Hongyan Xu, PhD
Associate Professor

Department of Biostatistics
College of Graduate Studies
Georgia Health Sciences University

Synopsis of Faculty Activities Profile

Synopsis of Faculty Activities Profile

Hongyan Xu, Ph.D.

This report summaries my faculty activities at my current position, including teaching, research and scholarly activities. I am currently a tenure-tracked associate professor in the Department of Biostatistics at the Georgia Health Sciences University (GHSU) in Augusta, GA. I received my Ph.D. in human population genetics from the Graduate School of Biomedical Sciences, the University of Texas Health Science Center at Houston. Then I underwent further post-doctoral training in statistical genetics/genetic epidemiology at the University of Texas MD Anderson Cancer Center. I joined the Department of Biostatistics at the Georgia Health Sciences University in November, 2005 and was promoted to associate professor in July 2011. At my current position, I have mentored 9 graduate students in Biostatistics program, and 2 students in the Clinical and Translational Science program. I am the main instructor of 3 graduate courses for the Biostatistics students and co-instructor for 1 graduate course for both students in the Biostatistics program and the Genomic Medicine PhD program, and 1 graduate course for the Genomic Medicine PhD program. I have 29 peer-reviewed publications, almost half of which are published when I am at my current position, mostly as leading or corresponding author. I am the principle investigator of a Scientist Training Program project (STP00004) from GHSU, which supports my investigation of statistical properties of population structure and its relation with genetic association studies. Based on the results from the STP, I am awarded an R21 grant (R21 NS057506) from NIH to investigate the impact of hidden population structure and admixture on the statistical testing of disease genes associations, especially in large scale genome-wide studies. Through the grant, I have developed novel statistical approaches for large-scale genetic association studies accounting for hidden population structure, which is especially critical for genome-wide association studies in recently admixed populations such as African Americans and Hispanics. I have extended my research on the genomics and epi-genomics and submitted an R01 grant on methodology development for studying methylation pattern for complex diseases based on functional data analysis. I serve as co-investigator on one NIH/NHLBI funded P01 project, one NIH/NCMHD funded P20 project, one NIH/NCI funded R01 project and one NIH/NCI funded R21 project. I also serve as co-investigators on 5 pending NIH grant applications, principle investigator of one pending subcontract from Department of Defense grant application. I have constantly provided statistical consulting services to both internal and external clients through the departmental Consulting and Survey Center. I have served at NIH special emphasis review panels for 2011 and 2012. I have served from two years as the chair of the departmental committee for Seminar and Journal Club. I am now the chair of the departmental committee for computing. I am also a member of GHSU's Institutional Animal Care and Use Committee (2010-2012) and a member of the Data Collection and Harmonization group of the Registry and Surveillance System for Hemoglobinopathies (RuSH) program by NIH and CDC. I am a member of the American Statistical Society (ASA), American Society of Human Genetics, and Sigma-Xi, the Scientific Research Society. I attend and present at the annual meetings and ASA local chapter meetings regularly. I have reviewed papers for journals such as *Biometrics*, *PloS One*, *PloS Genetics*, and *American Journal of Human Genetics*. As a summary of effort distributions, 65% of my efforts go to research, 25% to teaching, 5% to consulting, and 5% to intramural and extramural services.

TEACHING

At my position, I have served as advisor or the committee member of 9 MS/PhD students in the Biostatistics program. The names of the students are listed in the Educator's Profile. Currently, I am a dissertation committee member of Mr. Daniel Linder, a PhD student in Biostatistics. I am also the biostatistics advisor of two CTS students, Dr. Michelle Reid-Nicholson and Dr. Cheedy Jaja. Together with Dr. Waller, the graduate program director, we make sure they are on the right track in his study and resolve any issues in a timely fashion. I also serve as an advisory committee member for graduate students in the Biomedical Ph.D. in Genomic Medicine program.

I have taught the four courses for the students in the Biostatistics Program.

