

HAO WU

Last updated June 7, 2022

<i>Mailing address</i>	Department of Biostatistics and Bioinformatics Rollins School of Public Health Emory University 1518 Clifton Rd., NE Atlanta, GA 30322
<i>Phone</i>	(404)727-8633
<i>Email</i>	hao.wu@emory.edu
<i>Website</i>	http://haowulab.org

Education

- **B.S. Electrical Engineering**, 1996, Tsinghua University, Beijing, China.
- **M.S. Electrical Engineering**, 2000, Iowa State University, Ames, Iowa.
Minor in Economics.
- **MHS Bioinformatics**, 2010. Johns Hopkins University, Baltimore, MD.
- **Ph.D. Biostatistics**, 2010. Johns Hopkins University, Baltimore, MD.
Advisors: Rafael A. Irizarry, Hongkai Ji.

Professional Experience

- **July 1996–July 1998, Research Assistant**, Institute of Electrical Engineering, Chinese Academy of Science, Beijing, China.
- **February 2000–July 2001, Staff Engineer**, GE-Harris Energy Control Systems (Now GE Network Solution), Melbourne, Florida.
- **July 2001–August 2005, Scientific Software Engineer**, The Jackson Laboratory, Bar Harbor, Maine.
- **July 2010–August 2016, Assistant Professor**, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia.
- **August 2011–present, Affiliated faculty**, Graduate Program in Population Biology, Ecology, & Evolution (PBEE), Emory University, Atlanta, Georgia
- **September 2016–present, Associate Professor (with tenure)**, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia.
- **September 2016–present, Associated Associate Professor**, Department of Computer Science, Emory University, Atlanta, Georgia.
- **May 2022–present, Professor**, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia.

Research Interests

- Statistical Methodology: bioinformatics, computational biology, high-throughput omics data, biomarker discovery, machine learning and large-scale data mining.
- Applications: neurodegenerative diseases (Alzheimer’s Disease, Fragile X syndrome), cancers, cell differentiation/development/reprogramming.

Publications

(* corresponding author; §: equal contribution; underline: lab member, advisee, or student under my guidance.)

Google scholar citation statistics (as of December 2021): All citations: 13500+. h-index: 44. i10-index: 72.

Peer-Reviewed Journal - methodology

1. Broman KW, **Wu H**, Sen S, Churchill GA (2003) *R/qtl: QTL mapping in experimental crosses*. **Bioinformatics** 19:889-890.
2. Yandell BS, Mehta T, Banerjee S, Shriner D, Venkataraman R, Moon JY, Neely WW, **Wu H**, von Smith R, Yi N (2007) *R/qtlbim: QTL with Bayesian Interval Mapping in experimental crosses*. **Bioinformatics** 23(5):641-643.
3. Frangakis CE, **Wu H** (2007) *The geometry of inadmissibility of independent observations for estimating a single parameter in two-parameter ordered symmetric problems*. **Metron** Vol. LXV n.3.
4. Irizarry RA, **Wu H**, Feinberg AP (2009) *A species-generalized probabilistic model-based definition of CpG islands*. **Mammalian Genome** 20:674–680.
5. **Wu H**, Jaffe HA, Feinberg AP, Irizarry RA (2010) *Redefining CpG Islands Using a Hidden Markov Model*. **Biostatistics**, 11(3): 499–514.
6. **Wu H**, Irizarry, R.A. and Bravo, H.C. (2010) *Intensity normalization improves color calling in SOLiD sequencing*, **Nature Method** 7:336–337.
7. **Wu H**, Ji H (2010) *JAMIE: Joint analysis of multiple ChIP-chip experiments*. **Bioinformatics** 26:1864–1870.
8. **Wu H**, Wang C, Wu Z (2012) *A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data*. **Biostatistics**, 14(2):232-43. doi: 10.1093/biostatistics/kxs033.
9. **Wu H***, Qin ZS* (2013) *Exploring the Cooccurrence Patterns of Multiple Sets of Genomic Intervals*. **BioMed Research International**. 2013:617545. doi: 10.1155/2013/617545.
10. **Wu H***, Ji HK* (2014) *PolyaPeak: detecting transcription factor binding sites from ChIP-seq using peak shape information*. **PLoS One**. 9(3): e89694. doi:10.1371/journal.pone.0089694.
11. Feng H, Conneely KN*, **Wu H*** (2014) *A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data*. **Nucleic Acids Research**. 42(8):e69. doi: 10.1093/nar/gku154.
12. **Wu H***, Wang C, Wu Z* (2014) *PROPER: Comprehensive Power Evaluation for Differential Expression using RNA-seq*. **Bioinformatics**. doi: 10.1093/bioinformatics/btu640.
13. Chen L, Wang C, Qin Z, **Wu H*** (2015) *A novel statistical method for quantitative comparison of multiple ChIP-seq datasets*. **Bioinformatics**. doi: 10.1093/bioinformatics/btv094.
14. Xu T, Li B, Zhao M, Szulwach LE, Street RC, Lin L, Yao B, Jin P*, **Wu H***, Qin Z* (2015) *Base-resolution methylation patterns accurately predict transcription factor bindings in vivo*. **Nucleic Acids Research** 43(5), 2757-2766. doi: 10.1093/nar/gkv151.
15. Zhang N, Wu H-J, Zhang W, Wang J, **Wu H***, Zheng X* (2015) *Predicting tumor purity from methylation microarray data*. **Bioinformatics** 31(21), 3401-3405. doi: 10.1093/bioinformatics/btv370.

16. **Wu H***, Xu T, Feng H, Chen L, Li B, Yao B, Qin S, Jin P, Conneely KN (2015) *Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicate*. **Nucleic Acids Research**. doi: 10.1093/nar/gkv715.
17. Chong E, Huang Y, **Wu H**, Ghasemzadeh N, Uppal K, Quyyumi A, Jones D, Yu, T. (2015) *Local false discovery rate estimation using feature reliability in LC/MS metabolomics data*. **Scientific Report** 5:17221. doi: 10.1038/srep17221.
18. Park Y*, **Wu H*** (2016) *Differential methylation analysis for BS-seq data under general experimental design*. **Bioinformatics** 32 (10), 1446-1453. doi:10.1093/bioinformatics/btw026.
19. Wei Y*, **Wu H*** (2016) *Measuring the spatial correlations of protein binding sites*. **Bioinformatics**. doi:10.1093/bioinformatics/btw058.
20. Qin Z, Li B, Conneely KN, **Wu H**, Hu M, Ayyala D, Park Y, Jin VX, Zhang F, Zhang H, Li L, Lin S (2016) *Statistical challenges in analyzing methylation and long-range chromosomal interaction data*. **Statistics in Biosciences**. doi: 10.1007/s12561-016-9145-0.
21. Wang F, Zhang N, Wang J, **Wu H***, Zheng X* (2016) *Tumor purity and differential methylation in cancer epigenomics*. **Briefings in Functional Genomics**. doi: 10.1093/bfpg/elw016.
22. Liao P, **Wu H***, Yu T*. (2016) *ROC Curve Analysis in the Presence of Imperfect Reference Standards*. **Statistics in Biosciences**. DOI: 10.1007/s12561-016-9159-7.
23. Wang H, Horbinski C, **Wu H**, Liu Y, Sheng S, Liu J, Weiss H, Stromberg AJ, and Wang C. (2016) *NanoStringDiff: A Novel Statistical Method for Differential Expression Analysis Based on NanoString nCounter Data*. **Nucleic Acids Research**. doi: 10.1093/nar/gkw677.
24. Zheng X*, Zhang N, Wu HJ, **Wu H***. (2017) *Estimating and accounting for tumor purity in cancer methylation microarray analysis*. **Genome Biology** 18:17. doi: 10.1186/s13059-016-1143-5.
25. Hong C, Ning Y, Wang S, **Wu H**, Carroll RJ, Chen Y*. (2017) *PLEMT: A novel pseudolikelihood based EM test for homogeneity in generalized exponential tilt mixture models*. **J. Am. Statist. Assoc.** doi:10.1080/01621459.2017.1280405.
26. Wang Y, **Wu H***, Yu T*. (2017) *Differential Gene Network Analysis from Single Cell RNA-Seq*. **Journal of Genetics and Genomics**. doi:10.1016/j.jgg.2017.03.001.
27. Zhang W, Feng H, **Wu H***, Zheng X*. (2017) *Accounting for Tumor purity improves cancer subtype classification from DNA methylation data*. **Bioinformatics**. doi: 10.1093/bioinformatics/btx303.
28. Feng H, Jin P, **Wu H***. (2018) *Disease Prediction by Cell-Free DNA Methylation*. **Briefings In Bioinformatics**. doi:10.1093/bib/bby029.
29. Wu Z*, Zhang Y, Stitzel ML, **Wu H***. (2018) *Two-phase differential expression analysis for single cell RNA-seq*. **Bioinformatics**. doi: 10.1093/bioinformatics/bty329.
30. Xu T, Zheng X, Li B, Jin P, Qin Z, **Wu H***. (2018) *A comprehensive review of computational prediction of genome-wide features*. **Briefings In Bioinformatics**. doi: 10.1093/bib/bby110.
31. Li Z, Wu Z, Jin P, **Wu H***. (2019) *Dissecting differential signals in high-throughput data from complex tissues*. **Bioinformatics**. doi: 10.1093/bioinformatics/btz196.

