Maximum entropy framework for inference of cell population heterogeneity in signaling networks

Purushottam Dixit[1], Eugenia Lyashenko[1], Mario Niepel[2], and Dennis Vitkup[1,3]

Short Abstract — **Predictive models of signaling networks are key to our understanding of cellular function and in designing rational interventions in disease. However, using network models to predict signaling network behavior is challenging due to inherent cell-to-cell variability of network parameters, such as reaction rates and species concentrations that govern network dynamics and parameter non-identifiability. In this work, we present an inference framework based on the principle of maximum entropy (ME) to estimate the joint probability distribution over network parameters that is consistent with experimentally observed cell-to-cell variability in concentrations of network species. We apply the framework to study the signaling cascade activated by the epidermal growth factor (EGF) resulting in phosphorylation of protein kinase B (Akt); a central signaling hub in mammalian cells. Notably, the inferred parameter distribution allows us to computationally predict the distributions of phosphorylated Akt (pAkt) levels at early and late times after EGF stimulation as well as the distribution of cell surface EGF receptors (sEGFRs) after prolonged stimulation with EGF. We discuss how the developed framework can be generalized and applied to problems beyond signaling networks.**

Keywords — **cell-to-cell heterogeneity, signaling networks, parameter estimation, maximum entropy**

I. PURPOSE

Signaling networks within individual cells in a population often respond to extracellular stimuli in a heterogeneous manner even if the cells are isogenic. This heterogeneity has important functional consequences.

Several complementary experimental techniques have been developed to probe cell-to-cell variability in intracellular concentrations of biochemical species participating in signaling networks. Unfortunately, it is often challenging to estimate the distribution over network parameters (such as reaction rates and species abundances) and to predict the time evolution of species abundances that is consistent with experimentally measured cell-to-cell variability in network dynamics.

Over the last decade, computational methods have been developed to estimate the joint distribution $P(\overline{\theta})$ of network parameters consistent with experimental data. Here, $\overline{\theta} = \{\theta_1, \theta_2, \dots\}$ is a vector of parameters that describe the dynamics of the signaling network. The current approaches rely on several simplifying assumptions. For example, the parameter distributions are restricted to a pre-defined functional family, such as the multivariate log-normal distribution. Or, data collected at different time points and experimental conditions are assumed to be statistically independent of each other thus simplifying the probability of observing the entire data as a product of the probability of observing individual data points.

Building on our previous work[1,2], we present here a maximum entropy (ME) based framework to infer $P(\overline{\theta})$. Particularly, our approach circumvents the aforementioned simplifications.

We apply the ME framework to estimate parameter distributions in the signaling network leading to phosphorylation of protein kinase B (Akt) induced by the epidermal growth factor (EGF)-dependent activation of its receptor (EGFR); a central mammalian signaling cascade implicated in many cellular processes and diseases. Based on measured cell-to-cell variability in phosphorylated Akt (pAkt) and cell surface EGFR (sEGFR) levels in MCF 10A cells after stimulation with multiple EGF doses, we infer $P(\overline{\theta})$ for the pathway. $P(\overline{\theta})$ allows us to computationally predict the cell-to-cell variability in pAkt levels at unobserved early and late time points as well as cell-to-cell variability in cell surface EGFRs at steady state in response to a constant stimulation. We also discuss how to generalize the framework to other type of biological systems and experimental data.

REFERENCES

[1] Dixit PD (2003) Quantifying extrinsic noise in gene expression using the maximum entropy framework. Biophys. J. (104) 2743-2750 Lastone AB, Lasttwo AB (1993) *Book style.* Wadsworth, Belmont, CA.

[2] Eydgahi H. et al (2013) Properties of cell death models calibrated and compared using Bayesian approaches. Mol. Sys. Biol. (9) 644

[1]Department of Systems Biology, Columbia University. E-mail: dixitpd@gmail.com

[2]Department of Systems Biology, Harvard Medical School

[3] E-mail: dv2121@cumc.columbia.edu