© 2023. Published by The Company of Biologists Ltd | Development (2023) 150, dev201280. doi:10.1242/dev.201280

RESEARCH ARTICLE

A dynamical systems treatment of transcriptomic trajectories in hematopoiesis

Simon L. Freedman[1], Bingxian Xu[2,3], Sidhartha Goyal[4,5,*] and Madhav Mani[2,3,6,*]

ABSTRACT

Inspired by Waddington's illustration of an epigenetic landscape, cell-fate transitions have been envisioned as bifurcating dynamical systems, wherein exogenous signaling dynamics couple to the enormously complex signaling and transcriptional machinery of a cell to elicit qualitative transitions in its collective state. Single-cell RNA sequencing (scRNA-seq), which measures the distributions of possible transcriptional states in large populations of differentiating cells, provides an alternate view, in which development is marked by the variations of a myriad of genes. Here, we present a mathematical formalism for rigorously evaluating, from a dynamical systems perspective, whether scRNA-seq trajectories display statistical signatures consistent with bifurcations and, as a case study, pinpoint regions of multistability along the neutrophil branch of hematopoeitic differentiation. Additionally, we leverage the geometric features of linear instability to identify the low-dimensional phase plane in gene expression space within which the multistability unfolds, highlighting novel genetic players that are crucial for neutrophil differentiation. Broadly, we show that a dynamical systems treatment of scRNA-seq data provides mechanistic insights into the high-dimensional processes of cellular differentiation, taking a step toward systematic construction of mathematical models for transcriptomic dynamics.

KEY WORDS: Differentiation, Bifurcation, Single-cell RNA-seq, Pseudotime, Waddington

INTRODUCTION

During development and tissue regeneration, it is envisioned that cells progress through multiple transitions to ultimately adopt a distinguishable function. Although each transition en route to a terminal fate involves the coordination of myriads of molecules and complex gene regulatory networks interacting with external factors, there is a common view that they depend on significantly fewer control parameters. This view was notably explicated by Conrad Waddington in an illustration of an epigenetic space as a tilted, bifurcating landscape, where a vast number of nodes (genes) provide the scaffold for the smooth hills and valleys (cell state) down which a pebble (cell) can reliably roll until it finds a resting position (terminal fate) (Fig. 1A) (Waddington, 1957).

Many of the characteristics of Waddington's landscape have been codified into the language of dynamical systems, including that cell fates resemble valleys (attractors) in gene expression or transciptomic space (Huang et al., 2005; Corson and Siggia, 2012; Slack et al., 1991; Camacho-Aguilar et al., 2021), that a small amount of stable states can emerge from large interconnected Boolean networks (Kauffman, 1969), and that known genetic interactions can yield multiple cell fates (bistability) (Huang et al., 2007; Weston et al., 2018). Waddington's illustration has also motivated analysis of the wealth of data captured in single-cell RNA-sequencing (scRNA-seq), in which the transcriptome of individual cells are measured, often at multiple time-points, as they differentiate. For example, fitting a mathematical model of a pitchfork bifurcation to scRNA-seq data yields predictions for developmental perturbations (Marco et al., 2014), reducing the dimensionality of large transcriptomic matrices can enhance the resolution of bifurcations to precisely determine the genes enabling a cell fate decision (Setty et al., 2016; Tusi et al., 2018), and well characterized cell-lineage relationships can be used to extract predictive models of gene regulation (Furchtgott et al., 2017; Qiu et al., 2020; Wang et al., 2022). While these studies generally characterize cell fate decisions as bifurcations of an underlying developmental landscape, other studies formulate cell fate transitions as stochastic jumps between co-existing states of a multimodal cell-fate landscape that can occur even in the absence of bifurcations, to infer lineage relationships and state transition probabilities (Weinreb et al., 2018a; Zhou et al., 2021; Lange et al., 2022).

In the face of these contrasting views, it remains unclear when, during development, transcriptomes undergo bifurcations and whether they can be identified purely from statistical analyses of single-cell expression data alone. To address these unknowns, we note that as a bifurcation is a qualitative augmentation of the steady state solutions, or branches, of a dynamical system that occurs as a control parameter varies, detecting bifurcations from transcriptomic data requires that steady states and control parameters exist, and that their dynamics can be identified from the data. We hypothesized an association between cell fates and transcriptomic steady states in scRNA-seq data, as the dynamic molecular processes that lead to transcriptomic changes, such as signal transduction and transcription, generally occur in the order of seconds and minutes (Shamir et al., 2016), whereas cell fates change over the course of hours or days (Slack et al., 1991), yielding a significant separation between the time scales of molecular mechanisms and data collection. We also hypothesized that inferred developmental time (pseudotime) could be used as a high resolution readout of a biological control parameter to pinpoint developmental bifurcations,

[1]Illumina, San Diego, CA 92122, USA. [2]NSF-Simons Center for Quantitative Biology, Northwestern University, Evanston, IL 60208, USA. [3]Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA. [4]Department of Physics, University of Toronto, Toronto, ON M5R 2M8, Canada. [5]Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5R 2M8, Canada. [6]Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208, USA.

*Authors for correspondence (goyal@physics.utoronto.ca; madhav.mani@gmail.com)

S.G., 0000-0002-7452-892X; M.M., 0000-0002-5812-4167

Handling Editor: Paul François
Received 9 September 2022; Accepted 11 April 2023

Development (2023) 150, dev201280. doi:10.1242/dev.201280

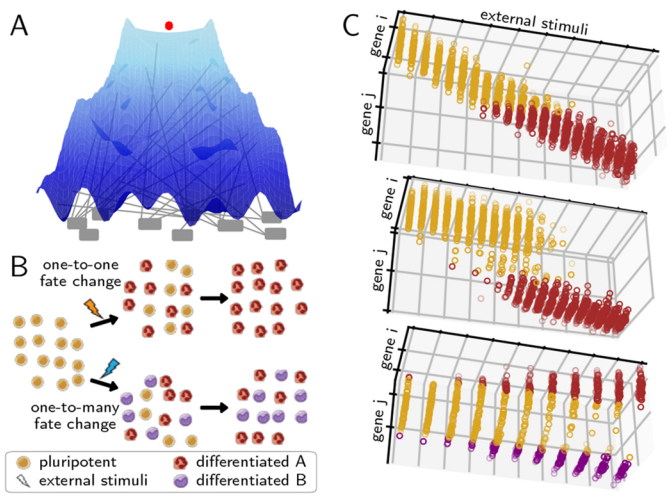

Fig. 1. Cell type differentiation as a dynamical process.
(A) Reimagination of Waddington's landscape of cell fate commitment in which cell fates are represented as valleys, commitment barriers as hills and gene activity as pegs underneath that control the heights of hills and valleys. (B) Schematic cell-population snapshots of maturation (top), in which one cell fate transitions to a different one, and a cell fate decision (bottom), in which a pluripotent cell differentiates to either of two lineages. Cell type images by A. Rad and M. Häggström. CC-BY-SA 3.0 license. (C) Gene expression trajectories for cells (dots) at varying levels of a differentiating stimuli for cases where the differentiation landscape does not bifurcate (top), undergoes a saddle-node bifurcation (middle) or undergoes a pitchfork bifurcation (bottom).

as it coincides well with intrinsic cellular dynamics, and may therefore correlate with known biological control parameters, such as morphogen concentration (Trapnell et al., 2014; Setty et al., 2016; Street et al., 2018).

Here, we use these hypotheses to lay out and demonstrate a statistical formalism for detecting and interrogating bifurcations in developmental fate transitions directly from transcriptomic pseudotime trajectories. Contrasting previous studies (Marco et al., 2014; Setty et al., 2016; Huang et al., 2007; Weston et al., 2018), we do not assume any specific mathematical form for the underlying genetic interactions, nor do we assume the shape, or even existence, of an underlying cell-fate landscape (Fig. 1B) (Chen et al., 2012; Mojtahedi et al., 2016) as it is not our goal to discern a specific model. Instead we rigorously query whether the necessary statistical signatures of bifurcations are present in a developmental timecourse. We build on and compare with similar styles of approach, which use correlation structure to detect signatures and molecular mechanisms of disease (Chen et al., 2012; Liu et al., 2012), analyze differentiation processes in temporal and pseudotemporal gene expression trajectories (Mojtahedi et al., 2016; Chen et al., 2018), and characterize reversibility in saddle-node bifurcations (Li et al., 2019). We show that our dynamical systems-driven approach enables us to distinguish between three different types of transcriptomic variation directly from systems-level data: a non-bifurcative cell fate change that is due to continuous changes in gene expression (Fig. 1C, top); a cell fate change that is due to a one-to-one state transition (Fig. 1C, middle), such as those that may occur during terminal-fate maturation (Ferrell, 2012); and a cell fate change that is due to a one-to-many state transition, e.g. those that occur when pluripotent cells decide between multiple cell lineages (Fig. 1C, bottom). We apply our framework to a class of *in silico* high-dimensional genetic networks to demonstrate its ability to recover the salient features of a bifurcating dynamical system, and examine the effects of high dimensionality and noise. We demonstrate the utility of our framework in the context of a recently published scRNA-seq exploration of hematopoiesis (Weinreb et al., 2020), and show that cell-fate bifurcations can be pinpointed and analyzed in scRNA-seq data, even without detailed knowledge of the dynamics and controls of the underlying system. Finally, we demonstrate that our framework allows us to identify a low-dimensional phase plane in which the dynamics unfolds, and can be used to distinguish new cellular clusters and extract genetic relationships that are pivotal to the bifurcative cell-fate change.

RESULTS

In this section, we show how the Continuous Time Lyapunov (CTL) equation (Appendix S1, section 1) can be used to investigate bifurcations in transcriptomic trajectories. An advantage of this framework is that we do not have to posit any specific functional form for the dynamical processes that yields a transcriptomic state or a shape, or even the existence of an underlying developmental landscape, only that the dynamical processes are (1) stochastic and Markovian (Rosenfeld et al., 2005; Raj and Van Oudenaarden, 2008; Gregor et al., 2007; Tkačik et al., 2008; Weinreb et al., 2018b; Zhou et al., 2021) ; and (2) occur at significantly faster timescales (seconds to minutes) than the timescales over which transitions in cellular fates are observed (hours to days) (Shamir et al., 2016; Slack et al., 1991). A consequence of these two assumptions (see details in the Materials and Methods section 'Continuous time Lyapunov equation for transcriptomic matrices') is that the local time evolution of the transcriptomic profile of a cell is controlled by a single matrix, the Jacobian ($\boldsymbol{J}$), where $J_{ij} = \partial \dot{g}_i / \partial g_j$ is the effect of the amount of gene j on the dynamics of gene i. Generically, the local geometry of a dynamical system can be obtained from its diagonalization, $\boldsymbol{J}=\boldsymbol{P}\Lambda\boldsymbol{P}^{-1}$, where Λ is a diagonal matrix of eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_{n_g}\}$ and $\boldsymbol{P^T}$ is the square matrix of eigenvectors $\{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_{n_g}\}$. If the system has a single stable transcriptomic state, then $\lambda_i < 0 \; \forall \; i$, in the same way that highly convex curvature is associated with a single fixed point (see Appendix S1, section 1 and Fig. S1, for an example). Conversely, if the system is undergoing a bifurcation, than the largest Jacobian eigenvalue, which we refer to as λ_d, and points in the $\vec{p}_d$ direction, must approach 0 from below, in the same way that flat curvature enables a fixed point exchange (Appendix S1, section 1). In the absence of a model, the eigenvalues and eigenvectors of the Jacobian are generally inaccessible from the CTL equation, even if the stochasticity is parameterized, as the covariance is a symmetric matrix and the Jacobian is asymmetric, yielding twice as many unknowns as there are equations. However, at a bifurcation, the CTL simplifies considerably, such that

$$\lim_{\lambda_d \to 0} C_{ij} \propto \frac{\vec{p}_d^{\,i} \vec{p}_d^{\,j}}{2\lambda_d}, \tag{1}$$

where C_{ij} is the covariance of gene i with gene j (Oku and Aihara, 2018). This simplification yields three key insights into the eigen-decomposition of the Jacobian directly from the eigen-decomposition of the covariance, $C_{ij} = \sum_{k=1}^{n_g} \omega_k \vec{s}_k^{\,i} \vec{s}_k^{\,j}$, where $\{\omega_1, \omega_2, \ldots, \omega_{n_g}\}$ are its eigenvalues and $\{\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_{n_g}\}$ are its eigenvectors. First, as all $\vec{s}_i$, by definition, normalize to 1, for at least one covariance eigenvalue (ω_1, without loss of generality):

$$\lim_{\lambda_d \to 0} \omega_1 = \infty, \tag{2}$$

i.e. the covariance diverges along the principal direction $\vec{s}_1$. Second, it

Development (2023) 150, dev201280. doi:10.1242/dev.201280

can be shown (see Materials and Methods section 'Bifurcation eigenvector equivalence') that

$$\lim_{\lambda_d \to 0} \vec{s}_1 = \pm \vec{p}_d, \tag{3}$$

meaning the direction of maximal covariance is identical to the direction of the bifurcation! Third, a direct result of Eqn. 1 is that

$$\lim_{\lambda_d \to 0} \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} = \pm 1, \tag{4}$$

meaning the Pearson's correlation coefficient $R_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$ of the data along axes i and j becomes maximal, provided that their corresponding loadings on the eigen-vector ($\overrightarrow{p_d}^{i,j}$) are non-zero.

The correlation structure and expansion of correlation coefficients at a bifurcation (Eqn. 4) has been used previously in the theory of Dynamical Network Biomarkers, or DNB (Chen et al., 2012), to explore bifurcations in cases where it can be determined which state variables are mechanistically involved in the bifurcation, including single-cell data (Mojtahedi et al., 2016). In our analysis, we analytically (Appendix S1, section 2) and empirically compare with this method, but note that unlike DNB analysis, it is not necessary to delineate which genes drive the bifurcation, as Eqn. 2 is a global feature of the covariance. Additionally, in contrast to previous studies that focus on correlation coefficients, we explore how covariance eigenvectors (Eqn. 3) provide direct insight into the underlying mechanisms driving developmental bifurcations.

Thus, three specific changes to the transcriptomic covariance data, Eqns. 2-4, that can be determined from observations of state variables, can inform us of the salient features of the system, its bifurcations, even when we have no direct access to the generative model for the dynamics or to its corresponding underlying geometry. Notably, these features rely only on the transcriptomic data being sampled from the vicinity of a steady state, and do not rely on special circumstances, such as the Jacobian being symmetric, or on the noise being of a particular nature. We first use theoretical models of noisy, high-dimensional genetic networks to demonstrate how this approach can be leveraged to detect and assess bifurcations of an underlying dynamical system from observations of state-variables alone (for a procedural outline, see Materials and Methods section 'Analysis pipeline'). Following this, we emphasize the power of this approach by directly applying it to scRNA-seq data for the neutrophil lineage in the hematopoietic system.

Covariance analysis recovers salient features of a high-dimensional *in silico* gene regulatory network

To better understand our mathematical framework in the context of scRNA-seq data, where the large number of discordant genes and biological noise may obfuscate the predicted covariance signal that is indicative of a bifurcation, and cell fate changes may take different geometric forms, we tested the framework on a noisy, high-dimensional, gene-regulatory network (GRN), illustrated in Fig. 2A. The deterministic aspects of this GRN are governed by a set of explicit ordinary differential equations, $\dot{\boldsymbol{G}} = \boldsymbol{F}(\boldsymbol{G})$, and stochasticity is incorporated by simulating the GRN with Poissonian noise (see Materials and Methods section 'Simulation methodology'). In the GRN, cell fate transitions result from two mutually inhibiting 'driver' genes, g_1 and g_2, via their dynamics:

$$\dot{g}_1 = \frac{m_1}{(1+g_2^2)} - k_D g_1, \tag{5}$$

where k_D are their degradation rate, and $m_{1,2}$ determine the scales of their synthesis (Gardner et al., 2000). Varying the control parameter m_1 yields a saddle-node bifurcation in gene-expression while varying k_D yields a pitchfork bifurcation (see Appendix S1, section 3 and Fig. S2). Similar networks have been analyzed to provide insight into gene inhibition and activation (Gallivan et al., 2020), and into a diversity of biological systems, such as the lac-operon (Ozbudak et al., 2004) and cell-cycle control (Novak and Tyson, 1993).

