J Community Genet (2017) 8:255–266
DOI 10.1007/s12687-017-0316-6

ORIGINAL ARTICLE

Diversity and inclusion in genomic research: why the uneven progress?

Amy R. Bentley[1] · **Shawneequa Callier**[1,2] · **Charles N. Rotimi**[1]

Received: 13 March 2017 / Accepted: 29 June 2017 / Published online: 18 July 2017

Abstract Conducting genomic research in diverse populations has led to numerous advances in our understanding of human history, biology, and health disparities, in addition to discoveries of vital clinical significance. Conducting genomic research in diverse populations is also important in ensuring that the genomic revolution does not exacerbate health disparities by facilitating discoveries that will disproportionately benefit well-represented populations. Despite the general agreement on the need for genomic research in diverse populations in terms of equity and scientific progress, genomic research remains largely focused on populations of European descent. In this article, we describe the rationale for conducting genomic research in diverse populations by reviewing examples of advances facilitated by their inclusion. We also explore some of the factors that perpetuate the disproportionate attention on well-represented populations. Finally, we discuss ongoing efforts to ameliorate this continuing bias. Collaborative and intensive efforts at all levels of research, from the funding of studies to the publication of their findings, will be necessary to ensure that genomic research does not conserve historical inequalities or curtail the contribution that genomics could make to the health of *all* humanity.

Amy R. Bentley and Shawneequa Callier contributed equally to the manuscript.

This article is part of the Topical Collection on *Inclusion of Diverse Populations in Genomics Research and Health Services: A Scientific and Health Equity Imperative.*

✉ Charles N. Rotimi
rotimic@mail.nih.gov

[1] Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

[2] The George Washington University School of Medicine and Health Sciences, Washington, DC, USA

Keywords Genomics · Diversity · Inclusion · Africa

Introduction

The importance of diversity and inclusion in genomic research has long been appreciated. Promoting genomic research in diverse populations could be described as predominantly motivated by two goals. First, as a matter of justice, individuals are expected to benefit most from genomic research conducted in individuals with a similar ancestral background to them. Failure to fully engage diverse populations at all levels of genomic research perpetuates already considerable health disparities. Second, including diverse populations in genomic research is not just the right thing to do for reasons of equity, it is a scientific imperative. The genomes of diverse individuals harbor a treasure trove of humanity's responses to challenges experienced by some or all populations: changes in climate, infectious diseases, diet, etc. It is beyond doubt that understanding how our bodies have responded to these challenges in a variety of circumstances could have significant impact on our understanding of biology and developing clinical interventions. Meanwhile, limiting our investigation into these genomic responses to scientifically well-represented populations certainly curtails progress in the entire field of genomics.

Achieving the aim of increasing diversity in genomic research, however, has been slow and uneven. As the literature warns, researchers should include underrepresented populations more often and materially in their research, describe diverse cohorts in specific and detailed ways, and engage marginalized communities meaningfully in the research process. These are important steps needed to achieve adequate diversity and sustained

participation in genomic research. In appreciation of the necessity of including diverse populations in research, legislation was introduced in the USA in 1993 that provided guidelines for the inclusion of women and minority groups in federally funded or approved clinical research [NIH Revitalization Act of 1993 Public Law 103-43. Federal Register, 59FR14508]. Despite this legislation, participation in health-related genetic research among members of ethnic groups experiencing the greatest disparities, such as African Americans and Latinos, has remained limited (Oh et al. 2016). A 2009 analysis reported the shocking statistic that 96% of the genomic studies were carried out in populations of European descent (Bustamante et al. 2011; Need and Goldstein 2009). When this issue was revisited in 2016, encouragingly, the proportion of genome-wide association studies (GWAS) in samples from underrepresented populations had increased to 19%. Most of this increase, however, was limited to Asian ancestry samples, with other ethnic groups experiencing minimal increases (from 1 to 4%) (Popejoy and Fullerton 2016). Now, considering the pace of change in the 20 years since the NIH mandated inclusion of diverse populations, it seems unlikely that current remedies are sufficient or that this dynamic will self-correct. Rather, it is imperative to examine the factors that appear to be perpetuating this inequality.

Scholars of various disciplines including genetics, ethics, and sociology have called attention to research norms that insufficiently prioritize the collection, annotation, and reporting of genomic variation among diverse populations over time, yet few effective changes have been made. In this article, we review the rationale for including diverse populations in genomic research, with a focus on African ancestry populations as an example, consistent with the authors' expertise. Then, given the modest advances despite the acknowledged need for an increase of diversity in genomic research, we provide an assessment of the barriers that are limiting genomic analyses in African ancestry populations. Finally, we report on ongoing efforts to increase diversity in genomic research and comment on other opportunities to promote this goal.

Why do genomic research in diverse populations?

Motivations to conduct research in the context of genetic diversity are numerous. Increased inclusion facilitates the understanding of health disparities, new discoveries in biology, more accurate matching of diverse patients with safe and effective treatments, improved interpretation of genetic tests, and better tracing of human history. Examples of each of these contributions of genomic research in diverse populations are described as follows.

Novel insights into health disparities

In recent years, genome scientists have begun to outline a role for genomic research to help reduce health disparities. As once estimated by the former US Surgeon General, Dr. David Satcher and colleagues, over 83,000 deaths occur every year because of health disparities between black and white patient populations (Satcher et al. 2005). Genomic research can help shed light on the genetic influences on health disparities, as is the case with kidney disease, which occurs at a much higher prevalence among individuals with African ancestry (Friedman and Pollak 2011; Genovese et al. 2010; Tarver-Carr et al. 2002). Variants in *APOL1* are associated with dramatically increased risk of kidney disease of varying etiologies (Freedman et al. 2014; Genovese et al. 2010; Kasembeli et al. 2015; Kopp et al. 2011; Parsa et al. 2013) with odds ratios that may be the highest reported for a common variant: 29 and 89 for risk of HIV-associated nephropathy in African Americans (Kopp et al. 2011) and South African Blacks (Kasembeli et al. 2015), respectively, and 17 for focal segmental glomerulosclerosis (Kopp et al. 2011). These variants are common among individuals with African ancestry, but absent among those without African ancestry. It is important to note that all individuals with African ancestry and, thus, potentially this variant, may not self-identify or appear to have African descent. These kidney disease risk alleles are thought to be at high frequency, because they confer resistance against human African trypanosomiasis ("African Sleeping Sickness"), perhaps in addition to other infectious diseases (Thomson et al. 2014). The frequency of these variants and the magnitude of the effect translate into not only a large public health burden but also a significant potential to ease this burden if targeted interventions are discovered. Also, uncovering this genetic risk factor has led to significant advances in understanding the pathophysiology of kidney disease (Julian et al. 2016; Ku et al. 2017; Ma et al. 2016; Peralta et al. 2016). *APOL1* exemplifies how a genetic variant can contribute to ethnic disparities in disease risk. Long-observed ethnic disparities in kidney transplantation outcomes have been attributed to this variant (Reeves-Daniel et al. 2011).

