About Georgia State University

Georgia State University, founded in 1913, is a Carnegie Doctoral/Research Extensive university. Located in the heart of downtown Atlanta, this major research university has an enrollment of more than 27,000 undergraduate and graduate students in six colleges. Georgia State is the second largest university in the state, with students coming from every county in Georgia, every state in the nation, and from over 145 countries. Georgia State is currently embarking on a record $1 billion campus expansion to provide space for the estimated 50,000 students who will attend the university by 2015.

About the Department of Computer Science

The Department of Computer Science offers programs leading to the B.S., M.S., and Ph.D. degrees in computer science. The Ph.D. program, which first admitted students in 2001, has produced 17 graduates so far and has a current enrollment of about 70.

Active research areas in the department include bioinformatics, graphics and visualization, networks and distributed systems, and algorithms. The department emphasizes interdisciplinary research. Current projects involve collaborations with the Departments of Biology, Chemistry, Physics and Astronomy, and Mathematics and Statistics, among others. In addition to running its own bioinformatics program, the Department of Computer Science participates in Georgia State’s Molecular Basis of Disease and the Brains and Behavior programs.

Many department faculty are leading experts in their fields; they serve on panels and study sections at agencies such as NIH and NSF, on the editorial board of journals, and as officers in professional societies. The department is also home to the editorial offices of four book series and two journals. Faculty members are supported by many grants from NSF, NIH, and other agencies. The department’s annual research budget exceeds one million dollars.

The department’s distinguished faculty includes department chair Dr. Yi Pan, the editor-in-chief of International Journal of Bioinformatics Research and Applications; Dr. Scott Owen, who is serving a three-year term as the president of ACM SIGGRAPH; Dr. Robert Harrison, who was named a Distinguished Cancer Scientist by the Georgia Cancer Coalition; and Dr. Yingshu Li, who recently received the prestigious NSF CAREER award.
Message from the General Co-Chairs

Welcome to Atlanta, Georgia – the big city with lots of Southern hospitality – and the 2007 International Symposium on Bioinformatics Research and Applications (ISBRA 2007). We are honored to serve the international community by bringing together researchers, scientists, and students from academia, laboratories, and industry to this premier meeting in Atlanta, the home of 13 Fortune 500 and 24 Fortune 1000 companies.

The technical program assembled for this symposium is of very high quality. We thank Program Chairs Ion Mandoiu and Alexander Zelikovsky for their efforts in putting together this incredible collection of papers. Not only do we recognize the program committee and reviewers for their hard work, but we also thank each author who submitted a paper whether or not it was selected for presentation this week. Due to the very competitive nature of the selections this year, many strong papers could not be accommodated in the final program. We hope that you will enjoy the paper and keynote presentations this week.

We are especially grateful to the Organizing Chairs, Robert Harrison and Yanqing Zhang, who have overseen many aspects of the conference, and who always help out before we even know to ask. The Publication Chair, Raj Sunderraman, has done an excellent job of creating the submission system and collecting and verifying the camera-ready manuscripts for the proceedings. As General Co-Chairs, we thank the Finance Chair, Anu Bourgeois, for her excellent job in creating the bank account and the payment system for our symposium. We thank the Publicity Chairs, Kim King and Yingshu Li, for their assistance in publicizing the conference. Finally, we thank the Poster Chairs, Gulsah Altun and Dumitru Brinza, for their role in collecting many poster papers and organizing the poster sessions.

We recognize Georgia State University for hosting ISBRA 2007, and thank our host coordinator, Tammie Dudley, for assisting the General Co-Chairs and other organizers during the preparations for this year’s conference and for organizing the staff and student volunteers. Thanks also go to Georgia State’s Molecular Basis of Disease Program for its financial support. Enjoy the symposium!

Dan Gusfield, University of California, Davis, USA
Yi Pan, Georgia State University, USA
On behalf of the Program Committee, we would like to welcome you to the 2007 International Symposium on Bioinformatics Research and Applications (ISBRA 2007). The ISBRA symposium is the successor to the International Workshop on Bioinformatics Research and Applications (IWBRA), held on May 22-25, 2005 in Atlanta, GA, and on May 28-31, 2006 in Reading, UK, in conjunction with the International Conference on Computational Science.

