

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>

Benjamin J. Raphael (Ed.)

Research in Computational Molecular Biology

22nd Annual International Conference, RECOMB 2018
Paris, France, April 21–24, 2018
Proceedings

Editor

Benjamin J. Raphael
Computer Science Department
Princeton University
Princeton, NJ
USA

ISSN 0302-9743

ISSN 1611-3349 (electronic)

Lecture Notes in Bioinformatics

ISBN 978-3-319-89928-2

ISBN 978-3-319-89929-9 (eBook)

<https://doi.org/10.1007/978-3-319-89929-9>

Library of Congress Control Number: 2018940142

LNCS Sublibrary: SL8 – Bioinformatics

© Springer International Publishing AG, part of Springer Nature 2018

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 the registered company Springer International Publishing AG
part of Springer Nature

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains 38 extended and short abstracts presented at the 22nd International Conference on Research in Computational Molecular Biology (RECOMB) 2018, which was hosted by the Pierre et Marie Curie Campus of Sorbonne Université in Paris, April 21–24.

These 38 contributions were selected from a total of 193 submissions. Each submission was reviewed by three members of the Program Committee (PC), who in many cases solicited additional advice from external reviewers. Following the initial reviews, final decisions were made after an extensive discussion of the submissions among the members of the PC. Reviews and discussions were conducted through the EasyChair Conference Management System.

While RECOMB 2018 did not allow parallel submissions, authors of accepted papers were given the option to publish short abstracts in these proceedings and submit their full papers to a journal. In addition, several accepted papers were invited to submit revised manuscripts for consideration in *Cell Systems*. Papers accepted for oral presentation that were subsequently submitted to a journal are published as short abstracts and were deposited on the preprint server arxiv.org or biorxiv.org. All other papers that appear as long abstracts in the proceedings were invited for submission to the RECOMB 2018 special issue of the *Journal of Computational Biology*.

In addition to presentations of these contributed papers, RECOMB 2018 also featured six invited keynote talks given by leading scientists. The keynote speakers were Peter Campbell (Sanger Institute), Nevan Krogan (University of California, San Francisco), Ron Shamir (Tel Aviv University), François Spitz (Institut Pasteur), Sarah Teichmann (Sanger Institute), and Tandy Warnow (University of Illinois, Urbana Champaign).

RECOMB 2018 also featured highlight talks of computational biology papers that were published in journals during the previous 18 months. Of the 27 highlight submissions, seven were selected for oral presentation at RECOMB.

The success of RECOMB depends on dedicated efforts and a substantial investment of time from many colleagues. I especially thank the co-chairs of the Organizing Committee, Mireille Régnier (Ecole Polytechnique) and Yann Ponty (CNRS, Ecole Polytechnique), for hosting RECOMB; the Steering Committee and especially its chair, Bonnie Berger (MIT), for help, advice, and support throughout the process; Cenk Sahinalp (Indiana University), the Program Chair of RECOMB 2017, for answering my many questions; Ziv Bar Joseph (CMU) for chairing the highlights track; Christina Boucher (University of Florida) for chairing the poster track; Mohammed El-Kebir (UIUC) for serving as the publications chair; Alessandra Carbone (Sorbonne Université) for acting as the publicity chair; Hélène Touzet (CNRS, Université Lille I) for organizing the RECOMB Satellites; Mark Chaisson (University of Southern California), Rayan Chikhi (CNRS, University of Lille, France), Valentina Boeva (Institut Cochin and Inserm), Moritz Gerstung (EMBL-EBI), Hugues Aschard (Institut

Pasteur), Simon Gravel (McGill University), Olivier de Fresnoye, Julio Saez-Rodriguez, Pablo Meyer-Rojas, Gustavo Stolovitzky, and Elise Blaese for chairing the RECOMB Satellite Workshops on Massively Parallel Sequencing, Computational Cancer Biology, Genetics, and DREAM Challenges.

I also thank the 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.

Finally, I also thank our sponsors, including the International Society of Computational Biology (ISCB), who supported student travel fellowships, as well as the Centre National de la Recherche Scientifique (CNRS), IBM Research, Ecole Polytechnique, Réseau Francilien en Sciences Informatiques (RFSI), Région Ile de France, Frontiers, PRABI, Inria, and GDR BIM, and Sorbonne Université.

