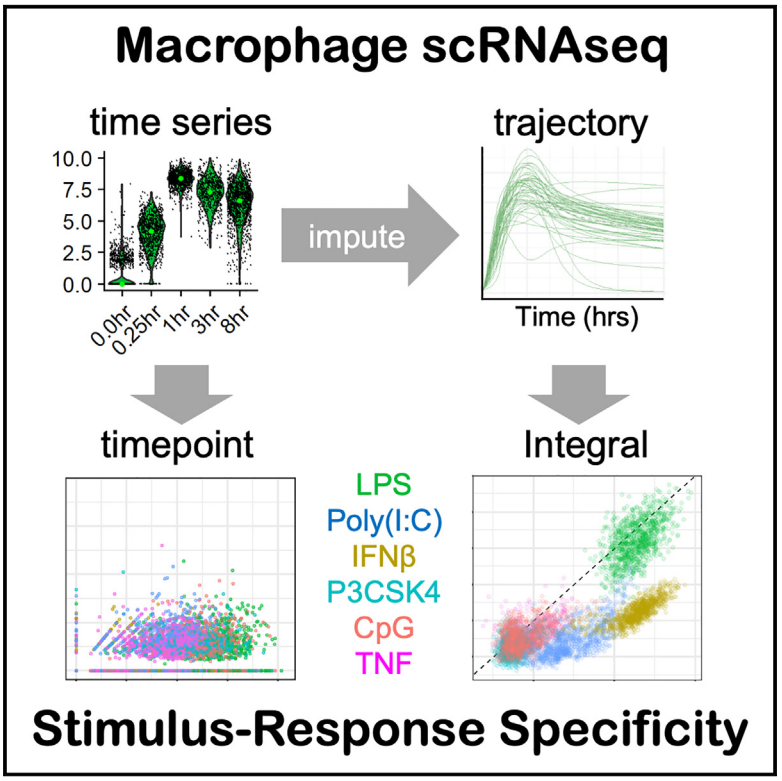
Molecular Cell

Article

Single-cell stimulus-response gene expression trajectories reveal the stimulus specificities of dynamic responses by single macrophages

Graphical abstract



Authors

Katherine M. Sheu, Aditya Pimplaskar, Alexander Hoffmann

Correspondence

ahoffmann@ucla.edu

In brief

Because RNA measurements are cell destructive, it is unclear how variable stimulus-induced dynamic gene expression trajectories (GETs) are. Sheu et al. developed a method to assess single-cell gene expression trajectories (scGETs) in macrophages responding to stimuli and found scGETs to be much more informative than any single time-point measurement.

Highlights

- Imputation of stimulus-induced single-cell gene expression trajectories (scGETs)
- Trajectory features reveal unexpected stimulus specificity in macrophage responses
- Expression integrals identify cell-cell correlations and co-regulated gene sets
- Stimulus-response expression trajectories readily characterize cell functional states

Sheu et al., 2024, Molecular Cell *84*, 4095–4110
November 7, 2024
https://doi.org/10.1016/j.molcel.2024.09.023

CellPress

Molecular Cell

CellPress
OPEN ACCESS

Article

Single-cell stimulus-response gene expression trajectories reveal the stimulus specificities of dynamic responses by single macrophages

Katherine M. Sheu,[1] Aditya Pimplaskar,[1] and Alexander Hoffmann[1,2,*]
[1]Department of Microbiology, Immunology, and Molecular Genetics, and Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, 611 Charles E. Young Dr S, Los Angeles, CA 90093, USA
[2]Lead contact
*Correspondence: ahoffmann@ucla.edu
https://doi.org/10.1016/j.molcel.2024.09.023

SUMMARY

Macrophages induce the expression of hundreds of genes in response to immune threats. However, current technology limits our ability to capture single-cell inducible gene expression dynamics. Here, we generated high-resolution time series single-cell RNA sequencing (scRNA-seq) data from mouse macrophages responding to six stimuli, and imputed ensembles of real-time single-cell gene expression trajectories (scGETs). We found that dynamic information contained in scGETs substantially contributes to macrophage stimulus-response specificity (SRS). Dynamic information also identified correlations between immune response genes, indicating biological coordination. Furthermore, we showed that the microenvironmental context of polarizing cytokines profoundly affects scGETs, such that measuring response dynamics offered clearer discrimination of the polarization state of individual macrophage cells than single time-point measurements. Our findings highlight the important contribution of dynamic information contained in the single-cell expression responses of immune genes in characterizing the SRS and functional states of macrophages.

INTRODUCTION

Macrophages are immune sentinel cells that deploy their diverse functions in response to exposure to pathogens or activating stimuli.[1,2] Immune response genes not only differ in the magnitudes of induction but also show distinctive temporal dynamics evident in the gene expression trajectory (GET). GETs reveal that some genes respond rapidly and often transiently, with others being delayed constituting a second phase and still others being activated into a near permanent state.[3–5] These diverse dynamics relate to phasing different immune functions, starting with the initiation of tissue inflammation and ending in resolution and wound healing.[6]

Recent studies have furthermore revealed that macrophage responses are stimulus specific as evidenced by transcriptomic response to stimulation with defined pathogen-associated molecular patterns (PAMPs) or cytokines.[7] Indeed, the stimulus-response specificity (SRS) of macrophages, which describes how tightly regulated is the deployment of only stimulus-appropriate immune functions, is thought to be a functional hallmark of healthy macrophages.[8] Yet, quantifying the SRS of gene expression responses is challenging as it requires not only measures of the heterogeneity of cells but also their full GETs over the time course.

Thus far, studies of inducible GETs at the transcriptome level have largely been confined to bulk assays of a population of cells given the destructive nature of single-cell RNA sequencing (scRNA-seq).[9–11] Thus attempts to quantify SRS of gene expression responses have been confined to single time-point snapshot assays, which revealed that SRS is affected by microenvironmental context either via defined polarizing cytokines or the undefined milieu of obese and aged mice.[12] But they also showed that single cells within a population show substantial heterogeneity in gene expression and that much stimulus information is lost, presumably for two reasons: (1) snapshot expression values do not allow us to consider dynamic trajectory features of GETs, and (2) technical noise of the assay could not be mitigated by repeated sampling of the same cell. In contrast, recent single-cell signaling studies, enabled by fluorescent reporter proteins and non-destructive live-cell imaging, have shown that there is substantial SRS information in the temporal dynamics of nuclear factor κB (NF-κB) and mitogen-activated protein kinase (MAPK)[13–15] and that these are modulated by polarizing cytokines.[16] Yet these studies only report on one or two signaling intermediates rather than the hundreds of immune response genes that are also subject to chromatin epigenetic heterogeneity.

Studying stimulus-response GETs at the single-cell level is particularly important in macrophages, as they generally function

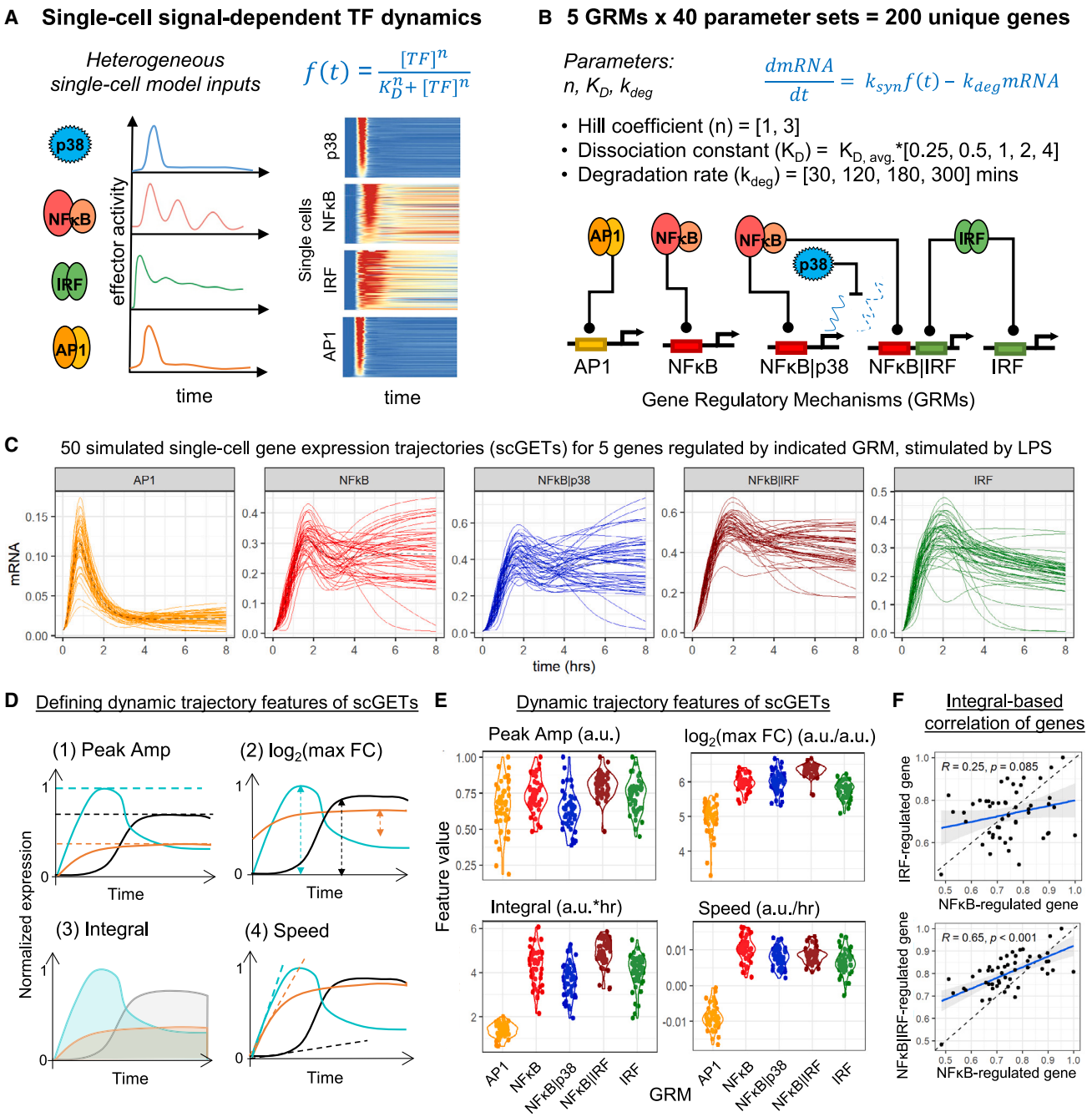


Figure 1. Mechanistic models simulate the full trajectories of single-cell gene expression responses

(A) Schematic of single-cell signaling activity used as signal-dependent transcription factor (SDTF) inputs into the mathematical models of gene expression. Single-cell signaling activity trajectories were derived from measured averages.[10]

(B) Mathematical models are used to simulate unique genes with different gene regulatory mechanisms (GRMs) and activation/degradation characteristics. Transcriptional activation and mRNA degradation rates are determined by three parameters, the dissociation constant K_D (affinity $K_{TF} = 1/K_D$), the Hill coefficient n, and the mRNA degradation rate k_{deg}, as described by Cheng et al.[10]

(C) Example single-cell gene expression trajectories (scGETs) simulated from the models described in (B). Five example genes are shown that have different GRMs, with 50 single-cell trajectories shown per gene. Genes were defined by the following parameters (genes shown have parameters taken from GRM cluster averages determined in Cheng et al.[10]): AP1-regulated gene (n = 1, $K_{TF,AP1}$ = 0.48, mRNA half-life 30 min); NF-κB-regulated gene (n = 1, $K_{TF,NF\text{-}κB}$ = 1.3, mRNA half-life 30 min); NF-κB|p38-regulated gene (n = 1, $K_{TF,NF\text{-}κB}$ = 1.3, mRNA half-life p38 dependent [Figure S1B]); NF-κB|IRF-regulated gene (n = 1, $K_{TF,NF\text{-}κB}$ = 1.3, $K_{TF,IRF}$ = 1.25, mRNA half-life 30 min, governed by OR gate regulatory logic); and IRF-regulated gene (n = 1, $K_{TF,IRF}$ = 1.25, mRNA half-life 30 min).

(legend continued on next page)

as individuals rather than within homotypic assemblages. Thus, aberrant outlier responses by individual cells may drive pathology, such as the failure to resolve the inflammation associated with an immune response event. Indeed, the need for studying single-cell GETs (scGETs) is broadly appreciated and new experimental modalities have been pioneered but come with limitations: Fluorescent reporters of mRNA abundance require genetic perturbation of the chromosomal locus and are limited to one or two genes at a time.[17–21] Microscopically sampling cytoplasmic RNA repeatedly from the same cell requires long recovery times and is associated with unavoidable technical noise.[22] Metabolic labeling allows for measurements of mRNA synthesis rate and abundance,[23] but it does not quantify GETs that allows determination of dynamic features.

Can scGETs be computationally inferred from scRNA-seq data? Multiple computational methods identify GETs across pseudotime, by leveraging the heterogeneity of expression states to order cells along a pseudotime axis. Pseudotime has been applied to understand biological processes such as cell differentiation,[24] disease courses,[25,26] or progressive changes in immune cell gene expression after stimulation[27] (Table S1). However, pseudotime approaches cannot provide scGETs of macrophage stimulus responses where cells respond to a perturbation with gene expression dynamics that are more rapid than the heterogeneity in their timing.

Here, to study the SRS of macrophage gene expression responses to immune stimuli, we developed an experimental workflow and computational approach to impute the ensemble of true-time scGETs. We evaluated the scGET imputation approach using hundreds of scGETs simulated from mathematical models, such that the ground truth is known. We then applied the method to new single-cell datasets of immune response genes across thousands of macrophages of three polarization states, whose expression is measured over five time points after stimulation with six distinct pathogen or cytokine stimuli. We found that trajectory expression features from individual macrophage cells were much more informative about the stimulus than single time-point measurements. The trajectory features of distinct combinations of genes are coordinated to provide stimulus distinction, and the differences in gene regulatory mechanisms (GRMs) (for example, between *Nfkbia* and *Tnf*) were more apparent when considering trajectory integrals rather than time-point data. Polarizing cytokine microenvironments altered the dynamics and variation in macrophage-inducible gene expression, such that individual cell functional states could be characterized more precisely than with single time-point measurements.

