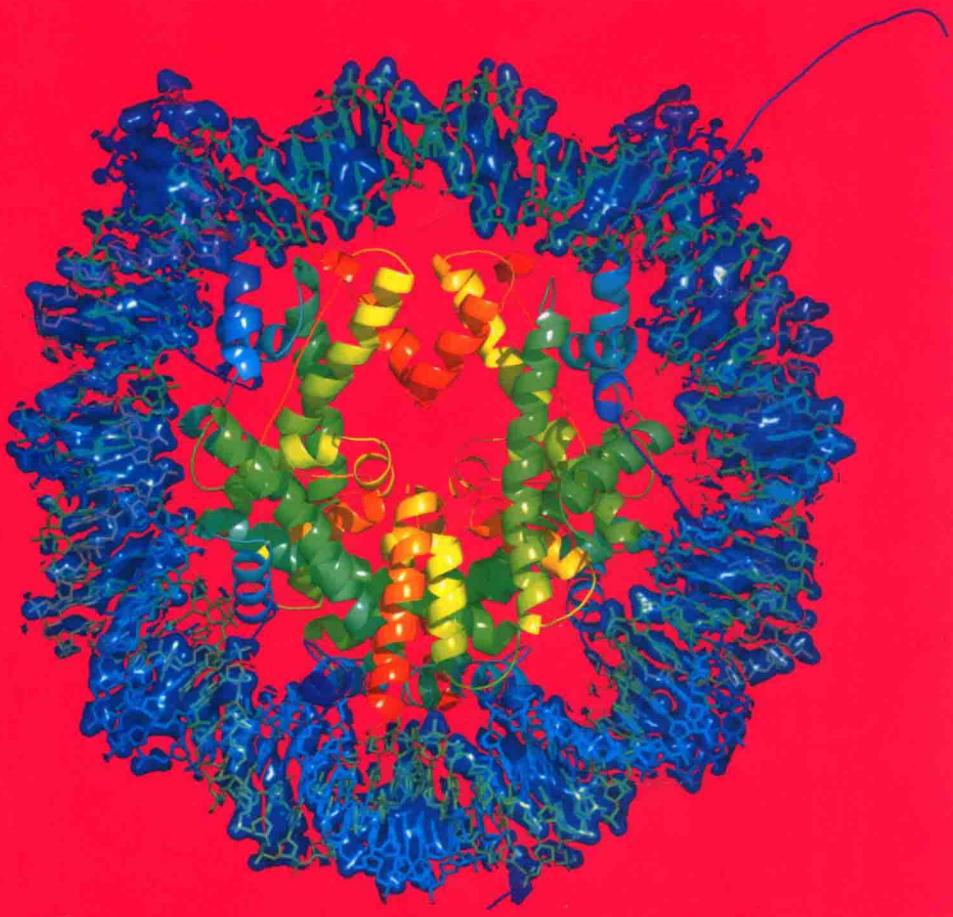
PACIFIC SYMPOSIUM ON BIOCOMPUTING 2007



Edited by

Russ B. Altman, A. Keith Dunker,
Lawrence Hunter, Tiffany Murray & Teri E. Klein

PACIFIC SYMPOSIUM ON BIOCOMPUTING 2007

Maui, Hawaii
3–7 January 2007

Edited by

Russ B. Altman
Stanford University, USA

A. Keith Dunker
Indiana University, USA

Lawrence Hunter
University of Colorado Health Sciences Center, USA

Tiffany Murray
Stanford University, USA

Teri E. Klein
Stanford University, USA

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published 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.

BIOCOMPUTING 2007

Proceedings of the Pacific Symposium

ISBN 981-270-417-5

Printed in Singapore by Mainland Press

PACIFIC SYMPOSIUM ON BIOCOMPUTING 2007

PACIFIC SYMPOSIUM ON BIOCOMPUTING 2007

Biomedical computing has become a key component in the biomedical research infrastructure. In 2004 and 2005, the U.S. National Institutes of Health established seven National Centers for Biomedical Computation, focusing on a wide range of application areas and enabling technologies, including simulation, systems biology, clinical genomics, imaging, ontologies and others (see http://www.bisti.nih.gov/ncbc/). The goal of these centers is to help seed an information infrastructure to support biomedical research. The Pacific Symposium on Biocomputing (PSB) presented critical early sessions in most of the areas covered by these National Centers, and we are proud to continue the tradition of helping to define new areas of focus within biomedical computation.

