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
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
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
Adam Siepel Cornell University, Ithaca, USA
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
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
Greg Tucker-Kellogg Lilly Singapore Centre for Drug Discovery, Singapore
Martin Vingron Max Planck Institute for Molecular Genetics, Berlin, Germany
(L-co-chair)
Limsoon Wong National University of Singapore, Singapore
(L-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 National University of Singapore, Singapore
Lay Khim Chng National University of Singapore, Singapore
Kwok Pui Choi National University of Singapore, Singapore
Alexia Leong National University of Singapore, Singapore
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
Siang Yong Yap National University of Singapore, Singapore
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 BioInformatics Institute, Singapore

<table>
<thead>
<tr>
<th>Date / Location</th>
<th>Hosting Institution</th>
<th>Program Chair</th>
<th>Conference Chair</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 20-23, 1997</td>
<td>Sandia National Lab</td>
<td>Michael Waterman</td>
<td>Sorin Istrail</td>
</tr>
<tr>
<td>Santa Fé, NM, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 22-25, 1998</td>
<td>Mt. Sinai School of Medicine</td>
<td>Pavel Pevzner</td>
<td>Gary Benson</td>
</tr>
<tr>
<td>New York, NY, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 22-25, 1999</td>
<td>INRIA</td>
<td>Sorin Istrail</td>
<td>Mireille Regnier</td>
</tr>
<tr>
<td>Lyon, France</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 8-11, 2000</td>
<td>University of Tokyo</td>
<td>Ron Shamir</td>
<td>Satoru Miyano</td>
</tr>
<tr>
<td>Tokyo, Japan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 22-25, 2001</td>
<td>Université de Montréal</td>
<td>Thomas Lengauer</td>
<td>David Sankoff</td>
</tr>
<tr>
<td>Montréal, Canada</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 18-21, 2002</td>
<td>Celera</td>
<td>Gene Myers</td>
<td>Sridhar Hannenhalli</td>
</tr>
<tr>
<td>Washington, DC, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 10-13, 2003</td>
<td>Max Planck Institute for Molecular Genetics</td>
<td>Webb Miller</td>
<td>Martin Vingron</td>
</tr>
<tr>
<td>Berlin, Germany</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 27-31, 2004</td>
<td>UC San Diego</td>
<td>Dan Gusfield</td>
<td>Philip E. Bourne</td>
</tr>
<tr>
<td>San Diego, CA, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 14-18, 2005</td>
<td>Broad Institute of MIT and Harvard</td>
<td>Satoru Miyano</td>
<td>Jill P. Mesirov and Simon Kasif</td>
</tr>
<tr>
<td>Boston, MA, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 2-5, 2006</td>
<td>University of Padova</td>
<td>Alberto Apostolico</td>
<td>Concettina Guerra</td>
</tr>
<tr>
<td>Venice, Italy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 21-25, 2007</td>
<td>California Institute for Quantitative Biomedical Research</td>
<td>Terry Speed</td>
<td>Sandrine Dudoit</td>
</tr>
<tr>
<td>Oakland, CA, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The RECOMB 2008 Program Committee gratefully acknowledges the valuable input received from the following external reviewers:

Sarah Aerni
Edoardo Airoldi
Frank Alber
Andrey Alexeyenko
Adrian Altenhoff
Peter Arndt
Lars Arvestad
George Asiminos
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 Bockmayr
Wouter Boomsma
Laurent Brehelin
Brona Brejova
Dan Brown
Brian Browning
Jocelyne Bruand
Sebastian Böcker
Gal Chechik
Ming-Chiang Chen
Kuan-Yu Chen
Xiaoyu Chen
Benny Chor
Han-Yu Chuang
Trees-Juen Chuang
Inbar Cohen-Gihon
Miklos Csuro
Tim Danford
Ofir Davidovich
Minghua Deng
Zhihong Ding
Chuong Do
Norbert Dojer
Francisco Domingues
Benn 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 Guedj
Anna Gambin
Irit Gat-Viks
Guy Geva
Manuel Gil
Uri Gophna
Pawel Gorecki
Assaf Gottlieb
Ewing Greg
Frauke Gräter
Clemens Gröpl
Karlebach Guy
Naomi Habib
Eran Halperin
Yonit Halperin
Thomas Hamelryck
Buhm Han
Greg Hannum
Morihiro Hayashida
Jacques van Helden
Torsten Herrmann
Andreas Hildebrandt
Ivo Hofacker
Barbara Holland
Liisa Holm
Judie Howrylak
Yao-Ting Huang
Xiaoqiu Huang
Katharina Huber
Man Hung
Daniel Huson
Seiya Imoto
Anders Jacobsen
Katharina Jahn
Ariel Jaimovich
Rui Jiang
Inge Jonassen
Maxim Kalaev
Lukas Kall
Hans-Michael
Kaltenbach
Hetunandan Kamisetty
Hyum 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
Organizers

NUS
National University of Singapore
School of Computing

NUS
Bioinformatics Programme

Supporting Organizers

Institute for Infocomm Research
Bioinformatics Institute

Platinum Sponsors

Lilly
Answers That Matter.

