ANNUAL REVIEWS

Annual Review of Genomics and Human Genetics

Five Priorities of African Genomics Research: The Next Frontier

Ambroise Wonkam,[1,2] Nchangwi S. Munung,[1] Collet Dandara,[1] Kevin K. Esoh,[1] Neil A. Hanchard,[3] and Guida Landoure[4]

[1]Division of Human Genetics, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa; email: nchangwi.munung@uct.ac.za, collet.dandara@uct.ac.za, eshkev001@myuct.ac.za

[2]Current affiliation: McKusick-Nathans Institute of Genetic Medicine and Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; email: awonkam1@jhmi.edu

[3]Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; email: neil.hanchard@nih.gov

[4]Faculty of Medicine and Odontostomatology, University of Sciences, Techniques, and Technology of Bamako, Bamako, Mali; email: glandoure@gmail.com

Annu. Rev. Genom. Hum. Genet. 2022. 23:499–521

First published as a Review in Advance on May 16, 2022

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

https://doi.org/10.1146/annurev-genom-111521-102452

Keywords

human genomic variations, Africa, evolutionary history, natural selection, equity

Abstract

To embrace the prospects of accurately diagnosing thousands of monogenic conditions, predicting disease risks for complex traits or diseases, tailoring treatment to individuals' pharmacogenetic profiles, and potentially curing some diseases, research into African genomic variation is a scientific imperative. African genomes harbor millions of uncaptured variants accumulated over 300,000 years of modern humans' evolutionary history, with successive waves of admixture, migration, and natural selection combining with extensive ecological diversity to create a broad and exceptional genomic complexity. Harnessing African genomic complexity, therefore, will require sustained commitment and equitable collaboration from the scientific community and funding agencies. African governments must support academic public research and industrial partnerships that build the necessary genetic medicine workforce, utilize the emerging genomic big data to develop expertise in computer science and bioinformatics, and evolve national and global

governance frameworks that recognize the ethical implications of data-driven genomic research and empower its application in African social, cultural, economic, and religious contexts.

BACKGROUND

As we begin the third decade since the completion of the Human Genome Project, there is increasing evidence that to embrace the lofty prospects of accurately diagnosing thousands of monogenic conditions, predicting disease risks for most common complex traits and diseases, tailoring treatment to individuals' pharmacogenetic profiles, and potentially curing some diseases, research into ancestral African genomic variation is a scientific imperative; however, equitable application of these ideals remains a major challenge (73, 131). Despite advances in genetic technologies and exponential drops in costs, inequalities in healthcare systems, deficits in genetic research workforces, and a lack of access to research funding have prevented knowledge produced by genomic research from truly informing the global public good, particularly in Africa (104). The World Health Organization publication *Genomics and World Health* recognized this concern two decades ago (143). Moreover, most genomic research on the African continent over the past two decades has been driven largely by agendas defined by outside investigators (5, 49, 124), and priority areas may not necessarily reflect the current needs of the continent (83). As an illustration, most published research using DNA samples collected from Cameroonians in Cameroon is not associated with local institutions or authors (134).

This preponderance of helicopter research, a term coined to label such practices, perpetuates a form of scientific imperialism (110). In addition, in Africa, the available medical workforce has limited knowledge in genetics (138), exemplified by misunderstanding and uncertainty during the informed-consent processes used when enrolling participants in genomic research; as a result, research participants often struggle to understand the full implications of their consent (9, 72). Recent reports on the perception of fragile X syndrome—a rare X-linked monogenic condition associated with mental retardation—in rural Cameroon indicated that the predominant belief was that it resulted from a curse laid upon children by the chief of the village in retaliation for his wives not mourning his intellectually disabled servant (55). The complex clinical presentations of rare genetic diseases, coupled with the inherent difficulty in assigning a definitive diagnosis and a limited genetics knowledge base, mean that these diseases are often attributed to the supernatural, leading to the stigmatization of patients and their families, as seen with oculocutaneous albinism in some African countries (31).

Notably, frameworks for the ethical, legal, and social implications of the governance, access, sharing, intellectual property regulation, and commercialization of products associated with African genomics research are often absent, which recently led to international controversy (78). Despite these challenges, established genetic medicine services, with academic training programs in medical genetics and genetic counseling, do exist in Africa, but on a limited scale, most notably in South Africa (90, 94), where the South African government has also launched the nationally funded Southern African Human Genome Programme, a rarity in Africa (95).

The expansion of endogenous efforts on the continent has been greatly enhanced by the activities of the African Society of Human Genetics, established in 2003 (101). This organization's fundamental goals are to improve education, enhance networking, and build capacity to undertake genetic research in Africa, and its work is facilitated by annual meetings in different African countries (33, 82, 88, 115, 128). Critically, it has facilitated African-led genomic research. Close collaboration with international funders (the National Institutes of Health and Wellcome Trust) led to the establishment of the Human Heredity and Health in Africa (H3Africa) consortium

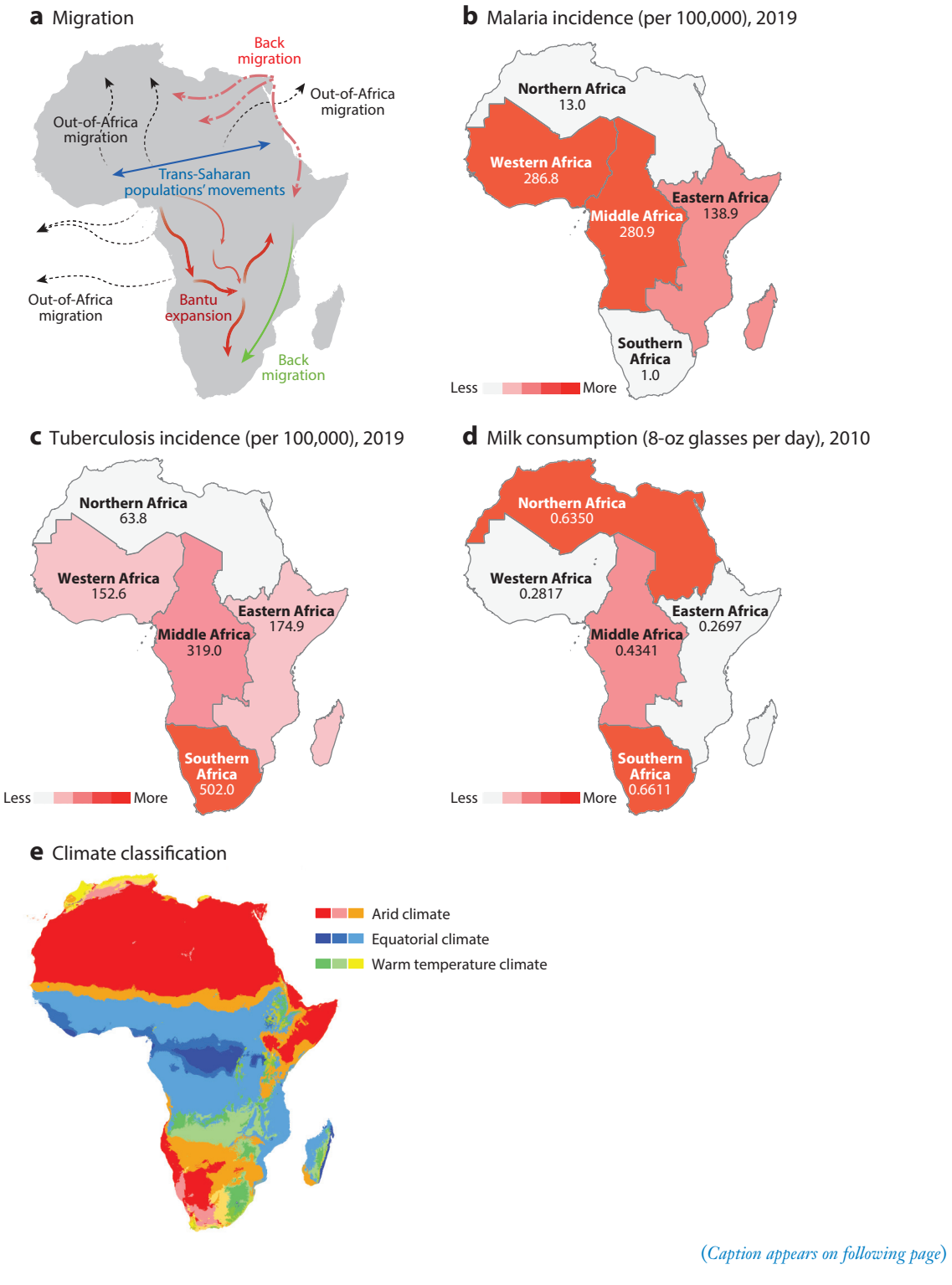
(102), which, alongside more recent genomic research initiatives in Africa, is generating substantial amounts of African human genomic data. The supported activities actively engage young and established African scientists in genomic research, as illustrated by emerging high-impact publications from the grantees (51). The increased availability of African genomes from these initiatives (19) will improve our understanding of genomic variation associated with complex traits and augment the clinical diagnosis of monogenic conditions in all populations. We propose that studying African genomic variation represents the next frontier of genetic medicine for three major reasons: ancestry, ecology, and equity (131). In this article, we discuss five priorities to ensure that the scientific imperative and the global rush for African genomic variations will concurrently enhance genomic research and medicine in Africa and around the world.

EVOLUTIONARY GENETICS: UNDERSTANDING ARCHAIC DNA, HUMAN MIGRATION AND ADMIXTURE, AND NATURAL SELECTION IN HEALTH AND DISEASE

There is evidence for missing or uncaptured information from African genomes in the publicly available databases. For example, many DNA variants in a pan-genome generated from 910 individuals of African descent were not found in the human reference genome (111). Whole-genome sequencing analyses of 426 individuals—comprising 50 ethnolinguistic groups, including previously unsampled populations—to explore the breadth of genomic diversity across Africa uncovered more than 3 million previously undescribed variants, most of which were found in individuals from newly sampled ethnolinguistic groups (19). Moreover, it is estimated that 2–19% of the genomes of ancestral Africans is derived from poorly investigated archaic populations that diverged before the split of Neanderthals and modern humans (32). The generation of high-coverage reference genomes for archaic hominid species such as Neanderthals and Denisovans (located in Eurasia) has made it feasible to identify, in present-day Europeans, archaic-introgressed segments that make up approximately 2% of the Neanderthal genome and are ostensibly enriched for variation in genes involved in dermatological phenotypes, neuropsychiatric disorders, and immunological functions, including host susceptibility to coronavirus disease 2019 (COVID-19) (117, 146). The largest challenge in expanding this area of research on the continent has been the inability to obtain high-quality ancient DNA from regions with a tropical climate, such as those across equatorial Africa. Once these technical challenges are overcome, however, it is anticipated that there will be novel associations between variants in ancient African DNA and human traits and diseases that will provide insights of benefit to all modern-day humans.

