BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Yun Li	POSITION TITLE Assistant Professor of Genetics and Biostatistics
eRA COMMONS USER NAME yun_li	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
Shanghai JiaoTong University, Shanghai, China	BS	07/01	English, Computer Sci.
Bowling Green State University, Bowling Green, OH	MA	08/02	Communications Studies
Bowling Green State University, Bowling Green, OH	MS	08/02	Applied Statistics
University of Michigan, Ann Arbor, MI	PhD	12/09	Biostatistics

A. Personal Statement

The focus of my research is on the development of statistical methods and their application to the genetic dissection of complex diseases and traits. In particular, I have developed a genotype imputation method (implemented in software MaCH) that has become standard in the analysis of genome-wide association scans. I have developed methods for meta-analysis, imputation and region-based association analysis of rare variants in both genetically homogeneous populations and in admixed populations, and assessed different approaches to handle imputation uncertainty in subsequent association analysis. I have worked on genomewide scans for genetic variants underlying several metabolic, auto-immune and cardiovascular diseases and related quantitative traits. In addition, I have developed methods to accommodate low-coverage sequencing data (implemented in software *thunder*) and have been actively involved in a number of next-generation sequencing (NGS) based studies including the 1000 Genomes Project (Project Leader on calling SNP genotypes from low-coverage pilot), identification of RNA-DNA differences (RDDs), targeted sequencing of selected exons in >14,000 individuals, the WHI whole exome sequencing project (WHISP), and whole genome sequencing based studies for type 2 diabetes, for cannabis and stimulant dependence, and for blood lipid levels.

B. Positions and Honors.

Professional Experience

2004-2009 Research Assistant, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI
2009- Faculty Member, Curriculum in Bioinformatics and Computational Biology, University of North Carolina, Chapel Hill (UNC-CH), Chapel Hill, NC
2009- Faculty Member, Carolina Center for Genome Sciences, UNC-CH
2009- Assistant Professor, Department of Biostatistics, UNC-CH
2009- Assistant Professor, Department of Genetics, UNC-CH
2009- Adjunct Assistant Professor, Department of Computer Science, UNC-CH

Other Experience and Professional Memberships

2002- Member, American Statistical Association
2003-2005 Member, American Society for Quality
2005- Member, American Society of Human Genetics

2007-2009 Member, American Association for the Advancement of Science
2007- Manuscript Reviewer, *American Journal of Human Genetics*, *American Journal of Public Health*, *Annals of Applied Statistics, Annals of Neurology*, *Bioinformatics*, *Biostatistics*, *BMC Bioinformatics*, *BMC Genetics*, *BMC Genomics, European Journal of Human Genetics*, *Frontiers of Medicine*, *Frontiers in Statistical Genetics and Methodology*, *Genetic Epidemiology*, *Genetics*, *Genome Research*, *Human Heredity*, *Human Molecular Genetics*, *International Journal of Biostatistics*, *Journal of Bioinformatics and Computational Biology*, *Nature Communications*, *Nature Methods*, *Pacific Symposium on Biocomputing, PLoS Genetics*, *PLoS ONE, Statistical Applications in Genetics and Molecular Biology*, *Theoretical Population Biology*
2012- Ad Hoc Grant Reviewer for Barts and The London Charity Grant, GCAT study section, Wellcome Trust and Royal Society Sir Henry Dale Fellowship, BDMA study section, ERC (European Research Council) Consolidator Grant
2010- Editorial Board, *Frontiers in Statistical Genetics and Methodology*
2011- Academic Editor in Editorial Board, *PLoS ONE*

Honors and Awards

2003 Wray Jackson Smith Scholarship, American Statistical Association
2004 Ronald Benton Scholarship, Toledo Section, American Society for Quality
2005 Best Performance on the Qualifying Examination, Dept. of Biostatistics, University of Michigan
2007 March of Dimes Scholarship on Medical and Experimental Mammalian Genetics
2008 Rackham Predoctoral Fellowship, University of Michigan
2008 Trainee Award in Predoctoral Basic, American Society of Human Genetics
2008 Rackham One-Term Dissertation Fellowship, University of Michigan
2012 Jefferson-Pilot Fellowship in Academic Medicine, School of Medicine, UNC-CH
2013 Junior Faculty Development Award, UNC-CH
2014 Thomson Reuters Highly Cited Researcher 2014

