Xiaowei Wu

Curriculum Vitae

Department of Statistics, Virginia Tech
406-A Hutcheson, Blacksburg, VA 24061
Tel: (540)231-0023, Email: xwwu@vt.edu

Education

Ph.D.	Statistics, Rice University	2010
M.S.	Electrical Engineering, Wuhan University, China	2003
B.S.	Electrical Engineering, Wuhan University, China	1997

Academic Employment

Assistant Professor, Virginia Polytechnic Institute and State University
Aug 2012 – present

Postdoctoral Scholar, The University of Chicago
Nov 2009 – Jul 2012

Research Interests

- Statistical Genetics and Bioinformatics
- Bayesian Statistics and Machine Learning
- Branching Processes and Population Dynamics

Publications

Peer Reviewed ($^{\#}$ co-first author)

[1] **Wu, X.** and McPeek, M.S. (2018). L-GATOR: Genetic association testing for a longitudinally-measured quantitative trait in samples with related individuals. Accepted. *The American Journal of Human Genetics*[1].

[2] Luo, Y., He, J., Xu, X., Sun, M., **Wu, X.**, Lu, X. and Xie, H. (2018). Integrative single-cell omics analyses reveal epigenetic heterogeneity in mouse embryonic stem cells. Accepted. *PLoS Computational Biology.*

[3] **Wu, X.**, Guan, T., Liu, J.D., Leon-Novelo, L.G. and Bandyopadhyay, D. (2018). Adaptive-weight burden test for associations between quantitative traits and genotype data with complex correlations. Accepted. *Annals of Applied Statistics.*

[4] Rodrigues, R., Rodgers, N., **Wu, X.** and Williams, M. (2018). COREMIC: a web-tool to search for a niche associated CORE MICrobiome. *Peer J*, 6:e4395, doi: 10.7717/peerj.4395.

[5] Tran, H., Zhu, H., **Wu, X.**, Kim, G., Clarke, C.R., Larose, H., Haak, D.C., Askew, S.D., Barney, J.N., Westwood, J.H. and Zhang, L. (2018). Identification of differentially methylated sites with weak methylation effect. *Genes*, 9(2), 75, doi: 10.3390/genes9020075

---

[1] Impact factor (2016): 9.025

[6] Hasan, M.S., **Wu, X.**, Watson, L.T. and Zhang, L. (2017). UPS-indel: a universal positioning system for indels. *Scientific Reports*, 7(1):14106. PMCID: PMC5658412, doi:10.1038/s41598-017-14400-1.

[7] Jin, X., Xiang, X., Guan, X., Bai, Q., **Wu, X.** and Zhou, Y. (2017). Assessing the relationship between the spatial distribution of land consolidation projects and farmland resources in China, 2006-2012. *Food Security*, 9(5):889–905, doi:10.1007/s12571-017-0719-6.

[8] Zhu, H., Caspers, P., Morris, J.S., **Wu, X.** and Müller, R. (2017). A unified analysis of structured sonar-terrain data using Bayesian functional mixed models. *Technometrics*, 60(1):112–123, doi:10.1080/00401706.2016.1274681.

[9] Zhang, S., You, Z. and **Wu, X.** (2017). Plant disease leaf image segmentation based on superpixel clustering and EM algorithm. *Neural Computing and Applications*, doi:10.1007/s00521-017-3067-8.

[10] León-Novelo, L.G., Womack, A., Zhu, H. and **Wu, X.** (2017). A Bayesian analysis of quantal bioassay experiments incorporating historical controls via Bayes factors. *Statistics in Medicine*, 36(12):1907–1923, doi:10.1002/sim.7218.

[11] Zhang, S., Zhu, Y., You, Z. and **Wu, X.** (2017). Fusion of superpixel, expectation maximization and PHOG for recognizing cucumber diseases. *Computers and Electronics in Agriculture*, 140:338–347, doi:10.1016/j.compag.2017.06.016.

[12] Wang, Z., Xu, K., Zhang, X., **Wu, X.** and Wang, Z. (2017). Longitudinal SNP-set association analysis of quantitative phenotypes. *Genetic Epidemiology*, 41(1):81–93. PMCID: PMC5154867, doi:10.1002/gepi.22016.