As Africa is the home of modern humans, studies of ancestral African populations will continue to shape our knowledge of human evolutionary history. Genetic studies in such populations have reconstructed a multitude of quasi-known migrations and movements of people across the continent, particularly the so-called Bantu expansion (13, 19, 44). Over many centuries, these movements (**Figure 1*a***) have generated a rich tapestry of admixture that represents both recent and more ancient introgressions. As people moved, however, they were faced with new cultures, diets, environments, and often new pathogens, all of which are potential drivers of genetic adaptation (**Figure 1*b–d***). Thus, the range of climates, cultures, and biodiversity across the continent, both in the present and in the past—for example, the now barren desert of the Sahara was once a lush equatorial rain forest (20)—has, in turn, shaped differential genomic variant frequencies among populations.

In turn, differential population genomic variant frequencies are shaped by natural selection as an adaptation to environmental pressures. There are several well-established examples of variants driven to high frequency in specific African populations as a result of such natural selection;



(*Caption appears on following page*)

Figure 1 (*Figure appears on preceding page*)

Intersecting factors contributing to complexity in African genomes. Migration and admixture events over many years [e.g., the Bantu expansion (*solid red arrows*), back migration (*dot-dashed red* and *solid green arrows*), out-of-Africa migrations (*dotted black arrows*), and trans-Saharan populations' movements (*solid blue arrow*) in panel *a*] combined with variability in infectious disease risk (panels *b* and *c*), diet (panel *d*), and ecology (panel *e*) have had far-reaching effects on African genomes. Data for panels *b* and *c* are from Reference 142; data for panel *d* are from Reference 114; panel *e* adapted from Reference 11 (CC BY 4.0).

quintessential examples include (*a*) variants in and around the beta-globin gene (*HBB*) that cause sickle cell disease (SCD) and related hemoglobinopathies but confer resistance to malaria (112), (*b*) variants in *APOL1* that are protective against trypanosomes (the parasites that cause sleeping sickness) (40) but also increase susceptibility to chronic kidney disease in populations of African ancestry (39, 40), (*c*) variants of *OSBPL10* and *RXRA* that protect against dengue fever (113), and (*d*) variants in genes that play a role in lactose tolerance (74).

Still more variants of relevance to persons of African ancestry are being added to the above examples (64, 127), and a recent survey of selection signatures in seven African populations identified 60 novel coding loci and 34 novel noncoding loci, with evidence of selection in at least one group (19). The function and selection driver for most of these loci remain unclear, but they hint at genes involved in viral pathogen infection and basic cellular metabolism (19, 35) and suggest that the potential adaptive pleiotropy seen at *HBB* and *APOL1* might also be at work in diseases such as chronic kidney disease (39, 40) and uterine fibroids, which are seen predominantly among persons of African ancestry (42, 50, 54).

Evidently, a real promise of improved healthcare for all populations lies in ancestral African population sequencing at the scale of population cohorts of tens of millions of individuals, covering various ethnolinguistic and geographical locations, for unbiased and comprehensive representation of the continent's diversity (131). Such efforts promise an expanded catalog of selected loci, which, alongside detailed studies of immunogenetics and host–pathogen interactions in African populations, have the potential to further our understanding of common human diseases and host responses to emerging infectious diseases such as COVID-19.

COMPLEX DISEASE AND POLYGENIC RISK SCORES: APPROPRIATE DNA ARRAY TECHNOLOGIES AND BIOINFORMATIC ANALYTIC APPROACHES

Current iterations of polygenic risk scores (PRSs) aim to predict an individual's risk for a specific disease based on the genetic variants harbored by that individual. The foundation of PRSs is information gleaned from genome-wide association studies (GWASs), including summary statistics thereof. The success of this approach is predicated on the ability to proxy the genome by genotyping informative single-nucleotide polymorphisms (SNPs) that are considered redundant to remaining SNPs, and/or having a robust reference of population haplotypes from which to infer or impute untyped variants. This approach reflects the nonrandom associations between markers along a chromosome known as linkage disequilibrium. As has been well documented, the vast majority of GWASs have been undertaken in populations of European (or at least non-African) ancestries—at last count, nearly 80% of individuals in GWASs are of European ancestry, and only approximately 2% are of African ancestry.

As a result of the long period of time over which modern humans have been in Africa, African populations are the most genetically diverse in the world. On average, a genome of African ancestry carries three to five times more variants than those from other ancestries (10). This observation is partially reflective of multiple partial migratory waves out of Africa approximately 70,000 years ago. The deep evolutionary tree within Africa, with more time for recombination to

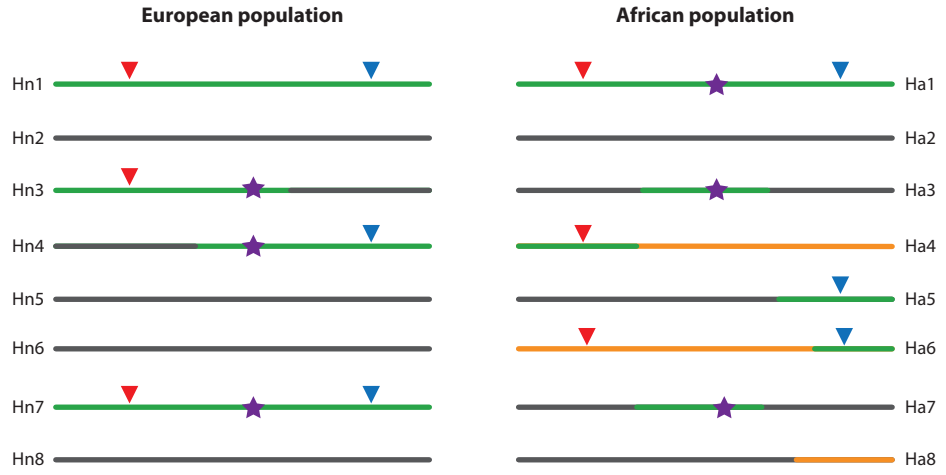


Figure 2

Variants, haplotypes, and linkage disequilibrium in African and non-African populations. The figure illustrates African (Ha1–8) and non-African European (Hn1–8) haplotypes in each population. The strongest associated genome-wide association study variant (putative causal variant) is shown as a purple star and occurs predominantly on a haplotype background in green (Hn1) alongside two genotyped variants (*red* and *blue triangles*); this haplotype is distinct from non-disease-variant-carrying haplotypes in gray (e.g., Hn2). In African populations, even though the frequencies of the two genotyped alleles are the same, recombination over generations (e.g., Ha5 and Ha7) and the introduction of additional ancestral haplotypes through admixture (*orange haplotypes*—Ha4, Ha6, and Ha8) have eroded the nonrandom association (linkage disequilibrium) between the alleles, and the causal allele is at a lower frequency (potentially as a result of selection or genetic drift). Thus, detecting association and developing genetic risk scores is easier in non-African European populations.

separate linked alleles, combined with complex patterns of admixture across the continent creates highly heterogenous haplotypes with limited linkage disequilibrium (**Figure 2**). Thus, existing PRSs, built on a Eurocentric backbone, exhibit a predictable bias when it comes to usability and transferability across populations, particularly African populations—most current-day PRSs do not account for untyped allelic associations that have low linkage disequilibrium with existing GWAS variants or associations that are limited to African populations, particularly if they have a high frequency or effect size among Africans. As a consequence, PRSs estimated using European data do not accurately predict phenotypes and disease risk in non-European populations, performing worst in individuals with African ancestry (71). These biases limit the broad applicability of PRSs in ethnically diverse populations and could thus exacerbate health inequities across populations (116).

At the same time, the extensive genetic and phenotypic diversity observed in African populations makes these populations particularly informative for GWASs—for example, reduced linkage disequilibrium between markers improves fine mapping and identification of putatively causative variants. Accordingly, while only 2.4% of participants in large GWASs are individuals of African ancestry, they account for 7% of all associations (47). A GWAS on the genetic susceptibility to type 2 diabetes identified a previously unreported African-specific significant locus while still demonstrating the transferability of 32 established type 2 diabetes loci (6). In addition, nonsense mutations in *PCSK9* that are found commonly among Africans but rarely in Europeans (21) are associated with a 40% reduction in plasma levels of low-density lipoprotein. This observation was central to support for *PCSK9* as an effective target for dyslipidemia and led to the development

of new statin-adjunctive medications for this global problem. In the largest GWAS meta-analysis for 34 complex traits, conducted in 14,345 Africans, several loci had limited transferability among cohorts (48), further illustrating the genomic variation among Africans. At the other end of the variant frequency spectrum, whole-exome sequencing of nearly 1,000 African study participants of Xhosa ancestry with schizophrenia found very rare damaging mutations in multiple genes (46), a finding that was replicated in a Swedish cohort of approximately 5,000 individuals (41). In comparison, results for the Xhosa cohort yielded larger effect sizes, which shows that for the same number of cases and controls, the greater genetic variation in African populations provides more power to detect genotype–phenotype relationships.

These results suggest that hundreds of thousands if not millions of African genomes—chosen to be fully representative of the diverse regional and ethnolinguistic backgrounds across the continent—should be sequenced, not only to fully harness the potential of African genomes to inform complex disease generally, but also to limit already gaping health inequities. In concert, there is a similar imperative to develop faster and better methods of accounting for admixture by teasing apart local ancestry (i.e., the likely ancestral parental haplotypes at a locus) to account for variants carried on haplotypes specific to particular regions or groups. Advances in these areas, alongside more comprehensive genome sequencing that facilitates interrogation of the true genome sequence and elucidation of the full spectrum of genetic variation (e.g., including copy number and repeat variants) (7, 76), will go a long way toward the goal of inclusiveness in human genetics articulated by major funding bodies (8, 86). These tools, alongside an increase in the number of (and number of samples within) medical biobanks in Africa with access to electronic health records, similar to the UK Biobank, would greatly facilitate mapping of hundreds of genetic associations with complex traits and diseases. This, in turn, would provide fodder for improved knowledge of the pathogenesis (and ultimately treatment) of such diseases in the long term while allowing for the broader application of PRSs to diverse populations in the short term.