RESULTS

Simulating the dynamics of scGETs

Single-cell signaling studies of macrophages have captured the inducible dynamics of transcription factor or kinase activity via live-cell microscopy.[13,28–31] The features of these signaling protein trajectories are dependent on the stimulus but also exhibit cell-to-cell heterogeneity (Figure 1A).[10,14] Yet, within each cell, the dynamic activity of signaling proteins results in the further downstream activation of hundreds of unique genes. To explore what dynamic gene expression might look like in single cells, we constructed and simulated mathematical models of immune response genes informed by prior literature, each with their synthesis and degradation rates regulated by gene-specific parameters and the previously measured activities of signaling effector molecules after lipopolysaccharide (LPS) stimulation (Figures S1A and S1B).[10,32,33] Genes were classified into five previously characterized GRMs (genes regulated by Activator Protein-1 [AP1], NF-κB, interferon regulatory factor [IRF], NF-κB|p38, or NF-κB|IRF),[10,12,32] where genes within each GRM differed in which effectors controlled their activation and degradation (Figure 1B). Each of the six stimuli activated signaling effectors with stimulus-specific dynamics (Figure S1C). In response to LPS, the resulting simulations illustrated the full induction dynamics of 200 unique genes for each individual cell, with cell-to-cell heterogeneity derived from single-cell signaling heterogeneity (Figures 1C and S1A).

To quantify the dynamical patterns of gene expression, we considered a set of metrics that characterizes GETs: relative peak amplitude (Peak Amp), maximum log fold change from time zero (Max LFC), total expression over the time course (Integral), and activation speed (Speed) (Figure 1D; Table S2). Calculating these features for the simulated ground truth scGETs summarized the single-cell heterogeneity of trajectories with a set of interpretable values (Figure 1E). Notably, as each cell contains the trajectories of hundreds of genes, gene-gene correlations across single cells can be determined. As expected, across single cells, genes that were regulated by GRMs that shared transcription factors had highly correlated trajectory features, while genes that are regulated by different transcription factors were less correlated (Figure 1F).

An imputation method identifies scGETs from time series data

To determine whether scGETs can be imputed from scRNA-seq time series data, we developed a method that encompasses two distinguishing steps: direct linkage of measured cells across adjacent time points and interpolation across unmeasured time frames (Figure 2A). Similar to other scRNA-seq analysis methods,[34–37] we employed principal-component analysis (PCA) on data from all time points, with the goal of assembling individual cell trajectories before then recovering individual gene trajectories over hundreds of cells (Figure 2B; STAR Methods). To ameliorate the sparsity of single-cell measurements, cell archetypes (metacells) were constructed via k-means clustering. Cell archetypes are linked across time points using random walks weighted by a transition probability matrix, inversely proportional to the Euclidean distance between

(D) scGETs can be decomposed into dynamical features (arbitrary units [a.u.]; scGETs normalized to the maximum over all trajectories).
(E) Violin plots showing the single-cell distributions of dynamical feature values for the five example genes in (C).
(F) Gene-gene correlation of the Integral values across single cells are indicative of related GRMs. Top: NF-κB-regulated gene vs. IRF-regulated gene (same genes as C). Bottom: same NF-κB-regulated gene vs. NF-κB|IRF-regulated gene (same genes as C).

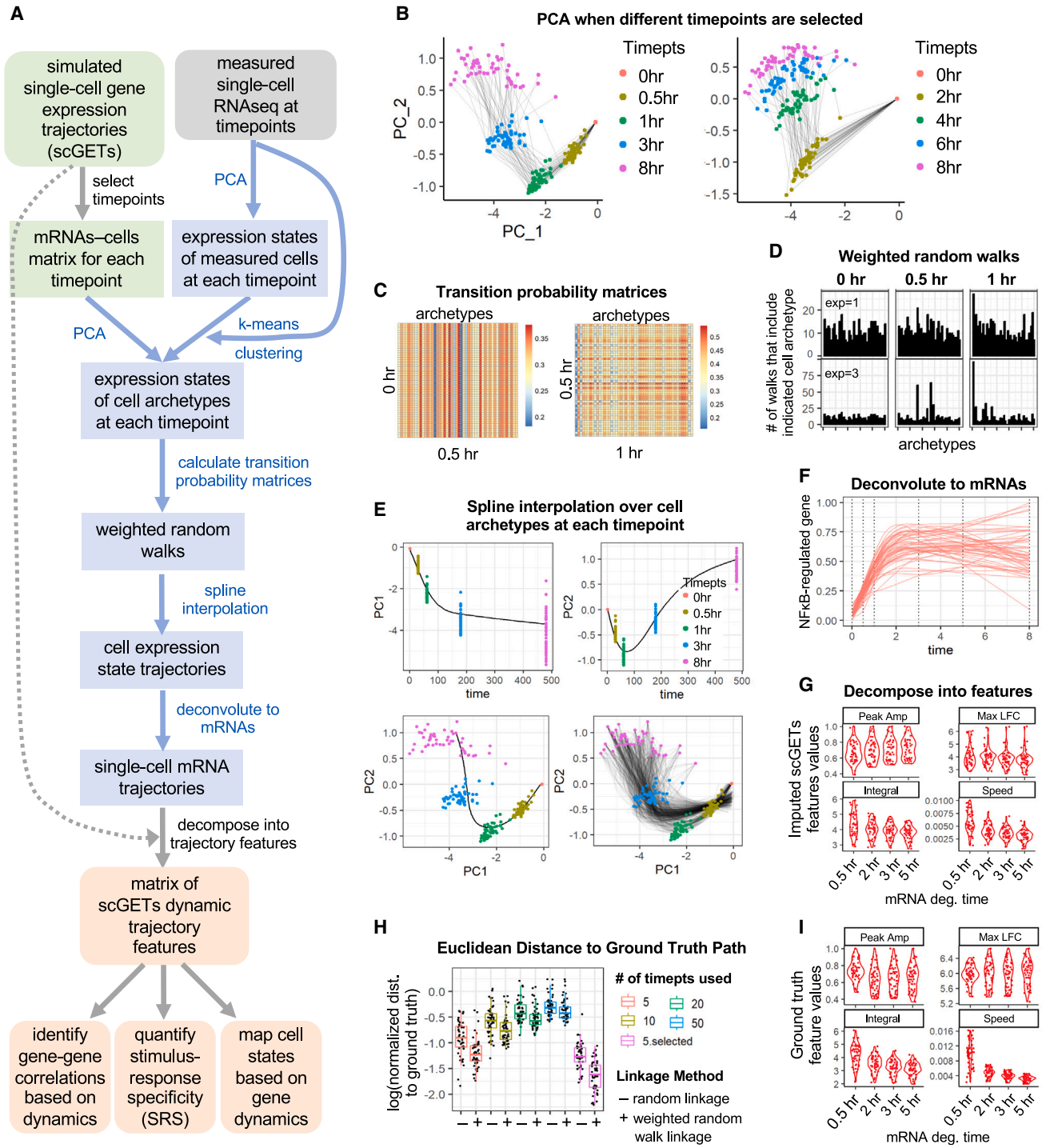


Figure 2. An imputation method identifies scGETs that match time series data distributions

(A) Schematic of a workflow to impute the ensemble of single-cell gene expression trajectories (scGETs) and quality control its performance. Key steps of the imputation workflow are depicted in blue. This workflow begins with measured scRNA-seq data at several time points (gray) or for quality control, simulated scGETs (green). The resulting scGETs are then decomposed in dynamic trajectory features, which may be used to address three questions (orange) with higher precision than single time-point measurements.

(B) PCA on simulated data for selected time points, either five front-loaded time points (referred to as 5.selected) or five evenly spaced time points, shows that selecting front-loaded time points produces smoother cell-state trajectories. For simulated data, the same number of cells was present at each time point, and k-means clustering was not used. Gray lines represent ground truth connections over time.

(legend continued on next page)

cell archetype pairs over adjacent time points (Figure 2C). The weighted random walks make every permutation of cell-cell connections possible but reduce the probability of trajectories across less likely paths (Figure 2D); exponentiation of the transition probability matrices makes inclusion of certain cell archetypes more determined but reduces tolerance to noise (Figure S2A). To interpolate between measured time points and obtain time course single-cell trajectories, splines were then fit across each principal component. The result is a set of trajectory profiles decomposed over principal components, with each component's spline describing a feature of the cell's transcriptomic trajectory (Figure 2E). Using the original PCA loadings, the cell trajectories are then projected back to the original gene expression space to recover correlated transcriptome-scale GETs (i.e., the expression dynamics of hundreds of individual genes per cell over real time) for hundreds of cells per gene (Figure 2F). scGETs can then be described by decomposing them into dynamical features (Figure 2G).

Evaluating the imputed expression trajectories against a model-simulated ground truth

To evaluate the imputed trajectories, at each step, we compared output to the ground truth dynamics provided by model simulations. We compared random walks weighted by transition probability matrices vs. completely random cell-cell connections, and we found that using weighted random walk generated paths closer to the ground truth for any number of time points used (Figure 2H). We also compared the effect of selecting different distributions of time points on cell-cell linkage, and we found that five time points selected to be front-weighted performed better than five evenly spaced time points (Figure 2H). We varied the number of components used to calculate the cell-cell linkages, and we found that more components did not significantly improve performance but that using weighted random walks and five front-loaded time points resulted in paths closer to the ground truth across any number of components (Figure S2B).

The trajectory features of the imputed ensemble of trajectories, compared with those of the ground truth, were concordant across a range of simulated genes with different mRNA degradation rates (Figure 2I). We explored different interpolation options to smooth the trajectories over selected single-cell paths (Figure S2C), as this choice might affect the quantification of trajectory features, including locally estimated scatterplot smoothing (LOESS),[38] cubic splines with three (spline.df3) or five (spline.df5) degrees of freedom, cubic splines using leave-one-out cross validation (spline.cv),[39] and monotone Hermite splines (spline.mono).[40] We found that each reasonably recapitulated the different trajectory features of the ground truth across a range of genes with different synthesis and degradation characteristics (Figure S2D). Across interpolation methods, however, gene trajectories with rapid degradation and high dissociation constants had Integrals that were less accurately recapitulated (Figure S2D). We proceeded with cubic splines fitted via leave-one-out cross validation as they provided both gene-specific features and maintained single-cell heterogeneity in comparison to the ground truth distribution of features.

Time series scRNA-seq unveils time-dependent heterogeneity in macrophage responses

Based on the performance analysis of the imputation method, we generated time series scRNA-seq data from thousands of macrophage cells at five relatively front-loaded time points, following addition of the stimulus (0, 0.25, 1, 3, and 8 h). We stimulated macrophages of three different polarization states with a panel of six immune ligands, encompassing cytokines or distinct bacterial or viral components, which each bind distinct Toll-like receptors (TLRs) or cytokine receptors: LPS (TLR4), poly(I:C) (PIC) (TLR3), Pam3CSK4 (P3C) (TLR2), CpG (TLR9), tumor necrosis factor (TNF) (TNF receptor [TNFR]), and interferon (IFN)-β (interferon-α/β receptor [IFNAR]) (Figure 3A). This resulted in over 72 scRNA-seq samples consisting of in total >100,000 individual macrophages (Figure S3A). We noted that gene expression heterogeneity was highly time-point dependent, and thus no single time point accurately represented the heterogeneity of macrophage responses to stimuli (Figure 3B). To reduce drop-out rate and improve sequencing depth for genes of interest such as cytokine genes (Figure 3B), sequencing was performed on a targeted set of 500 immune response genes previously identified as providing high SRS in macrophages[12] (Figure S3B; STAR Methods).

We collected replicate data from both overlapping and complementary time points for naive M0 macrophages (0, 0.5, 3, 5,

(C) Heatmaps of the transition probability matrices weighted by Euclidean distance of every pair of cells from one time point to the next. Shown are the heatmaps for the first two time steps from 0 to 0.5 h and from 0.5 to 1 h. Blue boxes indicate lower Euclidean distance and thus higher probability linkages.
(D) Random walks are weighted by inverse distance, so that some cell archetypes are passed through more frequently. Exponentiating the probability matrix decreases the randomness of the random walks and identifies the high probability linkages (in applications of the method in the paper, the probability matrix was not exponentiated). Two archetypes in particular are included at higher frequency at 0.5 h and 1 archetype at 1 h, corresponding to the archetypes in (C) that have low pairwise Euclidean distance across adjacent time points.
(E) Spline interpolation over the paths determined by weighted random walks. Top: PC1 (left) or PC2 (right) for a single-cell trajectory with respect to real time. Colored points indicate the distribution of PC scores of single-cell gene expression values collected at each time point. Bottom: PC1 vs. PC2 plotted for a single cell (left) or hundreds of cells (right), given the distribution of single-cell gene expression values collected at each time point.
(F) Recovery of single-cell trajectories for individual genes (scGETs) from the interpolated PC scores matrix, using the loadings matrix from the original PCA. Shown is an NF-κB-regulated gene (gene defined by parameters $n = 1$, $K_{TF,NF\text{-}\kappa B} = 1.3$, mRNA half-life 30 min).
(G) Violin plots of the distribution of trajectory features obtained from decomposing imputed scGETs (illustrated in Figure 1D and defined in STAR Methods) for the NF-κB-regulated gene in (F) and three other genes that only vary in mRNA half-life.
(H) Based on Euclidian distance, weighted random walks are closer to ground truth paths than completely random connections, given an equivalent number and spacing of time points, across a range of time points used. The number of principal components is held constant at 50. Five selected front-loaded time points (5.selected) are closer to the ground truth than five evenly spaced time points (5).
(I) Dynamical features calculated from the ground truth model-simulated scGETs show similar distributions to those imputed (G).

