Yi-Juan Hu (updated on 2021-4-6)

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University
Atlanta, Georgia USA 30322

Phone: (404) 712-4466
E-mail: yijuan.hu@emory.edu

Research Interests

Statistical methods for analyzing microbiome data; Statistical genetics. My research focuses on the development of statistical methods and software programs for analyzing high-throughput microbiome data and genetic data in human epidemiological and clinical studies.

Professional Experience

Associate Professor, September 2017 – present
Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University

Rollins Assistant Professor (tenure-track, endowed position), August 2011 – September 2017
Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University

Education

Ph.D., Biostatistics, May 2011
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC

- *Dissertation*: Statistical Analysis of Haplotypes, Untyped SNPs, and CNVs in Genome-Wide Association Studies. (*Advisor*: Dr. Danyu Lin, Dennis Gillings Distinguished Professor)
  [⋆⋆ **Barry H. Margolin Doctoral Dissertation Award, Biostatistics, UNC-CH** ⋆⋆]

B.S., Statistics, Honor Graduate, July 2005
School of Mathematics, Peking University, Beijing, China

B.S., Economics, July 2005
China Center for Economic Research, Peking University, Beijing, China

Research Grants

Active Support

R01 GM116065-01A1 (PI: Hu) NIH/NIGMS 09/18/15–08/31/21 (NCE)
Association Tests of Rare Variants Using Sequence Reads without Calling Genotypes
Role: PI

R01 DK124821 (MPI: Pacifici, Jones & Bilezikian) NIH 02/01/21-01/31/25
Role of the gut microbiome in the bone loss induced by hyperparathyroidism in mice humans
Role: Co-I

R01 AI151075 (PI: Dean) NIH/NIAID 03/20/20–02/28/25
Natural History of C. trachomatis urogenital and rectal infections
Role: Co-I

R01 AI147839-01 (PI: Kelley) NIH/NIAID 08/01/19-07/31/24
Understanding the Rectal Mucosal Effects of Cross-Sex Hormone Therapy among US and Thai Transgender Women
Role: Co-I

5UH 3OD023318-05S1 (MPI: Dunlop, Corwin & Brennan) NIH 09/21/16–08/31/23
The impact of the intrauterine and early childhood environments on neurocognitive and metabolic development in African American youth: focus on the gut-brain axis
Role: Co-I

R00 NR017897 (PI: Bai) NIH/NINR 09/03/20–07/31/23
Exploring the Microbiome-Gut-Brain Axis in Psychoneurological Symptoms for Children with Solid Tumors
Role: Co-I

U01 AI138909-01 (PI: Larsen) NIH/NIAID 08/01/18–07/31/23
Selective CD28 blockade in renal transplant recipients
Role: Co-I

R01 AI139188 (PI: David) NIH/NIAID 08/16/18–07/31/23
Genomics of S. Aureus colonization after initial and recurrent skin infections and the impact of antibiotics
Role: Co-I

P30 AI050409-19 (PI: Del Rio) NIH/NIAID 08/01/17–07/31/22
Emory/Atlanta Center for AIDS Research (CFAR)
Role: Co-I

R01 R01CA207467-01 (PI: Fedirko) NIH/NCI 07/01/17–06/30/22
Gut Microbiome, Antibiotic Use & Colon Cancer Recurrence
Role: Co-I

R01 HD092033-03 (PI: Kelley) NIH/NICHD 04/01/17–03/31/22(NCE)
The Rectal Mucosa and STI: Implications for HIV Transmission and Prevention
Role: Co-I

R01AI128799-01 (PI: Kelley) NIH/NIAID 12/01/16–11/30/21
Understanding Rectal HIV transmission among At-risk MSM: Age, Intercourse, and Mucosal Injury
Role: Co-I

Publications

Methodological papers [† PhD trainee] [*corresponding author]

1. **Hu, Y. J.***, Satten, G. A. (2021) A rarefaction-without-resampling extension of PERMANOVA for testing presence-absence associations in the microbiome. ***bioRxiv***.
   Software developed: R package LDM

2. Zhu, Z.†, Satten, G. A., Mitchell, C., and **Hu, Y. J.*** (2021) Analyzing matched sets of microbiome data using the LDM and PERMANOVA. ***Microbiome***, in press.
   Software developed: R package LDM

3. **Hu, Y. J.***, Lane, A.†, Satten, G.A. (2021) A Rarefaction-Based Extension of the LDM for Testing Presence-Absence Associations in the Microbiome. ***Bioinformatics***, https://doi.org/10.1093/bioinformatics/btab012.
   Software developed: R package LDM

4. Li, Y.†, **Hu, Y. J.***, and Satten, G. A. (2020) A Bottom-up Approach to Testing Hypotheses That Have a Branching Tree Dependence Structure, with Error Rate Control. ***Journal of American Statistical Association (T&M)***, DOI: 10.1080/01621459.2020.1799811.
   Software developed: R package BOUTH

5. **Hu, Y. J.**, Satten, G. A. Testing hypotheses about microbiome using the linear decomposition model (LDM) (2020). ***Bioinformatics***, 36(14):4106–4115. https://doi.org/10.1093/bioinformatics/btaa260
   Software developed: R package LDM

6. Liao, P.†, Satten, G. A., and **Hu, Y. J.*** (2017) Robust inference of population structure from next-generation sequencing data with systematic differences in sequencing. ***Bioinformatics***, 34(7):1157–1163. https://doi: 10.1093/bioinformatics/btx708.
Software developed: TASER-PC

7. **Hu, Y. J.**, Schmidt, A. F., Dudbridge, F., Holmes, M. V., Brophy, J. M., Tragante, V., Li, Z., Liao, P., McCubrey, R. O., Horne, B. D., Hingorani, A. D., Asselbergs, F. W., Patel, R., Long, Q. on behalf of the GENIUS-CHD Consortium (2017). The impact of selection bias on estimation of subsequent event risk. ***Circulation: Cardiovascular Genetics***, 10(5):e001616.

