Comment

ACEGID

Christian Happi at Redeemer's University in Ede, Nigeria, plans to sequence human genomes.

Sequence three million genomes across Africa

Ambroise Wonkam

Capture the full scope of variation to improve health care, equity and medical research globally.

Two decades after the completion of the Human Genome Project (HGP), there is still much to do to ensure that genomics works for the global public good. The focus on populations from high-income countries has come at the cost of understanding health and disease that might benefit the world. Less than 2% of human genomes analysed so far have been those of African people[1], despite the fact that Africa, where humans originated, contains more genetic diversity than any other continent. Too little of the knowledge and applications from genomics have benefited the global south because of inequalities in health-care systems, a small local research workforce and lack of funding.

The African Society of Human Genetics (AfSHG), which I currently lead, was established in 2003 to help address disparities, improve education, enhance networking and build research capacity in Africa (www.afshg.org). Despite recent progress and investments, too much of the genomic research done in Africa has been driven by European and American investigators[2]. Why is this a problem? Their priorities could be detached from what people on the continent need and want. Testing new treatments are more likely to yield high-profile papers that advance academic careers, but testing more-effective delivery methods for existing treatments is often more likely to save lives and ease suffering.

The reference genome sequences built from the HGP are missing many variants from African ancestral genomes. A 2019 study estimated that a genome representing the DNA of the African population would have about 10% more DNA than the current reference[3]. Last year, analyses of whole-genome sequences of just 426 people across 50 ethnolinguistic groups in Africa revealed more than 3 million variants that were previously unknown[4].

Those variants were identified as part of a US$180-million, 10-year initiative, the Human Heredity and Health in Africa (H3Africa) consortium (https://h3africa.org). It supports institutes across 30 African countries and is facilitated by the AfSHG in collaboration with the US National Institutes of Health (NIH) and the UK biomedical funder Wellcome. H3Africa is now winding up, so it is time to think about what should happen next.

A rough estimate for capturing the full scope of Africa's genetic variation would require sequencing some three million individuals, carefully selected across Africa to cover ethnolinguistic, regional and other groups. Therefore, we aim to start such a project, called the Three Million African Genomes (3MAG), which would build capacity on the continent — in genomics research and its applications, and governance. The findings would bring benefits worldwide, including some that are hard to anticipate. In a similar way, much knowledge put to use during the COVID-19 pandemic — from public communication to sharing biological samples and data — was hammered out in Ebola outbreaks during the past few years.

The development of 3MAG will probably take around a decade. We estimate core funding would need to be roughly $450 million per year (about $1,500 per participant in total). That would cover setting up and running biorepositories and developing data infrastructure and technology. We plan on sequencing and phenotyping about 300,000 African genomes in the first year.

Those who think this too daring are forgetting the ambition required to launch the HGP. That took more than 13 years, and infused genetic science across all areas of health care. Today, the cost of sequencing a genome is less than $1,000 — building the first draft reference genome cost around $300 million (see go.nature.com/3pfy2kh). Thirty years

ago, at the inception of the HGP, the US National Human Genome Research Institute distributed some 95% of the total NIH funds to human genomics research projects. Last year, that number was 10%, because all the other NIH institutes also support discipline-specific research grounded in genetics[5].

3MAG would aim to sequence enough genomes across Africa to build a representative human reference genome and to establish a pan-African biobank of clinical information and samples. To put this into perspective, the UK Biobank is midway through a 27-month project to sequence 500,000 genomes (see go.nature.com/3ciohcj), and the United Kingdom's population is about 5% of Africa's.

Research benefits

African genomes can reveal genes and variants that contribute to health and disease not found in previous, Eurocentric studies. (That said, certain European populations from a small pool of founders, such as Iceland's, are useful for different reasons: genetic homogeneity can help to reveal environmental factors and single-gene variations that have a strong effect.) Populations of African ancestry are the most genetically diverse in the world. They collectively have more genetic variation and less intermixing with other, non-African populations, which makes it easier to find variants likely to contribute to specific conditions.

For example, variants in the gene *PCSK9* are extremely rare in Europeans (less than 0.1%) but relatively common in African Americans (around 2%). The gene is correlated with much lower levels of certain blood lipids[6], a finding that has led to at least one new medication (for example, evolucomab) for the blood condition dyslipidaemia, associated with heart attack and stroke, which affects populations worldwide. A study of about 900 Africans of Xhosa ancestry with schizophrenia matched 900 Africans without it and found many rare mutations that contribute to the disorder, along with mechanistic insights[7]. A 2016 study in a Swedish population identified many of these same mutations — but required a sample more than four times the size[8].

