

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@jhu.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
07/2017 – present	Graduate Program Director, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
08/2015 – 07/2017	Graduate Program Co-director, 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 and Leadership

Program Chair-Elect (ASA Section on Statistics in Genomics and Genetics, 2018)
American Statistical Association (ASA)
Institute of Mathematical Statistics (IMS)
International Chinese Statistical Association (ICSA)
International Society for Computational Biology (ISCB)

Program or Project Development

2009 Session Organizer, ENAR
2010 Session Organizer, ICSA
2010 Session Chair, ENAR
2011 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
2014 Session Organizer and Session Chair, ENAR
2014 ASA White Paper for “Discovery with Data: Leveraging Statistics and Computer Science to Transform Science and Society”
2016 Session Organizer and Session Chair, ENAR
2016 Session Chair, ICSA Applied Statistics Symposium
2016 Session Chair, IMS Asia Pacific Rim Meeting
2016 Session Chair, JSM
2016 Session Organizer and Session Chair, The 10th ICSA International Conference
2017 Session Organizer and Session Chair, ENAR
2017 Editor, Special Issue on Statistics and Genomics for Statistics in Biosciences

EDITORIAL ACTIVITIES

Editorial Board Membership

Associate Editor, Statistical Applications in Genetics and Molecular Biology (SAGMB), 2011-present
Associate Editor, Statistics in Biosciences (SIBS), 2015-present

Peer Review Activities

Annals of Applied Statistics (AOAS)

Bioinformatics
Biometrics
Biostatistics
BMC Bioinformatics
BMC Genomics
Cancer Research
Cell Research
Cell Stem Cell
Computational Statistics
Computational Statistics and Data Analysis (CSDA)
F1000Research
Genome Biology
Genome Research
Genomics
IEEE/ACM Transactions on Computational Biology and Bioinformatics
IEEE Transactions on NanoBioscience
ISMB
Journal of Computational Biology (JCB)
Journal of the American Statistical Association (JASA)
Journal of the Royal Statistical Society, Series B (JRSSB)
Nature Biotechnology
Nature Communications
Nature Methods
Nucleic Acids Research (NAR)
Pacific Symposium on Biocomputing (PSB)
PLoS Computational Biology
Proceedings of the National Academy of Sciences of the USA (PNAS)
Scientific Reports
Statistica Sinica
Statistical Analysis and Data Mining (SADM)
Statistical Applications in Genetics and Molecular Biology (SAGMB)
Statistics in Biosciences (SIBS)
Wiley Interdisciplinary Reviews: Systems Biology and Medicine

Book Review

Numerical Analysis for Statisticians (Second Edition), Springer

Review of Proposals

08/2007	BBSRC Grant Review, UK
05/2010	NIH Challenge Grant Review, US
11/2011	NIH Special Emphasis Review Panel, ZHG1 HGR-M (J2), US
06/2014	NIH Study Section, GCAT, US
07/2014	NIH Special Emphasis Review Panel, Endocrinology, Metabolism, Nutrition and Reproductive Sciences IRG, US

10/2014 NIH Study Section, BMRD, US
 10/2014 NIH Member Conflict Review Panel, Genes, Genomes, and Genetics IRG, US
 12/2014 NIH Member Conflict Review Panel, Development, Differentiation and Modulation of Immune Responses, US
 02/2016 NIH Study Section, BDMA, US
 08/2016 MRC Grant Review, UK
 07/2017 NIH NIGMS Review Panel for the Centers of Biomedical Research Excellence (COBRE) Program, US
 10/2017 NIH Study Section, MABS, US

HONORS AND AWARDS

Honors

2017 Excellence in Teaching, Johns Hopkins Bloomberg School of Public Health
 2016 Excellence in Teaching, Johns Hopkins Bloomberg School of Public Health
 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

2016 Johns Hopkins Catalyst Award
 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; blue: advisees and lab members)

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: 270+

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: 780+

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: 220+

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.

Google Scholar Citations: 220+

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: 700+

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, Niranjan 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: 190+

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.

Google Scholar Citations: 200+

22. **Chen L**, **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

Google Scholar Citations: 90+

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.
35. **Wu H***, **Ji HK*** (2014) PolyPeak: detecting transcription factor binding sites from ChIP-seq using peak shape information. *PLoS ONE*. 9:e89694.