8. Liao, P.†, Satten, G. A., and **Hu, Y. J.*** (2017) PhredEM: a Phred-score-informed genotype-calling approach for next-generation sequencing studies. ***Genetic Epidemiology***, 41(5):375-387.
Software developed: PhredEM

9. **Hu, Y. J.**, Liao, P., Johnston, H. R., Allen, A. S., and Satten, G. A. (2016). Testing rare-variant association without calling genotypes allows for systematic differences in sequencing between cases and controls. ***PLoS Genetics***, 12(5): e1006040. doi:10.1371/journal.pgen.1006040.
Software developed: TASER

10. **Hu, Y. J.**, Sun, W., Tzeng, J. Y., Perou, C. M. (2015). Proper use of allele-specific expression improves statistical power for *cis*-eQTL mapping with RNA-seq data. ***Journal of American Statistical Association (A&C)***, 110(511): 962–974.
Software developed: TRECASE-MLE

11. **Hu, Y. J.**, Li, Y., Auer, P. L., Lin, D. Y. (2015). Integrative analysis of sequencing and array genotype data for discovering disease associations with rare mutations. ***Proceedings of the National Academy of Sciences***, 112(4): 1019–1024.
Software developed: SEQGWAS

12. Yuan, S., Johnston, H. R., Zhang, G., Li, Y., **Hu, Y. J.**, Qin, Z. (2015). One size doesn't fit all - RefEditor: building personalized diploid reference genome to improve read mapping and genotype calling in next generation sequencing studies. ***PLoS Computational Biology***, 11(8): e1004448. doi: 10.1371/journal.pcbi.1004448.

13. Johnston, H. R., **Hu, Y. J.**, and Cutler, D. J. (2015) Population genetics identifies challenges in analyzing rare variants. ***Genetic Epidemiology***, 39(3):145–148.

14. **Hu, Y. J.**, Lin, D. Y., Sun, W., and Zeng, D. (2014). A likelihood-based framework for association analysis of allele-specific copy numbers. ***Journal of American Statistical Association (T&M)***, 109(508): 1533–1545.
Software developed: CNVstat

15. **Hu, Y. J.**, Hui, Q., and Sun Y. V. (2014). Association analysis of whole-genome sequence data accounting for longitudinal and family designs. ***BMC Proceedings***, 8(Suppl 1): S89.

16. Wu, Z.*, **Hu, Y. J.***, Melton, P. E. (2014). Longitudinal data analysis for genetic studies in the whole-genome sequencing era. ***Genetic Epidemiology***, 38 Suppl 1:S74–80. (*co-first authors)

17. **Hu, Y. J.**, Berndt, S. I., Gustafsson, S., Ganna, A., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Hirschhorn, J., North, K. E., Ingelsson, E., and Lin, D. Y. (2013). Meta-analysis of gene-level associations with rare variants based on single-variant statistics. ***American Journal of Human Genetics***, 93: 236–248.
Software developed: MAGA

18. Sun, W., **Hu, Y. J.** (2012). eQTL mapping using RNA-seq data. ***Statistics in Biosciences***, doi:10.1007/s12561-012-9068-3.

19. **Hu, Y. J.**, Lin, D. Y. (2010). Analysis of untyped SNPs: maximum likelihood and imputation methods. ***Genetic Epidemiology***, 34: 803–815.
Software developed: tagIMPUTE

20. **Hu, Y. J.**, Lin, D. Y., Zeng, D. (2010). A general framework for studying genetic effects and gene-environment interactions with missing data. ***Biostatistics***, 11: 583–598.
Software developed: SNPMStat
[⋆⋆ **An earlier version won 2010 ENAR Distinguished Student Paper Award** ⋆⋆]

21. Lin, D. Y., **Hu, Y. J.**, Huang, B. E. (2008). Simple and efficient analysis of disease association with missing genotype data. ***American Journal of Human Genetics***, 82: 444–452.

22. Lin, D. Y., **Hu, Y. J.** (2008). Reply to Marchini and Howie. ***American Journal of Human Genetics***, 83: 539–540.

Collaborative papers

1. Yang, I., Arthur, R. A., Zhao, L., Clark, J., **Hu, Y. J.**, Corwin, E. J., Lah, J. (2021) The Oral Microbiome and Inflammation in Mild Cognitive Impairment. ***Experimental Gerontology***, in press.

