

Lecture Notes in Bioinformatics

4023

Edited by S. Istrail, P. Pevzner, and M. Waterman

Editorial Board: A. Apostolico S. Brunak M. Gelfand
T. Lengauer S. Miyano G. Myers M.-F. Sagot D. Sankoff
R. Shamir T. Speed M. Vingron W. Wong

Subseries of Lecture Notes in Computer Science

Eleazar Eskin Trey Ideker
Ben Raphael Christopher Workman (Eds.)

Systems Biology and Regulatory Genomics

Joint Annual RECOMB 2005 Satellite Workshops
on Systems Biology and on Regulatory Genomics
San Diego, CA, USA, December 2-4, 2005
Revised Selected Papers

Series Editors

Sorin Istrail, Brown University, Providence, RI, USA

Pavel Pevzner, University of California, San Diego, CA, USA

Michael Waterman, University of Southern California, Los Angeles, CA, USA

Volume Editors

Eleazar Eskin

Ben Raphael

University of California

Department of Computer Science

San Diego, CA, USA

E-mail: {eeskin,braphael}@cs.ucsd.edu

Trey Ideker

Christopher Workman

University of California

Department of Bioengineering

San Diego, CA, USA

E-mail: {trey,cworkman}@bioeng.ucsd.edu

Library of Congress Control Number: 2006940071

CR Subject Classification (1998): F.2, G.3, E.1, H.2.8, J.3

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-540-48293-8 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-48293-2 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2007

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11916338 06/3142 5 4 3 2 1 0

Preface

It has become increasingly evident that the use of large-scale experimental data and the application of principles from systems biology are gaining widespread acceptance in mainstream biology. Systems biology involves the use of global cellular measurements, i.e., genomic, proteomic, and metabolomic, to construct computational models of cellular processes and disease. These approaches involve an integration of experimental and computational techniques and may include: 1) developing models of cellular processes, 2) measuring the response to perturbations of model components, and 3) iteratively formulating and testing new hypotheses for unexpected observations.

A research area that is particularly important to systems biology is the study of gene regulatory networks. Although genome sequencing efforts have been tremendously successful, much is unknown about the regulation of these sequenced genomes. Automatic methods for helping decipher the regulatory mechanism are crucial for understanding the regulatory network as a whole. However, many new challenges are presented when analyzing complete genomes. These challenges include motif discovery in large genomes, leveraging information from multiple genomes, detection of weak signals and incorporating different types of genomic data such as protein localization data and gene expression. Novel methodology will be particularly relevant given the hypothesis that the observed phenotypic differences between organisms with very similar genomes may be due to variations in the gene regulation.

The amount of research in both of these areas has exploded in recent years, as witnessed by the number of research presentations at meetings such as RECOMB, ISMB, PSB, the Biopathways Consortium, and ICSB. The jointly held RECOMB Satellite on Systems Biology and the RECOMB Satellite on Regulatory Genomics provide a forum for addressing these challenges.

This year's workshops also included a special session on Computational Developmental Biology organized by David Gifford. Unique challenges posted by developmental biology include: 1) computational model representations that can express the execution of complex natural programs over time, 2) the identification of key developmental state variables and experimental methods to reliably observe these variables, and 3) the use of computational models to understand developmentally related disease, and to help develop therapeutics including the programming of stem cells as therapeutic agents. Invited speakers at this year's special session addressed the question: "What key developmental biology problems can now be examined from a systems biology perspective, and what data are necessary to do so?". The goal of this special session was to help computational and systems biologists understand both the challenge and excitement of working on development.

Organization

Steering Committee: Regulatory Genomics

Pierre Baldi	University of California, Irvine
Michael Eisen	Lawrence Berkeley National Lab
Eleazar Eskin	University of California, San Diego
Pavel Pevzner	University of California, San Diego

Steering Committee: Systems Biology

Leroy Hood	Institute for Systems Biology
Trey Ideker	University of California, San Diego
Douglas Lauffenburger	MIT
Satoru Miyano	University of Tokyo
Ron Shamir	Tel Aviv University

Organizing Committee

Eleazar Eskin	University of California, San Diego
David Gifford	Massachusetts Institute of Technology
Trey Ideker	University of California, San Diego
Teresa M. Przytycka	NIH/NLM/NCBI
Ben Raphael	University of California, San Diego
Cenk Sahinalp	Simon Fraser University
Samantha Smeraglia	University of California, San Diego
Christopher Workman	University of California, San Diego

Program Committee

John Aitchison	Institute for Systems Biology
Adam Arkin	University of California, Berkeley
Gary Bader	Memorial Sloan-Kettering Cancer Center
Pierre Baldi	University of California, Irvine
Yoseph Barash	Hebrew University
Mathieu Blanchette	McGill University
Gal Chechik	Stanford University
Francis Chin	University of Hong Kong
Michael Cusick	Harvard Medical School
Eric Davidson	CalTech
Michael Eisen	Lawrence Berkeley National Lab
Eleazar Eskin	University of California, San Diego

