Title

The effect of genetic ancestry on the genetic architecture of complex traits in admixed populations

Author

Spear, Melissa Lee

Publication Date

2019

The effect of genetic ancestry on the genetic architecture of complex traits in admixed populations

by

Melissa Spear

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Elad Ziv — Elad Ziv

D11C633BFE5F4F9...

Chair

DocuSigned by:

Ryan Hernandez — Ryan Hernandez

D7[illegible]4D8...

DocuSigned by:

Dara Torgerson — Dara Torgerson

[illegible]A437...

DocuSigned by:

Noah Zaitlen — Noah Zaitlen

23A5C8621D564DA...

Committee Members

To my family, for everything.

ACKNOWLEDGEMENTS

I would not be here today if it were not for the amazing mentors, colleagues, family, and friends that have supported me every step of this journey.

First to my family, who have been my rock throughout this entire journey: I know this journey was difficult at times and hard to understand, but know that I always loved you throughout it and was forever grateful to have you as my parents and sister. To my closest friends and loved ones who witnessed my growth and supported me throughout my PhD, thank you. Without your support, this journey would have been much more difficult and a lot less fun.

To Lillian So and the SOfit community, thank you for pushing me to always work on aspiring to be the best version of myself. This gem of a community was my home outside of science and I would not be the person I am today without it. I would not have been able to get through graduate school had I not discovered this amazing community who continuously pushed me to do the work, trust myself and overcome self-doubt.

At UCSF, to the office of the Biomedical Sciences Graduate Program, thank you for the support that helped me complete my research project. Thank you to the many members of the Asthma Genetics Lab for your support while I worked on the first chapter of my thesis. Specifically to Maria Pino-Yanes, it was an honor to work alongside you during my time in the lab and I learned so much from you.

To the Hernandez Lab, thank you for creating an awesome research environment over the years for me to train in. You guys were awesome and I learned so much from each and every one of you. To the additional members of my thesis committee, Noah Zaitlen and Elad Ziv, thank you for the scientific guidance and feedback that helped move my project forward. To the Gravel Lab at McGill University, thank you for welcoming me and being a second home while the Hernandez Lab started its new home at McGill.

To D’Anne Duncan, Assistant Dean of Diversity and Learner Success for the Graduate Division at UCSF, thank you for pushing me to persevere when things got hard and to continue

fighting the fight. To the members of the SACNAS at UCSF community, thank you for creating a home away from home and reminding me why my work matters and I as a Latina scientist, matter in the science community. It was an utmost honor to serve and be a part of this community during my time at UCSF.

Finally to my research advisors, Ryan Hernandez and Dara Torgerson. Ryan, thank you for your unwavering support, guidance and patience throughout these years. You took on a rotation student with very little programming experience, practically no background in population genetics and statistics and gave me the opportunity to learn and flourish. Dara, thank you for everything since day one. Your constant support, guidance, mentorship, advice, patience and belief in me have meant absolutely everything. Ryan and Dara, I can never thank the both of you enough.

ACKNOWLEDGEMENTS OF PREVIOUSLY PUBLISHED MATERIALS AND RESEARCH CONTRIBUTIONS

Chapter 2 of this dissertation was previously published. The text in Chapter 2 is a reprint of the materials as it appears in Spear et al. 2018. Donglei Hu, Maria Pino-Yanes, and Scott Huntsman assisted with the statistical analyses. Albert M. Levin, Victor E. Ortega, Elizabeth Ampleford, and Stephen P. Peters performed the replication. Celeste Eng, Marquitta J. White, Meghan E. McGarry, Angel C. Y. Mak, and Sam S. Oh were involved with data generation and/or research assistance. Neeta Thakur, Joshua M. Galanter, Adam Davis, Rajesh Kumar, Harold J. Farber, Kelly Meade, Pedro C. Avila, Denise Serebrisky, Michael A. Lenoir, Emerita A. Brigino-Buenaventura, William Rodriquez Cintron, Shannon M. Thyne, Jose R. Rodriguez-Santana, Jean G. Ford, Rocio Chapela, Andrés Moreno Estrada, Karla Sandoval, Max A. Seibold, L. Keoki Williams, and Cheryl A. Winkler contributed data to this work. Dara Torgerson, Ryan Hernandez and Esteban Gonzalez Burchard supervised the research that forms the basis of this chapter. The primary analysis was performed by Melissa Spear who also wrote the manuscript and created all figures and tables.

The contents of Chapter 3 and 4 will be submitted for review to a peer review journal. Alex Diaz-Papkovich, Simon Gravel, and Elad Ziv contributed data and analysis tools to this work. Figure B5 was created by and used with permission of Alex Diaz-Papkovich. Dara Torgerson and Ryan Hernandez supervised the research that forms the basis of this chapter. The primary analysis was performed by Melissa Spear who also wrote the manuscript and created all figures and tables outside of Figure B5.

REFERENCES

Spear, ML, Hu, D, Pino-Yanes, M, Huntsman, Huntsman, S, Eng, C, Levin, Levin, AM, White, MJ, McGarry, ME, et al. A Genome-wide Association and Admixture Mapping Study of Bronchodilator Drug Response in African Americans with Asthma. *The Pharmacogenomics Journal* (2018). doi: 10.1101/157198

ABSTRACT

The effect of genetic ancestry on the genetic architecture of complex traits in admixed populations

Melissa Lee Spear

Understanding the genetic basis of complex phenotypes is a critical problem in medical and evolutionary genetics. The evolutionary forces of natural selection and demography have shaped patterns of worldwide genetic variation, which in turn have shaped the genetic architecture of human phenotypic variation. Admixed populations, including African Americans and Latinos, have recent ancestry from two or more ancestral groups and are highly underrepresented populations in human genetics research. As a result, the genetic variation that contributes to the genetic architecture of complex traits in these populations has largely been undefined. Here through a combination of data analysis, population genetic modeling and statistical genetics, we further our understanding of admixed populations and highlight the importance of studying diverse populations. First, in a study of bronchodilator drug response (BDR), we identified both population specific and shared genetic variants associated with differences in BDR in African American and Latino children with asthma. Second, in a study of Hispanics/Latinos, we show that admixture has been a dynamic process in the recent history of Mexican Americans, with ancestry proportions changing over time due to a complex mixture of small effects from several population and cultural factors. Finally, we draw attention to the biases and potential for continued health disparities that persist when utilizing genomic prediction based only on large samples of European individuals in Mexican Americans. Through these studies, we improved upon our understanding of the genetic diversity within admixed populations, its effects on human phenotypic diversity, and subsequently our ability to understand genetic contributions to complex traits and disease.

TABLE OF CONTENTS

Appendices

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

The field of human genetics has exploded over the past two decades as a result of new genetic sequencing technologies and insights gleaned from medical and population genetic studies. From these studies, we have learned about the rich genetic diversity of populations around the world [1, 2] and how the evolutionary forces of demography and natural selection have shaped the genetic architecture of human phenotypic variation [3-8]. Despite the progress that we have made, the majority of these studies have been performed in individuals of European descent and this disparity has continued despite several calls for action [9-11]. Populations vary in terms of allele frequencies, linkage disequilibrium, and biological effect sizes of variants that affect the identification and importance of risk variants, thus European populations contain only a subset of the human genetic variation relevant to complex disease across the world (and many risk variants that are exclusive to European ancestries).

Throughout this thesis I focus on developing new insights into the genetics of a specific group of underrepresented populations: admixed populations. The genomes of recently admixed individuals are mosaics of ancestry segments from two or more previously diverged populations. The underrepresentation of diverse populations in biomedical research, including admixed populations, has impeded our ability to fully understand the genetic architecture of complex traits in these populations, and continued underrepresentation may result in further exacerbating health disparities. The majority of GWAS have been performed in European populations [9-10] and their findings have been leveraged in the clinic through use of polygenic risk scores [11]. However due to differences in the demographic history of populations around the world, these findings in Europeans do not readily translate to individuals with other ancestries [11].

In this thesis I highlight new insights into the genetics of admixed populations. First, I explore the genetics of differences in bronchodilator drug response in African Americans with asthma. I next transition to investigating the admixture dynamics of Hispanic/Latino populations living in the United States. I then conclude with examining the biases that exist with using publicly available European based genome-wide association study (GWAS) summary statistics in Mexican American individuals.

Asthma and Bronchodilator Drug Response

Asthma is a complex respiratory disease influenced by social, environmental, and genetic factors. An estimated 300 million people worldwide suffer from asthma, but prevalence varies widely between populations [12]. It is the most common chronic disease among children worldwide [13]. Among U.S. children, asthma prevalence is highest in Puerto Ricans (18.4%), followed by African Americans (14.6%), Whites (8.2%) and Mexican-Americans (4.8%) [14, 15].

Albuterol, a short-acting β_2-adrenergic receptor agonist (SABA), is the most commonly prescribed asthma medication to individuals of all racial/ethnic groups [16, 17]. Response to albuterol is known as bronchodilator response (BDR), a complex trait involving interactions among various tissues and cells, including inflammatory [18], airway epithelium [19], smooth muscle [20], and the autonomic nervous system [21]. BDR is quantitatively assessed as a change in forced expiratory volume in one second (FEV_1) after administration of a SABA. Among children in the US, there are great differences in BDR between patients and between racial/ethnic groups. African Americans and Puerto Ricans suffer the greatest morbidity due to asthma, and poor BDR likely contributes to these disparities. African Americans, in particular, have lower BDR compared to European populations even after controlling for asthma severity [22]. Differences in BDR between racial/ethnic groups may be due to environmental factors or varying frequencies of genetic variants affecting BDR. Their genomes are comprised of both African and European ancestry.

Knowledge of genetic variation that contributes to BDR in African Americans is limited [23]. Previous genetic studies in European populations have identified five candidate genes [24-30] and two genes from GWAS [31, 32] has being associated with BDR. Altogether, however, these results provide little information about the contribution of genetic variation to BDR, and there are likely to be additional variation in genes that contribute to variation in BDR in non-European populations.

Genetics of Hispanic/Latino populations

The genomes of Hispanic/Latino populations are mosaics of ancestry from three different continental populations: Native Americans who founded the Americas, Europeans as a result of colonization, and Africans from the subsequent African slave trade [33-36]. Proportions of genetic ancestry derived from each of these populations varies considerably, and is dependent on regional differences in large scale continental migrations. A limited number of studies of Latin American populations have revealed differences in disease prevalence, population specific genetic associations with disease phenotypes, and admixture dynamics at initial European colonization [33-35, 37-41].

The evolutionary forces of natural selection and demography, including migration, have shaped patterns of worldwide genetic variation, which in turn have shaped the genetic architecture of human phenotypic variation. In the United States, population demography has changed immensely over the 20th century as a result of immigration and this will continue to be one of the primary modes of population growth as the US approaches a “minority-majority” country [42]. Hispanics/Latinos are the largest and fastest growing of these groups [42]. The effect of these large-scale migrations in contributing to shaping genetic variation and subsequently phenotypic variation is unknown.

Summary

In summary, this thesis contains three pieces of work, each involved in studying the genetics of admixed populations. In chapter 2, I conduct a genome wide association and admixture mapping study of bronchodilator drug response in African Americans with asthma. This study was, to my knowledge, the first genome-wide association or admixture mapping study of BDR in African Americans with asthma, to date. Through these analyses, I identified both population specific and shared genetics variants associated with differences in BDR in African American and Latino children with asthma.

In chapters 3 and 4, I investigate how the admixture dynamics of Hispanics/Latinos in the US have changed over time and the implications this has for the genetic architecture of complex traits. Specifically in chapter 3, I highlight how global Native American ancestry has increased over time in Mexicans Americans and how this is due to a mixture of multiple genetic and social factors. In chapter 4, I examine the relationship between global Native American ancestry and multiple complex traits. Through these analyses, I identify a correlation between many of these traits and global Native American ancestry. For height specifically, I demonstrate that polygenic risk scores for height in Mexican Americans utilizing European GWAS summary statistics perform poorly and are biased based on proportions of global European ancestry. Both of these results illustrate the importance of including diverse populations in biomedical research or risk increasing health disparities.

References

1. Genomes Project, C., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
2. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
3. Agarwala, V., et al., *Evaluating empirical bounds on complex disease genetic architecture.* Nat Genet, 2013. **45**(12): p. 1418-27.
4. Eyre-Walker, A., *Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies.* Proc Natl Acad Sci U S A, 2010. **107 Suppl 1**: p. 1752-6.
5. Maher, M.C., et al., *Population genetics of rare variants and complex diseases.* Hum Hered, 2012. **74**(3-4): p. 118-28.
6. Simons, Y.B., et al., *The deleterious mutation load is insensitive to recent population history.* Nat Genet, 2014. **46**(3): p. 220-4.
7. Uricchio, L.H., et al., *Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants.* Genome Res, 2016. **26**(7): p. 863-73.
8. Yang, J., et al., *Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index.* Nat Genet, 2015. **47**(10): p. 1114-20.
9. Bustamante, C.D., E.G. Burchard, and F.M. De la Vega, *Genomics for the world.* Nature, 2011. **475**(7355): p. 163-5.
10. Popejoy, A.B. and S.M. Fullerton, *Genomics is failing on diversity.* Nature, 2016. **538**(7624): p. 161-164.
11. Martin, A.R., et al., *Clinical use of current polygenic risk scores may exacerbate health disparities.* Nat Genet, 2019. **51**(4): p. 584-591.

12. Masoli, M., et al., *The global burden of asthma: executive summary of the GINA Dissemination Committee report.* Allergy, 2004. **59**(5): p. 469-78.

13. Adams, P.F., et al., *Current estimates from the National Health Interview Survey, 1996.* Vital Health Stat 10, 1999(200): p. 1-203.

14. Moorman, J.E., et al., *National surveillance of asthma: United States, 2001-2010.* Vital Health Stat 3, 2012(35): p. 1-67.

15. Moorman, J.E., et al., *Current asthma prevalence - United States, 2006-2008.* MMWR Surveill Summ, 2011. **60 Suppl**: p. 84-6.

16. Palmer, L.J., et al., *Pharmacogenetics of Asthma.* American Journal of Respiratory and Critical Care Medicine, 2002. **165**(7): p. 861-866.

17. Nelson, H.S., *Beta-adrenergic bronchodilators.* N Engl J Med, 1995. **333**(8): p. 499-506.

18. Loza, M.J. and R.B. Penn, *Regulation of T cells in airway disease by beta-agonist.* Front Biosci (Schol Ed), 2010. **2**: p. 969-79.

19. Salathe, M., *Effects of beta-agonists on airway epithelial cells.* J Allergy Clin Immunol, 2002. **110**(6 Suppl): p. S275-81.

20. Shore, S.A. and P.E. Moore, *Regulation of beta-adrenergic responses in airway smooth muscle.* Respir Physiol Neurobiol, 2003. **137**(2-3): p. 179-95.

21. Jartti, T., *Asthma, asthma medication and autonomic nervous system dysfunction.* Clin Physiol, 2001. **21**(2): p. 260-9.

22. Naqvi, M., et al., *Ethnic-specific differences in bronchodilator responsiveness among African Americans, Puerto Ricans, and Mexicans with asthma.* J Asthma, 2007. **44**(8): p. 639-48.

23. Padhukasahasram, B., et al., *Gene-based association identifies SPATA13-AS1 as a pharmacogenomic predictor of inhaled short-acting beta-agonist response in multiple population groups.* Pharmacogenomics J, 2014.

24. Martinez, F.D., et al., *Association between genetic polymorphisms of the beta2-adrenoceptor and response to albuterol in children with and without a history of wheezing.* J Clin Invest, 1997. **100**(12): p. 3184-8.

25. Choudhry, S., et al., *Pharmacogenetic differences in response to albuterol between Puerto Ricans and Mexicans with asthma.* Am J Respir Crit Care Med, 2005. **171**(6): p. 563-70.

26. Silverman, E.K., et al., *Family-based association analysis of beta2-adrenergic receptor polymorphisms in the childhood asthma management program.* J Allergy Clin Immunol, 2003. **112**(5): p. 870-6.

27. Poon, A.H., et al., *Association of corticotropin-releasing hormone receptor-2 genetic variants with acute bronchodilator response in asthma.* Pharmacogenet Genomics, 2008. **18**(5): p. 373-82.

28. Tantisira, K.G., et al., *Molecular properties and pharmacogenetics of a polymorphism of adenylyl cyclase type 9 in asthma: interaction between beta-agonist and corticosteroid pathways.* Hum Mol Genet, 2005. **14**(12): p. 1671-7.

29. Litonjua, A.A., et al., *ARG1 is a novel bronchodilator response gene: screening and replication in four asthma cohorts.* Am J Respir Crit Care Med, 2008. **178**(7): p. 688-94.

30. Duan, Q.L., et al., *A polymorphism in the thyroid hormone receptor gene is associated with bronchodilator response in asthmatics.* Pharmacogenomics J, 2013. **13**(2): p. 130-6.

31. Himes, B.E., et al., *Genome-wide association analysis in asthma subjects identifies SPATS2L as a novel bronchodilator response gene.* PLoS Genet, 2012. **8**(7): p. e1002824.

32. Israel, E., et al., *Genome-wide association study of short-acting beta2-agonists. A novel genome-wide significant locus on chromosome 2 near ASB3.* Am J Respir Crit Care Med, 2015. **191**(5): p. 530-7.

33. Gravel, S., et al., *Reconstructing Native American migrations from whole-genome and whole-exome data.* PLoS Genet, 2013. **9**(12): p. e1004023.

34. Homburger, J.R., et al., *Genomic Insights into the Ancestry and Demographic History of South America.* PLoS Genet, 2015. **11**(12): p. e1005602.

35. Moreno-Estrada, A., et al., *Reconstructing the population genetic history of the Caribbean.* PLoS Genet, 2013. **9**(11): p. e1003925.

36. Reich, D., et al., *Reconstructing Native American population history.* Nature, 2012. **488**(7411): p. 370-4.

37. Pino-Yanes, M., et al., *Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos.* J Allergy Clin Immunol, 2015. **135**(6): p. 1502-10.

38. Pino-Yanes, M., et al., *Genetic ancestry influences asthma susceptibility and lung function among Latinos.* J Allergy Clin Immunol, 2015. **135**(1): p. 228-35.

39. Moreno-Estrada, A., et al., *Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits.* Science, 2014. **344**(6189): p. 1280-5.

40. Drake, K.A., et al., *A genome-wide association study of bronchodilator response in Latinos implicates rare variants.* J Allergy Clin Immunol, 2014. **133**(2): p. 370-8.

41. Galanter, J.M., et al., *Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study.* J Allergy Clin Immunol, 2014. **134**(2): p. 295-305.

42. Colby, S.L.O., J.M., *Projections of the Size and Compositon of the U.S. Population: 2014 to 2060*. 2015, U.S. Census Bureau.

Chapter 2: A Genome-wide Association and Admixture Mapping Study of Bronchodilator Drug Response in African Americans with Asthma

INTRODUCTION

Albuterol, a short-acting β_2-adrenergic receptor agonist (SABA), is the most commonly prescribed asthma medication worldwide [1]. SABAs cause rapid smooth muscle relaxation of the airways. Bronchodilator drug response (BDR) is a measure of a patient's clinical response to SABA treatment and is quantitatively assessed as a change in forced expiratory volume in one second (FEV_1) after administration of a SABA. BDR is a complex trait involving interactions among inflammatory cells [2], airway epithelium [3], smooth muscle cells [4], and the autonomic nervous system [5]. Variation in BDR is likely influenced by both population-specific and shared environmental and genetic factors [6-8]. In the United States (U.S.), BDR in children with asthma differs significantly between racial/ethnic groups [9]. Specifically, African Americans have lower BDR compared to European populations even after controlling for asthma severity [10]. Compared to European Americans, African Americans suffer increased asthma morbidity and mortality [7, 9, 11] and decreased BDR likely contributes to these disparities in disease progression and outcomes. The extensive use of albuterol as a first-line therapy for asthma, coupled with the decreased drug response (BDR) and increased disease burden in African Americans underscores the importance of identifying genetic factors that influence BDR in African American children with asthma. Once identified, these factors may lead to the generation of novel therapies and targeted interventions that will serve to improve patient care and asthma outcomes in an over-burdened and under-studied population.

To date, knowledge of genetic variation that contributes to BDR in African Americans is limited to a single genome-wide association study (GWAS) in 328 individuals [12]. Previous GWAS and candidate gene studies performed in populations of predominantly European ancestry with asthma have identified several BDR candidate genes [8, 13-22]. A recent study in Latinos with asthma replicated a number of these findings, and also identified novel population-specific associations with BDR [6]. Genetic effects identified in one population are not always generalizable across populations and several population-specific asthma-risk variants have been discovered in African-descent populations (i.e. African Americans and Latinos) [23-25]. Additionally, previous studies have shown that the varying degrees of African and European ancestry present in the African American population can be leveraged, through a technique known as admixture mapping, to identify the missing heritability of complex traits [26]. Admixture mapping is a genome-wide approach that uses the variable allele frequencies of multiple SNPs between different ancestral populations to test for an association between local ancestry and phenotype [26]. The likelihood of population-specific effects, the limited number, and scale, of prior studies of BDR performed in African Americans, and ability to perform admixture mapping analysis highlights the possibility of gaining novel information through evaluating the impact of common genetic factors on BDR in African American children with asthma.

