



Published in final edited form as:

Circ Res. 2020 June 05; 126(12): 1816–1840. doi:10.1161/CIRCRESAHA.120.315893.

Importance of Genetic Studies of Cardiometabolic Disease in Diverse Populations

Lindsay Fernandez-Rhodes, PHD MSPH^{1,*}, Kristin L Young, PhD MS-CR^{2,*}, Adam G Lilly, BA^{3,4}, Laura M Raffield, PhD⁵, Heather M Highland, PhD², Genevieve L Wojcik, PhD MHS⁶, Cary Agler, PhD^{2,7}, Shelly-Ann M Love, PhD MS², Samson Okello, MBChB, MMed MSc^{8,9,10}, Lauren E Petty, MS^{11,12}, Mariaelisa Graff, PhD², Jennifer E Below, PhD^{11,12}, Kimon Divaris, DfDS PhD^{2,7}, Kari E. North, PhD^{2,13}

¹Department of Biobehavioral Health, College of Health and Human Development, Pennsylvania State University, University Park, PA ²Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC ³Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC ⁴Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC ⁵Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC ⁶Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD ⁷Department of Pediatric and Public Health, Adams School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, NC ⁸Department of Internal Medicine, Mbarara University of Science and Technology, Uganda ⁹University of Virginia, Charlottesville, VA ¹⁰Harvard TH Chan School of Public Health, Boston, MA ¹¹Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Vanderbilt, TN ¹²Department of Genetic Medicine, Vanderbilt University, Vanderbilt, TN ¹³Carolina Center for Genome Sciences, Chapel Hill, NC

Abstract

Genome wide association studies have revolutionized our understanding of the genetic underpinnings of cardiometabolic disease (CMD). Yet, the inadequate representation of individuals of diverse ancestral backgrounds in these studies may undercut their ultimate potential for both public health and precision medicine. The goal of this review is to describe the imperativeness of studying the populations who are most affected by CMD, to the aim of better understanding the genetic underpinnings of the disease. We support this premise by describing the current variation in the global burden of CMD and emphasize the importance of building a globally and ancestrally-representative genetics evidence base for the identification of population-specific variants, fine-mapping, and polygenic risk score estimation. We discuss the important ethical, legal, and social implications of increasing ancestral diversity in genetic studies of CMD and the challenges that arise from the 1) lack of diversity in current reference populations and available analytic samples, and the 2) unequal generation of health-associated genomic data and

Corresponding Author: Kari E. North, PhD, 123 W Franklin St, CB# 8050, Chapel Hill, NC 27516-8050, Kari_north@unc.edu.
*Co-starred first authors.

Disclosures: HMH is a statistical editor for *Circulation Research*. JEB is an editor for *Circulation Research*. We have no other disclosures.

their prediction accuracies. Despite these challenges, we conclude that additional, unprecedented opportunities lie ahead for public health genomics and the realization of precision medicine, provided that the gap in diversity can be systematically addressed. Achieving this goal will require concerted efforts by social, academic, professional and regulatory stakeholders and communities, and these efforts must be based on principles of equity and social justice.

Keywords

diversity; cardiometabolic disease; race/ethnicity; GWAS; genetic studies

Subject Code:

Race and Ethnicity; Genetic; Association Studies; Epidemiology; Cardiovascular Disease; Genetics

Introduction

Genome wide association studies (GWAS) have transformed our understanding of the genetic underpinnings of human health and disease. Generally relying on genotyping microarrays that assess from ~100,000 to 2.5 million genetic variants across the genome, these studies often employ imputation to infer genotypes at untyped loci based on whole genome sequenced reference panels with a larger number of variants.¹ Initial genotyping microarrays and the first available reference panels were designed to assess common variation (i.e. genetic variants with a minor allele frequency, $MAF > 5\%$ in a population) in European populations, and as such our understanding of genetic variation across diverse global populations has been historically limited.² Indeed, as of 2016, most GWAS – including those of Cardiometabolic diseases (CMD) – had been conducted in European descent populations, with only 5% of participants in these studies representing Hispanic/Latino, Pacific Islander, Arab & Middle Eastern, and other Native peoples.³ The newly released GWAS Diversity Monitor tracks participant diversity in the GWAS catalog in real time, and as of this writing, non-Europeans make up between 11-24% of participants in CMD related trait GWAS, with the vast majority of non-European participants being of Asian descent (see Figure 1 for the latest data on diversity in GWAS of CMD).⁴ Although European ancestry GWAS were originally justified as a practical decision to boost power due to their relative homogeneity and comparatively large samples with genotypic data, it is now understood how problematic this lack of diversity in genetic studies is. This is particularly true for CMD, given its unequal burden across global populations. In fact, inadequate representation of individuals of diverse ancestral backgrounds in genomic studies may inadvertently undercut the potential benefits of precision medicine in the near future, and in particular for populations disproportionately impacted by CMD.⁵

The overarching goal of this review is to describe the relevance and, arguably, necessity of studying ancestrally diverse populations to gain a better understanding of the genetic underpinnings of CMD. To achieve this, we begin by summarizing key concepts regarding genetic diversity and then explain the importance of including global populations for CMD

genomics research (see Key Terms Related to the Genetic Epidemiology of CMD in Box 1). We support this premise by describing the global variation of prevalence of CMD and emphasize its pertinence for the creation of a globally-representative genetics evidence base. Subsequently, we review some of the main benefits of increasing diversity in genomic studies, including the identification of CMD genetic variants that are population specific, the importance of fine-mapping, and estimation of widely-generalizable polygenic risk scores (PRS). Although our review is comprehensive with respect to the need for diversity in genetic studies of CMD and the ethical, legal and social implications for CMD research, we do not address strategies for increasing genetic resources and developing the necessary infrastructure to incorporate diversity into future genomics research, which have been described in detail elsewhere.⁶⁻⁹ Indeed, the inclusion of diverse populations in genomics research has already yielded scientific insights for chronic kidney disease and low LDL in African descent populations, as well as T2D in Mexican ancestry groups (See Importance of variants specific to a population section below, and Table 1).¹⁰⁻¹² While our review is framed around CMD, we note that our main points are generalizable to many other complex traits and chronic diseases (e.g. Schizophrenia,¹³⁻¹⁷ Osteoporosis,^{18,19} and Asthma²⁰⁻²²). We conclude that additional, unprecedented opportunities lie ahead for the realization of precision medicine assuming that the issue of diversity can be systematically addressed by concerted efforts from key stakeholders.

How do we Describe Human Genetic Variation?

Human variation or variability refers to the range of all possible values for any phenotype, and can be attributed to genetics, environmental factors, and their interactions. It is now understood that the role of genes, although crucial, is dynamic and modifiable. This perspective contrasts with the misconception that all genetic inheritance is static and entirely deterministic. Genetic variation can take many forms and broadly refers to differences in structure (e.g., chromosomal rearrangements or abnormalities) and composition (e.g., DNA sequence) of the genome between individuals and populations. Genetic variation occurs in both germ cells and somatic cells, but only variation that arises in germ cells can be inherited. Types of common human genetic variation include single nucleotide variants (SNV) or polymorphisms (SNPs), insertions and deletions (indels), substitutions, inversions, and copy number variants (CNVs).²⁹ Genetic variants can be found at different frequencies in a population, ranging from private (the only copy), *de novo* familial (a few copies), rare (MAF< 1.0%), low-frequency (MAF=1% - 5%), and common (MAF>5%). Most variants are hypothesized to be neutral in terms of functional implications.²⁹ While no empirical information exists about the functional impact of the vast majority of the estimated 10 million SNPs in the human genome, prediction algorithms such as PolyPhen-2,³⁰ SIFT,³¹ FATHMM-XF,³² MutationTaster,³³ and Combined annotation Dependent Depletion (CADD)³⁴ have been developed for this purpose and are freely available. It is important to acknowledge that more information exists for common variants, due to the design of genotyping microarrays and the greater statistical power to detect common genetic variant associations. With the decreasing cost of sequencing, the field of genetic epidemiology is increasingly able to identify low frequency, rare and *de novo* familial variants for CMD, as well as aspects of genomic structural variation.

Here, we primarily focus on the role of ancestral genetic differences in CMD. In this review, we consider the genetic origins of differences in population CMD burden and related health parameters, which may in part track with ancestry or with the socially and culturally defined constructs like race or ethnicity (Box 2 for **Key Terms Related to Ancestral Diversity**). To describe the importance of ancestral diversity for the quantification of the influence of genetic factors on CMD, we briefly define additional key terms related to ancestral diversity (Box 2), and acknowledge the lack of gold standard scientific definitions.^{35,36} As noted above, we use ancestry to refer to the continental origin of the ancestors of an individual or population, and to a lesser extent to the population dynamics within each continent that shaped the observed patterns of genetic variation. We note that genetic ancestry is often estimated via comparison of participants' genotypes to continental reference populations, so incomplete representativeness, availability and small sample sizes of these reference populations are important limitations for the field of genetic epidemiology.³⁷ Moreover, discrete labeling of ancestral populations by continent or other means vastly oversimplifies genetic variation.

When describing the burden of disease in this review, we present both country-specific burden estimates in Figure 2 and refer to the Global Burden of Disease (GBD) regions as constructed by the Institute for Health Metrics and Evaluation (IHME). For simplicity in the text, we highlight the burden of CMD using GBD regions that we expect to have some amount of shared ancestral heritage (e.g. Western Europe, East Asia, Sub-Saharan Africa). Then we point out GBD regions that are comprised of countries, which may have more ancestral diversity given their recent demographic histories (e.g. the United States, US, and Canada, Australia or New Zealand). Nonetheless we recognize that grouping human populations in a geographic manner, e.g. by country or region, may inadvertently oversimplify human genetic diversity. Thus, in an effort to further unpack ancestral diversity within a country like the US, which is the primary focus of this review, we also refer to common categorizations for US race/ethnic minorities as proxy groupings of individuals who may have high proportions of non-European ancestry. However, we recognize that in the US many ancestrally diverse populations, such as racial/ethnic minorities or immigrant populations, may prefer other conceptualizations of race/ethnicity than those currently used in the US.³⁸ For example, in the US the term 'Hispanic/Latino' is defined by the Office of Management and Budget as a union of Spanish language use and heritage from Latin America and the Caribbean (only countries with Spanish cultural origins). When self-identified US Hispanic/Latinos were asked to mark their race on the 2010 US Census using five US racial categories, 48.9% of Hispanic/Latinos identified as being of "Some other Race" (30.5%) using written descriptors such as Mexican, Puerto Rican, Latin American, 5.4%, identified as being of "Two or More Races" (including the five US racial categories and "Some other Race"), and another 13.0% choose to not respond to the race question, making the non-response rate for self-identified non-Hispanic/Latinos three times higher than for the total US population.³⁹

Importance of global populations for CMD research.

Ancestral background may influence one's individual CMD risk as well as the population-level differences in CMD burden seen both across and within countries. Disability adjusted

life years (DALYs) are a common epidemiological measure of overall disease burden, as they account for years of life lost due to premature mortality and years of life lost due to disability from a specific condition.⁴⁸ For example, when comparing age-adjusted estimates of DALYs due to ischemic heart disease in 2017, the burden is greatest in a number of countries within the GBD regions of Oceania, Central Asia and Eastern Europe; in general males have a higher age-adjusted burden of CMD than females (Figure 2A-B).⁴⁹

Between 1990-2017 a number of GBD regions showed an intractably high burden of ischemic heart disease as measured by DALYs (e.g. Oceania, to a lesser extent South Asia), whereas others have experienced either steady declines (e.g. Central Europe, North Africa and the Middle East) or experienced intermittent declines during the same time period (e.g. Central Asia and Eastern Europe, Online Figure I).⁴⁹ Similar disparities in CMD burden are seen across time globally with respect to hypertensive heart disease, ischemic stroke, type 2 diabetes (T2D), and chronic kidney disease (CKD) (Online Table I).⁵⁰ For example, hypertensive heart disease has the highest burden in Central Sub-Saharan Africa, followed by Oceania, and the other regions in Africa and the Middle East (Online Figure II).⁴⁹ In contrast, ischemic stroke is most common in Eastern Europe, Oceania, Central and East Asia (Online Figure III).⁴⁹ T2D and CKD are both most common in Oceania, Central Latin America and Mexico, Central and Southern Sub-Saharan Africa (Online Figures IV-V).⁴⁹

Within the US, African Americans have the highest prevalence of hypertension and related conditions such as coronary artery disease (CAD), ischemic stroke, heart failure, and CKD.⁵¹ In fact, hypertension may account for roughly 50% of the disparity in life expectancy between African Americans and European Americans.⁵² Broadly speaking, the prevalence of adult obesity, T2D, and related complications are highest among individuals of Native American, African American, and Hispanic/Latino ancestry, and lowest in those of European and East Asian descent.^{53,54} Even within commonly used US race/ethnic groupings, notable differences in disease susceptibility and incidence exist. For example, Puerto Rican and Mexican American adults have a greater burden of cardiovascular disease (CVD) risk factors like obesity, than South Americans,⁵⁵ and Indian and Filipino Americans have a higher burden of obesity than Chinese Americans.⁵⁶ Asian Indian and Filipino Americans have the highest prevalence of diagnosed T2D among Asians (13% and 10%, respectively), and Mexican Americans and Puerto Ricans have the highest prevalence of diagnosed T2D than any other Hispanic/Latino backgrounds (14% and 12%, respectively).⁵⁴ While lifestyle, cultural norms, healthcare access, psychosocial and socioeconomic stressors are undeniably important contributors to the disproportionate disease burden across ancestrally diverse populations, some of these health disparities persist even after accounting for differences in social and environmental exposures for diseases.⁵⁷⁻⁵⁹ This observation further suggests that some of the susceptibility to CMD-related traits or diseases may be influenced by genetic factors, which may be ancestry-specific or have complex interactions with environmental factors that are patterned across racial/ethnic groups.⁶⁰

In light of the differing burden of CMD both globally and within countries with diverse populations, efforts to broaden the diversity of populations studied in genetic research have become an imperative for advancing clinical research and public health. The more inclusive genomic studies are, the more effective they will be at expanding the scope of known human

genomic variation and bolstering our understanding of disease etiology to be able to improve population health both globally and locally.

Importance of diverse studies for the evaluation of differential allele frequencies.

Human variation is the result of non-genetic and genetic forces, e.g. natural selection and genetic drift. For example, variation seen in current human populations has been heavily influenced by the Out-of-Africa migration of anatomically modern humans. The movement of relatively small groups out of the African continent over time is an example of a population bottleneck, where the groups that moved into Europe and Asia contained only a subset of the genetic diversity present in the entire continent of Africa.²⁷ Efforts to characterize global human genetic variation (e.g., the 1000 Genomes Project,⁶¹ H3 Africa⁶²), have demonstrated differences in allele frequencies between populations with differing continental ancestries.^{61,63,64} These differences vary based on the evolutionary age of the derived variant and the demographic history of the population. Historically, population allele frequency differences were attributed to genetic processes such as natural selection; however, we now have evidence that widespread allele frequency differences were created by the out-of-Africa bottleneck.⁶⁵ Indeed, the majority of all genomic variants are rare and display allele frequency differences, including those variants that are private to a single continental population (or are population specific).^{61,66} These differences have the potential to inform future treatments and recommendations, for example the discovery of *PCSK9* loss-of-function variants in African Americans that led to the development of new therapies to treat high LDL, as well as others (see Importance of variants specific to a population below). Medical genomics is increasingly conducting whole exome and genome sequencing of patients to identify disease susceptibility variants. However, identifying disease-relevant sequence variants has been difficult, partly due to lack of consensus on variant annotation. One factor that influences variant annotation is allele frequency estimates. Because of this, a priority for investigators has been the development of standard approaches for sharing genomic and phenotypic data provided by clinicians, researchers, and patients through centralized databases, such as ClinVar⁶⁷ and the University of Chicago's Geography of Genome Variation browser.⁶⁸ For example, the Clinical Genome Resource (ClinGen) Variant Curation Interface is a publicly-available curated resource for clinicians and researchers that pulls frequency data from numerous sequencing efforts, including gnomAD⁶⁹ (<https://gnomad.broadinstitute.org/>), PAGE⁵, 1000 Genomes Project (<https://www.internationalgenome.org/>), and the Exome Sequencing Project⁷⁰ (ESP; <https://evs.gs.washington.edu/EVS/>). These frequencies are also available on the National Center for Biotechnology Information (NCBI) dbSNP database⁷¹ ([ncbi.nlm.nih.gov/snp/](https://www.ncbi.nlm.nih.gov/snp/)), with the addition of the Vietnamese Genetic Variation Database⁷², Northern Sweden, the Avon Longitudinal Study of Parents and Children⁷³ (<https://www.ncbi.nlm.nih.gov/bioproject/PRJEB7217>), the UK10K Study⁷⁴ (<https://www.ncbi.nlm.nih.gov/bioproject/PRJEB7218>), and Trans-Omics for Precision Medicine (TOPMed)⁷⁵ (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA400167>). There are also many efforts conducted by industry to increase diverse representation, such as Regeneron's DRIFT Consortium⁷⁶ ([*Circ Res.* Author manuscript; available in PMC 2021 June 05.](https://www.regeneron.com/sites/all/themes/regeneron_corporate/files/science/DRIFT-Consortium-</p></div><div data-bbox=)

Factsheet-Backgrounder-July-FINAL.pdf) and 23andMe's Populations Collaborations Program for genotype data⁷⁷ (<https://research.23andme.com/populations-collaborations/>). However, these last two data sources are not currently publicly available and therefore not useful to the research and clinical communities for the adjudication of risk variants based on population frequencies.

