

CURRICULUM VITAE

Simon Kasif, Ph.D.

Professor of Biomedical Engineering, Bioinformatics and Computer
Science,

Co-Director, (CAGT) Center for Advanced Genomic Technology
Boston University
44 Cummington St.
Boston, MA 02215

Phone: 617 358 1845 Email: kasif@bu.edu Fax: 617-353-6766
<http://cagt.bu.edu>

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

(advised by Jack Minker and Azriel Rosenfeld)

3. Research Interests

Computational Biology, Functional and Comparative Genomics, Biotechnology, Health-care, Gene Function Prediction, Algorithms, Data mining, Machine Learning.

Broad research interests: transformative or disruptive methodologies for biological and medical problems by combining computational, analytical and technological innovations. Our approach relies on integrative frameworks that combine multiple datasets of different types and scales such as protein-interaction networks, microarray measurements, genomic and epigenetic analysis, prior knowledge, clinical data and sophisticated graph theoretical and probabilistic algorithms aimed to produce meaningful and predictive hypotheses, biomarkers or discoveries that can be validated in the laboratory or clinic.

4. Professional Experience

June 2002-present: Founding Co-Director, Center for Advanced Genomic Research (CAGT), Boston University.

April 2005-present: Staff Scientist, Children's Hospital, Harvard-MIT Division of Health Sciences and Technology (part time).

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.

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

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 1998).

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 - Jun 95: 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).

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

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

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

Selected Journal Citations (Google Scholar)

Microbial Gene Finding (Glimmer): (680+, 490+)

Comparative Genomics (Mummer): 450+

Decision Tree Induction (OC1): 510+

Init. Sequence of the Human Genome (Public Consortium): 5900+

Numerous papers with roughly 60+ citations

One of the few researchers in the US with a consistent stream of publications in the most competitive venues in Computer Science, Biotechnology and Life Sciences (AAAI, ICML, AI Journal, STOC/FOCS, SIAM J. of Computing, PNAS, Nature Publications, PLoS).

Selected Research Systems Design and Development (more than twenty systems co-developed)

Bioinformatics System: GLIMMER

Glimmer is the most widely used gene discovery system in microbial DNA sequences. Joint work with S. Salzberg, O. White and A. Delcher. Achieves over 98% accuracy on most bacterial genomes. Glimmer is routinely used around the globe to annotate the DNA sequences of bacterial genomes that include pathogens that cause Lyme disease, Syphilis, Tuberculosis or bacteria used for Bio-energy applications. A eukaryotic variant has been used for Malaria. (1996-).

Distributed Data Mining:

One of the first implemented distributed data mining systems. Awarded High Performance Terabyte Challenge Award, Supercomputing 1998. (with R. Grossman and S. Bailey, et al).

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 and Richard Beigel).

Bioinformatics System: Mummer

Mummer is one of the first systems for high-resolution comparison of whole genomes. The system was implemented by Art Delcher in joint work with Steven Salzberg and has been used by the Institute for Genomic Research to perform whole-genome alignment of two different strains of microbial organisms (1998) searching for polymorphisms that contribute to virulence or pathogenicity. Mummer was used for the minimal organism project at TIGR. Other applications include SNP detection and comparison of mouse and human DNA, genomic re-sequencing and other.

Multiplex PCR: MUPLEX

MUPLEX is a first freeware for multi-plexing biological assays. Applications include genotyping, pathogen detection, forensic PCR, disease diagnosis. The system was implemented by a former PhD student John Rachlin. Muplex has hundreds of users.

High Performance Search and Planning

The parallel dynamic programming system finds chess end games that require more than **223 moves** to capture a piece and win. The system was implemented in the Connection machine 2 (roughly 64000 processors). This is considered a major extension of Ken Thompson's work at Bell Labs on 4 piece chess end-games. The system was designed and implemented by a former Ph.D. student Lewis Stiller. The results have been described in the popular press: New York Times, London Times, Washington Post, Scientific American (1992).

Data mining System: OC1

OC1 is a widely used 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).

Parallel deductive database system PRISM

PRISM is the first or one of the first deductive database systems that has been implemented and mapped to a multiprocessor, ZMOB. Joint work with Jack Minker and Madhur Kohli (1980-1983).

