

Curtis Huttenhower

Assistant Professor of Computational Biology and Bioinformatics
Department of Biostatistics, School of Public Health, Harvard University
655 Huntington Avenue • Boston, MA 02115 • 617-432-4912 • chuttenh@hsph.harvard.edu

Academic Appointments

July 2009 - Present	<i>Department of Biostatistics, Harvard School of Public Health</i> Assistant Professor of Computational Biology and Bioinformatics
---------------------	--

Education

November 2008 - June 2009	<i>Lewis-Sigler Institute for Integrative Genomics, Princeton University</i> Supervisor: Dr. Olga Troyanskaya Postdoctoral Researcher
August 2004 - November 2008	<i>Computer Science Department, Princeton University</i> Adviser: Dr. Olga Troyanskaya Ph.D. in Computer Science, November 2008; M.A., June 2006
August 2002 - May 2004	<i>Language Technologies Institute, Carnegie Mellon University</i> Adviser: Dr. Eric Nyberg M.S. in Language Technologies, December 2003
August 1998 - November 2000	<i>Rose-Hulman Institute of Technology</i> B.S. summa cum laude, November 2000 Majored in Computer Science, Chemistry, and Math; Minored in Spanish
August 1996 - May 1998	<i>Simon's Rock College of Bard</i> A.A., May 1998

Awards, Honors, and Scholarships

- Quantitative and Computational Biology Program Training Fellowship (Princeton University, 2006-2008)
- Association of Princeton Graduate Alumni Teaching Award (Princeton University, 2006)
- Addison-Wesley Computer Science Award (Rose-Hulman Institute of Technology, 2000)
- William Albert Noyes, Sr. Award in Chemistry (Rose-Hulman Institute of Technology, 2000)
- EDS Computer Science Scholarship (Rose-Hulman Institute of Technology, 1998)
- Acceleration to Excellence Program full two-year scholarship (Simon's Rock College of Bard, 1996)

Academic Affiliations

- Systems Biology Program, Harvard Medical School (2011-Present)
- Broad Institute (2010-Present)
- NIH Human Microbiome Project: co-chair, Whole-Genome Shotgun Analysis and Metabolic Reconstruction; member, Steering and Publications (2010-Present)
- Harvard Microbial Sciences Initiative (2010-Present)
- Center for the Study of Inflammatory Bowel Diseases (2010-Present)
- Program in Biological Sciences in Public Health, Harvard School of Public Health (2009-Present)
- Dana-Farber/Harvard Cancer Center (2009-Present)

Teaching Experience

Genomic Data Manipulation (BIO508)

Harvard School of Public Health, Boston, MA

Spring 2010

- Introduction to tools and techniques needed to obtain, analyze, and interpret modern genome-scale data. Includes a brief overview of Python and statistical methods for high-dimensional data, geared toward biological investigators interpreting their own data or integrating it with results from public repositories.

Instructor

The Johns Hopkins University Center for Talented Youth, Baltimore, MD

June 2006 - August 2006, June 2005 - August 2005, June 2004 - August 2004, June 2000 - August 2000,

Supervisor: Dr. Ronald Bridwell

- Instructed gifted students, ages 12 through 16, in introductory computer science. Developed, prepared, and presented complete three-week curricula covering an introduction to the field of computer science and an introduction to the theory of computation; authored two short textbooks for the courses.

Mentor

Princeton Summer Undergraduate Research Experience through The Leadership Alliance, Princeton, NJ

June 2007 - August 2007, Supervisor: Dean David Redman

- Mentored undergraduate students individually and in groups to prepare them for careers in academia and education. Provided assistance in organizing the summer program and in preparing conference presentations, posters, and research papers.

Industry Experience

Software Design Engineer

Microsoft, Redmond, WA

January 2001 - August 2002, Supervisor: Dr. Douglas Potter

- Worked with the Natural Language Group Engine Team on the Microsoft Natural Language Development Platform. Developed spelling and grammar checking tools, updated and improved language detection software, and implemented a novel morphological processing environment.

