

Chapter 30

Unsupervised Methods to Identify Cellular Signaling Networks from Perturbation Data

Madhusudan Natarajan
Pfizer, USA

ABSTRACT

The inference of cellular architectures from detailed time-series measurements of intracellular variables is an active area of research. High throughput measurements of responses to cellular perturbations are usually analyzed using a variety of machine learning methods that typically only work within one type of measurement. Here, summaries of some recent research attempts are presented—these studies have expanded the scope of the problem by systematically integrating measurements across multiple layers of regulation including second messengers, protein phosphorylation markers, transcript levels, and functional phenotypes into signaling vectors or signatures of signal transduction. Data analyses through simple unsupervised methods provide rich insight into the biology of the underlying network, and in some cases reconstruction of key architectures of the underlying network from perturbation data. The methodological advantages provided by these efforts are examined using data from a publicly available database of responses to systematic perturbations of cellular signaling networks generated by the Alliance for Cellular Signaling (AfCS).

INTRODUCTION

How does cellular machinery function? What is the network of molecular entities that governs equally well resting basal homeostasis, as well as specific and precise response to cellular stimuli that drive a diverse range of cellular function?

The dissection of cellular architectures and accompanying function has followed those precise routes – to first identify molecular entities, followed by a reductionist approach to place it in a simplified physiological context. With the advent of high-throughput experimentation and rapid technological advances, focused biochemical experiments are being replaced with high con-

DOI: 10.4018/978-1-4666-3604-0.ch030

tent experiments. These experiments permit the sampling of large numbers of molecular entities with sufficient resolution and accuracy to develop a comprehensive parts list. However, quantitative descriptions of both (relative) amounts and activity levels of a large number of molecular entities within cells pose novel problems. First, identification of multiple causal relationships between vast numbers of measured molecular entities is needed to define the architecture of the interactions between molecules. Second, the relative contribution of a molecule in the context of a network of interactions is needed to delineate functional consequences.

The inferencing of cellular architectures and functions from biological measurements has rich precedents in engineering - especially in systems identification in the fields of systems engineering and control systems, but mostly in non-biological areas (Ljung, 1999). The goal in systems identification is to be able to build dynamical models from measured data. Therefore, the characterization of cellular networks using mathematical methods is an extension of systems identification theories in that it is simply an attempt to formalize and summarize the system's resting behavior and its response to perturbation. In an ideal scenario, the cellular architecture inferred from the data is isomorphic (i.e., structurally identical) and is identifiable - given sufficient observations from the system, it is possible to uniquely infer the parameters of the model producing the data. In reality, these are almost impossible to achieve - first, cellular architectures are vastly more complex than engineered systems such that even if resource limitations are not considered, sufficient observations may not be available to satisfy the identifiability criterion. Second, while we have made remarkable progress in the inference of biological networks and functions from data, the algorithmic methods are still far from perfect. The availability of additional data is currently not a bottleneck. Thus, state of the art is still focused on developing better methods to deal with existing data sets.

One classical method of exposition of biological function is through perturbation analysis. Classical biological perturbation studies have typically attempted to make incremental change to biological systems through tools such as pharmacological agents, environmental stressors, or more recently anti-sense technologies and measure functional performance after each change. With high-throughput data available to monitor responses to each such perturbation, studies have shown that responses to perturbation are directly amenable to mapping network topology and can yield significant insight into network architecture (Tegner, Yeung, Hasty, & Collins, 2003; Yeung, Tegner, & Collins, 2002).

The problem of inferring cellular networks from measured data is certainly not novel, and has been an active area of interest for several years now (Friedman, 2004; Kalir & Alon, 2004; Nachman, Regev, & Friedman, 2004; Ronen, Rosenberg, Shraiman, & Alon, 2002). In the past, multiple studies have simplified the problem by typically focusing on one type of measurement that is amenable to an analysis method. For example, investigators have reconstructed gene expression regulatory networks or protein-protein interaction networks using statistical and machine learning methods (Friedman, Linial, Nachman, & Pe'er, 2000; Kim, Imoto, & Miyano, 2003; Nariai, Kim, Imoto, & Miyano, 2004; Patil & Nakamura, 2005; Perrin et al., 2003). The use of deterministic modeling of enzyme kinetics and metabolic networks has provided significant insight into signaling machineries (Hendriks et al., 2006; Janes & Lauffenburger, 2006; Palsson, Price, & Papin, 2003). Restricting the focus to only one such measurement allows easy comparison within datasets and avoids the problems faced with integration of disparate data-types. For example, in measurements of transcriptional regulation, investigators can identify en-bloc patterns such as correlated or anti-correlated gene expression and attempt to describe functional linkages between genetic elements (Niehrs & Pollet, 1999). While such descriptions are excellent first pass estimations

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/unsupervised-methods-identify-cellular-signaling/76083

Related Content

Spatial Resolution of Shapes in Gamma Camera Imaging Using an Exact Formula for Solid Angle of View

S. Zimeras (2012). *International Journal of Systems Biology and Biomedical Technologies* (pp. 35-51).
www.irma-international.org/article/spatial-resolution-shapes-gamma-camera/63045

Key Issues and Challenges for Managing and Evaluating B2B E-Commerce Projects within the Australian Pharmaceutical Supply Chain

Chad Linand Geoffrey Jalleh (2012). *Pharmacoinformatics and Drug Discovery Technologies: Theories and Applications* (pp. 54-71).
www.irma-international.org/chapter/key-issues-challenges-managing-evaluating/64066

Information Needs of Bioinformatics Researchers

Manlunching (2017). *Library and Information Services for Bioinformatics Education and Research* (pp. 34-44).
www.irma-international.org/chapter/information-needs-of-bioinformatics-researchers/176136

Drug Optimization for Cystic Fibrosis Patients Based on Disease Pathways Crosstalk

Shuting Linand Yifei Wang (2021). *International Journal of Applied Research in Bioinformatics* (pp. 1-11).
www.irma-international.org/article/drug-optimization-for-cystic-fibrosis-patients-based-on-disease-pathways-crosstalk/267820

Efficient Fault Tolerant Algorithms for Internet Distributed Systems

Swati Mishraand Sanjaya Kumar Panda (2017). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 71-90).
www.irma-international.org/article/efficient-fault-tolerant-algorithms-for-internet-distributed-systems/178608