32. Li Z, **Wu H***. (2019) *TOAST: improving reference-free cell composition estimation by cross-cell type differential analysis*. **Genome Biology**. **20** (1), 1-17
33. Feng H, **Wu H***. (2019) *Differential methylation analysis for bisulfite sequencing using DSS*. **Quantitative Biology**. doi:10.1007/s40484-019-0183-8.
34. Li Z*, Guo Z, Cheng Y, Jin P, **Wu H***. (2020) *Robust partial reference-free cell composition estimation from tissue expression*. **Bioinformatics**. doi:10.1093/bioinformatics/btaa184.
35. Wu Z*, **Wu H**. (2020) *Accounting for cell-type hierarchy in evaluating single cell RNA-seq clustering*. **Genome Biology**. **21** (1), 1-14
36. Su K, Wu Z, **Wu H***. (2020) *Simulation, Power Evaluation, and Sample Size Recommendation for Single Cell RNA-seq*. **Bioinformatics**. doi: 10.1093/bioinformatics/btaa607.
37. Shi M, Tsui SKW, **Wu H** and Wei Y*. (2020) *Pan-cancer analysis of differential DNA methylation patterns*. **BMC Medical Genomics**. doi: 10.1186/s12920-020-00780-3.
38. Zhang W, **Wu H***, Li Z*. (2020) *Complete deconvolution of DNA methylation signals from complex tissues: a geometric approach*, **Bioinformatics**. DOI: 10.1093/bioinformatics/btaa930.
39. Su K, Yu T, **Wu H***. (2021) *Accurate Feature Selection Improves Single Cell RNA-seq Cell Clustering*, **Briefings in Bioinformatics**. doi:10.1093/bib/bbab034.
40. Guo Z, Shafik AM, Jin P, Wu Z, **Wu H***. (2021) *Detecting m⁶A methylation regions from Methylated RNA Immunoprecipitation Sequencing*, **Bioinformatics**. DOI: 10.1093/bioinformatics/btab181.
41. Wu X, **Wu H**, Wu Z*. (2021) *Penalized Latent Dirichlet Allocation Model in Single Cell RNA Sequencing*, **Statistics in Biosciences**. 1–20.
42. Wu Z*, Su K, **Wu H**. (2021) *Non-linear Normalization for Non-UMI Single Cell RNA-Seq*, **Frontiers in Genetics**. 12:612670. doi: 10.3389/fgene.2021.612670.
43. Ma W, Su K, **Wu H***. (2021) *Evaluation of some aspects in supervised cell type identification for single-cell RNA-seq: classifier, feature selection, and reference construction*, **Genome Biology**, <https://doi.org/10.1186/s13059-021-02480-2>.
44. Sun X, Lin X, Li Z, **Wu H***. (2022) *A comprehensive comparison of supervised and unsupervised methods for cell type identification in single cell RNA-seq*, **Briefings in Bioinformatics**. doi: 10.1093/bib/bbab567.
45. Wei X, Li Z, Ji H, **Wu H***. (2022) *EDClust: An EM-MM hybrid method for cell clustering in multiple-subject single-cell RNA sequencing*, **Bioinformatics**. doi: 10.1093/bioinformatics/btac168.

Peer-Reviewed Journal - project-targeted methodology

Collaborative researches that require methodological development and extensive data analyses

46. Wen B[§], **Wu H[§]**, Bjornsson H, Green RD, Irizarry RA, Feinberg AP (2008) *Overlapping euchromatin/heterochromatin marks are enriched in imprinted gene regions and predict allele-specific modification*. **Genome Research** 18(11):1806–1813.
47. Wen B, **Wu H**, Irizarry RA, Shinkai Y, Feinberg AP (2009) *Large histone H3 lysine-9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells*. **Nature Genetics** 41:246–250. Cover story.

48. McDonald OG, **Wu H**, Timp W, Doi A, Feinberg AP (2011), *Genome-scale epigenetic reprogramming during epithelial to mesenchymal transition*. **Nature Structural & Molecular Biology** 18(8):867-74.
49. **Wu H**, Wu MC, Zhi D, Santorico SA, Cui X (2012). *Statistical analysis for next generation sequencing – meeting report*. **Frontiers in Genetics**, 3:128.
50. Wen B[§], **Wu H**[§], Loh Y[§], Briem E, Daley GQ, Feinberg AP (2012) *Euchromatin islands in large heterochromatin domains are enriched for CTCF binding and differentially DNA-methylated regions*. **BMC Genomics**, 13:566. doi: 10.1186/1471-2164-13-566.
51. Wang T, **Wu H**, Li Y, Szulwach KE, Lin L, Li X, Chen P, Goldlust IS, Chamberlain SJ, Dodd A, Gong H, Ananiev G, Han JW, Yoon Y, Rudd MK, Yu M, Song CX, He C, Chang Q, Warren ST, Jin P (2013) *Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency*. **Nature Cell Biology**. 15(6):700-11. doi: 10.1038/ncb2748.
52. Tilghman J, **Wu H**, Sang Y, Shi X, Guerrero-Cazares H, Quinones-Hinojosa A, Eberhart C, Laterra J, Ying M (2014) *A HMMR maintains the stemness and tumorigenicity of glioblastoma stem-like cells*. **Cancer Research**. 74(11):3168-79. .
53. Fu Y, Lv P, Yan G , Fan H, Cheng L, Zhang F, Dang Y, **Wu H**, Wen B. (2015) *MacroH2A1 associates with nuclear lamina and maintains chromatin architecture in mouse liver cells*. **Scientific Report** 5:17186. doi:10.1038/srep17186.
54. Yao B, Cheng Y, Wang Z, Li Y, Chen L, Huang L, Zhang W, Chen D, **Wu H**, Tang B, Jin P (2017) *DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress*. **Nature Communication**. 8(1):1122. doi:10.1038/s41467-017-01195-y.
55. Fan H, Lv P, Huo X, Wu J, Wang Q, Cheng L, Liu Y, Tang Q, Zhang L, Zhang F, Zheng X, **Wu H**, Wen B (2018) *The nuclear matrix protein HNRNPU maintains 3D genome architecture globally in mouse hepatocytes*. **Genome Research**. doi: 10.1101/gr.224576.117.
56. Cheng Y, Li Z, Manupipatpong S, Lin L, Li X, Xu T, Jiang YH, Shu Q, **Wu H**, Jin P (2018) *5-hydroxymethylcytosine alterations in the human postmortem brains of autism spectrum disorder*. **Human Molecular Genetics**. doi: 10.1093/hmg/ddy193.
57. Yao B, Li Y, Wang Z, Chen L, Poidevin M, Zhang C, Lin L, Wang F, Bao H, Jiao B, Lim J, Cheng Y, Huang L, Phillips BL, Xu T, Duan R, Moberg KH, **Wu H**, Jin P (2018) *Active N⁶-Methyladenine Demethylation by DMAD Regulates Gene Expression by Coordinating with Polycomb Protein in Neurons*. **Molecular Cell**. doi: 10.1016/j.molcel.2018.07.005.

