Preface

The biotechnological advances of the last decade have forever changed the landscape of biological research. We are confronted daily with a deluge of data resulting from our present ability to measure cellular activities at the molecular level in genome-wide fashion. The task of organizing and structuring this information so that it can provide a reasonably coherent biological picture is both extremely exciting and truly daunting. Mathematical representation to model biological systems has become a very helpful step in integrating large amounts of varied data types, and a growing number of bioinformatics approaches have been developed to discover emerging biological properties in those systems. Hundreds of bioinformatics algorithms have been developed to tackle biological questions; the assessment of those remains an issue. A number of community efforts have emerged to address the ensuing complexities in data analysis, modeling, experimental design, and wet lab experimentation that no single research lab can handle alone. This volume presents a snapshot of some of the work developed under the umbrella of two such community efforts: the European Network of Excellence (ENFIN) and the Dialogue for Reverse Engineering Assessments and Methods (DREAM) project.

ENFIN develops infrastructure, tools, and methods to enhance integrative systems biology in Europe. ENFIN projects address three important fields of research: (1) function prediction of discrete molecular entities (such as protein or protein families); (2) network reconstruction; and (3) systems level modeling. One concept of the ENFIN network is the strong collaboration between dry and wet laboratories, which cycle data between computational predictions and experimental validations. Although wet experiments are used to validate chosen computational predictions, they often do not allow assessing the quality of computational methods, because of limited scale, technology, and resources. Motivated by the growing number of bioinformatics tools available in systems biology, researchers connected to the ENFIN organization are developing strategies to assess the accuracy of the computational predictions.

The Dialogue for Reverse Engineering Assessments and Methods (DREAM) initiative is a project designed to bridge the gap between computational systems biology algorithms (that, in turn, feed from high-throughput data sets) and biological knowledge. The missing link in this pipeline is the relative absence of experimental validation of the high-throughput hypotheses generated by computational tools. It is not that the same high-throughput technologies could not generate additional experimental support for these predictions—indeed they could. However, experimental biologists have not been keen to invest money and effort to validate large sets of computational predictions, in part, because these predictive methods still draw significant skepticism from that community. Thus, while ENFIN funds joint projects that cycle between computational predictions and experimental validations, DREAM aspires to create a formal self-assessment methodology in which models, in particular, molecular interaction networks, are inferred from data with a clear quantitative sense of the accuracy of the predictions. We believe

The Challenges of Systems Biology: Ann. N.Y. Acad. Sci. 1158: ix–xii (2009).
doi: 10.1111/j.1749-6632.2009.04470.x

this is critical to convince biologists that computational predictions are becoming as sensitive and specific as directed high-throughput assays and to create increasing symbiosis between the computational and experimental communities. To accomplish this goal, DREAM challenges systems-biology researchers to predict biological networks that have already been accurately mapped and validated—"gold standard" networks—but that are known only to a few evaluators who do not participate in the analysis. These evaluators could then compare the actual network with the predicted ones in order to assess, in a truly objective fashion, the extent to which reverse engineering algorithms deliver on their promise to quickly and accurately unravel the topology of cellular interactions.

On December 3 and 4, 2007, the New York Academy of Sciences hosted the second DREAM meeting, which came to be known as DREAM2. The conference featured plenary and invited talks, as well as contributed talks and posters, illuminating various aspects of present day systems biology research. The centerpiece of the DREAM2 meeting was a set of five "challenges"—or DREAM2-challenges—that constituted the first set of challenges to assess the ability of scientists to infer networks from experimental data, by comparing their predictions to gold-standard networks whose structure is known (the fact that the first set of challenges got to be known as the DREAM2 challenges is a somehow humorous incongruence that has generated many a misunderstanding). For the DREAM2 challenges, participants tried to replicate various types of known networks from specified data. The five challenges included identifying targets of the transcription factor BCL6, determining which proteins of a group interact, and inferring the topology of a variety of networks, including a five-gene synthetic network in yeast, several more complex, computer-generated networks, and a documented gene regulatory network in a bacterium.

Also in 2007, ENFIN and the DREAM initiative joined forces to co-organize a conference in Europe. ENFIN and DREAM investigators and other researchers gathered at the Spanish National Cancer Research Centre (CNIO) in Madrid, on April 28–29, 2008 for a conference on the assessment of computational methods in systems biology. Different groups tackling the issue of assessment of bioinformatics methods presented their approaches: the ENFIN consortium, the CASP experiment (Critical Assessment of Techniques for Protein Structure Prediction), and DREAM.

