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

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

Single-cell genomics datasets offer vast new resources with which to study cells, but their potential to inform parameter inference of cell dynamics has yet to be realized. Here we develop methods for Bayesian parameter inference with data that jointly measure gene expression and fluorescent reporter dynamics in single cells. By ordering cells by their transcriptional similarity, the posterior distribution of one is used to inform the prior distribution of its neighbor via transfer learning along a cell chain. In application to Ca^{2+} signaling dynamics, we fit the parameters of thousands of models of variable single-cell responses. We show that transfer learning accelerates inference, although constructing cell chains by gene expression does not improve over randomly ordered cells. Clustering cell posterior distributions reveals that only using similarity-based chains – and not randomly sampled chains – can we distinguish Ca^{2+} dynamic profiles and associated marker genes. Both at the global and the individual-gene level, transcriptional states predict features of the Ca^{2+} response. We discover complex and competing sources of cell heterogeneity: parameter covariation can diverge between the intracellular and intercellular contexts. Single-cell parameter inference thus offers broad means to quantify relationships between transcriptional states and the dynamic responses of cells.

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 distribution characterizes the most likely parameter values to give rise to the data as well as the uncertainty that we have regarding those parameters. Thus, parameter inference provides a map from the dynamic phenotypes that we observe in experiments to the parameters of a mathematical model.

Single-cell genomics technologies have revealed a wealth of information about the states of single cells that was not previously accessible [6]. This ought to assist with the characterization of dynamic phenotypes. However, it is much less clear how to draw maps between dynamic phenotypes of the cell and single-cell states as quantified via genomic measurements. The challenge in part lies in the combinatorial complexity: even if a small fraction of genes contain information regarding the phenotype of interest, say a few hundred, this remains more than enough to characterize any feasible number of states of an arbitrarily complex dynamical process.

This leads us to the central question of this work: can the integration of single-cell gene expression data into a framework for modeling and inference improve our understanding of the cellular phenotypes of interest? We address this question below by developing methods for single-cell data-informed parameter inference using data that jointly measure dynamics and gene expression in the same single cells [7]. We apply this framework to study Ca^{2+} signaling dynamics and signal transduction in response to adenosine triphosphate (ATP) in human breast MCF10A cells.

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 molecular secretion, as well as collective behaviors from organogenesis to wound healing [8]. In response to 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 almost certainly faster than gene transcription, enabling the study of links between the dynamic Ca^{2+} responses to ATP and the transcriptional state of the cell.

Our ability to measure gene expression in thousands of single cells per experiment has not only led to new discoveries but has also fundamentally changed how we identify and characterize cell states [11]. Technologies used to quantify gene expression in single cells include sequencing and fluorescent imaging. The latter permits the measurement 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 by protocols such as MERFISH [12] and seqFISH [13]. Moreover, by coupling multiplexed smFISH with fluorescent

imaging of Ca^{2+} dynamics using a GFP reporter in MCF10A cells, we are able to jointly capture the dynamic cell responses and the single-cell gene expression in the same single cells [7]. These data offer new potential to study the relationships between transcriptional states of cells and the dynamic phenotypes these may produce.

Models of gene regulatory networks and cellular signaling pathways described by ordinary differential equations (ODEs) capture the interactions between gene transcripts, proteins, or other molecular species and their impact on cellular dynamics. Well-established dynamical systems theory offers a wide range of tools with which to analyze the transient and equilibrium behavior of ODE models [14]. Although it remains to some extent under investigation whether or not such equilibrium approximations of behavior are appropriate for cells [15]. Here, we model Ca^{2+} dynamics via ODEs based on previous work [16, 17] to study the relationships between single-cell transcriptional states and dynamic cell responses. We develop a parameter inference scheme to fit the single-cell Ca^{2+} dynamics using information from cell predecessors as captured through the construction of 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 implemented for parameter inference using Hamiltonian Monte Carlo in Stan [18]. We go on to study the results of inference: we discover that priors informed by cell predecessors accelerate parameter inference, but that cell chains with randomly sampled predecessors perform as well as those with transcriptional similarity-informed predecessors. Analysis of hundreds of fitted single cells reveals that cell-intrinsic vs. cell-extrinsic posterior parameter relationships can differ widely, indicative of fundamentally different sources of underlying variability. The analysis of posterior distributions offers an intuitive way to assess parameter sensitivities in response to Ca^{2+} dynamics. We show that variability in single-cell gene expression is associated with variability in posterior parameter distributions, both by analysis of individual gene-parameter pairs and globally, by dimensionality reduction of the posterior space. We cluster cells using their posterior parameters, and discover that for cell chains derived from transcriptional similarity there are pronounced relationships between single-cell gene expression states and the dynamic cell phenotypes. Beyond the insight gained into the dynamics of the Ca^{2+} pathway, the modeling and inference framework we present is broadly applicable to other contexts in which one seeks to quantify relationships between dynamic phenotypes and single-cell gene expression states.

2 Materials and Methods

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

We model Ca^{2+} signaling pathway responses in MCF10A human epithelial cells using nonlinear ordinary differential equations (ODEs), as previously developed [16, 17]. 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, to characterize dynamic responses in MCF10A cells. 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 responses following ATP binding to purinergic receptors: the activations of PLC, IP3, the IP_3R channel on the surface of the ER, and finally the release of Ca^{2+} from the ER into the cytoplasm [17]. Ca^{2+} may also enter the ER through the IP_3R channel and the SERCA pump [17]. Our model differs Yao et al. [17] in that we combine the product of two parameters in the previous model, $K_{\text{on, ATP}}$ and ATP, into a single parameter, ATP. This reduction of the model parameter space removed the redundancy that would otherwise exist in the distributions of $K_{\text{on, ATP}}$ and 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 consist of a joint assay measuring Ca^{2+} dynamics and gene expression via multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) [12]. Ca^{2+} dynamics in a total of 5128 human MCF10A cells are measured via imaging for 1000 seconds (ATP stimulation at 200 seconds) using a GCaMP5 biosensor. Immediately following this step, 336 genes are measured by MERFISH [7]. The Ca^{2+} trajectories are smoothed using a moving average filter with a twenty-second window size. After smoothing, data points occurring before ATP stimulation are removed. Data points for each Ca^{2+} trajectory after t=300 are downsampled by a factor of 10; the trajectories are at or close to steady state by this time. Single-cell gene expression data is collected using MERFISH after the Ca^{2+} imaging as previously described [7, 12].