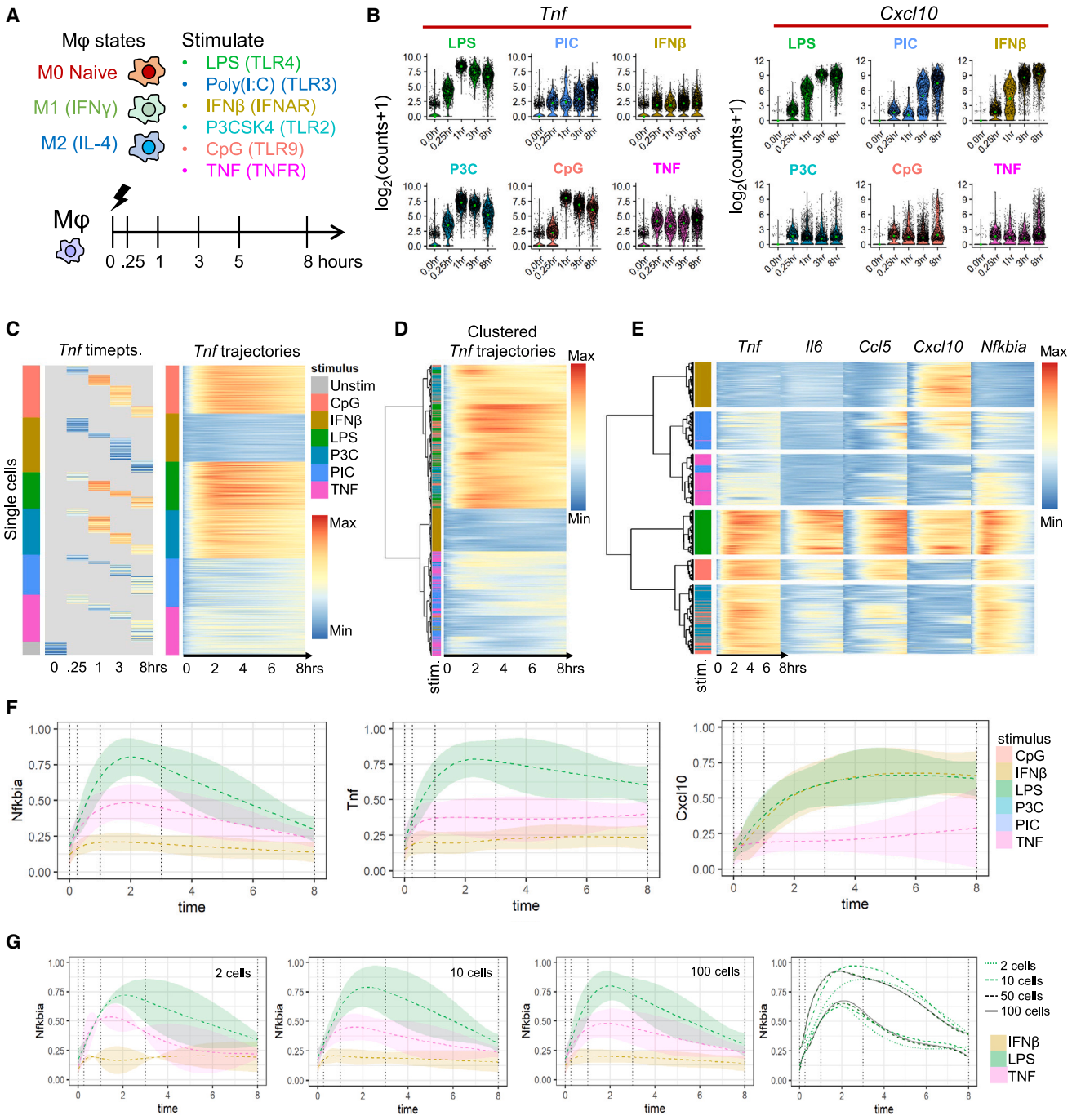


Figure 3. Imputation of scGETs in single macrophage cells responding to immune threats

(A) Schematic of time series scRNA-seq data collected for 3 macrophage states responding to 6 different immune stimuli, totaling over 72 scRNA-seq samples (3 macrophage states × 6 stimuli × at least 4 time points).

(B) Violin plots illustrate the stimulus-response distributions of single-cell *Tnf* vs. *Cxcl10* expression for M0 macrophages responding to each of the labeled stimuli, at each of the measured time points. Black points represent single cells. Green points represent the median across cells.

(C) Left: heatmap of measured time series data for *Tnf*. Once the cell is sampled, it is destroyed, leading to gray boxes (no measurement) across time. Right: imputed scGETs for *Tnf* expression in stimulated M0 macrophages. scGETs are displayed as a heatmap scaled to max expression across all stimuli. Row annotation colors indicate the stimulus each cell received.

(D) Hierarchical clustering of single cells using the *Tnf* expression trajectory. Row colors indicate the stimulus each cell received. scGETs scaled to max expression across all stimuli.

(legend continued on next page)

and 8 h) to evaluate data quality and reproducibility. Replicate time points produced concordant data (Figure S3C). Pseudobulk patterns of the single-cell data were also concordant with previously published population-level macrophage stimulus-response RNA-seq data[10] (Figure S3D).

Imputation of scGETs in single macrophage cells responding to immune threats

As time series scRNA-seq data are snapshots, they fail to capture a fundamental aspect of a single cell's immune response transcriptome—its inducible dynamics. We applied both pseudotime algorithms[36,41] and the scGET imputation method to our time series scRNA-seq data of macrophages stimulated with LPS (Figure S4). Unlike pseudotime trajectories (Table S1),[27,34,36,42,43] which provide the order of single cells along a sometimes branching trajectory (Figures S4A and S4B), scGETs instead provide heterogeneous gene induction characteristics such as fold changes, peak amplitudes, induction speeds, or accumulated total in mRNA abundance (Figures S4C and S4D). Plotting the pseudotime trajectories vs. single-cell real-time dynamics of a few genes illustrates this contrast (Figures S4E and S4F). Across multiple genes, pseudotime places single cells in order based on their expression profiles, and genes can be clustered based on their pseudotime pattern (Figure S4G); in imputing scGETs, the order is known as the data were collected across a time series, and single-cell archetypes are explicitly linked across adjacent time points, enabling clustering of single cells based on scGETs (Figure S4H).

Applying the scGET imputation method to time series scRNA-seq data from populations of M0 macrophages stimulated with six different ligands, we recovered continuous single-cell expression dynamics for the 8 h of stimulation, such as for *Tnf* (Figure 3C). We evaluated the accuracy of the imputed single-cell trajectories for each gene by comparing the mean and variance of the inferred trajectory ensemble to the measured mean and variance in the scRNA-seq data, at each of the measured time points (Figures S5A and S5B). Notably, some genes were more accurately imputed than others. *Tnf* was among the more poorly imputed genes, with relatively higher normalized root-mean-square deviations (RMSDs) for mean and variance (Figures S5C and S5D). However, even *Tnf* trajectories still reasonably matched the data and maintained the expected SRS (Figure 3C). Mean and variance of other genes more closely matched the data at the measured time points, such as *Ccl5* and *Cxcl10* (Figures S5C and S5D). In addition, the induction patterns of scGETs were consistent with the known GRM for each gene,[10,32] while providing an estimate of the cell-to-cell heterogeneity in the expression dynamics.

Importantly, the imputed scGETs allowed single macrophage cells to be clustered using information from the entire trajectory of gene expression, rather than solely their gene expression at a particular time point. The scGET of a single cytokine gene, *Tnf*, revealed that macrophages responded distinctly to at least three stimuli groups—IFN-β, bacterial PAMPs (LPS, P3C, CpG), and TNF/PIC—using solely *Tnf* as a readout (Figure 3D). In contrast, *Tnf* expression measured at only a single time point, such as 3 h post-stimulation, does not distinguish even two stimuli groups,[12] supporting the notion that a tailored response to immune threats relied on *Tnf* expression dynamics and not simply on expression level.

Furthermore, imputing scGETs recovered the co-regulated expression dynamics of all measured genes for each cell, enabling an assessment of stimulus-specific responses based on the heterogeneity of multigene time courses (Figure 3E). Hierarchical clustering on the combination of dynamic information from five genes, pro-inflammatory cytokines *Tnf* and *Il6*, lymphocyte chemokines *Ccl5* and *Cxcl10*, and NF-κB negative feedback regulator *Nfkbia*, further separated single cells responding to different stimuli, indicating that the robustness of distinct macrophage cell responses was through multi-gene complementary dynamical features of expression. Cells stimulated with TNF vs. PIC, which were undistinguishable by *Tnf* dynamics alone (Figure 3E), could now be distinguished by the late induction of *Cxcl10* and *Ccl5* in response to PIC only (Figure 3E). Measuring a single early time point such as at 1 h of stimulation would not have captured this temporal distinction in gene expression. Plotting expression distribution boundaries further illustrated that specificity between stimulus pairs is reliant on the regulation of heterogeneity of specific genes over the time course, for example, with LPS vs. TNF response distributions partially overlapping at late time points for *Nfkbia* but distinct for *Tnf* and *Cxcl10* (Figure 3F). We noted stable distribution boundaries from the imputation of ∼100 cells, suggesting that a sufficient number of scGETs were imputed (1,000 cells) to well capture the characteristics of stimulus-specific response distributions (Figure 3G).

Dynamical features of scGETs convey stimulus information

To quantify the dynamical patterns of scGETs, we again considered the presence of four dynamical features that characterize gene expression responses: Peak Amp, Max LFC, Integral, and Speed (Table S2). For *Tnf*, *Nfkbia*, and *Cxcl10*, plotting the single-cell distributions of trajectory features suggested that Peak Amp and Integral were more stimulus specific than Max LFC or Speed, with less overlapping single-cell distributions (Figure 4A). To quantify the SRS, we calculated the maximum mutual

(E) Hierarchical clustering of the response trajectories of five genes from each cell that have different gene regulatory strategies, which further distinguishes single-cell responses to each stimulus. Row annotation colors indicate the stimulus each cell received. scGETs are scaled to max expression across all stimuli.
(F) Single-cell gene expression response trajectory distributions, with distribution boundaries drawn at ±2 standard deviations, for three genes with different gene regulatory strategies. Vertical dotted lines at 0, 0.25, 1, 3, and 8 represent time points at which single-cell distributions were measured via scRNA-seq. Left: NF-κB-only gene. Center: NF-κB gene also affected by p38 activity. Right: IRF gene also affected by NF-κB activity. scGETs are scaled to max expression across all stimuli; only three stimulus condition distributions are shown for clarity.
(G) Distribution of scGETs for 3 example genes, for 3 different stimulus conditions, when 2, 10, or 100 cells are imputed. Far right: for LPS-stimulated cells, distribution boundaries drawn at ±2 standard deviations stabilize between 50 and 100 imputed cells.

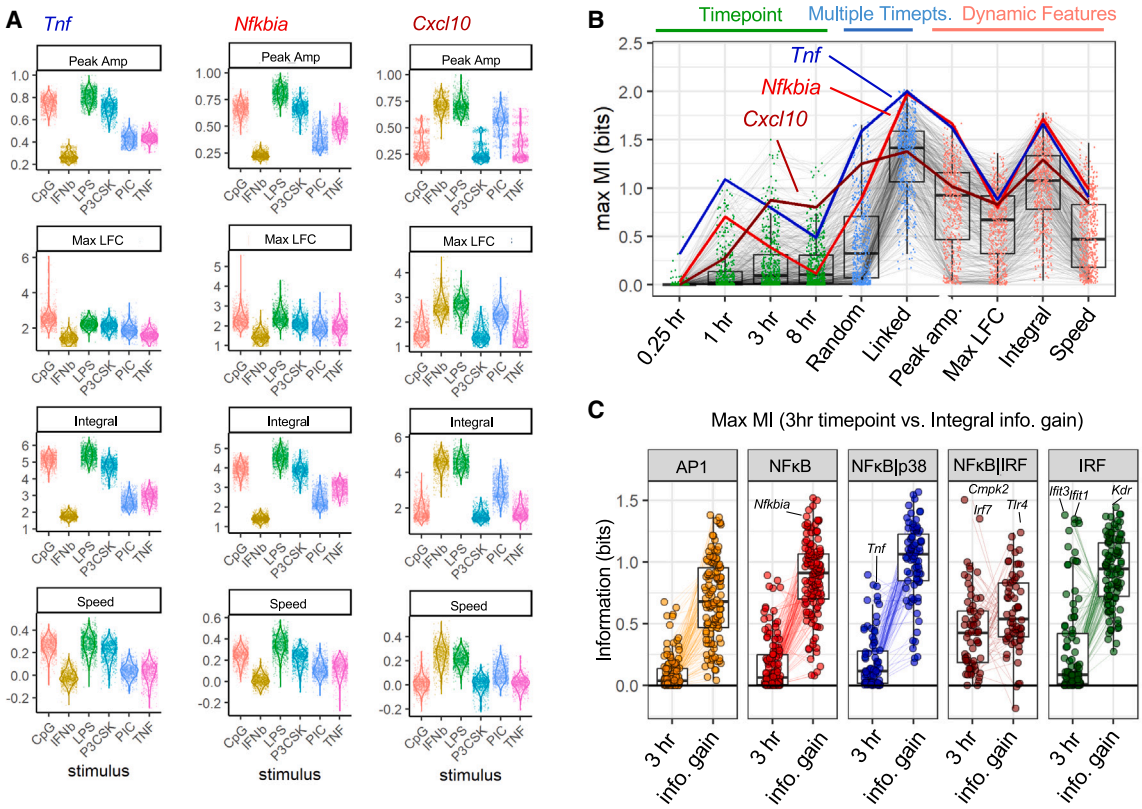


Figure 4. Dynamical features of imputed scGETs contain more stimulus information than single time-point values

(A) Distributions of the four dynamical features of scGETs: Peak Amp, Max LFC, Integral, and Speed. Each point represents a single cell's *Tnf*, *Nfkbia*, or *Cxcl10* dynamical feature value in response to the indicated stimuli.

(B) Stimulus information (quantified by max MI) within a gene when considering single time points (green) vs. linked time points (blue) vs. dynamical features (red), for all individual genes (points). Max MI values for each gene are connected by black lines across each feature type. Three known stimulus-response genes, *Tnf*, *Nfkbia*, and *Cxcl10*, are highlighted as examples.

(C) Paired boxplots of the max MI for a gene at 3 h post-stimulation vs. the corresponding information gain by considering the trajectory Integral. Genes were categorized by their reported GRM, resulting in five categories as also mathematically modeled in Figure 1. The trajectory Integral contains more stimulus information than the 3-h time point across most genes. Genes regulated by NF-κB|IRF are less reliant on the temporal dynamics of gene expression; the estimated SRS of NF-κB|IRF genes does not increase as much when considering trajectory Integrals.

information (max MI) between the six-stimulus input and the expression output. Higher max MI indicates more information about the stimulus and that the gene's dynamical feature is more stimulus specific.[44] We found that the max MI of each gene at a single measured time point was consistently less, compared with when trajectory information was available in the form of dynamical features (Figure 4B). In addition, we noted that using four randomly linked time points improved the max MI slightly over single time points, but four time points linked via weighted random walks revealed significantly greater SRS for all genes (Figure 4B).

To examine how much the features of scGETs reveal about the SRS, we next assessed the max MI of each feature, for each gene (Table S3). We found that Integral and Peak Amp had on average the highest max MI, while Max LFC and Speed were less informative (Figure 4B). Interestingly, certain genes more greatly relied on the differences in dynamical features to retain stimulus-specific information. For example, the NF-κB target gene *Nfkbia* was quantified as considerably less stimulus specific when considering expression at a time point than when considering either multiple time points linked via the scGET imputation method or *Nfkbia* dynamical features (Figure 4B). This finding aligns with prior work on the importance of the induction dynamics of *Nfkbia*,[5] which encodes a key NF-κB feedback protein, in generating appropriate responses to immune threats.

