resuspend. Incubate for 5 min at 25°C.

⚠ CRITICAL: make sure that the magnetic beads are wet before removing the Eppendorf tube from the magnet. Not doing so can result in a loss of sample.

m. Place the sample on the magnet for 5 min.
n. Remove the supernatant (contains the cDNA library) to a new 1.5 mL DNA LoBind Eppendorf tube.
    i. Discard the 2 mL tube containing the bead pellet.

Note: It is okay if a small amount of beads are transferred with the sample.

96. Perform the second purification of the cDNA library.
    a. Add 180 μL (0.9 times the sample volume) of AMPure XP SPRI beads to the sample. Mix well by pipetting up and down, and then incubate at 25°C for 5 min.
    b. Incubate the sample on the tube magnet for 5 min. At the end of the 5 min, tighten the cluster of beads using the 96-well plate magnet.
    c. Remove the supernatant, being careful not to disturb the beads.
    d. Wash the beads twice with 1 mL of 80% EtOH, making sure that the beads are covered.
        i. For each wash, rotate the Eppendorf tube at least twice on the magnet.
    e. After the second wash, remove all of the EtOH.
        i. Ensure that all EtOH is removed by aspirating any final drops left in the tube with a 20 μL tip.
    f. Dry the beads for 10 min, or until cracks start to appear in the cluster.
        i. During the drying period, cover the tube loosely with the lid of a pipette tip box.
    g. Elute the library in 200 μL of TE buffer, remove the sample from the magnet, and pipette up and down to resuspend.
    h. Incubate for 5 min at 25°C.
    i. Place the sample on the magnet for 5 min.
    j. Remove the supernatant (contains the purified library) to a new 1.5 mL DNA LoBind Eppendorf tube.
97. Perform library quality control.
    a. Dilute the library 1:10 in water, and analyze the size distribution on an Agilent Bioanalyzer using the Agilent DNA 1000 Kit according to the manufacturer's protocol.
        i. Example traces are shown below in Figure 6.
    b. Determine the concentration of the library using a Qubit™ dsDNA HS kit.
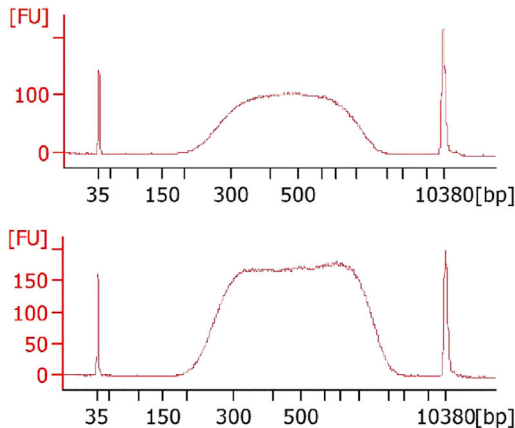
Figure 6. Example Bioanalyzer traces from two different bulk neutrophil bulk RNAseq Nextera libraries
Note the broad peak, with a broad size distribution of around 250–1,200 bp.

    i. Prepare 700 μL of Qubit working solution by diluting the appropriate volume of Qubit™ dsDNA HS Reagent 1:200 in dsDNA HS Buffer.
    ii. For the cDNA library, combine 199 μL of Qubit working solution and 1 μL of sample in a Qubit assay tube.

  ⚠ CRITICAL: Hold the tube containing the cDNA library on the tube magnet while transferring the sample. Any bead contamination will mess up the results of the Qubit assay.

    iii. Prepare standards 1 and 2 by combining 190 μL of Qubit working solution with 10 μL of the appropriate standard in Qubit assay tubes.
    iv. Make sure to vortex samples and standards well, then spin for 5 s to get rid of any bubbles.
    v. Select the dsDNA High Sensitivity program on the Qubit fluorometer.
    vi. Calibrate with standards 1 and 2, and then read the concentration of the cDNA library.

  ⏸ Pause point: the library can be stored at −20°C until sequencing.

98. Perform paired-end sequencing of cDNA libraries from enriched neutrophils using the Illumina NovaSeq S4 flowcell and NovaSeq 6000 sequencer or an equivalent platform.
   a. For the NovaSeq S4, use the following read parameters for 300 total cycles: 2 × 150 bp read 1/read 2 sequencing reads, and 8 bp i7/i5 index reads.

