

Lecture Notes in Bioinformatics

8394

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

Roded Sharan (Ed.)

Research in Computational Molecular Biology

18th Annual International Conference, RECOMB 2014
Pittsburgh, PA, USA, April 2-5, 2014
Proceedings



Springer

Volume Editor

Roded Sharan
Tel-Aviv University
School of Computer Science
Tel-Aviv 69978, Israel
E-mail: roded@post.tau.ac.il

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-319-05268-7

e-ISBN 978-3-319-05269-4

DOI 10.1007/978-3-319-05269-4

Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014932543

LNCS Sublibrary: SL 8 – Bioinformatics

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

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

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The RECOMB conference series – with the full name of the Annual International Conference on Research in Computational Molecular Biology – was started in 1997 by Sorin Istrail, Pavel Pevzner, and Michael Waterman. The 18th RECOMB conference, RECOMB 2014, was held in Pittsburgh, PA, USA during April 2–5, 2014. It was hosted by Carnegie Mellon University and the University of Pittsburgh.

This volume contains the 35 extended abstracts selected for oral presentation at RECOMB 2014 by the Program Committee (PC) out of 154 submissions. Each submission was assigned to at least three PC members and reviewed with the help of many external reviewers. Following the initial reviews, there was an extensive discussion of the submissions among the members of the PC, leading to the final decisions.

This year RECOMB teamed up with *PLoS Computational Biology* to allow parallel submissions to the proceedings and the journal in a coordinated manner. I would like to thank Thomas Lengauer, deputy editor of *PLoS Computational Biology*, for the countless hours he spent on coordinating the review process and ensuring the success of this partnership. In total, 55 papers were submitted for consideration by RECOMB and *PLoS Computational Biology*. Eighteen of those papers passed an initial pre-screening for the journal and underwent a full review process. Five papers were subsequently accepted to both venues and appear as two-page abstracts in this volume; a few additional papers were accepted to only one of the venues. The five papers were published in full in a special RECOMB 2014 section of *PLoS Computational Biology*. All other papers that were accepted to RECOMB 2014 were invited for submission of an edited journal version to a special issue of the *Journal of Computational Biology*.

In addition to the paper presentations, RECOMB 2014 featured six invited keynote talks by leading scientists world-wide. The keynote speakers were Ian T. Baldwin (Max Planck Institute for Chemical Ecology), Atul Butte (Stanford University), James J. Collins (Harvard University), Trey Ideker (University of California at San Diego), Tom Mitchell (Carnegie Mellon University), and Sarah A. Tishkoff (University of Pennsylvania). Following the tradition started at RECOMB 2010, RECOMB 2014 also featured a special highlights track containing computational biology papers that were published in journals during the last 18 months. There were 48 submissions, eight of which were selected for oral presentation.

The success of RECOMB depends on the effort, dedication, and devotion of many colleagues who contributed to the organization of the conference. We thank the PC members and the external reviewers for the timely review of the assigned papers despite their busy schedules; Teresa Przytycka for chairing the highlights track; Carl Kingsford for chairing the posters track; the Steering Committee and

its chair, Martin Vingron, for many excellent suggestions on the organization of the conference; the local Organizing Committee members, especially Thom Gulish and the Co-chairs Russell Schwartz and Panayiotis (Takis) Benos, for hosting the conference and providing the administrative, logistic, and financial support; and the authors of the papers, highlights, and posters and all the attendees for their enthusiastic participation in the conference. We also thank our generous sponsors, including the International Society of Computational Biology (ISCB), the US National Science Foundation, Biomed Central GigaScience, Carnegie Mellon's Lane Center for Computational biology, and University of Pittsburgh's Department of Computational and Systems Biology. Finally, I would like to thank Yehuda Afek (Tel Aviv University), Ron Shamir, and Benny Chor (Program Chairs of RECOMB 2000 and 2012, respectively) for their support and advice.