Peer-Reviewed Journal - collaborative

58. Irizarry RA, Ladd-Acosta C, Carvalho B, **Wu H**, Brandenburg SA, Wen B, Feinberg AP (2008) *Comprehensive high-throughput arrays for relative methylation (CHARM)*. **Genome Research** 18(5):780-790.
59. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, **Wu H**, Briem E, Irizarry RA, Feinberg AP (2011) *Generalized Loss of Stability of Epigenetic Domains Across Cancer Types*. **Nature Genetics** 26;43(8):768-75.
60. Szulwach KE, Li X, Li Y, Song CX, **Wu H**, Dai Q, Irier H, Upadhyay AK, Gearing M, Levey AI, Vasanthakumar A, Godley LA, Chang Q, Cheng X, He C, Jin P (2011) *5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging*. **Nature Neuroscience** 14:1607-1616, doi:10.1038/nn.2959.

61. Lu C, Lin L, Tan H, **Wu H**, Sherman SL, Gao F, Jin P, Chen D (2012) *Fragile X premutation RNA is sufficient to cause primary ovarian insufficiency in mice.* **Human Molecular Genetics**, **21**(23):5039-47. doi: 10.1093/hmg/dds348.
62. Wang T, Pan Q, Lin L, Szulwach KE, Song C, He C **Wu H**, Warren ST, Jin P, Duan R, Li X (2012) *Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum.* **Human Molecular Genetics**, **21**(26):5500-10. doi: 10.1093/hmg/dds394.
63. Song C, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin, M, **Wu H**, Gao J, Liu P, Li L, Xu GL, Jin P, He C (2013) *Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming.* **Cell**. **153**(3):678-91. doi: 10.1016/j.cell.2013.04.001.
64. Kusters A, Sun D, **Wu H**, Tian F, Felix JC, Li W, Karpen SJ (2013) *Sexually Dimorphic Genome-Wide Binding of Retinoid X Receptor alpha (RXR α) Determines Male-Female Differences in the Expression of Hepatic Lipid Processing Genes in Mice.* **PLoS One**. **8**(8):e71538. doi: 10.1371/journal.pone.0071538.
65. Bai B, Hales CM, Chen PC, Gozal Y, Dammer EB, Fritz JJ, Wang X, Xia Q, Duong DM, Street RC, Cantero G, Cheng D, Jones DR, Wu Z, Li Y, Diner I, Heilman CJ, Rees HD, **Wu H**, Lin L, Szulwach KE, Gearing M, Mufson EJ, Bennett DA, Montine TJ, Seyfried NT, Wingo TS, Sun YE, Jin P, Hanfelt J, Willcock DM, Levey A, Lah JJ, Peng J (2013) *U1 Small Nuclear Ribonucleoprotein Complex and RNA Splicing Alterations in Alzheimer's Disease.* **Proc Natl Acad Sci**. **8**;110(41):16562-7. doi: 10.1073/pnas.1310249110.
66. Singh AM, Chappell J, Trost R, Lin L, Wang T, Tang J, **Wu H**, Zhao S, Jin P, Dalton S (2013), *Cell-Cycle Control of Developmentally Regulated Transcription Factors Accounts for Heterogeneity in Human Pluripotent Cells.* **Stem Cell Reports**. **1**(6): 532–544.
67. Yao B, Lin L, Street RC, Zalewski ZA, Galloway JN, **Wu H**, Nelson DL, Jin P (2014) *Genome-wide alteration of 5-hydroxymethylcytosine in a mouse model of fragile X-associated tremor/ataxia syndrome.* **Human Molecular Genetics**. **23**(4):1095-107. doi: 10.1093/hmg/ddt504.
68. Potts LF, **Wu H**, Singh A, Marcilla I, Luquin MR, Papa SM (2014) *Modeling Parkinson's disease in monkeys for translational studies, a critical analysis.* **Experimental neurology**. **256**, 133-143.
69. Sun X, Fu X, Li J, Xing C, Frierson HF, **Wu H**, Ding X, Ju T, Cummings RD, Dong JT (2014) *Deletion of atbf1/zfx3 in mouse prostate causes neoplastic lesions, likely by attenuation of membrane and secretory proteins and multiple signaling pathways.* **Neoplasia**. **16**(5):377-89. doi: 10.1016/j.neo.2014.05.001.
70. Lapp S, Mok S, Lei Z, **Wu H**, Preiser P, Bozdech ZB, Galinski M (2015) *Plasmodium knowlesi gene expression differs in ex vivo compared to in vitro blood-stage cultures.* **Malaria Journal**. **14** (1), 110.
71. Halliley JL, Tipton C, Liesved J, Rosenber AF, Darce J, Gregoret IV, Papova L, Kaminiski D, Fucile C, Albizua-Santin I, Kyu S, Chiang KY, Bradley K, Burack R, Slifka M, Hammarlund E, **Wu H**, Zhao L, Walsh EE, Flasey AR, Randall TD Cheun WC, Sanz I, Lee FE (2015) *Identification of Long-lived Plasma Cells in Human Bone Marrow.* **Immunity**. doi:10.1016/j.immuni.2015.06.016.
72. Archer KJ, Dobbin K, Biswas S, Day RS, Wheeler DC, **Wu H** (2015) Computer Simulation, Bioinformatics, and Statistical Analysis of Cancer Data and Processes. **Cancer Informatics** **14**(Suppl 2):247-51.

73. Zeng Y, Yao B, Shin J, Lin L, Kim N, Song Q, Liu S, Su Y, Guo JU, Huang L, Wan J, **Wu H**, Qian J, Cheng X, Zhu H, Ming GL, Jin P, Song H. (2015) *Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression*. **Molecular Cell** 61 (1), 153-160 (cover story). doi: <http://dx.doi.org/10.1016/j.molcel.2015.11.020>.
74. Qian X, Nguyen HN, Song MM, Hadiono C, Ogden SC, Hammack C, Yao B, Hamersky G, Jacob F, Zhong C, Yoon K, Jeang W, Li L, Li Y, Thakor J, Berg D, Zhang C, Kang E, Chickering M, Naeun D, Wen Z, Christian KM, Ho CY, Shi PY, Maher B, **Wu H**, Jin P, Tang H, Song H and Ming GL. (2016) *Brain Region-specific Organoids using a Miniaturized Spinning Bioreactor and Modeling of Zika Virus Exposure*. **Cell** 165 (5), 1238-1254.
75. Zhang F, Hammack C, Ogden SC, Cheng Y, Lee EM, Wen Z, Qian X, Nguyen HN, Li Y, Yao B, Xu M, Xu T, Chen L, Wang Z, Feng H, Huang WK, Yoon KJ, Shan C, Huang L, Qin Z, Christian KM, Shi PY, Xu M, Xia M, Zheng W, **Wu H**, Song H, Tang H, Ming GL, Jin P. (2016) *Molecular signatures associated with ZIKV exposure in human cortical neural progenitors*. **Nucleic Acids Research**. doi: 10.1093/nar/gkw765.
76. Rao A, Kosters A, Mells JE, Zhang W, Setchell KDR, Amanso AM, Xu T, Wynn GM, Keller BT, Yin H, Banton S, Jones DP, **Wu H**, Dawson PA, Karpen SJ. (2016) *Inhibition of Ileal Bile Acid Uptake Protects Against Non-alcoholic Fatty Liver Disease in High Fat Diet-fed Mice*. **Science Translational Medicine**. Vol. 8, Issue 357, pp. 357ra122. doi: 10.1126/scitranslmed.aaf4823.
77. McDonald OG, Li X, Saunders T, Tryggvadottir R, Mentch SJ, Warmoes MO, Word AE, Carrer A, Salz TH, Natsume S, Stauffer KM, Makohon-Moore A, Zhong QY, **Wu H**, Wellen KE, Locasale JW, Iacobuzio-Donahue C, Feinberg AP. (2017) *Large-scale epigenetic reprogramming is punctuated late during the evolution of pancreatic cancer progression*. **Nature Genetics** 49, 367–376. doi:10.1038/ng.3753.
78. Zhao J, Zhu Y, Yang J, Li L, **Wu H**, De Jager PL, Jin P, Bennett DA. (2017) *A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease*. **Alzheimer's & Dementia**. doi: 10.1016/j.jalz.2016.10.004.
79. Li X, Yao B, Chen L, Kang Y, Li Y, Cheng Y, Li L, Wang Z, Wang M, Pan F, Dai Q, Zhang W, **Wu H**, Shu Q, Qin Z, He C, Xu M and Jin P. (2017) *Ten-eleven translocation 2 interacts with Forkhead box O3 and regulates adult neurogenesis*. **Nature Communication**. doi: 10.1038/ncomms15903.
80. Li L, Zang L, Zhang F, Chen J, Shen H, Shu L, Liang F, Feng C, Chen D, Tao H, Xu T, Li Z, Kang Y, **Wu H**, Tang L, Zhang P, Jin P, Shu Q, Li X. (2017) *Fat mass and obesity-associated (FTO) protein regulates adult neurogenesis*. **Human Molecular Genetics**. doi: 10.1093/hmg/ddx128.
81. Wingo TS, Duc M, Duong DM, Zhou M, Dammer EB, **Wu H**, Cutler DJ, Lah JJ, Levey AI, and Seyfried NT. (2017) *Integrating Next-Generation Genomic Sequencing and Mass Spectrometry to Estimate Allele-Specific Protein Abundance in Human Brain*. **Journal of Proteome Research**. doi: 10.1021/acs.jproteome.7b00324.
82. Oh Y, Zhang F, Wang Y, Lee EM, Choi IY, Lim H, Mirakhori F, Li R, Huang L, Xu T, **Wu H**, Li C, Qin CF, Wen Z, Wu QF, Tang H, Xu Z, Jin P, Song H, Ming GL, Lee G. (2017) *Zika virus directly infects peripheral neurons and induces cell death*. **Nature Neuroscience**. doi: 10.1038/nn.4612.