C. Selected publications (out of 66, in chronological order).

1. **Li Y**, Willer CJ, Sanna S, Abecasis GR (2009). Genotype imputation. *Annual Review Genomics and Human Genetics,* 10:387-406. PMCID: PMC2925172.
2. Willer CJ, **Li Y**, Abecasis GR (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics,* 26:2190-1. PMCID: PMC2922887.
3. The 1000 Genomes Project. (2010). A map of human genome variation from population scale sequencing. *Nature,* 467:1061-73. PMCID: PMC3042601.
4. **Li Y**, Byrnes AE, Li M (2010). To identify associations with rare variants, just WHaIT: Weighted haplotype and imputation-based tests. *The American Journal of Human Genetics,* 87:728-35. PMCID: PMC2978961.
5. **Li Y**, Willer CJ, Scheet P, Ding J, and Abecasis GR (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology,* 34:816-34. PMCID: PMC3175618.
6. **Li Y**, Sidore C, Kang HM, Boehnke M, Abecasis GR (2011). Low-coverage sequencing: implications for design of complex trait association studies. *Genome Research,* 21:940-51. PMCID: PMC3106327.
7. Li M, Wang IX, **Li Y**, Bruzel A, Richards AL, Toung JM, Cheung VG (2011). Widespread RNA and DNA sequence differences in the human transcriptome. *Science,* 333(6038):53-8. PMCID: PMC3204392.
8. Wu MC, Lee S, Cai T, **Li Y**, Boehnke M, Lin X (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics,* 89:82-93. PMCID: PMC3135811.
9. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu SA, Fraser D, Warren L, Aponte J, Zawistowski M, Liu X, Zhang H, Zhang Y, Li J, **Li Y**, Li L, Woollard P, Topp S, Hall MD, Nangle K, Wang J, Abecasis G, Cardon LR, Zollner S, Whittaker JC, Chissoe SL, Novembre J, Mooser V (2012). An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. *Science,* 337(6090):100-4. PMID: 22604722.
10. Chen W, Li B, Zeng Z, Sanna S, Sidore C, Busonero F, Kang HM, **Li Y**, Abecasis G (2012). Genotype calling and haplotyping in parent-offspring trios. *Genome Research,* 23:142-51.
11. Liu EY, Li M, Wang W, **Li Y** (2012). MaCH-Admix: Genotype Imputation for Admixed Populations.

Genetic Epidemiology, 37:25-37. PMCID: PMC3524415.
12. Auer PL, Johnsen JM, Johnson AD, Logsdon BA, Lange LA, Nalls MA, Zhang G, Franceschini N, Fox K, Lange EM, Rich SS, O'Donnell CJ, Jackson RD, Wallace RB, Chen Z, Graubert TA, Wilson JG, Tang H, Lettre G, Reiner AP, Ganesh SK, **Li Y** (2012). Imputation of Exome Sequence Variants into Population-Based Samples and Blood-Cell-Trait-Associated Loci in African Americans: NHLBI GO Exome Sequencing Project. *The American Journal of Human Genetics,* 91:794-808. PMCID: PMC3487117.
13. Duan Q, Liu EY, Croteau-Chonka DC, Mohlke KL, **Li Y** (2013). A comprehensive SNP and indel imputability database. *Bioinformatics,* 29(4):528-31. PMID: 23292738. PMCID: PMC3570215.
14. Kang J, Huang KC, Xu Z, Wang Y, Abecasis GR, **Li Y** (2013). AbCD: arbitrary coverage design for sequencing-based genetic studies. *Bioinformatics,* 29:799-801. PMCID: PMC3597143.
15. Duan Q, Liu EY, Auer PL, Zhang G, Lange EM, Jun G, Bizon C, Jiao S, Buyske S, Franceschini N, Carlson CS, Hsu L, Reiner AP, Peters U, Haessler J, Curtis K, Wassel CL, Robinson JG, Martin LW, Haiman CA, Le Marchand L, Matise TC, Hindorff LA, Crawford DC, Assimes TL, Kang HM, Heiss G, Jackson RD, Kooperberg C, Wilson JG, Abecasis GR, North KE, Nickerson DA, Lange LA, **Li Y** (2013). Imputation of Coding Variants in African Americans: Better Performance using Data from the Exome Sequencing Project. *Bioinformatics,* [Epub ahead of print]. PMID: 23956302.