We noticed that for subsets of genes, the trajectory features revealed much more stimulus information than expression at a single time point, while other genes showed smaller gains in information, such as *Cxcl10* (Figure 4B). To investigate this further, we categorized the immune response genes based on which signaling pathways activate them, according to prior literature, resulting in five categories (AP1, NF-κB, IRF, NF-κB "AND" p38, NF-κB "OR" IRF).[10,12,32] Interestingly, genes that were regulated by single transcription factors such as AP1, IRF, or NF-κB showed an information gain of ∼1 bit when considering expression Integral rather than expression level, but genes regulated by an NF-κB|IRF "sequential OR" gate[10]

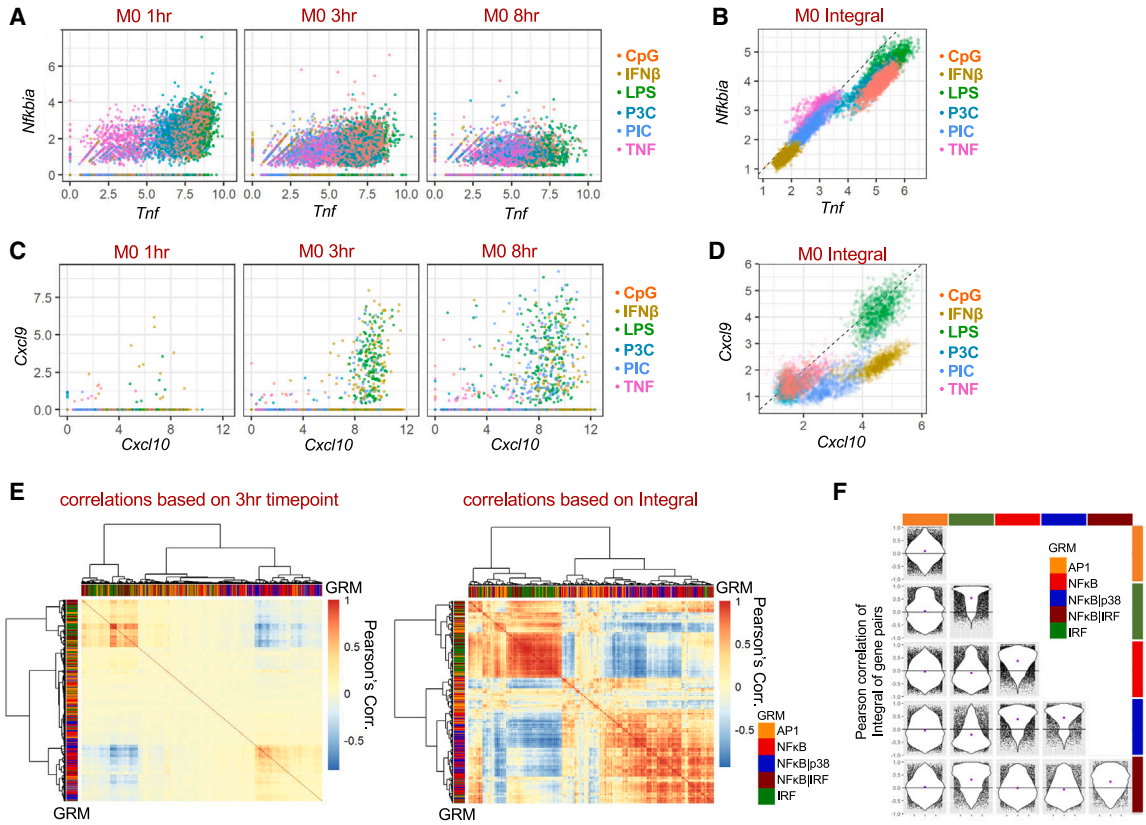


Figure 5. Dynamical features of imputed scGETs identify gene-gene correlations more reliably than single time-point scRNA-seq data

(A) Scatterplot of *Tnf* vs. *Nfkbia* expression from scRNA-seq data, at three of the time points measured after stimulation. Colors represent the stimulus the cell received. Only weak correlations are observed between these two NF-κB-regulated genes at any of the response time points.

(B) Scatterplot of *Tnf* vs. *Nfkbia*, using the dynamical feature Integral (total activity), reveals much stronger correlations. The dynamical feature Integral reveals the known regulation of *Tnf* by NF-κB|p38 while *Nfkbia* is NF-κB-only. The activation of MAPKp38 by the bacterial ligands LPS, CpG, and P3C results in a shift of *Tnf* expression from the identity line.

(C) Scatterplot of *Cxcl10* vs. *Cxcl9* expression from scRNA-seq data, at three response time points. Single-cell within-stimulus and across-stimulus correlations are time-point dependent, and responses appear uncorrelated.

(D) Scatterplot of *Cxcl10* vs. *Cxcl9*, using the Integral (total activity) values from scGETs, reveals within-stimulus correlations of gene expression across single cells and across stimulus response specificity. Co-regulated genes within single cells can appear uncorrelated at time points due to different timing of activation, but Integral reveals their correlation.

(E) Left: heatmaps of hierarchically clustered Pearson correlations across all genes at the 3-h time point. Right: hierarchical clustering of Pearson's correlations across all genes, using Integral values from scGETs.

(F) Pearson correlation values between the Integral values of scGETs in response to all six stimuli, of all gene pairs, separated by GRM. Genes of the same GRM or sharing the same transcription factors are correlated across single cells, and positive and negative correlations across GRMs are also detected.

showed gains of only ~0.5 bits (Figure 4C). This pattern suggested that the combinatorial activity of two transcription factors at the gene promoter may allow for stimulus distinction through combinatorial control with less reliance on the temporal dynamics of gene expression, while genes induced by a single transcription factor are more likely to show SRS through dynamic control (Figure 4C).

Trajectory features expose gene correlations not evident in time-point measurements

In transcriptome measurements of single cells, target genes of different signaling pathways may be correlated due to similar chromatin or signaling environments within each cell. To determine dynamics-dependent gene-gene correlations across single cells, we compared correlations of gene pairs that are targets of the same or different transcription factors, for either single time-point scRNA-seq data or scGET features. Two prominent NF-κB-regulated genes, *Tnf* and *Nfkbia*, showed weak to no correlation at any single response time point of 1, 3, or 8 h (Figure 5A). The lack of correlation at a single time point could be attributed to differences in the timing of activation: *Tnf* expression was more sustained, while *Nfkbia* more front-loaded, and thus total expression was not well captured by any single time point. However, considering the trajectory feature Integral revealed high gene-gene correlations both across stimuli and within each stimulus condition (Figure 5B). Examining the single-cell Integral values of these two genes also highlighted an increase in *Tnf* expression relative to *Nfkbia*

for cells stimulated with CpG, P3C, or LPS (Figure 5B). CpG, P3C, and LPS are immune ligands that signal through Myd88 to activate MAPKp38, resulting in an increase in NF-κB|p38-regulated *Tnf* expression relative to the NF-κB-only regulated *Nfkbia*. Only scGET features could clearly reveal this subtle shift in the *Tnf-Nfkbia* correlation that reflected the known regulation of these two genes.

We then examined two genes that encode Cxcr3 ligands, *Cxcl9* and *Cxcl10*, both chemokines produced by macrophages that lead to the recruitment of T cells (Figure 5C).[10,12] At single time points after stimulation, *Cxcl9* and *Cxcl10* were not well correlated across single cells, with only *Cxcl10* showing early activation by LPS-, IFN-β- or PIC-induced IRF signaling (Figure 5C). However, the Integrals of *Cxcl9* and *Cxcl10* were correlated to a much greater extent than evident from time-point data (Figure 5D). Furthermore, inspecting the Integral showed that *Cxcl9* and *Cxcl10* were positively correlated across single cells within each population of cells responding to the same stimulus, suggesting a co-regulation achieved through either shared activation of signaling pathways or similarly responsive chromatin environments. More globally, the correlation heatmap of all genes at a single time point of 3 h showed weak correlations between genes regulated by the same GRMs, while clustering genes based on Integral recovered stronger correlations that better distinguished NF-κB- vs. IRF-regulated gene clusters (Figure 5E). Gene pair Integrals showed high correlations when regulated by the same GRM or by GRMs with shared transcription factors, but some gene pairs that controlled different GRMs were also correlated (Figure 5F). Inspection of these pairs revealed correlations between, for example, NF-κB-regulated cytokines (e.g., *Il1a*, *Ccl5*) and IRF-regulated genes such as the virally induced *Heatr9*, which is involved in feedforward regulation of cytokines.[45,46] The full dynamic trajectory of single-cell gene induction was thus critical for capturing many additional correlative relationships between gene pairs, identifying the effect of shared chromatin or signaling environments.

Polarization alters which immune response genes mediate macrophage specificity

How do dynamical feature distributions of scGETs change when macrophages are polarized by microenvironmental cytokines? We next imputed scGETs using time series scRNA-seq data from thousands of M1(IFN-γ)- and M2(interleukin [IL]-4)-polarized macrophages, again stimulated with each of the six immune ligands (Figures S6A–S6C). Initial inspection of cells hierarchically clustered based on scGETs revealed expected changes to the average GETs of several immune response genes (Figure 6A). For example, in response to the bacterial PAMPs LPS, CpG, and P3C, the NF-κB target gene *Nfkbia* exhibited more transient activity in M1(IFN-γ) macrophages, which was diminished by 3 h, in contrast to more sustained activity in M0 macrophages. NF-κB|IRF target genes *Ccl5* and *Cxcl10* were more rapidly induced in M1(IFN-γ) macrophages, while exhibiting much slower speed of induction in M2(IL-4) macrophages (Figure 6A). The clustering of both M1(IFN-γ) and M2(IL-4) macrophage cells showed that responses remained stimulus specific, but the specificity could be attributed to different aspects of gene dynamics.

To identify genes with scGET dynamical features that were most variable across polarization states, we performed PCA on the scGET dynamical features using cells from all polarization states. We found that genes regulated by certain GRMs were weighted more highly in the PC loadings. For Max LFC, NF-κB- and NF-κB|p38-regulated genes were weighted most strongly along PC1, while IRF- and NF-κB|IRF-regulated genes were weighted most strongly along PC2, among them *Mx1* and *Ifit1* (Figure 6B; Table S4). Plotting *Mx1* and *Ifit1* Max LFC for all single cells showed that differences in the stimulus-response distributions of these IRF genes were lost in M1(IFN-γ) macrophages, most prominently between two bacterial stimuli like CpG and LPS (Figure 6C). This loss of specificity in IFN-γ-polarized macrophages was due to either more similar average responses across stimuli (e.g., *Mx1* Max LFC) or increased heterogeneity of responses (e.g., *Ifit1* Max LFC). A similar analysis of Speed showed that specificity produced in M0 macrophages by response Speed (e.g., *Mx1* or *Ifit1* Speed) was lost in M2(IL-4) macrophages, most notably between IFN-β and all other stimuli (Figure 6D). Taken together, the microenvironmental cytokines assessed here (IFN-γ and IL-4) had a notable impact on the heterogeneity of trajectory features, such as in Max LFC and Speed of IRF genes.

As polarization profoundly affected the expression dynamics of many genes, we asked how much these changes impacted macrophage SRS. We identified the gene combinations whose dynamical expression features were most informative of stimulus identity (Table S5; STAR Methods). Considering the Integral of scGETs in M0 macrophages, max MI approached ~2.5 bits with just three genes (*Mx2*, *Nfkbiz*, *Egr3*) (combination of an IRF gene, a NF-κB|p38 gene, and an AP1 gene), near the theoretical maximum of 2.58 bits (distinguishing six stimuli, as $2^1 = 2$, $2^2 = 4$, $2^{2.58} \approx 6$). A max MI near the theoretical ceiling could also be achieved with three genes in M1(IFN-γ) (*Mx2*, *Tnf*, *Gna15*) and M2(IL-4) macrophages (*Cited*, *Ehd1*, *Pou2f2*) (Figure 6E). In contrast, the top three gene combinations using a single time point (3 h) reached at most 2 bits of information (Figure 6E). Plotting the Integral of the selected genes corroborated the conclusion that the dynamic trajectory features separated the six stimuli distinctly, although some stimulus information was still lost due to overlap in CpG and P3C distributions (Figure 6F). Thus, while SRS of particular genes like *Cxcl10* might be diminished in polarized macrophages, macrophages maintain SRS upon polarization but rely on different gene combinations that mediate the specificity.

scGETs effectively distinguish polarization states

The context-dependent cell state of macrophages is typically assessed by protein markers or single-cell transcriptomics in their quasi-steady state. Yet, macrophage functions are deployed in response to stimuli. Therefore, we hypothesized that stimulus-response gene expression dynamics may provide more accurate markers of a cell's functional state. To examine the difference between steady-state measurements and scGET features in distinguishing polarization states, we focused on canonical marker genes that are used to identify M1(IFN-γ) or M2(IL-4) polarization, *Nos2* (M1 marker), *Cd86* (M1 marker), and *Retnla* (M2 marker). Plotting their expression values at the unstimulated