February 2018

Ben Raphael

Organization

Program Committee

Max Alekseyev	George Washington University, USA
Rolf Backofen	Albert-Ludwigs-University Freiburg, Germany
Vineet Bafna	University of California San Diego, USA
Ziv Bar-Joseph	Carnegie Mellon University, USA
Bonnie Berger	Massachusetts Institute of Technology, USA
Mathieu Blanchette	McGill University, Canada
Lenore Cowen	Tufts University, USA
Raluca Gordan	Duke University, USA
Fereydoun Hormozdiari	University of Washington, USA
John Kececioğlu	University of Arizona, USA
Carl Kingsford	Carnegie Mellon University, USA
Gunnar W. Klau	Heinrich Heine University of Düsseldorf, Germany
Jens Lagergren	SBC and CSC, KTH, Sweden
Mark Leiserson	University of Maryland, USA
Ming Li	University of Waterloo, Canada
Veli Mäkinen	University of Helsinki, Finland
Jian Ma	Carnegie Mellon University, USA
Paul Medvedev	The Pennsylvania State University, USA
Bernard Moret	Ecole Polytechnique Fédérale de Lausanne, Switzerland
Layla Oesper	Carleton College, USA
Laxmi Parida	IBM, USA
Itzik Pe'er	Columbia University, USA
Jian Peng	University of Illinois at Urbana-Champaign, USA
Yann Ponty	CNRS/LIX, Polytechnique, France
Teresa Przytycka	National Institutes of Health, USA
Ben Raphael	Princeton University, USA
Mireille Regnier	Inria, France
Knut Reinert	FU Berlin, Germany
S. Cenk Sahinalp	Indiana University Bloomington, USA
Michael Schatz	Cold Spring Harbor Laboratory, USA
Alexander Schoenhuth	Vrije Universiteit Amsterdam, 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

Jens Stoye	Bielefeld University, Germany
Fengzhu Sun	University of Southern California, USA
Wing-Kin Sung	National University of Singapore
Ewa Szczurek	University of Warsaw, Poland
Glenn Tesler	University of California San Diego, USA
Tamir Tuller	Tel Aviv University, Israel
Alfonso Valencia	Barcelona Supercomputing Center, Spain
Fabio Vandin	University of Padova, Italy
Jean-Philippe Vert	Ecole des Mines de Paris, France
Martin Vingron	Max Planck Institut für molekulare Genetik, Germany
Jerome Waldispuhl	McGill University, Canada
Tandy Warnow	University of Illinois at Urbana-Champaign, USA
Sebastian Will	University of Vienna, Austria
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
Michal Ziv-Ukelson	Ben-Gurion University of the Negev, Israel

Additional Reviewers

Adam, Alex	Bosio, Mattia	Dojer, Norbert
Aganezov, Sergey	Bruner, Ariel	Dunin-Horkawicz, Stanislaw
Akbari, Ali	Bryant, David	Durham, Timothy
Alavi, Amir	Burch, Kathryn	Eaton, Jesse
Alexeev, Nikita	Canzar, Stefan	Engler, Martin
Almodaresi, Fatemeh	Casale, Paolo	Eyras, Eduardo
Amodio, Matthew	Chaisson, Mark	Fan, Jason
Avdeyev, Pavel	Chang, Kingsley	Fang, Han
Aviran, Sharon	Chhangawala, Sagar	Feinberg, Adam
Avni, Eliran	Chor, Benny	Feizi, Soheil
Bakhtiari, Mehrdad	Ciccolella, Simone	Fernandez-Recio, Juan
Balvert, Marleen	Cirillo, Davide	Fleischauer, Markus
Bankevich, Anton	Costa, Fabrizio	Ford, Michael
Basu, Aritra	Cui, Xuefeng	Frenkel, Vladimir
Batut, Bérénice	Dadi, Temesgen	Frishberg, Amit
Beissbarth, Tim	Dahl, Andy	Gaudelet, Thomas
Benner, Philipp	Danko, David	Gawronski, Alexander
Bepler, Tristan	Davila, Jose	Gawrychowski, Pawel
Bernstein, Ryan	Deblasio, Dan	Ghareghani, Maryam
Bilenne, Olivier	Dilthey, Alexander	Giambartolomei, Claudia
Biran, Hadas	Ding, Jun	Gigante, Scott
Bockmayr, Alexander	Dirmeier, Simon	Gogolewski, Krzysztof
Boix, Carles	Doerr, Daniel	