- STAT 8880 Special Topics: Introduction to Statistical Genetics. This is a new course. The goal of the course is to create a new statistical genetics course for all second-year biostatistics students and students who perceive as relevant. This is the first time such a course has been offered. The predecessor course consistently received “very poor student evaluations”. I developed the course, organized lectures and weekly homework problems for 3 credit hour course to introduce basic concepts of Statistical Genetics in Fall 2009.
- STAT 9150 Advanced Statistical Methods in Genetic Analysis. This is also a new course. It is for PhD students who have passed the candidacy examination. The goal is to introduce advanced statistical and computational methods that are commonly used in genetic analysis. It is a 3 credit hour course. I worked with Dr. Varghese George as the main instructors for the course in Spring 2009.
- STAT 8510 Programming for Data Analysis. This is a core course for PhD/MS students in the first year. The goal is to provide a hands-on exposure to programming, data management and report generation with two of the most popular statistical software packages, SAS and R. It is a 3 credit hour course. I was the course director in Fall 2010 and Fall 2011.
- STAT 8550 High Throughput Data Analysis. This is a new course. It is a core course for PhD/MS students. Statistical analytic methods for high throughput data are in high demand in biostatistics for the current biomedical data generated from high throughput platforms. The goal of the course is to prepare students the required statistical techniques for the analysis of such data. It is a 3 credit hour course. I am the course director and main instructor. I developed the course, organized lectures and weekly homework problems in Summer 2012.

In addition, I co-taught a 3 credit hour course GNMD8050 Computational Methods in Genetics & Genomics, which is offered for both students in the Biostatistics Program and the Genomic Medicine program. I also taught in another 3 credit hour course: GNMD8051 Translational Genomics and Proteomics for the Genomic Medicine Program students.

RESEARCH and SCHOLARLY ACTIVITIES

My methodological research is in the field of statistical genetics/genomics and bioinformatics. Specifically, I am interested in developing and implementing novel statistical and computational methods for the analysis of human genetic/genomic data. Genetic dissection of complex diseases through gene mapping is an important theme in this field, often referred to

as genetic epidemiology. The data could be family-based or population-based. Recent development of high-throughput genotyping technology and next-generation sequencing (NGS) generate genetic data at unprecedented genomic level, which produce a big challenge in data analysis and call for new statistical and bioinformatics methods. I have worked on several methodological research projects, including (1): genome-wide association methods to find susceptibility genes for complex disease using family data; (2): developing methodology in genome-wide association study accounting for the effects of population structure in admixed populations; (3): methods for an ascertainment bias correction; (4): a markov chain approach for haplotype-based association; (5): new formal tests for linkage disequilibrium for genotype data; (6): statistical test for genetic association using sequence data from NGS accounting for rare genetic variants; and (7) statistical tests for differential methylation pattern for cancer and complex diseases.

My second important research topic is in evolutionary genetics and genomics. The new statistical methods such as those for detecting signatures of natural selection utilize genomic data and statistical testing to find regions of the genome that are under Natural Selection. They are of importance for functional genomics because only functional loci will be subject to the forces of Natural Selection. The data could come from humans or multiple related species. If genomic data from multiple species are involved, it is often referred to as comparative genomics. Naturally, since human diseases, especially complex diseases are an important force of Natural Selection, research topics in evolutionary genetics are complementary to the topics in dissecting the genetics basis of human diseases. Major evolutionary forces at the population level such as population admixture and structure, population size changes play important role in determine the disease gene frequencies and distribution in different human populations, which in turn determine the disease prevalence and affects the statistical power in disease gene mapping.

Since my appointment at my current position, I have been award the following grants for my methodological research:

STP 00004, Georgia Health Sciences University, 06/01/06-05/31/09
Effects of Population Structure on Genetic Association Studies
Scientist Training Program: to determine the role of population structure on genetic association studies.
Role: PI

R21 NS057506, NIH/NINDS, 04/01/08 - 03/31/10
Association Study of Stroke Risk in an Admixed Population of African Americans
This study is to develop new methodology in large-scale association studies of stroke risk in admixture African American populations.
Role: PI

I have set up a Genomics Interest Group since 2011 consisting of several researchers around the campus. We are working on the methodological and applied research on the genomic data from NGS and Illumina 450K for CLL and obesity. We have submitted 3 manuscripts summarizing the results from our research so far. We are also submitting an R01 proposing further methodological research for detecting differential methylation pattern in a genomic feather of interest such as a CpG island or promoter and enhancer regions.