Additional Academic Experience

Editorial Boards

BMC Bioinformatics, Associate Editor

Seminar Series Director

Microbiome Research in the Boston Area (MiRiBA), Boston, MA

Princeton Program in Integrative Information, Computer and Application Sciences (PICASso), Princeton, NJ

Academic Society Memberships

Life Sciences Society, July 2009 - Present

International Society for Computational Biology, June 2007 - Present

Genetics Society of America, June 2007 - Present

Referee

Nature, Nature Biotechnology, Nature Methods, Molecular Systems Biology, Genome Research, Genome Biology, Nucleic Acids Research, PLoS Computational Biology, Bioinformatics, Journal of Bioinformatics and Computational Biology, BMC Genomics, BMC Bioinformatics, Genomics

Publications

1. Ballal SA, Gallini CA, Segata N, **Huttenhower C**, Garrett WS. "Host and gut microbiota symbiotic factors: lessons from inflammatory bowel disease and successful symbionts." *Cellular Microbiol* 2011, 13(4):508-17
2. Morikawa T, Baba Y, Yamauchi M, Kuchiba A, Nosho K, Shima K, Tanaka N, **Huttenhower C**, Frank DA, Fuchs CS, Ogino S. "STAT3 Expression, Molecular Features, Inflammation Patterns, and Prognosis in a Database of 724 Colorectal Cancers." *Clin Cancer Res* 2011, 17(6):1452-62
3. Wardwell LH, **Huttenhower C**, Garrett WS. "Current concepts of the intestinal microbiota and the pathogenesis of infection." *Curr Infect Dis Rep* 2011, 13(1):28-34
4. Park CY, Hess DC, **Huttenhower C***, Troyanskaya OG*. "Simultaneous Genome-Wide Inference of Physical, Genetic, Regulatory, and Functional Pathway Components." *PLoS Computational Biology* 2010, 6(11)
5. Pop A*, **Huttenhower C***, Iyer-Pascuzzi A, Benfey PN, Troyanskaya OG. "Integrated functional networks of process, tissue, and developmental stage specific interactions in *Arabidopsis thaliana*." *BMC Systems Biology* 2010, 4:180
6. Sathirapongsasuti JF, Sathira N, Suzuki Y, **Huttenhower C**, Sugano S. "Ultraconserved cDNA segments in the human transcriptome exhibit resistance to folding and implicate function in translation and alternative splicing." *Nucleic Acids Research* 2011, 39(6):1967-79
7. Tanaka N, **Huttenhower C**, Nosho K, Baba Y, Shima K, Quackenbush J, Haigis KM, Giovannucci EL, Fuchs CS, Ogino S. "Novel Application of Structural Equation Modeling to Correlation Structure Analysis of CpG Island Methylation in Colorectal Cancer." *American Journal of Pathology* 2010, 177(6):2731-40
8. Baba Y, Nosho K, Shima K, **Huttenhower C**, Tanaka N, Hazra A, Giovannucci EL, Fuchs CS, Ogino S. "Hypomethylation of the IGF2 DMR in Colorectal Tumors, Detected by Bisulfite Pyrosequencing, is Associated with Poor Prognosis." *Gastroenterology* 2010, 139(6):1855-64
9. **Huttenhower C**, Hofmann O. "A quick guide to large scale genomic data mining." *PLoS Computational Biology* 2010, 6(5)
10. Baba Y, **Huttenhower C**, Nosho K, Tanaka N, Shima K, Hazra A, Schernhammer ES, Hunter DJ, Giovannucci EL, Fuchs CS, Ogino S. "Epigenomic diversity of colorectal cancer indicated by LINE-1 methylation in a database of 869 tumors." *Molecular Cancer* 2010, 9:125
11. **Huttenhower C***, Mehmood SO*, Troyanskaya OG. "Graphlet: Interactive exploration of large, dense graphs." *BMC Bioinformatics* 2009, 10(417)
12. **Huttenhower C***, Mutungu KT*, Indik N, Yang W, Schroeder M, Forman JJ, Troyanskaya OG[†], Collier HA[†]. "Detailing regulatory networks through large scale data integration." *Bioinformatics* 2009, 25(24):3267-74
13. **Huttenhower C***, Hibbs MA*, Myers CL*, Caudy AA, Hess DC, Troyanskaya OG. "The impact of incomplete knowledge on evaluation: an experimental benchmark for protein function prediction." *Bioinformatics* 2009, 25(18):2404-10
14. Chikina MD, **Huttenhower C**, Troyanskaya OG[†], Murphy CT[†]. "Global prediction of tissue-specific gene expression and context-dependent gene networks in *Caenorhabditis elegans*." *PLoS Computational Biology* 2009, 5(6)
15. **Huttenhower C***, Haley EM*, Hibbs MA, Dumeaux V, Barrett DR, Collier HA[†], Troyanskaya OG[†]. "Exploring the human genome with functional maps." *Genome Research* 2009, 19(6):1093-106

