# Lecture Notes in Bioinformatics

7262

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

Benny Chor (Ed.)

# Research in Computational Molecular Biology

16th Annual International Conference, RECOMB 2012 Barcelona, Spain, April 21-24, 2012 Proceedings



#### 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

Benny Chor Tel-Aviv University School of Computer Science 69978 Tel-Aviv, Israel E-mail: benny@cs.tau.ac.il

ISSN 0302-9743 e-ISSN 1611-3349 ISBN 978-3-642-29626-0 e-ISBN 978-3-642-29627-7 DOI 10.1007/978-3-642-29627-7 Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012935493

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

LNCS Sublibrary: SL 8 – Bioinformatics

© Springer-Verlag Berlin Heidelberg 2012

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.

The use of general descriptive names, registered names, trademarks, 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.

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**

This volume contains the papers presented at RECOMB 2012, the 16th Annual International Conference on Research in Computational Molecular Biology, held in Barcelona during April 21–24, 2012. The RECOMB conference series was started in 1997 by Sorin Istrail, Pavel Pevzner, and Michael Waterman. RECOMB 2012 was hosted by the Centre for Genomic Regulation (CRG), and took place at the Fira Palace hotel, Barcelona. This year, 31 papers were accepted for presentation out of 200 submissions. The papers presented were selected by the Program Committee (PC), with the help of many external reviewers. Each paper was reviewed by at least three members of the PC, or by external reviewers. Following the initial reviews, there was a vivid and extensive e-mail-based discussion over a period of two weeks, leading to the final decisions. Accepted papers were also invited for submission to a special issue of the Journal of Computational Biology. The highlights track, first introduced during RECOMB 2010, continued. Works published or accepted for publication in a journal during the last 15 months were eligible for submission. There were 42 submissions, and 6 were selected for oral presentation.

In addition to the contributed and highlight talks, RECOMB 2012 featured five keynote addresses on a broad theme of life science and computational biology. The keynote speakers were Richard Durbin (The Wellcome Trust Sanger Institute, UK), Eileen Furlong (Genome Biology Unit, EMBL, Germany), Thomas Gingeras (Cold Spring Harbor Laboratory, USA), Alfonso Valencia (National Cancer Research Centre, Spain), and Ada E. Yonath (Department of Structural Biology, Weizmann Institute of Science, Israel). Finally, RECOMB 2012 featured an NSF-sponsored special session on emerging areas and challenges in computational biology. This session was initiated by Mitra Basu (NSF, USA) and Mona Singh (Princeton University, USA). Five speakers were invited to give talks in this session and participate in a follow-up panel. They were Dirk Evers (UK), Christina Leslie (Memorial Sloan-Kettering Cancer Center, USA), Laxmi Parida (IBM, USA), Teresa Przytycka (NIH–NLM–NCBI, USA), and Amos Tanay (Weizmann Institute of Science, Israel).

RECOMB 2012 would not have been possible without the dedication, hard work, and substantial investment of many individuals and organizations. We thank the PC and external reviewers who helped form a high-quality conference program, the Organizing Committee, coordinated by Roderic Guigó, the Organization and Finance Chair, and Blanka Wysocka, the event coordinator, for hosting the conference and providing the administrative, logistic, and financial support. We also thank our sponsors, including the CRG (Centre for Genomic Regulation), the MICINN (Spanish Ministry of Science and Innovation), the ISCB (International Society for Computational Biology), BioMed Central, CRC Press, Cambridge University Press, the INB (Spanish Institute

of Bioinformatics), and the NSF (USA National Science Foundation). Without them the conference would not have been financially viable. We thank all the authors who contributed papers and posters, as well as the attendees of the conference for their enthusiastic participation. Finally, I would like to personally thank Roderic Guigó and Blanka Wysocka for the smooth cooperation we had regarding various organization details, and Michal Linial (Steering Committee), Martin Vingron (Chair of the Steering Committee), Bonnie Berger, and Vineet Bafna (Program Committee Chairs of RECOMB 2010 and 2011, respectively), for numerous discussions, helpful tips, and advice.