In this study, we performed a GWAS and admixture mapping study of bronchodilator drug response in 949 African American children with asthma from the Study of African Americans, Asthma, Genes & Environments (SAGE I and II) [27]. To increase power and distinguish between population-specific vs. shared associations, we also performed a trans-ethnic meta-analysis across our SAGE I and SAGE II participants and 1840 Latinos from GALA II (Genes-environments and Admixture in Latino Americans) studies [24], respectively (total N=2,789). We further attempted replication of our population-specific and trans-ethnic meta-analysis results in 416 Latinos from the Genetics of Asthma in Latino Americans study (GALA I) [7, 28], 1,325 African Americans from the Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity (SAPPHIRE) [29] and 290 African Americans from the Severe Asthma Research Program (SARP) [30, 31].

METHODS

Study subjects from the Study of African Americans, Asthma, Genes & Environments

The Study of African Americans, Asthma, Genes & Environments (SAGE) is an ongoing case-control study of asthma in children and adolescents recruited from the San Francisco Bay Area in California [27]. Subjects were eligible if they were 8-21 years of age and self-identified all four grandparents as African American. Exclusion criteria included: (1) 10 or more pack-years of smoking; (2) any smoking within 1 year of recruitment date; (3) pregnancy in the third trimester; or (4) history of one of the following conditions: sickle cell disease, cystic fibrosis, sarcoidosis, cerebral palsy, or history of heart or chest surgery. Asthma was defined by physician diagnosis, asthma medication use and reported symptoms of coughing, wheezing, or shortness of breath in the 2 years preceding enrollment. Detailed clinical measurements were recorded for each individual whom DNA was collected from. In addition, trained interviewers administered questionnaires to obtain baseline demographic data, as well as information on general health, asthma status, social, and environmental exposures. Pulmonary function testing was conducted with a KoKo® PFT Spirometer (nSpire Health Inc., Louisville, CO) according to American Thoracic Society recommendations [32], to obtain forced expiratory volume in one second (FEV_1) in addition to other standard measurements of airway obstruction. Subjects with asthma were instructed to withhold their bronchodilator medications for at least 8 hours before testing. After completing baseline spirometry, subjects were given albuterol administered through a metered-dose inhaler (90 mcg/puff) with a spacer, and spirometry was repeated after 15 minutes to obtain post-bronchodilator measurements. The dose of albuterol was different in early stages of SAGE recruitment (2001-2005: SAGE I) than in more recent participants (2006-present: SAGE II). In SAGE I, post-bronchodilator FEV_1 values were measured after providing the participants 2 puffs of albuterol (180 μg) if they were younger than 16 years of age and 4 puffs of albuterol (360 μg) if they were 16 years of age or older. In SAGE II, two doses of albuterol were delivered. For the first dose, 4 puffs of albuterol (360 μg) were provided independently of the age of the participant.

For the second dose, two puffs (180ug) for children < 16 years old were administered and 4 puffs for subjects older ≥ 16 years.

Body mass index (BMI) was calculated for each participant using weight and height measures and converted to a categorical scale of underweight, normal, overweight, and obese according to the Centers for Disease Control and Prevention. For participants under 20 years old, standardized sex- and age-specific growth charts were used to calculate BMI percentiles (http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm) and categorize their BMI as: underweight (BMI percentile<5th), normal (5th≤BMI<85th), overweight (85th≤BMI<95th), and obese (BMI≥95th). For participants older than 20 years old, BMI categories (http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html - interpretedAdults) were defined as: underweight (BMI<18.5), normal (18.5≤BMI≤24.9), overweight (25≤BMI≤29.9) and obese (BMI≥30). Further information about SAGE can be found in Appendix A.

Institutional review boards approved the study and all subjects/parents provided written assent/consent, respectively.

Genotyping and quality control (SAGE)

A total of 1,819 samples (1,011 asthma cases and 810 controls) were genotyped with the Axiom® World Array 4 (Affymetrix, Santa Clara, CA) at ~800,000 SNPs. Quality control was performed by removing SNPs that failed manufacturer's quality control, had genotyping call rates below 95%, and/or had a deviation from Hardy-Weinberg equilibrium ($p<10^{-6}$) within controls. 772,135 genotyped SNPs passed quality control. Samples were filtered based on discrepancy between genetic sex and reported gender and cryptic relatedness (PI_HAT>0.3). We excluded 3 subjects who were outliers for BDR (BDR of >60, or <-10). After sample quality control we included 759 SAGE II and 190 SAGE I asthma cases, for a total of 949 individuals with both genome-wide SNP data and measurements of bronchodilator drug response in the current study (Table 2.1).

Table 2.1. Descriptive statistics of SAGE I, SAGE II, & GALA II asthma cases. Values shown are the means, with the standard deviation in parentheses.

	SAGE I	**SAGE II**	**GALA II**
Total (N)	190	759	1830
Age (year)	18 (9.3)	14 (3.6)	13 (3.2)
<18 years (%)	64%	86%	93%
Sex (%Male)	41%	52%	55%
Race/Ethnicity	African American	African American	Latino
Global African Ancestry	0.81 (0.13)	0.72 (0.12)	0.15 (0.13)
Global Native American Ancestry	-	-	0.30 (0.25)
BMI			
<20 years	25 (7.3) (N=132)	25 (7.2) (N=722)	23 (6.5) (N=1782)
>20 years	31 (7.8) (N=58)	29 (7.0) (N=37)	30 (6.6) (N=48)
Pulmonary Function			
Pre-FEV1 % Predicted	92 (16)	99 (14)	91 (16)
Pre-FVC % Predicted	100 (17)	104 (13)	95 (16)
BDR (%)	9 (9.1)	9.5 (6.9)	11 (8.2)

Phasing of genotyped SNPs was performed using SHAPE-IT [33], and genotype imputation was performed using IMPUTE2 [34, 35] using all populations from 1000 Genomes Project Phase III [36] as a reference. Following imputation, a total of 9,573,507 genotyped and imputed (info score >0.3) SNPs with a MAF>0.05 were analyzed for SAGE II and 9,605,653 were analyzed for SAGE I.

Study subjects from the Genes-environments & Admixture in Latino Americans study

A total of 1,830 Latino children with asthma genotyped with the Axiom LAT1 array (World Array 4, Affymetrix) for the Genes-environments and Admixture in Latino Americans (GALA II) study were included in our analysis (Table 2.1). Asthma cases were defined as participants with a history of physician diagnosed asthma and the presence of two or more symptoms of coughing, wheezing, or shortness of breath in the 2 years preceding enrollment. Detailed clinical measurements were recorded for each individual whom DNA was collected from and each individual underwent spirometry. BDR was calculated as the percentage change in FEV_1 after 2 doses of albuterol (post-FEV_1) compared with base line values before administration of albuterol (pre-FEV_1). Specifically, post-bronchodilator FEV_1 values were measured after providing the participants 2 doses of albuterol, with a 15-minute waiting period after each dose. A total of 6 (if <16 years of age) to 8 (if ≥16 years of age) puffs of albuterol were administered. A total of 408 patients from the Centro de Neumologia Pediatrica in Puerto Rico were recruited based on having a BDR of at least 8%; of these, 121 patients were recruited based on having a BDR of at least 12%. Further details about GALA II are described the Appendix A and in depth elsewhere [6]. Imputation procedures identical to those described above for SAGE I and SAGE II were implemented, resulting in a total of 7,498,942 genotyped and imputed (info score >0.3) SNPs with a MAF>0.05.

Study subjects from the Genetics of Asthma in Latino Americans study

For our replication phase, 247 Mexican and 169 Puerto Rican asthma cases genotyped with the Genome-Wide Human SNP Array 6.0 (Affymetrix) for the Genetics of Asthma in Latino Americans (GALA I) study were included. Children were included in the study if they were between the ages of 8-40 with physician diagnosed mild to moderate-severe asthma and had experienced two or more symptoms in the previous two years at time of recruitment (including wheezing, coughing and/or shortness of breath.) BDR was measured in a similar way to GALA II, but with a lower dosage of albuterol. Specifically, post-FEV1 values were measured after only a single dose of albuterol (compared with 2 doses in GALA II). Two (if <16 years of age) to 4 (if ≥16 years of age) total puffs of albuterol were administered (compared with 4 [if <16 years of age] and 6 [if ≥16 years of age] in GALA II). Further details of the study are described in the Appendices A and elsewhere [8, 28].

Study subjects from the Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity

For additional replication, we included 1,325 Africans Americans with asthma from the Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity (SAPPHIRE) [29] genotyped with the Genome-Wide Human SNP Array 6.0 (Affymetrix). Subjects met the following criteria: age 12-56 years, had a diagnosis of asthma (based on both patient report and documentation in the medical record), and did not have a prior diagnosis of chronic obstructive pulmonary disease or congestive heart failure (CHF), a baseline FEV1 between 40-90% predicted, >12% baseline bronchodilator reversibility, no smoking in the preceding year or <10 pack-year smoking history total, no oral or inhaled corticosteroid use in the 4 weeks preceding screening, and not pregnant at the time of enrollment and not intending to get pregnant during the study period. Lung function testing was performed using a Fleisch-type pneumotachometer (KoKo PFT Spirometer®, nSpire Health Inc., Louisville, CO) and following 2005 ATS/ERS spirometry recommendations.(27;28) Patients using inhaled bronchodilators were asked to withhold these medications for the 12 hours prior to lung function tests. To assess bronchodilator response a 360 microgram (mcg) dose (i.e., 4 puffs) of inhaled albuterol sulfate hydrofluoroalkane (HFA) (GlaxoSmithKline, Research Triangle Park, NC) from a standard metered dose inhaler (MDI) using an AeroChamber Plus Flow-Vu® spacer (Monahan Medical Corp., Plattsburgh, NY) was administered to participants. Pulmonary function was reassessed 15 minutes after administering albuterol. Bronchodilator response was measured as the change in forced expiratory volume at one second (FEV1) between the baseline (pre-bronchodilator) measure and post-bronchodilator FEV1. Estimates of local ancestry were obtained using RFMix [37].

Study subjects from the Severe Asthma Research Program

We included 290 African Americans with mild to severe asthma from the Severe Asthma Program (SARP) genotyped with the Illumina 1Mv1 platform [23]. Subjects met the American Thoracic Society (ATS) definition of severe persistent asthma [38] and were characterized according to asthma severity (see [30, 31]).

Assessment of genetic ancestry

Genotypes from two populations were used to represent the ancestral haplotypes of African Americans for estimating local ancestry: HapMap European (CEU) and HapMap Africans (YRI). For Latinos, genotypes from 71 Native Americans were used as an additional ancestral population [39]. These 71 individuals included: 14 Zapotec, 2 Mixe, and 11 Mixtec from the southern State of Oaxaca [40] and 44 Nahua individuals from Central Mexico [3]. Global ancestry was estimated using ADMIXTURE [41] in a supervised analysis assuming two ancestral populations for African Americans and three ancestral populations for Latinos. A union set of SNPs was obtained by merging genotyped SNPs in SAGE and the ancestral populations (CEU/YRI). Local ancestry was estimated using the program LAMP-LD [40] in the GALA and SAGE studies and with RFMix [37] in SAPPHIRE.

Genotype association testing

All statistical analyses were conducted using R (version 2.15.3). For SAGE individuals, we used standard linear regression to test for an association between BDR and allele dosage at each individual SNP, adjusting for age, sex, BMI category, and both global and local African ancestry. A GWAS of BDR in GALA II has been previously published [6], however, this previous work did not include adjustment for BMI; in this study we re-ran the GWAS using a new reference imputation panel and further adjusted by BMI [42]. For GALA II individuals, we adjusted for age, sex, BMI category, ethnicity, global Native American and African ancestry, and local ancestry. All analyses were performed using imputed genotypes from 1000 Genomes phase III. Using the fixed-effects model implemented in METAL [43], we performed a meta-analysis of common variants (MAF ≥ 5%) across African Americans (SAGE I and SAGE II) and Latinos (GALA II). We selected variants that were common (MAF ≥ 5%) within each individual study and then took the intersection of SNPs for the meta-analysis.

Admixture mapping

We used local ancestry estimates generated across the genome to perform admixture mapping in African Americans. Linear regression models adjusted for age, sex, BMI category, and global African ancestry were used to identify significant associations between local ancestry estimates and BDR. The threshold for genome-wide significance was calculated using the empirical autoregression framework with the package *coda* in R to estimate the total number of ancestral blocks [44, 45]. The Bonferroni threshold was calculated as $\alpha=2.4\times10^{-4}$ based on 245 ancestral blocks. For African Americans, admixture mapping was performed separately in SAGE I and SAGE II and combined in a meta-analysis using METAL [43]. An admixture mapping study of BDR in GALA II has been previously published [6], however, this previous work did not include adjustment for BMI; in this study we re-ran the admixture mapping study further adjusting by BMI [42] to be consistent with the SAGE I and SAGE II analyses. For GALA II Latinos, linear regression models adjusted for age, sex, ethnicity, BMI category, global Native American and African ancestry were used to identify significant associations between local ancestry estimates and BDR. We further combined the African ancestry results of SAGEI, SAGE II and GALA II in a meta-analysis using METAL [43].

Replication in GALA I, SAPPHIRE, and SARP

We attempted replication of significant population-specific (SAGE I and SAGE II) and cosmopolitan (SAGE I, SAGE II, GALA II) associations with BDR in the GALA I, SAPPHIRE, and SARP studies. Replication in GALA I was performed using genotype imputation (i.e. *in silico* replication), followed by an examination at a locus-wide level for SNPs within +/- 50 kb. We imputed 100 kb regions around each SNP using the program IMPUTE2 for Mexican and Puerto Rican participants run separately using 1000 Genomes phase 3 haplotypes as a reference. Linear regression was used to test for an association between allele dosage and BDR separately in Mexicans and Puerto Ricans, adjusting for age, sex, BMI category, global and local ancestry. Replication in SAPPHIRE was performed using linear regression to test for an association between allele dosage and BDR in African Americans while adjusting for age, sex, BMI category, and global and local African ancestry. Replication in SARP was performed using linear regression to test for an association between allele dosage and BDR in African Americans while adjusting for age, sex, BMI, and global African ancestry. For GALA I and SAPPHIRE replication, statistical significance at the SNP level was evaluated at p<0.05, and at the locus-wide level was established using a conservative Bonferroni correction adjusting by the number of SNPs within +/- 50 kb of the original candidate SNP. For SARP replication, statistical significance was evaluated at p<0.05 at the SNP level only.

RESULTS

GWAS results

After filtering variants with a MAF ≥ 5% and with imputation quality score (info score) ≥ 0.3, we tested for an association of BDR at a total of 9,190,349 SNPs in 949 African Americans with asthma (λ = 1.006). We identified a single genome-wide significantly associated SNP within an intergenic region on chromosome 9 (rs73650726, imputation quality score=0.86) (Figures 2.1, 2.2, Figure A.1, Table 2.2). At this variant, additional copies of the A1 allele (A), was associated with decreased drug response (β=-3.8, p=7.69x10^{-9}) (Table 2.2 & Table A.2).

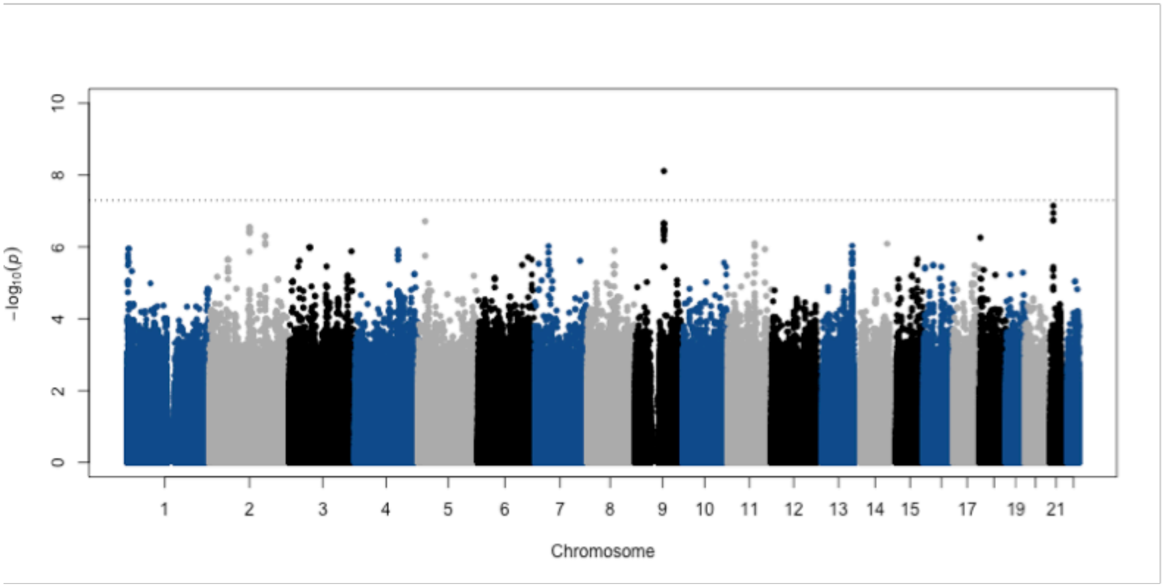

Figure 2.1. Meta-analysis of genome-wide association studies with BDR in African Americans. Association testing for BDR was performed using linear regression including age, sex, BMI category, local and global ancestry as covariates separately in SAGE I and II and combined in a meta-analysis. Dotted line indicates the genome-wide significance threshold of 5 × 10^{-8}.

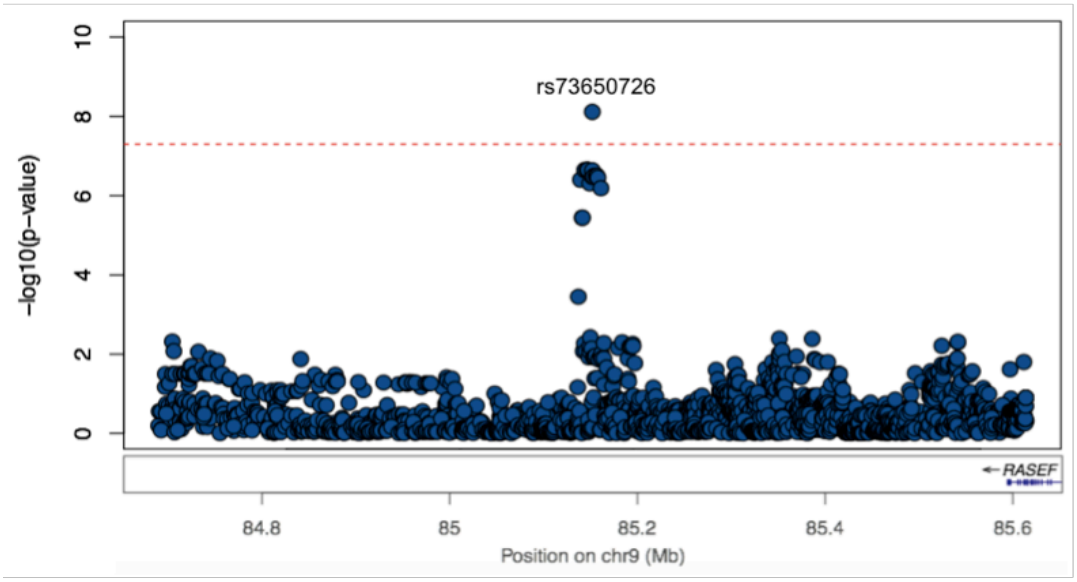

Figure 2.2. LocusZoom plot of chr9:84653000–85653000. Region includes genotyped and imputed variants from 1000 Genomes phase 3. Blue = variants common in SAGE I and II. Dotted line indicates the genome-wide significance threshold of 5 × 10^{-8}.

Table 2.2: Genome-wide significant associations identified through a meta-analysis within African Americans (SAGE I and II), and within African Americans and Latinos (SAGE I, SAGE II, and GALA II). Under 'Direction' the first symbol refers to SAGE I, second to SAGE II, and third to GALA II. 0 = absent/rare in study.

African Americans (SAGE I and II):								
Chr	**SNP**	**Position (hg19)**	**A1**	**A2**	**Effect (A1)**	**StdErr**	**Pvalue**	**Direction**
9q21	rs73650726	85152666	A	G	-3.8	0.66	7.69x10^{-9}	--0
African Americans + Latinos (SAGE I, SAGE II, GALA II):								
Chr	**SNP**	**Position (hg19)**	**A1**	**A2**	**Effect (A1)**	**StdErr**	**Pvalue**	**Direction**
10q21	rs7903366	53689774	T	C	1.23	0.22	3.94x10^{-8}	+++
10q21	rs7070958	53691116	A	G	-1.24	0.23	4.09x10^{-8}	---
10q21	rs7081864	53690331	A	G	1.23	0.22	4.94x10^{-8}	+++

The SNP rs73650726 is common in African Americans but rare in Latinos, with a minor allele frequency of 8% in both SAGE studies, but at a frequency of 1% in GALA II. This is consistent with allele frequencies observed in the 1000 Genomes Project, where the variant is common in African populations (8%), rare in Latino populations (1-2%), and absent in European and Asian populations (Figure 2.3) [46].

