CURRICULUM VITAE

PERSONAL INFORMATION

Name: **Yun Li**
e-mail: yunli@med.unc.edu

EDUCATION

PhD, University of Michigan, Ann Arbor, MI, 2004-2009, Biostatistics
MS, Bowling Green State University, Bowling Green, OH, 2002-2004, Applied Statistics
MA, Bowling Green State University, Bowling Green, OH, 2001-2002, Communication Studies
B.S., Shanghai Jiaotong University, China, 1997-2001, English (Finance and Business)
Second B.S., Shanghai Jiaotong University, China, 1998-2001, Computer and Application

PROFESSIONAL EXPERIENCE

Associate Professor of Genetics, UNC, Chapel Hill, NC, 2015-
Associate Professor of Biostatistics, UNC, Chapel Hill, NC, 2015-
Adjunct Assistant Professor of Computer Science, UNC, Chapel Hill, NC, 2012-
Assistant Professor of Genetics, UNC, Chapel Hill, NC, 2009-2015
Assistant Professor of Biostatistics, UNC, Chapel Hill, NC, 2009-2015
Member, Carolina Center for Genome Sciences, UNC, Chapel Hill, NC, 2009-

HONORS AND AWARDS

2015	Faculty Member, Theta Chapter of the Delta Omega Society
2014	Thomson Reuters Highly Cited Researcher
2013	Junior Faculty Development Award, UNC
2012	Jefferson-Pilot Fellowship in Academic Medicine, School of Medicine, UNC
2008-2009:	Rackham Predoctoral Fellowship, University of Michigan
2008:	ASHG Trainee Award in Predoctoral Basic
2008:	Rackham One-Term Dissertation Fellowship, University of Michigan
2007:	March of Dimes Scholarship on Medical & Experimental Mammalian Genetics
2006, 2007:	Rackham Travel Grant, University of Michigan
2005:	Best Performance on Qualifying Examination, University of Michigan
2004:	Robert A. Patton Book Scholarship, Bowling Green State University
2004:	Ronald Benton Scholarship, Toledo Section, American Society for Quality
2003:	Wray Jackson Smith Scholarship, American Statistical Association
2001:	Level 1 (highest) in International Japanese Proficiency Test
1997-2001:	Annual Academic Scholarship, Shanghai Jiaotong University
1998-1999:	Harler Scholarship for Excellence in German Language Studies
1997-1998:	Yan Kuanhu Fund Scholarship for Excellent Performance

1997-1998: Scholarship of the Metrobank Foundation

BIBLIOGRAPHY

Refereed original research:

* indicates first or co-first authorship
# indicates corresponding authorship
underscore indicates lab member

1. 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* [epub ahead of print]. PMID: 27153668.

2. Zhang G, Huang KC, Xu Z, Tzeng JY, Conneely KN, Guan W, Kang J, **Li Y#** (2016) Across-Platform Imputation of DNA Methylation Levels Incorporating Nonlocal Information Using Penalized Functional Regression. *Genet Epidemiol*. 40(4):333-40. PMID: 27061717. PMCID: PMC4862742.

3. Xu Z, Zhang G, Duan Q, Chai S, Zhang B, Wu C, Jin F, Yue F, **Li Y#**, Hu M (2016) HiView: an integrative genome browser to leverage Hi-C results for the interpretation of GWAS variants. *BMC Res Notes* 9(1): 159. PMID: 26969411. PMCID: PMC4788823.

4. Naik RP, Wilson JG, Ekunwe L, Mwasongwe S, Duan Q, **Li Y**, Correa A, Reiner AP (2016) Elevated D-dimer levels in African Americans with sickle cell trait. *Blood* 127(18): 2261-3. PMID: 26968536. PMCID: PMC4859200.

5. Below JE, Parra EJ, Gamazon ER, Torres J, Krithika S, Candille S, Lu Y, Manichakul A, Peralta-Romero J, Duan Q, **Li Y**, Morris AP, Gottesman O, Bottinger E, Wang XQ, Taylor KD, Ida Chen YD, Rotter JI, Rich SS, Loos RJ, Tang H, Cox NJ, Cruz M, Hanis CL, Valladares-Salgado A (2016) Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. *Scientific Reports* 6:19429. PMID: 26780889. PMCID: PMC4726092.

6. Urrutia E, Lee S, Maity A, Zhao N, Shen J, **Li Y**, Wu MC (2015) Rare variant testing across methods and thresholds using the multi-kernel sequence kernel association test (MK-SKAT). *Stat Interface* 8(4):495-505. PMID: 26740853. PMCID: PMC4698916.

7. Fan R, Wang Y, Chui CY, Chen W, Ren H, **Li Y**, Boehnke M, Amos CI, Moore JH, Xiong M (2015) Meta-analysis of Complex Diseases at Gene Level with Generalized Functional Linear Models. *Genetics* 202(2): 457-70. PMID: 26715663. PMCID: PMC4788228.

8. Lange EM, Ribado J, Zuhlke KA, Johnson A, Keele G, Li J, Wang Y, Duan Q, Li G, Gao Z, **Li Y**, Xu J, Zheng SL, Cooney KA (2015) Assessing the Cumulative Contribution of New and Established Common Genetic Risk Factors to Early-Onset Prostate Cancer. *Cancer Epidemiol Biomarkers Prev.* 25(5): 766-72. PMID: 26671023. PMCID: PMC4873425.

9. 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.

10. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526(7571):68-74. PMID: 26432245. PMCID: PMC4750478.

11. Li J, Lange LA, Sabourin J, Duan Q, Valdar W, Willis MS, **Li Y**, Wilson JG, Lange EM (2015) Genome- and exome-wide association study of serum lipoprotein (a) in the Jackson Heart Study. *J Hum Genet*. 60(12):755-61. PMID: 26377243.

12. Yuan S, Johnston HR, Zhang G, **Li Y**, Hu YJ, Qin ZS (2015) One Size Doesn't Fit All - RefEditor: Building Personalized Diploid Reference Genome to Improve Read Mapping and Genotype Calling in Next Generation Sequencing Studies. *PLoS Computational Biology* 11(8):e1004448. PMID: 26267278. PMCID: PMC4534450.

13. Buchkovich ML, Eklund K, Duan Q, **Li Y**, Mohlke KL, Furey TS (2015) Removing reference mapping biases using limited or no genotype data identifies allelic differences in protein binding at disease-associated loci. *BMC Med Genomics* 8:43. PMID: 26210163. PMCID: PMC4515314.

14. Lanktree MB, Elbers CC, **Li Y**, Zhang G, Duan Q, Karczewski KJ, Guo Y, Tragante V, North KE, Cushman M, Asselbergs FW, Wilson JG, Lange LA, Drenos F, Reiner AP, Barnes MR, Keating BJ (2015) Genetic meta-analysis of 15,901 African Americans identifies variation in EXOC3L1 is associated with HDL concentration. *J Lipid Res.* 56(9):1781-6. PMID: 26199122. PMCID: PMC4548782.

15. Fan R, Wang Y, Boehnke M, Chen W, **Li Y**, Ren H, Lobach I, Xiong M (2015) Gene Level Meta-Analysis of Quantitative Traits by Functional Linear Models. *Genetics* 200(4):1089-104. PMID: 26058849. PMCID: PMC4574252.

16. Yan S, Yuan S, Xu Z, Zhang B, Zhang B, Kang G, Byrnes A, **Li Y#** (2015) Likelihood-based complex trait association testing for arbitrary depth sequencing data. *Bioinformatics* 31(18):2955-62. PMID: 25979475. PMCID: PMC4668777.

17. Yu D, Zhang G, Huang X, Wu C, Tan W, Qiao Y, Chang J, Zhao H, Bi X, Cai J, **Li Y**#, Lin D (2015) Relatively Small Contribution of Methylation and Genomic Copy Number Aberration to the Aberrant Expression of Inflammation-Related Genes in HBV-Related Hepatocellular Carcinoma. *PLoS One* 10(5):e0126836. PMID: 25965583. PMCID: PMC4429029.

18. Bi W, Kang G, Zhao Y, Cui Y, Yan S, **Li Y**, Cheng C, Pounds SB, Borowitz MJ, Relling MV, Yang JJ, Liu Z, Pui CH, Hunger SP, Hartford CM, Leung W, Zhang JF (2015) SVSI: Fast and Powerful Set-Valued System Identification Approach to Identifying Rare Variants in Sequencing Studies for Ordered Categorical Traits. *Ann Hum Genet.* 79(4):294-309. PMID: 25959545. PMCID: PMC4474746.

19. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, Hedman ÅK, Sandling JK, Li LA, Irvin MR, Zhi D, Deloukas P, Liang L, Liu C, Bressler J, Spector TD, North K, **Li Y**, Absher DM, Levy D, Arnett DK, Fornage M, Pankow JS, Boerwinkle E (2015) Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Human Molecular Genetics* 24(15):4464-79. PMID: 25935004. PMCID: PMC4492394.

20. Coram MA, Candille SI, Duan Q, Chan KH, **Li Y**, Kooperberg C, Reiner AP, Tang H (2015) Leveraging Multi-ethnic Evidence for Mapping Complex Traits in Minority Populations: An Empirical Bayes Approach. *The American Journal of Human Genetics* 96(5):740-52. PMID: 25892113. PMCID: PMC4570551.

21. Xu Z, Duan Q, Yan S, Chen W, Li M, Lange E, **Li Y#** (2015) DISSCO: direct imputation of summary statistics allowing covariates. *Bioinformatics* 31(15):2434-42. PMID: 25810429. PMCID: PMC4514926.

22. Wang X, Zhang S, **Li Y**, Li M, Sha Q (2015) A powerful approach to test an optimally weighted combination of rare variants in admixed populations. *Genetic Epidemiology* 39(4):294-305. PMID: 25758547.

23. Abdo N, Xia M, Brown CC, Kosyk O, Huang R, Sakamuru S, Zhou YH, Jack JR, Gallins P, Xia K, **Li Y**, Chiu WA, Motsinger-Reif AA, Austin CP, Tice RR, Rusyn I, Wright FA (2015) Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 genomes high-throughput screening study. *Environ Health Perspect* 123(5):458-66. PMID: 25622337. PMCID: PMC4421772.

24. Hu YJ, **Li Y**, Auer PL, Lin DY (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):1019-24. PMID: 25583502. PMCID: PMC4313847.

25. Franceschini N, Hu Y, Reiner AP, Buyske S, Nalls M, Yanek LR, **Li Y**, Hindorff LA, Cole SA, Howard BV, Stafford JM, Carty CL, Sethupathy P, Martin LW, Lin DY, Johnson KC, Becker LC, North KE, Dehghan A, Bis JC, Liu Y, Greenland P, Manson JE, Maeda N, Garcia M, Harris TB, Becker DM, O'Donnell C, Heiss G, Kooperberg C, Boerwinkle E (2015) Prospective associations of coronary heart disease loci in African Americans using the MetaboChip: the PAGE study. *PLoS One* 9(12):e113203. PMID: 25542012. PMCID: PMC4277270.

26. Huang KC, Sun W, Wu Y, Chen M, Mohlke, Lange LA, **Li Y**# (2014) Association studies with imputed variants using expectation-maximization likelihood-ratio tests. *PLoS One* 9(11):e110679. PMID: 25383782. PMCID: PMC4226494.

27. Li J, Lange LA, Duan Q, Lu Y, Singleton AB, Zonderman AB, Evans MK, **Li Y**, Taylor HA, Willis MS, Nalls M, Wilson JG, Lange EM (2015) Genome-wide admixture and association study of serum iron, ferritin, transferrin saturation and total iron binding capacity in African

Americans. *Human Molecular Genetics* 24(2):572-581. PMID: 25224454. PMCID: PMC4334832.

28. Ma Y, Zhao J, Wong JS, Ma L, Li W, Fu G, Xu W, Zhang K, Kittles RA, **Li Y**, Song Q (2014) Accurate inference of local phased ancestry of modern admixed populations. *Scientific Reports* 4:5800. PMID: 25052506. PMCID: PMC4107375.

29. Lange EM, Johnson AM, Wang Y, Zuhlke KA, Lu Y, Ribado JV, Keele GR, Li J, Duan Q, Li G, Gao Z, **Li Y**, Xu J, Isaacs WB, Zheng S, Cooney KA (2014) Genome-wide association scan for variants associated with early-onset prostate cancer. *PLoS One* 9(4):e93436. PMID: 24740154. PMCID: PMC3989171.

30. Han B, Luo H, Raelson J, Huang J, **Li Y**, Tremblay J, Hu B, Qi S, Wu J (2014) TGFBI (βIG-H3) is a diabetes-risk gene based on mouse and human genetic studies. *Hum Mol Genet.* 23(17):4597-611. PMID: 24728038. PMCID: PMC4119410.

31. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature Genetics* 46(3):234-44. PMID: 24509480. PMCID: PMC3969612. (*I worked on performing imputation and analysis for the FUSION cohorts.*)

32. Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang ZZ, Bizon C, Lange EM, Smith JD, Turner EH, Jun G, Kang HM, Peloso G, Auer P0, Li KP, Flannick J, Zhang J, Fuchsberger C, Gaulton K, Lindgren C, Locke A, Manning A, Sim X, Rivas MA, Holmen OL, Gottesman O, Lu Y, Ruderfer D, Stahl EA, Duan Q, **Li Y**, Durda P, Jiao S0, Isaacs A, Hofman A, Bis JC, Correa A, Griswold ME, Jakobsdottir J, Smith AV, Schreiner PJ, Feitosa MF, Zhang Q, Huffman JE, Crosby J0, Wassel CL, Do R, Franceschini N, Martin LW, Robinson JG, Assimes TL, Crosslin DR, Rosenthal EA, Tsai M, Rieder MJ, Farlow DN, Folsom AR, Lumley T, Fox ER, Carlson CS0, Peters U0, Jackson RD0, van Duijn CM, Uitterlinden AG, Levy D, Rotter JI, Taylor HA, Gudnason V Jr, Siscovick DS, Fornage M, Borecki IB, Hayward C, Rudan I, Chen YE, Bottinger EP, Loos RJ, Sætrom P, Hveem K, Boehnke M, Groop L, McCarthy M0, Meitinger T, Ballantyne CM, Gabriel SB, O'Donnell CJ, Post WS, North KE, Reiner AP, Boerwinkle E0, Psaty BM, Altshuler D, Kathiresan S, Lin DY, Jarvik GP0, Cupples LA, Kooperberg C0, Wilson JG, Nickerson DA, Abecasis GR, Rich SS, Tracy RP, Willer CJ; NHLBI Grand Opportunity Exome Sequencing Project (2014) Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *The American Journal of Human Genetics* 94(2):233-45. PMID: 24507775. PMCID: PMC3928660. (*I helped with quality control and rare variant association analysis of the sequencing data.*)

33. Bizon C, Spiegel M, Chasse SA, Gizer IR, **Li Y**, Malc EP, Mieczkowski PA, Sailsbery JK, Wang X, Ehlers CL, Wilhelmsen KC (2014) Variant calling in low-coverage whole genome sequencing of a Native American population sample. *BMC Genomics* 15(1):85. PMID: 2447956. PMCID: PMC3914019.

34. Yan S, **Li Y**# (2014) BETASEQ: a powerful novel method to control type-I error inflation in partially sequenced data for rare variant association testing. *Bioinformatics* 30(4):480-7. PMID: 2433664. PMCID: PMC3928526.

35. Wu Y, Marvelle AF, Li J, Croteau-Chonka DC, Feranil AB, Kuzawa CW, **Li Y**, Adair LS, Mohlke KL (2013) Genetic association with lipids in Filipinos: waist circumference modifies an APOA5 effect on triglyceride levels. *J Lipid Res.* 54(11):3198-3205. PMID: 24023260. PMCID: PMC3793624.

36. Franceschini N, Fox E, Zhang Z, Edwards TL, Nalls MA, Sung YJ, Tayo BO, Sun YV, Gottesman O, Adeyemo A, Johnson AD, Young JH, Rice K, Duan Q, Chen F, **Li Y**, Tang H, Fornage M, Keene KL, Andrews JS, Smith JA, Faul JD, Guangfa Z, Guo W, Liu Y, Murray SS, Musani SK, Srinivasan S, Velez Edwards DR, Wang H, Becker LC, Bovet P, Bochud M, Broeckel U, Burnier M, Carty C, Chasman DI, Ehret G, Chen WM, Chen G, Chen W, Ding J, Dreisbach AW, Evans MK, Guo X, Garcia ME, Jensen R, Keller MF, Lettre G, Lotay V, Martin LW, Moore JH, Morrison AC, Mosley TH, Ogunniyi A, Palmas W, Papanicolaou G, Penman A, Polak JF, Ridker PM, Salako B, Singleton AB, Shriner D, Taylor KD, Vasan R, Wiggins K, Williams SM, Yanek LR, Zhao W, Zonderman AB, Becker DM, Berenson G, Boerwinkle E, Bottinger E, Cushman M, Eaton C, Nyberg F, Heiss G, Hirschhron JN, Howard VJ, Karczewsk KJ, Lanktree MB, Liu K, Liu Y, Loos R, Margolis K, Snyder M; Asian Genetic Epidemiology Network Consortium, Psaty BM, Schork NJ, Weir DR, Rotimi CN, Sale MM, Harris T, Kardia SL, Hunt SC, Arnett D, Redline S, Cooper RS, Risch NJ, Rao DC, Rotter JI, Chakravarti A, Reiner AP, Levy D, Keating BJ, Zhu X (2013) Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *The American Journal of Human Genetics* 93(3):545-54. PMID: 23972371. PMCID: PMC3769920. (*I contributed to the imputation and association analysis in the following cohorts: ARIC, CARDIA, JHS, Mesa and Mt. Sinai Study*.)

37. 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): 2744-2749. PMID: 23956302. PMCID: PMC3799474.

38. Byrnes AE, Wu MC, Wright FA, Li M, **Li Y**# (2013) The Value of Statistical or Bioinformatics Annotation for Rare Variant Association With Quantitative Trait. *Genetic Epidemiology* 37(7):666-674. PMID: 23836599. PMCID: PMC4083762.

39. Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, Johnson NA, Ochs-Balcom HM, Donlon TA, Martin LW, Eaton CB, Robinson JG, Risch NJ, Zhu X, Kooperberg C, **Li Y**, Reiner AP, Tang H (2013) Genome-wide Characterization of Shared and Distinct Genetic Components that Influence Blood Lipid Levels in Ethnically Diverse Human Populations.

The American Journal of Human Genetics 92(6): 904-916. PMID: 23726366. PMCID: PMC3675231.

40. Chang J, Huang Y, Wei L, Ma B, Miao X, **Li Y**, Hu Z, Yu D, Jia W, Liu Y, Tan W, He Z, Ke Y, Wu T, Shen H, Zeng Y, Wu C, Lin D (2013) Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population. *Carcinogenesis* 34(8):1782-1786. PMID: 23536576.

41. Wu C, Chang J, Ma B, Miao X, Zhou Y, Liu Y, **Li Y**, Wu T, Hu Z, Shen H, Jia W, Zeng Y, Lin D, Kraft P (2013) The case-only test for gene-environment interaction is not uniformly powerful: an empirical example. *Genetic Epidemiology* 37(4):402-407. PMID: 23595356.

42. Ramasamy A, Trabzuni D, Gibbs JR, Dillman A, Hernandez DG, Arepalli S, Walker R, Smith C, Ilori GP, Shabalin AA, **Li Y**, Singleton AB, Cookson MR; NABEC, Hardy J; UKBEC, Ryten M, Weale ME (2013) Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. *Nucleic Acids Res.* 41(7):e88. PMID: 23435227. PMCID: PMC3627570.

43. Kang J*, Huang KC*, Xu Z, Wang Y, Abecasis GR, **Li Y#** (2013) AbCD: arbitrary coverage design for sequencing-based genetic studies. *Bioinformatics* 29(6):799-801. PMID: 23357921. PMCID: PMC3597143.

44. 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.

45. The 1000 Genomes Project. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56-65. PMID: 23128226. PMCID: PMC3498066. (*total 692 co-authors. I was in charge of helping the Michigan team developing and using the pipeline for SNP detection and genotype calling for integrated genomic data.*)

46. Butler AM, Yin X, Evans DS, Nalls MA, Smith EN, Tanaka T, Li G, Buxbaum SG, Whitsel EA, Alonso A, Arking DE, Benjamin EJ, Berenson GS, Bis JC, Chen W, Deo R, Ellinor PT, Heckbert SR, Heiss G, Hsueh WC, Keating BJ, Kerr KF, **Li Y**, Limacher MC, Liu Y, Lubitz SA, Marciante KD, Mehra R, Meng YA, Newman AB, Newton-Cheh C, North KE, Palmer CD, Psaty BM, Quibrera PM, Redline S, Reiner AP, Rotter JI, Schnabel RB, Schork NJ, Singleton AB, Smith JG, Soliman EZ, Srinivasan SR, Zhang ZM, Zonderman AB, Ferrucci L, Murray SS, Evans MK, Sotoodehnia N, Magnani JW, Avery CL (2012) Novel Loci Associated With PR Interval in a Genome-Wide Association Study of 10 African American Cohorts. *Circ Cardiovasc Genet.* 5(6):639-646. PMID: 23139255. PMCID: PMC3560365. (*I led genotype imputation and association analysis in the following cohorts: ARIC, JHS, MESA and WHI.*)

47. 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. PMID: 23103231. PMCID: PMC3487117. (This manuscript is highlighted by *Nature Reviews Genetics*: http://www.nature.com/nrg/journal/v14/n1/full/nrg3387.html)

48. Liu EY, Li M, Wang W, **Li Y#** (2013) MaCH-Admix: Genotype Imputation for Admixed Populations. *Genetic Epidemiology* 37:25-37. PMID: 23074066. PMCID: PMC3524415.

49. Chen W, Li B, Zeng Z, Sanna S, Sidore C, Busonero F, Kang HM, **Li Y**, Abecasis G (2013) Genotype calling and haplotyping in parent-offspring trios. *Genome Research* 23(1):142-51. PMID: 23064751. PMCID: PMC3530674.

50. Mao X, **Li Y**, Liu Y, Lange L, Li M (2012) Testing Genetic Association With Rare Variants in Admixed Populations. *Genetic Epidemiology* 37:38-47. PMID: 23032398. PMCID: PMC3524352.

51. Wu C, Kraft P, Zhai K, Chang J, Wang Z, **Li Y**, Hu Z, He Z, Jia W, Abnet CC, Liang L, Hu N, Miao X, Zhou Y, Liu Z, Zhan Q, Liu Y, Qiao Y, Zhou Y, Jin G, Guo C, Lu C, Yang H, Fu J, Yu D, Freedman ND, Ding T, Tan W, Goldstein AM, Wu T, Shen H, Ke Y, Zeng Y, Chanock SJ, Taylor PR, Lin D (2012) Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nature Genetics* 44:1090-1097. PMID: 22960999. PMCID: PMC3513832.

52. Huang J, Liu EY, Welch R, Willer C, Hindorff LA, **Li Y#** (2012) WikiGWA: an Open Platform for Collecting and Using Genome-Wide Association Results. *European Journal of Human Genetics* 21(4):471-473. PMID: 22929026. PMCID: PMC3598322.

53. **Li Y#**, Chen W, Liu EY, Zhou YH (2012) Single Nucleotide Polymorphism (SNP) Detection and Genotype Calling from Massively Parallel Sequencing (MPS) Data. *Statistics in Biosciences* 5(1): 3-25. NIHMSID: 400379. (This manuscript ranks the 2nd most downloaded paper in the past 90 days of the journal as of 2012/12/29)

54. 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):100-104. PMID: 22604722.

55. Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Scott RA, Henneman P, Heid IM, Kizer JR, Lyytikäinen LP, Fuchsberger C, Tanaka T, Morris AP, Small K,Isaacs A, Beekman M, Coassin S, Lohman K, Qi L, Kanoni S, Pankow JS, Uh HW, Wu Y, Bidulescu A, Rasmussen-Torvik LJ, Greenwood CM, Ladouceur M,Grimsby J, Manning AK, Liu CT, Kooner J, Mooser VE, Vollenweider P, Kapur KA, Chambers J, Wareham

NJ, Langenberg C, Frants R, Willems-Vandijk K, Oostra BA, Willems SM, Lamina C, Winkler TW, Psaty BM, Tracy RP, Brody J, Chen I, Viikari J, Kähönen M, Pramstaller PP, Evans DM, St Pourcain B, Sattar N, Wood AR, Bandinelli S, Carlson OD, Egan JM, Böhringer S, van Heemst D, Kedenko L, Kristiansson K, Nuotio ML, Loo BM, Harris T, Garcia M, Kanaya A, Haun M,Klopp N, Wichmann HE, Deloukas P, Katsareli E, Couper DJ, Duncan BB, Kloppenburg M, Adair LS, Borja JB; DIAGRAM+ Consortium; MAGIC Consortium; GLGC Investigators; MuTHER Consortium, Wilson JG, Musani S, Guo X, Johnson T, Semple R, Teslovich TM, Allison MA, Redline S, Buxbaum SG, Mohlke KL,Meulenbelt I, Ballantyne CM, Dedoussis GV, Hu FB, Liu Y, Paulweber B, Spector TD, Slagboom PE, Ferrucci L, Jula A, Perola M, Raitakari O, Florez JC, Salomaa V, Eriksson JG, Frayling TM, Hicks AA, Lehtimäki T, Smith GD, Siscovick DS, Kronenberg F, van Duijn C, Loos RJ, Waterworth DM, Meigs JB, Dupuis J, Richards JB, Voight BF, Scott LJ, Steinthorsdottir V, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Hofmann OM, Segrè AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R,Boerwinkle E, Bonnycastle LL, Boström KB, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Crawford G, Doney AS, Elliott KS,Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B,Jackson AU, Johnson PR, Jørgensen T, Kao WH, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T,Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Petersen AK, Platou C, Proença C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR,Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparsø T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B,Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Morris AD, Palmer CN, Rudan I, Sijbrands E,Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Pedersen O, Barroso I, Groop L,Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI, Soranzo N, Wheeler E, Glazer NL,Bouatia-Naji N, Mägi R, Randall J, Elliott P, Rybin D, Dehghan A, Hottenga JJ, Song K, Goel A, Lajunen T, Doney A, Cavalcanti-Proença C, Kumari M, Timpson NJ,Zabena C, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Roccasecca RM, Pattou F, Sethupathy P, Ariyurek Y, Barter P, Beilby JP, Ben-Shlomo Y, Bergmann S, Bochud M, Bonnefond A, Borch-Johnsen K, Böttcher Y, Brunner E, Bumpstead SJ, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Crisponi L,Day IN, de Geus EJ, Delplanque J, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Grundy S, Gwilliam R,Hallmans G, Hammond N, Han X, Hartikainen AL, Hayward C, Heath SC, Hercberg S, Hillman DR, Hingorani AD, Hui J, Hung J, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le Bacquer O, Lecoeur C, **Li Y**, Mahley R, Mangino M, Martínez-Larrad MT,McAteer JB, McPherson R, Meisinger C, Melzer D, Meyre D, Mitchell BD, Mukherjee S, Naitza S, Neville MJ, Orrù M, Pakyz R, Paolisso G, Pattaro C, Pearson D,Peden JF, Pedersen NL, Pfeiffer AF, Pichler

I, Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Rayner NW, Rice K, Ripatti S, Rivadeneira F,Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA, Scheet P, Seedorf U, Sharp SJ, Shields B, Sigurðsson G, Sijbrands EJ, Silveira A, Simpson L,Singleton A, Smith NL, Sovio U, Swift A, Syddall H, Syvänen AC, Tönjes A, Uitterlinden AG, van Dijk KW, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N,Waeber G, Wagner PJ, Walley A, Ward KL, Watkins H, Wild SH, Willemsen G, Witteman JC, Yarnell JW, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC;DIAGRAM Consortium; GIANT Consortium; Global B Pgen Consortium, Borecki IB, Meneton P, Magnusson PK, Nathan DM, Williams GH, Silander K, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR, Cooper C, Serrano-Ríos M, Lind L, Palmer LJ, Hu FB 1st, Franks PW, Ebrahim S, Marmot M, Kao WH,Pramstaller PP, Wright AF, Stumvoll M, Hamsten A; Procardis Consortium, Buchanan TA, Valle TT, Rotter JI, Penninx BW, Boomsma DI, Cao A, Scuteri A,Schlessinger D, Uda M, Ruokonen A, Jarvelin MR, Peltonen L, Mooser V, Sladek R; MAGIC investigators; GLGC Consortium, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Chasman DI, Johansen CT, Fouchier SW, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Feitosa MF, Orho-Melander M,Melander O, Li X, Li M, Cho YS, Go MJ, Kim YJ, Lee JY, Park T, Kim K, Sim X, Ong RT, Croteau-Chonka DC, Lange LA, Smith JD, Ziegler A, Zhang W, Zee RY,Whitfield JB, Thompson JR, Surakka I, Spector TD, Smit JH, Sinisalo J, Scott J, Saharinen J, Sabatti C, Rose LM, Roberts R, Rieder M, Parker AN, Pare G,O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, McArdle W, Masson D, Martin NG, Marroni F, Lucas G, Luben R, Lokki ML, Lettre G, Launer LJ,Lakatta EG, Laaksonen R, Kyvik KO, König IR, Khaw KT, Kaplan LM, Johansson Å, Janssens AC, Igl W, Hovingh GK, Hengstenberg C, Havulinna AS, Hastie ND,Harris TB, Haritunians T, Hall AS, Groop LC, Gonzalez E, Freimer NB, Erdmann J, Ejebe KG, Döring A, Dominiczak AF, Demissie S, Deloukas P, de Faire U,Crawford G, Chen YD, Caulfield MJ, Boekholdt SM, Assimes TL, Quertermous T, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Taylor HA Jr, Gabriel SB, Holm H, Gudnason V, Krauss RM, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Strachan DP, Reilly MP, Samani NJ,Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, Kathiresan S (2012) Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genetics* 8(3): e1002607. PMID: 22479202. PMCID: PMC3315470. (*I contributed to data analysis in the FUSION study.*)

56. Palmer ND, McDonough CW, Hicks PJ, Roh BH, Wing MR, An SS, Hester JM, Cooke JN, Bostrom MA, Rudock ME, Talbert ME, Lewis JP; DIAGRAM Consortium;MAGIC Investigators, Ferrara A, Lu L, Ziegler JT, Sale MM, Divers J, Shriner D, Adeyemo A, Rotimi CN, Ng MC, Langefeld CD, Freedman BI, Bowden DW, Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segrè AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R,Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Boström KB, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ,Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S,Hassanali N, Herder C, Isomaa B, Jackson

AU, Johnson PR, Jørgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proença C,Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparsø T,Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB,Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ,Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ,Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC,Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI, Soranzo N,Wheeler E, Glazer NL, Bouatia-Naji N, Mägi R, Randall J, Johnson T, Elliott P, Rybin D, Henneman P, Dehghan A, Hottenga JJ, Song K, Goel A, Egan JM, Lajunen T, Doney A, Kanoni S, Cavalcanti-Proença C, Kumari M, Timpson NJ, Zabena C, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Roccasecca RM,Pattou F, Sethupathy P, Ariyurek Y, Barter P, Beilby JP, Ben-Shlomo Y, Bergmann S, Bochud M, Bonnefond A, Borch-Johnsen K, Böttcher Y, Brunner E,Bumpstead SJ, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Crisponi L, Day IN, de Geus EJ, Delplanque J, Fedson AC, Fischer-Rosinsky A, Forouhi NG,Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Grundy S, Gwilliam R, Hallmans G, Hammond N, Han X, Hartikainen AL, Hayward C, Heath SC,Hercberg S, Hicks AA, Hillman DR, Hingorani AD, Hui J, Hung J, Jula A, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le Bacquer O, Lecoeur C, **Li Y**, Mahley R, Mangino M, Manning AK, Martínez-Larrad MT, McAteer JB, McPherson R, Meisinger C,Melzer D, Meyre D, Mitchell BD, Mukherjee S, Naitza S, Neville MJ, Oostra BA, Orrù M, Pakyz R, Paolisso G, Pattaro C, Pearson D, Peden JF, Pedersen NL,Perola M, Pfeiffer AF, Pichler I, Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Psaty BM, Rayner NW, Rice K, Ripatti S, Rivadeneira F, Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA, Scheet P, Seedorf U, Sharp SJ, Shields B, Sijbrands EJ, Silveira A, Simpson L, Singleton A, Smith NL, Sovio U,Swift A, Syddall H, Syvänen AC, Tanaka T, Tönjes A, Uitterlinden AG, van Dijk KW, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N, Waeber G, Wagner PJ,Walley A, Ward KL, Watkins H, Wild SH, Willemsen G, Witteman JC, Yarnell JW, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC, Borecki IB, Loos RJ,Meneton P, Magnusson PK, Nathan DM, Williams GH, Silander K, Salomaa V, Smith GD, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR, Cooper C,Dedoussis GV, Serrano-Ríos M, Lind L, Palmer LJ, Franks PW, Ebrahim S, Marmot M, Kao WH, Pramstaller PP, Wright AF, Stumvoll M, Hamsten A, Buchanan TA, Valle TT, Rotter JI, Siscovick DS, Penninx BW, Boomsma DI, Deloukas P, Spector TD, Ferrucci L, Cao A, Scuteri A, Schlessinger D, Uda M, Ruokonen A,Jarvelin MR, Waterworth DM, Vollenweider P, Peltonen L, Mooser V, Sladek R (2012) A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS One* 7(1): e29202. PMID: 22238593. PMCID: PMC3251563. (*I contributed to statistical genetics analysis*.)

57. 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(7): 801-805. PMID: 22293688. PMCID: PMC3376268.

58. Liu EY, Buyske S, Aragaki AK, Peters U, Boerwinkle E, Carlson CS, Carty C, Crawford DC, Haessler J, Hindorff LA, Marchand LL, Manolio TA, Matise T, Wang W, Kooperberg C, North KE, **Li Y#** (2012) Genotype Imputation of Metabochip SNPs Using a Study-Specific Reference Panel of ~4,000 Haplotypes in African Americans From the Women's Health Initiative. *Genetic Epidemiology* 36:107-117. PMID: 22851474. PMCID: PMC3410659.

59. Cho YS, Chen CH, Hu C, Long J, Ong RT, Sim X, Takeuchi F, Wu Y, Go MJ, Yamauchi T, Chang YC, Kwak SH, Ma RC, Yamamoto K, Adair LS, Aung T, Cai Q, Chang LC, Chen YT, Gao Y, Hu FB, Kim HL, Kim S, Kim YJ, Lee JJ, Lee NR, **Li Y**, Liu JJ, Lu W, Nakamura J, Nakashima E, Ng DP, Tay WT, Tsai FJ, Wong TY, Yokota M, Zheng W, Zhang R, Wang C, So WY, Ohnaka K, Ikegami H, Hara K, Cho YM, Cho NH, Chang TJ, Bao Y, HedmanaK, Morris AP, McCarthy MI; DIAGRAM Consortium; MuTHER Consortium, Takayanagi R, Park KS, Jia W, Chuang LM, Chan JC, Maeda S, Kadowaki T, Lee JY, Wu JY, Teo YY, Tai ES, Shu XO, Mohlke KL, Kato N, Han BG, Seielstad M (2012) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nature Genetics* 44(1): 67-72. PMID: 22158537. (*I led imputation in the CLHNS study*.)

60. Wu Y, McDade TW, Kuzawa CW, Borja J, **Li Y**, Adair LS, Mohlke KL, Lange LA (2012) Genome-wide Association with C-Reactive Protein Levels in CLHNS: Evidence for the CRP and HNF1A Loci and their Interaction with Exposure to a Pathogenic Environment. *Inflammation* 35(2):574-83. PMID: 21647738. PMCID: PMC3193904.

61. Croteau-Chonka DC, Wu Y, **Li Y**, Fogarty MP, Lange LA, Kuzawa CW, McDade TW, Borja JB, Luo J, Abdelbaky O, Combs TP, Adair LS, Lange EM, Mohlke KL (2011) Population-specific coding variant underlies genome-wide association with adiponectin level. *Human Molecular Genetics* 21(2): 463-71. PMID: 22010046. PMCID: PMC3276282.

62. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, Holm H, Sanna S, Kavousi M, Baumeister SE, Coin LJ, Deng G, Gieger C, Heard-Costa NL, Hottenga JJ, Kuhnel B, Kumar V, Lagou V, Liang L, Luan J, Vidal PM, Mateo Leach I, O'Reilly PF, Peden JF, Rahmioglu N, Soininen P, Speliotes EK, Yuan X, Thorleifsson G, Alizadeh BZ, Atwood LD, Borecki IB, Brown MJ, Charoen P, Cucca F, Das D, de Geus EJ, Dixon AL, Doring A, Ehret G, Eyjolfsson GI, Farrall M, Forouhi NG, Friedrich N, Goessling W, Gudbjartsson DF, Harris TB, Hartikainen AL, Heath S, Hirschfield GM, Hofman A, Homuth G, Hypponen E, Janssen HL, Johnson T, Kangas AJ, Kema IP, Kuhn JP, Lai S, Lathrop M, Lerch MM, **Li Y**, Liang TJ, Lin JP, Loos RJ, Martin NG, Moffatt MF, Montgomery GW, Munroe PB, Musunuru K, Nakamura Y, O'Donnell CJ, Olafsson I, Penninx BW, Pouta A, Prins BP, Prokopenko I, Puls R, Ruokonen A, Savolainen MJ, Schlessinger D, Schouten JN, Seedorf U, Sen-Chowdhry S, Siminovitch KA, Smit JH, Spector TD, Tan W, Teslovich TM, Tukiainen T, Uitterlinden AG, Van der Klauw MM, Vasan RS, Wallace C, Wallaschofski H, Wichmann HE, Willemsen G, Wurtz P, Xu C, Yerges-Armstrong LM; AlcGen

Consortium; DIAGRAM+ Study; GIANT Consortium; Global Lipids Genetics Consortium; GOLD Consortium; ICBP-GWAS; MAGIC, Abecasis GR, Ahmadi KR, Boomsma DI, Caulfield M, Cookson WO, van Duijn CM, Froguel P, Matsuda K, McCarthy MI, Meisinger C, Mooser V, Pietilainen KH, Schumann G, Snieder H, Sternberg MJ, Stolk RP, Thomas HC, Thorsteinsdottir U, Uda M, Waeber G, Wareham NJ, Waterworth DM, Watkins H, Whitfield JB, Witteman JC, Wolffenbuttel BH, Fox CS, Ala-Korpela M, Stefansson K, Vollenweider P, Volzke H, Schadt EE, Scott J, Jarvelin MR, Elliott P, Kooner JS (2011) Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature Genetics* 43(11): 1131-8. PMID: 22001757. (*I performed genotype imputation and association analysis*.)

