

Curriculum Vitae

DAVID HAUSSLER

Howard Hughes Medical Institute Investigator

Professor, Biomolecular Engineering

Director, Center for Biomolecular Science and Engineering

1156 High Street, MS: CBSE/ITI

Engineering 2 Building, Suite 501

University of California

Santa Cruz, CA 95064

Tel: (831) 459-2105

Fax: (831) 459-1809

haussler@soe.ucsc.edu

<http://www.cbse.ucsc.edu/staff/haussler.shtml>

Affiliate of Crown College

EMPLOYMENT

- 2004- Professor, Biomolecular Engineering (Transferred to new department)
- 2000- Howard Hughes Medical Institute Investigator
- 1993- 04 Professor, Computer Science, University of California, Santa Cruz
- 1989- 93 Associate Professor, Computer Science, University of California, Santa Cruz
- 1986- 89 Assistant Professor, Computer Science, University of California, Santa Cruz
- 1982 -86 Assistant Professor, Mathematics and Computer Science, University of Denver, CO

EDUCATION

- Ph.D. 1982 University of Colorado at Boulder, Computer Science
Received Graduate Student Research Award
- M.S. 1979 California Polytechnic State University at San Luis Obispo, Applied Mathematics
Received Mathematics Award
- B.A. 1975 Connecticut College, New London, CT, Mathematics
Magna Cum Laude, Phi Beta Kappa, received Julia Bower Mathematics Award

HONORS & AWARDS

- 2006 Elected as a member of the National Academy of Sciences
- 2006 Elected as a Fellow of the American Academy of Arts and Sciences
- 2006 Dickson Prize in Science, Carnegie Mellon University
- 2005 World Technology Network Award, IT Software Category

- 2005 Classic Paper Award, American Association of Artificial Intelligence, for “Quantifying the inductive bias in concept learning,” 1986
- 2005 Distinguished Engineering Alumni Award, University of Colorado, Boulder
- 2004 Allen Newell Award from the Association for Computing Machinery and the American Association for Artificial Intelligence
- 2003 Tech Award Laureate, San Jose Tech Museum of Innovation
- 2003 Elected as a Fellow of the American Association for the Advancement of Science
- 2003 Boston Biomedical/Clinical Ligand Assay Society Distinguished Scientist of the Year Award
- 2001- 02 UCSC Faculty Research Lecturer
- 2001 Incyte Genomics Featured Scientist
- 2001 Elected Fellow of the California Academy of Sciences
- 2001 *Research and Development Magazine* Scientist of the Year
- 2000- 03 Awarded UC Presidential Chair of Computer Science
- 2000 Selected as Howard Hughes Medical Institute Investigator
- 1992- Elected as a Fellow of the American Association of Artificial Intelligence

PUBLICATIONS

Refereed Journal Papers

Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. 2008 Jan 24; [Epub ahead of print]

Zeng J, Yan J, Wang T, Mosbrook-Davis D, Dolan KT, Christensen R, Stormo GD, Haussler D, Lathrop RH, Brachmann RK, Burgess SM. Genome wide screens in yeast to identify potential binding sites and target genes of DNA-binding proteins. *Nucleic Acids Res*. 2008 Jan;36(1):e8. Epub 2007 Dec 17.

Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F, Kober KM, Miller W, Pedersen JS, Pohl A, Raney BJ, Rhead B, Rosenbloom KR, Smith KE, Stanke M, Thakapallayil A, Trumbower H, Wang T, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res*. 2008 Jan;36(Database issue):D773-9. Epub 2007 Dec 17.

Zhu J, Sanborn JZ, Diekhans M, Lowe CB, Pringle TH, Haussler D. Comparative genomics search for losses of long-established genes on the human lineage. *PLoS Comput Biol*. 2007 Dec 14;3(12):e247 [Epub ahead of print]

Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci U S A*. 2007 Nov 20;104(47):18613-8. Epub 2007 Nov 14.

Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW, Crosby MA, Rasmussen MD, Roy S, Deoras AN, Ruby JG, Brennecke J; Harvard FlyBase curators; Berkeley Drosophila Genome Project, Hodges E, Hinrichs AS, Caspi A, Paten B, Park SW, Han MV, Maeder ML, Polansky BJ, Robson BE, Aerts S, van Helden J, Hassan B, Gilbert

DG, Eastman DA, Rice M, Weir M, Hahn MW, Park Y, Dewey CN, Pachter L, Kent WJ, Haussler D, Lai EC, Bartel DP, Hannon GJ, Kaufman TC, Eisen MB, Clark AG, Smith D, Celniker SE, Gelbart WM, Kellis M. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature.* 2007 Nov 8;450(7167):219-32. PMID: 17994088
Karolchik D, Bejerano G, Hinrichs AS, Kuhn RM, Miller W, Rosenbloom KR, Zweig AS, Haussler D, Kent WJ. Comparative genomic analysis using the UCSC Genome Browser. *Methods Mol Biol.* 2007;395:17-34.

Karolchik D, Bejerano G, Hinrichs AS, Kuhn RM, Miller W, Rosenbloom KR, Zweig AS, Haussler D, Kent WJ. Comparative Genomic Analysis Using the UCSC Genome Browser. *Methods Mol Biol.* 2007;395:17-34.

Siepel A, Diekhans M, Brejová B, Langton L, Stevens M, Comstock CL, Davis C, Ewing B, Oommen S, Lau C, Yu HC, Li J, Roe BA, Green P, Gerhard DS, Temple G, Haussler D, Brent MR. Targeted discovery of novel human exons by comparative genomics. *Genome Res.* 2007 Dec;17(12):1763-73. Epub 2007 Nov 7.

Miller W, Rosenbloom K, Hardison RC, Hou M, Taylor J, Raney B, Burhans R, King DC, Baertsch R, Blankenberg D, Kosakovsky Pond SL, Nekrutenko A, Giardine B, Harris RS, Tyekucheveva S, Diekhans M, Pringle TH, Murphy WJ, Lesk A, Weinstock GM, Lindblad-Toh K, Gibbs RA, Lander ES, Siepel A, Haussler D, Kent WJ. 28-Way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Res.* 2007 Dec;17(12):1797-808. Epub 2007 Nov 5.

Yeang CH, Haussler D. Detecting coevolution in and among protein domains. *PLoS Comput Biol.* 2007 Nov 2;3(11):e211. Epub 2007 Sep 18.

Dreszer TR, Wall GD, Haussler D, Pollard KS. Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. *Genome Res.* 2007 Oct;17(10):1420-30. Epub 2007 Sep 4.

Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, Wilson RK, Salama SR, Haussler D. Human genome ultraconserved elements are ultraselected. *Science.* 2007 Aug 17;317(5840):915.

Yeang CH, Darot JF, Noller HF, Haussler D. Detecting the coevolution of biosequences—an example of RNA interaction prediction. *Mol Biol Evol.* 2007 Sep;24(9):2119-31. Epub 2007 Jul 17

ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007 Jun 14;447(7146):799-816.

Margulies EH, Cooper GM, Asimenos G, Thomas DJ, Dewey CN, Siepel A, Birney E, Keefe D, Schwartz AS, Hou M, Taylor J, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Brown JB, Bickel P, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Stone EA, Rosenbloom KR, Kent WJ, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Hinrichs A, Trumbower H, Clawson H, Zweig A, Kuhn RM, Barber G, Harte R, Karolchik D, Field MA, Moore RA, Matthewson CA, Schein JE, Marra MA, Antonarakis SE, Batzoglou S, Goldman N, Hardison R, Haussler D, Miller W, Pachter L, Green ED, Sidow A. Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res.* 2007 Jun;17(6):760-74.

Lowe CB, Bejerano G, Haussler D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci U S A.* 2007 May 8;104(19):8005-10. Epub 2007 Apr 26.

Rhesus Macaque Genome Sequencing and Analysis Consortium. Evolutionary and biomedical insights from the rhesus macaque genome. *Science.* 2007 Apr 13;316(5822):222-34.

Ohlson J, Pedersen JS, Haussler D, Ohman M. Editing modifies the GABAA receptor subunit alpha 3. *RNA.* 2007 May;13(5):698-703. Epub 2007 Mar 16.

Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, Siepel A, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pedersen JS, Hsu F, Hinrichs AS, Harte RA, Diekhans M, Clawson H, Bejerano G, Barber GP, Baertsch R, Haussler D, Kent WJ. The UCSC genome browser database: update 2007. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D668-73. Epub 2006 Nov 16.

Thomas DJ, Rosenbloom KR, Clawson H, Hinrichs AS, Trumbower H, Raney BJ, Karolchik D, Barber GP, Harte RA, Hillman-Jackson J, Kuhn RM, Rhead BL, Smith KE, Thakkapallayil A, Zweig AS; ENCODE Project Consortium; Haussler D, Kent WJ. The ENCODE Project at UC Santa Cruz. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D663-7. Epub 2006 Dec 13.

Thomas DJ, Trumbower H, Kern AD, Rhead BL, Kuhn RM, Haussler D, Kent WJ. Variation resources at UC Santa Cruz. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D716-20. Epub 2006 Dec 6.

Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, Rosenbloom K, Kent J, Haussler D. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* 2006; 2(10):e168.

Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, Kent WJ, Blanchette M, Haussler D, Miller W. Reconstructing contiguous regions of an ancestral genome. *Genome Res.* 2006 Dec;16(12):1557-65. Epub 2006 Sep 18.

Pollard KS, Salama SR, Lambert N, Lambot M-A, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, Kern AD, Dehay C, Igel H, Ares M Jr, Vanderhaeghen P, Haussler D. A rapidly evolving region in the human genome is an RNA gene expressed during early neocortical development. *Nature.* 2006 Sep 14;443(7108):167-72. Epub 2006 Aug 16.

Wu J, Haussler D. Coding exon detection using comparative sequences, *J Comput Biol.* 2006 Jul-Aug;13(6):1148-64.

Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, Haussler D. The UCSC Known Genes. *Bioinformatics.* 2006 May 1;22(9):1036-46. Epub 2006 Feb 24.

Pedersen JS, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, Lander ES, Kent J, Miller W, Haussler D. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol.* 2006 Apr 21;2(4):e33.

Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature.* 2006 May 4;441(7089):87-90. Epub 2006 Apr 16.

Sugnet CW, Srinivasan K, Clark TA, O'Brien G, Cline MS, Wang H, Williams A, Kulp D, Blume JE, Haussler D, Ares M. Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput Biol.* 2006 Jan;2(1):e4. Epub 2006 Jan 20.

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M., Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* 2006 Jan;34(Database issue):D590-8.

Lucena B, Haussler D. Counterexample to a claim about the reconstruction of ancestral character states. *Syst Biol.* 2005 Aug;54(4):693-5.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 2005 Aug;15(8):1034-50.

Bejerano G, Siepel AC, Kent WJ, Haussler D. Computational screening of conserved genomic DNA in search of functional noncoding elements. *Nat Methods.* 2005 Jul;2(7):535-45.

The Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome, *Nature*, 437(7005): 69-87 (2005)

Karchin R, Diekhans M, Kelly L, Thomas DJ, Pieper U, Eswar N, Haussler D, Sali A. LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources. *Bioinformatics*. 2005 Jun 15;21(12):2814-20.

Kent WJ, Hsu F, Karolchik D, Kuhn RM, Clawson H, Trumbower H, Haussler D. Exploring relationships and mining data with the UCSC Gene Sorter. *Genome Res*. 2005 May;15(5):737-41.

Robertson MP, Igel H, Baertsch R, Haussler D, Ares M Jr, Scott WG. The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol*. 2005 Jan;3(1):e5.

Hsu F, Pringle TH, Kuhn RM, Karolchik D, Diekhans M, Haussler D, Kent WJ. The UCSC Proteome Browser. *Nucleic Acids Res*. 2005 Jan 1;33(Database issue):D454-8.

Blanchette M, Green ED, Miller W, Haussler D. Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res*. 2004 Dec;14(12):2412-23.

International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 2004 Dec 9;432(7018):695-716.

MGC Project Team. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res*. 2004 Oct;14(10B):2121-7.

ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*. 2004 Oct 21;306(5696):636-40.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004 Oct 21;431(7011):931-45.

Furey TS, Diekhans M, Lu Y, Graves TA, Oddy L, Randall-Maher J, Hillier LW, Wilson RK, Haussler D. Analysis of human mRNAs with the reference genome sequence reveals potential errors, polymorphisms, and RNA editing. *Genome Res*. 2004 Oct;14(10B):2034-40.

She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L, Clark R, Graves T, Gulden CL, Alkan C, Bailey JA, Sahinalp C, Rocchi M, Haussler D, Wilson RK, Miller W, Schwartz S, Eichler EE. The structure and evolution of centromeric transition regions within the human genome. *Nature*. 2004 Aug 19;430(7002):857-64.

Jojic V, Jojic N, Meek C, Geiger D, Siepel A, Haussler D, Heckerman D. Efficient approximations for learning phylogenetic HMM models from data. *Bioinformatics*. 2004 Aug 4;20 Suppl 1:I161-8.

Bejerano G, Haussler D, Blanchette M. Into the heart of darkness: large-scale clustering of human non-coding DNA. *Bioinformatics*. 2004 Aug 4;20 Suppl 1:I40-8.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. Ultraconserved elements in the human genome. *Science*. 2004 May 28;304(5675):1321-5.

Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*. 2004 Apr 1;428(6982):493-521.

Yang S, Smit AF, Schwartz S, Chiaromonte F, Roskin KM, Haussler D, Miller W, Hardison RC. Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes. *Genome Res*. 2004 Apr;14(4):517-27.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res*. 2004 Apr;14(4):528-38.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* 2004 Apr;14(4):708-15.

Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE. Hotspots of mammalian chromosomal evolution. *Genome Biol.* 2004;5(4):R23.

Roskin KM, Diekhans M, Haussler D. Score functions for determining regional conservation in two-species local alignments. *J Comput Biol.* 2004 Mar;11(2-3):395-411.

Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. *J Comput Biol.* 2004 Mar;11(2-3):413-28.

Siepel A, Haussler D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol Biol Evol.* 2004 Mar;21(3):468-88.

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004 Jan 1;32(Database issue):D493-6.

Margulies EH, Blanchette M, Haussler D, Green ED; NISC Comparative Sequencing Program. Identification and characterization of multi-species conserved sequences. *Genome Res.* 2003 Dec;13(12):2507-18.

Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A.* 2003 Sep 30;100(20):11484-9.

Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin SQ, Maduro VV, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghghi P, Ho SL, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantripop S, Tionson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang LH, Osogawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green ED. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature.* 2003 Aug 14;424(6950):788-93.

Furey TS, Haussler D. Integration of the cytogenetic map with the draft human genome sequence. *Hum Mol Genet.* 2003 May 1;12(9):1037-44.

Winters-Hilt S, Vercoutere W, DeGuzman VS, Deamer D, Akeson M, Haussler D. Highly accurate classification of Watson-Crick basepairs on termini of single DNA molecules. *Biophys J.* 2003 Feb;84(2 Pt 1):967-76.

Hardison RC, Roskin KM, Yang S, Diekhans M, Kent WJ, Weber R, Elnitski L, Li J, O'Connor M, Kolbe D, Schwartz S, Furey TS, Whelan S, Goldman N, Smit A, Miller W, Chiaromonte F, Haussler D. Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res.* 2003 Jan;13(1):13-26.

Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. Human-mouse alignments with BLASTZ. *Genome Res.* 2003 Jan;13(1):103-7.

Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ. The UCSC Genome Browser Database. *Nucleic Acids Res.* 2003 Jan;31(1):51-4.

The Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002 Dec 5;420(6915):520-62.

Cline MS, Karplus K, Lathrop RH, Smith TF, Rogers RG Jr, Haussler D. Information-theoretic dissection of pairwise contact potentials. *Proteins*. 2002 Oct 1;49(1):7-14.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res*. 2002 Jun;12(6):996-1006.

Karchin R, Karplus K, Haussler D. Classifying G-protein coupled receptors with support vector machines. *Bioinformatics*. 2002 Jan;18(1):147-59.