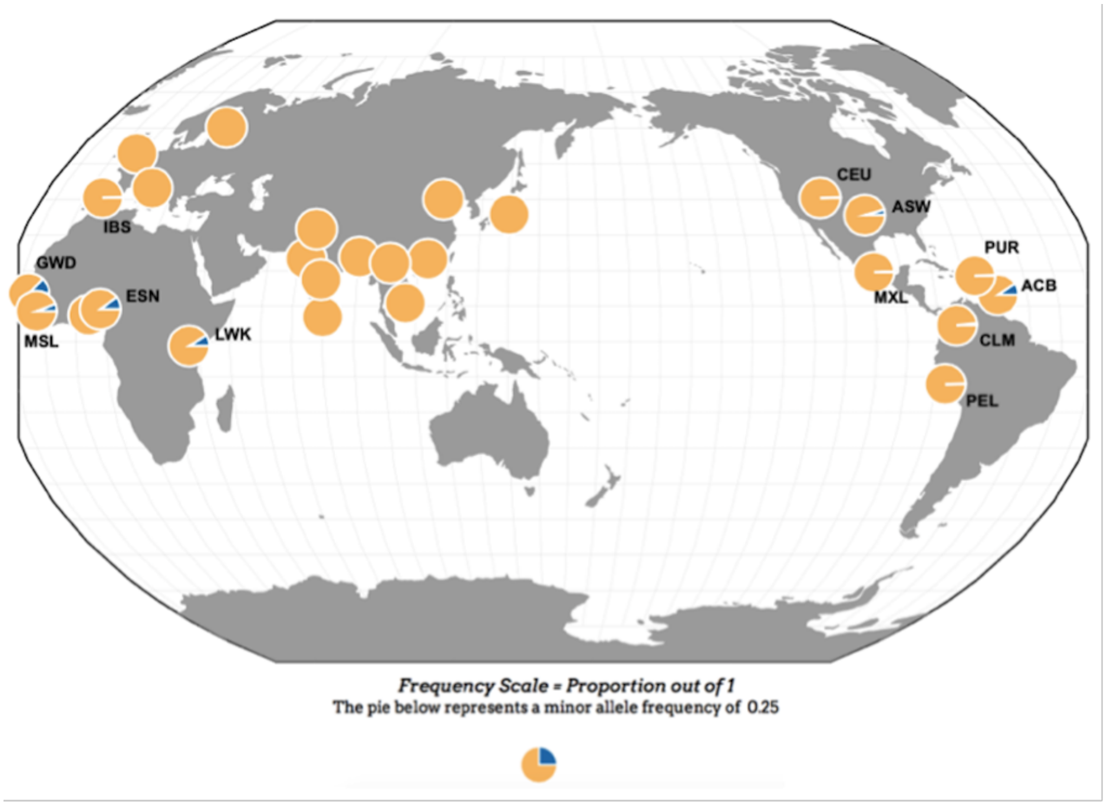
IBS
GWD
ESN
MSL
LWK
CEU
ASW
PUR
ACB
MXL
CLM
PEL
Frequency Scale = Proportion out of 1
The pie below represents a minor allele frequency of 0.25

Figure 2.3. Geographic distribution of allele frequencies of rs73650726. Each pie chart refers to a population from the 1000 Genomes Project phase 3. Yellow= Major allele (A), blue = minor allele (G). rs73650726 is common only in populations with African ancestry.

In order to increase power and identify associations shared between populations we performed a trans-ethnic meta-analysis across African American, and Latino participants from SAGE I, SAGE II, and GALA II, respectively. Following quality control and filtering on variants common in each study (MAF ≥ 5%), we took the overlap between the three studies and we performed a meta-analysis on 6,570,864 SNPs. We identified genome-wide significant associations at three SNPs located on chromosome 10 within the intron of *PRKG1*: rs7903366 (β=1.23, p=3.94x10^{-8}), rs7070958 (β=-1.24, p=4.09x10^{-8}), and rs7081864 (β=1.23, p=4.94x10^{-8}) (imputation quality scores > 0.98, Figures 2.4 & 2.5, Table 2.2, Table A.2). All three SNPs are eQTLs for *PRKG1* in lung tissue from the GTEx database (Table 2.3) [47], with the minor allele associated with decreased expression.

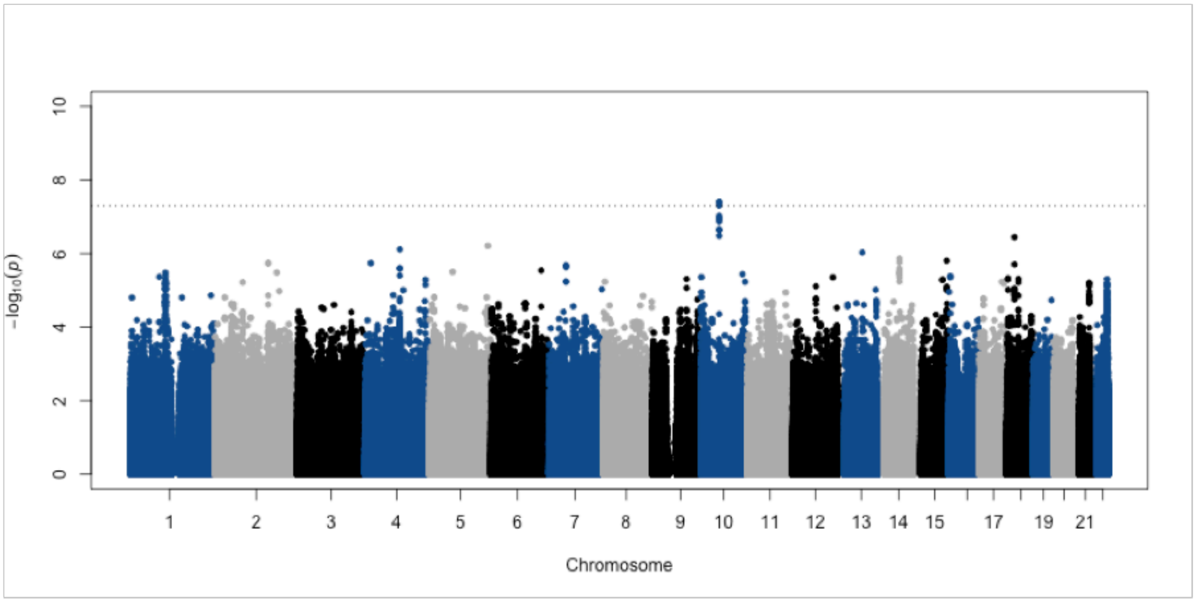

Figure 2.4. Meta-analysis of genome-wide association studies with BDR in African Americans and Latinos. Association testing for BDR was performed using linear regression including age, sex, BMI category, local and global ancestry as covariates; including ethnicity for GALA II. Dotted line indicates the genome-wide significance threshold of 5 × 10^{-8}.

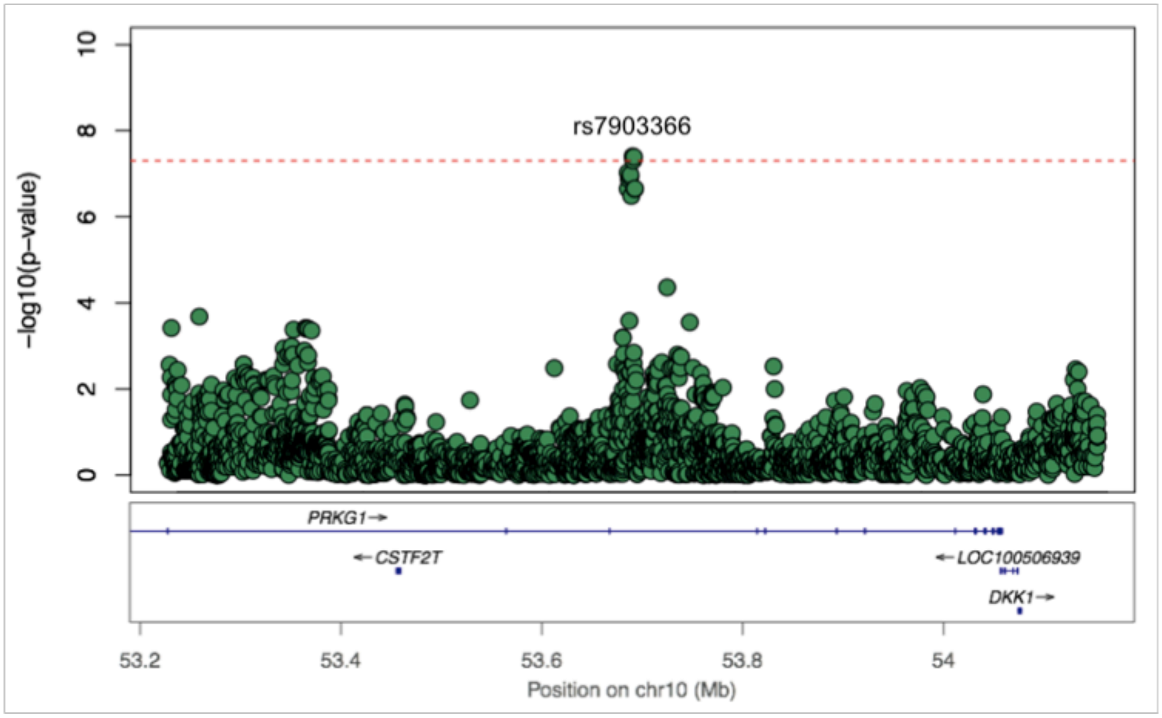

Figure 2.5. LocusZoom plot of chr10:53200000–54200000. Region includes genotyped and imputed variants from 1000 Genomes phase 3. Green = variants common in SAGE I, SAGE II and GALA II. Dotted line indicates the genome-wide significance threshold of 5 × 10^{-8}.

Replication of African American population-specific (rs73650726) and shared (rs7903366, rs7070958, rs7081864) was attempted in two independent Latino (GALA I) and African American (SAPPHIRE) studies. Although none of the identified associations replicated in either study population, the African American population-specific association between rs73650726 and BDR, identified in the SAGE studies, displayed a similar trend in direction of effect in Puerto Ricans (GALA I; β = -6.2) and in African Americans (SAPPHIRE, β = -0.65) (Table A.3). In addition, none of the SNPs within 50 kb of the four original SNPs were significantly associated with BDR following Bonferroni correction (Table A.4). Lastly, we evaluated previously identified candidate SNPs from prior candidate gene and GWAS with BDR in patients with asthma. After accounting for fifteen comparisons, no SNPs met the statistical significance threshold ($p<3.33x10^{-3}$) (Table A.5); only rs9551086 in *SPATA13* had a p-value below 0.05 (p=0.02).

Admixture mapping results

We tested for an association of BDR with local genetic ancestry inferred at 478,441 SNPs in 949 African Americans with asthma (190 from SAGE I and 759 from SAGE II) (Figures A.2 & A.3). A meta-analysis across both studies yielded no significant associations with ancestry ($p < 2.4 \times 10^{-4}$) (Figure A.4). The most significant peak was located on chromosome 8p11 where African ancestry was associated with higher BDR ($\beta=1.49$, $p=6.34 \times 10^{-4}$) (Table A.6). A meta-analysis across SAGE I, SAGE II and GALA II yielded results consistent with previous findings in the original admixture mapping study of GALA II (see [6]) (Figure A.5).

DISCUSSION

We performed a genome-wide association study for bronchodilator drug response in African Americans, and identified a population-specific association between rs73650726, located on chromosome 9, and BDR. Specifically, we discovered that the G allele of rs73650726 was associated with increased BDR and is more common in African Americans compared to European populations (Figure 2.3). The variant rs73650726, located on chromosome 9, does not map to any gene, but SNPs in high linkage disequilibrium ($r^2 \geq 0.8$) with this marker are located in enhancer histone marks in lung tissues [47].

Table 2.3: Correlation between the expression of PRKG1 in the lung and minor alleles at three intronic SNPs associated with BDR (cis-eQTLs). Data is from the GTEx database.

SNP	Ref Allele	Pvalue	Effect (Ref Allele)	T-Statistic	StdErr	Tissue	Gene
rs7903366	C	0.00051	-0.12	-3.5	0.034	Lung	*PRKG1*
rs7070958	A	0.00046	-0.12	-3.6	0.034	Lung	*PRKG1*
rs7081864	G	0.00052	-0.12	-3.5	0.034	Lung	*PRKG1*

Our results demonstrate that population-specific genetic variation contributes to variation in BDR in African American children with asthma. We further combined our results in a meta-analysis for BDR in African Americans and Latinos and identified multiple intronic variants in *PRKG1* that were associated with BDR in both populations. Overall our results demonstrate that population-specific and shared genetic factors contribute to variation in BDR among African American children with asthma.

Three of our significantly associated variants fell within the intronic region of an annotated gene, Protein Kinase, CGMP-Dependent, Type I (*PRKG1*). *PRKG1* encodes for a cyclic GMP-dependent protein kinase, which phosphorylates proteins involved in regulating platelet activation and adhesion [48], gene expression [49, 50], vascular smooth muscle cell contraction [51], and feedback of the nitric-oxide (NO) signaling pathway [52]. Notably, the NO pathway is a key pathway in modulating vasodilation in response to beta-agonists such as albuterol via beta 2-adrenergic receptors [53], making *PRKG1* a highly plausible BDR candidate gene. The three SNPs are in high linkage disequilibrium ($r^2 \geq 0.8$) with variants known to be functional [54], and are all associated with the expression of *PRKG1* in the lung – a tissue highly relevant to BDR. From the Genotype-Tissue Expression (GTEx) project database, the reference allele for all three SNPs was associated with decreased expression of the gene in lung tissue [47]. Thus, additional studies are required to identify the causal underlying variation at this locus, such as direct sequencing of this locus, and how the expression of *PRKG1* may be related to differences in BDR.

We attempted to replicate our study findings and candidate SNPs previously found to be associated with BDR, however we found no significant associations following multiple testing corrections. This could be due to differences in study design (Figure A.6), the presence of population specific differences in genetic contributions to BDR, lack of power due to small populations sizes, and/or varying patterns of linkage disequilibrium between populations. Furthermore, we were limited in sample size in GALA I [7, 28] to evaluate associations at low frequency variants, and note that SAPPHIRE is comprised of mainly adults [29] in comparison to SAGE and GALA II that are comprised of mainly children.

In conclusion, we identified two novel loci with biological plausibility whereby genetic variation is associated with differential response to albuterol, the most commonly prescribed asthma medication. One of these loci contains variation associated with BDR that is common to African Americans, a population that has historically been understudied in genetic studies [55-57]. Further genetic studies in African Americans are essential for identifying a more comprehensive set of genetic variants that contribute to differences in BDR, which in turn will lead to a better understanding of the pharmacogenetic response to asthma therapies. This will provide the foundation for genetic risk profiling and precision medicine, identifying novel genes and pathways associated with BDR, and the development of novel asthma therapies.

References:

1. Nelson, H.S., *Beta-adrenergic bronchodilators.* N Engl J Med, 1995. **333**(8): p. 499-506.
2. Loza, M.J. and R.B. Penn, *Regulation of T cells in airway disease by beta-agonist.* Front Biosci (Schol Ed), 2010. **2**: p. 969-79.
3. Kumar, R., et al., *Factors associated with degree of atopy in Latino children in a nationwide pediatric sample: the Genes-environments and Admixture in Latino Asthmatics (GALA II) study.* J Allergy Clin Immunol, 2013. **132**(4): p. 896-905 e1.
4. Shore, S.A. and P.E. Moore, *Regulation of beta-adrenergic responses in airway smooth muscle.* Respir Physiol Neurobiol, 2003. **137**(2-3): p. 179-95.
5. Jartti, T., *Asthma, asthma medication and autonomic nervous system dysfunction.* Clin Physiol, 2001. **21**(2): p. 260-9.
6. Drake, K.A., et al., *A genome-wide association study of bronchodilator response in Latinos implicates rare variants.* J Allergy Clin Immunol, 2014. **133**(2): p. 370-8.
7. Burchard, E.G., et al., *Lower bronchodilator responsiveness in Puerto Rican than in Mexican subjects with asthma.* Am J Respir Crit Care Med, 2004. **169**(3): p. 386-92.
8. Choudhry, S., et al., *Pharmacogenetic differences in response to albuterol between Puerto Ricans and Mexicans with asthma.* Am J Respir Crit Care Med, 2005. **171**(6): p. 563-70.
9. Naqvi, M., et al., *Ethnic-specific differences in bronchodilator responsiveness among African Americans, Puerto Ricans, and Mexicans with asthma.* J Asthma, 2007. **44**(8): p. 639-48.
10. Blake, K., et al., *Population pharmacodynamic model of bronchodilator response to inhaled albuterol in children and adults with asthma.* Chest, 2008. **134**(5): p. 981-9.

11. Gorina, Y., *QuickStats:asthma*death rates, by race and age group - United States, 2007–2009.* In (MMWR) MaMWR (ed.). Centers for Disease Control and Prevention, 2012.

12. Padhukasahasram, B., et al., *Gene-based association identifies SPATA13-AS1 as a pharmacogenomic predictor of inhaled short-acting beta-agonist response in multiple population groups.* Pharmacogenomics J, 2014.

13. Martinez, F.D., et al., *Association between genetic polymorphisms of the beta2-adrenoceptor and response to albuterol in children with and without a history of wheezing.* J Clin Invest, 1997. **100**(12): p. 3184-8.

14. Silverman, E.K., et al., *Family-based association analysis of beta2-adrenergic receptor polymorphisms in the childhood asthma management program.* J Allergy Clin Immunol, 2003. **112**(5): p. 870-6.

15. Poon, A.H., et al., *Association of corticotropin-releasing hormone receptor-2 genetic variants with acute bronchodilator response in asthma.* Pharmacogenet Genomics, 2008. **18**(5): p. 373-82.

16. Tantisira, K.G., et al., *Genomewide association between GLCCI1 and response to glucocorticoid therapy in asthma.* N Engl J Med, 2011. **365**(13): p. 1173-83.

17. Litonjua, A.A., et al., *ARG1 is a novel bronchodilator response gene: screening and replication in four asthma cohorts.* Am J Respir Crit Care Med, 2008. **178**(7): p. 688-94.

18. Duan, Q.L., et al., *A polymorphism in the thyroid hormone receptor gene is associated with bronchodilator response in asthmatics.* Pharmacogenomics J, 2013. **13**(2): p. 130-6.

19. Duan, Q.L., et al., *A genome-wide association study of bronchodilator response in asthmatics.* Pharmacogenomics J, 2014. **14**(1): p. 41-7.

20. Reihsaus, E., et al., *Mutations in the gene encoding for the beta 2-adrenergic receptor in normal and asthmatic subjects.* Am J Respir Cell Mol Biol, 1993. **8**(3): p. 334-9.

21. Israel, E., et al., *Genome-wide association study of short-acting beta2-agonists. A novel genome-wide significant locus on chromosome 2 near ASB3.* Am J Respir Crit Care Med, 2015. **191**(5): p. 530-7.

22. Himes, B.E., et al., *Genome-wide association analysis in asthma subjects identifies SPATS2L as a novel bronchodilator response gene.* PLoS Genet, 2012. **8**(7): p. e1002824.

23. Torgerson, D.G., et al., *Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations.* Nat Genet, 2011. **43**(9): p. 887-92.

24. Galanter, J.M., et al., *Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study.* J Allergy Clin Immunol, 2014. **134**(2): p. 295-305.

25. White, M.J., et al., *Novel genetic risk factors for asthma in African American children: Precision Medicine and the SAGE II Study.* Immunogenetics, 2016. **68**(6-7): p. 391-400.

26. Winkler, C.A., G.W. Nelson, and M.W. Smith, *Admixture mapping comes of age.* Annu Rev Genomics Hum Genet, 2010. **11**: p. 65-89.

27. Nishimura, K.K., et al., *Early-life air pollution and asthma risk in minority children. The GALA II and SAGE II studies.* Am J Respir Crit Care Med, 2013. **188**(3): p. 309-18.

28. Torgerson, D.G., et al., *Case-control admixture mapping in Latino populations enriches for known asthma-associated genes.* J Allergy Clin Immunol, 2012. **130**(1): p. 76-82 e12.

29. Gould, W., et al., *Factors predicting inhaled corticosteroid responsiveness in African American patients with asthma.* J Allergy Clin Immunol, 2010. **126**(6): p. 1131-8.

30. Moore, W.C., et al., *Characterization of the severe asthma phenotype by the National Heart, Lung, and Blood Institute's Severe Asthma Research Program.* J Allergy Clin Immunol, 2007. **119**(2): p. 405-13.

31. Moore, W.C., et al., *Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program.* Am J Respir Crit Care Med, 2010. **181**(4): p. 315-23.

32. *Standardization of Spirometry, 1994 Update. American Thoracic Society.* Am J Respir Crit Care Med, 1995. **152**(3): p. 1107-36.

33. Delaneau, O., J.F. Zagury, and J. Marchini, *Improved whole-chromosome phasing for disease and population genetic studies.* Nat Methods, 2013. **10**(1): p. 5-6.

34. Howie, B., J. Marchini, and M. Stephens, *Genotype imputation with thousands of genomes.* G3 (Bethesda), 2011. **1**(6): p. 457-70.

35. Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.* PLoS Genet, 2009. **5**(6): p. e1000529.

36. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

37. Maples, B.K., et al., *RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference.* Am J Hum Genet, 2013. **93**(2): p. 278-88.

38. *Proceedings of the ATS workshop on refractory asthma: current understanding, recommendations, and unanswered questions. American Thoracic Society.* Am J Respir Crit Care Med, 2000. **162**(6): p. 2341-51.

