REVIEW ARTICLE   OPEN

Evaluating the promise of inclusion of African ancestry populations in genomics

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

The lack of representation of diverse ancestral backgrounds in genomic research is well-known, and the resultant scientific and ethical limitations are becoming increasingly appreciated. The paucity of data on individuals with African ancestry is especially noteworthy as Africa is the birthplace of modern humans and harbors the greatest genetic diversity. It is expected that greater representation of those with African ancestry in genomic research will bring novel insights into human biology, and lead to improvements in clinical care and improved understanding of health disparities. Now that major efforts have been undertaken to address this failing, is there evidence of these anticipated advances? Here, we evaluate the promise of including diverse individuals in genomic research in the context of recent literature on individuals of African ancestry. In addition, we discuss progress and achievements on related technological challenges and diversity among scientists conducting genomic research.

INTRODUCTION

The genomic research landscape has transformed dramatically since the completion of the human genome project in 2003, but focus in the field has been constrained, in terms of worldwide populations included in this research. Representation in genomics studies has been dominated by populations descending from Europe. The concentration of research and resources within one ancestral group has the potential to further perpetuate and even exacerbate known disparities that negatively affect populations with ancestry that is underrepresented in research (see Box 1: "What do we mean by ancestry?"). There has been increasing recognition of the scientific and equity imperatives to increase the representation of diverse populations in genomic research,[1–9] and initiatives are underway to address past shortcomings. As these initiatives are beginning to yield results and more findings from studies including diverse populations are accumulating, it is worthwhile to revisit the promises of this inclusion to evaluate whether the anticipated advances are, indeed, emerging. With a focus on the African continent, consistent with the authors' expertise, we evaluate the evidence to support the expected progress from including diverse individuals in genomic research relative to: novel insights into human biology, improvements in clinical care, and enhanced understanding of health disparities (Table 1). The call for increased diversity in genomic research has been accompanied by the need for overcoming related technological challenges and for greater diversity among those conducting research. We also comment on the progress and achievements on these issues.

For clarity, and to be consistent with recently published guidelines for characterizing ancestry in genomic research,[10] we here use "individuals of African ancestry" to describe sub-Saharan Africans and those who identify as African American or Afro-Caribbean. Similarly, we use "European ancestry individuals" to describe those who identify as "white" from Europe or America. However, the concepts presented may extend directly to all individuals with any degree of a particular ancestry, as ancestry at a genomic locus of interest may have important health effects, regardless of self-identification.

INSIGHTS INTO HUMAN BIOLOGY

A substantial proportion of human genomic variation was left behind during the out-of-Africa migration.[11] Therefore, insights into the implications of this variation can only come from research conducted in those of African ancestry. This reservoir of relatively untested African sequence variation has fostered high hopes regarding the potential for novel findings when better representation in genomic research was achieved.[6] Notably, expected advances based on findings in those of African ancestry are not limited to those of that ancestry, as can be seen with the discovery of variants in *PCSK9* that dramatically reduced low-density lipoprotein cholesterol concentrations. Although these variants were discovered in samples from African Americans, the benefit of drugs targeting *PCKS9* extends beyond ancestry-defined subgroups.[12,13] In general, common complex diseases require the aggregation of large datasets to yield results. Achieving the needed datasets of participants with African ancestry has been a challenge due to the relatively limited numbers and sample sizes of cohorts from diverse populations. Now that studies of African ancestry individuals are increasing in sample size, are we seeing the expected novel results? A recent evaluation of data included in the NHGRI-EBI GWAS Catalog, which is the most complete record of published genome-wide association studies, found that only 2.4% of the individuals included in the catalog were of African ancestry.[10] Even at such a relatively small proportion, the promise of new findings with greater inclusion is clear: they contributed a larger than expected proportion of associations in the catalog (7%).[10]

New initiatives and large consortia with a focus on increasing the representation of diverse populations in research, such as TopMED and PAGE II, are beginning to bear fruit (Fig. 1; Table 2). In addition to these large-scale efforts, the inclusion of individuals with African ancestry in other studies has led to the identification

¹Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. ²Department of Clinical Research and Leadership, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA. *email: rotimic@mail.nih.gov

Table 1. Progress and challenges for achieving scientific promise of diversity and inclusion in genomics.

| Area	Example of recent success	Remaining challenges
Insights into human biology	New loci and new variants within known loci in large projects including African ancestry individuals	Number of included individuals remains relatively low Even with increasing numbers of diverse individuals, preferential analysis and reporting of EUR results, which have achieved even larger sample sizes Because of the diversity among AFR populations and the reduced linkage disequilibrium across the genome, studies of AFR individuals will require even greater sample sizes than EUR to properly interrogate variation present
Improvements in clinical care	Identification of new pharmacogenomic loci relevant for individuals with African ancestry through consortia and new tools	Continued underrepresentation of non-European ancestry individuals in clinically important databases limits diagnostic accuracy for these individuals Need for more pharmacogenomic research in diverse populations Need for more data from which to derive polygenic risk scores specific to ancestry groups
Improved understanding of health disparities	Identification of genetic variation that may be contributing to "racial" differences in disease-relevant traits	Identifying appropriate biomarkers, genetic variants, and/or environmental factors for distinguishing risk instead of race categories
Infrastructure to include diverse voices	Infrastructure development within H3Africa to support the full participation of African scientists in high-level genomic research	Sustainability of current initiatives will require further commitment on the part of funding agencies and governments
Overcoming technological challenges	Development of a genotyping array for improved interrogation of genetic variation present in populations with primarily African ancestry	Analysis of populations of diverse and admixed ancestral backgrounds may require the development of novel techniques
Development of ethical guidelines	Development of ethical guidelines and training modules related to conducting genomic research through the H3Africa initiative	Sustaining mature and informed institutional review boards and ethics committees to address ethical issues that arise with the expansion of genomic research in Africa will require continued efforts
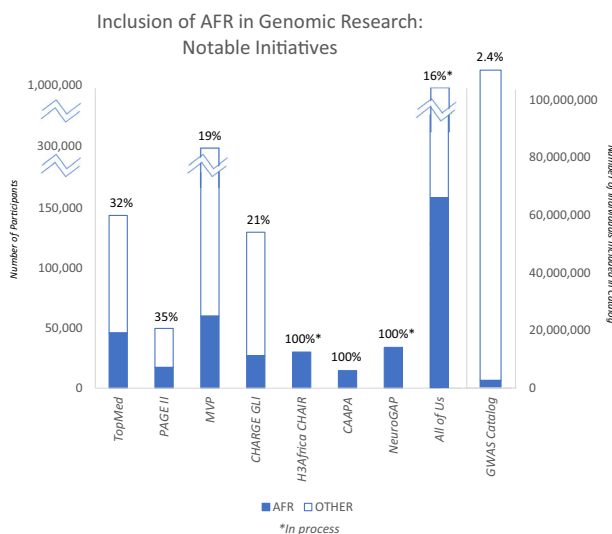
Fig. 1 Inclusion of AFR in genomic research: notable initiatives. Data for figures taken from the following sources: TOPMed,[106] PAGE II,[107] Million Veteran Program (MVP),[108] CHARGE Gene–Lifestyle Interactions (GLI),[109] Cardiovascular H3Africa Innovation Resource (CHAIR),[35] CAAPA,[110] NeuroGAP-Psychosis,[36] All of Us,[23] and the GWAS Catalog[10].

