

CURRICULUM VITAE

HONGKAI JI

PERSONAL DATA

Mailing Address: Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street, Room E3638
Baltimore, MD 21205

Phone: 410-955-3517
Fax: 410-955-0958
E-Mail: hji@jhsph.edu
Web Page: <http://www.biostat.jhsph.edu/~hji>

EDUCATION AND TRAINING

Ph.D.	2007	Harvard University, Cambridge, MA, USA	Statistics
	2004-2007	Visiting Student, Stanford University, Stanford, CA, USA	
M.A.	2004	Harvard University, Cambridge, MA, USA	Statistics
M.E.	2002	Tsinghua University, Beijing, P.R. China	Pattern Recognition
B.E.	1999	Tsinghua University, Beijing, P.R. China	Automation

PROFESSIONAL EXPERIENCE

09/2013 – present	Associate Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
09/2007 – 08/2013	Assistant Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
12/2013 – present	Affiliated Faculty, The Institute for Data Intensive Engineering and Science, Johns Hopkins University
03/2012 – present	Affiliated Faculty, Center for Computational Biology, Johns Hopkins University
01/2011 – present	Affiliated Faculty, High Throughput Biology HiT Center, Johns Hopkins School of Medicine
04/2011 – 07/2011	Visiting Scholar, Department of Statistics, University of California, Berkeley, CA, USA
07/2009	Visiting Scholar, Beijing Institute of Genomics Chinese Academy of Sciences, Beijing, China

PROFESSIONAL ACTIVITIES

Society Membership

American Statistical Association (ASA)
Institute of Mathematical Statistics (IMS)
International Chinese Statistical Association (ICSA)
International Society for Computational Biology (ISCB)

Program or Project Development

Session Organizer, ENAR 2009
Session Organizer, ICSA 2010
Session Chair, ENAR 2010
Session Organizer and Session Chair, ENAR 2011
Organizing Committee, Statistical Methods for Very Large Datasets Conference 2011
Program Director for the Johns Hopkins – Nanjing University Exchange Program in
Statistical and Data Sciences 2011

EDITORIAL ACTIVITIES

Editorial Board Membership

Associate Editor, Statistical Applications in Genetics and Molecular Biology, 2011-
present

Peer Review Activities

Annals of Applied Statistics (AOAS)
Bioinformatics
Biometrics
Biostatistics
BMC Bioinformatics
BMC Genomics
Cell Research
Cell Stem Cell
Computational Statistics
Computational Statistics and Data Analysis (CSDA)
Genome Biology
Genome Research
Genomics
IEEE/ACM Transactions on Computational Biology and Bioinformatics
ISMB
Journal of Computational Biology (JCB)
Journal of the American Statistical Association (JASA)
Journal of the Royal Statistical Society, Series B (JRSSB)

Nature Biotechnology
Nucleic Acids Research (NAR)
Pacific Symposium on Biocomputing (PSB)
PLoS Computational Biology
Proceedings of the National Academy of Sciences of the USA (PNAS)
Statistica Sinica
Statistical Analysis and Data Mining (SADM)
Statistical Applications in Genetics and Molecular Biology (SAGMB)
Statistics in Biosciences (SIB)
Wiley Interdisciplinary Reviews: Systems Biology and Medicine

Book Review

Numerical Analysis for Statisticians (Second Edition), Springer

Review of Proposals

BBSRC Grant Review, UK, August 2007.
NIH Challenge Grant Review, US, May, 2010.
NIH Study Section, ZHG1 HGR-M (J2) Special Emphasis Panel, US, Nov 29-30, 2011.

HONORS AND AWARDS

Honors

2011 Excellence in Teaching, Johns Hopkins Bloomberg School of Public Health
2010 Excellence in Teaching, Johns Hopkins Bloomberg School of Public Health
2004 Certificate of Distinction in Teaching, Harvard University
2002 Outstanding Master's Thesis, Tsinghua University
2000 Outstanding Graduate of Tsinghua University
1999 First Place of "Challenge Cup" Student Extra-Curricular Science and Technology Project Competition, Tsinghua University
1999 Third Place of NHK Robocon' 99 (Robosoccer) International Robotics Design Contest for Students, Japan

Awards

2007 Faculty Innovation Award, Johns Hopkins Bloomberg School of Public Health
2005 Chinese Medical Science and Technology Award (No. 20050319P0802), Chinese Medical Association
2005 Beijing Science and Technology Award (No. 2005Med-2-024-02), Beijing Municipal People's Government, China
2002-2004 Student Fellowship, Harvard University
1995-2002 Tsinghua University Excellent Student Scholarship

Named Lectureships

06/03/2010 Constance van Eeden Distinguished Speaker Seminar, Department of Statistics, University of British Columbia, Vancouver, BC, Canada

PUBLICATIONS

Journal Articles (Peer Reviewed)

(* corresponding author; § equal contribution; **bold**: lab members or advisees)

1. **Ji HK**, Zhou Q, Wen F and Ji L (2001) Research on multiple alignments for alternative splicing. *Journal of Tsinghua University (Science and Technology)*, 41(9): 111-114.
2. **Ji HK**§, Zhou Q§, Wen F, Xia HY, Lu X and Li YD (2001) AsMamDB: an alternative splice database of mammals. *Nucleic Acids Research*, 29: 260-263.
3. **Chinese Schizophrenia Consortium** (2004) Association study of an SNP combination pattern in the dopaminergic pathway in paranoid schizophrenia: a novel strategy for complex disorders. *Molecular Psychiatry*, 9: 510-521. (**Joint First Author with Xu Q, Yuan YB and Yu YQ**)

Highlighted in News & Commentary: Irizarry KJL and Galbraith SJ, *Mol Psychiatry* 9: 431-432 (2004).

