ACCURATE INFORMATION TRANSMISSION THROUGH DYNAMIC BIOCHEMICAL SIGNALING NETWORKS

BERNIE PINKENZON-HOWARD

07.12.16
CONTENTS

• Common grounds – Useful Definitions
• Introduction
• Experiment
• Results
• Further Research
COMMON GROUNDS

- **Cellular Noise**
  - Intrinsic – creates differences between identically-controlled genes within a single cell
  - Extrinsic – affects genes equally inside a cell. Creates Differences between cells.

- **Ligands** - a substance that forms a complex with a biomolecule to serve a biological purpose

- **Effectors** – a molecule that binds to a protein and regulates its biological activity

- **Enhancer** - short DNA seq. that increase the likelihood to transcript a gene
COMMON GROUNDS

• Signaling pathways – molecule signaling process that triggers a function.
  • cell death, cell division etc.

• Extracellular signal–regulated kinases (ERK)
  • An enzyme that modifies other proteins.
  • Activates transcription factors.
  • Eventually causing gene expression
COMMON GROUNDS

• **Nuclear factor kappa-B (NF-κB)**
  - Discovered as a nuclear factor that bind to an enhancer in a certain part in B Cells.
  - Proteins which harbor this DNA binding activity are expressed in nearly all cell types.
  - They regulate many target genes with a whole variety of functions.

• **Ca2+ signaling**
  - Binding it triggers changes in protein shape and charge
  - Used to change a protein's location in the cell from the cytoplasm to a membrane surface
• Signaling networks are designed to reliably transmit info about the extracellular environment, and accordingly change its state
• Signal transaction, i.e. signals passing along a signalling pathway, can be degraded by noise.
• Previous experiments suggested that most of the information lost traveling through the network regards the concentration of ligands outside the cell
EARLIER EXPERIMENTS

• We can estimate the maximal information transmission capacity in a noisy signaling network.

• Signaling networks analysis have been based before on scalar measurements performed at a single time point. That being contrary to the fact information on activating ligands are represented as a multi-variable vector, which is a dynamic signal.

• The vector contains the cell’s response at multiple time points.
HYPOTHESIS AND GOALS

• Temporal signal modulation, i.e. dynamics, can be used to reduce loss of such information

• Our main goal is to overcome, and maybe eliminate noise-induced information loss in the signaling pathways mentioned using dynamics

• Live measurements were performed on all three pathways mentioned earlier, i.e. ERK, Ca$^{2+}$, NF$\kappa$B, with among different levels and types of ligands.
EXPERIMENT

• Information transmission capacity of a single dynamic signaling network is calculated and referred as the maximal possible mutual information between a measured response and the activating ligand concentration.
PHASE 1

- Mutual info between Dynamic response and ligand concentration
- The algorithm uses continuous multidimensional response data and a k–nearest-neighbor approach to estimate the conditional probability density for each cell’s response.
PHASE 1 - ALGORITHM

• Signal S is defined by m discrete levels of extracellular ligand concentration ($S = [s_1, s_2, \ldots, s_m]$).

• For each input signal $s_i$ we have $n_i$ output protein trajectories ($R_i = [r_{i1}, r_{i2}, \ldots, r_{in_i}]$).

• We have $m$ trajectories
PHASE 1 - ALGORITHM

• The estimation to the information input (S) and output (R) is denoted $I(R; S) = H(R) - H(R|S)$ where the H’s are Shannon entropies.

• If we denote $q_i = P(S = s_i)$, $C(R; S) = \max_Q \{I(R; S)|Q \in \{q_1 ... q_m\}\}$ is the maximum information transfer.

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

$$H_{\text{diff}}(R|S) = - \sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} \log_2 \left( \frac{k}{n_i V_d(z(r_{ij}|R_i))^d} \right)$$
Fig. S7
General scheme for estimation of information transmission based on experimentally obtained conditional responses (R) to scalar input levels (S).
Information transmission capacity calculated from static response vs multivariate dynamic responses as a function of the dimension of the multivariate vector.
EXPERIMENT

• It was discovered that dynamic response had significantly higher information transmission capacity than several scalar responses

• Why?
  • Let's find out!
• SNR – Signal to Noise ratio = $\frac{\sigma_r^2}{\sigma_n^2}$

• As mentioned before - there are two kinds of noises – extrinsic and intrinsic

• IER – Intrinsic noise to Extrinsic noise ratio
PHASE 2.1 - SIGNAL-TO-NOISE RATIO (SNR) IN SINGLE-CELL

• Given a scalar signal $S$ and a scalar response $R$, the signaling pathway is described as a joint distribution $P(R,S)$

• We will denote $\langle \bar{R} \rangle$ as the avg response from multiple cells driven by the same input $S$. We define the signal magnitude as the variance of $R$ around its mean value $\langle \bar{R} \rangle$ over all possible $S$.