83. Weng YL, Wang X, An R, Cassin J, Vissers C, Liu Y, Liu Y, Xu T, Wang X, Wong SZH, Joseph J, Dore LC, Dong Q, Zheng W, Jin P, **Wu H**, Shen B, Zhuang X, He C, Liu K, Song H, Ming GL. (2018) *Epitranscriptomic m6A Regulation of Axon Regeneration in the Adult Mammalian Nervous System*. **Neuron**. 97(2):313-325.e6.
84. Whatney WE, Gandhi NR, Lindestam Arlehamn CS, Nizam A, **Wu H**, Quezada MJ, Campbell A, Allana S, Kabongo MM, Khayumbi J, Muchiri B, Ongalo J, Tonui J, Sasser LE, Fergus TJ, Ouma GS, Ouma SG, Beck AA, Mulligan MJ, Oladele A, Kaushal D, Cain KP, Waller L, Blumberg HM, Altman JD, Ernst JD, Rengarajan J, Day CL; TBRU-ASTRA Study Group. (2018) *A High Throughput Whole Blood Assay for Analysis of Multiple Antigen-Specific T Cell Responses in Human Mycobacterium tuberculosis Infection*. **The Journal of Immunology**. doi: 10.4049/jimmunol.1701737.
85. Hertzberg VS, Weiss H, Elon L, Si W, Norris SL; FlyHealthy Research Team. (2018) *Behaviors, movements, and transmission of droplet-mediated respiratory diseases during transcontinental airline flights*. **Proc Natl Acad Sci**. doi: 10.1073/pnas.1711611115.
86. Zhang F, Kang Y, Wang M, Li Y, Xu T, Yang W, Song H, **Wu H**, Shu Q, Jin P. (2018) *Fragile X mental retardation protein modulates the stability of its m6A-marked messenger RNA targets*. **Human Molecular Genetics**. doi: 10.1093/hmg/ddy292.
87. Cheng Y, Sun M, Chen L, Li Y, Lin L, Yao B, Li Z, Wang Z, Chen J, Miao Z, Xin N, Huang L, Allen EG, **Wu H**, Xu X, Jin P. (2018) *Ten-Eleven Translocation Proteins Modulate the Response to Environmental Stress in Mice*. **Cell Report**. doi: 10.1016/j.celrep.2018.11.061.
88. Johnson D, Huang L, Li R, Li Y, Yang Y, Kim HR, Grant C, **Wu H**, Whitsel EA, Kiel DP, Baccarelli AA, Jin P, Murabito N, Conneely KN. (2019) *Age-related DNA hydroxymethylation is enriched for gene expression and immune system processes in human peripheral blood*. **Epigenetics**. doi: 10.1080/15592294.2019.1666651.
89. Qin L, Xu Q, Li Z, Chen L, Li Y, Yang N, Liu Z, Guo J, Shen L, Allen EG, Chen C, Ma C, **Wu H**, Zhu X, Jin P, Tang B (2019) *Ethnicity-specific and overlapping alterations of brain hydroxymethylome in Alzheimer's disease*. **Human Molecular Genetics**. <https://doi.org/10.1093/hmg/ddz273>.
90. Qian X, Su Y, Adam C, Deutschmann A, Pather SR, Goldberg EM, Su K, Li S, Lu L, Jacob F, Nguyen PT, HuhS, Hoke A, Swinford-Jackson SE, Wen Z, Gu X, Pierce RC, **Wu H**, Briand LA, Chen HI, Wolf JA, Song H, Ming G (2020) *Sliced Human Cortical Organoids for Modeling Distinct Cortical Layer Formation*. **Cell Stem Cell**. <https://doi.org/10.1016/j.stem.2020.02.002>.
91. Haddad NS, Nozick S, Kim G, Ohanian S, Kraft CS, Rebolledo PA, Wang YF, **Wu H**, Bressler A, Le SNT, Kuruvilla ME, Cannon LE, Lee FE, Daiss J (2020) *Novel Immunoassay for Diagnosis of Ongoing Clostridioides difficile Infections Using Serum and Medium Enriched for Newly Synthesized Antibodies (MENSA)*, **J Immunol Methods**. doi: 10.1016/j.jim.2020.112932.
92. Shafik AM, Zhang F, Guo Z, Dai Q, Pajdzik K, Li Y, Kang Y, Yao B, **Wu H**, He C, Allen EG, Duan R, Jin P. (2021) *N6-methyladenosine dynamics in neurodevelopment and aging, and its potential role in Alzheimer's disease*. **Genome Biology**. doi:10.1186/s13059-020-02249-z.
93. Jiao B, Wang M, Feng H, Bao H, Zhang F, **Wu H**, Wang J, Tang B, Jin P, Shen L. (2021) *Downregulation of TOP2 modulates neurodegeneration caused by GGGGCC expanded repeats*. **Human Molecular Genetics**. doi: 10.1093/hmg/ddab079.

94. Haddad NS, Nguyen DC, Kuruvilla ME, Morrison-Porter A, Anam F, Cashman KS, Ramonell RP, Kyu S, Saini AS, Cabrera-Mora M, Derrico A, Alter D, Roback JD, Horwath M, O’Keefe JB, Wu HM, Wong AI, Dretler AW, Gripaldo R, Lane AN, **Wu H**, Chu HY, Lee S, Hernandez M, Engineer V, Varghese J, Patel R, Jalal A, French V, Guysenov I, Lane CE, Mengistsu T, Normile KE, Mnzava O, Le S, Sanz I, Daiss JL, Lee FE. (2021) *One-Stop Serum Assay Identifies COVID-19 Disease Severity and Vaccination Responses*. **Immunohorizons**. 5(5):322-335. doi: 10.4049/immunohorizons.2100011.
95. Corrado A, Ramonell RP, Woodruff MC, Tipton C, Wise S, Levy J, DelGaudio J, Kuruvilla ME, Magliocca KR, Tomar D, Garimalla S, Scharer CD, Boss JM, **Wu H**, Gumber S, Fucile C, Gibson G, Rosenberg A, Sanz I, Lee FE. (2021) *Extrafollicular IgD+ B cells generate IgE antibody secreting cells in the nasal mucosa*. **Mucosal Immunol**. doi: 10.1038/s41385-021-00410-w.
96. Kang Y, Zhou Y, Li Y, Han Y, Xu J, Niu W, Li Z, Liu S, Feng H, Huang W, Duan R, Xu T, Raj N, Zhang F, Dou J, Xu C, **Wu H**, Bassell GJ, Warren ST, Allen EG, Jin P, Wen Z. (2021) *A human forebrain organoid model of fragile X syndrome exhibits altered neurogenesis and highlights new treatment strategies*. **Nature Neuroscience**. doi:https://doi.org/10.1038/s41593-021-00913-6.
97. Kyu S, Ramonell RP, Kuruvilla M, Kraft CS, Wang YF, Falsey AR, Walsh EE, Daiss JL, Paulos S, Rajam G, **Wu H**, Velusamy S, Lee FE. (2021) *Diagnosis of Streptococcus pneumoniae infection using circulating antibody secreting cells*. **PLoS One**. doi: 10.1371/journal.pone.0259644.
98. Haddad NS, Nozick S, Kim G, Ohanian S, Kraft CS, Rebolledo PA, Wang Y, **Wu H**, Bressler A, Le SNT, Kuruvilla M, Runnstrom MC, Ramonell RP, Cannon LE, Lee FE, Daiss JL. (2022) *Detection of Newly Secreted Antibodies Predicts Non-recurrence in Primary Clostridioides difficile Infection*. **Journal of Clinical Microbiology**. DOI: 10.1128/jcm.02201-21.