February 2014

Roded Sharan

Conference Organization

Program Committee

Tatsuya Akutsu	Kyoto University, Japan
Max Alekseyev	University of South Carolina, USA
Joel Bader	Johns Hopkins University, USA
Vineet Bafna	University of California at San Diego, USA
Nuno Bandeira	University of California at San Diego, USA
Mukul S. Bansal	Massachusetts Institute of Technology, USA
Ziv Bar-Joseph	Carnegie Mellon University, USA
Serafim Batzoglou	Stanford University, USA
Asa Ben-Hur	Colorado State University, USA
Bonnie Berger	Massachusetts Institute of Technology, USA
Michael Brudno	University of Toronto, Canada
Sebastian Böcker	Friedrich Schiller University Jena, Germany
Brian Chen	Lehigh University, USA
Ting Chen	University of Southern California, USA
Yi-Ping Phoebe Chen	La Trobe University, USA
Chakra Chennubhotla	University of Pittsburgh, USA
Benny Chor	Tel Aviv University, Israel
Lenore Cowen	Tufts University, USA
Vincent Danos	Centre National de la Recherche Scientifique, France
Minghua Deng	Peking University, China
Colin Dewey	University of Wisconsin-Madison, USA
Nadia El-Mabrouk	University of Montreal, Canada
Sarel Fleishman	Weizmann Institute, Israel
Irit Gat-Viks	Tel Aviv University, Israel
Mikhail Gelfand	Institute for Information Transmission Problems, Russia
David Gifford	Massachusetts Institute of Technology, USA
Assaf Gottlieb	Stanford University, USA
Bjarni Halldorsson	Reykjavik University, Iceland
Wenlian Hsu	Academia Sinica, Taiwan
Sorin Istrail	Brown University, USA
Rui Jiang	Tsinghua University, China
Tao Jiang	University of California at Riverside, USA

Igor Jurisica	Ontario Cancer Institute, Canada
Tamer Kahveci	University of Florida, USA
Simon Kasif	Boston University, USA
Seyoung Kim	Carnegie Mellon University, USA
Carl Kingsford	Carnegie Mellon University, USA
	(Posters Chair)
Mehmet Koyuturk	Case Western Reserve University, USA
Rui Kuang	University of Minnesota Twin Cities, USA
Jens Lagergren	KTH Royal Institute of Technology, Sweden
Hyunju Lee	Gwangju Institute of Science and Technology, South Korea
Michal Linial	The Hebrew University of Jerusalem, Israel
Jinze Liu	University of Kentucky, USA
Stefano Lonardi	University of California at Riverside, USA
Tijana Milenkovic	University of Notre Dame, USA
Satoru Miyano	University of Tokyo, Japan
Bernard Moret	Ecole Polytechnique Federale de Lausanne, Switzerland
T. M. Murali	Virginia Polytechnic Institute and State University, USA
William Stafford Noble	University of Washington, USA
Arlindo Oliveira	IST/INESC-ID and Cadence Research Laboratories, Portugal
Laxmi Parida	IBM T. J. Watson Research Center, USA
Itsik Pe'er	Columbia University, USA
Teresa Przytycka	National Institute of Health, USA
	(Highlights Chair)
Ben Raphael	Brown University, USA
Knut Reinert	Free University Berlin, Germany
Marie-France Sagot	Inria Grenoble Rhône-Alpes and Université de Lyon 1, France
S. Cenk Sahinalp	Simon Fraser University, Canada
Dina Schneidman-Duhovny	University of California at San Francisco, USA
Denise Scholtens	Northwestern University, USA
Russell Schwartz	Carnegie Mellon University, USA
Roded Sharan	Tel Aviv University, Israel (Chair)
Mona Singh	Princeton University, USA
Andrew Smith	University of Southern California, USA
Berend Snel	Utrecht University, The Netherlands
Kai Tan	University of Iowa, USA

Alfonso Valencia	Spanish National Cancer Research Centre, Spain
Fabio Vandin	University of Southern Denmark, Denmark
Jean-Philippe Vert	Ecole des Mines de Paris, France
Martin Vingron	Max Planck Institute for Molecular Genetics, Germany
Lusheng Wang	City University of Hong Kong, Hong Kong
Tandy Warnow	University of Texas at Austin, USA
Haim J. Wolfson	Tel Aviv University, Israel
Limsoon Wong	National University of Singapore, Singapore
Xiaohui Xie	University of California at Irvine, USA
Dong Xu	University of Missouri-Columbia, USA
Jinbo Xu	Toyota Technology Institute at Chicago, USA
Yuzhen Ye	Indiana University, USA
Esti Yeger-Lotem	Ben-Gurion University, Israel
Sungroh Yoon	Seoul National University, South Korea
Louxin Zhang	National University of Singapore, Singapore
Xuegong Zhang	Tsinghua University, China
Hongyu Zhao	Yale University, USA