Nir Friedman	Hebrew University
Irit Gat-Vicks	Tel Aviv University
Mikhail Gelfand	Moscow State University
David Gifford	Massachusetts Institute of Technology
Sridhar Hannenhalli	University of Pennsylvania
Jeff Hasty	University of California, San Diego
Leong Hon Wai	National University of Singapore
Leroy Hood	Institute for Systems Biology
Trey Ideker	University of California, San Diego
Richard Karp	University of California, Berkeley
Uri Keich	Cornell University
Douglas Lauffenburger	Massachusetts Institute of Technology
Christina Leslie	Columbia University
Mike Levine	University of California, Berkeley
Hao Li	University of California, San Francisco
Nick Luscombe	European Bioinformatics Institute
Satoru Miyano	University of Tokyo
Dana Pe'er	Harvard Medical School
Pavel Pevzner	University of California, San Diego
Tzachi Pilpel	Weizmann Institute of Science
Teresa M. Przytycka	NIH/NLM/NCBI
Ben Raphael	University of California San Diego
Bing Ren	University of California, San Diego
Aviv Regev	Harvard Medical School
Mireille Regnier	INRIA
Marie-France Sagot	INRIA
Cenk Sahinalp	Simon Fraser University
Eran Segal	Rockefeller University
Ron Shamir	Tel Aviv University
Roded Sharan	Tel Aviv University
Amos Tanay	Tel Aviv University
Alfonso Valencia	Centro Nacional de Biotecnologia
Marc Vidal	Harvard Medical School
Christopher Workman	University of California, San Diego
Eric Xing	Carnegie Mellon University
Zohar Yakhini	Agilent
Ralf Zimmer	LMU, Institut für Informatik

Sponsoring Institutions

Industry-University Cooperative Research Program, The UC Discovery Grant
 California Institute for Telecommunications and Information Technology, Cal-(IT)²

Table of Contents

An Interactive Map of Regulatory Networks of <i>Pseudomonas aeruginosa</i> Genome	1
<i>Weihui Wu, Yongling Song, Shouguang Jin, and Su-Shing Chen</i>	
The Pathalyzer: A Tool for Analysis of Signal Transduction Pathways	11
<i>David L. Dill, Merrill A. Knapp, Pamela Gage, Carolyn Talcott, Keith Laderoute, and Patrick Lincoln</i>	
Decomposition of Overlapping Protein Complexes: A Graph Theoretical Method for Analyzing Static and Dynamic Protein Associations	23
<i>Elena Zotenko, Katia S. Guimarães, Raja Jothi, and Teresa M. Przytycka</i>	
Comparison of Protein-Protein Interaction Confidence Assignment Schemes	39
<i>Silpa Suthram, Tomer Shlomi, Eytan Ruppin, Roded Sharan, and Trey Ideker</i>	
Characterization of the Effects of TF Binding Site Variations on Gene Expression Towards Predicting the Functional Outcomes of Regulatory SNPs	51
<i>Michal Lapidot and Yitzhak Pilpel</i>	
Comparative Systems Biology of the Sporulation Initiation Network in Prokaryotes	62
<i>Michiel de Hoon and Dennis Vitkup</i>	
Improvement of Computing Times in Boolean Networks Using Chi-square Tests	70
<i>Haseong Kim, Jae K. Lee, and Taesung Park</i>	
Build a Dictionary, Learn a Grammar, Decipher Stegoscripts, and Discover Genomic Regulatory Elements	80
<i>Guandong Wang and Weixiong Zhang</i>	
Causal Inference of Regulator-Target Pairs by Gene Mapping of Expression Phenotypes	95
<i>David Kulp and Manjunatha Jagalur</i>	
Examination of the tRNA Adaptation Index as a Predictor of Protein Expression Levels	107
<i>Orna Man, Joel L. Sussman, and Yitzhak Pilpel</i>	

Improved Duplication Models for Proteome Network Evolution	119
<i>Gürkan Bebek, Petra Berenbrink, Colin Cooper, Tom Friedetzky, Joseph H. Nadeau, and S. Cenk Sahinalp</i>	
Application of Expectation Maximization Clustering to Transcription Factor Binding Positions for cDNA Microarray Analysis	138
<i>Chih-Yu Chen, Von-Wun Soo, and Chi-Li Kuo</i>	
Combinatorial Genetic Regulatory Network Analysis Tools for High Throughput Transcriptomic Data	150
<i>Elissa J. Chesler and Michael A. Langston</i>	
Topological Robustness of the Protein-Protein Interaction Networks	166
<i>Chien-Hung Huang, Jywe-Fei Fang, Jeffrey J.P. Tsai, and Ka-Lok Ng</i>	
A Bayesian Approach for Integrating Transcription Regulation and Gene Expression: Application to <i>Saccharomyces Cerevisiae</i> Cell Cycle Data	178
<i>Sudhakar Jonnalagadda and Rajagopalan Srinivasan</i>	
Probabilistic <i>in Silico</i> Prediction of Protein-Peptide Interactions	188
<i>Wolfgang Lehrach, Dirk Husmeier, and Christopher K.I. Williams</i>	
Improved Pattern-Driven Algorithms for Motif Finding in DNA Sequences	198
<i>Sing-Hoi Sze and Xiaoyan Zhao</i>	
Annotation of Promoter Regions in Microbial Genomes Based on DNA Structural and Sequence Properties	212
<i>Huiquan Wang and Craig J. Benham</i>	
An Interaction-Dependent Model for Transcription Factor Binding	225
<i>Li-San Wang, Shane T. Jensen, and Sridhar Hannenhalli</i>	
Computational Characterization and Identification of Core Promoters of MicroRNA Genes in <i>C. elegans</i> , <i>H. sapiens</i> and <i>A. thaliana</i>	235
<i>Xuefeng Zhou, Jianhua Ruan, Guandong Wang, and Weixiong Zhang</i>	
A Comprehensive Kinetic Model of the Exocytotic Process: Evaluation of the Reaction Mechanism	249
<i>Aviv Mezer, Eran Bosis, Uri Ashery, Esther Nachliel, and Menachem Gutman</i>	
Author Index	259