

Lecture Notes in Bioinformatics

5265

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

Madhu Chetty Alioune Ngom
Shandar Ahmad (Eds.)

Pattern Recognition in Bioinformatics

Third IAPR International Conference, PRIB 2008
Melbourne, Australia, October 15-17, 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

Madhu Chetty

Faculty of Information Technology

Monash University, Gippsland Campus

VIC-3842, Australia

E-mail: madhu.chetty@infotech.monash.edu.au

Alioune Ngom

School of Computer Science

University of Windsor

Windsor, Ontario, Canada

E-mail: angom@cs.uwindsor.ca

Shandar Ahmad

National Institute of Biomedical Innovation

Saito Asagi, Ibaraki-shi, Osaka, Japan

E-mail: shandar@nibio.go.jp

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.5, J.3, I.2, H.3, F.2.2

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-540-88434-3 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-88434-7 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

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: 12530489 06/3180 5 4 3 2 1 0

Preface

In the post-genomic era, a holistic understanding of biological systems and processes, in all their complexity, is critical in comprehending nature's choreography of life. As a result, bioinformatics involving its two main disciplines, namely, the life sciences and the computational sciences, is fast becoming a very promising multidisciplinary research field. With the ever-increasing application of large-scale high-throughput technologies, such as gene or protein microarrays and mass spectrometry methods, the enormous body of information is growing rapidly. Bioinformaticians are posed with a large number of difficult problems to solve, arising not only due to the complexities in acquiring the molecular information but also due to the size and nature of the generated data sets and/or the limitations of the algorithms required for analyzing these data. Although the field of bioinformatics is still in its embryonic stage, the recent advancements in computational and information-theoretic techniques are enabling us to conduct various *in silico* testing and screening of many lab-based experiments before these are actually performed *in vitro* or *in vivo*. These *in silico* investigations are providing new insights for interpretation and establishing a new direction for a deeper understanding. Among the various advanced computational methods currently being applied to such studies, the *pattern recognition* techniques are mostly found to be at the core of the whole discovery process for apprehending the underlying biological knowledge. Thus, we can safely surmise that the ongoing bioinformatics *revolution* may, in future, inevitably play a major role in many aspects of medical practice and/or the discipline of life sciences.

The aim of the Pattern Recognition in Bioinformatics (PRIB) conference is to provide an opportunity for academia, researchers, scientists and industry professionals to present their latest research in pattern recognition and computational intelligence-based techniques applied to problems in bioinformatics and computational biology. It also provides them with an excellent forum to interact with each other and share experiences. The conference is organized jointly by Monash University, Australia, and the IAPR (International Association for Pattern Recognition) Bioinformatics Technical Committee (TC-20).

This volume presents the proceedings of the Third IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB 2008), held in Melbourne, October 15–17, 2008. It includes 39 technical contributions that were selected by the International Program Committee from 121 submissions. Each of these rigorously reviewed papers was presented orally at PRIB 2008. The proceedings consists of six parts:

- Part 1** Protein: Structure, Function, and Interaction
- Part 2** Learning, Classification, and Clustering
- Part 3** Bio-Molecular Networks and Pathways Analysis
- Part 4** Microarray and Gene Expression Analysis

Part 5 Data Mining and Knowledge Discovery**Part 6** Applications of High-Performance Computing

Part 1 of the proceedings contains eight chapters on *Protein: Structure, Function, and Interaction*. Gromiha et al. propose a method based on a decision tree for discriminating the stabilizing and destabilizing mutants and predicting protein stability changes upon single-point mutations. The chapter also includes methods developed for discriminating thermophilic proteins from mesophilic ones. In the next chapter by Kumar et al., a new approach to locating the occurrences of user-defined motifs in a specified order in large proteins and in nucleotide sequence databases is proposed. Bauer et al. explore the nature of post-translation modifications (SUMOylation) using non-local sequence and structural properties, including secondary structure, solvent accessibility and evolutionary profiles. Hoque et al. combine a genetic algorithm with depth-first search for the solution of the protein structure prediction problem. Lonquety et al. present a stability-based analysis of the protein-folding nucleus to increase the recall and precision of two well-known protein mutant stability prediction methods. Kato et al. report a dynamic programming algorithm for an up-down class of antiparallel protein β -sheet, which can also be extended for more general classes of β -sheets. In Li et al., the concept of multi-scale glide zoom window feature extraction is used for predicting protein homo-oligomers. Koizumi et al. propose a method of searching for and comparing concave structures in protein-binding sites.

