

# Lecture Notes in Bioinformatics

4983

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

Ion Măndoiu Raj Sunderraman  
Alexander Zelikovsky (Eds.)

# Bioinformatics Research and Applications

Fourth International Symposium, ISBRA 2008  
Atlanta, GA, USA, May 6-9, 2008  
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 Editors

Ion Măndoiu

University of Connecticut

Computer Science and Engineering Department

Storrs, CT 06269, USA

E-mail: ion@enr.uconn.edu

Raj Sunderraman

Alexander Zelikovsky

Georgia State University

Computer Science Department

Atlanta, GA 30303, USA

E-mail: {raj,alexz}@cs.gsu.edu

Library of Congress Control Number: 2008925339

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

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-540-79449-2 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-79449-3 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

springer.com

© Springer-Verlag Berlin Heidelberg 2008

Printed in Germany

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

Printed on acid-free paper SPIN: 12262007 06/3180 5 4 3 2 1 0

# Preface

The 4<sup>th</sup> edition of the International Symposium on Bioinformatics Research and Applications (ISBRA 2008) was held on May 6-9, 2008 at Georgia State University in Atlanta, Georgia. The symposium provides a forum for the exchange of ideas and results among researchers, developers, and practitioners working on all aspects of bioinformatics and computational biology and their applications.

The technical program of the symposium included 35 contributed papers, selected by the Program Committee from a number of 94 full submissions received in response to the call for papers. The technical program also included six papers contributed to the First International Workshop on Optimal Data Mining in Gene Expression Analysis (ODGEA 2008), which was held in conjunction with ISBRA 2008. In addition to the contributed papers, the symposium included tutorials and poster sessions and featured invited keynote talks by six distinguished speakers. Andrew Scott Allen from Duke University and Dan Nicolae from the University of Chicago spoke on novel analysis methods for genome-wide association studies; Kenneth Buetow, director of the National Cancer Institute Center for Bioinformatics, spoke on the cancer Biomedical Informatics Grid; Andrey Gorin from Oak Ridge National Laboratory spoke on peptide identification from mass spectrometry data; Yury Khudyakov from the Center for Disease Control and Prevention spoke on integrative viral molecular epidemiology; and Kwok Tsui from Georgia Institute of Technology spoke on data mining and statistical methods for analyzing microarray experiments.

We would like to thank the Program Committee members and external reviewers for volunteering their time to review and discuss symposium papers. We would also like to thank the Chairs and the Program Committee of ODGEA 2008 for enriching the technical program of the symposium with a workshop on an important and active area of bioinformatics research. We would like to extend special thanks to the General Chairs of the symposium for their continued leadership, and to the Local Organization, Publications, Finance, Publicity, and Posters Chairs for their hard work in making ISBRA 2008 a successful event. Last but not least we would like to thank all authors for presenting their work at the symposium.

May 2008

Ion Măndoiu  
Raj Sunderraman  
Alexander Zelikovsky

# Symposium Organization

## General Chairs

Dan Gusfield, University of California, Davis, USA  
Yi Pan, Georgia State University, USA

## Program Chairs

Ion Mandoiu, University of Connecticut, USA  
Raj Sunderraman, Georgia State University, USA  
Alexander Zelikovsky, Georgia State University, USA

## Local Organization Chairs

Robert Harrison, Georgia State University, USA  
Yanqing Zhang, Georgia State University, USA

## Workshop Chairs

Luonan Chen, Osaka Sangyo University, Japan  
Lonnie Welch, Ohio University, USA

## Finance Chairs

Anu Bourgeois, Georgia State University, USA  
Akshaye Dhawan, Georgia State University, USA

## Publicity Chairs

Dumitru Brinza, University of California, San Diego, USA  
Yingshu Li, Georgia State University, USA

## Poster Chairs

Gulsah Altun, Georgia State University, USA  
Stefan Gremalschi, Georgia State University, USA

## Website

Web Design: Diana Mohan Babu, University of California, San Diego, USA  
Web Master: Gulsah Altun, Georgia State University, USA