Book Chapters

1. **Wu H**, Kerr K and Churchill GA(2002), *MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments*, **The analysis of gene expression data: methods and software**, Springer.
2. **Wu H**, Ji H (2011) *JAMIE: A Software Tool for Jointly Analyzing Multiple ChIP-chip Experiments*. **Next Generation Microarray Bioinformatics, Methods Mol Biol**. Humana Press, Springer.
3. Wu Z , **Wu H** (2016) *Experimental Design and Power Calculation for RNA-seq Experiments*. **Methods in Statistical Genomics, Methods Mol Biol**. Humana Press, Springer
4. Feng H, Conneely K, **Wu H** (2022) *Differential Methylation Analysis for Bisulfite Sequencing (BS-Seq) Data*. **Epigenome-Wide Association Studies**. Humana Press, Springer.

Thesis

1. **Wu H** (2000) *Enhancing Electricity Auction Mechanism with FACTS devices*, Master’s thesis, Iowa State University.
2. **Wu H** (2010) *Three Novel Statistical Applications in Genomics: Redefining CpG Island, Peak Detection from Multiple ChIP-chip Experiments, and Data Pre-Processing for ABI/SOLiD Second Generation Sequencing Technology*, Ph.D. thesis, Johns Hopkins University.

Talks and lectures (Since July 2010, only include the invited talks)

Invited talks at professional conferences or workshops

1. Modeling intensity data from ABI SOLiD sequencing. *ENAR spring meeting, Miami, FL*. March 22, 2011.
2. Modeling intensity data from ABI SOLiD sequencing. *ICSA meeting, New York, NY*. June 28, 2011.
3. Differential expression in RNA-seq data. *Workshop on Biostatistics and Bioinformatics. Department of Statistics, Georgia State University, Atlanta, GA*. May 5th, 2012.
4. A new shrinkage method for detecting differential expression in RNA-seq data. *ICSA meeting, Boston, MA*. June 24th, 2012.
5. Evaluating the correlations of protein binding sites. *Joint Statistical Meeting, San Diego, CA*. August 2, 2012.
6. Dispersion shrinkage methods for differential analysis in secondgeneration sequencing data. *BIT's 4th World DNA and Genome Day, Nanjing, China*. April 26, 2013.
7. A Bayesian Hierarchical Model to Detect Differentially Methylated Loci from Single Nucleotide Resolution Sequencing Data. *ICSA and KISS Joint Applied Statistics Symposium, Portland, OR*. June 16, 2014.
8. A Bayesian Hierarchical Model to Detect Differentially Methylated Loci from Single Nucleotide Resolution Sequencing Data. *Third Joint Biostatistics symposium, Chengdu, China*. June 28, 2014.
9. Differential methylation analysis from whole-genome bisulfite sequencing: a matter of spatial correlation, coverage depth and biological variance, *SAMSI epigenetics workshop, Duke University, Durham, NC*. March 10, 2015.
10. Differential methylation analysis for BS-seq data, *Cancer Genetics & Epigenetics Program Retreat 2015, Winship Cancer Institute, Emory University, Atlanta, GA*. May 21, 2015.
11. Differential methylation analysis for BS-seq data, *2015 Tsinghua Summer Workshop on Modern Statistics, Tsinghua University, Beijing, China*. June 26, 2015.
12. Differential methylation analysis for BS-seq data, *Bioinformatics workshop, Tongji University, Shanghai, China*. July 3, 2015.
13. Differential methylation analysis for BS-seq data under general experimental design, *Nankai statistics forum, Nankai University, Tianjin, China*. July 7, 2015.
14. Base-resolution methylation patterns accurately predict transcription factor bindings in vivo, *Joint Statistical Meeting, Seattle, WA*. August 9, 2015.
15. Estimating and accounting for tumor purity in DNA methylation data analysis, *The 4th Institute of Mathematical Statistics Asia Pacific Rim Meeting, Hongkong*. June 29, 2016.
16. Estimating and accounting for tumor purity in DNA methylation data analysis, *Joint Statistical Meeting, Chicago, IL*. August 3, 2016.
17. Redefining differential expression with single cell RNA-seq, *Dahshu 2017: Data Science & Computational Precision Health, San Francisco, CA*. February 21, 2017.
18. Tumor purity in DNA methylation data analysis, *Joint Statistical Meeting, Baltimore, MD*. August 3, 2017.

19. Mining large-scale biomedical data in neurodegenerative diseases, *Alliance for Neurodegenerative Diseases of the National Clinical Research Center for Geriatric Disorders, Changsha, China*. December 16, 2017.
20. Normalization and reproducibility in single cell RNA-seq, *2018 ICSA China Conference with the Focus on Data Science, Qingdao, China*. July 2, 2018.
21. Disease diagnosis for AD using cell-free DNA, *Alzheimer's Disease Centers Program Fall Meeting, Atlanta, GA*. October 20, 2018.
22. Normalization and differential expression in single cell RNA-seq, *The 11th International Conference of the ERCIM WG on Computational and Methodological Statistics, Pisa, Italy*. December 16, 2018.
23. Dissecting high-throughput (epi)genomics signals from heterogeneous samples, *Joint Statistical Meeting*. August 6, 2020.
24. Dissecting high-throughput (epi)genomics signals from heterogeneous samples, *ENAR spring meeting*. March 15, 2021.
25. On the strategy for supervised cell type identification in single-cell RNA-seq, *ICSA 2021 Applied Statistics Symposium*. September 14, 2021.
26. On the strategy for supervised cell type identification in single-cell RNA-seq, *ENAR 2022*. March 28, 2022.

Invited seminars at other academic institutes

27. Genomic bump finding. *Center for Bioinformatics and Computational Genomics, School of Biology, Georgia Institute of Technology, Atlanta, GA*. December 12, 2010.
28. Intensity pre-processing and quality assessment for ABI/SOLiD second generation sequencing data. *Computational Systems Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA*. February 17, 2011.
29. Genomic bump finding. *Department of Statistics, University of Georgia, Athens, GA*. March 3, 2011.
30. Genomic bump finding. *Department of Statistics, Georgia State University, Atlanta, GA*. March 10, 2011.
31. Modeling intensity data from ABI SOLiD sequencing. *Department of Statistics, Pennsylvania State University, State college, PA*. October 27, 2011.
32. Differential expression in RNA-seq data. *The Jackson Laboratory, Bar Harbor, ME*. March 2nd, 2012.
33. Differential expression in RNA-seq data. *Department of Biostatistics, University of Kentucky, Lexington, KY*. April 13th, 2012.
34. On the detection of differentially methylated loci. *Department of Bioinformatics, Tongji University, Shanghai, China*. April 22, 2013.
35. A novel statistical method for quantitative comparison of multiple ChIP-seq datasets. *Department of Biostatistics, Brown University, Providence, RI*. October 21, 2013.
36. Differential methylation analysis from whole-genome bisulfite sequencing: a matter of spatial correlation, coverage depth and biological variance, *Institute of Bioinformatics, University of Georgia, Athens, GA*. November 7, 2014.

