The ability to sense the environment is heterogeneously distributed in cell populations

Andrew Goetz[1*], Hoda Akl[2*], and Purushottam Dixit[1,2,3,4]

[1]Department of Biomedical Engineering, Yale University, New Haven CT
[2]Department of Physics, University of Florida, Gainesville FL
[3]Systems Biology Institute, Yale University, West Haven, CT
[4]Corresponding author: purushottam.dixit@yale.edu
[*]Contributed equally

Abstract

Channel capacity of signaling networks quantifies their fidelity in sensing extracellular inputs. Low estimates of channel capacities for several mammalian signaling networks suggest that cells can barely detect the presence/absence of environmental signals. However, given the extensive heterogeneity and temporal stability of cell state variables, we hypothesize that the sensing ability itself may depend on the state of the cells. In this work, we present an information theoretic framework to quantify the distribution of sensing abilities from single cell data. Using data on two mammalian pathways, we show that sensing abilities are widely distributed in the population and most cells achieve better resolution of inputs compared to an "*average cell*". We verify these predictions using live cell imaging data on the IGFR/FoxO pathway. Importantly, we identify cell state variables that correlate with cells' sensing abilities. This information theoretic framework will significantly improve our understanding of how cells sense in their environment.

Introduction

In cell populations, there is a significant overlap in responses to environmental stimuli of differing strengths. This raises a fundamental question[1]: do signaling networks in cells relay accurate information about their environment to take appropriate action? And if not, where along the signal transduction pathway is the information lost[2, 3]? Mutual information (MI) quantifies the information content in an intracellular output about extracellular inputs. For an input u (e.g., concentration of a ligand) and an output x (e.g., intracellular species concentration[4-6] or a cellular phenotype[7, 8]), the MI is defined as[9]

$$I = \sum_{x,u} p(x|u)p(u) \log_2 \frac{p(x|u)}{\sum_{u'} p(x|u')p(u')}. \quad (1)$$

Experimental single cell methods such as flow cytometry[10], immunofluorescence[10], mass spectrometry[11], or live cell imaging[12] allow us to estimate response histograms $p(x|u)$ across several inputs. Using these distributions, we can estimate the maximum of the MI (the channel capacity, CC) by optimizing Eq. 1 over input distributions $p(u)$. The CC quantifies fidelity of signal transduction[1, 2]. For example, a CC of 1 bit implies that that the cells can *at best* distinguish between two input levels (e.g. presence versus absence), with higher CCs indicating that cells can resolve multiple input states. Importantly, CC can be used to identify bottlenecks in signaling[2, 13].

The CC has been estimated for input-output relationships in several mammalian signaling networks[4-6, 14-16]. When the output is defined as levels of a single protein a fixed time, the CC was found to be surprisingly low, $\sim 1 - 1.5$ bits. These estimates have been improved by considering multidimensional outputs[5] or time varying inputs[17]. While these modifications led to somewhat higher CC estimates, the overall conclusion that cells know little about their environment remains well established. In contrast, significantly higher CC estimates have been found when the output at the level of cell population averages are considered[6], suggesting that the only way to overcome low sensing fidelity is population average response.

These previous calculations estimated one channel capacity for all cells in a population, implicitly assuming that individual cells have similar sensing abilities. However, we now know that cells exhibit extensive heterogeneity in cell state variables[18, 19] such as abundances of key signaling proteins[20, 21], mRNA abundances[22], and chromatin conformation[23] and accessibility[24]. Notably, the time scale of fluctuations in these variables can be significantly slower than relevant signaling time scales[25, 26], sometimes extending across multiple generations[27]. Heterogeneity in cell state variables leads to a

heterogeneity in response to extracellular cues, including chemotherapeutic drugs[25, 26, 28], mitogens[29], hormones[30], chemotactic signals[7, 27, 31, 32], and other electrical and chemical stimuli[33, 34]. Therefore, we hypothesize that that the ability to sense the environment varies from cell-to-cell in populations, in a cell state dependent manner.

There is no conceptual framework to estimate the variation in sensing abilities in cell populations and its dependence on cell state variables. To that end, we introduce an information theoretic quantity CeeMI: **Ce**ll sta**te** dependent **M**utual **I**nformation. We show that using typically collected single cell data and computational models of signaling networks, we can estimate the distribution $p_{\mathrm{CeeMI}}(I)$ of single cell signaling fidelities (single cell mutual information values). We also show that we can identify cell state variables that make some cells better and others worse at sensing their environment.

Using an illustrative example, we show that in heterogeneous cell populations, estimates of mutual information that average over cell states can be significantly lower than the mutual information of signaling networks in typical cells. Next, using previously collected experimental data, we estimate $p_{\mathrm{CeeMI}}(I)$ for two important mammalian signaling pathways[35]; the Epidermal growth factor (EGF)/EGF receptor pathway and the Insulin like growth factor (IGF)/Forkhead Box protein O (FoxO) pathway, we show that while the cell state agnostic CC estimates for both pathways are ~ 1 bit, most individual cells are predicted to be significantly better at resolving different inputs. Using live cell imaging data for the IGF/FoxO pathway, we show that our estimate of variability in sensing abilities matches closely with experimental estimates. Importantly, for this pathway, we also verify our prediction that specific cell state variables dictate cells' sensing abilities. Finally, using a simple receptor/ligand model, we show how the time scales cell state dynamics affects cells' individuality and therefore sensing abilities. We believe that CeeMI will be an important tool to explore variability in cellular sensing abilities and in identifying what makes some cells better and others worse in sensing their environment.

Results

Conditional mutual information models single cells as cell state dependent channels

Consider a cell population where cells are characterized by state variables $\boldsymbol{\theta}$. These include abundances of signaling proteins and enzymes, epigenetic states, etc. We assume that cell states are temporally stable, that is, $\boldsymbol{\theta}$ remains constant over a time scale that is longer than typical fluctuations in

environmental inputs. Later, using a toy simple model, we explicitly study the effect of cell state dynamics on cells' ability to sense their environment.