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, 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. (2nd Printing, February 1999).

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

1. A Critical Assessment of Genomic and Clinical Breast Cancer Biomarkers. Chang-Jiun Wu¹, Tianxi Cai, Zoltan Szallasi, Kamila Naxerova, Martin Steffen, L.J. Wei, Isaac S. Kohane*, Simon Kasif*, co-corresponding authors, in review 2009
2. A Predictive Phosphorylation signature of lung cancer. Chang-Jiun Wu, Tianxi Cai, Klarisa Rikova, David Merberg, Simon Kasif,* Martin Steffen*, co-corresponding authors, in review 2009.
3. Towards a Personal Repeatome: A Genome-Wide Survey of Length Variation in Short Tandem Repeats in Human Gene Transcripts. Molla M, Delcher A, Cantor CR, Kasif S, Proc. of National Academy of Sciences, to appear 2009.
4. Seeing the Forest for the Trees: Using the Gene Ontology to restructure Hierarchical clustering. Dotan-Cohen D, Kasif S, Melkman AA, Kasif S, Bioinformatics, to appear.
5. 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
6. 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
7. 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
8. 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
9. 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
10. 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
11. 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

12. 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
13. Context specific protein function prediction. Nariai N, Kasif S. *Genome Inform*. 2007; 18:173-82. PMID: 18546485
14. 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;
15. 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
16. Nariai N, Kolaczyk ED, Kasif S. Probabilistic protein function prediction from heterogeneous genome-wide data. *PLoS ONE*. 2007 Mar 28;2:e337. PMID: 17396164
17. 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
18. 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
19. 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
20. 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
21. Murali TM, Wu CJ, Kasif S. The art of gene function prediction. *Nat Biotechnol*. 2006 Dec;24(12):1474-5. PMID: 17160037
22. 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
23. Lee S, Kasif S. The complete genome sequence of a dog: a perspective. *Bioessays*. 2006 Jun;28(6):569-73. PMID: 16700068
24. 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
25. 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
26. 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

27. 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
28. 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
29. 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
30. 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
31. 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
32. Noga Alon, Richard Beigel and Simon Kasif and Steven Rudich and Benny Sudakov, "Learning a Hidden Matching", *SIAM Journal of Computing*, 2004.
33. 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
34. 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
35. 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
36. 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
37. 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
38. Wu CJ, Fu Y, Murali TM, Kasif S. Gene expression module discovery using Gibbs sampling. *Genome Inform*. 2004;15(1):239-48. PMID: 15712126
39. 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
40. 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
41. 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

42. 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
43. 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
44. 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
45. Letovsky S, Kasif S. Predicting protein function from protein/protein interaction data: a probabilistic approach. *Bioinformatics.* 2003;19 Suppl 1:i197-204. PMID: 12855458
46. 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
47. 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
48. 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
49. 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
50. Pavlovic V, Garg A, Kasif S. A Bayesian framework for combining gene predictions. *Bioinformatics.* 2002 Jan;18(1):19-27. PMID: 11836207
51. Lander ES et al. Initial sequencing and analysis of the human genome. *Nature.* 2001 Feb 15;409(6822):860-921.
52. Cai D, Delcher A, Kao B, Kasif S. Modeling splice sites with Bayes networks. *Bioinformatics.* 2000 Feb;16(2):152-8. PMID: 10842737
53. 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
54. 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
55. 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
56. 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
57. Bright, J., Kasif, L. Stiller, "Data Intensive Algorithms for Search Problems", accepted for publication in *IEEE Transactions on Computers*.
58. Kasif, S., "Datascope: Mining Biological Sequences", special issue on Data Mining, *IEEE Intelligent Systems*, pp. 38--45, 1999.

