CURRICULUM VITAE

PERSONAL INFORMATION

Name: **Yun Li**
Office address: 5090 Genetic Medicine Building, 120 Mason Farm Rd., Chapel Hill, NC 27599
Phone: 919-843-2832
E-mail address: yunli@med.unc.edu
Lab website: https://yunliweb.its.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

Professor of Genetics, UNC, Chapel Hill, NC,	2021-
Professor of Biostatistics, UNC, Chapel Hill, NC,	2021
Adjunct Associate Professor, Applied Physical Sciences, UNC, Chapel Hill, NC	2019-
Director, Data Science Core, Intellectual and Developmental Disabilities Research Center (IDDRC) , UNC, Chapel Hill, NC,	2015-
Associate Professor of Genetics, UNC, Chapel Hill, NC,	2015-2021
Associate Professor of Biostatistics, UNC, Chapel Hill, NC,	2015-2021
Adjunct Assistant Professor of Computer Science, UNC, Chapel Hill, NC,	2012-2018
Member, Carolina Center for Genome Sciences, UNC, Chapel Hill, NC,	2009-
Assistant Professor of Genetics, UNC, Chapel Hill, NC,	2009-2015
Assistant Professor of Biostatistics, UNC, Chapel Hill, NC,	2009-2015

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
1998-1999: Harler Scholarship for Excellence in German Language Studies
1997-2001: Annual Academic Scholarship, Shanghai Jiaotong University
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

As of 10/29/2020 on Google Scholar, publications below have a total of 66,902 citations, with an H-index score of 64 and an i10-index score of 114.

1. Song M, Pebworth MP, Yang X, Abnousi A, Fan C, Wen J, Rosen JD, Choudhary MNK, Cui X, Jones IR, Bergenholtz S, Eze UC, Juric I, Li B, Maliskova L, Lee J, Liu W, Pollen AA, **Li Y**, Wang T, Hu M, Kriegstein AR, Shen Y (2020) Cell-type-specific 3D epigenomes in the developing human cortex. *Nature* 587(7835):644-649. PMID: 33057195.

2. Li G, Raffield L, Logue M, Miller MW, Santos HP Jr, O'Shea TM, Fry RC, **Li Y#** (2020) CUE: CpG impUtation ensemble for DNA methylation levels across the human methylation450 (HM450) and EPIC (HM850) BeadChip platforms. *Epigenetics* doi: 10.1080/15592294.2020. PMID: 33016200.

3. Van Buren E, Hu M, Weng C, Jin F, Li Y, Wu D#, **Li Y**# (2020) TWO-SIGMA: A novel two-component single cell model-based association method for single-cell RNA-seq data. *Genet Epidemiol* doi: 10.1002/gepi.22361. PMID: 32989764.

4. Spracklen CN, Iyengar AK, Vadlamudi S, Raulerson CK, Jackson AU, Brotman SM, Wu Y, Cannon ME, Davis JP, Crain AT, Currin KW, Perrin HJ, Narisu N, Stringham HM, Fuchsberger C, Locke AE, Welch RP, Kuusisto JK, Pajukanta P, Scott LJ, **Li Y**, Collins FS, Boehnke M, Laakso M, Mohlke KL (2020) Adiponectin GWAS loci harboring extensive allelic heterogeneity exhibit distinct molecular consequences. *PLoS Genet.* 16(9):e1009019. PMID: 32915782.

5. Justice AE, Chittoor G, Gondalia R, Melton PE, Lim E, Grove ML, Whitsel EA, Liu CT, Cupples LA, Fernandez-Rhodes L, Guan W, Bressler J, Fornage M, Boerwinkle E, **Li Y**, Demerath E, Heard-Costa N, Levy D, Stewart JD, Baccarelli A, Hou L, Conneely K, Mori TA, Beilin LJ, Huang RC, Gordon-Larsen P, Howard AG, North KE (2020) Methylome-wide

association study of central adiposity implicates genes involved in immune and endocrine systems. Epigenomics 12(17):1483-1499. PMID: 32901515.

6. Chen MH*, Raffield LM*, Mousas A*, Sakaue S, Huffman JE, Moscati A, Trivedi B, Jiang T, Akbari P, Vuckovic D, Bao EL, Zhong X, Manansala R, Laplante V, Chen M, Lo KS, Qian H, Lareau CA, Beaudoin M, Hunt KA, Akiyama M, Bartz TM, Ben-Shlomo Y, Beswick A, Bork-Jensen J, Bottinger EP, Brody JA, van Rooij FJA, Chitrala K, Cho K, Choquet H, Correa A, Danesh J, Di Angelantonio E, Dimou N, Ding J, Elliott P, Esko T, Evans MK, Floyd JS, Broer L, Grarup N, Guo MH, Greinacher A, Haessler J, Hansen T, Howson JMM, Huang QQ, Huang W, Jorgenson E, Kacprowski T, Kähönen M, Kamatani Y, Kanai M, Karthikeyan S, Koskeridis F, Lange LA, Lehtimäki T, Lerch MM, Linneberg A, Liu Y, Lyytikäinen LP, Manichaikul A, Martin HC, Matsuda K, Mohlke KL, Mononen N, Murakami Y, Nadkarni GN, Nauck M, Nikus K, Ouwehand WH, Pankratz N, Pedersen O, Preuss M, Psaty BM, Raitakari OT, Roberts DJ, Rich SS, Rodriguez BAT, Rosen JD, Rotter JI, Schubert P, Spracklen CN, Surendran P, Tang H, Tardif JC, Trembath RC, Ghanbari M, Völker U, Völzke H, Watkins NA, Zonderman AB; VA Million Veteran Program, Wilson PWF, **Li Y**, Butterworth AS, Gauchat JF, Chiang CWK, Li B, Loos RJF, Astle WJ, Evangelou E, van Heel DA, Sankaran VG, Okada Y, Soranzo N, Johnson AD, Reiner AP, Auer PL, Lettre G (2020) Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* 182(5):1198-1213. PMID: 32888493.

7. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, Chen MH, Raffield LM, Tardaguila M, Huffman JE, Ritchie SC, Megy K, Ponstingl H, Penkett CJ, Albers PK, Wigdor EM, Sakaue S, Moscati A, Manansala R, Lo KS, Qian H, Akiyama M, Bartz TM, Ben-Shlomo Y, Beswick A, Bork-Jensen J, Bottinger EP, Brody JA, van Rooij FJA, Chitrala KN, Wilson PWF, Choquet H, Danesh J, Di Angelantonio E, Dimou N, Ding J, Elliott P, Esko T, Evans MK, Felix SB, Floyd JS, Broer L, Grarup N, Guo MH, Guo Q, Greinacher A, Haessler J, Hansen T, Howson JMM, Huang W, Jorgenson E, Kacprowski T, Kähönen M, Kamatani Y, Kanai M, Karthikeyan S, Koskeridis F, Lange LA, Lehtimäki T, Linneberg A, Liu Y, Lyytikäinen LP, Manichaikul A, Matsuda K, Mohlke KL, Mononen N, Murakami Y, Nadkarni GN, Nikus K, Pankratz N, Pedersen O, Preuss M, Psaty BM, Raitakari OT, Rich SS, Rodriguez BAT, Rosen JD, Rotter JI, Schubert P, Spracklen CN, Surendran P, Tang H, Tardif JC, Ghanbari M, Völker U, Völzke H, Watkins NA, Weiss S; VA Million Veteran Program, Cai N, Kundu K, Watt SB, Walter K, Zonderman AB, Cho K, **Li Y**, Loos RJF, Knight JC, Georges M, Stegle O, Evangelou E, Okada Y, Roberts DJ, Inouye M, Johnson AD, Auer PL, Astle WJ, Reiner AP, Butterworth AS, Ouwehand WH, Lettre G, Sankaran VG, Soranzo N (2020) The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* 182(5):1214-1231. PMID: 32888494.

8. Matoba N, Liang D, Sun H, Aygün N, McAfee JC, Davis JE, Raffield LM, Qian H, Piven J, **Li Y**, Kosuri S, Won H, Stein JL (2020) Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Translational psychiatry* 10, 265. PMID: 32747698.

9. Colicino E, Marioni R, Ward-Caviness C, Gondalia R, Guan W, Chen B, Tsai PC, Huan T, Xu G, Golareh A, Schwartz J, Vokonas P, Just A, Starr JM, McRae AF, Wray NR, Visscher

PM, Bressler J, Zhang W, Tanaka T, Moore AZ, Pilling LC, Zhang G, Stewart JD, **Li Y**, Hou L, Castillo-Fernandez J, Spector T, Kiel DP, Murabito JM, Liu C, Mendelson M, Assimes T, Absher D, Tsaho PS, Lu AT, Ferrucci L, Wilson R, Waldenberger M, Prokisch H, Bandinelli S, Bell JT, Levy D, Deary IJ, Horvath S, Pankow J, Peters A, Whitsel EA, Baccarelli A (2020) Blood DNA methylation sites predict death risk in a longitudinal study of 12, 300 individuals. *Aging* 12(14): 14092-14124. PMID: 32697766.

10. Qian H, Kowalski MH, Kramer HJ, Tao R, Lash JP, Stilp AM, Cai J, **Li Y**, Franceschini N (2020) Genome-Wide Association of Kidney Traits in Hispanics/Latinos Using Dense Imputed Whole-Genome Sequencing Data: The Hispanic Community Health Study/Study of Latinos. Circ Genom Precis Med 13(4):e002891. PMID: 32600054.

11. Lu L, Liu X, Huang WK, Giusti-Rodríguez P, Cui J, Zhang S, Xu W, Wen Z, Ma S, Rosen JD, Xu Z, Bartels CF, Kawaguchi R, Hu M, Scacheri PC, Rong Z, **Li Y**, Sullivan PF, Song H, Ming GL, Li Y, Jin F (2020) Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. Mol Cell 79(3):521-534. PMID: 32592681.

12. Yang Y, Li Gang, Qian H, Wilhelmsen KC, Shen Y, **Li Y#** (2020) SMNN: batch effect correction for single-cell RNA-seq data via supervised mutual nearest neighbor detection. *Briefings in Bioinformatics,* doi.org/10.1093/bib/bbaa097. PMID: 32591778.

13. Yang X, Bergenholtz S, Maliskova L, Pebworth MP, Kriegstein AR, **Li Y**, Shen Y (2020) SMART-Q: An Integrative Pipeline Quantifying Cell Type-Specific RNA Transcription. *PLoS One* 15(4):e0228760. PMID: 32348304.

14. Yang Y, **Li Y**, Sancar A, Oztas O (2020) The circadian clock shapes the Arabidopsis transcriptome by regulating alternative splicing and alternative polyadenylation. *The Journal of Biological Chemistry*, doi:10.1074/jbc.RA120.013513. PMID: 32303634.

15. Halvorsen M*, Huh R*, Oskolkov N, Wen J*, Netotea S, Giusti-Rodriguez P, Karlsson R, Bryois J, Nystedt B, Ameur A, Kähler AK, Ancalade N, Farrell M, Crowley JJ, **Li Y**, Magnusson PKE, Gyllensten U, Hultman CM, Sullivan PF, Szatkiewicz JP (2020) Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. *Nat Commun.* 11(1):1842. PMID: 32296054. (*My team has contributed to statistical and multi-omics integrative analysis, particularly relating to the analysis of whole genome sequencing data and genome-wide chromatin spatial organization data*.)

