Editorial: Current progress in Bioinformatics 2010

In this issue, we provide the next installment of our annual feature ‘Current Progress in Bioinformatics’. Each year, we invite leaders in bioinformatics to summarize the progress and challenges in hot and emerging areas of inquiry. We ask the authors specifically to summarize the past 12–18 months of important literature, and to provide a perspective in the field. The fields are chosen based on many different considerations: we start by studying the newest work in both bioinformatics journals as well as application-area journals in biology and medicine. We also consider the papers presented at bioinformatics conferences throughout the year, and the general ‘buzz’ of scientific areas. This year, the areas really jumped out without us having to work too hard. The field of bioinformatics is now a relatively stable discipline, with journals, conferences and a professional society all going back more than 10 years. Nonetheless, it is amazing to see how certain problems can ‘sneak up’ on us—the emergence of the data handling and analysis challenges associated with next-generation sequencing data is a prime example—many of us would have said that sequence analysis and genome assembly was a relatively dormant field, having solved most of the key problems for analysis and search that confronted biologists. Wow, were we wrong! Within months, many biological lab infrastructures were brought to their knees by the volume of data created by the new sequencing machines. Basic capabilities like storage, raw data triage and certainly search all were stressed by these data and the issues of basic sequence analysis have bounced back as a very busy area of bioinformatics. Similar ‘surprises’ have emerged from translational bioinformatics, synthetic biology and others. So we are pleased to have selected a few key areas where we can present a summary of progress. As usual, we organize these reviews in the rough order of the central dogma from DNA to RNA to proteins to networks to organs to organisms.

Dalca and Brudno present a summary of next-generation sequencing informatics challenges in ‘Genome variation discovery with next generation sequencing data’. The throughput of these new sequencing technologies is breathtaking, but so are the challenges in terms of noise. After first providing a brief review of the basic sequencing technologies, they review methods for mapping reads on to existing assemblies, SNP and indel discovery, and discovery of structural variation on a larger scale. They also mention software systems that are emerging for providing a platform for the continuing analysis of these important data.

The study of biology from a systems perspective has emerged as an important theme of bioinformatics in the last 5 years (often under the rubric of ‘systems biology’ but more generally as network and graph-based views of biological systems). The initial approaches often used static views of interactions in an ‘equilibrium’ view of biological homeostasis. In ‘Towards the dynamic interactome: it’s about time’ Przytycka, Singh and Slonim provide a review of recent methods to look at the evolution of interactions over time. Clearly, in the context of developmental biology (including stem cell biology) and dynamic response to perturbations such as drugs or other environmental inputs, it is absolutely critical to understand the dynamic response modes of cellular networks. In this review, the authors review the importance of time, space and context in understanding biological responses, the data sources that are relevant to these questions and the informatics techniques that are useful. It is exciting to imagine our static pathways ‘coming to life’ as we overlay data sets that shed light on dynamics. They end with the suggestion that our understanding of human genetic variation will necessarily require dynamic models of human biology.

A critical technical capability for the analysis of biological systems has become the ability to integrate large data sets. The initial enthusiasm for any particular type of data (e.g. expression data) typically yields to the recognition that biological inferences are most robust when they stem from multiple data sources with different sources of bias and noise. In ‘Knowledge-based data analysis comes of age’,
Michael Ochs summarizes the progress in overcoming the curse of dimensionality by using powerful mathematical (often explicitly probabilistic, Bayesian) models to normatively combine our \textit{a priori} probabilities with data to extract accurate \textit{a posteriori} estimates of biological parameters of interest. The high false positive rate of many high-throughput experimental technologies demands robust formal methods for integration of data.

In ‘Pathway tools version 13.0: integrated software for pathway/genome informatics and systems biology’, Karp and coworkers provide a useful update on the latest updates to the Pathway Tools environment for capturing and analyzing genetic, protein, metabolic and regulatory networks in over 800 organisms. This provides a key integrating piece of technology for the study of molecular systems biology.

The ultimate test of our understanding of biological systems is to physically manipulate them to create new capabilities. Many consider the emerging field of ‘synthetic biology’ to be an entirely experimental venture, but Alterovitz, Muso and Ramoni in ‘The challenges of informatics in synthetic biology: from biomolecular networks to artificial organisms’ argue that \textit{in silico} support for synthetic biology is absolutely critical and forms the beginning of the CAD/CAM (computer-aided design/computer-aided manufacturing) era of biology and bioinformatics! To get things started, the authors review the emergence of standards for describing biological parts, and the major experimental areas of investigation, including synthesis of DNA sequences, gene expression control constructs and systems biology approaches to engineering new pathways.

The rise in interest in ‘translational research’ in which basic science insights are brought to impact patient care has naturally lead to the emergence of ‘translational bioinformatics’ in which bioinformatics tools are used to study clinical/medical phenomena. Translational bioinformatics can be used to characterize the genetic underpinnings of disease, as well as to understand the molecular response to drugs, and (more generally) the link between molecular biology and clinical phenotypes. In ‘Advances in translational bioinformatics: computational approaches for the hunting of disease genes’, Kann provides an overview of how the combination of genome sequencing and high-throughput functional genomics data has enabled new understanding of the molecular bases for disease and its therapy. She summarizes how bioinformatics technologies can find new genetic modulators of disease, as well as the important variations associated with these genes.

In many ways, one ultimate goal of bioinformatics is a working \textit{in silico} model of a cell, organ or organism so that medical interventions can be tested on models before they are used on humans. The field of patient-specific modeling is just now emerging, and Neal and Kerckhoffs provide a summary in ‘Current progress in patient-specific modeling’. Currently, the field is dominated by ‘top down’ models based on 3D image data. Important progress has been made in modeling blood vessels and cardiovascular physiology, the heart (particularly for device modeling) and the musculoskeletal system. The opportunities for creating multiscale models that link molecular information to these higher level physiological systems are myriad.

Not part of our ‘current progress’ series, but also included in this issue are four valuable additional reviews. Duval and Hao discuss recent progress in gene selection for accurate classification. Sloot and Hoekstra review multiscale modeling in a piece that is very complementary to the one by Neal and Kerckhoffs. The review of cellular reaction systems by Kirkilionis beautifully complements that of Przytycka, Singh and Slonim. Finally, Kim and Park provide a useful review of a new book ‘Modern Genome Annotation’.

In summary, the field of bioinformatics continues to evolve to face the challenges that emerge across biology and medicine. We have shown remarkably flexibility in focusing on the data analysis needs of these communities, even when they take us a little by surprise! We hope you will enjoy reading about these important trends, and imagining where they will take us in the coming years.

\textit{Russ B. Altman}

\textit{Stanford University}