39. Pino-Yanes, M., et al., *Genetic ancestry influences asthma susceptibility and lung function among Latinos.* J Allergy Clin Immunol, 2015. **135**(1): p. 228-35.

40. Baran, Y., et al., *Fast and accurate inference of local ancestry in Latino populations.* Bioinformatics, 2012. **28**(10): p. 1359-67.

41. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals.* Genome Res, 2009. **19**(9): p. 1655-64.

42. McGarry, M.E., et al., *Obesity and bronchodilator response in black and Hispanic children and adolescents with asthma.* Chest, 2015. **147**(6): p. 1591-8.

43. Willer, C.J., Y. Li, and G.R. Abecasis, *METAL: fast and efficient meta-analysis of genomewide association scans.* Bioinformatics, 2010. **26**(17): p. 2190-1.

44. PLUMMER M, B.N., COWLES K, VINES K, *CODA: Convergence Diagnosis and Output Analysis for MCMC.* R News, 2012(6): p. 7-11.

45. Sobota, R.S., et al., *Addressing population-specific multiple testing burdens in genetic association studies.* Ann Hum Genet, 2015. **79**(2): p. 136-47.

46. Marcus, J.H. and J. Novembre, *Visualizing the geography of genetic variants.* Bioinformatics, 2016.

47. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-5.

48. Li, Z., et al., *A stimulatory role for cGMP-dependent protein kinase in platelet activation.* Cell, 2003. **112**(1): p. 77-86.

49. Tamura, N., et al., *cDNA cloning and gene expression of human type Ialpha cGMP-dependent protein kinase.* Hypertension, 1996. **27**(3 Pt 2): p. 552-7.

50. Orstavik, S., et al., *Characterization of the human gene encoding the type I alpha and type I beta cGMP-dependent protein kinase (PRKG1).* Genomics, 1997. **42**(2): p. 311-8.

51. Burgoyne, J.R., et al., *Cysteine redox sensor in PKGIa enables oxidant-induced activation.* Science, 2007. **317**(5843): p. 1393-7.

52. Pfeifer, A., et al., *Defective smooth muscle regulation in cGMP kinase I-deficient mice.* EMBO J, 1998. **17**(11): p. 3045-51.

53. Dawes, M., P.J. Chowienczyk, and J.M. Ritter, *Effects of inhibition of the L-arginine/nitric oxide pathway on vasodilation caused by beta-adrenergic agonists in human forearm.* Circulation, 1997. **95**(9): p. 2293-7.

54. Boyle, A.P., et al., *Annotation of functional variation in personal genomes using RegulomeDB.* Genome Res, 2012. **22**(9): p. 1790-7.

55. Bustamante, C.D., E.G. Burchard, and F.M. De la Vega, *Genomics for the world.* Nature, 2011. **475**(7355): p. 163-5.

56. Popejoy, A.B. and S.M. Fullerton, *Genomics is failing on diversity.* Nature, 2016. **538**(7624): p. 161-164.

57. Editors, P.M., et al., *Towards Equity in Health: Researchers Take Stock.* PLoS Med, 2016. **13**(11): p. e1002186.

Chapter 3: Characterization of the recent demographic history and population structure of Mexican Americans in the United States

Introduction

Hispanics/Latinos living in the United States are culturally, phenotypically and genetically diverse populations. Individuals who identify as Hispanic/Latino have varying proportions of Native American, African, and European genetic ancestry, each with its own unique continental demographic history. Demographic forces such as population bottlenecks and expansions, migration and adaptation to novel environments resulted in observable differences in continental patterns of genetic variation [1-3]. These differing patterns were shaped by many historical events of migration which included the founding of the Americas by Native American populations, the colonization by Europeans, and the subsequent African slave trade [4-8]. These large scale migrations and additional demographic events shaped the genetic diversity of individuals living today within the United States [9-11].

Demographic history has shaped the genetic architecture of modern human phenotypic variation [12-17], and is critical to consider in the search for the genetic basis of complex diseases. The demography of the United States has changed drastically over the 20th century, and by 2044 is predicted to become a 'minority-majority' country whereby no one racial/ethnic group comprises more than 50% of the population [18]. By 2060 Hispanics/Latinos are projected to make up the largest of that share at 29% or 119 million individuals [18]. However, to date, population-based medical genomics research [and its subsequent benefits, including polygenic risk score (PRS) profiling] have been disproportionately focused on individuals of European descent, with the findings primarily benefiting these populations [19, 20]. Despite the increases in sample sizes, rates of discovery, and traits studied, Hispanic or Latin American participation in genome-wide association studies (GWAS) has continued to hover around 1% [21, 22]. This shows a continuing trend of leaving this population particularly vulnerable to falling behind in receiving the benefits of the precision medicine revolution [20, 21].

In this study we utilize the largest genetic study of Hispanics/Latinos in the U.S. to date -- the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) [10] -- to understand how patterns of genetic variation in Hispanic/Latino populations in the United States have changed over the last century.

Methods

Study dataset and initial quality control

The HCHS/SOL study is a community-based cohort study of self-identified Hispanic/Latino individuals from four US metropolitan areas with the general goal of identifying risk and protective factors for various medical conditions including cardiovascular disease, diabetes, pulmonary disease, and sleep disorders [23]. 12,434 participants with birth year estimates between 1934-1993 who self-identified as being of Cuban, Dominican, Puerto Rican, Mexican, Central American, or South American background consented to genetics studies and posting of their genetic and phenotype data on the publicly available Database of Genotypes and Phenotypes (dbGaP) through Study Accession phs000810.v1.p1. Samples were genotyped on an Illumina custom array, SoL HCHS Custom 15041502 array (annotation B3, genome build 37), consisting of the Illumina Omni 2.5M array and 148,353 custom single nucleotide polymorphisms (SNPs) [10]. Data posted to dbGaP had passed initial sample quality control filters, including removing samples with differences in reported vs. genetic sex, call rates > 95%, and evidence for sample contamination (e.g. heterozygosity and sample call rates). For initial SNP quality control, we filtered out SNPs that were monomorphic, positional duplicates, or Illumina technical failures, as well as SNPs that had cluster separation <= 0.3, call rate <=2%, >2 disconcordant calls in 291 duplicate samples, >3 Mendelian errors in parent-offspring pairs/trios, Hardy-Weinberg Equilibrium combined P-value $<10^{-5}$, and sex differences in allele frequency ≥0.2. Our filtering resulted in 1,763,935 genotyped SNPs with minor allele frequency (MAF) >0.01.

Additional sample quality control performed in the HCHS/SOL dataset included filtering out samples with 1) large chromosomal anomalies, 2) substantial Asian ancestry as previously identified in HCHS/SOL (*12*) and 3) individuals with up to third degree genetic relatedness in the dataset as inferred by REAP [24]. For genetic relatedness filtering, individuals from pairs were kept to maximize representation of the birth year distribution, which resulted in 10,268 unrelated remaining samples.

From the original HCHS/SOL analysis, individuals were classified into genetic-analysis groups, similar to self-identified background groups in that they share cultural and environmental characteristics, but are also more genetically homogenous [10]. Individuals that had been classified as "other" were excluded from any further analyses.

Birth year for all individuals was estimated by subtracting the difference between date of first clinic visit for the baseline examination [23] and age.

Global, local and parental ancestry inference

All ancestry analyses were restricted to the 211,152 autosomal SNP markers that overlapped between the study and reference panel genotyping array. For the HCHS/SOL dataset, global African, European, and Native American ancestries were inferred with ADMIXTURE, in an unsupervised manner, with K=3. HCHS/SOL samples with greater than 95% of either African, European, or Native American ancestry were filtered out resulting in 9,913 samples: 1,099 Central American, 1,536 Cuban, 954 Dominican, 3,622 Mexican, 1,783 Puerto Rican, 652 South American and 267 "Other" individuals. Ancestral tracts, known as 'local' ancestry, along the genome for all HCHS/SOL samples were inferred using RFMix [25] and a three population reference panel, comprised of 315 individuals: 104 HapMap phase 3 CEU (European) and 107 YRI (African) samples [26] and 112 Native American samples from throughout Latin America [8]. The reference panel was limited to samples with 99% continental ancestry as inferred by unsupervised ADMIXTURE [27]. Prior to local ancestry inference, HCHS/SOL samples were merged with the reference panels and then phased using SHAPEIT2 [28]. For all HCHS/SOL Mexican American samples, parental genomic ancestry was inferred with ANCESTOR [29] using the local ancestry estimates generated by RFMix.

Uniform Manifold Approximation and Projection (UMAP)

Principal components for HCHS/SOL and the reference panel were computed using smartPCA [30]. UMAP was run using the Python script freely available at https://github.com/diazale/gt-dimred with parameter specification set at 15 nearest neighbours and a minimum distance between points of 0.5.

Admixture mapping

Local ancestry estimates for 211,151 SNPs across the genome were used to perform admixture mapping in HCHS/SOL Mexican Americans to determine if younger individuals harbored excess Native American ancestry in certain regions of the genome. Admixture mapping was performed applying two different models: 1) a linear regression model with age as the dependent variable adjusting for global Native American ancestry, sampling weight and center and 2) a logistic regression model dividing the HCHS/SOL Mexican cohort in to an older vs younger generation with 1965 set as the dividing point while also adjusting for global Native American ancestry, sampling weight and center. The threshold for genome-wide significance, 1.38×10^{-4} was calculated using the empirical autoregression framework with the package *coda* in R to estimate the total number of ancestral blocks [31, 32].

Multiple Regression Model for Tract Lengths

The model: $\log(f) = \beta_0 + \beta_1 T + \beta_2 A + \beta_3 TA + \varepsilon$, where f is a matrix containing the proportion of lengths of all ancestral tracts across the genome for all 3622 Mexican American individuals, T the tract length bin and A decade of birth year bin, was used to test for an effect of birth decade on the proportion of Native American ancestral tract lengths.

Diversity Calculations

Subcontinental ancestry was assessed using the diversity measurements π and F_{ST}. π was calculated as the average number of pairwise genetic differences among all pairs of overlapping Native American ancestry tracts across individuals. F_{ST} was calculated as:

$$F_{ST} = (H_T - H_S) / H_T$$

where H_T was equal to the average heterozygosity among total populations and H_S was equal to the average heterozygosity within subpopulations.

Inference of Runs of Homozygosity

Runs of homozygosity (ROH) were called using the program GARLIC v1.1.4 [33] on 211,152 sites for the Mexican American individuals. An analysis window size of 50 SNPs and an overlap fraction of 0.25 were both chosen using GARLIC's rule of thumb parameter estimation. GARLIC chose a LOD score cutoff of 0. Using a three-component Gaussian mixture, GARLIC determined class A/B (short/medium) and class B/C (medium/long) size boundaries as 845,097 bp and 2,501,750 bp, respectively.

Health and Retirement Study (HRS)

For replication, we used genotype data from 705 self-identified Mexican-Americans from the Health and Retirement Study (HRS) [34], genotyped on the Illumina Human Omni 2.5M platform. HRS data was made available under IRB Study No. A11-E91-13B - The apportionment of genetic diversity within the United States. Estimated global ancestry proportions for the Mexican American population in the HRS were calculated as in Baharian et al. [35], which used an alternative reference panel and alternative ancestry inference approach. Briefly, RFMix was used to infer local ancestry estimates across the genome utilizing CHS, YRI, and CEU individuals from the 1000 Genomes Project as reference populations for Native American/Asian, African, and

European ancestries, respectively. Global ancestry estimates were calculated using the summed RFMix calls.

Statistical Analyses and Plots

Statistical analyses and plot generation were performed within Rstudio using Version 1.1.463 and R version 3.5.3. ternary and ggridges/ggplot2 packages were used to create the simplex and ridgeline plots.

Results

Global ancestry proportions among HCHS/SOL Hispanic/Latino Populations

Using the subset of sites that overlapped with our African, European, and Native American reference panels, we called 3-way global ancestry estimates for 10,268 unrelated HCHS/SOL individuals (see methods). Figure 3.1A summarizes the global ancestry proportions shaded by admixture estimates in a ternary plot, recapitulating the original HCHS/SOL analysis of continental ancestry [10]. However, while several population groups appear to have overlapping ancestry proportions, this analysis masks more subtle structure in subcontinental ancestry. To investigate subtle population structure across these self-identified population groups, we performed UMAP on the top 3 principal components (see methods), and find substantial structure across self-identified groups (Figure 3.1B and Figure B.1B). We find that Dominicans, who have the highest average proportions of African ancestry, are in the middle, with Puerto Ricans and Cubans, diverging in opposite directions (Figure B.1B) with clines of increasing European ancestry proportions (Figure 3.1B). Further, while self-identified Mexican, Central, and South American groups appear to have overlapping ancestry proportions in Figure 3.1A, UMAP reveals that Mexican Americans and Central/South American groups form large, separate wings that diverge from self-identified Cubans and Dominicans, with both clusters diverging with clines of increasing ancestry toward different Native American populations (Figure 3.1B and Figure B.1B).

Dynamic Global Ancestry Proportions in Mexican Americans

For each of the HCHS/SOL Latino populations, we evaluated differences in global ancestry estimates over time while accounting for the sampling method (referred to as “sampling weight”, see methods) used for the design of the HCHS/SOL study [23]. We found that in all populations, the effect size for Native American ancestry on birth year is positive, though only statistically significant after multiple testing in the Mexican American (β =0.0023; P=3.58x10^{-22}; Figure 3.1C) and Central American (β =0.0013; P=0.0013) cohorts (Table B.1). Due to the larger sample size, magnitude of the effect, and statistical significance, we shift our focus to Mexican Americans. In Mexican Americans, the increase in Native American global ancestry over time was consistent across multiple data stratifications including recruitment region, US born or not US born, educational attainment, and sex (Table 3.1), and was robust to alternative methods for estimating global ancestry proportions (e.g. based on the summation of RFMix local ancestry estimates; Figures B.2 and B.3).

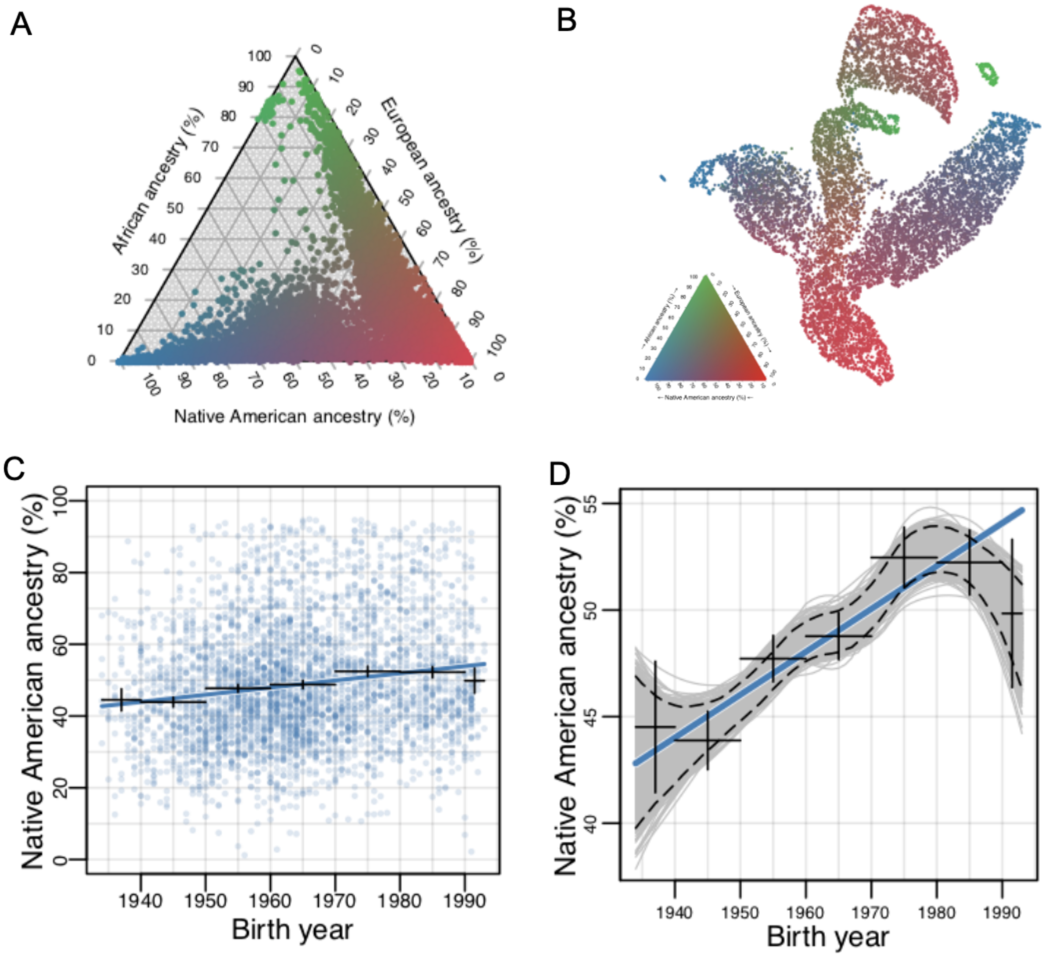

Figure 3.1. Recent dynamics continually shape the continuum of continental ancestry Hispanic/Latino populations. A. Ternary plot of HCHS/SOL (n=10,268) colored by admixture proportions. B. Uniform Manifold Approximation and Projection (UMAP) plot depicting the genetic diversity of HCHS/SOL and the reference panel (n=10,591) using 3 principal components, colored by admixture proportions (see Supplemental Fig 1 for population labels). C. Global Native American ancestry proportions plotted by birth year for Mexican Americans (n=3,622). Fitted line is multiple regression of Native American ~ birth year + sampling weight. Bars represents 95% confidence intervals for individuals grouped by decade. D. Bootstrap resampling (n=1000 iterations) of Native American global ancestry for the Mexican American individuals with a fitted loess regression line for each iteration. Dashed lines represent the 95% confidence interval and the blue line represents the fitted regression line from Figure 1C.

Table 3.1. Relationship of Native American global ancestry and birth year for Mexican Americans stratified by recruitment region, US born vs migrant status, sex and educational attainment. For recruitment region, data stratification was limited to Chicago and San Diego as sample size for the Bronx and Miami was limited: 124 and 25 individuals, respectively. Education attainment was categorized as either less than a high school diploma or equivalent degree (<HS), equal to a high school diploma or equivalent degree (=HS), or post secondary education (>HS).

Category	N	Mean	Median	R2	Effect	Std.Err	P
All	3622	0.489	0.468	0.027	0.0023	0.0002	3.58×10^{-22}
Chicago	1310	0.562	0.550	0.017	0.0016	0.0005	0.0006
San Diego	2163	0.428	0.422	0.012	0.0012	0.0002	4.29×10^{-7}
US born	634	0.427	0.418	0.063	0.0027	0.0004	1.77×10^{-10}
Not US born	2987	0.502	0.481	0.050	0.0032	0.0003	1.38×10^{-30}
Male	1500	0.494	0.475	0.038	0.0028	0.0004	3.83×10^{-14}
Female	2122	0.485	0.462	0.022	0.0019	0.0003	3.07×10^{-10}
<HS	1518	0.520	0.500	0.045	0.0026	0.0004	1.39×10^{-12}
=HS	960	0.501	0.479	0.022	0.0018	0.0005	0.0003
>HS	1140	0.436	0.422	0.045	0.0027	0.0004	6.53×10^{-13}

We performed bootstrap resampling (n=1000) of global Native American ancestry for the Mexican Americans and observed a consistent increase in Native American ancestry with fitted loess smoothing (Figure 3.1D) and when individuals were binned by birth year decades (Figure B.4). On average, global Native American ancestry has increased ~20% over the last 50 years in Mexican Americans.

We replicated our original findings of the increase in global Native American ancestry over time in a smaller, separate cohort of self-identified Mexican Americans (n=705) from the Health and Retirement Study (HRS) [34]. The HRS Mexican Americans in this study are older compared to the HCHS/SOL Mexican Americans (birth year distribution: 1915-1981; mean=1943, median:1942) and have lower levels of global Native American ancestry on average (mean=0.29), but we still observed an increase in global Native American ancestry over time ((β =0.00082; P=0.02;SE=0.0003673) (Figure B.5A). We performed 1000 bootstrap resampling iterations of the linear regression model (global Native American ancestry ~ birth year) fitted to the data. From these resampling iterations, the average β=0.00083 and 61.5% of the regression p-values were less than 0.05 as illustrated in (Figures B.5B-B.5D).

Highlighting the diversity within ancestry tracts

Native American ancestry lengths

We next sought to test whether differences at the local ancestry level could explain the shift in global Native American ancestry over time in the Mexican Americans. We calculated the length of each RFMix inferred local ancestry tract in each Mexican American individual, and tested for differences in the distribution of tract lengths across birth-decade using a multiple linear regression model (see methods). We found no significant associations between the decade bin and the proportion of Native American ancestral tracts at various lengths (Figure 3.2A), even when testing for violations of model assumptions (e.g. normalizing the tracts per bin by the number of individuals, or excluding the 1930s and/or 1990s individuals due to the small sample size in each bin).

