ISSN : 1001-1781
Journal of Clinical Otorhinolaryngology, Head, and Neck Surgery

GENOME-WIDE ASSOCIATION STUDIES (GWAS) HAVE REVOLUTIONIZED OUR VIEW OF HUMAN HEALTH AND DISEASE GENETICS AND OFFERED NOVEL GENE THERAPY TARGETS

Moataz Dowaidar [1,2]

[1] Department of Bioengineering, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia.

[2] Interdisciplinary Research Center for Hydrogen and Energy Storage (IRC-HES), King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, 31261, Saudi Arabia.

Abstract

Knowing how genetic, behavioural, and sociocultural factors influence each person's risk for C. It is imperative that a larger, more diverse set of genetic studies be done in order to be able to close the analysis of CMD distance In terms of disease prevention, there is a lot of interest in CMD genomic research. This potential can only be achieved if Ancestry DNA-style data like PRS is successfully gathered from the population. Inadequate participation is a big issue in current CMD genetic research. Differential minorities in the United States and Canada have set forth some important steps to improve their access to genetic research. To make sure that this will not happen again, these activities include discovering the issues and using community-based participatory interventions and benefits-sharing mechanisms. People underrepresented in the world of genetics will require more services to support them.

CMD and other genomics markers have been successfully identified and created novel avenues for human and population health change. In addition, it has complicated matters with regard to how this data would affect the broader healthcare system. What are the main questions: disproportionate difficulty in the CMD genotype-phenotype database; confounded research on disease heritability; ethnicity may not be well described, making estimating disease heritability difficultDespite these roadblocks, genome-informed inclusive data will bear unprecedented promise for bringing down CMD and improving wellbeing. A large-scale data unification has already occurred, as mentioned in this blog post to CMD Data sharing, however, is a project that must be done on a small scale in order to gain initial traction. This is also applicable to GWAS research on self-identified ethnicity. Although race and ethnicity are socially and culturally constructed, the use of self-identifying categories in genetic studies still endures.

Introduction

Genome-wide correlation studies (GWAS) have revolutionized our view of human health and disease genetics. These studies usually use imputation to predict genotypes at untyped loci based on whole genome sequenced reference panels with a greater number of variants, depending on genotyping microarrays that evaluate from 100 000 to 2.5 million genetic variants around the genome. 1 Due to the fact that the first genotyping microarrays and comparison panels were developed to measure widespread variation (i.e., genetic variants with

a minor allele frequency [MAF] > 5% in a population) in European populations, our understanding of genetic variation across diverse global populations has been historically limited. 2 Indeed, most GWAS, including those on cardiometabolic diseases (CMDs), were performed in European descent populations as of 2016, with just 5% of participants representing Hispanic/Latino, Pacific Islander, Arab and Middle Eastern, and other native groups in these studies. 3 As of this writing, non-Europeans make up between 11 percent and 24 percent of participants in CMD-related trait GWAS, with the vast majority of non-European participants being of Asian origin, according to the recently published GWAS Diversity Monitor, which monitors participant diversity in the GWAS catalog in real time. 4 While European ancestry GWAS was initially justified as a practical decision to increase control due to their relative homogeneity and large surveys of genotypic results, it is now known how problematic this lack of diversity in genetic studies is. This is especially true for CMD, which has an uneven burden around the world's populations. In reality, a lack of diverse ancestral history representation in genomic studies may unintentionally undercut the potential benefits of precision medicine in the immediate future, particularly for populations overwhelmingly affected by CMD. 5

The key aim of this analysis is to explain the significance and, in certain cases, the need for researching ancestrally diverse groups in order to better understand the genetic underpinnings of CMD. To do so, we start by summarizing core concepts of genetic diversity and then explain why incorporating global populations in CMD genomics research is important. We back up this point by explaining the global variance in CMD prevalence and stressing how important it is to construct a globally representative genetics evidence base. Following that, we go through some of the major advantages of increasing variation in genomic research, such as the discovery of population-specific CMD genetic variations, the significance of fine-mapping, and the calculation of broadly generalizable polygenic risk ratings (PRS). Despite the fact that our analysis is detailed in terms of the need for diversity in CMD genetic studies and the ethical, legal, and social implications for CMD science, we do not discuss solutions for increasing genetic capital and building the necessary infrastructure to integrate diversity into potential genomics research, which have been discussed in depth elsewhere.(6) – (9) Indeed, using diverse populations in genomics studies has already yielded clinical insights for chronic kidney disease and low LDL (low-density lipoprotein) in African-descent populations, as well as type 2 diabetes mellitus (T2D) in Mexican-descent populations (See Section Importance of Variants Specific to a Population below and 2). ten to twelve Although our study focuses on CMD, I'd like to point out that our key points relate to a broad variety of other complex traits and chronic diseases (e.g., schizophrenia,13–17 osteoporosis,18,19, and asthma20–22). Assuming that the problem of diversity can be consistently resolved by coordinated efforts by key stakeholders, we believe that additional, unforeseen prospects for the realization of precision medicine lie ahead.

What Is the Best Way to Describe Human Genetic Variation?

Individual variance, also known as diversity, refers to the spectrum of all potential values for any phenotype which can be due to evolution, environmental influences, and associations between the two. The function of genes is now understood to be complex and modifiable,

notwithstanding their importance. This viewpoint differs from the popular belief that all hereditary inheritance is fixed and deterministic. Person and population genetic variation can take several forms, but it generally refers to variations in the structure (e.g., chromosomal rearrangements or abnormalities) and composition (e.g., DNA sequence) of the genome. While both germ cells and somatic cells have genetic variation, only germ cell variation can be inherited. SNPs (single nucleotide polymorphisms), insertions and deletions (indels), substitutions, inversions, and copy number mutations are all examples of normal human genetic variance. 29 Private (the only copy), de novo familial (a few copies), uncommon (MAF1.0 percent), low-frequency (MAF=1 percent –5 percent), and normal (MAF > 5 percent) genetic variations can all be present throughout the same community. In terms of functional ramifications, most variants are thought to be functionally neutral. 29 Despite the lack of observational data on the practical effects of the overwhelming majority of the approximate 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 Depletion34 have been established and are readily accessible. Because of the design of genotyping microarrays and the increased statistical ability to identify common genetic variant interactions, more detail about common variants occurs. The area of genetic epidemiology is progressively able to distinguish low frequency, uncommon, and de novo familial variants for CMD, as well as facets of genomic structural heterogeneity, thanks to declining sequencing costs.

I'm mainly interested in the importance of ancestral genetic variations in CMD in this paper. In this study, I look at the genetic basis of variations in population CMD burden and associated health criteria, which may be attributed to ancestry or socially and culturally defined constructs such as race or ethnicity (3 for key terms related to ancestral diversity). I briefly identify additional core specific to ancestral diversity (3) and accept the absence of gold standard science concepts to explain the relevance of ancestral diversity for quantifying the effect of genetic factors on CMD. 36,35 As previously mentioned, I use the word ethnicity to refer to an individual's or population's continental ancestors, as well as, to a lesser degree, the population dynamics within each continent that influenced the observed patterns of genetic variation. We point out that since genetic ancestry is often measured by comparing participants' genotypes to continental reference populations, these reference populations' limited representativeness, availability, and limited sample sizes are significant limitations for the field of genetic epidemiology. 37 Furthermore, marking ancestral groups discretely by continent or other means vastly simplifies genetic variation.

In this study, I present both country-specific and global burden and use the Institute for Health Metrics and Evaluation's Global Burden of Disease (GBD) regions to describe disease burden. I use GBD regions that I expect to have some common ancestral roots to illustrate the strain of CMD in the text for clarification (eg, Western Europe, East Asia, Sub-Saharan Africa). Then I found out GBD regions are made up of countries that, with their recent population histories, might have more ancestral diversity (eg, the United States and Canada, Australia or New Zealand). Nonetheless, I understand that categorizing human populations geographically, such as by nation or area, can oversimplify human genetic diversity inadvertently. Thus, in an effort to further unpack ancestral diversity within a country like the United States, which is the

primary focus of this review, I also refer to common categorizations for US racial/ethnic minorities as proxy groupings of individuals who may have high proportions of non-European ancestry. However, I acknowledge that many ancestrally diverse demographics in the United States, such as racial/ethnic minorities or immigrant communities, may prefer different conceptualizations of race/ethnicity than those commonly used in the world. 46 In the United States, for example, the word 'Hispanic/Latino' is described by the Office of Management and Budget as a mixture of Spanish language usage and Latin American and Caribbean heritage (only countries with Spanish cultural origins). When self-identified US Hispanic/Latinos were asked to mark their race on the 2010 US Census using 5 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 2 or more races (including the 5 US racial categories and some other race), and another 13.0% choose to not respond to the race question, making the nonresponse rate for self-identified non-Hispanic/Latinos 3× higher than for the total US population. 47

Global Populations are Critical for CMD Research

Individual CMD risk, as well as population-level variations in CMD burden seen both across and within countries, can be affected by one's ancestors. Years of life lost due to early death and years of life lost due to injury from a specific illness are accounted for in disability-adjusted life years (DALYs), which are a standard epidemiological indicator of total disease burden. 48 When comparing age-adjusted estimates of DALYs due to ischemic heart disease in 2017, a number of countries in the GBD regions of Oceania, Central Asia, and Eastern Europe have the highest prevalence; in general, males have a higher age-adjusted burden of CMD than females. 49

Between 1990 and 2017, a number of GBD regions (e.g., Oceania, and to a lesser extent, South Asia) had an intractably high burden of ischemic heart disease as measured by DALYs, while others had either steady declines (e.g., Central Europe, North Africa, and the Middle East) or intermittent declines 49 Globally, similar differences in CMD burden can be seen over time in hypertension, ischemic stroke, type 2 diabetes, and chronic kidney disease (CKD; Online I). 50 For example, hypertensive heart disease is most prevalent in Central Sub-Saharan Africa, followed by Oceania, and other African and Middle Eastern regions. 49 Ischemic stroke, on the other hand, is most frequent in Eastern Europe, Oceania, and Central and East Asia. 49

Oceania, Central Latin America, and Mexico, as well as Central and Southern Sub-Saharan Africa, have the largest concentrations of T2D and CKD. 49

Blacks have the highest incidence of hypertension and associated diseases such as coronary artery disease (CAD), ischemic stroke, heart failure, and chronic kidney disease (CKD) in the United States. 51 In particular, hypertension could be responsible for roughly half of the difference in life expectancy between African-Americans and European-Americans. 52 Individuals with Native American, Black, and Hispanic/Latino ancestry have the greatest incidence of adult obesity, T2D, and associated complications, whereas those with European and East Asian ancestry have the lowest. 54,53 There are significant variations in disease

resistance and prevalence even among widely used US race/ethnic groups. For example, adults from Puerto Rico and Mexico are more likely than South Americans to have cardiovascular disease (CVD) risk factors such as obesity,55 and Indian and Filipino Americans are more likely than Chinese Americans to be obese. 56 Asian Indian and Filipino Americans have the highest prevalence of diagnosed T2D among Asians (13 percent and 10%, respectively), while Mexican Americans and Puerto Ricans have the highest prevalence of diagnosed T2D of any Hispanic/Latino community (14 percent and 12 percent, respectively). 54 Despite the fact that diet, cultural expectations, healthcare availability, psychosocial, and socioeconomic stressors are undeniably significant contributors to the unequal disease burden among ancestrally diverse communities, some of these health inequalities exist well after accounting for discrepancies in disease social and environmental exposures. 57–60 This finding also indicates that genetic factors, which may be ancestry-specific or have complex associations with environmental factors that are patterned through racial/ethnic groups, may affect any vulnerability to CMD-related traits or diseases. 60

Given the varying prevalence of CMD around the world and within countries with different communities, attempts to increase the variety of populations surveyed in genetic science have become critical for clinical research and public health. The more inclusive genomic studies are, the more successful they will be at extending the reach of understood human genomic heterogeneity and bolstering our knowledge of disease etiology, allowing us to enhance global and local public health.

