Single-cell Ca^{2+} parameter inference reveals how transcriptional states inform dynamic cell fates

Xiaojun Wu[1], Roy Wollman[2], and Adam L. MacLean[1,*]

[1]Department of Quantitative and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA
[2]Departments of Integrative Biology and Physiology and Chemistry and Biochemistry, University of California Los Angeles. Los Angeles, CA
[*]Correspondence: macleana@usc.edu

Abstract

Inferring the parameters of dynamic models is a cornerstone of systems biology. Large single-cell transcriptomic datasets have opened up many possibilities for new analyses, but their potential to inform parameter inference of molecular or cellular dynamics has not yet been realized. Here, making use of coupled data: single-cell gene expression and dynamic molecular measurements, we develop new methods for parameter inference. We construct cell chains in which the posterior distribution of a cell is used to inform the prior of the subsequent cell in the chain. In application to the Ca^{2+} signaling pathway, we show that cell predecessor-informed priors accelerate inference of the Ca^{2+} model parameters in single cells. Though use of cell chains informed by single-cell gene expression does not improve sampling relative to random chain assignment, we show that the posteriors produced via gene expression-informed cell chains capture distinct properties of the dynamic Ca^{2+} response. By clustering posterior parameters we can identify markers genes that correspond with variable Ca^{2+} responses. Additionally, through analysis of the posterior distributions of hundreds of single cells, we discover that divergent co-variation of parameters within and between cells, highlighting the complex and competing sources of cell heterogeneity. Through the analysis of large populations of posterior distributions we are able to quantify the relationships between single-cell transcriptional states and dynamic cellular responses, paving the way for more detailed mappings between gene expression states and dynamic cell fates.

1 Introduction

Models in systems biology span systems from the scale of protein/DNA interactions to cellular, organ, and whole organism phenotypes. Their assumptions and validity are assessed through their ability to describe biological observations, often accomplished by simulating models and fitting them to data [1, 2, 3, 4]. Under the framework of Bayesian parameter inference and model selection, the available data is used along with prior knowledge to infer a posterior parameter distribution for the model [5]. The posterior characterizes not only the most likely parameter values to give rise to the data but also the uncertainty that we have regarding those parameters. Thus parameter inference provides a map from the dynamic phenotypes that we observe from experiments and the parameters of a mathematical model. However, much less available are tools that give the ability to draw a map between the dynamic phenotypes of interest that are modeled, and the current state of a cell, as quantified by high-throughput genomic or proteomic measurements.

Single-cell genomic measurement technologies have given us access to a wealth of information on the states of single cells that was not previously accessible [6]. Take single-cell gene expression as an example: even if the overwhelming majority of the genes measured in a given experiment contain very little information about the phenotypes under investigation (say 90-95%, for sake of argument) that still leaves hundreds or thousands of genes that do contain relevant information. Even with only a few hundred genes retained, this is more than enough in theory to characterize any possible number of cell states or arbitrarily complex dynamical processes.

This leads us to the central question underlying this work: can the integration of single-cell gene expression data into a framework for dynamic modeling and inference improve our understanding of the cellular phenotypes of interest? We address this question below using novel computational methods for single-cell data-informed parameter inference of Ca^{2+} dynamics coupled with an innovative dataset jointly measuring dynamics and gene expression in the same single cells [7].

Ca^{2+} signaling is a highly conserved pathway that regulates a host of cellular responses in epithelial cells: from death and division to migration and molecule secretion, as well as collective behaviors from organogenesis to wound healing [8]. In response to adenosine triphosphate (ATP) binding to purinergic receptors, a signaling cascade is initiated whereby phospholipase C (PLC) is activated and in turn hydrolyzes phosphatidylinositol 4,5-bisphosphate (PIP_2), producing inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). The endoplasmic reticulum (ER) responds to IP_3 by the activation of Ca^{2+} channels: the subsequent release of calcium from the ER into the cytosol produces a spiked calcium response. To complete the cycle and return cytosolic calcium levels to steady state, the sarco/ER Ca^{2+}-ATPase (SERCA) channel pumps the Ca^{2+} from the cytosol back into the ER [9, 10]. This dynamic Ca^{2+} response to ATP stimulus occurs quickly: on a timescale that is faster than that of transcription, permitting the direct study of links between the dynamic phenotype of Ca^{2+} response to ATP stimulus and the transcriptional state of the cell.

The ability to measure gene expression global in many single cells per experiment has not only led to many new discoveries but has also fundamentally changed the means by which we identify and characterize cell states [11]. Technologies used to measure gene expression in single cells include sequencing and fluorescent imaging. The latter permits the quantification of

hundreds of genes in spatially-resolved populations of single cells. Small molecule fluorescence in situ hybridization (smFISH) can be multiplexed to achieve this high-resolution picture of gene expression in single cells either via MERFISH [12] or seqFISH [13]. By coupling fluorescent imaging of Ca^{2+} responses using a GFP reporter with multiplexed smFISH in MCF10A cells, we are able to jointly capture dynamic cell responses and single-cell gene expression in the same single cells [7]. These data offer great potential to study the relationship between the transcriptional states of cells and their corresponding phenotypes.

Models of gene regulatory networks and cellular signaling pathways built using ordinary differential equations (ODEs) can describe the dynamic interactions between gene transcripts, proteins, or other molecular species and their impact on cellular phenotypes. Well-established dynamical systems theory offers a wide range of tools with which to analyze the transient and equilibrium behavior of ODE models [14]; it remains under investigation the extent to which equilibrium behaviors are appropriate for the characterization of (dynamic) cells [15]. Here we derive an ODE model of Ca^{2+} dynamics based on previous work [16, 17] in order to study the informativeness of single-cell transcriptional states on dynamic cell responses. We developed a parameter inference scheme to fit the Ca^{2+} dynamics in single cells informed by cell predecessors in a cell chain. We use this framework to assess the extent to which transcriptionally similar cell states inform dynamic cellular responses.

In the next section we present the model and the methods associated with parameter inference using Hamiltonian Monte Carlo in Stan [18]. We go on to assess the inference framework: we discover that informative priors (informed by cell predecessors) accelerate parameter inference in single cells, but that cell chains where predecessors are sampled randomly perform as well those in which predecessor are based on transcriptional similarity. We go on to analyze the results of fitting the Ca^{2+} dynamics in hundreds of single cells and discover that cell-intrinsic vs. cell-extrinsic posterior parameter correlations can differ widely, indicative of fundamentally different sources of underlying variability. Analysis of posterior distributions offers an intuitive way to perform parameter sensitivity analysis in response to Ca^{2+} spiking, through which we identify a wide range of sensitivities. We show that variability in single-cell gene expression is associated with variability in posterior parameter distributions, both via targeted analysis of variable gene-parameter pairs and via global dimensionality reduction in posterior space. Based on these findings, we cluster cells based on their posterior distributions, and discover that clustering the posterior of cell chains derived by transcriptional similarity reveals deep insights into the information conveyed about dynamic phenotypes by single-cell gene expression. In addition to the new insight gained about Ca^{2+} signaling dynamics, the modeling and inference framework we present can be applied broadly to cellular systems biology in other contexts to link dynamic phenotypes with transcriptional states of single cells.

2 Materials and Methods

2.1 A model of Ca^{2+} dynamics in response to ATP

We model the Ca^{2+} pathway using nonlinear ordinary differential equations (ODEs), as previously developed [17, 16]. The model consists of four state variables: phospholipase C (PLC), inositol 1,4,5-trisphosphate (IP3), the fraction of IP3-activated receptor (h), and cytoplasmic Ca^{2+}. The four variables are associated with a system four nonlinear ODEs describing the rates

of change of the Ca^{2+} pathway species following ATP stimulation. The equations are given by:

$$\begin{aligned}
\frac{d[\text{PLC}]}{dt} &= \text{ATP} \cdot e^{-K_{\text{ATP}}t} - K_{\text{off, ATP}}[\text{PLC}] \\
\frac{d[\text{IP3}]}{dt} &= V_{\text{PLC}} \frac{[\text{PLC}]^2}{K_{\text{IP3}}^2 + [\text{PLC}]^2} - K_{\text{off, IP3}}[\text{IP3}] \\
\frac{dh}{dt} &= a([\text{Ca}^{2+}] + d_{\text{inh}}) \left(\frac{d_{\text{inh}}}{[\text{Ca}^{2+}] + d_{\text{inh}}} - h \right) \\
\frac{d[\text{Ca}^{2+}]}{dt} &= \beta \left(\varepsilon(\eta_1 m_\infty^3 h^3 + \eta_2)(c_0 - (1+\varepsilon)[\text{Ca}^{2+}]) - \eta_3 \frac{[\text{Ca}^{2+}]^2}{k_3^2 + [\text{Ca}^{2+}]^2} \right) \\
\beta &= \left(1 + \frac{K_e[B_e]}{(K_e + [\text{Ca}^{2+}])^2} \right)^{-1} \\
m_\infty &= \left(\frac{[\text{IP3}]}{d_1 + [\text{IP3}]} \right) \left(\frac{[\text{Ca}^{2+}]}{d_5 + [\text{Ca}^{2+}]} \right).
\end{aligned}$$

The equations describe a chain of activities following ATP binding on purigenic receptor, including activation of PLC, increase in IP3, activation of IP_3R channel on ER, and finally releasing of Ca^{2+} from ER into cytoplasm [17]. In addition, Ca^{2+} may also enter ER through the IP_3R channel and the SERCA pump on ER [17]. Our model differs from the model in Yao et al. [17] in that the product of two parameters, $K_{\text{on, ATP}}$ and ATP, is combined into one parameter ATP. A description of each of the parameters in the model is given in (Table 1), where reference values for each of the model parameters are found in Lemon et al. [16] and Yao et al. [17].