36. Kuang Z, Cai L, Zhang XK, Ji HK, Tu BP, Boeke JD (2014) High-temporal-resolution view of transcription and chromatin states across distinct metabolic states in budding yeast. *Nature Structural & Molecular Biology*. 21: 854-863. doi:10.1038/nsmb.2881.
37. Ying MY, Tilghman J, Wei YY, Guerrero-Cazares H, Quinones-Hinojosa A, Ji HK, Laterra J (2014) KLF9 inhibits glioblastoma stemness through global transcription repression and integrin- $\alpha 6$ inhibition. *Journal of Biological Chemistry*. 289: 32742-32756. doi: 10.1074/jbc.M114.588988
38. Wei YY, Tenzen T, Ji HK* (2015) Joint analysis of differential gene expression in multiple studies using correlation motifs. *Biostatistics*. 16: 31-46. doi:10.1093/biostatistics/kxu038.
39. Zhang SL, Du F, Ji HK* (2015) A novel DNA sequence motif in human and mouse genomes. *Scientific Reports*. 5: 10444 doi:10.1038/srep10444.
40. Wamaitha SE, del Valle I, Cho LTY, Wei YY, Fogarty NME, Blakeley P, Sherwood RI, Ji HK, Niakan KK (2015) Gata6 potently initiates reprogramming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes & Development*. 29: 1239-1255
41. Jin K, Park S, Teo WW, Korangath P, Cho SS, Yoshida T, Györfy B, Goswami CP, Nakshatri H, Cruz LA, Zhou WQ, Ji HK, Su Y, Ekram M, Wu Z, Zhu T, Polyak K, Sukumar S (2015) HOXB7 is an ER α cofactor in the activation of HER2 and multiple ER target genes leading to endocrine resistance. *Cancer Discovery*. 5: 944-959. doi: 10.1158/2159-8290.CD-15-0090
42. Lewandowski JP, Du F, Zhang SL, Powell MB, Falkenstein KN, Ji HK, Vokes SA (2015) Spatiotemporal regulation of GLI target genes in the mammalian limb bud. *Developmental Biology*. 406: 92-103. doi: 10.1016/j.ydbio.2015.07.022
43. Ji ZC, Vokes SA, Dang CV, Ji HK* (2016) Turning publicly available gene expression data into discoveries using gene set context analysis. *Nucleic Acids Research*. 44(1): e8. doi: 10.1093/nar/gkv873
44. Li Q, Lex RK, Chung HW, Giovanetti SM, Ji ZC, Ji HK, Person MD, Kim J, Vokes SA (2016) NANOG binds to GLI proteins and represses hedgehog-mediated transcription. *Journal of Biological Chemistry*. 291:7171-7182. doi: 10.1074/jbc.M116.714857
45. Hong XM, Ladd-Acosta C, Hao K, Sherwood B, Ji HK, Keete CA, Kumarf R, Caruso D, Liu X, Wang GY, Chen Z, Ji YL, Mao GY, Walker SO, Bartell TR, Ji ZC, Sun YF, Tsai HJ, Pongracic JA, Weeks DE, Wang XB (2016) Epigenome-wide

- association study links site-specific DNA methylation changes with cow's milk allergy. *Journal of Allergy and Clinical Immunology*. 138 (3): 908 - 911.e9
46. Wang GY, Hu FB, Mistry KB, Zhang CL, Ren FZ, Huo Y, Paige D, Bartell T, Hong XM, Caruso D, **Ji ZC**, Chen Z, Ji YL, Pearson C, **Ji HK**, Zuckerman B, Cheng TL, Wang XB. (2016) Associations of maternal prepregnancy BMI and plasma folate concentrations with child metabolic health. *JAMA Pediatrics*. 170(8):e160845
 47. Wang J, Xia SL, Arand B, Zhu H, Machiraju R, Huang K, **Ji HK**, Qian J (2016) Single-cell co-expression analysis reveals distinct functional modules, co-regulation mechanisms and clinical outcomes. *PLoS Computational Biology*. 12(4): e1004892.
 48. **Ji ZC**, **Ji HK*** (2016) TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Research*. 44(13):e117. doi: 10.1093/nar/gkw430
 49. Mao G, Nachman RM, Sun Q, Zhang X, Koehler K, Chen Z, Hong XM, Wang GY, Caruso D, Zong G, Pearson C, **Ji HK**, Biswal S, Zuckerman B, Wills-Karp M, Wang XB (2016) Individual and Joint Effects of Early-Life Ambient PM2.5 Exposure and Maternal Pre-Pregnancy Obesity on Childhood Overweight or Obesity. *Environmental Health Perspectives*. 125(6):067005
 50. **Zhou WQ**, **Sherwood B**, **Ji HK*** (2016) Global prediction of the functional genomic landscape: applications, methods and challenges. *Human Heredity*. 81(2): 88-105.
 51. Norrie JL, Li Q, Co S, Huang BL, **Ding D**, Uy JC, **Ji ZC**, Mackem S, Bedford MT, Galli A, **Ji HK** and Vokes SA (2016) PRMT5 is essential for the maintenance of chondrogenic progenitor cells in the limb bud. *Development*. 143(24): 4608-4619.
 52. White DT, Eroglu AU, Wang G, Zhang L, Sengupta S, **Ding D**, Rajpurohit SK, Walker SL, **Ji HK**, Qian J, Mumm JS (2016) ARQiv-HTS, a versatile whole-organism screening platform enabling in vivo drug discovery at high-throughput rates. *Nature Protocols* 11:2432–2453
 53. Hong XM, **Sherwood B**, Ladd-Acosta C, Peng S, **Ji H**, Hao K, Burd I, Bartell TR, Wang G, Tsai HJ, Liu X, Ji Y, Wahl A, Caruso D, Lee-Parritz A, Zuckerman B, Wang X (2017) Genome-wide DNA methylation associations with spontaneous preterm birth in US blacks: findings in maternal and cord blood samples. *Epigenetics* doi: 10.1080/15592294.2017.1287654.
 54. White DT, Sengupta S, Saxena MT, Xu Q, Hanes J, **Ding D**, **Ji HK**, Mumm JS (2017) Immunomodulation-accelerated neuronal regeneration following selective rod photoreceptor cell ablation in the zebrafish retina. *Proc. Natl. Acad. Sci. USA*. 114:E3719-E3728.