The difficulty in determining the pathogenicity of a rare variant is compounded by clinical laboratories' labeling putatively deleterious non-synonymous calls as variants of unknown significance, a phenomenon that occurs at higher rates for individuals of non-European descent, especially since these variants have been studied and characterized less frequently.⁷⁸ Alternatively, diverse genetic data have improved clinical knowledge by leading to a reclassification of putatively causal pathogenic variants for hypertrophic cardiomyopathy, which were later determined to be benign as a result of being over-represented in African Americans.⁷⁹ For-profit companies are also venturing into the business of variant reclassification. For example, Blueprint Genetics offers a variant classification service⁸⁰ that allows the sequencing data from a previous exome to be re-analyzed, in its entirety, to search for new clinically relevant variants that may explain or contribute to a patient's diagnosis.

Differences in allele frequencies across global populations have also driven a SNP ascertainment bias in genotype array data.^{81,82} Genotype arrays, especially older ones (e.g. Affymetrix 5.0, Illumina Goldengate), were developed based on European ancestry sequence data,⁸³ a feature that has contributed to the observed biased range of allele frequencies in GWAS of non-European populations. This is becoming a major stumbling block, as results from GWAS are combined to generate PRS (Polygenic Risk Scores, sometimes also called genetic risk scores), which are now being used to generate personalized CMD risk estimates in both clinical prognosis and personalized intervention/treatment plans.⁸⁴ Developing a successful PRS depends on maximizing the proportion of the total variance for a particular phenotype that is explained by a set of identified genetic variants. In research, it has become a widely-accepted practice to incorporate all measured variants (many of which are correlated) when calculating the proportion of variance explained, as this tends to improve prediction accuracy in complex traits.^{84,85} Such PRS are being widely applied, but SNP ascertainment bias can lead to a model with vastly different risk estimates across ancestries and poor prediction accuracy. Moreover, our recent work has shown that a PRS ascertained using standard methods in one population can yield unpredictable biases in the distributions of scores in other groups, with patterns fluctuating dramatically across traits.⁸⁶ This also suggests that many causal variants, particularly in non-European ancestries, remain undiscovered. The only equitable use of PRS is one that ensures that scores can be calculated accurately for everyone, meaning the genomic data used must be fully-representative of all human genetic variation. Any genetically-informed personalized medicine approach that fails to take this into account is at risk of misinterpreting the underlying data.

Importance of variants specific to a population.

Due to the historical over-representation of European ancestry individuals in current GWAS, to date we have just begun to identify associations that are uncommon in populations of

European ancestry but are common in others, due to either non-genetic or genetic factors that shift allele frequencies and differences in linkage disequilibrium (LD) patterns between populations. Below, we highlight a non-exhaustive list of examples of genetic variant associations for ischemic and hypertensive heart disease, stroke, T2D and CKD and summarize them in Table 1. We then describe the generalizability of mostly European ancestry discovered variants across ancestrally diverse populations. Although there is a paucity of genomic studies in ancestrally diverse populations, several notable large genomic studies and consortia have been assembled to focus on advancing the state of genetic research in these groups. The examples below are not meant to serve as a comprehensive accounting of all ancestrally diverse genomic studies of CMD and related traits, but rather showcase the breadth of discovery that is possible in such diverse studies.

Ischemic and Hypertensive Heart Disease and Stroke.

As described above, the burden of DALYs due to ischemic or hypertensive heart disease is greatest in Oceania, and concerningly high in Central Asia and Eastern European (ischemic heart disease) and several regions in Africa (hypertensive heart disease) (Online Figures I-II).⁴⁹ In contrast, ischemic stroke is most common in Eastern Europe, followed by Oceania and Central Asia (Online Figure III).⁴⁹ Between 1990-2017, most global regions have experienced a general decline in both ischemic and hypertensive heart disease, but the specific trajectories vary greatly. Differences in genetic ancestry and variation with respect to plasma lipid levels, hypertension or other CMD-related traits may contribute to these population-level differences. Several examples are highlighted below.

Plasma lipid levels and *PCSK9*, *CD36*, and *APOC3*.

Ancestry-specific variants associated with blood lipid levels were first described with the seminal 2005 sequencing of African ancestry participants of the Dallas Heart Study,¹² describing several loss-of-function *PCSK9* variants (e.g., rs28362286, MAF~1%, and rs67608943, MAF~0.3%) that were associated with approximately 40% lower low-density lipoprotein (LDL) cholesterol levels.¹² Both variants are found at lower frequencies in the African ancestry samples of gnomAD (Table 1). At the time, *PCSK9* had been identified as an autosomal dominant hypercholesterolemia gene (gain-of-function mutations), but the observations for *PCSK9* loss-of-function variants, with resulting large decreases in CAD risk, helped support the successful development of *PCSK9* inhibitors.⁸⁷

Other plasma lipid examples include a loss-of-function variant in *CD36* (rs3211938, Table 1) that is common only in African ancestry populations (MAF=9%) and under selective pressure due to malaria.⁸⁸ The variant is associated with higher high-density lipoprotein (HDL)⁸⁹⁻⁹¹ and lower triglycerides, as well as platelet traits,⁹² red cell distribution width,⁹³ C-reactive protein,⁹⁰ and other measures relevant to CMD.

Additionally, carriers of a triglyceride lowering null mutation (rs10892151, MAF ~5%) in *APOC3* are common in the Lancaster Old Order Amish, which has allowed for the identification of *APOC3* loss-of-function as cardioprotective.⁹⁴ This observation has since been confirmed in large meta-analyses for rs76353203 (Table 1).^{95,96}

GWAS of Blood Pressure Traits.

Recent GWAS of a pooled, ancestrally-diverse sample conducted as part of the Population Architecture using Genomics and Epidemiology (PAGE) study have analyzed the genetic etiology of systolic blood pressure (SBP), diastolic blood pressure, and 24 other complex traits.⁵ They discovered a novel variant at *GYPC* for SBP (rs28515082, Table 1) that was most common in their sample in Native and Hispanic/Latino Americans (MAF=10-13%), and is particularly rare in East and South Asian populations (MAF<0.5%). Although this variant is common in European descent groups (MAF=16%), it was first identified in association with blood pressure in the context of a diverse genetic sample. They also described a secondary signal for SBP at *GPR20* (rs111409240, Table 1), which is independent of the previously identified European signal⁹⁷ (rs34591516). The variant leading this novel secondary signal is common in African Americans (MAF=20%), low frequency in the other diverse populations analyzed (MAF<6%), and rare in European descent individuals (MAF<1%). Other large trans-ethnic electronic health record analyses and meta-analyses of ancestrally diverse samples have also highlighted the added value of diversity in studies of blood pressure variation.^{98,99}

JRKL and Hypertension.

Also in the PAGE study, an indel (rs145054295) in *JRKL* was associated for the first time with hypertension ($\beta=-0.43$, $P=3.70\times 10^{-9}$) and effect sizes at this site were shown to be different across ancestries ($P=0.025$).⁵ In the PAGE study sample the variant was monomorphic in European populations;⁶¹ yet, the minor allele was found at 2.4% frequency in African Americans ($\beta=-0.36$, $P=1.96\times 10^{-5}$), 0.4% in Hispanic/Latinos ($\beta=-0.52$, $P=5.08\times 10^{-3}$), and 0.5% in Native Americans ($\beta=-1.82$, $P=0.058$). The variant is at even lower frequencies among primarily East Asians (MAF=0.01%, $\beta=-2.39$, $P=0.32$) and Native Hawaiians (MAF=0.01%, $\beta=14.90$, $P=0.08$) in PAGE. These differential effect sizes by ancestry are likely largely due to allele frequency differences (also reflected in Table 1), further highlighting the importance of studying diverse populations for discovery in trait mapping and at scale, as when CMD-relevant variants are this rare, large sample sizes are required to robustly estimate effect sizes.

PRKCH and Stroke.

A missense variant (rs2230500) in *PRKCH* has been associated with increased risk of ischemic stroke in Japanese¹⁰⁰⁻¹⁰² and Chinese populations.¹⁰² In a meta-analysis of five study populations comprised of Chinese and Japanese individuals (3,686 cases and 4,589 controls), individuals with the GA or AA genotype had approximately 34% greater odds of developing ischemic stroke than those with the GG genotype.¹⁰² The SNP is common in Asian populations (Japanese, MAF=24%; Han Chinese, MAF=18%),¹⁰⁰ but rare in European and African descent populations (MAF 1%, Table 1). *PRKCH* is a member of the protein kinase C (PKC) family and is involved in the development and progression of atherosclerosis in humans.¹⁰⁰ The A allele results in the amino acid substitution Val 374 Ile within the ATP-binding site of PKC-delta, thereby enhancing PKC activity.¹⁰³

Sickle cell trait and Stroke.

Sickle cell trait (i.e. individuals who have only one copy of the sickle cell variant at rs334, which is more common in individuals of African ancestry; Table 1), has been associated in some studies with risk of stroke¹⁰⁴, though this link has been disputed by a recent meta-analysis.¹⁰⁵ More consistent associations have been observed with thrombosis and hemostasis biomarker D-dimer.¹⁰⁶⁻¹⁰⁹ Venous thromboembolism (particularly pulmonary embolism) has also been associated with sickle cell trait;^{110,111} more analysis is needed to understand these associations in larger sample sizes, as well as elucidate the underlying mechanisms.

Type 2 Diabetes.

Unlike the trends observed for heart disease burden globally, DALYs for T2D have generally remained more intractably high between 1990-2017. For example, the burden of T2D is highest for Oceania, where it increased and then plateaued from 1990-2017. High T2D burden is also found in Southern Sub-Saharan Africa, and Central Latin America and Mexico (Online Figure IV).⁴⁹ Genetic variation related to glycemic regulation, obesity, or other CMD-related traits may explain some of these global disparities in risk.

***SLC16A11* and T2D.**

Studies of non-European ancestry populations have also revealed T2D-related genetic variants with population frequency differences. For example, several admixed Mexican ancestry populations carry a T2D risk haplotype with variants in *SLC16A11* with four amino acid substitutions; subsequent lookups of this haplotype revealed a ~50% frequency in Native American and ~10% in East Asian study participants, but rare in participants of European or African descent.¹¹ This haplotype explains roughly 20% of the increased prevalence of T2D in Mexico, and expression of *SLC16A11* in heterologous cells (non-human cells that do not usually express this gene) alters lipid metabolism, causing an increase in intracellular triacylglycerol levels.¹¹² Despite T2D having been well-studied by GWAS previously, this analysis in Mexican ancestry individuals identified *SLC16A11* as a novel finding possibly involved in triacylglycerol metabolism.

T2D and Obesity in Founder Populations.

A nonsense variant in *TBC1D4* was associated with a large increase in T2D risk (Odds Ratio=10.3 in homozygous carriers) as well as a large decrease in glucose uptake in muscle in a founder population of Greenlandic Inuit among 2,575 participants (rs61736969, MAF=17%, Table 1).¹¹³

Another recent GWAS of a Greenlandic Inuit population has also reported a large effect size variant for height and weight in the *FADS* gene cluster (rs7115739), which is highly prevalent in the Inuit,¹¹⁴ likely as an adaptive mechanism to a diet rich in polyunsaturated fatty acids. Subsequent analyses revealed that this variant also influences height in European descent populations. The effect sizes for the weight finding differed between Greenlandic and a previous European ancestry GWAS,¹¹⁵ likely due to its lower frequency in Europeans (MAF<5%, Table 1).

Pancreatic Acinar Function and T2D.

In a meta-analysis of four GWAS of T2D in individuals with Japanese ancestry, three variants were found to be in moderate LD ($r^2 > 0.6$) with previously unreported missense variants [p.Val282Met in *GP2* (rs78193826), p.Ala341Thr in *CPA1* (rs77792157) and p.Arg131Gln in *GLP1R* (rs3765467) – all of which are more common in East Asian versus European populations (Table 1). These variants are in genes that have been previously related to pancreatic acinar cell function (e.g. *CPA1* and *GP2*) and insulin secretion (e.g. *GLP1R*).¹¹⁶ In previous work, another coding variant in *PAX4* (Arg192His, rs2233580), an important transcription factor for islet function, was also more common in East Asian populations (MAF=11%, Table 1) than any other ancestral group and was found to be associated with T2D.¹¹⁷

CREBRF, T2D and Obesity.

A large-effect BMI increasing missense variant in *CREBRF* (rs373863828) was also recently identified, as it is common in Samoan populations (MAF ~25%) and rare in other global populations outside of Oceania (Table 1).^{118,119} The variant has a larger effect size ($\beta = 1.36 \text{ kg/m}^2$) than many common BMI loci, including *FTO* [rs1558902, the largest effect size variant reported in European GWAS¹²⁰ ($\beta = 0.39 \text{ kg/m}^2$)], and is associated with decreased risk of T2D. In contrast to the majority of previous GWAS findings for BMI and T2D, this variant appears to increase BMI while decreasing T2D risk. Functional analyses suggest that the variant can decrease energy use and increase fat storage in adipocyte cell models. Due to its large effect size and common allele frequency in Samoans, this variant was detectable in a discovery and replication sample of ~5,000 individuals, which is much smaller than samples generally required for GWAS.

HbA1c and T2D.

There is also a growing awareness of the relationship between variants of differing allele frequencies across global populations and the accuracy of HbA1c as a measure of long term glycemic control.¹²¹ Recent analyses using assays robust to previously known assay interference effects have found that sickle cell trait (Table 1) can lower HbA1c relative to fasting glucose levels.¹²² These effects may be somewhat assay dependent, however, as in the Diabetes Prevention Program sickle cell trait was shown to be associated with higher HbA1c.¹²³ As shown in Table 1, common (rs1050828, MAF=12%) and rare (rs76723693, MAF=0.5%) missense variants at the *G6PD* locus in African ancestry populations may also influence the accuracy of HbA1c as a test for glycemic control.^{124,125} Similar to sickle cell trait, the geographic distribution of *G6PD* deficiency parallels the distribution of malaria endemic regions,¹²⁶ highlighting the need for further examination of HbA1c accuracy in other populations exposed to endemic malaria, e.g. Southeast Asia. Recent analyses suggest that alpha thalassemia (based on CNV esv2676630 carrier status) may also be associated with higher HbA1c¹²⁷ and may statistically interact with sickle cell trait to influence clinical parameters.¹²⁸ The fact that the recent identification of these HbA1c findings for common genetic variants came decades after the clinical use of HbA1c as a measure of long term glycemic control¹²¹ and an initial round of GWAS analyses for HbA1c¹²⁹ and other traits

illustrates the clinical significance of what has been missed by focusing GWAS on exclusively European ancestry populations.

Chronic Kidney Disease.

DALYs due to CKD dynamically changed from 1990-2017 (Online Figure V).⁴⁹ Similar to the global trends in burden for T2D, the burden of CKD is led by Oceania, followed by Central Latin America and Mexico, and Central Sub-Saharan Africa. Below, we highlight examples of how global variation in CKD burden may reflect differences in genetic ancestry.

APOL1 and CKD.

African Americans are more than twice as likely to develop end-stage renal disease as European Americans; this observation eventually led to the discovery of the G1 (rs73885319) and G2 (rs71785313) risk variants in *APOL1* that are more common among individuals of African ancestry (Table 1),¹⁰ likely due to selective pressure from African trypanosomiasis.¹³⁰ These variants have a large impact on risk in carriers of two risk alleles, with a reported Odds Ratios of 17 for focal segmental glomerulosclerosis, 29 for Human Immunodeficiency Virus-associated nephropathy¹³¹, and at least 15% lifetime risk of CKD in risk allele carriers.¹³² These variants are also associated with faster progression of disease in other African-admixed groups, such as Hispanic/Latinos.¹³³ The association was first found in African Americans because the relevant variants had reached higher allele frequency in that population (rs73885319 MAF = 23%, rs71785313 MAF = 13%), yielding higher power for a given sample size. Given that Hispanic/Latinos are an ancestrally-diverse ethnic group, it is not surprising that the association replicated in some Hispanic/Latino backgrounds (e.g., specifically those with a higher proportion of African ancestry, such as individuals from the Caribbean), but not others.

Sickle cell trait and CKD.

Initial reports of differential susceptibility to CKD for sickle cell trait from small studies¹³⁴ have since been confirmed by larger cohort studies.^{135,111} In fact, sickle cell trait may have a similar effect size for progression to end stage renal disease as the well-known *APOL1* high risk genotypes [Hazard Ratio of 1.8 for *APOL1* versus 2.0 for sickle cell trait in the REasons for Geographic and Racial Differences in Stroke study].¹³⁶

Summary.

The above examples highlight the need for more genetic discovery studies in ancestrally diverse and admixed populations in order to identify novel susceptibility variants that may be rare or absent in GWAS of European descent populations. In addition, there is rising concern that findings in one ancestral group may not have the same effect sizes in other ancestries. The PAGE Study has explored this question for several CMD traits; in a seminal paper they presented an analysis of BMI, T2D, and lipid levels, and compared the direction and magnitude of effects for GWAS-identified variants in multiple non-European ancestry populations against European ancestry findings. They found an overall dilution of effect sizes across ancestries.⁸⁶ The PAGE study further tested for evidence of diminished effect sizes in a diverse sample at previous GWAS findings for 26 traits related to CMD and other

complex diseases.⁵ This experiment demonstrated that previously-reported GWAS findings, derived from predominantly European ancestry samples, have significantly attenuated effect sizes on average in other ancestral groups. For example, the correspondence between previously-reported GWAS effect sizes and effects seen among Hispanic/Latinos was 0.86 (95% confidence interval= 0.83-0.90) and 0.54 (0.50-0.58) among African Americans.⁵ The observed weaker effect sizes in non-European populations may be a function of differential LD, allelic heterogeneity, gene-gene or gene-environment interactions, and can further widen the gap between the impact of known GWAS findings on CMD. Regardless of the origin of the differential effects, caution should be exercised in applying any genetic risk prediction model based on SNP association findings beyond the ancestry group in which they were identified.