Process sequencing data

⏱ Timing: 1–2 days

The alignment and quantification methods used herein are based on the GTEx/TOPMed RNA-seq pipeline from the Broad Institute, with custom modifications.

99. Obtain paired-end FASTQ files from the sequencer.

  Note: Each sample should have one R1 and one R2 fastq file. Depending on the sequencer or data delivery platform used, it may be necessary to concatenate files from multiple lanes of a flow cell that correspond to the same sample.

100. Install components of the pipeline (https://github.com/broadinstitute/gtex-pipeline/blob/master/TOPMed_RNAseq_pipeline.md).

101. Install FastQC by downloading the software and unzipping it in a suitable location.
102. Run FastQC on the FASTQ samples to test their quality.

```
>fastqc *.fastq
```

103. Align FASTQ files to the genome using STAR (Spliced Transcripts Alignment to a Reference).
    a. Using the GTEx pipeline Docker image established in "before you begin", run the following code.
    b. "star_index" is the path to the STAR index with the appropriate overhang generated in step 11 of "before you begin", "fastq1_abs" and "fastq2_abs" are the paths to the paired FASTQ files, and "prefix" is the name to be given to the BAM files.

```
>/src/run_STAR.py \

>  star_index fastq1_abs fastq2_abs prefix \

>  -output_dir star_out \

>  -outFilterMultimapNmax 20 \

>  -alignSJoverhangMin 8 \

>  -alignSJDBoverhangMin 1 \

>  -outFilterMismatchNmax 999 \

>  -outFilterMismatchNoverLmax 0.1 \

>  -alignIntronMin 20 \

>  -alignIntronMax 1000000 \

>  -alignMatesGapMax 1000000 \

>  -outFilterType BySJout \

>  -outFilterScoreMinOverLread 0.33 \

>  -outFilterMatchNminOverLread 0.33 \

>  -limitSjdbInsertNsj 1200000 \

>  -outSAMstrandField intronMotif \

>  -outFilterIntronMotifs None \

>  -alignSoftClipAtReferenceEnds Yes \

>  -quantMode TranscriptomeSAM GeneCounts \

>  -outSAMattrRGline ID:rg1 SM:sm1 \

>  -outSAMattributes NH HI AS nM NM ch \

>  -chimSegmentMin 15 \

>  -chimJunctionOverhangMin 15 \

>  -chimOutType Junctions WithinBAM SoftClip \

>  -chimMainSegmentMultNmax 1 \

>  -threads 8
```

104. Use Picard MarkDuplicates to identify PCR artifacts and generate the BAM index.
     a. Duplicates are defined as reads coming from the same fragment of DNA. "input_bam" is the output of the previous step.

```
>python3 -u /src/run_MarkDuplicates.py input_bam prefix \

>   -memory 4 \

>   -max_records_in_ram 50000 \

>   samtools index output_bam
```

105. Use RSEM (RNA-Seq by Expectation-Maximization) to quantify transcript abundance from the BAM file with duplicates marked.

```
>/src/run_RSEM.py \

>   -max_frag_len 1000 \

>   -estimate_rspd true \

>   -threads 2 \

>   rsem_reference output_bam sample_id

>gzip *.results

>   samtools index output_bam
```

106. Obtain quality control metrics of the alignment using RNA-SeQC2.
     a. RNA-SeQC2 takes the duplicate-marked BAM file, sample_id, and the collapsed genes GTF generated in "before you begin" as inputs.

```
>rnaseqc genes_gtf output_bam . -s sample_id -vv

>gzip *.gct
```

107. Aggregate results for RSEM and RNA-SeQC2 separately.
     a. tpm_gcts, count_gcts, exon_count_gcts, and metrics_gcts are the listed outputs of RNA-SeQC2, while rsem_isoforms and rsem_genes are the listed outputs of RSEM.

```
>mkdir individual_outputs

>mv tpm_gcts individual_outputs/

>mv count_gcts individual_outputs/

>mv exon_count_gcts individual_outputs/

>mv metrics_tsvs individual_outputs/

>python3 -m rnaseqc aggregate \

>   -parquet \

>   -o . \

>   individual_outputs \

>   prefix
```

```
>

>python3 /src/aggregate_rsem_results.py rsem_isoforms TPM IsoPct expected_count prefix

>python3 /src/aggregate_rsem_results.py rsem_genes TPM expected_count prefix
```