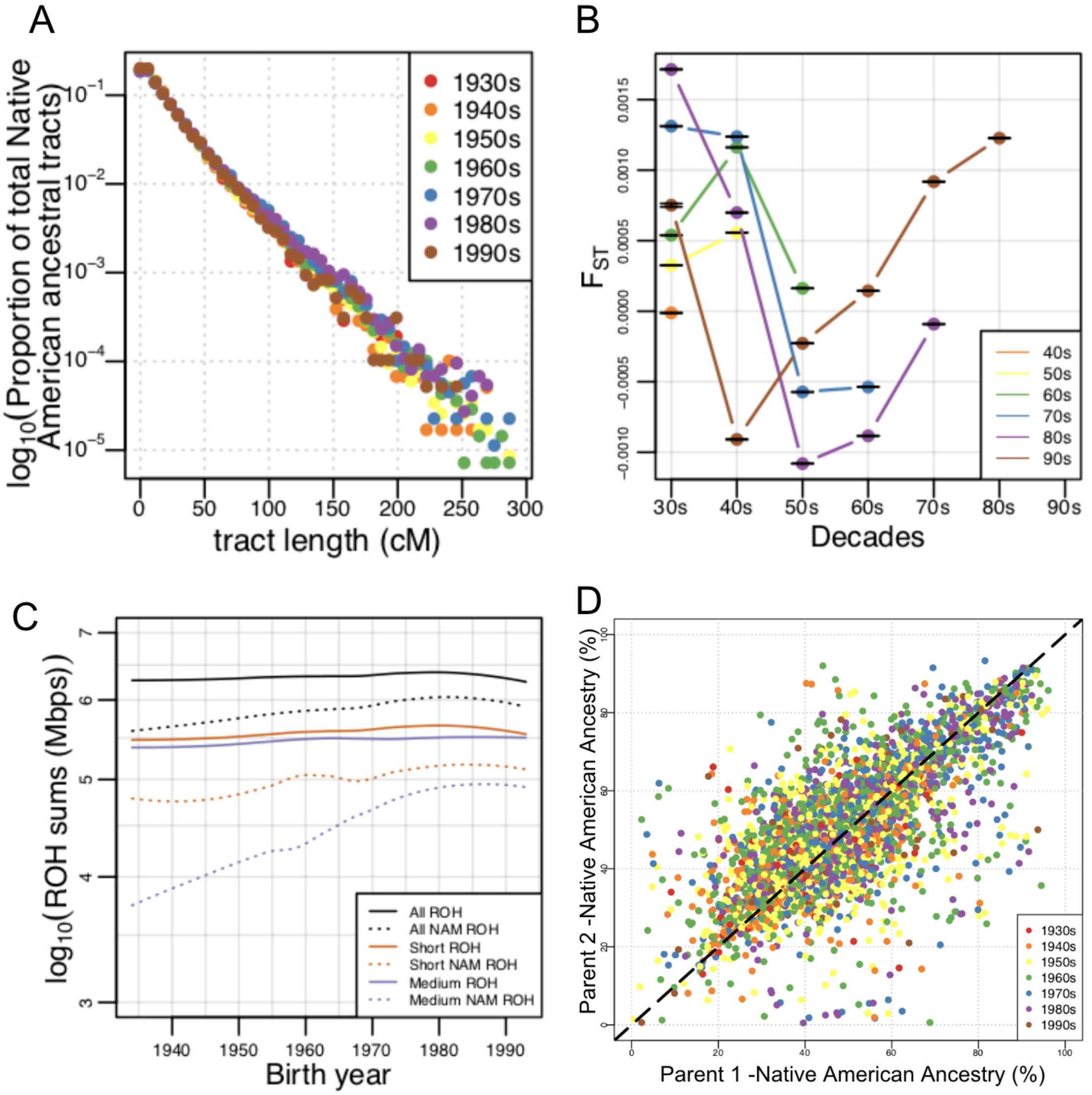

Figure 3.2. Diversity of and within Native American ancestral tracts. A) Proportion of total Native American ancestral tracts in the HCHS/SOL Mexican American population by decade. B) F_{ST} estimates calculated between each decade group. Bars represent the 95% CI. C) Loess regression of the log of the sum of total ROH and ROH overlapping Native American ancestral tracts separated by ROH class. Total long ROH is not represented as an individual group due to the high number of individuals missing long ROH (1694 for long ROH across ancestries and 1987 for long NAM ROH) but was included in the sum of "All ROH" and "All NAM ROH". D) Correlation of parent's inferred global Native American ancestries using ANCESTOR.

Grouping tracts by decade did not reveal any significant effects of birth year on Native American ancestral tract distribution, so we next investigated the individual level. We tested for a relationship between birth year and the proportion of long tracts per individuals and we found the two to be correlated (τ=0.092, P=2.91x10^{-8}, Kendall's rank correlation). This correlation was consistent when we tested different starting cutoffs of "long" tracts beginning with 50cM and ranging to 90 cM. These cutoffs were chosen based on tract separation between the birth year decades in Figure 3.2A.

Subcontinental ancestry

It is possible that the increase in global Native American ancestry over time could be biased by changes in the specific subcontinental Native American ancestries over time (though such an effect is not visible in our UMAP analysis, Figure 3.1B). If it were the case, then we would expect subtle signals of genetic divergence in Native American ancestry tracts over time. To investigate this, we calculated F_{ST} between all pairs of birth-decades (see methods). Figure 3.2B shows all pairwise comparisons among birth-decades, and demonstrates that while the estimates of F_{ST} are negligible (with many estimates below 0), there is a subtle trend of increasing F_{ST} as birth-decade differences increase (though individuals born in the 80s and 90s show a conflicting pattern). We further investigated this pattern using genetic diversity, π, in Native American ancestry tracts for each birth-decade (see methods). We hypothesized that if there were increased migration from multiple Native American source populations (coupled with rapid population growth in Mexican American communities), then genetic diversity should be increasing over time. We found the opposite: Figure B.6 shows a subtle decrease in genetic diversity (π) over time from the 1930s to the 1980s in non-US born Mexican Americans, and a subtle decrease in US born Mexican Americans from the 70s to the 90s (while remaining roughly constant from the 30s to the 70s).

Runs of homozygosity

Since genetic diversity has decreased over time in the Native American ancestry tracts of Mexican Americans (despite rapid population growth), it is possible that this population has also experienced increased haplotype homozygosity over time. We investigated this possibility by exploring runs of homozygosity (ROH) in Native American ancestry tracts in each of the 3622 Mexican Americans. We classified ROH into three categories: short, medium, and long, based on the length distribution in the population. Generally, short ROH are tens of kilobases in length and likely reflect the homozygosity of old haplotypes; medium ROH are hundreds of kilobases in length and likely reflect background relatedness in the population; and long ROH are hundreds of kilobases to several megabases in length and are likely the result of recent parental relatedness. Figure 3.2C shows a fitted loess curve to the log of the total length of ROH summed across each Mexican American's genome as a function of their birth year, broken down by ROH size class (as well as the total of each size class that overlaps Native American ancestry tracts; see also Figure B.7A). We identified a significant positive correlation between birth year and the total summed ROH across size classes ($P=6.115x10^{-5}$, $\boldsymbol{\tau}=0.0449$, Kendall's rank correlation). When stratified by size class, the associations (all Kendall's rank correlation) in ROH was primarily driven by the short ($P=9.449x10^{-14}$,$\boldsymbol{\tau}=0.0833$), and medium ($P=1.46x10^{-10}$,$\boldsymbol{\tau}=0.0718$) size classes. The long ROH had a negative correlation with birth year, but was insignificant after multiple testing ($P=0.01499$, $\boldsymbol{\tau}=-0.0291$; not that 1694 individuals did not have any long ROH calls in their genome). These correlations are much stronger and more significant when we restricted ROH calls to regions that overlapped with Native American ancestral tracts - total summed Native American ROH of all size classes: $p=9.457x10^{-15}$, $\boldsymbol{\tau}=0.0873$, total summed short ROH ($P<2.2x10^{-16}$, $\boldsymbol{\tau}=0.107$), total summed medium ROH ($P<2.2x10^{-16}$, $\boldsymbol{\tau}=0.1003$), and again there was no significant association between long Native American ROH calls and birth year.

Strong ancestry-related assortative mating in HCHS/SOL Mexicans

Given that short and medium length ROH have increased over time, it appears that background relatedness within Native American ancestry in Mexican Americans has increased over time. One way for this to occur is if individuals with similar ancestry patterns tend to mate with one another more often than expected under a model of random mating (i.e. assortative mating). To measure assortative mating, we estimated the ancestral proportions of the mother and father of each HCHS/SOL Mexican American (see methods). With individuals from all decades pooled together, we found the inferred parental Native American ancestries to be significantly correlated (Figure 3.2D, r=0.708, 95% CI:0.69-0.72, P<2.2x10^{-16} Pearson correlation). Stratified by decade, the correlation in inferred parental Native American global ancestry ranged from 0.65 to 0.74 (Figure B.8). This shows that there was strong spousal ancestry correlation in the Mexican Americans over different generations. However, since there is no trend in long ROH with birth year (and an overall low rate of long ROH among Mexican Americans), this signature of assortative mating is not due to recent parental relatedness.

Admixture mapping

We used local ancestry estimates generated across the genome to perform admixture mapping in HCHS/SOL Mexican Americans to determine if younger individuals harbored excess Native American ancestry in certain regions of the genome. Although we tried two different models (see methods), we did not find any loci to be significantly associated with birth year across the genome (Figure B.9).

Discussion

The United States is a dynamic, rapidly changing population, and this will continue to occur as the population size grows [18]. Hispanics/Latinos are the largest and fastest growing minority group, and are projected to comprise over 25% of the US population by 2060. They are a genetically and phenotypically diverse population as a result of extensive admixture between Native Americans and immigrants from multiple geographic locations around the world. In this study, we identified additional population substructure complexities that may contribute to phenotypic variation within Hispanics/Latinos.

Specifically, we demonstrated how the admixture dynamics of Mexican Americans have changed over time, resulting in an increase of ~20% Native American ancestry on average over the 50 year period studied. This change in ancestry is equivalent to a mean increase in Native American ancestry of ~0.4% per year. While the effect sizes vary to some extent, we replicate the underlying pattern across multiple data stratifications (two metropolitan cities, US born and non-US born) and also replicate this feature in an independent cohort of Mexican Americans. Further, we find that a similar trend holds across multiple self-identified Hispanic/Latino populations in the US (and is statistically significant in Central Americans). This effect does not appear to have a simple explanation: we do not see any statistically significant increases at individual loci, we do not see a strong signature of increased migration, and we do not see more than a negligible degree of population differentiation over time. We do, however, find that as Native American ancestry has increased, genetic diversity within Native American ancestry tracts across Mexican Americans has decreased over time, and is associated with increased short and medium length ROH over time. This suggested that there could be increased relatedness within Native American ancestries within Mexican Americans, and we confirmed that there is a very high degree ancestry-based assortative mating within the Mexican American population.

What could be driving the increased Native American ancestry in Mexican Americans? Population genetic theory suggests that while assortative mating could result in increased ROH and decreased genetic diversity, ancestry-based assortative mating alone should not result in mean changes in global ancestry proportions. Conceptually, in the absence of fecundity differences, reproduction among individuals with high Native American ancestry should be balanced by reproduction among individuals with low Native American ancestry.

While we have shown a dramatic shift in ancestry proportions in US Hispanic/Latinos, one of the caveats of this study is that the HCHS/SOL cohort is not representative of all US Hispanics/Latinos. HCHS/SOL participants were recruited at four primary centers: Bronx, Chicago, Miami, and San Diego. There may be additional genetic diversity that has not been captured by this dataset and trends exhibited in this dataset may not translate to Hispanic/Latino populations living in other regions of the US. With better population genetic modeling we will be able to improve our understanding of the genetic diversity within Hispanic/Latino populations, its effects on human phenotypic diversity, and subsequently our ability to understand genetic contributions to complex traits and disease. These insights will lead to optimization of population sampling for the design of future medical genetic studies, the identification of disease risk variants, and ultimately, precision medicine for all.

References:

1. Nelson, M.R., et al., *The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research.* Am J Hum Genet, 2008. **83**(3): p. 347-58.
2. Genomes Project, C., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
3. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
4. Gravel, S., et al., *Reconstructing Native American migrations from whole-genome and whole-exome data.* PLoS Genet, 2013. **9**(12): p. e1004023.
5. Homburger, J.R., et al., *Genomic Insights into the Ancestry and Demographic History of South America.* PLoS Genet, 2015. **11**(12): p. e1005602.
6. Moreno-Estrada, A., et al., *Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits.* Science, 2014. **344**(6189): p. 1280-5.
7. Moreno-Estrada, A., et al., *Reconstructing the population genetic history of the Caribbean.* PLoS Genet, 2013. **9**(11): p. e1003925.
8. Reich, D., et al., *Reconstructing Native American population history.* Nature, 2012. **488**(7411): p. 370-4.
9. Bryc, K., et al., *The genetic ancestry of African Americans, Latinos, and European Americans across the United States.* Am J Hum Genet, 2015. **96**(1): p. 37-53.
10. Conomos, M.P., et al., *Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos.* Am J Hum Genet, 2016. **98**(1): p. 165-84.
11. Han, E., et al., *Clustering of 770,000 genomes reveals post-colonial population structure of North America.* Nat Commun, 2017. **8**: p. 14238.

12. Agarwala, V., et al., *Evaluating empirical bounds on complex disease genetic architecture.* Nat Genet, 2013. **45**(12): p. 1418-27.

13. Eyre-Walker, A., *Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies.* Proc Natl Acad Sci U S A, 2010. **107 Suppl 1**: p. 1752-6.

14. Maher, M.C., et al., *Population genetics of rare variants and complex diseases.* Hum Hered, 2012. **74**(3-4): p. 118-28.

15. Simons, Y.B., et al., *The deleterious mutation load is insensitive to recent population history.* Nat Genet, 2014. **46**(3): p. 220-4.

16. Uricchio, L.H., et al., *Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants.* Genome Res, 2016. **26**(7): p. 863-73.

17. Yang, J., et al., *Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index.* Nat Genet, 2015. **47**(10): p. 1114-20.

18. Colby, S.L.O., J.M., *Projections of the Size and Compositon of the U.S. Population: 2014 to 2060*. 2015, U.S. Census Bureau.

19. Bustamante, C.D., E.G. Burchard, and F.M. De la Vega, *Genomics for the world.* Nature, 2011. **475**(7355): p. 163-5.

20. Martin, A.R., et al., *Clinical use of current polygenic risk scores may exacerbate health disparities.* Nat Genet, 2019. **51**(4): p. 584-591.

21. Popejoy, A.B. and S.M. Fullerton, *Genomics is failing on diversity.* Nature, 2016. **538**(7624): p. 161-164.

22. Mills, M.C. and C. Rahal, *A scientometric review of genome-wide association studies.* Commun Biol, 2019. **2**: p. 9.

23. Sorlie, P.D., et al., *Design and implementation of the Hispanic Community Health Study/Study of Latinos.* Ann Epidemiol, 2010. **20**(8): p. 629-41.

24. Thornton, T., et al., *Estimating kinship in admixed populations.* Am J Hum Genet, 2012. **91**(1): p. 122-38.

25. Maples, B.K., et al., *RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference.* Am J Hum Genet, 2013. **93**(2): p. 278-88.

26. International HapMap, C., et al., *Integrating common and rare genetic variation in diverse human populations.* Nature, 2010. **467**(7311): p. 52-8.

27. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals.* Genome Res, 2009. **19**(9): p. 1655-64.

28. Delaneau, O., J.F. Zagury, and J. Marchini, *Improved whole-chromosome phasing for disease and population genetic studies.* Nat Methods, 2013. **10**(1): p. 5-6.

29. Zou, J.Y., et al., *Inferring parental genomic ancestries using pooled semi-Markov processes.* Bioinformatics, 2015. **31**(12): p. i190-6.

30. Patterson, N., A.L. Price, and D. Reich, *Population structure and eigenanalysis.* PLoS Genet, 2006. **2**(12): p. e190.

31. PLUMMER M, B.N., COWLES K, VINES K, *CODA: Convergence Diagnosis and Output Analysis for MCMC.* R News, 2012(6): p. 7-11.

32. Sobota, R.S., et al., *Addressing population-specific multiple testing burdens in genetic association studies.* Ann Hum Genet, 2015. **79**(2): p. 136-47.

33. Szpiech, Z.A., A. Blant, and T.J. Pemberton, *GARLIC: Genomic Autozygosity Regions Likelihood-based Inference and Classification.* Bioinformatics, 2017. **33**(13): p. 2059-2062.

34. Fisher, G.G. and L.H. Ryan, *Overview of the Health and Retirement Study and Introduction to the Special Issue.* Work Aging Retire, 2018. **4**(1): p. 1-9.

35. Baharian, S., et al., *The Great Migration and African-American Genomic Diversity.* PLoS Genet, 2016. **12**(5): p. e1006059.

36. Collaboration, N.C.D.R.F., *A century of trends in adult human height.* Elife, 2016. **5**.

37. Yengo, L., et al., *Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry.* Hum Mol Genet, 2018. **27**(20): p. 3641-3649.

38. Pasaniuc, B. and A.L. Price, *Dissecting the genetics of complex traits using summary association statistics.* Nat Rev Genet, 2017. **18**(2): p. 117-127.

39. Maas, P., et al., *Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States.* JAMA Oncol, 2016. **2**(10): p. 1295-1302.

40. Schumacher, F.R., et al., *Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci.* Nat Genet, 2018. **50**(7): p. 928-936.

41. Sharp, S.A., et al., *Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis.* Diabetes Care, 2019. **42**(2): p. 200-207.

42. Martin, A.R., et al., *Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations.* Am J Hum Genet, 2017. **100**(4): p. 635-649.

Chapter 4: Leveraging the genetic ancestry of Mexican Americans to understand the genetic architecture of complex traits

Introduction

Genome-wide association studies (GWAS) have been the primary analyses performed to identify risk variants across a wide range of disease phenotypes [1]. Despite the thousands of GWAS studies and millions of individuals that have participated in these studies, representation of Hispanics/Latinos has continued to hover around 1% [2] . Within studies performed in Latinos, associations with Native American ancestry, with baseline lung function [3], asthma [4] and gallbladder cancer [5] have been identified, solidifying the importance of inferring population structure to understand disease etiology.

Results for many GWAS have been made readily available on public databases as summary association statistics that can be leveraged to build predictions of genetic risk through polygenic risk scores (PRS) [6]. Combined with other risk factors, researchers and clinicians have been able to demonstrate their use through PRS informed therapeutic intervention, disease screening and life planning [6].

One of the classic polygenic traits, height, has provided insights into what the genetic architecture of common human traits and diseases might look like, as well as given us insights into the prospects and challenges of different methods used to identify genetic risk factors [7, 8]. However, as mentioned, the vast majority of these studies have been performed in individuals with European ancestry and a previous study identified that human demographic history impacts genetic risk prediction across diverse populations [9] justifying the need to determine how well GWAS findings are going to translate to Hispanic/Latino populations.

In this study we assess the relationship between global Native American genetic ancestry and various complex traits for 3622 Mexican Americans from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) [10] -- to understand the role of Native American ancestry in the genetic architecture of complex traits. We further investigate the predictive power of polygenic risk scores for height in this cohort.

Methods

Imputation

Imputation for 3622 HCHS/SOL Mexican Americans was performed using IMPUTE2 [11] with the 1000 Genomes Project Phase 3 haplotypes used as a reference panel [12]. After filtering on an info score cutoff of 0.3, this resulted in 33,041,084 SNPs.

Polygenic Risk Score Calculations

Polygenic risk scores for height were calculated using the publicly available UK Biobank (UKBB) GWAS Round 2 Summary Statistics retrieved from http://www.nealelab.is/uk-biobank. Briefly, for sample quality control, sample inclusion was limited to unrelated samples who passed the sex chromosome aneulploidy filter. British ancestry was determined using the 1st 6 PCs; individuals more than 7 standard deviations away from the 1st 6 PCs were excluded. Further filtering included limiting to self -reported 'white-British' / 'Irish' / 'White' resulting in a QCed sample count of 361,194 samples https://github.com/Nealelab/UK_Biobank_GWAS#imputed-v3-sample-qc. An imputation panel of ~90 million SNPs from HRC, UK10K and 1KG were used to impute genotypes. 13.7 million autosomal and X-chromosome SNPs passed quality control thresholds including Info score>0.8, MAF>0.0001> HWE p-value>1e-10. For the phenotype, a linear regression model in Hail (linreg) was run for all individuals (both sexes) adjusting by the first 20 PCs + sex + age + age^2 + (sex*age) + (sex*age)^2. For height, there was complete phenotype information for 360,388 samples.

Risk scores were calculated by extracting the overlapping genome-wide significant hits initially discovered in the UKBB GWASs of height and selecting SNPs with the lowest p-value in each 1Mb window across the genome. For height this resulted in a dataset of 1,103 overlapping SNPs that were present in our dataset of genotyped and imputed SNPs.

Results

Genetic correlation of global Native American traits with biomedical traits

To further our understanding of the genetic architecture of complex traits in Mexican Americans, we investigated the relationship between Native American ancestry and various complex traits. Specifically, we tested for a correlation (Kendall's) between 66 complex traits from the HCHS/SOL phenotypic dataset and global Native American ancestry. As illustrated in Figure 4.1, we identified many of these traits to be significantly correlated (P<0.00076) after Bonferroni correction to account for multiple tests. We further investigated these traits using multiple regression to account for age, sex, center, and the sampling weight, and the effect of global Native American ancestry on many of these phenotypes persisted (Table C.1), highlighting the need for increased investigation into the role of Native American genetic ancestry in admixed populations such as Mexican Americans.