2.3 Generating cell chains via cell-cell similarity

Cell-cell similarity is quantified via single-cell transcriptional states, i.e. by comparing x_m^i and x_m^j, the expression of m genes in cells i and j. We obtain a symmetric cell-cell similarity matrix, W, from the log-transformed MERFISH expression data via optimization in SoptSC [20]: entries $W_{i,j}$ denote the similarity between cells i and j. To create a chain of cells linked through their similarity in gene expression space, we:

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 $G = (V, E)$; each node is a cell and an edge is placed between two cells if they have a similarity score above a given threshold;
2. For a choice of initial (root) cell, traverse G and record the order of cells traversed.

Ideally, each cell would be visited exactly once, however this amounts to finding a Hamiltonian path in G, an NP-complete problem. Therefore, as a heuristic solution we use a depth-first search (DFS), which can be completed in linear time. From the current node, randomly select an unvisited neighbor node and set this as the next current node, recording it once visited (pre-order DFS). If the current node has no unvisited neighbors, 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. Given cases where the similarity matrix is sparse (as we have here), the DFS generates a tree that is very close to a straight path.

2.4 Bayesian parameter inference with posterior-informed prior distributions

We seek to infer dynamic model parameters in single cells, informed by cell-cell similarity via the position of a cell in a cell chain. We use the Markov chain Monte Carlo (MCMC) implementation: Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS) in

Stan [21, 18]. HMC improves upon the efficiency of other MCMC algorithms by treating the sampling process as a physical system and employing conservation laws [22]. From an initial distribution, the algorithm proceeds through intermediate phase of sampling (warmup) until (one hopes) convergence to the stationary distribution. During warmup, NUTS adjusts the HMC hyperparameters automatically [21].

The prior distribution over parameters is a multivariate normal distribution, with dimensions θ_j, $j = 1, \ldots, m$, where m is the number of parameters. Let f be a function characterizing the interactions between ODE state variables, 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 we truncate the prior so that each θ_i is bounded by 0 from below.

For the first cell in a chain, we use a relatively uninformative prior, the "Lemon" prior (Table 1), derived from parameter value estimates in previous work [16, 19, 17]. For the ith cell in a chain ($i > 1$), the prior distribution is constructed from the posterior distribution of the $(i-1)^{th}$ cell (Section 2.5). For each cell, NUTS is run for four independent chains with the same initialization. To simulate $\widehat{y}(t)$ during sampling, we use the implementation of fourth and fifth order Runge-Kutta in Stan [18].

Convergence of NUTS chains is evaluated using the $\widehat{R}$ statistic: the ratio of between-chain variance to within-chain variance [18, 23]. A typical heuristic used is $\widehat{R}$ between 0.9 and 1.1 indicates that for this set of chains the stationary distributions reached for a given parameter are well-mixed. There are two caveats on our use of $\widehat{R}$ in practice:

1. For our model, we observe that well-fit (i.e. not overfit) Ca^{2+} trajectories did not require $\widehat{R} \in (0.9, 1.1)$ for all parameters. Thus we assess $\widehat{R}$ only for the log posterior, using a more tolerant upper bound of 4.0.
2. There are cases where one chain diverges but 3/4 are well-mixed. In such cases, we choose to retain the three well-mixed chains as a sufficiently successful run. Thus if $\widehat{R}$ is above the threshold, before discarding the run, we compute $\widehat{R}$ for all three-wise combinations of chains, and retain the run if there exist three well-mixed chains.

2.5 Constructing and constraining prior distributions

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 variance of the prior for θ_j is derived from $\sigma_j^{(i-1)}$, the posterior variance of θ_j from the $(i-1)^{th}$ cell. To a) sufficiently explore the parameter space, and b) prevent instabilities (rapid growth or decline) in marginal parameter posterior values along the cell chain, we scale each $\sigma_j^{(i-1)}$ 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 variance for θ_j for the ith cell.

2.6 Dimensionality reduction and sensitivity analyses

To compare posterior samples from different cells, we use principal component analysis (PCA). Posterior samples are projected onto a subspace by first choosing a cell (the focal cell) and normalizing the posterior samples from other cells against the focal cell, either by min-max or z-score normalization. Min-max normalization transforms a vector x to $\frac{x-x_{\min}}{x_{\max}-x_{\min}}$, where $x_{\min}$ is the minimum and $x_{\max}$ the maximum of x. z-score normalization transforms x to $\frac{x-\mu_x}{\sigma_x}$, where μ_x is the mean and σ_x is the standard deviation of x. Normalizing to the focal cell amounts to setting $x_{\min}, x_{\max}, \mu_x, \sigma_x$ to be the values corresponding to the focal cell for all cells normalized. We perform PCA on the normalized focal cell posterior samples and project them into the subspace spanned by the first two principal components. The normalized samples from all other cells are projected onto the PC1-PC2 subspace of the focal cell.

We develop methods for within-posterior sensitivity analysis to assess how perturbations of model parameters within the bounds of the posterior distribution affect Ca^{2+} responses. Given $\tilde{\theta}$, the posterior distribution 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 $\tilde{\theta}_{\cdot,j}$. Nine "evenly spaced" samples are drawn from the posterior range of $\tilde{\theta}$ for 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. The Euclidean distances between simulated trajectories and data are used to quantify the sensitivity of each parameter perturbation.

2.7 Correlation analysis and cell clustering of MERFISH data

Correlations between single-cell gene expression values and posterior parameters from the Ca^{2+} pathway model are determined for variable genes. We calculate the z-scores of posterior means for each parameter of a cell sampled from a population, and remove that cell if any of its parameters has a posterior mean z-score smaller than -3.0 or greater than 3.0. PCA is 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" of gene i 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 calculate the Pearson correlation between its log-normalized expression value and the posterior means of individual model parameters. Gene-parameter pairs are ranked by their absolute Pearson correlations and the top 30 are selected for analysis. Gene-parameter pair relationships are quantified by linear regression using a Huber loss, which is more robust to outliers than mean squared error.

To cluster cells using their single-cell gene expression, raw count matrices are normalized, log-transformed, and scaled to zero mean and unit variance before clustering using the Leiden algorithm at 0.5 resolution [25], implemented in Scanpy 1.8 [26]. Marker genes for each cluster are determined by a t-test.

2.8 Clustering of cell posterior parameter distributions

Cells are clustered according to their posterior distributions. For each parameter, the posterior means for each cell are computed and scaled to $[0, 1]$. The distance between two cells is 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 is performed using SciPy 1.7 [27]. Marker genes for each cluster identified are determined using a t-test.