Quality control analysis

⏱ Timing: 2 h–2 days

108. Import the count matrix, transcripts-per-million (TPM) matrix, quality control metrics, and other metadata into R.
109. Log transform the TPM data and run a principal component analysis (PCA) (Figure 7).
     a. Plot the various quality control metrics over the PCA to see if any parameters are driving outlier behavior and which should be filtered on.

```
>genepc <- read.delim(paste0(prefix,"Ensembl_to_Symbol.txt"))

>logTPM <- log2(TPM + 1)

>proteincoding <- genepc$Gene.stable.ID[genepc$Gene.type == "protein_coding"]

>logPCG <- logTPM[which(rownames(logTPM) %in% proteincoding),]

>logPCG <- logPCG[which(rowSums(logPCG) > 0),]

>pcaResults <- prcomp(t(logPCG), center = T, scale. = T)
```

110. Filter out low quality samples.
     a. The exact parameters for filtration will depend on the source and inherent quality of the dataset.
     b. For this neutrophil dataset, the criteria we find to retain the most high-quality samples is:
        i.   Percent of mitochondrial reads < 20%
        ii. Genes Detected > 10,000.
        iii. Median Exon CV < 1.
        iv. Exon CV MAD < 0.75.
        v.   Exonic rate > 25%.
        vi. Median 3′ bias < 0.9.
        vii. Other quality control metrics may be considered depending on the context; for example, one could generate a "housekeeping gene" score based on genes that viable cells should express and remove samples below a threshold.
111. Run the PCA once more, this time using only the high-quality samples (Figure 7). Check that technical bias does not have a major impact on the data distribution.
112. Check that the sample filtration procedure has not introduced bias into the dataset.
     a. For example, it is possible that samples from patients with severe disease could have lower quality and be systematically filtered, biasing the data towards patients with mild disease.
     b. Similar bias could unintentionally be introduced on the basis of age, gender, race, ethnicity, disease status, or comorbidities.
     c. To test whether there is an association between sample exclusion and one of these variables, perform a Fisher test on the two-way table comparing inclusion status versus the categories of the clinical variable.

Obtain cell type proportion estimates from CIBERSORTx
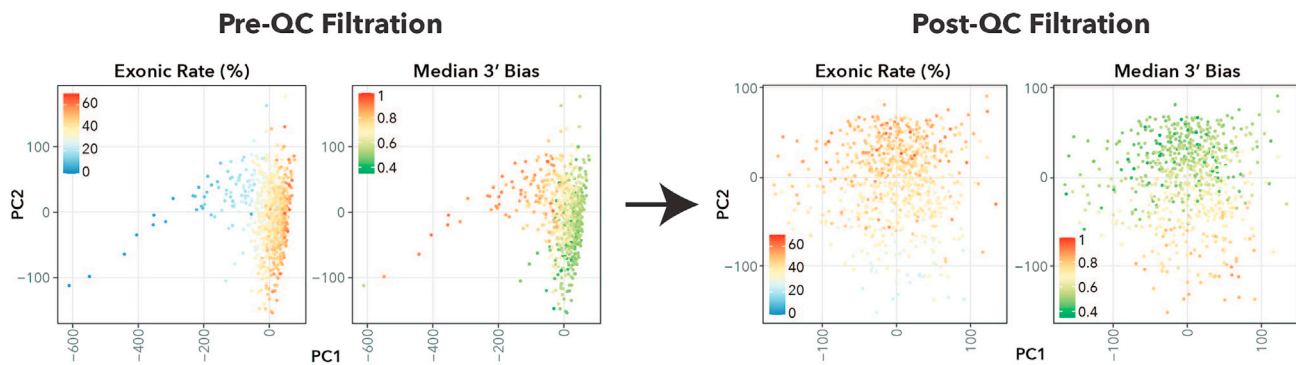
⏱ Timing: 2–6 h

