Supporting Online Material for

Information Transduction Capacity of Noisy Biochemical Signaling Networks

Raymond Cheong, Alex Rhee, Chiaochun Joanne Wang, Ilya Nemenman, Andre Levchenko*

*To whom correspondence should be addressed. E-mail: alev@jhu.edu

Published 15 September 2011 on *Science* Express
DOI: 10.1126/science.1204553

This PDF file includes:

TABLE OF CONTENTS 1

1. MATERIALS AND METHODS

1.1 Cell culture

Wildtype and A20$^{-/-}$ 3T3-immortalized mouse embryonic fibroblasts (kind gift from A. Hoffmann, Univ. of California, San Diego) were maintained in low glucose Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% calf bovine serum (American Type Culture Collection) and 10 U/mL each of penicillin and streptomycin (Invitrogen). Cells were seeded at a density of approximately 150 cells/mm^2 onto 15mm diameter circular coverslips (Fisher Scientific) coated with 0.1% gelatin (Sigma) placed in a 35mm diameter dish, then serum starved in medium with reduced serum concentration (0.1%) overnight before experimentation.

1.2 Immunocytochemistry

After exposure to murine TNF (Roche) or murine PDGF-BB (Sigma) at the specified concentrations and duration, the cells were washed 3 times with ice-cold phosphate buffered saline (PBS, Invitrogen), and fixed in 4% paraformaldehyde (Electron Microscopy Sciences) for 20 minutes. The cells were then permeabilized in 0.1% triton X-100 (Sigma) for 5 minutes, and blocked in 10% goat serum (Invitrogen) for 60 minutes. Next, the cells were incubated in primary antibody solution consisting of 1:100 rabbit anti-p65 antibody (Santa Cruz), 1:100 mouse anti-phospho-ATF-2 antibody (Santa Cruz), and 2 μg/mL Hoechst-33258 (Sigma) for 60 minutes. Finally, the cells were incubated in secondary antibody solution consisting of 1:200 Alexa Fluor 488-conjugated goat anti-rabbit and 1:200 Alexa Fluor 594-conjugated goat anti-mouse antibodies (Invitrogen) for 60 minutes. All solutions were made in PBS, and cells were washed with PBS in between each step. To minimize experimentally-induced variability and to enable quantitative comparisons across conditions, all concentrations of TNF and all cell lines were assayed at the same time using common reagents. Finally, the stained coverslips were mounted on glass microscope slides and imaged on an Axiovert 200M inverted epifluorescence microscope (Zeiss) equipped with Slidebook 4.2 (Intelligent Imaging Innovations). On average, 350 cells were imaged per experimental condition.

In Fig. S4, cells were exposed to the indicated inhibitors (kind gift from J. Zhang, Johns Hopkins Univ.) beginning 1 hour before the addition of TNF. In Fig. M3, immortalized human umbilical vein endothelial cells (kind gift from the late J. Folkman, Harvard) expressing GFP (*42*), were stained with 1:100 mouse anti-GFP antibody (Roche) paired with 1:200 Alexa Fluor 594-conjugated goat anti-mouse antibodies.

1.3 NF-κB reporter gene

Wildtype 3T3 mouse embryonic fibroblasts were infected with lentiviruses containing a gene for Turbo GFP whose promoter was under the control of NF-κB (Cignal lenti NF-κB reporter, from SA Biosciences). Lentiviral transfection was performed according to the manufacturer's recommendation using a multiplicity of infection of ~200 in the presence of 1 μg/mL polybrene (Sigma), followed by selection in 6 μg/mL puromycin (Sigma). After two rounds of infection, cells were clonally seeded in a 48-well plate and tested for response to TNF. Cells that displayed high levels of GFP fluorescence were individually selected and cultured to create clonal lines of cells. GFP expression was monitored in live cells on a Zeiss Axiovert 200M microscope, or measured in cells that were fixed by exposure to 4% paraformaldehyde for 20 minutes. Reporter gene copy number was determined by quantitative polymerase chain reaction ($\Delta\Delta C_t$ method) to amplify Turbo GFP from purified genomic DNA and benchmarked against endogenous genes of known copy number (Charles River Laboratory custom service, Troy, NY).

1.4 ERK2 translocation

Nuclear translocation of ERK2 was measured using ERK2-YFP clone C7, which is a H1299 human non-small cell lung cancer cell line clone expressing YFP-tagged ERK2 and mCherry-tagged CBX5 (chromobox 5), a protein with persistent nuclear localization and unconnected to ERK2 signaling (*43*) (generous gift from Drs. C. Cohen-Saidon and U. Alon, Weizmann Institute). The cell line was maintained as described in (*23*). Prior to experimentation, the cells were seeded into a 4-well LabTek optical chamber coated with fibronectin (Sigma) and allowed to attach in serum starved conditions for 5 hours. Within the LabTek well, the cells were maintained in transparent medium consisting of a riboflavin- and phenol red-free formulation of the RPMI medium (Athena Enzyme Systems custom medium) supplemented with 10 U/mL each of penicillin and streptomycin. ERK2 and CBX5 expression was monitored in live cells on a Zeiss Axiovert 200M microscope. Measurements were made for 5 minutes to establish a baseline (zero concentration) and for 40 minutes following the addition of EGF (Peprotech) in transparent medium to the well via syringe pump. Information theoretic calculations were performed for individual cell responses at 10 minutes EGF exposure, the time at which ERK2 nuclear translocation peaked.

1.5 Image and data analysis

Image processing, data analysis, and information theoretic calculations were performed using Matlab R2006a (MathWorks). Background correction, nucleus segmentation, and quantification of nuclear concentrations of NF-κB and phospho-ATF-2 were performed as described previously (*16*). Programs are available upon request.

2. NUMERICAL COMPUTATIONS OF MUTUAL INFORMATION

2.1 Bias correction and error estimate

Mutual information between two variables can be computed from discretized data using the standard formula (*9*):

$$\begin{aligned} I(R;S) &= H(R) - H(R \mid S) \\ &= \sum_j -P(R=r_j)\log_2 P(R=r_j) \\ &\quad -\sum_i P(S=s_i)\left(\sum_j -P(R=r_j \mid S=s_i)\log_2 P(R=r_j \mid S=s_i)\right), \end{aligned} \tag{2.1.1}$$

where H is the entropy functional. The marginal distribution of the response is given by

$$P(R=r_j) = \sum_i P(S=s_i)P(R=r_j \mid S=s_i), \tag{2.1.2}$$

where the values of R (i.e., r_j) are discretized into N_R bins and the values of S (i.e. S_i) are discretized into N_S bins. In the case that the response R is, for example, a two-dimensional vector then each element of R is discretized into N_R bins for N_R^2 bins in total. The formula for mutual information, written in the form shown in Eq. 2.1.1, highlights the dependence on $P(R|S)$ which is given by the single cell response data, and $P(S)$ which is chosen or assumed.

In the limit of infinitely small bins but infinitely many datapoints per bin, the discrete mutual information computed using Eq. 2.1.1 converges to the true continuous value. However, given finite (limited) data, direct estimates of mutual information using Eq. 2.1.1 are biased (*44*). Bias likewise contaminates estimates of the maximum mutual information, also known as the channel capacity (*9*). Since we are able to obtain large samples, typically consisting of ~350 single cell responses per signal value, we are far away from the severely undersampled regime (*40, 45*), and the bias resulting from finite sample size can be corrected by adapting universal estimators described in (*41, 46*).

In particular, we consider the series expansion of the mutual information in terms of inverse powers of sample size:

$$I_{\text{biased}} = I_\infty + \frac{a_1}{N} + \frac{a_2}{N^2} + \cdots, \tag{2.1.3}$$

where I_{biased} is the biased estimate of the mutual information, I_∞ is the unbiased estimate of the mutual information, N is the total number of samples, and the a_i are coefficients that depend on underlying distribution of the signal and the response. The quantity I_∞, which we wish to estimate, may be the value of the mutual information under a specific distribution of the signal, or the maximum value under all possible distributions of the signal. When N is sufficiently large, as in our case, terms of second order or larger are negligible in comparison to the first order term $\sim 1/N$, and the estimated mutual information is a linear function of inverse sample size.

We used jackknife sampling to estimate this linear function. In particular, we sampled fractions of the data, ranging from ~60% to 100%, without replacement and computed the discretized mutual information, I_{biased}. Notably, when computing the discretized mutual information, the boundaries of the bins were chosen so that each bin of the marginal distribution $P(R)$ has approximately equal density (under the assumption that $P(S)$ is uniformly distributed), as in (*47*). Then, we plotted the mutual information with respect to inverse sample size, and extrapolated to infinite sample size, i.e. $1/N \to 0$, to obtain I_∞ (Fig. M1A).

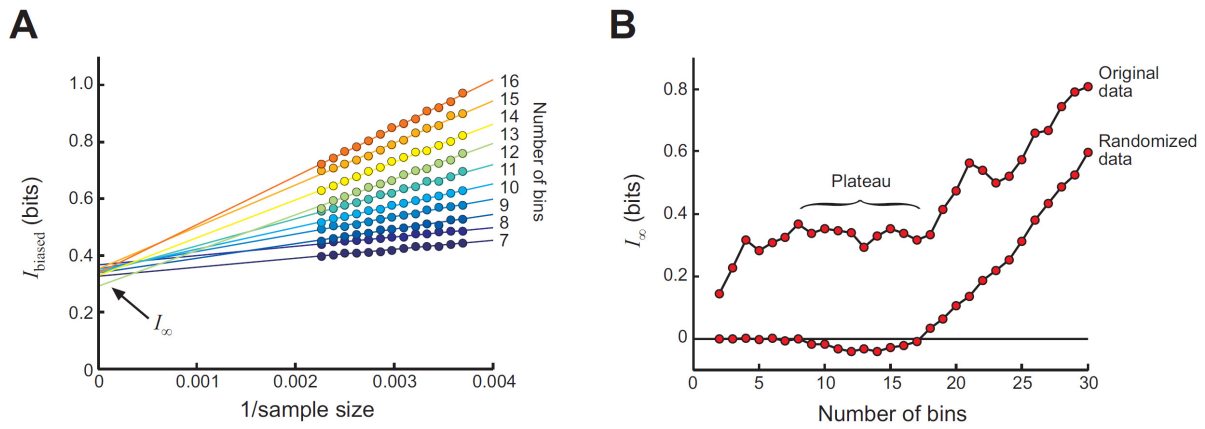

Figure M1: Determination of unbiased mutual information. *A*, Linear extrapolation to infinite sample size to determine I_∞ (see Eq. 2.1.3). *B*, I_∞ plateaus for those numbers of bins for which I_∞ computed for randomized data is slightly, but not statistically significantly, negative. The estimate and error for the unbiased mutual information are taken as the mean and standard deviation, respectively, of the I_∞ values within the plateau. The data shown in this figure illustrate the computation of I(NF-κB;ATF-2|TNF=50ng/mL) at the 30 min. timepoint.

The extrapolation procedure was performed for different numbers of response bins. When the number of bins is small, I_∞ is an underestimate because differential responses are not distinguished by the coarse discretization. For a moderate number of bins, I_∞ is constant, indicating that the unbiased mutual information is captured. The range of bin numbers for which this occurs is also known as the "plateau" region (*46*). For a large number of bins, I_∞ increases because the sample size is not large enough to support very fine discretization, and the linear approximation breaks down. Other popular approaches for selection of the appropriate coarseness of the data binning (*47, 48*) are conceptually very similar.

When computing the channel capacity (see Section 2.2) for a single response (scalar), e.g. the maximum value of I(NF-κB;TNF), we observed that the plateau region extended to at least 50 bins, a result of the large sample size (~350 cells per TNF concentration). The mutual information and its error was estimated as the average and standard deviation, respectively, of the values of I_∞ obtained from 10 to 50 bins, inclusive. When computing the maximum channel capacity for two responses (vector), e.g. the maximum value of I(NF-κB,ATF-2;TNF), the plateau region was typically between 4 and 15 bins (Fig. M1B). The plateau region was smaller due to the larger ratio between response space and the number of datapoints for two responses which scales as the square of the number of bins, compared to that for a single response which scales linearly in the number of bins. Furthermore, for the channel capacity of either single or multiple responses, for some bin numbers the value of I_∞ computed on data randomized by shuffling pairings of signals and responses can be negative, though not statistically significantly different than zero (*49*). Empirically, we found that these bin numbers reliably indicated the plateau region. The value and error of the mutual information was likewise

taken as the average and standard deviation, respectively, of the values of I_∞ computed on the non-randomized data in the plateau (Fig. M1B).

2.2 Computing the channel capacity given $P(R|S)$

In this section, we describe the methods used to determine the channel capacity of the signaling unit, $C(R;S)$, that is, the maximum value of $I(R;S)$ under all possible input distributions $P(S)$, given the experimental conditional response data $P(R|S)$. Formally, this can be stated as an optimization problem:

$$C(R;S) = \max_{P(S)} I(R;S) \text{ such that } \begin{cases} \sum_i P(S_i) = 1, \\ P(S_i) \geq 0. \end{cases} \tag{2.2.1}$$

The constraints ensure that the probability associated with each signal bin is between 0 and 1 inclusive, and the total probability sums to 1. Importantly, since $I(R;S)$ is a concave function of $P(S)$, and the constraints are also concave (linear) functions of $P(S)$, there is a single global maximum for $I(R;S)$ (*9*).

