

Subseries of Lecture Notes in Computer Science

LNBI 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

LNBI Editorial Board

Søren Brunak

Technical University of Denmark, Kongens Lyngby, Denmark

Mikhail S. Gelfand

IITP, Research and Training Center on Bioinformatics, Moscow, Russia

Thomas Lengauer

Max Planck Institute for Informatics, Saarbrücken, Germany

Satoru Miyano

University of Tokyo, Tokyo, Japan

Eugene Myers

Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

Marie-France Sagot

Université Lyon 1, Villeurbanne, France

David Sankoff

University of Ottawa, Ottawa, Canada

Ron Shamir

Tel Aviv University, Ramat Aviv, Tel Aviv, Israel

Terry Speed

Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia

Martin Vingron

Max Planck Institute for Molecular Genetics, Berlin, Germany

W. Eric Wong

University of Texas at Dallas, Richardson, TX, USA

More information about this series at <http://www.springer.com/series/5381>

S. Cenk Sahinalp (Ed.)

Research in Computational Molecular Biology

21st Annual International Conference, RECOMB 2017
Hong Kong, China, May 3–7, 2017
Proceedings

Editor
S. Cenk Sahinalp
Indiana University Bloomington
Bloomington, IN
USA

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Bioinformatics
ISBN 978-3-319-56969-7 ISBN 978-3-319-56970-3 (eBook)
DOI 10.1007/978-3-319-56970-3

Library of Congress Control Number: 2017936939

LNCS Sublibrary: SL8 – Bioinformatics

© Springer International Publishing AG 2017

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.

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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

RECOMB, the Annual International Conference on Research in Computational Molecular Biology started in 1997, under the leadership of Sorin Istrail, Pavel Pevzner, and Michael Waterman. This year marks its 21st anniversary. RECOMB 2017 was hosted by The University of Hong Kong and The Chinese University of Hong Kong during May 3–7, 2017. This volume contains the 38 extended or short abstracts selected for oral presentation at RECOMB 2017 by the Program Committee (PC). Each of the 184 submissions consisted of a full paper, and was assigned to at least three PC members and reviewed with the help of many external reviewers. Following the initial reviews, final decisions were made after an extensive discussion of the submissions among the members of the PC.

Even though RECOMB 2017 did not allow parallel submissions, authors of accepted papers were given the option to publish short abstracts in the proceedings and submit their full papers to a journal. The papers for which the proceedings feature short abstracts had appeared in a journal by the time of the conference and were to be deposited in the preprint server arxiv.org. All other papers that appear as long abstracts in the proceedings were invited for submission to RECOMB 2017 special issues by either *Cell Systems* or the *Journal of Computational Biology*.

In addition to the paper presentations, RECOMB 2017 featured six invited keynote talks by leading scientists worldwide. The keynote speakers were Colin Collins (Vancouver Prostate Centre), Joe Gray (Oregon Health Sciences University), Wang Jun (iCarbonX), Laxmi Parida (IBM T.J. Watson Research Center), Ben Raphael (Princeton University), and Michael Schnell Levin (10X Genomics).

Following the tradition started at RECOMB 2010, RECOMB 2017 also featured highlight talks presenting computational biology papers that were published in journals during the last 18 months. There were 32 highlight submissions, six of which were selected for oral presentation at the main conference.

The success of RECOMB depends on the effort, dedication, and devotion of many colleagues. I especially thank the Organizing Committee chair, Siu Ming Yiu (The University of Hong Kong), and Kevin Yip (The Chinese University of Hong Kong), for hosting RECOMB and RECOMB-Seq 2017, the RECOMB satellite meeting on Massively Parallel Sequencing; Sumaiya Nazeen (MIT) for website design and technical support; the Steering Committee and especially its chair, Bonnie Berger (MIT), for help, advice, and support throughout the process; Mona Singh, the Program Chair of RECOMB 2016 (Princeton), for answering my many questions; Mathieu Blanchette (McGill) for chairing the highlights track; Paul Medvedev (Penn State) for acting as the publications chair; Alex Schoenhuth (CWI) for acting as the publicity chair; Fereydoun Hormozdiari (UC Davis) and Jian Ma (CMU) for chairing RECOMB-Seq; the main conference and RECOMB-Seq PC members and external reviewers for their timely reviews of assigned papers despite their busy schedules; the authors of the papers, highlights, and posters for their scientific contributions; and all the attendees for their

enthusiastic participation in the conference. We also thank the International Society of Computational Biology (ISCB) for student support and the Croucher Foundation for additional sponsorship.

