A TWENTIETH ANNIVERSARY TRIBUTE TO PSB

DARLA HEWETT[1], MICHELLE WHIRL-CARRILLO[1], LAWRENCE E HUNTER[2], RUSS B ALTMAN[1], TERI E KLEIN[1]

[1]*Stanford University, Shriram Center for Bioengineering and Chemical Engineering*
443 Via Ortega, Stanford, CA 94305
Email: teri.klein@stanford.edu
[2]*University of Colorado School of Medicine, Computational Bioscience Program,*
Aurora CO 80045

PSB brings together top researchers from around the world to exchange research results and address open issues in all aspects of computational biology. PSB 2015 marks the twentieth anniversary of PSB. Reaching a milestone year is an accomplishment well worth celebrating. It is long enough to have seen big changes occur, but recent enough to be relevant for today. As PSB celebrates twenty years of service, we would like to take this opportunity to congratulate the PSB community for your success. We would also like the community to join us in a time of celebration and reflection on this accomplishment.

1. PSB's Influence

PSB is one of the world's leading conferences in computational biology. It is where top researchers present and discuss current research in the theory and application of computational methods in problems of biological significance. The following facts, computed October 2014, highlight how PSB has impacted and supported our community:

- PSB has accepted and is tracking 887 papers.
- 81% of the papers submitted to PSB have been cited in Google Scholar.
- On the average, if a PSB paper has citations recorded in Google Scholar, it is cited 45 times.
- There are currently 32,504 citations of PSB Papers recorded in Google Scholar.
- The most highly cited PSB paper has 978 citations recorded in Google Scholar.
- 538 organizations have contributed to PSB.
- 2583 individuals have contributed 3671 times to PSB as an author, a session leader, or a workshop leader.

2. PSB Top 10 Papers

PSB has advanced science in many ways but these top ten highly cited papers, and the topics they represent, have been of particular importance to our community.

| Session	Title	Authors	Description	Year
Genetic Relationships	REVEAL, A General Reverse Engineering Algorithm for Inference of Genetic Network Architectures	S Liang S Fuhrman R Somogyi	An algorithm for inferring genetic network architectures from state transition tables, which correspond to time series of gene expression patterns, using the Boolean network model.	1998
DNA Patterns	BioProspector: Discovering Conserved DNA Motifs in Upstream Regulatory Regions of Co-Expressed Genes	X Liu DL Brutlag JS Liu	BioProspector examines the upstream region of genes in the same gene expression pattern group and looks for regulatory sequence motifs.	2001
Protein Classification	The Spectrum Kernel: A String Kernel for SVM Protein Classification	C Leslie E Eskin WS Noble	String-based kernels, in conjunction with SVMs (support vector machines), offer a viable and computationally efficient alternative to other methods of protein classification and homology detection.	2002
Genetic Relationships	Principal Components Analysis to Summarize Microarray Experiments: Application To Sporulation Time Series	S Raychaudhuri JM Stuart RB Altman	Application of PCA to expression data provides a summary of the ways in which gene responses vary under different conditions. This analysis clarifies the relationship between previously reported clusters and is used to examine the relationships and differences between genes.	2000
Genetic Relationships	Mutual Information Relevance Networks: Functional Genomic Clustering Using Pairwise Entropy Measurements	AJ Butte IS Kohane	A technique that computes comprehensive pair-wise mutual information for all genes in RNA expression data to find functional genomic clusters.	2000
Gene Expression	Modeling Gene Expression with Differential Equations	T Chen HL He GM Church	A differential equation model for gene expression with two methods to construct the model from a set of temporal data.	1999
Genetic Relationships	Identification of Genetic Networks from a Small Number of Gene Expression Patterns Under the Boolean Network Model	T Akutsu S Miyano S Kuhara	REVEAL: improvement and mathematical proof.	1999