February 2012 Benny Chor

# Conference Organization

## Program Committee

Tatsuya Akutsu Kyoto University

Rolf Backofen University of Freiburg, Germany

Vineet Bafna UC San Diego, USA Nuno Bandeira UC San Diego, USA

Vikas Bansal Scripps Translational Science Institute, USA Bonnie Berger Massachusetts Institute of Technology, USA

Mathieu Blanchette McGill University, Canada
Elhanan Borenstein University of Washington, USA
Michael Brudno University of Toronto, Canada
Benny Chor (Chair) Tel Aviv University, Israel
Lenore Cowen Tufts University, USA
Eleazar Eskin UC Los Angeles, USA

Nir Friedman The Hebrew University of Jerusalem, Israel David Gifford Massachusetts Institute of Technology, USA

Bjarni Halldorsson Reykjavik University, Iceland

Eran Halperin Tel Aviv University, Israel, and ICSI, Berkeley,

USA

Alexander Hartemink Duke University, USA

Barbara Holland University of Tasmania, Australia Katharina Huber University of East Anglia, UK Daniel Huson University of Tübingen, Germany

Sorin Istrail Brown University, USA

Mehmet Koyuturk

Jens Lagergren

Thomas Lengauer

Michal Linial

Navodit Misra

Case Western Reserve University, USA

KTH Royal Institute of Technology, Sweden

Max Planck Institute for Informatics, Germany

The Hebrew University of Jerusalem, Israel

Max Planck Institute for Molecular Genetics,

Germany

Satoru Miyano University of Tokyo, Japan Luay Nakhleh Rice University, USA

William Stafford Noble University of Washington, USA

Nadia Pisanti Università di Pisa, Italy

Teresa Przytycka NIH, USA

Mark A. Ragan The University of Queensland, Australia

Ben Raphael Brown University, USA

Knut Reinert Freie Universität Berlin, Germany

Marie-France Sagot Université de Lyon, France

#### VIII Conference Organization

S. Cenk Sahinalp

David Sankoff

Russell Schwartz

Charles Semple

Simon Fraser University, Canada

University of Ottawa, Canada

Carnegie Mellon University, USA

University of Canterbury, New Zealand

Ron Shamir Tel Aviv University, Israel
Mona Singh Princeton University, USA
Sagi Snir University of Haifa, Israel
Jens Stoye Bielefeld University, Germany

Fengzhu Sun University of Southern California, USA Amos Tanay The Weizmann Institute of Science, Israel

Haixu Tang Indiana University, USA Anna Tramontano Sapienza University, Italy

Martin Vingron Max Planck Institute for Molecular Genetics,

Germany

Jerome Waldispuhl McGill University, Canada Yufeng Wu University of Connecticut, USA Jinbo Xu Toyota Tech. Inst. at Chicago, USA

Zohar Yakhini Agilent Technologies and the Technion, Israel

Michal Ziv-Ukelson Ben-Gurion University, Israel

## Steering Committee

Vineet Bafna UC 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 (Chair) Max Planck Institute for Molecular Genetics,

Germany

## Organizing Committee

Josep F Abril Universitat de Barcelona, Spain

Francisco Cámara CRG, Spain Eduardo Eyras UPF, Spain Pedro Ferreira CRG, Spain Toni Gabaldón CRG, Spain Roderic Guigó (Chair) CRG, Spain Cédric Nortedame CRG, Spain