Since my appointment, I also actively participated in collaborative research. I work collaboratively with the researchers in the Program Project grant from Georgia Prevention Institute, which is funded by NIH/NHLBI. I am a member of the statistical core of the grant. I'm collaborating with Dr. Abdullah Kutlar from the GHSU Comprehensive Sickle Cell Center on his P20 grant as a co-Investigator in the administrative core for statistical needs of the 3 sub-projects. Recently, we submitted one R21 grant and one U54 grant to NIH, in which I serve as co-investigator.

I am also working with Dr. Huidong Shi and Dr. Keith Robertson from the GHSU Cancer Center on their projects involving bioinformatics research on epigenomics of cancer. I am the co-Investigators on two of their grants currently and we have submitted two grants application with me. I'm collaborating with another GHSU Cancer Center investigator, Dr. Lesleyann Hawthorn on her projects of triple-negative breast cancer research and have submitted two NIH grants as a co-Investigator. I am also collaborating with Dr. Nahid Mivechi from the GHSU Center for Molecular Chaperone/Radiobiology and Cancer Virology. We submitted a P01 Program Project Grant together, which is currently under review.

These collaborative research projects are mutually beneficial and very productive. They not only result in grants and publications, they exposed me to new applied research field and new data sets, which could potentially expand my methodological research. The following grants are supporting these collaborative research efforts:

P01 HL069999, NIH/NHLBI PI: Harshfield, 10/01/07 – 09/30/12
Stress-Related Mechanisms of Hypertension Risk in Youth
Objectives: The overarching goal of the PPG is to evaluate the interrelationships among stress related behavioral, biological and genetic factors pertaining to the pathogenesis of hypertension.
Role: Co-investigator

P20 MD003383, NIH/MCMHD, PI: Kutlar and Gibson, 05/28/09 – 12/31/13
NCMHD Southeastern Exploratory Sickle Cell Center of Excellence
Objectives: Relieving the health disparity of SCD patients is the primary goal of this project.
Role: Co-investigator

R01 CA114229 , NIH/NCI, PI: Robertson, 09/01/05 – 06/30/11
De Novo Methyltransferase Function in Chromatin and Cancer
Objectives: The major goal is to test the hypothesis that de novo methyltransferase DNMT3B is a major regulator of genomic DNA methylation patterns in normal cells and that disruption of its functions contributes to DNA methylation defects in cancer.
Role: Co-investigator

R03 CA123565, NIH/NCI, PI: Shi, 07/01/07 – 03/31/10
Integrated Genetic and Epigenetic Biomarkers for Molecular Epidemiology
Objectives: The major goal is to understand the relationship between genetic variation, global methylation patterns and regional hypermethylation of tumor suppressor genes
Role: Co-investigator

I have over 30 peer-reviewed publications. They are listed below in chronological order. Those that are published since appointment are in italalics.

1. Bao Y, Lu D, Xu H, Shi Q, Qiu X, Xue J. Polymorphism of DXS102 locus in Chinese population and its application to gene diagnosis in hemophilia B family. Chin Med J (Engl) 111:527-230, 1998.

2. Bao Y, Lu D, Shi Q, Xu H, Qiu X, Xue J. [Determination of the polymorphism of DXS102 locus and its application in gene diagnosis]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 15:27-30, 1998.

3. Zhang W, Hu F, Xiao J, Xu H, Lu D-R, Jin L. The Distribution of a 3'A polymorphism of SDF-1 gene in a Chinese random population. Journal of Fudan Univeristy (Natural Science) 37:317--318, 1998.

4. Luo J, Ji Y, Peng Y, Xiao J, Yao Y, Xu H, Yang M, Zhen J, Lu D, Jin L. [Linkage analysis of chromosome 5 and asthma in a Chinese
population]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 16:318-320, 1999.

