

# CURRICULUM VITAE

**Simon Kasif, Ph.D.**

**Professor of Biomedical Engineering and Bioinformatics  
Director, (AI2BIO Laboratory)  
Former Co-Director Center for Advanced Genomic Technology  
(CAGT)  
Boston University  
44 Cummington St.  
Boston, MA 02215  
Phone: 617 358 1845 Email: [kasif@bu.edu](mailto:kasif@bu.edu) Fax: 617-353-6766  
<http://sites.bu.edu/phenogeno>**

## **Brief Biosketch:**

Dr. Kasif contributed, co-lead or lead several high impact and/or pioneering projects, methodologies and systems across a wide range of applications in both academia and industry. Dr. Kasif was co-founder and co-director of the Center for Advanced Genomic Technology (CAGT) with Dr. Charles DeLisi, co-founder of COMBEX: COMPUTATIONAL BRIDGE TO EXPERIMENTS PROJECT with Drs. Richard J. Roberts and Martin Steffen, [combrex.bu.edu](http://combrex.bu.edu), and a founding member of the I2B2 Center (Informatics for Integrating Biology and Bedside) ([www.i2b2.org](http://www.i2b2.org), PI: Dr. Isaac Kohane) at Harvard Partners. For the I2B2 consortium he co-lead the Diabetes Biological Driving Projects. He also directed the regional Diabetes Systems Biology and Bioinformatics Core at Harvard Medical School and Boston University as part of a National Diabetes Center. His group contributed to the Human Genome Consortium that produced the first public draft of the human genome. His contributions to Computational Biology include the co-development of the early versions of multiple pioneering systems and frameworks that include: Glimmer, one of the most widely used microbial gene finders in the community (with Drs. Steven Salzberg, Arthur Delcher and Owen White), Mummer, a widely used open access whole genome microbial comparative alignment system (with Drs Salzberg and Delcher), early advocacy and popularization of the broad use of Bayesian probabilistic networks (graphical models), the leading methodology for biological network and data integration (starting 1992), introduction of AI based probabilistic functional linkage networks (with Dr. Stan Letovsky in 2003, at Boston University (BU)) - pioneering AI based network based protein function prediction, pioneering (graph theory based) biological context networks (with Noga Alon et al 2005) deployment of AI based network signatures of disease for diabetes and insulin resistance (with Drs. Isaac Kohane, Ronald Kahn and Manway Liu in 2007, at BU and Harvard Medical School). Last but not last, in 1992-1994 Dr. Kasif and his AI group at Johns Hopkins University were among the earliest pioneers that advocated the application of AI based graphical causal models for biology generally and synthetic biology (directed evolution) in particular. Most recently, together with Richard Roberts he co-pioneered a one of a kind project integrating AI, Community Science, Active Learning to automatically

drive and validate discovery in biological sciences. Earlier in his academic career he founded the AI Laboratory at Johns Hopkins University where he and his collaborators produced a number of theoretical, methodological AI and ML advances and widely used AI systems.

### **Detailed Summary:**

Prof. Kasif genomics, systems biology and AI2BIO groups have been developing advanced network and general AI based predictive methodologies for protein function prediction, modeling and analysis of metabolic and signaling networks and their dis-regulation in diabetes, biomarker discovery, obesity, aging, wellness and health span. He is also involved in biotechnology development. He has an exceptionally strong track record in establishing successful collaborations with both senior biologists and computational scientists. Since arrival in Boston he directly collaborated with over 20 distinct large laboratories in the Boston area in multiple institutions. These are **direct** collaborations in basic research independent of his multi-investigator and consortium work in several projects that include hundreds of scientists. Dr. Kasif has been building numerous bridges to experimental validation and testing of computational predictions. He has recently worked, consulted or advised a number of companies that deploy genomic and data science technology for clinical applications or new technology development. Most recently he returned to AI consulting.

Previously in computer science he and his collaborators pioneered several concepts, algorithms and widely used systems in machine learning, deductive databases, Artificial Intelligence and high-performance parallel computing. He is not a novice to BIG Data as he co-designed and co-implemented the first working **parallel** deductive database in 1983 and received the Terabyte performance award for **distributed data mining** (with Dr. Bob Grossman in 1998). The parallel deductive database / logic programming system pioneered by Drs. Simon Kasif, Madhur Kohli and Minker preceded the major investment of Japan in the Fifth Generation Project. Lewis Stiller, Kasif's PhD student at Johns Hopkins produced one of the most striking demonstration of the power of parallelism for **AI planning** and **large databases** by discovering chess end games that require 221 moves to capture a piece and win (using a Connection Machine architecture with 65000 processors). Kasif and his collaborators at Johns Hopkins University co-pioneered one of the earliest applications of **data streaming** in machine learning and **data mining** (with Drs. Heath, Salzberg, Kosaraju) and efficient probabilistic databases based on novel logarithmic time queries and updates in **graphical models** and Bayesian networks (with Drs. Judea Pearl, Adam Grove and Arthur Delcher). These novel algorithms were applied to the analysis of functional impact of coding mutations via in-silico mutagenesis (1993-1995).

Dr. Kasif received his BS in Mathematics from Tel Aviv University and a PhD in Computer Science (Artificial Intelligence) from University of Maryland where he was fortunate to work with two legendary AI scientists -- Profs. Azriel Rosenfeld – one of the fathers of **Computer Vision** and Jack Minker – the father of **Deductive Databases**. He held (overlapping) faculty positions at Johns Hopkins University (12 years), Princeton University (Computer Science), University of Illinois, University of Chicago, MIT Genome Center (Whitehead Institute), Boston University Biomedical Engineering (20 years), Children's Hospital (CHIP, Division of the MIT/Harvard Science and Technology Program, 14 years), Joslin Diabetes Center, Harvard Medical School and others. Prior to entering academic research, he served seven years in the Air Force.

Dr Kasif is currently a Professor of Biomedical Engineering, Bioinformatics and Computer Science at Boston University. He was the PI on the Regional Systems Biology (Bioinformatics) Core of the Diabetes Research Center funded by NIH at Joslin. He has been on the editorial board of numerous journals in computational biology, computer science, AI, Machine Learning and Personalized Medicine. He co-chaired major conferences and symposiums in the field, on the advisory board of several national centers and organizations, and was elected as a fellow of American Institute for Medical and Biological Engineers.

He introduced and co-chaired the Disease Models area at the ISMB conference, the major conference in Bioinformatics and Computational Biology (with Drs. Isaac Kohane and David Heckerman).

He is teaching several novel courses, emphasizing experiential student driven learning. Some Examples include AI and Systems Biology, Network Biology Disease, Systems Biology and Genomics, Next Generation Sequencing: Technology, Analysis and Applications. His educational philosophy aims to challenge the students to ask creative integrative questions, produce “crazy solutions”, examining real world medical, biological and technological problems that involve teamwork, scientific discovery tools, data science, technology, open access analysis and modeling tools, and broad familiarity with both science and technology. He also emphasizes ethical and inclusion/equality issues in biology, AI, computation and medicine. He is deeply committed to advancement of under-represented minorities and women in STEM disciplines.

Last but not least he advised and mentored a number of young scientists that ended up making significant contributions to the computational biology field and AI. The full list of scientists mentored can be found here (over 60). <http://sites.bu.edu/phenogeno/people/>. He is fully committed to diversity.

## **1. Citizenship: USA**

## **2. Education**

<u>Degree</u>	<u>Institution</u>	<u>Date</u>	<u>Major</u>
B.Sc.	Tel Aviv University	July 1979	Mathematics
M.S.	University of Maryland	May 1983	Computer Science
Ph.D.	University of Maryland	May 1985	Computer Science

(co-advised by Profs. Jack Minker and Azriel Rosenfeld)

## **3. Current and Past Research Interests**

Broad research interests: transformative, creative or disruptive methodologies for biological and biomedical problems by combining computational, analytical, biomedical and technological innovations.

Driving Biomedical Science and Engineering with Artificial Intelligence, Scientific Reproducibility, Intelligent Inspectors, Pheno-genomic profiling and Analysis, Genomic Systems Biology, Biotechnology, Network Signatures of Disease, Functional and Comparative Genomics, Health-care delivery, Insulin Resistance, Systems Biology of Wellness and Aging, Disease Prevention, Network Biology, Gene Function Prediction, Algorithms, Data Science, Machine Learning. Past: Artificial Intelligence: Theory and Systems, High Performance Parallel Computing.

## **4. Professional Experience**

August 2022- Visiting Professor, Tel Aviv University

Dec 2019- Jan 2020 Visitor Tel Aviv University

2017 – 2018: Member of the Boston University Presidential Advisory Committee on Trans-Disciplinary Research in Genetics, Epigenetics, Systems Biology and Public Health, Boston University.

2016 – 2017: Member of the Provost Advisory Committee on Data Science, Boston University.

Dec 2017 - Jan 2018: Visitor Safra Bioinformatics Center, Tel Aviv University

June 2015- 2016: Senior Iacocca Fellow (sabbatical) at Joslin Diabetes Center, Harvard Medical School

Sept 2011- 2018: Adjunct Investigator, Joslin Diabetes Center, Harvard Medical School and PI, Regional Systems Biology and Informatics Core of the Diabetes Research Center, Member Center Executive Committee.

June 2002-2013: Founding Co-Director (with Charles DeLisi), Center for Advanced Genomic Research (CAGT), Boston University.

Summer 2009: Visiting Professor, Israel Institute of Technology, Technion, Haifa

Summer 2008: Visiting Professor, Israel Institute of Technology, Technion, Haifa.

April 2005-2017: Staff Scientist, Children's Hospital, Harvard-MIT Division of Health Sciences and Technology (consulting position).

Fall 2005- co-PI and senior investigator, I2B2 National Center for Biomedical Computing, (Harvard Medical School, MIT, BU).

Oct 2000 – present: Prof. of Biomedical Engineering, Bioinformatics & Computer Science, Boston University.

Sept 1999 – Aug 2000: Research Scientist, Cambridge Research Labs, Lead Scientist Computational Biology, Compaq Computers and HP Research Labs (on academic leave).

Sept 1999 – Aug 03: Visiting Scientist, MIT Genome Center, Analysis Core, International Human Genome Consortium (1999-2000) (on academic leave).

Aug 1997 – May 99: Founding Associate Director, National Center for Data Mining, University of Illinois at Chicago, Professor, Dept. of Electrical Engineering and Computer Science, joint appointment in the Department of Bioengineering (since 1999).

Aug 1996 – May 99: Adjunct Faculty, Johns Hopkins University

July 1994 - Jun 95: Visiting Associate Professor, Computer Science Department, Princeton University, Princeton, NJ (sabbatical leave from JHU).

July 1994 - Consultant to the Vice President / President, NEC Research Institute, Princeton NJ.