63. Li L, **Li Y**, Browning SR, Browning BL, Slater AJ, Kong X, Aponte JL, Mooser VE, Chissoe SL, Whittaker JC, Nelson MR, Ehm MG (2011) Performance of Genotype Imputation for Rare Variants Identified in Exons and Flanking Regions of Genes. *PLoS One* 6(9):e24945. PMID: 21949800. PMCID: PMC3176314.

64. 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. PMID: 21737059. PMCID: PMC3135811.

65. 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. PMID: 21596952. PMCID: PMC3204392.

66. **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(6):940-51. PMID: 21460063. PMCID: PMC3106327.

67. Zheng J, **Li Y**, Abecasis GR, Scheet P (2011) A comparison of approaches to account for uncertainty in analysis of imputed genotypes. *Genetic Epidemiology* 35(2): 102-110. PMID: 21254217. PMCID: PMC314371.

68. Ding J, Gudjonsson JE, Liang L, Stuart PE, **Li Y**, Chen W, Weichenthal M, Ellinghaus E, Franke A, Cookson W, Nair RP, Elder JT, Abecasis GR (2010) Gene expression in skin and lymphoblastoid cells: refined statistical method reveals extensive overlap in cis-eQTL signals. *The American Journal of Human Genetics* 87: 779-789. PMID: 21129726. PMCID: PMC2997368.

69. Zawistowski M, Gopalakrishnan S, Ding J, **Li Y**, Grimm S, Zollner S (2010) Extending rare variant testing strategies: analysis of non-coding sequence and imputed genotypes. *The American Journal of Human Genetics* 87: 604-617. PMID: 21070896. PMCID: PMC2978957.

70. **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: 816-834. PMID: 21058334.

71. **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* 87:728-735. PMID: 21055717. PMCID: PMC2978961.

72. The 1000 Genomes Project. (2010) A map of human genome variation from population scale sequencing. *Nature* 467: 1061-1073. PMID: 20981092. PMCID: PMC3042601. (*total 544 co-authors, among them 33 are indicated as Principal Investigators and 8 as Project Leader. I am one of the 8 project leaders. Specifically I was in charge of SNP detection and genotype calling for the low coverage data*.)

73. Wassel CL, Lange LA, Keating BJ, Taylor KC, Johnson AD, Palmer CD, Ho LA, Smith NL, Lange EM, **Li Y**, Yang Q, Delaney JA, Tang W, Tofler G, Redline S, Taylor HA Jr, Wilson JG, Tracy RP, Jacobs DR Jr, Folsom AR, Green D, O'Donnell CJ, Reiner AP (2010) Association of genomic loci from a cardiovascular gene SNP array with fibrinogen levels in European Americans and African-Americans from six cohort studies: the Candidate gene Association Resource (CARe). *Blood* 117(1):268-75. PMID: 20978265. PMCID: PMC3037748.

74. Stuart PE, Nair RP, Ellinghaus E, Ding J, Tejasvi T, Gudjonsson JE, **Li Y**, Weidinger S, Eberlein B, Gieger C, Wichmann HE, Kunz M, Ike R, Krueger GG, Bowcock AM, Mrowietz U, Lim HW, Voorhees JJ, Abecasis GR, Weichenthal M, Franke A, Rahman P, Gladman DD, Elder JT (2010) Genome-wide association analysis identifies three psoriasis susceptibility loci. *Nature Genetics* 42(11): 1000-4. PMID:20953189. PMCID: PMC2965799.

75. Ellinghaus E, Ellinghaus D, Stuart PE, Nair RP, Debrus S, Raelson JV, Belouchi M, Fournier H, Reinhard C, Ding J, **Li Y**, Tejasvi T, Gudjonsson J, Stoll SW, Voorhees JJ, Lambert S, Weidinger S, Eberlein B, Kunz M, Rahman P, Gladman DD, Gieger C, Wichmann HE, Karlsen TH, Mayr G, Albrecht M, Kabelitz D, Mrowietz U, Abecasis GR, Elder JT, Schreiber S, Weichenthal M, Franke A (2010) Genome-wide association study identifies a psoriasis susceptibility locus at *TRAF3IP2*. *Nature Genetics* 42: 991-995. PMID: 20953188. PMCID: PMC3136364.

76. Wu Y, **Li Y**, Lange EM, Croteau-Chonka DC, Kuzawa CW, McDade TW, Qin L, Curocichin G, Borja JB, Lange LA, Adair LS, Mohlke KL (2010) Genome-wide association study for adiponectin levels in Filipino women identifies *CDH13* and a novel uncommon haplotype at *KNG1-ADIPOQ*. *Human Molecular Genetics* 19: 4955-4964. PMID: 20876611. PMCID: PMC2989895.

77. Kapur K, Johnson T, Beckmann ND, Sehmi J, Tanaka T, Kutalik Z, Styrkarsdottir U, Zhang W, Marek D, Gudbjartsson DF, Milaneschi Y, Holm H, Diiorio A, Waterworth D, **Li Y**, Singleton AB, Bjornsdottir US, Sigurdsson G, Hernandez DG, Desilva R, Elliott P, Eyjolfsson GI, Guralnik JM, Scott J, Thorsteinsdottir U, Bandinelli S, Chambers J, Stefansson K, Waeber G, Ferrucci L, Kooner JS, Mooser V, Vollenweider P, Beckmann JS, Bochud M, Bergmann S (2010) Genome-wide meta-analysis for serum calcium identifies

significantly associated SNPs near the calcium-sensing receptor (*CASR*) gene. *PLoS Genetics* 6: e1001035. PMID: 20661308. PMCID: PMC2908705.

78. Willer CJ, **Li Y**, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190-2191. PMID: 20616382. PMCID: PMC2922887.

79. Sanna S, Pitzalis M, Zoledziewska M, Zara I, Sidore C, Murru R, Whalen MB, Busonero F, Maschio A, Costa G, Melis MC, Deidda F, Poddie F, Morelli L, Farina G, **Li Y**, Dei M, Lai S, Mulas A, Cuccuru G, Porcu E, Liang L, Zavattari P, Moi L, Deriu E, Urru MF, Bajorek M, Satta MA, Cocco E, Ferrigno P, Sotgiu S, Pugliatti M, Traccis S, Angius A, Melis M, Rosati G, Abecasis GR, Uda M, Marrosu MG, Schlessinger D, Cucca F (2010) Variants within the immunoregulatory *CBLB* gene are associated with multiple sclerosis. *Nature Genetics* 42: 495-497. PMID: 20453840.

80. Chambers JC, Zhang W, Lord GM, van der Harst P, Lawlor DA, Sehmi JS, Gale DP, Wass MN, Ahmadi KR, Bakker SJ, Beckmann J, Bilo HJ, Bochud M, Brown MJ, Caulfield MJ, Connell JM, Cook HT, Cotlarciuc I, Davey Smith G, de Silva R, Deng G, Devuyst O, Dikkeschei LD, Dimkovic N, Dockrell M, Dominiczak A, Ebrahim S, Eggermann T, Farrall M, Ferrucci L, Floege J, Forouhi NG, Gansevoort RT, Han X, Hedblad B, Homan van der Heide JJ, Hepkema BG, Hernandez-Fuentes M, Hypponen E, Johnson T, de Jong PE, Kleefstra N, Lagou V, Lapsley M, **Li Y**, Loos RJ, Luan J, Luttropp K, Maréchal C, Melander O, Munroe PB, Nordfors L, Parsa A, Peltonen L, Penninx BW, Perucha E, Pouta A, Prokopenko I, Roderick PJ, Ruokonen A, Samani NJ, Sanna S, Schalling M, Schlessinger D, Schlieper G, Seelen MA, Shuldiner AR, Sjögren M, Smit JH, Snieder H, Soranzo N, Spector TD, Stenvinkel P, Sternberg MJ, Swaminathan R, Tanaka T, Ubink-Veltmaat LJ, Uda M, Vollenweider P, Wallace C, Waterworth D, Zerres K, Waeber G, Wareham NJ, Maxwell PH, McCarthy MI, Jarvelin MR, Mooser V, Abecasis GR, Lightstone L, Scott J, Navis G, Elliott P, Kooner JS (2010) Genetic loci influencing kidney function and chronic kidney disease. *Nature Genetics* 42: 373-375. PMID: 20383145.

81. Lange LA, Croteau-Chonka DC, Marvelle AF, Qin L, Gaulton KJ, Kuzawa CW, McDade TW, Wang Y, **Li Y**, Levy S, Borja JB, Lange EM, Adair LS, Mohlke KL (2010) Genome-wide association study of homocysteine levels in Filipinos provides evidence for *CPS1* in women and a stronger *MTHFR* effect in young adults. *Human Molecular Genetics* 19: 2050-2058. PMID: 20154341. PMCID: PMC2860887.

82. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Magi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparso T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proenca C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Payne F, Roccasecca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A,

Bonnycastle LL, Borch-Johnsen K, Bottcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, de Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllensten U, Hadjadj S, Hallmans G, Hammond N, Han X, Hartikainen AL, Hassanali N, Hayward C, Heath SC, Hercberg S, Herder C, Hicks AA, Hillman DR, Hingorani AD, Hofman A, Hui J, Hung J, Isomaa B, Johnson PR, Jorgensen T, Jula A, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le Bacquer O, Lecoeur C, **Li Y**, Lyssenko V, Mahley R, Mangino M, Manning AK, Martínez-Larrad MT, McAteer JB, McCulloch LJ, McPherson R, Meisinger C, Melzer D, Meyre D, Mitchell BD, Morken MA, Mukherjee S, Naitza S, Narisu N, Neville MJ, Oostra BA, Orrù M, Pakyz R, Palmer CN, Paolisso G, Pattaro C, Pearson D, Peden JF, Pedersen NL, Perola M, Pfeiffer AF, Pichler I, Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Psaty BM, Rathmann W, Rayner NW, Rice K, Ripatti S, Rivadeneira F, Roden M, Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA, Scheet P, Scott LJ, Seedorf U, Sharp SJ, Shields B, Sigurethsson G, Sijbrands EJ, Silveira A, Simpson L, Singleton A, Smith NL, Sovio U, Swift A, Syddall H, Syvanen AC, Tanaka T, Thorand B, Tichet J, Tonjes A, Tuomi T, Uitterlinden AG, van Dijk KW, van Hoek M, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N, Waeber G, Wagner PJ, Walley A, Walters GB, Ward KL, Watkins H, Weedon MN, Wild SH, Willemsen G, Witteman JC, Yarnell JW, Zeggini E, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium, Borecki IB, Loos RJ, Meneton P, Magnusson PK, Nathan DM, Williams GH, Hattersley AT, Silander K, Salomaa V, Smith GD, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR, Cooper C, Dedoussis GV, Serrano-Ríos M, Morris AD, Lind L, Palmer LJ, Hu FB, Franks PW, Ebrahim S, Marmot M, Kao WH, Pankow JS, Sampson MJ, Kuusisto J, Laakso M, Hansen T, Pedersen O, Pramstaller PP, Wichmann HE, Illig T, Rudan I, Wright AF, Stumvoll M, Campbell H, Wilson JF; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators, Bergman RN, Buchanan TA, Collins FS, Mohlke KL, Tuomilehto J, Valle TT, Altshuler D, Rotter JI, Siscovick DS, Penninx BW, Boomsma DI, Deloukas P, Spector TD, Frayling TM, Ferrucci L, Kong A, Thorsteinsdottir U, Stefansson K, van Duijn CM, Aulchenko YS, Cao A, Scuteri A, Schlessinger D, Uda M, Ruokonen A, Jarvelin MR, Waterworth DM, Vollenweider P, Peltonen L, Mooser V, Abecasis GR, Wareham NJ, Sladek R, Froguel P, Watanabe RM, Meigs JB, Groop L, Boehnke M, McCarthy MI, Florez JC, Barroso I (2009) Novel genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics* 42: 105-116. PMID: 20081858. PMCID: PMC3018764. (*I performed imputation and association analysis in the FUSION cohorts*.)

83. Chambers JC, Zhang W, **Li Y**, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L, Freimer NB, Srai SK, Maxwell PH, Sternberg MJE, Ruokonen A, Abecasis G, Jarvelin MR, Scott J, Elliott P, Kooner JS (2009) Genome-wide association study identifies variants in *TMPRSS6* associated with hemoglobin levels. *Nature Genetics* 41: 1170-1172. PMID: 19820698. PMCID: PMC3178047.

