

**SONY**

Watch and learn how the ID7000 software enables users at all expertise levels to acquire and analyze high parameter data

[View Tutorial Videos](#)

*The Journal of*  
**Immunology**

RESEARCH ARTICLE | AUGUST 15 2022

## Single-Cell Analysis Reveals the Range of Transcriptional States of Circulating Human Neutrophils

Gustaf Wigerblad; ... et. al

*J Immunol* (2022) 209 (4): 772–782.

<https://doi.org/10.4049/jimmunol.2200154>

### Related Content

#### Top Reads

*J Immunol* (August,2022)

Positioning of the HA Globular Head within the TLR5 Agonist, Flagellin, Modulates the Immunogenicity and Reactogenicity of a Novel-H1 Pandemic Influenza Vaccine (136.13)

*J Immunol* (April,2010)

# Single-Cell Analysis Reveals the Range of Transcriptional States of Circulating Human Neutrophils

Gustaf Wigerblad,<sup>\*,1</sup> Qilin Cao,<sup>\*,1</sup> Stephen Brooks,<sup>†</sup> Faiza Naz,<sup>‡</sup> Manasi Gadkari,<sup>\*</sup> Kan Jiang,<sup>†</sup> Sarthak Gupta,<sup>\*</sup> Liam O'Neil,<sup>\*,2</sup> Stefania Dell'Orso,<sup>‡</sup> Mariana J. Kaplan,<sup>\*,3</sup> and Luis M. Franco<sup>\*,3</sup>

Neutrophils are the most abundant leukocytes in human blood and are essential components of innate immunity. Until recently, neutrophils were considered homogeneous and transcriptionally inactive cells, but both concepts are being challenged. Single-cell RNA sequencing (scRNA-seq) offers an unbiased view of cells along a continuum of transcriptional states. However, the use of scRNA-seq to characterize neutrophils has proven technically difficult, explaining in part the paucity of published single-cell data on neutrophils. We have found that modifications to the data analysis pipeline, rather than to the existing scRNA-seq chemistries, can significantly increase the detection of human neutrophils in scRNA-seq. We have then applied a modified pipeline to the study of human peripheral blood neutrophils. Our findings indicate that circulating human neutrophils are transcriptionally heterogeneous cells, which can be classified into one of four transcriptional clusters that are reproducible among healthy human subjects. We demonstrate that peripheral blood neutrophils shift from relatively immature (Nh0) cells, through a transitional phenotype (Nh1), into one of two end points defined by either relative transcriptional inactivity (Nh2) or high expression of type I IFN-inducible genes (Nh3). Transitions among states are characterized by the expression of specific transcription factors. By simultaneously measuring surface proteins and intracellular transcripts at the single-cell level, we show that these transcriptional subsets are independent of the canonical surface proteins that are commonly used to define and characterize human neutrophils. These findings provide a new view of human neutrophil heterogeneity, with potential implications for the characterization of neutrophils in health and disease. *The Journal of Immunology*, 2022, 209: 772–782.

The understanding of heterogeneity and plasticity in hematopoietic cells is changing rapidly. Historically, a combination of cell surface markers, transcription factors, and profiles of secreted cytokines has been employed to classify cells of similar histological appearance and ontogeny into discrete groups. The implicit assumption of this categorization, that the resulting cell “populations” or “subsets” represent polarized and fixed states, has been questioned for the past two decades by an extensive body of evidence. This is exemplified by the recent evidence that T cells and macrophages, which have long been classified in terms of subsets, can convert from one state to another and display mixed or partial profiles (1–3). This evidence suggests that hematopoietic cells may be best understood

along a continuum of differentiation and activation states. In this context, single-cell RNA sequencing (scRNA-seq) has been an important addition to the set of analytical tools, because cells in different states may express different sets of genes and scRNA-seq offers a less biased view of cells along a continuum of transcriptional states. Our understanding of the spectrum of cell states at baseline, in response to specific stimuli, and in disease remains limited. The extent to which different transcriptional states correspond to older classifications based on a limited number of proteins is also unclear for most cell types.

Neutrophils are the most abundant leukocytes in human blood and are essential components of the innate immune system. Until recently, they were thought to be a fairly homogeneous and

<sup>\*</sup>Systemic Autoimmunity Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD; <sup>†</sup>Biodata Mining and Discovery Section, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD; and <sup>‡</sup>Genomic Technology Section, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD

<sup>1</sup>G.W. and Q.C. share first authorship.

<sup>2</sup>Current address: Department of Internal Medicine, University of Manitoba, Winnipeg, MB, Canada.

<sup>3</sup>M.J.K. and L.M.F. jointly supervised the work and share senior authorship.

ORCIDs: 0000-0002-1798-4347 (G.W.); 0000-0001-9065-4574 (K.J.); 0000-0002-0540-5942 (L.M.F.).

Received for publication February 23, 2022. Accepted for publication May 23, 2022.

This work was supported by the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases at the National Institutes of Health (NIH) Grant ZIA AR041199. L.M.F. also receives research support from the Division of Intramural Research at the National Institute of Allergy and Infectious Diseases at the NIH.

G.W. and Q.C.: conceptualization, data curation, formal analysis, investigation, visualization, and writing – original draft. S.B. and K.J.: data curation, formal analysis, and methodology. F.N. and M.G.: investigation. L.O., S.G., and S.D.: conceptualization

and investigation. M.J.K. and L.M.F.: conceptualization, supervision, and writing – review & editing.

The sequences presented in this article have been submitted to the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE188288>) under accession number GSE188288.

Address correspondence and reprint requests to Dr. Luis M. Franco or Dr. Mariana J. Kaplan, Functional Immunogenomics Unit, Systemic Autoimmunity Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, 9000 Rockville Pike, Building 10, Room 13C101A, Bethesda, MD 20892 (L.M.F.) or Systemic Autoimmunity Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, 9000 Rockville Pike, Building 10, Room 12N248C, Bethesda, MD 20892 (M.J.K.). E-mail addresses: luis.franco@nih.gov (L.M.F.) or mariana.kaplan@nih.gov (M.J.K.)

The online version of this article contains supplemental material.

Abbreviations used in this article: BART, binding analysis for regulation of transcription; cat. no., catalog number; COVID-19, coronavirus disease 2019; LDG, low-density granulocyte; scRNA-seq, single-cell RNA sequencing; SLE, systemic lupus erythematosus; UMAP, uniform manifold approximation and projection for dimension reduction; UMI, unique molecular identifier.

This article is distributed under The American Association of Immunologists, Inc., [Reuse Terms and Conditions for Author Choice articles](#).

transcriptionally inactive cell type, but both concepts have been convincingly challenged in recent years (4, 5). Although human neutrophils have lower total RNA content per cell than macrophages (6) and other hematopoietic cell types (Supplemental Table 1), they express a broad range of genes in resting conditions (7, 8), and their transcriptome is strongly reactive to environmental stimuli (9–11). Neutrophils have been characterized based on discrete parameters, including cell-surface markers, buoyancy, histological characteristics associated with maturation status, or tissue localization. These observations have led to the emergent concept of neutrophil heterogeneity, which has been the subject of recent reviews (4, 5). In these, it has been proposed that single-cell sequencing technology is a promising avenue for a more comprehensive and less biased characterization of neutrophil states. Recent studies in mouse models, with a limited number of human samples for comparison, have applied scRNA-seq to the study of circulating and bone marrow neutrophils and have indeed documented the existence of a range of transcriptional states (12, 13). Direct evidence for distinct transcriptional subsets of human neutrophils has also been provided, by our group and others, in scRNA-seq studies of sex differences in the neutrophils of healthy donors (14) and in patients with lung cancer (15) or coronavirus disease 2019 (COVID-19) (16, 17). However, given their low per-cell RNA content, scRNA-seq in neutrophils remains technically challenging, explaining in part the paucity of scRNA-seq reports describing human neutrophils compared with other hematopoietic cell types. To address this, we first evaluated the technical aspects of scRNA-seq data generation and analysis. We found that a modified pipeline is necessary for proper identification of neutrophils in scRNA-seq data. We then applied such a pipeline to the transcriptional characterization of human circulating neutrophils from multiple healthy donors at the single-cell level.

## Materials and Methods

### Cell purification

Human venous peripheral blood samples from healthy donors were obtained from the Department of Transfusion Medicine at the National Institutes of Health Clinical Center. For neutrophil purification, whole blood samples were collected in vacutainer glass blood collection tubes with acid citrate dextrose. Neutrophils were isolated with the EasySep Direct Human Neutrophil Isolation Kit (cat. no. 19666; STEMCELL Technologies).

For granulocyte purification, whole blood was collected in heparinized tubes. Granulocytes were isolated by dextran sedimentation of RBC pellets as previously described (18). In brief, cells were first layered on a Ficoll/Hypaque gradient (cat. no. 17144003; GE Healthcare). The granulocyte/RBC fraction was then enriched by dextran sedimentation followed by RBC lysis using hypotonic solution. Granulocytes were then washed with PBS. For WBC purification, whole blood samples were collected in heparinized tubes. WBCs were isolated with the Erythroclear RBC Depletion Reagent Kit (cat. no. 01738; STEMCELL Technologies).

