

Full Title

Inferring the structures of signaling motifs from paired dynamic traces of single cells

Short Title

Motif inference from single cell dynamics

R. A. Haggerty^{1,2,3} and Jeremy E. Purvis^{1,2,3,4,*}

¹Department of Genetics

²Computational Medicine Program

³Curriculum for Bioinformatics and Computational Biology

⁴Lineberger Comprehensive Cancer Center

University of North Carolina, Chapel Hill

116 Manning Drive

11108A Mary Ellen Jones Building

Chapel Hill, NC 27599-7488

*Corresponding Author:

Jeremy Purvis

Mary Ellen Jones Building 11018C, CB#7488

116 Manning Drive

Chapel Hill, NC 27599-7488

jeremy_purvis@med.unc.edu

Abstract

Individual cells show variability in their signaling dynamics that often correlates with phenotypic responses, indicating that cell-to-cell variability is not merely noise but can have functional consequences. Based on this observation, we reasoned that cell-to-cell variability under the same treatment condition could be explained by a single signaling motif that deterministically maps different upstream signals into a corresponding set of downstream responses. If this assumption holds, then repeated measurements of upstream and downstream signaling dynamics in single cells could provide information about the underlying signaling motif for a given pathway, even when no prior knowledge of that motif exists. To test these two hypotheses, we developed a computer algorithm called MISC (Motif Inference from Single Cells) that infers the underlying signaling motif from paired time-series measurements from individual cells. When applied to measurements of transcription factor and reporter gene expression in the yeast stress response, MISC predicted signaling motifs that were consistent with previous mechanistic models of transcription. The ability to detect the underlying mechanism became less certain when a cell's upstream signal was randomly paired with another cell's downstream response, demonstrating how averaging time-series measurements across a population obscures information about the underlying signaling mechanism. In some cases, motif predictions improved as more cells were added to the analysis. These results provide evidence that mechanistic information about cellular signaling networks can be embedded within the dynamical patterns of single cells.

Author Summary

Cells use molecular signaling networks to translate dynamically changing stimuli into appropriate downstream responses. Specialized network structures, or motifs, allow cells to properly decode a variety of temporal input signals. In this paper, we present an algorithm that infers signaling motifs from multiple examples of an upstream signal paired with its downstream response in the same cell. We compare the predictive power of single-cell versus averaged time-series traces and the incremental benefit of adding more single-cell traces to the algorithm. We use this approach to understand how yeast respond to environmental stresses.

Introduction

Cells interpret complex temporal patterns of molecular signals to execute appropriate downstream responses such as changes in gene expression or cell fate [1-3]. The molecular factors that participate in these signaling networks are often organized into specialized network structures, or motifs, that carry out a specific signal-processing function [4-6]. For example, a positive feedback loop can facilitate strong and irreversible responses to an upstream signal such as the commitment to cell division [7]. A negative feedback loop, such as the metabolic response to changes in blood insulin, allows cells to adapt to different levels of an upstream signal [8, 9] or to filter signaling noise [10]. More complex network motifs, such as coupled positive and negative feedback, can lead to oscillations [11, 12]. Here, we use the term “upstream signals” to refer to the inputs that initiate signaling in a particular pathway. Examples of an upstream signal include the activity or expression level of a receptor, kinase, or second messenger. These signals are decoded by specialized motifs into “downstream responses” such as changes in gene expression or epigenetic state (Figure 1A). Understanding the signaling motifs that decode upstream signals into downstream responses is a major goal of systems biology because these mechanisms define the dynamic relationships among signaling components and provide quantitative predictions about the cellular response to pharmacological intervention [13].

Interestingly, not all cells respond to the same upstream signal in an identical way. Previous studies have shown that individual cells show considerable heterogeneity in their dynamic responses to the same input stimulus (Figure 1B) [14, 15]. Stimulation with epidermal growth factor (EGF), for example, leads to differences in extracellular signal-related kinase (ERK2) activity [16, 17]. Similarly, uniform induction of DNA damage leads to heterogeneous patterns of p53 dynamics [18]. In many cases, these differences in signaling dynamics are correlated with distinct cell fate decisions. For example, the temporal pattern of cyclin-dependent kinase 2 (CDK2) activity after mitosis—although variable from cell to cell—predicts whether a cell will proceed directly into S-phase or spend additional time in G1 [19]. Similarly, sustained ERK2 activity following serum addition indicates that cells will enter S-phase [16]. Such correlations suggest that differences in upstream signaling patterns play an important role in driving cells toward specific downstream responses.

These collective observations suggest a testable hypothesis: if differences between individual cells have predictable effects on downstream responses, then perhaps all cells in that population share a common signaling motif that consistently interprets each cell’s unique signaling dynamics (Figure 1C). In other words, upstream signals are consistently “mapped” to

downstream responses. If such a single mechanism exists in all cells, then a second hypothesis follows: by using many examples of the upstream signals and downstream responses, it may be possible to infer the underlying signaling motif without any prior knowledge of its structure. That is, single-cell signaling patterns gathered under the same experimental conditions could reveal the structure of the molecular mechanism that produced those patterns.

Here, we test these two hypotheses by analyzing paired sets of upstream and downstream signals in single cells. We focus specifically on inferring the signaling motifs that control the stress response of budding yeast [20, 21]. We introduce a computational approach called Motif Inference from Single Cells (MISC) that determines the mechanistic relationship between measurements of two fluorescent reporters in the same cell. MISC exploits the heterogeneity among identically treated cells by finding one or more signaling motifs that best explain the relationship between the paired upstream signals and downstream responses across all cells in a population. We believe this to be the first tool that infers dynamic networks from paired time-series traces from single cells. Our results provide evidence that single-cell dynamics contains information that reveals the underlying signaling mechanisms that produced them.

Results

Signaling motifs convert upstream signals into downstream responses

We hypothesized that cell-to-cell variation in an upstream signal can be decoded by a common signaling motif that produces an accompanying set of downstream responses. Under this model, the original source of variability in the upstream signal is not considered. Instead, we focus on the mechanistic relationship between the upstream signal and the downstream response. To explore this, we first considered how variation in an upstream signal might be propagated by a single signaling motif. We chose a simple upstream signaling pattern represented by a rapid 5-fold increase above basal levels after a 2 min delay and a return to basal levels after 6 minutes. To simulate realistic cell-to-cell variability in this signaling pattern, we allowed each cell to deviate from the average behavior in its amplitude and delay time (see Materials and Methods). This procedure allowed us to generate an arbitrarily large number of unique upstream signaling patterns that closely resembled experimental measurements from individual cells (Figure 2A).

We next generated 50 individual upstream signals, X , and asked how these simulated signaling patterns could be decoded by a common signaling motif (Figure 2B). We selected the incoherent feedforward loop (IFFL) as the signaling motif because of its well-characterized behavior [22, 23]. An IFFL involves three signaling factors (X , Y , and Z). X positively regulates

Y and Z, and Y negatively regulates Z. The mechanism is represented as a set of ordinary differential equations (ODEs) that allow calculation of the downstream response given a particular upstream signal. We used the Hill-Langmuir equations to represent positive and negative interactions between X, Y, and Z but derived new expressions for the right hand side of the ODEs to accommodate fold-change representations of X, Y, and Z rather than absolute quantities (Supplementary Information).

As expected, the average response to the step increase in X is a transient increase in Z that returns to basal levels as the accumulation of Y dampens the increase in Z and when the signal in X disappears, the cells overcorrect and drop below basal, and then return to basal levels. At the single-cell level, each cell showed a slightly varying downstream response in accordance with the upstream signal. When the upstream signal for each cell was plotted against its downstream response at 2.5 min, these differences showed a moderately positive correlation (Figure 2C, black dots). This analysis is analogous to performing immunohistochemistry to detect levels of X and Z across a large population of cells at a fixed time point. Importantly, this correlation was destroyed when the upstream signal of a given cell was randomly paired with the downstream response of a different cell (Figure 2C, gray dots). Thus, for this analysis, the correlation between upstream signals and downstream responses among individual cells reflects a mechanistic relationship between these two signaling factors.

These results show how a single signaling motif can lead to different downstream responses by propagating variation in an upstream signal. Importantly, the signaling motif is deterministic: it does not invoke any stochastic processes to generate different downstream responses. Instead, cell-to-cell differences are each mapped to unique downstream responses.