of novel loci in recent studies of obesity,[14] type 2 diabetes,[15,16] metabolic syndrome,[17] skin pigmentation,[18] suicide,[19] multiple sclerosis,[20] cleft palate,[21] and Epstein-Barr virus immune response.[22] The National Institutes of Health's recently launched All of Us Precision Medicine Initiative aims to recruit 1 million Americans, with a particular focus on representing the nation's diversity (http://allofus.nih.gov).[23] With estimates of early recruitment showing up to 75% from groups who are underrepresented

in health research, this effort promises to yield a considerable number of participants of diverse ancestries.[24] Although the development of large, multi-ethnic resources is a clear advance, diligence on the part of researchers, funders, and reviewers is required to ensure that these resources are used to their fullest extent. It is common to exclude non-European ancestry samples from analyses, despite the presence of a sufficient sample size for stratified analysis (such as in UK Biobank),[25] or to limit analyses in these samples to replication or confirmation of findings from the analysis of European ancestry samples.

The above projects largely reflect inclusion of African Americans in genomic research, but representation on the African continent, the birthplace of humanity and home to the greatest genomic diversity on the planet, has been more limited. In response to this need, the Wellcome Trust and the National Institutes of Health have supported African-led genomic research through the Human Heredity and Health in Africa (H3Africa) initiative.[26,27] Now in its 6th year, H3Africa has received over 170 million USD and sponsored nearly 50 research projects.[28] At the time of the writing of this manuscript, H3Africa had reached 54,000 research participants, and convened meetings and workshops with over 2000 attendees.[28] This initiative has produced 382 trainees and 219 publications.[28] While primary results from the H3Africa projects for the study of complex disease traits have not yet been published, they are poised to make substantial contributions to increasing diversity in genomic research given the following enrollment numbers: a study of over 10,000 individuals for cardiometabolic disease risk (AWI-GEN),[29,30] a study of 6000 T2D cases and 6000 controls (H3Africa T2D Study),[31] a study of kidney disease ($n = 4000$ cases and 4000 controls) (H3Africa Kidney Disease Research Network),[32] a study of 3000 stroke cases and 3000 controls (SIREN),[33] a study of over 11,000 women on Human Papilloma Virus and cervical cancer risk (African Collaborative Center for Microbiome and Genomics Research),[34] and a combined resource for cardiovascular studies with over 30,000

Table 2. Some insights into human biology from initiatives prioritizing inclusion of participants of African ancestry.

Initiative	Novel loci and novel associations within known loci[a]
Trans-Omics for Precision Medicine (TOPMed)	Serum lipids[111] Lipoprotein(a)[112] Chronic obstructive pulmonary disease[113] Rh antigens[114]
Population Architecture using Genomics and Epidemiology II (PAGE II) study	Adiposity traits[115–117] Serum lipids[118] Glycemic traits[119] QT interval[120] Age of reproductive events among women[121] C-reactive protein[122] Blood pressure[123] Complex traits[107]
The Million Veteran's Program (MVP)	Serum lipids[124] Blood pressure[108]
CHARGE Gene–Lifestyle Interactions[125]	Blood pressure[109,126,127] Serum lipids[128–130]
The Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA)	Understanding of genetic diversity among individuals with African ancestry[131,132] asthma risk[133]
MalariaGEN	Severe malaria risk[134] Cerebral malaria and alphathalassemia[135] Malaria resistance[136] G6PD and malaria risk[137] Population genetics[138,139]

[a]Including novel loci for the studied trait as well as novel associations within known loci

participants (CHAIR).[35] Another noteworthy initiative is the Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis) project, a study of the genetics of schizophrenia and bipolar disorder recruiting 34,000 participants from Ethiopia, Kenya, South Africa, and Uganda.[36] These genomic studies and associated consortia will undoubtedly greatly enhance the detection of novel loci for complex disease risk.

IMPROVEMENTS IN CLINICAL CARE

A key motivation for conducting genomic research is to facilitate improvements in clinical care. A promising outcome of the success of genome-wide association studies (GWAS) of common diseases is the development of polygenic risk scores (PRS): scores that can be calculated for an individual based on the presence or absence of risk variants identified in large GWAS studies. The promise of these scores for clinical use is gaining increased attention.[37,38] For instance, risk scores developed using known risk variants have been able to identify the 8.0% of the population with more than a three-fold increased risk of coronary artery disease.[39] However, PRS depend on underlying genomic research, which has predominantly been conducted in European ancestry individuals. European ancestry individuals are overrepresented in PRS research at 460% of what it would be if representation matched the proportion of the worldwide population. In contrast, African ancestry individuals are underrepresented (only 17% of expected proportion based on worldwide population).[40] As a result, these PRS represent only a subset of the global human population and have limited portability across ancestries.[1,41] Some of this lack of portability reflects differences in characteristics of the genome that have accumulated as a result of migration patterns. For
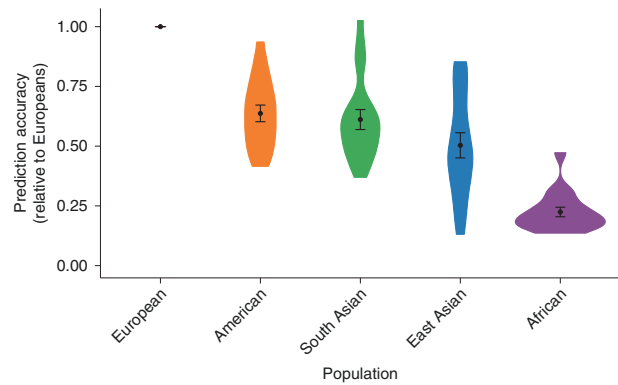

Fig. 2 Prediction accuracy relative to European-ancestry individuals across 17 quantitative traits and 5 continental populations in the UKBB: All phenotypes shown here are quantitative anthropometric and blood-panel traits. Prediction target individuals do not overlap with the discovery cohort and are unrelated. Violin plots show distributions of relative prediction accuracies, points show mean values, and error bars show s.e.m. values. [Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Nature Genetics, "Clinical use of current polygenic risk scores may exacerbate health disparities", Alicia R. Martin et al.[42]].