16. Hess DC, Myers CL*, **Huttenhower C***, Hibbs MA*, Hayes AP, Paw J, Clore JJ, Mendoza RM, San Luis B, Nislow C, Giaever G, Costanzo M, Troyanskaya OG[†], Caudy AA[†]. "Computationally driven, quantitative experiments discover genes required for mitochondrial biogenesis," *PLoS Genetics* 2009, 5(3)
17. Hibbs MA*, Myers CL*, **Huttenhower C***, Hess DC, Li K, Caudy AA, Troyanskaya OG. "Directing experimental biology: a case study in mitochondrial inheritance," *PLoS Computational Biology* 2009, 5(3)
18. Airoidi EM*, **Huttenhower C***, Gresham D, Lu C, Broach JR, Botstein D[†], Troyanskaya O[†], "Predicting the rate of cellular growth from gene expression states," *PLoS Computational Biology* 2009, 5(1)
19. **Huttenhower C**, Troyanskaya OG. "Assessing the functional structure of genomic data," *Bioinformatics* 2008, 24(13):i330-8
20. **Huttenhower C**, Schroeder M, Chikina MD, Troyanskaya OG. "The Sleipnir library for computational functional genomics," *Bioinformatics* 2008, 24(13):1559-61
21. Brauer MJ, **Huttenhower C***, Airoidi EM*, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. "Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast," *Molecular Biology of the Cell* 2008, 19(1):352-67
22. Hibbs MA, Hess DC, Myers CL, **Huttenhower C**, Li K, Troyanskaya OG, "Exploring the functional landscape of gene expression: directed search of large microarray compendia," *Bioinformatics* 2007, 23(20):2692-9
23. **Huttenhower C**, Flamholz AI, Landis JN, Sahi S, Myers CL, Hibbs MA, Siemens NO, Troyanskaya OG, Collier HA. "Nearest Neighbor Networks: clustering expression data based on gene neighborhoods." *BMC Bioinformatics* 2007, 8(250)
24. Chi A*, **Huttenhower C***, Geer LY, Coon JJ, Syka JE, Bai DL, Shabinowitz J, Burke DJ, Troyanskaya OG, Hunt DF. "Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry." *PNAS* 2007, 104(7):2193-8
25. Sealfon RS, Hibbs MA, **Huttenhower C**, Myers CL, Troyanskaya OG. "GOLEM: an interactive graph-based gene-ontology navigation and analysis tool," *BMC Bioinformatics* 2006, 7(443)
26. **Huttenhower C**, Hibbs MA, Myers CL, Troyanskaya OG. "A scalable method for integration and functional analysis of multiple microarray data sets," *Bioinformatics* 2006, 22(23):2890-7
27. Myers CL, Barrett DR, Hibbs MA, **Huttenhower C**, Troyanskaya OG. "Finding function: evaluation methods for functional genomic data," *BMC Genomics* 2006, 7(1)
28. **Huttenhower C**, Troyanskaya OG. "Bayesian data integration: a functional perspective," *Proceedings of the Computational Systems Bioinformatics Conference* 2006, 341-51
29. Nyberg E, Mitamura T, Callan J, Carbonell J, Frederking R, Collins-Thompson K, Hiyakumoto L, Huang Y, **Huttenhower C**, Judy S, Ko J, Kupse A, Lita L, Pedro V, Svoboda D, Van Durme B. "The JAVELIN question-answering system at TREC 2003: a multi-strategy approach with dynamic planning." 2003
30. **Huttenhower C**, Kinley A. "Development of an intelligent system for organic compound analysis." *Proceedings of the Eleventh Midwest Artificial Intelligence and Cognitive Science Conference* 2000, 85-92
31. **Huttenhower C**, Sherman G. "Extending cwatsets to higher dimensions: an examination of the statistical properties of cwatsets with dimension greater than two." *Spring meeting of the Indiana Section of the Mathematical Association of America* 2000