The chapters in this book, which have been selected from a large number of submissions to the aforementioned conferences in a peer review process by a distinguished program committee, address the contents of some of the most salient topics discussed in the DREAM2 and the ENFIN-DREAM conferences. The volume is divided into two parts. Part I features two papers selected from the ENFIN-DREAM conference, followed by papers selected from the DREAM2 conference submissions that discuss mathematical and algorithmic techniques to analyze transcription, signaling, and metabolic pathways. The last papers of Part I provide a discussion of the state of the art in reverse engineering biological networks—its theory, methods, and tools. The second half of the volume is devoted to the DREAM2 challenges. The first chapter of Part II describes the five DREAM2 challenges and provides a comparative study of the more than 100 submissions that we received in response to the challenges. The remaining chapters of Part II, which form the core of the second half of the volume, present the best performing teams discussing their best performing strategies.

This volume would not have been possible without the encouragement, support, and dedication of a number of friends and colleagues. The ENFIN-DREAM conference was ably hosted at the CNIO by Alfonso Valencia and Ana Rojas Mendoza. The New York Academy of Sciences has nurtured the DREAM initiative from its inception, and we thank its president, Ellis Rubinstein, for his encouragement, and the former NYAS director of the Frontiers of Science Program, Jeremy Paul, for his genuine interest and concrete help with the logistics of the organization of the DREAM2 conference.

Without the scientists who generously provide the data and help us conceive the challenges, the whole DREAM concept would be impossible. For the DREAM2 challenges, we were extremely fortunate to receive support and intellectual advise from Andrea Califano, Riccardo Dalla Favera, Kai Wang, and Adam Margolin (the BCL6 target inference challenge); Joel Bader, Marc Vidal, Fritz Roth, and Haiyuan Yu (the protein–protein interaction network challenge); Diego di Bernardo, Maria Pia Cosma, and Irene Cantone (the five-gene network challenge); Pedro Mendes (the *in silico* network challenge); and Tim Gardner and Boris Hayete (the genome-scale challenge). We counted on an extremely capable program committee to select the papers for the conferences and to help improve the best performance papers: Alberto Ambesi, Gary Bader, Mukesh Bansal, Andrea Califano, Joaquin Dopazo, Edward Dougherty, Frank Doyle, Tim Elston, Jean-Loup Faulon, Alexander Hartemink, Pablo Iglesias, In Sock Jang, Igor Jurisica, Winfried Just, Pascal Kahlem, Yannis Kevrekidis, Seungchan Kim, Yuval Kluger, Christina Leslie, Andre Levchenko, Wei Keat Lim, Avi Ma'ayan, Kathleen Marchal, Adam Margolin, Luis Mendoza, Satoru Miyano, Madan-Babu Mohan, Ilya Nemenman, Theodore Perkins, Timothy Ravasi, Jeremy Rice, Fritz Roth, Michael Samoilov, Ilya Shmulevich, Mona Singh, Eduardo Sontag, Brandilyn Stigler, Pavel Sumazin, Jesper Tegnér, Denis Thieffry, Vesteinn Thorsson, Yuhai Tu, Jose Vilar, John Wagner, Kai Wang, Harel Weinstein, Ioannis Xenarios, and Sean Zhou.

The DREAM2 conference ran smoothly thanks to the constant problem solving and professionalism of Desi Tahiraj from Columbia University and Crystal Ocampo from the New York Academy of Sciences. A considerable amount of information technology infrastructure was built for the DREAM2 challenges. The challenge registration process, format checking, and prediction submissions ran successfully, thanks to the dedication of Ethan Fuchs, Bernd Jagla, and Aris Floratos. This volume was made possible thanks to the constant support and encouragement from Kirk Jensen and his extremely able team, Steve Bohall and Ralph Brown. Special thanks go to the sponsors of the DREAM2 conference: the NIH Roadmap Initiative, the Columbia University Center for Multiscale Analysis Genomic and Cellular Networks (MAGNet), the IBM Computational Biology Center, Merck Research Laboratories, Pfizer, and the New York Academy of Sciences Systems Biology Discussion Group.

Finally, our heartfelt thanks go to the 79 teams that downloaded the DREAM2 data and the 36 teams that submitted the 110 predictions. Without this community of researchers, DREAM2 would not exist.

The Challenges of Systems Biology: Community Efforts to Harness Biological Complexity constitutes an overview of the state of the art in important subdisciplines within systems biology. It was conceived to provide a clearer understanding of the strengths and weaknesses of

algorithm development, the art and science of modeling in present day biology, and the role of community efforts in current day information-intensive biology.

GUSTAVO STOLOVITZKY
IBM Computational Biology Center
Yorktown Heights, New York

PASCAL KAHLEM
European Bioinformatics Institute
Cambridge, United Kingdom

ANDREA CALIFANO
Columbia University, New York, New York