Studies in African genomes will also help to correct injustice. Estimates of genetic risk scores for people of African descent that predict, say, the likelihood of cardiomyopathies or schizophrenia can be unreliable or even misleading using tools that work well in Europeans[9]. To promote discovery and produce reliable clinical tools, genotyping and analysis must be re-optimized using genomes from more populations.

Work on hearing impairment[10] and sickle-cell disease (SCD)[11] (the focus of my research) exemplifies many benefits that could come from studying diseases caused by a single gene. People with two copies of the sickle mutation have deformed, banana-shaped red blood cells. These clump together in blood vessels, stopping the distribution of oxygen to tissues. People with SCD often live into their fifties in high-income countries. In many poorer nations, they die in childhood, from bacterial infection, anaemia, lung diseases, stroke or other complications.

People who have SCD are much healthier if they carry variants in other genes that, say, prolong the production of fetal haemoglobin or, ironically, that contribute to a form of anaemia called α-thalassaemia. By contrast, variants in the gene *APOL1* increase susceptibility to kidney disease[12].

"How can expansive genomic sequencing be justified when people still die of malnutrition, malaria and HIV?"

Most genetic modifiers of SCD have been identified from studies in Europe and the United States, usually using gene chips developed to find variants common in European populations[13]. Of the approximately 300,000 babies born with the mutation each year, about 75% are in Africa, where genetic diagnoses and insights play almost no part in medical care. A multi-centre, well-coordinated longitudinal study across the continent could discover numerous variants. This would help to predict the disease course, suggest new routes for therapies[11] (perhaps including gene editing), provide better advice for parents when prenatal genetic testing is available, and help people to manage the disease. Such a study could also become a model for understanding how genomic variants influence other monogenic diseases.

Three priorities

To sequence three million African genomes will require support from African governments, academics and international organizations.

Collaborations. The most straightforward starting point is to establish more collaborations across the globe and across African countries, including both academic and corporate research. We can leverage technologies, workflows and best practices from H3Africa and a host of other projects (see 'Paving the way' and page 220). For instance, 54gene, a company based in Lagos, Nigeria, is setting up facilities to sequence the genomes of 100,000 Nigerians — and reports that Silicon Valley investors have put in $4.5 million to establish a biobank (https://54gene.com).

Workforce. Another requirement is to train a medical and technical research workforce, emphasizing human genetics, informatics and computer science. Ideally, this will happen by establishing graduate studies in African universities, and genetic-medicine centres of excellence in African health-care facilities.

Governance. Most challenging will be developing clinical, governmental and social infrastructure to embrace the diverse cultures and nations across Africa. Knowledge that emerges from 3MAG and associated biobanks will have profound ethical implications. Currently, there is a lack of research on what people in Africa

BRENT STIRTON/REPORTAGE BY GETTY

Xhosa women in Qunu, South Africa. Rare mutations exist in people with Xhosa ancestry.

PAVING THE WAY: THREE MILLION AFRICAN GENOMES CAN DRAW FROM EXISTING PROJECTS

Programmes	Launch (duration)	Funding	Purpose	Products
Human Heredity and Health in Africa (H3Africa)	2011 (10 years)	US$180 million, US National Institutes of Health (NIH), Wellcome (partnership with the African Society of Human Genetics)	Builds collaborations and genetics research led by African scientists for Africans.	Genome-wide and sequencing data on 79,254 individuals across 30 African countries. Three Biorepositories (Nigeria, Uganda and South Africa).
Malaria Genomic Epidemiology Network (MalariaGEN)	2005 (ongoing)	Wellcome, UK Medical Research Council, Bill & Melinda Gates Foundation, NIH	Connects genomics researchers with clinicians in malaria-endemic countries.	Genome-wide data on 17,000 individuals from 39 countries (12 in Africa).
Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa)	2020 (5 years)	$58 million, NIH	Advances data science in Africa to benefit clinical care, public health and research.	Pending
Trans-Omics for Precision Medicine (TOPMed)	2014 (ongoing)	NIH	Integrates sequencing, molecular and clinical data to individualize treatments.	>90,000 whole genomes sequenced and genome-wide data for 155,000 individuals from >80 different studies; 47,020 participants of African ancestry.
Genome Aggregation Database (gnomAD)	2017 (ongoing)	Broad Institute	Aggregates and harmonizes data from large sequencing projects in a public portal	>140,000 exomes (protein-encoding regions) and genomes from a variety of sequencing projects; 20,744 African/ African American participants.
UK Biobank	2006 (ongoing)	$332.3 million (multiple UK agencies)	Provides biomedical and genetic data from many individuals for medical research.	Clinical and genome-wide data from 500,000 participants, 8,066 Black African; 200,000 participant exomes.

