A mathematical framework for human neutrophil state transitions inferred from single-cell RNA sequence data

Gustaf Wigerblad[1,*], Jonathan Carruthers[2,*], Sumanta Ray[3,4], Thomas Finnie[2], Grant Lythe[5], Carmen Molina-París[5,6,†], Saumyadipta Pyne[1,3,7,†], and Mariana J. Kaplan[1,†]

[1] Systemic Autoimmunity Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS/NIH), Bethesda, MD, USA

[2] Analysis and Intelligence Assessment Directorate, UK Health Security Agency, Porton Down, UK

[3] Health Analytics Network, Pittsburgh, PA, USA

[4] Department of Data Science, The West Bengal University of Juridical Sciences, Kolkata, India

[5] Department of Applied Mathematics, School of Mathematics, University of Leeds, Leeds, UK

[6] T-6, Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA

[7] Department of Statistics and Applied Probability, University of California Santa Barbara, Santa Barbara, CA, USA

Corresponding authors: spyne@ucsb.edu, molina-paris@lanl.gov, mariana.kaplan@nih.gov

[*,†] These authors contributed equally.

Abstract

Neutrophils, the most abundant immune cells in the human circulation, play a central role in the innate immune system. While neutrophil heterogeneity is a topic of increasing research interest, few efforts have been made to model the dynamics of neutrophil population subsets. We develop a mathematical model to describe the dynamics that characterizes the states and transitions involved in the maturation of human neutrophils. We use single-cell gene expression data to identify five clusters of healthy human neutrophils, and pseudo-time analysis to inform model structure. We find that precursor neutrophils transition into immature neutrophils, which then either transition to an interferon-responsive state or continue to mature through two further states. The key model parameters are the transition rates (the inverse of a transition rate is the mean waiting time in one state before transitioning to another). In this framework, the transition from the precursor to immature state (mean time less than an hour) is more rapid than subsequent transitions (mean times more than 12 hours). Approximately a quarter of neutrophils are estimated to follow the interferon-responsive path; the remainder continue along the standard maturation pathway. We use Bayesian inference to describe the variation, between individuals, in the fraction of cells within each cluster.

Introduction

Neutrophils are the most abundant immune cell subset in the human circulation and play a central role in the rapid innate immune response against microbes and other danger signals [1]. Neutrophils are endowed with efficient strategies to neutralize micro-organisms, including phagocytosis, degranulation, and the formation of neutrophil extracellular traps (NETs). Conversely, neutrophil dysregulation or enhanced activation can contribute to auto-inflammatory and auto-immune diseases and promote end-organ damage, while decreases in neutrophil function or numbers can lead to enhanced risk for infections [2, 3]. The estimated half-life of human mature neutrophils in circulation is considered to be less than a day, which contributes to the challenges to fully characterize and study them [4].

Indeed, despite being such a crucial component of the innate immune system, neutrophil physiology and dynamics remain less well characterized than what is known for other immune cell types. This is likely due to technical difficulties in the study of neutrophils, stark differences in neutrophil behavior across certain species, and the short life in circulation of these cells. Recent technological advances allow for the investigation of transcriptional states of single neutrophils, which has begun to elucidate their heterogeneity and complexity [5, 4]. We recently showed that neutrophils can be described along a transcriptional continuum, with discrete clusters based on expression of certain genes [6]. However, accurate knowledge of their cell state dynamics remains incomplete, and is important to characterize the functionality of different neutrophil subsets.

Mathematical modeling has been used to describe the growth, differentiation, and decay of immune cell populations. These models are often based on different hypotheses about the biological mechanisms involved [7]. By comparing model predictions with experimental data, researchers cannot only evaluate the plausibility of these hypotheses but also guide the design of future experiments. So far, much of this work has focused on the generation and maintenance of immunological memory, particularly through models of T-cell populations [8]. This includes models of $CD8^+$ effector T-cell production, as well as more recent approaches that use single-cell clustering and pseudo-time trajectory analysis to track how tumor cells develop drug resistance [9, 10]. In contrast, relatively little effort has been made to mathematically model the transcriptional changes in neutrophils or to quantify their movement through the body beyond what has been observed *in vitro* [11].