Inferring the signaling motif from simulated single-cell dynamics

For many signaling pathways, the signaling motif that converts upstream signals into downstream responses is unknown [16, 19, 24]. We developed a computer algorithm called Motif Inference from Single Cells, or MISC, that infers signaling motifs from measured single-cell dynamics. Before applying the algorithm to an experimental data set, we first tested MISC on the synthetic data set generated by a known signaling motif described above. Figure 3A shows the upstream signaling dynamics and downstream responses for 50 simulated cells using an IFFL as the signaling motif. To infer the signaling motif in an unbiased way, we considered all possible structures of signaling motifs that could explain the relationship between X and Z, including mechanisms that involve a potential intermediate signaling factor, Y. Although it is possible to include more than one unknown factor in our search, we limited our search to

networks of 3 signaling factors as proof of principle and because the additional node, Y, can represent a subnetwork of one or more factors. We did not consider feedback to X because X is explicitly provided to the algorithm and reports on its own feedback

The complete set of signaling motifs can be succinctly represented as a heat map (Figure 3B). In this visual format, the regulatory relationships between each pair of factors is represented by a black or red box indicating positive or negative regulation, respectively. After eliminating redundant or trivial motif structures (e.g., motif containing only Y), there are 402 possible signaling motifs for a 3-factor signaling motif that allows distinct positive and negative interactions. Each mechanism was converted to set of ODEs that were used to calculate the downstream response Z for a given upstream signal X (see Supplementary Information). We fixed the sign (positive or negative) of all regulatory interactions in the mechanism and used 40,000 randomly selected, bounded, parameter sets to parameterize the ODEs (see Supplementary Information). We used a Monte Carlo approach rather than fitting parameters because we want to determine which motifs are robust to random parameterization. Next, each of the dynamic traces for this calculated response, Z' , were compared to the original downstream response to calculate an error for each signaling motif (Figure 3C). Using this approach, all signaling motifs receive an error score. Comparing these errors provides the bases for inferring the underlying signaling motif: predicted responses that best matched the original downstream response were hypothesized to be more structurally similar to the original (i.e., unknown) motif.

Figure 3D shows the errors for all randomly parameterized motifs, ranked in ascending order of error when the output of each motif was compared to the synthetic data. We found that the distribution of ranked error scores rapidly approached an asymptote of the worst possible score, which represents the output of a motif with no signaling output (i.e., a flat line at $Z' = 1$). This score represents the performance of a motif with no predictive power. The vast majority of the motifs produced errors that were close to the error asymptote. To identify well-performing motifs in a principled way, we then looked for the “elbow” of the graph in which the curve began to flatten out and the errors approached the asymptote. Algorithmically, we do this by looking at the first derivative and finding the point where the slope is 1% of the maximum slope (Figure 3D, inset). This is the location where the slope begins to level off. After separating the motifs into well-performing and poorly-performing motifs, we look among the well-performing motifs for the most overrepresented motifs. We do this by breaking down each motif into its component motifs (e.g., the IFFL contains a 1-edge motif represented by X activating Z) and calculating the enrichment of each sub-motif among the well-performing group. Our inference approach

predicted a small set of structurally similar signaling motifs including the correct motif structure. The IFFL was the top-ranking motif among those containing three interactions, present in 79% of the well-performing group. Furthermore, the top graphs also included additional motifs known to create adaptive behavior. For example, a similar structure containing positive feedback from Y to itself, was the highest ranking motif with four-interactions. Additionally, MISC predicted several motifs with negative feedback from Z to itself through Y, another motif which creates adaptive behavior in the downstream response (Figure 3F). In this manner, MISC was used to successfully infer the correct signaling motif for a synthetic data set in which the underlying motif structure was already known.

Different yeast stress response pathways show use of the same signaling motif

We next applied MISC to determine the underlying signaling motifs from experimental data. We focused on the transcriptional response to environmental stress in budding yeast, which has been studied extensively using time-lapse fluorescence microscopy [21, 25, 26]. To predict the signaling motifs at work in the yeast stress response, we used previously published single-cell data in which two signaling activities were measured in the same cell over time [21]. Here, the upstream signal X is the transcriptional activity of the transcription factor MSN2. Upon stress, MSN2 is translocated to the nucleus where it promotes transcription of multiple target genes (Figure 4A). The downstream response Z is a fluorescent protein driven by several DNA stress response elements (STREs). Thus, X and Z represent the signaling pathway that transmits upstream stress signals to the downstream expression of target genes that allow yeast to appropriately cope with stress.

We first examined the response to nutrient stress with 0.1% sugar in 60 single cells (Figure 4B). When the activity of all cells was averaged, MSN2 showed a transient 18-fold increase followed by a return to near-basal levels. As reported previously, many cells showed a prominent pulse of MSN2 nuclear localization followed by smaller, more sporadic pulses. The averaged downstream response of STRE gene expression showed a gradual, almost linear, increase in expression. Individual cells showed considerable variability in downstream response, with some cells displaying very little changes in expression. To predict the underlying signaling motif for this stress response, we applied the MISC algorithm to the full single-cell data set. The MISC fingerprint suggested several significant motif structures that were consistent with the measured relationships between X and Z. The top ranking sub-motif was an indirect connection from X to Z mediated through Y, showing up in 95% of the top motifs. Other well performing top results included the same connection from X to Z but mediated through Y with

feedback either to Y from itself (55%) or from the STRE to itself (49%). These predicted motif structures are similar to the previously published two-step model of transcription, in which the STRE promoter must first transition from an inactive to active state before productive transcription can begin [21]. Negative feedback on Z may represent first-order degradation of the transcription product, which is also present in previous models of transcription [20, 21, 26].

We next considered the yeast response to osmotic stress, using a smaller data set consisting of 28 single cells that were treated with 0.5 M KCl (Figure 4C). Under this condition, MSN2 levels showed a transient 12-fold increase in nuclear localization followed by a near-perfect adaptation to basal levels, leveling off around 3-fold. The downstream response showed a sudden increase in expression around 20 minutes after treatment that reached sustained 7-fold levels within the first 60 minutes. As with sugar stress, individual cells showed varying behavior and sporadic pulses of MSN2 activity. Applying MISC to this data set revealed the same top sub-motifs as for the glucose stress, but at slightly different percentages (93%, 64% and 55%).

Finally, we considered the yeast stress response to oxidative stress using a data set of 57 individual cells (Figure 4D). Under this experimental condition, MSN2 levels showed a more prolonged response to stress, reaching 14-fold change that adapted more slowly than either salt or sugar stress. The downstream response of the STRE reporter showed a biphasic response, characterized by a gradual increase that peaked at 1 h, followed by a transient decrease and subsequent increase in expression. When these data were analyzed by MISC, the top sub-motif was the same of X to Z mediated through Y at 65%. No other sub-motifs showed up in more than 50% of the top ranking graphs, however the sub-motif with the feedback from the STRE to itself was the next highest and showed up in 37% of the graphs. Notably, the top motifs for all three stresses were similar, suggesting that even though the stresses and the responses were different, the core machinery in place to decode the stresses is the same [21].

Single-cell dynamics contain information about the underlying motif structure that is not present in the population-averaged response

Previous studies have predicted motif structures based on a single time series [23]. It is unclear what additional value, if any, is provided by multiple single-cell measurements. To answer this question, we performed MISC on the same stress response data but permuted the data so that the upstream signal X for a given cell was randomly paired with the downstream response Z of another cell from the same experiment. Importantly, the population-averaged

response remains the same under this perturbation, but the relationship between MSN2 and STRE at the level of individual cells is lost (Figure 5A).

As proof of principal, we tested this permutation strategy on the synthetic single-cell data generated by the IFFL (Figures 2 and 3). Predictions from matched cells were consistently better at predicting the correct structure ($77.5 \pm 0.1\%$) than randomly paired cells, which predicted the correct motif in less than half of the well-performing motif structures (Figure 5B). When tested on the yeast stress response data, however, the permutation analysis showed a variety of trends. For glucose stress, random pairing of cells showed no difference in the distribution of the predictions. (KS test, $p = 0.86$). For both H_2O_2 and KCl, however, incorrect pairing of cells led to a difference in the distributions of motif predictions (KS test, $p = 0.0069$ and 0.0069) in which the paired cells performed better than their shuffled counterparts. For H_2O_2 , the variance decreased when paired, and for KCl, the variance decreased and the mean increased when paired. Taken together, these results show that at least part of the predictive power of MISC comes from the observed relationships between the dynamics of X and Z in individual cells. This result suggests that information about the underlying signaling motif can be embedded in cell-to-cell variability.

We next asked whether adding additional cells to the MISC analysis could improve its predictive power. To address this question, we calculated the percentage of time the correct sub-motif appeared in the top-performing motifs as a function of the number of cells included in the analysis. As a benchmark, we subsampled the synthetic IFFL—starting with five cells—and then incrementally adding more cells until we reached the total number of cells collected under each condition. For each of these subsampled sets of cells, we ran the MISC algorithm and calculated what percentage the correct motif (as determined by using all cells) appeared among the top-scoring motifs. We performed five replicates per sample size, choosing a random subset of cells for each replicate. The IFFL showed no gain in predictive power as more cells were added to the analysis (Figure 6A). This result was expected because, in the synthetic data set, each cell's downstream response is deterministically generated by its upstream response. The absence of intrinsic noise in the synthetic data renders each cell a perfect predictor of the underlying motif structure [29, 30].

