Pierre Baldi

School of Information and Computer Sciences
Department of Computer Science
Department of Biological Chemistry
Institute for Genomics and Bioinformatics
California Institute for Telecommunications and Information Technology
University of California, Irvine
Irvine, CA 92697-3435
(949) 824-5809
(949) 824-9813 FAX
Assistant: Katarina Fletcher (fletcher@ics.uci.edu)
pfbaldi@ics.uci.edu
www.ics.uci.edu/~pfbaldi
www.igb.uci.edu

Professional Experience

- October 2006 to present: Chancellor's Professor
- September 2006 to present: Associate-Director, Center for Machine Learning and Intelligent Systems.
- January 2001 to present: Director Institute for Genomics and Bioinformatics.
- June 2001 to present: Professor Department of Information and Computer Science, University of California, Irvine. [Joint appointment in the Department of Biological Chemistry, College of Medicine and the Department of Developmental and Cell Biology, School of Biological Sciences].
- January 2002 to December 2005: Application Layer Leader (Digitally Enabled Medicine) for the California Institute for Telecommunications and Information Technology [Calit2^2].
- July 1999 to May 2001: Associate Professor, Department of Information and Computer Science, University of California, Irvine. [Joint appointment in the Department of Biological Chemistry, College of Medicine and the Department of Developmental and Cell Biology, School of Biological Sciences].
- 1991 to June 1999: Chairman and CEO, Net-ID, Inc.
- January 1999: Visiting Professor, Department of Computer Science, University of Florence.
- 1995 to 1996: Member of the Professional Staff, Division of Biology, California Institute of Technology.
- 1988 to 1995: Member of the Technical Staff in the Nonlinear Science and Information Processing Group at the Jet Propulsion Laboratory, and Visiting Research Associate, Division of Biology, California Institute of Technology.
- Summer 1988: Visiting Research Mathematician, Department of Mathematics, University of California, San Diego.
- 1986-1988: Visiting Lecturer, Department of Mathematics, University of California, San Diego.

Education
- 1986: Ph.D. Mathematics, California Institute of Technology.
- 1983: MS Computer Science and Engineering, ENSTA, Paris.
- 1981: D.E.A Mathematics, University of Paris VII.
- 1980: MS Mathematics, University of Paris VII.
- 1980: MS Psychology, University of Paris X.

Honors
- 2010 Eduardo R. Caianiello Prize for Scientific Contributions to the Field of Neural Networks from the Italian Neural Network Society (SIREN).
- 2009 Dean's Award for Research (UCI).
- 2008 Elected Fellow American Association Advancement of Science (AAAS).
- 2007 Elected Fellow Association Advancement Artificial Intelligence (AAAI).
- 2006 Chancellor's Professor.
- 2006 Microsoft Faculty Research Award.
- 2006 IEEE Senior Member.
- 2001 Certificate of Recognition for Neural Network Invention by NASA.
- 1999 Laurel Wilkening Faculty Innovation Award at UCI.
- 1998 Certificate of Service for the Caltech Alumni Fund for the 1997-98 fund raising effort.
- Recipient of the 1993 Lew Allen Award at JPL.
- 1992 Certificate of Service with Distinction as Associate Editor of the *IEEE Transactions on Neural Networks*.

PUBLICATIONS

Refereed Journals

- J.146 R. Nasr, D. Hirschberg, and **P. Baldi**. Hashing Algorithms and Data Structures for Rapid Searches of Fingerprint Vectors. *Journal of Chemical Information and Modeling*, in press, (2010).
- J145. **P. Baldi** and R. Nasr. When is Chemical Similarity Significant? The Statistical Distribution of Chemical Similarity Scores and Its Extreme Values. *Journal of Chemical Information and Modeling*, in press, (2010).
- J. 144 Li Liang, Diana Leng, Chad Burk, Rie Nakajima-Sasaki, Matt Kayala, Vidya L. Atluri, Jozelyn Pablo, Berkay Unal, Thomas A. Ficht, Eduardo Gotuzzo, Mayuko Saito, John Morrow, Xiaowu Liang, Pierre Baldi, Joseph M. Vinetz, Philip L. Felgner, Renee M. Tsolis. Large Scale Immune Profiling of Infected Humans and Goats Reveals Differential Recognition of *Brucella melitensis* Antigens. *PLoS Neglected Tropical Diseases*, in press, (2010).
- J. 143 P. Crompton, M. Kayala, B. Traore, K. Kayentao, A. Ongoiba, G. Weiss, D. Molina, C. Burk, M. Waisberg, A. Jasinskas, X. Tan, S. Doumbo, D. Doumtabe, Y. Kone, D. Narum, X. Liang, O. Doumbo, L. Miller, D. Doolan, **P. Baldi**, P. Felgner, S. Pierce. A Prospective Analysis of the Antibody Response to *Plasmodium falciparum* Before and After a Malaria Season by Protein Microarray. *PNAS*, in press, (2010).
- J. 142 S. J. Swamidass, C. Azencott, K. Daily, and P. Baldi. A CROC Stronger than ROC: Measuring, Visualizing, and Optimizing Early Retrieval. *Bioinformatics*, in press, (2010).

- J. 141 S. Javanmardi, C. Lopes, and **P. Baldi**. Modeling User Reputation in Wikis. *Statistical Analysis and Data Mining*, 3, 2, 126-139, (2010).
- J. 140 A. B. Mochon1, J. Ye, M. A. Kayala, J. R. Wingard, C. J. Clancy, M. H. Nguyen, P. Felgner, P. Baldi, and H. Liu. Serological Profiling of a *Candida albicans* Protein Microarray Reveals Permanent Host-Pathogen Interplay & Stage-Specific Responses during Candidemia. *PLoS Pathogens*, in press, (2009).
- J. 139 **P. Baldi** and L. Itti. Of Bits and Wows: A Bayesian Theory of Surprise with Applications to Attention. *Neural Networks*, 23, 649—666, (2010).
- J. 138 D. M. Molina, S. Pal, A. Teng, M. Kayala, **P. Baldi**, P. Felgner, X. Liang, L. M. de la Maza. Identification of Immunodominant Antigens of Chlamydia Trachomatis Using Proteome Microarrays. *Vaccine*, in press, (2009).
- J. 137 D. Guzman, A. Randall, **P. Baldi**, and Z. Guan. Computational and single-molecule force studies of a macro domain protein reveal a key molecular determinant for mechanical stability. *Proceedings of the National Academy of Sciences USA*, 107, 1989—1994, (2010).
- J. 136 J. Chen and **P. Baldi**. No Electron Left-Behind: a Rule-Based Expert System to Predict Chemical Reactions and Reaction Mechanisms. *Journal of Chemical Information and Modeling*, 49, 9, 2034-2043, (2009).
- J. 135 **P. Baldi** and D. Hirschberg. An Intersection Inequality Sharper than the Tanimoto Triangle Inequality for Efficiently Searching Large Databases. *Journal of Chemical Information and Modeling*, 49, 8, 1866-1870, (2009).
- J. 134 C. N. Magnan, A. Randall, and **P. Baldi**. SOLpro: Accurate Sequence-Based Prediction of Protein Solubility. *Bioinformatics*, 25, 17, 2200-2207, (2009).
- J. 133 R. J. Nasr, S. J. Swamidass, and **P.Baldi**. Large Scale Study of Multiple-Molecule Queries. *Journal of Chemoinformatics*, 1:7, 2009.
- J. 132 M. C. Brandon, D. C. Wallace, and **P. Baldi**. Data Structures and Compression Algorithms for Genomic Sequence Data. *Bioinformatics*, 25, 1731-1738, (2009).
- J131. P.Felgner, M. Kayala, A. Vigil, C. Burk, R. Nakajima-Sasaki, J. Pablo, D. Molina, S. Hirst, J. Chew, D. Wang, G. Tan, M. Duffield, R. Yang, J. Neel, N. Chantratita, G. Bancroft, G. Lertmemongkolchai, D. Davies, P. Baldi, S. Peacock , and R. Titball. A Burkholderia pseudomallei protein array reveals serodiagnostic and cross-reative antigens. *PNAS*, 106, 13499-13504, (2009).
- J. 130 S. J. Swamidass, C. Azencott, H. Gramajo, S. Tsai, and **P. Baldi**. The Influence Relevance Voter: An Accurate and Interpretable Virtual High Throughput Screening Method. *Journal of Chemical Information and Modeling*, 49, 4, 756--766, (2009).
- J 129 X. Xie, P. Rigor, and P. Baldi. MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics*, 25, 167-174, (2009).
- J. 128 M. J. Sweredoski and **P. Baldi**. COBEpro: A Novel System for Predicting Continuous B-cell Epitopes. *Protein Engineering Design and Selection*, 22, 3, 113-120, (2009).
- J 127 A. Randall and **P. Baldi**. SELECTpro: Effective Protein Model Selection Using a Structure-Based Energy Function Resistant to Blunders. BMC Structural Biology,8:52, (2008).
- J126. L. Itti and **P. Baldi**. Bayesian Surprise Attracts Human Attention. *Vision Research*, 49, 1295-1306, (2009).
- J125 E. Linstead, S. Bajracharya, T. Ngo, P. Rigor, C. Lopes, **P. Baldi.** Sourcerer: Mining and Searching Internet-Scale Software Repositories. *Journal of Datamining and Knowledge Discovery*, 18, 2, 300-336, (2009).
- J124. M. Brandon,E. Ruiz-pesini, D. Mishmar, V. Procaccio, M.T. Lott, K. C. Nguyen, S. Spolim, U. Patil, **P. Baldi**, and D. Wallace. MITOMASTER: A Bioinformatics Tool For the

Analysis of Mitochondrial DNA Sequences. *Human Mutation*, Database Issue, 30, 1, 1-6, (2009).