[13] Zhang, S., **Wu, X.** and You, Z. (2017). Jaccard distance based weighted sparse representation for coarse-to-fine plant species recognition. *PLoS One*, 12(6):e0178317, doi:10.1371/journal.pone.0178317.

[14] Zhang, S., **Wu, X.**, You, Z. and Zhang, L. (2017). Leaf image based cucumber disease recognition using sparse representation classification. *Computers and Electronics in Agriculture*, 134:135–141, doi:10.1016/j.compag.2017.01.014.

[15] Sun, M., Sun, Z., **Wu, X.**, Rajaram, V., Keimig, D., Lim, J., Zhu, H. and Xie, H. (2016). Mammalian brain development is accompanied by a dramatic increase in cell-subset specific DNA methylation. *Scientific Reports*, 6:32298. PMCID: PMC5010034.

[16] Jin, X., Zhang, Z., **Wu, X.**, Xiang, X., Sun, W., Bai, Q. and Zhou, Y. (2016). Co-ordination of land exploitation, exploitable farmland reserves and national planning in China. *Land Use Policy*, 57, 682–693.

[17] Tran, H., **Wu, X.**, Tithi, S., Sun, M., Xie,H. and Zhang, L. (2016). A Bayesian assignment method for ambiguous bisulfite short reads. *PLoS One*, 11(3):e0151826. PMCID: PMC4806927.

[18] **Wu, X.** and Zhu, H. (2015). A Bayesian analysis of copy number variations in array comparative genomic hybridization data. *International Journal of Biomedical Data Mining*, 4:116.

[19] Hasan, M.S., **Wu, X.** and Zhang, L. (2015). Performance evaluation of indel calling tools using real short-read data. *Human Genomics*, 9:20.

[20] **Wu, X.**, Sun, M., Zhu, H. and Xie, H. (2015). Nonparametric Bayesian clustering to detect bipolar methylated genomic loci. *BMC Bioinformatics*, 16:11. PMCID: PMC4302125.

[21] Karunasena, E., McIver, L.J., Bavarva, J.H., **Wu, X.**, Zhu, H. and Garner, H.R. (2015). 'Cut from the same cloth': Shared microsatellite variants among cancers link to ectodermal tissues-neural tube and crest cells. *Oncotarget*, 6(26):22038–22047. PMCID: PMC4673144.

[22] **Wu, X.** and Zhu, H. (2015). Fast maximum likelihood estimation of mutation rates using a birth-death process. *Journal of Theoretical Biology*, 366, 1–7.

[23] Li, Z., **Wu, X.**$^{\#}$, He, B., and Zhang, L. (2014). Vindel: a simple pipeline for checking indel redundancy. *BMC Bioinformatics*, 15:359.

[24] Karunasena, E., McIver, L.J., Rood, B. R., **Wu, X.**, Zhu, H., Bavarva, J.H. and Garner, H.R. (2014). Somatic intronic microsatellite loci differentiate glioblastoma from lower-grade gliomas. *Oncotarget*, 5(15):6003–6014. PMCID: PMC4171608.

[25] **Wu, X.** and Kimmel, M. (2013). Modeling neutral evolution using infinite-allele Markov branching processes. *International Journal of Stochastic Analysis*, Volume 2013, Article ID 963831.

[26] **Wu, X.** and Kimmel, M. (2010). A note on the path to extinction of critical Markov branching processes. *Statistics and Probability Letters*, 80, 263–269.

[27] **Wu, X.**, Strome, E.D., Meng, Q., Hastings, P.J., Plon, S.E. and Kimmel, M. (2009). A robust estimator of mutation rates. *Mutation Research*, 661, 101–109.

[28] Strome, E.D., **Wu, X.**, Kimmel, M. and Plon, S.E. (2008). Heterozygous screen in *Saccharomyces Cerevisiae* identifies dosage-sensitive genes that affect chromosome stability. *Genetics*, 178, 1193–1207.

Submitted and Under Revision

[29] **Wu, X.** Borrowing strength from complex genotypic correlations: multiple genetic variant association testing for quantitative traits. Submitted. *ICSA Springer Book.*

[30] Banerjee, S., Zhu, H., Tang, M., Feng, W., **Wu, X.** and Xie, H. Identifying transcriptional regulatory modules among different chromatin states in mouse neural stem cells. Under revision. *Nucleic Acids Research.*

Manuscripts († corresponding author)

[31] Tang, M., Hasan, M. S., Zhang, L., Zhu, H. and **Wu, X.**† vi-HMM: A novel HMM-based method for sequence variant identification in short read data.