Gabriel Valiente Univ. Politècnica de Catalunya, Spain

Blanka Wysocka CRG, Spain

### Previous RECOMB Meetings

| Dates                  | Hosting Institution            | Program Chair      | Conference Chair  |
|------------------------|--------------------------------|--------------------|-------------------|
| January 20-23, 1997    | Sandia National Lab            | Michael Waterman   | Sorin Istrail     |
| Santa Fe, NM, USA      |                                |                    |                   |
| March 22-25, 1998      | Mt. Sinai School of Medicine   | Pavel Pevzner      | Gary Benson       |
| New York, NY, USA      |                                |                    |                   |
| April 22-25, 1999      | INRIA                          | Sorin Istrail      | Mireille Regnier  |
| Lyon, France           |                                |                    |                   |
| April 8-11, 2000       | University of Tokyo            | Ron Shamir         | Satoru Miyano     |
| Tokyo, Japan           |                                |                    |                   |
| April 22-25, 2001      | Université de Montreal         | Thomas Lengauer    | David Sankoff     |
| Montreal, Canada       |                                |                    |                   |
| April 18-21, 2002      | Celera                         | Gene Myers         | Sridhar           |
| Washington, DC, USA    |                                |                    | Hannenhalli       |
| April 10-13, 2003      | German Federal Ministry for    | Webb Miller        | Martin Vingron    |
| Berlin, Germany        | Education and Research         |                    |                   |
| March 27-31, 2004      | University of California,      | Dan Gusfield       | Philip E. Bourne  |
| San Diego, USA         | San Diego                      |                    |                   |
| May 14-18, 2005        | Broad Institute of MIT and     | Satoru Miyano      | Jill P. Mesirov   |
| Boston, MA, USA        | Harvard                        |                    | and Simon Kasif   |
| April 2-5, 2006        | University of Padova           | Alberto Apostolico | Concettina Guerra |
| Venice, Italy          |                                |                    |                   |
| April 21-25, 2007      | QB3                            | Terry Speed        | Sandrine Dudoit   |
| San Francisco, CA, USA |                                |                    |                   |
| March 30-April 2, 2008 | National University of         | Martin Vingron     | Limsoon Wong      |
| Singapore, Singapore   | Singapore                      |                    |                   |
| May 18-21, 2009        | University of Arizona          | Serafim Batzoglou  | John Kececioglu   |
| Tucson, AZ, USA        |                                |                    |                   |
| August 12-15, 2010     | INESC-ID and Instituto         | Bonnie Berger      | Arlindo Oliveira  |
| Lisbon, Portugal       | Superior Técnico               |                    |                   |
| March 28-31, 2011      | Lab for Computational Biology, | Vineet Bafna       | S. Cenk Sahinalp  |
| Vancouver, Canada      | Simon Fraser University        |                    |                   |

#### **External Reviewers**

Ben-Dor, Amir

Akev, Josh Ben-Hur, Asa Alkan, Can Benyamini, Hadar Allali, Julien Bernstein, Lars Altmann, Andre Berry, Vincent Birmele, Etienne Ander, Christina Andreotti, Sandro Bliven, Spencer Arndt, Peter Böcker, Sebastian Arvestad, Lars Caglar, Mehmet Umut Aviran, Sharon Capra, Tony Ay, Ferhat Carmel, Liran Bachmat, Eitan Carr, Rogan Backoffen, Rolf Catanzaro, Daniele Bader, Garv Chen, Quan Bansal, Mukul Chen, Yang-Ho Bar, Amir Chinappi, Mauro Baumbach, Jan Chiu, Hsuan-Chao Beiko, Robert Cho, Dongyeon Ben Bassat, Ilan Cho, Sungje

Chowdhury, Salim

Cloonan, Nicole Cohen, Reuven Colbourn, Charles Comin, Matteo Costa, Fabrizio Da Fonseca, P.G.S.

Dao, Phuong

Dasgupta, Bhaskar Daskalakis, Constantinos

David, Matei
Davis, Melissa
Denise, Alain
Diallo, Abdoulaye
Dinitz, Yefim
Donmez, Nilgun
Du, Xiangjun
Dörr, Daniel

Edwards, Matthew El-Mabrouk, Nadia Emde, Anne-Katrin

Erten, Sinan Farahani, Hossein Ferreira, Rui

Fertin, Guillaume Frånberg, Mattias Furlotte, Nicholas Furman, Ora Gaspin, Christine Geiger, Dan Giegerich, Robert

Goeke, Jonathan

Golan, David Gophna, Uri Gorecki, Pawel Gottlieb, Assaf Greenblum, Sharon Grossi, Roberto Guidoni, Leonardo

Guindon, Stephane Guthals, Adrian Habin, Naomi Hach, Faraz

Hajirasouliha, Iman Halloran, John Hamilton, Nicholas Han, Buhm

Hashimoto, Tatsunori Hayashida, Morihiro

He, Dan Hogan, James Hoinka, Jan

Holtgrewe, Manuel Hormozdiari, Farhad Hormozdiari, Fereydoun

Howbert, Jeff Hu, Jialu Huang, Yang Imoto, Seiya

Jaenicke, Sebastian Joo, Jong Wha Kamburov, Atanas Kang, Eun Yong Kang, John Kaplan, Tommy Kato, Yuki

Kayano, Mitsunori Keasar, Chen Kenigsberg, Effi Kim, Sangtae Kim, Yoo-Ah Kinsella, Marcus Kirkpatrick, Bonnie Klaere, Steffen Kostem, Emrah Krause, Roland

Kundu, Kousik Laframboise, Thomas

Landan, Gilad Lange, Sita Lasserre, Julia Lee, Heewook Lee, Hyunju Lee, Su-In Lempel, Noa Levin, Alex