3 Results

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

The dynamic responses of Ca^{2+} to ATP stimulation were modeled via Eqns. (1-4), and fit to data in single cells using Bayesian parameter inference (Figure 1A). Only those MCF10A cells classified as "responders" to ATP were studied, i.e. we sought to explain the different levels and types of response; cells with Ca^{2+} peak heights less than 1.8 were removed prior to inference. To assess the use of transfer learning to inform prior distributions — i.e. the cell chain approach — we fit cells using either informative priors via the cell chain, or using uninformative priors, i.e. each cell is fit individually using an identical prior each time. For individually fit cells we used the "Lemon" prior (Table 1), which is also used for the first cell in a cell chain. In studying the effects of different prior choices, we found that scaling and clipping the prior standard deviation was necessary for successively stable marginal posterior distributions along a chain (Figure S1).

Inferring the parameter of the Ca^{2+} model via a cell chain resulted in more efficient and more accurate parameter inference, with shorter computational times and higher overall posterior model probabilities (Figure 1B–C). To investigate further, 500 cells were fit using a similarity-based cell chain (see Methods) with 500 warmup steps and compared to cells fit independently using either 500 or 1000 warmup steps; the longer warmup is required to produce fits of comparable quality to the similarity-based run (see Table S1 for a summary of the chains run here and below).

The posterior probabilities for models fit to cells from the similarity-based run are higher than those from individually fitted cells (Figure 1B; Table S2), with sampling times between 2x to 25x faster than individually fit cells (Figure 1C). Model fits from the similarity-based run (as quantified by the $\widehat{R}$ statistic) were better overall than those from individually fitted cells (Table S3). These trends are consistent across multiple runs each consisting of hundreds of fitted cells (Figure 1D). The use of informative priors (relative to individually fit cells) improves the efficiency and the accuracy of parameter inference.

We compared inference along chains with priors informed by cell-cell similarity with chains that are constructed by a random ordering of cells. We discovered that the performance of the random cell chains – evaluated by computational efficiency (sampling times) and accuracy of fits (model posterior probabilities) – was not significantly different than that of the similarity-based chains (Table S4). Therefore, although the use of priors informed by cell predecessors accelerated inference relative to individually fit cells, the choice of cell predecessors (similarity-

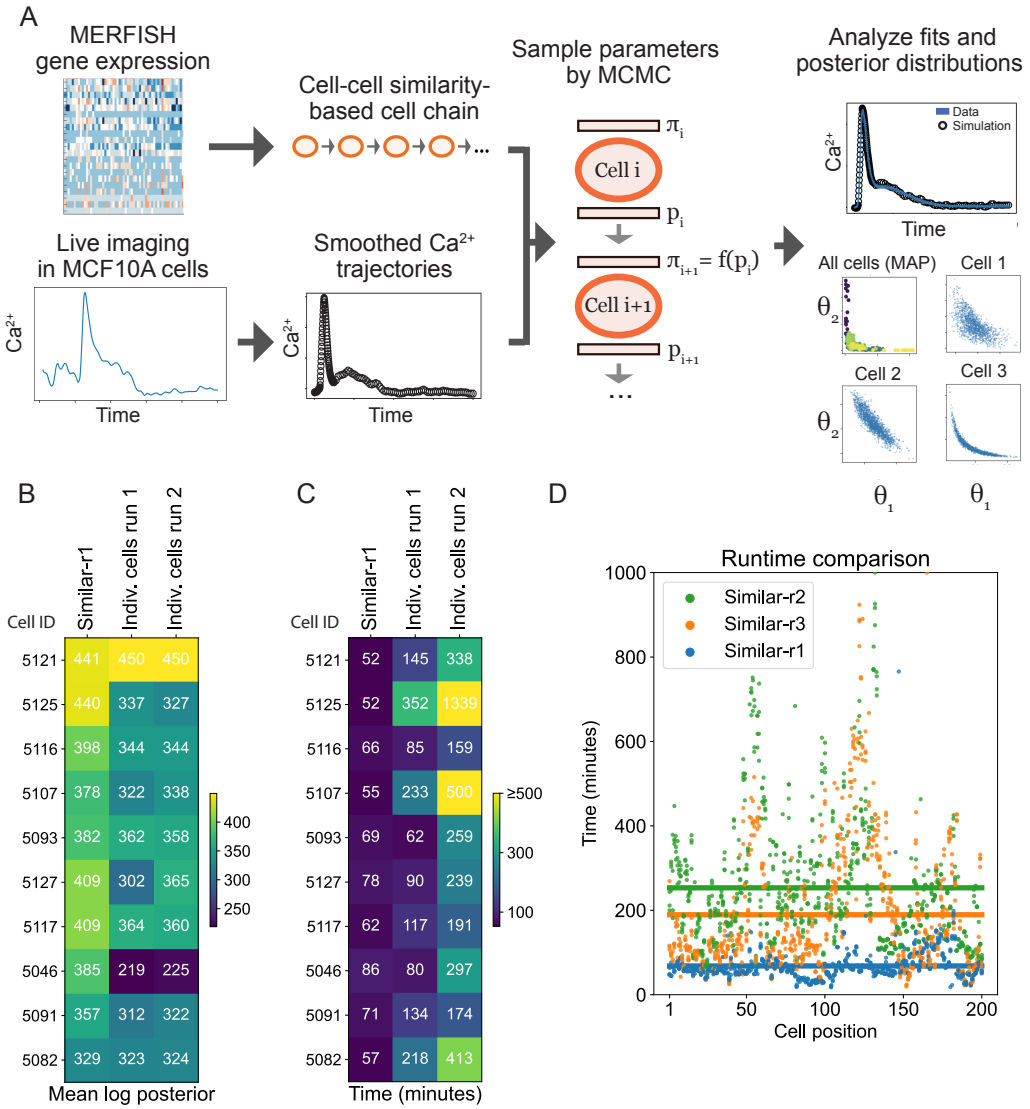

Figure 1: **Cell chains boost performance for single-cell parameter inference. A**: Workflow for single-cell parameter inference along a cell chain. π_i and p_i respectively denote the prior and the posterior distributions of cell i. MAP: maximum a posteriori value. **B-C**: Comparison of inference in a similarity-based cell chain (first column) vs. individually fitted cells (second two columns) by the fit quality: mean log posterior (B) and sampling time (C). Each row represents a single fitted cell. Individually fitted cells run-1 has 500 warmup steps and run-2 has 1000. **D**: Runtime comparison of HMC parameters. Accurate inference using cell chains is possible with fewer warmup steps and lower maximum tree depths, resulting in lower runtimes. Parameters: *(num. warmup steps; max. tree depth)* for r1: $(500, 10)$, r2: $(1000, 15)$, r3: $(500, 15)$.

based vs. randomly assigned) does not affect computational efficiency or the accuracy of fits.