Steering Committee

Vineet Bafna	University of California at San Diego, USA
Serafim Batzoglou	Stanford University, USA
Bonnie Berger	Massachusetts Institute of Technology, USA
Sorin Istrail	Brown University, USA
Michal Linial	The Hebrew University of Jerusalem, Israel
Martin Vingron	Max Planck Institute for Molecular Genetics, Germany (Chair)

Organizing Committee

Russell Schwartz	Carnegie Mellon University, USA (Co-chair)
Panayiotis (Takis) Benos	University of Pittsburgh, USA (Co-chair)
Thom Gulish	Carnegie Mellon University, USA (Local Arrangements Coordinator)
Carl Kingsford	Carnegie Mellon University, USA
Darya Filippova	Carnegie Mellon University, USA
Nichole Merrit	Carnegie Mellon University, USA

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 Kececiloglu
August 12-15, 2010 Lisbon, Portugal	INESC-ID and Instituto Superior Técnico	Bonnie Berger	Arlindo Oliveira
March 28-31, 2011 Vancouver, Canada	Lab for Computational Biology, Simon Fraser University	Vineet Bafna	S. Cenk Sahinalp
April 21-24, 2012 Barcelona, Spain	Centre for Genomic Regulation	Benny Chor	Roderic Guigó
April 7-10, 2013 Beijing, China	Tsinghua University	Fengzhu Sun	Xuegong Zhang
April 2-5, 2014 Pittsburgh, PA, USA	Carnegie Mellon University and University of Pittsburgh	Roded Sharan	Russell Schwartz and Panayiotis (Takis) Benos

External Reviewers

Aganezov, Sergey
 Aguiar, Derek
 Ali, Raja Hashim
 Antonov, Ivan
 Arvestad, Lars
 Atias, Nir
 Ayati, Marzieh
 Bandyopadhyay, Nirmalya
 Bebek, Gurkan
 Behnam, Ehsan
 Beissbarth, Tim
 Ben-Bassat, Ilan
 Bienkowska, Jadwiga
 Biesinger, Jacob
 Botzman, Maya

Brodt, Avital
 Bucher, Philipp
 Carmi, Shai
 Chen, Caster
 Chiu, Jimmy Ka Ho
 Cho, Dongyeon
 Chowdhury, Salim
 Chung, Ho-Ryun
 Cicek, Ercument
 Csürös, Miklós
 Daley, Timothy
 Daniels, Noah
 Dao, Phuong
 Dolzhenko, Egor
 Donmez, Nilgun

Dror, Ron	Lavenier, Dominique
Duma, Denise	Leiserson, Mark
Edwards, Matt	Li, Cong
Emde, Anne-Katrin	Li, Ming
Eskin, Eleazar	Li, Wei
Faisal, Fazle	Li, Yi
Farahani, Hossein S.	Libbrecht, Max
Frishberg, Amit	Lin, Yen Yi
Frånberg, Mattias	Lynn, Ke-Shiuan
Fujita, Andre	Ma, Jianzhu
Gabr, Haitham	Ma, Wenxiu
Gao, Long	Ma, Xiaoke
Ghersi, Dario	Makeev, Vsevolod
Gitter, Anthony	Mariadassou, Mahendra
Greenman, Christopher	McPherson, Andrew
Guo, Yuchun	Melsted, Páll
Guthals, Adrian	Meusel, Marvin
Hach, Faraz	Mirebrahim, Seyed
Haiminen, Niina	Mohammadi, Shahin
Hajirasouliha, Iman	Morris, Quaid
Hasan, Mahmudul	Na, Seungjin
Haws, David	Nachshon, Aharon
He, Bing	Navlakha, Saket
He, Dan	Newburger, Daniel
He, Xin	Niida, Atsushi
He, Zhiquan	Nikolenko, Sergey
Hildebrandt, Andreas	Numanagic, Ibrahim
Hodzic, Ermin	O'Donnell, Charles W.
Hufsky, Franziska	Oesper, Layla
Hui, Ken	Ounit, Rachid
Hulovatyy, Yuriy	Palmer, Cameron
Jiang, Shuai	Park, Heewon
Jiang, Yue	Park, Seunghyun
Jónsson, Björn	Pauwels, Edouard
Kalinina, Anastasiya	Platt, Daniel
Kashef, Dorna	Polishko, Anton
Keles, Sunduz	Popic, Victoria
Kim, Yoo-Ah	Pritykin, Yuri
Korobeynikov, Anton	Prjibelski, Andrey D.
Kreimer, Anat	Przytycki, Pawel
Kuang, Rui	Qu, Jenny
Kuleshov, Volodymyr	Quang, Daniel
Kwon, Sunyoung	Rajasekaran, Rajalakshmi
Kyriazopoulou-Panagiotopoulou, Sofia	Richard, Hugues
Lafond, Manuel	Ritz, Anna