16. Zhong W, Dong L, Poston TB, Darville T, Spracklen CN, Wu D, Mohlke KL, **Li Y**, Li Q, Zheng X (2020) Inferring Regulatory Networks From Mixed Observational Data Using Directed Acyclic Graphs. *Front Genet* 11, 8, doi:10.3389/fgene.2020.00008 (2020). PMID: 32127796.

17. Dong M, Thennavan A, Urrutia E, **Li Y**, Perou CM, Zou F, Jiang Y (2020) SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Brief Bioinform*, doi:10.1093/bib/bbz166 (2020). PMID: 31925417.

18. Gondalia R, Holliday KM, Baldassari A, Justice AE, Stewart JD, Liao D, Yanosky JD, Engel SM, Jordahl KM, Bhatti P, Horvath S, Assimes TL, Pankow JS, Demerath EW, Guan W, Fornage M, Bressler J, North KE, Conneely KN, **Li Y**, Hou L, Baccarelli AA, Whitsel EA (2020) Leukocyte Traits and Exposure to Ambient Particulate Matter Air Pollution in the Women's Health Initiative and Atherosclerosis Risk in Communities Study. *Environ Health Perspect.* 128(1):17004. PMID: 31903802.

19. Raffield LM, Iyengar AK, Wang B, Gaynor SM, Spracklen CN, Zhong X, Kowalski MH, Salimi S, Polfus LM, Benjamin EJ, Bis JC, Bowler R, Cade BE, Choi WJ, Comellas AP, Correa A, Cruz P, Doddapaneni H, Durda P, Gogarten SM, Jain D, Kim RW, Kral BG, Lange LA, Larson MG, Laurie C, Lee J, Lee S, Lewis JP, Metcalf GA, Mitchell BD, Momin Z, Muzny DM, Pankratz N, Park CJ, Rich SS, Rotter JI, Ryan K, Seo D, Tracy RP, Viaud-Martinez KA, Yanek LR, Zhao LP, Lin X, Li B, **Li Y**, Dupuis J, Reiner AP, Mohlke KL, Auer PL; TOPMed Inflammation Working Group; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium (2020) Allelic Heterogeneity at the CRP Locus Identified by Whole-Genome Sequencing in Multi-ancestry Cohorts. *Am J Hum Genet*. 106(1):112-120. PMID: 31883642.

20. Kowalski MH*, Qian H*, Hou Z, Rosen JD, Tapia AL, Shan Y, Jain D, Argos M, Arnett DK, Avery C, Barnes KC, Becker LC, Bien SA, Bis JC, Blangero J, Boerwinkle E, Bowden DW, Buyske S, Cai J, Cho MH, Choi SH, Choquet H, Cupples LA, Cushman M, Daya M, de Vries PS, Ellinor PT, Faraday N, Fornage M, Gabriel S, Ganesh SK, Graff M, Gupta N, He J, Heckbert SR, Hidalgo B, Hodonsky CJ, Irvin MR, Johnson AD, Jorgenson E, Kaplan R, Kardia SLR, Kelly TN, Kooperberg C, Lasky-Su JA, Loos RJF, Lubitz SA, Mathias RA, McHugh CP, Montgomery C, Moon JY, Morrison AC, Palmer ND, Pankratz N, Papanicolaou GJ, Peralta JM, Peyser PA, Rich SS, Rotter JI, Silverman EK, Smith JA, Smith NL, Taylor KD, Thornton TA, Tiwari HK, Tracy RP, Wang T, Weiss ST, Weng LC, Wiggins KL, Wilson JG, Yanek LR, Zöllner S, North KE, Auer PL; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium; TOPMed Hematology & Hemostasis Working Group, **Raffield LM**#, Reiner AP, **Li Y**# (2019) Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet*. 15(12):e1008500. PMID: 31869403.

21. Huh R, Yang Y, Jiang Y, Shen Y, **Li Y#** (2019) SAME-clustering: Single-cell Aggregated Clustering via Mixture Model Ensemble. *Nucleic Acids Res*. [Epub ahead of print] PMID: 31777938.

22. Gorkin DU, Qiu Y, Hu M, Fletez-Brant K, Liu T, Schmitt AD, Noor A, Chiou J, Gaulton KJ, Sebat J, **Li Y**, Hansen KD, Ren B (2019) Common DNA sequence variation influences 3-dimensional conformation of the human genome. *Genome Biol* 20(1):255. PMID: 31779666.

23. Zhao B, Zhang J, Ibrahim JG, Luo T, Santelli RC, **Li Y**, Li T, Shan Y, Zhu Z, Zhou F, Liao H, Nichols TE, Zhu H (2019) Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Mol Psychiatry.* [Epub ahead of print] PMID: 31666681.

24. Zhao B, Luo T, Li T, **Li Y**, Zhang J, Shan Y, Wang X, Yang L, Zhou F, Zhu Z; Alzheimer's Disease Neuroimaging Initiative; Pediatric Imaging, Neurocognition and Genetics, Zhu H (2019) Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet* 51(11):1637-1644. PMID: 31676860.

25. Johnson ND, Huang L, Li R, **Li Y**, Yang Y, Kim HR, Grant C, Wu H, Whitsel EA, Kiel DP, Baccarelli AA, Jin P, Murabito JM, Conneely KN (2019) Age-related DNA hydroxymethylation is enriched for gene expression and immune system processes in human peripheral blood. *Epigenetics*. PMID: 31506003.

26. Agha G, Mendelson MM, Ward-Caviness CK, Joehanes R, Huan T, Gondalia R, Salfati E, Brody JA, Fiorito G, Bressler J, Chen BH, Ligthart S, Guarrera S, Colicino E, Just AC, Wahl S, Gieger C, Vandiver AR, Tanaka T, Hernandez DG, Pilling LC, Singleton AB, Sacerdote C, Krogh V, Panico S, Tumino R, **Li Y**, Zhang G, Stewart JD, Floyd JS, Wiggins KL, Rotter JI, Multhaup M, Bakulski K, Horvath S, Tsao PS, Absher DM, Vokonas P, Hirschhorn J, Fallin MD, Liu C, Bandinelli S, Boerwinkle E, Dehghan A, Schwartz JD, Psaty BM, Feinberg AP, Hou L, Ferrucci L, Sotoodehnia N, Matullo G, Peters A, Fornage M, Assimes TL, Whitsel EA, Levy D, Baccarelli AA (2019) Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease. *Circulation* 140(8):645-657. PMID: 31424985. (*My team processed and analyzed DNA methylation data for the WHI cohort.*)

27. Song M, Yang X, Ren X, Maliskova L, Li B, Jones IR, Wang C, Jacob F, Wu K, Traglia M, Tam TW, Jamieson K, Lu SY, Ming GL, **Li Y**, Yao J, Weiss LA, Dixon JR, Judge LM, Conklin BR, Song H, Gan L, Shen Y (2019) Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* 51(8):1252-1262. PMID: 31367015.

28. Zhou X, Chen Y, Mok KY, Kwok TCY, Mok VCT, Guo Q, Ip FC, Chen Y, Mullapudi N; Alzheimer's Disease Neuroimaging Initiative, Giusti-Rodríguez P, Sullivan PF, Hardy J, Fu AKY, **Li Y**, Ip NY (2019) Non-coding variability at the APOE locus contributes to the Alzheimer's risk. *Nat Commun* 10(1):3310. PMID: 31346172.

29. Kim ES, Fong K, Lee L, Spiro A, Schwartz J, Whitsel E, Horvath S, Wang C, Hou L, Baccarelli AA, **Li Y**, Stewart J, Manson JE, Grodstein F, DeMeo DL, Kubzansky LD (2019) Optimism is not associated with two indicators of DNA methylation aging. *Aging* 11(14):4970-4989. PMID: 31322503.

30. Gondalia R, Baldassari A, Holliday KM, Justice AE, Méndez-Giráldez R, Stewart JD, Liao D, Yanosky JD, Brennan KJM, Engel SM, Jordahl KM, Kennedy E, Ward-Caviness CK, Wolf K, Waldenberger M, Cyrys J, Peters A, Bhatti P, Horvath S, Assimes TL, Pankow JS, Demerath EW, Guan W, Fornage M, Bressler J, North KE, Conneely KN, **Li Y**, Hou L, Baccarelli AA, Whitsel EA (2019) Methylome-wide association study provides evidence of particulate matter air pollution-associated DNA methylation. *Environ Int.* PMID: 31208937.

31. Zhong W, Spracklen CN, Mohlke KL, Zheng X#, Fine J#, **Li Y#** (2019) Multi-SNP mediation intersection-union test. *Bioinformatics* 35(22):4724-4729. PMID: 31099385.

32. Juric I, Yu M, Abnousi A, Raviram R, Fang R, Zhao Y, Zhang Y, Qiu Y, Yang Y, **Li Y**, Ren B, Hu M (2019) MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLoS Comput Biol* 15(4):e1006982. PMID: 30986246.

33. Fang Z, Weng C, Li H, Tao R, Mai W, Liu X, Lu L, Lai S, Duan Q, Alvarez C, Arvan P, Wynshaw-Boris A, **Li Y**, Pei Y, Jin F, Li Y (2019) Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key β-Cell-Specific Disease Genes. *Cell Rep* 26(11):3132-3144. PMID: 30865899.

34. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, **Li Y**, Stewart JD, Whitsel EA, Assimes TL, Ferrucci L, Horvath S (2019) DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* 11(2):303-327. PMID: 30669119.

35. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, Sunkin SM, Li Z, Shin Y, Zhu Y, Sousa AMM, Werling DM, Kitchen RR, Kang HJ, Pletikos M, Choi J, Muchnik S, Xu X, Wang D, Lorente-Galdos B, Liu S, Giusti-Rodríguez P, Won H, de Leeuw CA, Pardiñas AF; BrainSpan Consortium; PsychENCODE Consortium; PsychENCODE Developmental Subgroup, Hu M, Jin F, **Li Y**, Owen MJ, O'Donovan MC, Walters JTR, Posthuma D, Reimers MA, Levitt P, Weinberger DR, Hyde TM, Kleinman JE, Geschwind DH, Hawrylycz MJ, State MW, Sanders SJ, Sullivan PF, Gerstein MB, Lein ES, Knowles JA, Sestan N (2018) Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362(6420). PMID: 30545854. (*I contributed genome-wide chromatin interactome information, derived from adult and fetal brain Hi-C data to this study for integrative analysis.*)

36. Wang Y, Zhang B, Zhang L, An L, Xu J, Li D, Choudhary M, **Li Y**, Hu M, Hardison R, Wang T, Yue F (2018) The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biology* 19(1):151, PMID: 30286773.

37. Yang Y, Huh R, Culpepper HW, Lin Y, Love MI, **Li Y#** (2018) SAFE-clustering: Single-cell Aggregated (From Ensemble) Clustering for Single-cell RNA-seq Data. *Bioinformatics* 35(8):1269-1277. PMID: 30202935.

38. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, Felton S, Matsuyama M, Lowe D, Kabacik S, Wilson JG, Reiner AP, Maierhofer A, Flunkert J Aviv A, Hou L,

Baccarelli AA, **Li Y**, Stewart JD, Whitsel EA, Ferrucci L, Matsuyama S, Raj K (2018) Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging* 10(7):1758-1775. PMID: 30048243. PMCID: PMC6075434.

39. Zhao B, Ibrahim JG, **Li Y**, Li T, Wang Y, Shan Y, Zhu Z, Zhou F, Zhang J, Huang C, Liao H, Yang L, Thompson PM, Zhu H (2018) Heritability of Regional Brain Volumes in Large-Scale Neuroimaging and Genetic Studies. *Cereb Cortex* 29(7):2904-2914. PMID: 30010813.

40. Bi W, **Li Y**, Smeltzer MP, Gao G, Zhao S, Kang G (2018) STEPS: an efficient prospective likelihood approach to genetic association analyses of secondary traits in extreme phenotype sequencing. *Biostatistics* [Epub ahead of print]. PMID: 30007308.

41. Story Jovanova O, Nedeljkovic I, Spieler D, Walker RM, Liu C, Luciano M, Bressler J, Brody J, Drake AJ, Evans KL, Gondalia R, Kunze S, Kuhnel B, Lahti J, Lemaitre RN, Marioni RE, Swenson B, Himali JJ, Wu H, **Li Y**, McRae AF, Russ TC, Stewart J, Wang Z, Zhang G, Ladwig KH, Uitterlinden AG, Guo X, Peters A, Räikkönen K, Starr JM, Waldenberger M, Wray NR, Whitsel EA, Sotoodehnia N, Seshadri S, Porteous DJ, van Meurs J, Mosley TH, McIntosh AM, Mendelson MM, Levy D, Hou L, Eriksson JG, Fornage M, Deary IJ, Baccarelli A, Tiemeier H, Amin N (2018) DNA Methylation Signatures of Depressive Symptoms in Middle-aged and Elderly Persons: Meta-analysis of Multiethnic Epigenome-wide Studies. *JAMA Psychiatry* 75(9):949-959. PMID: 29998287. PMCID: PMC6142917.

42. Franceschini N, Kopp JB, Barac A, Martin LW, **Li Y**, Qian H, Reiner AP, Pollak M, Wallace RB, Rosamond WD, Winkler CA (2018) Association of APOL1 With Heart Failure With Preserved Ejection Fraction in Postmenopausal African American Women. *JAMA Cardiol.* 3(8):712-720. PMID: 29971324. PMCID: PMC6143074.

43. Lin Y, Liu L, Yang S, **Li Y**, Lin D, Zhang X, Yin X (2018) Genotype imputation for Han Chinese population using Haplotype Reference Consortium as reference. *Hum Genet.* 137(6-7):431-436. PMID: 29855708.

44. Hu Y, Raffield LM, Polfus LM, Moscati A, Nadkarni G, Preuss MH, Zhong X, Wei Q, Rich SS, **Li Y**, Wilson JG, Correa A, Loos RJF, Li B, Auer PL, Reiner AP; NHLBI Trans-Omics for Precision Medicine Consortium (2018) A common TCN1 loss-of-function variant is associated with lower vitamin B12 concentration in African Americans. *Blood* 131(25):2859-2863. PMID: 29764838. PMCID: PMC6014360.

45. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, Bacanu SA, Bækvad-Hansen M, Beekman AFT, Bigdeli TB, Binder EB, Blackwood DRH, Bryois J, Buttenschøn HN, Bybjerg-Grauholm J, Cai N, Castelao E, Christensen JH, Clarke TK, Coleman JIR, Colodro-Conde L, Couvy-Duchesne B, Craddock N, Crawford GE, Crowley CA, Dashti HS, Davies G, Deary IJ, Degenhardt F, Derks EM, Direk N, Dolan CV, Dunn EC, Eley TC, Eriksson N, Escott-Price V, Kiadeh FHF, Finucane HK, Forstner AJ, Frank J, Gaspar HA, Gill M, Giusti-Rodríguez P, Goes FS,

Gordon SD, Grove J, Hall LS, Hannon E, Hansen CS, Hansen TF, Herms S, Hickie IB, Hoffmann P, Homuth G, Horn C, Hottenga JJ, Hougaard DM, Hu M, Hyde CL, Ising M, Jansen R, Jin F, Jorgenson E, Knowles JA, Kohane IS, Kraft J, Kretzschmar WW, Krogh J, Kutalik Z, Lane JM, Li Y, **Li Y**, Lind PA, Liu X, Lu L, MacIntyre DJ, MacKinnon DF, Maier RM, Maier W, Marchini J, Mbarek H, McGrath P, McGuffin P, Medland SE, Mehta D, Middeldorp CM, Mihailov E, Milaneschi Y, Milani L, Mill J, Mondimore FM, Montgomery GW, Mostafavi S, Mullins N, Nauck M, Ng B, Nivard MG, Nyholt DR, O'Reilly PF, Oskarsson H, Owen MJ, Painter JN, Pedersen CB, Pedersen MG, Peterson RE, Pettersson E, Peyrot WJ, Pistis G, Posthuma D, Purcell SM, Quiroz JA, Qvist P, Rice JP, Riley BP, Rivera M, Saeed Mirza S, Saxena R, Schoevers R, Schulte EC, Shen L, Shi J, Shyn SI, Sigurdsson E, Sinnamon GBC, Smit JH, Smith DJ, Stefansson H, Steinberg S, Stockmeier CA, Streit F, Strohmaier J, Tansey KE, Teismann H, Teumer A, Thompson W, Thomson PA, Thorgeirsson TE, Tian C, Traylor M, Treutlein J, Trubetskoy V, Uitterlinden AG, Umbricht D, Van der Auwera S, van Hemert AM, Viktorin A, Visscher PM, Wang Y, Webb BT, Weinsheimer SM, Wellmann J, Willemsen G, Witt SH, Wu Y, Xi HS, Yang J, Zhang F; eQTLGen; 23andMe, Arolt V, Baune BT, Berger K, Boomsma DI, Cichon S, Dannlowski U, de Geus ECJ, DePaulo JR, Domenici E, Domschke K, Esko T, Grabe HJ, Hamilton SP, Hayward C, Heath AC, Hinds DA, Kendler KS, Kloiber S, Lewis G, Li QS, Lucae S, Madden PFA, Magnusson PK, Martin NG, McIntosh AM, Metspalu A, Mors O, Mortensen PB, Müller-Myhsok B, Nordentoft M, Nöthen MM, O'Donovan MC, Paciga SA, Pedersen NL, Penninx BWJH, Perlis RH, Porteous DJ, Potash JB, Preisig M, Rietschel M, Schaefer C, Schulze TG, Smoller JW, Stefansson K, Tiemeier H, Uher R, Völzke H, Weissman MM, Werge T, Winslow AR, Lewis CM, Levinson DF, Breen G, Børglum AD, Sullivan PF; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 50(5):668-681. PMID: 29700475. PMCID: PMC5934326. (*I contributed to functional annotations and interpretations of the association results.*)

46. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, **Li Y**, Whitsel EA, Wilson JG, Reiner AP, Aviv A, Lohman K, Liu Y, Ferrucci L, Horvath S (2018) An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 10(4):573-591. PMID: 29676998. PMCID: PMC5940111.

47. Napier MD, Franceschini N, Gondalia R, Stewart JD, Méndez-Giráldez R, Sitlani CM, Seyerle AA, Highland HM, **Li Y**, Wilhelmsen KC, Yan S, Duan Q, Roach J, Yao J, Guo X, Taylor KD, Heckbert SR, Rotter JI, North KE, Reiner AP, Zhang ZM, Tinker LF, Liao D, Laurie CC, Gogarten SM, Lin HJ, Brody JA, Bartz TM, Psaty BM, Sotoodehnia N, Soliman EZ, Avery CL, Whitsel EA (2018) Genome-wide association study and meta-analysis identify loci associated with ventricular and supraventricular ectopy. *Sci Rep*. 8(1):5675. PMID: 29618737. PMCID: PMC5884864.

48. Raffield LM, Ulirsch JC, Naik RP, Lessard S, Handsaker RE, Jain D, Kang HM, Pankratz N, Auer PL, Bao EL, Smith JD, Lange LA, Lange EM, **Li Y**, Thornton TA, Young BA, Abecasis GR, Laurie CC, Nickerson DA, McCarroll SA, Correa A, Wilson JG; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Hematology & Hemostasis, Diabetes, and

Structural Variation TOPMed Working Groups, Lettre G, Sankaran VG, Reiner AP (2018) Common α-globin variants modify hematologic and other clinical phenotypes in sickle cell trait and disease. *PLoS Genet*. e1007293 PMID: 29590102.

49. Zhou X, Chen Y, Mok KY, Zhao Q, Chen K, Chen Y, Hardy J, **Li Y**, Fu AKY, Guo Q, Ip NY; Alzheimer's Disease Neuroimaging Initiative (2018) Identification of genetic risk factors in the Chinese population implicates a role of immune system in Alzheimer's disease pathogenesis. *Proc Natl Acad Sci U S A*. 115(8):1697-1706. PMID: 29432188. PMCID: PMC5828602.

50. Raffield LM, Ellis J, Olson NC, Duan Q, Li J, Durda P, Pankratz N, Keating BJ, Wassel CL, Cushman M, Wilson JG, Gross MD, Tracy RP, Rich SS, Reiner AP, **Li Y**, Willis MS, Lange EM, Lange LA (2018) Genome-wide association study of homocysteine in African Americans from the Jackson Heart Study, the Multi-Ethnic Study of Atherosclerosis, and the Coronary Artery Risk in Young Adults study. *J Hum Genet*. 63(3):327-337. PMID: 29321517. PMCID: PMC5826839.

51. Lu AT, Xue L, Salfati EL, Chen BH, Ferrucci L, Levy D, Joehanes R, Murabito JM, Kiel DP, Tsai PC, Yet I, Bell JT, Mangino M, Tanaka T, McRae AF, Marioni RE, Visscher PM, Wray NR, Deary IJ, Levine ME, Quach A, Assimes T, Tsao PS, Absher D, Stewart JD, **Li Y**, Reiner AP, Hou L, Baccarelli AA, Whitsel EA, Aviv A, Cardona A, Day FR, Wareham NJ, Perry JRB, Ong KK, Raj K, Lunetta KL, Horvath S (2018) GWAS of epigenetic aging rates in blood reveals a critical role for TERT. *Nat Commun*. 9(1):387. PMID: 29374233. PMCID: PMC5786029.

52. Luo Y, Maity A, Wu MC, Smith C, Duan Q, **Li Y**, Tzeng JY (2018) On the substructure controls in rare variant analysis: Principal components or variance components? *Genet Epidemiol.* 42(3):276-287. PMID: 29280188. PMCID: PMC5851819.

