The critical needs and challenges for genetic architecture studies in Africa

Alicia R. Martin[1,2,3], Solomon Teferra[4,5], Marlo Möller[6], Eileen G. Hoal[6], Mark J. Daly[1,2,3]

[1] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA

[2] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[3] Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[4] Department of Psychiatry, School of Medicine, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia.

[5] Department of Epidemiology, Harvard T. H. Chan School of Public Health, Harvard University, Boston, USA

[6] DST/NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Cape Town, South Africa

Abstract

Human genetic studies have long been vastly Eurocentric, raising a key question about the generalizability of these study findings to other populations. Because humans originated in Africa, these populations retain more genetic diversity, and yet individuals of African descent have been tremendously underrepresented in genetic studies. The diversity in Africa affords ample opportunities to improve fine-mapping resolution for

 doi:10.20944/preprints201806.0201.v1

associated loci, discover novel genetic associations with phenotypes, build more generalizable genetic risk prediction models, and better understand the genetic architecture of complex traits and diseases subject to varying environmental pressures. Thus, it is both ethically and scientifically imperative that geneticists globally surmount challenges that have limited progress in African genetic studies to date while meaningfully including African investigators, as greater inclusivity and enhanced research capacity affords enormous opportunities to accelerate genomic discoveries that translate more effectively to all populations. We review the advantages and challenges of studying the genetic architecture of complex traits and diseases in Africa. For example, with greater genetic diversity comes greater ancestral heterogeneity; this higher level of understudied diversity can yield novel genetic findings, but some methods that assume homogeneous population structure and work well in European populations may work less well in the presence of greater diversity and heterogeneity in African populations. Consequently, we advocate for methodological development that will accelerate studies important for all populations, especially those currently underrepresented in genetics.

Keywords: genetic architecture; Africa; GWAS; health disparities

Historical biases in genetic studies

Nearly a decade ago, 96% of participants in genome-wide association studies (GWAS) were of European descent [1]. While the representation of non-European participants in genetic studies is growing—European descent individuals now account for 81% of GWAS participants—the proportion of participants with African ancestry has increased

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

negligibly from 0.57% to 3% [2]. This bias results in many medically relevant consequences emanating from an interpretability gap that varies across ancestries [3,4]. For example, based on standard practice at leading genetic testing laboratories, African Americans were more likely than European descent patients to be incorrectly told they have a genetic mutation that increases their risk of hypertrophic cardiomyopathy, an early-onset life-threatening heart disease [5]. Such genetic misdiagnoses are preventable, the authors point out, with the inclusion of even a small number of ancestry-matched controls [5]. Additionally, drug metabolism genes such as *CYP3A4* contain mutations that alter dosage requirements (e.g., for warfarin); but with genetic variation disproportionately uncatalogued in terms of frequency and clinical consequence among African populations and individuals [6], genotype-based dosage adjustment would be less valuable than in Europeans. In the US, the National Human Genome Research Institute (NHGRI) at the National Institutes of Health (NIH) has emphasized the need to increase diversity in genetic studies [7]. This prioritization is an important step that, if heeded, will aid interpretations in medical genomics for all ethnicities (including Europeans) [8]. Greater inclusivity of African populations in medical genomics endeavors is important for accelerating genomic discoveries, enabling the reconstruction of modern human origins, producing results that can be translated across populations more accurately, identifying genetic associations between traits and genetic variation present in some populations but not others, and building research capacity in Africa.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

These study biases have not happened in a vacuum (see "*Existing challenges to surmount for African genetics studies*"), but have had widespread consequences for the tools and resources available to perform GWAS in African populations. Genotyping arrays, because they are often based on European sequencing resources, have traditionally been biased towards alleles most frequent and imputable in European populations [9,10], compounding biases in which GWAS identify associated variants that are most common in the populations in which traits are studied [11,12]. In contrast, array backbones that prioritize SNPs that are polymorphic and maximize tagged variants across all populations improve imputation performance, providing more even genomic coverage [13]. Perhaps more importantly, the largest imputation panels are vastly Eurocentric, which shortchanges representation of the greater haplotypic diversity present in Africans as a consequence of more rapid linkage disequilibrium (LD) decay from deeper recombination history [14,15]GenomesProjectConsortium:2015gk}. Furthermore, the most widely available sequenced resources that act as ancestry references for globally diverse populations tend to have biased representation towards African Americans and West Africans, leaving huge swaths of African diversity uncatalogued [8,12].

Existing challenges to surmount for African genetics studies

To empower African genetic studies and build research capacity that will aid biological understanding across a diverse swath of humanity, we review many challenges relating to history, infrastructure, funding, ethics, and respect that need to be confronted and continually addressed.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

Historical

Africa has been subjected to a long, violent, and oppressive colonial history that has bred suspicion and an anticipation of exploitation of resources. This understandable mistrust continues to create strain in ongoing relations, with new actors such as China in addition to European groups scrambling for African resources [16,17]. The impact on research collaborations is evident, with some authors calling it 'neo-colonial science' [18]. Such strained relations are more pronounced in collaborations involving genetic studies, especially when shipping samples out of Africa and the Global South [19]. Some authors argue for maintaining the 'genomic sovereignty' of Africans, emphasizing the need to ensure such collaborations benefit Africans and address issues around ownership of African genetic material [20]. Proponents of international collaborations argue that working with high income countries will eventually ensure equity, justice, and benefit to Africans, with capacity building for genomic research being an immediate benefit for African institutions, although concerns have been raised about the sustainability of these efforts [21]. Ongoing tensions weigh the big benefit to Africans from the inclusion of more African DNA in global studies and African researchers involved in these collaborations against the challenges of promoting science in Africa in a way that accelerates the integration and importing of the best science around the world into Africa.

