Supplementary Materials for

Accurate Information Transmission Through Dynamic Biochemical Signaling Networks

Jangir Selimkhanov*, Brooks Taylor*, Jason Yao, Anna Pilko, John Albeck, Alexander Hoffmann, Lev S. Tsimring, Roy Wollman

Correspondence to: rwollman@ucsd.edu

This PDF file includes:

1 Materials and Methods

1.1 Cell culture

MCF10a cells (ATCC CRL-10317) were maintained in F12/DMEM media (Life Technologies) supplemented with serum, insulin and hydrocortisol according to (*29*). For the analysis of the ERK pathway, MCF10a constitutively expressed the FRET sensor EKARev-NES (*28*). This sensor has been optimized for dynamic range (*33*) and indeed we see up to 200% increase on the FRET/CFP ratio. Externally delivered EGF (PeproTech Cat. No. AF-100-15) was used to generate EKR response. Analysis of Ca2+ pathway was done by loading cells with 1 μM Fluo-4 using PowerLoad (Life Technologies) loading solution according to manufacturer's instructions. Calcium pathway was activated through externally delivered ATP (Teknova Cat. No. A1204). All experiments were performed in Hepes buffer containing 5 mM KCl, 125mM NaCl, 1.5 mM MgCl2, 1.5mM CaCl2, 50 mM Hepes, 10mM Glucose, pH 7.4, at 37oC.

Before cell activation, MCF10A were cultured overnight in 96 well plate in assay media (as previously described in (*29*). The following day, cells were stained with $0.5\mu M$ Hoechst (Invitrogen Cat. No. 21492) in assay media for 30 min before imaging. For the inhibitor study, indicated doses of UO126 (Cell Signaling Technology) were also added. After 30 min, assay media was replaced with Hepes buffer (supplemented with UO126 in inhibitor study) before the cells were imaged on a prewarmed stage (37C). Before adding EGF (or ATP) cells were imaged for 10 (or 1) minutes to establish a baseline. To correct for possible cell response to the shear forces generated from pipetting during media switch, Hepes buffer of the same volume as stimuli was added to the control wells.

RAW267.4 cells ATCC TIB-71) were maintained in a DMEM (CellGro 10-013) medium supplemented with 10% fetal bovine serum, 20mM HEPES, and 1x penicillin/streptomycin. Analysis of NFκB was done by sequentially transducing RAW cells to create a stable line with lentivirus vectors containing EYFP-RelA (driven by the native RelA promoter) and H2B-mCherry. Double-stable lines were made by successive selection, then further FACS-enriched. Cells were used in imaging from passage 16 to passage 20, then discarded. 20 hours prior to experiment start, cells were replated in Ibidi 8-well -slide at a density of 50,000/sq. cm. 2 hours before the experiment, 1/3 of the total media volume was drawn off and mixed with stimulus (LPS, Sigma B5:055 Cat.No. L2880), which was then injected into the chamber precisely at experiment start.

1.2 Image acquisition and analysis of single cell responses

Analysis of information transmission of dynamic vector responses can only be done from single live cell measurements. Furthermore, the analysis algorithm developed in this report is based on the estimation of the probability density for each cellular response based on its degree of similarity to the response of cells with very similar responses. While the algorithm corrects for sample size using a jackknife procedure to eliminate possible bias, we have empirically found that robust estimation substantially benefits from a large sample size. Therefore, a key technical aspect of this work is the ability to collect very large numbers of single cell responses. To that goal we have optimized both the image acquisition and analysis to increase the sample size without compromising on measurement accuracy.

Below we outline the key steps that were taken to allow the collection of the signaling responses of 910,121 cells overall.

1.2.1 Image acquisition of MCF10A cells Ca2+ and ERK pathways

The image acquisition was done in a 96-well plate format (Coaster 3094) on a Nikon Ti microscope using a 10x 0.45NA Plan Apo objective. To increase the number of cell images the camera that was used (Zyla 5.5 sCMOS camera, Andor) has a large sensor size (21.8 mm). Furthermore, an additional de-magnifying 0.7x optovar was in the light path. The microscope was controlled through custom software written in Matlab that uses the open-source micro-manager (*30*) as the driver layer that controls the hardware. Image acquisition was completely automated and relied on the Nikon Perfect Focus (PFS) to perform multi-well imaging while maintaining focus. The custom software performed periodic checks to verify that focus is maintained. To speed up image acquisition and minimize photo-damage and photo-bleaching, the acquisition rate of each channel was determined based on the need, where the channels that were used for measurement were acquired at every time point, whereas the channels that were used for image segmentation were acquired at 8x slower rate. To minimize background, cells were imaged in Hepes buffer for short period of time (< 1 hour).

1.2.2 Image acquisition of RAW 267.4 cells NFκB pathway

Cells were incubated and imaged every 5 minutes using a Zeiss AxioVert fitted with a 40x oil-immersion objective, LED fluorescence excitation, and CoolSnap HQ2 camera. We collected DIC, mCherry, and YFP images over 18.5 hours at 12 stage positions per experimental condition.

1.2.3 Image analysis of MCF10A Ca2+ and ERK pathways

Fully automated image analysis pipelines are essential to maintain throughput of data collection (*39*) and extract single cell response profiles from the raw images. An overview of the analysis of MCF10a data is shown in Fig. S1. The analysis was based on three stages: image correction (I), image segmentation and tracking (II), and cell measurements (III).

I. Image Correction. Images were corrected for three well known issues during acquisition: uneven illumination, background fluorescence, and stage jitter/drift. Uneven illumination is a result of the large imaging area. To correct for uneven illumination, each combination of channel and objective on the microscope was calibrated to estimate the spatial distribution of excitation intensity using wells containing an Alexa dye that matched the excitation and emission filters for that channel. Images were corrected for systemic shift (Zyla 5.5 adds 100 to all pixels to avoid negative values) and were divided by the estimated value of illumination. The resulting images had flat intensity as was verified by looking at the average intensity of cell stained with CellTracker dyes at multiple wavelengths. To remove any signal not resultant from cellular fluorescence we further estimated and subtracted the remaining background. Background estimation was done by first dividing the image into blocks of approximately 128x128 pixels each. In each of the blocks the mini-

mum intensity was measured to produce a 16x16 matrix that was then resized using bilinear interpolation to produce the estimated background. To correct for possible drift over time images were registered to the initial image. To register the image we identified the translation matrix using normalized cross correlation analysis of the center of the Hoechst image as it provided the most robust signal. Since Hoechst was imaged at lower frequency than other channels, drift was assumed to be of constant speed and channels that were imaged more frequently than Hoechst were corrected accordingly.

II. Image segmentation and tracking. Images were segmented to first identify the nuclei based on Hoechst staining. First, each image was converted into a binary mask using Otsu's method (*37*) on the log transform of the intensity. Since cells were imaged at confluence, separation of nuclei was not complete, i.e. there were situation where nuclei were overlapping. Therefore we utilized a marker controlled watershed approach. The nuclei image was first eroded using a 5x5 local neighborhood as the structural element. The eroded image intensity was higher at the center of the nuclei than at its edges. Further smoothing the eroded image with a 2D Gaussian filter converted each nuclei into a local hill. Simple peak detection, followed by rejection of peaks of small magnitude allowed us to find a single marker pixel per nuclei. Geodesic distance transformation from the marker pixels followed by watershed operation over the binary image created from Hoescht staining segmented overlapping nuclei into individual nuclei with separate regions in the label matrix. Figure S2 shows a region of the nuclei image going through several of the above transformation. In the case of ERK pathway, additional segmentation step was taken to assign cytoplasmic pixels to each nucleus based on proximity (Fig. S3). The proximity measure allowed accurate assignment of more than 50% of the pixels in each cell. Using median statistics for the measurements of intensity, together with the fact that MCF10a cells grown to confluence show relatively constant intensity across the cell area, allowed for accurate measurement of individual cell intensity. To make sure this was the case, cells whose cytoplasmic EKARev intensity was too small compared to the nuclear signal or showed high variability were excluded from the analysis. Cells where tracked over time to allow us to create the multivariable vector responses for each cells that represent their signaling response. We took advantage of the fact that MCF10a grown to confluence are grid-locked and do not migrate. Since the images were registered using nearest neighbor criteria, we achieved very good matching of cells between frames. Furthermore, over the period that we imaged for (1 hour) the percent of cells that divide was small. Cells whose temporal trajectories were not compete (i.e. were lost during tracking) were removed from any additional analysis.

III. Cell measurements. For Ca2+ measurements, we used the fluorogenic indicator Fluo-4. Loading 1 μM of Fluo-4 allowed for very high signal to noise levels and accurate Ca2+ measurements. Since Fluo-4 and Ca2+ increase in the cytoplasm as well as the nucleus, we used the nuclear labels segmented based on Hoescht staining as described above as the region for measurement. Acquisition of Ca2+ data was done one well at a time for 15 min each.

For ERK measurements, the median FRET signal per cell was calculated based on the measured FRET channel (CFP excitation, YFP emission) corrected for bleed-through from Cyan and Yellow channels based on previous calibration of bleed-through in our system divided by the CFP intensity in the Cyan channel. We found that the expressed CFP contributes substantially to the FRET channel whereas YFP had only minimal contribution

($\sim 2\%$). To increase imaging throughput and minimize photo-damage to the cells, the Yellow channel was imaged at 8-fold lower frequency. The missing values for bleed-through correction were estimated by interpolation. This was partially justified by the very low (2%) bleed-through measured during calibration. Overall the experimental and analysis pipeline described above was optimized for accuracy and throughput and it allowed us to measure very large samples of the dynamic Ca2+ and ERK responses in thousands of cells per experiment.

1.2.4 Image analysis of RAW 267.4 cells NFκB pathway

Image analysis consisted of 5 major steps: cell identification (phase/DIC) (I), nucleus identification (II), nucleus tracking/error correction (III), shape-based segmentation (IV), and NFκB measurement (V).

I. Cell identification. Vertical and horizontal Sobel edge transformations were calculated of cells in brightfield images (Fig. S4A-B): a custom speckle-noise-based threshold of the combined edge magnitude image formed the basis for foreground-vs-background determination (Fig. S4C). To better identify spread cells with weak outer edges, we used the preliminary masked Sobel image to "mark" edges detected in Gaussian-smoothed images at multiple scales - marked edges were expanded, then we intelligently filled holes and gaps to form a final mask (Fig. S4D).

II. Nucleus identification. Due to cellular heterogeneity in both morphology and expression levels of the H2B-mCherry label, we employed a separate algorithm to identify the nucleus in H2B-mCherry-labeled macrophages. The search space for nuclei was constrained to cell-identified regions from step I. Nuclei were then identified in a two-step process: first, using the Sobel edge-magnitude transformed image, we iteratively stepped down the threshold, searching for appropriately round nuclear-shaped objects (Fig. S4E), and stopping the search in the affected area (identified using a watershed transformation on the nuclear image). Second, weaker nuclei were identified after searching and ranking pixels in each remaining watershed region (Fig. S4F)) - appropriately concentric regions were pulled out and combined with the first set of nuclei. In the final step, we discerned between pieces of individual nuclei that needed to be combined, versus overlapping nuclei that needed to be separated, by merging all possible combinations of overlapping objects, and minimizing the total morphological compactness and eccentricity of the result. In this way, we could accurately identify the borders of objects possessing a wide range of intensities, where a single threshold would have led to overestimation of bright objects, or loss of dim ones.

III. Nucleus tracking/error correction. To maintain cell lineages, we applied a tracking algorithm across 7 frames simultaneously. All possible links were computed between an object and its viable candidate object in subsequent images. These links were then resolved based on minimizing distance and morphological similarity (area and perimeter) of the objects, and combined into "blocks", the list of each object's likely location in each frame. After making and updating this list of blocks, we used the list to make decisions about cell fate: division, death, and drift in or out of the frame. Using a stack size of 7 frames made it possible to identify and correct errors in subsets of frames, such as false positives (Fig. S5A) or falsely combined nuclei (Fig. S5B-C), voting on fates using information from

previous and subsequent frames.

IV. Shape-based segmentation. Macrophages are extremely heterogeneous in size and shape, meaning that existing segmentation techniques were ill-suited to drawing boundaries between clusters of cells (Fig. S5E). To aid segmentation, we employed a shape-based approach where the morphological skeleton was calculated (Fig. S5F) and pruned to segments connecting nuclei of cells. A distance transform calculated along these pruned-skeleton segments shows distinct local maxima (Fig. S5G) that generally correspond to inflection points where one cell touches another. We used these inflection points to inform the segmentation decision, which was ultimately made by the "propagate" algorithm (*32*).

V. NFκB measurement. Raw nuclear intensity traces could not be used directly, as macrophages experience strong morphological changes in response to LPS. To account for these changes, we computed both the median value for the pixels contained in the nucleus of each cell (Fig. 6A), and a value representing the cytoplasm - we chose the upper mode of the generally bimodal distribution derived from cellular cytoplasmic regions, which corresponds to the brighter portion of the cell close to the nucleus. This cytoplasmic function is insensitive to nuclear translocation events (Fig. 6B), but does show pronounced changes that were assumed to be indicative of the overall morphological shape of each cell. We used the cytoplasmic trajectory to fit the changes in nuclear values of each cell - cells were assumed to be in an off/zero state at time t=0, and at extremely late timepoints (t=12+ hrs). Baseline cytoplasmic/nuclear values from these two timepoints were used to scale the (smoothed) cytoplasmic trajectory of each cell, such that this scaled cytoplasmic trajectory could then be subtracted from the nuclear trajectory to eliminate the nuclear baseline change (Fig. S6C). Corrected nuclear traces were then divided by initial cytoplasmic values, thereby giving the fold-change in nuclear intensity over time.

2 Information transfer estimation from experimental data

2.1 Algorithm derivation

To derive the algorithm for estimation of information transfer, we first considered the type of experimental data that we have acquired, which will guide our general approach. In our signal transduction networks, the input signal S is defined by m discrete levels of extracellular ligand concentration ($S = [s_1, s_2, ..., s_m]$). For each input signal s_i we have n_i output protein trajectories ($R_i = [r_{i1}, r_{i2}, ..., r_{in_i}]$), with each trajectory occupying a single point in continuous Euclidean space of dimension d, where d is the number of time points in each output trajectory. Combined, we have $N = \sum_i^m n_i$ trajectories in our response R array. The general breakdown of the data is as follows:

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_i \vdots \\ s_m \end{bmatrix}, R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_i \\ \vdots \\ R_m \end{bmatrix}, s_i \rightarrow R_i = \begin{bmatrix} r_{i1} \\ r_{i2} \\ \vdots \\ r_{ij} \\ \vdots \\ r_{in_i} \end{bmatrix}, r_{ij} = [r_{ij,1}, r_{ij,2}, \ldots, r_{ij,d}]$$

To estimate the information transfer (I) between an input (S) and an output (R) using well known formula

$$I(R; S) = H(R) - H(R|S). \tag{2.1}$$

we need to calculate Shannon entropies $H(R)$ and $H(R|S)$. The general scheme of our approach is shown in Fig. S7. First, given that our data is continuous, we need to define how we will estimate these entropies. For a continuous probability density $f(x)$ of some observable X, the Shannon entropy is defined as differential entropy

$$H_{\text{diff}}(X) = -\int_{-\infty}^{\infty} f(x) log_2(f(x)) dx. \tag{2.2}$$

Following change of variable of integration, Equation 2.2 becomes

$$H_{\text{diff}}(X) = -\int_0^1 log_2(f(x)) dy. \tag{2.3}$$

where $y = \int_{-\infty}^{x} f(t)) dt$ is the cumulative probability density. We can estimate y by the cumulative probability distribution of N_x observations using

$$H_{\text{diff}}(X) = -\sum_{j=1}^{N_x} \delta_j log_2(f(x_j)), \tag{2.4}$$

where δ_j is the probability of observing x_j, $P(X = x_j)$.

Using Equation 2.4 as our basis, we will now illustrate how to obtain $H_{\text{diff}}(R|S)$ and $H_{\text{diff}}(R)$, given that our experimental data only contains conditional probabilities of a responses.