53. Duan Q, Xu Z, Raffield LM, Chang S, Wu D, Lange EM, Reiner AP, **Li Y#** (2018) A robust and powerful two-step testing procedure for local ancestry adjusted allelic association analysis in admixed populations. *Genet Epidemiol.* 42(3):288-302. PMID: 29226381. PMCID: PMC5851818.

54. Seyerle AA, Sitlani CM, Noordam R, Gogarten SM, Li J, Li X, Evans DS, Sun F, Laaksonen MA, Isaacs A, Kristiansson K, Highland HM, Stewart JD, Harris TB, Trompet S, Bis JC, Peloso GM, Brody JA, Broer L, Busch EL, Duan Q, Stilp AM, O'Donnell CJ, Macfarlane PW, Floyd JS, Kors JA, Lin HJ, Li-Gao R, Sofer T, Méndez-Giráldez R, Cummings SR, Heckbert SR, Hofman A, Ford I, **Li Y**, Launer LJ, Porthan K, Newton-Cheh C, Napier MD, Kerr KF, Reiner AP, Rice KM, Roach J, Buckley BM, Soliman EZ, de Mutsert R, Sotoodehnia N, Uitterlinden AG, North KE, Lee CR, Gudnason V, Stürmer T, Rosendaal FR, Taylor KD, Wiggins KL, Wilson JG, Chen YD, Kaplan RC, Wilhelmsen K, Cupples LA, Salomaa V, van Duijn C, Jukema JW, Liu Y, Mook-Kanamori DO, Lange LA, Vasan RS, Smith AV, Stricker BH, Laurie CC, Rotter JI, Whitsel EA, Psaty BM, Avery CL (2018) Pharmacogenomics study of thiazide diuretics and QT interval in multi-ethnic populations:

the cohorts for heart and aging research in genomic epidemiology. *Pharmacogenomics J.* 18(2):215-226. PMID: 28719597. PMCID: PMC5773415.

55. Floyd JS, Sitlani CM, Avery CL, Noordam R, Li X, Smith AV, Gogarten SM, Li J, Broer L, Evans DS, Trompet S, Brody JA, Stewart JD, Eicher JD, Seyerle AA, Roach J, Lange LA, Lin HJ, Kors JA, Harris TB, Li-Gao R, Sattar N, Cummings SR, Wiggins KL, Napier MD, Stürmer T, Bis JC, Kerr KF, Uitterlinden AG, Taylor KD, Stott DJ, de Mutsert R, Launer LJ, Busch EL, Méndez-Giráldez R, Sotoodehnia N, Soliman EZ, **Li Y**, Duan Q, Rosendaal FR, Slagboom PE, Wilhelmsen KC, Reiner AP, Chen YD, Heckbert SR, Kaplan RC, Rice KM, Jukema JW, Johnson AD, Liu Y, Mook-Kanamori DO, Gudnason V, Wilson JG, Rotter JI, Laurie CC, Psaty BM, Whitsel EA, Cupples LA, Stricker BH (2018) Large-scale pharmacogenomic study of sulfonylureas and the QT, JT and QRS intervals: CHARGE Pharmacogenomics Working Group. *Pharmacogenomics J.* 18(1):127-135. PMID: 27958378. PMCID: PMC5468495.

56. Richard MA, Huan T, Ligthart S, Gondalia R, Jhun MA, Brody JA, Irvin MR, Marioni R, Shen J, Tsai PC, Montasser ME, Jia Y, Syme C, Salfati EL, Boerwinkle E, Guan W, Mosley TH Jr, Bressler J, Morrison AC, Liu C, Mendelson MM, Uitterlinden AG, van Meurs JB; BIOS Consortium, Franco OH, Zhang G, **Li Y**, Stewart JD, Bis JC, Psaty BM, Chen YI, Kardia SLR, Zhao W, Turner ST, Absher D, Aslibekyan S, Starr JM, McRae AF, Hou L, Just AC, Schwartz JD, Vokonas PS, Menni C, Spector TD, Shuldiner A, Damcott CM, Rotter JI, Palmas W, Liu Y, Paus T, Horvath S, O'Connell JR, Guo X, Pausova Z, Assimes TL, Sotoodehnia N, Smith JA, Arnett DK, Deary IJ, Baccarelli AA, Bell JT, Whitsel E, Dehghan A, Levy D, Fornage M (2017) DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. *Am J Hum Genet*. 101(6):888-902. PMID: 29198723. PMCID: PMC5812919.

57. Martin JS, Xu Z, Reiner AP, Mohlke KL, Sullivan P, Ren B, Hu M, **Li Y#**. HUGIn: Hi-C Unifying Genomic Interrogator (2017) *Bioinformatics* 33(23):3793-5. PMID: 28582503. PMCID: PMC5860315.

58. Grant CD, Jafari N, Hou L, **Li Y**, Stewart JD, Zhang G, Lamichhane A, Manson JE, Baccarelli AA, Whitsel EA, Conneely KN (2017) A longitudinal study of DNA methylation as a potential mediator of age-related diabetes risk. *Geroscience* 39(5-6):475-489. PMID: 29159506. PMCID: PMC5745220.

59. Raffield LM, Zakai NA, Duan Q, Laurie C, Smith JD, Irvin MR, Doyle MF, Naik RP, Song C, Manichaikul AW, Liu Y, Durda P, Rotter JI, Jenny NS, Rich SS, Wilson JG, Johnson AD, Correa A, **Li Y**, Nickerson DA, Rice K, Lange EM, Cushman M, Lange LA, Reiner AP; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Hematology & Hemostasis TOPMed Working Group (2017) D-Dimer in African Americans: Whole Genome Sequence Analysis and Relationship to Cardiovascular Disease Risk in the Jackson Heart Study. *Arterioscler Thromb Vasc Biol*. 37(11):2220-2227. PMID: 28912365. PMCID: PMC5658238.

60. Kerr KF, Avery CL, Lin HJ, Raffield LM, Zhang QS, Browning BL, Browning SR, Conomos MP, Gogarten SM, Laurie CC, Sofer T, Thornton TA, Hohensee C, Jackson RD, Kooperberg C, **Li Y**, Méndez-Giráldez R, Perez MV, Peters U, Reiner AP, Zhang ZM, Yao J, Sotoodehnia N, Taylor KD, Guo X, Lange LA, Soliman EZ, Wilson JG, Rotter JI, Heckbert SR, Jain D, Whitsel EA (2017) Genome-wide association study of heart rate and its variability in Hispanic/Latino cohorts. *Heart rhythm* 14(11): 1675-84. PMID: 28610988. PMCID: PMC5671896.

61. Graff M, Emery LS, Justice AE, Parra E, Below JE, Palmer ND, Gao C, Duan Q, Valladares-Salgado A, Cruz M, Morrison AC, Boerwinkle E, Whitsel EA, Kooperberg C, Reiner A, **Li Y**, Rodriguez CJ, Talavera GA, Langefeld CD, Wagenknecht LE, Norris JM, Taylor KD, Papanicolaou G, Kenny E, Loos RJF, Chen YI, Laurie C, Sofer T, North KE (2017) Genetic architecture of lipid traits in the Hispanic community health study/study of Latinos. *Lipids Health Dis*. 16(1):200. PMID: 29025430. PMCID: PMC5639746.

62. Du Y, Martin JS, McGee J, Yang Y, Liu EY, Sun Y, Geihs M, Kong X, Zhou EL, **Li Y#**, Huang J (2017) A SNP panel and online tool for checking genotype concordance through comparing QR codes. *PLoS One* 12(9):e0182438. PMID: 28926565. PMCID: PMC5604942.

63. Hui D, Fang Z, Lin J, Duan Q, **Li Y**, Hu M, Chen W (2017) LAIT: a local ancestry inference toolkit. *BMC Genet* 18(1):83. PMID: 28877673. PMCID: PMC5585928.

64. Christophersen IE, Magnani JW, Yin X, Barnard J, Weng LC, Arking DE, Niemeijer MN, Lubitz SA, Avery CL, Duan Q, Felix SB, Bis JC, Kerr KF, Isaacs A, Müller-Nurasyid M, Müller C, North KE, Reiner AP, Tinker LF, Kors JA, Teumer A, Petersmann A, Sinner MF, Buzkova P, Smith JD, Van Wagoner DR, Völker U, Waldenberger M, Peters A, Meitinger T, Limacher MC, Wilhelmsen KC, Psaty BM, Hofman A, Uitterlinden A, Krijthe BP, Zhang ZM, Schnabel RB, Kääb S, van Duijn C, Rotter JI, Sotoodehnia N, Dörr M, **Li Y**, Chung MK, Soliman EZ, Alonso A, Whitsel EA, Stricker BH, Benjamin EJ, Heckbert SR, Ellinor PT (2017) Fifteen Genetic Loci Associated With the Electrocardiographic P Wave. *Circ Cardiovasc Genet* 10(4) PMID: 28794112.

65. Cannon ME, Duan Q, Wu Y, Zeynalzadeh M, Xu Z, Kangas AJ, Soininen P, Ala-Korpela M, Civelek M, Lusis AJ, Kuusisto J, Collins FS, Boehnke M, Tang H, Laakso M, **Li Y**, Mohlke KL (2017) Trans-ancestry Fine Mapping and Molecular Assays Identify Regulatory Variants at the *ANGPTL8* HDL-C GWAS Locus. *G3* 7(9):3217-3227. PMID: 28754724. PMCID: PMC5592946.

66. Nolte IM, Munoz ML, Tragante V, Amare AT, Jansen R, Vaez A, von der Heyde B, Avery CL, Bis JC, Dierckx B, van Dongen J, Gogarten SM, Goyette P, Hernesniemi J, Huikari V, Hwang SJ, Jaju D, Kerr KF, Kluttig A, Krijthe BP, Kumar J, van der Laan SW, Lyytikäinen LP, Maihofer AX, Minassian A, van der Most PJ, Müller-Nurasyid M, Nivard M, Salvi E, Stewart JD, Thayer JF, Verweij N, Wong A, Zabaneh D, Zafarmand MH, Abdellaoui A, Albarwani S, Albert C, Alonso A, Ashar F, Auvinen J, Axelsson T, Baker DG, de Bakker PIW, Barcella M, Bayoumi R, Bieringa RJ, Boomsma D, Boucher G, Britton AR,