Importance of diversity for fine-mapping

After the initial success of GWAS in identifying genomic regions that are robustly associated with a wide array of diseases and related traits, the next major challenges for genetic epidemiology are identifying the underlying causal variants and target genes and translating these findings into clinical insights. Of the thousands of genomic regions associated with complex traits, over 90% are in non-coding, potentially regulatory regions of the genome.¹³⁷ While GWAS are an effective tool for identifying genomic regions associated with a particular phenotype, they are often unable to pinpoint the causal variant or even the implicated gene. Indeed, studies of one single ancestral population, and their LD signature, provide investigators with limited ability or power to identify causal variants.¹³⁸

In many cases, the causal variants underlying GWAS signals are shared across ancestry groups, and GWAS of multiple diverse populations (i.e. transethnic meta-analysis or pooled analysis of multiple ancestral groups) can help narrow the credible set of causal variants at a given locus due to their varied LD structure. For example, African ancestry populations have on average shorter LD blocks than those found in European populations, and this characteristic has been shown to be particularly helpful for narrowing the number of candidate causal variants at a given locus and prioritizing candidate variants for functional follow-up.^{24,139} In many recent analyses of CMD, transethnic fine-mapping has helped narrow lists of candidate variants for kidney function,¹⁴⁰ QT interval,¹⁴¹ lipid traits,¹⁴² and BMI.¹⁴³ Integration of transethnic fine-mapping analyses along with functional annotation can aid in variant selection for follow-up testing of differences in transcriptional activity and protein binding *in vitro*,¹⁴⁴ and can lead to the identification of putative causal variants at a previously described GWAS locus.

Importance of diversity in risk prediction and precision medicine.

As described above, PRS are being routinely used to aggregate effect sizes across the genome to estimate the overall contribution of genetics to the variability observed in a given phenotype. In practice, a PRS is computed for each genotyped individual in a target (testing) sample based on extant GWAS discovery results (training sample). PRSs have been used for both estimating the impact of a set of novel GWAS results in external cohorts and for providing individualized risk prediction.^{84,145} To draw inferences, the distribution of risk

scores is often divided into percentiles or other categorizations to compare the risk of the outcome for an individual, to others in the same sample. However, PRS often suffer from bias as the majority of the training GWAS data are derived from European ancestry populations,¹⁴⁶ leading to unpredictable differences in PRS estimation and model fit across populations due to differences in allele frequencies, effect sizes, or the underlying etiology of the trait.^{145,147} There are many examples where trying to incorporate European-derived PRSs in diverse populations results in poor model fit.^{145,146,148,149} Below we describe examples of such biases in the published literature and from our own work.

CAD PRS in European Ancestry Populations.

Research has demonstrated the potential benefit of including genetic risk scores, in the form of PRS, along with more traditional assessment of risk factors, such as the Framingham Risk Score (FRS), in predicting poor cardiovascular outcomes. One of the earliest PRS, from the Myocardial Infarction-Genes randomized placebo-control clinical trial, demonstrated that for patients receiving both PRS (based on 11 SNPs associated with CAD) and Framingham 10-year CVD risk score (FRS) information had lower LDL cholesterol levels and increased statin use compared to those individuals for whom only FRSs were used.^{150,151} In a more recent clinical trial, the GeneRisk study in Finland showed that providing personalized cardiovascular disease risk information, based on a combination of traditional risk data and PRS, motivated healthy behaviors.¹⁵² Physicians at Massachusetts General Hospital are launching a Preventive Genomics Clinic to help patients understand their monogenic and polygenic risk for a variety of health conditions and take steps to minimize their risk.¹⁵³ The hope is that this clinic will serve as a model for how individuals may be able to get an inexpensive report of monogenic and polygenic risk and incorporate this information into preventive measures. However, if the PRSs are developed in a single population, they will necessarily miss important genetic variants in other populations that contribute to risk.

Obesity PRS.

To understand the differences in PRS performance as a function of population architecture and epidemiology, the PAGE study assessed the performance of a recently published PRS for obesity¹⁵⁴ in the different populations of the PAGE study. Model fit diminished with genetic distance from European populations when correlating the PRS with BMI. The four PAGE populations (Hispanic/Latino, N=19,028; African American, N=16,093; Asian, N=4,155 (88% Japanese, 5% Filipino, 4% Chinese, <1% South Asian); and Native Hawaiian, N=2,502) had an adjusted R^2 ranging from 2.7% (Hispanic/Latino) to 0.3% (Native Hawaiian) (Figure 3A). The Asian group also had low performance with an adjusted R^2 of 1.7%. However, when we look at performance for predicting obesity (BMI ≥ 30 kg/m²), the Asian participants had the best model fit (although still relatively poor), with an area under the curve of 0.587 (Figure 3B). This appears to be due to the underlying distribution of BMI and obesity within the groups. Within PAGE, Asians had the lowest proportion of obesity at 11.4%, compared to 44.2% within African American, 40.4% in Hispanic/Latino and 35.5% in Native Hawaiian participants, reflecting obesity prevalence among these groups in the general US population. (Figure 3C) Therefore, the risk score distribution was better able to differentiate the few Asian individuals with high risk for obesity within the top percentiles. In contrast, even with a higher adjusted R^2 in Hispanic/

Latinos, African Americans, and Native Hawaiians, the PRS could not accurately distinguish these risk strata, because such a large proportion of participants were obese. This exemplifies the intersection of model fit due to heterogeneity in effect sizes (often from differential LD and allele frequencies) and differences in the prevalence of a trait, which both interact to complicate the translation of PRS to other populations.

In summary, the predictive value of PRSs depends both on the relevant characteristics of the target (testing) dataset and the statistical power of the discovery (training) dataset—specifically, the enrichment of the genome-wide distribution of association test statistics attributable to aggregate, additive genetic effects. To date, PRSs have been developed using available GWAS as the training data, which have much larger sample sizes in Europeans than any other population.¹⁵⁵ PRSs developed from these findings are not necessarily transferable to other ancestral populations.¹⁵⁶ In fact, PRS accuracy is a function of recent human demographic history, such that a greater proportion of phenotypic variance is explainable by the PRS in target populations that are genetically more similar to the population studied in the discovery GWAS. As genetic distance between two populations increases, the polygenic predictive value of PRS decreases. A practical question is how to construct polygenic scores for recently admixed individuals or for individuals who are genetically distant from those populations reflected by the largest existing GWAS of CMD. The use of transethnic results to derive appropriate weights for more ancestrally diverse samples may increase prediction accuracy,¹⁵⁷ and MultiPred is a methodologic approach that combines PRSs based on European training data with PRSs based on training data from other target populations.¹⁵⁵ Current methods development is focused on best practices for handling allele frequency and LD differences both within and across populations. Given the limitations in assessing and comparing PRS across populations, great caution is advised in interpreting differences in PRSs across ancestries.

The promise of increasing diversity for future CMD research.

The PAGE Study and Authors of this paper have been involved in various initiatives that aim to promote the characterization of genomic variation in ancestrally diverse populations (e.g. the Hispanic Community Health Study/Study of Latinos¹⁵⁸), to develop statistical methods to analyze ancestrally diverse data^{159,160} and to improve the accuracy of PRS. Other important efforts are also working to bring the promise of precision medicine to ancestrally diverse populations. One such longitudinal effort is the Multi-Ethnic Study of Atherosclerosis (MESA), which was designed to assess subclinical CVD and progression to incident CVD in a diverse population-based sample. At six recruitment centers, MESA recruited 6,814 participants from 2000-2002 across four race/ethnic groups (European (39%), African American (28%), Hispanic/Latino (22%), and Chinese American (12%)).¹⁶¹ Key findings from the study include the predictive power of coronary artery calcification (CAC) across ancestry for coronary events¹⁶², the association of air pollution with CAC progression¹⁶³, and extensive explorations of CVD biomarkers, such as lipoprotein-associated phospholipase A2 (Lp-PLA2)¹⁶⁴ and lipoprotein(a)¹⁶⁵, and optimal CVD risk thresholds for these biomarkers by race/ethnic group.¹⁶⁵ MESA and PAGE have been leaders in collaborative efforts in CVD genetic epidemiology, such as the National Heart Lung and Blood Institutes' Trans-Omics for Precision Medicine program (TOPMed)⁷⁵ and

the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE)¹⁶⁶, and MESA has been a pioneer in the generation of multi-omics data (such as gene expression and methylation) for multi-ethnic populations.¹⁶⁷

Additional efforts are ongoing to recruit new cohorts (<https://www.theruralstudy.org/about/>)¹⁶⁸ and biobank studies¹⁶⁹⁻¹⁷² with better representation of US and global populations. These include All of Us, funded by the US National Institutes of Health, which is building a large-scale biomedical data resource with the goal of reflecting the diversity of the US population.¹⁷³ The Million Veteran's Program is also recruiting a large and ancestrally diverse cohort of US Veterans.¹⁷⁴ The population genetics company Color also recently announced plans to enroll 100,000 volunteers from under-represented groups to better assess the risk of myocardial infarction from low coverage WGS. As described above, it is especially critical that we study African descent populations⁶⁴, as early human migrations out of Africa (both forced and voluntary) took a portion of genetic diversity into Europe, East Asia and eventually into the Americas. Therefore, large genetic studies of African descent populations will likely improve the accuracy of PRSs for all populations, as well as the ability of precision medicine to reach those facing the highest CMD burden.

As the cost of sequencing the human genome continues to trend lower, whole genome sequencing is becoming available on many thousands of individuals worldwide; Table 2 presents a non-exhaustive list of global genome sequencing projects. Importantly, the ancestry biases present in GWAS arrays will no longer be a concern with sequencing data, and considerable effort should be placed on bringing these data together for novel discoveries and public health utility. Many countries (i.e., Australia, China, Dubai, Denmark, Estonia, France, Japan, Qatar, Saudi Arabia, Singapore, Turkey), and continental (H3Africa, Genome Asia 100K) efforts are also working hard to increase diversity of available genome sequence data worldwide. Of course, accomplishing the goals of coordinated discovery and sharing of data will likely take many years and follow the long timeline of the GWAS that preceded them. In addition, it will take the concerted efforts and diligence on the part of researchers and funders to prioritize these resources and ensure that they are used to their fullest extent to ameliorate bias in genomic studies and benefit human health globally.¹⁷⁶

Ethical, Legal and Social implications of genetics research.

In addition to the implications of genetic research more generally, there are specific ethical, legal, and social concerns surrounding research in underrepresented populations (see Brothers and Rothstein for a comprehensive review).¹⁷⁷ These authors explain that the increase in genomics-enriched health information and the potential of personalized approaches to exacerbate health disparities are key issues that require attention. Moreover, it is critical that these discussions and eventual policy decisions prioritize the issues of privacy, potential for discrimination, and changes in physician-patient relationships and liability. Specifically, it is argued that the current availability of genotype-phenotype information, rising costs, and diminished access to health care, as well as information technology, are all possible sources of increasing health disparities.

As described above, studying diverse populations is necessary for scientific understanding and improved equity and inclusion in the field. However, researchers must carefully consider how they define and approach specific populations in order to avoid essentializing race and racism and hindering engagement. One prominent recent example of the essentialization of race in cardiovascular medicine was the US Food and Drug Administration's approval of BiDil to treat heart failure in African-Americans only¹⁷⁸, in the absence of pharmacogenomic variation to support its race-specific marketing.⁴⁴ This decision was met with widespread criticism, as the authoritative position of the US Food and Drug Administration could have given the biological reification of race an appearance of legitimacy, even though there was no biological basis for the race-specific approval.¹⁷⁹ A second example comes from the non-disclosure that individuals with certain *CYP2C19* alleles do not respond equally to Plavix (clopidogrel bisulfate), an inhibitor of platelet aggregation, as they have a reduced hepatic capability for 2C19 to convert it to its active metabolite.¹⁸⁰ The states of Hawaii and New Mexico have filed civil suits against Bristol-Myers in 2014 and 2016, respectively, for wrongfully acquiring profits from sales within their borders. The alleles in question [*CYP2C19**2, rs4244285 (c.681G>A) and *CYP2C19**3, rs4986893 (c.636G>A)] are found at higher prevalence in the populations of these states –specifically East Asians, Native Hawaiians, Pacific Islanders, Native Americans and Hispanic/Latinos – than in European descent populations (Table 1). Moreover, there is rising concern that race-specific drug development, labeling, or marketing, may reinforce race/ethnic health disparities by increasing drug costs and demands for evidence to support efficacy and necessity in under-represented groups.⁴⁴

Some social scientists have also expressed concerns about etiologic research that investigates the genetic origins of disease and how they differ by race or ethnic groupings, as proxies of shared ancestral background. They worry that this research may inadvertently prompt the public to think of racial groups as biologically distinct categories. If genetic differences at least partially impact differential risk for disease, the public may think that this must also be true for other human traits.¹⁸¹ This hypothesis was tested by Phelan and colleagues¹⁸² using a nationally-representative survey where respondents were randomized to one of four conditions, which consisted of reading a different vignette followed by an assessment of beliefs in biologically essential racial differences. The four conditions included 1) the “Backdoor” vignette describing a genetic variant that is more strongly associated with heart attack in African Americans than in European Americans, 2) a vignette describing race as entirely socially constructed 3) a vignette describing racial groups as broadly genetically distinct from one another, 4) and a no-vignette control group. Results showed that endorsement of essential racial differences was greater for individuals assigned to the “Backdoor” vignette than to the “race as a social construction” vignette or to the non-vignette control group. There was also no difference in endorsement of essential racial differences between the groups assigned the “Backdoor” vignette and the “race as essential biological category” vignette.¹⁸² These findings imply that researchers should anticipate misinterpretations of findings and explain clearly the conclusions that can and cannot be drawn from them.

It is becoming increasingly clear that the underlying evidence base of much genetic research insufficiently represents the same populations historically underrepresented in biomedical

research – and that both the quantity and quality of the available information may be issues.¹⁸³ Skepticism about and suspicion around the research enterprise by potential research participants^{184,185} has been offered as one source of the persistence of genomics-associated disparities in data availability. This is a critical issue because the development of precision health care and ‘learning healthcare systems’ require the inclusion of health and genomics information from currently underrepresented populations.¹⁸³ Principles of implementation science are likely well suited in this domain, aiming to improve inclusion and representation of the genomics evidence base in CMD and health care in general.¹⁸⁶

Generally, biomedical literature has focused on the ethical implications of genetic research as it relates to appropriate disclosure and accurate interpretation of primary research results and unintended findings.¹⁸⁷ In fact, there is rising concern that the use of genetic testing for CMD prevention may not result in measurable changes in either increased motivation for lifestyle change, or healthier lifestyles with respect to diet or physical activity.¹⁸⁸ Racial/ethnic minorities already face structural barriers to healthy lifestyles and unhealthy built environments in the US.¹⁸⁹ High PRS estimates for obesity, or a clinician’s perception thereof, have been shown to result in higher-quality patient-provider interactions as compared to patients with more environmentally driven obesity.¹⁹⁰ Increasing use of genetic testing may overemphasize the relative importance of the biological versus social determinants of health,¹⁹¹ which may divert limited resources to provide assistance to those without insurance or fail to address the social or structural determinants of health, which both differentially impact certain US groups.¹⁸⁹ Recent research demonstrates that a common prediction algorithm used to identify patients at ‘high need’ underestimated African American patients’ risk compared to that of European Americans. This was because the algorithm used health care costs as a proxy for health care needs and on average less resources are spent on African Americans.¹⁹² If uptake of CMD genetic testing becomes more common among European Americans, their average healthcare costs will increase and so will the underestimation of need among African American patients.

CMD genetic testing, PRS estimation, and other precision medicine activities using genetic information may have differential implications for racial/ethnic groups or other vulnerable populations who may have experienced discrimination or stigmatization in the past.¹⁸⁹ First, as CMD genetic research is deployed more frequently to predict risk and classify individuals as high risk, to identify those who can benefit from behavior change or therapeutic interventions,¹⁹³ or to find those individuals who may or may not respond to standard therapies,¹⁹⁴ such labels may compound the discrimination and stigmatization already experienced by individuals in underrepresented groups.¹⁸⁹ Second, the portability of the resulting predictions from PRS across ancestries and environmental exposures is a major concern.^{145,147} In part, this has led to direct-to-consumer prediction algorithms for complex conditions being advertised as restricted to one race.¹⁹⁵ Moreover, for ancestrally diverse individuals and populations, common methodologic approaches to estimate ancestry, identify ancestral outliers, and/or stratify samples into relatively homogenous ancestral groups, may conflict with their own self-identities,¹⁹⁶ and appear to endorse the essentialization of racial categories.

Additional legal challenges may arise as the range and scale of genetic testing advances. Currently healthcare maintenance organizations are working to integrate genetic screenings into their routine clinical care and medical record systems by following the American College of Human Genetics list of 59 genes (including several cardiovascular diseases with high penetrance) as actionable incidental findings¹⁹⁷ It stands to reason that routine genetic predictions for complex CMD traits and diseases may soon be added and this could lead to further legal challenges to the Genetic Information Nondiscrimination Act in the US, or similar legislation in other countries,¹⁹⁸ or be used to adjust health insurance premiums or deny coverage of the relatives of genotyped patients.

