bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Maximum Entropy Framework For Inference Of Cell Population Heterogeneity In Signaling Networks

Purushottam D. Dixit*[1], Eugenia Lyashenko*[1], Mario Niepel[2], and Dennis Vitkup[1,3,4]

Correspondence should be addressed to P. D. (dixitpd@gmail.com) or D.V. (dv2121@columbia.edu).

[1]Department of Systems Biology, Columbia University
[2]Department of Systems Biology, Harvard Medical School
[3]Department of Biomedical Informatics, Columbia University
[4]Center for Computational Biology and Bioinformatics, Columbia University
* Contributed equally

Abstract

Predictive models of signaling networks are essential tools for understanding cell population heterogeneity and designing rational interventions in disease. However, using network models to predict signaling dynamics heterogeneity is often challenging due to the extensive variability of signaling parameters across cell populations. Here, we describe a **M**aximum **E**ntropy-based f**R**amework for **I**nference of heterogeneity in **D**ynamics of s**I**g**A**ling **N**etworks (MERIDIAN). MERIDIAN allows us to estimate the joint probability distribution over signaling parameters that is consistent with experimentally observed cell-to-cell variability in abundances of network species. We apply the developed approach to investigate the heterogeneity in the signaling network activated by the epidermal growth factor (EGF) and leading to phosphorylation of protein kinase B (Akt). Using the inferred parameter distribution, we also predict heterogeneity of phosphorylated Akt levels and the distribution of EGF receptor abundance hours after EGF stimulation. We discuss how MERIDIAN can be generalized and applied to problems beyond modeling of heterogeneous signaling dynamics.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Signaling cascades in genetically identical cells often respond in a heterogeneous manner to extracellular stimuli (Raj and van Oudenaarden 2008). This heterogeneity arises largely due to cell-to-cell variability in biochemical signaling parameters such as reaction propensities and chemical species abundances (Albeck, Burke et al. 2008, Spencer, Gaudet et al. 2009, Meyer, D'Alessandro et al. 2012, Llamosi, Gonzalez-Vargas et al. 2016, Kallenberger, Unger et al. 2017). This variability in parameters can have important functional consequences, for example, in fractional killing of cancer cells treated with chemotherapeutic drugs (Albeck, Burke et al. 2008, Spencer, Gaudet et al. 2009). Therefore, the knowledge of the distribution over parameters is essential to understanding phenotypic heterogeneity in cell populations.

Several experimental techniques such as flow cytometry (Wu and Singh 2012), immunofluorescence (Wu and Singh 2012), and live cell assays (Meyer, D'Alessandro et al. 2012) have been developed to investigate the variability of biochemical species abundances. However, it is often difficult to estimate the distribution over biochemical parameters from these experimental measurements. The reasons for this challenge are primarily threefold. First, parameters such as protein abundances and reaction propensities vary substantially across cells in a population (Raj and van Oudenaarden 2008). For example, previous studies have reported the coefficients of variation of protein abundances in the range 0.1 – 0.6 (Niepel, Spencer et al. 2009). This would make the effective rates of signaling reactions vary substantially between cells as well (Chung, Sciaky et al. 1997, Meyer, D'Alessandro et al. 2012). Second, the multivariate parameter distribution can potentially have a complex shape. For example, as is often the case in single cell data, gene expression, and thus cellular abundance, of key signaling proteins may exhibit multimodality (Shalek, Satija et al. 2013). Finally, single cell measurements are typically not sufficient to uniquely infer the underlying parameter variability; the problem usually referred as parameter non-identifiability (Banks 2012).

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Over the last decade, several computational methods have been developed to estimate the joint distribution over parameters consistent with experimentally measured cell-to-cell variability in biochemical species (Waldherr, Hasenauer et al. 2009, Hasenauer, Waldherr et al. 2011, Zechner, Ruess et al. 2012, Hasenauer, Hasenauer et al. 2014, Zechner, Unger et al. 2014, Loos, Moeller et al. 2018, Waldherr 2018). Most of these methods circumvent the ill-posed inverse problem of estimation of the parameter distribution (Banks 2012) by making specific *ad hoc* choices about the shape of the distribution. For example, Hasenauer et al. (Hasenauer, Waldherr et al. 2011, Hasenauer, Hasenauer et al. 2014) (see also (Waldherr, Hasenauer et al. 2009, Loos, Moeller et al. 2018)) approximate the parameter distribution as a linear combination of predefined distribution functions. Waldherr et al. (Waldherr, Hasenauer et al. 2009) approximate the parameter distribution using Latin hypercube sampling. Similarly Zechner et al. (Zechner, Ruess et al. 2012, Zechner, Unger et al. 2014) assume that the parameters are distributed according to a log-normal or a gamma distribution. Consequently, it may be difficult for these previously developed approaches to infer a complex multivariate parameter distribution of an unknown shape.

Building on our previous work (Dixit 2013, Eydgahi, Chen et al. 2013), we developed **MERIDIAN:** a **M**aximum **E**ntropy-based f**R**amework for **I**nference of heterogeneity in **D**ynamics of s**I**gn**A**ling **N**etworks. Instead of enforcing a specific functional form of the parameter distribution *a priori*, MERIDIAN uses data-derived constraints to derive it *de novo*. The maximum entropy principle was first introduced more than a century ago in statistical physics (Dixit, Wagoner et al. 2018). Among all candidate distributions that agree with imposed constraints, the maximum entropy principle selects the one with the least amount of over-fitting. Maximum entropy-based approaches have been successfully applied previously to a variety of biological problems, including protein structure prediction (Weigt, White et al. 2009), protein sequence evolution (Mora, Walczak et al. 2010), neuron firing dynamics (Schneidman, Berry et al. 2006), molecular simulations (Dixit, Jain et al. 2015, Tiwary and Berne 2016), and dynamics of biochemical

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reaction networks (Dixit 2018).

Following a description of the key ideas behind MERIDIAN, we illustrate its performance using synthetic data on a simplified model of growth factor signaling. Next, we use the framework to study the heterogeneity in the signaling network leading to phosphorylation of protein kinase B (Akt). Epidermal growth factor (EGF)-induced Akt phosphorylation governs key intracellular processes (Manning and Toker 2017) including metabolism, apoptosis, and cell cycle entry. Due to its central role in mammalian signaling, aberrations in the Akt pathway are implicated in multiple diseases (Herbst 2004, Manning and Toker 2017). We apply MERIDIAN to infer the distribution over signaling parameters using experimentally measured levels of phosphorylated Akt (pAkt) and cell surface EGFR (sEGFR) in MCF10A cells (Soule, Maloney et al. 1990) following EGF stimulation. We then demonstrate that the obtained parameter distribution allows us to accurately predict the heterogeneity in single cell pAkt levels at late time points, as well as the heterogeneity in cell surface EGFRs in response to EGF stimulation. Finally, we discuss the generalizations of the framework to study problems beyond modeling heterogeneity in signaling networks.

Results

Outline of MERIDIAN

We consider a signaling network comprising *N* chemical species whose intracellular abundances we denote by $\bar{x} = \{x_1, x_2, \dots, x_N\}$. We assume that the molecular interactions among the species are described by a system of ordinary differential equations

$$\frac{d}{dt}\bar{x}(t, \bar{\theta}) = f(\bar{x}, \bar{\theta}) \tag{1}$$

where $f(\bar{x}, \bar{\theta})$ is a function of species abundances $\bar{x}$. Here, $\bar{\theta} = \{\theta_1, \theta_2, \dots\}$ is a vector of parameters that describe the dynamics of the signaling networks. We denote by $x_a(t, \bar{\theta})$ the solution of equations (1) for species "*a*" at time *t* with parameters $\bar{\theta}$.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Our computational approach is illustrated in Figure 1. We use experimentally measured cell-to-cell variability of protein species "*a*" at multiple experimental conditions, for example, several time points, (illustrated by histograms in Figure 1) to constrain the parameter distribution $P(\bar{\theta})$. Specifically, we first quantify the experimentally measured cell-to-cell variability by estimating bin fractions ϕ_{ik}. In our notation, the index *i* specifies the experimental condition (measurement time, measured species, input conditions, etc.) and *k* indicates bin number. Every distinct dynamical trajectory $x_a(t, \bar{\theta})$ (illustrated by red and blue curves in Figure 1) generated by specific parameter values $\bar{\theta}$ passes through a unique set of bins (red curve through red bins and blue curve through blue bins in Figure 1) at multiple experimental conditions. Using MERIDIAN, we find a corresponding probability distribution over signaling parameters $P(\bar{\theta})$ such that the distribution over trajectories $P[x_a(t, \bar{\theta})]$ is consistent with all experimentally measured bin fractions. Below, we present our development to derive the functional form of $P(\bar{\theta})$ that is consistent with these constraints.

Derivation of $P(\bar{\theta})$ using MERIDIAN

For simplicity, we consider the case when the distribution of cell-to-cell variability in one species x_a is available only at one time point *t* (for example, *t* = t_1 in Figure 1). We denote by $\bar{\phi} = \{\phi_1, \phi_2, \dots, \phi_B\}$ the fraction of cells whose experimental measurement of x_a lies in individual bins. Here, given that we are considering only one experimental condition, for brevity we use only one index to indicate the bin fractions. Below, we first assume that there are no experimental errors in determining $\bar{\phi}$. Later, we demonstrate how to incorporate known experimental errors both in the inference procedure and in making predictions using $P(\bar{\theta})$.

Given a parameter distribution $P(\bar{\theta})$, the predicted fractions $\bar{\psi} = \{\psi_1, \psi_2, \dots, \psi_B\}$ can be obtained as follows. Using Markov chain Monte Carlo (MCMC), we generate multiple parameter sets from $P(\bar{\theta})$. For each sampled set of parameters $\bar{\theta}$, we solve equations (1) and find $x_a(t, \bar{\theta})$, i.e. the predicted value

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of the abundance at time *t*. Then, using the samples from the ensemble of trajectories, we estimate ψ_k as the fraction of sampled trajectories where $x_a(t,\bar{\theta})$ passed through the kth bin. Mathematically,

$$\psi_k = \int I_k\left(x_a(t,\bar{\theta})\right) P(\bar{\theta}) d\bar{\theta} \tag{2}$$

where $I_k(x)$ is an indicator function; $I_k(x)$ is equal to one if *x* lies in the kth bin and zero otherwise.

The central idea behind MERIDIAN is to find the maximum entropy distribution $P(\bar{\theta})$ over parameters such that all predicted fractions ψ_k agree with those estimated from experimental data, ϕ_k. Formally, we seek $P(\bar{\theta})$ with the maximum entropy

$$S = -\int P(\bar{\theta}) \log \frac{P(\bar{\theta})}{q(\bar{\theta})} d\bar{\theta} \tag{3}$$

subject to normalization ($\int P(\bar{\theta})\, d\bar{\theta} = 1$) and data-derived constraints $\psi_k = \phi_k$ for all *k*. Here, $q(\bar{\theta})$ plays a role similar to the prior distribution in Bayesian approaches (Caticha and Preuss 2004). In this work, we choose $q(\bar{\theta})$ to be a uniform distribution within literature-derived ranges of parameters, but other choices can be implemented as well.

To impose aforementioned constraints and perform the entropy maximization, we use the method of Lagrange multipliers. To that end, we write the Lagrangian function

$$L = S + \eta\left(\int P(\bar{\theta})\, d\bar{\theta} - 1\right) - \sum_{k=1}^{B} \lambda_k \left(\int I_k\left(x_a(t,\bar{\theta})\right) P(\bar{\theta}) d\bar{\theta} - \phi_k\right) \tag{4}$$

where η is the Lagrange multiplier associated with normalization and λ_k are the Lagrange multipliers associated with fixing the predicted fractions ψ_k to their experimentally measured values ϕ_k in all bins. Differentiating equation (4) with respect to $P(\bar{\theta})$ and setting the derivative to zero, we obtain

$$P(\bar{\theta}|\bar{\lambda}) = \frac{1}{\Omega} q(\bar{\theta}) \exp\left(-\sum_{k=1}^{B} \lambda_k I_k\left(x_a(t,\bar{\theta})\right)\right) \tag{5}$$

where Ω is the partition function that normalizes the probability distribution. Equation (5) is a key conceptual foundation of this work. We use it to estimate

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the parameter distribution from user-specified constraints. Here, for the sake of notational simplicity, we have restricted our discussion using measurement at a single time point. Later, we discuss how to generalize the approach when abundances of multiple species are measured at several time points (Methods equations (S24)).

Given the high dimensional nature of the parameter space in models of biological systems, the collected data is usually not sufficient to fully constrain the multidimensional parameter distribution (Banks 2012). As a result, the distribution inferred by MERIDIAN reflects both the true biological variability in parameters as well as parameter non-identifiability.

Numerical estimation of Lagrange multipliers

The Lagrange multipliers in equation (5) need to be numerically optimized such that the predicted bin fractions are consistent with the experimentally estimated ones. Notably, the search for the Lagrange multipliers is a convex optimization problem (Methods) and we solve it using an iterative algorithm proposed in (Tkacik, Schneidman et al. 2006) (see Figure 2). Briefly, we start from a randomly chosen point in the space of Lagrange multipliers. In the nth iteration of the optimization algorithm, using the current vector of the Lagrange multipliers $\bar{\lambda}^{(n)}$, we estimate the predicted bin reactions using MCMC (Methods). Next, we estimate the error vector $\bar{\Delta}^{(n)} = \bar{\psi}^{(n)} - \bar{\phi}$ for the nth iteration. We then update the multipliers for the n+1st iteration as $\bar{\lambda}^{(n+1)} = \bar{\lambda}^{(n)} + \alpha^{(n)}\bar{\Delta}^{(n)}$ (see Figure 2). The positive "learning rate" $\alpha^{(n)}$ is chosen to minimize the error $\bar{\Delta}^{(n+1)}$ (Methods). Given that the data is likely to have experimental errors (see below) and that the signaling network model (described by equations (1)) is an approximation to a real biological system, it may not be possible to achieve a complete agreement between data and model fits. Thus, we terminate the iterative optimization procedure once the error reaches a predefined accuracy cutoff.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Making predictions using $P(\bar{\theta})$

Here, we show how to make predictions using the inferred parameter distribution. In the discussion so far, we assumed that experimentally measured cell-to-cell variability had no errors. However, single cell experiments are often subject to uncertainty. Thus, we consider that the measurements are characterized by their mean values $\bar{\phi}$ as well as the standard errors of the mean $\bar{\sigma}$, which are estimated using several experimental replicates.