Levitt, Michael Levy, Roie Li, Jing

Lingiaerde, Ole-Christian

Linial, Nathan

Linz, Simone
Lipson, Doron
Liu, Kevin
Liu, Zhenqiu
Loytynoja, Ari
Lu, Bingwen
Lubling, Yaniv
Luccio, Fabrizio
Mahmoody, Ahmad
Mahony, Shaun

Mandoiu, Ion

Margelevicius, Mindaugas

Maticzka, Daniel
Mayrose, Itay
Mcilwain, Sean
Mckenney, Mark
Mcpherson, Andrew
Medvedev, Paul
Melsted, Pall

Mendelson-Cohen, Netta

Mendes, Nuno Mezlini, Aziz Milazzo, Paolo Milo, Nimrod Misra, Navodit

Moll, Mark Mysickova, Alena Nachman, Iftach

Nair, Nishanth Navarro, Gonzalo Navon, Roy Niida, Atsushi

O'Donnell, Charles Oesper, Layla Oostenbrink, Chris

Numanagic, Ibrahim

Ophir, Ron Orenstein, Yaron Oshlack, Alicia Pagli, Linda Park, Hyun Jung Parrish, Nathaniel

Peng, Jian

Peterlongo, Pierre

Pasaniuc, Bogdan

Pfeifer, Nico
Picard, Franck
Pinhas, Tamar
Preisner, Robert
Press, Maximilian
Privman, Eyal
Przulj, Natasa
Qi, Yanjun

Rahnenführer, Jörg Raimondo, Domenico

Rawitz, Droro Renda, M. Elena Roch, Sebastien Rohs, Remo Rose, Dominic Roshan, Usman Rosset, Saharon Rozov, Roye

Rubinstein, Nimrod Ruffalo, Matthew Ruths, Troy Safer, Hershel

Sankararaman, Sriram

Savel, Daniel Schlicker, Andreas Schnall-Levin, Michael Schoenhuth, Alexander Scornavacca, Celine

Serin, Akdes Shatkay, Hagit Sheridan, Paul Shibuya, Tetsuo Shimamura, Teppei Shomron, Noam Sindi, Suzanne Singer, Meromit Singh, Ajit Sjöstrand, Joel Smoly, Ilan Snedecor, June Sommer, Ingolf Souaiaia, Tade

Spivak, Marina Stacho, Juraj Steel, Mike

#### XII Table of Contents

Steinfeld, Israel

Subramanian, Ayshwarya

Sul, Jae Hoon Swanson, Lucas Szczurek, Ewa Sze, Sing-Hoi Taft, Ryan

Tamada, Yoshinori Tamura, Takeyuki Tannier, Eric

Thevenin, Annelyse Thomaseth, Karl Timmer, Jens Tofigh, Ali Tolosi, Laura Tompa, Martin

Trappe, Kathrin Tsourakakis, Charalampos

Tu, Zhidong Tuller, Tamir Tzfadia, Oren Valas, Ruben Van Iersel, Leo Vandin, Fabio

Veksler-Lublinsky, Isana Vialette, Stéphane Wagner, Allon Wan, Lin Wang, Jian

Wang, Mingxun Wang, Xiaowo Wang, Zhanyong Warnow, Tandy Will, Sebastian Wittler, Roland Wojtowicz, Damian Wolber, Gerhard

Wolf, Lior Wolfson, Haim Wu, Yin

Wuchty, Stefan Xu, Andrew Wei Yamaguchi, Rui Yang, Sichun Yang, Wen-Yun Yassour, Moran Yasuda, Tomohiro Yavas, Gokhan Yeger- Lotem, Esti Yeger-Lotem, Esti

Yosef, Nir Zaitlen, Noah Zakov, Shay Zerbino, Daniel Zhang, Jin Zhang, Shaojie Zhou, Wanding Zhu, Yun Zimmer, Ralf Zotenko, Elena Zwickl, Derrick

