Life Sciences Society

COMPUTATIONAL SYSTEMS BIOINFORMATICS

CSB2007 CONFERENCE PROCEEDINGS

Volume 6

University of California
San Diego, USA
13–17 August 2007

Editors

Peter Markstein
Ying Xu

Imperial College Press

Life Sciences Society

COMPUTATIONAL SYSTEMS BIOINFORMATICS

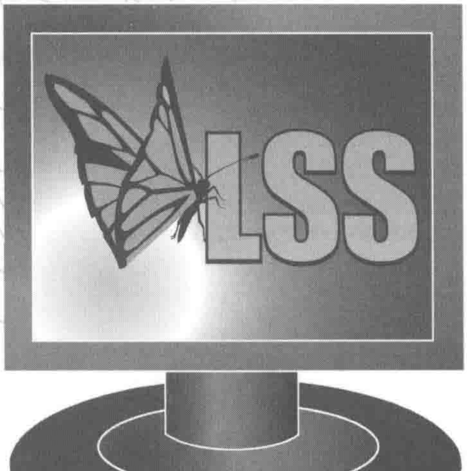
CSB2007 CONFERENCE PROCEEDINGS
Volume 6

University of California
San Diego, USA
13–17 August 2007

Editors

Peter Markstein
in silico Labs, LLC, USA

Ying Xu
University of Georgia, USA



ICP
Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

COMPUTATIONAL SYSTEMS BIOINFORMATICS
Proceedings of the Conference CSB 2007 — Vol. 6

ISBN-13 978-1-86094-872-5
ISBN-10 1-86094-872-3

Printed by FuIsland Offset Printing (S) Pte Ltd, Singapore

Life Sciences Society

COMPUTATIONAL SYSTEMS BIOINFORMATICS

Life Sciences Society

Thank You

CSB2007 Gold Sponsor

The Life Sciences Society, LSS Directors,
together with the CSB2007 Program Committee and
Conference Organizing Committee are extremely grateful to the

Hewlett-Packard Company

for their Gold Sponsorship of the
Sixth Annual Computational Systems Bioinformatics Conference, CSB2007
at University of California San Diego,
La Jolla, California, August 13-17, 2007

hp ®
invent

PREFACE

The 21st Century has seen the emergence of tremendous vigor and excitement at the interface between computing and biology. Following years of investment by industry and then private foundations, the US Federal Government has greatly increased its support. Increasingly, the experimental findings from all of the biological sciences are becoming data rich and their practitioners are turning to the use of computational methods to manage and analyze the data. In light of the growing opportunities and excitement at the frontier interface between computing and biology, a few scientists turned conference planners organized the first Computational Systems Bioinformatics (CSB) conference in 2001 at Stanford, CA; CSB continued each August at Stanford over the next five years. During this time, many computer scientists and other engineers, as well as biologists, have attended CSB meetings, which have particularly served to introduce cutting edge biological inquiry and challenge problems to investigators from quantitative science backgrounds.

CSB, more recently, became the public venue for the not-for-profit Life Sciences Society, or LSS, which was founded, in part, to enhance the opportunities at the interface between the quantitative / engineering sciences and the biological sciences. In 2006, LSS was honored to be invited to hold CSB2007 on the campus of the University of California at San Diego, UCSD. Some future meetings at UCSD, and ultimately, at other universities, as well as satellite sessions at bioinformatics meetings around the world, are anticipated over the next several years. The Stanford / Bay Area / Silicon Valley venue, along with presenting highlights in bioinformatics, has been especially valuable for connecting individuals from the computer and electronics industry with investigators in the pure and applied life sciences.