We also studied the effects of varying NUTS parameters on the sampling efficiency. We found that sampling times were much faster overall when we reduced the maximum tree depth (a NUTS parameter that controls the size of the search space) from 15 to 10, since rarely was a tree depth > 10 used in practice, so this reduction did not negatively impact the model fits (Table S5). We also found that a warmup period of 500 steps was sufficient for convergence. Setting the maximum tree depth to 10 and the number of warmup steps to 500, leading to much faster sampling times overall (Figure 1D).

3.2 Analysis of single-cell posteriors reveals divergent intracellular and intercellular sources of variability

The posterior distributions of hundreds of cells show interesting differences between marginal parameters: some are consistent across cells in a chain while others vary widely. To quantitatively assess this, we ran two cell chains with the final 100 cells ordered identically but with different initial cells. We found that while some marginal posterior parameters were similar for all cells (e.g. $K_{\text{off, ATP}}$, Figure 2A), others can diverge for the same cells along a chain (e.g. d_5, Figure 2B). Relative changes in marginal posteriors however seemed to be tightly correlated. We computed the fold changes in mean marginal posterior parameter values between consecutive cells along the chain (Figure 2A-B, second row): the majority of consecutive cell pairs were tightly correlated both in direction and magnitude. We obtained similar results for random cell chains run in parallel with different initial cells (Figure S2).

Further analysis of the marginal posterior distributions revealed two uninformative ("sloppy" [28]) parameters. The posterior distributions of B_e and η_1 drifted, i.e. varied along the chain independent of the particular cell (Figure S3A-B). Given these insensitivities, we studied model variants where either one or both of these parameters were set to a constant. Comparing chains of 500 cells each, the reduced models performed as well as the original in terms of sampling efficiency and convergence (Figure S3C-E, Table S6). Posterior predictive checks of the reduced models showed no significant differences in simulated Ca^{2+} trajectories. Thus, for further investigation into the parameters underlying single-cell Ca^{2+} dynamics, we analyzed the model with both B_e and η_1 set to a constant (chain *Reduced-3*).

We discovered striking differences between intracellular and intercellular variability through analysis of the joint posterior distributions of parameters in chain *Reduced-3* . Several parameter pairs were highly correlated, as can be expected given their roles in the Ca^{2+} pathway, e.g. as activators or inhibitors of the same species. However, comparison of parameter correlations within (intra) and between (inter) cells yielded stark differences. Some parameter pairs showed consistent directions of correlation intercellularly (along the chain) and within single cells. The Ca^{2+} pump permeability (η_3) and the concentration of free Ca^{2+} (c_0) were positively correlated both inter- and intracellularly (Figure 2C). Similarly, the ER-to-cytosolic volume (ϵ) and the ER permeability (η_2) were negatively correlated in both cases (Figure 2D). However, the ATP decay rate (K_{ATP}) and the PLC degradation rate ($K_{\text{off, ATP}}$) were positively correlated along the chain (posterior means) but – for many cells – negatively correlated within the cell (Figure 2E). The distribution of MAP values is well-mixed, i.e. there is no evidence of biases arising due to a cell's position in the chain: the variation observed in the posterior distributions represents

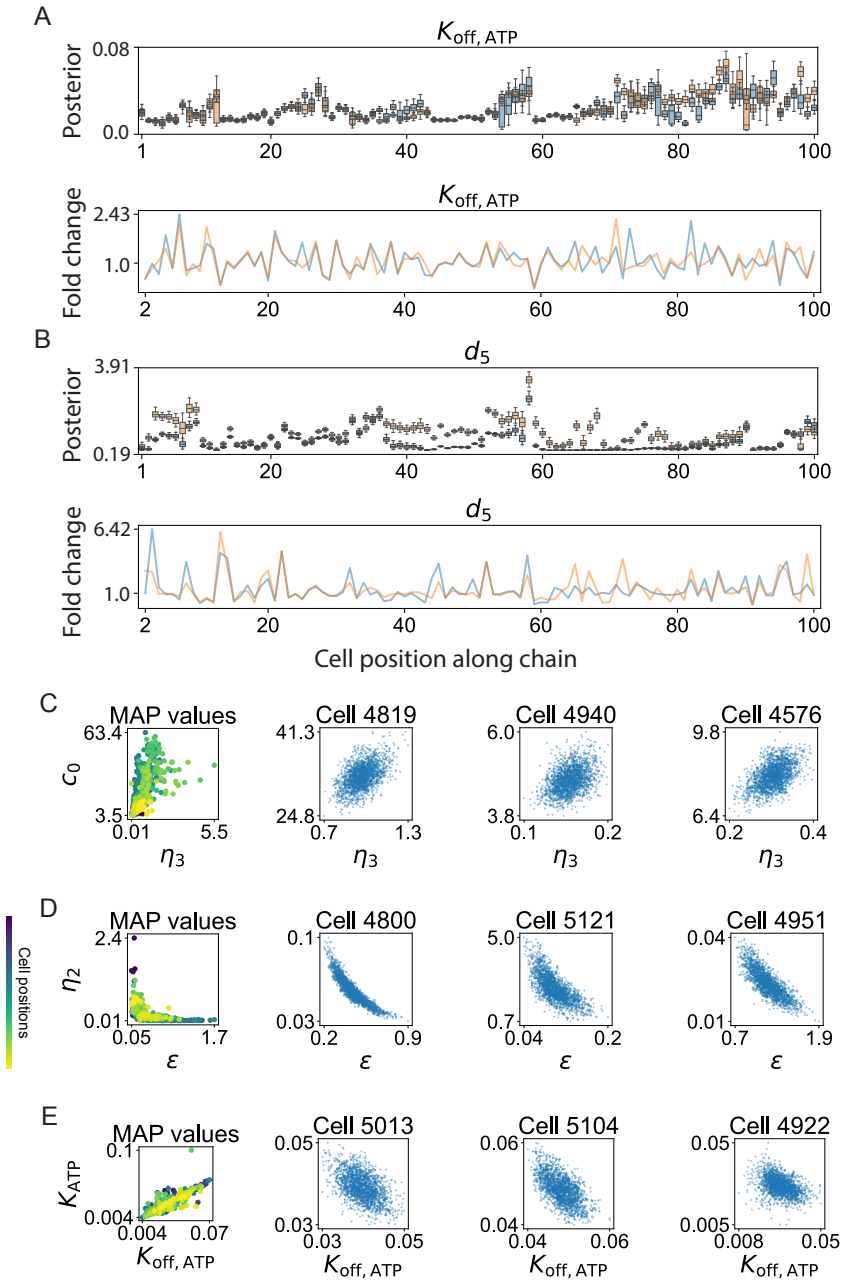

Figure 2: **Parameter dependencies revealed by the analysis of marginal posterior distributions along cell chains.** **A**: Marginal parameter posterior distributions for the PLC degradation rate ($K_{\text{off, ATP}}$) in two parallel runs of the same cell chain (upper), and the corresponding fold changes between consecutive cells in the chain (lower). **B**: As for (A) with the IP3 channel dissociation constant (d_5). **C**: Left: intercellular variability. Scatter plot of the maximum a posteriori (MAP) values of 500 cells, for η_3 and c_0, color indicates position along the chain. Right: intracellular variability. Scatter plots of 500 samples from the posterior of one cell from the chain; three representative cells are shown. **D**: As for (C) with parameters ϵ and η_2. **E**: As for (C) with parameters $K_{\text{off, ATP}}$ and K_{ATP}.