COMMON AND RARE MONOGENIC DISEASES: GENETIC AND LOCUS HETEROGENEITY

Common Monogenic Disease: Sickle Cell Disease

The greater availability of African genomes will also support research into common monogenic diseases, such as hemoglobinopathies. On this basis, one study estimated that the sickle cell mutation (*HBB-*β^S) emerged approximately 7,300 years ago (112), while another estimated that it originated approximately 22,000 years ago (63). However, a recent review (36) alluded to limitations in these studies; while the first assumed a complete recessive lethality of *HBB-*β^S (relative fitness = 0), contrary to the observed estimate of 0.8 (120), for its age estimate, the second utilized the average frequency (8.3%) of *HBB-*β^S across sub-Saharan Africa as the equilibrium frequency (11%). Thus, one would expect *HBB-*β^S to be older than 22,000 years, and this could be confirmed by more research on larger genomic data sets from the countries with the highest SCD prevalence. Nonetheless, the consensus of a single African origin of *HBB-*β^S not only suggested recent migration and admixture events between Africans and Mediterranean and/or Middle Eastern populations but also enhanced our understanding of genetic variation in general as well as its potential impact on hemoglobinopathies (36). Studies have shown that approximately 25% of the interindividual differences of severe malaria clinical expression can be explained by human genetics, with *HBB-*β^S alone contributing the largest proportion due to a single gene (up to 2%) in its heterozygous form, known as sickle cell trait (66), which affords up to 90% protection against severe *Plasmodium falciparum* malaria (69, 129). Consequently, the *HBB-*β^S allele has risen to appreciable frequencies in regions with historical malaria endemicity (**Figure 3*a***)—up to 20% in

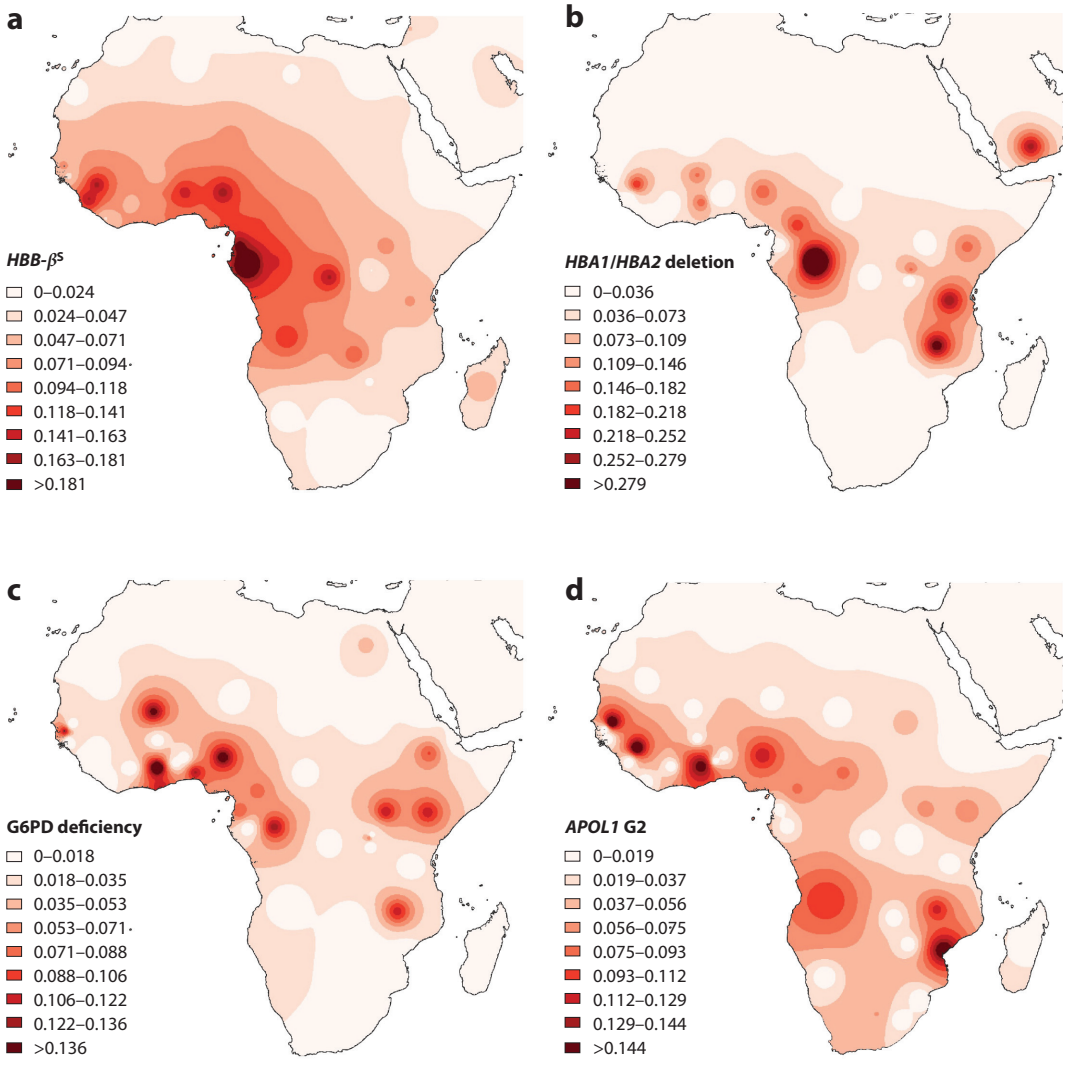


Figure 3

Potential coevolution of the *HBB-β*^{S} variant and other malaria- and trypanosome-associated gene variants in Africa, showing the distributions of (*a*) the *HBB-β*^{S} variant, (*b*) the 3.7-kb alpha-globin gene (*HBA1/HBA2*) deletion, (*c*) G6PD deficiency variants, and (*d*) the *APOL1* G2 variant. Figure adapted from Reference 36 (CC BY 4.0).

some regions, although statistical estimates have limited the highest frequency to 18% (96). The distribution of *HBB-β*^{S} outside of Africa has also improved our understanding of the influence of migration on the spread of pathogenic mutations, considering that it is most frequent in individuals of African descent—for instance, in the Mediterranean region, including in Padova and Monza in Italy (22); in the Americas, where it is absent in the indigenous population; and in regions where malaria is endemic or was historically endemic, such as in the Indian subcontinent, the Middle East, and the Mediterranean (97).

It is estimated that 75% of the 305,800 babies with SCD born each year are born in Africa. Simple clinical applications of DNA technology could improve healthcare in many African countries.

For example, as primary prevention, prenatal genetic diagnosis of SCD provides reproductive options. This has been an efficient point of entry of genetic medicine in targeted African settings (139). Moreover, knowledge of an individual's genetic variants can have an impact on secondary prevention of and treatment strategies for SCD. Indeed, the considerable observed interindividual heterogeneity in SCD clinical outcome has been linked to biological factors that are subject to genetic variations. Fetal hemoglobin (HbF) level—the most influential modifier of SCD clinical outcome, and a highly heritable quantitative trait (heritability of 89%) (75)—is associated with three major quantitative trait loci: *HBG2* (11p15.4), *BCL11A* (2p16.1), and *HBS1L-MYB* (6q24). Significant progress has been made toward uncovering the functionally relevant variants in these loci, such as rs7482144 in *HBG2*, rs1427407 in *BCL11A*, and rs66650371 in *HBS1L-MYB*. Two other loci have been found to alter HbF level: *FRMPD4* (Xp22.2) among Tanzanian SCD patients (125) and *BCL2L1* (20q11.21), which is yet to be seen in Africa (27). Nevertheless, many of the studies on polymorphisms in HbF quantitative trait loci have involved non-African populations or African Americans (38), Black British people of West African origin, and Afro-Caribbeans or Afro-Brazilians (24). Only a few studies have involved SCD patients from Africa [notably from Tanzania (67) and Cameroon (99, 137)], and these studies have jointly explained a maximum 20% of the total HbF trait heritability, compared with 50% in studies involving African Americans (38). This underscores fundamental disparities in the genetic architecture of populations of African ancestry that have had different niche constructions over approximately the past 400 years and implies that more than 60% of missing or hidden HbF trait heritability remains to be uncovered in sub-Saharan Africa (130). A whole-exome sequencing approach recently revealed genetic polymorphisms in genes involved in notable SCD pathophysiological pathways, including *CLCN6*, which is involved in chloride balance and implicated in regulation of blood pressure; *OGDHL*, which is important in arginine metabolism, a therapeutic target in SCD; and *SERPINC1*, which encodes antithrombin, a molecule involved in the clotting mechanism (132). These discoveries highlight the importance of a greater availability of African genomes to the understanding of the genetic modification of SCD clinical outcomes.

There is also considerable overlap in the geographical distributions of *HBB-*β^S and other gene variants that have come under enormous malarial pressure and have therefore been adaptively selected. One example is the 3.7-kb alpha-globin gene (*HBA1/HBA2*) deletion, which is especially common among patients with SCD in sub-Saharan Africa compared with ethnolinguistically matched controls (**Figure 3*b***). Coinheritance of this variant with *HBB-*β^S is associated with improved SCD hematological indices and low hospitalization rates (106). Another example involves the G6PD deficiency variants, which are associated with protection against severe malaria yet predispose SCD patients to hemolytic anemia (**Figure 3*c***). A similar overlap is observed between *HBB-*β^S and the *APLO1* G1 and G2 variants, which are prevalent in sub-Saharan Africa due to the protection they afford against African trypanosomiasis (sleeping sickness) but also predispose individuals of African ancestry to end-stage renal disease, which exacerbates SCD severity (23) (**Figure 3*d***). Therefore, combining individual genetic risk variants for predicting targeted endophenotypes and complications such as stroke or kidney disease can allow the use of genetics for secondary prevention and inform the pathobiology of this common condition in the global population.

