Oxford University Press

14th Annual International Conference on Intelligent Systems for Molecular Biology

ISMB 2006

August 6 - 10, 2006
Fortaleza, Brazil

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571
www.proceedings.com

ISBN: 978-1-60423-006-2

Some format issues inherent in the e-media version may also appear in this print version.

Copyright (2006) by Oxford Journals

All rights reserved.

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

Oxford Journals
Great Clarendon Street
Oxford OX2 6DP United Kingdom
Phone: +44 (0)1865 353695
Fax: +44 (0)1865 353835
www.oxfordjournals.org

BIOINFORMATICS CONTENTS

EDITORIAL

ISMB 2006 e1
P.E.Bourne and S.Brunak

ISMB 2006 Organization e3

ORIGINAL PAPERS

Automatic clustering of orthologs and inparalogs shared by multiple proteomes e9
A.Alexeyenko, I.Tamas, G.Liu and E.L.L.Sonnhammer

DynaPred: A structure and sequence based method for the prediction of MHC class I binding peptide sequences and conformations e16
I.Antes, S.W.I.Siu and T.Lengauer

ProfilePSTMM: capturing tree-structure motifs in carbohydrate sugar chains e25
K.F.Aoki-Kinoshita, N.Ueda, H.Mamitsuka and M.Kanehisa

The iRMSD: a local measure of sequence alignment accuracy using structural information e35
F.Armougom, S.Moretti, V.Keduas and C.Notredame

A model-based approach for mining membrane protein crystallization trials e40
S.Asur, P.Raman, M.E.Otey and S.Parthasarathy

Integrating structured biological data by Kernel Maximum Mean Discrepancy e49
K.M.Borgwardt, A.Gretton, M.J.Rasch, H.-P.Kriegel, B.Schölkopf and A.J.Smola

AClAP, Autonomous hierarchical agglomerative Cluster Analysis based protocol to partition conformational datasets e58
G.Bottegoni, W.Rocchia, M.Recanatini and A.Cavalli

A top-level ontology of functions and its application in the Open Biomedical Ontologies e66
P.Burek, R.Hoehndorf, F.Loebe, J.Visagie, H.Herre and J.Kelso

Comparative footprinting of DNA-binding proteins e74
B.Contreras-Moreira and J.Collado-Vides

A probabilistic approach to protein backbone tracing in electron density maps e81
F.DiMaio, J.Shavlik and G.N.Phillips

CONTRAfold: RNA secondary structure prediction without physics-based models e90
C.B.Do, D.A.Woods and S.Batzoglou

springScape: visualisation of microarray and contextual bioinformatic data using spring embedding and an 'information landscape' e99
T.M.D.Ebbels, B.F.Buxton and D.T.Jones

Analysis of sample set enrichment scores: assaying the enrichment of sets of genes for individual samples in genome-wide expression profiles e108
E.Edelman, A.Porrello, J.Guinney, B.Balakumaran, A.Bild,
P.G.Febbo and S.Mukherjee

Dense subgraph computation via stochastic search: application to detect transcriptional modules e117
L.Everett, L.-S.Wang and S.Hannenhalli

Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle e124
A.Fauré, A.Naldi, C.Chaouiya and D.Thieffry

Semi-supervised LC/MS alignment for differential proteomics e132
B.Fischer, J.Grossmann, V.Roth, W.Gruissem, S.Baginsky and J.M.Buhmann

Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE e141
B.C.Foat, A.V.Morozov and H.J.Bussemaker

MotifCut: regulatory motifs finding with maximum density subgraphs e150
E.Fratkin, B.T.Naughton, D.L.Brutlag and S.Batzoglou

Bistable switching and excitable behaviour in the activation of Src at mitosis e158
H.Fuß, W.Dubitzky, S.Downes and M.J.Kurth

Context-specific independence mixture modeling for positional weight matrices e166
B.Georgi and A.Schliep

Improved Pruning algorithms and Divide-and-Conquer strategies for Dead-End Elimination, with application to protein design e174
I.Georgiev, R.H.Lilien and B.R.Donald

Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks e184
O.Gevaert, F.De Smet, D.Timmerman, Y.Moreau and B.De Moor

ZPRED: Predicting the distance to the membrane center for residues in α-helical membrane proteins e191
E.Granseth, H.Viklund and A.Elofsson

Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data e197
J.Hertel and P.F.Stadler

Modelling sequential protein folding under kinetic control e203
F.P.E.Huard, C.M.Deane and G.R.Wood

BNTagger: improved tagging SNP selection using Bayesian networks e211
P.H.Lee and H.Shatkay

Finding the evidence for protein-protein interactions from PubMed abstracts e220
H.Jang, J.Lim, J.-H.Lim, S.-J.Park, K.-C.Lee and S.-H.Park

Learning MHC I—peptide binding e227
N.Jojic, M.Reyes-Gomez, D.Heckerman, C.Kadie and O.Schueler-Furman

Comparative genomics reveals unusually long motifs in mammalian genomes e236
N.C.Jones and P.A.Pevzner

Distance based algorithms for small biomolecule classification and structural similarity search e243
E.Karakoc, A.Cherkasov and S.C.Sahinalp

Rapid knot detection and application to protein structure prediction e252
F.Khatib, M.T.Weirauch and C.A.Rohl

Annotating proteins by mining protein interaction networks e260
M.Kirac, G.Ozsoyoglu and J.Yang

