Lecture Notes in Bioinformatics

4414

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

Sepp Hochreiter Roland Wagner (Eds.)

Bioinformatics Research and Development

First International Conference, BIRD 2007 Berlin, Germany, March 12-14, 2007 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

Sepp Hochreiter University of Linz, Institute of Bioinformatics Altenbergerstr. 69, 4040 Linz, Austria E-mail: hochreit@bioinf.iku.at

Roland Wagner University of Linz, Institute of FAW Altenbergerstrasse 69, 4040 Linz, Austria E-mail: rrwagner@faw.uni-linz.ac.at

Library of Congress Control Number: 2007921873

CR Subject Classification (1998): H.2.8, I.5, J.3, I.2, H.3, F.1-2

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-540-71232-1 Springer Berlin Heidelberg New York ISBN-13 978-3-540-71232-9 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 2007 Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India Printed on acid-free paper SPIN: 12029112 06/3142 5 4 3 2 1 0

Preface

This volume contains the papers which were selected for oral presentation at the first Bioinformatics Research and Development (BIRD) conference held in Berlin, Germany during March 12-14, 2007. BIRD covers a wide range of topics related to bioinformatics like microarray data, genomics, single nucleotide polymorphism, sequence analysis, systems biology, medical applications, proteomics, information systems.

The conference was very competitive. From about 140 submissions only 36 were selected by the Program Committee for oral presentation at BIRD and for publication in these proceedings. The acceptance rate was 1/4. The decisions of the Program Committee were guided by the recommendations of several reviewers for each paper. It should be mentioned that these proceedings have companion proceedings published by the Austrian Computer Society where selected poster presentations of the BIRD conference are included.

The invited talk titled "From Flies to Human Disease" by Josef Penninger, one of the leading researcher in genetic experiments for investigating disease pathogenesis, was very inspiring and gave new insights into future bioinformatics challenges.

Many contributions utilized microarray data for medical problems or for modeling biological networks. In the field of systems biology new approaches were introduced, especially for flux distribution, for regulatory modules where the inverse problem theory is applied, and for inferring regulatory or functional networks. Several contributions are related to medical problems: therapy outcome prognosis, SNP data, genome structure, genetic causes of addiction, HIV virus, MEDLINE documents, analyzing TP53 mutation data, investigation satellite-DNA data, or testing bioactive molecules. Another main topic was sequence analysis, where contributions propose new coding schemes, new profiles, and new kernels and where Markovian approaches for genome analysis and protein structure were developed. A major research field was proteomics, which can be sub-divided into measurement- and structure-related issues. Another research field focused on databases, text retrieval, and Web services.

The first BIRD conference was a great success to which many persons contributed. The editors want to thank first of all the authors, whose work made this volume possible. Then we want to thank once again the invited speaker, Josef Penninger. We thank especially the Program Committee members, who ensured the high quality of the conference and the publications in this volume, provided their experience and expert knowledge, and invested time to select the best submission. The conference was only possible by the excellent support of the local Organizing Committee Camilla Bruns and Klaus Obermayer (Technical University of Berlin), whom we want to thank. Many thanks also to the President of the Technical University of Berlin, Kurt Kutzler, for hosting this conference. We thank the Web administrators (A. Anjomshoaa, A. Dreiling, C. Teuschel) who took care of the online submission process and registration. Thanks to the "Arbeitskreis Bioinformatik" of the Austrian Computer Society for the support of

the BIRD conference. Many thanks also to Noura Chelbat, Djork-Arne Clevert, and Martin Heusel for their support in setting up this first BIRD conference.

Most of all we wish to express our gratitude to Gabriela Wagner (DEXA Society), who managed the conference, which includes organizing the submission and the review process, setting up and coordinating the decisions of the Program Committee, being responsible for the final paper versions, planning and supervising the technical program schedule including the banquet, looking after the editorial and printing issues, and many more. Last but not least our thanks to Rudolf Ardelt, Rector of the Johannes Kepler University of Linz, Austria, who enabled this conference by establishing Bioinformatics at the JKU Linz.

