# Lecture Notes in Bioinformatics

4955

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

Martin Vingron Limsoon Wong (Eds.)

# Research in Computational Molecular Biology

12th Annual International Conference, RECOMB 2008 Singapore, March 30 – April 2, 2008 Proceedings



#### 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

Martin Vingron
Max Planck Institute for Molecular Genetics
Ihnestr. 63-73, 14195 Berlin, Germany
E-mail: vingron@molgen.mpg.de

Limsoon Wong National University of Singapore, School of Computing Law Link, Singapore 117590 E-mail: wongls@comp.nus.edu.sg

Library of Congress Control Number: 2008922908

CR Subject Classification (1998): F.2.2, F.2, E.1, G.2, H.2.8, G.3, I.2, J.3

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-540-78838-7 Springer Berlin Heidelberg New York ISBN-13 978-3-540-78838-6 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 2008 Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India Printed on acid-free paper SPIN: 12247264 06/3180 5 4 3 2 1 0

#### **Preface**

This volume contains the papers presented at the 12th Annual International Conference on REsearch in COmputational Molecular Biology (RECOMB 2008) held in Singapore, March 30-April 2, 2008. The RECOMB conference series was founded in 1997 by Sorin Istrail, Pavel Pevzner and Michael Waterman. Its history is summarized in the table herein. The general development of the RECOMB conference series is guided by a Steering Committee. RECOMB 2008 was hosted by the National University of Singapore and held at the University Cultural Centre. The Conference Chair was Limsoon Wong, who was supported by a 16-member-strong Organizing Committee.

Out of 193 papers submitted to RECOMB 2008, the Program Committee (PC) selected 34 papers for presentation at the conference. The PC consisted of 46 members, who could further draw on the help of subreviewers. Three PC members with possible help of the external reviewers first reviewed each paper, subsequently a Web-based discussion led to the final selection of papers.

This year, RECOMB and the journal Genome Research teamed up to offer authors Genome Research as an alternative forum for publication. Approximately 55 of the RECOMB submissions, selected by a subset of the PC headed by Serafim Batzoglou, were forwarded to Genome Research for consideration. The reviewing procedures by RECOMB and Genome Research were conducted independently, leading to the acceptance of a small number of papers to both RECOMB and Genome Research. Other papers were accepted to Genome Research but not to RECOMB. The papers accepted by both were presented at the conference and are also summarized in this volume. Their final versions appear in Genome Research.

In addition to the contributed presentations, RECOMB 2008 featured seven keynote lectures, held by Howard Cedar (Hebrew University, Jerusalem), Vivian G. Cheung (University of Pennsylvania, Philadelphia), Suzanne Cory (Walter and Eliza Hall Institute of Medical Research, Melbourne), Sang Yup Lee (Korea Advanced Institute of Science and Technology, Daejeon), Edison Liu (Genome Institute of Singapore, Singapore), Andrei Lupas (Max Planck Institute for Developmental Biology, Tübingen), and Temple F. Smith (Boston University, Boston). The Ulam Lecture was delivered by Temple Smith, the Distinguished Technology Lecture by Edison Liu and the Distinguished Biology Lecture by Suzanne Cory.

In addition to the oral presentations, RECOMB 2008 also provided a forum for presenting approximately 150 posters. The poster sessions are an integral part of the conference and foster communication among participants.

A conference like RECOMB lives from the dedication of the people helping and supporting it. Thanks go to the RECOMB Steering Committee for their advice and guidance, to the PC and the external reviewers for their efforts, to

#### VI Preface

the Poster Committee for going through a large number of posters in short time, and to the local organizers for their hard work. The reviewing process profited greatly from the use of the EasyChair conference management system. Patricia Marquardt's help in putting together this proceedings volume is greatly appreciated. We are further indebted to the National University of Singapore, the Institute for Infocomm Research (A\*STAR), and the Bioinformatics Institute (A\*STAR) for their support of RECOMB 2008. We also appreciate the generous sponsorships from the Lilly Singapore Center for Drug Discovery, IBM, the International Society for Computational Biology, World Scientific Publishing Company, and Taylor & Francis Asia Pacific. Special thanks go to the participants, to those who contributed posters or oral presentations, and to the keynote speakers for their input. It is this community that makes RECOMB such an inspiring, lively, and enjoyable conference.