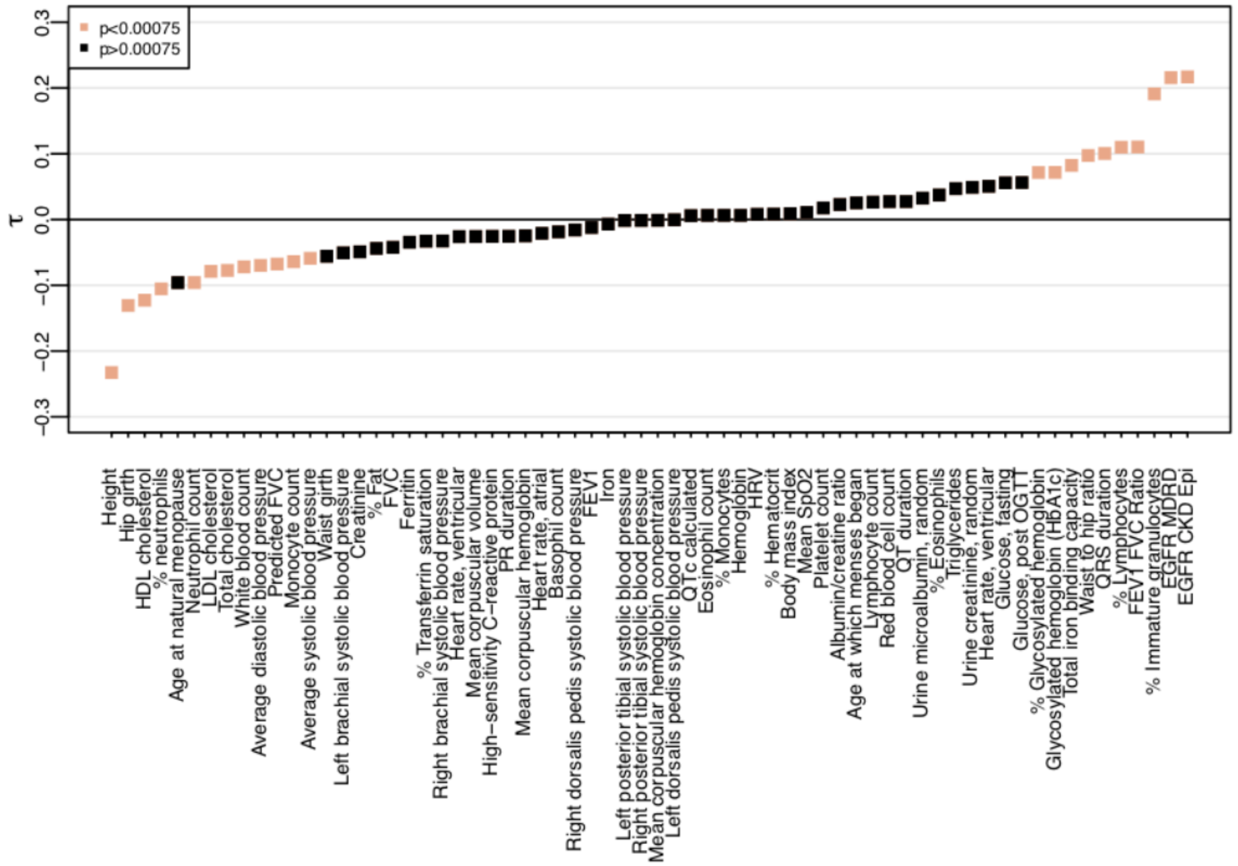

Figure 4.1. Correlation of 66 quantitative traits with global Native American ancestry. Significance level was determined using Bonferroni correction adjusting by the number of quantitative traits tested (0.05/66=0.00075).

Assessing the genetic contribution of Native American ancestry to height

Of the traits we tested for a correlation with global Native American ancestry, height had the strongest negative correlation, and our regression model indicated that height also had a strong positive relationship with birth year (reference plot and table again). Globally, populations have grown taller over time due to a variety of non-genetic, environmental factors [13]. We find a similar trend in the HCHS/SOL Mexican Americans (Figure 4.2A). Indeed, when we stratified individuals by quartiles of global Native American ancestry, we see that all quartiles have increased in height by a similar amount over the period investigated (though individuals with lower Native American ancestry were taller on average) (Table C.2).

Height is one of the most highly studied complex traits, with GWAS sample sizes numbering in the hundreds of thousands [14]. Results for many of these studies have been made readily available on public databases as summary association statistics that can be leveraged to build genetic predictions through polygenic risk scores (PRS) [6]. In Europeans, PRS have been shown to have great predictive power for several traits, including breast cancer, prostate cancer and type 1 diabetes [15-18]. PRS are most effective in populations of European descent as GWAS studies have been primarily performed in these populations [2, 15, 19] and are expected to be biased when applied to other populations due to differences in the genetic architecture of traits across diverse populations [9]. Since Mexican Americans have some fraction of European ancestry, we sought to determine whether PRS calculated utilizing GWAS summary statistics from European populations could still provide useful insight.

To evaluate the effectiveness of PRS for height (see methods), we first tested whether there was an association between the observed height and the predicted height estimates while controlling for sampling weight, sex, and recruitment center (see methods). We identified a significant association between observed height and predicted height for the population as a whole (β=0.0044881, P=2.19x10^{-12}; Figure 4.2B, Table C.3). However, when we stratified by quartiles of Native American global ancestry, the association only remained for the individuals in the lower two quartiles of global Native American ancestry (NAM<0.37: β=0.004, P=0.0008 and 0.36<NAM<0.46: β=0.004, P=0.003, Table C.3). The association between predicted height and observed height was no longer significant for individuals in the highest two quartiles of global Native American ancestry proportions (0.46<NAM<0.58 or 0.58<NAM, Table C.3).

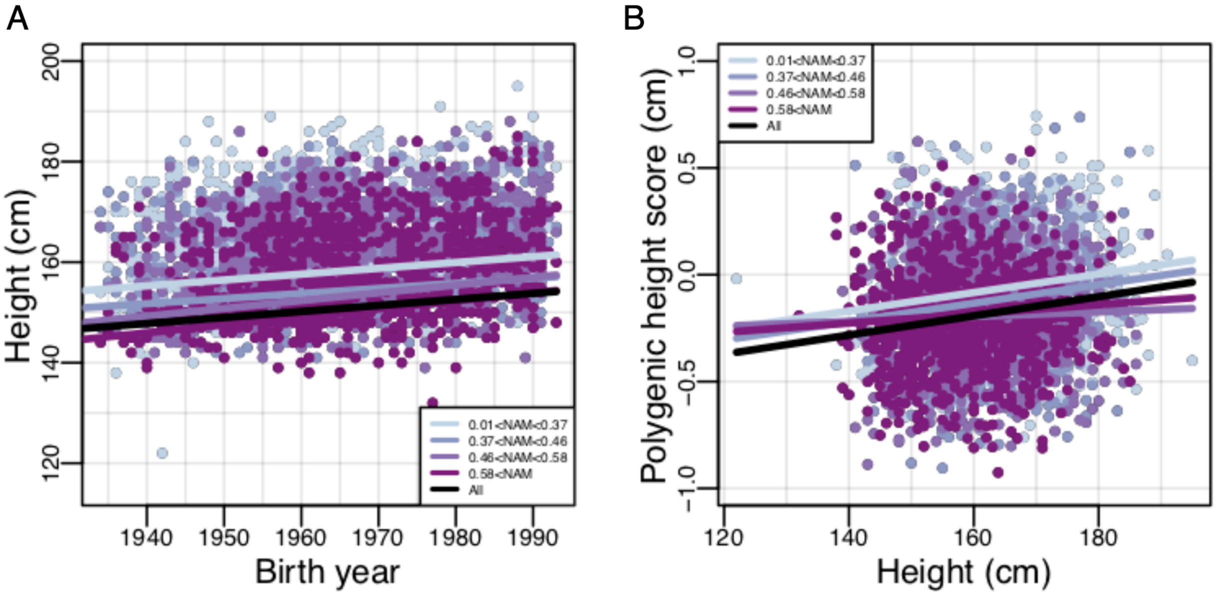

Figure 4.2: Height and global Native American ancestry in HCHS/SOL Mexican Americans. Each plot illustrates the relationship between A) Birth year and height B) Height and polygenic height score. The black line indicates the fitted linear model for all individuals. Each of the colors represents a different quartile of Native American global ancestry. Polygenic height scores were assessed utilizing UKBB summary statistics for 1,128 SNPs.

Discussion

We identified several biomedical traits that are correlated with Native American ancestry, and show that in the case of height, there are both ancestry and temporal effects. Further study is necessary to understand whether other biomedical traits are also changing over time as the genomic ancestry proportions change in this population.

In our study, we bring specific attention to the biases that continue to exist with using European GWAS summary statistics to calculate polygenic risk scores in admixed populations such as Mexican Americans that are comprised of European, Native American, and African genetic ancestries. In particular, in the case of height, we found that the PRS correlated with observed height only in the subset of individuals with the lowest levels of Native American ancestry (i.e. the subset of individuals with highest European ancestry). As the population dynamics of the US continue to change, it is imperative that we study diverse populations, or we risk exacerbating the health disparities that currently exist. To date, population-based medical genomics research (and its subsequent benefits) have been disproportionately focused on populations of European ancestry. In order to improve the design and implementation of medical genetics studies for the ethnically diverse U.S. population, we need detailed insights into the population history of diverse U.S. populations. This includes characterizing the admixture dynamics of Hispanic/Latino populations, as well as the evolutionary forces that shaped patterns of genetic variation of the ancestral populations that contributed to modern day Hispanic/Latino populations.

References

1. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.
2. Popejoy, A.B. and S.M. Fullerton, *Genomics is failing on diversity.* Nature, 2016. **538**(7624): p. 161-164.
3. Moreno-Estrada, A., et al., *Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits.* Science, 2014. **344**(6189): p. 1280-5.
4. Pino-Yanes, M., et al., *Genetic ancestry influences asthma susceptibility and lung function among Latinos.* J Allergy Clin Immunol, 2015. **135**(1): p. 228-35.
5. Lorenzo Bermejo, J., et al., *Subtypes of Native American ancestry and leading causes of death: Mapuche ancestry-specific associations with gallbladder cancer risk in Chile.* PLoS Genet, 2017. **13**(5): p. e1006756.
6. Pasaniuc, B. and A.L. Price, *Dissecting the genetics of complex traits using summary association statistics.* Nat Rev Genet, 2017. **18**(2): p. 117-127.
7. Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**(7538): p. 197-206.
8. Wood, A.R., et al., *Defining the role of common variation in the genomic and biological architecture of adult human height.* Nat Genet, 2014. **46**(11): p. 1173-86.
9. Martin, A.R., et al., *Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations.* Am J Hum Genet, 2017. **100**(4): p. 635-649.
10. Conomos, M.P., et al., *Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos.* Am J Hum Genet, 2016. **98**(1): p. 165-84.

11. Marchini, J. and B. Howie, *Genotype imputation for genome-wide association studies.* Nat Rev Genet, 2010. **11**(7): p. 499-511.

12. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

13. Collaboration, N.C.D.R.F., *A century of trends in adult human height.* Elife, 2016. **5**.

14. Yengo, L., et al., *Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry.* Hum Mol Genet, 2018. **27**(20): p. 3641-3649.

15. Martin, A.R., et al., *Clinical use of current polygenic risk scores may exacerbate health disparities.* Nat Genet, 2019. **51**(4): p. 584-591.

16. Maas, P., et al., *Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States.* JAMA Oncol, 2016. **2**(10): p. 1295-1302.

17. Schumacher, F.R., et al., *Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci.* Nat Genet, 2018. **50**(7): p. 928-936.

18. Sharp, S.A., et al., *Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis.* Diabetes Care, 2019. **42**(2): p. 200-207.

19. Bustamante, C.D., E.G. Burchard, and F.M. De la Vega, *Genomics for the world.* Nature, 2011. **475**(7355): p. 163-5.

APPENDIX A: Supplementary Material to Chapter 2

Additional information on medication use and medication withholding prior to spirometry testing:

Most of the participants in the three studies used for discovery were taking medications for their asthma symptoms (100%, 87%, and 91% in SAGE I, SAGE II, and GALA II, respectively). Regarding the use of ICS, 68%, 53%, and 66% were taking ICS alone or as part of a combo medication containing it in SAGE I, SAGE II, and GALA II, respectively. These patients were the ones that had persistent asthma.

Prior to spirometry testing, subjects were instructed to withhold the use of the following medications for the indicated time period:

- Withhold for >8 hours all short-acting bronchodilators, including short-acting β2-agonists (such as Albuterol, Alupent, Berotec, Brethaire, Bronkometer, Maxair, Proventil, Tornolate, Ventolin, Xopenex), non-prescription adrenaline inhaler (e.g. Primatime Mist), anticholinergics (e.g. Atrovent), and cromolyn (e.g. Intal).
- Withhold for 12 hours: caffeine (coffee), short acting theophylline, and aminophylline.
- Withhold for 24 hours: intermediate acting theophylline and aminophylline and oral β2-agonists.
- Withhold for 48 hours all long-acting bronchodilators, including long-acting theophylline, nedocromil, salmeterol, formoterol, tiotropium or combinations with these medications.

Asthma severity:

In addition, to show the distribution of severity of disease, treatment step was used as a proxy to classify the patients in categories of severity of disease, based on the medications the patients were taking in the last 12 months [1]. Step 1 includes those subjects taking short-acting β2-agonist as required (equivalent to intermittent asthma); step 2 comprises those taking inhaled corticosteroid plus short-acting β2-agonist as needed (equivalent to mild asthma); step 3 includes the patients taking the same medications as step 2 plus long-acting β2-agonist (equivalent to moderate asthma); and step 4 comprises the patients taking the mediations described in step 3 plus leukotriene antagonist (equivalent to severe asthma). This information has been included as Table A.1.

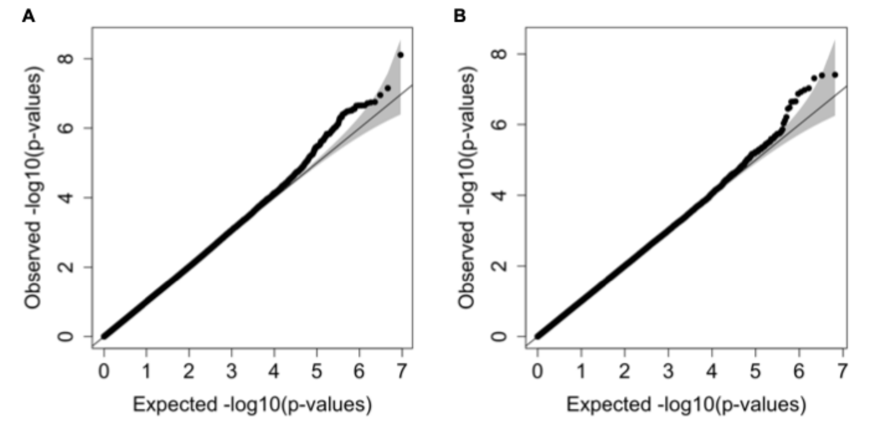

Figure A.1: Quantile-quantile plots for genome-wide allelic associations with BDR in a meta-analysis of A) SAGE I and II (inflation factor: λ = 1.006 for ~10 M common SNPs) **B)** SAGE I, SAGE II, and GALA II (inflation factor: λ = 1.004 for all common SNPs)

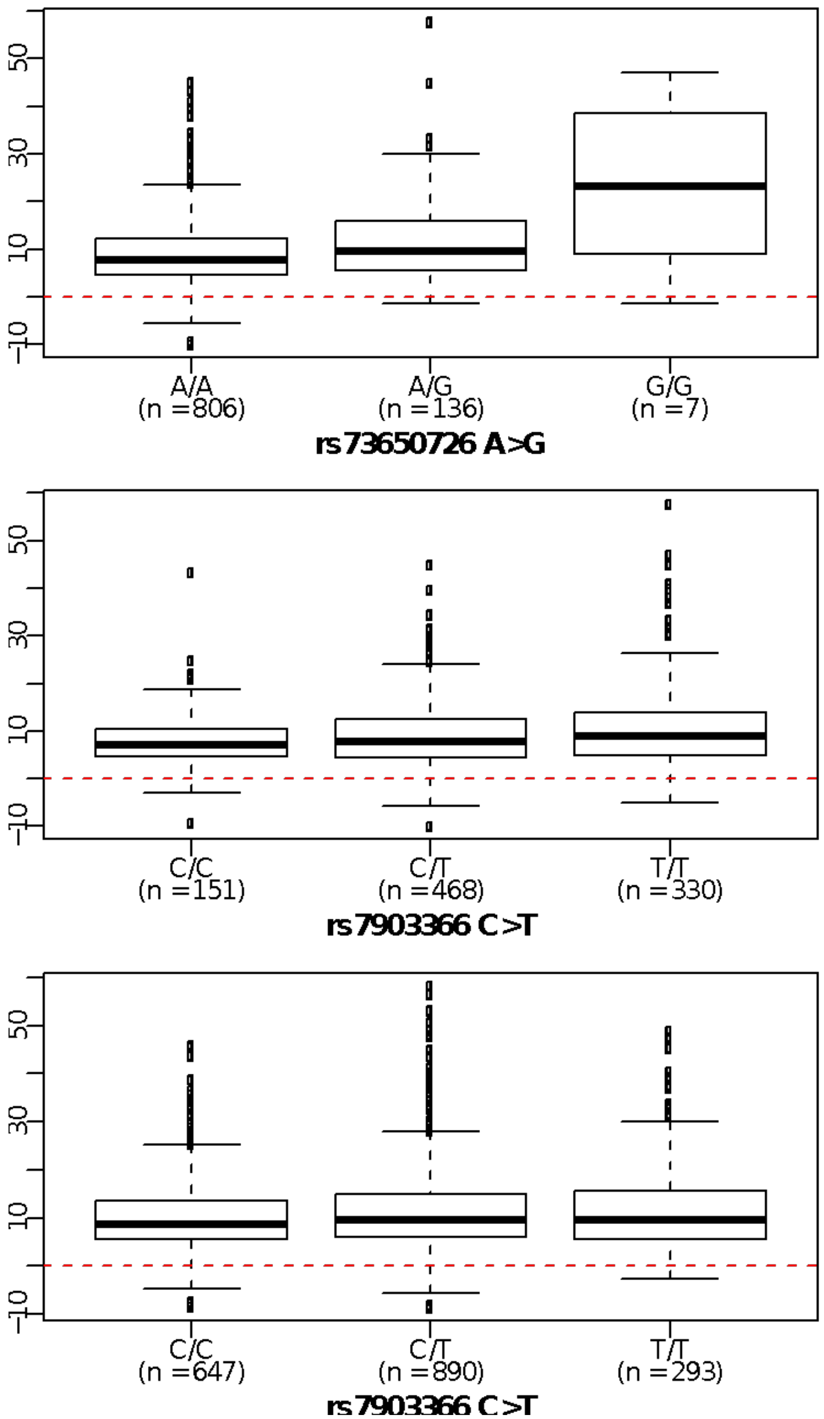

Figure A.2: Genotype-phenotype correlation: The box plot displays BDR with the three different genotypes at the: Top) rs73650726 in SAGE I and II Middle) rs7903366 in SAGE I and SAGE II Bottom) rs7903366 in GALA II.

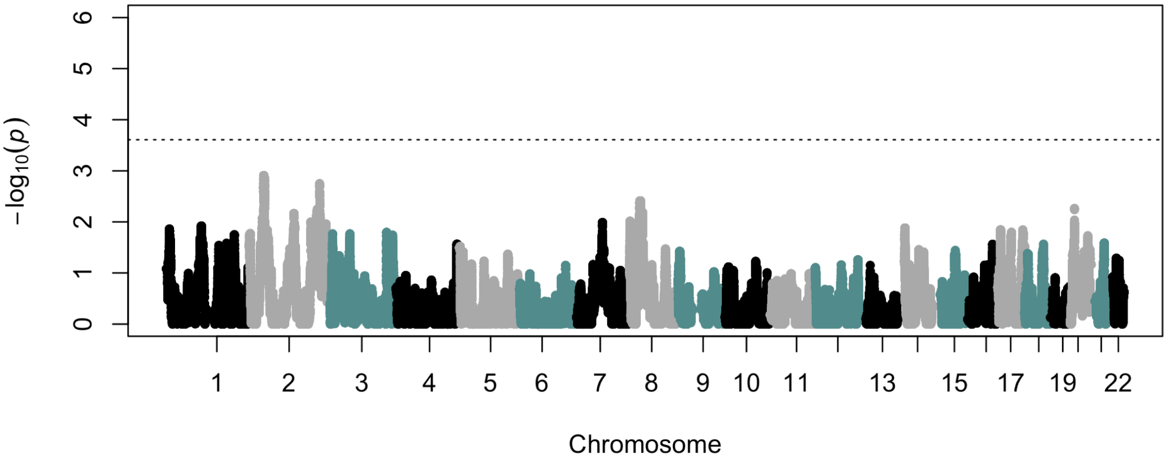

Figure A.3: Admixture mapping in SAGE II (n= 759) for African ancestry and BDR. Ancestry association testing was performed at 478,441 markers using linear regression including age, sex, BMI category, and global African ancestry covariates.