55. Hong XM, Hao K, **Ji HK**, Peng SN, **Sherwood B**, Narzo AD, Tsai HJ, Liu X, Burd I, Wang GY, Ji YL, Caruso D, Mao GY, Bartell TR, Zhang ZY, Pearson C, Hefferner L, Cerda S, Beaty T, Fallin MD, Lee-Parritz A, Zuckerman B, Weeks DE, Wang XB (2017) Genome-wide approach identified a novel gene-maternal pre-pregnancy BMI interaction on preterm birth. *Nature Communications*. 8:15608
56. **Ji ZC**[§], **Zhou WQ**[§], **Ji HK***. (2017) Single cell regulome data analysis by SCRAT. *Bioinformatics*. 33(18): 2930-2932
57. **Zhou WQ**, **Sherwood B**, **Ji ZC**, Xue Y, **Du F**, **Bai JW**, Ying MY, **Ji HK***. (2017) Genome-wide prediction of DNase I hypersensitivity using gene expression. *Nature Communications*. 8:1038.
58. Martos SN[§], Li T[§], Ramos RB, Lou D, Dai HZ, Xu JC, Gao GL, Gao Y, Wang QL, An C, Zhang XL, Jia YK, Dawson VL, Dawson TM, **Ji HK**, Wang ZB. (2017) Two approaches reveal a new paradigm of ‘switchable or genetics-influenced allele-specific DNA methylation (ASM) with potential in human disease. *Cell Discovery* Accepted
59. **Kuang Z**, **Ji ZC**, Boeke JD*, **Ji HK***. (2017) Dynamic motif occupancy (DynaMO) analysis identifies transcription factors and their binding sites driving dynamic biological processes. *Nucleic Acids Research*. gkx905. Accepted
60. **Kuang Z**, Pinglay S, **Ji HK***, Boeke JD*. (2017) Msn2/4 regulate expression of glycolytic enzymes and control transition from quiescence to growth. *eLife*. 6:e29938
61. Han F, **Ji HK**, **Ji ZC**, Wang HL. (2017) A provable smoothing approach for high dimensional generalized regression with an application in genomics. *Electronic Journal of Statistics*. Accepted
62. Surkan PJ, Dong LM, Ji YL, Hong XM, **Ji HK**, Kimmel M, Tang WY, Wang XB. (2017) Paternal involvement and support and risk of preterm birth: findings from the Boston Birth Cohort. *Journal of Psychosomatic Obstetrics & Gynecology*. Accepted

Articles and Editorials (Not Peer Reviewed)

63. Rudin C, Dunson D, Irizarry R, **Ji HK**, Laber E, Leek J, McCormick T, Rose S, Schafer C, van der Laan M, Wasserman L, Xue L. (2014) Discovery with data: Leveraging statistics with computer science to transform science and society. American Statistical Association White Paper

Thesis

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

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

Book Chapters

66. 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
67. **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
68. **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
69. **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
70. **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
71. **Ji HK** and **Wei YY** (2015) Integrative analysis of multiple ChIP-X data sets using correlation motifs. *Integrating Omics Data*, pp 110-134. Edited by George C. Tseng, Debashis Ghosh and Xianghong J. Zhou. Cambridge University Press.
72. **Wei YY**, Tenzen T and **Ji HK** (2016) Cormotif: An R package for jointly detecting differential gene expression in multiple studies. *Computational Biology and Bioinformatics*, pp 28-47. Edited by Ka-Chun Wong, CRC Press

Book Translation

73. 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

74. 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 Development Biology, San Francisco, CA. *Developmental Biology*, 283(2): 666

75. 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
76. 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
77. 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

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

Submitted Manuscripts

79. **Zhou WQ**, **Ji ZC**, **Ji HK***. Global prediction of chromatin accessibility using RNA-seq from small number of cells. Preprint: bioRxiv doi: <http://dx.doi.org/10.1101/035816>
80. Stephens KE, **Zhou WQ**, He SQ, **Ji ZC**, **Ji HK**, Guan Y, Taverna SD. Sex differences in gene regulation in the dorsal root ganglion after nerve injury
81. Wang J, Zibetti C, Shang P, Sripathi SR, Zhang PW, Cano1 M, **Ji HK**, Merbs SL, Zack DJ, Handa JT, Sinha D, Blackshaw S, Qian J. A widespread decrease of chromatin accessibility in age-related macular degeneration
82. Ji YL, Riley AW, Lee LC, Volk H, Hong XM, Wang GY, Angomas R, Stivers T, Wahl A, **Ji HK**, Bartell TR, Paige D, Fallin MD, Zuckerman B, Wang XB. Do maternal cholesterol levels affect ADHD risk in offspring? A prospective birth cohort study
83. **Zhang BY**, Hong XM, **Ji HK**, Tang WY, Kimmel M, Ji YL, Wang GY, Pearson C, Zuckerman B, Surkan PJ, Wang XB. Maternal smoking during pregnancy and cord blood DNA methylation: new insight on sex differences and effect modification by maternal folate

SOFTWARE AND DATABASE

BIRD: Big data regression for predicting DNase I hypersensitivity using gene expression data.

<https://github.com/WeiqiangZhou/BIRD>

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/>

110,000+ downloads, 700+ citations according to Google Scholar since 2008

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>

5,600+ downloads since 2012

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>

13,800+ downloads since 2012

dPCA: Software for analyzing differential protein-DNA interactions.

