

Epistemology of Computational Biology and Modeling of Complex Heterogeneous Systems

Ivan V. Ivanov

Veterinary Physiology & Pharmacology, TAMU
Center for Translational Environmental Health Research
Ivanov@cvm.tamu.edu

Keywords: *omics data, optimal Bayesian classification, Canonical Correlation Analysis, regulatory network modeling and control.*

In the last two decades, biological sciences have changed significantly with the advent of large amounts of *omics* data and the related development of post-genome technology. Bioinformatics and Computational Biology became important fields of research that require a different set of rules of engagement. The push for "big data" research requires careful considerations when designing experiments, and questions like "What is scientific knowledge" and "Does Computational Biology require its own epistemology" have to be addressed before one declares that data mining is all that is needed. While one cannot expect that a single lecture can suffice in properly addressing such questions a few examples of the current undertakings in molecular biology and translational science can hopefully initiate a discussion on these important topics. Examples of research projects focused on the highly complex ecosystem of epithelial cells and microbes in the human gut will be used to illustrate certain typical aspects of Bioinformatics and Computational Biology placed in the context of the two questions above. The examples will also be used to underline "What makes biology really difficult". Methods for noninvasive probing of the system as well as computational techniques for modeling the interactions among the components of this complex system will be presented. The roles of predictive models in biology and the related epistemology will be discussed. In particular, we will show how prior biological knowledge and model-based approaches can address the major challenge in biological sciences: high dimensionality of the space of measured variables/features vs. the typically low number of samples. We will briefly discuss recent developments and applications of optimal Bayesian classification, Canonical Correlation Analysis, and gene regulatory network modeling for the purposes of control and therapeutic intervention.