



# Proceedings

## 2007 IEEE International Conference on Bioinformatics and Biomedicine

2-4 November 2007 • Fremont, California



Los Alamitos, California  
Washington • Tokyo





# Proceedings

## BIBM 2007 Table of Contents

<b>Preface</b> .....	xi
<b>Organizing Committee</b> .....	xii
<b>Program Committee</b> .....	xiii
<b>Sponsors</b> .....	xvii

---

### Keynotes

Combining Simulation and Machine Learning to Recognize Function in 4D..... <i>Russ Biagio Altman</i>	3
Statistical Machine Learning and Computational Biology..... <i>Michael I. Jordan</i>	4
Computer Science as a Lens on the Sciences: The Example of Computational Molecular Biology..... <i>Richard M. Karp</i>	5

### Session 1: Biological Data Mining

Feature Cluster Selection for High-Throughput Data Analysis..... <i>Lei Yu and Hao Li</i>	9
GENMINER: Mining Informative Association Rules from Genomic Data..... <i>Ricardo Martinez, Claude Pasquier, and Nicolas Pasquier</i>	15

EGGS: Extraction of Gene Clusters Using Genome Context Based Sequence Matching Techniques .....	23
<i>Sun Kim, Ankita Bhan, Bharath K. Maryada, Kwangmin Choi, and Yves V. Brun</i>	
A Machine Learning Approach for Identification of Head and Neck Squamous Cell Carcinoma .....	29
<i>Mutlu Mete, Xiaowei Xu, Chun-Yang Fan, and Gal Shafirstein</i>	
Analysis of Protein Protein Dimeric Interfaces .....	35
<i>Feihong Wu, Fadi Towfic, Drena Dobbs, and Vasant Honavar</i>	
A Hybrid Abbreviation Extraction Technique for Biomedical Literature .....	42
<i>Min Song and Illhoi Yoo</i>	
Hypothesis-Driven Specialization of Gene Expression Association Rules .....	48
<i>Dharmesh Thakkar, Carolina Ruiz, and Elizabeth F. Ryder</i>	
A New Fuzzy ARTMAP Approach for Predicting Biological Activity of Potential HIV-1 Protease Inhibitors .....	56
<i>Răzvan Andonie, Levente Fabry-Asztalos, Lukas Magill, and Sarah Abdul-Wahid</i>	
Multi-topic Aspects in Clinical Text Classification.....	62
<i>Yutaka Sasaki, Brian Rea, and Sophia Ananiadou</i>	

## **Session 2: High Performance Bio-computing**

A Comparison of Unsupervised Dimension Reduction Algorithms for Classification .....	71
<i>Jaegul Choo, Hyunsoo Kim, Haesun Park, and Hongyuan Zha</i>	
Pathological Image Analysis Using the GPU: Stroma Classification for Neuroblastoma .....	78
<i>Antonio Ruiz, Olcay Sertel, Manuel Ujaldon, Umit Catalyurek, Joel Saltz, and Metin Gurcan</i>	

## **Session 3: Genomics, Comparative Genomics**

MAGMA: An Algorithm for Mining Multi-level Patterns in Genomic Data.....	89
<i>Winnie W. M. Lam, Keith C. C. Chan, David K. Y. Chiu, and Andrew K. C. Wong</i>	
Computational Identification of Protein-Coding Sequences by Comparative Analysis.....	95
<i>Arnaud Fontaine and H�el�ene Touzet</i>	
Multi-agent System for Translation Initiation Site Prediction.....	103
<i>Jia Zeng and Reda Alhajj</i>	

## **Session 4: Biological Databases, Data Integration and Visualization**

An Active Rule-Based Approach for Constraint Management of Biological Data in XML Documents.....	111
<i>Denise Jones, Ying Jin, and Huaqin Xu</i>	
Using a Novel Data Transformation Technique to Provide the EMBOSS Software Suite as Semantic Web Services.....	117
<i>Paul M. K. Gordon, Edward Kawas, Mark Wilkinson, and Christoph W. Sensen</i>	
Exploring and Organizing Spatiotemporal Features such as Waves in High Throughput Brain Recordings by Lifting to Feature Space.....	125
<i>Dragana Veljkovic, Kay A. Robbins, Doug Rubino, and Nicholas G. Hatsopoulos</i>	