The concavity of $I(R;S)$ enables easy identification of its maximum value and the corresponding $P(S)$ by a variety of algorithms. One fast and simple method to maximize the mutual information is the well-known Blahut-Arimoto algorithm (*9*), which by iteratively optimizing the mutual information over the marginal and conditional distributions of the input, converges on the input distribution that yields the maximum mutual information. The solution identified by the algorithm was checked using the Karush-Kuhn-Tucker conditions, which for this problem were both necessary and sufficient conditions satisfied by the optimal solution (*50*). The Blahut-Arimoto algorithm can further be run on jackknife samples as described above in Sec. 2.1, in order to obtain unbiased estimates of the maximum mutual information.

It is well-known that the $P(S)$ that yields the global maximum may be highly spiky or discontinuous, which might not represent a physically reasonable distribution. Hence, it is prudent to consider the maximum information that can be achieved when $P(S)$ is constrained to be "smooth" in some sense. Smoothness constraints are cumbersome to implement and not guaranteed to yield optimal solutions using a modified Blahut-Arimoto algorithm (*51*), but these difficulties can be surmounted using linear constraints and a gradient ascent method. In particular, in order to enforce additional constraints on $P(S)$, we utilized Matlab's `fmincon` function. (Technically, `fmincon` minimizes a function, but by using $-I(R;S)$ as the objective function, the maximum value of $I(R;S)$ is identified.) In the absence of additional constraints, `fmincon` and the Blahut-Arimoto algorithm yielded identical results.

We note that signals that are produced from multiple sources, as in the case of inflammatory signaling, can exhibit a unimodal (normal-like) shape, or they can be bimodal (e.g. inflammation that is either absent or present), with each of the modes having a similar shape for the same reason. This suggests using a definition of "smoothness" that is somewhat different from traditional constraints on derivatives of the distribution (see, e.g., (*52*)). Namely, we insist that the distribution $P(S)$ that attains the channel capacity is either unimodal or bimodal.

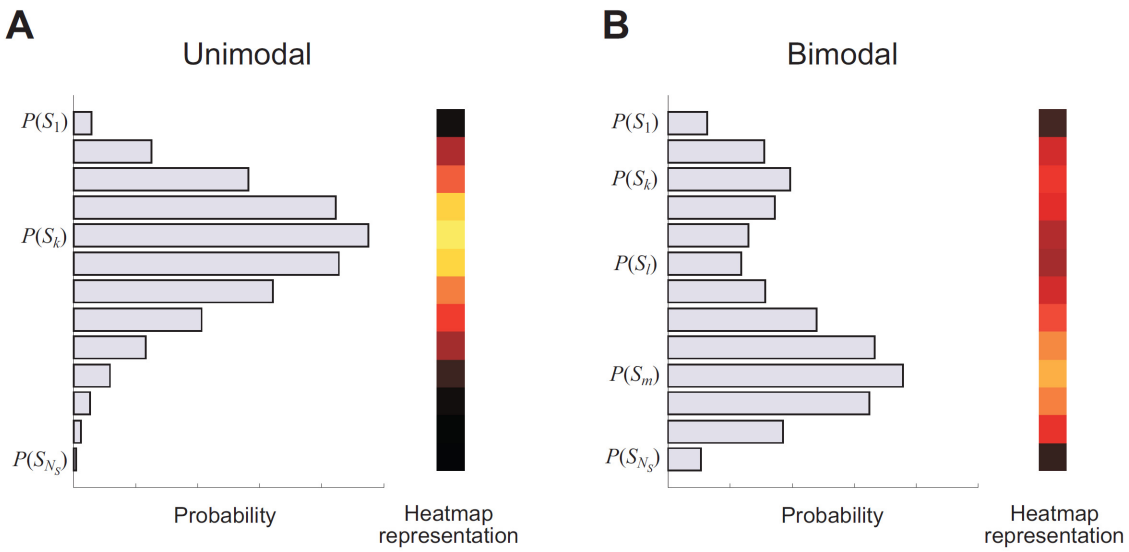

Figure M2: Schematic representation of unimodal and bimodal constraints. *A*, Unimodal probability distribution for the signal where the peak occurs at signal value S_k. *B*, Bimodal probability distribution for the signal where the peaks occur at S_k and S_m, with a local minimum at S_l. The corresponding heatmap representations are shown for comparison to Fig. S1.

First, we explored the information capacity that could be obtained if $P(S)$ was constrained to be a unimodal distribution (see Fig. S1). The corresponding optimization problem was written as:

$$\max_{P(S)} I(R;S) \text{ such that } \begin{cases} \sum_i P(S_i) = 1, \\ P(S_i) \geq 0, \\ P(S_1) \leq P(S_2) \leq \cdots \leq P(S_k), \\ P(S_k) \geq P(S_{k+1}) \geq \cdots \geq P(S_{N_S}). \end{cases} \tag{2.2.2}$$

for $1 \leq k \leq N_S$. The additional constraints ensured that the single peak of the input distribution is at $P(S_k)$ (Fig. M2A). The maximization was then performed for each of the N_S possible positions of the peak. For the TNF dose response experiments, the value of N_S was 13.

Second, we explored the mutual information that could be obtained if $P(S)$ was constrained to be a bimodal distribution (see Fig. S1). The corresponding optimization problem was:

$$\max_{P(S)} I(R;S) \text{ such that } \begin{cases} \sum_i P(S_i) = 1, \\ P(S_i) \geq 0, \\ P(S_1) \leq P(S_2) \leq \cdots \leq P(S_k), \\ P(S_k) \geq P(S_{k+1}) \geq \cdots \geq P(S_l), \\ P(S_l) \leq P(S_{l+1}) \leq \cdots \leq P(S_m), \\ P(S_m) \geq P(S_{m+1}) \geq \cdots \geq P(S_{N_S}). \end{cases} \tag{2.2.3}$$

for $1 \le k < l < m \le N_S$. These constraints ensured that the two peaks of the input distributions occur at $P(S_k)$ and $P(S_m)$ and the intervening local minimum occurred at $P(S_l)$ (Fig. M2B). The maximization was then performed for the $\binom{N_S}{3}$ possibilities for the locations of the two peaks and the local minimum. For the TNF dose response experiments, all $\binom{N_S}{3} = \binom{13}{3} = 286$ possibilities were tested.

For both the unimodal and bimodal constrained optimizations, we note that the added constraints are concave (linear) functions of $P(S_i)$. As a result, the Karush-Kuhn-Tucker conditions again guarantee existence of a unique global optimum and enable it to be verified (*50*).

To enable a fair comparison of the maximum mutual information under no, unimodal, or bimodal constraints (as shown in Fig. S1), we performed all calculations using $N_R = 10$ response bins without performing bias corrections. Due to the large sample size, we estimate that the bias is less than 0.017 bits (using the formulas of (*53*)), and thus does not affect the conclusions drawn. In all other figures and text, the maximum mutual information is reported without unimodal or bimodal constraints and is corrected for bias using the method described above in Sec. 2.1.

2.3 Computing $I(R_1;R_2|S)$ given $P(R_1,R_2|S)$

In this section, we describe the method used to compute, directly from the data, the mutual information between two responses resulting from a specific signal value. The corresponding formula is:

$$I(R_1;R_2 \mid S) = \sum_{R_1,R_2} P(R_1,R_2 \mid S) \log_2 \frac{P(R_1,R_2 \mid S)}{P(R_1 \mid S)P(R_2 \mid S)}. \qquad (2.3.1)$$

Notably, in comparison to the procedures used to maximize mutual information (Sec. 2.2), computing $I(R_1;R_2|S)$ can be performed solely with the conditional response data $P(R_1,R_2|S)$ and does not require any assumptions about other distributions. In particular, one does not need to assume the distribution $P(S)$. Nonetheless, bias correction must still be performed to yield reliable estimates of the mutual information.

The bias correction is performed similarly to the method described above (Sec. 2.1). The data is binned into N_R bins along the first response R_1 and N_R bins along the second response R_2, with the bin boundaries chosen so that the marginal distributions are equally partitioned into the bins. Jackknife samples are used to extrapolate to the mutual information I_∞ that would be obtained with infinite sample size, as $1/N \to 0$. Then I_∞ is plotted versus the number of bins, N_R, and the plateau region is identified as the bin numbers for which I_∞ computed on randomized data is slightly negative. The unbiased estimate of the mutual information and its error are taken as the average and standard deviation of I_∞ values within the plateau (as in Fig. M1B).

2.4 Computing $I(R_1,R_2;S)$ assuming conditionally independent responses given the signal

The key assumption of the bush network model (see Sec. 5.2) is that the responses are conditionally independent given the signal. For the case of two responses, R_1 and R_2, this implies that

$$P_{\text{bush}}(R_1, R_2 \mid S) = P(R_1 \mid S)P(R_2 \mid S)\,. \tag{2.4.1}$$

The joint conditional distribution, constructed in this way from the marginals, can be used to estimate the channel capacity that could be obtained if the responses were the result of signaling via a bush network. The computation is performed by maximizing the mutual information yielded by $P_{\text{bush}}(R_1,R_2|S)$ over all possible $P(S)$ using the algorithms described in Sec. 2.2 to yield unbiased estimates of the maximum mutual information.

3. EFFECT OF EXPERIMENTAL NOISE ON MUTUAL INFORMATION

In this section, we determine the amount of observed cell-to-cell variability that can be ascribed to true biological variability versus experimental noise, in order to evaluate the degree to which estimates of mutual information are affected by experimental noise. With respect to the experimental noise, we are primarily concerned with the accuracy with which concentrations of cellular species, particularly nuclear NF-κB, can be determined by immunofluorescence. Analogous to the method used to separate total noise into extrinsic and intrinsic noise (*5*), the total observed variability can be partitioned into true biological variability and immunochemical noise by simultaneous co-measurement of the species of interest.

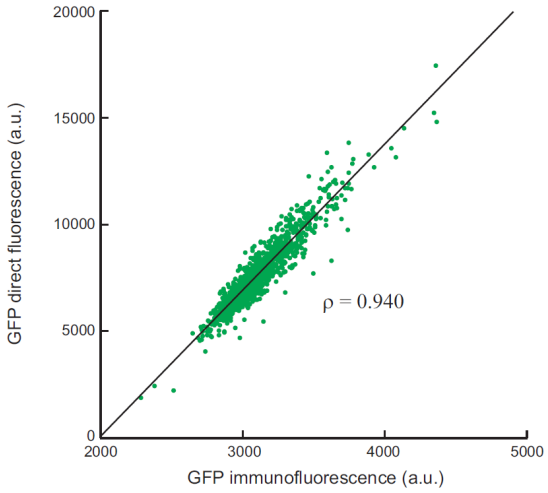

Figure M3: Experimental variability associated with immunofluorescence. Cells stably expressing GFP in the nucleus were fixed and immunostained for GFP. In each cell, nuclear GFP concentration was determined by measuring direct fluorescence from GFP and by GFP immunofluorescence. The graph shows the GFP measurements obtained for 1,096 cells. There is a tight linear relationship between direct fluorescence (proportional to GFP concentration) and immunofluorescence, with a correlation coefficient of 0.940.

First, we determined the level of experimental noise that can be generally ascribed to immunofluorescence. Using cells stably expressing GFP, we measured nuclear concentrations of GFP by direct measurement of GFP fluorescence and by immunofluorescence using GFP-specific antibodies (see Sec. 1.2 for detailed methods). We observed an excellent linear correspondence between the direct and stained GFP measurements, with a correlation coefficient of $\rho = 0.940$ (Fig. M3). Now, if we take the direct GFP measurement to be (proportional to) the true GFP concentration, then it is reasonable to define the experimental noise as the variance of the stained GFP measurement given the true value determined by direct fluorescence. Likewise, the total variability is given by the variance of the stained GFP measurement. Then, under Gaussian assumptions (cf. Eq. 5.1.4), the fraction of the total variability resulting from experimental noise is

$$\frac{\text{var}(\text{stained GFP} \mid \text{true GFP})}{\text{var}(\text{stained GFP})} = \frac{(1-\rho^2)\,\text{var}(\text{stained GFP})}{\text{var}(\text{stained GFP})} = 1-\rho^2 . \qquad (3.1)$$

Thus, about 12% ($1 - 0.940^2 = 0.116$) of the total observed variance resulted from immunofluorescence noise. In reality, the direct GFP fluorescence is a slightly noisy (due to shot noise, etc.) measurement of the true GFP concentration. This extra source of noise implies that 12% is a slight over-estimate of the actual portion of the total variance that results from immunofluorescence.