Let us denote by $\bar{\Lambda}$ the vector of optimized Lagrange multipliers (see Figure 2), and by $\bar{\psi}(\bar{\Lambda})$ the predicted bin fractions of species variability across cells (Figure 1). We represent the experimental errors in the estimated bin fractions using a distribution over the Lagrange multipliers themselves. We express the posterior distribution over of Lagrange multipliers $\bar{\lambda}$ as (Methods)

$$P(\bar{\lambda}) \propto \exp\left(-\sum_{k=1}^{B} \frac{\left(\psi_k(\bar{\lambda}) - \phi_k\right)^2}{2\sigma_k^2}\right). \tag{6}$$

In equation (6), we have assumed (1) that the errors are normally distributed and (2) a uniform prior over Lagrange multipliers; both these assumptions can in principle be relaxed. The distribution over model parameters taking into account the errors in the measurement is then given by

$$P(\bar{\theta}) = \int P(\bar{\theta}|\bar{\lambda}) P(\bar{\lambda})\, d\bar{\lambda} \tag{7}$$

where $P(\bar{\theta}|\bar{\lambda})$ is given by equation (5) and the full distribution $P(\bar{\theta})$ is given by averaging over the distribution of Lagrange multipliers $P(\bar{\lambda})$.

Equations (5-7) can be used, in principle, to make model predictions. For example, consider any quantity of interest $Y(\bar{\theta})$ that depends on model parameters. The mean m of $Y(\bar{\theta})$ is then given by

$$m = \int y(\bar{\lambda}) P(\bar{\lambda}) d\bar{\lambda}. \tag{8}$$

where

$$y(\bar{\lambda}) = \int Y(\bar{\theta}) P(\bar{\theta}|\bar{\lambda}) d\bar{\theta}. \tag{9}$$

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Similarly, the uncertainty s given by

$$s^2 = \left(\int y(\bar{\lambda})^2 P(\bar{\lambda}) d\bar{\lambda} \right) - m^2. \tag{10}$$

In practice it may be challenging to numerically integrate equations (8-10), because this requires sampling over multiple sets of Lagrange multipliers $\bar{\lambda}$, which in turn requires the estimation of $\bar{\psi}(\bar{\lambda})$ using MCMC (see equation (6)). However, estimates of the mean and the uncertainty can be analytically obtained if experimental errors are small (median error is 9% in our data). In this case, the model prediction m is simply given by:

$$m = y(\bar{\Lambda}) \tag{11}$$

In equation (11), $\bar{\Lambda}$ are the optimized Lagrange multipliers. Equation (11) is the maximum posterior estimate of equation (6). In Methods, we show that the estimate of the uncertainty s in the mean prediction is given by

$$s^2 = \sum_{k=1}^{B} \sum_{l=1}^{B} g_k g_l (\langle \delta\lambda_k \delta\lambda_l \rangle - \langle \delta\lambda_k \rangle \langle \delta\lambda_l \rangle). \tag{12}$$

In equations (12) $\delta\lambda_k = \lambda_k - \Lambda_k$ and $\langle \delta\lambda_k \delta\lambda_l \rangle - \langle \delta\lambda_k \rangle \langle \delta\lambda_l \rangle$ is the corresponding covariance matrix among Lagrange multipliers. In equation (12), the "couplings" g_k are given by

$$g_k = \langle I_k \left(x_a(t, \bar{\theta}) \right) Y(\bar{\theta}) \rangle - \psi_k(\bar{\Lambda}) y(\bar{\Lambda}). \tag{13}$$

g_k quantifies the correlation between the quantity of interest $Y(\bar{\theta})$ and the constraints $I_k \left(x_a(t, \bar{\theta}) \right)$. In Methods we show that

$$\langle \delta\lambda_k \delta\lambda_l \rangle - \langle \delta\lambda_k \rangle \langle \delta\lambda_l \rangle = \left((c^+) diag\left(\overline{\sigma^2}\right) (c^+)^T \right)_{kl}. \tag{14}$$

In equation (14) c^+ is the Moore-Penrose pseudo-inverse of the covariance matrix c. c is given by:

$$c_{kl} = \langle I_k \left(x_a(t, \bar{\theta}) \right) I_l \left(x_a(t, \bar{\theta}) \right) \rangle - \psi_k(\bar{\Lambda}) \psi_l(\bar{\Lambda}) \tag{15}$$

In equations (12-15), the averages denoted by angular brackets are calculated using equation (5) with the Lagrange multipliers $\bar{\Lambda}$ fixed at their optimized values.

To use the aforementioned approach in our calculations, we first estimate mean bin fractions $\bar{\phi}$ and the corresponding standard errors in the mean $\bar{\sigma}$ from

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

experimental data. Then, by using the iterative scheme described in Figure 2, we estimate the optimized Lagrange multipliers $\bar{\Lambda}$ in equation (5) and the corresponding predictions $\bar{\psi}(\bar{\Lambda})$. We then compute the covariance matrix c and its pseudo-inverse c^{+} (equation 15), and the covariance among the Lagrange multipliers (equation 14). Finally, for any quantity of interest $Y(\bar{\theta})$, we compute the couplings $\bar{g}$ (equation 13). The predicted mean value and the associated uncertainty are given by equation (11) and equation (12) respectively.

MERIDIAN performance on synthetic data

First, we use synthetic data to illustrate the utility of MERIDIAN and compare its performance with a previously developed discretized Bayesian (DB) approach by Hasenauer et al. (Hasenauer, Waldherr et al. 2011). In DB, a crucial step is to approximate the joint parameter distribution as a linear combination

$$P(\bar{\theta}) = \sum \gamma_j \Phi_j(\bar{\theta}) \tag{16}$$

where $\Phi_i(\bar{\theta})$ are a set of predefined distribution functions and $\gamma_i > 0$ are the corresponding weights. In principle, if a sufficiently large number of such basis functions are chosen, their linear combination can capture any complex probability distribution. DB discretizes the multidimensional parameter space using a Cartesian grid and assumes that $\Phi_i(\bar{\theta})$ are multivariate Gaussian distributions centered at grid points. The weights γ_i are sampled from their posterior distribution given the data. Because several other methods (Hasenauer, Waldherr et al. 2011, Hasenauer, Hasenauer et al. 2014, Loos, Moeller et al. 2018) also use linear combination of known functions to approximate the parameter distributions, a comparison with DB highlights the potential advantages of MERIDIAN.

To compare MERIDIAN with DB, we used a simplified growth factor network model (Methods, Supplemental Information Figure 1). Specifically, the network included three chemical species: the ligand *L*, ligand-free inactive receptors *R*, and ligand bound active receptors *P*. Ligand binding to receptors leads to their activation. Inactive receptors are constantly delivered to the cell

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

surface, and active and inactive receptors are removed from the cell surface at different rates (Methods). In this model, the dynamics of the species was described by five parameters.

Using the aforementioned model, we generated synthetic single cell data by varying two key parameters: (1) rate of degradation of inactive receptors k_{deg} and (2) rate of degradation of active receptors k^*_{deg}. Notably, it was demonstrated that the processes that govern receptor trafficking on the cell surface are correlated at the single cell level (Kallenberger, Unger et al. 2017). To account for these correlations, we sampled the two parameters from a correlated bivariate exponential distribution (Iyer 2002). Using the joint parameter distribution we generated distribution of activated receptor levels corresponding to four different experimental conditions (ligand *L* = 2 ng/ml, *L* = 10 ng/ml, t = 10 minutes and steady state, Methods). Normally distributed random errors were then added to individual cell measurements. Mean bin fractions and standard errors in bin fractions were obtained using 5 separate synthetic datasets. These data were then used to infer the joint distribution $P\left(k_{deg}, k^*_{deg}\right)$ using both MERIDIAN and DB.

There are two types of hyper-parameters in DB: (1) number of Cartesian grid points and (2) the width of the multivariate Gaussian distributions. We performed DB-inference using several sets of these hyper-parameters. For each set, we first found the maximum likelihood weights Γ_j. Next, we determined the optimal hyper-parameters using the Akaike information criterion (Akaike 1973). Using the optimal hyper-parameters, we sampled the posterior distribution over the weights γ_j and estimated the joint distribution over parameters $P\left(k_{deg}, k^*_{deg}\right)$ (Methods).

Notably, both MERIDIAN and DB were able to accurately fit the synthetic data (Methods, Supplemental Information Figure 2). The two approaches were also similar in capturing the Pearson correlation between the two rates ($\rho_{DB} \sim 0.19$, $\rho_{MERIDIAN} \sim 0.18$, and $\rho_{Data} \sim 0.3$). However, they predicted substantially different parameter distributions. While MERIDIAN accurately captured both the

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

distributions $P(k_{deg})$ and $P(k^*_{deg})$, the distribution inferred using DB were very different from exponential distributions (Figure 4).

Next, we obtained the inferred parameter distributions using DB for several sets of inference hyper-parameters by sampling over the posterior distribution over the weights (Supplemental Information Table 4). We quantified the similarity between the inferred and true distributions using three different cumulative-based metrics. Notably, MERIDIAN outperformed a vast majority of the tested choices of DB hyper-parameters (more than 93%) (Methods, Supplemental Information Table 4, and Supplemental Information Figure 3).

Finally, we note that applying the DB approach to realistic models of biological networks (with 10-30 parameters) may be computationally prohibitive. If we choose N_G grid points per dimension in an inference of a K dimensional parameter space, DB employs N_G^K basis functions. For example, for a signaling network described by 20 parameters, employing 5 grid points per dimension will require ~10^{14} basis functions.

Using MERIDIAN to study EGFR/Akt signaling

Computational model of the EGFR/Akt signaling network

Signal transduction in the EGFR/Akt network is illustrated in Figure 4. Following stimulation of cells with EGF, the ligand binds to the cell surface EGFRs. Ligand-bound receptors dimerize with other ligand-bound receptors as well as ligand-free receptors. EGFR dimers then phosphorylate each other and phosphorylated receptors (active receptors, pEGFRs) on the cell surface lead to downstream phosphorylation of Akt (pAkt). Both active and unphosphorylated (inactive) receptors are internalized with different rates from the cell surface through receptor endocytosis. After addition of EGF in the extracellular medium, pAkt levels increase transiently within minutes and then, as a result of receptor

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

endocytosis and action of phosphatases, both pAkt and surface EGFR (sEGFR) levels decrease within hours after EGF stimulation (Chen, Schoeberl et al. 2009).

To explore the cell-to-cell variability in this pathway, we used a dynamical model of EGF/EGFR dependent Akt phosphorylation based on Chen et al. (Chen, Schoeberl et al. 2009). The model includes reactions describing EGF binding to EGFR and subsequent dimerization, phosphorylation, dephosphorylation, internalization, and degradation of receptors. To keep the model relatively small, we simplified pEGFR-dependent phosphorylation of Akt by assuming a single step activation of Akt by pEGFR (Methods). We note that the first and second order rate constants employed the model should be treated as effective rates given that the law of mass action is only an approximation to describe the complex processes in the EGFR/Akt pathway. The model had 17 chemical species and 20 parameters. See Supplemental Information Table 2 for a list of model parameters and Supplemental Information Table 3 for a list of model variables. The model equations are given in Methods.

Numerical inference of the parameter distribution

To estimate the signaling parameter distribution consistent with experimental data using MERIDIAN, we used experimentally measured cell-to-cell variability in pAkt levels at early times after EGF stimulation. Specifically, we used measured pAkt levels after stimulation with five different EGF doses (0.1, 0.316, 3.16, 10, and 100 ng/ml) at 4 early time points (5, 15, 30, and 45 minutes) (Lyashenko, Niepel et al. 2017). Additionally, we used sEGFR levels without EGF stimulation and after 3 hours of EGF stimulation at 1 ng/ml. We used 11 bins to represent each experimentally measured distribution; the bin sizes and locations were chosen to cover the entire range of observed variability (Supplemental Information Table 1). There were a total of 264 bin fractions and corresponding 264 Lagrange multipliers. We numerically determined the optimal Lagrange multipliers corresponding to the bin fractions using the procedure described above (see Figure 2, Methods). It took approximately 90 hours to learn the optimal Lagrange multipliers.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Notably, the optimal Lagrange multipliers accurately reproduced the experimentally measured bin fractions (Pearson's r^2 = 0.9, $p < 10^{-10}$, median relative error ~ 14%). Furthermore, fitted bin fractions obtained in two independent calculations showed excellent agreement with each other as expected for a convex optimization problem (Pearson r^2 = 0.99, $p < 10^{-10}$, Supplemental Information Figure 4). In Figure 5, we show the temporal profile of measured cell-to-cell variability in pAkt levels (colored circles) at EGF stimulation of 10 ng/ml and the corresponding fits (dashed black lines) based on the inferred parameter distribution. The fits to all 24 distributions are given in Supplemental Information Figure 5. The marginal distributions of the individual parameters are given in Supplemental Information Figure 6 and the correlation structure amongst parameters is given in Supplemental Information Table 7.