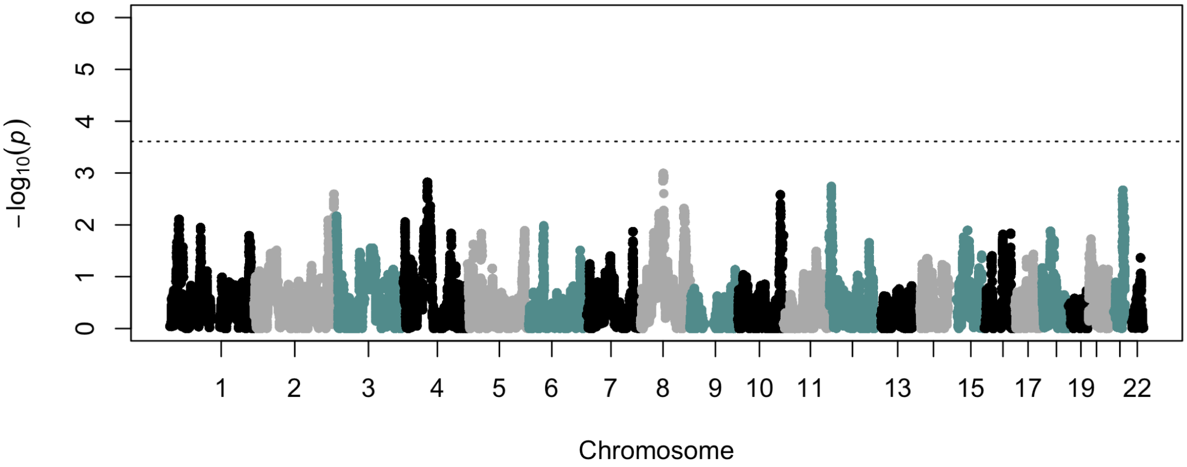

Figure A.4: Admixture mapping in SAGE I (n= 190) for African ancestry and BDR. Ancestry association testing was performed at 478,441 markers using linear regression including age, sex, BMI category, and global African ancestry covariates.

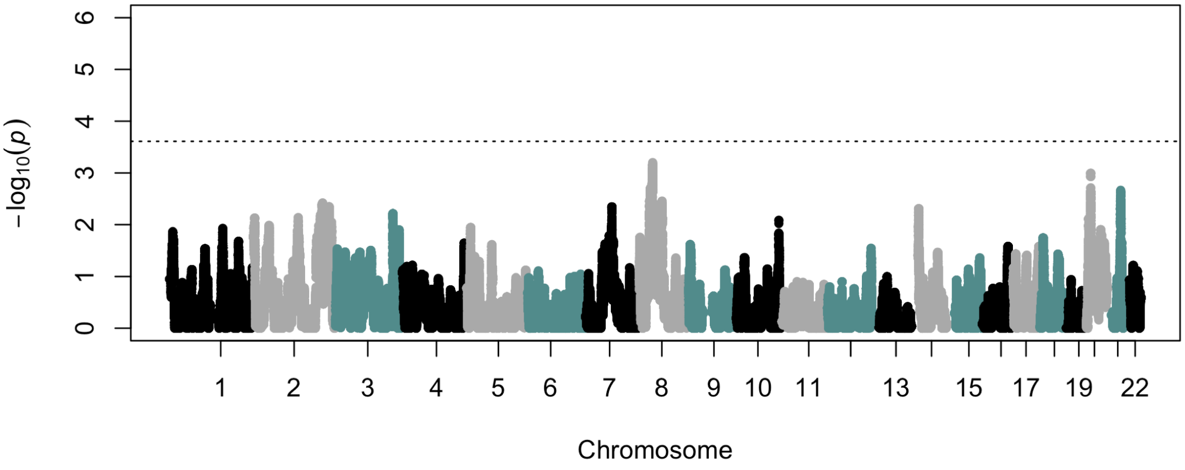

Figure A.5: Meta-analysis of admixture mapping in SAGE I and II (n=949) for African ancestry and BDR. Ancestry association testing was performed at 478,441 markers using linear regression including age, sex, BMI category, and global African ancestry covariates.

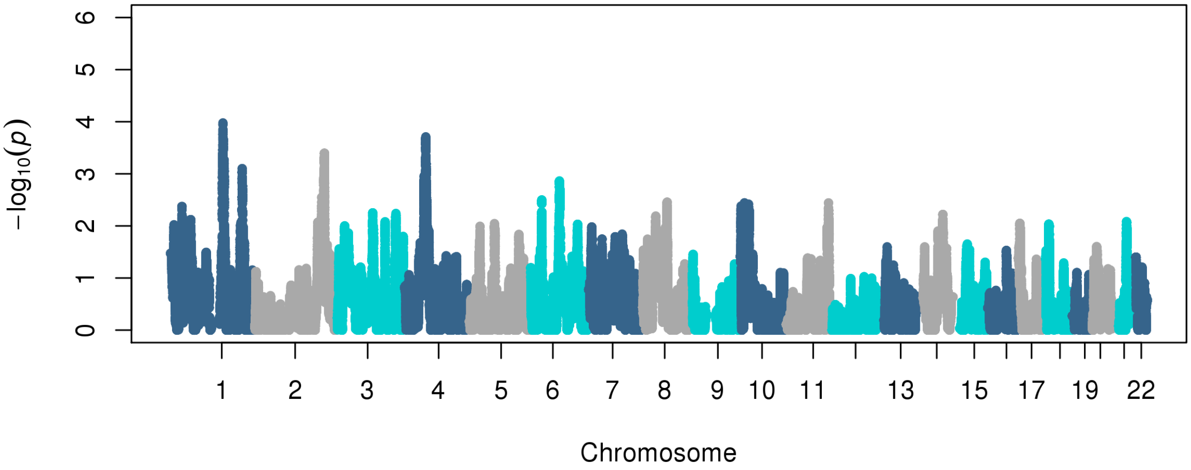

Figure A.6: Meta-analysis of admixture mapping in SAGE I, SAGE II, and GALA II (n=2,779) for African ancestry and BDR. For SAGE I and SAGE II, ancestry association testing was performed at 478,441 markers using linear regression including age, sex, BMI category, and global African ancestry covariates. For GALA II, ancestry association testing was performed using linear regression including age, sex, ethnicity, BMI category, global Native American ancestry and global African ancestry covariates. A meta-analysis was performed combining 362,528 markers.

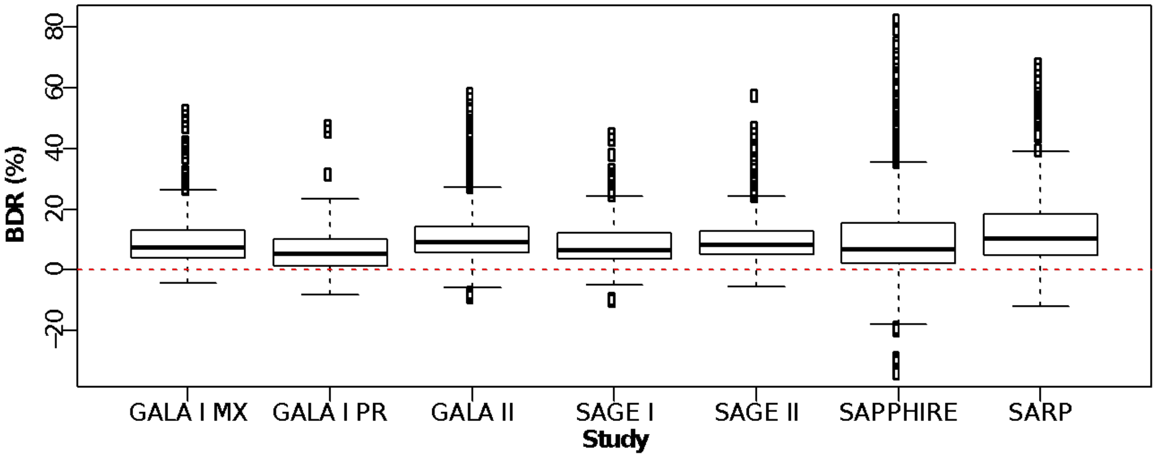

Figure A.7: Distribution of bronchodilator drug response measures for discovery and replication cohorts. Discovery cohorts = SAGE I (n=190), SAGE II (n=759) and GALA II (n =1,830); Replication cohorts = GALA I MX (n=247), GALA 1 PR (n=169), SAPPHIRE (n=1,325) and SARP (n=290).

Table A.1: Distribution of severity of disease in SAGE I, SAGE II, and GALA II.

Asthma medication (%)	SAGE I	SAGE II	GALA II
None	0%	13.10%	9.00%
Step 1	31.70%	30.60%	25.00%
Step 2	20.40%	29.60%	45.70%
Step 3	35.50%	19.80%	16.40%
Step 4	12.40%	6.90%	3.90%

Step 1: Short-acting β_2-agonist as required; Step 2: inhaled corticosteroid plus short-acting β_2-agonist as needed; Step 3: step 2 plus long-acting β_2-agonist; Step 4: step 3 plus leukotriene antagonist.

Table A.2: Individual genome-wide association study results from SAGE I, SAGE II, and GALA II.

SAGE I								
Chr	**SNP**	**Position (hg19)**	**A1**	**A2**	**Freq (A1)**	**Effect (A1)**	**StdErr**	**Pvalue**
9q21	rs73650726	85152666	A	G	0.92	-2.99	1.92	0.12
10q21	rs7903366	53689774	C	T	0.4	-0.54	1.02	0.6
10q21	rs7070958	53691116	A	G	0.39	-0.5	1.03	0.63
10q21	rs7081864	53690331	G	A	0.39	-0.46	1.03	0.65

SAGE II								
Chr	**SNP**	**Position (hg19)**	**A1**	**A2**	**Freq (A1)**	**Effect (A1)**	**StdErr**	**Pvalue**
9q21	rs73650726	85152666	A	G	0.92	-3.9	0.7	3.36x10^{-8}
10q21	rs7903366	53689774	C	T	0.41	-1.66	0.39	2.83x10^{-5}
10q21	rs7070958	53691116	A	G	0.41	-1.66	0.4	3.22x10^{-5}
10q21	rs7081864	53690331	G	A	0.41	-1.65	0.4	3.26x10^{-5}

GALA II								
Chr	**SNP**	**Position (hg19)**	**A1**	**A2**	**Freq (A1)**	**Effect (A1)**	**StdErr**	**Pvalue**
9q21	rs73650726	85152666	A	G	0.99	0.29	1.25	0.81
10q21	rs7903366	53689774	C	T	0.6	-1.06	0.28	1.74x10^{-4}
10q21	rs7070958	53691116	A	G	0.6	-1.08	0.28	1.57x10^{-4}
10q21	rs7081864	53690331	G	A	0.6	-1.06	0.28	1.79x10^{-4}

Table A.3: Replication results of candidate SNPs in GALA I, SAPPHIRE and SARP. Statistical significance was evaluated at Pvalue<0.05.

GALA I								
Population	**Chr**	**SNP**	**Position (hg19)**	**A1**	**A2**	**Effect (A1)**	**StdErr**	**Pvalue**
Mexicans	9q21	rs73650726	85152666	A	G	1.71	10.24	0.87
Mexicans	10q21	rs7903366	53689774	C	T	1.02	0.99	0.3
Mexicans	10q21	rs7070958	53691116	A	G	1.04	0.99	0.29
Mexicans	10q21	rs7081864	53690331	G	A	1.03	0.99	0.3
Puerto Ricans	9q21	rs73650726	85152666	A	G	-6.22	3.87	0.11
Puerto Ricans	10q21	rs7903366	53689774	C	T	-0.51	1.01	0.61
Puerto Ricans	10q21	rs7070958	53691116	A	G	-0.53	1	0.6
Puerto Ricans	10q21	rs7081864	53690331	G	A	-0.53	1	0.6

SAPPHIRE							
Chr	**SNP**	**Position (hg19)**	**A1**	**A2**	**Effect (A1)**	**StdErr**	**Pvalue**
9q21	rs73650726	85152666	A	G	-0.65	0.94	0.49
10q21	rs7903366	53689774	C	T	0.91	0.55	0.10
10q21	rs7070958	53691116	A	G	0.87	0.55	0.11
10q21	rs7081864	53690331	G	A	0.84	0.55	0.13

SARP							
Chr	**SNP**	**Position (hg19)**	**A1**	**A2**	**Effect (A1)**	**StdErr**	**Pvalue**
9q21	rs73650726	85152666	G	A	-6.12	2.981	0.04
10q21	rs7903366	53689774	C	T	-0.27	1.262	0.83
10q21	rs7070958	53691116	G	A	-0.4	1.275	0.76
10q21	rs7081864	53690331	A	G	-0.36	1.27	0.78

Table A.4: Top replication results (MAF >0.05) +/-50 kb of candidate SNPs in GALA I and SAPPHIRE.

GALA I									
Population	Chr	SNP	Position (hg19)	A1	A2	Freq (A1)	Effect (A1)	Std Err	Pvalue
Mexicans	9	rs34293766	85126730	T	G	0.94	5.74	2.27	0.01
Mexicans	10	rs59960792	53690251	C	T	0.80	2.97	1.12	0.01
Puerto Ricans	9	rs10780548	85167648	A	G	0.63	2.93	0.99	0.003
Puerto Ricans	10	rs6480581	53672743	T	C	0.74	-2.67	1.08	0.01

SAPPHIRE									
Population	Chr	SNP	Position (hg19)	A1	A2	Freq (A1)	Effect (A1)	Std Err	Pvalue
African Americans	9	rs62576848	85144677	T	C	0.94	3.03	1.23	0.01
African Americans	10	rs35969600	53662118	C	T	0.72	-1.49	0.59	0.01

Table A.5. Candidate SNP replication in meta-analysis of SAGE I and SAGE II. Under 'Direction' column, 1st symbols refer to SAGE I, second refer to SAGE II

GENE	SNP	Chr	Position	A1	A2	Effect	Pvalue	Direction
SOCS-ASB3 region	rs350729	2	52983773	T	G	0.22	0.53	-+
ADRB2	rs1042713	5	148206440	A	G	-0.13	0.7	--
ADRB2	rs1042714	5	148206473	C	G	-0.56	0.22	--
ADCY9	rs2230739	16	4033436	T	C	-0.14	0.77	+-
CRHR2	rs7793837	7	30726777	A	T	-0.26	0.58	--
ARG1	rs2781659	6	131891820	A	G	0.39	0.3	-+
SPATS2L	rs295137	2	201150040	T	C	0.4	0.23	-+
SPATS2L	rs295114	2	201195602	T	C	0.41	0.22	-+
THRB	rs892940	3	24538838	A	G	0.52	0.21	++
CRHR2	rs73294475	7	30701596	T	C	-0.32	0.51	+-
SPATA13	rs9507294	13	24823347	T	C	0.38	0.48	++
SPATA13	rs912142	13	24827500	A	G	-0.23	0.51	--
SPATA13	rs2248119	13	24827094	A	G	0.49	0.16	++
SPATA13	rs9551086	13	24830330	T	C	-1.36	0.02	--
SPATA13	rs9553225	13	24823006	A	G	-1.17	0.10	--

Table A.6: Meta-analysis of admixture mapping results within African Americans (SAGE I and II). Under 'Direction' the first symbol refers to SAGE I, second to SAGE II

Chr	SNP	Position (hg19)	Effect	StdErr	Pvalue	Direction
8p11	rs3927941	39805797	1.49	0.44	6.34×10^{-4}	++

Reference:

1. National Asthma E, Prevention P. Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. *J Allergy Clin Immunol* 2007; **120**(5 Suppl)**:** S94-138.

APPENDIX B: Supplementary Material to Chapter 3

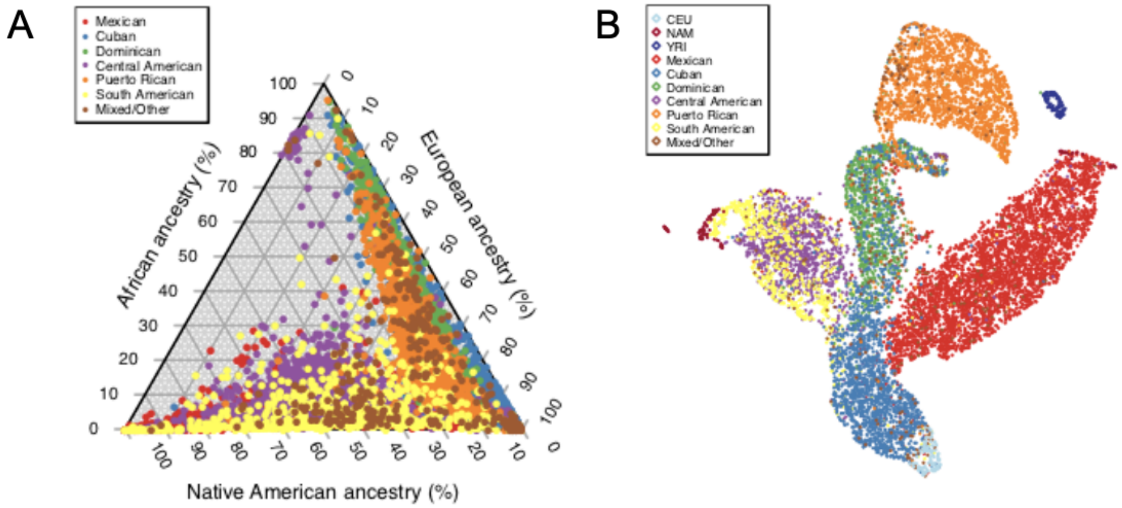

Figure B.1. Continental ancestral diversity of HCHS/SOL A) Ternary plot of global ancestry proportions colored by population for 10,268 HCHS/SOL samples B) Uniform Manifold Approximation and Projection (UMAP) plot of HCHS/SOL and the reference panel (n=10,591) using 3 principal components, colored by population.

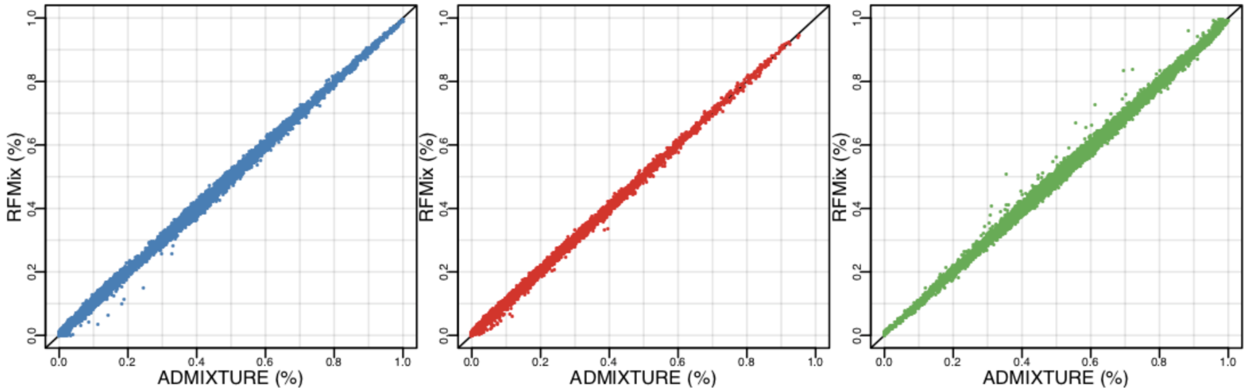

Figure B.2. Concordance of ADMIXTURE and RFMix global ancestry estimates. A) Native American ancestry B) African ancestry and C) European ancestry.

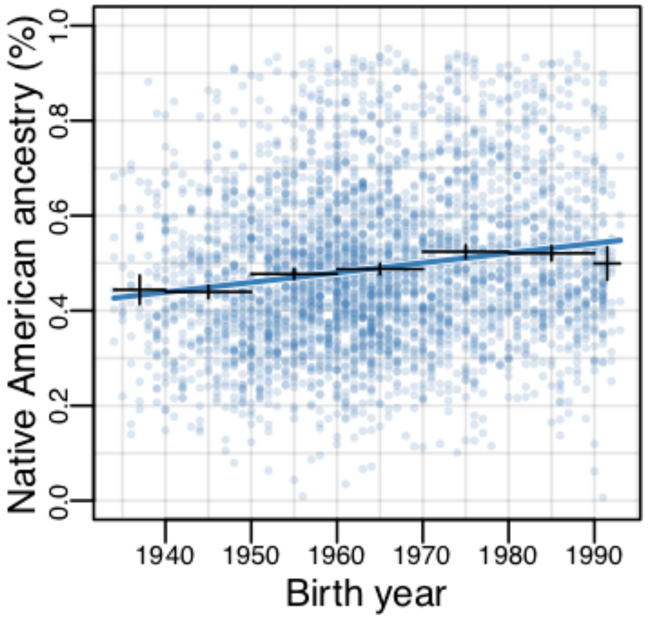

Figure B.3. RFMix inferred Native American global ancestry proportions plotted over time for HCHS/SOL Mexican Americans (n=3622).