<http://www.biostat.jhsph.edu/~hji/dpca/>

GSCA: Gene set context analysis. Using publicly available gene expression data to find new biological contexts of gene set activities.

<https://github.com/zji90/GSCA>

6,900+ downloads since 2014

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>

12,700+ downloads since 2012

JAMIE: Joint analysis of multiple ChIP-chip experiments.

<http://www.biostat.jhsph.edu/~hji/jamie/>

PDDB: Predicted DNase I hypersensitivity database.

<http://jilab.biostat.jhsph.edu/~bsherwo2/bird/index.php>

PolyaPeak: Detecting ChIP-seq peaks using peak shape information.
<http://web1.sph.emory.edu/users/hwu30/software/polyaPeak.html>

SCRAT: Single-cell regulome data analysis (single cell ATAC-seq, ChIP-seq, DNase-seq)
<https://github.com/zji90/SCRAT>

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

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

TSCAN [R/Bioconductor package]: Software tool for single cell RNA-seq analysis
<http://www.bioconductor.org/packages/release/bioc/html/TSCAN.html>
5,600+ downloads since 2014

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	2011-2013	Biostatistics
<u>Current Position:</u> Assistant Professor, University of Victoria, Canada		
Fang Du, PhD	2013-2014	Biostatistics
<u>Current Position:</u> Senior Data Scientist at Covance/LabCorp		
Benjamin Sherwood, PhD	2014-2016	Biostatistics
<u>Current Position:</u> Assistant Professor, University of Kansas		
Weiqiang Zhou, PhD	2013-present	Biostatistics

Advisees (PhD students)

Hao Wu, PhD	2007-2010	Biostatistics
<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>Award:</u> 2010 Margaret Merrell Award		
<u>Current Position:</u> Associate Professor, Emory University		
Yang Ning, PhD	2011-2012	Biostatistics
<u>Thesis:</u> Nonstandard likelihood based inference		
<u>Advisors:</u> Kung-Yee Liang and Hongkai Ji		
<u>Current Position:</u> Assistant Professor, Cornell University		
George Wu, PhD	2007-2013	Biostatistics
<u>Thesis:</u> Statistical methods for the integrative analysis of high-throughput genomic data		
<u>Current Position:</u> Scientist, Crescendo Bioscience		
Yingying Wei, PhD	2011-2014	Biostatistics
<u>Thesis:</u> Integrative statistical models for genomic signal detection		
<u>Awards:</u> 2012 Culley Award, 2013 ENAR Student Travel Award, 2013 Margaret Merrell Award		
<u>Current Position:</u> Assistant Professor, Chinese University of Hong Kong		
Bing He, PhD	2013-2017	Biostatistics

Thesis: FCAT: flexible classification-based signal detection in high-throughput sequencing data

Award: 2012 Kocherlakota Award

Current Position: Data Scientist, Vatic Labs

Dan Jiang, PhD	2013-2017	Biostatistics
<u>Thesis:</u> Adjustment procedure to permutation tests in epigenomic differential analysis		
<u>Current Position:</u> Data Scientist, Drawbridge		
Zhicheng Ji, PhD candidate	2015-present	Biostatistics
<u>Awards:</u> 2014 Kocherlakota Award, 2016 ASA SGG Distinguished Student Paper Award		
Weixiang Fang, PhD candidate	2016-present	Biostatistics
<u>Award:</u> 2015 Kocherlakota Award, 2017 award for outstanding performance in PhD comprehensive exam		
Boyang Zhang, PhD candidate	2017-present	Biostatistics
<u>Award:</u> 2016 Kocherlakota Award		
Yuting Xu, PhD (Academic advisor)	2012-2013	Biostatistics
Jeongyong Kim, PhD (Academic advisor)	2009-2011	Biostatistics
Fang Han, PhD (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> Assistant Professor, Auburn University		
Jiehuan Sun, ScM	2011-2012	Biostatistics
<u>Thesis:</u> Systematic annotation and exploration of Gene Expression Omnibus		
<u>Award:</u> 2011 Kocherlakota Award		
<u>Current Position:</u> PhD student, Biostatistics, Yale University		
Cheng Ran Huang, MHS	2010-2013	Public Health
Wei Wang, MHS	2011-2013	Public Health
Zheng Kuang, MHS	2012-2014	Public Health
Shilu Zhang, ScM	2013-2014	Biostatistics
<u>Thesis:</u> Transcription Factor Binding Motif Analyses in Two Biological Systems		
<u>Award:</u> 2013 Kocherlakota Award		
<u>Current Position:</u> PhD student, Systems Biology, Univ. of Wisconsin at Madison		

Zhicheng Ji, ScM transferred to PhD	2013-2015	Biostatistics
Weixiang Fang, ScM transferred to PhD	2014-2016	Biostatistics
Ding Ding, ScM <u>Thesis:</u> Spanning Tree Representation of High Dimensional Data <u>Current Position:</u> Research Fellow, The Johns Hopkins School of Medicine	2015-2017	Biostatistics
Yun Qing, MHS	2015-2016	Public Health
Boyang Zhang, ScM transferred to PhD	2015-2017	Biostatistics
Yifan Zhou, ScM <u>Thesis:</u> Statistical Methods to Detect Hierarchical Topological Domains in Chromatin <u>Current Position:</u> PhD student, Statistics, George Washington University	2015-2017	Biostatistics
Gege Gui, ScM candidate	2016-present	Biostatistics