instance, as a result of genetic drift and selection pressures, the presence and frequency of alleles vary across ancestry groups.[11] PRS can only capture the genetic associations observed in the population in which the underlying GWAS is conducted. An effect of variants that are not sufficiently frequent (or present) in that analysis may influence trait outcomes in a population of a different ancestry, yet the ancestry-unmatched PRS will not reflect those effects and the PRS will give less accurate results. Additionally, population history has led to a generally reduced linkage disequilibrium structure across the genome among those of African ancestry. Although an untested causal variant may be the same across populations, the variant which tags that effect may differ across ancestry groups. In this case, the risk associated with a particular variant may be misattributed and also lead to ancestry-unmatched PRS inaccuracies. The underperformance of the PRS when discovery and target populations are mismatched in ancestry has been demonstrated when PRS constructed based on GWAS findings in European ancestry individuals were tested in those of African ancestry. In UK Biobank data, prediction accuracy was 4.5-fold lower in African ancestry samples compared to European ancestry samples[42] (Fig. 2). Similarly, the median effect size for PRS for a variety of traits in African ancestry samples was only 42% of that in matched European ancestry samples in an analysis of published PRS studies.[40] In contrast, risk scores based on data derived in African Americans have been demonstrated to perform significantly better in African Americans than risk scores based on studies predominantly including those of European ancestry.[43] The PRS could be considered as a new disease biomarker to be evaluated alongside more traditional measures of risk. It is common that the distribution of disease biomarkers varies across ancestries, often leading to adjustments in biomarker calculations (for instance, predicted values for spirometry or estimated glomerular filtration rate). However, PRS uniformly performs more poorly in individuals with African ancestry (and in all understudied populations), and minor ancestry-specific adjustments, such as have been used with other biomarkers, will not be able to make up for this underperformance.[42] The complexities of PRS in the context of ancestry and the potential to exacerbate health disparities has been addressed in-depth in recent articles.[40,42] Importantly, in addition to limitations in transferability conferred by ancestry-related genomic differences, there can be substantial variation in the performance of PRS for complex genetic

Box 1 What do we mean by ancestry?

In the context of this paper, ancestry is defined using genetic variants based on the distribution of those variants in worldwide populations. An individual's genome is a mosaic of segments from different ancestral populations. By comparing segments of DNA with the distribution of genetic variants in worldwide populations, it is possible to determine the likely "parental" or source population for each segment of DNA, indicating a component of the individual's overall ancestry. Using this process to interrogate an individual's entire genome, the proportion of an individual's genomic inheritance from specific ancestral populations can be estimated. Importantly, someone's genetic ancestry may have little to do with their identity in terms of race or culture. An individual with a relatively small proportion of African ancestry may not self-identify as "Black" or "African American", yet they may have African ancestry at a specific region of the genome where that ancestry may confer risk, for instance, by carrying a pharmacogenomic risk allele that is more prevalent among those with African ancestry. Based on our most recent analysis of 282 global samples, there are at least 21 ancestries.[139] Notably, mixed ancestry is the norm across samples and continents, and the vast majority of individuals (an estimated 97.3%) show ancestral heterogeneity.[139] Finally, individuals' ancestry estimates may change over time as our reference datasets become more diverse.

traits based on methodological decisions[40] and on environmental context.[44] If PRS do become useful clinically, the limitations of applying data from a subset of the global population to all global populations are clear. Disparities in clinical care will persist for those with non-European ancestry: medical technology will be less capable of identifying risks, targets, and potential interventions for the majority of patients globally that represent greater genomic heterogeneity. While advances in understanding and adjusting for biases when PRS are applied across ancestries is useful, real progress in PRS for those of African ancestry is tied to increasing the number of large GWAS of African ancestry individuals on which to base these PRS.

Another important area of genomic research in which increased diversity is vital for equitable clinical care is pharmacogenomics: identifying genetic variants that may influence an individual's response to drugs. Depending on the pharmacogenomic variant carried by an individual, treatment outcomes could range from a lack of effect to potentially life-threatening adverse effects. Using genomic data to protect patients from harm or improve their treatment is a key motivation for precision medicine. As the genetic variants present in individuals may vary by their ancestry, limited pharmacogenomic research in diverse populations puts individuals with ancestry from an underrepresented group at more risk of an unanticipated drug response than those in better-studied populations. The African American Cardiovascular Pharmacogenetic Consortium (ACCOuNT) has been established to address this gap and aims to discover novel pharmacogenetic variants in African Americans and to incorporate these variants into clinical recommendations.[45] Recent efforts focusing on those of African ancestry have identified novel loci that may help to explain the diminished response to inhaled corticosteroids in African Americans with asthma,[46–48] though the low number of available studies with individuals of African ancestry to confirm findings was acknowledged as a continuing limitation.[47] The first GWAS of neutrophil counts during clozapine treatment (for schizophrenia) conducted in individuals of African ancestry identified a key role for the well-known African ancestry-specific Duffy allele in risk of neutropenia upon treatment, thereby improving our understanding of the differential rates of discontinuation of clozapine by ethnicity.[49] Major warfarin-associated bleeding occurs more commonly among those with African ancestry, though genetic risk factors associated with this serious outcome had not been well-studied in these individuals. A recent case-control study among individuals of African ancestry[50] was able to identify a locus that associated with more than eight times greater risk of major warfarin-associated bleeding. Risk variants at this locus are relatively common among individuals with African

ancestry (minor allele frequency = 0.05) but absent in populations without any African ancestry. Also, potentially relevant warfarin dosing haplotypes, common in Africans but rare in other ancestry populations, have not been systematically evaluated.[51,52] Although pharmaceutical drugs can be prohibitively expensive in many African countries, the potential for pharmacogenomic discoveries to increase safety and effectiveness should not be overlooked. Some countries, for instance, consider research findings related to pharmacogenomic alleles that have ramifications for their countries prior to selecting available medicines for use locally. As the prevalence of a CYP2D6 duplication that causes serious adverse outcomes with codeine use is 30% among Ethiopians, and broad genotyping is not feasible, the use of codeine has been banned in Ethiopia.[53] It is important to note, however, that because of the great genetic diversity within African ancestry individuals, care must be taken not to group those of African ancestry into broad categories with the expectation of genetic similarity. For instance, HLA-B*5701 is associated with increased incidence of abacavir hypersensitivity, a life-threatening condition among HIV patients being treated with this drug. The frequency of this variant ranges from virtually absent among the Yoruba in Nigeria to 13.6% among the Masai in Kenya. Among the Luhya, also in Kenya, the variant is only present in 3.3% of individuals. In this case, neither "African" nor "Kenyan" are sufficiently precise to describe risk at this pharmacogenomic locus.[54] While there are examples of recent progress, continued pharmacogenomics research is needed to form a clearer picture of the distribution of pharmacogenomics variants across diverse African populations.[3,54]

Inclusion and diversity are also of critical importance for the use of genomic information in clinical diagnosis. Translation of genomics to clinical practice depends upon what genomic studies have been conducted. Analyses of important databases for reporting genomic results illustrate the continuing underrepresentation of populations with significant ancestry outside of Europe: individuals of European ancestry constitute 55.8% of the observations in the Genome Aggregation Database (gnomAD);[2] and the Genome-wide Association Study Catalog and the database of Genotypes and Phenotypes both have significantly fewer studies of African, Latin American, and Asian ancestral populations compared to European ancestry populations.[2–4] This current reality highlights the comparatively reduced ability for clinicians to assess clinically significant variants in individuals with diverse ancestry, particularly when those variants are not shared across populations. Illustratively, genetic testing for cardiomyopathy in European ancestry individuals is more likely to positively identify pathogenic variants than testing conducted in underrepresented minorities in the US, while genetic testing in underrepresented minorities is more likely to be inconclusive.[5] African ancestry individuals were also more likely to have variants that were misclassified as pathogenic for hypertrophic cardiomyopathy because of a lack of ancestry-matched controls for comparison.[55]