2.2 Data collection and preprocessing

The data used in this work consists of a joint assay measuring Ca^{2+} trajectories in response to ATP stimulation and single-cell gene expression via multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) [12]. Ca^{2+} dynamics were measured via imaging for 1000 seconds using a GCaMP5 biosensor in a total of 5128 human MCF10a cells, immediately followed by measurement of 336 genes by MERFISH [7]. Each cell was stimulated by ATP after 200 seconds of imaging. To reduce the effects of signal noise, we smoothed the Ca^{2+} trajectory for each cell using a moving average filter with a twenty-second window size. After smoothing, we removed all data points occurring before ATP stimulation begins. Data points for each Ca^{2+} trajectory after t=300 were downsampled by a factor of 10, since the trajectories are at or close to steady state after this time. The single-cell gene expression data was collected using MERFISH after Ca^{2+} imaging [7, 12].

2.3 Generating cell chains through cell-cell similarity

We propose to characterize relationships between cells through their membership and location in cell chains: paths through cell-cell similarity space that connect cells based on their single-cell transcriptional states. To do so, we first obtain a cell-cell similarity matrix W from log-transformed MERFISH data using SoptSC [20]. Each entry $W_{i,j}$ is the similarity score between cell i and cell j. Given the similarity matrix, we generate a chain of cells in two steps:

Name	Description	Prior distribution
ATP	Concentration of ATP that activates PLC	$\mathcal{N}(5, 4)$
K_{ATP}	ATP decay rate	$\mathcal{N}(0.0083, 0.0025)$
$K_{\text{off, ATP}}$	PLC degradation rate	$\mathcal{N}(1.25, 1)$
V_{PLC}	Maximum velocity for IP3 generation	$\mathcal{N}(1, 1)$
K_{IP3}	Equilibrium constant for IP3 generation through PLC	$\mathcal{N}(0.5, 0.01)$
$K_{\text{off, IP3}}$	IP3 degradation rate	$\mathcal{N}(1.25, 1)$
a	Time constant of IP3 channel	$\mathcal{N}(1, 1)$
d_{inh}	Dissociation constant for IP3 channel calcium inhibiting subunit	$\mathcal{N}(0.4, 0.01)$
K_e	Dissociation constant for calcium buffer	$\mathcal{N}(10, 4)$
B_e	Concentration of calcium buffer	$\mathcal{N}(150, 25)$
d_1	Dissociation constant for IP3 channel IP3 activating subunit	$\mathcal{N}(0.13, 0.01)$
d_5	Dissociation constant for IP3 channel calcium activating subunit	$\mathcal{N}(0.0823, 0.01)$
ϵ	ER to cytosolic volume	$\mathcal{N}(0.185, 0.01)$
η_1	IP3 channel permeability constant	$\mathcal{N}(575, 625)$
η_2	ER leak permeability constant	$\mathcal{N}(5.2, 1)$
η_3	Ca^{2+} pump permeability constant	$\mathcal{N}(45, 25)$
c_0	Concentration of free Ca^{2+} in the ER	$\mathcal{N}(4, 1)$
k_3	SERCA pump dissociation constant	$\mathcal{N}(0.4, 0.01)$

Table 1: Definition and description of the ODE model parameters. Prior distributions are derived from [16, 19, 17].

1. Construct a graph $G = (V, E)$, where nodes are cells and edges are placed between two cells if they have a similarity score above a given threshold;
2. For a choice of initial (root) cell, traverse G, recording the order of cells while traversing.

Ideally, on this traverse each cell would be visited exactly once; this amounts to finding a Hamiltonian path in G. However, finding a Hamiltonian path is an NP-complete problem, which means there is no known time-efficient algorithm for this task. Therefore, we use the depth-first search (DFS) algorithm as a heuristic solution. From the current node (initialized at the root node), the DFS algorithm randomly selects an unvisited neighboring node. If the current node has no unvisited neighbors, a DFS backtracks until a node with unvisited neighbors is found. When there is no unvisited node left, every node in the graph has been visited exactly once. In particular, we use pre-order DFS, which records a node as soon as it is visited. For a sparse similarity matrix (as we have in this case), DFS generates a tree that is very close to a straight path. A significant advantage of a DFS, compared to the naïve approach of searching for a Hamiltonian path, is that DFS can be completed in linear time.

2.4 Bayesian parameter inference in Stan with priors informed by cell predecessors

We seek to infer the parameters of Ca^{2+} pathway dynamics in single cells, where the inference is informed by a cell's position in a cell chain. For parameter inference we use efficient Markov chain Monte Carlo (MCMC) methods as implemented via Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS) [21, 18]. HMC gains speedup relative to other MCMC algorithms by treating the sampling process as a physical system and employing conservation laws [22]. Both HMC and NUTS implementations start from an initial distribution and converge to the stationary distribution after an intermediate phase of iterations; this intermediate phase is called the warmup. During the warmup stage, NUTS adjusts the HMC configuration automatically [21].

For each parameter, we set its prior according to a Normal distribution, denoted θ_j for $j = 1, \ldots, m$, where m is the number of parameters. Let f be the function characterizing the trajectories of the state variables of the ODE model, and y_0 be the initial condition. Then, in each single cell, the Ca^{2+} response to ATP is generated by the following process:

$$\begin{aligned}
\theta_j &\sim \mathcal{N}(\mu_{\theta_j}, \sigma_{\theta_j}) \\
\widehat{y}(t) &= f(y_0, t; \theta) \\
\sigma &\sim \text{Cauchy}(0, 0.05) \\
y(t) &\sim \mathcal{N}(\widehat{y}(t), \sigma).
\end{aligned}$$

Where the noise around the Ca^{2+} trajectories is given by σ. In addition we truncate the prior so that each θ_i is bounded by 0 from below.

For the first cell in a cell chain, we use the Lemon prior (Table 1), a Gaussian prior of which the means are taken from Lemon et al. [16]. For the ith cell in a cell chain where $i > 1$, the prior distribution of its parameters is constructed from the posterior distribution for the $(i-1)$th cell, as detailed in Section 2.5. We pass the constructed prior and measured Ca^{2+} trajectory to NUTS, and run NUTS with 4 independent sampling chains of identical setup. For simulating $\widehat{y}(t)$ during sampling, we use Stan's implementation of fourth and fifth order Runge-Kutta method [18].

We evaluate convergence of NUTS chains using the $\widehat{R}$ statistic, the ratio of between-chain variance versus within-chain variance [18, 23]. We choose $\widehat{R}$ over the distance between input and fitted trajectories because it is much harder to determine a good criteria for convergence based on the latter. We follow the heuristic that if $\widehat{R}$ is between 0.9 and 1.1 for a parameter, this indicates that the set of chains run for that parameter (in practice we use 4 chains per simulation) are well-mixed. Two caveats on the use of $\widehat{R}$ in practice:

1. Due to the lack of identifiability in the parameter space of the model, well-fit Ca^{2+} trajectories do not require $\widehat{R}$ in $(0.9, 1.1)$ for all parameters. Thus, during simulation, we assess $\widehat{R}$ only for the log posterior, and we use a more tolerant upper bound of 4.0.
2. There are cases where 3/4 chains are well-mixed but one diverges. In such cases of harder-to-fit cells, we chose to retain the three well-mixed chains as sufficient of a good fit. Thus if $\widehat{R}$ is outside of the acceptable range, before discarding we compute the $\widehat{R}$ for all combinations of three chains in the run, and retain the run with the three well-mixed chains if it exists.

2.5 Stabilizing prior variance along cell chain

We construct the prior distribution of the ith cell from the posterior of the $(i-1)^{th}$ cell. The prior mean for each parameter θ_j for the ith cell is set to $\mu_j^{(i-1)}$, the posterior mean of θ_j from the $(i-1)^{th}$ cell. The standard deviation for the prior of each θ_j is derived from $\sigma_j^{(i-1)}$, the posterior standard deviation of θ_j from the $(i-1)^{th}$ cell. To prevent instabilities (rapid growth or decline) in posterior values along the cell chain, we scale $\sigma_j^{(i-1)}$ by a factor of 1.5 and clip the scaled value to be between 0.001 and 5. The scaled and clipped value is then set as the prior standard deviation for θ_j for the ith cell.

