

# **5th International Conference on Bioinformatics and Computational Biology 2013**

**(BICoB-2013)**

**Honolulu, Hawaii, USA  
4-6 March 2013**

ISBN: 978-1-62276-971-1

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2013) by the International Society for Computers and Their Applications  
All rights reserved. Reproduction in any form without the written consent of ISCA is prohibited.

Original ISBN: 978-1-880843-89-5 (Out of Print)  
Reprint ISBN: 978-1-62276-971-1

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the International Society for Computers and Their Applications  
at the address below.

International Society for Computers and Their Applications  
975 Walnut Street, Suite 132  
Cary, NC 27511-4216

Phone: (919) 467-5559  
Fax: (919) 467-3430

[isca@ipass.net](mailto:isca@ipass.net)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Table of Contents

<b>GENE EXPRESSION AND G.E. NETWORKS</b>	<b>1</b>
Towards Better Base Classifiers for Ensemble Classification of Gene Expression Data <i>William Duncan, Jian Zhang</i> .....	1
Inferring Directed-graph Patterns of Gene Responses to Treatments <i>Vinhthuy Phan, Nam S Vo, Thomas R Sutter</i> .....	7
Exploration of Gene Expression Data via Constrained Clustering <i>Marjan Trutschl, Phillip C.S.R. Kilgore, Urska Cvek, Robert E. Rhoads</i> .....	13
<b>DNA, RNA AND BIOLOGICAL SEQUENCES</b>	<b>19</b>
Finding Gapped Motifs using Particle Swarm Optimization Technique <i>U. Srinivasulu Reddy, Michael Arock, A.V. Reddy</i> .....	19
Flexible Approach for Novel Transcript Reconstruction from RNA-Seq Data using Maximum Likelihood Integer Programming <i>Serghei Mangul, Adrian Caciula, Sahar A. Seesi, Dumitru Brinza, Abdul R. Bandy, Rahul Kanadia, Ion Mandoiu, Alex. Zelikovsky</i> .....	25
Integrating RNA-seq transcript signals, Primary and Secondary Structure Information in Differentiating Coding and Non-coding RNA Transcripts <i>Ashis Kumer Biswas, Jean Gao, Xiaoyong Wu, Baoju Zhang</i> .....	35
<b>PROTEIN STRUCTURE AND FUNCTIONS – 1</b>	<b>41</b>
Protein Threading Based on Nonlinear Integer Programming <i>Wajeb Gharib Gharibi, Marwah Mohammed Bakri</i> .....	41
Detecting Intermediate Structures in Protein Conformational Pathways <i>Nurit Haspel, Dong Luo, Eduardo Gonzalez</i> .....	47
Efficient Coarse-Grained Geometry-Based Sampling of Protein Conformational Paths <i>Dong Luo, Nurit Haspel</i> .....	53
<b>CLASSIFICATION AND CLUSTERING</b>	<b>59</b>
Structured Motif Extraction using Affinity Based Cluster Analysis <i>Faisal Alobaid, Kishan Mehrotra, Chilukuri Mohan, Ramesh Raina</i> .....	59
Document Classification: A Novel Approach Based on SVM <i>Rania Kilany, Reda Ammar, Sanguthevar Rajasekaran</i> .....	67
Resolving Read Assignment Ambiguities in Metagenomic Clustering <i>Rahul Nihalani, Jaroslaw Zola, Srinivas Aluru</i> .....	73
<b>PROTEIN STRUCTURE AND FUNCTIONS – 2</b>	<b>81</b>
Ontology Based Semantic Similarity for Protein Interactions <i>Xiao Luo, Hisham Almubaid, Said Bettayeb</i> .....	81
A Coarse-grained Context-dependent Contact Potential for Protein Decoy Discrimination <i>Yaohang Li</i> .....	89
Applications of Conserved Indels and Signature Proteins for Microbial Phylogeny <i>Radhey S. Gupta</i> .....	95