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
Ashwini Patil	2013-2014	Bioinformatics
Yu-chi Chang	2016	Computer Science, undergraduate

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
Amanda Price	02/05/2014	Human Genetics
Zheng Kuang	06/09/2014	Human Genetics
Alexandra Maertens	06/17/2014	Environmental Health Sciences
Amanda Price	03/11/2015	Human Genetics
Suzanne Martos	04/21/2015	Environmental Health Sciences
Kipper Fletez-Brant	03/14/2016	Human Genetics
Suzanne Martos	03/27/2017	Environmental Health Sciences

Yuqi Tan	04/11/2017	Biochemistry, Cellular & Molecular Biol.
----------	------------	--

Preliminary Oral Participation

(*: I am an alternate)

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	Epidemiology
Yingying Wei	12/14/2011	Biostatistics
Taeyoung Hwang	11/09/2012	Biomedical Engineering
Man Li	12/20/2013	Epidemiology
Suzanne Martos	04/23/2014	Environmental Health Sciences
Leonardo Collado Torres	05/01/2014	Biostatistics
Bing He	04/24/2015	Biostatistics
Stephen Cristiano	03/27/2017	Biostatistics
Chang Shu	05/02/2017	Mental Health
Yuelong Ji	06/07/2017	Population, Family and Reproductive

Final Oral Participation

(*: I am an alternate)

Hao Wu	05/19/2010	Biostatistics
Tao Wu*	11/24/2010	Epidemiology
Simina Boca*	03/08/2011	Biostatistics
Yang Ning	04/10/2012	Biostatistics
Rahul Karnik	08/08/2012	Biomedical Engineering
George Wu	05/10/2013	Biostatistics
Yingying Wei	03/14/2014	Biostatistics
Alexandra Maertens	09/02/2014	Environmental Health Sciences
Jean-Philippe Fortin*	02/22/2016	Biostatistics
Hailun Liang*	03/01/2016	Health Policy and Management
Bing He	02/08/2017	Biostatistics
Suzanne Martos	08/09/2017	Environmental Health Sciences
Dan Jiang	08/25/2017	Biostatistics

PhD Thesis Reader

Jiehuan Sun	03/2017	Biostatistics, Yale University
-------------	---------	--------------------------------

Master Thesis Reader

Michael Ayars	04/2010	Molecular Microbiology and Immunology
---------------	---------	---------------------------------------

Jingyi Lu

12/2015

Biomedical Engineering

Classroom Instruction

Advanced Methods in Biostatistics III (140.753)	Principal Instructor, 2015-present
2015 Enrollment	9
2016 Enrollment	13
2017 Enrollment	10
Advanced Methods in Biostatistics IV (140.754)	Principal Instructor, 2015-present
2015 Enrollment	8
2016 Enrollment	13
2017 Enrollment	10
Methods in Biostatistics III (140.653)	Principal Instructor, 2013-2014
2013 Enrollment	32
2014 Enrollment	41
Methods in Biostatistics IV (140.654)	Principal Instructor, 2013-2014
2013 Enrollment	35
2014 Enrollment	36
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, 2010,2012,2014
Genomics (260.605)	Guest Lectures, 2009

Visiting Scholar

Hongqiang Lyu

04/2010

Assistant Professor, Xi'an Jiaotong Univ.

Others

1. Advised Yu Gu and Runzhe Li, two summer intern students from the University of Science and Technology of China and Tsinghua University 06/2017 – 09/2017
2. Advised Zhanhao Peng and Haomin Peng, two summer intern students from Tsinghua University 07/2015 - 09/2015
3. Advised Steffen Cornwell, a high school student, for the Center Scholar Program, Center for Talented Youth 06/2014 - 07/2014
4. Advised Alexander Smith, a high school student, for the Center Scholar Program, Center for Talented Youth 07/2012 - 08/2012
5. Taught a five-day summer course *Statistical Computing* for the Hopkins-Nanjing Exchange Program in Statistical and Data Sciences 07/2011
6. Advised Rachel Naphthal from Weston High School, MA for the Center Scholar Program, Center for Talented Youth 06/2009 - 08/2009
7. Offered online tutorial session “Introduction to CisGenome” at the *Affymetrix Workshop at EMBL*. 11/2007

RESEARCH GRANT PARTICIPATION

Principal Investigator Grants

(Current)

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

Principal Investigator: Ji, HK

Dates: 07/25/2012- 04/30/2018

Sponsoring Agency: NIH/NHGRI

Grant No.: R01HG006282

Award Amount: \$ 2,035,000 (Direct cost: 1,250,000)

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: Comparative Functional Genomics Based on Massive Amounts of Publicly Available Omics Data

Principal Investigator: Ji, HK

Dates: 08/2016-12/2017

Sponsoring Agency: Johns Hopkins Catalyst Award

Award Amount: \$75,000

Main Grant Objective: Build a computational framework for comparative functional genomics.

Effort Allocation: 1.2 calendar months (10%)

(Completed)

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

Principal Investigator: Ji, HK

Dates: 09/12/2012- 05/31/2016

Sponsoring Agency: NIH/NHGRI

Grant No.: R01HG006841

Award Amount: \$972,000 (Direct cost: \$600,000)

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: Genome-wide Prediction of DNase I Hypersensitivity and Transcription Factor Binding Sites Based on Gene Expression

Principal Investigator: Ji, HK

Dates: 05/01/2015- 04/30/2016

Sponsoring Agency: JHU/Institute for Data Intensive Engineering and Science

Grant No.: NA

Award Amount: \$ 25,000

Main Grant Objective: The objective of this proposal is to develop a data science approach to predict genome-wide DNase I hypersensitivity (DHS) and transcription factor binding sites (TFBSs) based on gene expression data.