2.6 Dimensionality reduction and sensitivity analyses

To visualize and compare posterior samples from the same cell chain, we used principal component analysis (PCA) implemented in scikit-learn 0.24 [24]. We projected all posterior samples onto the same subspace by first choosing a cell, which we call the focal cell, and preprocessing the posterior samples by normalizing them to this focal cell. We tested min-max normalization and z-score normalization. The min-max normalization transforms a vector x by $\frac{x-x_{\min}}{x_{\max}-x_{\min}}$, where $x_{\min}$ is the minimum of x and $x_{\max}$ the maximum. The z-score normalization transforms x by $\frac{x-\mu_x}{\sigma_x}$, where μ_x is the mean of x and σ_x the standard deviation. Given that we normalize to the focal cell, all cells used the same $x_{\min}, x_{\max}, \mu_x, \sigma_x$: the values from the focal cell. We then perform PCA on the normalized samples from the focal cell and created a subspace spanned by the first two principal components. The normalized samples from all other cells in the chain are then projected onto the PC1-PC2 subspace of the focal cell for comparison.

We implemented sensitivity analysis to assess the responses of Ca^{2+} trajectories to perturbations of various model parameters. Given the posterior distribution $\tilde{\theta}$ of a cell, each parameter θ_j is perturbed to two extreme values: $\tilde{\theta}_j^{(0.01)}$, the 0.01-quantile of $\tilde{\theta}_{\cdot,j}$, and $\tilde{\theta}_j^{(0.99)}$, the 0.99-quantile of $\tilde{\theta}_{\cdot,j}$. Before perturbing, we drew nine samples from $\tilde{\theta}$ that were "evenly spaced" over the marginal posterior distribution of the parameter of interest, $\tilde{\theta}_{\cdot,j}$: the kth draw corresponds to a sample $\tilde{\theta}_{i,\cdot}$ such that $\tilde{\theta}_{i,j} = \tilde{\theta}_j^{(0.1k)}$, the $0.1k$-quantile of $\tilde{\theta}_{\cdot,j}$. For each draw $\tilde{\theta}_{i,\cdot}$, we replace $\tilde{\theta}_{i,j}$ by either or $\tilde{\theta}_j^{(0.01)}$ or $\tilde{\theta}_j^{(0.99)}$ and then simulate a Ca^{2+} response. We use the Euclidean distances between simulated trajectories and measured trajectories to quantify the sensitivity of each parameter perturbation.

2.7 Analysis of MERFISH gene expression: correlation and cell clustering

Correlations between single-cell gene expression values and posterior parameters from the Ca^{2+} pathway model were determined for variable genes, chosen as follows. We calculated the z-scores of posterior means for each parameter of a cell sampled from a population of cells, and removed that cell if any of its parameters had a posterior mean z-score smaller than -3.0 or greater than 3.0. PCA was performed on log-normalized gene expression of remaining cells using scikit-learn 0.24 [24], which yields a loadings matrix A such that $A_{i,j}$ represents the "contribution" that gene i makes to component j. We designate gene i as variable if $A_{i,j}$ is ranked top 10 or bottom 10 in the jth column of A for any $j \leq 10$. For each variable gene, we calculated the Pearson correlation between its log-normalized expression value and the

posterior means of each model parameter. Gene-parameter pairs were ranked by their absolute Pearson correlations and the top 30 were selected.

To cluster cells according to their single-cell gene expression, the raw counts matrix was first normalized, log-transformed, and scaled to zero mean and unit variance before clustering, using processing routines in Scanpy 1.8 [25]. Clustering was then performed using the Leiden algorithm at 0.5 resolution, provided by Scanpy 1.8 [26]. Marker genes in each cluster are determined by t-test implemented in Scanpy 1.8 [25].

2.8 Clustering posterior parameter distributions

Cells were clustered according to their posterior distributions as output from NUTS. For each parameter, the posterior means of each cell were computed and scaled to $[0, 1]$. The distance between two cells was defined as the m-dimensional Euclidean distance between their posterior means (where m is the number of parameters). Given distances calculated between all pairs of cells, agglomerative clustering with Ward linkage was performed using Scipy 1.7 [27]. Marker genes for each cluster identified were determined using a t-test implemented in Scanpy 1.8 [25].

3 Results

3.1 Single-cell priors informed by cell predecessors enable computationally efficient parameter inference

The Ca^{2+} dynamic responses of single cells to ATP stimulation were modeled using Eqns. (1-4) and fit to data using a Bayesian parameter inference framework (Figure 1A). To quantify the effects of inference with cell chains (that is the use of prior distributions informed by cell predecessors) we fit cells using both cell chain-informed priors and individually, where the same prior is used for every cell. For individually fit cells, we used the "Lemon" prior, which was also used for the first cell in a cell chain (see details in Methods). We found that inferring single-cell Ca^{2+} parameters via a cell chain results in more efficient and more accurate parameter inference, with shorter computational times and higher overall posterior model probabilities (Figure 1B–C). Five hundred cells were fit using a similarity-based cell chain (see Methods) with 500 warmup steps, denoted *Similar-r1*, and compared to cells fit independently using either 500 or 1000 warmup steps; the longer warmup is required to produce fits of comparable quality (measured by the log posterior) to cells from *Similar-r1*. The posterior probabilities for models fit to cells from the *Similar-r1* are higher than those from individually fitted cells (Figure 1B; Table S1), with sampling times that were at least 2x and up to 25x faster than individually fitted cells (Figure 1C). In addition, the *Similar-r1* model fits as quantified by the $\widehat{R}$ statistic were better overall than those from individually fitted cells (Table S2). These trends were consistent across multiple runs consisting of hundreds of independent cells (Figure 1D). Thus, the use of informative priors derived from fits to other cells improves both the efficiency and the accuracy of inference.

We next investigated the effects of the choice of prior for cell chains on the inference results. We experimented with different ways of setting prior standard deviation and determined the scaling and clipping the prior standard deviation was necessary for stable posterior distribution in the long run (Figure S1D–I). We also compared single-cell parameter inference along cell

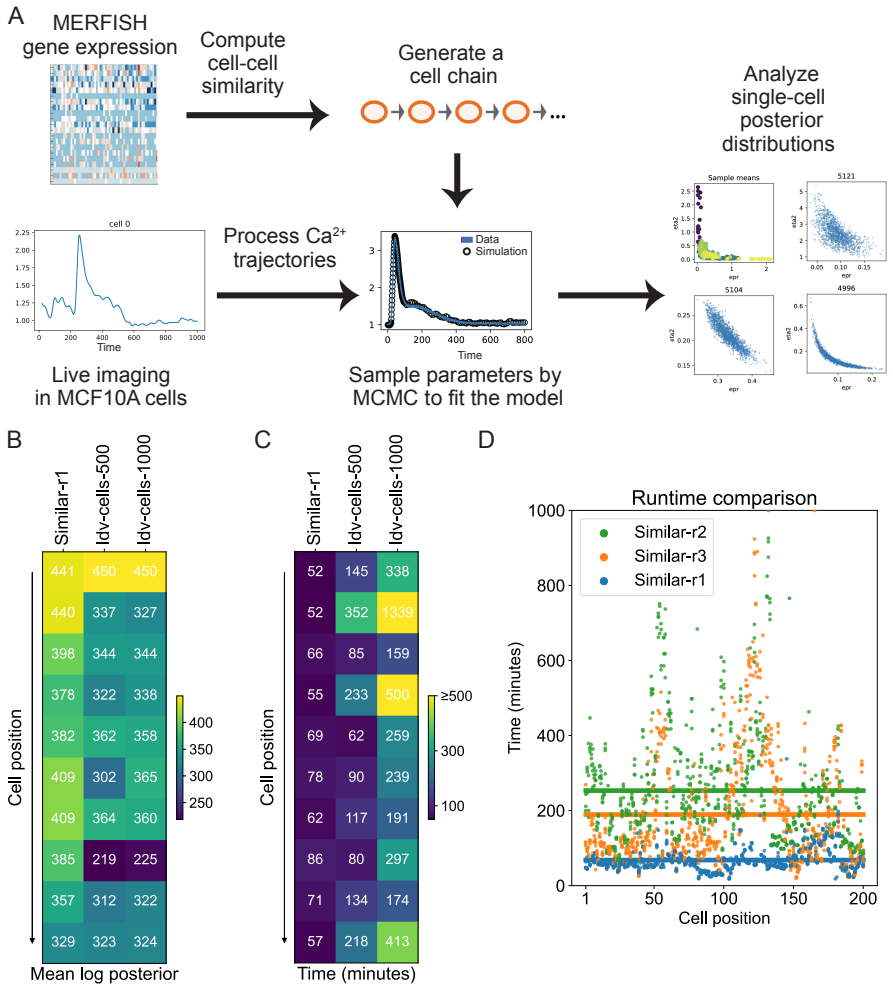

Figure 1: **Cell predecessor-informed priors boost performance for single-cell parameter inference. A**: Workflow for single-cell parameter inference along a cell chain. **B-C**: Comparison of cell chain vs. individually fit cells improves the log posterior likelihood (B) and the sampling efficiency (C). Each column represents a cell chain, and each row represents a single fitted cell. **D**: Optimizing HMC parameters, i.e. using fewer warmup iterations and lower maximum tree depths results in much lower computational runtimes. Maximum tree depth: $r1<r2=r3$; warmup iterations: $r1=r2<r3$.

