Lecture Notes in Bioinformatics 9874

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

Claudia Angelini
Paola M.V. Rancoita
Stefano Rovetta (Eds.)

Computational Intelligence Methods for Bioinformatics and Biostatistics

12th International Meeting, CIBB 2015
Naples, Italy, September 10–12, 2015
Revised Selected Papers

Springer

Editors
Claudia Angelini
Istituto per le Applicazioni del Calcolo
Consiglio Nazionale delle Ricerche
Naples
Italy

Paola M.V. Rancoita
Center for Statistics in the Biomedical Sciences
Vita-Salute San Raffaele University
Milano
Italy

Stefano Rovetta
DIBRIS
University of Genoa
Genova
Italy

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Bioinformatics
ISBN 978-3-319-44331-7 ISBN 978-3-319-44332-4 (eBook)
DOI 10.1007/978-3-319-44332-4

Library of Congress Control Number: 2016947208

LNCS Sublibrary: SL8 – Bioinformatics

© Springer International Publishing Switzerland 2016
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.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This volume contains a revised and selected version of the proceedings of the International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2015), which was in its 12th edition this year.

CIBB is a meeting with more than 10 years of history. Its main goal is to provide a forum open to researchers from different disciplines to present problems concerning computational techniques in bioinformatics, systems biology and medical informatics, to discuss cutting-edge methodologies and accelerate life science discoveries. Following this tradition and roots, this year's meeting brought together more than 80 researchers from the international scientific community interested in this field to discuss the advancements and the future perspectives in bioinformatics and biostatistics. Moreover, applied biologists participated in the conference in order to propose novel challenges aimed at having high impact on molecular biology and translational medicine. CIBB maintains a large Italian participation in terms of authors and conference venues but it has progressively become more international and more important in the current landscape of bioinformatics and biostatistics conferences.

This year the conference was organized in Naples (Italy) in the CNR research area during September 10–12, 2015. The topics of the conferences have kept pace with the appearance of new types of challenges in biomedical computer science, particularly with respect to a variety of molecular data and the need to integrate different sources of information. About 40 contributed papers were selected for presentation at the conference in the form of extended or short talks, either in the two main topic areas (i.e., bioinformatics and biostatistics) or in the five special sessions (The EDGE, enhanced definition of genomic entities for systems biomedicine in oncology; Multi-Omic metabolic models and statistical Bioinformatics of adaptations and biological associations; Large-Scale and HPC data analysis in bioinformatics: intelligent methods for computational, systems and synthetic biology; New knowledge from old data: power of data analysis and integration methods; Regularization methods for genomic data analysis). Each contributed paper received two reviews or more. Moreover, seven invited papers were presented in form of keynote talks. We deeply thank our invited speakers Michele Ceccarelli, Dario Greco, Dirk Husmeier, Wessel Van Wieringen, Cinzia Viroli, and Daniel Yekutieli. We are also indebted to the chairs of the very interesting and successful special sessions, which attracted very interesting contributions and attention.

All authors of contributed and invited papers were asked to submit an extended and revised paper for this volume. Afterward a further reviewing process took place, which led to the 21 papers that were selected to appear in this volume. The authors are spread over more than ten countries.

The editors would like to thank all the Program Committee members and the external reviewers of both the conference and post-conference versions of the papers for their valuable work.

A big thanks also to the sponsors, Gruppo Nazionale per il Calcolo Scientifico—GNCS INdAM, Bioinformatics Italian Society, Genomix4Life S.r.l., BMR Genomics S.R.L., M&M Biotech S.C.A.R.L., and in particular to the Istituto per le Applicazioni del Calcolo M. Picone and Institute of Genetics and Biophysics A. Buzzati Traverso that made this event possible. Finally, the editors would also like to thank all the authors for the high quality of the papers they contributed and for the interesting and stimulating discussion we had in Naples.

June 2016

Claudia Angelini
Paola Maria Vittoria Rancoita
Stefano Rovetta

Organization

CIBB2015 was jointly organized by: Istituto per le Applicazioni del Calcolo Mauro Picone, Istituto di Genetica e Biofisica Adriano Buzzati Traverso, and INNS International Neural Network Society SIG Bioinformatics, INNS International Neural Network Society SIG Bio-pattern, and the IEEE-CIS-BBCT Task Forces on Neural Networks and Evolutionary Computation.