We assume that cell state variables are distributed in the population according to a distribution $p(\boldsymbol{\theta})$. If x denotes an output of choice (e.g., phosphorylation levels of one or more protein(s) at one or more time point(s)) and u denotes the input (e.g., ligand concentration), the experimentally measured response distribution $p(x|u)$ can be decomposed as:

$$p(x|u) = \int p(x|u, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{2}$$

where $p(x|u, \boldsymbol{\theta})$ is the distribution of the output x conditioned on the input u *and* cell state variables $\boldsymbol{\theta}$. We note that in most cases, the same cell cannot be probed multiple times. Consequently, $p(x|u, \boldsymbol{\theta})$ may not be experimentally accessible. However, it is conceptually well defined and mathematically accessible when interactions amongst molecules are specified[36].

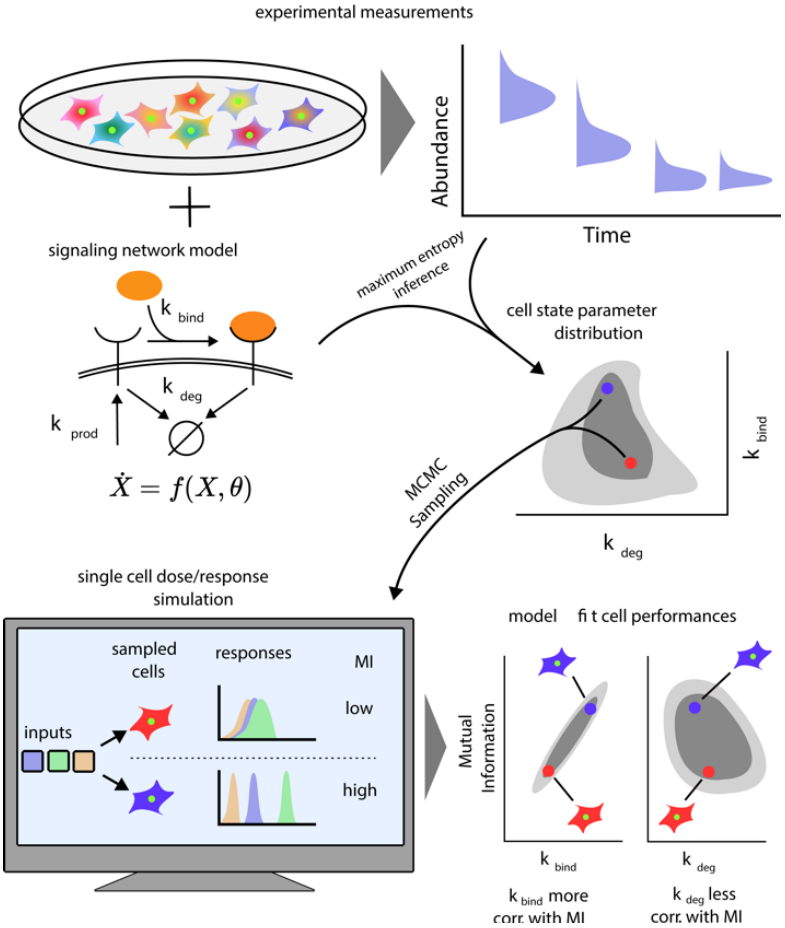

Figure 1. A schematic of our computational approach. (top) single cell data across different input conditions and time points are integrated with a stochastic model of a signaling network using a previous developed maximum entropy approach leading to a distribution over signaling network parameters $p(\boldsymbol{\theta})$ (middle). (bottom) In silico cells are generated using the inferred parameter distribution and cell-state specific mutual information $I(\boldsymbol{\theta})$ and

population distribution of cell performances $p_{\mathrm{CeeMI}}(I)$ is estimated. The model also evaluates the correlation between cells' performance and biochemical parameters.

Using $p(x|u,\boldsymbol{\theta})$, we can define the cell state dependent mutual information $I(\boldsymbol{\theta})$ for a fixed input distribution $p(u)$ analogously to Eq. 1:

$$I(\boldsymbol{\theta}) = \sum_{x,u} p(x|u,\boldsymbol{\theta})p(u)\log_2 \frac{p(x|u,\boldsymbol{\theta})}{\sum_{u'} p(x|u',\boldsymbol{\theta})p(u')}. \tag{3}$$

$I(\boldsymbol{\theta})$ quantifies individual cells' ability to sense their environment as a function of the cell state parameters $\boldsymbol{\theta}$. The distribution $p_{\mathrm{CeeMI}}(I)$ of single cell sensing abilities is

$$p_{\mathrm{CeeMI}}(I) = \int p(\boldsymbol{\theta})\,\delta\big(I - I(\boldsymbol{\theta})\big)d\boldsymbol{\theta} \tag{4}$$

where $\delta(\cdot)$ is the Dirac delta function. We can also compute the joint distribution between the single cell mutual information and any cell state variable of interest χ (e.g., abundance of cell surface receptors):

$$p_{\mathrm{CeeMI}}(I,\chi) = \int p(\boldsymbol{\theta})\delta\big(I - I(\boldsymbol{\theta})\big)\delta(\chi - \chi(\boldsymbol{\theta}))d\boldsymbol{\theta} \tag{5}$$

where $\chi(\boldsymbol{\theta})$ is the value of the biochemical parameter when cell state variables are fixed at $\boldsymbol{\theta}$. The distributions in Eq. 5 quantify the interdependency between a cell's signaling fidelity $I(\boldsymbol{\theta})$ and cell specific biochemical parameters $\chi(\boldsymbol{\theta})$. As we will see below, the distributions in Eq. 4 and Eq. 5 can be experimentally verified when appropriate measurements are available. Finally, we define the population average of the cell state dependent mutual information:

$$I_{\mathrm{Cee}} = \int I(\boldsymbol{\theta})\,p(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{6}$$

In information theory, I_{Cee} is known as the conditional mutual information[9] between the input u and the output x conditioned on $\boldsymbol{\theta}$. If cell state variables remain constant and if distribution over cell states is independent of the input distribution, it can be shown that $I_{\mathrm{Cee}} \geq I$ (SI Section 1). It then follows that at least some cells in the population have better sensing abilities compared to the cell state agnostic mutual information (Eq. 1).