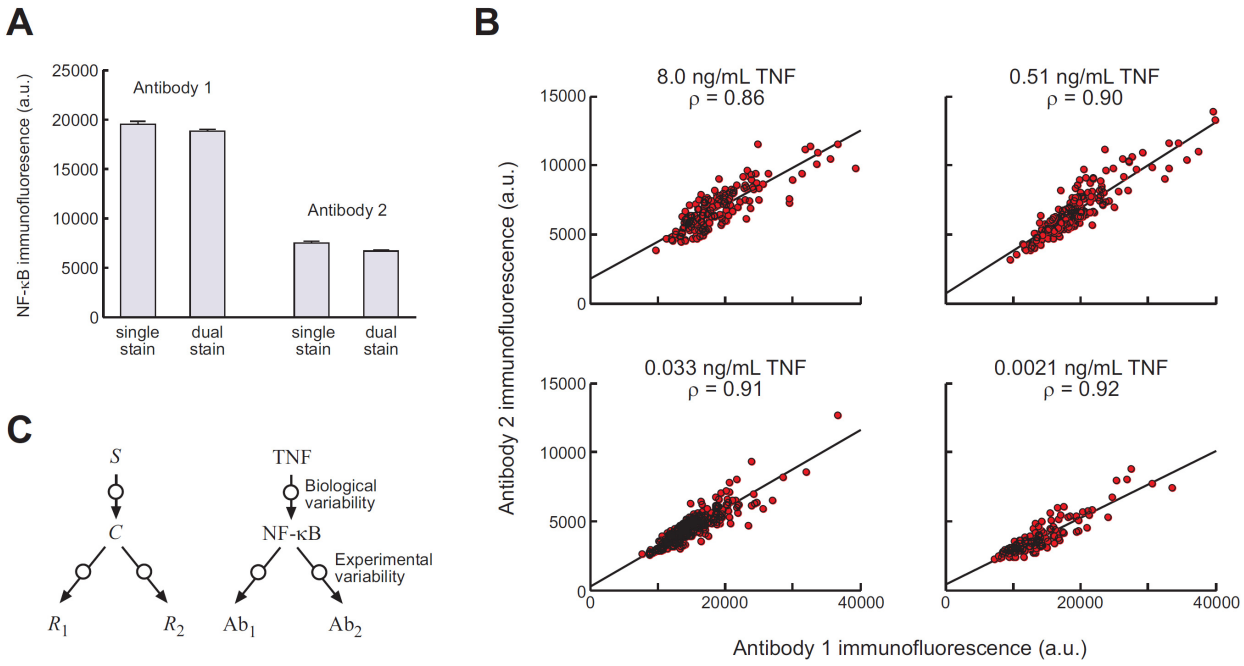

Figure M4: Experimental variability associated with NF-κB immunofluorescence. *A*, Wildtype fibroblasts exposed to 8.0 ng/mL TNF for 30 minutes were stained with two different antibodies specific to NF-κB applied individually (single stain) or simultaneously (dual stain). The average NF-κB immunofluorescence was similar for single and dual staining, indicating minimal interference between the two antibodies. *B*, Wildtype fibroblasts were exposed to the indicated concentrations of TNF for 30 minutes, then dual stained for NF-κB. At all concentrations, there was a tight linear relationship between the immunofluorescence of the two antibodies with a correlation coefficient of ~0.90. *C*, Variability in the dual staining experiment can be analyzed as a tree network. The trunk of the network transduces TNF dose into the true NF-κB concentration, and the branches transduce the true NF-κB concentration into the concentration measured by the antibodies (Ab_1, Ab_2) by immunofluorescence. The variability associated with the trunk represents the true biological variability, and the variability associated with the branches represents experimental noise.

To confirm this result specifically for immunofluorescence measurements of NF-κB, we performed another experiment in which the p65 subunit of NF-κB was simultaneously stained by two distinct antibodies. The antibodies were chosen to be specific to different termini of p65 to minimize interference with one another. We confirmed that dual staining did not substantially affect the measurements yielded by the individual antibodies (Fig. M4A). We found that, across a wide range of TNF concentrations, there was a linear correspondence between the two stained NF-κB measurements with a correlation coefficient of $\rho \approx 0.90$ (Fig. M4B). Since both stained measurements are affected by experimental noise, neither measurement should be taken to represent the true NF-κB concentration, and Eq. 3.1 does not apply. Instead, we note that, under Gaussian assumptions, the correlation between the joint measurements is the product of the correlations between each measurement and the true value:

$$\rho^2_{R_1,R_2|S} = \rho^2_{C,R_1|S}\rho^2_{C,R_2|S}\,, \tag{3.2}$$

where R_1 and R_2 are the measured levels of NF-κB and C is the actual level of NF-κB. (This expression can be obtained, for example, by considering a Gaussian tree network in which the trunk represents biological

variability and the branches represent experimental noise (Fig. M4C).) Since, in this experiment, the measurement noises both result from immunofluorescence, we expect that their contributions to the total variability are similar, i.e. $\rho^2_{C,R_1|S} \approx \rho^2_{C,R_2|S}$. Under this assumption, then, the fraction of the observed variance that can be ascribed to measurement noise is

$$\frac{\sigma^2_{C\to R_1}}{\mathrm{var}(R_1 \mid S)} \approx 1 - \rho_{R_1,R_2|S}\,. \tag{3.3}$$

In our experiment, this shows that ~10% (1 – 0.90) of the total observed variance is due to experimental noise and the rest is true biological variability. This result is consistent with the conservative estimate of 12% obtained from the GFP experiment above.

Next, we estimate the effect of this level of experimental noise on the measured amount of mutual information. We note that, for Gaussian communication channels, the mutual information is determined by the signal-to-noise ratio, ϕ, as in Eq. 6.1.1:

$$I(R;S) = \frac{1}{2}\log_2\left(1 + m^2\frac{\sigma_S^2}{\sigma^2_{S\to R}}\right) \quad\Rightarrow\quad \phi \equiv m^2\frac{\sigma_S^2}{\sigma^2_{S\to R}} = 2^{2I} - 1\,. \tag{3.4}$$

For the TNF-NF-κB pathway, whose maximum mutual information is I(NF-κB;TNF) = 0.916 bits, the corresponding signal-to-noise ratio is $\phi = 2.56$. The above experiments show that approximately 10% of the denominator of ϕ is due to experimental noise. Thus, continuing the Gaussian assumption, the true value could be as high as 2.56/(1 – 0.90) = 2.84. Plugging into Eq. 3.4, this implies that the true maximum mutual information may be 0.971 bits. Stated another way, the mutual information between the true p65 concentration and the antibody measurement is not smaller than $\sim \frac{1}{2}\log_2(1-0.90) = 1.66$ bits, which is substantially larger than the measured channel capacity of about 0.92 bits between the TNF signal and the antibody measurement. Hence, the measurement itself is not a bottleneck that substantially decreases the apparent value of the mutual information, whether the TNF-NF-κB relation is Gaussian or not.

Finally, we note that in the TNF-NF-κB pathway, accounting for experimental noise as an additive Gaussian process led to correcting the channel capacity by about 0.055 bits. For other signal-response pairs (e.g. Table S1) in which the initial estimate for mutual information is lower than that of the TNF-NF-κB pathway, accounting for experimental noise will lead to a smaller increase due to the monotonic relationship between mutual information and ϕ. Thus, in this study, 0.055 bits is the largest and most conservative value for the extent to which mutual information is underestimated due to experimental noise.

4. INFORMATION CAPTURED BY MULTIPLE VERSUS INDIVIDUAL RESPONSES

In this section, we show that the responses of multiple communication channels can obtain more information about a signal than the response of the individual channels. In particular, we explore the values for the mutual information resulting from two responses, $I(R_1,R_2;S)$, can attain relative to the mutual information resulting from the individual responses, $I(R_1;S)$ and $I(R_2;S)$. First, we prove that $I(R_1,R_2;S)$ is at least as large as the greater of $I(R_1;S)$ and $I(R_2;S)$. Then, we prove that if the responses result from independent signaling processes, then $I(R_1,R_2;S)$ is necessarily larger than $I(R_1;S)$ and $I(R_2;S)$. Finally, we show that $I(R_1,R_2;S)$ has no upper bound and can take on large values, for example, if the noise in the two responses is negatively correlated. The reader should consider exploring (*54*) for discussion of relations among mutual informations in more general multivariate dependencies models.

4.1 The lower bound of $I(R_1,R_2;S)$ is the greater of $I(R_1;S)$ and $I(R_2;S)$

The chain rule for mutual information gives the following relation:

$$I(R_1, R_2; S) = I(R_1; S) + I(R_2; S \mid R_1)\,. \tag{4.1.1}$$

Since mutual information is always non-negative, $I(R_2; S \mid R_1) \geq 0$. Thus $I(R_1, R_2; S) \geq I(R_1; S)$. By instead applying the chain rule conditioned on R_2, we can likewise show that $I(R_1, R_2; S) \geq I(R_2; S)$. The combination of these inequalities demonstrates that a lower bound for the information that two responses provide about a signal is

$$I(R_1, R_2; S) \geq \max[I(R_1; S), I(R_2; S)]\,. \tag{4.1.2}$$

This lower bound is achieved when either $I(R_2;S|R_1)$ or $I(R_1;S|R_2)$ equals zero, that is when one response is conditionally independent of the signal given the other response, implying no improvement in information using the two responses together. In other words, the information provided by the two responses together is not smaller than the information provided by the more informative individual response.

Notably, the proof of this lower bound does not depend on whether R_1 or R_2 are scalars or vectors, a fact that will be utilized in Section 4.2.

4.2 $I(R_1,R_2;S)$ is strictly greater than the lower bound if the responses are conditionally independent

In this section, we consider the case in which responses R_1 and R_2 are conditionally independent given the signal, corresponding to the scenario in which the responses are generated by signaling processes that do not interact, other than sharing a common signal. Below, we prove that conditional independence necessarily implies that the mutual information of the responses together is strictly greater than the lower bound, implying a gain of information compared to either response alone. The proof of this statement does not depend on whether R_1 and R_2 are scalars or vectors. Applying the proof to the case in which R_1 and R_2 are scalars implies that the responses of two signaling pathways considered together, one which yields output R_1 and the other which yields output R_2, is more informative about the signal than either pathway alone. If instead R_1 is a vector representing a set of outputs from some (arbitrarily complicated) signaling system then the proof implies that adding the

conditionally independent response R_2, representing either a scalar output of a separate pathway or a vector output of a separate signaling system, also increases the information about the signal.

Theorem: If $I(R_1;S) > 0$, $I(R_2;S) > 0$, and R_1 and R_2 are conditionally independent given S, then $I(R_1,R_2;S) > \max[I(R_1;S), I(R_2;S)]$ (strictly greater than the lower bound).

Proof: Suppose without loss of generality that R_1 is the most informative response, i.e. $I(R_1;S) \geq I(R_2;S) > 0$. The chain rule for mutual information allows us to write

$$I(R_1, R_2; S) = I(R_1; S) + I(R_2; S \mid R_1). \tag{4.2.1}$$

To prove that $I(R_1,R_2;S)$ is strictly greater than the lower bound, $I(R_1;S)$, we must prove that $I(R_2;S|R_1) > 0$. This can be proven by contradiction.

Mutual information cannot be negative, so assume that $I(R_2;S|R_1) = 0$. This implies that R_2 and S are conditionally independent given R_1, which implies that for any given values of R_1, R_2, and S, the following holds:

$$\begin{aligned} P(R_2, S \mid R_1) &= P(R_2 \mid R_1) P(S \mid R_1) \\ \frac{P(R_1, R_2, S)}{P(R_1)} &= P(R_2 \mid R_1) \frac{P(R_1, S)}{P(R_1)} \\ P(R_1, R_2 \mid S) P(S) &= P(R_2 \mid R_1) P(R_1 \mid S) P(S) \\ P(R_2 \mid S) &= P(R_2 \mid R_1), \end{aligned} \tag{4.2.2}$$

where in the last line we used the conditional independence of R_1 and R_2 given S. Since this holds for all values, we can sum the equation over all possible values of R_1, yielding

$$\begin{aligned} \sum_{R_1} P(R_1) P(R_2 \mid S) &= \sum_{R_1} P(R_1) P(R_2 \mid R_1) \\ P(R_2 \mid S) \sum_{R_1} P(R_1) &= P(R_2) \\ P(R_2 \mid S) &= P(R_2). \end{aligned} \tag{4.2.3}$$

Finally, this implies that

$$P(R_2, S) = P(R_2 \mid S) P(S) = P(R_2) P(S). \tag{4.2.4}$$

This shows that a necessary condition for $I(R_1,R_2;S)$ to equal the lower bound is that R_2 and S are unconditionally independent. However, this would imply that R_2 is not informative about S, contradicting the assumption that $I(R_2;S) > 0$. Therefore, the conditions of the claim imply that $I(R_1,R_2;S)$ is strictly greater than $I(R_1;S)$ and also strictly greater than $I(R_2;S)$. □

4.3 The upper bound of $I(R_1,R_2;S)$ is infinity

In this section we show that $I(R_1,R_2;S)$ has an infinite upper bound, by considering a simple example. Consider the case in which the responses are scalars given by the equations

$$\begin{cases} R_1 = S + \gamma_1 \\ R_2 = S + \gamma_2 \end{cases}, \tag{4.3.1}$$

where γ_1 and γ_2 are noise terms independent of S. If the noise terms have non-zero variance, then the information provided by each individual response, $I(R_1;S)$ and $I(R_2;S)$, is finite.

Now, suppose further that the noise terms are correlated. In the extreme, suppose that they are exactly negatively correlated such that $\gamma_1 = -\gamma_2$. Biologically, this situation might be approached if the there is strong mutually repressive crosstalk between the two pathway branches, or when both branches are competing for the same signaling molecule to activate them. Then, given knowledge of R_1 and R_2, their average yields:

$$\begin{aligned} \tfrac{1}{2}(R_1 + R_2) &= S + \tfrac{1}{2}(\gamma_1 + \gamma_2) \\ &= S. \end{aligned} \tag{4.3.2}$$

Hence, knowledge of R_1 and R_2 allows the noiseless recovery of the exact value of S. If S is a continuous variable, which requires an infinite number of bits to specify exactly, then $I(R_1, R_2; S) = \infty$.

More rigorously, using the methods of Sec. 5, one can show that the mutual information of the system described by Eq. 4.3.1 is $I(R_1, R_2; S) = \frac{1}{2}\log_2\left(1 + 2\frac{\sigma_S^2}{\sigma_{S\to R}^2}\frac{1}{1+\rho}\right)$ if S is a normally distributed stochastic variable with variance σ_S^2, and γ_1 and γ_2 are normally distributed each with variance $\sigma_{S\to R}^2$ and correlation ρ. (See also (*55*).) As $\rho \to -1$, it is easy to see that $I(R_1, R_2; S) \to \infty$. From this example, we conclude that $I(R_1, R_2; S)$ is unbounded from above.