- J123 R. Jurdak, **P. Baldi**, and C. Lopes. Software-Driven Sensor Networks for Short-Range Shallow Water Applications. .*Ad Hoc Networks Journal*, 7, 837-848, (2008).
- J122 J. Cheng, A. N. Tegge, and **P. Baldi**. Machine Learning Methods for Protein Structure Prediction. *IEEE Reviews in Biomedical Engineering*, 1, 1, in press, (2008).
- J121. J. Chen and **P. Baldi**. Synthesis Explorer: Organic Chemistry Tutorial System for Multi-Step Synthesis Design and Reaction Prediction. *Journal of Chemical Education*, 85, 12, 1699-1703, (2008).
- J120 L. Chang, S. Sundaresh, J. Elliott, P. A. Anton, **P. Baldi**, A. Licudine, M. Mayer, T. Vuong, M. Hirano, B. D. Naliboff, V. Z. Ameen, and E. A. Mayer, Dysregulation of the Hypothalamic-Pituitary-Adrenal (HPA) Axis in Irritable Bowel Syndrome. *Neurogastroenterology and Motility*, in press, (2008).
- J119 **P. Baldi** and R. W. Benz. BLASTing Small Molecules—Statistics and Extreme Statistics of Chemical Similarity Scores. *Bioinformatics*, 24(13):i357-i365, (2008).
- J118 M. J. Sweredoski and **P. Baldi**. PEPITO: Improved Discontinuous B-Cell Epitope Prediction Using Multiple Distance Thresholds and Half-Sphere Exposure. Bioinformatics, 24, 12, 1459-1460, (2008).
- J 117. D. L. Doolan, Y. Mu, B. Unal, S. Sundaresh, S. Hirst, C. Valdez, A. Randall, D. Molina, X. Liang, J. A. Oloo, P. L. Blair, J. C. Aguiar, **P. Baldi**, D. Huw Davies, and P. L. Felgner. Profiling Humoral Immune Responses to *P. falciparum* Infection with Protein Microarrays, *Proteomics*, 8, i22, 4680-4694, (2008).
- J 116. **P. Baldi**, D. S. Hirschberg and R. J. Nasr. Speeding Up Chemical Database Searches Using a Proximity Filter Based on the Logical Exclusive-OR. *Journal of Chemical Information and Modeling*,48, 7, 1367-1378, (2008).
- J115. A. G. Barbour, A. Jasinskas, M. Kayala, D. Huw Davies, A. C. Steere, **P. Baldi**, and P. Felgner. A genome-wide proteome array reveals a limited set of immunogens in natural infections of humans and white-footed mice with *Borrelia bugdorferi*. *Infection and Immunity*, 76, 3374-3389, (2008).
- J114. R. W. Benz, S. J. Swamidass, and **P. Baldi**. Discovery of Power-Laws in Chemical Space. *Journal of Chemical Information and Modeling*, 48, 6, 1138-1151, (2008).
- J113. L. Wu and **P. Baldi**. Learning to Play Go Using Recursive Neural Networks. *Neural Networks*, 21, 1392-1400, (2008).
- J112. A. Randall, J. Cheng, M. Sweredoski, and **P. Baldi**. TMBpro: Secondary Structure, Beta-Contact, and Tertiary Structure Prediction of Transmembrane Beta-Barrel Proteins. *Bioinformatics*, 24, 4, 513-520, (2008).
- J111. **P. Baldi**., R. W. Benz, D. S. Hirschberg, and S. Joshua Swamidass. Lossless Compression of Chemical Fingerprints Using Integer Entropy Codes Improves Storage and Retrieval. *Journal of Chemical Information and Modeling*, 47, 6, 2098-2109, (2007).
- J 110 S. Sundaresh; A. Randall; B. Unal; J. M. Petersen; J. T. Belisle; M. Gill Hartley; M. Duffield; R. W. Titball; D. Huw Davies; P. L. Felgner; and **P. Baldi**. From protein microarrays to diagnostic antigen discovery: a study of the pathogen *Francisella tularensis.* *Bioinformatics*, 23, i508-i518, (2007).
- J109. J. Chen, E. Linstead, S. J. Swamidass, D. Wang, and **P. Baldi**. ChemDB Update—Full-Text Search and Virtual Chemical Space. *Bioinformatics*, 23, 2348-2351, (2007).
- J108. M. Tress, J. Cheng, **P. Baldi**, K. Joo, J. Lee, J. Seo, J. Lee, D. Baker, D. Chivian, D. Kim, I. Ezkurdia. Assessment of Predictions Submitted for the CASP7 Domain Prediction Category. *Proteins*, 69, S8, 137-151, (2007).

- J107. M. Sweredoski, K. Donovan, B. Nguyen, A. J. Shaka, and **P. Baldi**. Minimizing the Overlap Problem in Protein NMR: A Computational Framework for Precision Amino Acid Labeling. *Bioinformatics*, 23, 21, 2829-2835, (2007).
- J106. N. Komarova, L. Wu, and **P. Baldi**. The fixed-size Luria-Delbruck Model with a Nonzero Death Rate. *Mathematical Biosciences*, 210, 1, 253-290, (2007).
- J105 J. Cheng and **P. Baldi**. Improved Residue Contact Prediction Using Support Vector Machines and a Large Feature Set. *BMC Bioinformatics*, 8, 113-121, (2007).
- J104. J. E. Eyles, B. Unal, M. Gill Hartley, S. L. Newstead, H. Flick-Smith, J. L. Prior, P. C. F. Oyston, A. Randal, Y. Mu, S. Hirst, D. M. Molina, D. Huw Davies, T. Milne, K. F. Griffin, **P. Baldi**, R. W. Titball and P. L. Felgner. Immunodominant *Francisella tularensis* antigens identified using proteome microarray. *Proteomics*, 7, 13, 2172-2183, (2007).
- J103. W. Einhäuser, T. N. Mundhenk, **P. Baldi**, C. Koch, and L. Itti. A bottom-up model of spatial attention predicts human error patterns in rapid scene recognition. *Journal of Vision*, 7, 10, article 6, 1-13, (2007).
- J102. S. J. Swamidass and **P. Baldi**. A Mathematical Correction for Fingerprint Similarity Measures to Improve Chemical Retrieval. *Journal of Chemical Information and Modeling*, 47, 3, 952-964, (2007).
- J101. S. J. Swamidass and **P. Baldi**. Bounds and Algorithms for Fast Exact Searches of Chemical Fingerprints in Linear and Sub-Linear Time. *Journal of Chemical Information and Modeling*, 47, 2, 302-317, (2007).
- J100. A. Sadovsky, **P. Baldi**, and F. Wan. A Theoretical Study of the In Vivo Mechanical Properties of Angiosperm Roots: Constitutive Theories and Methods of Parameter Estimation. *Journal of Engineering Materials and Technology*, 129, 3, 483-487, (2007).
- J99. R. Jurdak, **P. Baldi** , and C. Lopes**.** Adaptive Low Power Listening for Wireless Sensor Network， *IEEE Transactions on Mobile Computing*， 6, 8, 988-1004, (2007)．
- J98. E. Ruiz-Pesini, M. T. Lott, V. Procaccio, J. Poole, M. C. Brandon, D. Mishmar, C. Yi, J. Kreuziger, P. Baldi, D. C. Wallace. An Enhanced MITOMAP with a Global mtDNA Mutational Philogeny. *Nucleic Acids Research*, 35 (Database Issue), D823-D828, (2007).
- J97. L. Zicker-Larsen, M. Zhang, N. Beliakova-Bethell, V. Bilanchone, A. Lamsa, K. Nagashima, R. Najdi, K. Kosaka, V. Kovacevic, J. Cheng, **P. Baldi**, G. Wesley Hatfield, and S. Sandmeyer. Ty3 Capsid Mutations Reveal Early and Late Functions of the Amino-Terminal Domain. *The Journal of Virology*, 81, 13, 6957-6972, (2007).
- J96. N. Wang, **P. Baldi**, and B. Gaut. Phylogenetic Analysis, Genome Evolution, and the Rate of Gene Gain in the *Herpesviridae*. *Molecular Phylogenetics and Evolution*, 43, 3, 1066-1075, (2007).
- J95. C. Azencott, A. Ksikes, S. Joshua Swamidass, J. Chen, L. Ralaivola, and **P. Baldi**. One-to Four- Dimensional Kernels for Virtual Screening and the Prediction of Physical, Chemical, and Biological Properties. *Journal of Chemical Information and Modeling*, 47, 3, 965-974, (2007).
- J94. Y. Dou, K. Fox-Walsh, **P. Baldi**, and K. Hertel. Genomic Splice-Site Analysis Reveals Frequent Alternative Splicing Close to the Dominant Splice Site, *RNA*, 12, 12, 2047-2056, (2006).
- J93. Suman Sundaresh, Denise L. Doolan, Siddiqua Hirst, Yunxiang Mu, Berkay Unal, D. Huw Davies, Phil Felgner, and Pierre **Baldi**. Identification of Humoral Immune Responses in Protein Microarrays using DNA Microarray Data Analysis Techniques. *Bioinformatics*, 22,14,1760-1766, (2006).
- J92. M. Brandon, **P. Baldi**, and D. C. Wallace. Mitochondrial Mutations in Cancer. *Oncogene*, 25, 4647-4662, (2006).

- J91. J. Cheng and **P. Baldi**. A Machine Learning Information Retrieval Approach to Protein Fold Recognition. *Bioinformatics*, 22, 12, 1456-1463, (2006).
- J90. C. V. Lopes, A. Haghighat, A. Mandal, **P. Baldi**, and T. Givargis. Localization of Off-the-Shelf Mobile Devices Using Audible Sound: Architectures, Protocols and Performance Assessment. ACM SIGMOBILE Mobile Computing and Communications Review, 10, 2, 38-50, (2006).
- J89. C. Tagwerker, K. Flick1, M. Cui, C. Guerrero, Y. Dou, B. Auer, **P. Baldi**, L. Huang, and Peter Kaiser. A tandem-affinity tag for two-step purification under fully denaturing conditions: Application in ubiquitin profiling and protein complex identification combined with in vivo cross-linking. *Molecular and Cellular Proteomics*, 4, 737-748, (2006).
- J88. T. Lin, M. Melgar, S. J. Swamidass, J. Purdon, T. Tseng, G. Gago, D. Kurth, **P. Baldi**, H. Gramajo, and S. Tsai. Structure-Based Inhibitor Design of AccD5, an Essential acyl-CoA Carboxylase Carboxyltransferase Domain of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences USA*, 103, 9, 3072-3077, (2006).
- J87. G. Pollastri, A. Vullo, P. Frasconi, and **P. Baldi**. Modular DAG-RNN Architectures for Assembling Coarse Protein Structures. *Journal of Computational Biology*, 3, 631-650, (2006).
- J86 J. Cheng, M. J. Sweredoski, and **P. Baldi**. DOMpro: Protein Domain Prediction Using Profiles, Secondary Structure, Relative Solvent Accessibility, and Recursive Neural Networks. *Data Mining and Knowledge Discovery*, 13, 1, 1-10, (2006).
- J85. J. Cheng, A. Randall, and **P. Baldi**. Prediction of Protein Stability Changes for Single Site Mutations Using Support Vector Machines. *Proteins*, 62, 4, 1125-1132, (2006).
- J84 E. T. Wang, G. Kodama, **P. Baldi**, and R. K. Moyzis. Global Landscape of Recent Inferred Darwinian Selection for *Homo Sapiens*. *Proceedings of the National Academy of Sciences USA*, 103, 135-140, (2006).
- J83. J. Cheng, H. Saigo, and **P. Baldi**. Large-Scale Prediction of Disulphide Bridges Using Kernel Methods, Two-Dimensional Recursive Neural Networks, and Weighted Graph Matching. *Proteins*, 62, 3, 617-629, (2006).
- J82. S. A. Danziger, S. J. Swamidass, J. Zeng, L. R. Dearth, Q. Lu, J. H. Chen, J. Cheng, V. P. Hoang, H. Saigo, R. Luo, **P. Baldi**, Rainer K. Brachmann, and Richard H. Lathrop. Functional Census of Mutation Sequence Spaces: The Example of p53 Cancer Rescue Mutants. *IEEE Transactions on Computational Biology and Bioinformatics*, 3, 2, 114-125, (2006).
- J81 Kristi L. Fox-Walsh1, Yimeng Dou, Bianca J. Lam1, She-pin Hung, Pierre F. Baldi, and Klemens J. Hertel. The Architecture of pre-mRNAs Affects Mechanisms of Splice-Site Pairing. *Proceedings of the National Academy of Sciences USA,* 102, 16176-16181, (2005).
- J80. J. Chen, S. J. Swamidass, Y. Dou, J. Bruand, and **P. Baldi**. ChemDB: A Public Database of Small Molecules and Related Chemoinformatics Resources. *Bioinformatics*, 21, 4133-4139, (2005).
- J79. D. M. Chung, Y. Dou, **P. Baldi**, and J. S. Nowick. The Absence of Favorable Aromatic Interactions between Beta-Sheet Peptides. *Journal of the American Chemical Society*,127 (28), 9998-9999, (2005).
- J78. J. Cheng and **P. Baldi**. Three-Stage Prediction of Protein Beta-Sheets by Neural Networks, Alignments, and Graph Algorithms. Proceedings of the 2005 Conference on Intelligent Systems for Molecular Biology, ISMB 05. *Bioinformatics*, 21, Supplement 1, i75-84, (2005).
- J77. S. J. Swamidass, J. Chen, P. Phung, J. Bruand, L. Ralaivola, and **P. Baldi**. Kernels for Small Molecules and the Prediction of Mutagenicity, Toxicity, and Anti-Cancer Activity.