### Documentation of cell purity and viability

Flow cytometry was used to assess the purity and viability of purified neutrophils. The cells were stained with a panel of mAbs containing ECD CD16 clone 3G8 (cat. no. A33098; Beckman Coulter), BV711 CD45 clone HI30 (cat. no. 564357; BD Biosciences), and FITC CD66b clone G10F5 (cat. no. 305104; BioLegend). The LIVE/DEAD Fixable Dead Cell Stain Kit with aqua fluorescent reactive dye (cat. no. L34957; Thermo Fisher Scientific) was used to assess cell viability. PE-Annexin V (cat. no. 640908; BioLegend) was used to assess early apoptosis activity. UltraComp eBeads Compensation Beads (cat. no. 01-2222-42; Thermo Fisher Scientific) were used to perform spectral compensation. Data were collected by a BD Biosciences FACSCelesta flow cytometer and later analyzed with FlowJo software (v10). The purity of neutrophils was specifically defined by cell-lineage markers as the proportion of CD66b<sup>+</sup>CD16<sup>+</sup> events among CD45<sup>+</sup> events.

### scRNA-Seq

From each cell purification sample, ~50,000 cells were centrifuged at 300 × g for 5 min at 4°C and washed twice with PBS with 0.02% BSA. To

obtain single-cell gel beads-in-emulsion, we resuspended cells at a concentration of 1000 cells/μl and added 1 μl of RNase Inhibitor (cat. no. 10777-019; Invitrogen) before loading the mix on a Chromium Comptroller Instrument (10× Genomics). Single-cell cDNAs and libraries were prepared with a Chromium Single Cell 3' Library & Gel Bead Kit v3.1 (cat. no. 1000121; 10× Genomics). In brief, gel beads-in-emulsion–reverse transcription incubation was performed in a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (cat. no. 1851197; Bio-Rad): 53°C for 45 min, 85°C for 5 min, held at 4°C. Single-strand cDNAs were purified with DynaBeads MyOne Silane Beads (cat. no. 37002D; Thermo Fisher Scientific) and amplified with the C1000 Touch Thermal cycler with 96-Deep Well Reaction Module: 98°C for 3 min; 13 cycles of 98°C for 15 s, 63°C for 20 s, and 72°C for 1 min; 72°C for 1 min; held at 4°C. Amplified cDNA products were cleaned with 0.6× DynaBeads MyOne Silane Beads (cat. no. 37002D; Thermo Fisher Scientific). Quality and quantity of the cDNAs were assessed on a 4200 Tape Station (Agilent Technologies) with High Sensitivity D5000 DNA Screen Tape (cat. no. 5067-5592; Agilent). The final material was amplified as follows: 98°C for 45 s; 16 cycles of 98°C for 20 s, 54°C for 30 s, 72°C for 20 s; 72°C for 1 min; held at 4°C. Libraries were diluted to the same molarity and pooled for sequencing on NextSeq500 (Illumina) or NovaSeq6000 (Illumina) sequencers. Sequencing read lengths were 28 bp for read 1, 8 bp for the i7 index, and 91 bp for read 2.

Protease and RNase activity are known to be highly active in neutrophils (19). It should be noted, however, that the addition of a protease inhibitor (cat. no. A32963; Thermo Fisher Scientific) or an RNase inhibitor (cat. no. AM2682; Ambion), individually or in combination, to the standard 10× Genomics protocol for cell capture and library preparation did not increase the final cDNA concentration at the end of the library construction phase of the protocol.

### Cellular Indexing of Transcriptomes and Epitopes by Sequencing

TotalSeq-B oligonucleotide-conjugated Abs (BioLegend), compatible with the 10× Genomics 3' scRNA-seq chemistry, were used according to the manufacturer's protocol. The panel for common markers of circulating neutrophils included Abs targeting CD45, CD14, CD33, CD11c, CD10, CD16, CD107a, HLA-DR, CD11b, CD66b, CD35, CD24, CD184, and CD15.

### Processing and analysis of scRNA-seq data

Illumina run folders were demultiplexed and converted to FASTQ format with Cell Ranger mkfastq version 4.0.0 and Illumina bcl2fastq version 2.20. Reads were further counted and analyzed with Cell Ranger count version 4.0.0 and the refdata-gex-GRCh38-2020-A reference, to generate raw and filtered matrix files. The data can be accessed at GEO accession number GSE188288 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE188288>).

Matrix files were imported into the R package Seurat version 4.0.1 (20) for downstream processing. From the raw matrices, cells with a gene number between 100 and 2500 and a mitochondrial gene proportion < 0.1 were selected for downstream analysis. The matrices were then normalized by the LogNormalize method. The FindVariableFeatures() function was used to select the top 2000 variable genes with the versust selection method. Scaling was performed by the function ScaleData() regressing out the mitochondrial gene content. Principal component analysis and clustering were then performed on the scaled data. UMAP (version 0.2.7.0) was used for visualization, and SingleR (version 1.4.1) was used for cell identification.

After neutrophils were identified in the dataset corresponding to each sample, they were integrated. First, contaminants were removed if they had gene expression values > 1 for three marker genes specific to RBCs (HBA2), T cells (CD3G), and cells with a predominance of rRNA (RPS8). Second, genes that were shared among all datasets were identified for downstream integration. Anchors were identified with the FindIntegrationAnchors() function, and these anchors were used to integrate the neutrophils together with the function IntegrateData(). Finally, uniform manifold approximation and projection for dimension reduction (UMAP) was performed on the top eight principal components from the integrated data, and the resolution was set to 0.2 for visualization of the four clusters identified.

To examine the effect of sequencing depth on clustering and other downstream analyses, we down-sampled selected single samples by 50%. Typical single-cell Illumina runs consisted of two lanes of a flow cell sequencing the same pooled libraries. Fifty percent down-sampling was accomplished by analyzing the data from a single lane.

To study neutrophil cell-state trajectories, we used the analysis toolkit Monocle3, which is implemented as an R package (version 0.2.3.0) (21). A principal graph was learned on the UMAP projection of the cells with the learn\_graph() function. To generate a pseudotime axis, we then ordered the cells with the order\_cells() function. To identify genes that vary between groups of cells in UMAP space, we used the graph\_test() function; this function employs the spatial autocorrelation analysis statistic Moran's I, which has been shown to be effective for identifying genes that vary in scRNA-seq

datasets (22). The genes found to be variable were then grouped into modules with the function `find_gene_modules()`, which employs the Louvain method for community detection (23), to identify clusters of genes with a similar pattern of expression. To infer which transcriptional regulators are active in the cells, we used the module gene lists as input for the binding analysis for regulation of transcription (BART) pipeline (24). Transcription factors associated with *cis*-regulatory elements most likely to regulate the input gene lists (Irwin–Hall  $p < 0.01$ ) were used for further analysis with the Ghent University Bioinformatics and Evolutionary Genomics custom Venn diagram tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

### Single-cell Western blotting

Purified neutrophils were loaded on the scWest chip (ProteinSimple), allowed to settle for 20 min, and treated according to the manufacturer's instructions. In brief, the chip was placed in the Milo instrument (ProteinSimple) for 15-s lysis, 45-s separation, and 4-min ultraviolet exposure. The chip was then probed using Abs against the proteins ISG15 (cat. no. 2758; Cell Signaling) and GAPDH (cat. no. 5174; Cell Signaling), labeled with Alexa 488/Alexa 594, and scanned in an array scanner (Molecular Devices). Chips were then stripped and reprobed with Abs against IFITM3 (cat. no. 59212; Cell Signaling) and rescanned. Analysis of the images was done in Scout software (ProteinSimple), where GAPDH was used as loading control cell marker and data presented as % of total cells positive for ISG15 and/or IFITM3. A total of 3300 neutrophils were used from two different donors.

## Results

### *A modified analysis pipeline is required for the adequate identification of neutrophils in scRNA-seq data*

The standard analysis pipeline for scRNA-seq data generated with the 10× Genomics platform and Illumina short-read sequencing is implemented in the widely used Cell Ranger software (25). This pipeline involves grouping of the sequencing reads by their cell of origin (barcode) and RNA molecule of origin (unique molecular identifier [UMI]). This is followed by a cell-calling step, in which individual barcodes are determined to be empty (not corresponding to any cell) or to represent a captured cell. The current cell-calling algorithm employed by Cell Ranger is based on the EmptyDrops method (26). In the first step, the algorithm sets a threshold based on the number of UMIs associated with each barcode, and those that pass this threshold are classified as cells. In the second step, a set of barcodes with low UMI counts is selected, and a background model is generated. The RNA profile of each barcode that was not called as a cell in the first step is then compared against the background model, and those whose profile disagrees with that of the background model are called as cells. The resulting barcodes are then output in the form of a filtered matrix of the UMI counts corresponding to each gene in each called cell. The goal of the second step is to identify cells that may have lower RNA content than those identified in the first step.

