Yijuan Hu

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University
1518 Clifton Road NE, 3rd floor
Atlanta, Georgia USA 30322

Cell: (919) 357-6882
Office: (404) 712-4466
Fax: (404) 727-1370
E-mail: yijuan.hu@emory.edu

Research Interests

Statistical genetics and genomics. My research focuses on the development of statistical methods for analysis of high-throughput genetic, genomic and other "omics" data in epidemiological and clinical sciences. Specific topics include: meta-analysis; integrative analysis of sequencing and GWAS array data; analysis of sequencing reads without calling genotypes; eQTL mapping with RNA-seq data; analysis of DNA methylation data in EWAS and meQTL studies; and integrative analysis of multi-omics data.

Professional Experience

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

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

- 08/2014–07/2016. R03AI111396-01 (Hu) National Institutes of Health. *Epigenome-wide Association Study of Asthma in Populations of African Descent.* Role: Principal Investigator.
- 05/2014-04/2018. R01GM105785-01A1 (Sun) National Institutes of Health. *Statistical Methods for RNA-seq Data Analysis.* Role: co-Investigator.
- 08/2012-07/2017. P30AI050409-15 (Curran) National Institutes of Health. *Emory/Atlanta Center for AIDS Research (CFAR).* Role: co-Investigator.

Pending

- 04/2015-03/2020. R01GM116065-01 (Hu) National Institutes of Health. *Association Tests of Rare Variants Using Sequence Reads without Calling Genotypes.* Role: Principal Investigator.

Publications

Published/In Press

- **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*, in press.
- **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.

- 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)

- 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*, in press.

- Johnston, H. R., **Hu, Y. J.**, and Cutler, D. J. Population genetics identifies challenges in analyzing rare variants. *Genetic Epidemiology*, in press.

- 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. (2014). Genetic variation in estrogen and progesterone pathway genes and breast cancer risk: an exploration of tumor subtype-specific effects'. *Cancer Causes and Control*, in press.

- 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.

- **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.

- 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.

- 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.

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

- 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.

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

- **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.
  [⋆⋆ **An earlier version won 2010 ENAR Distinguished Student Paper Award** ⋆⋆]

- 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.

- 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.

- 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.

- 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.

- 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.

- 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.

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

Under Revision

- **Hu, Y. J.**, Li, Y., Auer, P. L., Lin, D. Y. Integrative analysis of sequencing and array genotype data for discovering disease associations with rare mutations. Under revision for *Proceedings of the National Academy of Sciences*.

- **Hu, Y. J.**, Sun, W., Tzeng, J. Y., Perou, C. M. Proper use of allele-specific expression improves statistical power for *cis*-eQTL mapping with RNA-seq data. Under revision for *Journal of American Statistical Association*.

- Adekambi, T., Ibegbu, C. C., Cagle, S., Kalokhe, A. S., Wang, Y. F., **Hu, Y. J.**, Ray, S. M., and Rengarajan, J. Host biomarkers for diagnosis of active tuberculosis and monitoring response to treatment. Under revision for *Journal of Clinical Investigation*.

Submitted

- Mathias, R. A., Taub, M., O'Connor, T. D., Fu, W., et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome.

- Yuan, S., Johnston, H. R., Zhang, G., Li, Y., **Hu, Y. J.**, Qin, Z. Using customized diploid reference genome improves reads mapping and genotype calling in whole genome sequencing studies.

- Prucha, M. S., Jiang, J., **Hu, Y. J.**, Chi, T., Moran, S., Rahim, T., Li, C., Conneely, K., Mao, H., Zhang, X., Bachevalier, J., Chan, A. W. S. Longitudinal metabolomics profiling and neuroimaging studies in transgenic

- Chan, A. W. S., Jiang, J., Chen, Y., Li, C., **Hu, Y. J.**, Chi, T., Moran, S., Rahim, T., Li, S., Li, X, Zola, S. M., Testa, C., Mao, H., Smith, Y., Zhang, X. and Bachevalier, J. Progressive cognitive deficit, motor impairment and neural anatomical changes in Huntingtons disease monkeys.

- Jiang, J., **Hu, Y. J.**, Rahim, T., Mao, H., Zhang, X., and Chan, A. W. S. White matter degeneration and volumetric loss in Huntingtons disease monkeys: a longitudinal neuroimaging study.

Book Chapters

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

Teaching

Department of Biostatistics and Bioinformatics, Emory University

- BIOS 508 Introduction to Categorical Data Analysis, Fall 2013, Fall 2014.
- BIOS 516 Introduction to Large-Scale Biomedical Data Analysis, Fall 2013.
- BIOS 770 Advanced Statistics Genetics, Spring 2013, Spring 2015.
- BIOS 570 Introduction to Statistics Genetics, Spring 2013, Spring 2015.
- BIOS 731 Advanced Statistical Computing, Fall 2012.