February 2017

S. Cenk Sahinalp

Organization

Program Committee

Max Alekseyev	George Washington University, USA
Rolf Backofen	Albert Ludwigs University of Freiburg, Germany
Vineet Bafna	University of California, San Diego, USA
Chris Bailey-Kellogg	Dartmouth College, USA
Nuno Bandeira	UCSD, USA
Ziv Bar-Joseph	Carnegie Mellon University, USA
Niko Beerenwinkel	ETH Zurich, Switzerland
Bonnie Berger	Massachusetts Institute of Technology, USA
Mathieu Blanchette	McGill University, USA
Sebastian Böcker	Friedrich Schiller University of Jena, Germany
Lenore Cowen	Tufts University, USA
Nadia El-Mabrouk	University of Montreal, Canada
Irit Gat-Viks	Tel Aviv University, Israel
David Gifford	MIT, USA
Raluca Gordan	Duke University, USA
Fereydoun Hormozdiari	University of California, Davis, USA
Trey Ideker	University of California, San Diego, USA
Tao Jiang	University of California, Riverside, USA
Vladimir Jovic	University of North Carolina, USA
John Kececioglu	University of Arizona, USA
Manolis Kellis	MIT, USA
Carl Kingsford	Carnegie Mellon University, USA
Gunnar W. Klau	CWI, The Netherlands
Jens Lagergren	SBC and CSC, KTH
Max Leiserson	Tufts University, USA
Ming Li	University of Waterloo, Canada
Jian Ma	Carnegie Mellon University, USA
Paul Medvedev	Pennsylvania State University, USA
Bernard Moret	EPFL, Switzerland
Veli Mäkinen	University of Helsinki, Finland
William Stafford Noble	University of Washington, USA
Laxmi Parida	IBM T.J. Watson Research Center, USA
Bogdan Pasaniuc	UCLA, USA
Jian Peng	University of Illinois at Urbana-Champaign, USA
Yann Ponty	CNRS/LIX, Polytechnique
Teresa Przytycka	NIH, USA
Ben Raphael	Princeton University, USA
Knut Reinert	FU Berlin, Germany

S. Cenk Sahinalp	Indiana University, Bloomington, USA
Alexander Schoenhuth	Centrum Wiskunde and Informatica, The Netherlands
Russell Schwartz	Carnegie Mellon University, USA
Roded Sharan	Tel Aviv University, Israel
Mona Singh	Princeton University, USA
Donna Slonim	Tufts University, USA
Sagi Snir	Institute of Evolution
Leen Stougie	VU University
Jens Stoye	Bielefeld University, Germany
Fengzhu Sun	University of Southern California, USA
Wing-Kin Sung	Nuational University of Singapore, Singapore
Glenn Tesler	University of California, San Diego, USA
Tamir Tuller	Tel Aviv University, Israel
Alfonso Valencia	Spanish National Cancer Research Centre, Spain
Fabio Vandin	University of Padova, Italy
Martin Vingron	Max Planck Institut für molekulare Genetik, Germany
Jerome Waldispuhl	McGill University, Canada
Sebastian Will	University of Leipzig, Germany
Jinbo Xu	Toyota Technological Institute at Chicago, USA
Noah Zaitlen	University of California San Francisco, USA
Alex Zelikovsky	Georgia State University, USA
Jianyang Zeng	Tsinghua University, China
Louxin Zhang	National University of Singapore, Singapore
Xuegong Zhang	Tsinghua University, China