Future genetic approaches to SCD therapy may also involve RNA. Small noncoding RNAs called microRNAs gum up the production of proteins by binding to the transcription machinery in a cell. They could be used to target the entire pathway of HbF production—particularly to suppress the expression of HbF-inhibitor proteins such as *BCL11A*—which would have a much stronger effect than targeting a single gene. Research that identifies more candidate microRNAs that act on HbF production will provide an attractive route for future SCD therapies that mimic

hydroxyurea-induced HbF production (77, 98). Mediation of HbA or HbF production through the injection of mRNA, a process used in COVID-19 vaccines, could provide another RNA-based technique for SCD therapy (131). Therefore, SCD is also a very good exemplar of complex disease with extensive environmental influences—social, dietary, climatic, and cultural—all of which vary across geographies where SCD is common. Thus, research on SCD in Africa (and in populations of African ancestry) is highly responsive to increasing calls for a better understanding of gene–environment interactions in human disease, and the disease serves as a model for understanding the impact of genetic variation on common monogenic traits to illustrate the multiple layers of genomic research and genomic medicine implementations (105). Future initiatives in research and healthcare will be enabled by current initiatives such as the Sickle in Africa consortium, which has a growing pool of clinical data from more than 10,000 patients from Tanzania, Ghana, and Nigeria (68), with the ultimate goal of establishing newborn and long-term prospective SCD cohorts. Establishing a well-coordinated, multicenter prospective longitudinal SCD cohort in Africa is crucial to improving our genomics and pathophysiology understanding and thus for investigating new routes for novel therapeutic interventions, with the aim of reducing the substantial disease burden (135).

Rare Monogenic Diseases

With only approximately 21% ($n = 4,347$) of human genes found to have disease-associated mutations to date (52) and thousands of phenotypes showing Mendelian inheritance, many more disease genes remain to be identified. Recent advances in genetic technology, especially next-generation sequencing, have allowed quicker and less expensive discovery of several hundred disease-associated variants and genes. Yet there is a nearly complete absence of data on the contribution of genetics to rare monogenic diseases in sub-Saharan Africa—a well-documented severe inequity, as mentioned above, and one that causes several biases in the interpretation of genetic findings. Indeed, almost all of the well-characterized single-gene diseases and syndromes have been reported in developed countries, and only infrequent cases have been described in sub-Saharan Africa (62). Among the several matters that have led to the neglect of rare monogenic disease in sub-Saharan Africa are the high burden of infectious diseases, which attracts the most attention from policy makers, and the lack of local expertise in genetics and funding. Nonetheless, this situation offers tremendous opportunity. Importantly, even beyond the core factor of inequity, there are additional compelling reasons to study rare monogenic diseases in sub-Saharan Africa. As stated above, Africa harbors the greatest genetic diversity of any continent, which is influenced by multiple factors, including thousands of years of human history prior to dispersion and likely loss of variations due to genetic drift, a massive land area with a variety of terrain, and an array of climates and ecologies (15, 43, 73, 103). This vast genetic diversity may be associated with phenotypic variability that is greater than that in other populations. Combining this remarkable diversity with additional factors, such as high fertility rates (more than six children per woman in some countries in sub-Saharan Africa), high levels of consanguinity [e.g., 27% according to a Malian study (108) and up to 67% according to a study based in the western Kordofan state of Sudan (26)], and local genetic bottlenecks, suggests that sub-Saharan Africa is a fertile ground for the discovery of novel single-gene disease-causing variants. Thus, the elucidation of the genetic basis of rare monogenic diseases in sub-Saharan Africa would be directly relevant to clinical practice for patients in this region, and could lead to more broadly applicable insights into the underlying pathophysiology of these diseases and, potentially, to novel therapies.

Indeed, allelic and locus heterogeneity display important differences in African individuals compared with other populations—for example, mutations in *GJB2* account for nearly 50% of congenital nonsyndromic hearing impairment among Europeans and Asians but are nearly

nonexistent in most African populations (4). Moreover, gene panel testing in a wide range of well-characterized Malian families with hereditary neurological disorders has come to a diagnosis in only 40% of the cases, suggesting the genetic heterogeneity of that population. There is evidence that novel variants in known genes, as well as novel genes, are more likely to be found in monogenic conditions in African populations—for example, for hearing impairment (3, 140, 141) or neurogenetic diseases (45, 59, 60, 62, 145). To accomplish precision medicine for everyone in the developed world, research on African populations is also necessary to address genetic and genomic health needs for the ever-increasing population of the African diaspora. In fact, a Malian family with hereditary spastic paraplegia had a mutation in a new gene, *C19ORF12* (62). The variant in *C19ORF12* was found in 3 of 3,836 African American alleles and 0 of 8,222 European American alleles (62). Interestingly, the same mutation in *C19ORF12* was found in a Brazilian family with some clinical overlap, and haplotype reconstruction suggested that the Malian and Brazilian families may be related (61), which indicated the importance of this population for the African diaspora. Ultimately, increased availability of ancestral African exomes, perhaps through the development of an African exome aggregation database, will enhance disease–gene pair curation and will address existing challenges surrounding database biases and inference of variant deleteriousness, which have led to the misclassification of variants—for example, for inherited cardiomyopathies (70).

PHARMACOGENETICS

In Africa, where infectious diseases are highly prevalent, research on pathogen genomes has enhanced our understanding of disease transmission, virulence mechanisms, and avoidance of host defenses (37). Furthering such research in molecular infectiology and its related genetic information is expected to enable the development of new diagnostic tests, vaccines, and therapeutic agents; it is also likely to lead to new approaches for vector control and reveal why individuals and populations vary in their susceptibility to infectious diseases (143). In parallel to this, major pharmacogenetic research could contribute to reducing the cost and ultimately the burden of diseases such as AIDS, tuberculosis, and malaria on African healthcare systems.

Healthcare systems worldwide spend considerable sums of money each year in the provision of therapeutic drugs. For example, it is estimated that more than 1 billion people worldwide suffer from hypertension, and in the United States, on average, a hypertensive patient spends an additional $2,000 annually (57), costing the US economy more than $130 billion. In developing countries, the coexistence of a huge burden of infectious diseases and a rising incidence of noncommunicable diseases strains healthcare delivery systems. Therapeutic drug treatments are not 100% effective because of several factors, including genomic modulation of drug response. The current revolution in human genome characterization has shown that an individual's genomic profile influences 20–95% of their drug responses. Pharmacogenomics, the genetics of drug response, has the potential to improve a patient's clinical outcomes through stratification and optimization of drug type and dosage. Cost has previously been a deterrent to the implementation of pharmacogenomics, but the availability of high-throughput genomics technologies and the accompanying reduction in the cost of genomic characterization are providing opportunities for patients to obtain precision genomic information that can help inform clinical decision-making (12). Although the benefit of using pharmacogenomic data to optimize drug therapy is becoming increasingly accepted by healthcare ecosystems, data on African populations remain limited in genome databases even though African populations present the greatest genome diversity. Although a handful of studies have explored pharmacogenomics in African populations (29), the genomic determinants of drug response in Africans differ, both quantitatively and qualitatively, within Africa or when compared to other world populations (87, 118, 121). Indeed, *CYP2C9*2*, *CYP2C9*3*, and the *VKORC1* c.1639G>A SNP—the most important known genetic determinants

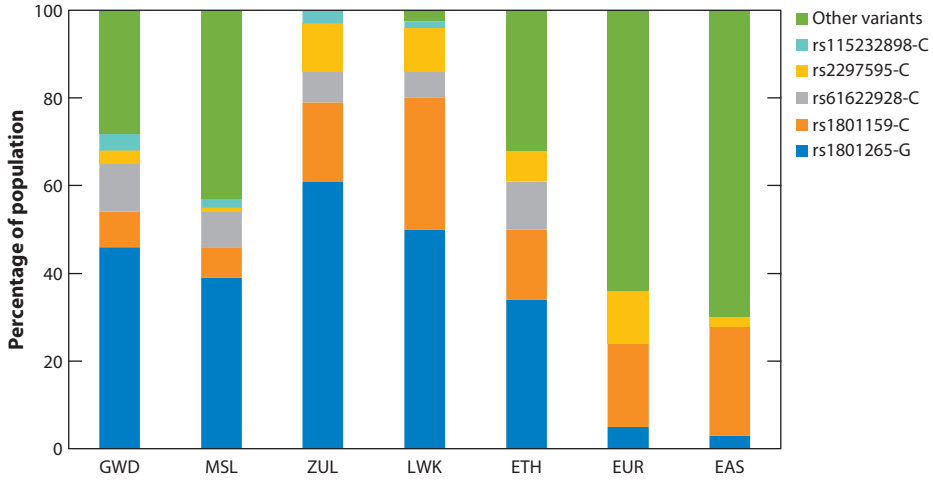


Figure 4

Quantitative and qualitative differences in the distributions of selected genetic variants in the dihydropyrimidine dehydrogenase (*DPYD*) gene across different populations. Abbreviations: GWD, Gambian; MSL, Mende; ZUL, Zulu; LWK, Luhya; ETH, Ethiopian; EUR, European; EAS, East Asian.

of warfarin response, which are part of the US Food and Drug Administration–approved warfarin drug label that provides a dosing table—are rare or absent among Africans (87). *SLCO1B1*5*, a determinant of statin responses, which is common among Europeans, also varies among African populations from 0% to 11% (118).

Figure 4 presents a comparison of selected variants in the dihydropyrimidine dehydrogenase (*DPYD*) gene, which encodes an enzyme that principally catabolizes fluoropyrimidines, such as fluorouracil, doxifluridine, and capecitabine. Fluoropyrimidines are used in the treatment of cancer, and most cancer treatments are known for their notoriously high toxicity and unpredictable variations among patients. A comparison of five common variants in the *DPYD* gene clearly shows that African populations (Gambian, Mende, Zulu, Luhya, and Ethiopian) have a different profile of variants than European and East Asian populations. Furthermore, among the African populations, the mosaic of variants differs when comparing populations from eastern Africa (Luhya and Ethiopian) with those from western Africa (Gambian and Mende) and southern Africa (Zulu) (25). The genetic variants presented for *DPYD* are mostly missense variants that result in loss of enzyme function; thus, increased plasma drug levels lead to toxicity. African populations seem to present with additional variants, such as rs61622928-T and rs115232898-C, that are absent or rare in both Europeans and East Asians. The profile of variants shown in **Figure 4** manifests in the phenotype as a range of responses by patients to the same drug. Thus, the pharmacogenomics of infectious diseases (e.g., AIDS, tuberculosis, malaria, and COVID-19) is important for African populations, while the pharmacogenomics of cardiovascular diseases is becoming important, as is the case for cancer and hypertension. Several studies have shown resistance due to genetic variations in two of the most expanded infectious diseases in Africa, AIDS (91) and tuberculosis (34, 58, 107).