CellPress
OPEN ACCESS

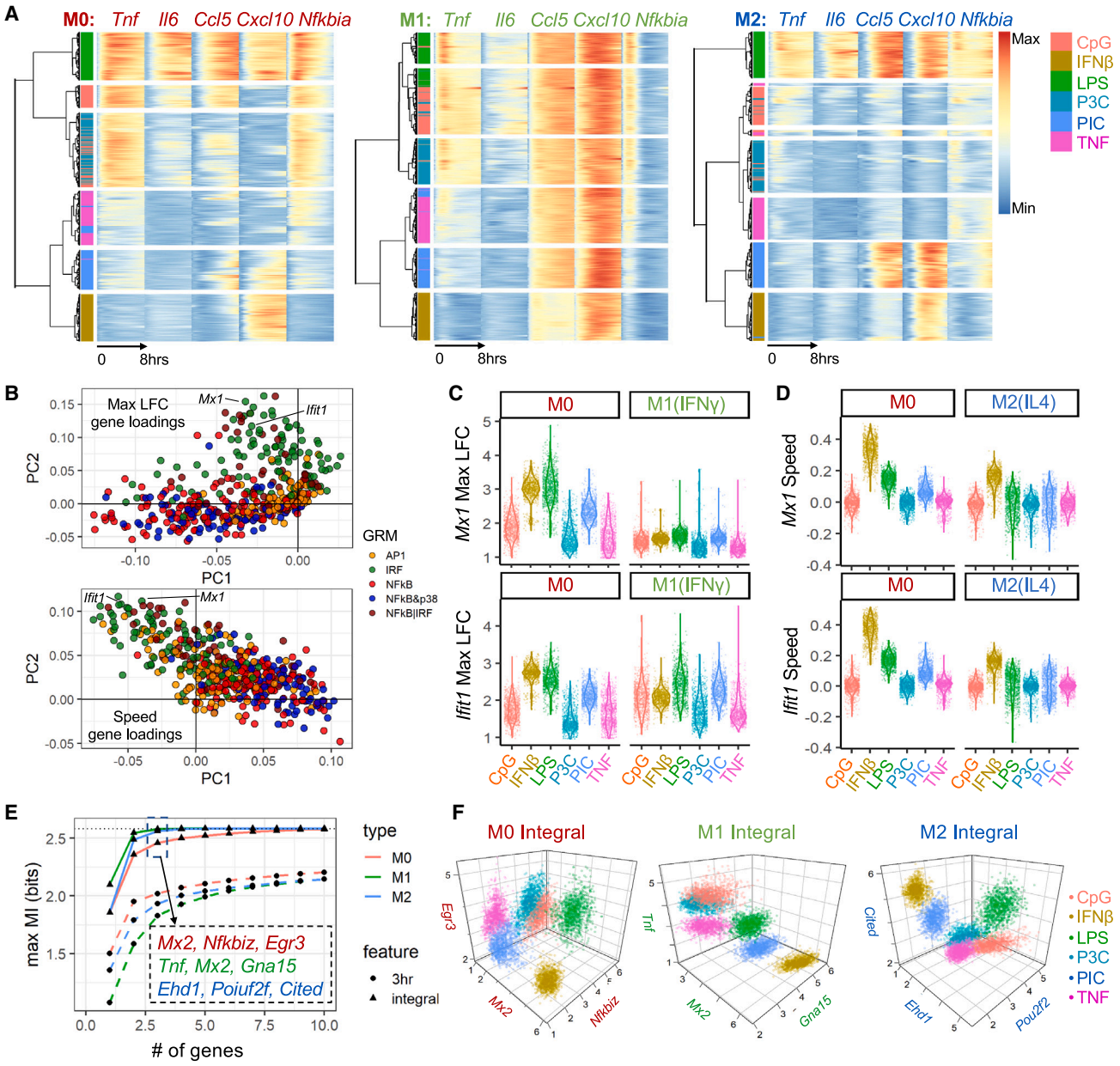


Figure 6. scGETs reveal how the SRS of macrophages is altered by polarization

(A) Hierarchical clustering of M1(IFN-γ) cells (center) and M2(IL-4) cells (right), compared with M0 (left), based on scGETs after stimulation, for a subset of cytokine and feedback regulator genes. Each row represents the imputed scGETs of a cell stimulated with the ligand indicated by the color code.

(B) PC loadings obtained from performing PCA on the dynamical features Max LFC (top) or Speed (bottom) for all genes across all macrophage cells. Genes with high loading values are genes with highly variable dynamical feature values across polarization states, identifying genes with loss of specificity in either Max LFC or Speed.

(C) Violin plots of *Ifit1* and *Mx1* Max LFC distributions for M0 vs. M1(IFN-γ) macrophages illustrate that loss of SRS in the fold change of IRF genes occurs in M1(IFN-γ) macrophages.

(D) Violin plots of *Ifit1* and *Mx1* Speed distributions for M0 vs. M2 (IL-4) macrophages illustrate that loss of SRS in the activation speed of IRF genes occurs in M2(IL-4) macrophages.

(E) SRS is preserved in polarized macrophages. Line plot of SRS (max MI) for the most informative combination of genes, when considering scGET Integrals (solid lines) vs. scRNA-seq data at the 3-h time point (dotted lines). (Inset displays the genes selected as most informative in combination.) When considering Integral, the best 3-gene combination can account for nearly the maximum possible SRS of macrophage, ~2.58 bits, whereas the best 3-gene combination when considering a single time point post-stimulation (3 h) does not reveal as high a SRS.

(F) Scatterplots of scGET Integrals for each polarization state show that the different combinations of three complementary genes do indeed provide high SRS.

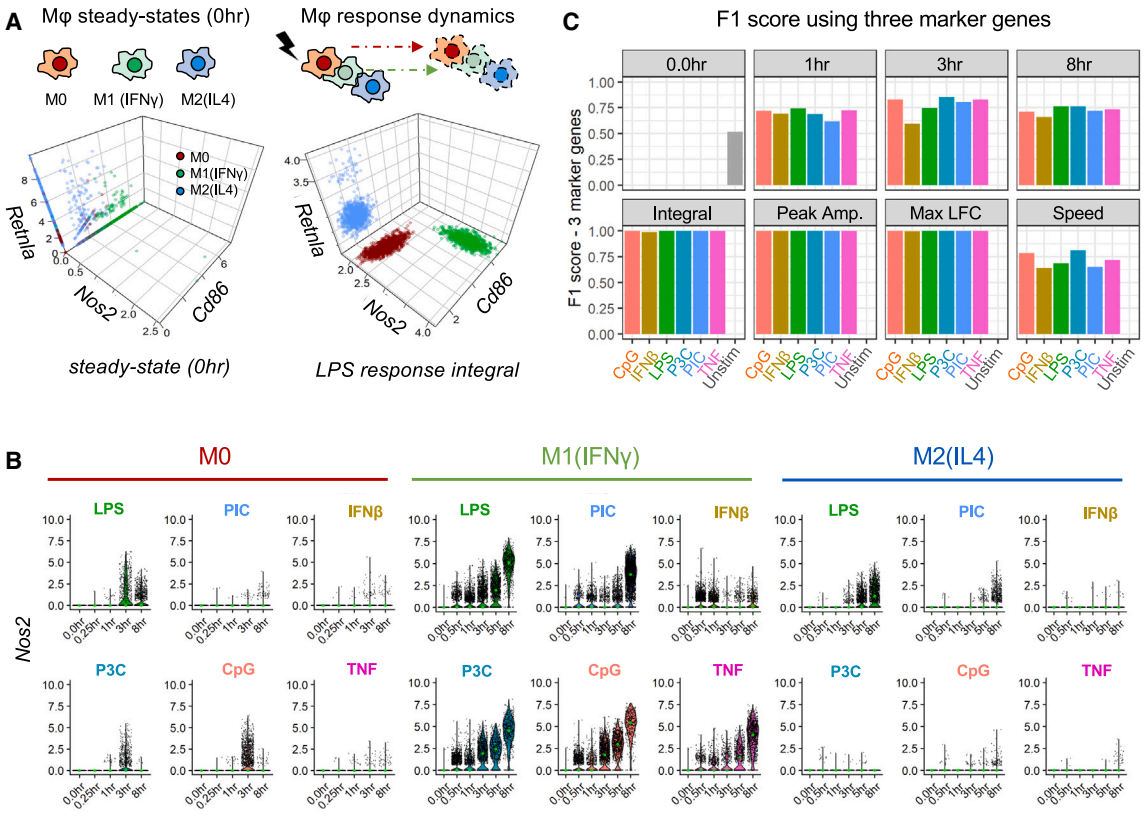


Figure 7. The scGETs of marker genes distinguish macrophage polarization states more precisely than time-point measurements

(A) Scatterplots of scRNA-seq expression values of three canonical M1(IFN-γ) or M2(IL-4) marker genes at steady state (unstimulated) vs. the LPS response Integral of scGET dynamics after 8 h of LPS stimulation. The response Integral of these canonical markers better distinguishes polarization states.

(B) Violin plots of time series scRNA-seq of one of the canonical M1(IFN-γ) macrophage markers, *Nos2*, across polarization states.

(C) F1 score of polarization state classification accuracy when considering the three marker genes for unstimulated macrophages at steady state (0 h) and stimulus-response time points (1, 3, and 8 h) or for scGETs' stimulus-response dynamics/dynamical features (Integral, Peak Amp, Max LFC, Speed). Classification accuracy is greater for dynamical features than for either expression at single time points or expression at steady state (0 h).

steady state vs. their LPS response Integral revealed that the scGET Integral distinguished all three macrophage polarization states more distinctly (Figure 7A). On average, *Nos2* was expressed more highly in M1 macrophages, but single-cell expression distributions of *Nos2* overlapped across polarization states, such that not all single cells could be identified as M1-based one time point alone (Figure 7B). A classifier for polarization state on these three marker genes showed that the trajectory features, especially Integral, Peak Amp, or Max LFC, drastically improved polarization state identification (higher F1 score) over steady-state values or responses measured at a single time point (Figure 7C).

We asked whether other genes at steady-state distinguished polarization states better than the canonical marker genes. Based on MI analysis, the top three genes that individually best distinguished unstimulated M0, M1(IFN-γ), and M2(IL-4) macrophages were *Irf1*, *Fgl2*, and *Tgtp1* (Figure S7A). While steady-state expression values poorly distinguished M0 and M2(IL-4) macrophages, the LPS response Integral of the same three genes perfectly distinguished the three polarization states (Figure S7B), pointing to the importance of response dynamics of genes encoding TFs (*Irf1*) or secreted proteins (*Fgl2*) in also specifying cell functional states.

To examine the relative number of genes needed to explain differences in polarization states, we next built a multinomial least absolute shrinkage and selection operator (LASSO)-penalized regression model using either steady-state or stimulus-response data. We found that performing LASSO-regularized regression on steady-state expression values (0 h) resulted in a model containing ∼150 genes, those built on single time-point response values (1, 3, and 8 h) ∼50–120 genes, and those built using a trajectory feature such as Integral only ∼25–30 genes (Figure S7C), again indicating the increased information content of GETs in specifying single-cell functional states.

DISCUSSION

To study macrophage SRS we present here an integrated experimental and computational approach to impute an ensemble of scGETs from time series scRNA-seq data. Stimulus-response gene expression is dynamic and heterogeneous, but current single-cell RNA measurement technology has not been able to

capture both characteristics simultaneously, as scRNA-seq is destructive to cells. Imputing scGETs from high-resolution time series data revealed the heterogeneity of dynamic responses of macrophages, thus quantifying macrophage SRS more completely. When considering dynamical expression features, SRS was calculated to be much higher, and scGETs were found to be both dramatically and subtly altered by the microenvironmental context of polarizing cytokines, indeed so distinctive that response trajectories were much more informative of cell functional state than steady-state mRNA abundances that are typically measured by scRNA-seq.

We developed the scGET imputation method to quantify SRS when considering single-cell gene expression response dynamics in a manner that is not possible with previously reported methods (Table S1). For specific signaling proteins, methods for imputing single-cell trajectories of protein activity have involved fitting mechanistic models of signaling networks to measured time series distributions,[47–51] but models of gene expression remain coarse and insufficiently precise. Further, the gene expression data consist of hundreds of data points per cell from many genes, posing additional challenges in fitting mechanistic models but enabling a statistical approach for constructing a transition matrix. Experimentally, the primary methodology for measuring scGETs is the MS2 reporter, which allows for live-cell imaging of fluorescently tagged mRNA transcripts. However, this technology only images one or two genes at a time, precluding analysis of correlations or biological importance of genes expressed in combination. In addition, it requires genetic engineering of the genes of interest, which necessitates cells that are immortal or have long lifespans (as opposed to primary macrophages) and may affect chromatin-determined expression dynamics.

As cells must adapt to changes in their environment, SRS is an important biological characteristic, particularly in immune sentinel cells such as macrophages. Several studies have investigated the stimulus information transmitted in biochemical signaling networks through live-cell imaging of kinase activities or transcription factor nuclear translocation, including NF-κB in macrophages. The trajectory vector of NF-κB signaling was shown to distinguish different doses of ligands much more accurately than single time points of NF-κB activity, and dynamical features of the NF-κB trajectories were identified to best distinguish different ligands,[14] analogous to what we found with gene expression.[13,52] However, estimations of stimulus information within the dynamics of a single transcription factor like NF-κB were consistently lower than information within the dynamics of the most informative genes in our study. That may not be surprising as many of the GETs we examined are the outcome of multiple non-redundant signaling pathways that operate on individual genes with different timescales, such as *Tnf* (NF-κB|p38) or *Cxcl10* (NF-κB|IRF). In addition, stimulation induces hundreds of individual genes per cell, enabling identification of multi-gene dynamics that together can result in high SRS, while measuring the trajectories of more than a few signaling proteins simultaneously in a single cell has not been feasible.[53]

A hallmark of healthy macrophages is SRS as a function of microenvironmental context and the history of exposure.[8] Macrophage stimulus-response gene expression dynamics are determined by (1) the activities of upstream signaling effectors (e.g., transcription factors NF-κB, IRF, AP1), whose abundances and nuclear translocation dynamics are strongly affected by contextual cytokines,[54–56] and (2) the mechanisms and molecules that interpret information within transcription factor activation, including chromatin opening mechanisms, nucleosome dynamics, and histone modifications,[30,57,58] which can also be significantly altered by contextual cytokines that either prime or repress gene regulatory elements. For example, an IFN-γ polarization context not only increases the nuclear availability of NF-κB[16,54,55] but also alters the epigenomic landscape of open chromatin and histone modifications.[59–61] Thus, the cell's response to stimuli not only encodes information about the stimulus but also about the cell's functional state.

Why might the functional state of a cell be predicted more accurately when accounting for dynamical features of stimulus-response expression trajectories rather than steady-state gene expression? Conceptually, response dynamics are more informative than steady state because the functional state of a cell is not only defined by abundances of molecules but also by the rate constants that determine synthesis, degradation, association, dissociation, and catalysis. For example, two cells may have exactly the same abundance of a ligand receptor, but if one cell has higher synthesis and degradation of the receptor than the other, they would be in different cell functional states, and the two would respond with different gene expression dynamics to the same ligand concentration.[15,62] Measuring stimulus responses reveals both the kinetic and abundance information of a cell's functional state at the time of stimulation, including availability of receptors, activities of kinases, and dynamics and localization of transcription factors. Capturing responses at the level of gene expression dynamics also captures the cell's chromatin state.

In this study, we focused on the dynamics of single-cell-inducible gene expression in macrophages, key to the initiation of immune responses. Response trajectories were critical for revealing the high SRS that is essential to macrophage function and supplied the kinetic information necessary for defining cell functional states. Future applications of defining cell states by stimulus-response trajectories could allow more robust clustering and subtyping for macrophages derived from undefined cytokine environments of inflammatory diseases, or they could enable the identification of rare subsets of single cells in other diseases where the function of outlier cells is pivotal, such as the responses of individual cancer cells to drugs or cell damage signals.

Limitations of the study

This study focuses on imputing rather than measuring scGETs. That is because there is no reliable technology to measure multiple scGETs in primary immune cells. In developing and evaluating the imputation method, we identified several limitations. First, we developed the experimental and computational pipeline for the particular biological application of innate immune cells responding to a defined stimulus, to quantify their SRS. Substantial biological knowledge is available about this system, including which genes are induced, the duration of the response, and population-level trajectories of gene expression, all of which

informed the experimental design and the imputation method. The described approach may not work well for other biological systems where less prior knowledge is available. Second, while we quality controlled the method on a defined population of cells all stimulated at the same time, it may not perform well on samples with mixed cell populations or cells receiving different or differently timed stimuli. Third, when the imputation method relies on a limited number of measured points, the imputed scGETs are generally less complex than the ground truth. In consideration of this, we used finer time points during the early phase when gene expression levels change more rapidly, and we decomposed scGETs into trajectory features for quantifying SRS.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Alexander Hoffmann (ahoffmann@ucla.edu).