A decompositional approach to parameter estimation in pathway modeling: a case study of the Akt and MAPK pathways and their crosstalk e271
G.Koh, H.F.C.Teong, M.-V.Clément, D.Hsu and P.S.Thiagarajan

Finding novel genes in bacterial communities isolated from the environment e281
L.Krause, N.N.Diaz, D.Bartels, R.A.Edwards, A.Pühler, F.Rohwer, F.Meyer and J.Stoye

A combinatorial pattern discovery approach for the prediction of membrane dipping (re-entrant) loops e290
G.Lasso, J.F.Antoniw and J.G.L.Mullins

Interpreting anonymous DNA samples from mass disasters—probabilistic forensic inference using genetic markers e298
T.Lin, E.W.Myers and E.P.Xing

Peptide sequence tag-based blind identification of post-translational modifications with point process model e307
C.Liu, B.Yan, Y.Song, Y.Xu and L.Cai

Identifying cycling genes by combining sequence homology and expression data e314
Y.Lu, R.Rosenfeld and Z.Bar-Joseph

Quantification of transcription factor expression from Arabidopsis images e323
D.L.Mace, J.-Y.Lee, R.W.Twigg, J.Colinas, P.N.Benfey and U.Ohler

Mutation parameters from DNA sequence data using graph theoretic measures on lineage trees e332
R.Magori-Cohen, Y.Louzoun and S.H.Kleinstein

Indel seeds for homology search e341
D.Mak, Y.Gelfand and G.Benson

Efficient identification of DNA hybridization partners in a sequence database e350
T.P.Mann and W.S.Noble

An experimental metagenome data management and analysis system e359
V.M.Markowitz, N.Ivanova, K.Palaniappan, E.Szeto, F.Korzeniewski, A.Lykidis, I.Anderson, K.Mavrommatis, V.Kunin, H.G.Martin, I.Dubchak, P.Hugenholtz and N.C.Kyrpides

An equilibrium partitioning model connecting gene expression and *cis*-motif content e368
J.Mellor and C.DeLisi

Identification of metabolic units induced by environmental signals e375
J.C.Nacher, J.-M.Schwartz, M.Kanehisa and T.Akutsu

Informative priors based on transcription factor structural class improve *de novo* motif discovery e384
L.Narlikar, R.Gordân, U.Ohler and A.J.Hartemink

Apples to apples: improving the performance of motif finders and their significance analysis in the Twilight Zone e393
P.Ng, N.Nagarajan, N.Jones and U.Keich

Create and assess protein networks through molecular characteristics of individual proteins e402
Y.Ofran, G.Yachdav, E.Mozes, T.Soong, R.Nair and B.Rost

BaCelLo: a balanced subcellular localization predictor e408
A.Pierleoni, P.L.Martelli, P.Fariselli and R.Casadio

Semi-supervised analysis of gene expression profiles for lineage-specific development in the *Caenorhabditis elegans* embryo e417
Y.Qi, P.E.Missiuro, A.Kapoor, C.P.Hunter, T.S.Jaakkola, D.K.Gifford and H.Ge

Decoding non-unique oligonucleotide hybridization experiments of targets related by a phylogenetic tree e424
A.Schliep and S.Rahmann

Integrating copy number polymorphisms into array CGH analysis using a robust HMM e431
S.P.Shah, X.Xuan, R.J.DeLeeuw, M.Khojasteh, W.L.Lam, R.Ng and K.P.Murphy

Relative contributions of structural designability and functional diversity in molecular evolution of duplicates e440
B.E.Shakhnovich

Integrating image data into biomedical text categorization e446
H.Shatkay, N.Chen and D.Blostein

On counting position weight matrix matches in a sequence, with application to discriminative motif finding e454
S.Sinha

An ontology for a Robot Scientist e464
L.N.Soldatova, A.Clare, A.Sparkes and R.D.King

ARTS: accurate recognition of transcription starts in human e472
S.Sonnenburg, A.Zien and G.Rätsch

A computational approach toward label-free protein quantification using predicted peptide detectability e481
H.Tang, R.J.Arnold, P.Alves, Z.Xun, D.E.Clemmer,
M.V.Novotny, J.P.Reilly and P.Radivojac

An integrative approach for causal gene identification and gene regulatory pathway inference e489
Z.Tu, L.Wang, M.N.Arbeitman, T.Chen and F.Sun

Computational inference of the molecular logic for synaptic connectivity in *C. elegans* e497
V.Varadan, D.M.Miller III and D.Anastassiou

Novel Unsupervised Feature Filtering of Biological Data e507
R.Varshavsky, A.Gottlieb, M.Linial and D.Horn

Constructing Near-Perfect Phylogenies with multiple homoplasy events e514
R.V.Satya, A.Mukherjee, G.Alexe, L.Parida and G.Bhanot

SNP Function Portal: a web database for exploring the function implication of SNP alleles e523
P.Wang, M.Dai, W.Xuan, R.C.McEachin, A.U.Jackson,
L.J.Scott, B.Athey, S.J.Watson and F.Meng

Protein classification using ontology classification e530
K.Wolstencroft, P.Lord, L.Tabernero, A.Brass and R.Stevens

Inferring Functional Pathways from Multi-Perturbation Data e539
N.Yosef, A.Kaufman and E.Ruppin

Accessing bioscience images from abstract sentences e547
H.Yu and M.Lee

A sequence-based filtering method for ncRNA identification and its application to searching for riboswitch elements e557
S.Zhang, I.Borovok, Y.Aharonowitz, R.Sharan and V.Bafna

Author Index e567