2. Smith, S. A., Murray, P. M., Amancha, P. K., Ackerley, C. G., **Hu, Y. J.**, Amara, R. R., Kelley, C. F. (2020) Ex vivo rectal explant model reveals potential opposing roles of Natural Killer cells and Marginal Zone-like B cells in HIV-1 infection. ***Scientific Report***, doi: 10.1038/s41598-020-76976-5.

3. Huang, Y., Hui, Q., Gwinn, M, **Hu, Y. J.**, Quyyumi, A. A., Vaccarino. V., Sun. Y. V. (2020) Sexual Differences in Genetic Predisposition of Coronary Artery Disease. ***Circulation: Genomic and Precision Medicine***, doi.org/10.1161/CIRCGEN.120.003147.

4. Yang, I., **Hu, Y. J.**, Corwin, E.J., Dunlop, A.L. (2020) Exploring the maternal infant oral microbiome. ***Journal of Perinatal and Neonatal Nursing***, doi: 10.1097/JPN.0000000000000511.

5. Tsementzi, D., Pena-Gonzalez, A., Bai, J., **Hu, Y. J.**, Patel, P., Shelton, J., Dolan, M., Arluck, J., Khanna, N., Conrad, L., Scott, I., Eng, T. Y., Konstantinidis, K. T., Bruner, D. W. (2020) Comparison of vaginal microbiota in gynecologic cancer patients pre- and post-radiation therapy and healthy women. ***Cancer Medicine***, doi: 10.1002/cam4.3027.

6. Amancha, P. K., Ackerley, C. G., Duphare, C., Lee, M., **Hu, Y. J.**, Amara, R. R., Kelley C. F. (2019). Distribution of functional CD4 and CD8 T cell subsets in blood and rectal mucosal tissues. ***Scientific Report***, doi: 10.1038/s41598-019-43311-6.

7. Daya, M., Barnes, K., Boorgula, M. P., Campbell, M., Chavan, S., Lange, L., Lange, E., Shetty, A., Rafaels, N., . . . , Barnes, K. C., and CAAPA[#] (2019) Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. ***Nature Communications***, https://doi.org/10.1038/s41467-019-08469-7. [[#] Hu, Y. J. included as consortium co-authors]

8. Bai, J., **Hu, Y. J.**, Bruner, D. W. (2018) Composition of gut microbiota and its association with body mass index and lifestyle factors in a cohort of 7-18 years old children from the American Gut Project. ***Pediatric Obesity***, doi: 10.1111/ijpo.12480.

9. Haaland, R. E., Fountain, J., Hu, Y, Holder, A., Dinh, C., Hall, L., Pescatore, N. A., Heeke, S., Hart, C. E., Xu, J., **Hu, Y. J.**, Kelley, C. F. (2018) Repeated rectal application of a hyperosmolar lubricant is associated with microbiota shifts but does not affect PrEP drug concentrations: results from a randomized trial in men who have sex with men. ***Journal of the International AIDS Society***, https://doi.org/10.1002/jia2.25199.

10. Lai, L., Rouphael, N., Xu, Y., Natrajan, M. S., Beck, A., Hart, M., Feldhammer, M., Feldpausch, A., Hill, C., Wu, H., Fairley, J. K., Lankford-Turner, P., Kasher, N., Rago, P., **Hu, Y. J.**, Edupuganti, S., Patel, S. M., Murray, K. O., Mulligan, M. J., Emory Zika Patient Study Team (2018) Innate, T-, and B-Cell Responses in Acute Human Zika Patients. ***Clinical Infectious Diseases***, 66(1):1–10.

11. Johnston, H. R., **Hu, Y. J.**, Gao, J., OConnor, T. D., Abecasis, G., Wojcik, G. L., Gignoux, C. R., Gourraud, P. A., Lizee, A., Hansen, M., Genuario, Bullis, R. D., Lawley, C., Kenny, E. E., Bustamante, C., Mathias, R., Barnes, K. and Qin, Z. S. on behalf of the CAAPA Consortium. (2017) Identifying tagging SNPs for African specific genetic variation from the African Diaspora Genome. ***Scientific Report***, 21;7:46398.

12. Kessler, M. D., Yerges-Armstrong, L., Taub, M. A., Shetty, A. C., Maloney, K., Jeng, L. J., Ruczinski, I., Levin, A. M., Williams, L. K., Beaty, T. H., Mathias, R. A., Barnes, K. C.; Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA)$^{\#}$, O'Connor, T. D. (2016) Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry. ***Nature Communications***, 11;7:12521. [$^{\#}$ Hu, Y. J. included as consortium co-authors]

13. Mathias, R. A., Taub, M. A., Gignoux, C. R., Fu, W., Musharoff, S., O'Connor, T. D., ..., CAAPA$^{\#}$, Barnes, K. C. (2016) A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. ***Nature Communications***, 11;7:12522. [$^{\#}$ Hu, Y. J. included as consortium co-authors]

14. Colasanti, J., Kelly, J., Pennisi, E., **Hu, Y. J.**, Hughes, D., Root, C., del Rio, C., Armstrong, W. (2015). Continuous retention and viral suppression provide further insights into the HIV care continuum compared to the cross-sectional HIV care cascade. ***Clinical Infectious Diseases***, doi: 10.1093/cid/civ941.