For the conditional case, $H_{\text{diff}}(R|S)$, since all n_i responses in R_i are equally likely, $\delta_j = \frac{1}{n_i}$, we can estimate probability density of a single response r_{ij} directly from all the other responses to $S = s_i$,

$$H_{\text{diff}}(R_i|S = s_i) = -\sum_{j=1}^{n_i} \frac{1}{n_i} log_2(f(R_i = r_{ij}|S = s_i)). \tag{2.5}$$

Here $f(R_i = r_{ij}|S = s_i)$ represents the probability density of response r_{ij} in R_i given all the other responses r_i (in R_i) to the signal $S = s_i$. We will explain how to estimate $f(R_i = r_{ij}|S = s_i)$ later in the derivation. With the probability of a given signal, $q_i = P(S = s_i)$, we can then sum the conditional entropies of each signal to get overall conditional entropy,

$$H_{\text{diff}}(R|S) = \sum_{i=1}^{m} q_i H_{\text{diff}}(R_i|S = s_i) = -\sum_{i=1}^{m} q_i \sum_{j=1}^{n_i} \frac{1}{n_i} log_2(f(R_i = r_{ij}|S = s_i)). \tag{2.6}$$

The case of estimating $H_{\text{diff}}(R)$ requires special attention, since we do not have access to non-conditional probabilities of responses. The difficulty arises from the fact that our estimate of non-conditional density of a single response, $f(R = r)$ is dependent on the probability of the input signals that generated all other responses. One approach is to estimate probability density that a given response r occurred in response to a given input signal (s_w), for each of the signals in S, by effectively placing that response into R_w and estimating the probability density for r is if it were also a response to s_w. Using total probability, for every response r in R, we can estimate the probability density within each set of responses R_w and sum over m such densities multiplied by the probability q_w of the signal that generated responses in R_w, as follows

$$f(R = r) = \sum_{w=1}^{m} q_w f(R = r|S = s_w). \tag{2.7}$$

Plugging 2.7 into 2.4, we get

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m} \sum_{j=1}^{n_i} \delta_{ij} log_2(f(R = r_{ij})). \tag{2.8}$$

The key difference between $f(R_i = r_{ij}|S = s_i)$ and $f(R = r_{ij})$, is that the former calculates the conditional probability density of a response among all other responses to the same signal, while the latter estimates non-conditional probability density of the response, combining law of total probability and conditional probability density of the response belonging to each of the subsets of responses R_i. To get δ_k we must account for the different probabilities associated with observing responses from different input signals and the number of responses n_i obtained for each input signal:

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2(f(R = r_{ij})). \tag{2.9}$$

Possibly a more intuitive presentation of the above formula using matrix representation of our data would be

$$f(R) = \begin{bmatrix} \left.\begin{matrix} f(R=r_{11}) \\ \vdots \\ f(R=r_{1j}) \\ \vdots \\ f(R=r_{1n_1}) \end{matrix}\right\} n_1 \\ \left.\begin{matrix} f(R=r_{21}) \\ \vdots \\ f(R=r_{2n_2}) \end{matrix}\right\} n_2 \\ \vdots \\ f(R=r_{ij}) \\ \vdots \end{bmatrix} = \left(\begin{bmatrix} f(r_{11}|S=s_1) & \cdots & f(r_{11}|S=s_m) \\ \vdots & \ddots & \vdots \\ f(r_{1j}|S=s_1) & \cdots & f(r_{1j}|S=s_m) \\ \vdots & \ddots & \vdots \\ f(r_{1n_1}|S=s_1) & \cdots & f(r_{1n_1}|S=s_m) \\ f(r_{21}|S=s_1) & \cdots & f(r_{21}|S=s_m) \\ \vdots & \ddots & \vdots \\ f(r_{ij}|S=s_1) & \cdots & f(r_{ij}|S=s_m) \\ \vdots & \ddots & \vdots \\ f(r_{ij}|S=s_1) & \cdots & f(r_{ij}|S=s_m) \\ \vdots & \ddots & \vdots \end{bmatrix} \cdot \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{bmatrix} \right)$$

$$H_{\text{diff}}(R) = -\left[\overbrace{\frac{q_1}{n_1}, \frac{q_1}{n_1}, \ldots, \frac{q_1}{n_1}}^{n_1}, \frac{q_2}{n_2}, \ldots, \frac{q_i}{n_i}, \ldots \right] \cdot \begin{bmatrix} \left.\begin{matrix} log_2(f(R=r_{11})) \\ \vdots \\ log_2(f(R=r_{1j})) \\ \vdots \\ log_2(f(R=r_{1n_1})) \end{matrix}\right\} n_1 \\ log_2(f(R=r_{21})) \\ \vdots \\ log_2(f(R=r_{ij})) \\ \vdots \end{bmatrix}$$

Now that we have formulas for $H_{\text{diff}}(R)$ and $H_{\text{diff}}(R|S)$, we need to estimate the probability densities $f(R = r|S = s_i)$ in 2.7 and $f(R_i = r_{ij}|S = s_i)$ in 2.5. This can be accomplished with the k-nearest neighbor (KNN) estimator,

$$f(x_j|X) = \frac{k}{N_x V_d z(x_j|X)_k^d} \tag{2.10}$$

where

$$V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)} \tag{2.11}$$

is the the volume of a unit sphere of dimension d (also dimension of x_j), N_x is the number of x_j in X, and $z(x_j)_k$ is the Euclidean distance to the kth nearest neighbor in X from x_j (*35*).

Applying this estimator to Equations 2.5 and 2.7, we get

$$H_{\text{diff}}(R|S) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2\left(\frac{k}{n_i V_d z(r_{ij}|R_i)_k^d}\right). \tag{2.12}$$

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2(\sum_{w=1}^{m} q_w \frac{k}{n_w V_d z(r_{ij}|R_w)_k^d}) \tag{2.13}$$

where $z(r_{ij}|R_i)_k$ is the distance from response r_j in R_i to the kth nearest neighbor in R_i, while $z(r_{ij}|R_w)_k$ is the distance from response r_{ij} in R_i to the kth nearest neighbor in R_w (Fig. S8). These formulas are very similar to Eqn. 20 in Kraskov *et al* (*34*), with the main difference coming from small error in our estimation of $\psi(k)$ and $\psi(n)$, where $\psi(x)$ is the digamma function. These estimation errors cancel out, however, when we calculate the mutual information by taking the difference between conditional and non-conditional entropies.

Without the knowledge of q_i, we are unable to estimate the information transfer $I(R;S)$ using Equations 2.13 and 2.12. However, the maximum information transfer (C) can be calculated with

$$C(R;S) = \max_{Q}\{I(R;S)\}, \tag{2.14}$$

where $Q = [q_1, q_2, ..., q_m]$, such that $\sum_{i=1}^{m} q_i = 1$ and $q_i \geq 0$ (Elements of Information Theory, 2nd ed.). This corresponds to the maximum possible information transfer between input S and output R.

2.2 Information transfer estimate validation

To test the accuracy of our algorithm, we decided to compare our estimate to the analytical value of mutual information. To get the analytical value, we considered 2-input system with the outputs being multivariate Gaussians with all sigmas equal to 1 and separation between the two multivariate Gaussians equal to 2 in one of the dimensions and 0 in all other dimensions (Fig. S9 A). To get the true value of conditional entropy we numerically integrated the analytical probability density in the dimension where the two multivariate Gaussians are separated and used the well known formula for the entropy of a multivariate Gaussian in the other directions: $H = \frac{1}{2} log_2(2\pi e|\Sigma|)$, where $|\Sigma|$ is the determinant of the covariance matrix. For our estimates, we used $k = 10$ for KNN calculation and considered two sizes of wells ($N = 500$ and $N = 4000$) corresponding to the number of cells we looked at experimentally. First, we looked at the effect of dimension of our estimate of mutual information up to dimension of 15 (Fig. S9 B). We find pretty good agreement between our estimate and the true value of mutual information. The absolute error of mutual information is much less than that of conditional and non-conditional entropy estimates. (Fig. S9 C). As expected there appears to be a bias in estimation of entropies. However, this bias results in less than 10% error in mutual information estimation.

Since our estimate introduces a new approach to estimation of non-conditional probability, we wanted to make sure that the algorithm is not affected by different size wells. To this effect, for dimension 10, we ran our algorithm for varying sized of well 1 (Fig. S9 D). There appears no bias in this estimate, except for inaccuracies due to small sample size in bin 1. Next, we considered the separation between the two multivariate Gaussians of dimension 10 (Fig. S9 D). Again, we find good agreement between our estimate and the true value of mutual information. Our algorithm appears to slightly overestimate MI for

small separation and underestimate for large separation. These effect do not appear to be significant. Finally, considering different variances in the first dimension of D-10 Gaussians (separation of 2 in the same dimension), we find that our algorithm underestimates MI in cases where one of the dimensions appears to have very low variance (Fig. S9 F).

To calculate the maximum information transfer, or channel capacity (CC), we used fmincon function in Optimization Toolbox^{TM} (MATLAB®) to optimize over possible q_i. We compared our "bin-less" algorithm in the case of one dimensional Gaussians against "binned" algorithm described by Cheong *et al* (*12*), for increasing separation between the two Gaussians, again finding very comparable results (Fig. S10A). To correct for sample size bias, we adapted jackknife sampling procedure similar to Cheong *et al* (*12*). Specifically, we computed information transfer (CC) for sampled fractions of the data ($60\% - 100\%$). Then, plotting CC relative to the inverse of sample size (defined as faction of all available trajectories), we fitted a straight line to obtain the y-intercept corresponding to the information transfer for infinite sample size (Fig. S10B). Statistical analysis of jackknife shows that our sampling leads to very small variability in channel capacity estimate (Fig. S11).

3 Information-theoretic analysis of dynamic signaling

3.1 Signal-to-Noise Ratio (SNR) in single-cell experiments

Suppose that the signaling pathway is characterized by the joint distribution of the scalar input S and the scalar output R, $P(R,S)$. Let us first compute the non-dimensional signal-to-noise ratio (SNR) according to the experimental algorithm (see Section 4.3.1). in which the observed output "signal" $\bar{R}$ is found by averaging responses from multiple cells within the same well, i.e. driven by the same input S. Therefore, the signal is given by the formula:

$$\bar{R} = \int dR\, RP(R|S) \tag{3.1}$$

where $P(R|S)$ is the conditional distribution of cellular responses given the input signal S, $P(R|S) = P(R,S)/P(S)$. The average value of output signals for all input signals S

$$\langle \overline{R} \rangle = \int dS\, P(S) \int dR\, RP(R|S) = \iint dR dS\, RP(R,S) \tag{3.2}$$

We define the signal magnitude as the variance of $\bar{R}$ around its mean value $\langle \overline{R} \rangle$ over all possible input values S:

$$\sigma_r^2 = \int dS\, P(S) \left[\int dR\, RP(R|S) \right]^2 - \left[\iint dR dS\, RP(R,S) \right]^2 \tag{3.3}$$

The noise magnitude is defined as the average width of the distribution of cellular responses for a given input S, averaged over S. Experimentally it is computed by first finding a variance of outputs R within the same well, and then averaging this variance over all wells. Therefore, mathematically it can be defined as

$$\sigma_n^2 = \int dS\, P(S) \left[\int dR\, R^2 P(R|S) - \left[\int dR\, RP(R|S) \right]^2 \right] \tag{3.4}$$

Then the signal-to-noise ratio SNR is given by the fraction σ_r^2/σ_n^2.

3.2 Intrinsic noise and extrinsic variability

Generally, the response R within a single cell is a sum of the deterministic part $g(S,\mathbf{P})$ that depends on the input S and the K-dimensional vector of parameters $\mathbf{P}$, and the zero-mean fluctuating component ξ that we will call *intrinsic noise*

$$R = g(S,\mathbf{P}) + \xi \tag{3.5}$$

Note that intrinsic noise distribution may also depend on the input signal and the system parameters. Every cell in a population shares the same functional structure of the signaling function $g(s,\mathbf{P})$, but the parameters $\mathbf{P}$ are *a priori* unknown and vary from cell to cell. We associate this uncertainty with *extrinsic noise*.

If the input signal and cell parameters deviate weakly from their respective means, we can linearize the system response function with respect to deviations of the input, parameters, and the output from their respective mean values:

$$r = \frac{\partial g}{\partial S} s + \sum_{k=1}^{K} \frac{\partial g}{\partial P_k} p_k + \xi \tag{3.6}$$

Here $r = R - \langle \overline{R} \rangle, s = S - \langle S \rangle, p_k = P_k - \langle P_k \rangle$. We assume for simplicity that s, p_k, and ξ are independent Gaussian random variables with variances $\sigma_s^2, \sigma_{p_k}^2, \sigma_\xi^2$. Then according to (3.3),(S21), we obtain the following formulas for the signal magnitude:

$$\sigma_r^2 = \left(\frac{\partial g}{\partial S} \right)^2 \sigma_s^2 \tag{3.7}$$

extrinsic noise magnitude

$$\sigma_e^2 = \sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k} \right)^2 \sigma_{p_k}^2 \tag{3.8}$$

the total noise magnitude

$$\sigma_n^2 = \sigma_e^2 + \sigma_\xi^2, \tag{3.9}$$

and the signal-to-noise ratio

$$SNR \equiv \frac{\sigma_r^2}{\sigma_n^2} = \frac{\left(\partial g / \partial S \right)^2 \sigma_s^2}{\sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k} \right)^2 \sigma_{p_k}^2 + \sigma_\xi^2} \tag{3.10}$$

We can also introduce the Intrinsic-to-Extrinsic Ratio (IER)

$$IER \equiv \frac{\sigma_\xi^2}{\sigma_e^2} = \frac{\sigma_\xi^2}{\sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k} \right)^2 \sigma_{p_k}^2} \tag{3.11}$$

3.3 Information capacity of scalar signaling

Let us compute the mutual information between an input S and a single output measurement R. According to the standard definition (*31*), the mutual information is given by the formula

$$I(R;S) = \int dS dR \; P(R,S) \log_2 \frac{P(R,S)}{P(R)P(S)} \tag{3.12}$$

For Gaussian distributions of R and S this general expression reduces to

$$I(R;S) = \frac{1}{2} \log_2 \left(\frac{\sigma_R^2}{\sigma_{R|S}^2} \right) \tag{3.13}$$

where σ_R is the overall variance of responses and $\sigma^2_{R|S}$ is the variance of the responses conditioned by the signal, and then averaged over all signals:

$$\sigma^2_R = \iint dSdR\, R^2 P(R,S) - \left[\iint dSdR\, RP(R,S)\right]^2 \tag{3.14}$$

$$\begin{aligned}\sigma^2_{R|S} &= \int dS P(S)\left[\int dR\, R^2 P(R|S) - \left[\int dR\, RP(R|S)\right]^2\right] \\ &= \iint dSdR\, R^2 P(R,S) - \int dS\, P(S)\left[\int dR\, RP(R|S)\right]^2. \end{aligned} \tag{3.15}$$

Using (3.6) again, we arrive at the following expressions

$$\sigma^2_R = \sigma^2_r + \sigma^2_n, \quad \sigma^2_{R|S} = \sigma^2_n \tag{3.16}$$

Substituting these in Eq. (3.13), we obtain the standard formula

$$I(R;S) = \frac{1}{2}\log_2\left(1 + \frac{\sigma^2_r}{\sigma^2_n}\right) = \frac{1}{2}\log_2(1+SNR) \tag{3.17}$$

Note that the scalar mutual information depends only on the overall SNR and is independent of IER.

3.4 Information capacity of vector signaling

We consider a vector signaling pathway as an input-output system that has a single scalar input S and generally multi-dimensional vector of outputs $\{R_1, ..., R_N\}$. These outputs may represent either values of a single variable taken at different time points, if the signaling cascades exhibits a non-trivial dynamic behavior in response to the input signal, or different physical variables (e.g. concentrations of multiple transcription factors activating downstream circuits), or possibly both (time courses of multiple output variables). We assume that system depends on K internal parameters $\{P_1, ..., P_K\}$ which can vary from cell to cell however remain fixed over the time (or, more generally, among multiple measurements) for the duration of a particular signaling event. We also take into consideration that every output variable is perturbed by stochastic fluctuations, and that the downstream cascades may measure the output with a finite precision. In the following, for simplicity we assume that these sources of intrinsic noise generate additive and uncorrelated stochastic contributions $\{\xi_1, ..., \xi_N\}$. This system can be written as

$$\mathbf{R} = \mathbf{g}(S, \mathbf{P}) + \boldsymbol{\xi} \tag{3.18}$$

where $\mathbf{R} \in \mathbb{R}^N$ is the vector of outputs, $\mathbf{P} \in \mathbb{R}^K$ is the vector of parameters, and $\boldsymbol{\xi} \in \mathbb{R}^N$ the vector of intrinsic noise which we assume to be Gaussian and statistically independent among different components $\langle \xi_i \xi_j \rangle = \sigma^2_{\xi i}\delta_{ij}$.

The fidelity of the signaling pathway can be characterized by the amount information about the input that is contained in the response, or the mutual information between $\mathbf{S}$ and $\mathbf{R}$. This information is generally limited not only by the intrinsic noise, but also (and often

even stronger) by the uncertainty of the system parameters $\mathbf{P}$, which may depend on the cell size, microenvironment, and other extrinsic, epigenetic factors.

In this section we will demonstrate quantitatively that multi-dimensional output can reduce both types of uncertainty and increase the mutual information for the same SNR. Importantly, while the effect of intrinsic noise can only be diminished gradually (logarithmically with the number of measurements), the uncertainty related to the unknown system parameters can be reduced dramatically or even completely eliminated with a finite number of independent output channels.

3.4.1 Dynamic measurements: extrinsic noise only

Let us first ignore the intrinsic noise completely, then the input-output relationship (3.18) has a form of a system of N deterministic algebraic equations for $K+1$ unknowns $S, P_1, ..., P_K$:

$$R_i = g_i(S, \mathbf{P}), \quad i = 1, ..., N \tag{3.19}$$

From this interpretation it is clear that for sufficiently large $N \geq K+1$ number of non-redundant equations, this system has at most a discrete set of solutions (not a continuous family) for both the system parameters and the input signal. Evidently, knowledge of these discrete solutions corresponds to an infinite amount of information about the input signal. As an illustration, in Fig. S12 we show the case of single parameter ($K = 1$). Then the output R_1 can be described by a two-dimensional surface in 3-dimensional space (S, P_1, R_1), where for a measured R_{1*} the input S and the parameter P_1 form a two-dimensional curve as an intersection of that surface and plane $R_1 = g_1(S, P_1)$. Thus, without knowledge of P_1, the input S cannot be determined uniquely. However, a second measurement, $R_2 = g_2(S, P_1)$ forms a different two-dimensional curve on the plane (S, P_1). The intersection of these two curves yields a single (or possibly multiple but still discrete) value of the input S that means infinite mutual information between S and R.