January 2007

Sepp Hochreiter Roland Wagner

Organization

Program Committee

General Chair:

Roland Wagner, FAW, University of Linz, Austria

Program Chair:

Sepp Hochreiter, University of Linz, Austria

Local Organizing Committee:

Klaus Obermayer, TU Berlin, Germany Camilla Bruns, TU Berlin, Germany

Program Committee (Steering):

Amos Bairoch, Swiss Institute of Bioinformatics, Switzerland Pierre Baldi, University of California, Irvine, USA

Philip E. Bourne, UCSD, USA

Bruno Buchberger, RISC, Linz

Cornelius Frömmel, Georg-August-Universität Göttingen, Germany

David Gilbert, University of Glasgow, UK

Amarnath Gupta, SDSC of the University of California San Diego, USA

Knut Reinert, FU Berlin, Germany

Erik Sonnhammer, Stockholm University, Sweden

Hershel Safer, Weizmann Institute of Science, Israel

Poster Session Chair:

Josef Küng, FAW, Austria Jürgen Palkoska, FAW, Austria

Program Committee:

Jörg Ackermann, Univesity of Applied Sciences, Berlin, Germany

Tatsuya Akutsu, Kyoto University, Japan

Mar Alba, IMIM Barcelona, Spain

Patrick Aloy, University of Barcelona, Spain

Francois Artiguenave, INRA, France

Brian Aufderheide, Keck Graduate Institute of Applied Life Sciences, USA

Joel S. Bader, Johns Hopkins University, USA

Amos Bairoch, Swiss Institute of Bioinformatics, Switzerland

Pierre Baldi, University of California, Irvine, USA

Simon Barkow, ETH Zurich, Switzerland

Leonard Barolli, Fukuoka Institute of Technology, Japan

Christian Bartels, Novartis Pharma AG, Switzerland

Khalid Benabdeslem, LORIA & IBCP, France

Paul Bertone, EMBL-EBI, UK

Howard Bilofsky, University of Pennsylvania, USA

Christian Blaschke, Bioalma Madrid, Spain

Jacek Blazewicz, Poznan University of Technology, Poland

Andreas M. Boehm, Rudolf Virchow Center for Experimental Biomedicine, Germany

Uta Bohnebeck, TTZ Bremerhaven, Germany

Philip E. Bourne, UCSD, USA

Eric Bremer, Children's Memorial Hospital Chicago, USA

Bruno Buchberger, University of Linz, Austria

Philipp Bucher, Swiss Institute of Bioinformatics, Switzerland

Gregory Butler, Concordia University, Canada

Rita Casadio, University of Bologna, Italy

Michele Caselle, INFN, Italy

Sònia Casillas Viladerrams, Universitat Autònoma de Barcelona, Spain

Alistair Chalk, RIKEN Yokohama Institute, Genome Sciences Center, Japan

Noura Chelbat, University of Linz, Austria

Phoebe Chen, Deakin University, Australia

Francis Y.L. Chin, The University of Hong Kong, Hong Kong

Jacques Colinge, University of Applied Sciences Hagenberg, Austria

James M. Cregg, Keck Graduate Institute of Applied Life Sciences, USA

Charlotte Deane, Oxford University, UK

Coral del Val Muñoz, University of Granada, Spain

Sabine Dietmann, GSF - National Research Center for Environment and Health, Germany

Zhihong Ding, University of California, Davis, USA

Anna Divoli, SIMS UC Berkeley, USA

Arjan Durresi, Louisiana State University, USA

Jacob Engelbrecht, KVL - The Royal Veterinary and Agricultural University, Denmark