15. Zhang, P., Huang, C., Fu, C., Tian, Y., **Hu, Y. J.**, Wang, B., Strasner, A., Song, Y., Song, E. (2015). Cordycepin (3'-deoxyadenosine) suppressed HMGA2, Twist1 and ZEB1-dependent melanoma invasion and metastasis by targeting miR-33b. ***Oncotarget***, 6(12): 9834–9853.

16. Zhang, P., Fu, C., **Hu, Y. J.**, Dong, C., Song, Y., Song, E. (2015). C_6-ceramide nanoliposome suppresses tumor metastasis by eliciting PI3K and PKCζ tumor-suppressive activities and regulating integrin affinity modulation. ***Scientific Reports***, DOI: 10.1038/srep09275.

17. Adekambi, T., Ibegbu, C. C., Cagle, S., Kalokhe, A. S., Wang, Y. F., **Hu, Y. J.**, Ray, S. M., and Rengarajan, J. (2015) Host biomarkers for diagnosis of active tuberculosis and monitoring response to treatment. ***Journal of Clinical Investigation***, doi: 10.1172/JCI77990.

18. Chan, A. W., Jiang, J., Chen, Y., Li, C., **Hu, Y. J.**, Chi, T., Moran, S., Rahim, T., Li, S., Li, X, Zola, S. M., Testa, C. M., Mao, H., Villalba, R., Smith, Y., Zhang, X. and Bachevalier, J. (2015) Progressive cognitive deficit, motor impairment and striatal pathology in a transgenic Huntington disease monkey model from infancy to adulthood. ***PLoS One***, doi: 10.1371/journal.pone.0122335.

19. Nyante, S. J., Gammon, M. D., Kaufman, J. S., Bensen, J. T., Lin, D. Y., Barnholtz-Sloan, J. S., **Hu, Y. J.**, He, Q., Luo, J., Millikan, R. C. (2015). Genetic variation in estrogen and progesterone pathway genes and breast cancer risk: an exploration of tumor subtype-specific effects. ***Cancer Causes and Control***, 26(1):121–131.

20. Franceschini, N., **Hu, Y. J.**, Reiner, A. P., Buyske, S., Nalls, M., Yanek, L. R., Li, Y., Hindorff, L. A., Cole, S. A., Howard, B. V., Andrews, J. S., *et al.* (2014). Prospective associations of coronary heart disease loci in African Americans using the MetaboChip: the PAGE study. ***PLoS One***, DOI: 10.1371/journal.pone.0113203.

21. Tan, F., Mosunjac, M., Adams, A. L., Adade, B., Taye, O., **Hu, Y. J.**, Rizzo, M. and Ofori-Acquah, S. F. (2014). Enhanced down-regulation of ALCAM/CD166 in African-American breast cancer. ***BMC Cancer***, 14:715.

22. Berndt, S. I., Gustafsson, S., Mgi, R., Ganna, A., Wheeler, E., Feitosa, M. F., Justice, A. E., Monda, K. L., Croteau-Chonka, D. C., Day, F. R., Esko, T., Fall, T., Ferreira, T., Gentilini, D., Jackson, A. U., Luan, J., Randall, J. C., Vedantam, S., Willer, C. J., Winkler, T. W., Wood, A. R., Workalemahu, T., **Hu, Y. J.**, et al. (2013). Large-scale genome-wide meta-analysis identifies 11 novel loci for anthropometric traits and provides new insights on the genetic architecture of the extremes of the distribution. ***Nature Genetics***, 45: 501–512.

23. Adisa, O. A., **Hu, Y. J.**, Ghosh, S., Aryee, D., Osunkwo, I., Ofori-Acquah, S. F. (2013). Association between plasma free haem and Incidence of vaso-occlusive episodes and acute chest syndrome in children with sickle cell disease. ***British Journal of Haematology***, 162: 702–705.

24. Nyante, S. J., Gammon, M. D., Kaufman, J. S., Bensen, J. T., Lin, D. Y., Barnholtz-Sloan, J. S., Luo, J., **Hu, Y. J.**, He, Q., Millikan, R. C. (2011). Common genetic variation in adiponectin, leptin, and leptin receptor and association with breast cancer subtypes. ***Breast Cancer Research and Treatment***, 129: 593–606.

25. Nambi, V., Chambless, L., Folsom, A. R., He, M., **Hu, Y. J.**, Mosley, T., Volcik, K., Boerwinkle, E., Ballantyne, C. M. (2010). Carotid intima-media thickness and presence or absence of plaque improves prediction of coronary heart disease risk: the ARIC (Atherosclerosis Risk In Communities) Study. ***Journal of the American College of Cardiology***, 55: 1600–1607.

26. Nambi, V., Chambless, L., Folsom, A. R., He, M., **Hu, Y. J.**, Mosley, T., Volcik, K., Boerwinkle, E., Ballantyne, C. M. (2010). Reply. ***Journal of the American College of Cardiology***, 56: 1069–1070.

27. Sullivan, P. F., de Geus, E. J., Willemsen, G., James, M. R., Smit, J. H., Zandbelt, T., Arolt, V., Baune, B. T., Blackwood, D., Cichon, S., Coventry, W. L., Domschke, K., Farmer, A., Fava, M., Gordon, S. D., He, Q., Heath, A. C., Heutink, P., Holsboer, F., Hoogendijk, W. J., Hottenga, J. J., **Hu, Y. J.**, Kohli, M., Lin, D., *et al.* (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. ***Molecular Psychiatry***, 14: 359–375.