IMPROVED UNDERSTANDING OF HEALTH DISPARITIES

Although the social and economic factors are the most significant contributors to health disparities, genomic research has the potential to help unravel disparities in health outcomes that are inappropriately attributed to race. The discovery of kidney disease risk variants in the APOL1 gene that are found predominantly in individuals with African ancestry demonstrated this potential. These variants are thought to have risen in frequency among some African populations through selection pressure due to protection from Human African Trypanomiasis, yet these variants are associated with a marked increase in the risk of kidney diseases of varying etiologies.[56] Well-known differences in kidney transplant outcomes that had been attributed to race were found to be better described by APOL1 genotypes. For instance, kidney

transplant survival is poorer when the donor is African American compared to European American, but after accounting for those African American donors with the high-risk genotype, African ancestry is no longer associated with such outcomes.[57–59] While stories like that of *APOL1* and kidney disease are likely uncommon, other health disparities are being investigated for genomic contributions, fueled by the increasing availability of genomic data for diverse populations. A *G6PD* variant that is common among those with African ancestry (but not those without African ancestry) leads to decreased HbA1C levels, irrespective of blood glucose, which may lead to an estimated 2% of African Americans with T2D to remain unidentified by HbA1c screening.[60] A separate study reported a role of genetic factors in the difference in distribution of HbA1c by ancestry, also highlighting an association between the variant for hemoglobin S ("sickle cell") in this trait.[61] Additional discoveries have been made for the genetic factors underlying differences in the frequency of certain cancers by ancestry. For instance, in a pan-cancer analysis,[62] higher degrees of chromosomal instability were observed among African American (vs. European American) patients with breast, head and neck, and endometrial cancers, while less chromosomal instability was observed in African American patients with kidney cancers.[62] A study of triple-negative breast cancer identified different mutation profiles in the tumors of African American vs. European American patients, informing the higher prevalence of this subtype of breast cancer observed among African Americans.[63] A recent study of prostate cancer disparities identified a protective allele that rose to high frequency in Europe due to linkage with a locus for skin pigmentation under selection pressure.[64] The distribution of D-dimer, a cardiovascular disease risk biomarker, is higher among African Americans than among individuals with only European ancestry, and some genetic factors have been identified that contribute to that difference.[65] Outcomes with treatment for chronic infection with hepatitis C virus vary, with African Americans having lower rates of treatment response compared to other ethnic groups. The C allele of rs12979860, a variant near *IL28B*, was found to be associated with improved treatment response among African Americans, European Americans, and Hispanics, and the inter-ancestry differences in frequency of this allele appear to explain approximately half of the disparities in treatment response rate.[66] This variant was also associated with spontaneous clearance of the hepatitis C virus, and worldwide allele frequencies suggest that it may have been under selection in human history.[67] Fortunately, newer treatment options have increased response to ~90% of patients, regardless of ancestral background.[68] With continuing work on understanding how genetic factors contribute to health disparities, inadequate characterizations based on race can be replaced by screening for relevant markers, with improved targeting and the potential for novel biological insights.

OVERCOMING TECHNOLOGICAL CHALLENGES

To efficiently capture the wide range of genomic diversity among African participants, genotyping tools must be optimized for interrogating African genomes. GWAS arrays developed using genetic information derived largely from populations of European ancestry are inefficient when used within the context of research involving non-European populations, particularly the diverse groups that comprise Africa. There have been several notable efforts to improve the genomic diversity represented in commercially available genotyping arrays, including the Multi-Ethnic Genotyping Array (MEGA) produced by Illumina in partnership with the PAGE and CAAPA Consortia[69] and the Affymetrix Axiom PanAFR Array, with a focus on coverage of 1000 Genomes AFR populations.[70] More recently, the H3Africa Consortium and Illumina have developed a GWAS array with 2.5 million markers

based on whole genome-sequencing data from thousands of samples originating from African research participants.[3] This genotyping array better represents and reflects the range of common variants in populations across the continent, facilitating the interrogation of variants that had not been previously evaluated as they were not captured by previous tools.

After collection of genome-wide genotype data, it is common practice to impute additional genotypes prior to analysis. Accurate outcomes for imputation are dependent on the selection of the appropriate reference panel, as imputation algorithms incorporate linkage disequilibrium patterns, which vary across ancestries. Due to the generally reduced linkage disequilibrium and greater genetic diversity among African ancestry individuals, imputation accuracy using a comparably sized reference panel is reduced for those of African vs. other ancestries.[71] Additionally, the reference panels publicly available for imputing African ancestry samples have been limited, predominated by the 1000 Genomes Project data. The situation has improved, however, with additional reference panels that include African ancestry individuals. High performance and accuracy for imputation with African Americans was observed for reference panels from 1000 Genomes,[11] the Haplotype Reference Consortium (HRC, which includes 1000 Genomes data and a large number of predominantly European ancestry samples[72]), and the Consortium on Asthma among African-Ancestry Populations in the Americas (CAAPA[73]).[74] These reference panels are available at both of the commonly used publicly available imputation servers: the Sanger Imputation Service (1000 Genomes and HRC: https://www.sanger.ac.uk/science/tools/sanger-imputation-service) and the Michigan Imputation Server (1000 Genomes, HRC, and CAAPA: https://imputationserver.sph.umich.edu/index.html#!). Additionally, the Michigan Imputation Server is preparing to offer a reference panel using TOPMed data. The African Genome Resource (AGR), available through the Sanger Imputation Service, is currently the largest publicly available reference panel of African ancestry individuals, including all 1000 Genomes individuals and an additional ~2000 samples from Uganda and ~100 samples each from 5 populations in Ethiopia, Egypt, Namibia, and South Africa. Notably, imputation using the AGR and the 1000 Genomes reference panel was shown to achieve high quality and accuracy even in the presence of five-way admixture in a South African population.[75] Additionally, it is anticipated that data from H3Africa will be made available as an imputation resource.[76]

As the price of sequencing technologies continue to decrease[77] and bioinformatic advances reduce the time and expertise burdens associated with processing these data, it is possible that sequencing may become more common. Importantly, the ancestry biases present in GWAS arrays and imputation are not a concern when using sequencing technology.