## Program Committee

Gabriela Alexe  
Broad Institute of MIT and Harvard, USA

Yonatan Aumann  
Bar Ilan University, Israel

Zvi Bar-Joseph  
Carnegie Mellon University, USA

Danny Barash  
Ben-Gurion University, Israel

Anne Bergeron  
Université du Québec à Montréal, Canada

Tanya Berger-Wolf  
University of Illinois at Chicago, USA

Piotr Berman  
Pennsylvania State University, USA

Olivier Bodenreider  
National Library of Medicine, NIH, USA

Jeremy Buhler  
Washington University in St. Louis, USA

Paola Bonizzoni  
Univ. de Studi di Milano-Bicocca, Italy

Mark Borodovsky  
Georgia Institute of Technology, USA

Daniel Brown  
University of Waterloo, Canada

Liming Cai  
University of Georgia, USA

Doina Caragea  
Kansas State University, USA

Luonan Chen  
Osaka Sangyo University, Japan

Yixin Chen  
Washington University, USA

Peter Damaschke  
Chalmers University, Sweden

Bhaskar Dasgupta  
University of Illinois at Chicago,  
USA

Sorin Draghici  
Wayne State University, USA

Colin Dewey  
University of Wisconsin-Madison,  
USA

Liliana Florea  
George Washington University,  
USA

Andrey Gorin  
Oak Ridge National Laboratory,  
USA

Jun-Tao Guo  
University of North Carolina at  
Charlotte, USA

Jieyue He  
Southeast University, China

Vasant Honavar  
Iowa State University, USA

Hae-Jin Hu  
National Cancer Center,  
South Korea

Xiaohua (Tony) Hu  
Drexel University, USA

Ming-Yang Kao  
Northwestern University, USA

Marek Karpinski  
University of Bonn, Germany

George Karypis  
University of Minnesota, USA

John Kececioglu  
University of Arizona, USA

Ed Keedwell  
University of Exeter, UK

Yury Khudyakov  
CDC, USA

Sun Wing Kin  
NUS, Singapore

Istvan Ladunga  
University of Nebraska-Lincoln, USA

Guojun Li  
University of Georgia, USA

Jing Li  
Case Western Reserve University, USA

Yiming Li  
National Chiao Tung University, Taiwan

Guohui Lin  
University of Alberta, Canada

Shiyong Lu  
Wayne State University, USA

Jingchu Luo  
Peking University, China

Osamu Maruyama  
Kyushu University, Japan

Satoru Miyano  
University of Tokyo, Japan

Bernard Moret  
Ecole Poly. Fed. de Lausanne, Switzerland

Kayvan Najarian  
Virginia Commonwealth University, USA

Giri Narasimhan  
Florida Int. University, USA

Craig Nelson  
University of Connecticut, USA

Laxmi Parida  
IBM T.J. Watson Res. Ctr, USA

Andrey A. Perelygin  
Georgia State University, USA

Itzik Pe'er  
Columbia University, USA

Mihai Pop  
University of Maryland, USA

Alex Pothen  
Old Dominion University, USA

Teresa Przytycka  
NCBI, USA

Sven Rahmann  
Technical University of Dortmund,  
Germany

Sanguthevar Rajasekaran  
University of Connecticut, USA

Ben Raphael  
Brown University, USA

David Sankoff  
University of Ottawa, Canada

Russell Schwartz  
Carnegie Mellon University, USA

Hagit Shatkay  
Queen's University, Canada

Jens Stoye  
Universität Bielefeld, Germany

Sing-Hoi Sze  
Texas A&M University, USA

Haixu Tang  
Indiana University, USA

Hannu Toivonen  
University of Helsinki, Finland

Esko Ukkonen  
University of Helsinki, Finland

Ugo Vaccaro  
Università di Salerno, Italy

Gwenn Volkert  
Kent State University, USA

Jianxin Wang  
Central South University, China

Liangjiang Wang  
Clemson University, USA

Li-San Wang  
University of Pennsylvania, USA

Zidong Wang  
Brunel University, UK

Fang Xiang Wu  
University of Saskatchewan, Canada

Hongwei Wu  
University of Georgia, USA

Weili Wu  
University of Texas at Dallas, USA

Yufeng Wu  
University of Connecticut, USA

Dong Xu  
University of Missouri, USA

Mohammed Zaki  
Rensselaer Polytechnic Institute,  
USA

Kaizhong Zhang  
University of West Ontario,  
Canada

Xuegong Zhang  
Tsinghua University, China

Si-qing Zheng  
University of Texas at Dallas, USA

Wei-Mou Zheng  
Chinese Academy of Sciences,  
China

Wei Zhong  
University of South Carolina,  
Upstate, USA

Leming Zhu  
George Washington University,  
USA

Ying Zhu  
Georgia State University, USA

## External Reviewers

Gulsah Altun  
Amin Assareh  
Kajia Cao  
Vineet Chaoji  
Shihyen Chen  
Wenan Chen

