

PACIFIC SYMPOSIUM ON BIOCOMPUTING 2005

2005 marks the tenth year of the Pacific Symposium on Biocomputing (PSB). The first meeting was in 1996 and was organized by Teri Klein and Larry Hunter. They brought Keith Dunker and Russ Altman on the team for 1997, and the organizing group has been in place and stable since then. PSB is a unique conference among those in bioinformatics and computational biology: the session topics are proposed each year and selected in a competitive process. They are chosen based on the opportunity to bring together a critical mass of scientists with shared interests in hot or emerging areas. Grass-roots session organizers who make successful proposals then solicit papers and moderate their review. The PSB organizers take pride in the fact that PSB is often the first venue to cluster papers related to a particular theme. With only a few exceptions, all PSB published manuscripts are available online at <http://psb.stanford.edu/psb-online/>. In fact, a look at the sessions over the last ten years is a useful way to understand the evolution of our field. A complete listing of the sessions is presented at the end of this preface. With the publication of this issue of the proceedings, PSB has now published more than 500 manuscripts. They are all in PubMed and can be retrieved using the NIH abbreviation for the proceedings "Pac Symp Biocomput."

What has been the impact of PSB papers? PSB is not indexed by the Thomson ISI service, and so overall impact factor is not available (we are working on this!). However, there are a few ways to gauge our impact. Informally, we have frequently noticed publications in other major scientific journals with references to PSB papers. This is particularly true in emerging areas that PSB "covered" before they hit mainstream, and includes gene expression array analysis, network reconstruction, and text analysis. The Citeseer program (<http://citeseer.ist.psu.edu/>) principally tracks references in computer science, and even then has only partial coverage of all citations. Nonetheless, it provides some measure of impact. The coverage of citations in biology is not good (the Journal of Molecular Biology has only 244 citations!), but in the computer science literature, PSB proceedings have been cited 567 times, and there are 71 papers with two or more citations. The top 10 papers in terms of impact in Citeseer are also provided at the end of this preface. We also have provided a histogram showing the relationship of citation frequency to date of publication.

In the biology literature, one measure of impact of PSB is the number of citations in the online Highwire Press collection of journals (<http://highwire.stanford.edu/>). A search for “Pac Symp Biocomput” in Highwire shows 390 Highwire journal manuscripts referring to PSB papers. Highwire provides an automated indexing of these papers, showing which keyword/topics appear most frequently. The list of topics with greater than five papers is also provided below. The three most common topics of papers citing PSB are “gene expression,” “networks,” and “alignment.” Again, the information from the Highwire analysis is limited because it only contains a fraction of all relevant biological publications.

In any case, it is clear that PSB papers are getting attention and are being cited in the archival literature. The organizers are working on linking the online proceedings to PubMed, a task which is straightforward but requires resources that we have so far been unable to muster. In addition, we are working on convincing the Thomson ISI folks to include PSB in their impact factor ratings.

The PSB organizers would like to thank the National Library of Medicine/National Institutes of Health and the U.S. Department of Energy for support. Applied Biosystems continues to sponsor PSB, and as a result, we are able to provide travel grants to many meeting participants. PSB values its relationship with the International Society for Computational Biology (<http://www.iscb.org/>), and is pleased to offer registration discounts to members.

We look forward to the keynote addresses by David Eisenberg of UCLA and Arti Rai of Duke University. Tiffany Jung has expertly managed the process of assembling the proceedings, in addition to many other organization tasks. Of course, the meeting exists because of the dedicated efforts of the session organizers. These busy scientists moderate a high quality review process that must be completed in less than two months, with an overall accept rate of less than 35%. They are:

Carlos Bustamante, Shamil Sunyaev and Matt Dimmic
Inferring SNP Function Using Evolutionary, Structural and Computational Methods

Alexander Tropsha and Herbert Edelsbrunner
Biogeometry: Applications of Computational Geometry to Molecular Structure

Patsy Babbitt, Philip Bourne, and Sean Mooney
Inferring Function from Structural Genomics Targets

Marylyn D. Ritchie, Michelle W. Carrillo, and Russell Wilke
Computational Approaches for Pharmacogenomics

Alex Hartemink and Eran Segal
Joint Learning from Multiple Types of Genomic Data

Olivier Bodenreider, Joyce Mitchell, and Alexa McCray
Biomedical Ontologies

PSB 2005 will also feature four tutorials: *Use of Multi-locus Data in Gene Mapping Studies* by Jo Knight; *Association Mapping: Design Issues and Data Analysis Approaches* by Jotun Hein and Leif Schausser; *Data Analysis and Sharing with Web Services* by Michael Jensen and Timothy Patrick; and *Function Prediction: From High Throughput to Individual Proteins* by Yanay Ofran, Marco Punta, and Burkard Rost.