4. Hellmann I, Prüfer K, **Ji HK**, Zody MC, Pääbo S and Ptak SE (2005) Why do human diversity levels vary at a megabase scale? *Genome Research*, 15: 1222-1231.
5. **The Chimpanzee Sequencing and Analysis Consortium** (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437: 69-87. (Listed among 67 authors).

Comment in News and Views: Li WH and Saunders MA, *Nature*, 437: 50-51 (2005).

Google Scholar Citations: 1300+

6. **Ji HK** and Wong WH (2005) TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics*, 21: 3629-3636.

Google Scholar Citations: 200+

7. Kim RS, **Ji HK** and Wong WH (2006) An improved distance measure between the expression profiles linking co-expression and co-regulation in mouse. *BMC Bioinformatics*, 7:44. doi:10.1186/1471-2105-7-44.

8. **Ji HK** and Wong WH (2006) Computational biology: toward deciphering gene regulatory information in mammalian genomes. *Biometrics*, 62: 645-663.
9. **Ji HK**, Vokes SA and Wong WH (2006) A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. *Nucleic Acids Research*, 34: e146. doi:10.1093/nar/gkl803.
10. Paik JH, Kollipara R, Chu G, **Ji HK**, Xiao Y, Ding Z, Miao L, Tothova Z, Horner JW, Carrasco DR, Jiang S, Gilliland DG, Chin L, Wong WH, Castrillon DH and DePinho RA (2007) FoxOs are lineage-restricted redundant tumor suppressors and regulate endothelial cell homeostasis. *Cell*. 128 (2): 309-323.

Google Scholar Citations: 410+

11. Vokes SA, **Ji HK**, McCuine S, Tenzen T, Giles S, Zhong S, Longabaugh WJ, Davidson EH, Wong WH and McMahon AP (2007) Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. *Development*, 134: 1977-1989.

Google Scholar Citations: 90+

12. Vokes SA, **Ji HK**, Wong WH and McMahon AP (2008) A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog mediated patterning of the mammalian limb. *Genes & Development*. 22: 2651-2663.
13. **Ji HK**, Jiang H, Ma W, Johnson DS, Myers RM and Wong WH (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nature Biotechnology*. 26: 1293-1300.

Research Highlights in *Nature Methods* 6: 9 (2009).

Highlighted in Park PJ, *Nature Reviews Genetics* 10: 669-680 (2009).

Google Scholar Citations: 330+

14. Nishi Y, **Ji HK**, Wong WH, McMahon AP and Vokes SA (2009) Modeling the spatio-temporal network that drives patterning in the vertebrate central nervous system. *BBA - Gene Regulatory Mechanisms*. 1789: 299-305.
15. **Judy JT** and **Ji HK*** (2009) TileProbe: modeling tiling array probe effects using publicly available data. *Bioinformatics*. 25: 2369-2375.
16. Sun Y, Fan XY, Cao DM, Tang W, He K, Zhu JY, He JX, Bai MY, Zhu S, Oh E, Patil S, Kim TW, **Ji HK**, Wong WH, Rhee SY, Wang ZY. (2010) Integration of

- brassinosteroid signal transduction with the transcription network for plant growth regulation in Arabidopsis. *Developmental Cell*. 19: 765-777.
17. **Ji HK*** and Liu XS* (2010) Analyzing 'omics data using hierarchical models. *Nature Biotechnology*. 28: 337-340.
 18. Lee EY, **Ji HK**, Ouyang ZQ, Zhou BY, Ma WX, Vokes SA, McMahon AP, Wong WH, Scott MP (2010) Hedgehog pathway-regulated gene networks in cerebellum development and tumorigenesis. *Proc. Natl. Acad. Sci. USA*. 107: 9736-9741.
 19. **Huang CR**, Schneider AM, Lu Y, Niranjana T, Shen P, Robinson MA, Steranka JP, Valle D, Civin CI, Wang T, Wheelan SJ, **Ji HK**, Boeke JD, Burns KH (2010) Mobile interspersed repeats are major structural variants in the human genome. *Cell*. 141: 1171-1182.
- Comment* in Lupski JR, *Cell*, 141: 1110-1112 (2010).
- Huang CR** is a PhD student in human genetics (advisor: Boeke JD) enrolled concurrently in the biostatistics MHS program. I am her advisor for the MHS degree. We developed the computational tool for TIP-chip data analysis in this paper.
- Google Scholar Citations:* 90+
20. **Wu H** and **Ji HK*** (2010) JAMIE: joint analysis of multiple ChIP-chip experiments. *Bioinformatics*. 26: 1864-1870.
 21. Niakan KK, **Ji HK**[§], Maehr R[§], Vokes SA[§], Rodolfa KT, Sherwood RI, Yamaki M, Dimos JT, Chen AE, Melton DA, McMahon AP and Eggan K (2010) Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes & Development*. 24: 312-326.
 22. **Li C**, **Wu G**, **Ji HK*** (2011) hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data. *Bioinformatics*. 27: 1447-1448. PMID:PMC3087956
 23. Wang QF, **Wu G**, Mi SL, He FH, Wu J, Dong JF, Luo R, Mattison R, Kaberlein J, Prabhakar S, **Ji HK**, Thirman MJ. (2011) MLL fusion proteins preferentially regulate a subset of wild type MLL target genes in the leukemic genome. *Blood*. 117: 6895-6905. PMID:PMC3128481
 24. **Ji HK***, **Wu G**, Zhan X, Nolan A, Koh C, De Marzo A, Doan, HM, Fan JS, Cheadle C, Fallahi M, Cleveland JL, Dang CV*, Zeller K*. (2011) Cell-type independent MYC target genes reveal a primordial signature involved in biomass accumulation. *PLoS ONE*. 6:e26057. PMID: PMC3198433