Proceedings of the 2005 Conference on Intelligent Systems for Molecular Biology, ISMB 05. *Bioinformatics*, 21, Supplement 1, i359-368, (2005).

- J76. J. Cheng, M. J. Sweredoski, and **P. Baldi**. Accurate Prediction of Protein Disordered Regions by Mining Protein Structure Data. *Data Mining and Knowledge Discovery*, 11, 3, 213-222, (2005).

- J75. J. Cheng, A. Z. Randall, M. Sweredoski, and **P. Baldi**. SCRATCH: a Protein Structure and Structural Feature Prediction Server. *Nucleic Acids Research*, 33, 9, Web Server issue, W72-76, (2005).

- J74. **P. Baldi** and M. Rosen-Zvi. On the Relationship Between Deterministic and Probabilistic Directed Graphical Models: from Bayesian Networks to Recursive Neural Networks. *Neural Networks,* special issue on Neural Networks and Kernel Methods for Structured Domains, 18, 8, (2005).

- J73. L. Ralaivola, J. S. Swamidass, H. Saigo, and **P. Baldi**. Graph Kernels for Chemical Informatics. *Neural Networks,* special issue on Neural Networks and Kernel Methods for Structured Domains, 18, 8, 1093-1110, (2005).

- J72. B. Irwin, M. Aye, **P. Baldi**, N. Beliakova-Bethell, H. Cheng, Y. Dou, W. Liou, and S. Sandmeyer. Retroviruses and Yeast Retrotransposons Use Overlapping Sets of Host Genes. *Genome Research*, 15, 641-654, (2005).

- J71. K. A. Salmon, S. Hung, N. R. Steffen, R. Krupp, **P. Baldi**, G. Wesley Hatfield, and R. P. Gunsalus. Global Gene Expression Profiling in *Escherichia coli* K12: The Effects of Oxygen Availability and ArcA. *Journal of Biological Chemistry*, 280, 15, 15084-15096, (2005).

- J70. R. Jurdak, P. Baldi, and C. Videira Lopes U-MAC: A Proactive and Adaptive UWB Medium Access Control Protocol. *Wireless Communications and Mobile Computing Journal*, Special Issue on UWB Communications, 5, 5, 551-566, (2005).

- J69. S. Sundaresh, S. Hung, G. W. Hatfield, and **P. Baldi**. How Noisy and Replicable are DNA Microarray Data? *International Journal of Bioinformatics Research and Applications*, 1, 1, 31-50, (2005).

- J68. H. Davies, X. Liang, J. E. Hernandez, K. M. Romero, T. T. Nguyen, M. Kalantari-Dehaghi, S. Hirst, Y. Mu, A. Randall, **P. Baldi**, L. Villarreal, P. L. Felgner. Profiling the Humoral Response to Infection by Using Proteome Microarrays: High-Throughput Vaccine and Diagnostic Antigen Discovery. *Proceedings of the National Academy of Sciences USA,* 102, 3, 547-552, (2005).

- J67. J. Cheng, L. Scharenbroich, **P. Baldi**, and E. Mjolsness. Sigmoid: Towards a Generative, Scalable Software Infrastructure for Pathway Bioinformatics and Systems Biology. *IEEE Intelligent Systems*, 20, 3, 68-75, (2005).

- J66. M. Brandon, M. Lott, K. Nguyen, S. Spolim, S. Navathe, **P. Baldi** , and D. Wallace. MITOMAP: a human mitochondrial genome database - 2004 update. *Nucleic Acids Research*, 33 (Database issue), D611-613, (2005).

- J65. I. Sokolchik, T. Tanabe, **P. Baldi**, and J. Y. Sze. Polymodal sensory function of the C. elegans OCR-2 channel arises from distinct intrinsic determinants within the protein and is selectively conserved in human TRPV2. *Journal of Neuroscience*, 25, 4, 1015-1023, (2005).

- J64. S. Hampson, B. S. Gaut, and **P. Baldi**. Statistical Detection of Chromosomal Homology Using Shared-Gene Density Alone. *Bioinformatics*, 21, 8, 1339-1348, (2005).

- J63. R. Jurdak, C. Videira Lopes, and **P. Baldi**. A Survey, Classification, and Comparative Analysis of Medium Access Control Protocols for Ad Hoc Networks. *IEEE Communications Surveys and Tutorials*, 6, 1, 2-16, (2004).

- J62. A. Z. Randall, **P. Baldi**, and L. P. Villareal. Structural Proteomics of the Poxvirus Family. *Artificial Intelligence in Medicine,* special issue on "Data mining in Genomics and Proteomics", 31, 2, 105-115, (2004).

- J61. Y. Dou, P. Baisnee, G. Pollastri, Y. Pecout, J. Nowick, and **P. Baldi** ICBS: A Database of Interactions Between Protein Chains Mediated by Beta-Sheet Formation. *Bioinformatics*, 20, 16, 2767-2777, (2004).

- J60. A. McLysaght, **P. Baldi**, and B. S. Gaut. Extensive Gene Gain Associated With Adaptive Evolution of Poxviruses. *Proceedings of the National Academy of Sciences USA*, 100, 26, 15655-15660, (2003).

- J59. K. Salmon, K. Mekjian, R. P. Gunsalus, S. Hung, **P. Baldi**, and G. Wesley Hatfield. Global Gene Expression Profiling in *Escherichia coli* K12: The Effects of Oxygen Availability and FNR. *Journal of Biological Chemistry*, 278, 32, 29837-29855, (2003).

- J58. S. Hampson, A. McLysaght, B. Gaut, and **P. Baldi** . LineUp : Statistical Detection of Chromosomal Homology with Applications to Plant Comparative Genomics. *Genome Research*, 13, 5, 999-1010, (2003).

- J57. G. A. Weiss, T. A. Roth, **P. Baldi**, and S. S. Sidhu Comprehensive Mutagenesis of the C-terminal Domain of the M13 Gene-3 Minor Coat Protein: The Requirements for Assembly into the Bacteriophage Particle. *Journal of Molecular Biology*, 332, 4, 777-782, (2003).

- J56. **P. Baldi** and G. Pollastri. The Principled Design of Large-Scale Recursive Neural Network Architectures—DAG-RNNs and the Protein Structure Prediction Problem. *Journal of Machine Learning Research*, 4, 575-602, (2003).

- J55. Y. Guermeur, G. Pollastri, A. Elisseeff, D. Zelus, H. Paugam-Moisy, and **P. Baldi.** Combining Protein Secondary Structure Prediction Models With Ensemble Methods of Optimal Complexity. *Neurocomputing*, Vol. 56C, 305-327, (2003) (partially presented at JOBIM 2001 conference).

- J54. **P. Baldi** and W. H. Batchelder. Bounds on Variances of Estimators for Multinomial Processing Tree Models. *Journal of Mathematical Psychology*, 47, 467-470, (2003).

- J53. **P. Baldi**, L. De Nardis, and M. G. Di Benedetto. Modeling and Optimization of UWB Communication Networks Through a Flexible Cost Function. *IEEE Journal on Selected Areas in Communications*, 20, 9, 1733-1744, (2002).

- J52. S. Hung, **P. Baldi**, and G. Wesley Hatfield. Global Gene Expression Profiling in *Excherichia coli* K12: The Effects of Leucine-responsive Regulatory Protein. *Journal of Biological Chemistry*, 277, 43, 40309-40323, (2002).

- J51. P. Baisnee, S. Hampson, and **P. Baldi**. Why Are Complementary DNA Strands Symmetric? *Bioinformatics*, 18, 1021-1033, (2002).

- J50. G. Wesley Hatfield, S. Hung, and **P. Baldi**. Differential Analysis of DNA Microarray Gene Expression Data. *Molecular Microbiology*, 47, 4, 871-877, (2003).

- J49. G. Pollastri and **P. Baldi**. Prediction of Contact Maps by GIOHMMs and Recurrent Neural Networks Using Lateral Propagation From All Four Cardinal Corners. Proceedings of the 2002 Conference on Intelligent Systems for Molecular Biology, ISMB 02. *Bioinformatics*, 18, Supplement 1, S62-S70, (2002). Nominated for Best Paper Award.

- J48. **P. Baldi** and G. Pollastri. A Machine Learning Strategy for Protein Analysis. *IEEE Intelligent Systems*. Special Issue on Intelligent Systems in Biology, 17, 2, 28-35, (2002).

- J47. S. Hampson, D. Kibler, and **P. Baldi.** Distribution Patterns of Over-Represented k-mers in Non-Coding Yeast DNA. *Bioinformatics*, 18, 4, 611-626, (2002).

- J46. G. Pollastri, **P. Baldi**, P. Fariselli, and R. Casadio. Prediction of Coordination Number and Relative Solvent Accessibility in Proteins. *Proteins*, 47, 142-153, (2002).