Part 2 of the proceedings contains seven chapters on *Learning, Classification, and Clustering*. Yang et al. propose a hybrid system for analyzing high-dimensional mass spectrometry data. Medvés et al. propose a modified Markov clustering algorithm for an efficient clustering of large protein sequence databases, based on a previously evaluated sequence similarity criteria. Stiglic et al. present a classification ensemble of decision trees called Rotation Forest and evaluate its classification performance on small subsets of ranked genes for 14 genomic and proteomic classification problems. Al Seesi et al. describe a new inference algorithm, based on tree adjoining grammars, for RNA pseudoknot structure identification. Mundra et al. propose to use support vector points for computation of t-scores for gene ranking. Anand et al. consider two sets of features based on DNA sequences and their physicochemical properties and applied a one-versus-all support vector machine with class-wise optimized features to identify transcription factor family-specific features in DNA sequences. Ji et al. present a novel protein classifier, the gapped Markov Chain with Support Vector Machine, that models the structure of a protein sequence by measuring the transition probabilities between pairs of amino acids.

Part 3 of the proceedings contains six chapters on *Bio-Molecular Networks and Pathways Analysis*. Zhao et al. propose a novel discriminative method for predicting domain–domain interactions in protein pairs by making use of interacting and non-interacting protein pairs, which improves the prediction reliability. Jancura et al. develop an algorithm for dividing protein–protein interaction networks that combines the graph theoretical property of articulation with a

biological property of orthology. Ram et al. demonstrate the application of a Markov blanket learning algorithm to gene regulatory networks, enhanced further by application of a proposed constraint logic minimization technique. Chaturvedi et al. model time delayed interactions in gene regulatory networks using a skip-chain model which finds missing edges between non-consecutive time points based on protein-protein interaction networks. Zhou et al. propose a new pattern recognition technique to help represent metabolic networks as weighted vectors. Ram et al. present an approach for synthetically generating gene regulatory networks using causal relationships.

Part 4 of the proceedings contains seven chapters on *Microarray and Gene Expression Analysis*. Huerta et al. introduce a new wrapper approach to the difficult task of microarray data gene selection, where a genetic algorithm is combined with Fisher's linear discriminant analysis. Wang et al. present a new heuristic approach for finding near-minimal non-unique probe sets for oligonucleotide microarray experiments. Using a well-known yeast cell cycle data set, Pittelkow et al. compare a method being used for finding genes following a periodic time series pattern with a method for finding genes having a different phase pattern during the cell cycle. Bedo explores the design problem of selecting a small subset of clones from a large pool for creation of a microarray plate. Nagarajan et al. use two distinct approaches, namely, the classical order zero-crossing count and the Lempel-Ziv complexity, in identifying non-random patterns from temporal gene expression profiles. Ooi et al. propose to determine the theoretical basis for the concept of differential prioritization through mathematical analyses of the characteristics of predictor sets found using different values of the degree of differential prioritization from realistic but toy datasets. Luo et al. propose a weighted top scoring pair method for gene selection and classification.

Part 5 of the proceedings contains seven chapters on *Data Mining and Knowledge Discovery*. Kasturi et al. present an algorithm to identify statistically significant and conserved discriminative motifs that distinguish between gene expression clusters. McGarry et al. describe methods to develop a reliable, automated method of detecting abnormal metabolite profiles from urinary organic acids, which can be used as GC-MS biomarkers. Girão et al. have applied multi-relational data mining methods with hidden Markov models and a Viterbi algorithm to mine tetratricopeptide repeat, pentatricopeptide and half-tetratricopeptide repeat in genomes of pathogenic protozoa *Leishmania*. Sehgal et al. present an enhanced heuristic non-parametric collateral missing value imputation algorithm which uses collateral missing value estimation as its core estimator and a heuristic non-parametric strategy to compute the optimal number of estimator genes to exploit optimally both local and global correlations. Han et al. develop a non-negative principal component analysis algorithm and propose a non-negative principal component analysis-based support vector machine algorithm with sparse coding in the cancer molecular pattern analysis of proteomics data. Macintyre et al. have developed a novel clustering algorithm which incorporates functional gene information from gene ontology into the clustering

process, resulting in more biologically meaningful clusters. Meydan et al. try evolutionary classification methods for selecting the important classifier genes in hexachlorobenzene toxicity using microarray data.