[32] **Wu, X.** and Zhu, H. A nonparametric Bayesian model for clustering inhomogeneous Poisson processes with application to detect transcriptional regulatory modules.

[33] Lu, R., Zhu, H. and **Wu, X.**† Using approximate Bayesian computation to estimate mutation rates in Markov branching process models.

[34] **Wu, X.** and Zhu, H. Dictionary learning for genetic data in related individuals.

Conference Proceedings

[35] Banerjee, S., Chen, X., **Wu, X.**, Xie, H., Xuan, J. and Feng, W. (2017). ChIP-GMM: A Gaussian mixture model for inferring binding regions in ChIP-seq profiles. *The 9h International Conference on Bioinformatics and Computational Biology.*

[36] Hasan, M. S., **Wu, X.**, Watson, L. T., Li, Z. and Zhang, L. (2016). UPS-indel: A better approach for finding indel redundancy. *IEEE 6th International Conference on Computational Advances in Bio and Medical Sciences.*

[37] Tran, H., **Wu, X.** and Zhang, L. (2015). A Bayesian method for assigning ambiguous bisulfite short reads. *The 7th International Conference on Bioinformatics and Computational Biology.*

[38] Zhu, H., **Wu, X.**, Xu, X. and Wu, C. (2007). Feature extraction and categorization of patient respiration pattern in radiation therapy. *The 15th International Conference on the Use of Computers in Radiation Therapy.*

[39] **Wu, X.**, Chan, Y., Preissl, H., Eswaran, H., Wilson, J., Murphy, P. and Lowery, C.L. (2004). Time-frequency and coherence analysis of fMEG signals. *The 14th International Conference on Biomagnetism.*

TALKS AND PRESENTATIONS

Invited Talks

[1] **Wu, X.** Learning patterns from genomics data through stochastic modeling. Department of Physics, Virginia Tech, Blacksburg, VA, USA, Oct 2017.

[2] **Wu, X.** and Zhu, H. Dictionary learning based genotype imputation to improve power for association testing. ICSA Applied Statistics Symposium, Chicago, IL, USA, Jun 2017.

[3] **Wu, X.** Multiple-variant association testing for quantitative traits with application to dental caries GWAS data. Department of Mathematics and Statistics, James Madison University, Harrisonburg, VA, USA, Feb 2017.

[4-5] **Wu, X.** Gene-based association and network analyses for dental caries GWAS data. Department of Biostatistics, University of Texas M. D. Anderson Cancer

Center, Houston, TX, USA, Oct 2016; Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA, Sep 2016.

[6] **Wu, X.** Principal component based adaptive-weight burden test for quantitative trait associations. ICSA Applied Statistics Symposium, Atlanta, GA, USA, Jun 2016.

[7] Zhu, H. and **Wu, X.** Fast Bayesian inference for complex, ultra-high dimensional functional data. ICSA Applied Statistics Symposium, Atlanta, GA, USA, Jun 2016.

[8] **Wu, X.** The optimal burden test on genotype data with bi-directional correlations. New England Statistics Symposium, New Haven, CT, USA, Apr 2016.

[9-10] **Wu, X.** Genetic association testing for a longitudinally-measured quantitative trait in samples with related individuals. Department of Mathematics, University of North Texas, Denton, TX, USA, Feb 2012; Department of Statistics, West Virginia University, Morgantown, WV, USA, Mar 2012.

[11-13] **Wu, X.** Bayesian analysis of array CGH data. Department of Statistics, University of Chicago, Chicago, IL, USA, July 2009; Department of Biostatistics, Epidemiology and Informatics, UPenn School of Medicine, Philadelphia, PA, USA, July 2009; Biostatistics Branch, NIEHS, Durham, NC, USA, June 2009.

Contributed Talks

[14] Guan, T. and **Wu, X.** Built-in BLUP genotype imputation of adaptive-weight genetic association burden test for data with complex correlations. The Joint Statistical Meeting, Baltimore, MD, USA, Aug 2017.