To re-iterate, multiple measurements $R_1, ..., R_N$ can be either different outputs of the signaling cascade, or different time points of a single output, provided that the output signal is *dynamic* and the time points are separated by sufficient time so that different time points provided sufficiently different input-output relations (3.19) for the accurate root-finding. A similar procedure is known in the nonlinear time series analysis as *time-delay embedding* (*38*) that according to Takens theorem and its generalizations allows one to unfold the "attractor" of a dynamical system and determine its parameters. While solving the nonlinear system (3.19) may prove difficult or impossible in cases when the structure of the pathway is poorly understood, this argument indicates that using multiple time point measurements can drastically increase the amount of information about the input signal in case of complex dynamical pathways. In the next section we illustrate this by computing the mutual information between a scalar input and a multidimensional output for a linear signaling cascade model.

3.4.2 Dynamic measurements: extrinsic and intrinsic noise

Now we return to the case when measurements are contaminated by both extrinsic and intrinsic noise. A quantitative characterization of the information gain from multi-

dimensional data can be achieved by linearization of the general nonlinear input-output relationship (3.18). Let us assume here that the dynamic ranges of the signal and parameters are small, such that the transfer function $\mathbf{g}(\cdot)$ can be linearized in the operational domain near the mean values S_0 and $\mathbf{P}_0$. Introducing deviations of these quantities from their mean values, $s = S - S_0$, $\mathbf{p} = \mathbf{P} - \mathbf{P}_0$ and the deviation of the output $\mathbf{r} = \mathbf{R} - \mathbf{R}_0$, we can re-write Eq. (3.18) in the following form

$$\mathbf{r} = \mathbf{J}_0 s + \hat{\mathbf{J}} \cdot \mathbf{p} + \boldsymbol{\xi} \tag{3.20}$$

where $\hat{\mathbf{J}}$ is the $N \times K$ Jacobian matrix of $\mathbf{g}$ with components $J_{ik} = \partial g_i / \partial p_k$, and $\mathbf{J}_0 = \partial \mathbf{g} / \partial S$ taken at $S = S_0$, $\mathbf{P} = \mathbf{P_0}$. We can also introduce the extended $(K+1)$-dimensional vector $\mathbf{p}_+ = \{s, \mathbf{p}\}$ and Jacobian $\hat{\mathbf{J}}_+ = \{\mathbf{J}_0 \ \hat{\mathbf{J}}\}$ of dimension $N \times (K+1)$. Then Eq. (3.20) can be re-written as

$$\mathbf{r} = \hat{\mathbf{J}}_+ \cdot \mathbf{p}_+ + \boldsymbol{\xi} \tag{3.21}$$

Assuming that signal s, parameters p_i and noise ξ_i are drawn from independent Gaussian distributions with zero mean and variances $\sigma_s^2, \sigma_{p_i}^2, \sigma_\xi^2$, we can compute the mutual information between the signal s and the response $\mathbf{r}$ using general formula

$$I(R_1, ..., R_K; S) = \frac{1}{2} \log_2 \left(\frac{|\boldsymbol{\Sigma}_R|}{|\boldsymbol{\Sigma}_{R|S}|} \right) \tag{3.22}$$

The components of the covariance matrices $\boldsymbol{\Sigma}_R$ and $\boldsymbol{\Sigma}_{R|S}$ read

$$(\Sigma_R)_{ij} = J_{i0} J_{j0} \sigma_s^2 + \sum_k J_{ik} J_{jk} \sigma_{p_k}^2 + \delta_{ij} \sigma_{\xi i}^2 \tag{3.23}$$

$$(\Sigma_{R|S})_{ij} = \sum_k J_{ik} J_{jk} \sigma_{p_k}^2 + \delta_{ij} \sigma_{\xi i}^2 \tag{3.24}$$

By introducing the renormalized vector $\mathbf{J}'_0$ with elements $J'_{i0} = J_{i0} \sigma_s$ and the Jacobian matrix $\hat{\mathbf{J}}'$ with elements $J'_{ik} = J_{ik} \sigma_{p_k}$, we can simplify the expressions for the covariance matrix elements to

$$\hat{\boldsymbol{\Sigma}}_R = \hat{\mathbf{J}}'_+ \cdot \hat{\mathbf{J}}'^T_+ + \hat{\mathbf{I}} \boldsymbol{\sigma}_\xi^2, \tag{3.25}$$

$$\hat{\boldsymbol{\Sigma}}_{R|S} = \hat{\mathbf{J}}' \cdot \hat{\mathbf{J}}'^T + \hat{\mathbf{I}} \boldsymbol{\sigma}_\xi^2. \tag{3.26}$$

The products of Jacobian matrices with their own transposes yield symmetric positive-semidefinite Gram matrices which we denote

$$\hat{\mathbf{G}} = \hat{\mathbf{J}}' \cdot \hat{\mathbf{J}}'^T, \quad \hat{\mathbf{G}}_+ = \hat{\mathbf{J}}'_+ \cdot \hat{\mathbf{J}}'^T_+, \tag{3.27}$$

Thus, the mutual information between the scalar signal s and the vector output $\mathbf{r}$ is given by

$$I = \frac{1}{2} \log_2 \frac{|\hat{\mathbf{G}}_+ + \hat{\mathbf{I}} \boldsymbol{\sigma}_\xi^2|}{|\hat{\mathbf{G}} + \hat{\mathbf{I}} \boldsymbol{\sigma}_\xi^2|} \tag{3.28}$$

Using the matrix determinant lemma

$$\left|\hat{\mathbf{J}}'_{+}\cdot\hat{\mathbf{J}}'^{T}_{+}+\hat{\mathbf{I}}\boldsymbol{\sigma}^2_{\xi}\right|\equiv\left|\mathbf{J}'_0\mathbf{J}'^{T}_0+\hat{\mathbf{J}}'\cdot\hat{\mathbf{J}}'^{T}+\hat{\mathbf{I}}\boldsymbol{\sigma}^2_{\xi}\right|=\left|\hat{\mathbf{J}}'\cdot\hat{\mathbf{J}}'^{T}+\hat{\mathbf{I}}\boldsymbol{\sigma}^2_{\xi}\right|\left(1+\mathbf{J}'^{T}_0(\hat{\mathbf{J}}\cdot\hat{\mathbf{J}}'^{T}+\hat{\mathbf{I}}\boldsymbol{\sigma}^2_{\xi})^{-1}\mathbf{J}'_0\right) \tag{3.29}$$

we obtain

$$I=\frac{1}{2}\log_2\left(1+\mathbf{J}'^{T}_0(\hat{\mathbf{G}}+\hat{\mathbf{I}}\boldsymbol{\sigma}^2_{\xi})^{-1}\mathbf{J}'_0\right) \tag{3.30}$$

Returning to the original notation, we can re-write it as

$$I=\frac{1}{2}\log_2\left(1+\sigma^2_s\mathbf{J}^T_0(\hat{\mathbf{J}}\hat{\mathbf{C}}\hat{\mathbf{J}}^T+\hat{\mathbf{I}}\boldsymbol{\sigma}^2_{\xi})^{-1}\mathbf{J}_0\right) \tag{3.31}$$

where $\hat{\mathbf{C}}$ is the diagonal $K\times K$ covariance matrix of parameters with diagonal elements $\sigma^2_{p_k}$.

Let us first consider $\boldsymbol{\sigma}^2_{\xi}=0$ (no intrinsic noise, only parameter variability). If the number of measurements is less than or equal to the number of independent parameters, $N\leq K$, and the measurements are linearly independent, the Gramian $|\hat{\mathbf{G}}|$ is non-zero, the inverse $\hat{\mathbf{G}}^{-1}$ exists, and the mutual information is finite. We can estimate it by performing eigendecomposition $\hat{\mathbf{G}}=\hat{\mathbf{Q}}\hat{\mathbf{\Lambda}}\hat{\mathbf{Q}}^T$, where $\hat{\mathbf{Q}}$ is the matrix composed of orthonormal eigenvectors $\mathbf{q}_i$ and $\mathbf{\Lambda}$ is the diagonal matrix of corresponding eigenvalues of the Gram matrix $\hat{\mathbf{G}}$. Then

$$\hat{\mathbf{G}}^{-1}=\hat{\mathbf{Q}}\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{Q}}^T \tag{3.32}$$

Substituting this expression in (3.30), we obtain

$$I=\frac{1}{2}\log_2\left(1+\sigma^2_s\sum_{i=1}^{N}\frac{d^2_i}{\lambda_i}\right) \tag{3.33}$$

where $d_i=\mathbf{J}_0\cdot\mathbf{q}_i$ are projections of vector $\mathbf{J}_0$ on the orthonormal eigenvectors of the Gram matrix $\hat{\mathbf{G}}$. Since Gram matrix is positive-semidefinite, its eigenvalues are non-negative, and so the mutual information is positive, as it should.

If the number of independent measurements is greater than the number of independent parameters, $N>K$, the Gramian $|\hat{\mathbf{G}}|=0$ (the rank R_G of the Gramian is at most K, which means that at least $N-K$ eigenvalues are zero. If additionally the corresponding projections d_i are non-zero, the mutual information becomes infinite. What is the meaning of non-orthogonality of $\mathbf{J}_0$ and $\mathbf{q}_i$, one of the the null-vectors of $\hat{\mathbf{G}}$? It is easy to see that if $\mathbf{J}_0\cdot\mathbf{q}_i=0$, then $\mathbf{q_i}$ is also the nullvector of $\hat{\mathbf{G}}_+$, i.e. the extended Gramian is also degenerate, so the measurements are linearly dependent. And the converse statement is also true, namely, if $|\hat{\mathbf{G}}|=0$ but the measurements are linearly independent, then the Gramian $\hat{\mathbf{G}}_+\neq 0$ which is only possible if $\mathbf{J}_0\cdot\mathbf{q}_i\neq 0$.

Equivalently, the condition for infinite mutual information can be formulated via the ranks $R_G, R_{G'}$ of the Gram matrices $\hat{\mathbf{G}}$, $\hat{\mathbf{G}}'$. As the number of measurements N increases beyond 1, so do the ranks $R_G, R_{G'}$. If all measurements are linearly independent, R_G $R_{G'}$ are equal to N until $N=K$. For all $N>K$, $R_G=K$ and $R_{G'}=K+1$. For all these $N>K$, the mutual information is infinite. However, the divergence of I does not always occur at $N=K$. More generally, the divergence of the mutual information occurs

at $N = N_*$ when for the first time $R_{G'} = R_G + 1$. If some of the measurements as functions of parameters and the signal are linearly dependent, N_* can be greater than K. On the other hand, If R_G is less than K, but $R_{G'}$ is still $R_G + 1$, then the divergence occur earlier than K, at $N = R_G$. This corresponds to the case when in some measurements, only parameters act as a a single combination, while the outputs as functions of parameters and the signal are still linearly independent.

These findings are consistent with the notion that in the absence of intrinsic noise, given enough independent measurements, the signal can be deduced exactly even when parameters of the system in individual cells are unknown. However, even when the number of measurements is less than the number of unobserved parameters, the information gain can be significant (see Eq. (3.41) below and Figure S14). On the other hand, if the measurements are linearly dependent, then additional measurements will not help to eliminate parameter uncertainty.

Let us now return to the case of non-zero intrinsic noise. If we additionally assume that the intrinsic noise amplitude is the same in all measurements ($\sigma_{\xi i}^2 = \sigma_\xi^2$), the above formula (3.33) can be easily generalized

$$I = \frac{1}{2}\log_2\left(1 + \sigma_s^2 \sum_{i=1}^{N} \frac{d_i^2}{\lambda_i + \sigma_\xi^2}\right) \tag{3.34}$$

so the MI remains finite for any N. When the number of measurements is greater than the number of parameters ($N > K$), for small intrinsic noise the main contribution to the information is given by the zero-eigenvalue terms:

$$I \approx \frac{1}{2}\log_2\left(1 + \frac{\sigma_s^2}{\sigma_\xi^2} \sum_{i=1}^{N-K} d_i^2\right) \tag{3.35}$$

(where the summation is performed only over the nullspace of $\hat{\mathbf{G}}$), and the extrinsic noise is to the large degree eliminated. Since the magnitudes of projections d_i can be estimated as $\mu_i|\partial g/\partial S|$ where $\mu_i = O(1)$ are system-dependent constants , we can rewrite this formula as

$$I \approx \frac{1}{2}\log_2\left(1 + \mu^2(N-K)SNR(1 + IER^{-1})\right) \tag{3.36}$$

where $\mu = O(1)$ and σ_e^2 is the variance of the extrinsic noise, see (3.8). Comparison of this formula with Eq. (3.17) for the scalar mutual information shows two distinct sources of the information gain from multi-dimensional measurements: First, if $IER \ll 1$ (small intrinsic noise), sufficient number of measurements $N = K + 1$ leads to complete elimination of extrinsic noise, and for large SNR the information gain is roughly $-0.5\log_2(IER)$ bits. Second, additional measurements beyond $N = K + 1$ suppress the remaining intrinsic noise and add approximately $0.5\log_2(N - K)$ bits to the mutual information. Figure 3A shows the two contributions to the information gain.

To illustrate these theoretical results, we computed average mutual information for a random selection of linear signaling systems with 3 parameters that produce up to 9 outputs. These outputs are different linear combinations of the input signal s, the three parameters p_{1-3}, and the intrinsic noise ξ that is different and independent in each measurement. We

assume that the extended Jacobian for each of the linear systems is a random 5×3 matrix with real entries taken from a uniform distribution between -0.5 and 0.5.

We also assume that the signal and all parameters are taken from Gaussian distributions with the same variance $\sigma_s^2 = \sigma_{p_i}^2 = 1$, and the intrinsic noise at each output has identical variance σ_ξ^2. For a typical matrix $\hat{\mathbf{J}}_+$ the rank of $\hat{\mathbf{G}}$ for $N \geq 3$ is 3, and the rank of $\hat{\mathbf{G}}_+$ for $N \geq 4$ is 4.

Figure S13 shows the mutual information between input signal and the N-dimensional vector output computed using Eq. (3.28) as a function of the output dimension N and averaged over 1000 realizations of $\hat{\mathbf{J}}_+$.

We also plot the average mutual information for different ratios of intrinsic-to-extrinsic noise (IER) which is determined by the level of intrinsic noise σ_ξ^2. When IER is large, the dynamical nature of the underlying system is unimportant, and the dynamical vector measurements produce the information gain on par with redundant measurement case (see below Sec. 3.4.3). For small IER, however, the information gain can be much higher as soon as $N > K$. The big jump in the mutual information between $N = 3$ and 4 is due to the appearance of the large term with zero eigenvalue $\lambda_4 = 0$ but $d_4 \neq 0$ in the sum of Eq.(3.34). After the big jump, the mutual information continues to rise more slowly (logarithmically) with the dimension due to additional statistical averaging of the intrinsic noise (multiple terms with zero λ_i and $d_i \neq 0$).

The big jump occurs at $N = N_*$ when for the first time $R_{G'} = R_G + 1$. Typically, $N_* = K$. However, if in a particular realization of $\hat{\mathbf{J}}_+$ some of the measurements are linearly dependent, N_* can be greater than K. This occurs when in some measurements, parameters and signal act as a single combination. On the other hand, if R_G is less than K, but $R_{G'}$ is still $R_G + 1$, then the big jump occurs at $N_* < K$, viz. at $N = R_G$. This occurs, for instance, when in some measurements, only parameters act as a single combination, while the outputs as functions of parameters and the signal are still linearly independent.

Average mutual information for random pathways in large K, small N limit.
The eigenvalues and eigenvectors of the Gram matrix are not universal, they depend on the specific structure of the signaling pathway. However, an analytical formula for the average mutual information can be derived averaging over an ensemble of such random systems if K is much larger than N. We assume that all elements of the Jacobian matrix are i.i.d with mean 0 and variance σ_j^2. Then the Gram matrix $\hat{\mathbf{G}}$ can be approximated by a diagonal matrix $\hat{\mathbf{G}} = K\sigma_j^2\delta_{ij}, i, j = 1, ..., N$ with N equal eigenvalues $\lambda_i = K\sigma_j^2$ and their corresponding projections $d_i = 1$. According to Eq. (3.34), the average mutual information is then given by

$$I = \frac{1}{2}\log_2\left(1 + \frac{N\sigma_s^2}{K\sigma_j^2 + \sigma_\xi^2}\right) \tag{3.37}$$

Note that in this case $K\sigma_j^2$ is equal to the variance of the extrinsic variability σ_e^2, so we obtain

$$I = \frac{1}{2}\log_2\left(1 + \frac{N\sigma_s^2}{\sigma_e^2 + \sigma_\xi^2}\right) = \frac{1}{2}\log_2\left(1 + N \cdot SNR\right) \tag{3.38}$$

Average mutual information for random pathways in large K, N limit.
If both $K, N \gg 1$ and $N = O(K)$, the approximation of the Gram matrix by a diagonal F fails for the estimation of its eigenvalues. However, in this limit the spectrum of eigenvalues of the Gram matrix for a Jacobian with random i.i.d. entries can still be found analytically. Indeed, according to the Marchenko-Postur formula (*36*), the asymptotic distribution of eigenvalues for $N < K$ is given by

$$P(\lambda) = \begin{cases} \frac{1}{2\pi\sigma_j^2} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\Delta\lambda}, & \text{for } K\sigma_j^2\lambda_- < \lambda < K\sigma_j^2\lambda_+ \\ 0 \quad \text{otherwise} \end{cases} \tag{3.39}$$

where $\lambda_\pm = (1 \pm \sqrt{\Delta})^2$ and $\Delta = N/K$. For $N > K$, a singular term $(1 - K/N)\delta(\lambda)$ should be added to expression (3.39) that accounts for the family of zero eigenvalues of $\hat{\mathbf{G}}$. It is also known that eigenvectors of the random Gram matrix $\mathbf{q}_i$ are distributed uniformly over a unit N-sphere independently of λ. It is easy to show that in this case the mean square value of the product $\mathbf{J}_0 \cdot \mathbf{q}_i$ where $\mathbf{J}_0$ is a fixed vector with random elements having zero mean and unit variance, is one.