Christophersen I, Dietrich A, Ehret GB, Ellinor PT, Eskola M, Felix JF, Floras JS, Franco OH, Friberg P, Gademan MGJ, Geyer MA, Giedraitis V, Hartman CA, Hemerich D, Hofman A, Hottenga JJ, Huikuri H, Hutri-Kähönen N, Jouven X, Junttila J, Juonala M, Kiviniemi AM, Kors JA, Kumari M, Kuznetsova T, Laurie CC, Lefrandt JD, Li Y, **Li Y**, Liao D, Limacher MC, Lin HJ, Lindgren CM, Lubitz SA, Mahajan A, McKnight B, Zu Schwabedissen HM, Milaneschi Y, Mononen N, Morris AP, Nalls MA, Navis G, Neijts M, Nikus K, North KE, O'Connor DT, Ormel J, Perz S, Peters A, Psaty BM, Raitakari OT, Risbrough VB, Sinner MF, Siscovick D, Smit JH, Smith NL, Soliman EZ, Sotoodehnia N, Staessen JA, Stein PK, Stilp AM, Stolarz-Skrzypek K, Strauch K, Sundström J, Swenne CA, Syvänen AC, Tardif JC, Taylor KD, Teumer A, Thornton TA, Tinker LE, Uitterlinden AG, van Setten J, Voss A, Waldenberger M, Wilhelmsen KC, Willemsen G, Wong Q, Zhang ZM, Zonderman AB, Cusi D, Evans MK, Greiser HK, van der Harst P, Hassan M, Ingelsson E, Järvelin MR, Kääb S, Kähönen M, Kivimaki M, Kooperberg C, Kuh D, Lehtimäki T, Lind L, Nievergelt CM, O'Donnell CJ, Oldehinkel AJ, Penninx B, Reiner AP, Riese H, van Roon AM, Rioux JD, Rotter JI, Sofer T, Stricker BH, Tiemeier H, Vrijkotte TGM, Asselbergs FW, Brundel BJJM, Heckbert SR, Whitsel EA, den Hoed M, Snieder H, de Geus EJC (2017) Genetic loci associated with heart rate variability and their effects on cardiac disease risk. *Nat Commun*. 8:15805. PMID: 28613276. PMCID: PMC5474732. (*I contributed to the analysis and interpretation of results.*)

67. Gondalia R, Avery CL, Napier MD, Mendez-Giraldez R, Stewart JD, Sitlani CM, **Li Y**, Wilhelmsen KC, Duan Q, Roach J, North KE, Reiner AP, Zhang ZM, Tinker LF, Yanosky JD, Liao D, Whitsel EA (2017) Genome-wide Association Study of Susceptibility to Particulate Matter-Associated QT Prolongation. *Environ Health Perspect* 125(6):067002. PMID: 28749367.

68. Chang J, Tan W, Ling Z, Xi R, Shao M, Chen M, Luo Y, Zhao Y, Liu Y, Huang X, Xia Y, Hu J, Parker JS, Marron D, Cui Q, Peng L, Chu J, Li H, Du Z, Han Y, Tan W, Liu Z, Zhan Q, **Li Y**, Mao W, Wu C, Lin D (2017) Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. *Nature communications* 8:15290. PMID: 28548104. PMCID: PMC5477513.

69. Noordam R, Sitlani CM, Avery CL, Stewart JD, Gogarten SM, Wiggins KL, Trompet S, Warren HR, Sun F, Evans DS, Li X, Li J, Smith AV, Bis JC, Brody JA, Busch EL, Caulfield MJ, Chen YI, Cummings SR, Cupples LA, Duan Q, Franco OH, Méndez-Giráldez R, Harris TB, Heckbert SR, van Heemst D, Hofman A, Floyd JS, Kors JA, Launer LJ, **Li Y**, Li-Gao R, Lange LA, Lin HJ, de Mutsert R, Napier MD, Newton-Cheh C, Poulter N, Reiner AP, Rice KM, Roach J, Rodriguez CJ, Rosendaal FR, Sattar N, Sever P, Seyerle AA, Slagboom PE, Soliman EZ, Sotoodehnia N, Stott DJ, Stürmer T, Taylor KD, Thornton TA, Uitterlinden AG, Wilhelmsen KC, Wilson JG, Gudnason V, Jukema JW, Laurie CC, Liu Y, Mook-Kanamori DO, Munroe PB, Rotter JI, Vasan RS, Psaty BM, Stricker BH, Whitsel EA (2017) A genome-wide interaction analysis of tricyclic/tetracyclic antidepressants and RR and QT intervals: a pharmacogenomics study from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. *J Med Genet*. 54(5):313-323. PMID: 28039329.

70. Hodonsky CJ, Jain D, Schick UM, Morrison JV, Brown L, McHugh CP, Schurmann C, Chen DD, Liu YM, Auer PL, Laurie CA, Taylor KD, Browning BL, **Li Y**, Papanicolaou G, Rotter JI, Kurita R, Nakamura Y, Browning SR, Loos RJF, North KE, Laurie CC, Thornton TA, Pankratz N, Bauer DE, Sofer T, Reiner AP (2017) Genome-wide association study of red blood cell traits in Hispanics/Latinos: The Hispanic Community Health Study/Study of Latinos. *PLoS Genet*. 13(4):e1006760. PMID: 28453575. PMCID: PMC5428979.

71. Ng MCY, Graff M, Lu Y, Justice AE, Mudgal P, Liu CT, Young K, Yanek LR, Feitosa MF, Wojczynski MK, Rand K, Brody JA, Cade BE, Dimitrov L, Duan Q, Guo X, Lange LA, Nalls MA, Okut H, Tajuddin SM, Tayo BO, Vedantam S, Bradfield JP, Chen G, Chen WM, Chesi A, Irvin MR, Padhukasahasram B, Smith JA, Zheng W, Allison MA, Ambrosone CB, Bandera EV, Bartz TM, Berndt SI, Bernstein L, Blot WJ, Bottinger EP, Carpten J, Chanock SJ, Chen YI, Conti DV, Cooper RS, Fornage M, Freedman BI, Garcia M, Goodman PJ, Hsu YH, Hu J, Huff CD, Ingles SA, John EM, Kittles R, Klein E, Li J, McKnight B, Nayak U, Nemesure B, Ogunniyi A, Olshan A, Press MF, Rohde R, Rybicki BA, Salako B, Sanderson M, Shao Y, Siscovick DS, Stanford JL, Stevens VL, Stram A, Strom SS, Vaidya D, Witte JS, Yao J, Zhu X, Ziegler RG, Zonderman AB, Adeyemo A, Ambs S, Cushman M, Faul JD, Hakonarson H, Levin AM, Nathanson KL, Ware EB, Weir DR, Zhao W, Zhi D; Bone Mineral Density in Childhood Study (BMDCS) Group, Arnett DK, Grant SFA, Kardia SLR, Oloapde OI, Rao DC, Rotimi CN, Sale MM, Williams LK, Zemel BS, Becker DM, Borecki IB, Evans MK, Harris TB, Hirschhorn JN, **Li Y**, Patel SR, Psaty BM, Rotter JI, Wilson JG, Bowden DW, Cupples LA, Haiman CA, Loos RJF, North KE (2017) Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet*. 13(4):e1006719. PMID: 28430825.

72. Yang L, Chang S, Lu Q, Zhang Y, Wu Z, Sun X, Cao Q, Qian Y, Jia T, Xu B, Duan Q, **Li Y**, Zhang K, Schumann G, Liu D, Wang J, Wang Y, Lu L (2018) A new locus regulating MICALL2 expression was identified for association with executive inhibition in children with attention deficit hyperactivity disorder. *Mol Psychiatry*. 23(4):1014-1020. PMID: 28416812.

73. Quach A, Levine ME, Tanaka T, Lu AT, Chen BH, Ferrucci L, Ritz B, Bandinelli S, Neuhouser ML, Beasley JM, Snetselaar L, Wallace RB, Tsao PS, Absher D, Assimes TL, Stewart JD, **Li Y**, Hou L, Baccarelli AA, Whitsel EA, Horvath S (2017) Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging* 9(2):419-46. PMID: 28198702.

74. Sun H, Wang Y, Chen Y, **Li Y**, Wang S (2017) pETM: a penalized exponential tilt model for analysis of correlated high-dimensional DNA methylation data. *Bioinformatics* 33(12):1765-1772. PMID: 28165116.

75. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, **Li Y**, Lin S, Lin Y, Barr CL, Ren B (2016) A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human. *Cell Reports* 17(8): 2042-2059. PMID: 27851967. PMCID: PMC5478386 (highlighted by *Nature Reviews Genetics* http://www.nature.com/nrg/journal/v18/n1/pdf/nrg.2016.161.pdf)

76. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai PC, Reynolds LM, Just AC, Duan Q, Boer CG, Tanaka T, Elks CE, Aslibekyan S, Brody JA, Kühnel B, Herder C, Almli LM, Zhi D, Wang Y, Huan T, Yao C, Mendelson MM, Joehanes R, Liang L, Love SA, Guan W, Shah S, McRae AF, Kretschmer A, Prokisch H, Strauch K, Peters A, Visscher PM, Wray NR, Guo X, Wiggins KL, Smith AK, Binder EB, Ressler KJ, Irvin MR, Absher DM, Hernandez D, Ferrucci L, Bandinelli S, Lohman K, Ding J, Trevisi L, Gustafsson S, Sandling JH, Stolk L, Uitterlinden AG, Yet I, Castillo-Fernandez JE, Spector TD, Schwartz JD, Vokonas P, Lind L, **Li Y**, Fornage M, Arnett DK, Wareham NJ, Sotoodehnia N, Ong KK, van Meurs JB, Conneely KN, Baccarelli AA, Deary IJ, Bell JT, North KE, Liu Y, Waldenberger M, London SJ, Ingelsson E, Levy D (2018) A DNA methylation biomarker of alcohol consumption. *Mol. Psychiatry* 23(2):422-433. PMID: 27843151. PMCID: PMC5575985.

77. Xu Z, Zhang G, Wu C, **Li Y#**, Hu M (2016) FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C data. *Bioinformatics* 32(17):2692-2695. PMID: 27153668. PMCID: PMC5013904.

78. 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.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. 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.

86. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526(7571):68-74. PMID: 26432245. PMCID: PMC4750478. (*I contributed to variant and genotyping calling from low-coverage sequencing data.*)

87. 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.

88. 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.

89. 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.

90. 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.

91. 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.

92. 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.

93. 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.

94. 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.

95. 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.

96. 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.

97. 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.

98. 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.

99. 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.

100. 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.

101. 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.

102. 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.

103. 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.

104. 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.

105. 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.

106. 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.

107. 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.*)

108. 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.*)

109. 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.

110. 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.

111. 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.

112. 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*.)

113. 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.

114. 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.

115. 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.

116. 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.

117. 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.

118. 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.

119. 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.

120. 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.

121. 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.*)

122. 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.*)

123. 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)

124. 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.

125. 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.

126. 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.

127. 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.

128. 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.

129. **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)

130. 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.

131. 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.*)

132. 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*.)

133. 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.

134. 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.

135. 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*.)

136. 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.

137. 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 (2012) Population-specific coding variant underlies genome-wide association with adiponectin level. *Human Molecular Genetics* 21(2): 463-71. PMID: 22010046. PMCID: PMC3276282.

138. 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*.)

139. 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.

140. Marth GT et al (2011) The functional spectrum of low-frequency coding variation. *Genome Biology* 12(9):R84. PMID: 21917140. PMCID: PMC3308047.

141. 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.

142. Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Bustamante CD (2011) Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences of the United States of America* 108(29):11983-8. PMID: 21730125 PMCID: PMC3142009. (*contribution: quality control and basic population genetics analysis*)

143. 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.

144. Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, Zilversmit M, Cartwright R, Rouleau GA, Daly M, Stone EA, Hurles ME, Awadalla P (2011) Variation in genome-wide mutation rates within and between human families. *Nature Genetics* 43(7):712-4. PMID: 21666693 PMCID: PMC3322360 (*Contribution: Quality control*, *Variant Calling, Genotype Calling, Mendelian Inconsistency and Haplotype Analysis*).

