



Editorial

African partnerships through the H3Africa Consortium bring a genomic dimension to longitudinal population studies on the continent

Michèle Ramsay^{1*} and Osman Sankoh^{2,3}, as members of the AWI-Gen study and the H3Africa Consortium

¹Sydney Brenner Institute for Molecular Bioscience and Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, ²INDEPTH Network, Kanda, Accra, Ghana and ³Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

*Corresponding author. E-mail: michele.ramsay@wits.ac.za

A health and epidemiological transition is enveloping the African continent from the southern and northern regions where the prevalence of obesity has rapidly increased over the past three decades.¹ In the wake of the transition to increased urbanization follow increased rates of hypertension, stroke and type 2 diabetes (T2D). Despite the widespread HIV, TB and malaria epidemics, age-standardized mortality for non-communicable diseases (the probability of dying from one of the four main NCDs—CVD, cancer, chronic respiratory disease and diabetes) between the ages of 30 and 70 years (comparable estimates for 2012) is over 25% in South Africa compared with less than 15% in North America and Europe.²

Good health-related epidemiological data from most African populations are sparse. When accessing global data on non-communicable diseases, it becomes clear that many African countries have no data; in some there is sporadic reporting on specific variables and then there are pockets of excellent data, albeit usually on smaller cohorts, or only in specific regions. For this reason, African health data are often modelled and predictions are based on models that are supported with little and sub-optimal information. This highlights an urgent need to support more systematic approaches to collecting epidemiological data in Africa.

The world is firmly in the post-genomic era and research studies that examine both genetic and environmental contributions to common and complex phenotypes now

require extensive genome data on large cohorts. Considerations of funding, infrastructure, a critical mass of researchers, analytical skills and political will have augured against this research being extensively performed in low- and middle-income countries (LMICs) and more specifically in Africa. As a consequence, much of our knowledge is based on studies involving populations of European and Asian origin. There is ample evidence that ethno-specific genetic differences may have a profound impact on health and should be considered in both public health genomics and precision medicine. The long-term aim of the genomic research is to translate the findings into a clinical application with individual utility for improving health and well-being across a wide spectrum of communities. The global experience has highlighted that this is not only an African challenge but is globally complex, even in highly resourced regions.³

The Human Heredity and Health in Africa Consortium, H3Africa, was born out of discussions highlighting the need for large genomics studies in Africa and with the vision to harness genomic technologies toward improving the health of Africans [<http://h3africa.org>].⁴ H3Africa was formally launched in October 2012 in Addis Ababa and is entering its 4th year. A marker paper was published in *Science* in June 2014, describing the consortium, its objectives and proposed outcomes.⁴ The consortium has grown and now includes eight collaborative centres, seven research projects, six ethics projects, three H3Africa

biorepository sites and a pan-African bioinformatics network. The research is performed by 26 research groups in 27 African countries and includes over 500 investigators who will be studying over 75 000 participants in Africa. It is time to reflect on the past 3 years and we share our observations.

The following challenges led to a longer than anticipated period for the consortium to start most of its research programmes: the enormous diversity across the Consortium; the relative inexperience of many groups in terms of team science; little experience in managing large grants; lack of access to sustainable infrastructure; and the lack of experience of many of the ethics review committees in evaluations genomics research. Having overcome many of the hurdles, H3Africa now has a set of excellent documents to guide good ethical practice (more specifically in terms of community engagement and broad informed consent) and to govern the processes of data and biospecimen sharing and access. These documents emanated from working groups with representation of members from different projects and were established early in the project. The members meet regularly and have been productive.

As an example, the pan-African bioinformatics network, H3ABioNet, has played a key role in capacity development. It has supported infrastructure development and skills building across the consortium through frequent workshops and meetings and has implemented an accreditation process for groups to enrich bioinformatics skills in genome-wide association studies and the analysis of next generation sequencing data.^{5,6} H3ABioNet is responsive to the needs of the consortium and will assist where necessary with data quality control (QC) and analysis. Genotype and phenotype data from the H3Africa research studies will be stored in the European Genome Phenome Archive (EGA) and will be accessible through a managed access process. Applications for access to data and biological samples will be evaluated by an independent H3Africa Data and Biospecimen Access Committee (DBAC), comprising mostly African members with expertise across relevant disciplines, who will consider the informed consent and stipulations of the original ethics committees who approved the studies, when granting access.

