

ISMB 2008

Open Access Articles

Bioinformatics Volume 24 #13 2008

Toronto, Canada
19-23 July 2008

ISBN: 978-1-61567-371-1

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© (2008) by Oxford University Press
All rights reserved.

Printed by Curran Associates, Inc. (2009)

For permission requests, please contact Oxford University Press
at the address below.

Oxford University Press
2001 Evans Road
Cary, North Carolina 27513

Phone: (919) 677-0977
Fax: (919) 677-1714

jnlorders@oup-usa.org

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
Web: www.proceedings.com

TABLE OF CONTENTS

Editorial	1
<i>A. Valencia, I. Cases</i>	
ISMB 2008 Organization	2
<i>N/A</i>	

ORIGINAL PAPERS

Sequence Analysis and Alignment

POIMs: Positional Oligomer Importance Matrices — Understanding Support Vector Machine-based Signal Detectors	6
<i>Soren Sonnenburg, Alexander Zien, Petra Philips, Gunnar Ratsch</i>	
The Effectiveness of Position - and Composition-specific Gap Costs for Protein Similarity Searches	15
<i>Aleksandar Stojmirovic, E. Michael Gertz, Stephen F. Altschul, Yi-Kuo Yu</i>	
ProSOM: Core Promoter Prediction Based on Unsupervised Clustering of DNA Physical Profiles	24
<i>Thomas Abeel, Yvan Saeys, Pierre Rouze, Yves Van de Peer</i>	
Optimal Pooling for Genome Re-sequencing with Ultra-high-throughput Short-read Technologies	32
<i>Iman Hajirasouliha, Fereydoun Hormozdiari, S. Cenk Sahinalp, Inanc Birol</i>	
Efficient Algorithms for Accurate Hierarchical Clustering of Huge Datasets: Tackling the Entire Protein Space	41
<i>Yaniv Loewenstein, Elon Portugaly, Menachem Fromer, Michal Linial</i>	
MicroRNA Prediction with a Novel Ranking Algorithm Based on Random Walks	50
<i>Yunpen Xu, Xuefeng Zhou, Weixiong Zhang</i>	
A Robust Framework for Detecting Structural Variations in a Genome	59
<i>Seunghak Lee, Elango Cheran, Michael Brudno</i>	
A Max-margin Model for Efficient Simultaneous Alignment and Folding of RNA Sequences	68
<i>Chuong B. Do, Chuan-Sheng Foo, Serafim Batzoglou</i>	

Comparative Genomics

MACHOS: Markov Clusters of Homologous Subsequences	77
<i>Simon Wong, Mark A. Ragan</i>	
Classification and Feature Selection Algorithms for Multi-class CGH Data	86
<i>Jun Liu, Sanjay Ranka, Tamer Kahveci</i>	
Guided Genome Halving: Hardness, Heuristics and the History of the Hemiascomycetes	96
<i>Chunfang Zheng, Qian Zhu, Zaky Adam, David Sankoff</i>	
A Model of Higher Accuracy for the Individual Haplotyping Problem Based on Weighted SNP Fragments and Genotype with Errors	105
<i>Minzhu Xie, Jianxin Wang, Jianer Chen</i>	

Evolution and Phylogeny

Estimating True Evolutionary Distances Under the DCJ Model	114
<i>Yu Lin, Bernard M.E. Moret</i>	
Efficient Inference of Bacterial Strain Trees from Genome-scale Multilocus Data	123
<i>C. Than, R. Sugino, H. Innan, L. Nakhleh</i>	
The Multiple Gene Duplication Problem Revisited	132
<i>Mukul S. Bansal, Oliver Eulenstein</i>	

Gene Regulation and Transcriptomics

Nucleosome Positioning from Tiling Microarray Data	139
<i>Moran Yassour, Tommy Kaplan, Ariel Jaimovich, Nir Friedman</i>	

Alignment and Classification of Time Series Gene Expression in Clinical Studies	147
<i>Tien-ho Lin, Naftali Kaminski, Ziv Bar-Joseph</i>	
Inferring Differentiation Pathways from Gene Expression	156
<i>Ivan G. Costa, Stefan Roepcke, Christoph Hafemeister, Alexander Schliep</i>	
Predicting Functional Transcription Factor Binding Through Alignment-free and Affinity-based Analysis of Orthologous Promoter Sequences	165
<i>Lucas D. Ward, Harmen J. Bussemaker</i>	
Combinatorial Influence of Environmental Parameters on Transcription Factor Activity	172
<i>T.A. Knijnenburg, L.F.A.Wessels, M.J.T. Reinders</i>	

Protein Structure and Function

Designing Succinct Structural Alphabets	182
<i>Shuai Cheng Li, Dongbo Bu, Xin Gao, Jinbo Xu, Ming Li</i>	
Predicting Protein Thermostability Changes from Sequence Upon Multiple Mutations	190
<i>Ludovica Montanucci, Piero Fariselli, Pier Luigi Martelli, Rita Casadio</i>	
Algorithm for Backrub Motions in Protein Design	196
<i>Ivelin Georgiev, Daniel Keedy, Jane S. Richardson, David C. Richardson, Bruce R. Donald</i>	
Contact Replacement for NMR Resonance Assignment	205
<i>Fei Xiong, Gopal Pandurangan, Chris Bailey-Kellogg</i>	
A Computational Framework to Empower Probabilistic Protein Design	214
<i>Menachem Fromer, Chen Yanover</i>	