- J45. G. Pollastri, D. Przybylski, B. Rost, and **P. Baldi**. Improving the Prediction of Protein Secondary Structure in Three and Eight Classes Using Recurrent Neural Networks and Profiles. *Proteins*, 47, 228-235, (2002).
- J44. G. Pollastri, **P. Baldi**, P. Fariselli, R. Casadio. Improved Prediction of the Number of Residue Contacts in Proteins by Recurrent Neural Networks. *Bioinformatics*, 17, S234-S242, (2001).
- J43. A. D. Long, H. J. Mangalam, B. Y. P. Chan, L. Tolleri, G. W. Hatfield, and **P. Baldi**. Improved Statistical Inference from DNA Microarray Data Using Analysis of Variance and a Bayesian Statistical Framework. *Journal of Biological Chemistry*, 276, 23, 19937-19944, (2001).
- J42. **P. Baldi** and A. D. Long. A Bayesian Framework for the Analysis of Microarray Expression Data: Regularized t-Test and Inference of Gene Changes. *Bioinformatic*s, 17, 6, 509-519, (2001).
- J41. P. Baisnée, **P. Baldi**, S. Brunak, and A. Gorm Pedersen. Flexibility of the Genetic Code with Respect to DNA Structure. *Bioinformatic*s, 17: 237-248, (2001).
- J40. **P. Baldi**, and P. Baisnée. Sequence Analysis by Additive Scales: DNA Structure for Sequences and Repeats of All Lenghts. *Bioinformatic*s, Vol 16, 10, 865-889, (2000).
- J39. **P. Baldi**, S. Brunak, Y. Chauvin, and H. Nielsen. Assessing the Accuracy of Prediction Algorithms for Classification: An Overview. *Bioinformatics*, Vol. 16, 5, 412-424, (2000).
- J38. **P. Baldi**. On the Convergence of a Clustering Algorithm for Protein-Coding Regions in Microbial Genomes, *Bioinformatics*, Vol. 16, 4, 367-371, (2000).
- J37. **P. Baldi**. Probabilistic Graphical Models in Computational Molecular Biology, *Journal of the Italian Association for Artificial Intelligence (AIIA)*, Vol. 1, 8-12, (2000).
- J36. **P. Baldi**, S. Brunak, Y.Chauvin, and A. Gorm Pedersen. Structural Basis for Triplet Repeat Disorders: A Computational Analysis, *Bioinformatics*, Vol. 15, 11, 918-929, (1999).
- J35 . **P. Baldi**, S. Brunak, P. Frasconi, G. Soda, and G. Pollastri. Exploiting the Past and the Future in Protein Secondary Structure Prediction, *Bioinformatics*, Vol. 15, 11, 937-946, (1999).
- J34. A. Gorm Pedersen, **P. Baldi**, S. Brunak, and Y. Chauvin. The Biology of Eukaryotic Promoter Prediction-a Review. *Computers and Chemistry, Special Issue on Genome and Informatics*, 23, 191-207, (1999).
- J33. A. Gorm Pedersen, **P. Baldi**, Y. Chauvin, and S. Brunak. DNA Structure in Human RNA Polymerase II Promoters, *Journal of Molecular Biology*, 281, 663-673, (1998).
- J32. **P. Baldi**, M. C. Vanier, and J. M. Bower. On the Use of Bayesian Methods for Evaluating Compartmental Neural Models, *Journal of Computational Neuroscience*, 5, 285-314, (1998).
- J31. **P. Baldi**, S. Brunak, Y. Chauvin, and A. Krogh. Naturally Occurring Nucleosome Positioning Signals in Human Exons and Introns, *Journal of Molecular Biology*, Vol. 263, 5, 503-510, (1996).
- J30. **P. Baldi** and Y. Chauvin. Hybrid Modeling, HMM/NN Architectures, and Protein Applications, *Neural Computation*, Vol. 8, 7, 1541-1565, (1996).
- J29. **P. Baldi**. Substitution Matrices and Hidden Markov Models, *Journal of Computational Biology*, Vol. 2, 3, 497-501, (1995).
- J28. **P. Baldi** and K. Hornik. Learning in Linear Networks: a Survey, *IEEE Transactions on Neural Networks*, Vol. 6, 4, 837-858, (1995)
- J27. **P. Baldi**. Gradient Descent Learning Algorithms Overview: A General Dynamical Systems Perspective, *IEEE Transactions on Neural Networks*, Vol. 6, 1, 182-195, (1995).
- J26. **P. Baldi** and Y. Chauvin. Hidden Markov Models of the G-Protein Coupled Receptor Family, *Journal of Computational Biology*, Vol. 1, 4, 311-335, (1994).

- J25. **P. Baldi**, Y. Chauvin, T. Hunkapiller, and M. A. McClure**.** Hidden Markov Models of Biological Primary Sequence Information, *PNAS USA*, Vol. 91, 3, 1059-1063, (1994).
- J24. **P. Baldi** and Y. Chauvin. Smooth On-Line Learning Algorithms for Hidden Markov Models, *Neural Computation*, Vol. 6, 2, 305-316, (1994).
- J23. **P. Baldi** and A. Atiya. How Delays Affect Neural Dynamics and Learning, *IEEE Transactions on Neural Networks*, Vol. 5, 4, 626-635, (1994).
- J22. **P. Baldi** and Y. Chauvin. Neural Networks for Fingerprint Recognition, *Neural Computation*, Vol. 5, 3, 402-418, (1993).
- J21. **P. Baldi** and S.S. Venkatesh. Random Interactions in Higher-Orders Neural Networks, *IEEE Transactions on Information Theory*, Vol. 39, 1, 274-283, (1993).
- J20. **P. Baldi** and F. Pineda. Contrastive Learning and Neural Oscillations, *Neural Computation*, Vol. 3, No. 4, 526-545, (1991).
- J19. **P. Baldi** and Y. Chauvin. Temporal Evolution of Generalization during Learning in Linear Networks, *Neural Computation*, Vol. 3, No. 4, 589-603, (1991).
- J18. E. A. Mayer, J. Furness, and **P. Baldi**. Can Regulatory Peptides Be Regarded as Words of a Biological Language?, *American Journal of Physiology*, Vol. 261(2), G171-G184, (1991).
- J17. S. S. Venkatesh and **P. Baldi**. Programmed Interactions in Higher-Orders Neural Networks: II. The Outer-Product Algorithm, *Journal of Complexity*, Vol. 7, 443-479, (1991).
- J16. S.S. Venkatesh and **P. Baldi**. Programmed Interactions in Higher-Orders Neural Networks: I. Maximal Capacity, *Journal of Complexity*, Vol. 7, 316-337, (1991).
- J15. **P. Baldi** and R. Meir. Computing with Arrays of Coupled Oscillators: an Application to Preattentive Texture Discrimination, *Neural Computation*, Vol. 2, No. 4, 458-471, (1990).
- J14. **P. Baldi**. On a Generalized Family of Colorings, *Graphs and Combinatorics*, Vol. 6, 95-110, (1990).
- J13. **P. Baldi** and E.C. Posner**.** Graph Coloring Bounds for Cellular Radio, *Computers and Mathematics with Applications*, Vol. 19, No. 10, 91-97, (1990).
- J12. **P. Baldi** and A. Atiya. Oscillations and Synchronizations in Neural Networks: an Exploration of the Labeling Hypothesis, *International Journal of Neural Systems*, Vol. 1, No. 2, 103-124, (1989).
- J11. **P. Baldi** and Y. Rinott. On Normal Approximations of Distributions in Terms of Dependency Graphs, *Annals of Probability*, Vol. 17, No. 4, 1646-1650, (1989).
- J10. **P. Baldi** and Y. Rinott. Asymptotic Normality of Some Graph Related Statistics, *Journal of Applied Probability*, Vol. 26, 171-175, (1989).
- J9. **P. Baldi** and K. Hornik. Neural Networks and Principal Component Analysis: Learning from Examples without Local Minima, *Neural Networks*, Vol. 2, No. 1, 53-58, (1988).
- J8. **P. Baldi** and W. Heiligenberg. How Sensory Maps Could Enhance Resolution Through Ordered Arrangements of Broadly Tuned Receivers, *Biological Cybernetics*, Vol. 59, No. 4,5, 313-318, (1988).
- J7. **P. Baldi**. Neural Networks, Acyclic Orientations of the Hypercube and Sets of Orthogonal Vectors, *SIAM Journal of Discrete Mathematics*, Vol. 1, No. 1 (1988).
- J6. **P. Baldi**. Group Actions and Learning for a Family of Automata, *Journal of Computer and System Sciences*, Vol. 36, No. 2, 1-15, (1988).
- J5. **P. Baldi**. Neural Networks, Orientations of the Hypercube and Algebraic Threshold Functions, *IEEE Transactions on Information Theory*, Vol. 34, No. 3, 523-530, (1988).
- J4. M. Aschbacher, **P. Baldi**, E. B. Baum, and R.M. Wilson. Embeddings of Ultrametric Spaces in Finite Dimensional Structures, *SIAM Journal of Algebraic and Discrete Methods*, Vol. 8, No. 4, 564-577, (1987).

- J3. **P. Baldi**. Symmetries and Learning in Neural Network Models, *Physical Review Letters,* Vol. 59, No. 17, 1976-1978, (1987).
- J2. **P. Baldi** and S.S. Venkatesh. Number of Stable Points for Spin Glasses and Neural Networks of Higher Orders, *Physical Review Letters*, Vol. 58, No. 9, 913-916, (1987).
- J1. **P. Baldi** and E.B. Baum. Bounds on the Size of Ultrametric Structures, *Physical Review Letters*, Vol. 56, No. 15, 1598-1600, (1986).

Books

- B4. **P. Baldi**, P. Frasconi, and P. Smyth. Modeling the Internet and the Web—Probabilistic Methods and Algorithms. *Wiley*, (2003). Japanese version in 2007.
- B3. **P. Baldi** and G. Wesley Hatfield. DNA Microarrays and Gene Regulation—From Experiments to Data Analysis and Modeling. *Cambridge University Press*, (2002). English version for China and electronic publishing version through eBrary in 2003 and 2004.
- B2. **P. Baldi**. The Shattered Self—The End of Natural Evolution, *MIT Press*, (2001). Paperback version in 2002.
- B1. **P. Baldi** and S. Brunak. Bioinformatics: the Machine Learning Approach. *MIT Press*, (1998). Second revised edition (2001). Indian and Chinese versions in 2003.

Databases and Web Servers

- Please see: www.ics.uci.edu/~pfbaldi and www.igb.uci.edu/servers/servers.html.
-

Protein Structure Prediction (SCRATCH suite):
SSpro: Protein secondary structure prediction server (3 classes).
SSpro8: Protein secondary structure prediction server (8 classes) .
ACCpro: Solvent accessibility prediction server.
CONpro: Residue contact number prediction server.
DIpro: Disulphide bridge prediction server.
MUpro: Single amino acid mutation stability prediction server.
DISpro: Disordered region prediction server.
DOMpro: Domain prediction server.
CMAPpro: Contact map prediction server.
CCMAPpro: Coarse contact map prediction server.
CMAP23Dpro: 3D reconstruction from contact map.
3Dpro: 3D structure prediction server.
TMBpro: Prediction of transmembrane beta-barrel features and tertiary structure.
SVMcon: Prediction of amino acid contact maps using Support Vector Machines.
SELECTpro: Protein model selection using a structure-based energy function.
SOLpro: Prediction of protein solubility.
BEpro: B-cell epitope prediction server.
COBEpro:B-cell continuous epitorpe prediction server.