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

Principal Investigator: Ji, HK

Dates: 07/01/2012- 06/29/2014

Sponsoring Agency: TEDCO/Maryland Stem Cell Research Fund

Grant No.: 2012-MSCRF-0135-00

Award Amount: \$ 230,000 (Direct cost: \$200,000)

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%)

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: 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.

(Pending)

Title of Grant: Analytical Infrastructure for Spatial Functional Genomics

Principal Investigator: Ji, HK

Grant Type: NIH R01

Title of Grant: Big Data Methods for Decoding Gene Regulation

Principal Investigator: Ji, HK

Grant Type: NIH R01

Collaborative Grants

(Current)

Title of Grant: Cis-regulatory Circuitry Underlying Hedgehog Mediated Limb Development

Principal Investigator: Vokes, SA

Requested Dates: 07/01/2012- 06/30/2018

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%)

Title of Grant: Prenatal Multi-level Stressors and Alterations in Maternal and Fetal Epigenomes

Principal Investigator: Surkan, PJ

Dates: 07/01/2016 - 06/30/2018

Sponsoring Agency: NIH/NICHD

Grant No.: R21HD085556

Main Grant Objective: This proposed study will investigate a broad spectrum of maternal psychosocial stressors in relation to both mothers' and children's DNA methylation profiles at birth.

Principal Responsibilities of Individual: Co-investigator

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Early Life Determinants of Obesity in U.S. Urban Low Income Minority Birth Cohort

Principal Investigator: Wang, XB

Dates: 01/01/2016 - 12/31/2020

Sponsoring Agency: NIH

Grant No.: 1R01HD086013

Main Grant Objective: This project aims to identify early life risk and protective factors for obesity development.

Principal Responsibilities of Individual: Co-investigator

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Exposure-Altered Gene Expression in Five Candidate Imprinted Loci for Adult Disease

Principal Investigator: Wang, ZB

Dates: 07/01/2016 - 03/31/2021

Sponsoring Agency: NIH/NIEHS

Grant No.: R01ES025761

Main Grant Objective: The goal of this project is to determine how toxicant Bisphenol A changes expression of genes within 17 novel candidate imprinted loci during mouse early embryonic development, therefore contributing to the pathogenesis of diseases in adulthood.

Principal Responsibilities of Individual: Co-investigator

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Preterm Birth, Maternal and Cord Blood Metabolome, and Child Metabolic Risk

Principal Investigator: Wang, XB/Hu, F

Dates: 05/03/2016 - 03/31/2021

Sponsoring Agency: NIH/NICHD

Grant No.: 2R01HD041702

Main Grant Objective: This proposal aims to investigate whether prematurity and maternal and fetal metabolic characteristics can jointly affect the future development of child adverse metabolic outcomes in a well-established U.S. urban low- income minority prospective birth cohort.

Principal Responsibilities of Individual: Co-investigator

Effort Allocation: 0.6 calendar months (5%)

Title of Grant: Chromatin Modifications in GBM-propagating Cells

Principal Investigator: Laterra, J

Dates: 03/01/2017 – 02/28/2022

Sponsoring Agency: NIH

Grant No.: 1R01NS096754-01A1

Main Grant Objective: The goals of this proposal are to identify mechanisms that induce GBM cells to express a GSC state and to develop novel approaches to target these mechanisms in order to inhibit GBM growth and improve survival of GBM- bearing animals.

Principal Responsibilities of Individual: Subcontract PI

Effort Allocation: 1.2 calendar months (10%)

Title of Grant: Molecular Genetics of Mammalian Larynx and Vocal Fold Development

Principal Investigator: Wallingford, JB/Vokes, SA

Requested Dates: 07/10/2017- 04/30/2022

Sponsoring Agency: NIH

Grant No.: 1R01HD090163-01

Main Grant Objective: This proposal seeks to fill a fundamental gap in our understanding of mammalian developmental biology by exploring the molecular genetics and morphogenesis of the mammalian larynx and vocal folds.

Principal Responsibilities of Individual: Subcontract PI

Effort Allocation: 0.6 calendar months (5%)

(Completed)

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: 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: 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: 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: 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: 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: 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%)

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: 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: 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-MSCRFII-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: 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: Genome-wide association study of preterm birth

Principal Investigator: Wang, XB

Dates: 01/01/2012 - 12/31/2015

Sponsoring Agency: NIH

Grant No.: R01HD041702

Main Grant Objective: The goal is to conduct genome-wide association study to identify susceptibility genes of preterm delivery in the Boston Birth Cohort.

Principal Responsibilities of Individual: Co-investigator

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: Modeling Aggressive Medulloblastoma using Human Induced Pluripotent Stem Cells

Principal Investigator: Ying, MY

Dates: 09/30/2015 - 09/29/2016

Grant No.: KKI subcontract 110866

Main Grant Objective: Computational analyses of the high-throughput genomic data for studying medulloblastoma using human iPS cells.

