Lecture Notes in Bioinformatics 3500

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

Satoru Miyano Jill Mesirov
Simon Kasif Sorin Istrail
Pavel Pevzner Michael Waterman (Eds.)

Research in Computational Molecular Biology

9th Annual International Conference, RECOMB 2005
Cambridge, MA, USA, May 14-18, 2005
Proceedings

Springer

Volume Editors

Satoru Miyano
University of Tokyo
Human Genome Center, Institute of Medical Science, Tokyo, Japan
E-mail: miyano@ims.u-tokyo.ac.jp

Jill Mesirov
Broad Institute of MIT and Harvard, Cambridge, USA
E-mail: mesirov@broad.mit.edu

Simon Kasif
Boston University
Computational Genomics Laboratory, Boston, USA
E-mail: kasif@bu.edu

Sorin Istrail
California Institute of Technology, Pasadena, USA
E-mail: istrail@caltech.edu

Pavel Pevzner
University of California, San Diego, USA
E-mail: ppevzner@cs.ucsd.edu

Michael Waterman
University of Southern California, Los Angeles, USA
E-mail: msw@usc.edu

Library of Congress Control Number: Applied for

CR Subject Classification (1998): F.2.2, F.2, E.1, G.2, H.2.8, G.3, I.2, J.3

ISSN 0302-9743
ISBN-10 3-540-25866-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-25866-7 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 is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11415770 06/3142 5 4 3 2 1 0

Preface

This volume contains the papers presented at the 9th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2005), which was held in Cambridge, Massachusetts, on May 14–18, 2005. The RECOMB conference series was started in 1997 by Sorin Istrail, Pavel Pevzner and Michael Waterman. The list of previous meetings is shown below in the section "Previous RECOMB Meetings." RECOMB 2005 was hosted by the Broad Institute of MIT and Harvard, and Boston University's Center for Advanced Genomic Technology, and was excellently organized by the Organizing Committee Co-chairs Jill Mesirov and Simon Kasif.

This year, 217 papers were submitted, of which the Program Committee selected 39 for presentation at the meeting and inclusion in this proceedings. Each submission was refereed by at least three members of the Program Committee. After the completion of the referees' reports, an extensive Web-based discussion took place for making decisions. From RECOMB 2005, the Steering Committee decided to publish the proceedings as a volume of *Lecture Notes in Bioinformatics* (LNBI) for which the founders of RECOMB are also the editors. The prominent volume number LNBI 3500 was assigned to this proceedings. 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. The RECOMB Program Committee consisted of 42 members, as listed on a separate page. I would like to thank the RECOMB 2005 Program Committee members for their dedication and hard work. Furthermore, I should thank the persons who assisted the Program Committee members by reviewing the submissions. In appreciation of their efforts, their names are given on a special page.

RECOMB 2005 invited eight keynote speakers: David Altshuler (Broad Institute of MIT and Harvard, Massachusetts General Hospital), Wolfgang Baumeister (Max Planck Institute for Biochemistry), James Collins (Boston University), Charles DeLisi (Boston University), Jonathan King (MIT), Eric Lander (Broad Institute of MIT and Harvard), Michael Levine (University of California at Berkeley), and Susan Lindquist (Whitehead Institute). The Stanislaw Ulam Memorial Computational Biology Lecture was given by Charles DeLisi. The Distinguished Biology Lecture was given by Jonathan King.

In addition to the keynote talks and the contributed talks, an important ingredient of the program was the lively poster session which was organized with tremendous efforts by the members of the RECOMB 2005 Organizing Committee, listed on a separate page.

I would like to thank the Organizing Committee for their contributions in making the RECOMB 2005 Conference a true Boston experience! Committee members are: Gary Benson (Boston University), Bonnie Berger (MIT),

Barb Bryant (Millennium Pharmaceuticals), Charlie Cantor (Boston University), George Church (Harvard University), Peter Clote (Boston College), Lenore Cowen (Tufts University), David Gifford (MIT), Jun Liu (Harvard University), Tommy Poggio (MIT), Aviv Regev (Harvard University), Brigitta Tadmor (MIT).