Guo, Yuchun	Lapierre, Nathan	Neubert, Kerstin
Gupta, Chhedi	Le, Thong	Ng, Bernard
Haghshenas, Ehsan	Lee, Heewook	Nguyen, Duong
Hammelman, Jen	Lei, Haoyun	Noutahi, Emmanuel
Harel, Tom	Levi, Maya	Ntranos, Vasilis
Harrison, Robert	Li, Han	Numanagic, Ibrahim
He, Liang	Li, Shuai Cheng	Nurk, Sergey
Heinig, Matthias	Lin, Chieh	Ochoa, Alex
Heinrich, Verena	Lin, Dejun	Ochoa, Idoia
Herman-Iżycka, Julia	Liptak, Zsuzsanna	Oren, Yael
Hescott, Benjamin	Liu, Ge	Osmanbeyoglu, Hatice
Hodzic, Ermin	Liu, Jie	Ouangraoua, Aida
Hormozdiari, Farhad	Liu, Yaping	Pal, Soumitra
Hou, Lei	Lu, Yang	Park, Jisoo
Hristov, Borislav	Ludwig, Marcus	Patkar, Sushant
Hu, Alex	Luo, Yunan	Payne, Samuel
Hu, Yangyang	Ma, Cong	Persikov, Anton
Huska, Matt	Malikic, Salem	Peterlongo, Pierre
Huynh, Linh	Malod-Dognin, Noel	Pietras, Christopher
Iranmehr, Arya	Mancuso, Nicholas	PirkI, Martin
Jahn, Katharina	Mandric, Igor	Pitkänen, Esa
Jain, Siddhartha	Mangul, Serghei	Pockrandt, Christopher
Jankowski, Aleksander	Mann, Martin	Posada-Céspedes, Susana
Jia, Peilin	Marass, Francesco	Preissner, Robert
Johnson, Ruth	Markowetz, Florian	Pritykin, Yuri
Joseph, Tyler	Marschall, Tobias	Pullman, Benjamin
Ju, Chelsea	Marz, Manja	Qu, Fangfang
Jünemann, Sebastian	Marçais, Guillaume	Raguideau, Sébastien
Kamath, Govinda	Maticzka, Daniel	Rahn, René
Keasar, Chen	May, Damon	Rajaby, Ramesh
Khabirova, Eleonora	Mazrouee, Sepideh	Rajaraman, Ashok
Kichaev, Gleb	Mefford, Joel	Ramisch, Anna
Kim, Jongkyu	Meleshko, Dmitrii	Rappoport, Nadav
Kim, Yoo-Ah	Mezlini, Aziz	Rashid, Sabrina
Kim, Younhun	Miladi, Milad	Rashidi Mehrabadi, Farid
Knijnenburg, Theo	Minkin, Ilia	Rautiainen, Mikko
Knyazev, Sergey	Momo, Remi	Razaviyayn, Meisam
Kockan, Can	Moon, Kevin	Reinharz, Vladimir
Komusiewicz, Christian	Moretti, Antonio	Ren, Jie
Koptagel, Hazal	Mueller, Jonas	Reyna, Matthew
Koslicki, David	Mueller, Teresa	Rhrissorkrai, Kahn
Kuipers, Jack	Müller, Tobias	Ricketts, Camir
Käll, Lukas	Na, Seungjin	Robinson, Welles
Köster, Johannes	Narasimhan, Giri	Rotman, Jeremy
Laehnemann, David	Naseri, Ardalan	Roytman, Megan
Lafond, Manuel	Navarro, Gonzalo	Ruffalo, Matthew

Russell, Nate	Starikovskaya, Tatiana	Wittler, Roland
Saez-Rodriguez, Julio	Sternberg, Barak	Wolf, Guy
Sahin, Merve	Steuerman, Yael	Wu, Guodong
Sahlin, Kristoffer	Strzalkowski, Alexander	Xia, Li
Salmela, Leena	Sun, Chen	Yang, Shuo
Sarmeshgi, Shahbeddin	Sundermann, Linda	Yankovitz, Gal
Sason, Itay	Swenson, Krister	Ye, Tiantian
Sauerwald, Natalie	Tang, Kujin	Yeo, Grace
Schmidt, Florian	Tejero, Héctor	Yu, Yun William
Schreiber, Jacob	Thomas, Marcus	Yuan, Han
Schrinner, Sven	Tran, Hieu Ngoc	Yuan, Jie
Schulte-Sasse, Roman	Tremblay-Savard, Olivier	Zamparo, Lee
Schulz, Tizian	Tsur, Dekel	Zehavi, Meirav
Sevillya, Gur	Tsyvina, Viachaslau	Zeng, Haoyang
Shao, Mingfu	van Dijk, David	Zhang, Jesse
Sheng, Wang	van Iersel, Leo	Zhang, Ruochi
Shi, Alvin	Veksler-Lublinsky, Isana	Zhang, Sai
Shi, Huwenbo	Vyatkina, Kira	Zhang, Zhizhuo
Shomorony, Ilan	Waldherr, Steffen	Zhong, Yi
Shrestha, Raunak	Wang, Hao	Zhou, Jian
Shteyman, Alan	Wang, Sheng	Zhu, Kaiyuan
Siegel, David	Wang, Weili	Zhu, Zifan
Silverbush, Dana	Wang, Yijie	
Skums, Pavel	Wang, Ying	
Solomon, Brad	Weimann, Oren	
Spang, Rainer	Wilenzik, Roni	
Srinivasan, Krishnan	Windels, Sam	
Standage, Daniel	Winkler, Jörg	