Since I_{Cee} depends on the input distribution $p(u)$, we can find an optimal input distribution $p(u)$ that maximizes I_{Cee} (SI Section 1). Going forward, unless an input distribution is specified, the distributions $p_{\mathrm{CeeMI}}(I)$ and $p_{\mathrm{CeeMI}}(I,\chi)$ are discussed in the context of this optimal input distribution.

Maximum entropy inference can estimate $p_{\mathrm{CeeMI}}(I)$

Estimation of $p_{\mathrm{CeeMI}}(I)$ requires decomposing the experimentally observed response $p(x|u)$ into cell specific output distributions $p(x|u,\boldsymbol{\theta})$ and the distribution of cell state variables $p(\boldsymbol{\theta})$ (Eq. 3 and 4). This problem is difficult to solve given that neither $p(x|u,\boldsymbol{\theta})$ nor $p(\boldsymbol{\theta})$ are accessible in typical experiments. However, for many signaling networks, stochastic biochemical models can be constructed to obtain the cell specific output distribution $p(x|u,\boldsymbol{\theta})$. Here, $\boldsymbol{\theta}$ comprise protein concentrations and effective rates of enzymatic reactions and serve as a proxy for cell state variables. Given the experimentally measured cell state averaged response $p(x|u)$ and the model predicted cell specific output distribution $p(x|u,\boldsymbol{\theta})$, we need a computational method to infer $p(\boldsymbol{\theta})$ (see Eq. 2). The problem of inference of parameter heterogeneity is a non-trivial inverse problem[37]. While there are several proposed methods to solve this problem (reviewed in[38]), most cannot efficiently infer $p(\boldsymbol{\theta})$ for signaling networks with even a moderate ($n \sim 10$) number of parameters. Here, we use our previously develop maximum entropy-based approach to infer $p(\boldsymbol{\theta})$[37]. This way, we can estimate $p_{\mathrm{CeeMI}}(I)$ using experimentally obtained cell state agnostic response $p(x|u)$ and a stochastic biochemical model $p(x|u,\boldsymbol{\theta})$ of the underlying signaling network.

An "average cell" can discern much less than a typical cell about the environment

To illustrate the effect of heterogeneity of cell state variables on the cell state agnostic estimate of mutual information (which we call I_{CSA} from now onwards), we consider a simple stochastic biochemical network of a receptor-ligand system. Details of the model and the calculations presented below can be found in SI Section 2. Briefly, extracellular ligand L binds to cell surface receptors R. Steady state levels of the ligand bound receptor B is considered the output. The signaling network is parametrized by several cell state variables $\boldsymbol{\theta}$ such as receptor levels, rates of binding/unbinding, etc. For simplicity, we assume that only one variable, steady state receptor level R_0 in the absence of the ligand, varies in the population. Calculations with variability in other parameters are presented in SI Section 2.

In this population, cells' response $p(B|L,R_0)$ is distributed as a Poisson distribution whose mean is proportional to the cell state variable R_0 (SI Section 2). That is, when all other parameters are fixed, a higher R_0 leads to lower noise (coefficient of variation). To calculate cell state dependent mutual information (Eq. 3), we assume that $p(L)$ is a gamma distribution. As expected, $I(R_0)$ (Eq. 3) between the output B and the input L increases monotonically with R_0 (inset in Fig. 2A). Moreover, given that R_0 varies in the population (also assumed to be a gamma distribution), the sensing ability varies in the population as well (Fig. 2A). Notably, the average I_{Cee} of $I(R_0)$ remains relatively robust to variation in

. At the same time, the traditional estimate , which is the mutual information between the input and the cell state agnostic population response (response of the "average" cell, Eq. 1 and Eq. 2), decreases as the population heterogeneity in increases. Importantly, is significantly lower than the sensing ability of most cells in the population (Fig. 2A). This is because the overlap in the population response distributions is significantly larger than that in single cell response distributions (Fig. 2B). This simple example illustrates that the traditional mutual information estimates may severely underestimate cells' ability to resolve inputs, especially when cell state variables are heterogeneously distributed. Moreover, it is crucial to explicitly account for heterogeneity in cell state variables when estimating fidelity of cellular communication channels.

