

## Supplement 1: Simon Kasif Key Work/Contributions Timeline

Year	Work	General AI/Biology/Medicine Topic	With	AI CS Innovation Biology Innovation
	<p>*Green: An early or pioneering contribution</p> <p>*Considered Textbook</p> <p>*Transformative Community Work or Widely used Systems</p> <p><b>Bold – highly cited or visible</b></p>		<p><b>Red : Lab Member</b></p> <p>*Blue: Major Researcher</p>	
1981	Pyramid Linking Converges	Early Unsupervised Deep Learning (AI theory)	A. Rosenfeld	XX
1983	First implemented Parallel Datalog System on a distributed memory machine	Parallel Joins Coupled with Datalog (AI system)	J. Minker et al	XX
1983	First Parallel Logic Programming AI System to be mapped and tested on a multiprocessor (ZMOB)	Parallel AI (system)	J. Minker et al	XX
1985	Arc Consistency / Discrete Relaxation is Inherently Sequential <b>HIGHLY CITED</b>	Parallel AI (AI theory)	Sole author	XX
1985-1987	Graph Separator based Parallel SAT Solving	Parallel AI (AI theory)	J. Reif D. Sherlekar	XX
1987	Provably Optimal Parallel Term Matching and Anti-Unification	Parallel AI (AI theory)	Art Delcher	XX
1991	An early introduction of Data Streaming into ML and Data Mining Mostly theory. Upper, Lower Bounds for Data Streaming Application in Cognitive Science	Preceded by several years the early Data Streaming Results that popularized the area (AI Theory)	D. Heath, R. Kosaraju, S. Salzberg	XX
1990	<b>223 moves to capture a piece and win</b> *Considered Textbook and seminal work in computer chess Follow up on Ken Thompson's 51 moves to win	Parallel AI System on the Connection Machine Revolutionized chess (AI system)	L. Stiller PhD Thesis	XX

1993	Introducing Randomization in Decision Tree (DT) Induction Prior to Random Forests <b>HIGHLY CITED</b>	Randomization in DTs, Random Projections (system)	D. Heath, S. Murthy, R. Beigel, S. Salzberg	XX
1993-	<b>Bayes Networks and Biology</b> <b>HIGHLY CITED</b>	Early proposal for causal BNs in Biology	A. Delcher	X X X
1994	<b>OC1 – Widely Used Open Access Decision Tree System</b> <b>HIGHLY CITED</b>	(Randomization, Scalability by Sorting on Attributes, Ensemble of DTs, AI system)	S. Murthy, S. Salzberg	XX
1994	Logarithmic Time Queries and Updates in Probabilistic Networks Factorization Extending on Miller-Reif in Bayes Nets	Novel factorization and compilation of BNs enabling parallelism and dynamic updates Application to In-silico mutagenesis First application of graphical models to synthetic biology	J. Pearl A. Grove, A. A. Delcher	X X X
1994-1996	First of a kind Bayes Nets Kernels (very early and perhaps the first of a kind EMPIRICAL work but lacks mathematical rigor found in the follow-up papers)	Producing a BN Kernel, integrating BN-s with k-NN, Testing on ML Benchmarks (system and theory)	J. Rachlin, D. Waltz S. Salzberg	X X X
1996	<b>AAAI Symposium Learning Complex Behaviors</b>	Moving ML from Toy Problems to Learning Complex Behaviors and Systems	S. Russell M. Jordan D. Koller Over 100 scientists	XX
1997	<b>Human Centered Systems: Information, Interactivity and Intelligence</b>	Launched the Human Centered Intelligence Initiative at NSF Early Warning about AI Safety and the problem of Control	J. Flannagan T. Huang P. Jones over 100 scientists	XX
1998	Computational Methods in Molecular Biology”, Elsevier Publ. Focus on AI and Biology including Bayes Nets Some textbook material – not all.	One of the first books in Computational Molecular Biology	S. Salzberg D. Searls	X X X