145. **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.

146. 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 (2011) 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.

147. 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.

148. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Sella G, Przeworski M (2011) Classic selective sweeps were rare in recent human evolution. *Science* 331(6019):920-4. PMID: 21330547 PMCID: PMC3669691. (*Contribution: Quality Control, Data Analysis in the 1000 Genomes Project*).

149. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, **Li Y**, Lin CY, Luo R, Mu XJ, Nemesh J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stütz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Ye K, Eichler EE, Gerstein MB, Hurles ME, Lee C, McCarroll SA, Korbel JO (2011) Mapping copy number variation by population-scale genome sequencing. *Nature* 470(7332):59-65. PMID: 21293372 PMCID: PMC3077050. (*Contribution: Quality Control, Variant Calling in the 1000 Genomes Project*).

150. 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.

151. 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.

152. **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.

153. **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.

154. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE (2010) Diversity of human copy number variation and multicopy genes. *Science* 330(6004):641-6. PMID: 21030649 PMCID: PMC3020103. (Contribution: Quality Control and Variant Calling within the 1000 Genomes Project.)

155. 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*.)

156. 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.

157. 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.

158. 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.

159. 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.

160. 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.

161. 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.

162. 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.

163. 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.

164. 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*.)

165. 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.

166. 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.

167. 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.

168. 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.

169. 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.

170. 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.

171. 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.

172. 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*.)

173. 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.*)

174. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, **Li Y**, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ,

Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449(7164): 913-8. PMID: 17943131. PMCID: PMC2687721.

175. 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*.)

176. 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*.)

177. 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*.)

Manuscripts Submitted/In Press

Omitted. Please visit https://yunliweb.its.unc.edu/publications.html for preprints.

Review articles:

178. **Li Y#**, Hu M, Shen Y (2018) Gene Regulation in the 3D Genome. *Hum Mol Genet*. 27(R2):R228-R233. PMID: 29767704. PMCID: PMC6061806.
179. **Li Y**, Willer CJ, Sanna S, Abecasis GR (2009) Genotype imputation. *Annual Review Genomics and Human Genetics* 10: 387-406. PMID: 19715440. PMCID: PMC2925172.

Invited oral presentations:

2020	2020 ICSA Applied Statistics Symposium, Virtual
2020	Colorado Center for Personalized Medicine, University of Colorado-Denver, Aurora, CO
2020	MidAtlantic Bioinformatics Conference, Virtual

2020 Department of Computational Medicine and Bioinformatics, University of Michigan, Virtual
2019 The 11th ICSA International Conference, Hanzhou, Zhejiang Province, China
2019 Biomedical Informatics, Ohio State University, Columbus, OH
2019 The 17th SCBA International Symposium, Kunming, Yunan Province, China
2019 ENAR 2019 Spring Meeting, Philadelphia, PA
2019 Duke Center for Statistical Genetics and Genomics, Inaugural Talk, Durham, NC
2019 BIRS Workshop, "Frontiers in Single-cell Technology, Applications and Data Analysis", Banff, Alberta, Canada
2018 Joint Statistical Meeting, Vancouver, British Columbia, Canada
2018 Workshop on high dimensional statistics: theory and applications, Changchun, Jilin Province, China
2018 Department of Statistics, University of California, Riverside, CA
2018 Joint GSP-TOPMed Analysis Workshop, Nashville, TN
2017 IBW2017, the Thirteenth International Bioinformatics Workshop, Haerbin, Heilongjiang Province, China
2017 2017 IMS-China International Conference on Statistics and Probability, Nanning, Guangxi Province, China
2017 Department of Public Health Sciences, University of Chicago, Chicago, IL
2017 Department of Preventative Medicine, University of Southern California, Los Angeles, CA
2017 Department of Biostatistics, University of Michigan, Ann Arbor, MI
2017 Department of Pediatrics, Baylor College of Medicine, Houston, TX
2016 Cancer Institute and Hospital, Chinese Academy of Medical Sciences, Beijing, China
2016 Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN
2016 School of Public Health, Shanghai Jiaotong University, Shanghai, China
2016 2016 Joint Statistical Meeting in Chicago, IL
2016 The 4th IBS-China International Biostatistical Conference, Shanghai, China
2016 Center for Statistical Genetics, University of Michigan, Ann Arbor, MI
2016 Division of Biostatistics, Department of Population Health, New York University School of Medicine, New York, NY
2015 Department of Biostatistics, University of Minnesota, Minneapolis, MN
2015 Department of Mathematics, Bowling Green State University, Bowling Green, OH
2015 Institute for Behavioral Genetics, University of Colorado, Boulder, CO
2015 Institute for Personalized Medicine (IPM), Penn State University, Hershey, PA
2015 ENAR, the International Biometric Society, Miami, FL
2014 Department of Biomathematics, University of California, Los Angeles, CA
2014 Department of Biostatistics, Harvard University, Boston, MA
2014 Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA
2014 Nanjing Medical University, Nanjing, Jiangsu Province, China
2014 Centre for Genomic Sciences, the University of Hong Kong, Hong Kong, China
2014 The Hong Kong University of Science and Technology, Hong Kong, China
2014 Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA
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

CLASSROOM TEACHING

Course/Module Director:

BCB723: Topics in Statistical Genetics and Genomics, 7.5 lecture hours
Spring 2020, 10 students and 1 auditor
Spring 2019, 9 students and 1 auditor

Spring 2018, 8 students
Bios782/BCB725: Statistical Methods in Genetic Studies,
Fall 2020, 19 students and 1 autitor, 22.5 lecture hours
BIOS782: Statistical Methods in Genetic Studies,
Fall 2018, 12 students and 2 auditors, 21 lecture hours
Fall 2016, 21 students, 22.5 lecture hours
BCB720: Introduction to Statistical Modeling, 12 lecture hours
Fall 2016, 24 students
EPI889: Advanced Genetic Epidemiology – High Throughput Data Analysis
12 lecture hours
Fall 2016, 12 students
Fall 2014, 5 students
Fall 2012, 18 students
BIOS735: Statistical Computing, 9 hours for lecturing/leading discussions
Fall 2015, 22 students
Fall 2013, 23 students
BCB725/BIOS681: Introduction to Statistical Genetics,
3 credit hours (24 hours for lecturing/leading discussions),
Spring 2015, 9 students
Spring 2012, 17 students

Guest Lectures at UNC:

BIOS/BCB785: Statistical Methods for Gene Expression Analysis,
1.5 lecture hours, Spring 2020
1.5 lecture hours, Spring 2019
BCB710: Bioinformatics Colloquium,
1.5 lecture hours, 2009, 2010, 2013, 2015, 2019, 2020
EPID743: Genetic Epidemiology, 1.5 contact hours, 2010-2020
BIOS781: Statistical Methods in Human Genetics, 4.5 contact hours, 2018
BCB720: Introduction to Statistical Modeling,
1.5 lecture hours, Fall 2015, 2017, 2018
BCB720: Introduction to Statistical Modeling,
12 lecture hours, Fall 2016, 24 students
BIOS784: Computational Biology, 4.5 contact hours, 2013, 2015
BIOS740: Statistical Methods for Genetic Association Studies,
4.5 lecture hours, 2012, 2014

Lectures outside UNC:

- ✧ 2019 BIRS Workshop on Single-cell Technology, Applications and Data Analysis
- ✧ 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

MENTORING

Undergraduate Student Advisor:

2020- Alice Y. Sun, B.S. student, Department of Biostatistics, Department of Biology
2019-2020 Daniel Malawsky, B.S. student, Department of Biostatistics
2018-2019 Wenwen Mei, B.S. student, Department of Biostatistics
2016-2017 Eric Zhou, B.S. student, Department of Computer Science

Graduate Student Thesis Advisor:

2020- Chanhwa Lee, Ph.D. student, Department of Biostatistics
2020- Jiawen Chen, Ph.D. student, Department of Biostatistics
2020- Le Huang, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2019- Quan Sun, Ph.D. student, Department of Biostatistics
2019- Weifang Liu, Ph.D. student, Department of Biostatistics
2018- Minzhi Jiang, Ph.D. student, Department of Applied Physical Sciences
2018- Bryce Rowland, Ph.D. student, Department of Biostatistics (NSF Graduate Research Fellowship recipient 2019-2022)
2017- Gang Li, Ph.D. student, Department of Statistics and Operation Research
2017- Amanda Tapia, Dr.PH student, Department of Biostatistics
2016- Taylor Lagler, Ph.D. student, Department of Biostatistics (NSF Graduate Research Fellowship recipient 2017-2020)
2016- Jon Rosen, Ph.D. student, Department of Biostatistics
2016- Yue Shan, Ph.D. student, Department of Biostatistics
2018-2020 Cheynna Crowley, Dr.PH student, Department of Biostatistics
2018-2020 Ai Ye, Ph.D. student, Department of Psychology and Neuroscience
2016-2020 Eric Van Buren, Ph.D. student, Department of Biostatistics
2017-2019 Huijun Qian, Ph.D. student, Department of Statistics and Operation Research
2017-2019 Madeline Kowalski, M.S. student, Department of Biostatistics
2016-2019 Wujuan Zhong, Ph.D. student, Department of Biostatistics
2016-2019 Ruth Huh, Ph.D. student, Department of Biostatistics
2016-2018 Cheynna Crowley, M.S. student, Department of Biostatistics
2016-2017 Evan Kwiatkowski, Ph.D. student, Department of Biostatistics
2015-2017 Yimeng Tiaoyao, M.S. student, Department of Statistics and Operation Research
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 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:

2019 Fall Yuriko Harigaya
2016 Winter Angel Wei Huang
2015 Winter Dan Liang
2014 Spring Shengjie Chai
2013 Spring Greg Keele
2012 Spring Guosheng Zhang
2011 Fall Qing Duan

Postdoctoral Scientist Advisor:

2018- Jia Wen, Ph.D. in Statistical Genomics
2017-2020 Yuchen Yang, Ph.D. in Biochemistry and Molecular Biology (promoted to faculty)
2016-2020 Laura Raffield, Ph.D. in Molecular Genetics and Genomics (promoted to faculty)
2018-2020 Munan Xie, Ph.D. in Botany
2016-2017 Josh Martin, Ph.D. in Theoretical Physics
2016-2017 Qing Duan, Ph.D. in Bioinformatics and Computational Biology.
2012-2016 Zheng Xu, Ph.D. in Statistics. Current position: Assistant professor, University of Nebreska, Lincoln
2012-2015 Song Yan, Ph.D. in Statistics. Current position: data scientist at Microsoft Bing Ads.