The current venue should provide some connections to telecommunications, while sustaining some of the earlier opportunities, but we anticipate an enhanced interaction with basic and applied biotechnology. The University of California at San Diego grew from the Scripps Institute of Oceanography and began as a graduate school with a strong focus on the natural sciences. The rich research culture around UCSD and many neighboring institutions would generally be termed the venue of the Torrey Pines Mesa, an area very rich in biotechnology research activities. Today, San Diego has the largest cluster of Life Sciences centers with 26 research institutes (including UCSD and a suite of Institutes: Salk, Neurosciences, Scripps Research, Burnham Medical, as well as smaller not for profits) located in an area less then 10 square miles. For more on San Diego's R&D Life Sciences Centers and the vibrant biotechnology and pharmacology efforts, do visit the BioCom website: **http://www.biocom.org/Portals/0/SanDiegoLifeScienceNumbers_Fall06.pdf**.

CSB2007 will continue to be a 5 day single track conference, with three core plenary presentations days sandwiched between a day of practical tutorials, long a very popular feature, and a day workshops exploring the future. Thus, CSB2007 includes several half day tutorials, 30 refereed papers plus keynote and invited speakers, and posters, during its five full days. Special events for the evenings are planned.

CSB2007, as in each of its previous years, owes a lot to its many hard working volunteers, who are listed under the Committees. The indefatigable energy of Vicky and Peter Markstein continues to sustain the magnitude and amplitude of the extraordinary science and technology vector that is CSB, and their partnership with Ying Xu also remains essential. The efforts to manage and enhance local arrangements, by Kayo Arima, Patrick Shih, Lydia Grech and Ed Buckingham, should also be acknowledged.

A few words, naturally, about SoCal: bring family and guests to enjoy San Diego's world-famous attractions as SeaWorld, the San Diego Zoo, the Wild Animal Park and LEGOLAND California, as well as historic cultural gems Balboa Park and Old Town, and of course, the “endless” beach.

John Wooley, General Conference Chair

COMMITTEES

Steering Committee

Phil Bourne - University of California, San Diego
Eric Davidson - California Institute of Technology
Steven Salzberg - The Institute for Genomic Research
John Wooley - University of California San Diego, San Diego Supercomputer Center

Organizing Committee

Kayo Arima – Universitye of California San Diego, Local Arangements
Pat Blauvelt - Communications
Ed Buckingham – LSS VP Conferences
Kass Goldfein - Finance Consultant
Lydia Grech – University of California San Diego, Local Arrangements
Fenglou Mao – University of Georgia, On-Line Registration and Refereeing Website
Vicky Markstein - Life Sciences Society, **Co-Chair**, LSS President
Patrick Shih – University of California San Diego, Local Arrangements
Jean Tsukamoto - Graphics Design
Bill Wang - Sun Microsystems Inc, LSS Information Technology
John Wooley - University of California San Diego, San Diego Supercomputer Center, **Co-Chair**

Program Committee

Tatsuya Akutsu - Kyoto University
Phil Bourne – University of California San Diego
Jake Chen - Indiana University
Amar Das - Stanford University
Chris Ding – Lawrence Berkeley Laboratory
Roderic Guigo, IMIM, Barcelona
Tao Jiang - University of California Riverside
Lydia Kavraki – Rice University
Hoong-Chien Lee - National Central University, Taiwan
Ann Loraine - University of Alabama
Michele Markstein – Harvard University
Peter Markstein - Hewlett-Packard Co., **Co-chair**
Satoru Miyano - University of Tokyo
Sean Mooney - Indiana University
Jan Mrazek - University of Georgia
Isidore Rigoutsos - IBM TJ Watson Research Center
Andrey Rzhetsky - Columbia University

Hershel M. Safer, Weizmann Institute of Science
David States - University of Michigan
Anna Tramontano - University of Rome
Olga Troyanskaya - Princeton University
Alfonso Valencia - Centro Nacional de Biotecnologia, Spain
Eberhard Voit - Georgia Tech
Limsoon Wong - Institute for Infocomm Research
Ying Xu - University of Georgia, **Co-chair**
Aidong Zhang - SUNY Buffalo
Michael Zhang - Cold Spring Harbor Laboratory
Xianghong Jasmine Zhou - University of Southern California
Yaoqi Zhou - Indiana University