5. INFORMATION THEORETIC ANALYSIS OF BUSH AND TREE NETWORKS

5.1 Preliminaries

In this section, we consider signaling networks that take a single signal S and broadcast the signal out to n communication channels yielding the responses $R_1, R_2, \ldots, R_n$. We are interested in the amount of information that the responses jointly yield about the signal, i.e. $I(R_1,\ldots,R_n;S)$. To gain semiquantitative insight into such pathways, we assume that $(R_1, R_2, \ldots, R_n, S)$ is a multivariate normal distribution of dimension $n + 1$, as detailed in the sections below. The Gaussian assumption enables the mutual information to be solved analytically. The resulting formulas allow us to understand the relative influences of the various sources of noise on the information gathering ability of the signaling network and to predict the value of the mutual information. In order to provide a self-contained description of the theoretical framework that is accessible to both specialists and non-specialists alike, we here provide a complete and detailed derivation of the formulas. However, we caution the reader that the formulas will not hold, in general, for non-Gaussian distributions of the variables.

First, we establish some mathematical formulas which will be used in the derivation of the mutual information for specific network structures. First, a well-known result in information theory is that a multivariate normal distribution of dimension n has an entropy of

$$H = \frac{1}{2}\log_2\left((2\pi e)^n |\Sigma|\right), \tag{5.1.1}$$

where $|\Sigma|$ is the determinant of the covariance matrix of the distribution (*9*). Since the marginal and conditional distributions of a multivariate normal distribution are themselves normal, it is easy to see that

$$\begin{aligned} I(R_1,\ldots,R_n;S) &= H(R_1,\ldots,R_n) - H(R_1,\ldots,R_n \mid S) \\ &= \frac{1}{2}\log_2\left((2\pi e)^n \left|\Sigma_{\vec{R}}\right|\right) - \frac{1}{2}\log_2\left((2\pi e)^n \left|\Sigma_{\vec{R}|S}\right|\right) \\ &= \frac{1}{2}\log_2\left(\frac{\left|\Sigma_{\vec{R}}\right|}{\left|\Sigma_{\vec{R}|S}\right|}\right), \end{aligned} \tag{5.1.2}$$

where $\left|\Sigma_{\vec{R}}\right|$ and $\left|\Sigma_{\vec{R}|S}\right|$ are the determinants of the covariance matrix of the responses and the responses given the signal, respectively.

If we consider just one response, R, then the determinants are given by

$$\left|\Sigma_R\right| = \mathrm{var}(R), \tag{5.1.3}$$

$$\left|\Sigma_{R|S}\right| = \mathrm{var}(R \mid S) = (1-\rho^2)\,\mathrm{var}(R), \tag{5.1.4}$$

yielding

$$I(R;S) = -\frac{1}{2}\log_2(1-\rho^2)\,. \tag{5.1.5}$$

As expected intuitively, when there is zero correlation between R and S, their mutual information is zero. In comparison, the information increases as the correlation approaches +1 or –1. If the correlation is perfect (exactly +1 or –1), the information is infinite. Note that this deterministic relation between the information and the correlation is a direct consequence of Gaussian assumption about the involved variables. In general, mutual information among two variables is not smaller than the value calculated using the Gaussian assumption.

Finally, we establish the following lemma, which enables us to compute the determinants for multiple responses resulting from either bush or tree signaling networks:

Lemma: The determinant of the $n \times n$ matrix $\mathbf{Q}$ whose entries are given by $q_{ij} = m_i m_j a + \delta_{ij} b_i$ is $\left(\prod_i b_i\right)\left(1+\sum_i m_i^2 \frac{a}{b_i}\right)$. Here, the Kronecker delta notation ($\delta_{ij} = 1$ if $i = j$, and is zero otherwise) indicates that b_i terms only appear in the diagonal elements of $\mathbf{Q}$.

Proof: A basic property of the matrix determinant is that it is invariant to elementary row addition and subtraction (also known as Gaussian elimination). That is, adding or subtracting a multiple of one row to/from another row does not change the determinant. Therefore, the determinant does not change if we subtract $\frac{m_i}{m_{i-1}}$ times row $i-1$ from row i, for each of the rows $i = n, n-1, \ldots, 2$. These operations yield:

$$\begin{aligned} |\mathbf{Q}| &= \begin{vmatrix} m_1^2 a + b_1 & m_1 m_2 a & m_1 m_3 a & m_1 m_4 a & \cdots \\ m_2 m_1 a & m_2^2 a + b_2 & m_2 m_3 a & m_2 m_4 a & \cdots \\ m_3 m_1 a & m_3 m_2 a & m_3^2 a + b_3 & m_3 m_4 a & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix} \\ &= \begin{vmatrix} m_1^2 a + b_1 & m_1 m_2 a & m_1 m_3 a & m_1 m_4 a & \cdots \\ -\frac{m_2}{m_1} b_1 & b_2 & 0 & 0 & \cdots \\ 0 & -\frac{m_3}{m_2} b_2 & b_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}. \end{aligned} \tag{5.1.6}$$

Next, we reduce row 2 by adding $\frac{\frac{m_2}{m_1} b_1}{m_1^2 a + b_1}$ times row 1 to row 2, yielding:

$$= \begin{vmatrix} d(1) & m_1 m_2 a & m_1 m_3 a & m_1 m_4 a & \cdots \\ 0 & \frac{d(2)}{d(1)} & \frac{m_2 m_3 a b_1}{d(1)} & \frac{m_2 m_4 a b_1}{d(1)} & \cdots \\ 0 & -\frac{m_3}{m_2} b_2 & b_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}. \tag{5.1.7}$$

where we have introduced the notation $d(n)=\left(\prod_{i=1}^{n} b_i\right)\left(1+\sum_{i=1}^{n} m_i^2 \frac{a}{b_i}\right)$. Now, we claim that after reduction of subsequent rows, that the diagonal element of row $k \geq 2$ equals $\frac{d(k)}{d(k-1)}$, and the elements of row k to the right in columns $h = k + 1, \ldots, n$ equal $\frac{m_k m_h a \prod_{i=1}^{k-1} b_i}{d(k-1)}$.

The claim can be proven by induction. Clearly the claim holds for row 2. Assume the claim holds for row k. Then, the reduction of row $k + 1$ is performed by multiplying row k by $\frac{\frac{m_{k+1}}{m_k} b_k}{d(k)/d(k-1)}$ and adding it to row $k + 1$. For the diagonal element of row $k + 1$, this yields

$$\underbrace{\frac{m_k m_{k+1} a \prod_{i=1}^{k-1} b_i}{d(k-1)}}_{\text{the element above the diagonal of row } k+1} \frac{\frac{m_{k+1}}{m_k} b_k}{d(k)/d(k-1)} + b_{k+1} = \frac{m_{k+1}^2 a \prod_{i=1}^{k} b_i + b_{k+1} d(k)}{d(k)}$$

$$= \frac{\left(\prod_{i=1}^{k+1} b_i\right)\left(m_{k+1}^2 \frac{a}{b_{k+1}} + 1 + \sum_{i=1}^{k} m_i^2 \frac{a}{b_i}\right)}{d(k)}$$

$$= \frac{d(k+1)}{d(k)}. \tag{5.1.8}$$

and for the element in column $h > k + 1$ to the right of the diagonal, the row reduction yields

$$\frac{m_k m_h a \prod_{i=1}^{k-1} b_i}{d(k-1)} \frac{\frac{m_{k+1}}{m_k} b_k}{d(k)/d(k-1)} = \frac{m_{k+1} m_h a \prod_{i=1}^{k} b_i}{d(k)}. \tag{5.1.9}$$

This proves the claim also holds for row $k + 1$, completing the induction.

Since the determinant of the fully row reduced (upper triangular) matrix is the product of its diagonal elements, the desired determinant telescopes to

$$|\mathbf{Q}| = d(1)\frac{d(2)}{d(1)}\frac{d(3)}{d(2)}\cdots\frac{d(n)}{d(n-1)} = d(n) = \left(\prod_{i=1}^{n} b_i\right)\left(1+\sum_{i=1}^{n} m_i^2 \frac{a}{b_i}\right) \tag{5.1.10}$$

as claimed. □

5.2 Information captured by a Gaussian bush network

In this section, we derive formulas for the mutual information, under Gaussian conditions, between a signal and multiple linear responses activated by a "bush" network. The key feature of a bush network is that the network branches into multiple signaling pathways at the level of the signal, so that each response is conditionally independent given the signal.

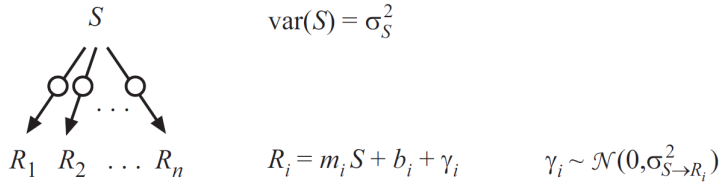

Figure M5: Model of a bush signaling network. Each pathway in the network transduces the signal into a linear response R_i, with gain m_i, bias b_i, and noise γ_i.

The formal formulation of this model is as follows (Fig. M5). The signal S is a normally distributed stochastic variable with variance σ_S^2. Each pathway $i = 1, 2, \ldots, n$ yields a linear response $R_i = m_i S + b_i + \gamma_i$ where m_i and b_i are the slope (gain) and intercept (bias) respectively between R_i and S in the absence of cellular variability, and γ_i is a stochastic variable representing cellular variability in the response R_i. We assume that γ_i is normally distributed from cell-to-cell with variance $\sigma_{S\to R_i}^2$ and that the γ_i terms are independent of each other. As a result, each of the R_i is normally distributed because each is the sum of two normally distributed variables, and the R_i are conditionally independent given the signal S. Note that in this general formulation that each pathway can have different values for the slope, intercept, and magnitude of noise.

Since the variance of independent variables add, the variance of each response is $\mathrm{var}(R_i) = m_i^2\sigma_S^2 + \sigma_{S\to R_i}^2$. Similarly, the covariance between any two responses is $\mathrm{cov}(R_i, R_j) = m_i m_j \sigma_S^2$ (for all $i \neq j$). Thus, using the lemma in Sec. 5.1, the determinant of the response covariance matrix is:

$$\left(\Sigma_{\vec{R}}\right)_{ij} = m_i m_j \sigma_S^2 + \delta_{ij}\sigma_{S\to R_i}^2 \Rightarrow \left|\Sigma_{\vec{R}}\right| = \left(\prod_{i=1}^{n}\sigma_{S\to R_i}^2\right)\left(1+\sum_{i=1}^{n} m_i^2 \frac{\sigma_S^2}{\sigma_{S\to R_i}^2}\right). \tag{5.2.1}$$

When S is given, the variance and covariance terms reduce to $\mathrm{var}(R_i \mid S) = \sigma_{S\to R_i}^2$ and $\mathrm{cov}(R_i, R_j \mid S) = 0$. Then, again using the lemma, the corresponding determinant evaluates to

$$\left(\Sigma_{\vec{R}|S}\right)_{ij} = \delta_{ij}\sigma_{S\to R_i}^2 \Rightarrow \left|\Sigma_{\vec{R}|S}\right| = \left(\prod_{i=1}^{n}\sigma_{S\to R_i}^2\right). \tag{5.2.2}$$

Finally, using Eq. 5.1.2, the mutual information between the responses together and the signal is

$$I(R_1,\ldots,R_n;S) = \frac{1}{2}\log_2\left(1+\sum_{i=1}^{n} m_i^2 \frac{\sigma_S^2}{\sigma_{S\to R_i}^2}\right). \tag{5.2.3}$$

The ratio $\sigma_S^2 / \sigma_{S\to R_i}^2$ can be considered to be a signal-to-noise ratio (*9*), where σ_S^2 represents the signal power and $\sigma_{S\to R_i}^2$ is the noise (variance) introduced in transmitting from *S* to R_i. The slope m_i can be considered to be a factor that normalizes $\sigma_{S\to R_i}^2$, or more specifically, allows the individual $\sigma_{S\to R_i}^2$ to be compared in similar units. Thus, the mutual information of the *n* responses together can be obtained by summing the signal-to-noise ratios of the *n* pathways, when those ratios are given in comparable units. The formula also enables determination of which pathways dominate the mutual information obtained by integrating multiple responses together.

When the *n* pathways are equivalent the formula simplifies to Eq. 2 in the main text. In particular, if all the $m_i = 1$ and the magnitude of the pathway variability is the same $\sigma_{S\to R}^2 = \sigma_{S\to R_i}^2$ for each pathway $i = 1, \ldots, n$, then the mutual information is:

$$I(R_1,...,R_n;S) = \frac{1}{2}\log_2\left(1 + n\frac{\sigma_S^2}{\sigma_{S\to R}^2}\right). \qquad (5.2.4)$$

As expected intuitively, the formula reveals that the information increases as the noise introduced by each branch ($\sigma_{S\to R}^2$) decreases with respect to the spread in the input (σ_S^2). Furthermore, the information grows logarithmically with the number of responses measuring the signal, in an unbounded fashion.

5.3 Information captured by a Gaussian tree network

In this section, we derive formulas for the mutual information, under Gaussian conditions, between a signal and multiple linear responses activated by a "tree" network. The key feature of a tree network is that the signal activates a common "trunk" before branching into the individual pathways. The trunk terminates at the point of branching denoted as *C*, i.e. the last common intermediate shared by the pathways. Thus, the responses are conditionally independent given *C*, but not conditionally independent given the signal. In comparison, responses of bush network are conditionally independent given the signal.

