BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Li, Yun

eRA COMMONS USER NAME (credential, e.g., agency login): yun_li

POSITION TITLE: Professor of Genetics and Biostatistics

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	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	08/04	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 genotype imputation methods (software MaCH and MaCH-Admix) that have become standard practice. I have also developed methods for meta-analysis, local ancestry inference, and region-based association of rare variants in both genetically homogeneous and in admixed populations, and proposed different approaches to handle imputation uncertainty in association analysis. I have worked on genome-wide scans for genetic variants underlying several metabolic, auto-immune, cardiovascular, neuropsychiatric diseases and related quantitative traits. In addition, I have developed methods to accommodate low-coverage sequencing data for genotype calling and for association testing (software *thunder* [now part of GotCloud]) 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), Exome Sequencing Project (ESP), and the Trans-Omics for Precision Medicine (TOPMed) project. In addition, I have developed methods for DNA methylation data and actively participated in multiple epigenome-wide association studies. Recently, I have developed methods for single-cell RNA-seq and spatial transcriptomics data, particularly on ensemble clustering, batch effect correction, and association in medical genetics context. I have also worked on method development and data analysis for Hi-C and derived data, particularly detection of long-range chromatin interactions and integration with GWAS and eQTL data.

B. Positions and Honors

Positions and Employment

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
2009-2015 Assistant Professor, Department of Biostatistics, UNC-CH
2009-2015 Assistant Professor, Department of Genetics, UNC-CH
2015-2021 Associate Professor, Department of Biostatistics, UNC-CH
2015-2021 Associate Professor, Department of Genetics, UNC-CH
2009- Adjunct Assistant Professor, Department of Computer Science, UNC-CH
2015- Director, Data Science Core, Intellectual and Developmental Disabilities Research Center

2021- Professor, Department of Biostatistics, UNC-CH
2021- Professor, Department of Genetics, 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*, *Nature Genetics, 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, Hong Kong Research Grant Council, Hong Kong Health Medical Research Fund, NIH special emphasis panels
2010- Editorial Board, *Frontiers in Statistical Genetics and Methodology*
2011- Academic Editor in Editorial Board, *PLoS ONE*
2017- Member, NIH GCAT Study Section
2018- Software section editor, *Human Genomics*
2021- Member, American Association for the Advancement of Science

Honors

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
2015 Faculty Member, Theta Chapter of the Delta Omega Society

C. Contribution to Science

Publications below were selected from a list of 194 publications in peer-reviewed journals; 66,902 citations, H-index score 64, i10-index score 114, as of 10/29/2020 on Google Scholar.
indicates corresponding authorship.

1. I have developed methods for haplotype inference and genotype imputation (implemented in MaCH) that has become widely used in genome wide association studies and meta-analysis to increase statistical power for gene mapping. I have also developed methods and software for subsequent association and meta-analysis (implemented in software Mach2dat, Mach2qtl, and METAL). I have conducted extensive research evaluating the impact of numerous factors influencing imputation quality, post-imputation quality control and subsequent downstream analysis.
   a. **Li Y**, Willer CJ, Sanna S, Abecasis GR (2009). Genotype imputation. *Annual Review Genomics and Human Genetics* 10:387-406. PMCID: PMC2925172. (*cited >1000 times*)
   b. Willer CJ, **Li Y**, Abecasis GR (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190-1. PMCID: PMC2922887. (*cited >2700 times*)
   c. **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. (*cited >1900 times*)

