

Whole-Genome Sequencing of Pharmacogenetic Drug Response in Racially Diverse Children with Asthma

Angel C. Y. Mak^{1*}, Marquitta J. White^{1*}, Walter L. Eckalbar^{1*}, Zachary A. Szpiech^{2*}, Sam S. Oh¹, Maria Pino-Yanes^{3,4}, Donglei Hu¹, Pagé Goddard¹, Scott Huntsman¹, Joshua Galanter^{1†}, Ann Chen Wu^{5,6}, Blanca E. Himes⁷, Soren Gerner⁸, Julia M. Vogel^{8§}, Karen L. Bunting⁸, Celeste Eng¹, Sandra Salazar¹, Kevin L. Keys¹, Jennifer Liberto¹, Thomas J. Nuckton¹, Thomas A. Nguyen¹, Dara G. Torgerson⁹, Pui-Yan Kwok^{10,11}, Albert M. Levin¹², Juan C. Celedón¹³, Erick Forno¹³, Hakon Hakonarson^{14,15}, Patrick M. Sleiman^{14,15}, Amber Dahlin⁵, Kelan G. Tantisira⁵, Scott T. Weiss⁵, Denise Serebrisky¹⁶, Emerita Brigino-Buenaventura¹⁷, Harold J. Farber¹⁸, Kelley Meade¹⁹, Michael A. Lenoir²⁰, Pedro C. Avila^{21||}, Saunak Sen^{1†}, Shannon M. Thyne²², William Rodriguez-Cintron²³, Cheryl A. Winkler²⁴, Andrés Moreno-Estrada²⁵, Karla Sandoval²⁵, Jose R. Rodriguez-Santana²⁶, Rajesh Kumar^{27,28}, L. Keoki Williams^{29,30}, Nadav Ahituv^{2,11}, Elad Ziv¹, Max A. Seibold³¹, Robert B. Darnell^{8,32,33}, Noah Zaitlen¹, Ryan D. Hernandez^{2,10,34}, and Esteban G. Burchard^{1,2}; on behalf of the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

¹Department of Medicine, ²Department of Bioengineering and Therapeutic Sciences, ⁹Department of Pediatrics, ¹⁰Cardiovascular Research Institute, ¹¹Institute for Human Genetics, and ³⁴Quantitative Biosciences Institute, University of California San Francisco, San Francisco, California; ³Research Unit, Hospital Universitario N. S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; ⁴CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; ⁵Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; ⁶Precision Medicine Translational Research (PRoMoTeR) Center, Department of Population Medicine, Harvard Medical School and Pilgrim Health Care Institute, Boston, Massachusetts; ⁷Department of Biostatistics, Epidemiology and Informatics and ¹⁵Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; ⁸New York Genome Center, New York, New York; ¹²Department of Public Health Sciences, ²⁹Department of Internal Medicine, and ³⁰Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, Michigan; ¹³Division of Pediatric Pulmonary Medicine, Allergy and Immunology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; ¹⁴Center for Applied Genomics, The Children's Hospital of Philadelphia Research Institute, Philadelphia, Pennsylvania; ¹⁶Pediatric Pulmonary Division, Jacobi Medical Center, Bronx, New York; ¹⁷Department of Allergy and Immunology, Kaiser Permanente Vallejo Medical Center, Vallejo, California; ¹⁸Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, Texas; ¹⁹Children's Hospital and Research Center, Oakland, California; ²⁰Bay Area Pediatrics, Oakland, California; ²¹Department of Medicine, Northwestern University, Chicago, Illinois; ²²Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California; ²³Veterans Caribbean Health Care System, San Juan, Puerto Rico; ²⁴Basic Science Laboratory, Center for Cancer Research, National Cancer Institute, Leidos Biomedical Research, Frederick National Laboratory, Frederick, Maryland; ²⁵National Laboratory of Genomics for Biodiversity (UGA-LANGEBIO), CINVESTAV, Irapuato, Guanajuato, Mexico; ²⁶Centro de Neumología Pediátrica, San Juan, Puerto Rico; ²⁷Division of Allergy and Immunology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; ²⁸Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois; ³¹Center for Genes, Environment and Health, Department of Pediatrics, National Jewish Health, Denver, Colorado; and ³²Laboratory of Molecular Neuro-Oncology and ³³Howard Hughes Medical Institute, The Rockefeller University, New York, New York

ORCID IDs: 0000-0002-5372-4198 (A.C.Y.M.); 0000-0001-6372-8224 (Z.A.S.); 0000-0002-2815-6037 (S.S.O.); 0000-0001-6497-9885 (E.F.).

Abstract

Rationale: Albuterol, a bronchodilator medication, is the first-line therapy for asthma worldwide. There are significant racial/ethnic differences in albuterol drug response.

Objectives: To identify genetic variants important for bronchodilator drug response (BDR) in racially diverse children.

Methods: We performed the first whole-genome sequencing pharmacogenetics study from 1,441 children with asthma from the tails of the BDR distribution to identify genetic association with BDR.

Measurements and Main Results: We identified population-specific and shared genetic variants associated with BDR, including genome-wide significant ($P < 3.53 \times 10^{-7}$) and suggestive ($P < 7.06 \times 10^{-6}$) loci near genes previously associated with lung capacity (*DNAH5*), immunity (*NFKB1* and *PLCB1*), and β -adrenergic signaling (*ADAMTS3* and *COX18*). Functional analyses of the

BDR-associated SNP in *NFKB1* revealed potential regulatory function in bronchial smooth muscle cells. The SNP is also an expression quantitative trait locus for a neighboring gene, *SLC39A8*. The lack of other asthma study populations with BDR and whole-genome sequencing data on minority children makes it impossible to perform replication of our rare variant associations. Minority underrepresentation also poses significant challenges to identify age-matched and population-matched cohorts of sufficient sample size for replication of our common variant findings.

Conclusions: The lack of minority data, despite a collaboration of eight universities and 13 individual laboratories, highlights the urgent need for a dedicated national effort to prioritize diversity in research. Our study expands the understanding of pharmacogenetic analyses in racially/ethnically diverse populations and advances the foundation for precision medicine in at-risk and understudied minority populations.

Keywords: albuterol; asthma; minority; *NFKB1*; Latinos

Asthma is a chronic inflammatory disorder characterized by recurrent respiratory symptoms and reversible airway obstruction. Asthma is the most common chronic childhood disease (1). In the United States, asthma prevalence is highest among Puerto Ricans (36.5%), intermediate among African Americans (13.0%) and European Americans (12.1%), and lowest among Mexican Americans (7.5%) (2). Asthma mortality is fourfold to fivefold higher in Puerto Ricans and African Americans compared with white persons and Mexican Americans (3).

Inhaled β_2 -agonists (e.g., albuterol) are the preferred treatment for acute asthma symptoms. Albuterol produces bronchodilation by causing rapid smooth

muscle relaxation in the airways. Albuterol is the most commonly prescribed asthma medication worldwide (4). Among low-income and minority populations, albuterol is often the only medication used for asthma regardless of asthma severity (5, 6).

Spirometry is used to quantify bronchodilator drug response (BDR), which varies significantly among individuals and between populations (7). Specifically, the populations with the highest asthma prevalence and mortality also have the lowest BDR: Puerto Rican and African American children have significantly lower BDR than white and Mexican American children (7, 8). This racial/ethnic variation in BDR may contribute to the observed

disparities in asthma morbidity and mortality (9, 10).

BDR is a complex trait, influenced by social, environmental, and genetic factors, with heritability estimates ranging from 47% to 92% (11–13). Genome-wide association studies (GWASs) have identified several common SNPs associated with BDR in populations of European descent (14–16). Only one GWAS of BDR has been conducted among African Americans (17). Although that study identified a novel BDR-associated locus, it did not replicate associations discovered in populations of European descent, suggesting that BDR may be partly determined by population-specific variants. Our genetic investigation of BDR among Latinos identified a

(Received in original form December 15, 2017; accepted in final form March 5, 2018)

*These authors contributed equally to this work.

†Current affiliation: Genentech, South San Francisco, California.

‡Current affiliation: The Scripps Research Institute, La Jolla, California.

§Current affiliation: Allergy and ENT Associates, The Woodlands, Texas.

¶Current affiliation: Division of Biostatistics, Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee.

Whole-genome sequencing for the TOPMed (Trans-Omics in Precision Medicine) program was supported by the NHLBI. Whole-genome sequencing for GALA II (NHLBI TOPMed: Genes-environments and Admixture in Latino Americans) Study (phs000920) and SAGE (NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environments) (phs000921) was performed at the New York Genome Center (3R01HL117004-01S3). Centralized read mapping and genotype calling, along with variant quality metrics and filtering, were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). This work was supported in part by the Sandler Family Foundation, the American Asthma Foundation, the RWJF Amos Medical Faculty Development Program, the Harry Wm. and Diana V. Hind Distinguished Professor in Pharmaceutical Sciences II, the NHLBI (R01HL117004, R01HL128439, R01HL135156, and X01HL134589), the National Institute of Environmental Health Sciences (R01ES015794 and R21ES24844), the National Institute on Minority Health and Health Disparities (NIMHD; P60MD006902, RL5GM118984, and R01MD010443), and the Tobacco-Related Disease Research Program (Award Number 24RT-0025). M.J.W. was supported by a diversity supplement of NHLBI (R01HL117004) and an Institutional Research and Academic Career Development Award (K12GM081266). K.L.K. was supported by a diversity supplement of NHLBI (R01HL135156). W.L.E. was supported by NHLBI (K99HL135403-01A1). M.P.-Y. was supported by the grant AC15/00015 by Instituto de Salud Carlos III within the ERACoSysMed first Joint Transnational Call (SysPharmPedia) from the European Union, under the Horizon 2020. A.C.W. was supported by the National Institute of Child Health and Human Development (R01HD0855993). B.E.H. was supported by NHLBI (R01HL133433). R.B.D., S.G., K.L.B., and J.M.V. were supported by the National Human Genome Research Institute (NHGRI) of the NIH (Award Number UM1HG008901). H.H. was supported by NHGRI (U01HG008684). C.A.W. was funded in part with federal funds from the National Cancer Institute, NIH (contract HHSN26120080001E), and by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. L.K.W. was supported by the National Institute of Allergy and Infectious Diseases (R01AI079139), NHLBI (R01HL118267), and National Institute of Diabetes and Digestive and Kidney Diseases (R01DK113003), American Asthma Foundation, and Fund for Henry Ford Hospital. N.A. was supported in part by NHGRI and Division of Cancer Prevention, National Cancer Institute (1R01CA197139), and NHLBI (1R01HL138424). N.Z. was supported by the NHLBI (K25HL121295) and the NHGRI (R01HG006399). M.A.S. was supported by NHLBI (P01-HL132821-01A1, R01HL135156-01, and R01HL128439-03), NIMHD (R01MD010443-02), and the Department of Defense (W81WH-16-2-0018). C.A.W. declares that the content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Author Contributions: A.C.Y.M., M.J.W., Z.A.S., D.H., M.P.-Y., J.G., J.L., P.G., and S.H. analyzed data. R.D.H., N.Z., M.A.S., E.G.B., D.G.T., M.J.W., A.C.Y.M., N.A., E.Z., R.B.D., and P.-Y.K. designed experiments or supervised research. A.C.Y.M., M.J.W., Z.A.S., S.S.O., M.P.-Y., E.G.B., J.L., P.G., T.J.N., and K.L.K. prepared tables and figures, and wrote the manuscript. All authors provided critical review of the manuscript. C.E., S.G., K.L.B., and J.M.V. generated data and supervised data generation. W.L.E., T.A.N., and B.E.H. performed validation experiments and analyses. M.J.W., M.P.-Y., L.K.W., A.M.L., J.C.C., E.F., H.H., P.M.S., A.C.W., A.D., K.G.T., and S.T.W. performed replication analyses. E.G.B., D.S., E.B.-B., H.J.F., J.R.R.-S., K.M., M.A.L., P.C.A., R.K., S. Sen, S.M.T., and W.R.-C. recruited study subjects. S. Salazar coordinated recruitment and maintained phenotype data quality. C.A.W., A.M.-E., and K.S. contributed Native American samples. TOPMed team provided variant annotation.

Correspondence and requests for reprints should be addressed to Angel C. Y. Mak, M.Phil., Ph.D., Department of Medicine, University of California San Francisco, 1550 4th Street, San Francisco, CA 94143. E-mail: angel.mak@ucsf.edu.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 197, Iss 12, pp 1552–1564, Jun 15, 2018

Copyright © 2018 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201712-2529OC on March 6, 2018

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Asthma is the most common chronic disease among children.

Albuterol, a bronchodilator medication, is the first-line therapy for asthma treatment worldwide. In the United States, asthma prevalence is the highest among Puerto Ricans, intermediate among African Americans and white persons, and lowest in Mexicans. Asthma mortality is fourfold to fivefold higher in Puerto Ricans and African Americans compared with Mexicans. Puerto Ricans and African Americans, the populations with the highest asthma prevalence and mortality, also have the lowest albuterol bronchodilator drug response.

What This Study Adds to the

Field: We conducted the largest pharmacogenetic study using whole-genome sequencing data from 1,441 minority children with asthma who had extremely high or low albuterol bronchodilator drug response. We identified population-specific and shared pharmacogenetic variants associated with bronchodilator drug response. We prioritized variants in *NFKB1* with multiple levels of existing biologic evidence and demonstrated their potential regulatory functions using chromatin immunoprecipitation sequencing, RNA sequencing, and luciferase enhancer assays. Our study reveals the challenges and importance of increasing diversity in research. Our findings help inform the direction of future development of asthma medications and advance the foundation of precision medicine for at-risk, yet understudied, racially/ethnically diverse populations.

significant contribution of population-specific rare variants (18).

Whole-genome sequencing (WGS) can identify rare and/or population-specific disease-causing genetic variants. WGS allows the detection of common and rare variants in coding and noncoding regions in a truly comprehensive and agnostic evaluation in the context of complex disease.

We performed WGS on 1,441 minority children with asthma from the tails of the

BDR distribution (see Figure E1 in the online supplement). We examined high and low drug responders from three ethnic groups: Puerto Ricans, Mexicans, and African Americans. Subject selection and the main analyses are summarized in Figure 1. We identified BDR-associated common and rare variants that are population-specific or shared among populations. This study represents a highly collaborative effort and is part of the NHLBI TOPMed (Trans-Omics for Precision Medicine Whole Genome Sequencing) program. Our study represents the first WGS study to investigate genetic variants of BDR in racially diverse children with asthma. Some of the results of this study have been previously reported in the form of conference abstracts (19–22).

Methods

Study Cohorts and Sample Details

This study examined a subset of subjects with asthma from SAGE II (Study of African Americans, Asthma, Genes and Environments) (23) and GALA II (Genes-Environments and Admixture in Latino Americans) studies (18).

A total of 1,441 individuals from three ethnic groups (483 Puerto Ricans, 483 Mexicans, and 475 African Americans) representing the tails of the BDR distribution were selected for WGS (Figure 1). Sequencing quality control metrics are summarized in Table E1. Subject selection and filtering processes are described in the online supplement. Descriptive data of study participants are summarized in Table 1. Detailed descriptions of BDR assessment and analysis are described in the online supplement.

WGS Data Generation, Processing, and Quality Control

Details regarding DNA processing, WGS, variant calling, data quality controls, and variant annotation are described in the online supplement.

Data Availability

TOPMed WGS data are available to download by submitting a data access request through dbGaP. The dbGaP study accession numbers for GALA II and SAGE II are phs000920.v1.p1 and phs000921.v1.p1, respectively.

BDR Association Testing and Replication

Single and multivariate testing for BDR association were performed. Logistic regression analysis (high vs. low responder status) was controlled for age, sex, body mass index category, and the first 10 principal components. We conducted population-specific analyses and also performed a transethnic meta-analysis using METASOFT (24). Multivariate analysis was also performed on individuals combined across all three populations, including local genetic ancestries as additional covariates. The contribution of individual common and rare variants to multivariate association significance was evaluated by excluding common or rare variants in reduced and drop-one analyses. Variation in BDR explained by identified associations was calculated using McFadden pseudo R^2 (25). Detailed descriptions of these analyses are described in the online supplement.

We were unable to identify another age- and population-matched asthma cohort of sufficient size to replicate our drug response findings. Replication was nevertheless attempted in smaller cohorts and is described in detail in the online supplement.

Variant Prioritization and Functional Validation

To prioritize BDR-associated variants for further evaluation, an H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) assay was performed to identify variants overlapping with regulatory regions in primary bronchial smooth muscle cells (BSMCs). The Diverse Convergent Evidence approach (26) was then used to score BDR-associated variants using multiple levels of observational, bioinformatic, and laboratory evidence. Prioritized variants were further validated by luciferase enhancer assays to verify enhancer activity and by RNA sequencing to identify potential expression quantitative trait locus function. Detailed descriptions of these procedures are described in the online supplement.

Results

Descriptive Characteristics of Study Subjects

Descriptive characteristics for all study subjects ($n = 1,441$) are summarized in Table 1. Covariates and demographic variables were assessed for significant

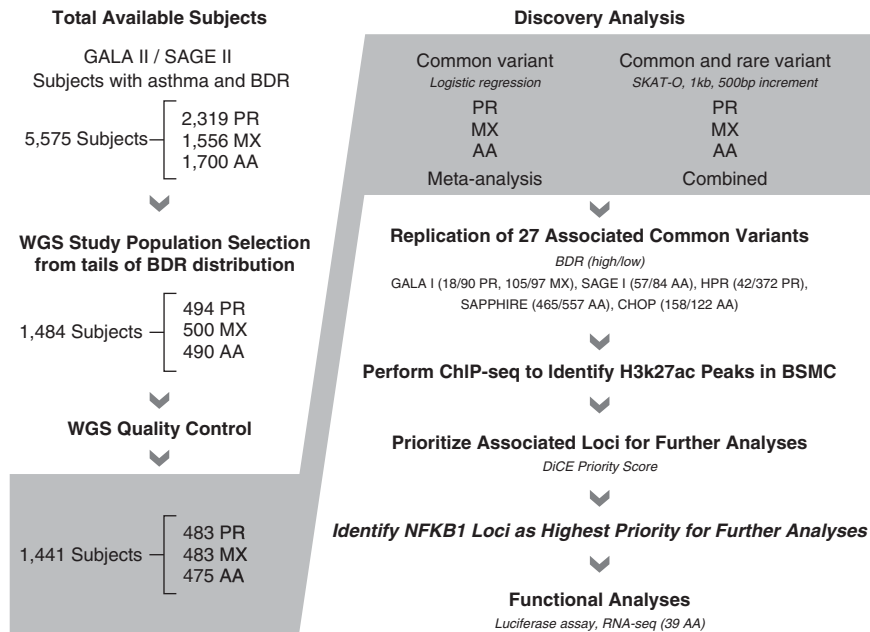


Figure 1. An overview of the main analyses performed in the current study. More detailed descriptions of the discovery and replication cohort demographics and analyses performed for common and rare variant analysis can be found in the METHODS section and the METHODS section in the online supplement. AA = African American; BDR = bronchodilator drug response; BSMC = bronchial smooth muscle cell; ChIP-seq = chromatin immunoprecipitation sequencing; CHOP = Children's Hospital of Philadelphia; DiCE = Diverse Convergent Evidence; GALA I = Genetics of Asthma in Latino Americans; GALA II = Genes-Environments and Admixture in Latino Americans; HPR = Hartford–Puerto Rico cohort; MX = Mexican; PR = Puerto Rican; SAGE = Study of African Americans, Asthma, Genes and Environments; SAPHIRE = Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity; WGS = whole-genome sequencing.

differences between high and low drug responders for each racial/ethnic group. Significant differences were found for age (Mexicans, $P < 0.001$), baseline lung function (pre-FEV₁% predicted, $P < 0.001$), total IgE ($P < 0.001$), and atopy. Pre-FEV₁% predicted was defined as the percentage of observed FEV₁ relative to the expected population average FEV₁ based on the Hankinson lung function prediction equations (27). Results and descriptions of genetic ancestry and genetic substructure are shown in Table 1, supplemental text E1, and Figures E2 and E3.

BDR Association Testing with Common Variants

Detailed descriptions of variant summary statistics are described in supplemental text E2, Figure E4, and Tables E2 and E3. Study design is described in Figure E5. All subsequent analyses in this study were performed only with biallelic SNPs. Throughout this study, low drug responders were assigned as the reference (i.e., control) group, and high drug responders were

classified as cases. We performed genome-wide association testing of common variants with BDR (dichotomized as high/low drug responders) for each population, adjusting by age, sex, body mass index category, and the first 10 principal components. We then performed a transesthetic meta-analysis across all three populations.

The commonly used GWAS P value threshold of 5×10^{-8} is derived from a Bonferroni correction under the assumption of 1,000,000 independent tests, based primarily on patterns of linkage disequilibrium (LD) from individuals of European descent. This threshold has been shown to be nongeneralizable for genetic studies among populations of non-European descent (28, 29). We therefore empirically calculated the effective number of independent tests for each population and for our transesthetic meta-analysis. The resulting genome-wide significance thresholds were 1.57×10^{-7} for Puerto Ricans, 2.42×10^{-7} for Mexicans, 9.59×10^{-8} for African Americans, and 3.53×10^{-7} for the transesthetic meta-analysis.

These numbers are highly concordant with WGS significance thresholds derived from the 1,000 Genomes sequencing data (28). Significance thresholds for discovery analyses in GWASs can often produce false-negative results. To minimize type II error, suggestive associations are often included in replication and functional validation studies. We identified suggestive associations based on the following widely used formula: $1/(\text{effective number of tests})$ (30).

Although no significant associations were identified from the population-specific analyses (see Figure E6), our transesthetic meta-analysis identified 10 unique loci (represented by 27 SNPs) significantly ($P < 3.53 \times 10^{-7}$) or suggestively ($P < 7.06 \times 10^{-6}$) associated with BDR status (Figure 2A and Table 2; see Table E4). After LD pruning, these 27 SNPs explain 23%, 16%, and 18% of the variation in BDR status in Puerto Ricans, Mexicans, and African Americans, respectively (see Table E5). To demonstrate that the results of our regression models were robust, a *post hoc* analysis was performed on the 27 SNPs by including Native American and African local ancestries as covariates. The association results before and after adjusting for local ancestry remained consistent (see Text E3, Table E6). We annotated all 27 SNPs by performing a thorough bioinformatic search in ENCODE, the NHGRI-EBL GWAS Catalog, and PubMed. Their previously reported lung-related phenotype associations and functional annotations are reported in Tables E7 and E8.

Two SNPs located on chromosome 5 (rs17834628 and rs35661809) were significantly associated with BDR ($P = 1.18 \times 10^{-8}$ and 3.33×10^{-8}). The direction of effect for these two variants is concordant across all three populations (Figure 2B; see Table E9). Figure 2C displays a LocusZoom plot of rs17834628 with 400-kb flanking regions. Three of the 27 SNPs were located within genes. Specifically, two SNPs are located in the third and fifth introns of *NFKB1* (rs28450894 and rs4648006), and a third SNP, rs16995064, mapped to intron 7 of *PLCB1* (Table 2). Based on 1,000 Genomes data, the low BDR-associated T allele of *NFKB1* (rs28450894) is found predominantly among African populations (minor allele frequency [MAF], 8.8–28.7%), followed by

Table 1. Study Population Description (N = 1,441)

Descriptive Statistics	Puerto Ricans (n = 483)			Mexicans (n = 483)			African Americans (n = 475)		
	High BDR	Low BDR	P Value	High BDR	Low BDR	P Value	High BDR	Low BDR	P Value
Number of subjects	239	244	—	243	240	—	233	242	—
Percent male	53.6	53.3	1.0	60.1	52.1	0.08	55.4	47.9	0.12
Median (IQR) age, yr	11.6 (9.7–14.8)	12.2 (10.1–15.2)	0.18	11.7 (9.6–14.0)	13.3 (10.6–16.0)	<0.001	13.8 (11.0–16.8)	13.8 (10.9–17.1)	0.48
Mean global ancestry proportions									
AFR	0.24	0.22	0.44	0.05	0.05	0.37	0.79	0.79	0.80
EUR	0.63	0.64	0.27	0.37	0.36	0.84	0.19	0.20	0.70
NAM	0.13	0.13	0.93	0.58	0.59	0.90	0.02	0.02	0.90
BMI category, n									
Obese, nonobese	76, 163	67, 177	0.32	100, 143	96, 144	0.85	82, 151	83, 159	0.85
Pre-FEV ₁ % predicted, n									
<80%, ≥80%	149, 90	56, 188	<0.001	43, 200	7, 233	<0.001	47, 186	6, 236	<0.001
Median (IQR) ΔFEV ₁ , %	21.2 (18.2–25.7)	5.0 (2.9–6.3)	—	12.7 (10.3–16.8)	3.6 (2.0–4.9)	—	15.5 (13.3–20.3)	3.3 (2.0–4.4)	—
Median (IQR) tlgE, ml	407.5 (126.8–952.8)	191.9 (50.5–542.2)	<0.001	247.5 (64.2–817.0)	105.7 (35.4–332.0)	<0.001	281.6 (97.3–552.4)	128.8 (36.6–351.3)	<0.001
Atopy, n	177	118	<0.001	155	117	<0.001	139	111	<0.001

Definition of abbreviations: AFR = African ancestry; BDR = bronchodilator drug response; BMI = body mass index; EUR = European ancestry; IQR = interquartile range; NAM = Native American ancestry; pre-FEV₁ % predicted = percentage of measured FEV₁ relative to predicted FEV₁ estimated by the Hankinson lung function prediction equations before administration of albuterol; tlgE = measure of total IgE from serum.

Atopy was defined as tlgE measurement ≥ 100 kU/L. ΔFEV₁ is a quantitative measure of BDR, measured as the percent change in baseline FEV₁ after administration of albuterol. High and low drug responders were chosen from the extremes of the BDR (ΔFEV₁) distribution.

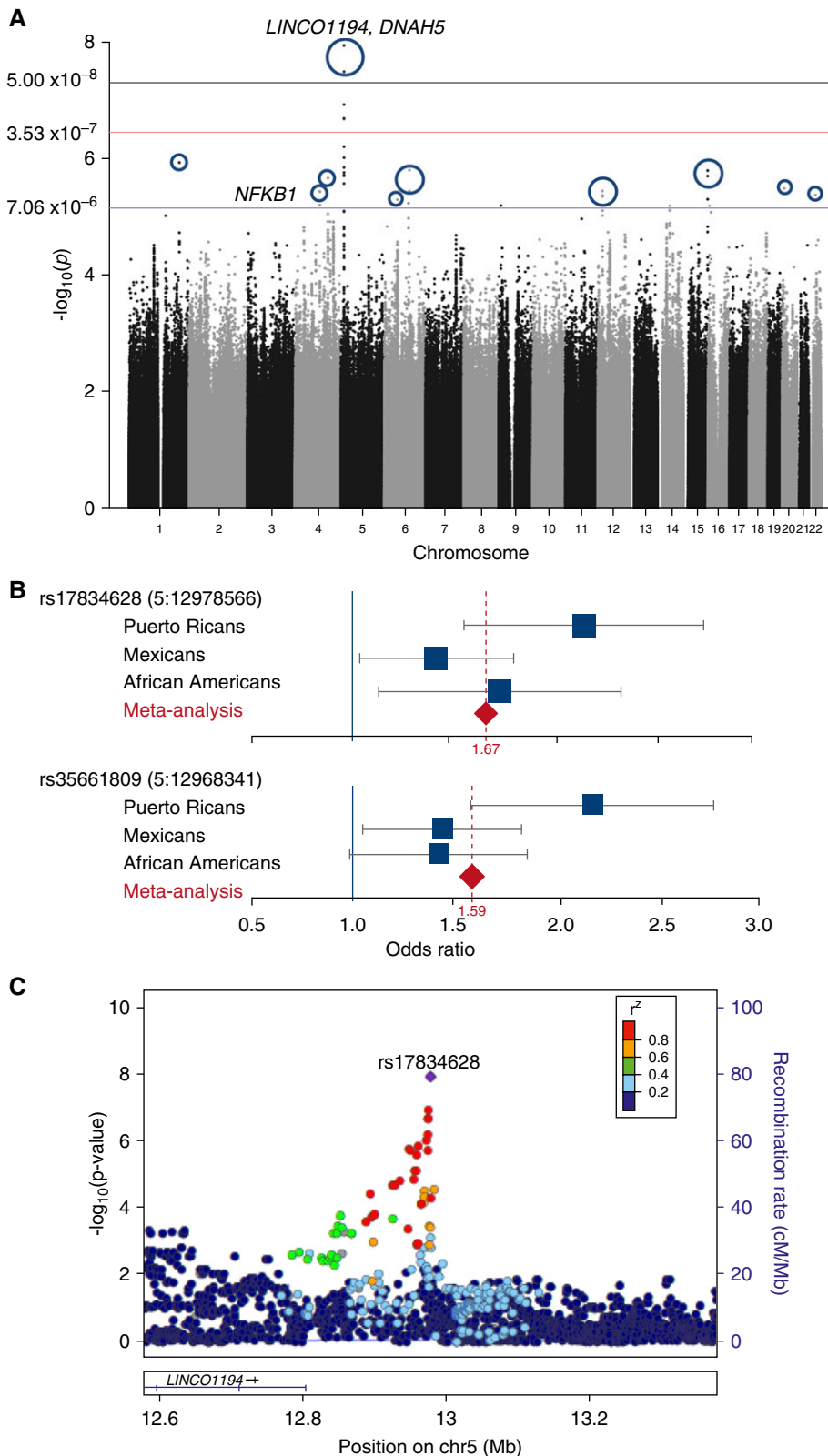


Figure 2. (A) Manhattan plot of the transethnic meta-analysis of single locus bronchodilator drug response association testing. Top 10 bronchodilator drug response–associated loci are circled. The black line indicates the universal genome-wide significance threshold (5.00×10^{-8}), the red line indicates the adjusted genome-wide significance threshold (3.53×10^{-7}), and the blue line indicates

European populations (MAF, 3.7–7.6%) and Puerto Ricans (MAF, 6.2%), and rare in Mexicans (MAF, 1.5%) (see Figure E7).

