

# Lecture Notes in Bioinformatics

6044

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

Bonnie Berger (Ed.)

# Research in Computational Molecular Biology

14th Annual International Conference, RECOMB 2010  
Lisbon, Portugal, April 25-28, 2010  
Proceedings



Springer

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 Editor

Bonnie Berger

Massachusetts Institute of Technology

Mathematics Department

Computer Science and Artificial Intelligence Laboratory

77 Massachusetts Avenue, Cambridge, MA 02139, USA

E-mail: bab@mit.edu

Library of Congress Control Number: 2010924885

CR Subject Classification (1998): J.3, H.2.8, I.2, F.2.2, I.6, F.2

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-642-12682-0 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-12682-6 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.com

© Springer-Verlag Berlin Heidelberg 2010

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper 06/3180

# Preface

This volume contains the papers presented at RECOMB 2010: the 14th Annual International Conference on Research in Computational Molecular Biology held in Lisbon, Portugal, during April 25–28, 2010. The RECOMB conference series was started in 1997 by Sorin Istrail, Pavel Pevzner, and Michael Waterman. RECOMB 2010 was hosted by INESC-ID and Instituto Superior Tecnico, organized by a committee chaired by Arlindo Oliveira and took place at the International Fair of Lisbon Meeting Centre.

This year, 36 papers were accepted for presentation out of 176 submissions. The papers presented were selected by the Program Committee (PC) assisted by a number of external reviewers. Each paper was reviewed by three members of the PC, or by external reviewers, and there was an extensive Web-based discussion over a period of two weeks, leading to the final decisions. RECOMB 2010 also introduced a Highlights Track, in which six additional presentations by senior authors were chosen from papers published in 2009. The RECOMB conference series is closely associated with the *Journal of Computational Biology*, which traditionally publishes special issues devoted to presenting full versions of selected conference papers.

RECOMB 2010 invited several distinguished speakers as keynotes and for special sessions on Genomics in Medicine and Regulatory RNAs. Invited speakers included Cecilia Arraiano (Instituto de Tecnologia Química e Biológica Universidade Nova de Lisboa), Chris Bakal (Institute of Cancer Research), David Bartel (MIT, Whitehead Institute and Howard Hughes Medical Institute), Jan Gorodkin (University of Copenhagen), Simon Kasif (Boston University), Isaac Kohane (Harvard Medical School), Doron Lancet (Weizmann Institute), Klaus Lindpainter (Roche Genetics), Patrice Milos (Helicos BioSciences Corporation), Norbert Perrimon (Harvard Medical School and Howard Hughes Medical Institute), Eitan Rubin (Ben Gurion University), and Mona Singh (Princeton University).

RECOMB 2010 was only possible through the dedication and hard work of many individuals and organizations. Special thanks go to the PC and external reviewers for helping to form a high-quality conference program, and the Organizing Committee, coordinated by Ana Teresa Freitas, for hosting the conference and providing the administrative, logistic, and financial support. We also thank our sponsors, including ISCB, FCT and FLAD. Without them the conference would not have been financially viable. We thank the RECOMB Steering Committee, chaired by Martin Vingron, for accepting the challenge of organizing this meeting in Lisbon. Finally, we thank all the authors who contributed papers and posters, as well as the attendees of the conference for their enthusiastic participation.

# Conference Organization

## Program Committee

Tatsuya Akutsu	Kyoto University, Japan
Nancy Amato	Texas A&M University, USA
Joel Bader	Johns Hopkins University, USA
Serafim Batzoglou	Stanford University, USA
Bonnie Berger (Chair)	Massachusetts Institute of Technology, USA
Phil Bradley	Fred Hutchinson Cancer Research Center, USA
Jadwiga Bienkowska	Biogenidec, USA
Michael Brent	Washington University, USA
Michael Brudno	University of Toronto, Canada
Lenore Cowen	Tufts University, USA
Colin Dewey	University of Wisconsin, Madison, USA
Dannie Durand	Carnegie Melon University, USA
Eleazar Eskin	University of California, Los Angeles, USA
Ana Teresa Freitas	INESC-ID / IST, Portugal
James Galagan	Boston University, USA
David Gifford	Massachusetts Institute of Technology, USA
Eran Halperin	University of California, Berkeley, USA
Des Higgins	University College Dublin, Ireland
Trey Ideker	University of California, San Diego, USA
Sorin Istrail	Brown University, USA
Tao Jiang	University of California, Riverside, USA
Simon Kasif	Boston University, USA
Isaac Kohane	Harvard Medical School, USA
Jens Lagergren	Royal Institute of Technology, Sweden
Thomas Lengauer	Max Planck Institute for Informatics, Germany
Michal Linial	Hebrew University, Israel
Satoru Miyano	Tokyo University, Japan
Bernard Moret	Swiss Federal Institutes of Technology, Switzerland
William Noble	University of Washington, USA
Arlindo Oliveira	INESC-ID / IST, Portugal
Dana Pe'er	Columbia University, USA
Pavel Pevzner	University of California, San Diego
Ron Pinter	Technion, Israel