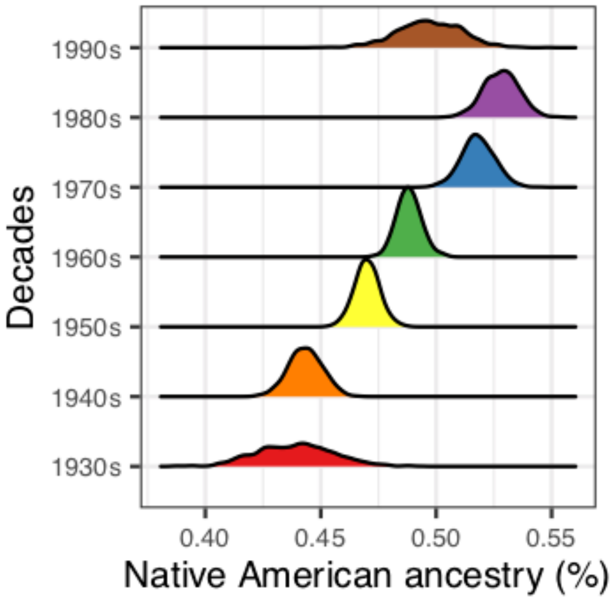

Figure B.4: Distributions of Native American global ancestry means generated by 1000 bootstrap resampling iterations within each decade of binned birth years.

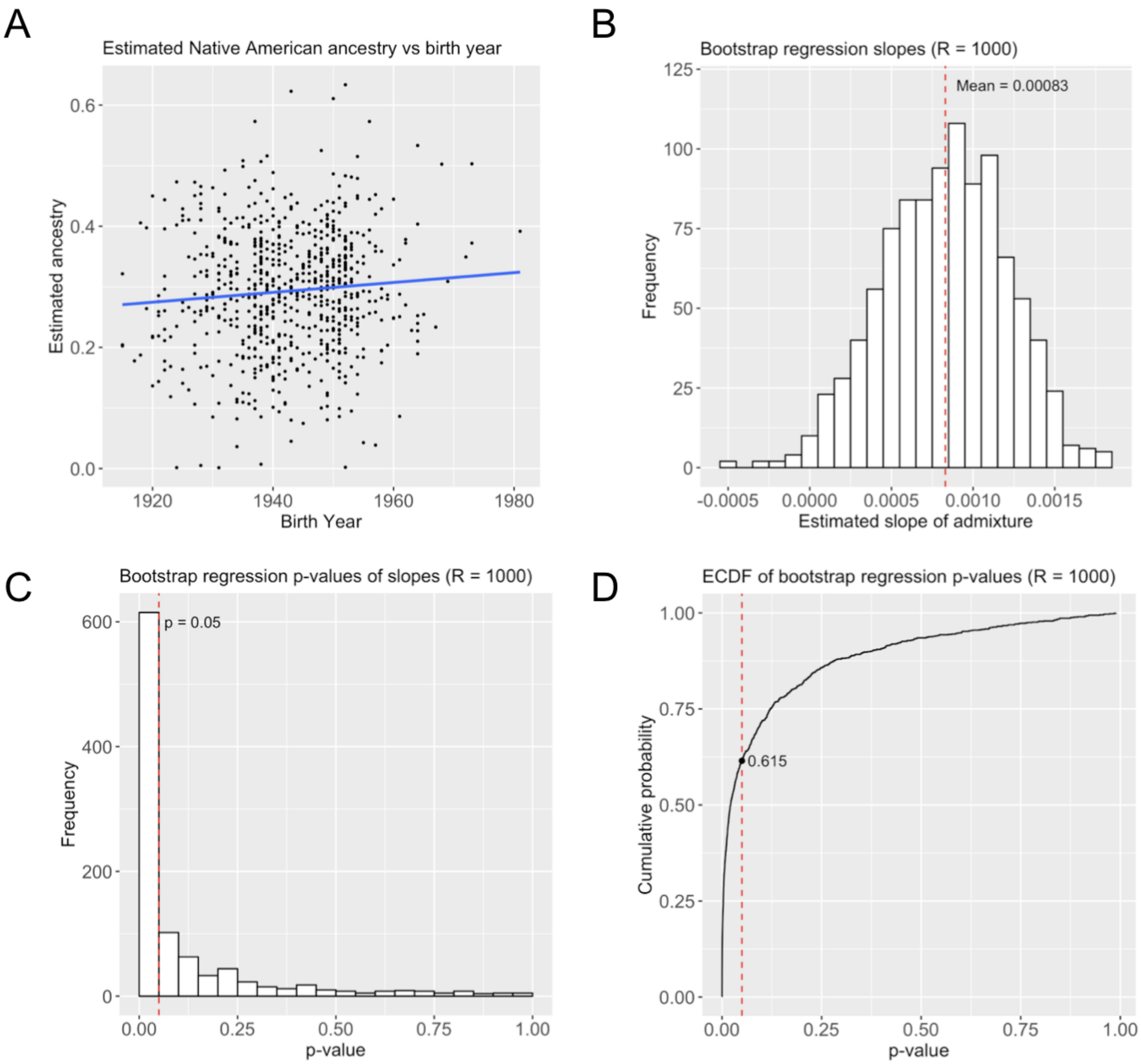

Figure B.5. Replication in the Health and Retirement Study for 705 self-identified Mexican Americans. A) Ancestry over time B) Distribution of regression slopes after 1000 bootstrap resampling iterations C) Distribution of bootstrap regression p-values D) ECDF of bootstrap regression p-values.

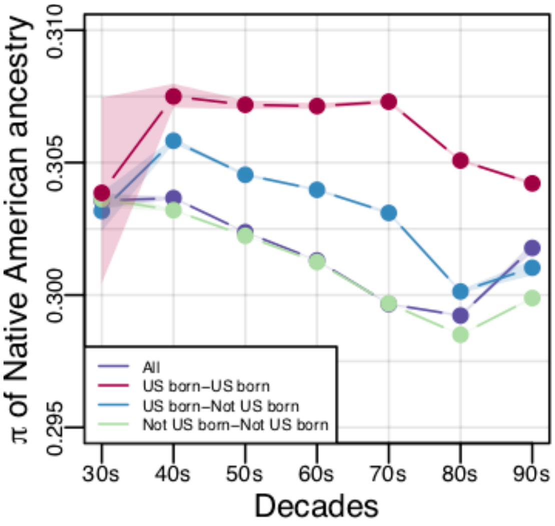

Figure B.6. Diversity of and within Native American ancestral tracts. Diversity (π) of subcontinental Native American ancestry stratified by US born/not US born status. π was calculated between pairs within each decade of birth years. 95% confidence intervals are highlighted by the shaded regions for each group.

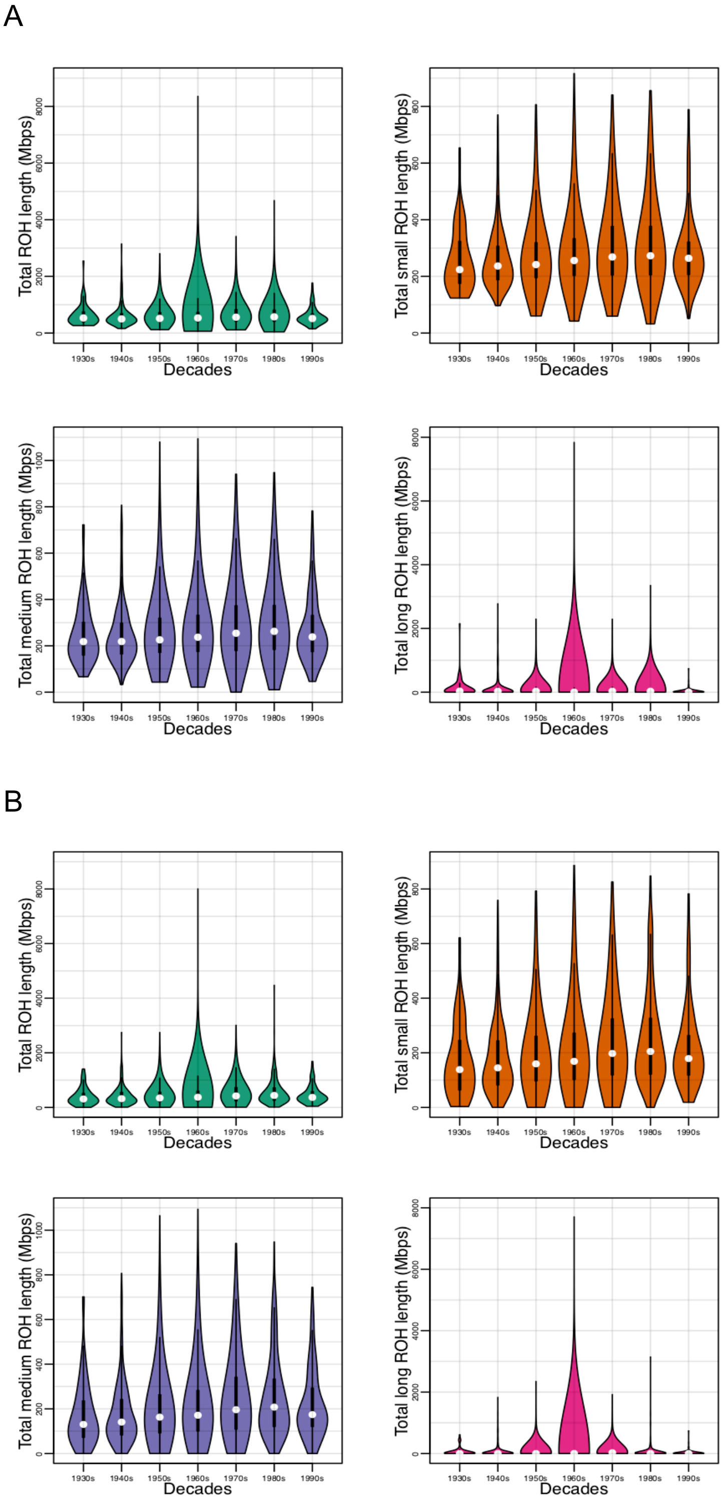

Figure B.7. Runs of homozygosity (ROH) in HCHS/SOL Mexican Americans. A) ROH across all ancestries separated by ROH class B) ROH overlapping Native American haplotypes separated by ROH clas

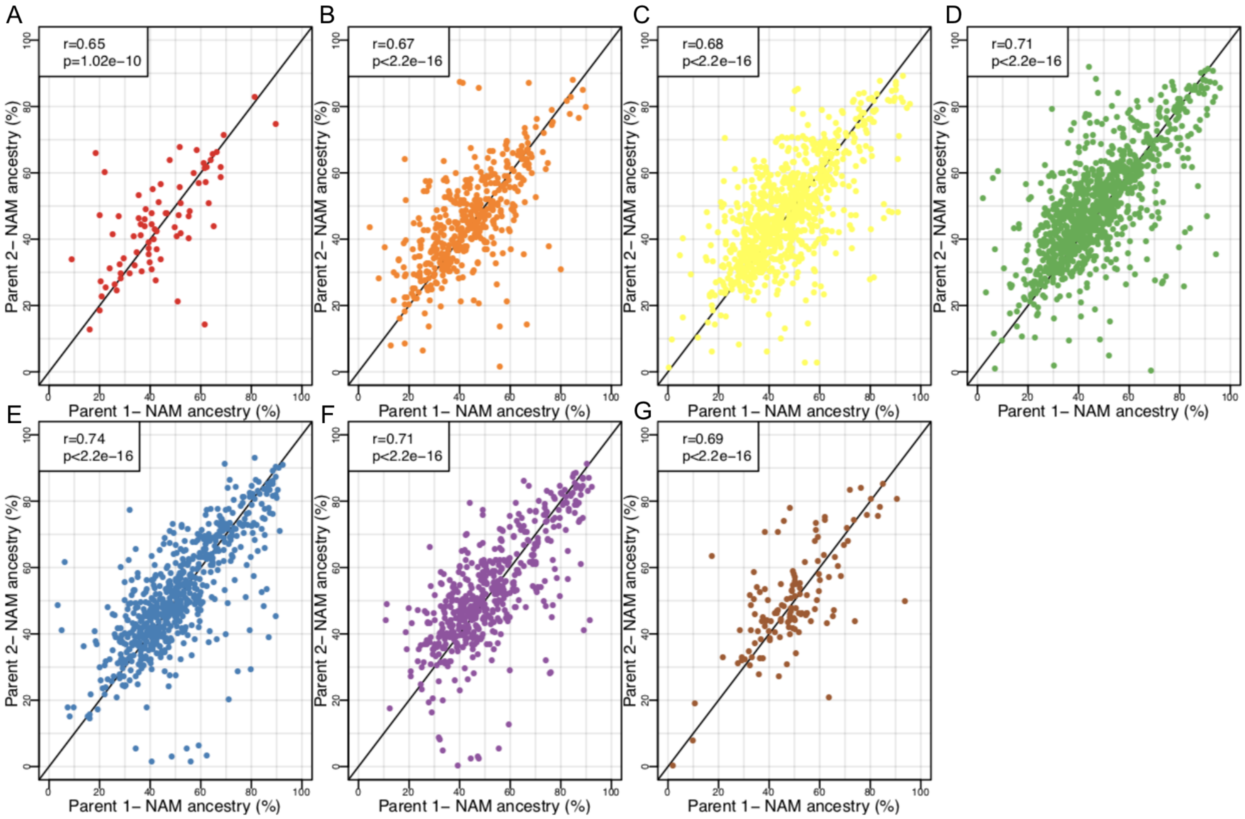

Figure B.8. Ancestry-related assortative mating in HCHS/SOL Mexican Americans separated by decade. Each plot represents the **c**orrelation of parent's inferred Native American ancestries using ANCESTOR by decade beginning with the 1930s (A) and ending with the 1990s (G). Each point corresponds to one Mexican American couple and the axes correspond to the inferred Native American ancestry of each partner.

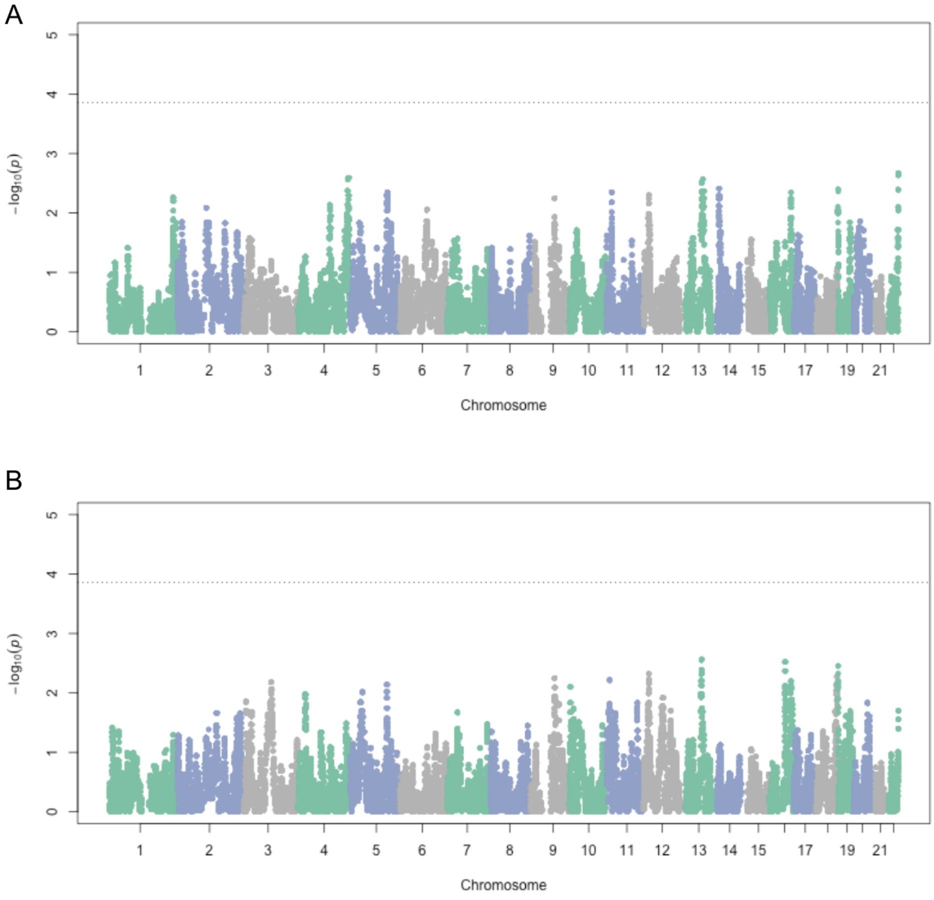

Figure B.9. Admixture mapping in HCHS/SOL Mexicans (n=3622) for Native American ancestry and A) birth year and B) Generation. Ancestry association testing was performed at 211,151 markers using A) linear regression and B) logistic regression, both including global Native American ancestry, sampling weight and center as covariates.

Table B.1: Association of global ancestries and birth year for all HCHS/SOL samples

Population	Ancestry	N	R2	Effect	Std.Err	Pvalue
Central American	NAM	1099	0.0139	0.0013	0.0004	0.0013
Central American	AFR	1099	0.0136	0.0002	0.0004	0.6561
Central American	EUR	1099	0.0138	-0.0015	0.0004	0.0002
Cuban	NAM	1536	0.0023	0.0002	0.0001	0.0938
Cuban	AFR	1536	0.0014	-0.0005	0.0004	0.1490
Cuban	EUR	1536	0.0005	0.0003	0.0004	0.3879
Dominican	NAM	954	0.0035	0.0002	0.0001	0.0663
Dominican	AFR	954	0.0030	-0.0007	0.0004	0.1287
Dominican	EUR	954	0.0022	0.0005	0.0004	0.2374
Mexican	NAM	3622	0.0268	0.0023	0.0002	0.0000
Mexican	AFR	3622	0.0008	0.0000	0.0000	0.4189
Mexican	EUR	3622	0.0285	-0.0023	0.0002	0.0000
Puerto Rican	NAM	1783	0.0014	0.0001	0.0001	0.1533
Puerto Rican	AFR	1783	0.0014	0.0003	0.0002	0.1743
Puerto Rican	EUR	1783	0.0027	-0.0005	0.0002	0.0355
South American	NAM	652	0.0110	0.0016	0.0007	0.0211
South American	AFR	652	0.0027	-0.0002	0.0004	0.5053
South American	EUR	652	0.0080	-0.0014	0.0006	0.0335

APPENDIX C: Supplementary Material to Chapter 4

Table C.1. Multiple regression table with traits that were significantly correlated with global NAM ancestry.

Trait	N	R2	Effect	Std.Err	P
Height	3615	0.588	-13.637	0.676	$1.01x10^{-85}$
Predicted FVC	3522	0.851	-695.376	36.867	$1.13x10^{-75}$
EGFR MDRD	3308	0.196	23.398	2.435	$1.39x10^{-21}$
EGFR CKD Epi	3308	0.441	14.752	1.581	$1.90x10^{-20}$
Waist to hip ratio	3617	0.195	0.043	0.007	$6.96x10^{-10}$
Glycosylated hemoglobin (HbA1c)	3609	0.081	9.184	1.59	$8.30x10^{-9}$
% Glycosylated hemoglobin	3609	0.08	0.837	0.146	$9.88x10^{-9}$
HDL cholesterol	3621	0.095	-6.74	1.372	$9.39x10^{-7}$
Total iron binding capacity	3620	0.066	23.54	5.467	$1.71x10^{-5}$
FEV1 FVC Ratio	3505	0.176	2.56	0.639	$6.37x10^{-5}$
% Lymphocytes	3442	0.021	3.768	1.023	$2.35x10^{-4}$
Hip girth	3617	0.077	-4.452	1.273	$4.78x10^{-4}$
% Neutrophils	3442	0.039	-3.649	1.178	$1.96x10^{-3}$
Monocyte count	3443	0.024	-0.048	0.019	$1.27x10^{-2}$
Neutrophil count	3442	0.043	-0.399	0.165	$1.59x10^{-2}$
LDL cholesterol	3529	0.047	-8.269	3.963	$3.70x10^{-2}$
Average diastolic blood pressure	3616	0.072	-2.067	1.136	$6.88x10^{-2}$
Total cholesterol	3622	0.065	-6.248	4.614	$1.76x10^{-1}$
White blood cell count	3442	0.032	-0.26	0.208	$2.12x10^{-1}$
Average systolic blood pressure	3619	0.211	1.59	1.712	$3.53x10^{-1}$
% Immature granulocytes	555	0.261	-0.09	0.11	$4.14x10^{-1}$
QRS duration	3596	0.168	0.062	1.268	$9.61x10^{-1}$

Table C.2. Height over time in HCHS/SOL Mexican Americans

Group	N	R2	Effect	Std.Err	P
All	3614	0.542	0.120	0.009	3.28×10^{-39}
NAM>0.58	929	0.575	0.159	0.018	2.73×10^{-18}
0.46 <=NAM<=0.58	955	0.578	0.154	0.017	3.07×10^{-19}
0.37<=NAM<0.46	842	0.547	0.101	0.017	9.73×10^{-9}
NAM<0.37	888	0.537	0.116	0.018	2.55×10^{-10}

Table C.3. Observed height vs. predicted height in HCHS/SOL Mexican Americans

Group	N	R2	Effect	Std.Err	P
All	3614	0.0249	0.0045	0.0006	2.19×10^{-12}
NAM>0.58	929	0.0072	0.0022	0.0012	7.79×10^{-2}
0.46<=NAM<=0.58	955	0.0058	0.0011	0.0013	3.90×10^{-1}
0.37<=NAM<0.46	842	0.0144	0.0043	0.0015	3.22×10^{-3}
NAM<0.37	888	0.0164	0.0043	0.0013	7.91×10^{-4}

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

October 14, 2019

Date