We were unable to identify age- and population-matched asthma cohorts with albuterol drug response data of sufficient sample size for replication. Nevertheless, we attempted to replicate all 27 SNPs in five independent populations (GALA I, SAGE I, HPR, SAPHIRE, and CHOP) separately and by meta-analysis (see Tables E10 and E11). Although none of the 27 SNPs were significantly associated with BDR status in our replication analyses (see Table E12), it is important to note that age-specific associations with asthma and asthma-related phenotypes have been reported (31, 32), and that children were not included in our largest replication cohort (SAPHIRE, median age >28) (see Table E10). All other replication populations were smaller than our discovery populations (ranging from 108 to 414 individuals per study), and some had an imbalance of cases and control subjects (see Table E10), which further diminished power for replication analyses.

Albuterol binds to β_2 -adrenergic receptors in BSMC, causing rapid onset of airway tissue relaxation and bronchodilation. BSMCs are therefore considered one of the most relevant cell types for molecular studies of BDR (33). We performed H3K27ac ChIP-seq experiments in primary BSMCs to identify potential regulatory regions marked by H3K27ac peaks. We then overlapped the H3K27ac peaks with the 27 BDR-associated SNPs and SNPs that were in high LD with them ($R^2 \geq 0.8$ within 1 Mb in any of our study populations). Variants overlapping with H3K27ac peaks may imply regulatory functions in BSMC (see Table E8).

We applied the Diverse Convergent Evidence (26) approach to prioritize the 27 BDR-associated SNPs for inclusion in further functional analyses (see Table E12). After integrating information from our WGS analysis, publicly available bioinformatics data, and ChIP-seq experiments in BSMCs, the *NFKB1* locus had the highest Diverse Convergent Evidence score, indicating that *NFKB1* had the strongest evidence of functional relevance to BDR variation (see Figure E8). Therefore, all further functional experiments were focused within this locus.

NFKB1 Functional Assays

Two H3K27ac ChIP-seq regions that overlapped with the BDR-associated *NFKB1* locus were tested for enhancer activity using luciferase enhancer assays in BSMCs (see Figure E9A and Table E13). One enhancer, *NFKB1* Region 2, showed significantly increased enhancer activity over empty vector ($\log_2$ 2.24-fold increase, $P = 1.59 \times 10^{-5}$, unpaired Wilcoxon test) (see Figure E9B).

Given the relevance of *NFKB1* in immune pathways and asthma, we also performed RNA sequencing experiments in African American children with asthma to verify whether the identified intronic *NFKB1* SNPs regulate expression of neighboring genes. Among genes within 1 Mb of rs28450894 meeting expression reliability cutoffs, we found that the low BDR-associated T allele of rs28450894 is significantly associated with decreased expression of *SLC39A8* in blood (see Figure E10) [$P = 0.0066$, FDR-adjusted $P = 0.0856$, $\log_2(\beta) = -0.327$].

Comparison with Previous BDR Association

We observed that two known BDR candidate genes, *ADCY9* and *CRHR2*, which replicated in a previous BDR GWAS performed in the full GALA II population (18), did not replicate in the current study (see Table E14). In that study, imputed GWAS array data were used to evaluate genetic associations with BDR measured as a continuous trait. To determine whether the discrepancy between findings was caused by data type (imputed array-based vs. WGS-based) or study design (continuous trait vs. extreme phenotype), the common variant analysis in the current analysis was repeated among the subset of samples with array-based and WGS data ($n = 1,414$ out of 1,441). Based on the top 1,000 BDR-associated SNPs from the current common variant analysis, there was perfect correlation between association P values generated from imputed array-based and WGS-based genotypes (Spearman correlation, 1.0), suggesting that data type is not the cause

of the observed discrepancy (see Figure E11A). Nearly all SNPs with high imputation R^2 exhibited high genotype concordance between array-based and WGS-based genotypes, confirming high imputation quality for most common SNPs ($\geq 99.7\%$; see Figures E11B and E11C). We also performed linear regression to analyze BDR as a continuous trait (ΔFEV_1) using imputed array-based data. The most significantly associated SNP identified in the current extreme phenotype analysis displayed the same direction of effect as analyzing BDR as a continuous trait (odds ratio, 1.67 in extreme phenotype analysis; $\beta = 0.51$ in continuous analysis). These observations indicate that the discrepancy between findings may be caused by differences in statistical power afforded by the different study designs (continuous trait vs. extreme phenotype). For common variant analyses, dichotomization of a continuous outcome results in a loss of statistical power (34). The opposite effect is observed in rare variant analyses. The extreme phenotype study design has been shown to increase power and the probability of identifying functional rare variants (35, 36). Also noteworthy is that the previously published results were discovered in one population (Puerto Ricans), whereas the results from our transethnic meta-analysis describe associations that are conserved across three populations (Puerto Ricans, Mexicans, and African Americans).

BDR Association Testing Using Common and Rare Variants

We tested the combined effects of common and rare variants on BDR using SKAT-O (37) to examine variants in 1-kb sliding windows with 500-bp increments. The same covariates used for common variant association testing were applied.

After determining the effective number of tests and adjusting for multiple comparisons on each population separately, we identified three population-specific loci associated with BDR at genome-wide significance levels; two were found in Mexicans on chromosome 1 and chromosome 11, and one in

African Americans on chromosome 19 (Figures 3A–3C and Table 3; see Table E15).

We also performed association testing across all three populations in a single analysis to maximize power. To minimize confounding by population substructure, association testing also included local genetic ancestry, defined as the proportions of Native American and African ancestries for the window under testing. Two loci on chromosomes 4 and 8 attained genome-wide significance ($P < 1.53 \times 10^{-7}$) (Figure 3D and Table 3). Sixty variants were identified from all SKAT-O regions reported in Table 3. Six of the 60 variants were located within predicted regulatory regions (see Table E16). Three variants on chromosome 11 identified in Mexicans overlap with a CTCF (transcriptional repressor) binding site and comprise a chromatin insulator region. The five regions identified in combined and population-specific SKAT-O analyses independently explained 4–8% of BDR variation in their respective populations (see Table E5).

To investigate whether common and rare variants both contributed to the BDR association P value, we performed reduced and drop-one SKAT analyses by excluding common or rare variants in the associated region one by one. Because excluding either all common or all rare variants would reduce the significance of the BDR association P value (see Table E17), the reduced (see Table E18) and drop-one (see Table E19) analyses indicated that both common and rare variants contribute to the significance of the BDR association.

Although we believe the sliding window method is the most appropriate approach for WGS data, we considered alternative grouping strategies for rare variants, including grouping by 1) genes from transcription start sites to transcription end sites, with or without 50-kb flanking regions; 2) transcription start sites with 20-kb flanking regions; and 3) H3K27ac ChIP-seq peaks from airway epithelial cells and airway smooth muscle cells. Association tests

Figure 2. (Continued). the suggestive significance threshold (7.06×10^{-6}). (B) Forest plot of the two most significantly associated SNPs, rs17834628 and rs35661809. The R^2 between these two SNPs is 0.93 in Puerto Ricans, 0.96 in Mexicans, and 0.66 in African Americans. (C) The most significantly associated SNP (rs17834628) is plotted with 400-kb flanking regions on either side. Dot color shows each SNP linkage disequilibrium with rs17834628 based on the 1,000 Genomes November 2014 admixed American population. Multiple SNPs in high linkage disequilibrium ($R^2 > 0.8$, red) reached a suggestive significance level.

Table 2. Results from Transethnic BDR Association Tests for Common Variants

Chr	Start	rsID	Effect Allele	OR (95% CI)	P Value	Effect Allele Frequency			Nearest Genes
						PR	MX	AA	
5	12978566	rs17834628	A	1.67 (1.29–2.16)	$1.18 \times 10^{-8*}$	0.32	0.42	0.17	LINC01194 (–173 kb), MIR4454 (–311 kb), CTNND2 (–1,074kb), DDAH5 (712 kb)
5	12968341	rs35661809	G	1.59 (1.20–2.10)	$3.33 \times 10^{-8*}$	0.34	0.43	0.24	LINC01194 (–163 kb), MIR4454 (–300 kb), CTNND2 (–1,064kb), DDAH5 (722 kb)
5	12975934	rs17237639	G	1.61 (1.30–2.00)	1.22×10^{-7}	0.31	0.43	0.16	LINC01194 (–171 kb), MIR4454 (–308 kb), CTNND2 (–1,072kb), DDAH5 (715 kb)
5	12975187	rs1017452	G	1.60 (1.31–1.96)	2.11×10^{-7}	0.31	0.43	0.16	LINC01194 (–170 kb), MIR4454 (–307 kb), CTNND2 (–1,071kb), DDAH5 (715 kb)
5	12975322	rs1017454	A	1.60 (1.31–1.96)	2.11×10^{-7}	0.31	0.43	0.16	LINC01194 (–170 kb), MIR4454 (–307 kb), CTNND2 (–1,071kb), DDAH5 (715 kb)
5	12975265	rs1017453	C	1.56 (1.25–1.95)	6.40×10^{-7}	0.31	0.42	0.16	LINC01194 (–170 kb), MIR4454 (–307 kb), CTNND2 (–1,071kb), DDAH5 (715 kb)
5	12972636	rs17237443	C	1.59 (1.28–1.97)	9.85×10^{-7}	0.29	0.42	0.11	LINC01194 (–170 kb), MIR4454 (–307 kb), CTNND2 (–1,071kb), DDAH5 (715 kb)
1	209324294	rs10746419	T	1.29 (0.75–2.25)	1.19×10^{-6}	0.49	0.54	0.53	MIR205HG (278 kb), MIR205 (281 kb), CAMK1G (433 kb), LAMB3 (464 kb)
5	12961545	rs17833938	A	1.56 (1.28–1.91)	1.45×10^{-6}	0.30	0.42	0.12	LINC01194 (–156 kb), MIR4454 (–294 kb), CTNND2 (–1,057kb), DDAH5 (729 kb)
6	104240500	rs13437006	C	1.56 (1.21–2.02)	1.61×10^{-6}	0.22	0.24	0.32	HACE1 (935 kb), LINC00577 (1,144kb), LIN28B (1,164kb)
15	101230457	rs1565749	A	1.66 (1.18–2.32)	1.64×10^{-6}	0.18	0.15	0.18	ASB7 (–39 kb), LINS1 (–88 kb), PRKXP1 (–131 kb)
5	12948369	rs34845041	T	1.56 (1.26–1.92)	1.77×10^{-6}	0.30	0.42	0.12	LINC01194 (–143 kb), MIR4454 (–280 kb), CTNND2 (–1,044kb), DDAH5 (742 kb)
5	12975108	rs1017451	T	1.55 (1.24–1.93)	1.96×10^{-6}	0.30	0.42	0.13	LINC01194 (–170 kb), MIR4454 (–307 kb), CTNND2 (–1,071kb), DDAH5 (715 kb)
5	12950432	rs62347395	G	1.55 (1.26–1.92)	2.02×10^{-6}	0.30	0.42	0.12	LINC01194 (–145 kb), MIR4454 (–282 kb), CTNND2 (–1,046kb), DDAH5 (740 kb)
15	101231049	rs57924834	A	1.59 (1.25–2.03)	2.04×10^{-6}	0.23	0.20	0.20	ASB7 (–39 kb), LINS1 (–89 kb), PRKXP1 (–132 kb)
4	137382142	rs17048684	A	1.8 (1.06–3.05)	2.20×10^{-6}	0.11	0.14	0.18	LINC00613 (–547 kb), PCDH18 (1,058kb)
5	12959598	rs1438293	G	1.55 (1.24–1.93)	2.73×10^{-6}	0.29	0.42	0.11	LINC01194 (–154 kb), MIR4454 (–292 kb), CTNND2 (–1,055kb), DDAH5 (731 kb)
20	8635168	rs16995064	G	1.96 (1.12–3.43)	3.30×10^{-6}	0.12	0.13	0.05	PLCB1 (intron 7), PLCB4 (415 kb)
12	19821401	rs66544720	T	0.66 (0.55–0.78)	3.66×10^{-6}	0.33	0.37	0.16	AEHP2 (–146 kb), PLEKHA5 (–292 kb)
6	104235591	rs6926020	C	1.57 (1.25–1.97)	3.68×10^{-6}	0.19	0.23	0.27	HACE1 (940 kb), LINC00577 (1,149kb), LIN28B (1,169kb)
4	103453535	rs28450894	T	0.47 (0.34–0.64)	3.75×10^{-6}	0.06	0.03	0.12	SLC39A8 (–187 kb), NFKB1 (intron 3), MANBA (99 kb)
4	103461559	rs4648006	T	0.47 (0.34–0.64)	3.75×10^{-6}	0.06	0.03	0.12	SLC39A8 (–195 kb), NFKB1 (intron 5), MANBA (91 kb)
22	27826429	rs60163793	G	2.01 (1.20–3.38)	4.30×10^{-6}	0.04	0.14	0.15	MM1 (318 kb), PTPN8 (421 kb)
12	19824386	rs7313907	C	0.66 (0.55–0.79)	4.35×10^{-6}	0.33	0.37	0.16	AEHP2 (–149 kb), PLEKHA5 (–295 kb)
12	19820677	rs11044754	A	0.66 (0.55–0.79)	4.54×10^{-6}	0.33	0.37	0.16	AEHP2 (–146 kb), PLEKHA5 (–291 kb)
15	10123236	rs5638658	A	1.61 (1.13–2.30)	5.08×10^{-6}	0.18	0.15	0.18	ASB7 (–41 kb), LINS1 (–91 kb), PRKXP1 (–134 kb)
6	54581204	rs13200833	A	0.66 (0.48–0.90)	5.15×10^{-6}	0.32	0.24	0.22	TINAG (–326 kb), MLIP (–450 kb), FAM83B (130 kb)

Definition of abbreviations: AA = African Americans; BDR = bronchodilator drug response; Chr = chromosome; CI = confidence interval; MX = Mexicans; OR = odds ratio; PR = Puerto Ricans. The top 10 unique loci (represented by 27 SNPs) significantly ($P < 3.53 \times 10^{-7}$) or suggestively ($P < 7.06 \times 10^{-6}$) associated with BDR status in our transethnic meta-analysis.

Chr and Start: chromosome locations of SNPs in GRCh37 coordinates. All significantly and suggestively associated common variants are presented. Nearest genes: the four nearest transcripts from RefSeq were identified and genes with multiple transcripts were reported only once, with the distance to the nearest transcript indicated in parentheses. Negative distances indicate upstream genes. Genes overlapping with BDR-associated SNPs are bold. High drug responders were assigned as cases.

*P values that achieve adjusted genome-wide significance for transethnic meta-analysis ($P < 3.53 \times 10^{-7}$).

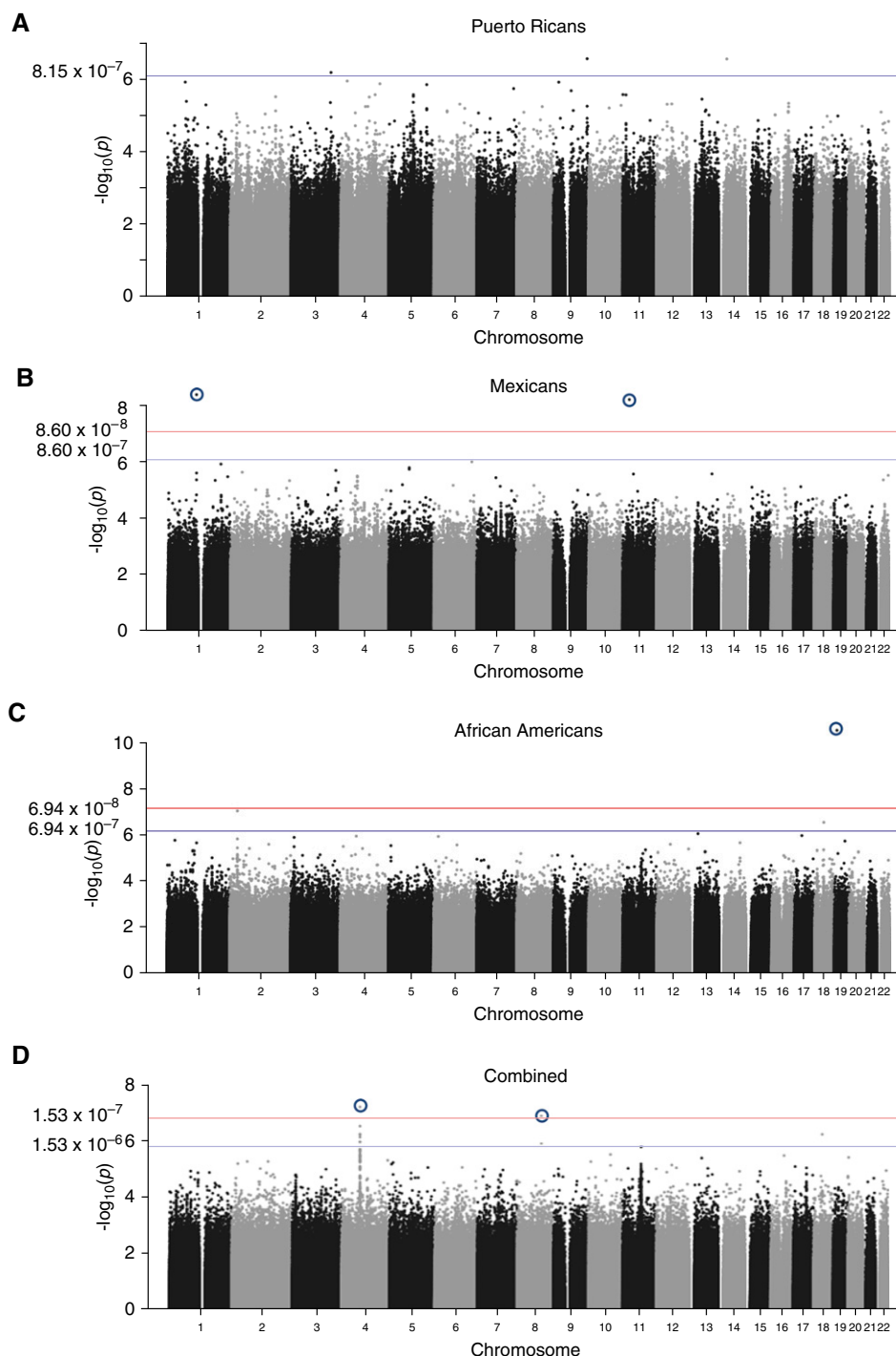


Figure 3. Manhattan plot of SKAT-O analysis of biallelic common and rare SNPs grouped by 1-kb windows sliding across chromosome 1–22 in (A) Puerto Ricans, (B) Mexicans, (C) African Americans, and (D) all populations combined. Bonferroni-corrected genome-wide and suggestive significance levels are marked by red and blue lines, respectively.

with these grouping strategies identified no further significant associations. Because these analyses were performed *post hoc*, multiple testing correction for testing alternative grouping strategies was not applied to the results of the sliding window approach.

Discussion

We identified population-specific and shared common and rare variants associated with BDR in three racially and ethnically diverse populations of children with

asthma. WGS provides comprehensive detection of common and rare variants in coding and noncoding regions. Combined, the 27 variants (after LD pruning) identified from our common variant analyses (Table 2) explained 23%, 16%,

Table 3. Association Testing for Combined Effects of Common and Rare Variants on BDR

Chr	Start	Stop	P Value	Population	nCommon	nRare	Nearest Genes
1	114177000	114178000	4.40×10^{-9}	MX	2	1	MAGI3 (intron 9) , <i>PHTF1</i> (62 kb), <i>RSBN1</i> (126 kb)
11	27507000	27508000	6.59×10^{-9}	MX	2	3	LOC105376671 (−3 kb), <i>LGR4</i> (−13 kb), <i>LIN7C</i> (8 kb)
19	10424000	10425000	3.12×10^{-11}	AA	1	2	ZGLP1 (−4 kb), <i>ICAM5</i> (−17 kb), FDX1L (intron 3) , <i>RAVER1</i> (2 kb)
4	73478000	73479000	6.25×10^{-8}	Combined	10	23	ADAMTS3 (−43 kb), <i>COX18</i> (441 kb)
8	97926000	97927000	1.32×10^{-7}	Combined	3	13	SDC2 (−302 kb), CPQ (intron 4) , <i>LOC101927066</i> (37 kb), <i>TSPYL5</i> (359 kb)

Definition of abbreviations: AA = African Americans; BDR = bronchodilator drug response; Chr = chromosome; MX = Mexicans; nCommon = number of common variants; nRare = number of rare variants.

Chr, Start, and Stop: GRCh37 chromosome coordinates. Combined: African Americans, Mexicans, and Puerto Ricans. Nearest genes: the four nearest transcripts from RefSeq were identified, and genes with multiple transcripts were reported once only, with distance to nearest transcripts indicated in parentheses. Negative distances indicate upstream genes. Genes that overlap with BDR-associated SNPs are bold. Details of all variants within the windows are presented in Table E16.

and 18% of the variation in BDR in Puerto Ricans, Mexicans, and African Americans, respectively (see Table E5). The five SKAT-O regions identified in our combined and population-specific analyses independently explained 4–8% of BDR variation (Tables 3; see E5). Our study represents an important investment from the NIH/NHLBI to improve racial and ethnic diversity in clinical and biomedical research.

Our transethnic common variant meta-analysis identified one locus on chromosome 5 that was associated with BDR at a genome-wide significance level ($P < 5 \times 10^{-8}$). The proximity of this BDR-associated locus to *DNAH5* and *LINC01194* is of particular interest. An SNP in *DNAH5* has been associated with total lung capacity in white subjects with chronic obstructive pulmonary disease (38). In a separate GWAS, the *DNAH5/LINC01194* locus was reported among Europeans to be associated with levels of IgE (39, 40), a biomarker associated with asthma endotypes. Baseline lung function (FEV_1) and total IgE levels are associated with asthma severity and can predispose an individual to lower BDR (7, 8, 41). We found two *NFKB1* intronic variants on chromosome 4 that were suggestively associated with BDR. The nuclear factor- κ B protein has a known role in allergic response, and various studies have demonstrated that the nuclear factor- κ B pathway is activated in patients with asthma, as reviewed by Edwards and colleagues (42).

ChIP-seq and functional enhancer assays in BSMCs imply that the regions containing the *NFKB1* intronic variants regulate expression of nearby

genes, but do not directly suggest that the genetic variants themselves alter the expression of nearby genes. However, the latter was supported by our RNA sequencing data, which showed that individuals with the low BDR-associated T allele displayed reduced expression of the neighboring *SLC39A8* gene.

Additionally, *SLC39A8* has been found to be responsive to cytokine treatment in airway epithelial cells (42) and had reduced expression in mice with allergic airway inflammation (43).

Studies have shown that up-regulation of *SLC39A8* is sufficient to protect the lung epithelium against tumor necrosis factor- α -induced cytotoxicity (44). Additionally, the higher frequency of the low BDR-associated allele (T allele of rs28450894 in *NFKB1*) in African populations suggests that the low BDR-associated allele tracks with African ancestry. This may explain why admixed populations with higher proportions of African ancestry (e.g., African Americans and Puerto Ricans) have lower BDR (8), and by extension may shed light on the higher asthma morbidity and mortality in these populations.

Another intronic variant (chromosome 20, rs16995064, *PLCB1* intron 7) suggestively associated with BDR is relevant to childhood asthma. *PLCB1* is differentially expressed among children with therapy-resistant asthma versus controlled persistent asthma or age-matched healthy control subjects (45). Silencing *PLCB1* inhibited the effect of lipopolysaccharide-induced endothelial cell inflammation by inhibiting expression of proinflammatory cytokines (46).

Additional functional studies are necessary

to establish the role of *NFKB1* and *PLCB1* on BDR.

We identified various combined effects of rare variants that were population-specific or shared across populations. Although some nearest genes show no known functional relationship to BDR (*MAGI3*, *LOC105376671*, *LIN7C*, and *CPQ*), the locus between *ADAMTS3* and *COX18* may be functionally relevant. The *ADAMTS3/COX18* locus was associated with β -adrenergic responses in murine cardiovascular-related traits (47). This locus was significantly associated with cardiac atrial weight in mice treated with the β -blocker atenolol and replicated under β -agonist isoproterenol treatment. These findings suggest that SNPs in this locus may modify β -adrenergic signaling pathways in BDR. In the present study, we also identified BDR association with rare variants within *CPQ*, which encodes a protein from the carboxypeptidase family. Although no previous BDR association has been identified for *CPQ*, another member of the carboxypeptidase family, carboxypeptidase A3, is expressed at higher levels in the airway epithelium among subjects with Th2-high asthma (48, 49). Further studies are necessary to determine the role of *CPQ* in BDR.

GWAS-based BDR-associated common variants in GALA II have previously been reported (18). However, those variants were not significantly associated with extreme BDR in the current study, likely because of differences in study design. The previous BDR GWAS used an array-based genotyping panel to examine children with asthma from the

entire BDR spectrum. In contrast, our current study sequenced the entire genome to investigate only the extremes of the BDR distribution. By repeating our current extreme phenotype analysis using a subset of individuals who had array and WGS data, we confirmed that the major discrepancy between the two studies was caused by study design instead of differences in data type. The contrast in results between GWAS and WGS caused by differences in study design implies that varied study designs are necessary for a comprehensive understanding of variants associated with asthma-related phenotypes and drug response. The extreme phenotype approach is recognized as one of the success factors in the study design of pharmacogenomic GWASs (50). Furthermore, the power gain from studying extreme phenotypes is much greater in analyses of rare variants compared with common variant studies (51). An extreme phenotype study design provides a cost-effective means for studying common and rare variant associations that may otherwise be missed when sampling across the entire phenotypic spectrum.

We did not identify BDR-associated variants from β_2 -adrenergic receptors signaling pathways. Instead, most of the BDR-associated genes identified in this study are related to lung function and allergic response, including total IgE levels and cytokine production in mast cells. This

suggests that BDR depends in part on the intrinsic state of airway smooth muscle cells. Genetic variation may determine individuals' intrinsic expression levels of candidate genes, which in turn determine whether their response to albuterol is beneficial.