In the limit $N \to \infty$, the sum over the eigenvalues in (3.34) can be replaced by an integral:

$$I = \frac{1}{2} \log_2 \left(1 + N\sigma_s^2 \int P(\lambda) \frac{d^2}{\lambda + \sigma_\xi^2} d\lambda \right) \tag{3.40}$$

where $P(\lambda)$ is the distribution of eigenvalues of the Gram matrix $\mathbf{G}$. The integral in (3.40) with $P(\lambda)$ given by (3.39) can be calculated analytically, and the resultant expression for the mutual information reads

$$I = \frac{1}{2} \log_2 \left(1 + K\ SNR(1 + 1/IER)W) \right) \tag{3.41}$$

where

$$W = \frac{1}{2} \left(\sqrt{(\lambda_- + IER)(\lambda_+ + IER)} - \sqrt{\lambda_-\lambda_+} - IER \right) + (1 - K/N)H(N/K - 1)$$

($H(x)$ is the Heaviside function). This expression becomes especially simple in the case $IER = 0$:

$$I = \frac{1}{2} \log_2 \left(1 + \frac{N}{1 - N/K} SNR \right) \tag{3.42}$$

Note that for $N \ll K$ this formula agrees with (3.38), as expected. It is evident from this formula that the mutual information diverges as $N \to K$. For non-zero IER, the divergence of mutual information at $N = K$ is lifted, and the mutual information remains finite. Figure S14 compares the mutual information as a function of N for $K = 10$ and five different IER obtained from analytical formulas (3.41) and by direct numerical computation using Eq.(3.34). The analytical formula agrees very well with the numerics already for these not so large K and N.

3.4.3 Redundant measurements

In this section we consider the case of *redundant* measurements, i.e. when all N measurements from the same cell have the same deterministic functional response $g(S, \mathbf{P})$, and therefore the same values of gain $\partial g/\partial S$ and parameter sensitivities $\partial g/\partial P_k$. Then the covariance matrix $\mathbf{\Sigma}_R$ has the components

$$(\Sigma_R)_{ij} = \left(\frac{\partial g}{\partial S}\right)^2 \sigma_s^2 + \sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k}\right)^2 \sigma_{p_k}^2 + \sigma_\xi^2 \delta_{ij} \tag{3.43}$$

and

$$\left(\Sigma_{R|S}\right)_{ij} = \sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k}\right)^2 \sigma_{p_k}^2 + \sigma_\xi^2 \delta_{ij} \tag{3.44}$$

These matrices have the form of a matrix $\hat{\mathbf{X}}$ with matrix elements $X_{ij} = a + b\delta_{ij}$. Using the matrix determinant lemma it is easy to show that $|\hat{\mathbf{X}}| = b^N(1 + Na/b)$, so

$$|\mathbf{\Sigma}_R| = \sigma_\xi^{2N} \left(1 + N\sigma_\xi^{-2} \left[\left(\frac{\partial g}{\partial S}\right)^2 \sigma_s^2 + \sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k}\right)^2 \sigma_{p_k}^2\right]\right) \tag{3.45}$$

and

$$\left|\mathbf{\Sigma}_{R|S}\right| = \sigma_\xi^{2N} \left(1 + N\sigma_\xi^{-2} \sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k}\right)^2 \sigma_{p_k}^2\right) \tag{3.46}$$

and the mutual information is then

$$I(R_1, ..., R_N; S) = \frac{1}{2} \log_2 \left(1 + \frac{N\left(\frac{\partial g}{\partial S}\right)^2 \sigma_s^2}{\sigma_\xi^2 + N\sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k}\right)^2 \sigma_{p_k}^2}\right) \tag{3.47}$$

or using the notation (3.10), (3.11),

$$I(R_1, ..., R_N; S) = \frac{1}{2} \log_2 \left(1 + SNR \frac{IER + 1}{IER/N + 1}\right) = \frac{1}{2} \log_2 \left(1 + \frac{SER}{IER/N + 1}\right) \tag{3.48}$$

As seen from this expression, the amount of information about the input gained from multi-dimensional redundant measurements depends not only on the signal-to-noise ratio, but also on the intrinsic-to-extrinsic ratio. For $N = 1$ this formula reduces to (3.17), as expected, but for large N it approaches the maximum value

$$\lim_{N\to\infty} I(R_1, ..., R_N; S) = \frac{1}{2} \log_2 (1 + SER) \tag{3.49}$$

where SER is the ratio of signal to extrinsic noise,

$$SER = \frac{\left(\frac{\partial g}{\partial S}\right)^2 \sigma_s^2}{\sum_{k=1}^{K} \left(\frac{\partial g}{\partial P_k}\right)^2 \sigma_{p_k}^2}, \tag{3.50}$$

Thus a large number of measurements effectively eliminates the information loss due to the intrinsic noise ξ. On the other hand, in the absence of intrinsic noise, N drops out of eq. (3.47), so the mutual information is independent of the number of measurements. This is also expected, since in this case all measurements from the same cell are identical, and therefore redundant measurements do not yield any additional information about the input signal.

While above results are obtained for Gaussian-distributed input signal, fluctuations, and responses, we expect them to hold semi-quantitatively for other continuous non-Gaussian distributions. Of course, for discrete input distributions (as in our experiments) the mutual information is limited from above by the entropy of the input distribution.

4 Additional calculations for model to data fitting

4.1 ERK Model Simulations

For numerical simulations, we adopted the ODE model of ERK signaling network from Sturm *et al* (*23*). The model incorporates dynamics from RasGTP through Raf and MEK down to ERK phosphorylation. Following (*23*), we used the input concentration of RasGTP as a proxy for extracellular EGF, varying its value over several orders of magnitude. The output was defined as double phosphorylated ERK (ERKpp), which serves as a proxy for ERK activity that is a monotonously increasing function with respect to ERKpp. We found that ERKpp dynamics generated by the model closely matched the dynamics of FRET signal recorded in experiments (Fig. S15). Given these realistic trajectories, we used the model to test some of the results predicted by the earlier discussed theory.

4.1.1 Simulations with extrinsic noise

To illustrate the effect of using multi-dimensional measurements to eliminate the contribution of extrinsic noise to the information transfer of the system, we calculated MI using model simulation trajectories of ERKpp as the response and the input RasGTP. The range of input RasGTP (2500 to 22500) was chosen to minimize saturation at both ends of response. Except for ERK and MEK, model parameter values were kept consistent with Sturm *et al* for all simulations. To generate extrinsic noise we randomly sampled initial concentrations (i.c.) of ERK and MEK values from uniform distributions ($\pm 20\%$) centered around the nominal values presented in Sturm *et al* . The model was allowed to reach a steady state with the chosen i.c. of ERK and MEK parameters at the lowest value of RasGTP (2500), before applying inducing amount of RasGTP. The model was simulated for 30min. As the number of input levels of RasGTP within the input range increased, we found that the multi-dimensional measurement of ERKpp ($0 : 3 : 30min$) resulted in MI equivalent to the number of input levels (orange curve in Fig. 3B). In stark contrast, the scalar measurement MI shows a saturation around 2 bits (purple curve in Fig. 3B), which is consistent with theoretical prediction that at least 3 measurements are required to completely remove extrinsic noise. We were also able to carry this analysis out with intrinsic noise, which was added in the form of Gaussian noise that scales with the response and additive Gaussian noise that is response independent (Fig. S16). The limit in number of trajectories per signal level makes estimation of MI at higher input levels more difficult with addition of intrinsic noise.

To further demonstrate the underlying principles that allow for elimination of extrinsic noise, we plot 50 ERKpp trajectories generated from two nearby input values of RasGTP (Fig. 3C). The overlap between these trajectories might suggest that it would be practically impossible to distinguish between the two input signal values. Furthermore, considering two different time points (vertical lines), this is exactly the case given the overlap between ERKpp response values at those time points (left and bottom 1-D histograms in Fig. 3D). If we consider those two points together, however, we can clearly see a separation between the two input levels as the 2-D histogram in Fig. 3D indicates.

4.1.2 Fitting vector MI vs SNR data

In Figure 4, for ERK response vector (green circles), we first considered the case that the mutual information as a function of signal-to-noise ration (SNR) can be explained by redundant measurements (i.e. no dynamics). Fitting IER in the theoretical formula for redundant measurements ($MI = 0.5log_2(1 + SNR(IER+1)/(IER/N+1))$ we find the best fit to be equal to 100 (higher bound on the fit). In actuality, the optimal IER is equal to infinity when extrinsic noise is equal to 0 (Fig. refMIvsSNR). Our IER estimates from this and one other study (SOM 4.3), lead to fits that more closely align with ERK scalar response data (green crosses) and theory (red curve).

To fit dynamic ERK response experimental inhibition data in Figure 4 (green filled circles), we applied a theoretical description of mutual information and SNR (see Section IIC: Extrinsic and intrinsic noise) to our ERK model. We calculated sensitivities of the ERK response at 10 equally spaced time points on the trajectory with respect to the signal (Ras_GTP) and 7 model protein parameters (MEK, ERK, Raf, Phase1, Phase2, Phase3, Phase4) near the middle induction level and nominal parameter values. Constructing Gram matrices with these sensitivities, we were then able to calculate MI as a function of SNR, according to Eq. 3.28 . We assumed equal coefficient of variation (CV) for all parameters. Intrinsic noise was calculated based on experimental IER ratio. Varying the number of parameters contributing to the extrinsic noise in the model, we were able to fit the mean MI vs SNR curve to the experimental data (Fig. S18). For a given number of parameters, we generated the mean MI vs SNR curve for all combinations of parameters. The best fit was obtained with two parameters accounting for the extrinsic noise in the model. This could be thought of as the number of effective system parameters contributing to extrinsic noise that our dynamic measurements can overcome. The fit is based on a linear Gaussian model where the parameters of the model are estimated with a non-linear dynamical model of ERK used in Fig. 3. It is possible that the goodness of fit of the curve to the data could further be improved with a more refined model that is not based on the linear Gaussian approximation.

4.2 Sampling dimension for vector response

To select the time points that should be part of the vector of dynamic responses for increasing dimension of the vector in Figure 4, we used a simple strategy. For a given time frame of the trajectory (60min for ERK, 15min for calcium, 5hrs for NFκB) and chosen dimension N, we selected $N+2$ equally spaced response values throughout the given time frame, removing the first and the last values. For example, for vector of dimension 1, we chose response value at the center of the given time frame, while for dimension 2, we chose values located at the one third and two thirds points of the time frame (Fig. S19).

4.3 Experimental noise analysis

4.3.1 Signal-to-Noise Ratio (SNR)

To calculate ERK signal-to-noise ratio (SNR), we defined the signal magnitude σ_r^2 as the variance of average responses over all m input levels of EGF:

$$\sigma_r^2 = \frac{1}{m}\sum_{i=1}^{m}((\frac{1}{m}\sum_{w=1}^{m}\frac{1}{n_w}\sum_{j=1}^{n_w} r_{wj}) - \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ij}))^2 \tag{4.1}$$

Noise magnitude was defined as the average of the variances of n_i responses to a single input level of EGF:

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m}(\frac{1}{n_i}\sum_{j=1}^{n_i}(\frac{1}{n_i}\sum_{w=1}^{n_i} r_{iw} - r_{ij})^2 \tag{4.2}$$

SNR is then σ_r^2/σ_n^2.

4.3.2 Autocorrelation of ERK response

We performed autocorrelation analysis on ERK trajectories to evaluate the time sampling of the data. According to the analysis, decay of autocorrelation function shows that on average self-correlation is lost after 11min (Fig. S20). This suggests that ERK measurements taken every 10min are on average independent from one another.

4.3.3 Intrinsic-to-Extrinsic Ratio (IER)

To calculate ERK intrinsic-to-extrinsic noise ratio (IER) from our data, we defined intrinsic noise as combination of stochasticity inherent to biochemical reactions in signal transduction and measurement noise, while extrinsic noise was defined by the variability in individual cell states. To estimate the upper bound on the experimental IER, we used the fact that our sampling of ERK response was faster than ERK dynamics (based on the autocorrelation of ERK response), to calculate intrinsic noise. Using second portion of ERK trajectory, where ERK levels did not change significantly between successive time point measurements, we estimated the intrinsic noise (σ_ξ^2) as the variance of the differences in ERK expression between successive time points (Fig. S21 red). To get the extrinsic noise (σ_e^2), we estimated the total noise according to equation 4.2 (Fig. S21 A cyan) for the second portion of ERK trajectory, and simply subtracted the intrinsic noise from the total noise. The mean IER (σ_ξ^2/σ_e^2) for all experimental conditions (69) was estimated to be 0.024. We also looked at increasing values of the time step for our intrinsic noise estimate finding that IER scaled with the time step, as expected. We found that even at a time step of 15 minutes the IER was only about 1 (Fig. S S21 C)

While IER estimate from our data best matches our theoretical analysis, it is limited by our assumption about ERK dynamics. For completeness, we decided to use previously obtained data (*25*) to estimated an upper bound on IER. We chose ERK data from Toettcher *et al* (*25*), who were able to measure multiple ERK responses within a single cell. To estimate IER from their data, we used a slightly different approach. Using a Hill function fit (exponent of 2 from (*25*) for each cell, we estimated intrinsic noise as the variance of the differences between experimental values and the model fit (Fig. S21 red). Similarly, to get the total noise, we fit the same function to all of the experimental points and calculated the variance of the differences between experimental values and the model fit (Fig. S21 cyan). To get extrinsic noise, we simply subtracted the intrinsic noise from the total noise. The

mean ratio of intrinsic to extrinsic noise was estimated to be 1.14. The higher value for IER in this estimate could in part come from the noise in membrane SOScat measurement as well.

To further show the benefit of dynamics on information transfer, we calculated channel capacity for 3 separate regions of ERK response trajectories (0-20min, 20-40min, and 40-60min). Comparing to the full trajectory maximal information transfer, we find that the first portion of the trajectory, which contains majority of the dynamic information contributes the most to the total information transfer (Fig.S22)

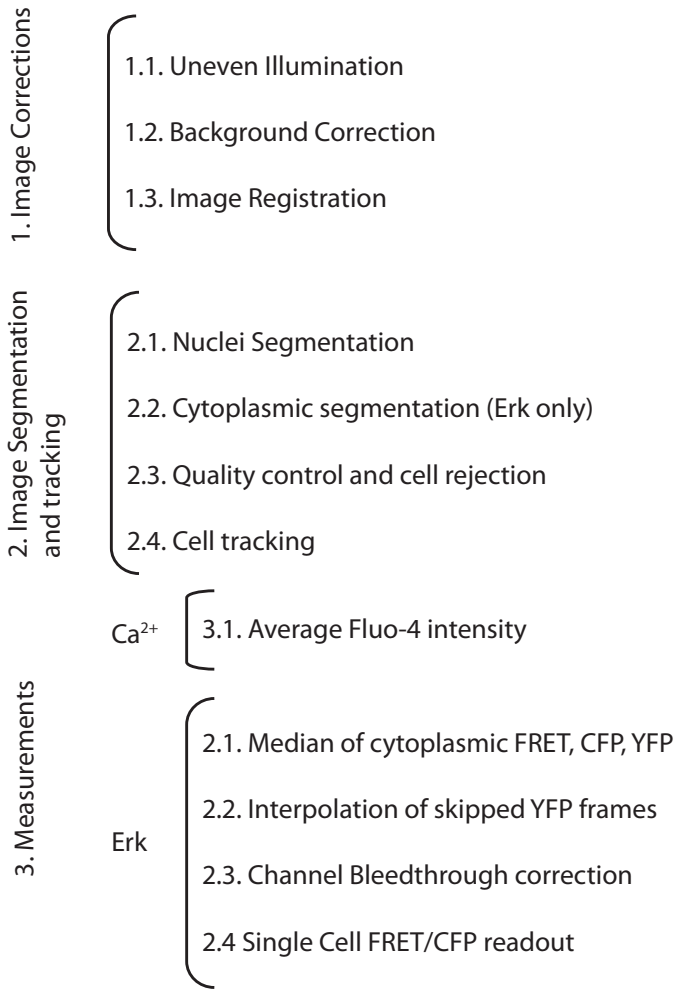

Fig. S1
Overview of the image analysis of MCF10A cells.

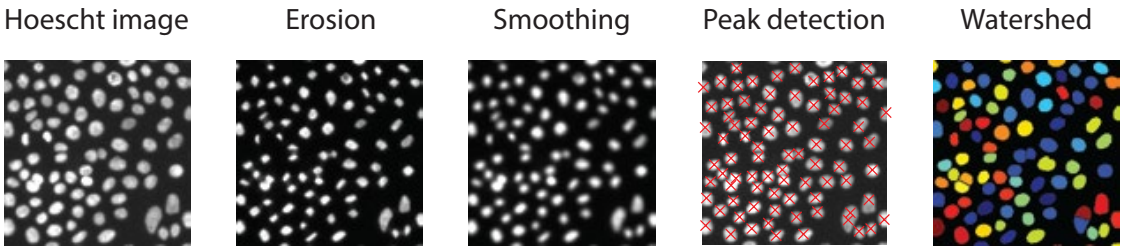

Fig. S2
Image analysis steps done to segment nuclei using Hoescht staining. Only a small part of an image is shown for brevity.