While previous work [6] identified four transcriptionally distinct neutrophil states using single-cell RNA sequencing, it did not provide a quantitative model of their transitions or kinetics. Our study builds on this by introducing a fifth precursor state and developing an integrated mathematical and statistical framework that models the dynamic transitions between neutrophil states. This enables estimation of transition rates, as defined by the mathematical model, and provides a mechanistic interpretation of cell state progression, which was not captured in earlier analyses [6].

Methods

Analysis of scRNA-seq data

Human peripheral neutrophils were purified from adult healthy volunteers as described previously [6]. Cells where captured and libraries made using the 10X 3' 3.1 chemistry. Matrix files were generated using Cellranger 8.0.1 for downstream analysis. The Python package scanpy (version 1.9.4) was used to import matrix files. Cells with fewer than 10^2 genes or a mitochondrial gene proportion greater than 0.1 were removed. Genes that were expressed in fewer than 5 cells were also omitted from the subsequent analyses.

Healthy volunteer samples were integrated using scvi-tools (version 1.3) and the top 3×10^3 variable genes. This creates a low-dimensional, latent representation of the raw gene expression data that attempts to capture the true biological variation while correcting for batch effects. The latent representation is then used to construct a neighborhood graph that identifies cells that are similar to each other and forms the basis of the clustering. Embedding of this graph in two dimensions is performed using Uniform Manifold Approximation and Projection (UMAP) and allows for better visualization of the clusters. Initial Leiden clustering at a resolution of 0.5 identified eight clusters. Cells in three of these clusters were removed as contaminants due to low expression of *CSF3R*, *FCGR3B* and *NAMPT*, genes that are expected to be highly expressed in human neutrophils. The scVI model was retrained on cells in the remaining clusters and a resolution of 0.42 was used to identify five clusters associated with different neutrophil transcriptional states.

Mathematical model

Let $n_i(t)$ denote the concentration of circulating neutrophils in state i (with $1 \leq i \leq N$) at time $t \geq 0$, where N is the number of states identified from the clustering analysis of individual cells. The influx of precursor neutrophils (or precursor state), n_0, is assumed to occur with constant rate ϕ, and neutrophils in any state are lost from circulation at rate μ. A neutrophil transitions from state i to state j with rate $\xi_{i,j}$. The changes in neutrophil populations associated with each state may then be described by the following system of ordinary differential equations (ODEs):

$$\frac{dn_0}{dt} = \phi - (\mu + \xi_{0,1})\, n_0 \ , \qquad \frac{dn_1}{dt} = \xi_{0,1} n_0 - (\mu + \xi_{1,2} + \xi_{1,3})\, n_1 \ ,$$
$$\frac{dn_2}{dt} = \xi_{1,2} n_1 - \mu n_2 \ , \qquad \frac{dn_3}{dt} = \xi_{1,3} n_1 - (\mu + \xi_{3,4})\, n_3 \ ,$$
$$\frac{dn_4}{dt} = \xi_{3,4} n_3 - \mu n_4 \ .$$

This system reflects our hypothesis that precursor neutrophils, n_0, enter state 1 before differentiating into either state 2 or state 3. The population of neutrophils in state 3 then proceed to the most mature state, represented here by state 4 (see Figure 3).

For healthy individuals, we assume that these populations are in equilibrium (stable steady state), and define f_i to be the fraction of circulating neutrophils in state i; that is, $f_i = n_i / \sum_j n_j$. The transition rates

may then be written in terms of these steady state fractions, all of which are measurable, and the loss rate, μ, as follows

$$\xi_{0,1} = \frac{\mu(1-f_0)}{f_0}\,, \qquad \xi_{1,2} = \frac{\mu f_2}{f_1}\,, \qquad \xi_{1,3} = \frac{\mu(f_3+f_4)}{f_1}\,, \qquad \xi_{3,4} = \frac{\mu f_4}{f_3}\,. \tag{1}$$

Statistical model and inference

To understand the variability in transition rates across a population of individuals, we begin by using a Dirichlet distribution, $\mathrm{Dir}(\boldsymbol{\alpha})$, to model the population-level variability in steady-state neutrophil fractions. A random variable following a Dirichlet distribution takes values in the form of N-dimensional vectors, with the conditions that each entry of the vector is non-negative and all entries sum to one. Since these conditions apply to the steady-state fractions of neutrophils, the Dirichlet distribution provides a natural way to model the between-individual variability in these fractions.