Our personal experience of the H3Africa initiative is through a fruitful partnership between the University of the Witwatersrand (Wits) in Johannesburg, including the Sydney Brenner Institute for Molecular Bioscience [<http://www.wits.ac.za/research/sbimb/>.html] and selected African centres from the International Network for the Demographic Evaluation of Populations and their Health in low- and middle-income countries (INDEPTH) [www.indepth-network.org].⁷ We have joined forces to establish the Africa Wits-INDEPTH partnership for genomic studies

on cardiometabolic diseases (AWI-Gen). This collaborative centre aims to build a research resource to understand the environmental and genomic contributions to body composition, cardiovascular diseases and metabolic diseases across African communities. It is a population cross-sectional study that draws its study participants from five INDEPTH health and demographic surveillance system (HDSS) sites across the African continent, ensuring a mix of participants from western, eastern and southern African communities from both rural and urban settings. The participating HDSSs are located in Nairobi (Kenya), Nanoro (Burkina Faso), Navrongo (Ghana), Agincourt (South Africa) and Dikgale (South Africa) and the study also includes participants from Soweto, South Africa (Development Pathways for Health Research Unit located at Chris Hani Baragwanath Hospital). The design of the AWI-Gen study leverages on established infrastructure, trained fieldworkers, long-standing community engagement and detailed longitudinal phenotypic data, focusing on obesity and cardiometabolic health at the INDEPTH Centres and in Soweto. Key strengths are: harmonized phenotyping across sites; building on strong existing longitudinal cohorts; and representation of the geographical and social variability of African populations.

The constellation of resources that are being developed under this partnership, in addition to other collaborations such as the NIH/NIA-funded study on Health and Ageing in Africa: Longitudinal Studies in INDEPTH Communities (HAALSI), provides a rich resource to examine risks for cardiovascular and metabolic diseases. More specifically, it is an ideal opportunity to lay a genomic foundation to the investigations of the effects of obesity and fat distribution on the risk for cardiovascular and metabolic diseases, and to develop longitudinal cohorts that could be followed as they age.

Ethical challenges

Being mindful of the unexplored nature of genomic studies on the continent, we have embarked on extensive interactions with ethics review committees and communities in the countries involved in the AWI-Gen study.⁸ Epidemiological and biomedical research involving infectious epidemics and biomarker studies have a long history on the African continent. However, adding an extensive genomic dimension and a request to participants to share their demographic, health, biomedical, anthropometric, social, environmental and genomic data raises questions with a new emphasis. Are we mindful enough of protecting individuals and communities from potential stigmatization? Despite de-identification of participants, is the potential or accidental identification of participants a real threat,

given that each individual has a unique identifiable genome? Have we considered how communities may realistically benefit from this research, given the resource constraints on the continent? These questions have been extensively probed within AWI-Gen, but also more widely in the H3Africa Consortium. A set of policies and guidelines have been developed by H3Africa to enhance good practice and to promote fairness.⁹

An African GWAS array enriched for common variants across multiple African populations

From the proposal stage, we realized the importance of understanding population structure in order to appropriately analyse genomic data. Resources were therefore dedicated to understanding biological heritage, as reflected in the genetic variation, and the context of the diverse and complex ethnolinguistic backgrounds within each of the communities in which we are working. The urban communities are particularly complicated as they represent an interesting convergence of the ethnolinguistic groups of a country and neighbouring regions, as individuals migrate in pursuit of employment and other opportunities. The genetic ancestry and population structure analyses will inform the strategies for the genetic association studies for cardiometabolic diseases. The H3Africa Consortium as a whole has embraced this objective with the aim of developing a single nucleotide variant (SNV) genotyping array which will have representation of common variants in populations across the continent. For AWI-Gen, we will use this genotyping array to produce a large dataset from which to perform genome-wide association studies (GWASs) for a multitude of phenotypes related to body composition and cardiometabolic risk. It will also provide data for complex modelling for genotype-phenotype, genotype-environment and three-way (or more) interactions. Since we will have a wealth of phenotype and environmental exposure data on the same individuals, these data could also be used to develop and validate potential Mendelian randomisation approaches to studying various risk factors in African populations.¹⁰

Synergies across H3Africa

A major objective for the H3Africa Consortium is to develop larger studies across several of its collaborative centres and research projects, building on their existing objectives. Together with AWI-Gen, five additional H3Africa projects involve cardiovascular disease-related research which will generate resources to examine cross-cutting questions where comparable data are being collected.

Given the increase in the prevalence of cardiovascular diseases across the continent, this is timely.¹¹ The H3Africa Consortium does not only provide opportunities for multi-study collaborations and joint data analyses, it also enables the garnering of additional funding to harmonize with other studies. There is enormous scientific merit to building effective synergies between studies. A good example is the AWI-Gen/HAALSI collaboration mentioned above. In addition to cardiometabolic diseases, this combined study of about 5000 participants from the Agincourt HDSS in South Africa, aged 40 years and older, will also examine other major disorders including dementias, neurological disorders and pulmonary disorders, with both an infectious and a non-infectious aetiology, and the effects of ageing on these processes.