84. Huang L, **Li Y**, Singleton AB, Hardy JA, Abecasis GR, Rosenberg NA, Scheet P (2009) Genotype imputation accuracy across worldwide human populations. *The American Journal of Human Genetics* 84: 235-250. PMID: 19215730. PMCID: PMC2668016.

85. Nair RP, Duffin KC, Helms C, Ding J, Stuard PE, Goldgar D, Gudjonsson JE, **Li Y**, Tejasvi T, Feng BJ, Ruether A, Schreiber S, Weichenthal M, Gladman D, Rahman P, Schrodi SJ, Prahalad S, Guthery SL, Fischer J, Liao W, Kwok P, Menter A, Lathrop GM, Wise C, Begovich AB, Voorhees JJ, Elder JT, Krueger GG, Bowcock AM, Abecasis GR (2009) Genome-wide scan reveals association of psoriasis with *IL-23* and *NF-κB* pathways. *Nature Genetics* 41: 199-204. PMID: 19169254. PMCID: PMC2745122.

86. Kathiresan S, Willer CJ, Peloso G, Demissie S, Musunuru K, Schadt E, Kaplan L, Bennett D, **Li Y**, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burtt N, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta K, Dupuis J, deBakker PI, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics* 41: 56-65. PMID: 19060906. PMCID: PMC2881676.

87. Yuan X, Waterworth D, Perry JR, Lim N, Song K, Chambers JC, Zhang W, Vollenweider P, Stirnadel H, Johnson T, Bergmann S, Beckmann ND, **Li Y**, Ferrucci L, Melzer D, Hernandez D, Singleton A, Scott J, Elliott P, Waeber G, Cardon L, Frayling TM, Kooner JS, Mooser V (2008) Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *The American Journal of Human Genetics* 83(4): 520-8. PMID: 18940312. PMCID: PMC2561937.

88. Gaulton KJ, Willer CJ, **Li Y**, Scott LJ, Conneely KN, Jackson AU, Duren WL, Chines PS, Narisu N, Bonnycastle LL, Luo J, Tong M, Sprau AG, Pugh EW, Doheny KF, Valle TT, Abecasis GR, Tuomilehto J, Bergman RN, Collins FS, Boehnke M, Mohlke KL (2008) Comprehensive association study of type 2 diabetes and related quantitative traits with 222 candidate genes. *Diabetes* 57: 3136-3144. PMID: 18678618. PMCID: PMC2570412.

89. Chambers JC, Elliott P, Zabaneh D, Zhang W, **Li Y**, Froguel P, Balding D, Scott J, Kooner JS (2008) Common genetic variation near the melanocortin-4 receptor gene is associated with waist circumference and insulin resistance. *Nature Genetics* 40: 716-718. PMID: 18454146.

90. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Boström KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jørgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H,

Lauritzen T, **Li Y**, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbæk A, Shields B, Sjögren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ; Wellcome Trust Case Control Consortium, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* 40: 638-645. PMID: 18372903. PMCID: PMC2672416. (*I performed imputation and statistical analysis in the FUSION sample*.)

91. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen W, **Li Y**, Scott LJ, Scheet P, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature Genetics* 40**:** 161-169. PMID: 18193043. (*I contributed to meta-analyzing results from several studies.*)

92. The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851-861. PMID: 17943122. PMCID: PMC2689609. (total 250 co-authors. *I performed basic quality control of genotypes from different labs*.)

93. Scott JL, Mohlke KL, Bonnycastle LL, Willer CJ, **Li Y**, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316: 1341-1345. PMID: 17463248. PMCID: PMC3214617. (*I performed genotype imputation and contributed to meta-analysis with two other studies*.)

94. Li M, Atmaca-Sonmez P, Othman M, Branham KE, Khanna R, Wade MS, **Li Y**, Liang L, Zareparsi S, Swaroop A and Abecasis GR (2006) *CFH* haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nature Genetics* 38**:** 1049-1054. PMID: 16936733. PMCID: PMC1941700. (*I performed haplotype association analysis*.)

Review articles:

1. **Li Y**, Willer CJ, Sanna S, Abecasis GR (2009) Genotype imputation. *Annual Review Genomics and Human Genetics* 10: 387-406. PMID: 19715440. PMCID: PMC2925172.

Articles in press:

1. Hu YJ, **Li Y**, Auer PL, Lin DY (2016) Integrative analysis of sequencing and array genotype data for discovering disease associations with rare mutations. (31 pages including 6 figures and 2 tables + 24 pages supplementary materials), *Proceedings of National Academy of Science*.

Articles submitted:

1. Xu Z, Zhang G, Wu C, **Li Y#**, Hu M# (2016+) FastHiC: a fast algorithm to detect long-range chromosomal interactions from Hi-C data. (2 pages including 1 figure and 1 table + 5 pages supplementary materials), under review with *Bioinformatics*.

Abstracts (published online only):

1. Xu Z, Zhang G, Schimdt A, Ren B, Hu M, **Li Y#**. Detecting Dynamic Long-Range Chromatin Interactions from Multiple Cell Types Hi-C Data Using a Bayesian Hierarchical Hidden Markov Random Field Model (2016) 2016 Joint Statistical Meeting in Chicago.
2. Xu Z, **Li Y#**. Rare Variant Test based on Next-Generation Sequencing Data with Arbitrary Length (2016) 2016 Joint Statistical Meeting in Chicago.
3. Hu Y, **Li Y**, Auer P, Lin DY. Integrative Analysis of Sequencing and Array Genotype Data for Discovering Disease Associations with Rare Mutations (2016) 2016 Joint Statistical Meeting in Chicago.
4. Nguyen S, Guan W, Roetker N, Grove ML, Bressler J, **Li Y**, Fornage M, Boerwinkle E, North KE, Pankow JS, Demerath EW. Obesity-related methylation variants as predictors of coronary heart disease in African Americans: The Atherosclerosis Risk in Communities (ARIC) Study (2016) The Obesity Society's Annual Meeting in New Orleans, LA October 31 - November 4, 2016.
5. Xu Z, Zhang G, Jin F, Chen M, Furey TS, Sullivan PF, Qin Z, Hu M#, **Li Y#**. A hidden Markov random field based Bayesian method for the detection of long-range chromosomal interaction in Hi-C data (2015) 65th meeting of the Am Soc Hum Genet.
6. Guan W, Wu C, Wu B, **Li Y**, Pankow J, Demerath EW, Bressler J, Fornage M, Grove ML, Boerwinkle E. Missing data imputation using genome-wide DNA methylation data (2015) 65th meeting of the Am Soc Hum Genet.
7. Cannon ME, Raulerson CK, Duan Q, Ko A, Pajukanta P, Laakso M, **Li Y**, Mohlke KL (2015) 65th meeting of the Am Soc Hum Genet.
8. Huang KC, Sun W, Wu Y, Chen MJ, Mohlke KL, Lange LA, **Li Y#**. Association Studies with Imputed SNPs Using Expectation-Maximization-Likelihood-Ratio Test (2014) 2014 Joint Statistical Meeting in Boston, MA.
9. Xu Z, Zhang G, Jin F, Hu M#, **Li Y#**. An HMRF-Based Bayesian Method for Chromatin Interaction Calling from Hi-C Data (2014) 2014 Joint Statistical Meeting in Boston, MA.
10. Yan S, **Li Y#** (2014) Likelihood based complex trait association testing for arbitrary depth sequencing data. ENAR 2014 in Baltimore, Maryland.
11. Huang KC, **Li Y#** (2014) Association studies with imputed SNPs using Expectation-Maximization likelihood ratio test. ENAR 2014 in Baltimore, Maryland.
12. Buchkovich ML, Eklund K, Duan Q, **Li Y**, Mohlke KL, Furey TS (2013) Accurate identification of allele imbalances in samples with limited genotype information available. 63rd meeting of the Am Soc Hum Genet.

13. Byrnes AE, Wu MC, Lange LA, Li M, **Li Y#** (2013) SKAT-Admix: Extending the SKAT method for rare variant association to admixed populations. 63rd meeting of the Am Soc Hum Genet.
14. Huang KC, **Li Y**# (2013) Association Studies with Imputed SNPs Using Expectation-Maximization-Likelihood-Ratio Test. 63rd meeting of the Am Soc Hum Genet.
15. Sun W, **Li Y**, Perou C. Using RNA-Seq Data to Study the Genetic Basis of Cancer Development (2013) 2013 Joint Statistical Meeting in Montreal, Canada.
16. **Li Y**, Yan S. A Powerful Method to Control Size of Unbalanced Sequencing Designs in Rare Variant Association Testing (2013) 2013 Joint Statistical Meeting in Montreal, Canada.
17. **Li Y**, Yan S. BETASEQ: A Powerful Novel Method to Control Type①]I Error Inflation in Partially Sequenced Data for Rare Variant Association Testing (2013) The Second Taihu International Statistics Forum.
18. **Li Y**, Yan S. Efficient Two-stage Association Test of Sequence Data without Genotype Calling (2013). ICSA/ISBS Joint Statistical Conference 2013 in Washington, D.C.
19. Chen W, Li B, Zeng Z, Sanna S, Sidore C, Busonero F, Kang HM, **Li Y**, Abecasis GR (2013) Genotype Calling and Haplotyping for Family-based Sequence Data. ENAR 2013 in Orlando, Florida.
20. Urrutia E, **Li Y**, Wu M (2013) Smoothed Stability Selection for Analysis of Sequencing Data. ENAR 2013 in Orlando, Florida.
21. **Li Y**, Li M, Liu Y, Mao X, Wang W (2013) Missing Genotype Inference and Association Analysis of Rare Variants in Admixed Populations. ENAR 2013 in Orlando, Florida.
22. Zhang G, Wu C, Yu D, Chang J, Sun W, Li M, Liang L, **Li Y#**, Lin D# (2012) A Comprehensive Genetic and Epigenetic Study of Hepatocellular Carcinoma with Microarrays. 62nd meeting of the Am Soc Hum Genet.
23. Yan S, **Li Y#** (2012) Efficient Two-Stage Analysis Approach for Complex Trait Association with Arbitrary Depth Sequencing Data. 62nd meeting of the Am Soc Hum Genet.
24. Roman TS, Fogarty MP, Vadlamudi S, Marvelle AF, Gaulton KJ, Gonzalez AJ, **Li Y**, Mohlke KL (2012) Identification of a regulatory variant that binds C/EBPβ at the *GALNT2* human high-density lipoprotein cholesterol locus. 62nd meeting of the Am Soc Hum Genet.
25. Liu EY, Li M, Wang W, **Li Y#** (2012) MaCH-Admix: Genotype Imputation for Admixed Populations. 62nd meeting of the Am Soc Hum Genet.
26. Hu YJ, **Li Y**, Lange LA, Lange EM, Bizon C, Auer P, Heiss G, Kooperberg C, Franceschini N, Peters U, Reiner AP, Hsu L, Jiao S, Carlson CS, North KE, Lin DY (2012) Integrative Analysis of Sequencing and GWAS Data Improves Statistical Power in Detecting Rare Variants Associated With Complex Diseases. 62nd meeting of the Am Soc Hum Genet.
27. Huang KC, Ehm MG, Bing N, Liu EY, Xu J, Slater A, Fraser D, Novembre J, Li J, Nelson MR, **Li Y#** (2012) Patterns of Indel Variation: Comparison of calling methods and LD with SNPs. 62nd meeting of the Am Soc Hum Genet.
28. Duan Q, Liu EY, Croteau-Chonka DC, Mohlke KL, **Li Y#**. Evaluation of SNP and indel imputation quality using reference haplotypes from the 1000 Genomes project. 62nd meeting of the Am Soc Hum Genet.
29. Byrnes AE, Li M, Wu MC, Wright FA, **Li Y#** (2012) Variable Selection Based Weighting Schemes for Implicating Rare Variants in Sequence Data. 62nd meeting of the Am Soc Hum Genet.
30. Ramasamy A, Trabzuni D, Gibbs RJ, Dillman A, Hernandez DG, Arepalli S, Walker R, Smith C, Shabalin AA, **Li Y**, Singleton AB, Cookson MR, Hardy J, Ryten M, Weale ME

