

CONTENTS

Preface	v
 BEYOND GAP MODELS: RECONSTRUCTING ALIGNMENTS AND PHYLOGENIES UNDER GENOMIC-SCALE EVENTS	
Session Introduction	1
<i>Michael Brudno, Bernard Moret, Randy Linder, and Tandy Warnow</i>	
FRESCO: Flexible Alignment with Rectangle Scoring Schemes	3
<i>A.V. Dalca and M. Brudno</i>	
Local Reliability Measures from Sets of Co-Optimal Multiple Sequence Alignments	15
<i>Giddy Landan and Dan Graur</i>	
The Effect of the Guide Tree on Multiple Sequence Alignments and Subsequent Phylogenetic Analysis	25
<i>S. Nelesen, K. Liu, D. Zhao, C. R. Linder, and T. Warnow</i>	
Sensitivity Analysis for Reversal Distance and Breakpoint Reuse in Genome Rearrangements	37
<i>Amit U. Sinha and Jaroslaw Meller</i>	
 COMPUTATIONAL CHALLENGES IN THE STUDY OF SMALL REGULATORY RNAs	
Session Introduction	49
<i>Doron Betel, Christina Leslie, and Nikolaus Rajewsky</i>	
Comparing Sequence and Expression for Predicting microRNA Targets Using GenMiR3	52
<i>J. C. Huang, B. J. Frey, and Q. D. Morris</i>	

Analysis of MicroRNA-Target Interactions by a Target Structure Based Hybridization Model	64
--	----

Dang Long, Chi Yu Chan, and Ye Ding

A Probabilistic Model for Small RNA Flowgram Matching	75
---	----

Vladimir Vacic, Hailing Jin, Jian-Kang Zhu, and Stefano Lonardi

COMPUTATIONAL TOOLS FOR NEXT-GENERATION SEQUENCING APPLICATIONS

Session Introduction	87
----------------------	----

Francisco M. De La Vega, Gabor T. Marth, and Granger Sutton

TRELLIS+: An Effective Approach for Indexing Genome-Scale Sequences Using Suffix Trees	90
--	----

Benjarath Phoophakdee and Mohammed J. Zaki

Pash 2.0: Scaleable Sequence Anchoring for Next-Generation Sequencing Technologies	102
--	-----

Cristian Coarfa and Aleksandar Milosavljevic

Population Sequencing Using Short Reads: HIV as a Case Study	114
--	-----

Vladimir Jojic, Tomer Hertz, and Nebojsa Jojic

Analysis of Large-Scale Sequencing of Small RNAs	126
--	-----

A. J. Olson, J. Brennecke, A. A. Aravin, G. J. Hannon, and R. Sachidanandam

KNOWLEDGE-DRIVEN ANALYSIS AND DATA INTEGRATION FOR HIGH-THROUGHPUT BIOLOGICAL DATA

Session Introduction	137
----------------------	-----

Michael F. Ochs, John Quackenbush, and Ramana Davuluri

SGDI: System for Genomic Data Integration	141
---	-----

V. J. Carey, J. Gentry, D. Sarkar, R. Gentleman, and S. Ramaswamy

Annotating Pathways of Interaction Networks	153
<i>Jayesh Pandey, Mehmet Koyutürk, Wojciech Szpankowski, and Ananth Grama</i>	
Integrating Microarray and Proteomics Data to Predict the Response of Cetuximab in Patients with Rectal Cancer	166
<i>Anneleen Daeman, Olivier Gevaert, Tijl de Bie, Annelies Debucquoy, Jean-Pascal Machiels, Bart de Moor, and Karin Haustermans</i>	
A Bayesian Framework for Data and Hypotheses Driven Fusion of High Throughput Data: Application to Mouse Organogenesis	178
<i>Madhuchhanda Bhattacharjee, Colin Pritchard, and Peter Nelson</i>	
Gathering the Gold Dust: Methods for Assessing the Aggregate Impact of Small Effect Genes in Genomic Scans	190
<i>Michael A. Province and Ingrid B. Borecki</i>	
Multi-Scale Correlations in Continuous Genomic Data	201
<i>R. E. Thurman, W. S. Noble, and J. A. Stamatoyannopoulos</i>	
Analysis of MALDI-TOF Spectrometry Data for Detection of Glycan Biomarkers	216
<i>Habtom W. Ressom, Rency S. Varghese, Lenka Goldman, Christopher A. Loffredo, Mohammed Abdel-Hamid, Zuzana Kyselova, Yehia Mechref, Milos Novotny, and Radoslav Goldman</i>	