This year, 146 papers were submitted in response to the call for papers. Following a rigorous review process, the program committee has selected 55 papers for publication in the proceedings and oral presentations at the symposium. The topics include clustering and classification, gene expression analysis, gene networks, genome analysis, motif finding, pathways, protein structure and interaction, phylogenetics, and software tools.

ISBRA 2007’s technical program also includes several tutorials and poster sessions, and features invited keynote talks by three distinguished speakers – Ming Li from the University of Waterloo, Laura L. Elnitski from the National Human Genome Research Institute, and Mark Borodovsky from the Georgia Institute of Technology.

We would like to thank all authors for submitting papers and presenting their work at the symposium. We would also like to thank the Program Committee members and external reviewers for volunteering their time and expertise to review and select symposium papers. We extend special thanks to the Organizing, Publications, Finance, Publicity, and Poster Chairs, all of whom are listed on the following pages, for their tremendous efforts in making ISBRA 2007 a great success. Last but not least, we would like to thank the General Chairs, Dan Gusfield and Yi Pan, for their leadership and guidance.

We hope you will find the technical program interesting and thought provoking and that attending ISBRA 2007 will provide you with stimulating ideas and ample opportunities to meet other researchers from around the world. Enjoy!

Ion Mandoiu, University of Connecticut, USA
Alexander Zelikovsky, Georgia State University, USA
Keynote Speakers

**Ming Li**
Canada Research Chair in Bioinformatics
Professor, David R. Cheriton School of Computer Science
University of Waterloo
http://www.cs.uwaterloo.ca/~mli/

**Laura L. Elnitski, Ph.D.**
Head, Genomic Functional Analysis Section
Genome Technology Branch
National Human Genome Research Institute
National Institutes of Health
http://www.genome.gov/12514761

**Mark Borodovsky,**
Regents' Professor
Director Center for the Bioinformatics and Computational Genomics
School of Biology, Georgia Institute of Technology
http://wwwbiology.gatech.edu/faculty/mark-borodovsky/
Symposium Organizers

General Chairs
Dan Gusfield, University of California, Davis, USA
Yi Pan, Georgia State University, USA

Program Chairs
Ion Mandoiu, University of Connecticut, USA
Alexander Zelikovsky, Georgia State University, USA

Organizing Chairs
Robert Harrison, Georgia State University, USA
Yanqing Zhang, Georgia State University, USA

Publication Chair
Raj Sunderraman, Georgia State University, USA

Finance Chair
Anu Bourgeois, Georgia State University, USA
Akshaye Dhawan, ACM Student Representative, Georgia State University, USA

Publicity Chairs
K.N. King, Georgia State University, USA
Yingshu Li, Georgia State University, USA

Poster Chairs
Dumitru Brinza, Georgia State University, USA
Gulsah Altun, Georgia State University, USA
Committee Members