Genetic Relationships	Linear Modeling of mRNA Expression Levels During CNS Development and Injury	P D'Haeseleer X Wen S Furhman R Somogyi	A linear modeling approach that allows one to infer interactions between all the genes included in a data set. The resulting model can be used to generate interesting hypotheses to direct further experiments.	1999
Gene Expression	Modeling Regulatory Networks with Weight Matrices	DC Weaver CT Workman GD Stormo	Representing regulatory relationships between genes as linear coefficients or weights, with the "net" regulation influence on a gene's expression being the mathematical summation of the independent regulatory inputs.	1999
Gene Expression	Using Graphical Models and Genomic Expression Data to Statistically Validate Models of Genetic Regulatory Networks.	AJ Hartemink DK Gifford T Jaakkola RA Young	A model-driven approach for analyzing genomic expression data that permits genetic regulatory networks to be represented in a biologically interpretable computational form.	2001

3. Twenty contributors to celebrate twenty years

Thousands of contributors have made significant and relevant contribution to PSB. However, we would like to highlight twenty people whom have recorded the most contributions to PSB. We believe these contributors should be counted on anyone's list of important contributors to the field of biological research. Since selection was based on contribution to PSB, not every important figure is included; and some people are included who would not describe themselves as important contributors despite their considerable impact on the field. However, based on the number and significance of their contributions to PSB, here is a list of twenty contributors who we would like to acknowledge as we celebrate the twentieth anniversary of PSB.

Russ Altman, MD, PhD
http://helix-web.stanford.edu/
Professor of Bioengineering, Genetics, and Medicine, (and Computer Science, by courtesy), Stanford University
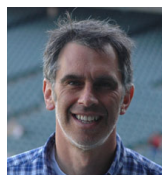
Russ Altman directs the Helix group. The Helix group focuses on the creation and application of computational tools to solve problems in biology and medicine.

Atul Butte, MD, PhD
https://buttelab.stanford.edu/

Associate Professor of Pediatrics & of Genetics and, by courtesy, of Computer Science, of Medicine & (BMIR) & of Pathology, Stanford University

The long-term research goal of the Butte Lab is to solve problems relevant to genomic medicine by developing new methodologies in translational bioinformatics.

Francisco M. De La Vega, D.Sc.
http://www.annaisystems.com/management-team-3/

Chief Scientific Officer, Anni Systems

Francisco De La Vega is a geneticist and computational biologist who has focused on analytical problems of large scale sequencing projects. Annai is a genomics management and analytics start-up, currently with a beachhead in cancer genomics.

Keith Dunker, PhD
http://www.compbio.iupui.edu/group/5/pages/about_us

Director, Center for Computational Biology and Bioinformatics, Professor, Biochemistry and Molecular Biology, Professor, School of Informatics, Indiana University
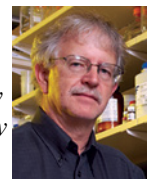
Keith Dunker and his collaborators were the first to consider intrinsically disordered proteins as a distinct class with important biological functions. Ongoing work suggests that the original proteins on earth were intrinsically disordered and that protein evolution followed a disorder to order pathway.

Graciela Gonzalez, PhD
https://webapp4.asu.edu/directory/person/869038

Associate Professor and Data Core Director for NIH/NIA supported Alzheimer's Disease Centers, Arizona State University

Graciela Gonzalez leads the DIEGO (Discovery through Integration and Extraction of Genomic knOwledge) lab focusing her research on translational applications of information extraction using Natural Language Processing techniques.

Alexander J. Hartemink, PhD
http://www.cs.duke.edu/~amink/

Professor of Computer Science, Statistical Science, and Biology Duke University

Alexander Hartemink's research interests are in computational systems biology and machine learning. Specifically, his work focuses on the development and application of new statistical learning algorithms to complex problems in systems biology.