Prediction of single cell dynamics

Akt is a key hub of mammalian cell signaling (Manning and Toker 2017). Naturally, sustained activity of phosphorylated Akt (pAkt) is implicated in diverse human diseases, such as psychiatric disorders (Gilman, Chang et al. 2012) and cancer (Vivanco and Sawyers 2002). Using the developed approach, we investigated whether we could predict pAkt levels hours after EGF stimulation using the parameter distribution inferred from pAkt variability at early times after EGF stimulation. To that end, we numerically sampled multiple parameter sets using the inferred parameter distribution and predicted pAkt levels at late time across a range of EGF stimulation levels corresponding to each parameter set. We compared predicted and experimentally observed distribution of pAkt levels across cells at late times (180 minutes) after sustained EGF stimulation (Figure 6a,b, Supplemental Information Figure 7). Our simulations correctly predicted that a significant fraction of cells have high pAkt levels hours after stimulation; the predicted and observed coefficient of variation (CV) of the pAkt distributions in cells stimulated with 10 ng/ml EGF for 180 minutes were in good agreement, 0.41 and 0.37 respectively. Notably, the inferred parameter distribution accurately captures the population mean and variability (Figure 6c) in pAkt levels

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

at late times across four orders of magnitude of EGF concentrations used to stimulate cells.

Importantly, MERIDIAN allowed us to investigate the biochemical parameters that significantly correlate with high pAkt levels at steady state. Interestingly, across all simulated trajectories, the levels of cell surface EGFR showed the highest correlation with pAkt levels among all receptor-related parameters (Supplemental Information Table 6, Pearson r = 0.4, EGF stimulation 10 ng/ml). This suggests that cells with high EGFR levels likely predominantly contribute to the sub-population of cells with high steady state pAkt activity. This demonstrates how MERIDIAN can be used to gain mechanistic insight into heterogeneity in signaling dynamics based on single cell data.

We next investigated whether MERIDIAN could predict the heterogeneity in EGFR levels after prolonged stimulation with EGF. To that end, we compared the predicted and the experimentally measured the steady state EGFR levels across EGF stimulation doses. Similar to pAkt, the simulations accurately captured both the population mean and variability of the EGFR receptor levels across multiple doses of EGF stimulations (Figure 7c). The simulations and experiments demonstrate that in agreement with model prediction that even hours after the growth factor stimulation there is a significant fraction of cells with relatively high levels of EGFR (Figure 7a,b).

Possible extensions of the MERIDIAN framework

Using MERIDIAN with inherently stochastic networks

A straightforward extension makes it possible to use the MERIDIAN framework for signaling networks when the time evolution of species abundances is intrinsically stochastic, for example, transcriptional networks and prokaryotic signaling networks with relatively small species abundances (Raj and van Oudenaarden 2008). To that end, we can modify the definition of the predicted bin fraction $\psi_k = \int P\left(x(t,\bar{\theta}) = x \middle| \bar{\theta}\right) d\bar{\theta}$ where $P\left(x(t,\bar{\theta}) = x \middle| \bar{\theta}\right)$ is the distribution of x values at time t with parameters $\bar{\theta}$. The distribution can be obtained numerically

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

using Gillespie's stochastic simulation algorithm (Gillespie 2007) and its fast approximations (Cao and Grima 2018) or approximated using moment closure techniques (Gillespie 2009). Notably, we have previously implemented this generalization of MERIDIAN to understand intrinsic and extrinsic noise in a simple gene expression circuit in *E. coli* (Dixit 2013).

Constraining moments in MERIDIAN

MERIDIAN can also be used to infer parameter distributions when, instead of the entire abundance distributions only a few moments of the distribution are available, such as average protein abundances measured using quantitative western blots or mass spectrometry (Shi, Niepel et al. 2016). For example, we consider the case where the population mean *m* and the variance *v* of one species *x* are measured at a fixed time point *t.* Instead of constraining fractions ψ_k that represent cell-to-cell variability in different bins of the relevant abundance distribution, we can constrain the population mean $\mu_1 = \int x(t,\bar{\theta})P(\bar{\theta})\,d\bar{\theta}$ and the second moment $\mu_2 = \int x(t,\bar{\theta})^2 P(\bar{\theta})\,d\bar{\theta}$ to their experimentally measured values *m* and $v+m^2$ respectively. Entropy maximization can then be carried out with these constraints. In this case, we have

$$P(\bar{\theta}) = \frac{1}{\Omega} q(\bar{\theta}) \exp(-\lambda_1 x(t,\bar{\theta}) - \lambda_1 x(t,\bar{\theta})^2) \tag{17}$$

Using MERIDIAN with high-dimensional data

MERIDIAN can be used to infer parameter distributions when multiple chemical species are experimentally measured in single cells at the same experimental condition. It may be difficult to accurately estimate the multidimensional bin counts from multidimensional data. Therefore, one can apply the following approach. For example, two species *x* and *y* simultaneously are measured across several cells, in addition to constraining the one-dimensional bin fractions $\bar{\phi}^{(x)}$ and $\bar{\phi}^{(y)}$, we can also constrain the cross-moment

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$r = \langle xy \rangle$. With these three types of constraints, the maximum entropy distribution is given by

$$P(\bar{\theta}) = \frac{1}{\Omega} q(\bar{\theta}) \exp\left(-\sum_{k=1}^{B_x} \lambda_k I_k^{(x)}(x) - \sum_{k=1}^{B_y} \kappa_k I_k^{(y)}(y) - \tau xy \right). \qquad (18)$$

In equation (18), $I_k^{(x)}$ and I_k^{y} are the indicator functions for species *x* and *y* respectively, B_x and B_y are the number of bins used in the *x*- and the *y*-dimension, $\bar{\lambda}$ and $\bar{\kappa}$ are Lagrange multipliers constraining the bin fractions $\bar{\phi}^{(x)}$ and $\bar{\phi}^{(y)}$ respectively, and τ is the Lagrange multiplier that constrains the cross-moment. By adding cross-moment constraints for each pair of species, equation (18) can be easily generalized to multiple dimensions; this will add ~$N^2/2$ Lagrange multipliers where *N* is the number of measured species.

Speeding up MERIDIAN inference using neural networks

A key numerical bottleneck in applying the MERIDIAN inference approach is the numerical optimization of a large number of Lagrange multipliers. It is a well-known problem in maximum entropy inference (Loaiza-Ganem, Gao et al. 2017). To address this problem, recently Loaiza-Ganem et al. (Loaiza-Ganem, Gao et al. 2017) proposed a maximum entropy flow network approach based on approximate deep generative modeling. Briefly, instead of finding the continuous density of the maximum entropy distribution in equation (5), they find the approximate maximum entropy distribution within a parametric family. The family is parameterized by several layers of a neural network and was shown to be sufficiently accurate in approximating true maximum entropy distributions. Moreover, a recent extension of this approach (Bittner and Cunningham 2019) enables fast simultaneous sampling of maximum entropy distributions along with a distribution of Lagrange multipliers (equation (7)). We believe that these fast methods will be crucial when using the MERIDIAN approach to study large signaling networks with several experimentally measured single cell distributions.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Using MERIDIAN with live cell imaging data

Finally, we discuss how we can use MERIDIAN to infer parameters from experiments where dynamics of species abundances within single cells are measured using live cell imaging (Meyer, D'Alessandro et al. 2012, Kallenberger, Unger et al. 2017). For example, consider that the time evolution of a species $x(t)$ is measured in n_c cells from time t = 0 to t = T. We can discretize the continuous time observations into K discrete times $\{t_1, t_2, \ldots, t_k\}$. At each time point t_i, we can then divide the range of observed abundances in B_i bins. Then, each individual dynamical trajectory $x(t)$ can be characterized by a vector of indices $x(t) \sim \{B_{1a_1}, B_{2a_2}, \ldots, B_{ka_k}\}$ where B_{ia_i} is the index of the abundance distribution bin through which the trajectory $x(t)$ passed at time point t_i. Given a sufficiently large number of trajectories, we then can constrain the fraction of trajectories that populate a given sequence of bins to infer the parameter distribution.

Discussion

Cells in a population exhibit heterogeneity in part because of heterogeneity in signaling network parameters (Albeck, Burke et al. 2008, Niepel, Spencer et al. 2009, Spencer, Gaudet et al. 2009, Meyer, D'Alessandro et al. 2012, Llamosi, Gonzalez-Vargas et al. 2016, Kallenberger, Unger et al. 2017). In this work, we developed a maximum entropy based approach to infer this parameter heterogeneity from single cell measurements of chemical species abundances. Notably, the inferred distribution combines two components: (1) the true biological parameter variability due to cell-to-cell heterogeneity and (2) the non-identifiability in parameter estimation given the single cell data. Consequently, the inferred distribution is likely to be broader compared to the true biological variability (Mukherjee, Seok et al. 2013). Notably, the non-identifiability component can be further minimized by (1) optimally designing experimental conditions to reduce non-identifiability (Bandara, Schloder et al.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2009, Kreutz and Timmer 2009) or by (2) directly including constraints on population average measurements of rate constants and other parameters of the signaling network.

We briefly discuss key differences between MERIDIAN and a previous maximum entropy based approach by Waldherr et al. (Waldherr, Hasenauer et al. 2009). Waldherr et al. employ the so-called Latin hypercube sampling (LHS) approach (Stein 1987). A potential advantage of LHS is that it avoids computationally expensive determination of the Lagrange multipliers. At the same time, LHS only sparsely samples the parameter space and generally cannot assign probabilities to arbitrary high dimensional parameter points. In contrast, an advantage of MERIDIAN is that the continuous density defined in equation (5) allows us to estimate the relative probability of any parameter point. Finally, unlike the LHS approach, MERIDIAN allows us to estimate the uncertainty in model predictions using measurement errors.

Recent developments in cytometry (Chattopadhyay, Gierahn et al. 2014) and single cell RNA sequencing (Saliba, Westermann et al. 2014) make it possible to simultaneously quantify multiple species abundances in single cells. Elegant statistical approaches have been developed to reconstruct trajectories of intracellular species dynamics consistent with time-stamped single cell abundance data (Gut, Tadmor et al. 2015, Mukherjee, Jensen et al. 2017, Mukherjee, Stewart et al. 2017). Complementary to these statistical methods, our approach (1) allows us to infer the distribution over signaling parameters that describe mechanistic interactions in the signaling network and moreover (2) the inferred parameter distribution can be used to predict the ensemble of single cell trajectories for time intervals and experimental conditions beyond the measured abundance distributions.

In this work, we applied the developed framework to signaling network data. However, it can also be used in other diverse research contexts. For example, the framework can be applied to computationally reconstruct the distribution of longitudinal behaviors from cross-sectional time-snapshot data in

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

fields such as public health, economics, and ecology or to estimate parameter distributions from lower dimensional statistics (Das, Mukherjee et al. 2015).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR. B. N. a. C. Petrov, F. Budapest.

Albeck, J. G., J. M. Burke, B. B. Aldridge, M. Zhang, D. A. Lauffenburger and P. K. Sorger (2008). "Quantitative analysis of pathways controlling extrinsic apoptosis in single cells." Mol Cell **30**(1): 11-25.

Bandara, S., J. P. Schloder, R. Eils, H. G. Bock and T. Meyer (2009). "Optimal experimental design for parameter estimation of a cell signaling model." PLoS Comput Biol **5**(11): e1000558.

Banks, H. a. K., ZR and Thompson WC (2012). "A review of selected techniques in inverse problem nonparametric probability distribution estimation." Journal of Inverse and Ill-Posed Problems **20**(4): 429-460.

Bertsimas, D. a. T. J. (1997). Introduction to linear optimization, Athena Scientific.

Bittner, S. and J. Cunningham (2019). Approximating exponential family models (not single distributions) with a two-network architecture. arXiv/1903.07515.

Cao, Z. and R. Grima (2018). "Linear mapping approximation of gene regulatory networks with stochastic dynamics." Nat Commun **9**(1): 3305.

Caticha, A. and R. Preuss (2004). "Maximum entropy and Bayesian data analysis: Entropic prior distributions." Phys Rev E Stat Nonlin Soft Matter Phys **70**(4 Pt 2): 046127.

Chattopadhyay, P. K., T. M. Gierahn, M. Roederer and J. C. Love (2014). "Single-cell technologies for monitoring immune systems." Nat Immunol **15**(2): 128-135.

Chen, W. W., B. Schoeberl, P. J. Jasper, M. Niepel, U. B. Nielsen, D. A. Lauffenburger and P. K. Sorger (2009). "Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data." Mol Syst Biol **5**: 239.

Chung, J. C., N. Sciaky and D. J. Gross (1997). "Heterogeneity of epidermal growth factor binding kinetics on individual cells." Biophys J **73**(2): 1089-1102.

Das, J., S. Mukherjee and S. E. Hodge (2015). "Maximum Entropy Estimation of Probability Distribution of Variables in Higher Dimensions from Lower Dimensional Data." Entropy (Basel) **17**(7): 4986-4999.

Dixit, P. D. (2013). "Quantifying extrinsic noise in gene expression using the maximum entropy framework." Biophys J **104**(12): 2743-2750.

Dixit, P. D. (2018). "Communication: Introducing prescribed biases in out-of-equilibrium Markov models." J Chem Phys **148**(9): 091101.

Dixit, P. D., A. Jain, G. Stock and K. A. Dill (2015). "Inferring Transition Rates of Networks from Populations in Continuous-Time Markov Processes." J Chem Theory Comput **11**(11): 5464-5472.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dixit, P. D., J. Wagoner, C. Weistuch, S. Presse, K. Ghosh and K. A. Dill (2018). "Perspective: Maximum caliber is a general variational principle for dynamical systems." J Chem Phys **148**(1): 010901.

Eydgahi, H., W. W. Chen, J. L. Muhlich, D. Vitkup, J. N. Tsitsiklis and P. K. Sorger (2013). "Properties of cell death models calibrated and compared using Bayesian approaches." Mol Syst Biol **9**: 644.

Gillespie, C. S. (2009). "Moment-closure approximations for mass-action models." IET Syst Biol **3**(1): 52-58.