(2012) The impact of polymorphisms within probe sequences on expression QTL studies. 62nd meeting of the Am Soc Hum Genet.
31. Urrutia G, **Li Y**, Wu MC (2012) Smoothed Stability Selection for Analysis of Sequencing Data. 2012 Joint Statistical Meeting in San Diego.
32. **Li Y**, Yan S (2012) Efficient Two-Stage Association Analysis with Arbitrary Depth Sequencing-Based Studies. 2012 Joint Statistical Meeting in San Diego.
33. Byrnes AE, Li M, Wu MC, Wright FA, **Li Y#** (2012) Variable Selection in Weighting Schemes for Implicating Rare Variants in Sequence Data. 2012 Joint Statistical Meeting in San Diego.
34. Byrnes A, Li M, **Li Y#** (2012) Association Analysis of Rare Variants in Sequencing-Based Studies for Complex Human Traits. The 2012 ICSA Applied Statistics Symposium.
35. **Li Y**, Liu EY, Li M, Wang W (2011) MaCH-Admix: Genotype Imputation for Admixed Populations. 61st meeting of the Am Soc Hum Genet.
36. Byrnes A, Li M, **Li Y#** (2011) A Comprehensive Evaluation of SNP Weighting Schemes for the Analysis of Uncommon Genetic Variants. 61st meeting of the Am Soc Hum Genet.
37. Liu EY, Buyske S, Aragaki AK, Peters U, Boerwinkle E, Carlson C, Carty C, Crawford DC, Haessler J, Hindorff LA, Marchand LL, Manolio TA, Matise T, Wang W, Kooperberg C, North KE, **Li Y#** (2011) Genotype Imputation of Metabochip SNPs Using a Study-Specific Reference Panel of 3,924 Haplotypes in African Americans from the Women's Health Initiative. 61st meeting of the Am Soc Hum Genet.
38. Bing N, Huang KC, Liu EY, Li J, Fraser D, Aponte J, Liu X, Zhang H, Slater A, Woollard P, Chissoe SL, Whittaker JC, Mooser VE, Ehm MG, Zolner S, Novembre J, **Li Y#**, Nelson MR (2011) Patterns of indel variation in 202 drug target genes from >14,000 individuals. 61st meeting of the Am Soc Hum Genet.
39. Mao X, **Li Y**, Li M (2011) Detecting Genetic Association with Rare Variants in Admixed Populations. 61st meeting of the Am Soc Hum Genet.
40. Franceschini N, Reiner A, Chi T, Stroller ML, Kahn A, Carty C, **Li Y**, Edwards TL, Jackson R (2011) Novel genetic locus for kidney stone using diverse ancestry populations: the Women's Health Initiative study. 61st meeting of the Am Soc Hum Genet.
41. Novembre J, Wegmann D, Zawistowski M, Rakshi A, Gopalakrishnan S, Kessner D, St. Jean P, Li L, Ehm MG, Li J, **Li Y**, Abecasis G, Whittaker JC, Chissoe SL, Mooser VE, Nelson MR, Zollner S (2011) The geographic structure of allele sharing and rare variant diversity assessed from re-sequencing of 202 genes in 15,000 individuals. 61st meeting of the Am Soc Hum Genet.
42. **Li Y**, Waterworth DM, Li L, Zhou Y, Vollenweider P, Waeber G, Mooser VE, Abecasis GR, Ehm MG (2010) Leveraging publicly available sequencing data in the post-GWAS era to identify novel significant association signals. 60th meeting of the Am Soc Hum Genet.
43. Li M, **Li Y**, Cheung VG (2010) Global patterns of RNA editing in humans. 60th meeting of the Am Soc Hum Genet.
44. Byrnes A, Li M, **Li Y** (2010) Haplotype and dosage-based tests for rare variant association. 60th meeting of the Am Soc Hum Genet.
45. Chen W, Li B, **Li Y**, Abecasis G (2010) An efficient LD-based variant calling and phasing method for next generation sequencing in trios. 60th meeting of the Am Soc Hum Genet.
46. Zhan X, Anderson P, **Li Y**, Abecasis G (2010) KARMA: K-tuple alignment with rapid matching algorithm. 60th meeting of the Am Soc Hum Genet.

47. Sidore C, Abecasis G, Kang HM, **Li Y**, Sanna S, Zollner S, Lo Y (2010) Software for generating linkage-disequilibrium aware genotype calls from next generation sequence data. 60th meeting of the Am Soc Hum Genet.
48. Wu Y, **Li Y**, Ethan EM, Croteau-Chonka DC, Kuzawa CW, McDade TW, Qin L, Curochicin G, Borja JB, Lange LA, Adair LS, Mohlke KL (2010) Genome-wide association study for adiponectin levels in Filipino women identifies CDH13 and a novel uncommon haplotype at *KNG1-ADIPOQ*. 60th meeting of the Am Soc Hum Genet.
49. Croteau-Chonka DC, Wu Y, **Li Y**, Lange LA, Kuzawa CW, McDade TW, Laakso M, Borja JB, Adair LS, Lange EM, Mohlke KL (2010) An uncommon SNP strongly associated with adiponectin levels in Filipinos is indirectly associated with a GWA signal 800 kb away at the *ADIPOQ* gene. 60th meeting of the Am Soc Hum Genet.
50. Fogarty MP, Bucholovich ML, Gaulton KJ, **Li Y**, Mohlke KL (2010) Evaluation of three molecular mechanisms for altered *MMAB* transcript level at a locus associated with high density lipoprotein cholesterol. 60th meeting of the Am Soc Hum Genet.
51. Roman TS, Marvelle AF, Gaulton KJ, Fogarty MP, Gonzalez AJ, **Li Y**, Mohlke KL (2010) Allele-specific regulatory activity of variants associated with human high-density lipoprotein cholesterol level at the *GALNT2* locus. 60th meeting of the Am Soc Hum Genet.
52. Lo YY, Sidore C, **Li Y**, Li B, Li J, Verzilli C, Nangle K, Chissoe SL, Nelson MR, Ehm MG, Abecasis G, Zollner S (2010) Imputation-based genotype calling in a worldwide sample of 15, 000 individuals. 60th meeting of the Am Soc Hum Genet.
53. Ehm MG, Li L, Song K, Bacanu SA, Cox C, Aponte J, Mitchell JK, Chissoe SL, Fraser D, Briley D, Yuan X, Verzilli C, Shen J, Nangle K, Vollenweider P, Waeber G, Cardon LR, Mooser VE, Waterworth DM, Whittaker JC, Nelson MR, **Li Y** (2010) Strengths and limitations of follow-up imputation to investigate variants originally identified through deep-sequencing experiments. 60th meeting of the Am Soc Hum Genet.
54. Nelson MR, Ehm MG, Warren L, Verzilli C, Shen J, Fraser D, Aponte J, Novembre J, Wegmann D, Li J, Zollner S, **Li Y**, St Jean P, Li L, Woollard P, Topp S, Hall M, Nangle K, Abecasis G, Cardon LR, Whittaker JC, Chissoe SL, Mooser V (2010) Deep resequencing study of 202 genes in 15, 000 individuals across 12 diseases to support drug repositioning. 60th meeting of the Am Soc Hum Genet.
55. Novembre J, Wegmann D, Gopalakrishnan S, Zawistowski M, St Jean P, Li L, Ehm MG, Li J, **Li Y**, Abecasis G, Whittaker JC, Chissoe SL, Mooser VE, Nelson MR, Zollner S (2010) The geographic structure of rare variant diversity. 60th meeting of the Am Soc Hum Genet.
56. Kapur K, Johnson T, Beckmann N, Sehmi J, Tanaka T, Kutalik Z, Styrkarsdottir U, Zhang W, Marek D, Gudbjartsson D, Milaneschi Y, Holm H, DiIorio A, Waterworth D, **Li Y**, Singleton A, Bjornsdottir U, Sigurdsson G, Hernandez D, DeSilva R, Elliott P, Eyjolfsson G, Guralnik J, Scott J, Thorsteinsdottir U, Bandinelli S, Chambers J, Stefansson K, Waeber G, Ferrucci L (2010) Genome-wide meta-analysis for serum calcium identifies significantly associated SNPs near the calcium-sensing receptor gene. 60th meeting of the Am Soc Hum Genet.
57. Bowden DW, Bielinski SJ, Kao L, Siscovick D, Patel SR, Zmuda JR, Meigs JB, Sims M, Sarpong D, Rich SS, Freedman BI, Goodarzi MO, Grant SFA, Langefeld CD, Allred ND, Pankow JS, **Li Y**, Lange LA, Wilson JG, Ng MC, and the Candidate Gene Association Resource (2010) Meta analysis of African American genomewide association studies of type 2 diabetes: The CARe T2D Plus Study. 60th meeting of the Am Soc Hum Genet.

58. Ellinghaus E, Ellinghaus D, Stuart PE, Nair RP, Debrus S, Raelson JV, Belouchi M, Fournier H, Reinhard C, Ding J, **Li Y**, Tejasvi T, Gudjonsson J, Stoll SW, Lambert S, Weidinger S, Eberlein B, Kunz M, Rahman P, Gladman D, Gieger C, Wichmann HE, Karlsen TH, Kabelitz D, Mrowietz U, Abecasis GR, Elder JT, Schreiber S, Weichenthal M, Franke A (2010) Genome-wide association study for psoriasis. 60th meeting of the Am Soc Hum Genet.
59. Sanna S, Pitzalis M, Zoledziewska M, Zara I, Sidore C, Murru R, Whalen MB, Scirru L, Secci MA, Deidda F, Deidda L, Barizzone N, Poddie F, Morelli L, Farina G, Dei M, Lai S, Mulas A, **Li Y**, Pugliatti M, Traccis S, Angius A, D'alfonso S, Melis M, Rosati G, Abecasis GR, Uda M, Marrosu MG, Schlessinger D, Cucca F (2010) Variants within the *CBLB* gene are associated with multiple sclerosis. 60th meeting of the Am Soc Hum Genet.
60. **Li Y**, Handsaker RE, Abecasis GR, McCarroll SA (2010) Accurate CNV Genotyping from Massively Parallel Sequencing Data. The Biology of Genomes, Cold Spring Harbor, NY.
61. **Li Y**, Abecasis GR (2009) Taking Advantage of Distant Relatedness: Genotype Imputation in the Resequencing Era. 59th meeting of the Am Soc Hum Genet.
62. Li B, **Li Y** , Sanna S, Schlessinger D, Najjar S, Scuteri A, Lakkata E, Boehnke M, Abecasis GR and Uda M for Sardinian Project (2009) Fine mapping of common and rare variants associated with low-density lipoprotein cholesterol (LDL-C) via sequencing candidate loci following genome-wide scans. 59th meeting of the Am Soc Hum Genet.
63. Lange LA, Marvelle AF, Croteau-Chonka D, Kuzawa C, McDade TW, **Li Y** , Levy S, Daniels M, Borja J, Lange EM, Adair LS and Mohlke K (2009) Genome-wide association study of homocysteine levels in Filipinos identifies a new locus (*CPS1*) and evidence for a stronger *MTHFR* effect in young adults than in their mothers. 59th meeting of the Am Soc Hum Genet.
64. **Li Y**, Abecasis GR (2008) Efficient Reconstruction of Whole Genomes Using Massively Parallel Shotgun Sequence Data. 58th meeting of the Am Soc Hum Genet.
65. Zheng J, **Li Y**, Abecasis GR, Scheet P (2008) A Comparison of Approaches to Account for Uncertainty in Analysis of Imputed Genotypes. 58th meeting of the Am Soc Hum Genet.
66. Huang L, **Li Y**, Singleton AB, Hardy JA, Abecasis GR, Rosenberg NA, Scheet P (2008) Genotype Imputation Accuracy Across World Human Populations. 58th meeting of the Am Soc Hum Genet.
67. Arnold S, Guy M, Kashuk K, **Li Y**, Abecasis GR, Chakravarti A (2008) Semaphorins as Candidate Genes in Hirschsprung Disease. 58th meeting of the Am Soc Hum Genet.
68. Xiang F, Scott LJ, **Li Y**, Jackson AU, Willer CJ, Stringham HM, Erdos MR, Bonnycastle LL, Kubalanza K, Swift AJ, Abecasis GR, Mohlke KL, Tuomilehto J, Bergman RN, Collins FS, Watanabe RM, Boehnke M (2008) Genome-wide Association for Insulin Resistance and Secretion in 542 Genotyped and Imputed Individuals. 58th meeting of the Am Soc Hum Genet.
69. Chen W, **Li Y**, Abecasis GR (2008) State Space Reduction in Hidden Markov Model for Haplotyping, Imputation and Analysis of Shotgun Sequence Data. 58th meeting of the Am Soc Hum Genet.
70. **Li Y**, Willer CJ, Ding J, Scheet P, Abecasis GR (2008) Rapid Genotype Imputation and Analysis of Resequencing Data Using Markov Models. 2008 Joint Statistical Meeting in Denver.
71. **Li Y**, Willer CJ, Ding J, Scheet P, Abecasis GR (2007) *In silico* Genotyping for Genome-wide Association Studies. 57th meeting of the Am Soc Hum Genet.