Many other important pharmacogenes, including *CYP2B6*, also exhibit polymorphisms that are important in the response to therapeutic drug treatment. *CYP2B6* is involved in the metabolism of essential drugs such as artemisinin, used in the treatment of malaria; cyclophosphamide, used in cancer therapy; and efavirenz, the most commonly used antiretroviral drug. A single-nucleotide substitution, *CYP2B6* c.516G>T, has been widely reported to affect responses to both efavirenz

and nevirapine (17). The c.516T variant, which is prevalent among African populations (121), apparently abolishes CYP2B6 enzyme activity. Although some of these studies were conducted in Africa, large-scale studies are needed in African subpopulations to optimize treatment of these diseases in Africa. Variants in *VKORC1* and *CYP2C9* have been associated with differential warfarin sensitivity in populations of European ancestry compared with populations of Asian and African ancestry, which has implications for therapeutic dosage (65). There are likely several Africa-specific pharmacogenetic variants that could have implications in the treatment and well-being of African populations and could only be uncovered through large-scale genomic studies. Lastly, several SNPs have been associated with antiepileptic drug resistance in populations of European descent; considering that approximately 80% of epilepsy cases are in developing countries, including African countries, identifying variants in African populations would help optimize treatment of this debilitating disease (16).

The limited involvement of African populations in pharmacogenomic studies has a huge effect on populations both in Africa and elsewhere. Currently, most (if not all) drugs are developed and tested in clinical trials using populations with European and Asian genomic backgrounds. Given the genetic diversity of African populations, not all adverse drug reactions are identified during drug development and clinical trials in these nearly homogeneous populations. Adverse drug reactions become apparent after the drugs come to market and are released to other populations, including Africans. The inequity in the involvement of African populations in drug development was apparent during the race to find a cure for COVID-19 (29). Although some strides in genomic research have been made through the H3Africa consortium, pharmacogenomics in Africa remains poorly funded, with the few studies on pharmacogenomics reporting on baseline frequencies of pharmacogene variants without relevant phenotypes. Including African populations in the early stages of drug discovery enables researchers to expose any potential drug to a diversity of underlying pharmacogene profiles, thus avoiding adverse drug reactions that are later identified after the drugs come to market, some of which lead to the drugs being recalled.

It must be borne in mind that most current medicines fail in a proportion of patients; thus, in an effort to mitigate failure of conventional therapies, patients are increasingly turning to the use of medicinal herbs. Herbal medicine research has lagged behind, yet the same pharmacogene enzymes that are important for conventional medicines also metabolize components of herbal medicine or are inhibited or induced by medicinal herbs (122). Pharmacogenomic research has huge potential to contribute to drug development among Africans, especially if herbal medicine gets attention and an influx of funding through integrated drug discovery (28). Funders, researchers, and policy makers should acknowledge the potential that pharmacogenomics has in precision medicine, as it enables preemptive decision-making before drug administration using knowledge of predicted drug–drug and herb–drug interactions. Studying genomic variability among Africans is beneficial to both Africans and the rest of the world as it allows the identification of the extent of the genomic markers of drug response across populations due to African genomic diversity.

A PATH FOR IMPLEMENTATION OF GENETIC MEDICINE: THE GENOMIC WORKFORCE, DATA-DRIVEN GENETIC MEDICINE IMPLEMENTATIONS, AND ETHICAL, LEGAL, AND SOCIAL IMPLICATIONS

Initiatives such as H3Africa have sown the seeds for capacity building, which needs to be nurtured by additional efforts from both the international community and, most importantly, African governments. Despite the progress made through these initiatives with the implementation of

high-standard research facilities and the ongoing training of the next generation of African scientists, challenges in terms of data storage, analysis, and interpretation for the African population and sustainability need to be mitigated. By implementing at least three of the key recommendations of the World Health Organization's *Genomics and World Health* report (143), these additional efforts could ensure the long-term sustainability of genomic medicine on the continent.

First, capacity building of a diverse African genomic workforce, including clinicians, scientists, bioinformaticians, and big data analysts, will be achieved most effectively through the development of academic public research and industrial partnerships between high- and low-income countries and among African countries themselves, encouraged by tax or other incentives as well as market incentives for the private sector to invest in genomic research, as these initiatives are now evolving in Africa [e.g., 54gene (**https://54gene.com**)]. Moreover, specific incentives should be directed toward neglected diseases of the world's poorest people, such as SCD, and through extensive networking within regions where there is evolving expertise in biotechnology (14). Developing an African genomics workforce will be necessary to meet the major need for research across the life span for cohorts of millions of individuals with complex or monogenic diseases (131), which appears to be an imperative to fill the gap of missing African variants in the global genomics landscape and the human reference genome. Therefore, the potential for genomic research to combat these diseases will be realized only through an enforceable global mechanism for public investment in health research and by providing equitable access and capacity to analyze genomic big data from all parts of the world, to ensure that samples and data providers from Africa can also reap the benefits of the third decade following the completion of the Human Genome Project.

Second, to be able to utilize the vast quantities of genomic data that are being generated, African countries must develop a critical mass of expertise in computer science and bioinformatics. For example, through successful H3Africa-funded research projects over the past 10 years, numerous research groups have generated large amounts of whole-exome sequencing data from multiplex and simplex families segregating monogenic conditions such as hearing impairment (3, 93, 136, 141), as well as in isolated cases of monogenic conditions such as SCD (132). Similarly, past and continuing efforts to elucidate the genetic basis of complex conditions such as stroke and other cardiovascular conditions (18), as well as infectious diseases such as AIDS (85) and COVID-19 (53, 126), will likely generate GWAS, whole-exome sequencing, and whole-genome sequencing data for more than 50,000 participants. Moreover, available H3Africa biorepositories have the potential to generate more data from various ethnolinguistically matched controls from numerous African countries. These data, like others available on the continent, will provide a strong basis to engage Africa in data-driven genomic research and genetic medicine implementations in Africa. Coupled with clinical data, building up population-scale databases of genomic plus clinical information will fuel the application of genomic findings for laboratory-based diagnosis, algorithms, and genetic medicine bioinformatic tool sets for better risk interpretation using PRSs (56).

Third, all African countries need to evolve appropriate national frameworks to consider the ethical implications of genomic research and its applications in their own unique social, cultural, economic, and religious contexts. H3Africa and other emerging big data–driven research initiatives in Africa, such as the Sickle in Africa consortium (68), are drawing increasing attention to several long-standing and emerging issues in genetic and genomic research, such as informed consent, privacy, and secondary uses of data (84, 89); community engagement (123); secondary use of genetic information; governance of biorepositories (1, 2); reciprocity and benefit sharing (30, 79); the return of research results and incidental findings (100, 133); questions of assent and reconsent (81); fears of exploitation in research collaborations; and the potential of research commercialization activities to further exacerbate inequalities in global health research (92) (**Figure 5**). Unfortunately, there is a dearth of published research on the views of African communities and

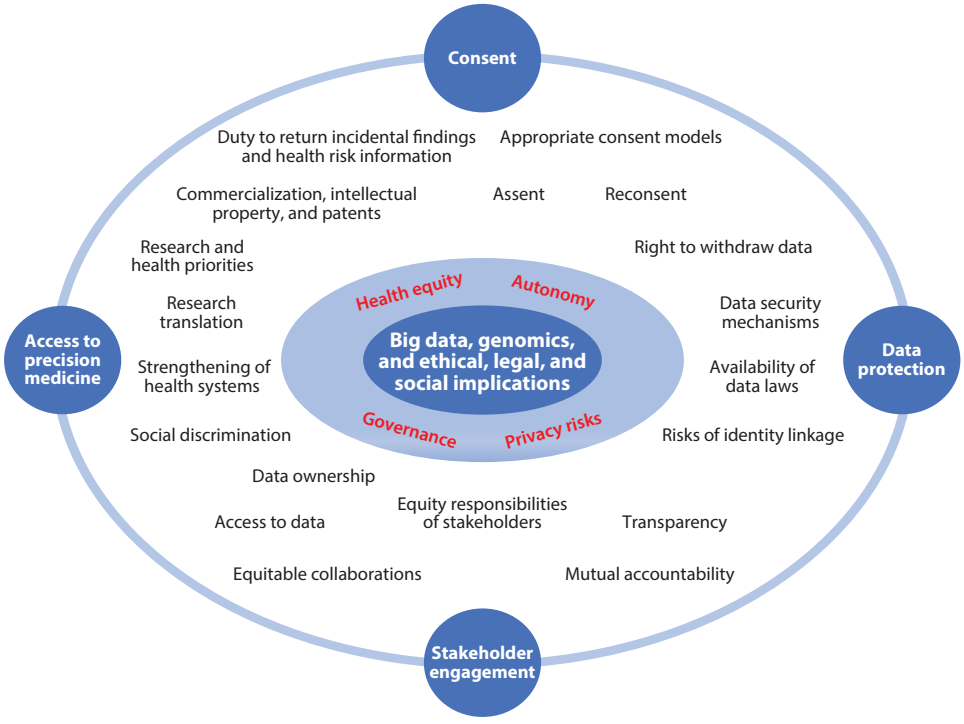


Figure 5

Ethical issues in big data and genomics in health and research.

stakeholders regarding their perspectives on several of these ethical and social issues. The impact of emerging guidelines for vulnerable populations, such as the San Code of Research Ethics (109) and the developing Protection of Personal Information Act in South Africa on data-driven research, is still to be investigated (119). Policies and guidelines for genomic research and medicine in Africa should urgently and empirically investigate equitable governance mechanisms for big data–driven genomic research. Principles that may provide the moral backbone for such frameworks have been identified through conceptual analysis of theories of justice (81) and stakeholder engagement exercises (144) and hinge on African communitarian values such as solidarity, reciprocity, deliberative and consensus-driven decision-making, open sharing, trust, and mutual accountability (80). However, further work is required on the realistic implementation of each of these principles in the context of genomic medicine and big data–driven research in Africa. More importantly, there is a need to give research participants a voice in decision-making on the use of their data and to devise mechanisms that will ensure that no stakeholder group is able to hoard genetic resources or products such as data, intellectual property, and patents for their personal gain at the expense of study communities and other stakeholder groups (80). Africa cannot afford to stand aside from the new genomic and genetic revolution, and as the land of the first human, it should take a bigger place in this endeavor so that the new knowledge will benefit its populations.