Gillespie, D. T. (2007). "Stochastic simulation of chemical kinetics." Annu Rev Phys Chem **58**: 35-55.

Gilman, S. R., J. Chang, B. Xu, T. S. Bawa, J. A. Gogos, M. Karayiorgou and D. Vitkup (2012). "Diverse types of genetic variation converge on functional gene networks involved in schizophrenia." Nat Neurosci **15**(12): 1723-1728.

Gut, G., M. D. Tadmor, D. Pe'er, L. Pelkmans and P. Liberali (2015). "Trajectories of cell-cycle progression from fixed cell populations." Nat Methods **12**(10): 951-954.

Hasenauer, J., C. Hasenauer, T. Hucho and F. J. Theis (2014). "ODE constrained mixture modelling: a method for unraveling subpopulation structures and dynamics." PLoS Comput Biol **10**(7): e1003686.

Hasenauer, J., S. Waldherr, M. Doszczak, N. Radde, P. Scheurich and F. Allgower (2011). "Identification of models of heterogeneous cell populations from population snapshot data." BMC Bioinformatics **12**: 125.

Hazoglou, M. J., V. Walther, P. D. Dixit and K. A. Dill (2015). "Communication: Maximum caliber is a general variational principle for nonequilibrium statistical mechanics." J Chem Phys **143**(5): 051104.

Herbst, R. S. (2004). "Review of epidermal growth factor receptor biology." Int J Radiat Oncol Biol Phys **59**(2): 21-26.

Iyer, S. a. M., D and Manivasakan, R (2002). "Bivariate exponential distribution using linear structures." Sankhya : The Indian Journal of Statistics **64**(A): 156-166.

Kallenberger, S. M., A. L. Unger, S. Legewie, K. Lymperopoulos, U. Klingmuller, R. Eils and D. P. Herten (2017). "Correlated receptor transport processes buffer single-cell heterogeneity." PLoS Comput Biol **13**(9): e1005779.

Kreutz, C. and J. Timmer (2009). "Systems biology: experimental design." FEBS J **276**(4): 923-942.

Llamosi, A., A. M. Gonzalez-Vargas, C. Versari, E. Cinquemani, G. Ferrari-Trecate, P. Hersen and G. Batt (2016). "What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast." PLoS Comput Biol **12**(2): e1004706.

Loaiza-Ganem, G., Y. Gao and J. Cunningham (2017). Maximum Entropy Flow Networks. ICLR 2017.

Loos, C., K. Moeller, F. Frohlich, T. Hucho and J. Hasenauer (2018). "A Hierarchical, Data-Driven Approach to Modeling Single-Cell Populations Predicts Latent Causes of Cell-To-Cell Variability." Cell Syst **6**(5): 593-603 e513.

Lyashenko, E., M. Niepel, P. D. Dixit, L. S.K., P. K. Sorger and D. Vitkup (2017). "Receptor-based mechanism of relative sensing in mammalian signaling networks." bioRxiv: 10.1101/158774.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Manning, B. D. and A. Toker (2017). "AKT/PKB Signaling: Navigating the Network." Cell **169**(3): 381-405.

Meyer, R., L. A. D'Alessandro, S. Kar, B. Kramer, B. She, D. Kaschek, B. Hahn, D. Wrangborg, J. Karlsson, M. Kvarnstrom, M. Jirstrand, W. D. Lehmann, J. Timmer, T. Hofer and U. Klingmuller (2012). "Heterogeneous kinetics of AKT signaling in individual cells are accounted for by variable protein concentration." Front Physiol **3**: 451.

Mora, T., A. M. Walczak, W. Bialek and C. G. Callan, Jr. (2010). "Maximum entropy models for antibody diversity." Proc Natl Acad Sci U S A **107**(12): 5405-5410.

Mukherjee, S., H. Jensen, W. Stewart, D. Stewart, W. C. Ray, S. Y. Chen, G. P. Nolan, L. L. Lanier and J. Das (2017). "In silico modeling identifies CD45 as a regulator of IL-2 synergy in the NKG2D-mediated activation of immature human NK cells." Sci Signal **10**(485).

Mukherjee, S., S. C. Seok, V. J. Vieland and J. Das (2013). "Cell responses only partially shape cell-to-cell variations in protein abundances in Escherichia coli chemotaxis." Proc Natl Acad Sci U S A **110**(46): 18531-18536.

Mukherjee, S., D. Stewart, W. Stewart, L. L. Lanier and J. Das (2017). "Connecting the dots across time: reconstruction of single-cell signalling trajectories using time-stamped data." Royal Society Open Science **4**(8).

Nicholson, K. M. and N. G. Anderson (2002). "The protein kinase B/Akt signalling pathway in human malignancy." Cell Signal **14**(5): 381-395.

Niepel, M., S. L. Spencer and P. K. Sorger (2009). "Non-genetic cell-to-cell variability and the consequences for pharmacology." Curr Opin Chem Biol **13**(5-6): 556-561.

Raj, A. and A. van Oudenaarden (2008). "Nature, nurture, or chance: stochastic gene expression and its consequences." Cell **135**(2): 216-226.

Saliba, A. E., A. J. Westermann, S. A. Gorski and J. Vogel (2014). "Single-cell RNA-seq: advances and future challenges." Nucleic Acids Res **42**(14): 8845-8860.

Schneidman, E., M. J. Berry, 2nd, R. Segev and W. Bialek (2006). "Weak pairwise correlations imply strongly correlated network states in a neural population." Nature **440**(7087): 1007-1012.

Shalek, A. K., R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park and A. Regev (2013). "Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells." Nature **498**(7453): 236-240.

Shi, T., M. Niepel, J. E. McDermott, Y. Gao, C. D. Nicora, W. B. Chrisler, L. M. Markillie, V. A. Petyuk, R. D. Smith, K. D. Rodland, P. K. Sorger, W. J. Qian and H. S. Wiley (2016). "Conservation of protein abundance patterns reveals the regulatory architecture of the EGFR-MAPK pathway." Sci Signal **9**(436): rs6.

Soule, H. D., T. M. Maloney, S. R. Wolman, W. D. Peterson, Jr., R. Brenz, C. M. McGrath, J. Russo, R. J. Pauley, R. F. Jones and S. C. Brooks (1990). "Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10." Cancer Res **50**(18): 6075-6086.

Spencer, S. L., S. Gaudet, J. G. Albeck, J. M. Burke and P. K. Sorger (2009). "Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis." Nature **459**(7245): 428-432.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stein, M. (1987). "Large Sample Properties of Simulations Using Latin Hypercube Sampling." Technometrics **29**(2): 143-151.

Tiwary, P. and B. J. Berne (2016). "Spectral gap optimization of order parameters for sampling complex molecular systems." Proc Natl Acad Sci U S A **113**(11): 2839-2844.

Tkacik, G., E. Schneidman, M. J. Berry, 2nd. and W. Bialek (2006). "Ising models for networks of real neurons." arXiv **q-bio/0611072**.

Vivanco, I. and C. L. Sawyers (2002). "The phosphatidylinositol 3-Kinase AKT pathway in human cancer." Nat Rev Cancer **2**(7): 489-501.

Waldherr, S. (2018). "Estimation methods for heterogeneous cell population models in systems biology." J R Soc Interface **15**(147).

Waldherr, S., J. Hasenauer and F. Allgower (2009). "Estimation of biochemical network parameter distributions in cell populations." Proc. 15th IFAC Symp Syst Ident(11265-1270).

Weigt, M., R. A. White, H. Szurmant, J. A. Hoch and T. Hwa (2009). "Identification of direct residue contacts in protein-protein interaction by message passing." Proc Natl Acad Sci U S A **106**(1): 67-72.

Wu, M. and A. K. Singh (2012). "Single-cell protein analysis." Curr Opin Biotechnol **23**(1): 83-88.

Zechner, C., J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros and H. Koeppl (2012). "Moment-based inference predicts bimodality in transient gene expression." Proc Natl Acad Sci U S A **109**(21): 8340-8345.

Zechner, C., M. Unger, S. Pelet, M. Peter and H. Koeppl (2014). "Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings." Nat Methods **11**(2): 197-202.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends

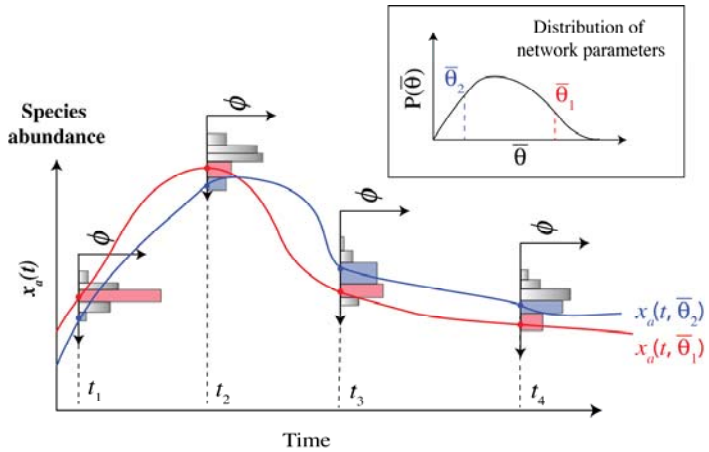


Figure 1. Illustration of the MERIDIAN inference approach. Cell-to-cell variability in protein "*a*" is measured at four time points t_1, t_2, t_3, and t_4. From the single cell data, we determine the fraction ϕ_{ik} of cells that populate the kth abundance bin in the ith experiment. The histograms show ϕ_{ik} at multiple experimental conditions. We find $P(\bar{\theta})$ with the maximum entropy while requiring that the corresponding distribution $P[x_a(t,\bar{\theta})]$ over trajectories of $x_a(t,\bar{\theta})$ simultaneously reproduces all bin fractions.

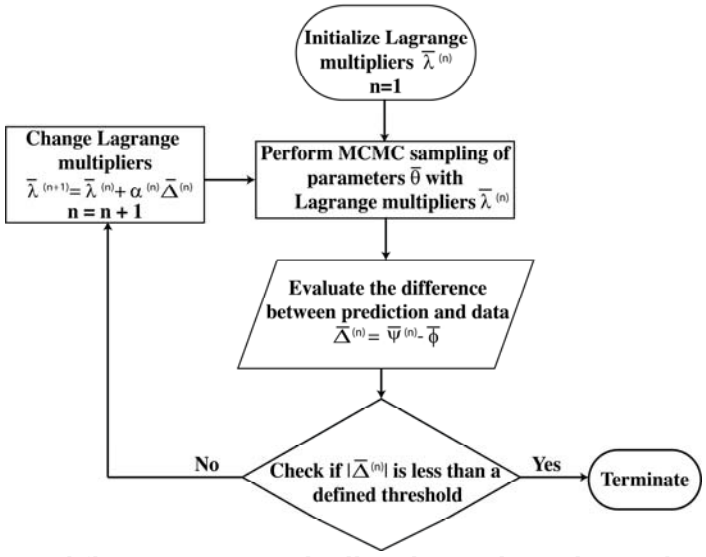


Figure 2. The workflow to numerically determine the values of Lagrange multipliers. In each iteration, we evaluate the error vector $\bar{\Delta}^{(n)}$ between the predicted bin fractions $\bar{\Delta}^{(n)}$ and the experimentally measured bin fractions using Markov chain Monte Carlo. We propose a new set of Lagrange multipliers based on the error vector. We repeat until the error reaches below a predefined accuracy cutoff.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

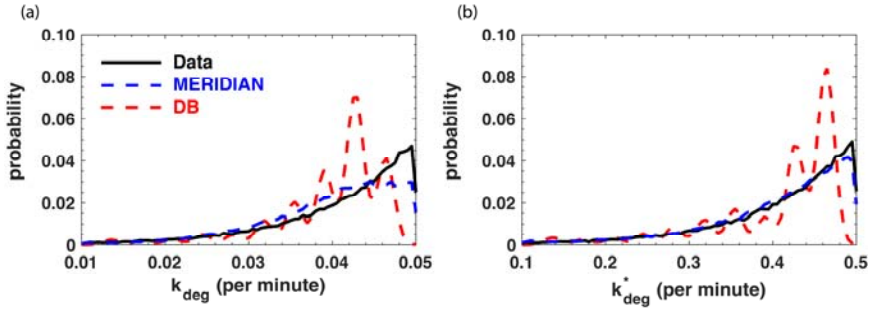


Figure 3. A comparison between the inferred parameter distributions and the data. We show the comparison between the true parameter distributions in panel a) and in panel b) (black lines) and the corresponding MERIDIAN-inferred distribution (blue lines) and the DB inferred distribution (dashed red lines).