Infrastructural

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

Conducting genetic studies in Africa is not an easy task. The infrastructural problems range from unreliable or no electricity in clinics and laboratories that process samples, through impassable roads in some rural areas, to crime or political instability making some areas dangerous and/or inaccessible for researchers. Many African countries do not have sufficient laboratory equipment or facilities for genomics research, and most require imported reagents. Importing is not only a lengthy process, but also costly—reagents are often many times more expensive in Africa than Western countries in real terms, not including shipment costs. There is also a lack of biobanks, partially due to power interruptions affecting storage and processing of samples. There are African institutions with experience in large-scale human genetic analyses, and the H3ABionet consortium has developed further core bioinformatics infrastructure in Africa [22]. However, high-speed internet connections and powerful computers are not always available to access large data files. Human resources issues can also be a challenge, namely lack of skilled personnel, high staff turnover with attrition due to inadequate pay, and competing demands for time from qualified staff. Relatedly, brain drain is a major issue, as many skilled African scientists leave the continent in search of greener pastures [23,24]. To be sensitive to these infrastructural challenges and the desire that African-based researchers get the first opportunity to analyze their own genetic data, some major international research initiatives such as H3Africa has required a relatively long embargo period on publication reserved only for African researchers [25]. Further efforts to connect African researchers to adequate computing power (e.g., wireless connections to cloud computing) may offer more direct means to facilitate and accelerate research. Compared with the relative ease of acquiring samples in areas

 doi:10.20944/preprints201806.0201.v1

close to major research institutes, with reliable roads, power, and social structure, it is no surprise that databanks have focused on European/white populations, but it is scientifically and ethically imperative that geneticists rise to these challenges for the benefit of all.

Funding

Genetic research is expensive, and a lack of attention from African policy makers in a resource-limited setting (i.e. the entire African continent having only a third the GDP of the United States) is primarily driven by competing priorities for more immediate public health concerns including infectious diseases over inherited conditions [26,27]. Data generation is still the most expensive part of genomics, whereas data analysis is more affordable and therefore a viable option for capacity building in resource-limited areas [28]. Furthermore, journals from the developed world typically exist behind expensive pay-walls that are not accessible to all researchers and do not always encourage publication of work from the Global South, often returning manuscripts without review citing a lack of “sufficient general interest.” Having fewer publications has a knock-on effect on future grant funding and attracting students.

Nearly all funding for genetics research comes from outside Africa, raising questions for African scientists about the utility of investigating disease genetics with less long-term funding security and intellectual freedom to prioritize their field of study. Incentives differ from the West, heavily favoring medicine over research training—clinical demands are heavier, PhD programs are scarce, and research often does not pay. However, some

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

external research project funding priorities in genetics and genomics are being led by African scientists, including most notably the Human Heredity and Health in Africa (H3Africa) Initiative—providing global funding by the NIH (US) and Wellcome Trust (UK) totaling more than $216 million in 2015 for 185 projects in 28 African countries [29]. The aim of H3Africa is to build the capacity for African scientists to conduct genomic research on heritable diseases afflicting Africans [30,31]. Additionally, the Bill & Melinda Gates Foundation has made massive investments to improve health in Africa, with agricultural genetics being a key component of their development strategy. This international support is essential for African geneticists to continue their research [21].

Ethical

Ethics review boards may lack capacity and familiarity with genomics research, which creates challenges for making informed decisions about long-term, large-scale collaborative genetic studies that in turn can delay funded projects [28]. These challenges are partially driven by restrictive ethical guidelines and uncertainty about the benefits of such studies specifically to African populations [28]. Unlike in the US, genetics projects are subject to ethics review both at the provincial and national levels as a legacy of colonialism, but this can lead to long delays on the order of years. Ethics approval by regulatory bodies in Africa is mostly restricted to project-specific research questions, often raising questions around 'broad consent' and 'indefinite storage' of samples that are not easy to answer. A primary concern about loss of control and ownership over the DNA samples arises when they are shipped abroad [32]. Burgeoning interest in building large-scale genomics collaborations in Africa has

resulted in a recent best practices ethical framework from H3Africa for genomics research and biobanking in Africa [25]. Some communities have set up local councils to oversee the research projects allowed in the community and the publications that result from the research [33]. While these are excellent in theory, in practice there can be long delays, misunderstanding due to unfamiliarity of lay people with scientific and genetic terms, and a lack of continuity in leadership. Consequently, even when extensive consultation on planned or existing research projects has taken place, this often needs to be repeated at each subsequent visit. New council leaders sometimes try to enforce sample destruction before allowing further sampling, even when consent forms specify long sample storage. A middle ground of community members more familiar with research methodology and terminology that is acceptable to the council would be the ideal solution, but is not always feasible. Furthermore, while returning scientific discoveries to communities or participants can and should be the norm, it is challenging to re-contact study participants in less stable communities as people lose cell phones or move for employment opportunities.