## VIII Organization

Teresa Przytcka	NIH NCBI, USA
Knut Reinert	Freie Universität Berlin, Germany
Marie-France Sagot	INRIA, France
Cenk Sahinalp	Simon Fraisher University, Canada
David Sankoff	University of Ottawa, Canada
Russell Schwartz	Carnegie Mellon University, USA
Eran Segal	Weizmann Institute, Israel
Roded Sharan	Tel Aviv University, Israel
Mona Singh	Princeton University, USA
Donna Slonim	Tufts University, USA
Terry Speed	UC Berkeley, USA
Peter Stadler	Universität Leipzig, Germany
Anna Tramontano	University of Rome, Italy
Alfonso Valencia	Spanish National Cancer Research Centre, Spain
Ana Tereza Vasconcelos	LNCC, Brazil
Martin Vingron	Max Planck Institute for Molecular Genetics, Germany
Jerome Waldispuhl	McGill University, Canada
Tandy Warnow	University of Texas, Austin, USA
Eric Xing	Carnegie Mellon University, USA
Jinbo Xu	Toyota Technology Institute Chicago, USA

### Steering Committee

Serafim Batzoglou	Stanford University, USA
Bonnie Berger	MIT, USA
Sorin Istrail	Brown University, USA
Thomas Lengauer	Max Planck Institute for Informatics, Germany
Michal Linial	The Hebrew University of Jerusalem, Israel
Pavel Pevzner	University of California, San Diego, USA
Terry Speed	University of California, Berkeley, USA
Martin Vingron (Chair)	Max Planck Institute for Molecular Genetics, Germany

### Organizing Committee

João Carriço	University of Lisbon, Portugal
Paulo Fonseca	INESC-ID / IST, Portugal
Ana Teresa Freitas	INESC-ID / IST, Portugal
José Pereira Leal	Instituto Gulbenkian da Ciencia, Portugal
Sara Madeira	INESC-ID / IST, Portugal
Arlindo Oliveira (Chair)	INESC-ID / IST, Portugal
Luís Rocha	Indiana University, Bloomington, USA
Sara Silva	INESC-ID, Portugal
Susana Vinga	INESC-ID / IST, Portugal

## Previous RECOMB Meetings

Dates	Hosting Institution	Program Chair	Conference Chair
January 20-23, 1997, Santa Fe, NM, USA	Sandia National Lab	Michael Waterman	Sorin Istrail
March 22-25, 1998 New York, NY, USA	Mt. Sinai School of Medicine	Pavel Pevzner	Gary Benson
April 22-25, 1999 Lyon, France	INRIA	Sorin Istrail	Mireille Regnier
April 8-11, 2000 Tokyo, Japan	University of Tokyo	Ron Shamir	Satoru Miyano
April 22-25, 2001 Montreal, Canada	Université de Montreal	Thomas Lengauer	David Sankoff
April 18-21, 2002 Washington, DC, USA	Celera	Gene Myers	Sridhar Hannenhalli
April 10-13, 2003 Berlin, Germany	German Federal Ministry for Education and Research	Webb Miller	Martin Vingron
March 27-31, 2004 San Diego, USA	University of California, San Diego	Dan Gusfield	Philip E. Bourne
May 14-18, 2005 Boston, MA, USA	Broad Institute of MIT and Harvard	Satoru Miyano	Jill P. Mesirov and Simon Kasif
April 2-5, 2006 Venice, Italy	University of Padova	Alberto Apostolico	Concettina Guerra
April 21-25, 2007 San Francisco, CA, USA	QB3	Terry Speed	Sandrine Dudoit
March 30-April 2, 2008 Singapore, Singapore	National University of Singapore	Martin Vingron	Limsoon Wong
May 18-21, 2009 Tucson, AZ, USA	University of Arizona	Serafim Batzoglou	John Kececioglu