Gabriela Alexe
IBM Research, USA

Yonatan Aumann
Bar Ilan University, Israel

Danny Barash
Ben-Gurion University, Israel

Anne Bergeron
Université du Québec à Montréal, Canada

Piotr Berman
Penn State, USA

Paola Bonizzoni
Università degli Studi di Milano-Bicocca, Italy

Mark Borodovsky
Georgia Institute of Technology, USA

Daniel Brown
University of Waterloo, Canada

Liming Cai
University of Georgia, USA

Jake Yue Chen
IUPUI, USA

Luonan Chen
Osaka Sangyo University, Japan

Yixin Chen
Washington University, USA

Peter Damaschke
Chalmers University, Sweden

Sorin Draghici
Wayne State University, USA

Bhaskar Dasgupta
University of Illinois at Chicago, USA

Jun-tao Guo
University of Georgia, USA

Hae-Jin Hu
Georgia State University, USA

Jieyue He
Southeast University, China

Tony Hu
Drexel University, USA

Chun-Hsi Huang
University of Connecticut, USA

Ming-Yang Kao
Northwestern University, USA

Marek Karpinski
University of Bonn, Germany

John Kececioglu
University of Arizona, USA

Ed Keedwell
University of Exeter, United Kingdom

Guojun Li
University of Georgia, USA

Jianzhong Li
Harbin Institute of Technology, China

Yiming Li
National Chiao Tung University, Taiwan

Yixue Li
Shanghai Center for Bioinformation Technology, China

Guohui Lin
University of Alberta, Canada

Xiaohui Liu
Brunel University, United Kingdom

Shiyong Lu
Wayne State University, USA

Jingchu Luo
Peking University, China

Osamu Maruyama
Kyushu University, Japan

Kayvan Najarian
University of North Carolina at Charlotte, USA

Giri Narasimhan
Florida International University, USA

Craig Nelson
University of Connecticut, USA

Laxmi Parida
IBM T.J. Watson Research Center, USA

Mihai Pop
University of Maryland, USA

Alex Pothen
Old Dominion University, USA

Teresa Przytycka
NCBI, USA

Sven Rahmann
Universität Bielefeld, Germany

Sanguthevar Rajasekaran
University of Connecticut, USA

David Sankoff
University of Ottawa, Canada

Russell Schwartz
Carnegie Mellon University, USA

Hagit Shatkay
Queen's University, Canada

Jens Stoye
Universität Bielefeld, Germany

Sing-Hoi Sze
Texas A&M University, USA

El-Ghaizali Talbi
Université des Sciences et Technologies de Lille, France

Esko Ukkonen
University of Helsinki, Finland

Ugo Vaccaro
Università di Salerno, Italy

Gwenn Volkert
Kent State University, USA

Jianxin Wang
Central South University, China

Zidong Wang
Brunel University, United Kingdom

Limsoon Wong
National University of Singapore, Singapore

Weili Wu
University of Texas at Dallas, USA

Fang Xiang Wu
University of Saskatchewan, Canada

Hongwei Wu
University of Georgia, USA

Dong Xu
University of Missouri, USA

Jack Y. Yang
Harvard University, USA

Mary Qu Yang
National Human Genome Research Institute, U.S. Dept. of Health & Human Services, USA

Kaizhong Zhang
University of West Ontario, Canada

Xuegong Zhang
Tsinghua University, China

Si-qing Zheng
University of Texas at Dallas, USA

Wei-Mou Zheng
Chinese Academy of Sciences, China

Wei Zhong
University of South Carolina, Upstate, USA

Ying Zhu
Georgia State University, USA
Conference Program

Monday, May 7, 2007

5:00 – 7:00 pm Reception at Atlanta Marriott Downtown (Olympic Ballroom)

Tuesday, May 8, 2007

9:00 – 10:30 am Parallel Sessions

1A: Gene expression analysis I (ALC 005)
Chair: Qihua Tan

9:00 am GFBA: A Biclustering Algorithm for Discovering Value-Coherent Biclusters, Xubo Fei, Shiyong Lu, Horia Pop, Lily Liang

9:30 am Significance Analysis of Time-Course Gene Expression Profiles, FangXiang Wu

10:00 am Data-driven Smoothness Enhanced Variance Ratio Test to Unearth Responsive Genes in 0-time Normalized Time-course Microarray Studies, Juntao Li, Jianhua Liu, R. Krishna Murthy Karuturi

10:30 am Coffee Break

11:00 – 12:30 pm Parallel Sessions

2A: Gene expression analysis II (ALC 005)
Chair: Seiya Imoto

11:00 am A Bootstrap Correspondence Analysis for Factorial Microarray Experiments with Replications, Qihua Tan

11:30 am Clustering Algorithms Optimizer: A Framework for Large Datasets, Roy Varshavsky, David Horn, Michal Linial

2B: Phylogenetics and genomic diversity (ALC 002)
Chair: Sanguthevar Rajasekaran

11:00 am Searching for Recombinant Donors in a Phylogenetic Network of Serial Samples, Patricia Buendia, Giri Narasimhan

12:00 pm Ranking Function Based on Higher Order Statistics (RF-HOS) for Two-Sample Microarray Experiments, Jahangheer Shaik, Mohammed Yeasin