59. 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.
60. Kasif, S., "Towards a Constraint-Based Engineering Framework for Algorithm Design and Application", *Journal of Constraints*, 1997.
61. 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.
62. Waltz, D. and S. Kasif, "On Reasoning from Data", *Computing Surveys*, 1996.
63. 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.
64. 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.
65. 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.
66. 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.
67. Kasif, S., "Optimal Parallel Algorithms for Quad-Tree Problems", *Journal of Computer Vision and Image Processing*, pp.281--285, May 1994.
68. Kasif, S. and A. Delcher, "Analysis of Local Consistency in Parallel Constraint Networks", *Artificial Intelligence*, 69, pp.307--327, 1994.
69. Heath, D. and S. Kasif, "On Voronoi Covers with Applications to Machine Learning", *Computational Geometry: Theory and Applications*, pp. 289-305, Nov. 1993.
70. 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.
71. Delcher, A. and S. Kasif, "Efficient Parallel Term Matching and Anti-Unification", *Journal of Automated Reasoning*, pp. 391--406, 1992.
72. Kasif, S., "On the Parallel Complexity of Discrete Relaxation in Constraint Networks", *Artificial Intelligence*, pp. 275-286, October 1990.
73. 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.
74. 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.
75. Kasif, S., L. Kitchen and A. Rosenfeld, "A Hough Transform Technique for Subgraph Isomorphism", *Pattern Recognition Letters*. Vol.2, pp.83--88, December 1983.
76. 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

77. Zhang, L., Pavlovic, E. Green, G. Cantor, C.R. and S. Kasif "Novel Human Gene Discovery using Cross-Species Analysis: Is the Mouse the Right Reference Point?".
78. B. Logan, P. Moreno, B. Suzek, Z. Weng and S. Kasif, "Learning Remote-Homology Using Probabilistic Feature Vectors", Cambridge Research Lab TR 2000.
79. Walker, M. and Kasif, S., "Comparative Analysis of lipid biosynthesis and metabolism component of the DnaA regulon".
80. 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.

6.3 Papers in Refereed Conferences and Invited Conference Presentations

Publications in very selective conferences with acceptance rates of 25% or less are marked with asterisks. These papers are typically considered as equal to a journal publication in Computer Science.

81. *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.
82. Kasif, S., M. Kohli and J. Minker, "PRISM---A Parallel Inference System Based on Logic", Logic Programming Workshop, pp.123--152, Portugal, June 1983.
83. *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.
84. 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.
85. *Kasif, S. and A. Rosenfeld, "The Fixpoints of Images and Scenes", Conf. on Computer Vision and Pattern Recognition, pp.454-456, June 1983.
86. *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.
87. *Kasif, S., "On the Parallel Complexity of Some Constraint Satisfaction Problems", National Conf. on Artificial Intelligence (AAAI-86), pp.349-353, August 1986.
88. 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.
89. 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.
90. Kasif, S., "Efficient Parallel Quad-Tree Algorithms", Proc. of 1988 ICAI, Tel Aviv, Israel, pp.353-363, December 1988.
91. Delcher, A. and S. Kasif, "Parallel Term Matching on Mesh Connected Computers", Proc. of 1988 ICAI, Tel Aviv, Israel, pp.~441-452, December 1988.

92. *Kasif, S., "Parallel Solutions to Constraint Satisfaction Problems", IEEE Conf. on Principles of Knowledge Representation and Reasoning, pp. 180-187, May 1989.
93. *Delcher, A. and S. Kasif, "Parallel Term Matching and Anti-Unification", International Conf. on Logic Programming, pp.355--369, June 1990.
94. *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.
95. *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.
96. *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.
97. D. Heath, S. Kasif and S. Salzberg, "Learning Oblique Decision Trees", Computational Learning Theory and Natural Learning Systems, 1992.
98. Kasif, S. and A. Delcher, "Analysis of Local Consistency in Parallel Constraint Networks", International Conference on Artificial Intelligence and Vision, pp. 217-231, 1992.
99. S. Kasif, "Iterative Focusing and Hashing: An Alternative to Alpha-Beta", International Conference on Artificial Intelligence and Vision, pp. 59-72, 1992.
100. *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.
101. *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.
102. 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.
103. *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.
104. *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.
105. *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.
106. *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.
107. *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.