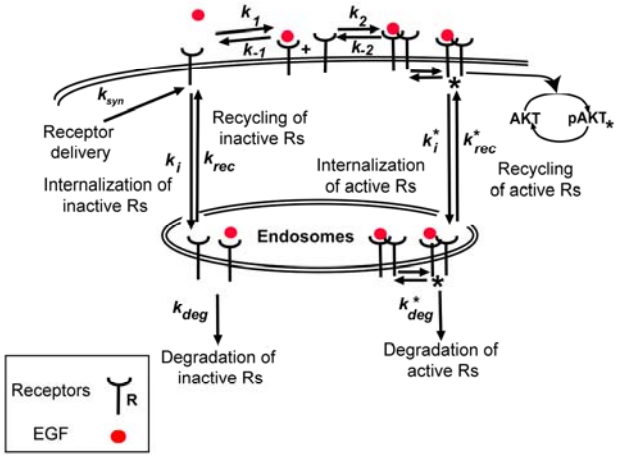


Figure 4. A schematic of the EGF/EGFR pathway leading to phosphorylation of Akt. Extracellular EGF binds to cell surface EGFRs leading to their dimerization. Dimerized EGFRs are autophosphorylated and in turn lead to phosphorylation of Akt. Receptors are also removed from cell surface through internalization into endosomes. See Methods for details of the model.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

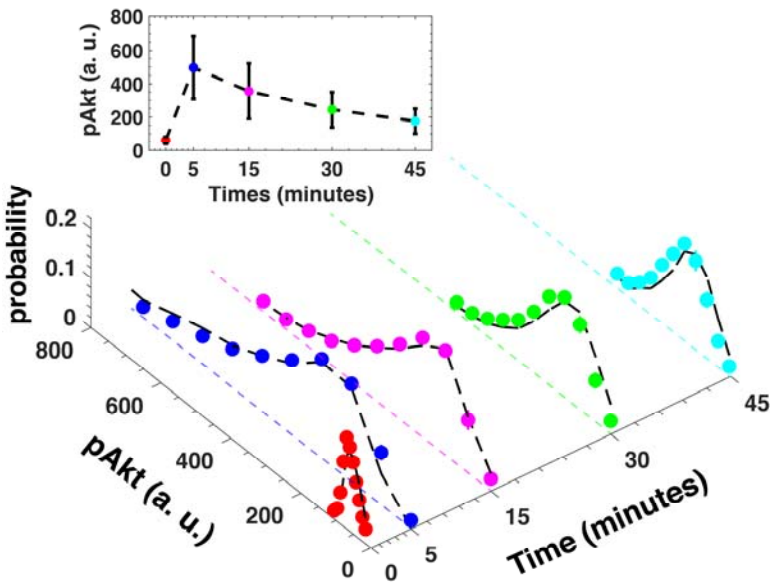


Figure 5. Experimental estimated cell-to-cell variability in pAkt levels used to infer the parameter distribution. We show distribution of pAkt levels at 0, 5, 15, 30, and 45 minutes after exposure to 10 ng/ml EGF. The colored circles represent the experimentally measured pAkt distributions used in the inference of the parameter distribution. The black dashed lines represent fitted distributions. The inset shows the experimentally measured population average pAkt levels at multiple time points. Error bars represent standard deviation. Error bars in the inset represent population standard deviations.

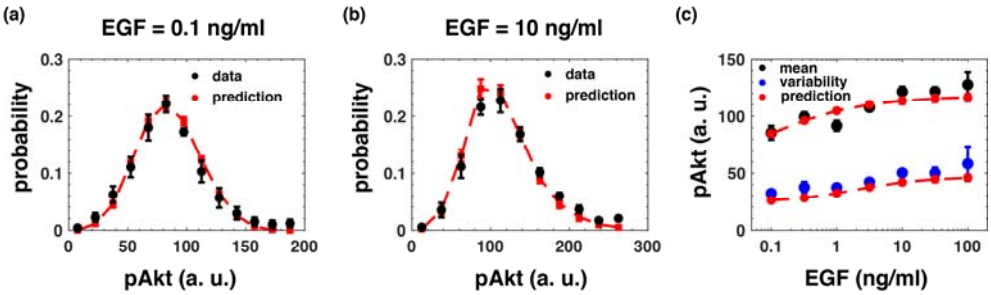


Figure 6. Prediction pAkt levels at late times. (a and b) Measured distributions (black circles and lines) and the corresponding predictions (red circles and dashed lines) of cell-to-cell variability in pAkt levels at 180 minutes after stimulation with 0.1 ng/ml and 10 ng/ml EGF respectively. (c) Measured mean pAkt levels (red circles) and measured standard deviation in pAkt levels (blue circles) at 180 minutes after sustained stimulation with EGF (x-axis) and the corresponding predictions (dashed red lines). The error bars in experimental data represent standard deviation. The error bars in model predictions represent the estimated uncertainty.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

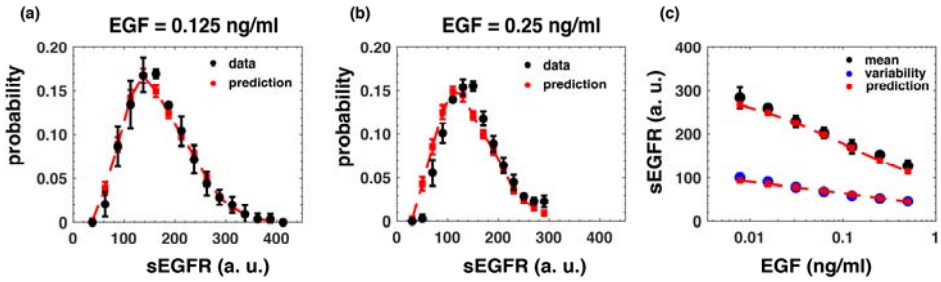


Figure 7. Prediction of sEGFR levels at late times. (a and b) Measured distributions (black circles and lines) and the corresponding predictions (red circles and dashed lines) of cell-to-cell variability in sEGFR levels at 180 minutes after stimulation with 0.125 ng/ml and 0.25 ng/ml EGF respectively. (c) Measured mean sEGFR levels (red circles) and measured standard deviation in sEGFR levels (blue circles) at 180 minutes after sustained stimulation with EGF and the corresponding predictions (dashed red lines). The error bars in experimental data represent standard deviation. The error bars in model predictions represent the estimated uncertainty.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Experimental Details

In this work, we used distributions of cell-to-cell variability in phosphorylated Akt levels as well as cell surface EGFR levels. We used the experimental data on pAkt levels previously measured in Lyashenko *et al*. (Lyashenko, Niepel et al. 2017) We measured cell-to-cell variability in sEGFR levels in this work. Briefly, we describe the experimental methods here.

MCF 10A cells (Soule, Maloney et al. 1990) were obtained from the ATCC. The cells were grown according to ATCC recommendations. We confirmed the cell identity by short tandem repeat (STR) profiling at the Dana-Farber Cancer Institute. We tested the cells with MycoAlert PKUS mycoplasma detection kit (Lonza) and ensured that they were free of mycoplasma infection. For the experiments, we coated 96 well plates (Thermo Fisher Scientific) with type I collagen from rat tail (Sigma-Aldrich) by incubating plates with 65 microliter of 4mg/ml collagen I solution in PBS for two hours at room temperature. We washed the plates twice with PBS using EL406 Microplate Washer Dispenser (BioTek) and sterilized them under UV light for 20 minutes prior to use. Cells were harvested during logarithmic growth. We dispensed 2500 cells per well into collagen-coated 96 well plates using a EL406 Microplate Washer Dispenser. We grew the cells in 200 microliter of complete medium for 24 hours. The cells were serum-starved twice in starvation media (DMEM/F12 lemented with 1% penicillin-streptomycin and 0.1% bovine serum albumin). Next, we incubated the cells in 200 microliter of starvation media for 19 hours and again for one more hour. This time point constituted t=0 for all experiments.

We created the EGF treatment solutions by dispensing the appropriate amounts of epidermal growth factor (EGF, Peprotech) into starvation media using a D300 Digital Dispenser (Hewlett-Packard). To fit the parameter distributions,

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

we used EGF concentrations of 0.1, 0.31, 3.16, 10, and 100 ng/ml for Akt phosphorylation measurements and EGF concentrations of 0, 1, and 100 ng/ml for surface EGFR measurements. At t=0 cells were stimulated with 100 microliter of 3x solution and incubated for indicated times (5, 15, 30, and 45 minutes for pAkt and 180 minutes for sEGFR). To test the model predictions, we collected pAkt distributions at 90 and 180 minutes after stimulation with 0.01, 0.031, 0.1, 0.31, 1, 3.16, 10, 31.6, and 100 ng/ml EGF. We also measured sEGFR distributions at 180 minutes after stimulation with 0.0078, 0.0156, 0.0312, 0.0625, 0.125, 0.25, 0.5, 1, and 100 ng/ml of EGF. All incubations were terminated by adding 100 µl of 12% formaldehyde solution (Sigma) in phosphate buffered saline (PBS) and fixing the cells for 30 min at room temperature.

We performed all subsequent washes and treatments with the EL406 Microplate Washer Dispenser. We washed the cells twice in PBS and permeabilized them with 0.3% Triton X-100 (Sigma-Aldrich) in PBS for 30 min at room temperature. Cells were washed once again in PBS, and blocked in 40 microliter of Odyssey blocking buffer (LI-COR Biotechnology) for 60 min at room temperature. Cells were incubated with 30 microliter of anti-phospho-Akt (Cell Signaling Technologies, #4060, 1:400) or anti-EGFR (Thermo Fisher Scientific, MA5-13319, 1:100) over night at 4°C. We then washed the cells once in PBS and three time in PBS with 0.1% Tween 20 (Sigma-Aldrich; PBS-T for 5 min each and incubated with 30 microliter of a 1:1000 dilution of Alexa Fluor 647 conjugated goat anti-rabbit or goat anti-mouse secondary antibody in Odyssey blocking buffer for 60 min at room temperature. Next we washed the cells two times in PBS-T, once with PBS, and stained for 30 min at room temperature with whole cell stain green (Thermo Fisher Scientific) and Hoechst (Thermo Fisher Scientific). Finally, cells were washed three times in PBS, covered in 200 microliter of PBS, and sealed for microscopy. We imaged cells with an Operetta high content imaging system (Perkin Elmer) and analyzed the resulting scans using the Columbus image data storage and analysis system (Perkin Elmer). We performed the experiments in biological triplicates for surface EGFR and

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

quadruplets for pAkt. To avoid potentially pathological bright cells, we removed the top 1% of the data in all single cell distributions.

Computational details

Incorporating experimental error in MERIDIAN

In this section, we expand on the mathematical details of how to incorporate experimental errors in the MERIDIAN inference procedure. As in the main text, we denote by $\bar{\phi}$ the experimentally estimated mean bin fractions and by $\bar{\sigma}$ the corresponding standard errors in the mean. We assume that following an iterative procedure described in Figure 2, we have obtained an optimal set of Lagrange multipliers $\bar{\Lambda}$. We denote by $\bar{\psi}(\bar{\Lambda})$ the corresponding model predicted bin fractions.

Any fixed set of Lagrange multipliers uniquely determines model predictions $\bar{\psi}$. Thus, the errors in experimental measurements are captured by a distribution over the Lagrange multipliers themselves. We write the probability of non-optimal Lagrange multipliers $\bar{\lambda} \neq \bar{\Lambda}$ as

$$P(\bar{\lambda}) \propto \exp\left(-\sum_{k=1}^{B} \frac{\left(\psi_k(\bar{\lambda}) - \phi_k\right)^2}{2\sigma_k^2}\right) \approx \exp\left(-\sum_{k=1}^{B} \frac{\left(\psi_k(\bar{\lambda}) - \psi_k(\bar{\Lambda})\right)^2}{2\sigma_k^2}\right) \quad (S1)$$

Equation (S1) assumes that the errors are normally distributed and that the residuals $\Delta_k = \psi_k(\bar{\Lambda}) - \phi_k$ are small. We have also neglected the Jacobian determinant associated with changing the variables from $\bar{\psi}(\bar{\lambda})$ to $\bar{\lambda}$. Sampling Lagrange multipliers from equation (S1) is in principle possible but may be numerically inefficient. This is because it requires on-the-fly estimation of predicted bin fractions $\psi_k(\bar{\lambda})$ for non-optimal Lagrange multipliers $\bar{\lambda} \neq \bar{\Lambda}$. However, if we are interested the first two moments (means and uncertainties), we can approximate the distribution over Lagrange multipliers as a multivariate Gaussian distribution. This is equivalent to assuming that the experimental errors σ_k are small compared to the mean values ϕ_k. In the EGFR/Akt data used in this

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

work, the standard errors in the mean are indeed small; median relative error is ~9% and the mean relative error is ~11%. To express the distribution in equation (S5) as a Gaussian, we first write

$$\psi_k(\bar{\lambda}) \approx \psi_k(\bar{\Lambda}) + \sum \left(\frac{d}{d\lambda_j} \psi_k(\bar{\lambda}) \right)_{\bar{\lambda}=\bar{\Lambda}} (\delta\lambda_j) \qquad (S2)$$

where $\delta\lambda_j = \lambda_j - \Lambda_j$ is the deviation in λ_j away from the optimal Lagrange multipliers. Using linear response theory from statistical physics (Hazoglou, Walther et al. 2015), the derivatives in equation (S2) can be expressed as ensemble average over the parameter space. We write

$$\left(\frac{d}{d\lambda_j} \psi_k(\bar{\lambda}) \right)_{\bar{\lambda}=\bar{\Lambda}} = \frac{d}{d\lambda_j} \frac{\int I_k\left(x_a(t,\bar{\theta})\right) q(\bar{\theta}) \exp\left(-\sum_{l=1}^{B} \Lambda_l I_l\left(x_a(t,\bar{\theta})\right)\right) d\bar{\theta}}{\int q(\bar{\theta}) \exp\left(-\sum_{l=1}^{B} \Lambda_l I_l\left(x_a(t,\bar{\theta})\right)\right) d\bar{\theta}} \qquad (S3)$$

$$\Rightarrow \left(\frac{d}{d\lambda_j} \psi_k(\bar{\lambda}) \right)_{\bar{\lambda}=\bar{\Lambda}} = -c_{jk} = -\left(\langle I_k\left(x_a(t,\bar{\theta})\right) I_j\left(x_a(t,\bar{\theta})\right) \rangle_{\bar{\lambda}=\bar{\Lambda}} - \psi_k(\bar{\Lambda})\psi_j(\bar{\Lambda}) \right) \qquad (S4)$$

In equation (S4), c_{jk} is the covariance matrix among the constraints. The average is computed using equation (5) in the main text with $\bar{\lambda} = \bar{\Lambda}$.