Respect and consent

To ensure mutual respect in collaborative studies of African genetics, it is important to avoid generalizing "African-ness" in such a vast continent, comprising not only more genetic diversity than the rest of the world, but also so many cultures, language groups, and world views, some of which may be marginalized or discriminated against. Consequently, it can be more challenging to establish and sustain multi-national African collaborations than in the West. However, it is important to obtain diverse perspectives

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

from continental Africans when communicating science broadly. Furthermore, meaningful engagement with African colleagues is vital to healthy collaborations and to avoid tokenism. Additionally, obtaining informed consent for genomics research can be complex in any setting, but poses even more of a challenge where there are lower income and literacy levels or language barriers. Furthermore, some diseases such as mental illness are subject to greater stigma in some African communities, requiring culturally sensitive enrollment and awareness of differences from the U.S. and Europe. Participants may misunderstand the study purpose or expect benefits that are not included, such as better disease treatment [34] or individual genetic results useful for land claims. Additionally, in some African societies, decisions to participate in research studies are made collectively as well as at the individual level [35], necessitating consultation with community leaders.

Communicating science respectfully can also be challenging when nomenclature is subject to sociopolitical debate. The naming of the descendants of the original hunter-gatherers of Southern Africa is a case in point. In an attempt to be politically correct, many population geneticists use the word "KhoeSan" to refer to the Khoe and San groups collectively. However, the San Council of Southern Africa prefers to keep these terms separate (i.e., San and Khoe or Nama) to denote the different cultural groups. Many of the "San" individuals prefer being called "Bushmen," while others consider the word to be pejorative. Labels are only useful insofar as they are universally informative, and it is important to decide on labels that are acceptable to the communities and

 doi:10.20944/preprints201806.0201.v1

investigators. Wide pre-publication consultation is obviously necessary to be respectful of research participants [36], but complete consensus is unlikely.

Genetic insights into population diversity, substructure, and origins in Africa

A clear understanding of African population structure is critical for the optimal design of genotype-phenotype studies, with low-hanging opportunities currently available due to a relative dearth of prior studies. Africa is the cradle of humanity, resulting globally in the highest levels of genetic diversity within and between groups [37]. Continental-level ancestral divergence has been impacted by demographically complex and regionally distinct differences in admixture patterns with multiple hunter-gatherer, pastoralist, and Eurasian populations [15,38-42], which is correlated with Afro-Asiatic, Nilo-Saharan, Khoisan, and Niger-Congo linguistic families (**Figure 1**) [43]. This complexity has hampered the investigation of health aspects likely to have a genetic basis, as fewer studies have deeply characterized the genetic make-up of those populations. In South Africa for example, efforts to admixture map tuberculosis susceptibility were waylaid for years by a detour to disentangle the very complex admixture history of study participants particularly given the relative dearth of reference genetic resources and ancestry studies [44].

The most widespread ancestry component in Africa is among the Niger-Congo language group, reflecting the Bantu Expansion [15]. This expansion of agro-pastoralist groups during the past ~5,000 years led to coexistence, assimilation, and displacement of hunter-gatherer groups, including the KhoeSan in southern Africa [42,45,46]. Within

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

and between African populations, fine-scale structure is widespread, reflecting differences in subsistence strategies, cultural, ecological, and geographical clines or boundaries [44]. Most recent descendants outside of Africa were Bantu speakers forcibly brought from West and Central Africa to the Americas during the African slave trade, and subsequent admixture events and movements trace structured patterns to their origins [47,48]. Genetic analysis indicates that originally, humans most likely migrated out of Africa via a predominantly northern route via Egypt [49].

Modern and ancient DNA (aDNA) studies are important for understanding the genetic architecture of traits and diseases because they provide insight into human origins and the demographic events in population history that shaped genetic diversity across populations. Recent work provides a framework for reconstructing the genomes of recent African ancestors to disentangle movements of diverse groups [38,50]. Additionally, the latest archeological excavations in and near Africa provide further evidence that early human evolution involved more than a single region of the continent [51-54]. While aDNA studies in Africa can be challenging due to warmer climates and thus poor preservation conditions, the identification of the petrous bone of the inner ear and cementum layer of teeth roots as sources of higher fidelity aDNA holds promise [55-57]. Two of the first aDNA studies in Africa [58,59] examined ancient Ethiopian and South African remains, finding highly complex demographic patterns reflecting the Bantu expansion, out-of-Africa and back-to-Africa migration patterns, and long-range KhoeSan migration events, with both studies highlighting yet-unexplainable complexities of varying genetic diversity as a function of time and space. These studies along with

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

recent archeological evidence also indicate that modern humans originated and diverged longer ago than previously thought (~250-350kya), indicating that multiple hominid species occupied Africa simultaneously during the early Middle Stone Age [58,59].

Evolutionary studies identify unique adaptations in Africa

Natural selection in response to different environments in Africa has led to adaptation, including to variable ultraviolet radiation exposure, immune response, diet, altitude, and other factors. Consequently, the genetic architecture of adaptive traits can vary considerably by population [60,61]. Some well-characterized examples of genetic adaptation in Africa have been reviewed previously [62]. Additional recent examples of variants associated with skin pigmentation include in/near *SLC24A5, TYRP1, SMARCA2/VLDLR, SNX13, MFSD12, DDB1, TMEM138, OCA,* and *HERC2* [60,63]. Other recent work compared haplotypes from diverse African populations to demonstrate a single origin ~7300 years ago (prior to the Bantu expansion) of the sickle cell allele, which confers disease in homozygous state but protects heterozygous individuals from *Plasmodium falciparum* [64].