Materials availability

This study did not generate new materials.

Data and code availability

- Data have been deposited to Zenodo and GEO (GEO: GSE224518) and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- All original code has been deposited to the GitHub repository https://github.com/KSheu/scResponseDynamics and is publicly available as of the date of publication. The DOI of an archived version is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

K.M.S. was supported by the UCLA Medical Scientist Training Program (T32-GM008042) and Systems and Integrative Biology Training Grant (T32-GM008185). The work was funded by R01AI127864 and R01GM12507 to A.H. Sequencing data from this paper were collected with the services of the UCLA Technology Center for Genomics and Bioinformatics Sequencing Core.

AUTHOR CONTRIBUTIONS

K.M.S. and A.H. conceptualized the study. K.M.S. designed and conducted all experiments. K.M.S. developed and implemented the analysis methods. A.P. assisted with data analysis. K.M.S. and A.H. wrote the manuscript. A.H. provided supervision and funding.

DECLARATION OF INTERESTS

K.M.S. and A.H. are authors of a patent (publication number 20230408493) that relates to the content of this manuscript.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Macrophage cell culture
- METHOD DETAILS
  - Macrophage scRNA-seq
  - Gene panel selection
  - scRNA-seq processing and analysis
  - Mathematical modeling
  - scGETs imputation algorithm
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - scGETs imputation applied to model-simulated trajectories
  - scGETs imputation method applied to macrophage scRNA-seq data
  - Calculation of trajectory features
  - Comparing scGETs imputed from timeseries data to scGETs from model-simulated ground truth
  - Calculation of maximum mutual information
  - Identification of optimized gene combinations
  - Statistical learning models to predict polarization state

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.09.023.

Received: November 3, 2023
Revised: July 5, 2024
Accepted: September 19, 2024
Published: October 15, 2024

REFERENCES

1. Wynn, T.A., Chawla, A., and Pollard, J.W. (2013). Macrophage biology in development, homeostasis and disease. Nature *496*, 445–455. https://doi.org/10.1038/nature12034.
2. Murray, P.J., and Wynn, T.A. (2011). Protective and pathogenic functions of macrophage subsets. Nat. Rev. Immunol. *11*, 723–737. https://doi.org/10.1038/nri3073.
3. Hao, S., and Baltimore, D. (2009). The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. Nat. Immunol. *10*, 281–288. https://doi.org/10.1038/ni.1699.
4. Smale, S.T., and Natoli, G. (2014). Transcriptional control of inflammatory responses. Cold Spring Harb. Perspect. Biol. *6*, a016261. https://doi.org/10.1101/cshperspect.a016261.
5. Hoffmann, A., Levchenko, A., Scott, M.L., and Baltimore, D. (2002). The IκB-NF-κB Signaling Module: Temporal Control and Selective Gene Activation. Science *298*, 1241–1245. https://doi.org/10.1126/science.1071914.
6. Rodríguez-Morales, P., and Franklin, R.A. (2023). Macrophage phenotypes and functions: resolving inflammation and restoring homeostasis. Trends Immunol. *44*, 986–998. https://doi.org/10.1016/j.it.2023.10.004.
7. Sheu, K.M., Luecke, S., and Hoffmann, A. (2019). Stimulus-specificity in the responses of immune sentinel cells. Curr. Opin. Syst. Biol. *18*, 53–61. https://doi.org/10.1016/j.coisb.2019.10.011.
8. Sheu, K.M., and Hoffmann, A. (2022). Functional Hallmarks of Healthy Macrophage Responses: Their Regulatory Basis and Disease Relevance. Annu. Rev. Immunol. *40*, 295–321. https://doi.org/10.1146/annurev-immunol-101320-031555.
9. Nau, G.J., Richmond, J.F.L., Schlesinger, A., Jennings, E.G., Lander, E.S., and Young, R.A. (2002). Human macrophage activation programs induced by bacterial pathogens. Proc. Natl. Acad. Sci. USA *99*, 1503–1508. https://doi.org/10.1073/pnas.022649799.
10. Cheng, C.S., Behar, M.S., Suryawanshi, G.W., Feldman, K.E., Spreafico, R., and Hoffmann, A. (2017). Iterative Modeling Reveals Evidence of Sequential Transcriptional Control Mechanisms. Cell Syst. *4*, 330–343.e5. https://doi.org/10.1016/j.cels.2017.01.012.
11. Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. (2009). Unbiased Reconstruction of a Mammalian Transcriptional Network Mediating

Pathogen Responses. Science *326*, 257–263. https://doi.org/10.1126/science.1179050.

12. Sheu, K.M., Guru, A.A., and Hoffmann, A. (2023). Quantifying stimulus-response specificity to probe the functional state of macrophages. Cell Syst. *14*, 180–195.e5. https://doi.org/10.1016/j.cels.2022.12.012.
13. Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. (2014). Systems biology. Accurate information transmission through dynamic biochemical signaling networks. Science *346*, 1370–1373. https://doi.org/10.1126/science.1254933.
14. Adelaja, A., Taylor, B., Sheu, K.M., Liu, Y., Luecke, S., and Hoffmann, A. (2021). Six distinct NFκB signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. Immunity *54*, 916–930.e7. https://doi.org/10.1016/j.immuni.2021.04.011.
15. Luecke, S., Sheu, K.M., and Hoffmann, A. (2021). Stimulus-specific responses in innate immunity: Multilayered regulatory circuits. Immunity *54*, 1915–1932. https://doi.org/10.1016/j.immuni.2021.08.018.
16. Singh, A., Sen, S., Iter, M., Adelaja, A., Luecke, S., Guo, X., and Hoffmann, A. (2024). Stimulus-response signaling dynamics characterize macrophage polarization states. Cell Syst. *15*, 563–577.e6. https://doi.org/10.1016/j.cels.2024.05.002.
17. Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., and Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. eLife *8*, e41769. https://doi.org/10.7554/eLife.41769.
18. Cawte, A.D., Unrau, P.J., and Rueda, D.S. (2020). Live cell imaging of single RNA molecules with fluorogenic Mango II arrays. Nat. Commun. *11*, 1283. https://doi.org/10.1038/s41467-020-14932-7.
19. Donovan, B.T., Huynh, A., Ball, D.A., Patel, H.P., Poirier, M.G., Larson, D.R., Ferguson, M.L., and Lenstra, T.L. (2019). Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. EMBO J. *38*, e100809. https://doi.org/10.15252/embj.2018100809.
20. Forero-Quintero, L.S., Raymond, W., Handa, T., Saxton, M.N., Morisaki, T., Kimura, H., Bertrand, E., Munsky, B., and Stasevich, T.J. (2021). Live-cell imaging reveals the spatiotemporal organization of endogenous RNA polymerase II phosphorylation at a single gene. Nat. Commun. *12*, 3158. https://doi.org/10.1038/s41467-021-23417-0.
21. Hao, N., and O'Shea, E.K. (2011). Signal-dependent dynamics of transcription factor translocation controls gene expression. Nat. Struct. Mol. Biol. *19*, 31–39. https://doi.org/10.1038/nsmb.2192.
22. Chen, W., Guillaume-Gentil, O., Rainer, P.Y., Gäbelein, C.G., Saelens, W., Gardeux, V., Klaeger, A., Dainese, R., Zachara, M., Zambelli, T., et al. (2022). Live-seq enables temporal transcriptomic recording of single cells. Nature *608*, 733–740. https://doi.org/10.1038/s41586-022-05046-9.
23. Cao, J., Zhou, W., Steemers, F., Trapnell, C., and Shendure, J. (2020). Sci-fate characterizes the dynamics of gene expression in single cells. Nat. Biotechnol. *38*, 980–988. https://doi.org/10.1038/s41587-020-0480-9.
24. Calvanese, V., Capellera-Garcia, S., Ma, F., Fares, I., Liebscher, S., Ng, E.S., Ekstrand, S., Aguadé-Gorgorió, J., Vavilina, A., Lefaudeux, D., et al. (2022). MAPPING HUMAN HAEMATOPOIETIC STEM CELLS FROM HAEMOGENIC ENDOTHELIUM TO BIRTH. Nature *604*, 534–540. https://doi.org/10.1038/s41586-022-04571-x.
25. Alpert, A., Nahman, O., Starosvetsky, E., Hayun, M., Curiel, T.J., Ofran, Y., and Shen-Orr, S.S. (2022). Alignment of single-cell trajectories by tuMap enables high-resolution quantitative comparison of cancer samples. Cell Syst. *13*, 71–82.e8. https://doi.org/10.1016/j.cels.2021.09.003.
26. Stephenson, E., Reynolds, G., Botting, R.A., Calero-Nieto, F.J., Morgan, M.D., Tuong, Z.K., Bach, K., Sungnak, W., Worlock, K.B., Yoshida, M., et al. (2021). Single-cell multi-omics analysis of the immune response in COVID-19. Nat. Med. *27*, 904–916. https://doi.org/10.1038/s41591-021-01329-2.
27. Alpert, A., Moore, L.S., Dubovik, T., and Shen-Orr, S.S. (2018). Alignment of single-cell trajectories to compare cellular expression dynamics. Nat. Methods *15*, 267–270. https://doi.org/10.1038/nmeth.4628.
28. Regot, S., Hughey, J.J., Bajar, B.T., Carrasco, S., and Covert, M.W. (2014). High-sensitivity measurements of multiple kinase activities in live single cells. Cell *157*, 1724–1734. https://doi.org/10.1016/j.cell.2014.04.039.
29. Batchelor, E., Loewer, A., Mock, C., and Lahav, G. (2011). Stimulus-dependent dynamics of p53 in single cells. Mol. Syst. Biol. *7*, 488. https://doi.org/10.1038/msb.2011.20.
30. Cheng, Q.J., Ohta, S., Sheu, K.M., Spreafico, R., Adelaja, A., Taylor, B., and Hoffmann, A. (2021). NF-κB dynamics determine the stimulus specificity of epigenomic reprogramming in macrophages. Science *372*, 1349–1353. https://doi.org/10.1126/science.abc0269.
31. Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., and Covert, M.W. (2010). Single-cell NF-κB dynamics reveal digital activation and analogue information processing. Nature *466*, 267–271. https://doi.org/10.1038/nature09145.
32. Tong, A.J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. Cell *165*, 165–179. https://doi.org/10.1016/j.cell.2016.01.020.
33. Wang, N., Lefaudeux, D., Mazumder, A., Li, J.J., and Hoffmann, A. (2021). Identifying the combinatorial control of signal-dependent transcription factors. PLoS Comput. Biol. *17*, e1009095. https://doi.org/10.1371/journal.pcbi.1009095.
34. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386. https://doi.org/10.1038/nbt.2859.
35. La Manno, G.L., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.
36. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982. https://doi.org/10.1038/nmeth.4402.
37. Song, Q., Wang, J., and Bar-Joseph, Z. (2022). scSTEM: clustering pseudotime ordered single-cell data. Genome Biol. *23*, 150. https://doi.org/10.1186/s13059-022-02716-9.
38. Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. J. Am. Stat. Assoc. *74*, 829–836. https://doi.org/10.1080/01621459.1979.10481038.
39. Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., and Schmid, M. (2019). A review of spline function procedures in R. BMC Med. Res. Methodol. *19*, 46. https://doi.org/10.1186/s12874-019-0666-3.
40. Fritsch, F.N., and Carlson, R.E. (1978). Piecewise Cubic Interpolation Methods (California University, Lawrence Livermore Laboratory).
41. Cannoodt, R., Saelens, W., Sichien, D., Tavernier, S., Janssens, S., Guilliams, M., Lambrecht, B., Preter, K.D., and Saeys, Y. (2016). SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. Preprint at bioRxiv. https://doi.org/10.1101/079509.
42. Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. Nat. Biotechnol. *37*, 451–460. https://doi.org/10.1038/s41587-019-0068-4.
43. Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. *37*, 547–554. https://doi.org/10.1038/s41587-019-0071-9.
44. Tang, Y., Adelaja, A., Ye, F.X.-F., Deeds, E., Wollman, R., and Hoffmann, A. (2021). Quantifying information accumulation encoded in the dynamics of biochemical signaling. Nat. Commun. *12*, 1272. https://doi.org/10.1038/s41467-021-21562-0.

45. Stairiker, C.J., van Meurs, M., Leon, L.G., Brouwers-Haspels, A.A., Rijsbergen, L., Mueller, Y.M., and Katsikis, P.D. (2020). Heatr9 is an infection responsive gene that affects cytokine production in alveolar epithelial cells. PLoS One *15*, e0236195. https://doi.org/10.1371/journal.pone.0236195.

46. Garcia, G., Irudayam, J.I., Jeyachandran, A.V., Dubey, S., Chang, C., Castillo Cario, S., Price, N., Arumugam, S., Marquez, A.L., Shah, A., et al. (2023). Innate immune pathway modulator screen identifies STING pathway activation as a strategy to inhibit multiple families of arbo and respiratory viruses. Cell Rep. Med. *4*, 101024. https://doi.org/10.1016/j.xcrm.2023.101024.

47. Dixit, P.D., Lyashenko, E., Niepel, M., and Vitkup, D. (2019). Maximum Entropy Framework for Predictive Inference of Cell Population Heterogeneity and Responses in Signaling Networks. Cell Syst. *10*, 204–212.e8. https://doi.org/10.1016/j.cels.2019.11.010.

48. Hasenauer, J., Waldherr, S., Doszczak, M., Radde, N., Scheurich, P., and Allgöwer, F. (2011). Identification of models of heterogeneous cell populations from population snapshot data. BMC Bioinformatics *12*, 125. https://doi.org/10.1186/1471-2105-12-125.

49. Hasenauer, J., Hasenauer, C., Hucho, T., and Theis, F.J. (2014). ODE Constrained Mixture Modelling: A Method for Unraveling Subpopulation Structures and Dynamics. PLoS Comput. Biol. *10*, e1003686. https://doi.org/10.1371/journal.pcbi.1003686.

50. Loos, C., and Hasenauer, J. (2019). Mathematical modeling of variability in intracellular signaling. Curr. Opin. Syst. Biol. *16*, 17–24. https://doi.org/10.1016/j.coisb.2019.10.020.

