

# **6th International Conference on Bioinformatics and Computational Biology**

**(BICoB 2014)**

**Las Vegas, Nevada, USA  
24 – 26 March 2014**

**Editors:**

**F. Saeed**

**B. DasGupta**

**ISBN: 978-1-63266-514-0**

**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© (2014) 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: 978-1-880843-94-9 (Out of Print)  
Reprint ISBN: 978-1-63266-514-0

Printed by Curran Associates, Inc. (2014)

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>KEYNOTE</b>	<b>1</b>
Computational Experiment Planning and the Future of Big Data <i>Christopher J. Lee</i> .....	1
<b>DATA MINING AND MACHINE LEARNING I</b>	<b>3</b>
Hierarchical Algorithm for Pattern Extraction from Biological Sequences <i>Faouzi Mhamdi, Mehdi Kchouk</i> .....	3
Discovering Compounds That Activate Plant Immunity Using Machine Learning <i>Takaya Yoshida, Nobutaka Kitahata, Hayato Ohwada, Kazuyuki Kuchitsu</i> .....	9
A K-means based approach for the design of Rotamer Libraries: Preliminary Results <i>Sara Hernandez, Carlos Brizuela</i> .....	15
Semi-supervised Classification of Protein Sequences Using Burrows Wheeler Transformation-based Features <i>Karthik Tangirala, Doina Caragea</i> .....	21
<b>RNA AND DNA I</b>	<b>27</b>
Transcriptomic Analysis by RNA-Seq and Gene Enrichment Analysis <i>Scott Nixon, Dianelys Gonzales Pena, Marcus Lawson, Robert McCusker, Jason O'Connor, Robert Dantzer, Keith W. Kelley, Sandra L. Rodriguez-Zas</i> .....	27
A Novel 2-Dimensional Graphical Representation of DNA Sequences Using Weighted Vector Assignments <i>Satoshi Mizuta, Kyohei Yamaguchi</i> .....	33
CircRNAFinder: a Tool for Identifying Circular RNAs Using RNA-Seq Data <i>Xing Fu, Renyi Liu</i> .....	39
Advanced statistical analysis of RNA-seq differential gene expression profiles of animal gene-environment interactions <i>Kelsey Caetano-Anolles, Scott Nixon, Sandra L. Rodriguez-Zas</i> .....	43
<b>MEDICAL INFORMATICS AND APPLICATIONS</b>	<b>49</b>
C-Cell Detector: High-Throughput Data Processing Framework for the Detection and Identification of Biological Targets using Solid-State Micropores <i>Abdul Hafeez, M.Mustafa Rafique, Ali R. Butt</i> .....	49
Semantics-Enabled Proactive and Targeted Dissemination of New Medical Knowledge <i>Ismailcem Arpinar, Asmita Rahman, Priya Wadhwa, Lakshmish Ramaswamy, Ugur Kursuncu</i> .....	55
In Search of Optimal Space-Filling Curves for 3-D to 1-D Mapping: Application to 3-D Brain MRI Data <i>Ünal Sakoğlu, Abdullah Arslan, Kushal Bohra, Heriberto Flores</i> .....	61
Potential Relationship Discovery Model Base on Context and Its Application in Lung Cancer Field <i>Haishan Chen, Jing Lan, Xiaoxia Zhang, Xiaojiao Xu, Cui Wang, Qingqiang Wu</i> .....	67
<b>GENES AND PROTEINS</b>	<b>73</b>
Analyzing Conservation Patterns and Its Influence on Identifying Protein Functional Sites <i>Chun Fang, Tamotsu Noguchi, Hayato Yamana</i> .....	73
New Weighting Method of Protein Sequences for Protein Classification <i>Faouzi Mhamdi, Salma Aouled Elhaj Mohamed</i> .....	81