Jun 1991 - Jun 96: Associate Professor, Dept of Computer Science, The Johns Hopkins University (joint appointment in Cognitive Science Department, Founding Member of the Mind-Brain Institute Task Force).

Sep 1987 - Jun 91: Assistant Professor, Dept of Computer Science, The Johns Hopkins University, joint appointment in Cognitive Science Center.

Sept 1987 – Co-founding faculty member of the Computer Science Department, Johns Hopkins University.

Jun 1987 - Aug 87: Visiting Scientist, Weizmann Institute of Science.

Sept 1985 – 1996: Founder of the AI Laboratory at Johns Hopkins University.

Sep 1985 - Sep 87: Assistant Professor, Dept. of Electrical Engineering and Computer Science, The Johns Hopkins University.

1973 –1980 Air Force

### **Selected OPEN ACCESS Research Systems Co-Design and/or Co-Development (more than twenty systems)**

#### **Computational Bridges to Experiments (COMBEX)**

Working closely with Sir Dr. Richard Roberts and Dr. Martin Steffen, Prof. Kasif co-founded a transformative Computational Bridges to Experiments (COMBEX) project and database to create an interdisciplinary community that aims to catalyze the broad translation of computational predictions into novel biology. This is a one of a kind project that deploys AI and Active Learning in Prioritizing Biological Experiments by a Community of Scientists. It's the first project where biological scientists were funded (via bids) to perform experimental validation and testing of high priority computational predictions selected by an Active Learning AI scheme. The initial project funded by NIH (2009) (PIs Simon Kasif, Richard J. Roberts Martin Steffen) focuses on biochemical gene function prediction and experimental testing. Prof. Kasif co-led the overall project (with Drs. Roberts, Steffen and multiple co-PIs) and directed the computational effort that involved development of a novel AI based experiment prioritization (for gene function determination in large protein families) based on active learning principles. In addition to the first of a kind active learning component, the project also produced a cost-efficient

database that hosted thousands of genomes and millions of genes, web portal, recruiting and storing predictions from multiple top teams, experimental bid processing and integration of numerous predictions and data sources. COMBREX is a novel instance of Amazon Turk and Citizen's Science projects.

### **Network Based Protein Function Prediction**

In the early 2000s Prof. Kasif and his group pioneered AI based network protein function prediction. This approach was influenced by AI ideas previously introduced in Computer Vision and Neural Networks and adapted in Computational Biology to protein and gene networks. These network propagation methodologies and their hundreds of variants have been improved by algorithmic ideas and become the leading framework in protein-function prediction. Multiple modern systems are currently using network based protein function prediction in papers published in prominent journals and benchmarking competitions. These systems are also used in prioritizing of gene targets in large genetic studies with potentially significant clinical implications.

### **Biological Context Networks**

In collaboration with Noga Alon and Charles Cantor Prof. Kasif and his group introduced the new framework of biological context networks (2005-). These networks allow to define multi-layer networks (each layer corresponds to a context such as tissue, cellular localization, stimuli, condition, environment and more). Connections are allowed within and across contexts. This framework has recently (last 10 years) become a large discipline that includes multi-layer networks and multiplex networks in physics and network science with numerous applications.

### **Bioinformatics System: GLIMMER**

Glimmer is one the **most widely** used gene discovery systems in microbial DNA sequences. Glimmer I was co-developed in joint work with S. Salzberg, O. White and A. Delcher. Glimmer is the first widely used system in bioinformatics to use variable length models pioneered by Ziv-Lempel for data compression and popularized in AI systems for Speech Analysis. Glimmer achieves over 98% accuracy on most bacterial genomes. Glimmer variants maintained and systematically maintained by Salzberg's group at JHU are routinely used around the globe to annotate the DNA sequences of hundreds of newly sequenced bacterial genomes that include major pathogens that cause Lyme disease, Meningitis, Tuberculosis or bacteria used for bio-energy applications. Glimmer and it's successors have been cited in thousands of publications and were used world-wide to

annotate thousands of microbial organisms. The bacterial enzymes **identified** by GLIMMER have endless applications in **medicine** (e.g., bacterial enzymes are used in CRISPR, **neuroscience** (optogenetics), biotechnology (e.g. clean energy), bacterial sensors and others. GLIMMER is based on machine learning principles previously introduced in speech recognition, natural language and AI.

### **Bioinformatics System for Whole Genome Comparison: Mummer**

Mummer is among the first open access **systems** for high-resolution comparison of closely related **whole microbial** genomes. The system was implemented by Dr. Arthur Delcher in joint work with Dr. Steven Salzberg and has been first used by the Institute for Genomic Research (TIGR) to perform whole-genome alignment of two different strains of microbial organisms (1998) searching for polymorphisms that contribute to **virulence, antibiotic resistance or pathogenicity**. Mummer was also used for the synthetic biology and minimal organism project at TIGR and foreign DNA detection. Other applications include **vaccine development**, SNP detection and comparison of mouse and human DNA, genomic re-sequencing and other. The Mummer System has been extended and its multiple successors have thousands of citations and users.

### **Graphical Models (Bayes Networks) and Molecular Biology**

In 1992 Prof. Kasif and his group at Johns Hopkins initiated one of the earliest efforts to apply Graphical Models and Bayes networks for analysis and modeling of molecular biological data. This work demonstrated that causal Bayes networks can be usefully deployed to model biological data and biological constraints, perform in-silico simulations and predict results of mutagenesis experiments. Today, Bayesian networks are among the most popular frameworks in systems biology, structural biology, integrative biology and other predictive biology applications. One of the key applications is integration of computational predictions from multiple biological data sources. (Kasif et al 2000). In 2003 jointly with Stan Letovsky, Prof. Kasif co-introduced and helped to disseminate the methodology of probabilistic functional linkage networks in the form of Markov Random Fields used to predict biological function for newly sequenced genes without any known annotation. AI variants of graphical models for proteins and DNA sequences advocated early by Kasif and others are now in common use by the community applying co-evolutionary analysis to proteins with multiple scientific and medical applications that include personalized medicine and analysis of mutations.

### **Integration of Networks and Genomic Data (Data Fusion)**

Prof. Kasif and his group were among the earliest efforts that helped popularize the **integration** of networks and genomic data starting from an early effort in 2001. Today data fusion and in particular integration of network and genomic data is a very active research area in molecular systems biology of all organisms used in genetics, genomics and systems biology of disease (cancer, autism and others).

## **Networks Signatures of Wellness via Network Anomaly**

The GNEA (Gene Network Enrichment System) implements integration of network and transcriptional data to produce unbiased network signatures of diabetes and insulin resistance (2007). Joint work with Manway Liu, Zak Kohane and Ron Kahn as part of the NIH Funded National Center for Biomedical Computing (I2B2). Insulin resistance is an important metabolic staging and risk biomarker for multiple diseases including Diabetes, Cardiovascular Diseases, Cancer and Alzheimer's among others. Insulin response networks also play an important role in longevity research in both model organisms and humans.

## **Distributed Data Mining:**

Dr. Kasif helped co-design a distributed data mining systems based on decision trees. Awarded High Performance Terabyte Challenge Award, Supercomputing 1998. (with R. Grossman and S. Bailey, et al). Dr. Kasif co-advised (with Dr. Kohane) several projects at Children's Hospital on Distributed Data Mining of Medical Records.

## **Genomic Technology: POMP- Pipette Optimized Multiplexed PCR:**

POMP is a new pooling procedure that helped the Institute for Genomic Research to close the gaps in the genome of Streptococcus Pneumonia. Used at TIGR (1999) and MIT Genome Center (2001). Joint work with Steven Salzberg, Herve Tettelin, Noga Alon, Richard Beigel and Lance Fortnow. The theoretical work helped start a new sub-area in Graph Theory, namely learning hidden graphs with pooled queries (with Noga Alon and Richard Beigel).

## **Multiplex PCR: MUPLEX**

MUPLEX was a first freeware for multiplexing biological assays. Applications include genotyping, pathogen detection, forensic PCR, disease diagnosis. The system was implemented by a former PhD student John Rachlin. Muplex had had hundreds of users. MUPLEX was also used to analyze multiplex PCR panels of the type deployed early on in prenatal DNA analysis in the pioneering efforts for detection of Down syndrome and other medical, diagnostic and biological applications. This work was part of the earliest Liquid Biopsy designs that became widely spread using broader genomic methods during the last 10 years.

## **High Performance Search and Planning**

The parallel AI system based on dynamic programming and super-efficient distributed storage finds chess end games that **require** more than **223 moves** to capture a piece and



win going way beyond human capability (1990). The system was implemented on the Connection machine 2 (over 64000 processors). This is considered a major extension of Dr. Ken Thompson's pioneering work at Bell Labs on 4 piece chess end-games. This pioneering system was designed and implemented by a Ph.D. student Lewis Stiller as part of his PhD thesis in the John Hopkins AI laboratory directed by Kasif. The results have been described in the popular press: New York Times (1991), London Times, Washington Post, Scientific American (1992).

### **Machine Learning System: OC1**

OC1 is a widely used open access data mining system based on decision trees. The system is readily available on the WEB and has been ftp-ed by thousands of users. Joint work with Steven Salzberg and S. Murthy (1993). One of the earliest systems to use sorting of attributes for scalability and more importantly randomization and random projections in practical and widely used decision tree systems.

### **Parallel deductive database system PRISM**

PRISM is one of the earliest and perhaps the first parallel (AI) deductive database systems that has been actually implemented and mapped to a multiprocessor, ZMOB. Joint work with Jack Minker, Madhur Kohli and others (1980-1983). This effort preceded the major Fifth Generation Computer Project in Japan that subsequently built a number of such machines.

## **6. Publications**

### **6.1 Books, Books Chapters and Special Reports**

1. S. Kasif and A. Delcher, "Analysis of Local Consistency in Parallel Constraint Networks", Principle and Practice of Constraint Programming, published by MIT press, editors Pascal van Henteryck and V.J. Saraswat, 1994.
2. Heath, D., S. Kasif and S. Salzberg, "Committees of Decision Trees", Cognitive Technology, North Holland Publishers, 1995
3. Simon Kasif and Stuart Russell (Eds.), Proceedings of the AAAI Fall Symposium on Learning Complex Behaviors, AAAI Press, 1996.
4. J. Flannagan, T. Huang, P. Jones, and S. Kasif, "Human Centered Systems: Information, Interactivity and Intelligence", Executive NSF Steering Committee Report, July, 1997.
5. R. Grossman, S. Kasif, J. Ullman, et al, "Perspectives on Data Mining", Executive Committee NSF Report of Four Workshops, 1998.
6. Kasif, S. and A. Delcher, "Biological Data Modeling using Probabilistic Networks", in Salzberg, S., D. Searls, and S. Kasif, "Computational Methods in Molecular Biology", Elsevier, Publ. 1998.