We then performed this analysis again on the experimentally-determined stress data, using the top performing sub-motif that encapsulates all higher ranking sub-motifs. Under glucose limitation, MISC also failed to show better predictive accuracy with increasing number of cells, although there was a trend toward less variance in the prediction (Figure 6B). This finding

suggests that additional cells are helpful in ruling out competing models for signal transduction. Under both salt and oxidative stress, however, we observed an increase in predictive accuracy as we increased the number of single cell traces subjected to MISC (Figure 6C-D). Given the limited number of cells gathered experimentally, it is unclear at what number of cells the predictive accuracy of MISC would have leveled off. Nevertheless, these results demonstrate that each pair of time-series traces from a single cell adds incremental value to the quality of the motif prediction.

Discussion

It is commonly stated in the field that single-cell measurements contain information that is obscured by population-averaged methods such as western blotting and quantitative PCR [31-34]. Indeed, single-cell approaches have revealed a staggering degree of heterogeneity among individual cells in terms of gene expression and protein dynamics. Here, we exploit those cell-to-cell differences by asking whether they arise from the same underlying generative process. Simulations show that we can predict a signaling motif based on paired sets of dynamic traces. We found that predictions were strengthened by considering the coupled relationships between two signals in individual cells, and that adding additional cells to the analysis can improve, in some cases, the certainty and accuracy of motif predictions. A similar strategy was used to discover signaling proteins that affect cell motility [35]. By screening a library of fluorescent fusion proteins for proteins that correlate with cell movement, this apparent “noise” in protein expression was used to identify novel regulators of motility.

Many biological network inference approaches have focused on predicting the topology of a pathway using static or population-averaged data that does not consider the single-cell dynamics of the system. Inference based on dynamical data was previously applied to global profiling data, in an effort to identify temporal transition nodes and edges [36]. MISC offers a novel inference strategy that examines a smaller scope of a biological system and aims to infer the mechanistic relationships among the measured components using single-cell dynamics. MISC allows nonlinear interactions between signaling factors. In addition, MISC allows the functional form of the governing equations to be modified to better suit various systems (Supplementary Information).

Another strength of MISC is the ability to infer components of the system that were not directly measured. Although it may be possible to measure the upstream signal and downstream response for a given signaling pathway, signaling motifs often involve additional molecular intermediates or regulatory relationships that are unmeasured or unknown. For

example, it may be possible to measure the activity of a transcription factor along with target gene expression in single cells [16, 20, 37]. However, additional factors, such as cofactors or chromatin modifying complexes, may modulate the downstream response [38]. Yet, knowledge of these signaling motifs is necessary to quantify how the upstream signal is decoded to produce the observed downstream responses. We present MISC as a tool that can hypothesize unknown signaling intermediates based on the dynamical behaviors of interacting signaling factors. As with other inference methods, the motif predictions made by MISC can be experimentally tested and used to make future predictions about the cell's behavior.

Finally, we show how increasing the number of single-cell traces improves the certainty and accuracy of motif prediction. Future studies are warranted to determine whether there is a most economical number of single-cell traces to collect under a certain experimental condition in order to identify the underlying signaling network that generated the traces. This is a practical consideration in many live-cell studies, in which accurate quantification of traces can be laborious and time-consuming, and for which there are often no best-practices for how many cells should be collected and analyzed. As such, MISC occupies a unique space in the field of biological network inference methods.

Materials and Methods

MATLAB scripts for all simulations, calculations, and figures are provided as Supplementary Information and are available at github.com/rahaggerty/MISC.

Acknowledgements

We thank Nan Hao and Erin O'Shea for providing raw single-cell data of the yeast stress response. We thank the laboratories of Tim Elston, Henrik Dohlman, Beverly Errede, and Galit Lahav for valuable criticism and suggested analyses. This work was supported by NIH grant DP2-HD091800 (J.E.P.) and fellowship F31-HL134336 (R.A.H.).

Author Contributions

R.A.H. developed the MISC software, performed simulations and analysis, derived the ODE expressions for the fold-change representation of Michaelis-Menten kinetics, co-designed the figures, and co-wrote the manuscript. J.E.P. conceived of the MISC inference method, co-designed the figures, and co-wrote the manuscript.

Figure Legends

Figure 1. Signaling motifs determine how upstream signals are converted into downstream responses.

(A) The same upstream signal, X , can produce different downstream responses, Z , depending on the signaling motif. Positive feedback leads to rapid amplification of Z following a delay in its induction. An incoherent feedforward loop (IFFL) allows Z to adapt to changes in X by first activating then dampening the downstream response. Coupled positive and negative feedback can lead to oscillations of Z . Signaling motifs often involve an intermediate signaling factor, Y , that is necessary to achieve the appropriate downstream response. Ordinary differential equations for each signaling motif are provided in the Supplemental Information.

(B) In response to a given stimulus, individual cells show heterogeneous signaling patterns. For many cellular signaling pathways, the variability of the upstream signal, X , is correlated with the downstream response, Z .

(C) Hypothetical model for a common signaling motif that explains the correlation between upstream signaling and downstream responses. Differences in upstream signal are mapped onto the downstream response. Under this model, it may be possible to infer the underlying structure by observing many examples of the upstream and downstream signaling patterns.

Figure 2. A single signaling motif can explain the correlation between upstream signals and downstream responses in single cells

(A) Experimentally measured values for cell-to-cell variation in signal height and delay were used to generate a large number of simulated signaling patterns. Each trace represents a unique upstream signal from an individual cell.

(B) Average (*top*) and single-cell (*bottom*) signaling patterns for upstream signals (*left*) and downstream responses (*right*) for 50 simulated cells. While X is generated using the approach in panel A, the downstream responses are calculated using the signaling motif (middle). Here, Z is produced by the incoherent feed forward loop.

A single signaling motif translates 50 upstream signals into 50 downstream responses. Each row corresponds to an individual cell. Color represents the fold change.

(C) There is a correlation between the upstream and downstream signals that is not present if the cells are shuffled. Black dots are from paired cells and grey dots are from shuffled cells. The red line is a correlation line for the paired cells. The unpaired cells do not have any correlation.

Figure 3. Inferring the underlying signaling motif from simulated single-cell dynamics

(A) An incoherent feedforward loop (IFFL) signaling motif was used to generate a synthetic data set for 50 single cells. Individual upstream signals X were generated as shown in [Figure 2A](#).

Each signal was then subjected to the IFFL signaling motif using ODEs to calculate a unique downstream response Z for each cell (see Supplementary Information). The resulting X and Z pairs for all cells were then analyzed to infer the underlying signaling motif.

(B) Heat map representation of motif structures for all possible 3-factor signaling motifs. (*left*) A conventional representation of motif structure consisting of circles and arrows can be represented by (*right*) a single column of shaded rectangles in which each rectangle denotes a specific interaction within the motif: black, positive regulation; red, negative regulation; white, no regulation. The heat map representation for the IFFL is shown next to the conventional diagram. (Right) All 402 possible unique motif structures are shown for a 3-component network (X , Y , and Z).

(C) To infer the mechanism underlying X and Z , a separate downstream response, Z' , is calculated for each possible parameterized signaling motif and compared to the original downstream response, Z . Each motif is assigned an error score that reflects the average difference in fold change between Z and Z' per cell per unit time.

(D) Distribution of error scores. The error scores are distributed such that they approach the asymptote of the worst possible score (i.e., the output of an unconnected graph). The top graphs are defined as those before the “elbow” of the graph and are determined by looking at the first derivative.

(E) The top-scoring motifs. The links for the incoherent feed forward loop, as well as its simpler sub-motif structures, are overrepresented among the best-performing motifs.

(F) Heat map representations of the top 10 sub-motifs (middle) and the percentage of time each motif was observed among the top-scoring results (top). Individual graphs are shown in the traditional format as well (bottom). Boxed is the IFFL, which is the top sub-motif containing all simpler sub-motifs that score better.

Figure 4. Motif inference for the yeast stress response pathway under different environmental perturbations.

(A) Schematic of the system showing the fluorescent reporter system used. See ref. 21 for experimental details. Experimental data were used with author permission.