Including admixed populations in WGS studies has important scientific implications. First, it facilitates discovery of genetic variation of multiple ancestral populations in a single study. Second, studying multiethnic admixed populations with ancestries that are underrepresented in existing genetic repositories can yield novel pharmacogenomic insights. For example, the widely popular PCSK9 inhibitors used to treat hypercholesterolemia were discovered by studying the genetics of African Americans but the drug development eventually benefited patients of all ethnicities (52). Finally, studying admixed populations, such as Mexicans, enhances the understanding of genetic variation in Native American ancestry, a population largely absent in major sequencing efforts.

We and others have documented the implications and challenges posed by the lack of non-European study populations in biomedical research (53–55). We made extensive efforts to test our top BDR-associated variants in other populations, but the unique characteristics of our discovery cohort (minority children with asthma who have BDR and WGS data)

posed significant challenges for finding comparable replication cohorts. Testing ancestry-dependent pharmacogenetic variants without age- and ethnicity-matched replication populations poses an impossible statistical bar for publication. We therefore conducted additional analyses involving bioinformatics and experimental assays (56). These challenges highlight the need to include more racially/ethnically diverse populations in all clinical and biomedical research.

In an era of precision medicine, addressing questions about the impact of genetic factors on therapeutic drug response in globally diverse populations is essential for making precision medicine socially and scientifically precise (57). This study advances the understanding of genetic analysis in admixed populations and helps to lay the foundation of precision medicine for understudied and racially and ethnically diverse populations. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank the New York Genome Center investigators and teams for whole-genome sequencing sample preparation, quality control, data generation, data processing, and initial joint genotyping. They gratefully acknowledge the studies and participants who provided biologic samples and data for TOPMed. They also gratefully acknowledge the contributions of the investigators of the NHLBI TOPMed Consortium (<https://www.nhlbiwgs.org/topmed-banner-authorship>).

References

1. World Health Organization. Asthma. 2017 [accessed 2017 Sep 12]. Available from: <http://www.who.int/mediacentre/factsheets/fs307/en/>.
2. Barr RG, Aviles-Santa L, Davis SM, Aldrich TK, Gonzalez F 2nd, Henderson AG, *et al*. Pulmonary disease and age at immigration among Hispanics: results from the Hispanic Community Health Study/Study of Latinos. *Am J Respir Crit Care Med* 2016;193:386–395.
3. Akinbami L. Asthma prevalence, health care use and mortality: United States, 2003–05. 2015 [accessed 2017 Sep 12]. Available from: <http://www.cdc.gov/nchs/data/hestat/asthma03-05/asthma03-05.htm>.
4. Palmer LJ, Silverman ES, Weiss ST, Drazen JM. Pharmacogenetics of asthma. *Am J Respir Crit Care Med* 2002;165:861–866.
5. Eggleston PA, Malveaux FJ, Butz AM, Huss K, Thompson L, Kolodner K, *et al*. Medications used by children with asthma living in the inner city. *Pediatrics* 1998;101:349–354.
6. Finkelstein JA, Lozano P, Farber HJ, Miroshnik I, Lieu TA. Underuse of controller medications among Medicaid-insured children with asthma. *Arch Pediatr Adolesc Med* 2002;156:562–567.
7. Burchard EG, Avila PC, Nazario S, Casal J, Torres A, Rodriguez-Santana JR, *et al*. Genetics of Asthma in Latino Americans (GALA) Study. Lower bronchodilator responsiveness in Puerto Rican than in Mexican subjects with asthma. *Am J Respir Crit Care Med* 2004;169:386–392.
8. Naqvi M, Thyne S, Choudhry S, Tsai HJ, Navarro D, Castro RA, *et al*. Ethnic-specific differences in bronchodilator responsiveness among African Americans, Puerto Ricans, and Mexicans with asthma. *J Asthma* 2007;44:639–648.
9. Wechsler ME, Castro M, Lehman E, Chinchilli VM, Sutherland ER, Denlinger L, *et al*. NHLBI Asthma Clinical Research Network. Impact of race on asthma treatment failures in the asthma clinical research network. *Am J Respir Crit Care Med* 2011;184:1247–1253.
10. Nelson HS, Weiss ST, Bleecker ER, Yancey SW, Dorinsky PM; SMART Study Group. The Salmeterol Multicenter Asthma Research Trial: a comparison of usual pharmacotherapy for asthma or usual pharmacotherapy plus salmeterol. *Chest* 2006;129:15–26.
11. McGeachie MJ, Stahl EA, Himes BE, Pendergrass SA, Lima JJ, Irvin CG, *et al*. Polygenic heritability estimates in pharmacogenetics: focus on asthma and related phenotypes. *Pharmacogenet Genomics* 2013;23:324–328.
12. Nieminen MM, Kaprio J, Koskenvuo M. A population-based study of bronchial asthma in adult twin pairs. *Chest* 1991;100:70–75.

13. Fagnani C, Annesi-Maesano I, Brescianini S, D'Ippolito C, Medda E, Nisticò L, *et al.* Heritability and shared genetic effects of asthma and hay fever: an Italian study of young twins. *Twin Res Hum Genet* 2008;11:121–131.
14. Himes BE, Jiang X, Hu R, Wu AC, Lasky-Su JA, Klanderman BJ, *et al.* Genome-wide association analysis in asthma subjects identifies SPATS2L as a novel bronchodilator response gene. *PLoS Genet* 2012;8:e1002824.
15. Duan QL, Lasky-Su J, Himes BE, Qiu W, Litonjua AA, Damask A, *et al.* A genome-wide association study of bronchodilator response in asthmatics. *Pharmacogenomics J* 2014;14:41–47.
16. Israel E, Lasky-Su J, Markezich A, Damask A, Szeffler SJ, Schuemann B, *et al.*; SHARP Investigators. Genome-wide association study of short-acting β_2 -agonists. A novel genome-wide significant locus on chromosome 2 near ASB3. *Am J Respir Crit Care Med* 2015;191:530–537.
17. Padhukasahasram B, Yang JJ, Levin AM, Yang M, Burchard EG, Kumar R, *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;14:365–371.
18. Drake KA, Torgerson DG, Gignoux CR, Galanter JM, Roth LA, Huntsman S, *et al.* A genome-wide association study of bronchodilator response in Latinos implicates rare variants. *J Allergy Clin Immunol* 2014;133:370–378.
19. Mak AC, White MJ, Eng C, Hu D, Huntsman S, Oh SS, *et al.* Whole genome sequencing to identify genetic variation associated with bronchodilator response in minority children with asthma [abstract]. *Am J Respir Crit Care Med* 2016;193:A3709.
20. Mak AC, White MJ, Szpiech ZA, Pino-Yanes M, Oh SS, Hu D, *et al.* Whole genome sequencing of pharmacogenetic drug response in ethnically diverse children with asthma. Presented at the American Society of Human Genetics Annual Meeting. Oct 18–22, 2016, Vancouver, Canada. Abstract 1660W.
21. Mak AC, White MJ, Szpiech ZA, Eckalbar W, Oh SS, Pino-Yanes M, *et al.* Whole genome sequencing of pharmacogenetic drug response in racially and ethnically diverse children with asthma. Presented at the American Society of Human Genetics Annual Meeting. Oct 17–21, 2017, Orlando, FL. Abstract 1941F.
22. Mak AC, White MJ, Szpiech ZA, Eckalbar W, Pino-Yanes M, Oh SS, *et al.* Whole genome sequencing study on bronchodilator drug response in ethnically diverse children with asthma [abstract]. *Am J Respir Crit Care Med* 2017;195:A4969.
23. Borrell LN, Nguyen EA, Roth LA, Oh SS, Tcheurekdjian H, Sen S, *et al.* Childhood obesity and asthma control in the GALA II and SAGE II studies. *Am J Respir Crit Care Med* 2013;187:697–702.
24. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* 2011;88:586–598.
25. McFadden D. Quantitative methods for analysing travel behavior of individuals: some recent developments. In: Hensher DA, Stopher PR, editors. Behavioural travel modelling. London: Croom Helm; 1979. pp. 279–318.
26. Ciesielski TH, Pendergrass SA, White MJ, Kodaman N, Sobota RS, Huang M, *et al.* Diverse convergent evidence in the genetic analysis of complex disease: coordinating omic, informatic, and experimental evidence to better identify and validate risk factors. *BioData Min* 2014;7:10.
27. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–187.
28. Sobota RS, Shriner D, Kodaman N, Goodloe R, Zheng W, Gao YT, *et al.* Addressing population-specific multiple testing burdens in genetic association studies. *Ann Hum Genet* 2015;79:136–147.
29. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008;32:381–385.
30. Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 2008;9:516.
31. Castro-Giner F, de Cid R, Gonzalez JR, Jarvis D, Heinrich J, Janson C, *et al.* Positionally cloned genes and age-specific effects in asthma and atopy: an international population-based cohort study (ECRHS). *Thorax* 2010;65:124–131.
32. White MJ, Risse-Adams O, Goddard P, Contreras MG, Adams J, Hu D, *et al.* Novel genetic risk factors for asthma in African American children: precision medicine and the SAGE II Study. *Immunogenetics* 2016;68:391–400.
33. Johnson M. The beta-adrenoceptor. *Am J Respir Crit Care Med* 1998;158:S146–S153.
34. Cumberland PM, Czanner G, Bunce C, Doré CJ, Freemantle N, García-Fiñana M; Ophthalmic Statistics Group. Ophthalmic statistics note: the perils of dichotomising continuous variables. *Br J Ophthalmol* 2014;98:841–843.
35. Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet Epidemiol* 2011;35:790–799.
36. Guey LT, Kravic J, Melander O, Burt NP, Laramie JM, Lyssenko V, *et al.* Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genet Epidemiol* 2011;35:236–246.
37. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 2012;13:762–775.
38. Lee JH, McDonald ML, Cho MH, Wan ES, Castaldi PJ, Hunninghake GM, *et al.*, COPDGen and ECLIPSE Investigators. DNAH5 is associated with total lung capacity in chronic obstructive pulmonary disease. *Respir Res* 2014;15:97.
39. Ortiz RA, Barnes KC. Genetics of allergic diseases. *Immunol Allergy Clin North Am* 2015;35:19–44.
40. Ramasamy A, Curjuric I, Coin LJ, Kumar A, McArdle WL, Imboden M, *et al.* A genome-wide meta-analysis of genetic variants associated with allergic rhinitis and grass sensitization and their interaction with birth order. *J Allergy Clin Immunol* 2011;128:996–1005.
41. Naqvi M, Choudhry S, Tsai HJ, Thyne S, Navarro D, Nazario S, *et al.* Association between IgE levels and asthma severity among African American, Mexican, and Puerto Rican patients with asthma. *J Allergy Clin Immunol* 2007;120:137–143.
42. Edwards MR, Bartlett NW, Clarke D, Birrell M, Belvisi M, Johnston SL. Targeting the NF-kappaB pathway in asthma and chronic obstructive pulmonary disease. *Pharmacol Ther* 2009;121:1–13.
43. Zhen G, Park SW, Nguyenvu LT, Rodriguez MW, Barbeau R, Paquet AC, *et al.* IL-13 and epidermal growth factor receptor have critical but distinct roles in epithelial cell mucin production. *Am J Respir Cell Mol Biol* 2007;36:244–253.
44. Besecker B, Bao S, Bohacova B, Papp A, Sadee W, Knoell DL. The human zinc transporter SLC39A8 (Zip8) is critical in zinc-mediated cytoprotection in lung epithelia. *Am J Physiol Lung Cell Mol Physiol* 2008;294:L1127–L1136.
45. Persson H, Kwon AT, Ramilowski JA, Silberberg G, Söderhäll C, Orsmark-Pietras C, *et al.* Transcriptome analysis of controlled and therapy-resistant childhood asthma reveals distinct gene expression profiles. *J Allergy Clin Immunol* 2015;136:638–648.
46. Lin YJ, Chang JS, Liu X, Tsang H, Chien WK, Chen JH, *et al.* Genetic variants in PLCB4/PLCB1 as susceptibility loci for coronary artery aneurysm formation in Kawasaki disease in Han Chinese in Taiwan. *Sci Rep* 2015;5:14762.
47. Hersch M, Peter B, Kang HM, Schüpfer F, Abriel H, Pedrazzini T, *et al.* Mapping genetic variants associated with beta-adrenergic responses in inbred mice. *PLoS One* 2012;7:e41032.
48. Dougherty RH, Sidhu SS, Raman K, Solon M, Solberg OD, Caughey GH, *et al.* Accumulation of intraepithelial mast cells with a unique protease phenotype in T(H)2-high asthma. *J Allergy Clin Immunol* 2010;125:1046–1053.
49. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, *et al.* Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc Natl Acad Sci USA* 2007;104:15858–15863.

50. Motsinger-Reif AA, Jorgenson E, Relling MV, Kroetz DL, Weinshilboum R, Cox NJ, *et al.* Genome-wide association studies in pharmacogenomics: successes and lessons. *Pharmacogenet Genomics* 2013;23:383–394.
51. Peloso GM, Rader DJ, Gabriel S, Kathiresan S, Daly MJ, Neale BM. Phenotypic extremes in rare variant study designs. *Eur J Hum Genet* 2016;24:924–930.
52. Hall SS. Genetics: a gene of rare effect. *Nature* 2013;496:152–155.
53. Oh SS, Galanter J, Thakur N, Pino-Yanes M, Barcelo NE, White MJ, *et al.* Diversity in clinical and biomedical research: a promise yet to be fulfilled. *PLoS Med* 2015;12:e1001918.
54. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016;538:161–164.
55. Hindorff LA, Bonham VL, Brody LC, Ginoza MEC, Hutter CM, Manolio TA, *et al.* Prioritizing diversity in human genomics research. *Nat Rev Genet* 2018;19:175–185.
56. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014;508:469–476.
57. Oh SS, White MJ, Gignoux CR, Burchard EG. Making precision medicine socially precise: take a deep breath. *Am J Respir Crit Care Med* 2016;193:348–350.

Whole genome sequencing of pharmacogenetic drug response in racially diverse children with asthma

AUTHORS

Angel C.Y. Mak^{1¶*}, Marquitta J. White^{1¶}, Walter L. Eckalbar^{1¶}, Zachary A. Szpiech^{2¶}, Sam S. Oh¹, Maria Pino-Yanes^{3,4}, Donglei Hu¹, Pagé Goddard¹, Scott Huntsman¹, Joshua Galanter^{1,#a}, Ann Chen Wu^{5,6}, Blanca E. Himes⁷, Soren Germer⁸, Julia M. Vogel^{8,#b}, Karen L. Bunting⁸, Celeste Eng¹, Sandra Salazar¹, Kevin L. Keys¹, Jennifer Liberto¹, Thomas J. Nuckton¹, Thomas A. Nguyen¹, Dara G. Torgerson⁹, Pui-Yan Kwok^{10,11}, Albert M. Levin¹², Juan C. Celedón¹³, Erick Forno¹³, Hakon Hakonarson^{14,15}, Patrick M. Sleiman^{14,15}, Amber Dahlin⁵, Kelan G. Tantisira⁵, Scott T. Weiss⁵, Denise Serebrisky¹⁶, Emerita Brigino-Buenaventura¹⁷, Harold J. Farber¹⁸, Kelley Meade¹⁹, Michael A. Lenoir²⁰, Pedro C. Avila^{21,#c}, Saunak Sen^{1,#d}, Shannon M. Thyne²², William Rodriguez-Cintron²³, Cheryl A. Winkler²⁴, Andrés Moreno-Estrada²⁵, Karla Sandoval²⁵, Jose R. Rodriguez-Santana²⁶, Rajesh Kumar^{27,28}, L. Keoki Williams^{29,30}, Nadav Ahituv^{2,11}, Elad Ziv¹, Max A. Seibold³¹, Robert B. Darnell^{8,32,33}, Noah Zaitlen¹, Ryan D. Hernandez^{2,10,34}, Esteban G. Burchard^{1,2} on behalf of the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium[^]

ONLINE DATA SUPPLEMENT

SUPPLEMENTARY TEXT

Text E1. Additional descriptive characteristics of study subjects

We estimated genetic ancestry for all participants. The major ancestry proportions in Puerto Ricans, Mexicans and African Americans are European, Native American and African ancestries, respectively (**Table 1, Figure E2**). Analysis of genetic substructure using principal component analysis (PCA) demonstrated that the three populations displayed the characteristic spectrum of ancestry found in admixed populations (**Figure E3**).

Text E2. Variant summary statistics

Genetic variant summary statistics revealed that the average number of variants by population corresponded to the proportion of African ancestry: the most variants were found among African Americans, followed by Puerto Ricans and Mexicans (**Figure E4A, Table E2**).

The majority of observed variants (>90%) were SNPs. The union of biallelic SNPs from all individuals in each population varied from 28.1M among Mexicans and 29.2M among Puerto Ricans, to 36.3M among African Americans. Approximately 65% of biallelic SNPs were rare (non-reference allele frequency < 1%, **Figure E4B, Table E2**). Biallelic SNPs that were population-specific (i.e., SNPs found in only one population) accounted for 23% (6.68M / 29.2M in Puerto Ricans) to 39% (14.1M / 36.3M in African Americans) of the biallelic SNPs observed in each population. Over 99% of the population-specific SNPs had a non-reference allele frequency less than

5% and the majority of these population-specific SNPs (69% to 82%) were also singletons (**Figure E4C, Table E2**). Based on dbSNP build 150, an average of 8% of biallelic SNPs were novel (**Figure E4D, Table E2**).

In all three populations, 99% of the biallelic SNPs were observed in noncoding regions. Based on the Combined Annotation Dependent Depletion (CADD) score (1), which estimates a variant's deleteriousness, over 99% of highly deleterious biallelic SNPs (CADD score ≥ 25) were observed in coding regions regardless of ethnicity (**Table E3**). This may be due to the relatively limited availability of functional annotations as positive training data for CADD to estimate deleteriousness in non-coding regions (2). The percentage of singletons in these highly deleterious biallelic SNPs varied from 51% (Puerto Ricans) to 70% (Mexicans).

Text E3. Common variant trans-ethnic BDR association tests adjusted and unadjusted for local ancestry

A post-hoc analysis including Native American and African local ancestries as covariates in the regression model for the 27 common SNPs listed in **Table 2**. The results of these analyses are shown in **Table E6**. The magnitude, direction, and significance of the effect estimates before and after adjusting for local ancestry remained consistent. Correlation tests of the p -values and odds ratios before and after adjustment for local ancestry were statistically significant (odds ratios: correlation = 0.999, correlation p -value = 2.20×10^{-16} ; p -values: correlation = 0.603, correlation p -value = 8.66×10^{-4}).

SUPPLEMENTARY METHODS

Subject selection

A total of 1,500 samples were selected from GALA II and SAGE II, with 500 samples each from Puerto Ricans, Mexicans and African Americans. For each population, we selected 250 samples from the tails of the BDR distribution (i.e., 250 high responders and 250 low responders). Sample quality control filtering was applied and resulted in 1,441 samples for downstream association analyses (see “WGS data processing and quality control”). For each of our three populations, we determined the highest ΔFEV_1 among those with low BDR status and the lowest ΔFEV_1 among those with high BDR status. These population-specific ΔFEV_1 thresholds were then used for determining high and low BDR status in the replication analyses.

Study design and power calculation

This project was conducted within the NHLBI TOPMed Consortium to generate whole genome sequencing (WGS) data for 1,500 samples. To determine power for variant discovery, we performed detailed population simulations that incorporated demographic history with admixture, and natural selection on deleterious nonsynonymous mutations (3-5). We discovered that sampling 500 individuals each from three different populations would lead to the discovery of a larger fraction of the genetic variation compared to sampling all individuals from a single population (Figure E5A). Previous research also suggested that the extreme phenotype study design would further increase our power and probability of identifying functional variants (6-8). Based on data restrictions from TOPMed and the rationale above, we

adopted the extreme phenotype study design and selected 500 individuals from both tails of the BDR distribution.

We evaluated power for pooled association testing in SKAT by randomly drawing 1,000 3 kb regions from a 1 Mb chromosome generated under a coalescent model (Figure E5B). Causal rare variants were assumed to be at frequencies $< 3\%$. Effect sizes of causal variants were modeled under a maximum effect size of 6 for the rarest of variants, with the effects of causal variants decreasing with increasing frequency. A maximum standard normal effect size of 6 was determined based on observed regression coefficients of rare coding and non-coding variants identified in our BDR GWAS in GALA II (9). Using a logistic model to compare BDR extremes (high vs. low) in $n=1,500$ individuals, we predicted 100% power to identify genes/loci at $p < 0.01$ with 25% causal rare variants, and 90% power to identify genes/loci at $p < 0.001$.

Bronchodilator response measurements

Spirometry was performed and BDR was calculated in GALA II (9) and SAGE II (10) as previously described. Briefly, BDR was calculated according to American Thoracic Society/European Respiratory Society guidelines (11). BDR was only measured in asthma cases (study participants who had been diagnosed with asthma by a physician). Forced expiratory volume in the first second (FEV_1) was obtained by spirometry prior to administering albuterol (pre- FEV_1) and again after two doses of albuterol (post- FEV_1), with a 15-minute waiting period between the two doses. Participants were given 4 puffs of albuterol for their first dose. The second dose was based on age: participants under age 16 years received 2 puffs while older

participants received 4. Spirometry was performed fifteen minutes after each dose of albuterol to obtain post-FEV₁ measurements. The higher of the two post-FEV₁ was used for determining BDR. BDR was calculated as the percent change in FEV₁ (Δ FEV₁) after albuterol administration; that is, $BDR (\Delta FEV_1) = (post-FEV_1 - pre-FEV_1) / pre-FEV_1$.

Figure E1 summarizes the BDR distribution of the 1,441 individuals who passed WGS data quality control (see “WGS data processing and quality control”). The Δ FEV₁ cutoffs for high and low responders are as follows: high responders (Δ FEV₁ > 16.29 for Puerto Ricans, > 8.55 for Mexicans and > 11.81 for African Americans); low responders (Δ FEV₁ < 7.23 for Puerto Ricans, < 6.05 for Mexicans and < 5.53 for African Americans).

Analysis on descriptive data of study subjects

Dichotomous variables were tested for association with BDR using Fisher’s exact test. Continuous variables were tested for normality using the Shapiro-Wilk test. Normally and non-normally distributed continuous variables were tested using Student’s t-test and the Wilcoxon rank-sum test, respectively.

Sample quality control and whole genome sequencing

Genomic DNA was extracted and purified from whole blood using Wizard® Genomic DNA Purification Kits (Promega, Madison, WI, USA). DNA samples were quantified by fluorescence using the Quant-iT PicoGreen dsDNA assay (ThermoFisher Scientific, Waltham, MA, USA) on a Spectramax fluorometer (Molecular Devices, Sunnyvale, CA, USA). Sample integrity was ascertained using the Fragment

Analyzer™ (Advanced Analytical Technologies, Inc., Ankeny, IA, USA). Samples passing QC were genotyped using the HumanCoreExome-24 array (Illumina®, San Diego, CA, USA). Genotyping results were analyzed using VerifyIDintensity (12) to flag sample contamination. Sequencing libraries were constructed using the TruSeq PCR-free DNA HT Library Preparation Kit (Illumina®, San Diego, CA, USA) with 500ng DNA input. Briefly, genomic DNA was sheared using a Covaris sonicator (Covaris, Woburn, MA, USA), followed by end-repair and bead-based size selection of fragmented molecules. Selected fragments were then A-tailed and sequence adaptors were ligated onto the fragments, followed by a final bead purification of the libraries. Final libraries were reviewed for size distribution using Fragment Analyzer and quantified by qPCR (Kapa Biosystems, Wilmington, MA, USA). Libraries were sequenced on a HiSeq X system (Illumina®, San Diego, CA, USA) with v2 chemistry, using a paired-end read length of 150 bp, to a minimum of 30x mean genome coverage.

WGS data processing and quality control

Sequencing data were demultiplexed using bcl2fastq version 2.16.0.10 (Illumina®, San Diego, CA, USA) and aligned to human reference hs37d5 with decoy sequences using BWA-MEM v0.7.8 (13). Data were further processed using the GATK best-practices v3.2-2 pipeline (14). Quality control procedures included marking of duplicate reads using Picard tools v1.83 (<http://picard.sourceforge.net>), realignment around indels, and base quality recalibration using 1000 Genomes Phase 1 high confidence SNPs, HapMap v3.3, dbSNP v137, 1000 Genomes omni2.5, 1000 Genomes Phase 1 indels, and both Mills and 1000 Genomes gold standard

indels. Single-sample genotypes were called using GATK HaplotypeCaller followed by joint genotyping of all subjects. The resulting multi-sample Variant Call Format (VCF) file was used for variant quality score recalibration (VQSR). A 99.8% truth sensitivity tranche level was used for SNPs and 99.0% for indel variants. SNP calls were used to check for sample contamination using VerifyBAMId (12), and sample identity was confirmed by requiring > 99.5% concordance with SNP array (HumanCoreExome-24 array) genotypes.

As part of NIH's Trans-Omics for Precision Medicine (TOPMed) Program, BAM files were submitted to the Informatics Resource Center (IRC) at the University of Michigan. All 1,484 samples sequenced passed TOPMed's IRC quality control metrics (mean genome coverage >30X; >95% of genome covered at >10X; and <3% contamination).

VCF-level variants were filtered by GATK version 3.4.46 and VCFtools version 0.1.14 (15). Variants were filtered according to the following procedures: (1) remove variants that were not indicated as "PASS" in the VCF FILTER column, (2) remove variants in low complexity regions (16) (downloaded from <https://github.com/lh3/varcmp/tree/master/scripts/LCR-hs37d5.bed.gz>), and (3) keep sample genotypes that have minimum read depths of 10 and genotype qualities of 20 ($DP \geq 10$ and $GQ \geq 20$). The ratio of homozygous to heterozygous variants (hom/het), ratio of transitions to transversions (Ti/Tv), and other variant summary statistics were generated using GATK VariantEval. VCF files were normalized using BCFtools from the Samtools suite (17) and converted into PLINK format using PLINK 1.9 software (18). Genotype consistency between WGS data and

previously published Axiom® Genome-Wide LAT 1 array (Affymetrix, Santa Clara, CA) genotype data (dbGaP phs000920.v1.p1 and phs000921.v1.p1) was assessed using VCFtools (15). Individuals with percentage consistency three S.D. below the mean (< 96.3%) were removed (N=7, **Figure E12**). Cryptic relatedness was detected using REAP (19). Global ancestry and allele frequency used by REAP were estimated using ADMIXTURE in supervised mode (20). Related individuals (kinship coefficient > 0.044, corresponding to a third degree relationship (21)) were excluded from further analysis (N=36), yielding a final sample size of 1,441 for downstream analysis. Downstream analyses were performed only on biallelic SNPs that passed all quality filters mentioned above and had less than 10% genotype missingness. The 10% genotype missingness filter was applied per population except for the rare variant analysis, which was performed with all three populations combined (see **Supplementary Methods** section “Multi-variant analyses of combined effects of rare BDR variants”).

In summary, 1,500 samples were selected for this study. We excluded 16 samples with WGS data previously sequenced in the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) (22) but not joint genotyped in TOPMed to avoid batch effects. Further sample quality control filtering resulted in a final study population of 1,441 samples for downstream analyses.

Principal component analysis

Principal component analysis (PCA) was performed to control for hidden population substructure using EIGENSTRAT’s smartpca program (23). After using PLINK 1.9 to remove biallelic SNPs with low minor allele frequency ($MAF \leq 0.05$) and in linkage

disequilibrium ($R^2 > 0.5$ in a 50-SNP window with a shift size of 5 SNPs), 710,256 variants remained for input into smartpca.

Local ancestry estimation

Reference genotypes for European and African ancestries were obtained from the Axiom® Genotype Data Set (24). SNPs with less than a 95% call rate were removed. Since no Native American reference samples are available in the HapMap database, reference genotypes for Native American ancestry were generated from 71 Native American individuals previously genotyped on the Axiom® Genome-Wide LAT 1 array (9).

To call local ancestry tracts, we first created a subset of our WGS data corresponding to sites found on the Axiom® Genome-Wide LAT 1 array, leaving 765,321 markers. Using PLINK 1.9, we merged these data with our European (CEU), African (YRI), and Native American (NAM) reference panels, which overlapped at 434,145 markers. After filtering multi-allelic SNPs and SNPs with > 10% missing data, we obtained a final merged dataset of 428,644 markers. We phased all samples using SHAPEIT2 (25) and called local ancestry tracts jointly with RFMix (26) under a three-way admixture model based on the African, European, and Native American reference genotypes described above.