February 2008

Martin Vingron

# Organization

#### Program Committee

Tatsuya Akutsu Kyoto University, Japan Serafim Batzoglou Stanford University, USA Bonnie Berger MIT, Cambridge, USA

Mathieu Blanchette McGill University, Montreal, Quebec, Canada

Joachim Buhmann ETH Zürich, Switzerland

Kun-Mao Chao National Taiwan University, Taipei, Taiwan Dannie Durand Carnegie Mellon University, Pittsburgh, USA Eleazar Eskin University of California, Los Angeles, USA Nir Friedman Hebrew University, Jerusalem, Israel

Dan Geiger Technion, Haifa, Israel Gaston Gonnet ETH Zürich, Switzerland

Dan Gusfield University of California, Davis, USA

Arndt von Haeseler Max F. Perutz Laboratories, Vienna, Austria

Alexander Hartemink Duke University, Durham, USA

Trey Ideker University of California, San Diego, USA Sorin Istrail Brown University, Providence, USA Tao Jiang University of California, Riverside, USA John Kececioglu University of Arizona, Tucson, USA

Manolis Kellis Massachusetts Institute of Technology, Cambridge, USA

Anders Krogh University of Copenhagen, Denmark

Jens Lagergren Royal Institute of Technology, Stockholm, Sweden Thomas Lengauer Max Planck Institute for Computer Science,

Saarbrücken, Germany

Arthur Lesk Pennsylvania State University, University Park, USA

Ming Li University of Waterloo, Canada Michal Linial Hebrew University, Jerusalem, Israel

Satoru Miyano Tokyo University, Japan

Adam Siepel

William Noble University of Washington, USA

Pavel Pevzner University of California, San Diego, USA
Knut Reinert Free University of Berlin, Germany
Stephan Robin AgroParisTech, Paris, France
Ron Shamir Tel Aviv University, Israel
Roded Sharan Tel Aviv University, Israel

Terence Speed University of California, Berkeley, USA

Peter Stadler University of Leipzig, Germany Jens Stoye University of Bielefeld, Germany

Fengzhu Sun University of Southern California, Los Angeles, USA

Cornell University, Ithaca, USA

#### VIII Organization

Wing-Kin Sung National University of Singapore, Singapore

Simon Tavaré Cornell University, Ithaca, USA Jerzy Tiuryn University of Warsaw, Poland

Martin Vingron (Chair) Max Planck Institute for Molecular Genetics,

Berlin, Germany

Lusheng Wang City University, Hong Kong

Michael Waterman University of Southern California, Los Angeles, USA

Limsoon Wong National University of Singapore, Singapore

Wing Hung Wong Stanford University, USA

Eric Xing Carnegie Mellon University, Pittsburgh, USA Zohar Yakhini Agilent Laboratories, Santa Clara, USA

#### Poster Committee

Hon-Nian Chua Institute for Infocomm Research, Singapore Neil Clarke Genome Institute of Singapore, Singapore

Frank Eisenhaber Bioinformatics Institute, Singapore

Wing-Kin Sung National University of Singapore, Singapore

Martin Vingron Max Planck Institute for Molecular Genetics,

(Co-chair) Berlin, Germany

Limsoon Wong National University of Singapore, Singapore

(Co-chair)

### Steering Committee

Serafim Batzoglou Stanford University, USA

Sorin Istrail Brown University, Providence, USA

Thomas Lengauer Max Planck Institute for Computer Science,

Saarbrücken, Germany

Michal Linial Hebrew University, Jerusalem, Israel Pavel Pevzner University of California, San Diego, USA

Ron Shamir Tel Aviv University, Israel

Terrence Speed University of California, Berkeley, USA Martin Vingron Max Planck Institute for Molecular Genetics,

Berlin, Germany

### Organizing Committee

Agnes Ang
Lay Khim Chng
Kwok Pui Choi
Alexia Leong
Hon Wai Leong
National University of Singapore, Singapore

Wai Kin Leong National University of Singapore, Singapore Stefanie Ng National University of Singapore, Singapore Wing-Kin Sung National University of Singapore, Singapore Martti Tammi National University of Singapore, Singapore Limsoon Wong (chair) National University of Singapore, Singapore National University of Singapore, Singapore Siang Yong Yap Xin Chen Nanyang Technological University, Singapore Chee Keong Kwoh Nanyang Technological University, Singapore Guillaume Bourque Genome Institute of Singapore, Singapore Alan Christoffels Temasek Lifesciences Laboratory, Singapore Dong-Yup Lee Bioprocessing Technology Institute, Singapore