(B) Oxidative Stress. 58 cells were treated with 0.1 mM H₂O₂. Far left, heatmap of the individual cell's Msn2 activation, and a plot of the mean and standard deviation of the cells. Left middle, STRE response as reported by the CFP reporter. Right middle, heatmap and ball and stick representations of the significant sub-motifs. Displayed as ball and stick representations are the top sub-motifs that contain higher ranking sub-motifs. Far right, Simulated STRE response of the top performing sub-motif.

(C) Nutrient Stress. 60 cells were treated with a 0.1% glucose media. Far left, heatmap of the individual cell's Msn2 activation, and a plot of the mean and standard deviation of the cells. Left middle, STRE response as reported by the CFP reporter. Right middle, heatmap and ball and stick representations of the significant sub-motifs.

(D) Osmotic Stress. 28 cells were treated with 0.5 M KCl. Far left, heatmap of the individual cell's Msn2 activation, and a plot of the mean and standard deviation of the cells. Left middle, STRE response as reported by the CFP reporter. Right middle, heatmap and ball and stick representations of the significant sub-motifs.

Figure 5. Single-cell dynamics can provide improved certainty or accuracy in motif prediction than population-averaged data

(A) In order to evaluate motif prediction based on single-cell versus population-averaged time-series traces, we performed MISC on pairs of time-series traces that were either correctly paired (*top*) or randomly shuffled (*bottom*). Note that in both cases the average time series trajectory is the same.

(B) Shuffled versus paired predictions for the IFFL and yeast stresses. Ten replicates were used for each condition. Note that MISC produces variability in predictions even for paired traces since parameter values on the motifs are randomly chosen. The synthetic data, oxidative stress and osmotic stresses showed differences in distributions between shuffled and paired, while nutrient stress did not.

Figure 6. Additional single-cell measurement can provide increasingly better predictions about the underlying signaling motif

MISC was performed on an increasing number of paired single-cell measurements and the accuracy of the prediction (based on the sub-motif predicted by the full data set) was calculated. (A) Synthetic Data, (B) Nutrient Stress, (C) Osmotic Stress, (D) Oxidative Stress. The Synthetic Data and Nutrient Stress did not increase, but Osmotic and Oxidative Stresses predictions increased in percentage as the number of cells increased.

References

1. Purvis JE, Lahav G. Encoding and decoding cellular information through signaling dynamics. *Cell*. 2013;152(5):945-56. doi: 10.1016/j.cell.2013.02.005. PubMed PMID: 23452846; PubMed Central PMCID: PMC3707615.
2. Kholodenko BN, Hancock JF, Kolch W. Signalling ballet in space and time. *Nature reviews*. 2010;11(6):414-26. Epub 2010/05/25. doi: nrm2901 [pii]
10.1038/nrm2901. PubMed PMID: 20495582; PubMed Central PMCID: PMC2977972.
3. Levine JH, Lin Y, Elowitz MB. Functional roles of pulsing in genetic circuits. *Science*. 2013;342(6163):1193-200. Epub 2013/12/07. doi: 10.1126/science.1239999
342/6163/1193 [pii]. PubMed PMID: 24311681.
4. Bhalla US, Iyengar R. Emergent properties of networks of biological signaling pathways. *Science*. 1999;283(5400):381-7. PubMed PMID: 9888852.
5. Alon U. Network motifs: theory and experimental approaches. *Nat Rev Genet*. 2007;8(6):450-61. Epub 2007/05/19. doi: nrg2102 [pii]
10.1038/nrg2102. PubMed PMID: 17510665.
6. Behar M, Hoffmann A. Understanding the temporal codes of intra-cellular signals. *Curr Opin Genet Dev*. 2010;20(6):684-93. Epub 2010/10/20. doi: S0959-437X(10)00166-8 [pii]
10.1016/j.gde.2010.09.007. PubMed PMID: 20956081; PubMed Central PMCID: PMC2982931.
7. Trunnell NB, Poon AC, Kim SY, Ferrell JE, Jr. Ultrasensitivity in the Regulation of Cdc25C by Cdk1. *Molecular cell*. 2011;41(3):263-74. doi: 10.1016/j.molcel.2011.01.012. PubMed PMID: 21292159; PubMed Central PMCID: PMC3060667.
8. Kubota H, Noguchi R, Toyoshima Y, Ozaki Y, Uda S, Watanabe K, et al. Temporal Coding of Insulin Action through Multiplexing of the AKT Pathway. *Molecular cell*. 2012;46(6):820-32. Epub 2012/05/29. doi: S1097-2765(12)00340-1 [pii]
10.1016/j.molcel.2012.04.018. PubMed PMID: 22633957.
9. Goentoro L, Kirschner MW. Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling. *Molecular cell*. 2009;36(5):872-84. doi: 10.1016/j.molcel.2009.11.017. PubMed PMID: 20005849; PubMed Central PMCID: PMC2921914.

10. Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. *Nature*. 2000;405(6786):590-3. doi: 10.1038/35014651. PubMed PMID: 10850721.
11. Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, et al. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet*. 2004;36(2):147-50. Epub 2004/01/20. doi: 10.1038/ng1293
ng1293 [pii]. PubMed PMID: 14730303.
12. Nelson DE, Ihekwaba AE, Elliott M, Johnson JR, Gibney CA, Foreman BE, et al. Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science*. 2004;306(5696):704-8. Epub 2004/10/23. doi: 10.1126/science.1099962. PubMed PMID: 15499023.
13. Behar M, Barken D, Werner SL, Hoffmann A. The dynamics of signaling as a pharmacological target. *Cell*. 2013;155(2):448-61. doi: 10.1016/j.cell.2013.09.018. PubMed PMID: 24120141; PubMed Central PMCID: PMC3856316.
14. Davis DM, Purvis JE. Computational analysis of signaling patterns in single cells. *Seminars in Cell and Developmental Biology*. 2014.
15. Regot S, Hughey JJ, Bajar BT, Carrasco S, Covert MW. High-sensitivity measurements of multiple kinase activities in live single cells. *Cell*. 2014;157(7):1724-34. Epub 2014/06/21. doi: 10.1016/j.cell.2014.04.039
S0092-8674(14)00655-2 [pii]. PubMed PMID: 24949979; PubMed Central PMCID: PMC4097317.
16. Albeck JG, Mills GB, Brugge JS. Frequency-Modulated Pulses of ERK Activity Transmit Quantitative Proliferation Signals. *Molecular cell*. 2013;49(2):249-61. Epub 2012/12/12. doi: 10.1016/j.molcel.2012.11.002
S1097-2765(12)00933-1 [pii]. PubMed PMID: 23219535.
17. Cohen-Saidon C, Cohen AA, Sigal A, Liron Y, Alon U. Dynamics and variability of ERK2 response to EGF in individual living cells. *Molecular cell*. 2009;36(5):885-93. Epub 2009/12/17. doi: 10.1016/j.molcel.2009.11.025. PubMed PMID: 20005850.
18. Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, et al. Oscillations and variability in the p53 system. *Mol Syst Biol*. 2006;2:2006 0033. Epub 2006/06/15. doi: 10.1038/msb4100068 [pii]

10.1038/msb4100068. PubMed PMID: 16773083; PubMed Central PMCID: PMC1681500.

19. Spencer SL, Cappell SD, Tsai FC, Overton KW, Wang CL, Meyer T. The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell*. 2013;155(2):369-83. doi: 10.1016/j.cell.2013.08.062. PubMed PMID: 24075009; PubMed Central PMCID: PMC4001917.

20. Hansen AS, O'Shea EK. Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Mol Syst Biol*. 2013;9:704. doi: 10.1038/msb.2013.56. PubMed PMID: 24189399.

21. Hao N, O'Shea EK. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat Struct Mol Biol*. 2012;19(1):31-9. Epub 2011/12/20. doi: 10.1038/nsmb.2192

nsmb.2192 [pii]. PubMed PMID: 22179789.

22. Goentoro L, Shoal O, Kirschner MW, Alon U. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular cell*. 2009;36(5):894-9. Epub 2009/12/17. doi: S1097-2765(09)00859-4 [pii]

10.1016/j.molcel.2009.11.018. PubMed PMID: 20005851; PubMed Central PMCID: PMC2896310.

23. Ma W, Trusina A, El-Samad H, Lim WA, Tang C. Defining network topologies that can achieve biochemical adaptation. *Cell*. 2009;138(4):760-73. Epub 2009/08/26. doi: S0092-8674(09)00712-0 [pii]

10.1016/j.cell.2009.06.013. PubMed PMID: 19703401; PubMed Central PMCID: PMC3068210.

24. Purvis JE, Karhohs KW, Mock C, Batchelor E, Loewer A, Lahav G. p53 dynamics control cell fate. *Science*. 2012;336(6087):1440-4. doi: 10.1126/science.1218351. PubMed PMID: 22700930; PubMed Central PMCID: PMCPMC4162876.