INFRASTRUCTURE TO INCLUDE DIVERSE VOICES

The predominance of European ancestry individuals is observed in those represented in genomic research as well as in the identities of those conducting research.[7,78–80] In the US, for example, the proportion of those in the academic doctoral workforce (for science, engineering, and health) who are African American has only increased from 5.8% in 1997 to 8.9% in 2017,[78] although African Americans represent 13.4% of the population.[81] Unfortunately, this lack of diversity among genomic researchers leads to the loss of perspectives that are important in developing hypotheses as well as directing research among diverse individuals, and disparity in terms of research leadership and output among African scientists remains.[82] Sustained genomic research capacity in any setting relies on engaged investigators and frameworks that can support the development of datasets relevant to the local population, infrastructure for computation and analysis, and expertise for the translation and implementation

of genomics and bioinformatics information into clinical practice. African countries have historically been challenged by poor research infrastructure and limited training and professional opportunities.[27] In appreciation of this obstacle faced by African researchers, and with an eye toward the elevation of researchers along with the increased inclusion of African samples, sustained capacity-building is a keystone of the H3Africa Initiative.[83] NeuroGAP has also established the Global Initiative for Neuropsychiatric Genetics Education in Research (GINGER), which will support the training of early-career investigators from the countries in which research is being conducted (Ethiopia, Kenya, South Africa, and Uganda).[84] The topics for training are determined through the participation of the fellows and their African mentors, so that their needs shape their experiences.[84]

H3Africa supports the development of large collaborations across multiple centers, leverages synergies between studies, and provides critical training to African researchers. The pan-African bioinformatics network, H3ABioNet, was designed to serve the consortium and assists with data quality control and analysis.[85] In addition, it has advanced infrastructure and professional skill development through workshops and meetings and accredited enrichment programs to advance skills related to the conduct of genome-wide association studies and the analysis of next generation sequencing data. Investigators have learned how to conduct GWAS and analyze genome-sequencing data. H3ABioNet has also established infrastructure for computing at different institutions across the continent. Through its monitoring and training efforts, the network has provided bioinformaticians, researchers, medical professionals, systems administrators, and software developers with the tools to utilize and maintain the computing infrastructure.[86]

One of the ways that the emergence of the voices of African researchers may be observed is in the focus on diseases and traits of significance to Africa among new Africa-led work. For instance, H3Africa-funded projects have supported work on Human African Trypanomiasis[87–92] and sickle cell disorders.[93–95] H3Africa provides a model that could be useful to include diversity in leadership in other parts of the world, including in America.

DEVELOPMENT OF ETHICAL GUIDELINES

As the African genomic research environment expands to examine genetic, environmental, and other contributions to health and diseases, new and familiar ethical questions and concerns will arise that warrant attention to African perspectives. Investigations that employ emerging research tools (e.g., mobile health applications, sensors, wearable devices), for instance, must consider concerns related to genomic research that incorporates behavioral and physiological data.[96] For instance, an individual's ability to consent may be complicated in some African cultures by expectations of involvement of family and community in decision-making,[97] as has been observed in indigenous people of the Americas.[98] Ethical guidelines for these types of scenarios may need to be reevaluated as genomic research capacity grows within African contexts and produces researchers, institutions, and findings that are more accessible locally. Importantly, African countries must have the opportunity to influence the policies for the design and dissemination of genomic research for collaboration to be sustainable. Concerns over parachute or safari research remain where African researchers are left out of the full research process, invited to collaborate only as is useful for sample collection, but without any benefit of the research returning to support African researchers or infrastructure. Community engagement is recognized as a critical strategy for addressing such challenges, understanding research priorities, respecting cultural concerns, and countering concerns of exploitation.[99] Others have highlighted the importance of tailoring community engagement

strategies to low resource settings, including members of disadvantaged groups in the engagement process, and structuring deliberations in ways that minimize power differentials in community settings.[100] Currently there are collaborations between African and non-African researchers "rooted" with a firm foundation within Africa that emphasize capacity building and shared scientific contribution, such as the Viral Hemorrhagic Fever Consortium (www.vhfc.org) and the African Center of Excellence for Genomics of Infections Disease (acegid.org).[101] H3Africa has the formation of equitable collaborations, African researcher-led science, capacity building, and infrastructure development among its central tenets.[76,83] Capacity building on the continent addresses concerns related to exploitation and parachute research through concomitant ethics and governance mechanisms, such as those created by H3Africa. Additional questions related to benefit sharing, privacy, confidentiality, and group harm should also be examined with emphasis on concerns specific to African populations.[102,103] To address these concerns, H3Africa has provided an excellent foundation by developing guidelines and ethics training modules for ethics committees on the continent that will assess ethical concerns.[104]

Given the importance of data sharing to genomic research, genotype and phenotype data from the H3Africa research studies are stored in the European Genome Phenome Archive (EGA) but access to data and biological samples is determined by an independent H3Africa Data and Biospecimen Access Committee (DBAC) comprised of mostly African members who consider the informed consent process and stipulations provided by the original ethics committees who approved the studies. Through collective action, members of H3Africa have developed ethical guidelines related to data sharing and processing, consent for the sharing and uses of African samples and data, and issues related to study coordination and the shipping of samples externally.[86] Example guidelines include rules and embargos that are sensitive to local infrastructural challenges, thus according to African investigators relatively sufficient time to publish the findings of their primary research questions first.[26,105] The training modules and guidelines provide a framework and foundation for existing and future ethics committee members. Strategies to ensure that ethics committee members remain engaged, well-informed, and committed to the committee will be critical to ongoing developments related to ethical guidance.

CONCLUSION

The need for inclusion of individuals of diverse ancestral backgrounds and identities in genomic research has long been understood, both from the perspective of scientific necessity and equity. Recent initiatives have embraced this call for diversity, including H3Africa, TopMed, CAAPA, PAGE II, and the Million Veteran Program, with more projects, such as large-scale H3Africa projects and the All of Us study seeking to represent US diversity, underway. As results from these and other efforts are reaching the literature, it is warranted to evaluate whether the anticipated benefits of the inclusion of diverse populations in genomics are being realized. Indeed, the increased number of participants with African ancestry is facilitating the discovery of novel genomic loci and establishing new variation of interest within known genomic loci. Importantly, the focus on increasing the number of African ancestry individuals in research is occurring alongside efforts to increase capacity and supporting infrastructure for diverse researchers, and to do so equitably. Clinical and pharmacogenomic studies are improving the data available to accompany integrated social and environmental research on the underlying causes of health disparities. Despite these advances, attaining meaningful representation of diverse populations across the breadth of genomic research that is adequately powered to inform health disparities research will require sustained and continuing efforts.