Understanding human biology

Including diverse populations in genomic research can help to facilitate new understanding of human biology important for clinical practice and public health. Variants that are present only or only at sufficient frequency in diverse populations, of course, can be evaluated exclusively or more efficiently in these populations. Thus, any insights latent in the association between these variants and traits of interest can only be

uncovered by studying diverse populations. For instance, rare nonsense variants (genetic alterations that cause the premature termination of a protein) in *PCSK9* found in higher frequency in African Americans are associated with dramatic reduction in low-density lipoprotein cholesterol concentration (LDLC; 28–40%) (Cohen et al. 2005, 2006) and concomitant decrements in coronary heart disease risk (88%) (Cohen et al. 2006). These variants were present in individuals of European descent, but in such limited numbers as to preclude analysis (0.006 vs. 2.6% carriers in African ancestry individuals) (Cohen et al. 2006). It has been suggested that these variants may be in higher frequency among African ancestry individuals as a result of selection pressures due to malaria, though this may also reflect genetic drift (Horton et al. 2007). Efforts to exploit this genetic phenomenon pharmacologically are promising. Two monoclonal antibody PCSK9 inhibitors, evolucumab and alirocumab, have been approved widely for use either alone or in combination with statins after demonstrating that they are well tolerated and effective at reducing LDLC (Gouni-Berthold et al. 2016; Roth et al. 2016) and, for evolucumab, risk of cardiovascular outcomes (Sabatine et al., 2017). Inclisiran, a synthetic small interfering RNA (siRNA) that reduces PCSK9, has recently been shown to be effective in reducing LDLC in a phase 2 clinical trial, with phase 3 trials underway (Ray et al. 2017; Fitzgerald et al. 2017; Sheridan 2013). Notably, despite the importance of African American genetic variation in identifying this locus, drugs targeting PCSK9 have the potential to benefit the large number of individuals, not solely African Americans, who are at risk of this leading cause of mortality.

Improving clinical care

Diversity and inclusion in genomic research can provide needed data and guidance to physicians to help them to make better decisions for patients of diverse ancestry at the point of clinical care. Pharmacogenomics is a promising area of genomic medicine that increasingly allows providers to choose particular drugs and dosages based on the patients' genetic profiles. One good example involves the ability of physicians now to predict the potential for a toxic effect and adverse reaction among patients who carry the HLA-B*5701 allele. Between 4 and 8% of the patients with HIV who are prescribed the drug abacavir experience abacavir hypersensitivity syndrome (AHS), a serious and potentially life-threatening response. Early clinical trials demonstrated that screening for the HLA-B*5701 allele substantially reduced the incidence of AHS. Since the incidence of AHS in European populations was 5 to 8% and much lower in other populations, including "black" and "African" individuals, it was recommended that screening for HLA-B*5701 occur in European populations. It was not until investigators explored the global distribution of the HLA-B*5701 variant that the inadequacy of the broad racial labels to describe AHS risk was apparent (Rotimi and Jorde 2010). The prevalence of HLA-B*5701 in the Kenyan Masai group was 13.6% (more than double that in European samples) and absent among the Yoruba in Nigeria. Genetically, and not racially, guided prescription of abacavir is now the standard of care for the treatment of HIV.

Informing genetic diagnoses

Including diverse populations in genomic research generates appropriate data for genetic diagnosis. Hypertrophic cardiomyopathy is a case in point, as it varies by ethnicity, with higher prevalence among African ancestry individuals (Maron et al. 1995; Movahed et al. 2010). The clinical presentation of hypertrophic cardiomyopathy can vary considerably, and identifying a pathogenic variant is a useful tool for establishing a diagnosis. Some of the pathogenic variants for this condition were found at a frequency in the African American general population that was unexpectedly high for a causative variant for this severe condition. Yet, it is likely that these variants were misclassified due to their rarity among white populations, in whom original analyses were conducted, while interrogation among African ancestry healthy controls would have established them as benign. Unfortunately, these historical failures to include diverse controls in analyses have led to individuals and their family members, of all ethnicities but more frequently of African ancestry, receiving an incorrect diagnosis with potentially significant impact on life choices (Manrai et al. 2016). This case exemplifies how not including diverse participants in genomic research can cause worse clinical care for individuals with African ancestry.

Describing human history

Increasing diversity in genomic research is critically important to the global efforts to develop a comprehensive catalog of genetic variations to address fundamental questions about human origin, peopling of the world, genetic admixture, and adaptive capabilities to different environmental conditions, all of which have played a role in shaping health and disease (Nielsen et al. 2017). Additional insights being gained include deeper understanding of the scope and extent of genetic diversity across the African continent despite current underrepresentation of Africans in genomic research (Genomes Project et al. 2015; Gurdasani et al. 2015; Shriner et al. 2014). Scientists are noting how highly subdivided African genomes are, with population structure across Africa currently recognized as comprising 11 ancestries that correspond to a combination of geographic and linguistic separation. By comparison, 12 ancestries have been identified in the rest of the world (Rotimi et al. 2016; Shriner et al. 2014). Furthermore, African populations display evidence of both ancient and modern migrations, including the well-described Bantu expansion with

movement of the speakers of this language from their homelands in the present-day areas of Nigeria and Cameroon to other regions of sub-Saharan Africa (Nielsen et al. 2017). Thus, application of the new high-throughput sequencing and genotyping technologies to a more systematic collection of DNA across Africa, especially the sub-Saharan region, promises to provide even greater insights into human history and how the evolutionary need to adapt to environmental challenges, such as climate, exposure to infectious agents, and diet has shaped human genomes (Genomes Project et al. 2015; Gurdasani et al. 2015; Rotimi et al. 2016).

Barriers to diversity and inclusion in genomic research

Given the scientific and equity motivations for including diverse populations in genomic research, and the contrasting underrepresentation in published results, it is reasonable to ask whether current approaches are enough to effect change. Some of the most critical barriers to diversity and inclusion, for instance, receive little attention in the public discourse. Further, proposed strategies for addressing these and other more prominent challenges have been limited. Key challenges to diversity and inclusion in genomic research are described as follows and summarized in Table 1.