72. Sanna S, Jackson AU, Usala G, Willer CJ, Dei M, Bonnycastle LL, Lai S, **Li Y**, Uda M, Erdos MR, Shen H, Shuldiner A, Cao A, Bergam RM, Schlessinger D, Collins FS, Boehnke M, Abecasis GR, Nagaraja R, Mohlke KL (2007) Genome-wide Association Scan for Height in 6,671 Individuals from Finland and Sardinia. 57th meeting of the Am Soc Hum Genet.
73. **Li Y**, Abecasis GR (2006) Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. 56th meeting of the Am Soc Hum Genet.
74. Absher D, Li J, Thompson RC, Burmeister M, Scott LJ, **Li Y**, Meng F, Guan W, Vawter MP, Choudary P, Tomita H, Evans SJ, Bunney WE, Jones EG, Barchas JD, Schatzberg A, Akil H, Watson SJ, Boehnke M, Myers RM (2006) An Association Study of Ninety-three Candidate Genes in Bipolar Disorder. 56th meeting of the Am Soc Hum Genet.
75. Gaulton KJ, Conneely KN, **Li Y**, Jackson AU, Scott LJ, Duren WL, Chines PS, Narisu N, Bonnycastle L, Swift A, Valle TT, Tuomilehto J, Bergman RN, Collins FS, Boehnke M, Mohlke KL (2006) Testing for Association between Type 2 Diabetes and 225 Candidate Genes in 2357 Finnish Cases and Controls. 56th meeting of the Am Soc Hum Genet.

Invited oral presentations:

2016	The 4th IBS-China International Biostatistical Conference, Shanghai, China
2016	Center for Statistical Genetics, University of Michigan
2016	Division of Biostatistics, Department of Population Health, New York University School of Medicine
2015	Department of Biostatistics, University of Minnesota
2015	Department of Mathematics, Bowling Green State University
2015	Institute for Behavioral Genetics, University of Colorado, Boulder
2015	Institute for Personalized Medicine (IPM), Penn State University
2015	ENAR, the International Biometric Society
2014	Department of Biomathematics, University of California, Los Angeles
2014	Department of Biostatistics, Harvard University
2014	Department of Biostatistics, University of Pittsburgh
2014	Nanjing Medical University
2014	Centre for Genomic Sciences, the University of Hong Kong
2014	The Hong Kong University of Science and Technology
2014	Department of Biostatistics, University of Pittsburgh
2014	Workshop on Emerging Statistical Challenges and Methods For Analysis of Massive Genomic Data in Complex Human Disease Studies, Banff International Research Station for Mathematical Innovation and Discovery, Banff, Canada.
2014	Workshop on Big Data in Genetics and Toxicogenomics, Durham, NC
2014	Department of Statistics, Purdue University, West Lafayette, IN
2014	Department of Biostatistics, Columbia University, New York
2013	2013 Triangle Statistical Genetics Conference, Durham, North Carolina
2013	2013 Joint Statistical Meeting in Montreal, Canada
2013	Department of Computer Science, Jiangnan University, Wuxi, China
2013	The Second Taihu International Statistics Forum
2013	ICSA/ISBS (International Chinese Statistical Association/International Society of Biopharmaceutical Statistics) Joint Statistical Conference in Washington, D.C.

2013 2013 ENAR (Eastern North American Region International Biometric Society, Orlando, Florida
2012 Department of Human Genetics, Emory University, Atlanta, Georgia
2012 Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio
2012 2012 ICSA Applied Statistics Symposium
2011 Quantitative Genomics Seminar Series, Division of Human Genetics at Cincinnati Children's Hospital, University of Cincinnati
2010 8th ICSA International Conference, Guangzhou, China
2010 Section of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
2010 Division of Biostatistics, Washington University School of Medicine
2010 2010 ICSA Applied Statistics Symposium
2010 National Institute of Environmental Health Sciences
2010 Center for Genomics and Personalized Medicine Research, Wake Forest University
2010 Cancer Institute and Hospital, Chinese Academy of Medical Sciences (CAMS)
2009 Genetics, University of North Carolina
2009 Lady Davis Institute of Medical Research, Department of Epidemiology, Biostatistics and Occupational Research, McGill University
2009 School of Public Health, Yale University
2009 Biostatistics, University of North Carolina
2009 Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York
2007 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University
2007 Pharmacogenetics, GlaxoSmithKline
2007 ICSA Symposium

TEACHING RECORD

Course/Module Director:

EPI 889: Topics in Epidemiology: Advanced Genetic Epidemiology – High Throughput Data Analysis, 12 contact hours, Fall 2012, 18 students; Fall 2014, 5 students.
BIOS 735: Statistical Computing (9 hours for lecturing/leading discussions), Fall 2013, 23 students; Fall 2015, 22 students
BCB 725/Bios681: Introduction to Statistical Genetics, 3 credit hours (24 hours for lecturing/leading discussions), Spring 2012, 17 students; Spring 2015, 9 students

Lectures at UNC:

BCB 720: Introduction to Statistical Modeling, 1.5 contact hours, 2015
BIOS 740: Statistical Methods for Genetic Association Studies, 4.5 contact hours, 2012, 2014
BIOS 784: Computational Biology, 4.5 contact hours, 2013, 2015
EPID 743: Genetic Epidemiology, 1.5 contact hours, 2010, 2011, 2012, 2013, 2014, 2016
BCB 710: Bioinformatics Colloquium, 1.5 contact hours, 2009, 2010, 2013, 2015

Lectures outside UNC:

✧ Joint Statistical Meeting (JSM) Professional Development Continuing Education Course on

Analysis of Genome-Wide Sequencing Association Studies, jointly taught with Xihong Lin and Michael Wu - 2014
- ✧ Short Course on Genetic Epidemiology, Nanjing Medical University, Nanjing, China - 2013
- ✧ 2nd Annual Short Course on Next Generation Sequencing: Technology and Statistical Methods, UAB in Birmingham, Alabama – 2012
- ✧ ICSA Short Course on the Analyses of Next Generation Sequencing Studies, Boston, 2012
- ✧ Statistical Genetics Short Course, Chapel Hill, North Carolina, 2010 summer
- ✧ 7th Course "SNP's and Human Diseases", Erasmus MC, Rotterdam, the Netherlands, 2010
- ✧ Graduate Student Instructor for Biostat605 (Introduction to SAS), University of Michigan, 2008

Postdoctoral Scientist Advisor:

2016- Laura Raffield, Ph.D. in Molecular Genetics and Genomics
2016- Qing Duan, Ph.D. in Bioinformatics and Computational Biology
2012-2016 Zheng Xu, Ph.D. in Statistics.
2012-2015 Song Yan, Ph.D. in Statistics. Current position: data scientist at Microsoft Bing Ads.

Graduate Student Thesis Advisor:

2012-2016 Guosheng Zhang, Ph.D. student, Curriculum in Bioinformatics and Computational Biology. First position: Software Engineer Internship at Google.
2012-2016 Qing Duan, Ph.D. student, Curriculum in Bioinformatics and Computational Biology.
2010-2015 Kuan-Chieh Huang, Ph.D. student, Department of Biostatistics. First position: Sr. Biostatistician at Gilead Sciences, Inc.
2009-2013 Andrea Byrnes, Ph.D. student, Department of Biostatistics. First position: Postdoctoral Fellow, Analytic and Translational Genetics Unit at Massachusetts General Hospital, the Broad Institute
2009-2013 Yi Liu, Ph.D. student, Department of Computer Science. First position: Applied Researcher with Microsoft

Graduate Student Thesis Advisory Committees:

2015- Christopher Bryant, Ph.D. student, Department of Biostatistics
2014- Xiaolei Zhou, Ph.D. student, Department of Biostatistics
2014- Robert Corty, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2015- Matthew Psioda, Ph.D. student, Department of Biostatistics
2015- Matthew Holt, Ph.D. student, Department of Computer Science
2015- Wei Xue, Dr.PH student, Department of Biostatistics
2015 Thomas Conrad, M.S. student, Department of Biostatistics
2015-2016 Ran Tao, Ph.D. student, Department of Biostatistics
2014-2016 Hojin Yang, Ph.D. student, Department of Biostatistics
2014-2015 Pratyaydipta Rudra, Ph.D. student, Department of Biostatistics
2013-2015 Jin Li, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2012-2015 Martin Buchkovich, Ph.D. student, Curriculum in Bioinformatics and Computational Biology

2013-2014 Shuai Yuan, Ph.D. student, Computer Science and Informatics, Emory University
2013-2014 Yunfei Wang, Dr.Ph. student, Department of Biostatistics
2012-2014 TingHuei Chen, Ph.D. student, Department of Biostatistics
2013-2014 Ni Zhao, Ph.D. student, Department of Biostatistics
2013-2014 Zhengzheng Tang, Ph.D. student, Department of Biostatistics
2013-2014 Gene Urrutia, Ph.D. student, Department of Biostatistics
2012-2013 Gregory Mayhew, Ph.D. student, Department of Biostatistics
2012-2013 Ni Zhao, M.S. student, Department of Biostatistics
2010-2014 Gabi Griffin, Ph.D. student, Curriculum in Genetics and Molecular Biology
2010-2012 Damien Croteau-Chonka, Ph.D. student, Curriculum in Genetics and Molecular Biology and Bioinformatics and Computational Biology Training Program
2010-2012 Yihui Zhou, Ph.D. student, Department of Biostatistics
2009-2010 Lindsey Ho, Dr.Ph. student, Department of Biostatistics
2009-2010 John Schwatz, Ph.D. student, Department of Biostatistics

Graduate Student Academic Advisor:
2016- Kayla Kilpatrick, Ph.D. student, Department of Biostatistics
2016- Hillary Heiling, Ph.D. student, Department of Biostatistics
2013-2015 Yinghao Pan, Ph.D. student, Department of Biostatistics
2012-2014 Xiaoqiang Xue, Dr.PH., Department of Biostatistics

Graduate Rotation Students:
2015 Winter Dan Liang
2014 Spring Shengjie Chai
2013 Spring Greg Keele
2012 Spring Guosheng Zhang
2011 Fall Qing Duan

Visiting Scholar:
2013-2014 Wei Chen
2014-2015 Suhua Chang

Programmer:
2012-2013 Yunfei Wang

Statistician:
2013-2014 Yurong Lu

GRANTS

ACTIVE

1R01HL129132 Li (PI) 07/15/16-02/29/20
NIH/NHLBI $658,579 20% Efforts
Genetic Studies of Blood Cell Traits in Multi-Ethnic Cohorts

The goal of this study is to map, annotate and validate genes for blood cell traits in multi-ethnic cohorts to increase our knowledge of blood cell trait genetics.
Role: PI

1R01HG006292 Li (PI) 08/23/11-05/31/17
NIH/NHGRI $250,000 30% Efforts
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

U54 HD079124 Piven(PI) 09/24/13-05/31/18
NIH $107,539 10% Efforts
Clinical Translational Research Center for Neurodevelopmental Disorders
Intellectual and Developmental Disabilities Research Center.
Role: Core Director

1R01DK101855 North(PI) 08/15/14-07/31/17
NIH/NIDDK $605,779 6% Efforts
Leveraging ancestral diversity to map adiposity loci in Hispanics
The goal of this project is to perform the first large-scale genomic study in search of obesity-susceptibility loci in HL populations.

5R01ES020836 Whitse(Pl) 08/06/12-04/30/17
NIH $343,515 5% Efforts
Epigenetic Mechanisms of PM-Mediated CVD risk
The goal of this project is to advance the understanding of epigenetic mechanisms underlying susceptibility to PM-mediated CVD risk in post-menopausal women

1R01DK093757-01 Mohlke (PI) 09/05/11-7/31/16
NIH/NIDDK $470,076 15% Efforts
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

PAST

1R01HG006703-01 Li (PI) 05/16/12-02/28/16
NIH/NHGRI $224,285 20% Efforts
Imputation and Analysis of Rare Variants in Admixed Populations
The goal of this study is to develop statistical methods and computational tools for the imputation of rare genetic variants in admixed populations.
Role: PI

Subcontract No. 3001352222 Li (PI) 12/01/09-12/11/14

GlaxoSmithKline/UMichigan $54,593 3% Efforts
The goal of this project is to develop and apply imputation based methodology to genome wide association and sequencing datasets.
Role: Subcontract PI

1R01 DA030976-01 Wilhelmsen (PI) 09/30/10-05/31/16
NIH $2,729,715 10% Efforts
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

5P01HD031921-15 Whitsel (PI) 08/01/10-03/15/13
NIH $493,201 8% Efforts
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

R01 MH090936-02 Rusyn (PI) 9/17/10-7/31/12
NIH/NIMH $220,904 4% Efforts
Facilitating GTEx, Disease, and GxE Analyses Via Fast Expression (e)QTL Mapping
The goal of this project is to develop new statistical tools and graphical user-friendly software to facilitate the analysis of eQTL studies.
Role: Co-Investigator

1RC2HL102924-01(NCE) Jackson, North (PI) 09/30/09-07/31/12
WHI Sequencing Project (WHISP)
NIH/Ohio State Univ. Sub $397,899 16% Efforts
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

5R01 HL095396-02 Knowles (PI) 09/24/08-07/31/12
NIH/NHLBI $515,788 10% Efforts
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