biological differences in the population. These differences may be in part explained by the differences in scale: intercellular parameter ranges are necessarily as large as (and sometimes many times larger than) intracellular ranges. On these different scales, parameters can be positively correlated over the large scale but negatively correlated locally, or vice versa. These divergent sources of variability at the inter- and intracellular levels highlight the complexity of the dynamics arising from a relatively simple model of Ca^{2+} pathway activation.

3.3 Quantifying the sensitivity of Ca^{2+} responses in a population of heterogeneous single cells

We conducted analysis of the sensitivity of Ca^{2+} responses to the model parameters within a population of cells. Typically, one defines a parameter sensitivity as the derivative of state variables with respect to that parameter [29, 30]. Here, we are interested in the sensitivity of Ca^{2+} responses to perturbations (small or large) within the confines of the parameter posterior distribution. I.e. we evaluate the sensitivity of the response to a given parameter by perturbing it not just locally, but across the distribution that parameter takes for the cell population. To do so, we sample from the posterior and alter the sample such that the parameter of interest is set to an extreme value according to its marginal posterior distribution (0.01-quantile or 0.99-quantile). We then simulate trajectories from these altered samples (Figure 3A).

The distance between simulated trajectories and the data was used to define the sensitivity of the Ca^{2+} response, where we take the mean distance of ten simulated trajectories. From the distribution of distances, it is clear that Ca^{2+} responses vary greatly: sensitive to some model parameters and insensitive to others (Figure 3B). Notably, the distances of the least sensitive parameters had mean values of close to 1.0: similar to the distances obtained from the best-fit posterior values (Table S5), i.e. the Ca^{2+} response is insensitive to these parameters across the whole posterior range. The insensitive parameters were not simply those which had the lowest posterior variance: there was little correlation between the inferred sensitivity and the posterior variance (Table S7), compare, e.g., parameters d_1 and d_5.

Analysis of the Ca^{2+} responses to parameter perturbations provides means to quantify how parameter perturbations affect the signaling pathway dynamics (Figure 3C-E). We compared an example of an insensitive parameter, d_{inh} (Figure 3C) with two sensitive parameters that control distinct aspects of the Ca^{2+} response. d_1 inversely controls the Ca^{2+} peak height: lower values of d_1 lead to consistently higher Ca^{2+} peaks and vice versa (Figure 3D). In contrast, η_2 controls the value of the Ca^{2+} steady state reached upon decay from the peak (Figure 3E). Higher values of η_2 led to consistently higher Ca^{2+} steady state values and vice versa. Of the most sensitive parameters, most control aspects of the Ca^{2+} dynamics directly, however notable exceptions include the IP3 degradation rate, $K_{\text{off, IP3}}$. The importance of IP3 in Ca^{2+} signal transduction is in agreement with the results of Yao et al. [17]; here we go further in that we can quantify the particular properties of the Ca^{2+} response affected by each parameter. In the case of $K_{\text{off, IP3}}$, similar to d_1 (although to a lesser extent), the main effect appears to be differences in the peak height of the Ca^{2+} response (Figure S4).