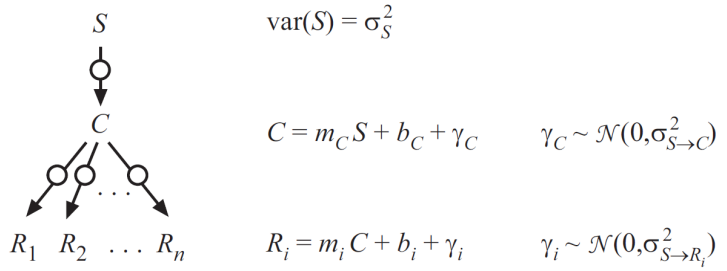

Figure M6: Model of a tree signaling network. The common trunk of the network transduces the signal into the intermediate linear response *C* with gain m_C, bias b_C, and noise γ_C. Each downstream pathway branch then transduces *C* into the linear response R_i with gain m_i, bias b_i, and noise γ_i.

The formal formulation of the tree network model (Fig. M6) is similar to that of bush network model. The signal *S* is a normally distributed stochastic variable with variance σ_S^2. *C* is the last common intermediate in the pathways measuring the signal, with $C = m_C S + b_C + \gamma_C$, where m_C and b_C are the slope (gain) and intercept (bias) respectively between *S* and *C* in the absence of cellular variability, and γ_C is a stochastic

variable representing cellular variability in the common trunk. In particular, we assume that γ_C is normally distributed from cell-to-cell with variance $\sigma^2_{S\to C}$.

Each downstream pathway yields a response $R_i = m_i C + b_i + \gamma_i$ where, similarly, m_i and b_i are the slope (gain) and intercept (bias) respectively between R_i and C in the absence of cellular variability, and γ_i is a stochastic variable representing cellular variability in the branch from C to R_i. We assume that γ_i is normally distributed from cell-to-cell with variance $\sigma^2_{S\to R}$. All of the noise terms γ_i and γ_C are independent of each other and independent of S.

Substituting the definition for R_i into the definition of C reveals that R_i is normally distributed, and on average a linear function of S with slope (gain) $m_C m_i$ and intercept (bias) $m_i b_C + b_i$:

$$R_i = m_C m_i S + (m_i b_C + b_i) + m_i \gamma_C + \gamma_i . \tag{5.3.1}$$

From this formula, it is easy to see that the variance of each response is $\mathrm{var}(R_i) = m_C^2 m_i^2 \sigma_S^2 + m_i^2 \sigma^2_{S\to C} + \sigma^2_{C\to R_i}$ and that the covariance between any two responses is $\mathrm{cov}(R_i, R_j) = m_C^2 m_i m_j \sigma_S^2 + m_i m_j \sigma^2_{S\to C}$ (for all $i \neq j$). Thus, using the lemma in Sec. 5.1, the determinant of the response covariance matrix is:

$$\left(\Sigma_{\vec{R}}\right)_{ij} = m_i m_j (m_C^2 \sigma_S^2 + \sigma^2_{S\to C}) + \delta_{ij} \sigma^2_{C\to R_i} \Rightarrow \left|\Sigma_{\vec{R}}\right| = \left(\prod_{i=1}^{n} \sigma^2_{C\to R_i}\right)\left(1 + \sum_{i=1}^{n} m_i^2 \frac{m_C^2 \sigma_S^2 + \sigma^2_{S\to C}}{\sigma^2_{C\to R_i}}\right). \tag{5.3.2}$$

When S is given, the variance and covariance terms reduce to $\mathrm{var}(R_i \mid S) = m_i^2 \sigma^2_{S\to C} + \sigma^2_{C\to R_i}$ and $\mathrm{cov}(R_i, R_j \mid S) = m_i m_j \sigma^2_{S\to C}$ (for all $i \neq j$). Then, again using the lemma, the determinant is:

$$\left(\Sigma_{\vec{R}|S}\right)_{ij} = m_i m_j \sigma^2_{S\to C} + \delta_{ij} \sigma^2_{C\to R_i} \Rightarrow \left|\Sigma_{\vec{R}|S}\right| = \left(\prod_{i=1}^{n} \sigma^2_{C\to R_i}\right)\left(1 + \sum_{i=1}^{n} m_i^2 \frac{\sigma^2_{S\to C}}{\sigma^2_{C\to R_i}}\right). \tag{5.3.3}$$

Finally, using Eq. 5.1.2, the mutual information between the responses together and the signal is

$$\begin{aligned} I(R_1, \ldots, R_n; S) &= \frac{1}{2} \log_2 \left(\frac{1 + \sum_{i=1}^{n} m_i^2 \frac{m_C^2 \sigma_S^2 + \sigma^2_{S\to C}}{\sigma^2_{C\to R_i}}}{1 + \sum_{i=1}^{n} m_i^2 \frac{\sigma^2_{S\to C}}{\sigma^2_{C\to R_i}}} \right) \\ &= \frac{1}{2} \log_2 \left(1 + \frac{\sum_{i=1}^{n} m_i^2 \frac{m_C^2 \sigma_S^2}{\sigma^2_{C\to R_i}}}{1 + \sum_{i=1}^{n} m_i^2 \frac{\sigma^2_{S\to C}}{\sigma^2_{C\to R_i}}} \right). \end{aligned} \tag{5.3.4}$$

Similar to the bush network, the information obtained from a tree network depends on signal-to-noise ratios. The information depends on two key ratios: (1) $\sigma_S^2 / \sigma_{C\to R_i}^2$, the signal power versus the noise in the downstream branches, and (2) $\sigma_{S\to C}^2 / \sigma_{C\to R_i}^2$, the noise in the trunk versus the noise in the downstream branches. The slope m_i can again be considered to be a factor that normalizes the noise in the downstream branch, $\sigma_{C\to R_i}^2$, enabling the noises to be compared in equivalent units. Likewise, the slope m_C normalizes the signal power σ_S^2. Thus, the formula enables determination of which sources of variability dominate the mutual information obtained by integrating multiple responses together.

Notably, the tree network contains a bush network embedded within, i.e. the network consisting of C and the downstream branches. The results for bush networks show that as the number of branches in the tree network grows, the information that the responses together yield about C grows without bound. However, the information that those responses yield about the signal S approaches a limit:

$$n \to \infty \;\Rightarrow\; \sum_{i=1}^{n} \frac{m_i^2}{\sigma_{C\to R_i}^2} \to \infty \;\Rightarrow\; I(R_1,...,R_n;S) \to \frac{1}{2}\log_2\left(1 + m_C^2 \frac{\sigma_S^2}{\sigma_{S\to C}^2}\right) = I(C;S)\,. \tag{5.3.5}$$

The equivalence to $I(C;S)$ can be seen by considering a bush network (Eq. 5.2.3) with a single branch from S to C with slope (gain) m_C and cellular variability magnitude $\sigma_{S\to C}^2$. (The data processing inequality (*9*) yields the same upper limit, i.e. if $S \to C \to (R_1, \ldots, R_n)$ form a Markov chain, then $I(R_1,\ldots,R_n;S) \le I(C;S)$, but Eq. 5.3.5 shows that the limit is actually approached through the use of many pathway branches.) Thus, many downstream branches allow a very accurate and informative estimate of C, but the information that these branches can obtain about S is limited by the bottleneck resulting from noise in the trunk portion of the pathway from S to C.

Finally, when the n downstream branches are equivalent the formula simplifies to Eq. 3 in the main text. In particular, if all the $m_i = 1$ and the magnitude of the variability in the branches is the same $\sigma_{C\to R}^2 = \sigma_{C\to R_i}^2$ for $i = 1, \ldots, n$, and we further assume for simplicity that $m_C = 1$, then the mutual information becomes:

$$I(R_1,...,R_n;S) = \frac{1}{2}\log_2\left(1 + \frac{n\sigma_S^2 / \sigma_{C\to R}^2}{1 + n\sigma_{S\to C}^2 / \sigma_{C\to R}^2}\right). \tag{5.3.6}$$

Again, the simplified formula highlights the dependence of the information on the two key signal-to-noise ratios and the number of downstream branches.

6. PREDICTIONS MADE BY THE BUSH AND TREE NETWORK MODELS

The Gaussian, linear response models for tree and bush networks described in Sec. 5.2 and 5.3 make specific quantitative predictions for mutual information. Both models make predictions for the information that multiple responses yield about the signal, based on the amount of information that the individual responses yield about the signal. The models also predict the mutual information between the responses. For the tree model, one can further predict the information capacity of the trunk. In this section, we derive formulas that enable such predictions. We illustrate the methods given experimental data for $n = 2$ responses, although they generalize to larger n.

6.1 Predicting $I(R_1,R_2;S)$ for the Gaussian bush network

Eq. 5.2.3 shows that the information captured by multiple responses emanating from a bush network depends on the sum of the signal-to-noise ratios for the individual branches. Reversing the relations, these ratios can be obtained from the information captured by the individual responses. In particular, we may compute ϕ_1, the signal-to-noise ratio for branch #1, as follows:

$$I_1 \equiv I(R_1;S) = \frac{1}{2}\log_2\left(1 + m_1^2 \frac{\sigma_S^2}{\sigma_{S\to R_1}^2}\right) \quad \Rightarrow \quad \phi_1 \equiv m_1^2 \frac{\sigma_S^2}{\sigma_{S\to R_1}^2} = 2^{2I_1} - 1 . \tag{6.1.1}$$

Likewise, for branch #2, $\phi_2 = 2^{2I_2} - 1$. Then, Eq. 5.2.3 predicts that the mutual information captured by the two responses together is simply:

$$I_{12} \equiv I(R_1,R_2;S) = \frac{1}{2}\log_2\left(1 + \phi_1 + \phi_2\right). \tag{6.1.2}$$

6.2 Predicting $I(R_1,R_2;S)$ for the Gaussian tree network

Eq. 5.3.4 shows that the information captured by multiple responses emanating from a tree network depends on the sums of two signal-to-noise ratios, namely $\sigma_S^2 / \sigma_{C\to R_i}^2$ and $\sigma_{S\to C}^2 / \sigma_{C\to R_i}^2$, whose values are normalized by the slopes (gains) m_i and m_C. For each branch, the latter ratio $\phi_{C,i} \equiv m_i^2 \sigma_{S\to C}^2 / \sigma_{C\to R_i}^2$ can be obtained by rearranging expressions given in Sec. 5.3 for the overall conditional variance and covariance of the responses:

$$\begin{aligned} \phi_{C,1} &= m_1^2 \frac{\sigma_{S\to C}^2}{\sigma_{C\to R_1}^2} = \frac{\frac{m_1}{m_2}\mathrm{cov}(R_1,R_2|S)}{\mathrm{var}(R_1|S) - \frac{m_1}{m_2}\mathrm{cov}(R_1,R_2|S)} \\ \phi_{C,2} &= m_2^2 \frac{\sigma_{S\to C}^2}{\sigma_{C\to R_2}^2} = \frac{\frac{m_2}{m_1}\mathrm{cov}(R_1,R_2|S)}{\mathrm{var}(R_2|S) - \frac{m_2}{m_1}\mathrm{cov}(R_1,R_2|S)}. \end{aligned} \tag{6.2.1}$$

The variance and covariance terms can be measured directly from the experimental data. The ratio of the slopes (gains) m_2/m_1 (or its inverse) can also be determined experimentally as the slope of the best fit line through the average values of R_2 plotted against the average values of R_1 that are induced by various levels of the signal S.

The other key ratio, $\phi_{S,i} \equiv m_1^2 m_C^2 \sigma_S^2 / \sigma_{C\to R_i}^2$, can be obtained from Eq. 5.3.4 using $\phi_{C,i}$. For branch #1, this is done as follows:

$$I_1 \equiv I(R_1;S) = \frac{1}{2}\log_2\left(1+\frac{m_1^2 m_C^2 \sigma_S^2 / \sigma_{C\to R_1}^2}{1+m_1^2\sigma_{S\to C}^2/\sigma_{C\to R_1}^2}\right) \quad \Rightarrow \quad \phi_{S,1} \equiv m_1^2 m_C^2 \sigma_S^2 / \sigma_{C\to R_1}^2 = (2^{2I_1}-1)(1+\phi_{C,1}). \quad (6.2.2)$$

Likewise, for branch #2, we have $\phi_{S,2} = (2^{2I_2}-1)(1+\phi_{C,2})$. Together, Eq. 5.3.4 then predicts that the mutual information captured by the two responses together is simply:

$$I_{12} \equiv I(R_1,R_2;S) = \frac{1}{2}\log_2\left(1+\frac{\phi_{S,1}+\phi_{S,2}}{1+\phi_{C,1}+\phi_{C,2}}\right). \quad (6.2.3)$$

6.3 Predicting $I(R_1;R_2|S)$ for the Gaussian bush and tree networks

The quantity $I(R_1;R_2|S)$ measures the amount of information one can obtain about a response R_1 with knowledge of the other response R_2, or vice versa, given the signal. It can be measured experimentally, e.g. by performing the computations of Sec. 2.3 on data obtained from single cells co-stained for multiple responses. These experimental measurements can then be compared to the values predicted from the bush and tree models.

The key assumption in the bush model is that the responses are conditionally independent given the signal. Therefore, the bush model predicts $I_{\text{bush}}(R_1;R_2|S) = 0$.