51. Loos, C., Moeller, K., Fröhlich, F., Hucho, T., and Hasenauer, J. (2018). A Hierarchical, Data-Driven Approach to Modeling Single-Cell Populations Predicts Latent Causes of Cell-To-Cell Variability. Cell Syst. *6*, 593–603.e13. https://doi.org/10.1016/j.cels.2018.04.008.

52. Cheong, R., Rhee, A., Wang, C.J., Nemenman, I., and Levchenko, A. (2011). Information Transduction Capacity of Noisy Biochemical Signaling Networks. Science *334*, 354–358. https://doi.org/10.1126/science.1204553.

53. Naigles, B., Narla, A.V., Soroczynski, J., Tsimring, L.S., and Hao, N. (2023). Quantifying dynamic pro-inflammatory gene expression and heterogeneity in single macrophage cells. J. Biol. Chem. *299*, 105230. https://doi.org/10.1016/j.jbc.2023.105230.

54. Mitchell, S., Mercado, E.L., Adelaja, A., Ho, J.Q., Cheng, Q.J., Ghosh, G., and Hoffmann, A. (2019). An NFκB Activity Calculator to Delineate Signaling Crosstalk: Type I and II Interferons Enhance NFκB via Distinct Mechanisms. Front. Immunol. *10*, 1425. https://doi.org/10.3389/fimmu.2019.01425.

55. Adelaja, A., and Hoffmann, A. (2019). Signaling Crosstalk Mechanisms That May Fine-Tune Pathogen-Responsive NFκB. Front. Immunol. *10*, 433. https://doi.org/10.3389/fimmu.2019.00433.

56. Begitt, A., Droescher, M., Meyer, T., Schmid, C.D., Baker, M., Antunes, F., Knobeloch, K.-P., Owen, M.R., Naumann, R., Decker, T., et al. (2014). STAT1-cooperative DNA binding distinguishes type 1 from type 2 interferon signaling. Nat. Immunol. *15*, 168–176. https://doi.org/10.1038/ni.2794.

57. Kang, K., Park, S.H., Chen, J., Qiao, Y., Giannopoulou, E., Berg, K., Hanidu, A., Li, J., Nabozny, G., Kang, K., et al. (2017). Interferon-γ Represses M2 Gene Expression in Human Macrophages by Disassembling Enhancers Bound by the Transcription Factor MAF. Immunity *47*, 235–250.e4. https://doi.org/10.1016/j.immuni.2017.07.017.

58. Kim, J., Sheu, K.M., Cheng, Q.J., Hoffmann, A., and Enciso, G. (2022). Stochastic models of nucleosome dynamics reveal regulatory rules of stimulus-induced epigenome remodeling. Cell Rep. *40*, 111076. https://doi.org/10.1016/j.celrep.2022.111076.

59. Qiao, Y., Giannopoulou, E.G., Chan, C.H., Park, S.-H., Gong, S., Chen, J., Hu, X., Elemento, O., and Ivashkiv, L.B. (2013). Synergistic activation of inflammatory cytokine genes by interferon-γ-induced chromatin remodeling and toll-like receptor signaling. Immunity *39*, 454–469. https://doi.org/10.1016/j.immuni.2013.08.009.

60. Kang, K., Bachu, M., Park, S.H., Kang, K., Bae, S., Park-Min, K.-H., and Ivashkiv, L.B. (2019). IFN-γ selectively suppresses a subset of TLR4-activated genes and enhancers to potentiate macrophage activation. Nat. Commun. *10*, 3320. https://doi.org/10.1038/s41467-019-11147-3.

61. Ivashkiv, L.B. (2018). IFNγ: signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy. Nat. Rev. Immunol. *18*, 545–558. https://doi.org/10.1038/s41577-018-0029-z.

62. Becker, V., Schilling, M., Bachmann, J., Baumann, U., Raue, A., Maiwald, T., Timmer, J., and Klingmüller, U. (2010). Covering a Broad Dynamic Range: Information Processing at the Erythropoietin Receptor. Science *328*, 1404–1408. https://doi.org/10.1126/science.1184913.

63. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software *28*, 1–26. https://doi.org/10.18637/jss.v028.i05.

64. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

65. Shum, E.Y., Walczak, E.M., Chang, C., and Christina Fan, H. (2019). Quantitation of mRNA Transcripts and Proteins Using the BD Rhapsody™ Single-Cell Analysis System. Adv. Exp. Med. Biol. *1129*, 63–79. https://doi.org/10.1007/978-981-13-6037-4_5.

66. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

67. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

68. Jetka, T., Nienałtowski, K., Winarski, T., Błoński, S., and Komorowski, M. (2019). Information-theoretic analysis of multivariate single-cell signaling responses. PLoS Comput Biol. *15*, e1007132. https://doi.org/10.1371/journal.pcbi.1007132.

69. Li, J., Bien, J., and Wells, M.T. (2018). rTensor: An R Package for Multidimensional Array (Tensor) Unfolding, Multiplication, and Decomposition. Journal of Statistical Software *87*, 1–31. https://doi.org/10.18637/jss.v087.i10.

70. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS. A Journal of Integrative Biology *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

71. Lin, L., Song, M., Jiang, Y., Zhao, X., Wang, H., and Zhang, L. (2020). Normalizing single-cell RNA sequencing data with internal spike-in-like genes. NAR Genomics and Bioinformatics *2*. https://doi.org/10.1093/nargab/lqaa059.

72. Sen, S., Cheng, Z., Sheu, K.M., Chen, Y.H., and Hoffmann, A. (2020). Gene Regulatory Strategies that Decode the Duration of NFκB Dynamics Contribute to LPS- versus TNF-Specific Gene Expression. Cell Syst. *10*, 169–182.e5. https://doi.org/10.1016/j.cels.2019.12.004.

73. Lugowski, A., Nicholson, B., and Rissland, O.S. (2018). Determining mRNA half-lives on a transcriptome-wide scale. Methods *137*, 90–98. https://doi.org/10.1016/j.ymeth.2017.12.006.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-mouse CD45-hashtag antibodies	BD Rhapsody	633793
Chemicals, peptides, and recombinant proteins		
LPS	Sigma, B5:055	L2880
murine TNF	Roche	11271156001
Pam3CSK4	Invivogen	tlrl-pms
low MW polyinosine-polycytidylic acid (Poly(I:C))	Invivogen	tlrl-picw
synthetic CpG ODN 1668	Invivogen	tlrl-1668
murine IFNβ	R&D	12401-1
murine IFNγ	R&D	485-MI
murine IL-4	R&D	404-ML
Critical commercial assays		
BD Rhapsody Express Single-Cell Analysis system	BD	633702
Targeted mRNA and AbSeq Reagent Kit 4 Pack	BD	633771
BD Rhapsody Cartridge Reagent Kit	BD	633731
BD Rhapsody Cartridge Kit	BD	633733
BD Rhapsody cDNA Kit	BD	633773
BD Rhapsody P5000M pipette	BD	633705
BD™ Stain Buffer (FBS)	BD Pharmigen	554656
Deposited data		
BD Rhapsody scRNA-seq	This paper	GSE224518
BMDM bulk RNA-seq data	Cheng et al.[10]	GSE68318
Experimental models: Cell lines		
Immortalized Myeloid Progenitor-derived macrophages	Singh et al.[16]	N/A
Oligonucleotides		
Rhapsody Custom Panel: ID 1330	BD	633743
Rhapsody Custom Panel: ID 1331	BD	633743
Rhapsody Custom Panel: ID 1332	BD	633743
Rhapsody Custom Panel: ID 1334	BD	633743
Rhapsody Custom Panel: ID 1341	BD	633743
Software and algorithms		
CARET	Kuhn et al.[63]	http://caret.r-forge.r-project.org/
Seurat	Stuart et al.[64]	https://www.rdocumentation.org/packages/Seurat/versions/3.1.4
BD Rhapsody Targeted Analysis Pipeline (version v1.0)	Shum et al.[65]	https://www.sevenbridges.com/
edgeR	Robinson et al.[66]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
HOMER	Heinz et al.[67]	http://homer.ucsd.edu/homer/
SLEMI	Jetka et al.[68]	https://cran.r-project.org/web/packages/SLEMI/index.html
rtensor	Li et al.[69]	https://cran.r-project.org/web/packages/rTensor/index.html
clusterProfiler	Yu et al.[70]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ISnorm	Lin et al.[71]	PMID 33575610, PMC7671304
Analysis code	This paper	https://github.com/KSheu/scResponseDynamics; Zenodo: https://doi.org/10.5281/zenodo.13370505

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Macrophage cell culture

Macrophages were generated by differentiating HoxB4-transduced myeloid precursors (hMPs) to HoxB4-transduced myeloid precursor-derived macrophages (hMPDMs), as previously described and characterized.[12,16] hMPs were initially thawed and expanded in 10cm non-adherent petri dishes for 3 days. Cells were then washed once and placed in differentiation media for a total of 10 days (DMEM/10% FBS, 30% L929 supernatant, 1% PS, 1% L-Glut, b-Me (1:1000)), with replating onto 6cm plates with new media on day 7, at a density of ~20k cells/cm^2. For polarized macrophages, cells were incubated in 50ng/ml IFNγ or 50ng/ml IL4 for 24 hours prior to stimulation on day 10. On day 10, macrophages were stimulated with one of six stimuli: 100ng/mL lipopolysaccharide (LPS, Sigma Aldrich B5:055 #L2880), 10ng/mL murine TNF (Roche #11271156001), 50μg/mL low molecular weight polyinosine-polycytidylic acid (Poly(I:C), Invivogen tlrl-picw), 100ng/mL Pam3CSK4 (P3C, Invivogen tlrl-pms), 100nM synthetic CpG ODN 1668 (CpG, Invivogen tlrl-1668), 500U/ml IFNβ (R&D #12401-1), or media only Untreated control. Samples were collected for scRNA-seq prior to stimulation, 0hrs, and at multiple timepoints post-stimulation, including 15 or 30 minutes, 1hr, 3hrs, 5hrs, and 8hrs.

METHOD DETAILS

Macrophage scRNA-seq

To collect the adherent macrophages for scRNA-seq using the BD Rhapsody platform, macrophage cells were washed 1x with cold PBS, then lifted into suspension by incubating at 37C for 5 minutes with Accutase, which resulted in cell viability typically >85%. Cells were centrifuged at 4C, 400g for 5 minutes, and resuspended in PBS + 2% FBS. Cells were hash-tagged with anti-CD45-hashtags (BD Rhapsody # 633793) and loaded onto the cartridge according to manufacturer's instructions (BD Rhapsody # 633771), with the following modifications, which helped ensure sufficient cell viability for the subsequent steps: Incubation with hashtags was performed for 30mins on ice, instead of 20mins at room temperature; only two washes were performed after hashtag incubation to minimize cell loss. Each cartridge was then loaded with a total of ~36k cells across 12 hash-tagged samples (~3k cells/sample). A set of 500 custom-designed primers were generated with the services of BD Rhapsody (Rhapsody Custom Panel: ID 1330, 1331, 1332, 1334, 1341), to preferentially amplify the gene set of interest for single-cell sequencing.[12] Libraries were prepared using the custom primer set according to manufacturer's instructions (BD Rhapsody # 633771) and sequenced 2x100 on Novaseq 6000.

Gene panel selection

To select genes for single-cell targeted gene profiling, we used a previously published algorithmic approach.[12] Briefly, we analyzed existing bulk transcriptomic profiling of macrophage responses (GSE68318)[10], to determine macrophage inducible genes across 14 stimulus conditions and 4 timepoints. From this set of ~1500 induced genes, we used PCA across all time points for the 14 stimuli in the dataset. The loadings matrix obtained from the PCA was used to calculate a rank score for each gene: $score_j = \sum_{x=1}^{20}(PCx_j)^2$, where PCx_j is the component x loadings value for gene j. The top 480 ranked genes were included in the panel, and the remaining 20 genes were manually selected to add genes such as cell type markers, macrophage polarization markers, and transcription factors.

scRNA-seq processing and analysis

Raw fastq files were processed using the BD Rhapsody™ Targeted Analysis Pipeline (version v1.0)[63] hosted on Seven Bridges Genomics. Distribution-Based Error Correction (DBEC)-adjusted UMI counts (molecules per cell) were used in the downstream analysis. Multiplets, cells with undetermined barcodes, and cells with less than 80 features were removed from the analysis. Because the selected 500-gene panel was comprised of largely inducible genes, the assumption that the total number of RNAs per cell is constant does not hold. Counts were therefore normalized using the package *Isnorm*,[64] rather than the more standard approach of dividing by total counts per cell. The internal-spike-in geneset calculated included 10 genes, among them a macrophage cell type gene *Adgre1*, further supporting the relevance of the selected spike-in geneset: *"Adgre1", "Alas1", "Casp4", "Gsr", "Ifrd1", "Nfkb1", "Ptafr", "Rela", "Sdc4", "Swap70"*. Bulk RNA-seq data from bone-marrow-derived macrophages was obtained from GSE68318 to compare it against time-series scRNA-seq pseudobulk data.

Mathematical modeling

Model equations

We used mechanistic models to simulate the five gene regulatory logics previously identified as regulating macrophage stimulus-response gene expression.[10] These gene regulatory mechanisms (GRMs) involve the dynamic activity of four key effector

proteins: AP1, NFκB, IRF, and MAPKp38. Three of the five GRMs involve only one transcription factor: AP1-only, NFκB-only, IRF-only, a fourth involves an NFκB-IRF OR gate, and the fifth an NFκB-p38 AND gate.[10]

To simulate dynamic gene expression resulting from each of the five GRMs, we solved ordinary differential equations that followed the same general form,

$$\frac{dmRNA}{\{dt\}} = k_{syn}f(t) - k_{deg} * mRNA,$$

with k_{syn} representing the gene synthesis rate constant, $f(t)$ representing time-varying transcription factor activity that affects synthesis rate, and k_{deg} representing the mRNA degradation rate that can be stabilized by time-varying MAPKp38 activity.

The function $f(t)$ was transcription factor dependent and governed by a Hill equation. For single transcription factor logic gates (AP1-only, NFκB-only, IRF-only),[10,33] $f(t)$ was written as:

$$f(t) = (1 - k_0)\frac{(K_{TF} * [TF(t)])^n}{1+(K_{TF} * [TF(t)])^n} + k_0,$$

where k_0 represents a small basal transcription rate in the absence of transcription factor activity, n represents the Hill coefficient, and $K_{TF} = 1/K_{D,TF}$, with K_D representing the dissociation constant of the transcription factor. As K_D decreases, the transcription factor's binding affinity K_{TF} increases and thus mRNA synthesis increases. $TF(t)$ is the time-dependent transcription factor activity measured over 8 hours.