12:30 pm Lunch

2:00 – 3:00 pm Plenary Session (ALC 005)
Chair: Yi Pan

Invited Keynote Talk: Modern Homology Search, Ming Li
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Title</th>
<th>Chair/Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00 pm</td>
<td>Coffee Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3B: Software tools (ALC 002) Chair: Lonnie Welch</td>
<td>NEUROgrid: A Toolkit for Generating Parameter-Space Maps using NEURON in a Grid Environment, Robert J Calin-Jageman, Chao Xie, Yi Pan, Art Vandenberg, Paul S. Katz</td>
</tr>
<tr>
<td>3:30 pm</td>
<td></td>
<td>Discovering Relations among GO-annotated Clusters by Graph Kernel Methods, Italo Zoppis, Daniele Merico, Marco Antoniotti, Bud Mishra, Giancarlo Mauri</td>
<td>An Adaptive Resolution Tree Visualization of Large Influenza Virus Sequence Datasets, Leonid Zaslavsky, Yiming Bao, Tatiana Tatusova</td>
</tr>
<tr>
<td>4:00 pm</td>
<td></td>
<td>Discovering Relations among GO-annotated Clusters by Graph Kernel Methods, Italo Zoppis, Daniele Merico, Marco Antoniotti, Bud Mishra, Giancarlo Mauri</td>
<td>Wavelet Image Interpolation (WII): A Wavelet-based Approach to Enhancement of Digital Mammography Images, Jordana Derado, DuBois Bowman, Rajan Patel, Mary Newell, Brani Vidakovic</td>
</tr>
<tr>
<td>4:30 pm</td>
<td></td>
<td>An Empirical Comparison of Dimensionality Reduction Methods for Classifying Gene and Protein Expression Datasets, George Lee, Anant Madabhushi, Carlos Rodriguez</td>
<td>High Level Programming Environment System for Protein Structure Data, Yanchao Wang, Rajshekhar Sunderraman, Piyaphol Phoungphol</td>
</tr>
<tr>
<td>5:00 pm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:30 pm</td>
<td>Coffee Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A2. Pair wise Alignment of Pathways, Qiong Cheng and Alexander Zelikovsky</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A3. Toward a Methodology for Discovery of Regulatory Motifs in Plant, Dazhang Gu, Klaus H. Ecker, Lonnie Welch, and Sarah Wyatt</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A4. Discovering Causal Sentences with Automatically Learned Patterns, Shreekanth Karvaje, Bharat Ravisekar, Baoli Li, and Ashwin Ram</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A5. An Imputation Method Based on Robust Regression Using Minimum Covariance Determinant Estimates for Cancer Gene Expression Data, Hyeonsoo Kim and Haejun Park</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A6. Design Pattern for Protein Identification, Jens Lichtenberg and Lonnie Welch</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A7. Characterizing Pseudobase and Predicting RNA secondary structure with simple H-type pseudoknots, Oyun-Erdene Namsrai</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A8. Mining MEDLINE for Gene Clustering: A Comparison of Feature Selection Approaches, Sailaja Pydimarri, Orlando</td>
<td></td>
</tr>
</tbody>
</table>
million SNP genotypes per experiment, with one million SNP genotypes per experiment expected in the very near future. Furthermore, due to decreasing genotyping costs, future association studies are expected to comprise thousands of typed individuals. Many of the commonly used analysis methods are vastly inadequate for handling datasets of the size envisioned to be produced by the next generation of genome-wide association studies.