Vanessa Clark  
Robert Day  
Riccardo Dondi  
Jason Ernst  
Lauri Eronen  
Jianwen Fang



Cen Gao  
Yonghua Han  
Mohammad Hasan  
Petteri Hintsanen  
Yang Huang  
Peter Husemann  
Katharina Jahn  
Soo-Yeon Ji  
Crystal Kahn  
Chris Kauffman  
RynagGuk Kim  
Weimin Li  
Yong Li  
Tien-ho Lin  
Yong Lu  
Chunmei Liu  
Jingping Liu  
Peng Liu  
Tobias Marschall  
Jonathan Myers  
Shilu Ni

Xia Ning  
Vinhthuy Phan  
YanJun Qi  
Anna Ritz  
Saeed Salem  
Petteri Sevon  
Quanhu Sheng  
Itamar Simon  
Rebecca Smith  
Yinglei Song  
Anuj Srivastava  
Deborah Stoffer  
Gianluca Della Vedova  
Lindi Wahl  
Nikil Wale  
Xiyin Wang  
Roland Wittler  
Yong Wu  
You Xu  
Zhiqiang Ye  
Weizhe Zhang

# Workshop Organization

## First International Workshop on Optimal Data Mining in Gene Expression Analysis (ODGEA 2008)

### General Chair

Lonnie R. Welch, Ohio University, USA

### Program Chairs

R. Krishna Murthy Karuturi, Genome Institute of Singapore, Singapore  
Qihua Tan, Odense University Hospital, Denmark  
Jahangheer Shareef Shaik, University of Memphis, USA

### Program Committee

Jing Hua Zhao, Medical Research Council in Cambridge, UK  
Xiaoli Li, Institute of Infocom Research, A-Star, Singapore  
Chengpeng Bi, University of Missouri - Kansas City, USA  
Talamuthu Anbupalam, Genome Institute of Singapore, A-Star, Singapore  
Vladimir Kuznetsov, Bioinformatics Institute, A-Star, Singapore  
Meena Sakharkar, Nanyang Technological University, Singapore  
Zhengxin Chen, University of Nebraska, Omaha, USA  
Vinhthuy Phan, University of Memphis, USA  
Parvathi Chundi, University of Nebraska, Omaha, USA  
Lih Deng, University of Memphis, USA  
Sach Mukherjee, University of Warwick, UK  
Yuri Orlov, Genome Institute of Singapore, A-Star, Singapore  
Vinsensius Vega, Genome Institute of Singapore, A-Star, Singapore

# Table of Contents

<i>Invited Keynote Talk: Set-Level Analyses for Genome-Wide Association Data (Abstract)</i> . . . . .	1
<i>Dan L. Nicolae, Omar De la Cruz, William Wen, Baoguan Ke, and Minsun Song</i>	
Hierarchical Clustering Using Constraints . . . . .	2
<i>Mariana Kant, Maurice LeBon, and David Sankoff</i>	
The Gene-Duplication Problem: Near-Linear Time Algorithms for NNI Based Local Searches . . . . .	14
<i>Mukul S. Bansal and Oliver Eulenstein</i>	
A Distance-Based Method for Detecting Horizontal Gene Transfer in Whole Genomes . . . . .	26
<i>Xintao Wei, Lenore Cowen, Carla Brodley, Arthur Brady, D. Sculley, and Donna K. Slonim</i>	
An Approach for Determining Evolutionary Distance in Network-Based Phylogenetic Analysis . . . . .	38
<i>Tingting Zhou, Keith C.C. Chan, Yi Pan, and Zhenghua Wang</i>	
Pairwise Statistical Significance Versus Database Statistical Significance for Local Alignment of Protein Sequences . . . . .	50
<i>Ankit Agrawal, Volker Brendel, and Xiaoqiu Huang</i>	
Estimating Pairwise Statistical Significance of Protein Local Alignments Using a Clustering-Classification Approach Based on Amino Acid Composition . . . . .	62
<i>Ankit Agrawal, Arka Ghosh, and Xiaoqiu Huang</i>	
Gapped Extension for Local Multiple Alignment of Interspersed DNA Repeats. . . . .	74
<i>Todd J. Treangen, Aaron E. Darling, Mark A. Ragan, and Xavier Messeguer</i>	
Improved Alignment of Protein Sequences Based on Common Parts . . . .	87
<i>David Hoksza</i>	
<i>Invited Keynote Talk: Computing P-Values for Peptide Identifications in Mass Spectrometry</i> . . . . .	100
<i>Nikita Arnold, Tema Fridman, Robert M. Day, and Andrey A. Gorin</i>	