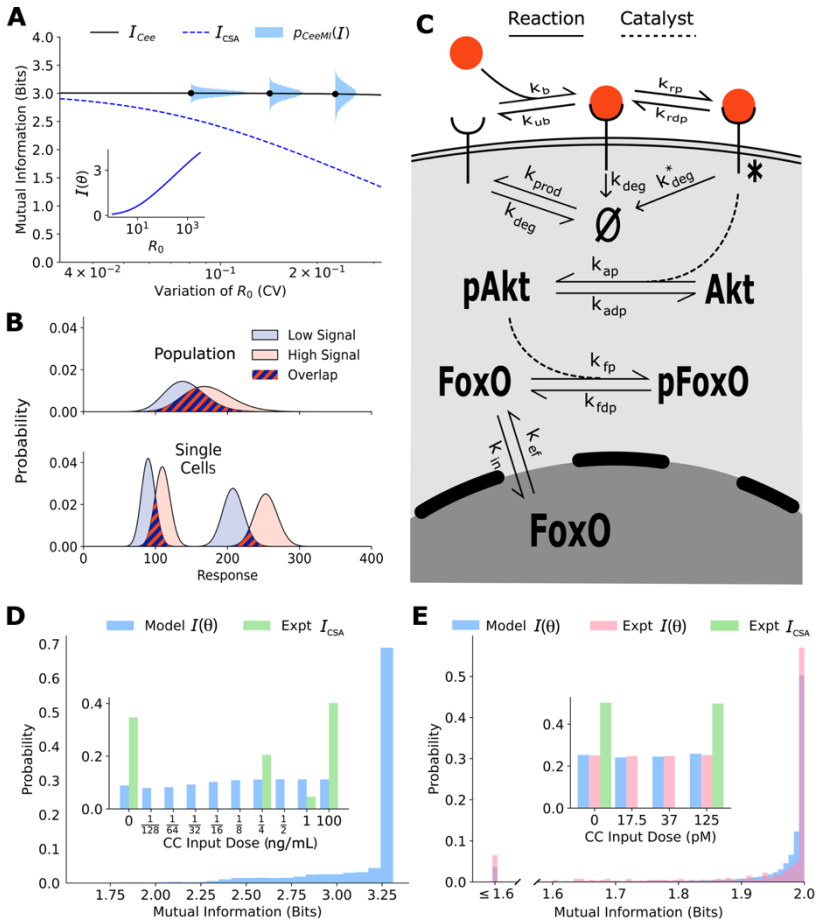

Figure 2. A. The distribution of single cell sensing abilities (horizontal blue histograms) and its average plotted as a function of the coefficient of variation of the distribution of one cell state variable, the cell surface receptor number The dashed blue lines show the traditional cell state averaged mutual information (Eq. 1). The inset shows the dependence between cell state specific mutual information and cell state variable The input distribution is assumed to be a gamma distribution. **B.** A schematic showing the effect of

heterogeneity in cell states on population level response. Even when individual cells have little overlap in their responses to extracellular signal (bottom), the population level responses could have significant overlap (top), leading to a low mutual information between cell state averaged response and the input. **C.** A combined schematic of the two growth factor pathways. Extracellular growth factor ligand (red circle) binds to cell surface receptors which are shuttled to and from the plasma membrane continuously. Ligand bound receptors are phosphorylated and activate Akt. Phosphorylated Akt leads to phosphorylation of FoxO which bars it from entering the nucleus. EGF/EGFR model is limited to the reactions on the plasma membrane. The corresponding cell state variables are given by: $\boldsymbol{\theta} \equiv \{k_{prod}, k_{bind}, k_{unbind}, k_p, k_{dp}, k_{deg}, k^*_{deg}\}$. The cell state variables for the IGF/FoxO model are given by: $\boldsymbol{\theta} \equiv \{k_{prod}, k_{bind}, k_{unbind}, k_p, k_{dp}, k_{deg}, k^*_{deg}, [Akt], k_{ap}, k_{adp}, k_{in}, k_{ef}, k_{fp}, k_{fdp}, [FoxO]\}$. **D**. The estimated distribution of single cell mutual information values $p_{\mathrm{CeeMI}}(I)$ for the EGF/EGFR pathway using a maximum entropy estimation of $p(\boldsymbol{\theta})$. The inset shows the input distribution $p(u)$ corresponding to the maximum of the average I_{Cee} of $p_{\mathrm{CeeMI}}(I)$ (blue), along with the input distribution corresponding to the channel capacity of I_{CSA} (green). **E**. Same as **D** for the IGF/FoxO pathway. We additionally show the experimentally estimated $p_{\mathrm{CeeMI}}(I)$ (pink).

Experimental verification of $p_{\mathrm{CeeMI}}(I)$ using growth factor signaling networks

In cell populations, state variables that govern signaling dynamics such as protein levels (receptors, kinases, dephosphatases, etc.)[20, 21] as well as effective rate constants such as endocytosis rates[39], ligand binding rates[40], etc. vary from cell to cell. Therefore, we expect the downstream phenotype of environmental sensing to be widely distributed as well. To experimentally verify the computational prediction of the distribution $p_{\mathrm{CeeMI}}(I)$ of sensing abilities, we need a system that allows us to approximate the cell state specific response distribution $p(x|u, \boldsymbol{\theta})$. The IGF/FoxO pathway[35] is an ideal system for these explorations for several reasons. First, following IGF stimulation, the transcription factor FoxO is pulled out of the nucleus. GFP-tagged variant of FoxO can be used to detect the dynamics of nuclear levels of FoxO at the single cell level[41]. Second, nuclear FoxO levels reach an approximate steady state within 30-45 minutes of constant IGF stimulation with FoxO levels decreasing with increasing IGF dose[41]. As a result, an approximate cell state specific distribution $p(x|u, \boldsymbol{\theta})$ of steady state levels of FoxO can be obtained by stimulating the same cell with successive IGF levels. Finally, the biochemical interactions in the IGF/FoxO are well studied (Fig. 2C), allowing us to a build stochastic biochemical model based on previous computational efforts[42, 43] that fits the single cell data accurately. Another system where $p_{\mathrm{CeeMI}}(I)$ can in principle be verified is the EGF/EGFR pathway (Fig. 2C). Here too, abundance of cell surface EGFR levels can be tracked in live cells following EGF stimulation[44], allowing us to obtain cell-state specific response distribution $p(x|u, \boldsymbol{\theta})$. Below, we show estimates of