25. Hwang W, Hackler Jr. L, **Wu G, Ji HK**, Zack DJ, Qian J. (2012) Dynamics of regulatory networks in the developing mouse retina. *PLoS ONE*. 7:e46521. PMID: PMC3463606
26. **Wei YY**[§], Li X[§], Wang QF, **Ji HK*** (2012) iASeq: integrating multiple ChIP-seq datasets for detecting allele-specific binding. *BMC Genomics*. 13:681. PMID: PMC3576346
27. Kano S, Colantuoni C, **Han F**, Zhou Z, Yuan Q, Wilson A, Takayanagi Y, Lee Y, Rapoport J, Eaton W, Cascella N, **Ji HK**, Goldman D, Sawa A. (2013) Genome-wide profiling of multiple histone methylations in olfactory cells: further implications for cellular susceptibility to oxidative stress in schizophrenia. *Molecular Psychiatry*. 18(7):740-2. Doi: 10.1038/mp.2012.
28. **Wei YY, Wu G, Ji HK*** (2013) Global mapping of transcription factor binding sites by sequencing chromatin surrogates: a perspective on experimental design, data analysis, and open problems. *Statistics in Biosciences*. 5:156-178.
29. Newman RH, Hu JF, Rho HS, Xie Z, Neiswinger J, Cooper C, Shirley M, Clark H, Hu SH, Hwang W, Jeong J, **Wu G**, Lin J, Gao XX, Ni Q, Goel R, Xia SL, **Ji HK**, Dalby K, Birnbaum MJ, Cole PA, Knapp S, Ryazanov A, Zack D, Blackshaw S, Pawson T, Gingras AC, Desiderio S, Pandey A, Turk B, Zhang J, Zhu H, Qian J. (2013) Construction of human activity-based phosphorylation networks. *Molecular Systems Biology*. 9:655.
30. **Wu G**[§], Yustein JT[§], McCall MN, Zilliox M, Irizarry RA, Zeller K, Dang CV, **Ji HK*** (2013) ChIP-PED enhances the analysis of ChIP-seq and ChIP-chip data. *Bioinformatics*. 29:1182-1189.
31. **Ji HK***, Li X, Wang QF, **Ning Y** (2013) Differential principal component analysis of ChIP-seq. *Proc. Natl. Acad. Sci. USA*. 110: 6789-6794. PMID: PMC3637734
32. Wang JY, Park JS, **Wei YY**, Rajurkar M, Cotton JL, Fan Q, Lewis BC, **Ji HK**, Mao JH (2013) TRIB2 acts downstream of Wnt/TCF in liver cancer cells to regulate YAP and C/EBPalpha function. *Molecular Cell*. 51:211-225.
33. **Wu G, Ji HK*** (2013) ChIPXpress: improved ChIP-seq and ChIP-chip target gene ranking using publicly available gene expression data. *BMC Bioinformatics*. 14:188.
34. Jaffe AE, Storey JD, **Ji HK**, Leek JT (2013) Gene set bagging for estimating the probability a statistically significant result will replicate. *BMC Bioinformatics*. 14:360.

Thesis

35. **Ji HK** (2002) Analysis of alternative splicing and complex disease based on population comparisons. Master's Dissertation. Tsinghua University.

36. **Ji HK** (2007) Decoding mammalian gene regulatory programs through efficient microarray, ChIP-chip and sequence analysis. PhD thesis, Harvard University

Book Chapters

37. Cullen P et al. including **Ji HK** (2004) Computational methods and bioinformatic tools. *Analysing Gene Expression*, edited by Lorkowski S and Cullen P, pp 769-904, Wiley-VCH Verlag GmbH & Co. KGaA
38. **Ji HK** (2010) Computational Analysis of ChIP-seq Data. In the *Computational Biology of Transcription Factor Binding, Methods in Molecular Biology*, 674: 143-159. Edited by Istvan (Steve) Ladunga, Humana Press, Springer
39. **Ji HK**, Jiang H, Ma WX and Wong WH (2011) Using CisGenome to Analyze ChIP-chip and ChIP-seq Data. *Current Protocols in Bioinformatics*. 2011 Mar, Chapter 2: Unit2.13. Wiley
40. **Ji HK** (2011) Computational Analysis of ChIP-chip Data. *Handbook of Computational Statistics*, edited by Henry Horng-Shing Lu, Bernhard Schölkopf and Hongyu Zhao. 2011, Part 2, 257-282, DOI: 10.1007/978-3-642-16345-6_12. Springer
41. **Wu H** and **Ji HK** (2012) JAMIE: A Software Tool for Jointly Analyzing Multiple ChIP-chip Experiments. In the *Next Generation Microarray Bioinformatics, Methods Mol Biol.*, 802: 363-75, Humana Press, Springer

Book Translation

42. Translated the *Bioinformatics (The Machine Learning Approach)* by Pierre Baldi and Soren Brunak into Chinese, joint work with Zhang DH, Huang Y, Cai J, Sun YF, Xia HY, Hu CF, Zhu ZH.

Conferences

43. Vokes SA, **Ji HK**, Wong WH, McMahon AP (2005) Identification of Gli target genes using chromatin immuno-precipitation with a genetically inducible system on genomic arrays. In the 64th Annual Meeting of the Society for Developmental Biology, San Francisco, CA. *Developmental Biology*, 283(2): 666
44. Vokes SA, **Ji HK**, Wong WH, McMahon AP (2006) A tale of two morphogen gradients: identifying Gli targets of hedgehog signaling. In the 65th Annual Meeting of the Society for Developmental Biology, Ann Arbor, MI. *Developmental Biology*, 295(1): 423
45. Wang QJ, **Wu G**, Mi SL, He FH, Dong JF, Mattison RJ, Kaberlein J, Wang QB, Prabhakar S, **Ji HK**, Thirman MJ (2009) MLL fusion proteins directly regulate a

small set of wild type MLL target genes. In the 51st Annual Meeting of the American Society of Hematology, New Orleans, LA. *Blood*, 114(22): 525

