PACIFIC SYMPOSIUM ON BIOCOMPUTING 2010

PACIFIC SYMPOSIUM ON BIOCOMPUTING 2010

Kamuela, Hawaii, USA
4–8 January 2010

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

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 2010

Proceedings of the Pacific Symposium

Copyright © 2010 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-4295-29-1

ISBN-10 981-4295-29-9

Printed in Singapore.

This page is intentionally left blank

PACIFIC SYMPOSIUM ON BIOCOMPUTING 2010

This year marks the 15th year of PSB. Started in 1996 by Teri Klein and Larry Hunter, the meeting was a session within the Hawaii International Conference on Systems Sciences Conference (HICSS) for a couple of years. The interest in biocomputing was great and was threatening to alter the balance of attendees within the HICSS meeting, and so the organizers asked Teri and Larry to find an alternative. Taking advantage of the nice meeting opportunities in Hawaii immediately following New Year's Day, they created PSB with a simple formula: oral presentation of peer reviewed papers in emerging areas of biocomputing, interactive poster sessions, and lively discussion sessions on these topics. They invited Keith Dunker and Russ Altman to join the team for the second PSB, and we have been working together ever since.

In fifteen years, we have had some papers that have made remarkable impact. Using Google Scholar, the top 15 papers, in terms of citations over the last 15 years are listed below (with number of citations listed as of September 2009, followed by the reference). Obviously, there are other great papers, especially ones that have been written in the last few years, that have not yet had time to rise to the top of this list. Nevertheless, the list provides highlights of the last 15 years in biocomputing challenges.

- (573) [REVEAL: A General Reverse Engineering Algorithm for Inference of Genetic Network Architectures](#) S. Liang, S. Fuhrman and R. Somogyi; Pacific Symposium on Biocomputing 3:18-29 (1998).
- (457) [Modeling Gene Expression with Differential Equations](#) T. Chen, H. L. He, and G.M. Church; Pacific Symposium on Biocomputing 4:29-40 (1999).
- (422) [Principal Components Analysis to Summarize Microarray Experiments: Application to Sporulation Time Series](#) S. Raychaudhuri, J.M. Stuart, and R.B. Altman; Pacific Symposium on Biocomputing 5:452-463 (2000).
- (419) [The Spectrum Kernel: A String Kernel for SVM Protein Classification](#) Leslie, E. Eskin, and W.S. Noble; Pacific Symposium on Biocomputing 7:566-575 (2002).
- (294) [EDGAR: Extraction of Drugs, Genes and Relations from the Biomedical Literature](#) T.C. Rindflesch, Lorraine Tanabe, John N. Weinstein, and L. Hunter; Pacific Symposium on Biocomputing 5:514-525 (2000).
- (270) [Mutual Information Relevance Networks: Functional Genomic Clustering Using Pairwise Entropy Measurements](#) A. J. Butte and I. S. Kohane; Pacific Symposium on Biocomputing 5:415-426 (2000).
- (245) [Biobibliometrics: Information Retrieval and Visualization from Co-Occurrences of Gene Names in Medline Abstracts](#) B.J. Stapley and G. Benoit; Pacific Symposium on Biocomputing 5:526-537 (2000).
- (218) [Hybrid Fold Recognition: Combining Sequence Derived Properties with Evolutionary Information](#) D. Fischer; Pacific Symposium on Biocomputing 5:116-127 (2000).
- (215) [Hybrid Petri Net Representation of Gene Regulatory Network](#) H. Matsuno, A. Doi, M. Nagasaki, and S. Miyano; Pacific Symposium on Biocomputing 5:338-349 (2000).
- (198) [Discovery of Regulatory Interactions Through Perturbation: Inference and Experimental Design](#) T.E. Ideker, V. Thorsson, and R.M. Karp; Pacific Symposium on Biocomputing 5:302-313 (2000).
- (152) [Cluster Analysis and Data Visualization of Large-Scale Gene Expression Data](#) G.S. Michaels, D.B. Carr, M. Askenazi, S. Fuhrman, X. Wen and R. Somogyi; Pacific Symposium on Biocomputing 3:42-53 (1998).
- (191) [ANN-Spec: A Method for Discovering Transcription Factor Binding Sites with Improved Specificity](#) C.T. Workman and G.D. Stormo; Pacific Symposium on Biocomputing 5:464-475 (2000).
- (137) [Development of a System for the Inference of Large Scale Genetic Networks](#) Y. Maki, D. Tominaga, M. Okamoto, S. Watanabe, and Y. Eguchi; Pacific Symposium on Biocomputing 6:446-458 (2001).