Nick Fankhauser, University of Bern, Switzerland

Domenec Farre, Center for Genomic Regulation, Spain

Pedro Fernandes, Inst.Gulbenkian de Ciência, Portugal

Milana Frenkel-Morgenstern, Weizmann Institute of Science, Israel

Rudolf Freund, European Molecular Computing Consortium, Austria

Christoph M. Friedrich, Fraunhofer SCAI, Germany

Holger Froehlich, University of Tübingen, Germany

Cornelius Frömmel, Georg-August-Universität Göttingen, Germany

Georg Fuellen, Ernst-Moritz-Arndt-University Greifswald, Germany

Toni Gabaldon, CIPF, Bioinformatics Department, Spain

Michal J. Gajda, InternationalInstitute of Molecular nd Cell Bology, Poland

David J. Galas, Battelle Memorial Institute, USA

Barbara Gawronska, School of Humanities and Informatics, Skövde, Sweden

Dietlind Gerloff, The University of Edinburgh, UK

David Gilbert, University of Glasgow, UK

Mark Girolami, University of Glasgow, UK

Aaron Golden, National University of Ireland, Ireland

Joaquin Goni, University of Navarra, Spain

Pawel Gorecki, Warsaw University, Poland

Gordon Gremme, Zentrum für Bioinformatik, Universität Hamburg, Germany

Roderic Guigo, IMIM Barcelona, Spain

Amarnath Gupta, SDSC of the University of California San Diego, USA

Hendrik Hache, Max Planck Institute for Molecular Genetics, Germany

John Hancock, MRC Mammalian Genetics Unit, UK

Artemis Hatzigeorgiou, Penn Center forBioinformatcs, USA

Amy Hauth, Université de Montréal, Canada

Marti Hearst, SIMS, UC Berkeley, USA

Jaap Heringa, Free University, Amsterdam, The Netherlands

Javier Herrero, EMBL-EBI, UK

Hanspeter Herzel, Humboldt University Berlin, Germany

Winston Hide, University of the Western Cape, South Africa

Des Higgins, University College Dublin, Ireland

Rolf Hilgenfeld, University of Lübeck, Germany

Ralf Hofestaedt, University of Bielefeld, Germany

Robert Hoffmann, Memorial Sloan-Kettering Cancer Center, USA

Hermann-Georg Holzhütter, Humboldt-Universität Berlin, Charité, Germany

Chun-Hsi Huang, University of Conneticut, USA

Tao-Wie Huang, National Taiwan University, Taiwan

Dan Jacobson, National Bioinformatics Network, South Africa

Markus Jaritz, Novartis Institutes for Biomedical Research GmbH & Co KG, Austria

Rafael C. Jimenez, University of the Western Cape, South Africa

Jaap Kaandorp, University of Amsterdam, The Netherlands

Lars Kaderali, Deutsches Krebsforschungszentrum, Germany

Alastair Kerr, University of Edinburgh, UK

Ju Han Kim, Seoul National University College of Medicine, Korea

Sung Wing Kin, National University of Singapore, Singapore

Juergen Kleffe, Charite, Germany

Erich Peter Klement, University of Linz, Austria

Ina Koch, Technical University of Applied Sciences Berlin, Germany

Martin Krallinger, National Center of Cancer Research (CNIO), Spain

Roland Krause, Max Planck Institute for Infection Biology, Germany

Judit Kumuthini, Cranfield University, UK

Josef Küng, University of Linz, Austria

Tony Kusalik, University of Saskatchewan, Canada

Michael Lappe, Max Planck Institute for Molecular Genetics, Germany

Pedro Larranaga, UPV/EHU, Spain

Gorka Lasso-Cabrera, University of Wales, Swansea, UK

Jorge Amigo Lechuga, Centro Nacional de Genotipado, Spain

Doheon Lee, KAIST, Korea

Sang Yup Lee, KAIST, Korea

Marc F. Lensink, SCMBB, Belgium

Christina Leslie, Columbia University, USA

Dmitriy Leyfer, Gene Network Sciences, Inc., USA

Guohui Lin, University of Alberta, Canada

Michal Linial, The Hebrew University of Jerusalem, Israel

Stefano Lise, University College London, UK

Mark Lohmann, University of Cologne, Germany

François Major, Université de Montréal, Canada

Paul Matthews, The European Bioinformatics Institute, Hinxton, UK

Patrick May, Konrad-Zuse-Zentrum Berlin, Germany

Shannon McWeeney, Oregon Health & Science University, USA

Engelbert Mephu Nguifo, Université d'Artois, France

Henning Mersch, Forschungszentrum Jülich, Germany

Satoru Miyano, University of Tokyo, Japan

Francisco Montero, Universidad Complutense Madrid, Spain

Yves Moreau, Katholieke Universiteit Leuven, Belgium

Robert F. Murphy, Carnegie Mellon University, USA

Sach Mukherjee, University of California, Berkeley, USA

Tim Nattkemper, University of Bielefeld, Germany

Jean-Christophe Nebel, Kingston University, UK

Steffen Neumann, Leibniz-Institut für Pflanzenbiochemie, Germany

See Kiong Ng, Institute for Infocomm Research, Singapore

Pierre Nicodeme, CNRS-LIX, France

Mario Nicodemi, Universitá di Napoli "Federico II", Italy

Grégory Nuel, CNRS, France

Klaus Obermayer, Technical University Berlin, Germany

Baldo Oliva, IMIM Barcelona, Spain

Bjorn Olsson, University of Skovde, Sweden

Philipp Pagel, GSF - National Research Center for Environment and Health, Germany