   d. Huang J, Ellinghaus D, Franke A, Howie B, **Li Y#** (2012). 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *European Journal of Human Genetics* 20:801-5. PMCID: PMC3376268. (*cited >130 times*)
2. I have developed methods for the analysis of next generation sequencing data, including both genotype calling from arbitrary depth sequencing data (software *thunder* [part of UMAKE and GotCloud], *triocaller*), design of sequencing based studies (toolkit *AbCD*) and rare variant association analysis (software WHaIT, SKAT). I have taken active parts in many sequencing-based studies.
   a. **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. (*cited 96 times*)
   b. **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. (*cited 297 times*)
   c. 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. (*cited 1937 times*)
   d. 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. (*cited 14 times*)
3. I have developed methods for genetic studies in admixed populations. Applications of the methods to large admixed cohorts have advanced gene mapping for multiple traits (e.g., blood cell traits, plasma lipid levels).
   a. Liu EY, Li M, Wang W, **Li Y#** (2012). MaCH-Admix: Genotype Imputation for Admixed Populations. *Genet Epidemiol.* 37:25-37. PMCID: PMC3524415. (*cited 122 times*)
   b. 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. (*cited 108 times*)
   c. 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* 29(21):2744-9. PMID: 23956302. PMCID: PMC3799474. (*cited 34 times*)
   d. Duan Q, Xu Z, Raffield LM, Chang S, Wu D, Lange EM, Reiner AP, **Li Y#** (2018) A Robust and Powerful Two-step Testing Procedure for Local Ancestry Adjusted Allelic Association Analysis in Admixed Populations. *Genet Epidemiol.* 42(3):288-302. PMID: 29226381. (*cited 4 times*)
4. I have worked on method development and analysis of Hi-C data. We have developed statistically rigorous hidden Markov random field based methods (HMRF-Bayes and FastHiC) for peak calling from Hi-C data. We have also analyzed a compendium of Hi-C data across 21 human cell lines and primary tissues. We have developed web resources (HiView and HUGIn) for investigators to access our results from the analysis of the Hi-C compendium data.
   a. Xu Z, Zhang G, Jin F, Chen M, Furey TS, Sullivan PF, Qin Z, Hu M, **Li Y#** (2015) A hidden Markov random field-based Bayesian method for the detection of long-range chromosomal interactions in Hi-C Data. *Bioinformatics* 32(5):650-6. PMID: 26543175. (*cited 27 times*)
   b. Xu Z, Zhang G, Wu C, **Li Y#**, Hu M (2016) FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C Data. *Bioinformatics* 32(17):2692-5. PMID: 27153668. PMCID: PMC5013904. (*cited 16 times*)
   c. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, **Li Y**, Lin S, Lin Y, Barr CL, Ren B (2016) A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* 17(8):2042-2059. PMID: 27851967. PMCID: PMC5478386. (*cited 400 times*)
   d. Martin JS, Xu Z, Reiner AP, Mohlke KL, Sullivan P, Ren B, Hu M, **Li Y#** (2017) HUGIn: Hi-C Unifying Genomic Interrogator. *Bioinformatics* 33:3793-5. PMID: 28582503. (*cited 40 times*)

5. Besides research in traditional statistical genetics focusing on the primary DNA sequence, I have also involved in the exploration of multi-omics data including DNA methylation, mRNA expression, both in bulk and in single cells.
   a. 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. (*cited 438 times*)
   b. Yang Y, Huh R, Culpepper HW, Lin Y, Love MI, **Li Y#** (2019) SAFE-clustering: Single-cell Aggregated (From Ensemble) Clustering for Single-cell RNA-seq Data. *Bioinformatics* 35(8):1269-77. PMID: 30202935. (*cited 26 times*)
   c. Zhong W, Spracklen CN, Mohlke KL, Zheng X, Fine J, **Li Y#** (2019) Multi-SNP mediation intersection-union test. *Bioinformatics* 35(22):4724-4729 PMID: 31099385. (*cited 5 times*)
   d. Huh R, Yang Y, Jiang Y, Shen Y, **Li Y#** (2020) SAME-clustering: Single-cell Aggregated Clustering via Mixture Model Ensemble. *Nucleic Acids Res* 48(1):86-95. PMID: 31777938. (*cited 4 times*)

Complete List of Published Work in MyBibliography:
https://www.ncbi.nlm.nih.gov/sites/myncbi/yun.li.1/bibliography/40364525/public/?sort=date&direction=descending

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support

1R01HL129132 Li,Reiner(NCE) 07/15/16-02/29/21
Genetic studies of blood cell traits in multi-ethnic cohorts
The goal of this study is to perform genetic studies of blood cell and related traits in multi-ethnic cohorts, particularly admixed populations including African Americans and Hispanics.
Role: contact PI

U24AR076730-01 LaVange, Ivanova 09/26/19-05/31/24
HEAL Initiative: Back Pain Consortium (BACPAC) Research Program Data Integration, Algorithm Development and Operations Management Center
The goal of this project is to set the stage for technology assessments, solicitation of patient input and utilities, and the evaluation of high-impact interventions through the innovative design and sound execution of clinical trials, leading to effective personalized treatment approaches for patients with chronic lower back pain.
Role: Co-Investigator and Leader of the System Biology and Bioinformatics Working Group

2R01DK093757 Mohlke 08/01/17-07/31/22
Genetic epidemiology of rare and regulatory variants for metabolic traits
The goal of this project is to identify novel genes for metabolic traits, discover pathogenic regulatory variants, and learn how environmental context can influence the dynamic range of gene regulation and the development of disease.
Role: Co-Investigator

1R01MD011609 Manuck 08/08/17-03/31/22
The Pharmacoepigenomics of recurrent preterm birth in non-Hispanic black women
The goal of this study is to provide immediate and sustained clinical and public health impact to reduce disparities in PTB outcomes in NH black women and infants, thereby reducing neonatal mortality and lifelong morbidity.
Role: Co-Investigator

U54 HD079124 Piven 09/24/13-06/30/25
Clinical Translational Research Center for Neurodevelopmental Disorders
Intellectual and Developmental Disabilities Research Center.
Role: Co-Investigator and Director of Data Science Core

U01 DA052713 Shen,Kriegstein 09/30/20-08/31/25
Charting the 3D Epigenome in Human Brain Development and Diseases
The goal of this 4DN Organization and Function project is to reveal new insight into the biological functions of the 3D epigenome in brain development and diseases.

Role: Co-Investigator

Completed Research Support

Omitted