Visiting Scholar:

2013-2014 Wei Chen
2014-2015 Suhua Chang

Programmer:

2012-2013 Yunfei Wang

Statistician:

2013-2014 Yurong Lu

Temporary Data Scientist:

2019 Huijun Qian, Ph.D. in Statistics
2019 Madeline Kowalski, M.S. in Biostatistics
2018-2019 Ye Su, M.S.in Bioinformatics

2018 Tong Shan, M.S. in Statistics

Graduate Student Thesis Advisory Committees:

2020- Kevin Donovan, Ph.D. student, Department of Biostatistics
2020- Nil Aygun, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2019- Sarah Brotman, Ph.D. student, Curriculum in Genetics and Molecular Biology
2018- Mike Lafferty, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2018- Darius Bost, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2018-2020 Dayne Filer, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2018-2020 Scott Van Buren, Ph.D. student, Department of Biostatistics
2018-2020 Arjun Bhattacharya, Ph.D. student, Department of Biostatistics
2018-2020 Sarah Reifeis, Ph.D. student, Department of Biostatistics
2018-2020 Pedro Baldoni, Ph.D. student, Department of Biostatistics
2017-2020 Yanwei Cai, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2017-2020 Kevin Currin, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2018-2019 Yue Jiang, Ph.D. student, Department of Biostatistics
2018-2019 Anqi Zhu, Ph.D. student, Department of Biostatistics
2017-2019 Wes Crouse, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2016-2019 Chelsea Raulerson, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2017-2018 James Xenakis, Ph.D. student, Department of Biostatistics
2016-2018 Vasyl Zhabotynsky, Ph.D. student, Department of Biostatistics
2014-2018 Xiaolei Zhou, Ph.D. student, Department of Biostatistics
2014-2018 Robert Corty, Ph.D. student, Curriculum in Bioinformatics and Computational Biology
2016-2017 Rachel Nethery, Ph.D. student, Department of Biostatistics
2015-2017 Christopher Bryant, Ph.D. student, Department of Biostatistics
2015-2016 Matthew Psioda, Ph.D. student, Department of Biostatistics
2015-2016 Matthew Holt, Ph.D. student, Department of Computer Science
2015-2016 Wei Xue, Dr.PH student, Department of Biostatistics
2015-2016 Ran Tao, Ph.D. student, Department of Biostatistics
2014-2016 Hojin Yang, Ph.D. student, Department of Biostatistics
2015 Thomas Conrad, M.S. 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

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
2012-2014 TingHuei Chen, Ph.D. student, Department of Biostatistics
2010-2014 Gabi Griffin, Ph.D. student, Curriculum in Genetics and Molecular Biology
2012-2013 Gregory Mayhew, Ph.D. student, Department of Biostatistics
2012-2013 Ni Zhao, M.S. student, Department of Biostatistics
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

GRANTS

ACTIVE

1R01HL129132 Li/Reiner (PI)(NCE) 07/15/16-02/29/21
NIH/NHLBI $621,504 10% Effort
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: contact PI

P50HD103573 Piven (PI) 07/01/20-06/30/25
NIH/NICHD $1,271,070 10% Effort
Clinical Translational Research Center for Neurodevelopmental Disorders
Intellectual and Developmental Disabilities Research Center.
Role: Director for the Data Science Core

U01 DA052713 Shen/Kriegstein PI 09/01/20-08/31/25
NIH/NIA $3,230,375 8% effort
Charting the 3D epigenome in human brain development and diseases
The goal of the project is to systematically profile the 3D epigenome in various cell types and conditions relevant to human brain development and diseases.
Role: Co-Investigator

R01 NR019245 Santos PI 07/01/20-06/30/25
NINR/NIH $1,917,272 10% Effort
Genetic and Epigenetic Effects on Childhood Cognitive Trajectories
The goal of the project is to advance our understanding of genetic and epigenetic factors related to preterm children's neurodevelopment.
Role: Co-Investigator

R01 MH123724 Sullivan (PI) 06/10/20-03/31/25
NIMH/NIH $4,401,278 5% Effort

A Trans-Nordic Study of Extreme Major Depression
The goal of the project is to conduct a trans-Nordic genetic study for extreme major depression.
Role: Co-Investigator

U24AR076730-01 LaVange, Ivanova (PI) 09/26/19-05/31/24
NIH $ 51,781,303 20% Effort
HEAL Initiative: Back Pain Consortium (BACPAC) Research Program Data Integration, Algorithm Development and Operations Management Center
The goal of this project is to set the stage for technology assessments, solicitation of patient input and utilities, and the evaluation of high-impact interventions through the innovative design and sound execution of clinical trials, leading to effective personalized treatment approaches for patients with chronic lower back pain.
Role: Co-Investigator, Leader of System Biology Group

R01MH118349 Stein/Love (PI) 12/10/18-11/30/23
NIH $493,569 4% Effort
PathQTL: Integrative Multi-Omics Causal Inference of Molecular Mechanisms Leading to Neuropsychiatric Illness
The goal of this study is to prioritize causal molecular pathways, to identify the relevant cell-type and developmental stage, and to shed light on mechanisms of neuropsychiatric disorders in an unbiased manner.
Role: Co-Investigator

R01MD013349 Aiello/Harris (PI) 08/14/18-03/31/23
NIH $678,479 5% Effort
The Add Health Epigenome Resource: Life course stressors and epigenomic modifications in adulthood
The overall goals of this project are to investigate the influence of life course psychosocial stressors on DNA methylation and gene expression that may influence cardiometabolic health and depression in a US representative study of young adults.
Role: Co-Investigator

R01GM105785 Sun (PI) 01/17/19-12/31/22
Fred Hutchinson/NIH $105,267 (Subcontract) 0% Effort
Statistical Methods for RNA-Seq Data Analysis
The goal of this study is to develop statistical/computational methods to study cell type composition or cell type-specific gene expression using bulk RNA-seq data, scRNA-seq data, or both bulk RNA-seq and scRNA-seq data.
Role: Co-Investigator

2R01DK093757 Mohlke (PI) 08/01/17-07/31/22
NIH/NIDDK $627,100 8% Effort
Genetic epidemiology of rare and regulatory variants for metabolic traits

The goal of this project is to identify novel genes for metabolic traits, discover pathogenic regulatory variants, and learn how environmental context can influence the dynamic range of gene regulation and the development of disease.
Role: Co-Investigator

R21AR075996 Sayed/Li/Mohlke PI 07/01/20-06/30/22
NIAMS/NIH $309,790 5% Effort
Evaluation of the Genetics of Hidradenitis Suppurativa
The goal of the project is to evaluate the genetics underlying Hidradenitis Suppurativa.
Role: MPI

1R01MD011609 Manuck (PI) 08/08/17-03/31/22
NIH/NIMHHD $759,176 4% Effort
The Pharmacoepigenomics of recurrent preterm birth in non-Hispanic black women
The goal of this study is to provide immediate and sustained clinical and public health impact to reduce disparities in PTB outcomes in NH black women and infants, thereby reducing neonatal mortality and lifelong morbidity.
Role: Co-Investigator

Cystic Fibrosis Foundation (5110544) Knowles (PI) 06/01/18-05/31/21
$711,233 20% Effort
Discovery of CF modifiers using whole genome sequencing-UNC
The overall goal of this project is to identify genetic modifiers for Cystic Fibrosis using whole genome sequencing data.
Role: Co-Investigator

PENDING

U01 HG011720 Li/Cox/Reiner PI 07/01/21-06/30/26
NIH/NHGRI $ 4,991,693 20% effort
Polygenic risk scores and health disparities: the role of blood cells immune response and evolutionary adaptation
The goal of the project is to refine and improve PRS for populations of diverse ancestry by integrating existing datasets with genomic and phenotype data for a broad range of complex diseases and traits.
Role: Contact PI

R01 Hu, PI 04/01/21-03/31/26
Cleveland Clinic/NIH $253,400 (UNC Subcontract) 20% effort
Genetics of mammalian chromatin interactome
The goal of the project is to study the genetics of chromatin interactome in mammalians, specifically in NeuN+ cortical neurons from Recombinant Inter Cross (RIX) mice, and to shed light on functional mechanisms between chromatin spatial organization, gene regulation, and human complex diseases and traits.
Role: Subcontract PI

COMPLETED

U54 HD079124 Piven (PI) 09/24/13-05/31/20
NIH/NICHD $1,271,070 10% Effort
Clinical Translational Research Center for Neurodevelopmental Disorders
Intellectual and Developmental Disabilities Research Center.
Role: Core Director

R01HL130733 Lange/Reiner (PI) 08/01/17-04/30/20
University of Colorado Denver/NIH $146,284 (Subcontract) 10% Effort
Sequence analysis of hematological traits in African Americans
The goal is to use sequence and genotyping data in African Americans to discover and functionally characterize novel genetic associations for blood cell traits.
Role: Co-Investigator

5216282-5500000980 Meigs (PI) 05/01/17-10/30/18
Broad Institute/NIH/NIDDK
TOPMed Whole Genome Sequence Analysis of Type 2 Diabetes and Related Traits
The goal of this project is to test the association of common, rare, and structural variation with type 2 diabetes and related traits using TOPMed Whole Genome Sequence data.
Role: Subcontract PI

5U01HL120393 Psaty (PI) 03/01/18-08/31/18
Rare variants and NHLBI traits in deeply phenotyped cohorts
The goal of this project is to leverage the deeply phenotyped cohort in NHLBI TOPMed cohorts to better understand biological pathways that may identify therapeutic targets.
Role: Subcontract PI

1R01DK101855 (NCE) North (PI) 08/15/14-7/31/18
Leveraging ancestral diversity to map adiposity loci in Hispanics
The goal of this project is to identify novel variants that influence traits related to obesity and related metabolic traits in under-represented Hispanic/Latino populations.
Role: Co-Investigator

5R01ES020836 (NCE) Whitsel (PI) 08/06/12-04/30/18
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.
Role: Co-Investigator

1R01DK093757 Mohlke (PI) 09/05/11-7/31/17
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

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

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

1R01HG006703-01 Li (PI) 05/16/12-02/28/16
NIH/NHGRI $224,285 20% Effort
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% Effort
The goal of this project is to develop and apply imputation based methodology to genome wide association and sequencing datasets.
Role: Subcontract PI

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

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% Effort
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% Effort
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% Effort
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% Effort
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 2017 - present: Regular member, Genomics, Computational Biology and Technology (GCAT) Study Section
Nov 2019: NIEHS Special Emphasis Panel for RIVER R35 Applications
Oct 2016: Behavioral Genetics and Epidemiology (BGES) Study Section
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:

Feb 2019: Genome Alberta, Canada
Oct 2017: Medical Research Council, UK
Dec 2016: Biomedical Research Fellowship Programme for India
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
PLoS Genetics (guest editor)
Human Genomics (Software section editor)

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:
Session Moderator, the American Society of Human Genetics Annual Meeting, 2017-present
Session Chair, Joint Statistical Meeting Invited Session, Baltimore, MD 2017
Session Chair, Joint Statistical Meeting Invited Session, Chicago, IL 2016
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:
2018- Biostatistics Graduate Student Recruitment Committee
2013- BBSP Recruitment Committee
2013-2019 Research Computing Advisory Committee
2017-2019 Genetics Department Advisory Committee
2016-2018 Data Science Initiative Committee
2014,2015,2016 Search Committee for Statistical Genetics Faculty Position
2012-2015 Course Organizing Committee: Statistical Software Development
2010-2016 Department of Biostatistics Computing committee
2010-2015 Bioinformatics and Computational Biology Curriculum advisory committee
2014 Search Committee for Computational Genomics Position
2014 Biostatistics Qualifying Exam Committee
2013-2014 Quantitative Sub Committee, Recruitment Committee for Biological and Biomedical Sciences Program (BBSP)
2013-2014 School of Public Health Awards Committee
2013-2014 Doctoral Examinations Committee, Department of Biostatistics
2013-2014 Award Committee, Department of Biostatistics
2013-2014 HHMI International Student Research Fellowship UNC Internal Review Committee
2013 Search Committee for Sequencing Informatics Position
2011-2012 Biostatistics Information Technology Committee, Department of Biostatistics
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

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 an integrative understanding of genetic mechanisms underlying complex human diseases and traits. Since promotion to tenure in 2015, my research has continued the approximate 50%-50% split between method development and real data applications.