5. Yuan W, Xu H, Zhao J, Ding W, Jiang H, Gu M, Xue J, Chen J, Fang F, Chen Z, Jin L, Huang W. [Information behavior of microsatellite loci in genome scanning]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 17:65-71, 2000.

6. Zhao J, Wang H, Xiong M, Huang W, Zuo J, Chen Z, Qiang B, Sun Q, Li Y, Liu Q, Du W, Chen J, Ding W, Yuan W, Zhao Y, Xu H, Jin L, Fang F. The localization of type 2 diabetes susceptibility gene loci in northern Chinese Han families. Chinese Science Bulletin 45:1792-1795, 2000.

7. Xiao J, Hu F, Xu H, Su B, Jiang Y, Luo J, Zhang W, Tan J, Jin L, Lu D. Provincial distribution of three HIV-1 resistant polymorphisms (CCR5-D32, CCR2-64I, and SDF1-3'A) in China. Science in China 43: 16-20, 2000.

8. Hong W, Cai G, Xu H, Chen H, Xiao J, Lu D, Xue J, Qiu X, Jin L. [Single nucleotide polymorphism in beta2-adrenoceptor gene and the distribution in Chinese Han ethnic group]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 18:1-3, 2001.

9. Wu H, Wang H, Li H, Oshuaakey J, Xiao F, Ke Y, Xu H, Xiao J, Lu D, Parra E, Shriver M, Xiong M, Barton SA, Hewett-Emmett D, Liu W, Jin L. Skin reflectance in the Han Chinese and Tibetan populations. Hum Biol 73: 461-466, 2001.

10. Xu H, Fu Y-X. Estimating Effective Population Size or Mutation Rate with Microsatellites. Genetics 166:555-563, 2004.

11. Cortes-Prieto L, Baltazar L, Perea F, Gallegos-Arreola M, Flores S, Sandoval L, Olivares1 N, Xu H, Barton S, Chakraborty R, Rivas F. HLA-DQB1, -DQA1, -DRB1 Linkage Disequilibrium Estimate from Segregating Haplotypes in Mestizo Families from Guadalajara, Mexico. Tissue Antigens 63:458-465, 2004.

12. Xu H, Wu X, Spitz MR, Shete S. Comparison of haplotype inference methods from

unrelated population genotype data. Human Heredity 58:63-68, 2004.

13. Zhao J, Xiong M, Huang W, Wang H, Zuo J, Wu GD, Chen Z, Qiang BQ, Zhang ML, Chen JL, Ding W, Yuan WT, Xu H, Jin L, Li YX, Sun Q, Liu QY,
Boerwinkle E, Fang FD. An autosomal genomic scan for loci linked to type 2 diabetes in Northern Han Chinese population. Journal of Molecular Medicine 83:209-215, 2005.

14. Xu H, Spitz MR, Amos CI, Shete S. Complex segregation analysis reveals a multigene model for lung cancer. Human Genetics, 116:121-127, 2005

15. Xu H, Chakraborty R, Fu Y-X. Mutation rate variation at human dinucleotide microsatellites. Genetics 170:305-312, 2005

16. Xu H, Shete, S. Effects of population structure on genetic association studies. BMC Genetics 6(Suppl 1):S109, 2005

17. Xu H, Shete, S. Mixed-effects Logistic Approach for Association Following Linkage Scan for Complex Disorders. Annals of Human Genetics, 71:230-237, 2006.

18. Shekhawat PS, Srinivas SR, Matern D, Bennett MJ, Boriack R, George V, Xu H, Prasad PD, Roon P, Ganapathy V. Spontaneous development of intestinal and colonic atrophy and inflammation in the carnitine-deficient jvs (OCTN2(-/-)) mice. Mol Genet Metab.92:315-324, 2007.