REFERENCES

1. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
2. Popejoy, A. B. et al. The clinical imperative for inclusivity: race, ethnicity, and ancestry (REA) in genomics. *Hum. Mutat.* **39**, 1713–1720 (2018).
3. Rotimi, C. N. et al. The genomic landscape of African populations in health and disease. *Hum. Mol. Genet.* **26**, R225–R236 (2017).
4. Landry, L. G., Ali, N., Williams, D. R., Rehm, H. L. & Bonham, V. L. Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. *Health Aff.* **37**, 780–785 (2018).
5. Landry, L. G. & Rehm, H. L. Association of racial/ethnic categories with the ability of genetic tests to detect a cause of cardiomyopathy. *JAMA Cardiol.* **3**, 341–345 (2018).
6. McClellan, J. M., Lehner, T. & King, M.-C. Gene discovery for complex traits: lessons from Africa. *Cell* **171**, 261–264 (2017).
7. Hindorff, L. A. et al. Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19**, 175 (2017).
8. Martin, A. R., Teferra, S., Möller, M., Hoal, E. G. & Daly, M. J. The critical needs and challenges for genetic architecture studies in Africa. *Curr. Opin. Genet. Dev.* **53**, 113–120 (2018).
9. Gurdasani, D., Barroso, I., Zeggini, E. & Sandhu, M. S. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* **20**, 520–535 (2019).
10. Morales, J. et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol.* **19**, 21 (2018).
11. The Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68 (2015).
12. Murphy, S. A. et al. Effect of the PCSK9 inhibitor evolocumab on total cardiovascular events in patients with cardiovascular disease: a prespecified analysis from the FOURIER trial. *JAMA Cardiol.* **4**, 613–619 (2019).
13. Sabatine, M. S. et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N. Engl. J. Med.* **376**, 1713–1722 (2017).
14. Chen, G. et al. Genome-wide analysis identifies an african-specific variant in SEMA4D associated with body mass index. *Obesity* **25**, 794–800 (2017).
15. Chen, J. et al. Genome-wide association study of type 2 diabetes in Africa. *Diabetologia* **62**, 1204–1211 (2019).
16. Adeyemo, A. A. et al. ZRANB3 is an African-specific type 2 diabetes locus associated with beta-cell mass and insulin response. *Nat. Commun.* **10**, 3195 (2019).
17. Tekola-Ayele, F. et al. Genome-wide association study identifies African-ancestry specific variants for metabolic syndrome. *Mol. Genet. Metab.* **116**, 305–313 (2015).
18. Crawford, N. G. et al. Loci associated with skin pigmentation identified in African populations. *Science* **358**, eaan8433 (2017).
19. Levey, D. F. et al. Genetic associations with suicide attempt severity and genetic overlap with major depression. *Transl. Psychiatry* **9**, 22 (2019).
20. Chi, C. et al. Admixture mapping reveals evidence of differential multiple sclerosis risk by genetic ancestry. *PLoS Genet.* **15**, e1007808 (2019).
21. Butali, A. et al. Genomic analyses in African populations identify novel risk loci for cleft palate. *Hum. Mol. Genet.* **28**, 1038–1051 (2018).
22. Sallah, N. et al. Whole-genome association study of antibody response to Epstein-Barr virus in an African population: a pilot. *Glob. Health Epidemiol. Genom.* **2**, e18–e18 (2017).
23. Precision Medicine Initiative (PMI) Working Group. The Precision Medicine Initiative Cohort Program—Building a Research Foundation for 21st Century Medicine (2015).
24. Late, M. Minority, underrepresented groups working to bring diversity to research: all of Us partners engage communities. *Nation's Health* **48**, 17–17 (2019).
25. Manolio, T. A. Using the data we have: improving diversity in genomic research. *Am. J. Hum. Genet.* **105**, 233–236 (2019).
26. Consortium, H. A. Enabling the genomic revolution in Africa. *Science* **344**, 1346 (2014).
27. Chu, K. M., Jayaraman, S., Kyamanywa, P. & Ntakiyiruta, G. Building research capacity in Africa: equity and global health collaborations. *PLoS Med.* **11**, e1001612 (2014).
28. https://h3africa.org/ (2019).
29. Ramsay, M. et al. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob. Health Epidemiol. Genom.* **1**, e20–e20 (2016).
30. Ali, S. A. et al. Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Glob. Health Action* **11**, 1507133–1507133 (2018).
31. Ekoru, K. et al. H3Africa multi-centre study of the prevalence and environmental and genetic determinants of type 2 diabetes in sub-Saharan Africa: study protocol. *Glob. Health Epidemiol. Genom.* **1**, e5–e5 (2016).
32. Osafo, C. et al. Genomic approaches to the burden of kidney disease in Sub-Saharan Africa: the Human Heredity and Health in Africa (H3Africa) Kidney Disease Research Network. *Kidney Int.* **90**, 2–5 (2016).
33. Akpalu, A. et al. Phenotyping stroke in sub-Saharan Africa: Stroke Investigative Research and Education Network (SIREN) phenomics protocol. *Neuroepidemiology* **45**, 73–82 (2015).
34. Adebamowo, S. N. et al. Cohort profile: African Collaborative Center for Microbiome and Genomics Research's (ACCME's) Human Papillomavirus (HPV) and Cervical Cancer Study. *Int. J. Epidemiol.* **46**, 1745–1745j (2017).
35. Owolabi, M. O. et al. Data resource profile: Cardiovascular H3Africa Innovation Resource (CHAIR). *Int. J. Epidemiol.* **48**, 366–367g (2018).
36. Stevenson, A. et al. Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya, South Africa and Uganda. *BMJ Open* **9**, e025469–e025469 (2019).
37. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392 (2016).
38. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
39. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
40. Duncan, L. et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).
41. 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 Treat. Diagn.* **2**, 1–6 (2018).
42. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
43. Onengut-Gumuscu, S. et al. Type 1 Diabetes risk in African-ancestry participants and utility of an ancestry-specific Genetic Risk Score. *Diab. Care* dc181727, https://doi.org/10.2337/dc18-1727 (2019).
44. Mostafavi, H., Harpak, A., Conley, D., Pritchard, J. K. & Przeworski, M. Variable prediction accuracy of polygenic scores within an ancestry group. *bioRxiv*, 629949. Preprint at https://doi.org/10.1101/629949 (2019).
45. Friedman, P. N. et al. The ACCOuNT Consortium: a model for the discovery, translation and implementation of precision medicine in African Americans. *Clin. Transl. Sci.* https://doi.org/10.1111/cts.12608 (2018).
46. Levin, A. M. et al. Integrative approach identifies corticosteroid response variant in diverse populations with asthma. *J. Allergy Clin. Immunol.* https://doi.org/10.1016/j.jaci.2018.09.034 (2018).
47. Mak, A. C. Y. et al. Whole-genome sequencing of pharmacogenetic drug response in racially diverse children with asthma. *Am. J. Respir. Crit. Care Med.* **197**, 1552–1564 (2018).
48. Hernandez-Pacheco, N. et al. Genome-wide association study of inhaled corticosteroid response in admixed children with asthma. *Clin. Exp. Allergy* https://doi.org/10.1111/cea.13354 (2019).
49. Legge, S. E. 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* https://doi.org/10.1038/s41380-018-0335-7 (2019).
50. De, T., Alarcon, C. & Hernandez, W. et al. Association of genetic variants with warfarin-associated bleeding among patients of african descent. *JAMA* **320**, 1670–1677 (2018).
51. Ndadza, A. et al. The genetics of warfarin dose–response variability in Africans: an expert perspective on past, present, and future. *OMICS* **23**, 152–166 (2019).
52. Ramos, E. et al. Pharmacogenomics, ancestry and clinical decision making for global populations. *Pharmacogenom. J.* **14**, 217 (2013).
53. Baker, J. L., Shriner, D., Bentley, A. R. & Rotimi, C. N. Pharmacogenomic implications of the evolutionary history of infectious diseases in Africa. *Pharmacogenom. J.* **17**, 112–120 (2017).
54. Rotimi, C. N. & Jorde, L. B. Ancestry and disease in the age of genomic medicine. *N. Engl. J. Med.* **363**, 1551–1558 (2010).
55. Manrai, A. K. et al. Genetic misdiagnoses and the potential for health disparities. *N. Engl. J. Med.* **375**, 655–665 (2016).