Jürgen Palkoska, University of Linz, Austria

Utz J. Pape, Max Planck Institute for Molecular Genetics, Germany

Jose M. Peña, Linköping University, Sweden

Graziano Pesole, University of Bari and Istituto Tecnologie Biomediche del C.N.R., Italy

Francisco Pinto, ITQB, University of Lisbon, Portugal

Thomas Ploetz, University of Dortmund, Germany

Adam Podhorski, CEIT, Spain

Heike Pospisil, University of Hamburg, Germany

Jagath Rajapakse, Nanyang Technological University, Singapore

Ari J Rantanen, University of Helsinki, Finland

Axel Rasche, Max Planck Institute for Molecular Genetics, Germany

Thomas Rattei, Technical University of Munich, Germany

Animesh Ray, Keck Graduate Institute, USA

Dietrich Rebholz, European Bioinformatics Institute, UK

Knut Reinert, FU Berlin, Germany

Stefan Rensing, University of Fribourg, Germany

Mark Robinson, University of Hertfordshire, UK

Peter Robinson, Humboldt-Universität, Germany

Hugues Roest Crollius, Dyogen, CNRS, France

Paolo Romano, National Cancer Research Institute (IST), Italy

Elsbeth Roskam, BIOBASE GmbH, Germany

Cristina Rubio-Escudero, University of Granada, Spain

Manuel Rueda, IRBB, Parc Cientific de Barcelona, Spain

Rob Russell, EMBL, Germany

Kim Rutherford, FlyMine (An Integrated DB for Drosophila and Anopheles Genomics), UK

Hershel Safer, Weizmann Institute of Science, Israel

Yasubumi Sakakibara, Keio University, Japan

Guido Sanguinetti, University of Sheffield, UK

David Sankoff, University of Ottawa, Canada

Clare Sansom, Birkbeck College, UK

Ferran Sanz Carreras, Biomedical Research Park of Barcelona, Spain

Reinhard Schneider, EMBL, Germany

Michael Schroeder, Technical University of Dresden, Germany

Joerg Schultz, University of Würzburg, Germany

Angel Sevilla, Universidad de Murcia, Spain

Ian Shadforth, Bioinformatics - Cranfield Health, UK

Amandeep S. Sidhu, University of Technology, Sydney, Australia

Florian Sohler, University of Munich, Germany

Erik Sonnhammer, Stockholm University, Sweden

Rainer Spang, Max Planck Institute for Molecular Genetics, Germany

William Stafford Noble, University of Washington, USA

Fran Supek, University of Zagreb, Croatia

Ashish V Tendulkar, Kanwal Rekhi School of Information Technology, India

Todt Tilman, HAN University, The Netherlands

A Min Tjoa, University of Technology Vienna, Austria

Andrew Torda, University of Hamburg, Germany

Todd Treangen, Universitat Politecnica de Catalunya, Spain

Oswaldo Trelles, University of Malaga, Spain

Michael Tress, CNIO, Spain

Elena Tsiporkova, R&D Group, Flemish Radio & Television, Belgium

Koji Tsuda, Max Planck Institutes Tübingen, Germany

Dave Ussery, The Technical University of Denmark, Denmark

P. Van der Spek, Erasmus MC, The Netherlands

Jean-Philippe Vert, Center for Computational Biology, Ecole des Mines de Paris, France

Allegra Via, University of Rome "Tor Vergata", Italy

Jordi Villà i Freixa, Research Unit on Biomedical Informatics (GRIB) of IMIM/UPF, Spain

Pablo Villoslada, University of Navarra, Spain

Susana Vinga, INESC-ID, Portugal

Dirk Walther, Max Planck Institute for Molecular Plant Physiology, Germany

Lusheng Wang, City University of Hong Kong, Hong Kong

Georg Weiller, Australian National University, Australia

Bertram Weiss, Schering AG, Germany

David Wild, Keck Graduate Institute, USA

XII Organization

Thomas Wilhelm, Fritz Lipmann Institute, Germany
Borys Wrobel, Polish Academy of Sciences, Poland
Weibing Xu, Cranfield University, UK
Paolo Zaccaria, Biocrates Life Sciences GmbH, Austria
Francisca Zanoguera, Serono Pharmaceutical Research Institute, Switzerland
Yanqing Zhang, Georgia State University, USA
Karl-Heinz Zimmermann, TU Hamburg-Harburg, Germany
Frank Gerrit Zoellner, University of Bergen, Norway
Moti Zviling, The Hebrew University of Jerusalem, Israel