Especially, I would like to thank the Organizing Committee Co-chairs Jill Mesirov (Broad Institute of MIT and Harvard) and Simon Kasif (Boston University) for their strong leadership and guidance, as well as Gus Cervini and Lynn Vitiello (Broad Institute of MIT and Harvard) for their tireless efforts over the past 18 months in bringing this conference to Boston/Cambridge. Special thanks are due to Gus Cervini for his marvelous work even on the day of a huge snowstorm.

RECOMB 2005 would like to thank the institutions and corporations who provided financial support for the conference: Broad Institute of MIT and Harvard, Boston University's Center for Advanced Genomic Technology; Affymetrix, Apple, Biogen, EMC, Glaxo-SmithKline, Hewlett-Packard, IBM, Mathworks, Millennium Pharmaceuticals, NetApp, Novartis, Pfizer, Rosetta BioSoftware, Wyeth.

I thank the members of the Steering Committee for inviting me to be the Program Chair of RECOMB 2005 and for their support and encouragement when I carried out this task.

I would also like to thank Hideo Bannai, Seiya Imoto and Mizuho Wada (Human Genome Center, Univerisity of Tokyo) for their efforts in handling the submissions and editing the final versions for publication.

In closing, I would like to thank all the people who submitted papers and posters and those who attended RECOMB 2005 with enthusiasm.

May 2005 Satoru Miyano

Program Committee Members

Tatsuya Akutsu	Kyoto University, Japan
Alberto Apostolico	Univeristy of Padova, Italy and Purdue University, USA
Vineet Bafna	University of California, San Diego, USA
Serafim Batzoglou	Stanford University, USA
Philip E. Bourne	University of California, San Diego, USA
Alvis Brazma	EBI, UK
Siren Brunak	Technical University of Denmark, Denmark
Tim Ting Chen	University of Southern California, USA
Andy Clark	Cornell University, USA
Gordon M. Crippen	University of Michigan, USA
Eleazar Eskin	University of California, San Diego, USA
Nir Friedman	Hebrew University, Israel
Sridhar Hannenhalli	University of Pennsylvania, USA
Trey Ideker	University of California, San Diego, USA
Sorin Istrail	Caltech, USA
Tao Jiang	University of California, Riverside, USA
Richard M. Karp	University of California, Berkeley, USA
Simon Kasif	Boston University, USA
Giuseppe Lancia	University of Padua, Italy
Thomas Lengauer	Max-Planck-Institut für Informatik, Germany
Michal Linial	The Hebrew University in Jerusalem, Israel
Jill Mesirov	Broad Institute of MIT and Harvard, USA
Satoru Miyano	Chair, University of Tokyo, Japan
Richard Mott	Oxford University, UK
Gene Myers	University of California, Berkeley, US
Pavel A. Pevzner	University of California, San Diego, USA
Tzachi Pilpel	Weizmann Institute of Science, Israel
John Quackenbush	TIGR, USA
Aviv Regev	Harvard University, USA
Burkhard Rost	Columbia University, USA
Walter L. Ruzzo	University of Washington, USA
David Sankoff	University of Montreal, Canada
Ron Shamir	Tel Aviv University, Israel
Steve Skiena	State University of New York at Stony Brook, USA
Temple F. Smith	Boston University, USA
Terry Speed	University of California, Berkeley, USA
Martin Vingron	Max Planck Institute for Molecular Genetics, Germany
Michael Waterman	University of Southern California, USA
Haim J. Wolfson	Tel Aviv University, Israel
Wing Wong	Harvard University, USA
Ying Xu	University of Georgia, USA
Golan Yona	Cornell University, USA

Steering Committee Members

Sorin Istrail	RECOMB General Vice-Chair, Caltech, USA
Thomas Lengauer	Max-Planck-Institut für Informatik, Germany
Michal Linial	The Hebrew University of Jerusalem, Israel
Pavel A. Pevzner	RECOMB General Chair, University California, San Diego, USA
Ron Shamir	Tel Aviv University, Israel
Terence P. Speed	University of California, Berkeley, USA
Michael Waterman	RECOMB General Chair, University of Southern California, USA