MOLECULAR BIOINFORMATICS FOR DISEASE: PROTEIN INTERACTIONS AND PHENOMICS

Session Introduction	228
<i>Yves A. Lussier, Younghee Lee, Predrag Radivojac, Yanay Ofran, Marco Punta, Atul Butte, and Maricel Kann</i>	
System-Wide Peripheral Biomarker Discovery Using Information Theory	231
<i>Gil Alterovitz, Michael Xiang, Jonathan Liu, Amelia Chang, and Marco F. Ramoni</i>	

Novel Integration of Hospital Electronic Medical Records and Gene Expression Measurements to Identify Genetic Markers of Maturation	243
<i>David P. Chen, Susan C. Weber, Philip S. Constantinou, Todd A. Ferris, Henry J. Lowe, and Atul J. Butte</i>	
Networking Pathways Unveils Association Between Obesity and Non-Insulin Dependent Diabetes Mellitus	255
<i>Haiyan Hu and Xiaoman Li</i>	
Extracting Gene Expression Profiles Common to Colon and Pancreatic Adenocarcinoma Using Simultaneous Nonnegative Matrix Factorization	267
<i>Liviu Badea</i>	
Integration of Microarray and Textual Data Improves the Prognosis Prediction of Breast, Lung and Ovarian Cancer Patients	279
<i>O. Gevaert, S. Van Vooren, and B. De Moor</i>	
Mining Metabolic Networks for Optimal Drug Targets	291
<i>Padmavati Sridhar, Bin Song, Tamer Kahveci, and Sanjay Ranka</i>	
Global Alignment of Multiple Protein Interaction Networks	303
<i>Rohit Singh, Jinbo Xu, and Bonnie Berger</i>	
Predicting DNA Methylation Susceptibility Using CpG Flanking Sequences	315
<i>S. Kim, M. Li, H. Paik, K. Nephew, H. Shi, R. Kramer, D. Xu, and T.-H. Huang</i>	

MULTISCALE MODELING AND SIMULATION SESSION: FROM MOLECULES TO CELLS TO ORGANISMS?

Session Introduction	327
<i>Jung-Chi Liao, Jeff Reinbolt, Roy Kerckhoffs, Anushka Michailova, and Peter Arzberger</i>	
Combining Molecular Dynamics and Machine Learning to Improve Protein Function Recognition	332
<i>Dariya S. Glazer, Randall J. Radmer, and Russ B. Altman</i>	

Prediction of Structure of G-Protein Coupled Receptors and of Bound Ligands with Applications for Drug Design	344
<i>Youyong Li and William A. Goddard III</i>	
Markov Chain Models of Coupled Intracellular Calcium Channels: Kronecker Structured Representations and Benchmark Stationary Distribution Calculations	354
<i>Hilary DeRemigio, Peter Kemper, M. Drew Lamar, and Gregory D. Smith</i>	
Spatially-Compressed Cardiac Myofilament Models Generate Hysteresis that Is Not Found in Real Muscle	366
<i>John Jeremy Rice, Yuhai Tu, Corrado Poggesi, and Pieter P. De Tombe</i>	
Modeling Ventricular Interaction: A Multiscale Approach from Sarcomere Mechanics to Cardiovascular System Hemodynamics	378
<i>Joost Lumens, Tammo Delhaas, Borut Kirn, and Theo Arts</i>	
Sub-Micrometer Anatomical Models of the Sarcolemma of Cardiac Myocytes Based on Confocal Imaging	390
<i>Frank B. Sachse, Eleonora Savio-Galimberti, Joshua I. Goldhaber, and John H. B. Bridge</i>	
Efficient Multiscale Simulation of Circadian Rhythms Using Automated Phase Macromodelling Techniques	402
<i>Shatam Agarwal and Jaijeet Roychowdhury</i>	
Integration of Multi-Scale Biosimulation Models via Light-Weight Semantics	414
<i>John H. Gennari, Maxwell L. Neal, Brian E. Carlson, and Daniel L. Cook</i>	
Comparisons of Protein Family Dynamics	426
<i>A. J. Rader and Joshua T. Harrell</i>	

