GENOME INFORMATICS SERIES (GIS)  
ISSN: 0919-9454  

The Genome Informatics Series publishes peer-reviewed papers presented at the International Conference on Genome Informatics (GIW) and some conferences on bioinformatics. The Genome Informatics Series is indexed in MEDLINE.

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Year</th>
<th>ISBN CI/Pa.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Genome Informatics Workshop I</td>
<td>1990</td>
<td>(in Japanese)</td>
</tr>
<tr>
<td>2</td>
<td>Genome Informatics Workshop II</td>
<td>1991</td>
<td>(in Japanese)</td>
</tr>
<tr>
<td>3</td>
<td>Genome Informatics Workshop III</td>
<td>1992</td>
<td>(in Japanese)</td>
</tr>
<tr>
<td>4</td>
<td>Genome Informatics Workshop IV</td>
<td>1993</td>
<td>4-946443-20-7</td>
</tr>
<tr>
<td>5</td>
<td>Genome Informatics Workshop 1994</td>
<td>1994</td>
<td>4-946443-24-X</td>
</tr>
<tr>
<td>6</td>
<td>Genome Informatics Workshop 1995</td>
<td>1995</td>
<td>4-946443-33-9</td>
</tr>
<tr>
<td>7</td>
<td>Genome Informatics 1996</td>
<td>1996</td>
<td>4-946443-37-1</td>
</tr>
<tr>
<td>8</td>
<td>Genome Informatics 1997</td>
<td>1997</td>
<td>4-946443-47-9</td>
</tr>
<tr>
<td>10</td>
<td>Genome Informatics 1999</td>
<td>1999</td>
<td>4-946443-59-2</td>
</tr>
<tr>
<td>11</td>
<td>Genome Informatics 2000</td>
<td>2000</td>
<td>4-946443-65-7</td>
</tr>
<tr>
<td>12</td>
<td>Genome Informatics 2001</td>
<td>2001</td>
<td>4-946443-72-X</td>
</tr>
<tr>
<td>13</td>
<td>Genome Informatics 2002</td>
<td>2002</td>
<td>4-946443-79-7</td>
</tr>
<tr>
<td>14</td>
<td>Genome Informatics 2003</td>
<td>2003</td>
<td>4-946443-82-7</td>
</tr>
<tr>
<td>15</td>
<td>Genome Informatics 2004 Vol. 15, No. 1</td>
<td>2004</td>
<td>4-946443-88-6</td>
</tr>
<tr>
<td>16</td>
<td>Genome Informatics 2004 Vol. 15, No. 2</td>
<td>2004</td>
<td>4-946443-91-6</td>
</tr>
<tr>
<td>17</td>
<td>Genome Informatics 2005 Vol. 16, No. 1</td>
<td>2005</td>
<td>4-946443-93-2</td>
</tr>
<tr>
<td>18</td>
<td>Genome Informatics 2005 Vol. 16, No. 2</td>
<td>2005</td>
<td>4-946443-96-7</td>
</tr>
<tr>
<td>19</td>
<td>Genome Informatics 2006 Vol. 17, No. 1</td>
<td>2006</td>
<td>4-946443-97-5</td>
</tr>
<tr>
<td>20</td>
<td>Genome Informatics 2006 Vol. 17, No. 2</td>
<td>2006</td>
<td>4-946443-99-1</td>
</tr>
<tr>
<td>21</td>
<td>Genome Informatics 2007 Vol. 18</td>
<td>2007</td>
<td>978-1-86094-991-3</td>
</tr>
<tr>
<td>23</td>
<td>Genome Informatics 2008 Vol. 20</td>
<td>2008</td>
<td>978-1-84816-299-0</td>
</tr>
<tr>
<td>24</td>
<td>Genome Informatics 2008 Vol. 21</td>
<td>2008</td>
<td>978-1-84816-331-7</td>
</tr>
</tbody>
</table>
Genome Informatics 2008
Proceedings of the 19th International Conference

Gold Coast, Queensland, Australia 1 – 3 December 2008

Editors

Jonathan Arthur
University of Sydney, Australia

See-Kiong Ng
Institute for Infocomm Research, Singapore
CONTENTS

Preface ix
Acknowledgments xi
Committees xiii

Part A  Full Papers 1

An Approach to Transcriptome Analysis of Non-Model Organisms Using Short-Read Sequences

Factoring Local Sequence Composition in Motif Significance Analysis
   P. Ng & U. Keich 15

A New Model of Multi-Marker Correlation for Genome-Wide Tag SNP Selection
   W.-B. Wang & T. Jiang 27