A number of specific ethical concerns remain around how individual genetic prediction using PRS is applied outside of boundaries of clinical research¹⁹⁹. Some direct-to-consumer genetic testing companies that provide these predictions may offer lower quality genetic testing and counseling than otherwise provided by laboratories used in clinical care. As an example, Genomic Prediction, with a Clinical Laboratory Improvement Amendments-certified laboratory, currently provides PRSs for a number of CMD diseases and traits (e.g. T2D, CAD, myocardial infarction, hypercholesterolemia and hypertension) for embryos prior to implantation.²⁰⁰ The broad use of PRS for embryonic selection for multifactorial traits and diseases like CMD still remains an ethical question for the field,²⁰¹ as it could further reinforce social intolerance of diversity.¹⁹¹

CONCLUSIONS

Understanding how genetic, environmental, and socio-cultural factors interact to influence disparities in CMD risk is imperative for improving individual and public health. There is a critical need for larger, more diverse genetic studies of CMD so that we can minimize the research gap between the global burden of CMD and the samples and findings available in current genetic resources and databases. CMD genetic research has great potential to inform prevention and personalized medicine. However, this potential will only be realized if CMD research findings, including PRS estimation, are derived from samples that reflect the true ancestral diversity of the target population. Similarly, the potential benefits of tailored lifestyle interventions or therapies cannot be fully realized until equitable representation across ancestrally diverse populations is achieved. Such an inclusive approach necessitates international collaboration, such as that exemplified by the GIANT²⁰² (Genetic Investigation of Anthropometric Traits - https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium), MAGIC²⁰³ (Meta-Analyses of Glucose and Insulin-related traits Consortium - <https://www.magicinvestigators.org>), GLGC²⁰⁴ (Global Lipids Genetics Consortium - <http://lipidgenetics.org>), and CHARGE consortia, as well as other efforts to increase diversity in genetic epidemiology studies.

Inadequate representation is a pressing limitation of current CMD genetic research, and it poses an additional challenge related to engagement of individuals from underrepresented populations. Many scholars have outlined several important steps to advancing engagement in genetic research for minority communities in the US and Canada, many of whom face CMD disparities.^{194,205,206} These steps include the need to acknowledge barriers to participation and to incorporate community-based participatory research approaches and

benefits-sharing models to avoid possible injustices around commercialization of data and medical innovations. Moreover, additional resources are needed to address barriers to higher education and to nurture the genetics careers of individuals from underrepresented groups in the field.²⁰⁷⁻²¹¹

The robust identification of genomic regions associated with CMD and other health outcomes has created major opportunities for the improvement of individual and population health. At the same time, it has generated new challenges that may affect the utility of this information and its equitable integration into the broader healthcare system. In this review, we identify the following key challenges: 1) lack of diversity in the genotype-phenotype evidence base for CMD; 2) lack of diversity in genomic reference panels that may be hampering our ability to identify causal variants for CMD; 3) lack of clearly defined designations for the classification of race/ethnicity and ancestry, which complicate analyses of the genetic contributions to disease; 4) the unequal generation of health-associated genomic data and prediction accuracies that may further exacerbate CMD disease disparities. Despite these challenges, unprecedented opportunities lie ahead for reducing the burden of CMD and improving health via the equitable generation and use of genomics data in diverse populations.

As we have documented in this review with respect to CMD, large scale sharing and harmonization of GWAS data is already occurring. However, these large scale initiatives are just beginning and it is our responsibility as researchers and stakeholders to insist on data integration. Related to this, and also of relevance for GWAS studies, is the issue of data harmonization around self-identified race/ethnicity. Although we have argued in this review that race and ethnicity are socially and culturally defined constructs, the continued use of self-identified categories in genetic studies may capture unexplained phenotypic variability. Novel machine learning approaches have been developed to address this problem^{212,213}, for example the HARE framework developed for the MVP.²¹³ HARE provides ancestry clustering that benefits from both self-identified labels and genetically-informed ancestry. Briefly, HARE combines self-identified race/ethnicity and supervised genetic clustering (determined via support vector machines) to improve clustering beyond that derived from either self-identified or genetic ancestry *only*.²¹³ This allows researchers to build predictors for consistent levels of ancestry, allowing for realistic harmonization of labels even within biobanks where self-identification may be limited or inconsistent.

We must make both the (1) increased generation and (2) increased integration of data in diverse populations a public health priority for CMD and other complex diseases. Achieving this goal will require concerted efforts by social, academic, professional and regulatory stakeholders, as well as the communities bearing the highest burden of CMD, and must be based on principles of equity and social justice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding

KLY is funded by NHLBI R21HL14041901. AGL is funded by NICHD training grant T32 HD091058. HMH and S-AML are funded by NHLBI training grant T32 HL129982. HMH is also funded by ADA Grant #1-19-PDF-045. LEP and JEB are funded by R01HL142302. KD is funded by U01DE025046.

Non-Standard Abbreviations and Acronyms:

| | |
|---------------|---|
| BMI | Body mass index |
| CAD | Coronary artery disease |
| CADD | Combined annotation Dependent Depletion |
| CKD | Chronic kidney disease |
| CMD | Cardiometabolic disease |
| CNV | Copy number variant |
| CVD | Cardiovascular disease |
| DALY | Disability adjusted life years |
| FRS | Framingham Risk Score |
| GWAS | Genome-wide association analysis |
| HDL | High density lipoprotein |
| LD | Linkage disequilibrium |
| LDL | Low density lipoprotein |
| MAF | Minor allele frequency |
| MESA | Multi-Ethnic Study of Atherosclerosis |
| PAGE | Population Architecture using Genomics and Epidemiology |
| PKC | Protein kinase C |
| PRS | Polygenic risk score |
| SBP | Systolic blood pressure |
| SNP | Single nucleotide polymorphism |
| SNV | Single nucleotide variant |
| TOPMed | Trans-Omics for Precision Medicine program |
| T2D | Type 2 diabetes |
| US | United States |

References

1. Das S, Abecasis GR, Browning BL. Genotype Imputation from Large Reference Panels. *Annu Rev Genomics Hum Genet.* 2018;19:73–96. [PubMed: 29799802]
2. Bien SA, Wojcik GL, Zubair N, Gignoux CR, Martin AR, Kocarnik JM, Martin LW, Buyske S, Haessler J, Walker RW, et al. Strategies for enriching variant coverage in candidate disease loci on a multiethnic genotyping array. *PLoS ONE.* 2016;11:e0167758. [PubMed: 27973554]
3. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature.* 2016;538:161–164. [PubMed: 27734877]
4. Mills MC, Rahal C. The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat Genet.* 2020;52:242–243. [PubMed: 32139905]
5. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature.* 2019;570:514–518. [PubMed: 31217584]
6. Popejoy AB. Diversity in precision medicine and pharmacogenetics: methodological and conceptual considerations for broadening participation. *Pharmgenomics Pers Med.* 2019;12:257–271. [PubMed: 31686892]
7. Jooma S, Hahn MJ, Hindorff LA, Bonham VL. Defining and achieving health equity in genomic medicine. *Ethn Dis.* 2019;29:173–178. [PubMed: 30906166]
8. Bentley AR, Callier S, Rotimi C. The emergence of genomic research in Africa and new frameworks for equity in biomedical research. *Ethn Dis.* 2019;29:179–186. [PubMed: 30906167]
9. Hindorff LA, Bonham VL, Brody LC, Ginoza MEC, Hutter CM, Manolio TA, Green ED. Prioritizing diversity in human genomics research. *Nat Rev Genet.* 2018;19:175–185. [PubMed: 29151588]
10. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, Bowden DW, Langefeld CD, Oleksyk TK, Uscinski Knob AL, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science.* 2010;329:841–845. [PubMed: 20647424]
11. SIGMA Type 2 Diabetes Consortium, Williams AL, Jacobs SBR, Moreno-Macías H, Huerta-Chagoya A, Churchhouse C, Márquez-Luna C, H García-Ortíz, Gómez-Vázquez MJ, Burt NP, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature.* 2014;506:97–101. [PubMed: 24390345]
12. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet.* 2005;37:161–165. [PubMed: 15654334]
13. Alkelai A, Lupoli S, Greenbaum L, Kohn Y, Kanyas-Sarner K, Ben-Asher E, Lancet D, Macciardi F, Lerer B. DOCK4 and CEACAM21 as novel schizophrenia candidate genes in the Jewish population. *Int J Neuropsychopharmacol.* 2012;15:459–469. [PubMed: 21682944]
14. Lam M, Chen C- Y, Li Z, Martin AR, Bryois J, Ma X, Gaspar H, Ikeda M, Benyamin B, Brown BC, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet.* 2019;51:1670–1678. [PubMed: 31740837]
15. Bigdeli TB, Genovese G, Georgakopoulos P, Meyers JL, Peterson RE, Iyegbe CO, Medeiros H, Valderrama J, Achtyes ED, Kotov R, et al. Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol Psychiatry.* 2019;
16. Legge SE, Pardiñas AF, Helthuis M, Jansen JA, Jollie K, Knapper S, MacCabe JH, Rujescu D, Collier DA, O'Donovan MC, et al. A genome-wide association study in individuals of African ancestry reveals the importance of the Duffy-null genotype in the assessment of clozapine-related neutropenia. *Mol Psychiatry.* 2019;24:328–337. [PubMed: 30647433]
17. Ikeda M, Takahashi A, Kamatani Y, Momozawa Y, Saito T, Kondo K, Shimasaki A, Kawase K, Sakusabe T, Iwayama Y, et al. Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect. *Schizophr Bull.* 2019;45:824–834. [PubMed: 30285260]
18. Guo Y, Tan L- J, Lei S- F, Yang T- L, Chen X- D, Zhang F, Chen Y, Pan F, Yan H, Liu X, et al. Genome-wide association study identifies ALDH7A1 as a novel susceptibility gene for osteoporosis. *PLoS Genet.* 2010;6:e1000806. [PubMed: 20072603]

19. Taylor KC, Evans DS, Edwards DRV, Edwards TL, Sofer T, Li G, Liu Y, Franceschini N, Jackson RD, Giri A, et al. A genome-wide association study meta-analysis of clinical fracture in 10,012 African American women. *Bone Rep.* 2016;5:233–242. [PubMed: 28580392]
20. Yan Q, Brehm J, Pino-Yanes M, Forno E, Lin J, Oh SS, Acosta-Perez E, Laurie CC, Cloutier MM, Raby BA, et al. A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. *Eur Respir J.* 2017;49.
21. Dahlin A, Sordillo JE, Ziniti J, Iribarren C, Lu M, Weiss ST, Tantisira KG, Lu Q, Kan M, Himes BE, et al. Large-scale, multiethnic genome-wide association study identifies novel loci contributing to asthma susceptibility in adults. *J Allergy Clin Immunol.* 2019;143:1633–1635. [PubMed: 30578877]
22. Daya M, Rafaels N, Brunetti TM, Chavan S, Levin AM, Shetty A, Gignoux CR, Boorgula MP, Wojcik G, Campbell M, et al. Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. *Nat Commun.* 2019;10:880. [PubMed: 30787307]
23. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. Global variation in copy number in the human genome. *Nature.* 2006;444:444–454. [PubMed: 17122850]
24. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet.* 2018;19:491–504. [PubMed: 29844615]
25. Chakravarti A Perspectives on Human Variation through the Lens of Diversity and Race. *Cold Spring Harb Perspect Biol.* 2015;7:a023358. [PubMed: 26330522]
26. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet.* 2010;11:499–511. [PubMed: 20517342]
27. Amos W, Hoffman JI. Evidence that two main bottleneck events shaped modern human genetic diversity. *Proc Biol Sci.* 2010;277:131–137. [PubMed: 19812086]
28. What are single nucleotide polymorphisms (SNPs)? - Genetics Home Reference - NIH [Internet]. [cited 2019 Dec 18]; Available from: <https://ghr.nlm.nih.gov/primer/genomicresearch/snp>
29. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009;10:241–251. [PubMed: 19293820]
30. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7:248–249. [PubMed: 20354512]
31. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc.* 2016;11:1–9. [PubMed: 26633127]
32. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics.* 2018;34:511–513. [PubMed: 28968714]
33. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods.* 2014;11:361–362. [PubMed: 24681721]
34. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47:D886–D894. [PubMed: 30371827]
35. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol.* 2002;3:comment2007. [PubMed: 12184798]
36. Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, Mountain JL, Pérez-Stable EJ, Sheppard D, Risch N. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med.* 2003;348:1170–1175. [PubMed: 12646676]
37. Royal CD, Novembre J, Fullerton SM, Goldstein DB, Long JC, Bamshad MJ, Clark AG. Inferring genetic ancestry: opportunities, challenges, and implications. *Am J Hum Genet.* 2010;86:661–673. [PubMed: 20466090]
38. Humes KR, Jones NA, Ramirez RR. Overview of race and Hispanic origin: 2010 2010 Census Briefs. Washington, DC: US Census Bureau 2011;
39. Ríos M, Romero F, Ramírez R. Race reporting among Hispanics: 2010. US Census Bureau, Population Division; 2013.

40. Fujimura JH, Rajagopalan R. Different differences: the use of “genetic ancestry” versus race in biomedical human genetic research. *Soc Stud Sci*. 2011;41:5–30. [PubMed: 21553638]
41. Serre D, Pääbo S. Evidence for gradients of human genetic diversity within and among continents. *Genome Res*. 2004;14:1679–1685. [PubMed: 15342553]
42. Lu Y- F, Goldstein DB, Angrist M, Cavalleri G. Personalized medicine and human genetic diversity. *Cold Spring Harb Perspect Med*. 2014;4:a008581. [PubMed: 25059740]
43. 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. [PubMed: 16862161]
44. Bonham VL, Callier SL, Royal CD. Will Precision Medicine Move Us beyond Race? *N Engl J Med*. 2016;374:2003–2005. [PubMed: 27223144]
45. Baker JL, Rotimi CN, Shriner D. Human ancestry correlates with language and reveals that race is not an objective genomic classifier. *Sci Rep*. 2017;7:1572. [PubMed: 28484253]
46. Williams DR, Priest N, Anderson NB. Understanding associations among race, socioeconomic status, and health: Patterns and prospects. *Health Psychol*. 2016;35:407–411. [PubMed: 27018733]
47. Williams DR, Mohammed SA. Discrimination and racial disparities in health: evidence and needed research. *J Behav Med*. 2009;32:20–47. [PubMed: 19030981]
48. Gold MR, Stevenson D, Fryback DG. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annu Rev Public Health*. 2002;23:115–134. [PubMed: 11910057]
49. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. Seattle, WA: IHME, University of Washington, 2018 [Internet]. [cited 2019 Dec 19]; Available from: <http://vizhub.healthdata.org/gbd-compare>
50. Institute for Health Metrics and Evaluation (IHME). GBD Results Tool. Seattle, WA: IHME, University of Washington, 2015 [Internet]. [cited 2020 Jan 9]; Available from: <http://ghdx.healthdata.org/gbd-results-tool>
51. Musemwa N, Gadegbeku CA. Hypertension in African Americans. *Curr Cardiol Rep*. 2017;19:129. [PubMed: 29081008]
52. Harper S, MacLehose RF, Kaufman JS. Trends in the black-white life expectancy gap among US states, 1990–2009. *Health Aff (Millwood)*. 2014;33:1375–1382. [PubMed: 25092839]
53. Hales CM, Fryar CD, Carroll MD, Freedman DS, Aoki Y, Ogden CL. Differences in Obesity Prevalence by Demographic Characteristics and Urbanization Level Among Adults in the United States, 2013–2016. *JAMA*. 2018;319:2419–2429. [PubMed: 29922829]
54. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017 [Internet]. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2017 [cited 2019 Dec 2]. Available from: <https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf>
55. Daviglius ML, Pirzada A, Talavera GA. Cardiovascular disease risk factors in the Hispanic/Latino population: lessons from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Prog Cardiovasc Dis*. 2014;57:230–236. [PubMed: 25242694]
56. Mui P, Hill SE, Thorpe RJ. Overweight and obesity differences across ethnically diverse subgroups of Asian American men. *Am J Mens Health*. 2018;12:1958–1965. [PubMed: 30095034]
57. Isasi CR, Parrinello CM, Ayala GX, Delamater AM, Perreira KM, Daviglius ML, Elder JP, Marchante AN, Bangdiwala SI, Van Horn L, et al. Sex Differences in Cardiometabolic Risk Factors among Hispanic/Latino Youth. *J Pediatr*. 2016;176:121–127.e1. [PubMed: 27344220]
58. Mitchell UA, Ailshire JA, Crimmins EM. Change in cardiometabolic risk among blacks, whites, and Hispanics: findings from the health and retirement study. *J Gerontol A Biol Sci Med Sci*. 2019;74:240–246. [PubMed: 29452335]
59. Harding S, Silva MJ, Molaodi OR, Enayat ZE, Cassidy A, Karamanos A, Read UM, Cruickshank JK. Longitudinal study of cardiometabolic risk from early adolescence to early adulthood in an ethnically diverse cohort. *BMJ Open*. 2016;6:e013221.
60. Collins FS. What we do and don’t know about “race”, “ethnicity”, genetics and health at the dawn of the genome era. *Nat Genet*. 2004;36:S13–5. [PubMed: 15507997]