On the other hand, the tree model assumes that the responses are not conditionally independent, and hence $I(R_1;R_2|S)$ is greater than zero. Since R_1 and R_2 are assumed to be jointly normally distributed, the mutual information can be predicted by considering the correlation between the responses (Eq. 5.1.5). In this case, the correlation is:

$$\begin{aligned}
\rho^2 &= \frac{\text{cov}^2(R_1,R_2\,|\,S)}{\text{var}(R_1\,|\,S)\,\text{var}(R_2\,|\,S)} \\
&= \frac{(m_1 m_2 \sigma_{S\to C}^2)^2}{(m_1^2\sigma_{S\to C}^2+\sigma_{C\to R_1}^2)(m_2^2\sigma_{S\to C}^2+\sigma_{C\to R_2}^2)} \\
&= \frac{(m_1^2\,\frac{\sigma_S^2}{\sigma_{C\to R_1}^2})(m_2^2\,\frac{\sigma_S^2}{\sigma_{C\to R_2}^2})}{(m_1^2\,\frac{\sigma_{S\to C}^2}{\sigma_{C\to R_1}^2}+1)(m_2^2\,\frac{\sigma_{S\to C}^2}{\sigma_{C\to R_2}^2}+1)} \\
&= \frac{\phi_{C,1}\phi_{C,2}}{(\phi_{C,1}+1)(\phi_{C,2}+1)}, \qquad (6.3.1)
\end{aligned}$$

where we used the ϕ notation of Sec. 6.2. The values of $\phi_{C,1}$ and $\phi_{C,2}$ can be obtained experimentally using the methods also described in Sec. 6.2. Then, plugging into Eq. 5.1.5 yields the predicted information:

$$I_{\text{tree}}(R_1;R_2|S) = -\frac{1}{2}\log_2\left(1-\rho^2\right) = \frac{1}{2}\log_2\left(1+\frac{\phi_{C,1}\phi_{C,2}}{1+\phi_{C,1}+\phi_{C,2}}\right). \quad (6.3.2)$$

6.4 Predicting *I*(*C*;*S*) for the Gaussian tree network

In a tree network, the common trunk from *S* to *C* sets a limit on the information about the signal that can be transmitted to the downstream branches, and this limit is given by *I*(*C*;*S*). Eq. 5.3.5 shows that for a Gaussian tree network, *I*(*C*;*S*) depends solely on the ratio $m_C^2\sigma_S^2/\sigma_{S\to C}^2$. By examining the definitions of $\phi_{S,1}$ and $\phi_{C,1}$ from Sec. 6.2 it can be easily seen that $m_C^2\sigma_S^2/\sigma_{S\to C}^2 = \phi_{S,1}/\phi_{C,1}$. Thus, the predicted value of *I*(*C*;*S*) is

$$I(C;S) = \frac{1}{2}\log_2\left(1+\frac{\phi_{S,1}}{\phi_{C,1}}\right). \quad (6.4.1)$$

I(*C*;*S*) can also be predicted from $\phi_{S,2}/\phi_{C,2}$ if a second response was measured (and so on for three or more responses), and the predicted values can be averaged together to yield a final prediction.

6.5 Predicting *I*(*R*$_1$,…,*R*$_n$;*S*) for the Gaussian tree network for an arbitrary number of identical branches

The mutual information for a tree network whose branches have identical levels of noise is given by Eq. 5.3.6

$$I(R_1,...,R_n;S) = \frac{1}{2}\log_2\left(1+\frac{n\sigma_S^2/\sigma_{C\to R}^2}{n\sigma_{S\to C}^2/\sigma_{C\to R}^2+1}\right), \quad (6.5.1)$$

where the values of m_C and m_i have been subsumed into σ_S^2 and $\sigma_{S\to C}^2$, respectively. This formula shows that the information essentially depends on just three parameters: *n*, $\sigma_S^2/\sigma_{C\to R}^2$, and $\sigma_{S\to C}^2/\sigma_{C\to R}^2$. Here, we show how to fit this equation to experimental data. To simplify the algebra, we will denote the noise ratios as $\phi_S \equiv \sigma_S^2/\sigma_{C\to R}^2$ and $\phi_C \equiv \sigma_{S\to C}^2/\sigma_{C\to R}^2$. Furthermore, we define θ_n to be a function of the mutual information resulting from *n* responses as:

$$\frac{n\phi_S}{n\phi_C+1} = 2^{2I(R_1,...,R_n;S)}-1 \equiv \theta_n. \quad (6.5.2)$$

Suppose that the mutual information has been experimentally measured for two different values of n (i.e., n_1 and n_2) and the ratio n_1/n_2 is also known. First, we will show how to extrapolate to $n \to \infty$. To do this we solve Eq. 6.5.2 for n, yielding

$$n = \frac{\theta_n}{\phi_S - \theta_n \phi_C}. \tag{6.5.3}$$

Then, by writing Eq. 6.5.3 for n_1 and n_2, dividing, and rearranging we obtain

$$\frac{n_1}{n_2} = \frac{\dfrac{\theta_{n_1}}{\phi_S - \theta_{n_1}\phi_C}}{\dfrac{\theta_{n_2}}{\phi_S - \theta_{n_2}\phi_C}} \Leftrightarrow \frac{\phi_S}{\phi_C} = \frac{\theta_{n_1}\theta_{n_2}(1-\frac{n_1}{n_2})}{\theta_{n_1} - \theta_{n_2}\frac{n_1}{n_2}}. \tag{6.5.4}$$

Thus, the ratio ϕ_S / ϕ_C depends only on experimentally accessible quantities. Examination of Eqs. 5.3.5 and 6.4.1 shows that this ratio allows us to directly compute the mutual information resulting from an infinite number of branches as:

$$n \to \infty \Rightarrow I(R_1,...,R_n;S) \to \frac{1}{2}\log_2\left(1+\frac{\phi_S}{\phi_C}\right). \tag{6.5.5}$$

Next, suppose that we wish to compute the mutual information for some other value of n (or, at least for some other value of n/n_2 if the exact value of n_2 is not known). Then, replacing n_1 with an arbitrary value $n > 0$ in Eq. 6.5.4 and solving for θ_n gives

$$\frac{\phi_S}{\phi_C} = \frac{\theta_n\theta_{n_2}(1-\frac{n}{n_2})}{\theta_n - \theta_{n_2}\frac{n}{n_2}} \Rightarrow \theta_n = \frac{\theta_{n_2}\frac{n}{n_2}\frac{\phi_S}{\phi_C}}{\theta_{n_2}(\frac{n}{n_2}-1)+\frac{\phi_S}{\phi_C}}, \tag{6.5.6}$$

which is a quantity consisting of all known values except n (or n/n_2). Thus, inverting the definition of θ_n gives the desired mutual information as a function of n (or n/n_2):

$$I(R_1,...,R_n;S) = \frac{1}{2}\log_2(1+\theta_n). \tag{6.5.7}$$

7. SUPPLEMENTARY FIGURES

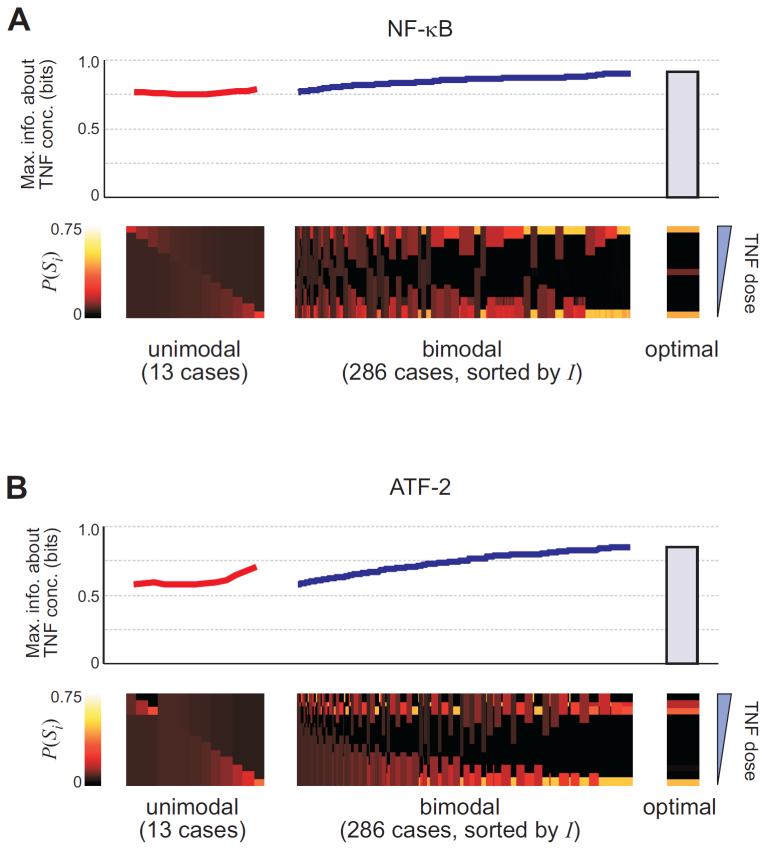

Fig. S1. Maximum mutual information about TNF concentration. (**A**) The top graph shows the maximum mutual information between TNF concentration and nuclear NF-κB concentration at 30 min. under a unimodal constraint (sorted in order of the 13 possible locations of the mode), bimodal constraint (testing all 286 possible locations of the two modes and the intervening minimum, sorted in increasing order of mutual information), and no constraint (optimal). The bottom heat maps show the signal distributions that yield the maximum mutual information under the various constraints. Each column in the heat map represents a signal distribution (a set of probabilities that sum to 1), each row corresponds to a specific signal value (TNF concentration), and the color indicates the probability associated with that signal value. The optimal value is approached by multiple bimodal distributions in which only very high and very low TNF concentrations are represented. (**B**) Same as panel A except the response analyzed is nuclear phospho-ATF-2.

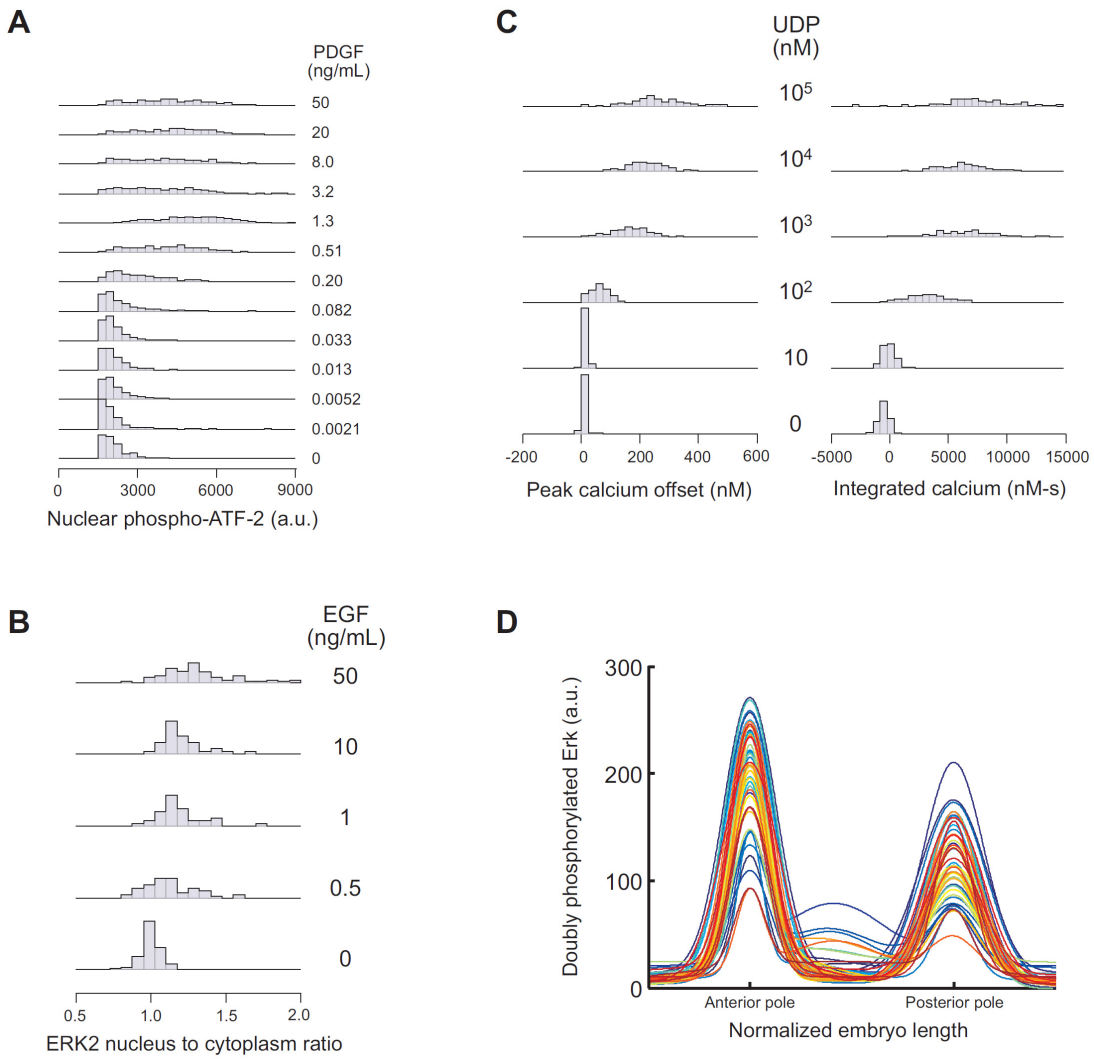

Fig. S2. Response distributions for various signaling systems. The data shown here were used to compute some of the channel capacity values reported in Table S1. (**A**) nuclear phospho-ATF-2 concentrations in mouse fibroblasts following 30 min. exposure to TNF at the indicated concentrations, as measured by immunofluorescence. (**B**) Fold-change in extracellular signal regulated kinase 2 (ERK2) nucleus to cytoplasm ratio in human lung cancer cells in response to 10 min. epidermal growth factor (EGF) exposure, as measured in single live cells (see SOM, Section 1.4). (**C**) Peak calcium concentration (left) and time-integrated calcium dynamics (right, integrated over 120 sec) in RAW264.7 macrophages following exposure to uridine diphosphate (UDP), a stimulus for the P2Y family of G protein-coupled receptors. Data was obtained courtesy of M. Simon (California Institute of Technology), see (*24*). (**D**) Concentrations of doubly phosphorylated Erk along the perimeter of wildtype *Drosophila melanogaster* embryos between nuclear cycles 10 and 14, as determined by immunofluorescence. Each curve is fitted to an individual embryo and normalized so that peak Erk activities occur at the anterior and posterior poles. Data was obtained courtesy of S. Shvartsman (Princeton), see (*25*).