The aim here is, therefore, to estimate the parameters of the Dirichlet distribution, $\boldsymbol{\alpha}$, using the cluster fractions derived from the analysis of the single-cell sequencing data, our observed fractions. Let $\boldsymbol{F}$ be a $M \times N$ matrix of these observed fractions, where $M = 7$ is the number of subjects we have samples for and $N = 5$ is the number of neutrophil states. The mth row, $\boldsymbol{F}^{(m)}$, then contains the cluster fractions for individual m ($1 \leq m \leq M$). For a particular value of $\boldsymbol{\alpha}$, the likelihood of our observed fractions is given by

$$L(\boldsymbol{\alpha}; \boldsymbol{F}) = \prod_{m=1}^{M} g\left(\boldsymbol{F}^{(m)}; \boldsymbol{\alpha}\right)\,,$$

where g is the probability density function of the Dirichlet distribution.

The frequentist approach of maximum likelihood estimation looks to find a single $\boldsymbol{\alpha}$ that maximizes the above likelihood, or equivalently, the value of $\boldsymbol{\alpha}$ that makes the observed data most probable. Here, however, we adopt a Bayesian approach that treats our parameters as random variables. In this case, the distribution of $\boldsymbol{\alpha}$ in the presence of our data, or the posterior distribution, is assumed to be proportional to the product of our prior beliefs and the above likelihood [12]. Often, as is the case here, precise expressions for the posterior distribution are difficult to obtain, so computational methods such as Markov Chain Monte Carlo (MCMC) are used to instead sample from it. The posterior sampling is performed here using the emcee Python package, version 3.1.4 [13]. To construct distributions for the transition rates, $\xi_{i,j}$, $\boldsymbol{\alpha}$ are first drawn from the MCMC posterior sample. Then, for each draw, steady-state fractions f_i, $1 \leq i \leq N$, are sampled from the corresponding $\mathrm{Dir}(\boldsymbol{\alpha})$ distribution and these sampled fractions are substituted into Equation (1). In this way, we account for both the uncertainty in our estimate of $\boldsymbol{\alpha}$, and the variability in steady-state fractions across a population of individuals.

Results

Clustering of neutrophils from seven human samples at single-cell level has led to identification of five clusters corresponding to different transcriptional states. These clusters are depicted in Figure 1, where the UMAP

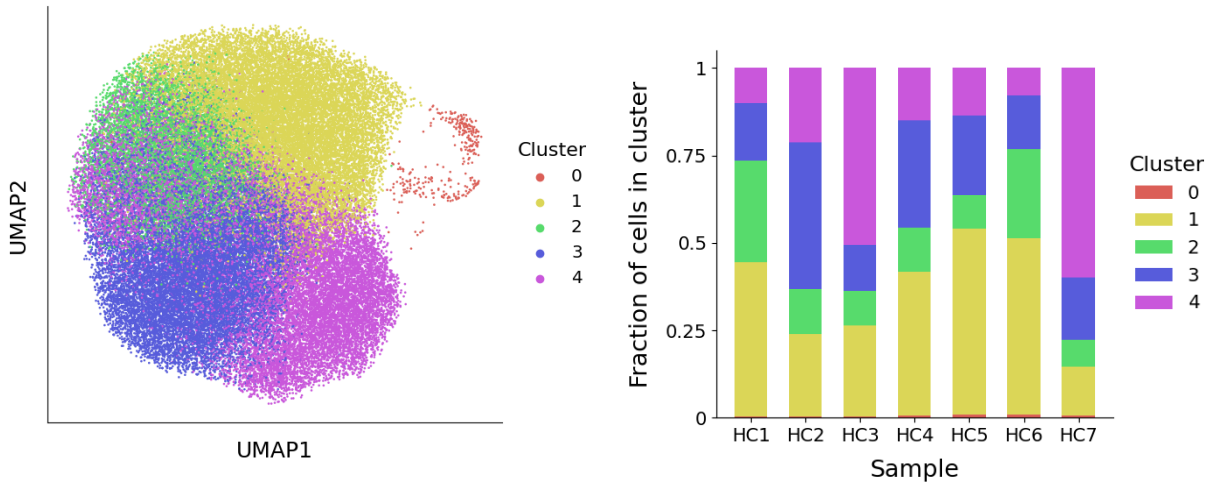

Figure 1: **Left**: A visualization of the five neutrophil clusters that uses UMAP to produce a two-dimensional representation of the raw gene expression data. **Right:** For each healthy subject (HC1-HC7), the fraction of neutrophils that belong to each cluster is depicted.

method of dimensionality reduction allows us to visualize the similarity between cells in two dimensions. Full details of the process used to derive these clusters is provided in the Methods section. To interpret the clusters, the average gene expression of cells in each cluster is evaluated for a subset of genes (Figure 2). This subset comprises the top marker genes for each of the clusters found here, as well as marker genes from a previous study of the same dataset [6].