7. S. Salzberg, D. Searls and S. Kasif, eds, "Computational Methods in Molecular Biology", Elsevier Publ. 1998. (2<sup>nd</sup> Printing, February 1999). **One of the earliest Books on Computational Biology and Bioinformatics.**
8. Rich Roberts, Peter Karp, Simon Kasif and Stuart Linn, "An Experimental Approach to Gene Function", Executive Report, American Academy for Microbiology.
9. Satoru Miyano Jill Mesirov, Simon Kasif, Sorin Istrail Pavel Pevzner, Michael Waterman, Research in Computational Molecular Biology: 9th Annual International Conference, RECOMB 2005, Cambridge, MA, USA, May 14-18, 2005, Proceedings Lecture Notes in Bioinformatics.

## 6.2 Journal Publications

Note: In Computer Science and AI authors are typically listed **alphabetically** with the exception of PhD students.

10. Does AI Need to Understand Nature to Pass the Turing Test: A Review of AI2BIO", Simon Kasif, in review, 2022.
11. "Biology Must Be Programmed", Simon Kasif and Rich Roberts, in review, 2022.
12. Targeted in silico characterization of fusion transcripts in tumor and normal tissues via Fusion Inspector, B. Haas, Alexander Dobin, Mahmoud Ghandi, Anne Van Arsdale, Timothy Tickle, James T. Robinson, Riaz Gillani, Simon Kasif, Aviv Regev in review, 2022.
13. "On Professors and Ranks", Simon Kasif, in review, 2022.
14. "Is it the racquet or the player: The case for long-term vision in academic institutions", Simon Kasif, in review 2022.
15. "Anomalies and beyond", Simon Kasif, in review 2022.
16. "The Anti-Vanity Mirror", Simon Kasif, in review 2022.
17. "Network Biology and Medicine: The Old Emperor Cloths Fit Better than Ever", Simon Kasif, in review 2022.
18. "On Modularity: A New Perspective", Simon Kasif, in preparation.
19. "On Convergence of Computation and Biomedical Sciences", Simon Kasif, in preparation, 2022.
20. Interpretable network propagation with application to expanding the repertoire of human proteins that interact with SARS-CoV-2. Law JN, Akers K, Tasnina N, Santana CMD, Deutsch S, Kshirsagar M, Klein-Seetharaman J, Crovella M, Rajagopalan P, Kasif S, Murali TM. Gigascience. 2021 Dec 29;10(12):giab082. doi: 10.1093/gigascience/giab082 PMID: 34966926
21. High-throughput mediation analysis of human proteome and metabolome identifies mediators of post-bariatric surgical diabetes control. Dreyfuss JM, Yuchi Y, Dong X, Efthymiou V, Pan H, Simonson DC, Vernon A, Halperin F, Aryal P, Konkar A, Sebastian Y, Higgs BW, Grimsby J, Rondinone CM, Kasif S, Kahn BB, Foster K, Seeley R, Goldfine A, Djordjilović V, Patti ME. Nature Commun. 2021 Nov 29;12(1):6951. doi: 10.1038/s41467-021-27289-2. PMID: 34845204

22. We need to keep a reproducible trace of facts, predictions, and hypotheses from gene to function in the era of big data. Kasif S, Roberts RJ. *PLoS Biol.* 2020 Nov 30;18(11):e3000999. doi: 10.1371/journal.pbio.3000999. eCollection 2020 Nov. PMID: 33253151
23. Artificial Tikkun Olam: AI Can Be Our Best Friend in Building an Open Human-Computer Society, Kasif, S.arXiv:2010.12015, 2020
24. Analysis of brain region-specific co-expression networks reveals clustering of established and novel genes associated with Alzheimer disease. Lancour D, Dupuis J, Mayeux R, Haines JL, Pericak-Vance MA, Schellenberg GC, Crovella M, Farrer LA, Kasif S. *Alzheimers Res Ther.* 2020 Sep 2;12(1):103. doi: 10.1186/s13195-020-00674-7.
25. Identifying Human Interactors of SARS-CoV-2 Proteins and Drug Targets for COVID-19 using Network-Based Label Propagation, Jeffrey Law et al, aRxiv, 2020
26. Rate Estimation and Identification of COVID-19 Infections: Towards Rational Policy Making During Early and Late Stages of Epidemics, Richard Beigel, Simon Kasif, medRxiv 2020
27. Single-cell transcriptional networks in differentiating preadipocytes suggest drivers associated with tissue heterogeneity. Ramirez AK, Dankel SN, Rastegarpanah B, Cai W, Xue R, Crovella M, Tseng YH, Kahn CR, Kasif S. *Nature Commun.* 2020 Apr 30;11(1):2117. doi: 10.1038/s41467-020-16019-9.PMID: 32355218
28. In the Land of Risks: Nature vs Man Made Threats, S. Kasif, in review.
29. Evolution of Knowledge Networks: Towards Reproducibility, Provenance and Causal Paths in Biology and Medicine, Kasif S. Roberts, RJ., in review.
30. Novel mediation analysis of human plasma proteome and metabolome reveals mediators of improved glycemia after gastric bypass surgery Jonathan Dreyfuss, Yixing Yang, Hui Pan, Xuehong Dong, Donald C Simonson, Ashley Vernon, Pratik Aryal, Anish Konkar, Yinong Sebastian, Brandon W Higgs, Joseph Grimsby, Christina M Rondinone, Simon Kasif, Barbara B Kahn, Kathleen Foster, Allison Goldfine, Mary Elizabeth Patti,BioRxiv, 817494, 2019
31. A Perspective on Biological Context Networks: A Simple but Powerful Language for Network Systems Biology, S. Kasif in preparation.
32. In the Land of Risks: Nature vs Man Made Threats, S. Kasif, in review.
33. Analysis of brain region-specific co-expression networks reveals clustering of established and novel genes associated with Alzheimer disease. Lancour D, Dupuis J, Mayeux R, Haines JL, Pericak-Vance MA, Schellenberg GC, Crovella M, Farrer LA, Kasif S. *Alzheimers Res Ther.* 2020 Sep 2;12(1):103. doi: 10.1186/s13195-020-00674-7.
34. Membrane metallo-endopeptidase (Neprilysin) regulates inflammatory response and insulin signaling in white preadipocytes. Ramirez AK, Dankel S, Cai W, Sakaguchi M, Kasif S, Kahn CR. *Mol Metab.* 2019 Apr;22:21-36. doi: PMID: 30795914
35. Not All Experimental Questions Are Created Equal: Accelerating Biological Data to Knowledge Transformation (BD2K) via Artificial Intelligence and Active Learning, S Kasif\*, S Letovsky, RJ Roberts\*, M Steffen, revision of bioRxiv, 155150 2017 (\*co-corresponding and listed alphabetically), in revision *Nature Systems Biology and Applications*.

36. Integrating Extracellular Flux Measurements and Genome-Scale Modeling Reveals Differences between Brown and White Adipocytes, Alfred K. Ramirez, Matthew D. Lynes, Farnaz Shamsi, C. Ronald Kahn\*, Simon Kasif\*, Jonathan M. Dreyfuss\* (\* co-corresponding, Cell Reports 2017) – best 2017 PhD paper award for Alfred Ramirez).
37. One for All and All for One: Improving Genetic Studies through Network Diffusion Daniel Lancour<sup>1,4</sup>, Adam Naj<sup>5</sup>, Richard Mayeux<sup>6</sup>, Jonathan L. Haines<sup>7</sup>, Margaret A. Pericak-Vance<sup>8</sup>, Gerard C. Schellenberg<sup>5</sup>, Mark Crovella\*, Lindsay A. Farrer\* and Simon Kasif\* (\* co-corresponding and listed alphabetically, PLoS Genetics 2018),
38. Unexpected Properties of Short Genomic Tandem Repeats I Glotova, M Molla, AL Delcher, S Kasif bioRxiv, 165308 2017
39. Defects in muscle branched-chain amino acid oxidation contribute to impaired lipid metabolism. Lerin C, Goldfine AB, Boes T, Liu M, Kasif S, Dreyfuss JM, De Sousa-Coelho AL, Daher G, Manoli I, Sysol JR, Isganaitis E, Jessen N, Goodyear LJ, Beebe K, Gall W, Venditti CP, Patti ME. Mol Metab. 2016 Aug 6;5(10):926-36. doi: 10.1016/j.molmet.2016.08.001. eCollection 2016 Oct. PMID: 27689005
40. COMBREX-DB: an experiment centered database of protein function: knowledge, predictions and knowledge gaps. Chang YC, Hu Z, Rachlin J, Anton BP, Kasif S\*, Roberts RJ\*, Steffen M\*. Nucleic Acids Res. 2016 Jan 4;44(D1):D330-5. doi: 10.1093/nar/gkv1324. Epub 2015 Dec 3. PMID: 26635392 (\* co-corresponding, listed alphabetically)
41. COMBREX-DB: “An Experiment-Centered Database of Protein Function: Knowledge, Predictions and Knowledge Gaps” Chang, Yi-Chien; Hu, Zhenjun; Rachlin, John; Anton, Brian; \*Kasif, Simon; \*Roberts, Rich\*; \*Steffen, Martin (\*senior authors listed alphabetically, NAR 2016), PMID: 26635392
42. Metabolic modeling of muscle metabolism identifies key reactions linked to insulin resistance phenotypes. Nogiec C, Burkart A, Dreyfuss JM, Lerin C, Kasif S, Patti ME. Mol Metab. 2015 Jan 2;4(3):151-63. doi: 10.1016/j.molmet.2014.12.012. eCollection 2015 Mar. PMID: 25737951 (Journal Cover).
43. Identification of rare germline copy number variations over-represented in five human cancer types. Park RW, Kim TM, Kasif S, Park PJ Mol Cancer. 2015 Feb 3;14(1):25. PMID: 25644941
44. Objective: biochemical function. Anton BP, Kasif S, Roberts RJ, Steffen M. Front Genet. 2014 Jul 8;5:210. doi: 10.3389/fgene.2014.00210. PMID: 25071837
45. M. Liu, J. Dreyfuss, M. Mori, P. Park, IS. Kohane, ME Patti, A. Goldfine, CR Kahn, S. Kasif, Network Signature of Wellness, in preparation.
46. Hypermutable DNA chronicles the evolution of human colon cancer. Naxerova K, Brachtel E, Salk JJ, Seese AM, Power K, Abbasi B, Snuderl M, Chiang, S, Kasif S, Jain RK. Proc Natl Acad Sci U S A. 2014 PMID: 24753616
47. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. Suvà ML, Rheinbay E, Gillespie SM, Patel AP, Wakimoto H, Rabkin SD, Riggi N, Chi AS, Cahill DP, Nahed BV, Curry WT, Martuza RL, Rivera MN, Rossetti N, Kasif S, Beik S, Kadri S, Tirosh I, Wortman I, Shalek AK, Rozenblatt-Rosen O, Regev A, Louis DN, Bernstein BE. Cell. 2014 Apr 24;157(3):580-94. doi: 10.1016/j.cell.2014.02.030. Epub 2014 Apr 10. PMID: 24726434

