2016 BCB Rotation Advertisement for Yun Li Group of Statistical and Computational Genetics and Genomics http://yunliweb.its.unc.edu/

Background: Li Group is recruiting rotation students from the BCB curriculum for year 2016-2017. My group develops statistical methods and computational tools for modern genetic, genomic, and epigenomic data. We do both method development and real data applications.

Qualifications: Solid background in Biology, Statistics and Computing. **Just kidding**: if you had all the above three, you can take my job. As a matter of fact, even I don't dare to say that I have "solid" background in all three. **The truth is** that I look for people who have motivation/willingness-to-learn and the ability to learn and to think. If you get excited with us (me and my lab members), we can teach you, very quickly and painlessly ^_^. And then you should start to teach us something back!

Group Website: NOT very well maintained/updated … http://yunliweb.its.unc.edu/

Lab Members or Alumni: (if you are interested in any project or in my lab in general, DO talk to them. I value their opinions A LOT!)

- **Qing Duan**, BCB PhD, qduan@email.unc.edu
- **Yimeng Tianyao**, STOR MS, yimeng@live.unc.edu
- **Laura Raffield,** postdoctoral fellow, raffield@email.unc.edu
- **Guosheng Zhang**, BCB alumni, graduated 2016 (first job: Software Engineer Internship at Google), gszhang@email.unc.edu
- **Kuan-Chieh Huang**, Biostatistics PhD alumni, graduated 2015 (first job: Sr. Biostatistician at Gilead Sciences), kchuang@live.unc.edu
- **Song Yan**, postdoctoral fellow alumni (Data Scientist at Bing Ads, Microsoft), song_yan@med.unc.edu
- **Zheng Xu**, postdoctoral fellow, xuzheng@email.unc.edu
- **Eric Yi Liu**, graduated 2013, PhD in Computer Science (Now Research Scientist at Facebook)
- **Andrea Byrnes**, graduated 2013, PhD in Biostatistics, now with the Broad Institute

Projects: There are MANY SUPER exciting projects (well, at least to me and most of my team members) including (but not limited to and for good or bad, I am VERY open to new areas of research):

- **Genetic association studies of blood cell traits in multi-ethnic cohorts**: Blood cell traits, including hemoglobin level, red blood cell (RBC), white blood cell (WBC), and platelet counts (PLT), are important intermediate clinical phenotypes for a variety of cardiovascular, hematologic, oncologic, immunologic and infectious diseases. We will use a combination of GWAS, WGS (whole genome sequencing) from 10,000-100,000s individuals together with other omics data to improve gene mapping for blood cell traits. **Talk to**: me (too obvious?), Qing Duan, Laura Raffield. Relevant publications: (Auer, Johnsen et al. 2012; Naik, Wilson et al. 2016)
- **DNA 3D structure**: Chromosomes are heavily packed; but they are not packed in a random way. The structure they take, which is dynamic and has some stochasticity, is heavily associated with their function. My group is developing methods and software for data from 3C (Chromosome Conformation Capture) derived technologies. **Talk to**: me, Guosheng Zhang, Zheng Xu. Relevant publications: (Xu, Zhang et al. 2016; Xu, Zhang et al. 2016; Xu, Zhang et al. 2016).
- **Genetic studies of Admixed Populations**: African Americans and Hispanics are admixed, i.e., have their chromosomes from more than one ancestral populations. They provide excellent opportunities to boost both power and resolution for gene mapping, i.e., finding genes or genetic variants associated with phenotypic traits like blood lipid levels, height, risk of type 2 diabetes, cardiovascular diseases etc. Relevant publications: (Auer, Johnsen et al. 2012; Liu, Li et al. 2013; Mao, Li et al. 2013; Ma, Zhao et al. 2014; Coram, Candille et al. 2015)
- **Epigenetic studies**: DNA methylation is very important to study: there must be good reasons that the copy from our dad was COMPLETELY (well, nothing is exactly 100% in biology) de-methylated before we were formed as a single cell. Relevant publications: (Demerath, Guan et al. 2015; Zhang, Huang et al. 2016)
- **Analysis of DNA sequencing data**. Although we are getting closer to $1,000 per genome, analysis still has a very high price tag. We have developed methods for genotype calling and haplotype construction in both unrelated and family data. We also developed online tools to help people with the design of sequencing-based studies. We have been developing methods for association analysis for both common and rare variants, for genetically both homogeneous and admixed populations. Relevant publications: (Li, Byrnes et al. 2010; Zawistowski, Gopalakrishnan et al. 2010; Li, Sidore et al. 2011; Wu, Lee et al. 2011; Byrnes, Wu et al. 2013; Huang, Sun et al. 2014; Yan and Li 2014; Fan, Wang et al. 2015; Fan, Wang et al. 2015; Hu, Li et al. 2015; Urrutia, Lee et al. 2015)
- **Genotype imputation**. In plain words, making up data, making up >90% of the data, and making them up smartly, quickly and accurately. We also

develop methods to take the uncertainty in making-up/guessing data into subsequent analysis. Relevant publications: (Huang, Li et al. 2009; Li, Willer et al. 2009; Li, Willer et al. 2010; Auer, Johnsen et al. 2012; Liu, Buyske et al. 2012; Duan, Liu et al. 2013; Duan, Liu et al. 2013; Liu, Li et al. 2013; Hu, Li et al. 2015)

- **Your favorite project** (as long as you can get my interest ^_^).