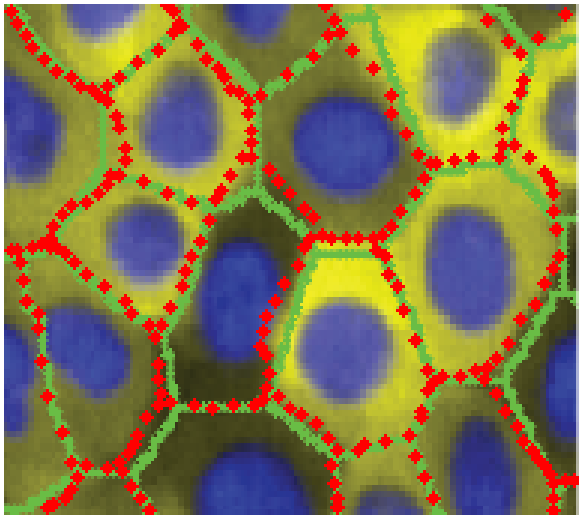

Fig. S3
MCF10a cells stained with Hoescht (blue) and expressing EKARev (yellow) with boundaries between cells determined by automatic segmentation marked in green. Red dots indicate ”true” boundaries identified by eye. The error in cell segmentation based automatically obtained and ”true” boundaries is $9\% \pm 2\%$ (SEM).

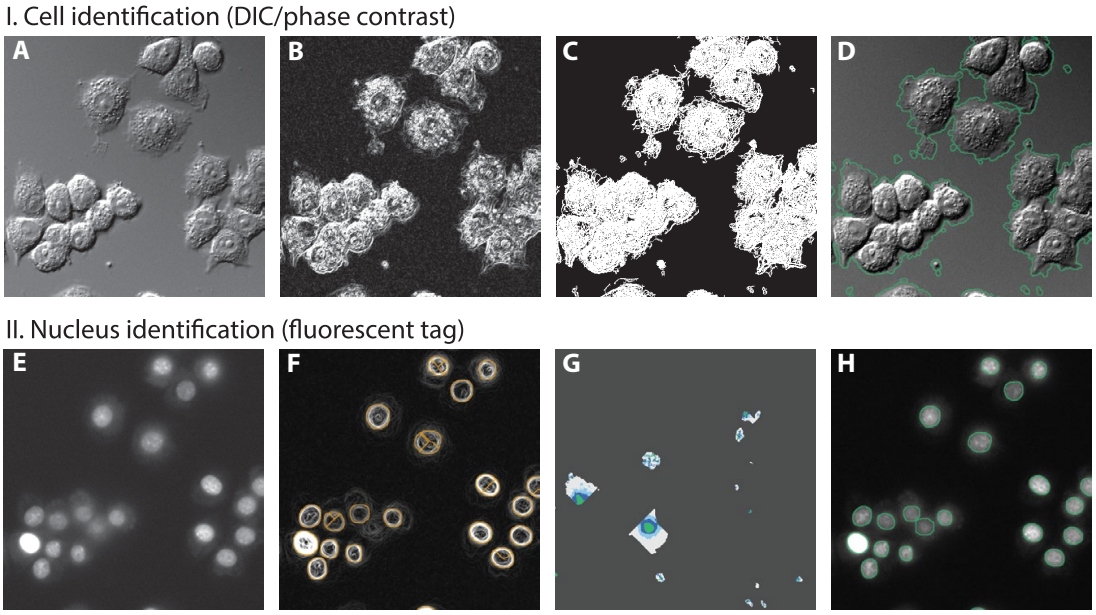

Fig. S4
Initial cell and nucleus identification. (A) Raw DIC image. (B) Sobel edge-magnitude image. (C) Thresholded edge image. (D) Final foreground/cellular boundaries. (E) Raw nuclear image (H2B-mCherry). (F) Strong objects found by scanning edge image (G) Weaker objects found in remaining candidate areas - pixels are ranked by intensity, and appropriately concentric objects are identified as nuclei. (H) Final nuclear boundaries.

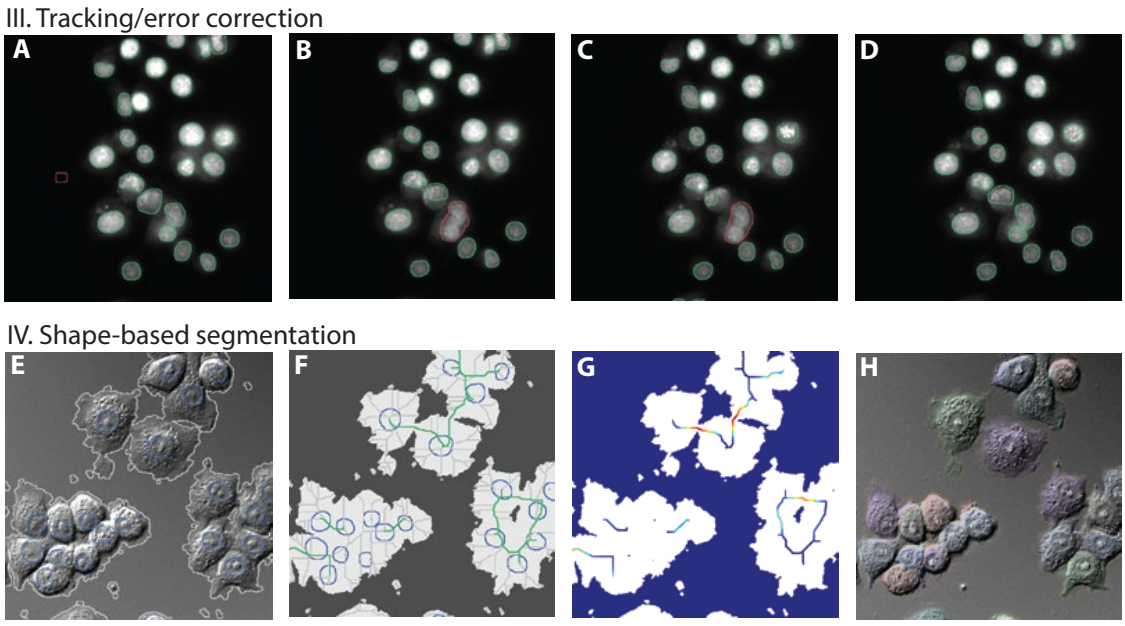

Fig. S5
Tracking and segmentation. (A-D) 4 consecutive images are individually processed, then tracked together. Voting on objects across the stack allows easy identification and correction of false positives (A, red) and false negatives (B and C, red). (E) Segmentation begins using the nuclear and cellular boundaries identified earlier. (F) The morphological skeleton is computed, then pruned to areas connecting each nuclei. (G) We identify the local maxima of the distance transformation along the pruned skeleton, as candidate splitting points. (H) Final segmented image.

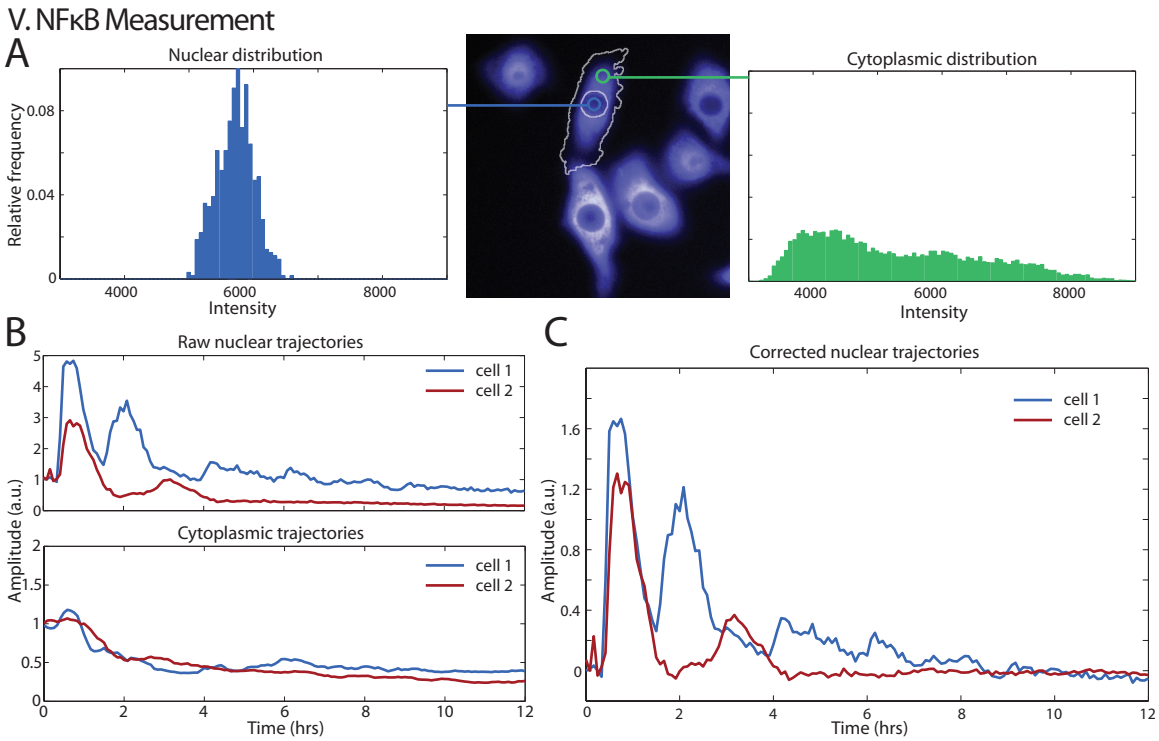

Fig. S6

NFκB measurement. (A) We calculate values that correspond to the nuclear (blue, median/mode) and cytoplasmic (green, higher mode) intensity distributions, which show identical decreasing trends over time (as a function of changing cell morphology). (B) Raw nuclear trajectories show cells that are at basal level before stimulation, and eventually stabilize after a maximum 10-14 hrs. We use this information, along with the shape computed from each cell's cytoplasmic trajectory, to calculate a true baseline for each cell. (C) Final, corrected and normalized nuclear trajectories can be directly compared.

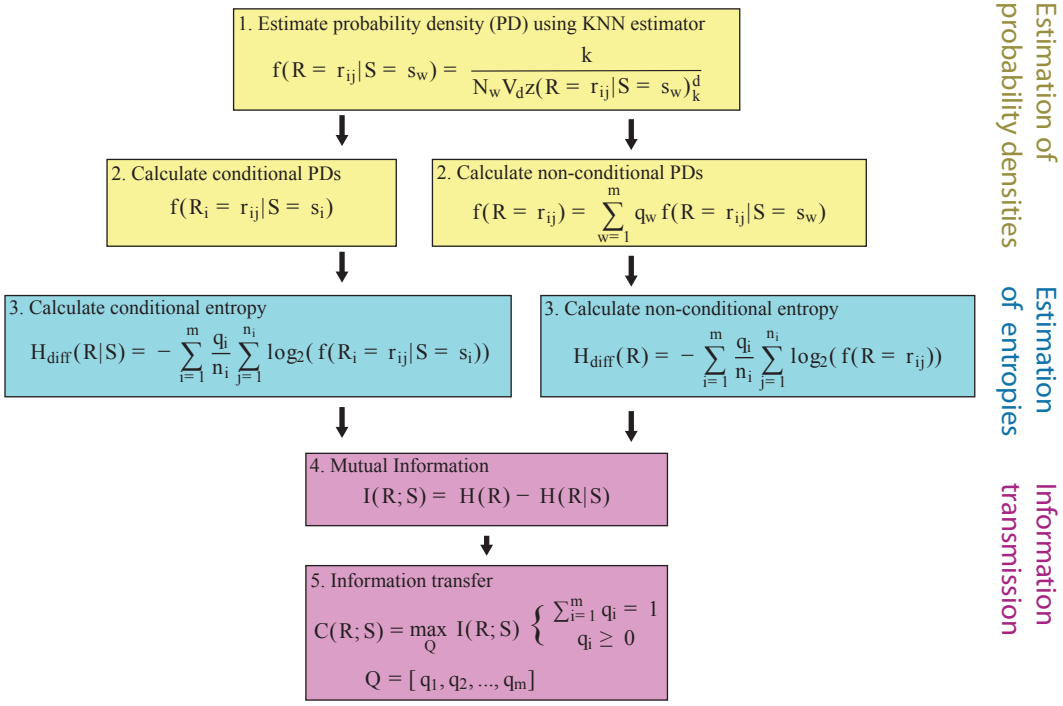

Fig. S7
General scheme for estimation of information transmission based on experimentally obtained conditional responses (R) to scalar input levels (S).

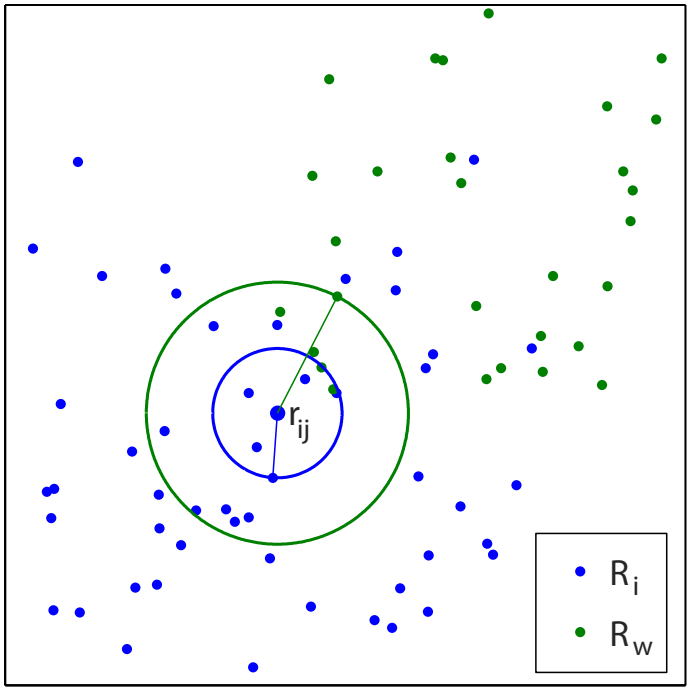

Fig. S8
Representation of k-nearest neighbor calculation for $k = 5$. The blue circle radius is the distance to the fifth closest neighbor within the same input response represented by blue points. The green circle radius is the distance to the fifth closest neighbor to a different input response (green points).

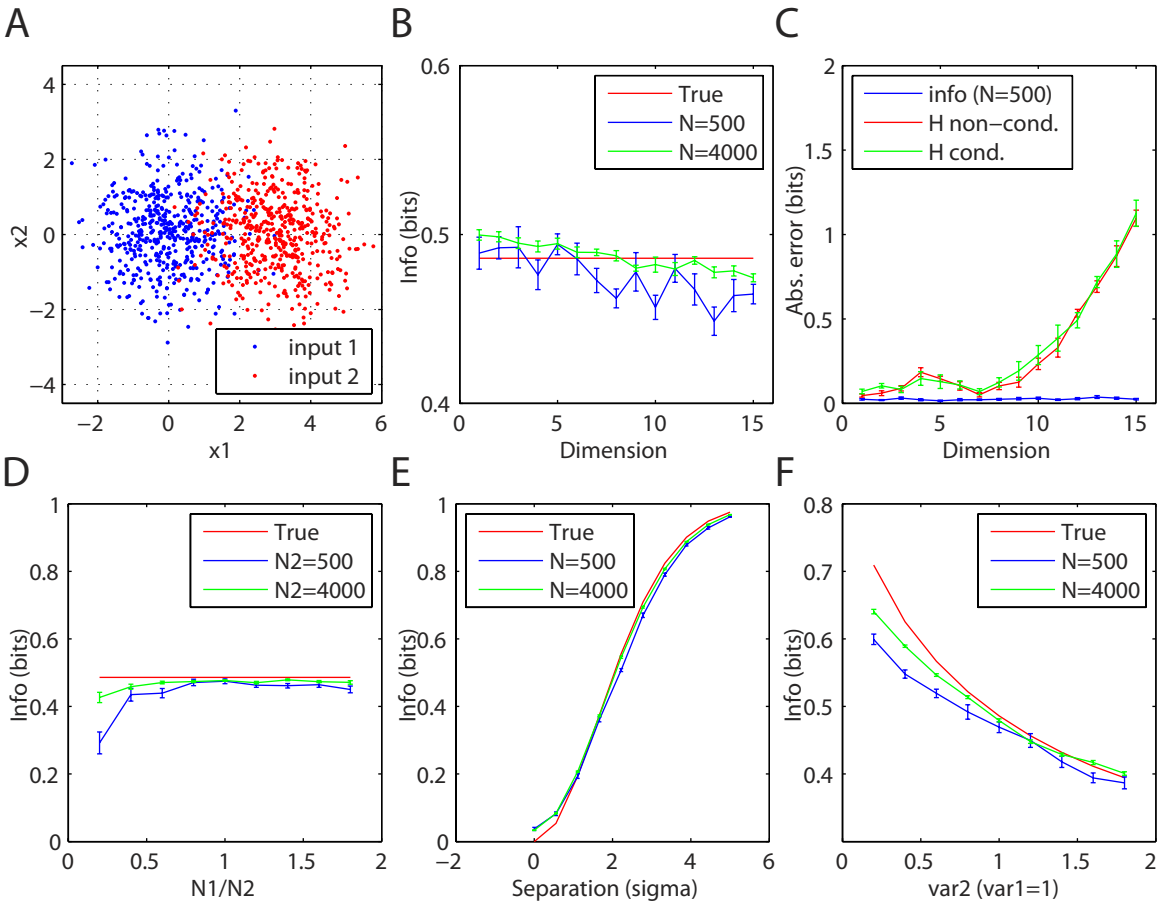

Fig. S9

(A) Sample of 2 input system with multivariate Gaussian output with separation of 2 in one of the dimensions. (B) True mutual information versus our estimate for two well sizes corresponding to number of cells measured in experiments. Our method shows good agreement up to dimension 15. (C) The absolute value of the error for conditional and non-conditional entropies due to algorithm bias increases with increasing dimension, while the error in mutual information appears to increase very slowly even in the case of 500 points per well. (D) The difference in the size of wells does not affect the accuracy of our estimate of mutual information except for when the size of those wells is too small (< 200). Dimension 10 and separation of 2 in one dimension. (E) Our algorithm estimate of MI matches true value of MI very well for varying separation of the two output multivariate Gaussians. The algorithm overestimates small separation and underestimates larges separation slightly. Dimension 10. (F) For varying variance of one of the wells, we find that our algorithm underestimates MI for low variance wells (< 0.5). Dimension 10, separation 2 in one dimension. Mean data for 10 iterations shown with error bars representing SEM.