[15] León-Novelo, L.G., Womack, A., Zhu, H. and **Wu, X.** Bayesian analysis of quantal bioassay experiments incorporating historical controls via Bayes factors. The Joint Statistical Meeting, Chicago, IL, USA, Jul 2016.

[16] **Wu, X.** and Zhu, H. Dictionary learning for genetic data in related individuals. The annual meeting of the Statistical Section of the Virginia Academy of Sciences, Blacksburg, VA, USA, May 2013.

[17] **Wu, X.** Bayesian analysis of array CGH data. The Joint Statistical Meeting, Washington, DC, USA, Aug 2009.

[18] **Wu, X.** and Kimmel, M. Variability study of a robust estimator of mutation rate in fluctuation analysis. The Joint Statistical Meeting, Salt Lake City, UT, USA, Jul 2007.

[19] **Wu, X.** and Kimmel, M. Some estimation problems in fluctuation analysis. The 3rd Erich L. Lehmann Symposium, Houston, TX, May 2007.

Posters

[20] **Wu, X.** Optimal burden test for association between quantitative traits and genotype data with bidirectional correlations. The Joint Statistical Meeting, Chicago, IL, USA, Aug 2016.

[21] **Wu, X.** and Zhu, H. Nonparametric Bayesian clustering to detect bipolar methylated genomic loci. The 4th Workshop on Biostatistics and Bioinformatics, Atlanta, GA, USA, May 2015.

[22] Zhu, H. and **Wu, X.** Wavelet-based imaging genomic modeling in Autism Spectrum Disorder study. The 4th Workshop on Biostatistics and Bioinformatics, Atlanta, GA, USA, May 2015.

[23] **Wu, X.** An optimal quasi-likelihood-based burden test for rare-variant association. The 63th Annual Meeting of the American Society of Human Genetics, Boston, MA, USA, Oct 2013.

[24] **Wu, X.** and Zhu, H. Bayesian dictionary learning in genetic studies on related individuals. The 22nd Annual Conference of the International Genetic Epidemiology Society, Chicago, IL, USA, Sep 2013.

[25] **Wu, X.** and McPeek, M.S. Genetic association testing for a longitudinally-measured trait in samples with related individuals. The 60th Annual Meeting of the American Society of Human Genetics, Washington, DC, USA, Nov 2010.

GRANTS AND PROPOSALS

Awarded

[1] Title: Statistical Methods to Unveil the Genetic Architecture of Dental Caries. Source of support: 4-VA Collaborative Research Grant Program. Role: PI. Amount: $25,000. Credit: 100%. Effective Date: 01/16/2018 - 01/15/2019.

[2] Title: Enhance Research Experiences for Underrepresented Undergraduate Students in Cross Disciplinary Research in the Areas of Bioinformatics and Statistics. Source of support: Virginia Tech Institute for Critical Technology and Applied Science (ICTAS). Role: Co-PI. Amount: $10,000. Credit: 50%. Effective Date: 11/16/2017 - 06/15/2018.

[3] Title: The Data & Decisions Destination Area Concept Proposal: A Scalable Analytical Framework for Ultra-High Dimensional Neuroimaging Data. Source of support: Virginia Tech. Role: Co-PI. Amount: $25,000. Credit: 5%. Effective Date: 12/01/2017 - 06/30/2018.

[4] Title: Imaging Genomic Modeling in Autism Spectrum Disorder Study. Source of support: Virginia Tech Center for Autism Research (VTCAR). Role: Co-PI. Amount: $5,000. Credit: 20%. Effective Date: 01/01/2015 - 06/15/2015.

[5] Title: Mentoring Award for New Tenure-track Assistant Professors. Source of support: Virginia Tech. Role: PI. Amount: $1,500. Credit: 100%. Effective Date: 03/01/2012 - 06/30/2014.

Service Grants (non-PI/Co-PI role)

[6] Traffic Prediction Based Radio Access Point Handover Control in Mobile Networks. Source of support: NSF-China. Role: Senior Personnel. Amount: ¥580,000. My share: International travel grant. Effective Date: 01/01/2017 - 12/31/2020.

Pending

[1] Understanding the Role of Genetic Variations in Cancer and Other Disease Risk via Statistical Modeling. Source of Support: NIH. Role: PI, Amount: $378,686. Credit: 85%.