25. Shellhammer JP, Pomeroy AE, Li Y, Dujmusic L, Elston TC, Hao N, et al. Quantitative analysis of the yeast pheromone pathway. *Yeast*. 2019. doi: 10.1002/yea.3395. PubMed PMID: 31022772.

26. Hansen AS, O'Shea EK. Encoding four gene expression programs in the activation dynamics of a single transcription factor. *Curr Biol*. 2016;26(7):R269-71. doi: 10.1016/j.cub.2016.02.058. PubMed PMID: 27046808.

27. Kawakami E, Singh VK, Matsubara K, Ishii T, Matsuoka Y, Hase T, et al. Network analyses based on comprehensive molecular interaction maps reveal robust control structures

in yeast stress response pathways. *NPJ Syst Biol Appl.* 2016;2:15018. doi:

10.1038/npsba.2015.18. PubMed PMID: 28725465; PubMed Central PMCID: PMCPMC5516916.

28. Goulev Y, Morlot S, Matifas A, Huang B, Molin M, Toledano MB, et al. Nonlinear feedback drives homeostatic plasticity in H₂O₂ stress response. *Elife.* 2017;6. doi: 10.7554/eLife.23971. PubMed PMID: 28418333; PubMed Central PMCID: PMCPMC5438251.

29. Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, et al. Systems biology. Accurate information transmission through dynamic biochemical signaling networks. *Science.* 2014;346(6215):1370-3. doi: 10.1126/science.1254933. PubMed PMID: 25504722; PubMed Central PMCID: PMCPMC4813785.

30. Handly LN, Yao J, Wollman R. Signal Transduction at the Single-Cell Level: Approaches to Study the Dynamic Nature of Signaling Networks. *J Mol Biol.* 2016;428(19):3669-82. doi: 10.1016/j.jmb.2016.07.009. PubMed PMID: 27430597; PubMed Central PMCID: PMCPMC5023475.

31. Spiller DG, Wood CD, Rand DA, White MR. Measurement of single-cell dynamics. *Nature.* 2010;465(7299):736-45. doi: 10.1038/nature09232. PubMed PMID: 20535203.

32. Wollman R. Robustness, Accuracy, and Cell State Heterogeneity in Biological Systems. *Curr Opin Syst Biol.* 2018;8:46-50. doi: 10.1016/j.coisb.2017.11.009. PubMed PMID: 29308439; PubMed Central PMCID: PMCPMC5752152.

33. Yuan GC, Cai L, Elowitz M, Enver T, Fan G, Guo G, et al. Challenges and emerging directions in single-cell analysis. *Genome Biol.* 2017;18(1):84. doi: 10.1186/s13059-017-1218-y. PubMed PMID: 28482897; PubMed Central PMCID: PMCPMC5421338.

34. Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences of the United States of America.* 2018;115(10):E2467-E76. doi: 10.1073/pnas.1714723115. PubMed PMID: 29463712; PubMed Central PMCID: PMCPMC5878004.

35. Farkash-Amar S, Zimmer A, Eden E, Cohen A, Geva-Zatorsky N, Cohen L, et al. Noise genetics: inferring protein function by correlating phenotype with protein levels and localization in individual human cells. *PLoS Genet.* 2014;10(3):e1004176. Epub 2014/03/08. doi: 10.1371/journal.pgen.1004176

PGENETICS-D-13-02140 [pii]. PubMed PMID: 24603725; PubMed Central PMCID: PMC3945223.

36. Kim Y, Han S, Choi S, Hwang D. Inference of dynamic networks using time-course data. *Briefings in bioinformatics*. 2014;15(2):212-28. doi: 10.1093/bib/bbt028. PubMed PMID: 23698724.
37. Loewer A, Batchelor E, Gaglia G, Lahav G. Basal dynamics of p53 reveal transcriptionally attenuated pulses in cycling cells. *Cell*. 2010;142(1):89-100. Epub 2010/07/06. doi: S0092-8674(10)00567-2 [pii] 10.1016/j.cell.2010.05.031. PubMed PMID: 20598361; PubMed Central PMCID: PMC3003696.
38. Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, van Oudenaarden A. Systematic identification of signal-activated stochastic gene regulation. *Science*. 2013;339(6119):584-7. Epub 2013/02/02. doi: 10.1126/science.1231456 339/6119/584 [pii]. PubMed PMID: 23372015.

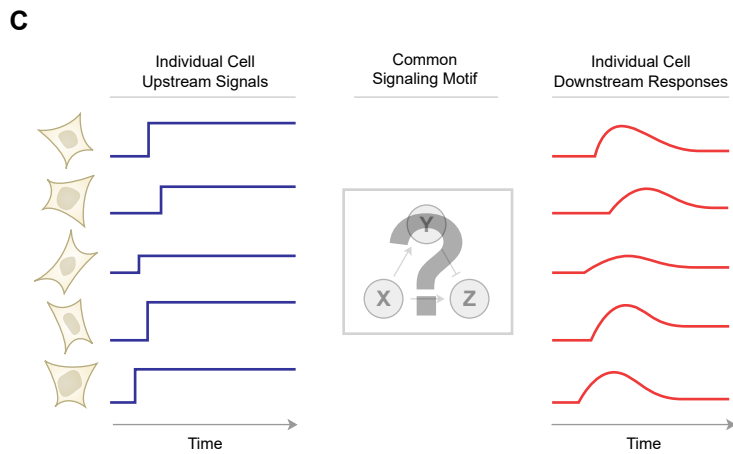
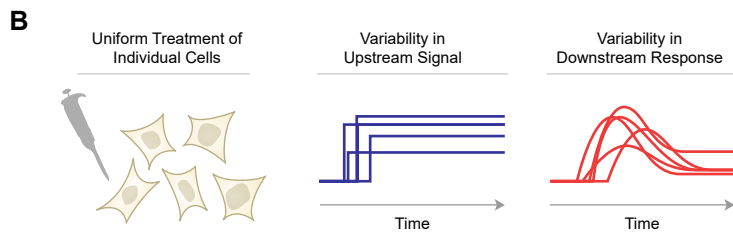
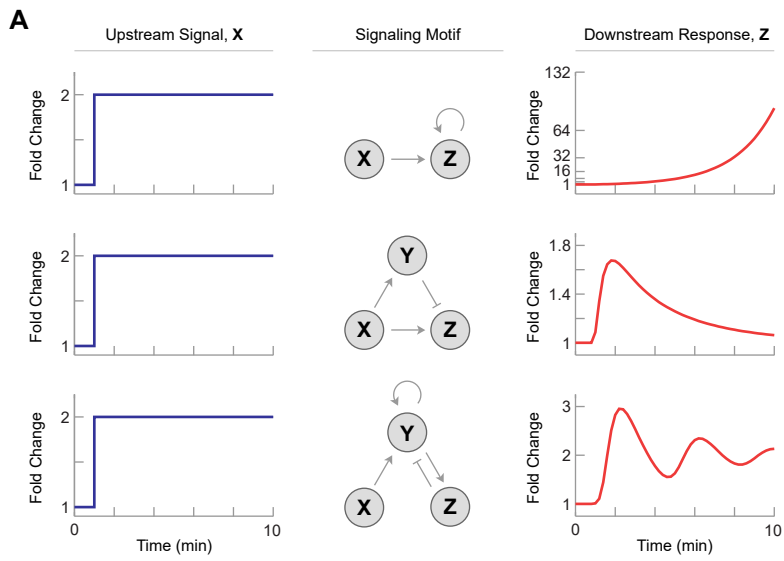