Additional Reviewers

Aganezov, Sergey	Biran, Hadas	Cunial, Fabio
Aguiar, Derek	Boix, Carles	Dahl, Andy
Alachiotis, Nikolaos	Bonora, Giancarlo	Daniels, Noah
Alexeev, Nikita	Bordewich, Magnus	Dao, Phuong
Alkhnabashi, Omer	Bryant, David	Deblasio, Dan
Andonov, Rumen	Canzar, Stefan	Ding, Jun
Artyomenko, Alexander	Cardner, Mathias	Dirmeier, Simon
Arvestad, Lars	Castillo, Omar	Doerr, Daniel
Avdeyev, Pavel	Chaisson, Mark	Doty, David
Backofen, Rolf	Cho, Hoon	Eetemadi, Ameen
Bankevich, Anton	Cho, Hyunghoon	El-Kebir, Mohammed
Bansal, Vikas	Chor, Benny	Emde, Anne-Katrin
Batu, Tugkan	Cichonska, Anna	Engler, Martin
Berry, Vincent	Cook, Kate	Eslami Rasekh, Marzieh
Bhadra, Sahely	Costa, Fabrizio	Fallmann, Joerg
Bhutani, Kunal	Crawford, Jake	Feijao, Pedro
Bielow, Chris	Csuros, Miklos	Frenkel, Zeev

Frishberg, Amit	Käll, Lukas	Palmer, Cameron
G. Costa, Ivan	Köster, Johannes	Park, Danny
Gagie, Travis	Lafond, Manuel	Park, Yongjin
Gao, Tianxiang	Langmead, Christopher	Peng, Yu
Glynn, Eric	Lee, Heewook	Persi, Erez
Golumbeanu, Monica	Lei, Jinzhi	Persikov, Anton
Gottlieb, Assaf	Lemaitre, Claire	Pham, Son
Gruenewald, Stefan	Lemieux, Sebastien	Pirkli, Martin
Gu, Jin	Levy, Maya	Pisanti, Nadia
Guo, Yuchun	Li, Wenyuan	Pittala, Srivamshi
Hach, Faraz	Li, Yue	Pockrandt, Christopher
Hajirasouliha, Iman	Libbrecht, Max	Pons Mayol, Joan Carles
Halldorsson, Bjarni	Limasset, Antoine	Przytycki, Pawel
Harel, Tom	Lin, Dejun	Pullman, Benjamin
Heller, David	Lin, Yen Yi	Rashid, Sabrina
Herman, Pawel	Liu, Yan	Reinharz, Vladimir
Hescott, Ben	Liu, Yaping	Ren, Jie
Hescott, Benjamin	Loh, Po-Ru	Reyna, Matthew
Hiaminen, Niina	Lu, Yang	Rhrissorakrai, Kahn
Hobolth, Asger	Ludwig, Marcus	Ruffalo, Matthew
Hodzic, Ermin	Luhmann, Nina	Rusch, Doug
Holley, Guillaume	Löytynoja, Ari	Saglam, Mert
Homilius, Max	Ma, Cong	Sahlin, Kristoffer
Hormozdiari, Farhad	Ma, Jianzhu	Salmela, Leena
Howbert, Jeff	Malikic, Salem	Sanguinetti, Guido
Hua, Kui	Mandric, Igor	Sarkar, Abhishek
Huang, Justin	Mann, Martin	Sauerwald, Natalie
Huska, Matt	Marcais, Guillaume	Schreiber, Jacob
Huynh-Thu, Vân Anh	May, Damon	Schöpflin, Robert
Jahn, Katharina	Mazza, Arnon	Scott, Camille
Jain, Siddhartha	Mefford, Joel	Shao, Mingfu
Joseph, Tyler	Meuleman, Wouter	Shi, Alvin
Jünemann, Sebastian	Minkin, Ilia	Shrestha, Raunak
Kamm, John	Mirzaei, Sajad	Shteyman, Alan
Kehr, Birte	Munro, Daniel	Silberberg, Yael
Keich, Uri	Na, Seungjin	Silverbush, Dana
Kelk, Steven	Nadimpalli, Shilpa	Simmons, Sean
Kim, Juho	Navlakha, Saket	Sindi, Suzanne
Kim, Yoo-Ah	Noutahi, Emmanuel	Singer, Jochen
Knyazev, Sergey	Numanagic, Ibrahim	Siragusa, Enrico
Kockan, Can	Nurk, Sergey	Solomon, Brad
Kopp, Wolfgang	Oesper, Layla	Stolzer, Maureen
Koyama, Taka	Ohler, Uwe	Stoye, Jens
Kuipers, Jack	Oren, Yael	Sun, Chen
Kundu, Kousik	Orenstein, Yaron	Sundermann, Linda
Kundu, Ritu	Ouangraoua, Aida	Swenson, Krister