Once again, we are fortunate to host two outstanding keynote speakers. Dr. Elizabeth Blackburn, Professor of Biology and Physiology in the Department of Biochemistry and Biophysics at the University of California, San Francisco will speak on "Interactions among telomeres, telomerase, and signaling pathways." Her work has led our understanding of overall organization and control of chromosomal dynamics. Our keynote speaker in the area of Ethical, Legal and Social implications of technology will be Marc Rotenberg, Executive Director of the Electronic Privacy Information Center (EPIC) in Washington, D.C. He will speak on "Data mining and privacy: the role of public policy." Many biomedical computation professionals have had and continue to grapple with privacy issues as interest in mining human genotype-phenotype data collections has increased.

PSB has a history of providing early sessions focusing on hot new areas in biomedical computation. These sessions are often conceived during the previous PSB meeting, as trends and new results are pondered and discussed. Very often, new sessions are lead by new faculty members trying to define a scientific niche and bring together leaders in the emerging areas. We are proud that many areas in biocomputing received their first significant focused attention at PSB. If you have an idea for a new session, we the organizers, are available to talk with you, either at the meeting or later by e-mail.

Again, the diligence and efforts of a dedicated group of researchers has led to an outstanding set of sessions, with associated introductory tutorials. These organizers provide the scientific core of PSB, and their sessions are as follows:

Indra Neil Sarkar
Biodiversity Informatics: Managing Knowledge Beyond Humans and Model Organisms

Bobbie-Jo Webb-Robertson & Bill Cannon
Computational Proteomics: High-throughput Analysis for Systems Biology

Martha Bulyk, Ernest Fraenkel, Alexander Hartemink, & Gary Stormo
DNA-Protein Interactions and Gene Regulation: Integrating Structure, Sequence and Function

Russ Greiner & David Wishart
Computational Approaches to Metabolomics

Pierre Zweigenbaum, Dina Demner-Fushman, Kevin Bretonnel Cohen, & Hong Yu
New Frontiers in Biomedical Text Mining

Maricel Kann, Yanay Ofran, Marco Punta, & Predrag Radivojac
Protein Interactions in Disease

In addition to the sessions and survey tutorials, this year's program includes two in depth tutorials. The presenters and titles of these tutorials are:

Giselle M. Knudsen, Reza A. Ghiladi, & D. Rey Banatao
Integration Between Experimental and Computational Biology for Studying Protein Function

Michael A Province & Ingrid B Borecki
Searching for the Mountains of the Moon: Genome Wide Association Studies of Complex Traits

We thank the Department of Energy and the National Institutes of Health for their continuing support of this meeting. Their support provides travel grants to many of the participants. Applied Biosystems and the International Society for Computational Biology continue to sponsor PSB, and as a result, we are able to provide additional travel grants to many meeting participants.

We would like to acknowledge the many busy researchers who reviewed the submitted manuscripts on a very tight schedule. The partial list following this preface does not include many who wished to remain anonymous, and of course we apologize to any who may have been left out by mistake.

Aloha!

Russ B. Altman
Departments of Genetics & Bioengineering, Stanford University

A. Keith Dunker
Department of Biochemistry and Molecular Biology, Indiana University School of Medicine

Lawrence Hunter
Department of Pharmacology, University of Colorado Health Sciences Center

Teri E. Klein
Department of Genetics, Stanford University

Pacific Symposium on Biocomputing Co-Chairs
September 28, 2006

Thanks to the reviewers...

Finally, we wish to thank the scores of reviewers. PSB requires that every paper in this volume be reviewed by at least three independent referees. Since there is a large volume of submitted papers, paper reviews require a great deal of work from many people. We are grateful to all of you listed below and to anyone whose name we may have accidentally omitted or who wished to remain anonymous.