28. Nambi, V., Hoogeveen, R. C., Chambless, L., **Hu, Y. J.**, Bang, H. J., Coresh, J., Ni, H. Y., Boerwinkle, E., Mosley, T., Sharrett, R., Folsom, A. R., Ballantyne, C. M. (2009). Lipoprotein-associated phospholipase A2 and high-sensitivity C-reactive protein improve the stratification of ischemic stroke risk in the Atherosclerosis Risk in Communities (ARIC) Study. ***Stroke***, 40: 376–381.

29. Yan, Y., **Hu, Y. J.**, North, K. E., Franceschini, N., Lin, D. Y. (2009). Evaluation of population impact of candidate polymorphisms for coronary heart disease in the Framingham heart study offspring cohort. ***BMC Proceedings***, 3(Suppl 7): S118.

Book Chapters

1. Sun, W., **Hu, Y. J.** (2014) Mapping of expression quantitative trait loci using RNA-seq data. *Statistical Analysis of Next Generation Sequencing Data*, Springer.

2. Sun, Y. V., **Hu, Y. J.** (2015) Integrative Analysis of Multi-omics Data for the Discovery and Functional Studies of Complex Human Diseases. *Advances in Genetics*, Elsevier.

Teaching

Department of Biostatistics and Bioinformatics, Emory University

- BIOS 709 Generalized Linear Models, Spring 2017, 2018, 2019, 2020, 2021.
- BIOS 508 Introduction to Categorical Data Analysis, Fall 2013, 2014, 2015, 2016.
- BIOS 516 Introduction to Large-Scale Biomedical Data Analysis, Fall 2013, 2015, 2017, 2019.
- BIOS 770 Advanced Statistics Genetics, Spring 2013, 2015, 2019.
- BIOS 570 Methods in Statistics Genetics, Spring 2013, 2015, 2017.
- BIOS 731 Advanced Statistical Computing, Fall 2012.

Graduate Students

Direction of Doctoral Research

- Yue, Ye (2020 – present). *Statistical Methods for Analyzing Microbiome Data.*
- Hu, Yingtian (2018 – present). *Statistical Methods for Compositional Analysis of Microbiome Data.*
- Zhu, Zhengyi (2017 – present). *Statistical Methods for Analyzing Clustered Microbiome Data.*
- Li, Yunxiao (2015 – 2019). *Statistical Methods for High-Dimentional Data.*
- Liao, Peizhou (2013 – 2017). *Statistical Methods for Next Generation Sequencing Studies.*
  - Livingston Award for a PhD student with excellent research potential, 2015.

Direction of Master's Research

- Jiahui, Xu (2016 – 2017). *Assessment of different pipelines for processing microbiome data*
- Sun, Zhe (2013 – 2014). *Estimating Genetic Effects When Stratification-Score Matching Is Used to Correct for Confounding by Population Stratification in Case-Control Studies.* Now Ph.D. student in the Department of Biostatistics, University of Pittsburgh.

Ph.D. Dissertation Committee

- Huang, Yunfeng (2018 – present; Chair: Yan V. Sun, Department of Epidemiology). *Gene-lifestyle Interactions in Coronary Artery Disease.*
- Daniel, Gaea Anita (2017 – 2019; Chair: Deborah Bruner in School of Nursing). *Vaginal microbiome and intravaginal practices.*
- Alhanti, Brooke (2014 – 2016; Chair: Howard Chang). *Methods for Estimating the Effect of Air Pollution on Asthma under a Changing Climate.*
- Yuan, Shuai (2012 – 2014; Chair: Zhaohui Qin). *Applying Diploid Method to Improve Read-mapping and Analysis Based on NGS Data.* Curren position: Microsoft, Seattle, WA.
- Jiang, Yunxuan (expected in 2015; Chair: Michael P. Epstein). *Statistical Methods for Rare-Variant Sequencing Studies in Pedigrees.*

Professional Activities

Journal Referee

- *Journal of American Statistical Association*, *Annals of Applied Statistics*, *Biometrics*, *Biostatistics*, *Genome Research*, *PLoS Genetics*, *Bioinformatics*, *Genetic Epidemiology*, *Statistica Sinica*, *Statistics in Medicine*, *PLoS One*, *Human Genetics*, *Statistics in Biosciences*, *Genomics*, *BMC Genetics*, *Genetic Analysis Workshop*, *Statistical Applications in Genetics and Molecular Biology*, *Environmental Microbiology*, *Frontiers in Microbiology*, *Journal of Applied Statistics*, *Frontiers in Genetics*

Panel Referee

- National Institutes of Health, BMRD (June 2020, March 2021)
- Special Emphasis Panel on Million Veteran Program (MVP), VA (July 2017)
- National Institutes of Health, National Cancer Institute, EDRN: Biomarker Developmental Laboratories (U01) (Jun 2015)
- Health and Medical Research Fund (HMRF), Hong Kong (2015, 2016, 2018)
- Department of Defense (DoD), Joint Warfighter Medical Research Program (JWMRP)– Clinical and Rehabilitative Medicine (CRM) (Jun 2014)
- Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) and Research Grants Council Earmarked Research Grants (RGC ERG) Joint Research Scheme, Hong Kong (Jul 2012)