Kent WJ, Haussler D. Assembly of the working draft of the human genome with GigAssembler. *Genome Res*. 2001 Sep;11(9):1541-8.

Vercoutere W, Winters-Hilt S, Olsen H, Deamer D, Haussler D, Akeson M. Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nat Biotechnol*. 2001 Mar;19(3):248-52.

The BAC Resource Consortium. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*. 2001 Feb 15;409(6822):953-8.

The International Human Genome Mapping Consortium. A physical map of the human genome. *Nature*. 2001 Feb 15;409(6822):934-41.

The International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860-921.

Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*. 2000 Oct;16(10):906-14.

Reese MG, Kulp D, Tammana H, Haussler D. Genie--gene finding in *Drosophila melanogaster*. *Genome Res*. 2000 Apr;10(4):529-38.

Jaakkola T, Diekhans M, Haussler D. A discriminative framework for detecting remote protein homologies. *J Comput Biol*. 2000 Feb-Apr;7(1-2):95-114.

Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, Haussler D. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A*. 2000 Jan 4;97(1):262-7.

Spingola M, Grate L, Haussler D, Ares M Jr. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA*. 1999 Feb;5(2):221-34.

Haussler D. Computational genefinding. *Trends in Biochemical Sciences, Supplementary Guide to Bioinformatics*. 1998;23(1):12-15.

Park J, Karplus K, Barrett C, Hughey R, Haussler D, Hubbard T, Chothia C. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *J Mol Biol*. 1998 Dec 11;284(4):1201-10.

Haussler D, Kivinen J, Warmuth M. Sequential prediction of individual sequences under general loss functions. *IEEE Transactions on Information Theory*, 1998 Sep;44(5):1906-25.

Cesa-Bianchi N, Haussler D. A graph-theoretic generalization of the Sauer-Shelah lemma. *Discrete Applied Mathematics*. 1998 Aug;86(1):27-35.

Haussler D, Opper M. Mutual information, metric entropy, and cumulative relative entropy risk. *Annals of Statistics*. 1997 Dec;25(6):2451-92.

Karplus K, Sjölander K, Barrett C, Cline M, Haussler D, Hughey R, Holm L, Sander C. Predicting protein structure using hidden Markov models. *Proteins*. 1997;29(Suppl 1):134-9.

Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997 Fall;4(3):311-23.

Cesa-Bianchi N, Freund Y, Haussler D, Helmbold D, Schapire R, Warmuth M. How to use expert advice. *J. ACM*. 1997 Jul;44(3):427-85.

Alon N, Ben-David S, Cesa-Bianchi N, Haussler D. Scale-sensitive dimensions, uniform convergence, and learnability. *Journal of the ACM* 1997 Jul;44(4):615-31.

Haussler D. A general minimax result for relative entropy. *IEEE Transactions on Information Theory*. 1997 Jul;43(4):1276-80.

Haussler D, Kearns M, Seung HS, Tishby N. Rigorous learning curve bounds from statistical mechanics. *Machine Learning*. 1996 Nov;25(2/3):195-236.

Fayyad U, Haussler D, Stolorz P. Mining scientific data. *Communications of the ACM*, 1996 Nov;39(11):51-7.

Sjölander K, Karplus K, Brown M, Hughey R, Krogh A, Mian IS, Haussler D. Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology. *Comput Appl Biosci*. 1996 Aug;12(4):327-45.

Opper M, Haussler D. Bounds for predictive errors in the statistical mechanics of supervised learning. *Phys Rev Lett*. 1995 Nov 13;75(20):3772-5.

Haussler D, Long PM. A generalization of Sauer's lemma. *Journal of Combinatorial Theory Series A*. 1995 Aug;71(2):219-40.

Knill E, Ehrenfeucht A, Haussler D. The Size of K-pseudotrees. *Discrete Mathematics* 1995 May;141(1-3):185-94.
Haussler D. Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. *Journal of Combinatorial Theory Series A*. 1995 Feb;69(2):217-32.

Ben-David S, Cesa-Bianchi N, Haussler D, Long P. Characterizations of learnability for classes of $\{0, \dots, n\}$ -valued functions. *Journal of Computer Systems Science*. 1995;50(1):74-86.

Haussler D, Littlestone N, Warmuth M. Predicting 0,1-functions on Randomly Drawn Points. *Information and Computation*. 1994 Dec;115(2):248-92.

Sakakibara Y, Brown M, Hughey R, Mian IS, Sjölander K, Underwood RC, Haussler D. Stochastic context-free grammars for tRNA modeling. *Nucleic Acids Res*. 1994 Nov 25;22(23):5112-20.

Krogh A, Mian IS, Haussler D. A hidden Markov model that finds genes in E. coli DNA. *Nucleic Acids Res*. 1994 Nov 11;22(22):4768-78.

Krogh A, Brown M, Mian IS, Sjölander K, Haussler D. Hidden Markov models in computational biology. Applications to protein modeling. *J Mol Biol*. 1994 Feb;235(5):1501-31.

Haussler D, Kearns M, Schapire R. Bounds on the sample complexity of Bayesian learning using information theory and the VC dimension. *Machine Learning*. 1994 Jan;4(1):83-114.

Haussler D. Decision Theoretic Generalizations of the PAC model for neural net and other learning applications. *Information and Computation*. 1992 Sept;100(1):78-150.

Opper M, Haussler D. Generalization performance of Bayes optimal prediction algorithm for learning a perceptron. *Physical Review Letters*. 1991 May;66(20):2677-81.

- Haussler D, Kearns M, Littlestone N, Warmuth M. Equivalence of models of polynomial learnability. *Information and Computation*. 1991 Dec;95(2):129-61.
- Pagallo G, Haussler D. Boolean feature discovery in empirical learning. *Machine Learning*, 1990 Mar;5(1):71-99.
- Ehrenfeucht A, Haussler D. Learning decision trees from random examples. *Information and Computation*. 1989 Sep;82(3):231-46.
- Blumer A, Ehrenfeucht A, Haussler D. Average sizes of suffix trees and DAWGs. *Discrete Applied Mathematics*. 1989;24(1):37-45.
- Baum E, Haussler D. What size net gives valid generalization. *Neural Computation*. 1989;1(1):151-60.
- Haussler D. Learning conjunctive concepts in structural domains. *Machine Learning*. 1989 Oct;4(1):7-40.
- Blumer A, Ehrenfeucht A, Haussler D, Warmuth M. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*. 1989 Oct;36(4):929-65.
- Ehrenfeucht A, Haussler D, Kearns M, Valiant L. A general lower bound on the number of examples needed for learning. *Information and Computation*. 1989 Sep;82(3):247-61.
- Ehrenfeucht A, Haussler D. A new distance metric on strings computable in linear time. *Discrete Applied Mathematics*. 1988;20(3):191-203.
- Haussler D. Quantifying inductive bias: AI learning algorithms and valiant's learning framework. *Artificial Intelligence*. 1988;36(2):177-221.
- Ehrenfeucht A, Haemer J, Haussler D. Quasi-monotonic sequences: theory, algorithms and applications. *SIAM J Alg and Disc Meth*. 1987;8(3):410-29.
- Blumer A, Ehrenfeucht A, Haussler D, Warmuth M. Occam's razor. *Inf Proc Let*. 1987 Apr;24(6):377-80.
- Haussler D, Welzl E. Epsilon-nets and simplex range queries. *Discrete and Computational Geometry*. 1987;2(2):127-51.
- Blumer A, Blumer J, Ehrenfeucht A, Haussler D, McConnell R. Complete inverted files for efficient text retrieval and analysis. *J. ACM*. 1987;34(3):578-89.
- Main MG, Bucher W, Haussler D. Applications of an infinite squarefree co-CFL. *Theoretical Computer Science*. 1987;49(2,3):113-20.
- Clift B, Haussler D, McConnell R, Schneider TD, Stormo GD. Sequence landscapes. *Nucleic Acids Res*. 1986 Jan 10;14(1):141-58.
- Edelsbrunner H, Haussler D. The complexity of cells in three dimensions. *Discrete Mathematics*. 1986;60(1):139-46.
- Bucher W, Ehrenfeucht A, Haussler D. On total regulators generated by derivation relations. *Theoretical Computer Science*. 1985;40(0):131-48.
- Haussler D. Another generalization of Higman's well-quasi-order result on a finitely generated free monoid. *Discrete Mathematics*. 1985;57(3):237-43.
- Blumer A, Blumer J, Haussler D, Ehrenfeucht A, Chen MT, Seiferas J. The smallest automaton recognizing the subwords of a text. *Theoretical Computer Science*. 1985;40(1):31-56.

