

XIN HE

Lane Center for Computational Biology
School of Computer Science
Carnegie Mellon University
Contact: (901) 336-6524, xinhe2@gmail.com
<http://www.cs.cmu.edu/~xinhe2/>

RESEARCH INTERESTS

- Statistical methods for genetic mapping of complex human diseases.
- Systems genetics: individual variation of molecular and epigenetic traits.
- Gene regulation: modeling and computational inference of gene regulatory networks.
- Evolutionary and comparative analysis of genomic data.

EDUCATION

- Oct 2009 **University of Illinois at Urbana-Champaign**
Ph.D. in Computer Science (Advisor: Prof. Saurabh Sinha)
Thesis: Computational models of function and evolution of *cis*-regulatory sequences
- 2003-2004 **University of North Carolina at Chapel Hill**
Ph.D. student in Program of Bioinformatics and Computational Biology
- Dec 2002 **Loyola University Chicago**
M.S. in Computer Science
- 1999-2001 **Northwestern University, School of Medicine**
Ph.D. student in Integrated Graduate Program in Life Sciences
- Jul 1999 **University of Science and Technology of China**
B.S. in Biochemistry (Special Class for Gifted Young)

PROFESSIONAL POSITIONS

Lane Fellow of Computational Biology Lane Center for Computational Biology, School of Computer Science, Carnegie Mellon University, (2011-present)
Work with Prof. Kathryn Roeder on statistic genetics and Prof. Ziv Bar-Joseph on gene regulation, network analysis and methods for next-generation sequencing.

Postdoctoral Researcher Dept. of Biochemistry and Biophysics, University of California at San Francisco, (2009-2011)
Work with Prof. Hao Li on evolution of gene regulatory networks and statistic genetics.

Research Assistant Dept. of Computer Science and Institute of Genomic Biology, University of Illinois at Urbana Champaign, (2005-2009)

Work with Drs. Bruce Schatz and Chengxiang Zhai on the Beespace literature-mining system.

Teaching Assistant Dept. of Biology, University of North Carolina at Chapel Hill, (2004)
Grading and discussion sessions for the course “Cellular and Developmental Biology”.

AWARDS

- Outstanding Postdoctoral Research Award, Association of Chinese Geneticists in America (ACGA), 2013
- Lane Fellowship of Computational Biology, Carnegie Mellon University, 2011
- Bioinformatics and Computational Biology Fellowship, University of North Carolina at Chapel Hill, 2003

PUBLICATIONS

Journal Articles under Review or in Preparation

- [1] **Xin He**, Li Liu, Bernie Devlin and Kathryn Roeder, Incorporating prior information of genetic variants improves gene association test. In preparation.
- [2] Saket Navlakha, **Xin He**, Christos Faloutsos and Ziv Bar-Joseph, Topological properties of robust networks. Under review.
- [3] Li Liu, Jing Lei, Stephan J Sanders, Jeremy A Willsey, Yan Kou, Ercument A Cicek, Lambertus Klei, Cong Lu, **Xin He**, Mingfeng Li, Rebecca A. Muhle, Avi Ma’ayan, James p. Noonan, Nenad Sestan, Kathryn McFadden, Matthew W. State, Joseph D. Buxbaum, Bernie Devlin and Kathryn Roeder, DAWN: A framework to identify autism genes and subnetworks using gene expression and genetics. Under review.

Journal Articles

- [4] Jeremy Willsey, Stephan J. Sanders, Mingfeng Li, Shan Dong, Andrew T. Tebbenkamp, Rebecca A. Muhle, Steven K. Reilly, Leon Lin, Sofia Fertuzinhos, Jeremy A. Miller, Michael T. Murtha, Candace Bichsel, Wei Niu¹, Justin Cotney, Gulhan Ercan-Sencicek, Jake Gockley, Abha Gupta, Wenqi Han, **Xin He**, Ellen Hoffman, Lambertus Klei, Jing Lei, Wenzhong Liu, Li Liu, Cong Lu, Xuming Xu, Ying Zhu, Shrikant M. Mane, Ed Lein, Liping Wei, James P. Noonan, Kathryn Roeder, Bernie Devlin, Nenad Sestan, Matthew W. State, Co-expression networks implicate human mid-fetal deep cortical projection neurons in the pathogenesis of autism. *Cell*, 2013 Nov 21;155(5):997-1007
- [5] Shan Zhong*, **Xin He*** and Ziv Bar-Joseph, Predicting tissue specific transcription factor binding sites. *BMC Genomics*, 2013 Nov 15;14(1):796