Figure 7. Principal component analysis (PCA) plots displaying Exonic Rate and Median 3′ bias of samples pre- and post-quality control filtration

113. Upload the quality control-filtered gene count matrix to CIBERSORTx as a "Mixture" file.
114. Select "2. Impute Cell Fractions" in "Custom" mode.
    a. Select the Signature Matrix file generated during "before you begin", and pick the "Mixture" file that has been uploaded in the previous step.
    b. Run the algorithm in "relative" mode (i.e., not absolute mode).
115. Download the cell type proportions as estimated by CIBERSORTx (Figure 8).
    a. Check for associations between the various estimates and the outcomes of interest to the study.

Note: For example, we identified an association between higher estimated T/NK cell fraction and non-severe COVID-19, which could potentially confound the results of a differential expression analysis between severe and non-severe disease. To control for this, the cell type proportion estimates can be used as covariates in downstream analyses.

EXPECTED OUTCOMES

This protocol allows the user to generate high-quality RNA-seq data from a large number of samples from neutrophils, a challenging yet important cell type to profile due to their involvement in inflammatory and infectious disease. The Smart-seq2 protocol allows for powerful, sensitive, and accurate full length cDNA sequencing, while the subsequent quality control filtration and cell type abundance estimates using CIBERSORTx helps to control for technical effects and contamination.

The final result of the protocol will be three matrices: a gene count by sample matrix, a gene TPM by sample matrix, and a matrix of cell type abundance estimates by sample. Combined with the user's metadata, these represent all necessary components for downstream analyses of interest.

LIMITATIONS

First, bulk RNA-seq has limited resolution compared to single-cell technologies, and signatures derived in downstream analyses will reflect a mixture of neutrophil states. Second, the protocol utilizes negative selection to enrich for neutrophils, but this does not result in 100% sample purity. Thus, expression of contaminating cell-type-specific genes cannot be excluded. Third, neutrophils have a short half-life and many will inevitably die over the course of the enrichment. This may result in elevation of mitochondrial and cell-death-associated gene expression. Thus, all samples need to be treated the same to minimize batch effects. Fourth, neutrophils inherently have lower RNA quality due to their high RNAse content and relatively lower transcript abundance. Fifth, the cell type proportion estimate relies on an accurate reference which incorporates all relevant cell types in the reference, and this can only provide an estimate. More accurate quantification would rely on flow cytometry measurements of all cell types present in each sample, but this is less feasible when
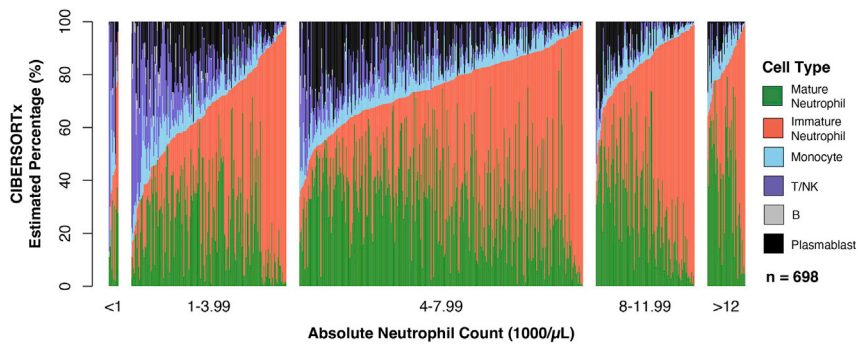
Figure 8. Bar plots displaying the CIBERSORTx estimated cell type percentages for each sample

Bar plots are separated by the absolute neutrophil count (ANC) range of the blood sample from which the neutrophils were isolated.

many samples are being processed in parallel and many neutrophils will die faster as the samples are stained.

TROUBLESHOOTING

Problem 1

The neutrophil cell count yield following the isolation protocol is low or the suspension contains many dead cells (step 33).

Potential solution

There are a variety of factors that can contribute to a low yield of neutrophils or to a large number of dead cells. In order to minimize cell death and to maximize the number of neutrophils enriched, the isolation procedure should be performed as efficiently as possible. Additionally, researchers can attempt to minimize the effects of the RBC depletion on neutrophil viability by trying different volumes of ACK lysis buffer and by minimizing the time that samples spend in ACK buffer.