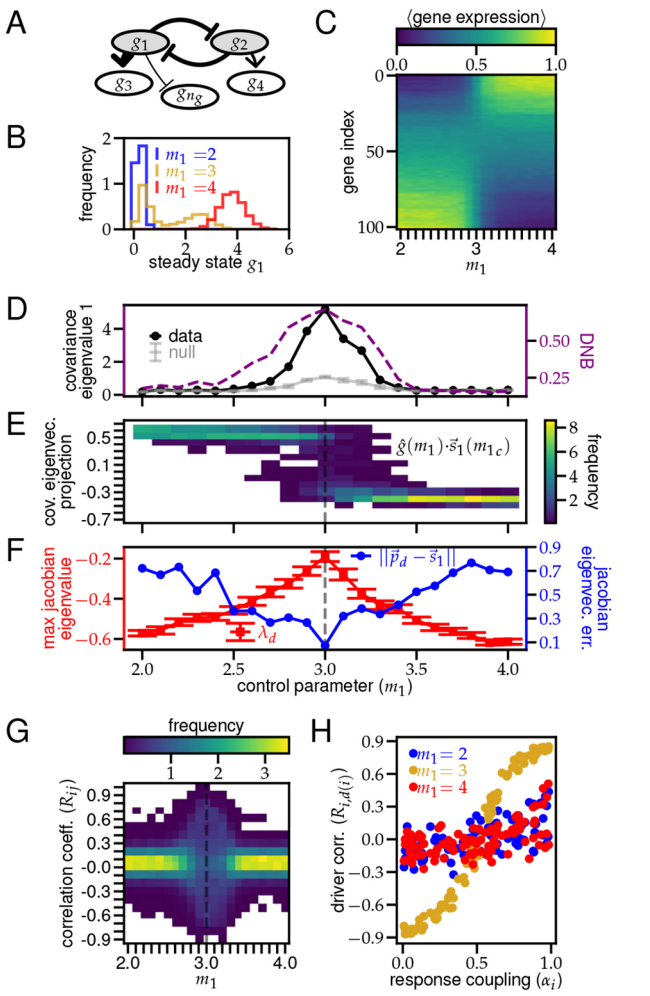

Fig. 2. Analysis of a gene regulatory network around a saddle-node bifurcation. (A) Schematic of the GRN for 5 of the 102 genes. Undisplayed nodes have a unidirectional arrow stemming from either g_1 or g_2. (B) Distributions of g_1 at steady state (see Materials and Methods section 'Simulation methodology') for three values of m_1. (C-G) GRN observations as a function of the bifurcating variable m_1 evaluated over a distribution of 100 cells (see Materials and Methods section 'Simulation methodology'). (C) Average final expression for each gene. Expression of driver genes, $g_{1,2}$ (bottom and top rows, respectively), are min-max normalized. Response genes are sorted by their corresponding driver [$\boldsymbol{d(i)}$] and activation level (α_i). (D) Black and gray indicate largest eigenvalue of covariance matrix shifted to have 0 min. Purple dashed line indicates DNB order parameter for gene expression matrix, where driver genes and tightly coupled responders $|\alpha_i - 0.5| < 0.25$ are in the DNB (Chen et al., 2012). (E) Distribution of normalized gene expression for each cell projected onto the bifurcating axis. (F) Red squares indicate the largest eigenvalue of the Jacobian matrix. Error bars are s.e.m. Blue circles indicate Euclidean distance from the corresponding Jacobian eigenvector ($\vec{\boldsymbol{p}}_{\boldsymbol{d}}$) to the principle covariance eigenvector $\vec{\boldsymbol{s}}_1$. (G) Distribution of Pearson's correlation coefficients for all gene pairs. (H) Correlation coefficients between responding genes and their driver as a function of the coupling coefficient $\boldsymbol{\alpha}$ (Eqn. 6) at three values of m_1. Each column in E and G integrates to 1.

Development (2023) 150, dev201280. doi:10.1242/dev.201280

As GRNs typically involve hundreds of genes, we include an additional n_g-2 genes in the network that respond variably to one of the two driver genes, according to

$$\dot{g}_i = \frac{\alpha_i g_{d(i)}^2 + 1 - \alpha_i}{1 + g_{d(i)}^2} - k_i g_i, \tag{6}$$

where $i \in [3, n_g]$, k_i is the degradation rate of i^{th} gene, g_i is its expression, $d(i)$ indicates the driver ($d(i)=1$ if g_i responds to g_1 and $d(i)=2$ if it responds to g_2) and $\alpha_i \in [0, 1]$ is the strength of the connection with its driver ($\alpha_i=0$ yields full inhibition and $\alpha_i=1$ yields full activation). Although this GRN can be made more complex, by including feedback from the responding genes to the two driver genes or by increasing the number of driver genes themselves, this simple model provides an interpretable demonstration of our proposed scheme.

We simulated this model for a fixed number of genes (n_g), statistical replicates or cells (n_c), noise scale (s), duration (N_t) and timestep (δt) for different values of the control parameters (m_1, k_D) (see Materials and Methods section 'Simulation methodology'). We define the $[n_c \times n_g]$ transcriptomic matrix $\boldsymbol{G}(m_1, k_D)$ once the system has reached steady state in the simulation. We observed that the steady-state distributions for individual genes (e.g. g_1, shown in Fig. 2B) shift their mean as the control parameter, m_1, is varied and exhibit bimodality at the bifurcation point, $m_1=m_{1c}=3$, as expected for saddle-node bifurcations.

Having verified that our model simulates a system that undergoes a high dimensional saddle-node bifurcation driven by a two-gene driver core, we used it to examine the effects of noise and a large number of responding genes on the theoretical predictions (Eqns 2-4). As predicted, we found that $\omega_1(m_1)$, the largest eigenvalue of the covariance of $\boldsymbol{G}(m_1)$, is maximal at the critical value m_{1c} (darker line in Fig. 2D), and the increase is significantly larger than can be obtained from a null distribution (lighter line in Fig. 2D) that lacks the correlations between genes of the model (see Appendix S1, section 4). This contrast between the data and the null can be understood by considering the bimodality of the transcriptomic distribution at the bifurcation. Far from the bifurcation, the transcriptomic distribution is unimodal, and all ω_i values scale with the noise scale s, which is undirected and therefore unaffected by resampling, yielding $\omega_1 \sim \langle \omega_1^{\text{null}} \rangle$ (Fig. S3, left and right panels). However, at the saddle-node bifurcation, the transcriptomic distribution is bimodal, so ω_1 scales with the distance between the two modes (Fig. S3, center top); marginal resampling of transcriptomes at the bifurcation yields new modes and the increased dimensionality of the bifurcation diminishes ω_1^{null}, compared with ω_1 (Fig. S3, center bottom). While Fig. S3 only demonstrates the bifurcation bimodality in $g_{1,2}$, the full transcriptomic bimodality can be visualized by computing $\hat{g}(m_1) \cdot \vec{s}_1(m_1 c)$, the normalized projection of the transcriptome of each cell along the principal covariance eigenvector. The distribution of this projection is densely centered around different fixed points to the right and left of m_{1c}, but widens significantly at m_{1c} as there is non-zero probability for both transcriptomic modes (Fig. 2E).

Although this toy model has an explicit bifurcation parameter, often the controls for specific developmental transitions are unknown in scRNA-seq data and the developmental time for each cell is inferred (pseudotime). To verify that similar spikes are expected even when the covariance is measured as a function of an indirect measure of a control parameter, such as pseudotime, we performed a pseudotime analysis on our toy model (Fig. S4A). In particular, we reduced the dimensionality of $\boldsymbol{G}$ using the SPRING method (Weinreb et al., 2018a), in which the (x, y) coordinates of each cell are determined by optimally placing each cell closest to its 4 nearest neighbors in the space of the top 10 gene-wise principal components (PC) of highly variable genes (Weinreb et al., 2018a), and computed their pseudotime using Slingshot (Street et al., 2018), in which pseudotime is approximated by the distance of the reduced-dimension data to a spline fit from one end of the data to the other (see Appendix S1, section 7.2 for details). We then binned the replicates (cells) by their pseudotime rank in 100 cell bins, and computed the principal covariance eigenvalue in each bin. We found that ω_1 exhibited a statistically significant spike precisely where the distance between the average control parameter in the pseudotime bin was closest to the critical parameter m_{1c} (Fig. S4B). This result suggests that our analytical framework may be directly applicable to detecting similar bifurcations in pseudotime-sorted scRNA-seq data.

Because, in this example, we have an explicit generative model ($\boldsymbol{F}(\dot{\boldsymbol{G}})$, given by Eqns 5 and 6), we can validate that just as m_{1c} resembles a bifurcation from analysis of the covariance matrix, it also resembles a bifurcation of the full noisy GRN, from analysis of the Jacobian. We show that the maximum negative eigenvalue (λ_d) of the Jacobian ($\boldsymbol{J}(m_1) = \frac{\partial \dot{\vec{g}}}{\partial \vec{g}}|_{\vec{g}(m_1)}$) for this network approaches 0 from below as $m_1 \to m_{1c}$ (Fig. 2F). We also show that at m_{1c}, the direction of maximal covariance, is given by the corresponding eigenvector of the Jacobian ($\vec{p}_d$), as the Euclidean distance between $\vec{s}_1$, the principal eigenvector of the covariance, and $\vec{p}_d$ approaches 0, as $m_1 \to m_{1c}$ (Fig. 2F). Thus, although the finite system size (n_c) prevents, or regularizes, ω_1 from diverging, and $|\vec{s}_1 - \vec{p}_d| > 0$, ω_1 is still at its largest and the eigenvectors are in closest correspondence at the bifurcation.

To empirically benchmark the principal covariance eigenvalue as a bifurcation indicator, we selected genes with strong connections ($|\alpha_i - 0.5| > 0.25$) and computed the DNB order parameter as a function of m_1 (see Appendix S1, section 2) (Liu et al., 2012; Chen et al., 2012; Li et al., 2019; Chen et al., 2018). We found that the variation of ω_1 matched the variation of the DNB order parameter (Fig. 2D), providing empirical support for using ω_1 to identify bifurcations. Importantly, computing ω_1 did not require preprocessing, while computing the DNB order parameter requires a preprocessing step to select the DNB genes (Chen et al., 2018). Additionally, ω_1 is computationally more efficient, as it is obtained via the singular value decomposition of the gene expression matrix, whereas the DNB order parameter is obtained via the correlation matrix across tens of thousands of genes (Hastie et al., 2009).

Although the covariance eigen-decomposition provides insight into the timing and direction of a bifurcation, Eqn. 4 predicts that the (Pearson) correlation coefficients between genes may help determine which genetic relationships are most critical for the dynamics at the bifurcation. We found that, for low and high m_1, when the network only has one fixed point, the distribution of correlation coefficients R_{ij} is strongly centered around 0 (Fig. 2G). However, at the bifurcation, this distribution spreads out to ± 1, as predicted in Eqn. 4 (Fig. 2G). To determine whether the gene pairs that yielded large R_{ij} corresponded with critical gene relationships in our network, we plotted $R_{i,d(i)}$: the correlation between all responder genes and their drivers, sorted by their connection strength α_i. We found that these correlation coefficients were much more strongly indicative of the responder-driver dependency (α_i) at bifurcation (Fig. 2H, green) rather than far away from the bifurcation (Fig. 2H, red and blue). Again, the GRN model makes it explicit that,

Development (2023) 150, dev201280. doi:10.1242/dev.201280

although the correspondence between geometry and dynamics is not universal, owing to the high-dimensionality of the system, in the vicinity of a bifurcation, a form of dimensionality reduction emerges that enables the gleaning of geometric characteristics directly from the dynamics. Thus, entries of a correlation matrix with high magnitude at a bifurcation may be reliable indicators of mechanistic gene-regulatory features.

We further used our GRN model to probe a pitchfork bifurcation induced by varying k_D (Fig. S5A). Unlike the example of a saddle-node bifurcation, we observed that ω_1 does not peak at the bifurcation parameter k_{Dc}=0.5, but rather begins to increase (Fig. S5B). This feature directly follows our interpretation that ω_1 corresponds to the distance between the two modes of transcriptomic distribution. Whereas the bimodality of the saddle-node bifurcation results from the discontinuous transition between states, the bimodality of pitchfork bifurcation emerges continuously from its root and becomes more pronounced as the control parameter is increased. Therefore, the distance between the modes (ω_1) increases with the control parameter. By clustering the cells according to their transcriptomic mode, or branches, we are able to recover the bifurcation signature predicted by Eqn. 2 (Fig. S5C), but we note that precise clustering requires prior knowledge (e.g. how many clusters there are).

As developmental decisions are often modeled as noise-induced state transitions between co-existing transcriptomic states (Weinreb et al., 2018b; Zhou et al., 2021), rather than bifurcations, we sought to determine whether our framework could distinguish these two possibilities. We used our model gene network (Fig. 2A) to explore the noise-induced transition possibility by varying the noise scale s of the network at fixed values of the bifurcation parameters (see Appendix S1, section 5 for details). We found that the principal covariance eigenvalue exhibited unique step-like dynamics as the stochasticity was varied (Fig. S6), which was unlike either the one-to-one (Fig. 2D) or one-to-many (Fig. S5B) bifurcation examples explored, demonstrating that analysis of the covariance dynamics in transcriptomics should enable distinguishing between noise-induced transitions and bifurcations.

In this example, we have demonstrated the applicability and power of the theoretical infrastructure outlined above to analyze a high-dimensional and noisy dynamical system undergoing a variety of bifurcations, by uncovering its crucial aspects, including its location, direction in gene space and influential genetic relationships. These calculations are also computationally simple; although covariance matrices can be cumbersome to compute for large numbers of genes and cells, reduced singular value decomposition can be used to determine quickly its largest eigenvalue and eigenvector, which is all our approach requires. Notably, our results apply only if the system is measured at steady state, otherwise there is no reason to anticipate clear divergences in the distribution of eigenvalues, transient bimodality or equivalence between the covariance and Jacobian principal directions (Fig. S7B-D).

Covariance analysis pinpoints a bifurcation in neutrophil development

Having verified that gene-gene covariance can be used to provide insight into transitions in a simulated genetic context, we applied our analysis framework to a recently published scRNA-seq data set of mouse hematopoietic stem cell (HSC) differentiation (Weinreb et al., 2020). In this experiment, HSCs were isolated *in vitro*, barcoded, plated in a media that supports multilineage differentiation (day 0) and subsequently sampled for single-cell sequencing using inDrops (Klein et al., 2015 ; Plasschaert et al., 2018) on days 2, 4 and 6. The resultant transcriptomic matrix (25,289 genes in 130,887 individual cells) was visualized in 2*D* using the SPRING method (Weinreb et al., 2018a) (Fig. 3A), using 4 nearest neighbors in the top 50 PC space for highly variable, non-cell cycle genes (Weinreb et al., 2018a). Each cell was then associated with one of 11 different cell types (annotations in Fig. 3A) based on its position in the SPRING plot and expression of cell type-specific marker genes (Weinreb et al., 2020). Cells that belonged to the developmental transition from multipotent progenitor (MPP) to neutrophil were identified by recategorizing cells as a cell-label distribution, and ranking cells by their similarity to fully committed neutrophils (see details in Appendix S1, section 7.1). The

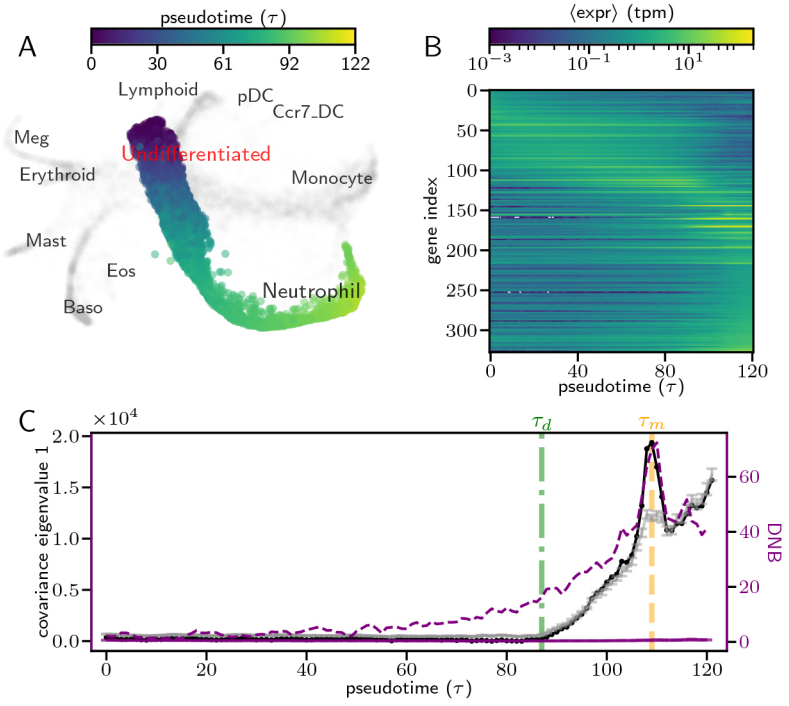

Fig. 3. Covariance analysis of temporal scRNA-seq data shows signatures of developmental bifurcations. (A) SPRING visualization for each cell (point) in an *in vitro* scRNA-seq experiment of mouse hematopoeitic stem cell differentiation (Weinreb et al., 2020). Cells in the pseudotime trajectory analyzed are colored accordingly (blue to yellow), whereas others are gray. SPRING coordinates, cluster labels, pseudotimes and a similar visualization were first reported by Weinreb et al. (2020). (B,C) Observations of neutrophil trajectory as a function of pseudotime calculated for each of 121 bins of pseudotemporally adjacent cells. All bins had 1000 cells except for the last one, which had 1310 cells, and had a 50% overlap with neighboring bins. (B) Average gene expression in pseudotemporal bins for highly expressed [max (⟨expr⟩)>1] and highly varying [coeff. of var.(⟨expr⟩)>0.5] genes. (C) Principal covariance eigenvalue (black) compared with a statistical null (gray, details in Appendix S1, section 4), shifted to have 0 min. Error bars of null are ±1s.d. DNB order parameter also shown, where the DNB comprises neutrophil marker genes (purple dashed line) or is the average of many random gene sets (purple solid line, see Fig. S9 for details). Green and yellow dashed lines indicate the developmental transition times τ_d and τ_m.