Part 6 of the proceedings contains four chapters on *Applications of High-Performance Computing*. Stamatakis et al. address parallelism issues via a thorough performance study by example of a widely used bioinformatics application for large-scale phylogenetic inference under the maximum likelihood criterion. Schröder et al. present an enhanced version of an existing DNA motif search algorithm tailored to fit on a massively parallel machine. Chen et al. present a novel approach to accelerate motif discovery based on commodity graphics hardware. Zhang et al. demonstrate how the PlayStation 3, powered by the Cell Broadband Engine, can be used as an efficient computational platform to accelerate the popular BLASTP algorithm.

Many have contributed directly or indirectly toward the organization and success of PRIB 2008 conference. We would like to thank all individuals and institutions, especially the authors for submitting the papers and the sponsors for generously providing financial support for the conference. We are very grateful to IAPR for the sponsorship and the IAPR Technical Committee (TC-20) on Pattern Recognition for Bioinformatics for their support and advice. Our gratitude goes to the Faculty of Information Technology, Monash University, Australia, and also to the Gippsland campus, Monash University, Australia, for supporting the conference in many ways.

We would like to express our gratitude to all PRIB 2008 International Program Committee members for their objective and thorough reviews of the submitted papers. We fully appreciate the PRIB 2008 Organizing Committee for their time, efforts, and excellent work. We would also like to thank Jagath Rajapakse, Program Co-chair, and Raj Acharya, General Co-chair, for their continuous support and guidance. We sincerely thank Shyh Wei Teng, Local Organization Chair, for his relentless work in managing various operational issues and finance matters related to the conference organization. We thank Dieter Bulich for organizing the conference sponsorship and the Publication Co-chair, Sy Loi Ho, for his hard work in getting the proceedings ready on time. We are also grateful to Tina Bradshaw, PRIB 2008 secretary, for coordinating all the logistics of the workshop and to Margot Schuhmacher for meticulously maintaining the PRIB 2008 conference website.

Last but not least, we wish to convey our sincere thanks to Springer for providing excellent professional support in preparing this volume.

October 2008

Madhu Chetty
Alioune Ngom
Shandar Ahmad

PRIB 2008 Organization

General Chair

Madhu Chetty Monash University, Australia

General Co-chair

Raj Acharya Pennsylvania State University, USA

Program Co-chairs

Alioune Ngom University of Windsor, Canada
Jagath C. Rajapakse Nanyang Technological University, Singapore

Technical Co-chairs

Elena Marchiori Vrije Universiteit, The Netherlands
Bertil Schmidt NICTA, Australia

Short Paper Co-chairs

Shandar Ahmad National Institute of Biomedical Innovation,
Japan
Ponraj Prabhakaran Duke University, USA
Michael Grominha Computational Biology Research Centre,
Japan

Special Session Chair

Alioune Ngom University of Windsor, Canada

Local Organization Chair

Shyh Wei Teng Monash University, Australia

Publicity Co-chairs

Shandar Ahmad National Institute of Biomedical Innovation,
Japan
Phoebe Chen Deakin University, Australia
Mariofanna Milanova University of Arkansas at Little Rock, USA
Md. Tamjidul Hoque Griffith University, Australia

Publication Co-chairs

Sy Loi Ho Nanyang Technological University, Singapore
Girija Chetty University of Canberra, Australia

Sponsorship Chair

Dieter Bulach CSIRO, Australia

Finance Manager

Narelle Wilkins Monash University, Australia

Secretariat

Tina Bradshaw Monash University, Australia

Webmaster

Margot Schuhmacher Monash University, Australia

Technical Manager

Glen Pringle Monash University, Australia

Program Committee

Raj Acharya	Pennsylvania State University, USA
Shandar Ahmad	National Institute of Biomedical Innovation, Japan
Tatsuya Akutsu	Kyoto University, Japan
James Bailey	University of Melbourne, Australia
Tim Bailey	University of Queensland, Australia
Sanghamitra Bandyopadhyay	Indian Statistical Institute, India
Wolfgang Banzhaf	Memorial University of Newfoundland, Canada
Justin Bedo	The Australian National University, Australia
Nitin Bhardwaj	University of Illinois at Chicago, USA
Mikael Boden	University of Queensland, Australia
Sebastian Boecker	Friedrich-Schiller-Universitaet Jena, Germany
Guillaume Bourque	Genome Institute of Singapore, Singapore
Roelof Brouwer	University of Stellenbosch, South Africa
Vladimir Brusic	Dana-Farber Cancer Institute, USA
Shakuntala Bulusu	DNA Research Centre, India
Conrad Burden	Australian National University, Australia
Yu-Dong Cai	Shanghai Institutes for Biological Science, China
Frederic Cazals	INRIA Sophia Antipoh's, France
Francis Chin	The University of Hong Kong, China
Peter Clote	Boston College, USA
Michael Cohen	Griffith University, Australia
Yun Cui	Nanyang Technological University, Singapore
Antoine Danchin	Institut Pasteur, France