Book Chapters, Reviews, and Technical Reports

- Waldron L, Collier HA, **Huttenhower C**. "Integrative Approaches for Microarray Data Analysis." *Methods in Molecular Biology*, in press
- Livstone MS, Oughtred R, Heinicke S, Vernet B, **Huttenhower C**, Durand D, Dolinski K. "Inferring protein function from homology using the Princeton Protein Orthology Database (P-POD)." *Curr Protoc Bioinformatics* 2011, 6:6.11
- **Huttenhower C**, Myers CL, Hibbs MA, Troyanskaya OG. "Computational analysis of the yeast proteome: understanding and exploiting functional specificity in genomic data." *Methods in Molecular Biology* 2009, 548:273-93
- **Huttenhower C**, Troyanskaya OG. "Analysis of large genomic data collections." Princeton University Doctoral Thesis 2008
- **Huttenhower C**, Nyberg E. "FLOOD: a planning framework for reasoning with linguistic data," Carnegie Mellon University Language Technologies Institute Master's Thesis 2003

Invited Presentations

- "Large scale genomic data integration for functional metagenomics," Keynote address at the Student Council Symposium of the 19th Annual International Conference on Intelligent Systems for Molecular Biology. Vienna, Austria, 2011
- "Metagenomic biomarker discovery and the human microbiome," Microbial Systems (and Beyond) Seminar @ Parsons Laboratory, MIT. Boston, MA, 2011 (presented by Nicola Segata)
- "Computational Metagenomics and the Human Microbiome", DFCI Center for Cancer Computational Biology seminar series. Boston, MA, 2011
- "Metabolic Reconstruction in Microbial Communities," Beyond Sequencing conference. San Francisco, CA, 2011
- "Understanding the Human Microbiome through Data Integration," Bio-IT World Conference and Expo. Boston, MA, 2011
- "Large scale genomic data integration for functional metagenomics," University of Maryland, Baltimore County. Baltimore, MD, 2010
- "Linking Microbiome Pathways to Disease in MetaHIT and the HMP," Center for the Study of Inflammatory Bowel Disease workshop. Boston, MA, 2010.
- "Understanding gene function in the human microbiome," *Frontiers in Mucosal Immunology* symposium. Boston, MA, 2010.
- "Large scale genomic data integration for functional metagenomics," Indiana University School of Informatics seminar. Bloomington, IN, 2010
- "Large scale genomic data integration for functional metagenomics," Johns Hopkins University Center for Computational Genomics seminar. Baltimore, MD, 2010
- "Supervised and unsupervised methods for large scale genomic data integration," Boston University Systems Biology seminar. Boston, MA, 2010
- "Personalized Medicine, Bioinformatics, and Biotechnology," Saint Francis University Science Day. Loretto, PA, 2009
- "Analyzing large genomic data collections," Georgia Tech - Oak Ridge National Lab International Conference on Bioinformatics. Atlanta, GA, 2009
- "Answering biological questions using large genomic data collections," Mount Sinai School of Medicine Systems Biology seminar. New York, NY, 2009
- "Microarray analysis," Cold Spring Harbor Laboratory Programming for Biology course. Cold Spring Harbor, NY, 2008

- "Integration and functional analysis of microarray datasets," Princeton Program in Integrative Information, Computer and Application Sciences (PICASs). Princeton, NJ, 2006

Presentations

- "Scalable metabolic reconstruction for metagenomic data and the human microbiome," 19th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Vienna, Austria, 2011
- "Metagenomic Biomarker Discovery and Explanation," 19th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Vienna, Austria, 2011 (presented by Nicola Segata)
- "Metabolic Reconstruction for Metagenomic Data and the Human Microbiome," Keystone Symposium on Microbial Communities. Breckenridge, CO, 2011
- "Metabolic Reconstruction in Microbial Communities," International Human Microbiome Congress. Vancouver, Canada, 2011
- "Using all the data: Large scale biological data mining for functional genomics and metagenomics," 18th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Boston, MA, 2010
- "Computational methodology for microbial and metagenomic characterization using large scale functional genomic data integration," Keystone Symposium on Biomolecular Interaction Networks: Function and Disease. Quebec, Canada, 2010
- "Exploring the human genome with functional maps," 17th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Stockholm, Sweden, 2009
- "Detailing regulatory networks through large scale data integration," 5th Annual RECOMB Satellite on Regulatory Genomics. Boston, MA, 2008
- "Assessing the functional structure of genomic data," 16th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Toronto, Canada, 2008
- "Growth-specific programs of gene expression," 16th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Toronto, Canada, 2008
- "*S. cerevisiae* mitochondria: validating predictions from microarray integration," 15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Vienna, Austria, 2007
- "Bayesian data integration: a functional perspective," Computational Systems Bioinformatics (CSB) Conference. Stanford, CA, 2006