Syed, Tahin	Wall, Timothy	Xin, Hongyi
Tang, Haixu	Wang, Lusheng	Yang, Shuo
Tang, Kujin	Wang, Mingxun	Yanover, Chen
Thankachan, Sharma V.	Wang, Sheng	Yardimci, Galip
Thurnherr, Thomas	Wang, Tina	Ye, Yuzhen
Tremblay-Savard, Olivier	Wang, Weili	Yeo, Grace
Utro, Filippo	Wang, Yijie	Yilmaz, Sule
Uurtio, Viivi	Wang, Ying	You, Xintian
Valenzuela, Daniel	Wetzel, Joshua	Yu, Michael
van Iersel, Leo	White, Tim	Yu, Ning
Varoquaux, Nelle	Wienbrandt, Lars	Zamalloa, Jose
Verma, Deeptak	Wilentzik, Roni	Zekic, Tina
Viduani Martinez, Fábio Henrique	Wise, Aaron	Zeng, Haoyang
von Kleist, Max	Wittler, Roland	Zhang, Mengge
	Wójtowicz, Damian	Zhang, Zhizhuo

Contents

Boosting Alignment Accuracy by Adaptive Local Realignment.	1
<i>Dan DeBlasio and John Kececioglu</i>	
A Concurrent Subtractive Assembly Approach for Identification of Disease Associated Sub-metagenomes	18
<i>Wontack Han, Mingjie Wang, and Yuzhen Ye</i>	
A Flow Procedure for the Linearization of Genome Sequence Graphs	34
<i>David Haussler, Maciej Smuga-Otto, Benedict Paten, Adam M. Novak, Sergei Nikitin, Maria Zueva, and Dmitrii Miagkov</i>	
Dynamic Alignment-Free and Reference-Free Read Compression	50
<i>Guillaume Holley, Roland Wittler, Jens Stoye, and Faraz Hach</i>	
A Fast Approximate Algorithm for Mapping Long Reads to Large Reference Databases	66
<i>Chirag Jain, Alexander Dilthey, Sergey Koren, Srinivas Aluru, and Adam M. Phillippy</i>	
Determining the Consistency of Resolved Triplets and Fan Triplets.	82
<i>Jesper Jansson, Andrzej Lingas, Ramesh Rajaby, and Wing-Kin Sung</i>	
Progressive Calibration and Averaging for Tandem Mass Spectrometry Statistical Confidence Estimation: Why Settle for a Single Decoy?	99
<i>Uri Keich and William Stafford Noble</i>	
Resolving Multicopy Duplications <i>de novo</i> Using Polyploid Phasing.	117
<i>Mark J. Chaisson, Sudipto Mukherjee, Sreeram Kannan, and Evan E. Eichler</i>	
A Bayesian Active Learning Experimental Design for Inferring Signaling Networks.	134
<i>Robert Osazuwa Ness, Karen Sachs, Parag Mallick, and Olga Vitek</i>	
<i>BBK*</i> (Branch and Bound over K^*): A Provable and Efficient Ensemble-Based Algorithm to Optimize Stability and Binding Affinity over Large Sequence Spaces.	157
<i>Adegoke A. Ojewole, Jonathan D. Jou, Vance G. Fowler, and Bruce R. Donald</i>	
Superbubbles, Ultrabubbles and Cacti	173
<i>Benedict Paten, Adam M. Novak, Erik Garrison, and Glenn Hickey</i>	

EPR-Dictionaries: A Practical and Fast Data Structure for Constant Time Searches in Unidirectional and Bidirectional FM Indices 190
Christopher Pockrandt, Marcel Ehrhardt, and Knut Reinert

A Bayesian Framework for Estimating Cell Type Composition from DNA Methylation Without the Need for Methylation Reference 207
Elior Rahmani, Regev Schweiger, Liat Shenhav, Eleazar Eskin, and Eran Halperin

Towards Recovering Allele-Specific Cancer Genome Graphs 224
Ashok Rajaraman and Jian Ma

Using Stochastic Approximation Techniques to Efficiently Construct Confidence Intervals for Heritability 241
Regev Schweiger, Eyal Fisher, Elior Rahmani, Liat Shenhav, Saharon Rosset, and Eran Halperin

Improved Search of Large Transcriptomic Sequencing Databases Using Split Sequence Bloom Trees 257
Brad Solomon and Carl Kingsford

AllSome Sequence Bloom Trees 272
Chen Sun, Robert S. Harris, Rayan Chikhi, and Paul Medvedev