Sequence Modeling and Analysis:
HMMpro: Hidden Markov Model simulator for biological sequence analysis, with graphical interface.

Comparative Genomics:
LineUp: Comparative genomics server (order + density).
CloseUp: Comparative genomics server (density alone).

DNA Microarray Analysis:
Cyber-T: DNA microarray gene expression analysis server.

Software Mining:
Sourcerer: Database and information retrieval system for source code Search engine for open source software

Databases:
ICBS: Inter-chain beta sheet database of protein-protein interactions.
 PSPDB: Poxvirus structural proteomics database.
Sigmoid Database: Database for molecular interactions and pathways (Systems Biology).
Sigmoid Architecture.
ChemDB: Chemoinformatics portal including: (1) a large database (5M) of organic compounds for molecular docking, drug screening, and retrosynthesis applications; (2) multiple Web server predictors of physical, chemical, and biological properties; (3) an organic chemistry expert system used to power several applications, such as Synthesis Explorer and Mechanism Explorer. Synthesis Explorer and Mechanism Explorer are interactive tutorial systems to learn undergraduate-level organic chemistry. Synthesis Explorer and Mechanism Explorer have been adopted and are currently used in all relevant undergraduate organic chemistry classes at UCI.
GOnet: Yeast database and visualization tool combining gene (SGD), gene ontology (GO), and gene interaction information (GRID).

High-Throughput Sequencing (HTS) pipeline:
This is a pipeline for analyzing data produced by the campus HTS instruments for the entire UCI campus. The pipeline is used to produce and store reads, map them to relevant genomes, and apply relevant quantitative analyses for sequencing, ChIP-seq, and RNA-seq projects.

Other
- R1. **P. Baldi** What Genes are Made Of. Review of edited book entitled: The Genomic Revolution: Unveiling the Unity of Life. *Pathways*, (2003).
- Several tutorials (e.g. Chemoinformatics Tutorial given at the 2006 ISMB Conference).

Book Chapters
- BC 20. J. Chen, M. A. Kayala, and **P. Baldi**. Reaction Explorer: Towards a Knowledge Map of Organic Chemistry to Support Dynamic Assessment and Personalized Instruction, in *Enhancing Learning with Online Resources, Social Networking and Digital Libraries*, American Chemical Society Symposium Series Book, R. E. Belford and J. Moore Editors, (2010).
- BC 19.A. Randall and **P. Baldi**. Transmembrane Beta-Barrel Protein Structure Prediction. In: *Structural Bioinformatics of Membrane Proteins*, Editor Dmitrij Frishman, Springer, in press, (2010).
- BC18. C. A. Azencott and **P. Baldi**. Virtual High-Throughput Screening with Two-Dimensional Kernels. In: *Hands-On Pattern Recognition. Challenges in Data Representation, Model Selection, and Performance Prediction,* I. Guyon , G. Cawley, G. Droor, and A. Saffari Editors, Lulu Press, (2007).
- BC17. S. J. Swamidass and **P. Baldi**. Statistical Distribution of Chemical Fingerprints. Proceedings of WILF2005, Springer Verlag. Andrea Tattamanzi Editor.
- BC16. R. Jurdak, C. V. Lopes, and **P. Baldi**. Battery Lifetime Estimation and Optimization for Underwater Sensor Networks. In *Sensor Network Operations*, published by IEEE Press,

edited by Shashi Phoha, Thomas F. La Porta and Christopher Griffin **,** 2004 (in press). Edited volume resulting from overflow to special issue on Sensor Networks in IEEE Transactions on Mobile Computing (2005).

- BC15. **P. Baldi**, G. Wesley Hatfield, and Li M. Fu. Clustering Algorithms for Gene Expression Analysis. Handbook of Computational Molecular Biology. S. Aluru Editor, Chapman & Hall/CRC Press, Computer and Information Science Series, (2005).
- BC14. S. Kremer and **P. Baldi**. Hidden Markov Models and Neural Networks. Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics. Wiley, (2004).
- BC13. **P. Baldi**. Surprise: A Shortcut for Attention? In: Neurobiology of Attention**.** L. Itti, G. Rees, and J. Tsotsos Editors, Elsevier, San Diego, CA, 24-28, (2005).
- BC12. S. P. Hung, G. Wesley Hatfield, S. Sundaresh, and **P. Baldi**. Understanding DNA Microarrays: Sources and Magnitudes of Variances in DNA Microarray Data Sets. In: Genomics, Proteomics, and Vaccines. G. Grandi Editor, John Wiley and Sons, (2003).
- BC11. **P. Baldi**. A Computational Theory of Surprise, in *Information, Coding, and Mathematics*. M. Blaum Editor, Kluwer, 1-25, (2002).
- BC10. **P. Baldi**, G. Pollastri, P. Frasconi, and A. Vullo. New Machine Learning Methods for the Prediction of Protein Topologies, in *Artificial Intelligence and Heuristic Methods in Bioinformatics*. P. Frasconi and R. Shamir Editors, IOS Press, 51-73, (2003).
- BC9. **P. Baldi**, S. Brunak, P. Frasconi, G. Pollastri, and G. Soda. Bidirectional Dynamics for Protein Secondary Structure Prediction, in *Sequence Learning: Paradigms, Algorithms, and Applications,* R. Sun and C. L. Giles Editors*,* Springer Verlag, 99-120, (2000).
- BC8. **P. Baldi**, S. Brunak, G. Pollastri, and P. Frasconi. Bidirectional IOHMMs and recurrent neural networks for protein secondary structure prediction, in *Protein Sequence Analysis in the Genomic Era*, Rita Casadio and Lanfranco Masotti Editors, CLUEB, Bologna, Italy, (2000).
- BC7. **P. Baldi**, Y. Chauvin, and V. Mittal Henkle. Software Foundation Libraries for the Design of Intelligent Systems. *Neural Nets WIRN Vietri 98. Proceedings of the 10-th Italian Workshop on Neural Nets*. M. Marinaro and R. Tagliaferri Editors. Springer Verlag Series on Perspectives on Neural Computing Series, J. Taylor Series Editor, 17-39, (1998).
- BC6. **P. Baldi**. Probabilistic Models of Neuronal Spike Trains, in *Adaptive Processing of Temporal Information*. C.L. Giles and M. Gori Editors, Lecture Notes in Computer Science, Springer Verlag, New York, 198-228, (1998).
- BC5. **P. Baldi**, Y. Chauvin, and K. Hornik. Back-propagation and unsupervised learning in linear networks, in *Back-Propagation: Theory, Architectures and Applications*. Y. Chauvin and D.E. Rumelhart Editors, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 389-432, (1995).
- BC4. **P. Baldi**. Gradient Descent Learning Algorithms: A Unified Perspective, in *Back-Propagation: Theory, Architectures and Applications*. Y. Chauvin and D.E. Rumelhart Editors, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 509-541, (1995).
- BC3. **P. Baldi** and Y. Chauvin. When Neural Networks Play Sherlock Holmes, in *Back-Propagation: Theory, Architectures and Applications*. Y. Chauvin and D.E. Rumelhart Editors, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 487-507, (1995).
- BC2. **P. Baldi** and Y. Chauvin. Statistical Models of Proteins: an Application to the G-Protein-Coupled Receptor Family, in *Modern Approaches in Molecular Bioinformatics*. S. Schulze-Kremer Editor, IOS Press, Amsterdam/New York, 53-102 (1994).
- BC1. **P. Baldi**, Y. Rinott, and C. Stein. A Normal Approximation for the Number of Local Maxima of a Random Function on a Graph, in *Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin*, T.W. Anderson, K.B. Athreya and D.L. Iglehart, Editors, Academic Press (1989).

Refereed Conference Proceedings

- C63. P. Baldi, C. Azencott, and S. J. Swamidass. Bridging the Gap Between Neural Network and Kernel Methods: Applications to Drug Discovery. Proceedings of the 22nd Italian Workshop on Neural Networks, WIRN 2010, IOS Press, (2010).
- C62. Linstead, E., Hughes, L., Lopes, C., Baldi, P. "Information-Theoretic Metrics for Project-Level Scattering and Tangling." International Conference on Software Engineering and Knowledge Engineering (SEKE). Redwood City, CA. July, 2010. (to appear)
- C61. F. Napolitano, R. Tagliaferri, and **P. Baldi**. A Scalable Reference-Point Based Algorithm to Efficiently Search Large Chemical Databases. IEEE Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN 2010), (2010).
- C60. E. Linstead, L. Hughes, C. Lopes, and **P. Baldi**. Software Analysis with Unsupervised Topic Models. Neural Information Processing Systems (NIPS) Conference, Workshop on "Applications for Topic Models", Whistler, Canada, December 2009.
- C59. E. Linstead, L. Hughes, and **P. Baldi**. The Evolution of Concerns, Scattering, and Tangling in Eclipse and ArgoUML. Third International Symposium on Empirical Software Engineering and Measurement (ESEM), Lake Buena Vista, Florida, October, submitted as a poster, (2009).
- C58. E. Linstead, L. Hughes, and **P. Baldi**. Information-Theoretic Metrics for Project-Level Scattering and Tangling. Third International Symposium on Empirical Software Engineering and Measurement (ESEM), Lake Buena Vista, Florida, October, submitted, (2009).
- C57. S. Javanmardi, Y. Ganjisaffar, C. Lopes, and **P. Baldi**. An Empirical Analysis of User Contributions in Wikipedia. WikiSym, Submitted, (2009).
- C57. S. Javanmardi, C. Loes, and P. Baldi. User Contribution and Trust in Wikipedia. Proceedings of the 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing, November 11-14, Washinton D.C., (2009).
- C56. E. Linstead and **P. Baldi**. Mining the Coherence of GNOME Bug Reports with Statistical Topic Models. MSR '09: Proceedings of the Sixth International Working Conference on Mining Software Repositories, Vancouver, BC, Canada, IEEE Computer Society, in press, (2009).
- C55. E. Linstead, L. Hughes, C. Lopes, and **P. Baldi**. Exploring Java Software Vocabulary: A Search and Mining Perspective. SUITE 2009: Proceedings of the First International Workshop on Search-Driven Development - Users, Tools, and Applications." Vancouver, BC. May 16, 2009.
- C54. E. Linstead, C. Lopes, and **P. Baldi**. An Application of Latent Dirichlet Allocation to Analyzing Software Evolution. International Conference on Machine Learning and Applications (ICMLA), San Diego,, (2008).
- C53. **P. Baldi**, C. Lopes, E. Linstead, S. Bajracharya. A Theory of Aspects as Latent Topics. 2008 ACM SIGPLAN Conference on Object-Oriented Programming Systems, Languages, and Applications. OOPSLA '08, October 19-23, Nashville, TN, (2008).
- C52. D. S. Hirschberg and P. Baldi. Effective Compression of Monotone and Quasi-Monotone Sequences of Integers. Proceedings of the 2008 Data Compression Conference (DCC 08), 520, Snowbird, Utah, J. A. Storer and M.W. Marcellin Editors, IEEE Computer Society Press, 520, (2008).
- C51. E. Linstead, P. Rigor, S. Bajracharya, C. Lopes, and **P. Baldi.** Mining Internet-Scale Software Repositories. *Advances in Neural Information Processing Systems* 20, NIPS 2007 , Y. Weiss, B. Scholkopf, and J. Platt Editors, 929-936, MIT Press, Cambridge, MA, (2008).

