ISMB/ECCB 2011
Open Access Articles

Bioinformatics Volume 27 #13

Vienna, Austria
17-19 July 2011

Editors:

Alex Bateman **Alfonso Valencia**

ISBN: 978-1-61839-142-1
ISSN: 1367-4803

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© (2011) by Oxford University Press
All rights reserved.

Printed by Curran Associates, Inc. (2011)

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

PROTEIN STRUCTURE AND FUNCTION

Sorting the Nuclear Proteome 1
D. C. Bauer, K. Willadsen, F. A. Buske, K.-A. Lê Cao, T. L. Bailey, G. Dellaire, M. Bodén

Environment Specific Substitution Tables Improve Membrane Protein Alignment 10
J. R. Hill, S. Kelm, J. Shi, C. M. Deane

Sequence-based Prediction of Protein Crystallization, Purification and Production Propensity 19
M. J. Mizianty, L. Kurgan

A Method for Probing the Mutational Landscape of Amyloid Structure 30
C. W. O'Donnell, J. Waldispühl, M. Lis, R. Halfmann, S. Devadas, S. Lindquist, B. Berger

Piecewise Linear Approximation of Protein Structures Using the Principle of Minimum Message Length 40
A. S. Konagurthu, L. Allison, P. J. Stuckey, A. M. Lesk

QAARM: Quasi-anharmonic Autoregressive Model Reveals Molecular Recognition Pathways in Ubiquitin 50
A. J. Savol, V. M. Burger, P. K. Agarwal, A. Ramanathan, C. S. Chennubhotla

Multi-view Methods for Protein Structure Comparison Using Latent Dirichlet Allocation 59
S. Shivashankar, S. Srivathsan, B. Ravindran, A. V. Tendulkar

Template-free Detection of Macromolecular Complexes in Cryo Electron Tomograms 67
M. Xu, M. Beck, F. Alber

SEQUENCE ANALYSIS

vipR: Variant Identification in Pooled DNA Using R 76
A. Altmann, P. Weber, C. Quast, M. Rex-Haffner, E. B. Binder, B. Müller-Myhsok

IPknot: Fast and Accurate Prediction of RNA Secondary Structures with Pseudoknots Using Integer Programming 84
K. Sato, Y. Kato, M. Hamada, T. Akutsu, K. Asai

Meta-IDBA: A De Novo Assembler for Metagenomic Data 94
Y. Peng, H. C. M. Leung, S. M. Yiu, F. Y. L. Chin

A Conditional Random Fields Method for RNA Sequence–structure Relationship Modeling and Conformation Sampling 102
Z. Wang, J. Xu

TEXT MINING

Discovering and Visualizing Indirect Associations Between Biomedical Concepts 112
Y. Tsuruoka, M. Miwa, K. Hamamoto, J. Tsujii, S. Ananiadou

MeSH: A Window Into Full Text for Document Summarization 121
S. Bhattacharya, V. Ha-Thuc, P. Srinivasan

SEQUENCE ANALYSIS

A Folding Algorithm for Extended RNA Secondary Structures 130
C. Höner zu Siederdissen, S. H. Bernhart, P. F. Stadler, I. L. Hofacker

Error Correction of High-throughput Sequencing Datasets with Non-uniform Coverage 138
P. Medvedev, E. Scott, B. Kakaradov, P. Pevzner

PROTEIN INTERACTIONS AND MOLECULAR NETWORKS

Generative Probabilistic Models for Protein–Protein Interaction Networks—The Biclique Perspective 143
R. Schweiger, M. Linial, N. Linial

RINQ: Reference-based Indexing for Network Queries 151
G. Gülsoy, T. Kahveci
Construction of Co-complex Score Matrix for Protein Complex Prediction from AP-MS Data 162
Z. Xie, C. K. Kwoh, X.-L. Li, M. Wu
Uncover Disease Genes by Maximizing Information Flow in the Phenome–interactome Network 171
Y. Chen, T. Jiang, R. Jiang
Physical Module Networks: An Integrative Approach for Reconstructing Transcription Regulation 181
N. Novershtern, A. Regev, N. Friedman
Identification of Metabolic Network Models from Incomplete High-throughput Datasets 190
S. Berthoumieux, M. Brilli, H. De Jong, D. Kahn, E. Cinquemani
TREEGL: Reverse Engineering Tree-evolving Gene Networks Underlying Developing Biological Lineages 201
A. P. Parikh, W. Wu, R. E. Curtis, E. P. Xing
Optimally Discriminative Subnetwork Markers Predict Response to Chemotherapy 210
P. Dao, K. Wang, C. Collins, M. Ester, A. Lapuk, S. C. Sahinalp