Development (2023) 150, dev201280. doi:10.1242/dev.201280

61,310 cells identified as belonging to the neutrophil transition were sorted into a neutrophil pseudotime trajectory (Fig. 3A) by ranking cells according to their similarity with the earliest pluripotent cells (see Appendix S1, section 7.1). This data-specific pseudotime algorithm was validated via the clonal barcodes of the cell, by ensuring that the MPP cells in the trajectory included neutrophil clones, and via the sequencing time, by ensuring that cells collected earlier were ranked earlier in the trajectory. Thus, several features of this trajectory make it ideal for applying our analysis framework: it includes a large number of cells, enabling statistically reliable covariance measurements; it is robust to the systematic, temporal controls of sequencing -time and cellular barcodes; and it is part of hematopoiesis, a well-characterized developmental process that enables the comparison of our findings with past work. We found that this trajectory was extremely dynamic, as the expression of hundreds of highly expressed genes is temporally variable, with large groups of genes either monotonically increasing or decreasing (Fig. 3B).

To determine whether the transitions from HSC to neutrophil were due to bifurcations in transcriptomic space, we split the neutrophil trajectory into overlapping bins of 1000 cells (last bin had 1310) and applied our covariance analysis to the full, row (cell)-normalized transcriptomic matrix at each bin $\boldsymbol{G}(\tau)$. We found that the largest eigenvalue of the covariance of the full gene expression matrix [$\omega_1(\tau)$, dark line in Fig. 3C] exhibited very little variation for $\tau<\tau_d$=85, but began to increase at τ_d and exhibited a significant spike at τ_m=109, which is indicative of a bifurcation. To determine whether ω_1 changes were statistically significant, we computed the corresponding statistical null (ω_1^{null}, lighter line in Fig. 3C; details in Appendix S1, section 4) and found that the large peak at τ_m was easily distinguishable from the null. To benchmark our result against established bifurcation identification methodologies, we computed the DNB order parameter (Chen et al., 2012) across pseudotime, using known marker genes for neutrophils and neutrophil progenitors (Weinreb et al., 2020) as the DNB (see details in Appendix S1, section 2). We found that the DNB order parameter (Fig. 3C, purple dashed line) exhibited similar dynamics to $\omega_1(\tau)$, but we emphasize that, unlike the computation of $\omega_1(\tau)$, computing the DNB requires gene filtering, as random sets of genes did not exhibit bifurcation signatures (Fig. 3C, purple solid line; Fig. S9). We also verified that bin size did not generally impact the dynamics of $\omega_1(\tau)$ (Fig. S10).

We first focus our attention at the dynamics at τ_m, after which we will address those observed at τ_d. As this pattern of a statistically significant spike following near-constant ω_1 echoed the observed behavior of a saddle-node bifurcation in our toy model (Fig. 2D), we speculated that at τ_m there was a one-to-one transcriptomic state transition. Additionally, to verify that the temporal trend in Fig. 2D was not limited to the diffusion-based pseudotime algorithm used by Weinreb et al. (2018a), we recalculated pseudotime using Slingshot (Street et al., 2018), and found the same rise and peak of ω_1 (see Appendix S1, section 7.2 and Fig. S11A,B).

We focus now on our observations in proximity to τ_d. As the increase in ω_1 at τ_d strongly resembled the pitchfork bifurcation of our toy model (Fig. S5B), as well as the proliferation of cell fates seen in high-resolution time-course scRNA-seq experiments (Nitzan and Brenner, 2021), we hypothesized that the increase of ω_1 at τ_d was also due to transcriptomic state changes. Further evidence of a developmental transition is that, at τ_d, the distribution of expression of each gene across cells begins to significantly shift toward higher values (Fig. S10C). However, the precise nature of this developmental transition is unclear, because, in contrast to our toy model, ω_1^{data} is nearly indistinguishable from ω_1^{null} during the increase.

To determine whether the transcriptomic state transitions we identified had biological significance, we compared our findings against the tree of cell fates for neutrophil development (Fig. 4A). We found (Fig. 4B) that τ_d, the moment ω_1 begins to increase, corresponded well with the moment in pseudotime that cells switch between the endpoints of this tree: from not expressing any terminal-fate marker genes to primarily expressing neutrophil cell-fate markers (Weinreb et al., 2020). At a more granular developmental level, the pseudotimes highlighted by our covariance analysis align with specific transitions between intermediate neutrophil progenitor states (Fig. 4A). These transitions include: (1) one-to-many cell fate changes (i.e. decisions), such as the transition between a granulocyte monocyte progenitor (GMP, or myeloblast) and any of its four terminal fates (neutrophil, monocyte, eosonophil and basophil); and (2) one-to-one cell fate changes (i.e. maturation), such as the transition between promyelocyte and myelocyte (Weinreb et al., 2020; Borregaard, 2010; Ostuni et al., 2016). τ_d corresponds well with the pseudotime at which promeylocyte marker genes are maximal (Fig. 4C and Fig. S11C), suggesting a connection between τ_d and the one-to-many change from GMP to promyelocyte. This can perhaps be understood in light of other one-to-many state transitions, such as a pitchfork bifurcation (Fig. S5A), where ω_1 increases steadily if different branches are left unclustered (Fig. S5B). Although the null and signal were significantly closer in the neutrophil trajectory than in the toy model, this discrepancy may be related to the difficulty in identifying one-to-many transcriptomic

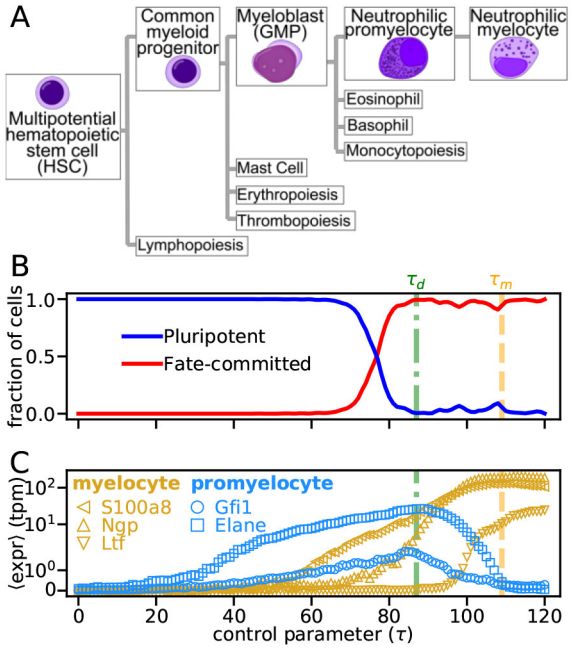

Fig. 4. Detected bifurcations correspond to biologically characterized developmental transitions. (A) Schematic of neutrophil development, beginning from hematopoietic stem cells and ending at the neutrophil myelocyte – a committed neutrophil progenitor. Lines indicate naturally occurring progeny, other than the cell type itself. Subsequent neutrophil-committed fates (neutrophil metamyelocyte, band cells and neutrophils) are not shown. Cell type images by A. Rad and M. Häggström. CC-BY-SA 3.0 license. (B) Fraction of cells in each cell type, based on annotated clustering in Weinreb et al. (2020). (C) Average expression of promyelocyte (blue) and myelocyte (gold) marker genes (Weinreb et al., 2020). Error bars (s.e.m.) are smaller than symbols.

Development (2023) 150, dev201280. doi:10.1242/dev.201280

changes in non-ideal conditions, compared with one-to-one transitions (see Appendix S1, section 6 and Fig. S8 for details). Accordingly, the increase of ω_1 at τ_d suggests that, although cells in the neutrophil trajectory are expected to include only the neutrophil lineage branch of GMP, they may in fact include other GMP lineages, such as eosonophils or basophils.

Conversely, τ_m corresponds well with the pseudotime when myelocyte marker genes are maximal and promyelocyte genes have reduced expression (Fig. 4C and Fig. S11C), suggesting that τ_m indicates the transition point between these two cell fates. Although the myelocyte marker genes begin increasing earlier than τ_m in the trajectory, this may be because of additional cellular processes that smooth out their dynamics over the course of a cell-fate transition. Alternatively, the marker gene dynamics may indicate that the transition at τ_d is connected to the transition at τ_m; e.g. the eigenvalue dynamics at τ_d may hint at an early bias toward the ultimate myelocyte transition, similar to other developmental biases that have recently been identified in hematopoiesis (Wang et al., 2022).

Thus, by using Eqn 2 to quantify the geometry of neutrophil development, we were able to recover the known GMP-neutrophil cell fate decision, qualify the trajectory as likely including other lineages and pinpoint a maturation step in neutrophil development. Importantly, this analysis also highlights the difficulties in using the principal covariance eigenvalue alone to characterize bifurcations, as one-to-many bifurcations in particular may be extremely sensitive to small errors or biases. To address these difficulties, and provide additional insights into possible dynamical transitions, we now leverage a key feature of the covariance analysis outlined above: that signatures of the underlying mechanisms driving a system through a bifurcation are evident in its principal covariance eigenvector.

Covariance eigenvectors provide interpretable low dimensional representations of neutrophil bifurcations

Perhaps the most surprising consequence of the Continuous Time Lyapunov equation, encapsulated in Eqn. 3, is that a high-dimensional bifurcation eigenvector, which is a characteristic of the underlying Jacobian of the system, is directly calculable from the transcriptomic-state data, as it equals, up to a sign, the principal covariance eigenvector $\vec{s}_1$. This result motivated us to probe $\vec{s}_1(\tau)$, the principal eigenvectors of the covariance matrix as a function of pseudotime, and in particular its structure in the vicinity of τ_d and τ_m, to glean further insight into the biological nature of our detected transition points.

We first sought to determine the uniqueness of $\vec{s}_1(\tau)$, as it is extremely high dimensional, by measuring how it varies across pseudotime, compared with average gene expression. We found that the correlation of average gene expression across pseudotime exhibits an approximate two-block structure throughout the trajectory, in which expression is well correlated in $\tau\in[0, 80]$ and again in $\tau\in[90, 121t]$ (Fig. 5A), hinting at the existence of two gene expression states. Interestingly, the correlation of $\vec{s}_1$ was significantly more detailed, exhibiting as many as six distinct blocks, with higher positive correlation and lower negative correlation than seen in expression (Fig. 5B). Importantly, τ_d and τ_m align well with transitions between blocks, further bolstering their significance as markers of developmental transitions. Thus, the variation of $\vec{s}_1(\tau)$, even when it is high dimensional, may reveal significant detail in the structure of a developmental trajectory.

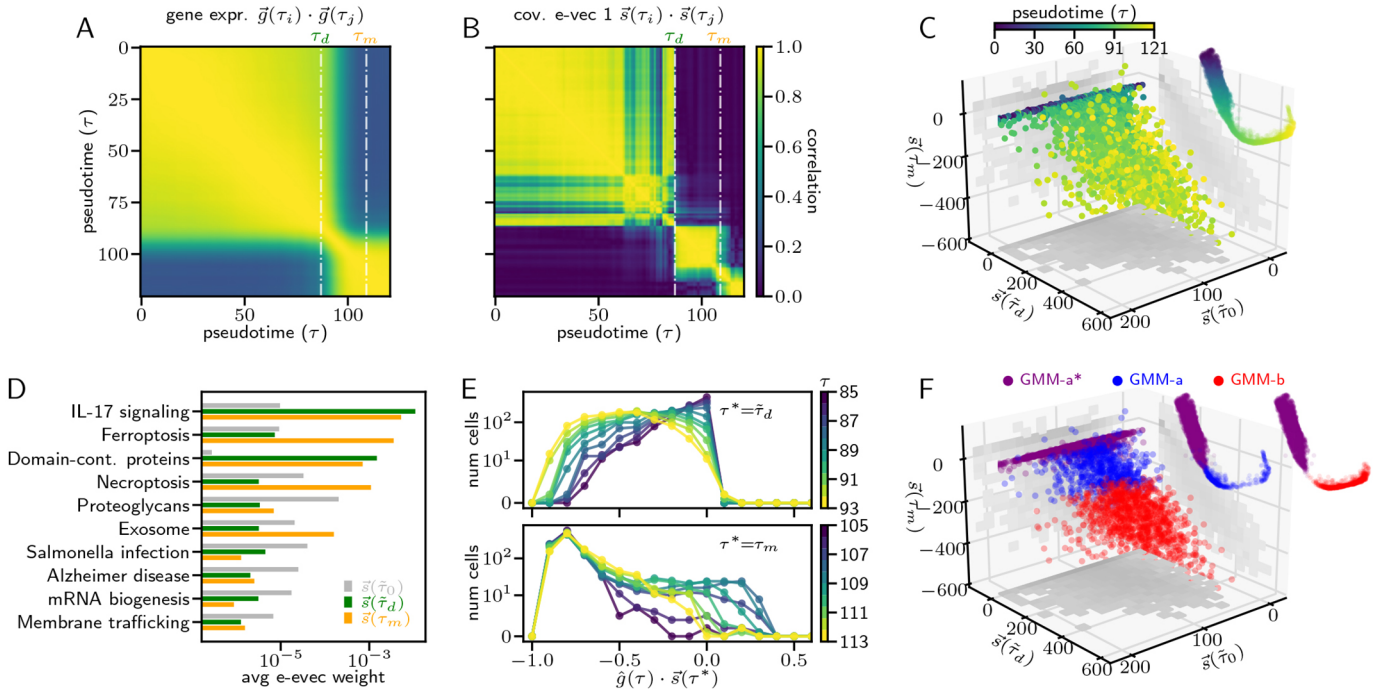

Fig. 5. Analysis of high-dimensional bifurcation directions. (A) Pairwise correlation of average normalized gene expression in each pseudotime bin. (B) Pairwise correlation of the principal covariance eigenvector of each pseudotime bin. (C) Projection of normalized gene expression along principal eigenvectors at τ_0, τ_d and τ_m. Each dot is a cell and its color indicates the pseudotime. Inset shows corresponding position of the cell in the SPRING plot (Fig. 4A). (D) Average weight, per gene, of the highest weighted categories in the KEGG database for *Mus musculus*, for the principal covariance eigenvectors at τ_0,τ_d, and τ_m. (E) Distribution of gene expression projected onto $\vec{s}_d$, near τ_d (top), and distribution of gene expression projected onto $\vec{s}_d$ near τ_m (bottom). (F) As in C, but the color of points (cells) indicates their cluster in the GMM; insets are split by cluster.