Assistants to the Program Co-Chairs

Ann Terka – University of Georgia
Joan Yantko – University of Georgia

Poster Committee

Nigam Shah - Stanford University, **Chair**
Patrick Shih – University of California San Diego

Tutorial Committee

Weizhong Li – University of California San Diego,
Al Shpuntoff – Syngenta Biotechnology Institute, **Chair**
John Wooley – University of California San Diego

Workshop Committee

Iddo Friedberg – University of California San Diego, **Co-Chair**
Weizhong Li – University of California San Diego, **Co-Chair**
Patrick Shih – University of California San Diego

REFEREES

Tatsuya Akutsu
Mar Alba

Takis Benos

Phil Bourne

Liming Cai
Ildefonso Cases
Robert Castelo
Dongsheng Che
Jake Chen
Liang Chen
David Chew
Young-Rae Cho
I-Chun Chou
Xiangqin Cui

PhuongAn Dam
Amar Das
David de Juan
Chris Ding

Iakes Ezkurdia

Matteo Floris
Sylvain Foissac

David Gilley
Gautam Goel
Roderic Guigo

Scott Harrison
Nurit Haspel
Jianjun Hu
Woochang Hwang

Seiya Imoto

Tao Jiang

Yuki Kato
Lydia Kavraki
Melissa Kemp

HC Lee
Hoong-Chien Lee
Haiquan Li
Jing Li
Xiaoman Shawn Li
Guohui Lin
Chun-Chi Liu
Guimei Liu
Huiqing Liu
Yunlong Liu
Ann Loraine

Michia Ma
Fenglou Mao
Peter Markstein
David Martin
Satoru Miyano
Mark Moll
Jan Mrazek

Masao Nagasaki
Luay Nakleh
Christoforous Nikolau
Juan Nunez-Iglesias

Victor Olman

Miguel Padilla
Grier Page
Florencio Pazos
Daniel Platt

Zhen Qi

Predrag Radivojac
Isidore Rigoutsos

Andrey Rzhetsky

Hershel M. Safer
Sudipto Saha
David States
Wing-Kin Sung

Takeyuki Tamura
Anna Tramontano
Olga Troyanskaya
Aristotelis Tsirigos

Alfonso Valencia
Siren Veflingstad
Eberhard Voit

John Wagner
Mingyi Wang
Limsoon Wong
Hongwei Wu
Jialiang Wu

Min Xu
Ying Xu

Weiwei Yin

Kangyu Zhang
Michael Zhang
Shiju Zhang
Fengfeng Zhou
Ruhong Zhou
Wen Zhou
Xianghong Jasmine Zhou
Yaoqi Zhou

CONTENTS

Biomedical Application

Pathways, Networks and Systems Biology

Genomics

Structural Bioinformatics

Ontology, Database and Text Mining

Biocomputing

Transcriptomics and Phylogeny

Computational Systems Bioinformatics 2007

Keynote Address

QUANTITATIVE ASPECTS OF GENE REGULATION IN BACTERIA: AMPLIFICATION, THRESHOLD, AND COMBINATORIAL CONTROL

Terry Hwa

Center for Theoretical Biological Physics and Department of Physics
University of California San Diego
9500 Gilman Drive
La Jolla, CA 92093-0374

Biological organisms possess an enormous repertoire of genetic responses to ever-changing combinations of cellular and environmental signals. Unlike digital electronic circuits however, signal processing in cells is carried out by a limited number of asynchronous devices in fluctuating aqueous environments.

In this talk, I will discuss the control of genetic responses in bacteria. Theoretical analysis of the known mechanisms of transcriptional control suggests "programmable" mechanisms for implementing a broad class of combinatorial control. Further analysis of post-transcriptional control suggests mechanisms for signal amplification, threshold response, and noise attenuation. I will present experimental characterization of some of these bio-computational "devices", as well as experiments illustrating how promoter sequences may be "trained" by directed evolution. Quantitative characterization and controlled manipulation of these devices may bring about predictive understanding of biological control systems, and reveal interesting, novel strategies of distributed computation.