To test the ability of this method to reliably identify human neutrophils in a mixed-cell population, we first generated scRNA-seq data from an RBC-depleted whole-blood sample and analyzed it with the standard Cell Ranger pipeline described earlier. Of the called cells in the filtered matrix, 27.2% were identified as neutrophils by an unbiased algorithm based on reference transcriptomic datasets (27), which was a clear underrepresentation of neutrophils in human whole blood (Fig. 1A). We then visualized the frequency distribution of the number of features per barcode (genes per cell), contrasting the filtered matrix with the unfiltered matrix (Fig. 1B). From this, it was clear that the filtered matrix excluded many events that were near the lower end of the distribution, yet formed a peak that is distinct from the null set of events with zero or near-zero features. We hypothesized that neutrophils, having lower overall transcript abundance than other cell types, could be enriched in this excluded cell population. To test this, we modified the analysis pipeline, departing from the unfiltered matrix and lowering the threshold of genes per cell based on the observed distribution. With

this modification, the proportion of cells that were identified as neutrophils increased to 58.7%, which is within the expected range of neutrophils in human whole blood (Fig. 1C). Correspondingly, the proportions of other nucleated cell types (T cells, B cells, NK cells, and monocytes) either decreased to, or remained within, their normal ranges in human peripheral blood, indicating a more expected representation of the cell composition of the sample. Overlaying the distribution of genes per cell of the events now identified as neutrophils on that of all events in the raw matrix indicated that a substantial proportion of the events in the distinct peak we had previously observed, in fact, correspond to neutrophils (Fig. 1D). To verify the identity of the cells that were rescued by the modified analysis pipeline as neutrophils, we used the bioinformatics tool NeutGX (14) and a publicly available dataset (GEO: GSE112101) of RNA-seq data in nine primary human immune cells (11) to identify genes that are highly expressed in neutrophils and specific to neutrophils (*FCGR3B*) or to myeloid cells (*CSF3R*, *NAMPT*). The expression of these neutrophil marker genes is high in the cells that were rescued by the modified analysis pipeline and classified as neutrophils, confirming their identity (Fig. 1E).

In practice, depending on the requirements of specific experimental settings, human neutrophils are purified by different methods. Therefore, we systematically compared the performance of the standard or modified pipelines for neutrophil scRNA-seq in cells purified by three common methods. Whole blood from each of seven healthy donors was simultaneously processed by three methods prior to cell capture for scRNA-seq: RBC-depleted whole blood, granulocytes from density-gradient centrifugation, and immunomagnetically purified neutrophils. In all sample types, the number of neutrophils detected was significantly higher with the modified scRNA-seq pipeline than with the standard pipeline (Fig. 1F).

We then asked whether the same principle could be applied to improve the identification of neutrophils in scRNA-seq experiments with samples from other tissues. To test this, we analyzed a recently published dataset (GEO: GSE145926) of bronchoalveolar lavage samples in patients with severe COVID-19 (28). With the standard analysis pipeline, 13.1% of cells were identified as neutrophils, compared with 55.8% of cells with the modified pipeline (Fig. 1G).

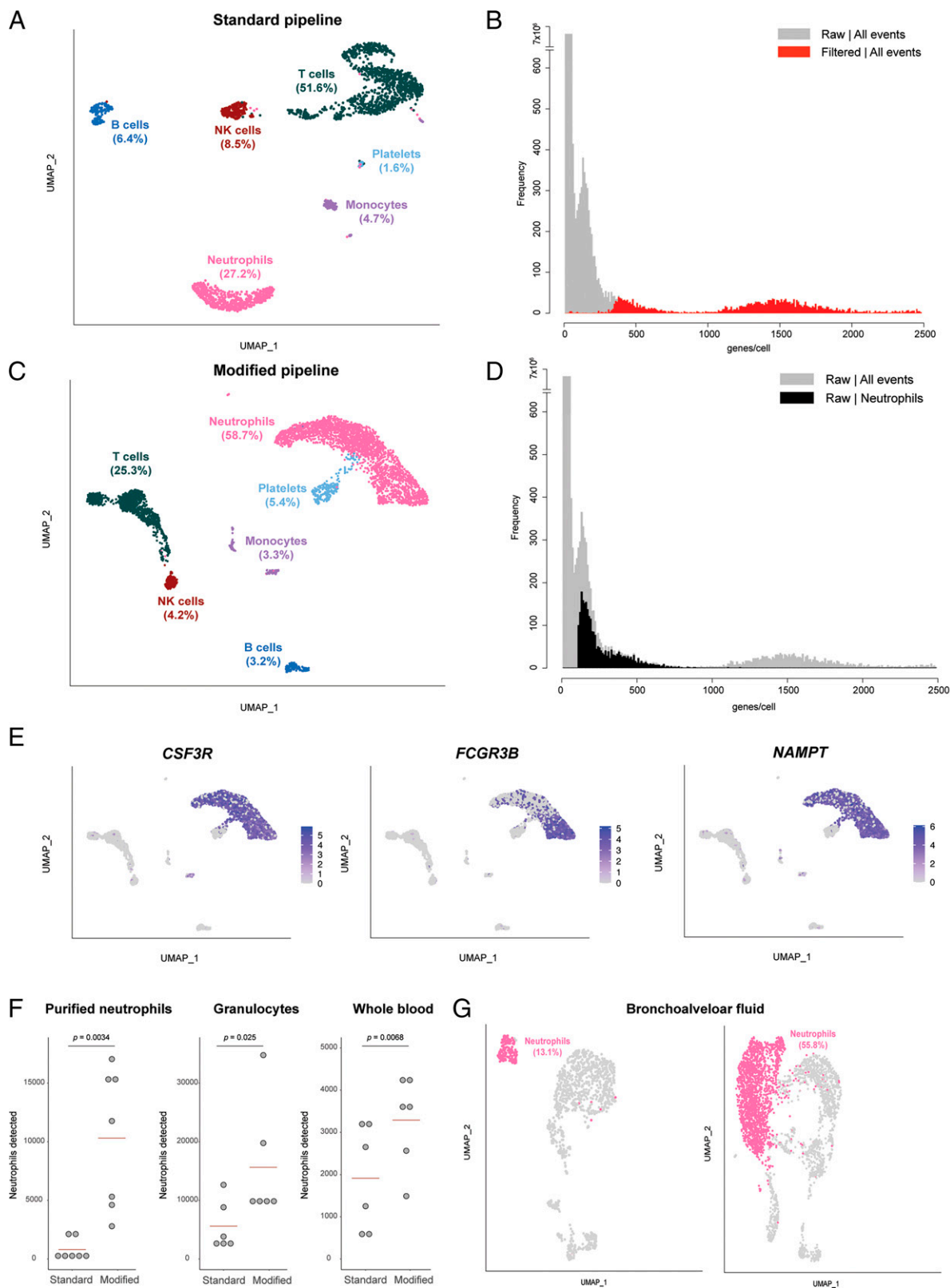
A related approach was proposed recently online (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/tutorials/neutrophils>, accessed February 9, 2022). It involves bypassing the second step of the standard cell-calling algorithm, forcing the Cell Ranger program to call a set number of events as cells, and including intronic reads. This is followed by filtering of noncell events based on the number of genes per cell. We performed a side-by-side comparison of this approach with our simpler, modified pipeline and found that both are capable of rescuing neutrophils in single-cell data and are comparable in terms of the specific cells and genes identified (Supplemental Fig. 1).

These results indicate that a modified analysis pipeline is required for adequate identification of neutrophils in scRNA-seq data, and that the cell-calling threshold along the frequency distribution of genes per cell is the key variable that has prevented standard analysis pipelines from identifying neutrophils.

### *Human circulating neutrophils consist of distinct and reproducible transcriptional subsets*

Recent observations by our group and others indicate that neutrophils from humans and mice exist in distinct transcriptional states (12–14, 16, 17). Taking advantage of the improved analysis pipeline, we directly evaluated this by performing scRNA-seq on purified and abundant neutrophil samples from healthy donors. To minimize the risk of potential changes in gene expression induced by gradient centrifugation, osmotic lysis of RBCs, or positive-