Organizing Committee Members

Jill P. Mesirov	Co-chair (Broad Institute of MIT and Harvard)
Simon Kasif	Co-chair (Boston University)
Gary Benson	Boston University
Bonnie Berger	Massachusetts Institute of Technology
Barb Bryant	Millennium Pharmaceuticals
Charlie Cantor	Boston University
Gus Cervini	Broad Institute of MIT and Harvard)
George Church	Harvard University
Peter Clote	Boston College
Lenore Cowen	Tufts University
David Gifford	Massachusetts Institute of Technology
Jun Liu	Harvard University
Tommi Poggio	Massachusetts Institute of Technology
Aviv Regev	Harvard University
Fritz Roth	Harvard University
Brigitta Tadmor	Massachusetts Institute of Technology
Lynn Vitiello	Broad Institute of MIT and Harvard

Previous RECOMB Meetings

Date/Location	Hosting Institution	Program Chair	Organizing 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 Montréal, Canada	Université de Montréal	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 & Research	Webb Miller	Martin Vingron
March 27–31, 2004 San Diego, USA	UC San Diego	Dan Gusfield	Philip E. Bourne

In addtion the Program Committee was greatly assisted by input from many people:

Tomohiro Ando
Iris Antes
Peter Arndt
George Asimenos
Vikas Bansal
Yoseph Barash
Aaron Birkland
Nikolaj Blom
Paola Bonizzoni
David Bryant
Jeremy Buhler
Tianxi Cai
Gal Chechik
Zhong Chen
Matteo Comin
Richard Coulson
Yan Cui
Vlado Danci
Sudeshna Das
Shailesh Date
Eugene Davydov
Milind Dawande
William Day
Michiel de Hoon
Arthur Delcher
Omkar Deshpande
Sharon Diskin
Chuong B. Do
Francisco Domingues
Riccardo Dondi
Pierre Donnes
Agostino Dovier
Gideon Dror
Steffen Durinck
Debojyoti Dutta
Nathan Edwards
Gal Elidan
Rani Elkon
Byron Ellis
Vince Emanuele
Anders Fausboll
Ari Frank
Jane Fridlyand
Iddo Friedberg
Carsten Friis
Tim Gardner
Olivier Gascuel
Irit Gat-Viks
Gary Glonek
Liang Goh
R. Gouveia-Oliveira
Apostol Gramada
Catherine Grasso
Clemens Groepl
Sam Gross
Steffen Grossmann
Jenny Gu
Brian Haas
Eran Halperin
Daniel Hanisch
Ke Hao
Boulos Harb
Tzvika Hartman

Anders Hinsby
Pengyu Hong
Katsuhisa Horimoto
Daniel Hsu
Haiyan Huang
Wolfgang Huber
Asa Ben Hur
Seiya Imoto
Ioannis Ioannidis
Rafael Irizarry
Ariel Jaimovich
Hanne Jarmer
Shane Jensen
Euna Jeong
Hongkai Ji
Yuting Jia
Hyun Min Kang
Ming-Yang Kao
Noam Kaplan
Tommy Kaplan
Ulas Karaoz
Uri Keich
Ryan Kelley
Wayne Kendal
Lars Kiemer
Gad Kimmel
Hironori Kitakaze
Satoshi Kobyashi
Oliver Kohlbacher
Mehmet Koyuturk
Chin-Jen Ku
Hiroyuki Kurata
Michal Lapidot
Zenia M. Larsen
Hyunju Lee
Christina Leslie
Haifeng Li
Jing Li
Lei Li
Ming Li
Xiaoman Li
Guohui Lin
Chaim Linhart
Yueyi Liu
Zhijie Liu
Ole Lund
Claus Lundegaard
Anne Mølgaard
Bin Ma
Bill Majoros
Craig Mak
Thomas Manke
Fenglou Mao
Osamu Maruyama
Hiroshi Matsuno
Kankainen Matti
Geoff McLachlan
Bud Mishra
Amir Mitchel
T.M. Murali
Iftach Nachman
Reiichiro Nakamichi
Naoki Nariai
H. Bjoern Nielsen
Mahesan Niranjan
Victor Olman
Sean O'Rourke
Paul S. Pang
Laxmi Parida
Dana Pe'er
Itsik Pe'er
Thomas N. Petersen
Ma Ping
Julia Ponomarenko
Liviu Popescu
Clint Potter
Steve Qin
Youxing Qu
Nusrat Rabbee
Jörg Rahnenführer
Gesine Reinert
Knut Reinert
Romeo Rizzi
Mark Robinson
Keith Robison
Wasinee Rungsarityotin
Gabriell Rustici
Yasubumi Sakakibara
Anastasia Samsonova
Oliver Sander
Simone Scalabrin
Thomas Schlitt
Russell Schwartz
Reut Shalgi
Roded Sharan
Itai Sharon
Kerby A. Shedden
Kirby Shedden
Noa Shefi
Tetsuo Shibuya
Ilya Shindyalov
Thomas Sicheritz-Ponten
Tobias Sing
Gordon Smyth
Lev Soinov
Ingolf Sommer
Rainer Spang
Eike Staub
Martin Steffen
Ulrike Stege
Israel Steinfeld
Nikola Stojanovic
Shih-Chieh Su
Zhengchang Su
Andreas Sundquist
Silpa Suthram
Amos Tanay
Haixu Tang
Martin Taylor
Elisabeth Tillier
Martin Tompa
Mercan Topkara
Umut Topkara
Nobuhisa Ueda
Igor Ulitsky
Dave Ussery
Gianluca Della Vedova
Stella Veretnik
Nicola Vitacolonna
Neils Volkmann
Yap Von Bing
Jeff Wall
Junwen Wang
Li-San Wang
Lusheng Wang