Event Organizer

- ENAR Invited Session: *New Topics in Analysis of Microbiome Data*. (March 2021). Speakers: Drs. Huilin Li, Ni Zhao, Zhigang Li, Jichun Xie.
- Southern Regional Council on Statistics (SRCOS), Kentucky (June 2019). Program co-chair.
- Banff International Research Station (BIRS) 5-day workshop: *Emerging Statistical Challenges and Methods for Analysis of Human Microbiome Data*. (9/15/19–9/20/19). Organizers: Yi-Juan Hu, Hong Gu, Shyamal Peddada, Michael Wu. Participants: 21.
- JSM Invited Session: *Quantification, association testing, and integration of micriobiome*. Washington DC (August 2017). Speakers: Drs. Michael C. Wu, Jun Chen, Benjamin Callahan, and Huilin Li.
- ENAR Invited Session: *Integrative Analysis of Multi-Omic Data for Understanding Complex Human Diseases*. Austin (March 2016). Speakers: Drs. Kathryn Roeder, Hongyu Zhao, Nicolas Larson, and Karen Conneely.
- JSM Invited Session: *Statistical Innovations for Genetic Association Studies*. Seattle (Aug 2015). Speakers: Drs. Danyu Lin, Xihong Lin, Nilanjan Chatterjee, and Andrew Allen.
- ENAR Invited Session: *Frontiers in Statistical Genetics for Complex Trait Association*. Baltimore (March 2014). Speakers: Drs. Nilanjan Chatterjee, Michael P. Epstein, Jung-Ying Tzeng, and Glen A. Satten.

Service

Department of Biostatistics and Bioinformatics, Emory University

- Faculty Search Committee: 2020
- Tenure and Tenure-Track Monthly Meetings (Coordinator): 2020 – Present
- Faculty Search Committee (Co-Chair): 2019
- Biostatistics High Performance Computing Committee (Co-Chair): 2018 – 2020
- Methods Curriculum Ad Hoc Committee (Chair): 2017 – 2020
- Research Computing Working Group (Chair): 2017 – 2018
- PhD Admissions Committee: 2011 – Present

- PhD Qualifying Exams Committee: 2015 – 2018
- Faculty Search Committee: 2015
- Donna J. Brogan Lecture Committee: 2015
- Student Recruitment Committee: 2014
- Departmental Website Committee: 2013 – 2014
- Seminar Chair: Fall 2012

Rollins School of Public Health, Emory University

- RSPH Computation and Data Science Advisory Committee (CDAG) (Co-chair): 2018 – Present
- Shepard Award Committee: 2015 – 2016
- Education Committee: 2013 – 2015

Professional Community

- ENAR Regional Advisory Board (RAB): 01/01/2016 – 12/31/2018

Honors

- Grizzle Distinguished Alumni Award, Department of Biostatistics, University of North Carolina, Chapel Hill, NC, 2021
- Finalist for James V. Neel Young Investigator Award, International Genetic Epidemiology Society, 2015
- Teaching Award Nominee, Department of Biostatistics and Bioinformatics, Emory, 2013, 2018
- Barry H. Margolin Doctoral Dissertation Award, Department of Biostatistics, University of North Carolina, Chapel Hill, NC, 2012
- Delta Omega Induction (awarded for outstanding academic achievement for students graduating in the previous year), 2011
- Distinguished Student Paper Award, Eastern North American Region (ENAR), International Biometric Society, 2010
- Award for Best Performance in Ph.D. Qualifying Exams, Department of Biostatistics, UNC-CH, 2006
- FRYER Fellowship, Department of Biostatistics, UNC-CH, 2005
- Honor Graduate, Peking University, 2005
- Outstanding Student Award, Peking University, 2001 – 2005
- Guang Hua Scholarship, Peking University, 2003
- May Fourth Scholarship, Peking University, 2002
- Third Prize, China Mathematics Olympiad (Winter Camp), P.R.China, 2001
- First Prize, National High-school Mathematics Competitions, P.R.China, 2000