CONCLUSION

As the world will be investing in ancestral African genomic variations as a scientific imperative, equitable access to data and analyses for African researchers, along with genetic medicine

implementations related to health issues in African populations, will be essential if African genomics is to reach its full potential as the next frontier of global genetic medicine. In the next decades, this will require an intentional commitment from the international scientific community to equitable collaboration modeled on the successes of initiatives such as H3Africa, and a sustained commitment from major global funding agencies to support genomics and big data–driven research. Most importantly, African governments must take steps to ensure long-term sustainability by implementing at least three key recommendations: supporting academic public research and industrial partnerships to build the genetic medicine workforce, utilizing the vast quantities of genomic big data that are being generated by African countries and internationally from populations of African ancestry in order to develop a critical mass of expertise in computer science and bioinformatics, and evolving appropriate national and global governance frameworks to consider the ethical implications of data-driven genomic research and its applications in these countries' unique social, cultural, economic, and religious contexts.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This article was possible thanks to funding from the National Institutes of Health (grants U01HG009716, 5U24HL135600, U01MH127692, and U54HG009790 to A.W.), the National Institute of Neurological Disorders and Stroke (grant U01HG007044 to G.L.), and the African Academy of Science and Wellcome Trust (grant H3A/18/001 to A.W.).

LITERATURE CITED

1. Abayomi A, Christoffels A, Grewal R, Karam LA, Rossouw C, et al. 2013. Challenges of biobanking in South Africa to facilitate indigenous research in an environment burdened with human immunodeficiency virus, tuberculosis, and emerging noncommunicable diseases. *Biopreserv. Biobank.* 11:347–54
2. Abimiku A, Mayne ES, Joloba M, Beiswanger CM, Troyer J, Wideroff L. 2017. H3Africa biorepository program: supporting genomics research on African populations by sharing high-quality biospecimens. *Biopreserv. Biobank.* 15:99–103
3. Adadey SM, Schrauwen I, Aboagye ET, Bharadwaj T, Esoh KK, et al. 2021. Further confirmation of the association of *SLC12A2* with non-syndromic autosomal-dominant hearing impairment. *J. Hum. Genet.* 66:1169–75
4. Adadey SM, Wonkam-Tingang E, Aboagye ET, Quaye O, Awandare GA, Wonkam A. 2022. Hearing loss in Africa: current genetic profile. *Hum. Genet.* 141:505–17
5. Adedokun BO, Olopade CO, Olopade OI. 2016. Building local capacity for genomics research in Africa: recommendations from analysis of publications in Sub-Saharan Africa from 2004 to 2013. *Glob. Health Action* 9:31026
6. Adeyemo AA, Zaghloul NA, Chen G, Doumatey AP, Leitch CC, et al. 2019. *ZRANB3* is an African-specific type 2 diabetes locus associated with beta-cell mass and insulin response. *Nat. Commun.* 10:3195
7. Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, et al. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061–73
8. Am. Soc. Hum. Genet. 2021. Mission & strategic plan. *American Society of Human Genetics*. **https://www.ashg.org/about/mission-strategic-plan**
9. Appiah-Poku J, Newton S, Kass N. 2011. Participants' perceptions of research benefits in an African genetic epidemiology study. *Dev. World Bioeth.* 11:128–35
10. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, et al. 2015. A global reference for human genetic variation. *Nature* 526:68–74

11. Beck HE, Zimmermann NE, McVicar T, Vergoplan N, Berg A, Wood EF. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* 5:180214. Correction. 2020. *Sci. Data* 7:274
12. Bielinski SJ, St Sauver JL, Olson JE, Larson NB, Black JL III, et al. 2020. Cohort profile: the Right Drug, Right Dose, Right Time: Using Genomic Data to Individualize Treatment Protocol (RIGHT Protocol). *Int. J. Epidemiol.* 49:23–24k
13. Bostoen K. 2018. The Bantu Expansion. In *The Oxford Research Encyclopedia of African History*. Oxford, UK: Oxford Univ. Press. **https://doi.org/10.1093/acrefore/9780190277734.013.191**
14. Bukhman G, Mocumbi AO, Gupta N, Amuyunzu-Nyamongo M, Echodu M, et al. 2021. From a *Lancet* Commission to the NCDI Poverty Network: reaching the poorest billion through integration science. *Lancet* 398:2217–20
15. Campbell MC, Tishkoff SA. 2008. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genom. Hum. Genet.* 9:403–33
16. Cárdenas-Rodríguez N, Carmona-Aparicio L, Pérez-Lozano DL, Ortega-Cuellar D, Gómez-Manzo S, Ignacio-Mejía I. 2020. Genetic variations associated with pharmacoresistant epilepsy (review). *Mol. Med. Rep.* 21:1685–701
17. Carr DF, Bourgeois S, Chaponda M, Takeshita LY, Morris AP, et al. 2017. Genome-wide association study of nevirapine hypersensitivity in a sub-Saharan African HIV-infected population. *J. Antimicrob. Chemother.* 72:1152–62
18. Ceballos FC, Hazelhurst S, Clark DW, Agongo G, Asiki G, et al. 2020. Autozygosity influences cardiometabolic disease-associated traits in the AWI-Gen sub-Saharan African study. *Nat. Commun.* 11:5754
19. Choudhury A, Aron S, Botigué LR, Sengupta D, Botha G, et al. 2020. High-depth African genomes inform human migration and health. *Nature* 586:741–48
20. Claussen M, Dallmeyer A, Bader J. 2017. Theory and modeling of the African Humid Period and the Green Sahara. In *The Oxford Research Encyclopedia of Climate Science*. Oxford, UK: Oxford Univ. Press. **https://doi.org/10.1093/acrefore/9780190228620.013.532**
21. 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.* 37:161–65. Corrigendum. 2005. *Nat. Genet.* 37:328
22. Colombatti R, Martella M, Cattaneo L, Viola G, Cappellari A, et al. 2019. Results of a multicenter universal newborn screening program for sickle cell disease in Italy: a call to action. *Pediatr. Blood Cancer* 66:e27657
23. Cooper A, Ilboudo H, Alibu VP, Ravel S, Enyaru J, et al. 2017. *APOL1* renal risk variants have contrasting resistance and susceptibility associations with African trypanosomiasis. *eLife* 6:e25461
24. Creary LE, Ulug P, Menzel S, McKenzie CA, Hanchard NA, et al. 2009. Genetic variation on chromosome 6 influences F cell levels in healthy individuals of African descent and HbF levels in sickle cell patients. *PLOS ONE* 4:e4218
25. da Rocha JEB, Lombard Z, Ramsay M. 2021. Potential impact of DPYD variation on fluoropyrimidine drug response in sub-Saharan African populations. *Front. Genet.* 12:626954
26. Daak AA, Elsamani E, Ali EH, Mohamed FA, Abdel-Rahman ME, et al. 2016. Sickle cell disease in western Sudan: genetic epidemiology and predictors of knowledge attitude and practices. *Trop. Med. Int. Health* 21:642–53
27. Dai Y, Shaikho EM, Perez J, Wilson CA, Liu LY, et al. 2019. *BCL2L1* is associated with γ-globin gene expression. *Blood Adv.* 3:2995–3001
28. Dandara C, Masimirembwa C, Haffani YZ, Ogutu B, Mabuka J, et al. 2019. African Pharmacogenomics Consortium: consolidating pharmacogenomics knowledge, capacity development and translation in Africa. *AAS Open Res.* 2:19
29. Dandara C, Swart M, Mpeta B, Wonkam A, Masimirembwa C. 2014. Cytochrome p450 pharmacogenetics in African populations: implications for public health. *Expert Opin. Drug Metab. Toxicol.* 10:769–85
30. Dauda B, Joffe S. 2018. The benefit sharing vision of H3Africa. *Dev. World Bioeth.* 18:165–70
31. de Vries J, Landouré G, Wonkam A. 2020. Stigma in African genomics research: gendered blame, polygamy, ancestry and disease causal beliefs impact on the risk of harm. *Soc. Sci. Med.* 258:113091

32. Durvasula A, Sankararaman S. 2020. Recovering signals of ghost archaic introgression in African populations. *Sci. Adv.* 6:5097
33. El-Kamah GY, Mohamed AM, Gad YZ, Abdelhak S, Hennig BJ, et al. 2020. Developing a road map to spread genomic knowledge in Africa: 10th Conference of the African Society of Human Genetics, Cairo, Egypt. *Am. J. Trop. Med. Hyg.* 102:719–23
34. Erwin ER, Addison AP, John SF, Olaleye OA, Rosell RC. 2019. Pharmacokinetics of isoniazid: the good, the bad, and the alternatives. *Tuberculosis* 116:S66–70
35. Esoh KK, Apinjoh TO, Nyanjom SG, Wonkam A, Chimusa ER, et al. 2021. Fine scale human genetic structure in three regions of Cameroon reveals episodic diversifying selection. *Sci. Rep.* 11:1039
36. Esoh KK, Wonkam A. 2021. Evolutionary history of sickle cell mutation: implications for global genetic medicine. *Hum. Mol. Genet.* 30:R119–28
37. Feero WG, Guttmacher AE, Collins FS. 2010. Genomic medicine—an updated primer. *N. Engl. J. Med.* 362:2001–11
38. Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G. 2010. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat. Genet.* 42:1049–51
39. Geard A, Pule GD, Chetcha Chemegni B, Ngo Bitoungui VJ, Kengne AP, et al. 2017. Clinical and genetic predictors of renal dysfunctions in sickle cell anaemia in Cameroon. *Br. J. Haematol.* 178:629–39
40. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, et al. 2010. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* 329:841–45
41. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, et al. 2016. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* 19:1433–41
42. Giri A, Edwards TL, Hartmann KE, Torstenson ES, Wellons M, et al. 2017. African genetic ancestry interacts with body mass index to modify risk for uterine fibroids. *PLOS Genet.* 13:e1006871
43. Gomez F, Hirbo J, Tishkoff SA. 2014. Genetic variation and adaptation in Africa: implications for human evolution and disease. *Cold Spring Harb. Perspect. Biol.* 6:a008524
44. Grollemund R, Branford S, Bostoen K, Meade A, Venditti C, Pagel M. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *PNAS* 112:13296–301
45. Guinto CO, Diarra S, Diallo S, Cissé L, Coulibaly T, et al. 2017. A novel mutation in *KIF5A* in a Malian family with spastic paraplegia and sensory loss. *Ann. Clin. Transl. Neurol.* 4:272–75
46. Gulsuner S, Stein DJ, Susser ES, Sibeko G, Pretorius A, et al. 2020. Genetics of schizophrenia in the South African Xhosa. *Science* 367:569–73
47. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. 2019. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* 20:520–35
48. Gurdasani D, Carstensen T, Fatumo S, Chen G, Franklin CS, et al. 2019. Uganda Genome Resource enables insights into population history and genomic discovery in Africa. *Cell* 179:984–1002.e36
49. Hedt-Gauthier BL, Jeufack HM, Neufeld NH, Alem A, Sauer S, et al. 2019. Stuck in the middle: a systematic review of authorship in collaborative health research in Africa, 2014–2016. *BMJ Glob. Health* 4:e001853
50. Hellwege JN, Jeff JM, Wise LA, Gallagher CS, Wellons M, et al. 2017. A multi-stage genome-wide association study of uterine fibroids in African Americans. *Hum. Genet.* 136:1363–73
51. Hum. Hered. Health Afr. (H3Africa). 2021. Publications. *H3Africa.* **https://h3africa.org/index.php/resource/publication-2**
52. Johns Hopkins Univ. 2021. OMIM gene map statistics. *Online Mendelian Inheritance in Man.* **https://www.omim.org/statistics/geneMap**
53. Kamulegeya R, Kateete DP, Bagaya BS, Nasinghe E, Muttamba W, et al. 2022. Biobanking: strengthening Uganda's rapid response to COVID-19 and other epidemics. *Biopreserv. Biobank.* 20:238–43
54. Keaton JM, Jasper EA, Hellwege JN, Jones SH, Torstenson ES, et al. 2021. Evidence that geographic variation in genetic ancestry associates with uterine fibroids. *Hum. Genet.* 140:1433–40
55. Kengne Kamga K, De Vries J, Nguefack S, Munung NS, Wonkam A. 2021. Explanatory models for the cause of Fragile X Syndrome in rural Cameroon. *J. Genet. Couns.* 30:1727–36

56. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, et al. 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* 50:1219–24
57. Kirkland EB, Heincelman M, Bishu KG, Schumann SO, Schreiner A, et al. 2018. Trends in healthcare expenditures among US adults with hypertension: national estimates, 2003–2014. *J. Am. Heart Assoc.* 7:e008731
58. Krishnakumariamma K, Ellappan K, Muthuraj M, Tamilarasu K, Kumar SV, Joseph NM. 2020. Molecular diagnosis, genetic diversity and drug sensitivity patterns of mycobacterium tuberculosis strains isolated from tuberculous meningitis patients at a tertiary care hospital in South India. *PLOS ONE* 15:e0240257
59. Landouré G, Dembélé K, Cissé L, Samassékou O, Diarra S, et al. 2019. Hereditary spastic paraplegia type 35 in a family from Mali. *Am. J. Med. Genet. A* 179:1122–25
60. Landouré G, Dembélé K, Diarra S, Cissé L, Samassékou O, et al. 2020. A novel variant in the spatacsin gene causing SPG11 in a Malian family. *Neurol. Sci.* 411:116675
61. Landouré G, Samassékou O, Traoré M, Meilleur KG, Guinto CO, et al. 2016. Genetics and genomic medicine in Mali: challenges and future perspectives. *Mol. Genet. Genom. Med.* 4:126–34
62. Landouré G, Zhu PP, Lourenço CM, Johnson JO, Toro C, et al. 2013. Hereditary spastic paraplegia type 43 (SPG43) is caused by mutation in *C19orf12*. *Hum. Mutat.* 34:1357–60
63. Laval G, Peyrégne S, Zidane N, Harmant C, Renaud F, et al. 2019. Recent adaptive acquisition by African rainforest hunter-gatherers of the Late Pleistocene sickle-cell mutation suggests past differences in malaria exposure. *Am. J. Hum. Genet.* 104:553–61
64. Ma S, Dubin AE, Zhang Y, Mousavi SAR, Wang Y, et al. 2021. A role of PIEZO1 in iron metabolism in mice and humans. *Cell* 184:969–982.e13
65. Ma Z, Cheng G, Wang P, Khalighi B, Khalighi K. 2019. Clinical model for predicting warfarin sensitivity. *Sci. Rep.* 9:12856
66. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN. 2005. Heritability of malaria in Africa. *PLOS Med.* 2:1253–59
67. Makani J, Menzel S, Nkya S, Cox SE, Drasar E, et al. 2011. Genetics of fetal hemoglobin in Tanzanian and British patients with sickle cell anemia. *Blood* 117:1390–92
68. Makani J, Sangeda RZ, Nnodu O, Nembaware V, Osei-Akoto A, et al. 2020. SickleInAfrica. *Lancet Haematol.* 7:e98–99
69. Malar. Genom. Epidemiol. Netw. 2014. Reappraisal of known malaria resistance loci in a large multicenter study. *Nat. Genet.* 46:1197–204
70. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, et al. 2016. Genetic misdiagnoses and the potential for health disparities. *N. Engl. J. Med.* 375:655–65
71. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. 2019. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51:584–91
72. Masiye F, Mayosi B, De Vries J. 2017. "I passed the test!" Evidence of diagnostic misconception in the recruitment of population controls for an H3Africa genomic study in Cape Town, South Africa. *BMC Med. Ethics* 18:12
73. McGuire AL, Gabriel S, Tishkoff SA, Wonkam A, Chakravarti A, et al. 2020. The road ahead in genetics and genomics. *Nat. Rev. Genet.* 21:581–96
74. McQuillan MA, Zhang C, Tishkoff SA, Platt A. 2020. The importance of including ethnically diverse populations in studies of quantitative trait evolution. *Curr. Opin. Genet. Dev.* 62:30–35
75. Menzel S, Garner C, Gut I, Matsuda F, Yamaguchi M, et al. 2007. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat. Genet.* 39:1197–99
76. Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, et al. 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585:79–84
77. Mnika K, Mazandu GK, Jonas M, Pule GD, Chimusa ER, et al. 2019. Hydroxyurea-induced miRNA expression in sickle cell disease patients in Africa. *Front. Genet.* 10:509
78. Moodley K, Kleinsmidt A. 2021. Allegations of misuse of African DNA in the UK: Will data protection legislation in South Africa be sufficient to prevent a recurrence? *Dev. World Bioeth.* 21:125–30

79. Munung NS, de Vries J. 2020. Benefit sharing for human genomics research: awareness and expectations of genomics researchers in sub-Saharan Africa. *Ethics Hum. Res.* 42:14–20
80. Munung NS, de Vries J, Pratt B. 2021. Genomics governance: advancing justice, fairness and equity through the lens of the African communitarian ethic of Ubuntu. *Med. Health Care Philos.* 24:377–88
81. Munung NS, Nembaware V, de Vries J, Bukini D, Tluway F, et al. 2019. Establishing a multi-country sickle cell disease registry in Africa: ethical considerations. *Front. Genet.* 10:943
82. Musanabaganwa C, Mihigo B, Tumusime R, Uwanyirigira M, Da Rocha J, et al. 2020. Building skills and resources for genomics, epigenetics, and bioinformatics research for Africa: report of the joint 11th Conference of the African Society of Human Genetics and 12th H3Africa Consortium, 2018. *Am. J. Trop. Med. Hyg.* 102:1417–24
83. Mutapi F. 2019. Africa should set its own health-research agenda. *Nature* 575:567
84. Mweemba O, Musuku J, Mayosi BM, Parker M, Rutakumwa R, et al. 2020. Use of broad consent and related procedures in genomics research: perspectives from research participants in the Genetics of Rheumatic Heart Disease (RHDGen) study in a University Teaching Hospital in Zambia. *Glob. Bioeth.* 31:184–99
85. Mwesigwa S, Williams L, Retshabile G, Katagirya E, Mboowa G, et al. 2021. Unmapped exome reads implicate a role for Anelloviridae in childhood HIV-1 long-term non-progression. *NPJ Genom. Med.* 6:24
86. Natl. Inst. Health. 2021. NIH-wide strategic plan. *National Institutes of Health.* **https://www.nih.gov/about-nih/nih-wide-strategic-plan**
87. Ndadza A, Thomford NE, Mukanganyama S, Wonkam A, Ntsekhe M, Dandara C. 2019. The genetics of warfarin dose-response variability in Africans: an expert perspective on past, present, and future. *OMICS* 23:152–66
88. Ndiaye Diallo R, Gadji M, Hennig BJ, Guèye MV, Gaye A, et al. 2017. Strengthening human genetics research in Africa: report of the 9th meeting of the African Society of Human Genetics in Dakar in May 2016. *Glob. Health Epidemiol. Genom.* 2:e10
89. Nembaware V, Johnston K, Diallo AA, Kotze MJ, Matimba A, et al. 2019. A framework for tiered informed consent for health genomic research in Africa. *Nat. Genet.* 51:1566–71
90. Nembaware V, Mulder N. 2019. The African Genomic Medicine Training Initiative (AGMT): showcasing a community and framework driven genomic medicine training for nurses in Africa. *Front. Genet.* 10:1209
91. Niama FR, Vidal N, Diop-Ndiaye H, Nguimbi E, Ahombo G, et al. 2017. HIV-1 genetic diversity and primary drug resistance mutations before large-scale access to antiretroviral therapy, Republic of Congo. *BMC Res. Notes* 10:243
92. Nordling L. 2021. Give African research participants more say in genomic data, say scientists. *Nature* 590:542
93. Oluwole OG, Esoh KK, Wonkam-Tingang E, Manyisa N, Noubiap JJ, et al. 2020. Whole exome sequencing identifies rare coding variants in novel human-mouse ortholog genes in African individuals diagnosed with non-syndromic hearing impairment. *Exp. Biol. Med.* 246:197–206
94. Ormond KE, Laurino MY, Barlow-Stewart K, Wessels TM, Macaulay S, et al. 2018. Genetic counseling globally: Where are we now? *Am. J. Med. Genet. C* 178:98–107
95. Pepper MS. 2011. Launch of the Southern African Human Genome Programme. *S. Afr. Med. J.* 101:287–88
96. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. 2010. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat. Commun.* 1:104
97. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. 2013. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet* 381:142–51
98. Pule GD, Mowla S, Novitzky N, Wonkam A. 2016. Hydroxyurea down-regulates *BCL11A*, *KLF-1* and *MYB* through miRNA-mediated actions to induce γ-globin expression: implications for new therapeutic approaches of sickle cell disease. *Clin. Transl. Med.* 5:15
99. Pule GD, Ngo Bitoungui VJ, Chetcha Chemegni B, Kengne AP, Antonarakis S, Wonkam A. 2015. Association between variants at *BCL11A* erythroid-specific enhancer and fetal hemoglobin levels among