46. Mi SL, He FH, Wu J, Zhou J, **Wu G**, Thirman MJ, **Ji HK**, Wang QF (2010) Genome wide location analysis reveals deregulated microRNA genes in MLL rearranged leukemic genome. In the 52nd Annual Meeting of the American Society of Hematology (ASH), Orlando, FL. *Blood*, 116(21): 1039

Technical Reports

47. **Ji HK** and Wong WH (2005) Increasing power of microarray gene selection: an empirical Bayes approach. Harvard Qualifying Paper

SOFTWARE AND DATABASE

CisGenome: An integrated software system with graphic user interface for analyzing ChIP-seq and ChIP-chip data, DNA motif discovery, and genome annotations.
<http://www.biostat.jhsph.edu/~hji/cisgenome/>
From 12/2008: >20,000 downloads; 330+ citations according to Google Scholar.

TileMap: Software for analyzing whole genome tiling array data.
<http://jilab.biostat.jhsph.edu/software/tilemap/index.htm>
From 10/2005: 200+ citations according to Google Scholar.

dPCA: Software for analyzing differential protein-DNA interactions.
<http://www.biostat.jhsph.edu/~hji/dpca/>

ChIP-PED: Using public gene expression data in GEO to boost the value of ChIP-seq and ChIP-chip experiments.
<http://www.biostat.jhsph.edu/~gewu/ChIPPED/>

ChIPXpress [R/Bioconductor package]: Improved ChIP-seq and ChIP-chip target gene ranking using publicly available gene expression data.
<http://www.biostat.jhsph.edu/~gewu/ChIPXpress/index.html>

Cormotif [R/Bioconductor package]: Integrative analysis of multiple gene expression datasets for simultaneously detecting differentially expressed genes and patterns.
<http://www.bioconductor.org/packages/release/bioc/html/Cormotif.html>

hmChIP: A database of public human and mouse ChIP-seq/ChIP-chip data.
<http://jilab.biostat.jhsph.edu/database/cgi-bin/hmChIP.pl>

iASeq [R/Bioconductor package]: Integrative analysis of allele-specificity in multiple ChIP-seq experiments.
<http://bioconductor.org/packages/2.11/bioc/html/iASeq.html>

JAMIE: Joint analysis of multiple ChIP-chip experiments.
<http://www.biostat.jhsph.edu/~hji/jamie/>

PolyaPeak: Detecting ChIP-seq peaks using peak shape information.
<http://userwww.service.emory.edu/~hwu30/polyaPeak.html>

TileProbe: Software for removing tiling array probe effects using public data in GEO
http://www.biostat.jhsph.edu/~hji/cisgenome/index_files/tileprobe.htm

PATENT

CN200310122497.6: Reagents and primers for predicting susceptibility to paranoid schizophrenia, by Xu Q, Shen Y, **Ji HK**, Ji L, Yuan YB, Shen YC (China).

CURRICULUM VITAE

HONGKAI JI

PART II

TEACHING

Advisees (Postdoctoral Fellow)

Xuekui Zhang, PhD <u>Current Position:</u> Eli Lilly and Company	2011-2013	Biostatistics
Weiqiang Zhou, PhD	2013-present	Biostatistics
Fang Du, PhD	2013-present	Biostatistics

Advisees (PhD students)

Hao Wu, PhD (Recipient of 2010 Margaret Merrell Award) <u>Thesis:</u> Three statistical applications in genomics: redefining CpG islands, peak detection from multiple ChIP-chips, and data normalization for second generation sequencing. <u>Advisors:</u> Rafael Irizarry and Hongkai Ji <u>Current Position:</u> Assistant Professor, Emory University	2007-2010	Biostatistics
Yang Ning, PhD <u>Thesis:</u> Nonstandard likelihood based inference <u>Advisors:</u> Kung-Yee Liang and Hongkai Ji <u>Current Position:</u> Postdoc, University of Waterloo	2011-2012	Biostatistics
George Wu, PhD <u>Thesis:</u> Statistical methods for the integrative analysis of high-throughput genomic data <u>Current Position:</u> Scientist, Crescendo Bioscience	2007-2013	Biostatistics
Yingying Wei, PhD candidate (Recipient of 2012 Culley Award, 2013 ENAR Student Travel Award)	2011-present	Biostatistics
Dan Jiang, PhD candidate	2013-present	Biostatistics
Yuting Xu, PhD candidate (Academic advisor)	2012-2013	Biostatistics
Jeongyong Kim, PhD candidate (Academic advisor)	2009-2011	Biostatistics

Fang Han, PhD candidate (Academic advisor)	2010-2011	Biostatistics
--	-----------	---------------

Advisees (Master's students)

Li Chen, MHS-Bioinformatics	2009-2011	Biostatistics
<u>Capstone Project:</u> hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data		
<u>Current Position:</u> PhD student, Computer Science, Emory University		

Jiehuan Sun, ScM	2011-2012	Biostatistics
<u>Thesis:</u> Systematic annotation and exploration of Gene Expression Omnibus		
<u>Current Position:</u> PhD student, Biostatistics, Yale University		

Cheng Ran Huang, MHS candidate	2010-2013	Public Health
--------------------------------	-----------	---------------

Wei Wang, MHS candidate	2011-present	Public Health
-------------------------	--------------	---------------

Zheng Kuang, MHS candidate	2012-present	Public Health
----------------------------	--------------	---------------

Shilu Zhang, ScM candidate	2013-present	Biostatistics
----------------------------	--------------	---------------

Zhicheng Ji, ScM candidate	2013-present	Biostatistics
----------------------------	--------------	---------------

Ashwini Patil, MS candidate	2013-present	Bioinformatics
-----------------------------	--------------	----------------