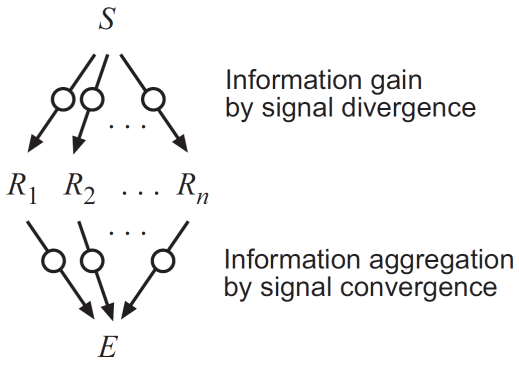

Fig. S3. Information flow through multiple communication channels that diverge then converge. Signaling through multiple communication channels to the responses R_1, R_2, …, R_n can increase the amount of information transduced about the input signal, S, as compared to the information transferred by an individual channel. This information can be aggregated through downstream convergence at a common effector, E.

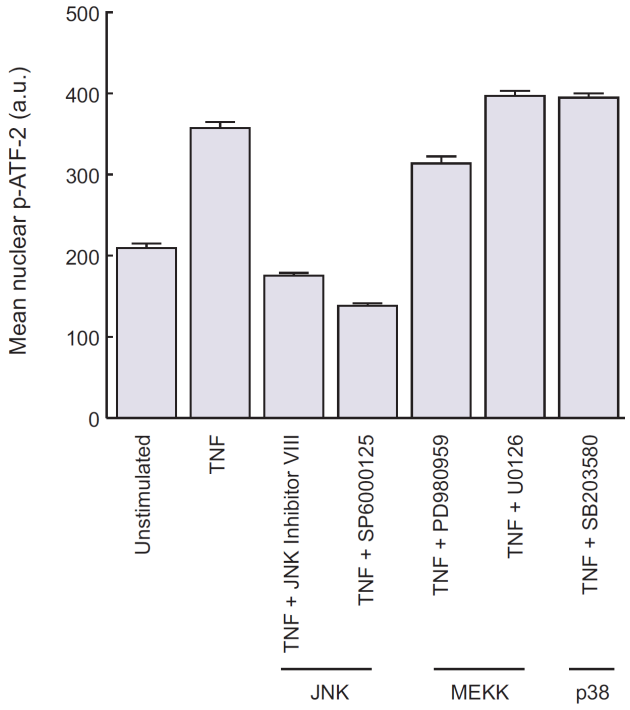

Fig. S4. Selective activation of ATF-2 by JNK in response to TNF. Plot shows the average nuclear phospho-ATF-2 concentration, as measured by immunofluorescence, of mouse fibroblasts incubated with the indicated JNK, mitogen activated protein kinase kinase kinase (MEKK), or p38 inhibitors for 1 hour prior to 30 min. stimulation with TNF. The JNK inhibitors, but not the MEKK and p38 inhibitors, were able to inhibit ATF-2 phosphorylation to levels at or below unstimulated cells, indicating that TNF-induced ATF-2 phosphorylation is mediated by JNK and not MEKK or p38.

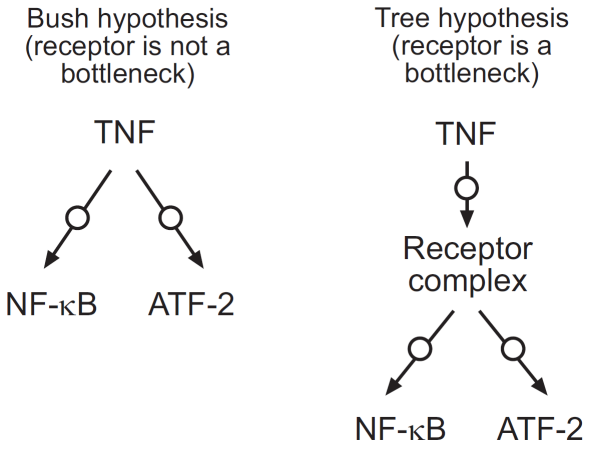

Fig. S5. Bush and tree representations of the TNF signaling network. Schematics of information flow through the TNF signaling network highlighting the experimentally testable hypotheses of whether the network lacks (bush model, left) or contains (tree model, right) an information bottleneck due to the steps of receptor complex activation common to multiple TNF signaling pathways.

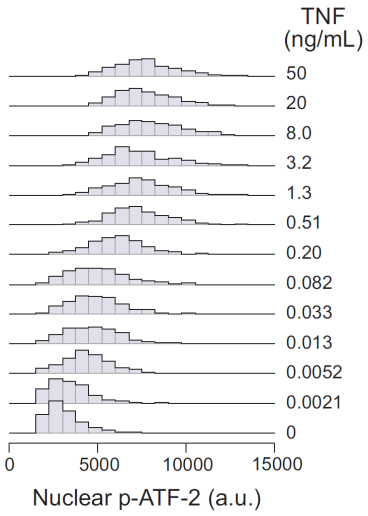

Fig. S6. Distribution of ATF-2 activity in response to TNF. Histograms showing the distribution of nuclear phospho-ATF-2 concentrations in mouse fibroblasts in response to 30 min. TNF exposure at the indicated concentrations.

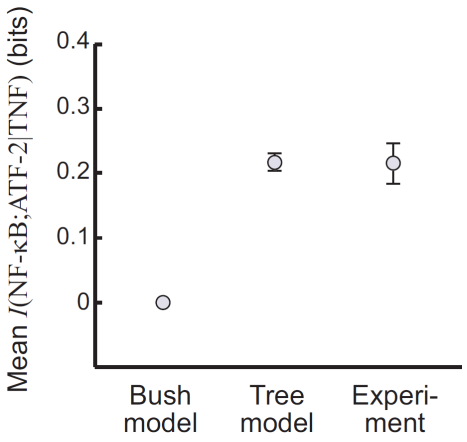

Fig. S7. Statistical dependence between NF-κB and ATF-2 responses to TNF. Plot shows the experimentally measured statistical dependence between the NF-κB and ATF-2 responses, as quantified by the mean value of I(NF-κB; ATF-2 | TNF) (see SOM, Sections 2.3 and 6.3), compared to values predicted by the bush and tree network models. The bush model predicts conditional independence between the responses and hence zero mutual information, but the tree model predicts conditional dependence resulting from the common trunk with mutual information of 0.22 ± 0.01 bits, which corresponds exactly with the experimentally observed value of 0.22 ± 0.03 bits. Conditional dependence between the responses may also arise from crosstalk between the pathways, but there is likely insufficient time for substantial crosstalk to occur following 30 min. TNF exposure.

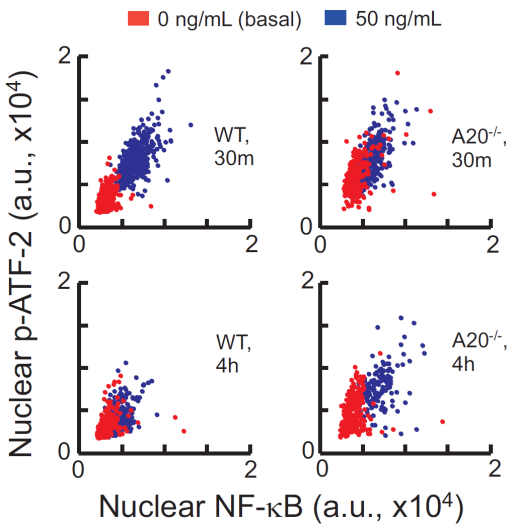

Fig S8. Responses to TNF with and without A20-mediated negative feedback. Plots show joint NF-κB and ATF-2 responses to TNF in wildtype and A20$^{-/-}$ cells at the indicated time points. Each datapoint represents a single cell. Only the responses to zero and saturating TNF concentrations are shown to clearly display observed changes in the dynamic range and noise.

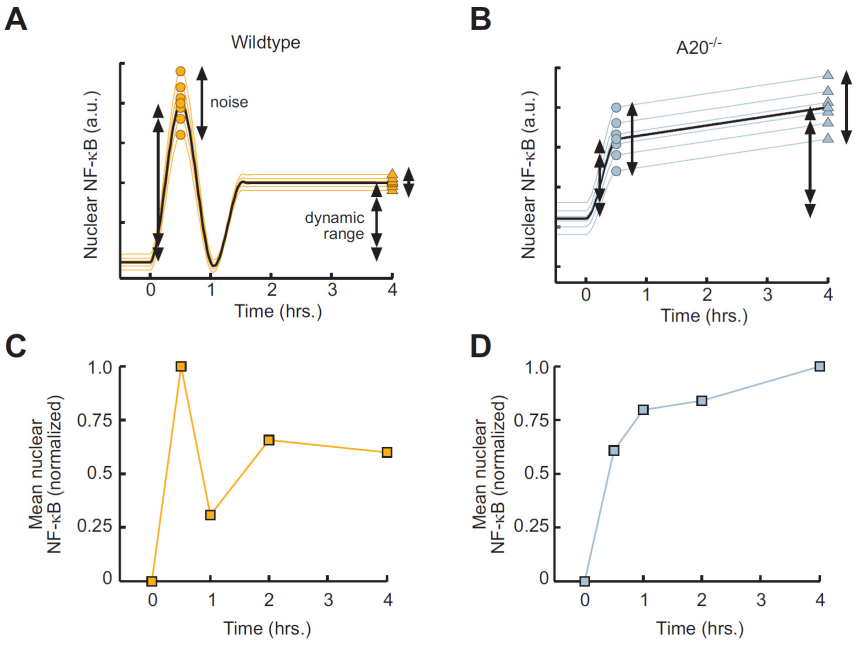

Fig. S9. Dynamics of NF-κB activity in wildtype versus A20$^{-/-}$ cells. (**A, B**) Plots reproduced from Fig. 3C and 3D depicting the dynamics of NF-κB responses in wildtype (panel A) and A20$^{-/-}$ (panel B) mouse fibroblasts exposed to saturating concentrations of TNF. Average dynamics (black) and the expected magnitudes of the dynamic range (double arrow) and noise (single arrow) are shown. (**C, D**) Time courses of the NF-κB response to 10 ng/mL TNF measured by immunocytochemistry, confirming that, on average, wildtype cells show biphasic dynamics consisting of an initial peak of activity lasting 1 hour, followed by a secondary steady phase lasting several hours (panel C), while A20$^{-/-}$ cells show a rapid increase in NF-κB activity in the first hour followed by a slower increase thereafter (panel D).

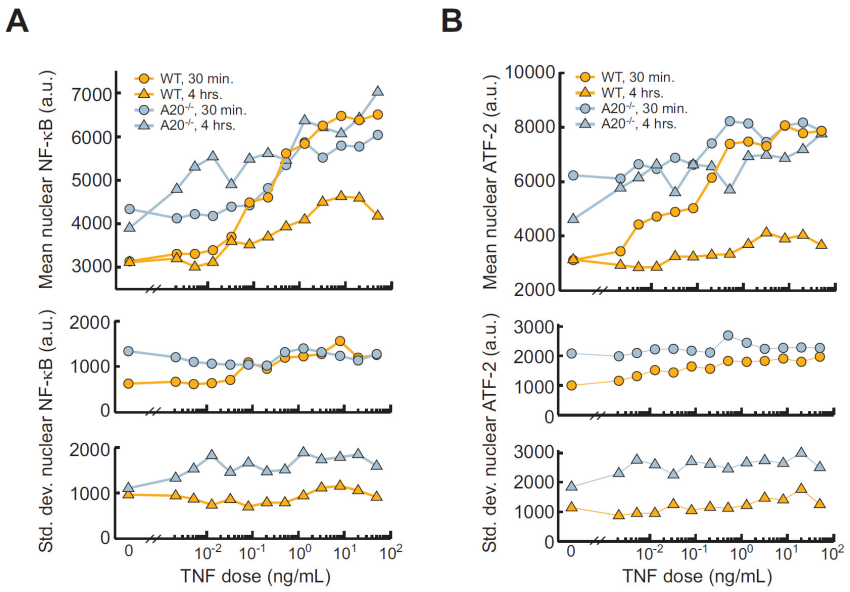

Fig. S10. Dynamic range and noise in responses to TNF. Dose response curves for mean nuclear NF-κB (**A**) and phospho-ATF-2 (**B**) concentration in response to TNF for the indicated duration in the indicated cells. The top plots demonstrate that, for either response, the dynamic range is greater at 30 min. but smaller at 4 hrs. in wildtype cells than in $A20^{-/-}$ cells. The bottom plots demonstrate that, for either response, the noise magnitude, measured as the standard deviation of the response, is greater (or at least no smaller) in $A20^{-/-}$ cells than in wildtype cells at all TNF concentrations and time points examined.