chains with priors that were informed by cell-cell similarity, such as the run *Similar-r1*, with runs using priors fit from cell predecessors that were random. For the latter, we generated two cell chains, *Random-1* and *Random-2*, each consisting of a cell ordering assigned randomly. With regards to the computational efficiency (sampling times) and accuracy of fits (model posterior probabilities), there were no significant differences between the two random chains and the similarity-informed chain (Table S3). Therefore, although the use of priors informed by cell predecessors accelerated inference relative to individually fit cells, the choice of cell predecessors (similarity-based vs. randomly assigned) did not affect computational efficiency or the accuracy of fits.

We also studied the effect of algorithmic parameters of the NUTS sampler on single-cell parameter inference. We found that in practice, even when the maximum tree depth was set to 15 (as for runs *Similar-r2* and *Similar-r3*), the actual tree depth used during sampling had a mean value close to 10 (Table S4). The accuracy of fits, as quantified by the mean log posterior values, the $\widehat{R}$ statistic, and the distances between sampled trajectories and the data were comparable (Figure 1C, Table S4). Thus, we chose to set the maximum tree depth to 10 in practice, leading to much faster overall sampling times (Figure 1E, Table S4).

3.2 Analysis of model posteriors reveals divergent sources of intracellular and intercellular parameter variability

Analysis of the posterior distributions generated by fitting hundreds of single cells sampled from cell chains revealed regions of non-identifiability in the parameter space. For inference of the parameters of cells using the full 18-dimensional parameter space, we compared two cell chains where the final 100 cells in the chain were ordered identically, but the initial cells were different between the chains (one has 6 initial cells, the other has 3 different ones). In comparing the posterior distributions of the final 100 cells from each chain, we see that although some marginal posterior parameters are similar for all cells (e.g. for $K_{\text{off, ATP}}$, Figure 2A), they often diverge (e.g. for d_5, Figure 2B). However, relative changes in parameters along the chain seemed to be tightly correlated. To quantify this we computed the fold changes in mean marginal posterior parameter values between consecutive cells along the chain (Figure 2A–B). We see that for the majority of consecutive cell pairs, these are tightly correlated both in direction and magnitude. Similar results were observed for random cell chains run in parallel with different initial cells (Figure S2). It is perhaps not surprising that there are identifiability challenges in an 18-dimensional parameter space when fitting to data for only one species, no matter how rich its dynamics.

Related to identifiability, we observed interesting behavior for the marginal posterior distributions of two parameters along cell chains. The marginal distributions of these parameters, B_e and η_1, drifted through parameter space, i.e. increased slowly but steadily in their mean values along the cell chain (Figure S3A–B). Given the apparent insensitivity of the Ca^{2+} dynamics to these parameters, we performed model reduction and ran cell chains on reduced models with either one or both of these parameters set to a constant. Comparing chains each of 500 cells, we saw that the reduced models had similar performance both in terms of sampling efficiency and convergence (Figure S3C–E, Table S5). Posterior predictive checks showed no significant differences in simulated Ca^{2+} trajectories using reduced models. Thus for further analysis of the parameters underlying single-cell Ca^{2+} dynamics, we used the reduced model (*Reduced-3*).

We next studied joint parameter posterior distributions in the *Reduced-3* cell chain, and discovered striking differences between intracellular and intercellular variability. Several parameter pairs are observed to be highly correlated, as is expected given their biological roles and impact on the Ca^{2+} pathway, e.g. as activators or inhibitors of the same species. However, comparison of parameter correlations within and between cells yielded surprising results. Some parameter pairs showed consistent correlations (to a varying degree) both between cells along the chain and within single cells. For example, Ca^{2+} pump permeability (η_3) and concentration of free Ca^{2+} (c_0) were positively correlated in posterior values (Figure 2C). Similarly, the ER-to-cytosolic volume (ϵ) and the ER permeability (η_2) were negatively correlated, both between cells (posterior means) and within cells (Figure 2D). Other parameter pairs displayed strikingly different cell-extrinsic vs. cell intrinsic effects. The parameters for the ATP decay rate (K_{ATP}) and the PLC degradation rate ($K_{\text{off, ATP}}$) were positively correlated in their posterior means along the cell chain but negatively correlated in their marginal distributions with a single cell, for many cells (Figure 2E). A possible interpretation of the observed differences in intracellular and intercellular parameter correlations is the different scales. The intracellular parameter range was always smaller then the intercellular one. These differences indicate that on a large scale parameters are positively correlated but these correlations could still be locally negative. These observations indicate that the underlying landscape for these parameters is complex and emerges from competing sources of biological variation within this simply represented signaling pathway at the intra- and inter-cellular levels. We note that the distribution of cells by MAP value is well-mixed, i.e. there is no evidence of significant biases in posterior parameters arising due to the construction of the cell chain, thus the variation captured in these posterior distributions represents true biological differences in the cell population.

3.3 Sensitivity analysis quantifies heterogeneity of Ca^{2+} responses in single cells

We conducted sensitivity analysis for model parameters using sampled posteriors. By convention, one would define sensitivity of a parameter as the derivative of state variables with respect to that parameter [28, 29]. Since our inferred parameters were output as a distribution rather than scalar values, we decided to set each parameter to some extreme value according to its marginal posterior distribution (0.01-quantile or 0.99-quantile) in select draws from posterior, and trajectories were simulated from those altered draws (Figure 3A).

The distances between simulated trajectories and measured trajectories were used to indicate how sensitive Ca^{2+} response is with respect to a tested parameter. For each parameter at an extreme value, we calculated the mean of those distances for each cell, which will be referred as sensitivity. From the distribution of sensitivities, it was obvious that Ca^{2+} response was very sensitive to some model parameters, while it was only moderately sensitive to some others (Figure 3B). The distribution of sensitivities (i.e. mean trajectory distances) of the least sensitive parameters were peaked around 1.0 (Figure 3B), which was not much higher than mean trajectory distances from actually sampled posterior (Table S4). For those parameters, even the 75th percentiles of sensitivities were quite low (Figure 3B). Indeed, for some cells, the simulated trajectories closely resembled the measured trajectories (Figure 3C). As for the more sensitive parameters, their sensitivities were high at both the 0.01- and 0.99-quantiles (Figure 3B). The sensitive parameters were also more likely to increase and shrink rapidly