IBM

iscb
International Society for Computational Biology

Gold Sponsors

World Scientific
Connecting Great Minds

Taylor & Francis
Taylor & Francis Group
# Table of Contents

Computational Biology: Its Challenges Past, Present, and Future (Keynote) ........................................... 1
   Temple F. Smith

Bootstrapping the Interactome: Unsupervised Identification of Protein Complexes in Yeast .............................................. 3
   Caroline C. Friedel, Jan Krumsiek, and Ralf Zimmer

CompostBin: A DNA Composition-Based Algorithm for Binning Environmental Shotgun Reads .................................... 17
   Sourav Chatterji, Ichitaro Yamazaki, Zhaojun Bai, and Jonathan A. Eisen

Reconstructing the Evolutionary History of Complex Human Gene Clusters ........................................................ 29
   Yu Zhang, Giltae Song, Tomáš Vinař, Eric D. Green, Adam Siepel, and Webb Miller

Ab Initio Whole Genome Shotgun Assembly with Mated Short Reads .......................................................... 50
   Paul Medvedev and Michael Brudno

Orchestration of DNA Methylation (Keynote) ................................. 65
   Howard Cedar

BayCis: A Bayesian Hierarchical HMM for Cis-Regulatory Module Decoding in Metazoan Genomes ................................. 66
   Tien-ho Lin, Pradipta Ray, Geir K. Sandve, Selen Uguroglu, and Eric P. Xing

A Combined Expression-Interaction Model for Inferring the Temporal Activity of Transcription Factors ................ 82
   Yanxin Shi, Itamar Simon, Tom Mitchell, and Ziv Bar-Joseph

A Fast, Alignment-Free, Conservation-Based Method for Transcription Factor Binding Site Discovery .......................... 98
   Raluca Gordân, Leelavati Narlikar, and Alexander J. Hartemink

The Statistical Power of Phylogenetic Motif Models ..................... 112
   John Hawkins and Timothy L. Bailey

Transcriptional Regulation and Cancer Genomics (Keynote) ......... 127
   Edison Liu
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Recognition of Cells (ARC) for 3D Images of <em>C. elegans</em></td>
<td>128</td>
</tr>
<tr>
<td><em>Fuhui Long, Hanchuan Peng, Xiao Liu, Stuart Kim, and Gene Myers</em></td>
<td></td>
</tr>
<tr>
<td>Spectrum Fusion: Using Multiple Mass Spectra for De Novo Peptide</td>
<td>140</td>
</tr>
<tr>
<td>Sequencing</td>
<td></td>
</tr>
<tr>
<td><em>Ritendra Datta and Marshall Bern</em></td>
<td></td>
</tr>
<tr>
<td>A Fragmentation Event Model for Peptide Identification by Mass</td>
<td>154</td>
</tr>
<tr>
<td>Spectrometry</td>
<td></td>
</tr>
<tr>
<td><em>Yu Lin, Yantao Qiao, Shiwei Sun, Chungong Yu, Gongjin Dong, and Dongbo Bu</em></td>
<td></td>
</tr>
<tr>
<td>A Bayesian Approach to Protein Inference Problem in Shotgun Proteomics</td>
<td>167</td>
</tr>
<tr>
<td><em>Yong Fuga Li, Randy J. Arnold, Yixue Li, Predrag Radivojac, Quanhu Sheng, and Haixu Tang</em></td>
<td></td>
</tr>
<tr>
<td>De Novo Sequencing of Nonribosomal Peptides</td>
<td>181</td>
</tr>
<tr>
<td><em>Nuno Bandeira, Julio Ng, Dario Meluzzi, Roger G. Linington, Pieter Dorrestein, and Pavel A. Pevzner</em></td>
<td></td>
</tr>
<tr>
<td>Systems Metabolic Engineering (Keynote)</td>
<td>196</td>
</tr>
<tr>
<td><em>Sang Yup Lee</em></td>
<td></td>
</tr>
<tr>
<td>Protein Function Prediction Based on Patterns in Biological Networks</td>
<td>197</td>
</tr>
<tr>
<td><em>Mustafa Kirac and Gultekin Ozsoyoglu</em></td>
<td></td>
</tr>
<tr>
<td>Automatic Parameter Learning for Multiple Network Alignment</td>
<td>214</td>
</tr>
<tr>
<td><em>Jason Flannick, Antal Novak, Chuong B. Do, Balaji S. Srinivasan, and Serafim Batzoglou</em></td>
<td></td>
</tr>
<tr>
<td>An Integrative Network Approach to Map the Transcriptome to the Phenome</td>
<td>232</td>
</tr>
<tr>
<td><em>Michael R. Mehan, Juan Nuñez-Iglesias, Mrinal Kalakrishnan, Michael S. Waterman, and Xianghong Jasmine Zhou</em></td>
<td></td>
</tr>
<tr>
<td>Fast and Accurate Alignment of Multiple Protein Networks</td>
<td>246</td>
</tr>
<tr>
<td><em>Maxim Kalaev, Vineet Bafna, and Roded Sharan</em></td>
<td></td>
</tr>
<tr>
<td>High-Resolution Modeling of Cellular Signaling Networks</td>
<td>257</td>
</tr>
<tr>
<td><em>Michael Baym, Chris Bakal, Norbert Perrimon, and Bonnie Berger</em></td>
<td></td>
</tr>
<tr>
<td>At the Origin of Life: How Did Folded Proteins Evolve? (Keynote)</td>
<td>272</td>
</tr>
<tr>
<td><em>Andrei Lupas</em></td>
<td></td>
</tr>
<tr>
<td>Locating Multiple Gene Duplications through Reconciled Trees</td>
<td>273</td>
</tr>
<tr>
<td><em>J. Gordon Burleigh, Mukul S. Bansal, Andre Wehe, and Oliver Eulenstein</em></td>
<td></td>
</tr>
</tbody>
</table>
Rapid and Accurate Protein Side Chain Prediction with Local Backbone Information ........................................ 285
Jing Zhang, Xin Gao, Jinbo Xu, and Ming Li

