

# **7th International Conference on Bioinformatics and Computational Biology**

**(BICoB 2015)**

**Honolulu, Hawaii, USA  
9-11 March 2015**

**Editors:**

**Fahad Saeed  
Nurit Haspel**

**ISBN: 978-1-5108-0013-7**

**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© (2015) 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: 978-1-880843-99-4 (Out of Print)  
Reprint ISBN: 978-1-5108-0013-7

Printed by Curran Associates, Inc. (2015)

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 Epigenomics and Regulatory Genomics for Deciphering the Non-coding Human Genome <i>Jason Ernst</i> .....	1
<b>DATA MINING AND MACHINE LEARNING - 1</b>	<b>3</b>
CentroidBLAST: Accelerating Sequence Search via Clustering <i>Wu-Chun Feng, Konstantinos Krommydas, Liqing Zhang</i> .....	3
Integrative Clustering of Cancer Genome Data using Infinite Relational Models <i>Yoichi Chikahara, Atsushi Niida, Rui Yamaguchi, Seiya Imoto, Satoru Miyano</i> .....	11
Combining two machine learning methods for predicting protein-ligand docking using structure and physiochemical properties <i>Tadasuke Ito, Hayato Ohwada, Shin Aoki</i> .....	19
Feature Weighting-based Classifier for Protein Subcellular Localization <i>Duong B. Nguyen, Hisham Al-Mubaid, Anurag Nagar</i> .....	25
<b>RNA AND DNA - 1</b>	<b>31</b>
Performance Evaluation of Parallel Genome Assemblers <i>Evaldo B Costa, Gabriel P. Silva, Marcello G. Teixeira</i> .....	31
MotifMutator: A Combinatoric Tool for Modeling Binding-Site Preference <i>Phillip Kilgore, Urška Cvek, Marjan Trutschl, Brandon Praslicka, Christopher Gissendanner</i> .....	39
Reducing Type I Errors in Tn-Seq Experiments by Correcting the Skew in Read Count Distributions <i>Michael Dejesus, Thomas Ioerger</i> .....	45
<b>MEDICAL INFORMATICS AND APPLICATIONS</b>	<b>51</b>
Validation of A Computational B-Cell Lymphoma Analysis by Flow Cytometry Data <i>Ming-Chih Shih, Shou-Hsuan Stephen Huang, Youli Zu, Ramesh Bhagat</i> .....	51
The Duplication and Intragenic Domain Expansion of Human C2H2 Zinc Finger Genes Are Associated with Transposable Elements And Relevant to The Expression-based Clustering <i>Wensheng Zhang, Andrea Edwards, Prescott Deininger, Kun Zhang</i> .....	57
Selection of Robust Reference Genes for Normalization of Quantitative RT-PCR Data from Differentiating Human Pluripotent Stem Cells <i>Gustav Holmgren, Xianmin Zeng, Jane Synnergren</i> .....	65
Information Distance Explains MHC II Supertypes <i>Shun Liao, Ying Fan, Lusheng Wang, Shuaicheng Li, Wenjun Shen</i> .....	71
<b>GENES AND PROTEINS - 1</b>	<b>77</b>
Gene-disease Relation Extraction and Gene Interaction Network Construction <i>Lei Hua, Changqin Quan, Fuji Ren</i> .....	77
Effective Prediction of Signaling Pathways from Protein-Protein Interaction Networks using Network Motifs <i>Yanan Xin, Young-Rae Cho</i> .....	85
Prediction of Plant Protein Subcellular Locations <i>Kofi Neizer-Ashun, Feng Yu, John Meinken, Xiangjia Min, Guang-Hwa Chang</i> .....	91

<b>BIOINFORMATICS APPLICATIONS - 1</b>	<b>97</b>
Ancestral Reconstruction with Duplications Using Binary Encoding and Probabilistic Model <i>Lingxi Zhou, Jijun Tang</i> .....	97
CSA-RRBS: A Comprehensive Streamlined Analysis of Reduced Representation Bisulfite Sequencing (RRBS) Data <i>Ting Gong, Sara Gaddis, Yue Lu, Kimie Kondo, Hongbo Zhao, Jianjun Shen, Marcelo Aldaz, Marcos Estecio</i> .....	105
ModFossa: A Python Library for Ion Channel Modeling <i>Gareth Ferneyhough, Corey Thibeault, Sergiu Dascalu, Fred Harris</i> .....	111
Comparison of Low- and High- Doses of Ionizing Radiation Using Networks of Co-regulated Biological Processes <i>YaredH Kidane, Francis A Cucinotta</i> .....	119
 <b>GENES AND PROTEINS - 2</b>	 <b>125</b>
AccuRefiner: A Machine Learning Guided Refinement Method for Protein-Protein Docking <i>Bahar Akbal-Delibas, Marc Pomplun, Nurit Haspel</i> .....	125
Consensus Properties of the Gene Duplication Problem for Enhanced Phylogenetic Inference <i>Harris T. Lin, Jucheol Moon, Oliver Eulenstein</i> .....	131
Learning Deep Architectures for Protein Structure Prediction <i>Kyungim Baek</i> .....	137
On the sampling of Big Mass Spectrometry Data <i>Muaaz Gul Awan, Fahad Saeed</i> .....	143
 <b>RNA AND DNA - 2</b>	 <b>149</b>
Optimizing Genomic Sequence Searches to Next-Generation Intel Architectures <i>Eduardo Ponce Mojica, Greg D Peterson, Bhanu Rekepalli</i> .....	149
pbSandwich: Scaffolding Draft Genomes with Long Reads <i>Aaron Steele, Scott Emrich</i> .....	155
A Bayesian Method for Assigning Ambiguous Bisulfite Short Reads <i>Hong Tran, Xiaowei Wu, Liqing Zhang</i> .....	161
Gene Ontology Summarization to Support Visualization and Quality Assurance <i>Christopher Ochs, Yehoshua Perl, Michael Halper, James Gellerand, Jane Lomax</i> .....	167
 <b>BIOINFORMATICS APPLICATIONS - 2</b>	 <b>175</b>
Visualization and Classification of DNA sequence using Pareto learning Self Organizing Maps for Short Sequences <i>Hiroshi Dozono</i> .....	175
Mebitoo: an Extensible Software Framework for Bioinformatics Analysis Workflow Automatization <i>Christian Spaniol, Mohamed Hamed, Johannes Trumm, Volkhard Helms Helms</i> .....	181
SNPwise: A SNP aware short read aligner <i>Saima Sultana Tithi, Lenwood Heath, Liqing Zhang</i> .....	187
Multi-genome Synteny for Assembly Improvement <i>Lauren Assour, Scott Emrich</i> .....	193
 <b>BIOINFORMATICS APPLICATIONS C3</b>	 <b>199</b>
The Unique Perfect Phylogeny Problem for 3-State Characters <i>Thong Le, Brad Shalters</i> .....	199
Latent Dirichlet Allocation on Top-Down Metabolic Pathway Analysis <i>Carlos M. Estévez Bretón, Liliana López, Luis F. Niño</i> .....	205
Displacement of the Tyrosyl Radical in RNR Enzyme: A Sophisticated Computational Approach to Analyze Experimental Data <i>Gajula MNV Prasad, Heinz-J. Steinhoff, Anuj Kumar, Ebrahimali A Siddiq, Anand K Polumetla, Friedhelm Lenzain</i> .....	211

Modeling Action Potentials of Body Wall Muscles in *C. elegans*: A Biologically Founded Computational Approach  
*Callen Johnson, Roger Mailler* ..... 219

**Author Index** ..... 227