PRESENTATIONS

Invited Departmental/Institutional Seminar

1. Department of Preventive Medicine - Biostatistics, Feinberg School of Medicine, Northwestern University, IL (Oct 2020). *Testing Presence-Absence Associations in the Microbiome.*
2. Department of Biostatistics, Virginia Commonwealth University, VA (Feb 2019). *Analyzing matched sets of microbiome data using the LDM and PERMANOVA.*
3. Division of Biostatistics, Department of Population Health, School of Medicine, New York University, NY (Mar 2018). *Testing Hypotheses about the Microbiome using an Ordination-Based Linear Decomposition Model.*
4. Institute of Bioinformatics, University of Georgia, Athens, GA (Sep 2017). *Testing Rare-Variant Association without Calling Genotypes Allows for Systematic Differences in Sequencing between Cases and Controls.*
5. Department of Biostatistics, Duke University, Durham, NC (Feb 2017). *Testing Rare-Variant Association without Calling Genotypes Allows for Systematic Differences in Sequencing between Cases and Controls.*
6. The 4th Workshop on Biostatistics and Bioinformatics, Georgia State University, Atlanta, GA (May 2015). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
7. Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC (May 2015). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
8. Department of Genetics, Rutgers University, New Brunswick, NJ (March 2015). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
9. Center for Integrative Proteomics Research, Rutgers University, New Brunswick, NJ (Feb 2015). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
10. Department of Statistics and Biostatistics, Rutgers University, New Brunswick, NJ (Feb 2015). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
11. Department of Statistics, Purdue University, West Lafayette, IN (Feb 2015). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
12. Department of Statistics, George Mason University, Fairfax, VA (Oct 2014). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
13. Department of Statistics, University of Georgia, Athens, GA (Apr 2014). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
14. Department of Statistics, Georgia State University, Atlanta, GA (Sep 2013). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.*
15. Fudan University, Shanghai, China (Jul 2013). *Integrative Analysis of Sequencing and GWAS Data in Detecting Rare Variants Associated With Complex Diseases.*
16. School of Lift Sciences and Technology, Tongji University, Shanghai, China (Jul 2013). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.*
17. Department of Mathematical Science, Middle Tennessee State University, Murfreesboro, TN (Apr 2013). *Meta-Analysis of Rare Variants in Genetic Association Studies.*
18. Cancer Genetics and Epigenetics Seminar, Winship Cancer Institute, Atlanta, GA (Jun 2012). *Integrative Analysis of Sequencing and GWAS Data Improves Statistical Power in Detecting Rare Variants Associated With Complex Diseases.*
19. Department of Biostatistics, Vanderbilt University, Nashville, TN (Mar 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*

20. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada (Mar 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
21. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA (Feb 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
22. Department of Biostatistics, University of Buffalo, Buffalo, NY (Feb 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
23. Department of Biostatistics, University of Michigan, Ann Arbor, MI (Jan 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
24. Statistical Genetics Seminar Series, University of North Carolina at Chapel Hill, NC (Oct 2009). *A Robust Statistical Method for Association Testing with Copy Number Variation.*

Conference/Workshop Talks

1. 13th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics), Virtual Conference (Dec 2020). *Testing Presence-Absence Associations in the Microbiome.* {**Invited**}
2. ENAR, Vanderbilt, TN (Mar 2020). *A logistic regression for testing differential abundance in compositional microbiome data.* {**Invited**}
3. Banff International Research Station (BIRS) 5-day workshop, (9/15/19–9/20/19). *Analyzing Matched Sets of Microbiome Data using LDM.* {**Contact organizer**}
4. JSM, Denver, CO (Jul 2019). *Analyzing Matched Sets of Microbiome Data using LDM.* {**Invited**}
5. WNAR, Portland, OR (June 2019). *Analyzing Matched Sets of Microbiome Data using LDM.* {**Invited**}
6. ENAR, Pennsylvania, PA (Mar 2019). *Analyzing Matched Sets of Microbiome Data using LDM.* {**Invited**}
7. 11th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics), Pisa, Italy (Dec 2018). *Analyzing Matched Sets of Microbiome Data.* {**Invited**}
8. American Society of Human Genetics (ASHG), San Diego, CA (Oct 2018). *Testing Hypotheses about the Microbiome using an Ordination-Based Linear Decomposition Model.*
9. Mathematical Statistics (IMS) Asia Pacific Rim Meeting (IMS-APRM2018), Singapore (June 2018). *Testing Presence-Absence Association in the Microbiome using the Linear Decomposition Model (LDM).* {**Invited**}
10. Southern Regional Council on Statistics (SRCOS), Virginia Beach, VA (June 2018). *Testing Hypotheses about the Microbiome using an Ordination-Based Linear Decomposition Model.* {**Invited**}
11. International Indian Statistical Association (IISA), Gainesville, FL (May 2018). *Testing Presence-Absence Association in the Microbiome using the Linear Decomposition Model (LDM).* {**Invited**}
12. American Society of Human Genetics (ASHG), Orlando, FL (Oct 2017). *Robust inference of population structure from next-generation sequencing data with systematic differences in sequencing.*
13. JSM, Baltimore, MA (Aug 2017). *Robust inference of population structure from next-generation sequencing data with systematic differences in sequencing* {**Invited**}
14. American Society of Human Genetics (ASHG), Vancouver, BC Canada (Oct 2016). *Testing rare-variant association without calling genotypes allows for systematic differences in sequencing between cases and controls.* {**Invited platform presentation**}