- C50. O. Lemos, S. Bajracharya, J. Ossher, R. Morla, P. Masiero, **P. Baldi**, and C. Lopes. CodeGenie: Using Test-Cases to Search and Reuse Source Code. *Proceedings of the 22nd IEEE/ACM International Conference on Automated Software Engineering (ASE'07), Tool Demonstration Session,* 525-526, Atlanta, GA, November 2007.
- C49. E. Linstead, P. Rigor, S. Bajracharya, C. Lopes, and **P. Baldi**. Mining Concepts from Code with Probabilistic Topic Models. *Proceedings of the 22nd IEEE/ACM International Conference on Automated Software Engineering,* (ASE'07), 461-464, Atlanta, GA, November 2007.
- C48. E. Linstead, P. Rigor, S. Bajracharya, C. Lopes, and **P. Baldi**. Mining Eclipse Developer Contributions via Author-Topic Models. *Proceedings of the MSR 2007: International Workshop on Mining Software Repositories,* 30-33, Minneapolis, MN, (2007). Best Paper Award, MSR "Scale" Challenge.
- C47. R. Jurdak, P.M.Q. Aguiar, **P. Baldi**, and C.V. Lopes. "Software Modems for Underwater Sensor Networks," *Proceedings of IEEE/OES Oceans'07*, June, 2007. Aberdeen, Scotland.
- C46. L. Wu and **P. Baldi**. "A Scalable Machine Learning Approach to Go", *Advances in Neural Information Processing Systems* 19, NIPS 2006 , Y. Weiss, B. Scholkopf, and J. Platt Editors, 1521-1528, MIT Press, Cambridge, MA, (2007).
- C45. R. Jurdak, C.V. Lopes, and **P. Baldi**. "Software Acoustic Modems for Short Range Mote-Based Underwater Sensor Networks," In Proceedings of IEEE Oceans – Asia Pacific, 1-7, Singapore, May, 2006.
- C44. R. Jurdak, C. V. Lopes, P. M. Q. Aguiar, and **P. Baldi**. "A Comparative Analysis and Experimental Study on Wireless Aerial and Underwater Acoustic Communications". Proceeding of the 2006 International Conference on Digital Telecommunications, 31, IEEE Computer Society, (2006).
- C43. **P. Baldi**. "Chemoinformatics, Drug Design, and Systems Biology", Proceedings of the 16th International Conference on Genome Informatics, (GIW 2005),Yokohama, Japan, December 2005.
- C42. **P. Baldi**. "Exploring Chemical Space with Computers: Informatics Challenges for AI and Machine Learning", BIOMAT V Symposium, Petrópolis, Brazil, December 2005, World Scientific Co.Pte.Ltd.
- C41 **P. Baldi** and L. Itti. "Attention: Bits versus Wows", *Proceedings of the International Conference on Neural Networks and Brain* (ICNN&B05), Edited by Mingsheng Zhao and Zhongzhi Shi, Vol. 1, PL-56-PL61, Beijing, China, IEEE Press, October 2005.
- C40. L. Itti and **P. Baldi**. "Bayesian Surprise Attracts Human Attention", *Advances in Neural Information Processing Systems* 18, NIPS 2005, Y. Weiss, B. Scholkopf, and J. Platt Editors, MIT Press, Cambridge, MA, 547--554, (2006).
- C39. R. Jurdak**, P. Baldi**, and C. V. Lopes. "State-Driven Energy Optimization in Wireless Sensor Networks," Proceedings of the IEEE/IEE International Conference on Sensor Networks (SENET), 356-363, Montreal, Canada. August, 2005.
- C38. R. Jurdak, **P. Baldi**, C. V. Lopes. "Energy-Aware Adaptive Low Power Listening for Sensor Networks", In Proceedings of the 2nd International Workshop on Networked Sensing Systems (INSS).). pp. 24-29, ISBN 0-9743611-2-7, San Diego, CA. June 2005.
- C37. R. Jurdak, C. V. Lopes, and **P. Baldi**. "Programming Model for Supporting Complex Optimizations in Sensor Networks," Proceedings of the Fourth International IEEE/ACM Conference on Information Processing in Sensor Networks (IPSN'05). Los Angeles, CA. April, 2005.
- C36. J. Cheng and **P. Baldi**. Three-Stage Prediction of Protein Beta-Sheets by Neural Networks, Alignments, and Graph Algorithms. Proceedings of the 2005 Conference on

Intelligent Systems for Molecular Biology, ISMB 05. *Bioinformatics*, Supplement 1, i75-84, (2005).

- C35. S. J. Swamidass, J. Chen, P. Phung, J. Bruand, L. Ralaivola, and **P. Baldi**. Kernels for Small Molecules and the Prediction of Mutagenicity, Toxicity, and Anti-Cancer Activity. Proceedings of the 2005 Conference on Intelligent Systems for Molecular Biology, ISMB 05. *Bioinformatics*, 21, Supplement 1, i359-368, (2005).

- C34. R. Jurdak, C. V. Lopes, **P. Baldi**. Adaptive Low Power Listening for Sensor Networks. 11ᵗʰ Annual International Conference on Mobile Computing and Networking.

- C33. L. Itti, **P. Baldi**. A Principled Approach to Detecting Surprising Events in Video. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2005.

- C32. L. Ralaivola, L. Wu, and **P. Baldi**. SVM and Pattern-Enriched Common Fate Graphs for the Game of GO. European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, April 2005. Accepted for Oral Persentation.

- C31. A. Mandal, C. V. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and **P. Baldi**. Beep: 3D Indoor Positioning Using Audible Sound. *Proceedings of the 2004 IEEE Consumer Communications and Networking Conference*, 348—353, (2004).

- C30. **P. Baldi**, J. Cheng, and A. Vullo. Large-Scale Prediction of Disulphide Bond Connectivity. *Advances in Neural Information Processing Systems* 17, NIPS 2004, in press, (2004).

- C29. R. Jurdak, C. Videira Lopes, and **P. Baldi**. An Identfication Scheme for Location Systems Using Robust Acoustic Signals. Proceeding of the ACS/IEEE International Conference on Pervasive Services, Beirut, Lebanon. IEEE Computer Society Press, (2004).

- C28. G. Pollastri, **P. Baldi**, A. Vullo, and P. Frasconi. Prediction of Protein Topologies Using Generalized IOHMMs and RNNs. *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer Editors, Volume 15, MIT Press, Cambridge, MA, (2003).

- C27. G. Pollastri and **P. Baldi**. Prediction of Contact Maps by GIOHMMs and Recurrent Neural Networks Using Lateral Propagation From All Four Cardinal Corners. Proceedings of the 2002 Conference on Intelligent Systems for Molecular Biology, ISMB 02. *Bioinformatics*, 18, Supplement 1, S62-S70, (2002). Nominated for Best Paper Award.

- C26. L. De Nardis, **P. Baldi**, and M. G. Di Benedetto. UWB Ad-Hoc Networks. Proceedings of the 2002 IEEE Conference on Ultra Wide Band Systems and Technologies (UWBST 2002), Baltimore, MD, 219-224, (2002).

- C25. . G. Pollasti, **P. Baldi**, P. Fariselli, R. Casadio. Improved Prediction of the Number of Residue Contacts in Proteins by Recurrent Neural Networks. *Proceedings of the 2001 Conference on Intelligent Systems for Molecular Biology, (ISMB01)*, Copenhagen, Denmark, AAAI Press (2001).

- C24. M. G. Di Benedetto and **P. Baldi**. A Model for Self-Organizing Large Scale Wireless Networks. International Symposium on 3G Infrastructure and Services. Athens, Greece, July 2001.

- C23. **P. Baldi**, G. Pollastri, C. A. F. Andersen, and S. Brunak. Matching Protein β-Sheet Partners by Feedforward and Recurrent Neural Networks. *Proceedings of the 2000 Conference on Intelligent Systems for Molecular Biology, (ISMB00)*, La Jolla, CA, 25-36, AAAI Press (2000).

- C22. S. Hampson, **P. Baldi**, D. Kibler, and S. Sandmeyer. Analysis of Yeast's ORFs Upstream Regions by Parallel Processing, Microarrays, and Computational Methods. *Proceedings of the 2000 Conference on Intelligent Systems for Molecular Biology, (ISMB00)*, La Jolla, CA, 190-201, AAAI Press (2000).