Lagging diversity within the scientific community

An overarching concern is that more diverse representation at all levels of research is sorely needed—from the participants included to the reviewers of proposals to the scientists conducting the research. Scientists from low-resource settings, minority backgrounds, and diverse ethnic groups are grossly underrepresented in all of these areas (Ginther et al. 2011; Oh et al. 2015; Panofsky and Bliss 2017; Science and inequality 2016). The underrepresentation of scientists from diverse ethnicities and environmental backgrounds may lead to the loss of hypotheses and priorities that experts with different perspectives would bring to genomic research. For instance, an Ethiopian scientist achieved the first GWAS of the neglected tropical disease, podoconiosis (Tekola Ayele et al. 2012), a significant cause of morbidity and stigma in Ethiopia, but rare among populations with thriving genomic research communities. The same researcher has been able to interpret signals of recent selection within the genomes of Ethiopian individuals in the context of dietary use of the crop enset (Tekola-Ayele et al. 2015a), an insight which would be difficult for a non-Ethiopian researcher to achieve. Since investigators may have personal connections and interests related to the communities that they aim to serve, diversity and inclusion among investigators, as well as incentives for investigators to return to their communities for ongoing research, access, dissemination, and benefit sharing, could open new areas of opportunity for important and sustained studies.

Limited engagement

When considering the shortfalls in genomic research participation among minority communities, scientists emphasize policies and barriers related to recruitment and sustained enrollment. Few models have been proposed, however, that stress sustained engagement and enthusiasm about genomic research (Tindana et al. 2015). Empirical evidence has shown,

Table 1 Key challenges and recommendations for promoting genomic research in diverse populations

Challenge	Recommendation	Example
Lagging diversity within the scientific community	Initiatives to foster the careers of researchers from diverse backgrounds	Human Health and Heredity in Africa (H3Africa)
Limited engagement	Empirical research on the participants' perspectives related to participating in research and remaining engaged throughout the research process. Follow-up with the participants after research is over. Discussion with communities about how members define engagement	MalariaGEN, The International Haplotype Mapping (HapMap) project
Preference by researchers to analyze data from well-characterized, well-powered, predominantly European ancestry cohorts	Increase the size and number of well-characterized cohorts of diverse populations Incentivize research in diverse samples/penalize lack of diversity at the stage of funding and publication	H3Africa, TOPMed, PAGE II
Difficulty in publication/funding due to relatively smaller sample sizes in diverse populations	Need for reviewers to appreciate the historic and current context of conducting genomic research in diverse populations	
Limitations in current genotyping technology to adequately capture variation among diverse individuals	Develop genotyping chips based on diverse samples	Affymetrix PanAFR, MEGA, H3Africa
Analytical challenges of diversity	Development of new strategies to address genomic and environmental diversity in analyses	Joint association and admixture mapping (BMIX)

for instance, that trust is a critical component in the participants' views regarding important ethics and legal challenges in genomic research, such as data sharing (Trinidad et al. 2010). Participants have emphasized the need for trustworthy governance policies to increase their comfort levels with genomic research, so that scientific advancement does not occur at the expense of participating communities (McDonald et al. 2008; Yarborough et al. 2009). Research should go further to gain insight into perspectives on partnerships, benefit sharing, and ongoing community engagement throughout the course of a genomic study (Tindana et al. 2015). Engagement with the participants and participating communities before and during the research process as well as after a particular research study is over could have long-term benefits for both genomic research and patient communities, especially with regard to communication and sustained interest among community partners.

Historical and international examples hold lessons for the genomic research community. In the USA, discoveries in Tay-Sachs disease research enabled clinicians along with synagogues and community centers to implement community-based Tay-Sachs disease carrier screening programs. The researchers' and clinicians' long-term involvement with at-risk families contributed to a screening program viewed by others as largely successful (Burke et al. 2011). On the other hand, misunderstandings in genetics related to the differences between sickle cell trait and sickle cell disease, racial tensions, lack of prenatal diagnostic testing at the time, and insufficient community involvement and support for many screening programs were some of the many reasons that carrier-screening programs for sickle cell disease failed in the 1970s (Burke et al. 2011). The legacy of sickle cell screening programs, the Tuskegee Syphilis experiments, and widely read books that draw attention to a divide that sometimes occurs between researchers' goals and research participants' expectations, such as *The Immortal Life of Henrietta Lacks* (Skloot 2011), continue to be cited as reasons for lack of trust and low recruitment rates among African Americans to genomic research. Enhanced engagement with diverse communities provides an opportunity to improve recruitment and build trust.

Around the globe, investigators have provided much needed examples of meaningful community engagement. The International Haplotype Mapping project (HapMap), for instance, incorporated community input into research recruitment and sampling processes (Rotimi et al. 2007). The Malaria Genomic Epidemiology Network (MalariaGEN), a scientific network of 36 countries investigating malaria and biology, provides a good example of an international project that aims to build long-lasting relationships to support collaborative science (Malaria Genomic Epidemiological 2017). Such partnerships are overcoming barriers to trust created by historical research abuses and exploitation internationally in underrepresented communities.

"Preferred Cohort" effect

A key factor in the perpetuation of disproportionate research participation is that European ancestry cohorts remain the "preferred cohort" for genomic analyses—the vast majority of the well-characterized, well-studied genomic research cohorts are of European ancestry populations. Historically, the genetic research community has focused on European ancestry populations, which has led to large available sample sizes from European ancestry cohorts. Sample size is a key factor on which genomic research is currently judged, particularly as focus shifts to lower frequency variants. Thus, in order to meet the expectations of reviewers, investigators are more likely to study cohorts that are predominantly of European ancestry. Furthermore, a study of a novel hypothesis or method may include a well-studied cohort or consortia so that results can be evaluated in terms of previous findings, instead of having the additional potential effect of ethnicity to evaluate. The scientific communities' emphasis on publishing high-quality research in a timely manner, though a noble goal, may also encourage the repeated use of well-characterized, predominantly European ancestry, cohorts.

In addition, the implications of genomic diversity, itself, may lead investigators to prefer to conduct studies in European ancestry populations. The relatively higher linkage disequilibrium in European ancestry populations means that, on average, a genotyped variant in these individuals "tags" a larger proportion of genetic variation than the same variant genotyped in African ancestry individuals. In the context of GWAS, then, conducting analyses in European ancestry populations is more efficient for the discovery of associations. Out of concern for errors due to population structure and to maintain the genetic "homogeneity" of the sample, even studies with some individuals of diverse ancestry may exclude them from analysis.

As GWAS of diverse populations are becoming increasingly available, consortia are including these populations in their analyses, but the preference for presenting European ancestry results is apparent. In many studies, although investigators may have a sufficient sample size from diverse populations to conduct scientifically valid analyses, the results are often summarized only as a comment on the generalizability of the "main" results from the European ancestry analyses. The data from analyses in diverse populations, meanwhile, may be presented only in the supplementary text. This represents a missed opportunity for the genomics of diverse populations. Additionally, it has the consequence of representing the findings in the larger European cohorts as "canon" or "standard," against which findings in other ancestry groups may be measured. As a result of this, novel findings in diverse populations face an additional burden of proof. Peer reviewers may ask, for instance, "why hasn't this already been found in the larger studies of European ancestry populations?" Where the

difference is not readily explained through interethnic allele frequency differences, novel findings in diverse populations may face additional hurdles in the peer review and publication process. For instance, a large meta-analysis of African Americans (Ng et al. 2014) discovered a novel type 2 diabetes locus at *INS-IGF2* but faced skepticism at peer review as this locus had not been previously identified in this context. Suggestive evidence for replication in Africans for this locus has now been observed (Adeyemo et al. 2015).