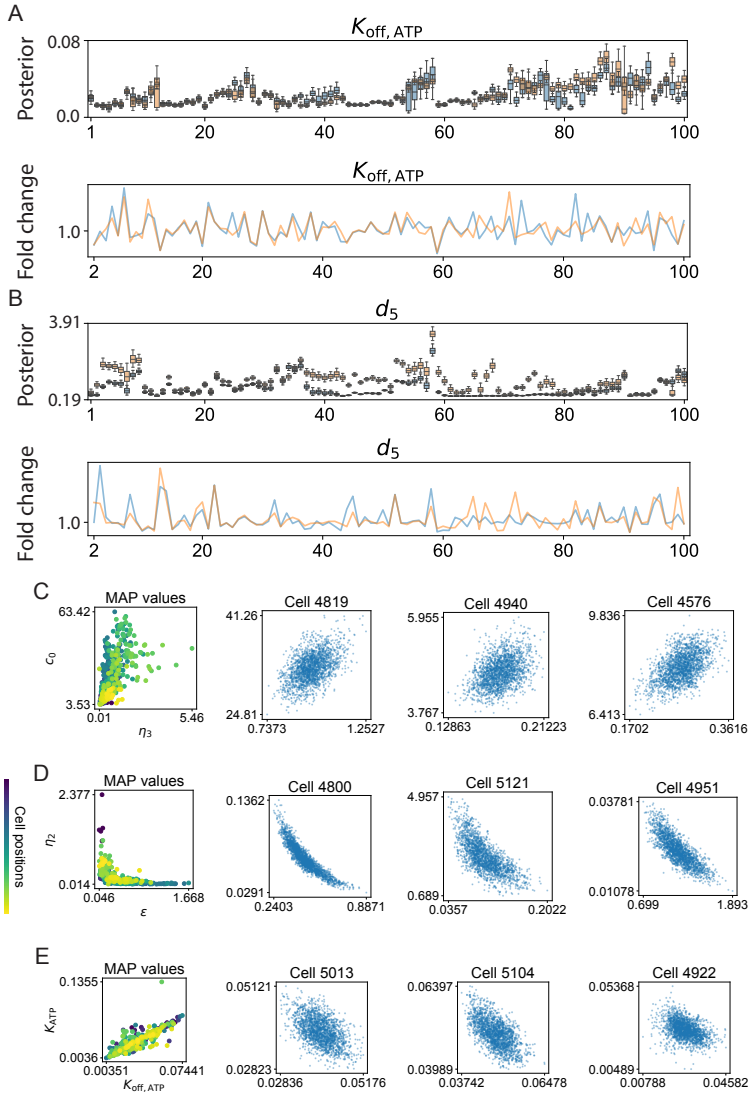

Figure 2: **Marginal posterior distributions along cell chains reveal parameter dependencies.** **A**: Marginal parameter posterior distributions in two parallel runs of an identical cell chain (upper), and the corresponding fold changes between consecutive cells (lower) are shown for $K_{\text{off, ATP}}$, the PLC degradation rate. **B**: As for (A) for the parameter d_5, the IP3 channel dissociation constant. **C**: Scatter plot of the maximum a posteriori (MAP) values for η_3 and c_0 for 500 cell posterior distributions from *Reduced-3* (left). Color indicates position along the cell chain. Scatter plots for the same parameter pain in three single cells are also shown: 500 samples from the cell's posterior distribution are shown in blue. **D**: As for (C) with parameters ϵ and η_2. **E**: As for (C) with parameters $K_{\text{off, ATP}}$ and K_{ATP}.

if their prior standard deviation had not been properly scaled and clipped before sampling (Figure S1G, I).

We noticed that Ca^{2+} response was sensitive to parameters in different fashions. For example, when d_1 was perturbed, the peak of Ca^{2+} response would change but the steady state would not (Figure 3D). However, perturbation of η_2 had much less impact on peak values than d_1 did, but it affected steady state considerably (Figure 3E). For the same parameter, going from one extreme value to another could cause opposite changes. Setting d_1 to the 0.01-quantile resulted in higher peaks as well as slower decrease after peaks (Figure 3D). On the other hand, setting d_1 to the 0.99-quantile caused lower peaks (Figure 3D). It is also worth noting that d_1 had the highest sensitivity at the 0.01-quantile, but it was only moderately sensitive at the 0.99-quantile (Figure 3B). η_2 showed opposite effects on steady state at the two extremes. The simulated trajectories had lower steady states at the 0.01-quantile but higher at the 0.99-quantile (Figure 3E).

3.4 Variability in Ca^{2+} dynamics is correlated with variability in gene expression

We extracted variable genes from the full gene expression data using principal component analysis (see details in Methods), and then sorted pairs of variable genes and and model parameters by the absolute values of Pearson correlations between gene expression and means of parameter posteriors in descending order. Table S6 shows the top 20 gene-parameter pairs for *Reduced-3*. Some genes like PPP1CC were highly correlated with more than one parameter. Meanwhile, some parameters, such as η_3, were also highly correlated to multiple genes. Since Pearson correlation is sensitive to outliers, we also ran linear regression with Huber loss, which is robust against outliers. The regression confirmed that those pairs of genes and parameters were definitely highly correlated (Figure 4A–D). For random cell chains, we also observed correlations between genes and parameters, but the correlations were not as strong as in similarity-induced chains (Table S7, Figure 4E–H). We compared top genes from correlation analysis for four cell chains: *Reduced-3*, *Similar-r1*, *Random-1*, *Random-2*. Gene-parameter pairs were sorted by absolute Pearson correlation in descending order and ranked genes by their appearance among sorted pairs. In total we identified 75 correlated gene-parameter pairs for the *Reduced-3* chain, applying a Bonferroni multiple testing correction (Figure S4). The top 30 genes ranked by highest gene-parameter correlation were chosen for further comparison. We found that 25 genes ranked in the top 30 in at least three cell chains (Figure 4I). Out of those 25 genes, 20 were also marker genes from Leiden clustering on all cells of the dataset (Figure 4I) [25, 26]. This high degree of overlap demonstrates the importance of these genes in explaining the variability in phenotypes, and hints at their information content pertaining to the dynamic Ca^{2+} cell phenotypes encapsulated by posterior parameter distributions.

Thus, across populations of cells, variable genes and variable marginal posterior parameter distributions can be highly correlated, but what is observed at the global cellular level? That is, what is the relationship between the full posterior parameter distribution of a cell and its global transcriptional state? To study this question, we analyzed the posterior distributions of cells using principal component analysis (PCA) for dimensionality reduction (while the size of the 16-dimensional posterior is small relative to the approx. 20,000 dimensions of a cell's transcriptional state, it is still unwieldy for analysis or visualization; the curse of dimensionality

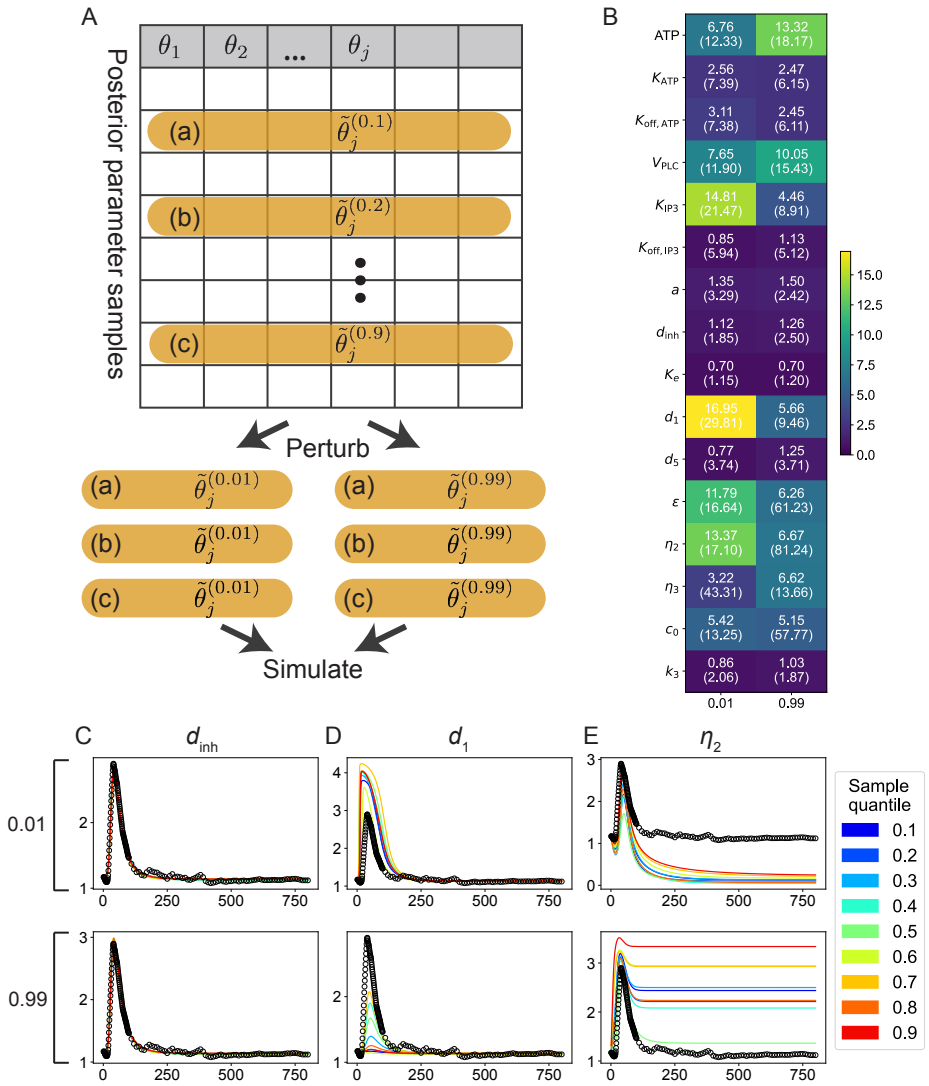

Figure 3: **Different aspects of the Ca^{2+} response are sensitive to specific parameters. A**: We determined sensitivity of Ca^{2+} response with respect to a parameter by perturbing the parameter along its marginal distribution. **B**: Parameters in the *Reduced-3* model had varying levels of sensitivity. **C**: Ca^{2+} response was not sensitive to perturbation in d_{inh}. **D**: Peak Ca^{2+} response changed dramatically when d_1 was perturbed. **E**: Steady state of Ca^{2+} response changed significantly when η_2 was perturbed.