(130) [A Comparison of Genetic Network Models](#) L.F.A. Wessels, E.P. Van Someren, and M.J.T. Reinders; Pacific Symposium on Biocomputing 6:508-519 (2001).

(126) [Detecting Gene Relations from MEDLINE Abstracts](#) M. Stephens, M. Palakal, S. Mukhopadhyay, R. Raje, and J. Mostafa; Pacific Symposium on Biocomputing 6:483-496 (2001).

We have seen PSB papers cited in the other “high impact journals” and it is satisfying to know that the work presented here is making a real difference. We have spoken with scientists who hesitate to submit their work to PSB for fear that it will not be subsequently consider eligible for a journal. This is an understandable concern, but is mitigated by the excellent penetration of PSB paper as they are listed in Medline and clearly generate some attention, if they are good.

We would like to thank our keynote speakers. Dr. Stanley Nelson, Professor of Human Genetics and Psychiatry at the David Geffen School of Medicine at UCLA will talk about “Whole genome sequencing: issues and progress towards common application.” Our keynote in the area of Ethical, Legal and Social Implications of Technology will be Dr. Isaac Kohane, Professor of Pediatrics and Medicine, Harvard Medical School who will discuss issues related to the ethical application of emerging technical capabilities.

PSB provides sessions focusing on emerging areas in biomedical computation. These sessions are frequently conceived at the meeting itself, during the discussion times. Again, the 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:

Bernard Moret, Webb Miller, Pavel Pevzner, and David Sankoff

[*Computational Challenges in Comparative Genomics*](#)

Rolf Backofen, Hamidreza Chitsaz, Ivo Hofacker, S. Cenk Sahinalp, and Peter F. Stadler

[*Computational studies of non-coding RNAs*](#)

Tanya Berger-Wolf, Teresa Przytycka, Mona Singh, and Donna Slonim

[*Dynamics of Biological Networks*](#)

Samuel Flores, Julie Bernauer, Xuhui Huang, Ruhong Zhou, and Seokmin Shin

[*Multi-resolution Modeling of Biological Macromolecules*](#)

Can Alkan, Michael Brudno, Evan E. Eichler, Maricel G. Kann, S. Cenk Sahinalp

[*Personal Genomics*](#)

Gil Alterovitz, Silvio Cavalcanti, Taro M. Muso, Marco F. Ramoni, and May Wang

[*Reverse Engineering and Synthesis of Biomolecular Systems*](#)

In addition, we welcome three satellite workshops to PSB this year:

Richard Goldstein, Phil Husbands, Chrisantha Fernando, and Dov Stekel

[*In silico biology*](#)

Peter Sterk, Lynette Hirschman, Dawn Field, John Wooley

[*Metagenomics, Metadata and Metaanalysis \(m3\)*](#)

Adrien Coulet, Nigam Shah, Larry Hunter, Chitta Baral and Russ B. Altman

[*GPD-Rxn Workshop: Genotype-Phenotype-Drug Relationship Extraction from Text*](#)

Tiffany Murray continues to run the peer review process and assembly of the proceedings. We thank the National Institutes of Health and the International Society for Computational Biology (ISCB) for travel grant support. We are particularly grateful to BJ McKay-Morrison at ISCB for her assistance. We also 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.

We look forward to a great meeting once again.

Aloha!

Pacific Symposium on Biocomputing Co-Chairs,
September 28, 2009

Russ B. Altman

Departments of Bioengineering, Genetics & Medicine, 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

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.

Gil Alterovitz
Nice-Sophia Antipolis
Zafer Aydin
Rolf Backofen
Tanya Berger-Wolf
Julie Bernauer
Olivier Bodenreider
Philip Bourne
Greg Bowman
Sharon Browning
Catarina D. Campbell
Alessandra Carbone
Bernard Maigret Carloni
Silvio Cavalcanti
Hamid Chitsaz
John Chodera
Yann Christinat
James Collins
Anne Condon
Gregory M. Cooper
Markus Covert
Anneleen Daemen
Frédéric Dayan
Lisette G. dePillis
Ruxandra Dima
Chuong Do
Son Doan
Nilgun Donmez
Ivan Erill
Nicolas Ferey
Christoph Flamm
Samuel Flores
Wolfgang Gatterbauer
Daniel Gautheret
Romain Gauthier
Robert Giegerich
Santhosh Girirajan
Jérôme Golebiowski
Jan Gorodkin
Andreas Gruber
Habiba
Joerg Hackermueller
Iman Hajirasouliha
Sihyun Ham
Ulrich Hansmann
Robert Harris
Steffen Heyne
Paul Higgs