61. 1000 Genomes Project Consortium, Auton, 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. [PubMed: 26432245]
62. H3Africa Consortium, Rotimi C, Abayomi A, Abimiku A, Adabayeri VM, Adebamowo C, Adebisi E, Ademola AD, Adeyemo A, Adu D, et al. Enabling the genomic revolution in Africa. *Science*. 2014;344:1346–1348. [PubMed: 24948725]
63. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 2008;319:1100–1104. [PubMed: 18292342]
64. Bentley AR, Callier SL, Rotimi CN. Evaluating the promise of inclusion of African ancestry populations in genomics. *NPJ Genom Med*. 2020;5:5.
65. Hofer T, Ray N, Wegmann D, Excoffier L. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann Hum Genet*. 2009;73:95–108. [PubMed: 19040659]
66. Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, 1000 Genomes Project, Bustamante CD. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci USA*. 2011;108:11983–11988. [PubMed: 21730125]
67. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46:D1062–D1067. [PubMed: 29165669]
68. Marcus JH, Novembre J. Visualizing the geography of genetic variants. *Bioinformatics*. 2017;33:594–595. [PubMed: 27742697]
69. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv*. 2019;
70. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA [Internet]. [cited 2020 Mar 18]; Available from: <http://evs.gs.washington.edu/EVS/>
71. Kitts A, Sherry S. The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation [Internet]. In: McEntyre J, Ostell J, editors. *The NCBI Handbook*. Bethesda (MD): National Center for Biotechnology Information (US); 2002 [cited 2020 Mar 18]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21088/>
72. Le VS, Tran KT, Bui HTP, Le HTT, Nguyen CD, Do DH, Ly HTT, Pham LTD, Dao LTM, Nguyen LT. A Vietnamese human genetic variation database. *Hum Mutat*. 2019;40:1664–1675. [PubMed: 31180159]
73. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: the ‘children of the 90s’--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013;42:111–127. [PubMed: 22507743]
74. UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JRB, Xu C, Futema M, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526:82–90. [PubMed: 26367797]
75. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Gagliano Taliun SA, Corvelo A, Gogarten SM, Min Kang H, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *BioRxiv*. 2019;
76. Shuldiner AR. The DRIFT Consortium: Discovery Research Investigating Founder Population Traits [Internet]. Regeneron Genetics Center. [cited 2020 Mar 18]; Available from: https://www.regeneron.com/sites/all/themes/regeneron_corporate/files/science/DRIFT-Consortium-Factsheet-Backgrounder-July-FINAL.pdf
77. 23andMe Populations Collaborations Program [Internet]. [cited 2020 Mar 18]; Available from: <https://research.23andme.com/populations-collaborations/>
78. Caswell-Jin JL, Gupta T, Hall E, Petrovchich IM, Mills MA, Kingham KE, Koff R, Chun NM, Levonian P, Lebensohn AP, et al. Racial/ethnic differences in multiple-gene sequencing results for hereditary cancer risk. *Genet Med*. 2018;20:234–239. [PubMed: 28749474]

79. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, Margulies DM, Loscalzo J, Kohane IS. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med*. 2016;375:655–665. [PubMed: 27532831]
80. Blueprint Genetics. Whole Exome Sequencing (WES) Variant Re-evaluation and WES Re-analysis Services [Internet]. 2019 [cited 2020 Jan 5]; Available from: <https://blueprintgenetics.com/wes-re-evaluation-re-analysis/>
81. Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J. Genetic disease risks can be misestimated across global populations. *Genome Biol*. 2018;19:179. [PubMed: 30424772]
82. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays*. 2013;35:780–786. [PubMed: 23836388]
83. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65. [PubMed: 23128226]
84. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219–1224. [PubMed: 30104762]
85. Zhang Y, Qi G, Park J-H, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet*. 2018;50:1318–1326. [PubMed: 30104760]
86. Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, Schumacher FR, Peters U, Franceschini N, Ritchie MD, Duggan DJ, Spencer KL, Dumitrescu L, Eaton CB, Thomas F, Young A, Carty C, Heiss G, Le Marchand L, Crawford DC, PAGE Consortium, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol*. 2013;11:e1001661. [PubMed: 24068893]
87. Sabatine MS. PCSK9 inhibitors: clinical evidence and implementation. *Nat Rev Cardiol*. 2019;16:155–165. [PubMed: 30420622]
88. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Vardilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. Positive natural selection in the human lineage. *Science*. 2006;312:1614–1620. [PubMed: 16778047]
89. Elbers CC, Guo Y, Tragante V, van Iperen EPA, Lanktree MB, Castillo BA, Chen F, Yanek LR, Wojczynski MK, Li YR, et al. Gene-centric meta-analysis of lipid traits in African, East Asian and Hispanic populations. *PLoS ONE*. 2012;7:e50198. [PubMed: 23236364]
90. Ellis J, Lange EM, Li J, Dupuis J, Baumert J, Walston JD, Keating BJ, Durda P, Fox ER, Palmer CD, Meng YA, et al. Large multiethnic Candidate Gene Study for C-reactive protein levels: identification of a novel association at CD36 in African Americans. *Hum Genet*. 2014;133:985–995. [PubMed: 24643644]
91. Love-Gregory L, Sherva R, Sun L, Wasson J, Schappe T, Doria A, Rao DC, Hunt SC, Klein S, Neuman RJ, et al. Variants in the CD36 gene associate with the metabolic syndrome and high-density lipoprotein cholesterol. *Hum Mol Genet*. 2008;17:1695–1704. [PubMed: 18305138]
92. Auer PL, Johnsen JM, Johnson AD, Logsdon BA, Lange LA, Nalls MA, Zhang G, Franceschini N, Fox K, Lange EM, et al. Imputation of exome sequence variants into population-based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. *Am J Hum Genet*. 2012;91:794–808. [PubMed: 23103231]
93. Chami N, Chen M-H, Slater AJ, Eicher JD, Evangelou E, Tajuddin SM, Love-Gregory L, Kacprowski T, Schick UM, et al. Exome Genotyping Identifies Pleiotropic Variants Associated with Red Blood Cell Traits. *Am J Hum Genet*. 2016;99:8–21. [PubMed: 27346685]
94. Pollin TI, Damcott CM, Shen H, Ott SH, Shelton J, Horenstein RB, Post W, McLenithan JC, Bielak LF, Peyser PA, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science*. 2008;322:1702–1705. [PubMed: 19074352]
95. Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med*. 2014;371:32–41. [PubMed: 24941082]

96. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute, Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitzel NO, Lange LA, Lu Y, Tang Z, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371:22–31. [PubMed: 24941081]
97. Surendran P, Drenos F, Young R, Warren H, Cook JP, Manning AK, Grarup N, Sim X, Barnes DR, Witkowska K, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet*. 2016;48:1151–1161. [PubMed: 27618447]
98. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok P- Y, Iribarren C, Chakravarti A, Risch N. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet*. 2017;49:54–64. [PubMed: 27841878]
99. Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, Torstenson ES, Kovesdy CP, Sun YV, Wilson OD, et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet*. 2019;51:51–62. [PubMed: 30578418]
100. Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, Matsushita T, Yamazaki K, Ohnishi Y, Saito S, et al. A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. *Nat Genet*. 2007;39:212–217. [PubMed: 17206144]
101. Serizawa M, Nabika T, Ochiai Y, Takahashi K, Yamaguchi S, Makaya M, Kobayashi S, Kato N. Association between PRKCH gene polymorphisms and subcortical silent brain infarction. *Atherosclerosis*. 2008;199:340–345. [PubMed: 18164711]
102. Li J, Luo M, Xu X, Sheng W. Association between 1425G/A SNP in PRKCH and ischemic stroke among Chinese and Japanese populations: a meta-analysis including 3686 cases and 4589 controls. *Neurosci Lett*. 2012;506:55–58. [PubMed: 22044875]
103. Matsuo R, Ago T, Hata J, Kuroda J, Wakisaka Y, Sugimori H, Kitazono T, Kamouchi M, FSR Investigators. Impact of the 1425G/A polymorphism of PRKCH on the recurrence of ischemic stroke: Fukuoka Stroke Registry. *J Stroke Cerebrovasc Dis*. 2014;23:1356–1361. [PubMed: 24534126]
104. Caughey MC, Loehr LR, Key NS, Derebail VK, Gottesman RF, Kshirsagar AV, Grove ML, Heiss G. Sickle cell trait and incident ischemic stroke in the Atherosclerosis Risk in Communities study. *Stroke*. 2014;45:2863–2867. [PubMed: 25139879]
105. Hyacinth HI, Carty CL, Seals SR, Irvin MR, Naik RP, Burke GL, Zakai NA, Wilson JG, Franceschini N, Winkler CA, et al. Association of Sickle Cell Trait With Ischemic Stroke Among African Americans: A Meta-analysis. *JAMA Neurol*. 2018;75:802–807. [PubMed: 29710269]
106. Naik RP, Wilson JG, Ekunwe L, Mwasongwe S, Duan Q, Li Y, Correa A, Reiner AP. Elevated D-dimer levels in African Americans with sickle cell trait. *Blood*. 2016;127:2261–2263. [PubMed: 26968536]
107. Raffield LM, Zakai NA, Duan Q, Laurie C, Smith JD, Irvin MR, Doyle MF, Naik RP, Song C, Manichaikul AW, Liu Y, Durda P, Rotter JI, Jenny NS, Rich SS, Wilson JG, Johnson AD, Correa A, Li Y, Nickerson DA, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Hematology & Hemostasis TOPMed Working Group*. D-Dimer in African Americans: Whole Genome Sequence Analysis and Relationship to Cardiovascular Disease Risk in the Jackson Heart Study. *Arterioscler Thromb Vasc Biol*. 2017;37:2220–2227. [PubMed: 28912365]
108. Folsom AR, Alonso A, George KM, Roetker NS, Tang W, Cushman M. Prospective study of plasma D-dimer and incident venous thromboembolism: The Atherosclerosis Risk in Communities (ARIC) Study. *Thromb Res*. 2015;136:781–785. [PubMed: 26337932]
109. Amin C, Adam S, Mooberry MJ, Kutlar A, Kutlar F, Esserman D, Brittain JE, Ataga KI, Chang J-Y, Wolberg AS, et al. Coagulation activation in sickle cell trait: an exploratory study. *Br J Haematol*. 2015;171:638–646. [PubMed: 26511074]
110. Noubiap JJ, Temgoua MN, Tankeu R, Tochie JN, Wonkam A, Bigna JJ. Sickle cell disease, sickle trait and the risk for venous thromboembolism: a systematic review and meta-analysis. *Thromb J*. 2018;16:27. [PubMed: 30305805]
111. Folsom AR, Tang W, Roetker NS, Kshirsagar AV, Derebail VK, Lutsey PL, Naik R, Pankow JS, Grove ML, Basu S, et al. Prospective study of sickle cell trait and venous thromboembolism incidence. *J Thromb Haemost*. 2015;13:2–9. [PubMed: 25393788]

112. Rusu V, Hoch E, Mercader JM, Tenen DE, Gymrek M, Hartigan CR, DeRan M, von Grotthuss M, Fontanillas P, Spooner A, et al. Type 2 Diabetes Variants Disrupt Function of SLC16A11 through Two Distinct Mechanisms. *Cell*. 2017;170:199–212.e20. [PubMed: 28666119]
113. Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliusen TS, Andersen MA, Nielsen TS, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature*. 2014;512:190–193. [PubMed: 25043022]
114. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, Korneliusen TS, Gerbault P, Skotte L, Linneberg A, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*. 2015;349:1343–1347. [PubMed: 26383953]
115. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46:1173–1186. [PubMed: 25282103]
116. Suzuki K, Akiyama M, Ishigaki K, Kanai M, Hosoe J, Shojima N, Hozawa A, Kadota A, Kuriki K, Naito M, et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet*. 2019;51:379–386. [PubMed: 30718926]
117. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536:41–47. [PubMed: 27398621]
118. Minster RL, Hawley NL, Su C- T, Sun G, Kershaw EE, Cheng H, Buhule OD, Lin J, Reupena MS, Viali S, et al. A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat Genet*. 2016;48:1049–1054. [PubMed: 27455349]
119. Hanson RL, Safabakhsh S, Curtis JM, Hsueh W- C, Jones LI, Aflague TF, Duenas Sarmiento J, Kumar S, Blackburn NB, et al. Association of CREBRF variants with obesity and diabetes in Pacific Islanders from Guam and Saipan. *Diabetologia*. 2019;62:1647–1652. [PubMed: 31280340]
120. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Mägi R, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010;42:937–948. [PubMed: 20935630]
121. Leong A, Wheeler E. Genetics of HbA1c: a case study in clinical translation. *Curr Opin Genet Dev*. 2018;50:79–85. [PubMed: 29522974]
122. Lacy ME, Wellenius GA, Sumner AE, Correa A, Carnethon MR, Liem RI, Wilson JG, Sacks DB, Jacobs DR, Carson AP, Luo X, Gjelsvik A, Reiner AP, Naik RP, Liu S, Musani SK, Eaton CB, Wu W- C. Association of sickle cell trait with hemoglobin a1c in African Americans. *JAMA*. 2017;317:507–515. [PubMed: 28170479]
123. Hivert M- F, Christophi CA, Jablonski KA, Edelstein SL, Kahn SE, Golden SH, Dagogo-Jack S, Mather KJ, Luchsinger JA, Caballero AE, Barrett-Connor E, Knowler WC, Florez JC, Herman WH. Genetic ancestry markers and difference in a1c between African American and white in the diabetes prevention program. *J Clin Endocrinol Metab*. 2019;104:328–336. [PubMed: 30358859]
124. Wheeler E, Leong A, Liu C- T, Hivert M- F, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, Chu AY, Zhang W, Wang X, Chen P, Maruthur NM, Porneala BC, Sharp SJ, Jia Y, Kabagambe EK, Chang L- C, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med*. 2017;14:e1002383. [PubMed: 28898252]
125. Sarnowski C, Leong A, Raffield LM, Wu P, de Vries PS, DiCorpo D, Guo X, Xu H, Liu Y, Zheng X, Hu Y, Brody JA, Goodarzi MO, Hidalgo BA, Highland HM, Jain D, Liu C- T, Naik RP, O'Connell JR, Perry JA, National Heart, Lung, and Blood Institute TOPMed Consortium. Impact of Rare and Common Genetic Variants on Diabetes Diagnosis by Hemoglobin A1c in Multi-Ancestry Cohorts: The Trans-Omics for Precision Medicine Program. *Am J Hum Genet*. 2019;105:706–718. [PubMed: 31564435]
126. Howes RE, Piel FB, Patil AP, Nyangiri OA, Gething PW, Dewi M, Hogg MM, Battle KE, Padilla CD, Baird JK, Hay SI. G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. *PLoS Med*. 2012;9:e1001339. [PubMed: 23152723]
127. Jun G, Sedlazeck FJ, Chen H, Yu B, Qi Q, Krasheninina O, Carroll A, Liu X, Mansfield A, Zarate S, Metcalf G, Muzny D, Lindstrom S, Selvin E, Kaplan R, Salerno WJ, Gibbs R, Boerwinkle E.

Identification of novel structural variations affecting common and complex disease risks with >16,000 whole genome sequences from ARIC and HCHS/SOL. Abstract. The American Society of Human Genetics; 2018.

128. Raffield LM, Ulirsch JC, Naik RP, Lessard S, Handsaker RE, Jain D, Kang HM, Pankratz N, Auer PL, Bao EL, Smith JD, Lange LA, Lange EM, Li Y, Thornton TA, Young BA, Abecasis GR, Laurie CC, Nickerson DA, McCarroll SA, Reiner AP. Common α -globin variants modify hematologic and other clinical phenotypes in sickle cell trait and disease. *PLoS Genet*. 2018;14:e1007293. [PubMed: 29590102]
129. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, Bouatia-Naji N, Langenberg C, Prokopenko I, Stolerman E, Sandhu MS, Heeney MM, Devaney JM, Reilly MP, Ricketts SL, Stewart AFR, Voight BF, Willenborg C, Wright B, Altshuler D, Meigs JB. Common variants at 10 genomic loci influence hemoglobin A₁(C) levels via glyceemic and nonglyceemic pathways. *Diabetes*. 2010;59:3229–3239. [PubMed: 20858683]
130. Cooper A, Ilboudo H, Alibu VP, Ravel S, Enyaru J, Weir W, Noyes H, Capewell P, Camara M, Milet J, Jamonneau V, Camara O, Matovu E, Bucheton B, MacLeod A. APOL1 renal risk variants have contrasting resistance and susceptibility associations with African trypanosomiasis. *elife*. 2017;6.
131. Kopp JB, Nelson GW, Sampath K, Johnson RC, Genovese G, An P, Friedman D, Briggs W, Dart R, Korbet S, Mokrzycki MH, Kimmel PL, Limou S, Ahuja TS, Berns JS, Fryc J, Simon EE, Smith MC, Trachtman H, Michel DM, Winkler CA. APOL1 genetic variants in focal segmental glomerulosclerosis and HIV-associated nephropathy. *J Am Soc Nephrol*. 2011;22:2129–2137. [PubMed: 21997394]
132. Dummer PD, Limou S, Rosenberg AZ, Heymann J, Nelson G, Winkler CA, Kopp JB. APOL1 kidney disease risk variants: an evolving landscape. *Semin Nephrol*. 2015;35:222–236. [PubMed: 26215860]
133. Parsa A, Kao WHL, Xie D, Astor BC, Li M, Hsu C, Feldman HI, Parekh RS, Kusek JW, Greene TH, Fink JC, Anderson AH, Choi MJ, Wright JT, Lash JP, Freedman BI, Ojo A, Winkler CA, Raj DS, Kopp JB, CRIC Study Investigators. APOL1 risk variants, race, and progression of chronic kidney disease. *N Engl J Med*. 2013;369:2183–2196. [PubMed: 24206458]
134. Key NS, Derebail VK. Sickle-cell trait: novel clinical significance. *Hematology Am Soc Hematol Educ Program*. 2010;2010:418–422. [PubMed: 21239829]
135. Naik RP, Derebail VK, Grams ME, Franceschini N, Auer PL, Peloso GM, Young BA, Lettre G, Peralta CA, Katz R, Hyacinth HI, Quarells RC, Grove ML, Bick AG, Fontanillas P, Rich SS, Smith JD, Boerwinkle E, Rosamond WD, Ito K, Reiner AP. Association of sickle cell trait with chronic kidney disease and albuminuria in African Americans. *JAMA*. 2014;312:2115–2125. [PubMed: 25393378]
136. Naik RP, Irvin MR, Judd S, Gutiérrez OM, Zakai NA, Derebail VK, Peralta C, Lewis MR, Zhi D, Arnett D, McClellan W, Wilson JG, Reiner AP, Kopp JB, Winkler CA, Cushman M. Sickle cell trait and the risk of ESRD in blacks. *J Am Soc Nephrol*. 2017;28:2180–2187. [PubMed: 28280138]
137. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kuttyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190–1195. [PubMed: 22955828]
138. Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, Gupta N, Neale BM, Daly MJ, Sklar P, Sullivan PF, Bergen S, Moran JL, Hultman CM, Lichtenstein P, Magnusson P, Purcell SM, Haas DW, Liang L, Sunyaev S, Price AL. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat Genet*. 2012;44:631–635. [PubMed: 22610117]
139. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*. 2008;9:403–433. [PubMed: 18593304]
140. Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, Smith GD, Mahajan A, Gaulton KJ, Nadkarni GN, Valladares-Salgado A, Wachter-Rodarte N, Mychaleckyj JC, Dueker