Cluster 0 is characterized by elevated expression of ribosomal protein genes (*RPL, RPS*), suggesting these are precursor neutrophils that have recently entered circulation from the bone marrow. Cluster 1 shows higher expression of S100 family genes and likely represents immature neutrophils. Cluster 2 defines a distinct state, marked by up-regulation of genes associated with a type I interferon (IFN) response. Clusters 3 and 4 correspond to increasingly mature circulating neutrophils, distinguished by elevated expression of *CXCR2* and the long non-coding RNAs, *MALAT1* and *NEAT1*, respectively. This interpretation of neutrophil states aligns with our previous study, which identified four transcriptionally distinct clusters corresponding to clusters 1-4 described here [6].

To better understand neutrophil dynamics across the identified clusters, we developed a mathematical model based on a system of ODEs, a standard approach for describing how quantities evolve over time. In this model, each variable represents the concentration of neutrophils in one of the five transcriptional states. The concentration within a state can increase due to incoming cells from other states and decrease due to transitions to other states or exit from circulation. As such, the model requires a defined ordering of states.

Previously, pseudo-time analysis was used to order neutrophils along a biological continuum based on gene expression profiles, offering a plausible path through transcriptional states [6]. We used this pseudo-time ordering to inform the structure of our model (illustrated in Figure 3). In this framework, precursor neutrophils (N_0) enter the circulation from the bone marrow and transition into immature neutrophils (N_1).

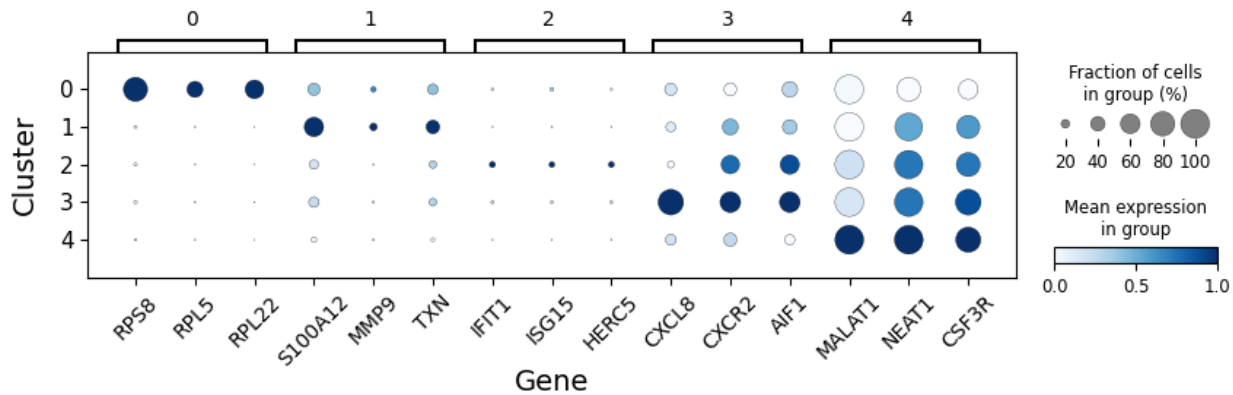

Figure 2: The average expression of select genes in neutrophils in each of the five clusters (states). These expression levels guide the interpretation of each cluster.