Flawed comparisons

In some ways, the very success of GWAS conducted in European ancestry populations inhibits the publication of studies in diverse populations. Although the first GWAS efforts in European ancestry populations were relatively small, large meta-GWAS in European ancestry publications have now aggregated data on hundreds of thousands of individuals for common traits. The reviewers' expectations of GWAS have advanced along with these newer publications, as would be expected. For GWAS of diverse populations, however, which have not yet reached such sample sizes, this change in perspective may prove to be a hardship. As an example, the first GWAS of type 2 diabetes in sub-Saharan Africans was moderately sized (n = 1775) in terms of early GWAS in European ancestry populations (Scott et al. 2007; Sladek et al. 2007), but was reviewed through the lens of current GWAS in more well-represented populations, and had to be repackaged as a replication effort for publication (Adeyemo et al. 2015). To date, no GWAS of type 2 diabetes in Africans has been published.

The observation that no GWAS of type 2 diabetes and most other common traits has yet been conducted in Africa is in contrast to the declaration that we are now in a "post-GWAS" era (Huang 2015; Li et al. 2012; Polychronakos and Alriyami 2015), an indication of the field's waning interest in this type of analysis. However, the promise of GWAS, which has yielded incredible successes (Price et al. 2015), is still largely unrealized among African ancestry populations. Categorizing GWAS in sub-Saharan Africans provides important context for this statement, in whom the spectrum of published GWAS predominantly reflects infectious disease-related traits (tuberculosis (Chimusa et al. 2014; Thye et al. 2012, 2010), malaria (Band et al. 2013; Jallow et al. 2009; Malaria Genomic Epidemiology et al. 2015; Milet et al. 2016), and HIV (Joubert et al. 2010; Lingappa et al. 2011; Petrovski et al. 2011)), although there are also GWAS for blood cell traits (Mtatiro et al. 2014; Ramsuran et al. 2011), height (Kang et al. 2010; N'Diaye et al. 2011), and the neglected tropical disease podoconiosis (Tekola Ayele et al. 2012). What remains grossly understudied are chronic disease-related outcomes, currently limited to BMI (Kang et al. 2010), plasma homocysteine (Kim et al. 2016), metabolic syndrome (Tekola-Ayele et al. 2015b), prostate cancer (Cook et al. 2014), and a replication and transferability analysis for type 2 diabetes (Adeyemo et al. 2015; Welter et al. 2013). While GWAS in European ancestry populations may have identified all of the common variants with appreciable influence on many traits, marking the declining usefulness of standard GWAS approaches, GWAS in diverse populations are far from exhaustive. Failure to judge efforts in diverse populations within the context of the literature on those populations inhibits the advancement of this research.

Technological limitations

The goal of adequately representing diverse populations in genomic research is hampered by technological challenges. Large-scale genomic research depends on adequate capture of the sequence variation present in the individuals under study. Many of the commercial chips that are used to genotype participants were developed using DNA samples from almost entirely non-African ancestry persons. These chips are therefore less efficient for interrogating the genomes of Africans, because they are less likely to adequately represent the common genetic variants seen in African ancestry populations. This introduces a significant limitation to evaluating the variation among African ancestry individuals that might be important in disease risk. The largest type 2 diabetes GWAS in African Americans was only able to cover 43.3% of the common variants expected based on sequence data of African Americans (Ng et al. 2014). An array specifically designed using studies of African genetic diversity (Affymetrix Axiom© PanAFR) dramatically increased coverage compared to previously available arrays, but the coverage was still markedly less than those achieved for those of Asian and European ancestry using the same chip (Ha et al. 2014). Thus, the proportion of genetic variation that has been evaluated with respect to disease risk has been less in African ancestry populations.

Analytical challenges of diversity

Finally, genomic diversity itself requires additional analytical consideration. The degree of genomic diversity is the highest among African ancestry populations, as a result of population history. As mentioned earlier, this diversity necessitates a larger number of variants in order to tag the same amount of variation as in European ancestry populations and the resulting decrease in power must be offset with a larger sample size. Thus, simply attaining a similar number of African ancestry individuals as are found in European ancestry studies may not yield a comparable number of findings. This, however, may not adequately address the implications of this diversity: given the remarkable genetic diversity among African ancestry populations, the monolithic categorization of these

individuals may be inappropriate (Rotimi et al. 2016). How should these individuals be meaningfully subdivided for genomic analysis? A study combining data from Nigeria and Ethiopia may be as analytically questionable as analyzing African and European Americans together. Similarly, great genetic diversity among African Americans has been noted (Mathias et al. 2016; Rotimi et al. 2016). How appropriate is the analysis of African Americans as a single entity given their descent from very diverse ethnic groups across the continent of Africa? The strategies for addressing this genomic diversity remain to be fully explored, but it is clear that "African ancestry" is an overly simplistic label for the complexity of the genomes of those of African descent.

In addition to genetic complexity represented by African ancestry populations, there is also considerable heterogeneity in environmental context among individuals in the diaspora, affecting factors that are certain contributors to disease risk: diet, physical activity, socioeconomic status, stress, infectious disease exposure, smoking, alcohol consumption, healthcare access, and healthcare beliefs, etc. (Bentley and Rotimi 2012; Rotimi et al. 2016). Although level of education may be collected in studies of both African Americans and West Africans, the meaning of these categorizations will be much different in terms of socioeconomic background in these two environments. Appropriately accounting for such factors is complex and, again, will necessitate careful sampling strategies by researchers who are knowledgeable of the communities under study.

The presence of admixture, often thought of in terms of African Americans in genomic research, has been found to be much greater than initially thought. In a study of worldwide populations, 97.3% of the individuals showed mixed ancestry (Baker et al. 2017; Shriner et al. 2014). New analytical strategies that account for, and take advantage of, admixture in the context of African Americans are emerging, such as joint association and admixture mapping (Shriner et al. 2011). Beyond admixed individuals who fall into categories that are represented in genomic research (e.g., African Americans and Hispanics), there are also an increasing number of individuals within the USA who self-identify with more than one racial or ethnic group and find census categories insufficient to capture their identities (Cohn 2015). As racial and ethnic identities continue to change and become more complex, new strategies to represent these individuals in genomic research are needed. Individuals who self-identify with multiple ancestries may be excluded from analyses as outliers (Bryc et al. 2015; Moore et al. 2017; Wells 2012). Separate analysis of these individuals is not reasonable using current strategies because of fewness of numbers, especially within categories of ancestry. Individuals of all ancestries, no matter how complex, deserve to be represented in genomic research, yet methods to achieve this goal may need to be developed.