Using the same notation, the logical OR gate that involved two transcription factors (i.e. NFκB **OR** IRF)[10,33] was written as,

$$f(t) = (1 - k_0)\frac{(K_{TF_1} * TF1(t))^n + (K_{TF_2} * TF2(t))^n + (K_{TF_1} * K_{TF_2} * TF1(t) * TF2(t))^n}{1+(K_{TF_1} * TF1(t))^n + (K_{TF_2} * TF2(t))^n + (K_{TF_1} * K_{TF_2} * TF1(t) * TF2(t))^n)} + k_0$$

For the logical AND gate where p38 controls mRNA half-life (i.e. NFκB **AND** p38),[10] we modified the degradation rate by extending the mRNA half-life as an exponential function of p38 activity levels. Deriving the expression for exponential decay, the decay constant λ is written as a function of half-life $t_{1/2}$, $\lambda = k_{deg} = ln(2)/t_{1/2}$. The decay rate is affected by two sources, the basal mRNA degradation rate, and the added stabilization from p38 activity. k_{deg} can thus be written as,

$$k_{deg} = \ln(2) \Big/ \left(t_{1/2} + t_{1/2}^{max} * p38\right)$$

where $t_{1/2}$ represented the half-life of the mRNA regardless of p38 activity, and $t_{1/2}^{max}$ represented the maximum half-life extension (8hrs) provided by dynamic p38 activity. p38 activity was modeled as a percentage of maximum p38 activity, such that the half-life extension provided by p38 varied between 0 and $t_{1/2}^{max}$.

Model simulations

1. *Model parameters*: Three parameters were varied in the simulations to produce genes/gene promoters with unique characteristics: the Hill coefficient of transcription factor binding (n), the dissociation constant of transcription factor binding (K_D), and the mRNA degradation rate (k_{deg}). The Hill coefficient was set to either 1 or 3, the K_D was varied with reference to the K_D of the averaged gene for each GRM, over two orders of magnitude from 0.25x to 4x, and the k_{deg} was varied to equate half-lives from 30 mins to 5hrs. The parameters previously fit to the average of clusters corresponding to each GRM were n =1, $k_{deg} = ln(2)/30$, $K_{TF,AP1} = 1/K_{D,AP1} = 0.48$, $K_{TF,NF\kappa B} = 1/K_{D,NF\kappa B} = 1.3$, $K_{TF,IRF} = 1/K_{D,IRF} = 1.25$[10]. All together the parameter combinations represented a total of 200 unique genes/gene promoters (2 Hill coefficient values, 5 K_D values, 4 k_{deg} values, 5 GRMs).
2. *Generating a library of single-cell signaling inputs*: To generate single-cell signaling inputs for the gene expression model, we used previously published average measured dynamic activity of each of the effector proteins AP1, NFκB, IRF, and p38 over the 8hour timecourse.[10] Single cell heterogeneity was generated by random sampling from a zero-truncated normal distribution, using the measured timepoints along both the value and time axes as the mean. Interpolating over the sampled values at measured timepoints provided the single-cell signaling input trajectories. The full signaling profile of a single cell encompasses the activities of all four effectors (AP1, NFkB, IRF, p38) and was determined by random sampling of a trajectory from the library of dynamic activity profiles of all four signaling proteins.
3. *Running the model*: The model was simulated using the single-cell signaling activity inputs of 100 individual cells, resulting in dynamic gene expression profiles of 200 genes in every cell. In these simulations, the cell-to-cell heterogeneity of gene expression is derived from the heterogeneity in the signaling activity inputs for each cell. To simulate transcriptional or measurement noise observed in scRNA-seq data, we added a random jitter of up to 10% of the modeled gene expression value. Simulations were performed in R using the packages ''deSolve'' and ''doParallel''.

scGETs imputation algorithm

In imputing scGETs, we aimed to link cell archetypes from timepoint to timepoint, to estimate the gene expression dynamics within individual cells over a short time course, with the assumption of minimal division or cell differentiation into other cell types. The key

elements of the imputation method were designed to address several challenges in a computational approach to estimating transcriptome trajectories: 1) maintaining measured gene-gene correlation within single-cell trajectories across time, 2) linking cells within distributions to reflect biophysically plausible continuity of gene expression, and 3) using minimal computational resources/time.

1. Dimensionality Reduction: We performed Principal Component Analysis (PCA) on the time-series single-cell RNA-seq data. PCA was performed across all stimuli simultaneously, generating a lower dimensional data embedding that captured the gene expression information of individual cells. The initial data matrix can be written as

$$\boldsymbol{M} = N \times P,$$

where N is the total number of measured genes and P is the number of single cells, including all stimulus conditions and timepoints together. The resulting matrix decomposition $\boldsymbol{S} = \boldsymbol{W}^T \times \boldsymbol{M}$, where $\boldsymbol{S}$ is the $r \times P$ component scores matrix for all single cells, and $\boldsymbol{W}$ the $N \times r$ loadings matrix for all genes, embeds the gene-gene relationships for each cell within the loadings matrix.

2. k-means Clustering: A conceptual assumption of the method was that an ensemble of cell archetypes at each timepoint accounted for the behavior of individual cells. Clustering reduces sparsity, as a large degree of the sparsity observed in scRNA-seq data arises from technical noise. The choice of k is user-defined and dependent on the number of cells measured and the structure of the data. At any time-point, a cell is classified as a member of one of k archetypes, and principal component scores of each cluster is summarized by a consensus value – the mean or median of individual cell PC scores,

$$PC_{k,n} = mean(C_k)_n,$$

where C_k is the set of cells C bellowing to archetype k, and n is the principal component.

3. Weighted Random Walks: The probability of archetype connections across consecutively measured timepoints was weighted by Euclidean distance. Thus, the linkage probability from a cell archetype k_i at timepoint i, to cell archetype $k_{\{i+1\}}$ at timepoint $\{i+1\}$, for all archetype pairs can be written as

$$P = \frac{1}{\sqrt{\sum_{j=1}^{n} \left(k_i - k_{\{i+1\}}\right)^2}},$$

where n is the number of principal components. Larger distances between cell archetypes across timepoints resulted in lower probability of linkage. Random walks through archetypes were weighted by these probabilities, and the cell path defined as the archetype's consensus principal component values at each given timepoint.

4. Spline interpolation: For each set of cell archetypes linked across timepoints, spline interpolation is used to estimate values across unmeasured timepoints. This provides a true-time continuous linkage of the lower dimensional cell-level data, resulting in a trajectory for each cell in PC space, with gene expression trajectories for each cell embedded and recoverable from the original PCA loadings.
5. Recovering gene trajectories: Individual gene trajectories for each cell can be recovered from the cell trajectories by using the loadings matrix to re-rotate the lower dimensional cell-based data back to the higher-dimensional gene expression space. In other words,

$$\boldsymbol{S_{new}} = \boldsymbol{W}^T \times \boldsymbol{M_{new}},$$

where $\boldsymbol{S_{new}}$ contains cell trajectories represented as principal component scores, $\boldsymbol{W}$ the original $N \times r$ loadings matrix for all genes, and $\boldsymbol{M_{new}}$ the desired matrix of single-cell gene expression for all genes over true-time.

QUANTIFICATION AND STATISTICAL ANALYSIS

scGETs imputation applied to model-simulated trajectories

Selecting time points

Model-simulated trajectories contain the full single-cell gene expression trajectories (scGETs) (ground truth); we can apply and test the scGETs imputation method by selecting different timepoints and comparing the resulting reconstructions to the ground truth trajectory. We selected five front-loaded timepoints, 0, 0.5, 1, 3, 8 hrs, and compared them to five evenly spaced timepoints, 0, 2, 4, 6, 8 hrs. Ten, 20, and 50 evenly-spaced timepoints were also compared.

Dimensionality reduction

PCA was performed on the selected timepoints. For model-simulated trajectories, k-means clustering to obtain cell archetypes was not necessary, as every timepoint already contained the same number of individual cells.

Weighted random walks

Transition probability matrices were calculated using Euclidean distance between every pair of archetypes across consecutive timepoints. The number of principal components used to generate the transition probability matrices were compared, to assess the impact of recapitulating ground truth connections, including 5, 10, 20, and 50 components.

Spline interpolation

Spline interpolation was performed using several different methods, to compare each to the shape of the ground truth trajectories. The R function *smooth.spline* was used to test the interpolation of third-order splines with different degrees of freedom. Spline functions that were trialed included third-order splines with 3 vs 5 degrees of freedom, or third-order splines fitted using leave-one-out cross validation. Other smoothing methods compared were the monotonic splines fitted using the method "monoH.FC", which computes a monotone Hermite spline according to the method of Fritsch and Carlson, and loess smoothing, with span set to 0.75.

scGETs imputation method note

In imputing scGETs, we aimed to provide the least complex ensemble of trajectories that accounted for the data distributions at all measured time points. This reduced the effect of technical noise (e.g. drop-outs in scRNA-seq), but it may also reduce biological noise. The method first relied on PCA to compress the information from 500 genes per cell into a cell score, with gene-gene correlations maintained in the PCA loadings. To reduce the impact of technical noise and to decrease computational time, cells at each timepoint were binned into cell archetypes. While specifying too few archetypes may miss outlier behavior of rare cells, we noted that 20^5 (5 timepoints) equated 3.2 million possible paths, far exceeding the number of cells measured. Thus, leveraging cell archetypes reduced the impact of technical noise at each timepoint, but still maintained sufficient coverage of possible trajectories. Using weighted random walks to link cell archetypes over timepoints resulted in trajectories that were both data-driven and biochemically plausible, as no trajectories were imputed with rapidly oscillatory characteristics that contradicted knowledge of mRNA half-lives.[72,73] This probabilistic approach to linking data across timepoints also ensured that all permutations of paths were possible, but some were less likely given the data. Identifying the trajectory paths for the entire cell in the latent space and then recovering individual gene trajectories by multiplying the PCA loadings matrix also allowed cellular gene-gene correlations to be preserved, which would not be the case if trajectory paths were identified for each gene independently

scGETs imputation method applied to macrophage scRNA-seq data

Dimensionality reduction

PCA was performed on all timepoints and all stimulus-conditions combined, for each macrophage polarization state. Timepoints used for M0 macrophages were 0, 0.25, 1, 3, and 8hrs. Timepoints used for M1(IFNγ) and M(IL4) macrophages were 0, 0.5, 1, 3, 5, and 8hrs.

k-means clustering

To obtain cell archetypes at each timepoint, we utilized k-means clustering with $k = 20$, after examining the elbow plots derived from the data. Each archetype was defined by the median of PC scores of all cells belonging to that archetype. Note that with archetype numbers $k > 10$, linked combinatorially across five timepoints, the number of possible single cell trajectories is 10^5 = 100,000, exceeding the number of individual single cells measured per stimulus. The use of k-means clustering to group similar cells into representative archetypes thus both decreases computational time and can theoretically capture the full range of heterogeneous behavior of individual cell trajectories.

Weighted random walks

Transition probability matrices were calculated using Euclidean distance between archetypes over timepoints over 50 principal components. The cell paths were realized via random walks over timepoints, weighted by the values specified by the transition probability matrices. We used 1000 random walks per stimulus to approximate the number of cells measured in the data.

Spline interpolation

We used the R function *smooth.spline*, setting {$cv = T$} to fit a smoothing spline using leave-one-out cross-validation.

Calculation of trajectory features

For comparing stimulus-specificity, trajectories were first scaled 0-1 across all stimuli for each polarization state separately, and the trajectory features were calculated on the scaled trajectories. By doing so we focused on gene expression expressed not in absolute units, but in relative counts or fold-changes. Scaling maintained the same relative amplitudes across genes and increased the comparability of trajectory features. For comparisons across polarization states, trajectories from all stimuli and all polarization states were scaled 0-1 across all cells together. Features were calculated in R, with the following definitions:

- Integral: $\int_0^8 f(x)dx$, definite integral from 0 to 8 hours.
- Peak amplitude: max $f(x)$, maximum value over 8 hours.
- Max log fold change: $log_2(\max f(x)/f(0))$, log of the max value divided by value at t=0.
- Activation speed $f'(x)|_{x=1}$, the derivative at 1 hour.

Comparing scGETs imputed from timeseries data to scGETs from model-simulated ground truth

Weighted random walks

To compare paths chosen by 1) random connections vs. 2) weighted random walks, to 3) the ground truth connection of cells over timepoints, we calculated the mean Euclidean distance of the imputed path to the ground truth path, across all the timepoints. We calculated these distances across each possible pair of an imputed path to the ground truth path; as imputed paths and ground truth paths are not 1-1 matched, across all possible pairs, we graphed the distance of the closest imputed path to each ground truth path.

Spline interpolation

To compare the impact of different spline interpolation options, we interpolated over the chosen paths with several types of splines, using the function smooth.spline, setting the function parameters as {*df*=*3*} or {*df*=*5*}. As the resulting scGETs are described by dynamical features, we compared the dynamical features of the trajectories for genes with different synthesis and degradation characteristics.

Calculation of maximum mutual information

An information theoretic approach was used to identify either individual genes or combinations of genes providing the highest maximum mutual information between ligand identity and gene expression. Estimation of maximum mutual information was implemented using the R package *SLEMI*,[67] which uses a statistical learning-based approach to more accurately and more efficiently estimate maximum mutual information for data types with higher dimensional outputs. The max MI was calculated for each gene individually, using all six stimuli (highest theoretical max MI = $log_2(6)$ = 2.58 bits).

Identification of optimized gene combinations

To estimate the maximum mutual information of the best combination of $1, 2, 3, \ldots, N$ genes for each dynamical features, we compiled the top 20 genes that individually had the best max MI value. For each of these single-dimension channels, we scanned every combination of two genes, calculated the max MI for the two-dimensional output, and again ranked the best combinations of two genes and retained the top 20, following approaches published previously.[12,14] This forward-selection process was repeated for each additional gene until the gain in max MI for each additional gene plateaued. Retaining only the top 20 sets at each dimension made the calculation computationally feasible, while still allowing the possibility for gene combinations that are not simply additive of the previous dimension's highest max MI combination.

Statistical learning models to predict polarization state

Statistical learning methods were applied to develop classification models to identify cell functional states from 1) steady-state abundances, 2) single timepoint responses, or 3) response trajectory dynamical features. Individual classification models for polarization state were trained for each stimulus. To determine classification accuracy of cell polarization state for each of the above three types of input, we trained LASSO (Least Absolute Shrinkage and Selection Operator)-regularized regression models, as implemented in the R package *glmnet*, on three canonical marker genes. Models were trained and then tested using a 70%/30% split of the data. After the model was trained, the remaining held-out data were tested, resulting in F1 scores and confusion matrices that summarized prediction accuracy for each stimulus used. We also fit a multinomial LASSO-penalized regression model using all measured genes as input and polarization state as output, to identify the number of genes/features selected for each of the above three types of data input.