Andreas Hildebrandt
Ivo Hofacker
Fereydoun Hormozdiari
David Hsu
Hsien-Da Huang
Xuhui Huang
Changbong Hyeon
Andy Itsara
Soonmin Jang
Joël Janin
Xiaoye Jiang
Helen Johnson
Mads Kaern
Ulas Karaoz
Rachel Karchin
Janet Kelso
Attila Kertesz-Farkas
Akio Kitao
Diana Kolbe
Robert Lacey
Mayank Lahiri
Doug Lauffenburger
Sonia Leach
Seunghak Lee
Yuhui Li
JC Liao
Pu Liu
Xiong Liu
Sébastien Loriot
Paolo Magni
Arun Maiya
François Major
Nawar Malhis
Leonardo Marino-Ramirez
Kshitij Marwah
David Mathews
Graham McVicker
Webb Miller
David Mobeley
Bernard Moret
Richard Morris
Alan Moses
Vincent Moulton
Yuguang Mu
Taro Muso
Chris J. Myers
Arcadi Navarro
Karen Nelson

Eric K. Neumann
Bill Noble
Ruth Nussinov
Yuko Okamoto
Anna Panchenko
Marc Parisien
Nicholas Pattengale
Dmitri Pervouchine
Pavel Pevzner
Son Pham
David Pincus
Jed W Pitera
Predrag Radivojac
Ben J. Raphael
Jens Reeder
Noah Rosenberg
Cenk Sahinalp
Yasubumi Sakakibara
Raheleh Salari
David Sankoff
Michael Schnieders
Gunnar Schroeder
Chaok Seok
Keith Sevcik
Saad Sheikh
Robert D. Skeel
Christina Smolke
Peter Stadler
Jens Stoye
Peter Sudmant
Ryan Taft
Shoji Takada
Chayant Tantipathananandh
Ron Taylor
May Wang
Todd Wareham
Robert Warren
Bruce Weir
John Wilbur
Sebastian Will
Tom Woolf
Eric Xing
Wei Yang
Yuan Yao
Seungtae Yoon
Michaela Zavolan
Troy Zerr
Ziuwei Zhang

CONTENTS

Preface.....	v
--------------	---

COMPUTATIONAL CHALLENGES IN COMPARATIVE GENOMICS

Session Introduction.....	1
---------------------------	---

Bernard Moret, Webb Miller, Pavel Pevzner, and David Sankoff

Accurate Taxonomic Assignment of Short Pyrosequencing Reads.....	3
--	---

José C. Clemente , Jesper Jansson, Gabriel Valiente

Benchmarking BLAST Accuracy of Genus/Phyla Classification of Metagenomic Reads.....	10
---	----

Steven D. Essinger, Gail L. Rosen

Guided genome halving: provably optimal solutions provide good insights into the preduplication ancestral genome of <i>Saccharomyces cerevisiae</i>	21
---	----

Haris Gavranovic, Eric Tannier

A Practical Algorithm for Estimation of the Maximum Likelihood Ancestral Reconstruction Error.....	31
--	----

Glenn Hickey, Mathieu Blanchette

Optimization methods for selecting founder populations for captive breeding of endangered species.....	43
--	----

Webb Miller, Stephen J. Wright, Yu Zhang, Stephan C. Schuster, Vanessa M. Hayes,

COMPUTATIONAL STUDIES OF NON-CODING RNAS

Session Introduction.....	54
---------------------------	----

Rolf Backofen, Hamidreza Chitsaz, Ivo Hofacker, S. Cenk Sahinalp, Peter F. Stadler

RNA Structural Segmentation.....	57
----------------------------------	----

Ivan Dotu, William A. Lorenz, Pascal Van Hentenryck, Peter Clote

RNAz 2.0: Improved Noncoding RNA Detection.....	69
---	----

Andreas R. Gruber, Sven Findeiß, Stefan Washietl, Ivo L. Hofacker, Peter F. Stadler

Identification and Classification of Small RNAs in Transcriptome Sequence Data.....	80
---	----