$$\sigma^2 = \int dS P(S) \left[ \int dR RP(R|S) \right]^2 - \left[ \int \int dRdS RP(R,S) \right]^2$$
PHASE 2.1 - SIGNAL-TO-NOISE RATIO (SNR) IN SINGLE-CELL

• The noise magnitude is defined as the average width of the distribution of cellular responses for a given input $S$, averaged over $S$

\[ \sigma_n^2 = \int dS \, P(S) \left[ \int dR \, R^2 \, P(R|S) - \left( \int dR \, R \, P(R|S) \right)^2 \right] \]
PHASE 2.2 - INTRINSIC NOISE AND EXTRINSIC VARIABILITY

• Response within a cell is the sum of two parts
  
  • \( G(S, P) \) – \( S \) is the input signal, \( P \) is a \( k \)-parameter vector of extrinsic noise which is unknown and vary from cell to cell, same function in each cell population
  
  • \( \zeta \) – Gaussian intrinsic noise vector

• IER will be accordingly \( \frac{\sigma^2_{\zeta}}{\sigma^2_{\nu}} \)

• We linearize the systems response function in order to prevent deviation

\[
\begin{align*}
  r &= \frac{\partial g}{\partial S} s + \sum_{k=1}^{K} \frac{\partial g}{\partial P_k} p_k + \xi \\

  SNR & \equiv \frac{\sigma^2_r}{\sigma^2_n} = \frac{\left( \frac{\partial g}{\partial S} \right)^2 \sigma^2_s}{\sum_{k=1}^{K} \left( \frac{\partial g}{\partial P_k} \right)^2 \sigma^2_{p_k} + \sigma^2_{\xi}}
\end{align*}
\]
We want to compute the mutual information between an input $S$ and a single output measurement $R$.

For Gaussian distribution values of $R$ and $S$ this general expression reduces to

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

Remembering $\sigma_R^2 = \sigma_r^2 + \sigma_n^2$ and $\sigma_{R|S}^2 = \sigma_n^2$ we get

$$I(R; S) = \frac{1}{2} \log_2 \left( 1 + \frac{\sigma_r^2}{\sigma_n^2} \right) = \frac{1}{2} \log_2 (1 + SNR).$$
We will demonstrate that multi-dimensional output can reduce both types of uncertainty and increase the mutual information for the same SNR.

A vector signaling pathway as an input-output system that has a single scalar input $S$ and generally multi-dimensional vector of outputs $\{R_1, \ldots, R_n\}$

$R = G(S, P) + \zeta$, when $\zeta = \{\zeta_1, \ldots, \zeta_n\}$ – vector of uncorrelated intrinsic noise, and $P \in \mathbb{R}^K$
These outputs may represent:

- Values of a single variable taken at different time points
- Different physical variables (e.g., concentrations of multiple transcription factors)
- Both

Case 1: let’s ignore the intrinsic noise completely.

- \( R_i = g_i(S|P) \mid i = i, \ldots, N \)
- What happens for \( N \geq K + 1 \)?
- What happens when we iterate?
PHASE 2.4 - INFORMATION CAPACITY OF VECTOR SIGNALING

• Solving the system may prove difficult or impossible in cases when the structure of the pathway is poorly understood.

• But! using multiple time point measurements can increase the amount of information about the input signal in complex pathways.
PHASE 2.4 - INFORMATION CAPACITY OF VECTOR SIGNALING

• Case 2: extrinsic and intrinsic noise are present

• We need to linearize the non-linear system $G$.

• We have two kinds of noises, so we’re creating two Jacobian matrices, so the output will be denoted as $r = J_0 s + \hat{f} \cdot p + \zeta$, when

  • $r = R - \langle R \rangle, s = S - \langle S \rangle, p = P - \langle P \rangle$; (×) means deviations of these quantities from their mean values

  • $J_0 = \frac{\partial g}{\partial s}$ taken at $(S), \hat{f} = \frac{\partial g_i}{\partial p_k}$
• $s, p_k, \zeta_i$ are all drawn from independent Gaussian distributions, so we can compute the mutual information between the signal $s$ and the response $r$:

$$(R_1 \ldots R_k \ | S) = \frac{1}{2} \log_2 \left( \frac{|\Sigma_R|}{|\Sigma_{R}\mid_S|} \right) \text{ when:}$$

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

$$(\Sigma_{R\mid S})_{ij} = \sum_k J_{ik}J_{jk}\sigma_{p_k}^2 + \delta_{ij}\sigma_{\zeta_i}^2$$
Thus, the mutual information between the scalar signal \( s \) and the vector output \( r \) is denoted \( I \). After some mathematics and assumptions:

- The intrinsic noise amplitude is the same in all measurements.
- When the number of measurements is greater than the number of parameters \( (N > K) \), and counting the greater contribution parameters.