Students Working on Projects

Jennifer T. Judy	2008-2009	Mental Health
Thai Le	2007-2008	Biomedical Engineering and Mathematics
Xueying Ji	2009	Geography and Environmental Engineering
Luolan Li	2012	Visiting student from Nanjing University
Jiawei Bai	2012	Biostatistics
Bing He	2013	Biostatistics

PhD Thesis Advisory Committee

Hao Wu	01/01/2008	Biostatistics
Simina Boca	03/11/2010	Biostatistics
Yang Ning	01/01/2011	Biostatistics
George Wu	09/01/2010	Biostatistics
Yingying Wei	12/01/2011	Biostatistics
David Gorkin	03/31/2011	Human Genetics

Rahul Karnik	01/01/2012	Biomedical Engineering
--------------	------------	------------------------

Preliminary Oral Participation

Hao Wu	01/10/2008	Biostatistics
Ji Li	06/09/2008	International Health
Simina Boca	12/10/2008	Biostatistics
George Wu	09/03/2010	Biostatistics
Yiyi Zhang	01/28/2011 (I am an alternate)	Epidemiology
Yingying Wei	12/14/2011	Biostatistics
Taeyoung Hwang	11/09/2012	Biomedical Engineering
Man Li	12/20/2013	Epidemiology

Final Oral Participation

Hao Wu	05/19/2010	Biostatistics
Tao Wu	11/24/2010 (I am an alternate)	Epidemiology
Simina Boca	03/08/2011 (I am an alternate)	Biostatistics
Yang Ning	04/10/2012	Biostatistics
Rahul Karnik	08/08/2012	Biomedical Engineering
George Wu	05/10/2013	Biostatistics

Classroom Instruction

Methods in Biostatistics III (140.653)	Principal Instructor, 2013-present
2013 Enrollment 32	
Methods in Biostatistics IV (140.654)	Principal Instructor, 2013-present
2013 Enrollment 35	
Advanced Statistical Computing (140.778)	Principal Instructor, 2008-2011
2008 Enrollment 5	
2009 Enrollment 5	
2010 Enrollment 20	
2011 Enrollment 10	
Special Topics in Genetics and Genomics (140.668)	Principal Instructor, 2009
2008 Enrollment 6	
Statistical Computing (140.776)	Principal Instructor, 2009-2011
2009 Enrollment 34	
2010 Enrollment 41	
2011 Enrollment 38	

Advanced Statistical Theory II (140.772)	Co-instructor, 2007
Statistics for Genomics (140.688)	Guest Lectures, 2010, 2012
Genomics (260.605)	Guest Lectures, 2009

Others

1. Advised Alexander Smith, a high school student, for the Center Scholar Program, Center for Talented Youth 07/2012 - 08/2012
2. Taught a five-day summer course *Statistical Computing* for the Hopkins-Nanjing Exchange Program in Statistical and Data Sciences 07/2011
3. Advised Rachel Napthal from Weston High School, MA for the Center Scholar Program, Center for Talented Youth 06/2009 - 08/2009
4. Offered online tutorial session “Introduction to CisGenome” at the *Affymetrix Workshop at EMBL*. 11/2007

RESEARCH GRANT PARTICIPATION

PI Grants

Title of Grant: Computational Tools for Mining Large Amounts of ChIP and Gene Expression Data

Principal Investigator: Ji, HK

Dates: 07/25/2012- 04/30/2017

Sponsoring Agency: NIH/NHGRI

Grant No.: R01HG006282

Award Amount: \$ 2,035,000 (Direct cost: 250,000 / year x 5 years)

Main Grant Objective: Develop computational and statistical tools that integrate large amounts of publicly available gene expression data with ChIP-seq and ChIP-chip data to systematically map functional contexts of gene regulatory pathways.

Effort Allocation: 3.0 calendar months (25%)

Title of Grant: Statistical and Computational Tools for Next-generation ChIP-seq Applications

Principal Investigator: Ji, HK

Dates: 09/12/2012- 05/31/2015

Sponsoring Agency: NIH/NHGRI

Grant No.: R01HG006841

Award Amount: \$972,000 (Direct cost: \$200,000 / year x 3 years)

Main Grant Objective: Develop computational and statistical tools for analyzing genome-wide gene regulation dynamics and allele-specific binding using ChIP-seq data.
Effort Allocation: 1.8 calendar months (15%)

Title of Grant: Global Prediction of Transcription Factor Binding Sites in Lineage Specific Neuronal Differentiation

Principal Investigator: Ji, HK

Dates: 07/01/2012- 06/30/2014

Sponsoring Agency: TEDCO/Maryland Stem Cell Research Fund

Grant No.: 2012-MSCRFE-0135-00

Award Amount: \$ 230,000 (Direct cost: \$100,000 / year x 2 years)

Main Grant Objective: We propose to develop computational methods to predict genome-wide transcription factor (TF) binding sites for many TFs and many biological conditions simultaneously based on exon array data. Applying this approach, we will study key regulatory programs in differentiation programs of human stem cell derived neuron progenitor cells.

Effort Allocation: 1.2 calendar months (10%)

Collaborative Grants

Title of Grant: Regulation of Neural and Neoplastic Stem Cells by Kruppel-Like Transcription Factors

Principal Investigator: Laterra, J

Dates: 07/01/2011 – 06/30/2014

Sponsoring Agency: TEDCO/Maryland Stem Cell Research Fund

Grant No.: 2011-MSCRFFII-0073-00

Main Grant Objective: We will analyze ChIP-seq data to identify a KLF target genes in neural and neoplastic stem cells and will participate in follow-up genomic analyses.

Principal Responsibilities of Individual: Subcontract PI

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Suppression of Glioblastoma Stem Cells by Kruppel-Like Factor 9

Principal Investigator: Laterra, J

Dates: 09/15/2011 – 07/31/2016

Sponsoring Agency: NIH

Grant No.: R01NS076759

Main Grant Objective: The research focuses on understanding glioblastoma stem cell regulation by the KLF9 transcription factor and to develop KLF9-based Brain cancer treatment strategies.