Warmuth M, Haussler D. On the complexity of iterated shuffle. *Journal Computer Systems Science*. 1984 Jun;28(3):345-58.

Ehrenfeucht A, Haussler D, Rozenberg G, Zeiger P. On DOS mappings and DOS languages. *Semigroup Forum*. 1984;29(1):123-48.

Ehrenfeucht A, Haussler D, Rozenberg G. On ambiguity in DOS systems. *RAIRO Informatique Theorique*. 1984;18(1):279-95.

Ehrenfeucht A, Haussler D, Rozenberg G. On regularity of context-free languages. *Theoretical Computer Science*. 1983;27(3):311-32.

Haussler D. Insertion languages. *Information Sciences*. 1983;31:77-89.

Haussler D. Model Completeness of an algebra of languages. *Proc. Amer. Math. Soc.* 1981;83(2):371-3.

Haussler D, Zeiger P. Very special languages and representations of recursively enumerable languages. *Information and Control*. 1980;47(3):201-11.

Brumbaugh P, Haussler D, Bressler R, Haussler M. Radioreceptor assay for 1 alpha, 25-dihydroxyvitamin D3. *Science*. 1974;183(3):1089-91.

Chapters in Books

Siepel A, Haussler D. Phylogenetic hidden Markov models. In Nielsen R. (ed): *Statistical Methods in Molecular Evolution*, Springer, 2004.

Haussler D, Opper M. Metric entropy and minimax risk in classification. In Mycielski J, Rozenberg G, Salomaa A (eds): *Lecture Notes in Computer Science: Studies in Logic and Computer Science*, 1997;1261:212-35.

Haussler D, Barron A. How well do Bayes methods work for on-line prediction of +1,-1 values? In *Proceedings of the Third NEC Symposium on Computation and Cognition*, SIAM, Princeton, NJ, 1992:74-100.

Haussler D. Decision theoretic generalizations of the PAC model for neural net and other learning applications. In *Workshop on Supervised Learning*, Santa Fe Institute Press, 1995.

Haussler D, Warmuth M. The probably approximately correct (PAC) and other learning models. In *Workshop on Supervised Learning, Santa Fe Institute Press*, 1995.

Haussler D, Warmuth M. The probably approximately correct (PAC) and other learning models. In Meyrowitz A, Chipman S (eds): *Machine Learning: Induction, Analogy and Discovery*.

Haussler D. Generalizing the PAC model. In Mehra P, Wah BW (eds): *Artificial Neural Networks: Concepts and Theory*, IEEE Society Press, 1992.

Haussler D. PAC learning and generalization. In Smolensky P, Mozer M, Rumelhart D (eds): *Mathematical Perspectives on Neural Networks*.

Haussler, D. Occam's razor; Quantifying inductive bias: AI learning algorithms and Valiant's learning framework; and What size net gives valid generalization. Reprinted in Shavlik J, Dietterich T (eds): *Readings in Machine Learning*, Morgan Kaufmann, Los Altos, CA, 1990.

Haussler D. Applying Valiant's learning framework to AI concept learning problems. In Michalski R, Kodratoff Y (eds): *Machine Learning: An Artificial Intelligence Approach*, Vol. III, Morgan Kaufmann, Los Altos, 1990:641-69.

Conference Papers

Darot J, Yeang CH, Haussler D. Detecting the dependent evolution of biosequences. Proceedings of the 10th International Conference on Research in Computational Molecular Biology, RECOMB, 2006.

Siepel A, Pollard KS, Haussler D. New methods for detecting lineage-specific selection. Proceedings of the 10th International Conference on Research in Computational Molecular Biology, RECOMB, 2006.

Hardison RC, Chiaromonte F, Kolbe D, Wang H, Petrykowska H, Elnitski L, Yang S, Giardine B, Zhang Y, Riemer C, Schwartz S, Haussler D, Roskin KM, Weber RJ, Diekhans M, Kent WJ, Weiss MJ, Welch J, Miller W. Global predictions and tests of erythroid regulatory regions. Cold Spring Harbor Symposium, Quant. Biology, 2003;68:335-44.

Chiaromonte F, Weber RJ, Roskin KM, Diekhans M, Kent WJ, Haussler D. The share of human genomic DNA under selection estimated from human-mouse genomic alignments. Cold Spring Harbor Symposium, Quant. Biology, 2004;68:245-54.

Siepel A, Haussler D. Computational identification of evolutionarily conserved exons. Proceedings of the 8th International Conference on Research in Computational Molecular Biology, RECOMB, 2004;177-86.

Jojic V, Jojic N, Meek C, Geiger D, Siepel A, Haussler D, Heckerman D. Efficient approximations for learning phylogenetic HMM models from data. Proceedings of ISMB 2004 and Bioinformatics, 2004.

Siepel A, Haussler D. Computational identification of evolutionarily conserved exons. Proceedings of the 8th Annual International Conference on Research in Computational Molecular Biology, RECOMB, 2004.

Sugnet CW, Kent WJ, Ares M Jr, Haussler D. Transcriptome and genome conservation of alternative splicing events in humans and mice. Pacific Symposium on Biocomputing (PSB), 2004. [Online Proceedings]

Wang H, Hubbell E, Hu JS, Mei G, Cline M, Lu G, Clark T, Saini-Rose MA, Ares M, Kulp DC, Haussler D. Gene-structure-based splice variant deconvolution using a microarray platform. Proceedings of ISMB and Bioinformatics, 2003;19(Suppl. 1):i315-22.

Roskin KM, Diekhans M, Haussler D. Scoring two-species local alignments to statistically separate neutrally evolving from selected DNA segments. Proceedings of the 7th Annual International Conference on Research in Computational Molecular Biology, RECOMB, 2003;257-66.

Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. Proceedings of the 7th Annual International Conference on Research in Computational Molecular Biology, RECOMB, 2003;277-86.

Rogic S, Furey TS, Kent J, Haussler D. Human genetic discovery using STS markers and the draft assembly browser. Proceedings of Genome Sequencing and Biology, Cold Spring Harbor Laboratory, New York, 2001 May;221.

Furey TS, Rogic S, Kent J, and Haussler D. From STS and cytogenetic maps to the human genome assembly and back. Proceedings of Genome Sequencing and Biology, Cold Spring Harbor Laboratory, New York, 2001 May;76.

Pavlidis P, Furey T, Liberto M, Haussler D, Grundy WN. Promoter region-based classification of genes. Proceedings of the Pacific Symposium on Biocomputing, Hawaii, 2000.

Haghighi F, Diekhans M, Haussler D, Grundy WN. Discriminative gene finding methods. Genome Sequencing and Biology, Cold Spring Harbor Laboratory, New York, May, 2000.

Kent J, Kulp D, Wheeler R, Reese M, Zahler A, Haussler D. Alternate splicing of human genes. Genome Sequencing and Biology, Cold Spring Harbor Laboratory, New York, May, 2000.

Jaakkola T, Diekhans M, Haussler D. Using the Fisher kernel method to detect remote protein homologies. Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology, AAAI Press, 1999;149-158. [Winner of best paper award]

Lazareva-Ulitsky B, Haussler D. A probabilistic approach to consensus multiple alignment. Proceedings of the Pacific Symposium on Biocomputing, World Scientific Press, 1999;150-61.

Jaakkola T, Haussler D. Probabilistic kernel regression models. Proceedings of Seventh Workshop on Artificial Intelligence and Statistics, 1998.

Jaakkola T, Haussler D. Exploiting generative models in discriminative classifiers. Advances in Neural Information Processing Systems, 1998.