Table of Contents

Efficient Algorithms for Detecting Signaling Pathways in Protein Interaction Networks
Jacob Scott, Trey Ideker, Richard M. Karp, Roded Sharan 1

Towards an Integrated Protein-Protein Interaction Network
Ariel Jaimovich, Gal Elidan, Hanah Margalit, Nir Friedman 14

The Factor Graph Network Model for Biological Systems
Irit Gat-Viks, Amos Tanay, Daniela Raijman, Ron Shamir 31

Pairwise Local Alignment of Protein Interaction Networks Guided by Models of Evolution
Mehmet Koyutürk, Ananth Grama, Wojciech Szpankowski 48

Finding Novel Transcripts in High-Resolution Genome-Wide Microarray Data Using the GenRate Model
Brendan J. Frey, Quaid D. Morris, Mark Robinson, Timothy R. Hughes 66

Efficient Calculation of Interval Scores for DNA Copy Number Data Analysis
Doron Lipson, Yonatan Aumann, Amir Ben-Dor, Nathan Linial, Zohar Yakhini 83

Keynote

A Regulatory Network Controlling Drosophila Development
Dmitri Papatsenko, Mike Levine 101

Keynote

Yeast Cells as a Discovery Platform for Neurodegenerative Disease
Susan Lindquist, Ernest Fraenkel, Tiago Outeiro, Aaron Gitler, Julie Su, Anil Cashikar, Smitha Jagadish 102

RIBRA – An Error-Tolerant Algorithm for the NMR Backbone Assignment Problem
Kuen-Pin Wu, Jia-Ming Chang, Jun-Bo Chen, Chi-Fon Chang, Wen-Jin Wu, Tai-Huang Huang, Ting-Yi Sung, Wen-Lian Hsu 103

Avoiding Local Optima in Single Particle Reconstruction
Marshall Bern, Jindong Chen, Hao Chi Wong 118

A High-Throughput Approach for Associating microRNAs with Their Activity Conditions
Chaya Ben-Zaken Zilberstein, Michal Ziv-Ukelson, Ron Y. Pinter, Zohar Yakhini 133

RNA-RNA Interaction Prediction and Antisense RNA Target Search
Can Alkan, Emre Karakoç, Joseph H. Nadeau, S. Cenk Şahinalp, Kaizhong Zhang 152

Consensus Folding of Unaligned RNA Sequences Revisited
Vineet Bafna, Haixu Tang, Shaojie Zhang 172

Keynote

Discovery and Annotation of Genetic Modules
Charles DeLisi 188

Efficient q-Gram Filters for Finding All ϵ-Matches over a Given Length
Kim R. Rasmussen, Jens Stoye, Eugene W. Myers 189

A Polynomial Time Solvable Formulation of Multiple Sequence Alignment
Sing-Hoi Sze, Yue Lu, Qingwu Yang 204

A Fundamental Decomposition Theory for Phylogenetic Networks and Incompatible Characters
Dan Gusfield, Vikas Bansal 217