DISEASE MODELS AND EPIDEMIOLOGY

Epistasis Detection on Quantitative Phenotypes by Exhaustive Enumeration Using GPUs 219
T. Kam-Thong, B. Pütz, N. Karbalai, B. Müller-Myhsok, K. Borgwardt
Detecting Epistatic Effects in Association Studies at a Genomic Level Based on an Ensemble Approach 227
J. Li, B. Horstman, Y. Chen

MASS SPECTROMETRY AND PROTEOMICS

Efficient Spatial Segmentation of Large Imaging Mass Spectrometry Datasets with Spatially Aware Clustering 235
T. Alexandrov, J. H. Kobarg

BIOIMAGING

Automatic 3D Neuron Tracing Using All-path Pruning 244
H. Peng, F. Long, G. Myers

EVOLUTION AND COMPARATIVE GENOMICS

Tanglegrams for Rooted Phylogenetic Trees and Networks 253
C. Scornavacca, F. Zickmann, D. H. Huson
Mapping Ancestral Genomes with Massive Gene Loss: A Matrix Sandwich Problem 262
H. Gavranović, C. Chauve, J. Salse, E. Tannier
Predicting Site-specific Human Selective Pressure Using Evolutionary Signatures 271
J. Sadri, A. B. Diallo, M. Blanchette
PhyloCSF: A Comparative Genomics Method to Distinguish Protein Coding and Non-coding Regions 280
M. F. Lin, I. Jungreis, M. Kellis

APPLIED BIOINFORMATICS

The Role of Proteosome-mediated Proteolysis in Modulating Potentially Harmful Transcription Factor Activity in Saccharomyces Cerevisiae 288
N. Bonzanni, N. Zhang, S. G. Oliver, J. Fisher
Mixed-model Coexpression: Calculating Gene Coexpression While Accounting for Expression Heterogeneity 294
N. A. Furlotte, H. M. Kang, C. Ye, E. Eskin
A Generalized Model for Multi-marker Analysis of Cell Cycle Progression in Synchrony Experiments 301
M. B. Mayhew, J. W. Robinson, B. Jung, S. B. Haase, A. J. Hartemink
Systematic Exploration of Error Sources in Pyrosequencing Flowgram Data 310
S. Balzer, K. Malde, I. Jonassen

An Enhanced Petri-net Model to Predict Synergistic Effects of Pairwise Drug Combinations from Gene Microarray Data 316
G. Jin, H. Zhao, X. Zhou, S. T. C. Wong

POPULATION GENOMICS

Accurate Estimation of Heritability in Genome Wide Studies Using Random Effects Models 323
D. Golan, S. Rosset
StructHDP: Automatic Inference of Number of Clusters and Population Structure from Admixed Genotype Data 330
S. Shringarpure, D. Won, E. P. Xing
Reconstruction of Genealogical Relationships with Applications to Phase III of HapMap 339
S. Kyriazopoulou-Panagiotopoulou, D. K. Haghighi, S. J. Aerni, A. Sundquist, S. Bercovici, S. Batzoglou

DATABASES AND ONTOLOGIES

ccSVM: Correcting Support Vector Machines for Confounding Factors in Biological Data Classification 348
L. Li, B. Rakitsch, K. Borgwardt
Ontology Patterns for Tabular Representations of Biomedical Knowledge on Neglected Tropical Diseases 355
F. Santana, D. Schober, Z. Medeiros, F. Freitas, S. Schulz

GENE REGULATION AND TRANSCRIPTOMICS

Detection and Interpretation of Metabolite–transcript Coresponses Using Combined Profiling Data 363
H. Redestig, I. G. Costa
From Sets to Graphs: Towards a Realistic Enrichment Analysis of Transcriptomic Systems 373
L. Geistlinger, G. Csaba, R. Küffner, N. Mulder, R. Zimmer
Small Sets of Interacting Proteins Suggest Functional Linkage Mechanisms Via Bayesian Analogical Reasoning 381
E. M. Airoldi, K. A. Heller, R. Silva
Characterization and Improvement of RNA-Seq Precision in Quantitative Transcript Expression Profiling 390
P. P. Labaj, G. G. Leparc, B. E. Linggi, L. M. Markillie, H. S. Wiley, D. P. Kreil
An Integrative Clustering and Modeling Algorithm for Dynamical Gene Expression Data 400
J. Sivriver, N. Habib, N. Friedman
A Novel Computational Framework for Simultaneous Integration of Multiple Types of Genomic Data to Identify MicroRNA-gene Regulatory Modules 409
S. Zhang, Q. Li, J. Liu, X. J. Zhou
Author Index