Opper M, Haussler D. Worst case prediction over sequences under log loss. The Mathematics of Information Coding, Extraction and Distribution, Cybenko G, O'Leary D, Rissanen J, eds., Springer Verlag, 1997.

Reese M, Eeckman F, Kulp D, Haussler D. Improved splice site detection in genie. First International Conference on Computational Molecular Biology, Santa Fe, New Mexico, 1997.

Kulp D, Haussler D, Reese M, Eeckman F. Integrating database homology in a probabilistic gene structure model. Pacific Symposium on Biocomputing, World Scientific Press, Hawaii, 1997.

Haussler D. A brief look at some machine learning problems in genomics. Proceedings of the 10th Annual Computational Learning Theory Conference, ACM Press, Santa Cruz, California, 1997;109-13.

Fayyad U, Haussler D, Stolorz P. KDD for science data analysis: issues and examples. Proc. Third International Conference on Knowledge Discovery and Data Mining, Portland, OR, 1996.

Kulp D, Haussler D, Reese M, Eeckman F. A generalized hidden Markov model for the recognition of human genes in DNA. Fourth International Conference on Intelligent Systems in Molecular Biology, St. Louis, MO, AAAI/MIT Press, June 1996.

Mirelli V, Haussler D. A hybrid parametric/case-based approach to object recognition using Bayes decision theory. Proceedings of SPIE Symposium on Electronic Imaging: Science and Technology, San Jose, CA, February 1996.

Gulko B, Haussler D. Using multiple alignments and phylogenetic trees to detect RNA secondary structure. Proceedings of the Pacific Symposium on Biocomputing, pp. 350-367, L. Hunter and T. Klein, eds., World Scientific Press, 1996 Jan; 350-67.

Haussler D. A generalized hidden Markov model for DNA parsing. extended abstract of talk for the Workshop on Gene-Finding and Gene Structure Prediction, Univ. of Pennsylvania, October 1995.

Opper M, Haussler D. General bounds for predictive errors in supervised learning. Proceedings of the Workshop on the Theory of Neural Networks: The Statistical Mechanics Perspective, World Scientific Press, 1995:51-8.

Haussler D, Opper M. Mutual information and Bayes methods for learning a distribution. Proceedings of the workshop on the Theory of Neural Networks: The Statistical Mechanics Perspective, World Scientific Press, 1995:42-50.

Haussler D, and Opper M. General bounds on the mutual information between a parameter and n conditionally independent observations. Proceedings of the Eighth Annual Computational Learning Theory Conference (COLT), Santa Cruz, CA, ACM Press, 1995;402-11.

Haussler D, Kivinen J, Warmuth M. Tight worst-case loss bounds for predicting with expert advice. Proc. of the European Conference on Computational Learning Theory (EUROCOLT), 1994.

Haussler D, Kearns M, Seung HS, Tishby N. Rigorous learning curve bounds from statistical mechanics. Proceedings of the Seventh ACM Conference on Computational Learning Theory, ACM Press, 1994.

Stormo G, Haussler D. Optimally parsing a sequence into different classes based on multiple types of evidence. Second International Conference on Intelligent Systems in Molecular Biology, Menlo Park, California, AAAI/MIT Press, 1994;369-75.

Grate L, Herbster M, Hughey R, Haussler D, Mian IS, Noller H. RNA Modeling using Gibbs sampling and stochastic context-free grammars. Second International Conference on Intelligent Systems in Molecular Biology, Menlo Park, CA, AAAI/MIT Press, 1994 Aug;138-46.

Sakakibara Y, Brown M, Hughey R, Mian IS, Sjölander K, Underwood R, Haussler D. Stochastic context-free grammars for modeling RNA. Proceedings of the Hawaii International Conference on System Sciences, Los Alamitos, CA, IEEE Computer Society Press, 1994.

Sakakibara Y, Brown M, Hughey R, Mian IS, Sjölander K, Underwood R, Haussler D. Recent methods for RNA modeling using stochastic context-free grammars. Proceedings of the Asilomar Conference on Combinatorial Pattern Matching, Springer-Verlag publisher, New York, NY, 1994.

Alon N, Ben-David S, Cesa-Bianchi N, Haussler D. Scale-sensitive dimensions, uniform convergence, and learnability. 34th IEEE Symposium on Foundations of Computer Science, Palo Alto, CA, October, 1993.

Haussler D. Dirichlet mixture priors for hidden Markov models. Joint Meeting of the American Statistical Association, San Francisco, CA, August 1993.

Brown M, Hughey R, Mian IS, Sjölander K, Underwood R, Haussler D. Using Dirichlet mixture priors to derive hidden Markov models for protein families. First International Conference on Intelligent Systems for Molecular Biology, Washington, DC, July 1993.

Barron A, Clarke B, Haussler D. Information bounds for the risk of Bayesian predictions and the redundancy of universal codes. International Symposium on Information Theory, San Antonio, TX, January 1993.

Krogh A, Mian IS, Haussler D. Parsing DNA with hidden Markov models. Alternative Readings of the Genetic Code, Parknasilla, County Kerry, Ireland, May 1993.

Haussler D, Krogh A, Brown M, Mian IS, Sjölander K. Protein modeling with hidden Markov models: an analysis of globins. 26th Hawaii Systems Conference, January 1993. [awarded best paper in AI methods in biotechnology track]

Cesa-Bianchi N, Freund Y, Helmbold D, Haussler D, Schapire R, Warmuth M. How to use expert advice. (extended abstract), 25th ACM Symposium on Theoretical Computer Science (STOC), 1993;382-91.

Haussler D, Krogh A. DNA Alignment and clustering. Neural Networks for Computing, Snowbird, Utah, Sandi von Pier (ed), AT&T Bell Laboratories, Crawfords Corner Road, Rm. 4E-422, Holmdel, NJ 07733. April 7, 1992.

Haussler D, Barron A. How well do Bayes methods work for on-line prediction of +1,-1 values? Third NEC Symposium on Computation and Cognition, Princeton, NJ, 1992.

Haussler D, Kearns M, Opper M, Schapire R. Estimating average-case learning curves using Bayesian, statistical physics and VC dimension methods. Fifth Conference on Neural Information Processing Systems, Denver, CO, 1991.

Freund Y, Haussler D. Unsupervised learning of distributions on binary vectors using two-layer networks. Fifth Conference on Neural Information Processing Systems, Denver, CO, 1991.

Haussler D, Kearns M, Schapire R. Bounds on the sample complexity of Bayesian learning using information theory and the VC dimension. Fourth Workshop on Computational Learning Theory (COLT), Santa Cruz, CA, 1991 Aug;61-74.

Opper M, Haussler D. Calculation of the learning curve of Bayes optimal classification algorithm for learning a perceptron with noise. Fourth Workshop on Computational Learning Theory (COLT), Santa Cruz, CA, 1991 Aug;75-87.

Haussler D. Uniting the VC and TLS theories of generalization. Neural Networks for Computing, Snowbird, UT, April 1991.

Haussler D, Long P. A generalization of Sauer's lemma. Southeastern Conference on Combinatorics, Graph Theory and Computing, Baton Rouge, LA, February 1991.

Haussler D. Probably approximately correct learning. AAAI, 1990;1101-8.

Haussler D. Sample size bounds for training multi-layer nets of quasi-linear, product and radial basis functions. Neural Networks for Computing, Snowbird, UT, April 1990.

Haussler D, Warmuth M. Analyzing the performance of learning algorithms. ONR Workshop on Knowledge Acquisition, Crystal City, VA, November 1989.

Milosavljevic A, Haussler D, and Jurka J. Parsimonious classification of aligned molecular sequences. Bio-Matrix, White Mountain Conference Center, Waterville Valley, NH, August 1989.

Haussler D. Generalizing the PAC Model: Sample size bounds from metric dimension-based uniform convergence results. 1989 IEEE Symp. on Foundations of Computer Science (FOCS), Research Triangle, NC, 1989 Oct;40-5.

Milosavljevic A, Haussler D, Jurka J. Informed parsimonious inference of prototypical genetic sequences. Second Workshop on Computational Learning Theory, Santa Cruz, CA, August 1989.