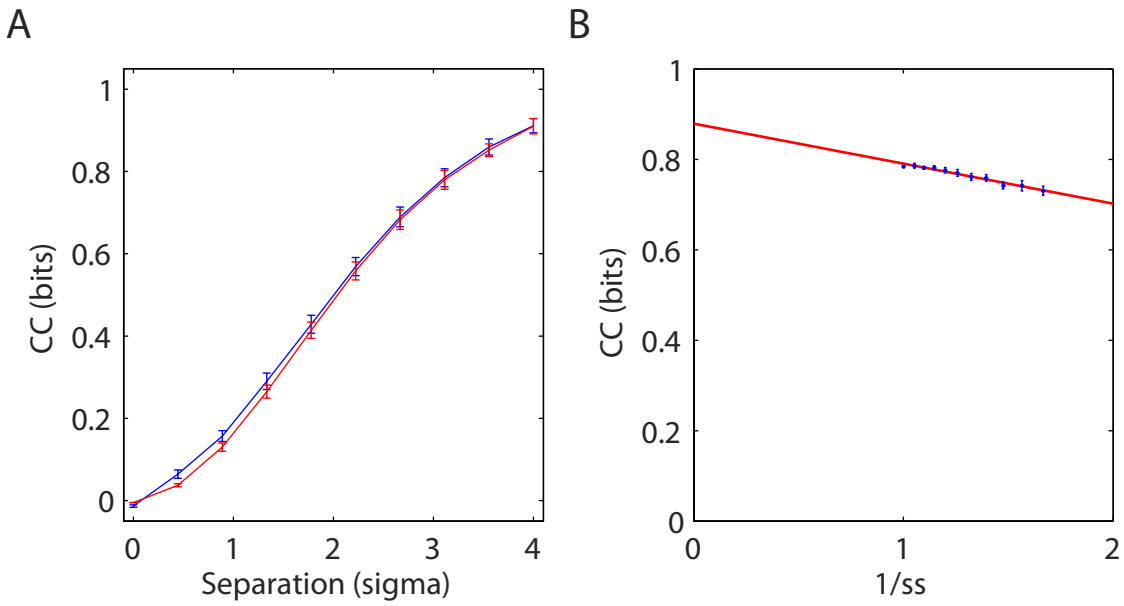

Fig. S10
(A) Comparison of our binless channel capacity calculation versus binned method (Cheong et al) for 2 Gaussians ($var = 1$) for increasing separation between their means showed similar results. Mean results shown with error bars representing SEM. (B) Applying jackknife sampling to the data, we can extrapolate the information transfer at infinite sample size.

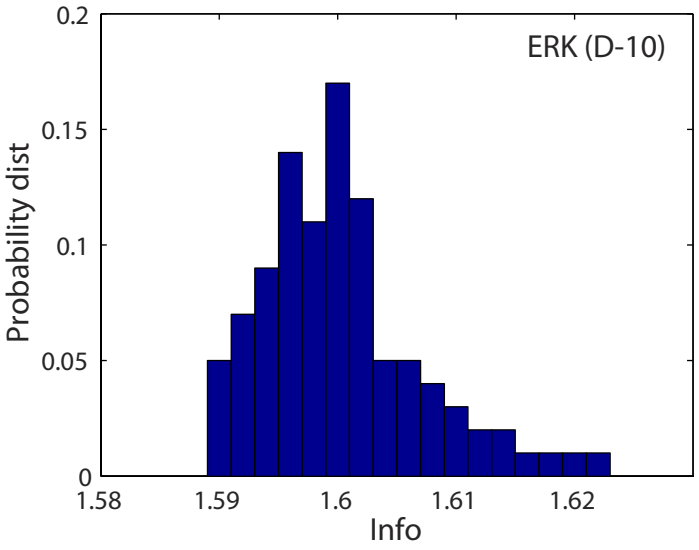

Fig. S11
Statistical analysis of our method shows that the sampling involved in jacknife estimation of channel capacity leads to very little uncertainty (STD = 0.007 bits). In this analysis we used the 10-dimensional response of ERK in Figure 2B.

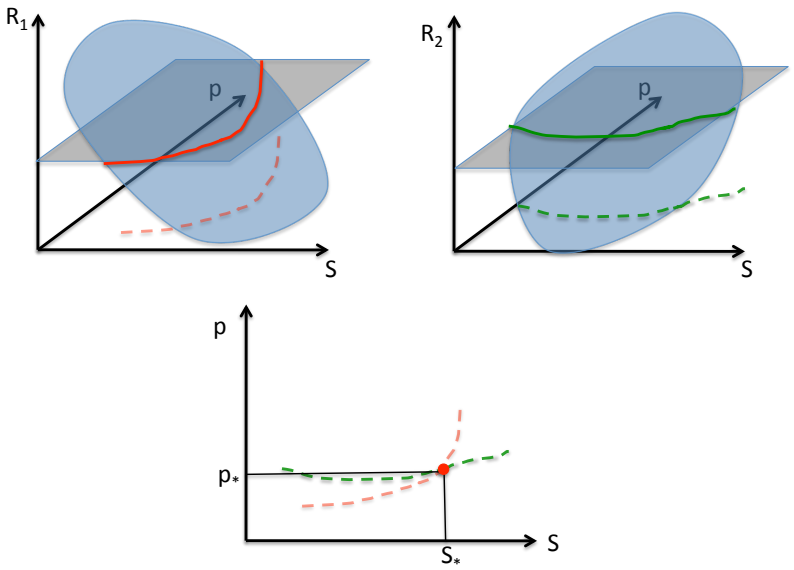

Fig. S12
Using two measurements for unique determination of the input signal for a one-parameter system.

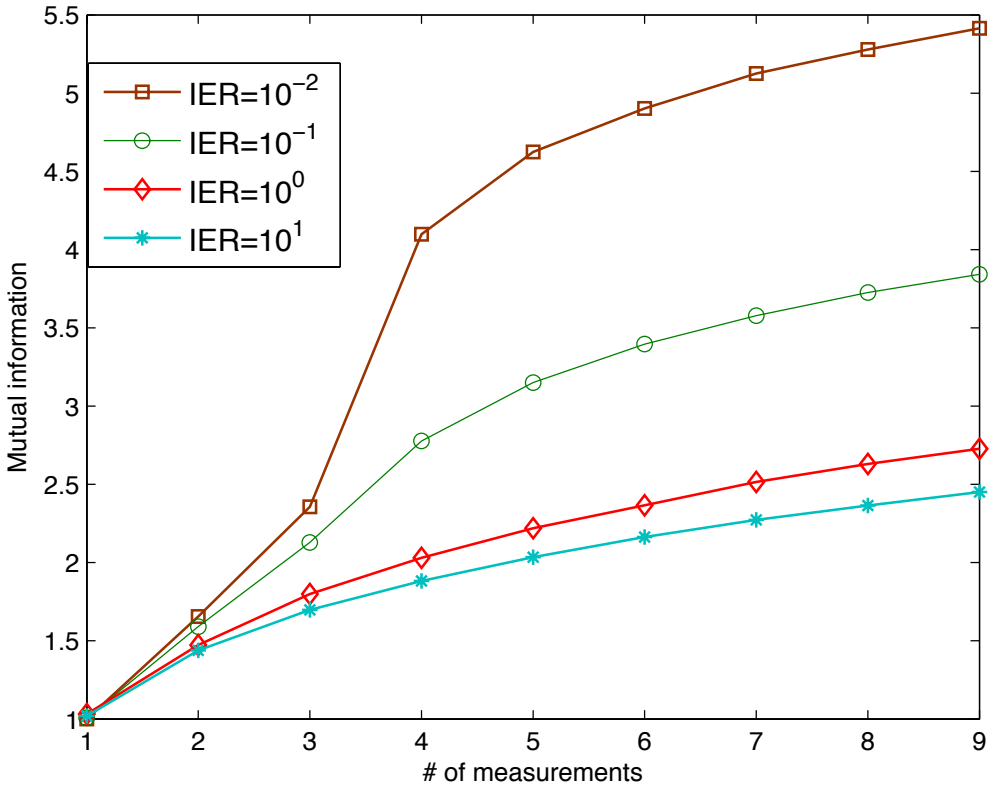

Fig. S13
Average mutual information as a function of the output dimension for different levels of intrinsic noise. The values of MI were averaged over 1000 realizations of random Jacobian matricis $\hat{\mathbf{J}}_+$ with entries uniformly distributed between -.5 and 0.5, SNR=3, and different values of IER.

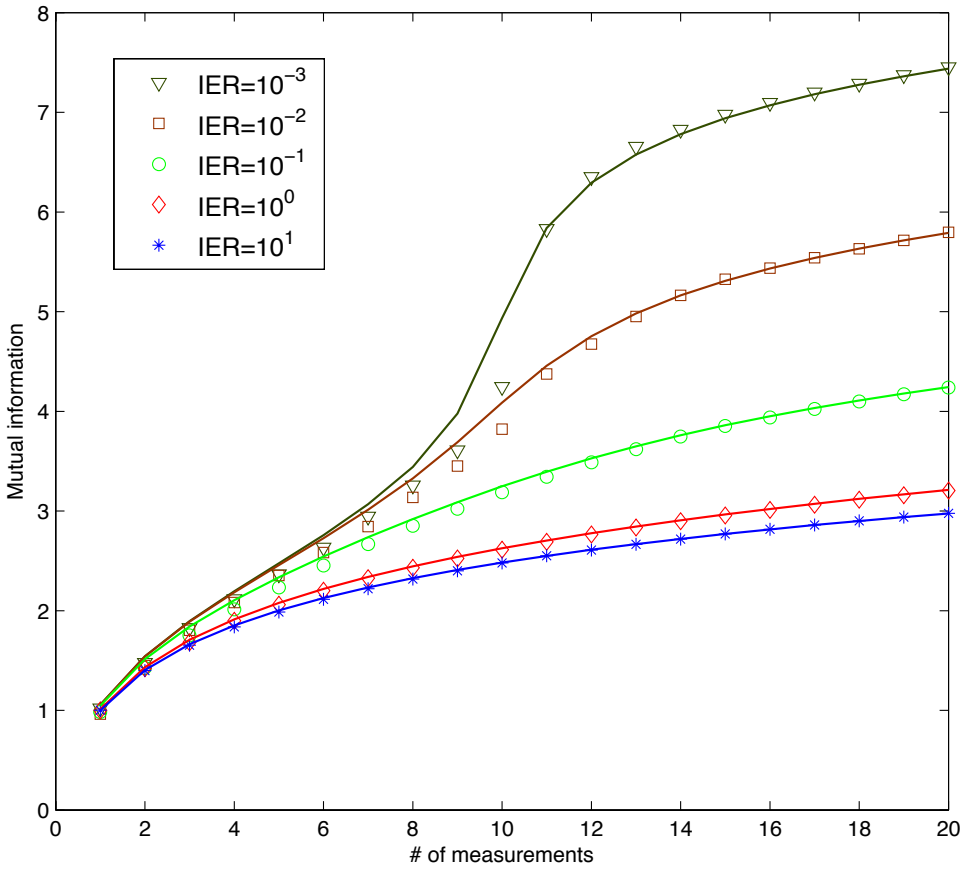

Fig. S14
Average mutual information as a function of the output dimension N for SNR$=3$, K$=10$ and five different IER obtained from formula (3.41) and by direct numerical calculations for 1000 random realizations of Jacobian matrices using Eq. (3.34).

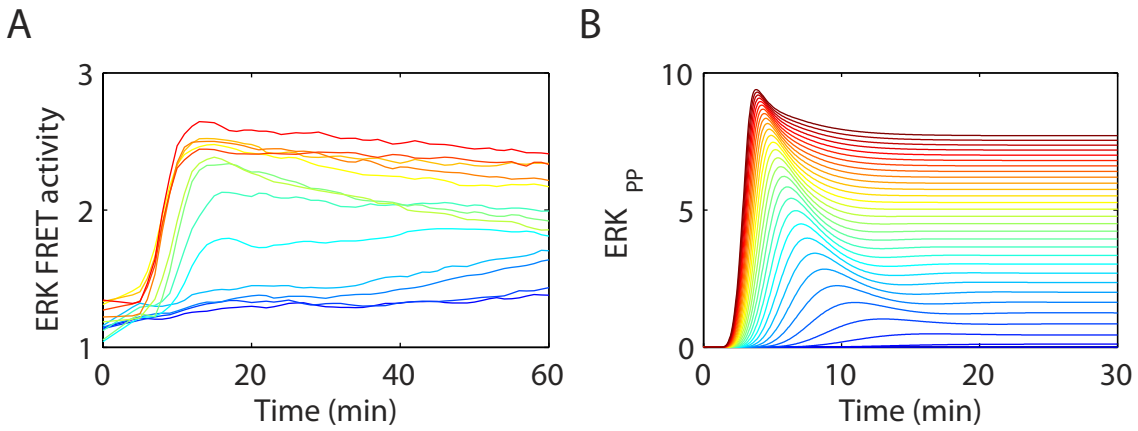

Fig. S15
Model simulation comparison to experimental ERK FRET trajectories. (A) Mean response of ERK FRET sensor to persistent EGF input. (B) ERKpp response trajectories from simulations of the ERK model (Sturm *et al*) for increasing amounts of RasGTP.

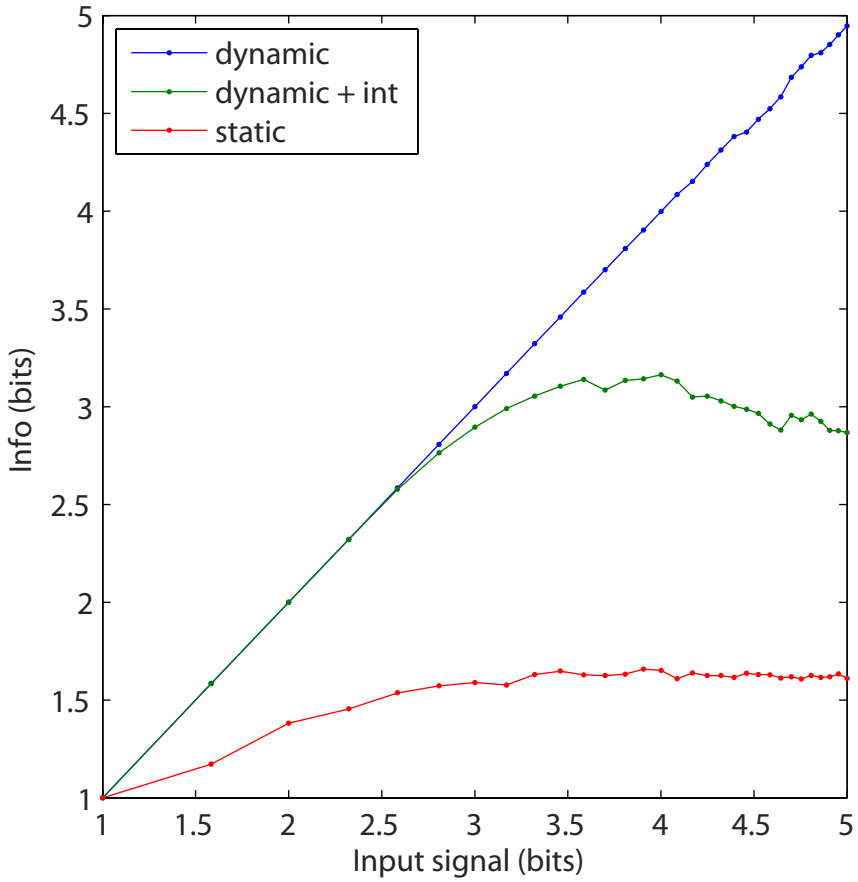

Fig. S16
Information transmission capacity of dynamic (blue, green) and static (red) calculated based on the full computational model of ERK where the extrinsic (all) noise and intrinsic (green) noise contributed to cell response variability.