Contents

Long Reads Enable Accurate Estimates of Complexity of Metagenomes	1
<i>Anton Bankevich and Pavel Pevzner</i>	
Chromatyping: Reconstructing Nucleosome Profiles from NOME Sequencing Data	21
<i>Shounak Chakraborty, Stefan Canzar, Tobias Marschall, and Marcel H. Schulz</i>	
GTED: Graph Traversal Edit Distance	37
<i>Ali Ebrahimipour Boroojeny, Akash Shrestha, Ali Sharifi-Zarchi, Suzanne Renick Gallagher, S. Cenk Sahinalp, and Hamidreza Chitsaz</i>	
Statistical Inference of Peroxisome Dynamics	54
<i>Cyril Galitzine, Pierre M. Jean Beltran, Ileana M. Cristea, and Olga Vitek</i>	
Loss-Function Learning for Digital Tissue Deconvolution	75
<i>Franziska Görtler, Stefan Solbrig, Tilo Wettig, Peter J. Oefner, Rainer Spang, and Michael Altenbuchinger</i>	
Inference of Population Structure from Ancient DNA	90
<i>Tyler A. Joseph and Itsik Pe'er</i>	
Using Minimum Path Cover to Boost Dynamic Programming on DAGs: Co-linear Chaining Extended	105
<i>Anna Kuosmanen, Topi Paavilainen, Travis Gagie, Rayan Chikhi, Alexandru Tomescu, and Veli Mäkinen</i>	
Modeling Dependence in Evolutionary Inference for Proteins	122
<i>Gary Larson, Jeffrey L. Thorne, and Scott Schmidler</i>	
Constrained <i>De Novo</i> Sequencing of neo-Epitope Peptides Using Tandem Mass Spectrometry	138
<i>Sujun Li, Alex DeCourcy, and Haixu Tang</i>	
Reverse de Bruijn: Utilizing Reverse Peptide Synthesis to Cover All Amino Acid <i>k</i> -mers	154
<i>Yaron Orenstein</i>	
Circular Networks from Distorted Metrics	167
<i>Sebastien Roch and Kun-Chieh Wang</i>	

A Nested 2-Level Cross-Validation Ensemble Learning Pipeline Suggests a Negative Pressure Against Crosstalk snoRNA-mRNA Interactions in <i>Saccharomyces Cerevisae</i>	177
<i>Antoine Soulé, Jean-Marc Steyaert, and Jérôme Waldispühl</i>	
Context-Specific Nested Effects Models.	194
<i>Yuriy Sverchkov, Yi-Hsuan Ho, Audrey Gasch, and Mark Craven</i>	
Algorithmic Framework for Approximate Matching Under Bounded Edits with Applications to Sequence Analysis.	211
<i>Sharma V. Thankachan, Chaitanya Aluru, Sriram P. Chockalingam, and Srinivas Aluru</i>	
Accurate Reconstruction of Microbial Strains from Metagenomic Sequencing Using Representative Reference Genomes.	225
<i>Zhemín Zhou, Nina Luhmann, Nabil-Fareed Alikhan, Christopher Quince, and Mark Achtman</i>	
Short Papers	
Targeted Genotyping of Variable Number Tandem Repeats with AdVNTR	243
<i>Mehrdad Bakhtiari, Sharona Shleizer-Burko, Melissa Gymrek, Vikas Bansal, and Vineet Bafna</i>	
Positive-Unlabeled Convolutional Neural Networks for Particle Picking in Cryo-electron Micrographs.	245
<i>Tristan Beppler, Andrew Morin, Alex J. Noble, Julia Brasch, Lawrence Shapiro, and Bonnie Berger</i>	
Designing RNA Secondary Structures Is Hard	248
<i>Édouard Bonnet, Paweł Rzqżewski, and Florian Sikora</i>	
Generalizable Visualization of Mega-Scale Single-Cell Data.	251
<i>Hyunghoon Cho, Bonnie Berger, and Jian Peng</i>	
Probabilistic Count Matrix Factorization for Single Cell Expression Data Analysis.	254
<i>G. Durif, L. Modolo, J. E. Mold, S. Lambert-Lacroix, and F. Picard</i>	
Fixed-Parameter Tractable Sampling for RNA Design with Multiple Target Structures.	256
<i>Stefan Hammer, Yann Ponty, Wei Wang, and Sebastian Will</i>	