48. The COMBREX project: design, methodology, and initial results. Anton BP, Chang YC, Brown P, Choi HP, Faller LL, Guleria J, Hu Z, Klitgord N, Levy-Moonshine A, Maksad A, Mazumdar V, McGettrick M, Osmani L, Pokrzywa R, Rachlin J, Swaminathan R, Allen B, Housman G, Monahan C, Rochussen K, Tao K, Bhagwat AS, Brenner SE, Columbus L, de Crécy-Lagard V, Ferguson D, Fomenkov A, Gadda G, Morgan RD, Osterman AL, Rodionov DA, Rodionova IA, Rudd KE, Söll D, Spain J, Xu SY, Bateman A, Blumenthal RM, Bollinger JM, Chang WS, Ferrer M, Friedberg I, Galperin MY, Gobeill J, Haft D, Hunt J, Karp P, Klimke W, Krebs C, Macelis D, Madupu R, Martin MJ, Miller JH, O'Donovan C, Palsson B, Ruch P, Setterdahl A, Sutton G, Tate J, Yakunin A, Tchigvintsev D, Plata G, Hu J, Greiner R, Horn D, Sjölander K, Salzberg, SL, Vitkup D, Letovsky S, Segrè D, DeLisi C, Roberts RJ, Steffen M, Kasif S. *PLoS Biology*. 2013 Aug;11(8):e1001638. doi: 10.1371/journal.pbio.1001638. Epub 2013 Aug 27. PMID: 24013487
49. An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. Rheinbay E, Suvà ML, Gillespie SM, Wakimoto H, Patel AP, Shahid M, Oksuz O, Rabkin SD, Martuza RL, Rivera MN, Louis DN, Kasif S, Chi AS, Bernstein BE. *Cell Reports*. 2013 May 30;3(5):1567-79. 2013.04.021. Epub 2013 May 23. PMID: 23707066
50. Systems biology characterization of engineered tissues. Rajagopalan P, Kasif S, Murali TM. *Annu Rev Biomed Eng*. 2013;15:55-70. doi: 10.1146/annurev-bioeng-071811-150120. PMID: 23862675
51. To supplement or not to supplement: a metabolic network framework for human nutritional supplements. Nogiec CD, Kasif S. *PLoS One*. 2013;8(8):e68751. doi: 10.1371/journal.pone.0068751. PMID: 23967053
52. Biochemical Characterization of Hypothetical Proteins from *Helicobacter pylori*. Choi HP, Juarez S, Ciordia S, Fernandez M, Bargiela R, Albar JP, Mazumdar V, Anton BP, Kasif S, Ferrer M, Steffen M. *PLoS One*. 2013 Jun 18;8(6):e66605. Print 2013. PMID: 23825549
53. Deep sequencing of the oral microbiome reveals signatures of periodontal disease, Bo Liu, Lina Faller, Niels Klitgord, Varun Mazumdar, Mohammad Ghodsi, Daniel D. Sommer, Theodore R. Gibbons, Todd J. Treangen, Yi-Chien Chang, Shan Li, O. Colin Stine, Hatice Hasturk, Simon Kasif, Daniel Segrè, Mihai Pop, Salomon Amar, *PLOS One*, 2012
54. Thousands of new genes with strong computational support found in bacterial genomes, Derrick E. Wood, Henry Lin, Ami Levy-Moonshine, Rajiswari Swaminathan, Yi-Chien Chang, Brian P. Anton, Lais Osmani<sup>4</sup>, Martin Steffen, Simon Kasif, and Steven L. Salzberg, *Biology Direct*, 2012.
55. Han-Pil Choi, Silvia Juarez, Sergio Ciordia, Marisol Fernandez, Rafael Bargiela, Juan P. Albar Varun Mazumdar, Brian P. Anton, Simon Kasif, Manuel Ferrer, Martin Steffen, *Biochemical Characterization Of Hypothetical Proteins From *Helicobacter pylori**, *PLoS One*, 2013.
56. A Critical Assessment of Genomic and Clinical Breast Cancer Biomarkers. Chang-Jiun Wu, Tianxi Cai, Zoltan Szallasi, Kamila Naxerova, Martin Steffen, L.J. Wei, Isaac S. Kohane\*, Simon Kasif\*, co-corresponding authors, in revision.
57. Integrative Pheno-genomic module discovery with application to breast cancer, Chang-Jiun Wu, Tianxi Cai, Isaac S. Kohane\*, Simon Kasif\*, in revision.

58. A Systems Biology Approach Links Inflammation to the Predisposition to Metabolic Diseases in Mouse Strains, Marcelo Mori\*, Manway Liu\*, Simon Kasif, Ron Kahn, Diabetes 2010.
59. Transcriptome Analysis Reveals Parallel Dysregulation of Oxidative Metabolism and Inflammation in Muscle and Adipose Tissue With Progression of Insulin Resistance in Humans, Mary-Elizabeth Patti, Manway Liu, Wanzhu Jin, Carles Lerin, Jonathan Dreyfuss, Martha Vokes, Joshua Schroeder, Elizabeth Tatro, Peter Park, Isaac Kohane, Igor Leykin, Simon Kasif, and Allison Goldfine, in preparation.
60. COMBREX: a project to accelerate the functional annotation of prokaryotic genomes. Roberts RJ, Chang YC, Hu Z, Rachlin JN, Anton BP, Pokrzywa RM, Choi HP, Faller LL, Guleria J, Housman G, Klitgord N, Mazumdar V, McGettrick MG, Osmani L, Swaminathan R, Tao KR, Letovsky S, Vitkup D, Segr D, Salzberg SL, Delisi C, Steffen M, Kasif S. Nucleic Acids Res. 2010 Nov 21. PMID: 21097892
61. Functional Characterization of the YmcB and YqeV tRNA Methyltransferases of *Bacillus subtilis* Brian P. Anton, Susan Russell, Patrick A. Limbach, Simon Kasif, Elisabeth A. Raleigh, and Richard J. Roberts, Nucleic Acid Research 2010.
62. Drug Response Phosphorylation Signatures of Lung Cancer, Chang-Jiun Wu, Martin Steffen, Simon Kasif (unpublished report).
63. The Evolution of Gene Annotation, Simon Kasif and Martin Steffen, Nature Chemical Biology, Jan 2010.
64. A Predictive phosphorylation signature of lung cancer. Chang-Jiun Wu, Tianxi Cai, Klarisa Rikova, David Merberg, Simon Kasif,\* Martin Steffen\*, co-corresponding authors, PLoS ONE, Nov 2009.
65. Triplet repeat length bias and variation in the human transcriptome, Molla M\*, Delcher A\*, Cantor CR, Kasif S, Proc. of National Academy of Sciences, Oct. 2009, PMID: 19805156
66. Integration of heterogeneous expression data sets extends the role of the retinol pathway in diabetes and insulin resistance. Park PJ, Kong SW, Tebaldi T, Lai WR, Kasif S, Kohane IS. Bioinformatics. 2009 Dec 1;25(23):3121-7. Epub 2009 Sep 28. PMID: 19786482
67. Seeing the Forest for the Trees: Using the Gene Ontology to Restructure Hierarchical clustering. Dotan-Cohen D, Kasif S, Melkman AA, Kasif S, Bioinformatics, 2009
68. Biological process linkage networks. Dotan-Cohen D, Letovsky S, Melkman AA, Kasif S. PLoS ONE. 2009;4(4):e5313. Epub 2009 Apr 23. PMID 19390589
69. Accelerated postnatal growth increases lipogenic gene expression and adipocyte size in low-birth weight mice. Isganaitis E, Jimenez-Chillaron J, Woo M, Chow A, DeCoste J, Vokes M, Liu M, Kasif S, Zavacki AM, Leshan RL, Myers MG, Patti ME. Diabetes. 2009 May;58(5):1192-200. Epub 2009 Feb 10. PMID: 19208909
70. Quantitative analysis of single nucleotide polymorphisms within copy number variation. Lee S, Kasif S, Weng Z, Cantor CR. PLoS ONE. 2008;3(12):e3906. Epub 2008 Dec 18. PMID: 19093001
71. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE. PLoS Genet. 2008 Oct;4(10):e1000242. Epub 2008 Oct 31. PMID: 18974828

72. Formula dissection: A parallel algorithm for constraint satisfaction. Reif, J. H., Kasif, S., and Sherlekar, D. *Comput. Math. Appl.* 55, 6 (Mar. 2008), 1170- 1177 (republished from a 1987 conference version).
73. Integration of relational and hierarchical network information for protein function prediction. Jiang X, Nariai N, Steffen M, Kasif S, Kolaczyk ED. *BMC Bioinformatics*. 2008 Aug 22; 9:350. PMID: 18721473
74. Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers. Naxerova K, Bult CJ, Peaston A, Fancher K, Knowles BB, Kasif S, Kohane IS. *Genome Biol.* 2008;9(7):R108. Epub 2008 Jul 8. PMID: 18611264
75. Anton BP, Saleh L, Benner JS, Raleigh EA, Kasif S, Roberts RJ. RimO, a MiaB-like enzyme, methylthiolates the universally conserved Asp88 residue of ribosomal protein S12 in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2008 Feb 12;105(6):1826-31. Epub 2008 Feb 5. PMID: 18252828
76. Context specific protein function prediction. Nariai N, Kasif S. *Genome Inform.* 2007; 18:173-82. PMID: 18546485
77. Tullai JW, Schaffer ME, Mullenbrock S, Sholder G, Kasif S, Cooper GM. Immediate-early and delayed primary response genes are distinct in function and genomic architecture, *J Biol Chem.* 2007 Jun 15;
78. Liu M, Liberzon A, Kong SW, Lai WR, Park PJ, Kohane IS, Kasif S. Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS Genet.* 2007 Jun 15;3(6):e96. PMID: 17571924
79. Nariai N, Kolaczyk ED, Kasif S. Probabilistic protein function prediction from heterogeneous genome-wide data. *PLoS ONE.* 2007 Mar 28;2:e337. PMID: 17396164
80. Zhang L, Kasif S, Cantor AC. Quantifying DNA-protein binding specificities by using oligonucleotide mass tags and mass spectroscopy. *Proc Natl Acad Sci U S A.* 2007 Feb 27;104(9):3061-6. Epub 2007 Feb 20. PMID: 17360609
81. Tullai JW, Chen J, Schaffer ME, Kamenetsky E, Kasif S, Cooper GM. Glycogen synthase kinase-3 represses cyclic AMP response element-binding protein (CREB)-targeted immediate early genes in quiescent cells. *J Biol Chem.* 2007 Mar 30;282(13):9482-91. Epub 2007 Feb 3. PMID: 17277356
82. Alon N, Asodi V, Cantor C, Kasif S, Rachlin J. Multi-node graphs: a framework for multiplexed biological assays. *J Comput Biol.* 2006 Dec;13(10):1659-72. PMID: 17238837
83. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-Scale Mapping and Validation of *Escherichia coli* Transcriptional Regulation from a Compendium of Expression Profiles., *PLoS Biol.* 2007 Jan 9;5(1):e8 PMID: 17214507
84. Murali TM, Wu CJ, Kasif S. The art of gene function prediction. *Nature Biotechnology.* 2006 Dec;24(12):1474-5. PMID: 17160037
85. Gustafson AM, Snitkin ES, Parker SC, DeLisi C, Kasif S. Towards the identification of essential genes using targeted genome sequencing and comparative analysis. *BMC Genomics.* 2006 Oct 19;7:265. PMID: 17052348
86. Lee S, Kasif S. The complete genome sequence of a dog: a perspective. *Bioessays.* 2006 Jun;28(6):569-73. PMID: 16700068