- ND, Guo X, Hai Y, Haessler J, Kamatani Y, Stilp AM, Zhu G, Franceschini N. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat Commun.* 2019;10:29. [PubMed: 30604766]
141. Avery CL, Wassel CL, Richard MA, Highland HM, Bien S, Zubair N, Soliman EZ, Fornage M, Bielinski SJ, Tao R, Seyerle AA, Shah SJ, Lloyd-Jones DM, Buyske S, Rotter JI, Post WS, Rich SS, Hindorff LA, Jeff JM, Shohet RV, North KE. Fine mapping of QT interval regions in global populations refines previously identified QT interval loci and identifies signals unique to African and Hispanic descent populations. *Heart Rhythm.* 2017;14:572–580. [PubMed: 27988371]
142. Zubair N, Graff M, Luis Ambite J, Bush WS, Kichaev G, Lu Y, Manichaikul A, Sheu WH- H, Absher D, Assimes TL, Bielinski SJ, Bottinger EP, Buzkova P, Chuang L- M, Chung R- H, Cochran B, Dumitrescu L, Gottesman O, Haessler JW, Haiman C, Carty CL. Fine-mapping of lipid regions in global populations discovers ethnic-specific signals and refines previously identified lipid loci. *Hum Mol Genet.* 2016;25:5500–5512. [PubMed: 28426890]
143. Fernández-Rhodes L, Gong J, Haessler J, Franceschini N, Graff M, Nishimura KK, Wang Y, Highland HM, Yoneyama S, Bush WS, Goodloe R, Ritchie MD, Crawford D, Gross M, Fornage M, Buzkova P, Tao R, Isasi C, Avilés-Santa L, Daviglus M, North KE. Trans-ethnic fine-mapping of genetic loci for body mass index in the diverse ancestral populations of the Population Architecture using Genomics and Epidemiology (PAGE) Study reveals evidence for multiple signals at established loci. *Hum Genet.* 2017;136:771–800. [PubMed: 28391526]
144. Cannon ME, Duan Q, Wu Y, Zeynalzadeh M, Xu Z, Kangas AJ, Soinen P, Ala-Korpela M, Civelek M, Lusi AJ, Kuusisto J, Collins FS, Boehnke M, Tang H, Laakso M, Li Y, Mohlke KL. Trans-ancestry Fine Mapping and Molecular Assays Identify Regulatory Variants at the ANGPTL8 HDL-C GWAS Locus. *G3 (Bethesda).* 2017;7:3217–3227. [PubMed: 28754724]
145. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51:584–591. [PubMed: 30926966]
146. Duncan L, Shen H, Gelaye B, Meijisen J, Ressler K, Feldman M, Peterson R, Domingue B. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun.* 2019;10:3328. [PubMed: 31346163]
147. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet.* 2017;100:635–649. [PubMed: 28366442]
148. Spaeth E, Starlard-Davenport A, Allman R. Bridging the Data Gap in Breast Cancer Risk Assessment to Enable Widespread Clinical Implementation across the Multiethnic Landscape of the US. *J Cancer Treatment Diagn.* 2018;2:1–6. [PubMed: 30662981]
149. Onengut-Gumuscu S, Chen W- M, Robertson CC, Bonnie JK, Farber E, Zhu Z, Oksenberg JR, Brant SR, Bridges SL, Edberg JC, Kimberly RP, Gregersen PK, Rewers MJ, Steck AK, Black MH, Dabelea D, Pihoker C, Atkinson MA, Wagenknecht LE, Divers J, Rich SS. Type 1 Diabetes Risk in African-Ancestry Participants and Utility of an Ancestry-Specific Genetic Risk Score. *Diabetes Care.* 2019;42:406–415. [PubMed: 30659077]
150. Ding K, Bailey KR, Kullo IJ. Genotype-informed estimation of risk of coronary heart disease based on genome-wide association data linked to the electronic medical record. *BMC Cardiovasc Disord.* 2011;11:66. [PubMed: 22151179]
151. Kullo IJ, Jouni H, Austin EE, Brown S- A, Kruisselbrink TM, Isseh IN, Haddad RA, Marroush TS, Shameer K, Olson JE, Broeckel U, Green RC, Schaid DJ, Montori VM, Bailey KR. Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation.* 2016;133:1181–1188. [PubMed: 26915630]
152. Snell K, Helén I. “Well, I knew this already” - explaining personal genetic risk information through narrative meaning-making. *Sociol Health Illn.* 2019;
153. Illumina. Polygenic risk: What’s the score? *Nature Research [Internet].* 2019 [cited 2019 Dec 1]; Available from: <https://www.nature.com/articles/d42473-019-00270-w>
154. Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, Distefano M, Senol-Cosar O, Haas ME, Bick A, Aragam KG, Lander ES, Smith GD, Mason-Suares H, Fornage M, Lebo M,

- Timpson NJ, Kaplan LM, Kathiresan S. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell*. 2019;177:587–596.e9. [PubMed: 31002795]
155. Márquez-Luna C, Loh P- R, South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol*. 2017;41:811–823. [PubMed: 29110330]
156. Knowles JW, Ashley EA. Cardiovascular disease: The rise of the genetic risk score. *PLoS Med*. 2018;15:e1002546. [PubMed: 29601582]
157. Grinde KE, Qi Q, Thornton TA, Liu S, Shadyab AH, Chan KHK, Reiner AP, Sofer T. Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genet Epidemiol*. 2019;43:50–62. [PubMed: 30368908]
158. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, Schneiderman N, Raij L, Talavera G, Allison M, Lavange L, Chambless LE, Heiss G. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20:629–641. [PubMed: 20609343]
159. Lin D- Y, Tao R, Kalsbeek WD, Zeng D, Gonzalez F, Fernández-Rhodes L, Graff M, Koch GG, North KE, Heiss G. Genetic association analysis under complex survey sampling: the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet*. 2014;95:675–688. [PubMed: 25480034]
160. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol*. 2015;39:276–293. [PubMed: 25810074]
161. Olson JL, Bild DE, Kronmal RA, Burke GL. Legacy of MESA. *Glob Heart*. 2016;11:269–274. [PubMed: 27741974]
162. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, Liu K, Shea S, Szklo M, Bluemke DA, O’Leary DH, Tracy R, Watson K, Wong ND, Kronmal RA. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med*. 2008;358:1336–1345. [PubMed: 18367736]
163. Kaufman JD, Adar SD, Barr RG, Budoff M, Burke GL, Curl CL, Daviglus ML, Diez Roux AV, Gassett AJ, Jacobs DR, Kronmal R, Larson TV, Navas-Acien A, Olives C, Sampson PD, Sheppard L, Siscovick DS, Stein JH, Szpiro AA, Watson KE. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. *Lancet*. 2016;388:696–704. [PubMed: 27233746]
164. Garg PK, McClelland RL, Jenny NS, Criqui MH, Greenland P, Rosenson RS, Siscovick DS, Jorgensen N, Cushman M. Lipoprotein-associated phospholipase A2 and risk of incident cardiovascular disease in a multi-ethnic cohort: The multi ethnic study of atherosclerosis. *Atherosclerosis*. 2015;241:176–182. [PubMed: 26004387]
165. Guan W, Cao J, Steffen BT, Post WS, Stein JH, Tattersall MC, Kaufman JD, McConnell JP, Hoefner DM, Warnick R, Tsai MY. Race is a key variable in assigning lipoprotein(a) cutoff values for coronary heart disease risk assessment: the Multi-Ethnic Study of Atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2015;35:996–1001. [PubMed: 25810300]
166. Psaty BM, Sitlani C. The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium as a model of collaborative science. *Epidemiology*. 2013;24:346–348. [PubMed: 23549178]
167. Liu Y, Ding J, Reynolds LM, Lohman K, Register TC, De La Fuente A, Howard TD, Hawkins GA, Cui W, Morris J, Smith SG, Barr RG, Kaufman JD, Burke GL, Post W, Shea S, McCall CE, Siscovick D, Jacobs DR, Tracy RP, Hoeschele I. Methyloomics of gene expression in human monocytes. *Hum Mol Genet*. 2013;22:5065–5074. [PubMed: 23900078]
168. About | RURAL Cohort Study [Internet]. [cited 2020 Mar 18]; Available from: <https://www.theruralstudy.org/about/>
169. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, Li L, China Kadoorie Biobank (CKB) collaborative group. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40:1652–1666. [PubMed: 22158673]

170. Gan W, Walters RG, Holmes MV, Bragg F, Millwood IY, Banasik K, Chen Y, Du H, Iona A, Mahajan A, Yang L, Bian Z, Guo Y, Clarke RJ, Li L, McCarthy MI, Chen Z, China Kadoorie Biobank Collaborative Group. Evaluation of type 2 diabetes genetic risk variants in Chinese adults: findings from 93,000 individuals from the China Kadoorie Biobank. *Diabetologia*. 2016;59:1446–1457. [PubMed: 27053236]
171. Tapia-Conyer R, Kuri-Morales P, Alegre-Díaz J, Whitlock G, Emberson J, Clark S, Peto R, Collins R. Cohort profile: the Mexico City Prospective Study. *Int J Epidemiol*. 2006;35:243–249. [PubMed: 16556648]
172. University of Oxford. The Mexico City Prospective Study [Internet]. UK Research and Innovation [cited 2020 Mar 18]; Available from: https://gtr.ukri.org/projects?ref=MC_UU_00017%2F2
173. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. *Genet Med*. 2017;19:743–750. [PubMed: 27929525]
174. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, Guarino P, Aslan M, Anderson D, LaFleur R, Hammond T, Schaa K, Moser J, Huang G, Muralidhar S, Przygodzki R, O'Leary TJ. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214–223. [PubMed: 26441289]
175. Regalado A China's BGI says it can sequence a genome for just \$100. *MIT Technology Review* [Internet]. 2020 [cited 2020 Mar 18]; Available from: <https://www.technologyreview.com/s/615289/china-bgi-100-dollar-genome/>
176. Stark Z, Dolman L, Manolio TA, Ozenberger B, Hill SL, Caulfield MJ, Levy Y, Glazer D, Wilson J, Lawler M, Boughtwood T, Braithwaite J, Goodhand P, Birney E, North KN. Integrating Genomics into Healthcare: A Global Responsibility. *Am J Hum Genet*. 2019;104:13–20. [PubMed: 30609404]
177. Brothers KB, Rothstein MA. Ethical, legal and social implications of incorporating personalized medicine into healthcare. *Per Med*. 2015;12:43–51. [PubMed: 25601880]
178. Temple R, Stockbridge NL. BiDil for heart failure in black patients: The U.S. Food and Drug Administration perspective. *Ann Intern Med*. 2007;146:57–62. [PubMed: 17200223]
179. Kahn J Misreading race and genomics after BiDil. *Nat Genet*. 2005;37:655–656. [PubMed: 15990879]
180. Wu AH, White MJ, Oh S, Burchard E. The Hawaii clopidogrel lawsuit: the possible effect on clinical laboratory testing. *Per Med*. 2015;12:179–181. [PubMed: 29771642]
181. Duster T *Backdoor to Eugenics*. Routledge; 2004.
182. Phelan JC, Link BG, Feldman NM. The Genomic Revolution and Beliefs about Essential Racial Differences: A Backdoor to Eugenics? *Am Sociol Rev*. 2013;78:167–191. [PubMed: 24855321]
183. Hindorff LA, Bonham VL, Ohno-Machado L. Enhancing diversity to reduce health information disparities and build an evidence base for genomic medicine. *Per Med*. 2018;15:403–412. [PubMed: 30209973]
184. Fullerton SM. The Input-Output Problem: Whose DNA Do We Study, and Why Does It Matter? In: Burke W, Edwards KA, Goering S, Holland S, Trinidad SB, editors. *Achieving Justice in Genomic Translation: Re-Thinking the Pathway to Benefit*. New York, NY: Oxford University Press; 2011 p. 40–55.
185. Wright GEB, Koornhof PGJ, Adeyemo AA, Tiffin N. Ethical and legal implications of whole genome and whole exome sequencing in African populations. *BMC Med Ethics*. 2013;14:21. [PubMed: 23714101]
186. Roberts MC, Kennedy AE, Chambers DA, Khoury MJ. The current state of implementation science in genomic medicine: opportunities for improvement. *Genet Med*. 2017;19:858–863. [PubMed: 28079898]
187. Callier SL, Abudu R, Mehlman MJ, Singer ME, Neuhauser D, Caga-Anan C, Wiesner GL. Ethical, legal, and social implications of personalized genomic medicine research: current literature and suggestions for the future. *Bioethics*. 2016;30:698–705. [PubMed: 27767224]
188. Li SX, Ye Z, Whelan K, Truby H. The effect of communicating the genetic risk of cardiometabolic disorders on motivation and actual engagement in preventative lifestyle

- modification and clinical outcome: a systematic review and meta-analysis of randomised controlled trials. *Br J Nutr.* 2016;116:924–934. [PubMed: 27405704]
189. Mensah GA, Jaquish C, Srinivas P, Papanicolaou GJ, Wei GS, Redmond N, Roberts MC, Nelson C, Aviles-Santa L, Puggal M, Green Parker MC, Minear MA, Barfield W, Fenton KN, Boyce CA, Engelgau MM, Khoury MJ. Emerging concepts in precision medicine and cardiovascular diseases in racial and ethnic minority populations. *Circ Res.* 2019;125:7–13. [PubMed: 31219738]
 190. Huibregtse BM, Boardman JD. Provider bias as a function of patient genotype: polygenic score analysis among diabetics from the Health and Retirement Study. *Obes Sci Pract.* 2018;4:448–454. [PubMed: 30338115]
 191. Palk AC, Dalvie S, de Vries J, Martin AR, Stein DJ. Potential use of clinical polygenic risk scores in psychiatry - ethical implications and communicating high polygenic risk. *Philos Ethics Humanit Med.* 2019;14:4. [PubMed: 30813945]
 192. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366:447–453. [PubMed: 31649194]
 193. Gibson G On the utilization of polygenic risk scores for therapeutic targeting. *PLoS Genet.* 2019;15:e1008060. [PubMed: 31022172]
 194. Godard B, Ozdemir V, Fortin M, Egalite N. Ethnocultural community leaders' views and perceptions on biobanks and population specific genomic research: a qualitative research study. *Public Understanding of Science.* 2010;19:469–485. [PubMed: 20977184]
 195. Regalado A White-people-only DNA tests show how unequal science has become. MIT Technology Review [Internet]. 2018 [cited 2019 Dec 18]; Available from: <https://www.technologyreview.com/s/612322/white-people-only-dna-tests-show-how-unequal-science-has-become/>
 196. Divers J, Redden DT, Rice KM, Vaughan LK, Padilla MA, Allison DB, Bluemke DA, Young HJ, Arnett DK. Comparing self-reported ethnicity to genetic background measures in the context of the Multi-Ethnic Study of Atherosclerosis (MESA). *BMC Genet.* 2011;12:28. [PubMed: 21375750]
 197. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med.* 2017;19:249–255. [PubMed: 27854360]
 198. Suter SM. GINA at 10 years: the battle over “genetic information” continues in court. *J Law Biosci.* 2018;5:495–526. [PubMed: 31143453]
 199. Lello L, Raben TG, Yong SY, Tellier LCAM, Hsu SDH. Genomic prediction of 16 complex disease risks including heart attack, diabetes, breast and prostate cancer. *Sci Rep.* 2019;9:15286. [PubMed: 31653892]
 200. Genomic Prediction. Frequently Asked Questions [Internet]. [cited 2019 Nov 8]; Available from: <https://genomicprediction.com/faqs/#faq-7.2>
 201. Karavani E, Zuk O, Zeevi D, Atzmon G, Barzilai N, Stefanis NC, Hatzimanolis A, Smyrnis N, Avramopoulos D, Kruglyak L, Lam M, Lencz T, Carmi S. Screening human embryos for polygenic traits has limited utility. *BioRxiv.* 2019;
 202. GIANT consortium [Internet]. 2019 [cited 2020 Mar 18]; Available from: https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium
 203. Magic Investigators - Home Page [Internet]. [cited 2020 Mar 18]; Available from: <https://www.magicinvestigators.org/>
 204. Global Lipids Genetics Consortium [Internet]. [cited 2020 Mar 18]; Available from: <http://lipidgenetics.org/>
 205. Buseh AG, Underwood SM, Stevens PE, Townsend L, Kelber ST. Black African immigrant community leaders' views on participation in genomics research and DNA biobanking. *Nurs Outlook.* 2013;61:196–204. [PubMed: 23219004]