Robin, Stéphane
Rubinov, Anatoly
Ruffalo, Matthew
Röst, Hannes
Sacomoto, Gustavo
Sadeh, Mohammad
Sahlin, Kristoffer
Salari, Raheleh
Savel, Daniel
Scheubert, Kerstin
Schloissnig, Siegfried
Schulz, Marcel
Sefer, Emre
Shen, Yang
Sheridan, Paul
Shiraishi, Yuichi
Song, Jimin
Srivastava, Gyan
Steuerma, Yael
Tegge, Allison
Teng, Li
Todor, Andrei
Ullah, Ikram
Uren, Philip

Utro, Filippo
Vinogradov, Dmitry
Wagner, Allon
Wang, Jian
Wang, Mingxun
Wang, Qingguo
Wang, Sheng
Wang, Zheng
Wang, Zhiyong
White, Tim
Wilenzik, Roni
Wise, Aaron
Wu, Hsin-Ta
Wu, Yu-Wei
Wójtowicz, Damian
Yang, Can
Yang, Ei-Wen
Yao, Qiuming
Yiu, Siuming
Yu, Seunghak
Yu, Zhaoxia
Zhang, Kui
Zhao, Feng
Zhong, Shan

Table of Contents

Tractatus: An Exact and Subquadratic Algorithm for Inferring Identical-by-Descent Multi-shared Haplotype Tracts	1
<i>Derek Aguiar, Eric Morrow, and Sorin Istrail</i>	
HapTree: A Novel Bayesian Framework for Single Individual Polyplootyping Using NGS Data	18
<i>Emily Berger, Deniz Yorukoglu, Jian Peng, and Bonnie Berger</i>	
Changepoint Analysis for Efficient Variant Calling	20
<i>Adam Bloniarz, Ameet Talwalkar, Jonathan Terhorst, Michael I. Jordan, David Patterson, Bin Yu, and Yun S. Song</i>	
On the Representation of de Bruijn Graphs	35
<i>Rayan Chikhi, Antoine Limasset, Shaun Jackman, Jared T. Simpson, and Paul Medvedev</i>	
Exact Learning of RNA Energy Parameters from Structure.....	56
<i>Hamidreza Chitsaz and Mohammad Aminisharifabad</i>	
An Alignment-Free Regression Approach for Estimating Allele-Specific Expression Using RNA-Seq Data.....	69
<i>Chen-Ping Fu, Vladimir Jojić, and Leonard McMillan</i>	
The Generating Function Approach for Peptide Identification in Spectral Networks	85
<i>Adrian Guthals, Christina Boucher, and Nuno Bandeira</i>	
Decoding Coalescent Hidden Markov Models in Linear Time	100
<i>Kelley Harris, Sara Sheehan, John A. Kamm, and Yun S. Song</i>	
AptaCluster – A Method to Cluster HT-SELEX Aptamer Pools and Lessons from Its Application	115
<i>Jan Hoinka, Alexey Berezhnoy, Zuben E. Sauna, Eli Gilboa, and Teresa M. Przytycka</i>	
Learning Sequence Determinants of Protein: Protein Interaction Specificity with Sparse Graphical Models.....	129
<i>Hetunandan Kamisetty, Bornika Ghosh, Christopher James Langmead, and Chris Bailey-Kellogg</i>	
On Sufficient Statistics of Least-Squares Superposition of Vector Sets ...	144
<i>Arun S. Konagurthu, Parthan Kasarapu, Lloyd Allison, James H. Collier, and Arthur M. Lesk</i>	