Problem 2

The cDNA yield following PCR amplification is low by Qubit (step 83).

Potential solution

One obvious cause of a low cDNA yield could be that the neutrophils were counted inaccurately and too few cells were lysed. Ensure that neutrophils are counted accurately prior to lysis so that enough mRNA is available for the RT reaction and subsequent amplification. Additionally, if the recommended number of cells are lysed and researchers are still obtaining lower than desirable cDNA yields, the number of cycles in the **PCR Preamplification** protocol should be increased. However, the number of cycles should not be too high or else the data will be further biased for overabundant transcripts. In general, it is better to avoid overamplification at the cost of losing a few lower-quality samples.

Problem 3

The cDNA Bioanalyzer trace following PCR amplification is abnormal (step 84).

Potential solution

In order to increase the quality of the cDNA generated, a few different approaches can be used. First, consider reducing the number of PCR cycles used for the cDNA amplification if an acceptable yield of cDNA can still be achieved in order to reduce the incidence of primer dimers. Primer dimers can also be used through an additional SPRI clean-up step (with 0.8 volumes of beads). If an unreasonable portion of the cDNA falls between 150–500 bp on the Bioanalyzer trace, the extent of RNA

degradation and overall RNA quality can be assessed via the Qubit RNA IQ Assay kit (ThermoFisher Scientific; CAT# Q33221) or the Bioanalyzer High Sensitivity RNA kit (Agilent; CAT# 5067-1513), according to the manufacturers' protocols.

Problem 4

The Nextera Library Bioanalyzer trace is skewed or not registering (step 97).

Potential solution

It is possible that the cDNA was over fragmented during the tagmentation step. To prevent this, researchers should be sure to keep the reaction on ice during library prep as indicated by the protocol. Additionally, it is important to keep the library plate on ice after the tagmentation is complete and to add the NT to quench the tagmentation as efficiently as possible.

Problem 5

There are outlier samples on the PCA (step 111).

Potential solution

Outlier samples on the PCA can arise for many reasons, some biological and some technical. It is important to identify whether the effect is technical so that it can be filtered out or regressed. As demonstrated in Figure 7, visualize quality control metrics on the PCA or plot individual PCs against quality control metrics to see whether any are significantly correlated. It may be necessary to use more stringent cutoffs for the quality control metrics that were suggested. In addition, the user can add other quality control metrics that were not used in this protocol example, such as a house-keeping gene score.

PCA outliers could also arise from contamination from other cell types. To check if this is the case, check whether individual gene markers of different cell types are correlated with the PCs. Also, check whether the individual CIBERSORTx fractions are correlated with the PCs.

Since there is often no definite cutoff for sample quality, the user will have to see which parameters make most sense with their dataset. Then they should check again that the sample filtration has not introduced bias into the data.

Problem 6

The CIBERSORTx neutrophil cell type purity is low (step 115).

Potential solution

We have found that patients with lower absolute neutrophil counts on blood draw generally have poorer enrichment. In addition, we have found that samples with high lymphocyte counts are also less pure following neutrophil enrichment. Also, patients with more severe disease may have more activated neutrophils which could undergo cell death faster than other contaminating cell types. Regardless of the cause, using more magnetic selection beads and/or selection antibodies may result in higher purity, or performing a second round of enrichment. In addition, being prepared for the next step such that no time is wasted reduces cell death over the course of the enrichment so neutrophils will not be depleted.

Problem 7

Despite correcting for CIBERSORTx cell type proportions, the downstream results still have traces of other cell types.