206. Hiratsuka V, Brown J, Dillard D. Views of biobanking research among Alaska native people: the role of community context. *Prog Community Health Partnersh*. 2012;6:131–139. [PubMed: 22820223]
207. University of Illinois. Summer Internship for Indigenous Peoples in Genomics (SING) [Internet]. Carl R. Woese Institute for Genomic Biology. [cited 2020 Jan 15]; Available from: <https://sing.igb.illinois.edu/>
208. SING Aotearoa. Summer Internship for Indigenous Genomics Aotearoa [Internet]. [cited 2020 Jan 15]; Available from: <https://www.singaotearoa.nz/>
209. SING Australia. Summer Internship for Indigenous Peoples in Genomics [Internet]. [cited 2020 Jan 15]; Available from: <https://www.singaustralia.org>
210. University of Alberta. Summer Internship for Indigenous Peoples in Genomics Canada (SING Canada) [Internet]. Indigenous STS (Science, Technology, and Society). 2020 [cited 2020 Jan 15]; Available from: <https://indigenousts.com/sing-canada/>
211. Wade L To overcome decades of mistrust, a workshop aims to train Indigenous researchers to be their own genome experts. *Science*. 2018;
212. Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: A Fast Distance-Based Method To Infer Subject Ancestry from Multiple Genotype Datasets Without Principal Components Analysis. *G3 (Bethesda)*. 2019;9:2447–2461. [PubMed: 31151998]
213. Fang H, Hui Q, Lynch J, Honerlaw J, Assimes TL, Huang J, Vujkovic M, Damrauer SM, Pyarajan S, Gaziano JM, DuVall SL, O'Donnell CJ, Cho K, Chang K- M, Wilson PWF, Tsao PS, VA Million Veteran Program, Sun YV, Tang H. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am J Hum Genet*. 2019;105:763–772. [PubMed: 31564439]

Box 1.**Key Terms Related to the Genetic Epidemiology of CMD**

| |
|---|
| Allele - alternate forms of a gene at a particular location in the genome. |
| Admixture (or gene flow) - refers to the evolutionary blending of populations. This term makes the assumption that the populations were separated for many generations before individuals began to “admix”, resulting in the current population’s cultural, linguistic or ancestral characteristics. |
| Copy number variant (CNV) - a one kilobase or larger DNA segment present at different amounts (number of copies) compared to a reference genome. ²³ |
| De novo familial variant - a genetic variant which arose in a single family through germ cell mutation. |
| Fine-mapping - a set of statistical procedures to identify potential causal variants in a genomic region associated with a particular trait. Diverse populations are particularly important here, as differences in linkage disequilibrium (LD) can narrow regions with fewer candidate causal variants for interrogation than in European descent populations. ²⁴ |
| Founder population - a group descended from a comparatively small number of individuals, which due to the reduced population size does not contain all of the genetic variation of the parental population. The anatomically modern human groups who migrated out of Africa are founder populations, as are many geographic or religious isolates. |
| Genetic drift - random changes in the frequency of genetic variants (alleles) in a population across time. |
| Genetic epidemiology - the study of the role of genetic factors (and their interaction with other genetic, environmental, and socio-cultural factors) on health and disease in human populations. |
| Genome-wide association study (GWAS) - a study which assesses the association of genetic variants across the genome with a trait of interest, generally using a separate linear or logistic regression model for each variant and adjusting for covariates as appropriate. Principal components are used to account for population stratification. |
| Gene by environment interaction study - a study that quantifies either the extent that a genetic effect varies across non-genetic (environmental or socio-cultural) factors, or the extent that the effect of a non-genetic factor varies by genotype. |
| Gene expression - transcription and translation of genetic code into phenotype. |
| Genetic variation - In this review, we define genetic variation in human populations in terms of a gene pool, or a group of organisms of the same species that live in the same area (may be micro- or macro-defined). ²⁵ |
| Genotype - the genetic makeup of an individual. |
| Imputation - a method for inferring genetic variants not included on genotyping arrays, based on LD patterns in known population reference panels. ²⁶ |
| Indel - an insertion or deletion of less than one kilobase of nucleotides in a genomic region. |
| Linkage disequilibrium (LD) - the non-random association of variants in a population, so that variants in LD tend to be inherited together. European descent populations generally have longer blocks of LD across the genome, such that GWAS associations in these populations can tag large regions of the genome including hundreds of potential causal variants. Diverse ancestry populations have different patterns of LD, and populations with recent admixture (African Americans and Hispanic/Latinos, for example) may have much shorter LD blocks, narrowing the tagging region of GWAS associations. |
| Mutation - the alteration of the DNA sequence of the genome of an organism or extrachromosomal DNA. Mutations may be characterized by the functional consequences, for example missense, nonsense, nonsynonymous, etc. |
| Natural selection (or selective pressure) - a change in the allele frequencies across generations that is driven by differential fertility or differential survival, as processes in which individuals with certain traits are able to differentially survive and reproduce in a given environment. |
| Nonsense variant - a mutation that results in the premature termination of a protein coding sequence. |
| Penetrance - describes the proportion of a population that have a particular genetic variant and also express the phenotype associated with that variant. |
| Phenotype - the observable traits of an individual, including physical appearance, behavior, biomarkers, and clinical measures that are a product of the genotype interacting with other genetic, environmental, and socio-cultural factors. |
| Polygenic risk scores (PRS) - a composite measure of multiple genetic risk factors for a particular trait or disease, typically calculated by summing the number of risk variants across the genome, with or without weighting by their effect sizes. PRS have primarily been developed to summarize genome-wide data from |

European descent populations, and their portability to other populations is complicated by differences in LD across populations.

Population bottleneck - a sharp reduction in population size, which can lead to a reduction in genetic variation. Bottlenecks can be caused by environmental conditions (natural disasters), disease (epidemics), or human behavior (migration or genocide). The Out-of-Africa migration of anatomically modern humans is an example of a bottleneck, where the relatively small groups that migrated out of Africa for Europe and Asia contained only a subset of the genetic diversity present in the entire continent of Africa.²⁷

Reference panel - a well-defined population, typically comprised of individuals whose four grandparents were all from a specific geographic area, used for imputation of genotype data in GWAS.

Single nucleotide polymorphism (SNP) - variation at a single nucleic acid position in the genome that occurs at a frequency of greater than 0.5% in a population. SNPs are the most common and well characterized type of genetic variation.²⁸

Single nucleotide variants (SNV) - variation at a single nucleic acid position in the genome with no frequency restrictions.

Box 2.**Key Terms Related to Describing Ancestral Diversity.**

The terms ancestry, diversity, ethnicity, and race have no gold standard scientific definitions.^{35,36} Indeed, often terms are erroneously used interchangeably, for example race and ethnicity.⁴⁰ Below, for clarity, we define these terms as used in this review.

Ancestry- We use ancestry (or ancestral background) to refer to the continental origin of the ancestors of an individual or population, and to a lesser extent to the population dynamics within each continent that shaped the observed genetic patterns. Genetic ancestry is often estimated via comparison of participants' genotypes to continental and/or global reference populations, so incomplete availability and the small sample sizes of these reference populations is an important concern.³⁷ Note that discrete labeling of ancestral populations oversimplifies genetic variation. However, given differences in allele frequencies and linkage disequilibrium (LD) across populations, estimating and accounting for ancestry (either discretely or continuously) is necessary for appropriately powered and calibrated genetic analyses.⁴¹⁻⁴³

Diversity- references many aspects and may include genetics, ancestry, gender/sexual orientation, age, culture, abilities/disabilities, geography, socioeconomics, etc. However, in this review we use the term diversity to refer to the genetic and ancestral diversity that drive differences in health, specifically CMD burden and related health parameters. We acknowledge that genetic or ancestral diversity are often conflated with the socially and culturally defined constructs of race and ethnicity.

Ethnicity- A socially and culturally heterogeneous term that represents how individuals and/or groups of individuals identify based on shared history, cultural traditions, and ancestry, which incorporates both societal norms and an individual's own self-perception. While ethnicity has been used as a proxy for health and disease risk at the population level, the term is heterogeneously defined worldwide; we acknowledge that within each group there is notable ancestral diversity and that individuals within the US or across the globe may prefer to use different constructs.

Race- A socially and culturally defined term used to refer to an individual and/or group of individuals. Like ethnicity, race is often used to refer to the shared history, language and ancestry of individuals and/or populations, but it does not singularly predict genetic susceptibility to disease, genotype, or drug response of an individual patient.^{44,45} Historically, the construct of race has been linked to visible physical attributes, and used as a justification for discrimination to establish and reinforce social inequities. As a result, environmental and sociocultural factors have been shown to differentially track with racial groupings; for example, disparities in access to goods and services has had a large impact on health and disease.^{46,47}

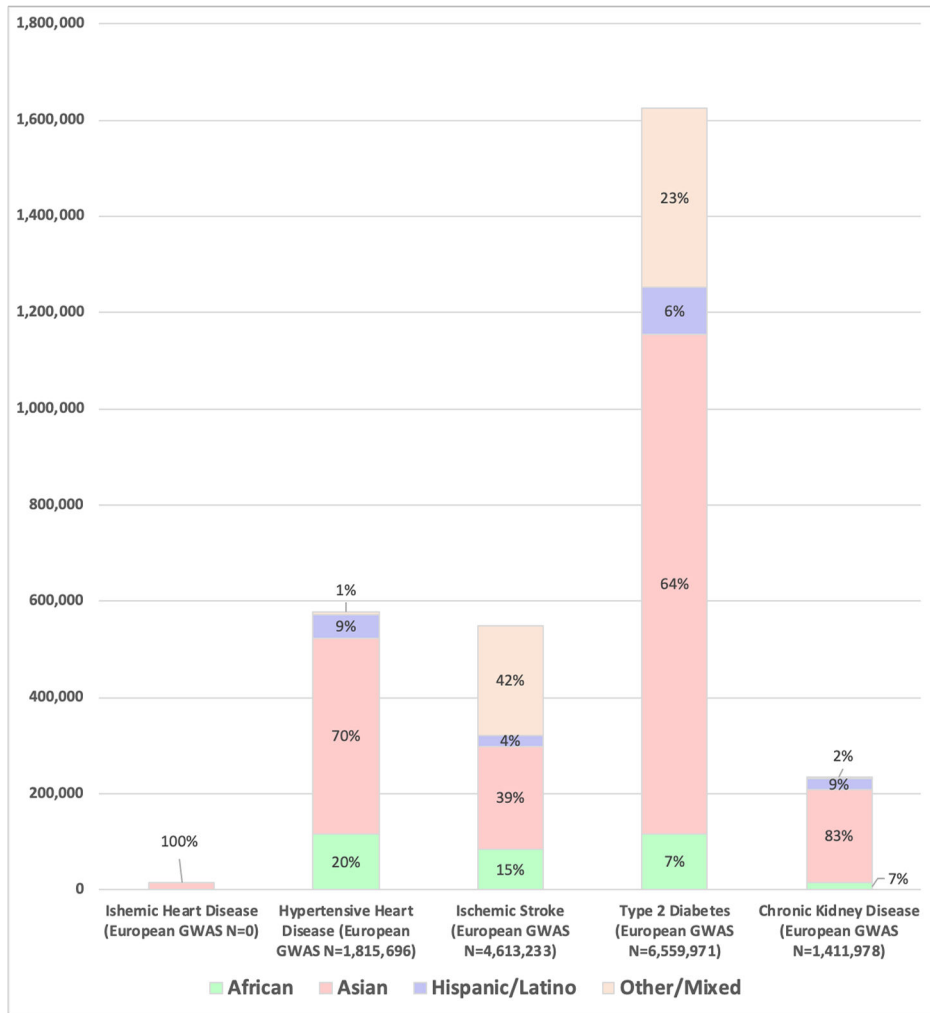


Figure 1. Participant Diversity in GWAS of Cardiometabolic Traits. Non-Europeans make up just 11% of GWAS of ischemic stroke, 17% of GWAS of chronic kidney disease, 20% of GWAS of Type 2 diabetes, and 24% of GWAS of hypertensive heart disease, and the vast majority of non-European GWAS participants are Asian. Trait definitions note: Ischemic Heart Disease included GWAS traits: Ischemic heart disease. Hypertensive Heart Disease definition included GWAS traits: hypertension, hypertension (young onset), and Medication use (antihypertensives). Ischemic Stroke definition included GWAS traits: Ischemic stroke, Ischemic stroke (cardioembolic), Ischemic stroke (large artery atherosclerosis), Ischemic stroke (non-cardioembolic), Ischemic stroke (small artery occlusion), Ischemic stroke (small-vessel), Ischemic stroke (undetermined subtype). Type 2 Diabetes definition included GWAS traits: Type 2 diabetes, Type 2 diabetes adjusted for BMI, prevalent Type 2 diabetes. Chronic Kidney Disease definition included GWAS traits: Chronic kidney disease, Incident chronic kidney disease, Renal function and chronic kidney disease, Chronic kidney disease and diabetic kidney disease in diabetes, Chronic kidney disease and diabetic kidney disease in type 2 diabetes, Chronic kidney disease in diabetes, Chronic kidney disease in type 2 diabetes, Chronic kidney disease (severe chronic kidney

disease vs normal kidney function) in type 1 diabetes, Chronic kidney disease (chronic kidney disease vs normal or mildly reduced eGFR) in type 1 diabetes, Chronic kidney disease (reduced eGFR or end stage renal disease) in type 1 diabetes, Chronic kidney disease (end stage renal disease vs. normal eGFR) in type 1 diabetes. Data from <https://gwasdiversitymonitor.com>, accessed March 18, 2020.⁴

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

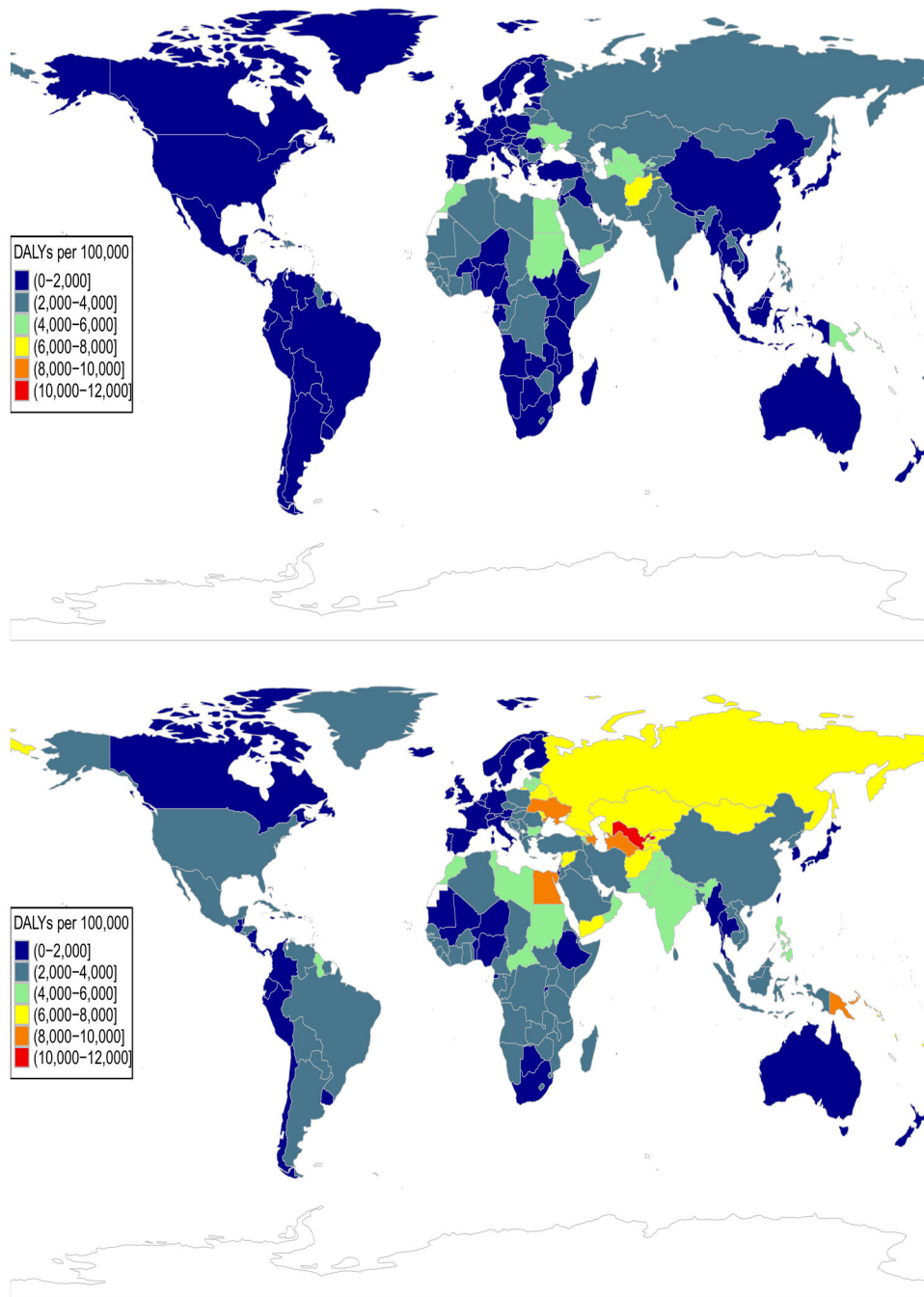
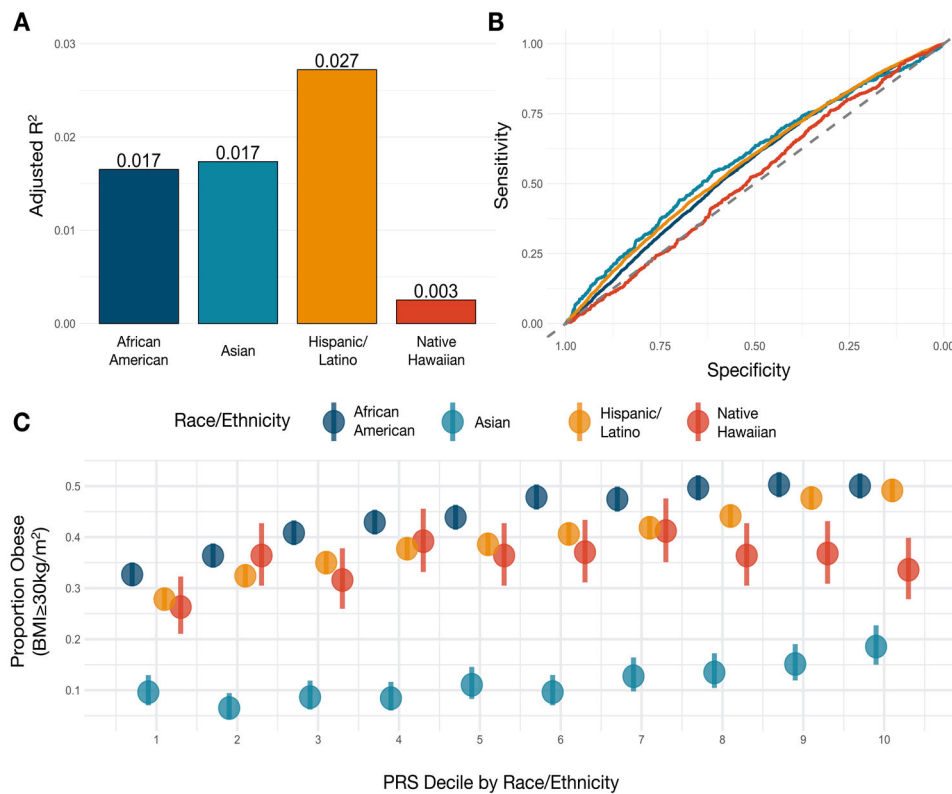