IDBA-MTP: A Hybrid MetaTranscriptomic Assembler Based on Protein Information	160
<i>Henry C.M. Leung, S.M. Yiu, and Francis Y.L. Chin</i>	
MRAlign: Protein Homology Detection through Alignment of Markov Random Fields	173
<i>Jianzhu Ma, Sheng Wang, Zhiyong Wang, and Jinbo Xu</i>	
An Integrated Model of Multiple-Condition ChIP-Seq Data Reveals Predeterminants of Cdx2 Binding	175
<i>Shaun Mahony, Matthew D. Edwards, Esteban O. Mazzoni, Richard I. Sherwood, Akshay Kakumanu, Carolyn A. Morrison, Hynek Wichterle, and David K. Gifford</i>	
PASTA: Ultra-Large Multiple Sequence Alignment	177
<i>Siavash Mirarab, Nam Nguyen, and Tandy Warnow</i>	
Fast Flux Module Detection Using Matroid Theory	192
<i>Arne C. Müller, Frank J. Bruggeman, Brett G. Olivier, and Leen Stougie</i>	
Building a Pangenome Reference for a Population	207
<i>Ngan Nguyen, Glenn Hickey, Daniel R. Zerbino, Brian Raney, Dent Earl, Joel Armstrong, David Haussler, and Benedict Paten</i>	
CSAX: Characterizing Systematic Anomalies in eXpression Data	222
<i>Keith Noto, Carla Brodley, Saeed Majidi, Diana W. Bianchi, and Donna K. Slonim</i>	
WHATSHAP: Haplotype Assembly for Future-Generation Sequencing Reads	237
<i>Murray Patterson, Tobias Marschall, Nadia Pisanti, Leo van Iersel, Leen Stougie, Gunnar W. Klau, and Alexander Schönhuth</i>	
Simultaneous Inference of Cancer Pathways and Tumor Progression from Cross-Sectional Mutation Data	250
<i>Benjamin J. Raphael and Fabio Vandin</i>	
DIPSPADES: Assembler for Highly Polymorphic Diploid Genomes	265
<i>Yana Safonova, Anton Bankevich, and Pavel A. Pevzner</i>	
An Exact Algorithm to Compute the DCJ Distance for Genomes with Duplicate Genes	280
<i>Mingfu Shao, Yu Lin, and Bernard Moret</i>	

HIT'nDRIVE: Multi-driver Gene Prioritization Based on Hitting Time	293
<i>Raunak Shrestha, Ermin Hodzic, Jake Yeung, Kendric Wang, Thomas Sauerwald, Phuong Dao, Shawn Anderson, Himisha Beltran, Mark A. Rubin, Colin C. Collins, Gholamreza Haffari, and S. Cenk Sahinalp</i>	
Modeling Mutual Exclusivity of Cancer Mutations	307
<i>Ewa Szczurek and Niko Beerenwinkel</i>	
Viral Quasispecies Assembly via Maximal Clique Enumeration	309
<i>Armin Töpfer, Tobias Marschall, Rowena A. Bull, Fabio Luciani, Alexander Schönhuth, and Niko Beerenwinkel</i>	
Correlated Protein Function Prediction via Maximization of Data-Knowledge Consistency	311
<i>Hua Wang, Heng Huang, and Chris Ding</i>	
Bayesian Multiple Protein Structure Alignment	326
<i>Rui Wang and Scott C. Schmidler</i>	
Gene-Gene Interactions Detection Using a Two-Stage Model	340
<i>Zhanyong Wang, Jae-Hoon Sul, Sagi Snir, Jose A. Lozano, and Eleazar Eskin</i>	
A Geometric Clustering Algorithm and Its Applications to Structural Data	356
<i>Shutan Xu, Shuxue Zou, and Lincong Wang</i>	
A Spatial-Aware Haplotype Copying Model with Applications to Genotype Imputation	371
<i>Wen-Yun Yang, Farhad Hormozdiari, Eleazar Eskin, and Bogdan Pasaniuc</i>	
Traversing the k -mer Landscape of NGS Read Datasets for Quality Score Sparsification	385
<i>Y. William Yu, Deniz Yorukoglu, and Bonnie Berger</i>	
Reconstructing Breakage Fusion Bridge Architectures Using Noisy Copy Numbers	400
<i>Shay Zakov and Vineet Bafna</i>	
Reconciliation with Non-binary Gene Trees Revisited	418
<i>Yu Zheng and Louxin Zhang</i>	

Learning Protein-DNA Interaction Landscapes by Integrating Experimental Data through Computational Models	433
<i>Jianling Zhong, Todd Wasson, and Alexander J. Hartemink</i>	
Imputation of Quantitative Genetic Interactions in Epistatic MAPs by Interaction Propagation Matrix Completion	448
<i>Marinka Žitnik and Blaž Zupan</i>	
Author Index	463