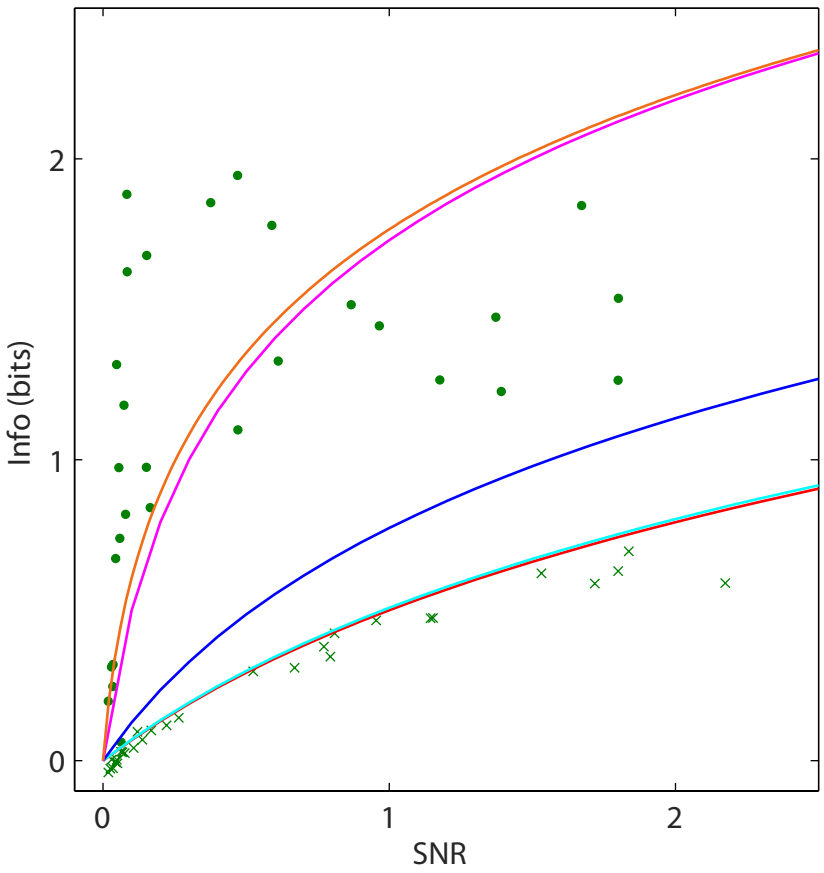

Fig. S17
Experimental measurement of the mutual information between ERK response (dimension 10) and EGF measured as a function of the response signal-to-noise ratio (SNR). Each marker represents calculations of SNR and mutual information from the dynamic (dot) and maximal scalar (cross) responses of cells from an 8-well dose-response experiment. Shown data are calculated based on 535,107 single cell responses from 29 experiments with six doses of MEK inhibitor U0126. Lines represent theoretical predictions of the mutual information as a function of SNR for three types of responses: static scalar (red line), redundant measurements where the multivariate response has no dynamics calculated based on two independent estimates of IER (0.024-light blue, 1.14-dark blue) and best fit IER=inf to the vector data (magenta), and dynamic response (orange) that can mitigate both intrinsic and extrinsic noise.

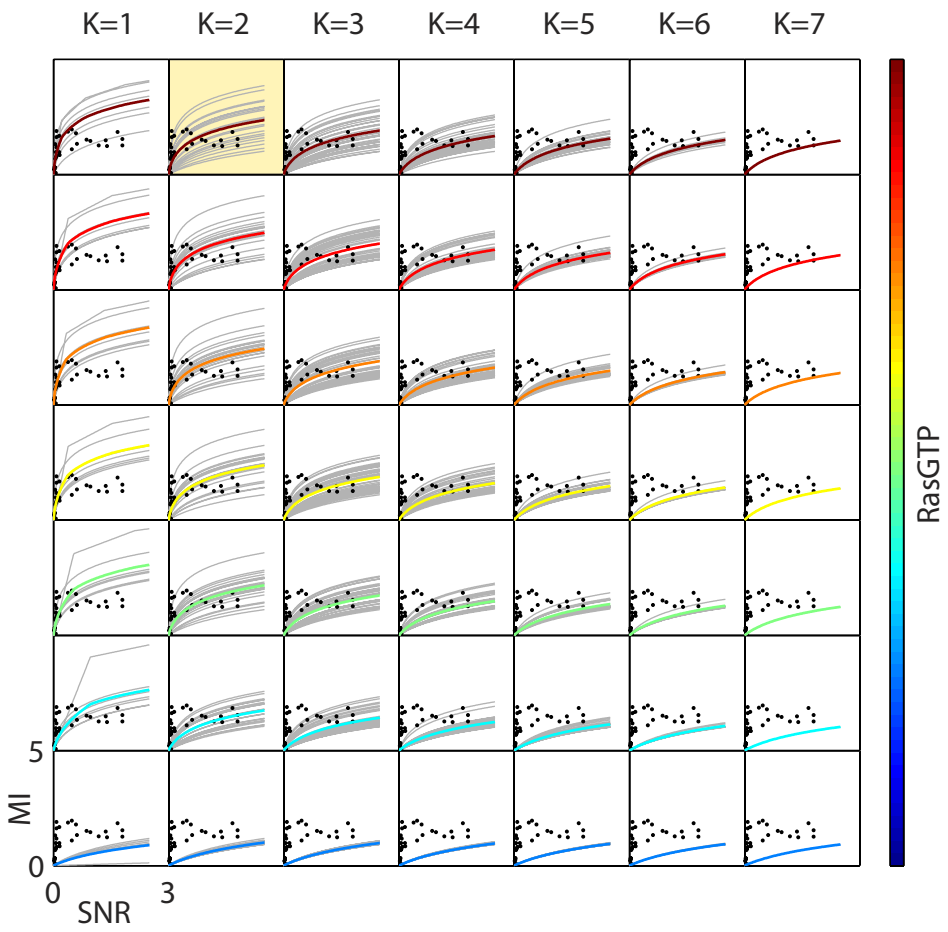

Fig. S18
Fitting ERK model to dynamic MI vs SNR data using analytic theory approach. We used K number of parameters and RasGTP input level as the two fitting parameters for the model. Grey MI vs SNR curves were obtained for are all combinations of K parameters (out of 7). Colored curves (input RasGTP level) correspond to the mean of all of those combinations of K parameters. The black dots represent experimentally obtained values for vector MI shown in Figure 4, that the model (colored curves) were fit to. Yellow highlighted box corresponds to the best fit for K and RasGTP level.

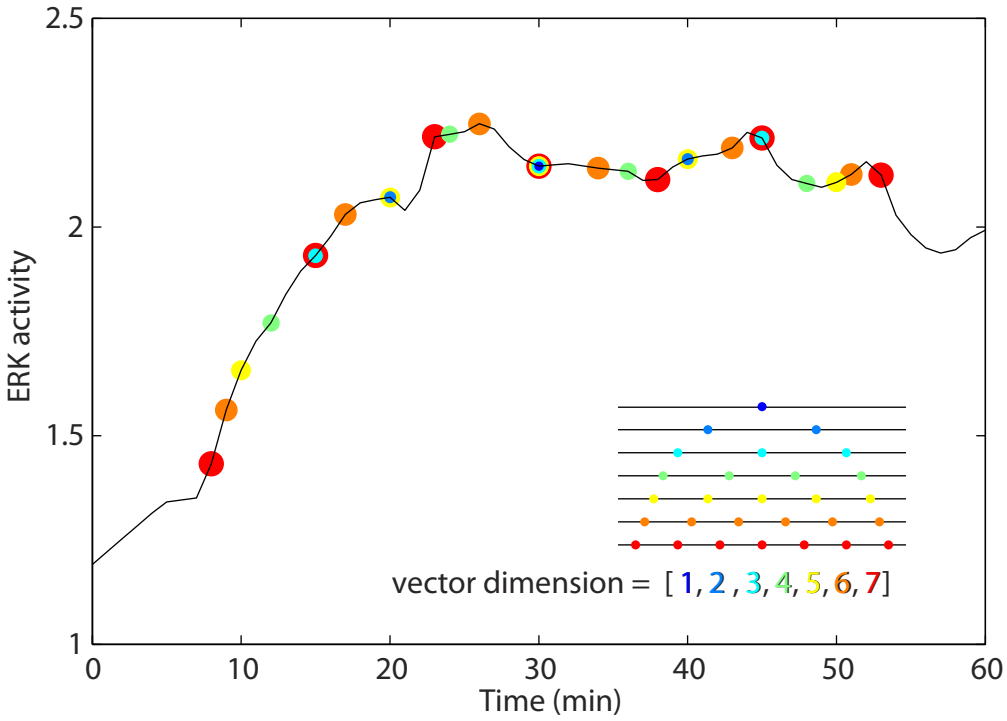

Fig. S19
Dimension sampling approach for vector dimension in Fig. 2B.

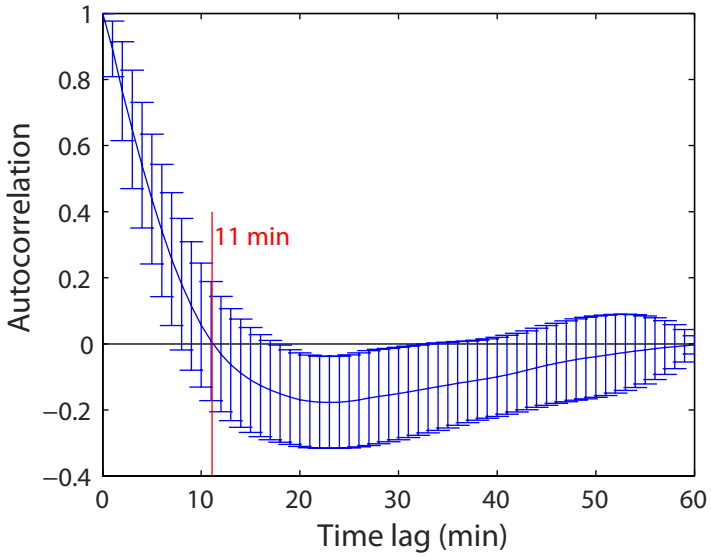

Fig. S20
Mean autocorrelation of ERK response trajectories (bars represent s.d.). Decay of autocorrelation function shows that on average self-correlation is lost after 11min.

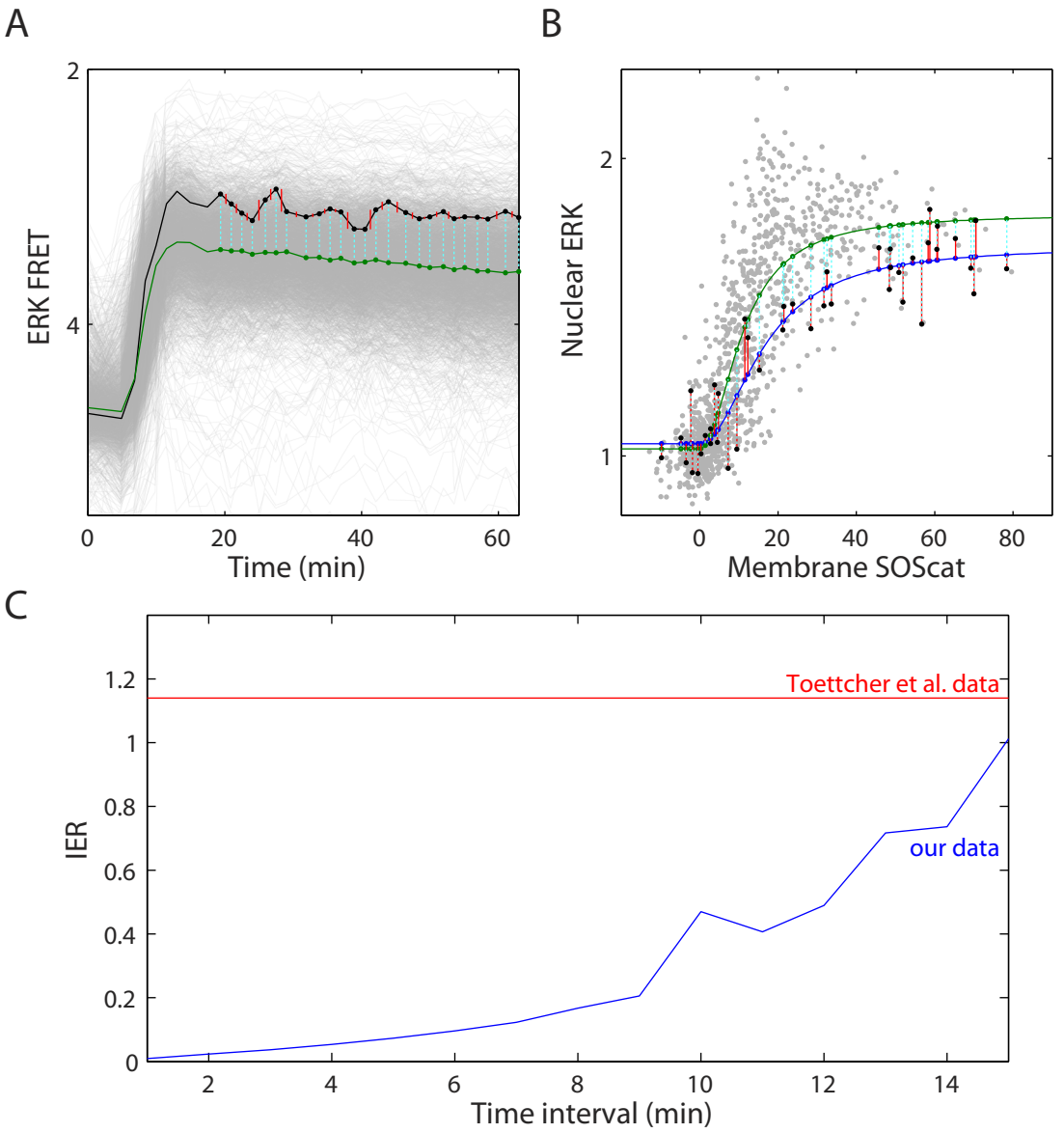

Fig. S21
Estimate of extrinsic and intrinsic noise in ERK data. (A) Using our data, intrinsic noise was estimated by the mean of the mean of squared errors between successive ERK trajectory points (red). Total noise was estimated by the mean of squared errors (cyan) between single ERK trajectory and average of all trajectories (green). Extrinsic ratio was obtained from the difference between total noise and intrinsic noise. The mean ratio of intrinsic to extrinsic noise was estimated to be 0.024. (B) Using Toettcher *et al* data, we fit a Hill function to the data and calculated the mean squared error between the fit for each cell (intrinsic noise) and between the fit for all points and each cell (total noise). The IER was estimated to be 1.14 (C) For increasing the time step in our estimate we find and increase in our estimate of IER, which never goes above Toettcher *et al* data IER estimate.

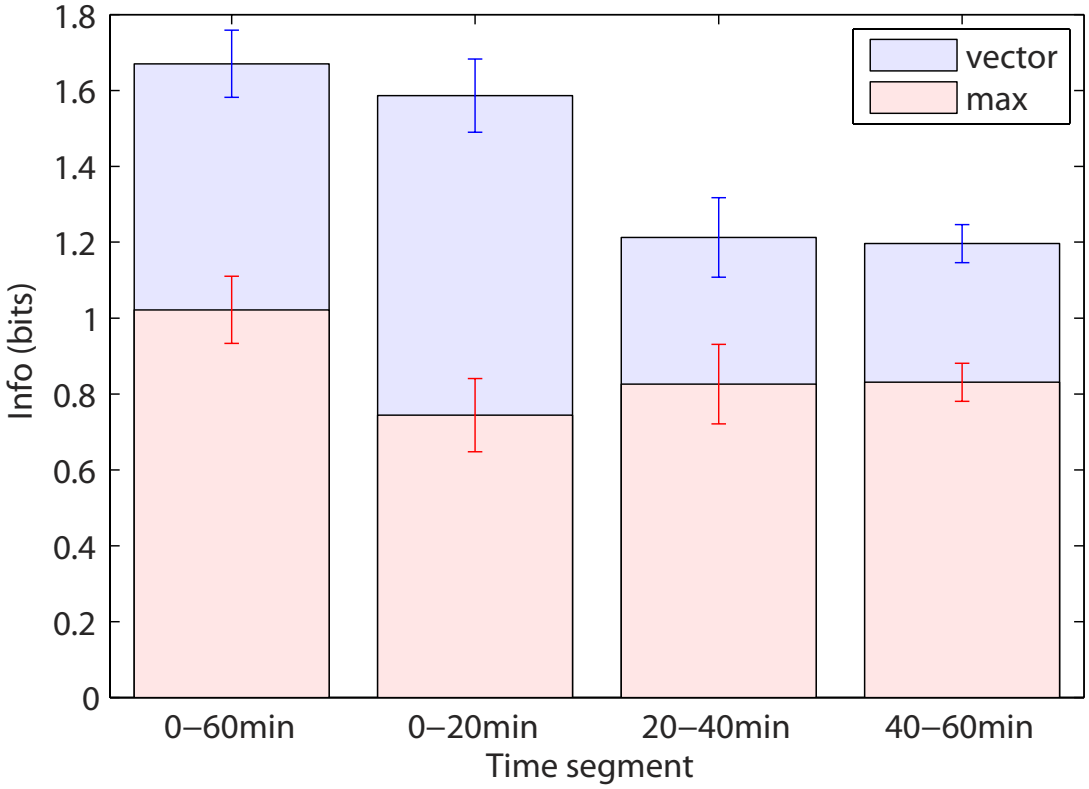

Fig. S22
Comparing the information transfer capacity of three equal and independent segments of ERK trajectory, we find that the first segment results in the highest vector information transfer. This can be explained by the fact that majority of the ERK dynamic response takes place in the first 20min, while the latter stages of the response are relatively static. Mean data shown with error bars representing SEM.