37. Base-resolution methylation patterns accurately predict transcription factor bindings *in vivo*, *Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA*. February 25, 2015.
38. Differential methylation analysis for BS-seq data, *Department of Statistics, The Chinese University of Hong Kong, Hong Kong*. June 23, 2015.
39. Base-resolution methylation patterns accurately predict transcription factor bindings *in vivo*, *Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, Shen Zhen, China*. June 23, 2015.
40. Differential methylation analysis for BS-seq data, *Department of Mathematics, Hong Kong Baptist University, Hong Kong*. June 24, 2015.
41. Differential methylation analysis for BS-seq data, *Department of Mathematics, Shanghai Normal University, Shanghai, China*. June 30, 2015.
42. Differential methylation analysis for BS-seq data, *Department of Biostatistics and Epidemiology, Georgia Regents University, Augusta, GA*. October 20, 2015.
43. Base-resolution methylation patterns accurately predict transcription factor bindings *in vivo*, *Department of Mathematics, Shanghai Normal University, Shanghai, China*. June 1, 2016.
44. Base-resolution methylation patterns accurately predict transcription factor bindings *in vivo*, *Department of Mathematics, Tianjin University, Tianjin, China*. Oct 13, 2016.
45. Looking for disease diagnostic models from biomedical data. *Xiangya hospital, Changsha, China*. Apr 19, 2017.
46. Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Icahn School of Medicine at Mount Sinai, New York, NY* . May 25, 2017.
47. Constructing disease diagnostic models from large-scale biomedical data. *Nankai University School of Medicine, Tianjin, China*. June 7, 2017.
48. Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Tianjin Medical University, Tianjin, China* . June 8, 2017.
49. Constructing disease diagnostic models from large-scale biomedical data. *Tianjin University, Tianjin, China*. June 14, 2017.
50. Constructing disease diagnostic models from large-scale biomedical data. *Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing, China*. June 15, 2017.
51. Constructing disease diagnostic models from large-scale biomedical data. *The second hospital of Jilin University, Changchun, China*. December 8, 2017.
52. Constructing disease diagnostic models from large-scale biomedical data. *Children's Hospital of Zhejiang University School of Medicine, Hangzhou, China*. December 11, 2017.
53. Constructing disease diagnostic models from large-scale biomedical data. *Chinese Center for Disease Control and Prevention, National Institute of Environmental Health, Beijing, China*. June 27, 2018.
54. Differential analysis in high-throughput data. *East China Normal University, Shanghai, China*. July 9, 2019.
55. Dissecting high-throughput (epi)genomics signals from heterogeneous samples. *CAS-MPG Partner Institute of Computational Biology, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai, China*. July 10, 2019.

56. Accounting for tissue heterogeneity in the analysis of bulk high-throughput omics data. *Fred Hutchinson Cancer Center*. November 3, 2020.
57. Deciphering cell type specific information from bulk high-throughput omics data. *Department of Biostatistics, University of Pennsylvania*. April 6, 2021.
58. Deciphering cell type specific information from bulk high-throughput omics data. *Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago*. October 20, 2021.

Invited lectures and short courses

59. Analysis of ChIP-seq Data. **Third Annual Short Course on Statistical Genetics and Genomics**, *University of Alabama at Birmingham, Birmingham, AL*. July 11, 2013.
60. Introduction to the analysis of ChIP-seq Data. **Third Annual Short Course on Next-Generation Sequencing: Technology and Statistical Methods**, *University of Alabama at Birmingham, Birmingham, AL*. December 18, 2013.
61. Introduction to next-generation sequencing data analysis: RNA-seq and ChIP-seq, **7th Annual Bayesian Biostatistics and Bioinformatics Conference**, *The University of Texas MD Anderson Cancer Center, Houston, TX*. February 12, 2014.
62. Statistical methods for genome wide regional analysis with next generation sequencing data. **ENAR 2014 spring meeting**. *Baltimore, MD*. March 16, 2014.
63. Introduction to R and Bioconductor. Fudan University, School of Medicine. *Shanghai, China*. June 20-26, 2014.
64. Introduction to statistical epigenomics. **Fourth Annual Short Course on Statistical Genetics and Genomics**, *University of Alabama at Birmingham, Birmingham, AL*. July 10, 2014.
65. Introduction to the analysis of ChIP-seq Data. **Fourth Annual Short Course on Next-Generation Sequencing: Technology and Statistical Methods**, *University of Alabama at Birmingham, Birmingham, AL*. December 17, 2014.
66. Introduction to statistical epigenomics. **Fifth Annual Short Course on Statistical Genetics and Genomics**, *University of Alabama at Birmingham, Birmingham, AL*. July 30, 2015.
67. Introduction to the analysis of ChIP-seq Data. **Fifth Annual Short Course on Next-Generation Sequencing: Technology and Statistical Methods**, *University of Alabama at Birmingham, Birmingham, AL*. December 17, 2015.
68. Introduction to statistical epigenomics. **Sixth Annual Short Course on Statistical Genetics and Genomics**, *University of Alabama at Birmingham, Birmingham, AL*. July 13, 2016.
69. Introduction to the analysis of ChIP-seq Data. **Sixth Annual Short Course on Next-Generation Sequencing: Technology and Statistical Methods**, *University of Alabama at Birmingham, Birmingham, AL*. December 14, 2016.
70. Introduction to single-cell sequencing. **Sixth Annual Short Course on Next-Generation Sequencing: Technology and Statistical Methods**, *University of Alabama at Birmingham, Birmingham, AL*. December 15, 2016.

71. Introduction to the analysis of ChIP-seq Data. **Seventh Annual Short Course on Next-Generation Sequencing**, *University of Alabama at Birmingham, Birmingham, AL*. December 21, 2017.
72. Introduction to single-cell sequencing. **Seventh Annual Short Course on Next-Generation Sequencing**, *University of Alabama at Birmingham, Birmingham, AL*. December 21, 2017.
73. Feature selection in Genomic data. **Short course for AI and machine learning in medicine**, *Fudan University School of Medicine, Shanghai, China*. July 12, 2019.
74. Single cell RNA-seq data analysis. **NHGRI-funded Clinical Research Education in Genome Science (CREiGS) Virtual Short Course**, *Icahn School of Medicine at Mount Sinai*. October, 2020.
75. The Analysis of Single-cell RNA-seq Data. **ENAR spring meeting**. March 14, 2021.

Research Funding

Current external funding

1R01GM141392 (PI: Wu) Dissecting epitranscriptomic signal from complex tissues Role: PI (15%). Total direct cost: around \$1.05 million	08/01/21 – 07/31/25
1R01GM122083 (PI: Wu) Statistical Methods for Single-Cell RNA-Seq Role: PI (20%). Total direct cost: around \$1.4 million	08/01/16 – 07/31/22
1R01GM124061 (PI: Kang; Subcontract PI: Wu) Bayesian Network Biomarker Selection in Metabolomics Data Role: Subcontract PI (5%). Total direct cost: around \$182,000	08/01/20 – 07/31/22
2 P50 AG025688 (PI: Levey) Alzheimer's Disease Research Centers Role: Biostatistician (5%)	04/01/05 – 04/30/25
2 P50 AG025688 (PI: Ramalingam) Emory University Lung Cancer SPORE Role: Biostatistician (5%)	07/10/19 – 5/31/24
4 UL1 TR000454 (PI: Taylor) Georgia Clinical & Translational Science Alliance (GaCTSA) Role: Biostatistician (10%)	09/22/17 – 6/30/22
5 P30 AI050409 (PI: Del Rio) Emory/Atlanta Center for AIDS Research (CFAR) Role: Co-Investigator (5%)	09/30/02 – 07/31/22
U19 AI111211 (PI: Blumberg)	08/11/14 – 07/31/22

Role of Antigen Specific T Cell Responses in the Control of TB

Role: Co-Investigator (10%)

1 P01 NS097206 (PI: Jin)

09/01/06 – 08/31/22

Epigenetic Regulation of Neurogenesis

Role: Bioinformatics core director (10%)

1U01MH116441 (PI: Jin)

08/01/18 – 04/30/23

Dynamic RNA Modifications in human brain development and neuropsychiatric disorder

Role: Co-Investigator (5%)

R01 NS107505 (PI: Mao & Wen)

07/01/18 – 06/30/23

Chaperone-mediated Autophagy and Synaptic Dysfunction in Parkinson's Disease

Role: Co-Investigator (5%)

W81XWH1910353 (PI: Wen)

07/01/19 – 06/30/22

Modeling TSC and translating for therapeutics with human cerebral organoids

Role: Co-Investigator (5%)