Gunaretnam

Rajagopal BioInformatics Institute, Singapore

## Previous RECOMB Meetings

Date / Location	Hosting Institution	Program Chair	Conference Chair
January 20-23, 1997	Sandia National Lab	Michael Waterman	Sorin Istrail
Santa Fé, NM, USA			
March 22-25, 1998	Mt. Sinai School	Pavel Pevzner	Gary Benson
New York, NY, USA	of Medicine		
April 22-25, 1999	INRIA	Sorin Istrail	Mireille Regnier
Lyon, France			
April 8-11, 2000	University of Tokyo	Ron Shamir	Satoru Miyano
Tokyo, Japan			
April 22-25, 2001	Université de Montréal	Thomas Lengauer	David Sankoff
Montréal, Canada			
April 18-21. 2002	Celera	Gene Myers	Sridhar Hannenhalli
Washington, DC, USA			
April 10-13, 2003	Max Planck Institute	Webb Miller	Martin Vingron
Berlin, Germany	for Molecular Genetics		
March 27-31, 2004	UC San Diego	Dan Gusfield	Philip E. Bourne
San Diego, CA, USA			
May 14-18, 2005	Broad Institute	Satoru Miyano	Jill P. Mesirov
Boston, MA, USA	of MIT and Harvard		and Simon Kasif
April 2-5, 2006	University of Padova	Alberto Apostolico	Concettina Guerra
Venice, Italy			
April 21-25, 2007	California Institute	Terry Speed	Sandrine Dudoit
Oakland, CA, USA	for Quantitative		
	Biomedical Research		

The RECOMB 2008 Program Committee gratefully acknowledges the valuable input received from the following external reviewers

Sarah Aerni Edoardo Airoldi Frank Alber Andrey Alexevenko Adrian Altenhoff Peter Arndt Lars Arvestad George Asimenos Manor Askenazi Chris Bailey-Kellogg Sourav Bandyopadhyay Hideo Bannai Vikas Bansal Avner Bar-Hen Nuno Barbosa-Morais Markus Bauer Jan Baumbach Amir Ben-Dor Sivan Bercovici Doron Betel Chris Bielow Alexander Bockmavr Wouter Boomsma Laurent Brehelin Brona Brejova Dan Brown Brian Browning Jocelvne Bruand Sebastian Böcker Gal Chechik Ming-Chiang Chen Kuan-Yu Chen Xiaovu Chen Benny Chor Han-Yu Chuang Trees-Juen Chuang Inbar Cohen-Gihon Miklos Csuros Tim Danford Morihiro Hayashida Ofir Davidovich Jacques van Helden Torsten Herrmann Minghua Deng

Zhihong Ding

Chuong Do Norbert Dojer Francisco Domingues Banu Dost Robin Dowell Janusz Dutkowski Eran Eden Nathan Edwards Tal El-Hay Ran Elkon Arne Elofsson Olof Emanuelsson Nikolas Fechner Bernd Fischer Jason Flannick Geoffrey Fox Ari Frank Eugene Fratkin Shiri Freilich Weijie Fu Mickael Guedi Anna Gambin Irit Gat-Viks Guv Geva Manuel Gil Uri Gophna Pawel Gorecki Assaf Gottlieb Ewing Greg Frauke Gräter Clemens Gröpl Karlebach Guv Naomi Habib Eran Halperin Yonit Halperin Thomas Hamelryck Buhm Han Greg Hannum

Andreas Hildebrandt

Ivo Hofacker Barbara Holland Liisa Holm Judie Howrylak Yao-Ting Huang Xiaoqiu Huang Katharina Huber Man Hung Daniel Huson Seiva Imoto Anders Jacobsen Katharina Jahn Ariel Jaimovich Rui Jiang Inge Jonassen Maxim Kalaev Lukas Kall Hans-Michael Kaltenbach Hetunandan Kamisetty

Hyun Min Kang Tommy Kaplan Yuki Kato Uri Keich Orgad Keller Pouya Kheradpour Seyoung Kim Gad Kimmel Bonnie Kirkpatrick Aaron Klammer Gunnar W. Klau Boguslaw Kluge Oliver Kohlbacher Rachel Kolodny Carolin Kosiol Dennis Kostka Roland Krause David Kreil Michael Lappe Berthold Lausen Chris Lee Hans-Peter Lenhof

Ulf Leser Mike Lin Stinus Lindgreen Chaim Linhart Simone Linz Hsiao-Fei Liu Yaniv Loewenstein Chin Lung Lu Xiaotu Ma Craig Mak John Marioni Paul Marjoram Tobias Marschall Troels Marstrand Marie-Laure Martin-Magniette Osamu Maruyama Tristan Mary-Huard Jochen Maydt Aurelien Mazurie Paul Medvedev Dirk Metzler Martin Middendorf

Istvan Miklos
Axel Mosig
Simon Myers
Cedric Notredame
Noa Novershtern
Sanne Nygaard
Charles O'Donnell
Sean O'Rourke
Peter Orbanz
Ivan Ovcharenko
Rod Page

Dana Pe'er Jakob Skou Pedersen Peter Pfaffelhuber Franck Picard Ron Pinter Robert Preissner