## **Session 5: Molecular Sequence Analysis**

A Geometric Representation of Protein Sequences.....	135
<i>Shengyin Gu, Olivier Poch, Bernd Hamann, and Patrice Koehl</i>	
A Divide-and-Conquer Implementation of Three Sequence Alignment and Ancestor Inference.....	143
<i>Feng Yue and Jijun Tang</i>	
SAIL-APPROX: An Efficient On-Line Algorithm for Approximate Pattern Matching with Wildcards and Length Constraints.....	151
<i>Dan He, Xindong Wu, and Xingquan Zhu</i>	
Predicting Markov Chain Order in Genomic Sequences.....	159
<i>Lenwood S. Heath and Amrita Pati</i>	
Micro-repetitive Structure of Genomic Sequences and the Identification of Ancient Repeat Elements.....	165
<i>Abanish Singh, Cedric Feschotte, and Nikola Stojanovic</i>	
The Normalized Similarity Metric and Its Applications.....	172
<i>Shihyen Chen, Bin Ma, and Kaizhong Zhang</i>	

## **Session 6: Healthcare Informatics**

Mining Rules for Risk Factors on Blood Stream Infection in Hospital Information System.....	181
<i>Kimiko Matsuoka, Shigeki Yokoyama, Kunitomo Watanabe, and Shusaku Tsumoto</i>	
A Comparative Medical Informatics Approach to Traumatic Pelvic Injuries.....	188
<i>Soo-Yeon Ji, Toan Huynh, and Kayvan Najarian</i>	
Automatic Identification of Temporal Sequences in Chewing Sounds.....	194
<i>Oliver Amft, Martin Kusserow, and Gerhard Tröster</i>	

Predicting Future High-Cost Patients: A Real-World Risk Modeling Application .....	202
<i>Sai T. Moturu, William G. Johnson, and Huan Liu</i>	
Human Disease-Gene Classification with Integrative Sequence-Based and Topological Features of Protein-Protein Interaction Networks .....	209
<i>Aaron Smalter, Seak Fei Lei, and Xue-wen Chen</i>	

## **Session 7: Proteomics, Computational Proteomics**

Prediction of Kinase-Specific Phosphorylation Sites by One-Class SVMs.....	217
<i>Tingting Li, Hu Fu, and Xuegong Zhang</i>	
Multi-stage Framework to Infer Protein Functional Modules from Mass Spectrometry Pull-Down Data with Assessment of Biological Relevance.....	223
<i>Byung-Hoon Park, Bing Zhang, Tatiana Karpinets, and Nagiza F. Samatova</i>	
Robust Estimation and Graph-Based Meta Clustering for LC-MS Feature Extraction .....	230
<i>Karin Noy and Daniel Fasulo</i>	
A Novel Classification Method for Analysis of Multi-stage Diseases via Mass Spectrometric Data .....	237
<i>Jung Hun Oh, Young Bun Kim, and Jean Gao</i>	

## **Session 8: Protein Structure**

Evaluation of Protein Hydropathy Scales .....	245
<i>Satu Jääskeläinen, Pentti Riikonen, Tapio Salakoski, and Mauno Vihinen</i>	
A Protocol to Detect Local Affinities Involved in Proteins Distant Interactions.....	252
<i>Christophe N. Magnan, Cécile Capponi, and François Denis</i>	
Improved Methods for Template-Matching in Electron-Density Maps Using Spherical Harmonics .....	258
<i>Frank DiMaio, Ameet Soni, George N. Phillips, and Jude W. Shavlik</i>	
Heuristic Strategy for Geometric Hashing Based Protein Structure Comparison of Ellipsoidal Representation.....	266
<i>Yhi Shiau, Jia-Nan Wang, Yu-Feng Huang, and Chien-Kang Huang</i>	

## **Session 9: Computational Systems Biology**

A Haplotyping Algorithm for Non-recombinant Pedigree Data Containing Missing Members .....	275
<i>Doan Dai Duong and Patricia A. Evans</i>	
Risk Factor Searching Heuristics for SNP Case-Control Studies .....	282
<i>Dumitru Brinza and Alexander Zelikovsky</i>	