Pagallo G, Haussler D. Two algorithms that learn DNF by discovering relevant features. Sixth International Workshop on Machine Learning, Cornell University, Ithaca, NY, July 1989.

Baum E, Haussler D. What size net gives valid generalization. IEEE Conference on Neural Information Processing Systems, Denver, CO, November 1988.

Haussler D, Littlestone N, Warmuth M. Predicting 0,1-functions on randomly drawn points. 29th IEEE Symposium on Foundations of Computer Science, White Plains, NY, 1988 Oct;100-9.

Haussler D, Kearns M, Littlestone N, Warmuth M. Equivalence for models of polynomial learnability. First Workshop on Computational Learning Theory, MIT, Cambridge, MA, August 1988.

Ehrenfeucht A, Haussler D. Learning decision trees from random examples. First Workshop on Computational Learning Theory, MIT, Cambridge, MA, August 1988.

Ehrenfeucht A, Haussler D, Kearns M, Valiant L. A general lower bound on the number of examples needed for learning. First Workshop on Computational Learning Theory, MIT, Cambridge, MA, August 1988.

Haussler D. Learning conjunctive concepts in structural domains. AAA, Seattle, WA, 1987 July;466-70.

Haussler D. Bias, Version spaces and Valiant's learning framework. 4th International Workshop on Machine Learning, Irvine, CA, June 1987.

Alon N, Haussler D, Welzl E. Partitioning and geometric embedding of range spaces of finite Vapnik-Chervonenkis dimension. 3rd International Conference on Computational Geometry, Waterloo, Canada, 1987 June;331-40.

Blumer A, Ehrenfeucht A, Haussler D. Average sizes of suffix trees and DAWGs. 1st Montreal Conference on Combinatorics and Computer Science, University of Montreal, Canada, May 1987.

Haussler D. Learning internal disjunctive concepts. 20th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, November 1986.

Haussler D. Quantifying the inductive bias in concept learning. AAAI, Philadelphia, PA, August, 1986.

Haussler D, Welzl E. Range spaces and Epsilon-nets. SIAM Workshop on Computational and Discrete Geometry, Santa Cruz, CA, July 1986.

Haussler D. A methodology for assessing the learnability of knowledge structures. 1st Annual Rocky Mountain Conference on Artificial Intelligence, Boulder, CO, June 1986.

Haussler D, Welzl E. Epsilon-nets and simplex range queries. 2nd International Conference on Computational Geometry, Yorktown Heights, NY, June 1986.

Blumer A, Ehrenfeucht A, Haussler D, Warmuth M. Classifying learnable geometric concepts with the Vapnik-Chervonenkis dimension. 18th ACM Symposium on Theoretical Computational Science, Berkeley, CA, 1986 May;273-82.

Bucher W, Ehrenfeucht A, Haussler D. On total regulators generated by derivation relations. Proc. 12th International Coll. Aut. Lang. Prog, Nafplion, Greece, 1985 July;71-9.

Main MG, Bucher W, Haussler D. Applications of an infinite squarefree co-CFL. Proc. 12th International Coll. Aut. Lang. Prog., Nafplion, Greece, 1985 July;404-12.

Blumer A, Blumer J, Ehrenfeucht A, Haussler D, McConnell R. Building the minimal DFA for the set of all subwords of a word on-line in linear time. Proc. 11th International Coll. Aut. Lang. Prog., Antwerp, Belgium, 1984 July;109-18.

Blumer A, Blumer J, Ehrenfeucht A, Haussler D, McConnell R. Building a complete inverted file for a set of text files in linear time. Proceedings of the 16th ACM Symposium on Theoretical Computational Science, Washington, D.C., 1984 May;349-58.

Ehrenfeucht A, Rozenberg G, Haussler D. Conditions enforcing regularity of context-free languages. Proc. 9th Int. Coll. Aut. Lang. Prog., Aarhus, Denmark, 1982 July;187-91.

Non-refereed Papers

Roskin KM, Diekhans M, Kent, WJ, Haussler D. Score functions for assessing conservation in locally aligned regions of DNA from two species. UCSC Tech Report UCSC-CRL-02-30. Sep, 2002.

Haussler D. The challenge of bioinformatics. R&D Magazine. 2001 Nov;43(11):8S-SC3 (in conjunction with Scientist of the Year Award).

Haussler D, Jaakkola T, Winters-Hilt S. Tradeoffs between generative and discriminative hidden Markov models. Computer Science Department, UC Santa Cruz. 1998.

PROFESSIONAL ACTIVITIES

Invited Conference Lectures

2007 “Reconstructing 100 million years of human evolutionary history,” honored speaker, NIH Intramural Sequencing Center’s 10 Anniversary Symposium, Bethesda, MC. October

“Rapidly evolving non-coding regions and brain evolution”, honored speaker, 57th Annual ASHG Meeting, San Diego, CA. August

“Reconstructing 100 million years of human evolutionary history,” Pacific Biosciences, Menlo Park, CA. July.

“Reconstructing 100 million years of human evolutionary history,” honored speaker, Vertebrate Comparative Genomics, US National Academy of Sciences and the Australian Academy of Science, Beckman Center of the National Academy of Sciences, Irvine, CA. May.

“Reconstructing the evolutionary history of the human genome,” Frontiers in Biology seminar series, Stanford University, Palo Alto, CA. May.

“The infinite sites model of genome evolution,” Biology of Genomes meeting, Cold Springs Harbor Conference, Cold Springs Harbor, NY. May.

“A rapidly evolved RNA gene may have played a role in the evolution of the cerebral cortex,” honored speaker, UCSD Project for Explaining the Origin of Humans Symposium, University of California, San Diego, CA. March.

2006 “Reconstructing the evolutionary history of the human genome,” 3rd University of California, Santa Cruz/QB3 Symposium of Bioinformatics, Santa Cruz, CA. December.

“Reconstructing 100 million years of human evolutionary history,” Center for Algorithmic and System Biology (CASB) Algorithmic Biology Meeting, University of California, San Diego, La Jolla, CA. November.

“Reconstructing 100 million years of human evolutionary history,” Johns Hopkins University, Baltimore, MD. November.

“Ultraconserved elements, living fossil transposons, and rapid bursts of change: reconstructing the uneven evolutionary history of the human genome,” International Symposium on Genomics, Hangzhou, China, October.

“Reconstructing 100 million years of human evolutionary history,” Beckman Center for Molecular and Genetic Medicine at Stanford, Stanford, CA. October.

“Reconstructing 100 million years of human evolutionary history,” honored speaker, Scientific Forum for Cabrillo Festival of Contemporary Music, world premiere of the Frans Lanting orchestral-photographic performance, “Life: a journey through time,” Santa Cruz, CA. July.

“Exploring the evolutionary history of the human genome over the last 100 million years,” keynote address, ACM SIGMOND International conference on Management of Data, Chicago, Illinois. June.

“Ultraconserved elements, living fossil transposons, and rapid bursts of change: reconstructing the uneven evolutionary history of the human genome,” 23rd International Conference on Machine Learning (ICML), Carnegie Mellon University, Pittsburgh, PA. June.

“Ultraconserved elements, living fossil transposons, distal enhancers, and mysterious RNA genes: reconstructing the detailed evolutionary history of the human genome,” Annual Society for Molecular Biology and Evolution (SMBE) Meeting, Genomes, Evolution, and Bioinformatics, Arizona State University. May.

“Ultraconserved elements, living fossil transposons, distal enhancers, and mysterious RNA genes: reconstructing the detailed evolutionary history of the human genome,” inaugural speech. Broad Distinguished Lecture Series in Computational Biology, Eli & Edythe L Broad Institute, Cambridge, MA. May.

“Feasibility of reconstructing the 100 million year history of the human genome,” Genomes and Biology, Cold Spring Harbor Laboratory, New York. May.

"Exploring the evolutionary history of the human genome over the last 100 million years," Human Evolution Symposium, University of California, San Francisco. May.

“Applying comparative genomics to assist in the completion of the Mammalian Gene Collection,” Mammalian Gene Collection Annual Meeting, Washington, DC. April.

“The infinite sites model of genome evolution,” Workshop on Reconstruction of Ancestral Genomes, Barbados. April.