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: 1.2 calendar months (10%)

Title of Grant: Highly Efficient Conversion of Human iPS Cells to Dopaminergic Neurons by Synthetic Modified mRNAs

Principal Investigator: Ying, MY

Dates: 07/01/2015 - 06/30/2017

Grant No.: KKI subcontract 613044

Main Grant Objective: Analyze high-throughput genomic data for studying efficient conversion of iPS cells to dopaminergic neurons.

Principal Responsibilities of Individual: Subcontract PI

Effort Allocation: 0.6 calendar months (5%)

ACADEMIC SERVICE

Department of Biostatistics

Director, Graduate Program & Admission Committee (2017 - present)

Co-director, Graduate Program & Admission Committee (2015 - 2017)

Member, Graduate Program & Admission Committee (2010 - present)

Member, Postdoctoral Fellow Committee (2007)

Member, Faculty Search Committee (2008)

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 Talks

Global prediction of chromatin accessibility using RNA-Seq from small number of cells. *JSM, Baltimore, MD, Aug 2, 2017*

Global prediction of chromatin accessibility using RNA-Seq from small number of cells. *2017 IMS China International Conference on Statistics and Probability, Nanning, Guangxi, China, June 30, 2017*

Global prediction of chromatin accessibility using RNA-Seq from single cell and small number of cells. *The 26th ICSA Applied Statistics Symposium, Chicago, IL, June 27, 2017*

Decoding gene regulation using single cell genomic data. *The Graybill Conference XIV on Statistical Genomics and Genetics, Fort Collins CO, June 6, 2017*

Single cell RNA-seq analysis by spanning trees. *The 31st New England Statistics Symposium, Storrs, CT, Apr 22, 2017*

Global prediction of chromatin accessibility using RNA-Seq from small number of cells. *ENAR, Washington DC, Mar 13, 2017*

Pseudo-time Reconstruction and Evaluation in Single-cell RNA-seq Analysis. *The 10th ICSA International Conference, Shanghai, China, Dec 19, 2016*

Genome-wide prediction of DNase I hypersensitivity using gene expression. *The 10th Annual Symposium and Poster Session on Genomics and Bioinformatics, JHU, Baltimore, MD, Oct 13, 2016*

Epigenome isoform analysis with applications. *JSM, Chicago, IL, Aug 3, 2016*

Big data regression and prediction in functional genomics. *The 4th IMS Asia Pacific Rim Meeting, Hong Kong, Jun 29, 2016*

Big data regression for predicting genome-wide functional genomic signals. *The 25th ICSA Applied Statistics Symposium, Atlanta, GA, Jun 13, 2016*

Big data regression for predicting genome-wide regulatory element activities. *Johns Hopkins Institute for Data Intensive Engineering and Science (IDIIES) Annual Symposium, Baltimore, MD, Oct 16, 2015*

A big data approach for integrating different functional genomic data types. *JSM, Seattle, WA, Aug 9, 2015*

Big data methods for dissecting variations in high-throughput genomic data. *Symposium on Advances in Genomics, Epidemiology, and Statistics (SAGES), Philadelphia, PA, May 29, 2015*

A big data method for decomposing variations in high-throughput genomic data. *ENAR, Miami, FL, Mar 18, 2015*

A big data method for decomposing variations in high-throughput genomic data. *The Third Joint Biostatistics Symposium, Chengdu, China, June 28, 2014*

A model-based approach for characterizing correlation in big genomic data. *The Annual Meeting of Statistical Society of Canada, Toronto, ON, Canada, May 27, 2014*

Big data methods for dissecting variations in high-throughput genomic data. *Tsinghua Chair Professors' Bioinformatics Workshop, Beijing, China, May 23, 2014*

Statistical analysis of time-course ChIP-seq data. *ENAR, Baltimore, MD, Mar 17, 2014*

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*

Invited Seminars

Big data regression for predicting functional genomic signals using gene expression. *University of Notre Dame, Notre Dame, IN, Mar 20, 2017*

Genome-wide prediction of DNase I hypersensitivity using gene expression. *University of Virginia, Charlottesville, VA, Nov 11, 2016*

Genome-wide prediction of DNase I hypersensitivity using gene expression. *MOFFITT Cancer Center, Tampa, FL, Sept 8, 2016*

Genome-wide prediction of DNase I hypersensitivity using gene expression. *UCLA, Los Angeles, CA*, May 16, 2016

Computational analysis of genome-wide chromatin immunoprecipitation data. *Brady Urological Institute, Johns Hopkins University, Baltimore, MD*, July 13, 2015

ChIP-seq: Unleashing its full potential through data integration. *University of South California, Los Angeles*, Sept 18, 2014

Crowd intelligence: a solution to new challenges in education? *The Institute of Economics of Education, Peking University, Beijing, China*, July 18, 2014

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.