19. Tan Y-D, Fonage M, George V, Xu H, Parent-child pair design for detecting gene-environment interactions in complex diseases. Human Genetics. 121:745-757, 2007

20. Xu H, George V, A new transmission test for affected sibpair families. BMC Proceedings, 1:S32, 2007.

21. Gao L, Xu H, Comparisons of Mutation Rate Variation at Genome-wide Microsatellites: Evolutionary Insights from Two Cultivated Rice and Their Wild Relatives, BMC Evolutionary Biology, 8:11, 2008

22. Tan YD, Fornage M, Xu H, Ranking analysis of F-statistics for microarray data. BMC Bioinformatics. 2008, 9:142.

23. Ellison GL, Weinrich SP, Lou M, Xu H, Powell IJ, Baquet CR., A randomized trial comparing web-based decision aids on prostate cancer knowledge for African-American men. Journal of National Medical Association, 2008, 100: 1139-1145

24. Xu H, Sarkar B, George V, A new measure of population structure using multiple single nucleotide polymorphisms and its relationship with FST, BMC Research Notes, 2009, 2:21.

25. Xu H, Mathew G, George V, Family-based genome-wide association study for Simulated Data of Framingham Heart Study. BMC Proceedings, 2009, 3:S124

26. Mathew G, Xu H, George V, Simultaneous Analysis of all SNPs in Genome-Wide Association Study of Rheumatoid Arthritis BMC Proceedings, 2009, 3:S11

27. Xu H, George V, Assessment of population structure and its effects on genome-wide association studies. Communications in Statistics – Theory and Methods, 2009, 38:2843-2855

28. Mukhopadhyay S, George V, Xu H, Variable selection method for quantitative trait analysis based on parallel genetic algorithm. Annals of Human Genetics, 2010, 74:88-96

29. Nandram B, Choi JW, Xu H, Maximum likelihood estimation for ascertainment bias in sampling siblings. Journal of Data Science, 2011, 9:23-41

30. Nandram B, Xu H. Bayesian Corrections of a Selection Bias in Genetics. Journal of Biometrics & Biostatistics 2: 112, DOI:10.4172/2155-6180.1000112, 2011.

31. Xu H, George V. A Monte Carlo test of linkage disequilibrium for single nucleotide polymorphisms. BMC Research Notes 4: 124, DOI:10.1186/1756-0500-4-124, 2011.

32. Xu H, George V. A gene-based approach for testing association of rare alleles. BMC Proceedings 5 Suppl 9:S7. PMID: 22373566, 2011

33. Jin B, Tiedemann RL, Xu H, Ernst J, Kellis M, Dalton S, Liu C, Choi JH, Robertson KD. Linking DNAmethyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. Cancer Research submitted

I have constantly provided consulting services to both internal and external clients through the departmental Consulting and Survey Center for both consulting projects and grant preparations. I have served from two years as the chair of the departmental committee for Seminar and Journal Club. I am now the chair of the departmental committee for computing. We worked with the ITSS to make sure the computing needs of the students, stuff and faculty members are met. I am also a member of MCG's Institutional Animal Care and Use Committee (2010-2012). I serve as a biostatistician, providing advices on statistical adequacy of the Animal Use Protocols.

I am the member of NIH special emphasis review panel 2011 and 2012. I participated in the study sections to help NIH for grants and contracts review. I am a member of the Data Collection and Harmonization group of the Registry and Surveillance System for Hemoglobinopathies (RuSH) program by NIH and CDC. I am a member of the American Statistical Society (ASA), American Society of Human Genetics, and Sigma-Xi, the Scientific Research Society. I attend and present at the annual meetings and ASA local chapter meetings regularly. I have reviewed papers for journals such as *Biometrics*, *Bioinformatics*, *PloS One*, *PloS Genetics*, and *American Journal of Human Genetics*.

EXPECTATION FOR THE POSITION AFTER TENURE

For the future position after tenure, I will consistently take active role in teaching, research, consulting and other scholarly activities such as services to the institution and professional organizations. I will continue to participate in local and national meetings and continue further educational activities to enhance my background in statistics. I am expecting to take active participation in national organizations and some external review panels. The time allocation regarding teaching, research and service should remain roughly the same.