Variant annotation

TOPMed freeze 2 and 3 variants were annotated using the WGSa annotation pipeline (27). Annotated VCF files were downloaded from the TOPMed Data Coordinating Center SFTP sites. dbSNP150 annotation was added separately by

using a VCF file downloaded from NCBI dbSNP ftp site (28). ENCODE (v4) annotations were downloaded as BED files from the UCSC Table Browser (Feb.2009 [GRCh37/hg19] assembly). The conversion from GRCh37 to GRCh38 coordinates was performed using liftOver from the UCSC Genome Browser Utilities (29).

Single locus BDR association testing on common variants

An additive logistic regression model was used for each population to evaluate the association of BDR with biallelic common variants (MAF > 1%) using PLINK 1.9 (18). Throughout this study, high drug responders were assigned as cases. Logistic regression models included age, sex and body mass index (BMI) category to account for previously reported confounders (30-37), and the first ten principal components (PCs) to correct for population substructure in admixed populations.

BMI and age- and sex-specific BMI percentiles (BMI-pct) were calculated as previously described (38) and used to assign BMI categories. For subjects aged 20 years and over, BMI categories were defined as follows: underweight (BMI < 18), normal ($18 \leq \text{BMI} < 25$), overweight ($25 \leq \text{BMI} < 30$) and obese (BMI ≥ 30). For subjects under 20 years of age, BMI categories were defined as follows: underweight (BMI-pct < 5), normal ($5 \leq \text{BMI-pct} < 85$), overweight ($85 \leq \text{BMI-pct} < 95$) and obese (BMI-pct ≥ 95).

Baseline lung function (pre-FEV₁) has a significant impact on BDR. Pre-FEV₁ was not included as a covariate in association analyses, as variation in pre-FEV₁ was indirectly captured by several of the principal components included in association models. A correlation matrix showing the relationship between pre-FEV₁, age, sex,

BMI status, mean global ancestry, tIgE, atopy and the top ten principal components is presented in **Table E20**. We included 10 PCs as covariates to control for the effects of known confounders, such as pre-FEV₁, tIgE and atopy. However, we acknowledge that PCs may not completely correct for all effects of each confounding variable and this should be taken into account when considering our study results.

Population-specific genome-wide significance thresholds for the single locus analyses were calculated based on genotypes using the autocorrelation-based *effectiveSize()* function in the R package '*coda*' as published by Sobota *et al.* (39). Population-specific genome-wide significance thresholds after adjusting for the effective number of tests (adjusted genome-wide significance) were 1.57×10^{-7} for Puerto Ricans, 2.42×10^{-7} for Mexicans, and 9.59×10^{-8} for African Americans. Suggestive significance thresholds were calculated as one divided by the effective number of tests (40). Linkage disequilibrium patterns (Genome build: hg19/1000 Genomes Nov 2014 AMR) of the flanking regions of BDR-associated SNPs were visualized using LocusZoom (41). Quantile-quantile (q-q) plots were generated assuming a uniform distribution for the expected *p*-value (**Figures E13A-C**). The genomic inflation factor (λ_{GC}) was calculated using the R package '*gap*'.

Trans-ethnic meta-analysis of common variant effects on BDR

A meta-analysis of the effects of common variants on BDR across the three populations was performed using METASOFT (42). We used the Han and Eskin random effects model optimized for detecting associations under heterogeneous genetic effects (42). The number of effective tests was estimated using *coda* as

described above, yielding adjusted genome-wide significant and suggestive thresholds of 3.53×10^{-7} and 7.06×10^{-6} , respectively. Worldwide allele frequency variation was visualized using the Geography of Genetic Variants Browser (GGV) beta v0.92 (43). The q-q plot and λ_{GC} were generated in the same way as described above (**Figure E13D**, see **supplementary methods** section, “Single locus BDR association testing on common variants”).

Calculation of variation in BDR explained by common variants

Significantly and suggestively associated variants in **Table 2** were pruned for LD using the LD prune function in PLINK 1.9 (R^2 cut-off: 0.6, window size = 50 SNPs, shift = 5 SNPs). Total variation in BDR explained was estimated by calculating McFadden’s pseudo R^2 (44) in each population separately, adjusting for age, sex, BMI category, and principal components 1-10.

Multi-variant analyses of combined effects of common and rare BDR variants

The combined effects of common and rare BDR variants were analyzed using SKAT-O (45) based on a window size of 1 kb and a step size of 500 base pairs using the GRCh37 genome. A total of 5.3 million windows were analyzed and *coda* was used to determine the number of effective tests based on autocorrelation of the association *p*-value, as described above. Adjusted genome-wide significance thresholds for Puerto Ricans (8.15×10^{-8}), Mexicans (8.60×10^{-8}), African Americans (6.94×10^{-8}) and for all three populations combined (1.53×10^{-7}) were used to identify windows of variants significantly associated with BDR. The same covariates used for common

variant association testing were used for analyses of individual populations. For analyses combined across all three populations, we avoided confounding from population substructure by including local ancestry as additional covariates, defined as the proportions of Native American and African ancestries for the window under testing. The q-q plots and λ_{GC} were generated as described above (**Figure E14**, see **Supplementary Methods** section, “Single locus BDR association testing on common variants”).

Reduced and drop-one SKAT analyses

In the reduced analysis, each SKAT-O region contained all common variants but only one rare variant from the original region was included. Results of the reduced analysis are presented in **Table E18**.

We also conducted another drop-one analysis in which each SKAT-O region contained all the rare variants but common variants from that region were excluded one at a time. Results of that drop-one analysis are presented in **Table E19**.

Variation in BDR explained by associated SKAT-O regions

Variation in BDR explained by SKAT-O regions was calculated as described above for common variant analyses using McFadden’s pseudo R^2 (44), but with one addition: variants were weighted using a weighted kernel as described in SKAT-O (45).

Single locus BDR association and trans-ethnic meta-analysis of array data

To address the discrepancy between our common variant analysis results with previously published BDR GWAS results (9), we used 1,414 of the 1,441 individuals who had both Axiom® Genome-Wide LAT 1 array (see **supplementary methods** section, “WGS data processing and quality control”) and WGS data available to rerun single locus BDR association testing and trans-ethnic meta-analysis. Array data were imputed to the Haplotype Reference Consortium (46) (HRC release 1) panel using the Michigan Imputation Server (47). We used the top 1,000 BDR-associated SNPs to examine the relationship between array-based and WGS-based association *p*-values, genotype discordance, and imputation R^2 . Correlation between array-based and WGS-based association *p*-values was determined by Spearman correlation. We also performed single locus BDR association testing and trans-ethnic meta-analysis by applying linear regression on 1,122 Puerto Ricans, 662 Mexicans and 1,105 African Americans using BDR (ΔFEV_1) as a continuous trait. HRC imputed array-based data and the same covariates as described above were used for the analysis.

Replication of top BDR-associated common variants

Replication cohorts included the Genetics of Asthma in Latino Americans Study (GALA I) (48, 49), the Study of African Americans, Asthma, Genes & Environments (SAGE I) (50), a case-control study of childhood asthma in Puerto Ricans (HPR) (51), the Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity (SAPPHIRE) (52) and a cohort from the Children’s Hospital of Philadelphia (CHOP) (53). Descriptive statistics for the replication cohorts are shown in **Table E10**.

Logistic regression was performed using the same population-specific extreme BDR cut-offs applied to the discovery analyses in the current study (see **supplementary methods** section, “Bronchodilator response measurements”). All 27 significantly and suggestively associated SNPs from the discovery analysis (**Table 2**) were assessed in each replication study population. An additive genetic model was assumed for each SNP tested. Association models were adjusted as in the discovery analysis for sex, age, BMI category, and the first ten principal components. Meta-analysis across all replication studies was performed using METASOFT as described above (see **supplementary methods** section, “Trans-ethnic meta-analysis of common variant effects on BDR”).

The GALA I and SAGE I replication cohorts included 108 Puerto Ricans, 202 Mexicans and 141 African Americans with BDR measurements and complete data for all regression covariates. Genotype data were imputed to the HRC panel using the Michigan Imputation Server (47). Replication in the HPR cohort involved 414 Puerto Rican subjects. Spirometry data were collected as previously described (54). Genome-wide genotyping was performed using the Illumina HumanOmni2.5 BeadChip platform (Illumina Inc., San Diego, CA) and processed as previously described (55). Genotype data were phased with SHAPE-IT (56) and imputation was performed with IMPUTE2 (57) using all populations from 1000 Genomes Project Phase 3 as reference (58). The SAPPHIRE replication cohort consisted of 1,022 African Americans with asthma. Genome-wide genotyping was performed using the Axiom® Genome-Wide AFR 1 array (Affymetrix Inc., Santa Clara, CA) as previously described (59). Genotype data were imputed to the cosmopolitan 1000 Genomes

Phase 1 version haplotypes using the Michigan Imputation Server (47). The CHOP replication cohort included 280 African Americans. Genotyping was performed as described (53), and genotype data were imputed to the HRC panel using the Sanger Imputation server (46).

Identification of nearest genes for BDR-associated loci

The four nearest transcripts to BDR-associated loci were identified using the “closest” command in BEDTools (60), with the parameters “-d -k 4”. RefSeq gene annotations (Feb.2009 [GRCh37/hg19] assembly) were downloaded in refFlat format from the UCSC Table Browser (61). Genes with multiple transcripts were reported only once. When reporting the nearest gene, the “closest” command in BEDTools with the parameters “-D a” was applied.

Primary bronchial smooth muscle cell culture

Cryopreserved primary human bronchial smooth muscle from two donors (Lonza catalog number CC-2576, lot number 0000212076 and ATCC catalog number PCS-130-011, lot number 62326179) was thawed and expanded in Lonza Smooth Muscle Growth Media (SmGM; catalog number CC-3182) in T75 flasks (E&K Scientific Products, catalog number 658175).

H3K27ac ChIP-seq assay

Upon reaching 80% confluency, BSMCs were serum-starved by replacing SmGM with smooth muscle basal media (SmBM) for 24 hours. BSMCs were then grown in

SmBM containing 5% FBS for 4 hours, fixed in 1% formaldehyde for 10 min and quenched with 0.125 M glycine for 5 minutes. Cells were removed from the T75 flasks by scraping in cold PBS containing sodium butyrate (20 mM, Diagenode, catalog number C12020010). Chromatin sheering was carried out using a Covaris S2 sonicator. Sheared chromatin was used for immunoprecipitation with antibodies against active chromatin marks (H3K27ac; Abcam, ab4729) using the Diagenode LowCell# ChIP kit (CAT#C01010072), following the manufacturer's protocol. Libraries were prepared using the Rubicon DNA-Seq kit (CAT#R400406) following the manufacturer's protocol and sequenced on an Illumina HiSeq 4000 using single-end 50-bp reads to a sequencing depth of at least 25 million reads (submitted under BioProject PRJNA369271). Uniquely mapping raw reads were aligned using Bowtie (62) under default settings. Peak calling was performed on each individual (BAM file) separately and then combined using MACS2 (63, 64) with a raw *p*-value of 0.001, resulting in three sets of peaks (donor 1, 2 and combined). The three sets of peaks were then used in the ENCODE IDR pipeline (65) to generate reproducible peaks using an IDR threshold of 0.1. MACS2 uses a sliding window approach to call peaks by finding regions significantly enriched in the ChIP DNA over the input DNA. An example of BSMC peaks identified by MACS2 and the ENCODE IDR pipeline is shown in the track "BSMC peaks" in Figure E9A. All identified H3K27Ac ChIP-seq peaks were overlapped with the 27 BDR-associated common variants listed in **Table 2**, as well as variants in high linkage disequilibrium with them ($R^2 \geq 0.8$ within 1 Mb in any of our study populations).

Diverse Convergent Evidence approach for variant prioritization

The Diverse Convergent Evidence (DiCE) approach is a logical heuristic framework for integrating multiple types of observational, bioinformatics, and laboratory evidence to prioritize variants discovered from high throughput genetic studies for further evaluation in functional experiments (66). Results from the trans-ethnic meta-analysis and from the replication analyses were considered observational data. Laboratory evidence was provided by the identified peaks in our ChIP-seq analyses performed in BSMCs. Informatic evidence was compiled using Ensembl, PubMed, the NHGRI-EBI GWAS Catalog, and ENCODE to identify previously reported associations with BDR or asthma-related phenotypes, and predicted biological functions associated with assessed loci. After compiling the observational, informatic and laboratory evidence for each suggestively or significantly associated variant, DiCE constructs an evidence matrix to estimate the strength of the information supporting each association. DiCE scores > 6 were considered strong evidence that a given locus was involved in the pathophysiology of BDR. After meeting this criterion, variants with the highest DiCE score were prioritized for downstream functional analyses.

Luciferase assays

NFKB1 candidate enhancer sequences were amplified from human genomic DNA (Roche) using oligonucleotides designed in Primer3 with 18 and 20 bp overhangs for forward and reverse primers, respectively (5'-GGCCTAACTGGCCGGTAC-3' and 5'-CGCCGAGGCCAGATCTTGAT-3'), complementary to sequences flanking the KpnI

and EcoRV sites in the pGL4.23 Gate A vector (Promega, Madison, WI, USA) using Phusion High-Fidelity PCR kit (NEB, catalog number M0531S). The DNA fragments used for the enhancer assays were amplified from a commercially available product containing DNA from a mixture of individuals (Roche). During cloning, one fragment was chosen at random and sequencing was carried out to confirm the correct region of interest was present in the vector. PCR primers were designed around the edges of these ChIP-seq peaks and the most complete successfully amplified fragment was used for luciferase assays. PCR products were then cloned into the pGL4.23 vector using the Gibson Assembly method (NEB, catalog number E2611S). The pGL4.23 vector contains a minimal promoter and a luciferase reporter gene. The pGL4.23 vector with the viral SV40 promoter was used as a positive control, and the pGL4.23 empty vector as a negative control. Smooth muscle cells were plated at 50-70% confluency in 24-well cell culture plates (Falcon, catalog number 353047) and grown to 80% confluency in SmGM. Transfections were carried out by combining polyethylenimine (PEI) with DNA vectors at a 1:1 ratio by weight in opti-MEM (Life Technologies, catalog number 31985070). The transfection mixture consisted of 225 ng of enhancer assay vectors and 25 ng of pGL4.74 (Renilla transfection efficiency control) with 250 ng of PEI in 50 μ L of opti-MEM. After 15 minutes, 500 μ L of SmBM was added to the transfection mixture and the combination added to cell culture. Cells were incubated for 4 hours in SmBM plus the transfection mixture, then media was replaced with SmGM for 24 hours. Cells were then washed with PBS and enhancer assay cells were lysed with 100 μ L of Passive Lysis Buffer (Promega, Madison, WI, USA). Reporter activity was measured using the Dual-Luciferase

Reporter Assay System (Promega, Madison, WI, USA) and measured on the Glomax 96 well plate luminometer (Promega, Madison, WI, USA). The luciferase assay was carried out in two separate experiments, with three independent replicates per experiment (three wells of cells were transfected per construct per experiment). Each well was then split into two technical replicates for luciferase activity measurements with a luminometer.

RNA extraction, library preparation and sequencing

Among African American subjects with WGS data in our study, 39 samples were selected for RNA-seq based on BDR status and the number of copies of low-BDR associated alleles at rs28450894. The number of samples in each category is shown in **Table E21**. Peripheral blood samples were collected into PAXgene Blood RNA tubes (PreAnalytiX, Hombrechtikon, Switzerland). Total RNA was extracted from PAXgene Blood RNA tubes using the MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit (CAT#4451894, Thermo Fisher Scientific, Waltham, MA, USA) according to manufacturer's protocols. RNA integrity and yield were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Globin depletion was performed using GLOBINclear™ Kit (CAT#AM1980, Thermo Fisher Scientific, Waltham, MA, USA). Library preparation and ribosomal depletion were performed using KAPA Stranded RNA-seq Kit with RiboErase (CAT#KK8483, Kapa Biosystems, Wilmington, MA, USA) according to the manufacturer's protocols. Each sample was uniquely barcoded with NEXTflex™ DNA Barcodes (CAT#514104, Bioo Scientific®, Austin, TX, USA). Barcoded libraries were pooled and sequenced on 4

lanes on a HiSeq 4000 sequencing system (Illumina®, San Diego, CA, USA) with a paired-end read length of 100 bp at the University of California, San Francisco's Center for Advanced Technology.

RNA-seq data processing and analysis

Raw sequencing reads were aligned to the human reference genome (hg19) using STAR (v2.4.2a) (67). Gene read counts were obtained from uniquely mapped reads based on Ensembl annotation (v75) (68). DESeq2 (69) was used to analyze read counts for differential gene expression changes between genotypes, including an interaction term with genotype and sex (genotype * sex). We used a linear model to account for sex, age and library prep batch, and a custom model matrix to correct for GC content difference between genes. After normalization for sequencing depth and GC percentage, genes with fewer than an average of five normalized read counts per sample and fewer than 20 samples with at least one read count were removed. This filtering process kept 19,592 Ensembl genes for analysis. Fold change, raw and FDR-adjusted *p*-values for the genotype term were reported. Genes were then further filtered to analyze the locus surrounding rs28450894 for differential gene expression by including all genes with a transcriptional start site within 1Mb of rs28450894. *p*-values were then corrected using the false discovery rate method to account for the 13 genes in this locus. FDR-adjusted *p*-values ≤ 0.1 were considered significant.

SUPPLEMENTARY REFERENCES

1. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310-315.
2. Gulko B, Hubisz MJ, Gronau I, Siepel A. A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat Genet* 2015;47:276-283.
3. Hernandez RD. A flexible forward simulator for populations subject to selection and demography. *Bioinformatics* 2008;24:2786-2787.
4. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 2009;5:e1000695.
5. Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, White TJ, Nielsen R, Clark AG, Bustamante CD. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 2008;4:e1000083.
6. Guey LT, Kravic J, Melander O, Burt NP, Laramie JM, Lyssenko V, Jonsson A, Lindholm E, Tuomi T, Isomaa B, Nilsson P, Almgren P, Kathiresan S, Groop L, Seymour AB, Altshuler D, Voight BF. Power in the phenotypic extremes: a simulation

study of power in discovery and replication of rare variants. *Genet Epidemiol* 2011;35:236-246.

7. Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet Epidemiol* 2011;35:790-799.

8. Lamina C. Digging into the extremes: a useful approach for the analysis of rare variants with continuous traits? *BMC Proc* 2011;5 Suppl 9:S105-6561-5-S9-S105.

9. Drake KA, Torgerson DG, Gignoux CR, Galanter JM, Roth LA, Huntsman S, Eng C, Oh SS, Yee SW, Lin L, Bustamante CD, Moreno-Estrada A, Sandoval K, Davis A, Borrell LN, Farber HJ, Kumar R, Avila PC, Brigino-Buenaventura E, Chapela R, Ford JG, Lenoir MA, Lurmann F, Meade K, Serebrisky D, Thyne S, Rodriguez-Cintron W, Sen S, Rodriguez-Santana JR, Hernandez RD, Giacomini KM, Burchard EG. A genome-wide association study of bronchodilator response in Latinos implicates rare variants. *J Allergy Clin Immunol* 2014;133:370-378.

10. McGarry ME, Castellanos E, Thakur N, Oh SS, Eng C, Davis A, Meade K, LeNoir MA, Avila PC, Farber HJ, Serebrisky D, Brigino-Buenaventura E, Rodriguez-Cintron W, Kumar R, Bibbins-Domingo K, Thyne SM, Sen S, Rodriguez-Santana JR, Borrell LN, Burchard EG. Obesity and bronchodilator response in black and Hispanic children and adolescents with asthma. *Chest* 2015;147:1591-1598.

11. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van der Grinten CP, Gustafsson P, Hankinson J, Jensen R, Johnson DC, MacIntyre N,

McKay R, Miller MR, Navajas D, Pedersen OF, Wanger J. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948-968.

12. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, Boehnke M, Kang HM. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 2012;91:839-848.

13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754-1760.

14. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491-498.

15. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156-2158.

16. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 2014;30:2843-2851.

17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078-2079.

18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7-015-0047-8. eCollection 2015.
19. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating kinship in admixed populations. *Am J Hum Genet* 2012;91:122-138.
20. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19:1655-1664.
21. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010;26:2867-2873.
22. Kessler MD, Yerges-Armstrong L, Taub MA, Shetty AC, Maloney K, Jeng LJ, Ruczinski I, Levin AM, Williams LK, Beaty TH, Mathias RA, Barnes KC, Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA), O'Connor TD. Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry. *Nat Commun* 2016;7:12521.
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909.

24. Affymetrix. Axiom genotype data set. 2014. [online] Available at: http://www.affymetrix.com/support/technical/sample_data/axiom_db/axiomdb_data.affx [Accessed 4/1 2014].
25. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013;10:5-6.
26. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* 2013;93:278-288.
27. Liu X, White S, Peng B, Johnson AD, Brody JA, Li AH, Huang Z, Carroll A, Wei P, Gibbs R, Klein RJ, Boerwinkle E. WGS: an annotation pipeline for human genome sequencing studies. *J Med Genet* 2016;53:111-112.
28. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308-311.
29. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform* 2013;14:144-161.
30. Yao TC, Ou LS, Yeh KW, Lee WI, Chen LC, Huang JL, PATCH Study Group. Associations of age, gender, and BMI with prevalence of allergic diseases in children: PATCH study. *J Asthma* 2011;48:503-510.
31. Nicolai T, Pereszlenyiova-Bliznakova L, Illi S, Reinhardt D, von Mutius E. Longitudinal follow-up of the changing gender ratio in asthma from childhood to

adulthood: role of delayed manifestation in girls. *Pediatr Allergy Immunol* 2003;14:280-283.

32. Joseph M, Elliott M, Zelicoff A, Qian Z, Trevathan E, Chang JJ. Racial disparity in the association between body mass index and self-reported asthma in children: a population-based study. *J Asthma* 2016;53:492-497.

33. Dixon AE, Shade DM, Cohen RI, Skloot GS, Holbrook JT, Smith LJ, Lima JJ, Allayee H, Irvin CG, Wise RA, American Lung Association-Asthma Clinical Research Centers. Effect of obesity on clinical presentation and response to treatment in asthma. *J Asthma* 2006;43:553-558.

34. Ullah MI, Newman GB, Saunders KB. Influence of age on response to ipratropium and salbutamol in asthma. *Thorax* 1981;36:523-529.

35. Mohamed MH, Lima JJ, Eberle LV, Self TH, Johnson JA. Effects of gender and race on albuterol pharmacokinetics. *Pharmacotherapy* 1999;19:157-161.

36. Carroll CL, Bhandari A, Zucker AR, Schramm CM. Childhood obesity increases duration of therapy during severe asthma exacerbations. *Pediatr Crit Care Med* 2006;7:527-531.

37. Forno E, Lescher R, Strunk R, Weiss S, Fuhlbrigge A, Celedon JC, Childhood Asthma Management Program Research Group. Decreased response to inhaled steroids in overweight and obese asthmatic children. *J Allergy Clin Immunol* 2011;127:741-749.

38. White MJ, Risse-Adams O, Goddard P, Contreras MG, Adams J, Hu D, Eng C, Oh SS, Davis A, Meade K, Brigino-Buenaventura E, LeNoir MA, Bibbins-Domingo K, Pino-Yanes M, Burchard EG. Novel genetic risk factors for asthma in African American children: Precision Medicine and the SAGE II Study. *Immunogenetics* 2016;68:391-400.
39. Sobota RS, Shriner D, Kodaman N, Goodloe R, Zheng W, Gao YT, Edwards TL, Amos CI, Williams SM. Addressing population-specific multiple testing burdens in genetic association studies. *Ann Hum Genet* 2015;79:136-147.
40. Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 2008;9:516-2164-9-516.
41. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336-2337.
42. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* 2011;88:586-598.
43. Marcus JH, Novembre J. Visualizing the geography of genetic variants. *Bioinformatics* 2017;33:594-595.

44. McFadden D. Quantitative methods for analysing travel behavior of individuals: Some recent developments. In: Hensher DA and Stopher PR, editors. Behavioural travel modelling. London: Croom Helm; 1979. p. 279-318.
45. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, NHLBI GO Exome Sequencing Project-ESP Lung Project Team, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 2012;91:224-237.
46. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson

CA, Myers RM, Boehnke M, McCarthy MI, Durbin R, Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48:1279-1283.

47. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. Next-generation genotype imputation service and methods. *Nat Genet* 2016;48:1284-1287.

48. Burchard EG, Avila PC, Nazario S, Casal J, Torres A, Rodriguez-Santana JR, Toscano M, Sylvia JS, Alioto M, Salazar M, Gomez I, Fagan JK, Salas J, Lilly C, Matallana H, Ziv E, Castro R, Selman M, Chapela R, Sheppard D, Weiss ST, Ford JG, Boushey HA, Rodriguez-Cintron W, Drazen JM, Silverman EK, Genetics of Asthma in Latino Americans (GALA) Study. Lower bronchodilator responsiveness in Puerto Rican than in Mexican subjects with asthma. *Am J Respir Crit Care Med* 2004;169:386-392.

49. Torgerson DG, Gignoux CR, Galanter JM, Drake KA, Roth LA, Eng C, Huntsman S, Torres R, Avila PC, Chapela R, Ford JG, Rodriguez-Santana JR, Rodriguez-Cintron W, Hernandez RD, Burchard EG. Case-control admixture mapping in Latino populations enriches for known asthma-associated genes. *J Allergy Clin Immunol* 2012;130:76-82.e12.

50. Nishimura KK, Galanter JM, Roth LA, Oh SS, Thakur N, Nguyen EA, Thyne S, Farber HJ, Serebrisky D, Kumar R, Brigino-Buenaventura E, Davis A, LeNoir MA, Meade K, Rodriguez-Cintron W, Avila PC, Borrell LN, Bibbins-Domingo K, Rodriguez-Santana JR, Sen S, Lurmann F, Balmes JR, Burchard EG. Early-life air pollution and asthma risk in minority children. The GALA II and SAGE II studies. *Am J Respir Crit Care Med* 2013;188:309-318.
51. Han YY, Forno E, Brehm JM, Acosta-Perez E, Alvarez M, Colon-Semidey A, Rivera-Soto W, Campos H, Litonjua AA, Alcorn JF, Canino G, Celedon JC. Diet, interleukin-17, and childhood asthma in Puerto Ricans. *Ann Allergy Asthma Immunol* 2015;115:288-293.e1.
52. Levin AM, Wang Y, Wells KE, Padhukasahasram B, Yang JJ, Burchard EG, Williams LK. Nocturnal asthma and the importance of race/ethnicity and genetic ancestry. *Am J Respir Crit Care Med* 2014;190:266-273.
53. Ong BA, Li J, McDonough JM, Wei Z, Kim C, Chiavacci R, Mentch F, Caboot JB, Spergel J, Allen JL, Sleiman PM, Hakonarson H. Gene network analysis in a pediatric cohort identifies novel lung function genes. *PLoS One* 2013;8:e72899.
54. Brehm JM, Ramratnam SK, Tse SM, Croteau-Chonka DC, Pino-Yanes M, Rosas-Salazar C, Litonjua AA, Raby BA, Boutaoui N, Han YY, Chen W, Forno E, Marsland AL, Nugent NR, Eng C, Colon-Semidey A, Alvarez M, Acosta-Perez E, Spear ML, Martinez FD, Avila L, Weiss ST, Soto-Quiros M, Ober C, Nicolae DL, Barnes KC, Lemanske RF, Jr, Strunk RC, Liu A, London SJ, Gilliland F, Sleiman P, March M, Hakonarson H, Duan

QL, Kolls JK, Fritz GK, Hu D, Fani N, Stevens JS, Almli LM, Burchard EG, Shin J, McQuaid EL, Ressler K, Canino G, Celedon JC. Stress and Bronchodilator Response in Children with Asthma. *Am J Respir Crit Care Med* 2015;192:47-56.