108. *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.
109. D. Dobkin, D. Gunopoulous, S. Kasif, “Induction of Low-Depth Decision Trees”, International Conference on Mathematics and Artificial Intelligence, 1996,
110. S. Weiss, S. Kasif, and E. Brill, “Towards a Framework for Adaptive Information Retrieval”, AAAI Spring Symposium on Information Retrieval (1996).
111. *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.
112. R. Grossman, S. Bailey, S. Kasif, “Papyrus: A System for Distributed Data Mining”, Workshop on Distributed Data Mining, NYC, 1998.
113. *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.
114. *Noga Alon, Richard Beigel and Simon Kasif and Steven Rudich and Benny Sudakov, “Learning a Hidden Matching”, Foundations of Computer Science, FOCS 2002.
115. T.M. Murali and S. Kasif, “Extracting Conserved Gene Expression Motifs from from Microarray Data”, Pacific Symposium on Biocomputing, January 2003.
116. 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).
117. D. Pervouchine, J. Graber, and S. Kasif, “Stable RNA Secondary Structure of Human Donor Splice Sites”, RECOMB 2002 (poster)
118. M. Walker, V. Pavlovic, and S. Kasif, “A Comparative Genomic Method for Computational Identification of Prokaryotic Translation Initiation Sites”, RECOMB 2002 (poster)
119. Y. Zheng, R. J. Roberts, and S. Kasif, “Computational Identification of Operons in Microbial Genomes”, RECOMB 2002 (poster).
120. Murali TM, Kasif S. Extracting conserved gene expression motifs from gene expression data. Pac Symp Biocomput. 2003;:77-88. PMID: 12603019
121. *S. Letovsky and S. Kasif, "A Probabilistic Approach to Gene Function Assignment and Propagation in Protein Interaction Networks", June 2003, ISMB 2003.
122. *Murali TM, Kasif S. Extracting conserved gene expression motifs from gene expression data. Pac Symp Biocomput. 2003;:77-88.
123. Rachlin J, Ding TM, Cantor CR, Kasif S., “The Limits of Multiplex PCR”, RECOMB 2005 (poster).
124. *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 and Research Associates

1. Art Delcher, Ph.D. 1989, M.A. in Mathematics, Johns Hopkins University, Research Full Professor, University of Maryland College Park. One of the architects of the Celera 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: Adaptive Modeling of Complex Systems, MA 1997 (MS)
5. Scott Weiss, BS in Computer Science, Carnegie-Mellon University, JHU Wolman Fellow, first position Mount Saint Mary's University.
6. Tarjei Mikkelsen, intern, BS/MS MIT, PhD with Eric Lander, current 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)
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.
9. D. Gunopulous, PhD Princeton University, (PhD with David Dobkin).
10. E. Li, postdoctoral associate in Computational Biology (1998-1999).
11. Yu Zheng, MS Yale, PhD 2004 (jointly with Richard J. Roberts), first position New England Biolabs.
12. John Zhang PhD 2005 (jointly with Charles Cantor), first position Center for Advanced Biotechnology.
13. Michael Schaffer, PhD 2006 (jointly with Geoff Cooper), 2006, Pfizer Discovery
14. John Rachlin, BS Cornell, PhD 2006 (jointly with Charles Cantor), 2006 Aventis
15. Megan Walker PhD 2006.
16. J. Faith (co-advisor, primary Tim Gardner) 2008, currently at Washington University.
17. Soo Lee (co-advisor, with Zhiping Weng and Charles Cantor), currently U. Mass. Medical School.
18. Terrence Wu 2008
19. Naoki Nariyai, 2009
20. Current Students: Manway Liu, BS MIT (jointly with Zak Kohane), Brian Anton, BS Georgia Tech, (jointly with Rich Roberts), Esther Rheinbau, BS Germany, MS MIT (jointly with Brad Bernstein).
21. Former and current postdoctoral associates and research faculty in the lab:
22. S. Murali (PhD Brown), currently tenured Virginia Tech.
23. V. Pavlovic (PhD Illinois), currently tenured Rutgers Univ.
24. J. Graber. (PhD Cornell), currently Jackson Labs
25. D. Pervouchine (PhD Moscow State), first position Center for Advanced Biotechnology.
26. John Zhang (PhD Boston University), first position Center for Advanced Biotechnology.

27. Arthur Liberzon (Weizmann Institute of Science), currently scientist Broad Institute of Harvard and MIT.
28. Current postdoctoral associates Jason Vertrese (PhD U. Texas), Terrence Wu, Michael Molla (PhD, U. Wisconsin, jointly with Jim Collins).