Past external funding

2 VA ORD (PI: Brandes)

01/01/14 – 12/31/15

VA IPA for Hao Wu: HDAC inhibition for the prevention of Lung Cancer

Role: PI. Total direct cost: \$10,000

Boeing RC187-S1 (PI: Weiss, Hertzberg)

07/30/13 – 07/30/15

Assessing the Rate of Infectious Disease Transmission in an Aircraft Cabin

Role: Co-Investigator

R21 HD073162 (PI: Jin)

08/01/10 – 07/31/14

Mapping 5-Hydroxymethylcytosine at Single-Base Resolution in Human ES Cells

Role: Co-Investigator

Simons Foundation 239320 (PI: Jin)

07/01/13 – 06/30/15

5-Hydroxymethylcytosine-Mediated Epigenetic Regulation in Autism

Role: Co-Investigator

U34 AG051418 (Murabito, Prime PI; Conneely, Emory PI)

09/30/15 – 05/31/17

Identifying Epigenetic Mechanisms Underlying Age-related Disease Risk in CHARGE

Role: Co-Investigator

R01 DK056239 (PI: Karpen)

09/01/12 – 08/31/18

Molecular Regulation of Hepatic Transporters

Role: Co-Investigator

U19 AI110483 (PI: Sanz)

05/01/14 – 04/30/19

B cell Autoimmunity in Human SLE

Role: Co-Investigator

R01 AI111948 (PI: Day) 04/01/14 – 03/31/19
NK Cell-mediated Regulation of T Cell Immunity in TB/HIV Co-infection
Role: Co-Investigator

R01 MH102690 (PI: Jin) 04/01/14 – 02/28/18
Tet-mediated Epigenetic Modulation in Autism
Role: Biostatistician

R01 HL130471 (PI: Lewis) 01/15/16 – 12/31/19
Expectations of Discrimination and CVD risk in African-American Women
Role: Co-Investigator

W81XWH1810526 (PI: Dong) 08/01/18 – 08/14/21
A novel regulator of cancer stem cell plasticity and its roles in prostate cancer
Role: Co-Investigator (4%)

Emory funding

Emory University Research Committee (URC) Grant 05/1/12 – 04/30/13
Statistical methods for data integration in genetic association studies
Role: PI. Total direct cost: \$30,000.

Emory University Alzheimer Disease Research Center pilot Grant 05/1/16 – 04/30/17
Methylation of Cell-free DNA in Alzheimer's Disease
Role: PI. Total direct cost: \$50,000.

Emory University WHSC 2018 Synergy Award 09/1/18 – 08/31/19
Developing cell-free DNA-based next-generation biomarkers for neurodegenerative disorders
Role: PI. Total direct cost: \$100,000.

Teaching at Emory University

- Fall 2011–2021: BIOS 555 *High-throughput data analysis using R and Bioconductor*, Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University.
- Fall 2012, 2014, 2016, 2018, 2020: BIOS 731 *Advanced Statistical Computing*, Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University.
- Fall 2015, 2017, 2019, 2021: BIOS 516 *Introduction to Large-Scale Biomedical Data Analysis*, Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University.
- Spring 2013–2017: BIOS 545 *Introduction to R programming*, Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University.

Advising at Emory University

Postdoc

- Ziyi Li, Department of Biostatistics and Bioinformatics. 2018 – 2020. **Current position:** Assistant Professor, MD Anderson Cancer Center.

Thesis advisees – Ph.D.

- Andrea Lane, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Zhenxing Guo, Ph.D. candidate, Department of Biostatistics and Bioinformatics. **2021 Michael Kutner Distinguished Doctoral Student Award.**
- Luxiao Chen, Ph.D. student, Department of Biostatistics and Bioinformatics.
- Wenjing Ma, Ph.D. student, Department of Computer Science.
- Kenong Su, Ph.D. 2021, Department of Computer Science. **2021 Schoettle Graduate Research Award. Current position:** Postdoc, University of Pennsylvania.
- Hao Feng, Ph.D. 2019, Department of Biostatistics and Bioinformatics. **2019 Michael Kutner Distinguished Doctoral Student Award. Current position:** Assistant Professor, Case Western Reserve University.
- Tianlei Xu, Ph.D. 2018, Department of Computer Science. **2018 Graduate Student Research Award.** Co-advise with Zhaohui Qin. **Current position:** Senior Data Scientist, KSQ therapeutics.
- Ben Li, Ph.D. 2017, Department of Biostatistics and Bioinformatics. Co-advise with Zhaohui Qin. **Current position:** Senior Biostatistician at Merck Research Laboratories.
- Li Chen, Ph.D. 2017, Department of Computer Science. Co-advise with Zhaohui Qin. **Current position:** Assistant Professor, Indiana University School of Medicine.

Thesis advisees – Master

- Feier Han, MSPH 2022, Department of Biostatistics and Bioinformatics.
- Xuan Li, MSPH 2022, Department of Biostatistics and Bioinformatics.
- Xin Wei, MSPH 2021, Department of Biostatistics and Bioinformatics. Finalist of the **Charles C. Shepard Awards** for the best master's thesis of the Rollins School of Public Health, Emory University.
- Xiaochu Lin, MSPH 2021, Department of Biostatistics and Bioinformatics.
- Hanyi Zheng, MSPH 2020, Department of Biostatistics and Bioinformatics.
- Jinjing He, MSPH 2020, Department of Biostatistics and Bioinformatics.
- Chunhui Gu, MSPH 2020, Department of Biostatistics and Bioinformatics.
- Can Li, MSPH 2020, Department of Biostatistics and Bioinformatics.
- Yuchen Yan, MSPH 2019, Department of Biostatistics and Bioinformatics.
- Xiaoran Meng, MSPH 2019, Department of Biostatistics and Bioinformatics. Co-advise with Jingjing Yang at Department of Human Genetics. Finalist of the **Charles C. Shepard Awards** for the best master's thesis of the Rollins School of Public Health, Emory University.
- Siyi Geng, MSPH 2019, Department of Biostatistics and Bioinformatics.

- Xisha Weng, MSPH 2019, Department of Biostatistics and Bioinformatics.
- Luxiao Chen, MSPH 2018, Department of Biostatistics and Bioinformatics.
- Haonan Feng, MSPH 2017, Department of Biostatistics and Bioinformatics.
- Hao Feng, MSPH 2012, Department of Biostatistics and Bioinformatics. Co-advise with Karen Conneely. Finalists of the **Charles C. Shepard Awards** for the best master's thesis of the Rollins School of Public Health, Emory University.

Academic advisees

- Yuxuan Chen, Ph.D. student, Department of Biostatistics and Bioinformatics.
- Luxiao Chen, Ph.D. student, Department of Biostatistics and Bioinformatics.
- Zhenxing Guo, Ph.D. student, Department of Biostatistics and Bioinformatics.
- Yingtian Hu, Ph.D. student, Department of Biostatistics and Bioinformatics.
- Ran Shi, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Yunxuan Jiang, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Jay Soh, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Hao Feng, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Ruizhe Wu, MSPH, Department of Biostatistics and Bioinformatics.
- Chenchen Yu, MSPH, Department of Biostatistics and Bioinformatics.
- Shiwei Tang, MPH, Department of Biostatistics and Bioinformatics.
- Ganzhong Tian, MSPH, Department of Biostatistics and Bioinformatics.

Practicum advisees

- Xiaochu Lin, MSPH 2021, Department of Biostatistics and Bioinformatics.
- Xisha Weng, MSPH 2019, Department of Biostatistics and Bioinformatics.
- Haonan Feng, MSPH 2017, Department of Biostatistics and Bioinformatics.
- Yufeng Huang, MSPH 2015, Department of Epidemiology.
- Purva Bulsara, MPH in Biostatistics, Georgia State University.

Thesis committees

- Yanting Huang, Ph.D. candidate, Department of Computer Science.
- Yutong Jin, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Ziyue Wu, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Zhengyi Zhu, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Yingtian Hu, Ph.D. candidate, Department of Biostatistics and Bioinformatics.
- Joshua Lukemire, Ph.D. 2021, Department of Biostatistics and Bioinformatics.
- Nicolas Johnson, Ph.D. 2021, Genetics and Molecular Biology Graduate Program.
- Yunchuan Kong, Ph.D. 2020, Department of Biostatistics and Bioinformatics.
- Crystal Grant, Ph.D. 2020, Genetics and Molecular Biology Graduate Program.
- Robert Petit, Ph.D. 2019, Population Biology, Ecology and Evolution Graduate Program.