Phenotype Profiling of Single Gene Deletion Mutants of E. coli Using Biolog Technology
   Y. Tohsato & H. Mori 42

Improved Algorithms for Enumerating Tree-Like Chemical Graphs with Given Path Frequency
   Y. Ishida, L. Zhao, H. Nagamochi & T. Akutsu 53

BSAlign: A Rapid Graph-Based Algorithm for Detecting Ligand-Binding Sites in Protein Structures
   Z. Aung & J. C. Tong 65
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein Complex Prediction Based on Mutually Exclusive Interactions in Protein Interaction Network</td>
<td>77</td>
</tr>
<tr>
<td>On the Reconstruction of the <em>Mus musculus</em> Genome-Scale Metabolic Network Model</td>
<td>89</td>
</tr>
<tr>
<td>L.-E. Quek &amp; L. Nielsen</td>
<td></td>
</tr>
<tr>
<td>Predicting Differences in Gene Regulatory Systems by State Space Models</td>
<td>101</td>
</tr>
<tr>
<td>Exploratory Simulation of Cell Ageing Using Hierarchical Models</td>
<td>114</td>
</tr>
<tr>
<td>M. Cvijovic, H. Soueidan, D. J. Sherman, E. Klipp &amp; M. Nikolski</td>
<td></td>
</tr>
<tr>
<td>Inferring Differential Leukocyte Activity from Antibody Microarrays Using a Latent Variable Model</td>
<td>126</td>
</tr>
<tr>
<td>J. W. K. Ho, R. Koundinya, T. S. Caetano, C. G. dos Remedios &amp; M. A. Charleston</td>
<td></td>
</tr>
<tr>
<td>Assessing and Predicting Protein Interactions Using Both Local and Global Network Topological Metrics</td>
<td>138</td>
</tr>
<tr>
<td>G. Liu, J. Li &amp; L. Wong</td>
<td></td>
</tr>
<tr>
<td>Modelling the Evolution of Protein Coding Sequences Sampled from Measurably Evolving Populations</td>
<td>150</td>
</tr>
<tr>
<td>M. Goode, S. Guindon &amp; A. Rodrigo</td>
<td></td>
</tr>
<tr>
<td>A Phylogenomic Approach for Studying Plastid Endosymbiosis</td>
<td>165</td>
</tr>
<tr>
<td>Cis-Regulatory Element Based Gene Finding: An Application in <em>Arabidopsis thaliana</em></td>
<td>177</td>
</tr>
<tr>
<td>Y. Li, Y. Zhu, Y. Liu, Y. Shu, F. Meng, Y. Lu, B. Liu, X. Bai &amp; D. Guo</td>
<td></td>
</tr>
<tr>
<td>Using Simple Rules on Presence and Positioning of Motifs for Promoter Structure Modeling and Tissue-Specific Expression Prediction</td>
<td>188</td>
</tr>
<tr>
<td>A. Vandenbon &amp; K. Nakai</td>
<td></td>
</tr>
</tbody>
</table>
Improving Gene Expression Cancer Molecular Pattern Discovery Using Nonnegative Principal Component Analysis
   X. Han

Simulation Analysis for the Effect of Light-Dark Cycle on the Entrainment in Circadian Rhythm
   N. Mitou, Y. Ikegami, H. Matsuno, S. Miyano & S.-I. T. Inouye

Part B Keynote Addresses

Sequencing the Transcriptome in toto
   S. M. Grimmond

Modern Homology Search
   M. Li

Modeling Human Genome-Wide Combinatorial Regulatory Networks Initiated by Transcription Factors and microRNAs Using Forward and Reverse Engineering
   Y.-X. Li

Reconstructing the Circuits of Disease: From Molecular States to Physiological States
   E. E. Schadt

The Emerging Generalizations of Prokaryotic Genomics
   E. V. Koonin

A New Understanding of the Human Genome
   J. Mattick

Author Index
This page intentionally left blank
PREFACE

This book contains papers presented at the Nineteenth International Conference on Genome Informatics (GIW 2008) held on the Gold Coast, Queensland, Australia on December 1st to 3rd, 2008.

The GIW series provides an international forum for presentation and discussion of original research papers on all aspects of bioinformatics, computational biology, and systems biology. Its scope includes biological sequence analysis, protein structure prediction, gene regulatory networks, clustering algorithms, comparative genomics, text mining, and many other areas. GIW has a history of 19 years and is the longest running international bioinformatics conference. The first GIW was held at Kikai Shinko Kaikan, Tokyo during December 3-4, 1990 as an open workshop just before the Japanese Human Genome Project started in 1991. GIW 2008 was the first time the conference has been held in Australia. This year it was hosted by Bioinformatics Australia, representing the bioinformatics community in Australia, and incorporated the annual Bioinformatics Australia conference. Bioinformatics Australia is organized within AusBiotech, the national peak body for biotechnology in Australia.