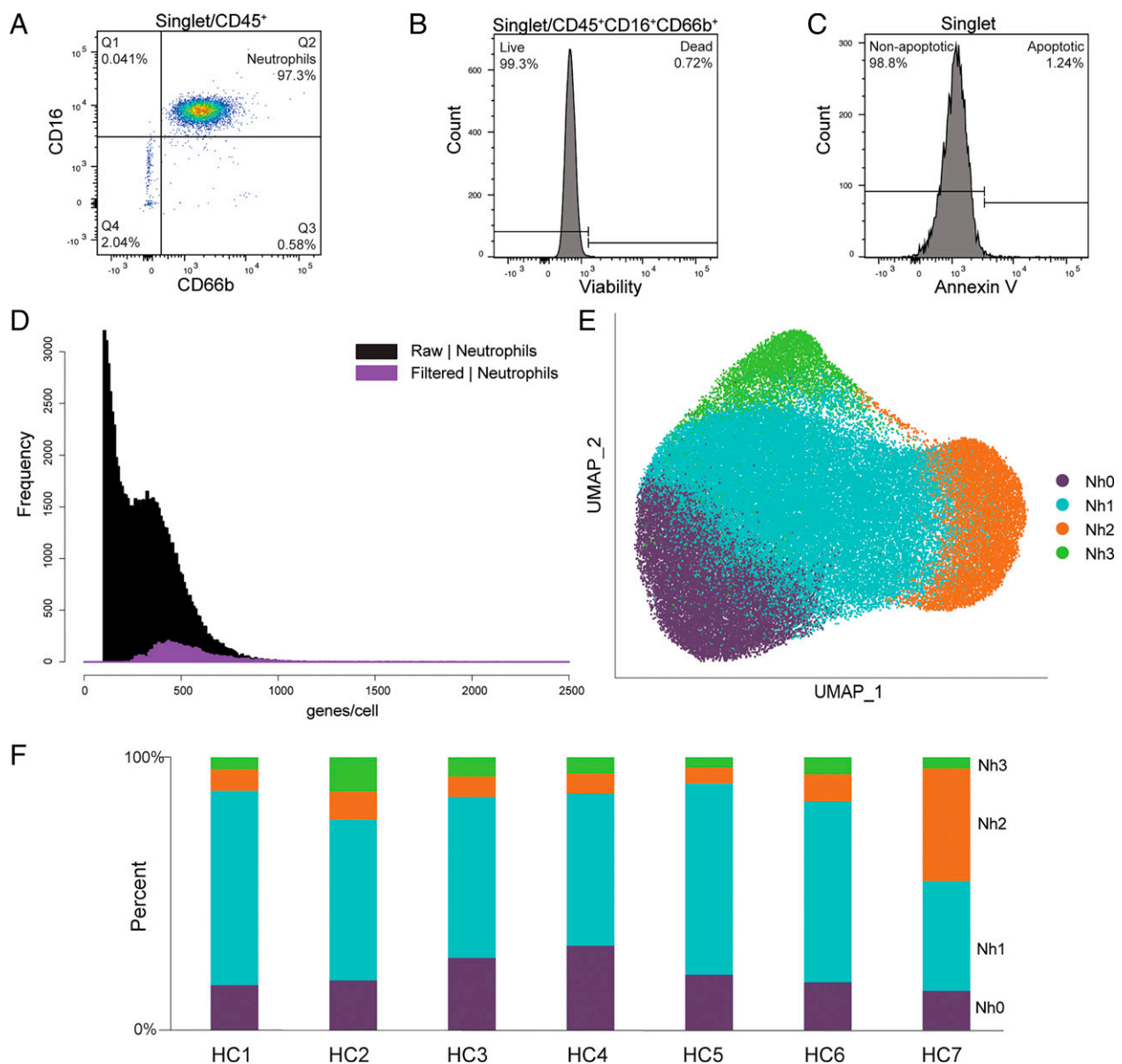


**FIGURE 1.** Pipeline for identification of neutrophils in scRNA-seq data. **(A)** Distribution of cell types identified in RBC-depleted whole blood in a scRNA-seq analysis performed with the filtered matrix output from Cell Ranger (standard pipeline). Data from one capture are shown. **(B)** Frequency distribution of the number of features per barcode (genes per cell) for the dataset shown in (A), comparing data from the filtered (red) versus raw (gray) matrices. **(C)** Distribution of cell types identified in the dataset shown in (A) when the analysis is performed with the raw matrix output from Cell Ranger (modified pipeline). **(D)** Frequency distribution of the number of features per barcode (genes per cell) for the dataset shown in (A) and (C), with the distribution for cells identified as neutrophils in the analysis of the raw matrix (modified pipeline) highlighted in black. **(E)** Feature plot on the UMAP shown in (C) for three genes expected to be highly expressed in human neutrophils. **(F)** Number of neutrophils detected by the standard or modified pipelines in samples from the same subjects processed by three methods. Each dot represents one biological replicate (one unrelated healthy donor). Statistical testing results are from a paired *t* test. **(G)** Proportion of neutrophils identified in a published scRNA-seq dataset of bronchoalveolar lavage fluid from patients with severe COVID-19 infection, comparing the results of the standard pipeline (left) with those of the modified pipeline (right).

selection Abs, we studied neutrophils purified directly from whole blood by immunomagnetic negative selection. Flow cytometry was performed on each sample to document purity, viability, and evidence of early apoptosis (Fig. 2A–C, Supplemental Table II). As expected, the modified analysis pipeline identified a high proportion of neutrophils that would have been excluded by the standard pipeline (Fig. 2D). A total of 72,183 purified circulating human neutrophils were analyzed. This analysis revealed four distinct transcriptional clusters (Fig. 2E), which were highly reproducible in samples obtained from seven unrelated healthy donors and processed independently (Fig. 2F). For clarity of display and to facilitate future comparisons of our data with those from other studies in humans or other species, we have classified these clusters as Nh0 (neutrophils, human, cluster 0) through Nh3. A table with the complete set of marker genes for each cluster is provided in Supplemental Dataset 1.

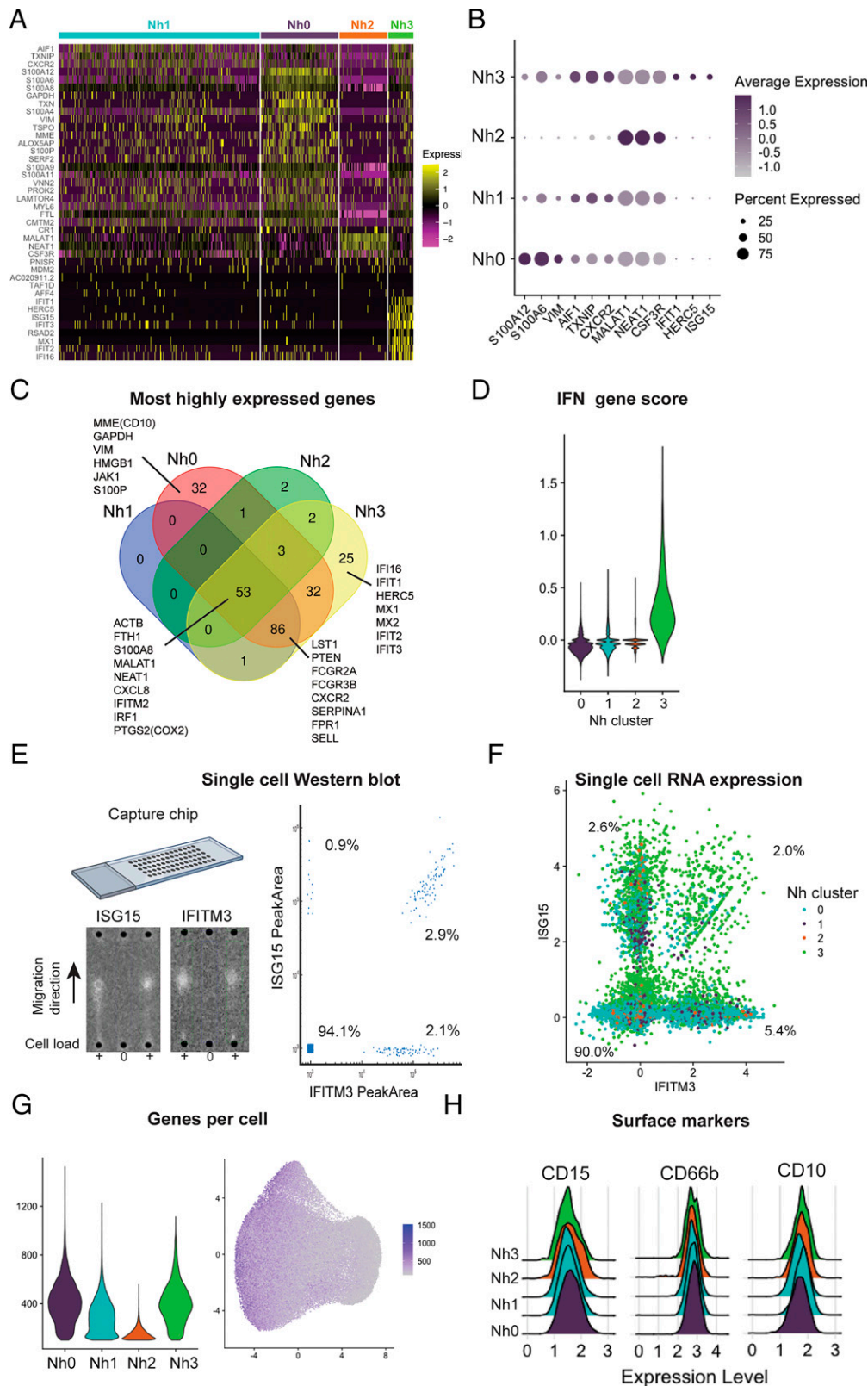
Nh0 neutrophils represent ~20% of circulating neutrophils (mean: 22.1%, range: 14.4–30.1%) and are characterized by higher

expression of genes that have been found to be characteristic of bone marrow neutrophils and are therefore associated with more immature neutrophil states (8, 13). These include the genes *MMP9*, *ITGAM*, *FCN1*, *CAMP*, *CYBB*, and *CST3*, which encode known or candidate neutrophil granule proteins (8, 29). The genes encoding vimentin (*VIM*), thioredoxin (*TXN*), and several proteins of the S100 family (*S100A6*, *S100A8*, *S100A9*, *S100A11*, and *S100A12*) are also differentially expressed in Nh0 cells compared with other clusters. Notably, the gene encoding the membrane metalloendopeptidase CD10 (*MME*), which at the protein level is associated with more mature neutrophils, is also more highly expressed in Nh0 cells, highlighting the complementary information offered by protein- and transcript-level measurements (Fig. 3A). Nh1 neutrophils represent the majority of circulating neutrophils (mean: 57.1%, range: 40.3–71.3%) and appear to be in a more mature state, as indicated by higher expression of the genes *AIF1*, *CXCR2*, and *TXNIP* (Fig. 3A, 3B). Compared with other clusters, Nh1 neutrophils have



**FIGURE 2.** Circulating human neutrophils consist of distinct transcriptional subsets. (A–C) Flow cytometry documentation of human neutrophil purity and viability. A representative sample is shown for each panel. Purity was defined as the proportion of CD66b<sup>+</sup>CD16<sup>+</sup> events among CD45<sup>+</sup> events, as shown in (A). Viability was assessed by uptake of an amine-binding dye, as shown in (B). Evidence of early apoptosis was assessed by Annexin V staining, as shown in (C). Results for each sample are in Supplemental Table II. (D) Frequency distribution of the number of features per barcode (genes/cell) in the purified neutrophils dataset, comparing data from the filtered (purple) and raw (black) matrices. (E) Two-dimensional projection (UMAP) of 72,183 purified circulating human neutrophils showing clusters Nh0–Nh3. (F) Bar graph showing the cluster proportion of the neutrophils from each of seven healthy controls (HC1–HC7).