PFPP: A Computational Framework for Phylogenetic Footprinting in Prokaryotic Genomes . . . . .	110
<i>Dongsheng Che, Guojun Li, Shane T. Jensen, Jun S. Liu, and Ying Xu</i>	
Accelerating the Neighbor-Joining Algorithm Using the Adaptive Bucket Data Structure . . . . .	122
<i>Leonid Zaslavsky and Tatiana A. Tatusova</i>	
Generalized Gene Adjacencies, Graph Bandwidth and Clusters in Yeast Evolution . . . . .	134
<i>Qian Zhu, Zaky Adam, Vicky Choi, and David Sankoff</i>	
Physicochemical Correlation between Amino Acid Sites in Short Sequences under Selective Pressure . . . . .	146
<i>David Campo, Zoya Dimitrova, and Yuri Khudyakov</i>	
HCV Quasispecies Assembly Using Network Flows . . . . .	159
<i>Kelly Westbrook, Irina Astrovskaya, David Campo, Yuri Khudyakov, Piotr Berman, and Alex Zelikovsky</i>	
A Dynamic Programming Algorithm for De Novo Peptide Sequencing with Variable Scoring . . . . .	171
<i>Matthew A. Goto and Eric J. Schwabe</i>	
<i>Invited Keynote Talk: Haplotype Sharing for Genome-Wide Case-Control Association Studies (Abstract)</i> . . . . .	183
<i>Andrew S. Allen</i>	
Incorporating Literature Knowledge in Bayesian Network for Inferring Gene Networks with Gene Expression Data . . . . .	184
<i>Eyad Almasri, Peter Larsen, Guanrao Chen, and Yang Dai</i>	
Integrative Network Component Analysis for Regulatory Network Reconstruction . . . . .	196
<i>Chen Wang, Jianhua Xuan, Li Chen, Po Zhao, Yue Wang, Robert Clarke, and Eric P. Hoffman</i>	
A Graph-Theoretic Method for Mining Overlapping Functional Modules in Protein Interaction Networks . . . . .	208
<i>Min Li, Jianxin Wang, and Jianer Chen</i>	
Identification of Transcription Factor Binding Sites in Promoter Regions by Modularity Analysis of the Motif Co-occurrence Graph . . . . .	220
<i>Alexandre P. Francisco, Arlindo L. Oliveira, and Ana T. Freitas</i>	
Mean Squared Residue Based Biclustering Algorithms . . . . .	232
<i>Stefan Gremalschi and Gulsah Altun</i>	