Principal Responsibilities of Individual: Subcontract PI

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Brain Cancer Stem Cell Reprogramming by c-Met

Principal Investigator: Laterra, J

Dates: 05/15/2012 – 04/30/2017

Sponsoring Agency: NIH
Grant No.: R01NS073611

Main Grant Objective: This research will study how the c-Met tyrosine kinase and reprogramming transcription factors regulate glioblastoma stem cells.

Principal Responsibilities of Individual: Subcontract PI

Effort Allocation: 0.6 calendar months (10%)

Title of Grant: Highly efficient conversion of human stem cells to dopaminergic neurons by proneural transcription factor Atoh1

Principal Investigator: Ying, MY

Dates: 07/01/2012- 06/30/2014

Sponsoring Agency: TEDCO/Maryland Stem Cell Research Fund

Grant No.: TBA

Main Grant Objective: This project will use high-throughput genomic technologies to study reprogramming programs in human stem cells.

Principal Responsibilities of Individual: Subcontract PI

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Cis-regulatory circuitry underlying Hedgehog mediated limb development

Principal Investigator: Vokes, SA

Requested Dates: 07/01/2012- 06/30/2017

Sponsoring Agency: NIH

Grant No.: 1R01HD073151-01

Main Grant Objective: Combine high-throughput genomic technologies with transgenic experiments to study gene regulatory programs in the hedgehog signaling pathway and limb development

Principal Responsibilities of Individual: Subcontract PI

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Center for systems biology of retrotransposition

Principal Investigator: Boeke, JD

Dates: 09/05/2013 – 05/31/2018

Sponsoring Agency: NIH

Grant No.: 1P50GM107632

Main Grant Objective: Develop systems approaches to understand and manipulate the vital, highly coordinated relationship between host and transposons in targeted ways.

Principal Responsibilities of Individual: Co-investigator

Effort Allocation: 0.6 calendar months (5%)

Completed Grants

Title of Grant: Decoding transcriptional regulatory programs in tumors dependent on hedgehog signaling

Principal Investigator: Ji, HK

Dates: 08/01/2007-08/01/2008

Sponsoring Agency: The Richard L. Gelb Cancer Research Fund, Johns Hopkins Bloomberg School of Public Health

Award Amount: \$ 10,000

Main Grant Objective: Identify genes that are direct targets of Gli proteins in Hh induced medulloblastoma, BCC and RMS, by utilizing genome-wide chromatin immunoprecipitation (ChIP) analysis coupled with genomic tiling (chip) arrays in a novel mouse model of Hh-related tumors. Develop statistical methods to analyze tiling array data and to compare TF-DNA binding activities across tumor samples.

Title of Grant: Center for the Epigenetics of Common Human Diseases

Principal Investigator: Feinberg, AP

Dates: 05/14/2007-04/30/2010

Sponsoring Agency: NIH/NHGRI

Grant No.: 5P50HG003233

Main Grant Objective: Develop high throughput tools for epigenome analysis. Develop statistical methods to analyze allele specific expression data.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 2.4 calendar months (20%)

Title of Grant: Bioinformatics and Resequencing in Mood Disorders

Principal Investigator: Potash, JB

Dates: 08/01/2008-07/31/2010

Sponsoring Agency: Buerger Fund for Bipolar Disorder

Main Grant Objective: Develop bioinformatics tools for analyzing massively parallel sequencing data for mood disorders.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 1.8 calendar months (15%)

Title of Grant: Microarray Detection of Retrotransposons

Principal Investigator: Boeke, JD

Dates: 09/22/2008-09/21/2010

Sponsoring Agency: JHMI Microarray Core

Main Grant Objective: Develop statistical models for detecting active transposon elements in human genome.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 1.2 calendar months (10%)

Title of Grant: Epigenetic Variation and Its Determinants in Depression

Principal Investigator: Potash, JB

Dates: 08/01/2008-07/31/2009

Sponsoring Agency: NIH/NIMH

Grant No.: R01MH074131

Main Grant Objective: The goal of this project is to find and measure epigenetic marks in subjects with major depressive disorder and to test potential genetic and environmental factors influencing epigenetic variation.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 1.2 calendar months (10%)

Title of Grant: Johns Hopkins Center in Urban Environmental Health

Principal Investigator: Groopman, J

Dates: 04/01/2008-03/31/2013

Sponsoring Agency: NIH/NIEHS

Grant No.: 2P30ES003819

Main Grant Objective: The Biostatistics Center consults with environmental health scientists including molecular toxicologists, physiologists, environmental engineers and epidemiologists.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 1.2 calendar months (10%)

Title of Grant: Characterizing Blood Progenitor Cells Differentiated from Human iPS and ES cells

Principal Investigator: Cheng, LZ

Dates: 09/30/2009-08/31/2011

Sponsoring Agency: NIH

Grant No.: RC2 GO grant HL101582 - 01

Main Grant Objective: Using genomic and epigenetic analysis tools to characterize similarities and differences between embryonic stem cells and iPS cells.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 1.2 calendar months (10%)

Title of Grant: Oral Clefts: Moving from Genome Wide Studies toward Functional Genomics

Principal Investigator: Beaty, T

Dates: 09/01/2009-08/31/2014

Sponsoring Agency: NIH/NIDCR

Grant No.: U01-DE-092003

Main Grant Objective: This study aims to investigate genes/regions yielding evidence of containing causal genes from a genome wide study of oral clefts as part of the FaceBase consortium.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Characterization of Novel Sequence Motifs Important for Stem Cell Biology

Principal Investigator: Zeller, KI

Dates: 09/01/2010 – 08/30/2012

Sponsoring Agency: TEDCO/Maryland Stem Cell Research Fund

Grant No.: 2010-MSCRF-0050-00

Main Grant Objective: Develop computational and experimental approaches to identify proteins that bind to novel DNA motifs in human ESC.