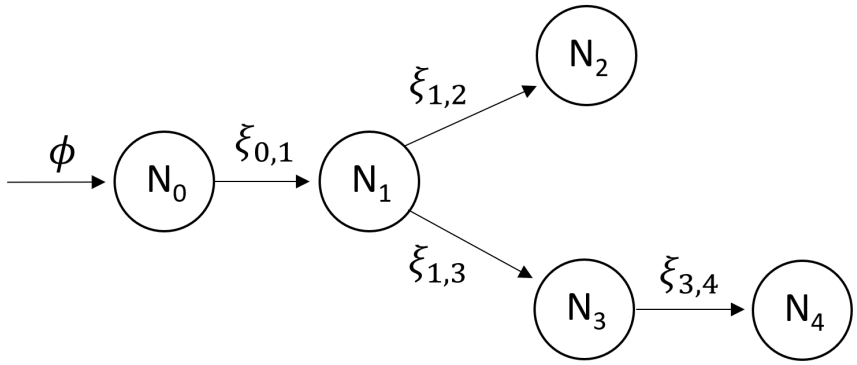

Figure 3: The proposed mathematical model for the transition of neutrophils between different transcriptional states. Neutrophils enter the circulation from the bone marrow at rate ϕ, with subsequent transitions between states marked by arrows. The state subscripts correspond to the cluster number and the parameter $\xi_{i,j}$ represents the rate at which a neutrophil transitions from state N_i to state N_j. Neutrophils in each state are cleared at rate μ, however, the arrows depicting these events are omitted from the diagram to improve clarity.

From there, cells either pass through an IFN-responsive state (N_2) or continue to mature through states N_3 and N_4. The model's key parameters are the transition rates, $\xi_{i,j}$, representing the rate at which neutrophils move from state N_i to state N_j. The inverse of these rates ($1/\xi_{i,j}$) corresponds to the average time spent in state N_i before transitioning to N_j. High transition rates imply short dwell times, whereas low rates indicate longer persistence in a state. To estimate these parameters, we assume that neutrophil populations in healthy individuals are at steady-state, meaning the number of cells in each state remains constant over time. Under this assumption, algebraic equations can be derived for the steady-state neutrophil fracions in each cluster, f_i, that depend solely on the transition rates and a fixed exit rate from circulation. This exit rate is set based on a presumed 12-hour half-life for circulating neutrophils [14]. We then estimate the transition rates of the seven healthy individuals by substituting their observed cell fractions (Figure 1, right) into these steady-state expressions.

Since cluster fractions vary across individuals, the corresponding estimates of transition rates also differ. However, due to the small sample size, these estimates alone do not capture how transition rates vary across the broader population. To address this, we developed a statistical model, described in detail in the Methods section, that provides us with predicted probability distributions for the steady-state cluster fractions (Figure 4). Using Bayesian inference, we obtain full posterior distributions, where each sample corresponds to a possible realization of cluster fractions. By drawing random samples from this statistical model, we simulate cluster fractions that are representative of a broader population. These samples are then used to infer corresponding probability distributions for the transition rates in the mathematical model (Figure 5). Together, the statistical and mathematical models link individual-level data to population-level estimates of neutrophil state transitions.

Median estimates from this combined framework suggest a rapid transition from precursor to immature neutrophils ($\xi_{0,1}$=4.9 h^{-1}), consistent with the low observed fraction of cells in cluster 0. Transition rates between subsequent states occur on slower but comparable time scales: ($\xi_{1,2}$=0.025 h^{-1}, $\xi_{1,3}$=0.076 h^{-1}, $\xi_{3,4}$=0.055 h^{-1}). From these estimates, the fraction of neutrophils that follow the IFN-responsive path (via cluster 2) can be calculated as: $\xi_{1,2}/(\xi_{1,2}+\xi_{1,3})$=0.24. This suggests that approximately 24% of neutrophils transition into the IFN-responsive state, while the remaining 76% continue along the standard maturation pathway.

Discussion

This study contributes to our understanding of neutrophil heterogeneity by introducing a novel mathematical framework to model transcriptional state dynamics in circulating neutrophils in adult healthy subjects. While mathematical models have been widely applied to study other immune cell populations, particularly T cells, relatively few efforts have addressed the transcriptional progression of neutrophils. Here, we leveraged single-cell RNA sequencing data from healthy individuals to identify five distinct transcriptional neutrophil states through clustering analysis. Building on previous work, and based on the expression of key marker genes, these clusters were interpreted to represent a progression from precursor cells to more mature states, with one cluster reflecting a type I IFN-responsive population [6, 15]

By integrating single-cell transcriptomic data with a system of ODEs, we provide a quantitative model for neutrophil state transitions. This model not only formalizes hypotheses about neutrophil maturation pathways but also enables the estimation of transition rates between transcriptional states. Despite a limited number of healthy subject samples, our statistical approach, based on Bayesian inference, allowed us to incorporate inter-individual variability and generate population-level estimates of state dynamics. This feature is especially valuable for extending insights beyond the specific individuals sampled and provides a framework that can scale with larger datasets of individuals affected by a number of disease states or physiological events [1, 3].