Auer, P. L., J. M. Johnsen, et al. (2012). "Imputation of Exome Sequence Variants into Population- Based Samples and Blood-Cell-Trait-Associated Loci in African Americans: NHLBI GO Exome Sequencing Project." Am J Hum Genet **91**(5): 794-808.

Byrnes, A. E., M. C. Wu, et al. (2013). "The value of statistical or bioinformatics annotation for rare variant association with quantitative trait." Genet Epidemiol **37**(7): 666-674.

Coram, M. A., S. I. Candille, et al. (2015). "Leveraging Multi-ethnic Evidence for Mapping Complex Traits in Minority Populations: An Empirical Bayes Approach." Am J Hum Genet **96**(5): 740-752.

Demerath, E. W., W. Guan, et al. (2015). "Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci." Hum Mol Genet **24**(15): 4464-4479.

Duan, Q., E. Y. Liu, et al. (2013). "Imputation of coding variants in African Americans: better performance using data from the exome sequencing project." Bioinformatics **29**(21): 2744-2749.

Duan, Q., E. Y. Liu, et al. (2013). "A comprehensive SNP and indel imputability database." Bioinformatics **29**(4): 528-531.

Fan, R., Y. Wang, et al. (2015). "Gene Level Meta-Analysis of Quantitative Traits by Functional Linear Models." Genetics **200**(4): 1089-1104.

Fan, R., Y. Wang, et al. (2015). "Meta-analysis of Complex Diseases at Gene Level by Generalized Functional Linear Models." Genetics.

Hu, Y. J., Y. Li, et al. (2015). "Integrative analysis of sequencing and array genotype data for discovering disease associations with rare mutations." Proc Natl Acad Sci U S A **112**(4): 1019-1024.

Huang, K. C., W. Sun, et al. (2014). "Association studies with imputed variants using expectation-maximization likelihood-ratio tests." PLoS One **9**(11): e110679.

Huang, L., Y. Li, et al. (2009). "Genotype-Imputation Accuracy across Worldwide Human Populations." American Journal of Human Genetics **84**(2): 235-250.

Li, Y., A. E. Byrnes, et al. (2010). "To Identify Associations with Rare Variants, Just WHaIT: Weighted Haplotype and Imputation-Based Tests." American Journal of Human Genetics **87**(5): 728-735.

Li, Y., C. Sidore, et al. (2011). "Low-coverage sequencing: implications for design of complex trait association studies." Genome Res **21**(6): 940-951.

Li, Y., C. Willer, et al. (2009). "Genotype imputation." Annual Review of Genomics and Human Genetics **10**: 387-406.

Li, Y., C. J. Willer, et al. (2010). "MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes." Genetic Epidemiology **34**(8): 816-834.

Liu, E. Y., S. Buyske, et al. (2012). "Genotype Imputation of MetabochipSNPs Using a Study-Specific Reference Panel of ~4,000 Haplotypes in African Americans From the Women's Health Initiative." Genetic Epidemiology **36**(2): 107-117.

Liu, E. Y., M. Li, et al. (2013). "MaCH-Admix: Genotype Imputation for Admixed Populations." Genet Epidemiol **37**(1): 25-37.

Ma, Y., J. Zhao, et al. (2014). "Accurate inference of local phased ancestry of modern admixed populations." Sci Rep **4**: 5800.

Mao, X., Y. Li, et al. (2013). "Testing genetic association with rare variants in admixed populations." Genet Epidemiol **37**(1): 38-47.

Naik, R. P., J. G. Wilson, et al. (2016). "Elevated D-dimer levels in African Americans with sickle cell trait." Blood **127**(18): 2261-2263.

Urrutia, E., S. Lee, et al. (2015). "Rare Variant Testing Across Methods and Thresholds Using the Multi-Kernel Sequence Kernel Association Test (MK-SKAT)."

Wu, M. C., S. Lee, et al. (2011). "Rare-variant association testing for sequencing data with the sequence kernel association test." American Journal of Human Genetics **89**(1): 82-93.

Xu, Z., G. Zhang, et al. (2016). "HiView: an integrative genome browser to leverage Hi-C results for the interpretation of GWAS variants." BMC Res Notes **9**(1): 159.

Xu, Z., G. Zhang, et al. (2016). "A hidden Markov random field-based Bayesian method for the detection of long-range chromosomal interactions in Hi-C data." Bioinformatics **32**(5): 650-656.

Xu, Z., G. Zhang, et al. (2016). "FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C data." Bioinformatics.

Yan, S. and Y. Li (2014). "BETASEQ: a powerful novel method to control type-I error inflation in partially sequenced data for rare variant association testing." Bioinformatics **30**(4): 480-487.

Zawistowski, M., S. Gopalakrishnan, et al. (2010). "Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes." American Journal of Human Genetics **87**(5): 604-617.

Zhang, G., K. C. Huang, et al. (2016). "Across-Platform Imputation of DNA Methylation Levels Incorporating Nonlocal Information Using Penalized Functional Regression." Genet Epidemiol **40**(4): 333-340.