[2] Statistical Modeling to Unveil Transcriptional Regulatory Mechanism in Embryonic Stem Cells. Source of Support: NIH. Role: PI, Amount: $152,520. Credit: 76%.

[3] ABI Innovation: SAMI: A New Platform and Toolkit for Statistical Analysis and Modeling of INDELs. Source of Support: NSF. Role: PI, Amount: $476,679. Credit: 80%.

[4] ABI Development: FastViromeExplorer-A Ultrafast and Accurate System for Virus Identification and Abundance Profiling in Metagenomics Data. Source of Support: NSF. Role: Co-PI, Amount: $560,135. Credit: 5%.

[5] ABI Development: A Code-Generation Framework and High-Performance Library for Biocomputing on Emerging Parallel Hardware. Source of Support: NSF. Role: Co-PI, Amount: $599,318. Credit: 4%.

[6] CDS&E: A Computational Framework for Parsimonious Sonar Sensing. Source of Support: NSF. Role: Co-PI, Amount: $698,955. Credit: 10%.

[7] Integrating Bayesian Statistics, Weighted Set Cover Algorithms, and Parallel Processing to Identify Combinations of Tumorigenic Mutations in Cancer. Source of Support: NIH. Role: Co-PI, Amount: $282,044. Credit: 10%.

COLLABORATIONS

Virginia Tech

Hehuang David Xie, Biocomplexity Institute

Liqing Zhang, Department of Computer Science

Amy Pruden, Department of Civil and Environmental Engineering

Mark Williams, Department of Horticulture

International

Xiaobin Jin, School of Geographic and Oceanographic Sciences, Nanjing University, China

Shaowei Wang, School of Electronic Science and Engineering, Nanjing University, China

TEACHING EXPERIENCE

- Applied Stochastic Processes: Spring 2015–2018
- Inference Fundamentals with Applications to Categorical Data: Fall 2014–2016
- Introduction to Statistical Genomics: Spring 2013, 2014
- Introduction to R Programming: Spring 2016–2018
- Probability and Statistics for Electrical Engineers: Fall 2012, 2013

ACADEMIC SERVICE

Journal Referee

Genetica, *Computational and Mathematical Methods in Medicine*, *BMC Genetics*, *Scientific Reports*, *Journal of Theoretical Biology*, *Bioinformatics*, *ICSA Springer Book*, *Computers and Electronics in Agriculture*, *Frontiers Plant Science*.

Journal Review Editor

Frontiers, the Editorial Board of Bioinformatics and Computational Biology

Conference Organizations

Organizing chair of invited session "Statistical Genetics" on the 2017 Annual meeting of International Chinese Statistical Association, 06/25/2017

Organizing chair of invited session "Statistical Genetics" on the 2016 Annual meeting of International Chinese Statistical Association, 06/13/2016

Faculty affiliate of the Virginia Tech interdisciplinary Ph.D. program in Genetics, Bioinformatics, and Computational Biology (GBCB). 2017–present

Faculty affiliate of the Virginia Tech Global Systems Science (GSS) destination area concept group. 2017–present

Advising

Ph.D. Students: Man Tang (STAT), Ting Guan (STAT), Sharmi Banerjee (ECE).

Ph.D. Dissertation Committee: Zhenguo Gao (STAT), Jacob Potter (CS).
Aritra Chatterjee (CEE), graduated.
Vinaya Vijayan (CS), graduated.
Mingming Liu (CS), graduated.
Rejesh Bawa (Forestry), graduated.
Hong Tran (CS), graduated.

Master Oral Exam: Jing Niu, Adam Edwards, Shuai Zhang, Elaine Perrin,
Ying Wang, Hong Tran, Weibin Xu, Taige Wang, Jia Liu,
Tianyi Wang, Xin Feng, Wenmeng Tian, Alexander Durbin,
Li Xu, Chenang Liu, Rejesh Bawa (DAAS), Lin Tan (DAAS),
Wei He, Man Zhang, Jaeo Han.

Department and University Committees

Qualifying Exam Committee, Department of Statistics, Virginia Tech.	2014–2017
Bioinformatics Committee, Department of Statistics, Virginia Tech.	2013–2016