233 $p_{\mathrm{CeeMI}}(I)$ for both pathways and an experimental verification of our estimates for the IGF/FoxO
234 pathway where live-cell imaging data was previously collected[41].
235
236 The details of the calculations presented below can be found in SI Sections 3 and 4. Briefly, we first
237 constructed stochastic biochemical models for the two pathways based on known interactions (Fig. 2C)
238 and previous models[42, 43]. The output for the EGFR pathway was defined as the steady state levels of cell
239 surface EGF receptors and the output for the IGF/FoxO pathway was defined as the steady state nuclear
240 levels of the transcription factor FoxO. Using previously collected single cell data on the two pathways[37,
241 41, 43] and our maximum entropy-based framework[37], we estimated the distribution over parameters $p(\boldsymbol{\theta})$
242 for the two pathways (SI Section 3). Using these inferred distributions, and the model-predicted cell
243 state specific response distribution $p(x|u,\boldsymbol{\theta})$, we could compute $p_{\mathrm{CeeMI}}(I)$ for any specified input
244 distribution $p(u)$. We choose the support of the input distribution as the ligand concentrations used in
245 the experimental setup. The estimates shown in Fig. 2D and Fig. 2E show $p_{\mathrm{CeeMI}}(I)$ corresponding to the
246 input distribution $p(u)$ that maximizes the conditional mutual information I_{Cee} (see Eq. 6). This input
247 distribution is shown in the inset of Fig. 2D and 2E.
248
249 Similar to the illustrative example (Fig. 2A), there is a wide distribution of single cell sensing fidelities in
250 real populations (Fig. 2D and Fig. 2E). Moreover, most cells are better sensors compared to the "average
251 cell", a cell whose response $p(x|u)$ is averaged over cell state variability which was estimated to be ~ 1
252 bit for both pathways. The cellular signaling fidelity skews towards the upper limit of 2 bits which
253 corresponds to the logarithm of the number of inputs used in the experiment. Indeed, as seen in the
254 insets of Fig. 2D and Fig. 2E, the input distribution corresponding to the maximum of the cell state
255 agnostic mutual information I_{CSA} is concentrated on the lowest and the highest input for both
256 pathways, indicating that cells may be able to detect only two input levels. In contrast, the input
257 distribution corresponding to the maximum of I_{Cee} is close to uniform, suggesting that individual cells
258 can in fact resolve different ligand levels.
259
260 To verify our computational estimate of $p_{\mathrm{CeeMI}}(I)$ for the IGF/FoxO pathway, we reanalyzed previously
261 collected data wherein several cells were stimulated using successive IGF levels[41]. The details of our
262 calculations can be found in SI Section 4. Briefly, the cells reached an approximate steady state within 60
263 minutes of each stimulation and nuclear FoxO levels measured between 60 to 90 minutes were used to
264 approximate an experimental cell state specific response distribution $p(x|u,\boldsymbol{\theta})$. The distribution
265 $p_{\mathrm{CeeMI}}(I)$ was then obtained by maximizing the average mutual information I_{Cee} (averaged over all

cells) with respect to the input distribution. As seen in Fig. 2D and Fig. 2E, the experimentally evaluated single cell signaling fidelities match closely with our computational estimates. Moreover, as predicted using our computational analysis, individual cells in the population were significantly better at sensing than what is implied by the maximum of I_{CSA}. Indeed, the distribution of steady state FoxO levels were found to be well-resolved at the single cell level as well (Fig. 2E). Live cell imaging data for the EGFR pathway was not available and we leave it to future studies to validate our predictions.

Our calculations show that real cell populations comprise cells that have differing sensing fidelities, individual cells are significantly better at sensing their environment than what traditional estimates would indicate, and importantly, the CeeMI approach can accurately estimate the variation in signaling performances using readily collected experimental data and stochastic biochemical models. Notably, the variability in cell states and therefore the heterogeneity in sensing abilities is likely to be stable over time; the same cell's FoxO responses to the same input were found to have significantly less variation compared to the variation within the population (SI Section 4).

CeeMI identifies biochemical parameters that determine cells' ability to sense their environment

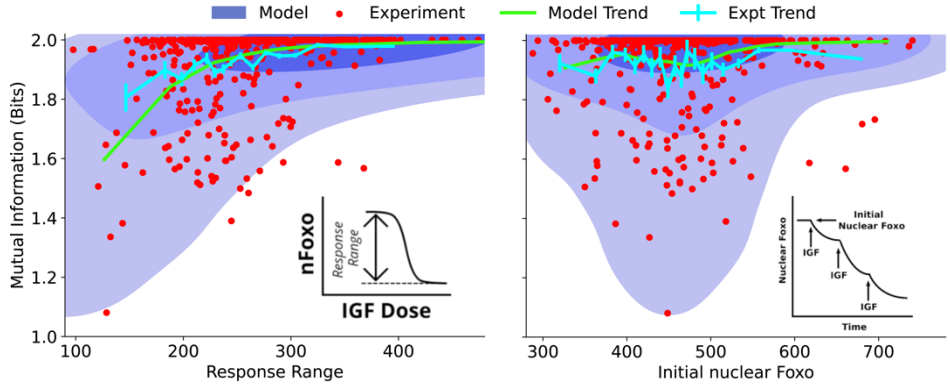

Figure 3. Dependence on cell state dependent mutual information on biochemical parameters. (left) The joint distribution $p_{\mathrm{CeeMI}}(I, \chi)$ of cell state specific mutual information and biochemical parameter χ chosen to be the single cell response range of nuclear FoxO levels (x-axis, see inset for a cartoon). The shaded blue regions are model predictions, and the green line is the model average. The darker shades represent higher probabilities. The red dots represent experimental cells. The cyan line represents experimental averages. (right) same as (left) with biochemical parameter χ chosen to be steady state nuclear Foxo levels in the absence of stimulation. The contours represent 1% to 10%, 10% to 50%, and 50% to 100% of the total probability mass (from faint to dark shading).

Like other phenotypes[21] and signaling dynamics[20, 30], we expect that cells' ability to sense their environment depends on their state variables. For example, cells with faster endocytosis rates[39] may integrate environmental fluctuations with higher accuracy[43]. Or cells with higher receptor numbers[21] will lead to a lowered relative noise in bound ligand concentration.

To systematically identify the variables that differentiate between cells' ability to sense the environment, we quantify the joint distribution $p_{\mathrm{CeeMI}}(I, \chi)$ of single cell signaling fidelity and biochemical state variables that take part in the signaling network. To test whether we can identify variables that are predictive of cellular fidelity, we estimated the joint distribution $p_{\mathrm{CeeMI}}(I, \chi)$ for two variables that were experimentally accessible, response range of nuclear FoxO (Fig. 3 left, see inset) and total nuclear FoxO levels prior to IGF stimulation (Fig. 3 right, see inset). In both figures, the shaded regions show computational estimates of the joint probability distributions, and the red circles represent real cells. The green and the cyan trend lines represent computational and experimental binned averages.