Figure 1

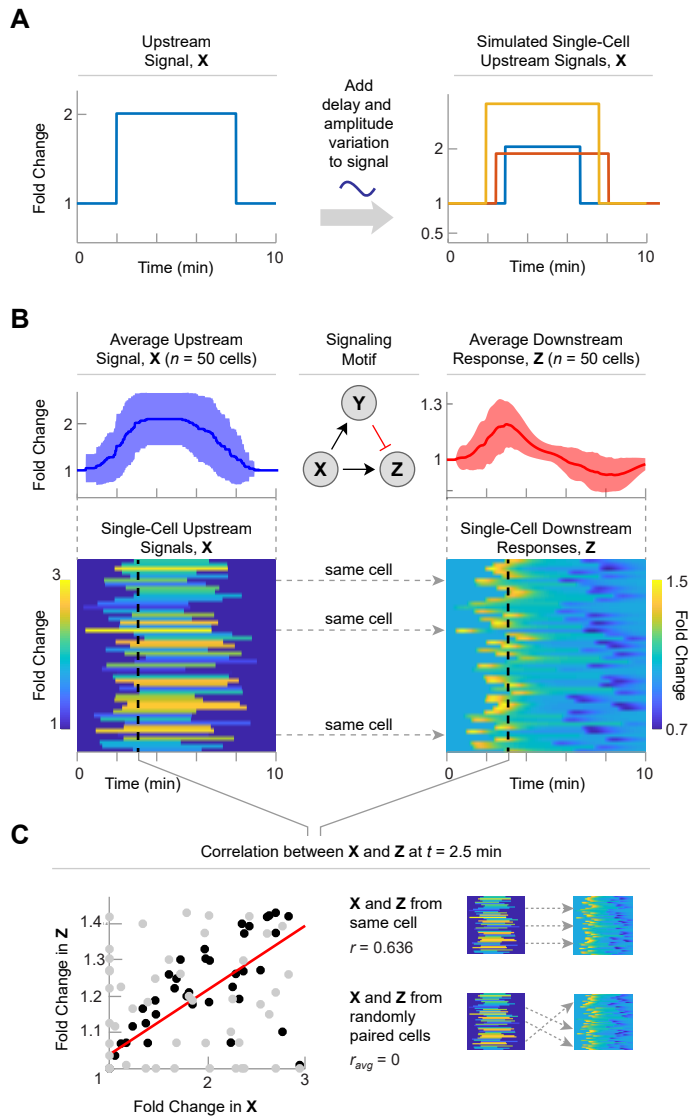
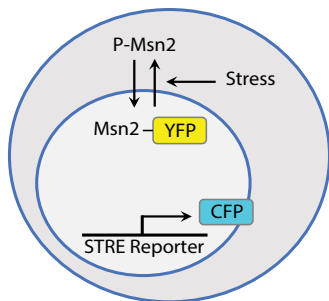


Figure 2

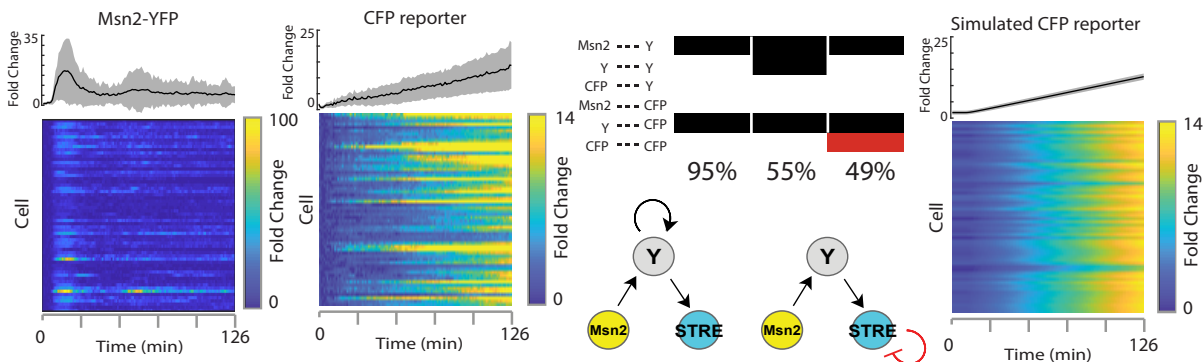
Figure 4

A



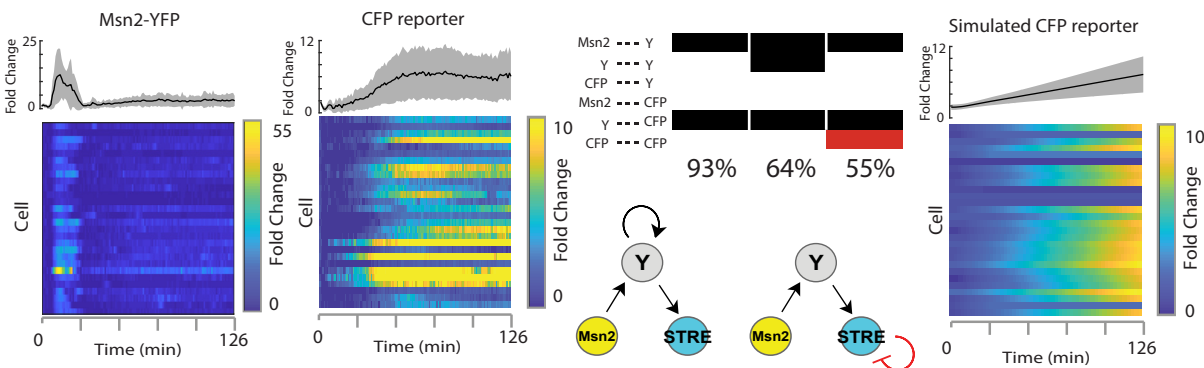
B

Nutrient Stress (0.1% glucose)



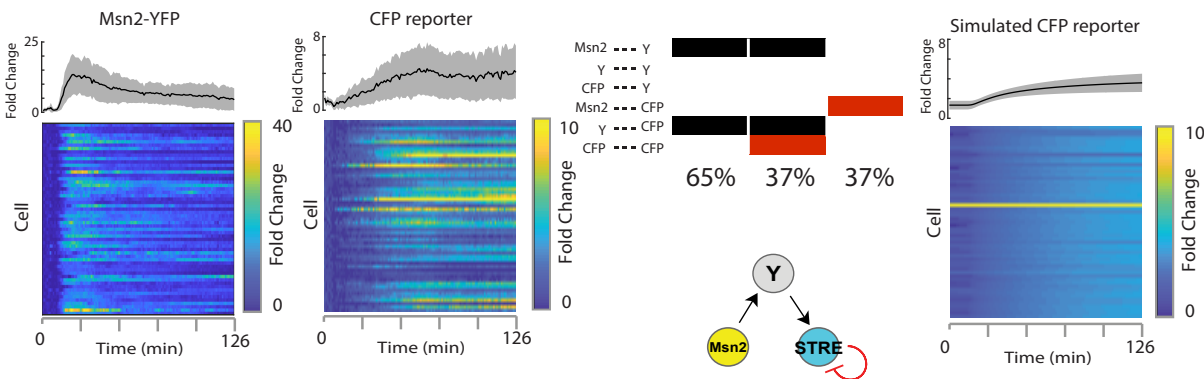
C

Osmotic Stress (0.5 M KCl)



D

Oxidative Stress (0.1 mM H₂O₂)



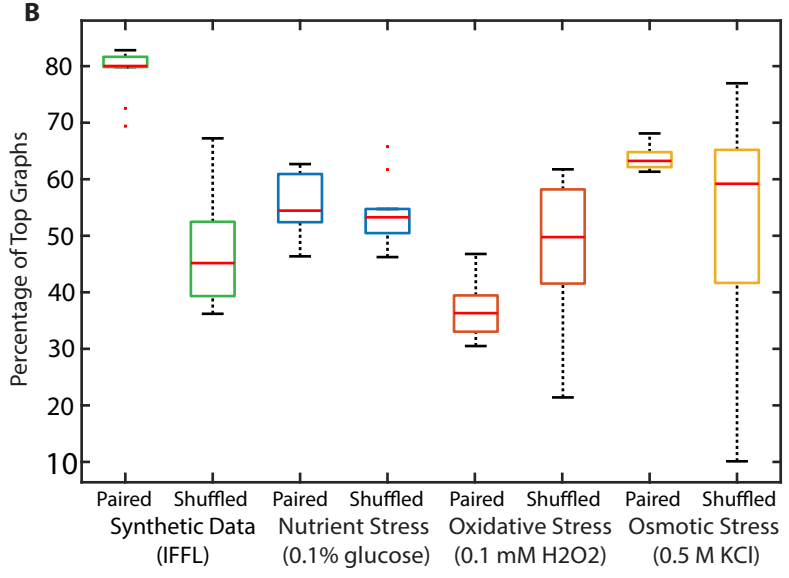
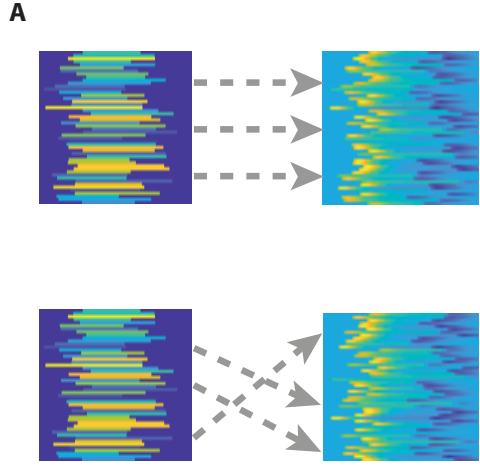