Finally, we get

\[
I \approx \frac{1}{2} \log_2 \left( 1 + \frac{\sigma_s^2}{\sigma_n^2} \sum_{i=1}^{N-K} d_i^2 \right)
\]

when \( d_i = J_0 \cdot q_i \). Meaning:

\[
I \approx \frac{1}{2} \log_2 \left( 1 + k \cdot (N - K)SNR(1 + IER^{-1}) \right), \quad K = O(1)
\]
PHASE 2 - INFORMATION CAPACITY OF VECTOR SIGNALING

• Observeing $I \approx \frac{1}{2} \log_2 (1 + k \cdot (N - K)SNR(1 + IER^{-1}))$ we notice that

• For $IER \ll 1$, elimination of extrinsic noise!

• For large $SNR$ values $I \approx -\frac{1}{2} \log_2 (N - K)$ bits.

• Additional measurements beyond $N=K+1$ suppress the remaining intrinsic noise and add approximately $\frac{1}{2} \log_2 (N - K)$ bits to the mutual information!
THEORETICAL CONCLUSIONS

• This theory gives a probable explanation for signal segregation caused by noise, both extrinsic and intrinsic

• Intrinsic noise adds to uncertainty in all dimensions (i.e., time points), independently from one another

• Extrinsic noise produces fluctuations that are diminished by the signalling network. These fluctuations have impact on one another
THEORETICAL CONCLUSIONS

• We also infer that intrinsic and extrinsic noise sources have different effects on the information transmission capacity of multivariate responses:

• In case of pure intrinsic noise - additional measurements increase the information deducted by ensemble averaging

• In case of pure extrinsic noise – if a number of dynamical measurement are accomplished, we can deduct information on the state the cell is in, and gain information on the activating ligand concentration
ASSESSMENT

• In order to assess the theory, simulations of ERK responses were made.
• ERK activity trajectories were simulated in response to an increasing number of ligand concentrations with MEK (Mitogen activated protein kinase kinase) added as extrinsic noise.
(A) gain in mutual information from overcoming intrinsic (cyan) and extrinsic (magenta) noise sources

(B) Information transmission capacity of dynamic (orange) and static (purple) responses calculated using trajectories from the model of ERK with only the extrinsic noise
ASSESSMENT

• The analytical prediction was correct!
  • The univariate response, based on maximal ERK dynamics, had limited information transmission capacity.
  • The dynamic multivariate response transmitted complete information about ligand concentration
- The trajectories of two populations of simulated responses of ERK activity to two input concentrations of epidermal growth factor (EGF) appear overlapping (C).
- In fact, they are completely separable, as shown in (D).
- This demonstrates how the extrinsic variability of a single parameter can be completely eliminated with measurements from only two time points.
PHASE 3

• The responses accuracy depends on SNR, noise and signal properties

• The theory predicts different relationship between mutual information and SNR for three different types of responses:
  • Scalar responses that do not distinguish between intrinsic and extrinsic noise
  • Multivariate responses without any dynamic component that can only reduce intrinsic noise
  • Dynamic responses mitigate both noises
PHASE 3

- Different SNR experiments were performed in the ERK network
  - Inhibition of the MEK using an inhibitor (U0126)
- At each U0126 level, mutual information and the SNR from single-cell responses were calculated
• Markers represent M.I. to SNR: dynamic (dot) and static (cross)
• Lines represent the theoretical predictions
  • Static scalar – red
  • Multivariate response without dynamics - blue
  • Dynamic response – orange
CONCLUSIONS

• For the scalar response, the theory is in very good agreement with the experimental measurements.

• The multivariate response requires knowledge of the IER, which has been calculated in two ways:
  • Quantifying fluctuations in the later portion of the response time series of the ERK data
    • required IER values that are two to four orders of magnitude higher than experimentally estimated IER values
  • using data for repeated measurements of single-cell responses
    • the theory is in very good agreement with the experimental measurements.
CONCLUSIONS

• Signaling dynamics allow biochemical networks to maximize the information transmission capacity of signaling networks

• The theory and observations presented focus on the information transmission capacity
  • Can be extended to multiple signalling molecules activated by one or more ligands
FUTURE RESEARCHES?

• Bioengineer cells that can store increasing quantity of transmitted information
• Amplify substance influence while decreasing dosage
• Deliver bogus information in order to fight tumors by causing cell suicide
• Pharmaceutical research regarding effectiveness of medicine combination as a function of time
• Double-Sided Channel creation for receiving intracellular information
Based on “Accurate information transmission through dynamic biochemical signaling networks”, Jangir Selimkhanov, Brooks Taylor, Jason Yao, Anna Pilko, John Albeck, Alexander Hoffmann, Lev Tsimring, Roy Wollman

Noise in Gene Expression: Origins, Consequences, and Control, Jonathan M. Raser and Erin K. O’Shea

The complexity of NF-κB signaling in inflammation and cancer, Bastian Hoesel and Johannes A Schmid

The MAPK (ERK) Pathway, Jack McCain

Calcium Signaling, David E. Clapham