Diverse Studies are Important for Evaluating Differential Allele Frequencies

Natural selection and genetic drift are examples of nongenetic and genetic factors that cause human variation. The Out of Africa influx of anatomically modern humans, for example, has had a significant impact on existing human population variation. A demographic bottleneck can be seen in the migration of relatively small populations out of Africa over time, with the groups that migrated into Europe and Asia representing only a fraction of the genetic diversity found on the entire continent of Africa. 27 The 1000 Genomes Project,61 H3 Africa62, and other attempts to classify global human genetic variation have revealed variations in allele frequencies between people of distinct continental ancestries. 61, 63, and 64 These variations vary depending on the evolutionary age of the derived variant and the population's historical background. Previously, population allele frequency variations were due to genetic factors such as natural selection. However, I now have proof that the out of Africa bottleneck caused widespread allele frequency differences. 65 Indeed, the majority of genetic variations, including those unique to a particular continental population, are uncommon and display allele frequency variations (or are population specific). 66,61 For example, the discovery of PCSK9 loss-of-function variants in Blacks contributed to the creation of new therapies to treat high LDL, among other things (see Importance of Variants Specific to a Population below).

Medical genomics is rapidly decoding patients' entire exomes and genomes to find disease resistance variants. However, owing to a lack of consensus on variant annotation, identifying disease-relevant sequence variants has proved difficult. Allele frequency estimates are one aspect that affects variant annotation. As a result, the implementation of mainstream methods

for exchanging genomic and phenotypic data generated by physicians, academics, and patients via unified databases, such as ClinVar67 and the University of Chicago's Geography of Genome Variation browser, has become a top priority for investigators. 68 The Clinical Genome Resource (ClinGen) Variant Curation Interface, for example, is a curated resource for clinicians and researchers that pulls frequency data from a variety of sequencing efforts, such as gnomAD69 (https://gnomad.broadinstitute.org/), PAGE,5 1000 Genomes Project (https://www.internationalgenome.org), and the Exome Sequencing Project70 (ESP; https://evs.gs.washingto.edu 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), and the UK10K Study74 (https://www.ncbi.nlm.nih.gov/bioproject/PRJEB7218), these frequencies are now available on the National Center for Biotechnology Information (NCBI) Regeneron's DRIFT Consortium76 (https://www.regeneron.com/sites/all/themes/regeneron corporate/files/science/DRIFT-Consortium-Factsheet-Backgrounder-July-FINAL.pdf) and 23andMe's Populations Partnerships Program for genotype data77 (https://research.23andme.com/populations-collaborations/) are two examples of industry initiatives However, these last two data sets are not actually widely accessible, making them ineffective for risk variant adjudication based on population frequencies in the scientific and clinical populations.

Clinical laboratories marking putatively deleterious nonsynonymous calls as variants of uncertain meaning, a phenomena that occurs at higher rates in individuals of non-European origin, particularly when these variants have been studied and characterized less regularly, adds to the difficulties of evaluating the pathogenicity of an uncommon variant. 78 Alternatively, a reclassification of putatively causal pathogenic forms for hypertrophic cardiomyopathy, which were subsequently determined to be safe as a consequence of becoming over-represented among Blacks, has enhanced scientific awareness. 79 For-profit businesses are now getting into the variant reclassification game. Blueprint Genetics, for example, has a variant classification service80 that allows the entire sequencing data from a previous exome to be re-analyzed to look for additional clinically significant variants that could clarify or lead to a patient's diagnosis.

SNP ascertainment bias in genotype array evidence has also been caused by differences in allele frequencies across global populations. 81,82 Genotype arrays, especially older ones (e.g.,

Affymetrix 5.0, Illumina Goldengate), were produced using European ancestry sequence data,83 which may have led to the observed skewed range of allele frequencies in non-European GWAS. When GWAS outcomes are merged to produce PRS (also known as genetic risk scores), and are also being used to generate customized CMD risk assessments of both clinical prognosis and personalized intervention/treatment programs, this is becoming a significant stumbling block. 84 Developing a good PRS requires optimizing the proportion of total variation described by a series of known genetic variants for a specific phenotype. When measuring the proportion of variation explained in complex characteristics, it has become common practice in analysis to include all calculated variations (many of which are correlated),

as this improves prediction precision. 85,84 SNP ascertainment bias will result in a model with dramatically different risk estimates across ancestries, as well as low prediction accuracy. Furthermore, recent research has shown that a PRS determined using standard methods in one population can result in unexpected biases in the distributions of scores in other populations, with trends varying significantly across traits. 86 This indicates that many causal variations, particularly in non-European ancestries, are still unknown. The only rational use of PRS is one that guarantees that scores can be measured correctly for all, which implies that the genomic data used be completely representative of all human genetic variation. Any genetically informed personalized medicine strategy that fails to account for this runs the risk of misinterpreting the results.

The Value of Population-Specific Variants

Due to the historical over-representation of European ancestry individuals in current GWAS, I have only recently begun to recognise correlations that are rare in European ancestry populations but normal in others, owing to nongenetic or genetic factors that change allele frequencies and LD trends across populations. I illustrate and outline a non-exhaustive list of examples of genetic variant correlations for ischemic and hypertensive heart disease, stroke, T2D, and CKD in 2. The generalizability of predominantly European ancestors discovered variants in ancestrally diverse populations is then identified. While there are few genomic studies in ancestrally diverse populations, a number of notable major genomic studies and consortia have been established to advance the state of genetic science in these populations. The examples below aren't intended to be exhaustive of all ancestrally diverse genomic studies of CMD and associated traits, but rather to highlight the scope of exploration that can be made in such studies.

Stroke and Ischemic and Hypertensive Heart Disease

As previously mentioned, the prevalence of DALYs caused by ischemic or hypertensive heart disease is highest in Oceania, and alarmingly high in Central Asia and Eastern Europe (ischemic heart disease), as well as many African areas (hypertensive heart disease). 49 Ischemic stroke, on the other hand, is most frequent in Eastern Europe, Oceania, and Central Asia. 49 Most global areas saw a general decrease in both ischemic and hypertensive heart disease between 1990 and 2017, but the particular trajectories varied greatly. These population-level variations may be attributed to differences in genetic heritage and variance in plasma lipid levels, asthma, or other CMD-related characteristics. A few illustrations are given below.

Plasma Lipid Levels and PCSK9, CD36, and APOC3 Ancestry-specific variants associated with blood lipid levels were first described in 2005 with the seminal Dallas Heart Study12 sequencing of African ancestry participants,12 identifying multiple loss-of-function PCSK9 variants (e.g., rs28362286, MAF1 percent, and rs67608943, MAF0.3 percent) that were associated with a 40% lower LDL cholesterol 12 Both forms are present at lower frequencies in gnomeAD samples of African ancestors (2). PCSK9 was discovered to be an autosomal dominant hypercholesterolemia gene (gain-of-function mutations) at the time, but the discovery of PCSK9 loss-of-function variants, which resulted in significant reductions in CAD risk, aided in the creation of PCSK9 inhibitors. 87

A loss-of-function variant in CD36 (rs3211938, 2), which is only present in African ancestry populations (MAF=9%) and is under selective pressure due to malaria, is another plasma lipid example. 88 Higher HDL (high-density lipoprotein)89–91 and lower triglycerides are linked to the variant, as are platelet traits,92 red cell distribution width,93 C-reactive protein,90 and other CMD-related steps.

Furthermore, in the Lancaster Old Order Amish, carriers of a triglyceride-lowering null mutation (rs10892151, MAF5 percent) in APOC3 are normal, allowing for the detection of APOC3 loss-of-function as cardioprotective. 94 Wide meta-analyses for rs76353203 have since validated this finding (2). GWAS with Blood Pressure Traits 95,96

The genetic etiology of systolic blood pressure, diastolic blood pressure, and 24 other complex traits was explored in a recent GWAS of a randomized, ancestrally diverse sample as part of the PAGE (Population Design using Genomics and Epidemiology) review. 5 They discovered a novel systolic blood pressure variant at GYPC (rs28515082, 2) that was most common in Native and Hispanic/Latino Americans (MAF=10% –13%) in their study, but was especially uncommon in East and South Asian populations (MAF0.5%). While this mutation is widespread in European descent groups (MAF=16%), it was first discovered in a complex genetic population in connection with blood pressure. They also found a new systolic blood pressure signal at GPR20 (rs111409240, 2) that is distinct from the previously identified European signal97 (rs34591516). The variant that leads to this novel secondary signal is normal in blacks (MAF=20%), uncommon in European descent individuals (MAF1%), and low frequency in the other diverse populations studied (MAF6%). Other major transethnic electronic health record surveys and meta-analyses of ancestrally representative populations have also shown that diversity adds importance to blood pressure variance studies. JRKL and Hypertension (98,99)

In the PAGE analysis, an indel (rs145054295) in JRKL was linked to hypertension for the first time (=0.43; P=3.70109), and impact sizes at this site were found to vary across ancestries (P=0.025). 5 While the minor allele was observed at 2.4 percent frequency in Blacks (=0.36; P=1.96105), 0.4 percent in Hispanic/Latinos (=0.52; P=5.08103), and 0.5 percent in Native Americans (=1.82; P=0.058) in the PAGE research sample, the mutation was monomorphic in European populations61. In PAGE, the variant is seen at much lower frequencies in East Asians (MAF=0.01%; =2.39; P=0.32) and Native Hawaiians (MAF=0.01%; =14.90; P=0.08). These variations in effect sizes by ancestry are most likely due to allele frequency differences (as seen in 2), emphasizing the significance of testing diverse populations for trait mapping and at scale, as when CMD-relevant variants are this uncommon, massive sampling sizes are needed to reliably estimate effect sizes.

PRKCH and Stroke

In Japanese 100–102 and Chinese communities, a missense variant (rs2230500) in PRKCH has been related to an elevated risk of ischemic stroke. 102nd Individuals with the GA or AA genotype had a 34% higher chance of experiencing ischemic stroke than those with the GG

genotype, according to a meta-analysis of 5 sample populations made up of Chinese and Japanese citizens (3686 cases and 4589 controls). 102nd The SNP is typical in Asian populations (Japanese, MAF=24%; Han Chinese, MAF=18%), but uncommon in European and African descent populations (MAF1%, 2). PRKCH is a protein kinase C (PKC) family member that plays a role in the formation and progression of atherosclerosis in humans. a hundred The A allele allows Val 374 Ile to be substituted in the ATP-binding site of PKC-delta, resulting in increased PKC activity. 102

Stroke and the Sickle Cell Trait

Some studies have linked the sickle cell phenotype (individuals with only one copy of the sickle cell variant at rs334, which is more common in people of African ancestry; 2) to an increased risk of stroke,104 but a recent meta-analysis has refuted this correlation. 105 The thrombosis and hemostasis biomarker D-dimer have shown more consistent correlations. 106–109 Venous thromboembolism (particularly pulmonary embolism) has also been related to sickle cell trait 110,111; further research is required to explain these correlations in larger sample sizes and to elucidate the mechanisms behind them.