9. Selected Recent Professional Recognition and Professional Activities

1. Computational Genomics Conference 2001, 2002, 2003, 2005, 2006, Program Committee.
2. RECOMB 2004,2005,2006, 2008, 2009 Program Committee
3. ISMB/ECCB 2007, 2008, 2009 Program Committee
4. AAAI-2009, AAAI-2002, UAI-2002, ICDM 2003, WABI 2003 Program Committee
5. International Conference on Machine Learning, ICML 2006. Program Committee
6. Advisory Board, Alberta ICORE.
7. Advisory Board, BLM Center, Language and Computational Biology, Carnegie Mellon University
8. Advisory Board, Columbia University, NIH National Center for Bioinformatics and Computational Biology.
9. Founding Member, Center Informatics to Bedside (I2B2), National Center for Biomedical Computing Center, Harvard Medical School/MIT/BU.
10. McCarthur Foundation, invited nominator and referee
11. Highly cited researcher (ISI)
12. General co-chair (with Jill Mesirov), RECOMB 2005.
13. Steering Committee: Experimental Approach to Gene Function (with Rich Roberts, Peter Karp, and Stuart Kim).
14. NIH Genome Section
15. NIH BISTI Section
16. International Public Human Genome Consortium, Hard Core Analysis Group
17. Co-chair, DARPA, Focus 2000, Mathematical/Language Modeling for Bio-Discovery (with Bob Berwick, MIT).
18. Member of Bioinformatics Task Force, University of Chicago.
19. NSF Panels
20. Chair, Health-care Track, NSF workshop on Human Centered Systems, 1998.
21. Co-chair, 1997 NSF sponsored workshop, Mathematical Techniques in Data Mining (with Robert Grossman, Chicago).
22. Member, National NSF Steering Committee on Human Centered Systems (with T. Huang, P. Jones and J. Flannagan), 1997.
23. Co-chair (with P. Jones), Communication and Collaboration, NSF workshop on Human Centered Systems (1997) (attended by over 100 leading scientists in engineering, computer science, and social sciences).
24. International Workshop on High Performance Data Mining, Program Committee.
25. AAAI-1998, National Conference on AI, Program Committee.
26. Editorial Board: Journal of Constraints (Kluwer Publ.).
27. Editorial Board: International Journal of Pattern Analysis and Applications (Elsevier).

28. Editorial Board, Machine Learning (Kluwer).
29. Editorial Board: ACM/IEEE Transactions on Computational Biology
30. Editorial Board: BMC BioData Mining.
31. Editorial Board: Open Access Bioinformatics
32. Editorial Board: Journal of Computational Biology (since July 2009)
33. Co-chair, AAAI 1996 Symposium on Learning Complex Behaviors (with Stuart Russell), Cambridge.
34. NSF Panel, 1996
35. AAAI-96, 1996 National Conference on AI, Program Committee.
36. AAAI-96, 1996 National Conference on AI Program Committee.
37. NSF Panel, 1995.
38. International Workshop on Constraints 1995, program committee.
39. Machine Learning Conference 1995, program committee.
40. International Workshop on Constraints 1995, program committee.
41. 1994 AAAI Program Committee Member.
42. 1993 International Conference on Pattern Recognition, Parallel Computing Track, Program Committee Member.
43. 1993 NSF Panel member, Knowledge Models and Cognitive Systems Program.
44. 1993 AFOSR initiative planning committee: AI and Operations Research.
45. Weizmann Institute of Science Youth Program.
46. NSF Initiation Award 1997.
47. Technical Innovation Award, GOI (1978)
48. Weizmann-Johns Hopkins Fellowship (Jun 1996)
49. High Performance Computing Terabyte Challenge Award (Supercomputing July 1998)
50. Knowledge and Distributed Intelligence NSF Award (Sep 1999)
51. ITR NSF Award, NSF (Sep 2004)
52. Research Initiation Award, NSF (1987)

10. Recent Grant History

Last five years roughly a \$1 million external funding per year (as a PI).

Multiple NIH grant applications pending.