Longitudinal Genotype-Phenotype Association Study via Temporal Structure Auto-learning Predictive Model 287
Xiaoqian Wang, Jingwen Yan, Xiaohui Yao, Sungeun Kim, Kwangsik Nho, Shannon L. Risacher, Andrew J. Saykin, Li Shen, Heng Huang, and for the ADNI

Improving Imputation Accuracy by Inferring Causal Variants in Genetic Studies 303
Yue Wu, Farhad Hormozdiari, Jong Wha J. Joo, and Eleazar Eskin

The Copy-Number Tree Mixture Deconvolution Problem and Applications to Multi-sample Bulk Sequencing Tumor Data 318
Simone Zaccaria, Mohammed El-Kebir, Gunnar W. Klau, and Benjamin J. Raphael

Quantifying the Impact of Non-coding Variants on Transcription Factor-DNA Binding 336
Jingkang Zhao, Dongshunyi Li, Jungkyun Seo, Andrew S. Allen, and Raluca Gordân

aBayesQR: A Bayesian Method for Reconstruction of Viral Populations Characterized by Low Diversity 353
Soyeon Ahn and Haris Vikalo

BeWith: A Between-Within Method for Module Discovery in Cancer using Integrated Analysis of Mutual Exclusivity, Co-occurrence and Functional Interactions (Extended Abstract) 370
Phuong Dao, Yoo-Ah Kim, Sanna Madan, Roded Sharan, and Teresa M. Przytycka

K-mer Set Memory (KSM) Motif Representation Enables Accurate Prediction of the Impact of Regulatory Variants 372
Yuchun Guo, Kevin Tian, Haoyang Zeng, and David K. Gifford

Network-Based Coverage of Mutational Profiles Reveals Cancer Genes. 375
Borislav H. Hristov and Mona Singh

Ultra-Accurate Complex Disorder Prediction: Case Study of Neurodevelopmental Disorders 377
Linh Huynh and Fereydoun Hormozdiari

Inference of the Human Polyadenylation Code 379
Michael K.K. Leung, Andrew Delong, and Brendan J. Frey

Folding Membrane Proteins by Deep Transfer Learning. 380
Zhen Li, Sheng Wang, Yizhou Yu, and Jinbo Xu

A Network Integration Approach for Drug-Target Interaction Prediction and Computational Drug Repositioning from Heterogeneous Information 383
Yunan Luo, Xinbin Zhao, Jingtian Zhou, Jinling Yang, Yanqing Zhang, Wenhua Kuang, Jian Peng, Ligong Chen, and Jianyang Zeng

Epistasis in Genomic and Survival Data of Cancer Patients 385
Dariusz Matlak and Ewa Szczurek

Ultra-Fast Identity by Descent Detection in Biobank-Scale Cohorts Using Positional Burrows-Wheeler Transform. 387
Ardalan Naseri, Xiaoming Liu, Shaojie Zhang, and Degui Zhi

Joker de Bruijn: Sequence Libraries to Cover All *k*-mers Using Joker Characters 389
Yaron Orenstein, Ryan Kim, Polly Fordyce, and Bonnie Berger

GATTACA: Lightweight Metagenomic Binning Using Kmer Counting 391
Victoria Popic, Volodymyr Kuleshov, Michael Snyder, and Serafim Batzoglou

Species Tree Estimation Using ASTRAL: How Many Genes Are Enough? 393
Shubhanshu Shekhar, Sebastien Roch, and Siavash Mirarab

Reconstructing Antibody Repertoires from Error-Prone Immunosequencing Datasets	396
<i>Alexander Shlemov, Sergey Bankevich, Andrey Bzikadze, Yana Safonova, and Pavel A. Pevzner</i>	
NetREX: Network Rewiring Using EXpression - Towards Context Specific Regulatory Networks	398
<i>Yijie Wang, Dong-Yeon Cho, Hangnoh Lee, Brian Oliver, and Teresa M. Przytycka</i>	
E Pluribus Unum: United States of Single Cells	400
<i>Joshua D. Welch, Alexander Hartemink, and Jan F. Prins</i>	
ROSE: A Deep Learning Based Framework for Predicting Ribosome Stalling	402
<i>Sai Zhang, Hailin Hu, Jingtian Zhou, Xuan He, Tao Jiang, and Jianyang Zeng</i>	
Author Index	405