## External Reviewers

Akavia, Uri David	Bilmes, Jeff
Alkan, Can	Brudno, Michael
Alterovitz, Gil	Buhler, Jeremy
Andreotti, Sandro	Cannistraci, Carlo
Anton, Brian	Carlbach, Eyal
AnunciaÃo, Orlando	Carvalho, Alexandra
Arndt, Peter	Chen, Bo-juen
Arvestad, Lars	Chuang, Han-Yu
Atias, Nir	Chung, Ho-Ryun
Aydin, Zafer	Codeco, Claudia
Bailey, Timothy L.	Dao, Phuong
Bandyopadhyay, Sourav	Darling, Aaron
Banks, Eric	David Amir, El-ad
Baran, Yael	David, Matei
Ben-Hur, Asa	DeConde, Robert
Benyamini, Tomer	Denise, Alain
Bertsch, Andreas	Domingues, Francisco
Bielow, Chris	Donald, Bruce Randall

Dotan, Dikla  
 Efros, Tal  
 Farahani, Hossein  
 Fonseca, Paulo  
 Fox, Andrew  
 Francisco, Alexandre  
 Frank, Ari  
 Fromer, Menachem  
 Fujita, Andre  
 Furlotte, Nicholas  
 Gitter, Anthony  
 Goeke, Jonathan  
 Gottlieb, Assaf  
 Gusfield, Dan  
 Haas, Stefan  
 Hach, Faraz  
 Hackermueller, Joerg  
 Hajiresouliha, Iman  
 Han, Buhm  
 Hannum, Gregory  
 He, Dan  
 Hescott, Ben  
 Hibbs, Matthew  
 Hoffman, Michael  
 Hofree, Matan  
 Homer, Nils  
 Hormozdiari, Fereydoun  
 Huang, Yang  
 Imoto, Seiya  
 Jaitly, Navdeep  
 Jeffery, Ian  
 Kang, Eun Yong  
 Kang, Hyun Min  
 Kapp, Eugene  
 Karakoc, Emre  
 Kehr, Birte  
 Khan, Zia  
 Khardon, Roni  
 Kim, Seyoung  
 Kirkpatrick, Bonnie  
 Klau, Gunnar W.  
 Klitford, Niels  
 Kohlbacher, Oliver  
 Kundaje, Anshul  
 Kuo, Dwight

Kuser, Paula  
 Langlois, Robert  
 Lasserre, Julia  
 Lee, Byoungkoo  
 Lee, Seunghak  
 Lengauer, Thomas  
 Lin, Yu  
 Litvin, Oren  
 Liu, Manway  
 Lonardi, Stefano  
 Lu, Hui  
 Ma, Jian  
 Magger, Oded  
 Mahony, Shaun  
 Manke, Thomas  
 Mann, Tobias  
 McIlwain, Sean  
 McLoughlin, Kevin  
 Medvedev, Paul  
 Melamed, Rachel  
 Mendes, Nuno  
 Miklos, Istvan  
 Misra, Navodit  
 Molla, Michael  
 Mongiovi, Misael  
 Nachman, Iftach  
 Nagasaki, Masao  
 Navarro, Gonzalo  
 Naxerova, Kamila  
 Neuburger, Daniel  
 Niida, Atsushi  
 Noboru, Sakabe  
 Noto, Keith  
 O'Rourke, Sean  
 Oliveira, Sara  
 Ortega, Miguel  
 Pasaniuc, Bogdan  
 Pe'er, Itsik  
 Pham, Son  
 Pisanti, Nadia  
 Ponty, Yann  
 Pop, Mihai  
 Puniyani, Kriti  
 Purdom, Elizabeth  
 Pushkarev, Dmitry

Qi, Yanjun  
Qi, Yuan  
Rajan, Vaibhav  
Ramoni, Marco  
Raskutti, Garvesh  
Reza-Chitsaz, Hamid  
Rheinbau, Esther  
Rodriguez, Jesse  
Rolfe, Alex  
Rosset, Saharon  
Russo, LuÃs  
Salari, Rahele  
Sanchez Garcia, Felix  
Sankararaman, Sriram  
Sargeant, Toby  
Schaub, Marc  
Schoenhuth, Alexander  
Serin, Akdes  
Sharon, Itai  
Shibuya, Tetsuo  
Shimamura, Teppei  
Shkolnisky, Yoel  
Shomron, Noam  
Shringarpure, Suyash  
Sievers, Fabian  
Singer, Amit  
Singh, Rohit  
Sommer, Ingolf  
Song, Le  
Sridhar, Srinath