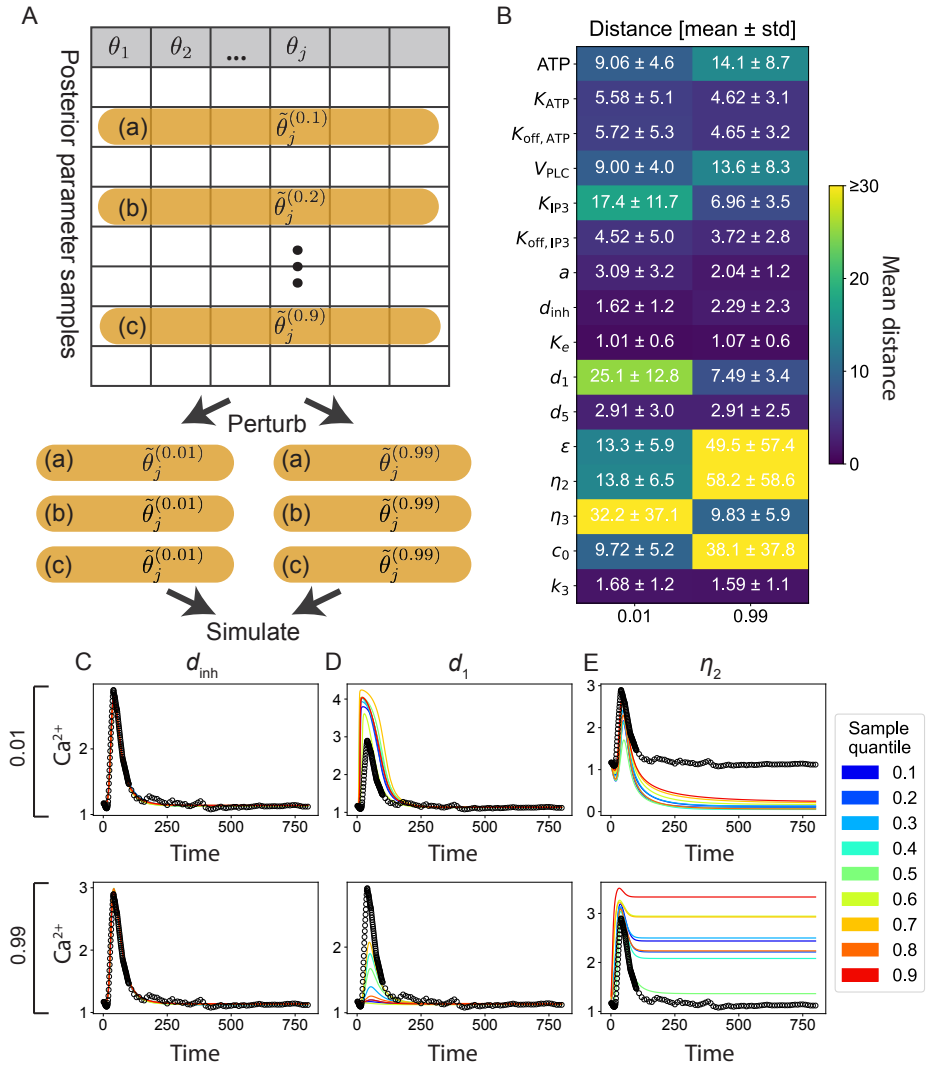

Figure 3: **Sensitivity of Ca^{2+} responses to parameter perturbations. A**: Sensitivities of Ca^{2+} responses with respect to parameters were determined by perturbing each parameter according to its marginal distribution (set values to 0.01- or 0.99-quantile). **B**: Parameter sensitivity across a population of cells from a similarity-induced chain (*Reduced-3* model). Sensitivities for the 0.01 and 0.99 quantile are reported as mean distances. std: standard deviation. **C**: Examples of simulated trajectories depicting responses to parameter perturbations: the Ca^{2+} response is insensitive to d_{inh}. **D**: Simulated trajectories in response to perturbations in d_1: in response the Ca^{2+} peak heights were altered. **E**: Simulated trajectories in response to perturbations in η_2: in response the steady states of Ca^{2+} were altered.

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

We studied variation between pairs of genes and parameters sampled from a cell population to assess whether relationships between them might exist. We found that several gene-parameter pairs were correlated. In general, the proportion of variance explained between a gene-parameter pair was low; this is to be expected given the many sources of variability in both the single-cell gene expression and the Ca^{2+} responses.

Analysis of the most highly correlated gene-parameter pairs (see Methods and Table S8) identified a number of genes that were correlated with multiple parameters, e.g. PPP1CC, as well as parameters that were correlated with multiple genes, e.g. η_3. Pairwise relationships were quantified via linear regression. The top four correlated gene-parameter pairs from a similarity-based cell chain are shown in Figure 4A-D. We performed the same analysis on a randomly ordered cell chain, where the same gene-parameter relationships were recapitulated, albeit with lower absolute correlation values (Figure 4E-H and Table S9). There is no discernable influence of a cell's position in a chain on the gene-parameter relationship, confirming that these correlations among a cell population reflects the variability in the population rather than any sampling artefacts.

We compared the top genes ranked by gene-parameter correlations for four populations: from two randomly sampled and two similarity-informed cell chains. Gene-parameter pairs were sorted by their absolute Pearson correlation coefficients, and the genes ranked by their position among sorted pairs. In total we identified 75 correlated gene-parameter pairs for the *Reduced-3* chain, applying a Bonferroni correction for multiple testing (Figure S5). Out of the top 30 of these, 25 appeared in the top 30 in at least 3/4 of the cell chains studied (Figure 4I). Of these 25 genes, 20 also appeared as top-10 marker genes from unsupervised clustering (into 3 clusters) of the gene expression data directly (Figure 4I). The high degree of overlap between these gene sets demonstrates that a subset of genes expressed in MCF10A cells explain not only their overall transcriptional variability but also their variability in Ca^{2+} model dynamics. These results are also suggestive of how information content pertaining to the heterogeneous Ca^{2+} cellular responses is encapsulated in the parameter posterior distributions.

Next, we turn our attention from the level of individual genes/parameters to that of the whole: what is the relationship between the posterior parameter distribution of a cell and its global transcriptional state? We used principal component analysis (PCA) for dimensionality reduction of the posterior distributions to address this question. We selected a cell (denoted the "focal cell") from a similarity-based cell chain (*Reduced-3*) and decomposed its posterior distribution using PCA. We projected the posterior distributions of other cells onto the first two components of the focal cell (Figure 4J-K and Figure S6A-B) to evaluate the overall similarity between the posterior distributions of cells relative to the focal cell. On PCA projection plots, posterior samples are colored based on gene expression: samples are derived from cells that are either transcriptionally similar to the focal cell, or share no transcriptional similarity. Comparison of similar and dissimilar cells from the same population showed that cells that were transcriptionally similar were located closer to the focal cell than dissimilar cells (Figure 4L-M and Figure S6C-D). In contrast, similar analysis of a random cell chain showed that transcriptionally similar cells were not located closer to the focal cell than dissimilar cells (Figure S7). Notably, proximity of posterior samples derived from transcriptionally similar cells