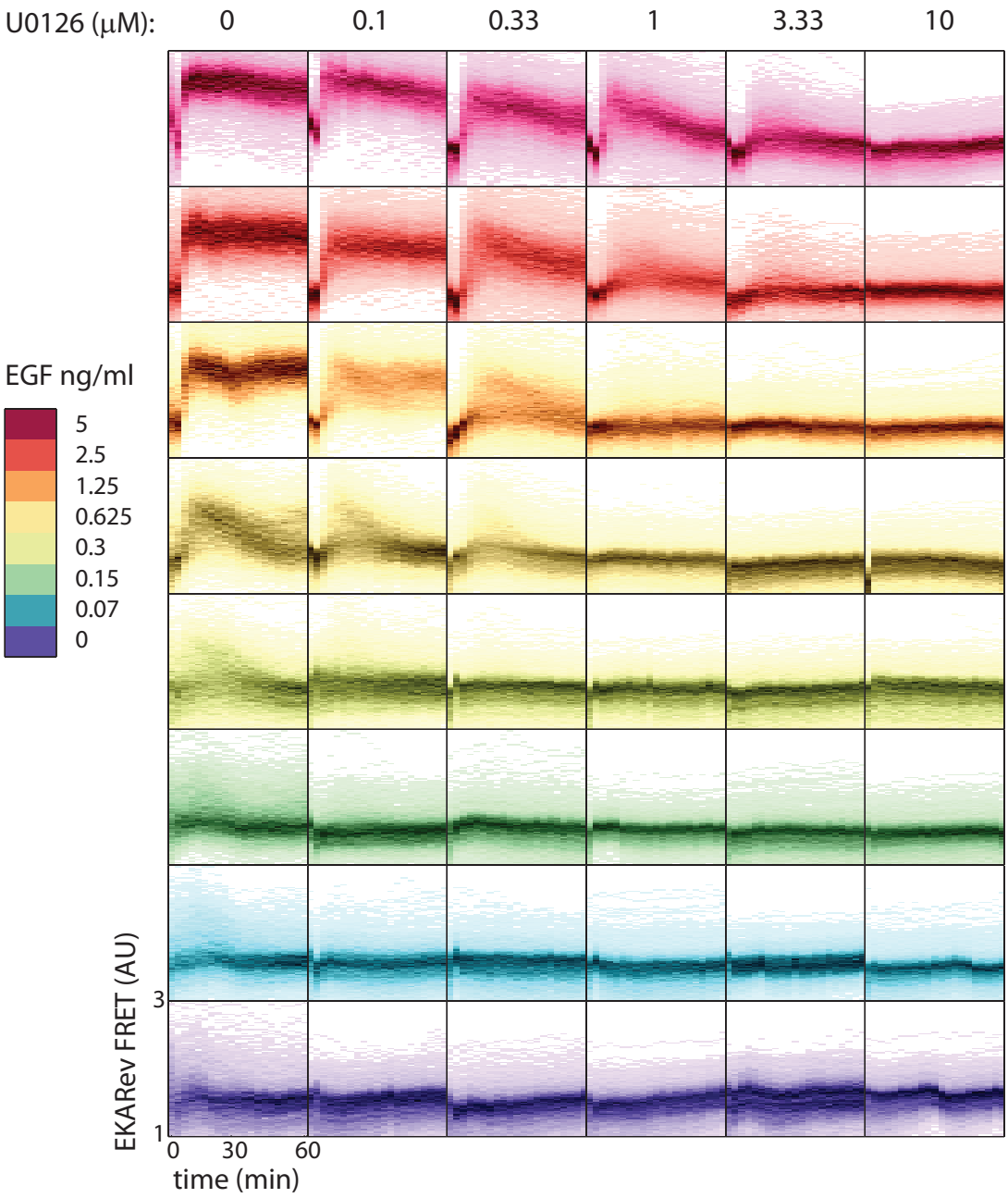

Fig. S23
Data from inhibitor experiments. The columns represent different MEK inhibitor (U0126) concentrations. The color-coded rows are different EGF induction levels. The plots show cell distribution with time where darker tint represents higher probability density. The addition of inhibitor leads to reduced ERK response due to decrease in signal propagation through the ERK pathway.

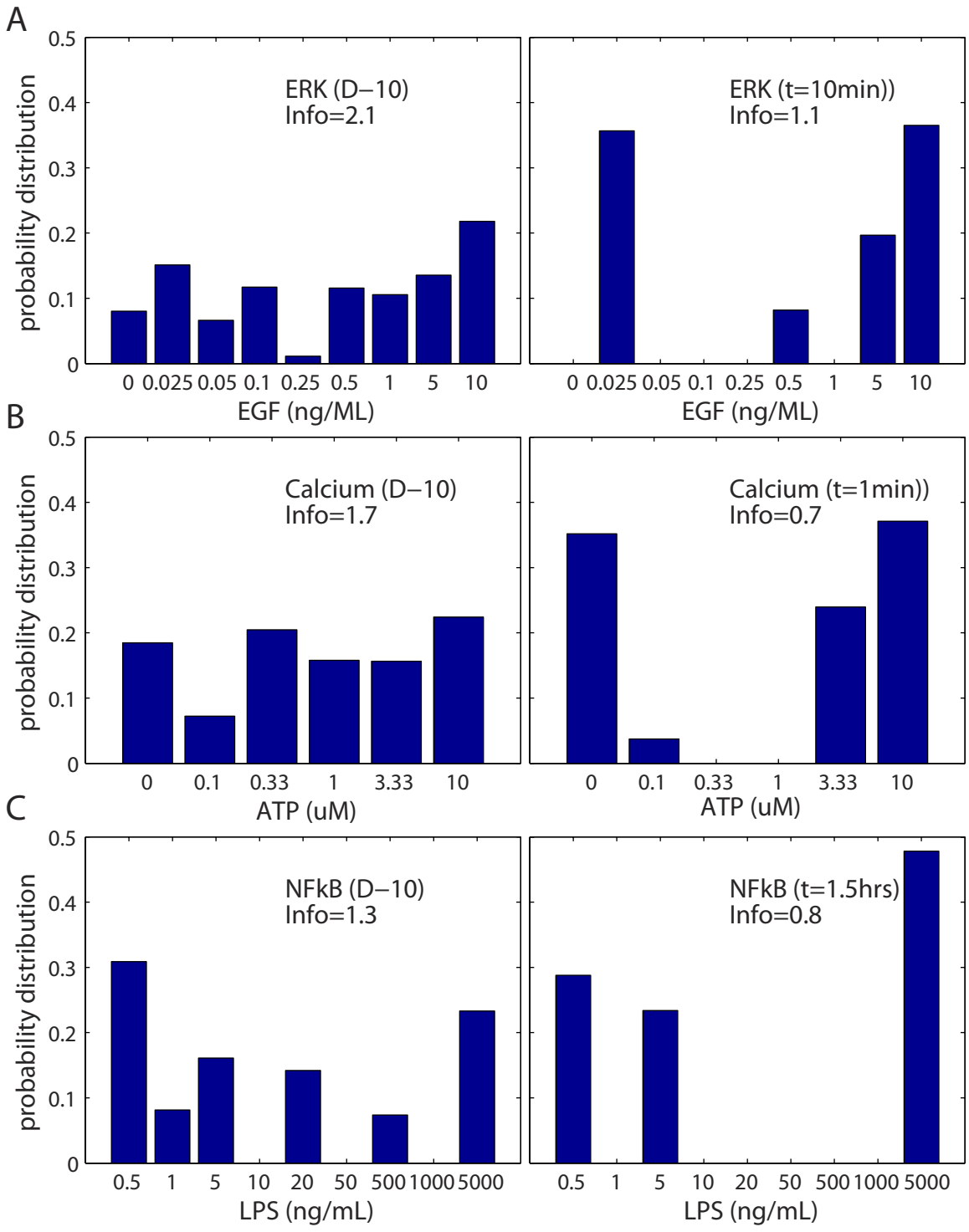

Fig. S24

Example of capacity-achieving signal distributions for ERK (A), Calcium (B) and NFκB (C) responses. The left plots show input distribution for vector calculation while the right plots show input distributions obtained for scalar response. The more uniform distribution across all input levels is indicative of higher information transfer.

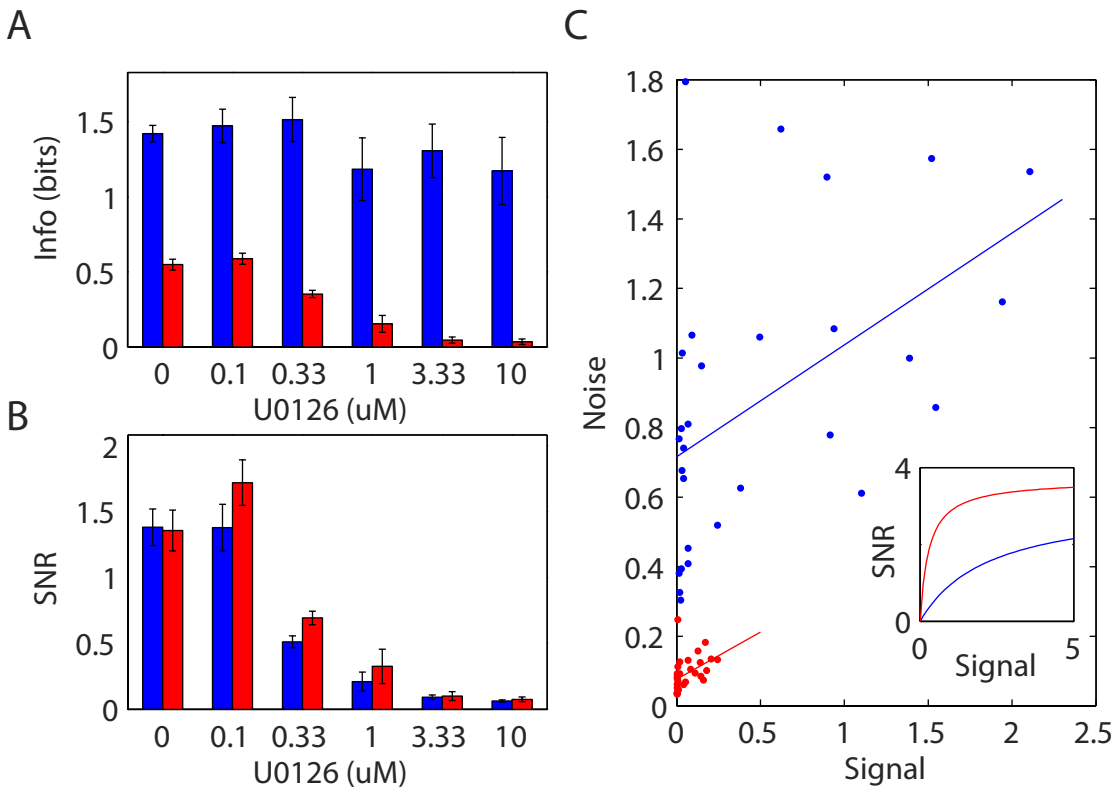

Fig. S25

Addition of increasing amounts of MEK inhibitor (U0126) does not affect information transfer calculated using vector (blue) as compared to max (red) ERK response, which shows decrease in information transfer (A). The signal-to-noise ratio (SNR) for both measurements decreases with increasing concentration of U0126 (B). Fitting a straight line to noise vs signal plot for each of the measurements (C), we find that there appears to be a limit to SNR for each of the measurements (insert). (A) and (B) show mean values with error bars representing SEM.

Table S1

Number of cell trajectories obtained in ERK experiments.

EGF												
replicate 1			replicate 2			replicate 3			replicate 4			replicate 5			replicate 6		
EGF (ng/mL)			EGF (ng/mL)			EGF (ng/mL)			EGF (ng/mL)			EGF (ng/mL)			EGF (ng/mL)		
0	4208		0	4081		0	5103		0	4871		0	2242		0	2736	
0.05	4381		0.03	3671		0.05	4937		0.05	5130		0.05	2854		0.05	2972	
0.1	4367		0.05	4580		0.1	5012		0.1	5041		0.1	2387		0.1	2999	
0.25	4457		0.1	4390		0.2	4812		0.2	4816		0.25	2510		0.25	2855	
0.5	4306		0.25	4285		0.3	5098		0.3	5100		0.5	2718		0.5	2806	
0.75	4266		0.5	3648		0.4	4956		0.4	4938		1	2490		1	2701	
1	4492		1	4209		0.5	5258		0.5	4985		5	2324		5	2708	
1.25	4499		5	3792		0.6	5263		0.7	4740		10	2165		10	2377	
1.5	4332		10	3708		0.7	5146		0.8	4717							
2	4288					0.8	5260		0.9	4710							
3	4557					0.9	5266		1	4721							
4	4332					1	5320		2	4653							
5	5030					2	5145		3	4564							
						3	4923		4	5023							
						4	4879		5	4821							
						5	4901										

Table S2

Number of cell trajectories obtained in Calcium experiments.

Calcium				
	replicate #			
ATP (uM)	1	2	3	4
0	3081	3279	3931	3139
0.1	3364	3460	3819	3710
0.33	3425	3399	3676	3299
1	3063	3419	2946	3639
3.33	3072	3366	3094	3618
10	3053	3090	2834	3790

Table S3

Number of cell trajectories obtained in NFκB experiments.

NFkB	
replicate 1	
LPS (ng/mL)	# cells
0.5	465
1	534
5	557
10	599
20	455
50	354
500	363
1000	555
5000	672

Table S4

Number of cell trajectories obtained in EGF and UO126 experiments.

EGF + U0126													
		replicate 1						replicate 2					
		U0126 (uM)						U0126 (uM)					
		0	0.1	0.33	1	3.33	10	0	0.1	0.33	1	3.33	10
ENG (mg/mL)	0	2445	2563	2912	2657	2710	2539	3094	3098	3044	3043	3161	2851
	0.07	2795	2392	2763	2478	2515	2318	3193	3243	3193	3243	3030	2843
	0.15	2653	2617	3031	2534	2685	2381	3089	3435	3106	3508	2967	2910
	0.3	2697	2687	3091	2520	2766	2483	3159	3294	3473	3547	3154	2482
	0.625	2759	3075	2661	2878	2699	2797	3028	3487	3306	3561	2858	2300
	1.25	2449	2918	2964	2686	2773	2414	3201	3325	2983	3275	2873	2472
	2.5	2492	2655	2876	2783	2464	2153	3147	3053	3034	3334	2652	2595
	5	2355	2211	2413	2447	2448	2353	3206	3188	3020	3129	2754	2806
		replicate 3						replicate 4					
		U0126 (uM)						U0126 (uM)					
		0	0.1	0.33	1	3.33	10	0	0.1	0.33	1	3.33	10
ENG (mg/mL)	0	1437	1713	1575	1653	1738	1762	2092	2451	2368	2549	2523	2533
	0.07	1854	1766	1658	1616	1708	1677	2078	2464	2904	2387	2666	2675
	0.15	1791	1788	1730	1696	1364	1345	1970	2197	2511	2601	2619	2552
	0.3	1877	1723	1829	1780	1589	1615	1896	2484	2506	2706	2415	2055
	0.625	1795	1689	1786	1705	1609	1659	1823	2349	2400	2703	2487	2396
	1.25	1890	1705	1759	1623	1716	1435	2006	2336	2583	2249	2361	2107
	2.5	1778	1692	1611	1597	1602	1643	1815	2055	2204	2475	724	2204
	5	1853	1532	1666	1725	1634	1558	1909	2090	2390	2267	2189	1926

Table S5

Number of cell trajectories obtained in ERK control experiments.

ERK control					
	replicate #				
EGF (mg/mL)	1	2	3	4	5
0	2633	1698	2958	2042	1772
0	2762	1918	3260	2080	2055
0	2397	1417	3568	2035	2062
0	2342	1669	3644	2566	1668
0	2624	1652	3225	2598	2159
0	2447	1542	3724	2273	1819

Table S6
Student's t-test p-values for Figure 3C (vector vs. scalar).

	A	**T**	**D**	**R**
ERK	1.43×10^{-5}	3.51×10^{-6}	5.57×10^{-6}	1.58×10^{-5}
Ca^{+2}	1.69×10^{-2}	2.67×10^{-2}	2.62×10^{-2}	3.32×10^{-2}
NFκB	1.99×10^{-9}	1.89×10^{-7}	5.84×10^{-10}	1.22×10^{-10}

References

28. J. G. Albeck, G. B. Mills, J. S. Brugge, Frequency-Modulated Pulses of ERK Activity Transmit Quantitative Proliferation Signals., *Mol. Cell* **49**, 24961 (2013).

29. J. Debnath, S. K. Muthuswamy, J. S. Brugge, Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures, *Methods* **30**, 256268 (2003).

30. A. Edelstein, N. Amodaj, K. Hoover, R. Vale, N. Stuurman, Computer control of microscopes using Manager., *Curr. Protoc. Mol. Biol.* **Chapter 14**, Unit14.20 (2010).

31. R. Gray, *Entropy and Information Theory* (Springer International Publishing, 2011).

32. T. R. Jones, A. Carpenter, P. Golland, Voronoi-Based Segmentation of Cells on Image Manifolds, , 535543 (2005).

33. N. Komatsu, K. Aoki, M. Yamada, H. Yukinaga, Y. Fujita, Y. Kamioka, M. Matsuda, Development of an optimized backbone of FRET biosensors for kinases and GT-Pases., *Mol. Biol. Cell* **22**, 464756 (2011).

34. A. Kraskov, H. Stgbauer, P. Grassberger, Estimating mutual information, *Phys. Rev. E* **69**, 066138 (2004).

35. D. Loftsgaarden, C. Quesenberry, A non parametric estimate of a multivariate density function, *Ann. Math. Stat.* **36**, 10491051 (1965).

36. V. Marchenko, L. Pastur, Distribution of eigenvalues for some sets of random matrices, *Mat. Sb.* **114**, 507536 (1967).

37. N. Otsu, A Threshold Selection Method from Gray-Level Histograms, *Automatica* **11**, 2327 (1979).

38. T. Schreiber, Interdisciplinary application of nonlinear time series methods, *Physics reports* **308**, 164 (1999).

39. R. Wollman, N. Stuurman, High throughput microscopy: from raw images to discoveries., *J. Cell Sci.* **120**, 371522 (2007).