A key insight from the model is the fast transition rate from the precursor to immature state ($\xi_{0,1}$),

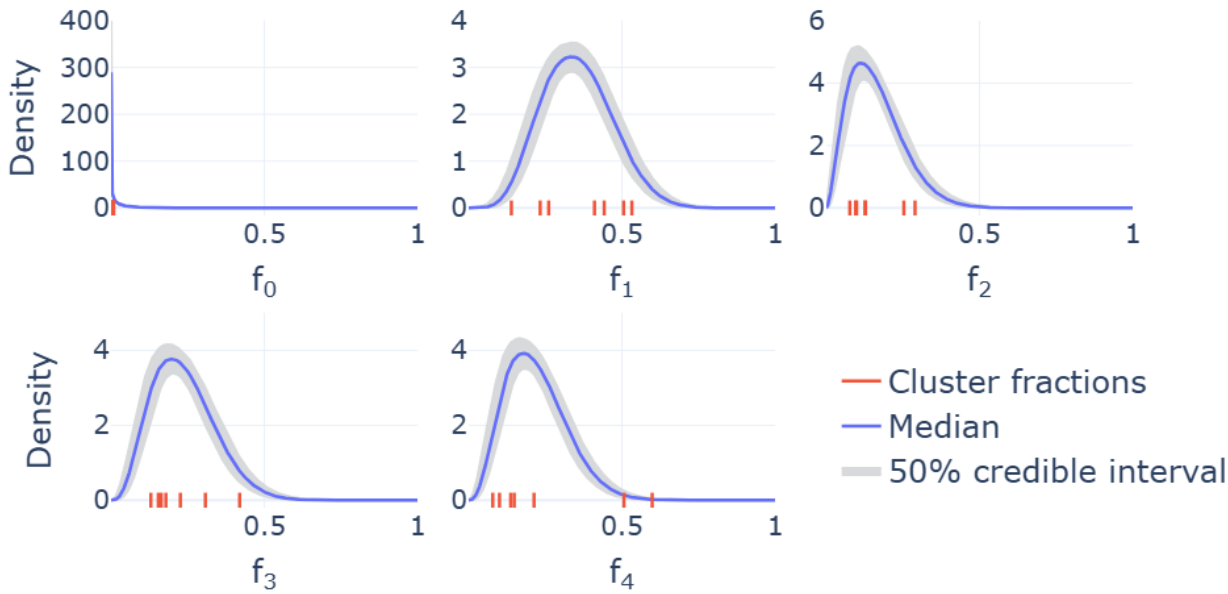

Figure 4: Marginal distributions describing the variability in cluster fractions across individuals. On each plot, the seven red ticks along the x-axis represent the cluster fractions for the seven individuals whose samples are included in the single-cell sequencing analysis. The median curves across these predicted densities are shown by solid lines, while the shaded 50% credible intervals indicate regions that half of these densities fall within.

consistent with the low fraction of neutrophils observed in the precursor cluster. However, we note that the inferred values for $\xi_{0,1}$ are highly sensitive to small variations in the precursor fraction (f_0), which is estimated from a sparse population ($\leq$1% of cells). Since $\xi_{0,1}$ is approximately equal to μ/f_0, small fluctuations in f_0 can lead to large changes in $\xi_{0,1}$. Thus, while the exact value of $\xi_{0,1}$ may be uncertain, the main conclusion, that this transition occurs rapidly, remains robust. In future studies, it may be reasonable to exclude the precursor cluster and return to the four-cluster model identified in previous work, particularly if more robust estimates of f_0 remain difficult to obtain [6].

We also assume a neutrophil clearance rate (μ) corresponding to a 12-hour half-life in circulation, consistent with established estimates. Since each transition rate in the model scales linearly with μ, uncertainty in μ propagates proportionally to the estimated rates. A +10% change in μ, for example, would result in a corresponding +10% change in all transition rate estimates. Moreover, the current statistical model assumes that all individuals are sampled from a single population. With a larger and more diverse dataset, it would be possible to incorporate covariates such as age, sex, or clinical status to refine these assumptions and better capture population-level variability.