Combining equations (S1), (S2), and (S4), we obtain the Gaussian approximation to the distribution over Lagrange multipliers,

$$P(\bar{\lambda}) \propto \exp\left(-\sum_{k=1}^{B} \frac{\left(-\sum_j c_{jk}(\lambda_j - \Lambda_j)\right)^2}{2\sigma_k^2} \right). \qquad (S5)$$

The multivariate Gaussian distribution in equation (S5) is fully determined by the means and the covariance matrix of the Lagrange multipliers. We determine these next.

Since we assume that the model can fit the data reasonably accurately, the average value of the deviation in Lagrange multipliers in equation (S5) is $\langle \delta\lambda_j \rangle = 0$. Next, we estimate the covariance matrix among the Lagrange multipliers. Let us consider a particular bin fraction ϕ_k. The model estimated uncertainty is given by

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$s_k^2 = \int \psi_k(\bar{\lambda})^2 P(\bar{\lambda}) d\bar{\lambda} - \psi(\bar{\Lambda})^2 \tag{S6}$$

$$\approx \langle \left(\psi_k(\bar{\Lambda}) - \sum_j c_{jk}(\delta\lambda_j) \right)^2 \rangle - \psi(\bar{\Lambda})^2 \tag{S7}$$

$$= \sum_{j,l} c_{jk} c_{lk} \left(\langle \delta\lambda_j \delta\lambda_l \rangle - \langle \delta\lambda_j \rangle \langle \delta\lambda_l \rangle \right) \tag{S8}$$

The experimentally estimated uncertainty in ϕ_k is σ_k^2. Equating the two, we have

$$\sigma_k^2 = \sum_{j,l} c_{jk} c_{lk} \left(\langle \delta\lambda_j \delta\lambda_l \rangle - \langle \delta\lambda_j \rangle \langle \delta\lambda_l \rangle \right) \tag{S9}$$

$$\Rightarrow \langle \delta\lambda_j \delta\lambda_l \rangle - \langle \delta\lambda_j \rangle \langle \delta\lambda_l \rangle = \left((c^+) diag(\overline{\sigma^2}) (c^+)^T \right)_{jl} \tag{S10}$$

In equation (S10), c^+ is the pseudoinverse of the covariance matrix.

These first two moments fully describe the multivariate Gaussian distribution over Lagrange multipliers (equation (S5)). Next, we show how to estimate mean predictions and uncertainty in model predictions.

Consider a variable $Y(\bar{\theta})$ that depends on model parameters $\bar{\theta}$. We are interested in estimating its mean predicted value "m" and the corresponding uncertainty "s". Let us denote by $y(\bar{\lambda}) = \int Y(\bar{\theta}) P(\bar{\theta}|\bar{\lambda}) d\bar{\theta}$ the model prediction when the Lagrange multipliers are fixed at $\bar{\lambda}$. We have the mean prediction

$$m = \int y(\bar{\lambda}) P(\bar{\lambda}) d\bar{\lambda} \approx y(\bar{\Lambda}) \tag{S11}$$

Next, we seek the estimated uncertainty (see equation 10 in the main text),

$$s^2 = \int y(\bar{\lambda})^2 P(\bar{\lambda}) d\bar{\lambda} - m^2 \tag{S11}$$

$$\approx \langle \left(y(\Lambda) - \sum g_j \delta\lambda_j \right)^2 \rangle_{\bar{\lambda}=\bar{\Lambda}} - y(\bar{\Lambda})^2 \tag{S12}$$

$$= \sum_{j,l} g_j g_l \left(\langle \delta\lambda_j \delta\lambda_l \rangle - \langle \delta\lambda_j \rangle \langle \delta\lambda_l \rangle \right). \tag{S13}$$

In equation (S13), the couplings g_j are given by

$$g_j = \left(\langle Y(\bar{\theta}) I_j \rangle_{\bar{\lambda}=\bar{\Lambda}} - \langle Y(\bar{\theta}) \rangle_{\bar{\lambda}=\bar{\Lambda}} \psi_j(\bar{\Lambda}) \right). \tag{S14}$$

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Equations (S11-S14)) show how to estimate model predictions and the corresponding uncertainty from the parameter distribution $P(\bar{\theta}|\bar{\Lambda})$ (equation (5) in the main text).

In the theoretical development above, we restricted the Taylor series expansion to the first order in $\bar{\lambda}$. More generally, higher order Taylor series expansions can also be included. Notably, similar to equation (S4), all higher order Taylor series coefficients can be estimated using MCMC calculations performed using the parameter distribution $P(\bar{\theta}|\bar{\Lambda})$.

Comparison between DB and MERIDIAN

Simplified signaling network

In this section, we describe the details of the simple growth factor model used to generate synthetic data and the procedure to fit data to infer the parameter distribution using the maximum entropy approach as well as the discretized Bayesian approach of Hasenauer et al. (Hasenauer, Waldherr et al. 2011).

The model comprised three species: the extracellular ligand *L*, inactive ligand-free cell surface receptors *R*, and active ligand-bound cell surface receptors *P*. The dynamics of the three variables was dictated by five parameters: the rate of ligand binding to inactive receptors k_1, the rate of ligand unbinding from active receptors k_{-1}, the rate of removal of inactive receptors k_{deg}, the rate of removal of active receptors k^*_{deg}, and the steady state cell surface receptor level in the absence of the ligand R_T. See Supplemental Information Figure 1 for an illustration.

In the model, the dynamics of the species are governed by the following two equations:

$$\frac{d[R]}{dt} = k_{del} - k_1[L][R] + k_{-1}[P] - k_{deg}[R] \quad \text{(S15)}$$

$$\frac{d[P]}{dt} = k_1[L][R] - k_{-1}[P] - k^*_{deg}[P] \quad \text{(S16)}$$

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We simulated the model to generate synthetic data as follows. We fixed three parameters $k_1 = \frac{1}{min \cdot \frac{ng}{ml}}$, $k_{-1} = 10/min$, and $R_T = 5$. We generated single cell distribution by varying the rates k_{deg} and k^*_{deg}. The rate of delivery of receptors to the cell surface was set equal to $k_{del} = R_T k_{deg}$. The rates were sampled from a correlated bivariate exponential distribution (Iyer 2002). k_{deg} was restricted between 0.01/min and 0.05/min and k^*_{deg} was restricted between 0.1/min and 0.5/min. The Pearson correlation between the sampled rates was set to be 0.28. Using 50000 sampled rates, we solved the differential equations (S15) and (S16) and generated single cell data on the number of activated receptors at four conditions (*L* = 2 ng/ml and 10 ng/ml, and *t* = 10 minutes, and steady state). A normally distributed random error was added to each single cell readout *P* that represented 5% of the magnitude of the data point ($P \rightarrow P * (1 + 0.05X)$ where *X* was a standard normal variable). The 50000 single cells were split into five experiments of 10000 cells each. From these five experiments, we estimated the mean bin fractions $\bar{\phi}$ and the corresponding standard errors of the means $\bar{\sigma}$. There were a total of 44 bin fractions (Supplemental Information Table 4).

We used these bin fractions to infer the parameter distribution using MERIDIAN and a previously developed discretized Bayesian (DB) approach.

Implementation of DB

In the DB approach, we approximated the parameter distribution as a linear combination

$$P\left(k_{deg}, k^*_{deg}\right) = \sum \gamma_j \Phi_j\left(k_{deg}, k^*_{deg}\right) \qquad (S17)$$

where $\Phi_j\left(k_{deg}, k^*_{deg}\right)$ are bivariate Gaussian distributions situated on an equally spaced *N*x*N* Cartesian grid and γ_j are the corresponding weights. Here, *N* is a hyper-parameter of the inference procedure. In order to test the accuracy of the DB approach, we used multiple values of the number of grid points (*N* = 5, 10,

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

15, and 20). Another choice of hyper-parameter is the covariance matrix of the Gaussians. We tested the covariance matrix of the Gaussians as: $\Sigma = diag\left(\sigma^2_{k_{deg}}, \sigma^2_{k^*_{deg}}\right)$ where $\sigma_{k_{deg}} = \frac{1}{\beta} \times \frac{0.04}{N}$ and $\sigma_{k^*_{deg}} = \frac{1}{\beta} \times \frac{0.4}{N}$. The hyper-parameter β controls the width of the bivariate Gaussian distribution. We scanned over multiple values of the parameter β (β = 1, 2, 4, 8, 16, and 32). Overall, we performed 24 DB inference calculations (4 N values x 6 β values).

In any given calculation, for every value of the index *j* (indicating one Gaussian basis function) we sampled $N_{sim} = 10^4$ parameter points from the bivariate Gaussian distribution $\Phi_j\left(k_{deg}, k^*_{deg}\right)$ and estimated the predicted bin fractions $\overline{\Psi}_j$. For any set of weights $\{\gamma_j\}$, the overall predicted bin fractions for the ith bin is then expressed as the linear combination

$$\psi_i = \sum_j \gamma_j \Psi_{ij} \qquad (S18)$$

In order to decide which DB hyper-parameters to use for further analysis, we used the Akaike information criterion (AIC) (Akaike 1973). The AIC balances the accuracy with which a model fits the data and the number of parameters in the model. We determined the AIC for a given set of DB hyper-parameters as follows. First, we found the maximum likelihood set of weights by maximizing the log-likelihood

$$LL = -\sum_i \frac{1}{2\sigma_i^2}\left(\sum_j \gamma_j \Psi_{ij} - \phi_i\right)^2 \qquad (S19)$$

In equation (S19), ϕ_i is the mean "experimental" bin fractions in the ith experiment and σ_i is the corresponding standard error of the mean. Next, we calculated the AIC as AIC = $2N^2 - 2LL$ where N^2 was the total number of free parameters in each model (the N^2 weights). For further analysis, we used the set of hyper-parameters that minimized the AIC. This corresponded to N = 10 grid points per dimension and β = 4. The statistical characterization of all tested models is presented in Supplemental Information Table 4.

To account for the non-identifiability in the parameter distribution, we sampled the full posterior distribution over the weights using MCMC. The

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

posterior distribution was given by the exponential of the log likelihood in equation (S19). For this calculation, we used a total of 4×10^7 MCMC steps in the space of weights γ_j . The first half of the simulation was discarded as equilibration. In the second half of the MCMC, we stored weights every 10000th step. For each set of weights, we sampled 250 bivariate Gaussian variables by first sampling 50 grid points according to a multinomial distribution with the weights as probabilities and then sampling 5 bivariate Gaussian variables per sampled grid point.

Implementation of MERIDIAN

The optimal Lagrange multipliers $\bar{\Lambda}$ for MERIDIAN were inferred using an iterative procedure described below. In the iterative search, the predicted bin fractions $\psi_k(\bar{\lambda})$ were estimated using direct integration using the trapezoidal rule. We note that integration using trapezoidal rule is not possible for a high dimensional parameter space. In order to incorporate the errors in experimental data, we sampled over the posterior distribution over Lagrange multipliers. First, we approximated the posterior distribution as a multivariate Gaussian distribution (see equation (S5), (S9), and (S14)). Next, we sampled 5000 sets of Lagrange multipliers using this multivariate Gaussian distribution. For each set of Lagrange multipliers, we performed MCMC calculations using the Metropolis criterion (see below for details) to samples through the parameter space using equation (5) in the main text. Briefly, 10^5 steps of MCMC were used for equilibration and 10^4 steps were used to save parameters. Parameters were saved every 500th step.

Comparing MERIDIAN and DB

We used three different metrics to compare the marginal parameter distributions $P(k_{deg})$ and $P(k^*_{deg})$ predicted using MERIDIAN and DB. First, we collected 50000 parameter samples from the parameter distributions using both MERIDIAN and DB. Next, we binned these samples using intervals 0.01:0.0005:0.05 and 0.1:0.005:0.5 for k_{deg} and k^*_{deg} respectively using the hist

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

function in MATLAB. Using these bins, we found the cumulative distributions $F_M(k_{deg})$ and $F_M(k^*_{deg})$ respectively. Here, the subscript "M" denotes the method (MERIDIAN and DB). We defined three different errors between the predicted distributions and the ground truth exponential distributions (GT).

(1) $E_1 = 0.5 * \left(\left\|F_M(k_{deg}) - F_{GT}(k_{deg})\right\|_2 + \left\|F_M(k^*_{deg}) - F_{GT}(k^*_{deg})\right\|_2\right)$ is the average L_2 norm of the difference between the predicted cumulative distribution $F_M(\cdot)$ and the ground truth $F_{GT}(\cdot)$. This is the square root of the Cramér-von Mises metric.

(2) $E_2 = 0.5 * \left(\sum\left|F_M(k_{deg}) - F_{GT}(k_{deg})\right| + \sum\left|F_M(k^*_{deg}) - F_{GT}(k^*_{deg})\right|\right)$ is the average sum of the absolute difference between the predicted cumulative distribution $F_M(\cdot)$ and the ground truth $F_{GT}(\cdot)$

(3) $E_3 = 0.5 * \left(\max\left|F_M(k_{deg}) - F_{GT}(k_{deg})\right| + \max\left|F_M(k^*_{deg}) - F_{GT}(k^*_{deg})\right|\right)$ is the average of the maximum difference between the predicted cumulative distribution $F_M(\cdot)$ and the ground truth $F_{GT}(\cdot)$. This is the Kolmogorov-Smirnov metric.

The performance of DB and MERIDIAN based on these three errors can be found in Supplemental Information Table 4.