Development (2023) 150, dev201280. doi:10.1242/dev.201280

We next sought to examine whether $\vec{s}_1(\tau)$ contain axes of a simplified space along which to examine the trajectory. As the correlation between eigenvectors exhibited distinct blocks, two of which had bifurcative dynamics, we used $\vec{s}_1$ near the beginning of the trajectory and near τ_d and τ_m to revisualize the neutrophil trajectory. For τ_0 and τ_d, we used an $\vec{s}_1$ toward the middle of its block (at $\tilde{\tau}_0$ and $\tilde{\tau}_d$, respectively; see Appendix S1, section 8 for details) to ensure that the direction had stabilized; however, for τ_m we used the vector precisely at τ_m, as the direction of a one-to-one transition may only be observable at the moment of bifurcation (Eqn. 3). We found that for the majority of the trajectory, the variation was localized to $\vec{s}_1(\tilde{\tau}_0)$, but starting near τ_d, cells began to vary along both $\vec{s}_1(\tilde{\tau}_d)$ and $\vec{s}_1(\tau_m)$ (Fig. 5C). As the variation along $\vec{s}_1(\tilde{\tau}_d)$ and $\vec{s}_1(\tau_m)$ coincided, and they had a high correlation of 0.67 (Fig. 5C), while both being nearly completely orthogonal to $\vec{s}_1(\tilde{\tau}_0)$, we sought to determine their differences. We found that whereas the fraction of variation of gene expression along $\vec{s}_1(\tilde{\tau}_d)$ begins to increase near τ_d and remains high even around τ_m (Fig. S12A), highlighting the importance of $\vec{s}_1(\widetilde{tau}_d)$ for the transition at τ_m, $\vec{s}_1(\tau_m)$ accounts for only a significant fraction of variation very close to τ_m (Fig. S12A), suggesting additional features that are distinct from $\vec{s}_1(\tilde{\tau}_d)$ (Fig. S12A). Additionally, the projection of the vector tangent to gene expression onto these eigenvectors appears high for both bifurcation eigenvectors after τ_d, suggesting that they are simultaneously involved in the gene expression dynamics (Fig. S12B) and that, at least from τ_d onward, the trajectory appears multi-dimensional. We found hints of the distinctly $\vec{s}_1(\tau_m)$ features in the gene expression projection (Fig. 5C), where towards the end of the trajectory, when cells have already deviated far from the $\vec{s}_1(\tau_m) = 0$ plane, some of the cells appear closer to that plane. Thus, the increased variance along $\vec{s}_1(\tau_m)$ reflects the spike of ω_1 at τ_m (Fig. 4C), and further illuminates it by showing that the direction of the bifurcation is toward the pluripotent state.

As these directions in transcriptomic space revealed new dynamics, we probed the loadings of each eigenvector to determine its functional relevance. To do so, we computed the average weight of each functional category in the KEGG database for *Mus musculus* in the eigenvector (Kanehisa, 2019), i.e.

$$W_c(t) = \frac{1}{|G_c|} \sum_{g \in \{G_c\}} s_g^2(t), \tag{7}$$

where $W_c(t)$ is the average weight of category c at pseudotime t, G_c is the set of genes in the transcriptomic matrix that map to category c, and $s_g(t)$ is the weight of gene g in the principal covariance eigenvector at pseudotime t. We show $W_c(t)$ for $t \in \{\tilde{\tau}_0, \tilde{\tau}_d, \tau_m\}$ in Fig. 5D for the five categories with highest W_c at each of those pseudotimes. We found that $W_c(\tilde{\tau}_d)$ and $W_c(\tau_m)$ were heavily weighted for interleukin 17 (IL17) signaling, a key pathway for controlling infection (Monin and Gaffen, 2018), which has been shown to be activated by neutrophils (Li et al., 2010), to promote neutrophil recruitment (Cua and Tato, 2010) and to aid in the formation of neutrophil exctracellular traps (Lin et al., 2011). Interestingly, $\vec{s}_1(\tau_m)$ is also highly weighted for regulated cell death mechanisms, including ferroptosis and necroptosis. The high weight for these mechanisms may indicate that the transition at τ_m includes a functional gain, as ferroptosis has recently been demonstrated as a mechanism that neutrophils use to combat glioblastoma cells (Yee et al., 2020) and is also associated with neutrophil extracellular traps (Chen et al., 2021). Alternatively, these mechanisms may have resulted from the activation of cell death within the neutrophil population itself, as necroptosis has been suggested as a population-control mechanism to prevent tissue damage that can occur from an overaccumulation of neutrophils at an infection site (Wang et al., 2018).

Last, we used $\vec{s}_1(\tau)$ to examine whether the increased ω_1 at τ_d and τ_m is due to the emergence of a second distinct cellular population or, alternatively, to an increase in the transcriptomic variance. Concretely, if the increased ω_1 is due to a second population of cells, then the projection of gene expression at τ along $\vec{s}_1(\tau)$ should appear bimodal. We found that this projection widened near both τ_d and τ_m (Fig. 5E and Fig. S12C), reflecting their increased ω_1, and exhibited bimodality at τ_m, suggesting the emergence of a second population of cells. This bimodality is also apparent from the gene expression distributions for many of the genes with highest weights in $\vec{s}_1(\tau_m)$, including Fth1, Psap and Ccl6 (Fig. S13). To further disentangle the two populations of cells, we fit a two-peak Gaussian Mixture Model (GMM) to $\boldsymbol{G}(\tau_m)$ (see Appendix S1, section 9 for details) (Pedregosa et al., 2011). The GMM clearly distinguished the two modes in bimodal gene expression distributions, and showed that many other genes, such as S100a9 and Ngp, which appear initially to have unimodal distributions, also exhibit bimodal structure (Fig. S13).

To evaluate the dynamics of the two cellular populations, we used the GMM to predict the cluster label for all cells in the full trajectory. We found that, for $\tau < \tau_d$, all cells belonged to the same cluster (GMM-a), and began to separate into two clusters (GMM-a and GMM-b) near τ_d (Fig. S14). Additionally, the two clusters exhibit contrasting ω_1 dynamics: GMM-a has increasing ω_1, while GMM-b has a spike in ω_1 at τ_m (Fig. S14). Importantly, the clustering split the cells along $\vec{s}_1(\tau_m)$, such that cells that were late in pseudotime, but close to the $\vec{s}_1(\tau_m) = 0$ plane, are part of cluster GMM-a, indicating that they are functionally earlier in the neutrophil specification path (Fig. 5F).

Summarizing, we found that the geometry of the data in transcriptomic space, explicated by the principal covariance eigenvector, yielded significant insight into developmental specification dynamics throughout pseudotime, especially at bifurcations. Throughout pseudotime, correlations between eigenvectors were able to pinpoint pivotal trajectory moments, including, and in addition to, those indicated by the principal covariance eigenvalue. Although the eigenvalue analysis alone did not clearly distinguish between the two transitions, the eigenvector analysis highlighted unique mechanistic features at each transition, suggesting they represent distinct developmental events. Eigenvectors at the bifurcations were also helpful in visualizing the trajectory, and inferring the molecular and cellular processes that reshape the developmental landscape. Furthermore, by viewing the bifurcating data along its eigenvector, we were able to discern bimodality and to distinguish between multiple cell fates within the same lineage. Thus, eigenvector analysis appears to be a promising new direction for analysis of pseudotime trajectories.

DISCUSSION

A singular challenge in understanding cellular fate transitions using transcriptomics has been dimensionality: cell fates are a low-dimensional functional description, a valley in Waddington's landscape, whereas gene-expression profiles are points in a myriad-dimensional space – how can gene expression possibly show the geometry of development? In this study, we have leveraged the continuous-time Lyapunov equation to show that the dynamics of state-variable covariance, even at high dimensionality, are sufficient to assess a crucial aspect of developmental geometry: when and how linear stability is lost to

Development (2023) 150, dev201280. doi:10.1242/dev.201280

yield a bifurcation. Our central and novel result, based on a restricted region of the transcriptomic trajectories present during the process of hematopoiesis, is that the requisite statistical signatures of a bifurcation are detectable and present during development, even through the complex, high-dimensional lens of sequencing. Although biases and imperfections in data may confound developmental bifurcations, pairing analysis of both the principal covariance eigenvalue and eigenvectors enable us to disentangle multiple transition points, and elucidate mechanistic features that are normally completely hidden in the absence of a candidate mathematical model. Thus, our results have important consequences for the theoretical understanding of developmental transitions, the specific biology of neutrophil development and the analysis of dynamic biological data.

Our finding that a transcriptomic trajectory can have distinct geometric signatures, including durations during which the principal covariance eigenvalue is constant or spikes, has considerable consequences for the theoretical understanding of developmental dynamics. That we saw any consistent behavior in the principal covariance eigenvalue lends significant support to our initial hypothesis that cell fate modifiers operate at a much slower rate than transcriptomic modifiers, because if these occurred on similar timescales, no such statistical signature would be evident, let alone those that align well with current understanding. Additionally, whereas previous statistical analyses of scRNA-seq data found that developmental trajectories appear as monotonic proliferations of cell fates (Nitzan and Brenner, 2021), our focus on a single developmental trajectory enables the distinction of multiple developmental epochs, including durations of development during which cell fates do not undergo qualitative changes, but proliferate (the GMP-to-promyelocyte transition) and change state (the promyelocyte-to-myelocyte transition). Finally, our evidence of bifurcations starkly contrasts with scRNA-seq visualizations that show gene expression varying smoothly along a developmental path, and underscores the importance of understanding both noise and non-linear dependencies when using transcriptomic profiles to classify the fate of a cell (Moris et al., 2016).

Our analysis of the data of Weinreb et al. (2020) also yielded intriguing implications regarding the specific geometry of neutrophil development in mice. In particular, some of the known cell fate changes in neutrophil development were not distinguishable in the covariance eigenvalue trajectory [e.g. from common myeloid progenitors (CMPs) to GMP], which indicates that these changes are less bifurcative than the GMP-to-promyelocyte or promyelocyte-to-myelocyte transitions. This could mean, for example, that even when CMPs differentiate to GMPs, the transition lacks commitment and is dependent on a sustained developmental signal, whereas once cells transition from GMP to promyelocyte, they are committed to becoming neutrophils regardless of an external signal. Alternatively, these non-bifurcative transitions may be driven by early fate biases (Wang et al., 2022), and further research may be necessary to robustly determine which progenitor cell fates are statistically stable. Additionally, in comparing the principal covariance eigenvectors throughout pseudotime (Fig. 5B), it became apparent that the direction along which that fate change happened was well aligned with the direction of the GMP-to-promyelocyte transition. This result may be a sign of distinct, soft directions in transcriptomic space along which cell fates are most likely to change. In addition, as the steady increase and spike in the neutrophil trajectory (Fig. 3C) did not resemble the covariance dynamics of noise-induced state transitions (Fig. S6B), our results suggest that the promyelocyte and myelocyte transitions likely occur due to a loss of stability between fixed points, rather than stochasticity alone.

Aside from these geometric implications, pinpointing bifurcations in pseudotime also enhances analysis of temporal biological data, as it enables the efficient identification of the genes and molecular mechanisms that drive a cell fate transition. At a bifurcation, the principal covariance eigenvectors can aid visualization and highlight critical mechanisms that distinguish clusters (Fig. 5). As the principal covariance eigenvector is equivalent to the Jacobian eigenvector at the bifurcation, and the Jacobian directly reflects gene dynamics, the eigenvector may also be useful for constraining an inferred global Jacobian (Nägele et al., 2014). Additionally, the correlation matrix at a bifurcation may aid in building regulatory network models when combined with previous protein-interaction data or new experimental perturbations (Sun et al., 2015). Furthermore, it may be possible to incorporate our covariance analysis into other indications of pseudotime rank, such as cellular barcodes and low-dimensional distance, to constrain developmental trajectories along bifurcative paths.

Although we focus here on scRNA-seq data, our approach is broadly applicable, and could, in principle, aid in illuminating other aspects of high-dimensional biological dynamics, such as the relationship between development and evolution, or the genomic structural modifications necessary for fate transitions (Buenrostro et al., 2013; Jia et al., 2018). Our analysis was only possible because scRNA-seq experiments can now measure the expression of tens of thousands of genes in hundreds of thousands of cells, enabling accurate covariance measurements. That we found bifurcative events in these data implies that there are low-dimensional, non-linear dynamical systems at play, and that sufficient biological sampling, coupled with physics-based analyses, can reveal the knobs to controllably tilt developmental landscapes.

MATERIALS AND METHODS

Continuous time Lyapunov equation for transcriptomic matrices

Let $\mathbf{G}$ be the steady-state transcriptomic matrix at a single developmental time with n_c rows (cells) and n_g columns (genes), and $\mathbf{F}$ be a set of differential equations describing the molecular interactions that generate $\mathbf{G}$, such that

$$\dot{\mathbf{G}} = \mathbf{F}(\mathbf{G}), \tag{8}$$

where $\dot{\mathbf{G}}$ is the derivative of $\mathbf{G}$ with respect to time. As all cells (columns) in $\mathbf{G}$ are at steady state at the same developmental time, τ, we assume (for the purpose of contradiction) that they are all statistical replicates of the same transcriptomic state, $\vec{g}^*$, and the full matrix, $\mathbf{G}$, is therefore in the vicinity of the hyperbolic fixed point:

$$E(\mathbf{G}) = \mathbf{G}^* = \vec{g}^* \otimes 1_{n_c}^T, \tag{9}$$

where 1_{n_c} is a vector of n_c values and E denotes the expectation operator. The dynamics of $\mathbf{G}$ can be by linearized by the distance to the fixed point $\mathbf{X} = \mathbf{G} - \mathbf{G}^*$, such that

$$\dot{\mathbf{G}} = \mathbf{F}(\mathbf{G}^* + \mathbf{X}) \approx \mathbf{F}(\mathbf{G}^*) + \sum_{i=1}^{n_g} \frac{\partial \mathbf{G}}{\partial \overrightarrow{G_i}} \mathbf{X}_i = \mathbf{J}\mathbf{X}, \tag{10}$$

where $\mathbf{J} = \dfrac{d\dot{\mathbf{G}}}{d\mathbf{G}}\Big|_{\mathbf{G}^*}$ is the Jacobian of $\mathbf{G}$ and we have used the fact that, at steady state, $\mathbf{F}(\mathbf{G}^*) = 0$.

If $\mathbf{F}$ is stochastic and Markovian, then the dynamics of $\boldsymbol{X}$ can be described as a discretized Ornstein-Uhlenbeck (OU) process:

$$\mathbf{X}_{t+\Delta t} = \mathbf{X}_t + \Delta t \mathbf{J}\mathbf{X}_t + \sqrt{\Delta t}\boldsymbol{\zeta}_t, \tag{11}$$

Development (2023) 150, dev201280. doi:10.1242/dev.201280

where Δt is the molecular interaction timescale and $\zeta_{t;i,j}$ is sampled from $N(0, \sigma_i)$, where σ_i is the variance of gene i. The gene-gene covariance matrix can then be defined as

$$\mathbf{C} = E((\mathbf{G} - E(\mathbf{G}))^T(\mathbf{G} - E(\mathbf{G})) = E(\mathbf{X}^T\mathbf{X}), \tag{12}$$

where the superscript T denotes transpose and we have approximated $E(\mathbf{G}) = \mathbf{G}^*$. The stationary condition for an OU process (i.e. that $\partial\mathbf{C}/\partial t = 0$) then yields

$$\frac{\partial \mathbf{C}}{\partial t} = \lim_{\Delta t \to 0} \frac{E(\mathbf{X}_{t+\Delta t}\mathbf{X}_{t+\Delta t}^T) - E(\mathbf{X}_t\mathbf{X}_t^T)}{\Delta t}$$

$$= \mathbf{J}\mathbf{C} + \mathbf{C}\mathbf{J}^T + \mathbf{D} = 0, \tag{13}$$

where $\mathbf{D} = E(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t^T)$; we have used the fact that $E(\zeta_t)$,$E(\mathbf{X}_t)$,$E(\boldsymbol{\zeta}_t\mathbf{X}_t)$ and $E(\mathbf{X}^T\boldsymbol{\zeta}^T)$ are all 0 (Oku and Aihara, 2018).