Table of Contents

Session 1: Microarray and Systems Biology I (Networks)	
Bayesian Inference of Gene Regulatory Networks Using Gene Expression Time Series Data	1
Biological Network Inference Using Redundancy Analysis	16
A Novel Graph Optimisation Algorithm for the Extraction of Gene Regulatory Networks from Temporal Microarray Data Judit Kumuthini, Lionel Jouffe, and Conrad Bessant	28
Session 2: Microarray and Systems Biology II	
Analysing Periodic Phenomena by Circular PCA	38
Identification of Cold-Induced Genes in Cereal Crops and Arabidopsis Through Comparative Analysis of Multiple EST Sets	48
Mining Spatial Gene Expression Data for Association Rules	66
Session 3: Medical, SNPs, Genomics I	
Individualized Predictions of Survival Time Distributions from Gene Expression Data Using a Bayesian MCMC Approach	77
Comparing Logic Regression Based Methods for Identifying SNP	
Interactions	90
Stochastical Analysis of Finite Point Sampling of 3D Chromatin Fiber	104
in Interphase Cell Nuclei E. Gladilin, S. Goetze, J. Mateos-Langerak, R. van Driel, K. Rohr, and R. Eils	104

Session 4: Medical, SNPs, Genomics II

Structural Screening of HIV-1 Protease/Inhibitor Docking by Non-parametric Binomial Distribution Test	119
satDNA Analyzer 1.2 as a Valuable Computing Tool for Evolutionary Analysis of Satellite-DNA Families: Revisiting Y-Linked Satellite-DNA Sequences of Rumex (Polygonaceae)	131
A Soft Hierarchical Algorithm for the Clustering of Multiple Bioactive Chemical Compounds	140
Session 5: Systems Biology	
A Novel Method for Flux Distribution Computation in Metabolic Networks	154
Inverse Bifurcation Analysis of a Model for the Mammalian G_1/S Regulatory Module	168
Weighted Cohesiveness for Identification of Functional Modules and Their Interconnectivity	185
Modelling and Simulation of the Genetic Phenomena of Additivity and Dominance via Gene Networks of Parallel Aggregation Processes Elena Tsiporkova and Veselka Boeva	199
Session 6: Sequence Analysis I (Coding)	
Protein Remote Homology Detection Based on Binary Profiles	212
Biological Sequences Encoding for Supervised Classification	224
Fast Search Algorithms for Position Specific Scoring Matrices	239

Session 7: Sequence Analysis II	
A Markovian Approach for the Segmentation of Chimpanzee Genome Christelle Melodelima and Christian Gautier	251
Synthetic Protein Sequence Oversampling Method for Classification and Remote Homology Detection in Imbalanced Protein Data	263
Stem Kernels for RNA Sequence Analyses	278
Prediction of Structurally-Determined Coiled-Coil Domains with Hidden Markov Models	292
Session 8: Proteomics I	
Patch Prediction of Protein Interaction Sites: Validation of a Scoring Function for an Online Server	303
Statistical Inference on Distinct RNA Stem-Loops in Genomic Sequences	314
Interpretation of Protein Subcellular Location Patterns in 3D Images Across Cell Types and Resolutions	328
Session 9: Proteomics II (Measurements)	
Bayesian Inference for 2D Gel Electrophoresis Image Analysis Ji Won Yoon, Simon J. Godsill, ChulHun Kang, and Tae-Seong Kim	343
SimShiftDB: Chemical-Shift-Based Homology Modeling	357
Annotation of LC/ESI-MS Mass Signals	371
Session 10: Proteomics III (Structure)	
Stochastic Protein Folding Simulation in the d-Dimensional HP-Model	381
11. December 11. Dealers, and 11.A. Allicette	

XVI Table of Contents

Enhancing Protein Disorder Detection by Refined Secondary Structure Prediction	395
Joining Softassign and Dynamic Programming for the Contact Map Overlap Problem Brijnesh J. Jain and Michael Lappe	410
Session 11: Databases, Web and Text Analysis	
An Evaluation of Text Retrieval Methods for Similarity Search of Multi-dimensional NMR-Spectra	424
Ontology-Based MEDLINE Document Classification	439
Integrating Mutations Data of the TP53 Human Gene in the Bioinformatics Network Environment	453
Efficient and Scalable Indexing Techniques for Biological Sequence Data	464
Author Index	481