Principal Responsibilities of Individual: Co-PI

Effort Allocation: 1.2 calendar months (10%)

Title of Grant: Preprocessing and Analysis Tools for Contemporary Microarray Applications

Principal Investigator: Irizarry, RA

Dates: 09/24/2007 – 08/31/2012

Sponsoring Agency: NIH

Grant No.: R01GM083084

Main Grant Objective: Develop statistical methods and bioinformatics tools for analyzing high-density tiling array data.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 1.2 calendar months (10%)

Title of Grant: Analysis Tools and Software for Second Generation Sequencing Data

Principal Investigator: Irizarry, RA

Dates: 12/01/2009 – 11/30/2013

Sponsoring Agency: NIH

Grant No.: R01HG005220

Main Grant Objective: Develop statistical methods and bioinformatics tools for analyzing high-throughput sequencing data.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 2.4 calendar months (20%)

Title of Grant: c-Myc Targets in the Pathogenesis of Human Cancers

Principal Investigator: Dang, CV

Dates: 07/01/2010 – 12/31/2014

Sponsoring Agency: NIH

Grant No.: R01CA57341

Main Grant Objective: Study roles of MYC oncogene in human cancers.

Principal Responsibilities of Individual: Co-Investigator

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Johns Hopkins – Nanjing Exchange Program in Statistical and Data Sciences

Principal Investigator: Ji, HK

Dates: 06/01/2011- 06/30/2013

Sponsoring Agency: JHU Benjamin and Rhea Yeung Center for Collaborative Chinese Studies

Award Amount: \$ 150,000

Main Grant Objective: We propose to establish the Johns Hopkins (JHU) – Nanjing University (NJU) exchange program in statistical and data sciences. The program involves bi-directional educational and research exchanges between JHU and NJU, and we aim to promote new cross-institutional and interdisciplinary collaborations.

Title of Grant: Flow Cytometry Analysis of Lyme Diseases

Principal Investigator: Soloski, M; Aucott J

Dates: 09/01/2011-08/31/2013

Sponsoring Agency: Non-sponsored

Main Grant Objective: This study aims to use flow cytometry and high-throughput genomics technologies to investigate immune response to lyme diseases.

Principal Responsibilities of Individual: Co-Investigator
Effort Allocation: 0.6 calendar months (5%)

ACADEMIC SERVICE

Department of Biostatistics

Member, Post Doctoral Fellow Committee (2007)
Member, Faculty Search Committee (2008)
Member, Graduate Program Admission Committee (2010 - present)
Member, Merrell Award Committee (2011)
Member, Statistical Genomics Bloomberg Professor Search Committee (2013-2014)

Bloomberg School of Public Health

Committee on Information Technology (2010)
Faculty Senate (2011 - 2013)

PRESENTATIONS

Scientific Meetings (Invited)

ChIP-seq: moving from single dataset peak calling to multiple experiment data integration. *ENAR, Orlando, FL, Mar 13, 2013*

Predicting functional targets of transcription factors by coupling ChIP-seq with GEO. *JSM, San Diego, CA, Aug 2, 2012.*

Differential Principal Component Analysis of ChIP-seq. *ENAR, Washington DC, Apr 2, 2012*

Integrating ChIP-seq data with other high dimension data types. Statistical analyses for next generation sequencing. *Birmingham, AL, Sep 26-27, 2011*

Differential Principal Component Analysis of ChIP-seq. *JSM, Miami, FL, July 31, 2011*

Differential Principal Component Analysis of ChIP-seq. *WNAR, San Luis Obispo, CA, June 21, 2011*

Functional interpretation of ChIP-seq using publicly available gene expression data. *ENAR, Miami, FL, Mar 21, 2011*

A latent mixture model for analyzing multiple related ChIP-chip and gene expression data sets. *ENAR, New Orleans, LA, Mar 24, 2010*

FlexModule: a flexible cis-regulatory module sampler. *IMS-China International Conference on Statistics and Probability, Weihai, Shandong, China, July 4, 2009*

A correlation motif based hidden Markov model for pooling information from multiple ChIP-chip experiments. *ENAR, San Antonio, TX, Mar 17, 2009*

CisGenome: an integrated tool for the analysis of ChIP-chip tiling array data. *The first North American ChIP-on-chip Scientific Meeting, Affymetrix, Boston, MA, Sept 14, 2007*

Scientific Meetings (Contributed)

Efficient exploration of multiple ChIP-seq and ChIP-chip data sets. *JSM, Vancouver, BC, Canada, Aug 4, 2010*

Invited Seminars

Differential Principal Component Analysis of ChIP-seq. *Department of Statistics, Pennsylvania State University, State College, PA, Oct 10, 2013.*

iASeq: integrative analysis of allele-specificity of protein-DNA interactions in multiple ChIP-seq datasets. *Beijing Institute of Genomics Chinese Academy of Sciences, Beijing, China, July 18, 2013.*

ChIP-seq: Unleashing its full potential through data integration. *University of Maryland, College Park, MD, June 6, 2013*

ChIP-seq: Unleashing its full potential through data integration. *Department of Biostatistics & Epidemiology, University of Pennsylvania, Philadelphia, PA, Apr 24, 2013*

Differential Principal Component Analysis of ChIP-seq. *Department of Statistics, Yale University, New Haven CT, Apr 16, 2012.*

Differential Principal Component Analysis of ChIP-seq. *Department of Biostatistics, Bioinformatics, Georgetown University, Washington DC, Mar 23, 2012.*

Differential Principal Component Analysis of ChIP-seq. *Workshop II: Transcriptomics and Epigenomics, Institute for Pure & Applied Mathematics, University of California, Los Angeles, CA, Oct 25-28, 2011.*

Differential Principal Component Analysis of ChIP-seq. *Beijing Institute of Genomics Chinese Academy of Sciences, Beijing, China, July 27, 2011.*

Integrative analysis of many ChIP-seq and ChIP-chip experiments. *Department of Statistics, University of California, Berkeley, CA, May 5, 2011.*

Integrative modeling and analysis of multiple ChIP-chip experiments. *National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC*, Jan 18, 2011.