Steel, Michael  
Steffen, Martin  
Swart, Estienne  
Swenson, Krister  
Szczyrek, Ewa  
Tamada, Yoshinori  
Tannier, Eric  
Tompa, Martin  
Tsai, Ming-Chi  
Tuller, Tamir  
Turcan, Sevin  
Turcotte, Marcel  
Ulitsky, Igor  
Uyar, Bora  
Veiga, Diogo  
Vernot, Benjamin  
Vilela, Marco  
Wall, Dennis  
Wei, Xintao  
Wilm, Andreas  
Xu, Andrew Wei  
Xuan, Jianhua  
Yamaguchi, Rui  
Yang, Shu  
Yanover, Chen  
Ye, Jieping  
Zaitlen, Noah  
Zemojtel, Tomasz  
Zhang, Xiuwei  
Zhou, Xianghong (Jasmine)

# Table of Contents

An Algorithmic Framework for Predicting Side-Effects of Drugs . . . . .	1
<i>Nir Atias and Roded Sharan</i>	
SubMAP: Aligning Metabolic Pathways with Subnetwork Mappings . . . .	15
<i>Ferhat Ay and Tamer Kahveci</i>	
Admixture Aberration Analysis: Application to Mapping in Admixed Population Using Pooled DNA . . . . .	31
<i>Sivan Bercovici and Dan Geiger</i>	
Pathway-Based Functional Analysis of Metagenomes . . . . .	50
<i>Sivan Bercovici, Itai Sharon, Ron Y. Pinter, and Tomer Shlomi</i>	
Hierarchical Generative Biclustering for MicroRNA Expression Analysis . . . . .	65
<i>José Caldas and Samuel Kaski</i>	
Subnetwork State Functions Define Dysregulated Subnetworks in Cancer . . . . .	80
<i>Salim A. Chowdhury, Rod K. Nibbe, Mark R. Chance, and Mehmet Koyutürk</i>	
Proteome Coverage Prediction for Integrated Proteomics Datasets . . . .	96
<i>Manfred Claassen, Ruedi Aebersold, and Joachim M. Buhmann</i>	
Discovering Regulatory Overlapping RNA Transcripts . . . . .	110
<i>Timothy Danford, Robin Dowell, Sudeep Agarwala, Paula Grisafi, Gerald Fink, and David Gifford</i>	
Alignment-Free Phylogenetic Reconstruction . . . . .	123
<i>Constantinos Daskalakis and Sebastien Roch</i>	
Inference of Isoforms from Short Sequence Reads (Extended Abstract) . . . . .	138
<i>Jianxing Feng, Wei Li, and Tao Jiang</i>	
The Clark Phase-able Sample Size Problem: Long-Range Phasing and Loss of Heterozygosity in GWAS . . . . .	158
<i>Bjarni V. Halldórsson, Derek Aguiar, Ryan Tarpine, and Sorin Istrail</i>	
A New Algorithm for Improving the Resolution of Cryo-EM Density Maps . . . . .	174
<i>Michael Hirsch, Bernhard Schölkopf, and Michael Habeck</i>	

Towards Automated Structure-Based NMR Resonance Assignment . . . . .	189
<i>Richard Jang, Xin Gao, and Ming Li</i>	
Gapped Spectral Dictionaries and Their Applications for Database Searches of Tandem Mass Spectra . . . . .	208
<i>Kyowon Jeong, Sangtae Kim, Nuno Bandeira, and Pavel A. Pevzner</i>	
naiveBayesCall: An Efficient Model-Based Base-Calling Algorithm for High-Throughput Sequencing . . . . .	233
<i>Wei-Chun Kao and Yun S. Song</i>	
Extracting Between-Pathway Models from E-MAP Interactions Using Expected Graph Compression . . . . .	248
<i>David R. Kelley and Carl Kingsford</i>	
Simultaneous Identification of Causal Genes and Dys-Regulated Pathways in Complex Diseases . . . . .	263
<i>Yoo-Ah Kim, Stefan Wuchty, and Teresa M. Przytycka</i>	
Incremental Signaling Pathway Modeling by Data Integration . . . . .	281
<i>Geoffrey Koh, David Hsu, and P.S. Thiagarajan</i>	
The Poisson Margin Test for Normalisation Free Significance Analysis of NGS Data . . . . .	297
<i>Adam Kowalczyk, Justin Bedo, Thomas Conway, and Bryan Beresford-Smith</i>	
Compressing Genomic Sequence Fragments Using SLIMGENE . . . . .	310
<i>Christos Kozanitis, Chris Saunders, Semyon Kruglyak, Vineet Bafna, and George Varghese</i>	
On the Genealogy of Asexual Diploids . . . . .	325
<i>Fumei Lam, Charles H. Langley, and Yun S. Song</i>	
Genovo: <i>De Novo</i> Assembly for Metagenomes . . . . .	341
<i>Jonathan Laserson, Vladimir Jojic, and Daphne Koller</i>	
MoGUL: Detecting Common Insertions and Deletions in a Population . . . . .	357
<i>Seunghak Lee, Eric Xing, and Michael Brudno</i>	
Generalized Buneman Pruning for Inferring the Most Parsimonious Multi-state Phylogeny . . . . .	369
<i>Navodit Misra, Guy Belloch, R. Ravi, and Russell Schwartz</i>	
Seed Design Framework for Mapping SOLiD Reads . . . . .	384
<i>Laurent Noé, Marta Gîrdea, and Gregory Kucherov</i>	
Accurate Estimation of Expression Levels of Homologous Genes in RNA-seq Experiments . . . . .	397
<i>Bogdan Paşaniuc, Noah Zaitlen, and Eran Halperin</i>	

