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
# Table of Contents

## KEYNOTE
Computational Experiment Planning and the Future of Big Data
*Christopher J. Lee* .......................................................... 1

## DATA MINING AND MACHINE LEARNING I
Hierarchical Algorithm for Pattern Extraction from Biological Sequences
*Faouzi Mhamdi, Mehdi Kchouk* ................................................ 3

Discovering Compounds That Activate Plant Immunity Using Machine Learning
*Takaya Yoshida, Nobutaka Kitahata, Hayato Ohwada, Kazuyuki Kuchitsu* .................................................. 9

A K-means based approach for the design of Rotamer Libraries: Preliminary Results
*Sara Hernandez, Carlos Brizuela* ............................................. 15

Semi-supervised Classification of Protein Sequences Using Burrows Wheeler Transformation-based Features
*Karthik Tangirala, Doina Caragea* ........................................... 21

## RNA AND DNA I
Transcriptomic Analysis by RNA-Seq and Gene Enrichment Analysis
*Scott Nixon, Dianelys Gonzales Pena, Marcus Lawson, Robert McCusker, Jason O’Connor, Robert Dantzer, Keith W. Kelley, Sandra L. Rodriguez-Zas* ........................................... 27

A Novel 2-Dimensional Graphical Representation of DNA Sequences Using Weighted Vector Assignments
*Satoshi Mizuta, Kyohei Yamaguchi* ........................................... 33

CircRNAFinder: a Tool for Identifying Circular RNAs Using RNA-Seq Data
*Xing Fu, Renyi Liu* ................................................................. 39

Advanced statistical analysis of RNA-seq differential gene expression profiles of animal gene-environment interactions
*Kelsey Caetano-Anolles, Scott Nixon, Sandra L. Rodriguez-Zas* ................................................................. 43

## MEDICAL INFORMATICS AND APPLICATIONS
C-Cell Detector: High-Throughput Data Processing Framework for the Detection and Identification of Biological Targets using Solid-State Micropores
*Abdul Hafeez, M.Mustafa Rafique, Ali R. Butt* ............................................ 49

Semantics-Enabled Proactive and Targeted Dissemination of New Medical Knowledge
*Ismailcem Arpinar, Asmita Rahman, Priya Wadhwa, Lakshmish Ramaswamy, Ugur Kursuncu* ............................................ 55

In Search of Optimal Space-Filling Curves for 3-D to 1-D Mapping: Application to 3-D Brain MRI Data
*Ünal Sakoğlu, Abdullah Arslan, Kushal Bohra, Heriberto Flores* ................................................................. 61

Potential Relationship Discovery Model Based on Context and Its Application in Lung Cancer Field
*Haishan Chen, Jing Lan, Xiaoxia Zhang, Xiaojiao Xu, Cui Wang, Qingqiang Wu* ................................................................. 67

## GENES AND PROTEINS
Analyzing Conservation Patterns and Its Influence on Identifying Protein Functional Sites
*Chun Fang, Tamotsu Noguchi, Hayato Yamana* ............................................ 73

New Weighting Method of Protein Sequences for Protein Classification
*Faouzi Mhamdi, Salma Aouled Elhaj Mohamed* ............................................ 81
Prediciting protein subcellular localization with reduced feature set
Hishan Al-Mubaid, Izzat Alsmadi ................................. 87

Understanding the impact of fractionation and dynamic exclusion on proteomics
Cynthia Zavala, Scott Nixon, Bruce Southey, Cong Wa, Elena Romanova, Jonathan V. Sweedler, Sandra L. Rodriguez-Zas ................................. 93

BIOINFORMATICS APPLICATIONS I
Improved Shortest Path Edit Distance for Synonyms Identification
Alex Rudyin, Min Song, James Geller ................................. 97

Evaluation of Significance Level Assignment of Database Search Programs using Monte Carlo Permutation Approach
Malik Nadeem Akhtar, Bruce Robert Southey, Per E. Andren, Jonathan V. Sweedler, Sandra L. Rodriguez-Zas ................................. 103

Formal Analysis on Automated Model Abstractions on a System of Biochemical Reactions under Uncertainty
Krishnendu Ghosh, John Schlipf ................................. 109

Impact of the Increased in the Biological Samples and Alternatives vs. Defaults Setting in the Transcripts Analysis.
Dianelys Gonzalez-Peña, Nixon E Scott, Bruce R. Southey, Marcus A. Lawson, Robert H. McCusker, Robert Dantzer, Keith W. Kelley, Sandra L. Rodriguez-Zas ................................. 115

Accurate Discrimination of Outer Membrane Proteins Using Amino Acid Composition and Auto Covariance Derived From PSI-BLAST Profiles
Lingyun Zou, Qingshan Ni ................................. 119

DATA MINING AND MACHINE LEARNING II
Bursty Topic Detection Based on K-States Automaton Model: an Emerging Trend in Cancer Field
Qingqiang Wu, Haibin Zhang, Qingqi Hong, Yan Zheng ................................. 125

Semi-supervised Self-training Approaches for Imbalanced Splice Site Datasets
Ana Stanescu, Doina Caragea ................................. 131

A Probabilistic Roadmap-based Method to Model Conformational Switching of a Protein Among Many Functionally-relevant Structures
Kevin Molloy, Amarda Shehu ................................. 137

Multi-Objective Optimization Techniques for Conformational Sampling in Template-Free Protein Structure Prediction
Brian Olson, Amarda Shehu ................................. 143

RNA AND DNA II
Mitochondrial DNA Variation Analysis of Mentawai Population
Todia Pediatama, Yoni Syukriani, Nurhalim Shahib ................................. 149

Decoding Methods for DNA Steganalysis
Marc Beck, Ahmed Desoky, Eric Roucheka, Roman Yampolskiy ................................. 155

Evaluating Assembly Pipeline for Transcriptomes
Natasha Pavlovikj, Sairam Behera, Kevin Begcy, Malachy Campbell, Harkamal Walia, Jitender S Deogun ................................. 163

Analysis of Cancer Somatic Mutations Taking into Consideration Human Genetic Variations
Bin He, Hanna Torkey, Azam Moosavi, Liqing Zhang ................................. 169

HIGH PERFORMANCE ALGORITHMS AND HIGH-THROUGHPUT PROCESSING
Optimal Alignment of Three Sequences On A GPU
Junjie Li, Sanjay Ranka, Sartaj Sahni ................................. 177

Efficient Next-Generation DNA Sequencing Analysis in Hadoop
Kareem Aggour, Dipen Sangurdekar, Vijay Kumar, Lee Newberg, Chinnappa Kodira ................................. 183

Scalable Genome-Wide Discovery and Presentation of Motifs
Marjan Trutschl, Phillip Kilgore, Urska Cvek, Rona Scott ................................. 191
BIOINFORMATICS APPLICATIONS II

Explore transcription factor profiles in human pluripotent stem cells
Jane Synnergren

Vision-based Breast Self-Examination Hand Interaction Tracking using Sparse Optical Flow and Genetic Algorithm
Melvin Cabatuan, Rey Masilang, Elmer Dadios, Laurence Gan Gim

A Novel Approach to Approximate Parikh Matching for Comparing Composition in Biological Sequences
Juan Mendivelso, Yoan Pinzon

Author Index