D. Langenberger, C.I. Bermudez-Santana, P.F. Stadler, S. Hoffmann

Improvement of Structure Conservation Index with Centroid Estimators.....	88
---	----

Yohei Okada, Kengo Sato, Yasubumi Sakakibara

Dynamic Programming Algorithms for RNA Structure Prediction with Binding Sites.....	98
---	----

Unyane Poolsap, Yuki Kato, Tatsuya Akutsu

An Algorithm for the Energy Barrier Problem Without Pseudoknots and Temporary Arcs.....108

Chris Thachuk, Jan Manuch, Arash Rafiey, Leigh-Anne Mathieson, Ladislav Stacho, Anne Condon

DYNAMICS OF BIOLOGICAL NETWORKS

Session Introduction.....120

Tanya Y. Berger-Wolf, Teresa M. Przytycka, Mona Singh, Donna Slonim

Local Optimization for Global Alignment of Protein Interaction Networks.....123

Leonid Chindelevitch, Chung-Shou Liao, Bonnie Berger

Identification of Coordinately Dysregulated Subnetworks in Complex Phenotypes.....133

Salim A. Chowdhury, Mehmet Koyuturk

Subspace Differential Coexpression Analysis: Problem Definition and a General Approach.....145

Gang Fang, Rui Kuang, Gaurav Pandey, Michael Steinbach, Chad L. Myers, Vipin Kumar

Estimation of Protein and Domain Interactions in the Switching Motility System of *Myxococcus xanthus*.....157

Faruck Morcos, Marcin Sikora, Mark Alber, Dale Kaiser, Jesus A. Izaguirre

Exploring Biological Network Dynamics with Ensembles of Graph Partitions.....166

Saket Navlakha, Carl Kingsford

Geometric Evolutionary Dynamics of Protein Interaction Networks.....178

Natasa Przulj, Oleksii Kuchaiev, Aleksandar Stevanovic, Wayne Hayes

The Steady States and Dynamics of Urokinase-mediated Plasmin Activation.....190

Lakshmi Venkatraman, Hanry Yu, Sourav S. Bhowmick, Forbes Dewey Jr., Lisa Tucker-Kellogg

MULTI-RESOLUTION MODELING OF BIOLOGICAL MACROMOLECULES

Session Introduction.....200

Samuel Flores, Julie Bernauer, Xuhui Huang, Ruhong Zhou, Seokmin Shin

Multi-Resolution Approach for Interactively Locating Functionally Linked Ion Binding Sites by Steering Small Molecules into Electrostatic Potential Maps Using a Haptic Device.....205

Olivier Delalande, Nicolas Férey, Benoist Laurent, Marc Guérout, Brigitte Hartmann, Marc Baaden

Predicting RNA Structure by Multiple Template Homology Modeling.....216

Samuel C. Flores, Yaqi Wan, Rick Russell, Russ B. Altman

Constructing Multi-Resolution Markov State Models (MSMs) to Elucidate RNA Hairpin Folding Mechanisms.....	228
Xuhui Huang, Yuan Yao, Gregory R. Bowman, Jian Sun, Leonidas J. Guibas, Gunnar Carlsson, Vijay S. Pande	
Multiscale Dynamics of Macromolecules Using Normal Mode Langevin.....	240
Jesus A Izaguirre, Christopher R. Sweet, Vijay S. Pande	
Insights into the Intra-Ring Subunit Order of TriC/CCT: Structural and Evolutionary Analysis.....	252
Nir Kalisman, Michael Levitt	
“Cross-Graining:” Efficient Multi-Scale Simulation via Markov State Models.....	260
Peter Kasson, Vijay Pande	
Toward Understanding Allosteric Signaling Mechanisms in the ATPase Domain of Molecular Chaperones.....	269
Ying Liu, Ivet Bahar	
3D-Blast: 3D Protein Structure Alignment, Comparison, and Classification Using Spherical Polar Fourier Correlations.....	281
Lazaros Mavridis. David W. Ritchie	
Structural Prediction of Protein-RNA Interaction by Computational Docking with Propensity-Based Statistical Potentials.....	293
Laura Pérez-Cano, Albert Solernou, Carles Pons, Juan Fernández-Recio	
PERSONAL GENOMICS	
Session Introduction.....	302
Can Alkan, Michael Brudno, Evan E. Eichler, Maricel G. Kann, S. Cenk Sahinalp	
Improving the Prediction of Pharmacogenes Using Text-Derived Gene-Drug Relationships.....	305
Yael Garten, Nicholas P. Tatonetti, Russ B. Altman	
Finding Unique Filter Sets in PLATO: A Precursor to Efficient Interaction Analysis in GWAS Data.....	315
Benjamin J. Grady, Eric Torstenson, Scott M. Dudek, Justin Giles, David Sexton, Marylyn D. Ritchie	
Enabling Personal Genomics with an Explicit Test of Epistasis.....	327
Casey S. Greene, Daniel S. Himmelstein, Heather H. Nelson, Karl T. Kelsey, Scott M. Williams, Angeline S. Andrew, Margaret R. Karagas, Jason H. Moore	
Loss of Post-Translational Modification Sites in Disease.....	337
Shuyan Li, Lilia M. Iakoucheva, Sean D. Mooney, Predrag Radivojac	