Potential solution

CIBERSORTx can only generate cell type estimates as well as they are represented in the reference dataset. For example, plasmablasts are not well represented in the reference dataset, but it is

possible that many have contaminated the sample and CIBERSORTx would not be the best tool to identify this. Instead, we suggest generating a signature of the contaminating cell type which can be used as a covariate to regress the contamination out of downstream analyses such as differential expression.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Moshe Sade-Feldman (msade-feldman@mgh.harvard.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The raw neutrophil RNA sequencing data used as the example in this study has not been deposited in a public repository because these data were collected at the beginning of the COVID-19 pandemic, and therefore, a waiver of informed consent was approved by the Massachusetts General Hospital governing institutional review board in compliance with the Code of Federal Regulation (45CFR 46, 2018 Common Rule). The read count matrix and TPM matrix are available in GEO: GSE212041. Any code required to run the quality control and pseudobulking for CIBERSORTx are available in the manuscript and has been deposited in Zenodo: https://doi.org/10.5281/zenodo.7030528. The example reference single cell dataset from Schulte-Schrepping et al.[12] is available through EGA: EGAS00001004571. All additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xpro.2023.102125.

AUTHOR CONTRIBUTIONS

Conceptualization, A.L.K.G., T.J.L., S.S.F., M.R., A.C.V., M.S.-F.; investigation, A.L.K.G., T.J.L., S.S.F., M.R., A.C.V., M.S.-F.; writing – original draft, A.L.K.G., T.J.L., M.S.-F.; writing – review & editing, A.L.K.G., T.J.L., S.S.F., M.R., N.H., A.C.V., M.S.-F.; funding acquisition, A.C.V., N.H.; supervision, M.S.-F.

DECLARATION OF INTERESTS

T.J.L. advises Chemomab Therapeutics. M.S.-F. received funding from Bristol Myers-Squibb and is a consultant for Galvanize Therapeutics. N.H. holds equity in Biontech and holds equity in and advises Danger Bio. M.R. is a current employee at Genentech.

REFERENCES

1. LaSalle, T.J., Gonye, A.L.K., Freeman, S.S., Kaplonek, P., Gushterova, I., Kays, K.R., Manakongtreecheep, K., Tantivit, J., Rojas-Lopez, M., Russo, B.C., et al. (2022). Longitudinal characterization of circulating neutrophils uncovers phenotypes associated with severity in hospitalized COVID-19 patients. Cell Rep. Med. 3, 100779. https://doi.org/10.1016/j.xcrm.2022.100779.

2. Boribong, B.P., LaSalle, T.J., Bartsch, Y.C., Ellett, F., Loiselle, M.E., Davis, J.P., Gonye, A.L.K., Sykes, D.B., Hajizadeh, S., Kreuzer, J., et al. (2022). Neutrophil profiles of pediatric COVID-19 and multisystem inflammatory syndrome in children. Cell Rep. Med. 3, 100848. https://doi.org/10.1016/j.xcrm.2022.100848.

3. Lahoz-Beneytez, J., Elemans, M., Zhang, Y., Ahmed, R., Salam, A., Block, M., Niederalt, C., Asquith, B., and Macallan, D. (2016). Human neutrophil kinetics: modeling of stable isotope labeling data supports short blood neutrophil half-lives. Blood 127, 3431–3438. https://doi.org/10.1182/blood-2016-03-700336.

4. Mitchell, M.J., Lin, K.S., and King, M.R. (2014). Fluid shear stress increases neutrophil activation via platelet-activating factor. Biophys. J. 106, 2243–2253. https://doi.org/10.1016/j.bpj.2014.04.001.

5. Begandt, D., Thome, S., Sperandio, M., and Walzog, B. (2017). How neutrophils resist shear stress at blood vessel walls: molecular mechanisms, subcellular structures, and cell-cell interactions. J. Leukoc. Biol. 102, 699–709. https://doi.org/10.1189/jlb.3MR0117-026RR.

6. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. 37, 773–782. https://doi.org/10.1038/s41587-019-0114-2.

7. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

8. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genomeBMC.

Bioinf 12, 323. https://doi.org/10.1186/1471-2105-12-323.

9. Graubert, A., Aguet, F., Ravi, A., Ardlie, K.G., and Getz, G. (2021). RNA-SeQC 2: efficient RNA-seq quality control and quantification for large cohorts. Bioinformatics 37, 3048–3050. https://doi.org/10.1093/bioinformatics/btab135.

10. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587. https://doi.org/10.1016/j.cell.2021.04.048.

11. Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods 10, 1096–1098. https://doi.org/10.1038/nmeth.2639.

12. Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al. (2020). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. Cell 182, 1419–1440.e23. https://doi.org/10.1016/j.cell.2020.08.001.