Applying MERIDIAN to EGFR/Akt pathway

Inference of Lagrange multipliers from data is convex

The entropy functional

$$S = -\int P(\bar{\theta}) \log P(\bar{\theta})\, d\bar{\theta} \qquad (S20)$$

is convex with respect to the probability distribution $P(\bar{\theta})$. Moreover, the constraints that impose normalization and bin fractions are linear with respect to the probability distribution and are thus convex with respect to $P(\bar{\theta})$ as well. Consequently, entire Lagrangian function (equation (4) of the main text)

$$L[P(\bar{\theta})] = S + \eta\left(\int P(\bar{\theta})\, d\bar{\theta} - 1\right) - \sum_{k=1}^{B} \lambda_k \left(\int I_k\left(x_a(t,\bar{\theta})\right) P(\bar{\theta}) d\bar{\theta} - \phi_k\right) \qquad (S21)$$

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is also convex. Let us consider the dual problem in the space of Lagrange multipliers. We substitute the maximum entropy probability distribution $P(\bar{\theta})$ from equation (5) of the main text. We have the dual

$$L[\bar{\lambda}] = \log \Omega\left(\bar{\lambda}\right) - \sum_{k=1}^{B} \lambda_k \phi_k. \qquad (S22)$$

Given that the original objective function is convex, the maximization of the dual (equation S22) is equivalent to the problem of maximizing the original objective function (the entropy). Moreover, since the original problem is convex, the dual is convex as well (Bertsimas 1997). This can be easily checked because the Hessian of the Lagrangian function in equation (S22) is simply

$$H\left(L[\bar{\lambda}]\right)_{kl} = \langle I_k I_l \rangle_{\bar{\lambda}} - \psi_k(\bar{\lambda})\psi_l(\bar{\lambda}) \qquad (S23)$$

the covariance matrix between bin fraction indicator functions evaluated at $\bar{\lambda}$. The covariance matrix is by construction a positive semi-definite matrix. This provides another evidence that the problem of finding Lagrange multipliers is a convex one.

Generalization of equation (5) for multiple species

Here, we give a generalization of equation (5) in the main text when the single cell distributions measured from multiple chemical species are used to constrain the parameter distribution. Consider that we have measured cell-to-cell variability in *n* different experimental conditions. The experimental conditions are identified by several indicators including identity of the measured species, input level, time of measurement, etc. We avoid multiple subscripts to specify these various indicators and denote the experimental conditions as $\{x^1, x^2, \ldots, x^n\}$. We consider that the single cell distribution at each measurement *"a"* is binned in B_a bins. The maximum entropy parameter distribution is given by (see equation (5) in the main text)

$$P(\bar{\theta}) = \frac{1}{\Omega} q(\bar{\theta}) \exp\left(-\sum_{a=1}^{n}\sum_{k=1}^{B_a} \lambda_{ak} I_{ak}\left(x^a(\bar{\theta})\right)\right). \qquad (S24)$$

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In equation (S24), $I_{ak}(x)$ is the indicator function corresponding to the kth bin for the ath experimental condition, B_a is the number of bins representing the ath experimental condition, and λ_{ak} is the corresponding Lagrange multipliers.

Model of the EGFR/Akt signaling pathway

In this section, we describe in detail the dynamical model used to simulate levels of phosphorylated Akt as well as cell surface EGFRs after stimulation of cells with EGF.

The model of EGF/EGFR dependent phosphorylation of Akt was based on the previous work of Chen et al. (Chen, Schoeberl et al. 2009). We retained the branch of the Chen et al. model that leads to phosphorylation of Akt subsequent to EGF stimulation. The model had 17 species and 20 parameters. The description of the species is given in Supplemental information Table 3. The description of the parameters is given in Supplemental information Table 2. A system of ordinary differential equations describing dynamics of concentrations of species participating in signaling is given below (equations S24-S39). The model described EGF binding to EGFRs, subsequent receptors dimerization, phosphorylation, dephosphorylation, receptors internalization, degradation and delivery to cell surface and activation of Akt. We denote by active receptors phosphorylated receptors and by inactive receptors all other receptor states. In agreement with the literature only cell surface-localized phosphorylated receptors were allowed to activate Akt (Nicholson and Anderson 2002). We simplified the phosphorylation of pAkt through pEGFR; we implemented direct interaction between pEGFR and Akt leading to phosphorylation of Akt.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$\frac{dR}{dt} = k_{syn} - k_1 uR + k_{-1}B - k_i R + k_{rec}R_i - k_2 RB + k_{-2}D_1 \tag{S25}$$

$$\frac{dR_i}{dt} = k_i R - k_{rec}R_i - k_{deg}R_i \tag{S26}$$

$$\frac{dB}{dt} = k_1 uR - k_{-1}B - k_2 RB + k_{-2}D_1 - 2k_2 B^2 + 2k_{-2}D_2 - k_i B + k_{rec}B_i \tag{S27}$$

$$\frac{dB_i}{dt} = k_i B - k_{rec}B_i - k_{deg}B_i \tag{S28}$$

$$\frac{dD_1}{dt} = k_2 RB - k_{-2}D_1 - k_{ap}D_1 + k_{dp}P_1 - k_1 uD_1 + k_{-1}D_2 - k_i D_1 + k_{rec}D_{1i} \tag{S29}$$

$$\frac{dD_{1i}}{dt} = k_i D_1 - k_{rec}D_{1i} - k_{deg}D_{1i} + k_{dp}P_{1i} - k_{ap}D_{1i} \tag{S30}$$

$$\frac{dD_2}{dt} = k_2 B^2 - k_{-2}D_2 - k_i D_2 + k_{rec}D_{2i} - k_{ap}D_2 + k_{dp}P_2 + k_1 uD_1 - k_{-1}D_2 \tag{S31}$$

$$\frac{dD_{2i}}{dt} = k_i D_2 - k_{rec}D_{2i} - k_{deg}D_{2i} + k_{dp}P2i - k_{ap}D_{2i} \tag{S32}$$

$$\frac{dP_1}{dt} = k_{ap}D_1 - k_{dp}P_1 - k_1 uP_1 + k_{-1}P_2 - k_i^* P_1 + k_{rec}^* P_{1i} - k_{bind}P_1 Akt + k_{db}P_{1Akt} + k_a P_{1Akt} \tag{S33}$$

$$\frac{dP_{1i}}{dt} = k_i^* P_1 - k_{rec}^* P_{1i} - k_{deg}^* P_{1i} - k_{dp}P_{1i} + k_{ap}D_{1i} \tag{S34}$$

$$\frac{dP_2}{dt} = k_{ap}D_2 - k_{dp}P_2 + k_1 uP_1 - k_{-1}P_2 - k_i^* P_2 + k_{rec}^* P_{2i} - k_{bind}P_2 Akt + k_{db}P_{2Akt} + k_a P_{2Akt} \tag{S35}$$

$$\frac{dP_{2i}}{dt} = k_i^* P_2 - k_{rec}^* P_{2i} - k_{deg}^* P_{2i} - k_{dp}P_{2i} + k_{ap}D_{2i} \tag{S36}$$

$$\frac{dP_{1Akt}}{dt} = k_{bind}P_1 Akt - k_{db}P_{1Akt} - k_a P_{1Akt} \tag{S37}$$

$$\frac{dP_{2Akt}}{dt} = k_{bind}P_2 Akt - k_{db}P_{2Akt} - k_a P_{2Akt} \tag{S38}$$

$$\frac{dpAkt}{dt} = k_a \left(P_{1Akt} + P_{2Akt}\right) - k_p pAkt \tag{S39}$$

$$\frac{dAkt}{dt} = -k_{bind}Akt\left(P_1 + P_2\right) + k_{db}\left(P_{1Akt} + P_{2Akt}\right) + k_p pAkt \tag{S40}$$

Binning single cell data

To infer the joint distribution over model parameters, we used 24 measured distributions of cell-to-cell variability (20 pAkt distributions, 1 pAkt background fluorescence distribution and 3 sEGFR distributions, see below). For each measured distribution we used 11 bins. The locations and widths of the bins were chosen to fully cover the observed abundance range while also ensuring reliable estimates of the bin fractions $\bar{\phi}$. See Supplementary Information Table 1 for bin locations and experimentally estimated bin fractions.

We detected a small but significant pAkt signal in the absence of EGF stimulation. This background fluorescence signal likely originated from off target binding of pAkt-detecting antibodies. We assumed that the fluorescent readouts

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of pAkt/sEGFR levels in individual cells were equal to the sum of EGF dependent pAkt/sEGFR levels as computed using the signaling network model and the cell-dependent, but time-independent background fluorescence signal. In case of pAkt levels, the distribution of the background fluorescence was fitted to the experimentally measured distribution of the background fluorescence (pAkt readout without EGF stimulation). Unlike pAkt levels that respond to stimulation with EGF, cells maintain a high number of EGF receptors on the cell surface in the absence of EGF. As a result, we did not have experimental access to 'background fluorescence' distribution for sEGFR-detecting antibodies. We determined the range of background sEGFR fluorescence levels as follows. At the highest saturating dose of EGF (100 ng/ml) majority of the cell surface EGFRs are likely to be removed from the cell surface and degraded. At this dose, we assumed that the sEGFR background fluorescence can account for half of the measured fluorescence. We did not fit the distribution of background sEGFR levels to a specific distribution.

Numerical inference of Lagrange multipliers

The numerical search for Lagrange multipliers that are associated with bin fractions is a convex optimization problem (see above). We resorted to a straightforward and stable algorithm proposed in (Tkacik, Schneidman et al. 2006). The algorithm proceeded as follows. We started the calculations with an initial guess for the Lagrange multipliers at zero for each of the 11 bins of the 24 fitted distributions. In the nth iteration, using the Lagrange multipliers $\bar{\lambda}^{(n)}$, we estimated the predicted bin fractions $\bar{\psi}^{(n)}$ using Markov chain Monte Carlo (MCMC) sampling.

MCMC sampling in each iteration was performed as follows. We propagated 50 parallel chains starting at random points in the parameter space. Individual MCMC chains in the parameter space were run as follows. In the MCMC, on an average 10 parameters were changed in a single Monte Carlo step. The parameters were constrained to be within the upper and lower limits determined individually for each parameter based on available literature

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

estimates (see Supplemental Information Table 2). Each chain was run for 50000 MCMC steps. At each step, we solved the system of differential equations given in equations (S25)-(S40) numerically with the proposed parameter assignment using the ode15s function of MATLAB. We evaluated the pAkt and sEGFR levels and accepted the proposed parameters using the Metropolis criterion applied to equation (5) in the main text. Briefly, for any set of parameters, we defined the energy

$$E(\bar{\theta}) = -\log P(\bar{\theta}) = \sum_{i=1}^{n} \sum_{k=1}^{B_i} \lambda_{ik} I_{ik}\left(x^i(\bar{\theta})\right) + const. \qquad (S41)$$

Starting from any parameter set $\bar{\theta}$, a new parameter set $\bar{\theta}'$ was proposed as described above. Then, the differential equations describing system dynamics were solved and the new energy $E(\bar{\theta}')$ was computed. The difference in energy $\delta E = E(\bar{\theta}') - E(\bar{\theta})$ was used to probabilistically accept/reject the new parameter set with an acceptance probability

$$p_{acc} = min(1, \exp(-\delta E)). \qquad (S42)$$

Parameter points that predicted pAkt and sEGFR levels outside of the ranges observed in experimental data were rejected (see Supplemental Information Table 5 for allowed ranges). We discarded the first 5000 steps as equilibration and saved parameter values every 50th iteration. At the end of the calculation, parameter samples from all MCMC chains were combined together. We also imposed a few realistic constraints on pAkt and sEGFR time courses predicted by the model. All parameter sets that did not satisfy these constraints were discarded. The constraints were as follows. (1) Given that EGF ligand induces receptor endocytosis, we required that the surface EGFR levels at 180 minutes of sustained stimulation with 100 ng/ml EGF to be lower than the steady state surface EGFR levels in the absence of EGF stimulation. (2) Similarly we required that pAkt levels at 45 minutes were lower than pAkt levels at 5 minutes for the highest EGF stimulation (100 ng/ml).

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Using the sampled parameters, we estimated the bin fractions $\bar{\psi}_n$ as well as the elements of the relative error vector $\bar{\Delta}^{(n)} = \bar{\psi}^{(n)} - \bar{\phi}^{(n)}$ in the nth iteration. For the n+1st iteration, we proposed new multipliers $\bar{\lambda}^{(n+1)} = \bar{\lambda}^{(n)} + \alpha^{(n)}\bar{\Delta}^{(n)}$. The multiplication constant $\alpha^{(n)}$ was chosen as follows. First, the approximate estimate of the predicted bin fractions for a given value of $\alpha^{(n)}$ was obtained using the Taylor series expansion

$$\bar{\psi}_{pred}^{(n+1)} \approx \bar{\psi}^{(n)} - \alpha^{(n)} c^{(n)} \bar{\Delta}^{(n)} \tag{S43}$$

where $c^{(n)}$ is the covariance matrix with entries

$$c_{kl}^{(n)} = \langle I_k I_l \rangle_{\bar{\lambda}^{(n)}} - \psi_k\left(\bar{\lambda}^{(n)}\right)\psi_l\left(\bar{\lambda}^{(n)}\right) \tag{S44}$$

when the Lagrange multipliers are fixed at $\bar{\lambda}^{(n)}$. We chose $\alpha^{(n)}$ in the interval [0.05, A_n] so as to minimize the predicted error $e_{pred}^{(n+1)} = \left|\bar{\psi}_{pred}^{(n+1)} - \bar{\phi}\right|$.

1000 MC steps took 5-10 minutes. At the end of the calculation, the numerically inferred distribution over parameters captured with high accuracy the individual bin fractions of the distributions that were used to constrain it (Pearson's r^2 = 0.9, $p < 10^{-10}$, median relative error = 14%). Notably, as seen in SI Figure 3, the predicted bin fractions from two independent calculations to determine the Lagrange multipliers were highly correlated with each other (Pearson's r^2 = 0.99, $p < 10^{-10}$) indicating that the calculations converged to the same parameter distribution.