think about certain issues, including informed consent, community engagement, privacy and confidentiality, and the use of genetic information. Nor is there knowledge of their views on the governance of biorepositories, benefit sharing and return of research results, or about their fears of exploitation in research collaborations and commercialization. These deficits caused controversy when the United Kingdom's Wellcome Sanger Institute planned to commercialize a gene-profiling kit based on insights drawn from African genomes (see go.nature.com/3r7elep). South Africa's legislature has been discussing the Protection of Personal Information Act for several years. But formal ethical, legal and social implication frameworks to cover these issues are needed urgently. Coordinated efforts should be established to build equity-oriented genomics research. These should draw on theories of global justice as well as on African-based concepts such as Ubuntu, which loosely translates as community spirit.

First steps

To set out comprehensive goals and plans, members from the African Society of Human Genetics, the African Academy of Sciences and the H3Africa consortium must work with academics, scientists, professional societies, government representatives, health workers, patient advocates and more. As with H3Africa, these goals could be refined through regular meetings of working groups (such as on medical genetics, training, biorepository, public engagement and sustainability), interspersed with broader gatherings.

Governments will need to commit to building data centres, developing genetic-medicine services and creating academic programmes. They will also need to facilitate public–private partnerships for research, development and translation to the clinic, establishing legal and ethical rules around personal data and consent, and more. It will be essential to have a committee within the World Health Organization or African Union with authority to aid and coordinate this infrastructure to streamline cross-country studies. Tax and market incentives should encourage the private sector to direct genomics research towards neglected diseases.

The project raises an obvious question: how can expansive genomic sequencing be justified when people still die of malnutrition, malaria and HIV?

I think that 3MAG will improve capacity in a whole range of biomedical disciplines that will equip Africa to tackle public-health challenges more equitably, and yield knowledge that could benefit vulnerable populations. Indeed, there is some evidence that effects and treatments of severe malnutrition are influenced by genetic variants[14]. And the HGP accelerated medicine in general, including molecular technology to diagnose HIV and tuberculosis that are widely used in Africa. It also informed preventive and therapeutic approaches for HIV.

3MAG could expand and extend these kinds of benefit. Genetic variants that influence the metabolism of HIV drugs have already been found in high-income populations. Up to 47% of African and African American populations carry a variant called *CYP2B6*6* that is associated with severe side effects to the HIV drug efavirenz[15], raising the likelihood that people skip doses and the emergence of viral resistance. Pharmacogenetic research should also be tasked further to match people with the most effective therapies in Africa.

3MAG must serve the populations it will study and bring better basic health care. What good is identifying variants for breast cancer or cardiovascular disease in people who can't get mammograms or checks for high blood pressure, and who probably won't have access to medical treatments? Ultimately, 3MAG, like H3Africa, must serve global genetic medicine, and African populations, well beyond genetics.

The author

Ambroise Wonkam is professor of medical genetics and deputy dean of research at the Faculty of Health Sciences, University of Cape Town, South Africa; president of the African Society of Human Genetics; and co-chair of the H3Africa consortium.
e-mail: ambroise.wonkam@uct.ac.za

1. Sirugo, G., Williams, S. M. & Tishkoff, S. A. *Cell* **177**, 26–31 (2019).
2. Adedokun, B. O., Olopade, C. O. & Olopade, O. I. *Glob. Health Action* **9**, 31026 (2016).
3. Sherman, R. M. *et al. Nature Genet.* **51**, 30–35 (2019).
4. Choudhury, A. *et al. Nature* **586**, 741–748 (2020).
5. Green, E. D. *et al. Nature* **586**, 683–692 (2020).
6. Cohen, J. *et al. Nature Genet.* **37**, 161–165 (2005).
7. Gulsuner, S. *et al. Science* **367**, 569–573 (2020).
8. Genovese, G. *et al. Nature Neurosci.* **19**, 1433–1441 (2016).
9. Curtis, D. *Psychiatr. Genet.* **28**, 85–89 (2018).
10. Wonkam, A., Manyisa, N., Bope, C. D., Dandara, C. & Chimusa, E. R. *Hum. Mol. Genet.* **29**, 3729–3743 (2020).
11. Wonkam, A. *et al. Clin. Transl. Med.* **10**, e152 (2020).
12. Geard, A. *et al. Br. J. Haematol.* **178**, 629–639 (2017).
13. Menzel, S. *et al. Nature Genet.* **39**, 1197–1199 (2007).
14. Schulze, K. V. *et al. Nature Commun.* **10**, 5791 (2019).
15. Swart, M. *et al. Pharmacogenet. Genomics* **23**, 415–427 (2013).