56. Freedman, B. I., Limou, S., Ma, L. & Kopp, J. B. APOL1-associated nephropathy: a key contributor to racial disparities in CKD. *Am. J. Kidney Dis.* **72**, S8–S16 (2018).

57. Reeves-Daniel, A. M. et al. The APOL1 gene and allograft survival after kidney transplantation. *Am. J. Transplant.* **11**, 1025–1030 (2011).

58. Freedman, B. I. et al. Apolipoprotein L1 gene variants in deceased organ donors are associated with renal allograft failure. *Am. J. Transplant.* **15**, 1615–1622 (2015).

59. Doshi, M. D. et al. APOL1 Genotype and renal function of Black living donors. *J. Am. Soc. Nephrol.* **29**, 1309–1316 (2018).

60. Wheeler, E. et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med.* **14**, e1002383 (2017).

61. Hivert, M.-F. et al. Genetic ancestry markers and difference in A1c between African-American and White in the Diabetes Prevention Program. *J. Clin. Endocrinol. Metab.* jc.2018-01416-jc.02018-01416, https://doi.org/10.1210/jc.2018-01416 (2018).

62. Yuan, J. et al. Integrated analysis of genetic ancestry and genomic alterations across cancers. *Cancer Cell* **34**, 549–560.e549 (2018).

63. Chang, C.-S., Kitamura, E., Johnson, J., Bollag, R. & Hawthorn, L. Genomic analysis of racial differences in triple negative breast cancer. *Genomics.* https://doi.org/10.1016/j.ygeno.2018.10.010 (2018).

64. Lachance, J. et al. Genetic hitchhiking and population bottlenecks contribute to prostate cancer disparities in men of African descent. *Cancer Res.* **78**, 2432 (2018).

65. Raffield, L. M. et al. D-Dimer in African Americans: whole genome sequence analysis and relationship to cardiovascular disease risk in the Jackson Heart Study. *Arterioscler. Thromb. Vasc. Biol.* **37**, 2220–2227 (2017).

66. Ge, D. et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399 (2009).

67. Thomas, D. L. et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798 (2009).

68. Vutien, P., Hoang, J., Brooks, L. Jr., Nguyen, N. H. & Nguyen, M. H. Racial disparities in treatment rates for chronic Hepatitis C: analysis of a population-based cohort of 73,665 patients in the United States. *Medicine* **95**, e3719–e3719 (2016).

69. Bien, S. A. et al. The future of genomic studies must be globally representative: perspectives from PAGE. *Annu. Rev. Genom. Hum. Genet.* **20**, null (2019).

70. *Data Sheet, Axiom Genome-Wide Pan-African Array Set.* http://tools.thermofisher.com/content/sfs/brochures/axiom_panafr_arrayplate_datasheet.pdf.

71. Huang, L. et al. Genotype-imputation accuracy across worldwide human populations. *Am. J. Hum. Genet.* **84**, 235–250 (2009).

72. Haplotype Reference Consortium et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279 (2016).

73. Mathias, R. A. et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat. Commun.* **7**, 12522 (2016).

74. Vergara, C. et al. Genotype imputation performance of three reference panels using African ancestry individuals. *Hum. Genet.* **137**, 281–292 (2018).

75. Schurz, H. et al. Evaluating the accuracy of imputation methods in a five-way admixed population. *Front. Genet.* **10**, https://doi.org/10.3389/fgene.2019.00034 (2019).

76. Mulder, N. et al. H3Africa: current perspectives. *Pharmacogenom. Personalized Med.* **11**, 59–66 (2018).

77. Wetterstrand, K. A. *DNA Sequencing Costs: Data.* https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data (2019).

78. National Science Foundation. *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019* (National Center for Science and Engineering Statistics, Alexandria, VA, 2019).

79. Ginther, D. K. et al. Race, ethnicity, and NIH research awards. *Science* **333**, 1015 (2011).

80. Hoppe, T. A. et al. Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Sci. Adv.* **5**, eaaw7238 (2019).

81. *Quick Facts: United States.* https://www.census.gov/quickfacts/fact/table/US/PST045218.

82. Uthman, O. A. et al. Increasing the value of health research in the WHO African Region beyond 2015—reflecting on the past, celebrating the present and building the future: a bibliometric analysis. *BMJ Open* **5**, e006340–e006340 (2015).

83. Bentley, A. R., Callier, S. & Rotimi, C. The emergence of genomic research in Africa and new frameworks for equity in biomedical research. *Ethn. Dis.* **29**, 179–186 (2019).

84. van der Merwe, C. et al. Advancing neuropsychiatric genetics training and collaboration in Africa. *Lancet Glob. Health* **6**, e246–e247 (2018).

85. Mulder, N. J. H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. *Genome Res.* **26**, 271–277 (2016).

86. Mulder, N. Development to enable precision medicine in Africa. *Personalized Med.* **14**, 467–470 (2017).

87. Ofon, E. et al. A polymorphism in the haptoglobin, haptoglobin related protein locus is associated with risk of human sleeping sickness within Cameroonian populations. *PLoS Neglect. Trop. Dis.* **11**, e0005979 (2017).

88. Kaboré, J. W. et al. Candidate gene polymorphisms study between human African trypanosomiasis clinical phenotypes in Guinea. *PLoS Negl. Trop. Dis.* **11**, e0005833–e0005833 (2017).

89. Ahouty, B. et al. Candidate genes-based investigation of susceptibility to Human African Trypanosomiasis in Côte d'Ivoire. *PLoS Negl. Trop. Dis.* **11**, e0005992–e0005992 (2017).

90. Kimuda, M. P. et al. No evidence for association between APOL1 kidney disease risk alleles and Human African Trypanosomiasis in two Ugandan populations. *PLoS Negl. Trop. Dis.* **12**, e0006300–e0006300 (2018).

91. Cooper, A. et al. APOL1 renal risk variants have contrasting resistance and susceptibility associations with African trypanosomiasis. *eLife* **6**, e25461 (2017).

92. Ilboudo, H. et al. Introducing the TrypanoGEN biobank: a valuable resource for the elimination of human African trypanosomiasis. *PLoS Negl. Trop, Dis.* **11**, e0005438–e0005438 (2017).

93. Dennis-Antwi, J. A. et al. Relation between religious perspectives and views on sickle cell disease research and associated public health interventions in Ghana. *J. Genet. Couns.* https://doi.org/10.1007/s10897-018-0296-7 (2018).

94. Landouré, G. et al. Neurological complications in subjects with sickle cell disease or trait: genetic results from Mali. *Glob. Heart* **12**, 77–80 (2017).

95. Wonkam, A. et al. Clinical and genetic factors are associated with pain and hospitalisation rates in sickle cell anaemia in Cameroon. *Br. J. Haematol.* **180**, 134–146 (2018).

