

CS341 SYSTEMS BIOLOGY I

LECTURE 1

1. INTRODUCTION

Stephanie described her research and the techniques she uses as examples of what we will learn in this class.

2. WHAT IS SYSTEMS BIOLOGY?

Here is my favorite definition.

Definition 1. Systems biology – Systems biology is the study of an organism, viewed as an *integrated* and *interacting network* of genes, proteins and biochemical reactions which give rise to life. (Institute for Systems Biology)

This means systems biology is trying to solve a multi-scale problem – we want to know how intercellular machinery works, how cells work together, how organs work, and how organisms work. At any level of detail we choose, there are interacting components. The approach taken by systems biology is to focus on all components and interactions among them rather than single components in isolation.

The Institute for Systems Biology goes on to say

Instead of analyzing individual components or aspects of the organism, such as sugar metabolism or a cell nucleus, systems biologists focus on all the components and the interactions among them, all as part of one system. These interactions are ultimately responsible for an organisms form and functions. For example, the immune system is not the result of a single mechanism or gene. Rather the interactions of numerous genes, proteins, mechanisms and the organisms external environment, produce immune responses to fight infections and diseases.

3. SYSTEMS BIOLOGY APPROACHES

3.1. Data-Driven Modeling. High throughput data, such as microarray data can provide information, but sometimes it is an overwhelming amount of information. To learn about a system for which there is a lot of data but not a lot of understand, it is useful to

construct data-driven models. As Janes and Yaffe (Nat Rev Mol Cell Biol, 2006) put it, “models can be constructed based solely on analysing the data itself, without having to make any assumptions about the underlying mechanisms.”

That analysis can then be used to make predictions or generate insight.

One form of analysis is clustering. Where we identify groups of data that behave similarly over time and in different experimental conditions.

Suppose we are interested in learning about a certain group of proteins and we have identified some of them. Suppose further that we have data for these proteins and lots and lots of others over time. Let’s analyze the data to determine which proteins behave similarly to our set A. Then, it is likely that they are being affected by the same things. In other words, they may share the same upstream regulators. Now, to learn more about the proteins in this group, we focus on the ones with similar timecourses – we have narrowed the field.

From Janes et al (Nat Rev Mol Cell Biol, 2006).

Clustering is most successful when the observations and variables in the data matrix contain a mix of known and unknown biological mechanisms. Recent phenotypic, pharmacological and RNA interference (RNAi)-based screens have used clustering to indicate the functions of unknown genes or small molecules by their association with recognized perturbations. Clustering time-course data with many variables can also be used to separate kinetic clusters that are possibly subject to the same upstream regulators. These “guilt by association” studies benefit from clustering by focusing the user’s attention to the subset of data within the clusters, rather than the entire starting data set.

Other forms of data analysis include principle component analysis (PCA) and partial least squares (PLS) regression. Both involve linear algebra techniques to reduce high-dimensional data to lower dimensional data.

3.2. Mechanistic Modeling.

Definition. Mechanistic Modeling – use prior knowledge to make specific molecular predictions and work best with pathways in which components and connectivity are relatively well established (Aldridge et al, Nat Rev Mol Cell Biol, 2006). Note this is also called biophysical and physicochemical modeling.

The circadian clock is a great example. A circadian clock coordinate the daily rhythms of organisms. To see a circadian clock in action, all you need to do is put a mouse in a dark cage, give it some food and a running wheel. Every 23.5 hours, it will get on to the wheel and run for approximately 8 hours. Why is this so repeatable? Because there is a clock in the brain that outputs signals indicating when it is time to wake up, to prepare for a meal, etc..

Through experimentation, biologists have determined that in the suprachiasmatic nucleus (in the hypothalamus), there are thousand of neurons, each of which contains a clock. The neurons signal each other to make sure their internal clocks are coordinated. So two important questions are: What are the intracellular clockworks? and what are the signals that are sent? Let's consider the first question. The clockworks involve genes and proteins. Certain genes code for proteins. The proteins are created, and then they return to the nucleus and stop production of additional proteins. This negative feedback loop causes the concentration of the clock proteins to rise and fall over time. It does this regularly, creating a clock.

Through knock-out experiments, several clock genes have been identified, and through analysis of the clock proteins, a basic mechanism has been proposed. And modeled. The purpose of the model is to further clarify the proposed mechanism and to simulate the results, to evaluate the plausibility of our understanding.

3.3. Web site. The web site is on the cs dept server. There will be journal articles linked to form the website, but who live in Moodle. This way, I am not publicly posting materials that are not publicly available. There will also be Zoom session recordings and potentially other movies. Those will be on Google Drive, but linked to from the web page.

3.4. Journal Articles. An important skill in any academic discipline is that of reading journal articles. In an interdisciplinary field like systems biology, it is important to learn to read journal articles published by those outside your field. This can be challenging and requires you to feel comfortable with a less-than-complete understanding. An important skill to learn is discernment – in this case how to discern between and important detail and an unimportant detail. With each reading in this course, I will give you a set of questions, that will make it clear what are the most important points for you to understand when reading a paper. Begin by focusing on those points. As you become more comfortable with reading articles, you can determine for yourself how much you need to understand.

3.5. Numpy in Jupyter Notebooks. We will be using Python, writing and running code in Jupyter notebooks. We will be writing relatively short programs that use Numpy, Scipy, and Matplotlib – python packages allowing numeric computation, scientific computing (in particular, numeric differential equation solvers), and plotting. The Jupyter notebooks allow us to interleave code, its output (e.g. graphs), and text.