Predicting protein subcellular localization with reduced feature set <i>Hishan Al-Mubaid, Izzat Alsmadi</i> .....	87
Understanding the impact of fractionation and dynamic exclusion on proteomics <i>Cynthia Zavala, Scott Nixon, Bruce Southey, Cong Wu, Elena Romanova, Jonathan V. Sweedler, Sandra L. Rodriguez-Zas</i> .....	93
<b>BIOINFORMATICS APPLICATIONS I</b>	<b>97</b>
Improved Shortest Path Edit Distance for Synonyms Identification <i>Alex Rudyin, Min Song, James Geller</i> .....	97
Evaluation of Significance Level Assignment of Database Search Programs using Monte Carlo Permutation Approach <i>Malik Nadeem Akhtar, Bruce Robert Southey, Per E. Andren, Jonathan V. Sweedler, Sandra L. Rodriguez-Zas</i> .....	103
Formal Analysis on Automated Model Abstractions on a System of Biochemical Reactions under Uncertainty <i>Krishnendu Ghosh, John Schlipf</i> .....	109
Impact of the Increased in the Biological Samples and Alternatives vs. Defaults Setting in the Transcripts Analysis. <i>Dianelys Gonzalez-Peña, Nixon E Scott, Bruce R. Southey, Marcus A. Lawson, Robert H. McCusker, Robert Dantzer, Keith W. Kelley, Sandra L. Rodriguez-Zas</i> .....	115
Accurate Discrimination of Outer Membrane Proteins Using Amino Acid Composition and Auto Covariance Derived From PSI-BLAST Profiles <i>Lingyun Zou, Qingshan Ni</i> .....	119
<b>DATA MINING AND MACHINE LEARNING II</b>	<b>125</b>
Bursty Topic Detection Based on K-States Automaton Model: an Emerging Trend in Cancer Field <i>Qingqiang Wu, Haibin Zhang, Qingqi Hong, Yan Zheng</i> .....	125
Semi-supervised Self-training Approaches for Imbalanced Splice Site Datasets <i>Ana Stanescu, Doina Caragea</i> .....	131
A Probabilistic Roadmap-based Method to Model Conformational Switching of a Protein Among Many Functionally-relevant Structures <i>Kevin Molloy, Amarda Shehu</i> .....	137
Multi-Objective Optimization Techniques for Conformational Sampling in Template-Free Protein Structure Prediction <i>Brian Olson, Amarda Shehu</i> .....	143
<b>RNA AND DNA II</b>	<b>149</b>
Mitochondrial DNA Variation Analysis of Mentawai Population <i>Todia Pediatama, Yoni Syukriani, Nurhalim Shahib</i> .....	149
Decoding Methods for DNA Steganalysis <i>Marc Beck, Ahmed Desoky, Eric Rouchka, Roman Yampolskiy</i> .....	155
Evaluating Assembly Pipeline for Transcriptomes <i>Natasha Pavlovikj, Sairam Behera, Kevin Begcy, Malachy Cambell, Harkamal Walia, Jitender S Deogun</i> .....	163
Analysis of Cancer Somatic Mutations Taking into Consideration Human Genetic Variations <i>Bin He, Hanna Torkey, Azam Moosavi, Liqing Zhang</i> .....	169
<b>HIGH PERFORMANCE ALGORITHMS AND HIGH-THROUGHPUT PROCESSING</b>	<b>177</b>
Optimal Alignment of Three Sequences On A GPU <i>Junjie Li, Sanjay Ranka, Sartaj Sahni</i> .....	177
Efficient Next-Generation DNA Sequencing Analysis in Hadoop <i>Kareem Aggour, Dipen Sangurdekar, Vijay Kumar, Lee Newberg, Chinnappa Kodira</i> .....	183
Scalable Genome-Wide Discovery and Presentation of Motifs <i>Marjan Trutschl, Phillip Kilgore, Urska Cvek, Rona Scott</i> .....	191

<b>BIOINFORMATICS APPLICATIONS II</b>	<b>197</b>
Explore transcription factor profiles in human pluripotent stem cells <i>Jane Synnergren</i> .....	197
Vision-based Breast Self-Examination Hand Interaction Tracking using Sparse Optical Flow and Genetic Algorithm <i>Melvin Cabatuan, Rey Masilang, Elmer Dadios, Laurence Gan Gim</i> .....	203
A Novel Approach to Approximate Parikh Matching for Comparing Composition in Biological Sequences <i>Juan Mendivelso, Yoan Pinzon</i> .....	209
 <b>Author Index</b>	 <b>215</b>