U01 DA024413 Costello (PI) 09/01/07-06/30/12
NIDA/Duke sub (Sullivan, UNC PI) $297,458 8% Efforts
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 DK078150-04 Mohlke (PI) 04/01/07-03/31/12
NIH/NIDDK $389,802 20% Efforts
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

3R01 CA082659-11S1 Lin (PI) 08/01/09-07/31/11
NIH/NCI $163,841 20% Efforts
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

PROFESSIONAL SERVICE

Service to Discipline:

NIH Study Section Reviewer:

Jul 2016: (co-chair) ZRG1 GGG-L (50) Study Section: Novel Genomic Technology Development
Feb 2016: Genomics, Computational Biology and Technology (GCAT) Study Section
Jun 2014: NHLBI Special Emphasis Panel ZHL1 CSR-X (O1)
Feb 2014: NHLBI Special Emphasis Panel ZHL1 CSR-X (M1)
Oct 2013: Biodata Management and Analysis (BDMA) Study Section
Feb 2012: Genomics, Computational Biology and Technology (GCAT) Study Section

Other Grant Reviews:

May 2016: Research Grant Counsel, Hong Kong
Mar 2015: Research Grant Counsel, Hong Kong
Jun 2013: ERC (European Research Council) Consolidator Grant Proposals
Jan 2013: Barts and the London Charity Grant Proposals
Apr 2012: Wellcome Trust and Royal Society Sir Henry Dale Fellowship

Editorial Board, Journals:

AIMS Genetics
Frontiers in Statistical Genetics and Methodology
PLoS ONE

Ad Hoc Reviewer, Journals:

The American Journal of Human Genetics

The 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 in Statistical Genetics and Methodology
Frontiers of Medicine
Genetic Epidemiology
Genetics
Genome Biology
Genome Research
Human Heredity
Human Molecular Genetics
International Journal of Biostatistics
Journal of Bioinformatics and Computational Biology
Nature Communications
Nature Genetics
Nature Methods
Pacific Symposium on Biocomputing (*PSB*)
PLoS Computational Biology
PLoS Genetics
PLoS ONE
Scientific Reports
Statistical Applications in Genetics and Molecular Biology
Theoretical Population Biology

Other:

Co-Organizer, Research Triangle Park Statistical Genetics Conference, 2011, 2012, 2013, 2014, 2015, 2016
Session Chair, Joint Statistical Meeting Invited Session, Miami Beach, FL 2011
Member, WIDTH Symposium Planning Committee, Ann Arbor, MI 2008

Service within UNC-Chapel Hill:

Member, 2013- Research Computing Advisory Committee
Member, 2012-2015 Course Organizing Committee: Statistical Software Development
Member, 2010- Department of Biostatistics Computing committee
Member, 2010-2015 Bioinformatics and Computational Biology Curriculum advisory committee
Member, 2014,2015 Search Committee for Statistical Genetics Faculty Position
Member, 2014 Search Committee for Computational Genomics Position
Member, 2014 Biostatistics Qualifying Exam Committee
Member, 2013-2014 Quantitative Sub Committee, Recruitment Committee for Biological

and Biomedical Sciences Program (BBSP)
Member, 2013-2014 School of Public Health Awards Committee
Member, 2013-2014 Doctoral Examinations Committee, Department of Biostatistics
Member, 2013-2014 Award Committee, Department of Biostatistics
Member, 2013-2014 HHMI International Student Research Fellowship
UNC Internal Review Committee
Member, 2013 Search Committee for Sequencing Informatics Position
Member, 2011-2012 Biostatistics Information Technology Committee,
Department of Biostatistics
Member, Spring 2011 BCB Statistics Course Committee,

Memberships in Professional Societies:

International Chinese Statistical Association, 2007-present
American Society of Human Genetics, 2005-present
American Statistical Association, 2002-present
American Association for the Advancement of Science, 2007-2008
American Society for Quality, 2003-2005

REFLECTIVE RESEARCH STATEMENT

My research interest is in statistical genetics, particularly for the dissection of genetic mechanisms underlying complex human traits. Specifically, my group focuses on the development, implementation, and application of statistical methods and computational tools for the understanding of human genetic variation, with a particular focus on complex human diseases and traits. My group has been focusing on the design and analysis of sequencing based studies for complex human traits.

Massively parallel sequencing has transformed the field of genomic studies. These technologies have resulted in the successful identification of causal variants for many rare Mendelian disorders. However, the development of robust statistical and computational methods has fallen behind the technological advances particularly for application to the study of complex human traits. My group has been attempting to tackle the following issues. First, we have been extending our genotype imputation methods (implemented in our software MaCH) to detect SNPs from a combination of sequencing and genotyping data, generate individual-level genotype calls with estimated uncertainty measures, and provide phased haplotypes (software *thunder* was developed). Our methods can accommodate low coverage sequencing data, *as well as be useful for high coverage data.* Second, our method, by allowing low coverage sequencing data, enables a wide range of alternative designs including low-depth sequencing in larger cohorts of individuals and staged designs involving subset of individuals sequenced at arbitrary depths. We have been examining these alternative designs in terms of variant discovery, accuracy of genotype calls at discovered loci, and effective sample sizes to aid study design and sample size calculations for association studies (online tool *AbCD* published). Third, we have been developing flexible region-based statistical methods for the analysis of rare genetic variants in both genetically homogeneous populations and in recently admixed populations including African Americans and Hispanics (WHaIT,SKAT, WHaIT-

admix/AWDS, MK-SKAT developed). Fourth, we have been developing computational more efficient genotype imputation, local ancestry inference, as well as association methods to handle ever-increasing reference panels and to accommodate admixed samples with more complex linkage disequilibrium patterns (software MaCH-Admix and aMAP developed, software SKAT-admix under development). The methods we propose can handle a combination of genotyping, sequencing, and imputed data, take various sources of uncertainty in the data into consideration, and easily incorporate prior biological knowledge and potential genetic, demographic or environmental confounders, and genetic variants that are associated with traits in different directions or in non-linear fashions, in both genetically homogenous as well as admixed populations.

In addition to method development, we also creatively apply the above methods to the genetic study of various human diseases and traits. One particular focus is on blood cell traits. Blood cell traits, including hemoglobin level, red blood cell (RBC), white blood cell (WBC), and platelet counts, are important intermediate clinical phenotypes for a variety of cardiovascular, hematologic, oncologic, immunologic and infectious disease. The distributions of these traits differ considerably across ethnicities. For example, further insight into the genetic determinants of “ethnic neutropenia” in African Americans (AA) may have implications for health disparities in risk of HIV infection, sickle cell disease severity, or mobilization of neutrophils or hematopoietic stem cells in cancer patients. Our preliminary data similarly suggest important differences in the distribution of blood cell traits in Hispanics compared to non-Hispanic whites. Investigations across multi-ethnicities are likely to have direct clinical impact by providing new avenues for treatment, particularly for minority patients with low blood counts, and illuminating new mechanisms by which genetic factors related to ethnic differences in blood cell counts contribute to U.S. health disparities for risk of chronic inflammatory and thrombotic diseases. We have assembled a large collection of AA and Hispanic/Latinos (HL) and will (1) map genes for blood cell traits in AA (n~34K) accounting for local ancestry; (2) conduct the first sequencing-aided GWAS of blood cell traits in HL (n~26K) accounting for 3-way local ancestry; and (3) meta-analyze across multi-ethnic datasets, annotate and validate novel association signals.

The above research has received three NIH R01 awards and will continue to be a main focus in my group in the next 4-5 years. In addition, another recent focus is on utilizing chromosome conformation capture (3C) derived technologies to understand the three dimensional (3D) structure of our genome and DNA looping regulatory interactions in general, and to aid the interpretation of regulatory variants identified in GWAS for prioritization in functional studies in particular. Data from 3C-derived technologies can provide information that can help shed light on potential functional mechanisms. For example, a recent *Nature* publication found long-range interactions between the obesity-associated intronic variants in *FTO* and *IRX3*, with the expression of *IRX3* rather than *FTO* being directly linked to body mass (Smemo et al *Nature* 2014, PMID: 24646999). Yet another recent proof-of-concept example was shown for the potential mechanistic elucidation of breast cancer GWAS variants (Dryden et al *Genome Research* 2014, published 13 August 2014; doi:10.1101/gr.175034.114). However, methods to analyze data from 3C-dereived technologies are truly in their infancy. We propose to: (1) develop statistically rigorous and computationally efficient methods to identify non-random interactions between pairs of loci in the genome; and (2) develop visualization

tools for the identifying potential target gene(s) of regulatory variant(s) of interest along with complementary epigenetic footprints.

My group implements all methods we propose into user friendly software which can benefit the community.

TEACHING STATEMENT

I love teaching, which is one major reason behind my decision to pursue an academic career. Teaching is not just a flow of knowledge from a teacher to students. Rather, it involves two-way learning, communication and interactions. A teacher is supposed to deliver knowledge to students, to encourage students to become motivated and learn independently, and to learn from students. In the following, I first describe my teaching experiences and then present my teaching philosophy.

(1) Teaching Experiences at UNC

After joining UNC, I have given guest lectures in several graduate level courses including BIOS 740 (Statistical Methods for Genetic Association Studies), BIOS735 (Statistical Computing and Software Development), EPID743 (Genetic Epidemiology), EPI889 (Advanced Genetic Epidemiology – High Throughput Data Analysis), and BCB710 (Bioinformatics Colloquium). It has been essential to design the lectures to cater the needs of students in each class. For example, for the epidemiology students, I try to teach the key concepts into context and focus more on interpretation; for the Biostatistics students, I give the mathematical and modeling details so that they can use similar models in other research problems; for the BCB students, I teach about implementation of the methods using different algorithms.

Besides guest lectures, I am the course director for a new course that I have opened with Dr. Ethan Lange: BCB725, Introduction to Statistical Genetics. The course has attracted a diverse audience when taught in Spring 2012: 17 registered students and 8 auditors. Among the registered students, 6 from BCB program, 6 from Computer Science, 3 from Biostatistics, and 2 from Epidemiology. We have designed the course to cater the needs of the diverse audience. In particular, (1) we have two lectures covering basic genetics and basic statistics related to the course at the beginning; (2) we provide knowledge that is beneficial only to some audiences (for example, mathematical derivations for Biostatistics students; source codes for BCB students) through the class website but do not attempt to cover in the lectures; (3) we have made some homework problems optional extra credits problems so that motivated students from relevant background are encouraged to tackle more challenging problems; and (4) give students flexible options for their final project. The course has been quite well received even with a potentially competing course BIOS740 that covers essentially the same topics but focus more on statistics. We will continue offering it in future years.

In addition to the regular teaching, I was also invited to teach in short courses, including the Statistical Genetics Short Course at Chapel Hill, North Carolina, in the summer of 2010; the 7th Course “SNP’s and Human Diseases”, Erasmus MC, Rotterdam, in the Netherlands, 2010; the short course on the Analyses of Next Generation Sequencing Studies, in Boston, 2012; the 2nd Annual Short Course on Next Generation Sequencing: Technology and Statistical Methods, UAB in Birmingham, Alabama; 2013 Short Course on Genetic Epidemiology, Nanjing Medical University, Nanjing, China; 2014 Joint Statistical Meeting (JSM) Professional Development Continuing Education Course on Analysis of Genome-Wide Sequencing Association Studies.

(2) Teaching Philosophy

I view learning as the responsibility of the student as well as the instructor. As an instructor, I would like to focus primarily on the responsibility on my side. The teacher serves a guide in students' pursuit of knowledge. The teacher has already traveled the intellectual ground that the students are covering. Because of this, he or she knows many of the pitfalls and dangers that exist. The teacher is also aware of many of the interesting and wonderful aspects that can be seen along the way. It is the responsibility of the teacher to point out as many of these as possible to the students.

When necessary, the teacher is also responsible for encouraging students to continue this quest beyond the course he or she is teaching. The teacher must be equipped with a myriad of techniques to accomplish this goal. The teacher must be willing to use both rewards and punishments to help the students continue. A teacher can encourage and inspire, but not force knowledge upon a student. A teacher should try his or her best to make the class instructive as well as full of fun. For the above reasons, I always try my best (1) to give a lecture focused on intuitions and concepts, rather than the technical details; (2) to clearly explain and deeply explore less/selected material rather than skimming over everything; and (3) to know my students well through in-class and out-of-class interactions so that I can work out different study plans for different students.

Finally, a teacher must be willing to learn more. Just because the teacher has traveled this intellectual ground before does not mean that he or she knows everything about it. The teacher must be willing to learn and adapt. The teacher should not only keep broadening his/her knowledge, but also keep sharpening his/her teaching techniques from students and peer evaluations.

In conclusion, a teacher is a guide who can aid and encourage students in their quest for knowledge. The teacher cannot force them to continue this quest, but can and should take great efforts to encourage them. The teacher must also be willing to learn from what his or her students are willing to teach them.