David Haussler, PhD

https://genomics.soe.ucsc.edu/haussler
Investigator, Howard Hughes Medical Institute, Director and Distinguished Professor, Center for Biomolecular Science and Engineering, Director, Cancer Genomics Hub, Scientific Co-Director, California Institute for Quantitative Biosciences (QB3), University of California, Santa Cruz

David Haussler is known for his work leading the team that assembled the first human genome sequence. He is credited with pioneering the use of hidden Markov models (HMMs), stochastic context-free grammars, and the discriminative kernel method for analyzing DNA, RNA, and protein sequences. He was the first to apply the latter methods to the genome-wide search for gene expression biomarkers in cancer, now a major effort of his laboratory.

Lawrence Hunter, PhD

http://compbio.ucdenver.edu/hunter/
Director, Center for Computational Pharmacology, Computational Bioscience Program, School of Medicine at the University of Colorado

Larry Hunter's laboratory is focused on knowledge-driven extraction of information from the primary biomedical literature, the semantic integration of knowledge resources in molecular biology, and the use of knowledge in the analysis of high-throughput data.

Teri E. Klein, PhD

https://med.stanford.edu/profiles/teri-klein
Director & Co-Principal Investigator, PharmGKB Stanford University

Teri Klein's research interests extend over the broad spectrum of pharmacogenetics, computational biology and bioinformatics. Applications include the development of a pharmcogenetics knowledge base, structure-function relationships, de novo modeling and the structural basis of disease.

Isaac S. Kohane, MD, PhD

http://chip.org/zak/
Professor of Pediatrics & Health Sciences Technology, Harvard Medical School

Isaac Kohane is transforming healthcare systems into discovery engines to accelerate innovation in clinical decision-support. He is working on neurodevelopmental diseases (NDD) and specifically autism spectrum disorder. Understanding the full landscape of its clinical manifestations, emphasizing early and systematic molecular-clinical diagnosis through the creation of an NDD information commons.

Satoru Kuhara, PhD

http://hyoka.ofc.kyushu-u.ac.jp/search/details/K000030/english.html

Professor, Bioscience & Biotechnology, Faculty of Agriculture, Kyushu University

Satoru Kuhara focus on the transcriptional regulation, the first step of the gene regulatory circuit and the study of the regulatory network in cells and in multi-cellular organization.

Shoudan Liang, PhD

http://profiles.gulfcoastconsortia.org/profilesystem/editprofile.php?pid=3461
Professor, Bioinformatics and Computational Biology, UT MD Anderson Cancer Center

Shoudan Liang creates algorithms to analyze microarray data with the aim of understanding the genetic circuitry that underlies diverse systems such as stem cell differentiation, development in animals, proliferation and apoptosis in immune and cancer cells.

Jun S. Liu, PhD
http://www.people.fas.harvard.edu/
~junliu/

Professor of Statistics
Harvard University

Jun Liu thinks he can solve some of the mysteries about genes quicker with statistics than biologists can with laboratory experiments. Using a method that can be illustrated on a blackboard, he has successfully predicted the locations of on/off switches for genes in a bacterium.

Satoru Miyano, PhD
http://dnagarden.hgc.jp/en/doku.php

Professor, Human Genome Center
Institute of Medical Science
The University of Tokyo

Satoru Miyano's mission is to create computational strategy for systems biology and medicine towards translational bioinformatics. The supercomputer system is the indispensable infrastructure for his mission.

Marylyn Ritchie, PhD
http://bmb.psu.edu/directory/mdr23

Director, Center for Systems Genomics
Professor of Biochemistry and Molecular
Biology,
The Pennsylvania State University

The mission of the Ritchie Lab is to improve our understanding of the underlying genetic architecture of common diseases. New methods focus on the detection of gene-gene interactions, gene-environment interactions, and network and/or pathway effects associated with human disease.

Gary Stormo, PhD
http://stormo.wustl.edu/

Professor, Department of Genetics
and the Center for Genome Sciences
and Systems Biology, Washington
University School of Medicine in
St Louis

The Stormo Lab's work is focused on understanding and modeling the regulation of gene expression, especially the basis of specificity in protein-nucleic acid interactions.