Teresa Przytycka

Utz Pape

Jian Qiu

Predrag Radivojac
Sven Rahmann
Jörg Rahnenführer
Ben Raphael
Matt Rasmusson

Ben Raphael
Matt Rasmussen
Tobias Rausch
Pradipta Ray
Sheila Reynolds
Markus Ringnér
Peter Robinson
Ingo Roeder
Volker Roth
Bashir Sadjad
Marie-France Sagot
Nikolaos Sahinidis
Albin Sandelin

Albin Sandelin Oliver Sander Marc Schaub Maya Schuldiner

Russell Schwartz Ariel Schwartz Bengt Sennblad

Oliver Serang Itai Sharon Maxim Shatsky Tetsuo Shibuya Teppei Shimamura Tomer Shlomi

Suyash Shringarpure Alexandra

Shulman-Peleg Sagi Snir

Kyung-Ah Sohn
Nan Song
Yun Song
Rainer Spang
Rohith Srivas
Alex Stark

Israel Steinfeld Christine Steinhoff

Jörg Stelling

Andreas Sundquist Silpa Suthram Alain Trubuil Takeyuki Tamura

Kai Tan
Eric Tang
Haixu Tang
Willie Taylor
Robert Thurman
Ali Tofigh
Martin Tompa
Zhidong Tu

Zhidong Tu
Eivind Valen
Roy Varshavsky
Balaji Venkatachalam
Stéphane Vialette
Antoine Vigneron
Jens Vilstrup Johansen

Tomas Vinar Le Sy Vinh

Jerome Waldispuhl

Lin Wan Li Wang

Hung-Lung Wang Ilan Wapinski David Weese Hugo Willy Ole Winther Roland Wittler Yufeng Wu Xiting Yan Zizhen Yao Yuzhen Ye Chun Ye

Daniel Yekutieli Nir Yosef Noah Zaitlen Alex Zelikovsky Tomasz Zemojtel Shaojie Zhang

Xianghong Jasmine Zhou

## **Organizers**





## **Supporting Organizers**





## **Platinum Sponsors**







### **Gold Sponsors**





# Table of Contents

(Keynote)	1
Bootstrapping the Interactome: Unsupervised Identification of Protein Complexes in Yeast	3
CompostBin: A DNA Composition-Based Algorithm for Binning Environmental Shotgun Reads	17
Reconstructing the Evolutionary History of Complex Human Gene Clusters	29
Ab Initio Whole Genome Shotgun Assembly with Mated Short Reads	50
Orchestration of DNA Methylation (Keynote)	65
BayCis: A Bayesian Hierarchical HMM for Cis-Regulatory Module Decoding in Metazoan Genomes	66
A Combined Expression-Interaction Model for Inferring the Temporal Activity of Transcription Factors	82
A Fast, Alignment-Free, Conservation-Based Method for Transcription Factor Binding Site Discovery	98
The Statistical Power of Phylogenetic Motif Models	112
Transcriptional Regulation and Cancer Genomics (Keynote) $Edison\ Liu$	127

Automatic Recognition of Cells (ARC) for 3D Images of C. elegans Fuhui Long, Hanchuan Peng, Xiao Liu, Stuart Kim, and Gene Myers	128
Spectrum Fusion: Using Multiple Mass Spectra for De Novo Peptide Sequencing	140
A Fragmentation Event Model for Peptide Identification by Mass Spectrometry	154
A Bayesian Approach to Protein Inference Problem in Shotgun Proteomics	167
De Novo Sequencing of Nonribosomal Peptides	181
Systems Metabolic Engineering (Keynote)	196
Protein Function Prediction Based on Patterns in Biological Networks	197
Automatic Parameter Learning for Multiple Network Alignment Jason Flannick, Antal Novak, Chuong B. Do, Balaji S. Srinivasan, and Serafim Batzoglou	214
An Integrative Network Approach to Map the Transcriptome to the Phenome	232
Fast and Accurate Alignment of Multiple Protein Networks	246
High-Resolution Modeling of Cellular Signaling Networks	257
At the Origin of Life: How Did Folded Proteins Evolve? (Keynote)	272
Locating Multiple Gene Duplications through Reconciled Trees J. Gordon Burleigh, Mukul S. Bansal, Andre Wehe, and Oliver Eulenstein	273

#### XVI Table of Contents

Increasing Power in Association Studies by Using Linkage Disequilibrium Structure and Molecular Function as Prior Information	434
Eleazar Eskin	101
Panel Construction for Mapping in Admixed Populations Via Expected Mutual Information	435
Constructing Level-2 Phylogenetic Networks from Triplets Leo van Iersel, Judith Keijsper, Steven Kelk, Leen Stougie, Ferry Hagen, and Teun Boekhout	450
Accurate Computation of Likelihoods in the Coalescent with Recombination Via Parsimony	463
Author Index	479