Sparse Decomposition of Gene Expression Data to Infer Transcriptional Modules Guided by Motif Information . . . . .	244
<i>Ting Gong, Jianhua Xuan, Li Chen, Rebecca B. Riggins, Yue Wang, Eric P. Hoffman, and Robert Clarke</i>	
A Novel Metric for Redundant Gene Elimination Based on Discriminative Contribution . . . . .	256
<i>Xue-Qiang Zeng, Guo-Zheng Li, Jack Y. Yang, and Mary Qu Yang</i>	
Network-Based Inference of Cancer Progression from Microarray Data . . . . .	268
<i>Yongjin Park, Stanley Shackney, and Russell Schwartz</i>	
<i>Invited Keynote Talk: Quiet Revolution: Connectivity in the Cancer Research Community (Abstract)</i> . . . . .	280
<i>Kenneth Buetow</i>	
Wavelet-Based 3-D Multifractal Spectrum with Applications in Breast MRI Images . . . . .	281
<i>Gordana Derado, Kichun Lee, Orietta Nicolis, F. DuBois Bowman, Mary Newell, Fabrizio F. Ruggeri, and Brani Vidakovic</i>	
Accurate Inverse Consistent Non-rigid Image Registration and Its Application on Automatic Re-contouring . . . . .	293
<i>Qingguo Zeng and Yunmei Chen</i>	
GlycoBrowser: A Tool for Contextual Visualization of Biological Data and Pathways Using Ontologies . . . . .	305
<i>Matthew Eavenson, Maciej Janik, Shravya Nimmagadda, John A. Miller, Krys J. Kochut, and William S. York</i>	
Pattern Matching in RNA Structures . . . . .	317
<i>Kejie Li, Reazur Rahman, Aditi Gupta, Prasad Siddavatam, and Michael Gribskov</i>	
The Use of a Conformational Alphabet for Fast Alignment of Protein Structures . . . . .	331
<i>Wei-Mou Zheng</i>	
On-the-Fly Rotamer Pair Energy Evaluation in Protein Design . . . . .	343
<i>Andrew Leaver-Fay, Jack Snoeyink, and Brian Kuhlman</i>	
<i>Invited Keynote Talk: Integrative Viral Molecular Epidemiology: Hepatitis C Virus Modeling</i> . . . . .	355
<i>James Lara, Zoya Dimitrova, and Yuri Khudyakov</i>	
Multiple Kernel Support Vector Regression for siRNA Efficacy Prediction . . . . .	367
<i>Shibin Qiu and Terran Lane</i>	

Hierarchical Clustering Support Vector Machines for Classifying Type-2 Diabetes Patients .....	379
<i>Wei Zhong, Rick Chow, Richard Stolz, Jieyue He, and Marsha Dowell</i>	
Computational Mutagenesis of <i>E. Coli Lac</i> Repressor: Insight into Structure-Function Relationships and Accurate Prediction of Mutant Activity .....	390
<i>Majid Masso, Kahkeshan Hijazi, Nida Parvez, and Iosif I. Vaisman</i>	
Evaluating Genetic Algorithms in Protein-Ligand Docking .....	402
<i>Rafael Ördög and Vince Grolmusz</i>	
A Hidden Markov Model Approach for Prediction of Genomic Alterations from Gene Expression Profiling .....	414
<i>Huimin Geng, Hesham H. Ali, and Wing C. Chan</i>	
Evolutionary Algorithm for Feature Subset Selection in Predicting Tumor Outcomes Using Microarray Data .....	426
<i>Qihua Tan, Mads Thomassen, Kirsten M. Jochumsen, Jing Hua Zhao, Kaare Christensen, and Torben A. Kruse</i>	
Incorporating Knowledge of Topology Improves Reconstruction of Interaction Networks from Microarray Data .....	434
<i>Peter Larsen, Eyad Almasri, Guanrao Chen, and Yang Dai</i>	
<i>Invited Keynote Talk: Data Mining and Statistical Methods for Analyzing Microarray Experiments (Abstract) .....</i>	<i>444</i>
<i>Shin-Lian Lo, Kwok-Leung Tsui, and Benjamin Barwick</i>	
Seven Variations of an Alignment Workflow - An Illustration of Agile Process Design and Management in Bio-jETI .....	445
<i>Anna-Lena Lamprecht, Tiziana Margaria, and Bernhard Steffen</i>	
Supporting Computational Systems Science: Genomic Analysis Tool Federations Using Aspects and AOP .....	457
<i>David Stotts, Keith Lee, and Ivan Rusyn</i>	
BioDQ: Data Quality Estimation and Management for Genomics Databases .....	469
<i>Alexandra Martinez, Joachim Hammer, and Sanjay Ranka</i>	
Stepped Linear Regression to Accurately Assess Statistical Significance in Batch Confounded Differential Expression Analysis .....	481
<i>Juntao Li, Jianhua Liu, and R. Krishna Murthy Karuturi</i>	
Bagging Multiple Comparisons from Microarray Data .....	492
<i>Dimitris N. Politis</i>	

Human Blood-Brain Differential Gene-Expression Correlates with  
Dipeptide Frequency of Gene Products ..... 504  
*Shandar Ahmad*

**Author Index** ..... 509