Contributed Talks

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

Posters

TSCAN: pseudo-time reconstruction and evaluation in single cell RNA-seq data. *ENAR, Austin, TX*, Mar 6, 2016

Global prediction of chromatin accessibility using RNA-seq in samples with small number of cells. *CSHL Systems Biology: Global Regulation of Gene Expression Meeting, Cold Spring Harbor, NY*, Mar 16, 2016

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 big and complex data, particularly high-throughput genomic data. I apply these tools to study gene regulatory programs in development and diseases. My research topics include:

- 1) *Analytical methods and software tools for high-throughput genomic data.* I develop statistical and computational methods for analyzing data from ChIP-seq, DNase-seq, ATAC-seq, RNA-seq, microarrays, tiling arrays, and other high-throughput sequencing and genomic technologies. I have created several popular methods and software tools for analyzing data from genome-wide chromatin immunoprecipitation (i.e. ChIP-seq and ChIP-chip) experiments, including **CisGenome** (Ji et al. *Nat Biotechnol.* 2008, 700+ citations), **TileMap** (Ji & Wong, *Bioinformatics*, 2005, 270+ citations), and **hmChIP** (Chen et al. *Bioinformatics*, 2011). I have also developed methods and tools for analyzing gene expression data (Wei et al., *Biostatistics*, 2015; Ji et al. *Nucleic Acids Res.* 2016), DNA sequence motifs (Ji et al. *Nat Biotechnol.* 2008; Ji et al. *Nucleic Acids Res.* 2006), TIP-chip data (Huang et al., *Cell*, 2010), and whole genome sequences (Hellmann et al. *Genome Res*, 2005; The Chimpanzee Sequencing and Analysis Consortium, *Nature*, 2005).
- 2) *Statistical and computational methods for data integration and data mining.* I develop scalable methods and tools for integrating high-throughput genomic data from multiple sources and mining massive amounts of data in public databases. Examples include differential Principal Component Analysis (**dPCA**) (Ji et al. *PNAS*, 2013), the correlation motif approach for scalable data integration (**CorMotif**) (Wei et al., *Biostatistics*, 2015), Gene Set Context Analysis (**GSCA**) for turning massive amounts of publicly available gene expression data into new discoveries (Ji et al. *Nucleic Acids Res.* 2016), and joint analysis of protein-DNA interactions (**JAMIE**) (Wu & Ji, *Bioinformatics*, 2010) and allele-specific events in multiple datasets (**iASeq**) (Wei et al. *BMC Genomics*, 2012). Most recently, we have used the massive amounts of data from the ENCODE project and the Gene Expression Omnibus (GEO) database to develop models that use gene expression data to predict genome-wide regulatory element activities. We developed big data regression (**BIRD**) to solve this challenging problem where both predictors and responses are ultra-high-dimensional (Zhou et al. *Nat Commun*, 2017, bioRxiv: <https://doi.org/10.1101/035808>).
- 3) *Single-cell genomics.* Recently, I started to work on single-cell genomics. My team has developed a cluster-based minimum spanning tree approach **TSCAN** for reconstructing cells' pseudo-temporal trajectories using single-cell RNA-seq data as well as methods for objectively evaluating computationally reconstructed pseudotime (Ji & Ji, *Nucleic Acids Res.* 2016). We have also developed **SCRAT**, the first integrated analysis tool with a graphical user interface, for analyzing single-cell ATAC-seq, single-cell ChIP-seq and single-cell DNase-seq data (Ji et al. *Bioinformatics*, 2017). Additionally, we demonstrate the feasibility to use single-cell RNA-seq data to predict genome-wide activities of gene regulatory elements (Zhou et al. bioRxiv: <https://doi.org/10.1101/035816>). Our ongoing research includes

development of new statistical methods for better signal extraction from single-cell genomic data and integrative analysis of multiple single-cell genomic data types.

- 4) *Gene regulation.* I closely collaborate with biomedical investigators to apply our methods and tools to decode gene regulatory circuitry in development and diseases. I have made contributions to elucidate regulatory networks in the Sonic hedgehog (Shh) signaling pathway (Vokes et al. *Development*, 2007; Vokes et al. *Genes Dev*, 2008; Lee et al. *PNAS*, 2010; Lewandowski et al. *Dev Biol*, 2015; Li et al. *J Biol Chem*, 2016), identify cell-type dependent and independent targets of MYC (Ji et al., *PLoS ONE*, 2011), and unravel regulatory programs in stem cell differentiation (Niakan et al. *Genes Dev*, 2010; Wamaitha *Genes Dev*, 2015), metabolic cell cycles (Kuang et al. *Nat Struct Mol Biol*, 2014) and cancers (Paik et al. *Cell*, 2007; Wang et al. *Mol Cell*, 2013; Ying et al. *J Biol Chem*, 2014; Jin et al. *Cancer Discovery*, 2015). Recently, my research team (adbc) have participated in the ENCODE-DREAM *in vivo* Transcription Factor Binding Site Prediction Challenge and ranked no. 4 among a total of 50 submitted solutions from 40 international teams.
- 5) *Genomics in public health.* In collaboration with epidemiologists, we combine high-throughput genomic approaches with epidemiology cohorts to investigate genetic, epigenetic and environmental factors that may influence human health. As an investigator in the Boston Birth Cohort (BBC), I contributed to the identification of genetic, epigenetic and environmental risk factors and their interactions that are associated with preterm birth (Hong et al. *Nat Commun*, 2017; Hong et al. *Epigenetics*, 2017), allergy (Hong et al. *J Allergy Clin Immunol*, 2016) and obesity (Wang et al. *JAMA Pediatr*, 2016; Mao et al. *Environ Health Perspect*, 2016). My previous research in genetic association studies of schizophrenia has also identified disease associating genetic variants in the Chinese Han population (Chinese Schizophrenia Consortium, *Mol Psychiatry*, 2004).

My long term goal is to build effective and efficient strategies for data collection, analysis, integration and interpretation, and to turn the massive amounts of data we have into new knowledge and discoveries to improve human health.

Keywords

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

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