The Program Committee of GIW 2008 received a total of 55 submissions from authors in 16 different countries around the world. Each submitted paper was peer-reviewed by at least three members of the Program Committee. Based on their reports, 18 papers were accepted (33%) for presentation at the conference. These 18 papers appear in this book and are indexed in Medline. In addition, this book contains abstracts from the six invited speakers: Sean Grimmond, University of Queensland (Australia), Eugene Koonin, National Centre for Biotechnology Information (USA), Ming Li, University of Waterloo (Canada), Yixue Li, Shanghai Jiaotong University (China), John Mattick, University of Queensland (Australia), and Eric Schadt, Rosetta Inpharmatics (USA).

The electronic versions of all the papers in this issue are also publicly available from the website of the Japanese Society for Bioinformatics (JSBi) (http://www.jsbi.org/journal.html).

Jonathan Arthur
See-Kiong Ng
GIW 2008 Program Committee Co-Chairs

Mark Ragan
GIW 2008 Conference Chair
ACKNOWLEDGMENTS

We thank all the authors for their efforts in preparing their manuscripts. We also appreciate the great efforts made by the Program Committee members in rigourously reviewing the manuscripts. The high quality of the papers presented by the authors provided a challenging task in selecting the very best for acceptance. We greatly appreciate the time and effort of both the authors and the Program Committee, in their respective contributions, to continuing the GIW tradition of a high quality, engaging scientific program.

We also acknowledge Bioinformatics Australia (within AusBiotech Ltd) for hosting GIW 2008 as well as the assistance of the National Organizing Committee, the Local Organizing Committee, and the Conference Organisers (Martin Lack and Associates) for the coordination of the conference.

We are grateful for the support of the Department of Innovation, Industry, Science and Research, the Queensland State Government, and:

- AIST Computational Biology Research Center
- ARC Research Network in Enterprise Information Infrastructure
- Australian Centre for Plant Functional Genomics
- Australian Genome Research Facility
- CSIRO
- NICTA
- Queensland Cyber Infrastructure Foundation
- SGI
- Sydney Bioinformatics
- University of Queensland

Finally, we give special thanks to those who presented papers or posters at GIW 2008, and those who attended the conference. GIW 2008 would not be a complete success without their enthusiastic participation.
This page intentionally left blank
PROGRAM COMMITTEE

Jonathan Arthur – University of Sydney, Australia; Co-Chair
See-Kiong Ng – Institute for Infocomm Research, Singapore; Co-Chair
Cathy Abbott – Flinders University, Australia
Gary Bader – University of Toronto, Canada
Vladimir Bajic – South African National Bioinformatics Institute, South Africa
Christopher Baker – Institute for Infocomm Research, Singapore
Guillaume Bourque – Genome Institute of Singapore, Singapore
Jung-Hsien Chiang – National Cheng Kung University, Taiwan
Francis YL Chin – University of Hong Kong, Hong Kong
Peter Clote – Boston College, USA
Aaron Darling – University of Queensland, Australia
Bhaskar DasGupta – University of Illinois, USA
Colin Dewey – University of Wisconsin, USA
Chris Ding – University of Texas at Arlington, USA
Roland Dunbrack – Fox Chase Cancer Center, USA
Jenny Graves – Australian National University, Australia
Win Hide – South African National Bioinformatics Institute, South Africa
Tamás Horváth – University of Bonn and Fraunhofer IAIS, Germany
Wen-Lian Hsu – Academia Sinica, Taiwan
Seiya Imoto – University of Tokyo, Japan
Lars Jermiin – University of Sydney, Australia
Minoru Kanehisa – Kyoto University, Japan
George Karypis – University of Minnesota, USA
Uri Keich – Cornell University, USA
Daisuke Kihara – Purdue University, USA
Edda Klipp – Max Planck Institute for Molecular Genetics, Germany
Stefen Kramer – Technische Universität München, Germany
Dong-Yup Lee – Bioprocessing Institute & National University of Singapore, Singapore
Sang Yup Lee – KAIST, Korea
Ming Li – University of Waterloo, Canada
Frédérique Lisacek – Swiss Institute of Bioinformatics, Switzerland
Hiroshi Mamitsuka – Kyoto University, Japan
Aleksandar Milosavljevic – Baylor College of Medicine, USA
Satoru Miyano – University of Tokyo, Japan
Bernard Moret – Swiss Federal Institute of Technology, Switzerland
Shin-ichi Morishita – University of Tokyo, Japan
Pablo Moscato – University of Newcastle, Australia
William Stafford Noble – University of Washington, USA
Laxmi Parida – IBM T. J. Watson Research Center, USA
Ron Pinter – Technion, Israel
Shoba Ranganathan – Macquarie University, Australia
Allen Rodrigo – University of Auckland, New Zealand
Rintaro Saito – Keio University, Japan
Yasubumi Sakakibara – Keio University, Japan
Christian Schönbach – Nanyang Technological University, Singapore
Tetsuo Shibuya – University of Tokyo, Japan
Mona Singh – Princeton University, USA
Wing Kin Sung – National University of Singapore, Singapore
Koji Tsuda – Max Planck Institute for Biological Cybernetics, Germany
Alfonso Valencia – Universidad Autonoma, Spain
Gabriel Valiente – Technical University of Catalonia, Spain
Jean-Philippe Vert – École des Mines de Paris, France
Lusheng Wang – The City University of Hong Kong, Hong Kong
Marc Wilkins – University of New South Wales, Australia
Michael Wise – University of Western Australia, Australia
Ying Xu – University of Georgia, USA
Gwan-Su Yi – Information & Communications University, Korea
Mohammed J. Zaki – Rensselaer Polytechnic Institute, USA