55. Chen W, Brehm JM, Lin J, Wang T, Forno E, Acosta-Perez E, Boutaoui N, Canino G, Celedon JC. Expression quantitative trait loci (eQTL) mapping in Puerto Rican children. *PLoS One* 2015;10:e0122464.

56. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2011;9:179-181.

57. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.

58. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* 2015;526:68-74.

59. Padhukasahasram B, Yang JJ, Levin AM, Yang M, Burchard EG, Kumar R, Kwok PY, Seibold MA, Lanfear DE, Williams LK. Gene-based association identifies SPATA13-AS1 as a pharmacogenomic predictor of inhaled short-acting beta-agonist response in multiple population groups. *Pharmacogenomics J* 2014;14:365-371.

60. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841-842.

61. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 2004;32:D493-6.
62. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25-2009-10-3-r25. Epub 2009 Mar 4.
63. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137-2008-9-9-r137. Epub 2008 Sep 17.
64. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 2012;7:1728-1740.
65. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shores N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* 2012;22:1813-1831.
66. Ciesielski TH, Pendergrass SA, White MJ, Kodaman N, Sobota RS, Huang M, Bartlett J, Li J, Pan Q, Gui J, Selleck SB, Amos CI, Ritchie MD, Moore JH, Williams SM. Diverse convergent evidence in the genetic analysis of complex disease:

coordinating omic, informatic, and experimental evidence to better identify and validate risk factors. *BioData Min* 2014;7:10-0381-7-10. eCollection 2014.

67. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15-21.

68. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt S, Johnson N, Juettemann T, Kahari AK, Keenan S, Kulesha E, Martin FJ, Maurel T, McLaren WM, Murphy DN, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat HS, Ruffier M, Sheppard D, Taylor K, Thormann A, Trevanion SJ, Vullo A, Wilder SP, Wilson M, Zadissa A, Aken BL, Birney E, Cunningham F, Harrow J, Herrero J, Hubbard TJ, Kinsella R, Muffato M, Parker A, Spudich G, Yates A, Zerbino DR, Searle SM. Ensembl 2014. *Nucleic Acids Res* 2014;42:D749-55.

69. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.

SUPPLEMENTARY FIGURES

Figure E1. Distribution of bronchodilator drug response (BDR) in Puerto Ricans, Mexicans (GALA II) and African Americans (SAGE II). The 1,441 subjects selected from the extreme of the BDR distribution for this study are highlighted.

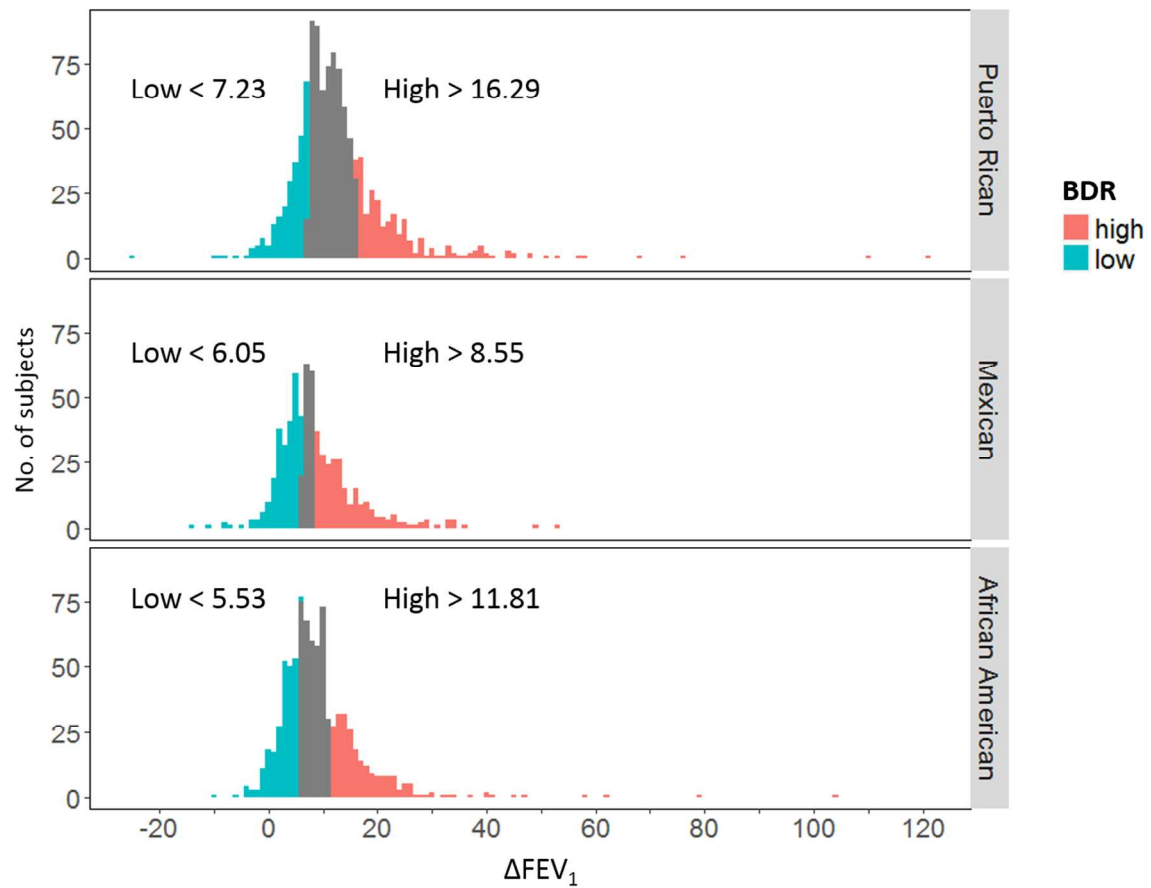


Figure E2. Global ancestry composition for **(A)** Puerto Ricans, **(B)** Mexicans and **(C)** African Americans. Each individual is represented by a vertical line and the ancestry composition is colored based on the proportion of African (red), European (blue) and Native American (green) ancestries.

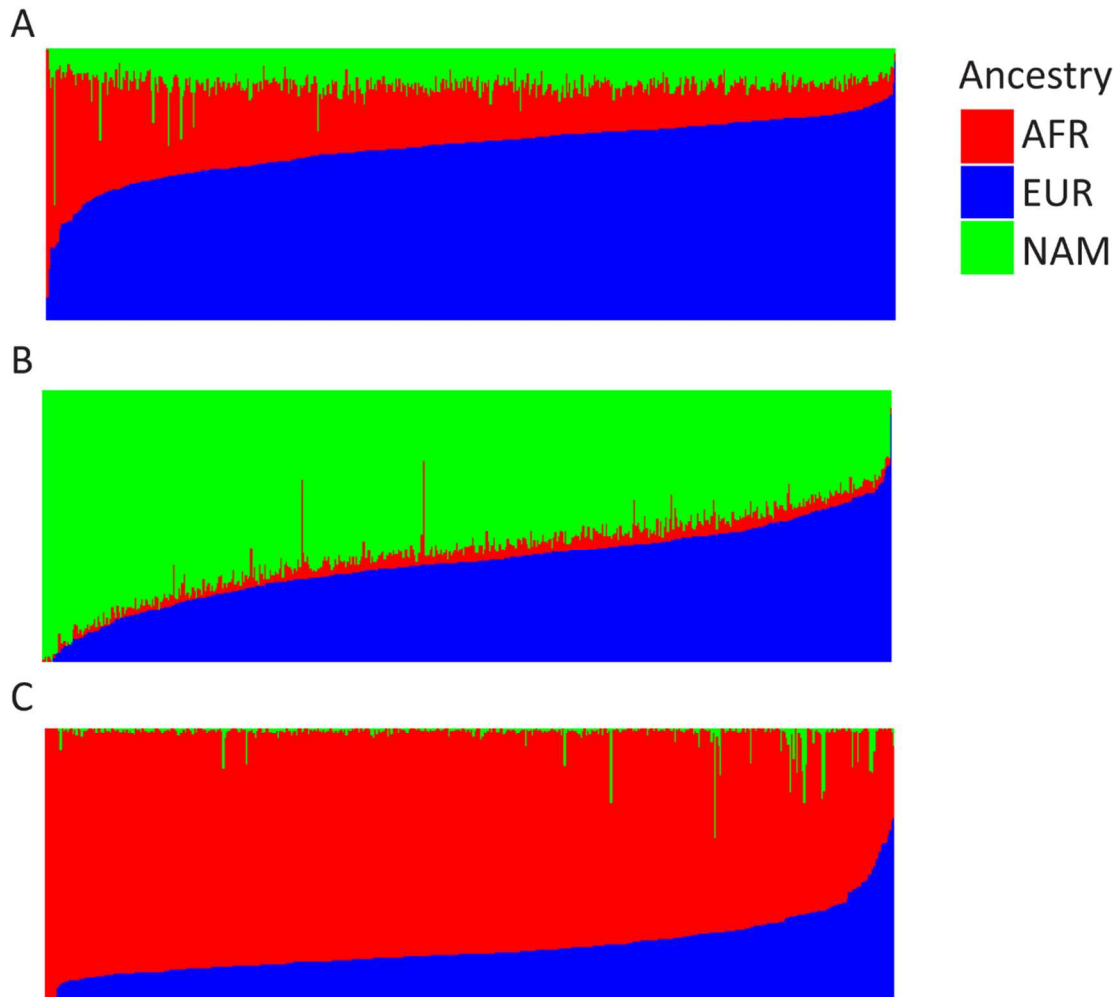


Figure E3. Plot of the first two principal components of variation based on WGS genotypes of 1,441 individuals.

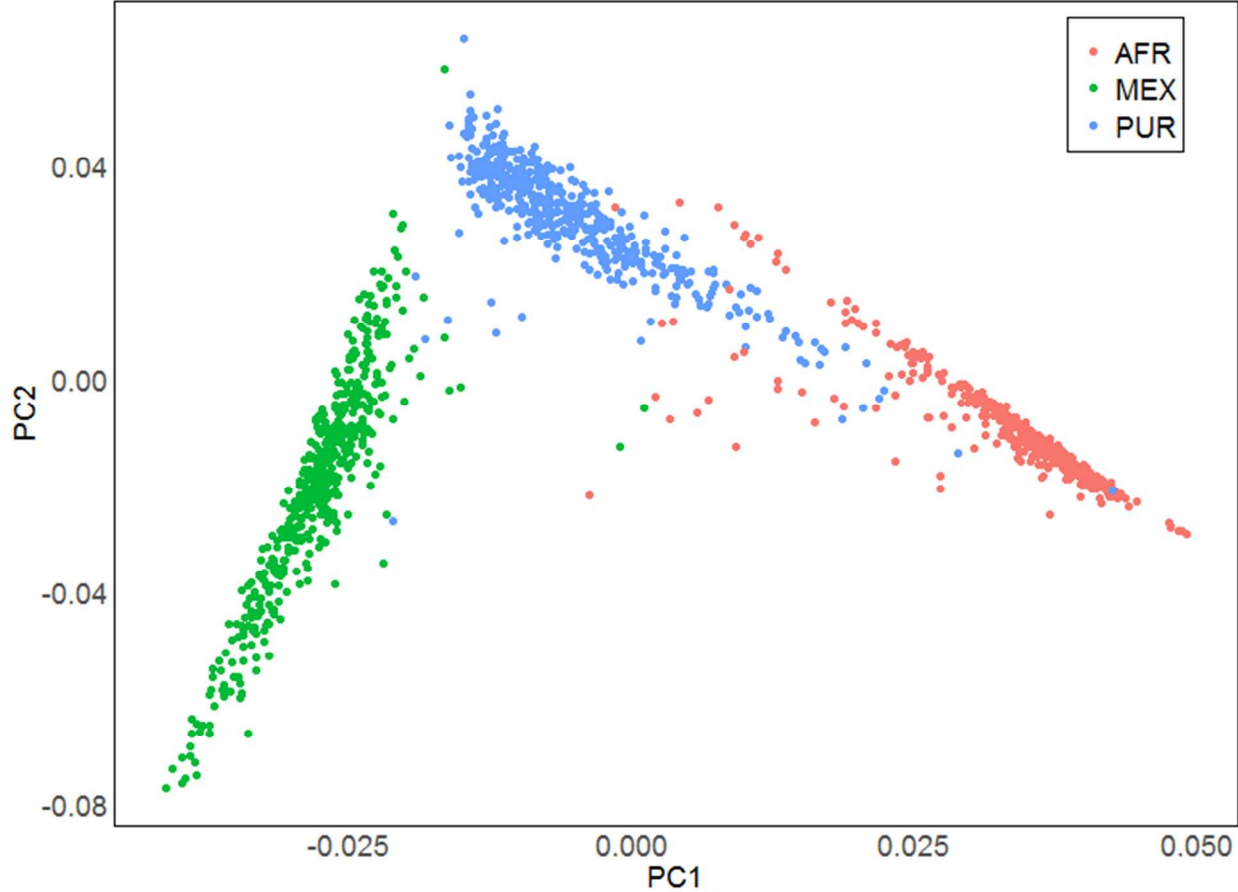


Figure E4. (A) Number of variants per sample. The bin size is 0.025M variants. (B) Allele frequency of biallelic SNPs (relative to GRCh37). (C) Allele frequency of population-specific biallelic SNPs. (D) Novel biallelic SNPs based on dbSNP build150.

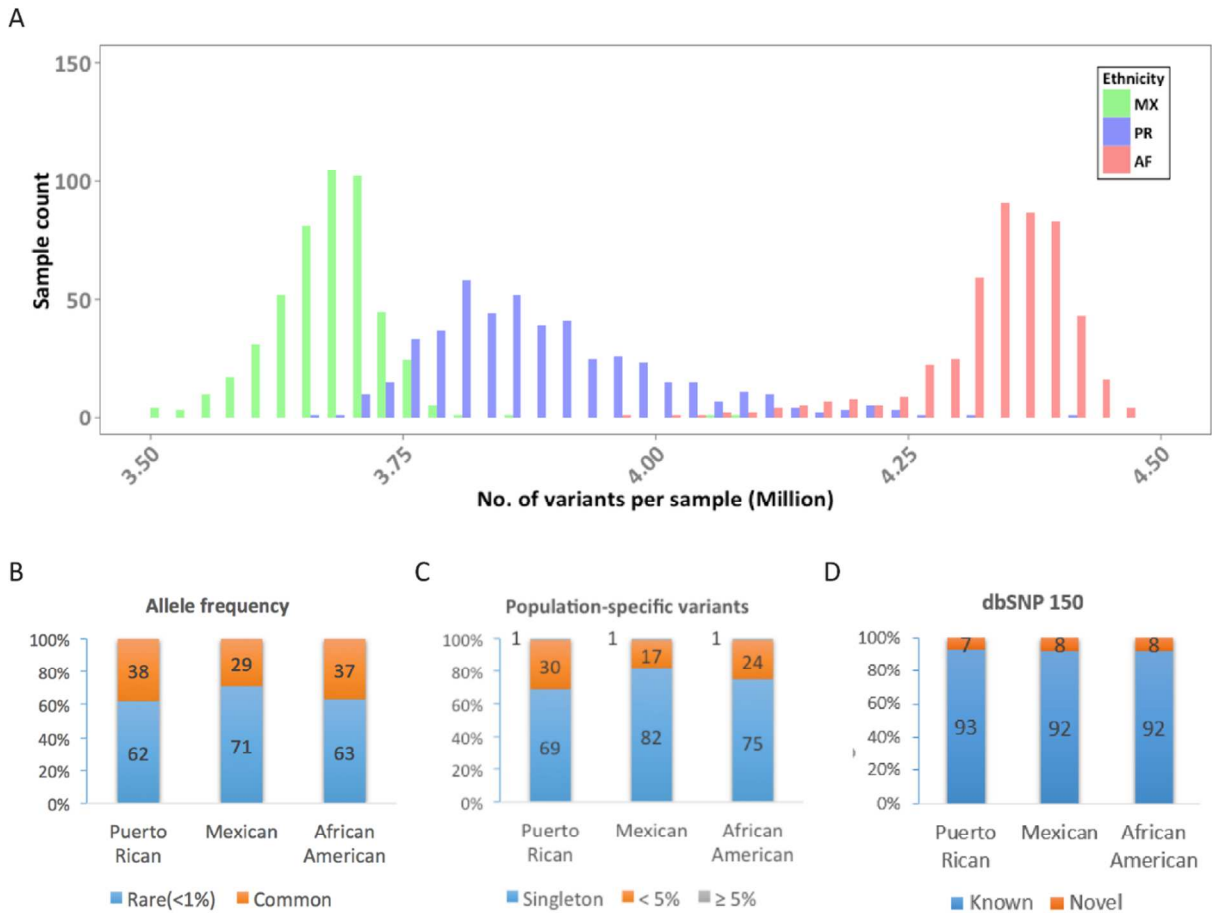
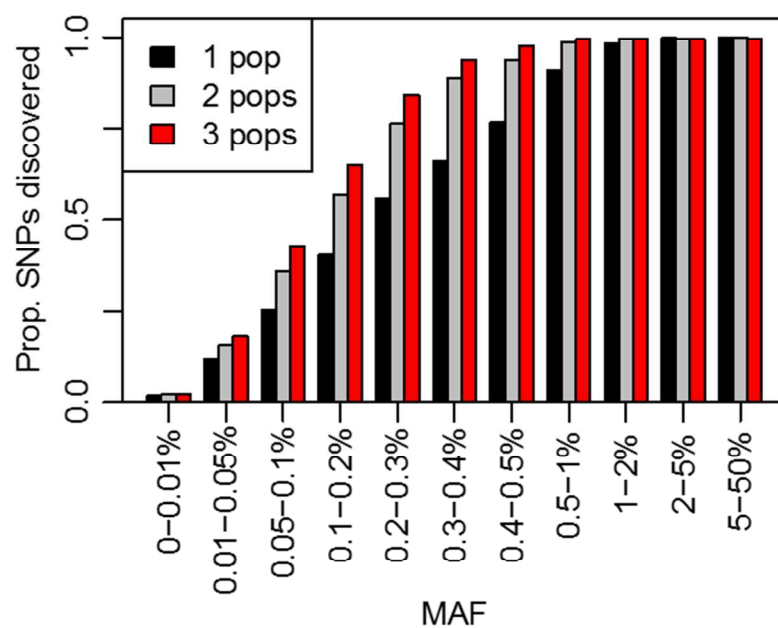


Figure E5. (A) The fraction of all variants we expect to discover by sequencing $n=1,500$ samples from a single population (black), $n=750$ samples from 2 populations (grey) or $n=500$ samples from three populations (red). Sampling from three populations maximizes our ability to capture rare variants. **(B)** Power estimation for extreme phenotype multi-allelic association testing using SKAT ($n=1,500$). A total of 1,000 3 kb regions were sampled with effect sizes of rare variants following a β distribution with a maximum effect size of 6, consistent with that observed in our previous GWAS.

A



B

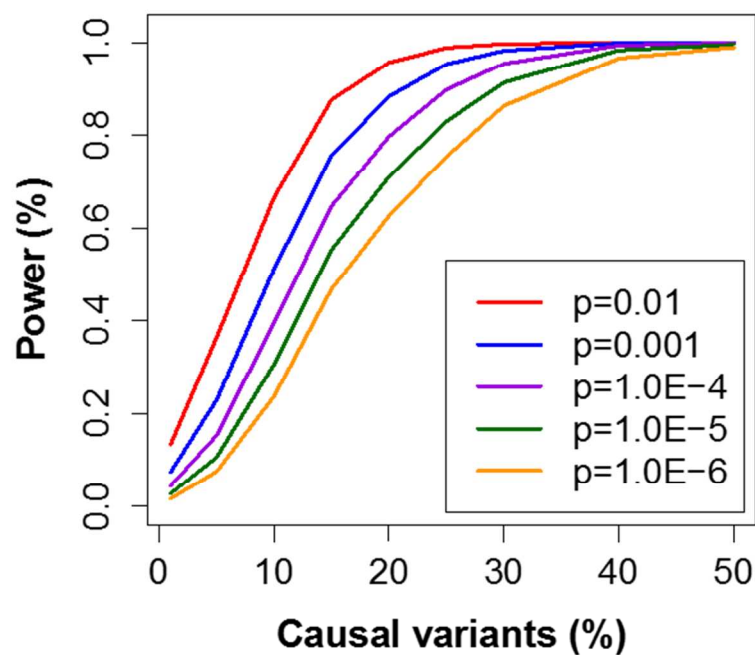


Figure E6. Manhattan plot of single locus BDR association testing for **(A)** Puerto Ricans, **(B)** Mexicans and **(C)** African Americans. The horizontal lines are colored blue for suggestive significance thresholds and red for Bonferroni-adjusted genome-wide significance thresholds. Since no associations were close to Bonferroni-significance in Mexicans and African Americans, the red line is only marked in Puerto Ricans.

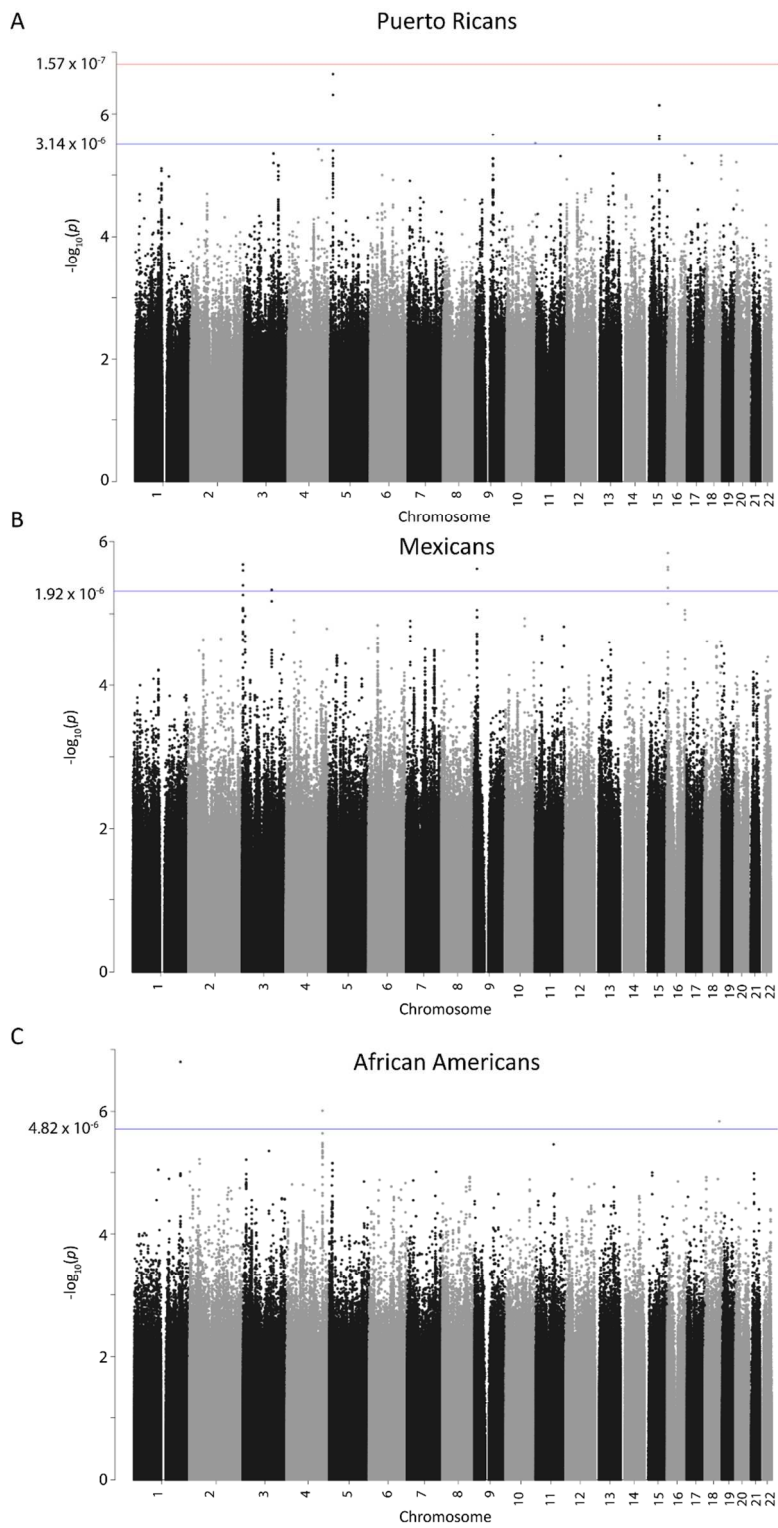


Figure E7. Allele frequency of rs2845894 for selected populations based on 1000 Genomes Project. European populations include CEU (Utah residents [CEPH] with northern and western ancestry), FIN (Finnish in Finland), GBR (British in England and Scotland), TSI (Toscani in Italia) and IBS (Iberian population in Spain). American populations include MXL (Mexican ancestry from Los Angeles USA), PUR (Puerto Ricans from Puerto Rico), CLM (Colombians from Medellin, Colombia) and PEL (Peruvians from Lima, Peru). African populations include ASW (Americans of African ancestry in SW USA), ACB (African Caribbeans in Barbados), GWD (Gambian in western divisions in the Gambia), MSL (Mende in Sierra Leone), YRI (Yoruba in Ibadan, Nigeria), ESN (Esan in Nigeria) and LWK (Luhya in Webuye, Kenya).

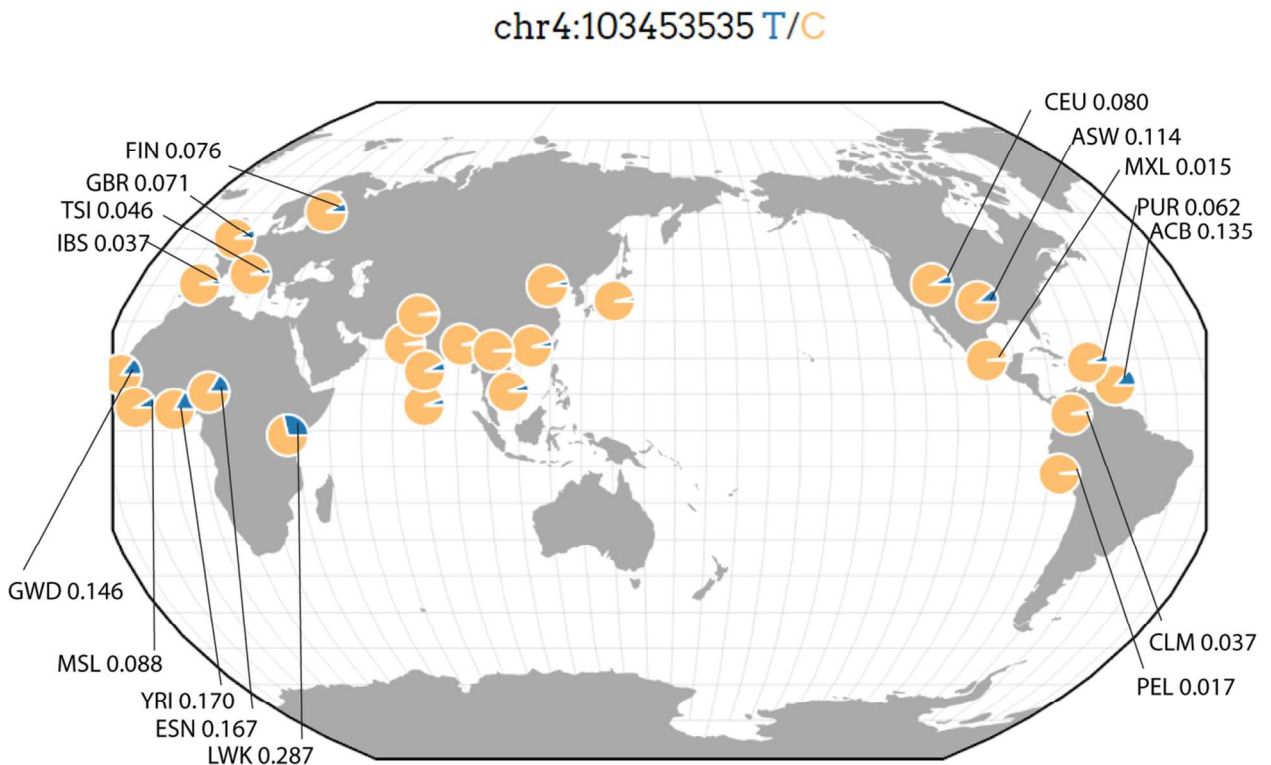


Figure E8. Diverse Convergent Evidence (DiCE) prioritization of 27 common variants (**Table 2**). Y-axis indicates the points allotted per SNP for each form of evidence according to **Table E12**: statistical evidence (grey), informatic evidence (blue), and experimental evidence (orange). Evidence-specific scoring rubric for each SNP is provided in **Table E12**. SNPs with a DiCE score ≥ 4 are labeled with the nearest gene.