The INDEPTH network which was formally established in 1998 has developed several strategies and policies for data management and sharing, including the INDEPTH Data Repository and its associated data visualization website INDEPTHStats.^{12,13} More recently it has proposed the Comprehensive Health and Epidemiological Surveillance System (CHESS) which should be capable of timely delivery of high-quality disease and pathogen-specific morbidity data, together with overall and cause-specific mortality data. In addition to disease aetiologies and morbidities, the new CHESS will include full risk factor surveillance and will address the full spectrum of the rapidly transitioning burden of disease, including non-communicable diseases and external causes and their associated morbidities. This experience and the resources now available have application in the context of cross-country epidemiological research and will be enriched with studies that include genetics and genomics.

The future of large genomic and epidemiology studies in Africa

H3Africa and its extensive web of investigators and collaborators, as well as additional genomic research on the continent (for example the MalariaGen consortium¹⁴), form the foundation of large cohort studies for exploring genotype and phenotype relationships among Africans, informed by an understanding of environmental context. The impact of HIV, malaria and tuberculosis epidemics and the effects of the infections on non-communicable diseases like cardiovascular, metabolic and pulmonary diseases and cancer, need to be better understood. Africa provides nature's experiment to investigate the clash between the host genomic diversity and exposure to infections, which makes it the continent with the highest per capita burden of disease. In addition to the obvious challenges that face researchers in LMICs, including poor

resources, poor infrastructure and a lack of critical mass of skilled investigators, we face the challenge of working together to identify the critical questions that are tractable, given the rich datasets we are developing. Our hope is that the research will inform our knowledge of the communities and their health, in the context of the transition, across the continent. Desirable outcomes would be greater insight into the planning of future biomedical research and targeted data and biospecimen collection, to ultimately lead to informing public health decision-making and contributing to feasible public health intervention strategies.

Acknowledgements

The AWI-Gen study was jointly planned by MR, OS, Stephen Tollman, Kathleen Kahn, Nigel Crowther, Shane Norris, Scott Hazelhurst, Zane Lombard, Himla Soodyall and Alisha Wade, as well as the Directors of INDEPTH member centres: Catherine Kyobutungi, Marianne Alberts, Halidou Tinto, Abraham Oduro and their teams. The AWI-Gen Collaborative Centre is funded by the NIH (1U54HG006938) as part of the H3Africa Consortium. MR is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology and administered by National Research Foundation of South Africa (NRF). OS is the Executive Director of the INDEPTH Network which receives core support funding from the Hewlett Foundation, Sida/Research Cooperation and Wellcome Trust. This editorial reflects the views of the authors and does not necessarily represent the official views of the National Institutes of Health or the NRF.

Conflict of interest: Osman Sankoh is the Executive Director of the INDEPTH Network. Michele Ramsay and Osman Sankoh are PI and co-PI of the NIH-Funded project AWI-Gen.

References

1. Stevens GA, Singh GM, Lu Y *et al*. National, regional, and global trends in adult overweight and obesity prevalences. *Population Health Metrics* 2012;10:22. P.
2. World Health Organization. *Global Status Report on Noncommunicable Diseases 2014*. Geneva: WHO, 2014.
3. Burke W, Korngiebel DM. Closing the gap between knowledge and clinical application: challenges for genomic translation. *PLoS Genet* 2015;11:e1004978.
4. Consortium HA, Rotimi C, Abayomi A *et al*. Research capacity. *Enabling the genomic revolution in Africa*. *Science* 2014;344:1346–48.
5. Tastan Bishop O, Adebisi EF, Alzohairy AM *et al*. Bioinformatics education - perspectives and challenges out of Africa. *Brief Bioinform* 2015;16:355–64.
6. Budd A, Corpas M, Brazas MD *et al*. A quick guide for building a successful bioinformatics community. *PLoS Comput Biol* 2015;11:e1003972.
7. Sankoh O, Byass P. The INDEPTH Network: filling vital gaps in global epidemiology. *Int J Epidemiol* 2012;41:579–88.
8. Ramsay M, de Vries J, Soodyall H, Norris SA, Sankoh O. Ethical issues in genomic research on the African continent: experiences and challenges to ethics review committees. *Hum Genomics* 2014;8:15.
9. de Vries J, Tindana P, Littler K *et al*. The H3Africa policy framework: negotiating fairness in genomics. *Trends Genet* 2015;31:117–19.
10. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89–98.
11. Owolabi MO, Mensah GA, Kimmel PL *et al*. Understanding the rise in cardiovascular diseases in Africa: harmonising H3Africa genomic epidemiological teams and tools. *Cardiovasc J S Afr* 2014;25:134–36.
12. Sankoh O, Herbst AJ, Juvekar S, Tollman S, Byass P, Tanner M. INDEPTH launches a data repository and INDEPTHStats. *Lancet Glob Health* 2013;1:e69.
13. Mullan Z. The good, the bad, and the neglected. *Lancet Glob Health* 2013;1:e55.
14. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet* 2010;11:149–60.