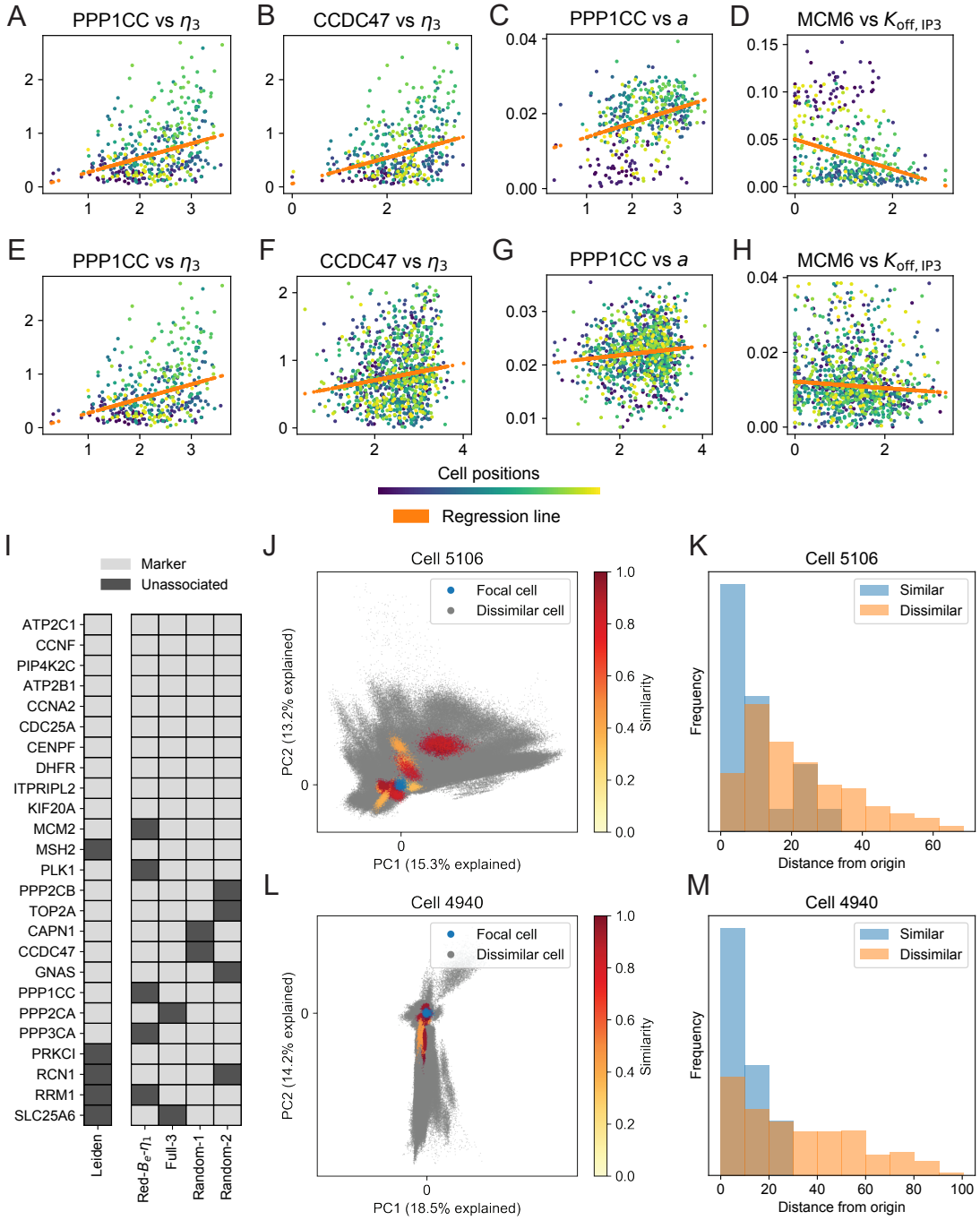

Figure 4: **Gene-parameter correlation analysis. A-D**: Highly correlated pairs of genes and parameters from *Reduced-3*. **E-H**: Highly correlated pairs of genes and parameters from *Random-2*. **I**: Consensus genes from gene-parameter correlation analysis and Leiden clustering. **J**: Projection of all cells onto first two PCs of cell 5106. **K**: Mean distances of projected samples from origin in (J). **L**: Projection of all cells onto first two PCs of cell 4940. **M**: Mean distances of projected samples from origin in (L).

strikes quickly). We chose a cell from the *Reduced-3* similarity-based cell chain and decomposed its posterior distribution using PCA, for which we visualize the first two principal components of this "focal cell" (Figure 4J, L and Figure S5 A, C). The posterior distributions of other cells from the same chain were then projected onto the principal components of the focal cell. In this way we can evaluate the positions of cell posterior distributions relative to the focal cell.

In Figure 4 and Figure S5 we color cells by their gene expression similarity. Comparison of similar and dissimilar cells from the same chain showed that cells that were similar based on their global gene expression states were closer overall to the focal cell than dissimilar cells. This result did not rely on the method by which posterior distributions were normalized (Figure 4K, M and Figure S5B, D). When we performed a similar analysis on cells from a random cell chain, similar cells were not located closer to the focal cell than non-similar cells (Figure S6). Notably, proximity of posterior distributions of similar cells was not driven by the location of a cell along the chain (i.e. block structure not observed in Figure S7), i.e. it is not due to a local cell-cell similarity effect, but rather reflects underlying associations between the global transcriptional state of a cell and its specific Ca^{2+} pathway dynamics. Hence, the use of similarity-based priors for single-cell parameter inference lead to a gain in information about the extent to which global transcriptional states and Ca^{2+} pathway parameters are associated.

3.5 Similarity-based posterior cell clustering reveals distinct transcriptional states underlying Ca^{2+} dynamics

To further characterize the extent to which single-cell posterior distributions can predict Ca^{2+} responses, we clustered 500 cells from the *Reduced-3* chain based on their inferred posterior distributions using hierarchical clustering (see Methods). Three clusters were obtained (Figure 5A) with distinct Ca^{2+} dynamics: "low-responders" exhibited lower overall Ca^{2+} peaks in response to ATP (Figure 5B); "early-responders" exhibited earlier overall Ca^{2+} peaks in response to ATP; and "late-high-responders" exhibited robust Ca^{2+} responses with peaks that were later and higher than cells from other clusters (Figure S8).

These dynamic profiles can be explained by the distinct parameter sets that give rise to each: low-responders are characterized by high concentration of free Ca^{2+} in ER (c_0) but low activation of IP3 receptor (Figure S8, Figure S9, Figure S10). Early-responders are characterized by parameters leading to faster and earlier IP3 and PLC dynamics (Figure S9, Figure S10); and late-high-responders are characterized by low d_1 (Figure S9). We can directly compare our posterior parameter clustering with that obtained by Yao et al. [17] using similar methods. Analysis of the parameters d_{inh}, d_1 and d_5 , which control the rates of Ca^{2+} release as regulated by the IP3R channel shows that d_{inh} is consistent across clustering results. In both cases d_{inh} is larger in cells that respond more strongly to ATP stimulation (Figure S10) [17]). In Yao et al. [17], both d_1 and d_5 were smaller in cells with stronger Ca^{2+} responses. We found that d_1 was smaller in the late-high-responders, but not in the early responders; and that d_5 was higher for the early-responders, in contrast with Yao et al. (Figure S10). We note that we set a stringent threshold for minimum peak Ca^{2+} response, i.e. we excluded non-responding cells, unlike Yao et al., thus in a direct comparison most of the cells in our population would belong to the "strong positive" cluster in Yao et al. [17].

To compare these Ca^{2+} profiles and the parameters that give rise to them, we performed two additional analyses. We clustered the same 500 cells based on their single-cell gene expression