Improving diversity and inclusion

The challenges outlined previously, although significant, are surmountable. There are initiatives underway that demonstrate successes in building capacity, developing resources, and overcoming technological limitations. These ongoing efforts are notable, yet other strategies, such as incentivizing diversity, may be necessary to develop an environment that fosters genomic research in diverse populations.

Building capacity in Asia

A success story in advancing genomics in non-European descent populations is taking place in Asia, where genomic research infrastructure and yield have been increasing rapidly. From 2009 to 2016, the proportion of GWAS participants that are of Asian ancestry has increased from 3 to 14% (Popejoy and Fullerton 2016), a remarkable achievement in terms of diversification of genomic research. Additionally, of GWAS involving only Asian participants, the vast majority (~93%) were conducted in Asian countries (Popejoy and Fullerton 2016). The meteoric rise in genomic capacity in China during that time appears to depend on four salient enabling features. First, and most importantly, the development in genomic resources is a priority of the Chinese government. BGI, formerly known as the Beijing Genomics Institute, received a 10-year US$1.58 billion loan from the China Development Bank in 2010 (Cyranoski 2012), which facilitated the largest purchase of sequencer instruments in history, and enabled BGI to become the world's most prolific DNA sequencer (Larson 2013). The Chinese government created the China National Genebank and other biobanks (Zhang et al. 2015). The government is also championing precision medicine (Cyranoski 2016) and recently announced a US$9.2 billion precision medicine initiative (Russell 2016). Secondly, genomic companies, particularly BGI, became significant providers of sequencing and analysis for institutions and companies worldwide. These efforts supply the company with steady and increasing economic resources for research and further development (Cyranoski 2012). Third, while the advances in genomic research are relatively recent, China already had a well-developed and stable academic, research, and medical infrastructure with an emphasis on centralization. In such an environment, additional resources could facilitate efficient development of genomic expertise and capacity. Finally, the value of genetic testing has been appreciated in China for many years, and prenatal genetic testing is widely available through public health services (Zhao et al. 2013). Thus, an understanding of the potential value of genomics was already established

to some degree. While the confluence of the factors supporting a genomics revolution in China may be unique to that country, it would be worthwhile to investigate how these successes might inform capacity building in other regions of the world.

Resource development

Notable efforts to increase representation of diverse participants in high-quality studies are currently underway, including the "All of Us" Research Program (Collins and Varmus 2015) in the USA and the Human Health and Heredity in Africa (H3Africa) Initiative (The H3Africa Consortium 2014), among others. Researchers in the USA will soon go beyond investigating medical interventions for the average patient to assessing how to tailor therapies to individuals based on their genomic data. The federally funded All of Us program, part of the federal Precision Medicine Initiative, envisions a large-scale research effort that will enroll one million or more volunteers to share their lifestyle, environment, and genetic information with investigators. Importantly, the organizers aim to build a cohort that reflects the diversity of the US population in terms of social, ethnic, ancestral, geographic, economic, and health status, as well as age (Collins and Varmus 2015). In Africa, considerable resource building is occurring as part of the H3Africa initiative, a collaboration among the African Society of Human Genetics, the National Institutes of Health, and the Wellcome Trust (Ramsay et al. 2015) (http://www.h3africa.org/ accessed March 8, 2017). These initiatives are enrolling more than 75,000 African participants from 27 African countries in genomic research on a diverse range of outcomes. These data are certain to change the landscape of genomic research. In addition to this groundbreaking work, there is also a notable focus on diversity in a range of new genomic initiatives, including the Trans-Omics for Precision Medicine (TOPMed) Program, which will provide a variety of -omics data for the Precision Medicine Initiative (https://www.nhlbi.nih.gov/research/resources/nhlbi-precision-medicine-initiative/topmed), and the Population Architecture using Genomics and Epidemiology (PAGE) Consortium, whose next phase (PAGE II), will genotype approximately 50,000 individuals from non-European cohorts (https://www.genome.gov/27541456/population-architecture-using-genomics-and-epidemiology/). These efforts are poised to provide greatly increased data to the field, undoubtedly improving subsequent genomic discoveries.

The H3Africa initiative represents a recognition that bringing Africa into the genomics revolution, however, necessitates more than the provision of samples. To facilitate genomic research in Africa, this initiative is addressing multiple levels of capacity building. H3Africa funds are given to African investigators working in African institutions. The initiative has funded research projects, collaborative centers, ethics projects, biorepositories, and a pan-African bioinformatics network (Mulder et al. 2016). Data governance policies encourage data sharing but also prioritize the investigators in Africa who collect and produce the data, so that they may develop their own research portfolios. Training of young African scientists is a priority of all funded projects. The effects of this initiative will be felt for many years, fostering of a vibrant research community of investigators on the continent who are world-class contributors to the field of genomics (Ramsay et al. 2015).

Advances in genotyping technology

The serious technological deficiency in coverage of the genetic diversity in African ancestry individuals is an area of considerable current effort. The relatively recent development of chips, such as the MEGA and the Affymetrix PanAFR arrays, based on DNA samples from more African ancestry populations, has improved coverage. The H3Africa initiative has taken these recent achievements several steps forward by spearheading the development of a new genome-wide chip based on the sequencing of Africans sampled from more ancestral backgrounds across sub-Saharan African populations than ever before (Ramsay et al. 2015). This unprecedented African-enriched genotyping chip promises to be more cost-effective and scientifically efficient than previous generations of genome-wide genotyping chips.

Incentivizing diversity

A key strategy for building diversity in genomic research that has not received enough attention is incentivizing diversity and discouraging lack of diversity. As mentioned earlier, while the NIH has passed research requirements to advance diversity and inclusion in research, as one prominent scientist has argued, they have been unsuccessful in enforcing these laws (Editors et al. 2016). Over time, the failure to hold researchers accountable has led to increased ethnic disparities (Editors et al. 2016). Given that the generally acknowledged need for increased diversity in genomic research has not, on its own, led to increased diversity, it may be reasonable to incentivize diversity at the level of the gatekeepers of genomic research: funders and editors. For instance, funders could reflect the potential of a proposed study to increase diversity in genomic research in the scoring for that study. Editors and reviewers should call researchers to account for submissions that either perform cursory analysis or reporting or entirely exclude diverse samples. Such a strategy has been successful in the transformation of the field in terms of data sharing. When the NIH mandated and top journals insisted on deposition of data to databases for broad sharing, that became the standard in genomic research. A similar insistence regarding

the attention to diversity could have an enormous impact on genomic research.