Diabetes Mellitus Type 2

DALYs for T2D have traditionally been more intractably elevated between 1990 and 2017, in contrast to global trends in heart disease burden. For eg, Oceania has the highest T2D burden, which increased and then plateaued between 1990 and 2017. Southern Sub-Saharan Africa, Central Latin America, and Mexico all have high T2D burdens. 49 Any of these global variations in vulnerability may be clarified by hereditary differences due to glycemic control, obesity, or other CMD-related characteristics.

T2D and SLC16A11

T2D-related genetic variations with population frequency variations have also been discovered in non-European ancestral groups. Several admixed Mexican ancestry groups, for example, bear a T2D vulnerability haplotype with four amino acid substitutions in SLC16A11; subsequent lookups of this haplotype showed a 50% prevalence in Native Americans and 10% in East Asian study participants, but was uncommon in European and African study participants. 11 This haplotype is responsible for around 20% of the increased prevalence of T2D in Mexico, and expression of SLC16A11 in heterologous cells (nonhuman cells that do not normally express this gene) changes lipid metabolism, resulting in an increase in intracellular triacylglycerol levels. No. 112 About the fact that GWAS has previously identified SLC16A11 as a novel finding likely involved in triacylglycerol metabolism, this study in Mexican ancestry individuals identified SLC16A11 as a novel finding possibly involved in triacylglycerol metabolism.

Obesity and T2D in Founder Groups

In a founder population of 2575 Greenlandic Inuit, a nonsense mutation in TBC1D4 was related to a significant increase in T2D risk (odds ratio=10.3 in homozygous carriers) as well as a large decrease in glucose uptake in muscle (rs61736969, MAF=17 percent; 2). 112

Another recent GWAS of a Greenlandic Inuit population found a strong impact size variant for height and weight in the FADS gene cluster (rs7115739), which is extremely prevalent in the Inuit114, perhaps as an adaptation mechanism to a polyunsaturated fatty acid-rich diet. Following more study, it was discovered that this variant has an impact on height in people of European descent. The impact sizes for the weight finding varied between Greenlandic and a prior European ancestry GWAS115, owing to its lower prevalence in Europeans (MAF5%; 2).

T2D and Pancreatic Acinar Function

Three variants were shown to be in intermediate LD ($r2 > 0.6$) of previously unreported missense variants (p.Val282Met in GP2 [rs78193826], p.Ala341Thr in CPA1 [rs77792157], and p.Arg131Gln in GLP1R [rs3765467]) in a meta-analysis of four GWAS of T2D in people of Japanese ethnicity (2). These mutations are in genes historically connected to pancreatic acinar cell activity (e.g., CPA1 and GP2) and insulin secretion (e.g., CPA1 and GP2) (eg, GLP1R). 116th Another coding variant in PAX4 (Arg192His, rs2233580), a significant transcription factor for islet function, was previously found to be more prevalent in East Asian populations (MAF=11 percent; 2) than any other ancestral group and was related to T2D. 116

T2D, CREBRF, and Obesity

A large-effect BMI-increasing missense variant in CREBRF (rs373863828) has also been discovered, which is normal in Samoan populations (MAF 25%) but uncommon in other global populations outside of Oceania (2). 118,119 individuals The variant is associated with a lower risk of T2D and has a greater effect size (=1.36 kg/m2) than many typical BMI loci, like FTO (rs1558902, the largest effect size variant recorded in European GWAS120 [=0.39 kg/m2]). This variant tends to increase BMI while decreasing T2D risk, in contrast to the majority of previous GWAS findings for BMI and T2D. In adipocyte cell models, functional analyses show that the variant can reduce energy consumption while increasing fat storage. Because of its large impact size and widespread allele occurrence in Samoans, this variant was detected in a discovery and replication study of 5000 people, which is far smaller than the typical GWAS sample size.

T2D and HbA1c

The association between variants of different allele frequencies across global populations and the precision of HbA1c as a test of long-term glycemic regulation is also becoming more well-known.

122 Sickle cell trait (2) has been shown to lower HbA1c as compared to fasting glucose levels in recent studies using assays that are resistant to previously reported assay interference effects. 122 However, these results may be assay based, as the sickle cell phenotype has been linked to higher HbA1c in the Diabetes Prevention Program. number 123 As seen in 2, missense variants at the G6PD locus in African ancestry populations can affect the accuracy of HbA1c as a glycemic control test (rs1050828, MAF=12 percent) and uncommon (rs76723693, MAF=0.5 percent). 124,125 The geographic spread of G6PD deficiency, including sickle cell trait, resembles the geographic distribution of malaria endemic regions,126 emphasizing the need for further research into HbA1c precision in other populations vulnerable to endemic malaria,

such as Southeast Asia. Recent research suggests that alpha thalassemia (based on carrier status of the copy number variant esv2676630) is linked to higher HbA1c127 levels and may statistically interfere with sickle cell trait to affect clinical parameters. 127 The therapeutic relevance of what has been overlooked by concentrating GWAS on predominantly European ancestry populations can be seen in the fact that these HbA1c results for common genetic variants come decades after the clinical use of HbA1c as a test of long-term glycemic control121 and an initial round of GWAS analyses for HbA1c129 and other traits.

Chronic Kidney Disease

From 1990 to 2017, the number of DALYs lost to CKD changed dramatically. 49 Oceania leads the burden of CKD, trailed by Central Latin America and Mexico, and Central Sub-Saharan Africa, similar to global patterns in T2D burden. Below are some examples of how global variation in CKD burden can represent genetic ancestry differences.

APOL1 and CKD Blacks are more than twice as likely as European Americans to experience end-stage renal disease; this result led to the discovery of the G1 (rs73885319) and G2 (rs71785313) risk variants in APOL1, which are more prevalent in people of African origin (2), likely due to selective pressure from African trypanosomiasis (2) ,10. 132 In carriers of two risk alleles, these variants have a major effect on risk, with recorded odds ratios of 17 for focal segmental glomerulosclerosis, 29 for HIV-associated nephropathy,131 and at least a 15% lifetime risk of CKD in risk allele carriers. The number 132 Other African-admixed populations, such as Hispanic/Latinos, have been linked to faster disease development through these variants. 132 Since the related variants had reached higher allele frequency in that population (rs73885319 MAF=23 percent, rs71785313 MAF=13 percent), higher power for a given sample size was observed first in Blacks. Given the ethnic diversity of Hispanic/Latinos, it's not shocking that the connection was repeated in some Hispanic/Latino backgrounds (e.g., some with a higher proportion of African ancestry, such as those from the Caribbean), but not in others.

CKD and Sickle Cell Trait

Small studies134 that suggested sickle cell trait differential vulnerability to CKD have since been validated by larger cohort studies. 111,135 people In reality, sickle cell phenotype can have a similar influence on 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). The number is 136.

In conclusion

More genetic discovery studies in ancestrally diverse and admixed populations are needed to recognize novel susceptibility variants that may be uncommon or missing in GWAS of European descent populations, as seen in the examples above. Furthermore, there is growing concern that observations in one ancestral group do not have the same magnitude of impact in other ancestries. The PAGE Study investigated this topic for many CMD traits; in a landmark paper, they compared the direction and extent of results for GWAS-identified variants in several non-European ancestry populations to European ancestry findings for BMI, T2D, and

lipid levels. The researchers discovered a general dilution of impact sizes across ancestries. 86 The PAGE thesis looked for evidence of smaller impact sizes in previous GWAS results for 26 characteristics linked to CMD and other complex diseases in a diverse population. 5 This experiment shows that previously published GWAS observations resulting from mainly European ancestry samples have substantially decreased impact sizes in other ancestral groups on average. The correlation between historically recorded GWAS impact sizes and results observed in Hispanic/Latinos was 0.86 (95 percent CI, 0.83–0.90) and 0.54 (0.50–0.58) among Blacks, respectively. 5 The smaller effect sizes found in non-European populations could be due to differences in LD, allelic variability, gene-gene, or gene-environment interactions, which could enlarge the distance between the effects of established GWAS results on CMD even further. Regardless of the source of the differential results, any genetic risk prediction model focused on SNP interaction observations outside the ancestry population in which they were found should be used with caution.

For fine-mapping, diversity is critical.

The next big obstacles for genetic epidemiology are determining the underlying causal variants and target genes, as well as transforming these results into therapeutic observations, following the initial progress of GWAS in identifying genomic regions that are robustly linked with a wide variety of diseases and related characteristics. About 90% of the tens of thousands of genomic regions related to diverse traits are noncoding, potentially regulatory regions of the genome. 137 GWASs are useful for recognizing genomic regions linked to a certain phenotype, but they often fail to identify the causal allele or even the involved gene. Indeed, analyses of a single ancestral group and their LD signature restrict the capacity or power of investigators to distinguish causal variants. The number is 138.

Because of their complex LD composition, GWAS of multiple diverse populations (ie, transethnic meta-analysis or pooled analysis of multiple ancestral groups) may help narrow the credible range of causal variants at a given locus. African heritage populations, for example, have shorter LD blocks on average than European populations, which has been shown to be especially useful for narrowing the number of candidate causal variants at a given locus and prioritizing candidate variants for practical follow-up. A total of 24,139 Transethnic fine-mapping has helped narrow lists of candidate variants for kidney function,140 QT interval,141 lipid traits,142 and BMI in several recent CMD studies. 142 Integration of transethnic fine-mapping studies with functional annotation can help with variant selection for in vitro research of variations in transcriptional activity and protein binding144 and can contribute to the discovery of putative causal variants at a previously identified GWAS locus.

In risk prediction and precision medicine, diversity is critical.

As previously mentioned, PRSs were widely used to aggregate impact sizes around the genome in order to approximate the total contribution of genetics to phenotypic heterogeneity. In practice, a PRS is calculated based on existing GWAS discovery findings for each genotyped organism in a target (testing) study (training sample). PRSs have been used to forecast individualized vulnerability and to estimate the effect of a series of novel GWAS findings on external cohorts. a total of 84,145 The distribution of risk scores is often divided into

percentiles or other categorizations to equate the risk of an occurrence for a person to that of those in the same sample in order to draw inferences. However, since the bulk of the training GWAS data was derived from European ancestry samples,146 PRS estimation and model fit will vary greatly across populations due to variations in allele frequencies, impact sizes, or the underlying etiology of the phenotype. 147,455 In many cases, attempting to integrate European-derived PRSs into different cultures has resulted in poor model fit.145-150. I provide explanations of such prejudices from the published literature as well as from our own studies.

CAD PRS in European Ancestry Populations Research has shown that mixing genetic risk ratings, in the form of PRS, with more conventional risk factor tests, such as the Framingham Risk Score, will better predict poor cardiovascular outcomes. One of the first PRS, from the Myocardial Infarction-Genes randomized placebo-controlled clinical trial, found that patients who received both PRS (based on 11 SNPs correlated with CAD) and Framingham 10-year CVD risk score information had lower LDL cholesterol levels and higher statin utilization than those who only received Framingham Risk Scores. 150, 151 The GeneRisk research in Finland found that delivering customized cardiovascular disease risk information based on a mixture of conventional risk data and PRS encouraged healthier behaviour. 152 Physicians at Massachusetts General Hospital are introducing a Preventive Genomics Clinic to help patients understand their monogenic and polygenic risk for a range of diseases and take measures to reduce the risk. 153 The goal is for this clinic to act as a blueprint for how individuals can obtain a low-cost report on monogenic and polygenic risk and use the knowledge to develop prevention strategies. However, if PRSs are produced in a single population, they will invariably lack essential genetic variants that lead to risk in other populations.