GWAS design challenges in Africa

Unlike most of the GWAS and complex trait studies that have been conducted in Europe, assumptions of largely homogeneous population structure are more likely to be violated in Africa, as few if any populations have remained isolated and unchanged even over the past 4000 years [65]. This higher level of diversity across African

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

populations relative to other continental populations [42,66] creates greater challenges when attempting to balance case/control collections at the outset of many studies due to greater complexities in population structure, including variable patterns of LD between study sites. Consequently, false positives are more likely to arise from confounding due to unaccounted population stratification, especially for rare variants, which principal components analysis is insufficient to correct [67,68]. Higher rates of genetic diversity also result in a larger number of effective genetic tests, meaning that the standard GWAS multiple testing threshold of $p < 5e-8$ needs to be roughly twice as stringent in Africa ($p < \sim 2.5e-8$) [69]. Furthermore, additional challenges arise from a dearth of large, easily accessible reference panels in Africa. While the African Genome Variation Project has worked to ameliorate this gap, data access is somewhat more challenging than e.g., the publicly available 1000 Genomes Project [15]. This difference in data availability re-emphasizes the tension between protected access for African scientists versus ensuring that African research and investigators are not left behind in global genetics efforts.

Some statistical methods for population and complex trait genetics that assume homogeneity may not work as well in Africans with higher diversity and more structure, and thus methodological advancements that explicitly account for heterogeneous population structure and admixture over a range of time periods will be especially helpful for African studies [70]. For example, heritability estimates in the presence of admixture can be inflated due to biased SNP effect size estimates as a consequence of different LD structure [71]. Alternatively, higher heritability estimates may be driven by

 doi:10.20944/preprints201806.0201.v1

higher relatedness among geographically proximal individuals. The presence of more population structure can create challenges disentangling the heritable component due to genetics versus similar environments [71]. Other methods for inferring heritability including LD score regression are suboptimal in the presence of admixture as LD from these populations are often not reflective of the study cohort and vary in local structure [72,73]. Other methods for inferring genetic architecture, including Bayesian linear mixed models (LMMs) such as the Bayesian sparse LMM (BSLMM), Bayes R, and BOLT-LMM, have been shown through simulations and applications in admixed populations to be effective at controlling population structure, cryptic relatedness, and further increase power in structured populations [74-77]. Additionally, studies of evolutionarily important traits such as malaria resistance may be further complicated by high levels of genetic diversity in the parasite, multiple independent origins of resistance loci, and allelic heterogeneity, necessitating specialized association approaches [78]. These studies demonstrate that more advanced GWAS methods may be more fruitful generally, but especially in Africa where higher rates of substructure are more typical than in Europe.

Advantages and opportunities for genetic architecture studies in Africa

While there are challenges to overcome for GWAS in Africa, the scientific and ethical opportunities for doing these large-scale studies should not be discounted. Growing inclusion of African American participants in large-scale medical genomics studies is crucial, but still leaves behind many populations and large swaths of sub-Saharan African genetic diversity, which may greatly increase our understanding of complex trait

 doi:10.20944/preprints201806.0201.v1

genetic architecture [79]. There is more genetic and often phenotypic diversity in Africa that has been understudied, meaning there is considerable low-hanging fruit that will ensure novel findings and insights into the genetic architectures and etiologies of complex traits. More rapid LD decay in Africa also means there is greater fine-mapping resolution to pinpoint causal variants influencing traits than will be discovered in any other global population [80], as reviewed recently [81]. For example, several variants in *TCF7L2* were associated with type 2 diabetes in European and East Asian populations in the early GWAS era, but the number of candidate loci were narrowed considerably via comparison with more diverse West African cohorts, even with smaller cohort sizes [82].

Major opportunities also present themselves in precision medicine. For example, polygenic risk scores have been of growing interest as the scale of GWAS now offers a low-cost test that, when coupled with other clinical factors, outpaces the clinical status quo. However, these polygenic risk scores generalize poorly across diverse populations [11]. Studies have now consistently demonstrated across phenotypes that European GWAS results predict genetic risk 2-5X less accurately in non-European populations, performing the worst in African Americans (and by extension, likely even worse in eastern, central, and southern African populations) [83-86]. A typical but somewhat misguided argument in favor of immediate implementation of polygenic risk scores in clinical practice is that general clinical lab tests (e.g., complete blood count, cholesterol, vitamin D, etc) are often interpreted differently across ethnicities and are sometimes more reliable in European descent populations. However, this interpretability gap for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2018 doi:10.20944/preprints201806.0201.v1

current clinical tests is less acutely and consistently worse in non-European populations than genetic risk prediction because the underlying biology remains the same, such that drugs for all diseases do not routinely work many-fold better in European than African-descent populations. Further, new methods for population-specific interpretation of common clinical lab tests will more easily ensure better prognostic value than existing reference intervals [87]. In contrast, the most significant and highest frequency genetic variants from GWAS used to predict genetic risk are *not* likely to be the same across populations, even when the underlying causal variants are the same. This is due to GWAS discovery biases, as variants used to predict risk tend to be more common and explain more phenotypic variation in the study population. We are developing new analytical methods that will help even the generalizability of genetic prediction accuracy by considering LD within and between populations; however, the only way genetic prediction power of inherited diseases in non-Europeans can truly be made equal is with massive investments to produce similar-sized GWAS of these phenotypes in non-European populations. Additionally, discoveries based on African genetics contribute to world knowledge, but most African population groups are sufficiently different that insights made from trans-ethnic studies (e.g. European and East Asian) can similarly be gained by analyzing multiple GWAS of different African populations.