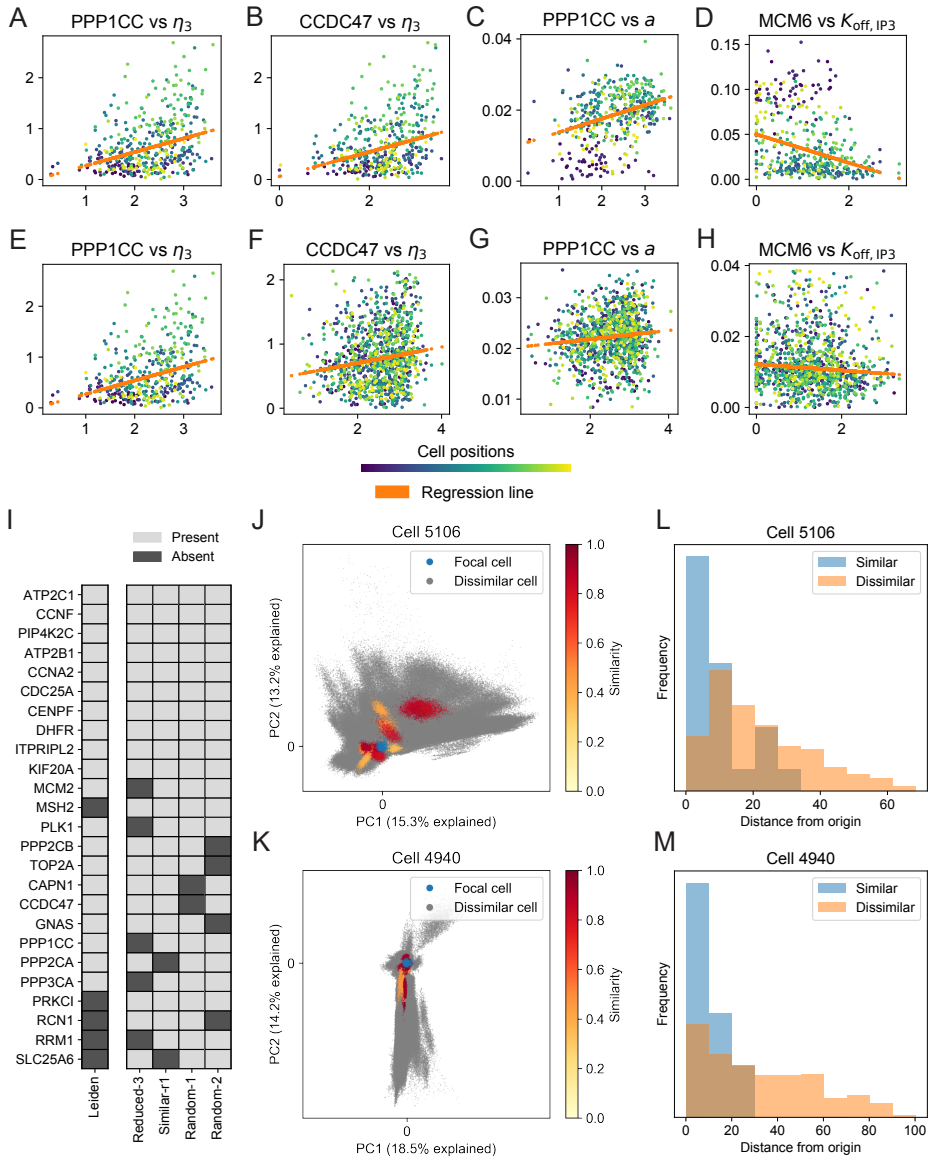

Figure 4: **Variability in Ca^{2+} model dynamics is associated with variability in gene expression.** **A-D**: Top four correlated gene-parameter pairs from similarity-informed chain (*Reduced-3*); color denotes cell position along the chain; cells are well mixed according to their chain positions. **E-H**: The same four gene-parameter pairs shown from a randomly ordered cell chain (*Random-2*). **I**: Consensus of genes identified as present/absent in the top 30 genes from gene-parameter correlation analysis and the top 30 marker genes from Leiden clustering. **J**: Projection of samples from the posterior (500 per cell) onto the first two principal components for the focal cell 5106, shown in blue. Posterior samples from dissimilar cells are gray and posterior samples from similar cell are yellow/red, where the shade indicates the global transcriptional similarity. **K**: As for (J) with PCA performed on cell 4940 as the focal cell. **L**: Histogram of mean distances between the origin (focal cell 5106) and samples from the surrounding cells in (J): similar or dissimilar. **M**: As for (L), with cell 4940 as the focal cells, i.e. samples from (K).

was not driven by a cell's position along the chain (no block structure observed; Figure S8). Similarities between posterior distributions of transcriptionally similar cells were thus not driven by local cell-cell similarity, but rather underlie a global effect and denote a relationship between the transcriptional states of cells and the Ca^{2+} pathway dynamics that they produce.

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

To characterize the extent to which we can predict Ca^{2+} responses from knowledge of the model dynamics, we clustered 500 cells from a similarity-based cell chain (*Reduced-3*) based on the single-cell posterior distributions using hierarchical clustering (see Methods). Three clusters were obtained (Figure 5A). Each cluster showed 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 S9).

The distinct dynamic profiles can be explained by the parameter sets that give rise to them: low-responders are characterized by high concentration of free Ca^{2+} in the ER (c_0) and low activation rates of IP_3R (Figure S9, Figure S10, Figure S12). Early-responders are characterized by parameters leading to faster and earlier IP3 and PLC dynamics, and late-high-responders are characterized by small values of d_1 (Figure S12).

Comparison of posterior parameter clustering with that performed by Yao et al. [17] shows that in both cases one of the three clusters obtained was characterized by stronger responses to ATP and correspondingly higher values of d_{inh} (Figure S12) [17]. In Yao et al., 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-responder cluster, but not in the early responders. In our results, d_5 was higher for the early-responders, in contrast with Yao et al. (Figure S12). 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 [17].

To understand the distinct Ca^{2+} dynamic profiles in light of single-cell gene expression, we performed two additional analyses. We clustered the same 500 cells based solely on their gene expression using community detection (Leiden algorithm in Scanpy [25]); we also clustered 500 cells from a randomly ordered cell chain using the same approach for hierarchical posterior clustering. For the cell clustering based on gene expression, as for the similarity-based cell clustering, distinct Ca^{2+} profiles are observed: in the case of the gene expression-based clustering these consist of "Ca-low", "Ca-mid", and "Ca-high" responses (Figure 5C). In contrast, no distinct Ca^{2+} dynamic responses could be observed for the posterior clustering based on the random cell chain (Figure 5D and Figure S11).

We performed differential gene expression analysis on each set of clusters (from the similarity-based chain, the randomly ordered chain, and from the gene expression-based clustering; Figure 5E-G). Distinct markers for each cluster were obtained for the similarity-based clustering and the gene expression-based clustering, but were not discernible for the random chain-based clustering. Clustering of cell posteriors from the randomly ordered chain was thus unable to distinguishable Ca^{2+} dynamic profiles nor gene expression differences. On the other hand, clustering posteriors from a similarity-based chain identified distinct gene