Data and biospecimen sharing

Like diversity and inclusion, responsible data and biospecimen sharing is a scientific imperative (Knoppers 2014) (https://genomicsandhealth.org/). Dr. George Gey shared cells generated from Henrietta Lacks' immortal cell line widely among investigators in the 1950s while in search of a cure for cancer (Skloot 2011). Funders and investigators involved with the Human Genome Project and genomic advancements since then continue to emphasize global data sharing and to require that data be distributed in shared open access and limited access databases, such as the database of Genotypes and Phenotypes (dbGaP) and the European Genotype Archive, for the benefit of all researchers globally (Kaye et al. 2009). All of the challenges described herein related to diversity and inclusion require a nuanced and careful data-sharing approach. Individual authorship, acknowledgement, and recognition of data collectors and principal investigators from low resource and underrepresented settings are necessary to provide diverse investigators adequate opportunity for professional advancement (Kaye et al. 2009; Rotimi and Mulder 2014). This is one reason why, as alluded to previously, H3Africa provides African researchers with 23 months of protected time prior to data publication by outside researchers to protect the original collectors of said data from unfair competition in publishing (de Vries et al. 2015; The 2014). Public trust, oversight, and long-lasting relationships with communities who participate in genomic research are required to advance both data sharing and diversity and inclusion—two major components of genomic research that must advance symbiotically for genomic research to benefit all.

Conclusions

As highlighted by the examples provided previously, insufficient diversity and inclusion in genomic research have ethical and scientific consequences. Individuals with African ancestry are not receiving the same level of care as individuals of European ancestry due to limitations in available data. Further, all populations are harmed when scientific advancements are stalled by gaps in research and policies that impair discovery. In the USA, where the vast majority of health research is federally funded through the NIH, the failure to adequately represent the diversity of the taxpayers raises fairness concerns.

Importantly, the disparities in genomic research, and in general biomedical research, that have been maintained, particularly with regard to African and Hispanic ancestry individuals, will not be ameliorated without concerted effort. While simply increasing the number of well-characterized diverse samples will be of tremendous importance moving forward, there remain other issues that need to be considered in order to effectively represent diverse genomes in research. This lack of diversity is a failing that the genomics community must address in order to ensure that the stories that we discover encoded in our DNA—of history, health risk, and adaptation—represent us all.

Acknowledgments The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official view of the National Institutes of Health. This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute in the Center for Research in Genomics and Global Health (CRGGH—Z01HG200362). CRGGH is also supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health.

Compliance with ethical standards This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare that they have no conflict of interest.

References

Adeyemo AA et al (2015) Evaluation of genome wide association study associated type 2 diabetes susceptibility loci in sub Saharan Africans. Front Genet 6:335. doi:10.3389/fgene.2015.00335

Baker JL, Rotimi CN, Shriner D (2017) Human ancestry correlates with language and reveals that race is not an objective genomic classifier. Sci Rep 7:1572. doi:10.1038/s41598-017-01837-7

Band G et al (2013) Imputation-based meta-analysis of severe malaria in three African populations. PLoS Genet 9:e1003509. doi:10.1371/journal.pgen.1003509

Bentley AR, Rotimi CN (2012) Interethnic variation in lipid profiles: implications for underidentification of African-Americans at risk for metabolic disorders. Expert Rev Endocrinol Metab 7:659–667. doi:10.1586/eem.12.55

Bryc K, Durand Eric Y, Macpherson JM, Reich D, Mountain Joanna L (2015) The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am J Hum Genet 96:37–53. doi:10.1016/j.ajhg.2014.11.010

Burke W, Tarini B, Press NA, Evans JP (2011) Genetic screening. Epidemiol Rev 33:148–164. doi:10.1093/epirev/mxr008

Bustamante CD, De La Vega FM, Burchard EG (2011) Genomics for the world. Nature 475:163–165

Chimusa ER et al (2014) Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet 23:796–809. doi:10.1093/hmg/ddt462

Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 37:161–165

Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart

disease. N Engl J Med 354:1264–1272. doi:10.1056/NEJMoa054013
Cohn DV (2015) Census considers new approach to asking about race—by not using the term at all. http://www.pewresearch.org/fact-tank/2015/06/18/census-considers-new-approach-to-asking-about-race-by-not-using-the-term-at-all/. Accessed 13 Mar 2017
Collins FS, Varmus H (2015) A new initiative on precision medicine. N Engl J Med 372:793–795. doi:10.1056/NEJMp1500523
Cook MB et al (2014) A genome-wide association study of prostate cancer in West African men. Hum Genet 133:509–521. doi:10.1007/s00439-013-1387-z
Cyranoski D (2012) Chinese genomics giant BGI plots commercial path. Nat Biotech 30:1159–1160. doi:10.1038/nbt1212-1159
Cyranoski D (2016) China embraces precision medicine on a massive scale. Nature 529:9–10. doi:10.1038/529009a
de Vries J et al (2015) The H3Africa policy framework: negotiating fairness in genomics. Trends in genetics : TIG 31:117–119. doi:10.1016/j.tig.2014.11.004
Editors PM et al (2016) Towards equity in health: researchers take stock. PLoS Med 13:e1002186. doi:10.1371/journal.pmed.1002186
Fitzgerald K et al (2017) A highly durable RNAi therapeutic inhibitor of PCSK9. N Engl J Med 376:41–51. doi:10.1056/NEJMoa1609243
Freedman BI et al (2014) End-stage renal disease in African Americans with lupus nephritis is associated with APOL1. Arthritis Rheumatol 66:390–396. doi:10.1002/art.38220
Friedman DJ, Pollak MR (2011) Genetics of kidney failure and the evolving story of APOL1. J Clin Invest 121:3367–3374. doi:10.1172/JCI46263
Genomes Project C et al (2015) A global reference for human genetic variation. Nature 526:68–74. doi:10.1038/nature15393
Genovese G et al (2010) Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science:329, 841–845. doi:10.1126/science.1193032
Ginther DK, Schaffer WT, Schnell J, Masimore B, Liu F, Haak LL, Kington R (2011) Race, ethnicity, and NIH research awards. Science 333:1015–1019. doi:10.1126/science.1196783
Gouni-Berthold I, Descamps OS, Fraass U, Hartfield E, Allcott K, Dent R, März W (2016) Systematic review of published Phase 3 data on anti-PCSK9 monoclonal antibodies in patients with hypercholesterolaemia. Br J Clin Pharmacol 82:1412–1443. doi:10.1111/bcp.13066
Gurdasani D et al (2015) The African genome variation project shapes medical genetics in Africa. Nature 517:327–332. doi:10.1038/nature13997
Ha NT, Freytag S, Bickeboeller H (2014) Coverage and efficiency in current SNP chips. Eur J Hum Genet 22:1124–1130. doi:10.1038/ejhg.2013.304
Horton JD, Cohen JC, Hobbs HH (2007) Molecular biology of PCSK9: its role in LDL metabolism. Trends Biochem Sci 32:71–77. doi:10.1016/j.tibs.2006.12.008
Huang Q (2015) Genetic study of complex diseases in the post-GWAS era. J Genet Genomics 42:87–98. doi:10.1016/j.jgg.2015.02.001
Jallow M et al (2009) Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet 41:657–665. doi:10.1038/ng.388
Joubert BR, Lange EM, Franceschini N, Mwapasa V, North KE, Meshnick SR, Immunology NCfHAV (2010) A whole genome association study of mother-to-child transmission of HIV in Malawi. Genome Med 2:17. doi:10.1186/gm138
Julian BA, Gaston RS, Brown WM, Reeves-Daniel AM, Israni AK, Schladt DP, Pastan SO, Mohan S, Freedman BI, Divers J (2016) Effect of replacing race with apolipoprotein L1 genotype in calculation of kidney donor risk index. Am J Trans doi: 10.1111/ajt.14113
Kang SJ et al (2010) Genome-wide association of anthropometric traits in African- and African-derived populations. Hum Mol Genet 19:2725–2738. doi:10.1093/hmg/ddq154
Kasembeli AN et al (2015) APOL1 risk variants are strongly associated with HIV-associated nephropathy in black South Africans. J Am Soc Nephrol 26:2882–2890. doi:10.1681/ASN.2014050469
Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P (2009) Data sharing in genomics—re-shaping scientific practice. Nat Rev Genet 10:331–335. doi:10.1038/nrg2573
Kim S et al (2016) Genetic influences on plasma homocysteine levels in African Americans and Yoruba Nigerians. J Alzheimers Dis 49:991–1003. doi:10.3233/JAD-150651
Knoppers BM (2014) Framework for responsible sharing of genomic and health-related data. HUGO J 8:3. doi:10.1186/s11568-014-0003-1
Kopp JB et al (2011) APOL1 genetic variants in focal segmental glomerulosclerosis and HIV-associated nephropathy. J Am Soc Nephrol 22:2129–2137. doi:10.1681/ASN.2011040388
Ku E et al (2017) Strict blood pressure control associates with decreased mortality risk by APOL1 genotype. Kidney Int 91:443–450. doi:10.1016/j.kint.2016.09.033
Larson C (2013) Inside China's genome factory. https://www.technologyreview.com/s/511051/inside-chinas-genome-factory/. Accessed 02/23/2017
Li MJ, Sham PC, Wang J (2012) Genetic variant representation, annotation and prioritization in the post-GWAS era. Cell Res 22:1505–1508. doi:10.1038/cr.2012.106
Lingappa JR et al (2011) Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. PLoS One 6:e28632. doi:10.1371/journal.pone.0028632
Ma L et al (2016) APOL1 renal-risk variants induce mitochondrial dysfunction. J Am Soc Nephrol. doi:10.1681/asn.2016050567
Malaria Genomic Epidemiological N (2017) Ethics of collaboration https://www.malariagen.net/ethics/ethics-collaboration. Accessed March 10 2017
Malaria Genomic Epidemiology N, Band G, Rockett KA, Spencer CC, Kwiatkowski DP (2015) A novel locus of resistance to severe malaria in a region of ancient balancing selection. Nature 526:253–257. doi:10.1038/nature15390
Manrai AK et al (2016) Genetic misdiagnoses and the potential for health disparities. N Engl J Med 375:655–665. doi:10.1056/NEJMsa1507092
Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT, Bild DE (1995) Prevalence of hypertrophic cardiomyopathy in a general population of young adults Echocardiographic Analysis of 4111 Subjects in the CARDIA Study 92:785–789 doi: 10.1161/01.cir.92.4.785
Mathias RA et al (2016) A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat Commun 7:12522. doi:10.1038/ncomms12522
McDonald M, Townsend A, Cox SM, Paterson ND, Lafreniere D (2008) Trust in health research relationships: accounts of human subjects. J Empir Res Hum Res Ethics 3:35–47. doi:10.1525/jer.2008.3.4.35
Milet J, Sabbagh A, Migot-Nabias F, Luty AJ, Gaye O, Garcia A, Courtin D (2016) Genome-wide association study of antibody responses to Plasmodium falciparum candidate vaccine antigens. Genes Immun 17:110–117. doi:10.1038/gene.2015.59
Moore JX, Chaudhary N, Akinyemiju T (2017) Metabolic syndrome prevalence by race/ethnicity and sex in the United States, National Health and Nutrition Examination Survey, 1988–2012. Prev Chronic Dis 14:E24. doi:10.5888/pcd14.160287
Movahed MR, Strootman D, Bates S, Sattur S (2010) Prevalence of suspected hypertrophic cardiomyopathy or left ventricular hypertrophy based on race and gender in teenagers using screening echocardiography. Cardiovasc Ultrasound 8:54. doi:10.1186/1476-7120-8-54
Mtatiro SN et al (2014) Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania. PLoS One 9:e111464. doi:10.1371/journal.pone.0111464

Mulder NJ et al (2016) H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. Genome Res 26:271–277. doi:10.1101/gr.196295.115

N'Diaye A et al (2011) Identification, replication, and fine-mapping of Loci associated with adult height in individuals of african ancestry. PLoS Genet 7:e1002298. doi:10.1371/journal.pgen.1002298

Need AC, Goldstein DB (2009) Next generation disparities in human genomics: concerns and remedies. Trends Genet 25:489–494. doi:10.1016/j.tig.2009.09.012

Ng MCY et al (2014) Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. PLOS Genetics 10:e1004517. doi:10.1371/journal.pgen.1004517

Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E (2017) Tracing the peopling of the world through genomics. Nature 541:302–310. doi:10.1038/nature21347

Oh SS et al (2015) Diversity in clinical and biomedical research: a promise yet to be fulfilled. PLoS Med 12:e1001918. doi:10.1371/journal.pmed.1001918

Oh SS, White MJ, Gignoux CR, Burchard EG (2016) Making precision medicine socially precise. Take a deep breath. Am J Respir Crit Care Med 193:348–350. doi:10.1164/rccm.201510-2045ED

Panofsky A, Bliss C (2017) Ambiguity and scientific authority: population classification in genomic science. Am Soc Rev 82:59-87. doi:10.1177/0003122416685812

Parsa A et al (2013) APOL1 risk variants, race, and progression of chronic kidney disease. N Engl J Med 369:2183–2196. doi:10.1056/NEJMoa1310345

Peralta CA, Bibbins-Domingo K, Vittinghoff E, Lin F, Fornage M, Kopp JB, Winkler CA (2016) APOL1 genotype and race differences in incident albuminuria and renal function decline. J Am Soc Nephrol 27:887–893. doi:10.1681/asn.2015020124