We would like to acknowledge the efforts of those who reviewed the submitted manuscripts on a very tight schedule. The partial list that is included after this preface does not include those who wished to remain anonymous. We apologize to others who we may have left off inadvertently.

We look forward to another exciting meeting, and encourage you to consider proposing a new session topic, teaching a new tutorial, or submitting your work for the conference as it enters its second decade.

Aloha!

Pacific Symposium on Biocomputing Co-Chairs

October 1, 2004

Russ B. Altman
Department of Genetics, Stanford University

A. Keith Dunker
Center for Computational Biology & Bioinformatics, Indiana University
School of Medicine

Lawrence Hunter
Department of Pharmacology, University of Colorado Health Sciences Center

Teri E. Klein
Department of Genetics, Stanford University

PSB Session Topics Over Last Ten Years (1996-2005)

1996

- The Evolution of Biomolecular Structures
- Discovering, Learning, Analyzing and Predicting Protein Structure.
- Stochastic Models, Formal Systems and Algorithmic Discovery for Genome Informatics.
- Interactive Molecular Visualization.
- Educational Issues in Biocomputing.
- Internet Tools for Computational Biology.
- Population Modelling.
- Hybrid Quantum and Classical Mechanical Methods for Studying Biopolymers in Solution.
- Control in Biological Systems.

1997

- Distributed and Intelligent Databases
- Modern Concepts in Molecular Modeling,
- Extracting Biological Knowledge from DNA Sequences
- Understanding and Predicting Protein Structure
- Biopolymer Structures: Where Do They Come From? Where Are They Going? Evolutionary Perspectives on Biopolymer Structure and Function
- Computing with Biomolecules
- Computation in Biological Pathways
- Biocomputing Education: Further Challenges

1998

- Gene expression and genetic networks
- Molecules to maps: tools for visualization and interaction
- Gene structure identification in large-scale genomic sequence
- Molecular modeling in drug design and biotechnology
- Protein structure prediction
- The relationships between protein structure and function
- Computing with biomolecules
- Complexity and information theoretic approaches to biology
- Distributed and intelligent databases
- Building bioinformation infrastructure in the Pacific Rim

1999

- Gene expression and genetic networks
- Data mining and knowledge discovery in molecular databases
- Computer modeling in physiology: from cell to tissue
- Information theoretic approaches to biology
- Molecules to maps: tools for visualization and interaction
- Computer-aided drug design
- Protein structure prediction
- Disorder in protein structure and function

2000

- Protein evolution and structural genomics
- Protein structure prediction in biology and medicine
- Molecules to maps: tools for visualization and interaction
- Molecular network modeling and data analysis
- Data mining and knowledge discovery in molecular databases
- Identification of coordinated gene expression and regulatory sequences
- Natural language processing for biology
- Computer-aided combinatorial chemistry and cheminformatics
- Applications of information theory in biology
- Human genome variation: analysis of SNP data

2001

- Human Genome Variation: Linking Genotypes to Clinical Phenotypes
- Disorder and Flexibility in Protein Structure and Function
- DNA Structure, Protein-DNA Interactions, and DNA-Protein Expression
- Structures, Phylogenies, and Genomes
- High Performance Computing for Computational Biology
- Natural Language Processing for Biology
- Genome, Pathway and Interaction Bioinformatics
- Phylogenetics in the Post-Genomic Era
- Bioethics, Fiction Science, and the Future of Mankind

2002

- Human Genome Variation: Disease, Drug Response, and Clinical Phenotypes
- Genome-Wide Analysis and Comparative Genomics
- Expanding Proteomics to Glycobiology
- Literature Data Mining for Biology
- Genome, Pathway and Interaction Bioinformatics
- Phylogenetic Genomics and Genomic Phylogenetics
- Proteins: Structure, Function and Evolution