CO-REVIEWERS

Satya Arjunan
Jun-tao Guo
Chris Kauffman
Tadahiko Sakiyama
Haibao Tang

Hong-Jie Dai
Kosuke Hashimoto
Ian Menz
Michael Shmoish
Katsuyuki Yugi

Kevin DeRonne
Rajaraman Kanagasabai
Nini Rao
Michihiro Tanaka
STEERING COMMITTEE

Minoru Kanehisa – Kyoto University, Japan
Satoru Miyano – University of Tokyo, Japan
Mark Ragan – University of Queensland, Australia
Toshihisa Takagi – University of Tokyo, Japan
Limsoon Wong – National University of Singapore, Singapore

CONFERENCE CHAIR

Mark Ragan – University of Queensland, Australia

NATIONAL ORGANIZING COMMITTEE

Cathy Abbott – Flinders University, Australia
Jonathan Arthur – University of Sydney, Australia
Tim Bailey – University of Queensland, Australia
Mark Baker – Australian Proteome Analysis Facility, Australia
Jeremy Barker – Queensland Facility for Advanced Bioinformatics, Australia
Matthew Bellgard – Murdoch University, Australia
Kevin Burrage – University of Queensland, Australia
Phoebe Chen – Deakin University, Australia
Ross Coppel – Monash University, Australia
Brian Dalrymple – CSIRO Livestock Industries, Australia
Simon Easteal – Australian National University, Australia
Dave Edwards – Australian Centre for Plant Functional Genomics, Australia
Sue Forrest – Australian Genome Research Facility, Australia
Bruno Gaëta – University of New South Wales, Australia
Jenny Graves – Australian National University, Australia
David Hansen – Australian e-Health Research Centre, Australia
James Hogan – Queensland University of Technology, Australia
Jonathan Keith – Queensland University of Technology, Australia
Vladimir Likic – University of Melbourne & Bio21, Australia
Committees

Tim Littlejohn – IBM Australia, Australia
John Mattick – University of Queensland, Australia
Geoff McLachlan – University of Queensland, Australia
Annette McGrath – Australian Genome Research Facility, Australia
David Mitchell – CSIRO CMIS, Australia
Pablo Moscato – University of Newcastle, Australia
Tuan Pham – James Cook University, Australia
Michael Poidinger – Johnson & Johnson, Australia
Mark Ragan – University of Queensland, Australia
Shoba Ranganathan – Macquarie University, Australia
Allen Rodrigo – University of Auckland, New Zealand
Rohan Teasdale – University of Queensland, Australia
Mervyn Thomas – Emphron Informatics, Australia
Matthew Wakefield – Walter & Eliza Hall Institute of Medical Research, Australia

Marc Wilkins – University of New South Wales, Australia
Sue Wilson – Australian National University, Australia
Michael Wise – University of Western Australia, Australia
Xiaofang Zhou – University of Queensland, Australia
Albert Zomaya – University of Sydney, Australia

LOCAL ORGANIZING COMMITTEE

Mark Ragan – University of Queensland, Australia
Tim Bailey – University of Queensland, Australia
Mikael Bodén – University of Queensland, Australia
Brian Dalrymple – CSIRO Livestock Industries, Australia
Dave Edwards – Australian Centre for Plant Functional Genomics, Australia
James Hogan – Queensland University of Technology, Australia
Rohan Teasdale – University of Queensland, Australia