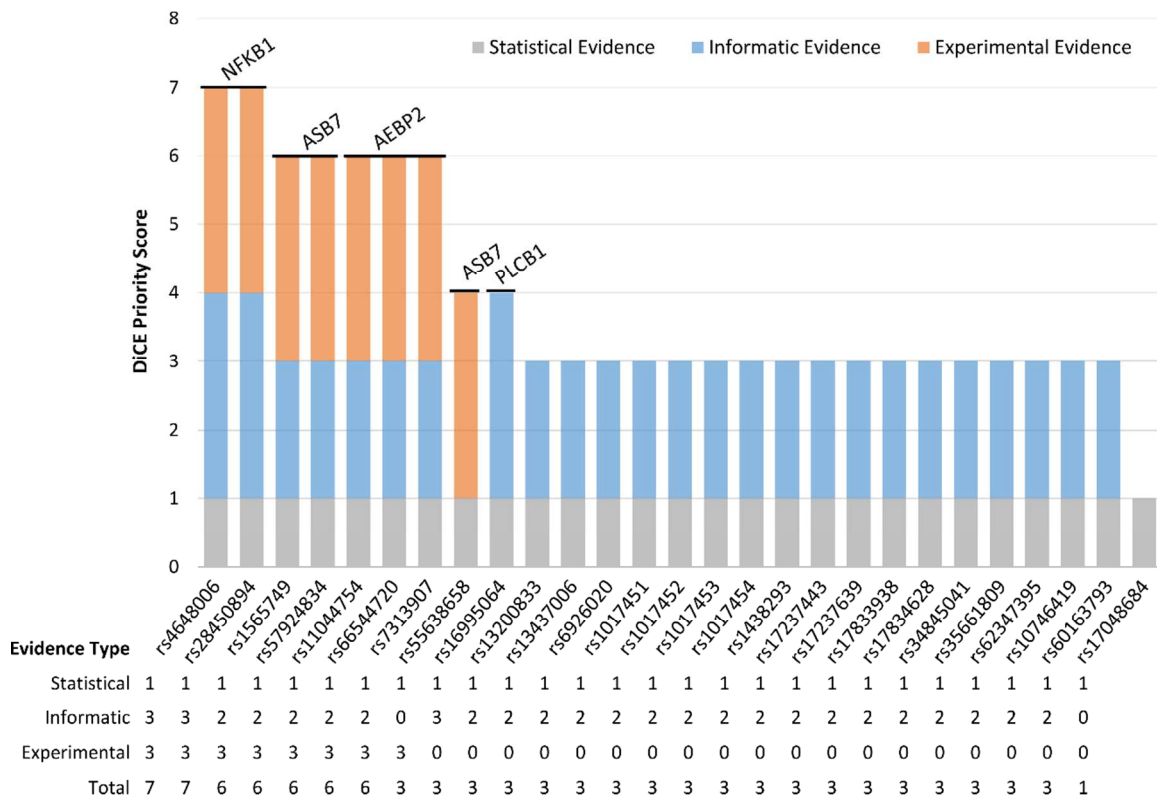
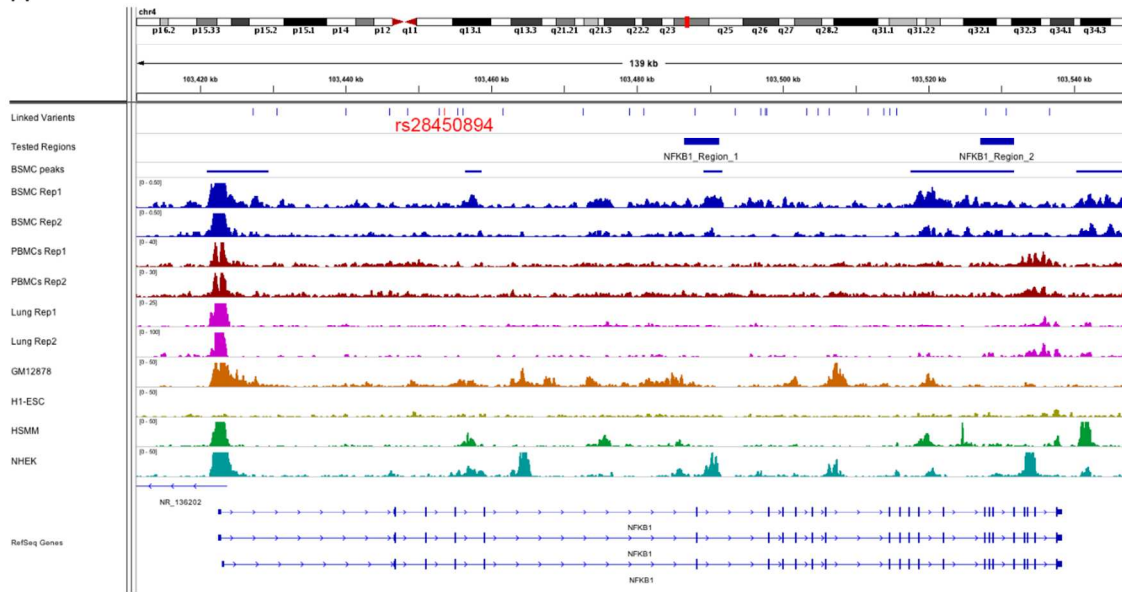


Figure E9. Regions overlapped with SNPs in LD with BDR-associated SNPs show enhancer activity in BSMCs. **(A)** H3K27ac ChIP-seq peaks in BSMCs overlap with SNPs in high LD ($R^2 \geq 0.8$ within 1 Mb) with rs28450894 (indicated in red). The GRCh37 coordinates are chr4:103486504-103491377 for Region 1 and chr4:103527184-103531814 for Region 2. The “BSMC peaks” track is derived from MACS2 broad peak calling followed by ENCODE’s IDR pipeline. The “BSMC rep 1 and 2” tracks are the normalized ChIP-seq read coverage graphs for biological replicates 1 and 2. In addition to BSMCs, H3K27ac data are shown for Roadmap Epigenomic peripheral blood mononuclear cells (PBMCs) and lung tissue, as well as ENCODE project data for GM12878, H1 human embryonic stem cell (H1-ESC) line, human skeletal muscle cells and myoblasts (HSMM) and normal human epidermal keratinocytes (NHEK). **(B)** Luciferase assay results for regions tested for enhancer activity in BSMCs. *NFKB1* Region 2 significantly increased the expression of luciferase over empty vector control ($\log_2$ fold change = 2.24, $p < 0.01$). The red line marks a fold change of one compared to the empty vector.

A



B

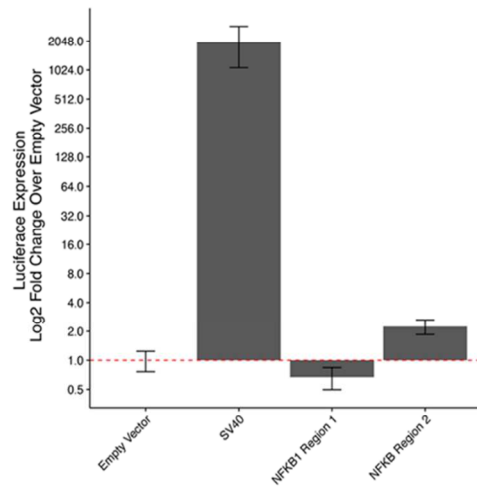


Figure E10. Boxplot showing increasing number of copies of low BDR-associated T allele of rs28450894 is associated with decreased expression of *SLC39A8* in blood regardless of sex ($p = 0.0066$, FDR adjusted $p = 0.0856$, $\log_2(\beta) = -0.327$).

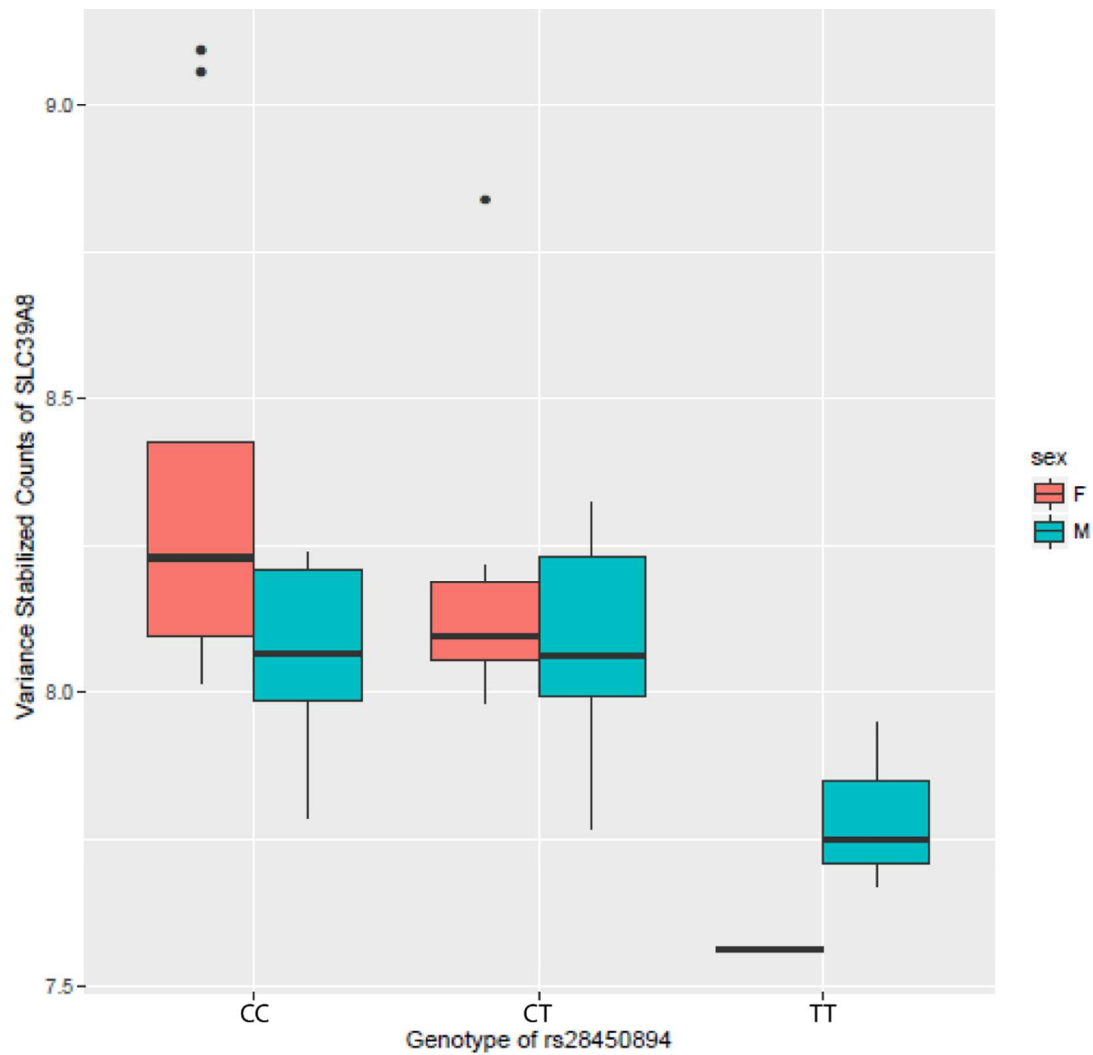
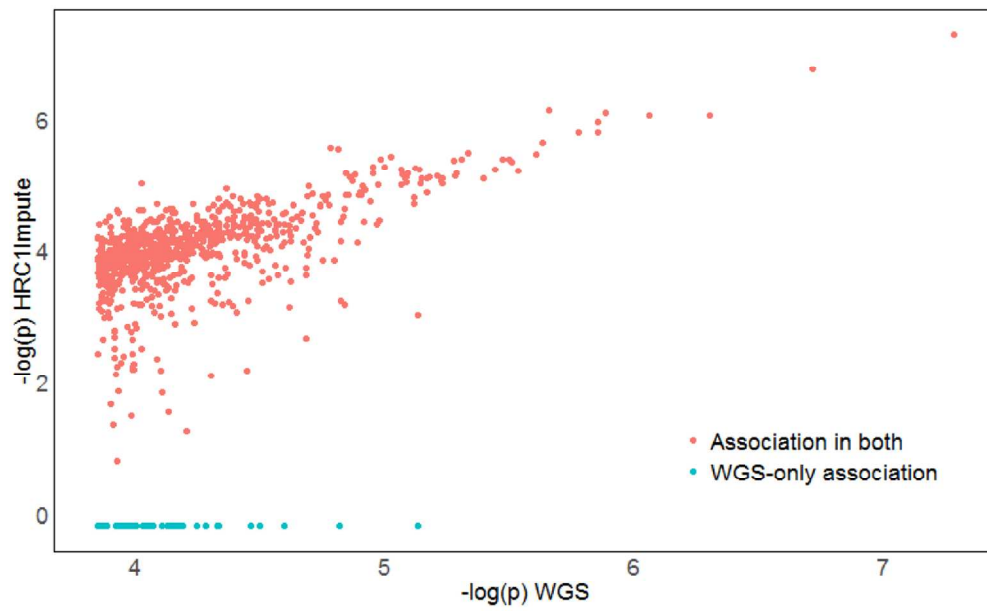
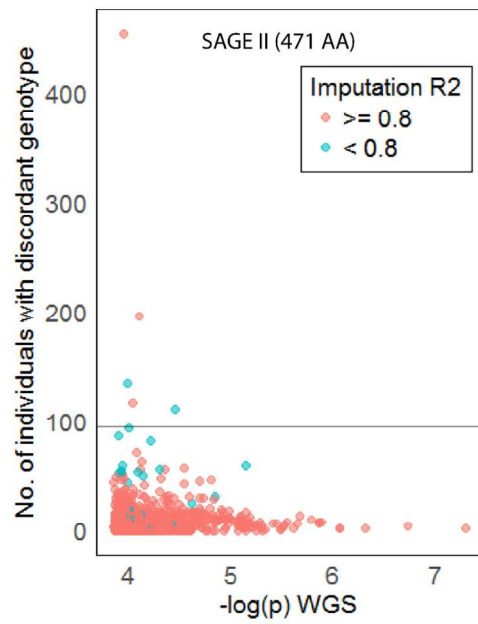


Figure E11. A comparison of the top 1000 BDR associations between the array-based and WGS-based genotype data. **(A)** A plot of association p -values of array-based and WGS-based data. The p -values from the two data types showed high correlation, especially for SNPs with more significant p -values. Genotype discordance of SAGE II **(B)** and GALA II **(C)** WGS SNPs. The corresponding imputation R^2 of the same SNPs in the HRC imputed array data is indicated by red ($R^2 \geq 0.8$) or blue ($R^2 < 0.8$).

A



B



C

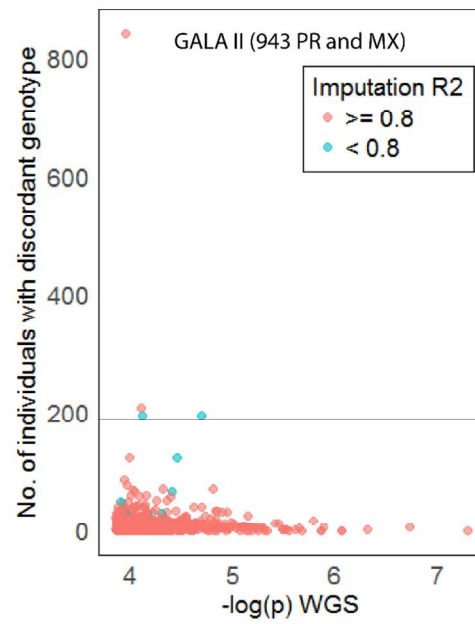


Figure E12. Percentage genotype concordance between Axiom LAT1 array and WGS genotypes. Grey horizontal lines mark one, two and three standard deviations (S.D.) from the mean percentage genotype concordance.

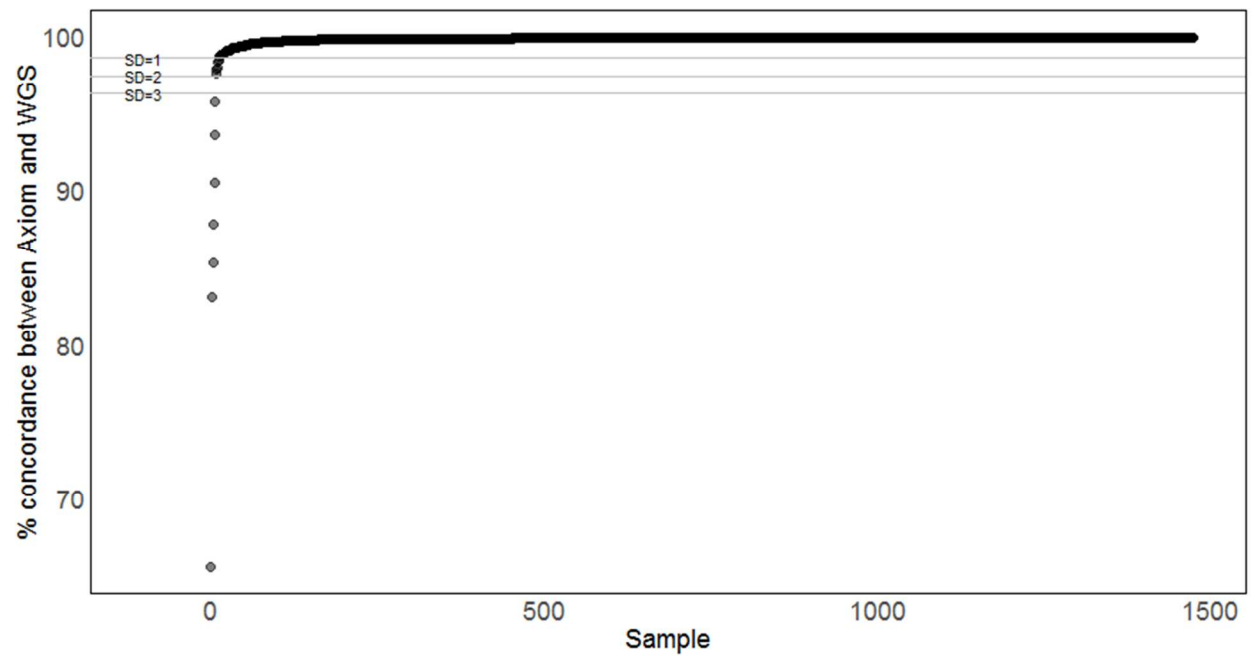


Figure E13. Quantile-quantile (q-q) plots and genomic inflation factors (λ_{GC}) for single locus BDR association tests among **(A)** Puerto Ricans, **(B)** Mexicans, **(C)** African Americans and **(D)** trans-ethnic meta-analysis.

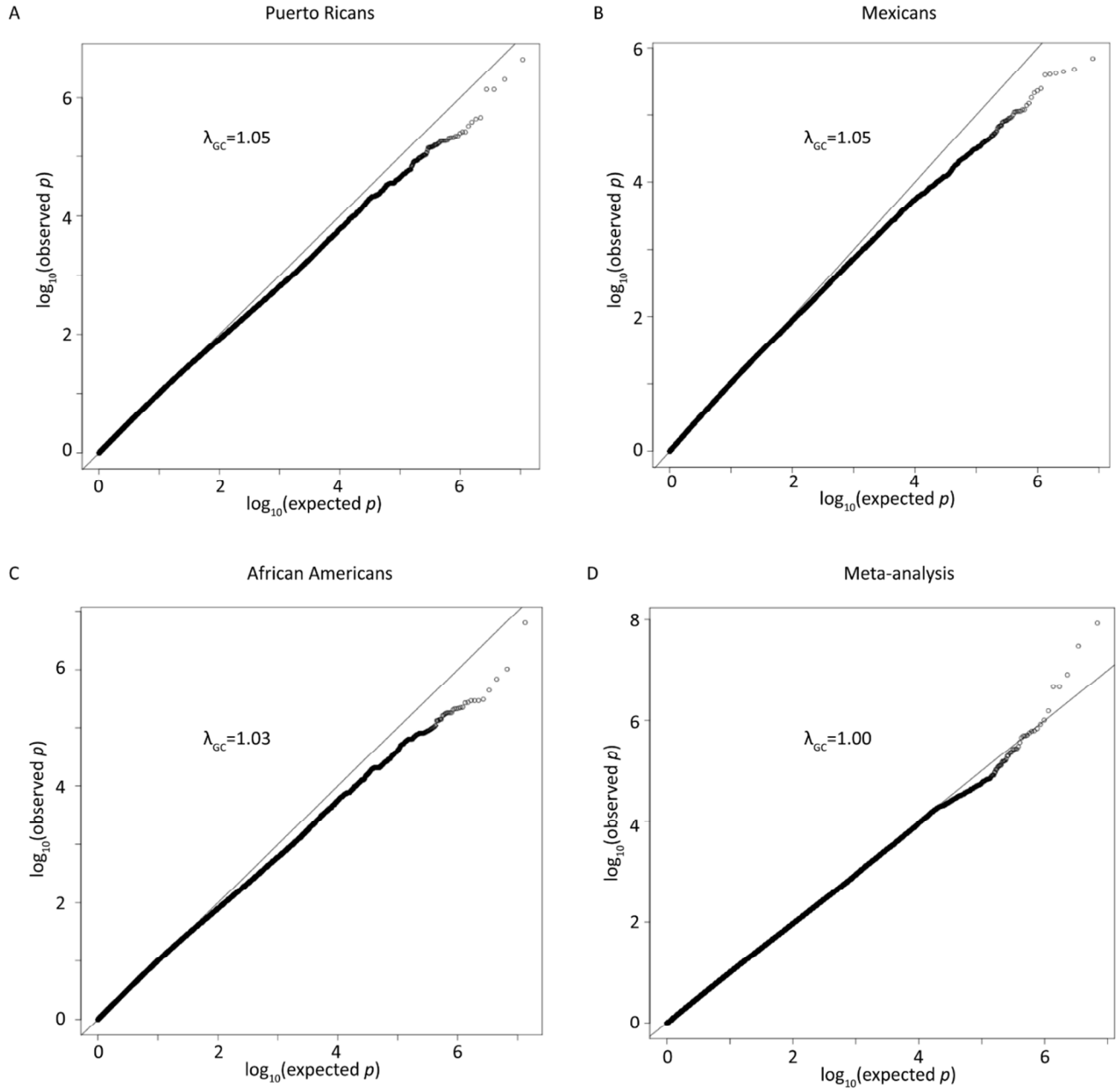
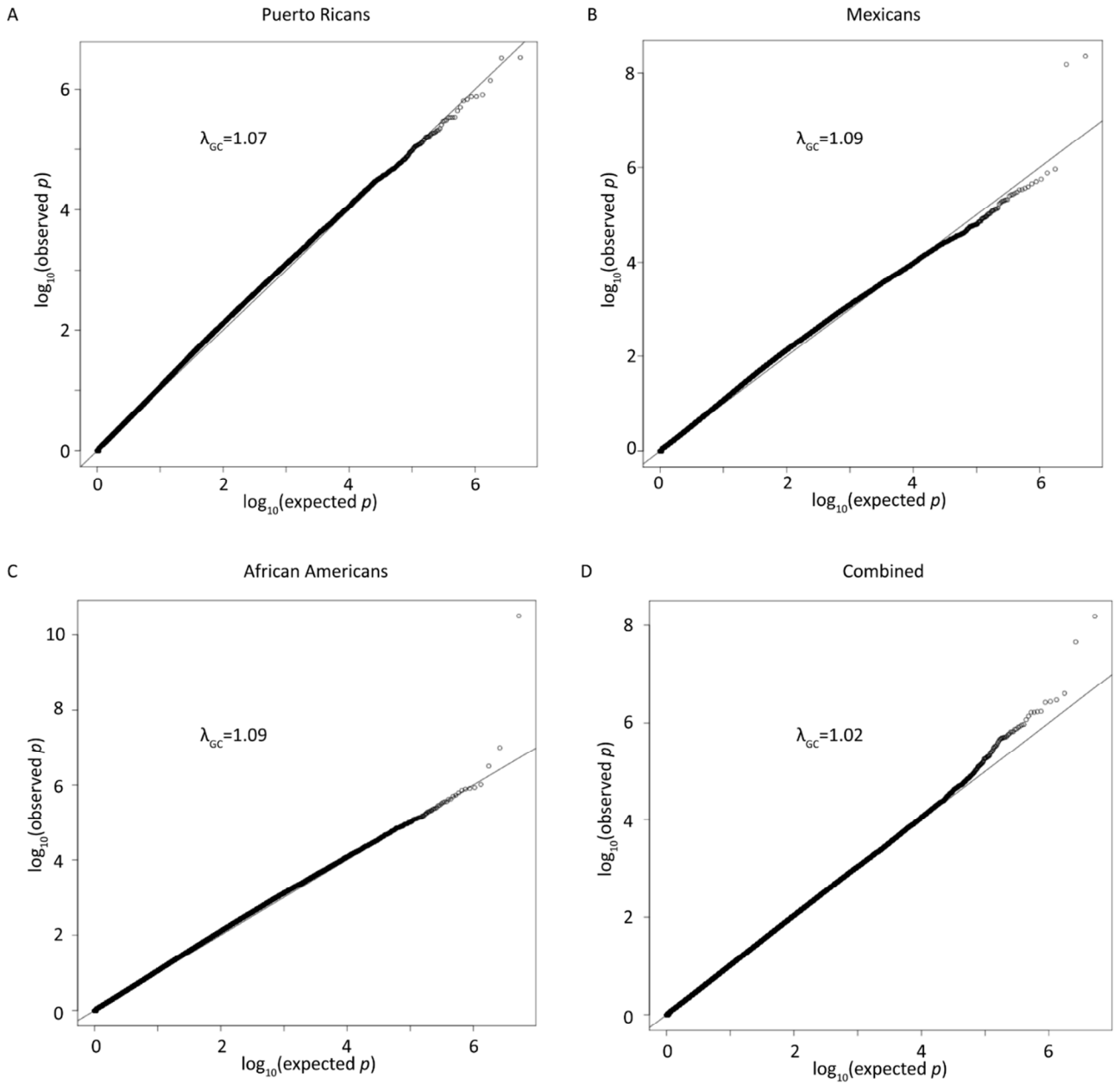


Figure E14. Quantile-quantile (q-q) plots and genomic inflation factors (λ_{GC}) for SKAT-O analysis of biallelic common and rare SNPs grouped by 1 kb windows for **(A)** Puerto Ricans, **(B)** Mexicans, **(C)** African Americans and **(D)** all three populations combined.



SUPPLEMENTARY TABLES

Table E1. Sequencing quality control metrics.

	Median	S.D.	Minimum	Maximum
Mean genome coverage*	35.2X	4.7X	29.7X	72.6X
Percentage of genome covered $\geq 20X$	92.7	1.9	86.2	97.9
Median insert size	380	23	307	497
Contamination**	0.0014	0.0021	0	0.0243
SNP array concordance	0.999	0.00062	0.994	1.000
Ti/Tv ratio (all variants)	1.94	0.011	1.9	2.05
Ti/Tv ratio (known variants)	2.05	0.006	2.04	2.09

* After removing duplicates.

** Estimated with VerifyBamID.

SNP array: internal SNP array (Illumina HumanCoreExome array) assayed by the New York Genome Center for sample quality control.

Table E2. Variant summary statistics.