Petrovski S et al (2011) Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. AIDS 25:513–518. doi:10.1097/QAD.0b013e328343817b

Polychronakos C, Alriyami M (2015) Diabetes in the post-GWAS era. Nat Genet 47:1373–1374. doi:10.1038/ng.3453

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

Price AL, Spencer CCA, Donnelly P (2015) Progress and promise in understanding the genetic basis of common diseases. Proc Royal Soc B Biol Sci 282 doi: 10.1098/rspb.2015.1684

Ramsay M, Sankoh O, study amotA-G, Consortium tHA (2015) African partnerships through the H3Africa Consortium bring a genomic dimension to longitudinal population studies on the continent. Int J Epidemiol. doi:10.1093/ije/dyv187

Ramsuran V et al (2011) Duffy-null-associated low neutrophil counts influence HIV-1 susceptibility in high-risk South African black women. Clin Infect Dis 52:1248–1256. doi:10.1093/cid/cir119

Ray KK et al. (2017) Inclisiran in patients at high cardiovascular risk with elevated LDL cholesterol N Engl J Med 376:1430-1440 doi:10.1056/NEJMoa1615758

Reeves-Daniel AM et al (2011) The APOL1 gene and allograft survival after kidney transplantation. Am J Transplant 11:1025–1030. doi:10.1111/j.1600-6143.2011.03513.x

Roth EM et al (2016) A phase III randomized trial evaluating alirocumab 300 mg every 4 weeks as monotherapy or add-on to statin: ODYSSEY CHOICE I. Atherosclerosis 254:254–262. doi:10.1016/j.atherosclerosis.2016.08.043

Rotimi C et al (2007) Community engagement and informed consent in the International HapMap project Community. Genet 10:186–198. doi:10.1159/000101761

Rotimi C, Mulder N (2014) Data acquisition and data/knowledge sharing in global genomic studies. Appl Transl Genomics 3:109–110. doi:10.1016/j.atg.2014.09.002

Rotimi CN, Jorde LB (2010) Ancestry and disease in the age of genomic medicine. N Engl J Med 363:1551–1558. doi:10.1056/NEJMra0911564

Rotimi CN, Tekola-Ayele F, Baker JL, Shriner D (2016) The African diaspora: history, adaptation and health. Curr Opin Genet Dev 41:77–84. doi:10.1016/j.gde.2016.08.005

Russell GW (2016) China plans leading role in 'precision medicine' field. The University of Hong Kong. http://jmsc.hku.hk/reportinghealth2016/2016/04/06/china-plans-leading-role-in-precision-medicine-field/. Accessed 02/23/2017 2017

Sabatine MS et al (2017) Evolocumab and clinical outcomes in patients with cardiovascular disease. N Engl J Med. doi:10.1056/NEJMoa1615664

Satcher D, Fryer GE Jr, McCann J, Troutman A, Woolf SH, Rust G (2005) What if we were equal? A comparison of the black-white mortality gap in 1960 and 2000. Health Aff (Millwood) 24:459–464. doi:10.1377/hlthaff.24.2.459

Science and inequality (2016) Nature 537:465. doi:10.1038/537465a

Scott LJ et al (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345. doi:10.1126/science.1142382

Sheridan C (2013) Phase 3 data for PCSK9 inhibitor wows. Nat Biotech 31:1057–1058. doi:10.1038/nbt1213-1057

Shriner D, Adeyemo A, Rotimi CN (2011) Joint ancestry and association testing in admixed individuals. PLoS Comput Biol 7:e1002325. doi:10.1371/journal.pcbi.1002325

Shriner D, Tekola-Ayele F, Adeyemo A, Rotimi CN (2014) Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. Sci Rep 4:6055. doi:10.1038/srep06055

Skloot R (2011) The immortal life of Henrietta Lacks, 1st edn. Broadway Paperbacks, New York

Sladek R et al (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881–885. doi:10.1038/nature05616

Tarver-Carr ME, Powe NR, Eberhardt MS, LaVeist TA, Kington RS, Coresh J, Brancati FL (2002) Excess risk of chronic kidney disease among African-American versus white subjects in the United States: a population-based study of potential explanatory factors. J Am Soc Nephrol 13:2363–2370

Tekola-Ayele F et al (2015a) Novel genomic signals of recent selection in an Ethiopian population. Eur J Hum Genet 23:1085–1092. doi:10.1038/ejhg.2014.233

Tekola-Ayele F et al (2015b) Genome-wide association study identifies African-ancestry specific variants for metabolic syndrome. Mol Gen Metab 116:305–313. doi:10.1016/j.ymgme.2015.10.008

Tekola Ayele F et al (2012) HLA class II locus and susceptibility to podoconiosis. N Engl J Med 366:1200–1208. doi:10.1056/NEJMoa1108448

The H3Africa Consortium (2014) Enabling the genomic revolution in Africa: H3Africa is developing capacity for health-related genomics research in Africa. Science (N Y NY) 344:1346–1348. doi:10.1126/science.1251546

Thomson R et al (2014) Evolution of the primate trypanolytic factor APOL1. Proc Natl Acad Sci U S A 111:E2130–E2139. doi:10.1073/pnas.1400699111

Thye T et al (2012) Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat Genet 44:257–259. doi:10.1038/ng.1080

Thye T et al. (2010) Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2 Nat Genet 42:739-741 doi:10.1038/ng.639

Tindana P et al (2015) Community engagement strategies for genomic studies in Africa: a review of the literature. BMC Med Ethics 16:24. doi:10.1186/s12910-015-0014-z

Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W (2010) Genomic research and wide data sharing: views of

prospective participants. Genet Med 12:486–495. doi:10.1097/GIM.0b013e3181e38f9e

Wells JC (2012) Ethnic variability in adiposity, thrifty phenotypes and cardiometabolic risk: addressing the full range of ethnicity, including those of mixed ethnicity. Obes Rev 13(Suppl 2):14–29. doi:10.1111/j.1467-789X.2012.01034.x

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson, H (2013) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42(D1):D1001–D1006. doi:10.1093/nar/gkt1229

Yarborough M, Fryer-Edwards K, Geller G, Sharp RR (2009) Transforming the culture of biomedical research from compliance to trustworthiness: insights from nonmedical sectors. Acad Med 84:472–477. doi:10.1097/ACM.0b013e31819a8aa6

Zhang Y, Li Q, Wang X, Zhou X (2015) China biobanking. In: Karimi-Busheri F (ed) Biobanking in the 21st century. Springer International Publishing, Cham, pp 125–140. doi:10.1007/978-3-319-20579-3_10

Zhao X, Wang P, Tao X, Zhong N (2013) Genetic services and testing in China. J Commun Genet 4:379–390. doi:10.1007/s12687-013-0144-2