Protein Interactions and Molecular Networks

Identifying Functional Modules in Protein–protein Interaction Networks: An Integrated Exact Approach	223
<i>Marcus T. Dittrich, Gunnar W. Klau, Andreas Rosenwald, Thomas Dandekar, Tobias Muller</i>	
Prediction of Drug–target Interaction Networks from the Integration of Chemical and Genomic Spaces	232
<i>Yoshihiro Yamanishi, Michihiro Araki, Alex Gutteridge, Wataru Honda, Minoru Kanehisa</i>	
Biomolecular Network Motif Counting and Discovery by Color Coding	241
<i>Noga Alon, Phuong Dao, Iman Hajirasouliha, Fereydoun Hormozdiari, S. Cenk Sahinalp</i>	
Protein Complex Identification by Supervised Graph Local Clustering	250
<i>Yanjun Qi, Fernanda Balem, Christos Faloutsos, Judith Klein-Seetharaman, Ziv Bar-Joseph</i>	

Text Mining

Annotation-based Inference of Transporter Function	259
<i>Thomas J. Lee, Ian Paulsen, Peter Karp</i>	
Detection of IUPAC and IUPAC - Like Chemical Names	268
<i>Roman Klinger, Corinna Kolarik, Juliane Fluck, Martin Hofmann-Apitius, Christoph M. Friedrich</i>	
Identifying Gene-disease Associations Using Centrality on a Literature Mined Gene-interaction Network	277
<i>Arzucan Ozgur, Thuy Vu, Gunes Erkan, Dragomir R. Radev</i>	
Integrating High Dimensional Bi-directional Parsing Models for Gene Mention Tagging	286
<i>Chun-Nan Hsu, Yu-Ming Chang, Cheng-Ju Kuo, Yu-Shi Lin, Han-Shen Huang, I-Fang Chung</i>	

Databases and Ontologies

The EXACT Description of Biomedical Protocols	295
<i>Larisa N. Soldatova, Wayne Aubrey, Ross D. King, Amanda Clare</i>	
Towards the Use of Argumentation in Bioinformatics: A Gene Expression Case Study	304
<i>Kenneth McLeod, Albert Burger</i>	
The Ontology of Biological Taxa	313
<i>Stefan Schulz, Holger Stenzhorn, Martin Boeker</i>	

GenoQuery: A New Querying Module for Functional Annotation in a Genomic Warehouse	322
<i>Frederic Lemoine, Bernard Labedan, Christine Froidevaux</i>	

Other Bioinformatics Applications and Methods

Assessing the Functional Structure of Genomic Data	330
<i>C. Huttenhower, O.G. Troyanskaya</i>	
Matching Isotopic Distributions from Metabolically Labeled Samples	339
<i>Sean McIlwain, David Page, Edward L. Huttlin, Michael R. Sussman</i>	
Modeling Peptide Fragmentation with Dynamic Bayesian Networks for Peptide Identification	348
<i>Aaron A. Klammer, Sheila M. Reynolds, Jeff A. Bilmes, Michael J. MacCoss, William Stafford Noble</i>	
BLASTING Small Molecules—Statistics and Extreme Statistics of Chemical Similarity Scores	357
<i>Pierre Baldi, Ryan W. Benz</i>	
A Maximum Common Substructure-based Algorithm for Searching and Predicting Drug-like Compounds	366
<i>Yiqun Cao, Tao Jiang, Thomas Girke</i>	
Classification of arrayCGH Data Using Fused SVM	375
<i>Franck Rapaport, Emmanuel Barillot, Jean-Philippe Vert</i>	

Bioinformatics of Disease

Assessment and Improvement of the Plasmodium Yoelii Yoelii Genome Annotation Through Comparative Analysis	383
<i>Ashley Vaughan, Sum-Ying Chiu, Gowthaman Ramasamy, Ling Li, Malcolm J. Gardner, Alice S. Tarun, Stefan H.I. Kappe, Xinxia Peng</i>	
Differential Variability Analysis of Gene Expression and Its Application to Human Diseases	390
<i>Joshua W.K. Ho, Maurizio Stefani, Cristobal G. dos Remedios, Michael A. Charleston</i>	
Selecting Anti-HIV Therapies Based on a Variety of Genomic and Clinical Factors	399
<i>Michal Rosen-Zvi, Andre Altmann, Mattia Prosperi, Ehud Aharoni, Hani Neuvirth, Anders Sonnerborg, Eugen Schuler, Daniel Struck, Yardena Peres, Francesca Incardona, Rolf Kaiser, Maurizio Zazzi, Thomas Lengauer</i>	
Reversible Jump MCMC Approach for Peak Identification for Stroke SELDI Mass Spectrometry Using Mixture Model	407
<i>Yuan Wang, Xiaobo Zhou, Honghui Wang, King Li, Lixiu Yao, Stephen T.C. Wong</i>	

Other Bioinformatics Applications and Methods

Multi-spectra Peptide Sequencing and Its Applications to Multistage Mass Spectrometry	414
<i>Nuno Bandeira, Jesper V. Olsen, Matthias Mann, Pavel A. Pevzner</i>	

Author Index