2003

- Gene Regulation
- Genome, Pathway, and Interaction Bioinformatics
- Informatics Approaches in Structural Genomics
- Genome-wide Analysis and Comparative Genomics
- Linking Biomedical Language, Information and Knowledge
- Human Genome Variation: Haplotypes, Linkage Disequilibrium, and Populations
- Biomedical Ontologies

2004

- Alternative Splicing
- Computational Tools for Complex Trait Gene Mapping
- Biomedical Ontologies
- Joint Learning from Multiple Types of Genomic Data
- Informatics Approaches in Structural Genomics
- Computational and Symbolic Systems Biology

2005

- Inferring SNP Function Using Evolutionary, Structural and Computational Methods
- Biogeometry: Applications of Computational Geometry to Molecular Structure
- Inferring Function from Structural Genomics Targets
- Computational Approaches for Pharmacogenomics
- Joint Learning from Multiple Types of Genomic Data
- Biomedical Ontologies

CiteSeer's Most Commonly Cited PSB Papers in Computer Science Literature

K. Fukuda, T. Tsunoda, A. Tamura, and T. Takagi. *Toward information extraction: Identifying protein names from biological papers*. 1998.

Liang S., Fuhrman S. and Somogyi R. REVEAL, *A General Reverse Engineering Algorithm for Inference of Genetic Network Architectures*. 1998.

A. Regev, W. Silverman, and E. Shapiro. *Representation and simulation of biochemical processes using the pi-calculus process algebra*. 2001

Matsuno, H., Doi, A., Nagasaki, M., and Miyano, S. *Hybrid Petri net representation of gene regulatory network*. 2000.

C. Leslie, E. Eskin, and W. S. Noble. *The spectrum kernel: A string kernel for SVM protein classification*. 2002

S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer, and K. Sonmez. *Pathway logic: Symbolic analysis of biological signaling*. 2002.

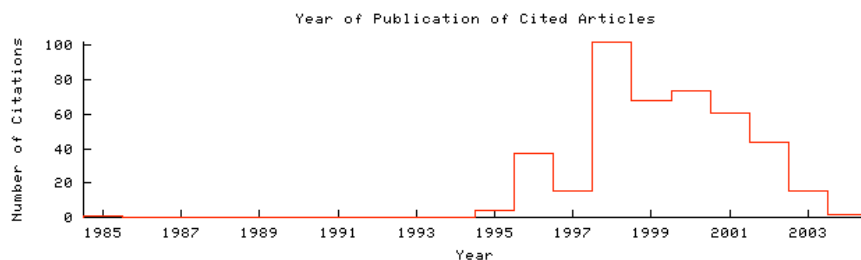
A. J. Hartemink, D. K. Gifford, T. S. Jaakkola, and R. A. Young, *Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks*. 2001

Raychaudhuri, S., Stuart, J. M. and Altman, R. B. *Principal components analysis to summarize microarray experiments: application to sporulation time series*. 2000.

James Thomas, David Milward, Christos Ouzounis, Stephen Pulman, and Mark Carroll. *Automatic extraction of protein interactions from scientific abstracts*. 2000.

Wroe CJ, Stevens RD, Goble CA, Ashburner M. *A methodology to migrate the Gene Ontology to a description logic environment using DAML+ OIL*. 2003.

CiteSeer number of citations in computer science literature plotted based on year of publication of PSB papers. There is a 4-5 year lag between having papers published and having them cited.



Highwire Press automated index of topics in papers referring to PSB proceedings. Number of citations discussing the topic in parentheses.

- | | |
|------------------------------|----------------------------|
| Gene expression (188) | Sample sizes (9) |
| Network (110) | Duplication; Evolution (9) |
| Alignment (88) | Precision And Recall (8) |
| Extracted (56) | Controlling Gene (7) |
| Ontology (35) | Simulation system (7) |
| Side Chain (30) | Molecular structure (7) |
| Disordered; Regions (27) | Cross-Linking (7) |
| Splicing (24) | Known 3D structure (7) |
| Metabolic pathways (16) | Pain; Modulation (6) |
| NMR (16) | Prediction Meta (6) |
| Support vector machines (14) | Promoter prediction (6) |
| Haplotype; Markers (12) | Protein Evolution (6) |
| Free energy (12) | Hormone; Temporal (6) |
| HMMs; Models (11) | Knowledge discovery (6) |
| Protein kinase (10) | Combinatorial Library (6) |
| Genome maps (10) | |