Figure 2 A-B.
 Global burden of ischemic heart disease
 Colors represent binned values of age-standardized disability adjusted life-years (DALYS) from ischemic heart disease per 100,000 population for females (Panel A) and males (Panel B). Data used for the figure was downloaded from the Institute for Health Metrics and Evaluation's Global Burden of Disease (GBD) Compare Data Visualization tool and is available at <http://vizhub.healthdata.org/gbd-compare>.⁴⁹

**Figure 3A-C.**

Performance and distribution of Polygenic Risk Score (PRS) for obesity in PAGE participants.

(A) The adjusted R^2 of the PRS developed by Khera and colleagues¹⁵⁴ in PAGE participants, stratified by self-identified race/ethnicity. The risk scores were standardized by self-identified race/ethnicity and outliers beyond 4 standard deviations were removed for these analyses.

(B) The performance of the PRS on obesity ($BMI \geq 30\text{kg/m}^2$), stratified by self-identified race/ethnicity. The highest area under the curve was found in East and South Asian participants.

(C) The distribution of obesity ($BMI \geq 30\text{kg/m}^2$) by race/ethnicity-stratified decile of the PRS, demonstrating the differential distributions of obesity by race/ethnicity.

Allele Frequencies from gnomAD browser v2.1.1 (as accessed December 19, 2019) and summary of cohorts included in the cited publications for variants discussed in this review article (in order of appearance).

Table 1.

| Annotated Gene | rsID | Publication | Included cohorts | Amino Acid Change | Trait and Additional Notes | Allele Frequencies, by Ancestral Population (as defined by gnomAD)* | | | | | | | | | |
|----------------|------------|-------------|---|-------------------|--|---|------------------|------------|--------------------|------------------------|--------|-------|-------------|-------|--|
| | | | | | | African | Ashkenazi Jewish | East Asian | European (Finnish) | European (non-Finnish) | Latino | Other | South Asian | | |
| <i>PCSK9</i> | rs28362286 | 12 | Resequencing study in 128 subjects from the Dallas Heart Study (50% African American) with low plasma levels of LDL, subsequent genotyping in n=3553 from Dallas Heart Study | p.Cys679Ter | Plasma Lipids | 0.80% | 0.00% | 0.00% | 0.00% | 0.00% | 0.02% | 0.01% | 0.00% | 0.00% | |
| <i>PCSK9</i> | rs67608943 | | | p.Tyr142Ter | Plasma Lipids | 0.24% | 0.00% | 0.00% | 0.00% | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% | |
| <i>CD36</i> | rs3211938 | 88-93 | Pleiotropic variant, with associations discovered in a number of single study efforts and meta-analyses including at least some African ancestry participants | p.Tyr325Ter | Plasma Lipids | 8.93% | 0.00% | 0.00% | 0.00% | 0.00% | 0.27% | 0.24% | 0.01% | 0.01% | |
| <i>APOC3</i> | rs76353203 | 94-96 | Discovered in a small cohort (n=809) of Lancaster Old Order Amish individuals, then followed up in meta-analyses of large population based European cohort studies for impacts on lipids and vascular disease | p.Arg19Ter | Plasma Lipids; common (~5%) in the Lancaster Old Order Amish | 0.01% | 0.00% | 0.01% | 0.02% | 0.05% | 0.08% | 0.12% | 0.18% | 0.18% | |
| <i>GYPC</i> | rs28515082 | 5 | Multi-ancestry GWAS meta-analysis of non- | | Blood Pressure | 7.84% | 7.78% | 0.00% | 3.24% | 6.68% | 3.30% | 5.14% | 0.00% | 0.00% | |

| Annotated Gene | rsID | Publication | Included cohorts | Amino Acid Change | Trait and Additional Notes | Allele Frequencies, by Ancestral Population (as defined by gnomAD)* | | | | | | | |
|-----------------|-------------|----------------------------|--|--|--|---|------------------|------------|--------------------|------------------------|--------|-------|-------------|
| | | | | | | African | Ashkenazi Jewish | East Asian | European (Finnish) | European (non-Finnish) | Latino | Other | South Asian |
| <i>GPR20</i> | rs111409240 | | European populations from the Population Architecture using Genomics and Epidemiology (PAGE) study (n=49,839) | | Blood Pressure | 20.13% | 0.69% | 0.00% | 0.00% | 0.07% | 2.13% | 1.11% | 0.00% |
| <i>JRKL</i> | rs145054295 | 5 | Multi-ancestry GWAS meta-analysis of non-European populations from the Population Architecture using Genomics and Epidemiology (PAGE) study (n=49,839) | | Hypertension | 2.47% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.09% | 0.00% |
| <i>PRKCH</i> | rs2230500 | 100-102 | Association with stroke reported in several studies and meta-analyses from Japanese and Chinese populations | p.Val374Ile | Stroke | 0.22% | 1.03% | 26.68% | 3.55% | 1.11% | 0.63% | 1.87% | 1.23% |
| <i>SLC16A11</i> | | 11 | Reported in a meta-analysis of Mexican studies (n= 3,848 cases and 4,366 controls) | Haplotype association, frequencies not in gnomAD | Type 2 Diabetes | | | | | | | | |
| <i>HBB</i> | rs334 | 104-111, 122, 123, 134-136 | Sickle cell anemia SNP, cited papers discuss effects on Stroke, HbA1c, chronic kidney disease in a number of African American cohorts | p.Glu7Val | HbA1c | 4.49% | 0.00% | 0.00% | 0.00% | 0.01% | 0.21% | 0.17% | 0.06% |
| <i>TBC1D4</i> | rs61736969 | 113 | Discovered in Greenlandic Inuit population (n=2,575) | | Type 2 Diabetes; 17% minor allele frequency in Greenland | 2.70% | 0.00% | 0.00% | 0.00% | 0.00% | 0.08% | 0.06% | 0.00% |

| Annotated Gene | rsID | Publication | Included cohorts | Amino Acid Change | Trait and Additional Notes | Allele Frequencies, by Ancestral Population (as defined by gnomAD)* | | | | | | | | | |
|-----------------------------|-------------|-------------|---|-------------------|--|---|------------------|------------|--------------------|------------------------|--------|--------|-------------|--|--|
| | | | | | | African | Ashkenazi Jewish | East Asian | European (Finnish) | European (non-Finnish) | Latino | Other | South Asian | | |
| <i>FA2S</i> gene cluster | rs71115739 | 114 | Candidate SNPs identified by selective pressure analyses in Greenlandic population, follow-up in three cohorts of Greenlanders or Greenlandic individuals who live in Denmark (n=2733, n=1331, and n=541) | | Type 2 Diabetes | 73.40% | 95.17% | 72.81% | 96.00% | 96.71% | 61.56% | 91.18% | 0.00% | | |
| | rs78193826 | 116 | Meta-analysis of four type 2 diabetes GWAS (36,614 cases and 155,150 controls of Japanese ancestry) | p.Val282Met | Type 2 Diabetes | 0.17% | 0.01% | 7.89% | 0.44% | 0.10% | 0.03% | 0.69% | 3.46% | | |
| | rs77792157 | | | p.Ala341Thr | Type 2 Diabetes | 0.00% | 0.00% | 0.23% | 0.00% | 0.00% | 0.00% | 0.01% | 0.00% | | |
| <i>GLPIR</i> | rs3765467 | | | p.Arg131Gln | Type 2 Diabetes | 0.22% | 0.19% | 23.34% | 0.03% | 0.27% | 4.53% | 1.87% | 8.63% | | |
| <i>PAX4</i> | rs2233580 | 117 | Exome sequencing in n=12,940 individuals from five ancestry groups, East Asian specific association | p.Arg192His | Type 2 Diabetes | 0.00% | 0.00% | 10.94% | 0.00% | 0.00% | 0.01% | 0.25% | 0.03% | | |
| <i>CREBRF</i> | rs373863828 | 118,119 | Discovery of BMI association in 3,072 Samoans, replication in 2,102 additional Samoans | p.Arg457Gln | Body Mass; common in Samoan populations (frequency ~25%) | 0.00% | 0.01% | 0.01% | 0.00% | 0.00% | 0.00% | 0.03% | 0.00% | | |
| | | | | | | | | | | | | | | | |
| <i>G6PD</i> | rs1050828 | 124,125 | Association with HbA1c discovered in trans-ethnic meta-analysis of n=159,940 (n=7,565 African ancestry participants), replicated in whole genome sequencing study | p.Val198Met | HbA1c | 11.64% | 0.00% | 0.00% | 0.00% | 0.02% | 0.41% | 0.30% | 0.04% | | |

| Annotated Gene | rsID | Publication | Included cohorts | Amino Acid Change | Trait and Additional Notes | Allele Frequencies, by Ancestral Population (as defined by gnomAD)* | | | | | | | |
|------------------|------------|-------------|--|--|-----------------------------------|---|------------------|------------|--------------------|------------------------|--------|--------|-------------|
| | | | | | | African | Ashkenazi Jewish | East Asian | European (Finnish) | European (non-Finnish) | Latino | Other | South Asian |
| <i>G6PD</i> | rs76723693 | 125 | including n=10,338 (n=3,123 African Americans) HbA1c association discovered in whole genome sequencing study including n=10,338 (n=3,123 African Americans) | p.Leu353Pro | HbA1c | 0.54% | 0.00% | 0.00% | 0.00% | 0.00% | 0.07% | 0.04% | 0.00% |
| <i>HBA1/HBA2</i> | esv2676630 | 127,128 | Association reported in two African American cohort studies | structural variant, not in gnomAD | HbA1c | | | | | | | | |
| <i>APOLI</i> | rs73885319 | 10, 130-133 | Cited papers discuss <i>APOLI</i> risk haplotype effects on various forms of chronic kidney disease in African ancestry and Hispanic/Latino populations | | Chronic Kidney Disease | 22.76% | 0.00% | 0.00% | 0.00% | 0.01% | 0.68% | 0.58% | 0.01% |
| <i>APOLI</i> | rs71785313 | | | Indel (G2 allele), not in gnomAD- approximate frequencies listed from 1000 Genomes | | 12.90% | | 0.00% | | 0.00% | 0.80% | | 0.00% |
| <i>CYP2C19</i> | rs4986893 | 180 | Cited article discusses differential allele frequencies by population and legal/ethical implications for these well known pharmacogenetic variants | p.Trp212Ter | Response to clopidogrel bisulfate | 0.04% | 0.00% | 6.26% | 0.01% | 0.03% | 0.02% | 0.24% | 0.40% |
| <i>CYP2C19</i> | rs4244285 | | | p.Pro227Pro | Response to clopidogrel bisulfate | 17.76% | 13.20% | 30.75% | 17.50% | 14.68% | 10.12% | 15.95% | 32.40% |

The gnomAD browser is available at: <https://gnomad.broadinstitute.org/>.

* The above populations are listed in alphabetically order and reflect several ancestral super-populations as defined by gnomAD: African (AFR), Ashkenazi Jewish (ASJ), East Asian (EAS), Finnish (FIN), Non-Finnish European (NFE), American Admixed/Latino (AMR), and South Asian (SAS). Other (OTH) is comprised of individuals who did not cluster with any super-population group.

Table 2. Non-exhaustive list of global whole genome sequencing efforts complete or currently underway.

| Study name | Location | Sample size | Ascertainment | Website |
|---|--------------------|--------------------------------|---|---|
| AFRICA | | | | |
| Africa Genome Variation Project | Sub-Saharan Africa | 320 | Population Genetics | https://ega-archive.org/studies/EGAS000001000960 |
| H3 Africa Whole Genome Sequencing Study | Africa | 350 | Population Genetics | https://h3africa.org |
| Southern African Human Genome Programme | South Africa | 24 | Population Genetics | https://www.sahgp.org/index.php |
| ASIA | | | | |
| China Precision Medicine Initiative | China | 100,000,000 | Population-based | http://en.most.gov.cn/eng/organization/Mission/index.htm |
| Genome Asia 100K | Asia | 100,000 | Population-based and clinical studies | https://genomeasia100k.org |
| Japan Genomic Medicine Program | Japan | 100,000 | Population-based and clinical cohorts, drug discovery | https://www.amed.go.jp/en/program/index05.html |
| Singapore 10k | Singapore | 10,000 | Random sample for Population genetics | https://www.sciencedirect.com/science/article/pii/S0092867419310700 |
| Singapore Sequencing Malay Project | Singapore | 100 | Random sample | https://blog.nus.edu.sg/sshsplphg/singapore-sequencing-malay/ |
| AUSTRALIA | | | | |
| Genomics Health Futures Mission | Australia | 200,000 | Population-based, clinical studies, and rare diseases | https://www.health.gov.au/initiatives-and-programs/genomics-health-futures-mission#what-are-the-goals-of-the-genomics-health-futures-mission |
| EUROPE | | | | |
| FarGen - Faroe Genome Project | Denmark | 1,500 | Population Genetics | https://www.fargen.fo/en/home/ |
| Genome Denmark | Denmark | ~60,000 | Population-based and clinical studies | http://www.genomedenmark.dk/english/ |
| Estonian Genome Project | Estonia | 52,000 | Cohort study | https://genomics.ut.ee/en/about-us/estonian-genome-centre |
| Genomics Medicine Plan 2025 | France | 235,000 per year | Cancer, rare and common diseases | https://aviesan.fr/fr/aviesan/accueil/route-l-actualite/plan-france-medicine-genomique-2025 |
| RADICON-CL | Netherlands | *Could not confirm sample size | Rare diseases | https://www.wgs-first.nl/en/project |
| Genomics England (GEL) | United Kingdom | 100,000 | Cancers, rare diseases | https://www.genomicsengland.co.uk |
| Scottish Genomes Partnership | United Kingdom | 2,588 | cancers, rare diseases, population study | https://www.scottishgenomespartnership.org |

| Study name | Location | Sample size | Ascertainment | Website |
|---------------------------------|-----------------|--------------------|--|---|
| MIDDLE EAST | | | | |
| Dubai Genomics | Dubai, UAE | 3,000,000 | Population-based and clinical studies | https://www.dha.gov.ae/en/Pages/DubaiGneomicsAbout.aspx |
| Qatar Genome Project | Qatar | 20,000 | Cohort study | https://qatargenome.org.qa |
| Saudi Arabia Genome Project | Saudi Arabia | > 100,000 | Population-based and clinical studies | https://genomics.saudigenomeprogram.org/en/ |
| Turkish Genome Project | Turkey | 100,000 | Population-based and clinical studies | https://www.bbmri-eric.eu/news-events/turkish-genome-project-launched/ |
| Study name | Location | Sample size | Ascertainment | Website |
| UNITED STATES | | | | |
| All of Us | United States | 1,000,000 | Cohort study | Sequencing tentative |
| NHGRI Genome sequencing project | United States | 141,000 | Mixed trait | https://www.genome.gov/Funded-Programs-Projects/NHGRI-Genome-Sequencing-Program |
| NHLBI TopMED | United States | 140,000 | Mixed trait | https://www.nhlbiwgs.org |
| NYCKidSeq | United States | 1,130 | Undiagnosed neurologic, cardiac, or immune disorders | https://nyckidseq.org |
| Undiagnosed Diseases Network | United States | 1,600 | Undiagnosed rare diseases | https://undiagnosed.hms.harvard.edu |