Obesity PRS

The PAGE research analyzed the performance of a recently released PRS for obesity154 in the various populations of the PAGE study to explain the variations in PRS performance as a result of population architecture and epidemiology. When comparing PRS to BMI, model fit decreased as genetic distance from European populations increased. The changed R2 for the four PAGE populations (Hispanic/Latino, N=19 028, black, N=16 093, Asian, N=4155 [88 percent Japanese, 5 percent Filipino, 4 percent Chinese, 1% South Asian]; and Native Hawaiian, N=2502) ranged from 2.7 percent (Hispanic/Latino) to 0.3 percent (Native Hawaiian) (Native Hawaiian). With a modified R2 of 1.7 percent, the Asian community scored poorly as well. When it came to forecasting obesity (BMI 30 kg/m2), however, the Asian participants had the better model fit (though it was still poor), with a region under the curve of 0.587. This seems to be due to the inherent BMI and obesity spread within the classes. Asians had the lowest obesity rate for PAGE users, at 11.4 percent, compared to 44.2 percent for Blacks, 40.4 percent for Hispanic/Latinos, and 35.5 percent for Native Hawaiians, representing obesity prevalence in the general US population. As a result, the risk score distribution was best able to identify the few Asians in the upper percentiles who were at high risk of obesity. The PRS, on the other hand, did not reliably differentiate these risk strata even with a higher modified R2 in Hispanic/Latinos, Blacks, and Native Hawaiians since such a significant proportion of participants were obese. Because of heterogeneity in impact sizes (often attributable to differing LD and allele frequencies) and discrepancies in the frequency of a

phenotype, which all combine to hinder the translation of PRS to other populations, this exemplifies the intersection of model fit.

The predictive importance of PRSs is calculated by the related characteristics of the aim (testing) dataset as well as the statistical strength of the discovery (training) dataset—specifically, the enrichment of the genome-wide distribution of interaction test statistics due to aggregate, additive genetic impact. PRSs have been produced so far using publicly accessible GWAS as training results, which have far greater sampling sizes in Europeans than in any other population. 155 PRSs derived from these discoveries may or may not be transferable to other ancestral populations. 156 In fact, PRS precision is a function of recent human demographic background, because in target groups that are genetically more similar to the population observed in the discovery of GWAS, the PRS can describe a greater proportion of phenotypic variation. The polygenic predictive value of PRS decreases as the genetic gap between two populations grows. A practical concern is how to establish polygenic scores for newly admixed individuals or individuals that are genetically distinct from the CMD populations represented by the largest GWAS currently available. Using transethnic data to calculate sufficient weights for more ancestrally diverse samples may improve prediction accuracy.157 MultiPred is a methodological technique in which PRSs based on European training data are combined with PRSs based on training data from other target populations. 155 The formulation of current approaches is based on best practices for dealing with allele frequency and LD variations within and across populations. Given the difficulties in measuring and evaluating PRS across cultures, understanding variations in PRSs across ancestries should be approached with care.

The Benefits of Increasing Diversity in CMD Research in the Future

The PAGE Study and the authors of this paper have been involved in a number of projects aimed at improving the quality of PRS and promoting the characterization of genomic heterogeneity in ancestrally diverse communities (e.g., the Hispanic Community Health Study/Study of Latinos158). Other critical initiatives are now underway to introduce the potential of precision medicine to people of various ethnic backgrounds. MESA (Multi-Ethnic Study of Atherosclerosis) is a prospective study that was intended to determine subclinical CVD and progression to event CVD in a representative population-based sample. Between 2000 and 2002, MESA hired 6814 participants across four race/ethnic groups (European [39 percent], blacks [28 percent], Hispanic/Latino [22 percent], and Chinese American [12 percent]) at eight recruiting centres. 164 The study's main findings include the predictive power of coronary artery calcification for coronary events through ancestry,162 the association of air pollution with coronary artery calcification progression,163, 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 a team of researchers. 165 MESA and PAGE have been pioneers in the production of multi-omics evidence (such as gene expression and methylation) for multi-ethnic communities, and MESA has been a leader in joint initiatives in CVD genetic epidemiology, such as the National Heart, Lung, and Blood Institutes' Trans-Omics for Precision Medicine program75 and the Cohorts for Heart and Aging Studies in Genetic Epidemiology,166. 167

More attempts are being made to attract new cohorts (https://www.theruralstudy.org/about/) 168 and biobank studies 169–172 for a more representative portrayal of US and global communities. The US National Institutes of Health is developing a large-scale biomedical data resource with the aim of representing the diversity of the US population, and it is sponsored by the US National Institutes of Health. 172 A massive and ethnically diverse cohort of US Veterans is being recruited for the Million Veteran's Program. 173 Color, a population genetics firm, recently unveiled plans to recruit 100,000 volunteers from underrepresented populations in order to help determine the risk of myocardial infarction associated with low coverage WGS. As previously stated64, studying African descent populations is especially important because early human migrations out of Africa (both forced and voluntary) brought a portion of genetic diversity into Europe, East Asia, and, eventually, the Americas.As a result, major genetic studies of African descent populations are expected to increase PRS accuracy around the board, as well as precision medicine's potential to target those with the greatest CMD load.

Whole genome sequencing is becoming possible for many thousands of people around the world as the cost of sequencing the human genome continues to fall. offers a nonexhaustive list of global genome sequencing initiatives. Importantly, the ancestry inequalities that plague GWAS arrays will no longer be an issue for sequencing data, and significant work should be put into putting these data together for new discoveries and public health applications. Many countries (e.g., Australia, China, Dubai, Denmark, Estonia, France, Japan, Qatar, Saudi Arabia, Singapore, Turkey) and continental initiatives (e.g., H3Africa, Genome Asia 100K) are still working hard to expand the variety of available genome sequence data globally. Of course, achieving the targets of coordinated exploration and data exchange would undoubtedly take several years, similar to the GWAS that came before them. Furthermore, it will require collective efforts and vigilance on the part of researchers and funders to prioritize these opportunities to ensure that they are used to their maximum potential in order to reduce prejudice in genomic studies and improve global human health. 175

Genetics Research's Ethical, Legal, and Social Consequences

In addition to the broader ethical, legal, and social ramifications of genetic testing, there are particular ethical, legal, and social issues about research in underrepresented communities (see Brothers and Rothstein176 for a comprehensive review). According to these writers, the growth of genomics-enriched health data and the potential for tailored interventions to intensify health inequalities are important concerns that must be tackled. Furthermore, it is important that the problems of anonymity, possible harassment, and improvements in physician-patient relationships and liability be prioritized in these consultations and subsequent policy decisions. It is argued that the current scarcity of genotype-phenotype information, rising prices, and reduced access to health services, as well as information technologies, are all potential causes of health inequality.

As previously stated, researching a variety of populations is important for science understanding as well as increased diversity and inclusion in the field. To prevent essentializing ethnicity and bigotry and hindering interaction, researchers must carefully analyze how they identify and address individual communities. In the absence of pharmacogenomic

heterogeneity to endorse race-specific ads, the US Food and Drug Administration licensed BiDil to treat heart disease in blacks only,177 in a recent example of the essentialization of race in cardiovascular medicine. 42 This decision was widely condemned, because the FDA's authoritative status may have given the biological reification of race the appearance of authority, despite the fact that the race-specific approval has no biological justification.178 A second example is the failure to disclose that people with certain CYP2C19 alleles do not react similarly to Plavix (clopidogrel bisulfate), a platelet aggregation inhibitor, since their hepatic ability to transform it into its active metabolite is limited. 179 In 2014 and 2016, the states of Hawaii and New Mexico brought civil suits against Bristol-Myers for improperly obtaining money from purchases within their territories. East Asians, Native Hawaiians, Pacific Islanders, Native Americans, and Hispanic/Latinos have a higher prevalence of the alleles in question [CYP2C19 * 2, rs4244285 (c.681G > A) and CYP2C19 * 3, rs4986893 (c.636G > A)] than European descent groups. Furthermore, there is growing concern that race-specific medication production, labeling, or promotion could intensify racial/ethnic health inequalities by raising drug prices and rising the need for evidence to justify effectiveness and need in under-represented communities. 42

Some social scientists have also voiced reservations about etiologic science, which looks at the hereditary causes of disease and how they vary by race or ethnicity as markers for shared ancestral history. They are afraid that the population will be driven to conclude that ethnic groups are biologically distinct categories as a result of this study. If genetic variations play a role in disease risk differentials, the public can assume that this holds true for other human traits as well. 181 Phelan et al181 used a nationally representative poll to test this theory. Respondents were randomly assigned to one of four circumstances, each of which consisted of reading a different vignette accompanied by a test of beliefs in biologically important racial distinctions. (1) the Backdoor vignette, which describes a genetic mutation that is more closely correlated with heart disease in Blacks than in European Americans, (2) a vignette describing

race as completely socially formed, (3) a vignette describing ethnic groups as narrowly genetically distinct from one another, and (4) a no-vignette control group Individuals assigned to the Backdoor vignette were more likely to support important racial distinctions than those assigned to the race as a social construction vignette or the nonvignette control category. Both groups assigned the Backdoor vignette and race as basic biological category vignettes, and there was no difference in support for essential racial distinctions.183 According to these results, researchers should expect misinterpretations of findings and specifically clarify what conclusions can and can not be derived from them.

It is becoming increasingly apparent that most genetic research's fundamental knowledge base may not adequately reflect the same groups that have traditionally been underrepresented in genomic research—and that both the amount and accuracy of accessible data can be problems. 182 Potential study subjects' skepticism and mistrust of the research enterprise 183,184 has been proposed as one reason for the prevalence of genomics-related data access inequalities. This is an important concern because the advancement of personalized health care and learning healthcare programs necessitates the incorporation of health and genomics data from

underrepresented communities. The number 182 Implementation science principles are certainly well suited in this domain, with the goal of improving genomics knowledge base participation and equity in CMD and health care in general. 185.

In general, biomedical science has concentrated on the legal aspects of genetic research in terms of acceptable transparency and correct analysis of primary research outcomes and unintended findings. 186 In particular, there is growing uncertainty that using genetic testing to avoid CMD would not result in measurable improvements in either improved enthusiasm for dietary improvement or healthy eating and physical activity habits. The number 187 In the United States, racial/ethnic minorities still face cultural obstacles to healthier lifestyles and unhealthy built environments. 188 As compared to patients with more environmentally driven obesity, elevated PRS estimates for obesity, or a clinician's understanding of obesity, have been found to result in higher-quality patient-provider experiences. Increased use of genetic testing can overemphasize the relative value of biological versus social determinants of health,190 diverting scarce resources away from those without insurance or failing to address the social or systemic determinants of health, which both affect some US groups differently. 188 According to a new study, a popular prediction algorithm used to classify patients in high demand underestimated the probability of black patients relative to European Americans. This was due to the fact that the algorithm used hospital costs as a metric for healthcare needs, and Blacks get fewer services on average. 192 If European Americans use CMD genetic testing more often, their overall healthcare expenses will rise, as will the underestimation of need for Black patients.

CMD genetic research, PRS prediction, and other precision medicine practices focused on genetic data may have various effects on racial/ethnic communities or other marginalized individuals that have historically faced prejudice or stigma.