Suash Deb	National Institute of Science and Technology, India
Omid Dehzangi	Nanyang Technological University, Singapore
Yong-Sheng Ding	Donghua University, China
M. Michael Gromiha	Computational Biology Research Center, Japan
Nadia El-Mabrouk	University of Montreal, Canada
Alexandru Floares	Oncological Institute Cluj-Napoca, Romania
Gary B. Fogel	Natural Selection Inc., USA
Mehdi Ghayoumi	Azad University, Iran
Prashant Goley	University Nanded Maharashtra, India
Iqbal Gondal	Monash University, Australia
Robin Gras	University of Windsor, Canada
Saman K. Halgamuge	University of Melbourne, Australia
Xiaoxu Han	Eastern Michigan University, USA
Jin-Kao Hao	University of Angers, France
Jaap Heringa	Vrije Universiteit, Belgium
Sy Loi Ho	Nanyang Technological University, Singapore
Md Tamjidul Hoque	Griffith University, Australia
Seiya Imoto	University of Tokyo, Japan
Shunsuke Inenaga	Kyushu University, Japan
R. Krishna Murthy Karuturi	Genome Institute of Singapore, Singapore
Nawaz Khan	Middlesex University, UK
Amit Kumar	DNA Research Centre, India
Lukasz Kurgan	University of Alberta, Canada
Adam Kowalczyk	NICTA, Australia
Vlad Kuznetsov	Genome Institute of Singapore, Singapore
Zoe Lacroix	Arizona State University, USA
Hon Wai Leong	National University of Singapore, Singapore
Jinyan Li	Nanyang Technological University
Vladimir Likic	University of Melbourne, Australia
Feng Lin	Nanyang Technological University, Singapore
Tsun-Chen Lin	Dahan Institute of Technology, Taiwan
Frederique Lisacek	Swiss Institute of Bioinformatics, Switzerland
Feng Liu	Vrije Universiteit, Belgium
Lifang Liu	Xi'an University, China
Weiguo Liu	National Technological University, Singapore
Geoff Macintyre	Victorian Research Lab, Australia
Niranjana Mahesan	University of Sheffield, UK
Veli Makinen	University of Helsinki, Finland
Elena Marchiori	Vrije Universiteit, The Netherlands
Hiroshi Matsuno	Yamaguchi University, Japan
Geoff McLachlan	University of Queensland, Australia
Aleksandar Milosavljevic	Baylor College of Medicine, USA
Martin Middendorf	University of Leipzig, Germany
Mariofanna Milanova	University of Arkansas at Little Rock, USA

Satoru Miyano	University of Tokyo, Japan
Krishna Mohan	NIBIO, Japan
Sukanta Mondal	NIBIO, Japan
Bernard Moret	Swiss Federal Institute of Technology, Switzerland
Narayana Nagesh	Centre for Cellular and Molecular Biology, India
See-Kiong Ng	Institute for Infocomm Research, Singapore
Alioune Ngom	University of Windsor, Ontario, Canada
Bing Niu	Shanghai University, China
Diana Oliveira	Universidade Estadual do Ceara, Brazil
Nikhil R. Pal	Indian Statistical Institute, India
Kuldip Paliwal	Griffith University, Australia
Mihail Popescu	University of Missouri-Columbia, USA
Jagath Rajapakse	Nanyang Technological University, Singapore
Ramesh Ram	Monash University, Australia
Alice Richardson	University of Canberra, Australia
Luis Rueda	Universidad de Concepción, Chile
Marie-France Sagot	INRIA, France
Meena Kishore Sakharkar	Nanyang Technological University, Singapore
Alexander Schliep	Max Planck Institute for Molecular Genetics, Germany
Bertil Schmidt	NICTA, Australia
Christian Schoenbach	Nanyang Technological University, Singapore
Jan Schröder	Christian Albrechts University of Kiel, Germany
Heiko Schroeder	RMIT University, Australia
Muhammad Shoaib Sehgal	University of Queensland, Australia
Mona Singh	Princeton, USA
Ingolf Sommer	Max Planck Institute for Informatics, Germany
Alexandros Stamatakis	Ludwig Maximilians University of Munich, Germany
Gregor Stiglic	University of Maribor, Slovenia
Wing Kin Sung	National University of Singapore, Singapore
Roberto Tagliaferri	Università di Salerno, Italy
Y-H. Taguchi	Chuo University, Japan
Shyh Wei Teng	Monash University, Australia
Qi-Chuan Tian	Taiyuan University of Science and Technology, China
Lokesh Tripathi	NIBIO, Japan
Lusheng Wang	City University of Hong Kong, China
Alan Wee-Chung Liew	Griffith University, Australia
Yanqing Zhang	Georgia State University, USA
Xingming Zhao	Shanghai University, China
Justin Zobel	NICTA, Australia