- Tao Wang, Ph.D. 2014, Department of Human Genetics.
- Nanxi Guo, MSPH 2021, Department of Biostatistics and Bioinformatics.
- Qi Yu, MSPH 2020, Department of Biostatistics and Bioinformatics.
- Zhenghao Chu, MSPH 2020, Department of Biostatistics and Bioinformatics.
- Yi Xiao, MSPH 2019, Department of Biostatistics and Bioinformatics.
- Jiani Hu, MSPH 2016, Department of Biostatistics and Bioinformatics.
- Jessica Chen, MSPH 2016, Department of Biostatistics and Bioinformatics.

Service at Emory University

- 2011–present, Affiliated faculty, Graduate Program in Population Biology, Ecology, & Evolution (PBEE).
- 2011–2012, BIOS seminar chair. Department of Biostatistics and Bioinformatics.
- 2012–2014, Woodruff Natural Science Fellowship selection committee, Emory University.
- 2011, 2013, 2018, 2020, Faculty Search Committee, Department of Biostatistics and Bioinformatics.
- 2011–2016, 2020, 2022, Curriculum Committee, Department of Biostatistics and Bioinformatics.
- 2011–2014, 2018–2020, Ph.D. admissions Committee, Department of Biostatistics and Bioinformatics.
- 2014–2017, Master admissions Committee, Department of Biostatistics and Bioinformatics.
- 2013–present, BIOS web site committee, Department of Biostatistics and Bioinformatics.
- 2018–2019, High performance computing committee, Department of Biostatistics and Bioinformatics.
- 2011–2018, High performance computing committee, Emory University.
- 2019, BIOS chair search committee, Rollins School of Public Health.
- 2020, Chair of the BIOS tenure track Assistant Professor search committee, Department of Biostatistics and Bioinformatics.
- 2021, BIOS tenure track Assistant/Associate Professor search committee, Department of Biostatistics and Bioinformatics.
- 2021, co-chair for 2021 Georgia Statistics Day.

Other professional Activities and Services

Editorial activities

- Associate editor: *Journal of Applied Statistics*.
- Review editor: *Frontiers in Epigenomics*.
- Guest editor: a supplement issue of *Cancer Informatics*.

- Article reviewer: *American Journal of Epidemiology; Annals of Applied Statistics; Bayesian Analysis; Bioinformatics; Biometrics; Biostatistics; Biosensors; BMC Bioinformatics; BMC Genomics; BMC Research Notes; Briefings in Bioinformatics; Cancer Research; Cell Discovery; Cell Death and Disease; Cell Research; Cell Reports Methods; CSH Protocols; G3: Genes/Genomes/Genetics; Genes; Genetic Epidemiology; Genome Biology; Genome Medicine; Genomics, Proteomics & Bioinformatics; Genome Research; IEEE/ACM Transactions on Computational Biology and Bioinformatics; Journal of Alzheimer's Disease; Journal of Genetics and Genomics; Journal of the Royal Statistical Society; Journal of the American Statistical Association; OBM Genetics; Plant Cell; PLoS One; PLoS Computational Biology; Proceedings of the National Academy of Sciences; Methods; Nature Communication; Nature Computational Science; Nature Genetics; Nature Protocols; New England Journal of Medicine; Nucleic Acids Research; SCIENCE CHINA Life Sciences; Science Translational Medicine; Scientific Report; Statistical Analysis and Data Mining; Statistics and Its Interface; Statistical Application in Genetics and Molecular Biology; Statistics in Medicine.*

Conference Program

- Organizer and Chair: invited session “Statistical Modeling and data mining of High-throughput Genomics Data”. *Third Joint Biostatistics symposium, Chengdu, China.* June 28, 2014.
- Chair: invited session “Modeling of long range interaction II”. *SAMSI epigenetics workshop*, Duke University, Durham, NC. March 10, 2015.
- Appointed Member: Distinguished Student Paper Award Committee, Section on Statistics in Genomics and Genetics, Joint Statistical Meeting, 2018.

Grant Review Services

- Member: NSF DMS/NIGMS grant review panel. December 2014, 2016.
- Member: American Heart Association: Artificial Intelligence, Machine Learning and Democratizing Data Grants. May 2018.
- Reviewer: Washington University Longer Life Foundation (LLF). July 2019.
- Reviewer: Woodruff Health Sciences Center (WHSC) Synergy Awards, Emory University. July 2019.
- Member: American Heart Association: Innovative Project Award: Clinical-Population Sciences. March 2020.
- Member: American Heart Association: Research Supplement to Promote Diversity in Science. May 2021.
- Member: NIH/NIEHS Special Emphasis Panel: *Functional Genomics for Interrogating G×E Interactions in Disease Review*. June 2021.
- Reviewer: Einstein Foundation Berlin. July 2021.

Other Services

- Vice president, Georgia chapter of the American Statistical Association. 2019–current.
- Provide Tenure and promotional letter for six faculty members at other institutes.

- Invited reviewer for JSM student paper awards, Section on Statistics in Genomics and Genetics, 2018.
- Panelist on *The 2nd Advancing Informatics in Government and Industry event: Too much data, Too little information?*, The Georgia Informatics Institutes. November 30th, 2018.

Software developed

On Bioconductor

- **DSS**: Dispersion shrinkage for Sequencing. An R/Bioconductor package for differential analysis from high-throughput sequencing data, including differential expression for RNA-seq and differential methylation for bisulfite-sequencing data. Available at <http://www.bioconductor.org/packages/release/bioc/html/DSS.html>.
- **TOAST**: TOols for the Analysis of heterogeneous Tissues: An R/Bioconductor package designed for the analyses of high-throughput data from heterogeneous tissues that are mixtures of different cell types. TOAST offers functions for detecting cell-type specific differential expression (csDE) or differential methylation (csDM), as well as improved reference-free deconvolution for mixing proportion estimation. Available at <http://www.bioconductor.org/packages/devel/bioc/html/TOAST.html>.
- **FEAST**: Feature selection for scRNA-seq clustering. An R/Bioconductor package for selecting the most representative features (genes) for scRNA-seq cell clustering. It can be used as a plug-in for existing clustering algorithms such as SC3, TSCAN, SHARP, SIMLR, and Seurat, and was shown to improve the results for these methods. Available at <http://bioconductor.org/packages/devel/bioc/html/FEAST.html>.
- **POWSC**: data simulation, power evaluation, and sample size calculation for single cell RNA-seq. <http://bioconductor.org/packages/devel/bioc/html/POWSC.html>.
- **PROPER**: PROspective Power Evaluation for RNA-seq. An R/Bioconductor package for power assessment and sample size determination for RNA-seq. Available at <http://www.bioconductor.org/packages/devel/bioc/html/PROPER.html>.
- **ChIPComp**: Quantitative comparison of multiple ChIP-seq datasets. An R/Bioconductor package for differential protein binding or histone modification analyses from ChIP-seq. Available at <http://bioconductor.org/packages/devel/bioc/html/ChIPComp.html>.
- **maanova**: An R package for analyzing gene expression microarray data using mixed effect model. <http://www.bioconductor.org/packages/release/bioc/html/maanova.html>.

Others

- **Wind**: weighted indexes for evaluating clustering results. <https://github.com/haowulab/Wind>.
- **SC2P**: two-phase differential expression for single-cell RNA-seq. <https://github.com/haowulab/SC2P>.
- **makeCGI**: defining CpG islands from DNA sequences using Hidden Markov Models. <http://www.haowulab.org/software/makeCGI/index.html>.
- **JAMIE**: An R package for jointly analyzing multiple ChIP-chip data sets and detecting transcription factor binding sites. <http://www.biostat.jhsph.edu/~hji/jamie/>.

Honors and Awards

- 2010, Margaret Merrell Award for excellence in research, Johns Hopkins University.
- 2011: Member of the Phi Beta Kappa Society.
- 2018: Emory Milipub Club - Having a publication with more than 1000 citations: *R/qtl: QTL mapping in experimental crosses*.
- 2019, BIOS Teaching Award, Department in Biostatistics and Bioinformatics at Emory.
- 2022, Faculty Mentor Award, Department in Biostatistics and Bioinformatics at Emory.

Professional membership

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