Methodologically, research in my group has focused on three major areas: (1) the development of integrative genomic methods to more powerfully map genes for complex diseases and traits, and to generate mechanistic hypotheses underlying these traits, in both genetically homogeneous populations as well as recently admixed populations including African Americans and individuals of Hispanic/Latino ancestry. Efforts along this line have resulted in eight publications on which I am the last author[1-8]; (2) proposing rigorous methods and tools for the analysis and visualization of genome-wide chromatin interactome data, resulting in another four published papers on which I am the last author[9-12]; and (3) developing efficient methods for single cell RNA sequencing (scRNA-seq) data, on which topic my team has published two papers[13,14]. Most of the work, albeit recent, has received attention from the community, partly manifested by the citations. For example, DISSCO[3] has been cited 19 times since 2015, HUGIn[9] 31 times since 2017, and SAFE[14] 18 times since 2018. Below, I will expand on two specific examples from our work mentioned above.

The first example is our most recent efforts to map genes underlying hematological traits[8]. In this study, we adopted a powerful imputation strategy, which my team has established long-term expertise on, with the original method[15] cited over 1800 times and the method specifically extended to accommodate admixed populations cited over 100 times. Specifically, for ~21,600 individuals of African ancestry and ~21,700 individuals of Hispanic/Latino ancestry, we started with genotype information measured, by commercially available DNA arrays, at ~1 million genetic markers. With our imputation strategy and latest resources obtained from the NHLBI funded Trans-Omics of Precision Medicine (TOPMed) Program, we were able to directly evaluate genetic association with hematological traits at up to 30-60 million variants, substantially improving genome coverage. The strategy offers a cost-effective way to study the minority populations in the United States without having to rely on still expensive whole genome sequencing approach. Using this approach, we were able to identity associations with two rare variants (with the minor form appearing at the frequency of 0.03% or 1.14% respectively) in the hemoglobin beta gene associated with total white blood cell counts, or hematocrit and hemoglobin levels. These variants have not been detected from studies of primary European ancestry (despite the large sample sizes of over half million individuals in those studies) and would not have been identified without our imputation strategy. Such an approach will be readily adopted by many other researchers studying the genetics of almost any complex traits. The second example is our SAFE-clustering method for single cell RNA-seq clustering[14]. Single cell profiling of the transcriptome has been transforming the field in many aspects. For most of the purposes, one almost indispensable component is to identify the number and nature of cell types in the data. Despite quite a few methods developed for clustering single cell RNA-seq data, there is surprisingly low concordance among them. As a

matter of fact, using standard concordance metrics that range from 0 to 1 (with 1 indicating complete concordance and 0 complete discordance), most of the popular methods result in an average concordance of 0.2-0.4 in 12-14 real datasets we evaluated! This motivated the development of our ensemble method, SAFE-clustering, which has been shown to be among the best performing method from an independent research group[16].

In terms of applications, my team has been focusing on collaborative efforts to study hematological traits, neuropsychiatric disorders and traits, as well as cardio-metabolic traits. For all these traits, the genetic basis has been well established with many being highly heritable. Recent genetic association studies have identified tens to tens of thousands of genetic variants reproducibly associated with these traits. However, the underlying mechanisms remain elusive. My group has been making serious efforts to address several major challenges in terms of biological interpretation of results from genomic studies, prioritization and functional validation of top candidates, as well as clinical translation of genetic association findings. Our efforts have lead to 69 publications since 2015, many in high profile journals, including one in *Nature* where I contributed variant and genotype calling to the 1000 Genomes Project[17], one in *Science* where I contributed data and analysis to the PsychENCODE consortium for integrative functional dissection of human brain development and neuropsychiatric disorders[18], three in *Nature Genetics* on genetics of metrics derived from brain imaging data[19], neural cell type specific epigenetic profiling[20], and genetic association of major depressive disorders[21] respectively, just to name a few.

The above research has received multiple NIH awards. These include three grants I have served as either the sole principal investigator (PI) or contact PI: (1) R01HG006703 on “Imputation and Analysis of Rare Variants in Admixed Populations”, and (2) R01HG006292 on “Design and Analysis of Sequencing-based Studies for Complex Human Traits”, on both of which I have served as the sole PI, and (3) R01HL129132 on “Genetic Studies of Blood Cell Traits in Multi-Ethnic Cohorts”, on which I am the contact PI. Moreover, for at least two projects where I am the co-investigator, I have been playing leadership roles serving as the Director of Data Science Core in the Intellectual and Developmental Disabilities Research Center; and the Leader of the System Biology Group for the HEAL Initiative: Back Pain Consortium (BACPAC) Research Program Data Integration, Algorithm Development and Operations Management Center, each of which is a multi-million project. In addition, I have actively participated, if not helped conceived, many other funded projects pertinent to either my methodological or application research.

Finally, I am putting in multiple R01 grant proposals on (1) Genetics of mammalian chromatin interactome where we aim to generate and analyze populations of mammalian (in collaborative cross mice, and in human lymphoblastoid cell lines) interactome, transcriptome and epigenome, to study genetics of a full spectrum of chromatin interactomic phenotypes, and to elucidate and validate contributions of iQTLs to target gene expression; (2) Model-based methods, analysis and applications of chromatin interactome data where we propose to develop model-based statistical methods to map chromatin interactions at kilobase resolution, to identify cell-type-shared and cell-type-specific chromatin interactions in diverse cell types, and to study temporal dynamics of chromatin interactions across different stages of cellular differentiation and/or disease development; and (3) Uncovering genetic mechanisms underlying Alzheimer's disease where we will first aquire and harmonize various in-house,

protected and public data, encompassing bulk and single cell RNA-seq data, GWAS summary statistics, array genotyping and whole genome sequencing data, as well as myriads of functional genomic data, then analyze them using a suite of computational methods and bioinformatics tools to generate cell-type-specific mechanistic hypotheses, and finally validate top findings in iPSC-derived neural cells (particularly excitatory neurons and microglia), as well as in iPSC-derived forebrain organoids involving diverse cell types.

References:

1. Du, Y. *et al.* A SNP panel and online tool for checking genotype concordance through comparing QR codes. *PLoS One* **12**, e0182438 (2017).
2. Duan, Q. *et al.* A robust and powerful two-step testing procedure for local ancestry adjusted allelic association analysis in admixed populations. *Genet Epidemiol* **42**, 288-302 (2018).
3. Xu, Z. *et al.* DISSCO: direct imputation of summary statistics allowing covariates. *Bioinformatics* **31**, 2434-42 (2015).
4. Yan, S. *et al.* Likelihood-based complex trait association testing for arbitrary depth sequencing data. *Bioinformatics* **31**, 2955-62 (2015).
5. Yu, D. *et al.* 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**, e0126836 (2015).
6. Zhang, G. *et al.* Across-Platform Imputation of DNA Methylation Levels Incorporating Nonlocal Information Using Penalized Functional Regression. *Genet Epidemiol* **40**, 333-40 (2016).
7. Zhong, W. *et al.* Multi-SNP mediation intersection-union test. *Bioinformatics* (2019).
8. Kowalski, M.H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet* **15**, e1008500 (2019).
9. Martin, J.S. *et al.* HUGIn: Hi-C Unifying Genomic Interrogator. *Bioinformatics* **33**, 3793-3795 (2017).
10. Xu, Z. *et al.* HiView: an integrative genome browser to leverage Hi-C results for the interpretation of GWAS variants. *BMC Res Notes* **9**, 159 (2016).
11. Xu, Z. *et al.* A hidden Markov random field-based Bayesian method for the detection of long-range chromosomal interactions in Hi-C data. *Bioinformatics* **32**, 650-6 (2016).
12. Xu, Z., Zhang, G., Wu, C., Li, Y. & Hu, M. FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C data. *Bioinformatics* **32**, 2692-5 (2016).
13. Huh, R., Yang, Y., Jiang, Y., Shen, Y. & Li, Y. SAME-clustering: Single-cell Aggregated Clustering via Mixture Model Ensemble. *Nucleic Acids Res* (2019).
14. Yang, Y. *et al.* SAFE-clustering: Single-cell Aggregated (From Ensemble) Clustering for Single-cell RNA-seq Data. *Bioinformatics* (2018).
15. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology* **34**, 816-34 (2010).
16. Duo, A., Robinson, M.D. & Soneson, C. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res* **7**, 1141 (2018).
17. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
18. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**(2018).

19. Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet* (2019).
20. Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* **51**, 1252-1262 (2019).
21. Wray, N.R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681 (2018).

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 several courses after joining UNC. First, I offered a new course together with Dr. Ethan Lange: BCB725, Introduction to Statistical Genetics, which was well attended with 9-17 students registered when we taught it. Second, after tenure in 2015 and after Dr. Lange moved out of UNC, I develop a new module with Dr. Fei Zou: BCB723, Topics in Statistical Genetics and Genomics. The course has attracted a diverse audience when first taught in Spring 2018 when we had 9 students and 1 auditor. We have designed the course to cater the needs of the diverse audience. In particular, (1) we have one lecture 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. Third, I have been co-directing a full 3-credit course BIOS782 (Statistical Methods in Genetic Studies) with Dr. Danyu Lin where we mainly teach advanced doctoral students in Biostatistics, Statistics, Bioinformatics and Computational Biology, as well as Computer Science students statistical methods underlying commonly used genetic tools, allowing them to master the methods behind as well as encouraging them to develop new or modify existing methods for their purposes. Last but not the least, I have co-directed EPID889 with Dr. Kari North on “Advanced Genetic Epidemiology – High Throughput Data Analysis” where we teach hands-on skills to analyze high throughput big and complex genetic data to students in Genetic Epidemiology, many of whom had no or limited exposure to high throughput computing when entering the class.

All courses have been quite well received even with other potential competing courses such as BCB720, BIOS740, BIOS735, BIOS781, BIOS784, BIOS785, EPID743 that have certain degree of overlap at least when judged solely based on topics covered.

We therefore will continue offer these courses. In the immediate future, we will offer BCB723 in Spring 2020 and BIOS782 in Fall 2020.

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; and the 2019 BIRS Workshop on "Frontiers in Single-cell Technology, Applications and Data Analysis".

(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. Teachers cannot force them to continue this quest, but can and should take great efforts to encourage them. Teachers must also be willing to learn from what his or her students are willing to teach them.