- C21. **P. Baldi**, G. Pollastri, C. A. F. Andersen, and S. Brunak. Protein β-Sheet Partner Prediction by Neural Networks. In: Artifical Neural Networks in Medicine and Biology. *Proceedings of the ANNIMAB-1 Conference*, Göteborg, Sweden, H. Malmgren, M. Borga and L. Niklasson Editors, Springer Verlag, (2000).
- C20. **P. Baldi**, S. Brunak, Y. Chauvin, and A. Gorm Pedersen. Computational Applications of DNA Structural Scales, *Proceedings of the 1998 Conference on Intelligent Systems for Molecular Biology, (ISMB98)*, Montreal, Canada, 35-42, AAAI Press (1998).
- C19. **P. Baldi**, A. Gorm Pedersen, S. Brunak, and Y. Chauvin. Characterization of Prokaryotic and Eukaryotic Promoters Using Hidden Markov Models, *Proceedings of the 1996 Conference on Intelligent Systems for Molecular Biology (ISMB96)*, Washington University, Saint Louis, 182-191, AAAI Press, (1996).
- C18. **P. Baldi**, S. Brunak, Y. Chauvin, and A. Krogh. Hidden Markov Models for Human Genes: Periodic Patterns in Exon Sequences, *Theoretical and Computational Methods in Genome Research*, S.Suhai Editor, Plenum, NY, (1996).
- C17. **P. Baldi** and K. Hornik. Universal Approximation and Learning of Trajectories Using Oscillators, *Advances in Neural Information Processing Systems 8*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo Editors, MIT Press, Cambridge MA, 451-457, (1996).
- C16. **P. Baldi** and Y. Chauvin. Protein Modeling with Hybrid Hidden Markov Model/Neural Network Architectures, *Proceedings of the 1995 Conference on Intelligent Systems for Molecular Biology (ISMB95)*, Cambridge (UK), 39-47, (1995).
- C15. **P. Baldi**, Y. Chauvin, S. Brunak, J. Engelbrecht, and A. Krogh. Periodic Sequence Patterns in Human Exons*, Proceedings of the 1995 Conference on Intelligent Systems for Molecular Biology (ISMB95)*, Cambridge (UK), 30-38, (1995).
- C14. P. Smyth, U. Fayadd, P. Perona, M. Burl, and **P. Baldi**. Inferring Ground Truth from Subjective Labelling of Venus Images, *Advances in Neural Information Processing Systems*, G. Tesauro, D. S. Touretzky and T. K. Leen Editors, MIT Press, Cambridge MA, 1085-1092, (1995).
- C13. **P. Baldi** and Y. Chauvin. Trading Decision Learning: from Theory to Personal Traders, *Proceedings of the Second International Conference on Neural Networks in the Capital Markets*, Pasadena, CA (1994).
- C12. **P. Baldi**, S. Brunak, Y. Chauvin, J. Engelbrecht, and A. Krogh. Hidden Markov Models of Human Genes, *Advances in Neural Information Processing Systems*, J. D. Cowan, G. Tesauro and J. Alspector Editors, Morgan Kauffmann Publishers, 761-768, (1994).
- C11. **P. Baldi** and Y. Chauvin. Discrimination of Tyrosine and Serine/Threonine Kinase Sub-Families by Hidden Markov Models, *Proceedings of The Third International Conference on Bioinformatics and Genome Research*, World Scientific Publishing Co., New Jersey, (1994).
- C10. **P. Baldi** and Y. Chauvin. Modeling Protein Families and Human Genes: Hidden Markov Models and a Little Beyond, *Proceedings of the 1994 Fifth Generation Computing Symposium, Workshop on Fusion of Molecular Biology and Knowledge Processing*, (Tokyo, Japan), (1994).
- C9. **P. Baldi**, Y. Chauvin, T. Hunkapiller, and M. A. McClure. Hidden Markov Models in Molecular Biology: New Algorithms and Applications, *Advances in Neural Information Processing Systems*, S. J. Hanson, J. D. Cowan and C. Lee Giles Editors, Morgan Kauffmann Publishers, 747-754, (1993).
- C8. **P. Baldi** and Nikzad Toomarian. Learning Trajectories with a Hierarchy of Oscillatory Modules, *Proceedings of the 1993 IEEE International Conference on Neural Networks*, San Francisco, CA, Volume III, 1172-1176, and also, in similar form, in *Proceedings of the 1993 European Symposium on Artificial Neural Networks* (ESANN93), Brussels, 183-188.

- C7. **P. Baldi** and Y. Chauvin. Trading Decision Learning, *Neural Networks for Computing Conference* (abstract), Snowbird, UT, (1992).
- C6. **P. Baldi**. A Modular Hierarchical Approach to Learning, *Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks*, (Iizuka, Japan), Vol. 2, 985-988, (1992).
- C5. **P. Baldi**. Computing with Arrays of Bell-Shaped and Sigmoid Functions. Bernstein Polynomials, the Heat Equation and Universal Approximation Properties, *Proceedings of the 1990 Conference on Neural Information Processing Systems*, Denver, CO, Morgan Kauffman Publishers, 735-742, (1991).
- C4. **P. Baldi**, Y. Rinott, and C. Stein. On the Distribution of the Number of Local Minima of a Random Function on a Graph, *Proceedings of the 1989 Conference on Neural Information Processing Systems*, Denver, CO, Morgan Kauffman Publishers, 727-732, (1990).
- C3. **P. Baldi**. Neural Networks and Principal Component Analysis: Landscapes and Algorithms, *Proceedings of the 1988 Conference on Neural Information Processing*, Denver, CO, Morgan Kauffman Publishers, 65-72, (1989).
- C2. **P. Baldi** and S.S. Venkatesh. On Properties of Networks of Neuron-Like Elements: Complexity and Capacity, *Proceedings of the IEEE Conference on Neural Information Processing Systems* (Denver 1987), published by the American Institute of Physics, (D.Z. Anderson, Editor).
- C1. **P. Baldi** and E.B. Baum. Caging and Exhibiting Ultrametric Structures, *Proceedings of the Conference on Neural Networks for Computing*, Snowbird, Utah, April 1986, published by the American Institute of Physics, 35-40, (1986), (John S. Denker, Editor).

Recent Invention Disclosure and Patent Applications

- Invention "Neural Network with Modular Hierarchical Learning" (filed by NASA).
- Invention "Structure-Based Drug Design of Anti-Tuberculosis Drug Targeted At Acyl-CoA Carboxylase of Mycobacterium Tuberculosis" (filed by UCI).
- Invention "Method of Treatment for Microbacterium Tuberculosis." PCT/US07/61859 (UC Case 2006-265-1)
- Invention "Human Genome Bioinformatics, Probabilistic Modeling of Biological Data", (Disclosure, UCI Office of Technology Alliances).
- Invention "Methods and Composition for Selecting and Using Single Nucleotide Polymorphisms" UC Case No. 2006-321-1
- Invention "Code Search: a Retrieval System for Software Programs", (Disclosure UCI Office of Technology Alliances). UC Case No. 2006-324-1
- Invention "Methods and Implementations for Rapid and Accurate Searches of Chemical Compounds", (Disclosure UCI Office of Technology Alliances). UC Case No. 2007-107-1
- Invention "Methods and Implementations for Improving the Retrieval of Chemical Compounds", (Disclosure UCI Office of Technology Alliances). C Case No. 2007-319-1
- Invention "Organic Chemistry Synthesis Explorer: Reaction Problem Generator and Tutorial System", (Disclosure UCI Office of Technology Alliances). UC Case No. 2007-378-1
- Invention "Methods and Implementations for Storing Sparse Vectors with Applications in Chemoinformatics and Other Areas". UC Case No. 2007-793-1
- Invention "Data Structures and Compression Methods for Biological Sequence Data". UC Case No. 2009-432-1

INVITED TALKS, TUTORIALS, PRESENTATIONS (Abridged Last Five Years)
- University of Florida, Gainesville
- MIT
- Genome Therapeutics
- University of Madrid, Spain
- University of Bologna, Italy
- Bioinformatics School, San Miniato, Italy
- Systems Biology Conference, Caltech
- ISMB Conference
- JOBIM Conference in Toulouse, France
- University of Paris VI, Paris, France
- Ecole Normale Superieure, Paris, France
- Fred Hutchinson Cancer Research Center
- University of Rome, Italy
- Interface Conference
- University of Rio de Janeiro, Brazil
- VII Brazilian Symposium on Neural Networks, Recife, Brazil
- Institute for Mathematics and its Applications (IMA), University of Minnesota
- University of California, San Diego
- RFIA Conference (Reconaissance des Formes et Intelligence Artificielle) , Toulouse, France
- Joint Symposium on Neural Computation
- International Joint Conference on Neural Networks

- University of Louisiana, Lafayette
- University of Algarve, Faro, Portugal
- IEEE CSB (Computational Systems Bioinformatics), Stanford
- University of Arizona, Tucson
- Genentech
- UCSD
- Invited Tutorial in Bioinformatics, GENSIP 2005, IEEE International Workshop on Genomic Signal Processing and Statistics, Newport, RI, (May 2005)
- Keynote Speaker, RFIA Conference (Reconnaissance des Formes et Intelligence Artificielle) , Nice, France (June 05)
- Keynote speaker, IJCNN, Montreal, Canada (August 05)
- Keynote speaker, Computer Vision and Pattern Recognition Conference (CVPR), San Diego, CA (July05)
- Invited speaker The Chinese University of Hong Kong (Distinguished Lecture Series)
- Keynote Speaker, International Conference on Neural Networks and Brain, Beijing, Chain (October 2005)
- Invited speaker NIHS Toxicology Division, Tokyo, Japan
- Invited speaker, Kyoto University (December 05)
- Keynote speaker, Second International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics Crema, ITALY -- September 15-17, 2005
- Keynote Speaker, Bioinformatics Conference, Atlanta, GA (November 2005)
- Keynote speaker BIOMAT Petropolis, Brazil (December 05)
- Keynote speaker for the 16th International Conference on Genome Informatics (GIW 2005) December 19-21, Yokohama, Japan.
- Invited speaker, Indiana University, Bloomington
- Invited speaker, Mitre Corporation
- Keynote speaker, Gubelkian Institute of Sciences, Portugal (June 06)
- Invited speaker, University of Rome, Italy (June 06)
- Invited speaker, University of Lisbon, Portugal (June 06)
- Chemoinformatics Tutorial ISMB (Intelligent Systems for Molecular Biology) Conference, Fortaleza, Brazil (August 2006)
- Invited speaker, University of Naples, Italy (October 06)
- Invited speaker, University of Benevento, Italy (October 06)
- Invited speaker, Iowa State University (March 2007)
- Keynote speaker 2007 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, Hawaii, April 2007-01-10
- Keynote speaker, Italian Conference on Bioinformatics, Naples, Italy (April 2007)
- Keynote speaker Physical and Chemical Foundations of Bioinformatics Methods, June 18-22, 2007, Dresden, Germany.
- Invited keynote speaker, Conference on Physical and Chemical Foundations of Bioinformatics Methods, Dresden, 2007
- Invited speaker, Iowa State University, 2007.
- Invite keynote speaker for the IEEE 2007 Symposium on Computational Intelligence in Bioinformatics and
- Computational Biology (CIBCB) , Honolulu, Hawaii, April 1-5, 2007.
- Invited keynote speaker Italian Bioinformatics Conference, Naples, April 26-28, 2007.