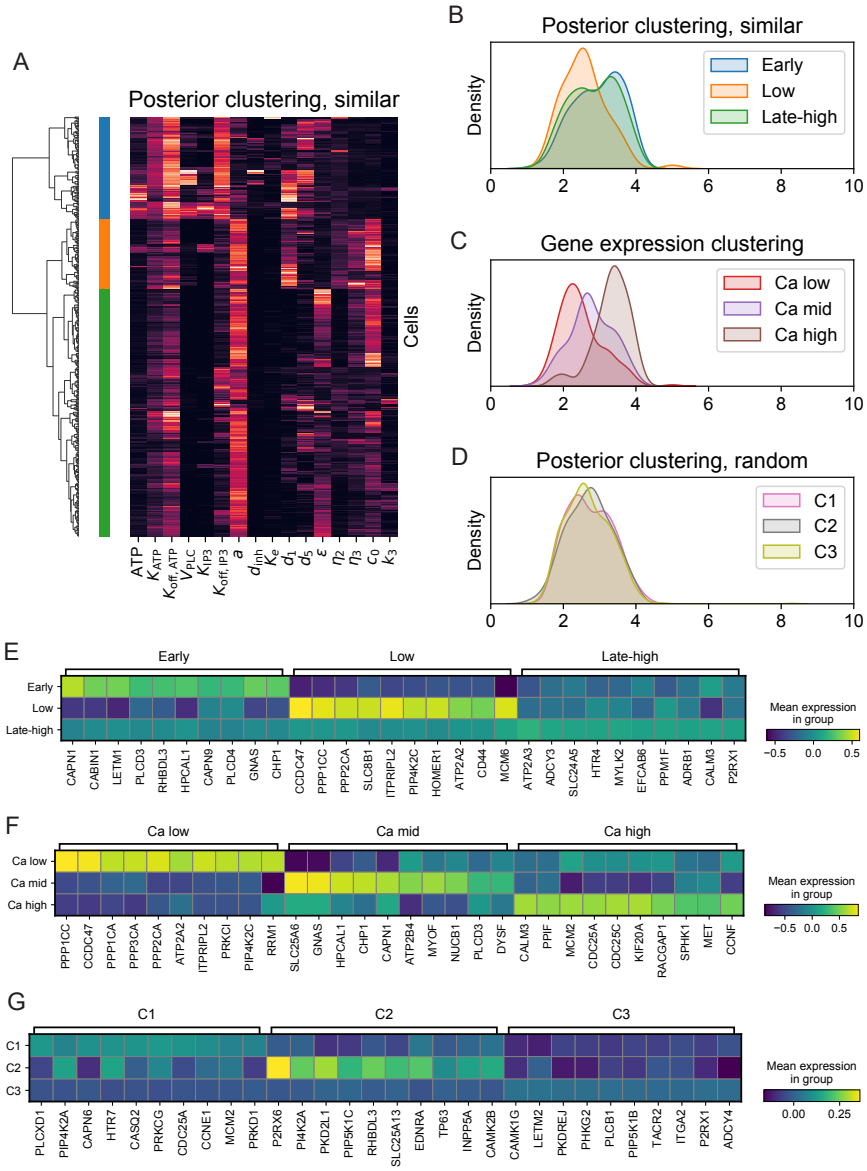

Figure 5: **Clustering of cell posterior distributions reveals marker genes for Ca^{2+} states.** **A**: Posterior means agglomeratively clustered using Ward linkage. **B**: Kernel density estimate of Ca^{2+} peaks from posterior clustering on *Reduced-3*. **C**: Kernel density estimate of Ca^{2+} peaks from gene expression clustering on *Reduced-3*. **D**: Kernel density estimate of Ca^{2+} peaks from posterior clustering on *Random-2*. **E**: Marker genes from posterior clustering on *Reduced-3*. **F**: Marker genes from gene expression clustering on *Reduced-3*. **G**: Marker genes from posterior clustering on *Random-2*.

using community detection (via the Leiden algorithm in Scanpy [25, 26]); we also clustered 500 cells from a random (i.e. non-similarity-based) cell chain using the same hierarchical posterior clustering methods as described above. For the cell clustering based on gene expression, as for the similarity-based cell clustering, distinct Ca^{2+} profiles could be observed, consisting in this case of "Ca-low", "Ca-mid", and "Ca-high" responses (Figure 5C). In contrast, no distinct Ca^{2+} dynamics could be observed for the posterior clustering based on the random cell chain (Figure 5D).

In order to study relationships between Ca^{2+} dynamic profiles and gene expression, we performed differential gene expression testing to obtain the top marker genes for each set of clusters obtained above (Figure 5E). Distinct markers corresponding to each cluster could be obtained for the similarity-based posterior clustering and the gene expression based clustering, but not for the random cell chain posterior clustering. Thus confirming that clustering cell posteriors derived from the random cell chain did not lead to distinguishable patterns in either Ca^{2+} dynamics nor gene expression space. On the other hand, clustering the cell posteriors from a similarity-induced chain led to distinct gene expression profiles per cluster, similar to the marker gene expression profiles obtained by clustering directly on the gene expression. Parameter inference of single-cell Ca^{2+} dynamics from a similarity-based chain enables the identification of sets of cells with different transcriptional profiles that also respond differently to ATP stimulation.

Analysis of the genes that are associated with specific Ca^{2+} profiles showed that the low-responder cells were characterized by up-regulation of CCDC47 and PP1 family genes PPP1CC and PPP2CA. Early-responder cells were characterized by up-regulation of CAPN1 and CHP1, among other genes. The late-high responder cells were characterized by up-regulation of CALM3, among others, although the marker gene signature for this cluster was not as clear relative to other clusters. Notably, there is considerable overlap in marker genes identified by the cell posterior clustering or directly by gene expression clustering: the early-responder cluster signature overlaps with the Ca-mid gene expression cluster, and the low-responder cluster overlaps with the Ca-low cluster. Thus Ca^{2+} model parameters contain significant information about gene expression. By clustering the posterior distributions of cells fit to similarity-based chains, but without any direct appeal to the gene expression, we are able to obtain many of the same marker genes that characterize the variability in Ca^{2+} responses between single cells. This demonstrates that the posterior distributions of cells fit from a similarity-based chain contain a significant amount of the same information regarding the distinct transcriptional states of single cells as can be gleaned by clustering directly on the gene expression.

4 Discussion

We have presented methods for inferring the parameters of a Ca^{2+} signaling pathway model, given data describing the dynamic Ca^{2+} responses in single cells coupled with the gene expression observed immediately following the Ca^{2+} response. We hypothesized that using the posterior distributions of previous cells as prior distributions for subsequent cells along a "cell chain" could lead to more efficient and more accurate inference of the model parameters. To the best of our knowledge, we present the first parameter inference framework for dynamic models that incorporates single-cell gene expression information into the inference framework. We

implemented this schema using the No-U-Turn Sampler [21], an efficient Hamiltonian Monte Carlo algorithm for MCMC sampling. We discovered that the use of cell predecessors to construct priors did indeed lead to much faster sampling of parameters. However, these improvements in computational efficiency did not rely on the use of single-cell gene expression to construct the priors: the performance of randomly sampled cell predecessors was equivalent. In the case that the cell chain was constructed using single-cell gene expression and transcriptional similarity, the resulting posterior parameter distributions contained more information about Ca^{2+} signaling dynamics. Through clustering of the posterior distributions, we were able to identify important relationships between gene expression and dynamic cell phenotypes, thus providing means to map from states to fates.

The model we chose to test parameter inference methods is a classical ODE model of Ca^{2+} signaling adapted from [19, 16] consists of 12 variables and (originally) a 40-dimensional parameter space. This was reduced to 19 parameters in Yao et al. [17] and 16 parameters in our work. Analysis of even a single 16-dimensional posterior distribution requires dimensionality reduction techniques, let alone the analysis of the posterior distributions obtained for populations of hundreds of single cells. By studying pairwise parameter posterior distributions, we discovered striking differences between intracellular and intercellular variability. We performed parameter sensitivity analysis by developing methods to perturb specific dimensions of the posterior parameter distribution. This was very informative: it allowed us to pinpoint the effects of specific parameter perturbations on the Ca^{2+} dynamic response. Indeed, we advocate for the use of sensitivity analysis more generally as means to distinguish and pinpoint the effects of different parameter combinations for models of complex biochemical signaling pathways.

By clustering posterior distributions, distinct patterns of Ca^{2+} dynamic responses to ATP were obtained, and could be mapped directly to variation in gene expression. These distinct patterns consisted of “early”, “low”, and “late-high” responders. In previous work using similar approaches for clustering [17], posterior parameter clusters predominantly revealed response patterns consisting of responders and non-responders. We excluded those cells that did not exhibit a robust response to ATP from inference, thus here we obtain subtler Ca^{2+} response dynamics and are able to predict the transcriptional states that drive them.

We were able to find sufficiently good fits to all single cells tested, but this came at the expense of model identifiability. With four variables and a 16-dimensional parameter space, the dimension of the model far exceeds that of the data: time series of Ca^{2+} responses in single cells. I.e. we have no data with which to constrain the three additional model species. As a result, we applied various techniques to minimize the consequences of this mismatch in dimensionality between model and data. We used the approach of “scaling and clipping” in prior construction: this constrained dramatic changes in the posterior variance, however it remains an ad hoc solution. More effective techniques might improve inference and could become necessary in the case of larger models than we consider here. These include (in order of sophistication): tailoring the scaling/clipping choices to be parameter-specific; tailoring the choice of prior variance based on additional sources of data; or performing model reduction/identifiability analysis to further constrain the prior space before inference. Constructing priors from cells with similar gene expression also helped to curb the curse of dimensionality: sampling cells sequentially places a constraint on the model. Nonetheless, in the future more directed approaches to tackle model identifiability ought to be considered.