First, as CMD genetic research is increasingly used to predict risk and classify individuals as high risk, to identify those who may benefit from behavior change or therapeutic interventions192, or to find those who may or may not respond to standard therapies193, such labels may exacerbate the discrimination and stigma already experienced by individuals in underrepresentation. number is 188. Second, the portability of PRS forecasts based on ancestors and environmental exposures is a major concern.147,455 As a consequence, direct-to-consumer prediction algorithms for complex conditions are often marketed as being exclusive to a single race. 192 Furthermore, traditional methodologic approaches to approximate ethnicity, classify ancestral outliers, and/or stratify samples into relatively homogeneous ancestral classes, for ancestrally diverse individuals and populations, can disagree with their own self-identities195 and tend to support the essentialization of racial categories.

When the spectrum and scope of genetic testing grows, new legal concerns can emerge. Currently, healthcare maintenance agencies are working to incorporate genetic screenings into their normal clinical treatment and medical record programs by using the American College of Human Genetics' inventory of 59 genes as actionable incidental discoveries (which includes

many cardiovascular disorders with high penetrance). 194 Routine genetic forecasts for diverse CMD characteristics and disorders are likely to be added in the near future, which may lead to new legal threats to the Genetic Information Nondiscrimination Act in the United States, or comparable laws in other countries197, or be used to change health care rates or restrict access to families of genotyped patients.

There are a variety of specific ethical questions regarding how human genetic prediction using PRS is used outside of clinical studies. 199 Any organization that offers these forecasts directly to customers may provide lower-quality genetic testing and counselling than laboratories that provide clinical treatment. Genomic Prediction, for example, uses a Scientific Laboratory Improvement Amendments-certified laboratory to provide PRSs for embryos before implantation for a range of CMD diseases and traits (e.g., T2D, CAD, myocardial infarction, hypercholesterolemia, and hypertension). 198 The common use of PRS for multifactorial traits and diseases like CMD remains an ethical issue for the field,200 as it risks reinforcing social intolerance of diversity. 191

Conclusion

Understanding how genetic, behavioral, and sociocultural influences work to affect CMD risk differences is critical for individual and public health. Larger, more diverse genetic studies of CMD are urgently needed so that I can close the study gap between the global burden of CMD and the samples and results available in existing genetic tools and databases. CMD genomic testing has a lot of promise in terms of preventing disease and providing personalized medicine. This opportunity, however, can only be realized if CMD research results, like PRS estimation, are obtained from surveys that accurately represent the target population's ancestral diversity.

Inadequate participation is a major disadvantage of modern CMD genetic studies, and it leads to the difficulties of recruiting people from underrepresented communities. Many academics have outlined some key measures for minority groups in the United States and Canada, many of whom face CMD inequalities, to advance their participation in genetic testing. To prevent future injustices around the commercialization of evidence and medical advances, these initiatives include identifying obstacles to access and implementing community-based participatory testing methods and benefits-sharing models. Furthermore, further resources are needed to tackle obstacles to higher education and to promote the employment of people from underrepresented communities in the field of genetics.

The precise detection of genomic regions related to CMD and other health results has opened up new pathways for human and population health improvement. Simultaneously, it has posed additional problems that will have an effect on the utility of this data and its equal inclusion into the wider healthcare system. I consider the following main issues in this review: (1) a lack of diversity in the genotype-phenotype database for CMD; (2) a lack of diversity in genomic comparison panels, which could limit our ability to classify causal variants for CMD; (3) a lack of well defined designations for race/ethnicity and heritage, which complicate studies of disease genetic contributions; (4) disproportionate generation of health-associate Despite these

obstacles, the inclusive production and usage of genomics data in different communities would have unparalleled prospects for reducing the burden of CMD and enhancing health.

Large-scale exchange and harmonization of GWAS data is already happening, as I reported in this analysis with respect to CMD. These large-scale projects, though, are only getting started, and it is up to us as researchers and consumers to focus on data sharing. The topic of data harmonization around self-identified race/ethnicity is related to this and also relevant for GWAS studies. Despite the fact that race and ethnicity are socially and culturally determined constructs, continued use of self-identified categories in genetic studies can catch unexplained phenotypic heterogeneity. To solve this issue, novel machine learning methods have been developed, such as the HARE system developed for the MVP 211,212. The number is 212. HARE uses both self-identified identifiers and scientifically aware ancestry to include ancestry clustering. In a nutshell, HARE uses a combination of self-identified race/ethnicity and controlled genetic clustering (determined using help vector machines) to boost clustering above that obtained solely by self-identified or genetic ancestors. The number is 212. This helps researchers to establish predictors for stable levels of ancestry, allowing for practical mark harmonization even inside biobanks with minimal or conflicting self-identification.

For CMD and other complex diseases, we must make (1) increased data generation and (2) increased data integration of different communities a public health priority. This goal would entail collaborative actions from social, scholarly, technical, and regulatory partners, as well as the communities most impacted by CMD, and must be focused on equality and social justice values. Nanomedicine applying gene therapy strategies 165-176 have elegantly endeavoured for treatment of various gene polymorphisms and mutations of complex disorders 177-191.

References

1. Das S, Abecasis GR, Browning BL. Genotype imputation from large reference panels. Annu Rev Genomics Hum Genet. 2018;19:73–96 doi: 10.1146/annurev-genom-083117-021602

2. Bien SA, Wojcik GL, Zubair N, Gignoux CR, Martin AR, Kocarnik JM, Martin LW, Buyske S, Haessler J, Walker RW, et alPAGE Study. Strategies for enriching variant coverage in candidate disease loci on a multiethnic genotyping array. PLoS One. 2016;11:e0167758 doi: 10.1371/journal.pone.0167758

3. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538:161–164 doi: 10.1038/538161a

4. Mills MC, Rahal C. The GWAS diversity monitor tracks diversity by disease in real time. Nat Genet. 2020;52:242–243 doi: 10.1038/s41588-020-0580-y

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 doi: 10.1038/s41586-019-1310-4

6. Popejoy AB. Diversity in precision medicine and pharmacogenetics: methodological and conceptual considerations for broadening participation. Pharmgenomics Pers Med. 2019;12:257–271 doi: 10.2147/PGPM.S179742

7. Jooma S, Hahn MJ, Hindorff LA, Bonham VL. Defining and achieving health equity in genomic medicine. Ethn Dis. 2019;29:173–178 doi: 10.18865/ed.29.S1.173
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 doi: 10.18865/ed.29.S1.179
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 doi: 10.1038/nrg.2017.89
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 doi: 10.1126/science.1193032
11. Williams AL, Jacobs SBR, Moreno-Macías H, Huerta-Chagoya A, Churchhouse C, Márquez-Luna C, García-Ortíz H, Gómez-Vázquez MJ, Burtt NP, Aguilar-Salinas CA, et alSIGMA Type 2 Diabetes Consortium. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature. 2014;506:97–101 doi: 10.1038/nature12828
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 doi: 10.1038/ng1509
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 doi: 10.1017/S1461145711000903
14. Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, Gaspar H, Ikeda M, Benyamin B, Brown BC, et alSchizophrenia Working Group of the Psychiatric Genomics Consortium; Indonesia Schizophrenia Consortium; Genetic REsearch on schizophreniA neTwork-China and the Netherlands (GREAT-CN). Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet. 2019;51:1670–1678 doi: 10.1038/s41588-019-0512-x
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 [published online October 7, 2019]. Mol Psychiatry.
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 doi: 10.1038/s41380-018-0335-7
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 doi: 10.1093/schbul/sby140

18. Guo Y, Tan LJ, Lei SF, Yang TL, Chen XD, 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 doi: 10.1371/journal.pgen.1000806
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 doi: 10.1016/j.bonr.2016.08.005
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;49:1601505 doi: 10.1183/13993003.01505-2016
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 doi: 10.1016/j.jaci.2018.11.037
22. Daya M, Rafaels N, Brunetti TM, Chavan S, Levin AM, Shetty A, Gignoux CR, Boorgula MP, Wojcik G, Campbell M, et alCAAPA. Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. Nat Commun. 2019;10:880 doi: 10.1038/s41467-019-08469-7
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 doi: 10.1038/nature05329
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 doi: 10.1038/s41576-018-0016-z
25. Chakravarti A. Perspectives on human variation through the lens of diversity and race. Cold Spring Harb Perspect Biol. 2015;7:a023358 doi: 10.1101/cshperspect.a023358
26. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11:499–511 doi: 10.1038/nrg2796
27. Amos W, Hoffman JI. Evidence that two main bottleneck events shaped modern human genetic diversity. Proc Biol Sci. 2010;277:131–137 doi: 10.1098/rspb.2009.1473
28. . What are single nucleotide polymorphisms (SNPs)? - Genetics Home Reference - NIH. 2019 Available at: https://ghr.nlm.nih.gov/primer/genomicresearch/snp. Accessed April 25, 2020
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 doi: 10.1038/nrg2554
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 doi: 10.1038/nmeth0410-248
31. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016;11:1–9 doi: 10.1038/nprot.2015.123
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 doi: 10.1093/bioinformatics/btx536

33. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11:361–362 doi: 10.1038/nmeth.2890
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 doi: 10.1093/nar/gky1016
35. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. Genome Biol. 2002;3:comment2007 doi: 10.1186/gb-2002-3-7-comment2007
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 doi: 10.1056/NEJMsb025007
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 doi: 10.1016/j.ajhg.2010.03.011
38. 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 doi: 10.1177/0306312710379170
39. Serre D, Pääbo S. Evidence for gradients of human genetic diversity within and among continents. Genome Res. 2004;14:1679–1685 doi: 10.1101/gr.2529604
40. Lu YF, Goldstein DB, Angrist M, Cavalleri G. Personalized medicine and human genetic diversity. Cold Spring Harb Perspect Med. 2014;4:a008581 doi: 10.1101/cshperspect.a008581
41. 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 doi: 10.1038/ng1847
42. Bonham VL, Callier SL, Royal CD. Will precision medicine move US beyond race? N Engl J Med. 2016;374:2003–2005 doi: 10.1056/NEJMp1511294
43. 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 doi: 10.1038/s41598-017- 01837-7
44. Williams DR, Priest N, Anderson NB. Understanding associations among race, socioeconomic status, and health: patterns and prospects. Health Psychol. 2016;35:407–411 doi: 10.1037/hea0000242
45. Williams DR, Mohammed SA. Discrimination and racial disparities in health: evidence and needed research. J Behav Med. 2009;32:20–47 doi: 10.1007/s10865-008-9185-0
46. Humes KR, Jones NA, Ramirez RR Overview Of Race And Hispanic Origin: 2010 Census Briefs. 2011 Washington, DC US Census Bureau
47. Ríos M, Romero F, Ramírez R Race Reporting Among Hispanics: 2010. 2013 US Census Bureau, Population Division
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 doi: 10.1146/annurev.publhealth.23.100901.140513

49. Institute for Health Metrics and Evaluation (IHME). . GBD compare data visualization. 2018 Seattle, WA IHME, University of Washington . Available at: http://vizhub.healthdata.org/gbd-compare. Accessed February 20, 2020

50. Institute for Health Metrics and Evaluation (IHME). . GBD results tool. 2015 Seattle, WA IHME, University of Washington . Available at: http://ghdx.healthdata.org/gbd-results-tool. Accessed February 20, 2020

51. Musemwa N, Gadegbeku CA. Hypertension in African Americans. Curr Cardiol Rep. 2017;19:129 doi: 10.1007/s11886-017-0933-z

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 doi: 10.1377/hlthaff.2013.1273

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 doi: 10.1001/jama.2018.7270

54. Centers for Disease Control and Prevention. . National diabetes statistics report, 2017. 2017 Atlanta, GA Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services . Available at: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf

55. Daviglus 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 doi: 10.1016/j.pcad.2014.07.006

56. Mui P, Hill SE, Thorpe RJ Jr.. Overweight and obesity differences across ethnically diverse subgroups of Asian American men. Am J Mens Health. 2018;12:1958–1965 doi: 10.1177/1557988318793259

57. Isasi CR, Parrinello CM, Ayala GX, Delamater AM, Perreira KM, Daviglus 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 doi: 10.1016/j.jpeds.2016.05.037

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 doi: 10.1093/gerona/gly026

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 doi: 10.1136/bmjopen-2016-013221

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–S15 doi: 10.1038/ng1436

61. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR1000 Genomes Project Consortium. . A global reference for human genetic variation. Nature. 2015;526:68–74 doi: 10.1038/nature15393

62. Rotimi C, Abayomi A, Abimiku A, Adabayeri VM, Adebamowo C, Adebiyi E, Ademola AD, Adeyemo A, Adu DH3Africa Consortium. . Enabling the genomic revolution in Africa. Science. 2014;344:1346–1348 doi: 10.1126/science.1251546

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 doi: 10.1126/science.1153717

64. Bentley AR, Callier SL, Rotimi CN. Evaluating the promise of inclusion of African ancestry populations in genomics. NPJ Genom Med. 2020;5:5 doi: 10.1038/s41525-019-0111-x

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 doi: 10.1111/j.1469-1809.2008.00489.x

66. Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Bustamante CD1000 Genomes Project. . Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci U S A. 2011;108:11983–11988 doi: 10.1073/pnas.1019276108

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 doi: 10.1093/nar/gkx1153

68. Marcus JH, Novembre J. Visualizing the geography of genetic variants. Bioinformatics. 2017;33:594–595 doi: 10.1093/bioinformatics/btw643

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 [published online August 13, 2019]. BioRxiv.

70. Exome Variant Server. . NHLBI GO exome sequencing project (ESP). 2020 Seattle, WA . Available at: http://evs.gs.washington.edu/EVS/. Accessed April 25, 2020

71. Kitts A, Sherry SMcEntyre J, Ostell J. The single nucleotide polymorphism database (dbSNP) of nucleotide sequence variation. In: The NCBI Handbook. 2002 Bethesda, MD National Center for Biotechnology Information (US) . Available at: 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 doi: 10.1002/humu.23835

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 doi: 10.1093/ije/dys064

74. Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JRB, Xu C, Futema M, Lawson D, et alUK10K Consortium. The UK10K project identifies rare variants in health and disease. Nature. 2015;526:82–90 doi: 10.1038/nature14962

75. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Gagliano Taliun SA, Corvelo A, Gogarten SM, et al Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program [published online March 6, 2019]. BioRxiv.
76. Shuldiner AR. The DRIFT Consortium: discovery research investigating founder population traits. 2020 Regeneron Genetics Center . Available at: https://www.regeneron.com/sites/all/themes/regeneron_corporate/files/science/DRIFT-Consortium-Factsheet-Backgrounder-July-FINAL.pdf. Accessed April 25, 2020
77. . 23andMe Populations Collaborations Program. 2020 Available at: https://research.23andme.com/populations-collaborations/. Accessed April 25, 2020
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 doi: 10.1038/gim.2017.96
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 doi: 10.1056/NEJMsa1507092
80. Blueprint Genetics. . Whole Exome Sequencing (WES) Variant Re-evaluation and WES Re-analysis Services. 2019 Available at: https://blueprintgenetics.com/wes-re-evaluation-re- analysis/. Accessed April 25, 2020
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 doi: 10.1186/s13059-018-1561-7
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 doi: 10.1002/bies.201300014
83. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA1000 Genomes Project Consortium. . An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65 doi: 10.1038/nature11632
84. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50:1219–1224 doi: 10.1038/s41588-018-0183-z
85. Zhang Y, Qi G, Park JH, 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 doi: 10.1038/s41588-018-0193-x
86. Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, Schumacher FR, Peters U, Franceschini N, Ritchie MD, et alPAGE Consortium. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. PLoS Biol. 2013;11:e1001661 doi: 10.1371/journal.pbio.1001661
87. Sabatine MS. PCSK9 inhibitors: clinical evidence and implementation. Nat Rev Cardiol. 2019;16:155–165 doi: 10.1038/s41569-018-0107-8

88. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. Positive natural selection in the human lineage. Science. 2006;312:1614–1620 doi: 10.1126/science.1124309

89. Elbers CC, Guo Y, Tragante V, van Iperen EP, 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 doi: 10.1371/journal.pone.0050198

90. Ellis J, Lange EM, Li J, Dupuis J, Baumert J, Walston JD, Keating BJ, Durda P, Fox ER, Palmer CD, 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 doi: 10.1007/s00439-014-1439-z

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 doi: 10.1093/hmg/ddn060

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 doi: 10.1016/j.ajhg.2012.08.031

93. Chami N, Chen MH, Slater AJ, Eicher JD, Evangelou E, Tajuddin SM, Love-Gregory L, Kacprowski T, Schick UM, Nomura A, et al Exome genotyping identifies pleiotropic variants associated with red blood cell traits. Am J Hum Genet. 2016;99:8–21 doi: 10.1016/j.ajhg.2016.05.007

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 doi: 10.1126/science.1161524

95. Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjærg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. N Engl J Med. 2014;371:32–41 doi: 10.1056/NEJMoa1308027

96. Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, Lange LA, Lu Y, Tang Z, Zhang H, Hindy G, et alTG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. N Engl J Med. 2014;371:22–31 doi: 10.1056/NEJMoa1307095

97. Surendran P, Drenos F, Young R, Warren H, Cook JP, Manning AK, Grarup N, Sim X, Barnes DR, Witkowska K, et alCHARGE-Heart Failure Consortium; EchoGen Consortium; METASTROKE Consortium; GIANT Consortium; EPIC-InterAct Consortium; Lifelines Cohort Study; Wellcome Trust Case Control Consortium; Understanding Society Scientific Group; EPIC-CVD Consortium; CHARGE+ Exome Chip Blood Pressure Consortium; T2D-GENES Consortium; GoT2DGenes Consortium; ExomeBP Consortium; CHD Exome+ Consortium. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat Genet. 2016;48:1151–1161 doi: 10.1038/ng.3654

98. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, 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 doi: 10.1038/ng.3715

99. Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, Torstenson ES, Kovesdy CP, Sun YV, Wilson OD, et alUnderstanding Society Scientific Group; International Consortium for Blood Pressure; Blood Pressure-International Consortium of Exome Chip Studies; Million Veteran Program. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat Genet. 2019;51:51–62 doi: 10.1038/s41588-018-0303-9

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 doi: 10.1038/ng1945

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 doi: 10.1016/j.atherosclerosis.2007.11.009

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 doi: 10.1016/j.neulet.2011.10.047

103. Matsuo R, Ago T, Hata J, Kuroda J, Wakisaka Y, Sugimori H, Kitazono T, Kamouchi MFSR 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 doi: 10.1016/j.jstrokecerebrovasdis.2013.11.011

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 doi: 10.1161/STROKEAHA.114.006110

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 doi: 10.1001/jamaneurol.2018.0571

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 doi: 10.1182/blood-2016-01-694422

107. Raffield LM, Zakai NA, Duan Q, Laurie C, Smith JD, Irvin MR, Doyle MF, Naik RP, Song C, Manichaikul AW, et alNHLBI 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 doi: 10.1161/ATVBAHA.117.310073

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 doi: 10.1016/j.thromres.2015.08.013

109. Amin C, Adam S, Mooberry MJ, Kutlar A, Kutlar F, Esserman D, Brittain JE, Ataga KI, Chang JY, Wolberg AS, et al Coagulation activation in sickle cell trait: an exploratory study. Br J Haematol. 2015;171:638–646 doi: 10.1111/bjh.13641
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 doi: 10.1186/s12959-018-0179-z
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 doi: 10.1111/jth.12787
112. Rusu V, Hoch E, Mercader JM, Tenen DE, Gymrek M, Hartigan CR, DeRan M, von Grotthuss M, Fontanillas P, Spooner A, et alMEDIA Consortium; SIGMA T2D Consortium. Type 2 diabetes variants disrupt function of SLC16A11 through two distinct mechanisms. Cell. 2017;170:199–212.e20 doi: 10.1016/j.cell.2017.06.011
113. Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliussen TS, Andersen MA, Nielsen TS, Krarup NT, et al A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature. 2014;512:190–193 doi: 10.1038/nature13425
114. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, Korneliussen TS, Gerbault P, Skotte L, Linneberg A, et al Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science. 2015;349:1343–1347 doi: 10.1126/science.aab2319
115. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, et alElectronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46:1173–1186 doi: 10.1038/ng.3097
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 doi: 10.1038/s41588-018-0332-4
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 doi: 10.1038/nature18642
118. Minster RL, Hawley NL, Su CT, 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 doi: 10.1038/ng.3620
119. Hanson RL, Safabakhsh S, Curtis JM, Hsueh WC, Jones LI, Aflague TF, Duenas Sarmiento J, Kumar S, Blackburn NB, Curran JE, et al Association of CREBRF variants with obesity and diabetes in Pacific Islanders from Guam and Saipan. Diabetologia. 2019;62:1647–1652 doi: 10.1007/s00125-019-4932-z
120. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Mägi R, et alMAGIC; Procardis Consortium. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010;42:937–948 doi: 10.1038/ng.686

121. Leong A, Wheeler E. Genetics of HbA1c: a case study in clinical translation. Curr Opin Genet Dev. 2018;50:79–85 doi: 10.1016/j.gde.2018.02.008
122. Lacy ME, Wellenius GA, Sumner AE, Correa A, Carnethon MR, Liem RI, Wilson JG, Sacks DB, Jacobs DR Jr, Carson AP, et al Association of sickle cell trait with hemoglobin A1c in African Americans. JAMA. 2017;317:507–515 doi: 10.1001/jama.2016.21035
123. Hivert MF, Christophi CA, Jablonski KA, Edelstein SL, Kahn SE, Golden SH, Dagogo-Jack S, Mather KJ, Luchsinger JA, Caballero AE, et al 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 doi: 10.1210/jc.2018-01416
124. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, et alEPIC-CVD Consortium; EPIC-InterAct Consortium; Lifelines Cohort Study. 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 doi: 10.1371/journal.pmed.1002383
125. Sarnowski C, Leong A, Raffield LM, Wu P, de Vries PS, DiCorpo D, Guo X, Xu H, Liu Y, Zheng X, et alTOPMed Diabetes Working Group; TOPMed Hematology Working Group; TOPMed Hemostasis Working Group; 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 doi: 10.1016/j.ajhg.2019.08.010
126. Howes RE, Piel FB, Patil AP, Nyangiri OA, Gething PW, Dewi M, Hogg MM, Battle KE, Padilla CD, Baird JK, et al G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. PLoS Med. 2012;9:e1001339 doi: 10.1371/journal.pmed.1001339
127. Jun G, Sedlazeck FJ, Chen H, Yu B, Qi Q, Krasheninina O, Carroll A, Liu X, Mansfield A, Zarate S, et al 2018Identification 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
128. Raffield LM, Ulirsch JC, Naik RP, Lessard S, Handsaker RE, Jain D, Kang HM, Pankratz N, Auer PL, Bao EL, et alNHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Hematology & Hemostasis, Diabetes, and Structural Variation TOPMed Working Groups. Common α-globin variants modify hematologic and other clinical phenotypes in sickle cell trait and disease. PLoS Genet. 2018;14:e1007293 doi: 10.1371/journal.pgen.1007293
129. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, Bouatia-Naji N, Langenberg C, Prokopenko I, Stolerman E, et alWTCCC. Common variants at 10 genomic loci influence hemoglobin A1(C) levels via glycemic and nonglycemic pathways. Diabetes. 2010;59:3229–3239 doi: 10.2337/db10-0502
130. Cooper A, Ilboudo H, Alibu VP, Ravel S, Enyaru J, Weir W, Noyes H, Capewell P, Camara M, Milet J, et al APOL1 renal risk variants have contrasting resistance and susceptibility associations with African trypanosomiasis. elife. 2017;6:e25461 doi: 10.7554/eLife.25461
131. Kopp JB, Nelson GW, Sampath K, Johnson RC, Genovese G, An P, Friedman D, Briggs W, Dart R, Korbet S, et al APOL1 genetic variants in focal segmental

glomerulosclerosis and HIV-associated nephropathy. J Am Soc Nephrol. 2011;22:2129–2137 doi: 10.1681/ASN.2011040388