Importantly, our model examines and assumes a steady-state neutrophil population (in the different states) for healthy individuals. Extending the framework to capture non-equilibrium dynamics, such as

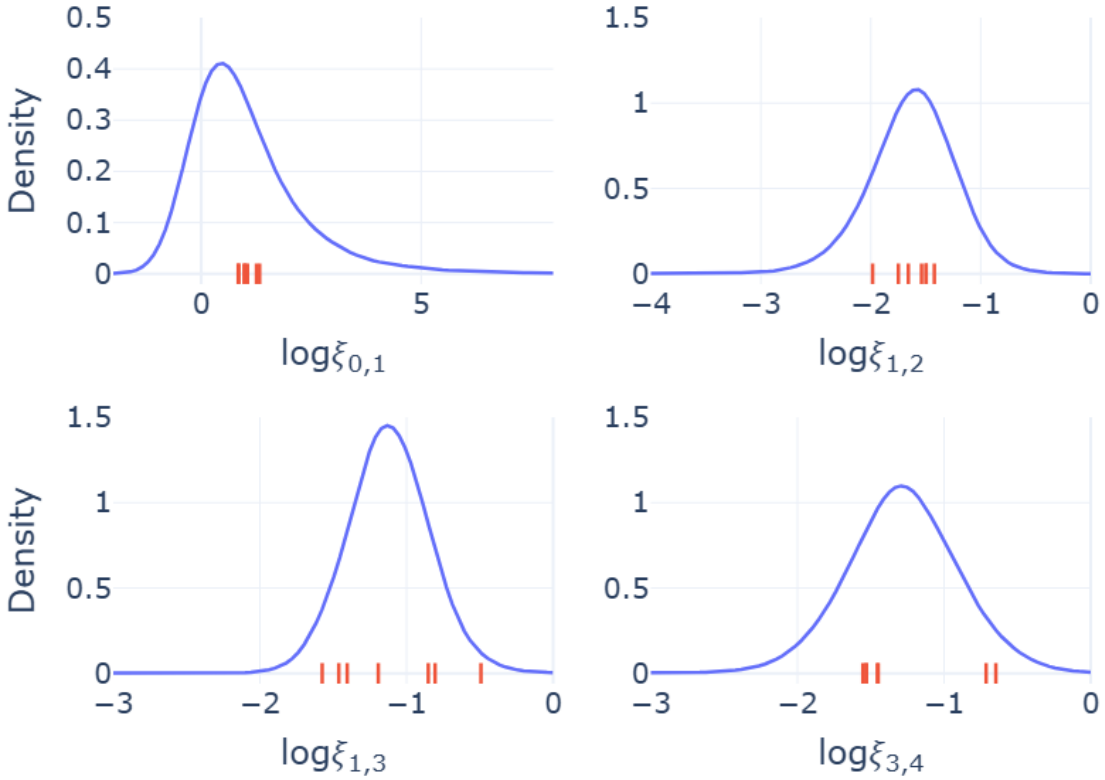

Figure 5: Probability distributions for the transition rates, $\xi_{i,j}$ between state i and state j. The red ticks on the x-axes indicate the values of each transition rate for the seven individuals included in the single-cell sequencing analysis. The solid lines are model-predicted densities and provide an indication of the likely values transition rates take across a broader population. Logarithmic values are in base 10.

those occurring during acute infection, chronic inflammation, or drug treatments, would represent a valuable future direction. In addition, while our model is built on transcriptional data, integrating complementary data types (*e.g.,* proteomics, epigenetics, functional assays, or tissue localization) would provide a more comprehensive view of neutrophil function and state transitions [11, 2]. We point out that our cluster interpretation is consistent with previous work [6], yet, the current study goes further by placing these clusters within a dynamic mathematical model. This framework provides not only a snapshot of neutrophil heterogeneity, but also insight into the timing and variability of transitions between states, offering a new dimension to understanding neutrophil maturation.

In summary, this study presents a foundational framework for modeling neutrophil transcriptional dynamics, addressing a significant gap in immunological modeling. Beyond improving our understanding of neutrophil behavior in healthy individuals, this approach lays the groundwork for investigating how neutrophil state dynamics are altered in disease contexts, including autoimmune and auto-inflammatory disorders, infections and metabolic diseases. Ultimately, the integration of mathematical modeling with high-dimensional

single-cell data can inform both mechanistic understanding and experimental design, advancing the study of neutrophil biology and its role in immune system function.