Xiaole Shirley Liu, PhD
http://liulab.dfci.harvard.edu/

Professor of Biostatistics
Harvard School of Public Health

X. Shirley Liu has helped develop a number of widely used algorithms for transcription factor motif finding, ChIP-chip/Seq and DNase-seq data analysis, and is developing more algorithms and data integration approaches for high throughput data in transcriptional and epigenetic gene regulation.

William Stafford Noble, PhD
http://noble.gs.washington.edu/

Professor of Genome Sciences
University of Washington

Bill Noble's research focuses on the development and application of machine learning and statistical methods for interpreting complex biological data sets. Currently, his research can be divided into three areas: predicting protein properties, chromatin and gene regulation, and analysis of mass spectrometry data.

Roland Somogyi, PhD
http://www.molecularmining.com/

Chief Scientific Officer,
Molecular Mining

Roland Somogyi is developing the tools that will help us understand how genes work together in gene networks. He is known for his work in using gene expression data to reverse-engineer gene networks.

Richard A. Young, PhD
http://younglab.wi.mit.edu/

Professor of Biology
Massachusetts Institute of Technology

The Young laboratory is mapping the regulatory circuitry that controls cell state and differentiation in mice and humans. The Young lab uses experimental and computational technologies to determine how signaling pathways, transcription factors, chromatin regulators and small RNAs control gene expression programs in embryonic stem cells and differentiated cells.

4. Acknowledging PSB's Supporting Organizations

We would also like to acknowledge and thank the 538 organizations that have supported the 2583 authors, session leaders, and workshop leaders of PSB. We do understand that behind the contributors to PSB there are many organizations supporting their contribution. Organizational support of our community has been vital to PSB's success.

5. Looking Ahead

In contemplating a vision for the future of biological research, it is appropriate to consider the remarkable path that has brought us here and the significant role PSB has had in the journey. While we celebrate our progress, on this twentieth anniversary we also want to look ahead to the future. Looking ahead we have the opportunity to explore a new vision, of transformative new approaches to achieve health benefits. Although innovative biological analysis methods are rapidly permeating biomedical research, the challenge of establishing robust paths from knowledge and information to improved human health remains. We look forward to the future, hearing from you, collaborating with you, reading your papers, and being challenged by your continued innovation.

Congratulations PSB.

References

1. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Liang S, Fuhrman S, Somogyi R*. Pac Symp Biocomput. 1998; :18-29.
2. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Liu X, Brutlag DL, Liu JS*. Pac Symp Biocomput. 2001 :127-38.
3. The spectrum kernel: a string kernel for SVM protein classification. *Leslie C, Eskin E, Noble WS*. Pac Symp Biocomput. 2002 :564-75.
4. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Raychaudhuri S1, Stuart JM, Altman RB*. Pac Symp Biocomput. 2000 :455-66.
5. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Butte AJ, Kohane IS*. Pac Symp Biocomput. 2000 :418-29.
6. Modeling gene expression with differential equations. *Chen T, He HL, Church GM*. Pac Symp Biocomput. 1999 :29-40.
7. Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Akutsu T, Miyano S, Kuhara S*. Pac Symp Biocomput. 1999 :17-28.
8. Linear modeling of mRNA expression levels during CNS development and injury. *D'haeseleer P, Wen X, Fuhrman S, Somogyi R*. Pac Symp Biocomput. 1999 :41-52.
9. Modeling regulatory networks with weight matrices. *Weaver DC, Workman CT, Stormo GD*. Pac Symp Biocomput. 1999 :112-23.
10. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. *Hartemink AJ, Gifford DK, Jaakkola TS, Young RA*. Pac Symp Biocomput. 2001 :422-33.
11. A Vision for the Future of Genomics Research. *Francis S. Collins, Eric D. Green, Alan E. Guttmacher and Mark S. Guyer*. Nature, Vol. 422, No. 6934, April 24, 2003, p. 835-847.