**FIGURE 3.** Neutrophil transcriptional subsets vary by type and number of genes expressed. **(A)** Heatmap of the top marker genes from each cluster. Each row represents one gene, and each column represents one cell. The cells corresponding to each cluster are grouped, as indicated by the colored bars. The top marker genes were defined by their adjusted  $p$  value and  $\log_2$  (fold difference) on differential expression analysis (expression in a cluster versus expression in all other clusters). Genes with adjusted  $p = 0$  and  $\log_2$  (fold difference)  $\geq 0.5$  in any cluster are shown. **(B)** Dot plot of the top three marker genes for each neutrophil cluster, showing the average expression level and the percent of cells expressing the gene in each cluster. **(C)** Venn diagram displaying the intersection of the top genes in each cluster by absolute expression. **(D)** Violin plot showing the score per cluster for a panel of IFN-related genes, as described by Aran et al. (27). **(E)** Single-cell Western blot on 3300 neutrophils, with Abs against the proteins ISG15 and IFITM3. A representative blot is shown on the left, and a bivariate plot of the estimated single-cell abundances (peak areas) is displayed on the right. **(F)** Neutrophil single-cell RNA expression of the same targets as in (E): ISG15 and IFITM3. **(G)** Violin plot of the number of genes per cell in each cluster (left) and distribution of the number of genes per cell on the UMAP projection (right). **(H)** Ridge plots showing the distribution of CD10, CD15, and CD66b surface protein expression among cells in each transcriptional cluster. Surface expression and RNA-seq were measured simultaneously, by Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq). Data for 11 additional neutrophil surface markers are shown in Supplemental Fig. 2.



a less distinct pattern of expression: contrary to other clusters, none of the top expressed genes in Nh1 are uniquely expressed in that cluster (Fig. 3C). Nh2 neutrophils, which represent ~14% of circulating neutrophils (mean: 13.6%, range: 5.8–41%), are characterized by higher expression of two specific long noncoding RNAs (*MALAT1* and *NEAT1*) and of the gene encoding the G-CSF receptor (*CSF3R*), relative to other clusters (Fig. 3B). Finally, Nh3

neutrophils, which correspond to ~7% of circulating neutrophils (mean: 7.2%, range: 3.8–12.6%), represent a very distinct cellular state, with substantially higher levels of expression of type I IFN-inducible genes, including *HERC5*, *IFI16*, *IFIT1*, *IFIT2*, *IFITM2*, *IFITM3*, and *ISG15* (Fig. 3A–D). Given that the marker genes for Nh3 neutrophils are primarily protein-coding genes expressed at very low levels in any of the other clusters, we tested whether this



IFN-regulated gene-high neutrophil phenotype was also detectable at the protein level. We performed single-cell Western blotting in purified neutrophils, using Abs recognizing ISG15 and IFITM3 (Fig. 3E). We found discrete sets of neutrophils that express these proteins at high levels, and the proportion of cells in which one or both proteins is detectable is within the percentage range for Nh3 neutrophils calculated from the gene expression data (Fig. 3E, 3F). Interestingly, most of the cells that express both of the ISG15 and IFITM3 transcripts are from the Nh3 cluster (Fig. 3F), indicating the enrichment for IFN-related genes in that cluster.

We compared the distribution of the number of genes detected in cells from each cluster and found that Nh2 neutrophils have a substantially lower number of genes per cell (Fig. 3G). We then asked whether this difference was the result of a true biological difference between the cells in that cluster or an artifact of the clustering algorithm, whereby cells with lower read counts were classified as a distinct group. To test the latter hypothesis, we performed a down-sampling analysis, in which we reran the entire analysis pipeline, but reducing the input reads in one of the samples by 50%. If the Nh2 cluster was in fact simply the result of cells with lower read counts being clustered together, then we would expect the down-sampling of input reads to result in a higher proportion of Nh2 neutrophils. We found no change in the proportion of Nh2 neutrophils after down-sampling, in the reduced sample or overall (Supplemental Fig. 2), indicating that this cluster is unlikely to represent a clustering artifact driven by cells with lower read counts and instead more likely represents a distinct cluster of neutrophils with higher expression of specific genes (Fig. 3B) but lower overall transcriptional output.

As with other hematopoietic cell types, neutrophils have been studied and classified almost exclusively in terms of discrete cell-surface markers measurable by flow cytometry or immunohistochemistry. To test whether the observed transcriptional subsets correlate to surface expression of one or more of the canonical proteins that have been used to characterize and group neutrophils, we performed Cellular Indexing of Transcriptomes and Epitopes by Sequencing, with a custom panel of oligonucleotide-conjugated Abs targeting CD10, CD11b, CD11c, CD14, CD15, CD16, CD24, CD33, CD35, CD45, CD66b, CD107a, CD184, and HLA-DR. This method allows simultaneous measurement of surface protein abundance and transcriptome characterization at the single-cell level (30). We found that the surface expression level of each of these proteins was similar among Nh0–Nh3 neutrophils (Fig. 3H, Supplemental Fig. 3), indicating that these four transcriptional subsets offer a view of circulating human neutrophil heterogeneity that is independent of the canonical surface proteins that are commonly used to define and characterize these cells.

#### *Nh2 and Nh3 cells are end points in the transcriptional trajectory of human neutrophils*

An important advantage of scRNA-seq is that it offers an opportunity to study cells along a range of transcriptional states, including those that fall between theoretically more stable end-point states. This has, in turn, offered the possibility of ordering single-cell states along pseudotemporal trajectories (pseudotime), which indicate how far a given cell has moved along a continuum of biological progress. We used the R package Monocle 3 (21) to construct a single-cell trajectory of circulating human neutrophils, with the immature (Nh0) neutrophils as the root. From this, it is evident that the Nh2 and Nh3 clusters represent distinct end points in the transcriptional trajectory of circulating neutrophils, while the Nh1 cluster represents an intermediate state (Fig. 4A). We then looked for genes that vary between clusters of circulating neutrophils and grouped these into modules that have a similar pattern of expression. This identified

five modules of coexpressed genes (Fig. 4B), which we mapped back to the trajectory map. Module 1 genes are most highly expressed in the immature (Nh0) neutrophil cluster, module 3 genes in the NEAT1/MALAT1 (Nh2) neutrophil cluster, and module 5 genes in the IFN (Nh3) neutrophil cluster. The genes in modules 2 and 4 are more highly expressed in the transitional (Nh1) neutrophil cluster, but they represent distinct regions along the trajectory: module 2 genes appear to characterize a transitional state between Nh0 and Nh2 neutrophils, whereas module 4 genes characterize a transitional state between Nh0 and Nh3 neutrophils (Fig. 4C).

Modules of coexpressed genes offer an opportunity to infer common transcriptional regulatory elements, without the assumptions and potential biases inherent to inference based on known functions or on genomic localization with respect to other genes or to DNA sequence motifs. To infer candidate transcription factors that regulate the sets of coexpressed genes in each neutrophil module, we used BART, a method that relies on experimental evidence of protein–DNA interactions for >400 known transcription factors across a variety of cell types (24). We then selected the transcription factors associated with *cis*-regulatory elements most likely to regulate the coexpressed genes from each module (Irwin–Hall  $p < 0.01$ ) and compared these across modules. The modules corresponding to Nh2 and Nh3 neutrophils have the highest number of predicted transcription factors uniquely associated with them (Fig. 4D). The transcription factors at the intersections of neutrophil modules are also informative, because the expression of genes encoding specific transcription factors varies along the transition from one neutrophil cluster to another. For example, the genes encoding the transcription factors FLI1, MAX, SPI1, and YY1 are expressed along the trajectory from the immature (Nh0) to the IFN (Nh3) states, but not in the MALAT1/NEAT1 (Nh2) neutrophil cluster (Fig. 4E, left). Similarly, the transition from the intermediate (Nh1) cluster to the IFN (Nh3) cluster is marked by increased expression of genes involved in NF- $\kappa$ B signaling (Fig. 4E, center). In contrast, the transition from the intermediate (Nh1) cluster to the NEAT1/MALAT1 (Nh2) cluster is characterized by increased expression of the genes encoding the transcriptional repressor FOXP1 and the methylcytosine dioxygenases TET2 and TET3. To validate the presence of these transcription factors at the protein level in circulating neutrophils from healthy donors, we analyzed mass spectrometry data from a recently published study of neutrophils obtained from five healthy donors (31). Most of the inferred transcription factors were detected at the protein level in bulk neutrophil preparations, with FCGR3A and FCGR3B as protein-abundance references (Fig. 4F).