Reconstruction of Reticulate Networks from Gene Trees
Daniel H. Huson, Tobias Klöpper, Pete J. Lockhart, Mike A. Steel 233

A Hybrid Micro-macroevolutionary Approach to Gene Tree Reconstruction
Dannie Durand, Bjarni V. Halldórsson, Benjamin Vernot 250

Constructing a Smallest Refining Galled Phylogenetic Network
Trinh N.D. Huynh, Jesper Jansson, Nguyen Bao Nguyen, Wing-Kin Sung 265

Keynote

Mapping Molecular Landscapes Inside Cells
Wolfgang Baumeister 281

Information Theoretic Approaches to Whole Genome Phylogenies
David Burstein, Igor Ulitsky, Tamir Tuller, Benny Chor 283

Maximum Likelihood of Evolutionary Trees Is Hard
Benny Chor, Tamir Tuller 296

Graph Theoretical Insights into Evolution of Multidomain Proteins
Teresa Przytycka, George Davis, Nan Song, Dannie Durand 311

Peptide Sequence Tags for Fast Database Search in Mass-Spectrometry
Ari Frank, Stephen Tanner, Pavel Pevzner 326

A Hidden Markov Model Based Scoring Function for Mass Spectrometry Database Search
Yunhu Wan, Ting Chen 342

EigenMS: De Novo Analysis of Peptide Tandem Mass Spectra by Spectral Graph Partitioning
Marshall Bern, David Goldberg 357

Keynote

Biology as Information
Eric Lander 373

Using Multiple Alignments to Improve Gene Prediction
Samuel S. Gross, Michael R. Brent 374

Learning Interpretable SVMs for Biological Sequence Classification
S. Sonnenburg, G. Rätsch, C. Schäfer 389

Segmentation Conditional Random Fields (SCRFs): A New Approach for Protein Fold Recognition
Yan Liu, Jaime Carbonell, Peter Weigele, Vanathi Gopalakrishnan 408

Rapid Protein Side-Chain Packing via Tree Decomposition
Jinbo Xu 423

Recognition of Binding Patterns Common to a Set of Protein Structures
Maxim Shatsky, Alexandra Shulman-Peleg, Ruth Nussinov, Haim J. Wolfson 440

Predicting Protein-Peptide Binding Affinity by Learning Peptide-Peptide Distance Functions
Chen Yanover, Tomer Hertz 456

Keynote

Amino Acid Sequence Control of the Folding of the Parallel β-Helix, the Simplest β-Sheet Fold
Jonathan King, Cammie Haase-Pettingell, Ryan Simkovsky, Peter Weigele 472

A Practical Approach to Significance Assessment in Alignment with Gaps
Nicholas Chia, Ralf Bundschuh 474

Alignment of Optical Maps
Anton Valouev, Lei Li, Yu-Chi Liu, David C. Schwartz, Yi Yang, Yu Zhang, Michael S. Waterman 489

Keynote

Engineering Gene Regulatory Networks: A Reductionist Approach to Systems Biology
James J. Collins 505

Modeling the Combinatorial Functions of Multiple Transcription Factors
Chen-Hsiang Yeang, Tommi Jaakkola 506

Predicting Transcription Factor Binding Sites Using Structural Knowledge
Tommy Kaplan, Nir Friedman, Hanah Margalit 522

Motif Discovery Through Predictive Modeling of Gene Regulation
Manuel Middendorf, Anshul Kundaje, Mihir Shah, Yoav Freund, Chris H. Wiggins, Christina Leslie 538

HAPLOFREQ – Estimating Haplotype Frequencies Efficiently
Eran Halperin, Elad Hazan 553

Improved Recombination Lower Bounds for Haplotype Data
Vineet Bafna, Vikas Bansal 569

A Linear-Time Algorithm for the Perfect Phylogeny Haplotyping (PPH) Problem
Zhihong Ding, Vladimir Filkov, Dan Gusfield 585

Keynote

Human Genome Sequence Variation and the Inherited Basis of Common Disease
David Altshuler 601

Stability of Rearrangement Measures in the Comparison of Genome Sequences
David Sankoff, Matthew Mazowita 603

On Sorting by Translocations
Anne Bergeron, Julia Mixtacki, Jens Stoye 615

Author Index 631