<b>BIOLOGICAL NETWORKS AND MODELING – 1</b>	<b>101</b>
A Graph Theoretic Mathematical Model for Alzheimer’s disease <i>C. Rose Kyrtos, John S. Baras</i> .....	101
Querying on Model Abstractions of Gene Regulation from Noisy Data <i>Krishnendu Ghosh, John Schlipf</i> .....	107
Computational Approaches for Predicting Interaction Sites of Cytochrome and Photosystem I <i>Wei Chen, Ali Sekmen, Barry D. Bruce, Khoa Nguyen, P. Mishra, L. Emujakporue, K. Wehbi</i> .....	113
<b>BIOINFORMATICS APPLICATIONS – 1</b>	<b>119</b>
Intellectual Property Protection for Bioinformatics and Computational Biology <i>Dennis Fernandez, Antonia Maninang, Shumpei Kobayashi</i> .....	119
Binary Classification of Compounds by Learning from Docking Software Results and Chemical Information <i>Masato Okada, Katsutoshi Kanamori, Hayato Ohwada, Shin Aoki</i> .....	125
Similarity Search in Protein Sequence Databases using Metric Access Methods <i>Ahmet Cetintas, Ahmet Sacan, Ismail H. Toroslu</i> .....	131
<b>ALGORITHMS IN BIOINFORMATICS</b>	<b>137</b>
PreGO: A Protein Function Prediction Algorithm Based on an Infinite Mixture of Hidden Markov and Bayesian Network Models <i>Takashi Kaburagi, Yukihiro Koizumi, Kousuke Oota, Takashi Matsumoto</i> .....	137
A Genetic Algorithm for the Reversal Median Problem <i>Nan Gao, Fei Hu, Jijun Tang</i> .....	145
Reference Matters: An Efficient and Scalable Algorithm for Large Multiple Structure Alignment <i>Jose S. Hleap, Khanh N. Nguyen, Alex Safatli, Christian Blouin</i> .....	153
<b>GENE EXPRESSION AND GENE SEQUENCE</b>	<b>159</b>
Gene expression network analysis reveals pathway interactions of the IGF axis in breast cancer <i>Emad Ramadan, Sudhir Perincheri, Emmett Sprecher, Lyndsay Harris, David Tuck</i> .....	159
Dynamic Bayesian Clustering of Gene Expression Data <i>Anna Fowler, Nicholas A. Heard</i> .....	165
Analysis of the Impact of Sequencing Errors on BLAST Using Fault Injection <i>So Youn Lee, Brielle J. Fischman, Steven S. Lumetta, Zbigniew Kalbarczyk, Ravishankar Iyer</i> .....	171
<b>MACHINE LEARNING AND PREDICTION IN BIOINFORMATICS</b>	<b>177</b>
Predicting ATP-Binding Pockets Based on Amino Acid Microenvironment <i>Jing Hu, Andrew S. Forcier, Changhui Yan</i> .....	177
Selecting Biomarkers that Better Predicts Cancer Outcome by Functional Transformation of Clinical Variables <i>Shang Gao, A. M. Elsheikh, Reda Alhadj, G. Wang, Jon Rokne</i> .....	181
An Efficient Comparative Machine Learning-based Metagenomics Binning Technique Via Optimal Feature-reduction Methods <i>Helal Saghir, Dalila B. Megherbi</i> .....	187
<b>BIOLOGICAL NETWORKS AND MODELING – 2</b>	<b>193</b>
Two-level mixed modeling of longitudinal pedigree data for genetic association analysis <i>Qihua Tan</i> .....	193
Multi-scale Modularity and Motif Distributional Effect in Metabolic Networks <i>Shang Gao, Omar Addam, Alan Chen, Ali Rahmani, Jia Zeng, Mehmet Tan, Reda Alhadj, Jon Rokne, Douglas Demetrick</i> .....	199
A Hybrid-System Model of the Coagulation Cascade <i>Joseph Makin, Srinu Narayanan</i> .....	205

<b>BIOINFORMATICS APPLICATIONS – 2</b>	<b>213</b>
Information theoretic feature selection for Weisfeiler-Lehman graph kernels <i>Mehmet Tan</i> .....	213
Automatic Annotation of GPI Structures Using Grid Computing <i>Clemente Aguilar-Bonavides, Gerardo A. Cardenas, Ernesto S. Nakayasu, Felipe Gazos-Lopes, Igor C. Almeida, Ming-Ying Leung</i> .....	219
Exact Solutions for Classic Gene Tree Parsimony Problems <i>Wen-Chieh Chang, Andre Wehe, Pawel Gorecki, Oliver Eulenstein</i> .....	225
<b>WORKING WITH BIOINFORMATICS DATA AND ANALYSIS</b>	<b>231</b>
Analysis and Visualization of Cell State Transitions During Differentiation <i>Ahmadreza Ghaarizadeh, Gregory J. Podgorski, Nicholas S. Flann</i> .....	231
Providing Cloud-based Metabolite Annotations for NMR Spectroscopic Data with Semantic Web Integration <i>Paul E. Anderson, Edward J. Pharr, Michael Peterson</i> .....	237
Kernel Based Difference Detection in LC-MS Data <i>Minh Nguyen, Xiaoyong Wu, Baoju Zhang, Jean X. Gao</i> .....	245
<b>BIOINFORMATICS APPLICATIONS – 3</b>	<b>251</b>
Sensitivity and block sensitivity of nested canalizing function <i>Yuan Li, John O. Adeyeye</i> .....	251
Physicochemical Determinants of Antimicrobial Activity <i>Daniel Veltri, Amarda Shehu</i> .....	253
Sorting Genomes Using Symmetric, Almost-Symmetric and Unitary Inversions <i>Ulisses Dias, Zanoni Dias</i> .....	261
<b>BIOINFORMATICS APPLICATIONS – 4</b>	<b>269</b>
A Synchronization Detection Approach for Identifying Rare Mutations underlying Common Disease <i>Jiayin Wang, Xuanping Zhang, Yanqin Liu, Jin Zhang, Yufeng Wu</i> .....	269
Simple and Accurate Trend Tests Using a Permutation Approximation <i>Yi-Hui Zhou, Fred A. Wright</i> .....	277
Computation verification of a potential pico-inversion with multi-species comparison <i>Minmei Hou, Ping Yao, Mitrick A. Johns</i> .....	283
<b>DATA INTENSIVE AND ANALYSIS APPLICATIONS</b>	<b>289</b>
Comparative Analysis of de novo Transcriptome Assembly <i>Kaitlin Clarke, Yi Yang, Ronald Marsh, LingLin Xie, Ke K Zhang</i> .....	289
Accelerating Data-Intensive Genome Analysis in the Cloud <i>Nabeel M Mohamed, Heshan Lin, Wu-chun Feng</i> .....	297
<b>Author Index</b>	<b>305</b>