If one of the major missions of genetic analysis is to understand the biological basis and evolutionary origins of diseases and traits and to use this knowledge to perform biologically-informed drug discovery, human evolution tells us that Africa has a huge role to play. Progress has so far been slower, but we emphasize that this is not due to

inherent differences in research ability, but rather a need for increased capacity and outreach efforts to collaboratively engage African investigators. Several outstanding examples of this great potential already exist, such as the Southern African Human Genome Programme (SAHGP), one of the first genetic architecture studies of African participants analyzed by African investigators and fully funded by an African government [45]. International collaborations have also blazed the trail for meaningful collaborations with deep investments in building research capacity in human genomics, such as partnerships between the African Center of Excellence for Genomics of Infectious Diseases (ACEGID) and the Broad Institute, as well as the Global Initiative for Neuropsychiatric Genetics Education in Research (GINGER) program. Calls from African researchers for funding and research capacity building [79] should be thoughtfully heeded to ensure that those countries with the greatest public health needs are not the last to benefit from genetics research.

Acknowledgments

We thank Brenna Henn and Lluis Quintana-Murci for inviting this review. We also thank Karestan Koenen, Lori Chibnik, and Anne Stevenson for their thoughtful comments.

Figure captions

Figure 1 – Map of publicly available African samples and corresponding language families from previous studies. Reference data comes from several previous studies [12,15,40-42,49,66,88,89].

Disclosures

The authors declare no conflict of interest.

References

1. Need AC, Goldstein DB: **Next generation disparities in human genomics: concerns and remedies**. *Trends in Genetics* 2009, **25**:489–494.

2. Popejoy AB, Fullerton SM: **Genomics is failing on diversity**. *Nature* 2016, **538**:161–164.

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

4. Petrovski S, Goldstein DB: **Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine [Internet]**. *Genome Biol.* 2016, **17**:157–3.

5. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, Margulies DM, Loscalzo J, Kohane IS: **Genetic Misdiagnoses and the Potential for Health Disparities**. *N Engl J Med* 2016, **375**:655–665.

6. Drögemöller BI, Wright GE, Niehaus DJ, Koen L, Malan S, Da Silva DM, Hillermann Rebello R, La Grange AM, Venter M, Warnich L: **Characterization of the genetic profile of CYP2C19in two South African populations**. *Pharmacogenomics* 2010, **11**:1095–1103.

7. Hindorff LA, Bonham VL, Brody LC, Ginoza MEC, Hutter CM, Manolio TA, Green ED: **Prioritizing diversity in human genomics research**. *Nature Publishing Group* 2017, doi:10.1038/nrg.2017.89.

   - Reviews NIH NHGRI efforts to prioritize diversity in human genomics research.

8. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al.: **Analysis of protein-coding genetic variation in 60,706 humans**. *Nature* 2016, **536**:285–291.

9. Albrechtsen A, Nielsen FC, Nielsen R: **Ascertainment biases in SNP chips affect measures of population divergence.** *Mol Biol Evol* 2010, **27**:2534–2547.

10. Martin AR, Tse G, Bustamante CD, Kenny EE: **Imputation-based assessment of next generation rare exome variant arrays.** *Pac Symp Biocomput* 2014, doi:10.1534/g3.113.007161.

11. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE: **Human Demographic History Impacts Genetic**

Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* 2017, **100**:635–649.

- Demonstrates that genetic risk prediction from Eurocentric studies is expected to generalize poorly to non-European populations.

12. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al.: **A global reference for human genetic variation.** *Nature* 2015, **526**:68–74.

13. Wojcik GL, Fuchsberger C, Taliun D, Welch R, Martin AR, Shringarpure S, Carlson CS, Abecasis G, Kang HM, Boehnke M, et al.: **Imputation aware tag SNP selection to improve power for multi-ethnic association studies**. 2017, doi:10.1101/105551.

14. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, et al.: **A reference panel of 64,976 haplotypes for genotype imputation**. *Nat Genet* 2016, **48**:1279–1283.

15. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, Karthikeyan S, Iles L, Pollard MO, Choudhury A, et al.: **The African Genome Variation Project shapes medical genetics in Africa.** *Nature* 2015, **517**:327–332.

16. Langan M: **Normative Power Europe and the Moral Economy of Africa–EU Ties: A Conceptual Reorientation of "Normative Power."** *New Political Economy* 2012, **17**:243–270.

17. Alden C: **China in Africa**. *Survival* 2005, **47**:147–164.

18. Dahdouh-Guebas F, Ahimbisibwe J, Van Moll R, Koedam N: **Neo-colonial science by the most industrialised upon the least developed countries in peer-reviewed publishing**. *Scientometrics* 2003, **56**:329–343.

19. Serwadda D, Ndebele P, Grabowski MK, Bajunirwe F, Wanyenze RK: **Open data sharing and the Global South—Who benefits?** *Science* 2018, **359**:642–643.

20. de Vries J, Pepper M: **Genomic sovereignty and the African promise: mining the African genome for the benefit of Africa**. *J Med Ethics* 2012, **38**:474–478.

21. Munung NS, Mayosi BM, de Vries J: **Equity in international health research collaborations in Africa: Perceptions and expectations of African researchers**. *PLoS ONE* 2017, **12**:e0186237–17.

- Collated African perspectives on genomics research equity in international collaborations.

22. Mulder NJ, Adebiyi E, Adebiyi M, Adeyemi S, Ahmed A, Ahmed R, Akanle B, Alibi M, Armstrong DL, Aron S, et al.: **Development of Bioinformatics Infrastructure for Genomics Research**. *Global Heart* 2017, **12**:91–98.

- Describes the development and impact of key bioinformatics infrastructure as part of H3Africa capacity building.