Covariance at bifurcation

If $\boldsymbol{J}$ is diagonalizable, such that

$$\mathbf{J} = \mathbf{P}\Lambda\mathbf{P}^{-1}, \tag{14}$$

where Λ is a diagonal matrix of eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_{n_g})$ and $\mathbf{P}^T$ is the square matrix of eigenvectors $(\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_{n_g})$, then Eqn 13, often referred to as the continuous-time Lyapunov (CL) equation, can be used to qualitatively assess $\mathbf{G}$. Left multiplying Eqn 13 by $\mathbf{P}^{-1}$ and right multiplying by $(\mathbf{P}^\dagger)^{-1}$, yields

$$\boldsymbol{\lambda}\tilde{\mathbf{C}} + \tilde{\mathbf{C}}\lambda^\dagger + \tilde{\mathbf{C}} = 0, \tag{15}$$

where † indicates conjugate transpose, $\tilde{\mathbf{C}} = \mathbf{P}^{-1}\mathbf{C}(\mathbf{P}^\dagger)^{-1}$ and $\tilde{\boldsymbol{D}} = \boldsymbol{P}^{-1}\boldsymbol{D}(\boldsymbol{P}^\dagger)^{-1}$. As Λ is diagonal, Eqn 15 can be rewritten elementwise:

$$\lambda_i\tilde{C}_{ij} + \bar{\lambda}_j\tilde{C}_{ij} + \tilde{D}_{ij} = 0$$

$$\rightarrow \tilde{C}_{ij} = \frac{-\tilde{D}_{ij}}{\lambda_i + \bar{\lambda}_j}, \tag{16}$$

which can be substituted to yield an expression for elements of the covariance

$$C_{ij} = \sum_{k=1}^{n_g}\sum_{l=1}^{n_g} P_{ik}\left(\frac{-\tilde{D}_{kl}}{\lambda_k + \bar{\lambda}_l}\right)P_{lj}, \tag{17}$$

as $\mathbf{C} = \mathbf{P}\tilde{\mathbf{C}}\mathbf{P}^\dagger$. At a bifurcation, $\max(\Lambda)=\lambda_d\rightarrow 0$, so the $k=l=d$ term in Eqn 17 becomes dominant and

$$C_{ij} = \left(\frac{-\tilde{D}_{dd}}{2\lambda_d}\right)\vec{p}_d^{\,i}\vec{p}_d^{\,j}, \tag{18}$$

where $\vec{p}_d$ is the d^{th} column of $\mathbf{P}$.

Bifurcation eigenvector equivalence

As $\mathbf{C}$ is real and symmetric, the eigenvalue decomposition can be written as a single sum:

$$\mathbf{C} = \mathbf{S}\boldsymbol{\Omega}\mathbf{S}^T$$

$$\rightarrow C_{ij} = \sum_{k=1}^{n_g}\omega_i\vec{s}_k^{\,i}\vec{s}_k^{\,j}, \tag{19}$$

where $\{\omega_1, \omega_2, \ldots, \omega_{n_g}\}$ are its eigenvalues, and $\{\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_{n_g}\}$ are its eigenvectors, which are normalized to 1. For Eqn 19 to be equivalent to Eqn 18 at a bifurcation, at least one eigenvalue $\omega_i\rightarrow\infty$, which we may, without loss of generality, refer to as ω_1. If $\omega_1 \gg \omega_i$ for $i\in[2\ldots n_g]$ then the $k=1$ dominates the sum in Eqn 19, and by equating with Eqn 18 we obtain

$$\vec{s}_1^{\,i}\vec{s}_1^{\,j} = \left(\frac{-\tilde{D}_{dd}}{2\lambda_d\omega_1}\right)\vec{p}_d^{\,i}\vec{p}_d^{\,j}$$

$$\rightarrow \frac{\vec{s}_1^{\,j}}{\vec{s}_1^{\,i}} = \left(\frac{-\tilde{D}_{dd}}{2\lambda_d\omega_1}\right)\frac{\vec{p}_d^{\,i}\vec{p}_d^{\,j}}{(\vec{s}_1^{\,i})^2} = \frac{\vec{p}_d^{\,i}\vec{p}_d^{\,j}}{(\vec{p}_d^{\,i})^2} = \frac{\vec{p}_d^{\,j}}{\vec{p}_d^{\,i}}$$

$$\rightarrow 1 = \sum_{j=1}^{n_g}(\vec{s}_1^{\,j})^2 = \sum_{j=1}^{n_g}\left(\frac{\vec{s}_1^{\,i}}{\vec{p}_d^{\,i}}\vec{p}_d^{\,j}\right)^2 = \left(\frac{\vec{s}_1^{\,i}}{\vec{p}_d^{\,i}}\right)^2$$

$$\rightarrow \vec{s}_1^{\,i} = \pm\vec{p}_d^{\,i}$$

$$\rightarrow \vec{s}_1^{\,j} = \vec{p}_d^{\,j}\vec{s}_1^{\,i}/\vec{p}_d^{\,i} = \pm\vec{p}_d^{\,j}$$

$$\rightarrow \vec{s}_1 = \pm\vec{p}_d, \tag{20}$$

where we have used the fact that $\vec{s}_1$ and $\vec{p}_d$ both normalize to 1. Importantly, it is also computationally advantageous to analyze the eigen decomposition of the covariance, rather than the covariance itself, because for large n_g, $\boldsymbol{\Omega}$ and $\boldsymbol{S}$ can be obtained directly from the singular value decomposition of $\boldsymbol{X}$.

Simulation methodology

To explore our analysis framework on a more biologically relevant gene network (Fig. 2A) we used a Focker-Plank simulation method. For each of the N_c=100 cells (N_c chosen by examining how many cells were necessary to accurately detect bifurcations in the neutrophil data (Fig. S10A,B), the expression of gene i $[g_i(t;m_1, m_2, k_D)]$ is initialized uniformly randomly in the interval [0,4]. The expression at subsequent time steps $[g_i(t+\Delta t)]$ is sampled from a Gaussian distribution $N(\mu, \sigma)$, where

$$\mu_i = g_i(t) + \dot{g}_i\, g_i(t))\Delta t \tag{21}$$

$$\sigma_i = \begin{cases} \sqrt{\dfrac{\Delta t}{s}\left(\dfrac{m_i}{1+g_{3-i}^2} + k_D g_i\right)} & i = 1,2 \\ \sqrt{\dfrac{\Delta t}{s}\left(\dfrac{\alpha_i g_{d(i)}^2 + 1 + \alpha_i}{(1+g_{d(i)}^2)} + k_i g_i\right)} & i = 3\ldots n_g \end{cases} \tag{22}$$

and bounded to be non-negative. The simulations ran for N_t=1e7 time steps, with Δt=0.01, at a noise scale of $1/s$=0.05, and the last timestep of each simulation is the steady-state expression $\boldsymbol{G}$. We verified that N_t was sufficiently large by averaging $\boldsymbol{G}(m_1)$ across cells, and observing that individual genes discontinuously, but predictably, switch their expression at m_{1c} [Fig. 2C; genes sorted by $d(g_i)$ and $_i$] compared with the continuous and unpredictable transitions observed with low N_t (Fig. S7A). In the saddle-node example (Fig. 2), the remaining parameters were k_D=1, m_2=3, $m_1\in$[2, 4], while in the pitchfork example (Fig. S5), $m_{1,2}$=1, $k_D\in$[0.24, 5].

Analysis pipeline

Eqns 2-4 imply an analysis pipeline characterizing bifurcations in high-dimensional temporal data, which we use in this article:

(1) Obtain highly sampled temporal data. Caveat: for data types such as scRNA-seq, where frequent sampling is difficult, and samples may include realizations from many different times, time may be inferable, using, for example, pseudotime inference (see Appendix S1, section 7.1).

(2) Bin the data along the temporal axis.

(3) Compute the largest eigenvalue of the covariance matrix (ω_1) in each bin (e.g. using an off-the-shelf PCA function).

(4) Evaluate whether a bifurcation occurs by comparing ω_1 with a suitable null (see Appendix S1, section 4): spike indicates a one-to-one bifurcation; steady increase indicates a one-to-many bifurcation.

(5) If a bifurcation is detected (e.g. at τ_c), compute and examine the principal covariance eigenvector at τ_c, as it reflects mechanistic aspects of the underlying dynamical system.

Acknowledgements

We thank Richard Carthew, Yogesh Goyal and Karna Gowda for reviewing the manuscript and providing suggestions. We thank Northwestern Information Technology for access to the Quest High-Performance Computing Cluster.

Competing interests
The authors declare no competing or financial interests.

Author contributions
Conceptualization: S.L.F., S.G., M.M.; Validation: S.L.F.; Formal analysis: S.L.F.; Investigation: S.L.F., B.X., S.G., M.M.; Writing - original draft: S.L.F., S.G., M.M.; Writing - review & editing: S.L.F., S.G., M.M.; Visualization: S.L.F.; Supervision: S.G., M.M.; Project administration: S.G., M.M.; Funding acquisition: S.G., M.M.

Funding
This work was supported by the National Science Foundation-Simons Center for Quantitative Biology at Northwestern University and the Simons Foundation (597491 to M.M.). S.G. acknowledges the University of Toronto's Medicine by Design initiative, which receives funding from the Canada First Research Excellence Fund, and a Natural Sciences and Engineering Research Council of Canada Discovery Grant. M.M. is a Simons Foundation Investigator. Open Access funding provided by Northwestern University. Deposited in PMC for immediate release.

Data availability
Instructions and Python code for reproducing all figures in this manuscript are available at http://github.com/simfreed/sc_bifurc_figs.

Peer review history
The peer review history is available online at https://journals.biologists.com/dev/lookup/doi/10.1242/dev.201280.reviewer-comments.pdf.

References
Borregaard, N. (2010). Neutrophils, from marrow to microbes. *Immunity* **33**, 657-670. doi:10.1016/j.immuni.2010.11.011

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dna-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213. doi:10.1038/nmeth.2688

Camacho-Aguilar, E., Warmflash, A. and Rand, D. A. (2021). Quantifying cell transitions in c. elegans with data-fitted landscape models. *PLoS Comput. Biol.* **17**, e1009034.doi:10.1371/journal.pcbi.1009034

Chen, L., Liu, R., Liu, Z.-P., Li, M. and Aihara, K. (2012). Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci. Rep.* **2**, 342.

Chen, Z., Bai, X., Ma, L., Wang, X., Liu, X., Liu, Y., Chen, L. and Wan, L. (2018). A branch point on differentiation trajectory is the bifurcating event revealed by dynamical network biomarker analysis of single-cell data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **17**, 366-375. doi:10.1109/TCBB.2018.2847690

Chen, X., Kang, R., Kroemer, G. and Tang, D. (2021). Ferroptosis in infection, inflammation, and immunity. *J. Exp. Med.* **218**, e20210518. doi:10.1084/jem.20210518

Corson, F. and Siggia, E. D. (2012). Geometry, epistasis, and developmental patterning. *Proc. Natl Acad. Sci. USA* **109**, 5568-5575. doi:10.1073/pnas.1201505109

Cua, D. J. and Tato, C. M. (2010). Innate il-17-producing cells: the sentinels of the immune system. *Nat. Rev. Immunol.* **10**, 479-489. doi:10.1038/nri2800

Ferrell, J. E., Jr (2012). Bistability, bifurcations, and waddington's epigenetic landscape. *Curr. Biol.* **22**, R458-R466. doi:10.1016/j.cub.2012.03.045

Furchtgott, L. A., Melton, S., Menon, V. and Ramanathan, S. (2017). Discovering sparse transcription factor codes for cell states and state transitions during development. *Elife* **6**, e20488. doi:10.7554/eLife.20488

Gallivan, C. P., Ren, H. and Read, E. L. (2020). Analysis of single-cell gene pair coexpression landscapes by stochastic kinetic modeling reveals gene-pair interactions in development. *Front. Genet.* **10**, 1387. doi:10.3389/fgene.2019.01387

Gardner, T. S., Cantor, C. R. and Collins, J. J. (2000). Construction of a genetic toggle switch in escherichia coli. *Nature* **403**, 339-342. doi:10.1038/35002131

Gregor, T., Tank, D. W., Wieschaus, E. F. and Bialek, W. (2007). Probing the limits to positional information. *Cell* **130**, 153-164. doi:10.1016/j.cell.2007.05.025

Hastie, T., Tibshirani, R., Friedman, J. H. and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 2. Springer.

Huang, S., Eichler, G., Bar-Yam, Y. and Ingber, D. E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 128701. doi:10.1103/PhysRevLett.94.128701

Huang, S., Guo, Y.-P., May, G. and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695-713. doi:10.1016/j.ydbio.2007.02.036

Jia, G., Preussner, J., Chen, X., Guenther, S., Yuan, X., Yekelchyk, M., Kuenne, C., Looso, M., Zhou, Y., Teichmann, S. et al. (2018). Single cell rna-seq and atac-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nat. Commun.* **9**, 4877. doi:10.1038/s41467-017-02088-w

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947-1951. doi:10.1002/pro.3715

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22**, 437-467. doi:10.1016/0022-5193(69)90015-0

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A. and Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201. doi:10.1016/j.cell.2015.04.044

Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H. B. et al. (2022). Cellrank for directed single-cell fate mapping. *Nat. Methods* **19**, 159-170. doi:10.1038/s41592-021-01346-6

Li, L., Huang, L., Vergis, A. L., Ye, H., Bajwa, A., Narayan, V., Strieter, R. M., Rosin, D. L. and Okusa, M. D. (2010). Il-17 produced by neutrophils regulates IFN-mediated neutrophil migration in mouse kidney ischemia-reperfusion injury. *J. Clin. Invest.* **120**, 331-342. doi:10.1172/JCI38702

Li, J. H., Ye, F. X.-F., Qian, H. and Huang, S. (2019). Time-dependent saddle–node bifurcation: Breaking time and the point of no return in a non-autonomous model of critical transitions. *Physica D* **395**, 7-14. doi:10.1016/j.physd.2019.02.005

Lin, A. M., Rubin, C. J., Khandpur, R., Wang, J. Y., Riblett, M., Yalavarthi, S., Villanueva, E. C., Shah, P., Kaplan, M. J. and Bruce, A. T. (2011). Mast cells and neutrophils release il-17 through extracellular trap formation in psoriasis. *J. Immunol.* **187**, 490-500. doi:10.4049/jimmunol.1100123

Liu, R., Li, M., Liu, Z.-P., Wu, J., Chen, L. and Aihara, K. (2012). Identifying critical transitions and their leading biomolecular networks in complex diseases. *Sci. Rep.* **2**, 1-9.

Marco, E., Karp, R. L., Guo, G., Robson, P., Hart, A. H., Trippa, L. and Yuan, G.-C. (2014). Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl Acad. Sci. USA* **111**, E5643-E5650. doi:10.1073/pnas.1408993111

Mojtahedi, M., Skupin, A., Zhou, J., Castaño, I. G., Leong-Quong, R. Y., Chang, H., Trachana, K., Giuliani, A. and Huang, S. (2016). Cell fate decision as high-dimensional critical state transition. *PLoS Biol.* **14**, e2000640. doi:10.1371/journal.pbio.2000640

Monin, L. and Gaffen, S. L. (2018). Interleukin 17 family cytokines: signaling mechanisms, biological activities, and therapeutic implications. *Cold Spring Harbor Perspect. Biol.* **10**, a028522. doi:10.1101/cshperspect.a028522

Moris, N., Pina, C. and Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nat. Rev. Genet.* **17**, 693-703. doi:10.1038/nrg.2016.98

Nägele, T., Mair, A., Sun, X., Fragner, L., Teige, M. and Weckwerth, W. (2014). Solving the differential biochemical jacobian from metabolomics covariance data. *PLoS ONE* **9**, e92299. doi:10.1371/journal.pone.0092299

Nitzan, M. and Brenner, M. P. (2021). Revealing lineage-related signals in single-cell gene expression using random matrix theory. *Proc. Natl Acad. Sci. USA* **118**, e1913931118. doi:10.1073/pnas.1913931118

Novak, B. and Tyson, J. J. (1993). Numerical analysis of a comprehensive model of m-phase control in xenopus oocyte extracts and intact embryos. *J. Cell Sci.* **106**, 1153-1168. doi:10.1242/jcs.106.4.1153

Oku, M. and Aihara, K. (2018). On the covariance matrix of the stationary distribution of a noisy dynamical system. *Nonlinear Theory and Its Applications, IEICE* **9**, 166-184. doi:10.1587/nolta.9.166

Ostuni, R., Natoli, G., Cassatella, M. A. and Tamassia, N. (2016). Epigenetic regulation of neutrophil development and function. *Semin. Immunol.* **28**, 83-93. doi:10.1016/j.smim.2016.04.002

Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of escherichia coli. *Nature* **427**, 737. doi:10.1038/nature02298

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825-2830.