One may expect that higher total nuclear FoxO levels could result in lower noise (coefficient of variation) and therefore better sensing abilities. However, we find that total nuclear FoxO levels only weakly correlate with cell state dependent mutual information (correlation coefficient $r = 0.16$ for computational estimates and $r = 0.04$ for experimental data). In contrast, cell state dependent mutual information depended strongly on the dynamic range of the response (correlation coefficient $r = 0.53$ for computational estimates and $r = 0.29$ for experimental data). Importantly, the model captured the observation that cells with a small response range had a variable sensing abilities while cells with a large response range all performed well in resolving extracellular IGF levels. Surprisingly, the total nuclear FoxO levels only weakly correlated with the cell specific mutual information. In SI Section 5, we show the predicted joint distributions $p_{\mathrm{CeeMI}}(I, \chi)$ for several other biochemical variables that can potentially govern single cells' response to extracellular IGF stimuli. This way, CeeMI can be used to systematically identify cell state variables that differentiate between good sensors and bad sensors of the environmental stimuli.

Time scale of stochastic dynamics of cell states dictates the divergence between state-specific and state-agnostic sensing abilities

A key limitation of the presented analysis is the assumption that cell state variables remain approximately constant over the time scale of typical environmental fluctuations. While many cell state variables change slowly, remaining roughly constant over the life spans of cells and beyond[25, 26], state changes may occur within the lifespan of a cell as well[45]. These dynamical changes can be

accommodated easily in our calculations. Here, instead of fixing the cell state variables $\boldsymbol{\theta}$, we can treat them as initial conditions and propagate them stochastically with their own dynamics. In the limit of very fast dynamics where individual cells rapidly transition cell states, we expect that the individuality of cells in a population vanishes and consequently cell state specific mutual information for each cell will agree with traditional cell state agnostic estimates of the channel capacity. In contrast, if the cell state dynamics are slow, our framework highlights differences between cells in the population.

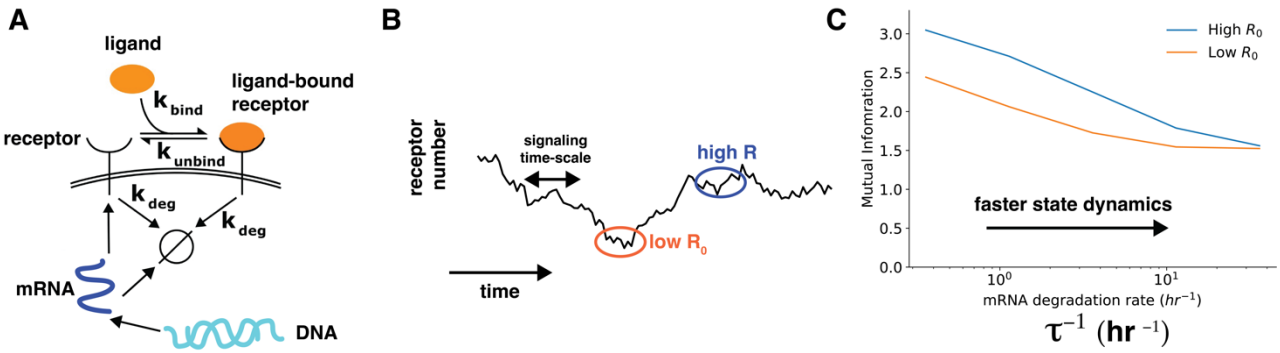

Figure 4. Cell state dynamics governs cell state conditioned mutual information. A. In a simple stochastic model, receptor mRNA is produced at a constant rate from the DNA and the translated into ligand-free receptors. The number of ligand-bound receptors after a short exposure to ligands is considered the output. **B.** A schematic showing dynamics of receptor numbers when mRNA dynamics are slower compared to signaling time scales. **C.** Conditioning on receptor numbers leads to differing abilities in sensing the environment when the time scale of mRNA dynamics is slow. In contrast, when the mRNA dynamics are fast (large τ^{-1}), conditioning on cell state variables does not lead to difference in sensing abilities.

To elucidate the role of cell state dynamics, we built a simple model of ligand-receptor system (SI Section 6, Fig. 4). Briefly, the model included production and degradation of the receptor mRNA, translation of the receptor protein from the mRNA, and ligand binding to the receptor (Fig.4A). The number of ligand-free receptors and the number mRNA molecules were together considered the state variables. We tuned the mRNA dynamics by keeping the average mRNA number constant while simultaneously changing mRNA production and degradation rates. The time scale of mRNA dynamics is denoted by τ.

When the mRNA dynamics are slower than translation, ligand binding, and receptor degradation, individual cells are effectively frozen around a particular receptor number (Fig. 4B). The distribution of the number of ligand-bound receptors after a short-term ligand exposure reflects this frozen state. As a result, the cell-state conditioned mutual information depends strongly the cell state, with higher mutual information for cells with higher receptor numbers. In contrast, as the mRNA dynamics become faster,

cells quickly lose the memory of cell state conditioning. As a result, the distribution of the number of ligand-bound receptors after a short-term ligand exposure reflects an ergodic averaging over the underlying mRNA dynamics. This averaging effectively washes out differences between cells (Fig. 4C).

This example shows that when cells states change at time scales slower than signaling dynamics, cells in a population can be identified by their cell state and their ability to sense the environment can differ from one another.

Discussion

Cell populations are characterized by heterogeneity in cell state variables[18, 19] that is responsible for important phenotypic differences and selective advantage for example in sensitivity to drugs[25, 26, 28], response to chemotactic signals[46], and following proliferation cues[29]. Therefore, it is reasonable to expect that cells' ability to sense their environment depends on their state and is therefore variable across cells in a population. To quantify this heterogeneity, here, we developed a novel information theoretic perspective that allowed us to quantify the distribution $p_{\mathrm{CeeMI}}(I)$ of single cell sensing abilities using easily measurable single cell data and stochastic models of signaling networks. We also quantified the joint distribution $p_{\mathrm{CeeMI}}(I, \chi)$ of cell specific sensing ability and biochemical cell state variable. Importantly, using two growth factor pathways, we showed that individual cells in real cell populations were much better at sensing their environment compared to what is implied by the traditional estimate of channel capacities of signaling networks. Typical single cell data are time stamped and do not give information about dynamics in the single cell. Moreover, these data do not provide information about dependence of cell-to-cell variability on cell state variables. Therefore, computational modeling of cellular trajectories from single cell data and stochastic models is essential in deciphering mechanistic origins of heterogeneity in cellular phenotypes.