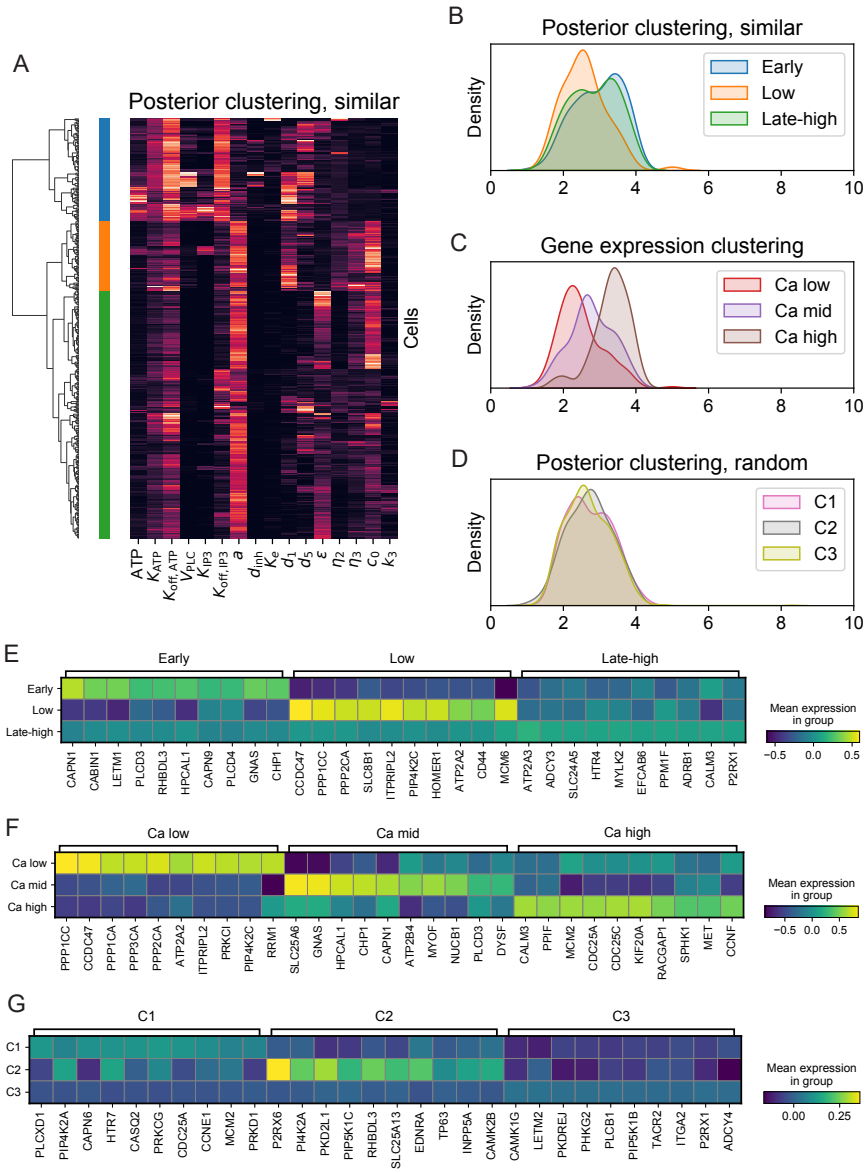

Figure 5: **Clustering of cell posterior distributions reveals marker genes for Ca^{2+} states.** **A**: Agglomerative clustering on posterior means from a similarity-based chain (*Reduced-3*) using Ward linkage ($k = 3$ clusters). **B**: Kernel density estimate of Ca^{2+} peak height from posterior clustering of a similarity-based chain. **C**: Kernel density estimate of Ca^{2+} peak height from gene expression clustering of the same cell population as in (A–B). **D**: Kernel density estimate of Ca^{2+} peak height from posterior clustering of a randomly ordered cell chain (*Random-2*). **E**: Top ten marker genes per cluster from similarity-based posterior clustering. **F**: Top ten marker genes per cluster from gene expression clustering of cells. **G**: Top ten marker genes per cluster from posterior clustering of randomly ordered cell chain.

expression profiles. Moreover, these overlapped with the marker gene profiles obtained by clustering on the gene expression directly. I.e. parameter inference of single-cell Ca^{2+} dynamics from a similarity-based chain enables the identification of cell clusters with distinct transcriptional profiles and distinct responses to ATP stimulation.

Analysis of the genes that are associated with each Ca^{2+} profile showed that low-responder cells were characterized by upregulation of CCDC47 and PP1 family genes (PPP1CC and PPP2CA). Early-responder cells were characterized by upregulation of CAPN1 and CHP1, among others. The late-high responder cells were characterized by increased expression CALM3 among others, although the marker genes for this cluster were less evident than the others. There was considerable overlap in the marker genes identified by posterior clustering vs. by gene expression clustering: the early-responder signature overlapping with the Ca-mid cluster, and the low-responder signature overlapping with the Ca-low cluster. The posterior distributions of cells fit from similarity-based (but not random) cell chains capture information regarding the transcriptional states of cells that overlaps with the information gained by gene expression clustering directly. Thus, specific Ca^{2+} model parameter regimes that characterize distinct Ca^{2+} responses contain information about distinct single-cell gene expression states.

4 Discussion

We have presented methods for inferring the parameters of a signaling pathway model, given data describing dynamics in single cells coupled with subsequent gene expression profiling. We hypothesized that via transfer learning we could use posterior information from a cell to inform the prior distribution of its neighbor along a "cell chain" of transcriptionally similar cells. To the best of our knowledge, this was the first parameter inference framework for dynamic models that incorporates single-cell gene expression information into the inference framework. Implemented using Hamiltonian Monte Carlo algorithm for MCMC sampling [21], we discovered that using cell predecessors to construct priors did indeed lead to faster sampling of parameters. However, these improvements did not rely on the use of gene expression to construct priors: the performance of randomly sampled cell predecessors was equivalent. In the case that cell chains were 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 mappings from state to dynamic cell fate.

The model studied here is described by ODEs to characterize the Ca^{2+} signaling pathway, adapted from [19, 16], consisting 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. Parameter sensitivity analysis highlighted the effects of specific parameter perturbations on the Ca^{2+} dynamic responses. 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 unsupervised clustering of the posterior distributions, we found that distinct patterns of

Ca^{2+} in response to ATP could be mapped to specific variation in the single-cell gene expression. In previous work using similar approaches for clustering [17], posterior parameter clusters predominantly revealed response patterns consisting of responders and non-responders; here we excluded those cells that did not exhibit a robust response to ATP. We are able to characterize subtler the Ca^{2+} response dynamics (described by "early", "low", and "late-high" responders) and predict which transcriptional states give rise to each. This approach is limited since relatively little gene expression variance is explained by an individual model parameter: it may be possible to address this in future work by surveying a larger range of cell behaviors, e.g. by including a wider range of cellular responses or by considering higher-level co-variance in the posterior parameter space. It also remains to be tested whether the given model of Ca^{2+} dynamics is appropriate to describe the signaling responses in cell types other than MCF10A cells.

Our ability to fit to of the single cells tested came potentially at the expense of an unwieldy model size. 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. Without data with which to constrain the three additional model species, we needed to constrain the model in other way. We used an approach of "scaling and clipping" for construction of the priors, i.e. setting ad hoc limits to control posterior variance. More effective (and less ad hoc) techniques could improve inference overall and may become necessary in the case of larger models. 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.

Connecting dynamic cell phenotypes to transcriptional states remains a grand challenge in systems biology. The limitations of deriving knowledge from gene expression data alone [31] have led to the proposal of new methods seeking to bridge the gap between states and fates [32]. Here, making use of technology that jointly measures Ca^{2+} dynamics and gene expression in single cells, we have shown that parameter inference informed by transcriptional similarity enables us to begin to make state-to-fate maps. 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] 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.

[26] 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.

[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] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Computational Biology*, 3(10):e189, October 2007. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0030189. URL https://dx.plos.org/10.1371/journal.pcbi.0030189.

[29] 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.

[30] 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.

[31] 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.

[32] 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.