Plasschaert, L. W., Žilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G., Klein, A. M. and Jaffe, A. B. (2018). A single-cell atlas of the airway epithelium reveals the cftr-rich pulmonary ionocyte. *Nature* **560**, 377-381. doi:10.1038/s41586-018-0394-6

Qiu, X., Rahimzamani, A., Wang, L., Ren, B., Mao, Q., Durham, T., McFaline-Figueroa, J. L., Saunders, L., Trapnell, C. and Kannan, S. (2020). Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell Syst.* **10**, 265-274. doi:10.1016/j.cels.2020.02.003

Raj, A. and Van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216-226. doi:10.1016/j.cell.2008.09.050

Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. and Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science* **307**, 1962-1965. doi:10.1126/science.1106914

Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., Choi, K., Bendall, S., Friedman, N. and Pe'er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34**, 637-645. doi:10.1038/nbt.3569

Shamir, M., Bar-On, Y., Phillips, R. and Milo, R. (2016). Snapshot: timescales in cell biology. *Cell* **164**, 1302-1302. doi:10.1016/j.cell.2016.02.058

Slack, J. M. W. (1991). *From Egg to Embryo: Regional Specification in Early Development*. Cambridge University Press. doi:10.1017/CBO9780511525322
Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E. and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477. doi:10.1186/s12864-018-4772-0
Sun, X., Länger, B. and Weckwerth, W. (2015). Challenges of inversely estimating jacobian from metabolomics data. *Front. Bioeng. Biotechnol.* **3**, 188.
Tkačik, G., Gregor, T. and Bialek, W. (2008). The role of input noise in transcriptional regulation. *PLoS ONE* **3**, e2774. doi:10.1371/journal.pone.0002774
Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381. doi:10.1038/nbt.2859
Tusi, B. K., Wolock, S. L., Weinreb, C., Hwang, Y., Hidalgo, D., Zilionis, R., Waisman, A., Huh, J. R., Klein, A. M. and Socolovsky, M. (2018). Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555**, 54-60. doi:10.1038/nature25741
Waddington, C. (1957). The strategy of the genes. Routledge.
Wang, X., Yousefi, S. and Simon, H.-U. (2018). Necroptosis and neutrophil-associated disorders. *Cell Death Dis.* **9**, 1-9. doi:10.1038/s41419-017-0012-9
Wang, S.-W., Herriges, M. J., Hurley, K., Kotton, D. N. and Klein, A. M. (2022). Cospar identifies early cell fate biases from single-cell transcriptomic and lineage information. *Nat. Biotechnol.* **40**, 1066-1074. doi:10.1038/s41587-022-01209-1
Weinreb, C., Wolock, S. and Klein, A. M. (2018a). Spring: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246-1248. doi:10.1093/bioinformatics/btx792
Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M. and Klein, A. M. (2018b). Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl Acad. Sci. USA* **115**, E2467-E2476. doi:10.1073/pnas.1714723115
Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. and Klein, A. M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381. doi:10.1126/science.aaw3381
Weston, B. R., Li, L. and Tyson, J. J. (2018). Mathematical analysis of cytokine-induced differentiation of granulocyte-monocyte progenitor cells. *Front. Immunol.* **9**, 2048. doi:10.3389/fimmu.2018.02048
Yee, P. P., Wei, Y., Kim, S.-Y., Lu, T., Chih, S. Y., Lawson, C., Tang, M., Liu, Z., Anderson, B., Thamburaj, K. et al. (2020). Neutrophil-induced ferroptosis promotes tumor necrosis in glioblastoma progression. *Nat. Commun.* **11**, 1-22. doi:10.1038/s41467-019-13993-7
Zhou, P., Wang, S., Li, T. and Nie, Q. (2021). Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. *Nat. Commun.* **12**, 5609.doi:10.1038/s41467-021-25548-w.

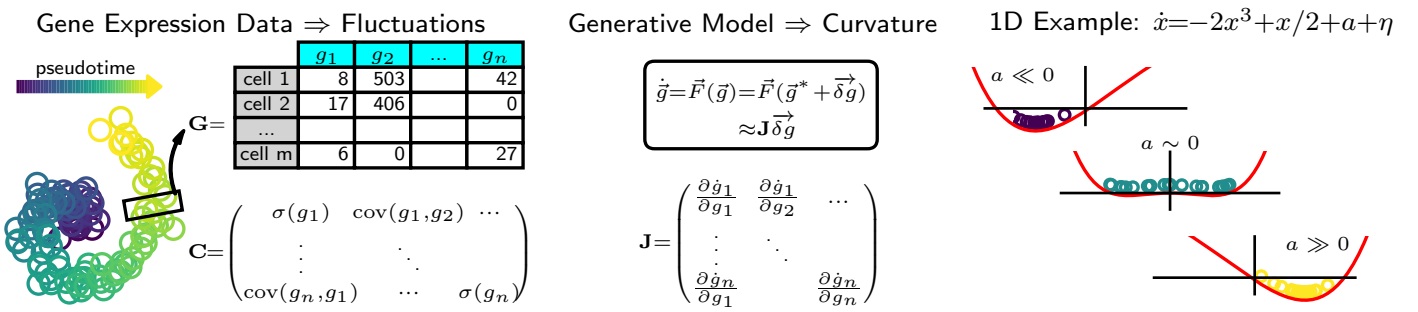

Fig. S1. The relationship between the covariance of a gene expression trajectory and its Jacobian. (A) Schematic of a single-cell RNA-seq dataset arranged by each cell's developmental (pseudo-) time. (Left) Visualization of dataset in two collective gene-expression dimensions. Gene expression matrix at the pseudotime indicated by the rectangle, and its corresponding covariance matrix. (Center) Schematic of a generative model ($\vec{F}$) that could yield the gene expression matrix in (A), and its connection to the Jacobian (**J**). In this model, δg is the deviation of the gene expression vector, $\vec{g}$ from the fixed point, $\vec{g}^*$. (Right) Snapshots of a collection of particles at steady state following the dynamical process defined by $\dot{x} = -2x^3 + x/2 + a$ and uniformly sampled noise for $a = -1$ (left), $a = 0$ (center) and $a = 1$ (right).

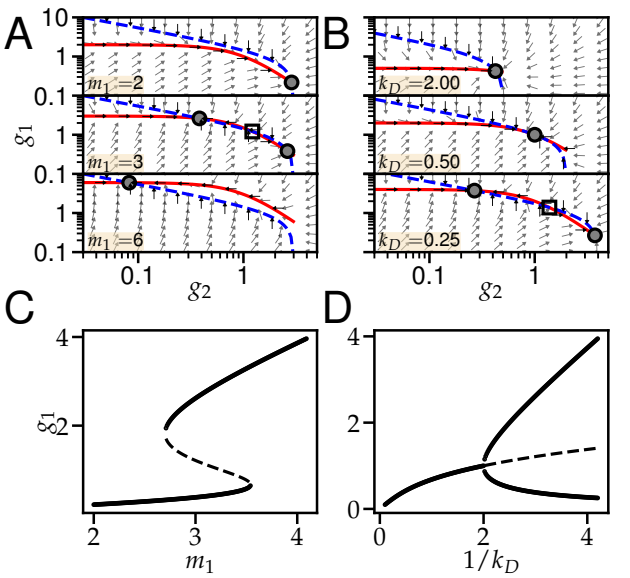

Fig. S2. Analysis of Eqn. 5. (A-B) Phase planes for different parameter sets yield a saddle-node bifurcation (A) or pitchfork bifurcation (B). Solid red line is given by Eqn. S4 while dashed blue line is given by Eqn. S5. Open squares are saddles while closed circles are nodes. Arrow angles are given by $tan^{-1}(\dot{g}1/\dot{g}2)$ and are uniform length. (C-D) Solutions to Eqn. S3 while varying $m1$ (C) or k_D (D). Solid lines are nodes and dashed lines are saddles. In (A,C) $k_D = 1$, $m2 = 3$ and in (B,D) $m_{1,2} = 1$.

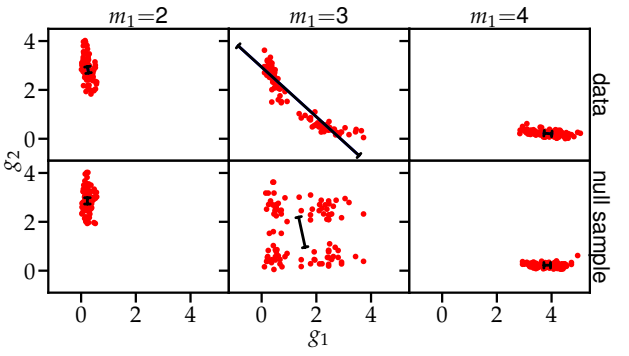

Fig. S3. Effect of resampling on principal eigenvector. Red dots are cells for the data (top) and a single marginal resampling (bottom) before the bifurcation (left), at the bifurcation (center) and after the bifurcation (right). Direction of the black lines corresponds to the principal eigenvector and length corresponds to the principal eigenvalue.

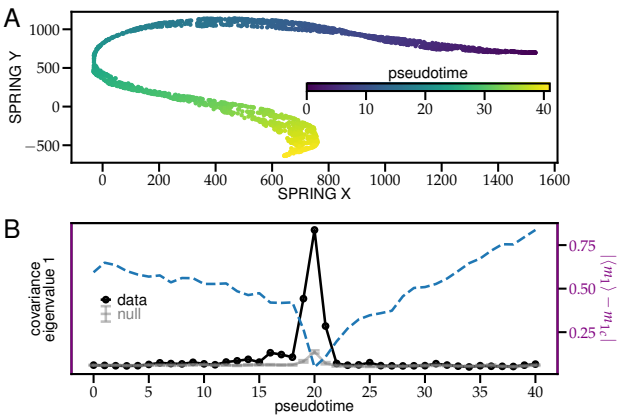

Fig. S4. Pseudotime analysis of saddle-node bifurcation. (A) Representation of simulated cells using SPRING dimen-sionality reduction [1] (min 1 cell, 50 PCs, and 10 nearest neighbors) and Slingshot pseudotime inference (sample size 2000 cells) [2]. (B) Principal covariance eigenvalue ($\omega 1$) plotted as a function of pseudotime (dark dots) as well as the distance of the bifurcation order parameter ($m1$ from its critical value (purple dashed line. DNB order parameter as functions of pseudotime. The peak of $\omega 1$ coincides with the minimum distance between $m1$ and its critical value.

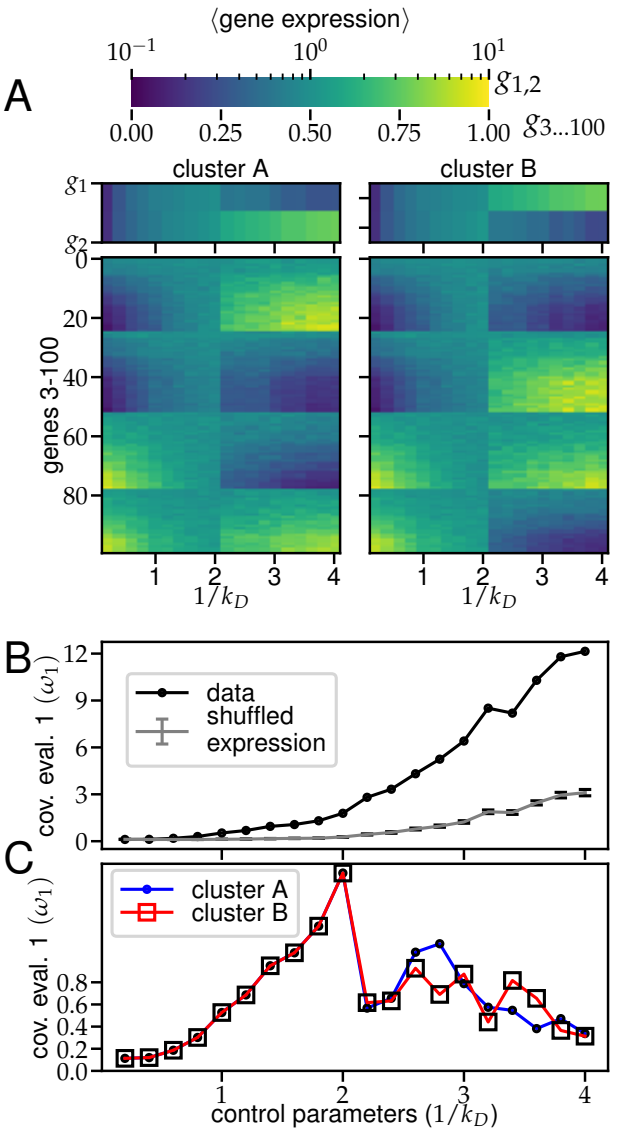

Fig. S5. Pitchfork bifurcation analysis. (A) Gene expression as function of the bifurcation variable τ, separated by cluster.(B) $\omega 1$ and null for unclustered data. (C) $\omega 1$ for clusters.

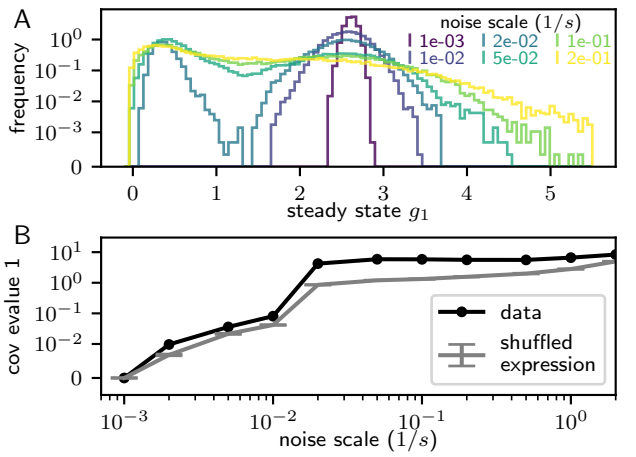

Fig. S6. Covariance eigenvalue signature for a noise induced transition. (A) Steady state distributions for the expression of $g1$ at varying noise scales. (B) Principal covariance eigenvalue as a function of noise.

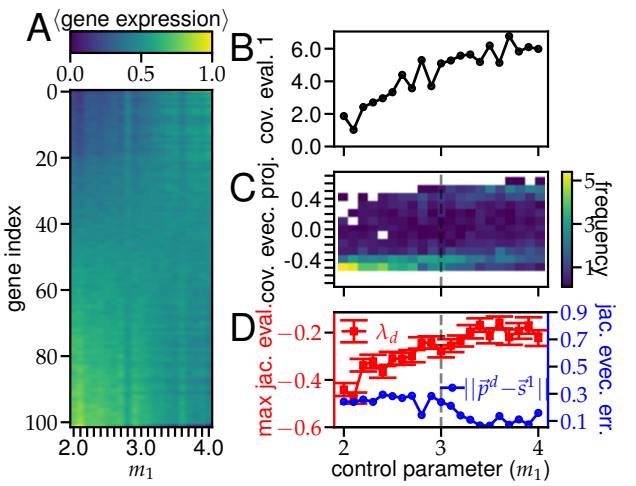

Fig. S7. Unequilibrated saddle-node bifurcation analysis. (A)-(D) Corresponding plots for Fig. 2 (C)-(F), respectively, when the simulation is runs for a $N_t = 500$ iterations rather than $N_t = 50000$.

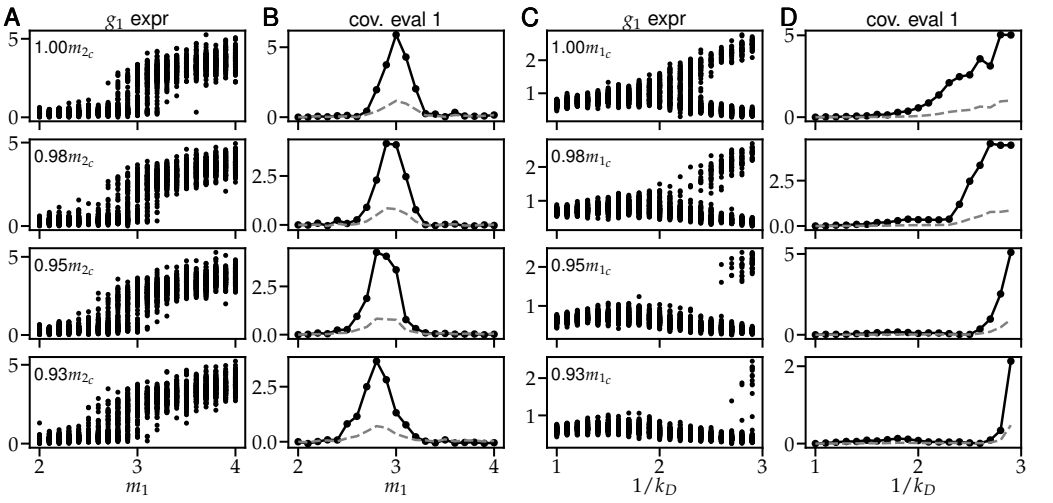

Fig. S8. The effect of small errors in parameter values for saddle node and pitchfork bifurcations. (A) Results of the saddle node bifurcation with 0,2%,5%, and 7% error in the value of the $m2$ parameter in Eqn. 5, while $m1$ is varied to induce a saddle-node bifurcation. (B) Principal covariance eigenvalue $\omega 1$ and its corresponding null, shifted to have min 0, for the data from part *a*. (C-D) Same as (A-B), but for errors in the $m1$ parameter, while k_D is varied to induce a pitchfork bifurcation.