“Ultraconserved elements, living fossil transposons, and rapid bursts of change: reconstructing the uneven evolutionary history of the human genome,” keynote address. 10th Annual Conference on Research in Computational Molecular Biology (RECOMB 2006), Venice Italy. March.

“Ultraconserved elements, living fossil transposons, and rapid bursts of change: reconstructing the uneven evolutionary history of the human genome,” Dickson Prize award talk, Carnegie-Mellon University, Pittsburg, PA. March.

“Ultraconserved elements, living fossil transposons, and rapid bursts of change: reconstructing the uneven evolutionary history of the human genome,” keynote address. Systems Biology: Global Regulation of Gene Expression, Cold Spring Harbor Laboratory, New York. March.

“Ultraconserved elements, living fossil transposons, distal enhancers, and mysterious RNA genes: reconstructing the detailed evolutionary history of the human genome,” Institute for Pure and Applied Mathematics, University of California, Los Angeles. January.

"Ultraconserved elements, living fossil transposons, distal enhancers, and mysterious RNA genes: reconstructing the detailed evolutionary history of the human genome," Integrative Graduate Education and Research Traineeships, University of Arizona. January.

2005 “Health, medicine, and biotechnology,” keynote address, World Technology Summit, San Francisco, CA. November.

“Using evolution to explore the human genome,” honored speaker, European Bioinformatics Institute, Hinxton, Cambridge, UK. November

“Using evolution to explore the human genome,” Evolutionary Genomics Meeting, The Stazione Zoologica Anton Dohrn, Naples, Italy. October.

“Using evolution to explore the human genome,” keynote address. Genome Informatics Conference, Cold Spring Harbor, New York. October.

“Using evolution to explore the human genome,” honored speaker, UC System-Wide Bioengineering Symposium, Santa Cruz. June.

“Using evolution to explore the human genome,” keynote address, IEEE Workshop on Computer Vision Methods for Bioinformatics (CVMB), San Diego. June.

“Computational reconstruction of an ancestral mammalian chromosome,” keynote address, Cold Spring Harbor Conference, Cold Spring Harbor, New York. May.

“Reconstructing an ancestral mammalian chromosome,” keynote address, Cold Spring Harbor Conference, Cold Spring Harbor, NY. May.

- “Ultraconserved elements in the human genome,” European Human Genetics Conference, Prague, Czech Republic. May.
- “Impact of human genome research: present and future,” honored speaker, Mervyn Young Memorial Lecture, University of Colorado, Boulder, CO. April.
- “Using evolution to explore the human genome,” honored speaker, Distinguished Engineering Alumni Award Seminar, University of Colorado, Boulder, CO. April.
- 2004 “Using evolution to explore the human genome,” honored speaker, University of Washington CSE Distinguished Lecture, Seattle, WA. December.
- “Using evolution to explore the human genome,” National Academy of Science Sackler Colloquium, Stanford University, Palo Alto, CA. October.
- “Comparative genomics to identify functional elements in the human genome,” Molecular Medicine Tri-Conference, San Francisco, CA March.
- “The power of multiple vertebrate genome sequences,” Keystone Symposia, Steamboat Springs, CO. March.
- “Reconstructing an ancestral mammalian genome in silico,” Gordon Research Conference in Molecular Evolution, Ventura, CA. February.
- “Comparing the human, chimp, mouse and rat genomes: using evolution to predict functional elements in the human genome,” Sali Labs, University of California, San Francisco, CA. January.
- 2003 “Using comparative genomics to predict functional elements in the human genome,” Sloan-Swartz Center for Theoretical Neurobiology Seminar, Distinguished speaker Series, Colorado State University, Ft. Collins, CO. November.
- “Using comparative genomics to predict functional elements in the human genome,” XIX International Congress of Genetics, Melbourne, Australia, July.
- “Identifying functional elements in the human genome by tracing the evolutionary history of the bases: a key challenge for comparative genomics,” Intelligent Systems for Molecular Biology, Brisbane, Australia, June.
- “Bioinformatics, genome evolution and the challenge of identifying functional elements in the human genome,” The Genome of Homo Sapiens. Symposium on Quantitative Biology, Cold Spring Harbor Laboratory, NY. May.
- “Computational analysis of the human and other mammalian genomes,” International Conference on Artificial Intelligence and Statistics,” Key West, FL, January.
- “Computational analysis of the human and other mammalian genomes,” Oncogenomics, Phoenix, AZ, January.
- 2002 “Initial computational analysis of the public working drafts of the human and mouse genomes, and the long road ahead,” Genomes and Biology Conference, Cold Spring Harbor Laboratory, NY, May.
- “Computational analysis of the human and mouse genomes,” Asilomar Conference on Signals, Systems, and Computers, Asilomar Conference Center, Pacific Grove, CA, November.
- “The working drafts of the human and mouse genomes,” Osong International Bio Conference, South Korea, October.

- “The working drafts of the human and mouse genomes,” IBC’s Post-Genomic Bioinformatics Conference, San Francisco, CA, June.
- “Initial computational analysis of the public working drafts of the human and mouse genomes, and the long road ahead,” Workshop on Theory of Computation and the Sciences, Berkeley, CA, May.
- “Assembly and initial analysis of the working draft of the human genome,” Clinical Ligand Assay Society, in conjunction with Boston Biomedica/CLAS Distinguished Scientist of the Year Award, 28th Annual Meeting, Houston, TX, May.
- “Status of the human and mouse working draft genomes, and some early comparisons,” Genes and Genomes in Health and Disease, HHMI, Washington, DC. March.
- “Status of the human genome working draft: map, assembly and web access,” Advances in Genome Biology & Technology, joint meeting with Automation in Mapping & DNA Sequencing, Marco Island, FL. February.
- 2001 “Overview of the human genome project and the construction of the working draft” and “Exploring the working draft of the human genome,” NATO ASI on Artificial Intelligence and Heuristic Methods for Bioinformatics, San Miniato, Italy, October.
- “Assembly and annotation of the public human genome working draft,” 13th International Genome Sequencing and Analysis Conference, San Diego, CA, October.
- “The public working draft of the human genome,” keynote address, CHI Bioinformatics and Genome Research, San Francisco, June.
- “Assembly and annotation of the working draft of the human genome,” Gordon Research Conference on Human Molecular Genetics, Newport, RI, August.
- “The public working draft of the human genome,” 33rd Annual Symposium on the Interface of Computer Science and Statistics, Costa Mesa, CA, May.
- “Assembly and initial analysis of a working draft of the human genome,” New Frontiers in Structural and Computational Biology, Howard Hughes Medical Institute, Chevy Chase, MD. March.
- “A working draft of the human genome,” Pacific Symposium on Biocomputing, Mauna Lani, HI, January.
- 2000 “A working draft of the human genome,” Biomedical Computation at Stanford 2000, Stanford University, Palo Alto, CA, October.
- “Hidden Markov models, Fisher kernels and support vector machines for biosequence analysis,” Bioinformatics 2000, Elsinore, Denmark, April.
- “Computational prediction of genes and gene function from high-throughput genomics data,” AAAS 2000, Washington D.C. February.
- 1999 “Hidden Markov models and Fisher kernels for biosequence analysis,” Learning 1999, 13th Annual Workshop, Snowbird, UT, April.
- “Combining discriminative classification methods with hidden Markov models for more effective biosequence analysis,” Eighth Bioinformatics and Genome Research Conference, San Francisco, CA, June.
- “Analysis of microarray gene expression data using support vector machines,” In Silico Biology: Sequence, Structure and Function, Atlanta, GA, November.