Graduate Students

Direction of Doctoral Research

- Liao, Peizhou (2013 – present). *Statistical Methods for Next Generation Sequencing Studies.*

Direction of Master's Research

- 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

- Yuan, Shuai (2012 – 2014; Chair: Zhaohui Qin). *Applying Diploid Method to Improve Read-mapping and Analysis Based on NGS Data.* 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*, *Biometrics*, *Biostatistics*, *Bioinformatics*, *Genetic Epidemiology*, *Human Genetics*, *Genome Research*, *Statistics in Biosciences*, *PLoS One*, *Statistica Sinica*, *Genomics*, *Annals of Applied Statistics*

Panel Referee

- Department of Defense (DoD), *Joint Warfighter Medical Research Program (JWMRP)– Clinical and Rehabilitative Medicine (CRM)* (Jun 2014)

Event Organizer

- Invited Session: *Frontiers in Statistical Genetics for Complex Trait Association.* ENAR, Baltimore (March 2014). Speaker list: Drs. Nilanjan Chatterjee, Michael P. Epstein, Jung-Ying Tzeng, and Glen A. Satten.
- Invited Session: *Statistical Innovations for Genetic Association Studies.* JSM, Seattle (Aug 2015). Speaker list: Drs. Danyu Lin, Xihong Lin, Nilanjan Chatterjee, and Andrew Allen.

Service

Department of Biostatistics and Bioinformatics, Emory University

- PhD Admissions Committee: 2011 – Present
- Seminar Chair: Fall 2012
- Departmental Website Committee: 2013 – Present
- Student Recruitement Committee: 2014 – Present

Rollins School of Public Health, Emory University

- Education Committee: 2013 – Present

Honors

- Rollins Assistant Professorship, Emory University, 2011 – present
- Teaching Award Nominee, Department of Biostatistics and Bioinformatics, Emory, 2013
- 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

- Department of Statistics, George Mason University, Fairfax, VA (Oct 2014). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
- Department of Statistics, University of Georgia, Athens, GA (Apr 2014). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
- Department of Statistics, Georgia State University, Atlanta, GA (Sep 2013). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.*
- Fudan University, Shanghai, China (Jul 2013). *Integrative Analysis of Sequencing and GWAS Data in Detecting Rare Variants Associated With Complex Diseases.*
- 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.*
- Department of Mathematical Science, Middle Tennessee State University, Murfreesboro, TN (Apr 2013). *Meta-Analysis of Rare Variants in Genetic Association Studies.*
- 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.*
- Department of Biostatistics, Vanderbilt University, Nashville, TN (Mar 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*

- 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.*
- Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA (Feb 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
- Department of Biostatistics, University of Buffalo, Buffalo, NY (Feb 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
- Department of Biostatistics, University of Michigan, Ann Arbor, MI (Jan 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
- 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

- American Society of Human Genetics (ASHG), San Diego, CA (Oct 2014). *Integrative Analysis of Sequencing and Genotype Array Data for Rare Variant Associations.*
- ENAR, Baltimore, MD (Mar 2014). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.* {Invited Poster}
- ASHG, Boston, MA (Oct 2013). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.*
- International Genetics Epidemiology Society (IGES), Chicago, IL (Sep 2013). *Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics.* {Invited}
- 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}
- IMS-China, Chengdu, China (Jul 2013). *Integrative Analysis of Sequencing & GWAS Data in Detecting Rare Variants Associated With Complex Diseases.* {Invited}
- ENAR, Orlando, FL (Mar 2013). *Analysis of Ultra High-Dimensional Data in Imaging Genetics Studies.*
- ASHG, San Francisco, CA (Nov 2012). *Integrative Analysis of Sequencing and GWAS Data Improves Statistical Power in Detecting Rare Variants Associated With Complex Diseases.*
- Genetic Analysis Workshop (GAW), Stevenson, WA (Oct 2012). *Genetic Association of Common and Rare SNPs with Longitudinal Measurements of Blood Pressure.*
- JSM, San Diego, CA (Aug 2012). *Accounting for Missing Data in eQTL Mapping.* {Invited}
- ENAR, Washington DC (Apr 2012). *Association Analysis of Rare Variants with Selective Sequencing.*
- International Congress of Human Genetics (ICHG), Montreal, Canada (Oct 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
- ICSA, New York, NY (Jun 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
- ENAR, Miami, FL (Mar 2011). *A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers.*
- ASHG, Washington, DC (Nov 2010). *Comprehensive Assessment of Methods for Untyped SNP Analysis.*
- JSM, Vancouver, BC Canada (Aug 2010). *Analysis of Untyped SNPs: Maximum Likelihood and Imputation Methods.*
- ENAR, New Orleans, LA (Mar 2010). *A General Framework for Studying Genetic Effects and Gene-Environment Interactions with Missing Data.*

- ASHG, Honolulu, HI (Oct 2009). *Analysis of Genetic Effects and Gene-Environment Interactions under A Flexible Model for Gene-Environment Dependence.*
- ENAR, St. Antonio, TX (Mar 2009). *Accounting for Gene-Environment Dependence in Genetic Association Studies.*
- IGES, St. Louis, MO (Sep 2008). *Likelihood-Based Inference on Haplotype Effects and Haplotype-Environment Interactions.*

Software

SEQGWAS (integrative analysis of SEQuencing and GWAS data)
- A C++ program for integrative analysis of sequencing and GWAS 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`

asSeq
- An R package for *cis*-eQTL mapping with allele-specific RNA-seq data.
  `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)