Table of Contents

Part I: Protein: Structure, Function, and Interaction

Sequence Based Prediction of Protein Mutant Stability and Discrimination of Thermophilic Proteins	1
<i>M. Michael Gromiha, Liang-Tsung Huang, and Lien-Fu Lai</i>	
A Method to Find Sequentially Separated Motifs in Biological Sequences (SSMBS)	13
<i>Chetan Kumar, Nishith Kumar, Sarani Rangarajan, Narayanaswamy Balakrishnan, and Kanagaraaj Sekar</i>	
Predicting SUMOylation Sites	28
<i>Denis C. Bauer, Fabian A. Buske, and Mikael Bodén</i>	
DFS Based Partial Pathways in GA for Protein Structure Prediction . . .	41
<i>Md Tamjidul Hoque, Madhu Chetty, Andrew Lewis, and Abdul Sattar</i>	
Evaluation of the Stability of Folding Nucleus upon Mutation	54
<i>Mathieu Lonquety, Zoé Lacroix, and Jacques Chomilier</i>	
Prediction of Protein Beta-Sheets: Dynamic Programming Versus Grammatical Approach	66
<i>Yuki Kato, Tatsuya Akutsu, and Hiroyuki Seki</i>	
Using Multi-scale Glide Zoom Window Feature Extraction Approach to Predict Protein Homo-oligomer Types	78
<i>QiPeng Li, Shao Wu Zhang, and Quan Pan</i>	
Extraction of Binding Sites in Proteins by Searching for Similar Local Molecular Surfaces	87
<i>Satoshi Koizumi, Keisuke Imada, Tomonobu Ozaki, and Takenao Ohkawa</i>	

Part II: Learning, Classification, and Clustering

A Clustering Based Hybrid System for Mass Spectrometry Data Analysis	98
<i>Pengyi Yang and Zili Zhang</i>	
A Modified Markov Clustering Approach for Protein Sequence Clustering	110
<i>Lehel Medvés, László Szilágyi, and Sándor M. Szilágyi</i>	

Feature Selection and Classification for Small Gene Sets	121
<i>Gregor Stiglic, Juan J. Rodriguez, and Peter Kokol</i>	
Pseudoknot Identification through Learning TAG _{RNA}	132
<i>Sahar Al Seesi, Sanguthevar Rajasekaran, and Reda Ammar</i>	
Support Vector Based T-Score for Gene Ranking	144
<i>Piyushkumar A. Mundra and Jagath C. Rajapakse</i>	
Prediction of Transcription Factor Families Using DNA Sequence Features	154
<i>Ashish Anand, Gary B. Fogel, Ganesan Pugalenthil, and P.N. Suganthan</i>	
g-MARS: Protein Classification Using Gapped Markov Chains and Support Vector Machines	165
<i>Xiaonan Ji, James Bailey, and Kotagiri Ramamohanarao</i>	

Part III: Bio-Molecular Networks and Pathways Analysis

Domain-Domain Interaction Identification with a Feature Selection Approach	178
<i>Xing-Ming Zhao and Luonan Chen</i>	
Dividing Protein Interaction Networks by Growing Orthologous Articulations	187
<i>Pavol Jancura, Jaap Heringa, and Elena Marchiori</i>	
Constraint Minimization for Efficient Modeling of Gene Regulatory Network	201
<i>Ramesh Ram, Madhu Chetty, and Dieter Bulach</i>	
Fusion of Gene Regulatory and Protein Interaction Networks Using Skip-Chain Models	214
<i>Iti Chaturvedi and Jagath C. Rajapakse</i>	
TopEVM: Using Co-occurrence and Topology Patterns of Enzymes in Metabolic Networks to Construct Phylogenetic Trees	225
<i>Tingting Zhou, Keith C.C. Chan, and Zhenghua Wang</i>	
Generating Synthetic Gene Regulatory Networks	237
<i>Ramesh Ram and Madhu Chetty</i>	