- “Computational analysis of high-throughput genomics data using hidden Markov models and support vector machines,” keynote lecture, Tenth Annual Genome Informatics Workshop, Tokyo, December.
- 1998 “Hidden Markov models for protein families,” Structure-Function Based Genomics Symposium, London, England, April.
- “Models and methods in biosequence analysis,” International Conference on Problems in Biophysics, Moscow, June.
- “Statistical methods in biosequence analysis: discriminative vs. generative models,” 2nd Annual Conference on Computational Genomics, Reston, VA, November.
- “A new discriminative methodology for detecting remote protein homologies,” First Annual Computational Structural Biology Research Meeting, Monterey, CA, December.
- 1997 “Design of the Genie genefinder,” Finding Genes: Computational Analysis of DNA Sequences, Cold Spring Harbor, NY, March.
- “A brief look at some machine learning problems in genomics,” Keynote Lecture, Tenth International Conference on Computational Learning Theory, Vanderbilt University, Nashville, TN, July.
- “Statistical genome analysis: hidden Markov methods,” Newton Seminar, Isaac Newton Institute, Cambridge, England, October.
- 1996 “Using hidden Markov models for biosequence analysis,” keynote lecture, Fourth International Conference on Intelligent Systems in Molecular Biology, St. Louis, MO, June.
- “Hidden Markov Models for Biosequence Analysis,” Workshop on Theoretical and Computational Biology, Lawrence Berkeley National Laboratory, CA, August.
- “VC dimension, covering numbers and worst-case prediction of individual sequences,” ICMS Workshop on the Vapnik-Chervonenkis Dimension, Edinburgh, Scotland, September.
- 1995 “Bounds on the mutual information between a parameter and a sequence of conditionally independent observations,” Workshop on Theory of Neural Networks, Pohang, Korea, February.
- “Hidden Markov models for proteins,” DIMACS Workshop on Sequence Based Methods for Protein Folding, Rutgers, NJ, March.
- “Hidden Markov models in biosequence analysis,” Workshop on Biosequence Analysis, Aspen Center for Physics, Aspen, CO, June.
- “Prediction, data compression and metric dimension,” 11th ACM Symposium on Computational Geometry, Vancouver, Canada, June.
- “Using hidden Markov models to search biosequence databases,” Workshop on Knowledge Discovery and Database Mining, Montreal, Canada, August.
- “Hidden Markov and related statistical models: how they have been applied to biosequence analysis,” Workshop on Uncertainty in Artificial Intelligence, Montreal, Canada, August.
- 1994 “Hidden Markov models for multiple alignments and database search for proteins,” Meeting on Critical Assessment of Techniques for Protein Structure Prediction, Asilomar, CA, December.

- “Using stochastic context-free grammars to fold, align and model homologous RNA sequences,” Workshop on the Fusion of Molecular Biology and Knowledge Information Processing, Tokyo, Japan, December.
- 1993 “Hidden Markov models and beyond,” *Macromolecules, Genes and Computers: Chapter Three*, Waterville, NH, August.
- “On-line prediction: models from computational learning theory and statistics,” Learning Days in Jerusalem Conference, Jerusalem, Israel, May.
- 1992 “How to use expert advice,” Sixth Conference on Neural Information Processing Systems, post-meeting workshop on Bayesian Methods, Vail, CO.
- “Hidden Markov models for protein families,” Sixth Conference on Neural Information Processing Systems, post-meeting workshop on hidden Markov Models, Vail, CO.
- “Hidden Markov models for protein families” and “Bayes Methods for Prediction,” Workshop on Supervised Learning, Santa Fe, NM.
- “How well do Bayes Methods work for on-line prediction of +1,-1 values?” Third NEC Symposium on Computation and Cognition, Princeton, NJ.
- 1991 “Uniting the VC and TLS theories of generalization,” Neural Networks for Computing, Snowbird, UT, April.
- “Computational learning theory,” Third Woodward Conference on Modeling Complex Phenomena, San Jose, CA, April.
- “How well do Bayes methods work?” Fifth Conference on Neural Information Processing Systems, post-meeting workshop on Bayesian methods, Vail, CO.
- “Information theory, VC dimension and the Bayesian approach to machine learning,” 2nd Workshop on Computational Learning Theory and Natural Learning Systems, Berkeley, CA, September.
- 1990 “Probably approximately correct learning,” National Conference on Artificial Intelligence (AAAI), Boston, MA, July.
- “Learnability and the metric dimension: decision theoretic generalizations of the PAC learning model,” First International Workshop on Algorithmic Theory, Tokyo, Japan, October.
- 1989 “Generalization in neural networks: a computational learning theory perspective,” Annual Meeting of the American Association for the Advancement of Science, San Francisco, CA, January.
- “Bounds on sufficient training set size when learning from random examples,” Neural Networks for Computing, Snowbird, UT, April.
- “Probably approximately correct learning,” *Discovery and Learning: Philosophical and Computational Perspectives*, Pittsburgh, PA, April.
- “Complexity issues in learning from random examples,” Annual Meeting of the Society for Industrial and Applied Mathematics, Minisymposium on Neural Computing, SIAM '89, San Diego, CA, July.
- 1988 “Theoretical results in machine learning,” 5th International Conference on Machine Learning, Ann Arbor, MI. June.

Selected Seminars

- 2005 “Using evolution to explore the human genome,” Computational and Genomic Biology Seminar Series, University of California, Berkeley. November.
- “Impact of human genome research: present and future,” Mervyn Young Memorial Lecture, University of Colorado, Boulder. April.
- “Using evolution to explore the human genome,” Distinguished Engineering Alumni Award Seminar, University of Colorado, Boulder. April.
- 2004 “Comparing the human, chimp, mouse and rat genomes: using evolution to predict functional elements in the human genome,” Chemistry and Genomics Seminar, Stanford University, Palo Alto, CA. March.
- “Comparing the human, chimp, mouse and rat genomes: using evolution to predict functional elements in the human genome,” Sali Labs, University of California, San Francisco. January.
- 2003 “Using comparative genomics to predict functional elements in the human genome,” Sloan-Swartz Center for Theoretical Neurobiology Seminar, Distinguished Speaker Series, Colorado State University. November.
- “Using comparative genomics to predict functional elements in the human genome,” 2003-04 Distinguished Seminar Series, Department of Computer Science, University of British Columbia. November.
- “Computational genomics on mammals: sequence, function, and evolution,” Intra- and Intercellular Communication Science Meeting, Howard Hughes Medical Institute, April.
- “Bioinformatics, genome evolution and the challenge of identifying functional elements in the human genome” Distinguished Science Seminar, Affymetrix. April.
- 2002 “Status of the human and mouse working draft genomes, and some early comparisons,” Genes and Genomes in Health and Disease, HHMI, Washington, DC. March.
- 2001 “Assembly and initial analysis of a working draft of the human genome,” New Frontiers in Structural and Computational Biology, Howard Hughes Medical Institute. March.
- 1998 “Using hidden Markov models for biosequence analysis: recent tests and new methods,” Understanding the Genome: Technological and Mathematical Challenges, Mathematical Sciences Research Institute, Berkeley, CA. May

Books Edited

Proceedings of the Fifth ACM Workshop on Computational Learning Theory, ACM, 1992.

Proceedings of the Second Workshop on Computational Learning Theory, Morgan Kaufmann, Los Altos, CA, 1989. (with R. Rivest and M. Warmuth)

Proceedings of the First Workshop on Computational Learning Theory, Morgan Kaufmann, Los Altos, CA, 1988. (with L. Pitt)

Journals Edited

2005- Associate Editor: Public Library of Science Computational Genomics

2001- Editorial Board: *Drug Discovery Today*

1996- Associate Editor: *Journal of Computational Biology*

- 1996-2002 Editorial Board: *Neural Computing Surveys*
- 1995-2002 Editorial Board: *Journal of Neurocomputing*
- 1993-95 Editorial Board: *Journal of Artificial Intelligence Research*
- 1988-97 Associate Editor: *Machine Learning*
- 1988-97 Guest editor: *Machine Learning*, special issue on recent theoretical directions in Machine Learning, Spring 1988

UNIVERSITY SERVICE

UC System-wide Service

- 2000 Scientific Co-Director: California Institute for Quantitative Biomedical Research (QB3)
- 1998 Member: UC System-Wide Life Science Informatics Working Group

Special UCSC Service

- 2005 Director, UCSC Training Program in the Systems Biology of Stem Cells
- 2000 Special Search Committee for the Dean of the School of Engineering
- 1999 Director of the UCSC Center for Biomolecular Science and Engineering
- 1997-98 Member, Chancellor's Millennium Committee