23. Ochola LI, Gitau E: **Challenges in retaining research scientists beyond the doctoral level in Kenya.** *PLoS Negl Trop Dis* 2009, **3**:e345.

24. Adebamowo SN, Francis V, Tambo E, Diallo SH, Landouré G, Nembaware V, Dareng E, Muhamed B, Odutola M, Akeredolu T, et al.: **Implementation of genomics research in Africa: challenges and recommendations.** *Glob Health Action* 2018, **11**:1419033.

25. **A welcome framework for research in Africa.** *Nature* 2018, **556**:274.

- An important ethics guideline developed through in-depth conversations with African researchers and ethics review boards to empower African researchers and communities by suggesting rules of engagement for international genomics partnerships.

26. Wonkam A, Mayosi BM: **Genomic medicine in Africa: promise, problems and prospects**. *Genome Med* 2014, **6**:11–3.

27. Uren C, Möller M, Helden PD, Henn BM, Hoal EG: **Population structure and infectious disease risk in southern Africa**. *Molecular Genetics and Genomics* 2017, **292**:499–509.

28. de Vries J, Bull SJ, Doumbo O, Ibrahim M, Mercereau-Puijalon O, Kwiatkowski D, Parker M: **Ethical issues in human genomics research in developing countries.** *BMC Med Ethics* 2011, **12**:5.

29. Coles E, Mensah GA: **Geography of Genetics and Genomics Research Funding in Africa**. *Global Heart* 2017, **12**:173–176.

- Breaks down international research funding and collaborations in Africa

30. H3Africa Consortium, Rotimi C, Abayomi A, Abimiku A, Adabayeri VM, Adebamowo C, Adebiyi E, Ademola AD, Adeyemo A, Adu D, et al.: **Research capacity. Enabling the genomic revolution in Africa.** *Science* 2014, **344**:1346–1348.

31. de Vries J, Tindana P, Littler K, Ramsay M, Rotimi C, Abayomi A, Mulder N, Mayosi BM: **The H3Africa policy framework: negotiating fairness in genomics**. *Trends in Genetics* 2015, **31**:117–119.

32. Tindana P, Molyneux CS, Bull S, Parker M: **Ethical issues in the export, storage and reuse of human biological samples in biomedical research: perspectives of key stakeholders in Ghana and Kenya**. *BMC Med Ethics* 2014, **15**:585–11.

33. Nordling L: **San people of Africa draft code of ethics for researchers**. *Science* 2017, doi:10.1126/science.aal0933.

34. Masiye F, Mayosi B, de Vries J: **"I passed the test!" Evidence of diagnostic misconception in the recruitment of population controls for an H3Africa genomic study in Cape Town, South Africa**. 2017, doi:10.1186/s12910-017-0175-z.

- Discusses consent issues and diagnostic misconceptions in resource-limited settings in Africa.

35. Rotimi CN, Marshall PA: **Tailoring the process of informed consent in genetic and genomic research.** *Genome Med* 2010, **2**:20.

36. Chennells R, Steenkamp A: *International Genomics Research Involving the San People*. Springer International Publishing; 2017.

37. Henn BM, Cavalli-Sforza LL, Feldman MW: **The great human expansion**. *Proc Natl Acad Sci USA* 2012, **109**:17758–17764.

38. Pickrell JK, Patterson N, Loh PR, Lipson M, Berger B, Stoneking M, Pakendorf B, Reich D: **Ancient west Eurasian ancestry in southern and eastern Africa**. *Proc Natl Acad Sci USA* 2014, **111**:2632–2637.

39. Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, Harris RS, Petersen DC, Zhao F, Qi J, et al.: **Complete Khoisan and Bantu genomes from southern Africa**. *Nature* 2010, **463**:943–947.

40. Schlebusch CM, Skoglund P, Sjodin P, Gattepaille LM, Hernandez D, Jay F, Li S, De Jongh M, Singleton A, Blum MGB, et al.: **Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History**. *Science* 2012, **338**:374–379.

41. Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al.: **Genomic Ancestry of North Africans Supports Back-to-Africa Migrations**. *PLoS Genet* 2012, **8**:e1002397.

42. Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodríguez-Botigué L, Ramachandran S, Hon L, Brisbin A, et al.: **Hunter-gatherer genomic diversity suggests a southern African origin for modern humans.** *Proc. Natl. Acad. Sci. U.S.A.* 2011, **108**:5154–5162.

43. Campbell MC, Tishkoff SA: **The Evolution of Human Genetic and Phenotypic Variation in Africa**. *Current Biology* 2010, **20**:R166–R173.

44. Uren C, Kim M, Martin AR, Bobo D, Gignoux CR, van Helden PD, Möller M, Hoal EG, Henn BM: **Fine-Scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries**. *Genetics* 2016, **204**:303–314.

- This study used genotype data from 22 southern African populations to assess fine-scale population structure, which was shown to have a complex relationship with geography, ecology, language, and subsistence strategy.

45. Choudhury A, Ramsay M, Hazelhurst S, Aron S, Bardien S, Botha G, Chimusa ER, Christoffels A, Gamieldien J, Sefid-Dashti MJ, et al.: **Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans**. *Nature Communications* 2017, doi:10.1038/s41467-017-00663-9.

- A pilot study of deep whole genome sequencing of 24 southern African individuals exhibiting complex, region-specific population structure. This study was led and funded entirely by African investigators and an African government, respectively.

46. Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Güldemann T, Kure B, Mpoloka SW, Nakagawa H, Naumann C, et al.: **The genetic prehistory of southern Africa**. *Nature Communications* 2012, **3**:1143.

47. Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, Bustamante CD, Kenny EE, Williams SM, Aldrich MC, et al.: **The Great Migration and African-American Genomic Diversity**. *PLoS Genet* 2016, **12**:e1006059.

48. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, Vergara C, Torgerson DG, Pino-Yanes M, Shringarpure SS, et al.: **A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome**. *Nature Communications* 2016, **7**:12522.

49. Pagani L, Schiffels S, Gurdasani D, Danecek P, Scally A, Chen Y, Xue Y, Haber M, Ekong R, Oljira T, et al.: **Tracing the Route of Modern Humans out of Africa by Using 225 Human Genome Sequences from Ethiopians and Egyptians**. *The American Journal of Human Genetics* 2015, **96**:986–991.

50. Jagadeesan A, Gunnarsdóttir ED, Ebenesersdóttir SS, Guðmundsdóttir VB, Thordardottir EL, Einarsdóttir MS, Jónsson H, Dugoujon J-M, Fortes-Lima C, Migot-Nabias F, et al.: **Reconstructing an African haploid genome from the 18th century.** *Nature Publishing Group* 2018, **50**:199–205.

- Reconstructed the genome of an African descent individual born in the Caribbean in 1784 who migrated to Iceland in 1802 by genotyping hundreds of

his descendants and inferring shared segments from successively smaller haplotype fragments over time.

51. Hublin J-J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, Bergmann I, Le Cabec A, Benazzi S, Harvati K, et al.: **New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens**. *Nature* 2017, **546**:289–292.

52. Stringer C, Galway-Witham J: **Palaeoanthropology: On the origin of our species.** *Nature* 2017, **546**:212–214.

53. Stringer C, Galway-Witham J: **When did modern humans leave Africa?** *Science* 2018, **359**:389–390.

54. Hershkovitz I, Weber GW, Quam R, Duval M, Grün R, Kinsley L, Ayalon A, Bar-Matthews M, Valladas H, Mercier N, et al.: **The earliest modern humans outside Africa.** *Science* 2018, **359**:456–459.

55. Hansen HB, Damgaard PB, Margaryan A, Stenderup J, Lynnerup N, Willerslev E, Allentoft ME: **Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum**. *PLoS ONE* 2017, **12**:e0170940–18.

56. Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpaslan-Roodenberg S, Gerritsen F, Moiseyev V, Gromov A, Raczky P, et al.: **Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone**. *PLoS ONE* 2015, **10**:e0129102–13.

57. Llorente MG, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, Curtis MC, Stock JT, Coltorti M, Pieruccini P, et al.: **Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa**. *Science* 2015, **350**:820–822.

58. Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, Munters AR, Vicente M, Steyn M, Soodyall H, et al.: **Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago**. *Science* 2017, **358**:652–655.

- Along with [59], one of the two first ancient DNA studies in Africa, which sequenced three ~2,000 year old hunter-gatherers to identify deep divergence and widespread admixture in Africa.

59. Skoglund P, Thompson JC, Prendergast ME, Mittnik A, Sirak K, Hajdinjak M, Salie T, Rohland N, Mallick S, Peltzer A, et al.: **Reconstructing Prehistoric African Population Structure**. *Cell* 2017, **171**:59–71.e21.

- Along with [58], one of the two first ancient DNA studies in Africa, which sequenced 16 ancient south-central and eastern African genomes up to ~8100 years old and showed that San ancestry was widespread in eastern Africa prior to the Bantu expansion.

60. Martin AR, Lin M, Granka JM, Myrick JW, Liu X, Sockell A, Atkinson EG, Werely CJ, Möller M, Sandhu MS, et al.: **An Unexpectedly Complex Architecture for Skin Pigmentation in Africans**. *Cell* 2017, **171**:1340–1353.e14.

- Along with [63], one of the two first studies of skin pigmentation in continental Africa, which showed that skin pigmentation is more polygenic in KhoeSan populations than described elsewhere previously.

61. Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, Campos J, Araújo II, Anderson TM, Vilhjálmsson BJ, et al.: **Genetic Architecture of Skin and Eye Color in an African-European Admixed Population**. *PLoS Genet* 2013, **9**:e1003372–15.

62. Rotimi CN, Tekola-Ayele F, Baker JL, Shriner D: **The African diaspora: history, adaptation and health.** *Current Opinion in Genetics & Development* 2016, **41**:77–84.

63. Crawford NG, Kelly DE, Hansen MEB, Beltrame MH, Fan S, Bowman SL, Jewett E, Ranciaro A, Thompson S, Lo Y, et al.: **Loci associated with skin pigmentation identified in African populations**. *Science* 2017, **358**:eaan8433–26.

- Along with [60], one of the two first studies of skin pigmentation in continental Africa, which showed functional evidence of a role for *MFSD12* in pigmentation.

64. Shriner D, Rotimi CN: **Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase.** *Am. J. Hum. Genet.* 2018, **102**:547–556.

- Analyzed 1000 Genomes Project data from 137 African sickle cell carriers to pinpoint the single origin of this mutation, most likely to the Green Sahara ~7,300 years ago.

65. Busby GB, Band G, Si Le Q, Jallow M, Bougama E, Mangano VD, Amenga-Etego LN, Enimil A, Apinjoh T, Ndila CM, et al.: **Admixture into and within sub-Saharan Africa.** *eLife* 2016, **5**: 56.