Inversion of covariance matrix (equation (S10))

In order to make predictions using MERIDIAN, we first sample several parameter points from the parameter distribution $P(\bar{\theta}|\bar{\Lambda})$ (equation (5) in the main text) using MCMC and the Metropolis criterion as described above. Using N_S parameter samples, we generate a sparse matrix with entries M_{ab} where *a* is the index of the sample point ($a \in (0, N_S)$) and *b* is the index of the bin (and the experiment). There are a total of 24x11 = 264 bins used in this work and the *b* index runs between 1 and 264. The entry $M_{ab} = 1$ only if the model solutions pass through the bth bin for any given set of parameters. From the matrix M, we

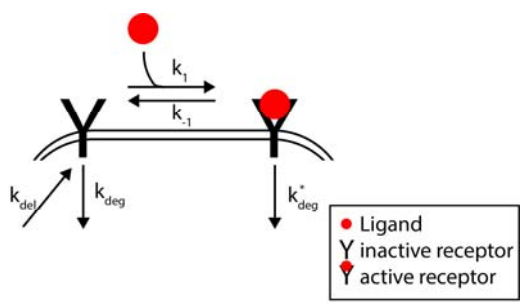
bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

estimate the 264x264 covariance matrix among the constraints. The entries of the convariance matrix are given by

$$c_{kl} = \langle I_k I_l \rangle_{\bar{\Lambda}} - \psi_k(\bar{\Lambda})\psi_l(\bar{\Lambda}) \tag{S45}$$

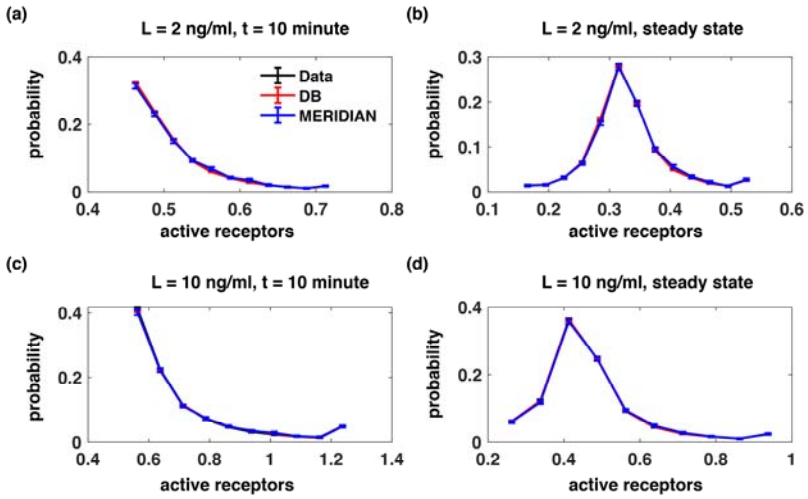
where $k, l \in [1, 264]$. Next, we compute the inverse of the covariance matrix. Since all bin fractions at any given experimental conditions add up to one by definition, the covariance matrix is not full rank. Indeed, it has a total of 24 zero eigenvalues corresponding to 24 redundancies in the constrained single cell distributions. When inverting the covariance matrix, we neglect these 24 zero eigenvalues. The resultant inverse c^+ is the so-called Moore-Penrose pseudoinverse of the matrix.

SI Figures



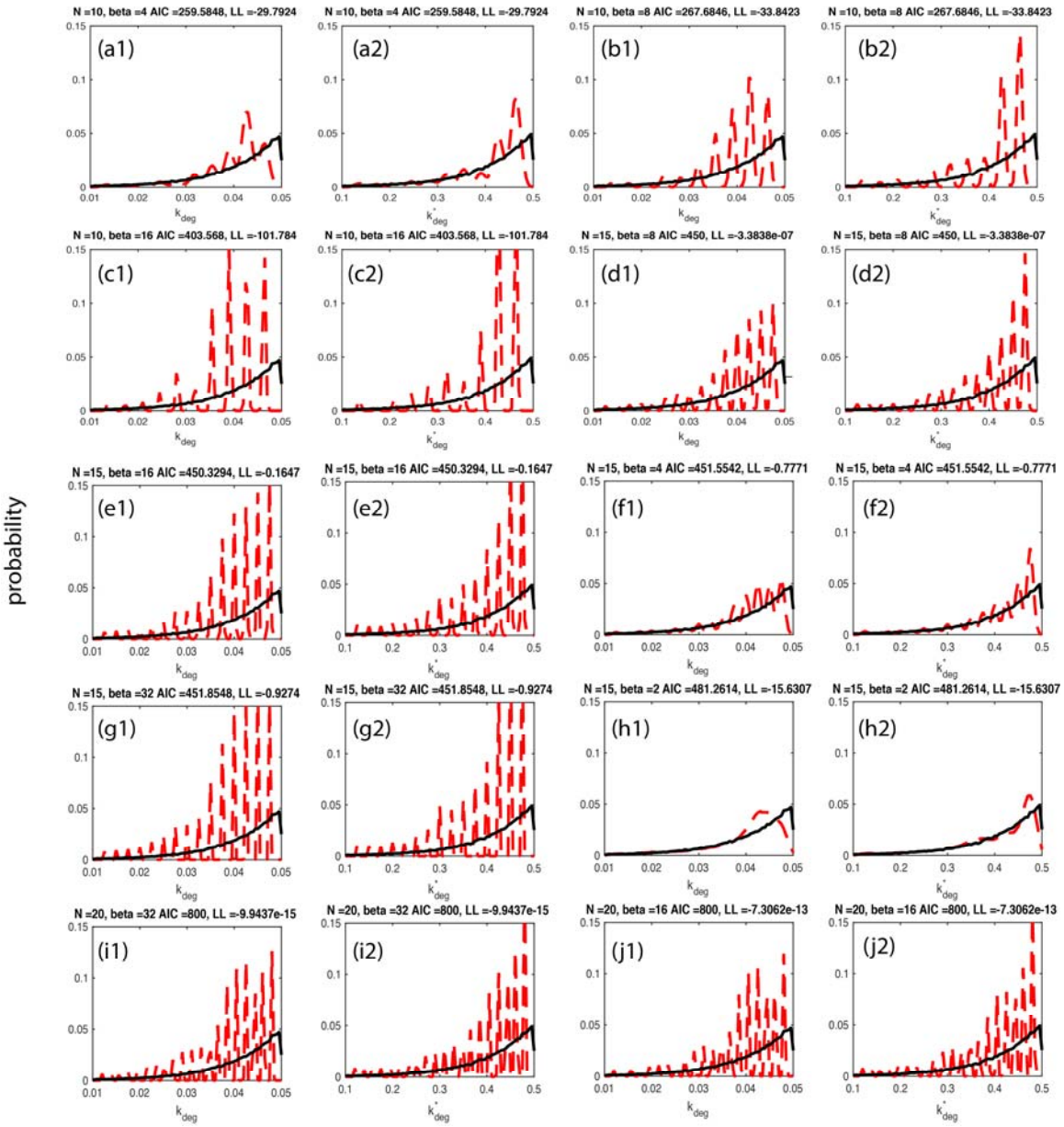
Supplemental Information Figure 1. Schematic of the simplified growth factor model. Inactive receptors are delivered from the intracellular medium on the plasma membrane. Receptors bind to the ligand and get activated. Both activated and inactivated receptors are removed from the cell surface and degraded albeit at different rates.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



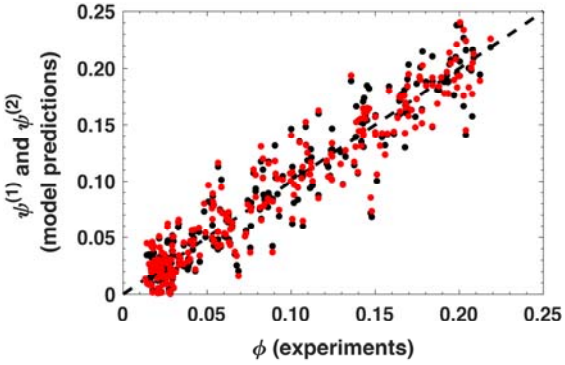
Supplemental Information Figure 2. Fitted bin fractions using DB and MERIDIAN. We show the "ground truth" bin fractions (black lines) and the corresponding fits using MERIDIAN (blue lines) and the DB approach (red lines) across four simulation conditions.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

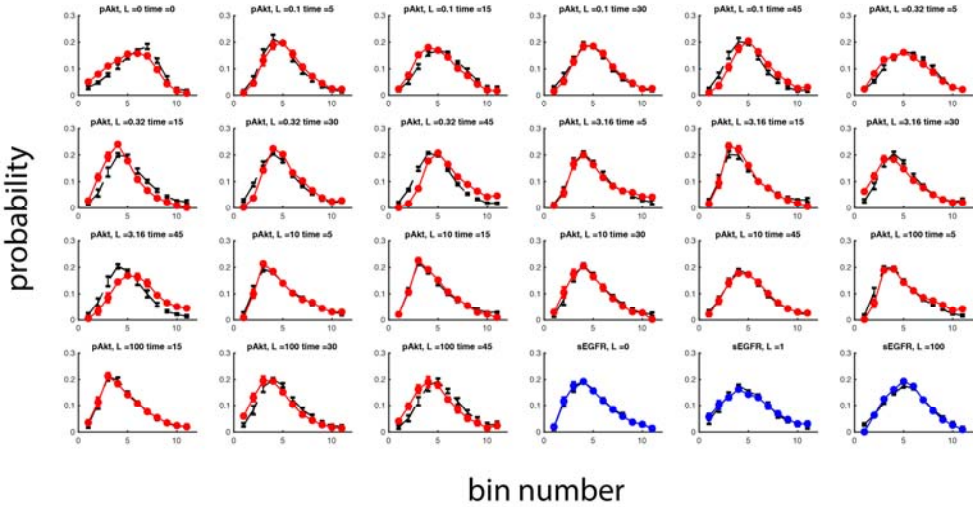


Supplemental Information Figure 3. Inferred parameter distribution for top 10 DB fits. Panels (a1-j1) show the inferred distribution $P(k_{deg})$ and panes (a2-j2) show the inferred distribution $P(k^*_{deg})$ for top ten choices of method-parameters *N* (number of grid points) and β (width of the Gaussian basis function) according to the Akaike information score (dashed red line). The Akaike score and the likelihood are shown on the top of each panel. Black lines indicate the "ground truth" distributions.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

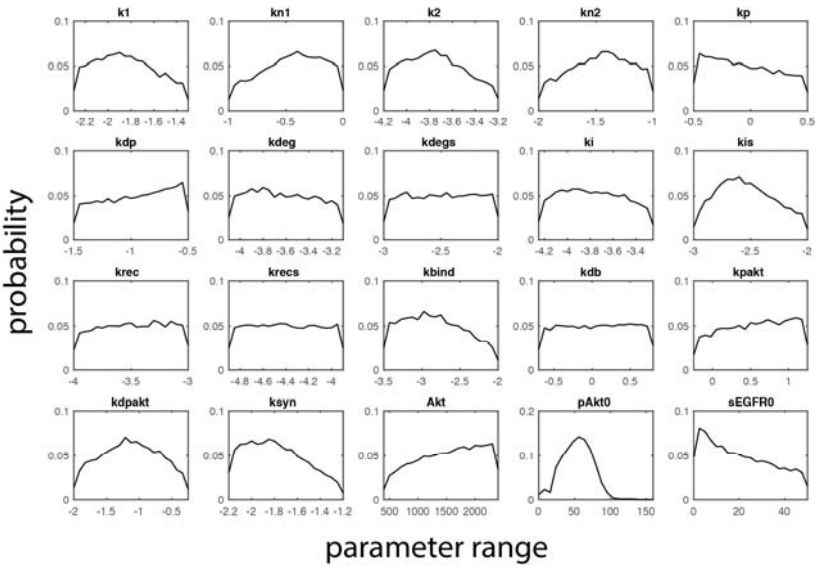


Supplemental Information Figure 4. Model predictions agree for two independent calculations. The correlation between experimentally estimated bin fractions (x axis) and predicted bin fractions (y axis) for two independent searches for the Lagrange multipliers (red and blue dots).

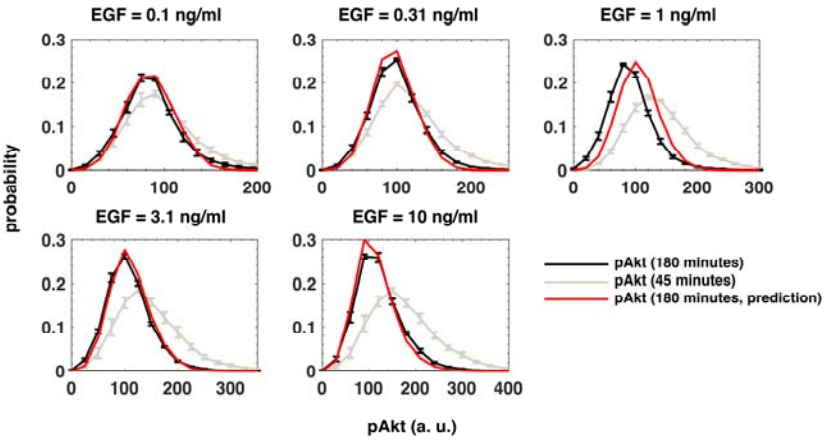


Supplemental Information Figure 5. 21 fitted pAkt distributions and 3 fitted sEGFR distributions used in parameter inference. We show all fitted single cell distributions (pAkt fits in red, sEGFR fits in blue, experimental data in black) used in the inference. Error bars represent experimentally estimated standard errors in the mean and model estimated uncertainties.

bioRxiv preprint doi: https://doi.org/10.1101/137513; this version posted April 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplemental Information Figure 6. Inferred marginal distributions of all 20 model parameters.



Supplemental Information Figure 7. Population dynamics has not reached steady state at 45 minutes. We plot single cell distributions of pAkt levels at 45 minutes (gray lines) and at 180 minutes (experiments in black lines, model fit in red lines) across a broad range of EGF doses. Error bars represent experimentally estimated standard errors in the mean. For simplicity, we do not show the model estimated uncertainties.