87. Rachlin J, Cohen DD, Cantor C, Kasif S. Biological context networks: a mosaic view of the interactome. *Nature/Embo. Mol Syst Biol.* 2006;2:66. Epub 2006 Nov 28. PMID: 17130868
88. Lee S, Kohane I, Kasif S. Genes involved in complex adaptive processes tend to have highly conserved upstream regions in mammalian genomes. *BMC Genomics.* 2005 Nov 27;6:168. PMID: 16309559
89. Zheng Y, Anton BP, Roberts RJ, Kasif S. Phylogenetic detection of conserved gene clusters in microbial genomes. *BMC Bioinformatics.* 2005 Oct 3;6:243. PMID: 16202130
90. Szustakowski JD, Kasif S, Weng Z. Less is more: towards an optimal universal description of protein folds. *Bioinformatics.* 2005 Sep 1;21 Suppl 2:ii66-ii71. PMID: 16204127
91. Rachlin J, Ding C, Cantor C, Kasif S. Computational tradeoffs in multiplex PCR assay design for SNP genotyping. *BMC Genomics.* 2005 Jul 25;6:102. PMID: 16042802
92. Wu CJ, Kasif S. GEMS: a web server for biclustering analysis of expression data. *Nucleic Acids Res.* 2005 Jul 1;33(Web Server issue):W596-9. PMID: 15980544
93. Rachlin J, Ding C, Cantor C, Kasif S. MuPlex: multi-objective multiplex PCR assay design. *Nucleic Acids Res.* 2005 Jul 1;33(Web Server issue):W544-7. PMID: 15980531
94. Zheng Y, Roberts RJ, Kasif S, Guan C. Characterization of two new aminopeptidases in *Escherichia coli*. *J Bacteriol.* 2005 Jun;187(11):3671-7. PMID: 15901689
95. Noga Alon, Richard Beigel and Simon Kasif and Steven Rudich and Benny Sudakov, "Learning a Hidden Matching", *SIAM Journal of Computing*, 2004.
96. Zheng Y, Roberts RJ, Kasif S. Identification of genes with fast-evolving regions in microbial genomes. *Nucleic Acids Res.* 2004 Dec 2;32(21):6347-57. Print 2004. PMID: 15576679
97. Zhang L, Kasif S, Cantor CR, Broude NE. GC/AT-content spikes as genomic punctuation marks. *Proc Natl Acad Sci U S A.* 2004 Nov 30;101(48):16855-60. Epub 2004 Nov 17. PMID: 15548610
98. Tullai JW, Schaffer ME, Mullenbrock S, Kasif S, Cooper GM. Identification of transcription factor binding sites upstream of human genes regulated by the phosphatidylinositol 3-kinase and MEK/ERK signaling pathways. *J Biol Chem.* 2004 May 7;279(19):20167-77. Epub 2004 Feb 9. PMID: 14769801
99. Zheng Y, Roberts RJ, Kasif S. Segmentally variable genes: a new perspective on adaptation. *PLoS Biol.* 2004 Apr;2(4):E81. Epub 2004 Apr 13. PMID: 15094797
100. Karaoz U, Murali TM, Letovsky S, Zheng Y, Ding C, Cantor CR, Kasif S. Whole-genome annotation by using evidence integration in functional-linkage networks. *Proc Natl Acad Sci U S A.* 2004 Mar 2;101(9):2888-93. Epub 2004 Feb 23. PMID: 4981259
101. Wu CJ, Fu Y, Murali TM, Kasif S. Gene expression module discovery using Gibbs sampling. *Genome Inform.* 2004;15(1):239-48. PMID: 15712126
102. Stitzel NO, Binkowski TA, Tseng YY, Kasif S, Liang J. topoSNP: a topographic database of non-synonymous single nucleotide polymorphisms with and without



- known disease association. *Nucleic Acids Res.* 2004 Jan 1;32(Database issue):D520-2. PMID: 14681472
103. Su Y, Murali TM, Pavlovic V, Schaffer M, Kasif S. RankGene: identification of diagnostic genes based on expression data. *Bioinformatics.* 2003 Aug 12;19(12):1578-9. PMID: 12912841
  104. Wu J, Kasif S, DeLisi C. Identification of functional links between genes using phylogenetic profiles. *Bioinformatics.* 2003 Aug 12;19(12):1524-30. PMID: 12912833
  105. Zhang L, Pavlovic V, Cantor CR, Kasif S. Human-mouse gene identification by comparative evidence integration and evolutionary analysis. *Genome Res.* 2003 Jun;13(6A):1190-202. Epub 2003 May 12. PMID: 12743024
  106. Pervouchine DD, Graber JH, Kasif S. On the normalization of RNA equilibrium free energy to the length of the sequence. *Nucleic Acids Res.* 2003 May 1;31(9):e49. PMID: 12711694
  107. Stitzel NO, Tseng YY, Pervouchine D, Goddeau D, Kasif S, Liang J. Structural location of disease-associated single-nucleotide polymorphisms. *J Mol Biol.* 2003 Apr 11;327(5):1021-30. PMID: 12662927
  108. Letovsky S, Kasif S. Predicting protein function from protein/protein interaction data: a probabilistic approach. *Bioinformatics.* 2003;19 Suppl 1:i197-204. PMID: 12855458
  109. Kasif S, Weng Z, Derti A, Beigel R, DeLisi C. A computational framework for optimal masking in the synthesis of oligonucleotide microarrays. *Nucleic Acids Res.* 2002 Oct 15;30(20):e106. PMID: 12384608
  110. Zheng Y, Roberts RJ, Kasif S. Genomic functional annotation using co-evolution profiles of gene clusters. *Genome Biol.* 2002 Oct 10;3(11):RESEARCH0060. Epub 2002 Oct 10. PMID: 12429059
  111. Zheng Y, Szustakowski JD, Fortnow L, Roberts RJ, Kasif S. Computational identification of operons in microbial genomes. *Genome Res.* 2002 Aug;12(8):1221-30. PMID: 12176930
  112. Walker M, Pavlovic V, Kasif S. A comparative genomic method for computational identification of prokaryotic translation initiation sites. *Nucleic Acids Res.* 2002 Jul 15;30(14):3181-91. PMID: 12136100
  113. Pavlovic V, Garg A, Kasif S. A Bayesian framework for combining gene predictions. *Bioinformatics.* 2002 Jan;18(1):19-27. PMID: 11836207
  114. Lander ES et al, Initial sequencing and analysis of the human genome. *Nature.* 2001 Feb 15;409(6822):860-921.
  115. Cai D, Delcher A, Kao B, Kasif S. Modeling splice sites with Bayes networks. *Bioinformatics.* 2000 Feb;16(2):152-8. PMID: 10842737
  116. Tettelin H, Radune D, Kasif S, Khouri H, Salzberg SL. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics.* 1999 Dec 15;62(3):500-7. PMID: 10644449
  117. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 1999 Dec 1;27(23):4636-41. PMID: 10556321
  118. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. Alignment of whole genomes. *Nucleic Acids Res.* 1999 Jun 1;27(11):2369-76. PMID: 10325427

119. Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 1998 Jan 15;26(2):544-8. PMID: 9421513
120. Bright, J., Kasif, L. Stiller, "Data Intensive Algorithms for Search Problems", accepted for publication in *IEEE Transactions on Computers*.
121. Kasif, S., "Datascope: Mining Biological Sequences", special issue on Data Mining, *IEEE Intelligent Systems*, pp. 38--45, 1999.
122. Kasif, S., Salzberg, S., Waltz, D., J. Rachlin, and Aha, D., "Towards of a Framework for Memory-Based Reasoning", *Artificial Intelligence*, pp. 287--311, 1998.
123. Kasif, S., "Towards a Constraint-Based Engineering Framework for Algorithm Design and Application", *Journal of Constraints*, 1997.
124. Delcher, A, A. Grove, S. Kasif and J. Pearl, "Logarithmic Time Queries and Updates in Probabilistic Networks", *Journal of Artificial Intelligence Research*, Vol. 4., pp. 37--59, 1996.
125. Waltz, D. and S. Kasif, "On Reasoning from Data", *Computing Surveys*, 1996.
126. Salzberg, S., D. Heath, A. Delcher and S. Kasif, "Best Case Analysis of Nearest Neighbours Algorithms", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17:6, pp.599--608, June 1995.
127. Delcher, A. and S. Kasif, "Term Matching on a Mesh-Connected Array of Processors", *Annals of Mathematics and Artificial Intelligence*, Volume 14, pp.177--186, 1995.
128. Murthy, S., S. Kasif and S. Salzberg, "System for Induction of Oblique Decision Trees", *Journal of Artificial Intelligence Research*, 2:1, pp.1--33, 1994.
129. Heath, D., S. Kasif, S. R. Kosaraju, S. Salzberg and G. Sullivan, "Learning Nested Concept Classes with Limited Memory", *Journal of Experimental and Theoretical AI*, 1996.
130. Kasif, S., "Optimal Parallel Algorithms for Quad-Tree Problems", *Journal of Computer Vision and Image Processing*, pp.281--285, May 1994.
131. Kasif, S. and A. Delcher, "Analysis of Local Consistency in Parallel Constraint Networks", *Artificial Intelligence*, 69, pp.307--327, 1994.
132. Heath, D. and S. Kasif, "On Voronoi Covers with Applications to Machine Learning", *Computational Geometry: Theory and Applications*, pp. 289-305, Nov. 1993.
133. Kasif, S., S. Banerjee, A. Delcher and G. Sullivan, "Some Results on the Complexity of Symmetric Connectionist Networks", *Annals of Mathematics and Artificial Intelligence*, pp.327-344, Nov. 1993.
134. Delcher, A. and S. Kasif, "Efficient Parallel Term Matching and Anti-Unification", *Journal of Automated Reasoning*, pp. 391--406, 1992.
135. Kasif, S., "On the Parallel Complexity of Discrete Relaxation in Constraint Networks", *Artificial Intelligence*, pp. 275-286, October 1990.
136. Delcher, A. and S. Kasif, "Some Results On the Complexity of Maintaining Data Flow Dependencies in Parallel Logic Programs", *Journal of Logic Programming*, pp.229--241, April 1989.
137. Kasif, S., "Control and Data Driven Execution of Logic Programs: A Comparison", *Journal of Parallel Programming*. Vol.15, No.1, pp.73--100, February 1987.