D. Research Support

ACTIVE

1R01HG006703 Li (PI) 05/16/12-02/28/15
Imputation and Analysis of Rare Variants in Admixed Populations
The goal of this study is to develop statistical methods and computational tools for imputation and association analysis of rare variants in admixed populations.
Role: PI

1R01HG006292 Li (PI) 08/23/11-05/31/16
Design and Analysis of Sequencing-based Studies for Complex Human Traits
The goal of this study is to establish a comprehensive statistical framework for the design and analysis of sequencing-based studies for complex human traits.
Role: PI

5R01 DK093757 Mohlke (PI) 09/05/11-07/31/16
Genetic epidemiology of rare and regulatory variants for metabolic traits
The goal of this project is to identify novel variants that influence traits related to diabetes, obesity and the metabolic syndrome and mechanisms by which DNA variants influence gene expression and disease.
Role: Co-Investigator

1R01 DA030976-01 Wilhelmsen (PI) 09/30/10-05/31/15
Deep Sequencing Studies for Cannabis and Stimulant Dependence
The goal of this proposal is to identify genes that affect susceptibility to stimulant and cannabis dependence using whole genome sequencing with genotype imputation.
Role: Co-Investigator

PAST

Subcontract No. 3001352222 Li (PI) 12/01/09-11/30/13
GlaxoSmithKline/UMichigan
The goal of this project is to develop and apply imputation based methodology to genome wide association and sequencing datasets.
Role: Subcontract PI

5R01 HL095396-02 Knowles (PI) 09/24/08-07/31/12
Molecular Phenotypes for Cystic Fibrosis Lung Disease
The goal of this project is to define a molecular phenotype for CF lung disease, which relates to prognosis, and new targets for therapy.

Role: Co-Investigator

RD83272001 Wright (PI) 10/01/05-09/30/11
Computational Toxicology: Environmental Bioinformatics Research
Role: Co-Investigator

U01 DA024413 Costello (PI) 09/01/07-06/30/12
A developmental model of gene-environment interplay in SUDs
The major goal of this study is to investigate genetic main effects and gene-environment interactions using GWAS data in longitudinal studies of substance initiation and progression.
Role: Co-Investigator

3R01 CA082659-11S1 Lin (PI) 08/01/09-07/31/11
Statistical Methods in Cancer Research
The goal of this project is the development of statistical methods for the designs and analysis of clinical and epidemiological cancer studies.
Role: Co-Investigator

3R01 DK078150-04 Mohlke (PI) 04/01/07-03/31/12
Genetic Epidemiology of Body Mass Index, Adiposity, and Weight Gain
The goals of this study are to test candidate genes for association with obesity-related traits and weight gain across 22 years in women from the Cebu Longitudinal Health and Nutrition Survey and to evaluate interactions with diet composition and physical activity.
Role: Co-Investigator

1RC2HL102924-01 Jackson, North (PI) 09/30/09-07/31/11
WHI Sequencing Project (WHISP)
The overall goal of this project submitted in response to NHLBI RC2 Topic 'Large-scale DNA Sequencing and Molecular Profiling of Well-phenotyped NHLBI Cohorts' (RFA-OD-09-004) is to identify putative functional variants for high-priority heart lung and blood phenotypes among American post-menopausal women from diverse ancestral and geographic backgrounds.
Role: Co-Investigator

5P01HD031921-15 Whitsel (PI) 08/01/10-05/31/13
Modification of PM-mediated Arrhythmogenesis in Populations
The goal of the project is to examine susceptibility to the arrhythmogenic effects of particulate matter (PM) air pollution contributed by common genetic variation.
Role: Co-Investigator