PROTEIN-NUCLEIC ACID INTERACTIONS: INTEGRATING STRUCTURE, SEQUENCE, AND FUNCTION

Session Introduction	438
<i>Martha L. Bulyk, Alexander J. Hartemink, Ernest Fraenkel, and Yael Mandel-Gutfreund</i>	
Functional Trends in Structural Classes of the DNA Binding Domains of Regulatory Transcription Factors	441
<i>Rachel Patton McCord and Martha L. Bulyk</i>	
Using DNA Duplex Stability Information for Transcription Factor Binding Site Discovery	453
<i>Raluca Gordân and Alexander J. Hartemink</i>	
A Parametric Joint Model of DNA-Protein Binding, Gene Expression and DNA Sequence Data to Detect Target Genes of a Transcription Factor	465
<i>Wei Pan, Peng Wei, and Arkady Khodursky</i>	
An Analysis of Information Content Present in Protein-DNA Interactions	477
<i>Chris Kauffman and George Karypis</i>	
Use of an Evolutionary Model to Provide Evidence for a Wide Heterogeneity of Required Affinities Between Transcription Factors and Their Binding Sites in Yeast	489
<i>Richard W. Lusk and Michael B. Eisen</i>	
Striking Similarities in Diverse Telomerase Proteins Revealed by Combining Structure Prediction and Machine Learning Approaches	501
<i>Jae-Hyung Lee, Michael Hamilton, Colin Gleeson, Cornelia Caragea, Peter Zaback, Jeffrey D. Sander, Xue Li, Feihong Wu, Michael Terribilini, Vasant Honavar, and Drena Dobbs</i>	

TILING MICROARRAY DATA ANALYSIS METHODS AND ALGORITHMS

Session Introduction	513
<i>Srinka Ghosh and Antonio Piccolboni</i>	
CMARRT: A Tool for the Analysis of CHIP-chip Data from Tiling Arrays by Incorporating the Correlation Structure	515
<i>Pei Fen Kuan, Hyonho Chun, and Sündüz Keleş</i>	
Transcript Normalization and Segmentation of Tiling Array Data	527
<i>Georg Zeller, Stefan R. Henz, Sascha Laubinger, Detlef Weigel, and Gunnar Rättsch</i>	
GSE: A Comprehensive Database System for the Representation, Retrieval, and Analysis of Microarray Data	539
<i>Timothy Danford, Alex Rolfe, and David Gifford</i>	

TRANSLATING BIOLOGY: TEXT MINING TOOLS THAT WORK

Session Introduction	551
<i>K. Bretonnel Cohen, Hong Yu, Philip E. Bourne, and Lynette Hirschman</i>	
Assisted Curation: Does Text Mining Really Help?	556
<i>Beatrice Alex, Claire Grover, Barry Haddow, Mijail Kabadjov, Ewan Klein, Michael Matthews, Stuart Roebuck, Richard Tobin, and Xinglong Wang</i>	
Evidence for Showing Gene/Protein Name Suggestions in Bioscience Literature Search Interfaces	568
<i>Anna Divoli, Marti A. Hearst, and Michael A. Wooldridge</i>	
Enabling Integrative Genomic Analysis of High-Impact Human Diseases Through Text Mining	580
<i>Joel Dudley and Atul J. Butte</i>	
Information Needs and the Role of Text Mining in Drug Development	592
<i>Phoebe M. Roberts and William S. Hayes</i>	

EpiLoc: A (Working) Text-Based System for Predicting Protein Subcellular Location	604
<i>Scott Brady and Hagit Shatkay</i>	
Filling the Gaps Between Tools and Users: A Tool Comparator, Using Protein-Protein Interactions as an Example	616
<i>Yoshinobu Kano, Ngan Nguyen, Rune Sætre, Kazuhiro Yoshida, Yusuke Miyao, Yoshimasa Tsuruoka, Yuichiro Matsubayashi, Sophia Ananiadou, and Junichi Tsujii</i>	
Comparing Usability of Matching Techniques for Normalising Biomedical Named Entities	628
<i>Xinglong Wang and Michael Matthews</i>	
Intrinsic Evaluation of Text Mining Tools May Not Predict Performance on Realistic Tasks	640
<i>J. Gregory Caporaso, Nita Deshpande, J. Lynn Fink, Philip E. Bourne, K. Bretonnel Cohen, and Lawrence Hunter</i>	
BANNER: An Executable Survey of Advances in Biomedical Named Entity Recognition	652
<i>Robert Leaman and Graciela Gonzalez</i>	