In this tutorial we review recent progress on scalable algorithms for genotype and haplotype analysis, including algorithms for haplotype inference, genotype error detection, genotype tagging and indexing, disease association search, and disease susceptibility prediction.
12:30 pm **Lunch**

2:00 – 3:00 pm **Plenary Session (ALC 005)**

   Chair: Jack Yang

   **Invited Keynote Talk:** A Computational Study of Bidirectional Promoters in the Human Genome, Laura L. Elnitski

3:00 pm **Coffee Break**

3:30 – 5:00 pm **Parallel Sessions**

<table>
<thead>
<tr>
<th>3:30 pm</th>
<th><strong>6A: Genome analysis II (ALC 005)</strong></th>
<th><strong>6B: Motif finding III (ALC 002)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The Identification of Antisense Gene Pairs through Available Software, Mark Lawson, Liqing Zhang</td>
<td>Predicting Palmitoylation Sites Using A Regularised Bio-Basis Function Neural Network, Ron Yang</td>
</tr>
</tbody>
</table>

| 4:00 pm | Inferring Weak Adaptations and Selection Biases in Proteins from Composition and Substitution Matrices, Steinar Thorvaldsen, Elinor Ytterstad, Tor Fla | A Novel Kernel-based Approach for Predicting Binding Peptides for HLA Class II Molecules, Hao Yu, Minlie Huang, Xiaoyan Zhu, Yabin Guo |

| 4:30 pm | Markov Model Variants for Appraisal of Coding Potential in Plant DNA, Michael Sparks, Volker Brendel, Karin Dorman | A Database for Prediction of Unique Peptide Motifs as Linear Epitopes, Tun-Wen Pai, Margaret Dah-Tsyr Chang, Wen-Shyong Tzou |

5:00 pm **Coffee Break**

5:15 – 6:45 pm **Poster Session B**

| B1. | CABIN: Collective Analysis of Biological Interaction Network, Mudita Singhal and Kelly Domico |
| B2. | Smoothing spline mixed effects modeling of multi-factorial gene expression profiles, Brandon Smith, Bruce Southey, and Sandra Rodriguez-Zas |
| B3. | DeltaProt: Molecular comparison of proteins based on sequence alignments, Steinar Thorvaldsen, Tor Fla, and Nils P Willassen |
| B4. | Using Logical Sets to Target Gene Expression Patterns, Timothy Tickle and M. Taghi Mostafavi |
| B5. | Database for Structural Analysis of HIV Protease, Yunfeng Tie, Hao Wang, Robert Harrison, and Irene Weber |
| B6. | A integrated solution based on distance method for reconstructing phylogenetic trees, Cristianno Vieira, Glauber Gontcalves, and Martha Torres |
| B8. | Ordered Combinatorial Feature Selection : An Information Portal for Multiple Indexing Sequence Alignment, Hsin-Wei Wang, Jian- | **Tutorial B (ALC 005)**

   **Protein Structural Prediction with Broad Initiatives in Bioinformatics Research and Applications**

   **Jack Yang and Mary Qu Yang**

   **Abstract.** Proteins are composed of one or more chains of amino acids, and exhibit several levels of structure. Many protein regions and some entire proteins lack specific 3-D structures, existing as dynamic, disordered ensembles under physiological conditions. Intrinsically Unstructured regions and disordered Proteins (IUP) affect protein folding pathways and ligand bindings. Recently, IUP are gaining more and more attention in medicinal and pharmaceutical studies, because IUP have been associated with a wide range of protein functions. IUP are also playing central roles in diseases characterized by protein misfolding and aggregation. Knowledge of IUP can help in determination of protein function and effective drug design and discovery.

   Although IUP can be identified by laborious and time-consuming methods such as X-ray crystallography, NMR and CDR, computational methods predicting IUP from the primary structure of a protein...
which are essential for automated structural and functional prediction and annotation of proteins as well as drug design and discovery. Our basic approach consists in applying a hybrid unsupervised-supervised classifier called the Recursive Maximum Contrast Tree (RMCT) classifier to this problem, in combination with novel feature generation, feature selection techniques and ensemble methods. This tutorial will discuss the effectiveness of the approaches to learning protein structural classes, and also provide comparisons to more traditional classifiers such as neural networks and support vector machines.