Detecting Genome-wide Haplotype Polymorphism by Combined Use of Mendelian Constraints and Local Population Structure.....	348
---	-----

Xin Li, Yixuan Chen, Jing Li

Sequence Feature Variant Type (SFVT) Analysis of the HLA Genetic Association in Juvenile Idiopathic Arthritis.....	359
--	-----

Glenys Thomson, Nishanth Marthandan, Jill A. Hollenbach, Steven J. Mack, Henry A. Erlich, Richard M. Single, Matthew J. Waller, Steven G.E. Marsh, Paula A. Guidry, David R. Karp, Richard H. Scheuermann, Susan D. Thompson, David N. Glass, Wolfgang Helmberg

ÇOKGEN: A Software for the Identification of Rare Copy Number Variation from SNP Microarrays.....	371
---	-----

Gökhan Yavaş, Mehmet Koyutürk, Meral Özsoyoğlu, Meetha P. Gould, Thomas Laframboise

REVERSE ENGINEERING AND SYNTHESIS OF BIOMOLECULAR SYSTEMS

Session Introduction.....	383
---------------------------	-----

Gil Alterovitz, Silvio Cavalcanti, Taro M. Muso, Marco F. Ramoni, May Wang

Co-Design in Synthetic Biology: A System-Level Analysis of the Development of an Environmental Sensing Device.....	385
--	-----

David A. Ball, Matthew W. Lux, Russell R. Graef, Matthew W. Peterson, Jane D. Valenti, John Dileo, Jean Peccoud

Critical Analysis of Transcriptional and Post-Transcriptional Regulatory Networks in Multiple Myeloma.....	397
--	-----

Marta Biasiolo, Mattia Forcato, Lino Possamai, Francesco Ferrari, Luca Agnelli, Marta Lionetti, Katia Todoerti, Antonino Neri, Massimo Marchiori, Stefania Bortoluzzi, Silvio Biciato

A Computational Model of Gene Expression in an Inducible Synthetic Circuit.....	409
---	-----

Francesca Ceroni, Simone Furini, Silvio Cavalcanti

Retrovirus HTLV-1 Gene Circuit: A Potential Oscillator for Eukaryotes.....	421
--	-----

Alberto Corradin, Barbara Di Camillo, Francesca Rende, Vincenzo Ciminale, Gianna Maria Toffolo, Claudio Cobelli

Emulsion Based Selection of T7 Promoters of Varying Activity.....	433
---	-----

Eric A. Davidson, Thomas Van Blarcom, Matthew Levy, Andrew D. Ellington

Clustering Context-Specific Gene Regulatory Networks.....	444
---	-----

Archana Ramesh, Robert Trevino, Daniel D. Von Hoff, Seungchan Kim

Writing and Compiling Code into Biochemistry.....	456
---	-----

Adam Shea, Brian Fett, Marc D. Riedel and Keshab Parhi

Synthesis of Pharmacokinetic Pathways through Knowledge Acquisition and Automated Reasoning.....	465
--	-----

Luis Tari, Saadat Anwar, Shanshan Liang, Jörg Hakenberg, Chitta Baral

WORKSHOPS

In silico Biology.....477

Richard Goldstein, Phil Husbands, Chrisantha Fernando, and Dov Stekel

Genomic Standards Consortium Workshop: Metagenomics, Metadata and Metaanalysis (M3).....481

Peter Sterk, Lynette Hirschman, Dawn Field, John Wooley

Extraction of Genotype-Phenotype-Drug Relationships from Text: From Entity Recognition to Bioinformatics
Application.....485

Adrien Coulet, Nigam Shah, Lawrence Hunter, Chitta Baral and Russ B. Altman