Cactus Graphs for Genome Comparisons . . . . .	410
<i>Benedict Paten, Mark Diekhans, Dent Earl, John St. John, Jian Ma, Bernard Suh, and David Haussler</i>	
IDBA - A Practical Iterative de Bruijn Graph De Novo Assembler . . . . .	426
<i>Yu Peng, Henry C.M. Leung, S.M. Yiu, and Francis Y.L. Chin</i>	
Predicting Nucleosome Positioning Using Multiple Evidence Tracks . . . . .	441
<i>Sheila M. Reynolds, Zhiping Weng, Jeff A. Bilmes, and William Stafford Noble</i>	
Dense Subgraphs with Restrictions and Applications to Gene Annotation Graphs . . . . .	456
<i>Barna Saha, Allison Hoch, Samir Khuller, Louiqa Raschid, and Xiao-Ning Zhang</i>	
Time and Space Efficient RNA-RNA Interaction Prediction via Sparse Folding . . . . .	473
<i>Raheleh Salari, Mathias Möhl, Sebastian Will, S. Cenk Sahinalp, and Rolf Backofen</i>	
HLA Type Inference via Haplotypes Identical by Descent . . . . .	491
<i>Manu N. Setty, Alexander Gusev, and Itsik Pe'er</i>	
Algorithms for Detecting Significantly Mutated Pathways in Cancer . . . . .	506
<i>Fabio Vandin, Eli Upfal, and Benjamin J. Raphael</i>	
Leveraging Sequence Classification by Taxonomy-Based Multitask Learning . . . . .	522
<i>Christian Widmer, Jose Leiva, Yasemin Altun, and Gunnar Rätsch</i>	
A Novel Abundance-Based Algorithm for Binning Metagenomic Sequences Using $l$ -Tuples . . . . .	535
<i>Yu-Wei Wu and Yuzhen Ye</i>	
A Markov Random Field Framework for Protein Side-Chain Resonance Assignment . . . . .	550
<i>Jianyang Zeng, Pei Zhou, and Bruce Randall Donald</i>	
Genomic DNA $k$ -mer Spectra: Models and Modalities (Abstract) . . . . .	571
<i>Benny Chor, David Horn, Nick Goldman, Yaron Levy, and Tim Massingham</i>	
Deciphering the Swine-Flu Pandemics of 1918 and 2009 (Abstract) . . . . .	572
<i>Richard Goldstein, Mario dos Reis, Asif Tamuri, and Alan Hay</i>	
Distinguishing Direct versus Indirect Transcription Factor-DNA Interactions (Abstract) . . . . .	574
<i>Raluca Gordân, Alexander J. Hartemink, and Martha L. Bulyk</i>	

A Self-regulatory System of Interlinked Signaling Feedback Loops Controls Mouse Limb Patterning (Abstract) . . . . .	575
<i>Jean-Denis Benazet, Mirko Bischofberger, Eva Tiecke, Alexandre Goncalves, James F. Martin, Aime Zuniga, Felix Naef, and Rolf Zeller</i>	
Automated High-Dimensional Flow Cytometric Data Analysis (Abstract) . . . . .	577
<i>Saamyadipta Pyne, Xinli Hu, Kui Wang, Elizabeth Rossin, Tsung-I Lin, Lisa Maier, Clare Baecher-Allan, Geoffrey McLachlan, Pablo Tamayo, David Hafler, Philip De Jager, and Jill Mesirov</i>	
Discovering Transcriptional Modules by Combined Analysis of Expression Profiles and Regulatory Sequences (Abstract) . . . . .	578
<i>Yonit Halperin, Chaim Linhart, Igor Ulitsky, and Ron Shamir</i>	
<b>Author Index</b> . . . . .	581