7:10 pm  
Banquet at Atlanta Marriott Downtown (Centennial Ballroom)

Thursday, May 10, 2007

9:00 – 10:30 am  
Parallel Sessions

7A: Sequence analysis (ALC 005)  
Chair: Mary Qu Yang

9:00 am  
A Novel Greedy Algorithm for the Minimum Common String Partition Problem, Dan He

9:30 am  
An Efficient Algorithm for Finding Gene-specific Probes for DNA Microarrays, Mun-Ho Choi, Seung-Ho Kang, In-Seon Jeong, Hyeong-Seok Lim

10:00 am  
Multiple Sequence Local Alignment Using Monte Carlo EM Algorithm, Chengpeng Bi

10:00 am  
Multiple Sequence Local Alignment Using Monte Carlo EM Algorithm, Chengpeng Bi

7B: Cancer Classification (ALC 002)  
Chair: Wei Zhong

Cancer Class Discovery using Non-negative Matrix Factorization based on Alternating Non-negativity-constrained Least Squares, Hyunsoo Kim, Haesun Park

A Support Vector Machine Ensemble for Cancer Classification using Gene Expression Data, Chen Liao, Shutao Li

10:00 am  
Combining SVM Classifiers Using Genetic Fuzzy Systems based on AUC for Gene Expression Data Analysis, Xiujuan Chen, Yichuan Zhao, Yan-Qing Zhang, Robert Harrison

10:30 am  
Coffee Break

11:00 – 12:30 pm  
Parallel Sessions

8A: RNA and protein structure (ALC 005)  
Chair: Luciano Margara

11:00 am  
A BP-SCFG Based Approach for RNA Secondary Structure Prediction with Consecutive Bases Dependency and Their Relative Positions Information, Dandan Song, Zhidong Deng

11:30 am  
Delta: a Toolset for the Structural Analysis of Biological Sequences on a 3D Triangular Lattice, Minghui Jiang, Martin Mayne, Joel Gillespie

8B: Clustering and Classification (ALC 002)  
Chair: Lawrence Hall

Coclustering Based Parcellation of Human Brain Cortex Using Diffusion Tensor MRI, Cui Lin, Shiyoug Lu, Danqing Wu

An Algorithm for Hierarchical Classification of Genes of Prokaryotic Genomes, Hongwei Wu, Fenglou Mao, Victor Olman, Ying Xu
### 12:00 pm
**Statistical Estimate for the Size of the Protein Structural Vocabulary, Xuezheng Fu, Bernard Chen, Yi Pan, Robert Harrison**  
**Using Multi Level Nearest Neighbor Classifiers for G-protein Coupled Receptor Sub-families Prediction, Mudassir Fayyaz, Adnan Mujahid Khan, Asifullah Khan, Alex Kavokin**

### 12:30 pm
**Lunch**

### 2:00 – 3:00 pm
**Plenary Session (ALC 005)**

Chair: Alex Zelikovsky

**Invited Keynote Talk:** Ab initio Gene Finding Engines: What is Under the Hood, Mark Borodovsky

### 3:00 pm
**Coffee Break**

### 3:30 – 5:00 pm
**Parallel Sessions**

**9A: Protein structure and nucleosome dynamics (ALC 005)**  
Chair: Wei Zhong

- **Reconstruction of 3D Structures From Protein Contact Maps, Marco Vassura, Luciano Margara, Filippo Medri, Pietro di Lena, Piero Fariselli, Rita Casadio**
- **Comparative Analysis of Gene-Coexpression Networks Across Species, Shiquan Wu, Jing Li**

**9B: Gene networks, pathways, and protein domain interactions (ALC 002)**  
Chair: Hongwei Wu

- **A Feature Selection Algorithm based on Graph Theory and Random Forests for Protein Secondary Structure Prediction, Gulsah Altun, Hae-Jin Hu, Stefan Gremalschi, Robert Harrison, Yi Pan**
- **Comparative Pathway Prediction via Unified Graph Modeling of Genomic Structure Information, Jizhen Zhao, Dongsheng Che, Liming Cai**

- **DNA Sites Buried in Nucleosome become Accessible at Room Temperature: A Discrete-Event-Simulation based Modeling Approach, Amin Mazloom, Kalyan Basu, Subhrangsu S. Mandal, Mehran Sorourian, Sajal Das**
- **Extending the Calculus of Looping Sequences to Model Protein Interaction at the Domain Level, Roberto Barbuti, Andrea Maggiolo Schettini, Paolo Milazzo**