96. Carter, A., Liddle, J., Hall, W. & Chenery, H. Mobile phones in research and treatment: ethical guidelines and future directions. *JMIR Mhealth Uhealth* **3**, e95–e95 (2015).

97. Adebamowo, S. N. et al. Implementation of genomics research in Africa: challenges and recommendations. *Glob. Health Action* **11**, 1419033 (2018).

98. Tsosie, K. S., Yracheta, J. M. & Dickenson, D. Overvaluing individual consent ignores risks to tribal participants. *Nat. Rev. Genet.* **20**, 497–498 (2019).

99. de Vries, J. & Munung, N. S. Ethical considerations in genomic research in South Africa %J SAMJ. *South Afr. Med. J.* **109**, 375–377 (2019).

100. Pratt, B. & de Vries, J. Community engagement in global health research that advances health equity. *Bioethics* **32**, 454–463, https://doi.org/10.1111/bioe.12465 (2018).

101. Yozwiak, N. L. et al. Roots, not parachutes: research collaborations combat outbreaks. *Cell* **166**, 5–8 (2016).

102. de Vries, J. et al. The H3Africa policy framework: negotiating fairness in genomics. *Trends Genet.* **31**, 117–119 (2015).

103. Ramsay, M. & Sankoh, O. African partnerships through the H3Africa Consortium bring a genomic dimension to longitudinal population studies on the continent. *Int. J. Epidemiol.* **45**, 305–308 (2015).

104. *High-Level Principles on Ethics, Governance and Resource Sharing.* https://h3africa.org/index.php/about/ethics-and-governance/.

105. Yakubu, A. et al. Model framework for governance of genomic research and biobanking in Africa? a content description [version 1; referees: 3 approved]. *AAS Open Res.* **1**, https://doi.org/10.12688/aasopenres.12844.1 (2018).

106. *NHLBI Trans-Omics for Precision Medicine: About TOPMed.* https://www.nhlbiwgs.org/#Study%20Characteristics.

107. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).

108. Giri, A. et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **51**, 51–62 (2019).

109. Sung, Y. J. et al. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. *Am. J. Hum. Genet.* **102**, 375–400 (2018).

110. Daya, M. et al. Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. *Nat. Commun.* **10**, 880 (2019).

111. Natarajan, P. et al. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat. Commun.* **9**, 3391 (2018).

112. Zekavat, S. M. et al. Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. *Nat. Commun.* **9**, 2606 (2018).

113. Prokopenko, D. et al. Whole-genome sequencing in severe chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* **59**, 614–622 (2018).

114. Wheeler, M. M. et al. Genomic characterization of the RH locus detects complex and novel structural variation in multi-ethnic cohorts. *Genet. Med.* **21**, 477–486 (2018).

115. Gong, J. et al. Trans-ethnic analysis of metabochip data identifies two new loci associated with BMI. *Int. J. Obes.* **42**, 384 (2017).

116. Fernández-Rhodes, L. 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.* **136**, 771–800 (2017).

117. Yoneyama, S. et al. Generalization and fine mapping of European ancestry-based central adiposity variants in African ancestry populations. *Int. J. Obes.* **41**, 324 (2016).

118. Zubair, N. et al. Fine-mapping of lipid regions in global populations discovers ethnic-specific signals and refines previously identified lipid loci. *Hum. Mol. Genet.* **25**, 5500–5512 (2016).

119. Bien, S. A. et al. Transethnic insight into the genetics of glycaemic traits: fine-mapping results from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. *Diabetologia* **60**, 2384–2398 (2017).

120. Avery, C. L. 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* **14**, 572–580 (2017).

121. Fernández-Rhodes, L. et al. The genetic underpinnings of variation in ages at menarche and natural menopause among women from the multi-ethnic Population Architecture using Genomics and Epidemiology (PAGE) Study: a trans-ethnic meta-analysis. *PLoS ONE* **13**, e0200486 (2018).

122. Kocarnik, J. M. et al. Discovery, fine-mapping, and conditional analyses of genetic variants associated with C-reactive protein in multiethnic populations using the Metabochip in the Population Architecture using Genomics and Epidemiology (PAGE) study. *Hum. Mol. Genet.* **27**, 2940–2953 (2018).

123. Franceschini, N. et al. Variant discovery and fine mapping of genetic loci associated with blood pressure traits in hispanics and African Americans. *PLoS ONE* **11**, e0164132–e0164132 (2016).

124. Klarin, D. et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat. Genet.* **50**, 1514–1523 (2018).

125. Rao, D. C. et al. Multiancestry study of gene–lifestyle interactions for cardiovascular traits in 610 475 individuals from 124 cohorts: design and rationale. *Circ. Cardiovasc. Genet.* **10**, e001649 (2017).

126. Feitosa, M. F. et al. Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570 K individuals across multiple ancestries. *PLOS ONE* **13**, e0198166 (2018).

127. Sung, Y. J. et al. A multi-ancestry genome-wide study incorporating gene-smoking interactions identifies multiple new loci for pulse pressure and mean arterial pressure. *Hum. Mol. Genet.* Preprint at https://doi.org/10.1093/hmg/ddz070 (2019). [Epub ahead of print].

128. Bentley, A. R. et al. Multi-ancestry genome-wide gene–smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nat. Genet.* **51**, 636–648 (2019).

129. de Vries, P. S. et al. Multi-ancestry genome-wide association study of lipid levels incorporating gene–alcohol interactions. *Am. J. Epidemiol.* **188**, 1033–1054 (2019).

130. Noordam, R. et al. Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration. *Nat. Commun.* **10**, 5121 (2019).

131. Johnston, H. R. et al. Identifying tagging SNPs for African specific genetic variation from the African Diaspora Genome. *Sci. Rep.* **7**, 46398–46398 (2017).

132. Kessler, M. D. et al. Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry. *Nat. Commun.* **7**, 12521 (2016).

133. Daya, M. et al. Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. *Nat. Commun.* **10**, 880 (2019).

134. Ndila, C. M. et al. Human candidate gene polymorphisms and risk of severe malaria in children in Kilifi, Kenya: a case-control association study. *Lancet Haematol.* **5**, e333–e345 (2018).

135. Opi, D. H. et al. Two complement receptor one alleles have opposing associations with cerebral malaria and interact with α+ thalassaemia. *eLife* **7**, e31579 (2018).

136. Leffler, E. M. et al. Resistance to malaria through structural variation of red blood cell invasion receptors. *Science* **356**, eaam6393 (2017).

137. Clarke, G. M. et al. Characterisation of the opposing effects of G6PD deficiency on cerebral malaria and severe malarial anaemia. *eLife* **6**, e15085 (2017).

138. Busby, G. B. J. et al. Admixture into and within sub-Saharan Africa. *eLife* **5**, e15266 (2016).

139. Baker, J. L., Rotimi, C. N. & Shriner, D. Human ancestry correlates with language and reveals that race is not an objective genomic classifier. *Sci. Rep.* **7**, 1572 (2017).

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Correspondence and requests for materials should be addressed to C.N.R.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.