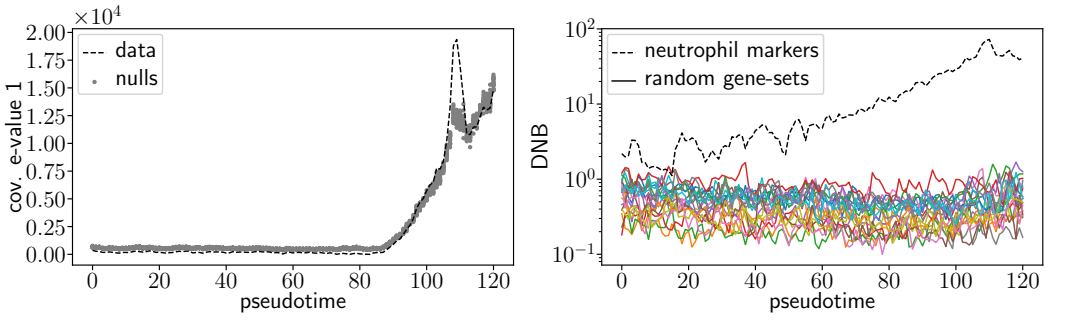

Fig. S9. Statistical nulls for bifurcation order parameters. Left: Principal covariance eigenvalue for the gene expression bins of the neutrophil trajectory (dashed line), and the randomly resampled gene expression bins (gray points). Right: DNB for neutrophil marker genes (dashed line) and for $n = 20$ different randomly selected gene sets. The randomly selected gene sets had the number of genes as the number of neutrophil markers.

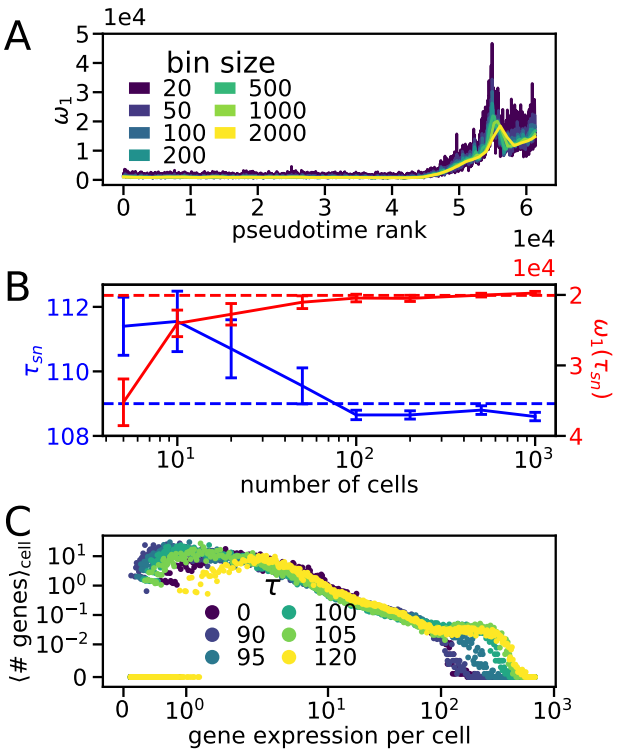

Fig. S10. Distributional properties of the neutrophil pseudotime trajectory. (A) Effect of varying the bin size on principal covariance eigenvalue. (B) Effect of undersampling a bin of 1000 cells on the detected saddle-node bifurcation time and magnitude. (C) Distribution of gene expression during the trajectory.

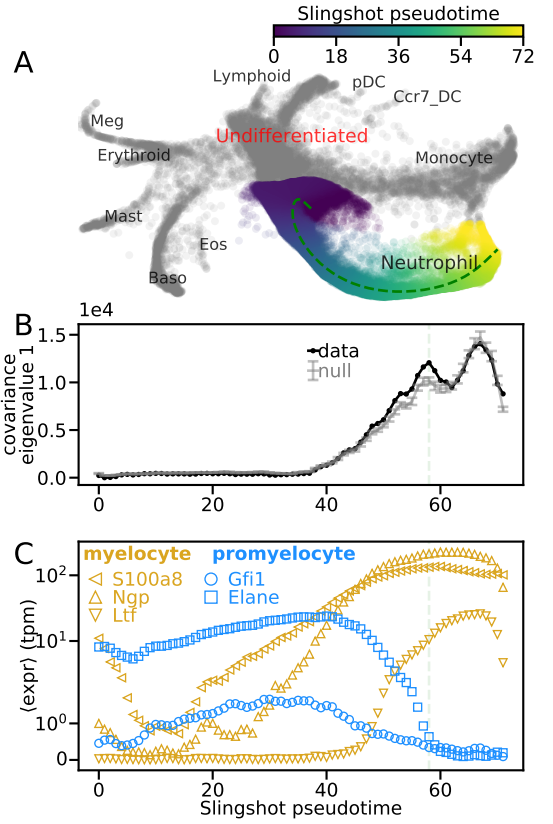

Fig. S11. Bifurcation characterization using Slingshot pseudotime algorithm. (A) Neutrophil development obtained by applying Slingshot to hematopoiesis scRNA-seq data [3]. Principal curves were approximated to 1000 points by setting approx points = 1000 in the slingshot function as increasing approx points further did not affect results. (B) Largest covariance eigenvalue (black) compared with a statistical null (gray, details in Section 4) in each 1000 cell pseudotemporal bin, shifted to have 0 min, using the Slingshot pseudotime ordering. Error bars of null are one SD. (C) Average expression of promyelocyte (blue) and myelocyte (gold) marker genes in Slingshot pseudotemporal bins [3]. SEM error bars are smaller than symbols. Light green line in (B-C) indicates peak of bifurcation window.

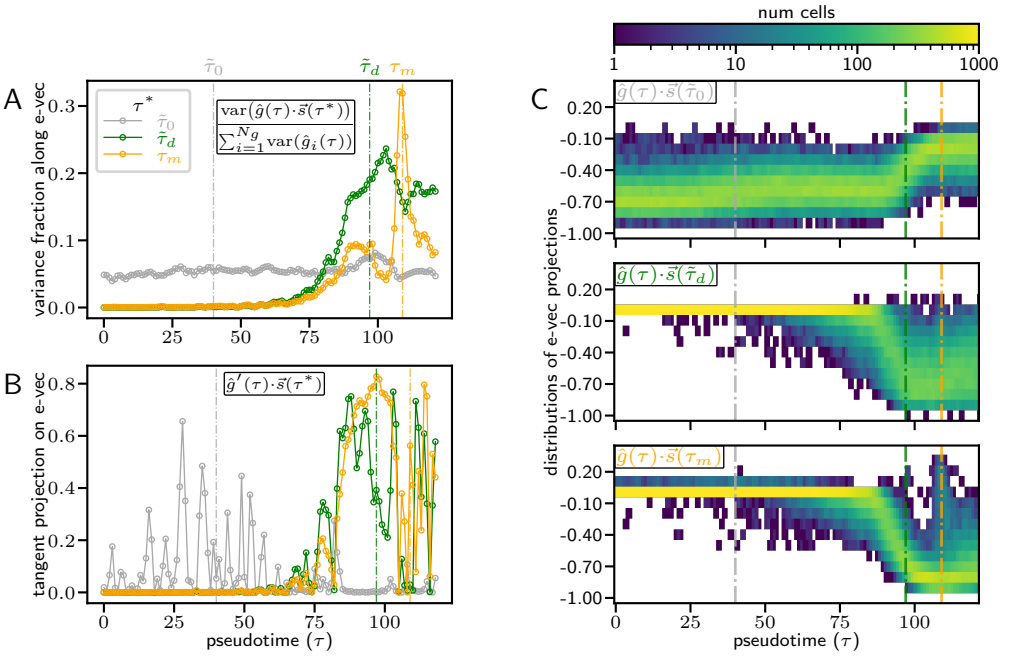

Fig. S12. Projections of gene expression onto bifurcation eigenvectors. (A) Fraction of variance along each of the three eigenvectors, as a function of pseudotime. (B) Mean projection of eigenvectors onto $g'(\tau)$, the normalized vector tangent to gene expression (obtained via finite difference). (C) Distribution of normalized gene expression for each cell projected onto the covariance eigenvector at τ_0 (left) τ_d (center) and τ_m (right).

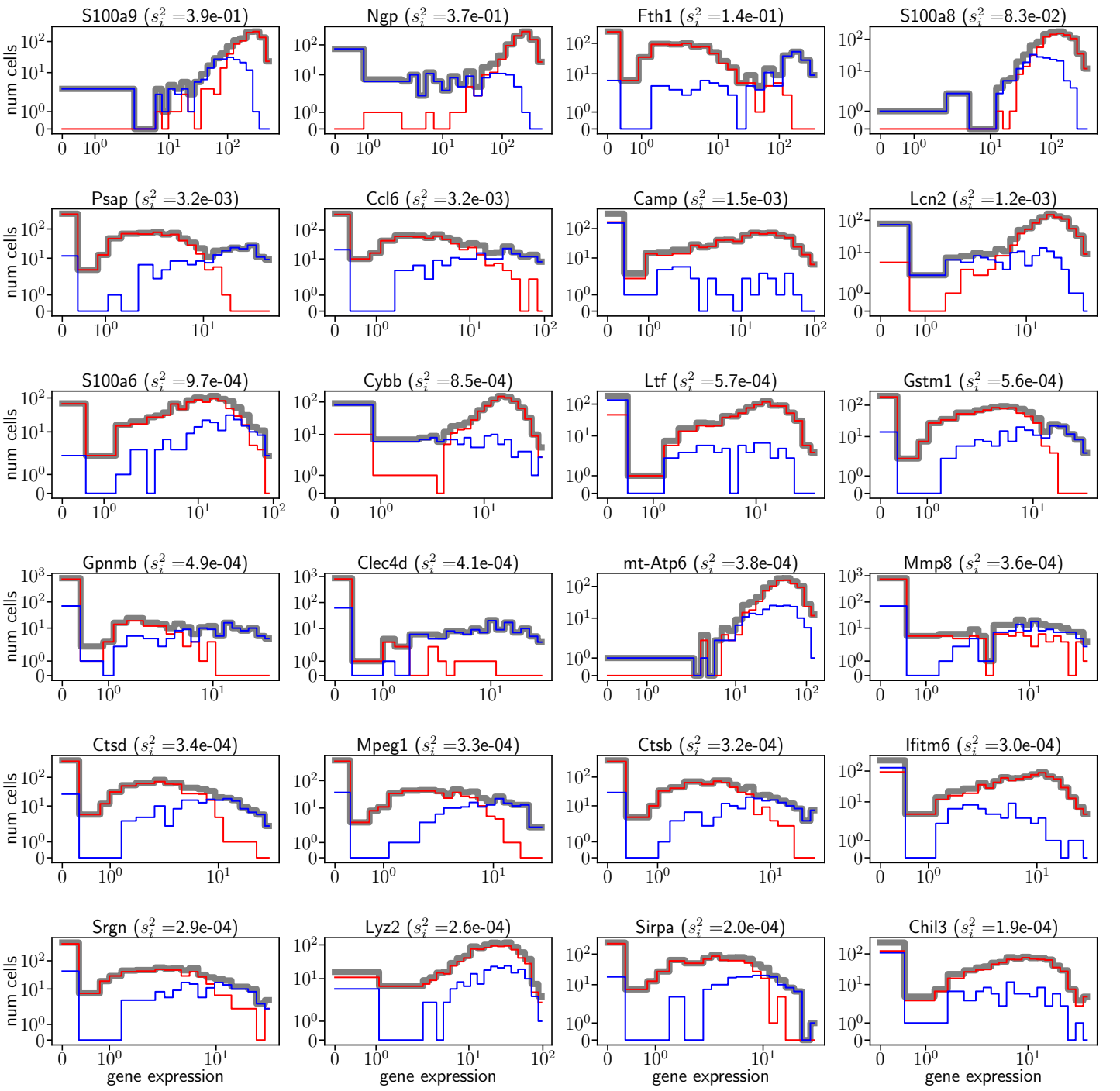

Fig. S13. Distribution of gene expression for genes with the highest absolute weight in $\vec{s}_m$. Red and blue indicate number of cells in each cluster (as in Fig. 5C) while gray indicates total.

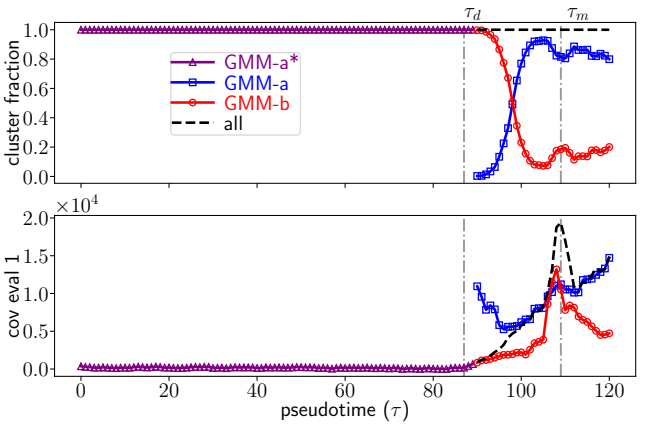

Fig. S14. Top: Fraction of cells in each of the two GMM clusters. Prior to the split between clusters, all cells are classified as GMM-a. Bottom: Covariance eigenvalue at pseudotimes for the full dataset (black) and each of the GMM clusters (red and blue). Dot-dashed vertical lines indicate τ_d and τ_m.

Appendix S1

1 The relationship between the Jacobian and the Covariance

Here, we outline the methodological framework that enables characterizing cell-fate transitions directly from scRNA-seq snapshots of a cell's transcriptomic state. A scRNA-seq measurement yields a transcriptomic matrix, where each row is a different cell and each column is a different gene (Fig. S1, left). This data is often visualized via dimensionality reduction algorithms, that reduce the $25,000$ dimensional gene space to two or three axes of variation, and sorted via parametric curve fitting tools, that show how the cells vary as a function of a control parameter, such as developmental time (pseudotime). Thus, one can compute statistics, such as the covariance C of the genes at a given pseudotime window.

Assuming that the underlying biochemical processes (1) are stochastic and Markovian and (2) occur at significantly faster timescales (seconds to minutes) than the timescales over which transitions in cellular fates are observed (hours to days), then the local time evolution of a cell's transcriptomic profile is controlled by a single matrix, the Jacobian ($\mathbf{J}$), where $J_{ij} = \partial \dot{g}_i / \partial g_j$ is the effect of the amount of gene j on the dynamics of gene i (Fig. S1, center). While $\mathbf{J}$, in general, changes with pseudotime, it relates to the covariance of gene expression at that pseudotime $\mathbf{C}$ through the continuous-time Lyapunov equation [4],

$$\mathbf{JC} + \mathbf{CJ}^T + \mathbf{D} = 0 \tag{S1}$$

where $\mathbf{D}$ is the expected noise amplitude for individual genes and their interactions (derivation in Methods: Continuous time Lyapunov equation for transcriptomic matrices) [5]. An important result from this relationship is that in the vincinity of bifurcations, the most salient properties of $\mathbf{J}$, corresponding to its eigen-decomposition, are inferrable from the eigendecomposition of $\mathbf{C}$.

We demonstrate the intuition behind the Eqn. S1 using a one-dimensional toy-model (Fig. S1, right). The slope of the potential function, drawn in red, provides the deterministic features of the system's dynamics. Parameter regimes ($a \ll 0$ and $a \gg 0$) where the potential has highly convex curvature exhibit stable fixed points, while parameter regimes near the bifurcation ($a \sim 0$), that have much flatter curvature, exhibit instability. Stochastic simulations of the system (drawn as open circles in Fig. S1– color corresponds to the value of the control parameter) demonstrate that owing to the reduction in curvature of the underlying potential, the data is spread maximally near the bifurcation, and narrows on either side of it.

This simple one-dimensional toy model captures the essence of the ideas used in this paper. If a complex high-dimensional dynamical system undergoes a bifurcation, then in its vicinity there must be, by definition, some direction in the high-dimensional space with greatly enhanced fluctuations. Thus bifurcations, and regions of multistability, can be located by finding the points along a developmental trajectory in transcriptomic space where the covariance eigenvalue spectrum is dominated by a single principle mode. Moreover, the direction of those fluctuations (the corresponding covariance eigenvector) is equivalient to the soft direction along which the system bifurcates (the corresponding eigenvector of the Jacobian), even in the $25,000$ dimension transcriptomic space.