Constraint-Based Simulation of Biological Systems Described by Molecular Interaction Maps .....	288
<i>Luca Bortolussi, Simone Fonda, and Alberto Policriti</i>	
A Multiscale Model for Efficient Simulation of a Membrane Bound Viral Fusion Peptide .....	294
<i>Yudong Sun, Steve McKeever, Kia Balali-Mood, and Mark S. P. Sansom</i>	
Mathematical Modeling of Phagocyte Chemotaxis toward and Adherence to Biomaterial Implants .....	302
<i>Jiaying Xue, Jean Gao, and Liping Tang</i>	
Modeling the Stochastic Dynamics of Gene Expression in Single Cells: A Birth and Death Markov Chain Analysis .....	308
<i>Samik Ghosh, Preetam Ghosh, Kalyan Basu, and Sajal K. Das</i>	

## **Session 10: Microarray Design and Data Analysis**

A Multi-metric Similarity Based Analysis of Microarray Data .....	317
<i>Fatih Altıparmak, Selnur Erdal, Ozgur Ozturk, and Hakan Ferhatosmanoglu</i>	
Smoothing Spline Mixed Effects Modeling of Multifactorial Gene Expression Profiles .....	325
<i>Brandon J. Smith, Bruce R. Southey, and Sandra L. Rodriguez-Zas</i>	

## **Session 11: BioOntologies**

A New Metric to Measure Gene Product Similarity .....	333
<i>Sachin Mathur and Deendayal Dinakarbandian</i>	
Kinetic Modeling Using BioPAX Ontology .....	339
<i>Oliver Ruebenacker, Ion I. Moraru, James C. Schaff, and Michael L. Blinov</i>	

## **Session 12: Gene Networks, Pathways, Networks, Systems Biology**

A Novel Methodology for Characterizing and Predicting Protein Functional Sites .....	349
<i>Leonardo Bobadilla, Fernando Niño, Edilberto Cepeda, and Manuel A. Patarroyo</i>	
BI-GRAPPIN: Bipartite GRAPh Based Protein-Protein Interaction Networks Similarity Search .....	355
<i>Valeria Fionda, Luigi Palopoli, Simona Panni, and Simona E. Rombo</i>	
Inference of Gene Pathways Using Gaussian Mixture Models .....	362
<i>Younhee Ko, Chengxiang Zhai, and Sandra L. Rodriguez-Zas</i>	
Graph Kernel-Based Learning for Gene Function Prediction from Gene Interaction Network .....	368
<i>Xin Li, Zhu Zhang, Hsinchun Chen, and Jiexun Li</i>	

An Algorithm for Assigning Unique Keys to Metabolic Pathways.....	374
<i>Fang Fang, Robert L. Grossman, and Xiangjun Liu</i>	

### **Session 13: Signal and Image Processing in BioMedicine**

Microarray Image Denoising Using a Two-Stage Multiresolution Technique.....	383
<i>Hara Stefanou, Thanasis Margaritis, Dimitris Kafetzopoulos, Konstantinos Marias, and Panagiotis Tsakalides</i>	

Selection of Tumor Areas and Segmentation of Nuclear Membranes in Tissue Confocal Images: A Fully Automated Approach.....	390
<i>Santa Di Cataldo, Elisa Ficarra, and Enrico Macii</i>	

### **Session 14: BioMedical Databases and Information Systems, BioMedical Information**

MDDM: A Resource for Differential Methylation in Cancer.....	399
<i>Chetan Mittal, Nevenka Dimitrova, and Vijaykumar Channakeshava</i>	

An Efficient Online Tool to Search Top-N Genes with Similar Biological Functions in Gene Ontology Database.....	406
<i>James Z. Wang, Zhidian Du, Philip S. Yu, and Chin-Fu Chen</i>	

Using SVG to Model and Query Image Annotations and Their History.....	412
<i>Fusheng Wang, Cornelius Rabsch, and Peiya Liu</i>	

### **Session 15: Intelligent Biomedical Knowledge Discovery**

A Semi-supervised Learning Approach to Disease Gene Prediction.....	423
<i>Thanh Phuong Nguyen and Tu Bao Ho</i>	

Mining Clinical Data with a Temporal Dimension: A Case Study.....	429
<i>Michele Berlingerio, Francesco Bonchi, Fosca Giannotti, and Franco Turini</i>	

Multiple Interacting Subcellular Structure Tracking by Sequential Monte Carlo Method.....	437
<i>Quan Wen, Jean Gao, and Kate Luby-Phelps</i>	

<b>Author Index</b> .....	443
---------------------------	-----