- Invited speaker workshop on Interdisciplinary Strategic Issues in e-Science and Cyber-Infrastructure, Caltech, Pasadena, California, June 13 - 14, 2007.
- Invited speaker, University of Naples, June 2007.
- Invited speaker, University of Benevento, June 2007.
- Invited keynote speaker, 2007 International Conference on Machine Learning and Applications (ICMLA'07), Cincinnati, Ohio, December 13-15, 2007.
- Invited distinguished speaker, Shumaker Bioinformatics Seminar of Informatics Institute at University of Missouri Columbia,
- Invited keynote speaker, BioSys Symposium on Human Post-Genomics, Evolution and the Future of Life held at the Royal Library, Copenhagen, Denmark, May 30th, 2008.
- Invited speaker, Biogen, Italy, June 2008.
- Invited speaker, University of Salerno, Italy, June 2008.
- Invited keynote speaker, 2008 International Conference on Artificial Neural Networks (ICANN 2008), Prague, Czech Republic, September 3-6, 2008.
- Invited member of the scientific committee of the 2nd International Workshops on Practical Applications of Computational Biology and Bioinformatics, Salamanca, Spain, October 2008.
- Distinguished speaker of the Shumaker Bioinformatics Seminar of Informatics Institute at University of Missouri Columbia (October 2007).
  Invited speaker, University College, Dublin, Ireland, (2008)
- Invited speaker, BioSys Symposium on Human Genomics, Evolution and the Future of Life
- Friday, May 30th, 2008 The Queen's Hall The Royal Library, Slotsholmen, Søren Kirkegaards Plads 1, DK - 1214 Copenhagen, Denmark.
- Invited speaker, Safra Program Distinguished Lectures Series, Tel Aviv University, Tel Aviv, Israel (March 1, 2010)
- Invited plenary speaker, Clinical Genomic Analysis Workshop, IBM Research, Haifa, Israel, March 2, 2010.
- Invited speaker, Hebrew University, Jerusalem, Israel, March 2, 2010.
- Invited speaker, University of Alberta, Alberta Ingenuity Centre for Machine Learning, Edmonton. Canada.

FUNDING

Grants (Abridged/Last Three years)

FUNDING SOURCE	ROLE	AMOUNT	DATES
NIH-NLM	PI	$4,664,422	7/1/02-6/30/07
UC Systems BREP	PI	$354,400	7/1/02-6/30/04
NIH-NCI	Co-PI	$994,404	5/1/02-4/30/03
Newkirk Center for Science and Society	PI	$16,000	3/1/02-2/28/03
Industry-University Cooperative Research Program	PI	$15,000	7/1/02-6/30/03
Subcontract from UCLA (NIH)-30896	PI	$30,000	9/1/01-8/31/03
UC Systems BREP	PI	$5,000	7/1/02-1/31/03
Calit2	PI	$4,000	7/1/02-6/30/03
NSF	Co-PI	$34,080	4/1/01-3/31/03
Center for Virus Research	PI	$6,000	7/1/02-6/30/03

Calit2 (fellowship for Marty Brandon)	PI	$10,000	7/1/02-6/30/03
Subcontract from Columbia Univ(NIH)	Co-PI	$250,670	9/15/02-6/30/04
Subcontract from UCLA(NIH)-31150	PI	$179,075	9/30/02-8/31/07
NIH	Investigator[1]	$8,552	5/01/03-4/30/04
Calit2	PI	$4,000	7/1/03-6/30/04
NSF EIA-0321390	PI	$269,989	10/1/03-9/30/06
NSF EIA-0321390 equipment cost sharing from Cal IT2, ICS, IGB, and RGS	PI	$115,714	10/1/03-9/30/06
NSF EF-0330786	Co-PI	$4,939,337	9/1/03-8/31/08
NIH	Investigator[2]	$22,725	7/1/03-6/30/05
Industry/UC Coop Research Initiative	PI	$5,000	1/26/04-6/30/05
Industry-University Cooperative Research Program	PI	$9,225	7/1/04-6/30/05
NIH AI-56464	Investigator[3]	$14,541.35	8/1/03-1/31/05
NIH AI-65359	Investigator[4]	$24,087	5/20/05-4/30/06
NIH AI-61363	Investigator[3]	$28,699	9/1/04-8/31/06
NSF IIS-0513376	PI	$293,538	7/15/05-6/30/08
NSF IIS-0514810	PI	$170,848	9/1/05-8/31/08
NIH CA112560	Co-PI	$1,809,537	8/1/05-5/31/10
Subcontract from UCSD 2005-3233 (DOD)	PI	$123,439	10/1/05-9/29/06
Opportunity Grant from Industry-University Cooperative Research Program	PI	$11,318	7/1/06-12/15/06
NIH GM079413	Investigator[5]	$28,548	3/12/07-2/28/09
UC Discovery from the Industry-University Cooperative Research Program and CODA Genomics, Inc.	Co-PI	$1,677,139	9/1/07-8/31/09
NSF CCF-0725370	Co-PI	$693,180	9/1/07-8/31/10
NIH-NLM 07443 (renewal)	PI	$5,644,874	7/1/07-6/30/12
NIH GM-076516	Investigator[6]	$11,532	8/1/07-7/31/12
NIH P50 GM076516	Investigator[7]	$10,000	5/1/08-7/31/08
NIH AI61363	Investigator[8]	$40,000	9/1/08-8/31/09
Camille and Henry Dreyfus Foundation	PI	$ 31,000	3/1/09-2/28/11

[1]PI (Hoda Anton-Culver), amount reflects total funding for my group only.

[2]PI (Suzanne Sandmeyer), amount reflects total funding for my group only.

[3]PI (Phil Felgner), amount reflects total funding for my group only.

[4]PI (Alan Barbour), amount reflects total funding for my group only.

[5]PI (Klemens Hertel), amount reflects first year funding for my group only.

[6]PI (Arthur Lander), amount reflects first year funding for my group only for the computational core.

[7]PI (Arthur Lander), amount reflects first year funding for Ivan Chang in my group.

[8]PI (Philip Felgner), amount reflects first year funding for my group only.

Gifts

Company	Amount	Received
Sun Microsystems	$5,000	10/02
Sun Microsystems	$47,000	10/02

Illumina	$1,000	9/03
Eli Lilly and Company	$5,000	9/04
Amgen	$1,000	9/04
Neurocrine Biosciences Inc.	$400	9/04
Fisher Scientific	$300	9/04
ImmPORT Therapeutics	$2,500	8/05
ImmPORT Therapeutics	$2,500	12/05
CODA Genomics	$5,000	8/06
Microsoft	$45,000	12/06
ImmPORT Therapeutics	$5,000	12/07
ImmPORT Therapeutics	$5,000	03/08
Antigen Discovery	$5,000	09/08
Antigen Discovery	$7,000	12/08

Paid Consultancies
- Banque de France
- TI Capital
- SmithKline-Beecham
- Paracel
- DE Shaw
- Calspan
- Ecobalance/Dames and Moore
- Colosseum Fund
- Banca del Salento
- Elitra
- Allergan
- Genome Canada
- San Antonio Life Sciences Institute
- Fundacion BBVA, Spain
- Italian Ministry for Education University and Research (MIUR)
- Pennsylvania Department of Health's Interim Performance Review
- MIT Press
- Cambridge University Press
- Wiley
- Mitre Corporation
- Center for Biological Sequence Analysis, DTU, Denmark
- Max Planck Institute for Biological Cybernetics
- Christian Doppler Forschungsgesellschaft, Austria
- Danish Agency for Science, Technology and Innovation
- Strategy and Funding Directorate at the Health Research Board, Ireland.

Other Activities

- Advisory Board, DNA Microarray Core, UCI.
- Associate Editor *Neural Networks.*
- Associate Editor *Data Mining and Knowledge Discovery*.
- Editorial Board *International Journal for Bioinformatics Research and Applications*.
- Editorial Board *Chemistry Central Journal*.
- Scientific Advisory Board, Max Plank Institute for Biological Cybernetics, Tubingen, Germany.
- Reviewer for several journals including Physical Review Letters, Neural Computation, Science, Journal of Computational Biology, Journal of Molecular Biology, Protein Science, CABIOS, IEEE Transactions on Information Theory, and IEEE Transactions on Neural Networks.
- Organizer of session "DNA Structure and Protein-DNA Interactions" at the Pacific Symposium on Biocomputing, Hawaii, January 3-7, 2001.
- Organizer of session "Bioethics and Fiction Science" at the Pacific Symposium on Biocomputing, Hawaii, January 3-7, 2001.
- On the organizing committee of the 33rd Symposium on the Interface of Computer Science and Statistics (Interface 01), Frontiers in Data Mining and Bioinformatics, June 13-16, Costa Mesa, CA, (2001). Organizer of special one-day workshop on Bioinformatics. Presentation of a tutorial on Bioinformatics.
- On the program committee of the 1987 to 1992 Conferences on Neural Information Processing (Denver, CO), of the 1990 and 1993 International Joint Conference on Neural Networks (Washington, D.C. and Portland, Oregon), and of the 1993-2008 Learning Conference (Snowbird, UT).
- Co-founder of Net-ID, Inc. in 1991.
- Member of ACM, IEEE, AAAI, AAAS, and ACS.
- Member of the International Society for Computational Biology
- Member Institute for Mathematical Behavioral Sciences, UCI
- Scientific Advisory Board Max Planck Institute for Biological Cybernetics, Tubingen
- Scientific Advisory Board of Sapient Discovery

Society Memberships

- Member of the American Association for the Advancement of Science (AAAS).
- Member and Fellow of the Association for the Advancement of Artificial Intelligence (AAAI).
- Member of the American Chemical Society (ACS).
- Member of Association for Computing Machinery (ACM).
- Senior Member of the Institute of Electrical and Electronic Engineers (IEEE) .
- Member of the International Society for Computational Biology (ISCB).

Current PhD Students, Programmers, and Postdoctoral Fellows
- Alessio Andronico
- Chloe Azencott
- Ivan Chang
- Kenneth Daily
- David Foshee
- Sholeh Forouzan
  Jordan Hayes
- Mat Kayala
- Christophe Magnan
- Ramzi Nasr
- Arlo Randall
- Paul Rigor

Recently Graduated PhD Students, Postodctoral Fellows, and Researchers

- Pierre-François Baisnée, Institut de Recherche pour le Dévelopment, France
- Ryan Benz, Applied Proteomics
- Martin Brandon, UCI
- Jonathan Chen, UCI
- Jianlin Cheng, University of Missouri
- Yimeng Dou, Verdezyne
- Steven Hampson, Deceased
- Raja Jurdak, CSIRO, Brisbane, Australia
- Erik Linstead, Boeing and Chapman University
- Antonio Maratea, University of Naples-Parthenope, Italy
- Gianluca Pollastri, University College Dublin, Ireland
- Liva Ralaivola, University of Provence/Aix-Marseille I, France
- Alex Sadovsky, Defense Logistics Agency
- Suman Sundaresh, NextBio
- S. Joshua Swamidass, Washington University, Saint Louis
- Mike Sweredoski, Caltech
- Eric Wang, Affymetrix
- Lin Wu, SAS
-

UCI Academic Service (examples)
- Conflict of Interest Oversight Committee. Received Certificate of Service with Distinction.

- Privilege and Tenure Committee
- Donald Bren Hall space committee.
- Founder new ORU (Organized Research Unit): Institute for Genomics and Bioinformatics (IGB).
- Committee for Statistics.
- Organizer of annual IGB Symposium and annual IGB Distinguished Speaker Series
- Helped develop and create new concentration: Informatics in Biology and Medicine
- In the process of developing new BS: Bioinformatics: Premedical

  .