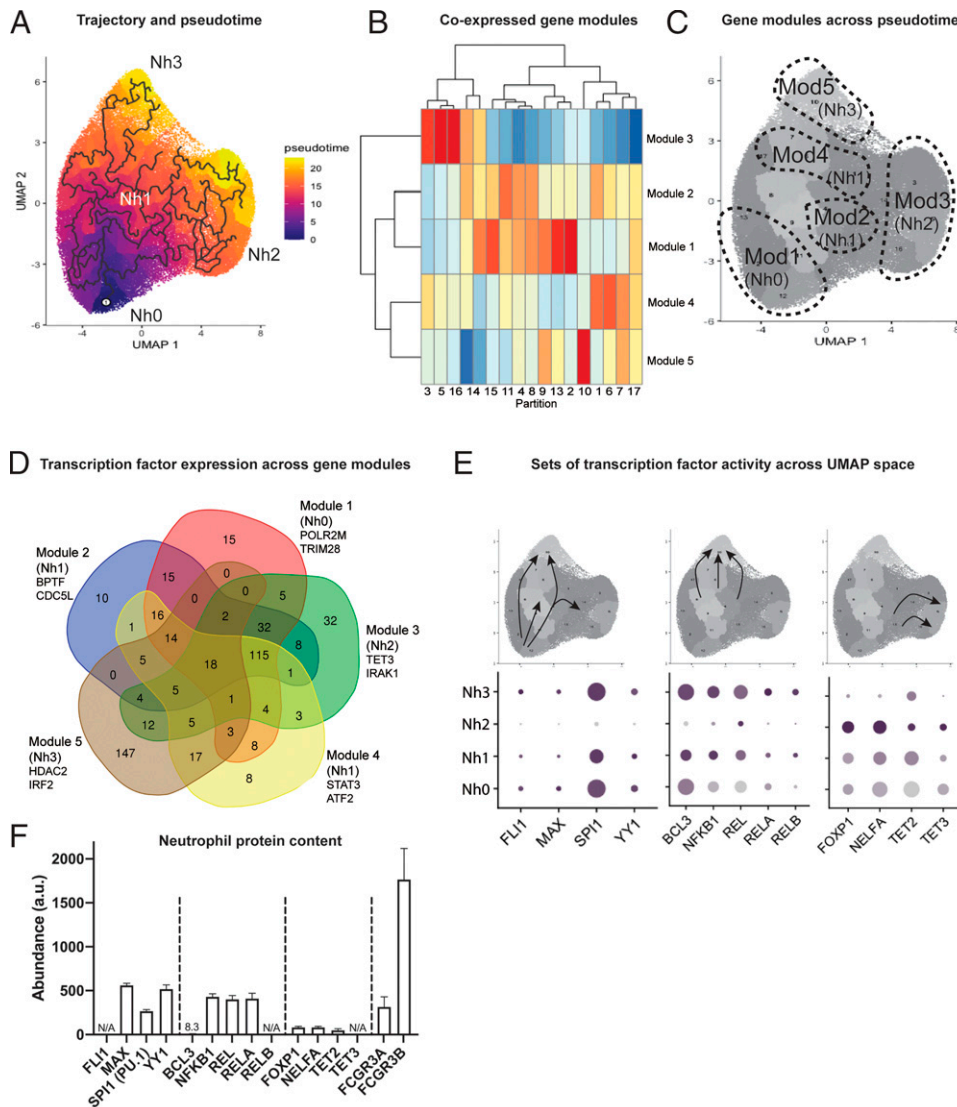
Our results support a model in which Nh2 and Nh3 cells represent end points in the transcriptional trajectory of circulating human neutrophils. Distinct sets of transcription factors, at least some of which are regulated at the level of transcription, orchestrate the transition from a less mature state (Nh0 cells) to one end-point state or the other, via an intermediate state (Nh1 cells) that corresponds to the majority of circulating neutrophils.

#### *A better understanding of neutrophil transcriptional states in healthy donors can facilitate the interpretation of data from patients*

An important potential advantage of a better understanding of neutrophil transcriptional states in healthy donors is the ability to contrast such states with those of neutrophils obtained from patients with different infectious or inflammatory diseases. To illustrate this, we applied our analysis pipeline and subset definitions to raw scRNA-seq data from low-density granulocytes (LDGs) obtained from patients with systemic lupus erythematosus (SLE) (32). These granulocytes layer with PBMCs in a density gradient, are prevalent in patients with autoimmune diseases like SLE, and have been associated with vasculopathy and immune stimulation. It could be



**FIGURE 4.** Nh2 and Nh3 cells are end points in the transcriptional trajectory of circulating human neutrophils. **(A)** Trajectory analysis showing the learned graph on the UMAP space with the pseudotime ordering by color. **(B)** Heatmap showing unsupervised classification of genes that vary across clusters of circulating neutrophils into five clusters of coexpressed genes. **(C)** Correspondence between the five modules of coexpressed genes and the four transcriptional clusters of circulating human neutrophils. **(D)** Venn diagram of transcription factors associated with *cis*-regulatory elements most likely to regulate the coexpressed genes in each module. Gene lists from the modules were used as input in BART. The overlap across modules for the top-ranking transcription factors (Irwin–Hall  $p < 0.01$ ) is shown. **(E)** Transcription factor gene expression changes along the transcriptional trajectory of circulating human neutrophils. Three patterns are shown: transcription factors expressed along the Nh0–Nh1–Nh3 trajectory, but not in Nh2 cells (left); transcription factors expressed in the transition from Nh1 to Nh3 cells (middle); and transcription factors expressed in the transition from Nh1 to Nh2 cells (right). **(F)** Mass spectrometry data from bulk neutrophil preparations obtained from five healthy donors, showing relative protein abundances for the transcription factors in (E), with FCGR3A and FCGR3B as abundance references.



reasonably hypothesized that LDGs would be strongly enriched for neutrophils in the Nh0 or Nh3 clusters. Integrating LDGs with healthy neutrophils, however, it is clear that they contain cells from all Nh clusters (Fig. 5A), and the proportions of LDGs in each Nh cluster are similar to those of healthy controls (Fig. 5B). Analyzing differentially expressed genes in a pseudobulk comparison, there is increased expression in the lupus LDGs of genes relating to IFN signaling and NF- $\kappa$ B signaling (Fig. 5C, 5D). This is consistent with the overactivity of type 1 IFN signaling that has been described by scRNA-seq mononuclear cells from patients with SLE (33). The large increase in the total IFN gene score seen in lupus LDGs (Fig. 5E, left) is apparent in all Nh subsets (Fig. 5E, right). These results indicate that LDGs in patients with lupus have similar relative representation of neutrophils in different transcriptional states as neutrophils from healthy donors. They also indicate that the high expression of IFN-stimulated genes in lupus LDGs results from an overall increase in transcript abundance for these genes, rather than an overrepresentation of Nh3 cells among the low-density fraction.

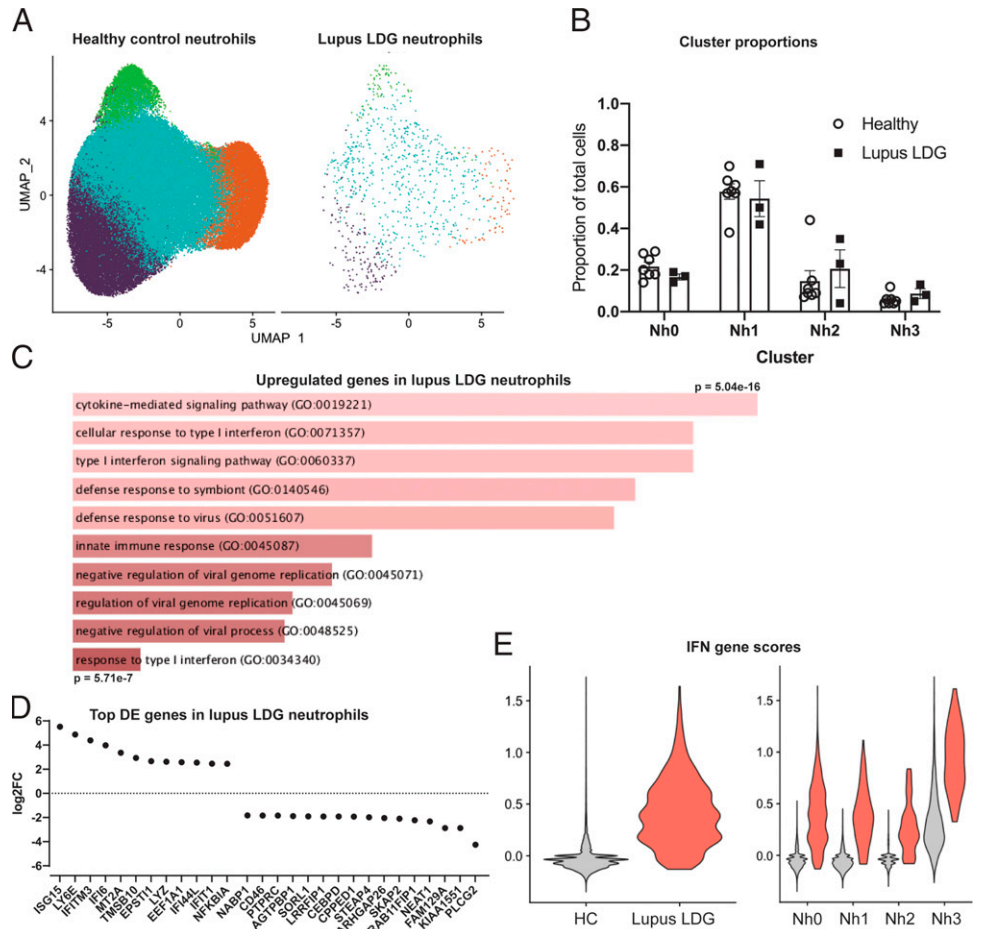
### Discussion

Our findings indicate that circulating human neutrophils are transcriptionally heterogeneous cells, which can be classified based on their transcriptional state into one of four clusters (Nh0–Nh3) that

are highly reproducible among healthy human subjects. We demonstrate that neutrophils transition transcriptionally from relatively immature (Nh0) cells, through an intermediate phenotype (Nh1), into one of two end points defined by either relative transcriptional inactivity (Nh2) or higher expression of IFN-induced genes (Nh3). More broadly, our findings demonstrate the feasibility of applying scRNA-seq to the study of human neutrophils obtained by different methods, by means of a modified analysis pipeline that significantly improves the identification of neutrophils in scRNA-seq datasets.