	Puerto Ricans	Mexicans	African Americans
Range in variants, per individual (mean)			
All	3.7M – 4.4 M (4.0 M)	3.5M – 4.1M (3.7M)	4.0M – 4.5M (4.3M)
SNPs	3.4M – 4.0 M (3.6M)	3.2M - 3.7M (3.4M)	3.6M – 4.1M (4.0M)
Indels	284,067 – 344,493 (305,618)	272,635 – 321,778 (289,096)	311,997 – 354,646 (339,570)
Others*	21,233 – 29,177 (24,728)	19,703 – 26,772 (23,190)	22,014 – 30,993 (27,911)
No. of biallelic SNPs, union**	29.2M	28.1M	36.3M
By allele frequency of data			
Rare (< 1%)	18,169,292 (62%)	20,029,291 (71%)	22,847,938 (63%)
Common (≥ 1%)	11,007,125 (38%)	8,092,157 (29%)	13,459,004 (37%)
Population-specific***	6,680,909	9,687,651	14,114,142
Singleton	4,616,383 (69%)	7,983,950 (82%)	10,574,879 (75%)
< 5%	6,676,286 (> 99%)	9,675,271 (> 99%)	14,096,844 (> 99%)
≥ 5%	4,623 (< 1%)	12,380 (< 1%)	17,298 (< 1%)
By novelty (dbSNP150)			
Known	27,034,179 (93%)	25,911,801 (92%)	33,540,387 (92%)
Novel	2,142,238 (7%)	2,209,647 (8%)	2,766,555 (8%)
By protein impact****			
Coding	284,269 (1%)	289,055 (1%)	362,823 (1%)
<i>Nonsynonymous</i>	157,541	164,312	203,029
<i>Stopgain / stoploss</i>	3,194	3,531	4,158
<i>Splicing</i>	2,111	2,214	2,620
<i>Other coding</i>	121,423	118,998	153,016
Noncoding	28,833,461 (99%)	27,769,434 (99%)	35,878,723 (99%)
Not annotated^	58,687 (< 1%)	62,959 (< 1%)	65,396 (< 1%)

* Includes multi-nucleotide polymorphism (MNPs), complex, symbolic and mixed variants as defined by GATK VariantEval.

** Biallelic SNPs with less than 10% genotype missingness per population were included.

*** Biallelic SNPs that are present in only one of the three studied populations.

**** Other coding variants include those annotated as exonic and synonymous in ANNOVAR.

^ Not annotated: biallelic SNPs not included in the annotation pipeline because they are not present in TOPMed freeze 2 and 3 data releases.

Table E3. CADD score summary statistics of biallelic SNPs*.

Population	CADD score	No. of SNPs*	% Coding	% Singletons
Puerto Ricans	0-9	26,936,285	0.5%	29%
	10-19	2,027,392	4%	32%
	20-24	123,481	32%	37%
	≥ 25	30,572	>99%	51%
Mexicans	0-9	25,932,006	0.5%	44%
	10-19	1,967,096	4%	49%
	20-24	124,173	34%	55%
	≥ 25	35,214	>99%	70%
African Americans	0-9	33,468,245	0.5%	36%
	10-19	2,573,523	4%	40%
	20-24	158,967	32%	46%
	≥ 25	40,811	>99%	64%

*Biallelic SNPs with less than 10% genotype missingness per population were included.

Table E4. Chromosomal location of BDR-associated common variants.

rsID	Chr	Chromosomal Location (bp)	
		GRCh37	GRCh38
rs10746419	1	209324294	209150949
rs17048684	4	137382142	136460987
rs28450894	4	103453535	102532378
rs4648006	4	103461559	102540402
rs1017451	5	12975108	12974996
rs1017452	5	12975187	12975075
rs1017453	5	12975265	12975153
rs1017454	5	12975322	12975210
rs1438293	5	12959598	12959486
rs17237443	5	12972636	12972524
rs17237639	5	12975934	12975822
rs17833938	5	12961545	12961433
rs17834628	5	12978566	12978454
rs34845041	5	12948369	12948257
rs35661809	5	12968341	12968229
rs62347395	5	12950432	12950320
rs13200833	6	54581204	54716406
rs13437006	6	104240500	103792625
rs6926020	6	104235591	103787716
rs11044754	12	19820677	19667743
rs66544720	12	19821401	19668467
rs7313907	12	19824386	19671452
rs1565749	15	101230457	100690252
rs55638658	15	101233236	100693031
rs57924834	15	101231049	100690844
rs16995064	20	8635168	8654521
rs60163793	22	27826429	27430468

Table E5. Variation in BDR status explained by significant and suggestively associated variants identified by trans-ethnic meta-analysis and SKAT-O.

Common Variants					
Population	R ²				
Puerto Ricans	0.23				
Mexicans	0.16				
African Americans	0.18				
Combined Common and Rare Variants					
Population	SKAT-O Regions (R ²)				
	Chr1	Chr11	Chr19	Chr4	Chr8
Puerto Ricans	--	--	--	--	--
Mexicans	0.08	0.08	--	--	--
African Americans	--	--	0.05	--	--
Combined	--	--	--	0.04	0.04

R² refers to McFadden's Pseudo R². Models used to calculate R² were adjusted for age, sex, BMI category, and the first 10 principal components. For common variants, the R² refers to the combined effect of all 27 common variants shown in **Table 2**. For SKAT-O regions, R² was calculated only for the population in which the significantly associated region was identified (**Table 3**) and describes the combined effect of all variants located within the specified region.

Table E6. Common variant trans-ethnic BDR association tests adjusted and unadjusted for local ancestry.

Chr	Start	rsID	Effect allele	Unadjusted for local ancestry		Adjusted for local ancestry	
				OR (95%CI)	<i>p</i>	OR (95%CI)	<i>p</i>
5	12978566	rs17834628	A	1.67 (1.29-2.16)	1.18E-08	1.65 (1.20-2.26)	4.73E-08
5	12968341	rs35661809	G	1.59 (1.20-2.10)	3.33E-08	1.57 (1.13-2.17)	1.02E-07
22	27826429	rs60163793	G	2.01 (1.20-3.38)	4.30E-06	2.19 (1.31-3.69)	4.93E-07
5	12975934	rs17237639	G	1.61 (1.30-2.00)	1.22E-07	1.59 (1.22-2.08)	5.41E-07
5	12975187	rs1017452	G	1.60 (1.31-1.96)	2.11E-07	1.58 (1.21-2.06)	9.85E-07
5	12975322	rs1017454	A	1.60 (1.31-1.96)	2.11E-07	1.58 (1.21-2.06)	9.85E-07
15	101231049	rs57924834	A	1.59 (1.25-2.03)	2.04E-06	1.62 (1.26-2.07)	1.35E-06
15	101230457	rs1565749	A	1.66 (1.18-2.32)	1.64E-06	1.66 (1.19-2.33)	1.56E-06
4	137382142	rs17048684	A	1.8 (1.06-3.05)	2.20E-06	1.80 (1.06-3.07)	2.01E-06
20	8635168	rs16995064	G	1.96 (1.12-3.43)	3.30E-06	1.97 (1.08-3.60)	2.45E-06
6	104240500	rs13437006	C	1.56 (1.21-2.02)	1.61E-06	1.56 (1.23-1.98)	2.55E-06
5	12975265	rs1017453	C	1.56 (1.25-1.95)	6.40E-07	1.53 (1.15-2.05)	2.64E-06
4	103453535	rs28450894	T	0.47 (0.34-0.64)	3.75E-06	0.46 (0.34-0.64)	2.76E-06
4	103461559	rs4648006	T	0.47 (0.34-0.64)	3.75E-06	0.46 (0.34-0.64)	2.76E-06
1	209324294	rs10746419	T	1.29 (0.75-2.25)	1.19E-06	1.27 (0.73-2.22)	2.86E-06
5	12972636	rs17237443	C	1.59 (1.28-1.97)	9.85E-07	1.55 (1.12-2.16)	3.65E-06
15	101233236	rs55638658	A	1.61 (1.13-2.30)	5.08E-06	1.62 (1.13-2.31)	5.00E-06
5	12961545	rs17833938	A	1.56 (1.28-1.91)	1.45E-06	1.52 (1.12-2.06)	5.99E-06
5	12975108	rs1017451	T	1.55 (1.24-1.93)	1.96E-06	1.51 (1.11-2.07)	6.64E-06
5	12948369	rs34845041	T	1.56 (1.26-1.92)	1.77E-06	1.52 (1.12-2.06)	7.17E-06
6	104235591	rs6926020	C	1.57 (1.25-1.97)	3.68E-06	1.56 (1.26-1.94)	7.25E-06
12	19821401	rs66544720	T	0.66 (0.55-0.78)	3.66E-06	0.66 (0.55-0.79)	7.73E-06
5	12950432	rs62347395	G	1.55 (1.26-1.92)	2.02E-06	1.51 (1.11-2.06)	8.05E-06
12	19824386	rs7313907	C	0.66 (0.55-0.79)	4.35E-06	0.66 (0.56-0.79)	9.08E-06
5	12959598	rs1438293	G	1.55 (1.24-1.93)	2.73E-06	1.50 (1.07-2.11)	9.43E-06
12	19820677	rs11044754	A	0.66 (0.55-0.79)	4.54E-06	0.66 (0.56-0.79)	9.65E-06
6	54581204	rs13200833	A	0.66 (0.48-0.90)	5.15E-06	0.66 (0.47-0.92)	1.40E-05

Table E7. Lung-related phenotypes previously reported for BDR-associated common variants and their nearest genes.

Gene	Common variants associated with BDR (Table 2)	Previously reported lung-related phenotypes for gene	PubMed ID	Population
<i>CTNND2</i>	rs17834628, rs35661809, rs17237639, rs1017452, rs1017454, rs1017453, rs17237443, rs17833938, rs34845041, rs1017451, rs62347395, rs1438293	Pulmonary function decline	PMC3340499	EUR
<i>DNAH5</i>	rs17834628, rs35661809, rs17237639, rs1017452, rs1017454, rs1017453, rs17237443, rs17833938, rs34845041, rs1017451, rs62347395, rs1438293	Post bronchodilator FEV1 and FEV1/FVC ratio in COPD, ciliary dyskinesia	PMC1995616 PMC4668640 PMC5355724 PMC4169636	EUR, AFR, EAS
<i>LAMB3</i>	rs10746419	Lung cancer	PMC3575388	EAS
<i>LIN28B</i>	rs13437006, rs6926020	Lung cancer	PMC5316988 PMC4155498	EUR, AFR
<i>PLCB1</i>	rs16995064	FEV1 and FEV1/FVC ratio in COPD, differentially expressed in therapy-resistant children, protects against endothelial cell inflammation	PMC3824029 PMC4593004 PMID: 25863981	EUR, AFR
<i>PLEKHA5</i>	rs66544720, rs7313907, rs11044754	COPD	PMC3137139	EUR
<i>SLC39A8</i>	rs28450894, rs4648006	Mucus regulation, protection against cytotoxicity in airway epithelium	PMC1899314 PMC3257720 PMID: 18390834	-
<i>NFKB1</i>	rs28450894, rs4648006	Inflammation and immune response	PMC444781 PMC5320420 PMC5289825 PMID: 18950657 PMID: 9895331	SAS, EUR
<i>MN1</i>	rs60163793	Lung function (FEV1)	PMID: 28166213	EUR
<i>PITPNB</i>	rs60163793	Lipid synthesis in lung maturation	PMC4012993	-
<i>FAM83B</i>	rs13200833	Lung squamous cell carcinoma	PMC4324586	EAS

BDR-associated common variants and their nearest genes (**Table 2**) were searched against the Ensembl, PubMed and GWAS Catalog databases to identify any associations with lung-related phenotypes. **EUR**: European; **AFR**: African; **EAS**: East Asian; **SAS**: South Asian.

Table E8. Functional annotations for BDR-associated common variants in Table 2.

rsID	Position (Chr:start)	Nearest Gene	Distance: SNP to feature (bp)			Chrom state	Overlap with H3K27ac ChIP-seq peaks in BSMC
			CpG	DHS	TFBS cluster		
rs10746419	1:209324294	MIR205HG (278k)	80552	-464	-364 (YY1)		
rs11044754	12:19820677	AEBP2 (-146k)	-226992	-1247	-132 (CEBPB)		Yes
rs66544720	12:19821401	AEBP2 (-146k)	-227716	1460	-856 (CEBPB)		Yes
rs7313907	12:19824386	AEBP2 (-149k)	-230701	355	171 (GATA3)		Yes
rs1565749	15:101230457	ASB7 (-39k)	-87320	0	0 (EP300)		Yes
rs57924834	15:101231049	ASB7 (-39k)	-87912	0	0 (GATA3)		Yes
rs55638658	15:101233236	ASB7 (-41k)	-90099	-466	-511 (EP300)		Yes
rs16995064	20:8635168	PLCB1 (intron 7)	413791	73	0 (SETDB1)		
rs60163793	22:27826429	MN1 (318k)	7997	377	1814 (JUN)		
rs28450894	4:103453535	NFKB1 (intron 3)	-30306	-25	1332 (SETDB1)		Yes
rs4648006	4:103461559	NFKB1 (intron 5)	-38330	1362	757 (GATA3)		Yes
rs17048684	4:137382142	LINC00613 (-547k)	-104655	-10692	1222 (SIN3A)		
rs34845041	5:12948369	LINC01194 (-143k)	-152722	-1039	623 (CEBPB)		
rs62347395	5:12950432	LINC01194 (-145)	-154785	1474	-2452 (CTCF)		
rs1438293	5:12959598	LINC01194 (-154k)	-163951	-6828	1657 (FOS)		
rs17833938	5:12961545	LINC01194 (-156k)	-165898	-8775	297 (MAFK)		
rs35661809	5:12968341	LINC01194 (-163k)	-172694	-15571	577 (FOS)		
rs17237443	5:12972636	LINC01194 (-170k)	-176989	-19866	4475 (FAM48A)		
rs1017451	5:12975108	LINC01194 (-170k)	-179461	20518	180 (FAM48A)		
rs1017452	5:12975187	LINC01194 (-170k)	-179540	20439	-2053 (FAM48A)		
rs1017453	5:12975265	LINC01194 (-170k)	-179618	20361	-2132 (FAM48A)		
rs1017454	5:12975322	LINC01194 (-170k)	-179675	20304	-2210 (FAM48A)		
rs17237639	5:12975934	LINC01194 (-171k)	-180287	19692	-2267 (FAM48A)		
rs17834628	5:12978566	LINC01194 (-173k)	-182919	17060	-2879 (FAM48A)		
rs13200833	6:54581204	FAM83B (130k)	129950	577	1136 (FAM48A)		
rs6926020	6:104235591	HACE1 (940k)	1071529	-3321	1391 (TEAD4)		
rs13437006	6:104240500	HACE1 (935k)	1066620	5201	-349 (JUN)		

Nearest gene and functional annotations were based on ENCODE annotations from the UCSC Table Browser (Feb.2009 [GRCh37/hg19] assembly). Negative distances indicate upstream features. SNPs within a gene (distance of 0 base pairs) are labeled with the intron or exon number. **CpG**: CpG islands. **DHS**: DNase I Sensitive Sites. **TFBS clusters**: Transcription Factor Binding Site regions aggregated across 91 cell types by ENCODE. For variants overlapping multiple TFBS, the sites are listed in order of descending ChIP-seq signal. Corresponding transcription factors are indicated in parentheses. **Chrom State**: chromatin state. Common chromatin state information determined by the BROAD Institute using a hidden Markov model to integrate ChIP-seq for nine canonical histone markers across nine tissue types, including lung fibroblasts.  = Heterochromatin  = Weak transcription markers  = Transcriptional elongation markers. **BSMC**: basal smooth muscle cells. H3K27ac ChIP-seq peaks determined experimentally within current study; H3K27ac overlapping was defined as any overlap of H3K27ac ChIP-seq peaks with variants reported in the table or with variants in high linkage disequilibrium ($R^2 \geq 0.8$ within 1 Mb) with the variants reported; Yes = overlap, otherwise no overlap.

Table E9. Common variant BDR association results by population.

Chr	Location (bp)	rsID	Effect allele	Puerto Rican N=483		Mexican N=483		African American N=475	
				OR (95%CI)	p	OR (95%CI)	p	OR (95%CI)	p
1	209324294	rs10746419	T	0.80 (0.62-1.04)	9.63E-02	1.29 (0.98-1.69)	7.11E-02	2.12 (1.60-2.80)	1.59E-07*
4	137382142	rs17048684	A	3.00 (1.88-4.77)	3.84E-06	1.77 (1.19-2.63)	5.15E-03	1.16 (0.82-1.63)	4.03E-01
4	103453535	rs28450894	T	0.51 (0.29-0.89)	1.83E-02	0.51 (0.23-1.15)	1.04E-01	0.44 (0.29-0.68)	1.86E-04
4	103461559	rs4648006	T	0.51 (0.29-0.89)	1.83E-02	0.51 (0.23-1.15)	1.04E-01	0.44 (0.29-0.68)	1.86E-04
5	12975108	rs1017451	T	1.89 (1.42-2.52)	1.44E-05	1.39 (1.05-1.83)	2.12E-02	1.34 (0.90-2.01)	1.54E-01
5	12975187	rs1017452	G	1.90 (1.43-2.53)	9.49E-06	1.37 (1.04-1.81)	2.60E-02	1.57 (1.08-2.29)	1.77E-02
5	12975265	rs1017453	C	1.91 (1.44-2.54)	7.12E-06	1.36 (1.03-1.79)	3.06E-02	1.44 (0.99-2.10)	5.63E-02
5	12975322	rs1017454	A	1.90 (1.43-2.53)	9.49E-06	1.37 (1.04-1.81)	2.60E-02	1.57 (1.08-2.29)	1.77E-02
5	12959598	rs1438293	G	1.88 (1.41-2.51)	1.96E-05	1.41 (1.07-1.86)	1.56E-02	1.29 (0.83-2.00)	2.51E-01
5	12972636	rs17237443	C	1.92 (1.43-2.56)	1.15E-05	1.39 (1.06-1.84)	1.94E-02	1.44 (0.92-2.26)	1.13E-01
5	12975934	rs17237639	G	1.93 (1.45-2.56)	5.45E-06	1.37 (1.04-1.81)	2.54E-02	1.59 (1.09-2.31)	1.58E-02
5	12961545	rs17833938	A	1.86 (1.40-2.48)	1.68E-05	1.40 (1.06-1.85)	1.84E-02	1.37 (0.90-2.09)	1.40E-01
5	12978566	rs17834628	A	2.08 (1.56-2.76)	4.90E-07*	1.37 (1.04-1.81)	2.56E-02	1.63 (1.13-2.35)	9.05E-03
5	12948369	rs34845041	T	1.86 (1.40-2.48)	1.68E-05	1.36 (1.03-1.79)	3.05E-02	1.44 (0.94-2.21)	9.25E-02
5	12968341	rs35661809	G	2.10 (1.59-2.78)	2.26E-07*	1.39 (1.06-1.83)	1.89E-02	1.36 (0.99-1.87)	5.60E-02
5	12950432	rs62347395	G	1.86 (1.40-2.48)	1.68E-05	1.36 (1.03-1.80)	3.02E-02	1.42 (0.93-2.17)	1.08E-01
6	54581204	rs13200833	A	0.50 (0.37-0.68)	1.00E-05	0.86 (0.63-1.17)	3.27E-01	0.66 (0.47-0.93)	1.59E-02
6	104240500	rs13437006	C	2.06 (1.46-2.89)	3.36E-05	1.28 (0.94-1.74)	1.19E-01	1.50 (1.13-1.98)	4.44E-03
6	104235591	rs6926020	C	2.02 (1.41-2.90)	1.40E-04	1.33 (0.97-1.83)	7.83E-02	1.51 (1.12-2.02)	6.63E-03
12	19820677	rs11044754	A	0.64 (0.48-0.85)	1.94E-03	0.63 (0.48-0.83)	1.29E-03	0.76 (0.53-1.08)	1.25E-01
12	19821401	rs66544720	T	0.63 (0.48-0.84)	1.75E-03	0.63 (0.48-0.83)	1.29E-03	0.75 (0.52-1.07)	1.13E-01
12	19824386	rs7313907	C	0.63 (0.48-0.84)	1.75E-03	0.63 (0.48-0.83)	1.29E-03	0.76 (0.53-1.09)	1.31E-01
15	101230457	rs1565749	A	1.96 (1.37-2.81)	2.51E-04	1.17 (0.81-1.69)	4.12E-01	1.96 (1.38-2.79)	1.66E-04
15	101233236	rs55638658	A	1.93 (1.34-2.76)	3.73E-04	1.12 (0.77-1.61)	5.56E-01	1.94 (1.36-2.77)	2.53E-04
15	101231049	rs57924834	A	1.70 (1.24-2.34)	1.09E-03	1.26 (0.90-1.75)	1.73E-01	1.90 (1.36-2.67)	2.00E-04
20	8635168	rs16995064	G	2.63 (1.69-4.08)	1.78E-05	1.19 (0.80-1.77)	3.89E-01	2.58 (1.38-4.82)	3.02E-03
22	27826429	rs60163793	G	3.53 (1.66-7.51)	1.08E-03	2.21 (1.48-3.32)	1.20E-04	1.30 (0.91-1.87)	1.51E-01

All models were adjusted for age, sex, BMI category, and the top 10 genome-wide principal components. ** Denotes genome-wide significance; *suggestive significance. ΔFEV_1 cutoff to define high and low responders and p -value significance thresholds are population-specific; **Puerto Rican**: high BDR > 16.29, low BDR < 7.23. ** $p \leq 1.57\text{E-}7$. * $p < 3.14\text{E-}06$; **Mexican**: high BDR > 8.55, low BDR < 6.05. ** $p \leq 2.42\text{E-}7$. * $p < 4.84\text{E-}06$; **African American**: high BDR > 11.81, low BDR < 5.53. ** $p \leq 9.59\text{E-}8$. * $p < 1.92\text{E-}06$.

Table E10. Descriptive statistics for replication cohorts.

Descriptive Statistics		SAPPHIRE African American N = 1,022			HPR Puerto Rican N = 414			CHOP African American N = 280			GALA I						SAGE I African American N = 141		
		High BDR	Low BDR	<i>p</i>	High BDR	Low BDR	<i>p</i>	High BDR	Low BDR	<i>p</i>	High BDR	Low BDR	<i>p</i>	High BDR	Low BDR	<i>p</i>	High BDR	Low BDR	<i>p</i>
Number of Subjects		465	557	-	42	372	-	158	122	-	105	97	-	18	90	-	57	84	-
% Male		38.30	35.40	0.37	59.50	53.50	0.46	50.00	51.00	0.90	47.60	58.80	0.11	61.11	61.11	1.00	49.10	35.70	2.9e-03
Median Age, yr (IQR)		30 (20-47)	28 (17-46)	0.23	9 (7-9)	10 (8-12)	0.01	10 (8-13)	11 (8-14)	0.77	12 (10-24)	13 (10-17)	0.33	12 (9-14)	11 (9-15)	0.89	17 (12-23)	15 (12-22)	0.24
Mean Global Ancestry Proportions	AFR	0.8	0.79	0.16	0.24	0.22	0.49	-	-	-	0.08	0.09	1.9e-03	0.2	0.2	0.71	0.83	0.81	0.74
	EUR	0.2	0.21	-	0.63	0.65	0.45	-	-	-	0.4	0.41	0.19	0.6	0.61	0.74	0.17	0.19	0.74
	NAM	-	-	-	0.12	0.13	0.58	-	-	-	0.52	0.5	0.12	0.2	0.19	0.37	-	-	-
BMI Category, N	Obese	188	266	1	12	103	0.9	27	24	0.64	28	39	0.05	5	34	0.59	21	42	0.17
	Non-Obese	277	331		30	269		131	98		77	58		13	56		36	42	
Pre-FEV ₁ % Predicted, N	< 80%	291	61	<0.001	25	123	<0.001	63	33	0.03	52	10	<0.001*	12	21	<0.001	23	7	<0.001
	≥ 80%	174	496	1	17	249		95	89		52	86		6	69		34	77	
Median ΔFEV ₁ , % (IQR)		19.53 (15.0 - 29.0)	1.4 (-1.2 - 3.4)	<0.001	20.2 (17.8 - 27.3)	1.0 (-2.4 - 3.9)	<0.001	18.48 (14.42 - 27.34)	3.24 (1.59 - 4.64)	<0.001	15.13 (11.11 - 24.23)	2.94 (1.30 - 4.64)	<0.001	18.95 (18.12 - 23.24)	1.29 (-1.43 - 3.34)	<0.001	16.67 (13.82 - 24.12)	3.16 (0.91 - 4.35)	<0.001
Median tIgE (IQR)		-	-	-	267 (66 - 612)	201 (56 - 611)	0.43	-	-	-	386 (160.5 - 692.75)	271 (90.85 - 465)	0.02	382 (175.25 - 663.25)	243 (59.77 - 499)	0.24	150.65 (51.1 - 393.97)	92.3 (33.65 - 399.5)	0.35
Atopy, N		-	-	-	29	236	0.47	-	-	-	45	50	0.01	14	54	0.41	31	39	0.29

BDR: bronchodilator drug response. **IQR:** interquartile range. **Pre-FEV₁ % predicted:** percentage of measured FEV₁ relative to predicted FEV₁ estimated by the Hankinson lung function prediction equations, prior to administration to albuterol. **ΔFEV₁:** the percent change in baseline FEV₁ after administration of albuterol. **tIgE:** measure of total Immunoglobulin E (IgE) from serum. **Atopy:** tIgE score greater than or equal to 100. *p*-value for percent male was determined by a proportion test; *p*-values for age, global ancestry, ΔFEV₁ %, and tIgE were calculated with Wilcoxon's rank-sum; *p*-values for BMI category, pre-FEV₁ % predicted status and atopy status were calculated with Fisher's exact test. *Two individuals (one BDR high, one BDR low) had incomplete information for this variable. - Data not available.

Table E11. Replication of common variant associations.

Chr	rsID	Effect allele	SAPPHIRE		HPR		CHOP		GALA I				SAGE I		Replication meta-analysis	
			African American (N=1,022)		Puerto Rican (N=414)		African American (N=280)		Mexican (N=202)		Puerto Rican (N=108)		African American (N=141)			
			OR	p	OR	p	OR	p	OR	p	OR	p	OR	p	OR	p
1	rs10746419	T	1.06	0.55	0.85	0.48	1.08	0.70	1.33	0.19	1.38	0.46	1.36	0.30	1.09	0.30
4	rs17048684	A	0.98	0.88	1.14	0.73	0.65	0.09	1.06	0.84	2.64	0.19	0.71	0.42	0.94	0.59
4	rs28450894	T	1.08	0.56	0.74	0.55	0.94	0.81	0.60	0.38	0.68	0.64	0.87	0.38	0.99	0.97
4	rs4648006	T	1.09	0.49	-	-	0.93	0.78	0.53	0.27	0.68	0.64	0.87	0.38	1.00	0.99
5	rs1017451	T	1.21	0.15	0.89	0.66	1.09	0.73	0.92	0.70	1.05	0.91	2.04	0.39	1.12	0.26
5	rs1017452	G	1.09	0.48	0.87	0.61	1.13	0.62	0.92	0.72	1.04	0.93	1.36	0.37	1.05	0.63
5	rs1017453	C	1.10	0.46	0.87	0.61	1.09	0.73	0.90	0.64	1.04	0.93	1.28	0.37	1.04	0.70
5	rs1017454	A	1.12	0.37	0.87	0.61	1.13	0.62	0.93	0.74	1.04	0.93	1.36	0.37	1.07	0.54
5	rs1438293	G	1.13	0.42	0.90	0.70	0.88	0.65	1.01	0.95	1.13	0.76	2.16	0.43	1.08	0.53
5	rs17237443	C	1.06	0.70	0.93	0.79	0.84	0.55	0.95	0.83	0.98	0.96	1.76	0.40	1.02	0.90
5	rs17237639	G	1.10	0.45	0.93	0.80	1.13	0.62	0.92	0.71	1.03	0.94	1.36	0.37	1.06	0.56
5	rs17833938	A	1.14	0.37	0.89	0.66	1.05	0.87	0.95	0.81	1.26	0.57	2.38	0.41	1.11	0.37
5	rs17834628	A	1.00	0.98	1.02	0.94	0.72	0.19	1.01	0.98	0.97	0.95	1.35	0.37	0.98	0.88
5	rs34845041	T	1.16	0.29	0.92	0.76	1.05	0.87	0.94	0.76	1.28	0.56	2.38	0.41	1.12	0.31
5	rs35661809	G	0.95	0.66	-	-	1.03	0.89	0.97	0.89	1.24	0.60	1.55	0.37	1.00	0.99
5	rs62347395	G	1.16	0.29	0.92	0.75	1.05	0.87	0.94	0.76	1.27	0.56	2.38	0.41	1.12	0.31
6	rs13200833	A	1.23	0.06	0.84	0.52	0.93	0.76	1.13	0.64	0.74	0.49	1.05	0.31	1.10	0.29
6	rs13437006	C	0.90	0.29	1.46	0.19	1.03	0.88	1.59	0.08	1.59	0.31	1.15	0.28	1.11	0.52
6	rs6926020	C	0.89	0.26	1.23	0.51	1.03	0.89	1.53	0.11	1.61	0.32	1.00	0.29	1.01	0.99
12	rs11044754	A	0.96	0.75	1.63	0.10	0.96	0.86	1.19	0.45	0.54	0.19	1.06	0.34	1.03	0.81
12	rs66544720	T	0.96	0.75	1.63	0.10	-	-	1.19	0.45	0.54	0.19	1.06	0.34	1.06	0.75
12	rs7313907	C	0.95	0.67	1.63	0.10	1.00	0.99	1.19	0.44	0.54	0.19	1.06	0.34	1.03	0.82
15	rs1565749	A	1.01	0.92	1.33	0.34	1.07	0.78	0.76	0.36	2.01	0.17	0.88	0.31	1.03	0.80
15	rs55638658	A	1.04	0.74	1.32	0.34	1.25	0.36	0.75	0.34	1.68	0.32	0.89	0.32	1.06	0.58
15	rs57924834	A	0.98	0.88	1.20	0.52	1.16	0.52	0.96	0.87	1.49	0.42	0.87	0.31	1.02	0.83
20	rs16995064	G	0.74	0.14	0.91	0.81	0.70	0.30	0.60	0.09	0.77	0.72	2.06	0.54	0.77	0.07
22	rs60163793	G	1.13	0.36	0.80	0.70	0.89	0.71	2.89	0.01**	-	-	1.17	0.37	1.20	0.22

Logistic regression and meta-analysis were performed with the same covariates and the same definition of high and low responders as in the discovery analysis.