Figure 5

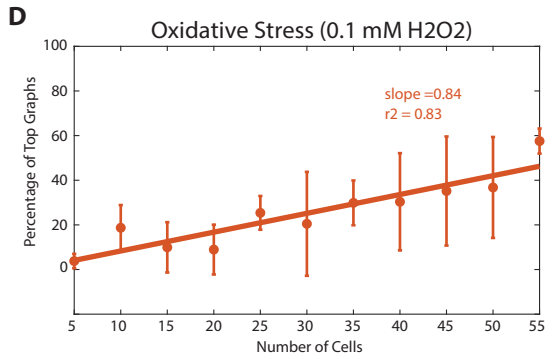
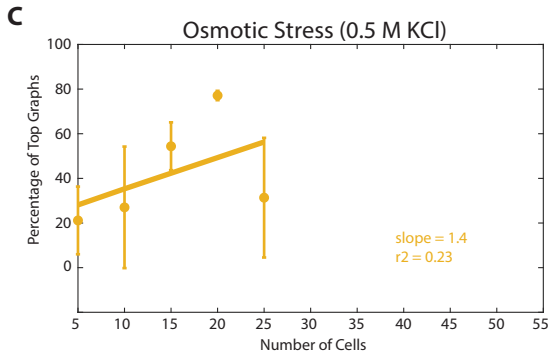
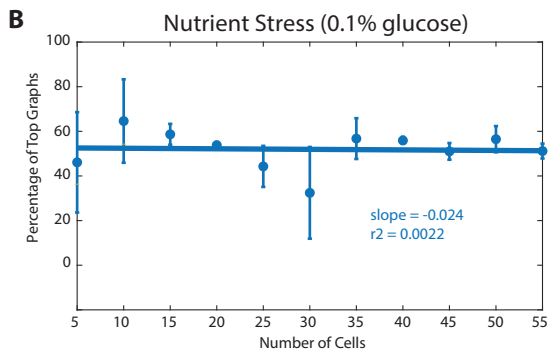
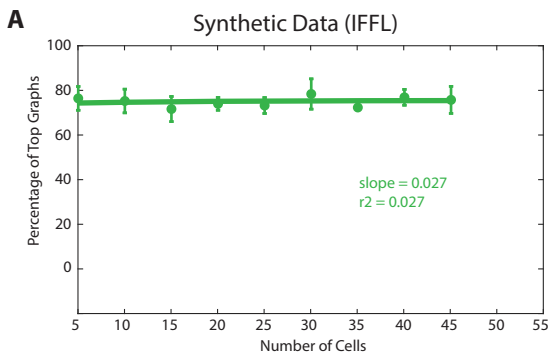


Figure 6

Supplemental Information

Inferring the structures of signaling motifs from paired dynamic traces of single cells

R. A. Haggerty^{1,2,3} and Jeremy E. Purvis^{1,2,3,4,*}

¹Department of Genetics

²Computational Medicine Program

³Curriculum for Bioinformatics and Computational Biology

⁴Lineberger Comprehensive Cancer Center

University of North Carolina, Chapel Hill

116 Manning Drive

11108A Mary Ellen Jones Building

Chapel Hill, NC 27599-7488

*Corresponding Author:

Jeremy Purvis

Mary Ellen Jones Building 11018C, CB#7488

116 Manning Drive

Chapel Hill, NC 27599-7488

jeremy_purvis@med.unc.edu

Materials and Methods:

Data Preparation:

Raw data was arrays of average arbitrary fluorescent units for each cell. The first X timepoints prior to the stimulation of the cells were used as the basal levels, and the average of those timepoints was used to find the foldchange of the cells by setting those to one.

Implementation details can be found in `dataprep.m`

Parameter Bounds:

a is bounded between 0 and 1 because the degradation rate must be a fraction between 0 and 1.

K, which represents the half max of the fold change is bounded between 0 and 100. This covers a broad biological range.

n is bounded between 1 and 4. The hill constant is usually between 1 and 4 for cooperative binding.

Coding:

All code was written in MATLAB 2018a and run on UNC Research Computing's Longleaf cluster's general compute nodes with the following specifications:

147 compute nodes, each with 24 physical cores, 2.50 GHz Intel processors, 30M cache (Model E5-2680 v3), 256-GB RAM, and 2x10Gbps NIC

and

30 compute nodes, each with 36 physical cores, 2.30 GHz Intel processors, 24.75M cache, 754-GB RAM, and 2x10Gbps NIC

ODE Derivation:

Hill Equations:

Hill equations are widely used in biochemistry, as they provide a more accurate model of biological processes than using simple zeroth or first order kinetics. The following are the typical forms of the hill equations for activation and repression:

Activation:

$$\frac{dZ}{dt} = \beta \frac{X^n}{K_{XZ}^n + X^n} - \alpha Z$$

Repression:

$$\frac{dZ}{dt} = \beta \frac{1}{1 + \left(\frac{Y}{K_{YZ}}\right)^n} - \alpha Z$$

Where:

β = Maximal production rate of promoter

K_{XZ} = Halfway activation point of Z by X

K_{YZ} = Halfway repression point of Z by Y

n = Hill Coefficient, cooperativity
 α = Degradation rate of protein

When multiple factors are affecting one node, the core of the equations are multiplied together into a compound equation to get the full effect, while the production rate and degradation rate remain the same. As an example, if Z is activated by X and repressed by Y, the total equation of the effect on Z would be:

$$\frac{dZ}{dt} = \beta \left(\frac{X^n}{K_{XZ}^n + X^n} \right) \left(\frac{1}{1 + \left(\frac{Y}{K_{YZ}} \right)^n} \right) - \alpha Z$$

Autogeneration of compound equations:

Given an array \vec{g} of the connections between each node to Y and Z, with one indicating activation, negative one indicating repression and zero indicating no connection, we can autogenerate the compound equations using \vec{g} to determine which equations to include.

$$\begin{aligned} \frac{dY}{dt} = & \left(\beta \left((eq_{activator})^{\vec{g}(1)^2 \left(\frac{\vec{g}(1)+1}{2} \right)} \left(\frac{1}{(eq_{repressor})^{\vec{g}(1)^2 \left(\frac{\vec{g}(1)-1}{2} \right)}} \right) \right. \right. \\ & * \left((eq_{activator})^{\vec{g}(2)^2 \left(\frac{\vec{g}(2)+1}{2} \right)} \left(\frac{1}{(eq_{repressor})^{\vec{g}(2)^2 \left(\frac{\vec{g}(2)-1}{2} \right)}} \right) \right. \\ & \left. \left. * \left((eq_{activator})^{\vec{g}(3)^2 \left(\frac{\vec{g}(3)+1}{2} \right)} \left(\frac{1}{(eq_{repressor})^{\vec{g}(3)^2 \left(\frac{\vec{g}(3)-1}{2} \right)}} \right) \right) \right) \right) \max(\vec{g}(1:3)^2) - \alpha Y \end{aligned}$$

$$\begin{aligned} \frac{dZ}{dt} = & \left(\beta \left((eq_{activator})^{\vec{g}(4)^2 \left(\frac{\vec{g}(4)+1}{2} \right)} \left(\frac{1}{(eq_{repressor})^{\vec{g}(4)^2 \left(\frac{\vec{g}(4)-1}{2} \right)}} \right) \right. \right. \\ & * \left((eq_{activator})^{\vec{g}(5)^2 \left(\frac{\vec{g}(5)+1}{2} \right)} \left(\frac{1}{(eq_{repressor})^{\vec{g}(5)^2 \left(\frac{\vec{g}(5)-1}{2} \right)}} \right) \right. \\ & \left. \left. * \left((eq_{activator})^{\vec{g}(6)^2 \left(\frac{\vec{g}(6)+1}{2} \right)} \left(\frac{1}{(eq_{repressor})^{\vec{g}(6)^2 \left(\frac{\vec{g}(6)-1}{2} \right)}} \right) \right) \right) \right) \max(\vec{g}(4:6)^2) - \alpha Z \end{aligned}$$

Where:

$$\vec{g} = [x \rightarrow y, y \rightarrow y, z \rightarrow y, x \rightarrow z, y \rightarrow z, z \rightarrow z]$$

With:

0 = no connection

1 = activation

-1 = repression

And:

eq_{activator} = equation for activation of the node

eq_{repressor} = equation for the repression of the node

When the connection is 1, the activator equation is selected for that connection. If it is -1, the repressor equation is selected. If it is 0, the connection becomes one. If all connections are zero, then the selection of the maximum for the connections ensures that the total effect is zero.

Fold Change Equations:

Hill equations are in units of concentration. However, when working with microscopy images, we are working in the realm of arbitrary fluorescent units and are unable to translate those into units of absolute concentration. We are able to get fold changes from a basal steady state. Therefore, it is necessary to get the hill equations into a form where they will work on fold change rather than absolute concentration.