The approach presented here will be useful in identifying bottlenecks in signal transduction. Many cellular phenotypes such as chemotaxis[7, 8] and cell proliferation[47] exhibit a weak correlation between cellular outputs (e.g., directional alignment with chemical gradients) with the input (e.g., gradient strength), resulting in a low channel capacity even for individual cells. In such cases, it is important to understand where exactly along the information transduction pathway is the information about the gradient is lost. If traditional calculations are pursued, for example, for movement of mammalian cells under growth factor gradients[7, 31], one may conclude that the information loss likely happens right at the

receptor level (Fig. 2D). In contrast, CeeMI will allow us to disentangle the effect of cell state heterogeneity and noisy cellular response to precisely pinpoint intracellular signaling nodes that are responsible for signal corruption.

How do we contrast our results with previous low estimates of channel capacities? There is extensive population heterogeneity in cell state variables and this heterogeneity often is stable over time. Nonetheless, this heterogeneity arises from stochasticity in underlying processes including dynamics of epigenetic transitions and gene expression. When one specifies state variables[19] $\boldsymbol{\theta}$, one effectively conditions on the stochastic trajectory of the underlying dynamics to a specific narrow window (Fig. 4). This extra conditioning narrows the cells' response distributions, revealing the heterogeneous nature of cell populations[19].

In summary, we showed that like other phenotypes, the ability to sense the environment is itself heterogeneously distributed in a cell population. Moreover, we also showed that when conditioned on cell state variables, mammalian cells appear to be significantly better at sensing their environment than what traditional mutual information calculation suggests. Finally, we showed that we could identify cell state variables that made some cells better sensors compared to others. We believe that CeeMI will be an important framework in quantifying fidelity of input/output relationships in heterogeneous cell populations.

Acknowledgments

AG, HA, and PD are supported by NIGMS grant R35GM142547. The authors would like to thank Andre Levchenko for useful discussions.

References

1. Levchenko A, Nemenman I. Cellular noise and information transmission. Curr Opin Biotechnol. 2014;28:156-64.
2. Rhee A, Cheong R, Levchenko A. The application of information theory to biochemical signaling systems. Phys Biol. 2012;9(4):045011.
3. ten Wolde PR, Becker NB, Ouldridge TE, Mugler A. Fundamental Limits to Cellular Sensing. J Stat Phys. 2016;162(5):1395-424.
4. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A. Information transduction capacity of noisy biochemical signaling networks. Science. 2011;334(6054):354-8.
5. Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, Tsimring L, Wollman R. Accurate information transmission through dynamic biochemical signaling networks. Science. 2014;346(6215):1370-3.
6. Suderman R, Bachman JA, Smith A, Sorger PK, Deeds EJ. Fundamental trade-offs between information flow in single cells and cellular populations. P Natl Acad Sci USA. 2017;114(22):5755-60.
7. Varennes J, Moon HR, Saha S, Mugler A, Han B. Physical constraints on accuracy and persistence during breast cancer cell chemotaxis. Plos Comput Biol. 2019;15(4). PubMed PMID: WOS:000467530600065.
8. Mattingly HH, Kamino K, Machta BB, Emonet T. Escherichia coli chemotaxis is information limited. Nat Phys. 2021;17(12):1426-+.
9. Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken, N.J.: Wiley-Interscience; 2006. xxiii, 748 p. p.
10. Wu M, Singh AK. Single-cell protein analysis. Curr Opin Biotechnol. 2012;23(1):83-8.
11. Slavov N. Single-cell protein analysis by mass spectrometry. Curr Opin Chem Biol. 2021;60:1-9.
12. Lemon WC, McDole K. Live-cell imaging in the era of too many microscopes. Curr Opin Cell Biol. 2020;66:34-42.
13. Hansen AS, O'Shea EK. Limits on information transduction through amplitude and frequency regulation of transcription factor activity. Elife. 2015;4.
14. Billing U, Jetka T, Nortmann L, Wundrack N, Komorowski M, Waldherr S, Schaper F, Dittrich A. Robustness and Information Transfer within IL-6-induced JAK/STAT Signalling. Commun Biol. 2019;2:27.
15. Garner KL, Perrett RM, Voliotis M, Bowsher C, Pope GR, Pham T, Caunt CJ, Tsaneva-Atanasova K, McArdle CA. Information Transfer in Gonadotropin-releasing Hormone (GnRH) Signaling: EXTRACELLULAR SIGNAL-REGULATED KINASE (ERK)-MEDIATED FEEDBACK LOOPS CONTROL HORMONE SENSING. J Biol Chem. 2016;291(5):2246-59.