138. Kasif, S., L. Kitchen and A. Rosenfeld, "A Hough Transform Technique for Subgraph Isomorphism", Pattern Recognition Letters. Vol.2, pp.83--88, December 1983.
139. Kasif, S. and A. Rosenfeld, "Pyramid Linking as a Special Case of Isodata", IEEE Transactions on Systems, Man and Cybernetics, Vol.SMC-13, No.1, January 1983.

#### Unpublished Technical Reports

140. B. Logan, P. Moreno, B. Suzek, Z. Weng and S. Kasif, "Learning Remote-Homology Using Probabilistic Feature Vectors", Cambridge Research Lab TR 2000. (Patent).
141. Joseph D. Szustakowski, Ulas Karaoz, Serafim Batzoglou, James Galagan, Tarjei Mikkelsen, Zhiping Weng, Joel H. Graber, S. Kasif, On the Organization of Ancient and Modern Genes in the Human Genome.
142. Zhang L, Pavlovic V, E. Green, G. Cantor, C.R. and S. Kasif "Novel Human Gene Discovery using Cross-Species Analysis: Is the Mouse the Right Reference Point?".
143. Walker M, Kasif S., "Comparative Analysis of lipid biosynthesis and metabolism component of the DnaA regulon".
144. Walker M, Wu CJ, Bender A, Kasif S, Descriptor Biclustering Improves Binary Compound Classification with Partial Least Squares, 2006.
145. Walker M. Kasif S., Active Learning for Sequential Screening of Compounds, 2006.
146. Wu CJ, Cai T, Szallasi Z, Naxerova K, Steffen M, Wei LJ, Kohane IS, Kasif S, A Critical Assessment of Genomic and Clinical Breast Cancer Biomarkers, 2007.
147. Wu, CJ, Cai T, Kohane IS, Kasif S, Integrative Pheno-Genomic Module Discovery with Application to Breast Cancer Prognosis, 2007.

### 6.3 Papers in Peer-Reviewed Conferences

Publications in very selective conferences with acceptance rates of 25% or less are marked with asterisks. These papers are typically considered equal to a journal publication in Computer Science. Authors are typically listed **alphabetically** with the exception of PhD students.

148. \*Eisinger, N., S. Kasif and J. Minker, "Logic Programming: A Parallel Approach", First International Logic Programming Conf., Faculte des Sciences de Luminy Marseille, France, pp.71--77, September 1982.
149. Kasif, S., M. Kohli and J. Minker, "PRISM---A Parallel Inference System Based on Logic", Logic Programming Workshop, pp.123--152, Portugal, June 1983.
150. \*Kasif, S., M. Kohli and J. Minker, "Control Facilities of PRISM---A Parallel Inference System Based on Logic", International Joint Conf. on Artificial Intelligence, August 1983.
151. Chakravarthy, U. S., S. Kasif, M. Kohli, J. Minker and D. Cao, "Logic Programming on ZMOB: A Highly Parallel Machine", Proc. 1982 International Conf. on Parallel Processing, IEEE Press, pp.347—349 New York, 1982.

152. \*Kasif, S. and A. Rosenfeld, "The Fixpoints of Images and Scenes", Conf. on Computer Vision and Pattern Recognition, pp.454-456, June 1983.
153. \*Kasif, S. and J. Minker, "The Intelligent Channel: A Scheme for Result Sharing in Parallel Logic Programs", International Joint Conf. on Artificial Intelligence, pp.29-31, August 1985.
154. \*Kasif, S., "On the Parallel Complexity of Some Constraint Satisfaction Problems", National Conf. on Artificial Intelligence (AAAI-86), pp.349-353, August 1986.
155. Kasif, S., J. Reif and D. Sherlekar, "Formula Dissection: A Parallel Algorithm for Constraint Satisfaction", IEEE Workshop on Computer Architecture for Pattern Analysis and Machine Intelligence, pp.51-57, October 1987.
156. Delcher, A. and S. Kasif, "On the Complexity of Incremental Parallel Computations in Artificial Intelligence", IEEE Workshop on Computer Architecture for Pattern Analysis and Machine Intelligence, pp.59-64, October 1987.
157. Kasif, S., "Efficient Parallel Quad-Tree Algorithms", Proc. of 1988 ICAI, Tel Aviv, Israel, pp.353-363, December 1988.
158. Delcher, A. and S. Kasif, "Parallel Term Matching on Mesh Connected Computers", Proc. of 1988 ICAI, Tel Aviv, Israel, pp.~441-452, December 1988.
159. \*Kasif, S., "Parallel Solutions to Constraint Satisfaction Problems", IEEE Conf. on Principles of Knowledge Representation and Reasoning, pp. 180-187, May 1989.
160. \*Delcher, A. and S. Kasif, "Parallel Term Matching and Anti-Unification", International Conf. on Logic Programming, pp.355--369, June 1990.
161. \*Heath, D., S. Kasif, S. R. Kosaraju, S. Salzberg and G. Sullivan, "Learning Nested Concept Classes with Limited Storage", Proceedings of the International Joint Conference on Artificial Intelligence(IJCAI-91), pp. 777-782, 1991.
162. \*Salzberg, S., D. Heath, A. Delcher and S. Kasif, "Learning with a Helpful Teacher", Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI-91), pp. 705-711, 1991.
163. \*S. Kasif and A. Delcher, "Improved Decision Making in Game Trees: Recovering from Pathology", Proceedings of the National Conference on Artificial Intelligence (AAAI-92), pp. 513-518, July 1992.
164. D. Heath, S. Kasif and S. Salzberg, "Learning Oblique Decision Trees", Computational Learning Theory and Natural Learning Systems, 1992.
165. Kasif, S. and A. Delcher, "Analysis of Local Consistency in Parallel Constraint Networks", International Conference on Artificial Intelligence and Vision, pp. 217-231, 1992.
166. S. Kasif, "Iterative Focusing and Hashing: An Alternative to Alpha-Beta", International Conference on Artificial Intelligence and Vision, pp. 59-72, 1992.
167. \*D. Heath, S. Kasif and S. Salzberg, "Learning Oblique Decision Trees", Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI 93), pp. 1002--1007, August 1993.
168. \*Murthy, S., S. Kasif, S. Salzberg and R. Beigel, "OC1: A Randomized Algorithm for Building Oblique Decision Trees", Proceedings of the National Conference on Artificial Intelligence, (AAAI-93), pp. 322--327, July 93.
169. Heath, D., S. Kasif and S. Salzberg, "k-DT: A Multi-Tree Learning Method", Proceedings of the Second International Workshop on Multi-Strategy Learning (pp. 138--149), Harpers Ferry, West Virginia, 1993.

170. \*Delcher, A., S. Kasif, H. Goldberg and W. Xsu, "Protein Secondary-Structure Modeling with Probabilistic Networks", International Conference on Intelligent Systems and Molecular Biology, pp. 109--117, 1993.
171. \*Delcher, A., S. Kasif, H. Goldberg and W. Xsu, "Application of Probabilistic Causal Trees to Analysis of Protein Secondary Structure", Proceedings of the National Conference on Artificial Intelligence, pp. 316--321, July 1993.
172. \*Bright, J., Kasif, L. Stiller, "Exploiting Algebraic Structure in Parallel State-Space Search", Proc. of the 11-th National Conf. on Artificial Intelligence, July 1994, (AAAI-94), pp. 1341-1346, preliminary version presented in the AAAI Symposium on Massively Parallel AI", March 1993.
173. \*Rachlin, J., S. Kasif, S. Salzberg and D. Aha, "Toward of a better understanding of Memory-Based Classifiers", (plenary talk), Proceeding of the 11-th Intern. Conf. on Machine Learning, pp. 242--250, July 1994.
174. \*Fulton, T., S. Kasif and S. Salzberg, "Efficient Algorithms for Finding Multi-Way Splits for Decision Trees", JHU TR, December 1993, Proceeding of the 12-th Intern. Conf. on Machine Learning, July 1995.
175. \*Delcher, A, A. Grove, S. Kasif and J. Pearl, "Logarithmic Time Queries and Updates in Probabilistic Networks", Proceedings of the 1995 Conference on Uncertainty in AI, August 1995.
176. D. Dobkin, D. Gunopoulous, S. Kasif, "Induction of Low-Depth Decision Trees", International Conference on Mathematics and Artificial Intelligence, 1996,
177. S. Weiss, S. Kasif, and E. Brill, "Towards a Framework for Adaptive Information Retrieval", AAAI Spring Symposium on Information Retrieval (1996).
178. \*T. Fulton, S., Kasif, S. Salzberg, and D. Waltz, "Local Induction of Decision Trees", Proceedings of the 1996 Conference of Knowledge Discovery in Databases", August 1996.
179. R. Grossman, S. Bailey, S. Kasif, "Papyrus: A System for Distributed Data Mining", Workshop on Distributed Data Mining, NYC, 1998.
180. \*Beigel, R., N. Alon, S. Apaydin, L. Fortnow, and S. Kasif, "An Optimal Multiplex PCR Protocol for Closing Gaps in Whole Genomes", RECOMB, April 2001.
181. \* Noga Alon, Richard Beigel and Simon Kasif and Steven Rudich and Benny Sudakov, "Learning a Hidden Matching", Foundations of Computer Science, FOCS 2002.
182. T.M. Murali and S. Kasif, "Extracting Conserved Gene Expression Motifs from from Microarray Data", Pacific Symposium on Biocomputing, January 2003.
183. B. Logan, P. Moreno, B. Suzek, Z. Weng, and S. Kasif, "Remote Homology Detection Using Feature Vectors Formed Using Alignments of Small Motifs", RECOMB 2002 (poster and patent 2000).
184. D. Pervouchine, J. Graber, and S. Kasif, "Stable RNA Secondary Structure of Human Donor Splice Sites", RECOMB 2002 (poster)
185. M. Walker, V. Pavlovic, and S. Kasif, "A Comparative Genomic Method for Computational Identification of Prokaryotic Translation Initiation Sites", RECOMB 2002 (poster)
186. Y. Zheng, R. J. Roberts, and S. Kasif, "Computational Identification of Operons in Microbial Genomes", RECOMB 2002 (poster).