Recent studies have applied scRNA-seq to the study of murine neutrophil development in states of health or experimental infection (12, 13), and have found clear evidence of neutrophil transcriptional heterogeneity. One of these studies also analyzed CD33<sup>+</sup> cells sorted from whole blood from a human donor (12), while the other analyzed a publicly available scRNA-seq dataset generated from human bone marrow neutrophils as part of the Human Cell Atlas (13), suggesting that human neutrophils also exhibit distinct patterns of transcriptional heterogeneity. Our group and others have also provided recent evidence for transcriptional subsets of human neutrophils in scRNA-seq studies of sex differences in neutrophils obtained from healthy donors (14) and in patients with lung cancer (15) or COVID-19 (16, 17). However, because of their lower RNA content relative to other cell types, scRNA-seq with human

**FIGURE 5.** LDGs from SLE patients show upregulated IFN-induced gene expression but normal proportions of Nh clusters. **(A)** UMAP of healthy control ( $n = 7$ ) neutrophils integrated with SLE LDGs ( $n = 3$ ), split between healthy control and lupus cells. **(B)** Nh cluster proportions in healthy donor neutrophils and lupus LDGs. **(C)** Gene Ontology terms for the top 50 upregulated genes in a pseudobulk comparison between healthy neutrophils and lupus LDGs. **(D)** Top upregulated and downregulated genes based on  $\log_2$  (fold change), in a pseudobulk comparison between healthy neutrophils and lupus LDGs. **(E)** Total IFN-gene score for healthy control and lupus LDG (left) and IFN-gene scores by cluster (right).



neutrophils remains technically challenging and not well standardized, and it is common in human scRNA-seq studies for neutrophils to be missing or drastically underrepresented with respect to their expected proportions (17, 28, 34, 35). One possibility is that nucleases or proteases in neutrophil granules could interfere with the standard cell capture, cell lysis, or library preparation steps in scRNA-seq. However, after testing several modifications to the standard 10 $\times$  Genomics chemistry, we did not find a clear benefit to the addition of nuclease or protease inhibitors. Another possibility is that the standard cell-calling algorithms that are routinely used by most laboratories are not optimal for the differentiation of neutrophils from the background distribution of empty capture beads, thus excluding most neutrophils from downstream analyses. We found this to be the most likely source of neutrophil underrepresentation and describe an alternative approach to data analysis that departs from the raw matrix of UMIs associated with each barcode and considers the observed frequency distribution of features per barcode (genes per cell). This simple modification to the analysis pipeline significantly increases the inclusion of cells that, based on their transcriptional profile, clearly represent neutrophils.

We applied the modified analysis pipeline to the study of human neutrophils purified by immunomagnetic negative selection, with high viability and without evidence of early apoptosis. We analyzed 72,183 cells and found that circulating human neutrophils can be consistently clustered into four distinct transcriptional states, which we have classified as Nh0–Nh3. The global pattern of gene expression in Nh0 cells is similar to what has been described in bone marrow neutrophils (8, 13), with higher relative expression of various granule proteins and of several members of the S100 family. Trajectory analysis indicates that circulating neutrophils develop from this

relatively immature state into a transitional cluster, Nh1, which is transcriptionally the least distinct cluster and accounts for a majority of the captured cells (~60%). From this cluster, the developmental trajectory diverges toward one of two end-point states: the Nh2 and Nh3 phenotypes. Nh2 cells are characterized by higher relative expression of specific noncoding (*NEAT1*, *MALAT1*) or coding (*CSF3R*) RNAs, but have a lower overall transcriptional output than other neutrophils. Accordingly, they also have higher expression of genes encoding active regulatory elements that are associated with epigenetic modulation of transcription in neutrophil development, including *TET2* and *NELFA* (36). In addition, the gene encoding the transcription factor SPI1 (*PU.1*), which is a central factor in myeloid development (37), is highly expressed in all clusters except Nh2. This end point, therefore, likely represents the mature and transcriptionally quiescent state that has been classically associated with all circulating neutrophils. The IFN gene-expressing Nh3 cluster is transcriptionally quite distinct from the Nh2 state. Nh3 cells express more genes, they have increased expression of IFN-inducible genes that are not significantly expressed by any other neutrophil cluster, and based on our results, their expression of key regulatory transcription factors is also distinct. The transition from Nh1 to Nh3 is associated with increased expression of genes in the NF- $\kappa$ B family of transcription factors, which are known to play a role in the regulation of neutrophil activation, apoptosis, and NADPH oxidase activity (38–40). The existence of a subset of circulating neutrophils that expresses increased levels of IFN-inducible genes is now a well-validated finding in mouse and human (12–17), and we had previously shown that there are gender differences in the expression of the genes in this cluster (14). Our single-cell Western blot results indicate that this cluster is also likely to be detectable at the protein

level. It is still unknown whether these cells represent neutrophils that have encountered a specific stimulus *in vivo* or if they are epigenetically committed from a precursor state. In either case, the fact that the proportion of this cluster is relatively stable among healthy donors suggests that they represent a steady state rather than an incidental finding related to a recent exposure. More studies in humans will be necessary, but data from *Escherichia coli*-challenged mice suggest that the equivalent cluster of IFN-high neutrophils might have different bone marrow precursors than other neutrophils (12).

An important advantage of defining the transcriptional states of circulating neutrophils in healthy donors is the possibility of contrasting these states to those of neutrophils from patients or from human subjects exposed to environmental or pharmacological stimuli that might alter neutrophil biology. We illustrate this by analyzing publicly available data we had previously generated on LDGs from patients with SLE. We found that these cells have a preserved distribution of Nh0–Nh3 neutrophil transcriptional states, and that the high expression of IFN-related genes in lupus LDGs occurs in all clusters and is not the result of expansion of the Nh3 cluster. Thus, by providing a careful description of neutrophil transcriptional states in healthy humans, we anticipate that our findings will contribute to the broader goal of understanding neutrophil heterogeneity in different contexts.

The limitations of this work can be considered in two categories. First, there are limitations related to the current state of scRNA-seq technology and data analysis methods. As with all available scRNA-seq technologies, we rely on a very shallow sampling of the transcriptome of any given cell (50,000 reads per cell in our case, but in many studies half of that or less). Data analysis methods in scRNA-seq also rely on linear (principal components analysis) and nonlinear (UMAP or *t*-distributed stochastic neighbor embedding) reductions from a high-dimensional ambient space into two-dimensional representations, with inevitable loss of potentially important relations between cells. The choice of clustering algorithms and parameters can also drastically affect the results, which highlights the need for standardized methods and clear reporting. Second, there are limitations related to the scope of our experiments. We focus on a single scRNA-seq chemistry (10× Genomics), which, although highly prevalent, is not the only one available. The extent to which our modified analysis pipeline can be extrapolated or adapted to other chemistries remains to be determined. Our study is also limited to circulating human neutrophils, which are of obvious biological importance but represent a minority of total neutrophils. Finally, the transcriptional subsets we describe appear to be offering a view of neutrophil heterogeneity that is independent of the very limited one afforded so far by a small set of cell-surface markers. There is, at this time, no reliable way to sort neutrophils based on transcriptional signatures while preserving viability. Therefore, experimental characterization of possible functional differences among the transcriptional subsets we have described is an important future goal.

Based on our results, we propose that human circulating neutrophils are transcriptionally dynamic cells that develop from a less mature state into one of two distinct transcriptional phenotypes that cannot be defined by common surface markers. We also propose that a modified analysis pipeline is necessary for proper representation of neutrophils in scRNA-seq studies. We hope that these findings will pave the way for better representation of neutrophils in scRNA-seq studies, to a better understanding of neutrophil heterogeneity, and to additional studies exploring the behavior of these transcriptional neutrophil subsets over time (circadian variation or variation over the human life span), in response to environmental or pharmacological stimuli, or under different pathological conditions.

## Acknowledgments

This study used the high-performance computing clusters of the National Institute of Allergy and Infectious Diseases Office of Cyber Infrastructure and Computational Biology and of the National Institute of Arthritis and Musculoskeletal and Skin Diseases Office of Science and Technology. We thank Thomas Lewis at the National Institutes of Health Clinical Center's Department of Transfusion Medicine for support in the obtention of human samples. We thank James Simone at the Flow Cytometry Section of the National Institute of Arthritis and Musculoskeletal and Skin Diseases Office of Science and Technology for technical expertise and assistance.

## Disclosures

The authors have no financial conflicts of interest.