NIH R01, 2005-2009, \$1,050,000, PI

NIH R33, 2004-2008, \$1,100,000, PI

Sequenom, 2007-2008, \$254,000, PI

Sequenom, 2006-2007, \$250,000, PI

Sequenom, 2006-2007, \$250,000, PI

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)

NSF ITR 2004-2009, \$1,900,000, PI

NSF KDI, 1999-2002, \$1,950,000, PI

NSF, 2003-2006, \$580,000.

Uninterrupted support from NSF from 1987, NIH from 2004.

Co-PI; Multiple Instrumentation grants, NSF.

Co-PI: Past and present training grants, NIH, and NSF IGERT.

Previous 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, CMU, Children's Hospital, University of Chicago, Columbia, start-ups, 3 patents granted and four 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 Santa Barbara, Computer Science 1985.
7. University of Virginia, Computer Science, 1985.
8. Johns Hopkins University
9. IBM Research, Yorktown, NY, May 1990.
10. Capital Area Theory Seminar, University of Maryland, College Park, April 1990.
11. Tel Aviv University, June 1990.
12. Department of Computer Science, Yale University, November 1990.
13. AAAI Symposium on Constraint Directed Reasoning, (invited plenary speaker), Stanford 1991.
14. Israeli Institute of Technology, (Technion), July 1991.
15. Department of Computer Science, UC Irvine, 1992.
16. AAAI Workshop on Constraint Directed Reasoning, Boston, August 1992.
17. Department of Computer Science, Hebrew University, Israel, 1994.
18. MIT, Spring 1994.
19. Bell Laboratories, Murray Hill, NJ., Spring 1994.
20. Navy Research Laboratories, Washington, DC, 1994.
21. National Institute of Health, 1994.
22. Boston, Invited Panel, AI Tools Conference, 1994.
23. Stanford University, Invited Panel, 1994.
24. NEC Research Laboratories, Princeton NJ, 1994.
25. Department of Computer Science, Princeton University, 1995.
26. University of Minnesota, Fall 1996, (invited Data Mining Workshop).
27. George Washington University, 1996.
28. Johns Hopkins University, Fall 1996.
29. Beckman Institute, University of Illinois, Winter 1997.
30. IBM Watson Research Labs, 1996.
31. Mathematical Techniques in Data Mining, DC (NSF sponsored workshop), 1997.

32. Human Centered Systems, DC (NSF sponsored workshop), 1997.
33. Bell Labs, Lucent, Murray-Hill (1998)
34. Bell Labs, Lucent, Chicago (1998)
35. University of Illinois, Medical School (1998)
36. Weizmann Institute of Science, Genome Center (1998)
37. University of Pennsylvania, February 1999
38. University of Chicago, March 1999
39. Snowbird Learning Workshop, April 1999.
40. Celera Genomics, 1999.
41. University of New Mexico, Summer 1999.
42. Computational Genomic III, Nov. 1999.
43. Stanford University, Fall 1999.
44. MIT Genome Center, Fall 1999.
45. Princeton Univ. Spring 2000.
46. Berkeley, Spring 2000.
47. Princeton, Institute for Advanced Studies, Spring 2002.
48. Gordon Conference, July 2002.
49. Children/CHIP 2003.
50. Tufts 2003.
51. MIT/HP Medical Informatics Seminar.
52. UCSD, Biomedical Engineering 2003.
53. MIT Bioinformatics Seminar 2004.
54. Massachusetts Biotechnology Council, Invited Panel 2006.
55. Tel Aviv University 2006.
56. Technion, Israel Institute of Technology 2006.
57. Boston University, Life Sciences Symposium, Keynote 2006.
58. Columbia Medical School, Center for Computational Biology and Bioinformatics 2007.
59. ICORE Summit Keynote, Branff 2007.
60. University of Alberta, 2007
61. Yale University, 2007.
62. University of Illinois, Chung Lecture 2008.
63. Rice University 2008.
64. Tufts University 2008.
65. Cell Signaling Technology, 2008
66. Technion, Israel Institute of Technology, Summer 2008
67. Tel Aviv University, Summer 2008
68. Accepted invited talks to be scheduled: Minnesota, MIT, Dartmouth, Princeton.