Acknowledgments

SP was supported by the NIH 2022 National Service DATA Scholarship. The study was also funded in part by the Intramural Research Program at NIAMS/NIH (ZIA AR-041199). The authors declare no conflict of interest.
This manuscript has been reviewed at Los Alamos National Laboratory and assigned report number LA-UR-24-27105.

References

[1] Garth Lawrence Burn, Alessandro Foti, Gerben Marsman, Dhiren Ferise Patel, and Arturo Zychlinsky. The neutrophil. *Immunity*, 54(7):1377–1391, 2021.

[2] Alejandra Aroca-Crevillén, Tommaso Vicanolo, Samuel Ovadia, and Andrés Hidalgo. Neutrophils in physiology and pathology. *Annual Review of Pathology: Mechanisms of Disease*, 19(1):227–259, 2024.

[3] Gustaf Wigerblad and Mariana J Kaplan. Neutrophil extracellular traps in systemic autoimmune and autoinflammatory diseases. *Nature Reviews Immunology*, 23(5):274–288, 2023.

[4] Miguel Palomino-Segura, Jon Sicilia, Iván Ballesteros, and Andrés Hidalgo. Strategies of neutrophil diversification. *Nature Immunology*, 24(4):575–584, 2023.

[5] Lai Guan Ng, Renato Ostuni, and Andres Hidalgo. Heterogeneity of neutrophils. *Nature Reviews Immunology*, 19(4):255–265, 2019.

[6] Gustaf Wigerblad, Qilin Cao, Stephen Brooks, Faiza Naz, Manasi Gadkari, Kan Jiang, Sarthak Gupta, Liam O'Neil, Stefania Dell'Orso, Mariana J Kaplan, et al. Single-cell analysis reveals the range of transcriptional states of circulating human neutrophils. *The Journal of Immunology*, 209(4):772–782, 2022.

[7] Rob J De Boer and Andrew J Yates. Modeling T-cell fate. *Annual Review of Immunology*, 41(1):513–532, 2023.

[8] Rustom Antia, Vitaly V Ganusov, and Rafi Ahmed. The role of models in understanding CD8+ T-cell memory. *Nature Reviews Immunology*, 5(2):101–111, 2005.

[9] H Hamlet Chu, Shiao-Wei Chan, John Paul Gosling, Nicolas Blanchard, Alexandra Tsitsiklis, Grant Lythe, Nilabh Shastri, Carmen Molina-París, and Ellen A Robey. Continuous effector CD8+ T cell production in a controlled persistent infection is sustained by a proliferative intermediate population. *Immunity*, 45(1):159–171, 2016.

[10] Shigeyuki Magi, Sewon Ki, Masao Ukai, Elisa Domínguez-Hüttinger, Atsuhiko T Naito, Yutaka Suzuki, and Mariko Okada. A combination approach of pseudotime analysis and mathematical modeling for understanding drug-resistant mechanisms. *Scientific Reports*, 11(1):18511, 2021.

[11] Wenbo Chen, Britton Boras, Tae Sung, Wenyue Hu, Mary E Spilker, and David Z D'Argenio. A whole-body circulatory neutrophil model with application to predicting clinical neutropenia from in vitro studies. *CPT: Pharmacometrics & Systems Pharmacology*, 10(7):671–683, 2021.

[12] Keegan E Hines. A Primer on Bayesian Inference for Biophysical Systems. *Biophysical Journal*, 108(9):2103–2113, 2015.

[13] Daniel Foreman-Mackey, David W Hogg, Dustin Lang, and Jonathan Goodman. emcee: the mcmc hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306, 2013.

[14] Erandi Pérez-Figueroa, Pablo Álvarez-Carrasco, Enrique Ortega, and Carmen Maldonado-Bernal. Neutrophils: Many ways to die. *Frontiers in Immunology*, 12:631821, 2021.

[15] Xuemei Xie, Qiang Shi, Peng Wu, Xiaoyu Zhang, Hiroto Kambara, Jiayu Su, Hongbo Yu, Shin-Young Park, Rongxia Guo, Qian Ren, et al. Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. *Nature immunology*, 21(9):1119–1133, 2020.