- [6] **Xin He**, Stephan J. Sanders, Li Liu, Silvia De Rubeis, Elaine T. Lim, James S. Sutcliffe, Gerard D. Schellenberg, Richard A. Gibbs, Mark J. Daly, Joseph D. Buxbaum Matthew W. State, Bernie Devlin, Kathryn Roeder, Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genetics*, 2013 Aug;9(8)
- Reported by SFARI News with interviews, “New model merges data streams to boost gene discovery” (SFARI - Simons Foundation Autism Research Initiative, is a major sponsor of autism research)
 - Recommended by Faculty of 1000.
 - Platform talk in American Society of Human Genetics Meeting 2013
- [7] **Xin He**, Chris K. Fuller, Yi Song, Qingying Meng, Bin Zhang, Xia Yang, Hao Li, Sherlock: Detecting Gene-Disease Associations by Matching Patterns of Expression QTL and GWAS. *Am J Hum Genet*, 2013 May 2;92(5):667-80
- Featured by *Am J Hum Genet* (May, 2013), “Sherlock Makes Identifying Gene-Disease Associations Elementary!”.
 - Highlighted in *Genetics* (May, 2013), “This Month in American Journal of Human Genetics”.
- [8] **Xin He***, Thyago S.P.C. Duque* and Saurabh Sinha, Evolutionary origins of transcription factor binding site clusters. *Mol Biol Evol*, 2012, 29(3):1059-70
- [9] Dan Xie*, Chieh-Chun Chen*, **Xin He**, Xiaoyi Cao and Sheng Zhong, Towards an evolutionary model of transcription networks. *PLoS Comput Biol*, 2011, 7(6)
- [10] Moushumi Sen Sarma, David Arcoleo, Radhika S Khetani, Brant Chee, Xu Ling, **Xin He**, Jing Jiang, Qiaozhu Mei, ChengXiang Zhai and Bruce Schatz, BeeSpace Navigator: exploratory analysis of gene function using semantic indexing of biological literature. *Nucleic Acids Res*, 2011, W462-9
- [11] **Xin He**, Md. Abul Hassan Samee, Charles Blatti and Saurabh Sinha, Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative Binding and short-range repression. *PLoS Comput Biol*, 2010, 6(9)
- [12] **Xin He**, Yanen Li, Radhika Khetani, Barry Sanders, Yue Lu, Xu Ling, ChengXiang Zhai and Bruce Schatz, BSQA: integrated text mining using entity relation semantics extracted from biological literature of insects. *Nucleic Acids Res*, 2010, 38, W175-81
- [13] **Xin He**, Moushumi Sen Sarma, Xu Ling, Brant Chee, Chengxiang Zhai and Bruce Schatz, Identifying overrepresented concepts in gene lists from literature: a statistical approach based on Poisson mixture model. *BMC Bioinformatics* 2010, 11:272
- [14] **Xin He** and Saurabh Sinha, ChIPs and Regulatory Bits. *Nature Biotechnology*, 2010, 28, 142-143
- [15] **Xin He**, Chieh-Chun Chen, Feng Hong, Fang Fang, Saurabh Sinha, Huck-Hui Ng and Sheng Zhong, A biophysical model for analysis of transcription factor interaction and binding site arrangement from genome-wide binding data. *PLoS ONE* 2009 4(12): e8155
- [16] Xu Ling, **Xin He** and Dong Xin, Detecting gene clusters under evolutionary constraint in a large number of genomes. *Bioinformatics*, 2009, 25(5):571-577