Algorithms for Joint Optimization of Stability and Diversity in Planning Combinatorial Libraries of Chimeric Proteins .................. 300
Wei Zheng, Alan M. Friedman, and Chris Bailey-Kellogg

Christophe Dessimoz, Daniel Margadant, and Gaston H. Gonnet

Computation of Median Gene Clusters .......................................................... 331
Sebastian Böcker, Katharina Jahn, Julia Mixtacki, and Jens Stoye

BCL-2: From Translocation to Therapy (Keynote) ................... 346
Suzanne Cory

Detecting Disease-Specific Dysregulated Pathways Via Analysis of Clinical Expression Profiles ........................................ 347
Igor Ulitsky, Richard M. Karp, and Ron Shamir

Constructing Treatment Portfolios Using Affinity Propagation ...... 360
Delbert Dueck, Brendan J. Frey, Nebojsa Jojic, Vladimir Jojic, Guri Giaever, Andrew Emili, Gabe Musso, and Robert Hegele

Bubbles: Alternative Splicing Events of Arbitrary Dimension in Splicing Graphs .......................................................... 372
Michael Sammeth, Gabriel Valiente, and Roderic Guigó

More Efficient Algorithms for Closest String and Substring Problems ... 396
Bin Ma and Xiaoming Sun

Disruption of a Transcriptional Regulatory Pathway Contributes to Phenotypes in Carriers of Ataxia Telangiectasia (Keynote) ............ 410
Denis Smirnov and Vivian G. Cheung

Accounting for Non-genetic Factors Improves the Power of eQTL Studies .......................................................... 411
Oliver Stegle, Anitha Kannan, Richard Durbin, and John Winn

Effects of Genetic Divergence in Identifying Ancestral Origin Using HAPAA .......................... 423
Andreas Sundquist, Eugene Fratkin, Chuong B. Do, and Serafim Batzoglou

On the Inference of Ancestries in Admixed Populations ............... 424
Sriram Sankararaman, Gad Kimmel, Eran Halperin, and Michael I. Jordan
Increasing Power in Association Studies by Using Linkage Disequilibrium Structure and Molecular Function as Prior Information .................................................. 434
   Eleazar Eskin

Panel Construction for Mapping in Admixed Populations Via Expected Mutual Information .................................................. 435
   Sivan Bercovici, Dan Geiger, Liran Shlush, Karl Skorecki, and Alan Templeton

Constructing Level-2 Phylogenetic Networks from Triplets ............ 450
   Leo van Iersel, Judith Keijsper, Steven Kelk, Leen Stougie, Ferry Hagen, and Teun Boekhout

Accurate Computation of Likelihoods in the Coalescent with Recombination Via Parsimony ........................................... 463
   Rune B. Lyngsø, Yun S. Song, and Jotun Hein

Author Index .................................................. 479