General Chairs

Claudia Angelini	Istituto per le Applicazioni del Calcolo, Italy
Adriano Decarli	University of Milan, Italy
Erik Bongcam-Rudloff	Swedish University of Agricultural Sciences, Sweden

Biostatistics Technical Chair

Paola M.V. Rancoita	Vita-Salute San Raffaele University, Italy

Bioinformatics Technical Chair

Stefano Rovetta	University of Genoa, Italy

Special Session Organizers

Elia Biganzoli	University of Milan, Italy
Clelia Di Serio	Vita-Salute San Raffaele University, Italy
Anagha Joshi	Roslin Institute, University of Edinburgh, UK
Tom Michoel	Roslin Institute, University of Edinburgh, UK
Franck Picard	CNRS LBBE, Lyon 1, France
Andrea Bracciali	University of Stirling, UK
Mario Rosario Guarracino	High Performance Computing and Networking Institute, CNR, Italy
Ivan Merelli	Institute for Biomedical Technologies, CNR, Italy
Claudio Angione	University of Cambridge, UK
Pietro Liò	University of Cambridge, UK
Sandra Pucciarelli	University of Camerino, Italy
Barbara Simionati	BMR Genomics, Italy

Program Committee

Fentaw Abegaz	University of Groningen, The Netherlands
Federico Ambrogi	University of Milan, Italy
Sansanee Auephanwiriyakul	Chiang Mai University, Thailand
Mario Cannataro	University Magna Grecia of Catanzaro, Italy

Hailin Chen	Qiagen, Inc., USA
Davide Chicco	University of Toronto, Canada
Federica Cugnata	University of Vita-Salute San Raffaele, Italy
Antonio Eleuteri	The Royal Liverpool and Broadgreen University Hospitals, UK
Enrico Formenti	University of Nice-Sophia Antipolis, France
Arief Gusnanto	University of Leeds, UK
Raffaele Giancarlo	University of Palermo, Italy
Javier Gonzalez	University of Sheffield, UK
Yin Hu	Sage Bionetworks, USA
Pawel Labaj	BOKU Vienna, Austria
Paulo Lisboa	Liverpool John Moores University, UK
Giosué Lo Bosco	University of Palermo, Italy
Anna Marabotti	Uviversity of Salerno, Italy
Giancarlo Mauri	University Bicocca of Milan, Italy
Luciano Milanesi	ITB-CNR, Italy
Marta Milo	University of Sheffield, UK
Paola Paci	IASI-CNR, Italy
Marianna Pensky	University of Central Florida, USA
Davide Risso	University of California, Berkeley, USA
Riccardo Rizzo	ICAR-CNR, Italy
Paolo Tieri	IAC-CNR, Italy
Maurizio Urso	ICAR-CNR, Palermo, Italy
Veronica Vinciotti	Brunel University, UK
Blaz Zupan	University of Ljubljana, Slovenia

Steering Committee

Pierre Baldi	University of California, USA
Elia Biganzoli	University of Milan, Italy
Clelia Di Serio	Vita-Salute San Raffaele University, Italy
Alexandru Floares	Oncological Institute Cluj-Napoca, Romania
Jon Garibaldi	University of Nottingham, UK
Nikola Kasabov	Auckland University of Technology, New Zealand
Francesco Masulli	University of Genoa, Italy and Temple University, USA
Leif Peterson	TMHRI, USA
Roberto Tagliaferri	University of Salerno, Italy

Local Organizing Committee

Claudia Angelini	Istituto per le Applicazioni del Calcolo, Italy
Valerio Costa	Institute of Genetics and Biophysics, Italy
Italia De Feis	Istituto per le Applicazioni del Calcolo, Italy
Angelo Facchiano	Istituto di Scienze dell'Alimentazione, Italy

Secretary and Administrative Chair

Patrizia Montanaro Istituto per le Applicazioni del Calcolo, Italy

Additional Reviewers

Abbruzzo, A.
Alfieri, R
Angelini, C.
Auephanwiriyaku, S.
Brilli, M.
Cannataro, M.
Chen, H.
Chicco, D.
Cugnata, F.
Cutillo, L.
De Canditiis, D.
De Feis, I.
Facchiano, A.
Fondi, M.
Formenti, E.
Hu, Y.
Labaj, P.
Lisboa, P.
Magillo, P.
Mahmoud, H.
Marabotti, A.
Masulli, F.
Mauri, G.
Miglio, R.
Pensky, M.
Rancoita, P.M.V.
Risso, D.
Romano, P.
Rovetta, S.
Stingo, F.
Tagliaferri, R.
Tarazona, S.
Tieri, P.
Zupan, B.