187. Murali TM, Kasif S. Extracting conserved gene expression motifs from gene expression data. *Pac Symp Biocomput.* 2003, 77-88. PMID: 12603019
188. \*S. Letovsky and S. Kasif, "A Probabilistic Approach to Gene Function Assignment and Propagation in Protein Interaction Networks", June 2003, ISMB 2003.
189. Rachlin J, Ding TM, Cantor CR, Kasif S., "The Limits of Multiplex PCR", RECOMB 2005.
190. \*Joseph D. Szustakowski, S. Kasif and Zhiping Weng, "A Dictionary of Protein Parts", European Conference on Computational Biology, 2005.

**8. Ph.D., Graduate Student, Interns, Research Associates, Research Scientists (over 60 mentored). Full list please see <http://sites.bu.edu/phenogeno/people/>**

1. Art Delcher, Ph.D. 1989, M.A. in Mathematics, Johns Hopkins University, Research Full Professor, Johns Hopkins Medical School. One of the computational architects of the Celera whole human shot-gun genome assembler.
2. David Heath, Ph.D. 1992, BS in Electrical Engineering, California Institute of Technology, JHU Wolman Fellow, first position faculty in Johns Hopkins Medical School, Visualization Research in Radiology, moved to a start-up in medical visualization.
3. Lewis Stiller, BS in Mathematics, Johns Hopkins University, Ph.D. May 1995, National Defense Fellow, Thesis area: Exploiting Symmetry in Parallel Computation, first position Postdoctoral Fellow, Computer Science Department Berkeley, CA.
4. Truxton Fulton, BS in Computer Science, Cal. Tech, JHU Wolman Fellow. Research area: Machine Learning, MA 1997 (MS)
5. Scott Weiss, BS in Computer Science, Carnegie-Mellon University, JHU Wolman Fellow, Machine Learning, first position Mount Saint Mary's University.
6. Tarjei Mikkelsen, intern, BS/MS MIT, PhD with Eric Lander (first position Broad Institute of MIT/Harvard, first author Chimp Genome paper.
7. S. Apaydin, intern PhD Stanford, with J.C. Latombe.
8. Baris Suzek, intern PhD Georgetown.
9. Ben Kao, BS in Biomedical Engineering, Duke University, Bioinformatics, PhD 2004 (jointly with Pete Nelson) (Machine Learning in Bioinformatics)
10. Deyou Cai, BS, MS, Taichin University and University of Chicago. (MS)
8. S. Murthy, (PhD Hopkins with S. Salzberg), senior scientist, IBM Research. PhD 1994. (Machine Learning)
9. D. Gunopoulous, PhD Princeton University, (PhD with David Dobkin) (Machine Learning)
10. E. Li, postdoctoral associate in Computational Biology (1998-1999).
11. Joseph D. Szustakowski (PhD Boston University), PhD with Zhiping Weng), currently Vice President Bristol Myers Squibb.
12. Yu Zheng, MS Yale, PhD 2004 (jointly with Richard J. Roberts), first position New England Biolabs (Computational Biology).
13. John Zhang PhD 2005 (jointly with Charles Cantor), first position Center for Advanced Biotechnology BU (Genomics).
14. Michael Schaffer, PhD 2006 (jointly with Geoff Cooper), 2006, (first position Pfizer Discovery) (Systems Biology).

15. Jason Laramie, PhD Boston University (with Rick Meyers) ,currently Vice President Novartis.
16. John Rachlin, BS Cornell, PhD 2006 (jointly with Charles Cantor), Machine Learning, Systems Biology, Genomic Technology) 2006, Assistant Professor, Northeastern University.
17. Megon Walker PhD 2006, Currently Patent Attorney (after Harvard Law School), (Active Learning for Drug Discovery).
18. J. Faith (PhD with Tim Gardner), PhD 2008, (first position Washington University, currently Assistant Professor Mount Sinai Genome Institute (Systems Biology).
19. Soo Lee (co-advisor, with Zhiping Weng and Charles Cantor), (first position U. Mass. Medical School, current Harvard Medical School). PhD 2009 (Genomics).
20. Terrence Wu, PhD 2008, (first position Dana Farber, Harvard Medical School), Bioinformatics and Machine Learning in Cancer.
21. Naoki Nariyai, BS University of Tokyo, PhD 2009, Bioinformatics and Machine Learning in Aging and Insulin Signaling).
22. Manway Liu, BS MIT, PhD 2010, (jointly with Zak Kohane, Children's), (first position Research Scientist, Novartis Research Institute), currently GoogleX (Systems Biology, Insulin Signaling and Inflammation).
23. Brian Anton, BS Georgia Tech, PhD 2010 (jointly with Rich Roberts), first position New England Biolabs) (Bioinformatics).
24. Esther Rheinbay, BS Germany, MS MIT (jointly with Brad Bernstein, Broad Institute), PhD May 2012, first position Scientist Broad Institute of MIT and Harvard (Epigenomics, Cancer, Bioinformatics), currently Assistant Professor, Harvard Medical School.
25. Kamila Naxerova, PhD with Rakej Jain, currently Assistant Professor at Harvard Medical School.
26. Chris Nogiec, BS Boston College, (jointly with ME Patti, Joslin, Harvard) (Metabolic Modeling, Systems Biology), currently Mount Sinai.
27. Richard Park, BS Brown University, (jointly with Peter Park, Harvard Medical School) (Bioinformatics, Genomics, Genomic Data Visualization). Second place in the Illumina Genome Visualization Competition.
28. Clark Freifeld, BS Yale, (jointly with John Brownstein, Harvard Medical School). Winner, Best Societal Engineering PhD thesis award, Currently Faculty Northeastern University, (Medical Informatics, Machine Learning).
29. Andy McMurry (jointly with Ben Reis and Zak Kohane, Harvard Medical School), (Distributed Data Mining of Medical Records, Medical Informatics).
30. Yi-Chien (jointly with Charles DeLisi), (Bioinformatics, Databases)
31. Alfred Ramirez BS MIT (jointly with Ron Kahn, Systems Biology)
32. Dan Lancour BS U. Wisconsin (jointly with Lindsay Farrer and Mark Crovella, Network Biology, Genetics).
33. Brian Haas (jointly with Aviv Regev, Genomics/Bioinformatics).

**Former postdoctoral associates and research faculty in the lab:**

34. S. Murali (PhD Brown), currently tenured full professor Virginia Tech (Bioinformatics, Data Mining).

35. V. Pavlovic (PhD Illinois), currently tenured full professor Rutgers Univ. (Graphics/Models, Data Integration).
36. J. Graber. (PhD Cornell), currently Jackson Labs (Bioinformatics, Genomics)
37. D. Pervouchine (PhD Moscow State), first position Center for Advanced Biotechnology (Bioinformatics, Genomics).
38. John Zhang (PhD Boston University), first position Center for Advanced Biotechnology (Genomics).
39. Arthur Liberzon (PhD Weizmann Institute of Science), currently scientist Broad Institute of Harvard and MIT supporting Gene Pattern and the GSEA Platform (Bioinformatics).
40. Michael Molla (PhD, U. Wisconsin), first position Director of Bioinformatics, Joslin Diabetes Center, Instructor Harvard Medical School (Genomics, Bioinformatics, Databases).
41. Jason Vertrees (PhD U. Texas), first position Schrodinger (Computational Biology, Data Visualization).
42. Revonda Pokrzywa (PhD Virginia Tech) (Bioinformatics).
43. Terrence Wu, MD, PhD (first position Dana Farber, Harvard Medical School) (Bioinformatics of Cancer) .
44. Ami Levy, PhD Israel, (first position the Broad Institute of Harvard and MIT), (Bioinformatics).
45. Jonathan Dreyfuss, PhD Boston University (currently Joslin Diabetes Center and Harvard Medical School), Bioinformatics, Systems Biology, Databases.
46. Hui Pan, PhD Genetics (first position, Joslin Diabetes Center) (Bioinformatics).
47. Many visiting scientists currently Vice Presidents in Biotech and Pharma.

### **Selected Research Assistants and Interns**

48. Phoebe Sengers, JHU, currently tenured professor Cornell University.
49. Sandeep Singhal, JHU currently senior scientist, Microsoft Research.
50. Kwaben Boahanen, JHU, currently tenured professor Stanford.
51. Willliam Hsu, JHU, currently tenured professor University of Kansas.
52. Tarjei Mikkelsen, MIT, first position, scientist Broad Institute (first author Chimp Genome paper).
53. Genevieve Housman, BU, currently University of Chicago
54. Catherine Dela Santana (BU, current in grad. program MIT)

Other co-mentored students, fellows and faculty can be found on <https://sites.bu.edu/phenogeno/people/>

### **9. Selected Professional Recognition and Professional Activities**

1. Presidential Advisory Committee on Trans-Disciplinary Research in Genetics, Epigenetics, Systems Biology and Public Health (BU), 2017-2018
2. Provost All University Advisory Data Science Committee (BU)



3. Senior Iacocca Fellowship, Harvard Medical School and Joslin Diabetes Center (2015-2016)
4. Advisor, Flybase, Harvard Medical School
5. Johns Hopkins University, Engineering School Advisory Committee.
6. Faculty of 1000 (F1000)
7. ISMB/ECCB 2015, Area Co-Chair for Disease Models
8. Joslin Diabetes National Research Center (Member Executive Committee)
9. ISMB 2014-2016, Area Co-Chair for Disease Models (with David Heckerman)
10. RECOMB 2014, Program committee.
11. AIMBE (Fellow Selection Committee for Bioinformatics 2015-2017)
12. Fellow of the American Institute for Medical and Biological Engineering, 2013.
13. ISMB/ECCB 2013, Area Chair Disease Models.
14. ISMB 2012, Area Co-Chair for Disease Models, (with Zak Kohane).
15. Computational Genomics Conference 2001, 2002, 2003, 2005, 2006, Program Committee.
16. Res. in Comp. Mol. Biology (RECOMB) 2004,2005,2006, 2008, 2009, 2010, 2011, 2012, 2013 Program Committee
17. Intern. Conf. on Data Mining, ICDM 2009 Program Committee
18. ISMB/ECCB 2007, 2008, 2009, 2010, 2011 Program Committee
19. ISMB 2012, ISMB/ECCB 2013 Area Chair: Disease Models.
20. AAI-2009, AAI-2002, UAI-2002, ICDM 2003, WABI 2003 Program Committee
21. International Conference on Machine Learning, ICML 2006. Program Committee
22. Yale University, Systems Biology Initiative, Advisory.
23. Advisory Board, Alberta ICORE
24. Advisory Board, Alberta Innovation Academy.
25. Advisory Board, NSF BLM Center, Carnegie Mellon University, PI Raj Reddy (Turing award winner in AI)
26. Advisory Board, Columbia University, NIH National Center for Multiscale Analysis of Genomic and Molecular Networks, PIs Andrea Califano and Barry Honig
27. Founding Member, Center Integrating Informatics and Bedside (I2B2), National Center for Biomedical Computing Center, Harvard Medical School/MIT/BU.
28. McArthur Foundation, invited nominator
29. Highly cited researcher (ISI)
30. General co-chair (with Jill Mesirov), RECOMB 2005.
31. Steering Committee: American Academy of Microbiology, Experimental Approach to Gene Function (with Rich Roberts, Peter Karp, and Stuart Kim).
32. NIH Centers for Excellence in Genomic Science (CEGS) panel
33. NIH Data Coordination Centers panel.
34. NIH Panels (BDMA, Genes and Genomes)
35. NIH Genome Section
36. NIH BISTI Section
37. 2003 International Conference on Data Mining, program committee.
38. International Public Human Genome Consortium, Hard Core Analysis Group (Lander et al).