Posters

- "Scalable metabolic reconstruction for metagenomic data and the human microbiome," 19th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Vienna, Austria, 2011
- "Metagenomic Biomarker Discovery and Explanation," 19th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Vienna, Austria, 2011 (presented by Nicola Segata)
- "A sequence-based method identifying core genes for evolutionary and functional microbial phylogenies," 19th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Vienna, Austria, 2011 (presented by Nicola Segata)
- "Optimized application of penalized regression methods to diverse genomic data," 19th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Vienna, Austria, 2011 (presented by Levi Waldron)
- "Whole-genome transcriptome profiling of archival tissue blocks by DASL," Keystone Symposium on the Cancer Genome. Boston, MA, 2011 (presented by Levi Waldron)
- "Metabolic reconstruction for metagenomic data and the human microbiome," Keystone Symposium on Microbial Communities. Breckenridge, CO, 2011

- "Scalable metabolic reconstruction for metagenomic data and the human microbiome," International Human Microbiome Congress. Vancouver, Canada, 2011
- "Metagenomic Biomarker Discovery and Explanation," International Human Microbiome Congress. Vancouver, Canada, 2011 (presented by Nicola Segata)
- "Robust biomarker development through integrative analysis of gene expression arrays," Program in Quantitative Genomics symposium. Boston, MA, 2010.
- "Computational methodology for microbial and metagenomic characterization using large scale functional genomic data integration," Keystone Symposium on Biomolecular Interaction Networks: Function and Disease. Quebec, Canada, 2010
- "Biological network integration and mining for microbial community analysis," Neural Information Processing Systems (NIPS). Vancouver, Canada, 2009
- "Orthology-based functional transfer in microbial communities," RECOMB Systems Biology Satellite. Boston, MA, 2009
- "HEFalMp: Integrating 30,000 experimental conditions to predict systems-level relationships in *H. sapiens*," 17th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Stockholm, Sweden, 2009
- "Assessing the functional structure of genomic data," 16th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Toronto, Canada, 2008
- "Predicting growth rate from gene expression signatures," Yeast Genetics and Molecular Biology Meeting. Toronto, Canada, 2008
- "*S. cerevisiae* Mitochondria: Validating Predictions from Microarray Integration," 15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). Vienna, Austria, 2007
- "A General Methodology for Integration of Microarray Data," Computational Systems Bioinformatics (CSB) Conference. Stanford, CA, 2005

Software

- *HUMANn* <http://huttenhower.sph.harvard.edu/humann>
Efficient metabolic and functional gene and pathway reconstruction from metagenomic data
- *LEfSe* <http://huttenhower.sph.harvard.edu/lefse>
Biomarker detection for metagenomic data incorporating statistical significance and biological consistency
- *HEFalMp* <http://function.princeton.edu/hefalmp>
Functional maps of the human genome generated by integrating ~30,000 experimental results
- *Sleipnir* <http://huttenhower.sph.harvard.edu/sleipnir>
Open-source C/C++ library for large scale computational functional genomics
- COALESCE <http://function.princeton.edu/coalesce>
Regulatory module inference and scalable expression data biclustering
- Graphle <http://function.princeton.edu/graphle>
Open-source Java applet and server for interactive exploration of large biological network compendia
- *Growth Rate* <http://function.princeton.edu/growthrate>
Tools for predicting a cellular culture's growth rate based on gene expression data
- *MEFIT* <http://function.princeton.edu/mefit>
Web-based interface to integration results from ~200 *S. cerevisiae* expression conditions
- *Nearest Neighbor Networks* <http://function.princeton.edu/nnn>
Open-source Java software for clustering based on reciprocal nearest neighborhoods

Patents

- 7490034, Lexicon with sectionalized data and method of using the same
- 7617089, Method and apparatus for compiling two-level morphology rules
- Application 20040193399, System and method for word analysis
- Application 20040148170, Statistical classifiers for spoken language understanding and command/control scenarios