# **Table of Contents**

| Protein Structure by Semidefinite Facial Reduction   | 1   |
|--|-----|
| Ancestry Inference in Complex Admixtures via Variable-Length Markov Chain Linkage Models                               | 12  |
| Charge Group Partitioning in Biomolecular Simulation   | 29  |
| Increased Methylation Variation in Epigenetic Domains across Cancer Types  | 44  |
| Estimating the Accuracy of Multiple Alignments and its Use in Parameter Advising                                       | 45  |
| Quantifying the Dynamics of Coupled Networks of Switches and Oscillators   | 60  |
| lobSTR: A Short Tandem Repeat Profiler for Personal Genomes  | 62  |
| Hap-seq: An Optimal Algorithm for Haplotype Phasing with Imputation Using Sequencing Data                              | 64  |
| Ballast: A Ball-Based Algorithm for Structural Motifs Lu He, Fabio Vandin, Gopal Pandurangan, and Chris Bailey-Kellogg | 79  |
| Evolution of Genome Organization by Duplication and Loss:  An Alignment Approach                                       | 94  |
| LoopWeaver – Loop Modeling by the Weighted Scaling of Verified Proteins  | 113 |

| A Robust Method for Transcript Quantification with RNA-seq Data  Yan Huang, Yin Hu, Corbin D. Jones, James N. MacLeod,  Derek Y. Chiang, Yufeng Liu, Jan F. Prins, and Jinze Liu | 127 |
|--|-----|
| Modeling the Breakage-Fusion-Bridge Mechanism: Combinatorics and Cancer Genomics   | 148 |
| TrueSight: Self-training Algorithm for Splice Junction Detection Using RNA-seq   | 163 |
| Synthetic Sequence Design for Signal Location Search   | 165 |
| The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation  | 180 |
| Discovery of Complex Genomic Rearrangements in Cancer Using High-Throughput Sequencing   | 181 |
| Network-Based Prediction and Analysis of HIV Dependency Factors<br>T.M. Murali, Matthew D. Dyer, David Badger, Brett M. Tyler, and Michael G. Katze                              | 183 |
| Structure-Guided Deimmunization of Therapeutic Proteins  | 184 |
| Evidence for Widespread Association of Mammalian Splicing and Conserved Long-Range RNA Structures  | 199 |
| Pathset Graphs: A Novel Approach for Comprehensive Utilization of Paired Reads in Genome Assembly  | 200 |
| Finding Maximum Colorful Subtrees in Practice  | 213 |
| Recovering the Tree-Like Trend of Evolution Despite Extensive Lateral Genetic Transfer: A Probabilistic Analysis   | 224 |

| RESQUE: Network Reduction Using Semi-Markov Random Walk Scores for Efficient Querying of Biological Networks (Extended Abstract) | 239 |
|--|-----|
| Sayed Mohammad Ebrahim Sahraeian and Byung-Jun Yoon  | 239 |
| Detecting SNP-Induced Structural Changes in RNA: Application to  | 941 |
| Disease Studies  | 241 |
| A Model for Biased Fractionation after Whole Genome Duplication<br>David Sankoff, Chunfang Zheng, and Baoyong Wang               | 244 |
| Exact Pattern Matching for RNA Structure Ensembles   | 245 |
| Reconstructing Boolean Models of Signaling   | 261 |
| Alignment-Free Sequence Comparison Based on Next Generation Sequencing Reads: Extended Abstract                                  | 272 |
| Differential Oestrogen Receptor Binding is Associated with Clinical Outcome in Breast Cancer                                     | 286 |
| Simultaneous Reconstruction of Multiple Signaling Pathways via the Prize-Collecting Steiner Forest Problem                       | 287 |
| Function-Function Correlated Multi-Label Protein Function Prediction over Interaction Networks                                   | 302 |
| Predicting Protein-Protein Interactions from Multimodal Biological Data Sources via Nonnegative Matrix Tri-Factorization         | 314 |
| CNVeM: Copy Number Variation Detection Using Uncertainty of Read Mapping   | 326 |

### XVI Table of Contents

| Structure-Based Whole Genome Realignment Reveals Many Novel   |     |
|---|-----|
| Non-coding RNAs   | 341 |
| Sebastian Will, Michael Yu, and Bonnie Berger                 |     |
| Probabilistic Inference of Viral Quasispecies Subject to      |     |
| Recombination   | 342 |
| Osvaldo Zagordi, Armin Töpfer, Sandhya Prabhakaran,           |     |
| Volker Roth, Eran Halperin, and Niko Beerenwinkel             |     |
| Simultaneously Learning DNA Motif along with Its Position and |     |
| Sequence Rank Preferences through EM Algorithm                | 355 |
| ZhiZhuo Zhang, Cheng Wei Chang, Willy Hugo,                   |     |
| Edwin Cheung, and Wing-Kin Sung                               |     |
| Author Index  | 371 |
|   | 011 |