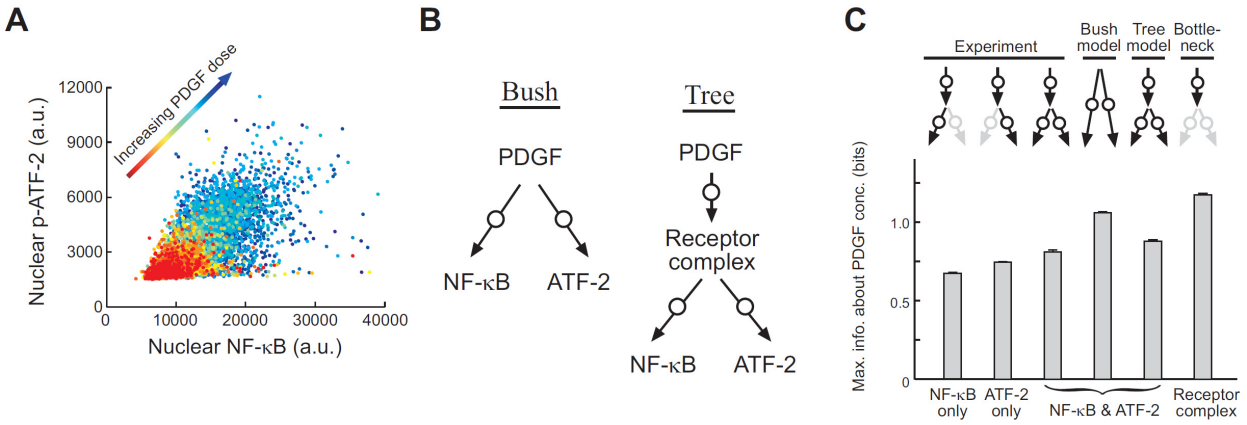

Fig. S11. The PDGF signaling network contains an information bottleneck. (**A**) Scatter plot showing nuclear NF-κB and ATF-2 responses to 30 min. stimulation of PDGF. Each datapoint represents a single cell, and each concentration of TNF examined is shown using a distinct color. (**B**) Schematics of information flow through the PDGF signaling network highlighting the experimentally testable hypotheses of whether the network lacks (bush model, left) or contains (tree model, right) an information bottleneck due to the steps of receptor complex activation common to multiple PDGF signaling pathways. (**C**) Comparison of bush and tree model predictions for the capacity of the PDGF network to experimental values. At 30 min., the NF-κB and ATF-2 pathways together capture more information about PDGF concentration than either pathway alone (bars 1-3), and the tree rather than bush model accurately predicts this increase (bars 3-5). The tree model further predicts a receptor level bottleneck of 1.18 ± 0.01 bits (bar 6).

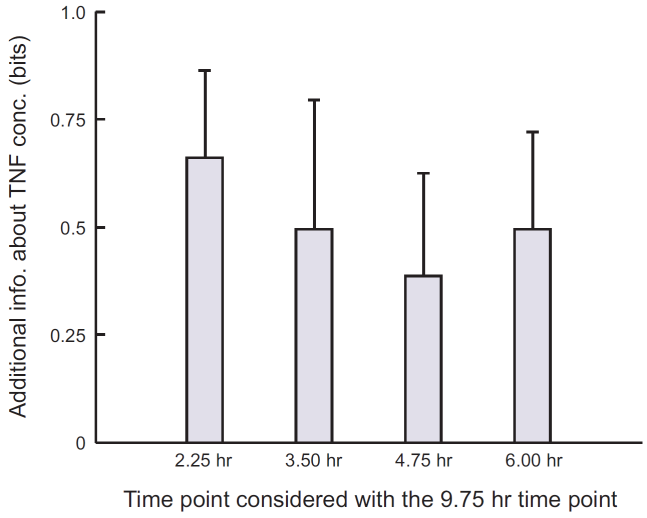

Fig. S12. Information increase attainable from temporal correlations in reporter gene expression. Plot showing the additional information that a cell can gain from the reporter gene expression level at 9.75 hrs TNF exposure and the expression level at an earlier time point (ranging from 2.25 to 6.00 hr), as measured by the mutual information between the expression levels at the two time points. The results indicate that a cell that can determine the reporter gene expression level at both an early and late time point can capture ~0.5 bit more information about TNF concentration than a cell that can only determine the expression level at the late time point.

8. SUPPLEMENTARY TABLES

Signal	Response	Maximum mutual information (bits)	Reference
TNF	NF-κB	0.92 ± 0.01	Figs. 1D, 2C
TNF	ATF-2	0.85 ± 0.02	Figs. S6, 2C
TNF	NF-κB and ATF-2	1.05 ± 0.02	Figs. 2C, 2D
PDGF	NF-κB	0.67 ± 0.01	Fig. S11A
PDGF	ATF-2	0.74 ± 0.01	Figs. S2A, S11A
PDGF	NF-κB and ATF-2	0.81 ± 0.02	Fig. S11A
EGF	Erk (fold-change)	0.60 ± 0.03	Fig. S2B, (*23*)
UDP	Peak Ca^{2+}	1.22 ± 0.03	Fig. S2C, (*24*)
UDP	Integrated Ca^{2+}	1.07 ± 0.02	Fig. S2C, (*24*)
Position	Doubly phosphorylated Erk	1.61 ± 0.05	Fig. S2D, (*25*)

Table S1. Experimentally measured channel capacity of various signaling pathways.

9. REFERENCES

1. J. G. Albeck, J. M. Burke, S. L. Spencer, D. A. Lauffenburger, P. K. Sorger, Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS Biol* **6**, 2831 (2008).
2. N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, Gene regulation at the single-cell level. *Science* **307**, 1962 (2005).
3. T. J. Perkins, P. S. Swain, Strategies for cellular decision-making. *Mol Syst Biol* **5**, 326 (2009).
4. W. J. Blake, K. A. M, C. R. Cantor, J. J. Collins, Noise in eukaryotic gene expression. *Nature* **422**, 633 (2003).
5. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell. *Science* **297**, 1183 (2002).
6. J. Paulsson, Summing up the noise in gene networks. *Nature* **427**, 415 (2004).
7. J. M. Pedraza, A. van Oudenaarden, Noise propagation in gene networks. *Science* **307**, 1965 (2005).
8. J. M. Raser, E. K. O'Shea, Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811 (2004).
9. T. M. Cover, J. A. Thomas, *Elements of information theory* (Wiley, New York, 1991), pp.
10. R. R. de Ruyter van Steveninck, G. D. Lewen, S. P. Strong, R. Koberle, W. Bialek, Reproducibility and variability in neural spike trains. *Science* **275**, 1805 (1997).
11. D. Fuller *et al.*, External and internal constraints on eukaryotic chemotaxis. *Proc Natl Acad Sci U S A* **107**, 9656 (2010).
12. E. Ziv, I. Nemenman, C. H. Wiggins, Optimal signal processing in small stochastic biochemical networks. *PLoS One* **2**, e1077 (2007).
13. G. Tkacik, C. G. Callan, Jr., W. Bialek, Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci U S A* **105**, 12265 (2008).
14. P. Mehta, S. Goyal, T. Long, B. L. Bassler, N. S. Wingreen, Information processing and signal integration in bacterial quorum sensing. *Mol Syst Biol* **5**, 325 (2009).
15. R. Cheong, A. Hoffmann, A. Levchenko, Understanding NF-kappaB signaling via mathematical modeling. *Mol Syst Biol* **4**, 192 (2008).
16. R. Cheong, C. J. Wang, A. Levchenko, High content cell screening in a microfluidic device. *Mol Cell Proteomics* **8**, 433 (2009).
17. S. L. Werner *et al.*, Encoding NF-kappaB temporal control in response to TNF: distinct roles for the negative regulators IkappaBalpha and A20. *Genes Dev* **22**, 2093 (2008).
18. S. Tay *et al.*, Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267 (2010).
19. L. Ashall *et al.*, Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription. *Science* **324**, 242 (2009).
20. D. E. Nelson *et al.*, Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science* **306**, 704 (2004).
21. R. Cheong *et al.*, Transient IkappaB kinase activity mediates temporal NF-kappaB dynamics in response to a wide range of tumor necrosis factor-alpha doses. *J Biol Chem* **281**, 2945 (2006).
22. A. Hoffmann, A. Levchenko, M. L. Scott, D. Baltimore, The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241 (2002).
23. C. Cohen-Saidon, A. A. Cohen, A. Sigal, Y. Liron, U. Alon, Dynamics and variability of ERK2 response to EGF in individual living cells. *Mol Cell* **36**, 885 (2009).
24. X. R. Bao, I. D. Fraser, E. A. Wall, S. R. Quake, M. I. Simon, Variability in G-protein-coupled signaling studied with microfluidic devices. *Biophys J* **99**, 2414 (2010).

25. M. Coppey, A. N. Boettiger, A. M. Berezhkovskii, S. Y. Shvartsman, Nuclear trapping shapes the terminal gradient in the Drosophila embryo. *Curr Biol* **18**, 915 (2008).
26. B. B. Averbeck, P. E. Latham, A. Pouget, Neural correlations, population coding and computation. *Nat Rev Neurosci* **7**, 358 (2006).
27. J. W. Pillow *et al.*, Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995 (2008).
28. E. Schneidman, W. Bialek, M. J. Berry II, Synergy, redundancy, and independence in population codes. *J Neurosci* **23**, 11539 (2003).
29. H. Wajant, K. Pfizenmaier, P. Scheurich, Tumor necrosis factor signaling. *Cell Death Differ* **10**, 45 (2003).
30. A. Becskei, L. Serrano, Engineering stability in gene networks by autoregulation. *Nature* **405**, 590 (2000).
31. I. Lestas, G. Vinnicombe, J. Paulsson, Fundamental limits on the suppression of molecular fluctuations. *Nature* **467**, 174 (2010).
32. R. C. Yu *et al.*, Negative feedback that improves information transmission in yeast signalling. *Nature* **456**, 755 (2008).
33. I. E. Wertz *et al.*, De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. *Nature* **430**, 694 (2004).
34. E. G. Lee *et al.*, Failure to regulate TNF-induced NF-kappaB and cell death responses in A20-deficient mice. *Science* **289**, 2350 (2000).
35. V. Shahrezaei, P. S. Swain, Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci U S A* **105**, 17256 (2008).
36. S. Krishna, M. H. Jensen, K. Sneppen, Minimal model of spiky oscillations in NF-kappaB signaling. *Proc Natl Acad Sci U S A* **103**, 10840 (2006).
37. S. Thierfelder, K. Ostermann, A. Gobel, G. Rodel, Vectors for glucose-dependent protein expression in Saccharomyces cerevisiae. *Appl Biochem Biotechnol* **163**, 954 (2011).
38. K. Francis, B. O. Palsson, Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc Natl Acad Sci U S A* **94**, 12258 (1997).
39. J. Parkin, B. Cohen, An overview of the immune system. *Lancet* **357**, 1777 (2001).
40. I. Nemenman, G. D. Lewen, W. Bialek, R. R. de Ruyter van Steveninck, Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS Comput Biol* **4**, e1000025 (2008).
41. S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, W. Bialek, Entropy and information in neural spike trains. *Phys Rev Lett* **80**, 197 (1998).
42. D. A. Freedman, J. Folkman, Maintenance of G1 checkpoint controls in telomerase-immortalized endothelial cells. *Cell Cycle* **3**, 811 (2004).
43. A. A. Cohen *et al.*, Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**, 1511 (2008).
44. L. Paninski, Estimation of entropy and mutual information. *Neural Comput* **15**, 1191 (2003).
45. I. Nemenman, F. Shafee, W. Bialek, in *Advances in Neural Information Processing Systems* T. G. Dietterich, S. Becker, Z. Ghahramani, Eds. (MIT Press, Cambridge, MA, 2002), vol. 14, pp. 95-100.
46. N. Slonim, G. S. Atwal, G. Tkacik, W. Bialek, Information-based clustering. *Proc Natl Acad Sci U S A* **102**, 18297 (2005).
47. A. A. Margolin *et al.*, ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7 (2006).
48. A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information. *Phys Rev E* **69**, 066138 (2004).

49. N. Slonim, G. S. Atwal, G. Tkacik, W. Bialek, Estimating mutual information and multi-information in large networks. *arXiv cs.IT/0502017* (2005).
50. E. K. P. Chong, S. H. Zak, *An introduction to optimization, 3rd ed.* (Wiley, Hoboken NJ, 2008), pp. 457-471.
51. G. Tkacik, C. G. Callan, Jr., W. Bialek, Information capacity of genetic regulatory elements. *Phys Rev E* **78**, 011910 (2008).
52. I. Nemenman, W. Bialek, Occam factors and model independent Bayesian learning of continuous distributions. *Phys Rev E* **65**, 026137 (2002).
53. S. Panzeri, R. Senatore, M. A. Montemurro, R. S. Petersen, Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* **98**, 1064 (2007).
54. A. A. Margolin, K. Wang, A. Califano, I. Nemenman, Multivariate dependence and genetic networks inference. *IET Syst Biol* **4**, 428 (2010).
55. T. Magesacher, P. Odling, J. Sayir, T. Nordstrom, paper presented at the IEEE International Symposium on Information Theory, Yokohama, Japan 2003.