2 Methodological relationship to Dynamical Network Biomarkers

Chen et al. [6], developed the concept of a dynamical network biomarker (DNB), a group of genes that drive a critical transition and are detectable from high dimensional gene expression datasets. In particular, they define an indicator function

$$I = \frac{SD_d \cdot |PCC_d|}{|PCC_o|} \tag{S2}$$

where SD_d is the average standard deviation of genes in the DNB, PCC_d is the average correlation coefficient between genes in the DNB, and PCC_o is the average correlation coefficient between genes in the DNB and genes outside the DNB [6]. At a critical state transition, or bifurcation, I is predicted to diverge, because SD_d and $|PCC_d|$ become large, while $|PCC_o|$ becomes small. Mathematically, the genes in the DNB correspond to those that have non-zero weight in the direction of the transition, i.e., $\vec{p}_d^{\,i} \neq 0$, where $\vec{p}_d$ is the principal eigenvector of the Jacobian, while genes outside of the DNB have $\vec{p}_d^{\,i} = 0$. This prediction is qualitatively similar, but not the same as Eqn. 2. In particular, while both SD_d and ω_1 increase at a bifurcation, they are not equivalent, as SD_d measures the variance of each individual gene, while ω_1 measures the variance across all genes, and therefore accounts for corrections to the total variance due to covariances between genes in the network. Therefore, for bifurcation detection, we focus solely on ω_1, instead of incorporating correlations into the indicator as in Eqn. S2.

As for determining which gene relationships are critical for the bifurcation, we take a similar approach to Refs. [6, 7], in focusing on the correlations that approach ± 1 at the bifurcation. This is justified via Eqn. 4, which yields that $R_{ij} \to \pm 1$ if $\vec{p}_d^{\,i} \neq 0$ and $\vec{p}_d^{\,j} \neq 0$. Interestingly, while we derived Eqn. 4 via the eigendecomposition of the covariance matrix, Refs. [6, 7] derived the same result form the covariance matrix itself, providing additional support to this method.

3 Bifurcations possibilities from two mutually inhibiting genes

At steady state, Eqn. 5 satisfies the quintic polynomial

$$g_1 = \frac{m_1/k_D}{\left(\frac{m_2/k_D}{g_1^2+1}\right)^2 + 1} \tag{S3}$$

which, depending on the parameter values, can have one real solution that is an attractor (e.g., if $m_{1,2} = 1$ and $k_D = 1$) or three real solutions, two attractors (nodes) and one repellor (saddle) (e.g., $m_{1,2} = 1, k_D = 1/3$). By examining the null clines,

$$g_1(g_2) = \frac{m_1/k_D}{g_2^2 + 1} \tag{S4}$$

$$g_1(g_2) = \sqrt{\frac{m_2}{k_D g_2} - 1} \tag{S5}$$

it can be deduced that varying m_1, while fixing τ and m_2 can yield a saddle-node bifurcation, as Eqn. S4 moves vertically while Eqn. S5 does not, allowing for either node to merge with the saddle (Fig. S2A).

Conversely, varying k_D, while fixing $m_{1,2}$ and m_2, can yield a pitchfork bifurcation, as both null clines move, such that above the bifurcation value, all three real solutions remain (Fig. S2B). Solving Eqn. S3 computationally via the Python function numpy.roots and plotting the real solutions (Fig. S2C-D) yields the bifurcations used in Fig. 2 and Fig. S5 [8].

4 Resampling principal eigenvalue

Given the transcriptomic matrix $\mathbf{G} = \left\{\vec{g_1}^T, \vec{g_2}^T, \ldots \vec{g_{n_g}}^T\right\}$, where $\vec{g}_i = \{G_{1,i}, G_{2,i}, \ldots, G_{n_c,i}\}$ and $G_{i,j}$ is the expression of the j^{th} gene in the i^{th} cell, we generate a null sample $\mathbf{G}^{\text{null}}$ by drawing each of its entries $G_{i,j}^{\text{null}}$ randomly, with replacement, from $\vec{g}_i$. In Fig. S10Fig. 2,Fig. 3, we compute the principal covariance eigenvalue ω_1^{null} for each of $n_s = 20$ samples, and compare this null distribution against the principal covariance eigenvalue of the data ω_1. This resampling technique has little impact on ω_1 for unimodal distributions as the scale of ω_1 is still determined by the system's noise (Fig. S3 left and right), but significantly decreases ω_1 for multimodal distributions (Fig. S3 center) since the structure of the multimodality is scrambled; thus we found it was an effective method for determining if a spike in ω_1 is due to multimodality or increased noise.

5 Noise induced transitions

To determine if a non-bifurcating noise-induced transition model [9, 10] could yield a similar covariance eigenvalue signature to a bifurcation, we ran the 102 gene network model (Fig. 2A) in a regime of the dynamical system that had two fixed-points ($m_{1,2} = 1$, $k_D = 1/3$) at varying noise scales s (see Fig. S2 and Eqn. 22 for details). To ensure a transition, we initialized all cells to populate the fixed point with higher g_1. We found that for low noise values ($1/s \leq 0.01$) the cells stayed near their initial fixed point, yielding a unimodal distribution for g_1 (Fig. S6A) and low principal covariance eigenvalue (Fig. S6B) while for high noise values ($1/s \geq 0.02$) the cells visited both fixed points, yielding a bimodal distribution for g_1, and a high principal covariance eigenvalue.

6 Effect of small errors

To better understand why the difference between ω_1 and its corresponding null was significantly more apparent at τ_m than τ_d (Fig. 3C), we exmained how small errors in the model parameters effect bifurcations. Specifically, we simulated the GRN model (Eqn. 5) with different amounts of error in other parameters. For the saddle-node bifurcation, in which m_1 is varied while τ_D and m_2 remain fixed, we perturbed m_2 by small amounts from its bifurcation value $m_{2_c} = 3$. We found (Fig. S8A) that the bifurcation was still largely detectable, and its eigenvalue still well distinguished from its null (Fig. S8B), at these small errors. For the pitchfork bifurcation, in which k_D is varied while m_1 and m_2 remain fixed, we perturbed m_1 by small amounts from its bifurcation value of $m_{1_c} = 1$. In this case, we found that the small perturbations biased the bifurcation toward one of the branches (Fig. S8C). This bias significantly reduces the difference ω_1 and its corresponding null (Fig. S8D). Our analysis suggests that small errors in the one-to-many bifurcating dynamical systems that appears present at τ_D may prevent it from being easily detectable, even when similar sized errors do not obscure the one-to-one bifurcation at ω_1.

7 Pseudotime inference

7.1 Algorithm for generating the pseudotime labels in Weinreb et al

SPRING (x-y) positions, cell type annotations, and pseudotime ranks for the data presented in Fig. 3A-B were downloaded from `https://github.com/AllonKleinLab/paper-data/tree/master/Lineage_tracing_on_transcriptional_landscapes_`

`links_state_to_fate_during_differentiation`. The algorithms to generate these values are described in detail in Ref. [3] (Supplementary Materials) and recapitulated here for completeness. Given the full *in-vitro* hematopoiesis transcriptomic matrix (all cells and all genes), the SPRING positions in Fig. 3A plot were generated using the following procedure.

1. A filtered transcriptomic matrix was generated which did not include genes that
   (a) had low variability as determined via the filter_genes function with parameters (85,3,3) from `https://github.com/AllonKleinLab/SPRING_dev/blob/master/data_prep/spring_helper.py` [1].
   (b) correlated highly ($R > 0.1$) across all cells with any of the following cell cycle genes: Ube2c, Hmgb2, Hmgn2, Tuba1b,Ccnb1, Tubb5, Top2a, and Tubb4b.
2. The top 50 principal components (PC) of the filtered transcriptomic matrix were computed.
3. 40,000 of the cells were selected randomly, and a k-nearest-neighbors (KNN) graph between those cells was constructed using the top 50 PC of the filtered transcriptomic matrix and k=4.
4. X-Y positions of these 40,000 cells were generated using the ForceAtlas2 algorithm with 500 steps [11].
5. Positions for each of the remaining 90,887 cells were computed as the average position of their 40 nearest neighbors (in the 50-PC space) among the initial 40,000 cells.

Cells were annotated with their cell types (cluster annotation in Fig. 3A) based on their position in the SPRING plot and their expression (terminal cell fates) or lack of expression (pluripotent) of pre-selected marker genes. Specifically the marker genes used to determine if cells were neutrophils were S100a9, Itgb2l, Elane, Fcnb, Mpo, Prtn3, S100a6, S100a8, Lcn2, and Lrg1.

Neutrophil pseudotime rank was then determined by smoothly interpolating between cells in the pluripotent and neutrophil clusters. The interpolation method used throughout this procedure is an iterative, diffusive process defined as

$$\begin{aligned} S_0(\mathbf{X}, b, i, k) =& \vec{x}_i \\ S_n(\mathbf{X}, b, i, k) =& bS_{n-1}(\mathbf{X}, b, i, k) \\ &+ \frac{1-b}{k} \sum_{j \in K_k(i)} S_{n-1}(\mathbf{X}, b, j, k) \end{aligned} \tag{S6}$$

where $\vec{x}_i$ is a vector quantity defined for cell i, $\mathbf{X} = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_{n_c}\}$ is the matrix of this quantity for all cells, $K_k(i)$ are the cell indices of the k nearest neighbors of cell i, $n > 0$ is the number of iterations, and b is the neighbor weight (low b and high n both yield high diffusion) [12]. The pseudotime ranking procedure is:

1. Cells are identified to be part of the neutrophil trajectory
   (a) Let $\vec{t}_i$ be an indicator vector for the cell type of i; i.e. $t_{ij} = 1$ if cell i is type j and 0 otherwise. Let $\mathbf{T} = \{\vec{t}_1, \vec{t}_2, \ldots, \vec{t}_{n_c}\}$ be the corresponding matrix for all cells.
   (b) Let $\mathbf{K}_{100}$ be the k-nearest-neighbor graph between cells for $k = 100$ using the top 50 PC.
   (c) Let $\hat{t}_i = S_{250}(\mathbf{T}, 0.1, i, 100)$ be the smooth cell type indicator.
   (d) Let $z_i = \sum_j a_j \hat{t}_{ij}$ be the weighted average cell type $\hat{t}_i$ where the weights for each cell type (j) are

$$a_j = \begin{cases} 0.1 & \text{if neutrophil or pluripotent} \\ -2 & \text{if megakaryocyte} \\ -1 & \text{otherwise} \end{cases} \tag{S7}$$

   (e) Let $\vec{c}_i$ be a neutrophil trajectory indicator such that $\vec{c}_i = \{1\}$ if $z_i > Q_{0.6}(z)$ and $\{0\}$ otherwise, where $Q_{0.6}(z)$ is the 60^{th} quantile of z. Let $\mathbf{C} = \{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_{n_c}\}$.
   (f) Let $\hat{c}_i = S_{50}(\mathbf{C}, 0.1, i, 100)$ be the smoothed neutrophil trajectory indicator.
   (g) Cells were considered part of the neutrophil trajectory if $\hat{c}_i > Q_{0.6}(\hat{c})$ where $Q_{0.6}(\hat{c})$ is the 60^{th} percentile of $\hat{c}$.
2. The $61,310$ cells identified as part of the neutrophil trajectory are sorted
   (a) Let $\vec{p}_i = \{1\}$ if a cell in the trajectory is pluripotent and 0 otherwise; i.e., it is an indicator for pluripotency. $\mathbf{P} = \{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_{n_c}\}$ is the corresponding matrix for all cells in the trajectory.
   (b) Let $\hat{p}_i = S_{300}(\mathbf{P}, 0.1, i, 100)$ be the smoothed pluripotency indicator.
   (c) The pseudotime of cell i is the rank (largest to smallest) of $\hat{p}_i$ among all $\hat{p}$.

7.2 Pseudotime inference in the absence of metadata

To test if the neutrophil bifurcation characterization was dependent on the choice of pseudotime algorithm, we used the Slingshot algorithm [2] to compute the pseudotime of each cell for its trajectory from the undifferentiated cluster to each of the terminal fate clusters. The input to Slingshot were the cells' cluster labels and their SPRING coordinates, and the output was a probability, or weight, that a cell belonged to each undifferentiated-to-terminal-fate trajectory, as well as its pseudotime along that trajectory. In Fig. S11A, we show the pseudotime of all cells that had $weight > 0$ for belonging to the trajectory that led from undifferentiated cells toward neutrophils. Unlike the pseudotime method described in Section 7.1, the origin of the trajectory does not coincide with the earliest sequenced cells, as time of sequencing and clonal barcode data could not be input to Slingshot. Nevertheless, we obtain a clear bifurcation signature in the principal covariance eigenvalue (Fig. S11B) at the point where promyelocyte gene expression decreases to 0 and myelocyte marker gene expression become maximal (Fig. S11C). This result supports our belief that the bifurcation characterization does not depend on the specific pseudotime calculation.

8 Determining the eigenvectors for analysis

In order to analyze the neutrophil trajectory in a native-space, we chose eigenvectors that were characteristic of the dynamics. Since τ_m coincides with a well defined eigenvalue peak in the neutrophil trajectory, it was natural to use $\vec{s}(\tau_m)$ to aid in visualizing the trajectory and further probe mechanisms. However, $\tau = 0$ and τ_d coincide with transition points between states (Fig. 5B), and mark the beginning of specific dynamics (i.e., the eigenvalue remaining constant, or increasing), and it the lower correlation on the edges of the blocks in Fig. 5B suggests that the eigenvectors at those points had not equiilibrated to their new positions. Therefore, we define $\tilde{\tau}_0$ and $\tilde{\tau}_d$ as the pseudotime bins with the eigenvector closest to the eigenvector at all other pseudotimes in that range, i.e.,

$$\tilde{\tau}_0 = \underset{0<=\tau<\tau_d}{\arg\min} \sum_{t=0}^{\tau_d - 1} ||\vec{s}(\tau) - \vec{s}(t)||^2 \tag{S8}$$

$$\tilde{\tau}_d = \underset{\tau_d<=\tau<\tau_m}{\arg\min} \sum_{t=\tau_d}^{\tau_m - 1} ||\vec{s}(\tau) - \vec{s}(t)||^2 \tag{S9}$$

and use these pseudotimes for downstream analysis.

9 Identifying clusters via Gaussian Mixture Models

As the distribution of gene expression projected onto $\vec{s}(\tau_m)$ exhibited bimodality (Fig. 5E, Fig. S12B), we used a Gaussian Mixture Model to separate the two modes. Specifically, we fit $\mathbf{G}(\tau_m)$, the normalized gene expression matrix at τ_m to a two component Gaussian Mixture Model using the `mixture.GaussianMixture` function from the `Python` package `scikit-learn` with $n\_components = 2$ and all other parameters set to their default [13]. We then used the `predict` function of our trained model to generate cluster labels for cells at all pseudotimes. We found that cells were predicted to belong to the same cluster (GMM-a) for $\tau \lesssim \tau_d$ (purple in Fig. 5F and Fig. S14). For $\tau \gtrsim \tau_d$, cells were split between the two clusters (red and blue in Fig. 5F and Fig. S14).

References

[1] Caleb Weinreb, Samuel Wolock, and Allon M Klein. Spring: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics*, 34(7):1246–1248, 2018.

[2] Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1):477, 2018.

[3] Caleb Weinreb, Alejo Rodriguez-Fraticelli, Fernando D Camargo, and Allon M Klein. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 367(6479), 2020.

[4] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[5] Makito Oku and Kazuyuki Aihara. On the covariance matrix of the stationary distribution of a noisy dynamical system. *Nonlinear Theory and Its Applications, IEICE*, 9(2):166–184, 2018.

[6] Luonan Chen, Rui Liu, Zhi-Ping Liu, Meiyi Li, and Kazuyuki Aihara. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Scientific reports*, 2(1):1–8, 2012.

[7] Mitra Mojtahedi, Alexander Skupin, Joseph Zhou, Ivan G Castaño, Rebecca YY Leong-Quong, Hannah Chang, Kalliopi Trachana, Alessandro Giuliani, and Sui Huang. Cell fate decision as high-dimensional critical state transition. *PLoS biology*, 14(12):e2000640, 2016.

[8] Charles R. Harris, K. Jarrod Millman, St'efan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fern'andez del R'ıo, Mark Wiebe, Pearu Peterson, Pierre G'erard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[9] Caleb Weinreb, Samuel Wolock, Betsabeh K Tusi, Merav Socolovsky, and Allon M Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, 2018.

[10] Peijie Zhou, Shuxiong Wang, Tiejun Li, and Qing Nie. Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. *bioRxiv*, 2021.

[11] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9(6):e98679, 2014.

[12] Betsabeh Khoramian Tusi, Samuel L Wolock, Caleb Weinreb, Yung Hwang, Daniel Hidalgo, Rapolas Zilionis, Ari Waisman, Jun R Huh, Allon M Klein, and Merav Socolovsky. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*, 555(7694):54–60, 2018.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.