Joshua Adkins
Eugene Agichtein
Gelio Alves
Sophia Ananiadou
Alan Aronson
Ken Baclawski
Joel Bader
Breck Baldwin
Ziv Bar-Joseph
Serafim Batzoglou
Asa Ben-Hur
Sabine Bergler
Olivier Bodenreider
Alvis Brazma
Kevin Bretonnel
Yana Bromberg
Harmen Bussemaker
Andrea Califano
Bob Carpenter
Michele Cascella
Saikat Chakrabarti
Shih-Fu Chang
Pierre Chaurand
Ting Chen
Hsinchun Chen
Nawei Chen
Praveen Cherukuri
Wei Chu
James Cimino
Aaron Cohen
Nigel Collier
Matteo Dal Peraro
Vlado Dancik
Rina Das
Tjil De Bie
Dina Demner-Fushman
Rob DeSalle
Luis DeSilva
Diego Di Bernardo
Chuong Do
Michel Dumontier
Mary G. Egan
Roman Eisner
Emilio Espisitio
Mark Fasnacht
Oliver Fiehn
Alessandro Flammini
Fabian Fontaine
Lynne Fox
Ari Frank
Kristofer Franzen
Tema Fridman
Carol Friedman
Robert Futrelle
Feng Gao
Adam Godzik
Roy Goodacre
Michael Grusak
Melissa A. Haendel
Henk Harkema
Marti Hearst
P. Bryan Heidorn
Bill Hersh
Lynette Hirschman
Terence Hwa
Sven Hyberts
Lilia Iakoucheva
Navdeep Jaitly
Helen Jenkins
Kent Johnson
Andrew Joyce
James Kadin
Martin R. Kalfatovic
Manpreet S. Katari
Sun Kim
Oliver King
Tanja Kortemme
Harri Lahdesmaki
Ney Lemke
Gondy Leroy
Christina Leslie
Li Liao
John C. Lindon
Chunmei Liu
Yves Lussier
Hongwu Ma
Kenzie MacIsaac
Tom Madej
Ana Maguitman
Askenazi Manor
Costas Maranas
Leonardo Marino
John Markley
Pedro Mendes
Ivana Mihalek

Leonid Mirny
Joyce Mitchell
Matthew Monroe
Sean Mooney
Rafael Najmanovich
Preslav Nakov
Leelavati Narlikar
Adeline Nazarenko
Jack Newton
William Noble
Christopher Oehmen
Christopher Oldfield
Zoltan Oltvai
Matej Oresic
Bernhard Palsson
Chrysanthi Paranavitana
Matteo Pellegrini
Aloysius Phillips
Paul J. Planet
Christian Posse
Natasa Przulj
Teresa Przytycka
Bin Qian
Weijun Qian
Arun Ramani
Kathryn Rankin
Andreas Rechtsteiner
Haluk Resat
Tom Rindflesch
Martin Ringwald
Elizabeth Rogers
Pedro Romero
Graciela Rosemblat
Andrea Rossi
Erik Rytting
Jasmin Saric
Indra Neil Sarkar
Yutaka Sasaki
Tetsuya Sato
Santiago Schnell
Rob Schumaker
Robert D. Sedgewick
Eran Segal
Kia Sepassi
Anuj Shah
Paul Shapshak
Hagit Shatkay
Mark Siddall
Mona Singh
Mudita Singhal
Saurabh Sinha
Thereza Amelia Soares
Bruno Sobral
Ray Sommorjai
Orkun Soyer
Irina Spasic
Padmini Srinivasan
Paul Stothard
Eric Strittmatter
Shamil Sunyaev
Silpa Suthram
Lorrie Tanabe
Haixu Tang
Igor Tetko
Jun'ichi Tsujii
Peter Uetz
Vladimir Uversky
Vladimir Vacic
Alfonso Valencia
Karin Verspoor
Mark Viant
K. Vijay-Shanker
Hans Vogel
Slobodan Vucetic
Alessandro Vullo
Wyeth Wasserman
Bonnie Webber
Aalim Weljie
John Wilbur
Kazimierz O. Wrzeszczynski
Dong Xu
Yoshihiro Yamanishi
Yuzhen Ye
Hong Yu
Peng Yue
Pierre Zweigenbaum

CONTENTS

COMPUTATIONAL APPROACHES TO METABOLOMICS

COMPUTATIONAL PROTEOMICS: HIGH-THROUGHPUT ANALYSIS FOR SYSTEMS BIOLOGY

DNA-PROTEIN INTERACTIONS: INTEGRATING STRUCTURE, SEQUENCE, AND FUNCTION

PROTEIN INTERACTIONS AND DISEASE

MARICEL KANN
National Center for Biotechnology Information, NIH
Bethesda, MD 20894, U.S.A.