Integrating ChIP-seq and ChIP-chip with publicly available microarray data provides a new way to make discoveries. *Beijing Institute of Genomics Chinese Academy of Sciences, Beijing, China*, Jun 10, 2010.

Joint analysis of multiple genome-wide chromatin immunoprecipitation experiments. *Department of Statistics, University of British Columbia, Vancouver, BC, Canada*, Jun 3, 2010. **(Constance van Eeden Distinguished Speaker Seminar)**

Integrative analysis of multiple chromatin immunoprecipitation data from public domains. *Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, WI*, Apr 23, 2010.

Improving high-throughput data analysis by using gene expression omnibus. *Department of Biostatistics, Harvard University School of Public Health, Boston, MA*, Oct 19, 2009.

Decoding gene regulatory Program using ChIP-chip and ChIP-seq. *Department of Biology, The Johns Hopkins University, Baltimore, MD*, Oct 15, 2009.

Decoding gene regulatory Program using ChIP-chip and ChIP-seq. *Bioinformatics Division, Tsinghua National Lab of Information Science and Technology, Beijing, China*, July 27, 2009.

Decoding gene regulatory Program using ChIP-chip and ChIP-seq. *Beijing Institute of Genomics Chinese Academy of Sciences, Beijing, China*, July 8, 2009.

Decoding gene regulatory Program using ChIP-chip and ChIP-seq. *National Heart Lung and Blood Institute, NIH, Bethesda, MD*, Feb 5, 2009.

CisGenome: an integrated system for analyzing ChIP-chip and ChIP-seq data. *Department of Bioinformatics and Computational Biology, George Mason University, Manassas, VA*, Oct 21, 2008.

Detecting gene regulatory signals from genome-wide ChIP-chip and sequence data. *Department of Biostatistics & Epidemiology, University of Pennsylvania*, Oct 23, 2007.

Analysis of ChIP-chip tiling array data for mammalian transcription factors. *Affymetrix Northern California ChIP-on-chip Workshop, Berkeley, CA*, Jul 24, 2007.

Detecting gene regulatory signals from genome-wide ChIP-chip and sequence data. *Department of Statistics, Purdue University*, Mar 7, 2007.

Detecting gene regulatory signals from genome-wide ChIP-chip and sequence data. *Department of Biostatistics, University of North Carolina at Chapel Hill*, Mar 5, 2007.

Detecting gene regulatory signals from genome-wide ChIP-chip and sequence data. *Department of Statistics, North Carolina State University*, Feb 23, 2007.

Detecting gene regulatory signals from genome-wide ChIP-chip and sequence data. *Department of Biostatistics, Johns Hopkins University*, Feb 19, 2007.

Detecting gene regulatory signals from genome-wide ChIP-chip and sequence data. *Department of Biostatistics, University of Washington*, Feb 15, 2007.

Detecting gene regulatory signals from genome-wide ChIP-chip experiments and sequence data: towards developing a strategy with increased statistical power, *Department of Statistics, Pennsylvania State University*, Feb 6, 2007.

Detecting gene regulatory signals from genome-wide ChIP-chip and sequence data. *Department of Statistics, Texas A&M University*, Jan 30, 2007.

Statistical analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. *The Stanford Workshop in Biostatistics, Stanford, CA*, Nov 16, 2006.

CisGenome – an *in silico* lab for the analysis of ChIP-chip tiling array data. *The Affymetrix GeneChip Exon and Tiling Arrays Data Analysis Workshop, Cambridge, MA*, Mar 23, 2006.

Others

The Future of Statistics in Biology. *The Simply Statistics Unconference on the Future of Statistics*, Oct 30, 2013.

ADDITIONAL INFORMATION

Personal statement of research and research objectives

I am interested in developing statistical and computational methods for analyzing high-throughput genomic data. I apply these tools to study gene regulatory programs in development and diseases. I have developed several popular tools for analyzing genome-wide tiling array and next-generation sequencing data, including CisGenome, an integrated software system for analyzing ChIP-seq and ChIP-chip data, and TileMap, a software tool for analyzing tiling array data. Recently, I have been working extensively on data integration. We have proposed new statistical methods and developed new software tools and a database for mining and utilizing massive amounts of gene expression, tiling array and ChIP-seq data. These include dPCA, TileProbe, JAMIE, hmChIP, iASeq, ChIP-PED, ChIPXpress, Cormotif, etc. By closely working with

collaborators, we have applied these methods to studying gene regulatory programs in mouse Sonic Hedgehog (SHH) signaling pathway, human MYC regulatory pathway in embryonic stem cells and cancers, mouse Sox17 regulation in stem cell differentiation, and gene regulatory programs in many other developmental and disease systems. My long term goal is to establish effective and efficient statistical strategies to integrate rich yet heterogeneous 'omics information for reconstructing complex gene regulatory networks in higher organisms. I wish that knowledge obtained from my research could help people to better understand human diseases and improve human health.

Keywords

Genomics, computational biology, bioinformatics, gene regulation, epigenomics, next-generation sequencing, microarray, tiling array, ChIP-seq, ChIP-chip, gene expression, motif, stem cell, development, cancer

Bayesian statistics, empirical Bayes, statistical computing, large data, hierarchical models, latent variable model, Markov Chain Monte Carlo, data integration, data mining