Mapping dynamic cell phenotypes to transcriptional states remains a central challenge

in systems biology. The limitations of deriving such knowledge from gene expression data alone [30] have led to the proposal of new methods that seek to bridge the gap between states and fates [31]. Here, making use of the technology that permits measurement of Ca^{2+} dynamics and gene expression in the same single cells, we have demonstrated that single-cell parameter inference informed by transcriptional similarity enables us to begin to draw state-to-fate maps, whereby dynamic properties of Ca^{2+} signaling can be inferred from specific gene expression states. More broadly, we expect the statistical framework presented here that uses single-cell gene expression to inform priors for Bayesian inference to be applicable across many domains. As a result, future models can more readily incorporate global or targeted transcriptional information to learn molecular and cellular dynamics.

Acknowledgments

This work was supported by an Andrew J. Viterbi Fellowship in Computational Biology and Bioinformatics (to X.W.), A.L.M. acknowledges support from the National Institutes of Health (R35GM143019) and the National Science Foundation (DMS2045327).

Data Availability

Parameter inference was developed in Python 3.6 and Stan 2.19. Posterior analyses were developed in Python 3.8. All code developed to simulate models and run parameter inference is released under an MIT license at: https://github.com/maclean-lab/singlecell-parinf

References

[1] Darren James Wilkinson. *Stochastic modelling for systems biology*. Chapman & Hall/CRC mathematical and computational biology. CRC Press, Taylor & Francis Group, Boca Raton London New York, third edition, first issued in paperback edition, 2020. ISBN 978-0-367-65693-5.

[2] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI–a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, December 2006. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btl485. URL `https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btl485`.

[3] Juliane Liepe, Sarah Filippi, Michał Komorowski, and Michael P. H. Stumpf. Maximizing the Information Content of Experiments in Systems Biology. *PLoS Computational Biology*, 9(1):e1002888, January 2013. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002888. URL `https://dx.plos.org/10.1371/journal.pcbi.1002888`.

[4] David J. Warne, Ruth E. Baker, and Matthew J. Simpson. Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts

to state-of-the-art. *Journal of The Royal Society Interface*, 16(151):20180943, February 2019. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2018.0943. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2018.0943.

[5] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P.H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31):187–202, February 2009. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2008.0172. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2008.0172.

[6] Allon Wagner, Aviv Regev, and Nir Yosef. Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnology*, 34(11):1145–1160, November 2016. ISSN 1087-0156, 1546-1696. doi: 10.1038/nbt.3711. URL http://www.nature.com/articles/nbt.3711.

[7] Robert Foreman and Roy Wollman. Mammalian gene expression variability is explained by underlying cell state. *Molecular Systems Biology*, 16(2), February 2020. ISSN 1744-4292, 1744-4292. doi: 10.15252/msb.20199146. URL https://onlinelibrary.wiley.com/doi/abs/10.15252/msb.20199146.

[8] Pavel A. Brodskiy and Jeremiah J. Zartman. Calcium as a signal integrator in developing epithelial tissues. *Physical Biology*, 15(5):051001, May 2018. ISSN 1478-3975. doi: 10.1088/1478-3975/aabb18. URL https://doi.org/10.1088/1478-3975/aabb18.

[9] G. Dupont, L. Combettes, G. S. Bird, and J. W. Putney. Calcium Oscillations. *Cold Spring Harbor Perspectives in Biology*, 3(3):a004226–a004226, March 2011. ISSN 1943-0264. doi: 10.1101/cshperspect.a004226. URL http://cshperspectives.cshlp.org/lookup/doi/10.1101/cshperspect.a004226.

[10] Muthu Periasamy and Anuradha Kalyanasundaram. SERCA pump isoforms: Their role in calcium transport and disease. *Muscle & Nerve*, 35(4):430–442, April 2007. ISSN 0148639X, 10974598. doi: 10.1002/mus.20745. URL https://onlinelibrary.wiley.com/doi/10.1002/mus.20745.

[11] Adam L. MacLean, Tian Hong, and Qing Nie. Exploring intermediate cell states through the lens of single cells. *Current Opinion in Systems Biology*, 9:32–41, June 2018. ISSN 24523100. doi: 10.1016/j.coisb.2018.02.009. URL https://linkinghub.elsevier.com/retrieve/pii/S2452310017302238.

[12] J.R. Moffitt and X. Zhuang. RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). In *Methods in Enzymology*, volume 572, pages 1–49. Elsevier, 2016. ISBN 978-0-12-802292-4. doi: 10.1016/bs.mie.2016.03.020. URL https://linkinghub.elsevier.com/retrieve/pii/S0076687916001324.

[13] Chee-Huat Linus Eng, Michael Lawson, Qian Zhu, Ruben Dries, Noushin Koulena, Yodai Takei, Jina Yun, Christopher Cronin, Christoph Karp, Guo-Cheng Yuan, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, 568(7751):235–239, April 2019. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-019-1049-y. URL http://www.nature.com/articles/s41586-019-1049-y.

[14] Steven H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview Press, a member of the Perseus Books Group, Boulder, CO, second edition edition, 2015. ISBN 978-0-8133-4910-7.

[15] Amy Y. Chang and Wallace F. Marshall. Dynamics of living cells in a cytomorphological state space. *Proceedings of the National Academy of Sciences*, 116(43):21556–21562, October 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1902849116. URL https://pnas.org/doi/full/10.1073/pnas.1902849116.

[16] G. Lemon, W.G. Gibson, and M.R. Bennett. Metabotropic receptor activation, desensitization and sequestration—I: modelling calcium and inositol 1,4,5-trisphosphate dynamics following receptor activation. *Journal of Theoretical Biology*, 223(1):93–111, July 2003. ISSN 00225193. doi: 10.1016/S0022-5193(03)00079-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0022519303000791.

[17] Jason Yao, Anna Pilko, and Roy Wollman. Distinct cellular states determine calcium signaling response. *Molecular Systems Biology*, 12(12):894, December 2016. ISSN 1744-4292, 1744-4292. doi: 10.15252/msb.20167137. URL https://onlinelibrary.wiley.com/doi/abs/10.15252/msb.20167137.

[18] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. *Stan* : A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 2017. ISSN 1548-7660. doi: 10.18637/jss.v076.i01. URL http://www.jstatsoft.org/v76/i01/.

[19] Yue-Xian Li and John Rinzel. Equations for InsP3 Receptor-mediated [Ca2+]i Oscillations Derived from a Detailed Kinetic Model: A Hodgkin-Huxley Like Formalism. *Journal of Theoretical Biology*, 166(4):461–473, February 1994. ISSN 00225193. doi: 10.1006/jtbi.1994.1041. URL https://linkinghub.elsevier.com/retrieve/pii/S0022519384710411.

[20] Shuxiong Wang, Matthew Karikomi, Adam L MacLean, and Qing Nie. Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Research*, 47(11):e66–e66, 2019. ISSN 0305-1048. doi: 10.1093/nar/gkz204. URL https://doi.org/10.1093/nar/gkz204.

[21] Matthew D. Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014. URL http://jmlr.org/papers/v15/hoffman14a.html.

[22] Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434 [stat]*, July 2018. URL http://arxiv.org/abs/1701.02434.

[23] Ravin Kumar, Colin Carroll, Ari Hartikainen, and Osvaldo Martin. ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, 4(33):1143, January 2019. ISSN 2475-9066. doi: 10.21105/joss.01143. URL http://joss.theoj.org/papers/10.21105/joss.01143.

[24] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al.

Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85): 2825–2830, 2011. URL http://jmlr.org/papers/v12/pedregosa11a.html.

[25] F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):15, December 2018. ISSN 1474-760X. doi: 10.1186/s13059-017-1382-0. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1382-0.

[26] V. A. Traag, L. Waltman, and N. J. van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233, December 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-41695-z. URL http://www.nature.com/articles/s41598-019-41695-z.

[27] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-019-0686-2. URL http://www.nature.com/articles/s41592-019-0686-2.

[28] Arvind Varma, Massimo Morbidelli, and Hua Wu. *Parametric Sensitivity in Chemical Systems*. Cambridge University Press, 1 edition, March 1999. ISBN 978-0-521-62171-7 978-0-521-01984-2 978-0-511-72177-9. doi: 10.1017/CBO9780511721779. URL https://www.cambridge.org/core/product/identifier/9780511721779/type/book.

[29] Micha\l Komorowski, Maria J. Costa, David A. Rand, and Michael P. H. Stumpf. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proceedings of the National Academy of Sciences*, 108(21):8645–8650, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1015814108. URL https://www.pnas.org/content/108/21/8645.

[30] 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, March 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1714723115. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1714723115.

[31] Xiaojie Qiu, Yan Zhang, Jorge D. Martin-Rufino, Chen Weng, Shayan Hosseinzadeh, Dian Yang, Angela N. Pogson, Marco Y. Hein, Kyung Hoi (Joseph) Min, Li Wang, et al. Mapping transcriptomic vector fields of single cells. *Cell*, page S0092867421015774, February 2022. ISSN 00928674. doi: 10.1016/j.cell.2021.12.045. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867421015774.