sickle cell disease patients in Cameroon: implications for future therapeutic interventions. *OMICS* 19:627–31
100. Ralefala D, Kasule M, Wonkam A, Matshaba M, de Vries J. 2020. Do solidarity and reciprocity obligations compel African researchers to feedback individual genetic results in genomics research? *BMC Med. Ethics* 21:112
101. Rotimi CN. 2004. Inauguration of the African Society of Human Genetics. *Nat. Genet.* 36:544
102. Rotimi CN, Abayomi A, Abimiku A, Adabayeri VM, Adebamowo C, et al. 2014. Enabling the genomic revolution in Africa. *Science* 334:1346–48
103. Rotimi CN, Bentley AR, Doumatey AP, Chen G, Shriner D, Adeyemo A. 2017. The genomic landscape of African populations in health and disease. *Hum. Mol. Genet.* 26:R225–36
104. Røttingen JA, Chamas C, Goyal LC, Harb H, Lagradae L, Mayosi BM. 2012. Securing the public good of health research and development for developing countries. *Bull. World Health Organ.* 90:398–400
105. Royal CDM, Babyak M, Shah N, Srivatsa S, Stewart KA, et al. 2021. Sickle cell disease is a global prototype for integrative research and healthcare. *Adv. Genet.* 2:e10037
106. Rumaney MB, Ngo Bitoungui VJ, Vorster AA, Ramesar R, Kengne AP, et al. 2014. The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival. *PLOS ONE* 9:e100516
107. Sandoval R, Monteghirfo M, Salazar O, Galarza M. 2020. Resistencia cruzada entre isoniacida y etionamida y su alta correlación con la mutación C-15T en aislamientos de *Mycobacterium tuberculosis* de Perú [Cross-resistance between isoniazid and ethionamide and its strong association with mutation C-15T in *Mycobacterium tuberculosis* isolates from Peru]. *Rev. Argent. Microbiol.* 52:36–42
108. Sangaré M, Hendrickson B, Sango HA, Chen K, Nofziger J, et al. 2014. Genetics of low spinal muscular atrophy carrier frequency in sub-Saharan Africa. *Ann. Neurol.* 75:525–32
109. Schroeder D, Chatfield K, Singh M, Chennells R, Herissone-Kelly P. 2019. The San code of research ethics. In *Equitable Research Partnerships: A Global Code of Conduct to Counter Ethics Dumping*, pp. 73–87. Cham, Switz.: Springer
110. Schroeder D, Cook J, Hirsch F, Fenet S, Muthuswamy V. 2018. Ethics dumping: introduction. In *Ethics Dumping: Case Studies from North-South Research Collaborations*, ed. D Schroeder, J Cook, F Hirsch, S Fenet, V Muthuswamy, pp. 1–8. Cham, Switz.: Springer
111. Sherman RM, Forman J, Antonescu V, Puiu D, Daya M, et al. 2019. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* 51:30–35
112. Shriner D, Rotimi CN. 2018. Whole-genome-sequence-based haplotypes reveal single origin of the sickle allele during the Holocene Wet Phase. *Am. J. Hum. Genet.* 102:547–56
113. Sierra B, Triska P, Soares P, Garcia G, Perez AB, et al. 2017. *OSBPL10*, *RXRA* and lipid metabolism confer African-ancestry protection against dengue haemorrhagic fever in admixed Cubans. *PLOS Pathog.* 13:e1006220
114. Singh GM, Micha R, Khatibzadeh S, Shi P, Lim S, et al. 2015. Global, regional, and national consumption of sugar-sweetened beverages, fruit juices, and milk: a systematic assessment of beverage intake in 187 countries. *PLOS ONE* 10:e0124845. Correction. 2019. *PLOS ONE* 14:e0214344
115. Sirugo G, Williams SM, Royal CDM, Newport MJ, Hennig BJ, et al. 2010. Report on the 6th African Society of Human Genetics (AfSHG) meeting, March 12–15, 2009, Yaoundé, Cameroon. *Am. J. Trop. Med. Hyg.* 83:226–29
116. Sirugo G, Williams SM, Tishkoff SA. 2019. The missing diversity in human genetic studies. *Cell* 177:26–31. Erratum. 2019. *Cell* 177:1080
117. Skov L, Coll Macià M, Sveinbjörnsson G, Mafessoni F, Lucotte EA, et al. 2020. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* 582:78–83
118. Soko ND, Chimusa E, Masimirembwa C, Dandara C. 2019. An African-specific profile of pharmacogene variants for rosuvastatin plasma variability: limited role for SLCO1B1 c.521T>C and ABCG2 c.421A>C. *Pharmacogenom. J.* 19:240–48
119. Staunton C, Adams R, Anderson D, Croxton T, Kamuya D, et al. 2020. Protection of Personal Information Act 2013 and data protection for health research in South Africa. *Int. Data Priv. Law* 10:160–79
120. Stephens C. 2007. Natural selection. In *Philosophy of Biology*, ed. M Matthen, C Stephens, pp. 111–27. Amsterdam: North Holland

121. Swart M, Skelton M, Ren Y, Smith P, Takuva S, Dandara C. 2013. High predictive value of *CYP2B6* SNPs for steady-state plasma efavirenz levels in South African HIV/AIDS patients. *Pharmacogenet. Genom.* 23:415–27
122. Thomford NE, Awortwe C, Dzobo K, Adu F, Chopera D, et al. 2016. Inhibition of CYP2B6 by medicinal plant extracts: implication for use of efavirenz and nevirapine based highly active anti-retroviral therapy (HAART) in resource-limited settings. *Molecules* 21:211
123. Tindana P, Campbell M, Marshall P, Littler K, Vincent R, et al. 2017. Developing the science and methods of community engagement for genomic research and biobanking in Africa. *Glob. Health Epidemiol. Genom.* 2:e13
124. Toivanen H, Ponomariov B. 2011. African regional innovation systems: bibliometric analysis of research collaboration patterns 2005–2009. *Scientometrics* 88:471–93
125. Urio F, Nkya S, Rooks H, Mgaya JA, Masamu U, et al. 2020. F cell numbers are associated with an X-linked genetic polymorphism and correlate with haematological parameters in patients with sickle cell disease. *Br. J. Haematol.* 191:888–96
126. Wilkinson E, Giovanetti M, Tegally H, San JE, Lessells R, et al. 2021. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science* 374:423–31
127. Williams LM, Qi Z, Batai K, Hooker S, Hall NJ, et al. 2018. A locus on chromosome 5 shows African ancestry-limited association with alloimmunization in sickle cell disease. *Blood Adv.* 2:3637–47
128. Williams SM, Tishkoff SA. 2011. Exploring genomic studies in Africa. *Genome Med.* 3:45
129. Williams TN, Mwangi TW, Wambua S, Alexander ND, Kortok M, et al. 2005. Sickle cell trait and the risk of *Plasmodium falciparum* malaria and other childhood diseases. *J. Infect. Dis.* 192:178–86
130. Wonkam A. 2020. Investigating the missing heritability of fetal haemoglobin level in Africa. *Br. J. Haematol.* 191:668–70
131. Wonkam A. 2021. Sequence three million genomes across Africa. *Nature* 590:209–11
132. Wonkam A, Chimusa ER, Mnika K, Pule GD, Ngo Bitoungui VJ, et al. 2020. Genetic modifiers of long-term survival in sickle cell anemia. *Clin. Transl. Med.* 10:e152
133. Wonkam A, de Vries J. 2020. Returning incidental findings in African genomics research. *Nat. Genet.* 52:17–20
134. Wonkam A, Kenfack MA, Muna WFT, Ouwe-Missi-Oukem-Boyer O. 2011. Ethics of human genetic studies in sub-Saharan Africa: the case of Cameroon through a bibliometric analysis. *Dev. World Bioeth.* 11:120–27
135. Wonkam A, Makani J. 2019. Sickle cell disease in Africa: an urgent need for longitudinal cohort studies. *Lancet Glob. Health* 7:E1310–11
136. Wonkam A, Manyisa N, Bope CD, Dandara C, Chimusa ER. 2021. Whole exome sequencing reveals pathogenic variants in *MYO3A*, *MYO15A* and *COL9A3* and differential frequencies in ancestral alleles in hearing impairment genes among individuals from Cameroon. *Hum. Mol. Genet.* 29:3729–43
137. Wonkam A, Ngo Bitoungui VJ, Vorster AA, Ramesar R, Cooper RS, et al. 2014. Association of variants at *BCL11A* and *HBS1L-MYB* with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. *PLOS ONE* 9:e92506
138. Wonkam A, Njamnshi AK, Angwafo FF. 2006. Knowledge and attitudes concerning medical genetics amongst physicians and medical students in Cameroon (sub-Saharan Africa). *Genet. Med.* 8:331–38
139. Wonkam A, Tekendo CN, Sama DJ, Zambo H, Dahoun S, et al. 2011. Initiation of a medical genetics service in sub-Saharan Africa: experience of prenatal diagnosis in Cameroon. *Eur. J. Med. Genet.* 54:e399–404
140. Wonkam-Tingang E, Schrauwen I, Esoh KK, Bharadwaj T, Nouel-Saied LM, et al. 2020. Bi-allelic novel variants in *CLIC5* identified in a Cameroonian multiplex family with non-syndromic hearing impairment. *Genes* 11:1249
141. Wonkam-Tingang E, Schrauwen I, Esoh KK, Bharadwaj T, Nouel-Saied LM, et al. 2021. A novel variant in *DMXL2* gene is associated with autosomal dominant non-syndromic hearing impairment (DFNA71) in a Cameroonian family. *Exp. Biol. Med.* 246:1524–32
142. World Health Organ. 2019. Malaria incidence (per 1 000 population at risk) (Malaria). *World Health Organization*. **https://apps.who.int/gho/data/node.imr.SDGMALARIA?lang=en**

143. World Health Organ. Advis. Comm. Health Res. 2002. *Genomics and world health: report of the Advisory Committee on Health Research*. Rep., World Health Organ., Geneva
144. Yakubu A, Tindana P, Matimba A, Littler K, Munung NS, et al. 2018. Model framework for governance of genomic research and biobanking in Africa – a content description. *AAS Open Res*. 1:13
145. Yalcouyé A, Diallo SH, Coulibaly T, Cissé L, Diallo S, et al. 2019. A novel mutation in the *GARS* gene in a Malian family with Charcot-Marie-Tooth disease. *Mol. Genet. Genom. Med.* 7:782
146. Zeberg H, Pääbo S. 2020. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 587:610–12