Sponsoring Institutions

Istituto per le Applicazioni del Calcolo M. Picone, Italy
Institute of Genetics and Biophysics A. Buzzati Traverso, Italy
Department of Clinical Sciences and Community Health, University of Milan, Italy
Gruppo Nazionale per il Calcolo Scientifico - GNCS INdAM
Bioinformatics Italian Society
Genomix4Life S.r.l.
BMR Genomics S.R.L.
M&M Biotech S.C.A.R.L.

Contents

A Commentary on a Censored Regression Estimator 1
Antonio Eleuteri

Selecting Random Effect Components in a Sparse Hierarchical Bayesian Model for Identifying Antigenic Variability 14
Vinny Davies, Richard Reeve, William T. Harvey, and Dirk Husmeier

Comparison of Gene Expression Signature Using Rank Based Statistical Inference 28
Kumar Parijat Tripathi, Sonali Gopichand Chavan, Seetharaman Parashuraman, Marina Piccirillo, Sara Magliocca, and Mario Rosario Guarracino

Managing NGS Differential Expression Uncertainty with Fuzzy Sets 42
Arianna Consiglio, Corrado Mencar, Giorgio Grillo, and Sabino Liuni

Module Detection in Dynamic Networks by Temporal Edge Weight Clustering 54
Paola Lecca and Angela Re

A Novel Technique for Reduction of False Positives in Predicted Gene Regulatory Networks 71
Abhinandan Khan, Goutam Saha, and Rajat Kumar Pal

Unsupervised Trajectory Inference Using Graph Mining 84
Leen De Baets, Sofie Van Gassen, Tom Dhaene, and Yvan Saeys

Supervised Term Weights for Biomedical Text Classification: Improvements in Nearest Centroid Computation 98
Mounia Haddoud, Aïcha Mokhtari, Thierry Lecroq, and Saïd Abdeddaïm

Alignment Free Dissimilarities for Nucleosome Classification 114
Giosué Lo Bosco

A Deep Learning Approach to DNA Sequence Classification 129
Riccardo Rizzo, Antonino Fiannaca, Massimo La Rosa, and Alfonso Urso

Clustering Protein Structures with Hadoop 141
Giacomo Paschina, Luca Roverelli, Daniele D'Agostino, Federica Chiappori, and Ivan Merelli

Comparative Analysis of MALDI-TOF Mass Spectrometric Data in Proteomics: A Case Study 154
Eugenio Del Prete, Diego d'Esposito, Maria Fiorella Mazzeo, Rosa Anna Siciliano, and Angelo Facchiano

Binary Particle Swarm Optimization Versus Hybrid Genetic Algorithm for Inferring Well Supported Phylogenetic Trees 165
Bassam Alkindy, Bashar Al-Nuaimi, Christophe Guyeux, Jean-François Couchot, Michel Salomon, Reem Alsrraj, and Laurent Philippe

Computing Discrete Fine-Grained Representations of Protein Surfaces 180
Sebastian Daberdaku and Carlo Ferrari

The Challenges of Interpreting Phosphoproteomics Data: A Critical View Through the Bioinformatics Lens 196
Panayotis Vlastaridis, Stephen G. Oliver, Yves Van de Peer, and Grigoris D. Amoutzias

Bioinformatics Challenges and Potentialities in Studying Extreme Environments 205
Claudio Angione, Pietro Liò, Sandra Pucciarelli, Basarbatu Can, Maxwell Conway, Marina Lotti, Habib Bokhari, Alessio Mancini, Ugur Sezerman, and Andrea Telatin

Improving Genome Assemblies Using Multi-platform Sequence Data 220
Pınar Kavak, Bekir Ergüner, Duran Üstek, Bayram Yüksel, Mahmut Şamil Sağıroğlu, Tunga Güngör, and Can Alkan

Validation Pipeline for Computational Prediction of Genomics Annotations . . . 233
Davide Chicco and Marco Masseroli

Advantages and Limits in the Adoption of Reproducible Research and R-Tools for the Analysis of Omic Data 245
Francesco Russo, Dario Righelli, and Claudia Angelini

Nucha*R*t: Embedding High-Level Parallel Computing in R for Augmented Hi-C Data Analysis 259
Fabio Tordini, Ivan Merelli, Pietro Liò, Luciano Milanesi, and Marco Aldinucci

A Web Resource on Skeletal Muscle Transcriptome of Primates 273
Daniela Evangelista, Mariano Avino, Kumar Parijat Tripathi, and Mario Rosario Guarracino

Author Index 285