15. American Society of Human Genetics (ASHG), Baltimore, MD (Oct 2015). *Proper Use of Allele-Specific Expression Improves Statistical Power for Cis-eQTL Mapping with RNA-seq Data.* {**Invited platform presentation**}
16. International Genetics Epidemiology Society (IGES), Baltimore, MD (Oct 2015). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.* {**Invited as one of the three finalists for James V. Neel Young Investigator Award**}
17. ICSA, Fort Collins, CO (Jun 2015). *Proper Use of Allele-Specific Expression Improves Statistical Power for Cis-eQTL Mapping with RNA-seq Data.* {**Invited**}
18. American Society of Human Genetics (ASHG), San Diego, CA (Oct 2014). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
19. ENAR, Baltimore, MD (Mar 2014). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.* {**Invited Poster**}
20. ASHG, Boston, MA (Oct 2013). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.*
21. International Genetics Epidemiology Society (IGES), Chicago, IL (Sep 2013). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.* {**Invited**}
22. The Second Taihu International Statistics Forum, Wuxi, Jiangsu, China (Jul 2013). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.* {**Invited**}
23. IMS-China, Chengdu, China (Jul 2013). *Integrative Analysis of Sequencing & GWAS Data in Detecting Rare Variants Associated With Complex Diseases.* {**Invited**}
24. ENAR, Orlando, FL (Mar 2013). *Analysis of Ultra High-Dimensional Data in Imaging Genetics Studies.*
25. ASHG, San Francisco, CA (Nov 2012). *Integrative Analysis of Sequencing and GWAS Data Improves Statistical Power in Detecting Rare Variants Associated With Complex Diseases.*
26. Genetic Analysis Workshop (GAW), Stevenson, WA (Oct 2012). *Genetic Association of Common and Rare SNPs with Longitudinal Measurements of Blood Pressure.*
27. JSM, San Diego, CA (Aug 2012). *Accounting for Missing Data in eQTL Mapping.* {**Invited**}
28. ENAR, Washington DC (Apr 2012). *Association Analysis of Rare Variants with Selective Sequencing.*
29. International Congress of Human Genetics (ICHG), Montreal, Canada (Oct 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
30. ICSA, New York, NY (Jun 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
31. ENAR, Miami, FL (Mar 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
32. ASHG, Washington, DC (Nov 2010). *Comprehensive Assessment of Methods for Untyped SNP Analysis.*
33. JSM, Vancouver, BC Canada (Aug 2010). *Analysis of Untyped SNPs: Maximum Likelihood and Imputation Methods.*
34. ENAR, New Orleans, LA (Mar 2010). *A General Framework for Studying Genetic Effects and Gene-Environment Interactions with Missing Data.*
35. ASHG, Honolulu, HI (Oct 2009). *Analysis of Genetic Effects and Gene-Environment Interactions under A Flexible Model for Gene-Environment Dependence.*
36. ENAR, St. Antonio, TX (Mar 2009). *Accounting for Gene-Environment Dependence in Genetic Association Studies.*
37. IGES, St. Louis, MO (Sep 2008). *Likelihood-Based Inference on Haplotype Effects and Haplotype-Environment Interactions.*

Software

LDM (Linear Decomposition Models)

- An R package that provides a single analysis path that includes distance-based ordination, global tests of any effect of the microbiome, and tests of the effects of individual OTUs (i.e., operational taxonomic units) with false discovery rate (FDR)-based correction for multiple testing.
  `http://github.com/yijuanhu/LDM`

BOUTH (Bottom-up Tree Hypothesis Tests)

- An R package for testing hypotheses that have a branching tree dependence structure in a bottom-up manner, with false discovery rate control.
  `https://github.com/yli1992/BOUTH`

TASER-PC (Test of Association using SEquencing Reads)

- A C++ program for robust inference of population structure from next-generation sequencing data with systematic differences in sequencing.
  `http://web1.sph.emory.edu/users/yhu30/software.html`

TASER (Test of Association using SEquencing Reads)

- A C++ program for testing association using sequencing reads without calling genotypes.
  `http://web1.sph.emory.edu/users/yhu30/software.html`

PhredEM

- A C++ program for phred-score-informed genotype calling.
  `http://web1.sph.emory.edu/users/yhu30/software.html`

TRECASE_MLE

- A C++ program for *cis*-eQTL mapping with RNA-seq data.
  `http://web1.sph.emory.edu/users/yhu30/software.html`

SEQGWAS (integrative analysis of SEQuencing and GWAS data)

- A C++ program for integrative analysis of sequencing and GWAS array data.
  `http://web1.sph.emory.edu/users/yhu30/software.html`

MAGA (Meta-Analysis of Gene-level Associations)

- A C++ program for meta-analysis of gene-level associations based on single-variant statistics of rare variants from participating studies.
  `http://web1.sph.emory.edu/users/yhu30/software.html`

CNVstat

- A C++ program for the genome-wide statistical association analysis of CNVs and Allele-Specific Copy Numbers
  `http://www.bios.unc.edu/~lin/software/CNVstat/`

SNPMStat

- A C++ program for the genome-wide statistical analysis of SNP-disease association with potentially missing genotype data, including untyped SNPs.
  `http://www.bios.unc.edu/~lin/software/SNPMStat/`

tagIMPUTE (tag-based IMPUTE)

- A C++ program for the genome-wide imputation of unobserved genotypes in SNP-disease association studies.
  `http://www.bios.unc.edu/~lin/software/tagIMPUTE/`

Computational Skills

- Programming Languages: proficient in C, C++, knowledge of PERL and PYTHON
- Software: proficient in R, SAS

- Operating Systems: Unix and Windows

Professional Memberships

- American Statistical Association (2010 – Present)
- Eastern North American Region, International Biometric Society (2007 – Present)
- American Society of Human Genetics (2008 – Present)
- International Genetic Epidemiology Society (2008 – Present)
- International Chinese Statistical Association (2011 – Present)
- International Indian Statistical Association (2011 – Present)