1998	<b>GLIMMER: widely Used Open Access AI System for Microbial Annotation</b> First variable length models in Bio <b>HIGHLY CITED</b>	Millions of new enzymes identified with massive impact on science and biotechnology (open access system)	A. Delcher S. Salzberg O. White	X X X
1999	<b>MUMMER: widely used and early system for whole genome comparison (bacteria)</b> First open access whole genome bacterial analysis <b>HIGHLY CITED</b>	Wide use in clinical and biological research with exceptional impact on biology and disease (open access system)	A. Delcher S. Salzberg	X X
1999	Multiplex PCR for Gap Closing in Genomes	Bridge between Complex Combinatorics and Genomics (theory and experimental implementation)	S. Salzberg H. Tettelin	
1999-2001	<b>Human Genome Project</b> <b>HIGHLY CITED</b>	<b>COMPUTATIONAL ANALYSIS TEAM</b>	Lander et al	XX X
2000	<b>DARPA SYSTEMS BIO AND SYNT. BIO SYMPOSIUM</b>	LANGUAGE MODELING FOR BIOLOGY	Co-organizer Over 100 scientists attended	X X X
2000	Bayes Nets and Genomic Integration		V. Pavlovic	X X X
2002	Center for Advanced Genomic Technology (at BU)		Charles Delisi	
2003-2005	<b>Network based Gene Function Prediction</b> Wide impact on most popular systems today <b>HIGHLY CITED</b>	Introducing Network Propagation into Function Prediction Established DISCIPLINE	S. Letovsky	X X X
2004	<b>Hopfield Networks and Gene Function Prediction</b>	Introducing Network propagation into Function Prediction	CM. Ding T.M. Murali C. Cantor	XXX
2004-2020	<b>Direct Involvement in Experimental Validation of Computational Predictions</b> <b>HIGHLY CITED</b>		R. Roberts C. Cantor G. Cooper M. Steffen R. Jain B. Bernstein	X

			A. Regev R. Kahn S. Tseng ME. Patti Combrex T. Gardner J. Collins	
2004	Learning Hidden Matchings in Graphs from queries	New mini-area in graph theory	N. Alon R. Beigel	XXX
2004	<b>Starting COMBREX CONSORTIUM Academy Meeting</b>	A Community Project using Active Learning for Recommending Experiments	Rich Roberts & > 50 scientists	XXX
2005	Multi-nodes graphs and Multiplex PCR <b>MUPLEX System</b>	Multiplex PCR analysis and open access system Motivated by non-invasive detection of fetal DNA in maternal blood (later liquid biopsy)	J. Rachlin CM. Ding C. Cantor N. Alon	XXX
2006	Biological Context Networks (BCN)-s	Pioneered a mini new area of network science	V. Asodi J. Rachlin N. Alon	XXX
2007-	Network Biology of Wellness (many talks given)	In progress	R. Kahn I. Kohane	XXX
2007	<b>Inference of Regulatory Networks in Bacteria with Validation</b> <b>HIGHLY CITED</b>	1700 citations	J. Faith T. Gardner J. Collins	XXX
2000 -	<b>Popularizing Data Integration in Biological Science Using Networks</b>	Large field now	With many independent contributions	XXX
2003-	Active Learning for Biology <b>THE COMBREX PROJECT</b>		Rich Roberts et al	XXX
2010-2017	Regional Bioinformatics and Systems Biology Core	NIH Center		XXX
2012	Driving Citizen Science with AI	NIH Common Fund Meeting	L. Griffith Meyer S. Seung	XXX
2013	<b>The COMBREX REPORT</b> <b>First community driven science by active learning paper (AI2BIO)</b>		COMBREX Consortium	XXX

2014	Reprogramming Stem Cells driven by computational analysis and epigenomics <b>HIGHLY CITED</b>	Highly cited	E. Rheinbay M. Suva B. Bernstein	X
2015	The Biomed Collaboration Initiative	Harvard Medical School		X X
2010-2020	The Provenance Initiative		Rich Roberts et al	X X X
2020-	AI and Biology Initiative	In progress		XXX
2020-	Automating COVID Drug Discovery (small)	In progress		XXX
2020-	Use of Complexity Theory in Pooling COVID-19 TESTS (small)	Theory and system	R. Beigel	XX
1996-2020	Many advisory boards and steering committees at both local, single university or national or international level including Hopkins, BU, Harvard Med, CMU, Columbia, Alberta Innovation Academy, U. Chicago, NSF, Joslin Diabetes Center, start-ups, conferences, industry and more)			
1985-2020	Over 60 students/fellows mentored or co-mentored, many minorities, women <a href="http://sites.bu.edu/phenogeno/people/">http://sites.bu.edu/phenogeno/people/</a>			
1985-2020	Over 70 direct small group collaborators (in addition to consortium memberships and large papers) <a href="http://sites.bu.edu/phenogeno/people/">http://sites.bu.edu/phenogeno/people/</a>			