467 16. Frick CL, Yarka C, Nunns H, Goentoro L. Sensing relative signal in the Tgf-beta/Smad
468 pathway. Proc Natl Acad Sci U S A. 2017;114(14):E2975-E82.
469 17. Lee JB, Caywood LM, Lo JY, Levering N, Keung AJ. Mapping the dynamic transfer
470 functions of eukaryotic gene regulation. Cell Syst. 2021;12(11):1079-93 e6.
471 18. Trapnell C. Defining cell types and states with single-cell genomics. Genome Res.
472 2015;25(10):1491-8.
473 19. Wollman R. Robustness, Accuracy, and Cell State Heterogeneity in Biological Systems.
474 Curr Opin Syst Biol. 2018;8:46-50.
475 20. Meyer R, D'Alessandro LA, Kar S, Kramer B, She B, Kaschek D, Hahn B, Wrangborg D,
476 Karlsson J, Kvarnstrom M, Jirstrand M, Lehmann WD, Timmer J, Hofer T, Klingmuller U.
477 Heterogeneous kinetics of AKT signaling in individual cells are accounted for by variable protein
478 concentration. Front Physiol. 2012;3:451.
479 21. Frankel NW, Pontius W, Dufour YS, Long J, Hernandez-Nunez L, Emonet T. Adaptability
480 of non-genetic diversity in bacterial chemotaxis. Elife. 2014;3.
481 22. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved,
482 highly multiplexed RNA profiling in single cells. Science. 2015;348(6233):aaa6090. Epub
483 20150409.
484 23. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A,
485 Fraser P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature.
486 2013;502(7469):59-64.
487 24. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY,
488 Greenleaf WJ. Single-cell chromatin accessibility reveals principles of regulatory variation.
489 Nature. 2015;523(7561):486-90. PubMed PMID: 26083756.
490 25. Flusberg DA, Roux J, Spencer SL, Sorger PK. Cells surviving fractional killing by TRAIL
491 exhibit transient but sustainable resistance and inflammatory phenotypes. Mol Biol Cell.
492 2013;24(14):2186-200.
493 26. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell
494 variability in TRAIL-induced apoptosis. Nature. 2009;459(7245):428-32.
495 27. Pleska M, Jordan D, Frentz Z, Xue B, Leibler S. Nongenetic individuality, changeability,
496 and inheritance in bacterial behavior. Proc Natl Acad Sci U S A. 2021;118(13).
497 28. Emert BL, Cote CJ, Torre EA, Dardani IP, Jiang CL, Jain N, Shaffer SM, Raj A. Variability
498 within rare cell states enables multiple paths toward drug resistance. Nat Biotechnol.
499 2021;39(7):865-76.
500 29. Matson JP, Cook JG. Cell cycle proliferation decisions: the impact of single cell analyses.
501 FEBS J. 2017;284(3):362-75.
502 30. Norris D, Yang P, Shin SY, Kearney AL, Kim HJ, Geddes T, Senior AM, Fazakerley DJ,
503 Nguyen LK, James DE, Burchfield JG. Signaling Heterogeneity is Defined by Pathway
504 Architecture and Intercellular Variability in Protein Expression. Iscience. 2021;24(2):102118.
505 31. Moon HR, Saha S, Mugler A, Han B. Signal processing capacity of the cellular sensory
506 machinery regulates the accuracy of chemotaxis under complex cues. Iscience. 2021;24(11).
507 32. Waite AJ, Frankel NW, Dufour YS, Johnston JF, Long J, Emonet T. Non-genetic diversity
508 modulates population performance. Mol Syst Biol. 2016;12(12):895.
509 33. Clark JA, Campbell SG. Diverse relaxation rates exist among rat cardiomyocytes isolated
510 from a single myocardial region. J Physiol. 2019;597(3):711-22.

511 34. Yao J, Pilko A, Wollman R. Distinct cellular states determine calcium signaling response.
512 Mol Syst Biol. 2016;12(12):894.
513 35. Lewis Cantley TH, Richard Sever, Jeremy Thorner Signal Transduction: Principles,
514 Pathways, and Processes 1st ed: Cold Spring Harbor Laboratory Press; 2013.
515 36. Feinberg AP, Levchenko A. Epigenetics as a mediator of plasticity in cancer. Science.
516 2023;379(6632):eaaw3835. Epub 20230210. doi: 10.1126/science.aaw3835. PubMed PMID:
517 36758093; PMCID: PMC10249049.
518 37. Dixit PD, Lyashenko E, Niepel M, Vitkup D. Maximum Entropy Framework for Predictive
519 Inference of Cell Population Heterogeneity and Responses in Signaling Networks. Cell Syst.
520 2020;10(2):204-12 e8.
521 38. Loos C H, J. Mathematical modeling of variability in intracellular signaling. Current
522 Opinion in Systems Biology. 2019;16:17-24.
523 39. Kallenberger SM, Unger AL, Legewie S, Lymperopoulos K, Klingmuller U, Eils R, Herten
524 DP. Correlated receptor transport processes buffer single-cell heterogeneity. Plos Comput Biol.
525 2017;13(9):e1005779.
526 40. Chung JC, Sciaky N, Gross DJ. Heterogeneity of epidermal growth factor binding kinetics
527 on individual cells. Biophys J. 1997;73(2):1089-102.
528 41. Gross SM, Dane MA, Bucher E, Heiser LM. Individual Cells Can Resolve Variations in
529 Stimulus Intensity along the IGF-PI3K-AKT Signaling Axis. Cell Syst. 2019;9(6):580-8 e4.
530 42. Wimmer RJ, Liu Y, Schachter TN, Stonko DP, Peercy BE, Schneider MF. Mathematical
531 modeling reveals modulation of both nuclear influx and efflux of Foxo1 by the IGF-I/PI3K/Akt
532 pathway in skeletal muscle fibers. Am J Physiol Cell Physiol. 2014;306(6):C570-84.
533 43. Lyashenko E, Niepel M, Dixit PD, Lim SK, Sorger PK, Vitkup D. Receptor-based
534 mechanism of relative sensing and cell memory in mammalian signaling networks. Elife. 2020;9.
535 44. Fortian A, Sorkin A. Live-cell fluorescence imaging reveals high stoichiometry of Grb2
536 binding to the EGF receptor sustained during endocytosis. J Cell Sci. 2014;127(Pt 2):432-44.
537 45. MacLean AL, Hong T, Nie Q. Exploring intermediate cell states through the lens of single
538 cells. Curr Opin Syst Biol. 2018;9:32-41.
539 46. Mattingly HH, Emonet T. Collective behavior and nongenetic inheritance allow bacterial
540 populations to adapt to changing environments. Proc Natl Acad Sci U S A.
541 2022;119(26):e2117377119.
542 47. Gross SM, Rotwein P. Unraveling Growth Factor Signaling and Cell Cycle Progression in
543 Individual Fibroblasts. J Biol Chem. 2016;291(28):14628-38.
544