## References

- Geginat, J., M. Paroni, S. Maglie, J. S. Alfen, I. Kastirz, P. Gruarin, M. De Simone, M. Pagani, and S. Abrignani. 2014. Plasticity of human CD4 T cell subsets. *Front. Immunol.* 5: 630.
- Muranski, P., and N. P. Restifo. 2013. Essentials of Th17 cell commitment and plasticity. *Blood* 121: 2402–2414.
- Sica, A., and A. Mantovani. 2012. Macrophage plasticity and polarization: *in vivo* veritas. *J. Clin. Invest.* 122: 787–795.
- Ng, L. G., R. Ostuni, and A. Hidalgo. 2019. Heterogeneity of neutrophils. *Nat. Rev. Immunol.* 19: 255–265.
- Silvestre-Roig, C., Z. G. Fridlender, M. Glogauer, and P. Scapini. 2019. Neutrophil diversity in health and disease. *Trends Immunol.* 40: 565–583.
- Marau, M., A. Gaillardet, A. Gally, P. Saas, and T. Cherrier. 2020. Human primary neutrophil mRNA does not contaminate human resolving macrophage mRNA after efferocytosis. *J. Immunol. Methods* 483: 112810.
- Itoh, K., K. Okubo, H. Utiyama, T. Hirano, J. Yoshii, and K. Matsubara. 1998. Expression profile of active genes in granulocytes. *Blood* 92: 1432–1441.
- Theilgaard-Mönch, K., L. C. Jacobsen, R. Borup, T. Rasmussen, M. D. Bjerregaard, F. C. Nielsen, J. B. Cowland, and N. Borregaard. 2005. The transcriptional program of terminal granulocytic differentiation. *Blood* 105: 1785–1796.
- Subrahmanyam, Y. V., S. Yamaga, Y. Prashar, H. H. Lee, N. P. Hoe, Y. Kluger, M. Gerstein, J. D. Goguen, P. E. Newburger, and S. M. Weissman. 2001. RNA expression patterns change dramatically in human neutrophils exposed to bacteria. *Blood* 97: 2457–2468.
- Kobayashi, S. D., J. M. Voyich, C. L. Buhl, R. M. Stahl, and F. R. DeLeo. 2002. Global changes in gene expression by human polymorphonuclear leukocytes during receptor-mediated phagocytosis: cell fate is regulated at the level of gene expression. *Proc. Natl. Acad. Sci. USA* 99: 6901–6906.
- Franco, L. M., M. Gadkari, K. N. Howe, J. Sun, L. Kardava, P. Kumar, S. Kumari, Z. Hu, I. D. C. Fraser, S. Moir, et al. 2019. Immune regulation by glucocorticoids can be linked to cell type-dependent transcriptional responses. *J. Exp. Med.* 216: 384–406.
- Xie, X., Q. Shi, P. Wu, X. Zhang, H. Kambara, J. Su, H. Yu, S.-Y. Park, R. Guo, Q. Ren, et al. 2020. Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. *Nat. Immunol.* 21: 1119–1133.
- Grieshaber-Bouyer, R., F. A. Radtke, P. Cunin, G. Stifano, A. Levescot, B. Vijaykumar, N. Nelson-Maney, R. B. Blaustein, P. A. Monach, and P. A. Nigrovic; ImmGen Consortium. 2021. The neutrotime transcriptional signature defines a single continuum of neutrophils across biological compartments. *Nat. Commun.* 12: 2856.
- Gupta, S., S. Nakabo, L. P. Blanco, L. J. O'Neil, G. Wigerblad, R. R. Goel, P. Mistry, K. Jiang, C. Carmona-Rivera, D. W. Chan, et al. 2020. Sex differences in neutrophil biology modulate response to type I interferons and immunometabolism. *Proc. Natl. Acad. Sci. USA* 117: 16481–16491.
- Zilionis, R., C. Engblom, C. Pfirschke, V. Savova, D. Zemmour, H. D. Saatchioglou, I. Krishnan, G. Maroni, C. V. Meyerovitz, C. M. Kerwin, et al. 2019. Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity* 50: 1317–1334.e10.
- Schulte-Schrepping, J., N. Reusch, D. Paclik, K. Baßler, S. Schlickeiser, B. Zhang, B. Krämer, T. Krammer, S. Brumhard, L. Bonaguro, et al.; Deutsche COVID-19 OMICS Initiative (DeCOI). 2020. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* 182: 1419–1440.e23.
- Combes, A. J., T. Courau, N. F. Kuhn, K. H. Hu, A. Ray, W. S. Chen, N. W. Chew, S. J. Cleary, D. Kushnoor, G. C. Reeder, et al. 2021. Global absence and targeting of protective immune states in severe COVID-19. [Published erratum appears in 2021 *Nature* 596: e8.] *Nature* 591: 124–130.
- Clark, R. A., and W. M. Nauseef. 2001. Isolation and functional analysis of neutrophils. *Curr. Protoc. Immunol.* 19: 7.23.
- Clancy, D. M., G. P. Sullivan, H. B. T. Moran, C. M. Henry, E. P. Reeves, N. G. McElvaney, E. C. Lavelle, and S. J. Martin. 2018. Extracellular neutrophil proteases are efficient regulators of IL-1, IL-33, and IL-36 cytokine activity but poor effectors of microbial killing. *Cell Rep.* 22: 2937–2950.
- Hao, Y., S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* 184: 3573–3587.e29.
- Trapnell, C., D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, and J. L. Rinn. 2014. The dynamics and



- regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32: 381–386.
22. Cao, J., M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, et al. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566: 496–502.
  23. Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008: P10008.
  24. Wang, Z., M. Civelek, C. L. Miller, N. C. Sheffield, M. J. Guertin, and C. Zang. 2018. BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics* 34: 2867–2869.
  25. Zheng, G. X. Y., J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8: 14049.
  26. Lun, A. T. L., S. Riesenfeld, T. Andrews, T. P. Dao, T. Gomes; participants in the 1st Human Cell Atlas Jamboree; and J. C. Marioni. 2019. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* 20: 63.
  27. Aran, D., A. P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R. P. Naikawadi, P. J. Wolters, A. R. Abate, et al. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20: 163–172.
  28. Liao, M., Y. Liu, J. Yuan, Y. Wen, G. Xu, J. Zhao, L. Cheng, J. Li, X. Wang, F. Wang, et al. 2020. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* 26: 842–844.
  29. Borregaard, N., O. E. Sørensen, and K. Theilgaard-Mönch. 2007. Neutrophil granules: a library of innate immunity proteins. *Trends Immunol.* 28: 340–345.
  30. Stoeckius, M., C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, and P. Smibert. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14: 865–868.
  31. Bashant, K. R., A. M. Aponte, D. Randazzo, P. Rezvan Sangsari, A. J. Wood, J. A. Bibby, E. E. West, A. Vassallo, Z. G. Manna, M. P. Playford, et al. 2021. Proteomic, biomechanical and functional analyses define neutrophil heterogeneity in systemic lupus erythematosus. *Ann. Rheum. Dis.* 80: 209–218.
  32. Carlucci, P. M., M. M. Purmalek, A. K. Dey, Y. Temesgen-Oyelakin, S. Sakhardande, A. A. Joshi, J. B. Lerman, A. Fike, M. Davis, J. H. Chung, et al. 2018. Neutrophil subsets and their gene signature associate with vascular inflammation and coronary atherosclerosis in lupus. *JCI Insight* 3: e99276.
  33. Nehar-Belaid, D., S. Hong, R. Marches, G. Chen, M. Bolisetty, J. Baisch, L. Walters, M. Punaro, R. J. Rossi, C.-H. Chung, et al. 2020. Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* 21: 1094–1106.
  34. Xie, X., M. Liu, Y. Zhang, B. Wang, C. Zhu, C. Wang, Q. Li, Y. Huo, J. Guo, C. Xu, et al. 2021. Single-cell transcriptomic landscape of human blood cells. *Nat. Sci. Rev.* 8: nwaal80.
  35. Mistry, P., S. Nakabo, L. O’Neil, R. R. Goel, K. Jiang, C. Carmona-Rivera, S. Gupta, D. W. Chan, P. M. Carlucci, X. Wang, et al. 2019. Transcriptomic, epigenetic, and functional analyses implicate neutrophil diversity in the pathogenesis of systemic lupus erythematosus. *Proc. Natl. Acad. Sci. USA* 116: 25222–25228.
  36. Li, C., Y. Lan, L. Schwartz-Orbach, E. Korol, M. Tahiliani, T. Evans, and M. G. Goll. 2015. Overlapping requirements for tet2 and tet3 in normal development and hematopoietic stem cell emergence. *Cell Rep.* 12: 1133–1143.
  37. Monticelli, S., and G. Natoli. 2017. Transcriptional determination and functional specificity of myeloid cells: making sense of diversity. *Nat. Rev. Immunol.* 17: 595–607.
  38. Khojraty, T. E., Z. Ai, I. Ballesteros, H. L. Eames, S. Mathie, S. Martín-Salamanca, L. Wang, A. Hemmings, N. Willemsen, V. von Werz, et al. 2021. Distinct transcription factor networks control neutrophil-driven inflammation. *Nat. Immunol.* 22: 1093–1106.
  39. Ward, C., E. R. Chilvers, M. F. Lawson, J. G. Pryde, S. Fujihara, S. N. Farrow, C. Haslett, and A. G. Rossi. 1999. NF-kappaB activation is a critical regulator of human granulocyte apoptosis in vitro. *J. Biol. Chem.* 274: 4309–4318.
  40. Anrather, J., G. Racchumi, and C. Iadecola. 2006. NF-kappaB regulates phagocytic NADPH oxidase by inducing the expression of gp91phox. *J. Biol. Chem.* 281: 5657–5667.