39. Co-chair, DARPA, Focus 2000 Symposium, Mathematical/Language Modeling for Bio-Discovery (with Bob Berwick, MIT). The FOCUS 2000 symposium broadly initiated the synthetic biology initiative at DARPA.
40. Member of the Advisory Bioinformatics Task Force, University of Chicago.
41. NSF Panels
42. Co-Chair, NIPS Workshop on Integration (with Bill Noble and Tommi Jaakkola)
43. Chair, Health-care Track, NSF Symposium on Human Centered Systems, 1998 (started the human centered systems initiative at NSF).
44. Co-chair, 1997 NSF sponsored workshop, Mathematical Techniques in Data Mining (with Robert Grossman, Chicago).
45. Member, National NSF Steering Committee on Human Centered Systems (with T. Huang, P. Jones and J. Flannagan, P. Jones), NSF workshop on Human Centered Systems (1997) (attended by over 100 leading scientists in engineering, computer science, and social sciences), initiated the Human Centered Systems initiative at NSF.
46. International Workshop on High Performance Data Mining, Program Committee.
47. AAAI-1998, National Conference on AI, Program Committee.
48. Editorial Boards (past and present):
  49. Journal of Constraints (Kluwer Publ.)
  50. International Journal of Pattern Analysis and Applications (Elsevier).
  51. Machine Learning (Kluwer).
  52. ACM/IEEE Transactions on Computational Biology
  53. BMC BioData Mining.
  54. Open Access Bioinformatics
  55. Journal of Computational Biology
  56. Genes
  57. Biology Direct
  58. Journal of Personalized Medicine
59. Co-chair, AAAI 1996 Symposium on Learning Complex Behaviors (with Stuart Russell), Cambridge.
60. NSF Panel, 1996
61. AAAI-97, 1996 National Conference on AI, Program Committee.
62. AAAI-96, 1996 National Conference on AI Program Committee.
63. NSF Panel, 1995.
64. International Workshop on Constraints 1995, program committee.
65. Machine Learning Conference 1995, program committee.
66. International Workshop on Constraints 1995, program committee.
67. 1994 AAAI Program Committee Member.
68. 1993 International Conference on Pattern Recognition, Parallel Computing Track, Program Committee Member.
69. 1993 NSF Panel member, Knowledge Models and Cognitive Systems Program.
70. 1993 AFOSR initiative planning committee: AI and Operations Research.
71. Weizmann-Johns Hopkins Fellowship (Jun 1996)
72. High Performance Computing Terabyte Challenge Award (Supercomputing July 1998)
73. Knowledge and Distributed Intelligence NSF Award (Sep 1999)
74. ITR NSF Award, NSF (Sep 2004)

75. Research Initiation Award, NSF (1987)
76. Technical Innovation Award, GOI (1978)
77. Weizmann Institute of Science Youth Program.

## **10. Selected Funding History**

- Hariri Grant-
- Diabetes Research Center, 2012-2017, PI Systems Biology Core
- NIH BD2K Center grant, co-PI, (PI, Zak Kohane) 2015-
- NIH P50, 2010-2015, center grant, I2B2: Informatics for Integrating Biology and Bedside, National Center for Biomedical Computing (first four funded NIH Road Map Centers in Biomedical Computing), co-PI, (PI Zak Kohane)
- NIH Grand Opportunity, 2009-2012, Multi-PI (with Rich Roberts and M. Steffen).
- NIH R01, 2005-2009, PI
- NIH R33, 2004-2008, PI
- Neuroscience Seed Grant: (PI)
- Sequenom, 2005-2009, PI
- BU, Center Funding
- NSF IGERT Training Grant, “Biological Networks”, co-PI.
- NIH P50, 2005-2009, center grant, I2B2: Informatics for Integrating Biology and Bedside, National Center for Biomedical Computing (first four funded NIH Road Map Centers in Biomedical Computing), co-PI, (PI Zak Kohane)
- NSF, 09/01/07 – 8/31/09, Acquisition of a Linux Cluster for Bioinformatics Research and Education, Co-PI, (PI Zhiping Weng)
- NIH, 09/01/02 – 08/31/05, Computational Methods for Cell Systems Analysis, Co-PI, (PI Charles DeLisi)
- NSF, 09/01/01 – 08/31/05, MRI: acquisition of computing infrastructure for Bioinformatics and education
- NSF ITR 2004-2009, PI
- NSF KDI, 1999-2002, PI
- NSF, 2003-2006.
- Co-PI: NSF IGERT Grant to Boston University.
- Co-PI: Past and present training grants, NIH,
- Previous Numerous Grants as PI and co-PI from DARPA, AFOSR, Pfizer and others.
- NSF, Research Initiation Award 1987.

## **11. Consulting, Industrial and Advisory Board Activities**

Digital Corp., Compaq, IBM, HP, Pfizer, NEC, Harvard Partners, Carnegie-Mellon University, Children’s Hospital, University of Chicago, Columbia University National Center for Biomedical Computing, State wide Alberta Innovation Academy, Cell

Signaling Technology, PARADIGM4 Database Company, multiple biotechnology start-ups, multiple patents granted and pending.

## **12. Selected Invited Talks**

1. Stanford University, Computer Science, 1985.
2. University of Wisconsin, Computer Science, 1985.
3. University of Indiana, Computer Science, 1985.
4. Duke University, Computer Science, 1985.
5. University of Minnesota, Computer Science, 1985.
6. University of California, Santa Barbara, Computer Science 1985.
7. University of Virginia, Computer Science, 1985.
8. Johns Hopkins University, 1985
9. Brandeis University, 1985.
10. IBM Research, Yorktown, NY, May 1990.
11. Capital Area Theory Seminar, University of Maryland, College Park, April 1990.
12. Tel Aviv University, June 1990.
13. Department of Computer Science, Yale University, November 1990.
14. AAAI Symposium on Constraint Directed Reasoning, (invited plenary speaker), Stanford 1991.
15. Israeli Institute of Technology, (Technion), July 1991.
16. Department of Computer Science, UC Irvine, 1992.
17. AAAI Workshop on Constraint Directed Reasoning, Boston, August 1992.
18. Department of Computer Science, Hebrew University, Israel, 1994.
19. MIT, Spring 1994.
20. Bell Laboratories, Murray Hill, NJ., Spring 1994.
21. Navy Research Laboratories, Washington, DC, 1994.
22. National Institute of Health, 1994.
23. Boston, Invited Panel, AI Tools Conference, 1994.
24. Stanford University, Invited Panel, 1994.
25. NEC Research Laboratories, Princeton NJ, 1994.
26. Department of Computer Science, Princeton University, 1995.
27. University of Minnesota, Fall 1996, (invited Data Mining Workshop).
28. George Washington University, 1996.
29. Johns Hopkins University, Fall 1996.
30. Beckman Institute, University of Illinois, Winter 1997.
31. IBM Watson Research Labs, 1996.
32. Mathematical Techniques in Data Mining, DC (NSF sponsored workshop), 1997.
- 33.** Human Centered Systems, DC (NSF sponsored workshop), 1997.
- 34.** Bell Labs, Lucent, Murray-Hill (1998)
- 35.** Bell Labs, Lucent, Chicago (1998)
36. University of Illinois, Medical School (1998)
37. Weizmann Institute of Science, Genome Center (1998)
38. University of Pennsylvania, February 1999
39. University of Chicago, March 1999
40. Snowbird Learning Workshop, April 1999.

41. Celera Genomics, 1999.
42. University of New Mexico, Summer 1999.
43. Computational Genomic III, Nov. 1999.
44. Stanford University, Fall 1999.
45. MIT Genome Center, Fall 1999.
46. Princeton Univ. Spring 2000.
47. Berkeley, Spring 2000.
48. Princeton, Institute for Advanced Studies, Spring 2002.
49. Gordon Conference, July 2002.
50. Children Hospital, Harvard Medical School (CHIP Talk) 2003.
51. Tufts 2003.
52. MIT/HP Medical Informatics Seminar.
53. UCSD, Biomedical Engineering 2003.
54. MIT Bioinformatics Seminar 2004.
55. Massachusetts Biotechnology Council, Invited Panel 2006.
56. Tel Aviv University 2006.
57. Technion, Israel Institute of Technology 2006.
58. Boston University, Life Sciences Symposium, 2006.
59. Columbia Medical School, Center for Computational Biology and Bioinformatics 2007.
60. ICORE Summit Keynote, Banff 2007.
61. University of Alberta, 2007
62. Yale University, 2007.
63. University of Illinois, Chung Lecture 2008.
64. Rice University 2008.
65. Tufts University 2008 (Network Anomaly Based Disease Biomarkers)
66. Cell Signaling Technology, 2008
67. Technion, Israel Institute of Technology, Summer 2008
68. Tel Aviv University, Summer 2008
69. Dartmouth, Jan 2010
70. University of Massachusetts Medical School, March 2010
71. RECOMB Invited Panel on Genomics and Medicine, 2010
72. Yale Systems Biology Symposium, Fall 2010
73. Arrowhead Microbial Meeting, Fall 2010
74. NIDDK Systems Biology of Diabetes: Drug Response Symposium, 2011
75. ISMB summer 2011, Vienna, Keynote.
76. ISMB Invited Panel
77. University of Maryland, 2013
78. NIH Common Fund Meeting, May 2013 (Citizen's and Community Science)
79. Harvard Medical School, "Network Biology of Neuropsychiatric Diseases", November 2015
80. Tel Aviv University, "Network Biology Axis of Wellness and Disease Prevention", December 2017
81. Carnegie Mellon University, Spring 2018
82. University of Illinois, Spring 2018
83. University of Maryland 2018

84. Synthetic Biology Symposium, Boston 2018
85. Florida Atlantic University 2018
86. Santa Fe Institute 2019
87. Tel Aviv University December 2019
88. MIT, April 2020 (Does AI Need to Understand Nature to Pass the Turing Test)
89. Brown U. (invited)
90. CMU (invited)
91. NEB (invited)