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 doi: 10.1016/j.semnephrol.2015.04.008

133. Parsa A, Kao WH, Xie D, Astor BC, Li M, Hsu CY, Feldman HI, Parekh RS, Kusek JW, Greene TH, et alAASK Study Investigators; CRIC Study Investigators. APOL1 risk variants, race, and progression of chronic kidney disease. N Engl J Med. 2013;369:2183–2196 doi: 10.1056/NEJMoa1310345

134. Key NS, Derebail VK. Sickle-cell trait: novel clinical significance. Hematology Am Soc Hematol Educ Program. 2010;2010:418–422 doi: 10.1182/asheducation-2010.1.418

135. Naik RP, Derebail VK, Grams ME, Franceschini N, Auer PL, Peloso GM, Young BA, Lettre G, Peralta CA, Katz R, et al Association of sickle cell trait with chronic kidney disease and albuminuria in African Americans. JAMA. 2014;312:2115–2125 doi: 10.1001/jama.2014.15063

136. Naik RP, Irvin MR, Judd S, Gutiérrez OM, Zakai NA, Derebail VK, Peralta C, Lewis MR, Zhi D, Arnett D, et al Sickle cell trait and the risk of ESRD in blacks. J Am Soc Nephrol. 2017;28:2180–2187 doi: 10.1681/ASN.2016101086

137. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337:1190–1195 doi: 10.1126/science.1222794

138. Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, Gupta N, Neale BM, Daly MJ, Sklar P, et al Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. Nat Genet. 2012;44:631–635 doi: 10.1038/ng.2283

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 doi: 10.1146/annurev.genom.9.081307.164258

140. Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, Smith GD, Mahajan A, Gaulton KJ, Nadkarni GN, et al Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. Nat Commun. 2019;10:29 doi: 10.1038/s41467-018-07867-7

141. Avery CL, Wassel CL, Richard MA, Highland HM, Bien S, Zubair N, Soliman EZ, Fornage M, Bielinski SJ, Tao R, et al 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 doi: 10.1016/j.hrthm.2016.12.021

142. Zubair N, Graff M, Luis Ambite J, Bush WS, Kichaev G, Lu Y, Manichaikul A, Sheu WH, Absher D, Assimes TL, et al 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 doi: 10.1093/hmg/ddw358

143. Fernández-Rhodes L, Gong J, Haessler J, Franceschini N, Graff M, Nishimura KK, Wang Y, Highland HM, Yoneyama S, Bush WS, et al 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 doi: 10.1007/s00439-017-1787-6
144. Cannon ME, Duan Q, Wu Y, Zeynalzadeh M, Xu Z, Kangas AJ, Soininen P, Ala-Korpela M, Civelek M, Lusis AJ, et al Trans-ancestry fine mapping and molecular assays identify regulatory variants at the ANGPTL8 HDL-C GWAS Locus. G3 (Bethesda). 2017;7:3217–3227 doi: 10.1534/g3.117.300088
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 doi: 10.1038/s41588-019-0379-x
146. Duncan L, Shen H, Gelaye B, Meijsen 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 doi: 10.1038/s41467-019-11112-0

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 doi: 10.1016/j.ajhg.2017.03.004
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 doi: 10.29245/2578-2967/2018/4.1137
149. Onengut-Gumuscu S, Chen WM, Robertson CC, Bonnie JK, Farber E, Zhu Z, Oksenberg JR, Brant SR, Bridges SL Jr, Edberg JC, et alSEARCH for Diabetes in Youth; Type 1 Diabetes Genetics Consortium. Type 1 diabetes risk in African-ancestry participants and utility of an ancestry-specific genetic risk score. Diabetes Care. 2019;42:406–415 doi: 10.2337/dc18-1727
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 doi: 10.1186/1471-2261-11-66
151. Kullo IJ, Jouni H, Austin EE, Brown SA, Kruisselbrink TM, Isseh IN, Haddad RA, Marroush TS, Shameer K, Olson JE, et al 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 doi: 10.1161/CIRCULATIONAHA.115.020109
152. Snell K, Helén I. 'Well, I knew this already' - explaining personal genetic risk information through narrative meaning-making. Sociol Health Illn. 2020;42:496–509 doi: 10.1111/1467- 9566.13018
153. Illumina. . Polygenic risk: what's the score? Nat Res. 2019 . Available at: 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, et al Polygenic prediction of weight and obesity trajectories from birth to adulthood. Cell. 2019;177:587–596.e9 doi: 10.1016/j.cell.2019.03.028
155. Márquez-Luna C, Loh PR, Price ALSouth Asian Type 2 Diabetes (SAT2D) Consortium; SIGMA Type 2 Diabetes Consortium. . Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol. 2017;41:811–823 doi: 10.1002/gepi.22083

156. Knowles JW, Ashley EA. Cardiovascular disease: the rise of the genetic risk score. PLoS Med. 2018;15:e1002546 doi: 10.1371/journal.pmed.1002546
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 doi: 10.1002/gepi.22166
158. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, Schneiderman N, Raij L, Talavera G, Allison M, et al Design and implementation of the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol. 2010;20:629–641 doi: 10.1016/j.annepidem.2010.03.015

159. Lin DY, Tao R, Kalsbeek WD, Zeng D, Gonzalez F 2nd, 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 doi: 10.1016/j.ajhg.2014.11.005
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 doi: 10.1002/gepi.21896
161. Olson JL, Bild DE, Kronmal RA, Burke GL. Legacy of MESA. Glob Heart. 2016;11:269– 274 doi: 10.1016/j.gheart.2016.08.004
162. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, Liu K, Shea S, Szklo M, Bluemke DA, et al Coronary calcium as a predictor of coronary events in four racial or ethnic groups. N Engl J Med. 2008;358:1336–1345 doi: 10.1056/NEJMoa072100
163. Kaufman JD, Adar SD, Barr RG, Budoff M, Burke GL, Curl CL, Daviglus ML, Diez Roux AV, Gassett AJ, Jacobs DR Jr, et al 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 doi: 10.1016/S0140-6736(16)00378-0
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 doi: 10.1016/j.atherosclerosis.2015.05.006
165. Guan W, Cao J, Steffen BT, Post WS, Stein JH, Tattersall MC, Kaufman JD, McConnell JP, Hoefner DM, Warnick R, et al 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 doi: 10.1161/ATVBAHA.114.304785
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 doi: 10.1097/EDE.0b013e31828b2cbb
167. Liu Y, Ding J, Reynolds LM, Lohman K, Register TC, De La Fuente A, Howard TD, Hawkins GA, Cui W, Morris J, et al Methylomics of gene expression in human monocytes. Hum Mol Genet. 2013;22:5065–5074 doi: 10.1093/hmg/ddt356
168. . About RURAL Cohort Study. 2020 Available at: https://www.theruralstudy.org/about/. Accessed April 25, 2020

169. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, Li LChina 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 doi: 10.1093/ije/dyr120

170. Gan W, Walters RG, Holmes MV, Bragg F, Millwood IY, Banasik K, Chen Y, Du H, Iona A, Mahajan A, et alChina 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 doi: 10.1007/s00125-016-3920-9

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 doi: 10.1093/ije/dyl042

172. University of Oxford. . The Mexico City Prospective Study [Internet]. 2020 UK Research and Innovation . Available at: https://gtr.ukri.org/projects?ref=MC_UU_00017%2F2. Accessed April 25, 2020

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 doi: 10.1038/gim.2016.183

174. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, et al Million Veteran Program: a mega-biobank to study genetic influences on health and disease. J Clin Epidemiol. 2016;70:214–223 doi: 10.1016/j.jclinepi.2015.09.016

175. Stark Z, Dolman L, Manolio TA, Ozenberger B, Hill SL, Caulfied MJ, Levy Y, Glazer D, Wilson J, Lawler M, et al Integrating genomics into healthcare: a global responsibility. Am J Hum Genet. 2019;104:13–20 doi: 10.1016/j.ajhg.2018.11.014

176. Brothers KB, Rothstein MA. Ethical, legal and social implications of incorporating personalized medicine into healthcare. Per Med. 2015;12:43–51 doi: 10.2217/pme.14.65

177. 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 doi: 10.7326/0003-4819-146-1-200701020-00010

178. Kahn J. Misreading race and genomics after BiDil. Nat Genet. 2005;37:655–656 doi: 10.1038/ng0705-655

179. 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 doi: 10.2217/pme.15.4

180. Duster T Backdoor to Eugenics. 2004 Routledge

181. 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 doi: 10.1177/0003122413476034

182. 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 doi: 10.2217/pme-2018-0037

183. Fullerton SMBurke W, Edwards KA, Goering S, Holland S, Trinidad SB. The input-output problem: whose dna do we study, and why does it matter? In: Achieving Justice in

Genomic Translation : Re-Thinking the Pathway to Benefit. 2011 New York, NY Oxford University Press:40–55

184. Wright GE, Koornhof PG, Adeyemo AA, Tiffin N. Ethical and legal implications of whole genome and whole exome sequencing in African populations. BMC Med Ethics. 2013;14:21 doi: 10.1186/1472-6939-14-21

185. 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 doi: 10.1038/gim.2016.210

186. 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 doi: 10.1111/bioe.12285

187. 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 doi: 10.1017/S0007114516002488

188. Mensah GA, Jaquish C, Srinivas P, Papanicolaou GJ, Wei GS, Redmond N, Roberts MC, Nelson C, Aviles-Santa L, Puggal M, et al Emerging concepts in precision medicine and cardiovascular diseases in racial and ethnic minority populations. Circ Res. 2019;125:7–13 doi: 10.1161/CIRCRESAHA.119.314970

189. 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 doi: 10.1002/osp4.293

190. 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 doi: 10.1186/s13010-019-0073-8

191. 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 doi: 10.1126/science.aax2342

192. Gibson G. On the utilization of polygenic risk scores for therapeutic targeting. PLoS Genet. 2019;15:e1008060 doi: 10.1371/journal.pgen.1008060

193. Godard B, Ozdemir V, Fortin M, Egalité N. Ethnocultural community leaders' views and perceptions on biobanks and population specific genomic research: a qualitative research study. Public Underst Sci. 2010;19:469–485 doi: 10.1177/0963662509104721

194. Regalado A. White-people-only DNA tests show how unequal science has become. 2018 MIT Technology Review [Internet]. Available at:

https://www.technologyreview.com/s/612322/white-people-only-dna-tests-show-how-unequal-science-has-become/

195. 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 doi: 10.1186/1471-2156-12-28

196. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, et al 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 doi: 10.1038/gim.2016.190
197. Suter SM. GINA at 10 years: the battle over 'genetic information' continues in court. J Law Biosci. 2018;5:495–526 doi: 10.1093/jlb/lsz002
198. 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 doi: 10.1038/s41598-019-51258-x
199. Genomic Prediction. . Frequently asked questions [Internet]. 2019 Available at: https://genomicprediction.com/faqs/#faq-7.2
200. Karavani E, Zuk O, Zeevi D, Barzilai N, Stefanis NC, Hatzimanolis A, Smyrnis N, Avramopoulos D, Kruglyak L, Atzmon G, et al Screening human embryos for polygenic traits has limited utility. Cell. 2019;179:1424–1435.e8 doi: 10.1016/j.cell.2019.10.033
201. . GIANT consortium [Internet]. 2019 Available at: https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium
202. . Magic Investigators - Home Page [Internet]. 2020 Available at: https://www.magicinvestigators.org/. Accessed April 25, 2020
203. . Global Lipids Genetics Consortium [Internet]. 2020 Available at: http://lipidgenetics.org/. Accessed April 25, 2020
204. 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 doi: 10.1016/j.outlook.2012.10.004
205. 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 doi: 10.1353/cpr.2012.0025
206. University of Illinois. . Summer Internship for Indigenous Peoples in Genomics (SING) [Internet]. 2020 Carl R. Woese Institute for Genomic Biology . Available at: https://sing.igb.illinois.edu/

207. SING Aotearoa. . Summer internship for indigenous genomics aotearoa [Internet]. 2020 Available at: https://www.singaotearoa.nz/. Accessed April 25, 2020
208. SING Australia. . Summer internship for indigenous peoples in genomics [Internet]. 2020 Available at: https://www.singaustralia.org
209. University of Alberta. . Summer internship for indigenous peoples in genomics Canada (SING Canada) [Internet]. 2020 Indigenous STS (Science, Technology, and Society) Available at: https://indigenoussts.com/sing-canada/. Accessed April 25, 2020
210. Wade L. To overcome decades of mistrust, a workshop aims to train indigenous researchers to be their own genome experts. Science. September 27, 2018 https://www.sciencemag.org/news/2018/09/overcome-decades-mistrust-workshop-aims-train-indigenous-researchers-be-their-own. Accessed April 25, 2020
211. 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 doi: 10.1534/g3.118.200925

212. Fang H, Hui Q, Lynch J, Honerlaw J, Assimes TL, Huang J, Vujkovic M, Damrauer SM, Pyarajan S, Gaziano JM, et alVA Million Veteran Program. Harmonizing genetic ancestry and self-identified race/ethnicity in Genome-wide Association Studies. Am J Hum Genet. 2019;105:763–772 doi: 10.1016/j.ajhg.2019.08.012

165. Moataz Dowaidar, Hani Nasser Abdelhamid, Ülo Langel. Improvement of transfection with PepFects using organic and inorganic materials.

166. Tomas Venit, Moataz Dowaidar , Maxime Gestin , Syed Raza Mahmood, Ülo Langel , Piergiorgio Percipalle. Transcriptional profiling reveals ribosome biogenesis, microtubule dynamics and expression of specific LncRNAs to be part of a common response to cell penetrating peptides. Biomolecules, 10(11), pp. 1-21, 1567, 2020, https://doi.org/10.3390/biom10111567.

167. HN Abdelhamid, Moataz Dowaidar, Ü Langel. Carbonized chitosan encapsulated hierarchical porous zeolitic imidazolate frameworks nanoparticles for gene delivery. Microporous and Mesoporous Materials, 110200, 2020, https://doi.org/10.1016/j.micromeso.2020.110200.

168. HN Abdelhamid, Moataz Dowaidar, M Hällbrink, Ü Langel. Gene Delivery Using Cell Penetrating Peptides-Zeolitic Imidazolate Frameworks. Microporous and Mesoporous Materials, 2020, 110173, ISSN 1387-1811, DOI:10.1016/j.micromeso.2020.110173.

169. HN Abdelhamid, Moataz Dowaidar, M Hällbrink, Ü Langel. Cell Penetrating PeptidesHierarchical Porous Zeolitic Imidazolate Frameworks Nanoparticles: An Efficient Gene Delivery Platform. Available at SSRN 3435895, 2019.

170. Moataz Dowaidar. (2018). Chimeric gene delivery vectors : Design, synthesis, and mechanisms from transcriptomics analysis [Department of Biochemistry and Biophysics, Stockholm University]. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1242000

171. Moataz Dowaidar, Hani Nasser Abdelhamid, Mattias Hällbrink, Ülo Langel, Xiaodong Zou: Chitosan enhances gene delivery of oligonucleotide complexes with magnetic nanoparticles–cell-penetrating peptide. Journal of Biomaterials Applications 09/2018; 33(3):392-401., DOI:10.1177/0885328218796623

173. Moataz Dowaidar, Hani Nasser Abdelhamid, Mattias Hällbrink, Krista Freimann, Kaido Kurrikoff, Xiaodong Zou, Ülo Langel: Magnetic Nanoparticle Assisted Self-assembly of Cell Penetrating Peptides-Oligonucleotides Complexes for Gene Delivery. Scientific Reports 08/2017; 7(1)., DOI:10.1038/s41598-017-09803-z

174. Moataz Dowaidar, Hani Nasser Abdelhamid, Mattias Hällbrink, Xiaodong Zou, Ülo Langel: Graphene oxide nanosheets in complex with cell penetrating peptides for

oligonucleotides delivery. Biochimica et Biophysica Acta (BBA) - General Subjects 07/2017; 1861(9)., DOI:10.1016/j.bbagen.2017.07.002

175. Moataz Dowaidar, Jakob Regberg, Dimitar A. Dobchev, Tõnis Lehto, Mattias Hällbrink, Mati Karelson, Ülo Langel: Refinement of a Quantitative Structure–Activity Relationship Model for Prediction of Cell-Penetrating Peptide Based Transfection Systems. International Journal of Peptide Research and Therapeutics 07/2016; 23(1)., DOI:10.1007/s10989-016- 9542-8

176. Maxime Gestin, Moataz Dowaidar, Ülo Langel: Uptake Mechanism of Cell-Penetrating Peptides. Peptides and Peptide-based Biomaterials and their Biomedical Applications, 10/2017: pages 255-264; ISBN: 978-3-319-66094-3, DOI:10.1007/978-3-319-66095-0_11

177. Ahmed A.A. Ali, Nahla M. Wassim, Moataz M. Dowaidar, Ahmed E. Yaseen: Genetic polymorphism of CYP2D6 gene among Egyptian hypertensive cases. The Journal of Basic & Applied Zoology, 08/2013; 66(4)., 228-233. DOI:10.1016/j.jobaz.2012.12.002

178. Ahmed A.A. Ali, Nahla M. Wassim, Moataz Dowaidar, Ahmed E. Yaseen: Association of eNOS (E298D) and CYP2J2 (-50 G/T) gene polymorphisms with hypertension among Egyptian cases. The Journal of Basic & Applied Zoology, 08/2013; 66(4)., 234-241. DOI:10.1016/j.jobaz.2012.12.001

179. Almeman Ahmad, Khalaf Hassan, Rasool Semaab, Moataz Dowaidar, Al Orainy Mohammad: The impact of CYP2C19 polymorphism on platelet reactivity for guiding

clopidogrel treatment and cost analysis. Journal of the Saudi Heart Association 04/2013; 25(2):107., DOI:10.1016/j.jsha.2013.03.005

180. Abdullah Algasham, Hisham Ismail, Moataz Dowaidar, Ahmad A Settin: Methylenetetrahydrofolate Reductase (MTHFR) and Angiotensin Converting Enzyme (ACE) Gene Polymorphisms among Saudi Population from Qassim Region. 01/2013; 5(2 Suppl 1):3-4. PMC3533339

181. Abdullah Alghasham, Ahmad A Settin, Ahmad Ali, Moataz Dowaidar, Hisham Ismail: Association of MTHFR C677T and A1298C gene polymorphisms with hypertension. 12/2012; 6(1):3-11., DOI:10.12816/0005968

182. Ahmad Ali, Abdullah Alghasham, Hisham Ismail, Moataz Dowaidar, Ahmad Settin: ACE I/D and eNOS E298D gene polymorphisms in Saudi subjects with hypertension. Journal of Renin-AngiotensinAldosterone System 10/2012; 14(4)., DOI:10.1177/1470320312459976

183. Abdullah Alghasham, Ahmad Ali, Hisham Ismail, Moataz Dowaidar, Ahmad A Settin: CYP2J2-50 G/T and ADRB2 G46A gene polymorphisms in saudi subjects with hypertension.

Genetic Testing and Molecular Biomarkers 06/2012; 16(9):1027-31., DOI:10.1089/gtmb.2012.0006

184. Ahmad A Settin, Abdullah Alghasham, Ahmad Ali, Moataz Dowaidar, Hisham Ismail: Frequency of thrombophilic genetic polymorphisms among Saudi subjects compared with other populations. Hematology (Amsterdam, Netherlands) 05/2012; 17(3):176-82., DOI:10.1179/102453312X13376952196575

185. HA Ismail, AA Alghasham, MM Dowaidar, AA Settin: Polymorophisms in MTHF and ace genes and the association with hypertension among Saudi population from Qassim region. 06/2011; 29(1)., DOI:10.4314/ejbmb.v29i1.67382

186. Badr Aljarallah, Ahmed Ali, Moataz Dowaidar, Ahmad Settin: Prevalence of α-1-Antitrypsin Gene Mutations in Saudi Arabia. Saudi Journal of Gastroenterology 03/2011; 17(4):256-60., DOI:10.4103/1319-3767.82580

187. Moataz Dowaidar, Ahmad Settin: Risk of Myocardial Infarction Related to Factor V Leiden Mutation: A Meta-Analysis. Genetic Testing and Molecular Biomarkers 08/2010; 14(4):493-8., DOI:10.1089/gtmb.2010.0017

188. Ahmad A Settin, Abdullah Algasham, Moataz Dowaidar, Hisham Ismail: Methylene Tetrahydrofolate Reductase and Angiotensin Converting Enzyme Gene Polymorphisms Related toOverweight/Obesity among Saudi Subjects from Qassim Region. Disease markers 11/2009; 27(2):97-102., DOI:10.3233/DMA-2009-0660

189. Ahmad Settin, Hala Almarsafawy, Ahmad Alhussieny, Moataz Dowaidar: Dysmorphic Features, Consanguinity and Cytogenetic Pattern of Congenital Heart Diseases: a pilot study from Mansoura Locality, Egypt. 07/2008; 2(2):101-11. PMC3068729

190. Ahmad Settin, Moataz Dowaidar, Rizk El-Baz, Ayman Abd-Al-Samad, Ibrahim El-Sayed, Mahmoud Nasr: Frequency of factor V Leiden mutation in Egyptian cases with myocardial infarction. Hematology (Amsterdam, Netherlands) 07/2008; 13(3):170-4., DOI:10.1179/102453308X316158

191. Ahmad Settin, Ibrahem S Abu-Saif, Rizk El-Baz, Moataz Dowaidar, Rabab Abu-Al Kasim, Shaimaa Shabana: Diagnosis of Sex Chromosome Disorders and Prenatal Diagnosis of Down Syndrome using Interphase Fluorescent In-Situ Hyperidization Technique. 07/2007; 1(2):203-9. PMID: 21475429 PMC: PMC3068641