YANAY OFRAN
Department of Biochemistry & Molecular Biophysics, Columbia University
New York, NY 10032, U.S.A.

MARCO PUNTA
Department of Biochemistry & Molecular Biophysics, Columbia University
New York, NY 10032, U.S.A.

PREDRAG RADIVOJAC
School of Informatics, Indiana University
Bloomington, IN 47408, U.S.A.

In 2003, the US National Human Genome Research Institute (NHGRI) articulated grand challenges for the genomics community in which the translation of genome-based knowledge into disease understanding, diagnostics, prognostics, drug response and clinical therapy is one of the three fundamental directions ("genomics to biology," "genomics to health" and "genomics to society").[1] At the same time the National Institutes of Health (NIH) laid out a similar roadmap for biomedical sciences.[2] Both the NHGRI grand challenges and the NIH roadmap recognized bioinformatics as an integral part in the future of life sciences. While this recognition is gratifying for the bioinformatics community, its task now is to answer the challenge of making a direct impact to the medical science and benefiting human health. Innovative use of informatics in the "translation from bench to bedside" becomes a key for bioinformaticians.

In 2005, the Pacific Symposium on Biocomputing (PSB) first solicited papers related to one aspect of this challenge, protein interactions and disease, which directly addresses computational approaches in search for the molecular basis of disease. The goal of the session was to bring together scientists interested in both bioinformatics and medical sciences to present their research progress. The session generated great interest resulting in a number of high quality papers and testable hypothesis regarding the involvement of proteins in various disease pathways. This year, the papers accepted for the session on Protein Interactions and Disease at PSB 2007 follow the same trend.

The first group of papers explored structural aspects of protein-protein interactions. Kelly et al. study ABC transporter proteins which are involved in substrate transport through the membrane. By investigating intra-transporter domain interfaces they conclude that nucleotide-binding interfaces are more conserved than those of transmembrane domains. Disease-related mutations were mapped into these interfaces. Pulim et al. developed a novel threading algorithm that predicts interactions between receptors (membrane proteins) and ligands. The method was tested on cytokines, proteins implicated in intra-cellular communication and immune system response. Novel candidate interactions, which may be implicated in disease, were predicted. Kasson and Pande use molecular dynamics to address high-order molecular organization in cell membranes. A large number of molecular dynamics trajectories provided clues into structural aspects of the insertion of about 20-residue long fusion peptide into a cell membrane by a trimer hemagglutinin of the influenza virus. The authors explain effects of mutations that preserve peptide's monomeric structure but incur loss of viral infectivity.

The second group of studies focused on analysis of protein interaction networks. Sam et al. investigate molecular factors responsible for the diseases with different causes but similar phenotypes and postulate that some are related to breakdowns in the shared protein-protein interaction networks. A statistical method is proposed to identify protein networks shared by diseases. Sridhar et al. developed an efficient algorithm for perturbing metabolic networks in order to stop the production of target compounds, while minimizing unwanted effects. The algorithm is aimed at drug development where toxicity of the drug should be reduced. Borgwardt et al. were interested in predicting clinical outcome by combining microarray and protein-protein interaction data. They use graph kernels as a measure of similarity between graphs and develop methods to improve their scalability to large graphs. Support vector machines were used to predict disease outcome. Gonzalez et al. extracted a large number of gene-disease relationships by parsing literature and mapping them to the known protein-protein interaction networks. They propose a method for ranking proteins for their involvement in disease. The method was tested on atherosclerosis. Valdivia-Granda et al. devised a method to integrate protein-protein interaction data along with other genomic annotation features with microarray data. They applied it to microarray data from a study of non-human primates infected with variola and identified early infection biomarkers. The study was complemented with a comparative protein domain analysis between host and pathogen. This work contributes to the understanding of the mechanisms of infectivity, disease and suggests potential therapeutic targets. Finally, Cook et al. worked on the novel ontology of biochemical pathways. They present Chalkboard, a tool for