Part IV: Microarray and Gene Expression Analysis

Gene Selection for Microarray Data by a LDA-Based Genetic Algorithm	250
<i>Edmundo Bonilla Huerta, Béatrice Duval, and Jin-Kao Hao</i>	

Sequential Forward Selection Approach to the Non-unique Oligonucleotide Probe Selection Problem	262
<i>Lili Wang, Alioune Ngom, and Luis Rueda</i>	
On Finding and Interpreting Patterns in Gene Expression Data from Time Course Experiments	276
<i>Yvonne E. Pittelkow and Susan R. Wilson</i>	
Microarray Design Using the Hilbert–Schmidt Independence Criterion	288
<i>Justin Bedo</i>	
Identifying Non-random Patterns from Gene Expression Profiles	299
<i>Radhakrishnan Nagarajan, Meenakshi Upreti, and Mariofanna Milanova</i>	
A Study on the Importance of Differential Prioritization in Feature Selection Using Toy Datasets	311
<i>Chia Huey Ooi, Shyh Wei Teng, and Madhu Chetty</i>	
Weighted Top Score Pair Method for Gene Selection and Classification	323
<i>Huaïen Luo, Yuliansa Sudibyó, Lance D. Miller, and R. Krishna Murthy Karuturi</i>	
Part V: Data Mining and Knowledge Discovery	
Identifying Conserved Discriminative Motifs	334
<i>Jyotsna Kasturi, Raj Acharya, and Ross Hardison</i>	
Exploratory Data Analysis for Investigating GC-MS Biomarkers	349
<i>Ken McGarry, Kim Bartlett, and Morteza Pourfarzam</i>	
Multi-relational Data Mining for Tetratricopeptide Repeats (TPR)-Like Superfamily Members in <i>Leishmania spp.</i> : Acting-by-Connecting Proteins	359
<i>Karen T. Girão, Fátima C.E. Oliveira, Kaio M. Farias, Italo M.C. Maia, Samara C. Silva, Carla R.F. Gadelha, Laura D.G. Carneiro, Ana C.L. Pacheco, Michel T. Kamimura, Michely C. Diniz, Maria C. Silva, and Diana M. Oliveira</i>	
Heuristic Non Parametric Collateral Missing Value Imputation: A Step Towards Robust Post-Genomic Knowledge Discovery	373
<i>Muhammad Shoaib B. Sehgal, Iqbal Gondal, Laurence S. Dooley, and Ross Coppel</i>	
Protein Expression Molecular Pattern Discovery by Nonnegative Principal Component Analysis	388
<i>Xiaoxu Han and Joseph Scazzero</i>	

Gene Ontology Assisted Exploratory Microarray Clustering and Its Application to Cancer 400
Geoff Macintyre, James Bailey, Daniel Gustafsson, Alex Boussioutas, Izhak Haviv, and Adam Kowalczyk

Discovery of Biomarkers for Hexachlorobenzene Toxicity Using Population Based Methods on Gene Expression Data 412
Cem Meydan, Alper Küçükural, Deniz Yörükoğlu, and O. Uğur Sezerman

Part VI: Applications of High Performance Computing

Exploiting Fine-Grained Parallelism in the Phylogenetic Likelihood Function with MPI, Pthreads, and OpenMP: A Performance Study 424
Alexandros Stamatakis and Michael Ott

Massively Parallelized DNA Motif Search on the Reconfigurable Hardware Platform COPACOBANA 436
Jan Schröder, Lars Wienbrandt, Gerd Pfeiffer, and Manfred Schimmler

GPU-MEME: Using Graphics Hardware to Accelerate Motif Finding in DNA Sequences 448
Chen Chen, Bertil Schmidt, Liu Weiguo, and Wolfgang Müller-Wittig

Accelerating BLASTP on the Cell Broadband Engine 460
Huiliang Zhang, Bertil Schmidt, and Wolfgang Müller-Wittig

Author Index 471