At steady state, the change in Z is zero. As stated above the change in Z is as follows:

$$\begin{aligned}\Delta Z &= \beta f(X)g(Y) - \alpha Z \\ 0 &= \beta f(X_{st})g(Y_{st}) - \alpha Z_{st}\end{aligned}$$

Solve for the steady state of Z:

$$\begin{aligned}\alpha Z_{st} &= \beta f(X_{st})g(Y_{st}) \\ Z_{st} &= \frac{\beta}{\alpha} f(X_{st})g(Y_{st})\end{aligned}$$

The fold change is equal to the current state of Z divided by the steady state of Z:

$$\begin{aligned}Z_{fc} &= \frac{Z}{Z_{st}} \\ \frac{\Delta Z}{Z_{st}} &= \frac{\beta f(X)g(Y) - \alpha Z}{\frac{\beta}{\alpha} f(X_{st})g(Y_{st})}\end{aligned}$$

Simplify:

$$\begin{aligned}\frac{\Delta Z}{Z_{st}} &= \frac{\left(\frac{\alpha}{\beta}\right) \beta f(X)g(Y)}{\left(\frac{\alpha}{\beta}\right) \frac{\beta}{\alpha} f(X_{st})g(Y_{st})} - \frac{\alpha Z}{\frac{\beta}{\alpha} f(X_{st})g(Y_{st})} \\ \frac{\Delta Z}{Z_{st}} &= \frac{\alpha f(X)g(Y)}{f(X_{st})g(Y_{st})} - \alpha \frac{Z}{Z_{st}} \\ \Delta Z_{fc} &= \frac{\alpha f(X)g(Y)}{f(X_{st})g(Y_{st})} - \alpha Z_{fc}\end{aligned}$$

Full Equations:

Putting together the fold change equations with the autogeneration of the compound equations, we get the following full equations:

$$\frac{dY}{dt} = \vec{a}(1) * \left(\left(\left(\frac{\vec{x}^{\vec{n}(1)}}{\vec{k}(1)^{\vec{n}(1)} + \vec{x}^{\vec{n}(1)}}} \right)^{\vec{g}(1)^2 * \left(\frac{\vec{g}(1)+1}{2}\right)} * \left(\frac{\vec{k}(1)^{\vec{n}(1)} + \vec{x}^{\vec{n}(1)}}{\vec{k}(1)^{\vec{n}(1)}} \right)^{\vec{g}(1)^2 * \left(\frac{\vec{g}(1)-1}{2}\right)} \right) \right. \\ \left. * \left(\frac{1}{\vec{k}(1)^{\vec{n}(1)} + 1} \right)^{\vec{g}(1)^2 * \left(\frac{\vec{g}(1)+1}{2}\right)} * \left(\frac{\vec{k}(1)^{\vec{n}(1)} + 1}{\vec{k}(1)^{\vec{n}(1)}} \right)^{\vec{g}(1)^2 * \left(\frac{\vec{g}(1)-1}{2}\right)} \right) \\ * \left(\left(\frac{\vec{y}^{\vec{n}(2)}}{\vec{k}(2)^{\vec{n}(2)} + \vec{y}^{\vec{n}(2)}}} \right)^{\vec{g}(2)^2 * \left(\frac{\vec{g}(2)+1}{2}\right)} * \left(\frac{\vec{k}(2)^{\vec{n}(2)} + \vec{y}^{\vec{n}(2)}}{\vec{k}(2)^{\vec{n}(2)}} \right)^{\vec{g}(2)^2 * \left(\frac{\vec{g}(2)-2}{2}\right)} \right) \\ \left. * \left(\frac{1}{\vec{k}(2)^{\vec{n}(2)} + 1} \right)^{\vec{g}(2)^2 * \left(\frac{\vec{g}(2)+1}{2}\right)} * \left(\frac{\vec{k}(2)^{\vec{n}(2)} + 1}{\vec{k}(1)^{\vec{n}(2)}} \right)^{\vec{g}(2)^2 * \left(\frac{\vec{g}(2)-1}{2}\right)} \right) \\ * \left(\left(\frac{\vec{z}^{\vec{n}(3)}}{\vec{k}(3)^{\vec{n}(3)} + \vec{z}^{\vec{n}(3)}}} \right)^{\vec{g}(3)^2 * \left(\frac{\vec{g}(3)+1}{2}\right)} * \left(\frac{\vec{k}(3)^{\vec{n}(3)} + \vec{z}^{\vec{n}(3)}}{\vec{k}(3)^{\vec{n}(3)}} \right)^{\vec{g}(3)^2 * \left(\frac{\vec{g}(3)-1}{2}\right)} \right) \\ \left. * \left(\frac{1}{\vec{k}(3)^{\vec{n}(3)} + 1} \right)^{\vec{g}(3)^2 * \left(\frac{\vec{g}(3)+1}{2}\right)} * \left(\frac{\vec{k}(3)^{\vec{n}(3)} + 1}{\vec{k}(3)^{\vec{n}(3)}} \right)^{\vec{g}(3)^2 * \left(\frac{\vec{g}(3)-1}{2}\right)} \right) - \vec{y} \right)$$

$$\frac{dZ}{dt} = \vec{a}(2) * \left(\left(\left(\frac{\vec{x}^{\vec{n}(4)}}{\vec{k}(4)^{\vec{n}(4)} + \vec{x}^{\vec{n}(4)}}} \right)^{\vec{g}(4)^2 * \left(\frac{\vec{g}(4)+1}{2}\right)} * \left(\frac{\vec{k}(4)^{\vec{n}(4)} + \vec{x}^{\vec{n}(4)}}{\vec{k}(4)^{\vec{n}(4)}} \right)^{\vec{g}(4)^2 * \left(\frac{\vec{g}(4)-1}{2}\right)} \right) \right. \\ \left. * \left(\frac{1}{\vec{k}(4)^{\vec{n}(4)} + 1} \right)^{\vec{g}(4)^2 * \left(\frac{\vec{g}(4)+1}{2}\right)} * \left(\frac{\vec{k}(4)^{\vec{n}(4)} + 1}{\vec{k}(4)^{\vec{n}(4)}} \right)^{\vec{g}(4)^2 * \left(\frac{\vec{g}(4)-1}{2}\right)} \right) \\ * \left(\left(\frac{\vec{y}^{\vec{n}(5)}}{\vec{k}(5)^{\vec{n}(5)} + \vec{y}^{\vec{n}(5)}}} \right)^{\vec{g}(5)^2 * \left(\frac{\vec{g}(5)+1}{2}\right)} * \left(\frac{\vec{k}(5)^{\vec{n}(5)} + \vec{y}^{\vec{n}(5)}}{\vec{k}(5)^{\vec{n}(5)}} \right)^{\vec{g}(5)^2 * \left(\frac{\vec{g}(5)-2}{2}\right)} \right) \\ \left. * \left(\frac{1}{\vec{k}(5)^{\vec{n}(5)} + 1} \right)^{\vec{g}(5)^2 * \left(\frac{\vec{g}(5)+1}{2}\right)} * \left(\frac{\vec{k}(5)^{\vec{n}(5)} + 1}{\vec{k}(5)^{\vec{n}(5)}} \right)^{\vec{g}(5)^2 * \left(\frac{\vec{g}(5)-1}{2}\right)} \right) \\ * \left(\left(\frac{\vec{z}^{\vec{n}(6)}}{\vec{k}(6)^{\vec{n}(6)} + \vec{z}^{\vec{n}(6)}}} \right)^{\vec{g}(6)^2 * \left(\frac{\vec{g}(6)+1}{2}\right)} * \left(\frac{\vec{k}(6)^{\vec{n}(6)} + \vec{z}^{\vec{n}(6)}}{\vec{k}(6)^{\vec{n}(6)}} \right)^{\vec{g}(6)^2 * \left(\frac{\vec{g}(6)-1}{2}\right)} \right) \\ \left. * \left(\frac{1}{\vec{k}(6)^{\vec{n}(6)} + 1} \right)^{\vec{g}(6)^2 * \left(\frac{\vec{g}(6)+1}{2}\right)} * \left(\frac{\vec{k}(6)^{\vec{n}(6)} + 1}{\vec{k}(6)^{\vec{n}(6)}} \right)^{\vec{g}(6)^2 * \left(\frac{\vec{g}(6)-1}{2}\right)} \right) - \vec{z} \right)$$

Where:

\vec{a} = vector with the degradation rates of y and z

\vec{k} = vector of the halfway activation points

\vec{n} = vector of the cooperativity

\vec{g} = vector of connectivity

\vec{x} = vector of factor x

\vec{y} = vector of factor y

\vec{z} = vector of factor z

Code Implementation:

Please see the included files dY_hill.m and dZ_hill.m to see the implementation of the full equations.