Contribution of Structural Variation to Genome Structure: TAD Fusion Discovery and Ranking	259
<i>Linh Huynh and Fereydoun Hormozdiari</i>	
Assembly of Long Error-Prone Reads Using Repeat Graphs.	261
<i>Mikhail Kolmogorov, Jeffrey Yuan, Yu Lin, and Pavel Pevzner</i>	
A Multi-species Functional Embedding Integrating Sequence and Network Structure.	263
<i>Mark D. M. Leiserson, Jason Fan, Anthony Cannistra, Inbar Fried, Tim Lim, Thomas Schaffner, Mark Crovella, and Benjamin Hescott</i>	
Deciphering Signaling Specificity with Deep Neural Networks	266
<i>Yunan Luo, Jianzhu Ma, Yang Liu, Qing Ye, Trey Ideker, and Jian Peng</i>	
Integrative Inference of Subclonal Tumour Evolution from Single-Cell and Bulk Sequencing Data	269
<i>Salem Malikic, Katharina Jahn, Jack Kuipers, S. Cenk Sahinalp, and Niko Beerenwinkel</i>	
Mantis: A Fast, Small, and Exact Large-Scale Sequence-Search Index	271
<i>Prashant Pandey, Fatemeh Almodaresi, Michael A. Bender, Michael Ferdman, Rob Johnson, and Rob Patro</i>	
Tensor Composition Analysis Detects Cell-Type Specific Associations in Epigenetic Studies.	274
<i>Elior Rahmani, Regev Schweiger, Saharon Rosset, Sriram Sankararaman, and Eran Halperin</i>	
Assembly-Free and Alignment-Free Sample Identification Using Genome Skims	276
<i>Shahab Sarmashghi, Kristine Bohmann, M. Thomas P. Gilbert, Vineet Bafna, and Siavash Mirarab</i>	
Efficient Algorithms to Discover Alterations with Complementary Functional Association in Cancer	278
<i>Rebecca Sarto Basso, Dorit S. Hochbaum, and Fabio Vandin</i>	
Latent Variable Model for Aligning Barcoded Short-Reads Improves Downstream Analyses	280
<i>Ariya Shajii, Ibrahim Numanagić, and Bonnie Berger</i>	
ModulOmics: Integrating Multi-Omics Data to Identify Cancer Driver Modules.	283
<i>Dana Silverbush, Simona Cristea, Gali Yanovich, Tamar Geiger, Niko Beerenwinkel, and Roded Sharan</i>	

SCIΦ: Single-Cell Mutation Identification via Phylogenetic Inference	285
<i>Jochen Singer, Jack Kuipers, Katharina Jahn, and Niko Beerenwinkel</i>	
AptaBlocks: Accelerating the Design of RNA-Based Drug Delivery Systems	287
<i>Yijie Wang, Jan Hoinka, Piotr Swiderski, and Teresa M. Przytycka</i>	
A Unifying Framework for Summary Statistic Imputation	289
<i>Yue Wu, Eleazar Eskin, and Sriram Sankararaman</i>	
Characterizing Protein-DNA Binding Event Subtypes in ChIP-Exo Data	291
<i>Naomi Yamada, William K. M. Lai, Nina Farrell, B. Franklin Pugh, and Shaun Mahony</i>	
Continuous-Trait Probabilistic Model for Comparing Multi-species Functional Genomic Data.	293
<i>Yang Yang, Quanquan Gu, Takayo Sasaki, Julianna Crivello, Rachel O'Neill, David M. Gilbert, and Jian Ma</i>	
Deep Learning Reveals Many More Inter-protein Residue-Residue Contacts than Direct Coupling Analysis	295
<i>Tian-Ming Zhou, Sheng Wang, and Jinbo Xu</i>	
Author Index	297