** $p < 0.01$.

– Data not available due to failing genotyping QC.

Table E12. Diverse Convergent Evidence (DiCE) scoring rubric.

Statistical (Observational) Evidence	Informatic (Biological Database) Evidence	Experimental (Laboratory) Evidence
<i>Significant/Suggestive Association in Discovery Population*</i>	<i>Evidence from Ensembl, PubMed or NHGRI-EBI GWAS databases linking 4 closest genes to BDR, asthma, or lung-related phenotype#</i>	<i>SNP overlap with H3K27ac ChIP-seq peaks in BSMC[†]</i>
Yes (1 point)	Yes (2 points)	Yes (3 points)
No (0 points)	No (0 points)	No (0 points)
<i>And either</i>	<i>or</i>	
<i>Standard Statistical Validation¹</i>	<i>Evidence from ENCODE Database that SNP overlaps a functionally active region[‡]</i>	
Yes (3 point)	Yes (2 points)	
No (0 points)	No (0 points)	
<i>or</i>	<i>or</i>	
<i>Alternative Statistical Validation²</i>	<i>Evidence from two or more independent sources of informatic evidence</i>	
Yes (2 points)	Yes (3 points) [§]	
No (0 points)		

*Refers to results from trans-ethnic meta-analysis (**Table 2**); ¹Significant *p*-value in at least one independent replication analysis (**Table E10**); ²Significant *p*-value in replication meta-analysis (**Table E11**); #Detailed information presented in **Table E7**; [†]Detailed information presented in **Table E8**. [§]SNPs with two or more independent sources of informatic evidence are prioritized with an additional point.

Table E13. DNA inserts used in luciferase assay and key variants included within those sequences.

DNA Insert					Key Variant(s)						
Region	Chr	Start	Stop	size (bp)	Chr	Start	Stop	rsID	R ² with BDR associated SNPs		
									PR	MX	AA
1	4	103486503	103491377	4874	4	103487982	103487983	rs2293970	0.78	0.82	0.48
2	4	103527183	103531814	4631	4	103527876	103527877	rs4648095	0.78	0.82	0.47
					4	103530642	1035307643	rs28557116	0.78	0.82	0.47

GRCh37 coordinates are shown. R² refers to pairwise correlation between the key variants in two *NFKB1* regions above and the two variants in the *NFKB1* region suggestively associated with BDR status, rs28450894 and rs4648006; As rs28450894 and rs4648006 are in complete LD (R² = 1) only one value is presented above.

Table E14. Association of previously reported BDR-associated SNPs in the current study by population and trans-ethnic meta-analysis.

PMID	Pop	Gene	rsID	Lowest <i>p</i>	Effect allele	Direction of effect	Current Study							
							PR		MX		AA		Meta-analysis	
							OR	<i>p</i>	OR	<i>p</i>	OR	<i>p</i>	OR	<i>p</i>
21545619	Korean	<i>ADCY9</i>	rs2230739	0.03 *	C	+	1.12	0.48	0.91	0.59	1.05	0.79	1.03	0.81
22792082	European	<i>SPATS2L</i>	rs295137	9.70E-7 §	T	+	0.93	0.61	1.00	0.98	1.28	0.07	1.06	0.48
19514054	European	<i>CPS1</i>	rs1047891 [∅]	0.03 *	A	-	0.91	0.51	1.27	0.12	0.92	0.56	1.01	0.96
23508266	European	<i>THRB</i>	rs892940	1.2E-3 *	G	+	1.11	0.46	0.92	0.57	0.95	0.77	1.00	0.99
19514054	African American	<i>GSNOR</i>	rs1154400	0.02 *	T	+	0.97	0.84	1.12	0.56	1.27	0.10	1.11	0.29
16931635	African American	<i>ADRB2</i>	rs1042714	0.03 *	C	+	1.09	0.53	0.96	0.79	1.14	0.33	1.03	0.85
15557128	Puerto Rican	<i>ADRB2</i>	rs1042713	1.00E-6 *	A	-	0.94	0.71	1.02	0.92	1.15	0.44	1.06	0.49
18617639	European	<i>ARG1</i>	rs2781659	0.05 †*	G	-	0.89	0.38	0.98	0.87	0.76	0.06	0.88	0.12
18408560	European	<i>CRHR2</i>	rs7793837	0.03 *	T	-	0.78	0.08	0.72	0.03	1.40	0.04	0.92	0.04
21991891	European	<i>GLCCI1</i>	rs37972	8.50E-3 †§	C	+	0.84	0.21	1.15	0.30	0.95	0.76	0.98	0.87
23508266	European	<i>COL22A1</i>	rs6988229	8.51E-6 §	T	+	1.12	0.49	0.89	0.49	1.28	0.07	1.11	0.27

+ ; allele associated with increased BDR (protective allele). – ; allele associated with decreased BDR. PR, MX, and AA represent population-specific results for the Puerto Rican, Mexican and African American groups, respectively.

[∅] Referenced study uses deprecated rsID (rs7422339) in place of rs1047891; the most up-to-date rsID has been provided in the table.

* *p*-value obtained from candidate gene association study

† *p*-value obtained from family-based association study

§ *p*-value obtained from GWAS

Table E15. Chromosomal location of identified SKAT-O regions by genome build.

Chr	GRCh37		GRCh38	
	Start	Stop	Start	Stop
1	114177000	114178000	113634378	113635378
11	27507000	27508000	27485453	27486453
19	10424000	10425000	10313324	10314324
4	73478000	73479000	72612283	72613283
8	97926000	97927000	96913772	96914772

Table E16. Functional annotations for variants within SKAT-O regions.

Chr	Region, bp (start – stop)	rsID	Position (Chr:Start)	MAF	Nearest gene	Distance: SNP to feature (bp)			Chrom state	Overlap with H3K27ac ChIP-seq peaks in BSMC
						CpG	DHS	TFBS cluster		
1	114177000 – 114178000	rs75791290	1:114177174	22.57	<i>MAGI3</i> (intron 9)	124341	-3259	-3159 (ZNF263)		
		NA	1:114177664	0.10	<i>MAGI3</i> (intron 9)	123851	-3749	-3649 (ZNF263)		
		rs200935852	1:114177765	12.22	<i>MAGI3</i> (intron 9)	123750	-3850	-3750 (ZNF263)		
11	27507000 – 27508000	NA	11:27507125	0.10	<i>LOC105376671</i> (-3k)	-12768	356	320 (CTCF)		
		rs142981556	11:27507149	9.94	<i>LOC105376671</i> (-3k)	-12792	332	296 (CTCF)		
		NA	11:27507540	0.10	<i>LOC105376671</i> (-3k)	-13183	0	0 (CTCF)		
								0 (SMC3)		
								0 (RAD21)		
		NA	11:27507584	0.10	<i>LOC105376671</i> (-4k)	-13227	0	0 (CTCF)		
19	10424000 – 10425000							0 (SMC3)		Yes
								0 (RAD21)		
		rs4922787	11:27507902	55.07	<i>LOC105376671</i> (-4k)	-13545	-87	-34 (CTCF)		
		NA	19:10424127	0.11	<i>FDX1L</i> (intron 3)	2255	1074	904 (CEBPB)		
19	10424000 – 10425000	NA	19:10424816	0.21	<i>FDX1L</i> (intron 3)	1566	385	215 (CEBPB)		Yes
		rs7408715	19:10424911	9.16	<i>FDX1L</i> (intron 3)	1471	290	120 (CEBPB)		
4	73478000 – 73479000	rs147407778	4:73478028	1.04	<i>ADAMTS3</i> (-43k)	-42707	1693	1737 (USF1)		
		rs72658525	4:73478042	0.03	<i>ADAMTS3</i> (-43k)	-42721	1679	1723 (USF1)		
		rs72859951	4:73478051	5.62	<i>ADAMTS3</i> (-43k)	-42730	1670	1714 (USF1)		
		rs566597853	4:73478157	0.17	<i>ADAMTS3</i> (-44k)	-42836	1564	1608 (USF1)		
		rs570801742	4:73478248	0.14	<i>ADAMTS3</i> (-44k)	-42927	1473	1517 (USF1)		
		NA	4:73478269	0.03	<i>ADAMTS3</i> (-44k)	-42948	1452	1496 (USF1)		
		rs72859952	4:73478282	5.62	<i>ADAMTS3</i> (-44k)	-42961	1439	1483 (USF1)		
		NA	4:73478288	0.03	<i>ADAMTS3</i> (-44k)	-42967	1433	1477 (USF1)		
		rs190134966	4:73478293	0.24	<i>ADAMTS3</i> (-44k)	-42972	1428	1472 (USF1)		
		NA	4:73478313	0.10	<i>ADAMTS3</i> (-44k)	-42992	1408	1452 (USF1)		
		rs72859955	4:73478320	5.63	<i>ADAMTS3</i> (-44k)	-42999	1401	1445 (USF1)		
		NA	4:73478334	0.03	<i>ADAMTS3</i> (-44k)	-43013	1387	1431 (USF1)		
		rs536408324	4:73478379	0.14	<i>ADAMTS3</i> (-44k)	-43058	1342	1386 (USF1)		
		rs182735555	4:73478401	0.07	<i>ADAMTS3</i> (-44k)	-43080	1320	1364 (USF1)		
		rs1878756	4:73478415	33.02	<i>ADAMTS3</i> (-44k)	-43094	1306	1350 (USF1)		
		NA	4:73478416	0.03	<i>ADAMTS3</i> (-44k)	-43095	1305	1349 (USF1)		
		rs540309770	4:73478417	0.03	<i>ADAMTS3</i> (-44k)	-43096	1304	1348 (USF1)		
		NA	4:73478435	0.03	<i>ADAMTS3</i> (-44k)	-43114	1286	1330 (USF1)		
		rs72859959	4:73478476	5.62	<i>ADAMTS3</i> (-44k)	-43155	1245	1289 (USF1)		

		NA	4:73478497	0.07	ADAMTS3 (-44k)	-43176	1224	1268 (USF1)	
		NA	4:73478561	0.03	ADAMTS3 (-44k)	-43240	1160	1204 (USF1)	
		rs77125504	4:73478577	6.91	ADAMTS3 (-44k)	-43256	1144	1188 (USF1)	
		rs16848301	4:73478729	5.63	ADAMTS3 (-44k)	-43408	992	1036 (USF1)	
		NA	4:73478778	0.03	ADAMTS3 (-44k)	-43457	943	987 (USF1)	
		NA	4:73478780	0.03	ADAMTS3 (-44k)	-43459	941	985 (USF1)	
		rs72859963	4:73478794	5.62	ADAMTS3 (-44k)	-43473	927	971 (USF1)	
		rs566651030	4:73478825	0.17	ADAMTS3 (-44k)	-43504	896	940 (USF1)	
		rs183467344	4:73478857	0.31	ADAMTS3 (-44k)	-43536	864	908 (USF1)	
		NA	4:73478858	0.03	ADAMTS3 (-44k)	-43537	863	907 (USF1)	
		NA	4:73478864	0.03	ADAMTS3 (-44k)	-43543	857	901 (USF1)	
		rs752852071	4:73478894	0.03	ADAMTS3 (-44k)	-43573	827	871 (USF1)	
		NA	4:73478902	0.03	ADAMTS3 (-44k)	-43581	819	863 (USF1)	
		rs72859965	4:73478921	5.63	ADAMTS3 (-44k)	-43600	800	844 (USF1)	
8	97926000 – 97927000	NA	8:97926031	0.07	CPQ (intron 3)	-268088	-341	-319 (FOXA1)	
		NA	8:97926255	0.03	CPQ (intron 3)	-268312	-565	-543 (FOXA1)	
		rs146400763	8:97926336	0.07	CPQ (intron 3)	-268393	605	595 (MAFK)	
		rs28610125	8:97926352	3.71	CPQ (intron 3)	-268409	589	579 (MAFK)	
		rs150046254	8:97926412	0.35	CPQ (intron 3)	-268469	529	519 (MAFK)	
		NA	8:97926430	0.03	CPQ (intron 3)	-268487	511	501 (MAFK)	
		NA	8:97926475	0.03	CPQ (intron 3)	-268532	466	456 (MAFK)	
		NA	8:97926488	0.03	CPQ (intron 3)	-268545	453	443 (MAFK)	
		rs537555951	8:97926707	0.03	CPQ (intron 3)	-268764	234	224 (MAFK)	
		rs189625312	8:97926795	0.21	CPQ (intron 3)	-268852	146	136 (MAFK)	
		rs73279841	8:97926796	5.41	CPQ (intron 3)	-268853	145	135 (MAFK)	
		rs182022255	8:97926843	0.94	CPQ (intron 3)	-268900	98	88 (MAFK)	
		rs367621180	8:97926856	0.07	CPQ (intron 3)	-268913	85	75 (MAFK)	
		NA	8:97926915	0.03	CPQ (intron 3)	-268972	26	16 (MAFK)	
		rs186209034	8:97926916	1.11	CPQ (intron 3)	-268973	25	15 (MAFK)	
		rs568121607	8:97926934	0.03	CPQ (intron 3)	-268991	7	0 (MAFK)	

Nearest gene and functional annotations were based on ENCODE annotations from the UCSC Table Browser (Feb.2009 [GRCh37/hg19] assembly). Negative distances indicate upstream features. SNPs within a gene (distance of 0 base pairs) are labeled with the intron or exon number. **CpG**: CpG islands. **DHS**: DNase I Sensitive Sites. **TFBS clusters**: Transcription Factor Binding Site regions aggregated across 91 cell types by ENCODE. For variants overlapping multiple TFBS, the sites are listed in order of descending ChIP-seq signal. Corresponding transcription factors indicated in parentheses. **Chrom State**: chromatin state. Common chromatin state information determined by the BROAD Institute using a hidden Markov model (HMM) to integrate ChIP-seq for nine canonical histone markers across nine tissue types, including lung fibroblasts.  = Heterochromatin  = Insulator chromatin  = Weak transcription markers. **BSMC**: Basal Smooth Muscle cells. H3K27ac ChIP-seq peaks determined experimentally within current study; H3K27ac overlapping was performed by intersecting the region start and stop with the H3K27ac ChIP-seq peaks; Yes = overlap, otherwise no overlap.

Table E17. Results of SKAT-O analyses on regions reported in **Table 3** after excluding all common or rare variants.

Chr	Start	Stop	Population	<i>p</i>		
				Original	Exclude common	Exclude rare
1	114177000	114178000	MX	4.40E-09	1.41E-01	6.85E-05
11	27507000	27508000	MX	6.59E-09	9.09E-02	3.16E-05
19	10424000	10425000	AF	3.12E-11	3.10E-01	2.33E-05
4	73478000	73479000	Combined	6.25E-08	2.13E-02	3.26E-06
8	97926000	97927000	Combined	1.32E-07	7.42E-05	3.90E-04

Exclude common (rare): *p* values obtained from SKAT-O analyses after excluding all common (or rare) variants. **Original:** the original SKAT-O *p*-value reported in **Table 3**.

Table E18. Results of the reduced SKAT-O analysis on rare variants.

Chr	Start	Stop	Pop	RareVar	<i>P</i>	
					Original Model	Reduced Model
1	114177000	114178000	MX	None	4.40E-09	6.85E-05
1	114177000	114178000	MX	1:114177664:A:G*	4.40E-09	4.40E-09
11	27507000	27508000	MX	None	6.59E-09	3.16E-05
11	27507000	27508000	MX	11:27507125:T:C	6.59E-09	1.21E-04
11	27507000	27508000	MX	11:27507540:T:A	6.59E-09	1.66E-05
11	27507000	27508000	MX	11:27507584:C:T	6.59E-09	1.66E-05
19	10424000	10425000	AF	None	3.12E-11	2.33E-05
19	10424000	10425000	AF	19:10424127:A:C	3.12E-11	3.18E-05
19	10424000	10425000	AF	19:10424816:C:T*	3.12E-11	6.35E-13
4	73478000	73479000	Combined	None	6.25E-08	3.26E-06
4	73478000	73479000	Combined	4:73478042:G:A	6.25E-08	3.18E-06
4	73478000	73479000	Combined	4:73478157:G:T	6.25E-08	1.94E-06
4	73478000	73479000	Combined	4:73478248:T:G	6.25E-08	3.25E-06
4	73478000	73479000	Combined	4:73478269:T:C	6.25E-08	3.36E-06
4	73478000	73479000	Combined	4:73478288:A:G	6.25E-08	3.41E-06
4	73478000	73479000	Combined	4:73478293:A:C	6.25E-08	1.14E-06
4	73478000	73479000	Combined	4:73478313:T:A	6.25E-08	1.96E-06
4	73478000	73479000	Combined	4:73478334:A:G	6.25E-08	2.21E-06
4	73478000	73479000	Combined	4:73478379:A:G	6.25E-08	2.09E-06
4	73478000	73479000	Combined	4:73478401:A:G	6.25E-08	3.40E-06
4	73478000	73479000	Combined	4:73478416:G:A	6.25E-08	2.23E-06
4	73478000	73479000	Combined	4:73478417:C:T	6.25E-08	2.16E-06
4	73478000	73479000	Combined	4:73478435:A:T	6.25E-08	2.34E-06
4	73478000	73479000	Combined	4:73478497:G:A	6.25E-08	2.04E-06
4	73478000	73479000	Combined	4:73478561:T:C	6.25E-08	3.38E-06
4	73478000	73479000	Combined	4:73478778:A:G	6.25E-08	2.15E-06
4	73478000	73479000	Combined	4:73478780:C:G	6.25E-08	2.28E-06
4	73478000	73479000	Combined	4:73478825:T:A	6.25E-08	1.99E-06
4	73478000	73479000	Combined	4:73478857:C:T	6.25E-08	3.31E-06
4	73478000	73479000	Combined	4:73478858:A:G	6.25E-08	3.37E-06
4	73478000	73479000	Combined	4:73478864:G:A	6.25E-08	2.20E-06
4	73478000	73479000	Combined	4:73478894:C:T	6.25E-08	3.40E-06
4	73478000	73479000	Combined	4:73478902:T:C	6.25E-08	2.20E-06
8	97926000	97927000	Combined	None	1.32E-07	3.90E-04
8	97926000	97927000	Combined	8:97926031:G:A	1.32E-07	7.49E-04
8	97926000	97927000	Combined	8:97926255:A:T	1.32E-07	6.19E-04
8	97926000	97927000	Combined	8:97926336:A:G	1.32E-07	2.20E-04
8	97926000	97927000	Combined	8:97926412:T:G*	1.32E-07	2.67E-05
8	97926000	97927000	Combined	8:97926430:A:C	1.32E-07	2.05E-04
8	97926000	97927000	Combined	8:97926475:A:C	1.32E-07	4.41E-04
8	97926000	97927000	Combined	8:97926488:G:A	1.32E-07	6.28E-04
8	97926000	97927000	Combined	8:97926707:A:G	1.32E-07	7.66E-04
8	97926000	97927000	Combined	8:97926795:C:T	1.32E-07	3.04E-04
8	97926000	97927000	Combined	8:97926843:G:T*	1.32E-07	1.22E-07
8	97926000	97927000	Combined	8:97926856:G:C	1.32E-07	2.67E-04
8	97926000	97927000	Combined	8:97926915:G:A	1.32E-07	2.66E-04
8	97926000	97927000	Combined	8:97926934:T:G	1.32E-07	6.43E-04

Each row is the result of one SKAT-O analysis containing all common variants. No more than one rare variant was included in each analysis. **RareVar**: the SNP ID of the rare variant included is shown in the format of hg19 chr:pos:ref:alt, where chr: chromosome; pos: chromosome position; ref: reference allele based on hg19; alt: alternative allele. *rare variants that contribute to ≥ 10 -fold change in the *p*-value compared to the analysis with common variant only. **Original Model**: SKAT-O *p*-value from **Table 3**. **Reduced Model**: *p*-value from the current one-rare-variant-per-window analysis.

Table E19. Results of the “drop-one” SKAT-O analysis on common variants.

Chr	Start	Stop	nCommon	Pop	CommonVar excluded	<i>p</i>	
						Original Model	Reduced Model
1	114177000	114178000	2	MX	None	4.40E-09	4.40E-09
1	114177000	114178000	1	MX	1:114177765:T:G*	4.40E-09	4.91E-02
1	114177000	114178000	0	MX	1:114177174:C:T	4.40E-09	1.41E-01
11	27507000	27508000	2	MX	None	6.59E-09	6.59E-09
11	27507000	27508000	1	MX	11:27507902:G:T	6.59E-09	6.90E-09
11	27507000	27508000	0	MX	11:27507149:C:A*	6.59E-09	9.09E-02
19	10424000	10425000	1	AF	None	3.12E-11	3.12E-11
19	10424000	10425000	0	AF	19:10424911:C:T*	3.12E-11	3.10E-01
4	73478000	73479000	10	Combined	None	6.25E-08	6.25E-08
4	73478000	73479000	9	Combined	4:73478921:G:A	6.25E-08	6.24E-08
4	73478000	73479000	8	Combined	4:73478794:A:G	6.25E-08	6.25E-08
4	73478000	73479000	7	Combined	4:73478729:C:T	6.25E-08	6.23E-08
4	73478000	73479000	6	Combined	4:73478577:A:G	6.25E-08	6.46E-08
4	73478000	73479000	5	Combined	4:73478476:C:T	6.25E-08	6.59E-08
4	73478000	73479000	4	Combined	4:73478415:C:T	6.25E-08	2.94E-07
4	73478000	73479000	3	Combined	4:73478320:G:A	6.25E-08	6.99E-07
4	73478000	73479000	2	Combined	4:73478282:A:G*	6.25E-08	7.08E-06
4	73478000	73479000	1	Combined	4:73478051:T:C*	6.25E-08	2.00E-03
4	73478000	73479000	0	Combined	4:73478028:T:A*	6.25E-08	2.13E-02
8	97926000	97927000	3	Combined	None	1.32E-07	1.32E-07
8	97926000	97927000	2	Combined	8:97926916:G:A*	1.32E-07	4.09E-06
8	97926000	97927000	1	Combined	8:97926796:A:G	1.32E-07	4.04E-06
8	97926000	97927000	0	Combined	8:97926352:T:C*	1.32E-07	7.42E-05

Each row is the result of one SKAT-O analysis with all rare variants. Common variants were excluded one at a time in each analysis. **nCommon**: the number of common variants in the analysis. **CommonVar excluded**: the SNP ID, shown in the format of hg19 chr:pos:ref:alt, of the common variant excluded from the nCommon+1 analysis. chr: chromosome; pos: chromosome position; ref: reference allele based on hg19; alt: alternative allele. *common variants that contribute to ≥ 10 fold change in the *p*-value after it is excluded. **Original Model**: SKAT-O *p*-value from **Table 3**. **Reduced Model**: *p*-value from the current drop-one-common-variant-per-window analysis.

Table E20. Correlation between baseline lung function, top ten principal components, and other covariates in discovery association analyses (N =1,441).

	Sex	Pre-FEV ₁ % predicted	Age	EUR	AFR	NAM	BMI	tIgE	Atopy	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Pre-FEV ₁ % predicted	0.00																	
Age	0.15***	-0.04																
EUR	-0.02	0.24***	-0.06*															
AFR	0.04	0.05	0.10***	-0.33***														
NAM	-0.03	-0.04	-0.12***	0.16***	-0.91***													
BMI	-0.01	-0.09**	0.05	-0.10***	-0.05	0.07**												
tIgE	-0.10***	-0.19***	0.02	0.10***	0.02	-0.04	0.03											
Atopy	-0.08**	-0.15***	0.03	0.04	0.07*	-0.07*	0.05	0.83***										
PC1	0.04	0.06*	0.11***	-0.27***	0.98***	-0.95***	-0.06	0.03	0.07*									
PC2	0.00	0.27***	-0.01	0.86***	0.13***	-0.30***	-0.13***	0.11***	0.07*	0.20***								
PC3	-0.05	-0.20***	0.05	0.00	-0.09**	-0.07*	0.05	-0.07*	-0.05	-0.03	-0.03							
PC4	0.05	-0.02	-0.02	-0.01	0.02	0.00	-0.02	0.04	0.02	0.01	0.00	-0.04						
PC5	0.03	0.05	0.00	0.04	-0.02	0.02	-0.07*	0.03	0.02	-0.01	0.03	0.00	-0.02					
PC6	-0.01	0.01	-0.05	0.02	-0.02	-0.00	0.01	-0.01	0.00	-0.01	0.02	0.00	-0.02	-0.02				
PC7	-0.01	-0.05	0.06*	0.04	-0.03	-0.05	0.00	0.06*	0.07*	-0.01	0.03	0.10***	-0.01	-0.00	0.00			
PC8	-0.01	-0.03	0.00	-0.01	0.05	-0.03	0.00	0.00	-0.02	0.04	0.01	-0.02	0.00	0.01	0.04	0.01		
PC9	0.01	-0.03	-0.01	0.00	-0.01	-0.01	-0.02	0.00	0.02	0.00	0.00	0.03	-0.02	0.05	0.02	0.02	-0.01	
PC10	-0.04	0.02	0.00	0.00	-0.00	0.00	-0.02	-0.02	-0.04	-0.01	0.00	-0.02	0.01	-0.04	-0.03	-0.02	-0.02	0.04

Each cell presents a Spearman's correlation coefficient; significance of correlations: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **PC**: principal component. Other variables are defined as in **Table 1**. **EUR**, **AFR** and **NAM**: global European, African and Native American ancestries, respectively. **Pre-FEV₁ % predicted**: percentage of measured FEV₁ relative to predicted FEV₁ estimated by the Hankinson lung function prediction equations. **tIgE**: measure of total Immunoglobulin E from serum in milliliters. **Atopy**: tIgE measurement greater than or equal to 100 kU/L.

Table E21. Number of African American samples selected for RNA-seq based on BDR status and number of copies of low BDR-associated allele.

BDR status	No. of copies of low BDR-associated allele	No. of samples
High	0	8
	1	8
	2	1
Low	0	10
	1	9
	2	3
Total		39