- Analyzed haplotype sharing across among 48 sub-Saharan African populations from the MalariaGEN consortium and publicly available data, which showed widespread admixture outside individuals' current geographic region and across ethno-linguistic groups within the last 4,000 years.

66. Altshuler DM, Gibbs RA, Dermitzakis E, Peltonen L, Dermitzakis E, Bonnen PE, de Bakker PIW, Deloukas P, Gabriel SB, Gwilliam R, et al.: **Integrating common and rare genetic variation in diverse human populations**. *Nature* 2010, **467**:52–58.

67. Mathieson I, McVean G: **Demography and the Age of Rare Variants**. *PLoS Genet* 2014, **10**:e1004528.

68. Mathieson I, McVean G: **Differential confounding of rare and common variants in spatially structured populations**. *Nat Genet* 2012, **44**:243–246.

69. Pe'er I, Yelensky R, Altshuler D, Daly MJ: **Estimation of the multiple testing burden for genomewide association studies of nearly all common variants.** *Genet. Epidemiol.* 2008, **32**:381–385.

70. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE, Baker DG, Beckham JC, Bierut LJ, Bisson J, et al.: **Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability.** *Molecular Psychiatry* 2017, doi:10.1038/mp.2017.77.

71. Browning SR, Browning BL: **Population Structure Can Inflate SNP-Based Heritability Estimates**. *The American Journal of Human Genetics* 2011, **89**:191–193.

72. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM: **LD Score regression distinguishes confounding from polygenicity in genome-wide association studies**. *Nat Genet* 2015, **47**:291–295.

73. Park DS, Brown B, Eng C, Huntsman S, Hu D, Torgerson DG, Burchard EG, Zaitlen N: **Adapt-Mix: learning local genetic correlation structure improves summary statistics-based analyses.** *Bioinformatics* 2015, **31**:i181–9.

74. Lloyd-Jones LR, Robinson MR, Moser G, Zeng J, Beleza S, Barsh GS, Tang H, Visscher PM: **Inference on the Genetic Basis of Eye and Skin Color in an Admixed Population via Bayesian Linear Mixed Models**. *Genetics* 2017, **206**:1113–1126.

- Applied Bayesian linear mixed models to an admixed population in Cape Verde, an island nation of West Africa, to account for population stratification, cryptic relatedness, and increase statistical power. These results have important implications for GWAS in structured populations.

75. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM: **Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model**. *PLoS Genet* 2015, **11**:e1004969–22.

76. Zhou X, Carbonetto P, Stephens M: **Polygenic modeling with bayesian sparse linear mixed models.** *PLoS Genet* 2013, **9**:e1003264.

77. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, et al.: **Efficient Bayesian mixed-**

model analysis increases association power in large cohorts. *Nat Genet* 2015, **47**:284–290.

78. Malaria Genomic Epidemiology Network: **A novel locus of resistance to severe malaria in a region of ancient balancing selection**. *Nature* 2015, **526**:253–257.

79. Dalvie S, Koen N, Duncan L, Abbo C, Akena D, Atwoli L, Chiliza B, Donald KA, Kinyanda E, Lochner C, et al.: **Large Scale Genetic Research on Neuropsychiatric Disorders in African Populations is Needed**. *EBIOM* 2015, **2**:1259–1261.

80. Wu Y, Waite LL, Jackson AU, Sheu WH-H, Buyske S, Absher D, Arnett DK, Boerwinkle E, Bonnycastle LL, Carty CL, et al.: **Trans-Ethnic Fine-Mapping of Lipid Loci Identifies Population-Specific Signals and Allelic Heterogeneity That Increases the Trait Variance Explained**. *PLoS Genet* 2013, **9**:e1003379.

81. Rotimi CN, Bentley AR, Doumatey AP, Chen G, Shriner D, Adeyemo A: **The genomic landscape of African populations in health and disease**. *Human Molecular Genetics* 2017, **26**:R225–R236.

82. Helgason A, Pálsson S, Thorleifsson G, Grant SFA, Emilsson V, Gunnarsdottir S, Adeyemo A, Chen Y, Chen G, Reynisdottir I, et al.: **Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution**. *Nat Genet* 2007, **39**:218–225.

83. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Genovese G, Loh P-R, Bhatia G, Do R, Hayeck T, et al.: **Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores**. *The American Journal of Human Genetics* 2015, **97**:576–592.

84. Ware EB, Schmitz LL, Faul JD, Gard A, Mitchell C, Smith JA, Zhao W, Weir D, Kardia SL: **Heterogeneity in polygenic scores for common human traits**. 2017, doi:10.1101/106062.

85. Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, et al.: **Genome-wide association study identifies 112 new loci for body mass index in the Japanese population**. *Nat Genet* 2017, **49**:1458–1467.

86. Li Z, Chen J, Yu H, He L, Xu Y, Zhang D, Yi Q, Li C, Li X, Shen J, et al.: **Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia**. *Nature Publishing Group* 2017, **49**:1576–1583.

87. Rappoport N, Paik H, Oskotsky B, Tor R, Ziv E, Zaitlen N, Butte AJ: **Creating ethnicity-specific reference intervals for lab tests from EHR data**. *bioRxiv* 2017, doi:10.1101/213892.

 doi:10.20944/preprints201806.0201.v1

88. Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H: **Genome-wide patterns of population structure and admixture among Hispanic/Latino populations**. *Proc Natl Acad Sci USA* 2010, **107**:8954–8961.

89. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, et al.: **Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation**. *Science* 2008, **319**:1100–1104.