- [17] **Xin He**, Xu Ling and Saurabh Sinha, Alignment and prediction of regulatory sequences based on a probabilistic model of evolution. *PLoS Comput Biol*, 2009, 5(3):e1000299
- Selected by Genome Web Daily Scan “What’s Worth Reading on the Web” (March 16, 2009)
- [18] Jaebum Kim*, **Xin He*** and Saurabh Sinha, Evolution of regulatory sequences in 12 *Drosophila* species. *PLoS Genetics* 2009 Jan;5(1):e1000330. (joint first author)
- [19] Xu Ling, **Xin He**, Dong Xin and Jiawei Han, Efficiently identifying max-gap clusters in pairwise genome comparison. *J Comput Biol* 2008 Jul-Aug;15(6):593-609.
- [20] Saurabh Sinha and **Xin He**, MORPH: probabilistic alignment combined with hidden Markov models of cis-regulatory modules. *PLoS Comput Biol* 2007 Nov;3(11):e216
- [21] Long Li, Qianqian Zhu, **Xin He**, Saurabh Sinha and Marc S Halfon, Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. *Genome Biol*, 2007;8(6):R101 (Highly accessed)
- [22] Yue Lu, **Xin He** and Sheng Zhong, Cross-species microarray analysis with the OSCAR system suggests an INSR-Pax6-NQO1 neuro-protective pathway in aging and Alzheimer’s disease. *Nucleic Acid Research* 2007 Jul;35:W105-14
- [23] Xu Ling, Jing Jiang, **Xin He**, Qiaozhu Mei, Chengxiang Zhai and Bruce Schatz, Generating gene summaries from biomedical literature: a study of semi-structured summarization. *Information Processing and Management*, 43, 2007, 1777-1791
- An earlier version accepted by Pacific Symposium on Biocomputing
- [24] **Xin He** and Michael Goldwasser, Identifying conserved gene clusters in the presence of homology families. *J Comput Biol*, 2005 Jul-Aug;12(6):638-56
- Accepted by Research in Computational Molecular Biology (acceptance rate 17%)

Conference Papers

- [25] Yue Lu, Jing Jiang, Xu Ling, **Xin He** and ChengXiang Zhai, Language models for genomics information retrieval: UIUC at TREC 2007 Genomics Track. Proceedings of the 16th Text REtrieval Conference (TREC’07)
- [26] Xu Ling, Jing Jiang, **Xin He**, Qiaozhu Mei, Chengxiang Zhai and Bruce Schatz, Automatically generating gene summaries from biomedical literature Proceedings of Pacific Symposium on Biocomputing (PSB), 2006, pages 40-51
- [27] Jing Jiang, **Xin He** and ChengXiang Zhai, Robust pseudo feedback estimation and HMM passage extraction: UIUC at TREC 2006 Genomics Track. Proceedings of the 15th Text REtrieval Conference (TREC’06)
- [28] **Xin He** and Michael Goldwasser, Identifying conserved gene clusters in the presence of orthologous groups Proceedings of the Eighth Annual International Conferences on Research in Computational Molecular Biology (RECOMB), 2004, pp. 272-280

Book Chapters

- [29] **Xin He** and Saurabh Sinha, Evolution of cis-regulatory sequences in *Drosophila*. Computational Biology of Transcription Factor Binding: Methods Mol Biol. Ed: Istvan Ladunga. Humana Press, 2010

INVITED TALKS AND CONFERENCE PRESENTATIONS

- “Integrated methods improve detection of risk genes and discover hidden associations for complex diseases”, Division of Psychiatric Genomics, Mount Sinai School of Medicine, November, 2013
- “Integrated methods improve detection of risk genes and discover hidden associations for complex diseases”, Department of Biostatistics, Emory University, November, 2013
- “Integrated model of multiple types of rare variants and prior information improves the power of detecting risk genes for autism”, ASHG Annual Meeting, Boston, October, 2013
- “Combining Transcription Factor Binding Site Clustering and Evolutionary Conservation for Predicting cis-Regulatory Modules”, RECOMB Regulatory Genomics, Boston, November, 2008
- “Automatically generating gene summaries from biomedical literature”, Pacific Symposium on Biocomputing, Maui, Hawaii, January, 2006
- “Identifying conserved gene clusters in the presence of orthologous groups”, Department of Computer Science, University of Illinois at Urbana-Champaign, January, 2005